



**HAL**  
open science

# Catching cellular heterogeneity and differentiation dynamics of normal and pathological airway epithelia through single cell transcriptional profiling

Sandra Ruiz Garcia

► **To cite this version:**

Sandra Ruiz Garcia. Catching cellular heterogeneity and differentiation dynamics of normal and pathological airway epithelia through single cell transcriptional profiling. Molecular biology. Université Côte d'Azur, 2018. English. NNT: 2018AZUR4204 . tel-02108264

**HAL Id: tel-02108264**

**<https://theses.hal.science/tel-02108264>**

Submitted on 24 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT

Appréhender l'hétérogénéité cellulaire et la  
dynamique de différenciation des  
épithéliums des voies aériennes au moyen  
de signatures transcriptionnelles sur cellule  
unique

**Sandra RUIZ GARCIA**

Physiologie génomique des eucaryotes

Présentée en vue de l'obtention  
du grade de docteur en Interactions  
moléculaires et cellulaires

d'Université Côte d'Azur

Dirigée par : Pascal Barbry

Co-encadrée par : Laure-Emmanuelle Zaragosi

Soutenue le : 18-December-2018

Devant le jury, composé de :

**Marc Chanson**, Professeur associé, Université  
de Genève

**Pascal Chanez**, Professeur PU-PH, Aix-  
Marseille Université

**David Garfield**, Chef d'équipe, Université  
Humboldt de Berlin

**Pascal Barbry**, Directeur de recherche,  
Université côte d'azur

**Laure-Emmanuelle Zaragosi**, Chargée de  
recherche, Université côte d'azur

# Appréhender l'hétérogénéité cellulaire et la dynamique de différenciation des épithéliums des voies aériennes au moyen de signatures transcriptionnelles sur cellule unique

Jury :

President du jury :

**Marc Chanson**, Professeur associé, Université de Genève

Rapporteurs :

**Pascal Chanez**, Professeur PU-PH, Aix-Marseille Université

**David Garfield**, Chef d'équipe, Université Humboldt de Berlin

Examineurs :

**Pascal Barbry**, Directeur de recherche, Université côte d'azur, Université Nice-Sophia-Antipolis

**Laure-Emmanuelle Zaragosi**, Chargée de recherche, Université côte d'azur, Université Nice-Sophia-Antipolis

## RESUME

Les voies aériennes humaines sont bordées d'un épithélium pseudostratifié composé principalement de cellules basales et de cellules pyramidales parmi lesquelles figurent les cellules sécrétrices de mucus et les cellules multiciliées. Toutes ces cellules contribuent à la clairance mucociliaire des voies respiratoires. Cet épithélium se régénère lentement dans des conditions homéostatiques, mais il est capable de se régénérer rapidement après agression grâce à des processus de prolifération, de migration, de polarisation et de différenciation.

Chez les patients atteints de maladies respiratoires chroniques telles que la broncho-pneumopathie chronique obstructive, l'asthme ou la mucoviscidose, la réparation tissulaire est souvent défectueuse, caractérisée par une perte de cellules multiciliées et une hyperplasie des cellules sécrétrices, ayant pour conséquence une clairance mucociliaire affectée.

La séquence des événements cellulaires conduisant à un tissu fonctionnel ou remodelé est encore mal décrite. Notre principal objectif a été d'identifier les types cellulaires successifs mis en jeu lors de la régénération tissulaire et les événements moléculaires responsables d'une régénération saine ou pathologique.

Nous avons analysé la composition cellulaire de l'épithélium des voies respiratoires à plusieurs stades de différenciation en utilisant un modèle de culture 3D *in vitro* qui reproduit la composition cellulaire *in vivo*. En appliquant une méthode de transcriptomique sur cellule unique couplée à des méthodes bioinformatiques, nous avons établi les hiérarchies cellulaires permettant de reconstruire les différentes trajectoires cellulaires mises en jeu lors de la régénération de l'épithélium des voies respiratoires humaines. Après avoir confirmé les lignages cellulaires qui ont été précédemment décrits, nous avons découvert une nouvelle trajectoire reliant les cellules sécrétrices de mucus aux cellules multiciliées. Nous avons également caractérisé de nouvelles populations cellulaires et de nouveaux acteurs moléculaires impliqués dans le processus de régénération de l'épithélium des voies respiratoires humaines. Enfin, grâce à ces approches, nous avons mis en évidence des réponses spécifiques de chaque type cellulaire survenant dans des situations pathologiques d'hyperplasie sécrétoire.

Ainsi, nos données, en apportant d'importantes contributions à la compréhension de la dynamique de différenciation de l'épithélium des voies respiratoires humaines, ouvrent de nouvelles voies vers l'identification de cibles thérapeutiques.

## MOTS CLES

Épithélium des voies respiratoires, régénération épithéliale, remodelage des voies respiratoires, transcriptomique sur cellule unique, multiciliogenèse, hyperplasie des cellules sécrétrices de mucus.

# Catching cellular heterogeneity and differentiation dynamics of normal and pathological airway epithelia through single cell transcriptional profilings

Jury:

President of the jury:

**Marc Chanson**, Professeur associé, Université de Genève

Reporters:

**Pascal Chanez**, Professeur PU-PH, Aix-Marseille Université

**David Garfield**, Chef d'équipe, Université Humboldt de Berlin

Reviewers:

**Pascal Barbry**, Directeur de recherche, Université côte d'azur, Université Nice-Sophia-Antipolis

**Laure-Emmanuelle Zaragosi**, Chargée de recherche, Université côte d'azur, Université Nice-Sophia-Antipolis

## **ABSTRACT**

Human airways are lined by a pseudostratified epithelium mainly composed of basal and columnar cells, among these cells we can find multiciliated, secretory cells and goblet cells. All these cells work together in the mucociliary clearance of the airways. This epithelium regenerates slowly under homeostatic conditions but is able to recover quickly after aggressions through proliferation, migration, polarization and differentiation processes.

However, in patients with chronic pulmonary diseases such as chronic obstructive pulmonary disease, asthma or cystic fibrosis, epithelial repair is defective, tissue remodeling occurs, leading to loss of multiciliated cells and goblet cell hyperplasia, impairing correct mucociliary clearance.

The sequence of cellular events leading to a functional or remodelled tissue are still poorly described. Hence, we aim at identifying the successive cell types appearing during tissue regeneration and the molecular events that are responsible for healthy or pathological regeneration.

We have analysed airway epithelial cell composition at several stages of differentiation using an in vitro 3D culture model which reproduces in vivo epithelial cell composition. Applying single cell transcriptomics and computational methods, we have identified cell lineage hierarchies and thus constructed a comprehensive cell trajectory roadmap in human airways. We have confirmed the cell lineages that have been previously described and have discovered a novel trajectory linking goblet cells to multiciliated cells. We have also discovered novel cell populations and molecular interactors involved in the process of healthy human airway epithelium regeneration. Using these approaches, we have finally shed light on cell-type specific responses involved in pathological goblet cell hyperplasia.

Our data, by bringing significant contributions to the understanding of differentiation's dynamics in the context of healthy and pathological human airway epithelium, may lead to the identification of novel therapeutic targets.

## **KEY WORDS**

Airway epithelium, epithelium regeneration, airway remodeling, single cell transcriptomics, multiciliogenesis, goblet cell hyperplasia.



*Above all, don't fear difficult moments.*

*The best comes from them.*

*Rita Levi-Montalcini*

## Acknowledgement

First my most sincere thanks to Pascal Barbry for welcoming me into his laboratory and proposing this thesis project. Thank you very much for being the boss you are, for making me feel at home reading the emotional pages of the Julio Iglesias biography, thank you infinitely, for always being attentive to the work of all team members even when you don't have time. It has been a pleasure for me to have the opportunity to do my PhD in this laboratory and not only my doctorate but also a post-doc period! Thank you for your confidence in me and in my work I hope to be able to continue learning from your wisdom.

Thanks to my supervisor Laure-Emmanuelle, for all that you have taught me during these years, have been complicated years of failed experiments (many), but you have always known how to teach me the positive side of things, you have made me realize that there is always something good and positive and you have to know how to see it! I hope to know how to put into practice all your good advice and your good work, sure that I will continue to apply your teachings in the future, you taught me how to put a bit of order in all the chaos in my head and that is something that has served me and will serve me forever!

Thanks to the members of the jury for their involvement in this final phase of my thesis, for their good advice and criticism that undoubtedly help me to continue learning.

I am very grateful to work in this team surrounded by such good people, no doubt working with you helps to overcome the difficult moments that every thesis entails.

Above all grateful to Signalfife for funding my studies, thanks to them I have had the great opportunity to be here and get the my PhD, within Signalfife an immense thanks to Konstanze Beck because she has taken care of all of us since we came to the arms of Signalfife, has guided us, has helped us and has supervised each one of us and we are not few! An immense and well done job, without you we would be lost in the networks of the French administration!

Thanks to Laura Moreno who helped me to have my "Spanish moment" during the day, what a relax listen to my mother tongue!

Thanks to Mr. Ponzio for filling the lab with good humor and songs! Thanks to Amélie for being a good colleague and friend these years, thanks for your advice and our discussions! It's a pleasure to be able to vent in difficult moments.

Thanks to Marie for your work and effort! Together we have been able to finish a beautiful story without your work that would not have been so beautiful. Thanks to MJ for working side by side with me even underwater! Thanks to Marine, Karine and Carol for the talks during the meals, small moments of relax during the day and that have helped to improve my French a lot!

Thanks at the BM team! Bernard Roger and George for their support and good advice for the presentation of my thesis, as well as Agnes Paquet who always has good advice thank you for knowing how to guide us and teach us to have clear ideas, thanks to the rest of the members of the

genomic platform Virginie, Geraldine and Kevin is a pleasure to work with you. Thanks to Brice and Rainer for their teachings and advice. Also big thanks to Marin for your last-minute data analysis.

Thanks to Sophie and Fred for teaching me all about microscopy! Good teachers!

Last but not least thanks to my family Papa, mama and Pablo for your support, for giving me the opportunity to lead a wonderful life, without you I would not have been able to be here and finish my studies, thank you for your work and effort to form this wonderful family of which I am more than proud, I feel fortunate to always have you by my side.

To finish thanks to my love and companion, infinite thanks, for the support you have given me, for taking care of me, for loving me even in my worst moments. This final period of the thesis has been very hard but I cannot even imagine what it would have been like if you had not been by my side, you have motivated me, consoled me with your shoulder to cry on so many occasions, you have given me calm and energy to continue, I know that also has been a hard period for you and I promise is the last PhD I do!

## PREFACE

The airway epithelium acts as a protective barrier defending us from pathogen infections, pollutants or external aggressions. It is a pseudostratified epithelium composed by many different cell types such as basal, secretory, goblet, multiciliated cells and rare cell populations such ionocytes, tuft or pulmonary neuro endocrine cells. In adult homeostasis there is a slow turnover of the epithelium ensuring the correct proportion of all the different cell types important for a proper functionality of the epithelium. Our laboratory is mainly interested in the study of the airway epithelium differentiation and regeneration. Deciphering the different mechanism involved in these processes is important for a proper understanding of the mechanisms that are impaired in lung diseases. Thanks to the development of the single-cell RNAseq technique our laboratory elaborated a project aiming to analyse the kinetics during differentiation of human airway epithelial cells cultured in vitro.

When I started my project as a PhD student in the laboratory of Pascal Barbry in 2015, the single cell technology was starting to be developed, as we said we were at the beginning of the single cell era. My role in this project was the development and settle of the different protocols needed to perform single-cell RNAseq in our model of cultured primary human airway epithelial cells. I developed protocols ensuring the best dissociation practices with the minimal perturbation of the cells and maximal cell viability; I learned how to perform cell isolation using different techniques and equipments such as the C1 device from the company Fluidigm or the Chromium from 10XGenomics. Huge work that was developed and tested during a period of two years. Thanks to this work our laboratory designed a new protocol for single cell RNAseq “low cost” (Arguel *et al.*, 2017) (ANNEX) with the publication of this work our team was the first one publishing single cell data in France.

After these two years a new technology arrived to the laboratory, the Chromium from 10XGenomics, and thanks to all the knowledge that was acquired from the previous years we could start working with this new technology and generate big amount of data very fast.

Different studies performed by my lab revealed a new gene involved in the multiciliogenesis, CDC20B (Revinski *et al.*, 2018) (ANNEX), Thanks to the single cell technology I could determine the expression of this gene in a very specific cell type, the multiciliated cells, as I will detail during my results. Functional analysis of this gene performed by the other authors demonstrated its role in the multiciliogenesis.

Our lasts work recently submitted (Ruiz Garcia, Deprez et al. 2018, *submitted BioRxiv*)(ANNEX) recapitulates a big part of this thesis project. My contribution, as first author, was the generation of all the single cell datasets from all the different models used in the study.

Thanks to the tight collaboration with bioinformaticians, notably working side by side with Marie Deprez, also first author of this work, we analyze all these different data sets; from this analysis we could define the different cell populations forming the airway epithelium, some of them already known but interestingly we describe new cell populations appearing during the process of airway epithelium differentiation and in homeostasis, we inferred cell trajectories describing the cell hierarchies during the airway epithelium regeneration, we determine the implication of different signalling pathways involved also in the process but more interestingly we describe new players in the regulation of the epithelium regeneration. Summarizing, this work describes the cell composition, cell hierarchies and signalling regulation of the process of regeneration of human airway epithelium in vitro and in vivo in healthy conditions.

In lung diseases there is a chronic inflammatory state and a chronic injury of the airways that lead to remodelling and mal functioning of the epithelial barrier. The second objective of my thesis project was to decipher the mechanism giving rise to the pathology, more concretely, the remodelling of the airway epithelium. Recent work that I have performed revealed new information about epithelial remodelling during inflammation. I will describe the different results and perspectives in the results section of this manuscript.

# ABBREVIATION LIST

3D-SIM	3D STRUCTURED ILLUMINATION MICROSCOPY
AGR	ANTERIOR GRADIENT
AHI	ABELSON HELPER INTEGRATION SITE
BB	BASAL BODY
BBS	BARDET-BIEDL SYNDROME
BC	BASAL CELL
BMP	BONE MORPHOGENETIC PROTEIN
BNC	BASONUCLIN
CBF	CILIARY BEAT FREQUENCY
CCDC	COILED-COIL DOMAIN CONTAINING
CCNO	CYCLIN-O
CCP110	CENTRIOLAR COILED-COIL PROTEIN 110
CCSP	CLARA CELL SECRETORY PROTEIN
CD	CENTRIOLAR DEPENDENT
CDC	CELL DIVISION CYCLE
CENPJ	CENTROMERE PROTEIN J
CENT	CENTRIN
CEP	CENTROSOMAL PROTEIN
CF	CYSTIC FIBROSIS
CFTR	CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR
CHIP-SEQ	CHROMATIN-IMMUNOPRECIPITATION-SEQUENCING
CNV	COPY NUMBER VARIATION
COACH	COACH SYNDROME
COPD	CHRONIC OBSTRUCTIVE PULMONARY DISEASE
CORD	CONE-ROD DYSTROPHY
COX	CYCLOOXYGENASE
CRISPR	CLUSTERED REGULARLY INTERSPACED SHORT PALINDROMIC REPEATS
DAPT	N-[N-(3,5-DIFLUOROPHENACETYL-L-ALANYL)]-(S)-PHENYLGLYCINE T-BUTYL ESTER
DASC	DISTAL AIRWAY STEM CELL
DD	DEUTEROSOME DEPENDENT
DEUP1	DEUTEROSOME ASSEMBLY PROTEIN 1
DLL	DELTA LIKE LIGAND
DNAH	DYNEIN AXONEMAL HEAVY CHAIN
EGF	EPIDERMA GROWTH FACTOR
EGFR	EPIDERMA GROWTH FACTOR RECEPTOR

EM	ELECTRON MICROSCOPE
FACS	FLUORESCENCE-ACTIVATED CELL SORTER
FC	FOLD CHANGE
FDR-Q	BENJAMINI-HOCHBERG FALSE DISCOVERY RATE
FOX	FORKHEAD BOX
GC	GOBLET CELL
GEM	GEMININ
GEM	GEL BEAD EMULSION
GM-CSF	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR
HAEC	HUMAN AIRWAY EPITHELIAL CELL
HCA	HUMAN CELL ATLAS
HDM	HOUSE DUST MITE
HES	HAIRY AND ENHANCER OF SPLIT
IFC	INTEGRATED FLUIDIC CIRCUIT
IFT	INTRA FLAGELAR TRANSPORT
IG	IMMUNOGLOBULINE
IL	INTERLEUKIN
ILC	INNATE LYMPHOID CELL
INC	INTERCALATING NON-CILIATED CELL
INDROP	INDEXIN DROPLETS
IPF	IDIOPATIC PULMONARY FIBROSIS
ITGA	INTEGRIN SUBUNIT ALPHA
IVT	IN VITRO TRANSCRIPTION
JAG	JAGGED
JAK	JANUS KINASE
JBTS	JOUBERT SYNDROME
KD	KNOCK-DOWN
KIF	KINESIS FAMILY
KO	KNOCK-OUT
KRT	KERATIN
LCA	LEBER CONGENITAL AMAUROSIS
LINNAEUS	LINEAGE TRACING BY NUCLEASE-ACTIVATED EDITING OF UBIQUITOUS SEQUENCES
LNA	LOCKED NUCLEIC ACID
MAP	MITOGEN ACTIVATED PROTEIN
MAPK	MITOGEN ACTIVATED PROTEIN KINASE
MARS	MASSIVE PARALELL RNA SINGLE-CELL

MC	MACULAR DEGENERATION
MCC	MULTICILIATED CELL
MCIDAS	MULTICILIATE DIFFERENTIATION AND DNA SYNTHESIS ASSOCIATED CELL CYCLE PROTEIN
MEF	MOUSE EMBRYONIC FIBROBLAST
MKS	MECKEL GRUBER SYNDROME
MKS	MECKEL SYNDROME
MMLV	MOLONEY LURINE LEUKEMIA VIRUS
MRC	MITOCHONDRIAL RICH CELL
MTEC	MURINE TRACHEAL EPITHELIAL CELL
MUC	MUCIN
NPHP	NEPHRONOPHTHISIS
ODF	OUTER DENSE FIBER OF SPERM TAILS 2
OFD4	OROFACIODIGITAL SYNDROME IV
OVA	OVALBUMIN
PCD	PRIMARY CILIARY DISKINESIA
PCL	PERCILIARY LIQUID
PCM	PERICENTRIOLAR MATERIAL
PCR	POLYMERASE CHAIN REACTION
PI3K	PHOSPHATIDYLINOSITOL-3 KINASE
PLK	POLO LIKE KINASE
PNEC	PULMONARY NEURO ENDOCRINE CELL
PSC	PROTON SECRETING CELL
RFX	REGULATORY FACTOR X
RP	RETINITIS PIGMENTOSA
RPKM	READS PER KILOBASE PER MILLION
RT	RETROTRANSCRIPTION
RT-QPCR	QUANTITATIVE REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION
SC	SECRETORY CELL
SCD	SECONDARY CILIARY DYSKINESIA
SCGB	SECRETOGLOBIN
SCUBA	SINGLE CELL USING BIFURCATION ANALYSIS
SEQ	SEQUENCING
SEV	SENDAI VIRUS
SLICE	SINGLE CELL LINEAGE INFERENCE USING CELL EXPRESSION SIMILARITY AND ENTROPY

SLSN	SENIOR-LOKEN SYNDROME
SNV	SINGLE NUCLEOTIDE VARIANTS
STAT	SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
SVZ	SUB-VENTRICULAR ZONE
TAC	TRANSITORY AMPLIFYING CELL
TCTN	TECTONIC
TGF	TRANSFORMING GROWTH FACTOR
TH	T-HELPER
TMEM	TRANSMEMBRANE PROTEIN
TNF	TUMOR NECROSIS FACTOR
TP	TUMOR PROTEIN
TSCAN	TOOLS FOR SINGLE CELL ANALYSIS
TSLP	THYMIC STROMAL LYMPHOPOIETIN
TSO	TEMPLATE SWITCHING OLIGO
TZ	TRANSITION ZONE
UMI	UNIQUE MOLECULAR IDENTIFIER
WHO	WORLD HEALTH ORGANIZATION
WTA	WHOLE TRANSCRIPTOME AMPLIFICATION
$\gamma$ -TURC	$\gamma$ TUBULIN RING COMPLEX



## Table of contents

<b>I.</b>	<b>The airway epithelium .....</b>	<b>4</b>
I.1	Cell composition of the airway epithelium.....	6
I.1.1	Basal Cells (BCs) .....	6
I.1.2	Secretory Cells (SCs) / Clara / Club cells .....	8
I.1.3	Goblet Cells (GCs) .....	10
I.1.4	RARE CELLS .....	11
<b>II.</b>	<b>The Cilia .....</b>	<b>16</b>
II.1	Cilia diversity .....	16
II.2	Ciliopathies .....	18
II.3	Ciliogenesis.....	21
II.3.1	Centrosomes.....	21
II.3.2	Centriole multiplication.....	23
II.3.3	The deuterosome .....	26
II.3.4	Centriole amplification by the DD pathway .....	26
II.3.5	Centriole elongation.....	28
II.3.6	Basal bodies (BBs) .....	30
II.3.7	The transition zone (TZ).....	30
II.3.8	Intraflagellar transport (IFT) .....	31
II.4	Regulatory mechanisms of multiciliogenesis .....	33
II.4.1	Role of the Geminin super family .....	33
II.4.2	E2f family of transcription factors.....	35
II.4.3	Foxj1 transcription factor.....	36
II.4.4	Rfx family of transcription factors.....	37
II.4.5	Tumor protein (Tp)73 .....	38
II.4.6	Myb transcription factor .....	38
<b>III.</b>	<b>Airway epithelium development, homeostasis and regeneration .....</b>	<b>40</b>
III.1	Differentiation of the airway epithelium in normal homeostasis, development and in regeneration conditions after injury in human and murine models.....	40
III.1.1	Basal cells progenitors of the airway epithelium.....	40
III.1.2	Secretory cells: Precursors of multiciliated cells .....	43
III.1.3	Goblet cell increased differentiation.....	44
III.2	The Notch pathway .....	47
III.2.1	Notch activation favors GC and inhibits MCC differentiation.....	48
III.2.2	BCs send a Notch signal to SCs .....	50
III.2.3	Notch3 over-activation depletes the BC pool by over-differentiation into SCs.....	51
III.2.4	The miR-449 family favors MCC differentiation through Notch pathway inhibition.....	52
III.3	EGF signalling pathway.....	53
III.3.1	EGFR signalling is implicated in the differentiation of GCs.....	54
III.4	TGF-B/BMP Pathway .....	55
III.4.1	BMP negatively regulate MCC differentiation.....	55
III.4.2	BMP maintains BCs in a quiescent state.....	56
<b>IV.</b>	<b>Airway remodelling .....</b>	<b>57</b>
IV.1	Secondary Ciliary Dyskinesia (SCD).....	57
IV.1.1	Asthma.....	57
IV.1.2	Chronic obstructive pulmonary disease (COPD) .....	58
IV.1.3	Cystic Fibrosis (CF).....	59

IV.2	Airway remodelling in disease.....	60
IV.2.1	<i>Mechanical stress and mucus over production triggers of inflammation</i> .....	60
IV.2.2	<i>Inflammation</i> .....	62
IV.3	Th2 cytokines.....	62
IV.3.1	<i>Interleukin-13 (IL-13) as mediator of allergic inflammation</i> .....	63
IV.3.2	<i>Effect of IL-13 in airway epithelium</i> .....	64
IV.3.3	<i>Molecular pathways involved in GC hyperplasia induced by Th2 type cytokine</i> .....	69
IV.3.4	<i>IgE/IgA in asthma</i> .....	73
IV.3.5	<i>Therapies targeting IL-13/IL-4 signalling</i> .....	74
V.	Single-cell transcriptomics: a powerful tool to analyze cell composition of tissues with high resolution.....	76
V.1	Challenges of scRNA-seq .....	79
V.1.1	<i>Tissue dissociation</i> .....	79
V.1.2	<i>Cell isolation</i> .....	80
V.1.3	<i>Cell barcoding</i> .....	81
V.1.4	<i>Untargeted amplification of transcriptomes</i> . .....	82
V.1.5	<i>Unique Molecular Identifiers</i> .....	83
V.2	Single cell RNA-seq protocols.....	83
V.3	Tang method .....	84
V.4	Quartz-seq.....	84
V.5	Smart-seq and smart-seq2 .....	85
V.6	Tag-based sequencing methods: CEL-seq, CEL-seq2 and MARS-seq .....	86
V.6.1	<i>CEL-seq: (Cell expression by linear amplification and sequencing)</i> .....	86
V.6.2	<i>CEL-seq2</i> .....	87
V.6.3	<i>MARS-seq (Massively parallel RNA Single-cell sequencing)</i> .....	87
V.6.4	<i>STRT-seq (Single-cell tagged reverse transcription sequencing)</i> .....	87
V.7	Array based and embulsion-based methods: Cyto-seq, Drop-seq, In-drops and 10x Genomics.....	88
V.7.1	<i>CytoSeq</i> .....	88
V.7.2	<i>InDrop (indexin droplets)</i> .....	88
V.7.3	<i>DropSeq</i> .....	88
V.7.4	<i>10X Genomics (Chromium device)</i> .....	89
VI.	Cell trajectories inference from single cell data sets .....	90
VI.1	Dimensionality reduction-based algorithms .....	90
VI.2	Nearest neighbour graph-based algorithms .....	92
VI.3	Connecting of cluster centers in high dimensional space .....	93
VII.	Combination of single cell transcriptomics and cell lineage labelling .....	95
VII.1	Prospective lineage-tracing .....	95
VII.2	Retrospective lineage-tracing.....	97
VIII.	Single cell studies in airway epithelium.....	99
IX.	RESULTS.....	103
IX.1	Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration .....	104
IX.2	Airway remodelling .....	121
IX.3	MATERIALS AND METHODS .....	151
X.	REFERENCES .....	153
XI.	ANNEX .....	197
XI.1	LIST OF PUBLICATIONS .....	198

XII. A cost effective 5', selective single cell transcriptome profiling approach with improved UMI design 199

XIII. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells .... 200

XIV. Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration' ..... 201

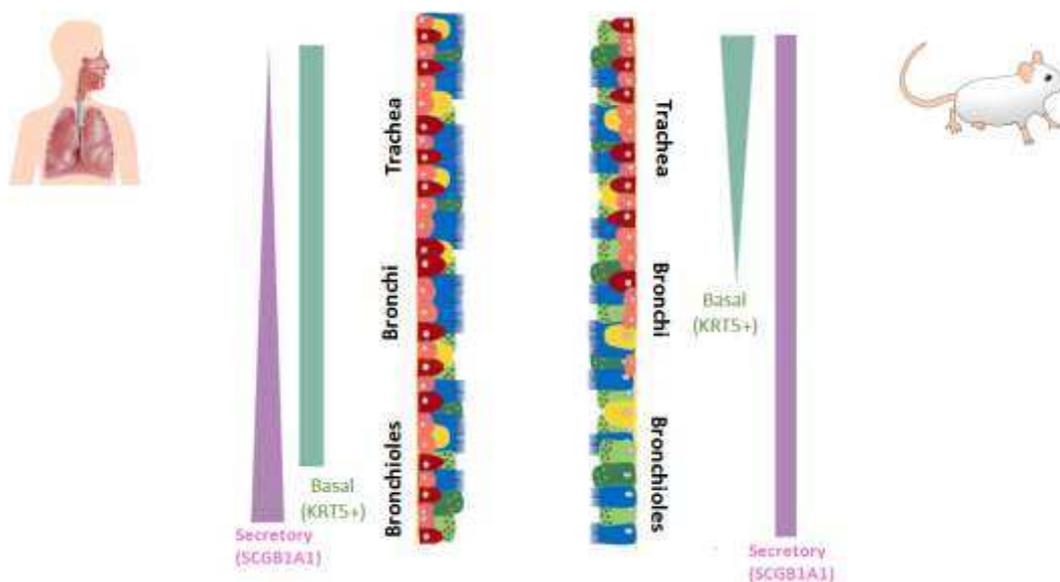
## *I. The airway epithelium*

---

The airway epithelium is a pseudostratified mucociliary epithelium which cellular composition and structure varies all along the respiratory track, from the nasal cavity to the terminal bronchioles. This epithelium is composed mainly by four cell types: (1) Basal Cells (BCs) present on the basal lamina of airway epithelium from the upper airways until the respiratory bronchioles in humans, but only in upper airways of mice (Rock, Rande and Hogan, 2010); (2) multiciliated cells (MCCs) harbouring hundreds of cilia in their apical surface, allowing mucociliary clearance; (3) goblet cells (GCs) secreting mucus and (4) secretory cells (SCs), commonly named Club or Clara cells, that secrete numerous peptides involved in immune response regulation, as well as anti-microbial peptides. Rare cell types such as pulmonary neuroendocrine cells (PNECs), Brush cells (also called Tuft cells) and ionocytes also compose this epithelium. All these different cell populations work together for a proper muco-clearance function of the airways, immune defense against pathogens, mucus composition and maintenance of pH levels and hydration of the airways.

A correct balance between these cell populations is essential to assure its correct function and integrity. Indeed, maintenance of the homeostasis of this epithelium during the whole lifetime of individuals is crucial and should be tightly regulated.

Humans and mice have a somewhat similar composition of the airway epithelium: both are composed by basal, multiciliated, secretory, goblet, neuroendocrine and intermediate cell populations (Rock, Rande and Hogan, 2010; Teixeira et al., 2013). GCs are present in humans but in mice that are maintained in laboratory conditions these cells are rare (Rock, Rande and Hogan, 2010). SCs are present lining all the surface of the mouse airways but in humans it has been described that SCs are enriched in smaller airways and absent in upper airways (Tata and Rajagopal, 2017). BCs are lining all the surface of the human airways from the nasal cavity to the distal lung until the terminal bronchioles, whereas in mouse BCs are restricted to upper airways (Fig.1)



**Fig. 1. Anatomical and cellular differences between murine and human lung.** The respiratory tree consists of two distinct regions: airways and alveoli. Within airways, distinct compartments contain different cell populations that vary along the proximodistal axis. In human airways, basal cells extend throughout the small airways, whereas in mouse, basal cells extend only up to the mainstem bronchi. In humans, cartilage rings span several generations of airway, whereas in mouse, cartilage rings are present only in the trachea and mainstem bronchi. In humans, submucosal glands are present throughout many small airways, whereas in mice submucosal glands are restricted to the proximal domains of the trachea. The distribution of other various cell types is also indicated in the schematics. Notably, whereas much of the small airway in mouse is composed of a cuboidal epithelium consisting mainly of ciliated and secretory cells, in the human such epithelium is restricted to only the very most distal cells of the bronchioalveolar duct junction. *modified from* (Tata and Rajagopal, 2017).

Although there is a different spatial distribution of the cell types present in the airways between humans and mice, all the different cell types are represented in both species, the mouse trachea is the part of the mouse airways more similar to human airways.

Most studies in the field have been performed with mouse trachea to analyse the development of the airway epithelium, but it is important to keep in mind possible regional differences when interpreting studies.

## I.1 Cell composition of the airway epithelium

### I.1.1 Basal Cells (BCs)

BCs are the first component of the epithelium that lines all the respiratory tract until the respiratory bronchioles where the pseudostratified epithelium is replaced by a cuboidal epithelium without BCs. BCs are directly in contact with the basal lamina, they present a high number of desmosomes and keratin fibres allowing the formation of a compact and tight epithelium and proper anchoring to the basal lamina creating a barrier with the external environment (Evans and Plopper, 1988; Evans *et al.*, 1989, 2001; Mercer *et al.*, 1994; Rock, Rande and Hogan, 2010).

In mouse the presence of the BC is restricted to the trachea and main bronchi and it is already in the main bronchi where the epithelium becomes simple columnar epithelium without BC (Evans *et al.*, 2001; Rock, Rande and Hogan, 2010; Tata and Rajagopal, 2017).

The number of basal cells per mm varies between species, from 39 basal cells per mm in hamster's trachea to 197 basal cells per mm in the sheep trachea (127 in human). The thickness of the basal cell layer depends on the number of basal cells for a given surface. The thickness of the epithelium is not dependent in the specie but varies according to the diameter of the airway: airways with larger diameters have a thicker epithelium and a higher number of basal cells than airways with smaller diameters (Evans and Plopper, 1988; Evans *et al.*, 2001).

The transcriptional profile of the BCs has been very well described, and BCs are characterized by the expression of the *TP63* (transcription factor transformation-related protein 63, TP63, *Trp63* in mouse) which is essential for the development of BCs, as shown by Daniely's lab with a *Trp63* null mice that was not able to develop BCs in tracheal epithelium (Daniely *et al.*, 2004). The hemidesmosome component integrin ITGA6 and basonuclin (zinc finger protein basonuclin-1 is a BC-specific transcription factor encoded by *Bnc1*) are also characteristic of BCs.

BCs are also characterized for the expression of cytokeratins 5 and 17 (*KRT5*, *KRT17*) and some BCs are positive also for cytokeratin 14 (*KRT14*) (James E Boers, Ambergen and Thunnissen, 1998; Nakajima *et al.*, 1998; Evans *et al.*, 2001; Kyung U Hong *et al.*, 2004; Kyung U. Hong *et al.*, 2004; Rock *et al.*, 2009; Hackett *et al.*, 2011; Watson *et al.*, 2015; N. F. Smirnova *et al.*, 2016). Hackett's group published in 2011 the transcriptome profile of BCs extracted from human trachea and they performed the differential analysis between the transcriptome of these BCs and the transcriptome of the luminal epithelium. Their study reports the expression of all the above-mentioned markers, but it also defines additional ones that were highly expressed in the BC population: cytokeratin 6A and 16 (*KRT6A*, *KRT16*), interleukin receptor-like 1, small proline-rich protein 1A and collagen type XVII alpha 1 (Hackett *et al.*, 2011). Their human airway BC signature largely

coincides with the one published by the group of Rock in 2009 (Rock et al., 2009). The latter provide new evidence that the BC population is indeed more heterogeneous with subgroups such as populations of KRT5<sup>+</sup> KRT14<sup>+</sup> cells found in mouse and already reported by Nakajima in 1998 (Nakajima et al., 1998), but also additional groups of BCs expressing distinct cytokeratins signatures. Recent evidences, such as the increase of KRT5 and KRT14 expressing BCs in idiopathic pulmonary fibrosis (IPF) (Ficial *et al.*, 2014; N. F. Smirnova *et al.*, 2016; Confalonieri *et al.*, 2017), suggest alterations in the abundance and expression of different cytokeratins are associated to lung diseases.

It has been shown also that an incorrect balance between BC proliferation and differentiation led to an altered regeneration of the epithelium causing BC hyperplasia or epithelial hypoplasia (Rock, Rande and Hogan, 2010; N. F. Smirnova *et al.*, 2016). In this context, alterations in the BC genome such as mutations or epigenetic modifications can induce chronic susceptibility to respiratory diseases.

#### 1.1.1.1 Function of BCs in the epithelium

BCs have a dual function in the airway epithelium, a function of anchorage, as mentioned before, and a function of development of the epithelium.

BCs have been considered as progenitor cells in central airways. In 1988-1989, the group of Inayama performed xenografts of rabbit cells suspensions with 90% of BCs over denuded tracheas of mice (Inayama et al., 1988) or in culture (Inayama et al., 1989) and showed re-epithelization with all tracheal cell types showing the progenitor cell features of BCs. However, cell renewal mechanisms at the steady state can differ from the mechanisms occurring after injury (Rawlins et al., 2009). BCs also have stem cell features, they maintain themselves by cell division and act as a source of the other cell populations in the airway epithelium (Breuer et al., 1990; Montoro et al., 2018).

The groups of Keenan, McDowell, Evans and Lum during the 80s performed different studies inducing epithelium regeneration after injury with very different methods such as mechanical injury, deprivation of vitamin A or exposition to nitric oxide (NO<sub>2</sub>), oxygen or ozone, they proposed that the mechanisms they observed for regeneration of the epithelium were the same after injury than at steady state. According to them, the major difference was the speed of the cell regeneration (Lum et al., 1978; Keenan, Combs and McDowell, 1982a, 1982b, 1982c; Keenan, Wilson and McDowell, 1983; Evans et al., 1986). Their study showed the importance of secretory cells (SCs) during epithelium regeneration, i.e. an observation that rather depreciated the role of BCs in this process. This assumption was challenged in 1990 by Breuer and Johnson who showed that BCs have an important role for normal growth and renewal of the epithelium at steady state since BCs are much more frequent (32-39%) than the other cell populations. It is only after injury that they detected an increase in the number of SC population, which were in that case the cells directly involved in the epithelium recovery (Breuer et al., 1990; Johnson et al., 1990).

It is also noteworthy that some experiments of transplantation performed in mice, with implantation of rabbit SCs, did not give rise to cilia or mucus cells and did not restore a functional mucociliary epithelium (Brody et al., 1987).

### 1.1.2 Secretory Cells (SCs) / Clara / Club cells

The non-ciliated SCs of the airways were first described by a Nazi histologist in 1937 called Max Clara who identified their distinct cytoplasmic granules, indicative of a secretory function. These cells were initially called Clara cells in his honour, but due to the implication of Clara during the 3<sup>rd</sup> Reich, they have been renamed SCs or Club cells (James E Boers, Ambergen and Thunnissen, 1998).

SCs represent a heterogeneous population which phenotype varies all along the axis of the airways from the proximal to the distal end of the system (Plopper et al., 1983). In human, they are mainly present in the distal part of the conducting airways in normal human lungs and their contribution to the proliferation compartment was 44% in the respiratory bronchioles to be compared with the 15% in terminal bronchioles, demonstrating a substantial role of the SC cells in the maintenance of the epithelium in the distal airways of human lungs (James E Boers, Ambergen and Thunnissen, 1998).

The morphological classification of SC distinguishes non-ciliated bronchiolar SCs from serous cells and GCs (Reynolds et al., 2002), but the more recent use of molecular classifiers slightly modified this classification. For instance, expression of Clara cell secretory protein (CCSP, CC10, CC16, PCB-BP and SCGB1A1), which was considered as a specific marker of SCs, is also detected in serous and GCs.

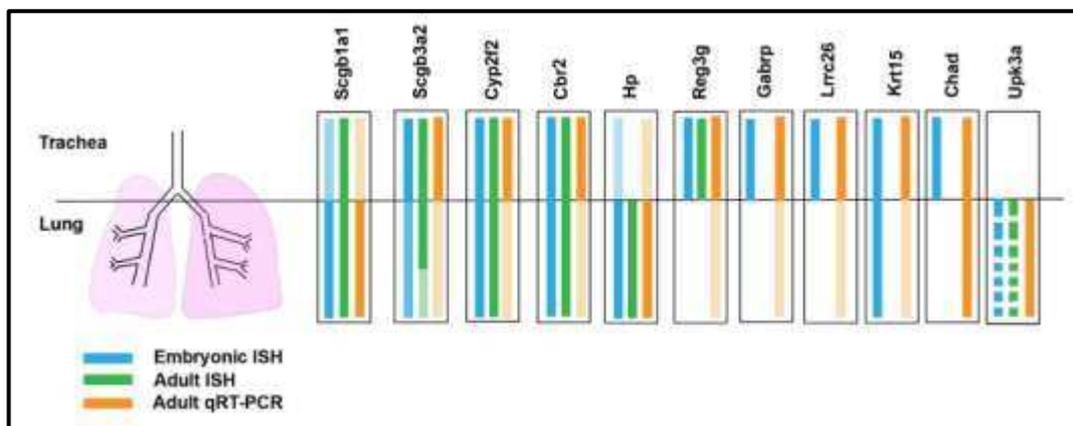
In 2002 the group of Stripp demonstrated that two members of the secretoglobin gene family, SCGB3A1 and SCGB3A2, define molecularly distinct subsets of SCs within the conducting airway epithelium. They identified SCGB3A2 as an early molecular marker for SCs all along the airway epithelium in mouse while, in contrast, SCGB3A1 was highly expressed in epithelial cells of the bronchi. They detected a similar behaviour in human, but the expression of secretory genes was more restricted to the distal part of the human airway system and were more associated with glands (Reynolds et al., 2002).

The group of Guha in 2014 analysed the transcriptome of SCs in embryonic mice to identify markers of SCs precursors and SC markers inhibiting Notch Pathway in developing lung and identified different markers for distinct subpopulations of SCs along the proximal-distal axis of the respiratory system (Guha et al., 2014) (Table 1 and Figure 2).

**Table 1. Microarray-based identification of mRNAs downregulated in Rbpjk<sup>NULL</sup> lungs at E18.5 and validation using qRT-PCR.** Genes downregulated greater than two-fold (p-value ,0.05 and FDR-q value ,0.28) are shown here. FC= fold change; FDR-q= Benjamini-Hochberg False Discovery Rate.

A total of n= 3 lungs per condition from control (CTRL) and Shh cre/+; Rbp Flox/Flox (RbpjkCNUL) were used for the respective analyses. From (Guha *et al.*, 2014).

Symbol	Description	Array			qRT-PCR	
		E18.5 p	E18.5 FDR q	E18.5 FC	E18.5 p	E18.5 FC
Scgb1a1	secretoglobin, family 1A, member 1 (uteroglobin)	9.42E-04	0.13	-41.31	1.43E-03	-285.8
Scgb3a2	secretoglobin, family 3A, member 2	6.76E-05	0.07	-15.44	1.13E-02	-665.9
Reg3g	regenerating islet-derived 3 gamma	2.19E-03	0.18	-9.84	8.25E-04	-318.8
Cyp2f2	cytochrome P450, family 2, subfamily f, polypeptide 2	7.96E-04	0.13	-4.20	1.28E-02	-12.5
Krt15	keratin 15	2.56E-05	0.05	-3.41	9.90E-04	-452.8
Chad	chondroadherin	2.21E-04	0.09	-2.88	6.46E-03	-76.7
Hp	haptoglobin	3.83E-05	0.06	-2.48	1.03E-02	-7.4
Cbr2	carbonyl reductase 2	3.40E-03	0.22	-2.44	1.66E-02	-7.7
Lrrc26	leucine rich repeat containing 26	2.87E-04	0.09	-2.25	3.68E-03	-30.9
Upk3a	uroplakin 3A	5.39E-04	0.11	-2.20	2.90E-03	-209.9
Gabrp	gamma-aminobutyric acid (GABA) A receptor, pi	5.42E-03	0.27	-2.19	3.68E-03	-30.9



**Fig2. Summary of gene expression patterns in embryonic and adult lung SCs.** Cartoon summarizing the domains of expressions of the genes identified in table.1 in the embryonic (E18.5) and adult lung based on ISH (blue, green) and qRT-PCR (orange) analysis. From (Guha *et al.*, 2014).

### 1.1.2.1 Function of SCs in the epithelium

SCs are differentiated luminal cells that have both secretory and detoxifying functions, they have a function plasticity crucial for adaptation after acute stress on the airways such as an infection or pollutant exposure. It has been shown that in cases of acute inflammation or in chronic inflammatory states, the SCs rapidly transform in mucus-producing GCs (Evans *et al.*, 2004; Chen *et al.*, 2009) . Together with BCs, SCs are potential progenitors of the epithelium. As mentioned before, SCs are the main cell population involved in epithelium regeneration after injury. Around the 90s Breuer and Evans, after injury induction, noticed that SCs were acting as progenitors (Evans *et al.*, 1986; Breuer *et al.*, 1990). Differentiation of SCs into goblet and ciliated cells during epithelium regeneration homeostasis will be further described in another part of the manuscript.

### 1.1.3 Goblet Cells (GCs)

GCs were identified by histologists mainly lining the gut, as cells having a cup-shape filled by secretory granules (Rogers, 1994a). In humans, GCs are found at the surface of the airway epithelium in the upper and lower respiratory track and in the submucosal glands and are absent in the respiratory bronchioles and not very frequent in the terminal bronchioles.

#### 1.1.3.1.1 Function of the GCs

GCs secrete high molecular weight mucous glycoproteins (mucins) that protect the airways against contacts with bacteria or external inhaled particles. They are involved, together with MCCs in mucociliary clearance (Rogers, 1994b, 2003; Cohn, 2006; Turner et al., 2011; Bonser and Erle, 2017).

The viscosity of the mucus is essential for the well-functioning of the ciliary/mucous protecting system. Mucins are the molecules giving the right viscosity to the mucous solution lining the airways. They are secreted by GCs of the conducting epithelium and also by the GCs in the submucosal glands, which are only present in the cartilaginous region of the airways (Rogers, 1994a). Thus, GC are the only source of mucins in the non-cartilaginous bronchioles, but they are absent in the smallest airways (<2mm in diameter). In healthy humans the submucosal glands provide more than the 95% of the upper airway mucus (Wine, 2004).

There are 21 genes *MUC* but not all of them code for “real” secreted mucins, the nomenclature of these genes groups together mucins and mucin-like proteins, the difference resides in the ability of generation of the three-dimensional network of the mucus gel. There are only 4 secreted mucins coded by the genes *MUC5AC*, *MUC5B*, *MUC2* and *MUC6*, from these mucins only *MUC5AC* and *MUC5B* are secreted in the airways (Porchet and Aubert, 2004).

A very recent publication described the localization of the expression of both mucins in the healthy human airways, the authors revealed that *MUC5B* was the dominant secreted mucin in the airways, in submucosal glands, but also at the surface of conducting airways, and is mainly restricted to distal airways. *MUC5B* is critical for homeostatic defence (Roy *et al.*, 2014; Okuda *et al.*, 2018). *MUC5AC* is related to airway hyperresponsiveness and mucus obstruction and is expressed predominantly in the upper airways, in the cartilaginous region (Evans *et al.*, 2015; Okuda *et al.*, 2018).

GCs they can be also positive for the SC marker CC10, 25% of the GCs in the bronchi were positive for CC10 expression, in the larger non-terminal bronchioles 34% and in the terminal bronchioles 45% (James E Boers, Ambergen and Thunnissen, 1998).

In pulmonary diseases such as Asthma or COPD there is a GC hyperplasia in the upper conducting airways, meaning that there is an increase in the number and size of the GCs with the consequent increase in mucus secretion. In the smallest airways, where usually we don't find these GCs, we talk about metaplasia of GCs that start to be present in these thinner ducts, leading to mucous trapping and airflow obstruction.

In experimental conditions secretion of mucins in animals or in cell cultures can be stimulated by exposure to inflammatory stimuli, bacterial products, with irritant gases or others such as oxidant stress.

#### I.1.4 RARE CELLS

##### I.1.4.1 Ionocytes

Ionocytes also called mitochondrion-rich cells (MRCs) or chloride cells, are cells specialized in ionic transport. Ionocytes have been described in fish at the level of the gills, in frog skin, as well as in several mammalian organs systems such as kidney, epididymis and endolymphatic duct of the inner ear (Jänicke, Carney and Hammerschmidt, 2007; Esaki et al., 2009; Quigley, Stubbs and Kintner, 2011).

##### I.1.4.1.1 Function of ionocytes

In fish living in fresh water, facing problems of losing salt from their bodies and gaining big amounts of water, ionocytes actively transport salt from the media of the fish into its body through the gills. In larval stages ionocytes can be found in the skin of the animal due to the non-mature function of the gills.

Two different types of ionocytes have been identified in the skin of zebrafish larvae depending on the ATPases they use to provide the driving forces of the ionic transport: (1) one rich in  $\text{Na}^+\text{-K}^+\text{-ATPase}$  with a minor subpopulation involved in  $\text{Ca}^{2+}$  absorption (2) the second one called  $\text{H}^+$ -pump-rich cells characterized by the expression of V-type  $\text{H}^+$  ATPase this second one is also rich in carbonic anhydrase 2 ATPase involved in  $\text{Na}^+$  intake and also has an ammonia transporter (*Rhcg1*) involved in ammonia secretion (Esaki *et al.*, 2006, 2009; Jänicke, Carney and Hammerschmidt, 2007). In the case of  $\text{H}^+$  ATPase ionocyte cells, it is known that differentiation is driven through the action of the transcription factor Foxi3a, orthologue of Foxi1 (Esaki et al., 2009; Quigley, Stubbs and Kintner, 2011; Montoro et al., 2018), and the fate of cells to become ionocytes is negatively regulated by the Delta-Notch competitive lateral inhibition, blocking Notch activity increases Foxi1 and increases the number of PSCs (Esaki et al., 2009).

In mammals, ionocytes are also called Proton-secreting cells (PSCs) as they drive transepithelial movements of ions required for pH osmoregulation using the  $\text{H}^+$  ATPase. PSCs have been classified depending in their gene expression, there are specific isoforms of the  $\text{H}^+$ ATPase such as *atp6v0a4* and *atp6v1b1*, or anion antiporters used to exchange bicarbonate  $\text{HCO}_3^-$ , or for  $\text{Cl}^-$  such as *Slc26a4* (pendrin), *Slc4a1* (AE1) or *Slc4a9*,

and carbonic anhydrases used to produce protons and  $\text{HCO}_3^-$  from water and carbon dioxide. As in zebrafish, the development of PSCs in mice requires of the transcription factor *Foxi1* (Vidarsson *et al.*, 2009).

Quigley *et al.* showed that there were two cell types arising in the *Xenopus laevis* embryos after blocking of Notch pathway, one of the cell types were MCCs and the other one they called Intercalating non-ciliated cells (INCs) or small SCs. They showed that this INCs expressed genes specific for PSCs and the population expressing these genes was heterogeneous. After a meticulous analysis they classified two subgroups of these INCs based in the localization of the  $\text{H}^+$  ATPase and the differential expression of *ae1* and *pendrin* (Quigley, Stubbs and Kintner, 2011).

Recent experiments of single cell transcriptomics of the airway epithelium have shown a cluster of cells expressing markers present in the previously described ionocytes such as the transcription factors *Foxi1*, *Ascl3*, and *Tfcp2l1*. These cells also expressed subunits of the  $\text{H}^+$ ATPase, and several subunits of  $\text{Cl}^-$  transport systems (*Clcnka*, *Clcnkb*, *Stap1*, *Bsnd*). Interestingly, this cluster of cells is also highly enriched in *Cftr* (Montoro *et al.*, 2018; Plasschaert *et al.*, 2018), which encodes for a chloride channel that is mutated in cystic fibrosis. In their work the authors demonstrated that blocking of Notch signalling pathway led to a decrease in the number of ionocytes. This behaviour is totally opposite to the one reported previously in the studies from Esaki (Esaki *et al.*, 2009)

The level of expression of total CFTR appeared correlated to the expression of the ionocytes (Plasschaert *et al.*, 2018).

Ionocytes seem to have a basal origin as shown by Plasschaert *et al.* and Montoro. The former saw a *Krt5+* *Foxi1+* cell population during the recovery of the epithelium after injury (Plasschaert *et al.*, 2018), and the latter performed over mouse tracheal cells single cell transcriptomics coupled with genetic cell lineage tracing over time and they describe cell lineages showing that ionocytes are immediately descendants of BCs (Montoro *et al.*, 2018).

Submucosal glands are the predominant site of CFTR expression in the human bronchia (Engelhardt *et al.*, 1992), so, probably a predominant site for ionocyte localization.

#### 1.1.4.2 Pulmonary Neuro Endocrine cells (PNECs)

PNECs constitute one of the minor cell populations, which has made difficult their analysis and characterization. These cells can be found in the airway epithelium forming clusters, which were called Neuroepithelial Bodies (NEBs) (Ito, 1999).

PNECs are marked by the expression of CGRP (calcitonin gene-related peptide: CALCA), chromogranin A and ASCL1 (achete-scute homolog1).

They derive from the same progenitors than ionocytes, tuft cells and SCs, i.e. BCs, but contrary to the others, PNECs are the first cells type differentiating from these progenitors in the embryonic airways (Ito,

1999; Tata and Rajagopal, 2017). Immunohistology studies show how the PNECs are able to migrate in the epithelium in order to group themselves together and form the NEBs. Lineage tracing studies indicated that PNECs are able to self-renew (Song et al., 2012).

#### 1.1.4.2.1 Function of the PNECs

Some authors mention the hypothesis that PNECs are niche of the Naphtalene-resistant SCs (Hong et al., 2001; Guha et al., 2012), other studies suggest an oxygen sensory function immune cell recruitment and chemosensing (Ito, 1999; Branchfield et al., 2016).

#### 1.1.4.3 Tuft cells

Tuft cells also referred as Brush, caveolated, multivesicular and fibrilovesicular cells, are chemosensory cells in the epithelia lining the surface of the respiratory tract and intestines. Their name come from the microvilli projecting in their surface. Tuft cells have a “pear” shape.

##### 1.1.4.3.1 Function of the tuft cells

There is not much information about the function of these cells in the airways, some authors have seen an increase of these cells after injury in mice and in humans, only in pathological conditions (Reid et al., 2005).

In recent single cell transcriptome analysis studies, the authors were able to detect the presence of these brush cells in the airway epithelium of mice and human in homeostatic adult tissue. Using a model of injury using Polidocanol in mice they revealed that these cell types were also restored to a normal value after regeneration of the epithelium. They identified a specific marker for tuft cells, *Rgs13*. A closer analysis of the tuft cells revealed three different clusters of tuft cells, immature tuft cell, tuft-1 and tuft-2, where tuft-1 expressed markers associated with taste transduction and tuft-2 expressing genes mediating in inflammation and asthma. Lineage tracing revealed a basal origin of these (Montoro et al., 2018; Plasschaert et al., 2018).

#### 1.1.4.4 Multiciliated cells (MCCs)

MCCs cells in the airways harbour around 250 of motile cilia in the apical part. They have a pyramidal or cylindrical shape and they are around 20  $\mu\text{m}$  tall and 7  $\mu\text{m}$  wide (Breeze and Wheeldon, 1977).

In mammals these cells are present in the airway epithelium, brain ventricles, oviduct and efferent ducts, they can be found also in kidney and oesophagus during embryonic development (Sorokin, 1968; Danielian *et al.*, 2016; Meunier and Azimzadeh, 2016). In other chapter of the manuscript I will define the process of multiciliogenesis by which MCCs are able to generate hundreds of motile cilia at their apical surface.

In *Xenopus*, MCCs are found in skin trachea and digestive tract, and also in pronephros of *Xenopus* and zebrafish (Gerlach and Wingert, 2014; Walentek and Quigley, 2017).

The differentiation of MCCs is negatively regulated by the Notch signalling pathway. Blocking of the pathway increase the number of MCCs (Guseh *et al.*, 2009a; Tsao *et al.*, 2009; Morimoto *et al.*, 2010; Kyrousi *et al.*, 2015). The superfamily of microRNAs (miRNAs) miR-34/449 regulates the process of MCC differentiation inhibiting the Notch signalling pathway targeting NOTCH1 and its ligand DLL1 (Marcet, Chevalier, Luxardi, *et al.*, 2011; Chevalier *et al.*, 2015). The BMP pathway interacts with Notch to also regulate the MCC differentiation (Cibois *et al.*, 2015).

The inhibition of the notch signalling pathway initiates a cascade of MCC differentiation with the activation of members of the geminin family *GEMC1* and *MCIDAS*, and the transcription factors *RFX2/3*, *FOXJ1*, *C-MYB* and *E2F4/5*. I will describe the process of MCC differentiation in other chapter of the manuscript.

#### 1.1.4.4.1 Function of MCCs

Combined with mucus produced by the GCs and the submucosal glands, MCCs are part of the crucial of the “mucociliary escalator”, also known as “mucociliary clearance”, which composes a first layer of protection against aerial aggressions. The hundreds of cilia of the MCCs beat in a coordinated fashion to generate a synchronic fluid flow of the mucus, in which all particles and microbes can be trapped in and then expelled out of the respiratory system. Defective mucociliary clearance can cause severe infectious diseases, such as in cystic fibrosis or in PCD (primary ciliary dyskinesia).

MCCs are not unique to the airways, and in mammals, MCCs are also found: (1) in the spinal cord and ventricles of the adult brain they drive the flow of the cerebrospinal fluid helping also with the neuronal migration (Sawamoto *et al.*, 2006), (2) in the oviducts and fallopian tubes where they are important for the ovum transport (Lyons, Saridogan and Djahanbakhch, 2006; Meunier and Azimzadeh, 2016; Spassky and Meunier, 2017), (3) In the efferent ducts of the testis allowing proper maturation of sperm (Danielian *et al.*, 2016; Terré *et al.*, 2018).

Origins of MCCs have been studied for a long time. They derive from TP63 expressing progenitors (BCs), and it has been shown that the absence of Notch in the system favours the fate of cell differentiation into multiciliated phenotype (Guseh *et al.*, 2009a; Tsao *et al.*, 2009; Morimoto *et al.*, 2010; Marcet, Chevalier, Luxardi, *et al.*, 2011; Danahay *et al.*, 2015; Lafkas *et al.*, 2015; Mori *et al.*, 2015).

Our laboratory has been focused for many years in deciphering the mechanisms involved in the regulation of the multiciliogenesis. During the introduction of this manuscript I will describe more extensively the generation of the MCCs in the airways with a first description of the cilia and ciliogenesis to better understand the process of MCCs differentiation in the context of regeneration of airway epithelium.



## II. The Cilia

---

In 1968, Sorokin made the first wide description of the generation of non-motile primary cilia in the mammal lung (Sorokin, 1968). This study was based on the observation of the different events of the ciliogenesis using electron microscopy. But it has been during the 21<sup>TH</sup> century that different studies showed the role of these cilia in physiological processes such as signal transduction, chemo- or mechano- detection, olfaction, vision, equilibrium, left/right asymmetry, fluid flow and locomotion.

Cilia are specialized cellular organelles, made of nine microtubule (MT) doublets assembled and projecting to the extracellular space. Their structure is highly conserved from protists to vertebrates. Unicellular organisms use motile cilia for locomotion, feeding and sensation. In vertebrates, most cells are able to generate a non-motile primary cilium, as an organelle involved in signal transduction during development and homeostasis (Goetz and Anderson, 2010). Some specialized cells are involved in a specific process of differentiation called multiciliation that results in the synthesis of hundreds of motile cilia at their apical side. These cells are called multiciliated cells (MCCs).

### II.1 Cilia diversity

The term cilia include two subtypes, motile and non-motile cilia. The structure of the motile cilia corresponds with the nine assembled doublets of MT and a central pair of MT in a conformation (9+2), whereas non-motile cilia lack the central pair of MT in a conformation (9+0) with some exceptions such as the neural motile cilia of the node in embryos with a conformation (9+0) (Sedykh *et al.*, 2016; Mitchison and Valente, 2017). There is a vast diversity of cilia. For example, primary cilia in human kidney cells have with a size ranging from 10 to 15  $\mu\text{m}$ , and the cilia of the sperm of *Drosophila melanogaster* reaching a length of 5 cm. Different types of cilia have been described:

a) Primary cilia: non-motile cilia, is present in almost all the human cells as a sensory cilia, this organelle is able to capture extracellular signals and integrate them to elaborate biological responses like regulation of the cell cycle, development and idfferntiation processes, signal transduction migration and polarity and maintenance of stme cells(Elliott and Brugmann, 2018)(Fig. 3a).

b) Flagella: The term flagella is also used to describe motile cilia. Flagella have usually the same structure as motile cilia, but their length is bigger. This organelle protrudes from the body of bacteria or some eukaryotic cells, it has a primary function for locomotion but also serves as a sensorial organelle to detect chemicals and temperatures of the extracellular environment (Fig 3b).

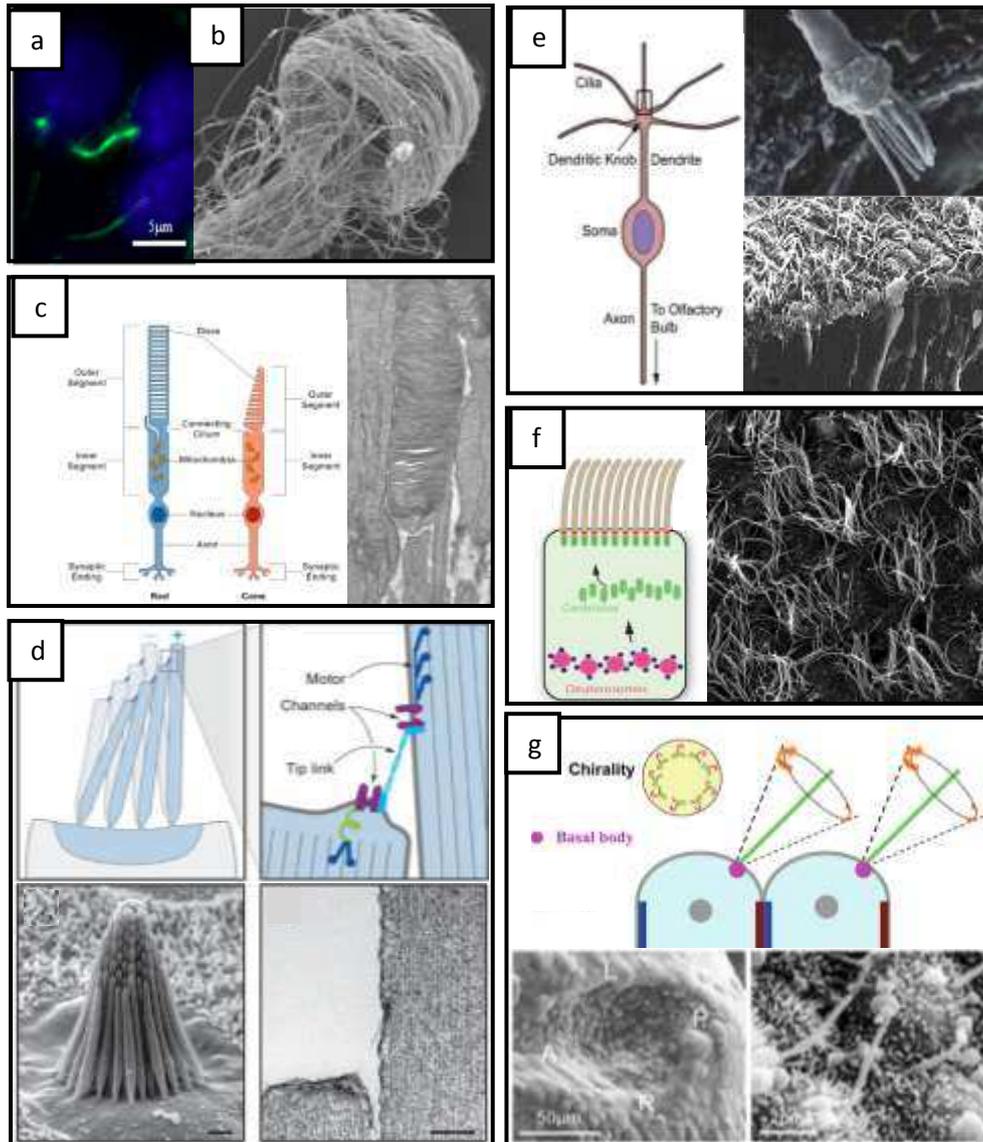
c) Ciliary photoreceptors “Rods & Cones”: photoreceptor cells in the retina of the eye. These cells harbour a modified cilium where the ciliary membrane is expanded and thrown into deep folds, this modified cilium is the light receptor and looks like a stack of discs, each of these cells (Rods and Cones) there are going to have different photosensitive pigments, the Rods have Rhodopsin in charge of the night vision due to its high sensitivity to light, Cones are less sensitive to light, they have three different kinds of photosensitive pigments in the group of the photopsins, they are less sensitive to light but they are more accurate, responsible of the day vision and the reception of the different colours (Fig 3c).

d) Kinocilium and stereocilium: Cilia present in the apical surface of the “Hair cells”, these cells are mechanosensitive cells. The stereocilium are sensory receptors of the auditory system and vestibular system in the ears of all vertebrates. Hair cells detect movement in their environment through mechanotransduction, they are located in the cochlea of the inner ear. The hair cells also amplify the low-level sound that enters in the cochlea (Fig 3d).

e) Olfactory cilia: Cilia of olfactory sensory neurons (OSN) necessary for detection of odorants. The dendrites of the OSN extend towards the nasal cavity ending in a knob containing multiple cilia projecting into the mucus of the nasal epithelium (Jenkins, McEwen and Martens, 2009) (Fig 3e).

f) Cilia of Multiciliated cells: Motile cilia in the apical surface of the MCCS, they are responsible of the generation of a fluid flow of the liquid lining the surface of the cells or locomotion in different animal species. In mammals these cells are present in the spinal cord and ventricles of the adult brain, in the airways and in the oviduct/fallopian tubes and efferent canals (Meunier and Azimzadeh, 2016) (Fig 3f).

g) The nodal cilia: motile cilia essential during the development of the embryo. The motile cilia in the node serves to set the correct left/right asymmetry of the body and all the organs (Nonaka et al., 1998) (Fig 3g).



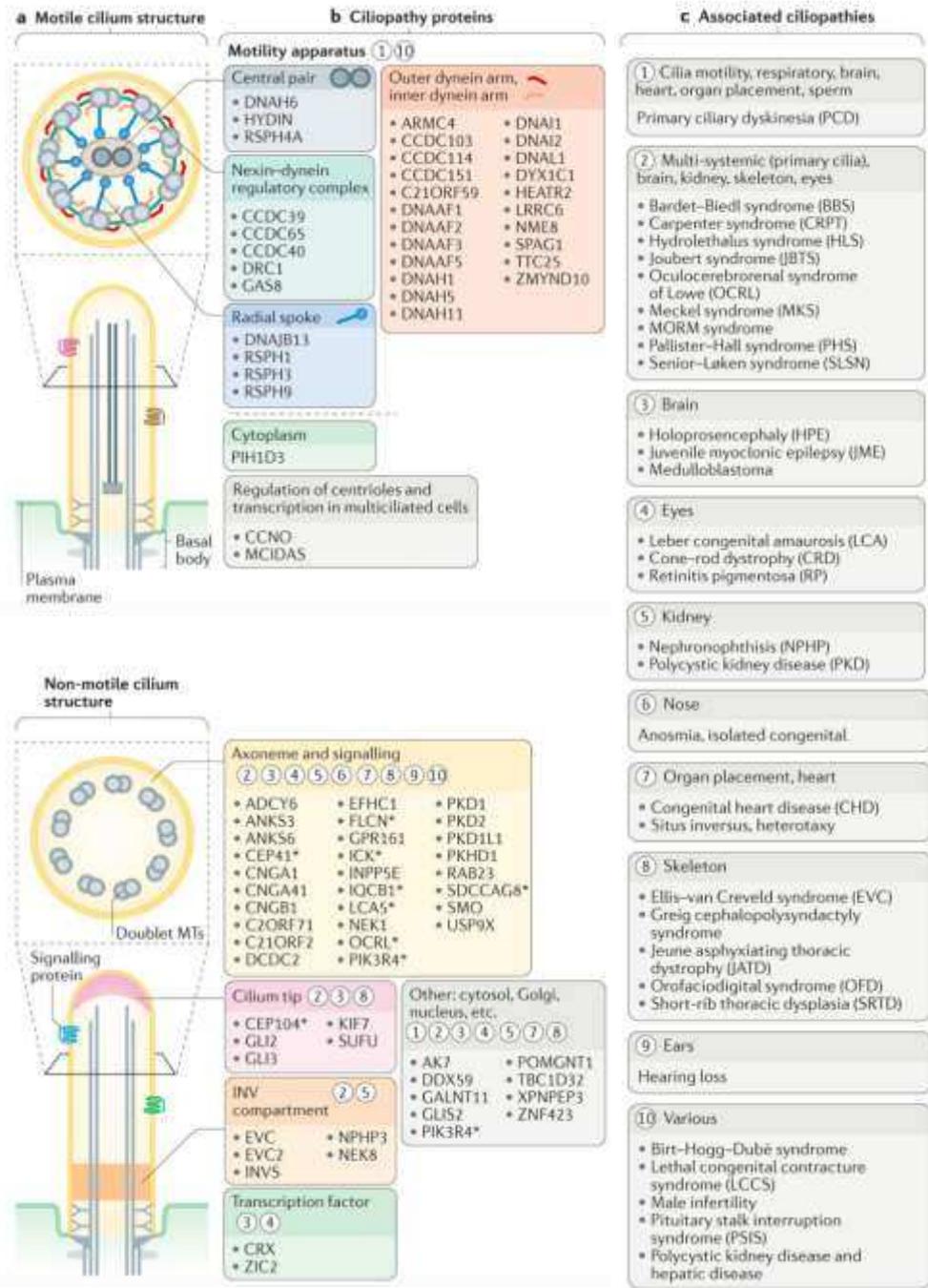
**Fig3. Cilia heterogeneity in mammals.** (a) Primary cilia, (b) Flagella, (c) Ciliary photoreceptors, (d) Kinocilium and stereocilium, (e) Olfactory cilia, (f) cilia of multiciliated cells, (g) Nodal cilia.

## II.2 Ciliopathies

Many components are required for the biogenesis, maintenance and function of cilia. Development of new high throughput sequencing techniques and proteomics allowed the identification of more than 2500 proteins (Inglis, Boroevich and Leroux, 2006) which are playing a role in ciliogenesis, also the work of Maxence V. Nachury described all the proteome of the cilia performing proximity labelling using cilia-APEX (Mick et al., 2015). These recent works eventually led to the generation of a database called “the ciliome”.

This knowledge will certainly have a huge impact on the research on ciliopathies, many of these diseases such as the Joubert syndrome, Bardet-Biedel (BBS), Meckel-Gruber (MKS), and Nephronophthisis (NPHP), are due to mutations in cilia-associated genes, illustrating the importance of understanding cilium structure and function and the mechanisms required for its assembly.

Recently Jeremy F. Reiter and Michel R. Leroux published a review describing all the different ciliopathy-associated proteins. Figure 4 summarizes the main proteins of components of the motile and non-motile cilia involved in ciliopathies (Fig4) (Reiter and Leroux, 2017).



**Fig. 4. Structural and functional features of motile and non-motile cilia are associated with ciliopathies.** (a) Representation of the major structures of motile and non-motile cilia. (b) Major sites of action for ciliopathy-associated proteins that are components of motile cilia (motility apparatus or transcription factors required for the generation of motile cilia) and sensory cilia (axonemal and signalling proteins, ciliary tip proteins or inversin (INV) compartment proteins). The asterisks indicate proteins that are also localized to other ciliary regions during ciliogenesis or ciliary trafficking. Circled numbers indicate one or more ciliopathies that result from defects in the different ciliary compartments and proteins. (c) Ciliopathies grouped into major categories that are associated with the proteins and ciliary regions shown in part (b). From (Reiter and Leroux, 2017).

To illustrate the extent of cilia-related disease-causing genes, Table 2 recapitulates all the proteins localized at the transition zone of the cilia which mutations are involved in human diseases. The transition zone is an important part of the cilium and corresponds to the proximal portion of the axoneme, I will describe the transition zone of the cilia in other section of the manuscript. (reviewed in Gonçalves and Pelletier, 2017).

**Table 2: TZ proteins – associated diseases and loss of function phenotypes in vertebrate systems.**

MKS: Meckel syndrome; JBTS: Joubert syndrome; BBS: Bardet-Biedl syndrome; NPHP: Nephronophthisis; OFD4: Orofaciodigital syndrome IV; COACH: COACH syndrome; SLSN: Senior-Loken syndrome; LCA: Leber congenital amaurosis; CORD: Cone-rod dystrophy; MC: Macular degeneration; RP: Retinitis pigmentosa; KO: Knock-out; KD: Knock-down; MEFS: mouse embryonic fibroblast (From Gonçalves and Pelletier, 2017).

Protein	Associated disease	Sub-cellular localization
MKS1	MKS, JBTS, BBS	Centrosome, TZ
B9D1	MKS, JBTS, BBS	Basal body, TZ, axoneme (Garcia-Gonzalo et al., 2011; Chih et al., 2012)
B9D2	MKS	Basal body, TZ, nucleus (Dowdle et al., 2011; Chih et al., 2012)
TCTN1	JBTS	Basal body, TZ (Garcia-Gonzalo et al., 2011)
TCTN2	MKS, JBTS	Basal body, TZ, axoneme (Garcia-Gonzalo et al., 2011)
TCTN3	JBTS, OFD4	Basal body, TZ, axoneme (Garcia-Gonzalo et al., 2011)
CC2D2A	MKS, JBTS, COACH	Centrosome, TZ (Chih et al., 2012)
TMEM17		TZ (Chih et al., 2012)
TMEM67	MKS, JBTS, COACH, NPHP, BBS	Basal body, TZ, axoneme (Garcia-Gonzalo et al., 2011)
TMEM107	MKS, JBTS, OFD	TZ (Lambacher et al., 2016)

TMEM216	MKS, JBTS	Basal body, TZ, axoneme, Golgi (Lee et al., 2012)
TMEM231	MKS, JBTS	TZ (Chih et al., 2012)
TMEM237	JBTS	TZ (Huang et al., 2011)
NPHP1	JBTS, NPHP, SLSN	TZ, cell junctions (Garcia-Gonzalo et al., 2011; Sang et al., 2011)
NPHP4	NPHP, SLSN	Centrosome, TZ, cilium, cell junctions (Garcia-Gonzalo et al., 2011; Sang et al., 2011)
CEP290	MKS, JBTS, LCA, BBS, SLS	Centriolar satellites, centrosome, TZ (Garcia-Gonzalo et al., 2011; Sang et al., 2011)
NPHP5	SLS	Centrosome, cilium (Sang et al., 2011)
RPGRIP1L	MKS, JBTS, COACH	Centrosome, basal body, TZ (Delous et al., 2007; Garcia-Gonzalo et al., 2011)
RPGRIP1	LCA	Centrioles, TZ, axoneme (Shu et al., 2005)
RPGR	CORD, MC, RP	Centrosome, TZ (Shu et al., 2005)
LCA5	LCA	Centrosome, TZ, MTs (den Hollander et al., 2007)
AHI1	JBTS	Mother centriole, basal body, TZ (Hsiao et al., 2009)
CEP162		TZ, MTs (Wang et al., 2013)
TMEM138	JBTS	Basal body, TZ, axoneme (Lee et al., 2012)
JBTS17	JBTS	TZ (Damerla et al., 2015)

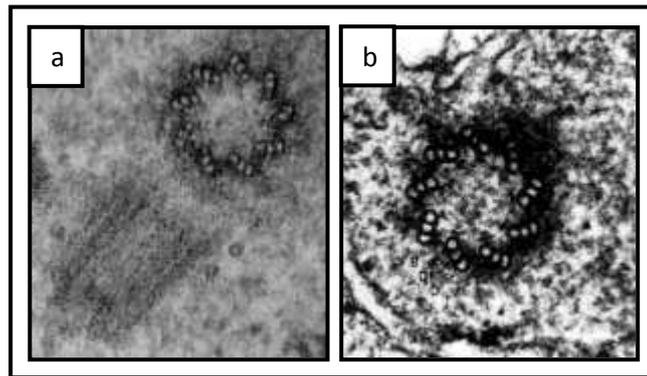
## II.3 Ciliogenesis

The following chapter will focus on the generation of the cilia, by describing the various structural elements of cilia such as centrioles and deuterosomes, basal bodies, transition zone, axoneme as well as proteins regulating ciliogenesis.

### II.3.1 Centrosomes

Centrosomes are the main MT organizing centre in animal cells; they are composed by two centrioles (mother and daughter centrioles), positioned orthogonally to each other, and pericentriolar material (PCM).

Centrioles are structures with cylindrical shape with 0.2  $\mu\text{m}$  of diameter and 0.5  $\mu\text{m}$  long. They are typically composed by 9 sets of short MT triplets (ABC), arranged in a cylinder (Fig.5). This structure is largely conserved in all the eukaryotes with some exceptions, for example in the worm *Caenorhabditis elegans* where centrioles are composed by a single set of MTs (Pelletier et al., 2006), or in *Drosophila melanogaster* embryos that show only nine doublets (Delattre, 2004).



**Fig.5. Centrosome electron microscopy.** (a) Electron micrograph showing two centrioles in cross section (upper right) and longitudinal section (lower left). (b) A high magnification electron micrograph showing a centriole in transverse section. Its wall is composed by 9 sets of triplets. Each triplet is composed of 3 subunits: a, b, and c, which form a MT.

### *II.3.1.1 The dual function of the centrosome*

The centrosome is responsible for both the organization of the MT network during cell division and for the formation of a primary cilium after migration to the plasma membrane and maturation into a basal body (Tang K. Tang, 2013; Loncarek and Bettencourt-Dias, 2018).

#### *II.3.1.1.1 Cell division*

During S phase, the centrosome duplicates, then each of the centrosomes migrates to the opposite cell poles, forming a mitotic spindle that is going to be responsible for the segregation of the sister chromatids in the two daughter cells. Defective duplication of the centrosomes, defective duplication of the chromatids or defective synchronization between the two will lead to defects in the DNA segregation into the daughter cells. This is a cause of genomic instability and loss of genome integrity.

An essential process that occurs at the end of mitosis corresponds to the disengagement of the two centrioles. It is a required step for the future cell division and is tightly regulated, especially through PLK1 and Separase. Pericentrin and CEP215 are pericentriolar proteins that are involved in the engagement of the centrioles. Pericentrin is cleaved by the protease Separase and phosphorylated by the kinase PLK1. This cleavage by separase promotes the removal of CEP215 and leads to the disengagement of the centrioles (Arbi et al., 2018). Mechanisms of disengagement between centrioles and between sister chromatids during cell replication are very similar. They both require to be tightly synchronized with specific components of the cell cycle and probably share some common regulators (Arbi et al., 2018).

#### II.3.1.1.2 Cilia formation

When exiting from cell cycle and becoming quiescent, some cells develop a primary cilium. To that end, the mature centriole of the centrosome (the mother centriole) migrates to the plasma membrane where it docks as a basal body (BB), allowing the elongation of an axoneme and resulting in the generation of a primary cilia. If the cell should re-enter the cell cycle, this cilium must be first resorbed (Arbi et al., 2018).

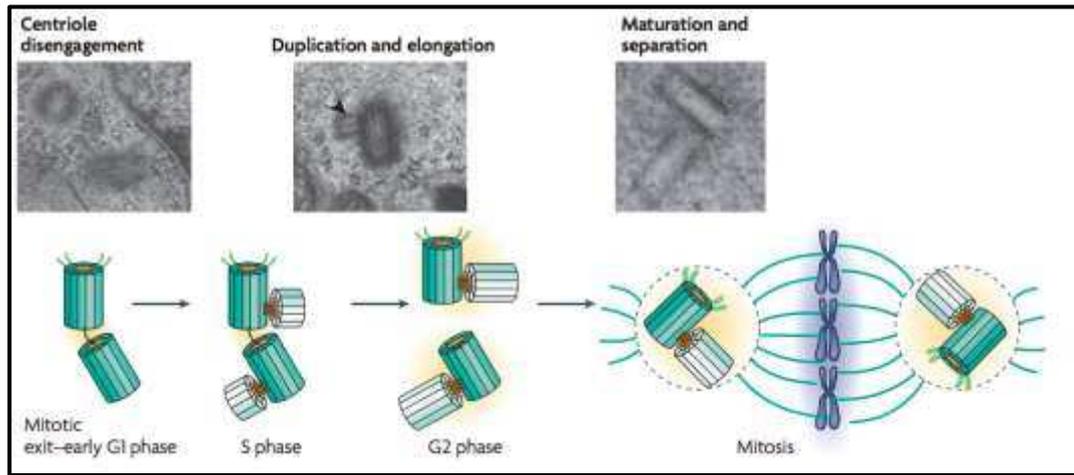
Some specialized cells become multiciliated after an extensive multiplication of the centrioles resulting in the formation of hundreds of BBs.

### II.3.2 Centriole multiplication

For multiciliogenesis the cell must perform a massive centriole amplification, through two distinct pathways, one centriolar pathway, which is common with duplication occurring during cell division, and an “acentriolar pathway”, which is specific to centriole amplification.

#### II.3.2.1 *The “Centriolar” pathway or the “mother centriole pathway”*

The duplication of the centrosomes is a very tightly regulated process. During S phase, centrioles of the centrosomes get disengaged, losing their orthogonal arrangement, in order to be duplicated only once during cell cycle, then a new centriole (daughter centriole) is generated orthogonally to its parental one (mother centriole), forming a new centrosome composed by mother/daughter centriole. Each daughter cell is going to inherit only one mother/daughter composed centrosome. This process is known as centriole-dependent (CD) pathway (Tang K. Tang, 2013; Arbi et al., 2018) and is tightly regulated in cycling cells (Fig. 6).

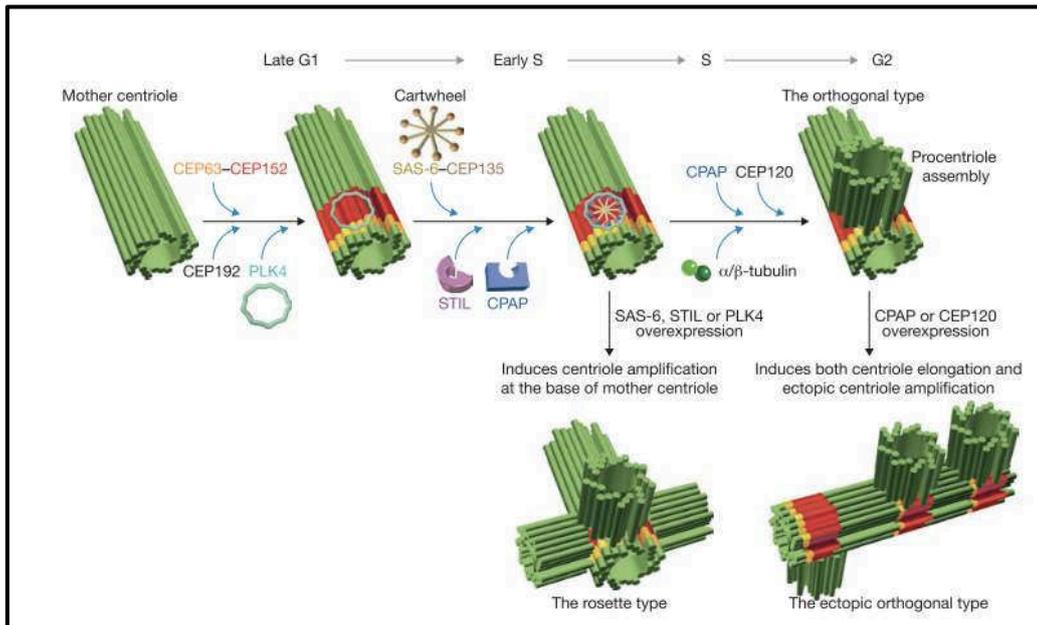


**Fig. 6. The centriole duplication cycle.** Electron microscopy micrographs from HeLa cells showing distinct steps in centriole duplication (also shown diagrammatically). The mother centriole is represented in dark green showing appendages. Daughter centrioles are shown in light green. At mitotic exit–early G1 phase, centrioles in a centrosome lose their orthogonal configuration. Next, duplication starts in late G1–S phase with the nucleation of daughter centrioles (see electron micrograph; the arrowhead shows a procentriole). The procentrioles elongate fully by late G2 phase or by the beginning of G1 phase of the next cell cycle. Last, maturation and separation of both centrosomes occur at the G2–M transition by the acquisition of maturation markers, the recruitment of pericentriolar material (PCM; yellow) and an increase in microtubule-organizing centre (MTOC) activity. From (Bettencourt-Dias and Glover, 2007).

Expression of molecules such as PLK4, SAS6 and STIL is related with the control of the copy number of the centrosome per cycle and the launching of the duplication process (Arbi et al., 2018). PLK4 is the main kinase regulating this process and it is recruited at the mother centriole by CEP152 which forms a ring like structure together with CEP63 (Zhao et al., 2013).

Three types of CD duplication and centriole amplification have been described, (recapitulated in Fig. 7) (Tang K. Tang, 2013):

1. The orthogonal (canonical centriole duplication) where a new centriole arises orthogonally to its mother centriole.
2. A rosette type replication where multiple new centrioles arise around the extremity of the mother centriole.
3. Ectopic orthogonal type where supernumerary centrioles are formed and vertically aligned along the axes of already pre-existing centrioles.



**Fig. 7. Multiple types of centriole-dependent centriole duplication or amplification in human cycling cells.** In the orthogonal type (canonical centriole duplication), the onset of centriole assembly is triggered by the activation of PLK4 and recruitment of CEP63 and CEP152 to the mother centriole. CEP63 first forms a ring-like structure with CEP152 at the surface of the proximal end of the mother centriole. PLK4 interacts with CEP152 and is recruited to the mother centriole by CEP152 and CEP192. STIL and the cartwheel proteins SAS6 and CEP135 are then recruited to the base of the nascent procentriole during the late G1 and early S phases. CEP135 directly binds to SAS6 and CPAP, linking the cartwheel to the outer microtubules. CPAP then cooperates with CEP120 to promote the assembly of nine triplet microtubules during centriole biogenesis. Overexpression of PLK4, SAS6, or STIL induces the rosette type of centriole amplification, whereas excess CPAP or CEP120 induces not only the ectopic orthogonal type of centriole amplification but also overly long centrioles in cycling cells. From (Tang K. Tang, 2013).

In the CD pathway, the molecules CEP-63 and CEP-152 form a complex surrounding the mother centriole, this complex recruits PLK4, an initiator of centriole formation. PLK4 binds and phosphorylates STIL allowing the association with SAS6. These three proteins form the cartwheel perpendicularly to the proximal wall of the mother centriole. The cartwheel is the structure that is going to guide the assembly of a procentriole that is going to elongate giving rise to the daughter centriole, driving the 9-fold symmetry (Leidel et al., 2005; Nakazawa et al., 2007; Guichard et al., 2010; Tang K. Tang, 2013; Arquint and Nigg, 2016; Loncarek and Bettencourt-Dias, 2018).

The 2nd and the 3rd type of centriole amplification (Rosette and Ectopic) lead to the generation of multiple centrioles and are related to the over expression of PLK4, SAS6 and STIL (Kleylein-Sohn *et al.*, no date; Strnad *et al.*, no date; Habedanck *et al.*, 2005; Leidel *et al.*, 2005; Rodrigues-Martins *et al.*, 2007; Tang *et al.*, 2011; Arquint *et al.*, 2012; Vulprecht *et al.*, 2012; Ohta *et al.*, 2014; Arquint and Nigg, 2016).

### II.3.2.2 Acentriolar pathway

An additional mode of centriole amplification occurs in MCCs together with CD pathway. The precursors of these MCCs, are able to produce from 30 to 300 new centrioles (Spassky and Meunier, 2017).

In the acentriolar pathway (also called Deuterosome-Dependent (DD) pathway) the centriole replicates “de novo” without the mother centriole and hundreds of centrioles are generated thanks to specific cytoplasmic organelles called deuterosomes (Eszter K. Vladar and and Tim Stearns, 2007; Zhao et al., 2013; Arbi et al., 2018). The DD pathway will be described more extensively below.

### II.3.3 The deuterosome

The deuterosome is a non-MT based organelle where centrioles can replicate (via the acentriolar pathway). These structures were initially described by Sorokin as electron-dense bodies (Sorokin, 1968). The DD pathway is the major source of centriole biogenesis in MCCs.

Even though deuterosomes were known since 1968, their function and molecular composition remained elusive until 2013 when Dehring (KlosDehring et al., 2013) identified the protein CCDC78 (coiled-coil domain containing protein 78) as a component of the deuterosome. They showed in embryonic skin of *Xenopus laevis* that the lack of this protein in MCCs was directly affecting the number of amplified centrioles, thereby influencing the fluid-flow function of these cells. Dehring et al also showed that: (1) some centriole duplication regulators, such as SAS6, CEP152 and PLK4, co-localized with the observed deuterosomes; (2) an over expression of CEP152 led to an increased number of centrioles and MCCs size, while a lack of CEP152 impeded the recruitment of PLK4 to the deuterosomes, CCDC78 being still detectable.

The same year, Zhao’s group (Zhao et al., 2013) identified in murine tracheal epithelial cells (MTECs) another protein component of the deuterosome, DEUP1 (CCDC67), a paralogue of CEP63, an actor of the centriole duplication in the (CD) pathway. Zhao showed an interaction between DEUP1 and CEP152 during the DD pathway.

Another important player for the deuterosome formation is the Cyclin O (CCNO). This protein localizes at the level of the deuterosome, and Funk ‘s group (Funk et al., 2015) showed that its deletion in mice reduced the number of deuterosomes relative to controls. The gene encoding for CCNO was seen to be mutated in patients with impaired mucociliary clearance, these mutations led to the reduction of the number of motile cilia in the airways (Wallmeier et al., 2014).

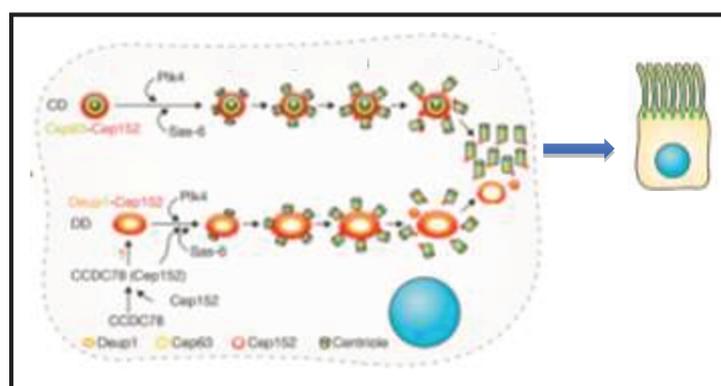
Recently our group described a new component of the deuterosome, CDC20B (Revinski *et al.*, 2018).

### II.3.4 Centriole amplification by the DD pathway

In the DD pathway, the formation of the deuterosome starts in a centriole-independent manner. The protein complexes of CCDC78-CEP152 and DEUP1-CEP152 are formed but are not surrounding the centriole, unlike in the centriolar pathway where the CEP63-CEP152 complex is surrounding the mother

centriole. DEUP1 seems to be the main skeleton of the deuterosome. The DEUP1–CEP152 complex may constitute major components of functional deuterosomes. As in the CD pathway, the next step consists in the activation of PLK4 and the recruitment of SAS6 forming the cartwheel the centrioles are elongated. Then the amplified mature centrioles migrate to the cell surface, mature into basal bodies, dock in the plasma membrane through a dense actin meshwork and elongate axonemes.

Despite all the differences between centrosome duplication during the cell cycle and multicilia formation, this duplication produces identical structures, so it is possible they are regulated by similar mechanisms (Fig8).



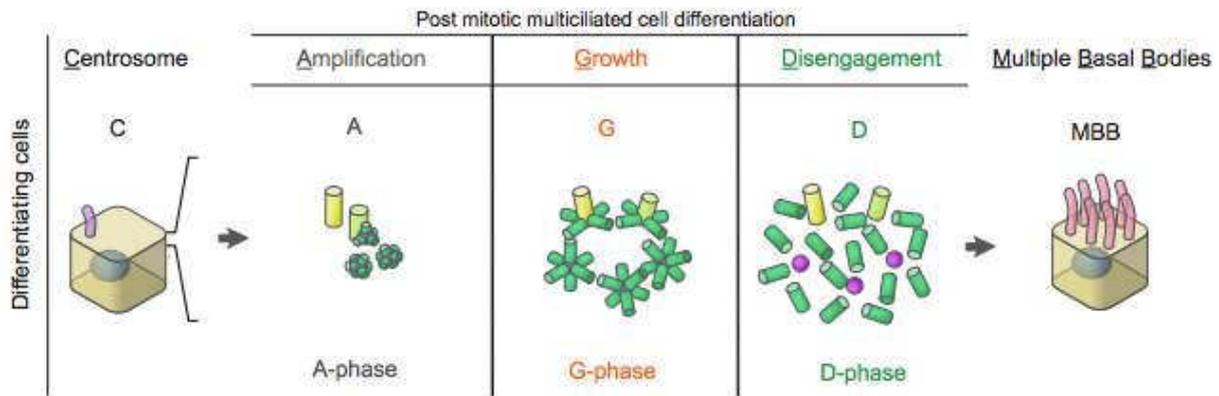
**Fig.8. Multicentriogenesis through CD and DD.** Model showing the CD and DD pathways in multiciliated cells. CD pathway is mediated by the Cep63–Cep152 complex. DD pathway is mediated by the Deup1–Cep152 and CCDC78–Cep152 complexes. From (Tang K. Tang, 2013).

The group of A. Meunier and N. Spassky revealed that centriolar and acentriolar pathways are both orchestrated by the centrosome. By a combination of 3D-SIM (3D structured illumination microscopy), EM (electronic microscopy) and live imaging, they have studied the massive amplification of centrioles that occurs in multiciliated ependymal cells from mouse brain. Differentiation of these cells and generation of hundreds of centrioles are orchestrated by both pathways. They described the accumulation of the deuterosomal proteins CCDC78 and DEUP1 at the daughter centriole and they could detect the release of “halos” from this daughter centriole, these “halos” presented clusters of DEUP1 in the center. A closer look into these “halos” by 3D-SIM revealed the expression of CETN2 subunits, that were, in fact procentrioles organized around spherical deuterosomes (AI Jord et al., 2014).

By studying the acentriolar pathway of centriole duplication (DD pathway), they then determined that the deuterosomes were also emanating from the wall of the daughter centriole, rather than being generated “de novo” in the cytoplasm as it was previously thought.

The group of Meunier described different sequential phases for centriole amplification in MCCs of mouse brain. First phase of amplification (represented in figure by A) where the procentrioles form around the deuterosome platforms. Second phase of growth (represented in figure by G) where the procentrioles grow synchronously, surprisingly at this stage the centrosomal centrioles segregate and the deuterosomes are

displayed all along the nuclear surface. Third phase of disengagement (represented in figure by D) centrioles detach from the deuterosomes and migrate and dock in the membrane becoming BBs nucleating the motile cilia (Fig. 9) (Al Jord *et al.*, 2017).



**Fig. 9. Post mitotic multiciliated cells differentiation.** A-Phase: amplification, G-Phase: growth, D-Phase: Disengagement. From (Al Jord *et al.*, 2017)

The cell cycle in dividing cells is controlled by the mitotic oscillator, centered on the cyclin-dependent kinase 1–anaphase-promoting complex/cyclosome (CDK1-APC/C) axis (where Cycling B1 and CDC20 are mitotic co-regulators) (Yang and Ferrell, 2013).

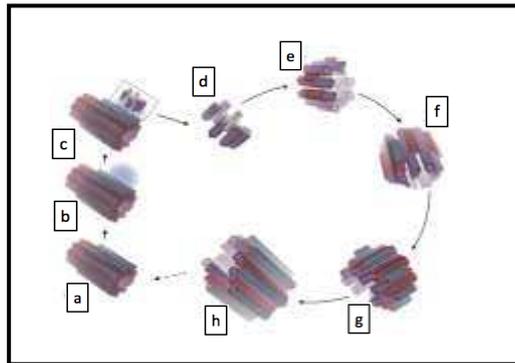
In cycling cells, centriole dynamics mirrors this spatiotemporal pattern where the duplicated centrosomes separate along the nuclear membrane at the G2/M transition of the cell cycle. Due to this similarities Meunier’s group tested whether the different regulators of the cell cycle had a role in the different phases of centriole amplification in MCCs precursors. They showed the involvement of some of this cell cycle regulators at different stages of centriole duplication. For instance, CDK1, Cyclin B1, PLK1, KI67, Cyclin B1 (pSer126) Lamin A+C (pSer392) APC 3 (pSer426) showed immunoreactivity for the transition from the phase A to the D phase. CDC20 and Histone 3 (pSer1) appeared in the transition between G and D phase, and Vimentin showed at the D phase. With this study they revealed that, in MCC precursors, the oscillators controlling the cell cycle were driving the ordered progression of the centriole amplification phases (Al Jord *et al.*, 2017).

### II.3.5 Centriole elongation

In all centriole duplication/amplification pathways, the generation of the new centriole starts with the cartwheel formation. After the cartwheel is formed, the procentriole starts the elongation of the MTs until complete assembly of the new centriole. MTs are composed by dimers of tubulin  $\alpha$  and  $\beta$  that form protofilaments which assemble. The cartwheel structure disappears after formation of a new procentriole,

meaning that cartwheel is only functional for the generation of the new procentriole but not for its maintenance.

The MT that attaches first to the cartwheel is called MT A and is nucleated by a gamma tubulin ring complex ( $\gamma$ -TuRC)-like structure, this structure appears like a cap closing the proximal end of the MT and this cap disappears in the mature centriole. The MTs B and C are next generated using the MT A and B as a template respectively. MT B and C do not have this cap-like structure and they elongate bidirectionally by both extremities (distal and proximal) (Fig. 10) (Guichard et al., 2010).



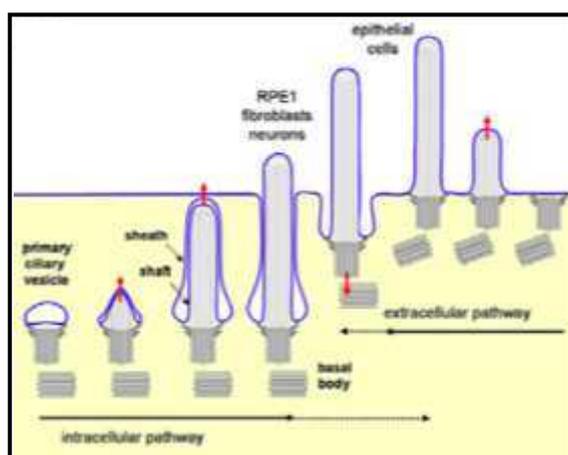
**Fig. 10. Model of human procentriole assembly.** (a) Mature centriole. (b, c) The first step of procentriole assembly, formation of a stalk and the central hub of the cartwheel. (c) The cartwheel assembles and organizes the procentriolar wall, A- MTs start growing from the  $\gamma$ -TuRCs. (d) Enlargement of the procentriole. (e) B-MTs start growing from the wall of A-MTs. (f) B-MTs grow bidirectionally, and C-MTs start growing from the wall of the B-micro- tubules. (g) C-MTs grow bidirectionally until the B- and C-MTs reach the proximal end of the A-MT. (h) elongation of the MT(Guichard et al., 2010).

MT elongation is controlled by several proteins including CEP97, which acts as a negative regulator, acting via a recruitment of the centriolar coiled-coil protein of 110KDa (CP110), which caps the distal extremity of the MTs and impairs further elongation (Spektor et al., 2007). Because localization of CP110 at the distal end of the MTs acts as a negative regulator of elongation, further elongation to generate an axoneme can only take place once CP110 has been removed. Accordingly, it has been shown that ectopic expression of CP110 prevents cilia formation. Surprisingly, it turned out that a knockdown of CP110 does not trigger cilia elongation but instead leads to a further elongation of the centrioles (Yadav et al., 2016). Using a mouse model in which Cp110 was lacking, Yadav's group described an impaired docking of the BB to the plasma membrane, through a delocalization of the proteins related to the formation of the appendages in the BB, suggesting that CP110 can also act as a positive regulator of ciliogenesis (Yadav et al., 2016).

### II.3.6 Basal bodies (BBs)

The structure of a primary cilia consists basically in an axoneme composed by 9 pairs of MTs that elongate from a modified centriole, called Basal Body (BB), which is docked in the plasma membrane by transition fibers. Docking of the BB in the membrane requires a series of appendages (Paintrand et al., 1992) that are essential not only for the docking but also for anchoring to the cytoplasmic MTs (Bornens, 2002). Several proteins regulate the function of these appendages, cenexin (ODF2), ninein (NIN), coiled-coil and C2 domain-containing protein 2A (CC2D2A),  $\epsilon$ -tubulin and centriolin (CNTRL).

The process of ciliogenesis after the docking of the BB to the cell surface continues with the interaction of a ciliary vesicle to the BB, which grows after the fusion of smaller vesicles. The fusion of this vesicle with the cell membrane generates the ciliary pocket and leads to the cilia protrusion from the cell surface as showed in the figure 8. The ciliary pocket is an endocytic membrane domain from where clathrin-coated vesicles are formed, which is also involved in the interaction with the actin-based cytoskeleton (Fig. 11) (Molla-Herman et al., 2010).



**Fig.11. Formation of the ciliary pocket.** In the intracellular pathway (left), a primary ciliary vesicle first interacts with the mother centriole, the axoneme then grows within this vesicle forming then the shaft and the sheath. The latter eventually fuses with the plasma membrane, allowing the distal part of the cilia to interact with the extracellular milieu. In the extracellular pathway (right), the mother centriole directly docks to the plasma membrane, with the cilium growing directly in the extracellular milieu. (Molla-Herman et al., 2010)

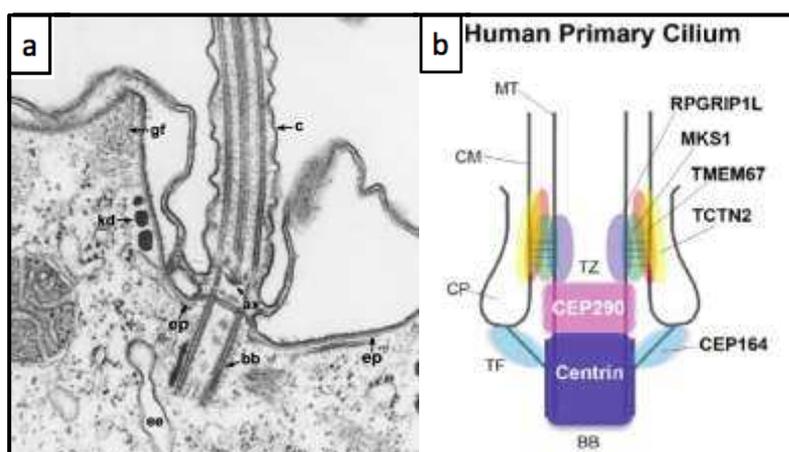
### II.3.7 The transition zone (TZ)

An important part of the cilium is the transition zone (TZ), which corresponds to the proximal portion of the axoneme, distal to the BB. This TZ is a “gatekeeping”. The cilia is an organelle shared by the intra and the extracellular compartments of a cell, here the TZ complex maintains the stability of the BB and the axoneme

in those different cell compartments and it is required for the compartmentalization of the cilium, controlling the protein composition of the cilia. Proteins of the TZ regulate ciliogenesis and ciliary membrane composition (Garcia-Gonzalo et al., 2011). This ciliary domain has gained considerable attention in recent years since most of its components have been associated with human diseases (Gonçalves and Pelletier, 2017).

The TZ is the area where the cilium starts elongating from the distal part of the BB where only the MTs A and B are starting to grow the axoneme.

The composition of the TZ has been extensively studied by approaches such as co-immunoprecipitations or BioID (proximity-dependent biotin identification) coupled to mass-spectrometry. These works establish molecules such as CEP162, RPGRIP1L, AHI1 and LCA5 as bona fide TZ components, and show their interactions with centriolar satellite proteins such as PCM1 and KIAA0753. Composition of TZ includes proteins such as TCTN1, TCTN3 and seven MKS proteins (MKS1, TMEM216, TMEM67, CEP290, CC2D2A, B9D1 AND TCTN2). These works also demonstrate that some protein complexes, such as B9D1-B9D2, NPHP1-NPHP4 or CEP290-NPHP5, are assembled before their incorporation into the TZ (Fig. 12) (Garcia-Gonzalo et al., 2011; Gupta et al., 2015).



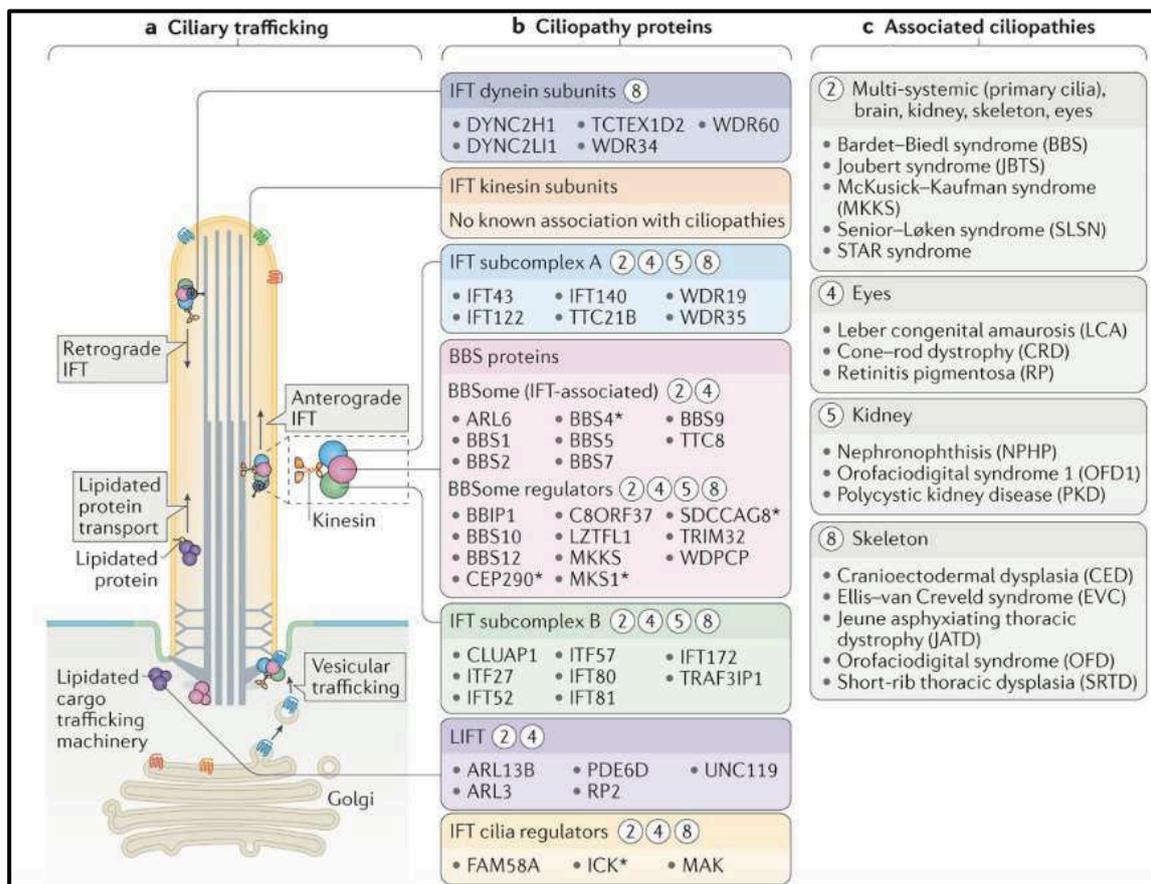
**Fig. 12. Transition zone, structure and protein composition.** (a) Electron micrograph of a longitudinal section of a cilium. (b) Localization of transition zone and basal body components in human cilia and flagella. MT – microtubules; CM – ciliary membrane; CP – ciliary pocket; TF – transition fibers; BB – basal body; TZ – transition zone; BP – basal plate; TP – terminal plate. From (Gonçalves and Pelletier, 2017)

### II.3.8 Intraflagellar transport (IFT)

IFT is the transport of molecular motors and protein complexes through the cilium or flagella. Thanks to this machinery, tubulins and other cilium components are transported all along the cilium for its assembly or maintenance (Czarnecki and Shah, 2012). The absence of this IFT machinery prevents the construction of

most cilia and flagella. Incidentally, many human genetic diseases are caused by perturbation of the IFT components leading to mal functioning or defects in the structure of the cilium (Bertiaux et al., 2018). There are two essential motor proteins in charge of the movement of the IFT along the cilium, the kinesin II and the dyneins.

Kinesin II is in charge of the transport of IFT complexes, these IFT complexes are “charged” with some other proteins that have to be transported to the tip of the cilium, such as tubulin for the axoneme assembly. The sense of the transport of the Kinesin II molecules is always from the base to the tip of the cilium. Also the dynein motor proteins are going to be transported in these IFT complexes from the base to the tip of the cilium. Once arrived to the tip, the protein complexes are loaded on the tip of the cilium. Now the “discharged” IFT are going to start the transport of other protein complexes not needed anymore in the tip of the cilium to the base, in this transport are the dyneins in charge of this retrograde transport (Fort and Bastin, 2014; Fort et al., 2016).



**Fig. 13. Links between ciliary trafficking and ciliopathies.** (a) The functional components of two ciliary trafficking pathways: intraflagellar transport (IFT) and lipidated protein intraflagellar targeting (LIFT). Ciliary proteins are trafficked from the Golgi or cytosol to the base of the cilium, after which they are transported into the ciliary compartment. IFT modules that mediate trafficking include anterograde and retrograde motors, IFT complexes, and an accessory module that contains Bardet–Biedl syndrome (BBS)

proteins (the BBSome). (b) Ciliopathy proteins that constitute, or are regulators of, the IFT and LIFT trafficking systems. Circled numbers indicate which ciliopathies (listed in part c) result from defects in these ciliary trafficking components. The asterisks indicate proteins that are also localized to other ciliary regions during ciliogenesis or ciliary trafficking. (c) Ciliopathies that result from defects in ciliary trafficking grouped into categories according to the tissues affected (Reiter and Leroux, 2017).

## II.4 Regulatory mechanisms of multiciliogenesis

Multiciliogenesis occurs in the airway epithelium but also in epithelial cells from fallopian tubes and the tubulus efferens of the testis, in radial glia in the brain ventricles (ependymal cells) where they sustain cerebrospinal fluid circulation.

### II.4.1 Role of the Geminin super family

The geminin superfamily consists of three proteins: (1) geminin, (2) geminin coiled-coil domain-containing protein 1 (GEMC1) and (3) multicilin (MCIDAS). The firstly identified member of the family was geminin, GEMC1 and MCIDAS being then identified based on their homology with geminin. All three proteins share several conserved coiled-coil domains, including a conserved domain located in the carboxy-terminal part of these coiled-coil regions. MCIDAS and geminin also share a common peptide sequence at the N-terminus of their coiled-coil domain. MCIDAS and GEMC1 share a common domain located in their carboxy-terminal region that is not present in geminin. These homologies and the fact that geminin is present in all metazoans, while MCIDAS and GEMC1 are only in vertebrates draws a picture of evolutionary divergence for this protein family (Arbi et al., 2018). Initially, they have been described as regulators of DNA-synthesis (Mcgarry and Kirschner, 1998; Balestrini et al., 2010; Pefani et al., 2011).

MCIDAS and GEMC1 are among the most upstream activators of multiciliogenesis. MCIDAS was the first of the geminin family to be described as a regulator of multiciliogenesis (Stubbs et al., 2012).

Stubbs et al. revealed that MCIDAS expression was regulated via Notch inhibition, promoting MCC differentiation in *Xenopus*. They showed how MCIDAS acts as a transcriptional regulator of genes required for centriole amplification and formation of motile cilia (Stubbs et al., 2012).

Ma et al. described the role of MCIDAS in the biogenesis of centrioles through the formation of a protein complex with E2F4 and E2F5 and DP1 (TfDp1 transcription factor that heterodimerize with E2F proteins to enhance their DNA-binding activity and promote transcription from E2F target genes). This resulted in the activation of the expression for genes related to (1) basal body assembly (PLK4, CEP152, STIL and DEUP1) and (2) transcription factors related to multiciliogenesis (FOXJ1, RFX2 and cMYB) (Ma et al., 2014; Arbi et al., 2018)

Experiments of *Mcidas* overexpression in developing cortex of mouse triggered the exit of cell cycle of progenitor cells and promoted multiciliated ependymal cells differentiation. On the other hand, depletion of *Mcidas* led to a reduction by half of the number of MCCs (Kyrousi et al., 2015).

GEMC1 expression has been localized at very high levels in tissues harbouring MCCs (F. Zhou et al., 2015; Kyrousi et al., 2015; Arbi et al., 2016; Terré et al., 2016). Experiments by Kyrousi and colleagues showed that an overexpression of GEMC1 maintained a part of the progenitors in an active cell cycle state while the rest of them overcome differentiation into ependymal MCCs. On the opposite, a depletion of GEMC1 led to a 50% reduction in the number of MCCs (Kyrousi et al., 2015).

Zhou et al. reported the role of *Gemc1* in MCCs formation in Zebrafish embryos. They used a morpholino approach and observed that *Gmc1* depletion led to an almost complete absence of MCCs in the pronephros. However, over expression of *Gemc1* did not give rise to an increase in MCCs (F. Zhou et al., 2015).

GEMC1 is essential for the proper expression of genes involved in multiciliogenesis but is not sufficient to trigger this process, except in *Xenopus* where the ectopic expression of *Gemc1* increases the levels of MCCs. This effect is perhaps related to a very high level of expression that can be obtained after an expression in *Xenopus*.

*Gemc1* depletion in mouse triggers hydrocephaly, growth impairment and infertility. Genetic depletion of this gene resulted in impairment of the MCC development because of the drastic reduction of genes involved in multiciliogenesis including *Mcidas*, *Foxj1*, *Ccno*, *Ccdc78* and *Deup1* with absence of MCCs in the airways, reproductive tissues, brain and defects in spermatogenesis and also defects in growth. *Gemc1* depleted mice died few days after birth due to defects in airway mucus clearance (Arbi et al., 2016; Terré et al., 2016).

To reveal the position of these genes in the regulatory cascade of multiciliogenesis, experiments have been done including knockout mice models of *Gemc1* and *Mcidas*. These experiments measured the protein expression and mRNA expression levels of different sets of genes involved in multiciliogenesis such as the transcription factors *FoxJ1* and *Myb*, and genes such as *Ccno*, *Deup1* and *Ccdc78*. The results showed that after depletion of *Gemc1* or *Mcidas* the expression of all these genes was highly decreased (F. Zhou et al., 2015; Kyrousi et al., 2015; Arbi et al., 2016)..

Moreover, the depletion of *Gemc1* or *Mcidas* revealed that *Gemc1* acts upstream *Mcidas*, overexpression of *Gemc1* increased the expression of *Mcidas*, while overexpression of *Mcidas* did not increase the expression of *Gemc1* (F. Zhou et al., 2015; Kyrousi et al., 2015; Arbi et al., 2016; Terré et al., 2016).

Work by Zhou and Kyrousi determined that *Gemc1* acts downstream the Notch signalling pathway, as inhibiting Notch in zebrafish embryos increased the level of MCCs, and invalidating *Gemc1* with morpholinos

decreased the levels of MCCs. As expected, ectopic expression of Notch inhibited MCC development and this effect was counteracted by *Gemc1* overexpression (F. Zhou *et al.*, 2015; Kyrousi *et al.*, 2015).

Terré and Arbi demonstrated that GEMC1 is able to form a complex with E2F4/5-DP1 and MCIDAS through different domains (Arbi *et al.*, 2016; Terré *et al.*, 2016).

#### II.4.2 E2f family of transcription factors

This family has primarily been implicated in the regulation of genes required for proliferation and passage through the cell cycle, *E2F4/5* are cell-cycle-repressing transcription factors. Depletion of *E2F4* alone in Mouse Embryonic Fibroblasts (MEFs) showed no alteration of the cell cycle but depletion of *E2f4* and *E2f5* together led to these defects showing a synergistic effect of *E2f5*. However, mice deficient for *E2f4* showed different defects such as chronic rhinitis, infertility and early death after birth caused by the rhinitis and infections caused by opportunistic bacteria (Humbert *et al.*, 2000), symptoms linking *E2f4* to multiciliogenesis. The loss of *E2f4* also led to an epithelium with basal and goblet cells, like in controls. In contrast, MCCs, usually abundant on all the epithelial surface in controls, were replaced by a columnar secretory cell type, the depletion of *E2f4* in mice revealed the decreased expression of genes involved in centriole duplication and early multiciliogenesis such as the transcription factor *Foxj1* (Mori *et al.*, 2017). The authors revealed that the loss of MCCs was not due to apoptosis nor proliferation. Accordingly, nasal epithelium of *E2f4* deficient mouse showed an increase in mucins secretion (Danielian *et al.*, 2007).

Ma *et al.* tested whether the complex formed by MCIDAS and GEMC1 with E2F4 and E2F5 was critical for differentiation. They generated truncated versions of E2F4 and E2F5 where 140 amino acids were lacking from the C-terminus. Expression of these truncated molecules in *Xenopus* embryos impaired centriole assembly and BB formation. RNA-sequencing performed to assess the genes affected by the truncated version of E2F4 together with coexpression of an inducible form of MCIDAS revealed that genes related to centriole duplication were altered such as: *Cent3*, *Cep135*, *Cep76*, *Plk4*, *Cep152*, *Sass6*, *Stil*, *Cenpj*, *Ccp110*, *Deup1*, *Cep63*, *Ccno*, *Plk1* (Ma *et al.*, 2014) and our re-analysis of the data showed also the presence of the recently described *Cdc20b* (Revinski *et al.*, 2018).

These results suggested that the MCIDAS/E2F4 association is required for the up-regulation of BB components and ciliogenic transcription factors. Together with the role of E2F4/5 in cell-cycle exit, MCIDAS/E2F4 interaction is proposed to induce the MCC terminal differentiation program (Meunier and Azimzadeh, 2016).

The groups of Lees and Cardoso studied the mechanisms related to E2F4 and E2F5 during multiciliogenesis. They showed that there is switch of E2F4 from the nucleus to the cytoplasm during multiciliogenesis, surprisingly this switch was essential for the induction of a transcriptional program for

centriole biogenesis and for centriole assembly. 3D-SIM revealed that during generation of MCCs, E2F4 was located at the procentrioles forming a ring-like structure with some centrosomal proteins (CEP152, PLK4, CEP63, DEUP1... among others) suggesting a role in assembly of the new centrioles. Co-immunoprecipitation experiments showed the interaction between E2F4 with DEUP1 (Danielian et al., 2016; Mori et al., 2017).

#### II.4.3 Foxj1 transcription factor

This transcription factor was described to control the generation of motile cilia by Brody, by showing that depleting *Foxj1* (first named as Hfh4 for Hepatocyte nuclear factor-3/forkhead homologue 4) in mice induced a lack of cilia in the airways and defects in left-right symmetry where half of the mice presented situs-inversus. Interestingly, the nodal cilium was present but not functional which suggests that these two types of cilia were generated following different pathways. In airways, the absence of *Foxj1* led to an impairment of BBs migration and docking into the membrane, but centriologensis was normal (Brody et al., 2000). *Foxj1* modulates the activity of RHOA and Ezrin that are important for actin cytoskeleton organization (Gomperts, 2004; Yu et al., 2008; Chevalier et al., 2015; Spassky and Meunier, 2017). Gomperts' group studied the role of *Foxj1* in the regulation of basal body anchoring to the apical cytoskeleton. They performed microarrays analysis on RNA extracted from lung tissues depleted for *Foxj1*. They reported a significant decrease of Calpastatin (protein with a role in cytoskeleton remodeling, and in numerous membrane fusion events, is a specific inhibitor of the protease calpain). Also, after *Foxj1* depletion, expression of Ezrin (linker molecule between the F- actin cytoskeleton and specific apical membrane proteins) was also decreased, which led them to relate this *Foxj1* to BBs in the mouse trachea (Gomperts, 2004).

Jacquet et al. tested the effect of depletion of *Foxj1* in in developing brain cortex in mice. They reported an absence of ependymal MCCs, but primary cilia were present in the apical surface of glial cells lining the brain ventricles, just like in wildtype animals. They identified candidate genes regulated by *Foxj1* using a microarray transcriptome analysis, revealing 55 genes regulated by *Foxj1* in the ependymal zone of the adult brain, including genes encoding cytoskeleton-associated molecular motors, and sperm/flagellar-associated proteins or MT-associated proteins. The list also includes members of the kinesin family of motor proteins (Kif6/9/27). These findings positioned *Foxj1* in the regulation of the expression of genes regulating the formation of the motility apparatus in the MCCS of the brain (Jacquet et al., 2009).

Studies in zebrafish confirmed that *Foxj1* is essential for motile cilia formation. *Foxj1* controls the formation of motile cilia through the regulation of the transcription of several genes involved in motile cilia biogenesis such as *Dnah9* and *Cetn2*. They showed the role of *Foxj1* regulating *Rfx2*, a transcription factor member of the RFX family which are also regulators of motile ciliogenesis. They found that *Foxj1* defective embryos showed a decrease in the *Rfx2* expression and contrary the overexpression of *Foxj1* driven *Rfx2* increased expression (Yu et al., 2008). However, there are other studies revealing a control of the RFX family

over *Foxj1* such as in the work of Didon where they revealed that *RFX3* can enhance *FOXJ1*-dependent transcription, and they described the interaction between *RFX2/3*–*FOXJ1* in human airways, suggesting a role of *RFX3* as a co-factor of *FOXJ1*. Thus RFX factors regulate core cilia genes and cooperate with *FOXJ1* to regulate multiciliogenesis (Didon *et al.*, 2013).

#### 11.4.4 Rfx family of transcription factors

This family has been identified as the second family of transcription factors regulating ciliogenesis in *Drosophila melanogaster*, *Caenorhabditis elegans* and vertebrates (Thomas *et al.*, 2010). It has been shown that RFXs are necessary for development of cilia in the sensory neurons in *Drosophila* (Dubruille, 2002). *RFX3* has been shown to be necessary for motile cilia biogenesis in ependymal cells (El Zein *et al.*, 2009) and in human airway epithelial cells (Didon *et al.*, 2013). In zebrafish, *Rfx2* was shown to regulate biogenesis of motile cilia (Liu *et al.*, 2007; Yu *et al.*, 2008).

RFX targets have been extensively studied and most of them are components of the IFT machinery, and all the BBS (Bardet-Biedl syndrome) proteins which are involved in the coordination of the IFT transport and regulation of the intra-cilia trafficking (Thomas *et al.*, 2010).

Chung *et al.* analyzed the role of *Rfx2* in *Xenopus*, they show by in situ hybridization the expression of *Rfx2* in all the tissues containing ciliated cells (Chung *et al.*, 2012). Other members of the family were also expressed in the neural tube (*Rfx1*, *Rfx3*, *Rfx4* and *Rfx5*). *Rfx3* was also expressed in MCCs in *Xenopus* skin together with *Rfx2* but there is a shift in the expression of *Rfx3* compared with *Rfx2* which is expressed earlier in the development. Depletion of *Rfx2* showed that this transcription factor is essential for the neural tube closing and cilia biogenesis. Analysis of this depletion in *Xenopus* skin revealed that *Rfx2* is not essential for the MCCs formation but it is essential for the motile cilia assembly in these cells. These depletion experiments also showed the role of *Rfx2* in the establishment of the Left/Right asymmetry by the regulation of the cilia in the node.

Quigley & Kintner analysed the interactions between *Rfx2* and *Foxj1*, performing RNAseq in *Xenopus* skin, and Chromatin-immunoprecipitation-sequencing (ChIP-seq). They revealed that *FOXJ1* was rarely bound to promoters of target genes if *Rfx2* was not present indicating that *FOXJ1* binding is stronger in the presence of *Rfx2* in promoters. To check if one of the functions of *Rfx2* is the stabilization of the binding of *Foxj1* to its targets, they knocked down *Rfx2* with a morpholino and performed ChIP-seq. Their results showed that, if in control tissue there were 15305 peaks identified as *FOXJ1* binding sites, in morphants, 6360 of these peaks were reduced 3-fold and 2084 were reduced 10-fold. They observed that the most reduced binding sites corresponded to promoters of genes implicated in MCCs biogenesis, meaning that *Rfx2* stabilizes the binding of *FOXJ1* to the promoters of its target genes (Quigley and Kintner, 2017).

#### II.4.5 Tumor protein (Tp)73

Homolog of *TP53*, *Tp73* is a member of the family of the tumor suppressor genes (*TP53*, *TP63* and *TP73*). It has been described as important regulator of neural stem cell maintenance and organization in the sub-ventricular zone (SVZ) germinal center (Gonzalez-Cano et al., 2015). The depletion of *Tp73* leads to defects in differentiation of ependymal cells generating immature cells, relating this defect in ependymal cell maturation to defects in cilia formation. More exhaustive analysis of ependymal cells in *Tp73* depleted mice showed the presence of clusters of BBs in the deuterosomes, meaning that the defects in cilia formation come from impairment of BB docking and cilia assembly.

Nemajerova et al. (2016, 2018) analysed the effect of the depletion of *Tp73* in mice airways. Depleted mice showed a severe phenotype of respiratory distress marked by chronic inflammation, macrophage infiltration and emphysema. Immunostaining analysis revealed a decrease in cilia number (cilia was almost absent). The fewer cilia that were still present had a decreased length in *Tp73* depleted mice compared to controls. They showed that the number of BBs generated was similar to controls but there were clear defects in BBs docking in the membrane and loss of polarity. These results situate the action of *Tp73* downstream to centriole duplication and BBs formation and upstream of BBs polarity and docking. Immunostainings revealed that TP73 was already expressed in FOXJ1 positive cells and that its expression was maintained in FOXJ1/RFX2/RFX3 positive cells. Moreover, depletion of *Tp73* resulted in a decreased expression of *FoxJ1*, *Rfx2* and *Rfx3* at mRNA and protein levels. They finally showed that ectopic expression of *MCIDAS* induced *Tp73* expression. These results indicate that *Tp73* acts downstream of *MCIDAS* and upstream of *FoxJ1*, *Rfx2*, *Rfx3*, and *Myb*. (Nemajerova et al., 2016, 2018). Similar results were obtained by Marshall and colleagues who provided evidence that *Tp73* is a direct regulator of *Foxj1*. Depletion of *Tp73* in mice airways led to a reduction in MCCs in the trachea and bronchia, a reduction in the BCs and an increase in SCs and GCs (Marshall et al., 2016).

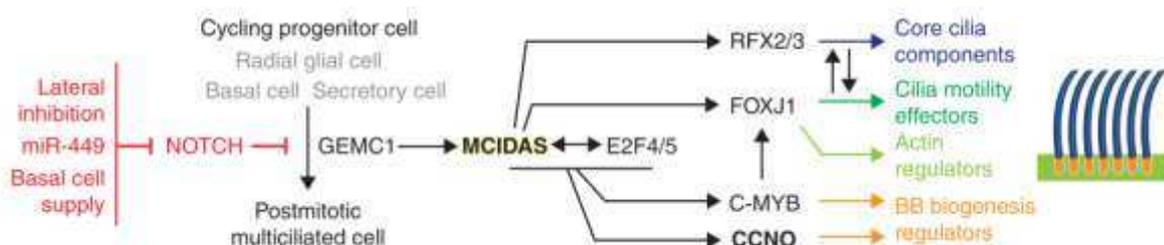
#### II.4.6 Myb transcription factor

Myb stands for Myeloblastosis proto-oncogene and is a transcription factor promoting the S phase of the cell cycle.

Tan et al. showed that Myb is expressed in developing MCCs in mouse airway epithelium just after the exit of cell cycle, and that it is necessary for the multiciliogenesis. They determined that Myb is acting upstream Foxj1 and downstream Mcidas and Notch in mouse and in *Xenopus*. Moreover, they revealed that Myb has a role in the regulation of centriole amplification during multiciliogenesis. Ectopic expression of MYB in *Xenopus* embryos lead to an increase in twice the number of MCCs (Tan et al., 2013).

Pan et al. analysed the role of Myb in human and mouse cultured primary cells from the airways. Following the expression of MYB through the process of differentiation, they detected Myb at early time

points during the differentiation. Its expression decreased drastically once the activation of the multiciliogenesis programme was established, as was determined by the centriole amplification program. They did not see expression of MYB in basal cells, thus positioning this transcription factor between the progenitor basal cells and the differentiated MCCs. Depleting Myb in their system led to a defect in centriole amplification and absence of cilia, together with a lack of FOXJ1 and RFX3 expression. They also revealed the role of Myb in the differentiation of secretory cells, after a reduction of Myb expression in SCGB1A1 positive cells (secretory cells). They tested the impact of Myb on secretory cell fate in a model of injury in mice cultured airway cells. They treated the mice with naphthalene, that specifically depleted the secretory cells in the tissue. Naphtalene treatment led to a reduction in Myb expression levels, during injury repair, they notice an increase in Myb expression and co-localization of MYB and SCGB1A1, meaning that Myb is regulating the differentiation of the secretory cells. Quantitative reverse transcription polymerase chain reaction (RT-qPCR) performed in Myb depleted tissues showed decrease in expression of mucous-cell marker genes such as Muc5ac, Spdef and Foxa3 (Pan et al., 2014).



**Fig. 14. Regulation of the multiciliated cells differentiation in vertebrates.** Notch inhibition triggers the activation of the regulatory cascade of multiciliogenesis which involves the geminin-related proteins GEMC1 and MCIDAS, the transcription factors E2F4/5, FOXJ1, FRX3/2 and C-MYB, and the protein cyclin-O (CCNO).

### III. *Airway epithelium development, homeostasis and regeneration*

---

The airway epithelium plays a critical role in the defence of the respiratory system against inhaled particles or pathogens avoiding damage or infections. It is replaced over time at a very slow-turnover, in order to maintain a correct cell balance and composition for a good functionality. However, after injury, this epithelium can rapidly regenerate and restore its function.

Respiratory diseases are the third leading cause of death in the industrialized world. The airways undergo many pathological changes in pulmonary disorders such as COPD, asthma or cystic fibrosis, where the epithelium is constantly under chronic injury conditions. This results in a pathological remodelling of the airways, leading to functionality loss, metaplasia, GC hyperplasia, hypoplasia, epithelial thickness, malignant transformation... Despite the importance of airway remodelling in lung disease, this process is still poorly understood, due in part to the lack of knowledge of the normal differentiation process.

Deciphering the process of differentiation, homeostasis and regeneration of the lung epithelia can have a tremendous impact in many life-threatening diseases (Rock, Rande and Hogan, 2010; Hogan et al., 2014; Volckaert and De Langhe, 2014).

Lots of different studies have been performed to this issue. With the development of techniques as cell lineage tracking and, more recently, single cell transcriptome, analysis can describe more comprehensively the process of differentiation of the airway epithelium during postnatal development, adult tissue homeostasis and repair.

#### III.1 Differentiation of the airway epithelium in normal homeostasis, development and in regeneration conditions after injury in human and murine models

##### III.1.1 Basal cells progenitors of the airway epithelium

BCs have been described as progenitors of the airway epithelium: Breuer et al. showed that BCs act as progenitors in the steady state in conducting airways, but are substituted by SCs as progenitor population for the maintenance of the epithelium in the smaller airways where BCs are rarer (Breuer et al., 1990). This is consistent with the studies performed by Reynolds in 2000 (Reynolds et al., 2000) showing that an ablation of SCs in mice causes a loss of MCCs in the terminal bronchioles but not in proximal airways. Different populations of cells affect therefore the fate of the epithelium, and these differences depend of regions of the airways that are considered.

BCs constitute a multipotent progenitor population that gives rise to other BCs, SCs and MCCs, playing an important role in the regeneration of the epithelium. Experiments of Hong and Gosh revealed a BC subset expressing KRT5 and KRT14 responsible of the regeneration of the epithelium after injury in mouse through naphthalene-induced ablation of the SC population (Kyung U Hong *et al.*, 2004; Kyung U. Hong *et al.*, 2004) or in increased proliferating cells due to remodelling events in cultured primary cells from cadaveric tracheobronchial tissue (Ghosh *et al.*, 2013).

Edith Puchelle and Christelle Coraux identified the BC population as the responsible of the differentiation a regeneration of the epithelium, they FACS-sorted BCs and columnar cells separately and analysed the properties of these cell suspensions. They showed that only cells of the BC suspension, once seeded in plastic or in denuded rat tracheas, were able to attach, proliferate and differentiate. They also studied the telomerase activity of the two cells suspension and they only detected active telomerase in the BC suspension, which is a feature of stem cells (Rodolphe Hajj *et al.*, 2007).

To characterize better the BC population and improve cell lineage tracing during epithelial homeostasis and regeneration, the group of Hogan genetically tagged KRT5+ cells in adult mouse under tamoxifen induction, and they reported that just after lineage induction only BCs were tagged. Over time, the percentage of tagged BCs declined, while the number of tagged SCs and MCCs increased (Rock *et al.*, 2009). Since the number of tagged SCs was considerably bigger than the number of MCCs, two possible hypotheses were raised: (1) MCCs have a longer life span than SCs, so BCs give rise to more SCs than MCCs in a certain period of time; (2) BCs give rise to SCs that slowly transitioned to ciliated cells. The latter hypothesis indeed agrees well with the results of Plopper's group (Evans *et al.*, 2001) and with their own study (Rawlins *et al.*, 2009) on the role of SCs during homeostasis and repair of lung airway. In the same way, they analysed the cellular composition and behaviour of BCs in postnatal growth and showed that BCs were able to self-renew and generate SC and MCCs.

Hogan's group also performed experiments in mice after injury using sulfur dioxide (SO<sub>2</sub>) and analysed the repair model with their KRT5 genetically tagged mouse line (Rock *et al.*, 2009). SO<sub>2</sub> caused an extensive damage in the airways. During repair they analysed the BC population and they showed that this population was highly proliferative and gave rise to patches of cilia and SCs. In contrast with the behaviour showed during homeostasis, the process of regeneration after injury by BCs gave rise to more cilia than SC cells. This plasticity suggests that the fate of the progeny of BCs is really influenced by the tissue conditions.

Already in 1986 Evans *et al.* used a model of "mild" injury in rats that consisted in the exposure to NO<sub>2</sub>. This kind of injury led to the destruction of the more outer cell layer, meaning that only GCs and MCCs were ablated. After injury they studied the regeneration of the epithelium, focusing on the region of the bronchi with diameters between 1.47 and 1.89, a level where BCs are still present. They revealed that the SC

population was the responsible of the restoration of the ablated cell types after the injury and not the BCs (Evans et al., 1986).

Teixeira and colleagues used a statistical lineage tracing model based in the study on mitochondrial mutations in the epithelial cells as a marker to analyse clonal expansion. They showed that in the human upper airways, the model of homeostatic maintenance of the epithelium was consistent with a model where a unique, multipotent progenitor cell is loosed and replaced in a stochastic manner, and this progenitor cell would be a BC (Teixeira et al., 2013).

Collectively, these results suggest that there are at least two populations of progenitor cells: one BC population that maintains the homeostasis at long-term and self-renews, and SCs that acts as facultative progenitors.

The statistical model of cell lineage developed by Texeira and colleagues revealed that the model of proliferation of the BC population led to three different outcomes for the daughter cells, where the division of the progenitor cell could give rise to two progenitor cells or to two cells that were committed to differentiation (symmetrical division), and could give rise to an asymmetric division generating one progenitor cell and one cell prepared to differentiate, this both kinds of division were also described later on by the group of Rawlins (Teixeira *et al.*, 2013; Watson *et al.*, 2015).

Boers's group in 1998 described the heterogeneity of the BC population describing the presence of BCs expressing KRT5 and KRT14 and other called "paraBCs" expressing KRT13. They did a measurement of the number of BCs expressing KRT5 and KRT14 and paraBCs expressing KRT13 in healthy human lungs and they quantified also the proportion of these cells contributing to the proliferative fraction of the epithelium in proximal and distal airways. In the proximal fraction of the airways there were 31% of BCs, 51% of them were proliferating and 7% of paraBCs with 33% proliferating. In distal airways they did not observe any paraBC and only a 6% of BCs with a 30% of proliferating BCs (James E Boers, Ambergen and Thunnissen, 1998). The group of Rawlins performed single cell RT-qPCR over 67 isolated single cells and described the presence of a BC population expressing *Krt5* but also the luminal keratin marker *Krt8* (Watson *et al.*, 2015). This KRT5+/KRT8+ population was also described by the group of Cardoso suggesting that these paraBCs were the progenitors of the differentiated luminal cells (Mori *et al.*, 2015).

Very recently a study of in vitro regeneration of airway epithelium in mouse, described the presence of a *KRT4+/KRT13+* population that could also correspond to supra-basal cells. During mouse regeneration of airway epithelium in vivo, the authors describe the presence of this population in "blocks" what they called "hillocks", they suggested that this population can act as an alternative way from BCs to SCs (Montoro *et al.*, 2018). Also very recently a single cell transcriptome analysis of Mouse Tracheal Epithelial Cells (MTECs) showed a discrete cluster of cells expressing *Krt4* and *Ktr13* (Plasschaert *et al.*, 2018). In our study of in vitro and in vivo HAECs differentiation by single cell RNAseq, our data showed a transitioning BC population precursor of the SC population (Ruiz Garcia *et al.*, 2018).

The groups of Wa Xian and Frank McKeon described the presence of clusters of Trp63<sup>+</sup> and Krt5<sup>+</sup> cells in distal airways after H1N1 influenza virus infection. They termed them “Distal Airway Stem Cell” (DASC<sup>P63/Krt5</sup>). These authors describe how these cells are able to proliferate and disperse to injured areas where they proliferate and differentiate into cells expressing genes linked to alveolar function. They demonstrate that DASC contribute to lung regeneration giving rise to multiple epithelial lineages including cell types of the distal lung (Kumar et al., 2011; Zuo et al., 2014).

The groups of John Ngai and Russell B. Fletcher used a murine model of injury to assess the role of the BCs during the recovery of the airway epithelium and the derivatives of these BCs, to do so they used single cell techniques and clonal lineage tracing of BCs. They confirmed the BC quiescence at the homeostatic level and that after injury, BCs show multipotency and divide in a symmetric or asymmetric fashion early in the regeneration. To study the cell fate time course of the BC after injury, they used single cell transcriptomic analysis of lineage traced cells at time before injury and after injury in an early and late state. They show a lineage trajectory where they describe the presence of different populations of BCs in the regenerating epithelium, these BC populations are transitioning states with different transcriptome level of expression of basal markers such as *P63*, *Krt5* and *Krt14*. Moreover, they see clear differences between the BCs in homeostasis and the BCs after injury, after injury the BCs shift to express genes involved in wound-response transcriptional program such as *Krt6a* and *Krt16* and members of the SPRR gene family (Gadye et al., 2017). However, our analysis of single cell transcriptome of HAECs showed the expression of these markers in the BC population in a condition of homeostasis *in vitro* and *in vivo* without injury (Ruiz Garcia *et al.*, 2018).

### III.1.2 Secretory cells: Precursors of multiciliated cells

As was described before, the BCs in mouse are restricted to the more wide ducts of the airways, meaning that they are only present in the upper part of the conducting system, whereas the SCs are lining all the surface of the airways, and in the smaller airways without BCs, the SCs are going to behave as the progenitors of the epithelial cell types in these areas (Breuer et al., 1990).

This was demonstrated by Rawlins and colleagues in 2009, they used a genetic SC tractable mouse model, where they genetically tagged the Scgb1a1+ cells, and study the behaviour of the SCs in the bronchi, these results confirmed the self-renewal capacity of the SCs in the bronchi and also, their ability to differentiate and give rise to MCCs (Rawlins et al., 2009).

The analysis of the SCs in the upper airways in homeostasis of the adult lung, confirmed that the population was not self-renewing and has to be replenished by an unlabelled population (BCs). They also analysed SCs during postnatal growth, where they showed self-renewing of this populations from day 2 until 3 weeks after birth, meaning that in these early stages of the development the SCs acted as a transitioning-amplifying cell population. With these experiments they confirmed the fact that MCCs are deriving from SCs, during the homeostasis and postnatal growth in trachea and bronchi

They also studied the behaviour of these SCs after injury in trachea and bronchi, in both cases they reported that remaining SCs were responsible for the restoration of the epithelium. Interestingly, in the latter experiments, they were able to identify a rare lineage-labelled BCs, meaning that depending on the conditions, SCs can certainly change their fate and even dedifferentiate to give rise in this case to basal stem cells (Rawlins et al., 2009).

The group of Rajagopal in 2013 wanted to clarify this possibility of SC dedifferentiation into BCs. They generated a mouse model where they were able to deplete the BC population specifically in the airways and from then, by immunocytochemistry or by cell lineage tracing, they could study the behaviour of this SCs. They showed that after ablation of the BC population, SCs were still able to dedifferentiate and restore the missing BC population. Moreover, these newly dedifferentiated BCs were losing the markers of SCs and were acquiring markers and functions of basal stem cells, being able to fully act as a normal BC population able to self-renew and differentiate (Tata *et al.*, 2013). During these experiments they revealed that only immature SCs were able to dedifferentiate, based on the appearance of certain markers such as SSEA (FUT4 a Fucosyltransferase) which is a cell surface marker of SCs, B1 (marker against a subunit of the H+ATPase gene (Atp6v1b1)) that is expressed only in mature SCs (Kim et al., 2012; Domenighetti et al., 2018).

### III.1.3 Goblet cell increased differentiation

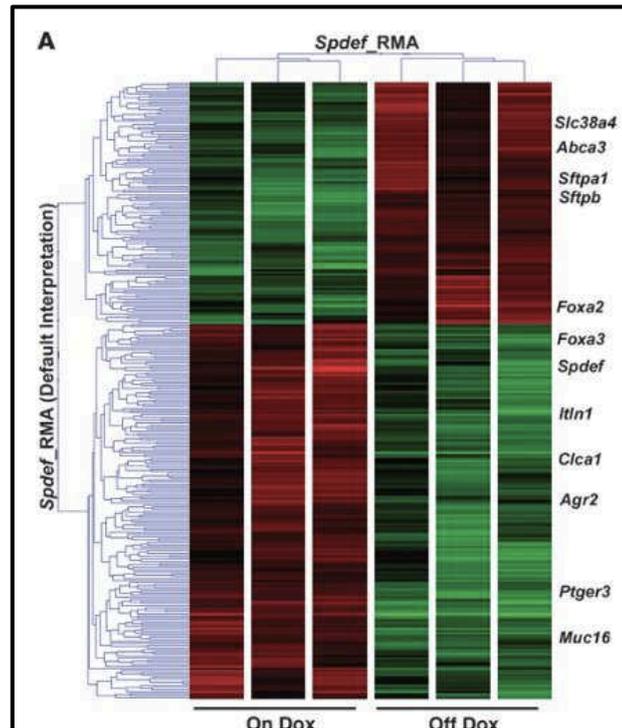
GCs are important cells of the airways, secreting mucins and working in the defence of the respiratory system. Under pathological conditions such Asthma or COPD the number of these GCs increase (hyperplasia) leading to an excess in the secretion of mucus what causes airway obstruction. A better knowledge of the origins of these cells could help defining new relevant therapeutic targets in the context of GC hyperplasia (Evans et al., 2004; Turner et al., 2011).

Little information is available regarding transcriptional programs controlling GC differentiation. In human, GCs are not abundant in absence of inflammation. After inflammation there is an increase of GCs mediated by Th2 type cytokines interleukin (IL)-4 and IL-13 (this point will be discussed more in details in other chapter) and this increase of GCs is correlated with a decrease of the transcription factor Forkhead Box A2 (*Foxa2*) (Kuperman et al., 1998; Wan et al., 2004; Zhen et al., 2007). In tissue from patients with different lung diseases, there is an inverse correlation between FOXA2 and GC hyperplasia (Wan et al., 2004). In 2004, the group of Whitsett generated a transgenic mouse with a conditional Cre/loxP recombination system to delete *Foxa2* in respiratory epithelial cells of the developing lung (*Foxa2<sup>loxP/loxP</sup>*). They showed that the mice lacking this transcription factor had an increase number of GCs which were almost undetectable in control mice. They also demonstrate together with the group of Kuperman, that the GC hyperplasia due to the cytokine treatment or the lack of FOXA2 was dependent on STAT6 as *Stat6<sup>-/-</sup>* mice did not present GC hyperplasia nor reduction of FOXA2 after treatment with IL-4. Interestingly, also thanks to their mouse model lacking *Foxa2* they showed that there was no increase in the transcription of cytokines involved in

inflammation and remodelling (IL4, IL5, IL6, IL9, IL10, IL11, IL13, IL17 and tissue necrosis factor  $\alpha$ ). They also showed that *Foxa2* inhibited the transcription of *Muc5ac* (Kuperman *et al.*, 1998; Wan *et al.*, 2004).

The SAM pointed domain-containing ETS-like transcription factor (SPDEF) is shown to be sufficient and necessary for GC hyperplasia (Bonser and Erle, 2017). This transcription factor was detected to be expressed in epithelial cells of very different tissues such prostate epithelium, intestine, and oviduct in the mouse, but it was not until 2007 when the group of Whitsett detected this molecule at protein and mRNA level in a subset of epithelial cells in tracheal glands and in the conducting airways of the foetal and adult mouse lung (Park *et al.*, 2007). Moreover, they detect a higher level of expression (mRNA and protein) in epithelial cells of the tracheal glands. To assess the role of this transcription factor in the differentiation of SCs, they conditionally expressed SPDEF under control of *Scgb1a1*/Clara cell secretory protein (CCSP-rtTA) and they showed that this conditional overexpression of SPDEF triggered the differentiation of SCs into GCs in large (cartilagenous), medium-sized (central), and lateral (distal) airways, but not in terminal bronchioles or in the alveoli. They also showed the absence of FOXA2 in these cells and that this GC hyperplasia was caused without the induction of any expression cytokines or inflammatory components.

The same group assessed the role of SPDEF in the differentiation of GCs. They performed lineage tracking labelling of SC and then induced GC differentiation using a model of injury where mice were sensitized to ovalbumin (OVA), which mimick allergic conditions. Their results showed that the new GCs came from a process of differentiation of the SCs rather than proliferation of GCs. They identify the mRNAs regulated during GC differentiation after conditional expression of SPDEF in SCs and they show that 306 genes were influenced by SPDEF expression such as genes involved in mucus production, protein glycosylation and mucin secretion (*Foxa3* and *Agr2*). They also identified the downregulation of genes involved in SC or alveolar cell differentiation such as *Foxa2* and *Titf1* (Chen *et al.*, 2009). (Fig.15).



**Fig15. Microarray analysis of bronchiolar epithelial cells.** comparing On Dox (overexpressed SPDEF) and Off Dox (control) green indicates those decreased and red indicates mRNAs increased. From (Chen et al., 2009)

Interestingly they detected co-localization of SPDEF, FOXA3 and AGR2 in areas of GC hyperplasia in bronchial tissues from patients with lung diseases such as CF or diseases caused by cigarette smoking together with the expression of MUC5AC. In vivo these markers co-localize in normal conditions in the GCs of the submucosal glands in humans and mice.

Different experiments studying the epithelium regeneration after injury in mouse revealed that the GCs were coming from the SCs through a process of differentiation (Reader *et al.*, 2003; Evans *et al.*, 2004; Hayashi *et al.*, 2004)

More surprisingly K. Park et al. identified a process of dedifferentiation of MCCs after naphthalene treatment in mouse bronchi. After this treatment they depleted the SC population in the airways, they analysed this depletion in the area of the bronchi and they noticed that instead of having a denuded basal lamina there was a thin layer of squamous cells, a deeper analysis of these cells revealed some characteristics of MCCs as the expression of FOXJ1,  $\beta$ -tubulin IV, and the presence of some disorganized fragments of cilia in the cytoplasm (Park et al., 2006). Few days after injury the analysis of the bronchiolar epithelium revealed the complete restoration of all the different cells types of the bronchi. These results suggested that the differentiated MCCs underwent squamous metaplasia, covering epithelial surface, and redifferentiated into both ciliated and nonciliated cell types.

Maintaining a correct balance between the different cell populations forming the airway epithelium is essential for the homeostasis in adult tissue in healthy conditions. After injury, a fast response to restore the epithelium cell population is needed. There are different signalling pathways involved in the differentiation and homeostasis of the airway epithelium and it is very important to understand their implication in the choice of fate of every cell in the system and their role in a post-injury response. The next chapters will briefly introduce the different signalling pathways involved in this process and their specific role during the differentiation of the airway epithelium will be later described.

### III.2 The Notch pathway

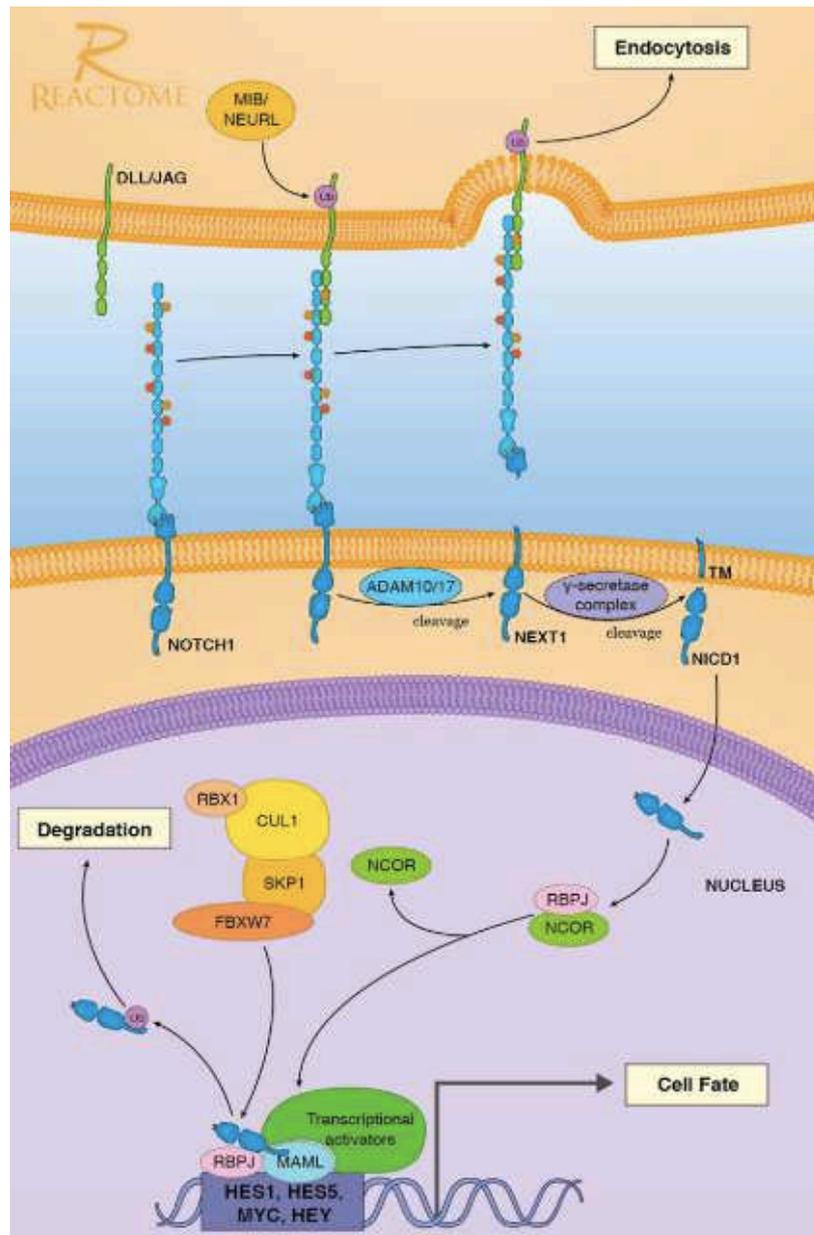
This signalling pathway plays an important role in different models and systems. It is highly conserved across evolution and plays an important role in tissue differentiation and homeostasis (Fortini, 2012; Guruharsha, Kankel and Artavanis-Tsakonas, 2012). A mal functioning of this pathway leads to an incredibly amount of different diseases.

Notch was discovered in 1913 by John Smith Dexter in a mutant of *Drosophila melanogaster*. The functioning of the Notch signalling pathway involves a signalling cascade between two neighboring cells: a first cell expresses Notch receptors on its surface while the second one expresses ligands of the Notch pathway.

In mammals the Notch receptors are NOTCH1, NOTCH2, NOTCH3 and NOTCH4, and the ligands of the Notch pathway in mammals belong to two different family proteins: the family of the Jagged protein family (JAG1 and JAG2) and the Delta-like protein family (DLL1, DLL3 and DLL4).

The ligand has to be activated by ubiquitination by MIB (Mind Bomb), and after activation, is able to link with the Notch receptor of the neighbour cell (Tsao et al., 2009).

Interactions between these two components lead to the release Notch intracellular domain (called NICD from Notch IntraCellular Domain) by proteolysis and its translocation to the nucleus where it binds to a recombining binding protein called “suppressor of hairless” (RBPJ) also called CSL (CBF1, Suppressor of Hairless, Lag-1) (Pursglove and Mackay, 2005)(Fortini, 2012). Target genes of Notch are transcriptionally repressed by the association of two proteins, CSL and NcoR (nuclear receptor co-repressor), which are located in an inhibitory transcription factor sitting on the promoter of these genes. The released NICD is going to break the binding between CSL and NcoR making CSL accessible to MAML and p300 that binds to it and makes a transcription activation complex, activating the transcription of targets genes of the pathway (Pursglove and Mackay, 2005; Tsao et al., 2009).



**Fig. 16. Model of the Notch pathway signalling.** The ligand DLL or JAG of one cell binds with the NOTCH receptor of the neighbour cell, this interaction lead to the cleavage of the Notch intracellular domine (NICD) from the cell membrane thanks to a secretase complex. NICD translocates into the nucleus where it binds the complex RBPJ/NCOR activating the gene transcription. Source Reactome.

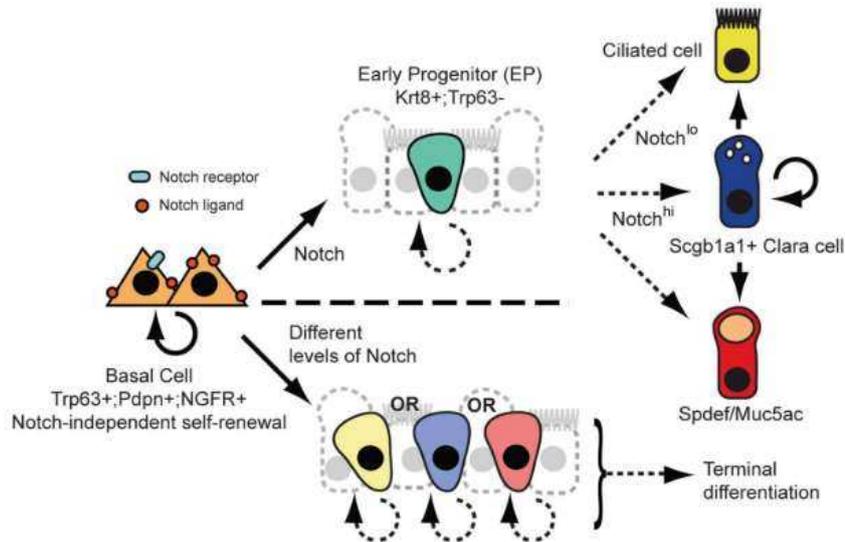
### III.2.1 Notch activation favors GC and inhibits MCC differentiation

Lineage commitment to either secretory or multiciliated fate has been shown to be regulated by the Notch signalling pathway and in an early stage in the differentiation (Guseh *et al.*, 2009b; Tsao *et al.*, 2009; Morimoto *et al.*, 2010, 2012; Danahay *et al.*, 2015).

Tsao, Guseh and Morimoto described the role of the Notch pathway in the control of the correct balance between the different cell populations in the airways. Blocking the Notch pathway induced a defect in

differentiation that led to an absence of SCs and an overpopulation of MCCs and neuroendocrine cells. On the contrary, activation of Notch led to GC hyperplasia and a decrease in the number of MCCs (Guseh *et al.*, 2009b; Tsao *et al.*, 2009; Morimoto *et al.*, 2010).

Rock *et al.* also studied the implication of Notch in the differentiation and homeostasis. To that end, they analysed the BC population that they isolated by FACS. They noticed that this population was highly enriched by components of the Notch pathway such as Notch1, Dll1, Jag2 and Jag1. They used a mouse transgenic Notch reporter (TNR) model to monitor the activity of the Notch pathway. Based on this approach, they determined that at steady state in adult mouse, the activation of notch was minimal, just to maintain the low turnover of normal homeostasis. A different situation was noticed after they treated the mouse with SO<sub>2</sub> to provoke injury. In this case, they detected an increase of the Notch pathway. Under these conditions, they decided to analyse this activation of Notch in the BC population and they discovered that the cells positive for the action of Notch were not the proliferative basal stem cells but, the before mentioned basal luminal cells or early progenitors characterized by the expression of KRT8 and a reduced expression of TP63. After further activating Notch in transgenic mice, they showed that the pool of basal luminal cells increased, meaning that activation of Notch increased the differentiation rate of basal stem cells into basal luminal cells early progenitor cells. They noticed that most of the cells positive for Notch expression were expressing the SC marker SCGB1A1 and half of these ones were positive for GC markers such as SPDEF and MUC5AC meaning that activation of Notch induces the differentiation into secretory and GCs (Rock *et al.*, 2011).



**Fig. 17. Models for Notch signalling in the adult pseudostratified airway epithelium.** the luminal differentiation of BC required canonical notch signalling. In the upper panel: a multipotent early progenitor (EP) has limited capacity of proliferation and gives rise to mature MCCs and SCs in respond to a second notch signal. In the lower panel: EPs are more lineage restricted and their fates are specified by Notch input at the time of BC division. From (Rock *et al.*, 2011)

### III.2.2 BCs send a Notch signal to SCs

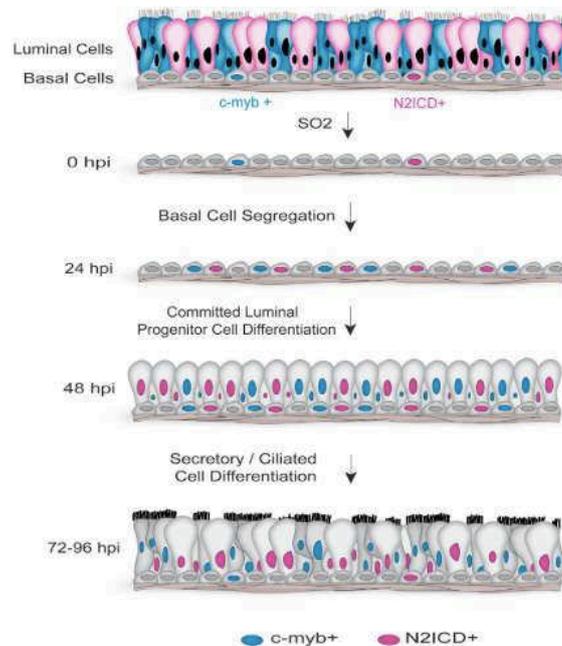
The group of Rajagopal and Lafkas in 2015 also studied the role of the Notch pathway in the differentiation of the epithelium (Lafkas *et al.*, 2015; Pardo-Saganta, Law, *et al.*, 2015; Pardo-Saganta, Tata, *et al.*, 2015). Pardo-Saganta *et al.* tested the expression of several Notch markers that were either related to secretory or MCC fate (Morimoto *et al.*, 2010, 2012). Interestingly they observed two subpopulations of BCs expressing low levels of c-myb (transcription factor acting downstream Notch) and N2ICD (the active Notch2 intracellular domain) and the expression of these markers was exclusive. Using two models of injury in mouse they revealed that there was an increase in the number of BCs expressing c-myb or N2ICD, and these populations were able to proliferate. After recovery of the epithelium they detected the presence of Scgb1a1+ expressing cells (SCs) where almost all of them co-expressed N2ICD and the presence of Foxj1+ cells (MCCs) where also almost all of them co-expressed c-myb. The inhibition of the Notch pathway in a transgenic mouse lacking *Rbpj* (CLS) provoked, after injury, a decrease in the number of BCs expressing N2ICD that led to a decrease in the final number of SCs. Moreover, they showed an increase in the number of c-myb+ BCs and a consequent increase in the number of MCCs.

Their results indicate that inside the BC population there are subpopulations of BCs expressing different markers that reveal a predisposition of the cells to be differentiated in MCCs or SCs.

They also tested the implication of the Notch pathway in the feedback signalling through the different cell types of the airway epithelium during homeostasis in the adult tracheal mice tissue and noticed that through ablation of the MCC population there was no change in the proliferation or differentiation rate of BCs, the epithelium kept its normal homeostatic turnover rate meaning that there is not signal transmitted by the MCCs to the progenitors (Pardo-Saganta, Tata, *et al.*, 2015).

Then they used the SC lineage traced mouse model to analyse the effect of the ablation of the BC population. They showed an increase of MCCs traced from SCs and they detected SCs losing markers such as SCGB1A1 and expressing markers such as FOXJ1, acetylated tubulin and C-MYB. meaning that the ablation of BCs led to an increase of the differentiation of the SCs into MCCs.

Analysing the expression of the different components of the Notch pathway in the different cell populations of the epithelium they detect the expression of the receptor Notch1 in the BCs and Notch2 and 3 in the SCs, whereas N2ICD is restricted to the SCs. To test whether the Notch pathway had an impact in the maintenance or differentiation of the SCs they ablated different components of the Notch pathway (RBPJ, Notch2, MIB, JAG2) and they showed the increase differentiation of the SCs into MCCs with the consequent decrease in the number of SCs. They confirmed that ligands emanating from basal stem cells are necessary for N2ICD activity in SCs (Pardo-Saganta, Tata, *et al.*, 2015). Lafkas obtained similar results after blocking JAG1 and JAG2 ligands of the Notch pathway, showing differentiation of SCs into MCCs (Lafkas *et al.*, 2015).



**Fig. 18. Model of the airway epithelial lineage hierarchy induced by injury.** Basal cells segregate into a N2ICD+ population and a c-myb+ population in response to injury. During regeneration, a continuum of differentiation is observed and basal cell markers and markers of differentiation are coexpressed between 48 and 72 hpi. N2ICD+ basal cells and c-myb+ basal cells subsequently express differentiation markers and basal cell markers are gradually lost. Suprabasal N2ICD+ cells then give rise to secretory cells, whereas c-myb+ cells give rise to ciliated cells. From (Pardo-Saganta, Tata, et al., 2015).

In 2015 the group of Danahay assess the role of the different receptor of Notch present in BCs isolated from a model of 3D culture of primary human airway BCs (bronchospheres). Blocking the different Notch receptors (Notch1, Notch2 or Notch3) using receptor-specific blocking antibodies, they showed that Notch2 blocking antibody inhibited all the GC markers (MUC5AC, MUC5B and FOXA3) and increased the MCC markers such as FOXJ1 (Danahay et al., 2015). These results agree with the results of Pardo-Saganta and rajagopal (Morimoto et al., 2010, 2012).

### III.2.3 Notch3 over-activation depletes the BC pool by over-differentiation into SCs

The group of Cardoso (Mori et al., 2015) described the role of the receptor NOTCH3 in the differentiation and homeostasis of the epithelium. Their results show the expression of NICD3 (Notch intracellular domain 3) and Hes1 (notch target) in a paraBC population KRT5+ TP63-, and that this receptor of NOTCH is involved in the control of the architectural arrangement of the epithelium regulating the pool of basal stem cells (KRT5+, TP63+) and also the undifferentiated parabasal population (KRT5+, TP63-). Inhibition of Notch pathway using DAPT (gamma-secretase inhibitor) leads to the proliferation of KRT5+ TP63+ basal stem cells and the downregulation of Notch3.

A previous report by Dang and colleagues (Dang et al., 2003), show how activation of Notch 3 in the embryonic lung epithelium leads to an undifferentiated state of the epithelium. Cardoso et al. activated Notch 3 constitutively and, contrary to Dang et al., they detected a decrease in the TP63+ population and an increase in cells expressing SCGB3A2, SC marker. So, activation of Notch3 leads to an excessive SC differentiation.

#### III.2.4 The miR-449 family favors MCC differentiation through Notch pathway inhibition

Our laboratory provided new information deciphering an important mechanism involved in the inhibition of the Notch pathway, by analysing the function of the miR-449 family in the development of MCCs, in HAECs and in the embryonic *Xenopus* skin (Marcet, Chevalier, Coraux, et al., 2011; Marcet, Chevalier, Luxardi, et al., 2011). MicroRNAs (miRNAs) are small non coding RNAs that regulates gene expression through post-transcriptional repression (Ambros, 2004; He and Hannon, 2004). In this work they characterized for the first time the involvement of the miR-34/449 family of miRNAs in the inhibition of the Notch pathway. The family of miR-34/449 is conserved in vertebrates, comprising three genomic loci, miR-34a, miR-34b/34c and miR-449c/449b/449a (miR449) which encode six homologous miRNAs (miR-34a, 34b, 34c, 449c, 449b and 449a). Marcet et al showed the increase of expression of miR-449 during the process of differentiation of the human airway epithelium, and they localized by in situ hybridization the expression of this miRNA in MCCs, but not in basal or GCs. They showed also how the expression of this miRNA is involved in the differentiation of MCCS at an early stage, before the centriologenesi, but is not involved in the expression of other markers such as FOXJ1, which is still expressed after blockage of miR-449, even though after this blockage the generation of MCCs is impaired. Identification of targets of this family or microRNAs shows an interaction between miR34/449 and some members of the Notch pathway, including NOTCH1 and Dll1, they demonstrated the direct repression of these Notch signalling components by miR-449 elucidating the mechanism of inhibition of Notch necessary for the development of MCCs in the human airway epithelium. Later on, the group of Lin He revealed cp110 as the major target of this miRNAs family (Song et al., 2014). Levels of cp110 have to be tight regulated during the process of multiciliogenesis, removal of cp110 from centrioles is essential for ciliation (Lai et al., 2011; Tsang and Dynlacht, 2013). They use a mouse model triple knock-out deficient for all miR34/449 loci, with this model they confirmed the deficient MCCs development in airways, they observed deficient basal body docking in the membrane, essential for multiciliogenesis, instead, they detected the basal bodies mislocalized in the cytoplasm. They also analysed the possible targets of miR-34/449 and reveal cp110 as an important candidate harbouring two binding sites for these miRNAs, they could reveal that cp110 expression is regulated by miR-34/449, they show using their mouse model triple knock-out, that after knock-down cp110 the multiciliogenesis was restored. They performed also overexpression of cp110 and obtained the same phenotype as in the miR-34/449 knockdown. However, our group tried to reproduce this targeting of cp110 by miR-449 without success (our unpublished data).

### III.3 EGF signalling pathway

Another important pathway implicated in the differentiation of the airway epithelium is the EGFR signalling pathway. This signalling pathway is expressed in epithelial cells, in epidermal cells of the skin and epithelial cells lining other organs such as the airways. When any of these system is damaged the epidermal growth factor (EGF) is release. EGF is a peptide molecule that acts as a primary messenger in the EGF signalling transduction pathway and it is going to activate the growth and division of epidermal and epithelial cells.

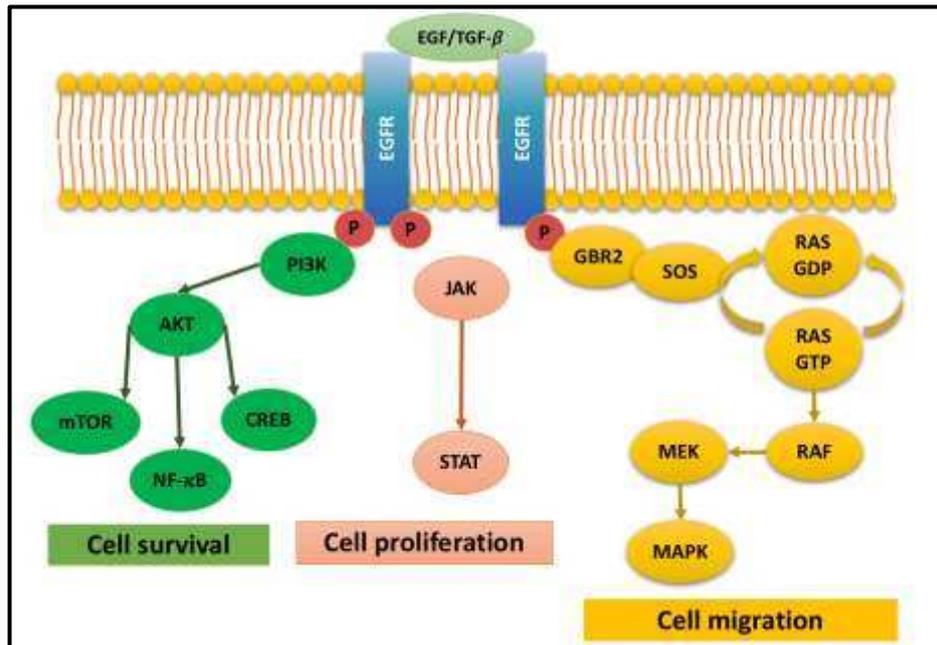
This pathway consists in the extracellular binding of EGF molecules to a transmembrane receptor which is a monomer in its inactivated state. After binding of the EGF, it is activated forming a dimer, and the intracellular part (the tyrosine kinase domain) is phosphorylated creating an attachment point for the rest of the components of the signalling pathways.

There are, at least, three signalling pathways that are activated upon activation of EGFR.

- The small GTPases pathway: An adaptor protein called GRB2 is going to attach to this phosphorylated domain of the EGFR and it is going to act as a bridge for the anchoring of other important protein called SOS that has a binding site for RAS a small G protein that allows the binding of GDP and GTP molecules. The binding of SOS to RAS activates RAS expelling a GDT and binding a GTP. Once RAS is activated can activate RAF. RAF activates protein kinases called MEKs these MEKs that activate other proteins called ERKs (Extracellular signal-Regulated Kinases), once ERKs are activated they translocate into the nucleus and stimulates transcription factors leading to the cell growth and proliferation.

- The PI3K/AKT pathway: Upon the activation of the tyrosine kinase complex of the EGF receptor, PI3K (Phosphoinositol 3 kinase) binds to the phosphorylated complex of the EGFR. The active PI3K activates AKT (protein kinase B) which is a proto-oncoprotein. One of the effects of this protein is the inhibition of the apoptosis programmed cell death, other is the activation of protein synthesis or translation through mTOR pathway.

- The JAK/ STAT pathway: (Janus Kinase/ Signal Transducer and Activator of Transduction): In this pathway two JAKs molecules bind directly the tyrosine kinase domain of the EGFR homodimer they trans-phosphorylate each other getting activated, they can activate STAT that can translocate to the nucleus and promote the gene transcription promoting cell growth and differentiation



**Fig. 19. Model of the EGFR pathway.** (green) The PI3K/AKT pathway. (orange) The JAK/STAT pathway. (yellow) The small GTPases pathway.

### III.3.1 EGFR signalling is implicated in the differentiation of GCs

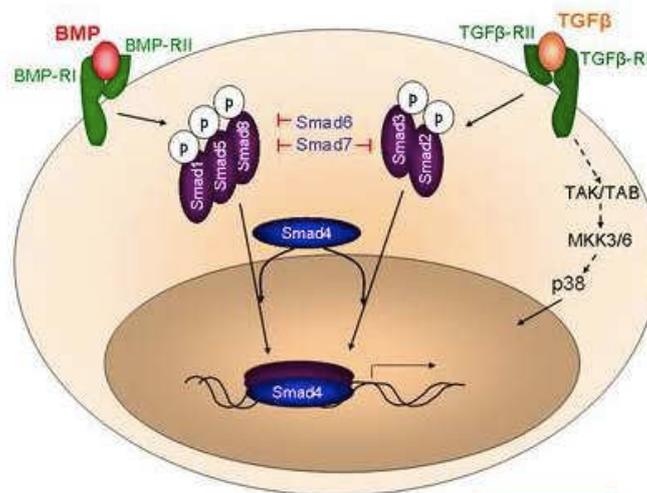
EGFR has been demonstrated to play a central role in epithelial repair processes and mucin synthesis (Takeyama et al., 1999; Burgel et al., 2000; Nadel and Burgel, 2001; Zhen et al., 2007).

In 1999 the group of Kateyama tested the effect of the different ligands of the EGFR pathway (EGF or TGFβ) in the airway epithelium of rats. They incubated airway epithelial cells with EGF or TGFβ and showed that the stimulation of EGFR by its ligands result in GC production with the consequent MUC5AC overproduction. In the other hand using inhibitors for the EGF-R Tyrosine Kinase they inhibited MUC5AC production and also Muc5ac expression was decreased (Takeyama et al., 1999). The use of pharmacologic inhibitors to block EGFR ligand binding and inhibitors of EGFR tyrosine kinase activity in cultured HAECs showed decreased expression of MUC5AC (Takeyama et al., 1999; Zhen et al., 2007). Surprisingly blockade of EGFR reduced drastically the expression of transcripts of MCCs and BCs (TUB-β4, KRT14 & KRT4)(Zhen et al., 2007). Nadel and Burgel checked the role of EGFR in the hypersecretion of mucus and in the GC hyperplasia in patients with polyps, they showed that the airway epithelium of these patients secreted more MUC5AC than controls and that the expression of EGFR was also enhanced both in gene and protein (Burgel et al., 2000; Nadel and Burgel, 2001).

These results shown that EGF signalling is a positive regulator of GC differentiation.

### III.4 TGF-B/BMP Pathway

BMP and TGF-B ligands belong to the “TGF-B superfamily”. They bind to heterodimeric transmembrane receptors triggering the phosphorylation of receptor subunit 1 by receptor subunit 2. The phosphorylated subunit 1 then phosphorylates an intracellular SMAD protein (SMAD1,5 when the ligand is BMP, SMAD 2,3 when the ligand is TGF-B). This phosphorylated SMAD binds to other SMAD (SMAD4) protein forming a transcriptional regulatory complex that enters the nucleus and activates or inactivates the transcription of target genes.



**Fig. 20. Model of the signalling pathways TGF-B and BMP.** From [www.humphath.com](http://www.humphath.com)

#### III.4.1 BMP negatively regulate MCC differentiation

Our team, in collaboration with Kodjabachian’s team described the role of the BMP pathway in the development of the mucociliary epithelium of *Xenopus* skin and human airways. They detected the increase of the phosphorylated form of SMAD1/5/8 (indicating activation of the pathway) in most of the non-neural ectodermal cells in *Xenopus* and also during differentiation of human airway epithelial cells (HAECs) until the onset of differentiation. Overexpression of BMP4 in *Xenopus* led to a decrease in MCCs, SCs and ionocytes whether GC numbers were not altered. In HAECs, incubation with BMP2 led to the decrease of MCCs and GCs, the epithelium suffered squamous metaplasia, meaning that the overexpression or over-activation of the BMP pathway led to a development of an altered and dysfunctional mucociliary epithelium. In the same line, inactivation of the BMP pathway in *Xenopus* skin, led to increase in MCCs, ionocytes and SCs, but the MCCs are not able to form cilia. In HAECs the inhibition of BMP using an antagonist of the pathway stimulates the MCCs formation. The authors revealed the relation between the Notch and the BMP pathways, showing that a tight regulation of BMP is required for Dll1 expression and the specification of MCCs, ionocytes, SCs and non-intercalating inner cells, through the Notch pathway. They showed that the inactivation of Notch pathway after over expression of BMP is enough to restore MCC levels. On the other

hand, activation of the Notch pathway lowers the high MCCs levels obtained after inactivation of BMP (Cibois et al., 2015).

In the same line, other authors using an in vivo mice model of injury and cultured primary cells from mouse trachea and from human biopsies, studied the differentiation of airway epithelium and showed that TGFb/BMP/SMAD signalling inhibition compromised terminal differentiation, and activation of SMAD via BMP or via TGF- $\beta$  triggers the differentiation of the epithelium (Mou et al., 2016).

#### III.4.2 BMP maintains BCs in a quiescent state

The group of Hogan and colleagues, using a model of regeneration of airway epithelium after injury in mouse, described that the activation of BMP4 inhibited the differentiation of the epithelium. BMP acts as a brake on cell proliferation and maintains the BC population quiescent. They showed that, after losing the luminal cells of the airway epithelium, due to injury, there was the accumulation and layering of BCs acting as a new progeny of the epithelium, this accumulation of BCs was possible upon inhibition of BMP signalling detected by the decreased levels of epithelial phospho-SMAD1/5/8 (Tadokoro et al., 2016).

## IV. *Airway remodelling*

---

Airway epithelium is constantly challenged with external signals. Inhalation of pathogens or toxic gases can cause damage of the epithelium. In normal conditions, the resulting injury of the epithelium leads to a succession of cellular events such as the loss of the surface epithelium integrity to partial shedding of the epithelium or eventually a complete denudation of the basal lamina, in these conditions, the response of the healthy epithelium results in a full repair and regeneration of the mucociliary clearance system.

Disorders in mucociliary clearance which are due to ciliary dysfunctions are grouped under the term “Primary Cilia Dyskinesia” (PCD). However these dysfunctions can also be caused by inflammation and increased viscosity of the airway epithelial secretions that are secondary to Cystic fibrosis (CF), asthma or chronic obstructive pulmonary disease (COPD). All these different disorders are included in a category of Secondary (or acquired) Ciliary Dyskinesia (SCD) (Bertrand et al., 2000). The injury and inflammatory events occurring in a chronic manner in these patients with SCD, lead to a remodelling of the airway epithelium, in association with mucin hypersecretion, increased viscosity and dehydration of the mucus, impairment of the beating of the motile cilia of the MCCs, colonization by opportunistic pathogens and development of recurrent infections.

### IV.1 Secondary Ciliary Dyskinesia (SCD)

The more frequent SCD are: Asthma, cystic fibrosis and Chronic pulmonary disease (COPD).

#### IV.1.1 Asthma

There are around 235 million people suffering from asthma all around the world (World Health Organization 2017, WHO). The term asthma rather of being a defined single disease, is a group of clinical and physiological characteristics that include different phenotypes, associated or not to allergy (Wenzel, 2012). It is a chronic pulmonary disease characterized by structural changes in the airways, considered as airway remodelling, these changes in a long-term have bad outcomes, and are associated with eosinophilic inflammation and airway hyperresponsiveness. Airway hyperresponsiveness is associated to a variety of “triggers” such as environmental allergens leading to a reversible airway inflammation and obstruction. After inflammation of the bronchia and bronchioles, the airway ducts are narrow and swell. Asthma disease is also characterized by GC hyperplasia associated with a hyper secretion of mucus, causing difficult breath difficult, wheezing, coughing and shortness of breath. Asthma is also associated with responsiveness to mechanical forces that occurs during bronchoconstriction episodes. In vitro studies indicate that this mechanical stress

also leads to remodelling of the airway epithelium in a mechanism that is independent of inflammation (Ressler et al., 2000; Grainge et al., 2011; Park and Fredberg, 2016).

Among different clinical phenotypes of asthma that have been described, lung inflammation can result from a non-allergic eosinophilic response (mainly triggered by viruses), a Th1-Th17 mediated inflammation induced by neutrophils rather than by eosinophils, a paucigranulocytic asthma, or, more frequently, a Th2 response mediated by the production of pro-inflammatory cytokines (such as IL 4/5, eotaxins and IL-13). These different subclasses of asthma differ by their responses to different classes of drugs. IL-13 is an important trigger, which is targeted by several specific drugs, and I will discuss more deeply its role in the inflammatory phenotype and in the hypersecretion of mucus (Lai and Rogers, 2010; Papi *et al.*, 2018).

While my presentation is mainly focussed on the epithelial component, it is also important to remember the importance of other cellular components in the disease. Thus, smooth muscle responds to the presence of different allergens or contracting-stimuli along the airway system, and this can lead to the shortening and constriction of the airway ducts. This can be caused due by secretion of neurotransmitters such as acetylcholine, and it is known that cytokines such as IL-13 and IL-17 are able to induce the expression of small GTPase RhoA, an important effector of the muscular contraction (Zhang et al., 2015). It has been shown that factors such as Wnt5a, produced by airway epithelial cells can trigger smooth muscle cells hyperreactivity (Koopmans *et al.*, 2016).

There are nowadays therapies that help to mitigate the symptoms such as daily inhaled or oral corticosteroids combined with long-acting  $\beta_2$  agonists (Chung, 2015), antileukotrienes or leukotriene modifiers, cromolyn sodium, methylxanthines or immunomodulators (American college of allergy, asthma & immunology 2014).

#### IV.1.2 Chronic obstructive pulmonary disease (COPD)

COPD is a chronic inflammatory disease that englobes emphysema, chronic bronchitis and refractory (non-reversible) asthma, causing obstruction of the airway ducts. It is mainly triggered by a long-term exposure to irritating gases or inhaled particles, and often is developed in smokers.

The symptoms are: difficulty to breath, GC hyperplasia with the consequent over-production of mucus, cough and wheezing. These symptoms do not appear until the damage is rather significant and usually worsen over time. People with COPD have a higher risk of developing heart disease and lung cancer. COPD has a lot of similarities with asthma, and people suffering of both diseases usually show worse outcomes. Some authors claim that asthma and COPD are the continuum of the same disease that results from same origins but develop in response to different stimuli. Other authors claim that both asthma and COPD have separate and independent pathogenic mechanisms (Maselli and Hanania, 2018).

COPD is characterized by a decrease in elasticity and diameter of the bronchia, which affects smaller ducts of the airways (<2mm) than asthma. It also differs by its own inflammatory response characterized by the presence of T/Th1 lymphocytes (Pizov, 2005).

Events linked to MCCs in COPD have been studied. Yaghi et al. studied the ciliary beat frequency (CBF) an important measure of cilia function in patients with COPD, and detected reduction in the CBF, impairing the mucociliary clearance function of the airway epithelium and contributing to retention of secretions and infection (Yaghi et al., 2012).

#### IV.1.3 Cystic Fibrosis (CF)

CF is an autosomal recessive disease that affects approximately 80,000 people worldwide, being the most common inherited recessive monogenic disease in the caucasian population (one CF patient out of 2500). CF is caused by mutations in the CFTR gene (cystic fibrosis transmembrane regulator). It affects the lungs but also other tissues and organs such as pancreas, intestine, liver, genital tract, and sweat glands. Lung disease is nowadays the most critical manifestation of the disease, associated with airway obstruction, difficulty to breath, excessive mucus production with increased viscosity, leading to chronic cycles of lung inflammation and infection that progressively affect lung function. Additional symptoms include sinus infection, poor growth, clubbing in the fingers and toes, infertility (caused by vas deferens agenesis).

The CFTR gene codes for an anionic channel present in the apical membrane of epithelial cells in the airway, intestine, pancreas, and reproductive tract. Its function is crucial in the homeostasis of the extracellular liquid layer in many organs (Machen, 2006). This channel is responsible  $\text{Cl}^-$  and  $\text{HCO}_3^-$  transport, contributing to the ionic composition of the airway surface liquid (Mall et al., 2004). It is known that CFTR-dependent bicarbonate secretion is required for normal expansion of mucins, a decrease in the transport of this ion leads to the aggregation of mucins making them non-soluble, in CF patients it has been seen that this transport of bicarbonate it is impaired. These abnormal properties of the liquid lining the airways provide a suitable niche for opportunistic pathogens such as *Pseudomonas aeruginosa* (Matsui et al., 2006; Zhou-Suckow et al., 2017).

The last 30 years have seen a considerable progress in the treatment of this disease, illustrated by a raise in life expectancy from 4-5 to 50 years. Recent achievements include the development of Ivacaftor, a compound that potentiates the activity of the CFTR molecules that are expressed at the cell surface, such as the G551D CFTR. It was shown by Accurso et al. that a treatment with Ivacaftor was improving by more than 10% the predicted forced expiratory volume in 1 second of G551D patients, the treatment was also normalizing their sweat electrolytes (Accurso *et al.*, 2010). This approach provided a rationale for a causative treatment of cystic fibrosis, through the administration of chemical modulators of CFTR activity. The nearly 2000 different CF mutations can be classified into 5 different classes, which are not mutually exclusive. Class I corresponds to mutations causing the production of a defective protein, due to production of nonsense mutations, large deletions or insertions; Class II are associated with a defective protein processing, such as the F508del, the most common of CF mutations, which is present in at least one allele in 85% of the CF

population; Class III includes CFTR proteins with defective gating, such as G551D; Class IV corresponds to mutants with defective protein conductance (R117H, R334W, R347P); Class V are mutants that generate reduced amount of functioning protein (2789+5G→A, 3849+10KbC→T, A455E). Recently, a triple combination of VX-445 (or VX-659) with Tezacaftor and Ivacaftor was tested in patients one or two Phe508del alleles. It resulted in an increased percentage of predicted FEV1 of up to 13.8 points in the Phe508del–MF group ( $P < 0.001$ ) (Davies *et al.*, 2018; Keating *et al.*, 2018).

CF provides a good example to clarify the relationships between inflammation and infection in the airways. It is logical to think that the inflammatory state is due to the pathogen infections, but, in fact, different studies indicated that an inflammatory state was possible in the absence of bacterial infection (Machen, 2006; Stoltz *et al.*, 2010; Pezzulo *et al.*, 2012). The group of C. Coraux used xenografts in mouse with human primary cells from CF or healthy patients, they used non-infectious environmental conditions to study the regeneration of the epithelium. Their results showed a pro-inflammatory state of the CF epithelium and the secretion of the cytokine IL-8 by the CF epithelial cells. Surprisingly, they did not observe a higher expression of MUC5AC in the CF cells, meaning that there was not GC hyperplasia in cell cultures from CF patients. They detected a decrease in the expression of MUC5B in comparison with controls. Same results were obtained using a model of human primary cell culture, using primary cells from CF or healthy patients. Cell composition of these cultures was analyzed and results showed that the CF epithelium was structurally higher, it had basal cell hyperplasia and no changes in the number of GCs. (R Hajj *et al.*, 2007; Adam *et al.*, 2015). These results suggest that the CF epithelium presents an inflammatory state even in the absence a given pathogenic infection, and that inflammatory properties are intrinsic to epithelial cells. It will be important to study this model at the light of the recent observation stating pulmonary ionocytes as a major site of CFTR expression in the airways.

## IV.2 Airway remodelling in disease

Airway remodeling is defined as the cellular and molecular architecture changes that result in structural alterations of the airway wall. These histologic and structural changes lead to a progressive loss of airway function and permanent and irreversible airway obstruction (Pascual and Peters, 2005; Ojiaku, Yoo and Panettieri, 2017). As it was mentioned before, all these structural changes can be triggered by inflammation or mechanical stress.

### IV.2.1 Mechanical stress and mucus over production triggers of inflammation

Is it known that mechanical stress leads to secretion of inflammatory mediators during bronchoconstriction caused by the contraction of the smooth fibres lining the airways (Swartz *et al.*, 2001; Park and Tschumperlin, 2009; Grainge *et al.*, 2011). In order to prove that mechanical stresses are indeed the

ones that trigger airway remodelling, some authors developed a system of primary cell culture from rats (Ressler et al., 2000) or humans (Tschumperlin et al., 2004) where they “squeezed” the cells causing a compressive stress. With these experiments, they showed that mechanical stress initiated the transcription of asthma associated mediators known to be implicated in remodelling of the airways such as Tumour Necrosis Factor (TNF- $\alpha$ ), early growth responsive protein 1, endothelin1 and transforming growth factor $\beta$ 1 (TGF- $\beta$ 1) (Park and Fredberg, 2016; Ojiaku, Yoo and Panettieri, 2017).

In asthma, the two major remodelling events are the thickening of the sub-epithelium and GC hyperplasia or metaplasia. Some authors showed that inferring mechanical stress on cultured airway epithelial cells is sufficient for them to mimic both events (Swartz et al., 2001; Tschumperlin et al., 2004; Park and Tschumperlin, 2009). Grainge and colleagues showed the same results in humans in vivo, detecting after repeated episodes of bronchoconstriction an increase in the number of GCs and a thickening of the subepithelium (Grainge et al., 2011).

In chronic lung diseases, MCCs are often reduced in number and mucus secretion often is increased, leading to an obstruction of the airway ducts and mucus plugging which results in defects in lung function and homeostasis. In a disease like CF, the obstruction caused by over secretion of mucus is the main cause of death. In CF, absence of functional CFTR leads to the dehydration of the airway surface which leads to the increased secretion of mucins (MUC5AC and MUC5B) resulting in osmotic pressure increase over the surface of the epithelium compressing the cilia, making them move slower impairing the proper mucus clearance.

The hyper-concentration of mucins could be caused not only by defects in the ion transport because of the mutation of CFTR in the CF patients, but also to inflammation events in patients with asthma or COPD (Button, Anderson and Boucher, 2016). Indeed, it was described by Puchelle and colleagues that abnormal expression or distribution of CFTR proteins can be due to the remodelling of the tissue where squamous metaplasia or/and GC hyperplasia disturb the functionality of the epithelium (Dupuit et al., 1995; Puchelle et al., 2006).

Using the murine model ENaC, where the subunit  $\beta$ ENaC of the Na<sup>+</sup> channel is overexpressed, several authors described that the mice had different phenotypes, First, they developed an early chronic airway inflammation characterized by airway neutrophilia, elevated levels of chemokines, robust macrophage activation and a type 2 inflammation with high levels of IL-13 and airway eosinophilia. Second, they developed airway remodelling with GC metaplasia and mucus hypersecretion characterized by the high expression of *Muc5ac* and *Muc5b*. Third, these mice showed high susceptibility to bacterial infections (Mall et al., 2008; Livraghi et al., 2009; Livraghi-Butrico et al., 2012; Zhou-Suckow et al., 2017). After breeding of these transgenic mice in germ-free conditions their airway epithelium showed inflammation, and had also

airway mucus obstruction, meaning that mucus obstruction and plugging is sufficient to trigger sterile airway inflammation (Livraghi-Butrico et al., 2012)

In conclusion Inflammation is not the only trigger of the airway remodelling.

Several authors described that one of the functions of the CFTR was the regulation of the epithelial Na<sup>+</sup> channel (ENaC) in airways and sweat glands (Stutts, Rossier and Boucher, 1997; Reddy, Light and Quinton, 1999). There was shown an abnormal Na<sup>+</sup> transport in CF-affected epithelia in vivo and in vitro in humans and in mouse(Grubb, Vick and Boucher, 1994; Mall *et al.*, 1998). However an study performed in three different laboratories reach the conclusion of the non inhibition of ENaC by CFTR (Nagel *et al.*, 2005).

#### IV.2.2 Inflammation

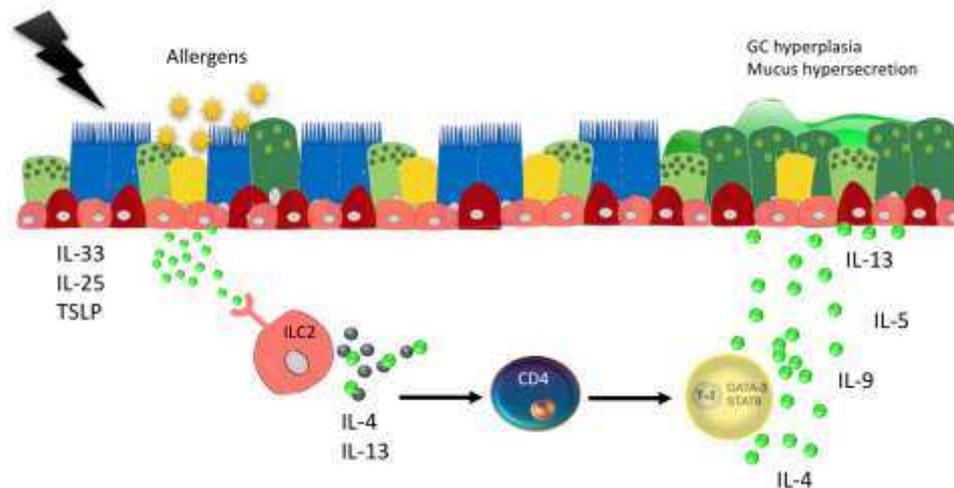
Inflammatory response to injury, pollutants, microbes, oxidative stress or allergens activate chemokines, interleukins (IL) and growth factors, all of them are secreted by a diversity of cell types such as mesenchymal or endothelial cells, lymphocytes, neutrophils, eosinophils and macrophages, but also are secreted by airway epithelial cells during injury and repair.

In the following sections of the manuscript I will focus in the inflammation driven by the Th2-type cytokines, describing the effect of this cytokines in the airway remodelling.

#### IV.3 Th2 cytokines

The group of the Th2 cytokines has a pleiotropic effect in the lung epithelium, such as inflammation with recruitment of lymphocytes and eosinophils, mucous cell metaplasia and hyperplasia, loss in MCCs, altered frequency in ciliary beating and fibrosis (Gomperts et al., 2007)

Secretion of pro-inflammatory cytokines by airway epithelial cells contributes to the pathology. In the inflammation driven by the production of Thymic Stromal Lymphopoietin (TSLP), Interleukin (IL)-33 and IL-25 lead to the induction of innate lymphoid cells (ILC2s) that secrete Th2 cytokines (IL-4, IL-5 and IL-13) and are going to promote the differentiation of the native CD4<sup>+</sup>Tcells into Th2 lymphocytes, contributing with the Th2 inflammatory response releasing different cytokines such as: IL-4, IL-9, IL-13, IL-5 (Divekar and Kita, 2015; Gras et al., 2017) (Fig. 22).



**Fig. 21. Model of Th2 inflammation in the airway epithelium.** Allergens provoke the release of pro-cytokines (IL-33, IL-25 and TSLP) by the airway epithelial cells, leading to the secretion of Th2 cytokines by the ILC2. The secretion of these cytokines provokes the maturation of the CD4 lymphocytes that contributes with the inflammation secreting Th2 cytokines IL-13, IL-4, IL-5 and IL-9.

IL-3, IL-9, IL-5 and IL-13 are Th2 cytokines up-regulating mucous production in asthmatic patients. IL-13 is crucial for regulation of mucus secretion and no other cytokine or mediator of mucus secretion can induce production of mucous in the absence of IL-13. Blocking IL-13 production can be sufficient to block the mucus production (Louahed et al., 2000; Parker et al., 2013; Qin et al., 2016). IL-31 was also related to inflammation in asthma but the group of Parker showed that this cytokine did not induce the asthmatic phenotype in HAECs (Parker et al., 2012).

#### IV.3.1 Interleukin-13 (IL-13) as mediator of allergic inflammation

IL-13 is a pleiotropic cytokine acting through the IL13RA1/IL4A membrane complex inducing activation responses contributing to the inflammatory diseases.

IL-13 together with IL-4 are key cytokines in the pathogenesis of asthma (Lee et al., 2001; Kondo et al., 2002; Oh, Geba and Molino, 2010). IL-4 and IL-13 induce B cell IgE class switch (Oettgen and Geha, 2001; Froidure et al., 2016). However, IL-13 is suspected to be a more central mediator of the pathophysiologic changes induced by allergic inflammation in many tissues.

Surprisingly both cytokines have different roles in airway epithelial differentiation as was shown by Kondo in 2002, where they used primary airway cells of guinea pigs cultured in 3D air liquid interface, they treated the cells with IL-13 or IL-4 during two weeks and showed that only cells treated with IL-13 showed an increase in GC differentiation and MUC5AC overproduction. And together with this GC hyperplasia they

corroborated the decrease in the number of MCCs (Kondo et al., 2002). The same results for IL-13 treatment were observed by Atherton, but in contrast their treatment with IL-4 also induced GC hyperplasia (Atherton, Jones and Danahay, 2003). This could be due to the fact that Kondo and colleagues worked on Pig and, as they said, there was no IL-13 nor IL-4 specific for this specie so they used the human cytokines for treatment, and the group of Atherton worked with human primary cells.

#### IV.3.2 Effect of IL-13 in airway epithelium

Numerous studies show the effects of IL13 treatment over cultured cells or in vivo, the GC hyperplasia and the MCCs decrease is one of the major deadly outcomes for patients with lung diseases. Deciphering the molecular and cellular mechanisms leading to these events is one of the major objectives for the lung research community.

In 2006 Tyner et al. described for the first time a process of “transdifferentiation”. The term transdifferentiation or lineage reprogramming is a process in which one differentiated cell transforms in other differentiated cells without the intervention of any intermediate state or progenitor cell type.

Analysing the effect of IL-13 treatment over cultured cells and in vivo in mice, they corroborated the increase in GCs and the decrease in MCCs. These results raised several questions about the origin of these new GCs or the fate of the MCCs.

They did not detect increase in MCCs apoptosis due to treatment. Surprisingly they showed cells sharing properties of goblet and MCCs as an early response to the IL-13 treatment. Analysing human fresh lung tissue coming from lung transplants of patients with COPD they could also detect these cilia-GCs. Their claim was that the increase on GCs was due to a transdifferentiation of MCCs to GCs (Tyner et al., 2006), and these results were repeated by the group of Brigitte and colleagues, after treatment of human airway cultured cells with IL-13, which showed co-stainings of MUC5AC and FOXJ1 (Gomperts et al., 2007). In order to check this assumption, Tyner et al. blocked the receptor of IL-13 using a IL-13R $\alpha$ 2 fusion protein, after this blockade and after treatment with IL-13 they abolished completely the GC metaplasia and increased the number of MCCs confirming that, in treatment with IL-13, MCCs transdifferentiated in GCs, moreover they showed that the number of CLUB cells was not altered during treatment with IL-13 in both conditions (blocking or not IL-13 receptor), meaning that the source of GC would be due more likely the transdifferentiation events (Tyner et al., 2006). Same results are obtained after inhibiting IL-13 with an anti-IL13 antibody (Kondo et al., 2006).

Interestingly, in 2011 the group of Turner performed a cell-lineage tagging approach in human bronchial epithelial cells. They used a lentiviral cotransduction approach using the promoter of the transcription factor *FOXJ1* to investigate the possible role of MCCs as a progenitors of GCs. With this lentiviral construction they could monitor via immunofluorescence all the cells expression that had expressed FOXJ1 at any moment in the development of the epithelium.

They monitored the histological and cellular changes of the human airway cultured epithelium after treatment with IL-13, showing that after treatment there were some positive cells for FOXJ1 lineage tracing and MUC5AC (6-fold more than in controls without treatment), and they didn't see any effect of the treatment in the basal cell or SC cell population meaning that one of the sources of GCs was the MCCs compartment (Turner et al., 2011).

Already in 2004 Hayashi described a very rare cell type harbouring cilia and mucus granules in the airway epithelium of a model of asthmatic mouse but they didn't define this cell type. Based on the results I showed before, these events seemed to be events of transdifferentiation of MCCs to GCs (Hayashi et al., 2004).

Further than having a role in the GC hyper/metaplasia, IL-13 has also an impact in the MCC population, Brigitte and colleagues studied the cellular and molecular mechanisms of cilia loss in response to IL-13 treatment. They used airway human differentiated cultured cells treated with IL-13, showing a decrease in MCCs and a delocalization of the basal bodies. They analysed the expression of the transcription factor *FOXJ1* and they detected a decreased expression after treatment. Their study revealed a molecular regulation of FOXJ1 by IL-13, via a consensus STAT-binding element they found on the human and mouse promoters of FOXJ1. So, the STAT protein activated after IL-13 treatment was able to bind the promoter of FOXJ1, leading to its inhibition (Gomperts et al., 2007).

Interestingly in 2008 Nakao et al. identified pendrin as a gene inducible by the action of IL-13, as it was mentioned before, pendrin is a molecule involved in the ionic transport in the membrane of airway epithelial cells necessary for a correct mucus formation (Nakao et al., 2008).

Other effect of IL-13 is the attenuation of BPIFA1 expression. BPIFA1 (bactericidal/permeability-increasing fold containing family A, member 1), or SPLUNC1, is expressed in human nasopharyngeal and respiratory epithelium and has demonstrated antimicrobial activity. This protein has been shown to be highly expressed in COPD and CF (Yeh, Lee and Hsu, 2010). Tsou et al. showed that lipopolysaccharide (LPS) induced BPIFA1 expression by activation of the JNK/c-Jun signalling pathway, and that BPIFA1 expression decreased in patients with Th2 inflammation. They then tested whether a IL-13 treatment could impact the expression of BPIFA1 and found that IL-13 suppressed the activation of the JNK/c-Jun pathway, thus decreasing BPIFA1 expression (Tsou et al., 2015).

Yeh and colleagues analysed the expression of BPIFA1 after treatment of differentiated cultured human nasal primary cells with IL-13. Their results also showed decreased expression of BPIFA1 after IL-13 treatment. Suggesting that the inflammatory state of the airways acts in detriment of the immune response of the airways given by the SCs, (Yeh, Lee and Hsu, 2010). These results were confirmed in single cell in vivo analysis of Chronic rhinosinusitis (CRS) patients with different grades of Th2 inflammation (Ordovas-Montanes et al., 2018)

Giovannini-Chami et al. performed an study of the gene expression of different children with allergic rinitis with or without asthma in nasal epithelial cells by microarray (GSE19190) (table 3). They showed the activation of Th2-type response in the allergic patients in comparison with healthy ones by the expression of some genes such as *POSTN*, *CD44*, *GSN*, *CST1*, *NTS* and *ALOX15* (Giovannini-Chami *et al.*, 2012). Also analyzed the differential gene expression between patients with controlled vs non controlled asthma and found a differential panel of genes expressed depending on the severity of the disease such as the up-regulation of *POSTN*, *FETUB* or *DPP4*, or the down-regulation of several INF-genes. The work of Woodruff also described the differentially expressed genes in patients with asthma compared with healthy patients and detected also the high expression of *CLCA1*, *POSTN*, *SERPINB2*, among other also expressed in the gene expression analysis of Giovannini-Chami, moreover, in other study Woodruff and colleagues defined two subgroups of asthma defined by the degree of the Th2 inflammation (Woodruff *et al.*, 2007, 2009).

**Table 3. Representation of the most differentially expressed genes in nasal epithelial cells from patients with allergic rinitis.** Analysis in GEO2R of the GSE19190 data set from (Giovannini-Chami *et al.*, 2012)

Gene.symbol	adj.P.Val	P.Value	logFC	Gene.title
CST1	2.52e-09	7.56e-14	734.912.287	cystatin SN
POSTN	1.12e-03	1.34e-07	392.084.238	periostin
ITLN1	1.38e-02	1.40e-05	319.749.415	intelectin 1
FETUB	1.64e-02	1.85e-05	234.457.095	fetuin B
CST2	5.42e-04	3.26e-08	218.192.784	cystatin SA
DPP4	1.12e-02	5.36e-06	191.148.059	dipeptidyl peptidase 4
CLC	1.83e-02	2.26e-05	182.846.972	Charcot-Leyden crystal galectin
SERPINB2	1.64e-02	1.87e-05	182.745.336	serpin family B member 2
CLCA1	2.39e-01	3.76e-03	175.886.571	chloride channel accessory 1
SLC5A5	1.10e-01	5.60e-04	163.420.268	solute carrier family 5 member 5
CDH26	1.16e-02	8.67e-06	161.704.843	cadherin 26
SLC9B2	4.04e-02	8.85e-05	154.165.317	solute carrier family 9 member B2
TFF3	4.70e-02	1.27e-04	146.724.448	trefoil factor 3

The group of Alevy performed a microarray analysis (GSE37693) over human tracheal and bronchial cells cultured in the liquid-liquid interface system with or without IL-13 in submerged conditions for 2 days and then in air-liquid interface for 3 weeks, they showed the over expression of several molecules related to the IL-13 treatment such as *CST1*, *CCL26*, *POSTN* related to inflammation and genes involved in mucus generation such as *SPDEF* and *FOXA3*, surprisingly there is not an increased expression of *MUC5AC*, probably due to the treatment with IL-13 before differentiation of the epithelium, for instance there is only one mucin over expressed and is the *MUC13* (Alevy *et al.*, 2012) (Table 4).

**Table 4. Representation of the most differentially expressed genes in airway epithelial cells cultured with IL-13.** Analysis in GEO2R of the GSE37693 data set from (Alevy et al., 2012)

GeneID	adj,P,Val	P,Value	logFC	Gene
CST1	5,27E-10	6,70E-14	875285338	cystatin SN
CCL26	2,36E-10	1,00E-14	850616306	C-C motif chemokine ligand 26
POSTN	1,91E-07	1,54E-10	634794614	periostin
ITLN1	8,00E-09	1,69E-12	588227151	intelectin 1
CA2	7,86E-08	4,67E-11	545197904	carbonic anhydrase 2
SERPINB2	2,65E-05	1,45E-07	522713497	serpin family B member 2
ALOX15	3,50E-06	7,54E-09	501357147	arachidonate 15-lipoxygenase
SPRR3	2,26E-04	3,03E-06	477237829	small proline rich protein 3
FOXA3	1,00E-06	1,61E-09	475026727	forkhead box A3
SPDEF	4,06E-08	1,89E-11	464514101	SAM pointed domain containing ETS transcription factor
CAPN14	3,21E-07	3,32E-10	457282801	calpain 14
CD44	1,19E-03	3,42E-05	222605232	CD44 molecule (Indian blood group)
GSN	2,35E-03	8,58E-05	212464412	gelsolin
MUC13	3,29E-07	4,18E-10	37114631	mucin 13, cell surface associated

Periostin, coded by *POSTN*, is a protein of the extracellular matrix, its expression is increased in airway tissues from patients with asthma and in less extend in COPD patients and is induced by IL-4 and IL-13 (Takayama et al., 2006; Sidhu et al., 2010; Górska et al., 2016).

Sehra et al. showed that periostin was already present at low levels in airway epithelium of wt mice, after OVA challenging over these mice there was a 2.5-fold increase of the expression of periostin. However, depletion of periostin in mice and posterior OVA challenging didn't affect the inflammatory response in those depleted animals but, in fact, there was an increase in airway resistance and an increase in IgE levels suggesting an enhanced Th2 response. Finally, they showed that depletion of periostin triggered GC hyperplasia and mucus over production related to the increased expression of Muc5ac and Gob5 (also called Clca1, involved in the regulation of mucus production by GCs induces Muc5ac), to test whether periostin has a role controlling mucus secretion inhibiting GC hyperplasia the authors performed IL-13 treatment over cultured tracheal epithelial cells from mice, after treatment they showed how the ectopic expression of periostin led to a decrease in the expression of Gob5 (Sehra et al., 2011).

Sidhu et al. revealed that periostin is secreted by airway epithelial cells as a major source of this protein, using cultured human airway primary cells they showed that after treatment with IL-13 there was a big increase of periostin secretion in the basal but not the luminal side of the epithelium. They showed that secreted Periostin activated airway fibroblast through TGF- $\beta$ , meaning that secretion of periostin induced changes at the underlying matrix of the epithelia. Periostin thus acted upstream TGF- $\beta$ , and was responsible for the

production of collagen type I, having a capacity of changing the biomechanical properties of the airway epithelia (Sidhu et al., 2010).

The group of Jia et al. analyzed the serum periostin levels of patients with asthma and different levels of eosinophilic airway inflammation, these levels were significantly increased in the subjects with higher eosinophilic profile compared with those with minimal eosinophilic airway inflammation. Their data suggested that measuring the levels of periostin in serum was a good predictor for the airway eosinophilia (Jia *et al.*, 2012).

A recent published work from Ordovas-Montanes et al. showed the analysis performed by single-cell RNAseq, of 12 samples of cell suspensions dissociated from resected sinus tissue, 6 from healthy patients and 6 from patients with polyps with different grades of chronic rhinosinusitis (CRS) (18624 cells in total were analyzed), aiming to reveal the molecular signature of the Th2-type inflammation related to this disease. They detected clusters of basal (KRT5) and apical (KRT8) cells, MCCs (FOXJ1) and glandular (LTF) cells, endothelial cells (DARC, also known as ACKR1), fibroblasts (COL1A2), plasma cells (CD79A), myeloid cells (HLA-DRA), T cells (TRBC2) and mast cells (TPSAB1). A closer look into each cell population showed certain heterogeneity. They showed a subgroup of mast cells expressing HPGDS and PTGS2, suggesting that they may be a dominant source of prostaglandin D2, which is implicated in activation of T helper 2 (Th2) cells. They also described the basal cells as responsible for the secretion of the pro-cytokines IL-25, IL-33 and TSLP.

Analysis of both samples showed the clustering of BCs, SCs, glandular cells, and MCCs. The analysis of the cell number of each population comparing healthy vs inflamed tissue revealed that BCs, SCs, and glandular cells showed the most significant links to the disease state.

SCs differential analysis showed that in the disease these cells supplant the role of secreting antimicrobial proteins for a role in tissue repair, as was shown by Tsou et al. and Yeh and colleagues (Yeh, Lee and Hsu, 2010; Tsou *et al.*, 2015).

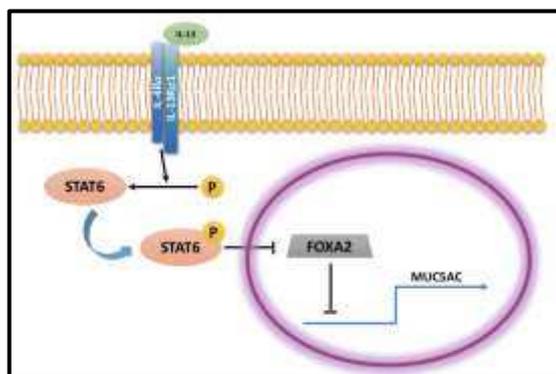
They analysed different grades of severity of the disease, sampling different patients in different inflammatory stages, and they showed that specific cell signature was associated to each level of severity, with a continuum from IFN- $\alpha$ / IFN- $\gamma$ -induced genes to IL-4/IL-13-induced genes at increasing grades of disease severity.

They performed hierarchical cell lineage reconstruction showing how BCs differentiate to mature SCs, they showed in the inflammatory state that the BCs remained proliferative via the expression of several transcription factors such as ATF3, AP-1, p63 and KLF5, which impaired differentiation (Ordovas-Montanes et al., 2018).

### IV.3.3 Molecular pathways involved in GC hyperplasia induced by Th2 type cytokine

#### IV.3.3.1 IL-13 and IL-4 induce GC hyperplasia

IL-4 operates through the IL-4 receptor (IL-4R), a heterodimer of IL-4R $\alpha$  and either  $\gamma$ c or IL-13R $\alpha$ 1, while IL-13 operates through IL-13R the is a heterodimer of IL-4R $\alpha$  and IL-13R $\alpha$ 1. The activity of IL-13 and IL-4 is mediated by the phosphorylation of STAT6 (Kondo et al., 2002; Atherton, Jones and Danahay, 2003; Oh, Geba and Molfino, 2010). STAT6 is a molecule activated by the IL-13 receptor and it is essential for the development of asthma. STAT6 is expressed in normal airway epithelial cells but in patients with asthma the expression is increased (Kuperman et al., 2002), the activation of STAT6 through its phosphorylation leads to the down-regulation of FOXA2, which in normal condition inhibits the transcription of Muc5ac (Kuperman et al., 1998; Wan et al., 2004). Foxa2 was also shown to be downregulated after treatment with IL-13 (Zhen et al., 2007).



**Fig. 22. Model of mucus hyperproduction through the action of Th2-type cytokines.**

Phosphorylation of STAT6 induces the inhibition of Foxa2 promoting the transcription of Muc5ac.

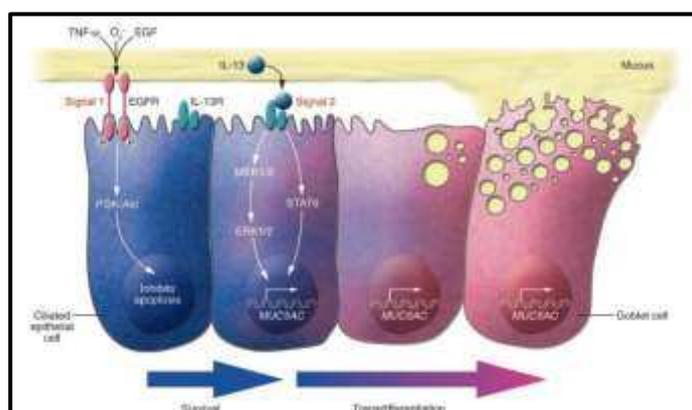
In 2002 Kuperman and colleagues studied the role of IL-13 in the induction of asthma phenotype. They generated a mice model STAT6<sup>-/-</sup> and analysed the effect of IL-13, they did not detect response to the IL-13 treatment, these mice were protected from IL-13 dependent mucus overproduction (Kuperman et al., 2002).

Already in 2003 the group of Atherton showed the implication of the ERK/MAP Kinases and Phosphatidylinositol-3 kinase (PI3K) pathways in the GC hyperplasia after treatment with IL-13 and surprisingly they show no involvement of the EGFR pathway in the IL-13 dependent GC hiperplasia (Atherton, Jones and Danahay, 2003).

In 2006 Tyner et al. described the function of EGFR in chronic pulmonary diseases as a critical component for remodelling towards a chronic asthma/bronchitis phenotype. They used a mice model

inoculated with common influenza virus (Sendai virus: SeV), these mice showed airway epithelial damage, immune-response gene expression, GC metaplasia and ciliated cell hyperplasia; they showed EGFR expressed in the apical surface of MCCs in both models, but, for instance, they detected a higher expression of the activated form of EGFR, phosphorylated EGFR (p-EGFR), in airway epithelium of mice inoculated with virus compared with controls. p-EGFR was located at the apical membrane and in the nucleus of MCCs, the same expression pattern for both molecules was shown in human tissues comparing healthy and asthma patients. Meaning that in MCCs of immune-challenged tissues EGFR is active. To assess the role of EGFR they inhibited EGFR specifically in airway cells using the inhibitor EKB-569, after this blockade they showed the inhibition of MCCs hyperplasia, a partial inhibition of GC metaplasia and a ciliated cell apoptosis. Meaning that the activation of EGFR has a main role in the remodelling of the airways, inhibiting apoptosis of MCCs (Tyner et al., 2006).

These results together with the results obtained after treatment with IL-13, predict a model of action where activation of EGFR inhibits MCCs apoptosis and the action of IL-13 causes the transdifferentiation of the MCCs in GCs (Cohn, 2006; Tyner et al., 2006). (fig.23)



**Fig.23 Ciliated cell differentiation into goblet cells in several steps.**

Step 1: activation of EGFR on ciliated cells and induction of EGFR phosphorylation and activation of PI3K/Akt inhibiting ciliated cell apoptosis.

Step 2: IL-13 binding to its receptor. Upon IL-13 receptor (IL-13R) activation and STAT6 signaling, ciliated cells begin to produce mucins, and lose their ciliated cell surface. Taken from L. Cohn (2006) (Cohn, 2006)

#### *IV.3.3.2 TNF- $\alpha$ signalling increases GC hyperplasia*

TNF- $\alpha$  is released by lymphocytes and neutrophils that are recruited at the airways in response to IL-13. Interaction of TNF- $\alpha$  with its receptor (TNFR1) leads to activation of the transcription factor NF- $\kappa$ B that regulates different genes including MUC genes such as Muc2 or Muc5ac. It has been shown that NF- $\kappa$ B is

essential for mucus production (Lai and Rogers, 2010). Dimers of NF- $\kappa$ B are released and translocated to the nucleus where bind to specific DNA regulatory sequences controlling the transcription of several targets.

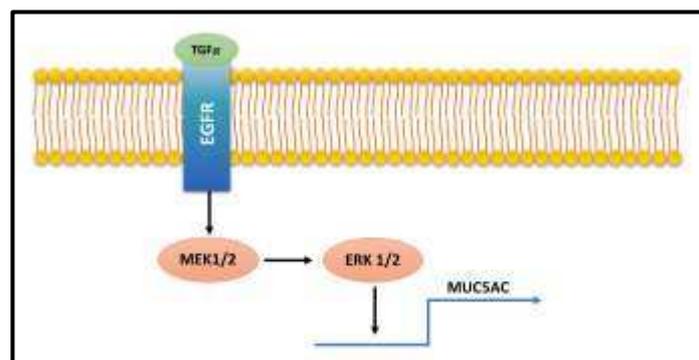
NF- $\kappa$ B was shown to be increased in tissue from patients with COPD and asthma. Blockade of several steps of the pathway of NF- $\kappa$ B demonstrated inhibition of mucus production after OVA challenge in mouse (Boucherat et al., 2013).

#### IV.3.3.3 IL-1 $\beta$ increases mucin synthesis

IL-1 $\beta$  cytokine produced by activated macrophages and eosinophils, induces mucus secretion through 2 different mechanisms, the first, indirect one, activating CD4+ lymphocytes promoting an inflammatory cascade. The second one, direct, IL-1 $\beta$  induces cyclooxygenase (COX)-2 and prostaglandin release. COX-2 interacts with ERK and MAPK cascades, the inhibition of these pathways suppress COX-2 expression and mucin synthesis.

#### IV.3.3.4 EGFR

This pathway had been already described in the chapter III.3, as a pathway for GC differentiation. In inflammatory conditions there is a hypersecretion and a release of cytokines such as TNF- $\alpha$ , this cytokine is responsible for the up-regulation of EGF-R expression leading to the activation of the EGF cascade for mucus production (Takeyama et al., 1999, 2001; Nadel and Burgel, 2001; Lai and Rogers, 2010).



**Fig. 24. Model of Mucus induction through EGFR signalling via PI3K/AKT.** Activation of the transcription of MUC5AC through the MEK/ERK signaling.

#### IV.3.3.5 SPDEF and ALOX15 trigger GC hyperplasia and inflammation

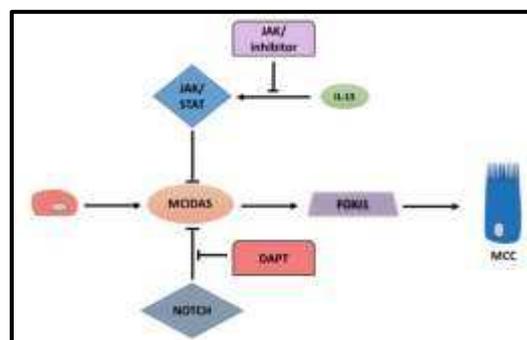
The group of Whitsett and colleagues described the role of SPDEF and FOXA3 in the induction of an inflammatory phenotype in mouse: they showed that expression of *Spdef* in SCs cells triggered extensive GC metaplasia, Th2 inflammation and increased airway hyperresponsiveness. In the same line, when *Foxa3* was expressed in SCs cells, it induced SPDEF expression, leading to the same phenotype. They showed how, in fact, *Spdef* is necessary for the inflammation, airway hyperresponsiveness and GC differentiation after exposure to “house dust mite” (HDM) to induce immune-response in the mice (Rajavelu et al., 2015).

Recently it has been described an effect of ALOX15 and its product 15-HETE in asthma, after the report of its increased expression in asthmatic lungs and after IL-13/IL-4 stimulation in vitro (Zhao et al., 2017). In the work of Zhao and colleagues, they described the effects of ALOX15 in GC hyperplasia and periostin expression. They analysed the effect of ALOX15 in cultured human bronchial cells after stimulation with IL-13, they showed the increase in the gene expression of FOXA3, MUC5AC, POSTN and ALOX15, and its protein products and a decrease in FOXA2. Inhibition of ALOX15 showed decrease in FOXA3, POSTN and MUC5AC and an increase in FOXA2.

Exogenous stimulation with 15-HETE (without previous treatment with IL-13) showed upregulation of MUC5AC and FOXA3.

All these results posit ALOX15 as a regulator of GC hyperplasia and periostin secretion in Th2 induced tissues (Zhao et al., 2017).

The effects of IL-13 over MCCs have been also intensively studied. B. Gerovac and N. Fregien showed that the inhibition of MCCs development was independent from Notch Pathway, in fact, it was through the JAK/STAT signalling pathway, inhibiting JAK in the presence of IL-13 partially restored ciliated cell differentiation (Gerovac and Fregien, 2016).



**Fig. 25. Model of inhibition of MCCs differentiation by Th2-type signature.** Inhibition of IL-13 downstream effectors using a JAK inhibitor restores the MCCs differentiation process.

#### IV.3.3.6 WNT signalling pathway altered in Th2 inflammation

The developmental pathway of WNT is altered in IPF, as was shown by Königshoff and colleagues analysing a mouse model of pulmonary fibrosis and human lung tissues of patients with IPF they detected elevated expression of WNT1-inducible signaling protein-1 (WISP1) which is coded by a target of the pathway (Königshoff *et al.*, 2009; Baarsma and Königshoff, 2017).

The work from Ordovas-Montanes *et al.* in their single-cell analysis of different patients with different grades of CRS, showed the implication of the WNT pathway in the disease. In their analysis they showed the high expression of CTNNB1 ( $\beta$ -catenin) which is a key effector of the WNT pathway and expression of CTGF (specific factor of the pathway) (Ordovas-Montanes *et al.*, 2018). These results confirmed the results of Königshoff, where they revealed and increased expression of *WNT-1*, *WNT-7B*, *WNT-10B*, *FZD2*, *FZD3*, *CTNNB1* and *LEF1* when comparing lung tissue from patients with IPF compared to controls (Königshoff *et al.*, 2008).

Regarding to asthma the gene expression of multiple WNT ligands is positively associated with a Th2 signature such as *WNT-3A*, *WNT-5A*, *WNT-6* and *WNT-10<sup>a</sup>*, and also negatively associated like *WNT-5B*.

Also, as I mentioned before, WNT5A is related to hyperreactivity of the smooth muscle cells in the airways of patients with asthma. WNT5A has been associated with a high levels of TH2-type inflammation (Koopmans *et al.*, 2016). Other authors revealed high expression levels of WNT5A in mononuclear blood cells from healthy donors treated with IL-13, suggesting that, among others, WNT5A could serve as biomarker for therapies against IL-13 (Syed *et al.*, 2007).

#### IV.3.4 IgE/IgA in asthma

As I described before induction of IL-4 and IL-13 increased the IgE levels (Oettgen and Geha, 2001; Sehra *et al.*, 2011; Chung, 2015).

Allergen specific IgE antibodies, produced by B cells, have been considered to play an important role in asthma

In 1998 Nahm studied the IgE antibodies in the secretion of airways from asthmatic patients, they showed how in asthmatic patients the secretion of these antibodies was increased in comparison with controls. IgE has emerged as a target for therapies in patients with severe asthma (Nahm and Park, 1998).

Ladjemi *et al.* analysed the expression of pIgR in asthmatic patients. pIgR is an immunoglobulin receptor that translocates IgA to the apical pole of the cell where it is released. The authors showed that protein expression of pIgR was reduced in tissues from asthmatic patients in comparison with controls. In vitro cultured bronchial cells treated with IL-13 and IL-4 also showed a reduction of the expression of the receptor. They checked the different mechanisms triggering these decrease of pIgR expression, based on previous

knowledge they checked whether the TGF family had a role on this mechanism, because the upregulation of member of this family after treatments with IL-13 or IL-4. They showed that an ectopic expression of TGF- $\beta$  or TGF- $\alpha$  completely or partially abolish expression of pIgR, respectively. Inhibition of STAT6 abrogated the effects of IL-13 and IL-4 over the expression of pIgR. The decrease of pIgR reduced the transport of IgA through the epithelium, IgA has a protective function against several pathogens, so a defect in IgA could lead to a limitation in the capacity of the airways to fight against inhaled pathogens (Ladjemi et al., 2018).

#### IV.3.5 Therapies targeting IL-13/IL-4 signalling

Given the potential role of the th2 cytokines in the development of inflammation and pathological outcomes such as asthma or COPD is logical to design therapies blocking the production of these cytokines reducing their symptoms.

Even that the use of corticosteroids is the most recurrent therapy for asthma, their effect result from the modulation of the Th2 cytokines and their associated inflammation, unfortunately their activity is too unspecific (Wenzel, 2012).

There are therapies directly blocking the activity of IL-13 such as the use of antibodies (lebrikizumab and tralokinumab), unfortunately these therapies did not success in abrogating completely the asthma exacerbations, they worked in the physiological response against allergens but not in cases of eosinophilic asthma.

Using antibodies against the receptor IL-4R $\alpha$  (dupilomab), which inhibits the activity of IL-13 and IL-4, has proven positive results in asthmatic patients.

After the discovery of the role of IL-5 in the differentiation of eosinophils new therapies have been developed against this cytokine such as mepolizumab, which is efficient in patients with eosinophilic asthma but not in patients with asthma due to the interaction with allergens (Wenzel, 2012; Walker and J McKenzie, 2017).

There are also therapies developed against IgE such as antibodies able to bind free IgE (Omalizumab, ligelizumab), reducing the amounts of IgE available to bind its receptors Fc $\epsilon$ RI, this therapy has a very good outcome for patients with moderate-severe allergic asthma (Chanez et al., 2010; Chung, 2015; Gauvreau et al., 2016)

Tiotropium is a long-acting anticholinergic agent, it has been used for patients with lung diseases such as COPD, emphysema, chronic bronchitis and asthma due to its effects over bronchoconstriction and inflammation. This drug is not acting against the cytokines but is reversing the IL-13-induced GC hyperplasia, preventing the decreased expression of FOXA2 and impairing the up regulation of FOXA3, whereas expression of SPDEF was not affected (Kistemaker et al., 2015).



## V. *Single-cell transcriptomics: a powerful tool to analyze cell composition of tissues with high resolution*

---

Evolutionary successes of animals and plants has resulted in a large extent to multicellularity, through which the same organism can develop specialized functions. This raises the question of the origin of the many different cells that can be generated by a same organism. Incidentally, it makes very difficult the identification and classification all cells from an organism, since there is at the moment no comprehensive suitable approach. The recent initiative to build an atlas of all cell types in several organisms (human, mouse, drosophila, axolotl...) represents an interesting option to tackle this problem. Before going into the details of these approaches, it is important to recognize the conceptual work made before by Leeuwenhoeck, Schleidel, Schwann, Virchow, and others, which settled a strong background.

Anton van Leeuwenhoek, known as the “father of the microbiology”, was a Dutch draper and scientist in the Golden Age of Dutch science and technology. He is one of the first microscopists, being best known for his pioneering work in microscopy and his contributions to the studies of microbes. Using a primitive microscope, he was the first to experiment with microbes (that he called animalcules). He was the first to determine their sizes. He was also a pioneer in describing muscle fibers, bacteria, spermatozoa, red blood cells, and blood flow in vessels.

Matthias Jakob Schleiden was a German botanist and a co-founder of the cell theory, along with Theodor Schwann and Rudolf Virchow. He studied the plant structure at the microscopic level and he stated that all the parts of the plant organism were composed of cells. He also recognized the importance of the cell nucleus (that was previously described by the botanist Robert Brown) and described its connections with cell division.

Theodor Schwann was a German physiologist. His many contributions to biology include the development of cell theory, the discovery of Schwann cells in the peripheral nervous system, the discovery and study of pepsin, the discovery of the organic nature of yeast, and the definition of the metabolism.

Rudolf Ludwig Karl Virchow was a German pathologist considered as the “father of the modern pathology”. He contributed to the formulation of the cell theory, being one of the first accepting that the origin of any cell is the cell division of a pre-existing one “omnis cellula e cellula”.

This historical point of view illustrates well the use at different times of biology of different approaches to define a cell, which resulted collectively in the elaboration of a “cell theory”. According to it, the cell is the core unit of all living organisms. Cell is key to our understanding of health and to the ways by which

molecular dysfunction leads to human disease. Distinct cells can be classified by the specific molecules that they express, whatever they are: proteins, lipids, sugars, metabolites, etc. From that perspective, RNA probably represents at the present time the most interesting proxy, since it can be easily quantified at a reasonably good level and low costs. Other options are still worth being considered, as illustrated by the transformative impact of the CD (cluster of differentiation) classification in immunology, which was entirely established through the use of specific antibodies against cell surface markers. Nowadays, the major challenge is to develop quantitative approaches that can be easily used at a single cell level in order to define specific types, states and transitions, in relation or not with particular structures or functions. Distinguishing two closely related cells depends entirely of our ability to identify the right proxies (genes, proteins, functions) that are able to set them apart.

According to the cell theory, each cell is derived by biogenesis from another parent cell. This initial view was in large part an answer to the wrong theory of the spontaneous generation that held that living creatures could arise from non-living matter and that such processes were commonplace and regular. On the contrary, cell theory now postulates a continuum between the primordial cell of an organism (zygote) and any of its descents. Considering that human adult cells all arise from the zygote, a challenging question is therefore to draw the different transitions that exist from one cell type to another, and from one state to another. Overall, different cell fates will depend on many regulators of gene expression that can govern these different choices (Trapnell, 2015a).

After understanding cell diversity through histological and imaging approaches, through the observation of tissue sections, as was done by Ramon Y Cajal around 1900, and then by analysing a small number of markers in each cell (usually proteins or glycoproteins specific to a cell type), we now have in hands the right tools to truly describe the molecular complexity of each cell, through analyses that can be conducted in parallel on a very large number of cells.

The development of next generation sequencing made possible the analysis of the transcriptome of complex tissues or organs at a very high depth and accuracy. This technique helped us to understand how complex are gene expression patterns, their regulations and networks, which are all orchestrated in the different organs and tissues at different developmental stages and homeostasis. Transcriptomics, defined as “high-throughput quantitative study of the total complement of cellular RNA molecules” (Kolodziejczyk and Lönnberg, 2017), is a powerful technique to analyse molecular changes or changes in signalling networks on cell state or activity in development, differentiation or after experimental perturbation (such as response to drugs or chemicals) (Kolodziejczyk and Lönnberg, 2017).

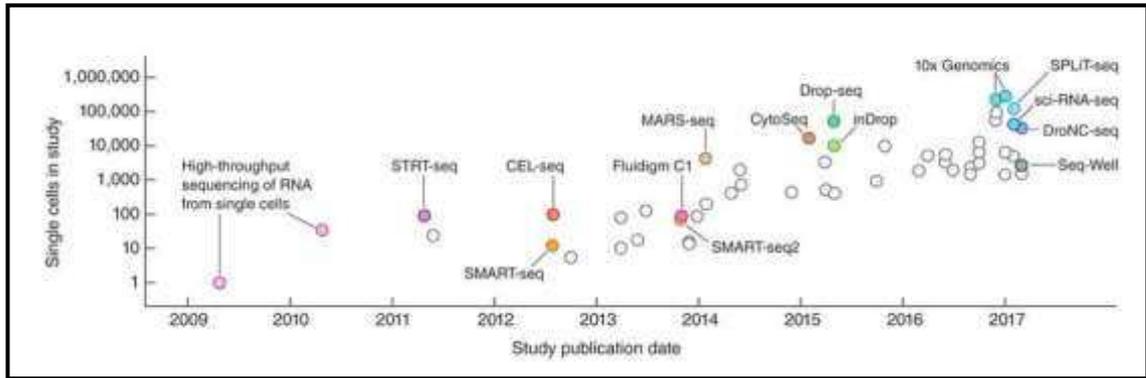
At the same time, the early vision drawn by transcriptomic approaches was biased by the use of large amount of RNA, corresponding in fact to a mixture of many different cells, from various origins and/or in several physiological situations. By mixing information coming from these different cells, effects that may be specific to a unique cell were potentially missed. This does not necessarily imply that all the works performed so far using what is called today “bulk transcriptomic” are necessarily wrong or uninformative, but it is important to understand well that the development of more sensitive approaches is necessary to understand

exactly what can happen in a specific cell. Using standard bulk RNA sequencing protocols, tens of nanograms RNA (and even several micrograms when considering direct RNA sequencing) are necessary to perform the analyses (Svensson, Vento-Tormo and Teichmann, 2018). This amount corresponds to the RNA content of hundreds, thousands, or even millions of cells, and these protocols are just inadequate for very small samples, such as embryos. Observations made with bulk approaches can be useful, but it is essential to keep in mind that they cannot detect any type of modification. For instance, it is still possible to study the expression of a gene that is highly specific in a small fraction of cells. On the contrary, some regulations affecting these same cells can be missed if the corresponding transcripts are more robustly expressed in a larger group of cells. This also introduces the notion that not all cells are behaving in a similar manner at the same time, even when considering homogeneous cell lines (Elowitz et al., 2002; Levsky and Singer, 2003). For all these different reasons, single cell mRNAseq approaches really arose as useful progresses to investigate better the transitional states between different cells in a complex system.

The next progresses were striking. In 2009, Tang et al. published a first description of the technique for sequencing single-cell messenger RNAs in the journal *Nature Methods*. In 2013, this technique was named "Method of the Year" by the same journal. In 2017, new protocols made it possible to sequence up to 50,000 cells per experiment in parallel - in just 24 hours (Tang et al., 2009).

From some perspective, single cell RNA sequencing is the logical development of the bulk transcriptomics approaches that emerged after the sequencing of complex genomes in the 1990<sup>th</sup> and 2000<sup>th</sup>. Single cell transcriptome analysis by mRNAseq arises as a key technique to investigate the transitional state of all the different cells and address this cell-type diversity inside a heterogeneous sample (Trapnell, 2015b).

In few years, many different single-cell RNA sequencing (scRNA-seq) protocols have been developed. The first experiments aiming to perform scRNA-seq were analysing very few cells and over the years the protocols have evolved aiming to analyse hundreds or thousands of cells in the same experiment, these improvements of the protocols were made possible by the development of new techniques to capture and isolate the cells, by the capacity of minimizing the reagent volume needed for the scRNAseq (made possible by progresses in microfluidics) and by improvements of molecular biology protocols, as illustrated by the creation of the cell barcoding (Svensson, Vento-Tormo and Teichmann, 2018) (Fig. 26).



**Fig. 26. Scaling of scRNA-seq experiments.** Cell numbers reported in representative publications by publication date, they key technologies are indicated (Svensson, Vento-Tormo and Teichmann, 2018).

The benefits of using single cell transcriptomics are enormous, one obvious advantage is the possibility of study rare cell types. Another benefit comes from the possible analysis of cell heterogeneity in complex systems such as hematopoietic lineages (Tsang et al., 2015), airway epithelium (Fletcher et al., 2017; Gadye et al., 2017) or brain (Zeisel et al., 2015) where it becomes possible to analyse different number of cells that collaborate together for the integrity and the functionality of the system. With single cell transcriptomics, the different populations can be analysed separately, thus avoiding the averaging effect that occurs in bulk analysis (Kolodziejczyk and Lönnberg, 2017).

## V.1 Challenges of scRNA-seq

### V.1.1 Tissue dissociation

Compared with the many improvements that have already been introduced in experimental protocols of single cell transcriptomics, cell dissociation is probably one of the most important and challenging steps of all the process, although it is essentially based on protocols that were elaborated long time ago, when people were developing cell culture protocols. From one sample to the other, cell dissociation can be easy or difficult, and there is no evidence that the chosen protocol is necessary able to extract quantitatively all the cells to be analyzed.

Various reasons can contribute to a bad cellular extraction: cell-cell interactions, the cells are making strong connections with their neighbours such as tight junctions, as in human adult kidney or lung, which makes enzymatic digestion particularly inefficient (dissociation can be less challenging in the case of developmental studies, made on fetal or infant tissue); cellular morphology, as in adipocytes, which become very fragile after dissociation; cellular ultrastructure, as in cardiac or in muscular tissues, with the fusion of myoblasts in myotubes; cellular size, as in neurons, where the isolation of just one cell can represent a very tedious work. In each of these situations, optimized protocols have to be set up, which have to take into account the specificities of the sample to be analysed. Tissue cell dissociation needs of an accurate and

efficient protocol designed for each type of tissue and cell type. It is important to avoid cell death and transcriptomic modifications due to the process of dissociation. Controls have to be performed in order to know the possible changes in the single cell profiling that can result from this dissociation process. For example, comparison to in situ measurements can help to determine the change in single-cell profiles that were introduced during dissociation.

When cell dissociation becomes ineffective, or too damaging to cells, an alternative is to perform isolation of nuclei, and perform single nucleus RNA-Seq (Habib et al., 2017). This technique, in which nuclei can be directly isolated from frozen tissue, quantify the nuclear RNA. Surprisingly, there is a good correlation between this type of measurement and the more classical quantification of cytoplasmic RNA (or more likely of the pool between cytoplasmic and nuclear RNA). The big advantage of this technique is to allow profiling of archived material.

### V.1.2 Cell isolation

The big challenge of the technique is, in fact, a quantitative isolation of the whole types of cells in a “healthy” form, so that their analysis can be undergone. The situation is very different between immune cells which are non-adherent cells, and most others cell types which must be dissociated from their tissues. The challenge is to dissociate the cells of the tissue of interest using enzymatic or mechanical techniques avoiding cell death and transcriptome modifications due to the dissociation.

Once a single cell suspension is obtained, there are different methods to isolate individual cells for their analysis.

a. The first protocols were using FACS sorting. A major advantage of this approach is to be able to choose the cells that will be analysed by using, for instance, immunofluorescent markers. It is then possible to isolate rare cell populations. There have been several protocols adapted to merge FACS sorting and scRNA-seq and the use of micro-well plates allows loading of cells into minuscule wells with the subsequent reduction in the reagent volume used. It remains that a main disadvantage of the approach comes from the relatively large volumes of reagents that are required at the different steps, which makes this approach quite expensive, despite a relatively low throughput (a few hundreds of cells).

b. Other systems developed for single cell isolation are based on microfluidics. One of them, the C1 Single-cell Autoprep from the company Fluidigm is able to isolate 96 or 800 cells in isolated chambers inside a chip (IFC, Integrated Fluidic Circuit) and inside these chambers the cells are going to be lysed, their mRNA is going to be retro-transcribed and amplified. The advantage of the C1 is the use of much smaller volumes

(nanoliters), which reduces a lot the use of the different reagents, hence the cost of each experiment and improving efficiency. Another advantage is that the possibility to image the cells inside their chambers, in order to control the occurrence of doublets. The C1 machine also offers the possibility to implement and develop novel protocols. It is with this device that our group has developed a cost and labor effective 5' selective single cell transcriptomic profiling approach suitable for Ion Torrent and Illumina sequencers (Arguel et al., 2017). Disadvantages of this system are: (1) the low capture efficiency: to only capture 96 cells, a suspension with thousands of cells is needed, which is not always possible; (2) the size restriction of the cells to load in the IFC, since there are different IFC depending on the size of the chambers, meaning that each cell suspension will have to be uniform in size; (3) High rate of doublets which can only be detected upon cautious imaging of each well, which is not possible using an 800-cell IFC.

c. Droplet based microfluidics methods such as inDrop (Klein et al., 2015), Drop-seq (Macosko et al., 2015) or Chromium from 10X genomics (Zheng et al., 2016). These methods load single cells in aqueous droplets which are emulsioned in an oil phase, then fused with other droplets containing barcoded beads and the different reagents necessary to perform cell lysis, reverse transcription and cDNA amplification. These methods are largely used nowadays. They allow the analysis of thousands of single cells in one single experiment, and work well with limited amounts of isolated cells.

### V.1.3 Cell barcoding.

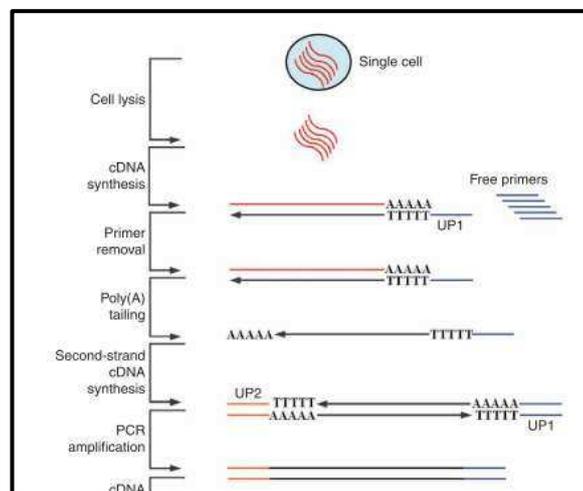
Adding a barcode to the all the cDNAs generated from each single cell allows multiplexing, meaning that at the end of sequencing, it will be possible to deconvolute the signal and attribute each transcript to an identified cell, in just one step. The approach reduces in an important manner the technical biases, since most of the steps are performed in a unique tube for all the cells. Less reagents and less processing are required and reproducibility is increased. Islam et al. published the first protocol or scRNA-seq using cell barcoding (Islam et al., 2011).

Barcoding of the cells captured in droplets or in micro-well plates is accomplished by the use of beads, these beads have the RT primers (generally a poly-T) containing the cell barcode sequence, this way each cDNA from every single cell can be barcoded in each isolate. The problem of multiplexing is the high number of barcodes needed (as much as cells you want to process), for large cell populations analysis could be a very costly step in the protocol. There are different ways to avoid this, one is to combine short barcodes into longer ones (H. C. Fan, Fu and Fodor, 2015; Klein et al., 2015). With this method, it is possible to reach a number of 147 456 barcodes. In a second approach, very long random barcodes are synthesized, they have to be long to avoid doublets in their synthesis (Macosko et al., 2015; Gierahn et al., 2017). This allows the synthesis of 16.7 million barcodes.

#### V.1.4 Untargeted amplification of transcriptomes.

As mentioned before, the amount of RNA needed to perform standard bulk RNAseq reaches the scale of nanograms to micrograms. For a unique single cell, the amount of RNA is in the range of 1 to 50 pg. A solution to get sufficient material from so little RNA is the retrotranscription of the mRNA to cDNA followed post amplification.

For the untargeted retrotranscription (RT) of the mRNA most of the techniques use the poly A tail of the mRNA to start the RT using a poly (T)-oligonucleotide that contains an adaptor sequence. Once the first strand of cDNA is generated, one possibility is that a transferase adds a polyA tail to the extremity 3' of the first strand so using other oligo-poly-(T), with other adaptor incorporated, the RT can start to synthesize the second strand (Tang et al., 2009).

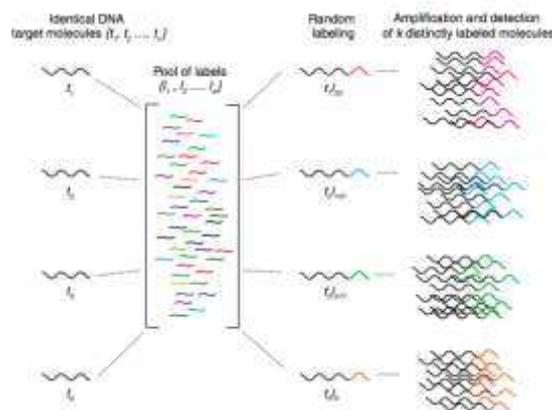


**Fig. 27. Schematic of the single-cell whole-transcriptome analysis.** After cell lysis, the mRNAs are reverse-transcribed into cDNAs using a poly(T) primer with anchor sequence (UP1) and unused primers are digested. Poly(A) tails are added to the first- strand cDNAs at the 3' end, and second-strand cDNAs are synthesized using poly(T) primers with another anchor sequence (UP2). Then cDNAs are evenly amplified by PCR using UP1 and UP2 primers, fragmented, and P1 and P2 adaptors are ligated to the ends. From (Tang et al., 2009).

The most used method to ensure full transcription of the mRNA is the use of a Template-Switching-Oligo (TSO). This method relies on the capacity of transcriptases from a “Moloney” murine leukemia virus to add some nucleotides, mostly cytosines, when the RT reaches the end of the first strand. This addition of a short sequence at the end of the strand allows the polymerase to start the RT of the second strand, this method was used by Islam and colleagues (Islam et al., 2011) in their single-cell tagged reverse transcription (STRT) protocol.

### V.1.5 Unique Molecular Identifiers

The group of Fu also developed UMI (Unique Molecular Identifiers) (Fu et al., 2011), which represents a powerful strategy to better control the biases introduced by the multiple cycle of polymerase chain reaction that were necessary to amplify the cDNA. To count single molecules, the technique entitled “stochastic labelling” consists in the attachment of a random set of labels at one extremity of the cDNA. With this technique they converted a population of identical DNA molecules in a population of distinct DNA molecules, differing by a random sequence located at one extremity of the DNA. UMIs are used as an internal validation control. During the RT, each molecule of cDNA is tagged with random sequence acting as a UMI. The UMI are going to reflect counting of the initial unamplified cDNA molecules that is much less biased than using number of reads per kilobase per million reads (RPKM). This approach was developed by Islam (Islam et al., 2014) and Jaitin (Jaitin et al., 2014).



**Fig. 28. A schematic representation of the labeling process.** An example showing four identical target molecules in solution. Each DNA molecule randomly captures and joins with a label by choosing from a large, nondepleting reservoir of  $m$  labels. Each resulting labeled DNA molecule takes on a new identity and is amplified to detect the number of  $k$  distinct labels. From (Fu et al., 2011).

## V.2 Single cell RNA-seq protocols

The vast majority of the scRNAseq protocols are designed to sequence the 3' extremity of the transcript, meaning a starting point of sequencing through a poly(T) sequence. The sequencing of repetitive tandems can be tricky, and usually the sequencing give rise to poor reads quality. Conversely sequencing of the 5' end of the cDNA does not present this problem, and allows the sequencing of starting transcription sites. The group of Islam (Islam et al., 2014) described a first approach to sequence through the 5' extremity of the transcripts, in which the cell barcode was inserted into the cDNA during the fragmentation step, which is really costly for the huge amount of transposase to use. They also used a very short UMI which became saturated in the case of very highly expressed transcripts. The protocol that we developed (Arguel et al., 2017) avoids these issues

thanks to a longer UMI design. While the other scRNAseq protocols were developed to be used in Illumina sequencers, our protocol was compatible with Illumina sequencers but also with the IonProton.

Here will be described the technical details of the main scRNA-seq protocols that have been used since the early times of single-cell developments.

### V.3 Tang method

The method developed by Tang and collaborators (Tang et al., 2009) is an improved version of other developed for single cell microarray analysis (Kurimoto *et al.*, 2006). Cells were manually picked under microscope, lysed and the polyadenylated RNA was reverse transcribed into cDNA using an oligo-dT primer carrying a specific anchor sequence (UP1). The obtained cDNAs were up to 3 kb length (0.85kb with the previous method). After the RT, a polyA tail was added to the 3'-end by using a terminal deoxynucleotidyl transferase and a second cDNA strand was synthesized using a poly-T primer carrying a second anchor sequence at 5' called UP2. The cDNA obtained then is carrying in both ends primer sequences suitable to be amplified by PCR. The adding the adaptors needed for sequencing they sequenced the cDNA using SOLID sequencer (Applied Biosystems).

The method still has the limitation of 3'-Bias, the majority of the reads are mapping in the 3' portion of the transcripts. Moreover, this method is not strand specific because this information is lost during the PCR amplification.

### V.4 Quartz-seq

It is an improved version of the Tang method (Sasagawa et al., 2013; Picelli, 2017) and consist in 5 main steps:

1. Reverse transcription with an RT primer to generate the first-strand cDNAs from the target RNAs.
2. Primer digestion with exonuclease I; this is one of the key steps to prevent the synthesis of byproducts.
3. Addition of a poly-A tail to the 3' ends of the first-strand cDNAs,
4. Second-strand synthesis using a tagging primer, which prepares the substrate for subsequent amplification.
5. PCR enrichment reaction with a suppression PCR primer to ensure that a sufficient quantity of DNA is obtained for the massively parallel sequencers or microarrays.

All five steps are completed in a single PCR tube without any purification.

The amplified cDNA contains WTA adaptor sequences from the RT primer and the tagging primer. The amplified cDNA is then used in a massively parallel sequencer (Quartz-Seq) and a microarray system (Quartz-Chip). For the Quartz-Seq, the amplified cDNA is fragmented using the Covaris shearing system which is a focused-ultrasonicator, it was the preferred method for cDNA fragmentation specially with long molecules. The fragmented cDNA is ligated to adaptors, which enables the multiplex production of paired-end (PE) sequences.

The main technique improvements are:

- Removal of the remaining primers after the RT procedure. The minimum primer concentration for the RT is used. After RT, remaining primer is removed by treating with exonuclease I, which digests single-strand DNAs such as primers. This exonuclease I digestion suppressed the synthesis of byproducts
- To prevent amplification of the modified primers, suppression PCR technology is used. Suppression PCR is very effective in the suppressing amplification of small-size DNA that contains complementary sequences at both ends of the template DNA.
- To shorten the length of the remaining primers modified by the terminal transferase and thus make them targets for suppression PCR, reaction time of the terminal transferase is minimized. This restriction suppresses the synthesis of byproducts.

## V.5 Smart-seq and smart-seq2

This method was developed to improve transcriptome coverage. It is based in the SMART reaction (Switching Mechanism At the end of the 5'-end of the RNA Transcript) (Ramsköld et al., 2012; Picelli, 2017).

As the Tang method the mRNA is primed with an oligo-dT primer. To avoid the bias due to the 3'-end, a reverse transcriptase from the Moloney Murine Leukemia Virus (MMLV) is used since it has the ability of inserting several nucleotides (3 or 5 cytosines) when it reaches the 5'-end of the cDNA template. These nucleotides are used for the annealing of an oligonucleotide, Template Switching Oligo (TSO), that carries 3 guanoses at its end 3' and helps to synthesize a cDNA template. This TSO allows the insertion of a known sequence in 5' of the cDNA; usually the same sequence as in 3' is inserted to perform PCR of all cDNAs in a cell using one single primer.

Even though this brings a clear improvement to the Tang method, it still presents some limitations such as:

- Low read coverage in 5'end of the transcripts.
- Under representation of transcripts with a high GC-content, probably due to the secondary structure of RNA that the DNA polymerase cannot overcome during the PCR.
- High cost of reagents to prepare the libraries for sequencing if the authors were planning to sequence hundreds of cells.

In order to improve the protocol a new one was proposed called SMARTSeq2. It is based in the previous one, it starts with the RT reaction, using a MMLV-based enzyme and an oligo-dT with a known sequence at its 5'-end. Since one of the most inefficient steps is the RT reaction, the researchers worked to improve as much as possible the processivity using different additives such as Betaine plus magnesium chloride (maximizing the cDNA yield). Also, the TSO was modified replacing the terminal riboguanosine with a locked nucleic acid (LNA)-modified deoxyguanosine (LNAs have a strong thermal stability and can anneal strongly to the untemplated 3' extension of the cDNA). The use of LNAs can increase the RT but can also create artefacts. Another critical point in the process is the PCR amplification step in which the amplification bias is considerable. The authors tested different polymerases to avoid bias in this step and they found that the KAPA High-Fidelity DNA polymerase performed the best. The biggest issue with the Smart-seq and the Smart-seq2 is that the samples cannot be pooled until just before the sequencing. So, library preparation is labor-intensive and costly.

## V.6 Tag-based sequencing methods: CEL-seq, CEL-seq2 and MARS-seq

### V.6.1 CEL-seq: (Cell expression by linear amplification and sequencing)

Hashimshony and collaborators used a different approach for amplifying the RNA (Hashimshony et al., 2012). Instead of doing a PCR-based amplification, they performed an IVT (in vitro transcription). This process, in single cells, is not always easy due to the low amount of RNA present in each cell in comparison with the amount needed for one round of amplification using the IVT protocol (400pg). In the CEL-seq method this problem is solved pooling barcoded samples, thus obtaining enough material to perform an IVT in an efficient way.

The CEL-seq method starts with a RT reaction using a composite primer carrying a T7 promoter sequence at the 5'-end followed by the 5'-Illumina sequencing adaptor sequence, a unique barcode and an anchored poly(T) at the 3'-end. After RT and second strand cDNA synthesis, the samples are pooled and undergo IVT reaction. The antisense RNA (aRNA) is then fragmented to the appropriate size for sequencing, a 3'-Illumina adaptor is added by ligation, the RNA is reverse transcribed to DNA and, finally, the 3'-most fragments (the only ones containing both Illumina adaptors and a barcode) are selected. Paired-end sequencing is performed, where the first read identifies the barcode, and thus the cell identity, while the second recovers the mRNA transcript.

The advantages of this method over the other ones published before are:

- Strand specificity (more of the 98% of exonic reads coming from the sense strand).
- Barcoding efficiency (>96%).
- Only the RNA fragments closest to the polyA tail are selected, so the estimation of the expression levels is much more accurate and there is no need of normalization by gene length.

The disadvantages:

- Strong 3' bias.
- Low sensitivity for low expressed transcripts.

#### V.6.2 CEL-seq2

It is an improved version of the CEL-seq method (Hashimshony et al., 2016). It uses UMI and eliminates the ligation step by inserting the illumina adapter directly during the RT step as a 5'-tail attached to the random hexamer primers. This method has been also implemented on the Fluidigm C1 system.

#### V.6.3 MARS-seq (Massively parallel RNA Single-cell sequencing)

The major advantage and innovation of this method is the introduction of UMI in the oligo-dT primer, this allows random barcoding during cDNA synthesis, which allows the counting of unique RNA molecules (Jaitin et al., 2014). This method contains three levels of multiplexing; molecular-, cellular- and plate- level tag to facilitate molecule counting with high degree of multiplexing. The strategy is to characterize cell subpopulations by first classifying single cells based on low-depth RNA sampling, and then study transcriptional profiles at high resolution by integrating data from dozens to hundreds of cells within each unsupervised class.

#### V.6.4 STRT-seq (Single-cell tagged reverse transcription sequencing)

It has been developed by Sten Linnarson's group (Islam et al., 2011). In this method, single cells are also reverse transcribed using an oligo-dT and a MMLV-based enzyme. It uses TSO for the template switching reaction. The TSO contains 6-base random barcode upstream the three riboguanosines and has 2 uraciles to make sure that it is degraded before PCR by the use of uracil-specific excision reagent (USER). The TSO also contains a biotine group in the 5' end to prevent secondary template switching reactions. The cDNA was amplified by PCR and retained with streptavidine beads making sure that only the 5' fragments are kept for the subsequent library preparation.

This protocol has also been implemented for the Fluidigm C1 system (STRT/C1) (Islam et al., 2014). One disadvantage of this method (and all the ones tag-based sequencing methods) is that is not possible to detect SNPs or splice variants located outside of the 5' portion of the transcript.

## V.7 Array based and embulsion-based methods: Cyto-seq, Drop-seq, In-drops and 10x Genomics

### V.7.1 CytoSeq

Cytoseq consists in loading a cell suspension on custom arrays (H Christina Fan, Fu and Fodor, 2015). Then the array is loaded with a magnetic bead library. The beads are coupled with hundreds of millions of oligonucleotides bearing an universal PCR priming sequence, a unique 8 bp cell label (identical for all the oligonucleotides on the same bead but unique for each bead), another 8 bp molecular index variable among oligonucleotides on the same bead, and a oligo-dT sequence at the 3'-end. In the end, all of the nucleotides in a bead carry the same cell label but a different index. A lysis reaction occurs and the polyA tail hybridates with the oligo-dT of the bead, then all the mRNA fragments are attached to the beads with the corresponding labels so the mix can be pooled. All mRNAs are retrotranscribed and amplified in the same tube to be eventually sequenced.

This approach is a powerful tool for studying the expression profiles on a large and heterogeneous cell population because it has no cell-size restriction.

Cyto-seq is able to retrieve only the 3'-most terminal portion of the transcripts.

### V.7.2 InDrop (indexin droplets)

Is similar to cyto-seq but with emulsion droplets instead of arrays. Each single cell is encapsulated in droplets together with lysis buffer, RT reagents and a barcoded hydrogel microsphere. Each microsphere is coupled with oligonucleotides encoding one out of 147 456 unique barcodes, up to 3000 cells can be barcoded in a unique way. The oligonucleotides coupled to the microsphere are photo-releasable so just with the exposure to the UV light the barcodes are released from the beads. Each barcode contains the promoter for the T7 RNA Polymerase, a 5'-Illumina sequencing primer, a cell barcode, a UMI and an oligo-dT tail. The RT reaction is performed inside the droplets, before all the droplets are broken up and the library preparation continues in a single tube according to the CELseq/ MARS-seq protocol.

One disadvantage of this method is the very low capture efficiency of transcripts (only 7%), which allows to reliably detect only transcripts present at 20-50 copies per cell (Klein *et al.*, 2015; Macosko *et al.*, 2015) .

### V.7.3 DropSeq

Drop-seq also requires a cell suspension and coencapsulates each single cell with a barcoded bead in nanoliter-scale droplets. The oligonucleotides on all beads contain a common sequence for PCR amplification, followed by a cell barcode (shared by all the oligonucleotides on the same bead), a UMI (different between all the oligonucleotides on each bead) and an oligo-dT sequence to capture mRNA molecules.

The cells are simply lysed after being isolated in droplets, so that the poly(A) tail of the mRNAs hybridize to the oligo-dT tail on the beads forming STAMPs (Single-cell Transcriptomes Attached to MicroParticles). The droplets are broken and the RT is performed in a single tube. The single-stranded cDNA is then amplified via PCR and undergo fragmentation with the commercial Nextera XT kit (Klein *et al.*, 2015; Macosko *et al.*, 2015) .

#### V.7.4 10X Genomics (Chromium device)

Droplet-based system enables 3' mRNA digital counting of up to tens of thousands of single cells. ~50% of cells loaded into the system can be captured, and up to 8 samples can be processed in parallel per run. Reverse transcription takes place inside each droplet, and barcoded cDNAs are amplified in bulk. The resulting libraries then undergo standard Illumina short-read sequencing (Zheng *et al.*, 2016). This protocol is based also in using gel bead emulsion (GEM), GEM generation takes place in a 8 channel microfluidic chip that encapsulates single gel beads.

Each gel bead is coupled with barcoded oligonucleotides that consist of:

- i) Sequencing adapters and primers (P7)
- ii) 14bp barcode drawn from ~750,000 95 designed sequences to index GEMs. (10XBC)
- iii) 10bp randomer to index molecules (unique molecular 96 identifier, UMI). (R2)
- iv) Anchored 30bp oligo-dT to prime poly-adenylated RNA transcripts. (Poly-(dT)VN)

Cell lysis begins immediately after encapsulation. Gel beads automatically dissolve to release their oligonucleotides for reverse transcription of poly-adenylated RNAs. Each cDNA molecule contains a UMI and shared barcode per GEM and ends with a template switching oligo at the 3' end. Next, the emulsion is broken, and barcoded cDNA is pooled for PCR amplification, using primers complementary to the switch oligos and sequencing adapters. Finally, adapters and sample indices are incorporated into the libraries to be compatible with next-generation short sequencing

This streamlined approach enables parallel capture of thousands of cells in each of the 8 channels for scRNA-seq analysis.

## VI. *Cell trajectories inference from single cell data sets*

---

One fundamental question is to understand how cell lineages from complex tissues or organism are formed. Finding an answer for this question is essential not only to monitor the normal tissue development and homeostasis but also is essential to develop our understanding of pathological situations such as cancers, or in our case pathological remodelling of the airway epithelium in chronic lung diseases.

Until now the different techniques that accomplish cell lineage tracing consist in the introduction of a heritable mark in a cell, this mark allows the researchers to follow the progeny of this cell. This technique makes possible the detection of a certain number of cells harbouring this mark, meaning that all these cells come from the same founder cell. By looking at the heterogeneity of the cell types found in the progeny, it is possible to determine the potency of the founder cell.

Nowadays, the development of single-cell transcriptome makes possible to perform “cell trajectories” independent from any genetic tracing. Lineages can be inferred from analyses of the transcriptome of the different single cells, without making any prior hypothesis about the populations of cells that are involved. In developing/differentiating tissues, some cells are transitioning from one cell state to another. If sufficient number of cells in these transitioning states are captured in the single cell experiment there may be enough information to place them on the differentiation trajectories. These computational methods are based on the hypothesis that cells with very similar gene expression profile arise from the same lineage, which can be assessed by the similarity of the transcriptome of these cells. It results from these analyses a clustering of these cells in different groups.

Over the last years, many computational methods have been developed to analyse single cell RNA seq experiments and perform lineage inference.

### VI.1 Dimensionality reduction-based algorithms

a) Monocle (Trapnell et al., 2014) was one of the first algorithms applied for differentiation trajectory construction. This algorithm is based on the use of an independent-component analysis (ICA) to project all the cells in a 2-dimensional space (usually only 2). The following step is the construction of a minimum spanning tree (MST) that is a subset of connected points in a graph, these connections should be the shorter as possible and not making circles, and the definition of a “track” connecting the two most different populations of cells. All other cells are going to be projected into this “track” resulting in a 1-dimensional ordering of all the cells, which is called “Pseudotime”. The pseudotime predicts the lineage trajectory of all the cells. When using Monocle, the “sense” of the lineage is not defined, so that some pre-knowledge is required in order to provide the right starting and ending points of the lineage. Monocle was not very efficient to track branches in a lineage, and the software was therefore restricted to the analysis of linear trajectories. This issue has been

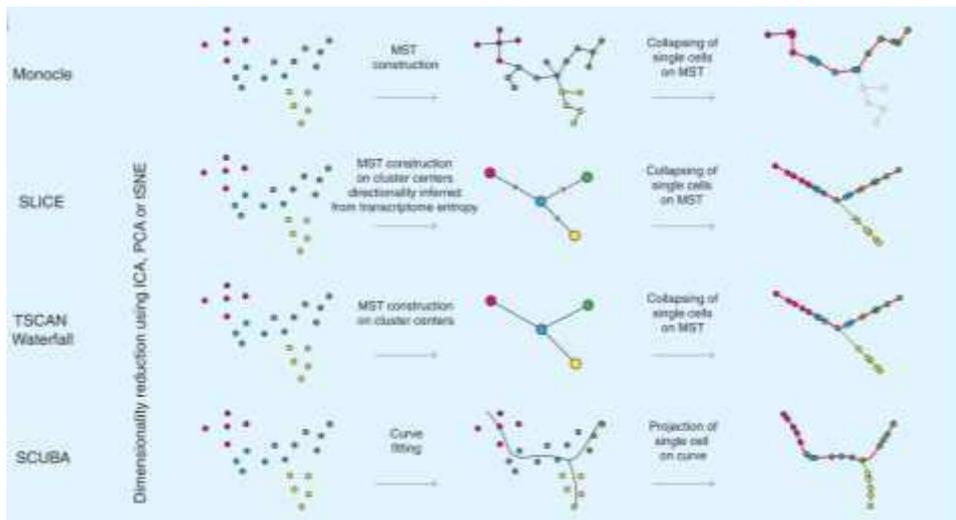
solved with the release of Monocle 2 (Qiu et al., 2017), which perform the analysis in a higher dimensional space: this approach is able to perform more intricate trajectory analysis (Fig. 29).

b) SLICE (Guo et al., 2017) (Single Cell Lineage Inference Using Cell Expression Similarity and Entropy). This algorithm uses predefined cluster of cells instead of the single cells that were considered by Monocle. The use of cell clusters largely simplifies the MST. SLICE consists in the use of transcriptomic entropy as a measure for differentiation. With this, it can detect the less differentiated cell population and create a starting point for the pseudotime. The algorithm can build complex branching trees that are defining different coexisting differentiation pathways (Fig. 29).

c) SCUBA (Marco et al., 2014) (Single Cell Using Bifurcation Analysis). This algorithm also uses predefined cluster of cells to create the MST. It reduces data dimensionality by using t-Stochastic Neighbor Embedding (tSNE), followed by fitting of a smooth curve. With this method the authors analysed two different data sets and successfully reconstructed the cellular hierarchy during development of mouse embryos. They defined the dynamic changes in the gene expression patterns and were able to predict possible perturbations in the trajectory with the perturbation of key transcriptional regulators (Fig. 29).

d) TSCAN (Ji and Ji, 2016) (Tools for Single Cell Analysis). This method also groups the cells into clusters then uses a MST approach to order the cells connecting the center of each cluster. The Pseudotime is obtained projecting every cell into the MST. The authors analysed a single-cell data set of hematopoietic cells and revealed the importance of a specific regulator (HOPX) in the formation of blood cells (Fig. 29).

e) Slingshot (Street et al., 2018) uses dimensional reduction, constructs an MST to identify the key elements of the global lineage structure, and then uses simultaneous principal curves to fit smooth branching curves to these lineages. The cells are then projected into the resulting tree in an ordered lineage trajectory that includes bifurcations.



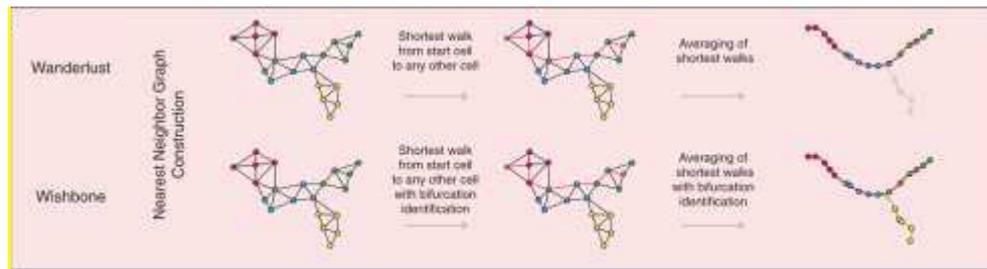
**Fig. 29. Overview of Lineage Reconstruction Algorithms.** Lineage reconstruction algorithms based on dimensional reduction. Monocle uses independent-component analysis, followed by the construction of a minimum spanning tree (MST) connecting all cells. Connecting the two cells furthest away from each other identifies a backbone. Directionality can be provided by the user through the identification of a root cell. Large side branches are excluded, and remaining cells are projected onto the pseudotime backbone. SLICE constructs a MST of cluster centers, and directionality is inferred from transcriptome entropy. Single cells are projected on the edges connecting the cluster centers. TSCAN and Waterfall also constructs a MST based on the cluster centers, followed by projection of the single cells onto the edges to align cells in pseudotime. SCUBA uses tSNE for dimensionality reduction followed by the fitting of a smooth curve. Single cells are projected on the smooth curve to order them in pseudotime. Monocle, TSCAN, Waterfall, and SCUBA all require user input to infer directionality. From (Kester and van Oudenaarden, 2018)

## VI.2 Nearest neighbour graph-based algorithms

In these algorithms each cell is connected to its nearest neighbors, linking together similar cells.

a) Wanderlust (Bendall et al., 2014) Consists in the ordering of the cells by determining each cell position by performing steps between neighboring cells. It generates a collection of the shortest “walks” from one cell to a neighbour. This way, several trajectories are obtained and the algorithm calculates the most probable one (Fig. 30).

b) Wishbone (Setty et al., 2016) Uses a cell ordering technique similar to that of Wanderlust, but is able to identify bifurcations/branches (Fig. 30).



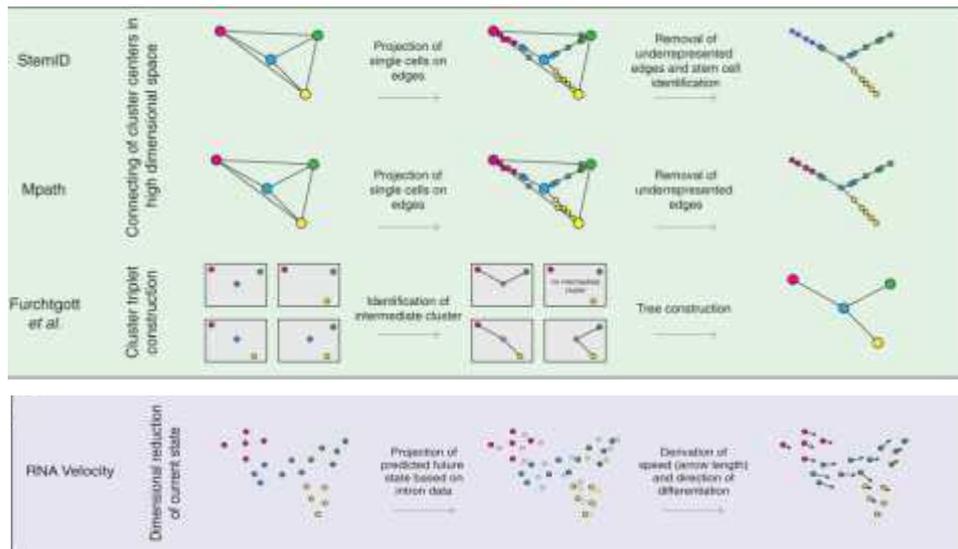
**Fig. 30. Overview of Lineage Reconstruction Algorithms.** Lineage reconstruction algorithms based on NNGs. Both Wanderlust and Wishbone start with the construction of a NNG. A collection of shortest walks, from a user-defined root cell to all other cells in the graph, is then used to construct the lineage trajectory. Wishbone has the added benefit that it can identify bifurcations in the lineage trajectory. From (Kester and van Oudenaarden, 2018)

### VI.3 Connecting of cluster centers in high dimensional space

a) StemID (Grün et al., 2016). This algorithm clusters the cells using k-medoid, the clusters are connected by their centers in a high dimensional space then the cells are projected in the edges between the clusters. Then StemID identifies a stem cell population from the entropy of the cluster determining the starting point of the network like in the SLICE method (Fig. 31).

b) Mpath (Chen et al., 2016). This algorithm clusters the cells using hierarchical clustering. As in the StemID algorithm, it connects the centers of the clusters, creating a network and then projecting all the rest of the cells in this network. Mpath can identify linear and branching lineage pathways, it does not need a lot of cells to already determine the trajectories and it can define the starting point of the network just using the gene signature of genes from the most differentiated cells (Fig. 31).

c) Cluster triplet construction. This algorithm was first used for bulk RNA-seq analysis (Heinäniemi et al., 2013) and later on was used by Furchtgott (Furchtgott et al., 2017) to analyse scRNA-seq data. It consists in the generation of cell clusters, then grouping these cell clusters in 3 and performing differential gene expression analysis to determine the order between clusters in a trajectory. after all the relations of all the triplets of cluster are disposed generating a lineage tree (Fig. 31).

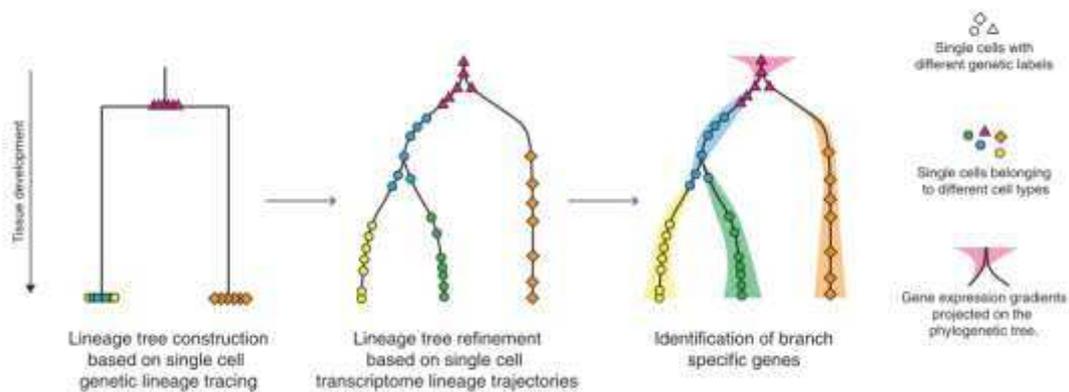


**Fig. 31. Overview of Lineage Reconstruction Algorithms.** Lineage reconstruction algorithms based on cluster networks. Both StemID and Mpath start by connecting all cluster centers in a high dimensional space. Single cells are then projected on the edges between the clusters, and underrepresented edges are removed from the graph. StemID identifies a potential stem cell population based on transcriptome entropy. The Furchtgott method infers the intermediate cluster (if possible) from each triplet of clusters in the data, followed by tree construction based on the triplet relations. From (Kester and van Oudenaarden, 2018)

d) Dimensional reduction of current state. RNA velocity (La Manno et al., 2018) “RNA velocity” is defined as the time derivative of the gene expression state. It can be inferred by the quantification of unspliced and spliced mRNAs, with the assumption that unspliced transcripts are a prediction of the state of a cell within the next hours. The unspliced fraction of all the transcripts of all the cells is going to be represented as a vector that points to the future state of a cell, spliced state. Thus, RNA velocity is a high-dimensional vector that predicts the future state of individual cells on a timescale of hours (Fig. 31).

## VII. Combination of single cell transcriptomics and cell lineage labelling

Cell lineage inferring is powerful to decipher cell lineages with no “a priori” but is not experimental-based. In order to demonstrate experimentally the origin of each cell in a given system, new experimental techniques allowing the combination of single cell transcriptomics and cell lineage labelling have been published. These new techniques have a powerful outcome, integrating genetic inheritance by lineage labelling and the differentiation trajectories derived from transcriptome analysis of single cells (Kester and van Oudenaarden, 2018). The combination of both approaches generates an integrative scheme of the differentiation process. (Fig. 32)



**Fig. 32. Combination of Single-Cell Genetic Lineage Tracing and Single-Cell Transcriptomics.** First a phylogenetic tree is constructed based on genetic labels identified in single cells. This tree can then be refined using transcriptomics-based lineage reconstruction algorithms. Finally, gene-expression gradients can be projected onto the phylogenetic tree to identify gene-expression dynamics throughout the system. From (Kester and van Oudenaarden, 2018)

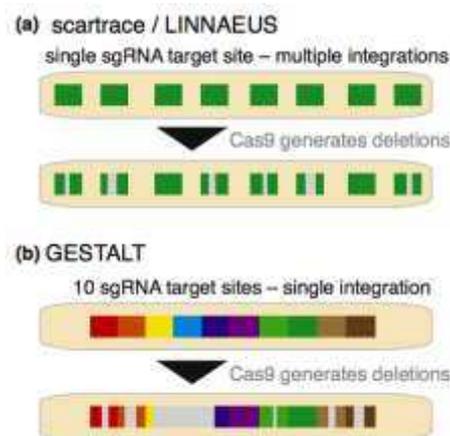
The different techniques combining both approaches are classified in two categories:

### VII.1 Prospective lineage-tracing

a) Model of Viral Barcoding-Based Lineage Tracing: mice cells can be infected with libraries of virus harbouring unique DNA barcodes. With this technique, founder cells are going to be labeled with unique barcodes. Founder cells are going to differentiate and progeny is going to be harvested and sequenced. The different clones can be identified through the barcode sequence and the lineage trajectory can be described (Fig34a). This method has been extensively used for example in hematopoietic cell lineage.

b) Polylox mouse model: based on the fluorescent label of founder cells through a system Cre-lox which can be activated at any moment during the process of interest for labelling them, this method has the limitation of the colour used to follow the lineage. Now, researchers have improved this technique instead of using fluorescent probes they use DNA barcodes (Fig. 34b).

c) CRISPR-Cas9 Genome Editing-Based Lineage Tracing: this method consists in the use of the technique CRISPR-Cas9 direct genome editing. The idea of CRISPR/Cas9 lineage tracing is to use Cas9 to create deletions or insertions in transgenic target sites in the genome, this modifications are random and they can be used as barcodes that are going to be inherited by the progeny of the cells giving the necessary tool to follow the cell lineage in the developing tissue (Spanjaard and Junker, 2017) (Fig. 34c). This approach has been used in very different ways. For example, McKenna (Raj et al., 2018)'s group has generated a new transgenic zebrafish line with synthetic concatemerized target sites in the 3' UTR of a GFP transgene, approach that the authors called GESTALT (Genome Editing of Synthetic Target Arrays for Lineage Tracing) (Fig. 33). Junker's group developed LINNAEUS (Fig. 33). They have used Cas9 targeting an RFP transgene that was existing in a zebrafish line, like this they were sure of not interfere with the normal development of the fish (Spanjaard and Junker, 2017; Kester and van Oudenaarden, 2018).



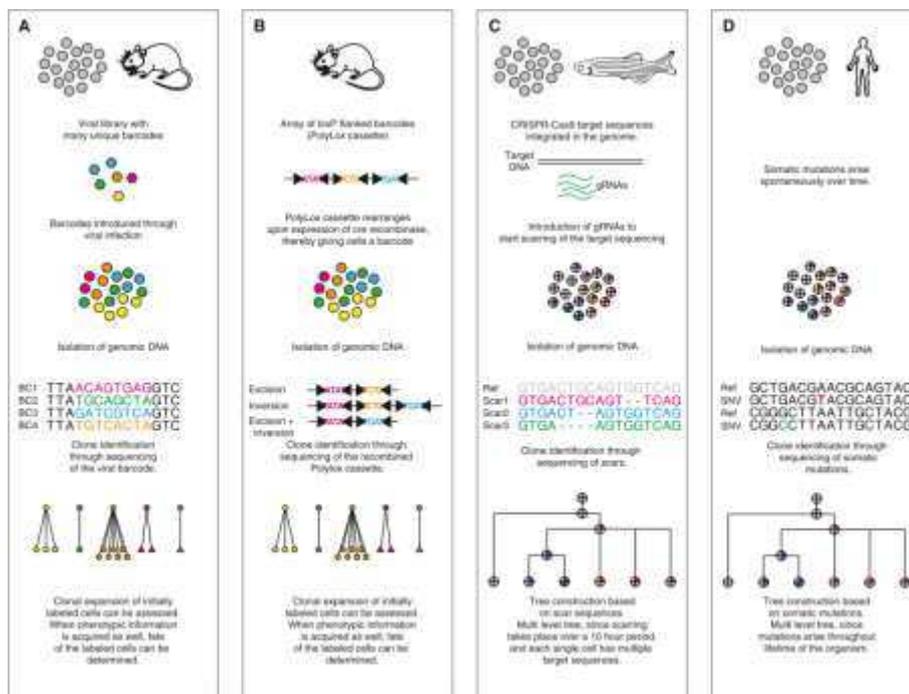
**Fig. 33. Massively parallel lineage tracing using the CRISPR/Cas9 system.** (a) In scartrace/LINNAEUS, an existing fish line with multiple integrations of a transgene is targeted by Cas9. The sequences of the resulting ‘genetic scars’ (light gray) are used as lineage markers. (b) GESTALT uses the same principle, but a new line with concatemerized target sites is used. Compared to scartrace/LINNAEUS, this has the advantage that the individual sites (different colors) can be distinguished. From (Spanjaard and Junker, 2017)

The problem with all these techniques is the necessity to introduce exogenous material, which in some cases is not easy due to the nature of the sample. That is why there was the necessity of implementing other techniques termed retrospective lineage tracing.

## VII.2 Retrospective lineage-tracing

Consist in the lineage tracing of natural occurring mutations such as somatic mutations or Copy Number Variations (CNV).

a) Tracking of somatic mutations: this technique arises due to the limitations of the above-described tracing methods, for example the limitation of lineage tracing in humans. The aim of this technique is the study of the spontaneous mutations occurring naturally in the cells, these mutations are inherited by the cell progeny and can be analysed (Kester and van Oudenaarden, 2018)(Fig. 34d).



**Fig. 34. Overview of Genetic Lineage Tracing Strategies.** (A) Lineage tracing through viral barcoding. Cells are infected with a virus library containing many different barcodes. After a period of time clones can be identified through the barcode sequence. (B) The Polylox mouse model was created by the introduction of a set of barcodes interspersed with loxP sites. Upon activation of the Cre recombinase, the Polylox cassette recombines, thereby producing unique cellular barcodes via a combination of losses and inversions of single barcodes. After a period of time the DNA is isolated and clone identification is done through the assessment of the combination of losses and inversions of barcodes. (C) Lineage tracing using CRISPR-Cas9 can be done in cultured cells and zebrafish. Introduction of CRISPR-Cas9 and gRNA into the cells results in the scarring of the target sequence during a given time window. After a period of time cells are harvested, genomic DNA is isolated, and scars are sequenced. The combination of scars in each cell produces a unique barcode, and the construction of multi-level phylogenetic trees is possible since scarring takes place over a long period of time and a portion of the scars will be shared between clones. (D)

Tracking of somatic mutations can be performed in any model organism. Somatic mutations arise spontaneously and accumulate over time, thereby marking single cells and all their progeny. Clones can be identified through whole-genome or targeted sequencing. Construction of multi-level phylogenetic trees is possible since mutations arise over a long period of time. From (Kester and van Oudenaarden, 2018)

b) Tracking CNV: the CNVs can easily be detected by sequencing. The problem is that in healthy tissues, they are not really common so the construction of phylogenetic trees in samples from healthy patients are not really possible. However, CNVs are abundant in cancer and change during tumour developing what makes them suitable for the lineage reconstruction in cancer. In the work of Navin and Wang (Navin et al., 2011; Wang et al., 2014) studying breast cancer, they found enough CNVs to generate the phylogenetic tree.

c) Tracking Single-Nucleotide Variants (SNV), indels and repeated regions: all exist in non-coding regions of the genome, but for instance, for their detection is only possible performing whole genome sequencing, making difficult the detection of these isolated events, and also the amplification step for the whole sequencing of the genome hamper the detection of the SNV. Beside SNV also retrotransposon elements and microsatellite repeats have been used for lineage tracing.

d) Tracing through the detection of Epigenetic marks: using DNA methylation or hydroxymethylation. Methylation of DNA is a regulator of gene expression and is maintained during cell division, the methylation of DNA can change in time depending of the cell requirements but, for instance, this changes are occurring in a very slow fashion, meaning that are a good source to detect cell lineage in a developing tissue.

The fast and vast development in single cell transcriptomics has revolutionized the scientific community. With this new tool we can go deeper in the understanding of cellular and molecular events regulating all the biological processes occurring in the different systems, and now, more than ever, we can describe and characterize dysfunctional processes that lead to disease.

The scientific community now work together in a challenging project: The Human Cell Atlas (HCA). The aim of this project is to define all human cell types in terms of their distinctive patterns of gene expression, physiological states, developmental trajectories, and location and the creation of a “comprehensive reference maps of all human cells as a basis for both understanding human health and diagnosing, monitoring, and treating disease” (Hon et al., 2017; Paper, 2017).

This project will help to the community being a reference, putting together the different discoveries and applications creating a new tool to fight against the different diseases tethering the human life, helping in the development of different treatments and in the evolution of a new era of the regenerative medicine.

## VIII. *Single cell studies in airway epithelium*

---

With the huge advances in the single cell transcriptome analysis and the different techniques that allow the identification of new populations in heterogeneous systems such as the airway epithelium, and the development of techniques for the inferring of cell lineage hierarchies, this technique has become really demanded.

Here I'm going to summarize the main experiments performed in the airway epithelium using scRNAseq.

In 2014 the group of Mark A. Krasnow and Stephen R. Quake performed a pioneer single-cell study in distal lung, they analysed a total of 198 single cells. They obtained single cell suspension from micro-dissected portions of distal lung of mice and they isolated the cells using a microfluidic device (C1 system from the company Fluidigm).

In a first analysis they performed scRNAseq over 80 cells and performing Principal component analysis (PCA), the most enriched genes were then analysed by unsupervised hierarchical clustering as well as PCA. Looking at the expression of different known markers they recognized four populations, SCs, MCCs, alveolar type-1 (AT1) and alveolar type-2 (AT2), a fifth population was characterized by the expression of markers of the AT1 and AT2 markers together revealing or a population transitioning from one to the other or a precursor population of both AT1 and AT2.

To reveal the cell trajectories of the differentiation of the cell populations in the distal lung they performed a second single cell analysis, this time performing the analysis over single cell suspensions collected in three different time points during the differentiation, early, intermediate and very late time point. Thanks to this analysis they performed lineage hierarchies of the alveolar cells in the distal lung, they confirmed the existence of a bipotential progenitor that triggered the differentiation of two branches in the lineage to AT1 or AT2 cells.

With their pioneer analysis they decipher these trajectories and also new markers for the different cell populations. Notably studying the different effectors that could have a role in the differentiation, they noticed that the differentiation from the early progenitors to the bipotential progenitors it was not caused by the increased expression of any gene but, in fact, they detected a global downregulation of markers such as the transcription factor *Sox1* (Treutlein *et al.*, 2014).

In 2016 the group of Whitsett performed scRNAseq analysis over single cell suspensions from distal lung dissections comparing healthy lung (3 patients, 215 cells) and IPF lungs (6 patients, 325 cells), they did a first purification of the cell suspensions by FACS, sorting all the epithelial cells positive for CD326 marker (EPCAM) marker for epithelial cells, then they isolated the cells using the C1 system from Fluidigm. Analyzing the single cell transcriptome from healthy and IPF patients they detected different cell clusters,

notably in the healthy samples all the cells belonging to the cluster of AT2 cells were almost not present in IPF patients.

The analysis of the expression pattern of different markers associated with chloride transporters such as CLCN2/4/5, SLC6A14, SLC26A4 and CFTR were highly decreased in IPF samples whereas sodium and bicarbonate transporters were increased. They also detected 26 different signaling pathways altered, TGF- $\beta$  & PI3K/AKT among others. The gene signature of the cell populations founded in the IPF patients showed a clear loss of cell identity, markers normally expressed in AT1 co-expressed with AT2 markers, markers such as SOX2 and SOX9 (usually expressed in very different compartments, conducting and peripheral bud of embryonic lung respectively) were found together in some single cells; all these results together showed that in fact in IPF patients a process of remodeling was taking place. In conclusion, thanks to the technique, they could establish the gene expression diversity of transitional states of IPF cells, providing new information to the biological process of the disease (Xu *et al.*, 2016).

The group of Nora A. Barrett and Alex K. Shalek performed seq-Well massively-parallel scRNAseq, using UMI and barcoded beads in a microarray over 12 samples of cell suspension dissociated from resected sinus tissue, 6 from healthy patients and 6 from patients with polyps with different grades of chronic rhinosinusitis (CRS)(18624 cells in total were analyzed), aiming to reveal the molecular signature of the TH2-type inflammation related to this disease.

They analysed the data performing dimensionality reduction and graph-based algorithms for the cell clustering. They detected clusters of basal and apical cells, MCCs and glandular cells, endothelial cells, fibroblasts, plasma cells, myeloid cells, T cells and mast cells. A closer look into each cell population showed certain heterogeneity. They describe the basal cells as responsible of the secretion of the pro-cytokines IL25, IL-33 and TSLP; they showed that that BCs, SCq, and glandular cells had the most significant links to the disease state.

They performed pseudotime mapping to aligning and reconstructing how BCs differentiate to mature SCs, they found in the inflammatory state that the BCs remained as proliferative by the expression of several transcription factors such as ATF3, AP-1, p63 and KLF5, impairing their differentiation. Analysing the different pathways involved in the disease they showed the activation of the WNT pathway through the high expression of CTNNB1 ( $\beta$ -catenin) which is a key effector of the WNT pathway and expression of CTGF (specific factor of the pathway) (Ordovas-Montanes *et al.*, 2018).Part of their results have been introduced along the introduction of the manuscript.

A very recent published work from the group of Rajagopal, studied the mouse tracheal epithelial cells, previously sorted by FACS, using a combinational method with scRNAseq and genetic cell lineage tracing, they used two different approaches for scRNAseq, first a massively parallel droplet-based 3' scRNAseq (10X genomics) where they analysed 7193 cells, and second full-length scRNAseq using Smart-Seq2 where they

analysed 301 cells. They found the known clusters of epithelial cells BCs, SC/GCs and MCCs and rare cell types such as tuft, ionocytes and PNECs.

They describe new markers for the different cell populations such *Nfia*, transcription factor expressed in SCs and regulates Notch signalling required for SCs maintenance, surprisingly in their data the MCCs show also expression of this marker. Or *Foxq1* expressed specifically in GCs (Montoro *et al.*, 2018).

Thanks to the scRNAseq technology the detection of rare cell types is possible. In this work Rajagopal and colleagues revealed the different markers defining the rare population of Tuft cells such as *Pou2f3*, *Gfi1b*, *Spib* and *Sox9* as transcription factors detected. Moreover, they could detect heterogeneity inside the tuft cell population and they defined three different groups of Tuft cells related with their different maturation points. They characterized also the ionocytes population expressing the V-ATPase subunits *Atp6v1c2* and *Atp6v0d2*, new gene markers for this population are, among others, *Ascl3*, *Foxi1*, *Foxi2* and *Pparg*, and they also express *Cftr*.

After the characterization of the ionocytes population founded through scRNAseq analysis of MTECs, did some functional analysis.

They found that ionocytes express high levels of *Cftr*. This cell type was detected in the submucosal glands and in nasal and olfactory epithelium. They used a transgenic mouse model KO for *Foxi1* and they showed that *Foxi1* it was essential for the expression of the ionocytes transcription factor *Asl3* and also for the expression of the Majority of the *Cftr*. They use also this mouse model defective for *Foxi1* to test the mucus viscosity, the airway surface liquid (ASL) height and the ciliary beating frequency (CBF) and they found similarities with the CF disease such as the viscosity of the mucus and also the CBF increased. For instance, the height of the epithelium or the pH were not altered (Montoro *et al.*, 2018). More of their results have been introduced along the manuscript.

A second very recent published work by the group of Jaffe perform scRNAseq analysis in MTECs (7662 cells) and in Human bronchial epithelial cells (HBECs) differentiated in vitro in a 3D system of air liquid interface (ALI) (2970 cells), they used a graph based algorithm, SPRING, to analyse the data. They identified the known cell clusters in both models; BCs, SCs, MCCs, tuft, PNECs, and as the results from Montoro *et al.*(Montoro *et al.*, 2018) they show also heterogeneity in the BC and SC populations. They found the expression of *FOXN4* as a marker of immature MCCs, what we called deuterosomal cells as I will discuss in the result part of the manuscript, this transcript is associated with the transcription of ciliated genes in *Xenopus*(Campbell, Quigley and Kintner, 2016).

They described also a cluster of Ionocytes expressing the transcription factors *Foxi1*, *Ascl3* and *Tfcp2l1*, that was also highly enriched in *Cftr*. They showed that the overexpression of *FOXII* in HBECs was sufficient to induce a larger number of cells clustered as ionocytes. In *Xenopus* the Ionocytes differentiation is modulated by notch signalling, and the authors detected some Notch target genes expressed in the ionocytes cell population, so they tested in HBECs the effect of inhibition of the Notch pathway using DAPT, they

corroborated the effect of DAPT over the MCCs population triggering a decrease in the number of MCCs, and also they detected a decrease of the number of ionocytes, contrary to what was seen in *Xenopus*.

They analysed the regeneration of MTECs after injury through scRNAseq, one day after injury they detected a basal cycling population (*Krt5+*) co-expressing other keratins such as *Krt14*, *Krt8* and *Krt13/Krt4*, that were never co-localizing in homeostasis. At two and three days after injury they detect the direct differentiation of MCCs from BCs, by-passing the SCs population. Seven days after injury they detected total recovery of all the cell populations detected in homeostasis (Plasschaert *et al.*, 2018).

The group of Ido Amit performed scRNAseq in distal lung characterizing immune and non-immune cells. They analyzed 50770 single cells recovered during different stages of lung development in mouse during embryogenesis and after birth. They found rare cell types that appeared in different moments during development such as a novel lung alveolar basophil that exist in the lung at early stage during embryonic development. The analysis of all the ligand-receptor interactions among all the cells revealed the cellular network in the lung during development (Cohen *et al.*, 2018).

## *IX. RESULTS*

---

## IX.1 Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration

### CONTEXT OF THE STUDY

As extensively described in the different introductory chapters, the epithelium of the airways serves as a protecting barrier against damage and infections, through the mucociliary clearance. Correct function of mucociliary clearance relies on the proper balance of cell populations composing the epithelium and the interactions between them. After aggressions such as inhaled pollutants and infections, part of the epithelium can be damaged. In healthy subjects, the epithelium is able to regenerate until complete restoration and recovery.

The events of chronic injury and inflammation triggered in the epithelium of patients with chronic lung diseases such as COPD, CF or asthma lead to remodelling, goblet cell hyperplasia and loss of MCCs, leading to the pathological situations such as mucus clogging, blocking of the airways and recurrent infections with opportunistic pathogens.

There is emergency in understanding the molecular and cellular events leading to goblet cell hyperplasia and remodelling of the epithelium. For that aim, we must first understand mechanisms occurring during the normal regeneration of healthy airways.

Various studies have tried to decipher the regeneration process. However, they have mostly been performed in mouse models which allow *in vivo* genetic cell lineages, and the study of airway regeneration after injuries. These studies have resulted in a relatively complete scheme of the cell trajectories during regeneration of the epithelium and of the regulatory molecular pathways, as described during the III chapter of the introduction. However, they include several inherent disadvantages. For instance, difference in epithelial cell composition between human and mouse species do not allow simple transposition of findings obtained in mouse to human.

Moreover, to perform mouse cell lineage studies, a prerequisite is to use one cell type marker as a tracer gene. This introduces an “a priori” information constraints the field of the study to cells expressing this given gene and does not consider the whole gene expression profile that makes a cell identity.

As exposed in the V introductory chapter, single-cell transcriptomics has greatly contributed to counteract this problem, by allowing, without previous knowledge of the system: (1) the study of complex systems composed heterogeneous cell populations, (2) lineage hierarchy inference and (3) identification of active molecular pathways

The aim of this project was to analyse the differentiation process in upper human airway epithelium using single cell transcriptomics in order to identify the distinct cell populations emerging during differentiation, infer their lineage relationships and determine regulatory molecular mechanisms.

We have used cultured human epithelial cells from resected turbinate. The culture system consists in dissociating basal cells from the turbinates and seeding them on culture inserts (Transwells) allowing air-liquid interface culture. Typically, basal cells are let to proliferate on the transwells for 4 to 6 days until confluency. Then the apical side of the cell layer is put in direct contact with the air, the basal side remaining soak in inductive defined medium. After about 20-21 days, a 3D airway epithelium has been regenerated.

We have analysed several time points during the generation of this epithelium. Time points were chosen as most representative of the known steps of airway regeneration: proliferation, polarization and specification. We have performed this experiment using two culture media which are commercially available and widely used by the community. Before performing single-cell experiments, we have validated proper *in vitro* regeneration with the use of two different media, but have also noticed that they induce variations in cell composition. This point will be further discussed in the result section. To complete our findings on human cells, we have performed single-cell transcriptome analysis on a variety of samples, from distinct origins (fresh tissue or *in vitro*) and from distinct organisms: cells dissociated from human nasal brushing, nasal turbinates or bronchial biopsies, pig trachea, as well as differentiating Mouse Tracheal Epithelial Cells (MTECs). We have applied computational methods such as differential expression-based cell clustering, cell lineage inference and RNA velocity to obtain the results which are detailed in the following publication: “Single-cell RNA-seq reveals novel cell differentiation dynamics during human airway epithelium regeneration”.

This work is presented in the submitted publication (Ruiz Garcia *et al.*, 2018) (ANNEX)

## RESULTS AND DISCUSSION

### *Cell trajectory reconstruction during airway regeneration*

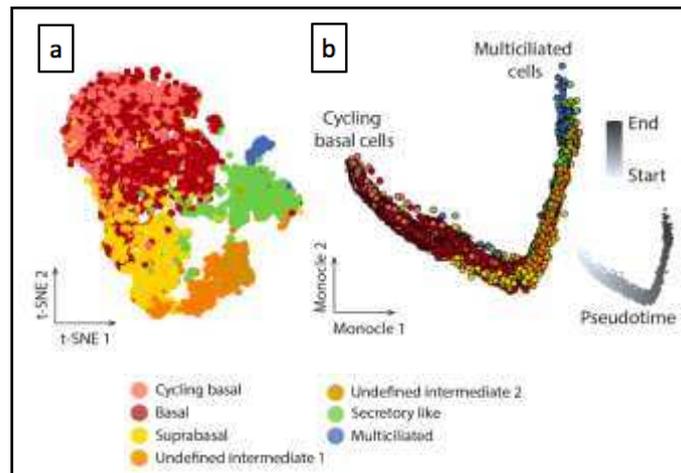
Our analysis of the transcriptome of several datasets at the single cell level led to the identification of the distinct known populations composing the airway epithelium through the detection of specific markers such as *TP63* and *KRT5* for basal cells (BCs), *SCGB1A1* for secretory cells (SCs), *MUC5AC* for goblet cells (GCs) and *FOXJ1* for multiciliated cells (MCCs). If proportions of all these cell types varied through the various samples we have analysed, we could detect all of the main airway epithelial cell types were detected. We have validated these proportions of the by performing cell identification at the protein level, which validated our approach.

Our first objective was to determine the distinct cell populations and molecular regulators emerging throughout the time course of in vitro differentiation, by performing scRNA-seq at several time points. We used two different media to induce differentiation, BEGM (Lonza) and PneumaCult (StemCell Technologies) and both resulted in the establishment of similar cell populations. We could distinguish 6 main populations: cycling and non-cycling BCs, a cluster of suprabasal cells (supraBCs) expressing *KRT5*, *KRT13* and *KRT4* but not *TP63*, SCs, GCs and MCCs. The proportion of these cell populations evolved during all the time course where in a starting point most of the cells were BCs, with an evolution towards the emergence of supraBCs, SCs, GCs then finally MCCs. Surprisingly, in the BEGM media condition, we were not able to detect any of the usual SC markers secretoglobins (*SCGB1A1*, *SCGB3A1*, *SCGB3A2*) nor the usual GC marker *MUC5AC*. Instead we found a secretory-like population with a gene expression profile which was very similar to that of SCs, at the exception of secretoglobins. We also detected two intermediate populations that we could not classify and we named them “undefined intermediates 1” and “2”.

Recently, new methods and protocols to perform large scale cell lineages tracing have arisen. Their principle is to combine single cell transcriptomics with cell lineage labelling, allowing a strict tracking of cell trajectories in complex samples and bypassing the dependency on only one gene marker, which is the flaw of classical lineage tracing studies. These approaches also allow to perform lineage tracing in vitro, which is very useful in organoid experiments such as ours. At the date of our experiment, these techniques were not available, and technical optimization is now necessary to apply it to our 3D model of differentiation. The only robust approach that was available was lineage inference through computational analysis of the single-cell transcriptomes. Here we used the algorithm of Monocle 2 (chapter VI.1) that allows the ordering of cells from the same sample into a “pseudotime” and the detection of branches composed of cells taking divergent trajectories during airway epithelium differentiation. Analysis of the final time point of differentiation was sufficient to reconstruct cell trajectories, as the fully regenerated tissue retains all transitory cells. However, to obtain the most extensive picture of cell intermediates as possible, we have performed the pseudotime analysis after aggregating all time points into one dataset.

### ***Cell trajectories in BEGM media***

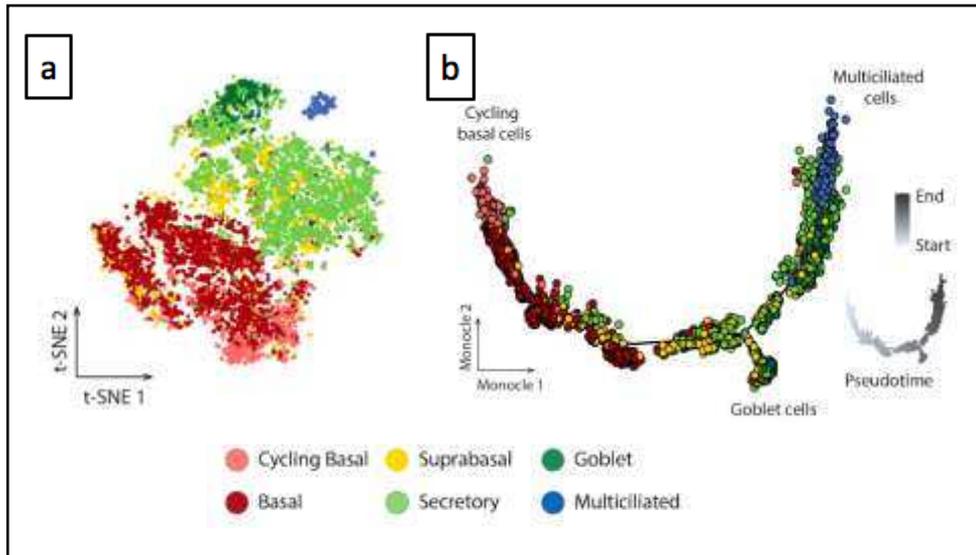
Lineage inference with Monocle 2 allowed us to establish the following cell lineage: cycling BCs (*MKI76+/KRT5+/TP63+*) give rise to non-cycling BCs (*MKI67-/KRT5+/TP63+*), which give rise to undefined intermediate 1, then undefined intermediate 2, from which by the secretory-like population emerges (*SCGB1A1-/BPIFA1+/KRT8+*) which in turn, gives rise to the MCC population (*FOXJ1+*). Thus, we have detected only one continuous branch of cell differentiation (Fig.35).



**Fig.35. Characterization of MCC cell lineages during airway epithelium regeneration using single cell RNA-Seq in BEGM medium. (a)** t-SNE plot of the aggregate of all cells from each time point. **(b)** Representation of the cell lineages inferred by Monocle 2 occurring during the upper airway epithelium regeneration (aggregate of all time points). Pseudotime evolution along the differentiation trajectory shown by white to grey gradient.

### ***Cell trajectories in PneumaCult media***

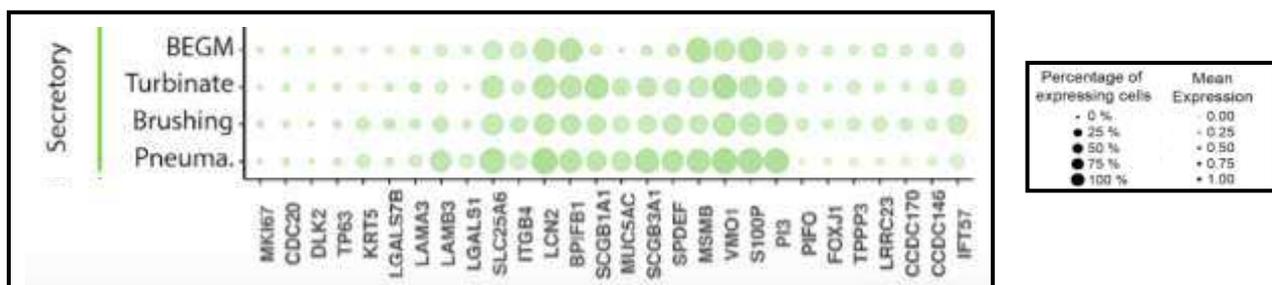
In PneumaCult media condition, we could detect all cell populations, including a canonical SC population expressing *SCBG1A1+* and a *MUC5AC+* GC population. Undefined intermediate populations were not detected. Monocle yielded a cell hierarchy starting with cycling BCs, going through BCs, supraBCs and SCs. From that point, two diverging trajectories were detected, with a short branch leading to GC, and a longer branch leading to MCCs, confirming that MCCs and GCs both come from a SCs as it was described before in the literature (Fig.36).



**Fig.36. Characterization of MCC and Goblet cell lineages during airway epithelium regeneration in vitro using single cell RNA-Seq in PneumaCult.** (a) Aggregate t-SNE plot of gene expression in 9826 cells. (b) Inference of goblet and MCC cell lineages by Monocle 2, based on an aggregate of the entire experiment. Color code is the same as in (a). Inset: pseudotime picturing by a white to grey gradient along the differentiation trajectory.

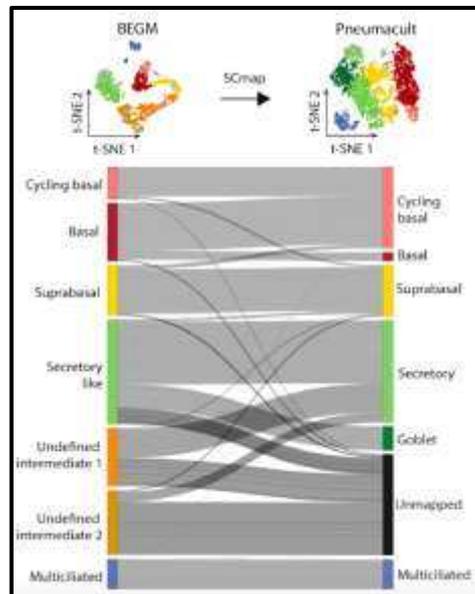
***Secretory-like cells, not canonical, but still functional differentiation intermediates***

Hence, in BEGM, no canonical SC population was detected. However, we have detected secretory-like cells which seemed unable to generate GCs but were still able to give rise to MCCs. It would be worth identifying factors which are responsible for this inconsistency. Differential expression analysis between secretory and secretory-like cells show that secretory-like cells do not express the genes *SCGB1A1*, *SCGB3A1*, *MUC5AC* and *SPDEF* (Fig.3). It is known that *SPDEF* is related to GC differentiation, and the downregulation of this gene correlates with the fact that in HACEs cultured with BEGM media we don't detect the expression of the goblet cell marker *MUC5AC*.



**Fig.37. Cell type composition comparison between homeostatic in vitro samples and fresh human tissues.** Dot plot of the main cell population marker genes. Dot size describe the percentage of cells expressing the respective marker genes and the average expression level of that gene based on UMI counts are shown by color intensity.

If we look BEGM cells onto Pneumacult cells with scmap, which finds similarities between cell population of different samples, we can see that the secretory-like population maps onto the secretory and goblet populations of PneumaCult (Fig.38).

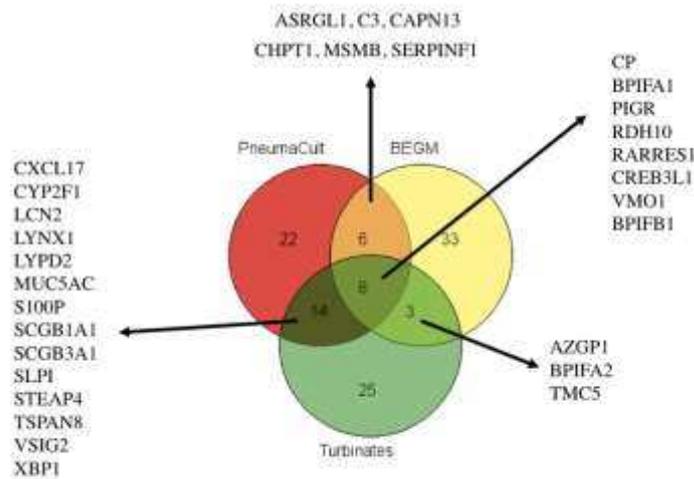


**Fig.38. Comparison of fully differentiated epithelia cell population between PneumaCult and BEGM media.** Sankey Network of the mapping of BEGM cells onto PneumaCult cells.

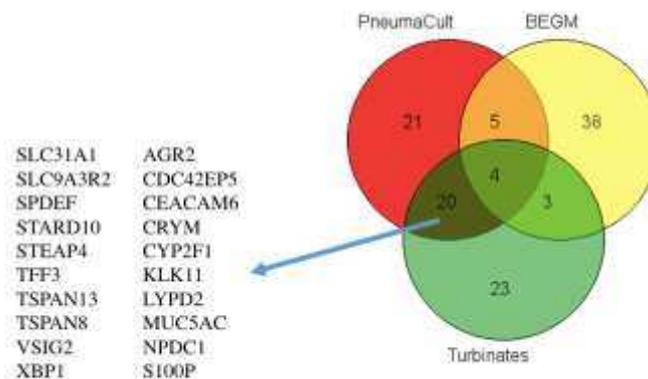
The secretory-like population has lost its potency to differentiate in goblet cells, but it remains functional playing a role in the immune defense of the epithelium as we detect by the expression of the immunoglobuline binding factor MSMB, or the genes coding for antimicrobial proteins such as *BPIFA1*, *BPIFA2* and *BPIFB1*. This fact lights up the inconvenience of having only one specific marker for the detection of a certain cell population, most the studies focused in the analysis of the SCs use the expression of SCGB1A1 for the detection of this cell population. Now we found new evidences of functional SCs SCGB1A1<sup>-</sup>. As was described in the work from Reynolds's lab, the non-homogeneous distribution of the SCGB1A1<sup>+</sup> population along the airway ducts in humans led to hypothesize about the possibility of having others SC populations that compensated the role of the lacking SCGB1A1<sup>+</sup> cells in the airways.

Analysis of the differences between the secretory-like and the secretory cells in the different human system in vitro and in vivo can decipher new actors in the differentiation of GCs. This analysis shows the expression of 14 common genes between the secretory cells from turbinates and from cultured HAECs in PneumaCult media that are not expressed in the secretory-like population in cultured HAECs in BEGM media (Fig.39). These genes could represent goblet cell markers expressed in the precursors of the GCs or in the mature GCs.

A closer analysis of the cells expressing the goblet cell marker *MUC5AC* in PneumaCult and in turbinates compared with the secretory-like cell population in BEGM show 20 common genes specifically expressed in the GCs of turbinates and PneumaCult cultured HAECs (Fig. 40). Future functional analysis of overexpression could reveal the involvement of these genes in the differentiation of the GBs.

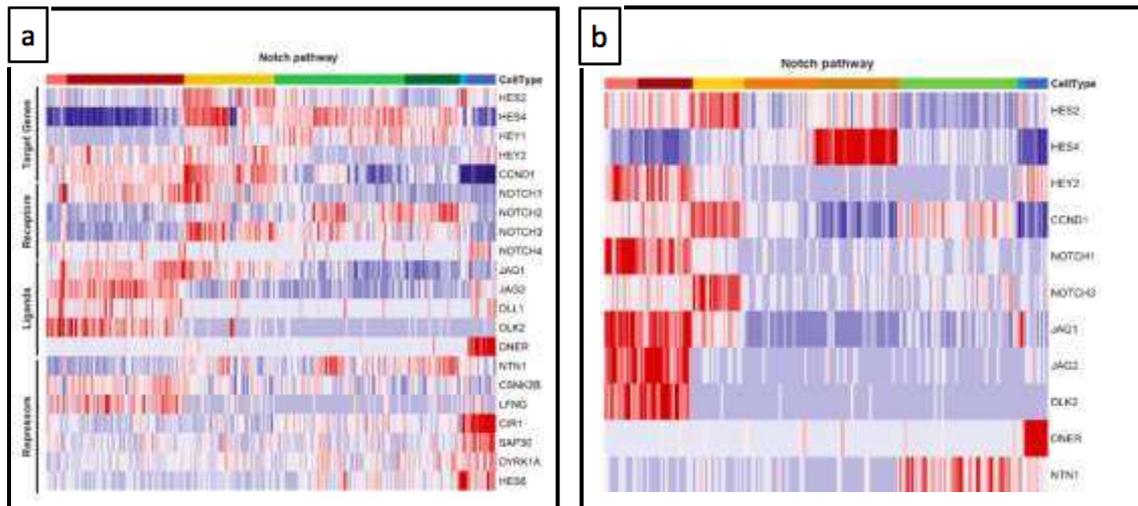


**Fig.39. Different gene expression between secretory and secretory-like cells.** Venn diagram showing the overlaps and differences existing between top gene markers of secretory cells (*BPIFA1+*) from Turbinates (green) secretory cells from in vitro HACEs cultured cells in PneumaCult media (red) and secretory-like cells (*BPIFA1+*) in HAECs cultured in BEGM media (yellow).



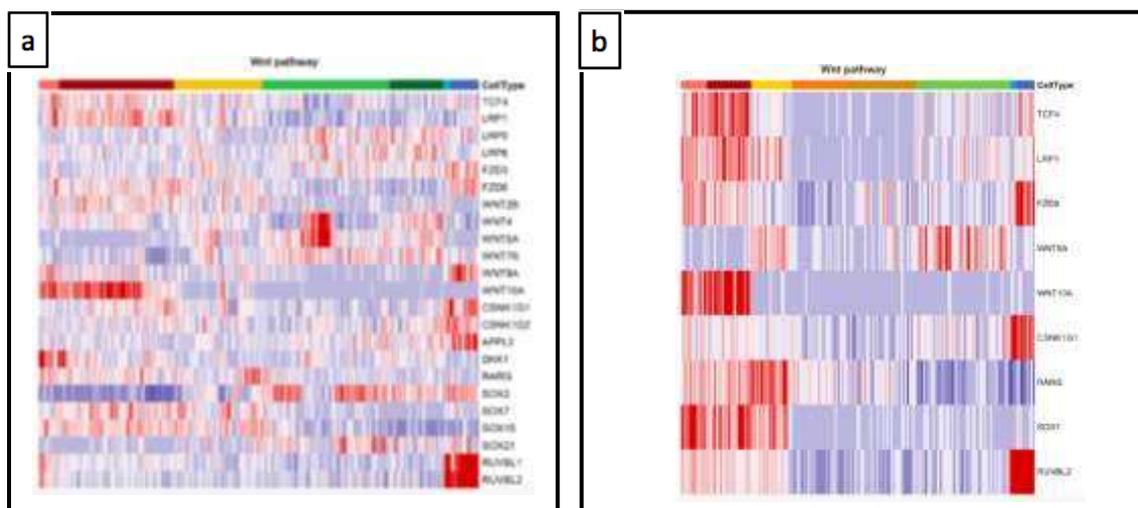
**Fig.40. Different gene expression between GCs and secretory-like cells.** Venn diagram showing the overlaps and differences existing between top gene markers of goblet cells (*MUC5AC+*) from Turbinates (green) secretory cells from in vitro HACEs cultured cells in PneumaCult media (red) and secretory-like cells (*BPIFA1+*) in HAECs cultured in BEGM media (yellow).

Regarding the pathways controlling the differentiation in both medias:



**Fig.41. Differential expression of the genes related to the NOTCH pathway.** (a) Heatmap of the genes related to the Notch pathway] with cells cultured in PneumaCult media ordered by cluster.(b) Heatmap of the genes related to the Notch pathway] with cells cultured in BEGM media ordered by cluster.

Regarding the pattern of expression of the Notch signalling components comparing both culture medias, we don't detect *NOTCH2* in the BEGM media whereas is highly expressed in the GC population in PneumaCult media. As was shown by Pardo-Saganta, Notch2 is necessary to maintain the secretory cell state and avoids the differentiation into MCCs, and Danahay showed that using antibodies blocking Notch2 receptor in MTECs inhibited all the GC markers (*MUC5AC*, *MUC5B* and *FOXA3*)(Danahay *et al.*, 2015; Pardo-Saganta, Tata, *et al.*, 2015). The non-expression of *NOTCH2* in the BEGM system is coherent with a loss in GCs (Fig.41)

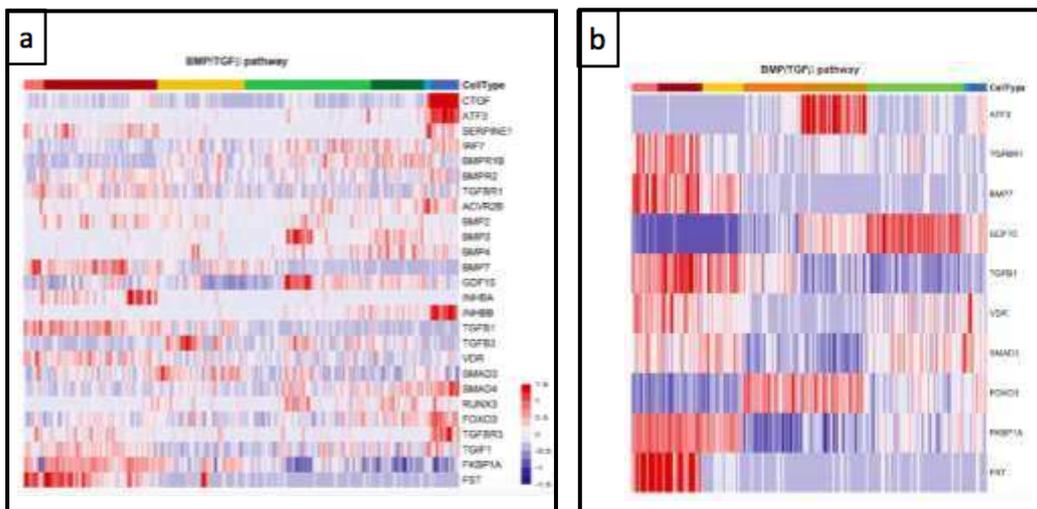


**Fig.42. Differential expression of the genes related to the WNT pathway.** (a) Heatmap of the genes related to the WNT pathway with cells cultured in PneumaCult media ordered by cluster.(b) Heatmap of the

genes related to the WNT pathway with cells cultured in BEGM media ordered by cluster.

For the WNT pathway we detect differential expression for some markers of the pathway comparing both medias, for instance, in the BEGM media we don't detect the expression of *WNT4* that is highly expressed in the SC population in the Pneuma media, meaning that the secretory-like in BEGM lacks the expression of *WNT4*, meaning that maybe *WNT4* is important for the goblet cell differentiation. *WNT4* has been shown to be upregulated in the epithelium of COPD patients and upregulates *IL8* and *CXCL8* gene expression in HAECs(Heijink *et al.*, 2012; Durham *et al.*, 2013), meaning that maybe the SCs have a role in immune protection that the Secretory-like cells don't.

The same for the transcriptional repressors *SOX21* and *SOX2* that are expressed in a subset of SCs in PneumaCult media and is not detected in BEGM, probably meaning that the wnt pathway remains active in the Secretory-like cells in BEGM media which impairs the maturation of SCs into GCs (Fig. 42).



**Fig.43. Differential expression of the genes related to the BMP/TGF-β pathway.** (a) Heatmap of the genes related to the BMP/TGF-β pathway with cells cultured in PneumaCult media ordered by cluster. (b) Heatmap of the genes related to the BMP/TGF-β pathway with cells cultured in BEGM media ordered by cluster.

Comparing the expression of the different components of the BMP/TGF-β pathway in HAECs cultured with PneumaCult or BEGM media we detect a strong expression of the ligand GDF15 for the TGF-β (Fig. 43), the protein coded by GSF15 is a cytokine and its secretion is related with inflammation or response after injury and mucus secretion. Is induced after cigarette smoke exposure and triggers inflammation and epithelial senescence(Wu, Jiang and Chu, 2012; Wu *et al.*, 2016; Verhamme *et al.*, 2017). Increased serum levels of GDF15 have been shown increased in patients with COPD. Several experiments in performed in mice revealed increased levels of GDF15 mRNA and protein after cigarette smoke exposure, thus the hypothesis

that the expression of GDF15 is related to immune response. Mice depleted for GDF15 and exposed to cigarette smoke showed a decrease of the recruitment of immune cells such as neutrophils, dendritic cells and lymphocytes B and T in the Bronchoalveolar lavage, and showed attenuated inflammation, meaning that GDF15 could be a mediator in the induced inflammatory processes (Verhamme *et al.*, 2017). The detection of this transcript in HAECs primary cultures show the potency of the SCs and secretory-like to respond to injuries.

We detect a different expression profile of the Bone morphogenic protein-3 (BMP3), is expressed in SCs of PneumaCult media but not in BEGM (Fig. 43). BMP3 mRNA was detected in developing lung from human and rat embryos, suggesting that BMP3 may have a role in the lung morphogenesis or function (Vukicevic, Helder and Luyten, 1994; Takahashi and Ikeda, 1996). Beside this role in embryogenesis little is known about the role of this protein in the airways. In our system BMP3 could be related to GC differentiation.

Interestingly we shown the different expression pattern of the transcription factor FOXO3 and the TGF- $\beta$  target gene ATF3. The activation of FOXO3 is related with the reduction of the process of inflammation, is downregulated in patients with COPD and in HAECs after cigarette smoke exposure (Di Vincenzo *et al.*, 2018).

FOXO3 has been shown to down regulate NFkB in lung adenocarcinoma cell lines after treatment with nicotine-derived nitrosaminoketone (NNK), tobacco carcinogen compound that triggers lung adenocarcinoma (Blake *et al.*, 2010).

In 2013 the group of Lee and Yang described the role of the NFkB in the regulation of COPD-related cytokines IL-6 and IL-8 (Lee and Yang, 2013; Wu *et al.*, 2017). Mice depleted for *Atf3* (*Atf3*<sup>-/-</sup>) attenuated the expression of IL-6 and IL-8 after cigarette smoke exposure inhibiting the NFkB pathway (Wu *et al.*, 2017). Previous studies also revealed attenuation of NFkB expression by ATF3 (Kwon *et al.*, 2015).

An overall hypothesis applied in our model would consist in the inhibition of the NFkB pathway in the BEGM medium by the increased expression of *ATF3* and *FOXO3* in the indetermined population that in the lineage trajectory is going to give rise to the secretory population, the inhibition of the NF-kB pathway will abrogate the differentiation of the SCs into GCs in the BEGM media. We could test this hypothesis inhibiting or overexpressing the NFkB pathway in the PneumaCult or the BEGM media respectively and check whether we inhibit or induce the GC differentiation.

### ***Goblet cells can be intermediates in the differentiation of the MCCs***

Looking closer at the trajectory data we have obtained in the PneumaCult condition, we could localize some goblet cells inside the MCC branch. So, we made the hypothesis that GCs can be facultative precursors of MCCs. In our model, we could not perform classical lineage tracing to bring direct evidence of this trajectory. However, several of our findings corroborate this hypothesis. First, when we clustered cell populations with the robust clustering method, it was impossible to separate SCs and GCs into 2 distinct clusters. We had to annotate GCs “by hand” if significant amount of *MUC5AC* levels were detected. Both these populations are very similar, both are expressing *SCGB1A1*. The only difference we could make was a stronger expression of several mucins including *MUC5AC* and *MUC5B* by GCs. Thus, GCs may just be “hyperactive” SCs, and the weak differences between GCs and SCs, may not keep GCs from giving rise to MCCs. Second, analysis of GCs which are located in the MCC branch are expressing MCC markers such as *FOXJ1*. This finding was confirmed at the protein levels by identifying cells co-expressing *MUC5AC* and Acetylated Tubulin. We also confirmed these results in fresh samples from human bronchial biopsies and in pig trachea. Third, we used RNA velocity (chapter VI.3) to determinate the fate of the different cells in the system, this analysis showed the trajectory of these GCs going towards the MCC population.

This is the first time that, in healthy conditions, is reported the involvement of the GCs in the generation of MCCs.

Some studies have revealed events of MCC/GC transdifferentiation after injury or inflammation. The groups of Tyner and Brigitte after IL-13 treatment over cultured MTECs or HAECs respectively showed some GCs showed cells sharing properties of goblet and MCCs as an early response of treatment with IL-13 (Tyner *et al.*, 2006; Gomperts *et al.*, 2007).

### ***Deuterosomal cells: a discrete novel cell population***

To analyse the possible heterogeneity inside each of the main cell population that we detected, we searched for intra-cluster heterogeneity. With this additional differential analysis, we could refine the clustering and we were able to detect a total of 12 different clusters from the 6 previously detected. This new vision of the clustering allowed us to identify a cluster of cells which is very closely related to MCCs as it is also expressing *FOXJ1*. This cell cluster specifically expresses *DEUPI*, marker of the deuterosomal state together with *CCNO*, *PLK4*, *CDC20B* and *CEP78*. All precursors of MCCs develop deuterosome structures to massively replicate their centrioles and generate the hundreds of BBs that are going to dock into the cell membrane and elongate axonemes. This cell population clustering, which is apart from the mature MCCs is robust and is detected in all the different data bases analysed in this study. Even in homeostatic adult human airway epithelium we could detect this cluster, meaning that in the normal slow turnover of this epithelium this deuterosomal population is present and necessary to maintain the pool of MCCs in the adult epithelium. Our study provides for the first time an extensive characterization of the gene signature of these “deuterosomal” cells. We could detect novel specific markers, such as *YPELI* (Yippee-like factor). Very little data has been published on this gene’s functions. It has been detected close to the centrosome, but also in the

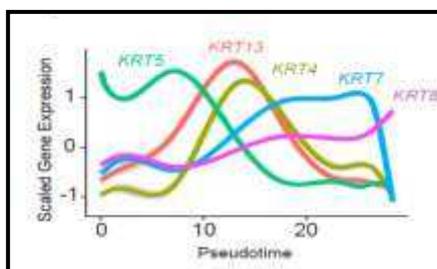
nucleus, where it may play a role in the regulation of cell division (Farlie *et al.*, 2001). YPEL1 is surprisingly highly conserved, even in *Drosophila* (Hosono *et al.*, 2004). Nucleic acid conservation in mouse is almost 100%, which is very uncommon. We have established a collaboration with Bénédicte Durand (Team “Cilia assembly and development” - SFR Santé Lyon-Est) who is using *Drosophila* as a model to study the function of putative ciliary/ basal body genes. They have acquired mutant *Drosophila* lines allowing to study the location of YPEL1 protein and the effect of its downregulation by siRNA.

We and others (Zepp *et al.*, 2017; Revinski *et al.*, 2018) have previously described the expression of cell cycle-related genes related by deuterosomal cells. Thanks to this new clustering apart from MCCs we can now go deeper in the analysis of specific cell cycle genes involved in the early stages of multiciliogenesis. We performed two differential analyses to detect genes specifically expressed in this population: (1) deuterosomal cells vs. cycling cells and (2) deuterosomal cells vs. mature MCCs. We found 152 genes specifically expressed in deuterosomal cells, and among them we found genes coding for centromeric proteins such as *CENPF*, *CENPU* and *CENPW*. We also found a very specific expression of *TOP2A* (topoisomerase 2) which is related to chromosome condensation, separation of chromatids in mitosis, relief of DNA torsional stress during DNA replication or transduction. Several cohesion complex components were also detected in deuterosomal cells (say which one). So, it seems that chromosome-separation mechanisms are active in deuterosomal cells. This is in line with our finding that Separase (*ESPL1*) is required for centriole disassembly at the deuterosome (Revinski *et al.*, 2018). Hence, we may have identified here additional actors taking part in this process.

One of the novel markers of deuterosomal cells is *ANLN*, which codes for Anillin Actin Binding Protein, that plays a role in cell growth and division, migration, cytokinesis and creation of intercellular junctions. The encoded protein is thought to regulate actin cytoskeletal dynamics. One study suggests a novel function for *ANLN* in the organization of the bronchiolar epithelium and an association to disease (Holopainen *et al.*, 2017). In this study, the authors aiming to identify a possible genetic cause of Acute Respiratory Distress Syndrome (ARDS) did a whole genome sequencing in Dalmatian dogs with ARDS and control dogs and they found a mutation in the sequence of the *ANLN* gene triggering a nonsense variant of the protein that completely abrogates the function of the protein. They related the defects showed in the affected dogs to failure in epithelia regeneration due to an impaired cell division program. Other hypothesis is the failure in the creation of intercellular junctions triggering the disorganization of the epithelium. Given the described function in the regulation of actin cytoskeletal dynamics, I would suggest a role for *ANLN* in the migration of the centrioles to the apical membrane and docking, and, unlike the other cell cycle-related genes I have mentioned earlier, not in centriole disengagement from deuterosomes. So, several cell cycle-related molecules are re-expressed during centriole amplification at deuterosomes and may take part in distinct mechanisms of this differentiation stage.

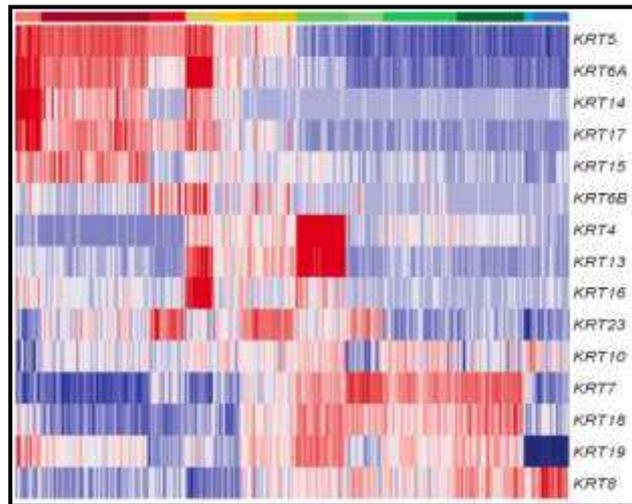
### ***Keratin switch pattern during airway regeneration.***

There are three different kind of proteins that compose the cytoskeleton: microfilaments of actin, microtubules and intermediate filaments (IF), cytokeratines are included in this last group of IF(Kanaji *et al.*, 2007). It is known that specific keratins are expressed during successive stages of epithelial cell development during embryogenesis(Kumar and Jagannathan, 2018). In the airways patterns of keratin expression have been mostly deduced from mouse studies, where KRT5 is clearly a BC marker, just like in other epithelia. It has been described that a subset of KRT5+ expresses KRT14 (James E. Boers, Ambergen and Thunnissen, 1998; Nakajima *et al.*, 1998; Evans *et al.*, 2001; Kyung U Hong *et al.*, 2004). The expression of KRT8 has been clearly associated with luminal differentiated cells. Some keratins have been related to diseases, for example, increase of KRT5 and KRT14 was detected in basal cells of patients with IPF(Ficial *et al.*, 2014; N F Smirnova *et al.*, 2016; Confalonieri *et al.*, 2017). So, the description of the pattern of expression of some cytokeratines in airways and their specificity in some cell populations may have an important role in the study of diseases and new treatments. Our scRNA-seq allowed us to detect very specific expression of many keratins in the different cell types, and we reported their expression pattern during airway regeneration process (Fig. 44 & 45).



**Fig.44. Plot of normalized gene expression of keratins according to pseudotime from scRNA-seq of cells differentiated in Pneumacult medium (ALI28).**

We could confirm expression of *KRT5* and *KRT14* in BCs, with variations depending on the proliferation status: the cycling BC population expresses more *KRT14* than the non-cycling population. We have also identified keratins that are specific to supraBCs, *KRT13* and *KRT4*, with *KRT13* being upregulated earlier in differentiation than *KRT4*. Very recently a study of in vitro regeneration of airway epithelium in mouse, described the presence of a *KRT4+ /KRT13+* population that could also correspond to supra-basal cells. During mouse regeneration of airway epithelium in vivo, the authors describe the presence of this population in “blocks” what they called “hillocks”, they suggested that this population can act as an alternative way from BCs to SCs (Montoro *et al.*, 2018).



**Fig.45. Keratins gene expression in the pseudotime.** Heatmap for scRNA-seq data from Pneumacult ALI28 showing gene expression for keratins.

We have also identified novel keratins associated with the SC population: *KRT7*, *18* and *19*. The expression of these keratins seems to follow that of the luminal keratin *KRT8*, but their expression dramatically and abruptly drops when cells reach the deuterosomal stage (Fig. 45). Few functional studies have been performed to investigate the specific roles of these keratins. Yin and colleagues showed a significant downregulation of *KRT18* in treatment-resistant prostate cancer cell and they directly associated low levels of *KRT18* to tumor aggressiveness (YIN *et al.*, 2016). Further investigations could reveal if these cytokeratines play a role in the determination of cell fate, if they are target genes for known regulators of cell fate, or/and only play a role in the stability and good functioning of SCs/GCs.

#### ***Signalling pathways during airway epithelium regeneration.***

In order to find new insights in the molecular regulation of the airway epithelium regeneration during the pseudotime, we have focused the analyses of our datasets on the identification of specific expression of signalling pathway components.

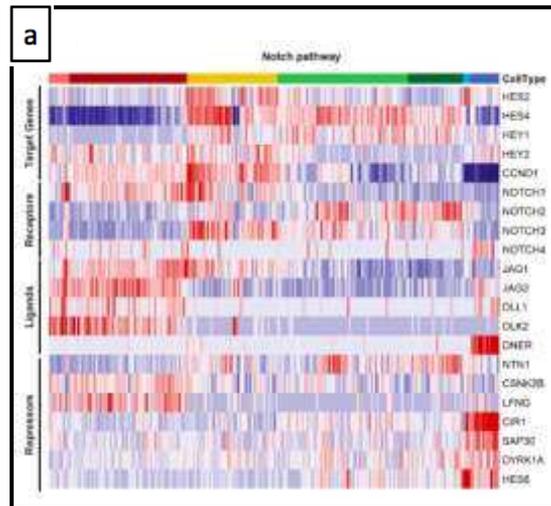
#### ***Notch Pathway.***

Notch activation can regulate proliferation or differentiation in a cell type-dependant manner. Analyses of the components of the Notch pathway confirmed the expected pattern of regulation during differentiation. In accordance with the work performed by Rock and colleagues (Rock *et al.*, 2011) we detected *NOTCH1*, *DLL1*, *JAG1* and *JAG2* but we were not able to detect expression of most targets genes meaning that in this basal cell population the pathway was inactive. We could detect expression of *HEY2* and *CCND1* though, but not of *HEY1*, *HES2* nor *HES4*, the latter being the most activated in our system. Rock et al. have described a

minimal activity of the pathway in homeostatic basal cells (Rock *et al.*, 2011). We detected the expression of *LFNG* (lunatic fringe) in the basal cell compartment (Fig. 46). It has been seen that *LFNG*, even though is not impairing the binding between *JAG1* and *NOTCH1*, somehow inhibits *NOTCH1* proteolysis, which is required for activation of the downstream signalling events of *NOTCH1* (Yang *et al.*, 2005; Zhang *et al.*, 2014). It was shown that specific *Lfng* gene depletion in mouse led to expansion of the basal layer, increased proliferation of luminal and basal cells in mice prostate (Zhang *et al.*, 2014). *LFNG* expression has been shown to be regulated by Notch. During mouse somite segmentation, *Lfng* displays an oscillatory expression pattern, as a consequence of a periodic activation and repression of the *Lfng* promoter. This promoter bears an enhancer/silencing region comprising a binding site for the complex CBF1/RBP-J mediator of Notch signalling (Holley, Geisler and Nüsslein-Volhard, 2000; Rida, Le Minh and Jiang, 2004). The oscillation seems to be regulated by *LFNG* itself in a negative feedback loop, activation of Notch signalling induces *Lfng* transcription but *LFNG* protein inhibits Notch signalling and thereby represses its own transcription (Bessho and Kageyama, 2003). In the airway epithelium, Notch signalling pathway may be activated or repressed in basal cells depending in the requirements of the epithelium during regeneration or during the slow turn-over occurring in adult homeostasis.

At the onset of the suprabasal cell population *NOTCH3* receptor expression is upregulated (Fig. 46). It has been described that *LFNG* enhances the activation of Notch3 (Zhang *et al.*, 2014). Interestingly, *LFNG* expression seems enriched in the latest BC appearing in the pseudotime. It would be of great interest to identify whether *LFNG* promotes *NOTCH3* upregulation in the supraBCs, thus taking part in the subsequent cell differentiation.

In the secretory and goblet cell compartment, we reported robust expression of the notch receptor *NOTCH2* and the notch ligand *NTN1* (Fig. 46). As *NTN1* is able to bind to the *NOTCH2*-*JAG1* complex and activate Notch signalling pathway (Ylivinkka *et al.*, 2013; Ylivinkka, Keski-Oja and Hyytiäinen, 2016), it could be an important activator of the pathway, together with the canonical ligands. In the deuterosomal and MCC compartment there is a clear shift where *NOTCH2*, *NOTCH3*, *HEY1* and *HES4* are downregulated and *NOTCH4* and *JAG2* are up regulated together with the ligands *DLL1* and *DNER1*. The absence of target genes detection fits well with the well-known downregulation of the Notch pathway which is necessary multiciliogenesis. Thus, if MCCs express some Notch ligands and one Notch receptor, the pathway remains inactive. This may be because of the expression of *CIR1* and *SAP30* which are transcriptional co-repressors and *DYRK1A* which are able to inhibit Notch signalling through phosphorylation of *NICD* preventing it from forming a complex with RBP-J impairing the activity as a transcription factor for *HES1* (Abbassi *et al.*, 2015). We can detect also the very specific expression of *HES6* which has been reported as a repressor of the Notch pathway (Yang *et al.*, 2005; Ball *et al.*, 2007).

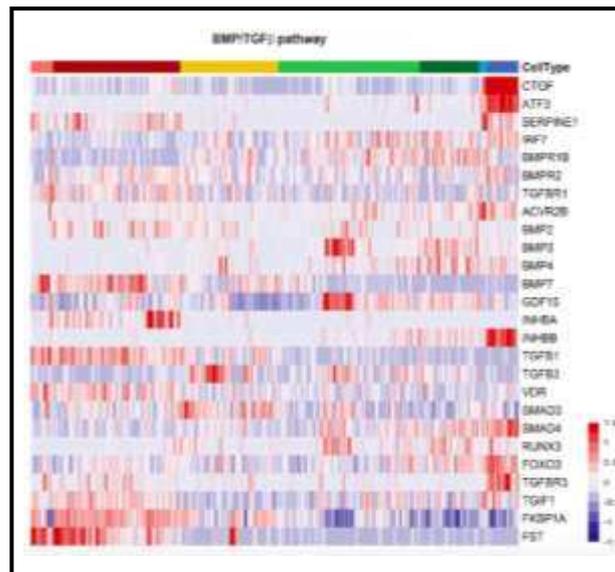


**Fig.46. Differential expression of the genes related to the NOTCH pathway.** Heatmap of the genes related to the Notch pathway with cells cultured in PneumaCult media ordered by cluster

#### *BMP AND TGF- $\beta$ pathways*

In the BC population we could detect the high expression of *FST* (Fig. 47), which is known to inhibit *Bmp2* (Maguer-Satta *et al.*, 2003; Akamatsu *et al.*, 2009). *Bmp2* has been related to suppress the differentiation and sustain the stem cell self-renewal of embryonic stem cells (Ying *et al.*, 2003) Interestingly, it has been reported that *Bmp2* upregulates *Hes1*, this is interesting because *Hes1* acts as a safeguard against irreversible cell cycle exit preventing from inappropriate differentiation (Mira *et al.*, 2010). In our system *FST* is expressed in BCs but also in cycling BCs, meaning that can be controlling the cell cycle exit of the basal cells pushing them into the differentiation program.

If we look the expression levels of the components of the TGF- $\beta$  pathway we can see an activation of the pathway at the level of the deuterosome and multiciliated cell compartment with the specific expression of genes such as the connective tissue factor (*CTGF*). TGF- $\beta$  regulates the formation of the motile cilia (Tözser *et al.*, 2015).



**Fig.47. Differential expression of the genes related to the BMP/TGF- $\beta$  pathway.** Heatmap of the genes related to the BMP/TGF- $\beta$  pathway with cells cultured in PneumaCult media ordered by cluster.

## CONCLUSION

With our study of human airway epithelium regeneration, we could determine the cell trajectories leading to the complete differentiation of the epithelium *in vitro*, and all the different cell types involved in the process. We determined new cell populations and new molecular interactors controlling the process of differentiation. With all these new data-sets we have a whole new horizon open to analyse the role of these cells and signalling molecules in healthy regeneration epithelium. Now, with all this new information, we can study the different defects during regeneration leading to the different respiratory diseases.

## IX.2 Airway remodelling

### CONTEXT OF THE STUDY

As mentioned before, injury and inflammatory events are occurring in a chronic manner in patients with SCD such as CF, COPD and asthma, leading to the remodelling of the airway epithelium and causing different traits associated with these diseases: mucin hypersecretion, increased viscosity and dehydration of the mucus, impairment of the beating of the motile cilia of the MCCs, loss of MCCs and colonization of opportunistic pathogens with the development of recurrent infections.

This remodelling can be caused by different “triggers” such as mechanical stress or inflammation.

IL-13 is a th2-type cytokine released during the inflammatory response to those “triggers”. This cytokine is central in asthma, COPD, idiopathic pulmonary fibrosis (IPF) and chronic rhinosinusitis (CRS). The analysis of different biopsies of the airway epithelium of patients with asthma revealed a considerable thickening of the epithelium compared to controls, that led to a narrowing of the airways and airflow limitations (Pascual and Peters, 2005; Burgel and Nadel, 2008).

As airway epithelial remodelling represents an important aspect of these diseases it is crucial to identify the regulators of this remodelling. The identification of the different cell types and signalling pathways that are involved can help to discover new therapeutic targets.

Many authors revealed the impact of the stimulation with the IL-13 cytokine in the airway epithelium. As it was described during the introduction of the manuscript, this overexpression of IL-13 led to GC hyperplasia with the consequent increased mucus secretion and loss in MCCs, decreased expression of BPIFA1 and increased expression of periostin (POSTN) (Laoukili *et al.*, 2001; Takayama *et al.*, 2006; Sidhu *et al.*, 2010; Yeh, Lee and Hsu, 2010; Sehra *et al.*, 2011; Alevy *et al.*, 2012; Tsou *et al.*, 2015; Górska *et al.*, 2016).

Our recent work studying the regeneration of human airway epithelium in healthy conditions (Ruiz Garcia, Deprez *et al.*, *submitted*), gave us a rationale to understand the process of regeneration of the epithelium, and to define the different cell types and molecular interactors involved in this process. We inferred cell trajectories during human airway epithelial regeneration *in vivo* and *in vitro*. Now we are aiming at applying our tools to analyse the process of remodelling of the airway epithelium in a context of increased levels of IL-13. Using single cell transcriptomics, we are now revealing the cellular characteristics and the molecular pathways involved in the remodelling through the investigation of differentiated HAECs after addition of the cytokine IL-13.

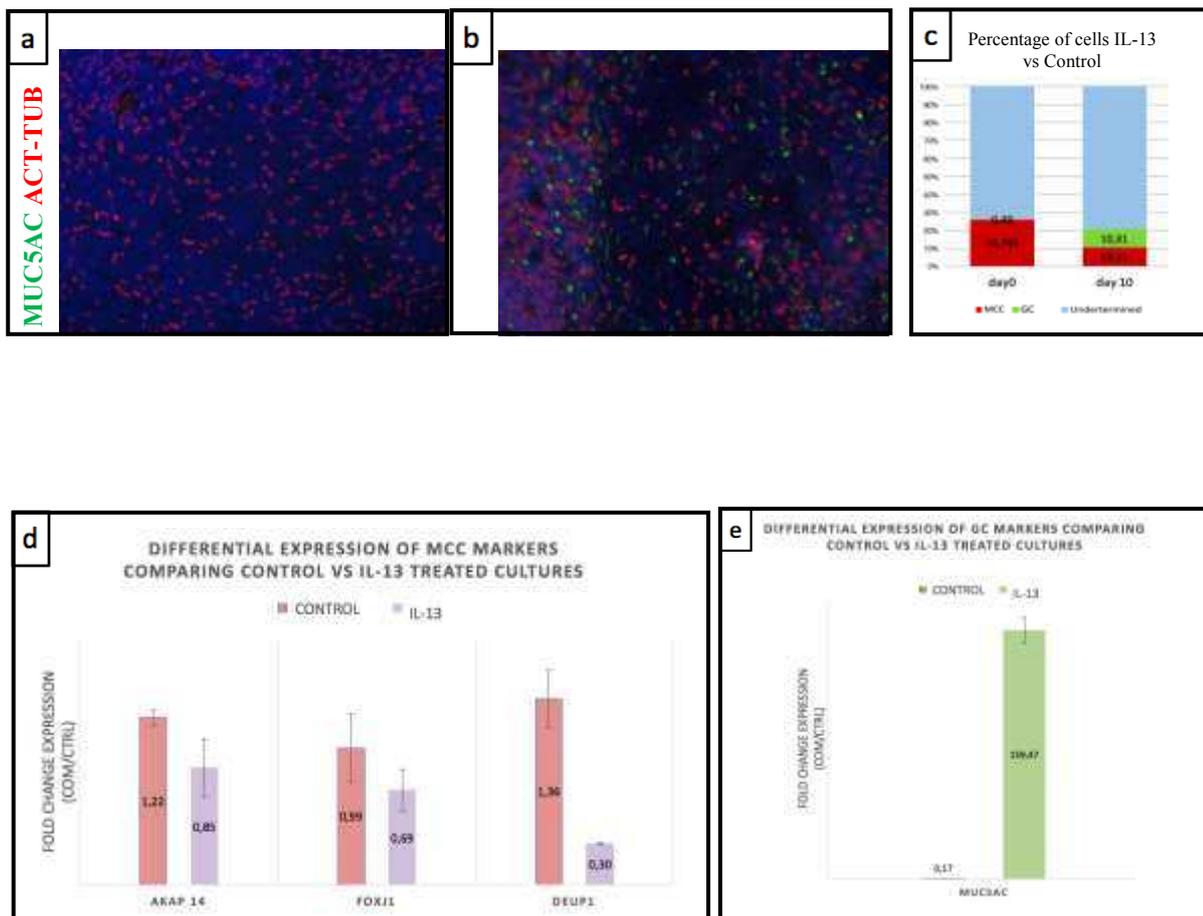
## RESULTS AND DISCUSSION

### *MCCs loss and goblet cell hyperplasia after IL-13 treatment*

All studies of airway remodelling by induction of an inflammatory state upon addition of IL-13 *in vitro* and *in vivo* showed the same effects regarding GC and MCCs (Kondo *et al.*, 2002, 2006; Atherton, Jones and Danahay, 2003; Tyner *et al.*, 2006). In order to demonstrate that the IL-13 treatment had the same effect in our experimental model of HAECs, we measured the levels of MCCs and Goblet cells through RT-qPCR gene expression analysis and Immunofluorescence analysis after 10 days of IL-13 treatment. We treated differentiated HAECs with 100 ng/ml added in the BEGM cultured media, in the basal compartment of the culture system described in (Ruiz Garcia, Deprez, 2018, *submitted BioRxiv*).

Immunostainings against MUC5AC and actetylated tubulin showed higher expression of MUC5AC compared with control (Fig. 48 a, b) and relative quantification of (Fig. 48 a, b) showed a decrease in MCCs by 15,47% and an increase of GCs by 9,82%.

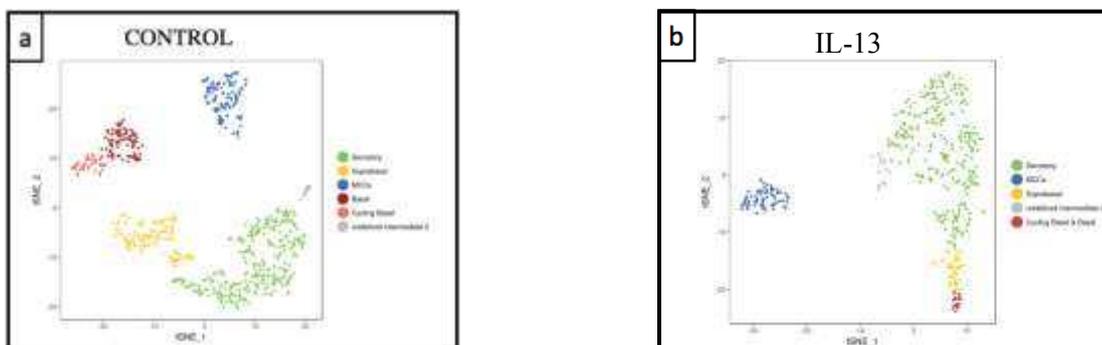
Analysis of the RNA expression level of AKAP14, FOXJ1 and DEUP1, which are markers of MCCs, confirmed the overall downregulation after treatment (Fig. 48d), whereas MUC5AC RNA expression was highly increased in treated cultures (Fig.48e).



**Fig. 48. MCC and GC number variations after IL-13 treatment.** (a and b) immunofluorescence staining for MUC5AC (green) and Acetylated tubulin (ACT-TUB) (red) in control condition (a) and after 10 days of IL-13 treatment (b). Nuclei are shown in blue (DAPI). (c) Quantification of relative number of MCCs (red) and GCs (green) from (a and b). (d and e) RT-qPCR of AKAP14, FOXJ1 and DEUP1, markers for MCCs (d), and MUC5AC marker of GCs (e) from primary HAECs in ALI culture after treatment with IL-13 at 100 ng/mL for 10 days. Control: non-treated.

### *Deciphering early and late remodelling effects of IL-13 treatment*

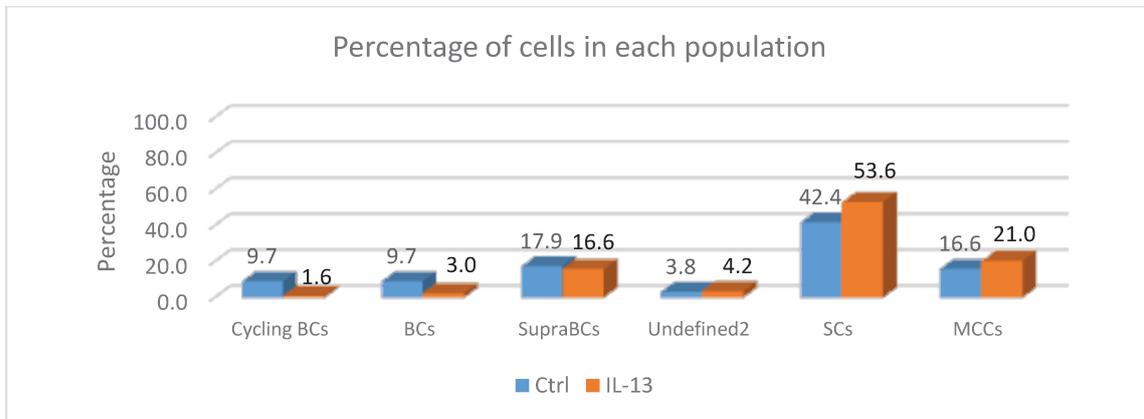
As it has been described in the literature the treatment of airway epithelial cells with IL-13 has different effects such as the decrease in MCCs and the increase in SCs and GCs. But the origin of this cell dynamics is still unclear. Our objective was to identify the events taking place at the very early stage of remodelling and in the late stage. We performed scRNA-seq on fully differentiated HAECs that were treated with IL-13 for 3 days. We have analyzed 429 treated cells and 663 control cells. The identification of each cell population was inferred from the expression of specific markers such KRT5 and TP63 for BCs, MKI67 for cycling BCs, KRT13 and KRT4 for supraBCs, BPIFA1 for SC-like, MUC5AC for GCs, and FOXJ1 for MCCs. We could identify all the cell populations that form the airway epithelium that we had described in our previous work (Ruiz Garcia, Deprez et al. 2018, *submitted ou BioRxiv*).



**Fig. 49. t-SNE plot of expression from scRNA-seq at early stage of remodelling.** (a) t-SNE plot of control cultured HAECs (663 cells) (b) t-SNE plot of three days IL-13 treated HAECs (429 cells). Colours indicates cell types.

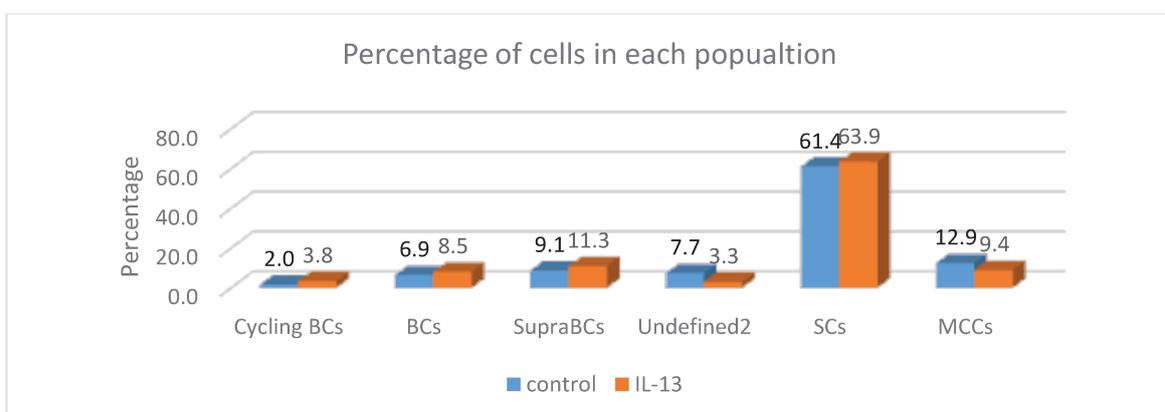
Quantification of cell populations obtained after the 3-day IL-13 treatment showed a decrease in cycling and non-cycling BCs (9,7% and 9,7% in controls, to be compared with 1,6% and 3,0% in treated cultures respectively) (Fig.49) and a slightly increase of SCs from 42% to 53% (Fig.49). These epithelium

modifications revealed that early remodelling events were affecting mostly the basal cell population compartment. At this early stage of remodelling we did not notice a significant decrease in the MCCs population (Fig. 50).

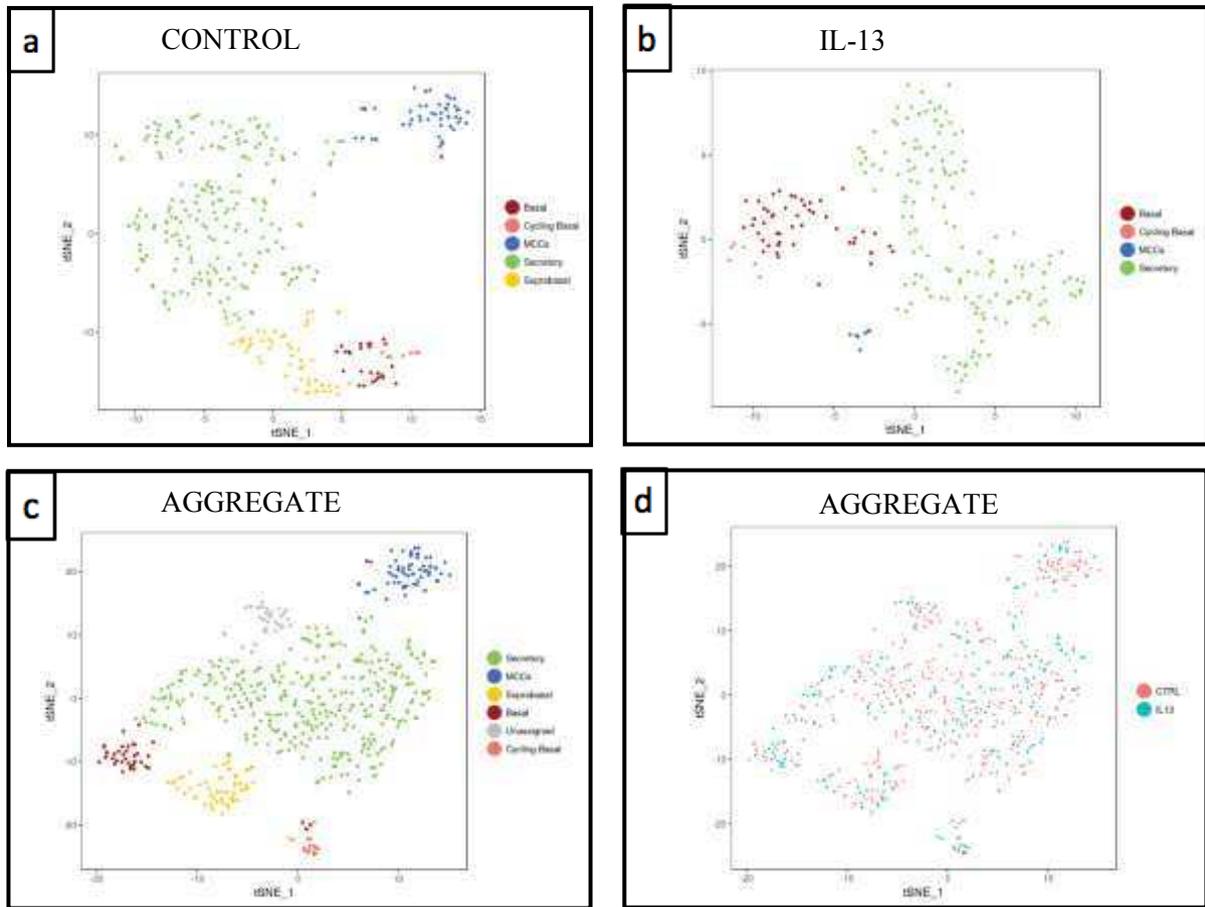


**Fig. 50. Relative abundance of the 6 main cell types at each time point after 3 days of IL-13 treatment of fully differentiated HAECs**

A similar analysis was performed over 213 epithelial differentiated cells from HAECs 3D cultures that were treated with IL-13 100 ng/ml during a period of thirty days after differentiation, and 350 epithelial differentiated cells from HAECs 3D cultures without treatment as controls. With this long-term treatment, we did not detect sharp differences in the number of cells in the different populations (Fig. 51), we detected all 6 different cell populations in treated and non-treated conditions (Fig.51). In the t-SNE plot representation of the different cell populations in control and IL-13 separately we didn't detect all the cell populations described at 3 days of treatment (Fig. 52 a,b), for instance, when we performed a t-SNE plot merging both conditions we detected that, in fact, all the cell populations are present (Fig. 52 c,d).

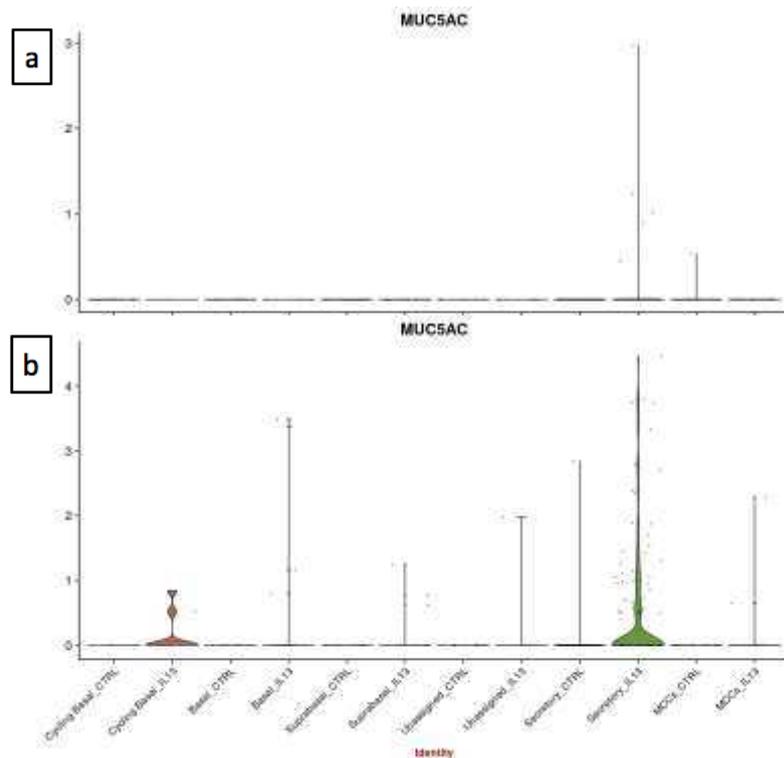


**Fig. 51. Relative abundance of the 6 main cell types at each time point after 30 days of IL-13 treatment of fully differentiated HAECs**



**Fig. 52. t-SNE plot of expression from scRNA-seq at late stage of remodelling.** (a) t-SNE plot of control HAECs (350 cells) (b) t-SNE plot of three days IL-13 treated HAECs (213 cells). Colours indicates cell types. (c and d) t-SNE of the aggregate (563 cells), (c) t-SNE plot representation of the 6 cell populations in the aggregate condition. Colours indicate cell types, (d) Projection in the t-SNE of the 350 cells of the control condition (pink) and 213 cells of the IL-13 treated condition (blue).

For the analysis of the effect of IL-13 treatment in the induction of goblet cell differentiation, we found that short treatment with IL-13 did not impact the amount of goblet cells measured by the levels of expression of MUC5AC (Fig 53a). However, a longer exposure to the cytokine gave rise to a high expression of MUC5AC in the SC population, surprisingly we detected also expression of MUC5AC in the cycling-BC population (Fig. 53b)



**Fig.53. Violin plots of normalized expression of *MUC5AC*.** (a and b) representation of the expression of *MUC5AC* in cell populations comparing control vs IL-13 treated cells. (a) Expression of *MUC5AC* after 3 days of IL-13 treatment at 100 ng/mL. (b) expression of *MUC5AC* after 30 days of IL-13 treatment at 100 ng/mL.

### ***Revealing the Th2-type inflammation molecular signature***

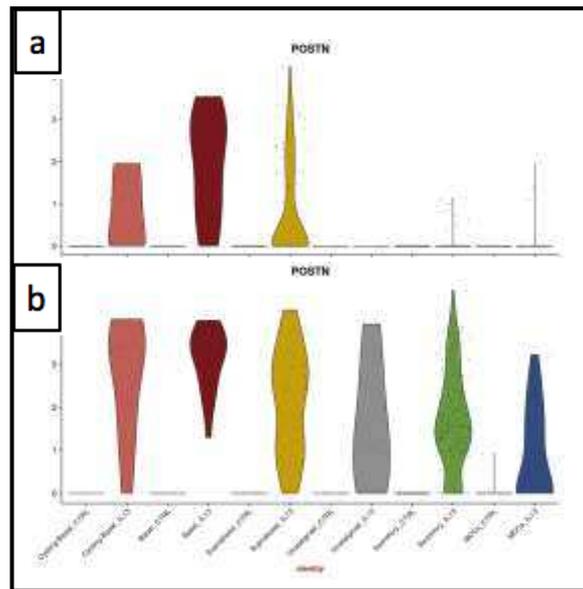
Different studies have revealed the molecular signature for the Th2-type inflammation in airways. Giovannini-Chami et al. performed microarray analysis over samples from nasal respiratory epithelium of healthy (11) or asthmatic (13) patients. Cells were recovered directly from the patients through brushings of the inferior turbinates. The analysis revealed the characteristic molecular signature of the asthmatic patients with overexpression of genes such as *CST1*, *POSTN*, *NTS*, *CD44*, *GSN* and *ALOX15* (Giovannini-Chami *et al.*, 2012).

Alevy et al. performed a microarray analysis of HAECs stimulated by IL-13, compared with non-treated cells they revealed the over expression of the same markers shown by Giovannini-Chami (Alevy *et al.*, 2012). Previous data obtained by our group (not published) have revealed the molecular signature of IL-13 treatment in our model of differentiated HAECs. They performed a IL-13 treatment during three days and bulk RNA-seq. The analysis showed increased expression of the markers already associated with the TH2-type inflammatory state such as *CST1*, *CCL26*, *POSTN*, *ITLN1*, *SERPINB2*, *FOXA3* and *SPDEF* among others (table 5).

**Table 5. Log fold change expression of the most expressed transcripts in the IL-13 condition compared with controls.**

Gene ID	logFC_IL-13_ctrl	adj.P.Val_IL-13_ctrl
<b>CST1</b>	8,904317232	0,015695006
<b>CCL26</b>	8,807876554	0,043976897
<b>POSTN</b>	5,212748864	0,037161468
<b>ITLN1</b>	2,81870569	0,061292592
<b>SERPINB2</b>	3,542325097	0,039647526
<b>SPRR3</b>	-1,024894187	0,083992908
<b>FOXA3</b>	0,82504905	0,347222693
<b>SPDEF</b>	5,41146176	0,02525545
<b>CAPN14</b>	5,152704905	0,039647526
<b>CD44</b>	1,786282244	0,046978627
<b>GSN</b>	1,75007453	0,042086757
<b>MUC13</b>	0,358969281	0,353121421
<b>NTS</b>	1,056160716	0,127075964

The analysis of the markers up-regulated in IL-13 treatment in our scRNA-seq datasets showed that their expression is related to specific cell populations in the epithelium. Our analysis revealed the expression of *POSTN* at initial stages of the remodelling mainly in the basal and suprabasal compartment of the epithelium (Fig. 53a), this is coherent with the results obtained by Sidhu that showed an increased secretion of periostin after treatment with IL-13 in cultured HAECs, and this secretion was towards the basal part of the epithelium rather to the luminal part (Sidhu *et al.*, 2010). Contrary to the results obtained by Sehra and colleagues who detected low levels of periostin in the airways of wt mice, we do not detect expression of *POSTN* in the control condition (Sehra *et al.*, 2011). Ordovas–Montanes and colleagues in their scRNAseq analysis of nasal HAECs from patients with polyps, also detected the expression of *POSTN* in the basal cell compartment(Ordovas-Montanes *et al.*, 2018). At later stages of the remodelling we detected the expression of *POSTN* disseminated through all the different cell populations in the IL-13 treated cells (Fig. 54). The work of Kelley Bentley *et al.* showed that periostin was necessary for the inflammatory response in mouse exposed to House Dust Mite (HDM) and a blockade of Periostin showed a decreased expression of several cytokines such as IL-13 and reduced presence of neutrophils, macrophage/monocytes, lymphocytes and eosinophils (Bentley *et al.*, 2014). Regarding these results, the question about the role of periostin in the remodelling of the airways arise, future analysis of our lab will consist in revealing the possible role of periostin in the GC hyperplasia.

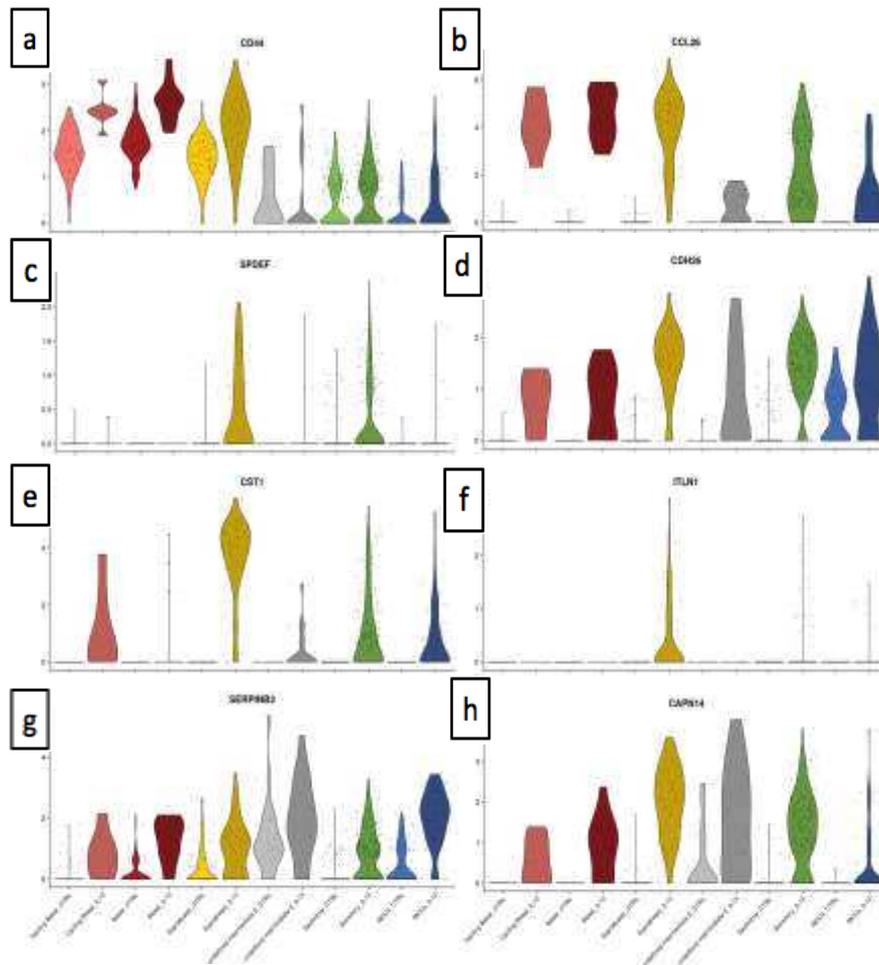


**Fig.54. Violin plots of normalized expression of *POSTN*.** (a and b) representation of the expression of *POSTN* in cell populations comparing control vs IL-13 treated cells. (a) Expression of *POSTN* after 3 days of IL-13 treatment at 100 ng/mL. (b) expression of *POSTN* after 30 days of IL-13 treatment at 100 ng/mL.

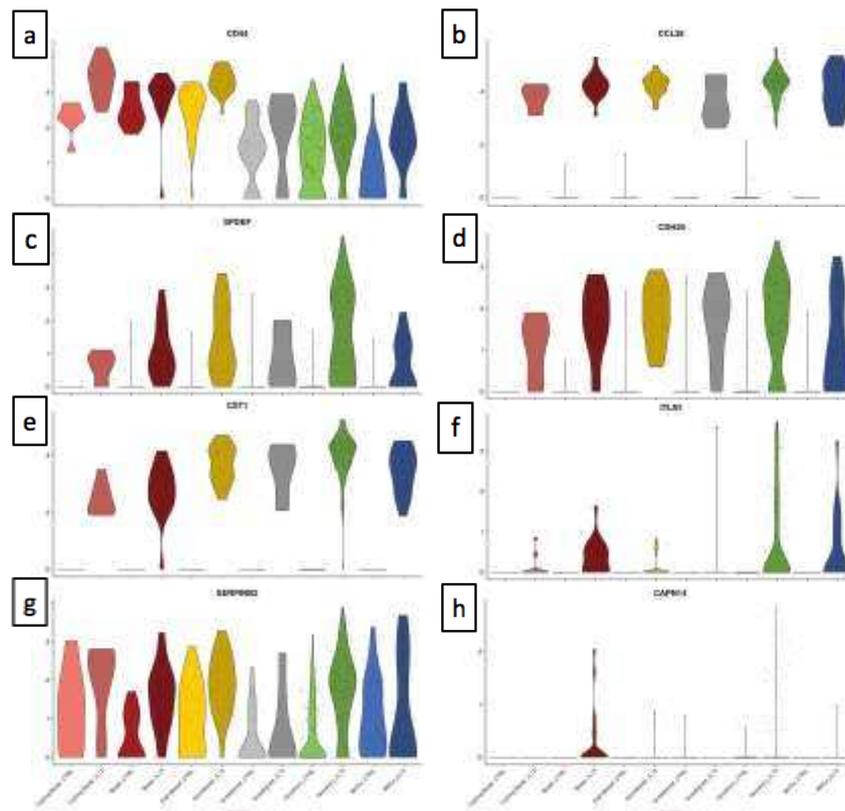
We assessed the expression of some of the genes differentially expressed in IL-13 condition (showed in Table 5) (Fig. 55 and 56). Regarding the expression of genes such as *CD44*, *SERPINB2*, *CCL26*, *CDH26*, *CST1* and *CAPN14*, their expression was clearly induced by the IL-13 treatment in all the different cell types, with a very homogeneous distribution in all the different cell populations after a 3-day or 30-day (Fig. 55 and 56). After a 30-day treatment, we detected a decrease in the expression of *CAPN14* compared with the 3-day treatment and this expression was restricted to the BC compartment (Fig. 55 and 56). It is known that *CAPN14* is upregulated in response to Th2-type inflammation in the oesophagus and is responsible of the impairment of the epithelial barrier (Litosh *et al.*, 2017). Its function in airway epithelium remodelling caused by Th2-type inflammation is unknown.

*SPDEF* expression in the IL-13 treated samples was restricted to the supraBC and SC populations after 3 days of IL-13 treatment (Fig.55c). This experiment was performed in BEGM media, in which no GC are detectable, hence in which expression of *SPDEF* in control cells is almost undetectable, as was previously shown in the results of our previous work (Ruiz Garcia, Deprez et al. 2018, *Submitted, BioRxiv*). Notably, after a 30-day of IL-13 treatment, expression of *SPDEF* was expanded to all cell populations (Fig. 56c). The pattern of expression of *ITLN1* was also intriguing. At early stage of remodelling its expression was detected in the supraBC population (Fig. 55f), a longer exposure to IL-13 showed a diffusion in the expression to BC, SC and MCCs (Fig.56f). Intelectin-1 (coded by *ITLN1*), also named as Omentin-A, is a protein constituent of

pathologic mucus, upregulated in Th2-type inflammation in asthma. It has a host defense function against pathogenic bacteria in the guts (Komiya, Tanigawa and Hirohashi, 1998; Kuperman *et al.*, 2005; Yi *et al.*, 2017). It has been suggested that intelectin-1 could be a component of pathologic mucus triggered by infections. Some intelectins have been reported to interact with mucins altering the biophysical properties of mucus. Pemberton *et al.* described that intelectin-2 interacted with MUC5AC mucin in the gastric mucus of sheep (Pemberton *et al.*, 2011). The study of Intelectin-1 in asthma patients revealed a prominent presence of the protein in sputum and in immunostained lung sections; the amount of protein detected was directly proportional to the severity of the disease, confirming the induction of the protein by the Th2-type inflammation. Intelectin-1 has been characterized as a lactoferrin receptor and interacts with lactoferrin in airway mucus in acute asthma (Kerr *et al.*, 2014). Interestingly, Intelectin-1 contributed in the up-regulated levels of expression of the pro-cytokines TSLP, IL-25 and IL-33 after exposure to allergen. An asthma mouse model defective for *Itln-1* showed protection from allergen-induced airway hyperresponsiveness and mucus metaplasia (Yi *et al.*, 2017). The up-regulation of *ITLN1* in the supraBC population in our model of early stage of remodelling (Fig.55f), suggested that the supraBC population could be the responsible for the early stages of the remodeling secreting the pro-cytokines TSLP, IL-13 and IL-25 in a Th2-type inflammation phenotype.



**Fig.55.** Violin plots of normalized expression for genes induced in the th2-type phenotype after 3 days of IL-13 treatment at 100 ng/mL. (a to h) representation of the expression of the different genes differentially expressed due to IL-13 treatment showed in Table 5 in the different cell populations comparing control vs IL-13 treated cells at early stage of remodelling. (a) *POSTN*, (b) *CCL26*, (c) *SPDEF*, (d) *CDH26*, (e) *CST1*, (f) *ITLN1*, (g) *SERPIN2*, (h) *CAPN14*.



**Fig.56. Violin plots of normalized expression for genes induced in the th2-type phenotype after 30 days of IL-13 treatment at 100 ng/mL.** (a to h) representation of the expression of the different genes differentially expressed due to IL-13 treatment showed in Table 5 in the different cell populations comparing control vs 3 days IL-13 treated cells. (a) *POSTN*, (b) *CCL26*, (c) *SPDEF*, (d) *CDH26*, (e) *CST1*, (f) *ITLN1*, (g) *SERPIN2*, (h) *CAPN14*.

#### ***Differential keratins expression after IL-13 treatment***

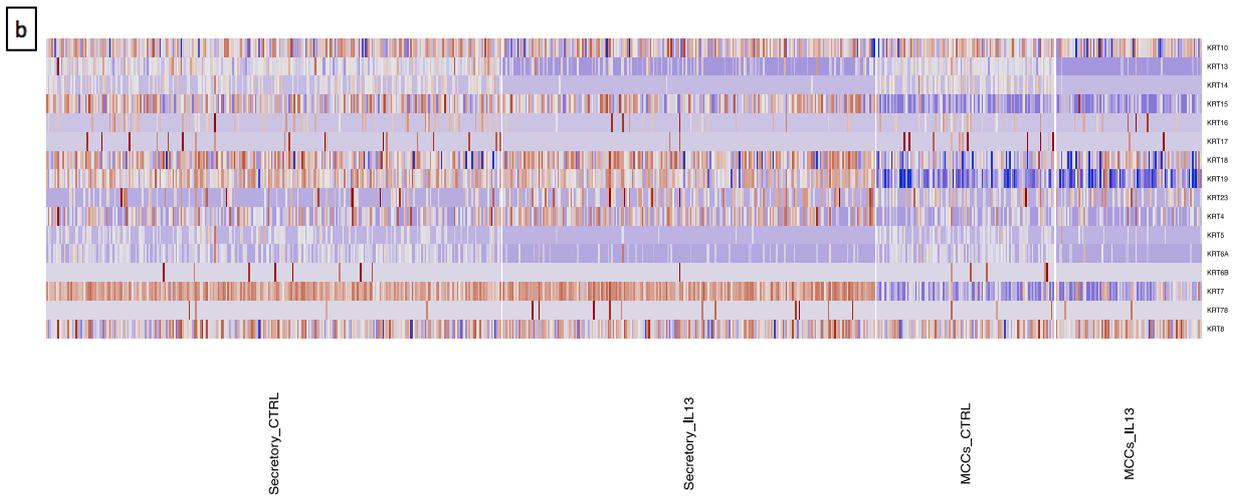
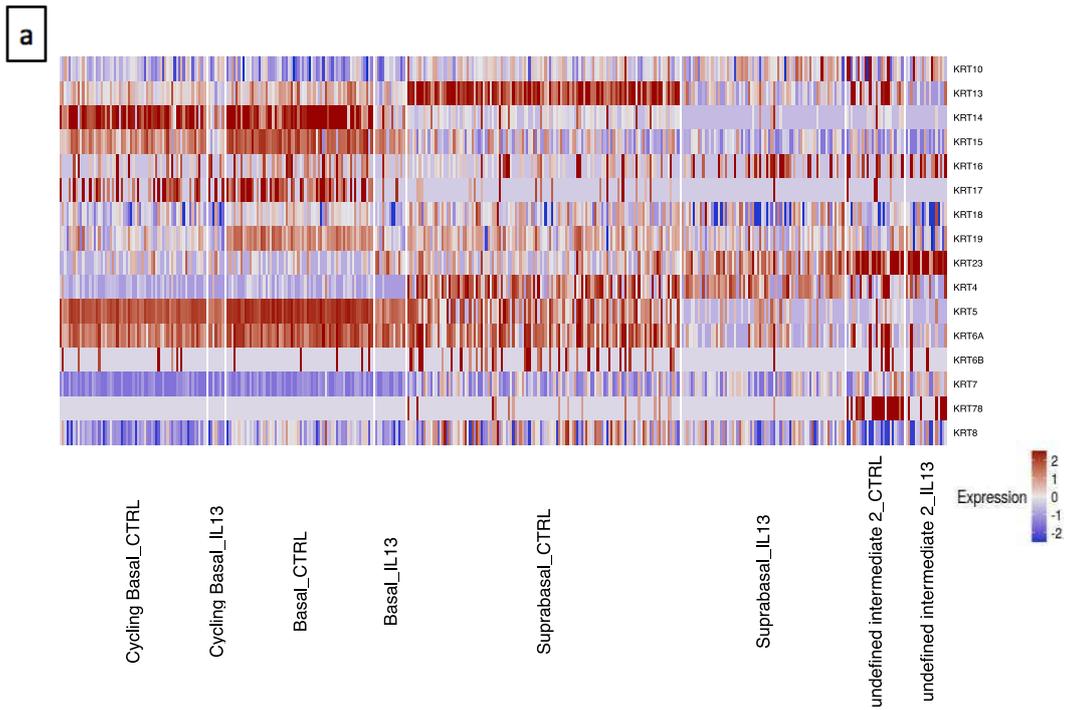
The characterization of cell populations can also be performed by analysing the expression of specific keratins. KRT5 and KRT14 are well known markers for basal cell populations in various tissues (Evans *et al.*, 2001; Kyung U. Hong *et al.*, 2004; Rock *et al.*, 2009; Ohashi *et al.*, 2010; Watson *et al.*, 2015). There are evidence about the direct relation of keratins expression and disease, such as the increased expression of KRT5 and KRKT14 in the IPF (Ficial *et al.*, 2014; N F Smirnova *et al.*, 2016; Confalonieri *et al.*, 2017). Altered keratin distribution has been associated with multiple atopic epithelial barrier disorders such as atopic dermatitis (AD); mutations in KRT5 and/or KRT14 are known to cause epidermolysis bullosa simplex (EBS) (Kiran, Rothenberg and Sherrill, 2015). KRT13 and its partner KRT4 are expressed in the suprabasal layers of the oesophagus (Ohashi *et al.*, 2010). Kiran *et al.* studied Eosinophilic esophagitis (EoE) which is a chronic allergic inflammatory disorder and is characterized by the implication of IL-13. This disease is characterized by defects in epithelial differentiation and barrier defects. One of the features of the disease is the decreased gene expression of epithelial keratins such as KRT14, KRT78 and KRT16. The blocking of this cytokine with anti-IL-13 therapy resulted in the rescue of the normal levels of these keratins. In vitro experiments showed

decreased keratin expression levels after IL-13 treatment of immortalized oesophageal human epithelial cells cultured in AIR-Liquid interface systems (Kiran, Rothenberg and Sherrill, 2015). The groups of Choi and Ye showed the overexpression of KRT19 in cultured HAECs after exposure to toluene diisocyanate (TDI), and the over expression of this cytokeratin in patients with asthma due to the exposure of this compound. More surprisingly they found serum IgGs against KRT19 in the serum of these patients, suggesting an autoimmune mechanism maybe involved in airway inflammation. One of the conclusions of these works is the capacity of diagnosis of this type of induced-asthma with the use of antibodies against KRT19 and KRT18 (Choi *et al.*, 2004; Ye *et al.*, 2006).

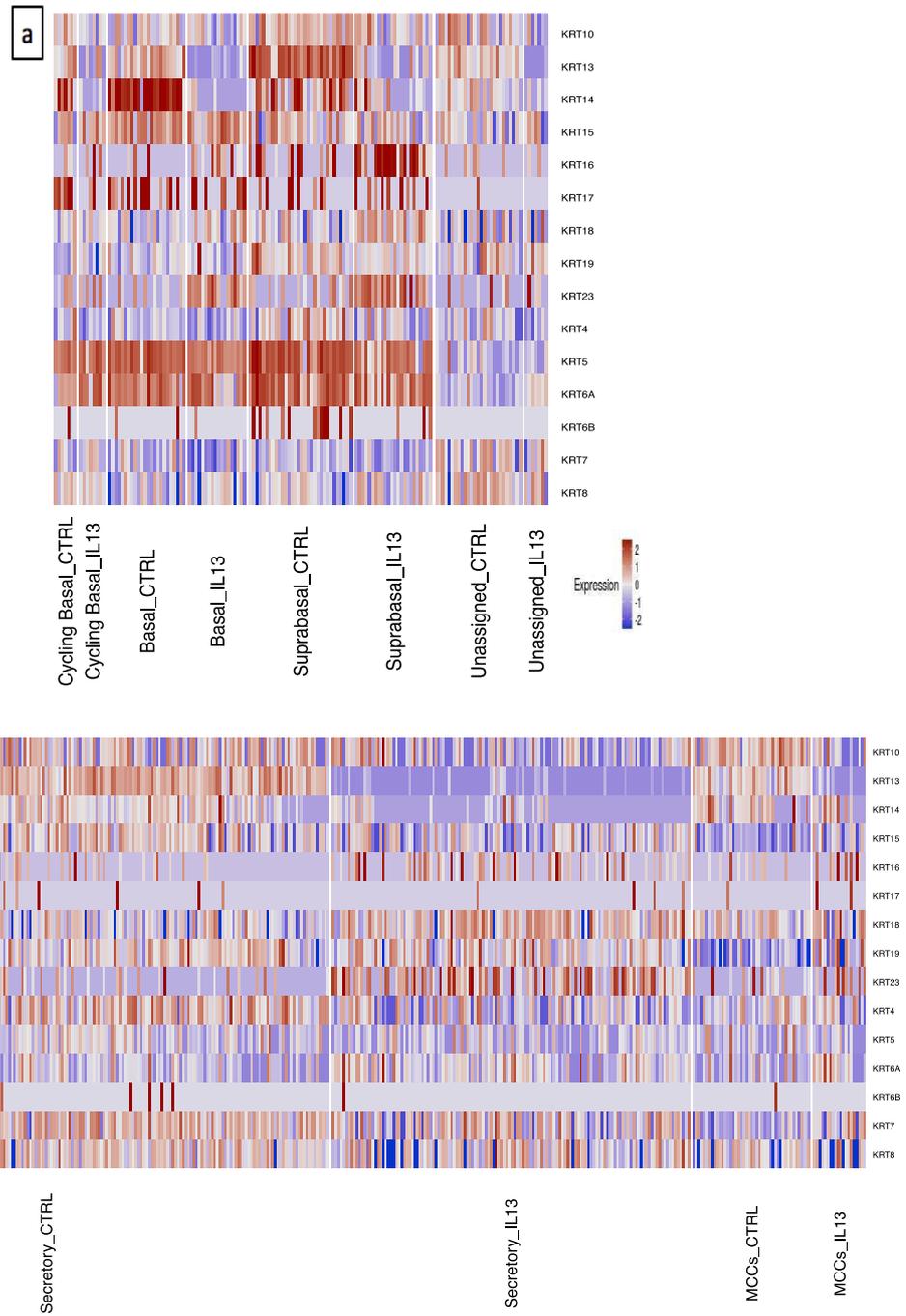
Our previous work (Ruiz Garcia, Deprez et al.2018, *as before*) defined a pattern of keratins expression specific to each population in the airway epithelium. We have also analysed the expression of keratins in our IL-13 treatment single cell datasets. After 3 days of IL-13 treatment, as mentioned above, the cycling and non-cycling BC populations were the ones with more changes in cell number.

Regarding the expression of keratins, there was an abrupt decrease of expression of *KRT14*, *KRT15* and *KRT17* in the remaining BCs after three days IL-13 treatment. The differential expression analysis in the suprabasal compartment showed the downregulation of *KRT13*, *KRT4*, *KRT6B* and *KRT6A*. Interestingly we found an up-regulation of *KRT23*, which is associated with the undefined intermediate population in both control and IL-13 conditions (Fig.57).

The keratins expression pattern after 30 days of treatment, among the BC population, KRT14 was the only keratin displaying a decrease upon IL-13 treatment. Differences were more noticeable in the supraBC and SC populations. Expression of *KRT23* in the supraBC population was increased. *KRT16* showed to be highly increased in the SC population of treated cells, together with a downregulation of *KRT13* and *KRT4*. This pattern of expression is continued in the SC population with a slight downregulation of *KRT10* that continues in the MCC population (Fig. 58).



**Fig.57. Keratins gene expression in the pseudotime after 3 days of IL-13 treatment.** (a,b) Heatmap for scRNA-seq data from control and IL-13 treated HAECs. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.



**Fig.58. Keratins gene expression in the pseudotime after 30 days of IL-13 treatment.** (a,b) Heatmap for scRNA-seq data from control and IL-13 treated HAECs. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.

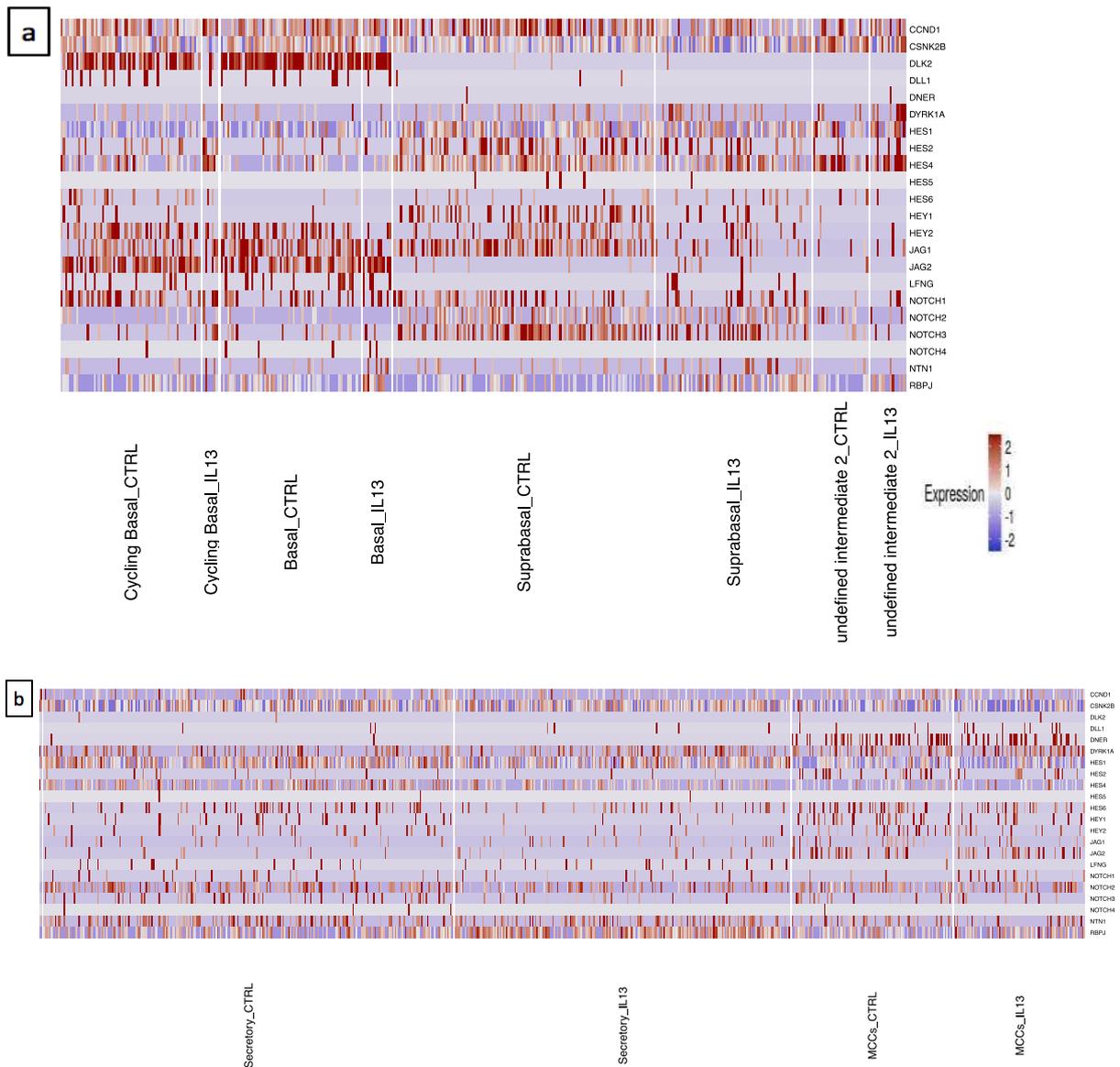
### ***Identification of molecular regulators of airway remodelling after IL-13 treatment***

To identify the molecular pathways involved in the airway epithelial remodelling after treatment with the Th2-type cytokine IL-13, we have analysed molecular regulators of the main pathways regulating normal regeneration of the epithelium as we described in our previous results (Ruiz Garcia, Deprez et al. 2018, *submitted BioRxiv*). However we only found remarkable differences between treated and non-treated conditions in the Notch and WNT signalling pathways.

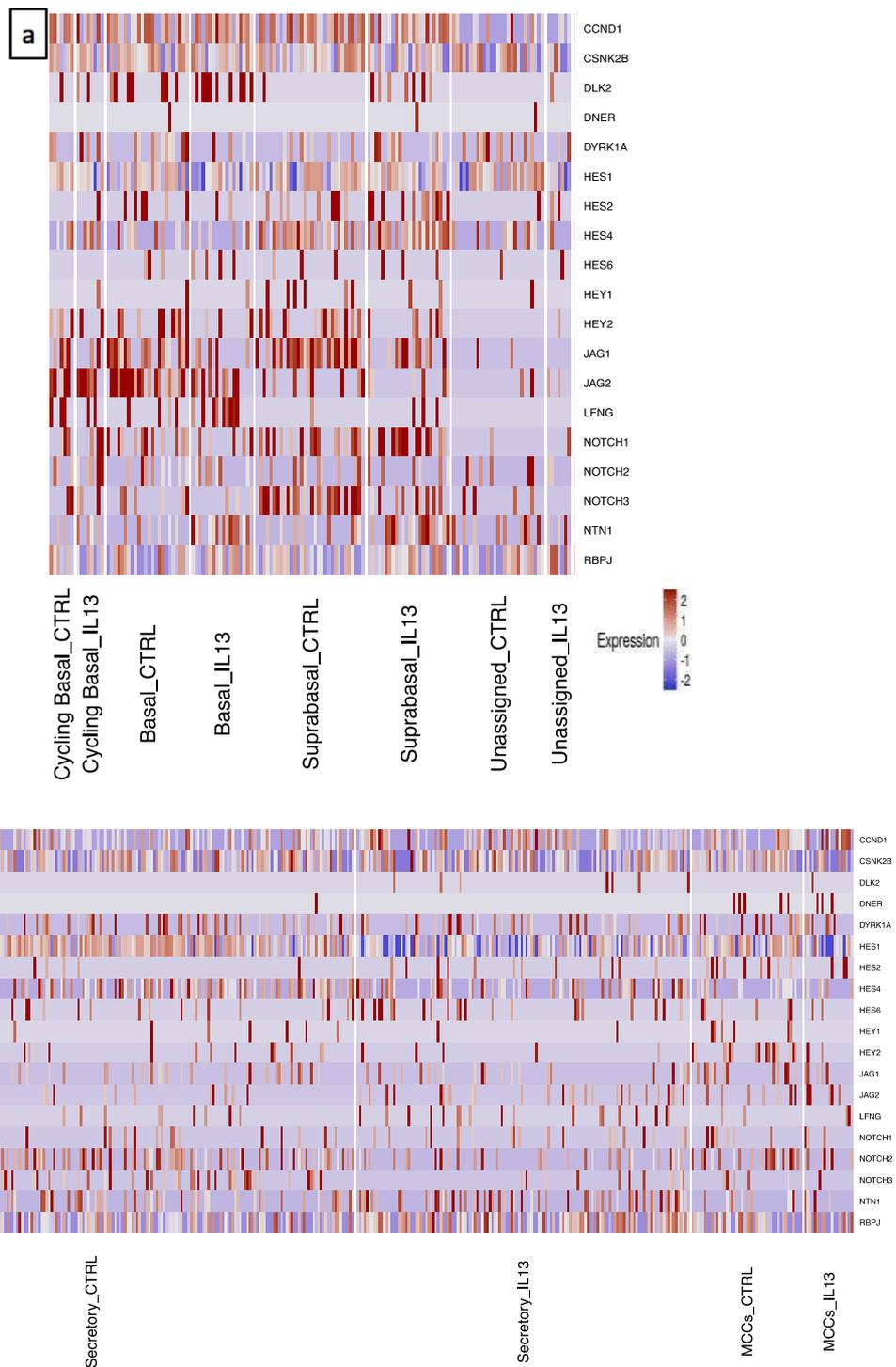
#### *Notch pathway*

After 3 days of IL-13 treatment, in the cycling BC population there was a slight increase of the Notch target gene *HES4* meaning an activation of this signalling pathway in the cycling BC compartment of the epithelium (Fig. 59). The supraBC population was characterized by the decreased expression of the *NOTCH3* receptor and the ligand *JAG1* and the up-regulation of *LFNG* meaning a repressor of the pathway (Fig. 59). We detected an increase in the expression of the transcription factor RBPJ after IL-13 treatment. *RBPJ* was demonstrated to be implicated in the activation of the Th2-type inflammatory cascade activation and in the differentiation of the Th2 cells. Depletion of the Notch pathway in mice (*Rbpj*<sup>-/-</sup>) demonstrated impaired differentiation of the Th2 cells and decreased inflammation and mucus secretion after induction of inflammation (M. Zhou *et al.*, 2015; Qu *et al.*, 2017). In our system the increased expression of RBPJ was mostly detected in the SC population. Taken together these results suggest that the secretory cells could increase their mucins secretion through a mechanism that requires activation of the Notch pathway.

After 30 days of IL-13 treatment, the cycling BC population did not show differences in the expression of the signalling pathway component compared with control condition. In the BC population there was an increased expression of *LFNG* followed by a decreased expression of *JAG1*. In the suprabasal cell population there was an activation of the pathway with the increased expression of *NOTCH1* (Fig. 60).



**Fig.59. Differential expression of the genes related to the NOTCH pathway after 3 days of IL-13 treatment.** Heatmap of the genes related to the Notch pathway with cells in control conditions and or IL-13 treated. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.



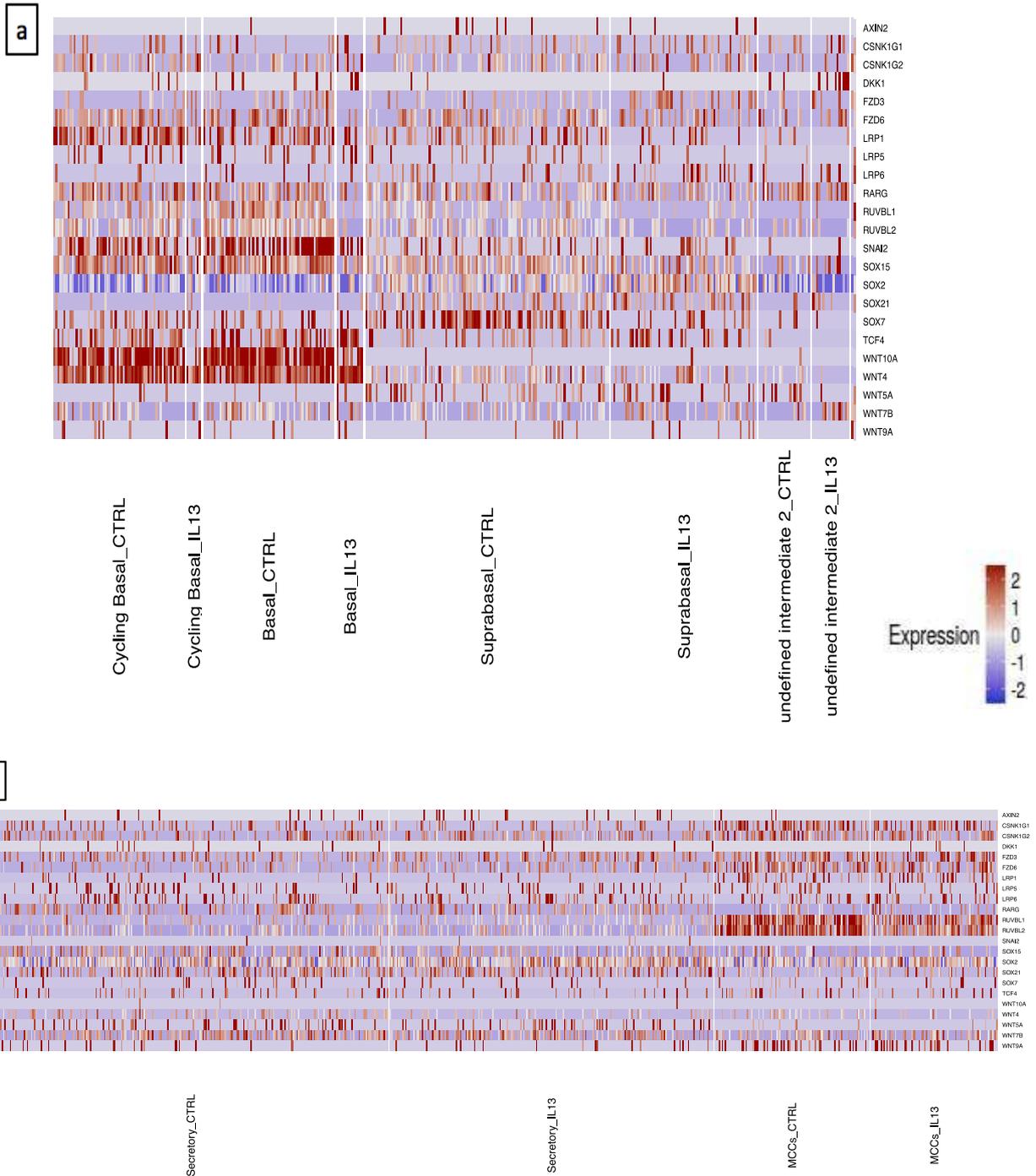
**Fig.60. Differential expression of the genes related to the NOTCH pathway after 30 days of IL-13 treatment.** Heatmap of the genes related to the Notch pathway with cells in control conditions and or IL-13 treated. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.

### *WNT pathway*

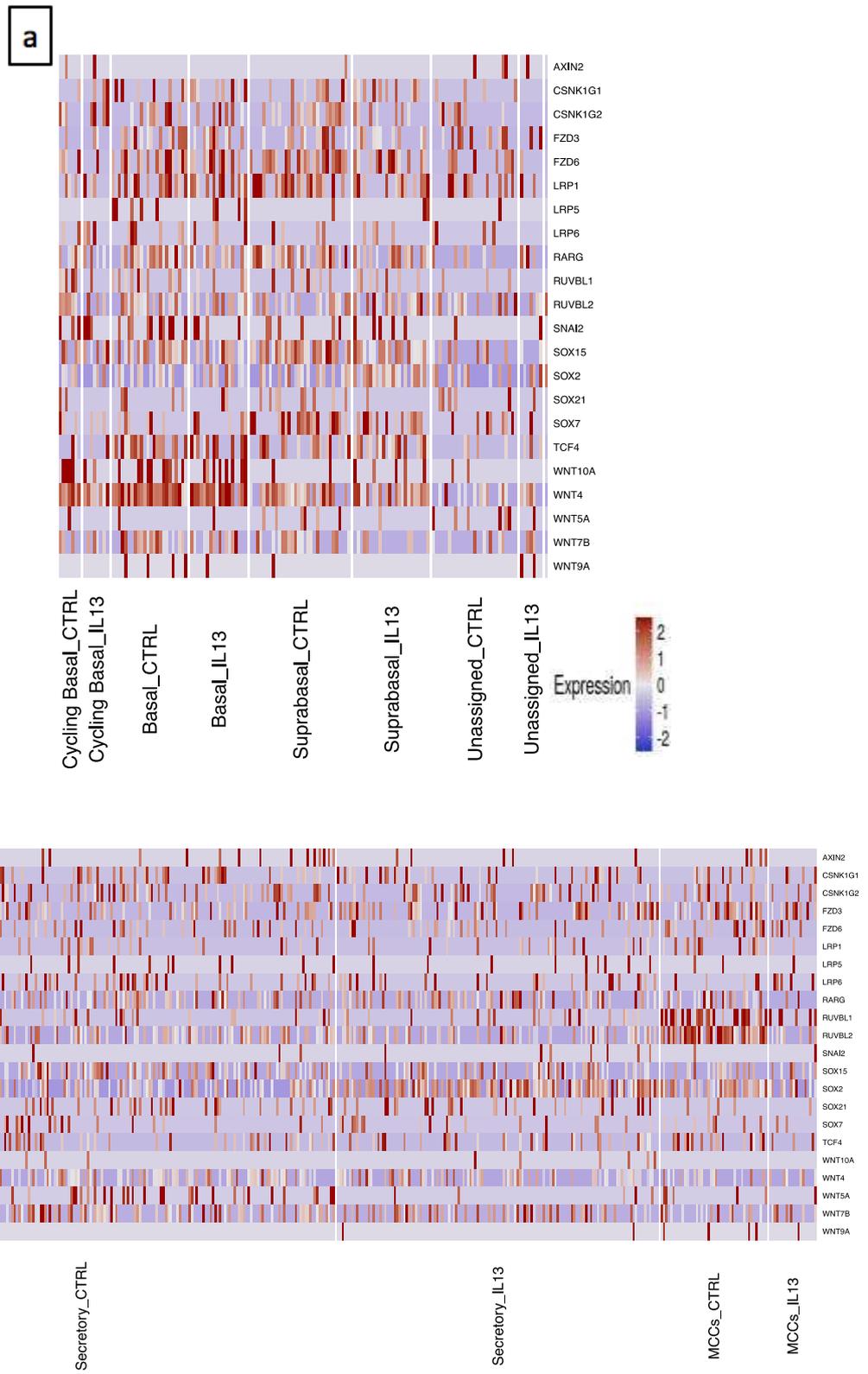
As I described in the introduction of the manuscript, the WNT signalling pathway is altered in lung diseases such as IPF, asthma or COPD. Our data showed a marked decreased expression of *WNT10A* and *TCF4* in the cycling basal cell population at 3 days of treatment with IL-13 (Fig. 61a). The decreased expression of *WNT10A* in treated cells in the cycling BC population is detectable also in the treatment during 30 days with IL-13. However, we did not detect differential expression of *WNT5A* in our system due to the treatment. We detected a decreased expression of *SOX7* in the cells treated with IL-13 during 3 days in the supraBC population (Fig. 61), this decreased expression is not detected in the longer treatment (Fig. 62). Also, the endogenous inhibitor *DKK1* is overexpressed in the undefined intermediate population in cells treated during 3 days.

In order to confirm the results obtained by other scRNAseq study performed in HAECs from patients suffering from CRS, where they described the activation of the WNT signalling pathway by the induction of the expression of a key effector of the pathway, *CTNNB1* ( $\beta$ -catenin) and expression of *CTGF* (specific factor of the pathway) (Ordovas-Montanes et al., 2018). We tested the expression levels of these genes in our system. Our results showed a slight increased expression of *CTNNB1* due to the IL-13 treatment at both stages of the remodelling, being more increased in the long treatment of 30 days (Fig 63 a,c). However, the expression of Connective Tissue Growth Factor (*CTGF*) was reduced in the treated cells (Fig 63 b,d), this protein is related to hypertrophy, proliferation and cellular matrix synthesis, and it was also shown to be up-regulated in a mouse model of asthma (Lin et al., 2017), maybe our system of in vitro culture of HAECs did not favor the transcriptional activation of this gene.

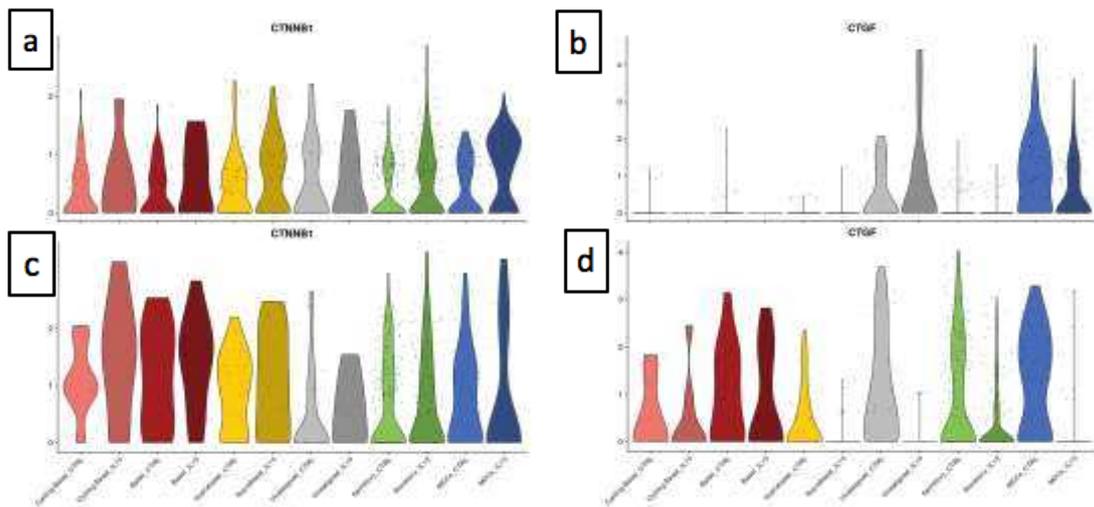
In conclusion, we did not detect the same expression patterns as the other studies in the field in relation with the different genes of the WNT pathway, however, there is a slight increase in the activation of the pathway due to the IL-13 treatment.



**Fig. 61. Differential expression of the genes related to the WNT pathway after 3 days of IL-13 treatment.** Heatmap of the genes related to the WNT pathway with cells in control conditions and or IL-13 treated. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.



**Fig. 62. Differential expression of the genes related to the WNT pathway after 30 days of IL-13 treatment.** Heatmap of the genes related to the WNT pathway with cells in control conditions and or IL-13 treated. (a) cycling-BC, BC, supraBC and undefined intermediate2 cell populations. (b) SC, MCC populations.



**Fig.63. Violin plots of normalized expression for genes coding CTNNB1 and CTGF.** for the receptors subunits of IL-13 induced in the Th2-type phenotype after 3 days (a,b,e,f) or 30 days (c,d) of IL-13 treatment. (a,c) *IL13RA1*, (b,d) *IL4R*, (e) *IL4I*,(f) *IL13RA2*.

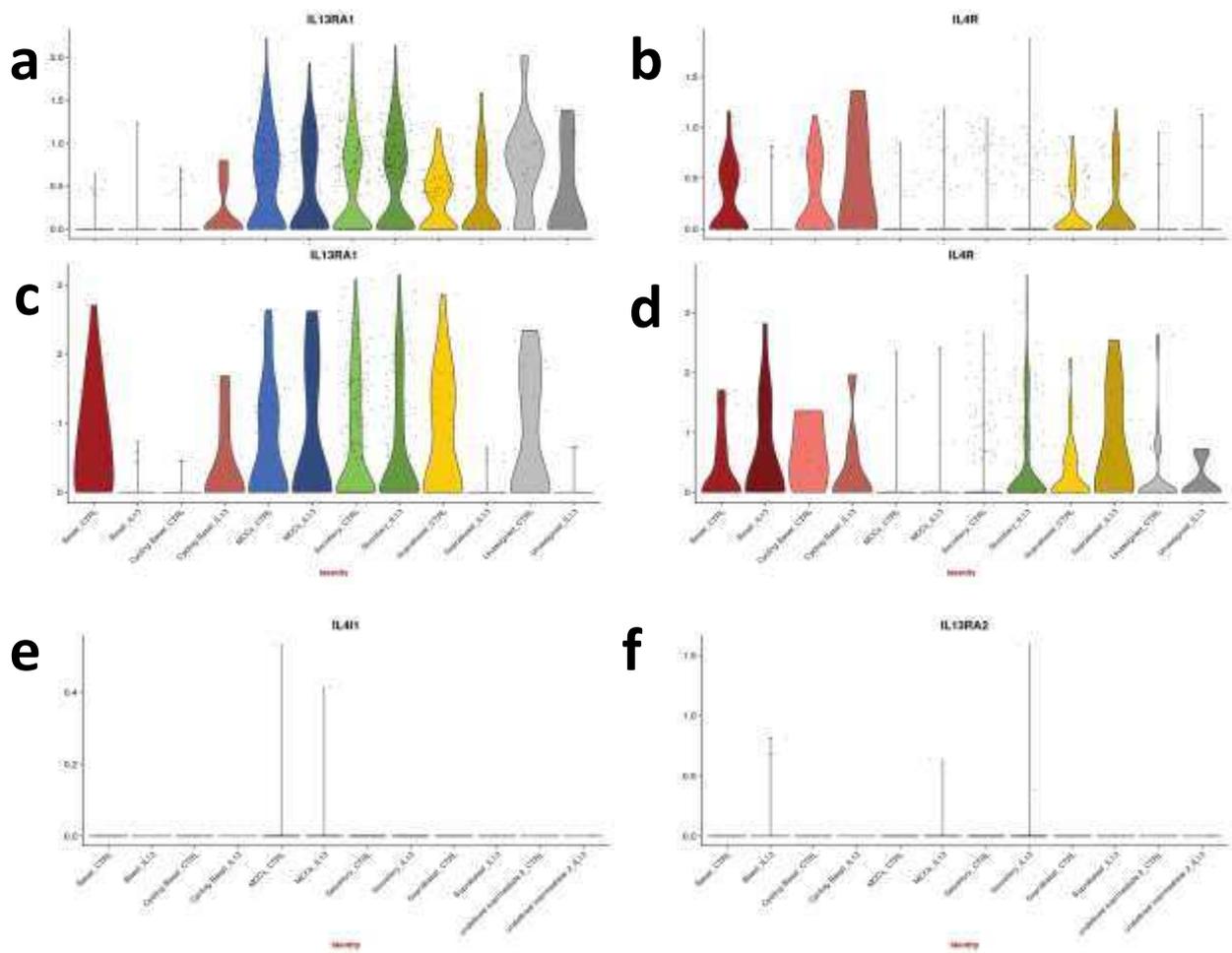
#### ***Expression of the IL-13 receptor subunits***

Aiming to reveal the cells directly involved in the initiation of the IL-13 cascade, we checked for the expression of the genes coding for the subunits of the IL-13 receptor (*IL13RA1* and *IL4R*) after 3 and 30 days of treatment.

We detected a wide expression of *IL13RA1* in mostly all the cells types at early remodelling state (fig. 64a), at late remodelling the expression was detected also distributed among almost all of the cell types (Fig. 64c). The expression of *IL-14R* was increased in the cycling BC and supraBC populations after 3 days of treatment (Fig. 64b), after 30 days of treatment the expression expands to the SC population (Fig. 64d). Meaning that the expression of the receptor subunit *IL4R* is increased after treatment with IL-13 in the BC, supraBC and SC populations.

We checked whether other genes coding for other genes non-specific to IL-13 receptor such *IL4I1* and *IL13RA2*, we detected no expression of these genes in none of the cells of our system in control or treated conditions (Fig. 64 e,f).

In conclusion the receptors *IL13RA1* subunits for the IL-13 cytokine are widely expressed in mostly all the cells of the human airway epithelium and their expression is not increased after induction of the th2-type inflammation conditions. However, the subunit *IL4R* of the IL-13 receptor is more expressed after induction of the inflammation phenotype.

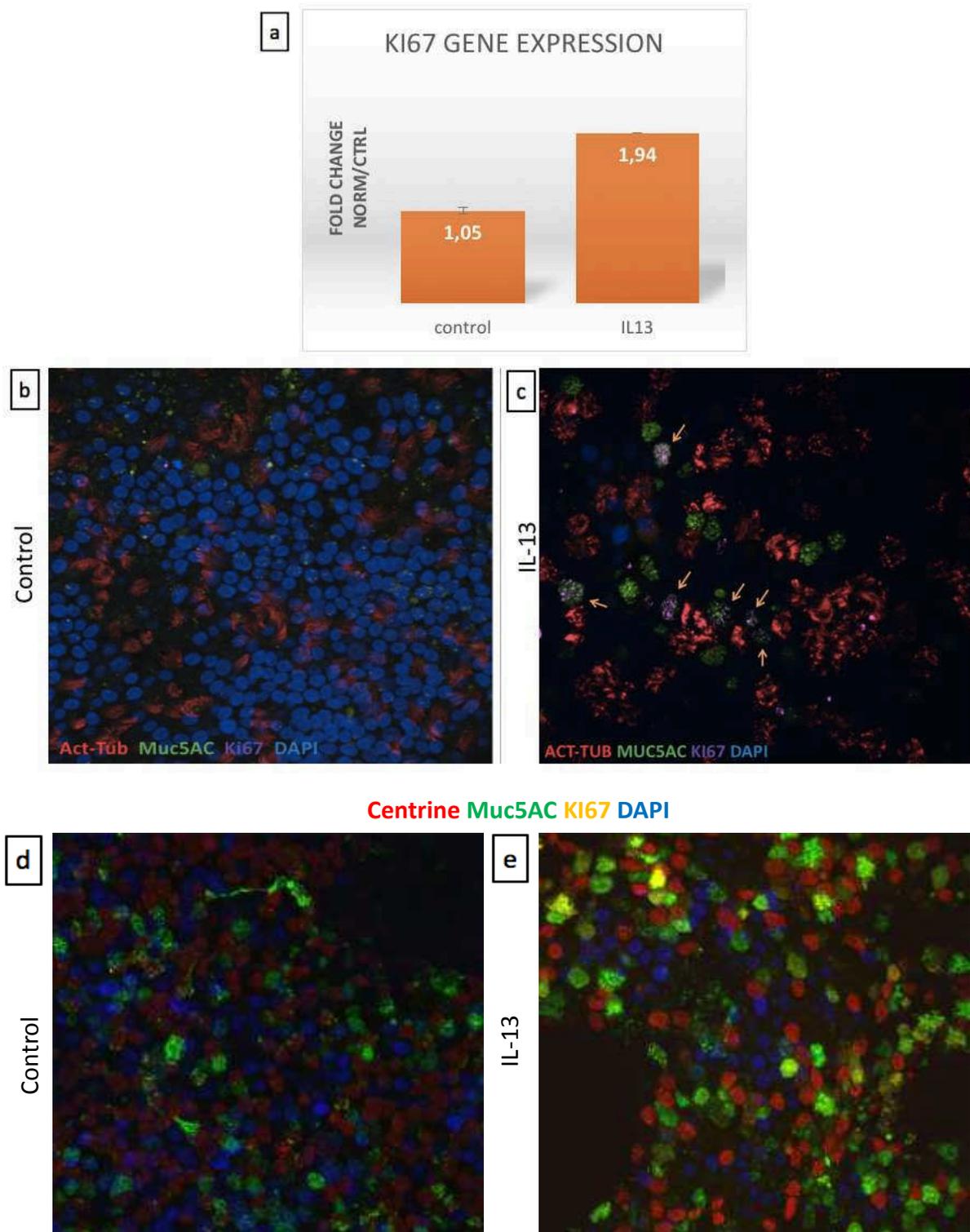


**Fig.64. Violin plots of normalized expression for genes coding for the receptors subunits of IL-13 induced in the Th2-type phenotype.** Differential gene expression after 3 days (a,b,e,f) or 30 days (c,d) of IL-13 treatment. (a,c) *IL13RA1*, (b,d) *IL4R*, (e) *IL4I1*,(f) *IL13RA2*.

### ***Ki67 increased after IL-13 treatment***

Several authors revealed a process of “transdifferentiation” of MCCs into GCs during inflammation, in mouse models of asthma and after addition of IL-13 in human or mouse airway epithelial cells (Tyner *et al.*, 2006; Gomperts *et al.*, 2007). They demonstrated that the GC hyperplasia triggered after induction of the inflammation was in part caused by this process of transdifferentiation. However, the process responsible of the transdifferentiation events in these studies is not clear. Moreover, Pardo Saganta *et al.* in a OVA challenge mouse model showed that the pool of MCCs did not cooperate in the generation of the new goblet cells (Pardo-Saganta *et al.*, 2013).

We tested whether in HAECs, the goblet cell hyperplasia observed after treatment was related to an increase of cell proliferation. After treatment of HAECs with IL-13 at 100 ng/mL for 10 days, we measured the levels of expression of KI67, known marker used for the detection of proliferating cells, at protein level by immunofluorescence and at mRNA level by RT-qPCR. We detected a slight increase in the expression of KI67 after treatment with IL-13 (Fig.65 a, c) compared to controls (Fig. 65 b). Surprisingly we detected KI67 expressed in the apical surface of cells expressing MUC5AC (Fig. 65 d). The detection of KI67 in the apical surface of the cells seemed to be mimicking the BB distribution in MCCs. We showed that KI67 was co-localizing with centrine2, a marker of centrioles, used as marker for BB in MCCs, in the IL-13 treated cells, apical expression of KI67 was not detected in controls (Fig. 65 d,e).



**Fig. 65. KI67 expression after IL-13 treatment.** (a) Fold change gene expression of KI67 analysed by RT-qPCR of CCP110 from primary differentiated HAECs in ALI culture after treatment with IL-13 during a period of 10 days vs non-treated as a control. Normalization with controls, mean  $\pm$ S.D. (b) Immunofluorescence staining for MUC5AC (green) and Acetylated tubulin (ACT-TUB) (red) in IL-13 treated condition after 10 days of treatment, orange arrow show co-localization of both proteins. (c,d) Immunofluorescence staining for MUC5AC (green), Acetylated tubulin (ACT-TUB) (red) and KI67

(magenta) control (c) and in IL-13 treated condition after 10 days of treatment, orange arrows show co-localization of MUC5AC and KI67 (d). Nuclei are shown in blue (DAPI). (e,f) Immunofluorescence staining for MUC5AC (green) and centrin-2 (red) and KI67 (yellow) in controls (e) and IL-13 treated condition after 10 days of treatment (f). Nuclei are shown in blue (DAPI).

#### ***New role of KI67 in airway remodelling.***

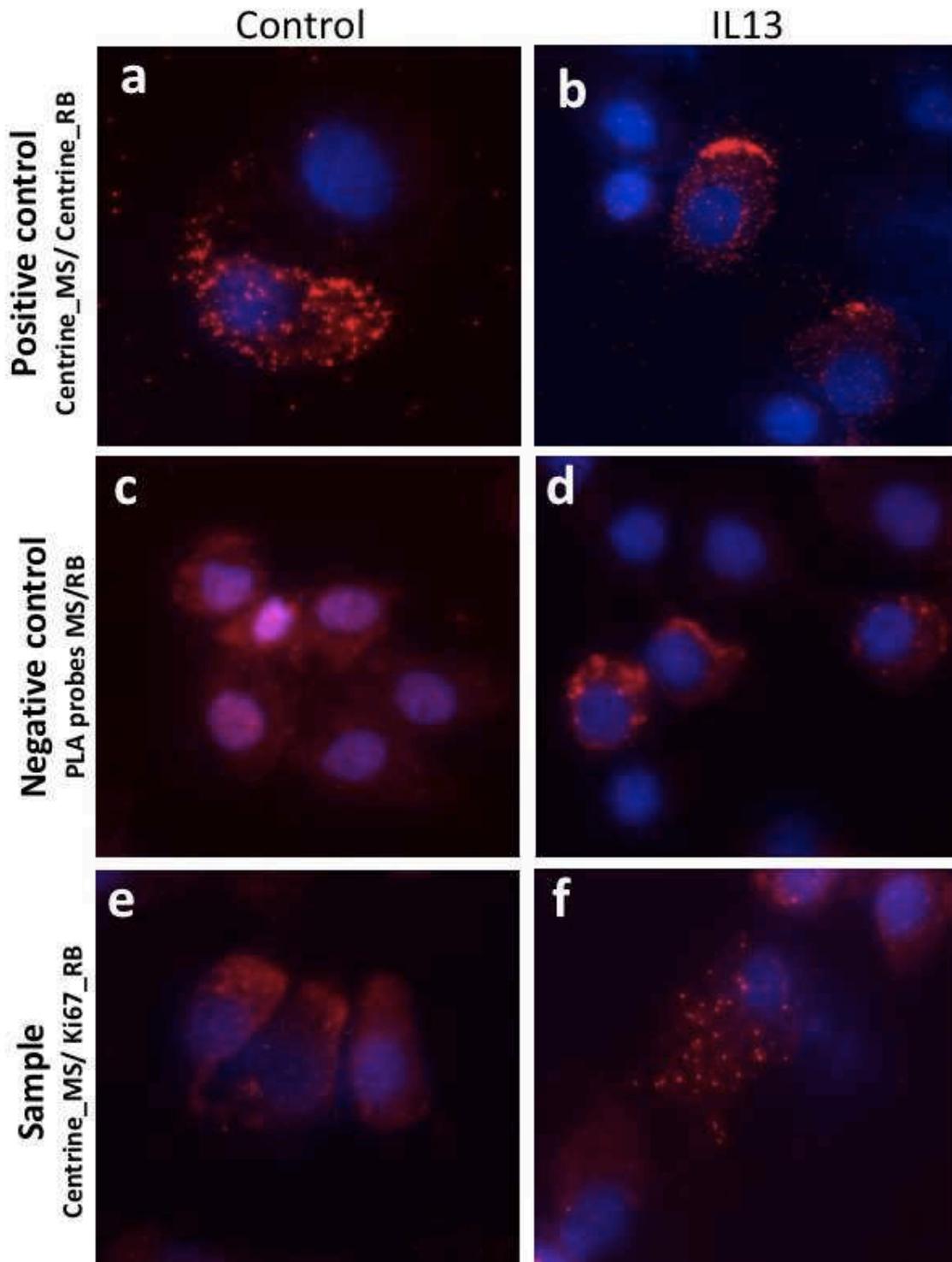
KI67 is widely used as a marker of proliferation and it has been suggested as a marker for cancer prognosis, predicting the clinical outcomes for patients with several types of cancer (Jaafari-Ashkavandi, Mehranmehr and Roosta, 2018; Jing *et al.*, 2018; Raghavan *et al.*, 2018), but it has been never related with cilia. KI67 was shown to interact with the centrosomal protein CCP110 (Hein *et al.*, 2015) which is related with the control of cilia formation. We showed that mRNA levels of CCP110 increased after IL13 treatment (Fig 66). We hypothesized the interaction between CCP110 and KI67 forming a complex involved in the process of reabsorption of cilia in MCCs. As was showed by Meunier *et al.* KI67 is detected during the process of centriole duplication through the deuterosome protein platform (Al Jord *et al.*, 2017).

The detection of KI67 in the apical surface of the cells resembled to be mimicking the BB distribution in MCCs. We showed that KI67 it was not co-localizing with centrin-2 (centriolar marker), used as marker for BB in MCCs, in the non-treated sample (Fig. 65d), whereas it seemed to co-localize with centrin-2 in the IL-13 treated cells (Fig.65e). We further confirmed the co-localization of ki67 with centrin-2 through a proximity ligation assay (Duolink) (Fig 68) we performed this assay in cellular suspensions of dissociated cells (cytospins), where the red dots appearing in the assay of the IL-13 treated sample revealed that these two proteins were proximal to each other (Fig 68 f). However, the staining did not show an apical co-localization, this could be due to the technique itself because the cytospin altered the structure of the cell, or could be due to the delocalization of the BBs in the MCCs giving rise to the image of the interacting proteins disperse in the cytoplasm.

These results led to the hypothesis of a new unknown role of the protein in the maintenance/reabsorption of cilia probably triggered by a process of metaplasia of the MCC population where the MCCs are suffering a process off de-differentiation.

Surprisingly we did not detect an increased expression of *KI67* in the MCCs due to treatment (Fig. 67). The analysis of the expression of *KI67* at three days of treatment did not showed any change comparing with controls, in both conditions *KI6* is expressed in the cycling basal cell population. However, we detected different pattern of expression in the cells treated during 30 days, surprisingly we detected an increased expression of KI67 in the SC population. This may suggest two hypotheses, or the SC population re-enters in the cell cycle and proliferates after the treatment or that the events of co-localization of KI67 are, in fact, dedifferentiation of MCCs in SCs and the gene expression signature already cluster those cells as SCs where they have lost their identity as MCCs.





**Fig. 68. Proximity ligand assay of KI67 and CCP110 in HAECs cells.** (a,b) Positive control of the proximity ligation assay using centrin-2 (host species mouse) and centrin-2 (host species rabbit) antibodies in control HAECs (a) and in IL-13 treated cells (b). (c,d) Negative control of the proximity ligation assay using the PLA probe mouse and the PLA probe rabbit in control HAECs (c) and in IL-13 treated cells (d). (e,f) Proximity ligation assay using the centrin-2 and KI67 antibodies in control HAECs (e) and in IL-13 treated cells (f).



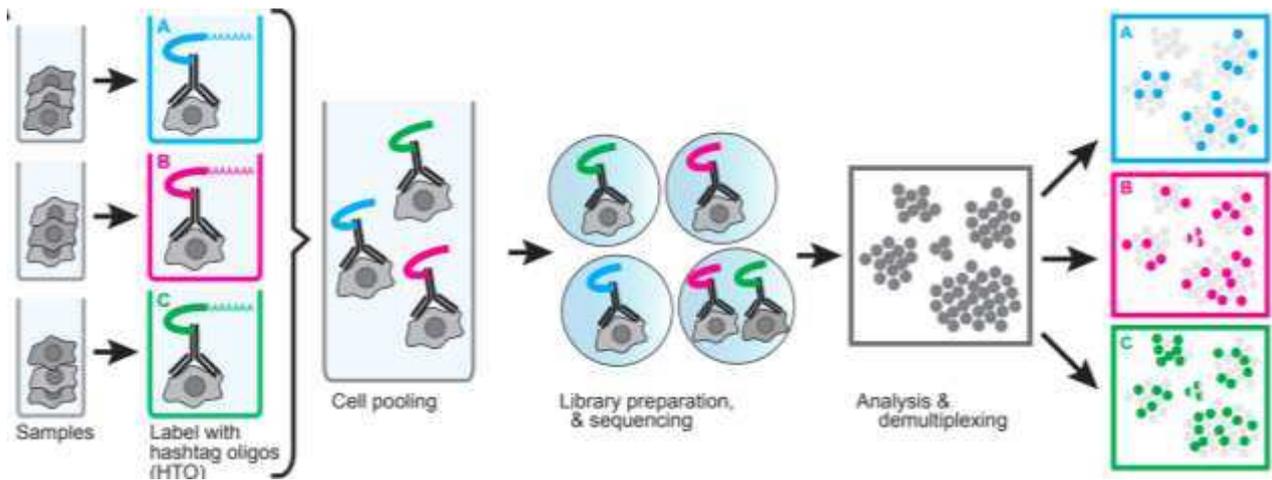
## CONCLUSION AND PERSPECTIVES

IL-13 induced the typical Th2 inflammatory phenotype in HAECs in vitro, showing an increased expression of genes related to inflammatory lung diseases. Thanks to our approach analysing the differential gene expression at the single cell level we could determine the specific expression of genes such as POSTN and ITLN1 in the cell populations that are expressed in the airways. Specific profile of expression of certain components of the signalling pathways Notch and Wnt indicate a role in the process of airway remodelling and goblet cell hyperplasia.

Future analysis of these datasets is still necessary. We will perform cell trajectory analysis as we showed for the normal differentiation study of the epithelium (Ruiz Garcia, Deprez et al. 2018, *submitted BioRxiv*), this analysis will show whether the SCs or GCs come from de-differentiation of the MCCs. Unfortunately, due to the low number of cells of these datasets the trajectory analysis could be more complicated.

Thanks to the new developments of different approaches coupling scRNAseq and cell lineage tracing we have the tools to infer cell lineage trajectories in human cultures primary cells, for instance, the use of “Linnaeus” (Spanjaard *et al.*, 2018), this technique consists in the use of CRISPR-Cas9 and the introduction of gDNA in the cells, targeting the CRISPR-Cas9 to this gDNA will create random cuts or “scars” that can be followed by sequencing, the combination of the different scars in each cell acts as unique barcodes, the construction of phylogenetic trees then is possible through the analysis of these barcodes and this is possible because the creation of scars is a process that takes long time so portions of the scars are shared between one cell and its progeny.

Also the development of “cell hashing” (Stoeckius *et al.*, 2017), this technique consists in the use of the barcoded antibodies. The aim of the technique is to target a determined cell type with a specific antibody, and this antibody is going to harbour a barcode DNA sequence that can be sequenced. Performing scRNAseq together with cell hashing would give us very different advantages such as: the possibility of pooling cell suspension from different donors, or different conditions but also cell hashing could enable us to distinguish the differentiation events that we hypothesized during treatment with IL-13 in HAECs.



**Fig. 68. Sample multiplexing using DNA-barcoded antibodies.** Schematic overview of sample multiplexing by cell hashing. Cells from different samples are incubated with DNA-barcoded antibodies recognizing ubiquitous cell surface proteins. Distinct barcodes (referred to as ‘hashtag’-oligos, HTO), on the antibodies allow pooling of multiple samples into one single cell RNA- sequencing experiment. After sequencing, cells can be classified to their sample of origin based on HTO levels. From (Stoeckius *et al.*, 2017)

In our laboratory we are also preparing a project to analyse by scRNAseq biopsies of patients with different grades of asthma. Asthma has a very wide spectrum of clinical phenotypes because of the different types of lung inflammation: non allergic eosinophilic response mainly caused by viruses, a Th2 response mediated through the production of inflammatory cytokines such as IL-4 or IL-13, caused by allergens, or a Th1-Th-17 neutrophilic inflammation (Papi *et al.*, 2018). Due to the variety of asthmatic phenotypes it is important to have an equal variety of therapies, however the different mechanism of the disease are not well described. With this project we expect to identify the sequence of events involved in airway remodelling in the disease. We will establish a cell atlas presented in pathological conditions as we did for healthy HAECs in our previous work (Ruiz Garcia, Deprez *et al.*, *submitted*).

## IX.3 MATERIALS AND METHODS

### **IL-13 treatment**

Primary culture of human airway epithelial cells (HAECs) was performed as described in (Ruiz Garcia, Deprez et al. 2018, *submitted BioRxiv*). For the present experiments, cells were cultured in BEGM médium (Lonza) until differentiation. Once reached the differentiation (30 days of culture) the cells were maintained in culture with or without adding the recombinant human IL-13 from PreproTech (cat: 200-13) at a concentration of 100ng/ml in the basal media of the cells, during a perios of 3 or 10 days depending on the experiment.

### **Cytospins**

Fully differentiated HAECs were dissociated by incubation with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in HBSS (Hanks' balanced salts) overnight at 4°C. Cells were gently detached from the Transwells® by pipetting and then transferred to a microtube. Cells were then cytocentrifuged at 72 g for 10 min onto SuperFrost™ Plus slides using a Shandon Cytospin™ 4 cytocentrifuge.

### **Immunostaining**

Cytospin™ slides or transwells were fixed for 10 min in 4% paraformaldehyde at room temperature or with methanol at -20°C during 10 minutes, and then permeabilized with 0.5% Triton X-100 in PBS for 10 min. Cells were blocked with 3% BSA in PBS for 30 min. For mounting on slides, Transwell® membranes were cut with a razor blade and mounted with ProLong™ Gold medium (Thermo Fisher Scientific).

The incubation with primary antibodies was carried out at 4°C overnight.

Primary antibodies: mouse monoclonal Acetylated Tubulin (1:500, Sigma-Aldrich clone 6-11B-1), mouse monoclonal MUC5AC (1:250, Abnova clone 45M1). Rabbit polyclonal Centrin-2 (1:250, Santa-Cruz clone sc-27793), Mouse monoclonal Centrin-2 (1:500, Millipore clone 20H5), Rabbit polyclonal KI67 (1:500, Abcam clone ab15580)

Secondary antibodies: Alexa Fluor 647 Goat anti-Mouse IgG2b (1:500, Fisher Scientific), Alexa Fluor 488 Goat anti-Mouse IgG1 (1:500, Fisher Scientific), Alexa Fluor 647 goat anti-rabbit (1:500; Thermo Fisher Scientific)

Incubation with secondary antibodies was carried out during 1h at room temperature. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI).

Images were acquired using the Olympus Fv10i or Leica sp5 confocal imaging systems.

### **Duolink proximity ligation assay**

To determine the interactions of proteins the Duolink Proximity ligation assay, Duolink® In Situ Red Starter Kit Mouse/Rabbit (Cat. DUO92101-1KT) was used.

The assay is based on the binding of PLA probes to the primary antibodies. If these are closer than 40 nm a signal is generated.

Cytospin slides with HACEs cultured with or without treatment were fixed with metanol blocked and incubated with the primary antibodies diluted in blocking solution were applied (centrinMS/centrinRB for positive control; centrinMS/KI67RB for the sample)

The slides were allowed to incubate overnight in a humidified chamber under continuous movement. The primary antibodies were tapped off and the slides were washed in PBS  $2 \times 10$  min. PLA probes detecting mouse and rabbit antibodies were diluted (1:5) in PBS.

For the incubation of the slides. Slides were incubated at  $37^{\circ}\text{C}$  in a humidified chamber for 1 h. The excess amount of PLA probes was tapped off and the samples were washed in Wash-Buffer A  $2 \times 10$  min. Duolink II Ligation stock (1:5) and Duolink Ligase (1:40) were diluted in high purity water and added to the slides. After incubation for 30 min in a humidified chamber at  $37^{\circ}\text{C}$ , the solution was tapped off and the slides were washed in Wash-Buffer A  $2 \times 5$  min. Duolink Polymerase (1:80) and Duolink Amplification and Detection stock (1:5) were diluted in high purity water and added to the samples. The slides were allowed to incubate for 100 min in a humidified chamber at  $37^{\circ}\text{C}$  in the dark. Finally, the slides were washed  $2 \times$  in Wash-Buffer B for 10 min and  $1 \times$  in  $0.01 \times$  Wash-Buffer B for 1 min. Nuclear staining was performed using DAPI. After drying for 30 min at room temperature in the dark they were mounted with Mowiol and stored at  $4^{\circ}\text{C}$  until examination. A negative control was also performed whit the adition of the PLA probes whitout the primary antibodies.

### Quantitative RT-PCR.

HAECs RNA was extracted using QIAzol Lysis Reagent from (QIAGEN Cat No./ID: 79306), followed by quantitative PCR on the Biomark system (Fluidigm) using SsoFast<sup>TM</sup> evaGreen® Supermix (Biorad). Primers:

Gene ID	Fw	Rv
CCP110	CTGACCAAACATGGCGAAAC	CTCCCAGGTTCAAGCAATTC
MKI67	CGACGGTCCCCACTTTCC	AATATTGCCTCCTGCTCATGG

## X. REFERENCES

---

- Abbassi, R., Johns, T. G., Kassiou, M. and Munoz, L. (2015) 'DYRK1A in neurodegeneration and cancer: Molecular basis and clinical implications', *Pharmacology & Therapeutics*, 151, pp. 87–98. doi: 10.1016/j.pharmthera.2015.03.004.
- Accurso, F. J., Rowe, S. M., Clancy, J. P., Boyle, M. P., Dunitz, J. M., Durie, P. R., Sagel, S. D., Hornick, D. B., Konstan, M. W., Donaldson, S. H., Moss, R. B., Pilewski, J. M., Rubenstein, R. C., Uluer, A. Z., Aitken, M. L., Freedman, S. D., Rose, L. M., Mayer-Hamblett, N., Dong, Q., Zha, J., Stone, A. J., Olson, E. R., Ordoñez, C. L., Campbell, P. W., Ashlock, M. A. and Ramsey, B. W. (2010) 'Effect of VX-770 in Persons with Cystic Fibrosis and the G551D- CFTR Mutation', *New England Journal of Medicine*. NIH Public Access, 363(21), pp. 1991–2003. doi: 10.1056/NEJMoa0909825.
- Adam, D., Roux-Delrieu, J., Luczka, E., Bonnomet, A., Lesage, J., M??rol, J. C., Polette, M., Ab??ly, M. and Coraux, C. (2015) 'Cystic fibrosis airway epithelium remodelling: Involvement of inflammation', *Journal of Pathology*, 235(3), pp. 408–419. doi: 10.1002/path.4471.
- Al Jord, A., Lemai, A., Delgehyr, N., Faucourt, M., Spassky, N. and Meunier, A. (2014) *Centriole amplification by mother and daughter centrioles differs in multiciliated cells*. doi: 10.1038/nature13770.
- Akamatsu, T., Azlina, A., Purwanti, N., Karabasil, M. R., Hasegawa, T., Yao, C. and Hosoi, K. (2009) 'Inhibition and transcriptional silencing of a subtilisin-like proprotein convertase, PACE4/SPC4, reduces the branching morphogenesis of and AQP5 expression in rat embryonic submandibular gland', *Developmental Biology*, 325(2), pp. 434–443. doi: 10.1016/j.ydbio.2008.10.015.
- Alevy, Y. G., Patel, A. C., Romero, A. G., Patel, D. A., Tucker, J., Roswit, W. T., Miller, C. A., Heier, R. F., Byers, D. E., Brett, T. J. and Holtzman, M. J. (2012) 'IL-13 – induced airway mucus production is attenuated by MAPK13 inhibition', 122(12). doi: 10.1172/JCI64896.that.
- Ambros, V. (2004) 'The functions of animal microRNAs', *Nature*, pp. 350–355. doi: 10.1038/nature02871.
- Arbi, M., Pefani, D. E., Taraviras, S. and Lygerou, Z. (2018) 'Controlling centriole numbers: Geminin family members as master regulators of centriole amplification and multiciliogenesis', *Chromosoma*, pp. 151–174. doi: 10.1007/s00412-017-0652-7.
- Arbi, M., Pefani, D., Kyrousi, C., Lalioti, M., Kalogeropoulou, A., Papanastasiou, A. D., Taraviras, S. and Lygerou, Z. (2016) 'GemC1 controls multiciliogenesis in the airway epithelium', *EMBO reports*, 17(3), pp. 400–413. doi: 10.15252/embr.201540882.

- Arguel, M. J., Lebrigand, K., Paquet, A., García, S. R., Zaragosi, L. E., Barbry, P. and Waldmann, R. (2017) 'A cost effective 5', selective single cell transcriptome profiling approach with improved UMI design', *Nucleic Acids Research*. Oxford University Press, 45(7), p. gkw1242. doi: 10.1093/nar/gkw1242.
- Arquint, C. and Nigg, E. A. (2016) 'The PLK4–STIL–SAS-6 module at the core of centriole duplication', *Biochemical Society Transactions*, 44(5), pp. 1253–1263. doi: 10.1042/BST20160116.
- Arquint, C., Sonnen, K. F., Stierhof, Y.-D. and Nigg, E. A. (2012) 'Cell-cycle-regulated expression of STIL controls centriole number in human cells', *Journal of Cell Science*, 125(5), pp. 1342–1352. doi: 10.1242/jcs.099887.
- Atherton, H. C., Jones, G. and Danahay, H. (2003) 'IL-13-induced changes in the goblet cell density of human bronchial epithelial cell cultures: MAP kinase and phosphatidylinositol 3-kinase regulation.', *American journal of physiology. Lung cellular and molecular physiology*, 285(3), pp. L730-9. doi: 10.1152/ajplung.00089.2003.
- Baarsma, H. A. and Königshoff, M. (2017) "'WNT-er is coming": WNT signalling in chronic lung diseases', *Thorax*, 72(8), pp. 746–759. doi: 10.1136/thoraxjnl-2016-209753.
- Balestrini, A., Cosentino, C., Errico, A., Garner, E. and Costanzo, V. (2010) 'GEMC1 is a TopBP1-interacting protein required for chromosomal DNA replication.', *Nat Cell Biol*. doi: 10.1038/ncb2050.
- Ball, A. J., Abrahamsson, A. E., Tyrberg, B., Itkin-Ansari, P. and Levine, F. (2007) 'HES6 reverses nuclear reprogramming of insulin-producing cells following cell fusion', *Biochemical and Biophysical Research Communications*, 355(2), pp. 331–337. doi: 10.1016/j.bbrc.2007.01.153.
- Bendall, S. C., Davis, K. L., Amir, E.-A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., Shenfeld, D. K., Nolan, G. P. and Pe'er, D. (2014) 'Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.', *Cell*, 157(3), pp. 714–25. doi: 10.1016/j.cell.2014.04.005.
- Bentley, J. K., Chen, Q., Hong, J. Y., Popova, A. P., Lei, J., Moore, B. B. and Hershenson, M. B. (2014) 'Periostin is required for maximal airways inflammation and hyperresponsiveness in mice', *Journal of Allergy and Clinical Immunology*, 134(6), pp. 1433–1442. doi: 10.1016/j.jaci.2014.05.029.
- Bertiaux, E., Mallet, A., Fort, C., Blisnick, T., Bonnefoy, S., Jung, J., Lemos, M., Marco, S., Vaughan, S., Tréput, S., Tinevez, J.-Y. and Bastin, P. (2018) 'Bidirectional intraflagellar transport is restricted to two sets of microtubule doublets in the trypanosome flagellum'. doi: 10.1083/jcb.201805030.

- Bertrand, B., Collet, C., Eloy, P. and Rombaux, P. (2000) 'Secondary ciliary dyskinesia in upper respiratory tract', *Acta oto-rhino-laryngologica Belgica*, pp. 309–316. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11082767> (Accessed: 16 October 2018).
- Bessho, Y. and Kageyama, R. (2003) 'Oscillations, clocks and segmentation', *Current Opinion in Genetics and Development*, pp. 379–384. doi: 10.1016/S0959-437X(03)00083-2.
- Bettencourt-Dias, M. and Glover, D. M. (2007) 'Centrosome biogenesis and function: Centrosomics brings new understanding', *Nature Reviews Molecular Cell Biology*, pp. 451–463. doi: 10.1038/nrm2180.
- Blake, D. C., Mikse, O. R., Freeman, W. M. and Herzog, C. R. (2010) 'FOXO3a elicits a pro-apoptotic transcription program and cellular response to human lung carcinogen nicotine-derived nitrosaminoketone (NNK)', *Lung Cancer*, 67(1), pp. 37–47. doi: 10.1016/j.lungcan.2009.03.013.
- Boers, J. E., Ambergen, A. W. and Thunnissen, F. B. (1998) 'Number and proliferation of clara cells in normal human airway epithelium.', *American journal of respiratory and critical care medicine*, 159(5 Pt 1), pp. 1585–91. doi: 10.1164/ajrccm.159.5.9806044.
- Boers, J. E., Ambergen, A. W. and Thunnissen, F. B. J. M. (1998) 'Number and proliferation of basal and parabasal cells in normal human airway epithelium', *American Journal of Respiratory and Critical Care Medicine*, 157(6 PART I), pp. 2000–2006. doi: 10.1164/ajrccm.157.6.9707011.
- Bonser, L. and Erle, D. (2017) 'Airway Mucus and Asthma: The Role of MUC5AC and MUC5B', *Journal of Clinical Medicine*, 6(12), p. 112. doi: 10.3390/jcm6120112.
- Bornens, M. (2002) 'Centrosome composition and microtubule anchoring mechanisms.', *Current opinion in cell biology*, 14(1), pp. 25–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11792541> (Accessed: 20 September 2018).
- Boucherat, O., Boczkowski, J., Jeannotte, L. and Delacourt, C. (2013) 'Cellular and molecular mechanisms of goblet cell metaplasia in the respiratory airways', *Experimental Lung Research*, 39, pp. 207–216. doi: 10.3109/01902148.2013.791733.
- Branchfield, K., Nantie, L., Verheyden, J. M., Sui, P., Wienhold, M. D. and Sun, X. (2016) 'Pulmonary neuroendocrine cells function as airway sensors to control lung immune response HHS Public Access', *Science*, 351(6274), pp. 707–710. doi: 10.1126/science.aad7969.
- Breeze, R. G. and Wheeldon, E. B. (1977) 'The Cells of the Pulmonary Airways<sup>1, 2</sup>', *American Review of Respiratory Disease*, 116(4), pp. 705–777. doi: 10.1164/arrd.1977.116.4.705.

- Breuer, R., Zajicek, G., Christensen, T. G., Lucey, E. C. and Snider, G. L. (1990) *Cell Kinetics of Normal Adult Hamster Bronchial Epithelium in the Steady State*, *Am. J. Respir. Cell Mol. Biol.* Available at: <https://www-atsjournals-org.gate2.inist.fr/doi/pdf/10.1165/ajrcmb/2.1.51> (Accessed: 4 August 2018).
- Brody, A. R., Hook, G. E., Cameron, G. S., Jetten, A. M., Butterick, C. J. and Nettesheim, P. (1987) 'The differentiation capacity of Clara cells isolated from the lungs of rabbits.', *Laboratory investigation; a journal of technical methods and pathology*, 57(2), pp. 219–29. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3613528> (Accessed: 4 August 2018).
- Brody, S. L., Yan, X. H., Wuerffel, M. K., Song, S. K. and Shapiro, S. D. (2000) 'Ciliogenesis and left-right axis defects in forkhead factor HFH-4-null mice.', *American journal of respiratory cell and molecular biology*, 23(1), pp. 45–51. doi: 10.1165/ajrcmb.23.1.4070.
- Burgel, P.-R. and Nadel, J. A. (2008) 'Epidermal growth factor receptor-mediated innate immune responses and their roles in airway diseases.', *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*, 32(4), pp. 1068–1081. doi: 10.1183/09031936.00172007.
- Burgel, P. R., Escudier, E., Coste, A., Dao-Pick, T., Ueki, I. F., Takeyama, K., Shim, J. J., Murr, a H. and Nadel, J. a (2000) 'Relation of epidermal growth factor receptor expression to goblet cell hyperplasia in nasal polyps.', *The Journal of allergy and clinical immunology*, 106(4), pp. 705–712. doi: 10.1067/mai.2000.109823.
- Button, B., Anderson, W. H. and Boucher, R. C. (2016) 'Mucus hyperconcentration as a unifying aspect of the chronic bronchitic phenotype', in *Annals of the American Thoracic Society*, pp. S156–S162. doi: 10.1513/AnnalsATS.201507-455KV.
- Campbell, E. P., Quigley, I. K. and Kintner, C. (2016) 'Foxn4 promotes gene expression required for the formation of multiple motile cilia.', *Development (Cambridge, England)*. Company of Biologists, 143(24), pp. 4654–4664. doi: 10.1242/dev.143859.
- Chanez, P., Contin-Bordes, C., Garcia, G., Verkindre, C., Didier, A., De Blay, F., Tunon De Lara, M., Blanco, P., Moreau, J.-F., Robinson, P., Bourdeix, I., Trunet, P., Le Gros, V., Humbert, M. and Molimard, M. (2010) 'Omalizumab-induced decrease of Fc3RI expression in patients with severe allergic asthma', *Respiratory Medicine*, 104, pp. 1608–1617. doi: 10.1016/j.rmed.2010.07.011.
- Chen, G., Korfhagen, T. R., Xu, Y., Kitzmiller, J., Wert, S. E., Maeda, Y., Gregorieff, A., Clevers, H. and Whitsett, J. A. (2009) 'SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production', *Journal of Clinical Investigation*. doi: 10.1172/JCI39731.

- Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. and Poidinger, M. (2016) 'Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development.', *Nature communications*, 7, p. 11988. doi: 10.1038/ncomms11988.
- Chevalier, B., Adamiok, A., Mercey, O., Revinski, D. R., Zaragosi, L. E., Pasini, A., Kodjabachian, L., Barbry, P. and Marcet, B. (2015) 'MiR-34/449 control apical actin network formation during multiciliogenesis through small GTPase pathways', *Nature Communications*, 6. doi: 10.1038/ncomms9386.
- Chih, B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L. G., Hass, P. E., Sandoval, W. and Peterson, A. S. (2012) 'A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain', *Nature Cell Biology*. Nature Publishing Group, 14(1), pp. 61–72. doi: 10.1038/ncb2410.
- Choi, J. H., Nahm, D. H., Kim, S. H., Kim, Y. S., Suh, C. H., Park, H. S. and Ahn, S. W. (2004) 'Increased levels of IgG to cytokeratin 19 in sera of patients with toluene diisocyanate-induced asthma', *Annals of Allergy, Asthma and Immunology*, 93(3), pp. 293–298. doi: 10.1016/S1081-1206(10)61504-9.
- Chung, K. F. (2015) 'Targeting the interleukin pathway in the treatment of asthma', *The Lancet*, 386(9998), pp. 1086–1096. doi: 10.1016/S0140-6736(15)00157-9.
- Chung, M.-I., Peyrot, S. M., LeBoeuf, S., Park, T. J., McGary, K. L., Marcotte, E. M. and Wallingford, J. B. (2012) 'RFX2 is broadly required for ciliogenesis during vertebrate development.', *Developmental biology*, 363(1), pp. 155–65. doi: 10.1016/j.ydbio.2011.12.029.
- Cibois, M., Luxardi, G., Chevalier, B., Thome, V., Mercey, O., Zaragosi, L.-E., Barbry, P., Pasini, A., Marcet, B. and Kodjabachian, L. (2015) 'BMP signalling controls the construction of vertebrate mucociliary epithelia', *Development*, 142(13), pp. 2352–2363. doi: 10.1242/dev.118679.
- Cohen, M., Giladi, A., Gorki, A.-D., Solodkin, D. G., Zada, M., Hladik, A., Miklosi, A., Salame, T.-M., Halpern, K. B., David, E., Itzkovitz, S., Harkany, T., Knapp, S. and Amit, I. (2018) 'Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting', *Cell*. doi: 10.1016/j.cell.2018.09.009.
- Cohn, L. (2006) 'Mucus in chronic airway diseases: sorting out the sticky details.', *The Journal of clinical investigation*, 116(2), pp. 306–8. doi: 10.1172/JCI27690.
- Confalonieri, M., Buratti, E., Grassi, G., Bussani, R., Chilosi, M., Farra, R., Abrami, M., Stuani, C., Salton, F., Ficial, M., Confalonieri, P., Zandonà, L. and Romano, M. (2017) 'Keratin14 mRNA expression in human pneumocytes during quiescence, repair and disease', *PLoS ONE*, 12(2). doi: 10.1371/journal.pone.0172130.
- Czarnecki, P. G. and Shah, J. V (2012) 'The ciliary transition zone: From morphology and molecules to medicine', *Trends in Cell Biology*, pp. 201–210. doi: 10.1016/j.tcb.2012.02.001.

- Damerla, R. R., Cui, C., Gabriel, G. C., Liu, X., Craige, B., Gibbs, B. C., Francis, R., Li, Y., Chatterjee, B., San Agustin, J. T., Eguether, T., Subramanian, R., Witman, G. B., Michaud, J. L., Pazour, G. J. and Lo, C. W. (2015) 'Novel Jbts17 mutant mouse model of Joubert syndrome with cilia transition zone defects and cerebellar and other ciliopathy related anomalies.', *Human molecular genetics*. Oxford University Press, 24(14), pp. 3994–4005. doi: 10.1093/hmg/ddv137.
- Danahay, H., Pessotti, A. D., Coote, J., Montgomery, B. E., Xia, D., Wilson, A., Yang, H., Wang, Z., Bevan, L., Thomas, C., Petit, S., London, A., LeMotte, P., Doelemeyer, A., Vélez-Reyes, G. L., Bernasconi, P., Fryer, C. J., Edwards, M., Capodiecici, P., Chen, A., Hild, M. and Jaffe, A. B. (2015) 'Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in the lung', *Cell Reports*, 10(2), pp. 239–252. doi: 10.1016/j.celrep.2014.12.017.
- Dang, T. P., Eichenberger, S., Gonzalez, A., Olson, S. and Carbone, D. P. (2003) 'Constitutive activation of Notch3 inhibits terminal epithelial differentiation in lungs of transgenic mice', *Oncogene*, 22(13), pp. 1988–1997. doi: 10.1038/sj.onc.1206230.
- Danielian, P. S., Bender Kim, C., Caron, A., Vasile, E., Bronson, R. and Lees, J. (2007) 'E2f4 is required for normal development of the airway epithelium', *Developmental biology*, 305(2), pp. 564–576. doi: 10.1115/1.3071969.Automating.
- Danielian, P. S., Hess, R. A., Lees, J. A., Danielian, P. S., Hess, R. A. and Ef, J. A. L. (2016) 'E2f4 and E2f5 are essential for the development of the male reproductive system', 4101(February). doi: 10.1080/15384101.2015.1121350.
- Daniely, Y., Liao, G., Dixon, D., Linnoila, R. I., Lori, A., Randell, S. H., Oren, M. and Jetten, A. M. (2004) 'Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium', *American Journal of Physiology-Cell Physiology*, 287(1), pp. C171–C181. doi: 10.1152/ajpcell.00226.2003.
- Davies, J. C., Moskowitz, S. M., Brown, C., Horsley, A., Mall, M. A., McKone, E. F., Plant, B. J., Prais, D., Ramsey, B. W., Taylor-Cousar, J. L., Tullis, E., Uluer, A., McKee, C. M., Robertson, S., Shilling, R. A., Simard, C., Van Goor, F., Waltz, D., Xuan, F., Young, T. and Rowe, S. M. (2018) 'VX-659–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles', *New England Journal of Medicine*, 379(17), pp. 1599–1611. doi: 10.1056/NEJMoa1807119.
- Delattre, M. (2004) 'The arithmetic of centrosome biogenesis', *Journal of Cell Science*, 117(9), pp. 1619–1630. doi: 10.1242/jcs.01128.

Delous, M., Baala, L., Salomon, R., Laclef, C., Vierkotten, J., Tory, K., Golzio, C., Lacoste, T., Besse, L., Ozilou, C., Moutkine, I., Hellman, N. E., Anselme, I., Silbermann, F., Vesque, C., Gerhardt, C., Rattenberry, E., Wolf, M. T. F., Gubler, M. C., Martinovic, J., Encha-Razavi, F., Boddaert, N., Gonzales, M., Macher, M. A., Nivet, H., Champion, G., Berthéléme, J. P., Niaudet, P., McDonald, F., Hildebrandt, F., Johnson, C. A., Vekemans, M., Antignac, C., Rütger, U., Schneider-Maunoury, S., Attié-Bitach, T. and Saunier, S. (2007) 'The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome', *Nature Genetics*. Nature Publishing Group, 39(7), pp. 875–881. doi: 10.1038/ng2039.

Didon, L., Zwick, R. K., Chao, I. W., Walters, M. S., Wang, R., Hackett, N. R. and Crystal, R. G. (2013) 'RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium.', *Respiratory research*, 14, p. 70. doi: 10.1186/1465-9921-14-70.

Divekar, R. and Kita, H. (2015) 'Recent advances in epithelium-derived cytokines (IL-33, IL-25, and thymic stromal lymphopoietin) and allergic inflammation', *Current Opinion in Allergy and Clinical Immunology*, pp. 98–103. doi: 10.1097/ACI.0000000000000133.

Domenighetti, A. A., Mathewson, M. A., Pichika, R., Sibley, L. A., Zhao, L., Chambers, H. G., Richard, X. and Lieber, L. (2018) 'Loss of myogenic potential and fusion capacity of muscle stem cells isolated from contractured muscle in children with cerebral palsy', *Am J Physiol Cell Physiol*, 315, pp. 247–257. doi: 10.1152/ajpcell.

Dowdle, W. E., Robinson, J. F., Kneist, A., Sirerol-Piquer, M. S., Frints, S. G. M., Corbit, K. C., Zaghoul, N. A., Zaghoul, N. A., van Lijnschoten, G., Mulders, L., Verver, D. E., Zerres, K., Reed, R. R., Attié-Bitach, T., Johnson, C. A., García-Verdugo, J. M., Katsanis, N., Bergmann, C. and Reiter, J. F. (2011) 'Disruption of a ciliary B9 protein complex causes Meckel syndrome.', *American journal of human genetics*. Elsevier, 89(1), pp. 94–110. doi: 10.1016/j.ajhg.2011.06.003.

Dubruille, R. (2002) 'Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation', *Development*, 129(23), pp. 5487–5498. doi: 10.1242/dev.00148.

Dupuit, F., Kälin, N., Brézillon, S., Hinrasky, J., Tümmeler, B. and Puchelle, E. (1995) 'CFTR and differentiation markers expression in non-CF and  $\Delta F$  508 homozygous CF nasal epithelium', *Journal of Clinical Investigation*, 96(3), pp. 1601–1611. doi: 10.1172/JCI118199.

Durham, A. L., McLaren, A., Hayes, B. P., Caramori, G., Clayton, C. L., Barnes, P. J., Chung, K. F. and Adcock, I. M. (2013) 'Regulation of Wnt4 in chronic obstructive pulmonary disease.', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. The Federation of American Societies for Experimental Biology, 27(6), pp. 2367–81. doi: 10.1096/fj.12-217083.

- Elliott, K. H. and Brugmann, S. A. (2018) 'Sending mixed signals: Cilia-dependent signaling during development and disease.', *Developmental biology*. doi: 10.1016/j.ydbio.2018.03.007.
- Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002) 'Stochastic gene expression in a single cell', *Science*, 297(5584), pp. 1183–1186. doi: 10.1126/science.1070919.
- Engelhardt, J. F., Yankaskas, J. R., Ernst, S. A., Yang, Y., Marino, C. R., Boucher, R. C., Cohn, J. A. and Wilson, J. M. (1992) 'Submucosal glands are the predominant site of CFTR expression in the human bronchus', *Nature Genetics*. Nature Publishing Group, 2(3), pp. 240–248. doi: 10.1038/ng1192-240.
- Esaki, M., Hoshijima, K., Kobayashi, S., Fukuda, H., Kawakami, K. and Hirose, S. (2006) 'Visualization in zebrafish larvae of Na<sup>+</sup> uptake in mitochondria-rich cells whose differentiation is dependent on foxi3a', *AJP: Regulatory, Integrative and Comparative Physiology*, 292(1), pp. R470–R480. doi: 10.1152/ajpregu.00200.2006.
- Esaki, M., Hoshijima, K., Nakamura, N., Munakata, K., Tanaka, M., Ookata, K., Asakawa, K., Kawakami, K., Wang, W., Weinberg, E. S. and Hirose, S. (2009) 'Mechanism of development of ionocytes rich in vacuolar-type H<sup>+</sup>-ATPase in the skin of zebrafish larvae', *Developmental Biology*, 329(1), pp. 116–129. doi: 10.1016/j.ydbio.2009.02.026.
- Eszter K. Vadar and Tim Stearns (2007) 'Molecular characterization of centriole assembly in ciliated epithelial cells', *The Journal of Cell Biology*, 178(1), pp. 31–42. doi: 10.1083/jcb.200703064.
- Evans, C. M., Raclawska, D. S., Ttofali, F., Liptzin, D. R., Fletcher, A. A., Harper, D. N., McGing, M. A., McElwee, M. M., Williams, O. W., Sanchez, E., Roy, M. G., Kindrachuk, K. N., Wynn, T. A., Eltzhig, H. K., Blackburn, M. R., Tuvim, M. J., Janssen, W. J., Schwartz, D. A. and Dickey, B. F. (2015) 'The polymeric mucin Muc5ac is required for allergic airway hyperreactivity', *Nature Communications*, 6. doi: 10.1038/ncomms7281.
- Evans, C. M., Williams, O. W., Tuvim, M. J., Nigam, R., Mixides, G. P., Blackburn, M. R., Demayo, F. J., Burns, A. R., Smith, C., Reynolds, S. D., Stripp, B. R. and Dickey, B. F. (2004) 'Mucin is produced by clara cells in the proximal airways of antigen-challenged mice', *American Journal of Respiratory Cell and Molecular Biology*, 31(4), pp. 382–394. doi: 10.1165/rcmb.2004-0060OC.
- Evans, M. J., Cox, R. A., Shami, S. G., Wilson, B. and Plopper, C. G. (1989) *The Role of Basal Cells in Attachment of Columnar Cells to the Basal Lamina of the Trachea*, *Am. J. Respir. Cell Mol. Biol.* Available at: <https://www-atsjournals-org.gate2.inist.fr/doi/pdf/10.1165/ajrcmb/1.6.463> (Accessed: 4 August 2018).

- Evans, M. J. and Plopper, C. G. (1988) 'The role of basal cells in adhesion of columnar epithelium to airway basement membrane.', *The American review of respiratory disease*, 138(2), pp. 481–3. doi: 10.1164/ajrccm/138.2.481.
- Evans, M. J., Shami, S. G., Cabral-Anderson, L. J. and Dekker, N. P. (1986) 'Role of nonciliated cells in renewal of the bronchial epithelium of rats exposed to NO<sub>2</sub>.', *The American journal of pathology*, 123(1), pp. 126–33. Available at: <https://www.ncbi.nlm.nih.gov/gate2.inist.fr/pmc/articles/PMC1888156/pdf/amjpathol00157-0132.pdf> (Accessed: 24 May 2018).
- Evans, M. J., Van Winkle, L. S., Fanucchi, M. V and Plopper, C. G. (2001) 'Cellular and molecular characteristics of basal cells in airway epithelium.', *Exp Lung Res*, 27(5), pp. 401–415. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11480582> (Accessed: 21 June 2018).
- Fan, H. C., Fu, G. K. and Fodor, S. P. A. (2015) 'Combinatorial labeling of single cells for gene expression cytometry', *Science*, 347(6222), pp. 1258367–1258367. doi: 10.1126/science.1258367.
- Fan, H. C., Fu, G. K. and Fodor, S. P. A. (2015) 'Expression profiling. Combinatorial labeling of single cells for gene expression cytometry.', *Science (New York, N.Y.)*, 347(6222), p. 1258367. doi: 10.1126/science.1258367.
- Farlie, P., Reid, C., Wilcox, S., Peeters, J., Reed, G. and Newgreen, D. (2001) 'Ypel1: a novel nuclear protein that induces an epithelial-like morphology in fibroblasts', *Genes to Cells*. Wiley/Blackwell (10.1111), 6(7), pp. 619–629. doi: 10.1046/j.1365-2443.2001.00445.x.
- Ficial, M., Antonaglia, C., Chilosi, M., Santagiuliana, M., Tahseen, A. O., Confalonieri, D., Zandoña, L., Bussani, R. and Confalonieri, M. (2014) 'Keratin-14 expression in pneumocytes as a marker of lung regeneration/repair during diffuse alveolar damage', *American Journal of Respiratory and Critical Care Medicine*, pp. 1142–1145. doi: 10.1164/rccm.201312-2134LE.
- Fletcher, R. B., Das, D., Gadye, L., Street, K. N., Baudhuin, A., Wagner, A., Cole, M. B., Flores, Q., Choi, Y. G., Yosef, N., Purdom, E., Dudoit, S., Risso, D. and Ngai, J. (2017) 'Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution', *Cell Stem Cell*, 20(6), p. 817–830.e8. doi: 10.1016/j.stem.2017.04.003.
- Fort, C. and Bastin, P. (2014) 'Élongation de l'axonème et dynamique du transport intraflagellaire', *Medecine/Sciences*, 30(11), pp. 955–961. doi: 10.1051/medsci/20143011008.
- Fort, C., Bonnefoy, S., Kohl, L. and Bastin, P. (2016) 'Intraflagellar transport is required for the maintenance of the trypanosome flagellum composition but not its length.', *Journal of cell science*. The Company of Biologists Ltd, 129(15), pp. 3026–41. doi: 10.1242/jcs.188227.

- Fortini, M. E. (2012) 'Introduction—Notch in development and disease', *Seminars in Cell and Developmental Biology*, 23, pp. 419–420. doi: 10.1016/j.semcd.2012.03.001.
- Froidure, A., Mouthuy, J., Durham, S. R., Chanez, P., Sibille, Y. and Pilette, C. (2016) 'Asthma phenotypes and IgE responses', *European Respiratory Journal*, 47(1), pp. 304–319. doi: 10.1183/13993003.01824-2014.
- Fu, G. K., Hu, J., Wang, P.-H. and Fodor, S. P. A. (2011) 'Counting individual DNA molecules by the stochastic attachment of diverse labels', *Proceedings of the National Academy of Sciences*. PNAS, 108(22), pp. 9026–9031. doi: 10.1073/pnas.1017621108.
- Funk, M. C., Bera, A. N., Menchen, T., Kualess, G., Thriene, K., Lienkamp, S. S., Dengjel, J., Omran, H., Frank, M. and Arnold, S. J. (2015) 'Cyclin O (Ccn0) functions during deuterosome-mediated centriole amplification of multiciliated cells.', *The EMBO journal*, 34, pp. 1078–89. doi: 10.15252/embj.201490805.
- Furchtgott, L. A., Melton, S., Menon, V. and Ramanathan, S. (2017) 'Discovering sparse transcription factor codes for cell states and state transitions during development', *eLife*, 6. doi: 10.7554/eLife.20488.
- Gadye, L., Das, D., Sanchez, M. A., Street, K., Baudhuin, A., Wagner, A., Cole, M. B., Choi, Y. G., Yosef, N., Purdom, E., Dudoit, S., Risso, D., Ngai, J. and Fletcher, R. B. (2017) 'Injury Activates Transient Olfactory Stem Cell States with Diverse Lineage Capacities', *Cell Stem Cell*, 21(6), p. 775–790.e9. doi: 10.1016/j.stem.2017.10.014.
- Garcia-Gonzalo, F. R., Corbit, K. C., Sirerol-Piquer, M. S., Ramaswami, G., Otto, E. A., Noriega, T. R., Seol, A. D., Robinson, J. F., Bennett, C. L., Josifova, D. J., García-Verdugo, J. M., Katsanis, N., Hildebrandt, F. and Reiter, J. F. (2011) 'A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition.', *Nature genetics*, 43(8), pp. 776–84. doi: 10.1038/ng.891.
- Gauvreau, G. M., Arm, J. P., Boulet, L.-P., Leigh, R., Cockcroft, D. W., Davis, B. E., Mayers, I., FitzGerald, J. M., Dahlen, B., Killian, K. J., Laviolette, M., Carlsten, C., Lazarinis, N., Watson, R. M., Milot, J., Swystun, V., Bowen, M., Hui, L., Lantz, A.-S., Meiser, K., Maahs, S., Lowe, P. J., Skerjanec, A., Drollmann, A. and O'Byrne, P. M. (2016) 'Efficacy and safety of multiple doses of QGE031 (ligelizumab) versus omalizumab and placebo in inhibiting allergen-induced early asthmatic responses', *Journal of Allergy and Clinical Immunology*, 138(4), pp. 1051–1059. doi: 10.1016/j.jaci.2016.02.027.
- Gerlach, G. F. and Wingert, R. A. (2014) 'Zebrafish pronephros tubulogenesis and epithelial identity maintenance are reliant on the polarity proteins Prkc iota and zeta.', *Developmental biology*. NIH Public Access, 396(2), pp. 183–200. doi: 10.1016/j.ydbio.2014.08.038.

- Gerovac, B. J. and Fregien, N. L. (2016) 'IL-13 Inhibits Multicilin Expression and Ciliogenesis via Janus Kinase/Signal Transducer and Activator of Transcription Independently of Notch Cleavage.', *American journal of respiratory cell and molecular biology*, 54(4), pp. 554–61. doi: 10.1165/rcmb.2015-0227OC.
- Ghosh, M., Ahmad, S., Jian, A., Li, B., Smith, R. W., Helm, K. M., Seibold, M. A., Groshong, S. D., White, C. W. and Reynolds, S. D. (2013) 'Human tracheobronchial basal cells: Normal versus remodeling/repairing phenotypes in vivo and in vitro', *American Journal of Respiratory Cell and Molecular Biology*, 49(6), pp. 1127–1134. doi: 10.1165/rcmb.2013-0049OC.
- Gierahn, T. M., Wadsworth, M. H., Hughes, T. K., Bryson, B. D., Butler, A., Satija, R., Fortune, S., Love, J. C., Shalek, A. K. and Shalek, A. K. (2017) 'Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput.', *Nature methods*. NIH Public Access, 14(4), pp. 395–398. doi: 10.1038/nmeth.4179.
- Giovannini-Chami, L., Marcet, B., Moreilhon, C., Chevalier, B., Illie, M. I., Lebrigand, K., Robbe-Sermesant, K., Bourrier, T., Michiels, J. F., Mari, B., Crénesse, D., Hofman, P., De Blic, J., Castillo, L., Albertini, M. and Barbry, P. (2012) 'Distinct epithelial gene expression phenotypes in childhood respiratory allergy', *European Respiratory Journal*, 39(5), pp. 1197–1205. doi: 10.1183/09031936.00070511.
- Goetz, S. C. and Anderson, K. V (2010) 'The primary cilium: A signalling centre during vertebrate development', *Nature Reviews Genetics*, pp. 331–344. doi: 10.1038/nrg2774.
- Gomperts, B. N. (2004) 'Foxj1 regulates basal body anchoring to the cytoskeleton of ciliated pulmonary epithelial cells', *Journal of Cell Science*, 117(8), pp. 1329–1337. doi: 10.1242/jcs.00978.
- Gomperts, B. N., Kim, L. J., Flaherty, S. A. and Hackett, B. P. (2007) 'IL-13 regulates cilia loss and foxj1 expression in human airway epithelium.', *American journal of respiratory cell and molecular biology*, 37(3), pp. 339–46. doi: 10.1165/rcmb.2006-0400OC.
- Gonçalves, J. and Pelletier, L. (2017) 'The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate', *Mol. Cells*, 40(4), p. 243. doi: 10.14348/molcells.2017.0054.
- Gonzalez-Cano, L., Fuertes-Alvarez, † S, Robledinos-Anton, N., Bizy, A., Villena-Cortes, A., Fari, I., Marques, M. M. and Marin, M. C. (2015) 'p73 is Required for Ependymal Cell Maturation and Neurogenic SVZ Cytoarchitecture', *Inc. Develop Neurobiol.* Wiley Periodicals, 76, pp. 730–747. doi: 10.1002/dneu.22356.
- Górska, K., Maskey-Warzęchowska, M., Nejman-Gryz, P., Korczyński, P., Prochorec-Sobieszek, M. and Krenke, R. (2016) 'Comparative study of periostin expression in different respiratory samples in patients with asthma and chronic obstructive pulmonary disease', *Polish Archives of Internal Medicine*, 126(3), pp. 124–137. doi: 10.20452/pamw.3299.

- Grainge, C. L., Lau, L. C. K., Ward, J. A., Dulay, V., Lahiff, G., Wilson, S., Holgate, S., Davies, D. E. and Howarth, P. H. (2011) *Effect of Bronchoconstriction on Airway Remodeling in Asthma, n engl j med*. Available at: <https://www-nejm-org.gate2.inist.fr/doi/pdf/10.1056/NEJMoa1014350> (Accessed: 7 September 2018).
- Gras, D., Martinez-Anton, A., Bourdin, A., Garulli, C., De Senneville, L., Vachier, I., Vitte, J. and Chanez, P. (2017) 'Human bronchial epithelium orchestrates dendritic cell activation in severe asthma', *European Respiratory Journal*, 49(3). doi: 10.1183/13993003.02399-2016.
- Grubb, B. R., Vick, R. N. and Boucher, R. C. (1994) 'Hyperabsorption of Na<sup>+</sup> and raised Ca(2<sup>+</sup>)-mediated Cl<sup>-</sup> secretion in nasal epithelia of CF mice.', *The American journal of physiology*, 266(5 Pt 1), pp. C1478-83. doi: 10.1152/ajpcell.1994.266.5.C1478.
- Grün, D., Muraro, M. J., Boisset, J.-C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., de Koning, E. J. P. and van Oudenaarden, A. (2016) 'De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data.', *Cell stem cell*, 19(2), pp. 266–277. doi: 10.1016/j.stem.2016.05.010.
- Guha, A., Vasconcelos, M., Cai, Y., Yoneda, M., Hinds, A., Qian, J., Li, G., Dickel, L., Johnson, J. E., Kimura, S., Guo, J., McMahon, J., McMahon, A. P. and Cardoso, W. V. (2012) 'Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways', *Proceedings of the National Academy of Sciences*, 109(31), pp. 12592–12597. doi: 10.1073/pnas.1204710109.
- Guha, A., Vasconcelos, M., Zhao, R., Gower, A. C., Rajagopal, J. and Cardoso, W. V. (2014) 'Analysis of notch signaling-dependent gene expression in developing airways reveals diversity of clara cells', *PLoS ONE*, 9(2), p. 88848. doi: 10.1371/journal.pone.0088848.
- Guichard, P., Chrétien, D., Marco, S. and Tassin, A. M. (2010) 'Procentriole assembly revealed by cryo-electron tomography', *EMBO Journal*, 29(9), pp. 1565–1572. doi: 10.1038/emboj.2010.45.
- Guo, M., Bao, E. L., Wagner, M., Whitsett, J. A. and Xu, Y. (2017) 'SLICE: determining cell differentiation and lineage based on single cell entropy.', *Nucleic acids research*, 45(7), p. e54. doi: 10.1093/nar/gkw1278.
- Gupta, G. D., Coyaud, É., Gonçalves, J., Mojarad, B. A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J. M., Cheung, S. W. T., Bashkurov, M., Hasegan, M., Knight, J. D., Lin, Z.-Y., Schueler, M., Hildebrandt, F., Moffat, J., Gingras, A.-C., Raught, B. and Pelletier, L. (2015) 'A dynamic protein interaction landscape of the human centrosome-cilium interface HHS Public Access', *Cell*, 163(6), pp. 1484–1499. doi: 10.1016/j.cell.2015.10.065.

- Guruharsha, K. G., Kankel, M. W. and Artavanis-Tsakonas, S. (2012) 'The Notch signalling system: Recent insights into the complexity of a conserved pathway', *Nature Reviews Genetics*, pp. 654–666. doi: 10.1038/nrg3272.
- Guseh, J. S., Bores, S. A., Stanger, B. Z., Zhou, Q., Anderson, W. J., Melton, D. A. and Rajagopal, J. (2009a) 'Notch signaling promotes airway mucous metaplasia and inhibits alveolar development.', *Development (Cambridge, England)*. Company of Biologists, 136(10), pp. 1751–9. doi: 10.1242/dev.029249.
- Guseh, J. S., Bores, S. A., Stanger, B. Z., Zhou, Q., Anderson, W. J., Melton, D. A. and Rajagopal, J. (2009b) 'Notch signaling promotes airway mucous metaplasia and inhibits alveolar development', *Development*. Company of Biologists, 136(10), pp. 1751–1759. doi: 10.1242/dev.029249.
- Habedanck, R., Stierhof, Y.-D., Wilkinson, C. J. and Nigg, E. A. (2005) 'the Polo kinase Plk4 functions in centriole duplication The human Polo-like kinase 1 (PLK1) and its functional homologues that are present in other eukaryotes have multiple, crucial roles in meiotic and mitotic cell division', 7(11). doi: 10.1038/ncb1320.
- Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S. R., Aguet, F., Gelfand, E., Ardlie, K., Weitz, D. A., Rozenblatt-Rosen, O., Zhang, F. and Regev, A. (2017) 'Massively parallel single-nucleus RNA-seq with DroNc-seq.', *Nature methods*, 14(10), pp. 955–958. doi: 10.1038/nmeth.4407.
- Hackett, N. R., Shaykhiev, R., Walters, M. S., Wang, R., Zwick, R. K., Ferris, B., Witover, B., Salit, J. and Crystal, R. G. (2011) 'The Human Airway Epithelial Basal Cell Transcriptome', *PLoS ONE*, 6(5). doi: 10.1371/journal.pone.0018378.
- Hajj, R., Baranek, T., Le Naour, R., Lesimple, P., Puchelle, E. and Coraux, C. (2007) 'Basal Cells of the Human Adult Airway Surface Epithelium Retain Transit-Amplifying Cell Properties', *STEM CELLS*. Wiley-Blackwell, 25(1), pp. 139–148. doi: 10.1634/stemcells.2006-0288.
- Hajj, R., Lesimple, P., Nawrocki-Raby, B., Birembaut, P., Puchelle, E. and Coraux, C. (2007) 'Human airway surface epithelial regeneration is delayed and abnormal in cystic fibrosis', *The Journal of Pathology*, 211(3), pp. 340–350. doi: 10.1002/path.2118.
- Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K. J., Rozenblatt-Rosen, O., Dor, Y., Regev, A. and Yanai, I. (2016) 'CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.', *Genome biology*. BioMed Central, 17, p. 77. doi: 10.1186/s13059-016-0938-8.
- Hashimshony, T., Wagner, F., Sher, N. and Yanai, I. (2012) 'CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification', *Cell Reports*, 2(3), pp. 666–673. doi: 10.1016/j.celrep.2012.08.003.

- Hayashi, T., Ishii, A., Nakai, S. and Hasegawa, K. (2004) 'Ultrastructure of goblet-cell metaplasia from Clara cell in the allergic asthmatic airway inflammation in a mouse model of asthma in vivo', *Virchows Archiv*, 444(1), pp. 66–73. doi: 10.1007/s00428-003-0926-8.
- He, L. and Hannon, G. J. (2004) 'MicroRNAs: Small RNAs with a big role in gene regulation', *Nature Reviews Genetics*, pp. 522–531. doi: 10.1038/nrg1379.
- Heijink, I. H., De Bruin, H. G., Van Den Berge, M., Bennink, L. J. C., Brandenburg, S. M., Gosens, R., Van Oosterhout, A. J. and Postma, D. S. (2012) 'Role of aberrant WNT signalling in the airway epithelial response to cigarette smoke in chronic obstructive pulmonary disease'. doi: 10.1136/thoraxjnl-2012-201667.
- Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A. and Mann, M. (2015) 'A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances', *Cell*, 163(3), pp. 712–723. doi: 10.1016/j.cell.2015.09.053.
- Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J. X., Kreisberg, R., Kauffman, S. A., Huang, S. and Shmulevich, I. (2013) 'Gene-pair expression signatures reveal lineage control', *Nature Methods*, 10(6), pp. 577–583. doi: 10.1038/nmeth.2445.
- Hogan, B. L. M., Barkauskas, C. E., Chapman, H. A., Epstein, J. A., Jain, R., Hsia, C. C. W., Niklason, L., Calle, E., Le, A., Randell, S. H., Rock, J., Snitow, M., Krummel, M., Stripp, B. R., Vu, T., White, E. S., Whitsett, J. A. and Morrissey, E. E. (2014) 'Repair and Regeneration of the Respiratory System: Complexity, Plasticity, and Mechanisms of Lung Stem Cell Function', *Cell Stem Cell*. Elsevier Inc., 15(2), pp. 123–138. doi: 10.1016/j.stem.2014.07.012.
- den Hollander, A. I., Koenekoop, R. K., Mohamed, M. D., Arts, H. H., Boldt, K., Towns, K. V, Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M., Dharmaraj, S., Lopez, I., Ivings, L., Williams, G. A., Springell, K., Woods, C. G., Jafri, H., Rashid, Y., Strom, T. M., van der Zwaag, B., Gosens, I., Kersten, F. F. J., van Wijk, E., Veltman, J. A., Zonneveld, M. N., van Beersum, S. E. C., Maumenee, I. H., Wolfrum, U., Cheetham, M. E., Ueffing, M., Cremers, F. P. M., Inglehearn, C. F. and Roepman, R. (2007) 'Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis', *Nature Genetics*. Nature Publishing Group, 39(7), pp. 889–895. doi: 10.1038/ng2066.
- Holley, S. A., Geisler, R. and Nüsslein-Volhard, C. (2000) 'Control of her1 expression during zebrafish somitogenesis by a delta-dependent oscillator and an independent wave-front activity.', *Genes & development*, 14(13), pp. 1678–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10887161> (Accessed: 14 October 2018).

- Holopainen, S., Hytönen, M. K., Syrjä, P., Arumilli, M., Järvinen, A.-K., Rajamäki, M. and Lohi, H. (2017) 'ANLN truncation causes a familial fatal acute respiratory distress syndrome in Dalmatian dogs'. doi: 10.1371/journal.pgen.1006625.
- Hon, C.-C., Shin, J. W., Carninci, P. and Stubbington, M. J. (2017) 'The Human Cell Atlas: Technical approaches and challenges', *Briefings in Functional Genomics*. doi: 10.1093/bfgp/elx029.
- Hong, K. U., Reynolds, S. D., Giangreco, A., Hurley, C. M. and Stripp, B. R. (2001) *Clara Cell Secretory Protein-Expressing Cells of the Airway Neuroepithelial Body Microenvironment Include a Label-Retaining Subset and Are Critical for Epithelial Renewal after Progenitor Cell Depletion*, *Am. J. Respir. Cell Mol. Biol.* Available at: www.atsjournals.org (Accessed: 16 August 2018).
- Hong, K. U., Reynolds, S. D., Watkins, S., Fuchs, E. and Stripp, B. R. (2004) 'Basal Cells Are a Multipotent Progenitor Capable of Renewing the Bronchial Epithelium', *American Journal of Pathology*, 164(2), pp. 577–588. doi: 10.1016/S0002-9440(10)63147-1.
- Hong, K. U., Reynolds, S. D., Watkins, S., Fuchs, E. and Stripp, B. R. (2004) 'In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations', *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 286(4), pp. L643–L649. doi: 10.1152/ajplung.00155.2003.
- Hosono, K., Sasaki, T., Minoshima, S. and Shimizu, N. (2004) 'Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species', *GENE*. Elsevier, 340(1), pp. 31–43. doi: 10.1016/j.gene.2004.06.014.
- Hsiao, Y.-C., Tong, Z. J., Westfall, J. E., Ault, J. G., Page-McCaw, P. S. and Ferland, R. J. (2009) 'Ahi1, whose human ortholog is mutated in Joubert syndrome, is required for Rab8a localization, ciliogenesis and vesicle trafficking.', *Human molecular genetics*. Oxford University Press, 18(20), pp. 3926–41. doi: 10.1093/hmg/ddp335.
- Huang, L., Szymanska, K., Jensen, V. L., Janecke, A. R., Innes, A. M., Davis, E. E., Frosk, P., Li, C., Willer, J. R., Chodirker, B. N., Greenberg, C. R., McLeod, D. R., Bernier, F. P., Chudley, A. E., Müller, T., Shboul, M., Logan, C. V, Loucks, C. M., Beaulieu, C. L., Bowie, R. V, Bell, S. M., Adkins, J., Zuniga, F. I., Ross, K. D., Wang, J., Ban, M. R., Becker, C., Nürnberg, P., Douglas, S., Craft, C. M., Akimenko, M.-A., Hegele, R. A., Ober, C., Utermann, G., Bolz, H. J., Bulman, D. E., Katsanis, N., Blacque, O. E., Doherty, D., Parboosingh, J. S., Leroux, M. R., Johnson, C. A. and Boycott, K. M. (2011) 'TMEM237 is mutated in individuals with a Joubert syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone.', *American journal of human genetics*. Elsevier, 89(6), pp. 713–30. doi: 10.1016/j.ajhg.2011.11.005.

- Humbert, P. O., Rogers, C., Ganiatsas, S., Landsberg, R. L., Trimarchi, J. M., Dandapani, S., Brugnara, C., Erdman, S., Schrenzel, M., Bronson, R. T. and Lees, J. A. (2000) 'E2F4 is essential for normal erythrocyte maturation and neonatal viability.', *Molecular cell*, 6(2), pp. 281–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10983976> (Accessed: 7 October 2018).
- Inayama, Y., Hook, G. E., Brody, A. R., Cameron, G. S., Jetten, A. M., Gilmore, L. B., Gray, T. and Nettesheim, P. (1988) 'The differentiation potential of tracheal basal cells.', *Laboratory investigation; a journal of technical methods and pathology*, 58(6), pp. 706–17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3379917> (Accessed: 4 August 2018).
- Inayama, Y., Hook, G. E., Brody, A. R., Jetten, A. M., Gray, T., Mahler, J. and Nettesheim, P. (1989) *In Vitro and In Vivo Growth and Differentiation of Clones of Tracheal Basal Cells*, *American Journal of Pathology*. Available at: <https://www.ncbi.nlm.nih.gov/gate2.inist.fr/pmc/articles/PMC1879538/pdf/amjpathol00123-0056.pdf> (Accessed: 4 August 2018).
- Inglis, P., Boroevich, K. and Leroux, M. (2006) 'Piecing together a ciliome', *Trends in Genetics*, 22(9), pp. 491–500. doi: 10.1016/j.tig.2006.07.006.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P. and Linnarsson, S. (2011) 'Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq', *Genome Research*, 21(7), pp. 1160–1167. doi: 10.1101/gr.110882.110.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P. and Linnarsson, S. (2014) 'Quantitative single-cell RNA-seq with unique molecular identifiers', *Nature Methods*, 11(2), pp. 163–166. doi: 10.1038/nmeth.2772.
- Ito, T. (1999) 'Differentiation and Proliferation of Pulmonary Neuroendocrine Cells', *Progress in Histochemistry and Cytochemistry*. Urban & Fischer, 34(4), pp. 247–320. doi: 10.1016/S0079-6336(99)80001-1.
- Jaafari-Ashkavandi, Z., Mehranmehr, F. and Roosta, E. (2018) 'MCM3 and Ki67 proliferation markers in odontogenic cysts and ameloblastoma', *Journal of Oral Biology and Craniofacial Research*, 9(1), pp. 47–50. doi: 10.1016/j.jobcr.2018.09.003.
- Jacquet, B. V., Salinas-Mondragon, R., Liang, H., Therit, B., Buie, J. D., Dykstra, M., Campbell, K., Ostrowski, L. E., Brody, S. L. and Ghashghaei, H. T. (2009) 'FoxJ1-dependent gene expression is required for differentiation of radial glia into ependymal cells and a subset of astrocytes in the postnatal brain.', *Development (Cambridge, England)*. Company of Biologists, 136(23), pp. 4021–31. doi: 10.1242/dev.041129.

- Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A. and Amit, I. (2014) 'Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types.', *Science (New York, N.Y.)*. NIH Public Access, 343(6172), pp. 776–9. doi: 10.1126/science.1247651.
- Jänicke, M., Carney, T. J. and Hammerschmidt, M. (2007) 'Foxi3 transcription factors and Notch signaling control the formation of skin ionocytes from epidermal precursors of the zebrafish embryo', *Developmental Biology*, 307(2), pp. 258–271. doi: 10.1016/j.ydbio.2007.04.044.
- Jenkins, P. M., Mcewen, D. P. and Martens, J. R. (2009) 'Olfactory cilia: Linking sensory cilia function and human disease', *Chemical Senses*. Oxford University Press, 34(5), pp. 451–464. doi: 10.1093/chemse/bjp020.
- Ji, Z. and Ji, H. (2016) 'TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis.', *Nucleic acids research*, 44(13), p. e117. doi: 10.1093/nar/gkw430.
- Jia, G., Erickson, R. W., Choy, D. F., Mosesova, S., Wu, L. C., Solberg, O. D., Shikotra, A., Carter, R., Audusseau, S., Hamid, Q., Bradding, P., Fahy, J. V., Woodruff, P. G., Harris, J. M., Arron, J. R. and Bronchoscopic Exploratory Research Study of Biomarkers in Corticosteroid-refractory Asthma (BOBCAT) Study Group (2012) 'Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients', *Journal of Allergy and Clinical Immunology*, 130(3), p. 647–654.e10. doi: 10.1016/j.jaci.2012.06.025.
- Jing, Y., Yang, Y., Hao, F., Song, Y., Zhang, X., Zhang, Y., Huang, X., Hu, Q. and Ni, Y. (2018) 'Higher Ki67 expression in fibroblast like cells at invasive front indicates better clinical outcomes in oral squamous cell carcinoma patients.', *Bioscience reports*, p. BSR20181271. doi: 10.1042/BSR20181271.
- Johnson, N. F., Wilson, J. S., Habbersett, R., Thomassen, D. G., Shopp, G. M. and Smith, D. M. (1990) 'Separation and characterization of basal and secretory cells from the rat trachea by flow cytometry', *Cytometry*. Wiley-Blackwell, 11(3), pp. 395–405. doi: 10.1002/cyto.990110310.
- Al Jord, A., Shihavuddin, A., Servignat d'Aout, R., Faucourt, M., Genovesio, A., Karaïskou, A., Sobczak-Thépot, J., Spassky, N. and Meunier, A. (2017) 'Calibrated mitotic oscillator drives motile ciliogenesis.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 358(6364), pp. 803–806. doi: 10.1126/science.aan8311.
- Kanaji, N., Bandoh, S., Fujita, J., Ishii, T., Ishida, T. and Kubo, A. (2007) '1. Kanaji, N. et al. Compensation of type I and type II cytokeratin pools in lung cancer. *Lung Cancer* 55, 295–302 (2007). Compensation of type I and type II cytokeratin pools in lung cancer', *Lung Cancer*, 55(3), pp. 295–302. doi: 10.1016/j.lungcan.2006.11.004.

Keating, D., Marigowda, G., Burr, L., Daines, C., Mall, M. A., McKone, E. F., Ramsey, B. W., Rowe, S. M., Sass, L. A., Tullis, E., McKee, C. M., Moskowitz, S. M., Robertson, S., Savage, J., Simard, C., Van Goor, F., Waltz, D., Xuan, F., Young, T. and Taylor-Cousar, J. L. (2018) 'VX-445–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles', *New England Journal of Medicine*, 379(17), pp. 1612–1620. doi: 10.1056/NEJMoa1807120.

Keenan, K. P., Combs, J. W. and McDowell, E. M. (1982a) 'Regeneration of hamster tracheal epithelium after mechanical injury. I. Focal lesions: quantitative morphologic study of cell proliferation.', *Virchows Archiv. B, Cell pathology including molecular pathology*, 41(3), pp. 193–214. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6191435> (Accessed: 5 August 2018).

Keenan, K. P., Combs, J. W. and McDowell, E. M. (1982b) 'Regeneration of hamster tracheal epithelium after mechanical injury. II. Multifocal lesions: stathmokinetic and autoradiographic studies of cell proliferation.', *Virchows Archiv. B, Cell pathology including molecular pathology*, 41(3), pp. 215–29. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6135268> (Accessed: 5 August 2018).

Keenan, K. P., Combs, J. W. and McDowell, E. M. (1982c) 'Regeneration of hamster tracheal epithelium after mechanical injury. III. Large and small lesions: comparative stathmokinetic and single pulse and continuous thymidine labeling autoradiographic studies.', *Virchows Archiv. B, Cell pathology including molecular pathology*, 41(3), pp. 231–52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6135269> (Accessed: 5 August 2018).

Keenan, K. P., Wilson, T. S. and McDowell, E. M. (1983) 'Regeneration of hamster tracheal epithelium after mechanical injury. IV. Histochemical, immunocytochemical and ultrastructural studies.', *Virchows Archiv. B, Cell pathology including molecular pathology*, 43(3), pp. 213–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6194612> (Accessed: 5 August 2018).

Kerr, S. C., Carrington, S. D., Oscarson, S., Gallagher, M. E., Solon, M., Yuan, S., Ahn, J. N., Dougherty, R. H., Finkbeiner, W. E., Peters, M. C. and Fahy, J. V. (2014) 'Intelectin-1 is a prominent protein constituent of pathologic mucus associated with eosinophilic airway inflammation in asthma', *American Journal of Respiratory and Critical Care Medicine*, pp. 1005–1007. doi: 10.1164/rccm.201312-2220LE.

Kester, L. and van Oudenaarden, A. (2018) 'Single-Cell Transcriptomics Meets Lineage Tracing', *Cell Stem Cell*. doi: 10.1016/j.stem.2018.04.014.

- Kim, J. K., Vinarsky, V., Wain, J., Zhao, R., Jung, K., Choi, J., Lam, A., Pardo-Saganta, A., Breton, S., Rajagopal, J. and Yun, S. H. (2012) 'In vivo imaging of tracheal epithelial cells in mice during airway regeneration', *American Journal of Respiratory Cell and Molecular Biology*, 47(6), pp. 864–868. doi: 10.1165/rcmb.2012-0164OC.
- Kiran, K. C., Rothenberg, M. E. and Sherrill, J. D. (2015) 'In vitro model for studying esophageal epithelial differentiation and allergic inflammatory responses identifies keratin involvement in eosinophilic esophagitis', *PLoS ONE*, 10(6). doi: 10.1371/journal.pone.0127755.
- Kistemaker, L. E. M., Hiemstra, P. S., Bos, I. S. T., Bouwman, S., van den Berge, M., Hylkema, M. N., Meurs, H., Kerstjens, H. A. M. and Gosens, R. (2015) 'Tiotropium attenuates IL-13-induced goblet cell metaplasia of human airway epithelial cells', *Thorax*, 70(7), pp. 668–676. doi: 10.1136/thoraxjnl-2014-205731.
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W. (2015) 'Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells', *Cell*. NIH Public Access, 161(5), pp. 1187–1201. doi: 10.1016/j.cell.2015.04.044.
- Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.-D. and Nigg, E. A. (no date) 'Plk4-Induced Centriole Biogenesis in Human Cells'. doi: 10.1016/j.devcel.2007.07.002.
- KlosDehring, D. A., Vladar, E. K., Werner, M. E., Mitchell, J. W., Hwang, P. and Mitchell, B. J. (2013) 'Deuterosome-Mediated Centriole Biogenesis', *Developmental Cell*, 27(1), pp. 103–112. doi: 10.1016/j.devcel.2013.08.021.
- Kolodziejczyk, A. A. and Lönnberg, T. (2017) 'Global and targeted approaches to single-cell transcriptome characterization', *Briefings in Functional Genomics*. doi: 10.1093/bfpg/elx025.
- Komiya, T., Tanigawa, Y. and Hirohashi, S. (1998) 'Cloning of the Novel Gene Intelectin, Which Is Expressed in Intestinal Paneth Cells in Mice', *Biochemical and Biophysical Research Communications*, 251(3), pp. 759–762. doi: 10.1006/bbrc.1998.9513.
- Kondo, M., Tamaoki, J., Takeyama, K., Isono, K., Kawatani, K., Izumo, T.-H. and Nagai, A. (2006) 'Elimination of IL-13 Reverses Established Goblet Cell Metaplasia into Ciliated Epithelia in Airway Epithelial Cell Culture', *Allergology international*, 55, pp. 329–336. doi: 10.2332/allergolint.55.329.
- Kondo, M., Tamaoki, J., Takeyama, K., Nakata, J. and Nagai, A. (2002) 'Interleukin-13 induces goblet cell differentiation in primary cell culture from Guinea pig tracheal epithelium.', *American journal of respiratory cell and molecular biology*. American Thoracic Society, 27(5), pp. 536–41. doi: 10.1165/rcmb.4682.

- Königshoff, M., Balsara, N., Pfaff, E.-M., Kramer, M., Chrobak, I., Seeger, W. and Eickelberg, O. (2008) 'Functional Wnt Signaling Is Increased in Idiopathic Pulmonary Fibrosis', *PLoS ONE*. Edited by H. H. H. W. Schmidt, 3(5), p. e2142. doi: 10.1371/journal.pone.0002142.
- Königshoff, M., Kramer, M., Balsara, N., Wilhelm, J., Amarie, O. V., Jahn, A., Rose, F., Fink, L., Seeger, W., Schaefer, L., Günther, A. and Eickelberg, O. (2009) 'WNT1-inducible signaling protein-1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis', *Journal of Clinical Investigation*, 119(4), pp. 772–87. doi: 10.1172/JCI33950.
- Koopmans, T., Kumawat, K., Halayko, A. J. and Gosens, R. (2016) 'Regulation of actin dynamics by WNT-5A: implications for human airway smooth muscle contraction', *Nature Publishing Group*. doi: 10.1038/srep30676.
- Kumar, A. and Jagannathan, N. (2018) 'Cytokeratin: A review on current concepts', *International Journal of Orofacial Biology*. *International Journal of Orofacial Biology*, 2, pp. 6–11. doi: 10.4103/ijofb.ijofb\_3\_18.
- Kumar, P. A., Hu, Y., Yamamoto, Y., Hoe, N. B., Wei, T. S., Mu, D., Sun, Y., Joo, L. S., Dagher, R., Zielonka, E. M., Wang, D. Y., Lim, B., Chow, V. T., Crum, C. P., Xian, W. and McKeon, F. (2011) 'Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection', *Cell*, 147(3), pp. 525–538. doi: 10.1016/j.cell.2011.10.001.
- Kuperman, D. A., Huang, X., Koth, L. L., Chang, G. H., Dolganov, G. M., Zhu, Z., Elias, J. A., Sheppard, D. and Erle, D. J. (2002) 'Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma', *Nature Medicine*, 8(8), pp. 885–889. doi: 10.1038/nm734.
- Kuperman, D. A., Lewis, C. C., Woodruff, P. G., Rodriguez, M. W., Yang, Y. H., Dolganov, G. M., Fahy, J. V and Erle, D. J. (2005) 'Dissecting asthma using focused transgenic modeling and functional genomics.', *The Journal of allergy and clinical immunology*, 116(2), pp. 305–11. doi: 10.1016/j.jaci.2005.03.024.
- Kuperman, D., Schofield, B., Wills-Karp, M. and Grusby, M. J. (1998) 'Signal Transducer and Activator of Transcription Factor 6 (Stat6)-deficient Mice Are Protected from Antigen-induced Airway Hyperresponsiveness and Mucus Production', *The Journal of Experimental Medicine*, 187(6), pp. 939–948. doi: 10.1084/jem.187.6.939.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K. D., Yamada, R. G., Ueda, H. R. and Saitou, M. (2006) 'An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis', *Nucleic Acids Research*. Oxford University Press, 34(5), p. e42. doi: 10.1093/nar/gkl050.

Kwon, J.-W., Kwon, H.-K., Shin, H.-J., Choi, Y.-M., Anwar, M. A. and Choi, S. (2015) 'Activating transcription factor 3 represses inflammatory responses by binding to the p65 subunit of NF- $\kappa$ B.', *Scientific reports*, 5(1), p. 14470. doi: 10.1038/srep14470.

Kyrousi, C., Arbi, M., Pilz, G.-A., Pefani, D.-E., Lalioti, M.-E., Ninkovic, J., Go tz, M., Lygerou, Z. and Taraviras, S. (2015) 'Mcidas and GemC1 are key regulators for the generation of multiciliated ependymal cells in the adult neurogenic niche', *Development*, 142(21), pp. 3661–3674. doi: 10.1242/dev.126342.

Ladjemi, M. Z., Gras, D., Dupasquier, S., Detry, B., Lecocq, M., Garulli, C., Fregimilicka, C., Bouzin, C., Gohy, S., Chanez, P. and Pilette, C. (2018) 'Bronchial epithelial IgA secretion is impaired in asthma role of IL-4/IL-13', *American Journal of Respiratory and Critical Care Medicine*, 197(11), pp. 1396–1409. doi: 10.1164/rccm.201703-0561OC.

Lafkas, D., Shelton, A., Chiu, C., De Leon Boenig, G., Chen, Y., Stawicki, S. S., Siltanen, C., Reichelt, M., Zhou, M., Wu, X., Eastham-Anderson, J., Moore, H., Roose-Girma, M., Chinn, Y., Hang, J. Q., Warming, S., Egen, J., Lee, W. P., Austin, C., Wu, Y., Payandeh, J., Lowe, J. B. and Siebel, christian W. (2015) 'Therapeutic antibodies reveal Notch control of transdifferentiation in the adult lung', *Nature*, 528(7580), pp. 127–131. doi: 10.1038/nature15715.

Lai, H. Y. and Rogers, D. F. (2010) 'Mucus hypersecretion in asthma: Intracellular signalling pathways as targets for pharmacotherapy', *Current Opinion in Allergy and Clinical Immunology*, pp. 67–76. doi: 10.1097/ACI.0b013e328334643a.

Lai, Y., Chen, B., Shi, J., Palmer, J. N., Kennedy, D. W. and Cohen, N. A. (2011) 'Inflammation-mediated upregulation of centrosomal protein 110, a negative modulator of ciliogenesis, in patients with chronic rhinosinusitis', *Journal of Allergy and Clinical Immunology*, 128(6). doi: 10.1016/j.jaci.2011.09.001.

Lambacher, N. J., Bruel, A.-L., van Dam, T. J. P., Szymańska, K., Slaats, G. G., Kuhns, S., McManus, G. J., Kennedy, J. E., Gaff, K., Wu, K. M., van der Lee, R., Burglen, L., Doummar, D., Rivière, J.-B., Faivre, L., Attié-Bitach, T., Saunier, S., Curd, A., Peckham, M., Giles, R. H., Johnson, C. A., Huynen, M. A., Thauvin-Robinet, C. and Blacque, O. E. (2016) 'TMEM107 recruits ciliopathy proteins to subdomains of the ciliary transition zone and causes Joubert syndrome.', *Nature cell biology*. Europe PMC Funders, 18(1), pp. 122–31. doi: 10.1038/ncb3273.

Laoukili, J., Perret, E., Willems, T., Minty, A., Parthoens, E., Houcine, O., Coste, A., Jorissen, M., Marano, F., Caput, D. and Tournier, F. (2001) 'IL-13 alters mucociliary differentiation and ciliary beating of human respiratory epithelial cells.', *The Journal of clinical investigation*, 108(12), pp. 1817–24. doi: 10.1172/JCI13557.

- Lee, I. T. and Yang, C. M. (2013) 'Inflammatory signalings involved in airway and pulmonary diseases', *Mediators of Inflammation*. Hindawi, p. 791231. doi: 10.1155/2013/791231.
- Lee, J. H., Kaminski, N., Dolganov, G., Grunig, G., Koth, L., Solomon, C., Erle, D. J. and Sheppard, D. (2001) 'Interleukin-13 induces dramatically different transcriptional programs in three human airway cell types.', *American journal of respiratory cell and molecular biology*. American Thoracic Society New York, NY, 25(4), pp. 474–85. doi: 10.1165/ajrcmb.25.4.4522.
- Lee, J. H., Silhavy, J. L., Lee, J. E., Al-Gazali, L., Thomas, S., Davis, E. E., Bielas, S. L., Hill, K. J., Iannicelli, M., Brancati, F., Gabriel, S. B., Russ, C., Logan, C. V, Sharif, S. M., Bennett, C. P., Abe, M., Hildebrandt, F., Diplas, B. H., Attié-Bitach, T., Katsanis, N., Rajab, A., Koul, R., Sztriha, L., Waters, E. R., Ferro-Novick, S., Woods, C. G., Johnson, C. A., Valente, E. M., Zaki, M. S. and Gleeson, J. G. (2012) 'Evolutionarily assembled cis-regulatory module at a human ciliopathy locus.', *Science (New York, N.Y.)*. NIH Public Access, 335(6071), pp. 966–9. doi: 10.1126/science.1213506.
- Leidel, S., Delattre, M., Cerutti, L., Baumer, K. and Gönczy, P. (2005) 'SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells', *NATURE CELL BIOLOGY*, 7(2). doi: 10.1038/ncb1220.
- Levsky, J. M. and Singer, R. H. (2003) 'Gene expression and the myth of the average cell', *Trends in Cell Biology*, 13(1), pp. 4–6. doi: 10.1016/S0962-8924(02)00002-8.
- Lin, S.-C., Chou, H.-C., Chiang, B.-L. and Chen, C.-M. (2017) 'CTGF upregulation correlates with MMP-9 level in airway remodeling in a murine model of asthma', *Archives of Medical Science*, 3(3), pp. 670–676. doi: 10.5114/aoms.2016.60371.
- Litosh, V. A., Rochman, M., Rymer, J. K., Porollo, A., Kottyan, L. C. and Rothenberg, M. E. (2017) 'Calpain-14 and its association with eosinophilic esophagitis', *Journal of Allergy and Clinical Immunology*, 139(6), p. 1762–1771.e7. doi: 10.1016/j.jaci.2016.09.027.
- Liu, Y., Pathak, N., Kramer-Zucker, A. and Drummond, I. a (2007) 'Notch signaling controls the differentiation of transporting epithelia and multiciliated cells in the zebrafish pronephros.', *Development (Cambridge, England)*, 134, pp. 1111–1122. doi: 10.1242/dev.02806.
- Livraghi-Butrico, A., Kelly, E. J., Klem, E. R., Dang, H., Wolfgang, M. C., Boucher, R. C., Randell, S. H. and O'Neal, W. K. (2012) 'Mucus clearance, MyD88-dependent and MyD88-independent immunity modulate lung susceptibility to spontaneous bacterial infection and inflammation', *Mucosal Immunology*, 5(4), pp. 397–408. doi: 10.1038/mi.2012.17.

- Livraghi, A., Grubb, B. R., Hudson, E. J., Wilkinson, K. J., Sheehan, J. K., Mall, M. A., O'neal, W. K., Boucher, R. C. and Randell, S. H. (2009) 'Airway and lung pathology due to mucosal surface dehydration in  $\beta$ -Epithelial Na<sup>+</sup> Channel-overexpressing mice: role of TNF $\alpha$  and IL-4R $\alpha$  signaling, influence of neonatal development, and limited efficacy of glucocorticoid treatment', *J Immunol*, 182(7), pp. 4357–4367. doi: 10.4049/jimmunol.0802557.
- Loncarek, J. and Bettencourt-Dias, M. (2018) 'Building the right centriole for each cell type', *Journal of Cell Biology*, pp. 823–835. doi: 10.1083/jcb.201704093.
- Louahed, J., Toda, M., Jen, J., Hamid, Q., Renauld, J. C., Levitt, R. C. and Nicolaides, N. C. (2000) 'Interleukin-9 upregulates mucus expression in the airways', *American Journal of Respiratory Cell and Molecular Biology*, 22(6), pp. 649–656. doi: 10.1165/ajrcmb.22.6.3927.
- Lum, H., Schwartz, L. W., Dungworth, D. L. and Tyler, W. S. (1978) 'A comparative study of cell renewal after exposure to ozone or oxygen. Response of terminal bronchiolar epithelium in the rat.', *The American review of respiratory disease*, 118(2), pp. 335–45. doi: 10.1164/arrd.1978.118.2.335.
- Lyons, R. A., Saridogan, E. and Djahanbakhch, O. (2006) 'The reproductive significance of human Fallopian tube cilia', *Human Reproduction Update*. Oxford University Press, 12(4), pp. 363–372. doi: 10.1093/humupd/dml012.
- Ma, L., Quigley, I., Omran, H. and Kintner, C. (2014) 'Multicilin drives centriole biogenesis via E2f proteins', *Genes and Development*, 28(13), pp. 1461–1471. doi: 10.1101/gad.243832.114.
- Machen, T. E. (2006) 'Innate immune response in CF airway epithelia: hyperinflammatory?', *American journal of physiology. Cell physiology*, 291(2), pp. 218–230. doi: 10.1152/ajpcell.00605.2005.
- Macosko, E. Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A. and McCarroll, S. A. (2015) 'Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets', *Cell*. NIH Public Access, 161(5), pp. 1202–1214. doi: 10.1016/j.cell.2015.05.002.
- Maguer-Satta, V., Bartholin, L., Jeanpierre, S., Ffrench, M., Martel, S., Magaud, J.-P. and Rimokh, R. (2003) 'Regulation of human erythropoiesis by activin A, BMP2, and BMP4, members of the TGFbeta family.', *Experimental cell research*, 282(2), pp. 110–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12531697> (Accessed: 15 October 2018).

- Mall, M. A., Harkema, J. R., Trojanek, J. B., Treis, D., Livraghi, A., Schubert, S., Zhou, Z., Kreda, S. M., Tilley, S. L., Hudson, E. J., O'Neal, W. K. and Boucher, R. C. (2008) 'Development of chronic bronchitis and emphysema in  $\beta$ -epithelial Na<sup>+</sup> channel-overexpressing mice', *American Journal of Respiratory and Critical Care Medicine*, 177(7), pp. 730–742. doi: 10.1164/rccm.200708-1233OC.
- Mall, M., Bleich, M., Greger, R., Schreiber, R. and Kunzelmann, K. (1998) 'The amiloride-inhibitable Na<sup>+</sup> conductance is reduced by the cystic fibrosis transmembrane conductance regulator in normal but not in cystic fibrosis airways.', *The Journal of clinical investigation*, 102(1), pp. 15–21. doi: 10.1172/JCI2729.
- Mall, M., Grubb, B. R., Harkema, J. R., O'Neal, W. K. and Boucher, R. C. (2004) 'Increased airway epithelial Na<sup>+</sup> absorption produces cystic fibrosis-like lung disease in mice', *Nature Medicine*, 10(5), pp. 487–493. doi: 10.1038/nm1028.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., Cramer, P., Adameyko, I., Linnarsson, S. and Kharchenko, P. V (2018) 'RNA velocity of single cells.', *Nature*, 560(7719), pp. 494–498. doi: 10.1038/s41586-018-0414-6.
- Marcet, B., Chevalier, B., Coraux, C., Kodjabachian, L. and Barbry, P. (2011) 'MicroRNA-based silencing of Delta/Notch signaling promotes multiple cilia formation', *Cell Cycle*, 10(January 2015), pp. 2858–2864. doi: 10.4161/cc.10.17.17011.
- Marcet, B., Chevalier, B., Luxardi, G., Coraux, C., Zaragosi, L. E., Cibois, M., Robbe-Sermesant, K., Jolly, T., Cardinaud, B., Moreilhon, C., Giovannini-Chami, L., Nawrocki-Raby, B., Birembaut, P., Waldmann, R., Kodjabachian, L. and Barbry, P. (2011) 'Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway', *Nature Cell Biology*, 13(6), pp. 693–701. doi: 10.1038/ncb2241.
- Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L. and Yuan, G.-C. (2014) 'Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(52), pp. E5643-50. doi: 10.1073/pnas.1408993111.
- Marshall, C. B., Mays, D. J., Beeler, J. S., Rosenbluth, J. M., Boyd, K. L., Santos Guasch, G. L., Shaver, T. M., Tang, L. J., Liu, Q., Shyr, Y., Venters, B. J., Magnuson, M. A. and Pietenpol, J. A. (2016) 'P73 Is Required for Multiciliogenesis and Regulates the Foxj1-Associated Gene Network', *Cell Reports*, 14(10), pp. 2289–2300. doi: 10.1016/j.celrep.2016.02.035.
- Maselli, D. J. and Hanania, N. A. (2018) 'Asthma COPD Overlap: Impact of Associated Comorbidities', *Pulmonary Pharmacology & Therapeutics*. doi: 10.1016/j.pupt.2018.08.006.

- Matsui, H., Wagner, V. E., Hill, D. B., Schwab, U. E., Rogers, T. D., Button, B., Taylor, R. M., Superfine, R., Rubinstein, M., Iglewski, B. H. and Boucher, R. C. (2006) 'A physical linkage between cystic fibrosis airway surface dehydration and *Pseudomonas aeruginosa* biofilms', *Proceedings of the National Academy of Sciences of the United States of America*, 103(48), pp. 18131–18136. doi: 10.1073/pnas.0606428103.
- McGarry, T. J. and Kirschner, M. W. (1998) *Geminin, an Inhibitor of DNA Replication, Is Degraded during Mitosis*, *Cell*. Available at: [https://ac-els-cdn-com.gate2.inist.fr/S009286740081209X/1-s2.0-S009286740081209X-main.pdf?\\_tid=96229015-5d84-46ac-98f7-8f8ce67793c6&acdnat=1540304112\\_d2727dc8f531a2caa8378ec9b453f562](https://ac-els-cdn-com.gate2.inist.fr/S009286740081209X/1-s2.0-S009286740081209X-main.pdf?_tid=96229015-5d84-46ac-98f7-8f8ce67793c6&acdnat=1540304112_d2727dc8f531a2caa8378ec9b453f562) (Accessed: 23 October 2018).
- Mercer, R. R., Russell, M. L., Roggli, V. L. and Crapo, J. D. (1994) 'Cell number and distribution in human and rat airways.', *American journal of respiratory cell and molecular biology*, 10(6), pp. 613–24. doi: 10.1165/ajrcmb.10.6.8003339.
- Meunier, A. and Azimzadeh, J. (2016) 'Multiciliated Cells in Animals.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 8(12), p. a028233. doi: 10.1101/cshperspect.a028233.
- Mick, D. U., Rodrigues, R. B., Leib, R. D., Adams, C. M., Chien, A. S., Gygi, S. P., Nachury Correspondence, M. V and Nachury, M. V (2015) 'Proteomics of Primary Cilia by Proximity Labeling Developmental Cell Proteomics of Primary Cilia by Proximity Labeling', *DEVCEL*, 35, pp. 497–512. doi: 10.1016/j.devcel.2015.10.015.
- Mira, H., Andreu, Z., Suh, H., Lie, D. C., Jessberger, S., Consiglio, A., San Emeterio, J., Hortigüela, R., Marqués-Torrejón, M. Á., Nakashima, K., Colak, D., Götz, M., Fariñas, I. and Gage, F. H. (2010) 'Signaling through BMPR-IA Regulates Quiescence and Long-Term Activity of Neural Stem Cells in the Adult Hippocampus', *Cell Stem Cell*, 7(1), pp. 78–89. doi: 10.1016/j.stem.2010.04.016.
- Mitchison, H. M. and Valente, E. M. (2017) 'Motile and non-motile cilia in human pathology: from function to phenotypes', *The Journal of Pathology*, 241(2), pp. 294–309. doi: 10.1002/path.4843.
- Molla-Herman, A., Ghossoub, R., Blisnick, T., Meunier, A., Serres, C., Silbermann, F., Emmerson, C., Romeo, K., Bourdoncle, P., Schmitt, A., Saunier, S., Spassky, N., Bastin, P. and Benmerah, A. (2010) 'The ciliary pocket: an endocytic membrane domain at the base of primary and motile cilia', *Journal of Cell Science*, 123(10), pp. 1785–1795. doi: 10.1242/jcs.059519.
- Montoro, D. T., Haber, A. L., Biton, M., Vinarsky, V., Lin, B., Birket, S. E., Yuan, F., Chen, S., Leung, H. M., Villoria, J., Rogel, N., Burgin, G., Tsankov, A. M., Waghray, A., Slyper, M., Waldman, J., Nguyen, L., Dionne, D. and Rajagopal, J. (2018) 'A revised airway epithelial hierarchy includes CFTR-expressing ionocytes', *Orit rozenblatt-rosen*, 4, p. 19. doi: 10.1038/s41586-018-0393-7.

Mori, M., Hazan, R., Danielian, P. S., Mahoney, J. E., Li, H., Lu, J., Miller, E. S., Zhu, X., Lees, J. A. and Cardoso, W. V (2017) 'Cytoplasmic E2f4 forms organizing centres for initiation of centriole amplification during multiciliogenesis', *Nature Communications*, 8. doi: 10.1038/ncomms15857.

Mori, M., Mahoney, J. E., Stupnikov, M. R., Paez-Cortez, J. R., Szymaniak, A. D., Varelas, X., Herrick, D. B., Schwob, J., Zhang, H. and Cardoso, W. V (2015) 'Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors', *Development*, 142(2), pp. 258–267. doi: 10.1242/dev.116855.

Morimoto, M., Liu, Z., Cheng, H.-T., Winters, N., Bader, D. and Kopan, R. (2010) 'Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate', *Journal of Cell Science*. Company of Biologists, 123(2), pp. 213–224. doi: 10.1242/jcs.058669.

Morimoto, M., Nishinakamura, R., Saga, Y. and Kopan, R. (2012) 'Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells.', *Development (Cambridge, England)*. Company of Biologists, 139(23), pp. 4365–73. doi: 10.1242/dev.083840.

Mou, H., Vinarsky, V., Tata, P. R., Brazauskas, K., Choi, S. H., Crooke, A. K., Zhang, B., Solomon, G. M., Turner, B., Bihler, H., Harrington, J., Lapey, A., Channick, C., Keyes, C., Freund, A., Artandi, S., Mense, M., Rowe, S., Engelhardt, J. F., Hsu, Y.-C. and Rajagopal, J. (2016) 'Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells', *Cell Stem Cell*, 19(2), pp. 217–231. doi: 10.1016/j.stem.2016.05.012.

Nadel, J. a and Burgel, P. R. (2001) 'The role of epidermal growth factor in mucus production.', *Current opinion in pharmacology*, 1(3), pp. 254–258. Available at: [https://ac-els-cdn-com.gate2.inist.fr/S1471489201000455/1-s2.0-S1471489201000455-main.pdf?\\_tid=ba027a76-baf1-483c-be4e-8261ca6a77e6&acdnat=1534417714\\_b9e05aeb7ba240e41c9d48c4554a7a6f](https://ac-els-cdn-com.gate2.inist.fr/S1471489201000455/1-s2.0-S1471489201000455-main.pdf?_tid=ba027a76-baf1-483c-be4e-8261ca6a77e6&acdnat=1534417714_b9e05aeb7ba240e41c9d48c4554a7a6f) (Accessed: 16 August 2018).

Nagel, G., Barbry, P., Chabot, H., Brochiero, E., Hartung, K. and Grygorczyk, R. (2005) 'CFTR fails to inhibit the epithelial sodium channel ENaC expressed in *Xenopus laevis* oocytes', *The Journal of Physiology*, 564(3), pp. 671–682. doi: 10.1113/jphysiol.2004.079046.

Nahm, D. H. and Park, H. S. (1998) 'Analysis of induced sputum for studying allergen-specific IgE antibodies in airway secretion from asthmatic patients.', *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, 28(6), pp. 686–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9677132> (Accessed: 9 October 2018).

- Nakajima, M., Kawanami, O., Jin, E., Ghazizadeh, M., Honda, M., Asano, G., Horiba, K. and Ferrans, V. J. (1998) 'Immunohistochemical and ultrastructural studies of basal cells, Clara cells and bronchiolar cuboidal cells in normal human airways', *Pathology International*, 48(12), pp. 944–953. doi: 10.1111/j.1440-1827.1998.tb03865.x.
- Nakao, I., Kanaji, S., Ohta, S., Matsushita, H., Arima, K., Yuyama, N., Yamaya, M., Nakayama, K., Kubo, H., Watanabe, M., Sagara, H., Sugiyama, K., Tanaka, H., Toda, S., Hayashi, H., Inoue, H., Hoshino, T., Shiraki, A., Inoue, M., Suzuki, K., Aizawa, H., Okinami, S., Nagai, H., Hasegawa, M., Fukuda, T., Green, E. D. and Izuhara, K. (2008) 'Identification of Pendrin as a Common Mediator for Mucus Production in Bronchial Asthma and Chronic Obstructive Pulmonary Disease', *The Journal of Immunology*, 180(9), pp. 6262–6269. doi: 10.4049/jimmunol.180.9.6262.
- Nakazawa, Y., Hiraki, M., Kamiya, R. and Hirono, M. (2007) 'SAS-6 is a Cartwheel Protein that Establishes the 9-Fold Symmetry of the Centriole', *Current Biology*, 17(24), pp. 2169–2174. doi: 10.1016/j.cub.2007.11.046.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., Muthuswamy, L., Krasnitz, A., McCombie, W. R., Hicks, J. and Wigler, M. (2011) 'Tumour evolution inferred by single-cell sequencing', *Nature*, 472(7341), pp. 90–94. doi: 10.1038/nature09807.
- Nemajerova, A., Amelio, I., Gebel, J., Dötsch, V., Melino, G. and Moll, U. M. (2018) 'Non-oncogenic roles of TAp73: From multiciliogenesis to metabolism', *Cell Death and Differentiation*, pp. 144–153. doi: 10.1038/cdd.2017.178.
- Nemajerova, A., Kramer, D., Siller, S. S., Herr, C., Shomroni, O., Pena, T., Suazo, C. G., Glaser, K., Wildung, M., Steffen, H., Sriraman, A., Oberle, F., Wienken, M., Hennion, M., Vidal, R., Royen, B., Alevra, M., Schild, D., Bals, R., Dönitz, J., Riedel, D., Bonn, S., Takemaru, K. I., Moll, U. M. and Lizé, M. (2016) 'TAp73 is a central transcriptional regulator of airway multiciliogenesis', *Genes and Development*, 30(11), pp. 1300–1312. doi: 10.1101/gad.279836.116.
- Nonaka, S., Tanaka, Y., Okada, Y., Takeda, S., Harada, A., Kanai, Y., Kido, M. and Hirokawa, N. (1998) 'Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein', *Cell*, 95(6), pp. 829–837. doi: 10.1016/S0092-8674(00)81705-5.
- Oettgen, H. C. and Geha, R. S. (2001) 'IgE regulation and roles in asthma pathogenesis', *Journal of Allergy and Clinical Immunology*, 107(3), pp. 429–441. doi: 10.1067/mai.2001.113759.
- Oh, C. K., Geba, G. P. and Molfino, N. (2010) 'Investigational therapeutics targeting the IL-4/IL-13/STAT-6 pathway for the treatment of asthma', *European Respiratory Review*, pp. 46–54. doi: 10.1183/09059180.00007609.

- Ohashi, S., Natsuizaka, M., Yashirohtani, Y., Kalman, R. A., Nakagawa, M., Wu, L., Kleinszanto, A. J., Herlyn, M., Diehl, J. A., Katz, J. P., Pear, W. S., Seykora, J. T. and Nakagawa, H. (2010) 'NOTCH1 and NOTCH3 coordinate esophageal squamous differentiation through a csl-dependent transcriptional network', *Gastroenterology*, 139(6), pp. 2113–2123. doi: 10.1053/j.gastro.2010.08.040.
- Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M. and Kitagawa, D. (2014) 'Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole', *Nature communications*. Nature Publishing Group, 5, p. 5267. doi: 10.1038/ncomms6267.
- Ojiaku, C. A., Yoo, E. J. and Panettieri, R. A. (2017) 'Transforming growth factor  $\beta$ 1 function in airway remodeling and hyperresponsiveness: The missing link?', *American Journal of Respiratory Cell and Molecular Biology*, 56(4), pp. 432–442. doi: 10.1165/rcmb.2016-0307TR.
- Okuda, K., Chen, G., Subramani, D. B., Wolf, M., Gilmore, R. C., Kato, T., Radicioni, G., Kesimer, M., Chua, M., Dang, H., Livraghi-Butrico, A., Ehre, C., Doerschuk, C. M., Randell, S. H., Matsui, H., Nagase, T., O'Neal, W. K. and Boucher, R. C. (2018) 'Localization of Secretory Mucins MUC5AC and MUC5B in Normal/Healthy Human Airways', *American Journal of Respiratory and Critical Care Medicine*, p. rccm.201804-0734OC. doi: 10.1164/rccm.201804-0734OC.
- Ordovas-Montanes, J., Dwyer, D. F., Nyquist, S. K., Buchheit, K. M., Vukovic, M., Deb, C., Wadsworth, M. H., Hughes, Travis K., Kazer, S. W., Yoshimoto, E., Cahill, K. N., Bhattacharyya, N., Katz, H. R., Berger, B., Laidlaw, Tanya M., Boyce, J. A., Barrett, N. A. and Shalek, A. K. (2018) 'Allergic inflammatory memory in human respiratory epithelial progenitor cells', *Nature*. doi: 10.1038/s41586-018-0449-8.
- Paintrand, M., Moudjou, M., Delacroix, H. and Bornens, M. (1992) 'Centrosome organization and centriole architecture: Their sensitivity to divalent cations', *Journal of Structural Biology*, 108(2), pp. 107–128. doi: 10.1016/1047-8477(92)90011-X.
- Pan, J. H., Adair-Kirk, T. L., Patel, A. C., Huang, T., Yozamp, N. S., Xu, J., Reddy, E. P., Byers, D. E., Pierce, R. A., Holtzman, M. J. and Brody, S. L. (2014) 'Myb permits multilineage airway epithelial cell differentiation', *Stem Cells*, 32(12), pp. 3245–3256. doi: 10.1002/stem.1814.
- Paper, W. (2017) *THE HUMAN CELL ATLAS*. Available at: [https://www.humancellatlas.org/files/HCA\\_WhitePaper\\_18Oct2017.pdf](https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf) (Accessed: 29 September 2018).
- Papi, A., Brightling, C., Pedersen, S. E. and Reddel, H. K. (2018) 'Asthma', *The Lancet*, 391, pp. 783–800. doi: 10.1016/S0140-6736(17)33311-1.

- Pardo-Saganta, A., Law, B. M., Gonzalez-Celeiro, M., Vinarsky, V. and Rajagopal, J. (2013) 'Ciliated Cells of Pseudostratified Airway Epithelium Do Not Become Mucous Cells after Ovalbumin Challenge', *American Journal of Respiratory Cell and Molecular Biology*, 48(3), pp. 364–373. doi: 10.1165/rcmb.2012-0146OC.
- Pardo-Saganta, A., Law, B. M., Tata, P. R., Villoria, J., Saez, B., Mou, H., Zhao, R. and Rajagopal, J. (2015) 'Injury Induces Direct Lineage Segregation of Functionally Distinct Airway Basal Stem/Progenitor Cell Subpopulations', *Cell Stem Cell*, 16(2), pp. 184–197. doi: 10.1016/j.stem.2015.01.002.
- Pardo-Saganta, A., Tata, P. R., Law, B. M., Saez, B., Chow, R. D.-W., Prabhu, M., Gridley, T. and Rajagopal, J. (2015) 'Parent stem cells can serve as niches for their daughter cells.', *Nature*. doi: 10.1038/nature14553.
- Park, J.-A. and Tschumperlin, D. J. (2009) 'Chronic Intermittent Mechanical Stress Increases MUC5AC Protein Expression', *American Journal of Respiratory Cell and Molecular Biology*, 41(4), pp. 459–466. doi: 10.1165/rcmb.2008-0195OC.
- Park, J. A. and Fredberg, J. J. (2016) 'Cell jamming in the airway epithelium', *Annals of the American Thoracic Society*, 13, pp. S64–S67. doi: 10.1513/AnnalsATS.201507-476MG.
- Park, K.-S., Korfhagen, T. R., Bruno, M. D., Kitzmiller, J. A., Wan, H., Wert, S. E., Khurana Hershey, G. K., Chen, G. and Whitsett, J. A. (2007) 'SPDEF regulates goblet cell hyperplasia in the airway epithelium', *The Journal of Clinical Investigation*, 117. doi: 10.1172/JCI29176.
- Park, K. S., Wells, J. M., Zorn, A. M., Wert, S. E., Laubach, V. E., Fernandez, L. G. and Whitsett, J. A. (2006) 'Transdifferentiation of ciliated cells during repair of the respiratory epithelium', *American Journal of Respiratory Cell and Molecular Biology*, 34(2), pp. 151–157. doi: 10.1165/rcmb.2005-0332OC.
- Parker, J. C., Thavagnanam, S., Skibinski, G., Lyons, J., Bell, J., Heaney, L. G. and Shields, M. D. (2013) 'Chronic IL9 and IL-13 exposure leads to an altered differentiation of ciliated cells in a well-differentiated paediatric bronchial epithelial cell model.', *PloS one*, 8(5), p. e61023. doi: 10.1371/journal.pone.0061023.
- Parker, J. C., Thavagnanam, S., Skibinski, G., McBrien, M., Heaney, L. G. and Shields, M. D. (2012) 'IL-31 does not induce normal human ciliated epithelial cells to differentiate into a phenotype consistent with the pathophysiology of asthma.', *Results in immunology*, 2, pp. 104–11. doi: 10.1016/j.rinim.2012.05.001.
- Pascual, R. M. and Peters, S. P. (2005) 'Airway remodeling contributes to the progressive loss of lung function in asthma: An overview', *Journal of Allergy and Clinical Immunology*, pp. 477–486. doi: 10.1016/j.jaci.2005.07.011.

- Pefani, D.-E., Dimaki, M., Spella, M., Karantzelis, N., Mitsiki, E., Kyrousi, C., Symeonidou, I.-E., Perrakis, A., Taraviras, S. and Lygerou, Z. (2011) 'Idas, a Novel Phylogenetically Conserved Geminin-related Protein, Binds to Geminin and Is Required for Cell Cycle Progression \* □ S'. doi: 10.1074/jbc.M110.207688.
- Pelletier, L., O'Toole, E., Schwager, A., Hyman, A. A. and Müller-Reichert, T. (2006) 'Centriole assembly in *Caenorhabditis elegans*', *Nature*, 444(7119), pp. 619–623. doi: 10.1038/nature05318.
- Pemberton, A. D., Verdon, B., Inglis, N. F. and Pearson, J. P. (2011) 'Sheep intelectin-2 co-purifies with the mucin Muc5ac from gastric mucus.', *Research in veterinary science*, 91(3), pp. e53-7. doi: 10.1016/j.rvsc.2011.03.004.
- Pezzulo, A. A., Tang, X. X., Hoegger, M. J., Abou Alaiwa, M. H., Ramachandran, S., Moninger, T. O., Karp, P. H., Wohlford-Lenane, C. L., Haagsman, H. P., Eijk, M. Van, Bánfi, B., Horswill, A. R., Stoltz, D. A., Mc Cray, P. B., Welsh, M. J. and Zabner, J. (2012) 'Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung', *Nature*, pp. 109–113. doi: 10.1038/nature11130.
- Picelli, S. (2017) 'Single-cell RNA-sequencing: The future of genome biology is now', *RNA Biology*, 14(5), pp. 637–650. doi: 10.1080/15476286.2016.1201618.
- Pizov, R. (2005) 'The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease', *Survey of Anesthesiology*, 49(2), pp. 95–96. doi: 10.1097/01.sa.0000158589.31056.51.
- Plasschaert, L. W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A. M. and Jaffe, A. B. (2018) 'A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte', *Nature*. Nature Publishing Group, p. 1. doi: 10.1038/s41586-018-0394-6.
- Plopper, C. G., Halsebo, J. E., Berger, W. J., Sonstegard, K. S. and Nettesheim, P. (1983) 'Distribution of nonciliated bronchiolar epithelial (Clara) cells in intra- and extrapulmonary airways of the rabbit.', *Experimental lung research*, 5(2), pp. 79–98. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6628348> (Accessed: 6 August 2018).
- Porchet, N. and Aubert, J. P. (2004) 'Les gènes MUC: Mucin or not mucin? That is the question', *Medecine/Sciences*, pp. 569–574. doi: 10.1051/medsci/2004205569.
- Puchelle, E., Zahm, J.-M., Tournier, J.-M. and Coraux, C. (2006) 'Airway Epithelial Repair, Regeneration, and Remodeling after Injury in Chronic Obstructive Pulmonary Disease', *Proceedings of the American Thoracic Society*, 3(8), pp. 726–733. doi: 10.1513/pats.200605-126SF.
- Pursglove, S. E. and Mackay, J. P. (2005) 'CSL: A notch above the rest', *International Journal of Biochemistry and Cell Biology*, pp. 2472–2477. doi: 10.1016/j.biocel.2005.06.013.

- Qin, Y., Jiang, Y., Sheikh, A. S., Shen, S., Liu, J. and Jiang, D. (2016) 'Interleukin-13 stimulates MUC5AC expression via a STAT6-TMEM16A-ERK1/2 pathway in human airway epithelial cells Extracellular regulated kinase Signal transducer and activator of transcription 6 Mucin 5AC', *International Immunopharmacology*, 40, pp. 106–114. doi: 10.1016/j.intimp.2016.08.033.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A. and Trapnell, C. (2017) 'Reversed graph embedding resolves complex single-cell trajectories', *Nature Methods*, 14(10), pp. 979–982. doi: 10.1038/nmeth.4402.
- Qu, S.-Y., He, Y.-L., Zhang, J. and Wu, C.-G. (2017) 'Transcription factor RBP-J-mediated signalling regulates basophil immunoregulatory function in mouse asthma model', *IMMUNOLOGY*. doi: 10.1111/imm.12753.
- Quigley, I. K. and Kintner, C. (2017) 'Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression.', *PLoS genetics*. Public Library of Science, 13(1), p. e1006538. doi: 10.1371/journal.pgen.1006538.
- Quigley, I. K., Stubbs, J. L. and Kintner, C. (2011) 'Specification of ion transport cells in the *Xenopus* larval skin', *Development*. Company of Biologists, 138(4), pp. 705–714. doi: 10.1242/dev.055699.
- Raghavan, S. S., Hong, E. K., Kim, Y. H. and Kim, J. (2018) 'Utility of CD30, Ki67 and p53 in assisting with the diagnosis Mycosis Fungoides with Large Cell Transformation.', *Journal of cutaneous pathology*. doi: 10.1111/cup.13375.
- Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., Gagnon, J. A. and Schier, A. F. (2018) 'Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain', *Nature Biotechnology*, 36(5), pp. 442–450. doi: 10.1038/nbt.4103.
- Rajavelu, P., Chen, G., Xu, Y., Kitzmiller, J. A., Korfhagen, T. R. and Whitsett, J. A. (2015) 'Airway epithelial SPDEF integrates goblet cell differentiation and pulmonary Th2 inflammation', *Journal of Clinical Investigation*, 125(5), pp. 2021–2031. doi: 10.1172/JCI79422.
- Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtkova, I., Loring, J. F., Laurent, L. C., Schroth, G. P. and Sandberg, R. (2012) 'Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.', *Nature biotechnology*. NIH Public Access, 30(8), pp. 777–82. doi: 10.1038/nbt.2282.
- Rawlins, E. L., Okubo, T., Xue, Y., Brass, D. M., Auten, R. L., Hasegawa, H., Wang, F. and Hogan, B. L. M. (2009) 'The Role of Scgb1a1+ Clara Cells in the Long-Term Maintenance and Repair of Lung Airway, but Not Alveolar, Epithelium', *Cell Stem Cell*, 4(6), pp. 525–534. doi: 10.1016/j.stem.2009.04.002.

- Reader, J. R., Tepper, J. S., Schelegle, E. S., Aldrich, M. C., Putney, L. F., Pfeiffer, J. W. and Hyde, D. M. (2003) 'Pathogenesis of Mucous Cell Metaplasia in a Murine Asthma Model', *American Journal of Pathology*, 162 N° 6, pp. 2069–2078. Available at: [https://www.ncbi.nlm.nih.gov.gate2.inist.fr/pmc/articles/PMC2216702/pdf/3638.pdf](https://www.ncbi.nlm.nih.gov/gate2.inist.fr/pmc/articles/PMC2216702/pdf/3638.pdf) (Accessed: 30 September 2018).
- Reddy, M. M., Light, M. J. and Quinton, P. M. (1999) 'Activation of the epithelial Na<sup>+</sup> channel (ENaC) requires CFTR Cl<sup>-</sup> channel function', *Nature*, 402(6759), pp. 301–304. doi: 10.1038/46297.
- Reid, L., Meyrick, B., Antony, V. B., Chang, L. Y., Crapo, J. D. and Reynolds, H. Y. (2005) 'The mysterious pulmonary brush cell: A cell in search of a function', in *American Journal of Respiratory and Critical Care Medicine*, pp. 136–139. doi: 10.1164/rccm.200502-203WS.
- Reiter, J. F. and Leroux, M. R. (2017) 'Genes and molecular pathways underpinning ciliopathies', *Nature Reviews Molecular Cell Biology*, pp. 533–547. doi: 10.1038/nrm.2017.60.
- Ressler, B., Lee, R. T., Randell, S. H., Drazen, J. M. and Kamm, R. D. (2000) 'Molecular responses of rat tracheal epithelial cells to transmembrane pressure.', *American journal of physiology. Lung cellular and molecular physiology*, 278(6), pp. L1264–L1272. Available at: <http://www.ajplung.org> (Accessed: 7 September 2018).
- Revinski, D. R., Zaragosi, L.-E., Boutin, C., García Ruiz, S., Deprez, M., Pons, N., Marcet, B., Kodjabachian, L. and Barbry, P. (2018) 'CDC20B is required for deuterosome-mediated centriole production in multiciliated cells', *bioRxiv preprint*, in press, pp. 1–50. doi: <https://doi.org/10.1101/218750>.
- Reynolds, S. D., Hong, K. U., Giangreco, A., Mango, G. W., Guron, C., Morimoto, Y. and Stripp, B. R. (2000) 'Conditional Clara cell ablation reveals a self-renewing progenitor function of pulmonary neuroendocrine cells', *American Journal of Physiology-Lung Cellular and Molecular Physiology*. American Physiological Society Bethesda, MD, 278(6), pp. L1256–L1263. doi: 10.1152/ajplung.2000.278.6.L1256.
- Reynolds, S. D., Reynolds, P. R., Pryhuber, G. S., Finder, J. D. and Stripp, B. R. (2002) 'Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways', *American Journal of Respiratory and Critical Care Medicine*. American Thoracic Society, 166(11), pp. 1498–1509. doi: 10.1164/rccm.200204-285OC.
- Rida, P. C. ., Le Minh, N. and Jiang, Y.-J. (2004) 'A Notch feeling of somite segmentation and beyond', *Developmental Biology*. Academic Press, 265(1), pp. 2–22. doi: 10.1016/J.YDBIO.2003.07.003.

- Rock, J. R., Gao, X., Xue, Y., Randell, S. H., Kong, Y. Y. and Hogan, B. L. (2011) 'Notch-dependent differentiation of adult airway basal stem cells', *Cell Stem Cell*, 8(6), pp. 639–648. doi: 10.1016/j.stem.2011.04.003.
- Rock, J. R., Onaitis, M. W., Rawlins, E. L., Lu, Y., Clark, C. P., Xue, Y., Randell, S. H. and Hogan, B. L. M. (2009) 'Basal cells as stem cells of the mouse trachea and human airway epithelium.', *Proceedings of the National Academy of Sciences of the United States of America*, 106, pp. 12771–5. doi: 10.1073/pnas.0906850106.
- Rock, J. R., Rande, S. H. and Hogan, B. L. M. (2010) 'Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling', *Disease Models & Mechanisms*. doi: 10.1242/dmm.006031.
- Rodrigues-Martins, A., Riparbelli, M., Callaini, G., Glover, D. M. and Bettencourt-Dias, M. (2007) 'Revisiting the role of the mother centriole in centriole biogenesis.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 316(5827), pp. 1046–50. doi: 10.1126/science.1142950.
- Rogers, D. F. (1994a) 'Airway goblet cells: responsive and adaptable front-line defenders', *Eur Respir J*, 7, pp. 1690–1706. doi: 10.1183/09031936.94.07091678.
- Rogers, D. F. (1994b) 'Airway goblet cells: responsive and adaptable front-line defenders', *Eur Respir J*, 7(7), pp. 1690–1706. doi: 10.1183/09031936.94.07091678.
- Rogers, D. F. (2003) 'The airway goblet cell', *The International Journal of Biochemistry & Cell Biology*, 35, pp. 1–6. Available at: [https://ac-els-cdn-com.gate2.inist.fr/S1357272502000833/1-s2.0-S1357272502000833-main.pdf?\\_tid=9ad7c6c2-c3e3-11e7-9f7f-00000aacb35f&acdnat=1510076978\\_a2cfbec55d1fdb07c2639e5f55ebfb98](https://ac-els-cdn-com.gate2.inist.fr/S1357272502000833/1-s2.0-S1357272502000833-main.pdf?_tid=9ad7c6c2-c3e3-11e7-9f7f-00000aacb35f&acdnat=1510076978_a2cfbec55d1fdb07c2639e5f55ebfb98) (Accessed: 7 November 2017).
- Roy, M. G., Livraghi-Butrico, A., Fletcher, A. A., McElwee, M. M., Evans, S. E., Boerner, R. M., Alexander, S. N., Bellinghausen, L. K., Song, A. S., Petrova, Y. M., Tuvim, M. J., Adachi, R., Romo, I., Bordt, A. S., Bowden, M. G., Sisson, J. H., Woodruff, P. G., Thornton, D. J., Rousseau, K., De La Garza, M. M., Moghaddam, S. J., Karmouty-Quintana, H., Blackburn, M. R., Drouin, S. M., Davis, C. W., Terrell, K. A., Grubb, B. R., O'Neal, W. K., Flores, S. C., Cota-Gomez, A., Lozupone, C. A., Donnelly, J. M., Watson, A. M., Hennessy, C. E., Keith, R. C., Yang, I. V., Barthel, L., Henson, P. M., Janssen, W. J., Schwartz, D. A., Boucher, R. C., Dickey, B. F. and Evans, C. M. (2014) 'Muc5b is required for airway defence', *Nature*, 505(7483), pp. 412–416. doi: 10.1038/nature12807.
- Ruiz Garcia, S., Deprez, M., Lebrigand, K., Paquet, A., Cavard, A., Arguel, M.-J., Magnone, V., Caballero, I., Leroy, S., Marquette, C. H., Marcet, B., Barbry, P. and Zaragoza, L.-E. (2018) 'Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration', *bioRxiv*. Cold Spring Harbor Laboratory, p. 451807. doi: 10.1101/451807.

Sang, L., Miller, J. J., Corbit, K. C., Giles, R. H., Brauer, M. J., Otto, E. A., Baye, L. M., Wen, X., Scales, S. J., Kwong, M., Huntzicker, E. G., Sfakianos, M. K., Sandoval, W., Bazan, J. F., Kulkarni, P., Garcia-Gonzalo, F. R., Seol, A. D., O'Toole, J. F., Held, S., Reutter, H. M., Lane, W. S., Rafiq, M. A., Noor, A., Ansar, M., Devi, A. R. R., Sheffield, V. C., Slusarski, D. C., Vincent, J. B., Doherty, D. A., Hildebrandt, F., Reiter, J. F. and Jackson, P. K. (2011) 'Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways.', *Cell*. NIH Public Access, 145(4), pp. 513–28. doi: 10.1016/j.cell.2011.04.019.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T. and Ueda, H. R. (2013) 'Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity.', *Genome biology*. BioMed Central, 14(4), p. R31. doi: 10.1186/gb-2013-14-4-r31.

Sawamoto, K., Wichterle, H., Gonzalez-Perez, O., Cholfin, J. A., Yamada, M., Spassky, N., Murcia, N. S., Garcia-Verdugo, J. M., Marin, O., Rubenstein, J. L. R., Tessier-Lavigne, M., Okano, H. and Alvarez-Buylla, A. (2006) 'New neurons follow the flow of cerebrospinal fluid in the adult brain.', *Science (New York, N.Y.)*, 311(5761), pp. 629–32. doi: 10.1126/science.1119133.

Sedykh, I., TeSlaa, J. J., Tatarsky, R. L., Keller, A. N., Toops, K. A., Lakkaraju, A., Nyholm, M. K., Wolman, M. A. and Grinblat, Y. (2016) 'Novel roles for the radial spoke head protein 9 in neural and neurosensory cilia', *Scientific Reports*. Nature Publishing Group, 6(1), p. 34437. doi: 10.1038/srep34437.

Sehra, S., Yao, W., Nguyen, E. T., Ahyi, A.-N. N., Barbe Tuana, F. M., Ahlfeld, S. K., Snider, P., Tepper, R. S., Petrache, I., Conway, S. J. and Kaplan, M. H. (2011) 'Periostin Regulates Goblet Cell Metaplasia in a Model of Allergic Airway Inflammation', *The Journal of Immunology*, 186(8), pp. 4959–4966. doi: 10.4049/jimmunol.1002359.

Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K., Bendall, S., Friedman, N. and Pe'er, D. (2016) 'Wishbone identifies bifurcating developmental trajectories from single-cell data.', *Nature biotechnology*, 34(6), pp. 637–45. doi: 10.1038/nbt.3569.

Shu, X., Fry, A. M., Tulloch, B., Manson, F. D. C., Crabb, J. W., Khanna, H., Faragher, A. J., Lennon, A., He, S., Trojan, P., Giessl, A., Wolfrum, U., Vervoort, R., Swaroop, A. and Wright, A. F. (2005) 'RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin', *Human Molecular Genetics*. Oxford University Press, 14(9), pp. 1183–1197. doi: 10.1093/hmg/ddi129.

Sidhu, S. S., Yuan, S., Innes, A. L., Kerr, S., Woodruff, P. G., Hou, L., Muller, S. J. and Fahy, J. V (2010) 'Roles of epithelial cell-derived periostin in TGF- activation, collagen production, and collagen gel elasticity in asthma', *Proceedings of the National Academy of Sciences*, 107(32), pp. 14170–14175. doi: 10.1073/pnas.1009426107.

- Smirnova, N. F., Schamberger, A. C., Nayakanti, S., Hatz, R., Behr, J. and Eickelberg, O. (2016) 'Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs', *Respiratory Research*, 17(1), p. 83. doi: 10.1186/s12931-016-0404-x.
- Smirnova, N. F., Schamberger, A. C., Nayakanti, S., Hatz, R., Behr, J. and Eickelberg, O. (2016) 'Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs', *Respiratory Research*, 17(1). doi: 10.1186/s12931-016-0404-x.
- Song, H., Yao, E., Lin, C., Gacayan, R., Chen, M.-H. and Chuang, P.-T. (2012) 'Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis', *Proceedings of the National Academy of Sciences*, 109(43), pp. 17531–17536. doi: 10.1073/pnas.1207238109.
- Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J. S., Dixon, G., Harland, R., Wan, Y., Lishko, P., Lize, M., Kessel, M. and He, L. (2014) 'MiR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110', *Nature*, 510(7503), pp. 115–120. doi: 10.1038/nature13413.
- Sorokin, S. P. (1968) 'RECONSTRUCTIONS OF CENTRIOLE FORMATION AND CILIOGENESIS IN MAMMALIAN LUNGS', *J. Cell Sci*, 3, pp. 207–230. Available at: <http://jcs.biologists.org.gate2.inist.fr/content/joces/3/2/207.full.pdf> (Accessed: 22 January 2018).
- Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N. and Junker, J. P. (2018) 'Simultaneous lineage tracing and cell-type identification using CrIsPr-Cas9-induced genetic scars', *Nature Biotechnology*, 36(5), pp. 469–473. doi: 10.1038/nbt.4124.
- Spanjaard, B. and Junker, J. P. (2017) 'Methods for lineage tracing on the organism-wide level', *Current Opinion in Cell Biology*, 49, pp. 16–21. doi: 10.1016/j.ceb.2017.11.004.
- Spassky, N. and Meunier, A. (2017) 'The development and functions of multiciliated epithelia', *Nature Reviews Molecular Cell Biology*, 18(7), pp. 423–436. doi: 10.1038/nrm.2017.21.
- Spektor, A., Tsang, W. Y., Khoo, D. and Dynlacht, B. D. (2007) 'Cep97 and CP110 Suppress a Cilia Assembly Program', *Cell*, 130(4), pp. 678–690. doi: 10.1016/j.cell.2007.06.027.
- Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B., Smibert, P. and Satija, R. (2017) 'Cell & hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics', *bioRxiv*. Cold Spring Harbor Laboratory, p. 237693. doi: 10.1101/237693.

- Stoltz, D. A., Meyerholz, D. K., Pezzulo, A. A., Ramachandran, S., Rogan, M. P., Davis, G. J., Hanfland, R. A., Wohlford-Lenane, C., Dohrn, C. L., Bartlett, J. A., Nelson IV, G. A., Eugene, C., Taft, P. J., Ludwig, P. S., Estin, M., Hornick, E. E., Launspach, J. L., Samuel, M., Rokhlina, T., Karp, P. H., Ostedgaard, L. S., Uc, A., Starnes, T. D., Horswill, A. R., Brogden, K. A., Prather, R. S., Richter, S. S., Shilyansky, J., McCray, P. B., Zabner, J. and Welsh, M. J. (2010) 'Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth', *Science Translational Medicine*, 2(29), pp. 29–31. doi: 10.1126/scitranslmed.3000928.
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018) 'Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics', *BMC Genomics*, 19, p. 477. doi: 10.1186/s12864-018-4772-0.
- Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A. and Gönczy, P. (no date) 'Regulated HsSAS-6 Levels Ensure Formation of a Single Procentriole per Centriole during the Centrosome Duplication Cycle'. doi: 10.1016/j.devcel.2007.07.004.
- Stubbs, J. L., Vladoj, E. K., Axelrod, J. D. and Kintner, C. (2012) 'Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation', *Nature Cell Biology*, 14(2), pp. 140–147. doi: 10.1038/ncb2406.
- Stutts, M. J., Rossier, B. C. and Boucher, R. C. (1997) 'Cystic fibrosis transmembrane conductance regulator inverts protein kinase A-mediated regulation of epithelial sodium channel single channel kinetics.', *The Journal of biological chemistry*, 272(22), pp. 14037–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9162024> (Accessed: 29 October 2018).
- Svensson, V., Vento-Tormo, R. and Teichmann, S. A. (2018) 'Exponential scaling of single-cell RNA-seq in the past decade', *Nature Protocols*, pp. 599–604. doi: 10.1038/nprot.2017.149.
- Swartz, M. A., Tschumperlin, D. J., Kamm, R. D. and Drazen, J. M. (2001) 'Mechanical stress is communicated between different cell types to elicit matrix remodeling.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(11), pp. 6180–6185. doi: 10.1073/pnas.111133298.
- Syed, F., Huang, C. C., Li, K., Liu, V., Shang, T., Amegadzie, B. Y., Griswold, D. E., Song, X.-Y. R. and Li, L. (2007) 'Identification of interleukin-13 related biomarkers using peripheral blood mononuclear cells', *Biomarkers*, 12(4), pp. 414–423. doi: 10.1080/13547500701192652.
- Tadokoro, T., Gao, X., Hong, C. C., Hotten, D. and Hogan, B. L. M. (2016) 'BMP signaling and cellular dynamics during regeneration of airway epithelium from basal progenitors', *Development*, 143(5), pp. 764–773. doi: 10.1242/dev.126656.

- Takahashi, H. and Ikeda, T. (1996) 'Transcripts for two members of the transforming growth factor- $\beta$  superfamily BMP-3 and BMP-7 are expressed in developing rat embryos', *Developmental Dynamics*, 207(4), pp. 439–449. doi: 10.1002/(SICI)1097-0177(199612)207:4<439::AID-AJA8>3.0.CO;2-I.
- Takayama, G., Arima, K., Kanaji, T., Toda, S., Tanaka, H., Shoji, S., McKenzie, A. N. J., Nagai, H., Hotokebuchi, T. and Izuhara, K. (2006) 'Periostin: A novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals', *Journal of Allergy and Clinical Immunology*, 118(1), pp. 98–104. doi: 10.1016/j.jaci.2006.02.046.
- Takeyama, K., Dabbagh, K., Lee, H.-M., Agusti, C., Lausier, J. A., Ueki, I. F., Grattan, K. M. and Nadel, J. A. (1999) *Epidermal growth factor system regulates mucin production in airways*, *Medical Sciences*. Available at: [www.pnas.org](http://www.pnas.org). (Accessed: 19 August 2018).
- Takeyama, K., Jung, B., Shim, J. J., Burgel, P.-R., Dao-Pick, T., Ueki, I. F., Protin, U., Kroschel, P. and Nadel, J. A. (2001) 'Activation of epidermal growth factor receptors is responsible for mucin synthesis induced by cigarette smoke', *Am J Physiol Lung Cell Mol Physiol*, 280(1), pp. L165-172. Available at: <http://www.ajplung.org> (Accessed: 3 September 2018).
- Tan, F. E., Vldar, E. K., Ma, L., Fuentealba, L. C., Hoh, R., Espinoza, F. H., Axelrod, J. D., Alvarez-Buylla, A., Stearns, T., Kintner, C. and Krasnow, M. A. (2013) 'Myb promotes centriole amplification and later steps of the multiciliogenesis program', *Development*. Company of Biologists, 140(20), pp. 4277–4286. doi: 10.1242/dev.094102.
- Tang, C.-J. C., Lin, S.-Y., Hsu, W.-B., Lin, Y.-N., Wu, C.-T., Lin, Y.-C., Chang, C.-W., Wu, K.-S. and Tang, T. K. (2011) 'The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation', *The EMBO Journal*, 30, pp. 4790–4804. doi: 10.1038/emboj.2011.378.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K. and Surani, M. A. (2009) 'mRNA-Seq whole-transcriptome analysis of a single cell', *Nature Methods*, 6(5), pp. 377–382. doi: 10.1038/nmeth.1315.
- Tang, K. Tang (2013) 'Centriole biogenesis in multiciliated cells', *Nature Publishing Group*, 15(4), pp. 10750–10758. doi: 10.1038/ncb2892.
- Tata, P. R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B. M., Vinarsky, V., Cho, J. L., Breton, S., Sahay, A., Medoff, B. D. and Rajagopal, J. (2013) 'Dedifferentiation of committed epithelial cells into stem cells in vivo', *Nature*. Nature Publishing Group, 503(7475), pp. 218–223. doi: 10.1038/nature12777.

- Tata, P. R. and Rajagopal, J. (2017) 'Plasticity in the lung: making and breaking cell identity', *Development*, 144(5), pp. 755–766. doi: 10.1242/dev.143784.
- Teixeira, V. H., Nadarajan, P., Graham, T. A., Pipinikas, C. P., Brown, J. M., Falzon, M., Nye, E., Poulosom, R., Lawrence, D., Wright, N. A., McDonald, S., Giangreco, A., Simons, B. D. and Janes, S. M. (2013) 'Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors', *eLife*, 2. doi: 10.7554/eLife.00966.001.
- Terré, B., Lewis, M., Gil-gómez, G. and Stracker, T. H. (2018) 'GEMC1 and CCNO are required for efferent duct development and male fertility', *bioRxiv*. doi: 10.1101/258418.
- Terré, B., Piergiovanni, G., Segura-Bayona, S., Gil-Gómez, G., Youssef, S. A., Attolini, C. S., Wilsch-Bräuninger, M., Jung, C., Rojas, A. M., Marjanović, M., Knobel, P. A., Palenzuela, L., López-Rovira, T., Forrow, S., Huttner, W. B., Valverde, M. A., de Bruin, A., Costanzo, V. and Stracker, T. H. (2016) 'GEMC1 is a critical regulator of multiciliated cell differentiation.', *Embo J*, 35(9), pp. 942–960. doi: 10.15252/embj.201592821.
- Thomas, J., Morlé, L., Soulavie, F., Laurençon, A., Sagnol, S. and Durand, B. (2010) 'Transcriptional control of genes involved in ciliogenesis: a first step in making cilia.', *Biology of the cell / under the auspices of the European Cell Biology Organization*, 102, pp. 499–513. doi: 10.1042/BC20100035.
- Tözser, J., Earwood, R., Kato, A., Brown, J., Tanaka, K., Didier, R., Megraw, T. L., Blum, M. and Kato, Y. (2015) 'TGF- $\beta$  Signaling Regulates the Differentiation of Motile Cilia', *Cell Reports*, 11(7), pp. 1000–1007. doi: 10.1016/j.celrep.2015.04.025.
- Trapnell, C. (2015a) 'Defining cell types and states with single-cell genomics', pp. 1491–1498. doi: 10.1101/gr.190595.115.
- Trapnell, C. (2015b) 'Defining cell types and states with single-cell genomics', *Genome Research*, pp. 1491–1498. doi: 10.1101/gr.190595.115.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014) 'The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells', *Nature Biotechnology*, 32. doi: 10.1038/nbt.2859.
- Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. a and Quake, S. R. (2014) 'Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq', *Nature*. Nature Publishing Group, 509(7500), pp. 371–375. doi: 10.1038/nature13173.

- Tsang, J. C. H., Yu, Y., Burke, S., Buettner, F., Wang, C., Kolodziejczyk, A. A., Teichmann, S. A., Lu, L. and Liu, P. (2015) 'Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells', *Genome Biology*, 16(1), p. 178. doi: 10.1186/s13059-015-0739-5.
- Tsang, W. Y. and Dynlacht, B. D. (2013) 'CP110 and its network of partners coordinately regulate cilia assembly.', *Cilia*, 2(1), p. 9. doi: 10.1186/2046-2530-2-9.
- Tsao, P.-N., Vasconcelos, M., Izvolsky, K. I., Qian, J., Lu, J. and Cardoso, W. V (2009) 'Notch signaling controls the balance of ciliated and secretory cell fates in developing airways', *Development*. Company of Biologists, 136(13), pp. 2297–2307. doi: 10.1242/dev.034884.
- Tschumperlin, D. J., Dal, G., Maly, I. V., Kikuchi, T., Lalho, L. H., McVittie, A. K., Haley, K. J., Lilly, C. M., So, P. T. C., Lauffenburger, D. A., Kamm, R. D. and Drazen, J. M. (2004) 'Mechanotransduction through growth-factor shedding into the extracellular space', *Nature*, 429(6987), pp. 83–86. doi: 10.1038/nature02543.
- Tsou, Y.-A., Lin, C.-D., Chen, H.-C., Hsu, H.-Y., Wu, L.-T., Chiang-Ni, C., Chen, C.-J., Wu, T.-F., Kao, M.-C., Chen, Y.-A., Peng, M.-T., Tsai, M.-H., Chen, C.-M. and Lai, C.-H. (2015) 'Interleukin-13 Inhibits Lipopolysaccharide-Induced BPIFA1 Expression in Nasal Epithelial Cells', *PLOS ONE*, 10(12), p. e0143484. doi: 10.1371/journal.pone.0143484.
- Turner, J., Roger, J., Fitau, J., Combe, D., Giddings, J., Heeke, G. Van and Jones, C. E. (2011) 'Goblet cells are derived from a FOXJ1-expressing progenitor in a human airway epithelium.', *American journal of respiratory cell and molecular biology*, 44(3), pp. 276–84. doi: 10.1165/rcmb.2009-0304OC.
- Tyner, J. W., Kim, E. Y., Ide, K., Pelletier, M. R., Roswit, W. T., Morton, J. D., Battaile, J. T., Patel, A. C., Patterson, G. A., Castro, M., Spoor, M. S., You, Y., Brody, S. L. and Holtzman, M. J. (2006) 'Blocking airway mucous cell metaplasia by inhibiting EGFR antiapoptosis and IL-13 transdifferentiation signals', *Journal of Clinical Investigation*, 116(2), pp. 309–321. doi: 10.1172/JCI25167.
- Verhamme, F. M., Seys, L. J. M., De Smet, E. G., Provoost, S., Janssens, W., Elewaut, D., Joos, G. F., Brusselle, G. G. and Bracke, K. R. (2017) 'Elevated GDF-15 contributes to pulmonary inflammation upon cigarette smoke exposure.', *Mucosal immunology*, 10(6), pp. 1400–1411. doi: 10.1038/mi.2017.3.
- Vidarsson, H., Westergren, R., Heglind, M., Blomqvist, S. R., Breton, S. and Enerbäck, S. (2009) 'The forkhead transcription factor Foxi1 is a master regulator of vacuolar H<sup>+</sup>-ATPase proton pump subunits in the inner ear, kidney and epididymis', *PLoS ONE*, 4(2). doi: 10.1371/journal.pone.0004471.

- Di Vincenzo, S., Heijink, I. H., Noordhoek, J. A., Cipollina, C., Siena, L., Bruno, A., Ferraro, M., Postma, D. S., Gjomarkaj, M. and Pace, E. (2018) 'SIRT1/FoxO3 axis alteration leads to aberrant immune responses in bronchial epithelial cells.', *Journal of cellular and molecular medicine*, 22(4), pp. 2272–2282. doi: 10.1111/jcmm.13509.
- Volckaert, T. and De Langhe, S. (2014) *Lung epithelial stem cells and their niches: Fgf10 takes center stage, Fibrogenesis & Tissue Repair*. doi: 10.1186/1755-1536-7-8.
- Vukicevic, S., Helder, M. N. and Luyten, F. P. (1994) 'Developing human lung and kidney are major sites for synthesis of bone morphogenetic protein-3 (osteogenin).', *Journal of Histochemistry & Cytochemistry*, 42(7), pp. 869–875. doi: 10.1177/42.7.8014470.
- Vulprecht, J., David, A., Tibelius, A., Castiel, A., Konotop, G., Liu, F., Bestvater, F., Raab, M. S., Zentgraf, H., Izraeli, S. and Kramer, A. (2012) 'STIL is required for centriole duplication in human cells', *Journal of Cell Science*, 125(5), pp. 1353–1362. doi: 10.1242/jcs.104109.
- Walentek, P. and Quigley, I. K. (2017) 'What we can learn from a tadpole about ciliopathies and airway diseases: Using systems biology in *Xenopus* to study cilia and mucociliary epithelia.', *Genesis (New York, N.Y. : 2000)*. NIH Public Access, 55(1–2). doi: 10.1002/dvg.23001.
- Walker, J. A. and McKenzie, A. N. (2017) 'TH2 cell development and function'. doi: 10.1038/nri.2017.118.
- Wallmeier, J., Al-Mutairi, D. A., Chen, C.-T., Loges, N. T., Pennekamp, P., Menchen, T., Ma, L., Shamseldin, H. E., Olbrich, H., Dougherty, G. W., Werner, C., Alsabab, B. H., Köhler, G., Jaspers, M., Boon, M., Griese, M., Schmitt-Grohé, S., Zimmermann, T., Koerner-Rettberg, C., Horak, E., Kintner, C., Alkuraya, F. S. and Omran, H. (2014) 'Mutations in CCNO result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia', *Nature Genetics*, 46(6), pp. 646–651. doi: 10.1038/ng.2961.
- Wan, H., Kaestner, K. H., Ang, S.-L., Ikegami, M., Finkelman, F. D., Stahlman, M. T., Fulkerson, P. C., Rothenberg, M. E. and Whitsett, J. A. (2004) 'Foxa2 regulates alveolarization and goblet cell hyperplasia.', *Development (Cambridge, England)*, 131(4), pp. 953–64. doi: 10.1242/dev.00966.
- Wang, W.-J., Tay, H. G., Soni, R., Perumal, G. S., Goll, M. G., Macaluso, F. P., Asara, J. M., Amack, J. D. and Tsou, M.-F. B. (2013) 'CEP162 is an axoneme-recognition protein promoting ciliary transition zone assembly at the cilia base.', *Nature cell biology*. NIH Public Access, 15(6), pp. 591–601. doi: 10.1038/ncb2739.

- Wang, Y., Waters, J., Leung, M. L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., Multani, A., Zhang, H., Zhao, R., Michor, F., Meric-Bernstam, F. and Navin, N. E. (2014) 'Clonal evolution in breast cancer revealed by single nucleus genome sequencing', *Nature*, 512(7513), pp. 155–160. doi: 10.1038/nature13600.
- Watson, J. K., Rulands, S., Wilkinson, A. C., Wuidart, A., Ousset, M., Van Keymeulen, A., Göttgens, B., Blanpain, C., Simons, B. D. and Rawlins, E. L. (2015) 'Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium', *Cell Reports*, 12, pp. 90–101. doi: 10.1016/j.celrep.2015.06.011.
- Wenzel, S. E. (2012) 'Asthma phenotypes: the evolution from clinical to molecular approaches', *Nature Medicine*, 18(5). doi: 10.1038/nm.2678.
- Wine, J. J. (2004) 'Submucosal Glands and Airway Defense', *Proceedings of the American Thoracic Society*, 1(1), pp. 47–53. doi: 10.1513/pats.2306015.
- Woodruff, P. G., Boushey, H. A., Dolganov, G. M., Barker, C. S., Yang, Y. H., Donnelly, S., Ellwanger, A., Sidhu, S. S., Dao-Pick, T. P., Pantoja, C., Erle, D. J., Yamamoto, K. R. and Fahy, J. V. (2007) 'Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids', *Proceedings of the National Academy of Sciences*, 104(40), pp. 15858–15863. doi: 10.1073/pnas.0707413104.
- Woodruff, P. G., Modrek, B., Choy, D. F., Jia, G., Abbas, A. R., Ellwanger, A., Arron, J. R., Koth, L. L. and Fahy, J. V. (2009) 'T-helper type 2-driven inflammation defines major subphenotypes of asthma', *American Journal of Respiratory and Critical Care Medicine*, 180(5), pp. 388–395. doi: 10.1164/rccm.200903-0392OC.
- Wu, Q., Jiang, D. and Chu, H. W. (2012) 'Cigarette smoke induces growth differentiation factor 15 production in human lung epithelial cells: implication in mucin over-expression.', *Innate immunity*, 18(4), pp. 617–26. doi: 10.1177/1753425911429837.
- Wu, Q., Jiang, D., Matsuda, J. L., Ternyak, K., Zhang, B. and Chu, H. W. (2016) 'Cigarette Smoke Induces Human Airway Epithelial Senescence via Growth Differentiation Factor 15 Production.', *American journal of respiratory cell and molecular biology*, 55(3), pp. 429–38. doi: 10.1165/rcmb.2015-0143OC.
- Wu, Y.-P., Cao, C., Wu, Y.-F., Li, M., Lai, T.-W., Zhu, C., Wang, Y., Ying, S.-M., Chen, Z.-H., Shen, H.-H. and Li, W. (2017) 'Activating transcription factor 3 represses cigarette smoke-induced IL6 and IL8 expression via suppressing NF-κB activation.', *Toxicology letters*, 270, pp. 17–24. doi: 10.1016/j.toxlet.2017.02.002.

- Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang, J., Wikenheiser-Brokamp, K. A., Perl, A.-K. T., Funari, V. A., Gokey, J. J., Stripp, B. R. and Whitsett, J. A. (2016) 'Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis.', *JCI insight*. American Society for Clinical Investigation, 1(20), p. e90558. doi: 10.1172/jci.insight.90558.
- Yadav, S. P., Sharma, N. K., Liu, C., Dong, L., Li, T. and Swaroop, A. (2016) 'Centrosomal protein CP110 controls maturation of the mother centriole during cilia biogenesis.', *Development (Cambridge, England)*, 143(9), pp. 1491–501. doi: 10.1242/dev.130120.
- Yaghi, A., Zaman, A., Cox, G. and Dolovich, M. B. (2012) 'Ciliary beating is depressed in nasal cilia from chronic obstructive pulmonary disease subjects', *Respiratory Medicine*, 106(8), pp. 1139–1147. doi: 10.1016/j.rmed.2012.04.001.
- Yang, L.-T., Nichols, J. T., Yao, C., Manilay, J. O., Robey, E. A. and Weinmaster, G. (2005) 'Fringe Glycosyltransferases Differentially Modulate Notch1 Proteolysis Induced by Delta1 and Jagged1', *Molecular Biology of the Cell*, 16, pp. 927–942. doi: 10.1091/mbc.
- Yang, Q. and Ferrell, J. E. (2013) 'The Cdk1–APC/C cell cycle oscillator circuit functions as a time-delayed, ultrasensitive switch', *Nature Cell Biology*. Nature Publishing Group, 15(5), pp. 519–525. doi: 10.1038/ncb2737.
- Ye, Y. M., Nahm, D. H., Kim, C. W., Kim, H. R., Hong, C. S., Park, C. S., Suh, C. H. and Park, H. S. (2006) 'Cytokeratin autoantibodies: Useful serologic markers for toluene diisocyanate-induced asthma', *Yonsei Medical Journal*, 47(6), pp. 773–781. doi: 10.3349/ymj.2006.47.6.773.
- Yeh, T. H., Lee, S. Y. and Hsu, W. C. (2010) 'Expression of SPLUNC1 protein in nasal polyp epithelial cells in air-liquid interface culture treated with IL-13', *American Journal of Rhinology and Allergy*, 24(1), pp. 17–20. doi: 10.2500/ajra.2010.24.3381.
- Yi, L., Cheng, D., Zhang, K., Huo, X., Mo, Y., Shi, H., Di, H., Zou, Y., Zhang, H., Zhao, J., Xu, Y., Erle, D. J. and Zhen, G. (2017) 'Intelectin contributes to allergen-induced IL-25, IL-33 and TSLP expression and type 2 response in asthma and atopic dermatitis HHS Public Access', *Mucosal Immunol*, 10(6), pp. 1491–1503. doi: 10.1038/mi.2017.10.
- YIN, B., ZHANG, M., ZENG, Y., LI, Y., ZHANG, C., SONG, Y. and Song, Y. (2016) 'Downregulation of cytokeratin 18 is associated with paclitaxel-resistance and tumor aggressiveness in prostate cancer', *International Journal of Oncology*, 48(4), pp. 1730–1736. doi: 10.3892/ijo.2016.3396.

- Ying, Q.-L., Nichols, J., Chambers, I. and Smith, A. (2003) 'BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3', *Cell*. Cell Press, 115(3), pp. 281–292. doi: 10.1016/S0092-8674(03)00847-X.
- Ylivinkka, I., Hu, Y., Chen, P., Rantanen, V., Hautaniemi, S., Nyman, T. A., Keski-Oja, J. and Hyytiäinen, M. (2013) 'Netrin-1-induced activation of Notch signaling mediates glioblastoma cell invasion', *Journal of Cell Science*. The Company of Biologists Ltd, 126(11), pp. 2459–2469. doi: 10.1242/jcs.120022.
- Ylivinkka, I., Keski-Oja, J. and Hyytiäinen, M. (2016) 'Netrin-1: A regulator of cancer cell motility?', *European Journal of Cell Biology*. Urban & Fischer, pp. 513–520. doi: 10.1016/j.ejcb.2016.10.002.
- Yu, X., Ng, C. P., Habacher, H. and Roy, S. (2008) 'Foxj1 transcription factors are master regulators of the motile ciliogenic program', *Nature Genetics*, 40(12), pp. 1445–1453. doi: 10.1038/ng.263.
- El Zein, L., Ait-Lounis, A., Morle, L., Thomas, J., Chhin, B., Spassky, N., Reith, W. and Durand, B. (2009) 'RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies', *Journal of Cell Science*, 122(17), pp. 3180–3189. doi: 10.1242/jcs.048348.
- Zeisel, A., M̄oz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., Manno, G. La, Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J. and Linnarsson, S. (2015) 'Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq', *Science*, 347(6226), pp. 1138–1142. doi: 10.1126/science.aaa1934.
- Zepp, J. A., Zacharias, W. J., Frank, D. B., Cavanaugh, C. A., Zhou, S., Morley, M. P. and Morrisey, E. E. (2017) 'Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung', *Cell*. Elsevier Inc, 170(6), p. 1134–1148.e10. doi: 10.1016/j.cell.2017.07.034.
- Zhang, S., Chung, W.-C., Wu, G., Egan, S. E. and Xu, K. (2014) 'Tumor-Suppressive Activity of Lunatic Fringe in Prostate through Differential Modulation of Notch Receptor Activation', *Neoplasia*, 16(2), pp. 158–167. doi: 10.1593/neo.131870.
- Zhang, W., Huang, Y., Wu, Y. and Gunst, S. J. (2015) 'A novel role for RhoA GTPase in the regulation of airway smooth muscle contraction', *Canadian journal of physiology and pharmacology*, 93(2), pp. 129–136. doi: 10.1139/cjpp-2014-0388.
- Zhao, H., Zhu, L., Zhu, Y., Cao, J., Li, S., Huang, Q., Xu, T., Huang, X., Yan, X. and Zhu, X. (2013) 'The Cep63 paralogue Deup1 enables massive de novo centriole biogenesis for vertebrate multiciliogenesis.', *Nature cell biology*. Nature Publishing Group, 15(12), pp. 1434–44. doi: 10.1038/ncb2880.

- Zhao, J., Minami, Y., Etling, E., Coleman, J. M., Lauder, S. N., Tyrrell, V., Aldrovandi, M., O'Donnell, V., Claesson, H. E., Kagan, V. and Wenzel, S. (2017) 'Preferential generation of 15-HETE-PE induced by IL-13 regulates goblet cell differentiation in human airway epithelial cells', *American Journal of Respiratory Cell and Molecular Biology*, 57(6), pp. 692–701. doi: 10.1165/rcmb.2017-0031OC.
- Zhen, G., Sung, W. P., Nguyenvu, L. T., Rodriguez, M. W., Barbeau, R., Paquet, A. C. and Erie, D. J. (2007) 'IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production', *American Journal of Respiratory Cell and Molecular Biology*, 36(2), pp. 244–253. doi: 10.1165/rcmb.2006-0180OC.
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Shuga, J., Montesclaros, L., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Ness, K. D., Beppu, L. W., Joachim Deeg, H., McFarland, C., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Hindson, B. J. and Bielas, J. H. (2016) 'Massively parallel digital transcriptional profiling of single cells 1 2', *Phone*. Cold Spring Harbor Labs Journals, (206), pp. 667–3170. doi: 10.1101/065912.
- Zhou-Suckow, Z., Duerr, J., Hagner, M. and Mall, M. A. (2017) 'Airway mucus, inflammation and remodeling: emerging links in the pathogenesis of chronic lung diseases', *Cell and Tissue Research*, 367(3), pp. 537–550. doi: 10.1007/s00441-016-2562-z.
- Zhou, F., Narasimhan, V., Shboul, M., Chong, Y. L., Reversade, B. and Roy, S. (2015) 'Gmnc Is a Master Regulator of the Multiciliated Cell Differentiation Program', *Current Biology*, 25(24), pp. 3267–3273. doi: 10.1016/j.cub.2015.10.062.
- Zhou, M., Cui, Z. lei, Guo, X. jun, Ren, L. pin, Yang, M., Fan, Z. wen, Han, R. chao and Xu, W. guo (2015) 'Blockade of Notch Signalling by  $\gamma$ -Secretase Inhibitor in Lung T Cells of Asthmatic Mice Affects T Cell Differentiation and Pulmonary Inflammation', *Inflammation*, 38(3), pp. 1281–1288. doi: 10.1007/s10753-014-0098-5.
- Zuo, W., Zhang, T., Wu, D. Z., Guan, S. P., Liew, A.-A., Yamamoto, Y., Wang, X., Lim, S. J., Vincent, M., Lessard, M., Crum, C. P., Xian, W. and McKeon, F. (2014) 'p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration.', *Nature*. Nature Publishing Group, 517(7536), pp. 616–620. doi: 10.1038/nature13903.

## *XI. ANNEX*

---

## XI.1 LIST OF PUBLICATIONS

Marie-Jeanne Arguel, Kevin LeBrigand, Agnès Paquet, **Sandra Ruiz García**, Laure-Emmanuelle Zaragosi, Pascal Barbry\* and Rainer WaldmannArguel (2017) ‘A cost effective 5', selective single cell transcriptome profiling approach with improved UMI design', *Nucleic Acids Research*. Oxford University Press, 45(7), p. gkw1242. doi: 10.1093/nar/gkw1242.

Revinski, D. R., Zaragosi, Laure-Emmanuelle, Boutin, Camille, **Sandra Ruiz-Garcia**, Marie Deprez, Virginie Thomé, Olivier Rosnet, Anne-Sophie Gay, Olivier Mercey, Agnès Paquet, Nicolas Pons, Gilles Ponzio, Brice Marcet, Laurent Kodjabachian & Pascal Barbry (2018) ‘CDC20B is required for deuterosome-mediated centriole production in multiciliated cells', *bioRxiv preprint*, in press, pp. 1–50. doi: <https://doi.org/10.1101/218750>.

**Sandra Ruiz Garcia**, Marie Deprez, Kevin Lebrigand, Agnès Paquet, Amélie Cavard, Marie-Jeanne Arguel, Virginie Magnone, Ignacio Caballero, Sylvie Leroy, Charles-Hugo Marquette, Brice Marcet, Pascal Barbry, Laure-Emmanuelle Zaragosi (2018) ‘Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration', *bioRxiv*. Cold Spring Harbor Laboratory, p. 451807. doi: 10.1101/451807.

*XII. A cost effective 5', selective single cell transcriptome profiling approach with improved UMI design*

---

---

# A cost effective 5' selective single cell transcriptome profiling approach with improved UMI design

Marie-Jeanne Arguel, Kevin LeBrigand, Agnès Paquet, Sandra Ruiz García, Laure-Emmanuelle Zaragosi, Pascal Barbry\* and Rainer Waldmann

Université Côte d'Azur, CNRS, Institut de Pharmacologie Moléculaire et Cellulaire, F06560 Sophia Antipolis, France

Received April 22, 2016; Revised November 14, 2016; Editorial Decision November 25, 2016; Accepted November 28, 2016

## ABSTRACT

Single cell RNA sequencing approaches are instrumental in studies of cell-to-cell variability. 5' selective transcriptome profiling approaches allow simultaneous definition of the transcription start size and have advantages over 3' selective approaches which just provide internal sequences close to the 3' end. The only currently existing 5' selective approach requires costly and labor intensive fragmentation and cell barcoding after cDNA amplification. We developed an optimized 5' selective workflow where all the cell indexing is done prior to fragmentation. With our protocol, cell indexing can be performed in the Fluidigm C1 microfluidic device, resulting in a significant reduction of cost and labor. We also designed optimized unique molecular identifiers that show less sequence bias and vulnerability towards sequencing errors resulting in an improved accuracy of molecule counting. We provide comprehensive experimental workflows for Illumina and Ion Proton sequencers that allow single cell sequencing in a cost range comparable to qPCR assays.

## INTRODUCTION

The cell is the minimal building block of any living organism. Investigating the properties of individual cells rather than the average of a group of seemingly identical cells provided important insights in various domains such as cancer (1), development (2,3), immunology (4) and neurobiology (5–7). Single cell transcriptome sequencing is a key technology to address this cellular heterogeneity.

Since the first sequencing of a single cell transcriptome (8), advances in library preparation techniques greatly improved both efficiency and throughput (9).

Most mammalian cells contain just a few hundred thousand mRNA molecules (10). In consequence, efficient conversion of mRNA into cDNA is crucial and was the focus of several recent studies (1,10,11). Most current single cell

mRNA cloning techniques exploit the template switching activity of reverse transcriptases (STRT-seq (12), Smart-seq (1)) to efficiently clone full length cDNA, which is subsequently amplified by PCR. The approach was further refined by Picelli *et al.* (11) (Smart-Seq2) and Islam *et al.* (10). Alternate approaches that use isothermal cRNA amplification were also developed (Mars-seq (13), CEL-seq (14,15)).

Early single cell profiling approaches processed single cells in tubes or in plates. Performing cDNA synthesis in tiny volumes either in microfluidic devices such as the Fluidigm C1 (2,3,10) or in microdroplets (16,17) was an important further development which increased throughput and reduced both reagent cost and labor. Single cell transcriptome profiling in microfluidic devices was also shown to yield increased transcript discovery rates and thus mRNA cloning efficiencies when compared to manual processing in tubes (18).

Amplification bias and library complexity are clearly issues that need to be considered since single cell library preparation requires huge amplification of tiny amounts (< 1pg) of cDNA. To address those issues, Islam *et al.* (10) and Jaitin *et al.* (13) stochastically tagged cDNA molecules during reverse transcription with short random nucleotide sequences (unique molecular identifiers, UMIs). The use of UMIs largely improved data quality since it allows counting of the initial unamplified cDNA molecules what is much less biased than counting transcript read numbers in heavily amplified cDNA.

UMIs are introduced during reverse transcription either at the 5' or 3' end of the cDNA. In consequence, only the UMI tagged extremity of the transcript is recovered and sequenced after cDNA fragmentation. The vast majority of currently used approaches introduce the UMI via the oligo-dT reverse transcription primer and sequence the 3' terminal, UMI tagged, fragment of the cDNA (14–17). Sequencing the actual 3' end of a transcript would require sequencing through the poly-dT stretch of the reverse transcription primer. Sequencing through such long repeats typically yields poor read qualities due to phasing issues and homopolymer length heterogeneity within flow cell clusters generated by polymerase slipping during amplification. In

\*To whom correspondence should be addressed. Tel: +334 93 95 77 00; Fax: +334 93 95 77 94; Email: barbry@ipmc.cnrs.fr

consequence, 3' selective single cell sequencing approaches don't sequence the actual 3' end of a transcript but rather the 5' end of the most 3' fragment of the cDNA obtained after cDNA fragmentation.

Conversely, 5' selective approaches do not have this limitation and allow sequencing of the actual 5' end of the cDNA and thus a simultaneous definition of mRNA expression levels and transcription start sites. However, the only currently published 5' selective approach (10) has essentially two drawbacks: First, the cell index is introduced during cDNA fragmentation with indexed transposons independently for each cell, what is labor intensive and costly when commercial transposase is used. Secondly, only short UMIs of 5 nucleotides were used, since extension of the template switching oligonucleotide is thought to adversely affect mRNA capture efficiency. However such short UMIs get saturated for highly expressed transcripts and efficient UMI sequencing error correction strategies (16) cannot be used, since they would further decrease an already low UMI complexity.

Another limitation of all current efficient single cell sequencing approaches is that they are only available for Illumina sequencers and not for other low cost benchtop sequencers such as the Ion Proton.

We addressed those issues and developed a highly efficient cost- and labor effective 5' selective single cell transcriptome profiling approach for both Ion Torrent and Illumina sequencers. Our method introduces cell barcodes by PCR prior to cDNA fragmentation and requires just one fragmentation and library preparation for the pooled cDNAs from the individual cells. We also present a novel UMI design that allows better error correction and thus more reliable molecule counting. We show that our barcoding strategy and UMI design allows robust and efficient single cell transcriptome profiling with the Fluidigm C1 at just a fraction of the cost of currently available commercial approaches.

## MATERIALS AND METHODS

### Cell culture

HEK293 cells were cultured in DMEM medium supplemented with glutamine and 10% fetal calf serum. Human airway epithelial cells were isolated and cultured as described by Marcet *et al.* (19).

### cDNA synthesis, PCR amplification—tube controls

Concentrations of reagents and enzymes as well as the reagent volume per cell were identical for tube controls and the Fluidigm C1 microfluidic device. Primer sequences are listed in Supplementary Table S1. A schematic workflow is provided in Supplementary Figure S11.

### Cell lysis

1000 HEK293 cells in 4.5  $\mu$ l C1 wash buffer (Fluidigm) were lysed with 9  $\mu$ l of lysis buffer (0.2% w/v Tween 20, 1 U/ $\mu$ l Promega RNasin RNase inhibitor, 2  $\mu$ M reverse transcription primer, 2.5 mM dNTPs, 1 $\times$  C1 loading reagent

(Fluidigm), ERCC Spike-In Mix 1 at 20 000 molecules/cell (Life Technologies). The sample was incubated for 10 min at room temperature followed by 3 min at 70°C and 3 min at 10°C.

### Reverse transcription

Reverse transcription was adapted from (20). 18  $\mu$ l of 1.75 $\times$  reverse transcription buffer (Thermo), 8.75 mM DTT, 1.75 M betaine, 10.5 mM MgCl<sub>2</sub>, 1.75  $\mu$ M template switching oligonucleotide, 0.5 U/ $\mu$ l Promega RNasin, 5.5 U/ $\mu$ l Life Technologies Superscript II reverse transcriptase, 1 $\times$  C1 loading reagent were added to the lysed cells and the sample was incubated for 10 min at 25°C, 90 min at 42°C, 15 min at 70°C and kept <10°C until PCR amplification.

### PCR amplification

One tenth of the reverse transcription (3.15  $\mu$ l, 100 cells) was mixed with 27  $\mu$ l 1.15 $\times$  KAPA HiFi HotStart Ready Mix, 55 nM barcode primer and 1.1  $\mu$ M biotinylated PCR primer. For Illumina sequencing forward and reverse primers for this PCR are distinct and an additional reverse PCR primer was added (1.1  $\mu$ M). For PCR amplification samples were incubated 3 min at 98°C followed by 18 cycles at 98°C for 20 s, 64°C for 15 s, 72°C for 6 min, and a final extension at 72°C for 5 min. Primers and small fragments were removed by cleanup with 1 vol. SpriSelect beads. Typical cDNA size distributions are shown in Supplementary Figure S13.

### cDNA synthesis, PCR amplification – microfluidic device

Lysis, reverse transcription and PCR mixes were the same as for the tube controls. Lysis (7  $\mu$ l) and reverse transcription mix (8  $\mu$ l) were added to the wells of the microfluidic device specified in the script for the C1. 6.5  $\mu$ l of PCR mix with one of the 96 barcodes was added to each of the outlet wells. The PCR mixes were backloaded from the outlet wells into the reaction chambers resulting in cell specific barcoding on the microfluidic chip during the PCR amplification. The script for the Fluidigm C1 has been submitted to the Fluidigm C1 OpenApp script repository (<https://www.fluidigm.com/c1openapp>).

### Library preparation

Since Ion Torrent requires smaller fragment sizes than Illumina sequencers and no suitable commercial kits were available, we adapted a fragmentation protocol of Picelli *et al.* (21) for the Ion Proton. Assembly of transposomes was performed following Wang *et al.* (22). Two reverse complementary oligonucleotides containing the Tn5 mosaic end sequence (upper: 5'-AGA TGT GTA TAA GAG ACA-G 3', lower: 5'-PhosCTG TCT CTT ATA CAC ATC T-3') were annealed at a concentration of 50  $\mu$ M each in TE (95°C 3 min, 70°C 3 min, cooling at 2°C min<sup>-1</sup> to 26°C) and subsequently diluted to 10  $\mu$ M in 50% glycerol. Annealed oligonucleotides and Ez-Tn5 transposase (1 U/ $\mu$ l, Epicentre) were mixed in a 4:1 ratio and incubated 30 min at room temperature for transposon assembly. The transposons were used immediately or stored at -20°C.

For the Fluidigm microfluidic chip, cDNA from the 96 output wells were pooled without quantification of the individual samples.

10 ng of cDNA in 9  $\mu$ l water were mixed with 4  $\mu$ l TAPS Buffer (50mM TAPS-NaOH, pH8.5 @ RT, 25 mM MgCl<sub>2</sub>), 2  $\mu$ l dimethylformamide and 5  $\mu$ l of transposon. After a 7 min incubation at 55°C, samples were cooled to 10°C and 4  $\mu$ l of 0.1% SDS was added, and the tubes were incubated for 10 min at 65°C to detach the transposase, and then cooled to 4°C. For Illumina sequencing, tagmentation was done for 5 min at 55°C. Alternatively, a commercial Nextera tagmentation kit (Illumina) can be used for Illumina library preparation.

Terminal fragments that were biotinylated during the PCR amplification were captured with 24  $\mu$ l Dynabeads<sup>®</sup> MyOne<sup>™</sup> Streptavidin C1 beads (Life Technologies) and washed following the manufacturer supplied protocol. Beads were suspended in 10.5  $\mu$ l water and 5' terminal fragments were amplified in 25  $\mu$ l with KAPA HiFi HotStart Ready mix and 0.5  $\mu$ M reverse library primer, 0.5  $\mu$ M forward library primer and 0.125  $\mu$ M extended forward library primer. For Illumina sequencing an extended reverse library primer was added (0.125  $\mu$ M). A gap filing step was done at 72°C during 3 min followed by 98°C for 30 sec, 15 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 30 s, and 72°C for 2 min. Ion Torrent libraries were size selected (200–350 pb) with SPRIselect<sup>®</sup> beads (Beckman Coulter) following the manufacturer supplied protocol. 100  $\mu$ l library in TE was incubated with 75  $\mu$ l of SPRIselect<sup>®</sup> beads to deplete fragments > 350 pb. Beads were discarded and an additional 15  $\mu$ l of SPRIselect<sup>®</sup> beads were added to capture fragments > 200 pb. Beads were washed with 85% EtOH and bound cDNA was eluted from the beads with 10  $\mu$ l water.

Illumina libraries (100  $\mu$ l) were just size selected for fragments > 200 bp with 90  $\mu$ l SPRIselect beads. Beads were recovered, washed and eluted as described above.

Quality and yield was determined with an Agilent Bionalyzer (Supplementary Figure S13).

## Sequencing

Libraries were sequenced either on a Proton Ion PI<sup>™</sup> Chip v3 (Thermo) or on a Nextseq 500 MID output flowcell (Illumina). For Illumina Nextseq sequencing, the custom sequencing primers listed in Supplementary Table S1 were added to the reagent cartridge following the 'NextSeq<sup>®</sup> System Custom Primers Guide' (Illumina Part # 15057456). Instructions for the use of custom sequencing primers with Illumina HiSeq sequencers are in Illumina document # 15061846. Our protocol for Illumina sequencers uses single indexing and either single or paired end sequencing. For Illumina sequencers, sequencing more libraries from more than 96 cells is possible. This can be done by adding a plate index during the final library preparation using an indexed extended reverse library primer (see Supplementary Table S1) and sequencing the plate index as 'index 2'.

## Read alignments and gene-expression analysis

Ion Torrent sequencers generate just one read that contains the barcode and the insert sequence. In consequence, the

read after barcode trimming starts with the TSO sequence (Supplementary Figure S3). We first examined whether the TSO sequence including the UMI [(ATCG)<sub>4</sub>(ATC)<sub>4</sub>] and the three Guanines following the UMI were correctly formatted and free of substitutions or indels (Supplementary Figure S3). Only reads with correctly formatted TSO sequences were processed further. UMI sequences were extracted and the TSO sequence was trimmed from the reads. Trimmed reads that were shorter than 26 nt were discarded. Typically 80 – 85% of the reads passed those filters.

In the case of Illumina sequencing, reads start with the UMI sequence and were just filtered for correct UMI formatting and the presence of three Guanines after the UMI.

Trimmed reads were mapped against the human genome (hg19) and ERCC sequences using STAR aligner (v2.4.0a), with default parameters. STAR indices were generated using Ensembl GTF file (release 75).

For molecule counting based on UMI counts, we used the Dropseq Core Computational Protocol version 1.0.1 (dropseq.jar) (16). Unless indicated otherwise, we used the `uniq` option (identical UMIs at two different transcript positions are only counted once) and `edit distance = 1` (UMIs for a transcript that are potentially a substitution mutant of another UMI with higher read coverage of the same transcript are discarded).

## Single cell quality filters

Capture sites were visually inspected for the presence and viability of cells after staining with the LIVE/DEAD<sup>®</sup> Viability/Cytotoxicity Kit for mammalian cells (Thermo) at 10 $\times$  magnification. Only capture sites with one live cell were retained for analysis. Additional quality control of the remaining libraries was performed using the R package 'Single-cell analysis toolkit for gene expression data in R' (scater version 1.0.4, <https://www.bioconductor.org>). Briefly, a set of cell quality indicators, such as the total number of UMIs for the cell, the total number of detected genes and the percentage of UMIs corresponding to mitochondrial genes was computed for each cell (23). Then, all cells flagged as outlier in a principal component analysis based on these quality measures were excluded. Furthermore, we excluded cells with a percentage of counts on ERCC spike-ins greater than the median + 4 times the median absolute deviation for the batch (`isOutlier` function of the scater R package). Special cases, such as identification of rare quiescent cells with low transcript numbers in a heterogeneous cell population might require fine-tuning of those filters.

## Statistical analysis

Statistical analysis was performed using the statistical package R version 3.3.1. ERCC capture efficiency was estimated as the intercept of a regression line with a constrained slope of 1 fitted between the expected number of ERCC molecules and the number of ERCC molecules counted. Only ERCC for which at least 10 molecules were spiked in were used for analysis. Correlation coefficients are calculated using Pearson's method. Single cell UMI count data were normalized for sequencing depth differences using the `normalize` function from the scran package version 1.0.4.

### Down-sampling of reads

After mapping, cell index and UMI extraction, a BAM file with reads for 47 single cells (mean =  $1.43 \times 10^6$  reads/cell) was down-sampled to 0.05, 0.1, 0.2, 0.4, 0.5, 0.75, 1.0, 1.25 million mean reads per cell. mRNA molecule counting was done with the dropseq.jar java pipeline (16).

### Analysis of public single cell RNAseq data

Data were downloaded from Gene Expression Omnibus, reads were matched to the reference genomes and UMIs were counted as described above.

HEK293 Dropseq data are from GEO accession GSE63473 (16). 259 cells were selected for further analysis using the same quality filtering as described above.

CEL-seq2 data of 96 cells processed in the Fluidigm C1 are from GEO accession GSE 78779 (samples GSM 2076519–GSM2076614) (14). The UMIs in this dataset have just five nucleotides and not six nucleotides as stated in Hashimshony *et al.* (14).

### Accession codes

Data were deposited in Gene Expression Omnibus (GSE79136).

## RESULTS

### Library preparation strategy

We sought to design a 5' selective library preparation strategy that uses UMIs and fulfills the following criteria: (i) unbiased introduction of cell indices before the costly and labor intensive fragmentation step; (ii) compatibility with the Fluidigm C1 microfluidic device design; (iii) essentially sequencing platform independent; (iv) cost and labor effective.

A simple option for pre-fragmentation barcoding is to introduce cell indices during reverse transcription. All current 3' selective single cell profiling approaches use this strategy and barcode via indexed reverse transcription primers. A similar strategy was also used in an early version of a 5' selective protocol by Islam *et al.* (12) who used barcoded TSOs to introduce the cell index at the 5' end of the cDNA during reverse transcription. However, introducing the cell index via barcoded TSOs has essentially two major disadvantages. First, an increase in TSO length has a negative impact on capture efficiency (10,24), a critical parameter in single cell transcriptome profiling. Secondly, use of different TSOs during reverse transcription was shown to cause differential capture of transcripts (24). Thus, the use of barcoded TSOs during reverse transcription would likely introduce bias that cannot be corrected with UMIs since UMIs are introduced during this step.

To avoid any barcode induced bias we opted for barcoding during PCR amplification of the cDNA, as eventual barcode dependent amplification bias can easily be detected and corrected by counting Unique Molecule Identifiers (UMIs) introduced during reverse transcription.

Performing reverse transcription in tiny volumes in Fluidigm microfluidic devices was shown to yield superior

mRNA capture than carrying out the same protocol in tubes or microtiter plates (18). Although the cost of the disposable microfluidic device is substantial, it is compensated by the >100-fold lower amount of required reagents and enzymes.

We designed a 5' single cell transcriptome sequencing workflow that is compatible with the Fluidigm C1 microfluidic device (Figure 1). We initially performed pilot experiments in tubes that mimicked the reaction conditions in the microfluidic device to select the optimal UMI and TSO design. The proposed protocol is therefore highly versatile and can easily be adapted to other instruments.

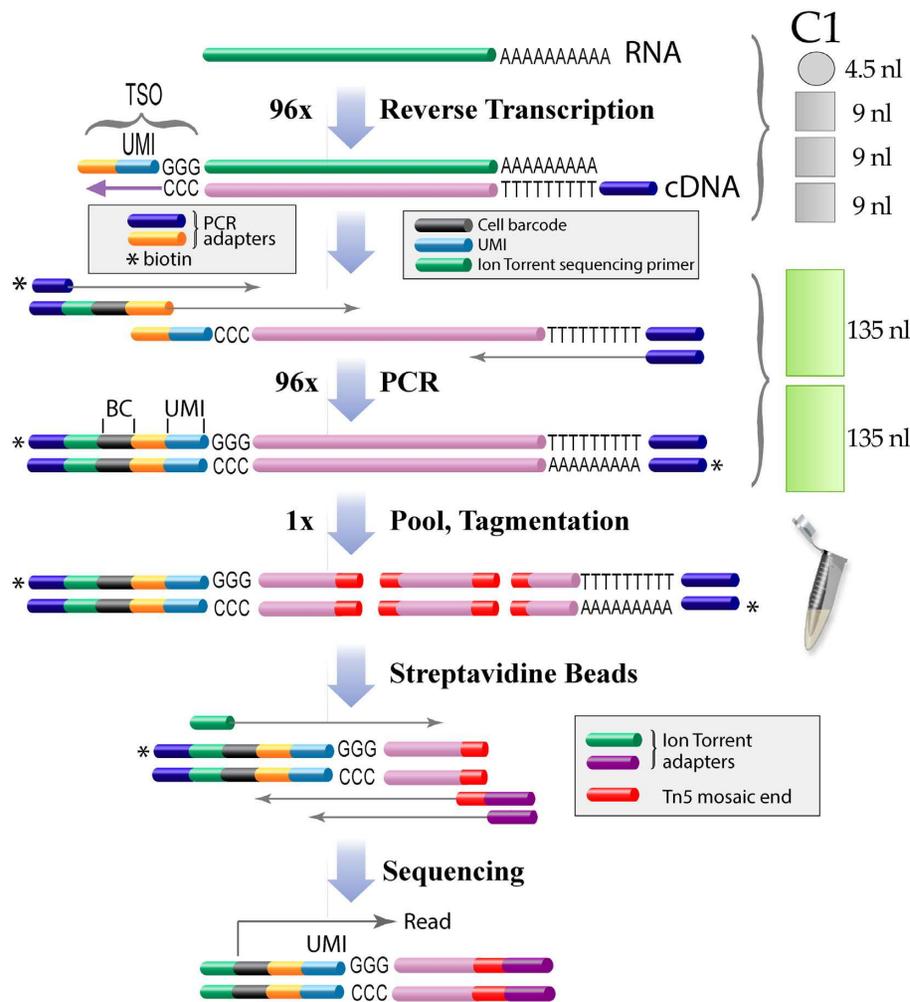
### TSO and UMI design

Rather short UMIs with five degenerate nucleotides were previously used by others for 5' selective single cell mRNA sequencing (7,10). However, the 1024 distinct sequences of a N<sub>5</sub> UMI are clearly insufficient to uniquely tag each copy of an abundant transcript with one and only one UMI. Several strategies were used by others to count abundant transcripts with short UMIs. One approach considers two reads with identical UMI of a given transcript as distinct molecules if both reads start at different positions on the transcript (10). Theoretical considerations of UMI usage saturation or UMI collision were also used to extrapolate the number of molecules for abundant transcripts (25). The used equations contain logarithms that tend to infinity and exaggeratedly overestimate the number of molecules when the number of detected UMIs approaches the maximal complexity of the UMI: they are thus not reliable for highly abundant transcripts.

To overcome the limitations of short UMIs, we rather increased the length of the UMI to 7 nucleotides. The complexity of a N<sub>7</sub> UMI ( $n = 16\,384$ ) should be sufficient to tag each copy of even abundant transcripts in a single cell with a unique UMI.

Ideally, UMIs should be introduced randomly, without bias for particular UMI sequences. However, we noticed that UMIs that are G-rich, particularly at the 3' end of the UMI, were highly enriched. Fifteen % of the UMI:transcript combinations were associated with a limited subset of just 100 G-rich UMIs (Figure 2a). This G-bias is likely due to the variable number of template independent nucleotides that are added by the reverse transcriptase. Previous studies showed that mainly three to four but sometimes up to six nucleotides, mainly cytosines, are added to the end of a cDNA (24) by the intrinsic terminal transferase activity of Superscript II. In the TSO, the seven Ns of the UMI are followed by three guanines to allow annealing to the 3' terminal cytosines of the cDNA. When more than three cytosines are added by the reverse transcriptase, a TSO with a longer stretch of 3' terminal Gs and thus UMIs that have Gs at their 3' terminus are likely selected, leading to a G bias at the 3' of the UMI.

To overcome such a G bias, we tested a N<sub>4</sub>H<sub>4</sub> UMI where the last four nucleotides are constrained to either A, T or C. The new UMI design resulted in a far better balanced usage of UMI sequences than the initial N<sub>7</sub> UMI (Figure 2B and C).



**Figure 1.** On chip barcoding workflow. After cell lysis in 4.5 nl poly-adenylated RNA is reverse-transcribed in 31.5 nl with an anchored oligodT primer. A PCR primer sequence and unique molecular identifiers (UMIs) are added to the 3' end of the cDNA via reverse transcriptase template switching. The cDNA is subsequently amplified and cell index sequences (barcode) as well as terminal biotins are introduced by PCR in the microfluidic device. The barcoded cDNAs are pooled, fragmented by tagmentation with Tn5 transposase and the biotinylated terminal fragments are isolated on streptavidin beads. 5' terminal fragments are selectively amplified and additional sequences required for Ion Torrent sequencers are introduced by PCR. For a detailed protocol see Supplementary Figure S11 and for Illumina sequencers see Supplementary Figure S8.

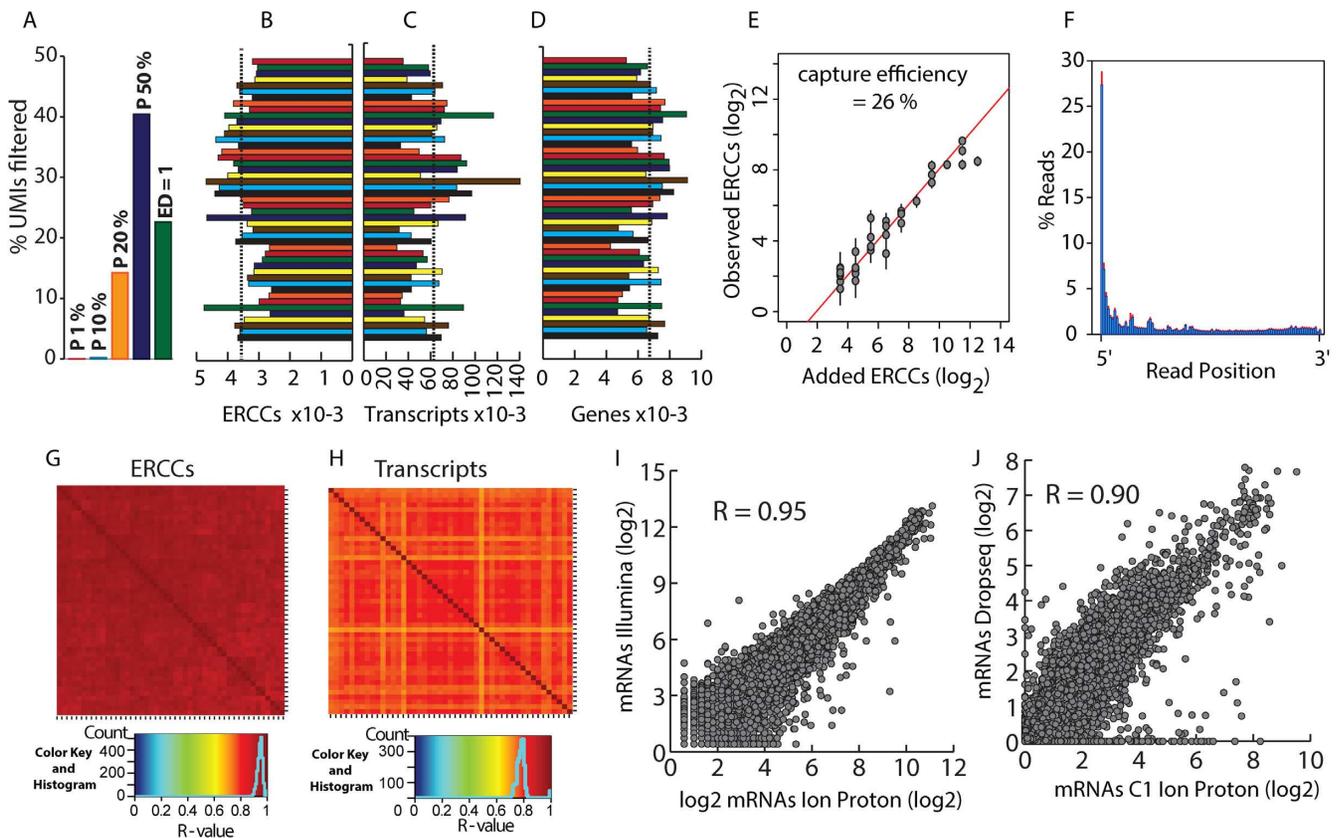
In single cell transcriptome profiling efficient transformation of a limited number of mRNA molecules into amplified cDNA is crucial. Most currently used highly efficient single cell transcriptome sequencing approaches exploit the template switching activity of reverse transcriptases to add a priming site required for subsequent PCR amplification to the 3' end of the cDNA (10,11,16). Different TSO designs were recently proposed for efficient template switching. Islam *et al.* (12) used a TSO with three 3' terminal riboguanosines (TSO\_rG3) while Picelli *et al.* (11) claimed superior template switching efficiency when the 3' terminal nucleotide of the TSO is a LNA base (TSO\_LNA). Conversely, another study reported superior efficiency of TSOs with three terminal riboguanosines over TSOs with LNA bases (26). In our experimental conditions both TSO designs performed rather similarly (cDNA yield TSO\_rG3/TSO\_LNA =  $1.03 \pm 0.26$  S.E.M.,  $n = 3$  means of triplicates). Since the UMI usage was slightly better balanced with the TSO\_rG3 (Figure 2B and C), we used this TSO

for all further experiments. The final protocol is highly reproducible with pools of HEK293 (correlation coefficients  $> 0.96$ , Figure 2D).

We next examined how our on chip barcoding protocol performs with single HEK293 cells in the Fluidigm C1 96 cell integrated fluidic circuit (IFC). The 96 amplified cDNAs were pooled without normalization, libraries were prepared and sequenced on an Ion Proton sequencer (Figure 1, Supplementary Figure S11, Materials and Methods section).

Introducing UMIs during cDNA synthesis theoretically allows correction of all bias induced by steps downstream of cDNA synthesis (e.g. PCR). However UMI counting and error correction strategies need to be critically considered to avoid bias.





**Figure 3.** Single cell sequencing. (A) Impact of UMI error filtering strategies. Percentage of filtered UMIs for different UMI error correction strategies. Filtering strategies were: Percentile, UMIs with a read coverage of less than the indicated fraction (P 1%, P 10%, P 20%, P 50%) of the average UMI read coverage of the corresponding gene were discarded; Edit distance (ED) = 1, UMIs that differ in just one nucleotide were merged into a single UMI. The percentages of eliminated UMIs were: P1%, 0.01%; P10%, 0.20%; P20%, 14.25%; P50%, 40.43%; ED = 1, 22.61%. Data are from one cell. (B, C) Number of ERCC (B) and transcript molecules (C) detected for each cell (means (dashed lines)/c.v.: ERCCs, 3558/15.7%; transcripts, 62 841/36.7%). (D) Number of genes detected for each cell (mean = 6679 (dashed line); c.v. = 16.9%). (E) Scatter plot showing the number of input ERCCs vs. the number of detected ERCCs (means  $\pm$  SD). The capture efficiency (26%) was calculated from the intercept of the regression line and the y-axis. (F) Distribution of read starts on annotated transcripts in one % bins between the 5' (0%) and the 3' end (100%). Data are means  $\pm$  SD (red bars) for 47 cells. (G) Heatmap of the pairwise correlation of ERCC molecules for 47 cells. (H) As (G) but for mRNAs. (I) Correlation between transcript (UMI) counts ( $\log_2(\text{counts} + 1)$ ) for pools of 100 HEK293 cells sequenced on an Ion Proton or Illumina Nextseq 500, respectively. Data are means from two pools of 100 HEK293 cells processed in tubes. (J) Correlation of HEK293 single cell transcript (UMI) counts ( $\log_2(\text{counts} + 1)$ ) between our Fluidigm C1 data and previously published Dropseq data (16). Transcript counts are means from 47 cells (Fluidigm) or 259 cells (Dropseq). Average numbers of transcript molecules detected per cell were: Fluidigm, 62,841; Dropseq, 36,746. R: Pearson correlation coefficient.

An alternative approach that does not rely on high sequencing depths was recently introduced (16). This approach, called 'edit distance filtering', merges UMIs of a given transcript when they differ by just one base and eliminates UMIs generated by substitution errors during PCR or sequencing. With our dataset, this filtering method eliminates 22.6% of the UMIs (Figure 3A) and is already effective at low sequencing depths. Edit distance filtering eliminated preferentially UMIs with low read coverage and increased the average UMI sequencing depth from 12.6 to 16.3 (Supplementary Figure S2b).

Despite an average UMI sequencing depth of 12.6 before filtering, 18% of the UMI:transcript combinations were read just once (Supplementary Figure S2b). A similar heterogeneity in UMI sequencing depth was reported by others (7). Yet, 66% of the UMIs covered by just one read are retained by the 'edit distance' filtering, an approach that eliminates UMIs with single substitution errors. In conse-

quence, the majority of those single read UMIs are likely not erroneous UMIs but rather correspond to real mRNA molecules. Conversely, those single read coverage UMIs are preferentially (10) or completely (7) discarded with UMI error filtering strategies that are simply based on UMI sequencing depth. Both the higher sensitivity at reasonable sequencing depths and the higher selectivity for erroneous UMIs led us to select the 'edit distance' filtering for UMI error correction. The Ion Proton sequencer adds a particular challenge since it generates a pretty high number of indel errors (up to 0.5%) which are hardly detected by the 'edit distance' UMI error filtering. To eliminate UMIs erroneously generated by indels, we took advantage of our UMI design where no guanosine is present in the last four bases of the UMI (Supplementary Figure S3). Any insertion or deletion in the UMI sequence thus results in a right or left shift of the first G following the UMI, respectively and can thus be filtered out.

## Performance and reproducibility

Analysis of the 47 cells that passed our quality filters (see methods section, Supplementary Table S2) showed that reads were preferentially located close to the 5' end of transcripts (27.3% of the reads started within the first percent of the mRNA, Figure 3F) what is consistent with our 5' selective library preparation strategy (Figure 1).

Individual ERCCs were highly correlated between cells (average  $R = 0.94$ , Figure 3G), gross ERCC molecule counts (Figure 3B) and UMI counts for individual ERCC spike in RNA (Figure 3E, Supplementary Figure S4) were similar for all cells indicating that mRNA was captured with comparable efficiency. The average ERCC cloning efficiency was 26% (Figure 3E), close to the efficiencies recently reported after stringent UMI error correction (7,16).

Conversely, gross counts for mRNA molecules (Figure 3C) and reads (Supplementary Figures S4b and S5) were more heterogeneous. The cell-to-cell differences in UMI and read counts likely represent real differences in the number of mRNA molecules rather than experimental variability: (i) ERCC cloning efficiencies were similar among cells (Figure 3B); (ii) while the number of transcript reads varied by as much as a factor of 6.8 between cells, the average transcript sequencing depth varied <2-fold (Supplementary Figure S5). In consequence, library normalization and generation of the same number of reads for each cell would likely lead to a higher heterogeneity in sequencing depth for the individual cells than simple library pooling without normalization.

Despite the rather high cell-to-cell heterogeneity of mRNA molecule (UMI) counts we observed a good correlation of mRNA expression between all cells (average  $R = 0.77$ , Figure 3H). The correlation between single cells was, as expected, somewhat lower than between pools of 100 HEK293 cells (Figure 2D) where cell-to-cell variations (e.g. cell cycle, etc.) are averaged out.

We performed two additional HEK293 cell single cell sequencing experiments to test the reproducibility of our protocol. Mean mRNA expressions from the three experiments correlated well (Supplementary Figure S6), despite the fact that the experiments were performed over a period of six months with HEK293 cells at different passage numbers.

To test our protocol in a biologically relevant context, we profiled human airway epithelial cells cultured at an air-liquid interface, a model that contains several distinct cell populations. The data from two independent primary cultures, which slightly differed in their cell culture conditions correlated well ( $R = 0.95$ , Supplementary Figure S7b). We anticipated the detection of at least two cell populations, namely multiciliated and basal cells. Hierarchical clustering of RNA sequencing data from two independent IFC runs (one for each cell culture) identified three main clusters that were further characterized based on the expression of specific markers (Supplementary Figure S7a). One cluster clearly corresponded to multiciliated cells, as evidenced by the expression of ciliated cells markers such as *TPP3*, *FOXJ1* and *ROPN1*. A second cluster was reminiscent of basal cells, as evidenced by a robust expression of basal cell markers such as *KRT5*, *KRT6A*, *KRT17* or *S100A2*. A third cluster is characterized by high levels of

*BPIFA1* and *BPIFB1*, which are associated to the innate immune response. Further experiments will be necessary to understand the cell types in this cluster. A secondary sub-clustering by cell culture / donor within those clusters is likely due to the different differentiation state of both cultures (culture 1, 52 days; culture 2, 33 days).

Taken together, the high reproducibility shown for pools of 100 HEK293 cells, for single HEK 293 cells and for primary epithelial cultures illustrates well the robustness of our SmartSeq based single cell library preparation technique.

While all other current single cell transcriptome profiling approaches were specifically designed for Illumina sequencers, our approach is essentially platform independent. After replacement of some oligonucleotides, the protocol designed for Ion Torrent sequencers was adapted for sequencing on Illumina sequencers (Supplementary Table S1, Supplementary Figure S8). Interestingly, the correlation between two distinct biological replicates sequenced on two different sequencers ( $R = 0.95$ ; Figure 3i) is close to what we obtained when replicates were sequenced on the same platform ( $R = 0.96-0.97$ ; Figure 2D). The precision of UMI based molecule counting is further illustrated by the high correlation ( $R = 0.90$ ) between our data and Dropseq single cell transcriptome data previously published for HEK293 cells (Figure 3J). This is particularly noteworthy, considering the use of two distinct single cell isolation approaches (Dropseq vs. microfluidic device), two different sequencing strategies (3' versus 5' end sequencing) and two different sequencer specific library preparations (Illumina versus Ion torrent).

Our on chip barcoding strategy reduces library preparation cost for the Fluidigm 96 cell IFC to essentially the cost of the microfluidic chip. With decreasing library preparation cost, sequencing of the libraries becomes the major cost factor in single cell transcriptome profiling. To estimate how many sequencing reads are required for profiling, we examined the impact of the number of sequencing reads on the number of detected transcript molecules and genes (Supplementary Figure S9). With an average of 1.43 million reads per cell, transcript and gene discovery rates approached a maximum with 62841 transcript molecules and 6679 expressed genes per HEK293 cell. The transcript detection rate is principally capped by the mRNA cloning efficiency, which is slightly above 26% with our protocol. (Figure 3e). Increasing further the sequencing depth would bring the transcript discovery rate somewhat closer to this limit but would also increase sequencing cost drastically. The number of required sequencing reads depends on the question to be addressed. Shallow sequencing with just 50 000 reads per cell was shown to be sufficient for cell type classification and biomarker identification (5). With 50 000 reads per cell we detect 54% and 28% of the maximally detected genes and transcript molecules, respectively. Reliable identification of expression changes for weakly expressed transcripts will require more reads. With 0.5–1 million reads we detect 80–93% of the transcript molecules and 90–97% of the expressed genes that we find at our maximal, almost saturating sequencing depth (Supplementary Figure S9). This should be sufficient for most routine single cell transcriptome profiling studies.

## DISCUSSION

We present a robust cost and labor effective 5' selective single cell transcriptome profiling approach where all the barcoding is done prior to fragmentation. This is to our knowledge the first 5' selective single cell mRNA sequencing protocol that allows pooling of the amplified, barcoded cDNA before fragmentation and does not require labor intensive and costly fragmentation of the individual libraries. We adapted the workflow for the Fluidigm integrated fluidic circuit (C1) and we detail workflows for Ion Torrent Proton and for Illumina sequencers. Yet, the method could likely be adapted for any other sequencing platform including long read sequencers. Adapters and barcodes that are used during cDNA amplification have just to be replaced by *ad hoc* sequences required for respective specific sequencer.

Most current single cell transcriptome profiling approaches use UMIs, which are introduced during reverse transcription. In consequence, any cell barcode bias introduced during reverse transcription remains uncorrected and should be avoided. Unlike all other currently popular single cell transcriptome profiling approaches that use UMIs (10,16,17), our method does not use barcoded primers during reverse transcription to exclude any barcode induced bias during this first highly critical step of mRNA capture and UMI tagging.

Another source of bias comes from false UMIs generated by sequencing errors. Such false UMIs can even outnumber real UMIs in heavily over-sequenced samples, for instance in experiments with low capture efficiency or for leaky cells with few mRNA molecules when the read number is boosted by library normalization. Since current UMI error correction strategies (16) hardly eliminate indel errors, we developed a novel UMI design that allows reliable identification and elimination of erroneous UMIs with indels (Supplementary Figure S3). This improved UMI design will be of interest not just for transcriptome profiling but also for the increasing number of NGS applications that rely on UMI based molecule counting or identification (27,28).

ERCC spike-in RNAs combined with UMIs were used to probe mRNA capture efficiencies in several studies. However, UMI lengths and UMI error correction strategies differ widely. We used an error correction strategy for our N<sub>4</sub>H<sub>4</sub> UMI that merges UMIs that differ in just one nucleotide (edit distance 1) and obtain 26% ERCC capture efficiency. Jaitin *et al.* (13) used the same UMI error filtering for a N<sub>4</sub> UMI in a 3' selective isothermal amplification based approach in microplates (Mars-seq) and claimed just 1–2% capture efficiency. This is likely highly underestimated since a N<sub>4</sub> UMI has a complexity of just 256 and edit distance filtering reduces the effective complexity even further, resulting in elimination of UMIs that correspond to real RNA molecules. Conversely, 22% capture efficiency was recently reported with a isothermal amplification approach (CEL-seq2) which is similar to the Mars-seq approach in a microfluidic device (14). However the authors did not correct for UMI errors and used UMI collision extrapolations to correct the UMI counts for abundant transcripts upwards. We reanalyzed the CEL-seq2 data (see methods section) with 'edit distance 1' UMI error filtering, the error correction used in our and previous studies (16) and obtained

13.8% capture efficiency. The highest capture efficiency reported for single cell transcriptome sequencing was 48% (10). However this value was obtained with the low stringency percentile filtering that barely filters any UMIs in our data. A more recent study by the same group used the same approach with more stringent UMI error filtering and reported 22% capture efficiency (7). Those differences in UMI design and error correction make any direct comparison of capture efficiencies reported in different studies difficult.

Our protocol and other published high efficiency single cell transcriptome profiling techniques (10,11,16) use a Smartseq based mRNA cloning strategy that relies on the template switching activity of reverse transcriptase, a process which is thought to favor capped RNAs, since reverse transcriptase preferentially adds non template dependent nucleotides to the cDNA when the RNA is capped (29). Thus, capture of mRNAs is likely cap selective but definitively not cap specific since the uncapped ERCC spike-in RNAs are also cloned with high efficiency by us (Figure 3) and others (10). ERCC RNAs (NIST #2374, [https://www-s.nist.gov/srmors/view\\_cert.cfm?srm=2374](https://www-s.nist.gov/srmors/view_cert.cfm?srm=2374)) are *in vitro* transcribed RNAs that all start with the same pT7T318 plasmid sequence including three 5' terminal Gs. The resulting cDNAs have 3' terminal Cs which are complementary to the 3' terminal Gs of the TSO. This might explain why template switching on uncapped ERCCs is efficient.

In our opinion, the ERCC capture efficiency should not be used to extrapolate absolute mRNA molecule counts from cDNA (UMI) counts since: (i) it is currently unknown whether those uncapped ERCCs are captured with the same efficiency as capped cellular mRNAs. (ii) Although capture of individual ERCCs (Figure 3E) and mRNAs (Figures 2D, and 3H–J) was highly reproducible in replicated experiments, the capture efficiency of two distinct but equally abundant ERCC molecules can vary almost by a factor of four in one sample (Figure 3E). Yet, the use of spike-in RNAs is crucial for the comparison of different protocols and the identification of badly performing samples or channels in a microfluidic device.

The transcript discovery rate did not completely saturate with an average of 1.43 million reads per cell (Supplementary Figure S9). This is essentially due to the high UMI sequencing depth heterogeneity. After UMI error filtering, 16% of the UMI::transcript combinations were sequenced just once despite a mean UMI sequencing depth of 16.3 (Supplementary Figure S2b). One likely reason for this broad heterogeneity is PCR amplification bias (30). Isothermal cRNA amplification, which is typically less biased than PCR, was recently proposed as an alternative to PCR in a 3' selective single cell sequencing approach (CEL-Seq, (15); Mars-Seq, (13), CEL-seq2 (14)). However, comparison of our data (PCR based approach) with recently published CEL-seq2 data (Isothermal amplification) reveals a similar UMI sequencing depth heterogeneity and thus amplification bias for both approaches (Supplementary Figure S10). Although bias downstream of reverse transcription is efficiently corrected by UMIs, reducing amplification and library preparation bias remains an important future challenge since this would profoundly reduce the required sequencing depth and sequencing cost. This will be of par-

ticular importance for high throughput droplet based approaches where thousands of cells are analyzed.

Recent developments in single cell transcriptome profiling focused on an increased throughput mainly with droplet based techniques. However, ERCC capture efficiencies are apparently lower for droplet based approaches (12.8% (16), 7.1% (17)) than for the Fluidigm microfluidic device (26% (this study), 22% (7)). The lower capture efficiency is not restricted to ERCC spike-in RNAs and was also observed for mRNAs. Macosko *et al.* (16) reported a mean Dropseq mRNA capture efficiency of 10.7%. Our observations are consistent with this. With our microfluidic approach we detected an average of 62 841 transcript molecules in a single HEK293 cell (Figure 3). In a published HEK293 Dropseq dataset (16) we identified 36 746 mRNA molecules per cell (Figure 3).

While droplet based techniques are currently clearly the method of choice when thousand of cells are analyzed, microfluidic devices are in our opinion better suited for small to medium size projects. With the on chip barcoding strategy we present, the Fluidigm C1 96 cell IFC allows a robust and highly efficient capture of the single cell transcriptome with little hands on time and negligible reagent cost. For routine single cell transcriptome profiling one Proton P1 chip ( $10^8$  reads) should be sufficient for a 96 cell microfluidic device. This reduces the overall cost of a 96 single cell transcriptome profiling study to about 1400 USD (Supplementary Table S3) and thus into a cost range where single cell transcriptome profiling becomes highly accessible and competitive with qPCR assays.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank all members from PB's group for discussions and Virginie Magnone for valuable advice on high throughput sequencing. This work was developed together with the functional genomics platform of Nice Sophia Antipolis, a partner of the National Infrastructure France Génomique.

## FUNDING

Cancéropôle PACA and the Commissariat aux Grands Investissements [ANR-10-INBS-09-03 and ANR-10-INBS-09-02]; Fondation pour la Recherche Médicale [DEQ20130326464 to P.B.]; Vaincre la Mucoviscidose [RF20140501158/1/1/70]; Agence Nationale pour la Recherche [ANR-12-BSVI-0023-02]; labex Signallife [ANR-11-LABX-0028-01]; Conseil Départemental 06. Funding for open access charge: ANR.

*Conflict of interest statement.* None declared.

## REFERENCES

- Ramskold,D., Luo,S., Wang,Y.-C., Li,R., Deng,Q., Faridani,O.R., Daniels,G.A., Khrebtkova,I., Loring,J.F., Laurent,L.C. *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.*, **30**, 777–782.
- Trapnell,C., Cacchiarelli,D., Grimsby,J., Pokharel,P., Li,S., Morse,M., Lennon,N.J., Livak,K.J., Mikkelsen,T.S. and Rinn,J.L. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, **32**, 381–386.
- Treutlein,B., Brownfield,D.G., Wu,A.R., Neff,N.F., Mantalas,G.L., Espinoza,F.H., Desai,T.J., Krasnow,M.A. and Quake,S.R. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, **509**, 371–375.
- Shalek,A.K., Satija,R., Adiconis,X., Gertner,R.S., Gaublot,J.T., Raychowdhury,R., Schwartz,S., Yosef,N., Malboeuf,C., Lu,D. *et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, **498**, 236–240.
- Pollen,A.A., Nowakowski,T.J., Shuga,J., Wang,X., Leyrat,A.A., Lui,J.H., Li,N., Szpankowski,L., Fowler,B., Chen,P. *et al.* (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.*, **32**, 1053–1058.
- Usoskin,D., Furlan,A., Islam,S., Abdo,H., Lonnerberg,P., Lou,D., Hjerling-Leffler,J., Haeggstrom,J., Kharchenko,O., Kharchenko,P.V. *et al.* (2015) Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.*, **18**, 145–153.
- Zeisel,A., Munoz-Manchado,A.B., Codeluppi,S., Lonnerberg,P., La Manno,G., Jureus,A., Marques,S., Munguba,H., He,L., Betsholtz,C. *et al.* (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, **347**, 1138–1142.
- Tang,F., Barbacioru,C., Wang,Y., Nordman,E., Lee,C., Xu,N., Wang,X., Bodeau,J., Tuch,B.B., Siddiqui,A. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, **6**, 377–382.
- Kolodziejczyk,A.A., Kim,J.K., Svensson,V., Marioni,J.C. and Teichmann,S.A. (2015) The technology and biology of single-cell RNA sequencing. *Mol. Cell*, **58**, 610–620.
- Islam,S., Zeisel,A., Joost,S., La Manno,G., Zajac,P., Kasper,M., Lonnerberg,P. and Linnarsson,S. (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods*, **11**, 163–166.
- Picelli,S., Björklund,Å.K., Faridani,O.R., Sagasser,S., Winberg,G. and Sandberg,R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
- Islam,S., Kjallquist,U., Moliner,A., Zajac,P., Fan,J.B., Lonnerberg,P. and Linnarsson,S. (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.*, **21**, 1160–1167.
- Jaitin,D.A., Kenigsberg,E., Keren-Shaul,H., Elefant,N., Paul,F., Zaretzky,I., Mildner,A., Cohen,N., Jung,S., Tanay,A. *et al.* (2014) Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science*, **343**, 776–779.
- Hashimshony,T., Senderovich,N., Avital,G., Klochender,A., de Leeuw,Y., Anavy,L., Gennert,D., Li,S., Livak,K.J., Rozenblatt-Rosen,O. *et al.* (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.*, **17**, 77.
- Hashimshony,T., Wagner,F., Sher,N. and Yanai,I. (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.*, **2**, 666–673.
- Macosko,E.Z., Basu,A., Satija,R., Nemes,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
- Klein,A.M., Mazutis,L., Akartuna,I., Tallapragada,N., Veres,A., Li,V., Peshkin,L., Weitz,D.A. and Kirschner,M.W. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
- Wu,A.R., Neff,N.F., Kalisky,T., Dalerba,P., Treutlein,B., Rothenberg,M.E., Mburu,F.M., Mantalas,G.L., Sim,S., Clarke,M.F. *et al.* (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods*, **11**, 41–46.
- Marcet,B., Chevalier,B., Luxardi,G., Coraux,C., Zaragosi,L.-E., Cibois,M., Robbe-Sermesant,K., Jolly,T., Cardinaud,B., Moreilhon,C. *et al.* (2011) Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nat. Cell Biol.*, **13**, 693–699.

20. Picelli,S., Faridani,O.R., Björklund,Å.K., Winberg,G., Sagasser,S. and Sandberg,R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.*, **9**, 171–181.
21. Picelli,S., Björklund,A.K., Reinius,B., Sagasser,S., Winberg,G. and Sandberg,R. (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.*
22. Wang,Q., Gu,L., Adey,A., Radlwimmer,B., Wang,W., Hovestadt,V., Bähr,M., Wolf,S., Shendure,J., Eils,R. *et al.* (2013) Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.*, **8**, 2022–2032.
23. Ilicic,T., Kim,J.K., Kolodziejczyk,A.A., Bagger,F.O., McCarthy,D.J., Marioni,J.C. and Teichmann,S.A. (2016) Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.*, **17**, 29.
24. Zajac,P., Islam,S., Hochgerner,H., Lönnerberg,P. and Linnarsson,S. (2013) Base Preferences in Non-Templated Nucleotide Incorporation by MMLV-Derived Reverse Transcriptases. *PLoS One*, **8**, e85270.
25. Fu,G.K., Hu,J., Wang,P.-H. and Fodor,S.P.A. (2011) Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9026–9031.
26. Harbers,M., Kato,S., de Hoon,M., Hayashizaki,Y., Carninci,P. and Plessy,C. (2013) Comparison of RNA- or LNA-hybrid oligonucleotides in template-switching reactions for high-speed sequencing library preparation. *BMC Genomics*, **14**, 1–6.
27. Borgstrom,E., Redin,D., Lundin,S., Berglund,E., Andersson,A.F. and Ahmadian,A. (2015) Phasing of single DNA molecules by massively parallel barcoding. *Nat. Commun.*, **6**, 7173.
28. Kinde,I., Wu,J., Papadopoulos,N., Kinzler,K.W. and Vogelstein,B. (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9530–9535.
29. Schmidt,W.M. and Mueller,M.W. (1999) CapSelect: A highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res.*, **27**, e31.
30. Kechschull,J.M. and Zador,A.M. (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.*, **43**, e143.

*XIII. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells*

---

---

ARTICLE

DOI: 10.1038/s41467-018-06768-z

OPEN

# CDC20B is required for deuterosome-mediated centriole production in multiciliated cells

Diego R. Revinski <sup>1</sup>, Laure-Emmanuelle Zaragosi <sup>2</sup>, Camille Boutin<sup>1</sup>, Sandra Ruiz-Garcia<sup>2</sup>, Marie Deprez<sup>2</sup>, Virginie Thomé<sup>1</sup>, Olivier Rosnet<sup>1</sup>, Anne-Sophie Gay<sup>2</sup>, Olivier Mercey<sup>2</sup>, Agnès Paquet<sup>2</sup>, Nicolas Pons<sup>2</sup>, Gilles Ponzio<sup>2</sup>, Brice Marcet<sup>2</sup>, Laurent Kodjabachian <sup>1</sup> & Pascal Barbry <sup>2</sup>

Multiciliated cells (MCCs) harbor dozens to hundreds of motile cilia, which generate hydrodynamic forces important in animal physiology. In vertebrates, MCC differentiation involves massive centriole production by poorly characterized structures called deuterosomes. Here, single-cell RNA sequencing reveals that human deuterosome stage MCCs are characterized by the expression of many cell cycle-related genes. We further investigated the uncharacterized vertebrate-specific *cell division cycle 20B* (*CDC20B*) gene, which hosts microRNA-449abc. We show that *CDC20B* protein associates to deuterosomes and is required for centriole release and subsequent cilia production in mouse and *Xenopus* MCCs. *CDC20B* interacts with PLK1, a kinase known to coordinate centriole disengagement with the protease Separase in mitotic cells. Strikingly, over-expression of Separase rescues centriole disengagement and cilia production in *CDC20B*-deficient MCCs. This work reveals the shaping of deuterosome-mediated centriole production in vertebrate MCCs, by adaptation of canonical and recently evolved cell cycle-related molecules.

<sup>1</sup>Aix Marseille Univ, CNRS, IBDM, Marseille, France. <sup>2</sup>Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France. These authors contributed equally: Diego R. Revinski, Laure-Emmanuelle Zaragosi, Camille Boutin. Correspondence and requests for materials should be addressed to B.M. (email: [marcet@ipmc.cnrs.fr](mailto:marcet@ipmc.cnrs.fr)) or to L.K. (email: [laurent.kodjabachian@univ-amu.fr](mailto:laurent.kodjabachian@univ-amu.fr)) or to P.B. (email: [barbry@ipmc.cnrs.fr](mailto:barbry@ipmc.cnrs.fr))

**M**ulticiliated cells (MCCs) are present throughout metazoan evolution and serve functions ranging from locomotion of marine larvae and flatworms, to brain homeostasis, mucociliary clearance of pathogens and transportation of oocytes in vertebrates<sup>1–3</sup>. The formation of MCCs requires the production of numerous motile cilia through a complex process called multiciliogenesis<sup>2,3</sup>. The transcriptional control of multiciliogenesis has been decrypted to a large extent, through studies in *Xenopus* and mouse<sup>2</sup>. Seating at the top of the cascade, the Geminin-related factors GemC1<sup>4–7</sup> and Multicilin<sup>8,9</sup> (MCIDAS in mammals) are both necessary and sufficient to initiate MCC differentiation. GemC1 and Multicilin in complex with E2F transcription factors have been reported to activate the expression of Myb, FoxJ1, Rfx2, and Rfx3, which collectively regulate the expression of a large body of effectors required for the formation of multiple motile cilia<sup>4,5,8–11</sup>. Recently, defective multiciliogenesis caused by mutations in MCIDAS and Cyclin O (CCNO) has been associated with congenital respiratory and fertility syndromes in human<sup>12,13</sup>.

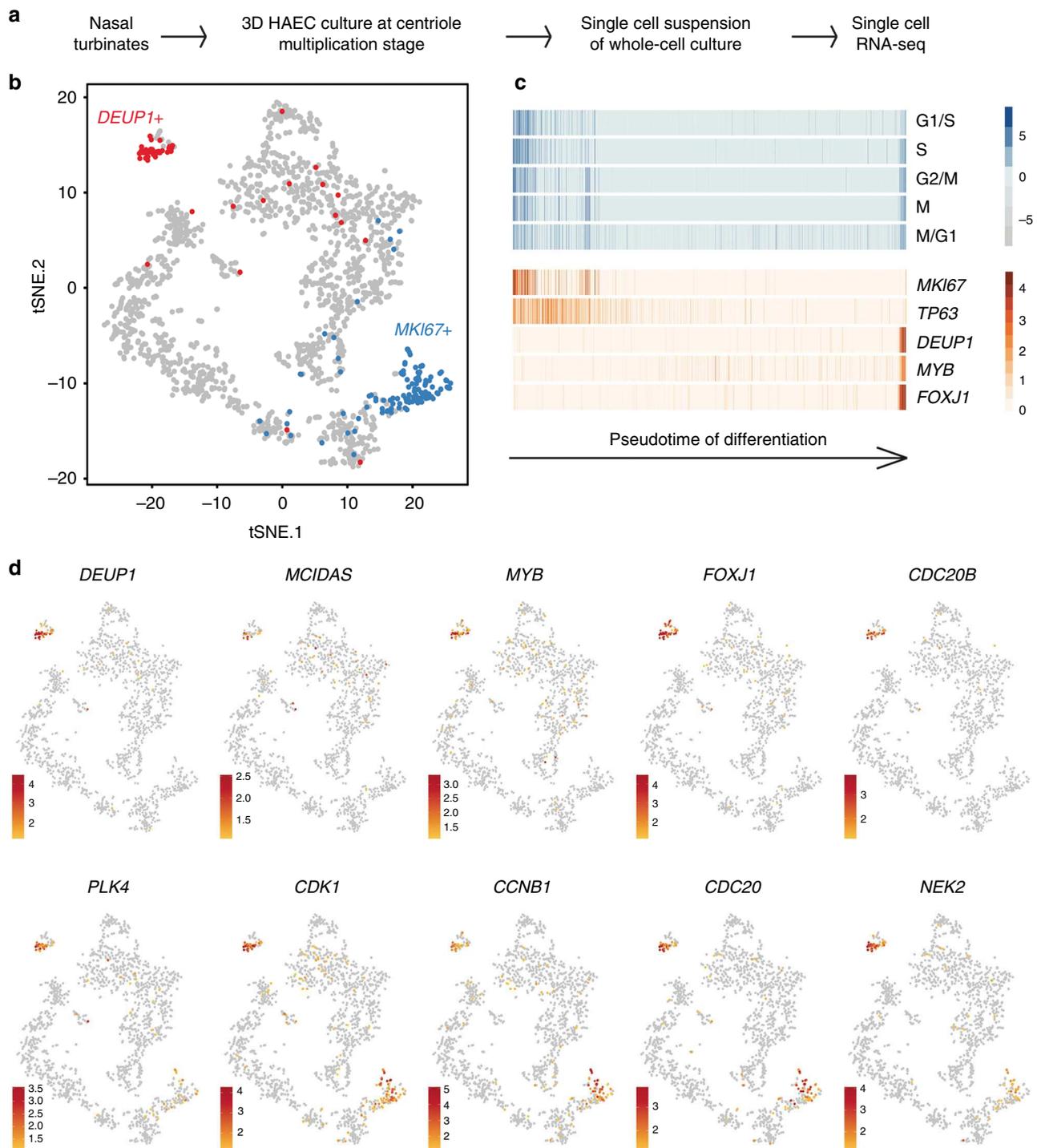
Each cilium sits atop a modified centriole, called a basal body (BB). After they exit from the cell cycle, maturing MCCs face the challenge of producing dozens to hundreds of centrioles in a limited time window. In vertebrate MCCs, bulk centriole biogenesis is mostly achieved through an acentriolar structure named the deuterosome, although canonical amplification from parental centrioles also occurs<sup>1–3</sup>. The deuterosome was first described in early electron microscopy studies of various multiciliated tissues including the mammalian lung<sup>14</sup> and oviduct<sup>15,16</sup>, the avian trachea<sup>17</sup>, and the *Xenopus* tadpole epidermis and trachea<sup>18</sup>. In mammalian MCCs, the deuterosome was described as a spherical mass of fibers organized into an inner dense region and an outer, more delicate, corona<sup>16</sup>. In *Xenopus*, deuterosomes were initially named pro-centriole organizers and were reported as dense amorphous masses<sup>18</sup>. Recent studies have revealed that deuterosome-mediated centriole synthesis mobilizes key components of the centriole-dependent duplication pathway of the cell cycle, including CEP152, PLK4, and SAS6<sup>19–21</sup>. However, the deuterosome itself differs from the centriole and may contain specific components. The identification of one such component, called DEUP1 for Deuterosome assembly protein 1, opened the possibility to investigate the deuterosome at the molecular level<sup>21</sup>. In mouse tracheal ependymal cells, DEUP1 was detected in the core of the deuterosome<sup>21</sup>. DEUP1, also known as CCDC67, is a conserved vertebrate paralogue of CEP63, itself known for its importance in initiation of centriole duplication during the cell cycle<sup>21,22</sup>. Consistently, DEUP1 was shown to be essential for centriole multiplication in mouse and *Xenopus* MCCs<sup>21</sup>. Both CEP63 and DEUP1 interact with CEP152, an essential event for centriole duplication and multiplication in cycling cells and MCCs, respectively<sup>21,22</sup>. Once centriole multiplication is over, neo-synthesized centrioles must disengage from deuterosomes and parental centrioles, convert into BBs and migrate apically to dock at the plasma membrane to initiate cilium elongation.

In this study, we aimed at better understanding deuterosome biology. We found that the gene *CDC20B* was specifically expressed in maturing MCCs during the phase of centriole multiplication. We established the corresponding CDC20B protein as an essential regulator of centriole-deuterosome disengagement. This work illustrates well the strong functional relationships that exist between centriole release from deuterosomes and centriole disengagement in mitotic cells. It also posits CDC20B as a component of a “multiciliary locus” that contains several gene products, either proteins, such as MCIDAS, CCNO or CDC20B itself, or microRNAs, such as miR-449abc, which are all actively involved into vertebrate multiciliogenesis.

## Results

**MCC single-cell transcriptome at deuterosome stage.** To identify regulators of centriole multiplication, we analyzed the transcriptome of human airway epithelial cells (HAECs) at the differentiation stage corresponding to active centriole multiplication<sup>23</sup> at the single-cell level (Fig. 1a). Gene expression data from 1663 cells were projected on a 2D space by *t*-distributed Stochastic Neighbor Embedding (tSNE) (Fig. 1b). We identified a small group of 37 cells corresponding to maturing MCCs engaged in deuterosome-mediated centriole amplification, as revealed by the specific expression of *MCIDAS*<sup>8</sup>, *MYB*<sup>24</sup>, and *DEUP1*<sup>21</sup> (Fig. 1c, d and Supplementary Figure 1). This subpopulation was characterized by the expression of known effectors of centriole synthesis, such as *PLK4*, *STIL*, *CEP152*, *SASS6*, but also of cell cycle regulators, such as *CDK1*, *CCNB1*, *CDC20*, *SGOL2*, and *NEK2* (Fig. 1d, Supplementary Figure 1 and Supplementary Table 1). We reasoned that uncharacterized cell cycle-related genes that are specific to this subpopulation could encode components of the deuterosome-dependent centriole amplification pathway. A particularly interesting candidate in this category was *CDC20B* (Fig. 1d), which is related to the cell cycle regulators *CDC20* and *FZR1*<sup>25</sup> (Supplementary Figure 2a). First, the *CDC20B* gene is present in the vertebrate genomic locus that also contains the key MCC regulators *MCIDAS*<sup>8</sup> and *CCNO*<sup>13</sup>. Co-expression of *CDC20B*, *MCIDAS*, and *CCNO* throughout HAEC differentiation was indeed observed in an independent RNA sequencing study, performed on a bulk population of HAECs (Supplementary Figure 2b). These results fit well with the observation that the promoter of human *CDC20B* was strongly activated by the MCIDAS partners E2F1 and E2F4 (Supplementary Figure 2c), as also shown in *Xenopus* by others<sup>9</sup> (Supplementary Figure 2d). Second, the *CDC20B* gene bears in its second intron the miR-449 microRNAs, which were shown to contribute to MCC differentiation<sup>23,26–30</sup>. Finally, in *Xenopus* epidermal MCCs, *cdc20b* transcripts were specifically detected during the phase of centriole amplification (Supplementary Figure 2e–m). This first set of data pointed out the specific and conserved expression pattern of *CDC20B* in immature MCCs. In the rest of this study, we analyzed the putative role of CDC20B in deuterosome-mediated centriole multiplication.

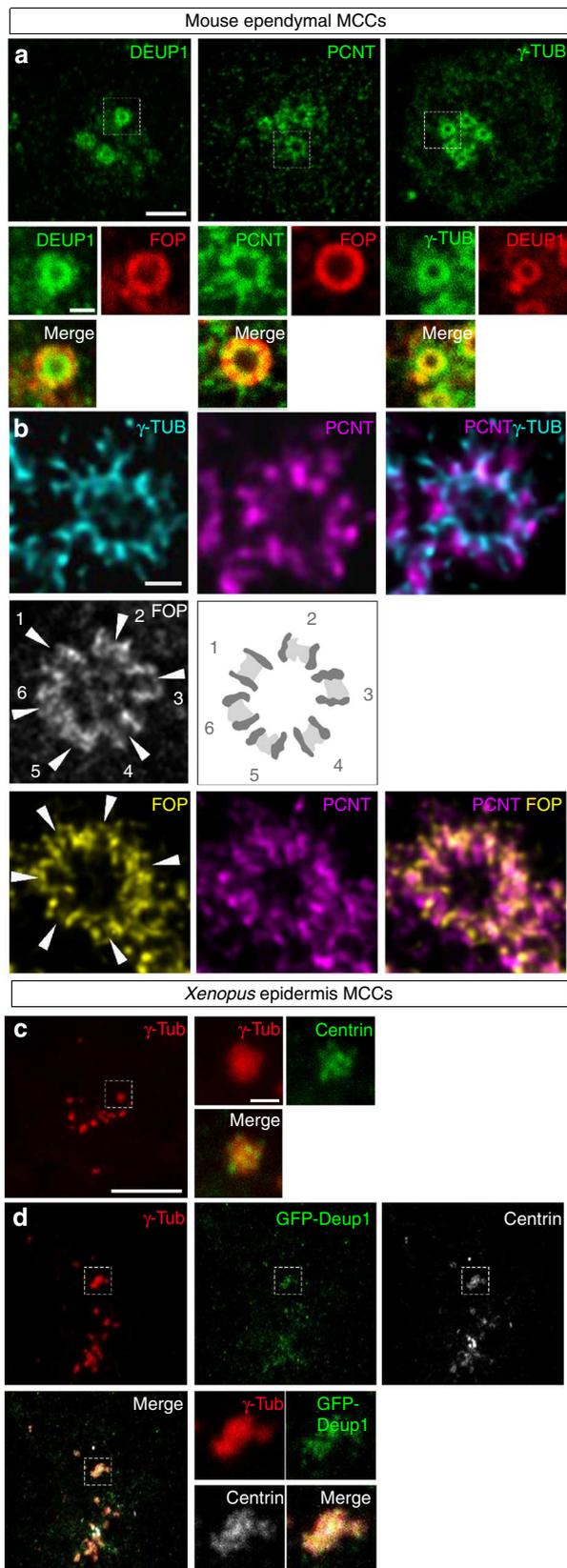
**Composition and organization of vertebrate deuterosomes.** We first conducted a series of immunofluorescence analyses to gain a better understanding of deuterosome organization in mouse ependymal and *Xenopus* epidermal MCCs as models. In whole-mounts of mouse ependymal walls, mature deuterosomes revealed by DEUP1 staining appeared as circular structures around a lumen (Fig. 2a). We noticed that DEUP1 also stained fibers emanating from the core into the corona. Nascent centrioles revealed by the marker FOP were organized around the DEUP1-positive core ring. STED super-resolution microscopy helped to better appreciate the regular organization of individual FOP-positive pro-centrioles (Fig. 2b). Proximity labeling assays have revealed that when ectopically expressed in centrosomes CCDC67/DEUP1 is found close to Pericentrin (PCNT) and  $\gamma$ -tubulin, two main components of the pericentriolar material (PCM)<sup>31</sup>. Interestingly, we found that PCNT was present in the deuterosome corona (Fig. 2a), and STED microscopy further revealed that PCNT formed fibers around growing pro-centrioles (Fig. 2b).  $\gamma$ -tubulin staining was detected in the DEUP1-positive deuterosome core, as well as in the corona (Fig. 2a). STED microscopy indicated that PCNT and  $\gamma$ -tubulin stained distinct interwoven fibers in the deuterosome corona. Next, we stained immature *Xenopus* epidermal MCCs with  $\gamma$ -Tubulin and Centrin to reveal centriole amplification platforms. These platforms



**Fig. 1** Single-cell RNA-seq analysis reveals MCC transcriptome at deuterosome stage. **a** Experimental design of the scRNA-seq experiment. **b** tSNE plot. Each point is a projection of a unique cell on a 2D space generated by the tSNE algorithm. Blue dots represent *MKI67*-positive proliferating cells, and red dots represent *DEUP1*-positive cells corresponding to maturing MCCs at deuterosome stage. **c** Cell cycle-related gene set expression in HAECs measured by scRNA-seq. Cells were ordered along a pseudotime axis, defined with the Monocle2 package. Phase-specific scores are displayed in the top heatmap. Expression of selected genes is displayed in the bottom heatmap. **d** tSNEs plots for a selection of genes specifically enriched in deuterosome stage cells. Note that *CDC20B* exhibits the most specific expression among deuterosome marker genes

displayed irregular shapes and sizes (Fig. 2c), in agreement with early electron microscopy studies<sup>18</sup>. Expression of low amounts of GFP-Deup1 in MCCs induced by Multicilin confirmed that active deuterosomes are embedded in  $\gamma$ -Tubulin-positive masses (Fig. 2d). Overall, this analysis is consistent with early ultra-structural studies, as the deuterosome core and corona can be

distinguished by the presence of *DEUP1* and *PCNT*, respectively. Moreover,  $\gamma$ -tubulin is a conserved marker of centriole amplification platforms in vertebrate MCCs. By analogy to the organization of the centrosome, we propose to coin the term perideuterosomal material (PDM) to describe the corona, as this region may prove important for deuterosome function.

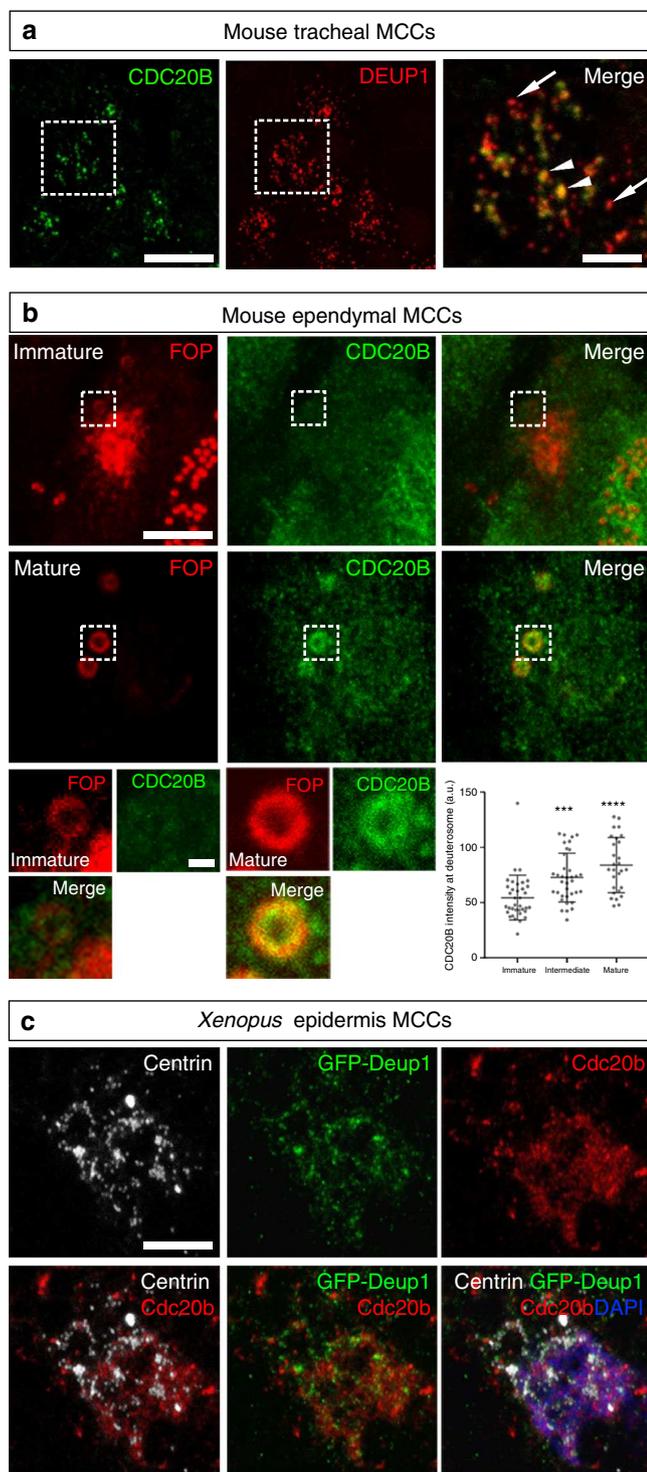


**Fig. 2** Composition and organization of vertebrate deuterosomes **a, b** Maturing mouse ependymal MCCs were immunostained as indicated, pictures were taken with confocal (**a**) or STED (**b**) microscope. **a** Individual deuterosomes (dashed boxes in top panels) are shown at higher magnification in bottom panels. DEUP1 stains the deuterosome core (ring) and a close fibrous area that defines the perideuterosomal region. The centriolar marker FOP reveals procentrioles arranged in a circle around the deuterosome. Pericentrin (PCNT) is enriched in the perideuterosomal region.  $\gamma$ -Tubulin ( $\gamma$ -TUB) stains the core as well as the periphery of the deuterosome. **b** STED pictures showing the organization of FOP, PCNT, and  $\gamma$ -TUB around deuterosomes. Individual centrioles identified by FOP staining are pointed out with arrowheads. The diagram was drawn from the adjacent FOP photograph to help reveal the regular concentric organization of nascent centrioles in a typical deuterosomal figure. **c** *Xenopus* embryos were immunostained for  $\gamma$ -Tubulin ( $\gamma$ -Tub) and Centrin and high magnification pictures of immature epidermal MCCs were taken. In these cells, Centrin-positive procentrioles grow around  $\gamma$ -Tubulin-positive structures. **d** *Xenopus* embryos were injected with *Multicilin-hGR* and *GFP-Deup1* mRNAs, treated with dexamethasone at gastrula st11 to induce *Multicilin* activity, and immunostained at neurula st18 for  $\gamma$ -Tubulin, GFP, and Centrin. In **c** and **d**, zooms (right panels) were made on regions identified by dashed boxes. Scale bars: 5  $\mu$ m (**a**, top), 500 nm (**a**, bottom), 500 nm (**b**), 10  $\mu$ m (**c**, **d**, large view), 1  $\mu$ m (**c**, **d**, high magnification)

We noticed that CDC20B tended to associate primarily to large DEUP1 foci. As deuterosomes grow as they mature<sup>21</sup>, this suggests that CDC20B may penetrate into the deuterosomal environment at a late stage of the centriole multiplication process. The same observation was made when comparing CDC20B staining in the region of immature and mature deuterosomes of mouse ependymal MCCs (Fig. 3b). As double DEUP1/CDC20B staining could not be performed on these cells, we analyzed CDC20B distribution relative to FOP-positive procentrioles. In early deuterosome stage MCCs, CDC20B was expressed at low levels and FOP staining was mostly concentrated in a large amorphous cloud (Fig. 3b). In such cells, no CDC20B staining was detected in association to FOP-positive procentrioles growing around deuterosomes. In contrast, in mature deuterosome stage MCCs, CDC20B was enriched in the innermost part of the PDM, probably very close to the deuterosome core (Fig. 3b). Further evidence was provided with a custom-made polyclonal antibody (Supplementary Figure 3b, c) used to analyze Cdc20b protein distribution in *Xenopus* epidermal MCCs. Here also, Cdc20b was found associated to Deup1-positive deuterosomes actively engaged in centriole synthesis (Fig. 3c). We finally analyzed the distribution of CDC20B in mature MCCs. As previously reported, the CDC20B protein was detected near BBs<sup>23</sup>, but also in cilia of fully differentiated human airway MCCs (Supplementary Figure 4a–c). This was confirmed by proximity ligation assays that revealed a tight association of CDC20B with Centrin2 and acetylated  $\alpha$ -Tubulin, in BBs and cilia, respectively (Supplementary Figure 4d–f). Fluorescent immunostaining also revealed the presence of Cdc20b in the vicinity of BBs in *Xenopus* epidermal MCCs (Supplementary Figure 4g–i). In contrast, no cilia staining was observed in these cells. Altogether, our analyses revealed that in three distinct types of MCCs in two distant vertebrate species, CDC20B is tightly associated to mature deuterosomes. We next investigated whether it may control their function.

**CDC20B associates to vertebrate deuterosomes.** We then analyzed the subcellular localization of CDC20B protein in deuterosome stage mouse and *Xenopus* MCCs. In immature mouse tracheal MCCs, double immunofluorescence revealed the association of CDC20B to DEUP1-positive deuterosomes (Fig. 3a).

**CDC20B is required for multiciliogenesis in vertebrates.** For that purpose, *Cdc20b* was knocked down in mouse ependymal MCCs, through post-natal brain electroporation of three distinct shRNAs. One of them, sh274, which targets the junction between exons 3 and 4, and can therefore only interact with mature



**Fig. 3** CDC20B associates to vertebrate deuterosomes. **a** Double immunofluorescence was performed on mouse tracheal MCCs after 3 days of culture in air-liquid interface. Low magnification confocal panels show coincident CDC20B and DEUP1 staining in several individual MCCs. High magnification on a single MCC reveals the prominent association of CDC20B to large deuterosomes marked by DEUP1 (arrowheads). Note that some smaller deuterosomes do not contain CDC20B (arrows). **b** Mouse ependymal MCCs were immunostained as indicated, and high magnification confocal pictures of cells with immature and mature deuterosomal figures were taken. In these cells, centrioles revealed by FOP form a ring around deuterosomes. CDC20B staining forms a ring inside the ring of FOP-positive procentrioles indicating that CDC20B is tightly associated to deuterosomes. Note that the CDC20B signal associated to deuterosome increased with their maturation (high magnification pictures of >25 cells per category from two different animals were quantified in the graph; mean values and standard deviations are shown). Unpaired *t* test vs immature:  $p = 0.0005$  (intermediate, \*\*\*);  $p < 0.0001$  (mature, \*\*\*\*). In **a** and **b**, zooms were made on regions identified by dashed boxes. **c** *Xenopus* embryos were injected with *GFP-Deup1* mRNA and immunostained at neurula st18 as indicated. Scale bars: 5 μm (**a**, **b**, large view), 1.5 μm (**a**, high magnification), 500 nm (**b**, high magnification), 10 μm (**c**)

that displayed centrioles still engaged on deuterosomes (Fig. 4f, g). Fifteen days after electroporation, a majority of CDC20B-deficient MCCs still showed a severely reduced number of released centrioles, and consequently lacked cilia (Fig. 4h–k).

*Cdc20b* was also knocked down in *Xenopus* epidermal MCCs, through injection of two independent morpholino antisense oligonucleotides targeting either the ATG (Mo ATG), or the exon 1/intron 1 junction (Mo Spl) (Supplementary Figure 5b). The efficiency of Mo ATG was verified through fluorescence extinction of co-injected *Cdc20b*-Venus (Supplementary Figure 5c). RT-PCR confirmed that Mo Spl caused intron 1 retention (Supplementary Figure 5d), which was expected to introduce a premature stop codon, and to produce a *Cdc20b* protein lacking 96% of its amino acids, likely to undergo unfolded protein response-mediated degradation. Thus, both morpholinos were expected to generate severe loss of *Cdc20b* function. Consistent with this interpretation, both morpholinos strongly reduced *Cdc20b* immunostaining in deuterosome stage MCCs (Supplementary Figure 5e). We verified that neither morpholinos caused *p53* transcript up-regulation (Supplementary Figure 5f), a non-specific response to morpholinos that is sometimes detected in zebrafish embryos<sup>32</sup>. Importantly, whole-mount in situ hybridization indicated that miR-449 expression was not perturbed in the presence of either morpholino (Supplementary Figure 5g). We found that *cdc20b* knockdown did not interfere with acquisition of the MCC fate (Supplementary Figure 6a–e), but severely impaired multiciliogenesis, as revealed by immunofluorescence and electron microscopy (Fig. 5a–i). This defect stemmed from a marked reduction in the number of centrioles, and poor docking at the plasma membrane (Fig. 5g–o and Supplementary Figure 6f–k). Importantly, centrioles and cilia were rescued in Mo Spl MCCs by co-injection of *cdc20b*, *venus-cdc20b* or *cdc20b-venus* mRNAs (Fig. 5j–o and Supplementary Figure 6f–k). In normal condition, *Xenopus* epidermal MCCs arise in the inner mesenchymal layer and intercalate into the outer epithelial layer, while the process of centriole amplification is underway<sup>33</sup>. To rule out secondary defects due to poor radial intercalation, we assessed the consequences of *cdc20b* knockdown in MCCs induced in the outer layer by Multicilin overexpression<sup>8</sup>. Like in natural MCCs, *Cdc20b* proved to be essential for the production of centrioles and cilia in response to Multicilin activity (Supplementary Figure 7a–g). We also noted that the apical actin network that

mRNA, was useful to rule out possible interference with the production of miR-449 molecules from the *Cdc20b* pre-mRNA (Supplementary Figure 5a). Five days after electroporation, all three shRNAs significantly reduced the expression of CDC20B in deuterosome stage MCCs (Fig. 4c), but did not alter MCC identity as revealed by FOXJ1 expression (Fig. 4a, b, d). Centriole production by deuterosomes was analyzed by FOP/DEUP1 double staining 9 days after electroporation. At this stage, control MCCs had nearly all released their centrioles and disassembled their deuterosomes (Fig. 4e, g). In sharp contrast, *Cdc20b* shRNAs caused a significant increase in the number of defective MCCs

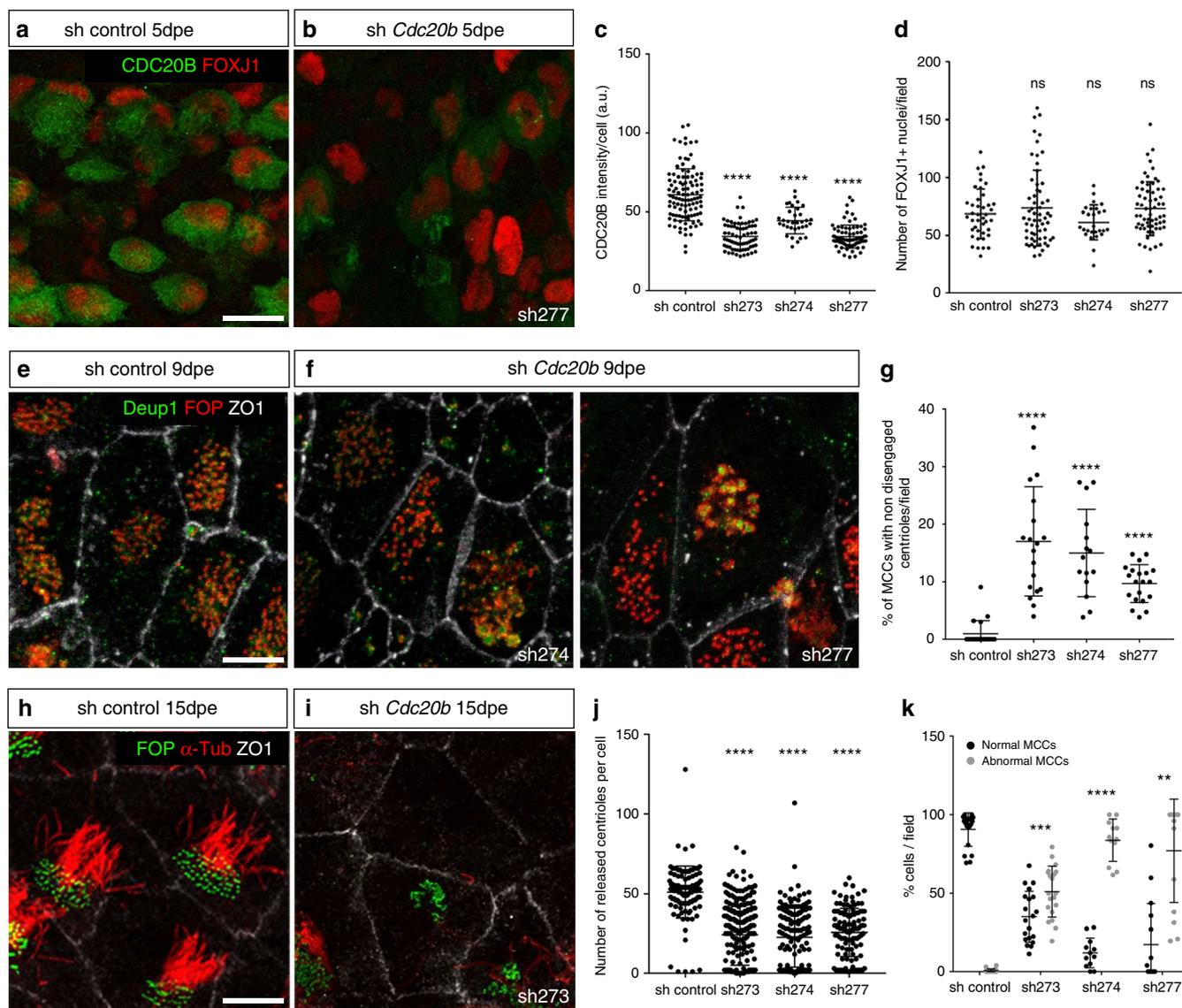
normally surrounds BBs was disrupted in absence of Cdc20b, although this defect could be secondary to the absence of centrioles (Supplementary Figure 7d–g). Centrioles in Cdc20b morphant cells often formed clusters, suggesting that disengagement from deuterosomes could have failed (Fig. 5l,m). To better assess this process, we injected GFP-Deup1 in Multicilin-induced MCCs and stained centrioles with Centrin. In mature control MCCs, deuterosomes were disassembled, centrioles were converted into BBs, had docked and initiated cilium growth (Fig. 5p, s). In contrast, both morpholinos caused a marked increase in the number of defective MCCs, which were devoid of cilia and displayed centrioles still engaged on deuterosomes (Fig. 5q–u). Altogether our functional assays in mouse and *Xenopus* indicate that CDC20B is required for centriole disengagement from deuterosomes and subsequent ciliogenesis in MCCs. We next investigated the molecular mechanism of action of CDC20B underlying its role in centriole release.

**Partners and effectors of CDC20B reveal its mechanism of action.** In mitotic cells, centriole disengagement is necessary to license centriole duplication in the following cell cycle<sup>34</sup>. This process is known to depend on the coordinated activities of the mitotic kinase PLK1 and the protease Separase<sup>35</sup>. One proposed mechanism involves the phosphorylation of PCNT by PLK1, which induces its cleavage by Separase, thereby allowing centriole disengagement through disassembly of the PCM<sup>36,37</sup>. Separase is known to be activated by the degradation of its inhibitor Securin, which is triggered by the Anaphase Promoting Complex (APC/C) upon binding to CDC20<sup>25</sup>. *PLK1*, *Separase (ESPL1)*, *Securin (PTTG1)*, *CDC20*, and *PCNT* were all found to be expressed in human deuterosome stage MCCs (Fig. 1d and Supplementary Figure 1). We have shown above that PCNT is present in the PDM and a recent study revealed the presence of CDC20 and the APC/C component APC3 in mouse ependymal MCCs at the stage of centriole disengagement<sup>38</sup>. Based on this large body of information, we hypothesized that centriole-deuterosome disengagement involves the coordinated activities of PLK1 and Separase, and that CDC20B would be involved in this scenario. *CDC20B* encodes a protein of about 519 amino acids largely distributed across the vertebrate phylum<sup>23</sup>. In its C-terminal half, CDC20B contains seven well conserved WD40 repeats, predicted to form a  $\beta$ -propeller, showing 49 and 37% identity to CDC20 and FZR1 repeats, respectively (Supplementary Figure 2a). However, CDC20B lacks canonical APC/C binding domains (Supplementary Figure 2a). Using mass spectrometry on immunoprecipitated protein complexes from transfected HEK cells, we could identify multiple APC/C components interacting with CDC20 but not with CDC20B (Supplementary Table 2). We conclude that CDC20B is probably incapable of activating APC/C. Interestingly, an unbiased interactome study reported association of CDC20B with PLK1<sup>39</sup>. Using reciprocal co-immunoprecipitation assays in HEK transfected cells, we confirmed that CDC20B and PLK1 could be found in the same complex (Fig. 6a and Supplementary Figure 8). This suggested that CDC20B could cooperate with PLK1 to trigger centriole disengagement. Consistent with this hypothesis, we found that PLK1 was enriched in the PDM of mature deuterosomes in mouse ependymal MCCs (Fig. 6b), in agreement with a recent report<sup>38</sup>. Another interesting partner of CDC20B identified in a second unbiased interactome study<sup>40</sup> was SPAG5 (Astrin), which was reported to control timely activation of Separase during the cell cycle<sup>41,42</sup>. Using the same strategy as above, we could detect CDC20B and SPAG5 in the same complex (Fig. 6c and Supplementary Figure 8). As SPAG5 was found associated to DEUP1 in a proximity labeling assay<sup>31</sup>, we assessed its localization in

deuterosomes. Strikingly, SPAG5 was detectable in mature deuterosomes of mouse ependymal MCCs, with a clear enrichment in the deuterosome core (Fig. 6d). Finally, reciprocal co-immunoprecipitations revealed that CDC20B and DEUP1 were detected in the same complex when co-expressed in HEK cells (Fig. 6e and Supplementary Figure 8). Consistent with this result, we observed that RFP-Cdc20b was recruited around spherical Deup1-GFP structures positive for  $\gamma$ -Tubulin and Centrin in *Xenopus* epidermal MCCs (Supplementary Figure 7h–m). This series of experiments suggested that CDC20B could participate in the assembly of a protein complex in mature deuterosomes, required to coordinate the activities of PLK1 and Separase for centriole disengagement. As Separase is the last effector in this scenario, we tested whether over-expressing human Separase in *Xenopus cdc20b* morphant MCCs could rescue centriole disengagement. In support to our hypothesis, over-expression of wild-type, but not protease-dead Separase, efficiently rescued centriole disengagement and cilia formation in *cdc20b* morphant MCCs (Fig. 7a–g and Supplementary Figure 7n–s). Separase could also rescue multiciliogenesis in Multicilin-induced MCCs injected with *cdc20b* Mos (Supplementary Figure 7t–z). We conclude that CDC20B is involved in Separase-mediated release of mature centrioles from deuterosomes in vertebrate MCCs (Fig. 7h).

## Discussion

In this study, we report the essential and conserved role of CDC20B in vertebrate multiciliogenesis. Our data suggest that the presence of CDC20B in the perideuterosomal region is necessary to allow centriole disengagement. We note, however, that our data, which are based on partial knockdowns, remain compatible with an earlier function of CDC20B in promoting deuterosome assembly and/or activity. A total genetic knockout of *Cdc20b* should help to assess this possibility in mouse tracheal and ependymal MCCs. By analogy to mitosis, we propose that CDC20B is involved in Separase-dependent proteolysis at deuterosomes, allowing the release of mature centrioles and subsequent ciliogenesis. This view is consistent with a recent report showing that centriole disengagement in murine ependymal MCCs involves the activities of PLK1, a partner of CDC20B, and APC/C, the activator of Separase<sup>38</sup>. The central question arising from our work then becomes: how are CDC20B and Separase activities integrated? The simple scenario of a CDC20-like function of CDC20B is very unlikely as it does not appear to bind APC/C (Supplementary Table 2). CDC20 was detected in cultured murine ependymal MCCs during the phase of centriole disengagement<sup>38</sup>, and FZR1 genetic ablation was reported to cause reduced production of centrioles and cilia in the same cells<sup>43</sup>. APC/C is therefore likely activated in maturing MCCs by its classical activators, CDC20 and/or FZR1, leading to Separase activation through degradation of its inhibitor Securin. In that context, we propose that additional factors linked directly or indirectly to CDC20B may contribute to activation of Separase. It was shown that SPAG5 inhibits or activates Separase depending on its status of phosphorylation<sup>41,42</sup>. As the phosphorylation status of SPAG5 was shown to be controlled by PLK1<sup>44</sup>, our data suggest that the CDC20B/PLK1/SPAG5 complex could control the timing of Separase activation locally in deuterosomes. It is therefore possible that multiple modes of activation of Separase may act in parallel to trigger the release of neo-synthesized centrioles in maturing MCCs. Alternatively, different pathways may be used in distinct species, or in distinct types of MCCs. An important question for future studies regards the identity of PLK1 and Separase substrates involved in centriole disengagement. Work on mitotic cells<sup>36,37</sup> and our own analysis suggest that PCNT may represent a prime target. Another potentially relevant

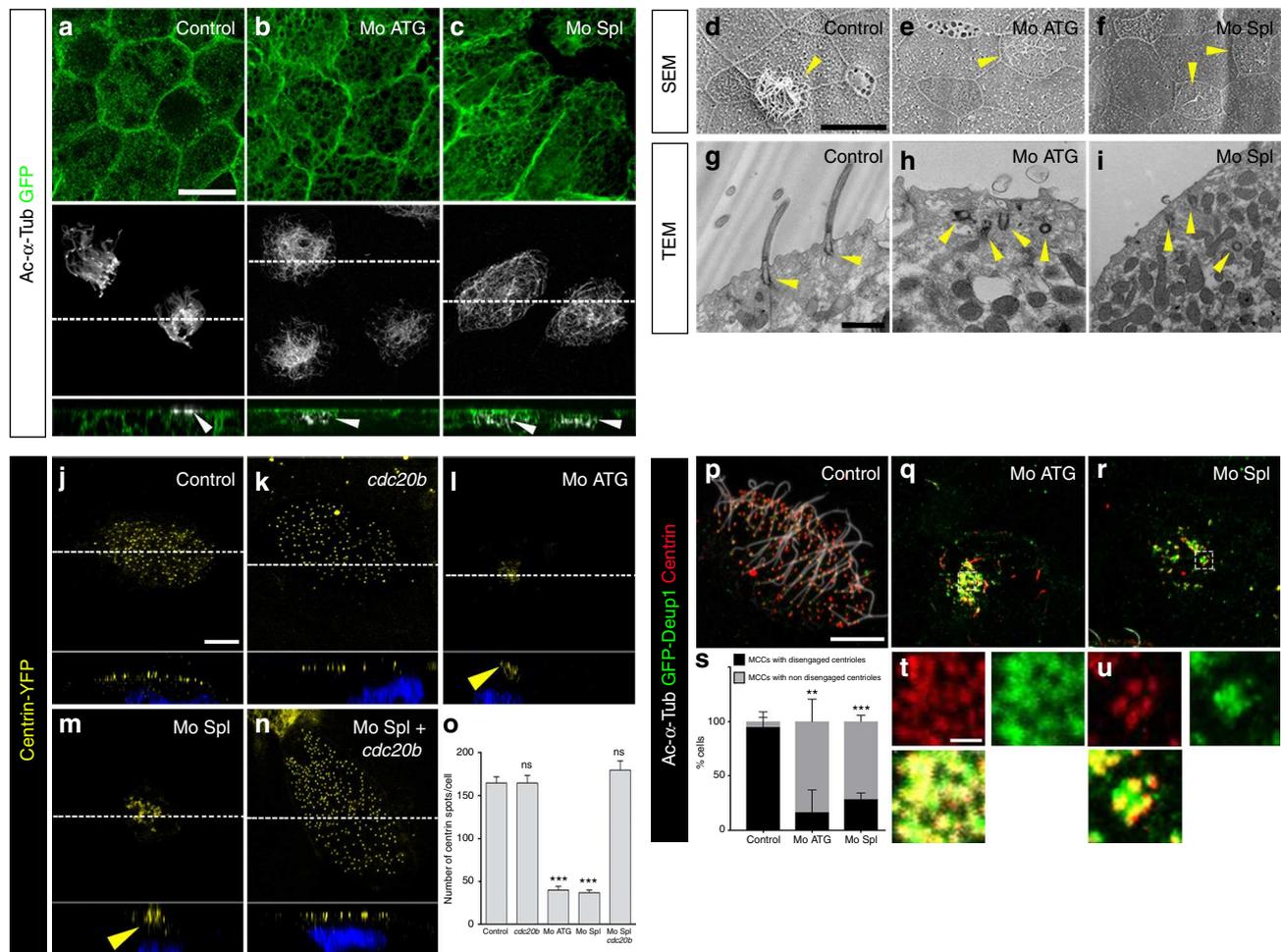


**Fig. 4** CDC20B knockdown impairs multiciliogenesis in mouse ependymal MCCs. **a, b** Ependyma were stained for CDC20B (green) and FOXJ1 (nuclear MCC fate marker, red) 5 days post electroporation (5dpe) of control shRNA (**a**) or *Cdc20b* shRNA (**b**). sh277 is exemplified here, but all three *Cdc20b* shRNAs produced similar effects. **c** Graph showing the quantification of CDC20B protein levels in cells at the deuterosomal stage at 5dpe from two experiments. Mean values and standard error are shown. Unpaired *t*-test: \*\*\*\* $p < 0.0001$ . **d** Dot plot showing the number of FOXJ1-positive nuclei observed for each field, with mean values and standard deviations from two experiments. Unpaired *t*-test:  $p = 0.3961$  (sh273, ns),  $p = 0.1265$  (sh274, ns),  $p = 0.3250$  (sh277, ns). No significant variations were observed between conditions, indicating that MCC fate acquisition was not affected by *Cdc20b* knockdown. **e, f** Confocal pictures of 9dpe ependyma electroporated with control shRNA (**e**) or *Cdc20b* shRNAs (**f**) and stained for DEUP1 (deuterosome, green), FOP (centrioles, red) and ZO1 (cell junction, white). DEUP1-positive deuterosomes with non-disengaged FOP-positive centrioles were observed much more frequently in MCCs electroporated with *Cdc20b* shRNAs compared to control. **g** Dot plot showing the percentage of MCCs with non-disengaged centrioles per field, with mean values and standard deviations. Two experiments were analyzed. Unpaired *t*-test: \*\*\*\* $p < 0.0001$ . **h, i** Confocal pictures of 15dpe ependyma stained for FOP (centrioles, green),  $\alpha$ -Tubulin ( $\alpha$ -TUB, cilia, red), and ZO1 (cell junction, white) showing the morphology of normal MCCs in shRNA control condition (**h**), and examples of defects observed in MCCs treated with sh *Cdc20b* (**i**). **j** Dot plot showing the number of released centrioles per cell, with mean values and standard deviations. **k** Dot plot showing the percentage of normal and abnormal MCCs per field of observation, with mean values and standard deviations. MCCs were scored abnormal when they did not display organized centriole patches associated to cilia. Three experiments were analyzed. Unpaired *t*-test:  $p = 0.0004$  (sh273, \*\*\*),  $p = 0.0001$  (sh274, \*\*\*\*),  $p = 0.0038$  (sh277, \*\*). Scale bars: 20  $\mu$ m (**a**), 5  $\mu$ m (**e, i**)

candidate could be DEUP1 itself as it is clear that deuterosomes are disassembled after the release of centrioles. In that respect, it is interesting to note the presence of multiple PLK1 consensus phosphorylation sites in human, mouse, and *Xenopus* DEUP1.

In this study, we have introduced the notion of perideuterosomal material, in analogy to the pericentriolar material. It is

striking that the two main components of the PCM, PCNT, and  $\gamma$ -Tubulin, are also present in the PDM, which begs the question whether additional PCM proteins may be present in the PDM. The PDM may constitute a platform to sustain procentriole growth, through the concentration and delivery of elementary parts. It could also have a mechanical role to hold in place the



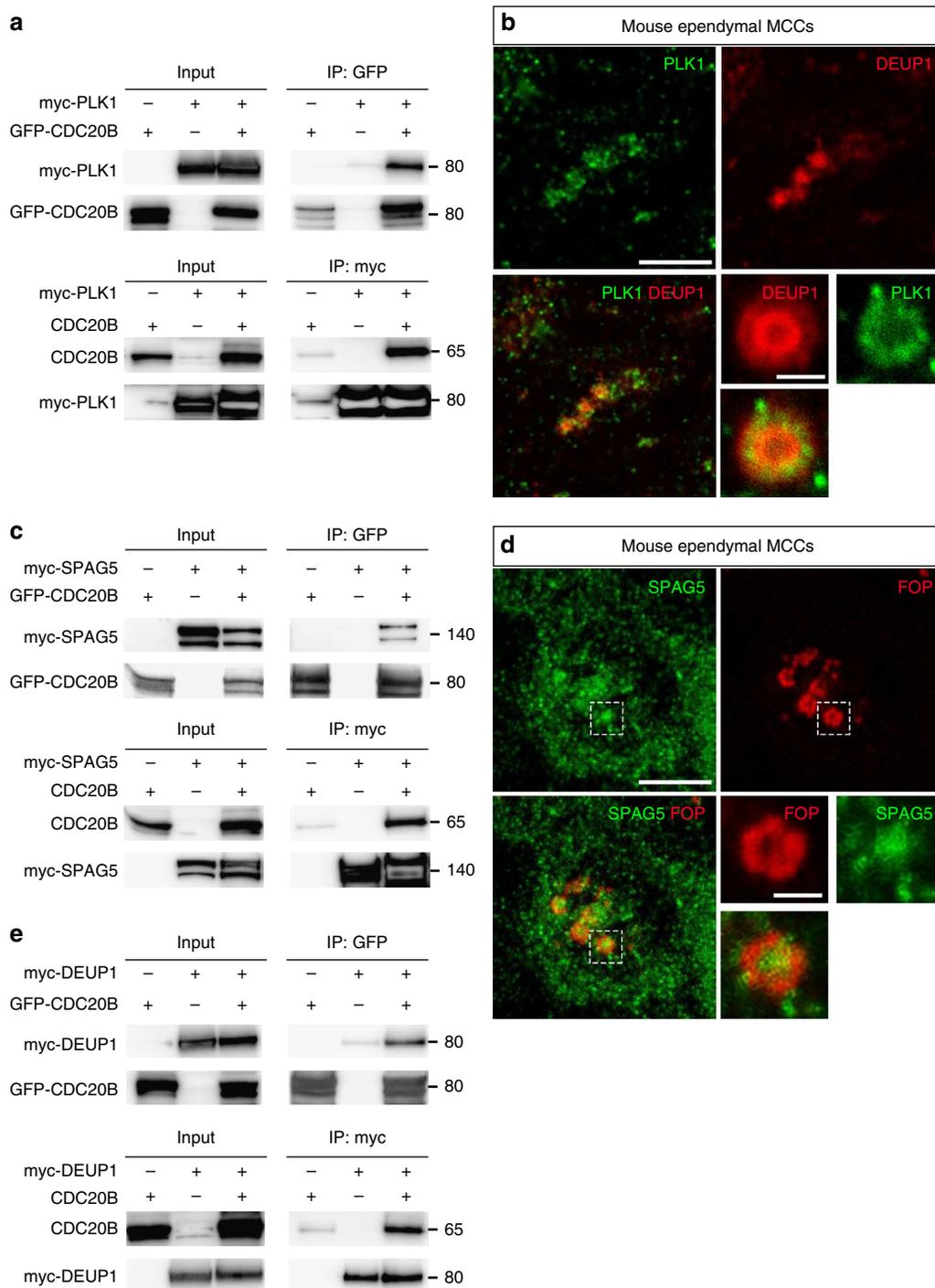
**Fig. 5** *cdc20b* knockdown impairs multiciliogenesis in *Xenopus* epidermal MCCs. **a–c** 8-cell embryos were injected in presumptive epidermis with GFP-CAAX mRNA and *cdc20b* morpholinos, as indicated. Embryos at tailbud st25 were processed for fluorescent staining against GFP (injection tracer, green) and Acetylated- $\alpha$ -Tubulin (Ac- $\alpha$ -Tub, cilia, white). White dotted lines indicate the position of orthogonal projections shown in bottom panels. Note that *cdc20b* morphant MCCs display cytoplasmic filaments but do not grow cilia (white arrowheads). **d–f** Scanning electron microscopy (SEM) of control (**d**) and *cdc20b* morphant (**e, f**) embryos at tadpole st31. Yellow arrowheads point at normal (**d**) and defective MCCs (**e, f**). **g–i** Transmission electron microscopy (TEM) of control (**g**) and *cdc20b* morphant (**h, i**) embryos at tailbud st25. Yellow arrowheads point at normally docked basal bodies supporting cilia (**g**) and undocked centrioles unable to support cilia (**h, i**). **j–n** 8-cell embryos were injected in presumptive epidermis with *centrin-YFP* mRNA, *cdc20b* morpholinos, and *cdc20b* mRNA, as indicated. Centrin-YFP fluorescence was observed directly to reveal centrioles (yellow). Nuclei were revealed by DAPI staining in blue. White dotted lines indicate the position of orthogonal projections shown in bottom panels. Yellow arrowheads point at undocked centrioles. **o** Bar graph showing the mean number of BBs per MCC, and standard error mean, as counted by Centrin-YFP dots. One-way ANOVA and Bonferroni's multiple comparisons test on two experiments, \*\*\* $p < 0.0001$ . *cdc20b* knockdown significantly reduced the number of BBs per cell, and this defect could be corrected by *cdc20b* co-injection with Mo Spl. **p–u** Embryos were injected with *Multicilin-hGR* and *GFP-Deup1* mRNAs, treated with dexamethasone at gastrula st11 to induce Multicilin activity, and immunostained at neurula st23 against Acetylated- $\alpha$ -tubulin (cilia, white), GFP (deuterosomes, green), and Centrin (centrioles, red). **p** Control cells showed individual centrioles, many of which had initiated ciliogenesis. Note that Deup1-positive deuterosomes were no longer visible at this stage. (**q, r, t, u**) *cdc20b* morphant MCCs showed procentrioles still engaged on deuterosomes and lacked cilia. In **t** and **u**, zooms were made on regions identified by dashed boxes in **q** and **r**. **s** Bar graph showing the mean percentage of cells that completed or not centriole disengagement with standard deviations. Three experiments were analyzed. Unpaired *t*-test:  $p = 0.0037$  (Mo ATG, \*\*),  $p = 0.0004$  (Mo Spl, \*\*\*). Scale bars: 20  $\mu\text{m}$  (**a, d**), 1  $\mu\text{m}$  (**g, t**), 5  $\mu\text{m}$  (**j, p**)

growing procentrioles. Future work should evaluate deuterosome-mediated centriole synthesis in absence of major PDM components.

We found that beyond its association to deuterosomes during the phase of centriole amplification, CDC20B was also associated to BBs and cilia in fully differentiated mammalian MCCs. This dual localization is consistent with failed ciliogenesis upon CDC20B knockdown in mouse ependymal MCCs. However, while we could detect Cdc20b near BBs of mature MCCs in *Xenopus*, we found no evidence of its presence in cilia. Furthermore, cilia were rescued by Separase overexpression in Cdc20b

morphant MCCs. This suggests that Cdc20b is not required for ciliogenesis in this species, although it could potentially contribute to cilium structure and/or function. Thus, refined temporal and spatial control of CDC20B inhibition will be needed to study its function beyond centriole synthesis.

This and previous studies<sup>23,26–28</sup> establish that the miR-449 cluster and its host gene *CDC20B* are commonly involved in multiciliogenesis. Consistent with its early expression, it was suggested that miR-449 controls cell cycle exit and entry into differentiation of MCCs<sup>23,27,30</sup>. This study reveals that CDC20B itself is involved in the production of centrioles, the first key step

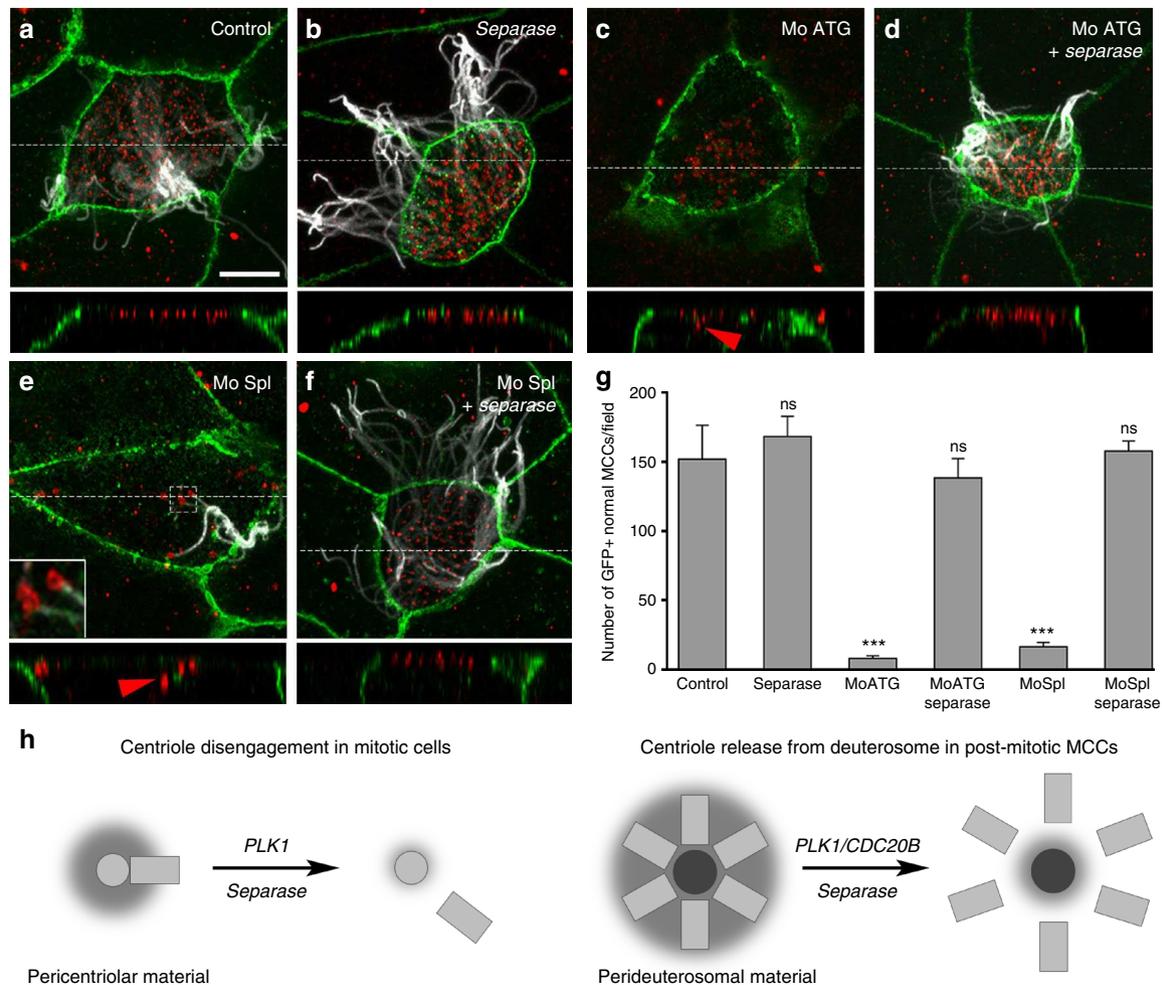


**Fig. 6** CDC20B interacts with PLK1, SPAG5, and DEUP1. **a, c, e** Co-immunoprecipitations of PLK1, SPAG5, and DEUP1 with CDC20B were tested after transfections of different constructs in HEK cells, indicated at the top of each panel. Proteins (left legend) were revealed by immunoblotting. Representative blots from 3 independent experiments. **b, d** Maturing mouse ependyma were immunostained for the indicated proteins, and pictures were taken with a confocal microscope. PLK1 and SPAG5 are expressed in maturing MCCs. In **b**, zoom was made on a different cell to reveal PLK1 enrichment in the perideuterosomal region. In **d**, zoom was made on the dashed box to reveal SPAG5 enrichment in the deuterosome core. Scale bars in **b, d**: 5  $\mu$ m (large view), 1  $\mu$ m (high magnification)

of the multiciliogenesis process. From that perspective, the nested organization of miR-449 and *CDC20B* in vertebrate genomes, which allows their coordinated expression, appears crucial for successful multiciliogenesis.

It is also noteworthy to point out the location of this gene in a genomic locus where congenital mutations in MCIDAS and CCNO were recently shown to cause a newly-recognized MCC-

specific disease, called reduced generation of multiple motile cilia (RGMC). RGMC is characterized by severe chronic lung infections and increased risk of infertility<sup>12,13</sup>. Its location in the same genetic locus as MCIDAS and CCNO makes *CDC20B* a putative candidate for RGMC. By extension, the deuterosome stage-specific genes uncovered by scRNA-seq in this study also represent potential candidates for additional RGMC mutations.



**Fig. 7** Separase overexpression rescues multiciliogenesis in absence of Cdc20b. **a–f** 8-cell *Xenopus* embryos were injected in the presumptive epidermis with *GFP-gpi* mRNA, *cdc20b* morpholinos, and human *Separase* mRNA, as indicated. Embryos were fixed at tailbud st25 and immunostained against GFP (injection tracer, green), Acetylated- $\alpha$ -Tubulin (cilia, white) and  $\gamma$ -Tubulin (BBs, red). White dotted lines indicate the position of orthogonal projections shown in bottom panels. Red arrowheads point undocked BBs. Left inset in **e** shows zoom on clustered centrioles. **g** Bar graph showing the mean number of properly ciliated MCCs among injected cells, per field of observation, with standard error mean, from two independent experiments. Counting was performed on pictures taken at low magnification ( $\times 20$ ), in order to score a large number of cells. Separase overexpression fully rescued multiciliogenesis in *cdc20b* morphant MCCs. One-way ANOVA and Bonferroni's multiple comparisons on two experiments,  $***p < 0.0001$ ,  $^{ns}p > 0.05$ . Scale bars:  $5 \mu\text{m}$  (**a**). **h** Model illustrating the analogy between centriole disengagement in mitotic cells and centriole release from deuterosomes in post-mitotic MCCs

Previous works have established the involvement of the centriole duplication machinery active in S-phase of the cell cycle, during centriole multiplication of vertebrate post-mitotic MCCs<sup>19–21</sup>. Our study further reveals a striking analogy between centriole disengagement from deuterosomes in MCCs, and centriole disengagement that occurs during the M/G1 transition of the cell cycle (Fig. 7g). Thus, it appears that centriole production in MCCs recapitulates the key steps of the centriole duplication cycle<sup>34</sup>. However, the cell cycle machinery must adapt to the acentriolar deuterosome to massively produce centrioles. Such adaptation appears to involve physical and functional interactions between canonical cell cycle molecules, such as CEP152 and PLK1, and recently evolved cell cycle-related deuterosomal molecules, such as DEUP1<sup>21</sup> and CDC20B. It remains to examine whether additional deuterosomal cell cycle-related molecules have emerged in the vertebrate phylum to sustain massive centriole production.

In conclusion, this work illustrates how coordination between ancestral and recently evolved cell cycle-related molecules can give rise to a novel differentiation mechanism in vertebrates.

## Methods

**Subjects/human samples.** Inferior turbinates were from patients who underwent surgical intervention for nasal obstruction or septoplasty (provided by L. Castillo, Nice University Hospital, France). Experiments involving human tissues were performed according to the guidelines of the Declaration of Helsinki, after approval by the institutional review board “Comité de Protection des Personnes Sud Méditerranée V” (06/16/2015). All patients gave their written informed consent.

## Single-cell RNA sequencing of human airway epithelial cells (HAECs)

HAECs cultures were derived from nasal mucosa of inferior turbinates. After excision, nasal inferior turbinates were immediately immersed in  $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free HBSS supplemented with 25 mM HEPES, 200 U/mL penicillin, 200  $\mu\text{g}/\text{mL}$  streptomycin, 50  $\mu\text{g}/\text{mL}$  gentamicin sulfate, and 2.5  $\mu\text{g}/\text{mL}$  amphotericin B (all reagents from Gibco). After repeated washes with cold supplemented HBSS, tissues were digested with 0.1% Protease XIV from *Streptomyces griseus* (Sigma) overnight at 4 °C. After incubation, fetal calf serum (FCS) was added to a final concentration of 10%, and nasal epithelial cells were detached from the stroma by gentle agitation. Cell suspensions were further dissociated by trituration through a 21 G-needle and then centrifuged at 150 $\times g$  for 5 min. The pellet was resuspended in supplemented HBSS containing 10% FCS and centrifuged again. The second cell pellet was then suspended in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 10% FCS and cells were plated (20 000 cells per  $\text{cm}^2$ ) on 75  $\text{cm}^2$ -flasks coated with rat tail collagen I (Sigma-Aldrich). Cells were incubated in a humidified atmosphere of 5%  $\text{CO}_2$  at 37 °C. Culture medium was replaced with Bronchial Epithelium Basal Medium (BEBM, Lonza) supplemented with BEGM SingleQuot Kit Supplements

(Lonza) on the day after and was then changed every other day. After 4 to 5 days of culture, after reaching about 70% confluence, cells were detached with trypsin-EDTA 0.05% (Gibco) for 5 min and seeded on Transwell® permeable supports (6.5 mm diameter; 0.4 µm pore size; Corning), in BEGM medium, with a density of 30,000 cells per Transwell®. Once the cells have reached confluence (typically after 5 days), they were induced to differentiate at the air–liquid interface by removing medium at the apical side of the Transwell®, and by replacing medium at the basal side with DMEM:BEEM (1:1) supplemented with BEGM SingleQuot Kit Supplements. Culture medium was changed every other day. Single-cell analysis was performed after 14 days of culture at the air–liquid interface, which corresponds to the maximum centriole multiplication stage. To obtain a single-cell suspension, cells were incubated with 0.1% protease type XIV from *S. griseus* in supplemented HBSS for 4 h at 4 °C. Cells were gently detached from Transwells® by pipetting and then transferred to a microtube. 50 units of DNase I (EN0523 ThermoFisher Scientific) per 250 µL were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150×g for 5 min) and resuspended in 500 µL supplemented HBSS containing 10% FCS, centrifuged again (150×g for 5 min) and resuspended in 500 µL HBSS before being mechanically dissociated through a 26 G syringe (4 times). Finally, cell suspensions were filtered through a Scienceware® Flowmi™ Cell Strainer (40 µm porosity), centrifuged (150×g for 5 min) and resuspended in 500 µL of cold HBSS. Cell concentration measurements were performed with Scepter™ 2.0 Cell Counter (Millipore) and Countess™ automated cell counter (ThermoFisher Scientific). Cell viability was checked with Countess™ automated cell counter (ThermoFisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10× genomics device, the cell concentration was adjusted to 300 cells/µL in HBSS aiming to capture 1500 cells. We then followed the manufacturer's protocol (Chromium™ Single Cell 3' Reagent Kit, v2 Chemistry) to obtain single cell 3' libraries for Illumina sequencing. Libraries were sequenced with a NextSeq 500/550 High Output v2 kit (75 cycles) that allows up to 91 cycles of paired-end sequencing: the forward read had a length of 26 bases that included the cell barcode and the UMI; the reverse read had a length of 57 bases that contained the cDNA insert. Cell Ranger Single-Cell Software Suite v1.3 was used to perform sample demultiplexing, barcode processing and single-cell 3' gene counting using default parameters and human build hg19. Additional analyses were performed using R. Pseudotemporal ordering of single cells was performed with the last release of the Monocle package<sup>45</sup>. Cell cycle scores were calculated by summing the normalized intensities of genes belonging to phase-specific gene sets then centered and scaled by phase. Gene sets for each phase were curated from previously described sets of genes<sup>46</sup> (Table S2). Data was submitted to the GEO portal under series reference GSE103518. Data shown in Fig. 1 is representative of four independent experiments performed on distinct primary cultures.

**RNA sequencing of HAECs.** For Supplementary Fig. 2B, three independent HAEC cultures (HAEC1, HAEC2, HAEC3) were triggered to differentiate in air–liquid interface (ALI) cultures for 2 days (ALI day 2, undifferentiated), ALI day 14 (first cilia), or ALI day 28 (well ciliated). RNA was extracted with the miRNeasy mini kit (Qiagen) following manufacturer's instructions. mRNA-seq was performed from 2 µg of RNA that was first subjected to mRNA selection with Dynabeads® mRNA Purification Kit (Invitrogen). mRNA was fragmented 10 min at 95 °C in RNaseIII buffer (Invitrogen) then adapter-ligated, reverse transcribed and amplified (6 cycles) with the reagents from the NEBNext Small RNA Library Prep Set for SOLiD. Small RNA-seq was performed from 500 ng RNA with the NEBNext Small RNA Library Prep Set for SOLiD (12 PCR cycles) according to manufacturer's instructions. Both types of amplified libraries were purified on Purelink PCR micro kit (Invitrogen), then subjected to additional PCR rounds (8 cycles for RNA-seq and 4 cycles for small RNA-seq) with primers from the 5500 W Conversion Primers Kit (Life Technologies). After Agencourt® AMPure® XP beads purification (Beckman Coulter), libraries were size-selected from 150 nt to 250 nt (for RNA-seq) and 105 nt to 130 nt (for small RNA-seq) with the LabChip XT DNA 300 Assay Kit (Caliper Lifesciences), and finally quantified with the Bioanalyzer High Sensitivity DNA Kit (Agilent). Libraries were sequenced on SOLiD 5500XL (Life Technologies) with single-end 50b reads. SOLiD data were analyzed with lifescape v2.5.1, using the small RNA pipeline for miRNA libraries and whole transcriptome pipeline for RNA-seq libraries with default parameters. Annotation files used for production of raw count tables correspond to Refseq Gene model v20130707 for mRNAs and miRBase v18 for small RNAs. Data generated from RNA sequencing were then analyzed with Bioconductor (<http://www.bioconductor.org>) package DESeq and size-factor normalization was applied to the count tables. Heatmaps were generated with GenePattern using the “Hierarchical Clustering” Module, applying median row centering and Euclidian distance.

**Re-analysis of *Xenopus* E2F4 Chip-seq and RNA-seq.** RNA-seq (samples GSM1434783 to GSM1434788) and ChIP-seq (samples GSM1434789 to GSM1434792) data were downloaded from GSE59309. Reads from RNA-seq were aligned to the *Xenopus laevis* genome release 7.1 using TopHat<sup>47</sup> with default parameters. Quantification of genes was then performed using HTSeq-count<sup>48</sup> release 0.6.1 with “-m intersection-nonempty” option. Normalization and statistical analysis were performed using Bioconductor package DESeq2<sup>49</sup>. Differential expression analysis was done between Multicilin-hGR alone versus Multicilin-hGR

in the presence of E2F4ΔCT. Reads from ChIP-seq were mapped to the *X. laevis* genome release 7.1 using Bowtie2<sup>50</sup>. Peaks were called and annotated according to their positions on known exons with HOMER<sup>51</sup>. Peak enrichments of E2F4 binding site in the promoters of centriole genes and cell cycle genes<sup>9</sup> were estimated in presence or absence of Multicilin and a ratio of E2F4 binding (Multicilin vs no Multicilin) was calculated.

**Promoter reporter studies.** The human *CDC20B* promoter was cloned into the pGL3 Firefly Luciferase reporter vector (Promega) with SacI and NheI cloning sites. The promoter sequenced ranged from −1073 to +104 relative to the transcription start site. 37.5 ng of pGL3 plasmid were applied per well. pCMV6-Neg, pCMV6-E2F1 (NM\_005225) and pCMV6-E2F4 (NM\_001950) constructs were from Origene. 37.5 ng of each plasmid was applied per well. 25 ng per well of pRL-CMV (Promega) was applied in the transfection mix for transfection normalization (Renilla luciferase). HEK 293T cells were seeded at 20,000 cells per well on 96-well plates. The following day, cells were transfected with the indicated plasmids (100 ng of total DNA) with lipofectamine 3000 (Invitrogen). After 24 h, cells were processed with the DualGlo kit (Promega) and luciferase activity was recorded on a plate reader.

**Proximity ligation assays.** Fully differentiated HAECs were dissociated by incubation with 0.1% protease type XIV from *S. griseus* (Sigma-Aldrich) in HBSS (Hanks' balanced salts) for 4 h at 4 °C. Cells were gently detached from the Transwells® by pipetting and then transferred to a microtube. Cells were then cytocentrifuged at 72×g for 8 min onto SuperFrostPlus slides using a Shandon Cytospin 3 cytocentrifuge. Slides were fixed for 10 min in methanol at −20 °C for Centri2 and ZO1 assays, and for 10 min in 4% paraformaldehyde at room temperature and then permeabilized with 0.5% Triton X-100 in PBS for 10 min for acetylated-α-tubulin assays. Cells were blocked with 3% BSA in PBS for 30 min. The incubation with primary antibodies was carried out at room temperature for 2 h. Then, mouse and rabbit secondary antibodies from the Duolink® Red kit (Sigma-Aldrich) were applied and slides were processed according to manufacturer's instructions. Images were acquired using the Olympus Fv10i confocal imaging systems with ×60 oil immersion objective and Alexa 647 detection parameters.

**Animals.** All experiments were performed following the Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. Experiments on *X. laevis* and mouse were approved by the ‘Direction départementale de la Protection des Populations, Pôle Alimentation, Santé Animale, Environnement, des Bouches du Rhône’ (agreement number F 13 055 21). Mouse experiments were approved by the French ethical committee no.14 (permission number: 62-12112012). Timed pregnant CD1 mice were used (Charles Rivers, Lyon, France).

**Immunostaining on mouse ependyma.** Dissected brains were subjected to 12 min fixation in 4% paraformaldehyde, 0.1% Triton X-100, blocked 1 h in PBS, 3% BSA, incubated overnight with primary antibodies diluted in PBS, 3% BSA, and incubated 1 h with secondary antibodies at room temperature. Ependyma were dissected further and mounted with Mowiol before imaging using an SP8 confocal microscope (Leica microsystems) equipped with a ×63 oil objective. The same protocol was used to prepare samples for super-resolution acquisition. Pictures were acquired with a TCS SP8 STED ×3 microscope equipped with an HC PL APO 93×/1.30 GLYC motCORR™ objective (Leica microsystems). Pericentrin was revealed using Alexa 514 (detection 535–564 nm, depletion 660 nm), γ-tubulin was revealed using Alexa 568 (detection 582–667 nm, depletion 775), and FOP was revealed using Alexa 488 (detection 498–531 nm, depletion 592 nm). Pictures were deconvoluted using Huygens software. Maximum intensity projection of 3 deconvoluted pictures is presented in Fig. 4g. Primary antibodies: rabbit anti-CDC20B (1:500; Proteintech, 133376-1-AP), mouse IgG anti-PLK1 (1:500; ThermoFisher, 33–1700), rabbit anti-Pericentrin (1:500, Abcam, ab4448), mouse IgG1 anti-FoxJ1 (1:1000; eBioscience, 14–9965), rabbit anti-Deup1 (1:1000; kindly provided by Dr Xueliang Zhu), rabbit anti-Deup1 (1:250; Proteintech, 24579-1-AP), mIgG1 anti-γ-Tubulin (clone GTU88) (1:250; Abcam, Ab 11316), rabbit anti-ZO1 (1:600; ThermoFisher Scientific, 61–7300), rabbit anti-Spag5 (1:500; Proteintech, 14726-1-AP), mouse IgG1 anti-ZO1 (1:600; Invitrogen, 33-9100), mouse IgG2b anti-FGFR1OP (FOP) (1:2000; Abnova, H00011116-M01), mouse IgG1 anti-α-tubulin (1:500; Sigma-Aldrich, T9026). Secondary antibodies: Alexa Fluor 488 goat anti-rabbit (1:800; ThermoFisher Scientific, A-11034), Alexa Fluor 647 goat anti-rabbit (1:800; ThermoFisher Scientific, A-21244), Alexa Fluor 514 goat anti-rabbit (1:800; ThermoFisher Scientific, A-31558), Alexa Fluor 488 goat anti-mouse IgG2b (1:800; ThermoFisher Scientific, A-21141), Alexa Fluor 568 goat anti-mouse IgG2b (1:800; ThermoFisher Scientific, A-21144), Alexa Fluor 488 goat anti-mouse IgG2a (1:800; ThermoFisher Scientific, A-21131), Alexa Fluor 568 goat anti-mouse IgG1 (1:800; ThermoFisher Scientific, A-21134), Alexa Fluor 647 goat anti-mouse IgG1 (1:800; ThermoFisher Scientific, A-21240).

**Mouse constructs.** Expression constructs containing shRNA targeting specific sequences in the *CDC20B* coding sequence under the control of the U6 promoter

were obtained from Sigma-Aldrich (ref. TRCN0000088273 (sh273), TRCN0000088274 (sh274), TRCN0000088277 (sh277)). PCX-mcs2-GFP vector (Control GFP) kindly provided by Xavier Morin (ENS, Paris, France), and U6 vector containing a validated shRNA targeting a specific sequence in the NeuroD1 coding sequence<sup>52</sup> (Control sh, ref. TRCN0000081777, Sigma-Aldrich) were used as controls for electroporation experiments.

**Post-natal mouse brain electroporation.** The detailed protocol for post-natal mouse brain electroporation established by Boutin and colleagues<sup>53</sup> was used with minor modifications. Briefly, P1 pups were anesthetized by hypothermia. A glass micropipette was inserted into the lateral ventricle, and 2  $\mu$ L of plasmid solution (concentration 3  $\mu$ g/ $\mu$ L) was injected by expiratory pressure using an aspirator tube assembly (Drummond). Successfully injected animals were subjected to five 95 V electrical pulses (50 ms, separated by 950 ms intervals) using the CUY21 edit device (Nepagene, Chiba, Japan), and 10 mm tweezer electrodes (CUY650P10, Nepagene) coated with conductive gel (Signagel, Parker laboratories). Electroporated animals were reanimated in a 37 °C incubator before returning to the mother.

**Statistical analyses of mouse experiments.** Analysis of CDC20B signal intensity in deuterosomes (dot plot in Fig. 3b). For each category, >25 cells from two different animals were analyzed. Deuterosome regions were delineated based on FOP staining and the intensity of CDC20B fluorescent immunostaining was recorded using ImageJ software, and expressed as arbitrary units. Unpaired *t* test vs immature: *p* = 0.0005 (intermediate, \*\*\*); *p* < 0.0001 (Mature, \*\*\*\*).

Analysis of *Cdc20b* shRNAs efficiency (Fig. 4c): For each cell at the deuterosomal stage, the intensity of CDC20B fluorescent immunostaining was recorded using ImageJ software and expressed as arbitrary units. Data are mean  $\pm$  sem. Two independent experiments were analyzed. A minimum of 35 cells per condition was analyzed. *n* = 3, 4, 5 and 5 animals for sh control, sh273, sh274, and sh277, respectively. Unpaired *t* test vs sh control: *p* < 0.0001 (sh273, sh274, and sh277 \*\*\*\*).

Analysis of the number of FOXJ1-positive cells at 5dpe (Fig. 4d): Unpaired *t* test vs sh control: 0.3961 (sh273, ns), 0.1265 (sh274, ns), 0.3250 (sh277, ns).

Analysis of the number of cells with non-disengaged centrioles at 9dpe (Fig. 4g): 15–20 fields were analyzed per condition. *n* = 4, 4, 3, and 4 animals for sh control, sh273, sh274, and sh277, respectively, from two independent experiments. Unpaired *t* test vs sh control: *p* < 0.0001 (sh273, sh274, sh277 \*\*\*\*).

Analysis of the number of centrioles per cell at 15dpe (Fig. 4j): > 100 cells were analyzed per condition. *n* = 3, 3, 3, and 3 animals for sh control, sh273, sh274, and sh277, respectively, from two independent experiments. Unpaired *t* test vs sh control: *p* < 0.0001 (sh273, sh274, sh277 \*\*\*\*).

Analysis of ependymal cell categories at 15dpe (Fig. 4k): Data are mean  $\pm$  sem from three independent experiments. More than 500 cells were analyzed for each condition. *n* = 4, 4, 3, and 3 animals for sh control, sh273, sh274, and sh277, respectively. Unpaired *t* test vs sh control: *p* = 0.0004 (sh273, \*\*\*), 0.0001 (sh274, \*\*\*\*), 0.0038 (sh277, \*\*).

**Mouse tracheal epithelial cells (MTECs).** MTECs cell cultures were established from the tracheas of 12 weeks-old mice. After dissection, tracheas were placed in cold DMEM:F-12 medium (1:1) supplemented with 15 mM HEPES, 100 U/mL penicillin, 100  $\mu$ g/mL streptomycin, 50  $\mu$ g/mL gentamicin sulfate, and 2.5  $\mu$ g/mL amphotericin B. Each trachea was processed under a binocular microscope to remove as much conjunctive tissue as possible with small forceps and was opened longitudinally with small dissecting scissors. Tracheas were then placed in supplemented DMEM:F-12 containing 0.15% protease XIV from *S. griseus*. After overnight incubation at 4 °C, FCS was added to a final concentration of 10%, and tracheal epithelial cells were detached by gentle agitation. Cells were centrifuged at 400  $\times$ g for 10 min and resuspended in supplemented DMEM:F-12 containing 10% FCS. Cells were plated on regular cell culture plates and maintained in a humidified atmosphere of 5% CO<sub>2</sub> at 37 °C for 4 h to allow attachment of putative contaminating fibroblast. Medium containing cells in suspension was further centrifuged at 400  $\times$ g for 5 min and cells were resuspended in supplemented DMEM:F-12 containing BEGM Singlequots kit supplements and 5% FCS. Cells were plated on rat tail collagen I-coated Transwell®. Typically, 5 tracheas resulted in 12 Transwells®. Medium was changed every other day. Air-liquid interface culture was conducted once transepithelial electrical resistance had reached a minimum of 1000 ohm/cm<sup>2</sup> (measured with EVOM2, World Precision Instruments).

Air-liquid interface culture was obtained by removing medium at the apical side of the Transwell®, and by replacing medium at the basal side with supplemented DMEM:F-12 containing 2% Ultraser-G™ (Pall Corporation). 10  $\mu$ M DAPT (*N*-[*N*-(3,5-difluorophenacetyl)-*L*-alanyl]-*S*-phenylglycine *t*-butyl ester) (Sigma) was added one day after setting-up the air-liquid interface.

**Immunostaining on HAECs and MTECs.** Three days after setting-up the air-liquid interface, MTECs on Transwell membranes were pre-extracted with 0.5% Triton X-100 in PBS for 3 min, and then fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. HAECs were treated 21 days after setting-up the air-liquid interface. They were fixed directly on Transwells® with 100% cold methanol for 10 min at –20 °C (for CDC20B and Centrin2 co-staining,

Supplementary Figure 4a, b) or with 4% paraformaldehyde in PBS for 15 min at room temperature (for CDC20B single staining, Supplementary Figure 4c). All cells were then permeabilized with 0.5% Triton X-100 in PBS for 5 min and blocked with 3% BSA in PBS for 30 min. The incubation with primary and secondary antibodies was carried out at room temperature for 2 h and 1 h, respectively. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI). Transwell® membranes were cut with a razor blade and mounted with ProLong Gold medium (ThermoFisher). Primary antibodies: rabbit anti-CDC20B (1:500; Proteintech, 133376-1-AP), rabbit anti-DEUP1 (1:500; Proteintech, 24579-1-AP), anti-Centrin2 (Clone 20H5, 1:500; Millipore, 04-1624). Secondary antibodies: Alexa Fluor 488 goat anti-rabbit (1:1000; ThermoFisher Scientific, A-11034), Alexa Fluor 647 goat anti-mouse (1:1000; ThermoFisher Scientific, A-21235). For co-staining of CDC20B and DEUP1, CDC20B primary antibody was directly coupled to CF™ 633 with the Mix-n-Stain™ kit (Sigma-Aldrich) according to the manufacturer's instruction. Coupled primary antibody was applied after secondary antibodies had been extensively washed and after a 30 min blocking stage in 3% normal rabbit serum in PBS.

**Western blot and immunofluorescence on transfected cells.** Cos-1 or HeLa cells were grown in DMEM supplemented with 10% heat inactivated FCS and transfected with Eugene HD (Roche Applied Science) according to manufacturer's protocol. Transfected or control cells were washed in PBS and lysed in 50 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, containing 1% NP-40 and 0.25% sodium deoxycholate (modified RIPA) plus a Complete Protease Inhibitor Cocktail (Roche Applied Science) on ice. Cell extracts separated on polyacrylamide gels were transferred onto Optitran membrane (Whatman) followed by incubation with rabbit anti-mouse CDC20B (1:500, Proteintech, 24579-1-AP) or homemade rabbit anti-*Xenopus* Cdc20b (1:300) antibody and horseradish peroxidase conjugated secondary antibody (Jackson ImmunoResearch Laboratories, 711-035-152 and 715-035-150). Signal obtained from enhanced chemiluminescence (Western Lightning ECL Pro, Perkin Elmer) was detected with MyECL Imager (ThermoFisher Scientific).

For immunofluorescence staining, transfected cells were grown on glass coverslips and fixed for 6 min in methanol at –20 °C. Cells were washed in PBS, blocked in PBS, 3% BSA and stained with rabbit anti-*Xenopus* Cdc20b (1:300) or rabbit anti-CFTR (1:200, Santa-Cruz Biotechnology, 10747) as a negative control, in blocking buffer. After washings in PBS 0.1% Tween-20, cells were incubated with Alexa fluor 488 donkey anti-rabbit antibody (ThermoFisher Scientific, R37118), washed, and DNA was stained with 250 ng/mL DAPI. Coverslip were then rinsed and mounted in Prolong Gold antifade reagent (ThermoFisher Scientific) and confocal images were acquired by capturing Z-series with 0.3  $\mu$ m step size on a Zeiss LSM 510 laser scanning confocal microscope.

**Co-immunoprecipitation studies.** Asynchronous HEK cells transfected with the plasmids described below, using lipofectamine 3000 according to manufacturer's instructions, were rinsed on ice with chilled Ca<sup>2+</sup> and Mg<sup>2+</sup> free Dulbecco's PBS (DPBS, Invitrogen), harvested using a cell scraper and lysed on ice for 5 min in lysis buffer (0.025 M Tris, 0.15 M NaCl, 0.001 M EDTA, 1% NP-40, 5% glycerol; pH 7.4) supplemented with EDTA and Halt™ Protease and Phosphatase Inhibitor Cocktail (Pierce, ThermoFisher). Lysates were clarified (12,000  $\times$ g, 4 °C, 10 min) and the protein concentrations were determined using the Bradford assay (Bio-Rad). Immunoprecipitations were performed with the Pierce co-immunoprecipitation kit (Pierce, ThermoFisher) according to the manufacturer's instructions. For each immunoprecipitation, 1–1.5 mg of total lysate was precleared on a control column, then incubated on columns coupled with 20  $\mu$ g of anti-GFP or anti-c-myc antibody (clone 9E10). Incubation was performed overnight at 4 °C. Columns were washed and eluted with 50  $\mu$ L elution buffer. Samples were denatured at 70 °C for 10 min with Bolt™ LDS Sample Buffer and Bolt reducing agent, then separated on 4–12% gradient Bolt precast gels (ThermoFisher), transferred onto nitrocellulose (Millipore), and subjected to immunoblot analysis using either anti-CDC20B (ProteinTech, 133376-1-AP, 1/500) or anti-c-myc antibody (clone 9E10, 1/1000). In Fig. 6, note that the high level of expression of myc-PLK1 (Fig. 6a) and myc-SPAG5 (Fig. 6b) drained out locally the ECL reagent at the peak of the protein. The resulting double bands correspond in fact to unique ones. Human SPAG5, subcloned into pCMV6-MT, was from OriGene. Human DEUP1 and PLK1 were cloned into pCS2-MT vector (Addgene). Human CDC20B was cloned into pEGFP-C1, pEGFP-N1 (Clontech) for the GFP fusion protein and pIRES-EYFP (Addgene) for the untagged protein.

**In-gel digestion, NanoHPLC, and Q-exactive plus analysis.** For mass spectrometry analysis, protein spots were manually excised from the gel and destained with 100  $\mu$ L of H<sub>2</sub>O/ACN (1/1). After 10 min vortexing, liquid was discarded and the procedure was repeated 2 times. They were rinsed with acetonitrile and dried under vacuum. Extracts were reduced with 50  $\mu$ L of 10 mM dithiothreitol for 30 min at 56 °C, then alkylated with 15  $\mu$ L of 55 mM iodoacetamide for 15 min at room temperature in the dark. They were washed successively by: (i) 100  $\mu$ L of H<sub>2</sub>O/ACN (1/1) (2 times) and (ii) 100  $\mu$ L of acetonitrile. Gel pieces were rehydrated in 60  $\mu$ L of 50 mM NH<sub>4</sub>HCO<sub>3</sub> containing 10 ng/ $\mu$ L of trypsin (modified porcine trypsin, sequence grade, Promega) incubated for one hour at 4 °C. After the

removal of trypsin, samples were incubated overnight at 37 °C. Tryptic peptides were extracted with: (i) 60 µL of 1% FA (formic acid) in water (10 min at RT), (ii) 60 µL acetonitrile (10 min at RT). Extracts were pooled, concentrated under vacuum, resuspended in 15 µL of aqueous 0.1% formic acid for NanoHPLC separation.

Separation was carried out using a nanoHPLC (Ultimate 3000, ThermoFisher Scientific). After concentration on a µ-Precolumn Cartridge Acclaim PepMap 100 C<sub>18</sub> (i.d. 5 mm, 5 µm, 100 Å, ThermoFisher Scientific) at a flow rate of 10 µL/min, using a solution of H<sub>2</sub>O/ACN/FA 98%/2%/0.1%, a second peptide separation was performed on a 75 µm i.d. × 250 mm (3 µm, 100 Å) Acclaim PepMap 100 C<sub>18</sub> column (ThermoFisher Scientific) at a flow rate of 300 nL/min. Solvent systems were: (A) 100% water, 0.1% FA, (B) 100% acetonitrile, 0.08% FA. The following gradient was used  $t = 0$  min 6% B;  $t = 3$  min 6% B;  $t = 119$  min, 45% B;  $t = 120$  min, 90% B;  $t = 130$  min 90% B (temperature at 35 °C).

NanoHPLC was coupled via a nanoelectrospray ionization source to the Hybrid Quadrupole-Orbitrap High Resolution Mass Spectrometer (ThermoFisher Scientific). MS spectra were acquired at a resolution of 70,000 (200  $m/z$ ) in a mass range of 300–2000  $m/z$  with an AGC target 3e6 value of and a maximum injection time of 100 ms. The 10 most intense precursor ions were selected and isolated with a window of 2  $m/z$  and fragmented by HCD (Higher energy C-Trap Dissociation) with normalized collision energy (NCE) of 27. MS/MS spectra were acquired in the ion trap with an AGC target 2e5 value, the resolution was set at 17 500 at 200  $m/z$  combined with an injection time of 100 ms.

Data were reprocessed using Proteome Discoverer 2.1 equipped with Sequest HT. Files were searched against the Swissprot Homo sapiens FASTA database (update of February 2016). A mass accuracy of ±10 ppm was used to precursor ions and 0.02 Da for product ions. Enzyme specificity was fixed to trypsin, allowing at most two miscleavages. Because of the previous chemical modifications, carbamidomethylation of cysteines was set as a fixed modification and only oxidation of methionine was considered as a dynamic modification. Reverse decoy databases were included for all searches to estimate false discovery rates, and filtered using the Percolator algorithm at a 1% FDR.

**Xenopus embryo injections, plasmids, RNAs, and morpholinos.** Eggs obtained from NASCO females were fertilized in vitro, dejellied and cultured using standard protocols<sup>54</sup>. All injections were done at the 8-cell stage in one animal-ventral blastomere (presumptive epidermis), except for electron microscopy analysis for which both sides of the embryo were injected, and for RT-PCR analysis for which 2-cell embryos were injected.

*cdc20b* riboprobe was generated from *X. laevis* cDNA. Full-length sequence was subcloned in pGEM<sup>+</sup>-T Easy Vector Systems (Promega). For sense probe, it was linearized by *SpeI* and transcribed by T7. For antisense probe it was linearized by *ApaI* and transcribed by Sp6 RNA polymerase. Synthetic capped mRNAs were produced with the Ambion mMESAGE mMACHINE Kit. pCS105/GFP-CAAX was linearized with *AseI* and mRNA was synthesized with Sp6 polymerase. pCS2-mRFP and pCS2-GFP-gpi were linearized with *NotI* and mRNA was synthesized with Sp6 polymerase. pCS-Centrin4-YFP (a gift from Reinhard Köster, Technische Universität Braunschweig, Germany) was linearized with *NotI* and mRNA was synthesized with Sp6 polymerase. pCS2-GFP-Deup1 and pCS2-Multicilin(MCI)-hGR were kindly provided by Chris Kintner; both plasmids were linearized with *ApaI*, and mRNAs were synthesized with Sp6 polymerase. Embryos injected with *MCI-hGR* mRNA were cultured in Dexamethasone 20 µM in MBS 0.1× from st11 until fixation. pCS2-Separase wild-type and phosphomutant 2/4 (protease dead, PD) were provided by Marc Kirchner and Olaf Stemann, respectively; plasmids were linearized with *NotI* and mRNAs were synthesized with Sp6 polymerase. *Venus-cdc20b*, *cdc20b-Venus*, and *cdc20b* were generated by GATEWAY<sup>™</sup> Cloning Technology (GIBCO BRL) from *Xenopus laevis cdc20b* cDNA. *cdc20b* was also subcloned in pCS2-RFP to make *RFP-cdc20b* and *cdc20b-RFP* fusions. All *cdc20b* constructs were linearized with *NotI* and mRNAs were synthesized with Sp6 polymerase. Quantities of mRNA injected: 500 pg for *GFP-CAAX*, *RFP*, *GFP-gpi*, *Separase* and *Separase(PD)*; 25 to 500 pg for *GFP-Deup1*; 40 to 500 pg for *MCI-hGR*; 1 ng for *Venus-cdc20b*, *cdc20b-Venus*, *cdc20b*, and *cdc20b-RFP*; 500 pg to 1 ng for *RFP-cdc20b*.

Two independent morpholino antisense oligonucleotides were designed against *cdc20b* (GeneTools, LLC). *cdc20b* ATG Mo: 5'-aaatcttctcaactccagtcac-3', *cdc20b* Spl Mo 5'-acacatggcacaacgtaccacac-3'. 20 ng of MOs was injected per blastomere or 10 ng of each Mo for co-injection.

**PCR and quantitative RT-qPCR.** *Xenopus* embryos were snap frozen at different stages and stored at -80 °C. Total RNAs were purified with a Qiagen RNeasy kit (Qiagen). Primers were designed using Primer-BLAST Software. PCR reactions were carried out using GoTaq<sup>®</sup> G2 Flexi DNA Polymerase (Promega). RT reactions were carried out using iScript<sup>™</sup> Reverse Transcription Supermix for RT-qPCR (BIO-RAD). qPCR reactions were carried out using SYBRGreen on a CFX Bio-rad qPCR cyler. To check *cdc20b* temporal expression by qPCR we directed primers to exons 9/10 junction (Forward: 5'-ggctatgattgggtgcc-3') and exons 10/11 junction (Reverse: 5'-gcaggagcagatctggg-3') to avoid amplification from genomic DNA. The relative expression of *cdc20b* was normalized to the expression of the house-keeping gene *ornithine decarboxylase* (*ODC*) for which primers were as follows: forward: 5'-gccattggaagctctctcattc-3'; reverse: 5'-ttcgggtgattccttgcac-3'.

To check the efficiency of Mo SPL, expected to cause retention of intron 1 in the mature mRNA of *cdc20b* we directed forward (5'-ctccccgagatgagga-3') and reverse (5'-gcatgtgtactcttctgctca-3') primers in exon 1 and exon2, respectively.

To check the expression of *p53* in morphants by qPCR, primers were as follows: forward: 5'-cgcagccgatgagatgatt-3'; reverse: 5'-cattctggcactaatggt-3'. The relative expression of *p53* was normalized to Histone4 expression (H4) for which primers were as follows: forward: 5'-gggtgatccctgattgtt-3'; reverse: 5'-ggcaaggaggaaaaggactg-3'.

**Immunostaining on *Xenopus* embryos.** Embryos were fixed in 4% paraformaldehyde (PFA) overnight at 4 °C and stored in 100% methanol at -20 °C. Embryos were rehydrated in PBT and washed in MABX (Maleic Acid Buffer + Triton X100 0.1% v/v). Next, embryos were incubated in Blocking reagent (Roche) 2% BR + 15% Serum + MABX with respective primary and secondary antibodies. The anti-*Xenopus laevis* CDC20B antibody was obtained by rabbit immunization with the peptide SPDQRRIFSAANGT (amino acids 495–509) conjugated to keyhole limpet hemocyanin, followed by affinity purification (Eurogentec). For immunofluorescence, embryos were fixed at RT in PFA 4% in PBS, and incubated in the CDC20B antibody diluted 1/150 in BSA 3% in PBS. For all experiments, secondary antibodies conjugated with Alexa were used. GFP-CAAX in Supplementary Figure 5g was revealed using a rabbit anti-GFP antibody together with a secondary antibody coupled to Alkaline Phosphatase (AP), which was revealed as follows: embryos incubated with the AP-conjugated antibody were washed twice in alkaline phosphatase buffer (PAB) (NaCl 0.1 M, Tris HCl pH 9.5 0.1 M, MgCl<sub>2</sub> 0.05 M, Tween 0.1%), 10 min each. Next, embryos were incubated in PAB with INT/BCIP substrate (Roche, REF:11681460001) until appropriate staining. Finally embryos were washed twice in MABX and fixed in MEMFA 30 min at RT. To mark cortical actin in MCCs, embryos were fixed in 4% paraformaldehyde (PFA) in PBT (PBS + 0.1% Tween v/v) for 1 h at room temperature (RT), washed 3 × 10 min in PBT at RT, then stained with phalloidin-Alexa Fluor 555 (Invitrogen, 1:40 in PBT) for 4 h at RT, and washed 3 × 10 min in PBT at RT. Primary antibodies: mouse anti-Acetylated- $\alpha$ -Tubulin (Clone 6-11B-1, Sigma-Aldrich, T7451, 1:1000), rabbit anti- $\gamma$ -Tubulin (Abcam, Ab 16504, 1:500), mouse anti- $\gamma$ -Tubulin (Clone GTU88, Ab 11316, Abcam, 1:500), Chicken anti-GFP (AVES, GFP-1020, 1:1000), rabbit anti-GFP (Torrey Pines Biolabs, TP401, 1:500), mouse anti-Centrin (Clone 20H5, EMD Millipore, 04-1624, 1:500). Secondary antibodies: donkey anti-rabbit-AP (Jackson ImmunoResearch, 711055152, 1:1000), Alexa Fluor 647 goat anti-mouse IgG2a (1:500; ThermoFisher Scientific, A-21241), Alexa Fluor 488 goat anti-chicken (1:500; ThermoFisher Scientific, A-11039), Alexa Fluor 568 goat anti-rabbit (1:500; ThermoFisher Scientific, A-11011).

**In situ hybridization on *Xenopus* embryos.** Whole-mount chromogenic in situ hybridization and whole-mount fluorescent in situ hybridization (FISH) was performed as detailed by Marchal and colleagues<sup>54</sup>, and Castillo-Briceno and Kodjabachian<sup>55</sup>, respectively. For single staining, all RNA probes were labeled with digoxigenin. For FISH on section, embryos were fixed in 4% paraformaldehyde (PFA), stored in methanol for at least 4 h at -20 °C, then rehydrated in PBT (PBS + Tween 0.1% v/v), treated with triethanolamine and acetic anhydride, incubated in increasing sucrose concentrations and finally embedded with OCT (VWR Chemicals). 12 µm-thick cryosections were made. Double FISH on sections was an adaptation of the whole-mount FISH method. 80 ng of *cdc20b* digoxigenin-labeled sense and antisense riboprobes and 40 ng of antisense  $\alpha$ -tubulin fluorescein-labeled riboprobe<sup>56</sup> were used for hybridization. All probes were generated from linearized plasmids using RNA-labeling mix (Roche). FISH was carried out using Tyramide Signal Amplification – TSA TM Plus Cyanine 3/Fluorescein System (Perkin Elmer). Antibodies: Anti-DigAP (Roche, 11266026, 1:5000), Anti-DigPOD (Roche, 11207733910, 1:500), Anti-FluoPOD (Roche, 11426346910, 1:500).

**Microscopy.** Confocal: Flat-mounted epidermal explants were examined with a Zeiss LSM 780 confocal microscope. Four-colors confocal z-series images were acquired using sequential laser excitation, converted into single plane projection and analyzed using ImageJ software. Scanning Electron Microscopy (SEM): stage 37 *Xenopus* embryos were fixed in 3% glutaraldehyde in 0.1 M phosphate buffer pH 7.4 (19 mL monosodium phosphate 0.2 M and 81 mL disodium phosphate 0.2 M) made with filtered (0.22 µm) bi-distilled water, during 4 h with vigorous agitation, then washed with phosphatase buffer and filtered bi-distilled water, to be successively dehydrated in ethanol at 25, 50, and 70% for 30 min each; then, embryos were stored in fresh ethanol 70% at 4 °C for 1–2 days before further processing. Embryos in 70% ethanol were further dehydrated with vigorous agitation in ethanol once at 90% and twice at 100% for 30 min each; they were subsequently subjected to CO<sub>2</sub> critical point drying (CPD030, Balzers) at 31 °C and 73 atm. Finally, samples were sputter-coated with gold (vacuum 1 × 10<sup>-12</sup> Torr, beam energy 3–4 keV) for immediate SEM digital imaging (FEI TENE0) of the skin epidermis. Transmission Electron Microscopy (TEM): stage 25 *Xenopus* embryos were fixed overnight at 4 °C in 2.5% glutaraldehyde, 2% paraformaldehyde, 0.1% tannic acid in a sodium cacodylate buffer 0.05 M pH 7.3. Next, embryos were washed 3 × 15 min in cacodylate 0.05 M at 4 °C. Post-fixation was done in 1% osmium buffer for 2 h. Next, embryos were washed in buffer for 15 min. Then, embryos were washed in water and dehydrated conventionally with alcohol, followed by a step in 70% alcohol containing 2% uranyl during 1 to 2 h at RT, or

overnight at 4 °C. Following three incubations in 100% alcohol, completed with three washes of acetone, embryos were included in classical epon resin, which was polymerized in oven at 60 °C for 48 h. Sections of 80 nm were made and analyzed into an FMI TECNAI microscope with acceleration of 200 kV.

**Statistical analysis of *Xenopus* experiments.** To quantify the effect of our different experiments, we applied one-way ANOVA analysis and Bonferroni's multiple comparisons test (*t* test). \*\*\**p* < 0.05; ns = not significant. Statistical analyses were done using GraphPad Prism 6.

Figure 5o and Fig. S6k: 10 cells per condition were analyzed and the total number of Centrin-YFP or  $\gamma$ -tubulin-positive spots per injected cell was counted.

Figure 7g: 5 fields ( $\times 20$  zoom) per condition were analyzed, and the total number of properly ciliated MCCs based on acetylated  $\alpha$ -tubulin staining among GFP positive cells per field was counted. Each field corresponded to a different embryo.

Figure 5s: 160–200 cells per condition were analyzed. *n* = 6, 8, and 10 embryos from three independent experiments for control, Mo ATG and Mo Spl, respectively. Unpaired *t* test vs control: *p* = 0.0037 (Mo ATG \*\*) and 0.0004 (Mo Spl \*\*\*).

## Data availability

scRNA-seq data were submitted to the GEO portal under series reference [GSE103518](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103518). Proteomics data are available via ProteomeXchange with identifier [PXD010629](https://www.ebi.ac.uk/psd/entry/PXD010629). All other relevant data are available from the authors.

Received: 6 November 2017 Accepted: 6 August 2018

Published online: 07 November 2018

## References

- Meunier, A. & Azimzadeh, J. Multiciliated cells in animals. *Cold Spring Harb. Perspect. Biol.* **8**, a028233 (2016).
- Spassky, N. & Meunier, A. The development and functions of multiciliated epithelia. *Nat. Rev. Mol. Cell Biol.* **18**, 423–436 (2017).
- Brooks, E. R. & Wallingford, J. B. Multiciliated cells. *Curr. Biol.* **24**, R973–R982 (2014).
- Kyrousi, C. et al. Mcdas and GemC1 are key regulators for the generation of multiciliated ependymal cells in the adult neurogenic niche. *Development* **142**, 3661–3674 (2015).
- Arbi, M. et al. GemC1 controls multiciliogenesis in the airway epithelium. *EMBO Rep.* **17**, 400–413 (2016).
- Terre, B. et al. GEMC1 is a critical regulator of multiciliated cell differentiation. *EMBO J.* **35**, 942–960 (2016).
- Zhou, F. et al. Gmnc is a master regulator of the multiciliated cell differentiation program. *Curr. Biol.* **25**, 3267–3273 (2015).
- Stubbs, J. L., Vladar, E. K., Axelrod, J. D. & Kintner, C. Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation. *Nat. Cell Biol.* **14**, 140–147 (2012).
- Ma, L., Quigley, I., Omran, H. & Kintner, C. Multicilin drives centriole biogenesis via E2f proteins. *Genes Dev.* **28**, 1461–1471 (2014).
- Quigley, I. K. & Kintner, C. Rfx2 stabilizes Foxj1 binding at chromatin loops to enable multiciliated cell gene expression. *PLoS Genet.* **13**, e1006538 (2017).
- Chung, M. I. et al. RFX2 is broadly required for ciliogenesis during vertebrate development. *Dev. Biol.* **363**, 155–165 (2012).
- Boon, M. et al. MCIDAS mutations result in a mucociliary clearance disorder with reduced generation of multiple motile cilia. *Nat. Commun.* **5**, 4418 (2014).
- Wallmeier, J. et al. Mutations in CCNO result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia. *Nat. Genet.* **46**, 646–651 (2014).
- Sorokin, S. P. Reconstructions of centriole formation and ciliogenesis in mammalian lungs. *J. Cell. Sci.* **3**, 207–230 (1968).
- Dirksen, E. R. Centriole morphogenesis in developing ciliated epithelium of the mouse oviduct. *J. Cell. Biol.* **51**, 286–302 (1971).
- Anderson, R. G. & Brenner, R. M. The formation of basal bodies (centrioles) in the Rhesus monkey oviduct. *J. Cell. Biol.* **50**, 10–34 (1971).
- Kalnings, V. I. & Porter, K. R. Centriole replication during ciliogenesis in the chick tracheal epithelium. *Z. Zellforsch. Mikrosk. Anat.* **100**, 1–30 (1969).
- Steinman, R. M. An electron microscopic study of ciliogenesis in developing epidermis and trachea in the embryo of *Xenopus laevis*. *Am. J. Anat.* **122**, 19–55 (1968).
- Al Jord, A. et al. Centriole amplification by mother and daughter centrioles differs in multiciliated cells. *Nature* **516**, 104–107 (2014).
- Klos Dehring, D. A. et al. Deuterosome-mediated centriole biogenesis. *Dev. Cell.* **27**, 103–112 (2013).
- Zhao, H. et al. The Cep63 paralogue Deup1 enables massive de novo centriole biogenesis for vertebrate multiciliogenesis. *Nat. Cell Biol.* **15**, 1434–1444 (2013).
- Sir, J. H. et al. A primary microcephaly protein complex forms a ring around parental centrioles. *Nat. Genet.* **43**, 1147–1153 (2011).
- Marcet, B. et al. Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nat. Cell Biol.* **13**, 693–699 (2011).
- Tan, F. E. et al. Myb promotes centriole amplification and later steps of the multiciliogenesis program. *Development* **140**, 4277–4286 (2013).
- Yu, H. Cdc20: a WD40 activator for a cell cycle degradation machine. *Mol. Cell* **27**, 3–16 (2007).
- Song, R. et al. miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature* **510**, 115–120 (2014).
- Otto, T. et al. Cell cycle-targeting microRNAs promote differentiation by enforcing cell-cycle exit. *Proc. Natl Acad. Sci. USA* **114**, 10660–10665 (2017).
- Wu, J. et al. Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. *Proc. Natl Acad. Sci. USA* **111**, E2851–E2857 (2014).
- Chevalier, B. et al. miR-34/449 control apical actin network formation during multiciliogenesis through small GTPase pathways. *Nat. Commun.* **6**, 8386 (2015).
- Mercey, O. et al. Characterizing isomiR variants within the microRNA-34/449 family. *FEBS Lett.* **591**, 693–705 (2017).
- Firat-Karalar, E. N., Rauniyar, N., Yates, J. R. 3rd & Stearns, T. Proximity interactions among centrosome components identify regulators of centriole duplication. *Curr. Biol.* **24**, 664–670 (2014).
- Robu, M. E. et al. p53 activation by knockdown technologies. *PLoS Genet.* **3**, e78 (2007).
- Werner, M. E. et al. Radial intercalation is regulated by the Par complex and the microtubule-stabilizing protein CLAMP/Spel1. *J. Cell. Biol.* **206**, 367–376 (2014).
- Firat-Karalar, E. N. & Stearns, T. The centriole duplication cycle. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369**, 20130460 (2014).
- Tsuo, M. F. et al. Polo kinase and separase regulate the mitotic licensing of centriole duplication in human cells. *Dev. Cell* **17**, 344–354 (2009).
- Kim, J., Lee, K. & Rhee, K. PLK1 regulation of PCNT cleavage ensures fidelity of centriole separation during mitotic exit. *Nat. Commun.* **6**, 10076 (2015).
- Matsuo, K. et al. Kendrin is a novel substrate for separase involved in the licensing of centriole duplication. *Curr. Biol.* **22**, 915–921 (2012).
- Al Jord, A. et al. Calibrated mitotic oscillator drives motile ciliogenesis. *Science* **358**, 803–806 (2017).
- Huttlin, E. L. et al. The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
- Rual, J. F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Thein, K. H., Kleylein-Sohn, J., Nigg, E. A. & Gruneberg, U. Astrin is required for the maintenance of sister chromatid cohesion and centrosome integrity. *J. Cell. Biol.* **178**, 345–354 (2007).
- Chiu, S. C. et al. The mitosis-regulating and protein-protein interaction activities of astrin are controlled by aurora-A-induced phosphorylation. *Am. J. Physiol. Cell Physiol.* **307**, C466–C478 (2014).
- Eguren, M. et al. The APC/C cofactor Cdh1 prevents replicative stress and p53-dependent cell death in neural progenitors. *Nat. Commun.* **4**, 2880 (2013).
- Chung, H. J., Park, J. E., Lee, N. S., Kim, H. & Jang, C. Y. Phosphorylation of astrin regulates its kinetochore function. *J. Biol. Chem.* **291**, 17579–17592 (2016).
- Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
- Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Boutin, C. et al. NeuroD1 induces terminal neuronal differentiation in olfactory neurogenesis. *Proc. Natl Acad. Sci. USA* **107**, 1201–1206 (2010).
- Boutin, C., Diestel, S., Desoeuvre, A., Tiveron, M. C. & Cremer, H. Efficient in vivo electroporation of the post-natal rodent forebrain. *PLoS ONE* **3**, e1883 (2008).

54. Marchal, L., Luxardi, G., Thome, V. & Kodjabachian, L. BMP inhibition initiates neural induction via FGF signaling and *Zic* genes. *Proc. Natl Acad. Sci. USA* **106**, 17437–17442 (2009).
55. Castillo-Briceno, P. & Kodjabachian, L. *Xenopus* embryonic epidermis as a mucociliary cellular ecosystem to assess the effect of sex hormones in a non-reproductive context. *Front. Zool.* **11**, 9 (2014).
56. Deblandre, G. A., Wettstein, D. A., Koyano-Nakagawa, N. & Kintner, C. A two-step mechanism generates the spacing pattern of the ciliated cells in the skin of *Xenopus* embryos. *Development* **126**, 4715–4728 (1999).

## Acknowledgements

We are grateful to Chris Kintner, Marc Kirschner, Olaf Stemmann, Reinhard Köster, Xavier Morin and Xueliang Zhu for reagents. Imaging in IBDM was performed on PiCSL-FBI core facility, supported by the French National Research Agency through the program “Investments for the Future” (France-BioImaging, ANR-10-INBS-04). Sequencing at UCAGENOMIX (IPMC), a partner of the National Infrastructure France Génomique, was supported by Commissariat aux Grands Investissements (ANR-10-INBS-09-03, ANR-10-INBS-09-02) and Canceropôle PACA. The authors thank Florian Roguet for *Xenopus* care, and Nathalie Garin from Leica Microsystems GmbH for technical advice on STED microscopy. We are grateful to Rainer Waldmann, Kévin Lebrigand, Virginie Magnone and Nicolas Nottet for fruitful discussions on single cell RNA sequencing, and Delphine Debayle for help with mass spectrometry experiments. We thank Julien Royet and Harold Cremer for insightful comments on the manuscript. This project was funded by grants from ANR (ANR-11-BSV2-021-02, ANR-13-BSV4-0013, ANR-15-CE13-0003), FRM (DEQ20141231765, DEQ20130326464, DEQ20180339158), Fondation ARC (PJA 20161204865, PJA 20161204542), the labex Signalife (ANR-11-LABX-0028-01), the association Vaincre la Mucoviscidose (RF20140501158, RF20120600738, RF20150501288), and the Chan Zuckerberg Initiative (Silicon Valley Foundation, 2017-175159-5022). OM, CB, and DRR were supported by fellowships from Ligue Nationale contre le Cancer (OM and CB), and Fondation ARC (DRR).

## Author contributions

P.B., B.M., and L.K. designed and supervised the study, and obtained funding. L.E.Z., S.R.G., and O.M. performed and analyzed human and mouse airway cells experiments. D.R.

R. and V.T. performed and analyzed *Xenopus* experiments. C.B. performed and analyzed all experiments on mouse ependymal MCCs and contributed to the description of *Xenopus* deuterosomes. O.R. characterized CDC20B antibodies. M.D. and A.P. performed the bioinformatic analysis. N.P. carried out scRNA-seq experiments. A.S.G. performed mass spectrometry analyses. G.P. designed and performed CDC20B interaction studies. All authors were involved in data interpretation. D.R.R., L.E.Z., and C.B. designed the figures. L.K. drafted the original manuscript. D.R.R., L.E.Z., C.B., B.M., P.B., and L.K. edited the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-06768-z>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

*XIV. Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration*

---

---

## **Title**

**Single-cell RNA sequencing reveals novel cell differentiation dynamics during human airway epithelium regeneration**

## **Authors:**

Sandra Ruiz Garcia<sup>1†</sup>, Marie Deprez<sup>1†</sup>, Kevin Lebrigand<sup>1</sup>, Agnès Paquet<sup>1</sup>, Amélie Cavard<sup>1</sup>, Marie-Jeanne Arguel<sup>1</sup>, Virginie Magnone<sup>1</sup>, Ignacio Caballero<sup>2</sup>, Sylvie Leroy<sup>1,3</sup>, Charles-Hugo Marquette<sup>3</sup>, Brice Marcet<sup>1</sup>, Pascal Barbry<sup>1</sup>, Laure-Emmanuelle Zaragosi<sup>1</sup>

## **Affiliations:**

<sup>1</sup>Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France.

<sup>2</sup>ISP, INRA, Université Tours, 37380, Nouzilly, France.

<sup>3</sup>Université Côte d'Azur, CHU de Nice, Pulmonology Department, Nice, France.

<sup>†</sup> These authors contributed equally

Correspondence to: [barbry@ipmc.cnrs.fr](mailto:barbry@ipmc.cnrs.fr)

## **Abstract**

### **Background:**

It is usually considered that the upper airway epithelium is composed of multiciliated, goblet, secretory and basal cells, which collectively constitute an efficient first line of defense against inhalation of noxious substances. Upon injury, regeneration of this epithelium through proliferation and differentiation can restore a proper mucociliary function. However, in chronic airway diseases, the injured epithelium frequently displays defective repair leading to tissue remodeling, characterized by a loss of multiciliated cells and mucus hyper-secretion. Delineating drivers of differentiation dynamics and cell fate in the human airway epithelium is important to preserve homeostasis.

### **Results:**

We have used single cell transcriptomics to characterize the sequence of cellular and molecular processes taking place during human airway epithelium regeneration. We have characterized airway subpopulations with high resolution and lineage inference algorithms have unraveled cell trajectories from basal to luminal cells, providing markers for specific cell populations, such as deuterosomal cells, i.e. precursors of multiciliated cells. We report that goblet cells, like secretory cells, can act as precursors of multiciliated cells. Our study provides a repertoire of molecules involved in key steps of the regeneration process, either keratins or components of the Notch, Wnt or BMP/TGF $\beta$  signaling pathways. Our findings were confirmed in independent experiments performed on fresh human and pig airway samples, and on mouse tracheal epithelial cells.

### **Conclusions:**

Our single-cell RNA-seq study provides novel insights about airway epithelium differentiation dynamics, clarifies cell trajectories between secretory, goblet and multiciliated cells, identifies novel cell subpopulations, and maps the activation and repression of key signaling pathways.

## **Keywords**

Airway epithelium, regeneration, single-cell RNA-seq, differentiation, multiciliated cells, secretory cells, goblet cells, mucus-secreting cells, basal cells, keratins, pathways, clustering.

## **Background**

In mammalian airways, a pseudo-stratified mucociliary epithelium constitutes an efficient first line of defense of the respiratory tract against a large panel of inhaled substances. This epithelium forms a complex ecosystem, mainly composed by: multiciliated cells (MCCs), projecting hundreds of motile cilia at their apical surface, goblet cells (GCs), secreting protective mucins on the luminal surface, secretory cells (SCs), producing anti-microbial and anti-inflammatory peptides, and basal cells, playing a role in adhesion and stability of the epithelium [1,2]. Altered balance between multiciliated and goblet lineages (i.e. decreased number of MCCs with increased number of GCs) is a hallmark of chronic respiratory diseases, such as chronic obstructive pulmonary disease, primary ciliary dyskinesia, asthma or cystic fibrosis. These diseases, which collectively affect hundreds of millions of people, are characterized by frequent injuries, repair defects, tissue remodeling and altered mucociliary clearance [3–5]. In order to accurately characterize the cellular and molecular identity of progenitor cells of the airway epithelium that contribute to efficient tissue regeneration and of the mechanisms that control the MCC/GC balance, key facts about lung epithelial cell fates have been investigated in mouse, mainly injuring the epithelium with noxious agents. Lineage tracing studies have identified basal cells as the main airway progenitor cells, as they display self-renewal capacities and ability to differentiate into MCCs and SCs [2,6,7]. Basal cells in mouse are abundant in the trachea and in the main bronchi, but they are absent in smaller airways [8]. In human, they populate the whole airways, though showing a 2-fold decrease in numbers in smaller airways, i.e. with diameters below 0.5mm [9]. The TP63 transcription factor, keratins 5 and 14 (KRT5 and KRT14), podoplanin (PDPN), nerve growth factor receptor (NGFR), galectin 1 (LGALS1), integrins  $\alpha 6$  and  $\beta 4$  (ITGA6 and ITGB4), laminins  $\alpha 3$  and  $\beta 3$  (LAMA3 and LAMB3) have been frequently used as selective markers of BCs [2,6]. While a direct differentiation of a subpopulation of BCs into MCCs has been reported upon injury [10], the consensus emerging from lineage tracing studies suggests that BCs usually differentiate first into SCs [11]. SCs, known as club (or Clara) cells, are widespread from mouse trachea to bronchioles, and are also found in the human airways, though at a much less abundance, being absent from the upper airways and enriched in the terminal and respiratory bronchioles [12]. SCs are detected at a luminal position of the epithelium, showing a characteristic columnar shape, being filled with secretory granules containing anti-microbial, anti-inflammatory peptides [13] and contributing to xenobiotic metabolism [14]. The most extensively studied marker of SCs is the anti-inflammatory secretoglobin SCGB1A1. Lineage tracing studies have used SCGB1A1 to determine the fate of SCs and have shown that they give rise to MCCs, identified by the specific expression of the transcription factor FOXJ1 [11,15] and to GCs, identified by the expression of specific mucin MUC5AC [2,16]. Molecular mechanisms regulating cell fate decisions in the airway epithelium lineages have been investigated over the past few years and

they have established the role of Notch signaling during commitment of BCs. It is now clear that Notch activation leads to the SC/GC lineages as Notch inhibition leads to commitment towards the MCC fate [17–20]. In agreement with these findings, our group has shown that Notch pathway inhibition by the miR-34/449 families of microRNAs is required for MCC differentiation [21,22]. Although *in vivo* lineage tracing studies have largely unveiled cell lineage hierarchies in the airway epithelium, they have some inherent limitations. They can only be performed in mouse, after specific, drastic, non-physiological forms of injuries which may not completely reveal physiological tissue turnover. Moreover, they rely on genetic cell labelling, usually *Krt5* for BCs and *Scgb1a1* for SCs, thus, they are not necessarily comprehensive and cannot provide a full picture of cell hierarchies in the airway epithelium. In human, there is no possibility to perform such studies, and cell lineage hierarchies must be inferred indirectly by *in vitro* approaches. Single-cell RNA-sequencing, coupled with computational methods measuring transition between cell states, has emerged as a complementary and powerful approach to reveal lineage hierarchies in many tissues [23–25] and even in a whole organism, by capturing undifferentiated, intermediate and terminally differentiated states [26]. In the lung, one pioneer single-cell study on 198 cells has delineated lineage hierarchies of alveolar cells in the mouse distal lung [27]. Recent studies have established first atlases of the airways in mouse [28] and human [29,30]. However, extensive characterization of cell population diversity and cell lineages in the human airways remains to be achieved. Here, we have performed and collected single-cell RNA-seq data from fresh human airway epithelial tissues and all along an experiment of 3D *in vitro* regeneration of human airway epithelium. Our work has resulted in a comprehensive cell trajectory roadmap of human airways, which identifies novel cell populations and offer new insights into molecular mechanisms taking place during the mucociliary epithelium regeneration.

## **Results**

### ***Reconstruction of cell lineage in regenerating airway epithelium by single-cell RNA-seq***

To identify trajectories of cells composing the human airway epithelium, we have analyzed single-cell transcriptomes obtained at successive stages during *in vitro* 3D regeneration of this tissue (Fig. 1a, b). We first validated that this *in vitro* model faithfully recapitulated native airway tissues by comparing cell population compositions of 3D regenerated airway tissues (Human Airway Epithelial Cells, HAECs) with that of fresh human airway tissues. We performed single-cell RNA-seq of epithelial cells dissociated from nasal brushing samples and from nasal turbinates which were resected during surgery, as well as single-cell RNA-seq of HAECs at late time point of *in vitro* air-liquid interface differentiation (3D cells). Our main results were obtained with HAECs that were cultured in a Pneumacult media (StemCell Technologies), which allows the production of multiciliated and goblet

cells, but additional experiments were also performed with HAECs cultured in BEGM (Lonza), which rather favors the production of multiciliated cells. Cell identity was inferred from the expression of specific marker genes, such as *KRT5* and *TP63* for basal cells (BCs), *SCGB1A1* for secretory cells (SCs), *MUC5AC* for goblet cells (GCs) and *FOXJ1* for multiciliated cells (MCCs). These cell types were robustly found in all samples at various proportions, validating the use of these *in vitro* models to trace airway cell lineages (Supplemental Figure S1a-c). We also confirmed that cell type proportions inferred from single-cell RNA-seq were correlated with cell type proportions inferred from protein measurements by performing immunostaining of selected population markers (Supplemental Figure S1d, e). In addition, we have evaluated the effect of cell dissociation on gene expression (Supplemental Figure S2a). We found that cell dissociation did not produce a major impact on gene expression with the exception of *FOS* and *FOSB* which were highly upregulated (Supplemental Figure S2b, c). Molecular function enrichment with Ingenuity Pathway Analysis (Qiagen) showed that “cell death and survival” and “cellular growth and proliferation” were the only molecular functions to be regulated with a p-value inferior to 0.001 (Supplemental Figure S2d).

We have generated single-cell transcriptomes of HAECs cultivated in Pneumacult medium at 3 time points of *in vitro* airway regeneration (ALI7, ALI12, and ALI28) (Fig. 1b). Time points were chosen as most representative of the known steps of airway regeneration: proliferation, polarization and specification [31]. This experiment was complemented by 6 additional time points of HAECs cultivated in BEGM medium (ALI2, ALI4, ALI7, ALI12, ALI17 and ALI22). In a first approach, each time point was analyzed independently. We did 10 random selections of cells, corresponding to subgroups containing 90% of the initial number of cells. The resulting gene expression submatrices were then iteratively clustered (10 times with varying parameters), and a census was applied to define the most robust cell types. We then studied the variations of these populations during the entire time course. Cells clustered in 6 main populations in the Pneumacult condition: (1) cycling (*MKI67+*) and (2) non-cycling (*MKI67-*) BCs (*KRT5+/TP63+*), (3) supraBCs (*KRT5+/ TP63- /KRT13+/KRT4+*), (4) SCs (*SCGB1A1+*), (5) GCs (*MUC5AC+*) and (6) MCCs (*FOXJ1+*) (Figure 1c). Cell population proportions evolved during the time course, with a global reduction of BCs and SCs, a first detection of supraBCs at ALI7, followed by an increase of the proportion of this cell population at ALI28 and a first detection of GCs and MCCs at ALI28 (Fig. 1d). In the BEGM condition, cells clustered in 7 cell populations (Supplemental Figure S3a, b). We did not detect SCs and GCs with this culture condition, but we found instead a cell population that we termed “Secretory-like cells”, given their high gene expression similarity with SCs, except for *SCGB1A1*, which was not detected (Supplemental Figure S4). Additional cell types were found in these samples: *KRT5-* supraBCs (*TP63-/KRT13+/KRT4+*) and 2 cell populations that we termed as “undefined intermediates 1” and “undefined intermediates 2” because their gene expression profile did not allow

unambiguous classification. Differential gene expression showed that these undefined intermediates 1 and 2 expressed specific genes, such as *SPINK1* and *SPINK7*, respectively. A representation of all cells analyzed in each culture was generated using t-SNE graphs on the aggregate of all time points for each medium condition (Fig. 1e for Pneumacult and Supplemental Figure S3c for BEGM). Cell trajectories and transitions from one cell population to another were deduced from a trajectory inference analysis using Monocle 2, followed by differential expression analysis between consecutive cell states in pseudotime using Seurat. In the BEGM condition, a unique cell trajectory was found (Supplemental Figure S3d), starting with cycling and non-cycling BCs at its beginning, followed by *KRT5*<sup>+</sup> and then *KRT5*<sup>-</sup> supraBCs cells, and with MCCs at its end. Before reaching the secretory-like cell state, the trajectory goes through the 2 “undefined intermediate” cell states. Despite the absence of *SCGB1A1* expression in secretory-like cells (*SCGB1A1*<sup>-</sup>/*BPIFA1*<sup>+</sup>/*KRT8*<sup>+</sup>), these cells were ordered in the pseudotime before MCCs, as expected for canonical SCs (Supplemental Figure S3d-f). A more complex trajectory was observed with the Pneumacult condition, in which Monocle2 detected a bifurcation into 2 distinct branches after the SC stage: the larger branch leading to *FOXJ1*<sup>+</sup> MCCs, and the smaller one leading to *MUC5AC*<sup>+</sup> GCs (Fig. 1f, g). A closer examination of the pseudotime ordering and differential gene expression (Fig. 1h) revealed that few *MUC5AC*<sup>+</sup> cells were found on the MCC branch, after the GC bifurcation and that some *FOXJ1*<sup>+</sup> cells retained expression of *MUC5AC*. Altogether, our findings confirm SCs as precursors of both MCCs and GCs. They also suggest that GCs can also act as MCC precursors in airway epithelial regeneration.

### ***Goblet cells can act as differentiation intermediates for multiciliated cells***

We further investigated the possibility that some GCs may correspond in fact to precursor cells for MCCs. A first evidence is coming from our robust clustering analyses, either when they were performed on cells coming from *in vitro* samples or from fresh tissues. The two populations of GCs and SCs displayed very similar gene expression profiles, and were discriminated based on their levels of expression of *MUC5AC* and *MUC5B* which were higher in GCs (Supplemental Figure S1a-c). In the Pneumacult experiment, 24 out of the 54 top genes for GCs were also associated with SCs (Fig. 2a). Among these transcripts was *SCGB1A1*, while expression of *MUC5AC* and *MUC5B* was more robust in GCs (Fig. 2b). When we directly assessed differential gene expression between cells located at the two ends of the GC branch (Fig. 1f, Fig. 2c), we confirmed the high similarity of gene expression existing between SCs and GCs. GCs differed from SCs by higher levels of expression for SC/GC specific genes. This was observed for molecules such as mucins (*MUC1*, *MUC4*, *MUC5B*, *MUC5AC*), secretoglobins (*SCGB1B1* and *SCGB3A1*), PLUNC antimicrobial factors (*BPFA1* and *BPIFB1*) and *SLPI*, the secretory

leukocyte protease inhibitor (Fig. 2c). These properties led us to consider GCs as ‘hyperactive’ SCs and led to the prediction that these cells could also function as MCC precursors. This point was tested by quantifying the cellular expression of *MUC5AC* and *FOXJ1*, and by measuring the percentage of doubly labeled cells. Our rationale was that a detection of cells that express at the same time *MUC5AC* and *FOXJ1* would suggest the existence of a transitory state between GCs and MCCs. Fig. 2d, 2g and 2j indeed show that 8.9% of GCs + MCCs express at the same time *MUC5AC* and *FOXJ1*. It also shows the existence of SCs/MCCs expressing both *SCGB1A1* and *FOXJ1*, which correspond to a more conventional type of precursors for MCCs (Fig. 2m). The presence of *MUC5AC*<sup>+</sup>/*FOXJ1*<sup>+</sup> and *SCGB1A1*<sup>+</sup>/*FOXJ1*<sup>+</sup> cells was not restricted to a cell culture model, as these transitional cells were also detected in fresh biopsies from human bronchi (Fig. 2e, h, k, n) and pig trachea (Fig. 2f, i, l, o). The presence of doubly-labeled cells was confirmed by qPCR in a fully independent dataset, derived from a HAEC culture (Supplemental Figure S5). In that other experiment, we isolated the cells with the C1 technology (Fluidigm) and quantified gene expression by quantitative RT-PCR with a Biomark (Fluidigm). Cells isolated with the C1 were visually inspected, and these experimental settings ensured the absence of cell doublets. We found that 4 cells out of 74 expressed GC specific genes (namely *MUC5AC*, *MUC5B* and *TFF3*), together with MCC specific genes (*FOXJ1*), and more specifically, immature MCC genes (*PLK4*, *MYB* and *CDC20B*) [32] (Supplemental Figure S5a, b). The result was further confirmed after reanalyzing a recent dataset published by others [29] (Supplemental Figure S5c). A further confirmation came from the detection at a protein level of cells that were labeled at the same time by *MUC5AC* and acetylated Tubulin, a specific protein marker of the cilia (Fig. 2p). A final point came after a survey of our data with “RNA velocity” [33]. Velocity can predict the fate of individual cells at a timescale of hours by distinguishing the expression of spliced and unspliced forms of transcripts. We used Velocity to analyze the behavior of three transcripts: *CEP41*, a specific marker of cells in an early phase of multiciliated differentiation, *SCGB1A1* and *MUC5B*. RNA velocity algorithm calculates a residual value of each gene, which indicates expected upregulation when it is positive and expected downregulation when it is negative. Positive residuals were found for transcripts of *CEP41* in the GC population, predicting an upregulation of this gene’s expression in the next hours. A different picture was observed for the transcripts of *SCGB1A1* and *MUC5B*, in which negative residuals were found in the GC and SC populations, indicating an expected downregulation of the corresponding transcripts in the next hours (Fig. 2q). Altogether, these data indicate that GCs can act as precursors for MCCs in normal *in vitro* and in homeostatic *in vivo* airway regeneration.

***Refining cell clustering identifies 6 additional clusters, including a discrete population of pre-MCC “deuterosomal” cells***

To gain further insight in the diversity of cell populations composing the airway epithelium and the transitional ones occurring during the regeneration, we considered additional clusters that could be derived from our sub-clustering analysis, by accepting less discriminations between them than between the initial 6 previously identified clusters. This deeper analysis defined 12 clusters, rather than the 6 initial ones (Fig. 3a, Supplemental Figure S6a). The non-cycling BC population was split into 2 clusters that we termed BC1 and BC2. The major difference between these 2 clusters was the higher level of expression for genes associated with cell migration: *FN1*, *VIM*, *SPARC* and *TAGLN* in the BC2 cluster. Analysis of enriched canonical pathways with Ingenuity Pathway Analysis showed enrichment for integrin, actin cytoskeleton and Rho GTPase signaling as well as the pathway “regulation of actin-based motility” in BC2 compared to BC1, suggesting an increased migratory activity in BC2 (Supplemental Figure S7). Since BC2 cells were associated with higher pseudotime values than BC1, BC2 likely correspond to the population of most differentiated BCs, just before reaching a suprabasal state. The *supraBC* and SC populations could also be further split into 3 distinct sub-populations, resulting in a total of 3 new populations of *supraBC* and 3 new populations of SC (Fig. 3a and Supplemental Figure S6a). Each of them displayed its own distinct gene set enrichment (Supplemental Figure S7). Among them, the SC2 subpopulation appeared particularly interesting, since it displayed a strong enrichment score for the feature “immune cell migration, invasion and chemotaxis”, and also a strong positive enrichment for canonical pathways such as “neuroinflammation signaling” and “dendritic cell maturation”. This was explained by the increased gene expression of targets for pro-inflammatory molecules such as TNF, IFNG, NFkB, IL1A/B, IL2, or IL6, as well as decreased gene expression for targets for the anti-inflammatory PPARG pathway (Supplemental Figure S7). This may confer to this subpopulation of SCs a unique relationship with the immune response. The MCC cluster, containing *FOXI1*+ cells, was further split in 2 discrete clusters: (1) the largest one is positive for mature MCC genes such as *DNAH5*, and corresponds to terminally differentiated MCCs; (2) the second one specifically expresses several molecules that are important for the biosynthesis of hundreds of motile cilia. Among them is *DEUP1*, a hallmark of massive centriole amplification at deuterosomes (Fig. 3b). This led us to coin these cells “deuterosomal” cells. This subpopulation is clearly distinct from mature MCCs in our clustering analysis (Fig. 3b) and expresses also highly specific markers such as *PLK4*, *CCNO*, and *CEP78* (Supplemental Figure S8). We have confirmed the existence of deuterosomal cells in three distinct species with single cell experiments performed on mouse tracheal epithelial cells (MTECs) dissociated at ALI3 (i.e. the time point of maximal centriole amplification in this model), newborn pig trachea and human bronchial biopsy from a healthy adult subject (Fig. 3c). In all samples, even under homeostatic conditions, we noticed the presence of deuterosomal cells that clustered independently of mature MCCs. This deuterosomal cell population was characterized by the expression of gene markers, either uniquely expressed in deuterosomal cells, or also expressed in MCCs and cycling BCs

(Fig. 3d). Our analysis shows that the group of deuterosomal cells specifically expressed 149 genes, while sharing 33 and 244 genes with cycling BCs and mature MCCs, respectively (Fig. 3e). Among the 33 genes in common with cycling BCs, we noticed the re-expression of several cell cycle-related genes, which are required for the massive amplification of centrioles that takes place [34,35]. The most specific genes are displayed in Fig. 3d. This analysis not only confirmed the known expression of *CDK1* in deuterosomal cells [35], as it also highlights the expression in deuterosomal cells of genes coding for centromere proteins (*CENPF*, *CENPU*, *CENPW*), securin (*PTTG1*), a core subunit of the condensin complex (*SMC4*) and cyclin-dependent kinases regulatory subunits (*CKS1B*, *CKS2*). We confirmed the deuterosomal-specific expression of *CDC20B*, the miR-449 host gene. We have recently shown that *CDC20B* is a key regulator of centriole amplification by deuterosomes [32]. Incidentally, we noticed the existence of a novel and short isoform of this gene, arising from alternative splicing and results in the inclusion of a novel exon after the location of the miR-449 family (Fig. 3b and Supplemental Figure S9a, b). We found that this short *CDC20B* isoform was also detectable in mouse RNA-seq data (Supplemental Figure S9c). Comparison of transcript abundance in several samples, including the Pneumacult ALI28 and the human bronchial biopsy showed higher levels for short *CDC20B* compared to *CDC20B* (Supplemental Figure S9d). This short *CDC20B* likely corresponds to the major source of miR-449 in deuterosomal cells. A list of novel markers of deuterosomal cells that are specifically expressed in this cell population is provided in Supplemental Table S1. Some of these genes had never been described before in the context of centriole amplification, such as the Yippee-like factor *YPEL1* or the Notch pathway related hairy-enhancer-of-split family of transcription factors *HES6* (Supplemental Figure S8). Gene set enrichment of the deuterosomal population specific genes (Fig. 3f) shows an enrichment for terms such as cilium assembly, centrosome maturation but also, cell-cycle mechanism-related terms such as resolution of sister chromatid cohesion, regulation of *AURKA* and *PLK1* activity, *CDH1* autodegradation. “Mitochondrial membrane part” was also among the enriched terms, suggesting an increase in mitochondria numbers at this stage. This deuterosome-specific signature perfectly delineates the events that are occurring at this MCC differentiation stage and provides an extensive repertoire of specific cell-cycle related genes which are re-expressed at the deuterosomal stage. This deuterosomal population is a consistent population, much larger than recently described rare cell populations such as ionocytes [28,29], which we also identified (Supplemental Figure S6c).

### ***Establishing a keratin switch pattern during airway regeneration***

An extremely diverse repertoire of keratins has been associated with different types of epithelial cells. Considering that expression of specific keratins varies with cell type, period of embryonic

development, stage of histologic differentiation, cellular growth environment, and disease state, we reasoned that it would be useful to compare our scRNA-seq data with the expression of different keratins. *KRT5* and *KRT14*, are largely used as BC markers in the airway and lung epithelia, and also in other epithelia such as bladder [36], prostate [37] or mammary gland [38]. *KRT8* is clearly associated with luminal cell types [6]. Besides, there is no clear repertoire of the other keratins that are associated with airway cell types. A recent study performed on mouse and human models of *in vitro* regeneration identified *KRT4* and *KRT13* in a subpopulation reminiscent of our supraBCs, as it emerges between BCs and SCs [29]. We have established an extensive single-cell repertoire of KRT expression during airway regeneration based on pseudotime ordering in our Pneumacult ALI28 and pig trachea datasets. Our analysis confirmed in both datasets the presence of *KRT5* and *KRT14* in BCs, of *KRT4* and *KRT13* in supraBCs, and the expression of *KRT8* in luminal cell types (SCs, GCs and MCCs) (Fig. 4a and e). We consistently noticed that expression profiles of *KRT13* and *KRT4* did not completely overlap, with *KRT13* detected at earlier pseudotimes than *KRT4*. This was confirmed at a protein level by a quantification by immunostainings of the proportion of *KRT5*+/*KRT13*+ and *KRT5*+/*KRT4*+ double positive cells (Fig. 4b). Fig. 4c shows that there was more *KRT5*+/*KRT13*+ (7.4%) than *KRT5*+/*KRT4*+ (4.9%) double positive cells, which is consistent with an earlier expression of *KRT13* compared to *KRT4*. In the pig trachea, we also found a very clear shift, with 16.8% and 11.2% of *KRT5*+/*KRT13*+ and *KRT5*+/*KRT4*+ double positive cells, respectively (Fig. 4d and Supplemental Figure S10). Our results show that *KRT4* and *KRT13* are not strictly expressed at the same time during airway regeneration and their expression delineate subtle differences in cell subpopulations. Additional keratins such as *KRT16* and *KRT23* displayed a specific supraBC expression (Fig. 4e). We also identified additional keratins that were more specifically associated to more differentiated cell types: *KRT7* and *KRT18* were strongly enriched in SCs, but their expression completely dropped in MCCs, while *KRT8* was still expressed (Fig. 4e). Expression patterns for these cell type specific keratins were confirmed in the pig trachea (Supplemental Figure S10c). The keratin switch pattern is indeed sufficiently specific to allow a reconstruction of the cell trajectories during airway regeneration.

### ***Establishing a combinatorial repertoire of signaling pathways during airway regeneration***

We have finally analyzed the cell specificity of expression of important signaling pathways, to catch mutual influences between distinct cells that could play a role in airway regeneration. Our investigation was focused on the Notch, BMP/TGF $\beta$  and Wnt pathways. For each different component, we classified them as ligands, receptors, or targets. The expression profiles are shown as heatmaps, cells being sorted by their subgroups, then by ascending pseudotime value (Fig. 5a-c).

Notch pathway (Fig. 5a). BCs express the ligands *DLL1*, *JAG1* and *JAG2*, as well as the receptor *NOTCH1*, as expected [6,29]. In this population, no target gene expression is detected, suggesting that the pathway is inactive. Interestingly, BCs also express *LFNG*, which has been described to inhibit *JAG1* signaling via *NOTCH1* [39]. SupraBCs cells express *NOTCH1*, *JAG1* and *JAG2* but, as opposed to BCs, they show clear activation of the Notch pathway by expression of the target genes *HEY1*, *HES2* and *HES4*. *NOTCH3* expression is turned on and is specific to this population. In SCs/GCs, *NOTCH2* is the major receptor to be detected and signal activation remains as evidenced by the expression of *HEY1* and *HES4*. SCs/GCs also express the non-canonical Notch ligand *NTN1*. In Deuterosomal cells/MCCs, a clear shift is observed. *NOTCH2*, *NOTCH3*, *HEY1* and *HES4* are turned down, and *NOTCH4* is specifically expressed. As previously described, *JAG2* [29], which is present in BCs, then absent in supraBCs and SCs/BCs, is re-expressed in the MCC compartment. We have found the same behavior for *DLL1* and the non-canonical ligand *DNER1*. Thus, MCC express some Notch ligands. Strikingly, a major inhibitory signature dominates in MCCs, with the expression of *CIR1* and *SAP30*, which are transcriptional co-repressors, as well as of *DYRK1A*, an inhibitor of NICD. *HES6*, which expression is not regulated by Notch signaling but which has been identified as a Notch pathway inhibitor through its binding to *HES1* [40], displays an expression pattern that is highly enriched in deuterosomal cells (Fig. 5a and Fig. 3d).

Wnt pathway (Fig. 5b). The Wnt target genes *SNAI2* and *TCF4*, indicators of an active pathway are mainly enriched in the BC population, especially in BC2 for *SNAI2*. In the BC population, *WNT10A* and *LRP1* are strongly enriched, and several *SOX* family members (*SOX2* and *SOX21*) are underrepresented, especially in the cycling BCs, suggesting an activation of the pathway in this compartment. In the MCC population, the situation is more complex. Despite a slight expression of *TCF4* together with positive regulators of the pathway such as *WNT9A*, *FZD6*, *APPL2*, *CSNK1G1* (a casein kinase component that can act as an activator or inhibitor of the pathway [41]), no *SNAI2* expression is detected and known repressors of the Wnt pathway are also overrepresented. Indeed, MCCs express significant levels of the transcriptional repressors *SOX2*, *SOX21* and display strong enrichments for the Reptin components *RUVBL1* and *RUVBL2* (Fig. 5b).

BMP/TGF $\beta$  pathway (Fig. 5c). BMP ligands, such as *BMP2*, and *BMP7*, are both enriched in the BC population, while *BMP3* and *BMP4* are both enriched in the SC/GC populations. A different picture was found for BMP receptors, for which we did not find specific cell population expression. The most striking observation is the specific expression in BCs of BMP inhibitors such as the inhibitory ligand *FST* and the intracellular inhibitor *FKBP1A* (also known as *FKBP12*). As BMP signaling is considered as a brake for proliferation, inhibition in the BC compartment is consistent with the maintenance of a proliferative potential of this progenitor population. Regarding the TGF $\beta$  pathway, a clear signal of activation is detected in the deuterosomal/MCC population, with specific expression of the target

genes *SERPINE1* (PAI-1), *CTGF*, *ATF3*, *TGFBR3* and *IRF7*, consistent with the previous finding that TGF $\beta$  pathway regulates motile cilia length by affecting the transition zone of the cilium [42]. We did not detect TGF $\beta$  ligands in the MCC population but rather found them expressed in BCs (TGFB1) and supraBCs (TGFB3).

We have confirmed the main distribution of the three pathway components sample differentiated with the BEGM medium (Supplemental Figure S11) and in 2 fresh tissue samples (human bronchial biopsy and nasal turbinate) for which a selection of genes is shown in Fig. 5d. Collectively, our data provide for the first time a detailed repertoire of signaling pathways at work during airway regeneration, with receptors and ligands specifically expressed at each cell stage.

## **Discussion**

In this study, we have established a comprehensive single-cell atlas throughout the entire time course of human airway regeneration *in vitro*. We carefully chose representative time points in the process and we quantified the proportion and identity of each cell population at different time points after the establishment of the air liquid interface. We provide here for the first time a comparison of the most widely used culture media in 3D culture of airway epithelial cells: BEGM, the most established commercial medium with which the majority of studies have been performed, and a more recent commercial medium, Pneumacult. In the BEGM medium, we have performed analyses at earlier time points, i.e. ALI2 and ALI4. These time points allowed us to measure the extent of cell proliferation during *in vitro* regeneration. Cycling basal cells account for approximately 40% of total cells at ALI2 and ALI4, and this number drops to 5% at ALI7. These early time points also showed that suprabasal cells appeared early in these conditions, as they were already detected at ALI4. With the BEGM medium, unlike with the Pneumacult medium, we never detected any goblet (*MUC5AC*<sup>+</sup>) nor “canonical” secretory (secretoglobin<sup>+</sup>) cell, even after long periods of time, and on several dozens of cultures coming from distinct donors (Supplemental Figure S1, S3, and personal observations). However, we found a cell population that we have termed “secretory-like”. “Secretory-like” cells express a gene pattern very similar to that of canonical secretory cells, and they can differentiate into MCCs. Interestingly, goblet cells were detected in BEGM medium after an IL-13 treatment [41, and our personal observations]. Future work should investigate whether secretory-like cells first evolve into canonical secretory cells and then goblet cells upon IL-13 treatment.

In the Pneumacult medium, but also in freshly dissociated human bronchial biopsy and pig trachea, we have detected cells expressing both *MUC5AC* and *FOXJ1*. We believe this finding is in line with our lineage inference and RNA velocity data showing that goblet cell can be precursors for MCCs. Other groups had previously detected cells expressing both these markers, in the context of GC

hyper/metaplasia induced by Sendai virus infection or after IL-13 treatment [44–46]. These findings led them claim that MCC can transdifferentiate into GCs. However, none of their data clearly showed a difference in the number of these double-positive cells between control and treated conditions. Our present finding probably provide a different answer to this conundrum, saying that these double-positive cells do exist in the absence of IL-13 stimulation. We noticed that their expression profiles posit them more naturally as precursors of MCCs than as trans-differentiated multiciliated cells. Turner and colleagues [45] proposed that MCCs could trans-differentiate into GCs after performing *in vitro* lentiviral transduction of HAECs with a vector containing a Cre recombinase under the control of the *FOXJ1* promoter. These findings were not confirmed by Rajagopal's group who showed no GC arising from MCC in a context of OVA-induced mucous metaplasia in mouse airways, using *in vivo* lineage tracing with *Foxj1-cre* mice [47].

Our datasets are not limited to HAECs, but also provide useful information about mouse, human, and pig tissues. They correspond to *in vitro* MTECs, as well as cells freshly isolated from human nasal brushing, turbinate, bronchial biopsy and finally pig trachea. We have used this variety of samples to confirm our findings with HAECs, and to invalidate the possibility of culture artefacts. Our study was not focused on rare cell types such as pulmonary neuroendocrine, brush cells or ionocytes, which have been recently described elsewhere. We have indeed detected cells, which displayed the gene signature of ionocytes, characterized by expression of *CFTR* and other ion transporters, as well as specific transcription factors such as *ASCL3* or *FOXJ1* [28,29]. Our investigation was rather focused on the diversity of the main cell types that compose the epithelium. We have identified 3 subtypes of basal cells among which cycling basal cells, and a group of basal cells expressing significantly higher levels of genes involved in extracellular matrix connection and actin-based motility. As these cells are the latest basal cells in our pseudotime, they likely represent an important transient migratory state on the way to the suprabasal state.

An additional peculiarity found within the secretory compartment comes from one of our three secretory subpopulations which displayed an immune-related gene signature. So far, diversity within the secretory compartment was established after expression of different members of the secretoblogin family [48] or via activation level of the Notch pathway [49]. We propose that diversity within the secretory cell compartment should also include specialized functions related to the interaction between the epithelium and immune cells. Additional experiments including protein labeling on fresh tissue sections have now to be performed in order to confirm this diversity and identify the spatial distribution of these subpopulations.

Our study has also provided a first extensive gene signature of the deuterosomal population, which play a key role during differentiation of multiciliated cells. In line with what has been shown recently by our group and others [32,35], cell-cycle related genes become re-expressed in this population on

non-cycling cells. We have confirmed the very specific expression of *CDC20B*, a key player of centriole amplification [32], and have identified, both in human and mouse, a novel isoform of this transcript which displays higher expression than the annotated long isoform. As the pre-mRNA corresponding to this short isoform comprises the *miR-449*-encoding intron, we suggest that this isoform should indeed be the major source of *miR-449* in deuterosomal cells. The alternative splicing that is responsible for this alternative isoform might represent an optimization of gene expression regulation to efficiently increase miR-449 levels.

Besides the identification of each cell type, we have also characterized the distribution of some important signaling pathway components. First, we have focused on the Notch pathway as it is already known as a major regulator of the mucociliary differentiation. We have confirmed the distribution of ligands and receptors described by others [19,20,29,50]. We have also confirmed the absence of Notch activation in BCs and MCCs, with *HES4* being the most representative target gene in our model. BCs express *NOTCH1* and NOTCH ligands. However, no clear Notch pathway activation can be detected within this cell population even in an uneven manner as it could be expected from Notch lateral inhibition. This could be solely the result of weak *NOTCH1* expression but could also be a consequence of specific expression of Notch inhibitors such as the ligand *LFNG* or Casein kinase II subunit beta (*CSNK2B*) [51,52]. At the other end of mucociliary differentiation, inhibition of the Notch pathway has been widely documented in MCCs: failure to inactivate Notch results in failure of MCC signaling differentiation. Here we suggest a novel mechanism for Notch pathway inactivation by showing specific expression of several Notch transcriptional inhibitors at the deuterosomal stage. These include *HES6*, an inhibitory HES acting through HES1 binding [40,53], *DYRK1A*, an inhibitor of Notch Intracellular Domain transcriptional activity [54], as well as *CIR1* and *SAP30* which are transcriptional repressors of Notch/CSL transcriptional complex [55]. As opposed to MCCs and BCs, SCs must undergo clear Notch activation to maintain their cell identity, and differentiate into GCs [18–20]. However, the onset of this signal's activation has not been widely studied. Mori and colleagues have described NOTCH3 expression in TP63-negative cells in a parabasal position of the epithelium, which likely correspond to the cells we, and others, have termed supraBCs [50]. We have confirmed that *NOTCH3* RNA is absent from BCs and becomes upregulated in supraBCs. We went further by showing *HES4* becomes expressed at this cell stage, confirming that Notch pathway activation starts at the supraBC stage. Thus, we emphasize here the importance of this intermediate cell population, though it has not been well characterized so far, for establishing Notch activation and subsequent differentiation.

Contrary to the Notch pathway, the Wnt/ $\beta$ -catenin pathway has not been extensively studied in the context of airway epithelium differentiation. Crosstalk between these 2 pathways has at least been suggested by other studies performed in other systems. For instance, in the hair follicle precortex,  $\beta$ -catenin stimulates Notch signaling by inducing *Jag1* transcription [56]. In the context of airway

epithelium, a very recent study demonstrated that  $\beta$ -catenin signaling is required for the early stages of mucus and ciliated cell differentiation, which they defined as “specification”, but was detrimental to the later “commitment” stages [57]. Wnt also seems to be related to epithelial remodeling upon inflammatory situations: Ordovas-Montanes and colleagues have recently shown that in nasal polyps, which are characterized by an inflammatory state favoring GCs at the expense of MCCs, there is an altered balance between Wnt and Notch signaling, in favor of Wnt signaling [30]. *WNT5A* has been associated with remodeling of airway smooth muscle cells in the context of airway hyperresponsiveness [58]. *WNT4* has been shown to be upregulated in the epithelium of patients suffering from chronic obstructive pulmonary disease and upregulates *IL8* and *CXCL8* gene expression in HAECs [59]. Interestingly, we found both *WNT5A* and *WNT4* specifically expressed by the subpopulation of SCs that may be involved in immune response. This finding again suggests a role for this SC population in the inflammation-induced airway remodeling.

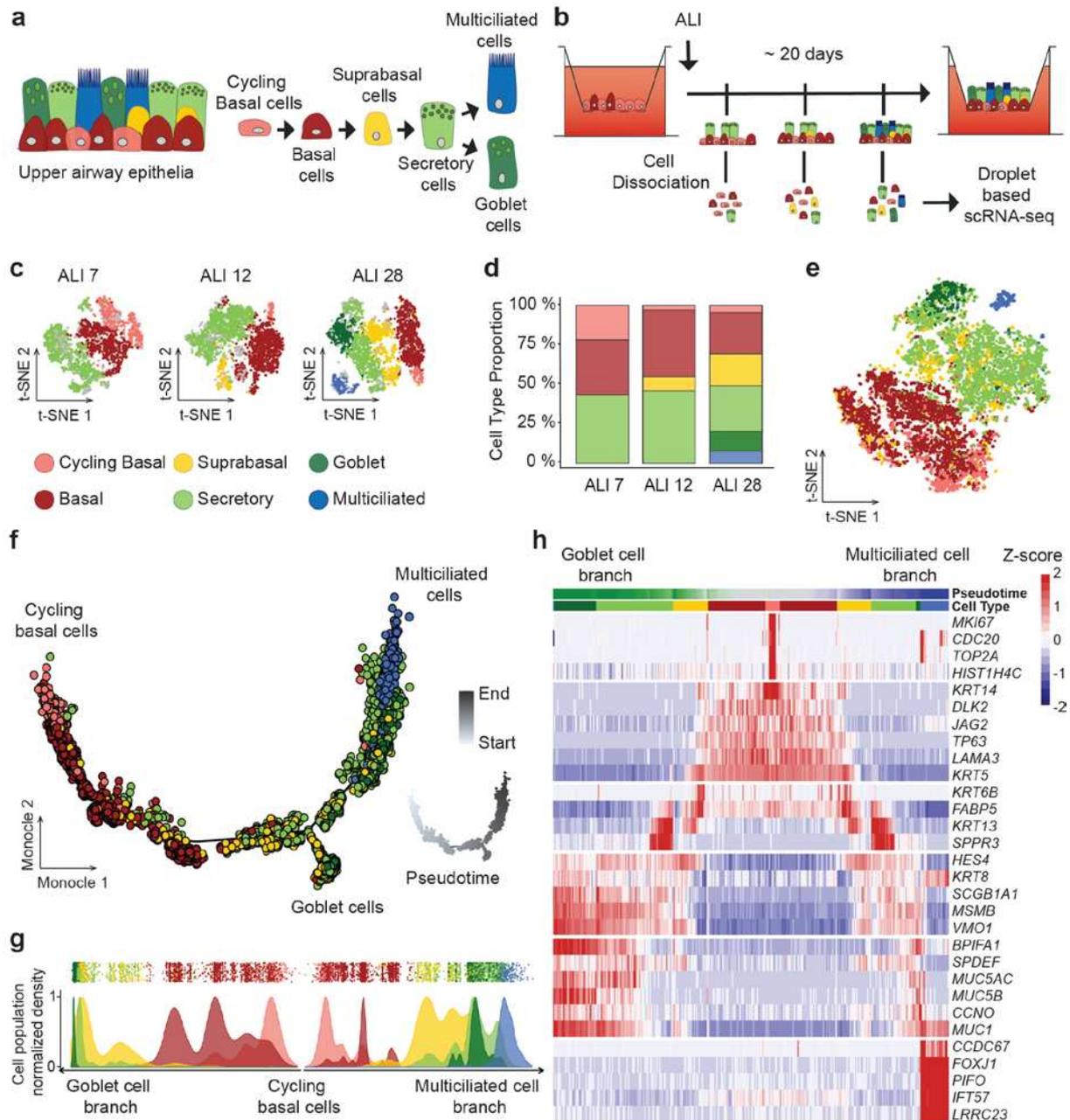
Based on expression of the target genes *TCF4* and *SNAI2*, in our study, activation of the Wnt pathway appears to be confined to the BC population. This population also expresses strongly and specifically the ligand *WNT10A*, suggesting an autocrine regulatory loop. *WNT10A* has been shown to be specific of BC in other epithelia, such as the mammary epithelium [60]. In fallopian organoids, Wnt has been shown to be essential for stemness [61] and for self-renewal, but not proliferation, in basal-like breast cancer cells [62]. Thus, autocrine *WNT10A* signaling may regulate self-renewal in the BC compartment of the airway epithelium. In MCCs, we have observed a specific expression of both members from the Reptin family which are ATP-dependent DNA helicases that have been identified as Wnt signaling repressors [63,64]. As Reynolds and colleagues recently showed that  $\beta$ -catenin was necessary for mucus and ciliated cell specification [57], additional investigations should certainly be carried out to characterize precisely the role of Wnt/ $\beta$ -catenin during airway epithelial regeneration.

## **Conclusions**

Our lineage inference study during the airway epithelium regeneration has provided novel insights in differentiation dynamics by positioning goblet cells as possible precursors for multiciliated cells. Thus, in the airway epithelium, unlike in most tissues, cells carrying specialized function, i.e. secretory and goblet cells, can constitute differentiation intermediates for other specialized cells, the multiciliated cells. We have also identified subpopulations within the basal, suprabasal, secretory and multiciliated cell compartments. In particular, we propose that if all secretory cells produce anti-microbial peptides, one subset of secretory cells may be more specifically engaged in immune-cell signaling. Our dataset also provides extensive characterization of the deuterosomal cell population. In addition, we have established an exhaustive repertoire of keratin expression and showed that observation of the “keratin switch” during differentiation could be self-sufficient to establish the identity of the different cell

populations. Finally, we have improved signaling pathway characterization by detecting putative Notch repressors that might participate in Notch signal shutdown at the deuterosomal stage, and reporting Wnt pathway activity within the basal cell compartment.

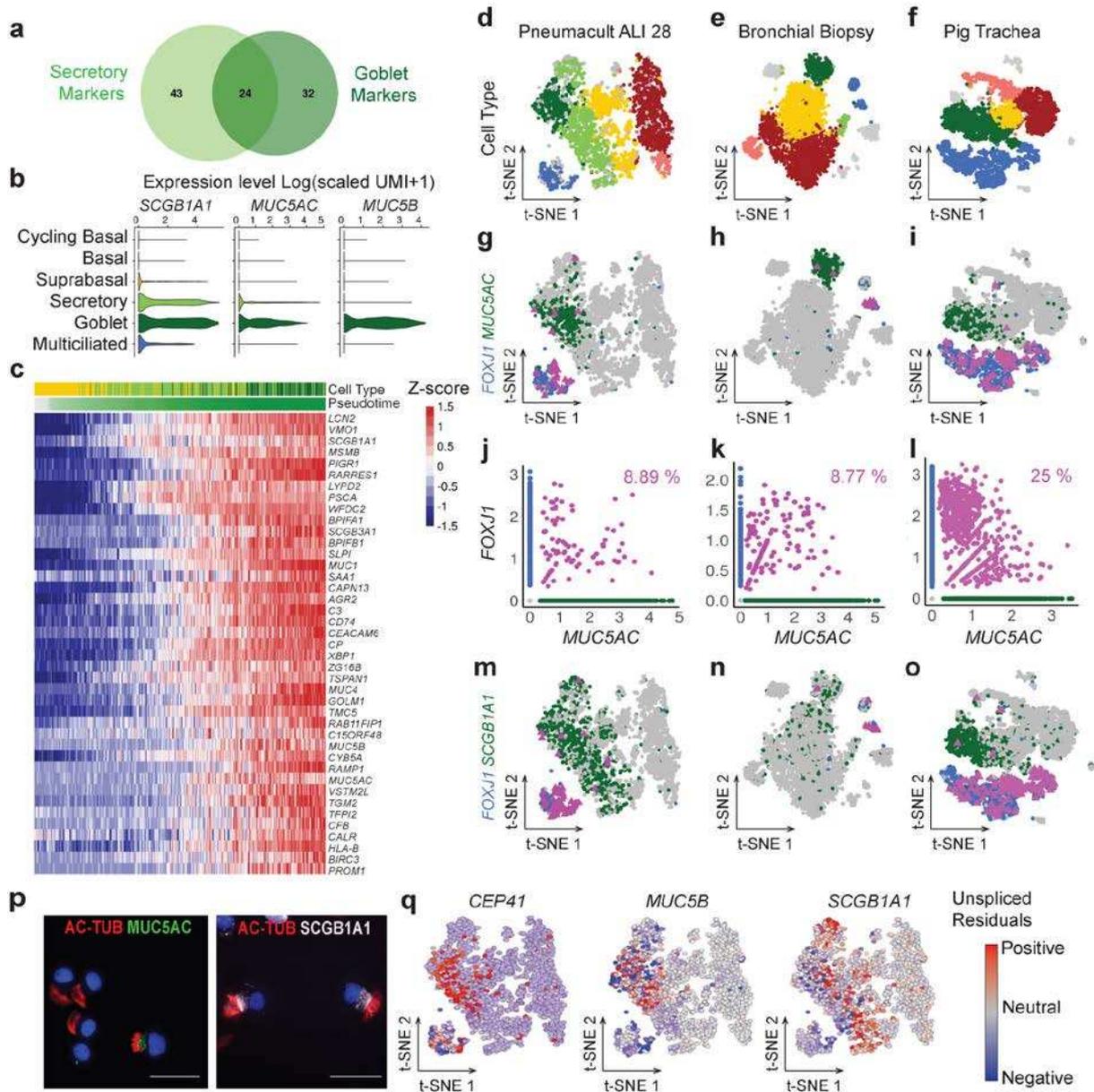
## Figures



**Figure 1: Characterization of MCC and Goblet cell lineages during airway epithelium regeneration using single cell RNA-Seq**

**a** Model of upper airway epithelium, based on 6 major types of epithelial cells, with consensus lineage hierarchy. **b** single-cell (sc)RNA-seq experimental design. Regenerating airway epithelia was dissociated at successive days (7, 12, 28) after a transition to an air-liquid interface (ALI). **c** t-SNE plots of the scRNA-seq expression data highlighting the main cell types observed at ALI7 (3426 cells), ALI12 (2785 cells), ALI28 (3615 cells) (grey: unassigned cells). **d** Relative abundance of the 6 main cell types at each time point. **e** Aggregate t-SNE plot of gene expression in 9826 cells. **f** Inference of goblet and MCC cell lineages by Monocle 2, based on an aggregate of the entire experiment. Color code is the

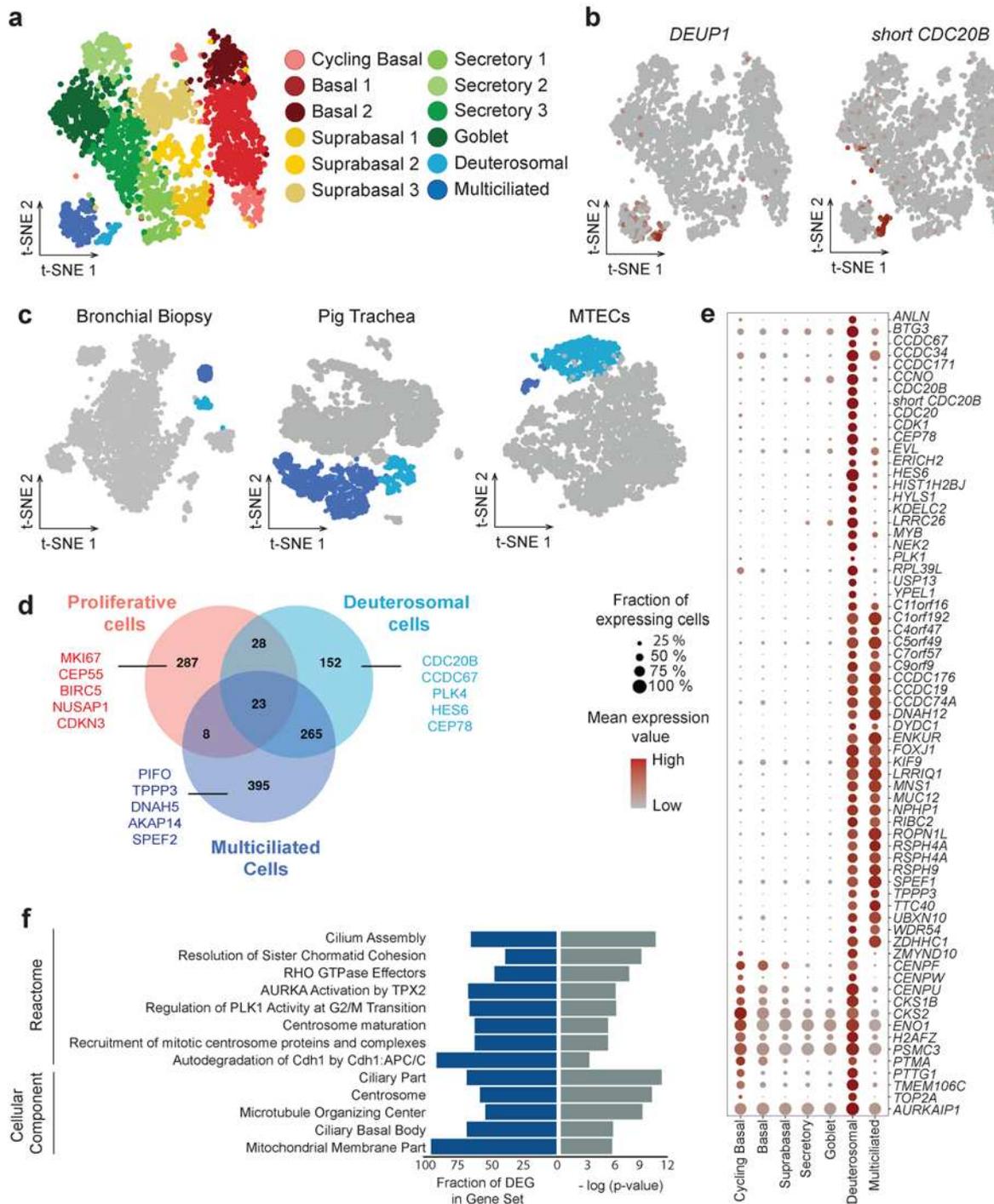
same as in c. Inset: pseudotime picturing by a white to grey gradient along the differentiation trajectory. **g** Distribution of the 6 main cell types in the pseudotime along the two branches of the trajectory (down: goblet cell branch; upper right: multiciliated cell branch). **h** Heatmap representing the smoothed temporal expression pattern of a representative list of cell type specific markers, with branch representation as in g. Cells were ordered by branch, then cluster appearance, then pseudotime.



**Figure 2: Goblet cells as differentiation intermediates for multiciliated cells**

**a** Venn diagram illustrating closeness of the best marker genes for secretory and goblet cells deduced from scRNA-seq of cells differentiated in Pneumacult medium (ALI28). **b** Violin plots of normalized expression for *SCGB1A1*, *MUC5AC* and *MUC5B*, three markers of secretory and goblet cells. **c** Heat map of the most differentially expressed genes between groups of suprabasal, secretory and goblet cells at key point in the pseudotime (before branching, start of the GC branch and end of the GC branch). Cells are ordered by pseudotime. Bars on the top of the heatmap indicate cell type and pseudotime. **d-e-f** t-SNE plots of expression from scRNA-seq of ALI28 (**d**), bronchial biopsy cells (**e**), pig tracheal cells (**f**). Colors indicates cell types as in Figure 1. **g-h-i** Highlights of gene expression for *FOXJ1*<sup>+</sup> cells (blue), *MUC5AC*<sup>+</sup> cells (green), and *FOXJ1*<sup>+</sup>/*MUC5AC*<sup>+</sup> cells (pink) in the same samples as in **d-e-f**. **j-k-l** Relationships between normalized expression of *MUC5AC* and *FOXJ1* in the three same samples. **m-n-**

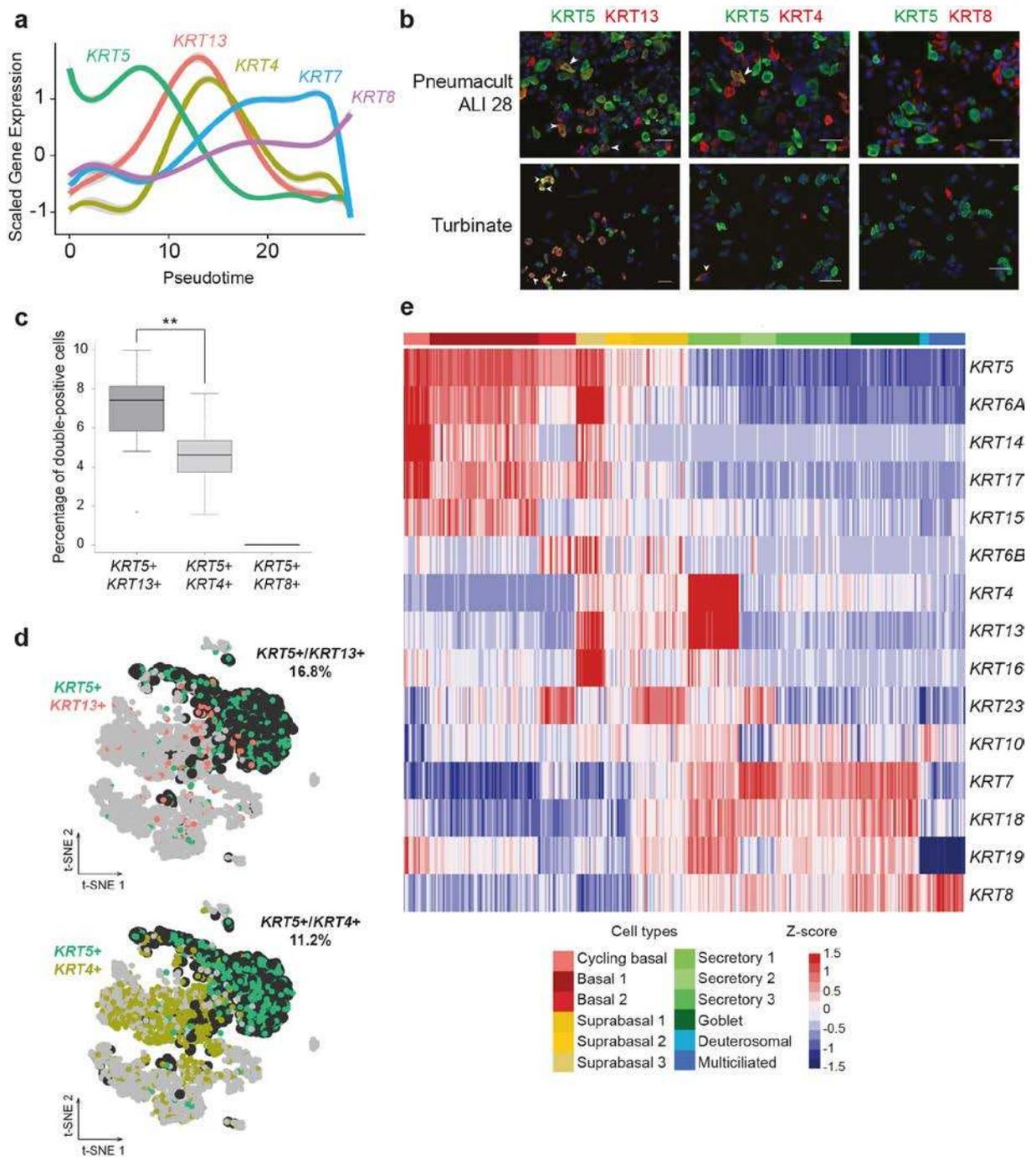
o Highlights of gene expressions for *FOXJ1*+ cells (blue), *SCGB1A1*+ cells (green), and *FOXJ1*/*SCGB1A1*+ cells (pink). p Immunodetection of cells co-expressing markers of multiciliated cells (Acetylated Tubulin) and of goblet cells (*MUC5AC*) (o) or of secretory cells (*SCGB1A1*) (p). q Representation by a t-SNE plot (scRNA-seq of cells differentiated in Pneumacult medium at ALI28) of the Velocity residuals colored according to estimates of the positive (red) and negative (blue) residues for a multiciliated cell marker (*CEP41*), a goblet cell marker (*MUC5B*) and a secretory cell marker (*SCGB1A1*).



**Figure 3: Deuterosomal cells form a discrete MCC intermediate population with a centriole amplification signature**

**a** Subcluterization of scRNA-seq from cells differentiated in Pneumacult medium (ALI28) into 12 cell types, deduced from intra-heterogeneity analysis of the 6 initial clusters. **b** Illustration of the specific expression of *DEUP1* and *short CDC20B* in the deuterosomal cell population (low to high expression, grey to red). **c** Identification of the cluster of deuterosomal cells in scRNA-seq data from a biopsy of human bronchi, pig trachea and mouse primary culture (MTEC, ALI3). Deuterosomal (light blue). Multiciliated cluster (dark blue). **d** Venn Diagram showing overlaps existing between top gene markers

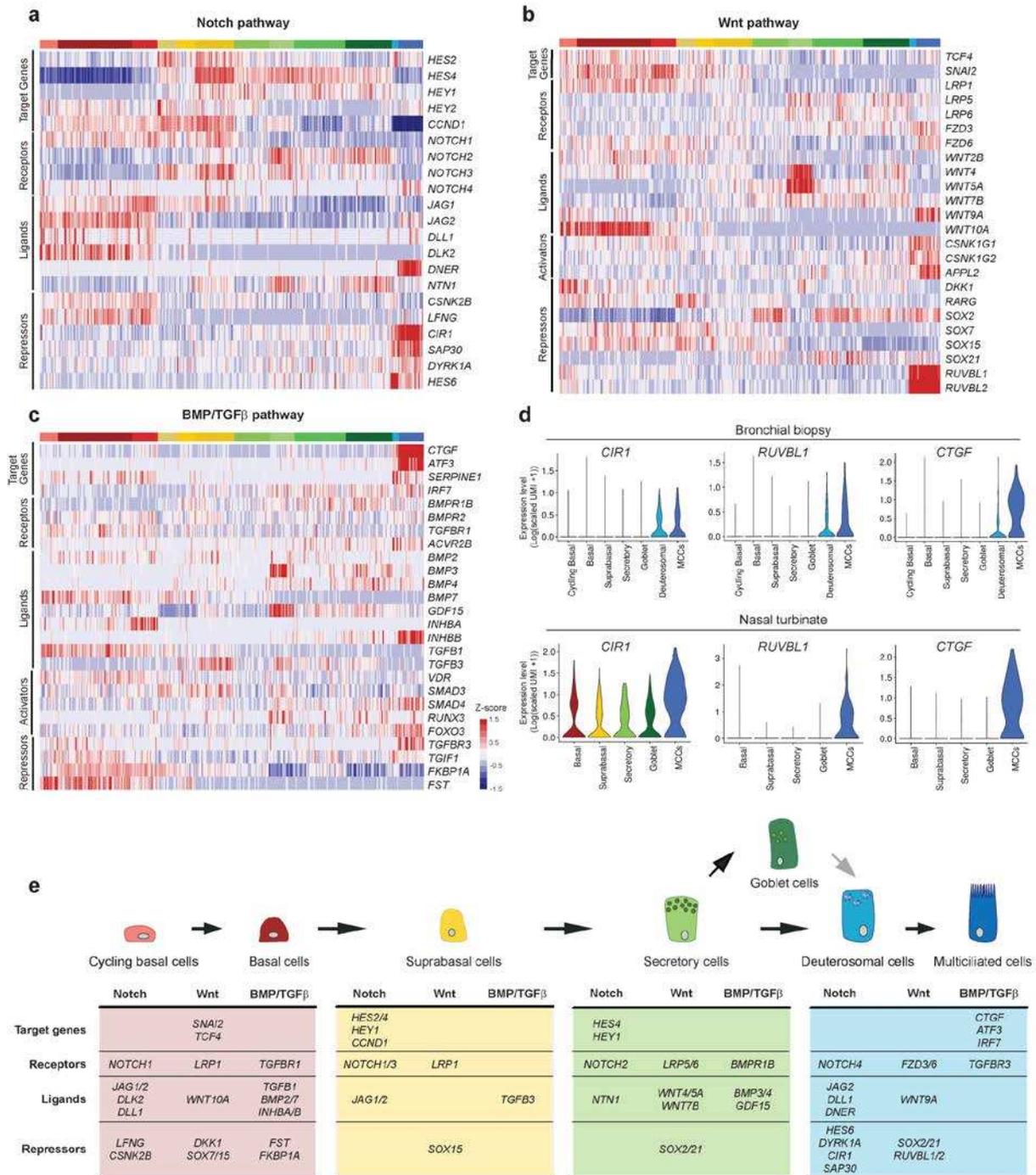
of deuterosomal cells (light blue) and those of proliferative (pink) or multiciliated cells (dark blue). **e** Dot plot of marker genes for the deuterosomal cell population. Color gradient (grey to red) and dot size indicate for each cluster the mean marker expression and the percentage of cells expressing the marker, respectively. **f** Enriched gene sets in deuterosomal cell marker genes.



**Figure 4: Keratin signatures switch during the differentiation process**

**a** Plot of normalized gene expression of keratins according to pseudotime from scRNA-seq of cells differentiated in Pneumacult medium (ALI28). **b** Double immunofluorescence staining for KRT5 and KRT13, KRT4 or KRT8. White arrowheads indicate doubly labelled cells (KRT5+/KRT13+, KRT5+/KRT4+, KRT5+/KRT8+). Nuclei are shown in blue (DAPI). **c** Quantification of double-positive cells from **b**. \*\*: pvalue<0.01 (Wilcoxon test). **d** tSNEs of scRNA-seq data from pig tracheal epithelial cells. KRT5+ cells are shown in green, KRT13+ are shown in red, KRT4+ are shown in light green, double-positive cells are

shown in black. The indicated percentage corresponds to double-positive cells. **e** Heatmap for scRNA-seq data from Pneumacult ALI28 showing gene expression for keratins.



**Figure 5: Single-cell expression of signaling pathways components during airway regeneration**

**a** Heatmap of the genes related to the Notch pathway] with cells ordered by cluster. **b** Heatmap of the genes related to the Wnt pathway with cells ordered by cluster. **c** Heatmap of the genes related to the BMP/TGFβ pathway with cells ordered first by cluster. **d** Violin plots for selected genes in the bronchial biopsy and nasal turbinate samples. **e** Summary of the major partners involved in specific cell types for the three pathways. Cell type colors are as in Fig. 3 and 4.

## **Methods**

### **Human Airway Epithelial Cells (HAECs) culture**

HAECs cultures were derived from nasal mucosa of inferior turbinates. After excision, nasal inferior turbinates were immediately immersed in Ca<sup>2+</sup>/Mg<sup>2+</sup>-free HBSS supplemented with 25 mM HEPES, 200 U/mL penicillin, 200 µg/mL streptomycin, 50 µg/mL gentamicin sulfate, and 2.5 µg/ml amphotericin B (all reagents from Gibco). After repeated washes with cold supplemented HBSS, tissues were digested with 0.1% Protease XIV from *Streptomyces Griseus* (Sigma-Aldrich) overnight at 4°C. After incubation, fetal calf serum (FCS) was added to a final concentration of 10%, and nasal epithelial cells were detached from the stroma by gentle agitation. Cell suspensions were further dissociated by trituration through a 21G-needle and then centrifuged at 150g for 5 min. The pellet was resuspended in supplemented HBSS containing 10% FCS and centrifuged again. The second cell pellet was then suspended in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 10% FCS and cells were plated (20 000 cells per cm<sup>2</sup>) on 75 cm<sup>2</sup>-flasks coated with rat-tail collagen I (Sigma-Aldrich). Cells were incubated in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C. Culture medium was replaced with Bronchial Epithelium Basal Medium (BEBM, Lonza) supplemented with BEGM SingleQuot™ Kit Supplements (Lonza) on the day after and was then changed every other day. After 4 to 5 days of culture, after reaching about 70% confluence, cells were detached with trypsin-EDTA 0.05% (Gibco) for 5 min and seeded on Transwell® permeable supports (6.5 mm diameter; 0.4 µm pore size; Corning), in BEGM medium, with a density of 30 000 cells per Transwell®. Once the cells have reached confluence (typically after 5 days), they were induced to differentiate at the air-liquid interface by removing medium at the apical side of the Transwell®, and by replacing medium at the basal side with either DMEM:BEBM (1:1) supplemented with BEGM SingleQuot™ Kit Supplements or with Pneumacult-ALI (StemCell Technologies) as indicated in the figure legends. Culture medium was changed every other day.

### **Mouse tracheal epithelial cells (MTECs)**

MTECs cell cultures were established from the tracheas of 12 weeks-old mice. After dissection, tracheas were placed in cold DMEM:F-12 medium (1:1) supplemented with 15 mM HEPES, 100 U/mL penicillin, 100 µg/mL streptomycin, 50 µg/mL gentamicin sulfate, and 2.5 µg/ml amphotericin B. Each trachea was processed under a binocular microscope to remove as much conjunctive tissue as possible with small forceps and was opened longitudinally with small dissecting scissors. Tracheas were then placed in supplemented DMEM:F-12 containing 0.15% protease XIV from *Streptomyces Griseus*. After overnight incubation at 4°C, FCS was added to a final concentration of 10%, and tracheal epithelial cells were detached by gentle agitation. Cells were centrifuged at 400g for 10 min and resuspended in supplemented DMEM:F-12 containing 10% FCS. Cells were plated on regular cell culture plates and maintained in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C for 4 hours to allow attachment of putative contaminating fibroblast. Medium containing cells in suspension was further centrifuged at 400g for 5 min and cells were resuspended in supplemented DMEM:F-12 containing BEGM Singlequot™ kit supplements and 5% FCS. Cells were plated on rat tail collagen I-coated Transwell®. Typically, 5 tracheas resulted in 12

Transwells<sup>®</sup>. Medium was changed every other day. Air-liquid interface culture was conducted once transepithelial electrical resistance had reached a minimum of 1000 ohm/cm<sup>2</sup> (measured with EVOM2, World Precision Instruments). Air-liquid interface culture was obtained by removing medium at the apical side of the Transwell<sup>®</sup>, and by replacing medium at the basal side with Pneumacult-ALI medium (StemCell Technologies).

#### **HAEC and MTEC dissociation for single-cell RNA-seq**

Single-cell analysis was performed at the indicated days of culture at the air-liquid interface. To obtain a single-cell suspension, cells were incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS for 4 hours at 4°C degrees. Cells were gently detached from Transwells<sup>®</sup> by pipetting and then transferred to a microtube. 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250 µL were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min) and resuspended in 500 µL supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min) and resuspended in 500 µL HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered through a 40 µm porosity Flowmi<sup>™</sup> Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500 µL of cold HBSS. Cell concentration measurements were performed with Scepter<sup>™</sup> 2.0 Cell Counter (Millipore) and Countess<sup>™</sup> automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess<sup>™</sup> automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 300 cells/µl in HBSS aiming to capture 1500 cells for HAECs and 5000 cells for MTECs.

#### **Turbinates epithelial cell dissociation**

To obtain a single-cell suspension from turbinates, the whole turbinate was incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS at 4°C degrees overnight. Epithelial cells were gently detached from the turbinate by washing with HBSS pipetting up and down and then transferred to a 50 ml Falcon tube. Cells were centrifuged (150g for 5 min at 4°C), after removing of the supernatant the cells were resuspended in 1 ml of HBSS, 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250 µL were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min at 4°C) and resuspended in 1 ml supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min at 4°C) and resuspended in 500 µL HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered 40 µm porosity Flowmi<sup>™</sup> Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500 µL of cold HBSS. Cell concentration measurements were performed with Scepter<sup>™</sup> 2.0 Cell Counter (Millipore) and Countess<sup>™</sup> automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess<sup>™</sup> automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/µl in HBSS aiming to capture 5000 cells.

### **Anesthetic procedure**

Intranasal anesthesia is performed with topical application (gauze) of 5% lidocaine (anesthetic) plus naphazoline (vasoconstrictor) solution (0.2 mg/ml). Laryngeal and endobronchial anesthesia is performed with topical application of 2% lidocaine through the working channel of a 4.9 mm outer diameter bronchoscope.

### **Nasal brushing**

Brushing was performed with a 2 mm cytology brush (Medi-Globe) in the inferior turbinate zone.

### **Bronchial biopsy**

Bronchial biopsy was performed at the spur between the left upper lobe and the left lower lobe with a 1.8mm-diameter Flexibite biopsy forceps (Medi-Globe) passed through the working channel of the bronchoscope (WCB).

### **Dissociation of nasal brushing**

The brush was soaked in a 5 mL Eppendorf containing 1 mL of dissociation buffer which was composed of HypoThermosol® (BioLife Solutions) 10 mg/mL protease from *Bacillus Licheniformis* (Sigma-Aldrich, reference P5380) and 0.5 mM EDTA. The tube was shaken vigorously and centrifuged for 2 min at 150 g. The brush was removed, cells pipetted up and down 5 times and then incubated cells on ice for 30 min, with gentle trituration with 21G needles 5 times every 5 min. Protease was inactivated by adding 200 µL of HBSS/2% BSA. Cells were centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet. Cells were resuspended in 500 µL of wash buffer (HBSS/0.05% BSA) and 2.250 mL of Ammonium Chloride 0.8% was added to perform red blood cell lysis. After a 10 min incubation, 2 mL of wash buffer were added, and cells were centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 1000 µL of wash buffer centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 1000 µL of wash buffer and passed through 40 µm porosity Flowmi™ Cell Strainer (Bel-Art), then centrifuged (400g for 5 min at 4°C). Supernatant was discarded leaving 10 µL of residual liquid on the pellet, cells were resuspended in 100 µL of wash buffer. Cell counts and viability were performed with Countess™ automated cell counter (Thermo Fisher Scientific). For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/µL in HBSS aiming to capture 5000 cells. All steps were performed on ice.

### **Dissociation of bronchial biopsy**

The biopsy was soaked in 1 mL dissociation buffer which was composed of DPBS, 10 mg/mL protease from *Bacillus Licheniformis* (Sigma-Aldrich, reference P5380) and 0.5 mM EDTA. After 1 h, the biopsy was finely minced with a scalpel, and returned to dissociation buffer. From this point, the dissociation procedure is the same as the one described in the “dissociation of nasal brushing” section, with an incubation time increased to 1h, and

omitting the red blood cell lysis procedure. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 300 cells/ $\mu$ l in HBSS aiming to capture 5000 cells. All steps were performed on ice.

### **Pig tracheal epithelial cell dissociation**

To obtain a single-cell suspension from pig trachea, whole clean tracheas were incubated with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in supplemented HBSS at 4°C degrees overnight. Epithelial cells were gently detached from the turbinate by washing with HBSS pipetting up and down and then transferred to a 50 ml Falcon tube. Cells were centrifuged (150g for 5 min at 4°C), after removing of the supernatant the cells were resuspended in 1 ml of HBSS, 50 units of DNase I (EN0523 Thermo Fisher Scientific) per 250  $\mu$ L were directly added and cells were further incubated at room temperature for 10 min. Cells were centrifuged (150g for 5 min at 4°C) and resuspended in 1 ml supplemented HBSS containing 10% FCS, centrifuged again (150g for 5 min at 4°C) and resuspended in 500  $\mu$ L HBSS before being mechanically dissociated through a 26G syringe (4 times). Finally, cell suspensions were filtered through 40  $\mu$ m porosity Flowmi™ Cell Strainer (Bel-Art), centrifuged (150 g for 5 min) and resuspended in 500  $\mu$ L of cold HBSS. Cell concentration measurements were performed with Scepter™ 2.0 Cell Counter (Millipore) and Countess™ automated cell counter (Thermo Fisher Scientific). Cell viability was checked with Countess™ automated cell counter (Thermo Fisher Scientific). All steps except the DNase I incubation were performed on ice. For the cell capture by the 10X genomics device, the cell concentration was adjusted to 500 cells/ $\mu$ l in HBSS aiming to capture 5000 cells.

### **Single-cell RNA-seq**

We followed the manufacturer's protocol (Chromium™ Single Cell 3' Reagent Kit, v2 Chemistry) to obtain single cell 3' libraries for Illumina sequencing. Libraries were sequenced with a NextSeq 500/550 High Output v2 kit (75 cycles) that allows up to 91 cycles of paired-end sequencing: Read 1 had a length of 26 bases that included the cell barcode and the UMI; Read 2 had a length of 57 bases that contained the cDNA insert; Index reads for sample index of 8 bases. Cell Ranger Single-Cell Software Suite v1.3 was used to perform sample demultiplexing, barcode processing and single-cell 3' gene counting using standards default parameters and human build hg19, pig build sus scrofa 11.1 and mouse build mm10.

### **Single-cell quantitative PCR**

HAECs were dissociated as described above, then single cells were separated with a C1™ Single-cell AutoPrep system (Fluidigm), followed by quantitative PCR on the Biomark system (Fluidigm) using SsoFast™ evaGreen® Supermix (Biorad) and the primers described in Supplementary Table S2.

### **RNA-seq on dissociated and non-dissociated HAECs**

Two Transwells® from fully differentiated HAECs from 2 distinct donors were each dissociated as described above. After the final resuspension, cells were centrifuged and resuspended in 800 µL Qiazol (Qiagen). Non-dissociated cells from 2 Transwells® were also lysed in 800 µL Qiazol. RNAs were extracted with the miRNeasy mini kit (Qiagen) according to manufacturer's instructions. Two micrograms from each RNA was used in RNA-seq library construction with the Truseq® stranded total RNA kit (Illumina). Sequencing was performed with a NextSeq 500/550 High Output v2 kit (75 cycles). Reads were aligned against hg19 human build using STAR aligner. Low expressed genes were filtered out, then paired differential analysis was performed with DESeq2 comparing dissociated vs non-dissociated samples from cultures generated from 2 different donors. P-values were adjusted for multiple testing using the false discovery rate (FDR). Top differentially expressed genes were selected using the following cutoffs: FDR < 0.001 and an absolute log<sub>2</sub>FC > 1.5

### **Cytospins**

Fully differentiated HAECs were dissociated by incubation with 0.1% protease type XIV from *Streptomyces griseus* (Sigma-Aldrich) in HBSS (Hanks' balanced salts) overnight at 4°C. Cells were gently detached from the Transwells® by pipetting and then transferred to a microtube. Cells were then cytocentrifuged at 72 g for 10 min onto SuperFrost™ Plus slides using a Shandon Cytospin™ 4 cytocentrifuge.

### **Tissue processing and OCT tissue embedding (see if we maintain figures containing this part)**

Nasal turbinates were fixed in PFA 4% at 4°C overnight then, for cryoprotection, tissues were soaked in a 15% sucrose solution until tissue sinks, then soaked in a 30% solution before tissue sinks. Tissue was embedded in "optimal cutting temperature" OCT medium (Thermo Fisher Scientific) at room temperature and then submerged in Isopentane previously tempered at -80°C. Cutting of frozen tissues was performed with a cryostat Leica CM3050 S.

### **Immunostaining**

Cytospin™ slides were fixed for 10 min in 4% paraformaldehyde at room temperature and then permeabilized with 0.5% Triton X-100 in PBS for 10 min. Cells were blocked with 3% BSA in PBS for 30 min. The incubation with primary antibodies was carried out at 4°C overnight.

Primary antibodies: mouse monoclonal KRT4 (1:50, Santa Cruz Biotechnology sc-52321), mouse monoclonal KRT8 (1:50, Santa Cruz Biotechnology, sc-58737), mouse monoclonal KRT13 (1:200, Sigma-Aldrich clone KS-1A3), rabbit KRT5 (1:2000, Biolegend BLE905501), Rabbit CC10 (SCGB1A1) (1:500, Millipore 07-623), mouse monoclonal Acetylated Tubulin (1:500, Sigma-Aldrich clone 6-11B-1), mouse monoclonal MUC5AC (1:250, Abnova clone 45M1).

Secondary antibodies: Alexa Fluor 488 goat anti-rabbit (1:500; Thermo Fisher Scientific), Alexa Fluor 647 goat anti-mouse (1:500; Thermo Fisher Scientific), Alexa Fluor 488 Goat anti-Mouse IgG1 (1:500, Fisher Scientific), Alexa Fluor 594 Goat anti-Mouse IgG2a (1:500, Fisher Scientific), Alexa Fluor 647 Goat anti-Mouse IgG2b (1:500,

Fisher Scientific). Incubation with secondary antibodies was carried out during 1h at room temperature. Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI).

When necessary, Acetylated Tubulin, Muc5AC and KRT5 antibodies were directly coupled to CF™ 594/488/488 respectively, with the Mix-n-Stain™ kit (Sigma-Aldrich) according to the manufacturer's instruction. Coupled primary antibodies were applied for 2 hours at room temperature after secondary antibodies had been extensively washed and after a 30 min blocking stage in 3% normal rabbit or mouse serum in PBS. MTECs immunostainings were directly performed on Transwell® membranes using a similar protocol. For mounting on slides, Transwell® membranes were cut with a razor blade and mounted with ProLong™ Gold medium (Thermo Fisher Scientific). Images were acquired using the Olympus Fv10i or Leica sp5 confocal imaging systems.

### **Time course sample analysis**

*Preprocessing.* For each sample, cells with levels in the top 5% or bottom 5% of distribution for the following quality metrics: number of expressed features, dropout percentage and library size (total UMI count) were filtered out. Additionally, cells with a percentage of mitochondrial genes > top 5% were also removed. Quality metrics were computed using the scater package (2.3.0) [65]. Genes expressed at less than 1 UMI in at least 5 cells were removed from further analysis.

*Normalization.* The scran package [66] was used to calculate cell-based scale factors and normalize cells for differences in count distribution. Each sample was normalized separately twice, first in an unsupervised manner, then after grouping cells of similar gene expression based on our robust clustering results.

*Clustering Robustness.* In order to best determine the key steps in the studied differentiation process, a customized method was implemented to analyze clustering robustness to dataset perturbation. For all possible number of clusters (from 2 to 9), multiple subsets of the studied dataset were created (10 subsets, with 10% of the cells randomly removed each time) and clustering was performed multiple times on each subset with changing settings of the seed parameter. The result of those clusterings were stored in a n cells x n cells stability matrix, containing for each pair of cells 1 or 0 if the cells are clustered together or not. This stability matrix was then transformed in a Euclidean distance matrix between cells and then divided into the used k number of clusters k using hierarchical clustering (hclust with 'average' method). To identify the optimal number of clusters, a visual inspection of the elbow plot of the average intra-stability (mean stability within each cluster) and the average inter-stability (mean stability between each cluster) was done. Cells with a stability metric less than 70% were labelled as Unassigned, due to their high clustering variability between each round of clustering, then removed from further analysis of the time course data. Cell clustering was performed using SIMLR (package version 1.4.1) [67].

*Differential analysis.* To further analyze the robustness of each step of the differentiation process, we tested the robustness of the cell type marker gene identification through differential gene expression analysis. Differential expression analysis was performed using edgeR (package version 3.22) [68]. In a one vs. all differential analysis, a pool of 100 cells from one cluster were analyzed against an equal mixture of cells from all other clusters. In a one vs one differential analysis pools of cells of the same size were compared. Those differential analysis were

performed multiple times (10 times) on different pool of cells and the DEG identified were compared between each pool of cells using the rank-rank hypergeometric overlap algorithm [69]. Unfortunately, this approach was too stringent, and only identified highly expressed marker gene as they are less likely to submit to dropout events. Thus, the Seurat FindAllMarkers function based on a non-parametric Wilcoxon rank sum test was used to identify cell type marker genes.

*Time points aggregation.* 10X datasets generated during the time course were aggregated using MNN correction [70] from the scran package.

*Trajectory Inference.* Trajectory inference was performed using monocle 2 (package version 2.8) [71]. Cell ordering was based on highly variable genes (~ 200 - 500 genes) selected by their expression dispersion. Monocle analysis on the aggregated time points was done on raw count after library size correction (downsampling). Branch building was performed using BEAM analysis within the monocle package, and corresponding differential analysis was done by cross comparison of group of cells along the pseudotime (before branching, after branching and at the branch end) using Seurat 1 vs 1 differential analysis.

*Cell type projection.* To compare cell types identified in distinct samples, cells were projected from one dataset onto the other using scmap R package version 1.1, scmapCluster function [72].

*Data Visualization.* All graphs were generated using R (ggplot2). Heatmaps were obtained using pheatmap (no clustering used, genes ordered by their expression in pseudotime or in cluster, cells ordered by pseudotime or cluster). Heatmaps show smoothed gene expression values: for each gene, normalized gene expression values were first transformed into z-scores, then averaged across 10 neighboring cells in the chosen ordering (pseudotime only or pseudotime in clusters). For single gene representation: only cells with expression levels above the top 50 percentiles for that gene are represented for clarity.

### **Individual sample analysis**

Each sample of our study was reanalyzed with less stringent parameters, to identify rare or transitory cell types or gene expression events

*Preprocessing, normalization and clustering.* Individual dataset analysis was performed using Seurat standard analysis pipeline. Briefly, cells were first filtered based on number of expressed features, dropout percentage, library size and mitochondrial gene percentage. Thresholds were selected by visually inspecting violin plots in order to remove the most extreme outliers. Genes expressing less than 5 UMI across all cells were removed from further analysis. Cell-level normalization was performed using the median UMI counts as scaling factor. Highly variable genes were selected for following analysis based on their expression level and variance. PCA analysis was performed on those hvgs, the number of PCs to use was chosen upon visual inspection of the PC variance elbowplot (~10 to 20 PCs depending on the dataset). Clustering was first performed with default parameter and then increasing the resolution parameter above 0.5 to identify small clusters (but with the knowledgeable risk of

splitting big cluster due to high gene expression variability). Differential analysis was again performed using Seurat FindAllMarkers and FindMarkers functions based on non-parametric Wilcoxon rank sum test. Gene Set Enrichment analysis was performed using fgsea R package with the following gene sets reactome.db (R package) and GO cellular component (Broad Institute GSEA MSigDB) genesets. Molecular function enrichment analysis was performed using Ingenuity Pathway Analysis (Qiagen).

*Cell Type Annotation.* Based on the time course experiment analysis and associated top ~15 marker genes identified, a score was computed to associate cell types to each cluster. Scoring method is based on Macosko. *et al.* cell cycle phase assignment [73]. For each cell it measures the mean expression of the top marker genes for each possible cell type, which results in a matrix c cell types per n cells. Then it performs a z-score of those mean expression for each cell, the top resulting score gives the matching cell type.

*Velocity.* RNA velocity was calculate using latest release of velocityto pipeline (<http://velocityto.org/>) using standard parameters: GTF file used for Cell Ranger analysis and the possorted\_genome\_bam.bam, Cell Ranger output alignment file. From the loom file which contains count table of spliced and unspliced transcript the gene.relative.velocity.estimates function was used on cell type marker genes. The resulting expression pattern of unspliced-spliced phase portraits shows the induction or repression of those marker genes from one cell type to the next. We used velocityto package version 0.5 [33].

#### **Plasscheart *et al.* dataset**

Plasscheart *et al.* data [29] were downloaded as processed data along with visualization coordinates and were used without further manipulation.

([https://kleintools.hms.harvard.edu/tools/springViewer\\_1\\_6\\_dev.html?datasets/reference\\_HBECs/reference\\_HBECs](https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?datasets/reference_HBECs/reference_HBECs))

## **Declarations**

### **Ethics approval and consent to participate**

Human samples were collected according to the guidelines of the Declaration of Helsinki, after approval by the institutional review board "Comité de Protection des Personnes Sud Est IV " (12/12/2017). All patients gave their written informed consent to participate. Mouse experiments were performed following the Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. The IPMC is authorized to use animals by the French agricultural ministry (Ministere de l'Agriculture et de l'Alimentation) and persons performing experiments were authorized by the French research ministry (Ministère de l'Education Supérieure, de la Recherche et de l'Innovation).

### **Consent for publication**

All patients gave their written informed consent for anonymous publication of data arising from their samples.

### **Availability of data and materials**

All datasets generated during this study were deposited to the Gene Expression Omnibus repository under the series number GSE121600.

### **Competing interests**

The authors declare no competing financial interests.

### **Funding**

This project was funded by grants from FRM (DEQ20180339158), the labex Signallife (ANR-11-LABX-0028-01), the association Vaincre la Mucoviscidose (RF20180502280), and the Chan Zuckerberg Initiative (Silicon Valley Foundation, 2017-175159 -5022). The UCAGenomiX platform, a partner of the National Infrastructure France Génomique, is supported by Commissariat aux Grands Investissements (ANR-10-INBS-09-03, ANR-10-INBS-09-02) and Canceropôle PACA.

### **Authors' contributions**

LEZ and PB designed and supervised the study and obtained fundings. SRG, MJA, AC, VM and LEZ performed the experiments. SL et CHM provided human fresh tissues. IC provided pig trachea. MD, KLB and AP performed the bioinformatics analysis. All authors were involved in data interpretation. MD and LEZ designed the figures. LEZ drafted the original manuscript. MD, PB, SRG, AP, KLB, BM and LEZ edited the manuscript.

## **Acknowledgements**

We are grateful to the UCAGenomiX platform and to Marin Truchi for fruitful discussions and technical help on single cell RNA sequencing.

## **References**

1. Gras D, Chanez P, Vachier I, Petit A, Bourdin A. Bronchial epithelium as a target for innovative treatments in asthma. *Pharmacol Ther* [Internet]. Elsevier Inc.; 2013;140:290–305. Available from: <http://dx.doi.org/10.1016/j.pharmthera.2013.07.008>
2. Kotton DN, Morrisey EE. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nat Med* [Internet]. 2014;20:822–32. Available from: <http://dx.doi.org/10.1038/nm.3642>
3. Cohn L. Mucus in chronic airway diseases: Sorting out the sticky details. *J Clin Invest*. 2006;116:306–8.
4. Merigo F, Benati D, Piacentini G, Boner A, Sbarbati A. The ultrastructure of nasal mucosa in children with asthma. *Ultrastruct Pathol*. 2002;26:293–8.
5. Curran DR, Cohn L. Advances in mucous cell metaplasia: A plug for mucus as a therapeutic focus in chronic airway disease. *Am J Respir Cell Mol Biol*. 2010;42:268–75.
6. Rock JR, Onaitis MW, Rawlins EL, Lu Y, Clark CP, Xue Y, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* [Internet]. 2009;106:12771–5. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2714281&tool=pmcentrez&rendertype=abstract>
7. Cole BB, Smith RW, Jenkins KM, Graham BB, Reynolds PR, Reynolds SD. Tracheal basal cells: A facultative progenitor cell pool. *Am J Pathol* [Internet]. American Society for Investigative Pathology; 2010;177:362–76. Available from: <http://dx.doi.org/10.2353/ajpath.2010.090870>
8. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, et al. Repair and regeneration of the respiratory system: Complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2014;15:123–38. Available from: <http://dx.doi.org/10.1016/j.stem.2014.07.012>
9. Boers JE, Ambergen AW, Thunnissen FB. Number and Proliferation of Basal and Parabasal Cells in Normal Human Airway Epithelium. *Am J Respir Crit Care Med* [Internet]. 1998;157:2000–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10228131> <http://www.atsjournals.org/doi/pdf/10.1164/10.1164/ajrccm.157.6.9707011>

10. Pardo-Saganta A, Law BMM, Tata PRR, Villoria J, Saez B, Mou H, et al. Injury Induces Direct Lineage Segregation of Functionally Distinct Airway Basal Stem/Progenitor Cell Subpopulations. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2015;16:184–97. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S193459091500003X>
11. Watson JKK, Rulands S, Wilkinson ACC, Wuidart A, Ousset M, Van Keymeulen A, et al. Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. *Cell Rep* [Internet]. 2015;12:90–101. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S2211124715006099>
12. Boers JE, Ambergen AW, Thunnissen FBJM. Number and Proliferation of Clara Cells in Normal Human Airway Epithelium. *Am J Respir Crit Care Med*. 1999;159.
13. Wang S-Z, Rosenberger CL, Bao Y-X, Stark JM, Harrod KS. Clara Cell Secretory Protein Modulates Lung Inflammatory and Immune Responses to Respiratory Syncytial Virus Infection. *J Immunol* [Internet]. 2003;171:1051–60. Available from: <http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.171.2.1051>
14. Jones KG, Holland JF, Foureman GL, Bend JR, Fouts JR. Xenobiotic metabolism in Clara cells and alveolar type II cells isolated from lungs of rats treated with beta-naphthoflavone. *J Pharmacol Exp Ther* [Internet]. 1983 [cited 2018 Jun 18];225:316–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6601704>
15. Rawlins EL, Okubo T, Xue Y, Brass DM, Auten RL, Hasegawa H, et al. The Role of Scgb1a1+ Clara Cells in the Long-Term Maintenance and Repair of Lung Airway, but Not Alveolar, Epithelium. *Cell Stem Cell* [Internet]. Elsevier Ltd; 2009;4:525–34. Available from: <http://dx.doi.org/10.1016/j.stem.2009.04.002>
16. Chen G, Korfhagen TR, Xu Y, Kitzmiller J, Wert SE, Maeda Y, et al. SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J Clin Invest*. 2009;119:2914–24.
17. Morimoto M, Liu Z, Cheng HT, Winters N, Bader D, Kopan R. Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate. *J Cell Sci* [Internet]. 2010/01/06. 2010;123:213–24. Available from: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=20048339](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20048339)
18. Tsao P-NN, Vasconcelos M, Izvolsky KI, Qian J, Lu J, Cardoso W V. Notch signaling controls the

- balance of ciliated and secretory cell fates in developing airways. *Development* [Internet]. 2009/06/09. The Company of Biologists Limited; 2009 [cited 2015 Jun 12];136:2297–307. Available from:  
[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=19502490](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19502490)
19. Pardo-Saganta A, Tata PR, Law BM, Saez B, Chow RD-W, Prabhu M, et al. Parent stem cells can serve as niches for their daughter cells. *Nature* [Internet]. 2015;523:597–601. Available from:  
<http://www.nature.com/doifinder/10.1038/nature14553>
20. Rock JR, Gao X, Xue Y, Randell SH, Kong Y-Y, Hogan BLM. Notch-Dependent Differentiation of Adult Airway Basal Stem Cells. *Cell Stem Cell* [Internet]. 2011 [cited 2015 Apr 18];8:639–48. Available from: <http://www.sciencedirect.com/science/article/pii/S1934590911001688>
21. Marcet B, Chevalier B, Luxardi G, Coraux C, Zaragosi L-EE, Cibois M, et al. Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nat Cell Biol* [Internet]. 2011/05/24. Nature Publishing Group; 2011;13:693–9. Available from:  
[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=21602795](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21602795)
22. Marcet B, Chevalier B, Coraux C, Kodjabachian L, Barbry P. MicroRNA-based silencing of Delta/Notch signaling promotes multiple cilia formation. *Cell Cycle* [Internet]. 2011/08/23. 2011;10:2858–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21857154>
23. Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, et al. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell* [Internet]. Elsevier Inc.; 2017;20:817–830.e8. Available from: <http://dx.doi.org/10.1016/j.stem.2017.04.003>
24. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol* [Internet]. 2018 [cited 2018 Jun 20];19:85–97. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29167569>
25. Pal B, Chen Y, Vaillant F, Jamieson P, Gordon L, Rios AC, et al. Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* [Internet]. Nature Publishing Group; 2017 [cited 2018 Jun 20];8:1627. Available from:  
<http://www.nature.com/articles/s41467-017-01560-x>
26. Plass M, Solana J, Alexander Wolf F, Ayoub S, Misios A, Glažar P, et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* (80- ). 2018;360.

27. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* [Internet]. Nature Publishing Group; 2014 [cited 2014 Apr 28];509:371–5. Available from: <http://www.nature.com/gate1.inist.fr/nature/journal/v509/n7500/full/nature13173.html>
28. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket S, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* [Internet]. Springer US; 2018;560:319–24. Available from: <http://dx.doi.org/10.1038/s41586-018-0393-7>
29. Plasschaert LW, Žilionis R, Choo-wing R, Savova V, Knehr J, Roma G, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*. 2018;560:377–381.
30. Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* [Internet]. Springer US; 2018; Available from: <http://www.nature.com/articles/s41586-018-0449-8>
31. Chevalier B, Adamiok A, Mercey O, Revinski DR, Zaragosi L-E, Pasini A, et al. MiR-34/449 control apical actin network formation during multiciliogenesis through small GTPase pathways. *Nat Commun*. 2015;6.
32. Revinski DR, Zaragosi L-E, Boutin C, Pons N, Marcet B, Kodjabachian L, et al. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nat Commun*. 2018;In Press.
33. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature* [Internet]. Springer US; 2018;560:494–8. Available from: <http://www.nature.com/articles/s41586-018-0414-6>
34. Zaragosi L, Revinski DR, Mercey O, Chevalier B, Kodjabachian L, Marcet B, et al. CDC20B, the host gene of miR-449, controls vertebrate multiciliogenesis. *Prep*. 2017;
35. Al Jord A, Shihavuddin A, Servignat d’Aout R, Faucourt M, Genovesio A, Karaiskou A, et al. Calibrated mitotic oscillator drives motile ciliogenesis. *Science* (80- ) [Internet]. 2017 [cited 2018 Jan 2];358:803–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28982797>
36. Colopy SA, Bjorling DE, Mulligan WA, Bushman W. A population of progenitor cells in the basal and intermediate layers of the murine bladder urothelium contributes to urothelial development and regeneration. *Dev Dyn* [Internet]. 2014 [cited 2018 Sep 21];243:988–98. Available from: <http://doi.wiley.com/10.1002/dvdy.24143>
37. Hudson DL, Guy AT, Fry P, O’Hare MJ, Watt FM, Masters JRW. *Epithelial Cell Differentiation*

Pathways in the Human Prostate: Identification of Intermediate Phenotypes by Keratin Expression. *J Histochem Cytochem* [Internet]. 2001 [cited 2018 Sep 21];49:271–8. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/11156695>

38. Jumppanen M, Gruvberger-Saal S, Kauraniemi P, Tanner M, Bendahl P-O, Lundin M, et al. Basal-like phenotype is not associated with patient survival in estrogen-receptor-negative breast cancers. *Breast Cancer Res* [Internet]. 2007 [cited 2018 Sep 21];9:R16. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1649>

39. Yang L, Nichols J, Yao C, Maniley J, Robey E, Weinmaster G. Fringe Glycosyltransferases Differentially Modulate Notch1 Proteolysis Induced by Delta1 and Jagged1. *Mol Biol Cell* [Internet]. 2004;16:927–42. Available from: <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E04-07-0614>

40. Bae S, Bessho Y, Hojo M, Kageyama R. The bHLH gene Hes6, an inhibitor of Hes1, promotes neuronal differentiation. *Development* [Internet]. 2000;127:2933–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10851137>

41. Cruciat CM. Casein kinase 1 and Wnt/ $\beta$ -catenin signaling. *Curr Opin Cell Biol* [Internet]. Elsevier Ltd; 2014;31:46–55. Available from: <http://dx.doi.org/10.1016/j.ceb.2014.08.003>

42. Tözser J, Earwood R, Kato A, Brown J, Tanaka K, Didier R, et al. TGF- $\beta$  Signaling Regulates the Differentiation of Motile Cilia. *Cell Rep*. 2015;11:1000–7.

43. Chand HS, Harris JF, Tesfaigzi Y. IL-13 in LPS-Induced Inflammation Causes Bcl-2 Expression to Sustain Hyperplastic Mucous cells. *Sci Rep* [Internet]. Nature Publishing Group; 2018 [cited 2018 Oct 13];8:436. Available from: <http://www.nature.com/articles/s41598-017-18884-9>

44. Tyner JW, Kim EY, Ide K, Pelletier MR, Roswit WT, Morton JD, et al. Blocking airway mucous cell metaplasia by inhibiting EGFR antiapoptosis and IL-13 transdifferentiation signals. *J Clin Invest*. 2006;116:309–21.

45. Turner J, Roger J, Fitau J, Combe D, Giddings J, Van Heeke G, et al. Goblet cells are derived from a FOXJ1-expressing progenitor in a human airway epithelium. *Am J Respir Cell Mol Biol*. 2011;44:276–84.

46. Gomperts BN, Kim LJ, Flaherty SA, Hackett BP. IL-13 regulates cilia loss and foxj1 expression in human airway epithelium. *Am J Respir Cell Mol Biol*. 2007;37:339–46.

47. Pardo-Saganta A, Law BM, Gonzalez-Celeiro M, Vinarsky V, Rajagopal J. Ciliated cells of pseudostratified airway epithelium do not become mucous cells after ovalbumin challenge. *Am J*

Respir Cell Mol Biol. 2013;48:364–73.

48. Reynolds SD, Reynolds PR, Pryhuber GS, Finder JD, Stripp BR. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. *Am J Respir Crit Care Med*. 2002;166:1498–509.

49. Guha A, Vasconcelos M, Zhao R, Gower AC, Rajagopal J, Cardoso W V. Analysis of notch signaling-dependent gene expression in developing airways reveals diversity of clara cells. *PLoS One*. 2014;9.

50. Mori M, Mahoney JE, Stupnikov MR, Paez-Cortez JR, Szymaniak AD, Varelas X, et al. Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development* [Internet]. The Company of Biologists Limited; 2015 [cited 2015 Jun 15];142:258–67. Available from: <http://dev.biologists.org/content/142/2/258.full>

51. Wang X, Gupta P, Fairbanks J, Hansen D. Protein kinase CK2 both promotes robust proliferation and inhibits the proliferative fate in the *C. elegans* germ line. *Dev Biol* [Internet]. Elsevier; 2014;392:26–41. Available from: <http://dx.doi.org/10.1016/j.ydbio.2014.05.002>

52. Cheng P, Kumar V, Liu H, Youn J-I, Fishman M, Sherman S, et al. Effects of Notch Signaling on Regulation of Myeloid Cell Differentiation in Cancer. *Cancer Res* [Internet]. 2014 [cited 2018 Oct 15];74:141–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24220241>

53. Nam SM, Kim YN, Kim JW, Kyeong DS, Lee SH, Son Y, et al. Hairy and Enhancer of Split 6 (Hes6) Deficiency in Mouse Impairs Neuroblast Differentiation in Dentate Gyrus Without Affecting Cell Proliferation and Integration into Mature Neurons. *Cell Mol Neurobiol*. Springer US; 2016;36:57–67.

54. Fernandez-Martinez J, Vela EM, Tora-Ponsioen M, Ocaña OH, Nieto MA, Galceran J. Attenuation of Notch signalling by the Down-syndrome-associated kinase DYRK1A. *J Cell Sci* [Internet]. The Company of Biologists Ltd; 2009 [cited 2018 Oct 15];122:1574–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19383720>

55. Hsieh JJ, Zhou S, Chen L, Young DB, Hayward SD. CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. *Proc Natl Acad Sci U S A* [Internet]. 1999;96:23–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9874765><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC15086>

56. Estrach S, Ambler CA, Lo Celso CL, Hozumi K, Watt FM. Jagged 1 is a -catenin target gene required for ectopic hair follicle formation in adult epidermis. *Development* [Internet].

2006;133:4427–38. Available from: <http://dev.biologists.org/cgi/doi/10.1242/dev.02644>

57. Malleske DT, Hayes D, Lallier SW, Hill CL, Reynolds SD. Regulation of human airway epithelial tissue stem cell differentiation by  $\beta$ -catenin, P300, and CBP. *Stem Cells* [Internet]. 2018 [cited 2018 Oct 15]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30171668>

58. Koopmans T, Kumawat K, Halayko AJ, Gosens R. Regulation of actin dynamics by WNT-5A: implications for human airway smooth muscle contraction. *Sci Rep* [Internet]. 2016 [cited 2018 Oct 15];6:30676. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27468699>

59. Durham AL, McLaren A, Hayes BP, Caramori G, Clayton CL, Barnes PJ, et al. Regulation of Wnt4 in chronic obstructive pulmonary disease. *FASEB J* [Internet]. 2013;27:2367–81. Available from: <http://www.fasebj.org/doi/10.1096/fj.12-217083>

60. Ji H, Goode RJAA, Vaillant F, Mathivanan S, Kapp EA, Mathias RA, et al. Proteomic profiling of secretome and adherent plasma membranes from distinct mammary epithelial cell subpopulations. *Proteomics* [Internet]. 2011 [cited 2018 Oct 15];11:4029–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21834135>

61. Kessler M, Hoffmann K, Brinkmann V, Thieck O, Jackisch S, Toelle B, et al. The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. *Nat Commun* [Internet]. Nature Publishing Group; 2015;6:1–11. Available from: <http://dx.doi.org/10.1038/ncomms9989>

62. DiMeo TA, Anderson K, Phadke P, Feng C, Perou CM, Naber S, et al. A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Res* [Internet]. 2009 [cited 2018 Oct 15];69:5364–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19549913>

63. Bauer A, Chauvet S, Huber O, Usseglio F, Rothbacher U, Aragnol D, et al. Pontin52 and Reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J* [Internet]. 2000 [cited 2018 Oct 15];19:6121–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11080158>

64. Weiske J, Huber O. The histidine triad protein Hint1 interacts with Pontin and Reptin and inhibits TCF- $\beta$ -catenin-mediated transcription. *J Cell Sci* [Internet]. 2005 [cited 2018 Oct 15];118:3117–29. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16014379>

65. McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* [Internet]. Oxford

University Press; 2017 [cited 2018 Oct 19];33:btw777. Available from:

<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw777>

66. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* [Internet]. 2016 [cited 2018 Oct 19];5:2122.

Available from: <https://f1000research.com/articles/5-2122/v2>

67. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat Methods* [Internet]. 2017 [cited 2018 Oct

19];14:414–6. Available from: <http://www.nature.com/articles/nmeth.4207>

68. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* [Internet]. 2010 [cited 2018 Oct 19];26:139–

40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19910308>

69. Plaisier SB, Taschereau R, Wong JA, Graeber TG. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res* [Internet]. 2010 [cited 2018 Oct 19];38:e169–e169. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/20660011>

70. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* [Internet]. 2018 [cited 2018 Oct

19];36:421–7. Available from: <http://www.nature.com/doi/10.1038/nbt.4091>

71. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14:979–82.

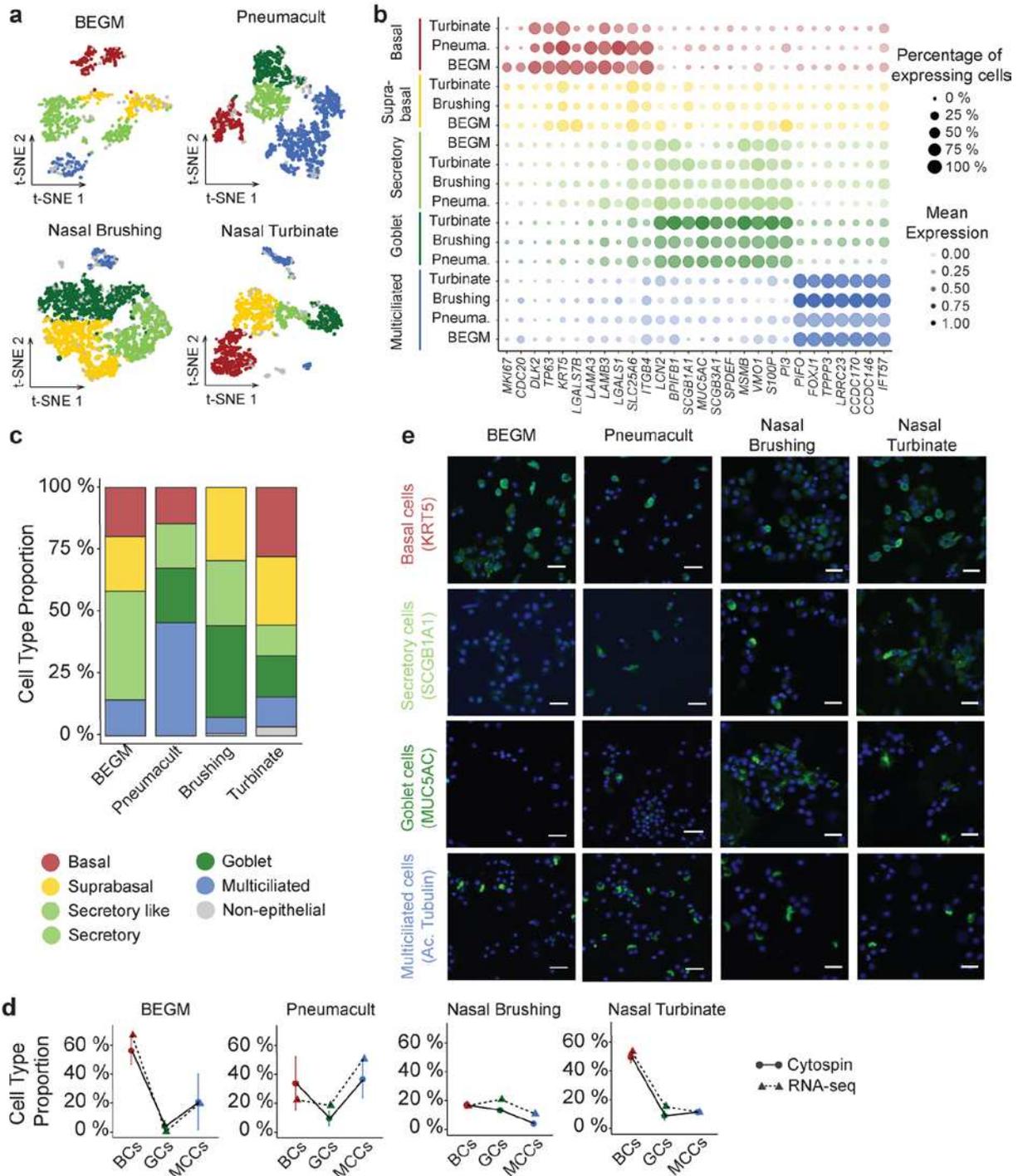
72. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. *Nat Methods* [Internet]. 2018 [cited 2018 Oct 19];15:359–62. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/29608555>

73. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* [Internet]. 2015 [cited 2018

Jan 25];161:1202–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26000488>

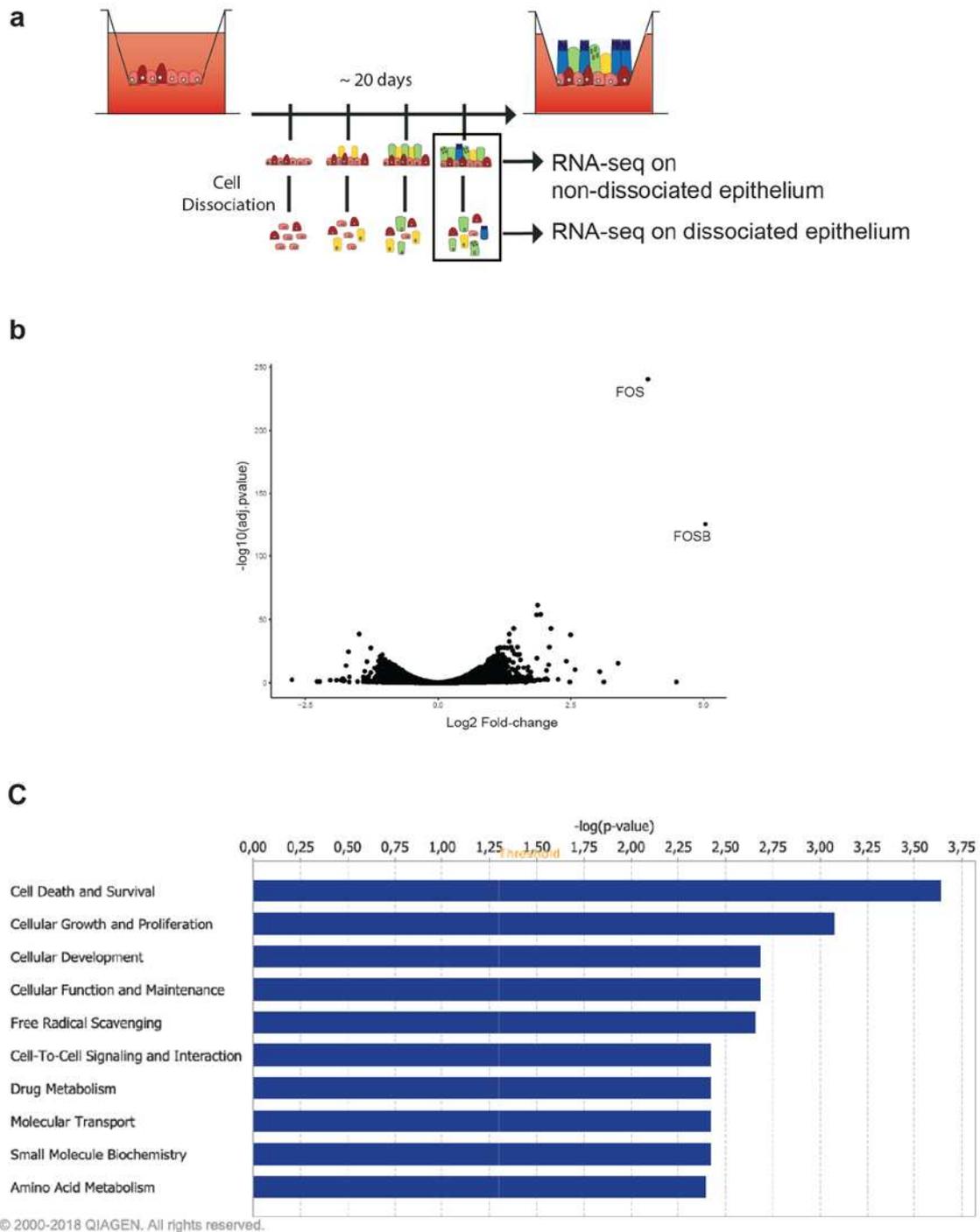
## Supplementary figures



**Supplementary figure 1: Cell type composition comparison between homeostatic *in vitro* samples and fresh human tissues**

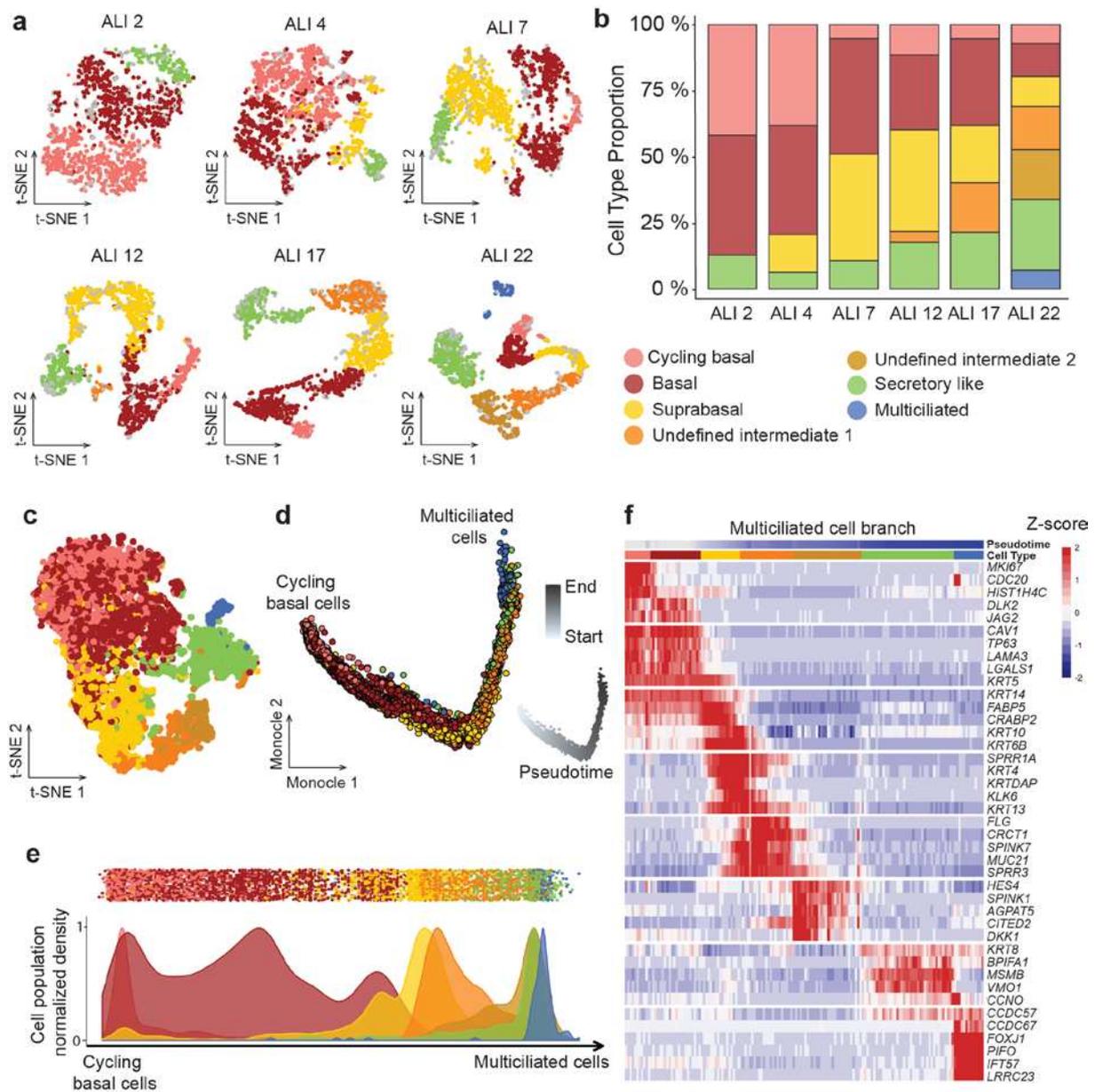
**a** t-SNE representing the distinct cell populations identified in each sample. **b** Dot plot of the main cell population marker genes. Dot size describe the percentage of cells expressing the respective marker genes and the average expression level of that gene based on UMI counts are shown by color intensity. **c** Relative abundance of cell types in each sample. **d** Scatter plot of the percentage of cells expressing

the selected marker genes for basal, goblet and multiciliated cells at the transcript (dotted line) and protein level (full line). **e** Immunostaining of the selected marker genes for basal, secretory, goblet and multiciliated in cytospin for each sample.



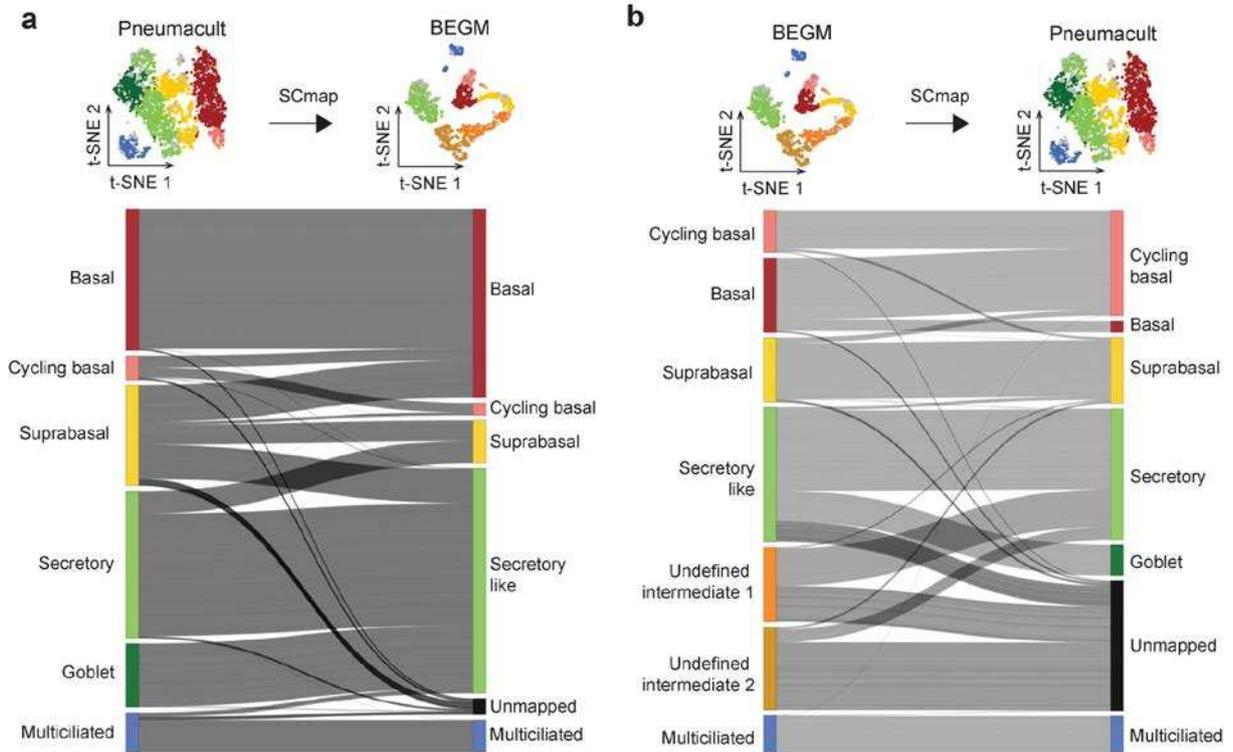
### Supplementary figure 2: Effect of cell dissociation on HAEC gene expression

**a** RNA-seq experimental design. Regenerating airway epithelia was dissociated at 28 days after a transition to an air-liquid interface. Dissociated and non-dissociated cultures from 2 donors were subjected to RNA-seq. **b** Volcano plot showing differential gene expression of dissociated vs. non-dissociated cell cultures. **c** Identification, with Ingenuity Pathway Analysis (Qiagen) of the molecular functions most significantly affected by dissociation, based on an analysis of 300 differential expressed genes ( $FDR < 0.01$  and  $abs(\log_2FC) > 1$ ).



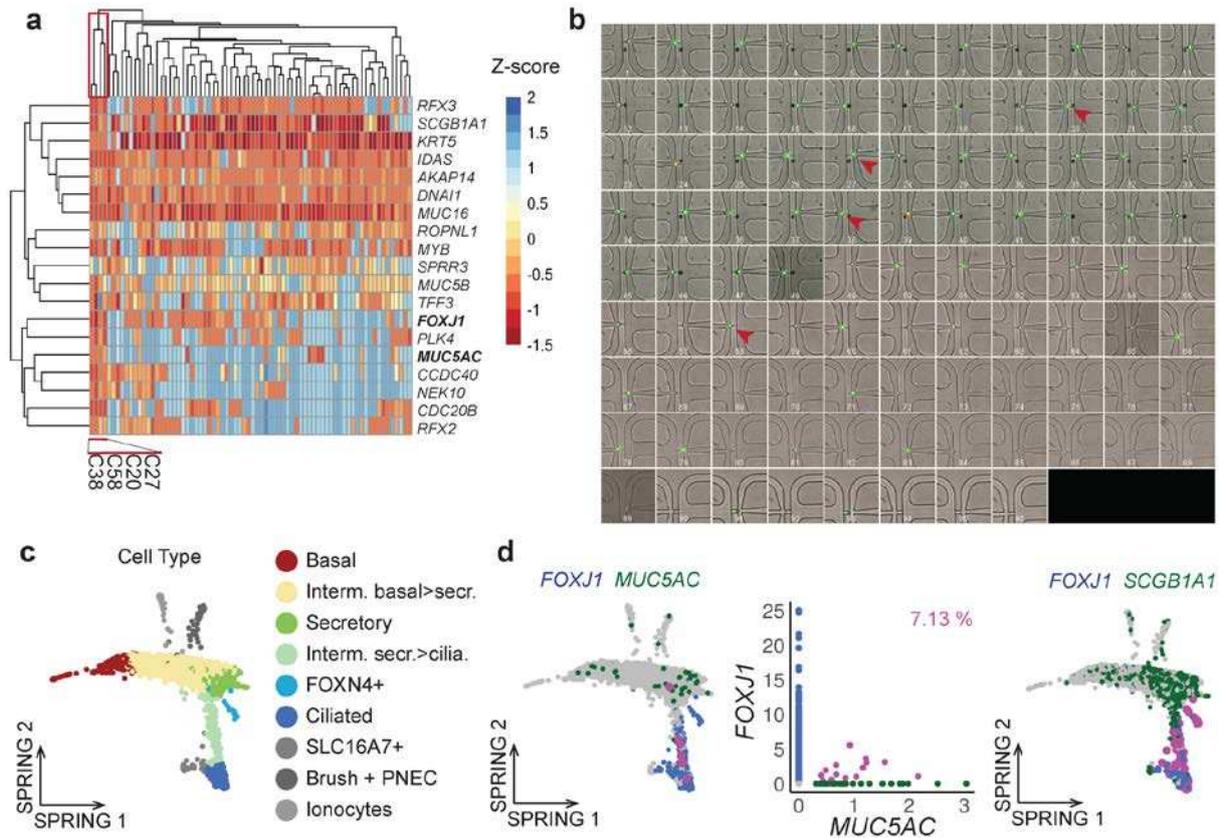
**Supplementary figure 3: Characterization of MCC cell lineages during airway epithelium regeneration using single cell RNA-Seq in BEGM medium.**

**a** t-SNE representing the distinct cell populations identified at each time points. **b** Relative abundance of cell types at each time point. **c** t-SNE plot of the aggregate of all cells from each time point. **d** Representation of the cell lineages inferred by Monocle 2 occurring during the upper airway epithelium regeneration (aggregate of all time points). Pseudotime evolution along the differentiation trajectory shown by white to grey gradient. **e** Distribution of the defined cell types in pseudotime. **f** Heatmap representing the smoothed temporal expression pattern of indicated cell type specific marker genes on MCC trajectory. Cells were ordered by cluster appearance in pseudotime.



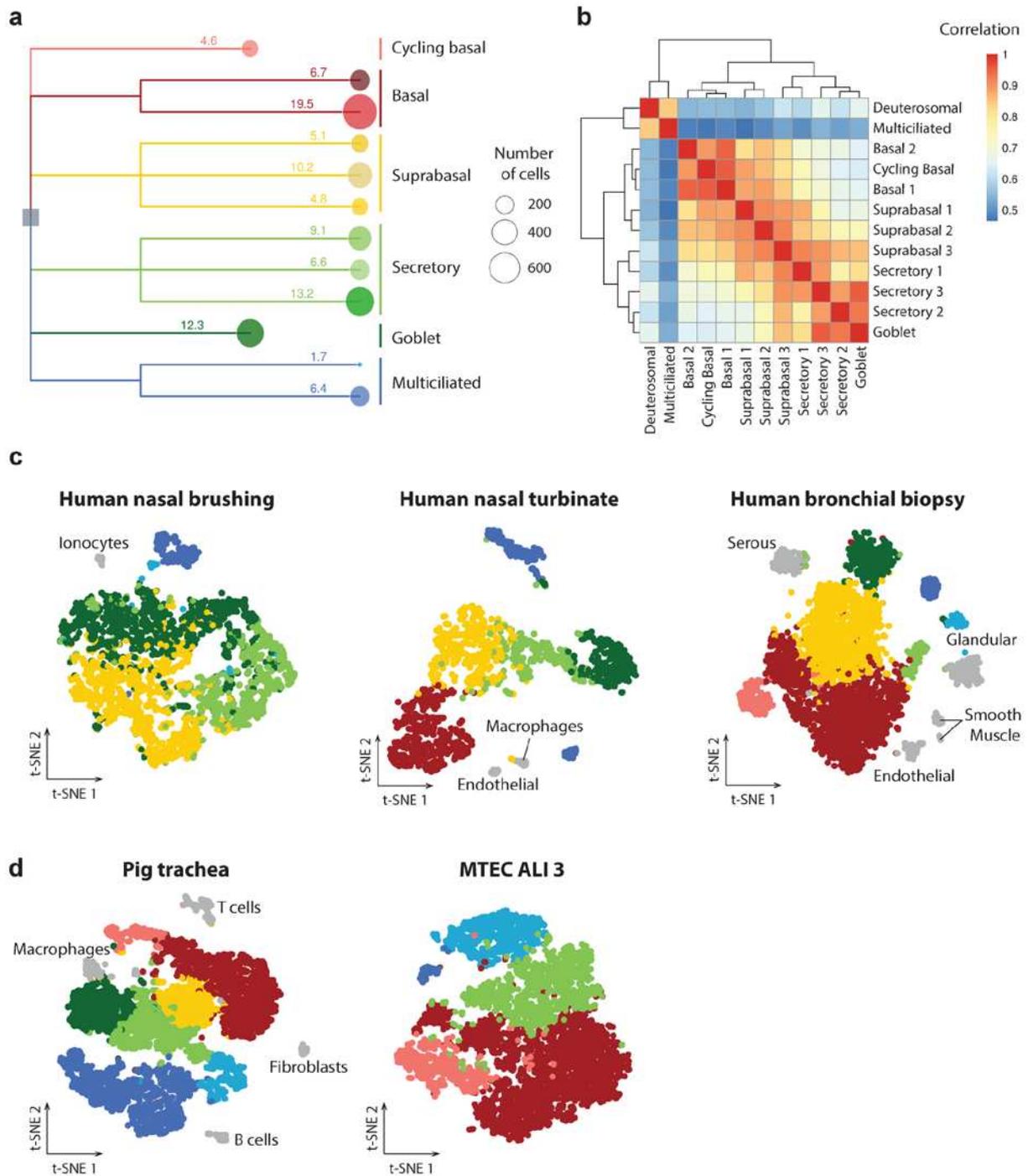
**Supplementary figure 4: Comparison of fully differentiated epithelia cell population between Pneumacult and BEGM media.**

**a** Sankey Network of the mapping of Pneumacult cells onto BEGM cells. **b** Sankey Network of the mapping of BEGM cells onto Pneumacult cells.



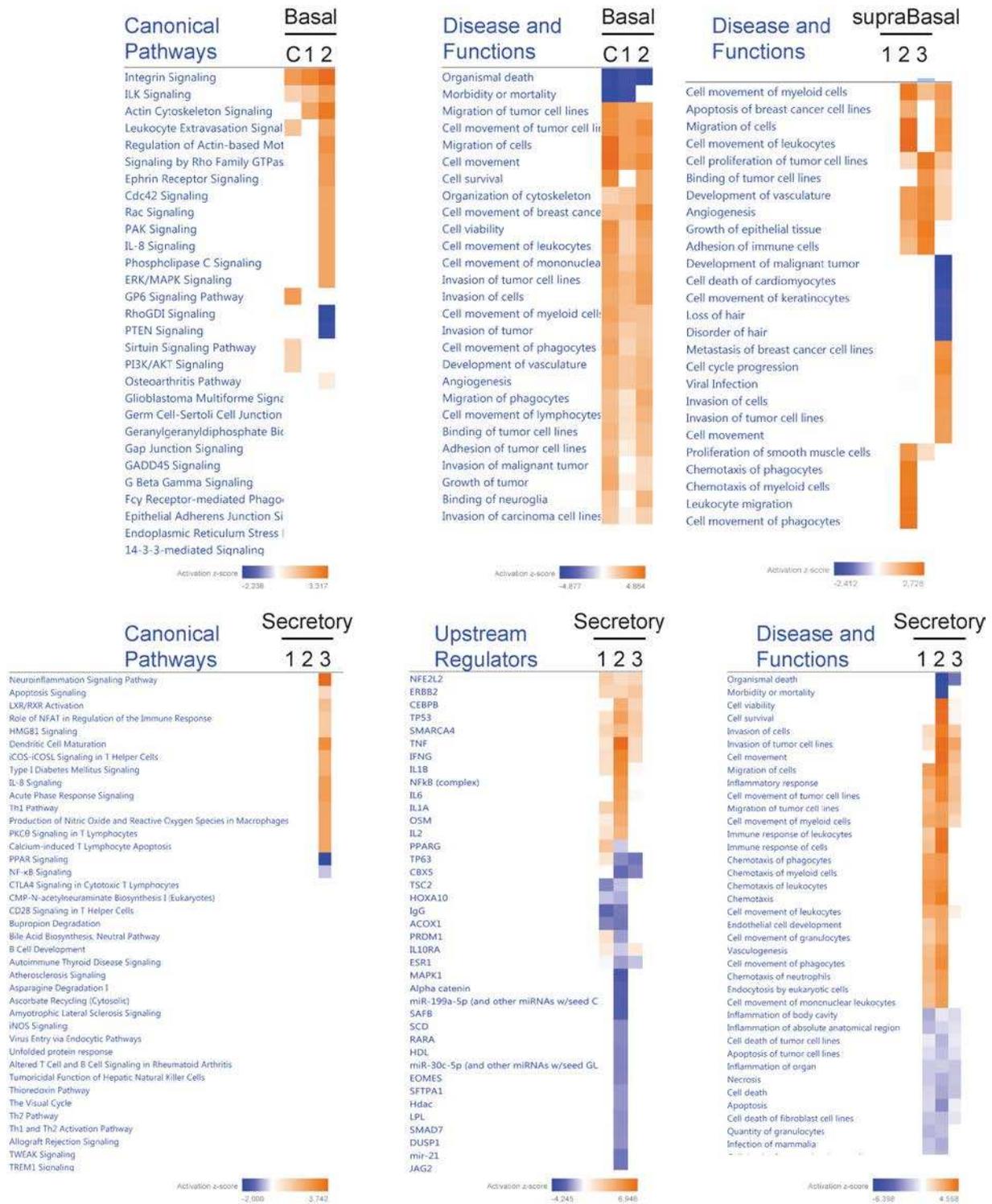
**Supplementary figure 5: Validation of Goblet cells markers in multiciliated cells in doublet free single cell transcriptomic datasets.**

**a** Heatmap colored by Z-score from C1 Biomark experiment in PneumaCult media, co-expression of Goblet and Multiciliated cell markers in single cells (red cluster). **b** C1™ Single-Cell Preamp IFC (10–17  $\mu$ m) imaging, Red arrowheads show cells expressing both cell type markers (*FOXJ1* and *MUC5AC*), one cell per chamber, green cells=living cells, red cells=dead cells. **c** SPRING representations of Plasschaert *et al.* dataset. Left panel colored by cell type. Center left panel colored by *FOXJ1*+ cells (blue), *MUC5AC*+ cells (green), co-expressing *FOXJ1* and *MUC5AC* cells (pink). Center right panel displays a scatter plot of normalized expression of *MUC5AC* and *FOXJ1* in cells (dot) colored by *FOXJ1*+ cells (blue), *MUC5AC*+ cells (green), co-expressing *FOXJ1* and *MUC5AC* cells (pink). Right panel colored by *FOXJ1*+ cells (blue), *SCGB1A1*+ cells (green), co-expressing *FOXJ1* and *SCGB1A1* cells (pink).



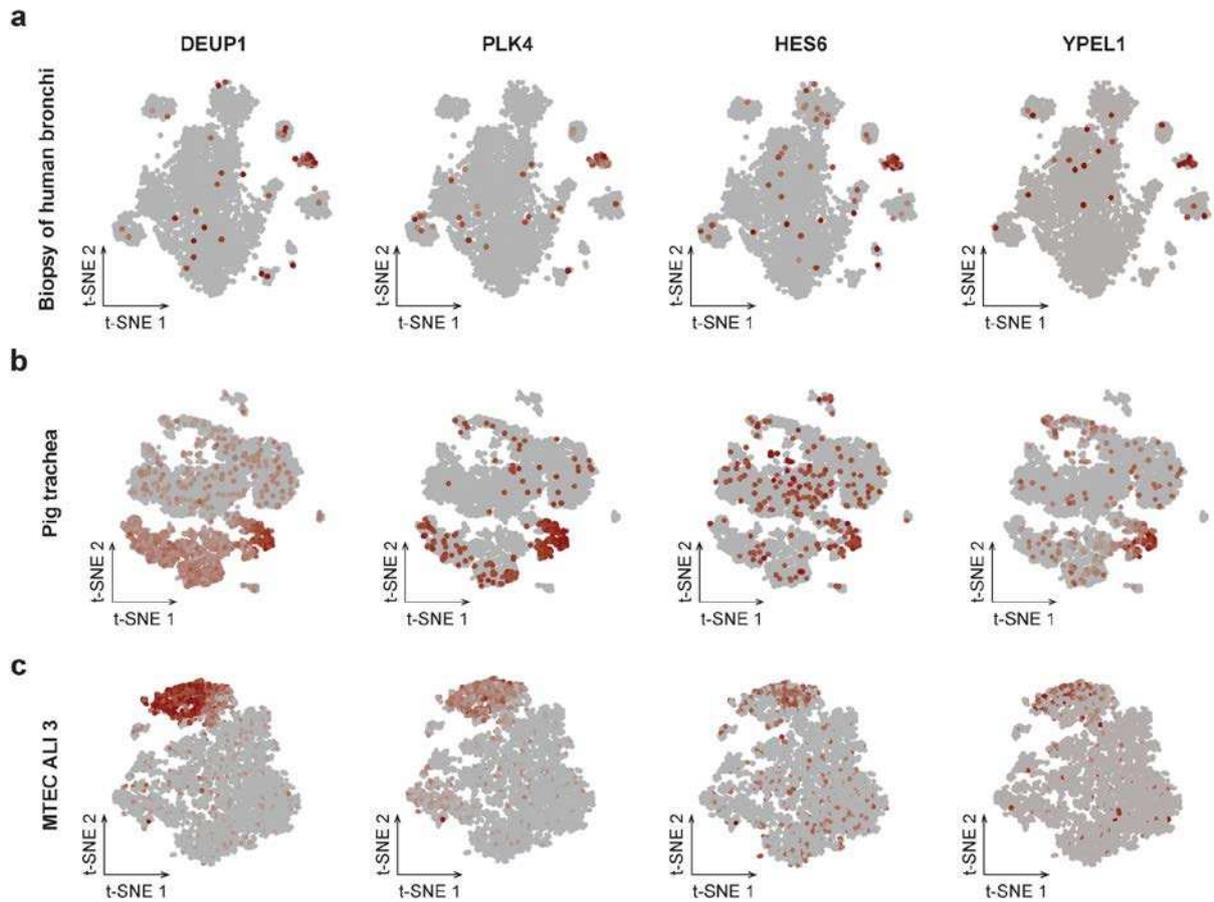
**Supplementary figure 6: Identification of subpopulations by high resolution cell clustering**

**a** Dendrogram of cell types identified through robust and high-resolution clustering of scRNA-seq data from Pneumacult ALI28 sample. **b** Correlation matrix of cell type identified in Pneumacult sample. **c** t-SNE representation of scRNA-seq data from human nasal brushing (left), nasal turbinate (center), bronchial biopsy (right) colored by cell type described previously (Fig. 1), in grey are rare or non-epithelial cell types. **d** t-SNE representation of scRNA-seq data from pig trachea (left) and mouse culture (right) colored by cell type previously described (Fig. 1).



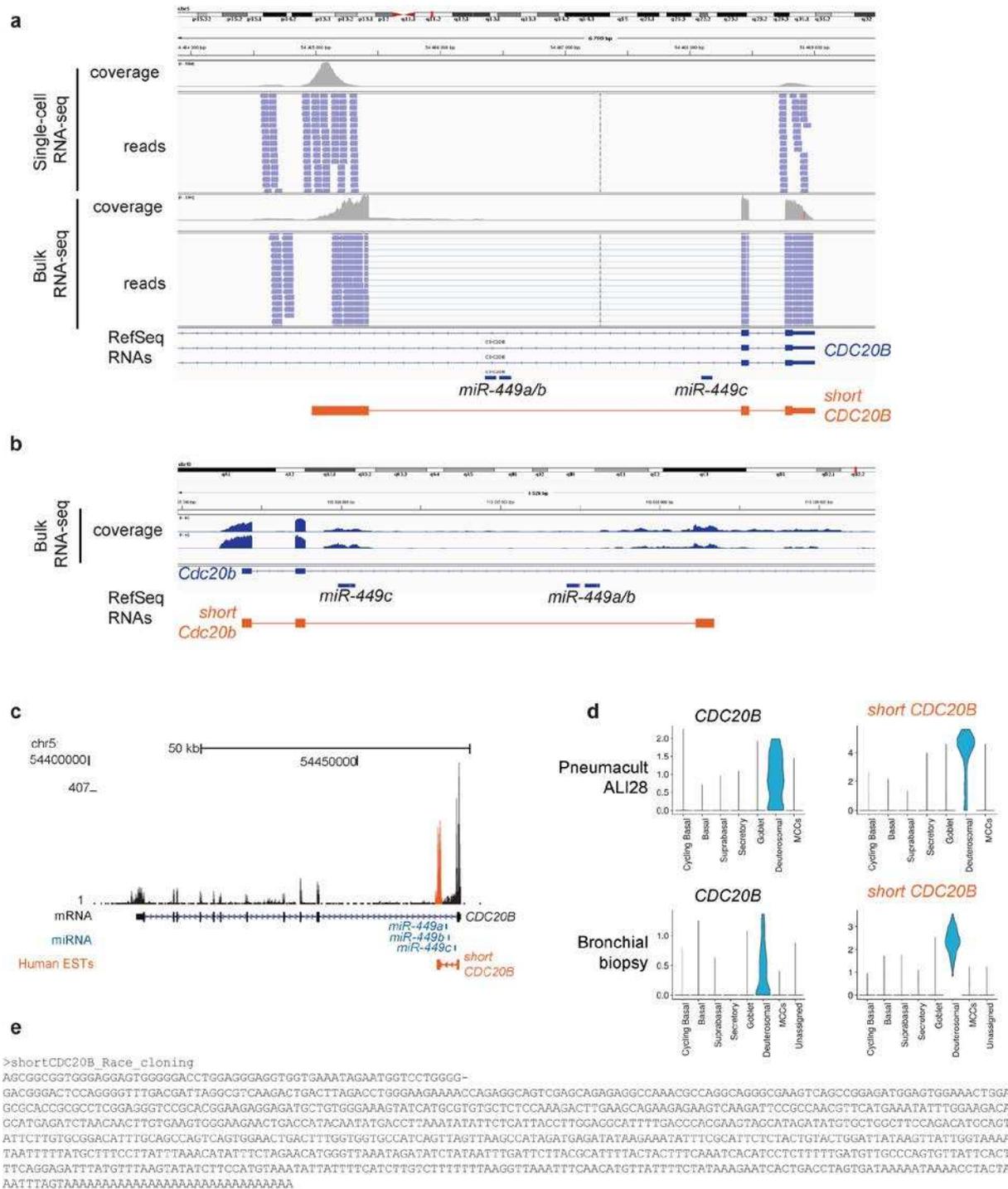
**Supplementary figure 7: Pathway enrichment comparison between basal, suprabasal and secretory cell subtypes**

Identification, with Ingenuity Pathway Analysis (Qiagen) of the enriched canonical pathways, upstream regulators and disease and functions in each of the basal and secretory subpopulations, based on an analysis of 300 differentially expressed genes (FDR<0.01 and log<sub>2</sub>FC>0.6 for BCs and SCs, and log<sub>2</sub>FC>0.4 for supraBCs). C: cycling basal cells.



**Supplementary figure 8: Robustness of Deuterosomal cells marker genes**

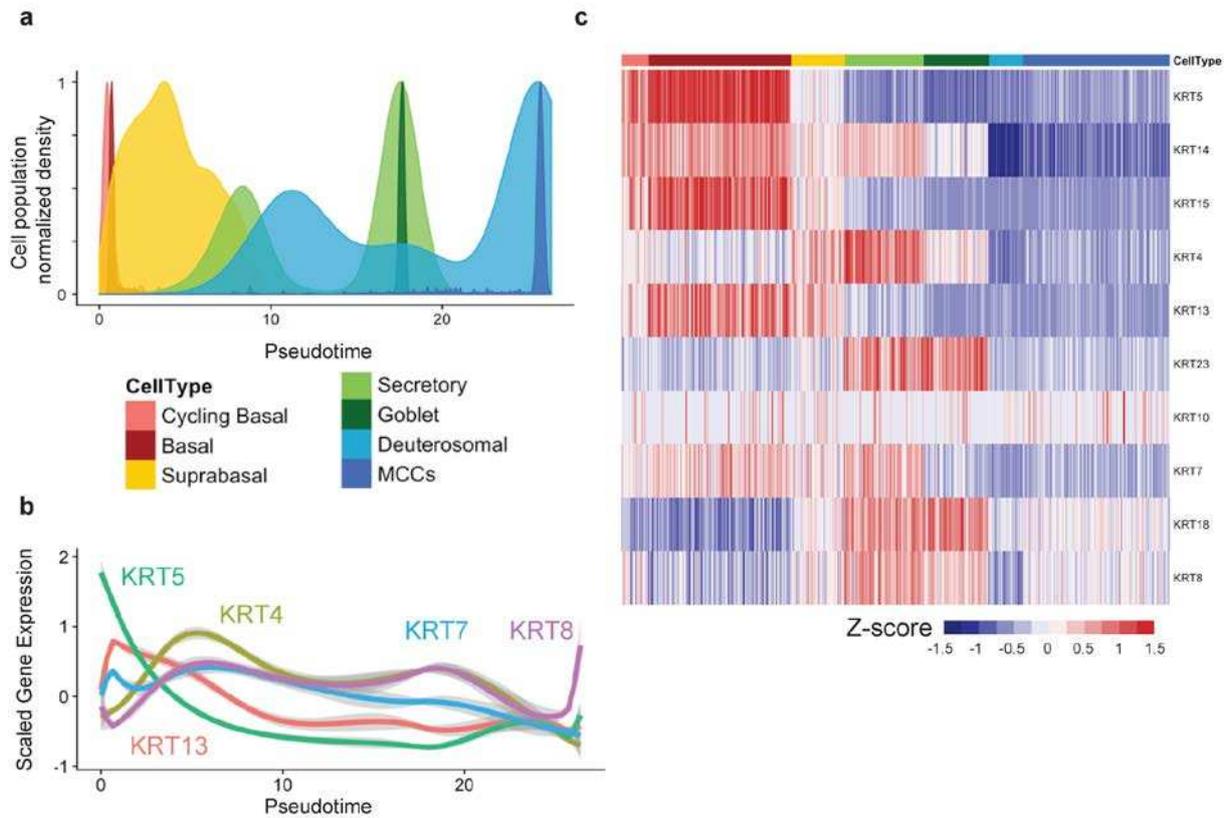
**a, b, c** t-SNE representation of Deuterosomal cell population marker genes expression (lowly to highly expressed, grey to red) in (a) Biopsy of human trachea, (b) Pig trachea, (c) MTECs ALI 3.



**Supplementary figure 9: Identification of short CDC20B, a novel isoform of a deuterosomal cell population marker gene.**

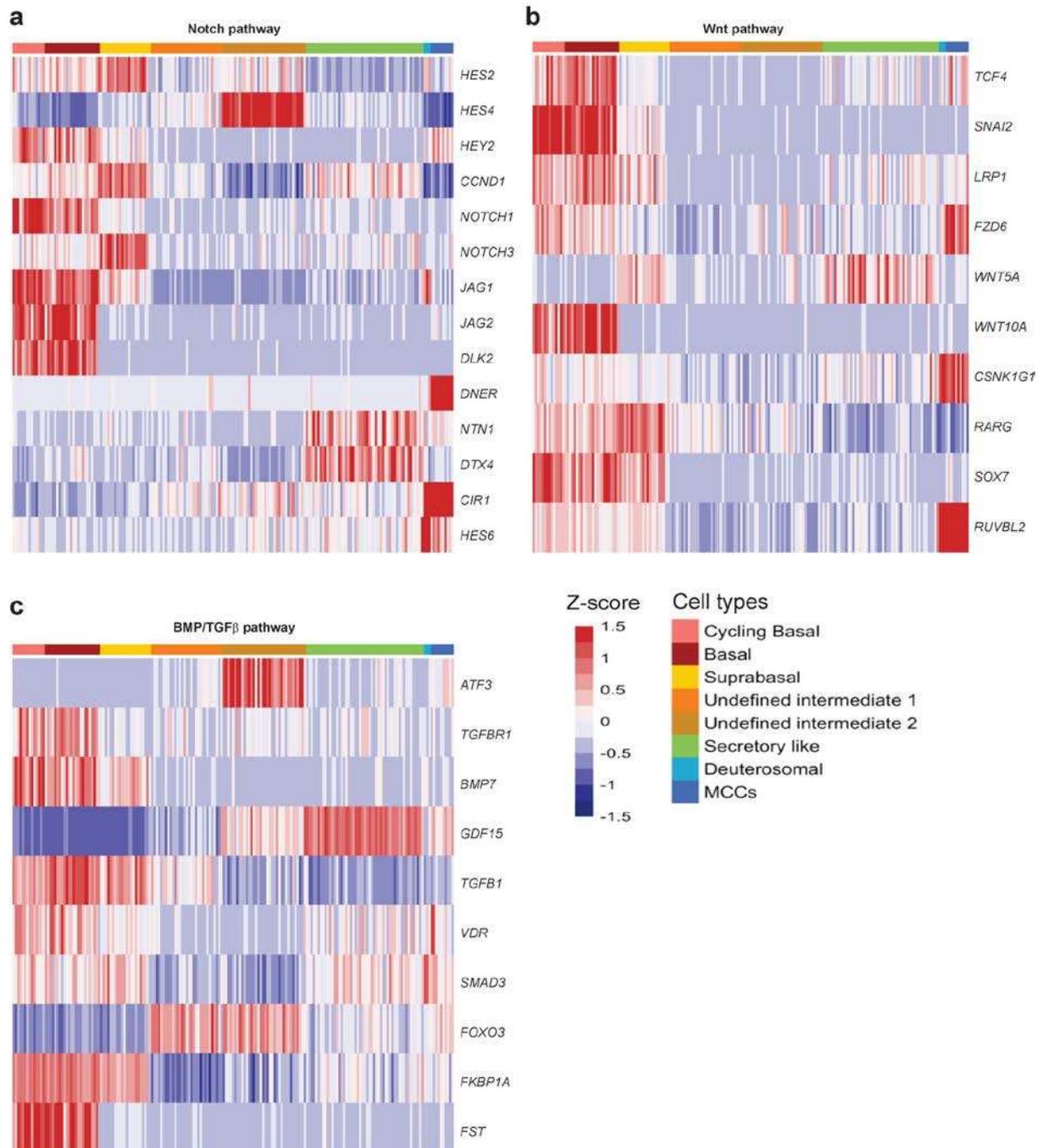
**a** Integrated Genome Viewer (hg19) view of the CDC20B gene with coverage and read alignment from bulk RNA-seq of well-differentiated HAECs and single-cell RNAseq of ALI14 differentiated HAECs. **b** Genome Viewer (mm10) view of the CDC20B gene with coverage from bulk RNA-seq of ALI 7 differentiated MTECs from the public dataset GSE75715. **c** UCSC Genome browser (hg19) view of the CDC20B gene with read alignment from bulk RNA-seq of ALI28 differentiated HAECs.

The alternative 3rd exon is shown in orange. **d** Violin plots for CDC20B and short CDC20B abundance in the Pneumacult ALI28 and the bronchial biopsies samples. **e** Alternative human short CDC20B sequence identified with 5' RACE cloning.



**Supplementary figure 10: Keratin expression in scRNA-seq from pig tracheal cells**

**a** Distribution of the 7 main cell types in the pseudotime from scRNA-seq of pig tracheal cells. **b** Plot of normalized gene expression of selected keratins according to pseudotime. **c** Heatmap representing the smoothed temporal expression pattern of regulated keratins. Cells were ordered by cluster appearance.



**Supplementary figure 11: Single-cell expression of signaling pathways components during airway regeneration in BEGM medium**

**a** Heatmap of the genes related to the NOTCH pathway with cells ordered by cluster. **b** Heatmap of the genes related to the WNT pathway with cells ordered by cluster. **c** Heatmap of the genes related to the BMP/TGF $\beta$  pathway with cells ordered by cluster.

## Supplementary tables

**Supplementary table 1: Differential Gene expression of Deuterosomal cluster vs. All clusters**

gene	avg_logFC	pct.1	pct.2	p_val_adj
CDC20Bshort	4.29173241	0.895	0.106	4.10E-87
CCNO	2.47379533	0.86	0.324	7.42E-28
HES6	2.18385143	0.965	0.112	8.70E-97
C5orf49	2.05693783	0.877	0.24	5.15E-36
PSIP1	2.01390765	0.982	0.205	1.19E-61
CCDC34	1.9809765	0.895	0.335	6.78E-32
KIF9	1.94960808	0.947	0.349	4.93E-34
TPPP3	1.92875042	0.825	0.308	4.35E-23
ROPN1L	1.89011241	0.789	0.206	1.05E-30
PSENN	1.86666232	0.965	0.624	2.30E-23
PCM1	1.84248476	1	0.64	6.52E-32
CCDC74A	1.84041155	0.825	0.211	1.13E-34
TUBB4B	1.81532629	0.965	0.795	1.36E-19
FOXJ1	1.80998137	0.947	0.097	7.16E-97
C20orf85	1.79683265	0.86	0.379	4.73E-19
C1orf192	1.73404906	0.86	0.136	1.11E-54
C11orf88	1.7232126	0.754	0.192	9.00E-30
ZMYND10	1.71516915	0.807	0.136	6.82E-50
RSPH9	1.71414237	0.754	0.12	4.12E-48
LRR1Q1	1.71044455	0.912	0.15	6.63E-58
HSP90AA1	1.70454418	1	0.997	2.63E-21
WDR54	1.67122137	0.702	0.215	1.25E-22
EFCAB1	1.65220289	0.754	0.156	3.48E-36
TUBA1B	1.64767051	1	0.692	2.18E-27
FAM183A	1.64248298	0.807	0.253	4.34E-25
C9orf9	1.6414399	0.772	0.133	2.01E-46
FAM229B	1.63041554	0.877	0.402	8.13E-21
C9orf116	1.62919072	0.93	0.283	2.56E-33
TCTEX1D2	1.62395152	0.93	0.496	5.39E-22

NUCB2	1.61638026	0.982	0.674	4.85E-28
MNS1	1.6004758	0.807	0.103	2.81E-63
MUC12	1.59921724	0.667	0.086	7.39E-51
NUDC	1.59651915	0.965	0.648	2.25E-24
NDUFAF3	1.59393536	0.965	0.712	3.13E-20
EVL	1.57890572	0.667	0.307	7.39E-11
RPL39L	1.56640794	0.772	0.228	1.49E-29
ZDHC1	1.51682303	0.579	0.108	8.37E-29
CCDC19	1.51575188	0.807	0.102	1.86E-64
RIBC2	1.51195731	0.772	0.079	2.82E-76
CCDC104	1.51138797	0.842	0.459	1.18E-18
CAPSL	1.50462124	0.737	0.2	1.03E-25
CCDC146	1.5016302	0.807	0.264	9.31E-25
BTG3	1.47849706	0.912	0.486	7.95E-24
CALM3	1.47740454	0.912	0.704	3.29E-18
CEP78	1.47525634	0.842	0.125	2.65E-63
RP11-620J15.3	1.46515404	0.789	0.201	6.34E-31
C22orf15	1.44376914	0.702	0.139	1.86E-32
ODF2L	1.44349845	1	0.663	1.77E-25
CCDC170	1.4402005	0.754	0.142	4.32E-37
RAMP1	1.42617804	0.947	0.324	1.54E-29
C1orf194	1.41626622	0.719	0.187	6.75E-24
RABL5	1.41085988	0.895	0.411	1.59E-23
RSPH1	1.40994819	0.86	0.302	6.72E-24
HSPH1	1.39112603	0.86	0.484	9.44E-18
PIFO	1.38292969	0.86	0.198	1.19E-36
RGCC	1.37321897	0.702	0.428	2.17E-08
HIST1H1C	1.36241342	0.842	0.63	2.97E-10
RP11-356K23.1	1.35863787	0.789	0.297	1.14E-17
RUVBL2	1.35693629	0.877	0.361	1.01E-24
BPIFB1	1.34577944	0.965	0.712	9.42E-16

CBY1	1.34356571	0.86	0.436	3.44E-20
IDH2	1.34082721	0.912	0.482	6.90E-22
DYX1C1	1.3387954	0.754	0.131	1.31E-43
CCDC181	1.33688299	0.789	0.124	2.37E-52
CCNA1	1.3354672	0.544	0.175	3.80E-13
FAM81B	1.33244038	0.737	0.098	3.55E-51
CALM2	1.32413545	1	0.983	8.24E-22
NPHP1	1.3177589	0.842	0.136	1.11E-55
IQCG	1.30701982	0.789	0.17	1.34E-36
ENKUR	1.29320702	0.772	0.103	1.47E-54
PTGES3	1.2881322	0.947	0.704	2.56E-22
B9D2	1.28680325	0.754	0.203	3.48E-28
C21orf59	1.28678643	0.93	0.429	1.85E-25
HSPA1A	1.28654824	0.737	0.393	4.02E-12
LRRC26	1.28611747	0.807	0.156	4.97E-44
CDK2AP2	1.27190382	0.86	0.522	1.43E-15
CARHSP1	1.27014247	0.947	0.542	8.33E-21
AKR7A2	1.25970564	0.86	0.549	1.18E-16
MORN2	1.25529358	0.947	0.579	4.71E-21
IFT27	1.25311043	0.807	0.457	3.00E-14
MLF1	1.24180441	0.789	0.17	6.30E-37
C11orf74	1.24151256	0.772	0.27	1.94E-21
TNFAIP8L1	1.23511791	0.807	0.119	4.50E-53
IFT57	1.23511155	0.895	0.435	2.20E-20
PPP1R2	1.22361833	0.807	0.502	3.22E-14
SPA17	1.22298091	0.842	0.255	5.24E-28
REPIN1	1.21730845	0.702	0.207	1.75E-22
TMEM106C	1.21602546	0.912	0.34	5.84E-30
CAPS	1.21510639	0.93	0.596	3.80E-17
CENPF	1.20650472	0.737	0.051	1.56E-100
AGR3	1.2009874	0.684	0.158	2.91E-25
DNALI1	1.19420228	0.789	0.144	7.65E-42
COPRS	1.19044873	0.895	0.428	1.51E-23

DNAAF1	1.18760976	0.684	0.087	1.39E-45
DNAH12	1.18723425	0.684	0.077	2.35E-55
P4HTM	1.17749297	0.807	0.208	7.23E-33
CDS1	1.1651768	0.789	0.304	3.14E-19
CCDC176	1.16312236	0.825	0.107	1.13E-60
CCDC67	1.16206565	0.561	0.026	1.26E-104
DYNLL1	1.16095038	1	0.998	2.61E-18
PPIL6	1.15746657	0.719	0.11	6.89E-43
SPAG6	1.15571312	0.649	0.077	7.90E-51
NEK2	1.14382991	0.667	0.01	1.37E-253
C14orf142	1.12998563	0.737	0.281	1.73E-16
DPCD	1.11695438	0.86	0.418	3.67E-18
MRPL43	1.11527385	0.807	0.743	6.40E-05
IFT43	1.11522388	0.825	0.491	3.92E-14
C11orf70	1.11311121	0.702	0.141	8.27E-33
CKS1B	1.10982697	0.947	0.516	6.97E-25
SMC4	1.10960774	0.807	0.275	3.38E-24
ENDOG	1.10892823	0.737	0.29	2.13E-16
IK	1.10174504	0.877	0.562	3.67E-15
CDC20	1.08621046	0.649	0.028	1.21E-130
SPAG1	1.08478524	0.842	0.218	6.73E-33
TRAF3IP1	1.08278447	0.737	0.231	1.50E-22
EFHC1	1.07961279	0.842	0.222	2.65E-30
EIF2S2	1.07650973	0.965	0.94	4.59E-20
TEKT1	1.07591039	0.649	0.11	4.64E-35
TUBA1A	1.07269853	0.754	0.401	4.14E-10
SPATA33	1.07039339	0.86	0.198	2.55E-40
C7orf57	1.06940792	0.596	0.053	4.15E-62
TXNRD1	1.06515922	0.842	0.368	2.66E-20
PLCE1	1.06316626	0.719	0.155	1.19E-31
LRRC23	1.06274825	0.789	0.203	1.03E-27
HSPA8	1.0625061	0.982	0.897	2.30E-15
MAP1A	1.06179327	0.649	0.074	5.78E-52

MUC1	1.0600729	0.965	0.66	1.18E-15
HSPBP1	1.05799385	0.789	0.255	1.47E-24
SPEF1	1.0573456	0.754	0.086	5.38E-62
CNTRL	1.05449509	0.702	0.151	8.76E-31
TMEM14B	1.05317828	0.947	0.846	2.30E-20
ANKRD36C	1.05092894	0.754	0.299	4.53E-14
YWHAH	1.05078326	0.842	0.468	1.17E-14
H2AFZ	1.05056193	1	0.977	8.47E-26
FANK1	1.04532129	0.702	0.143	3.13E-31
KLHDC9	1.0422098	0.702	0.141	7.43E-33
LRRC46	1.04043533	0.579	0.11	2.40E-26
RANBP1	1.04028579	0.965	0.818	7.39E-22
C9orf135	1.03949632	0.684	0.143	1.23E-27
SNTN	1.03863809	0.596	0.174	1.22E-14
CCDC171	1.0350885	0.684	0.071	1.95E-64
HIST1H2BJ	1.0304779	0.702	0.071	1.98E-70
HSPE1	1.02651432	0.982	0.973	1.98E-17
CDC20B	1.02530067	0.702	0.006	0
LINC01171	1.02417279	0.719	0.122	9.43E-37
H2AFV	1.0211016	0.982	0.775	3.50E-23
CHORDC1	1.01679794	0.772	0.274	1.39E-21
ARMC3	1.01478908	0.719	0.073	3.25E-65
SPAG16	1.01355423	0.825	0.453	3.56E-14
DPY30	1.0099647	0.895	0.793	2.62E-11
MRPL18	1.00932664	0.947	0.776	5.95E-13
CCP110	1.00845073	0.789	0.141	3.63E-43
CEP41	1.00636513	0.684	0.149	7.89E-31
CETN2	1.00537082	0.842	0.553	8.17E-13
ZNF487	1.00212949	0.684	0.101	4.78E-43
DNAJA1	1.00011384	0.93	0.663	3.22E-15

**Supplementary table 2: Primers used in the Biomark qRT-PCR**

<b>Gene</b>	<b>Sequence of forward primer</b>	<b>Sequence of reverse primer</b>
AKAP14	AGATTGTGGAAGAGGAGCGAAACCC	TACCATGCTGGAGTGGTCGGC
CCDC40	CATCCCACGGAGTCTTAGGC	TGGATCCTGTCAATCTGCC
CDC20B	CGGCTGAGAAATATGCTTGG	ATAAACACGCCCCAGTCTTG
DNAI1	AACGACGGCTGTCCCTAAAG	AGCCTACAAAACGCTCCCTC
FOXJ1	TGGATCACGGACAACCTTCTG	GAGGCACTTTGATGAAGCAC
IDAS	TTTCAGAGACACGGTGGATG	TGGTGATATGTCGCAAGGAG
KRT5	AGGAATGCAGACTCAGTGGAG	CAGAGGAAACACTGCTTGTGAC
MUC16	GTCAAGCCAGGCAGACAAGG	GGGATGTGCTGCTGGACTGC
MUC5AC	CCAAATACGCCAACAAGACC	ATTCCATGGGTGTCAGCTTG
MUC5B	TGCAACCGTCCC GGCTTCG	GTTGTGTTGCACACGCACACCG
MYB	ATGATGAAGACCCTGAGAAGGAAA	AACAGGTGCACTGTCTCCATGA
NEK10	GCAAGAAATCACCATCAGGGAC	GGCTGGAAGCTGTTGTTTGC
PLK4	TCCAACACAGGCACCAATC	GAGATGTCTGTTCCAGAAGCTG
RFX2	GCGACCACATCCTCTACCAG	AGTTACGGATGGCCTGTGTC
RFX3	AGCCAACATCATCAACAGTTTTT	TGCAGTGACTTGATATCCTCAA
ROPN1L	CTCGCATCCCCTTCAAGACG	TCTTCCTGGCGTCTATATTTCC
SCGB1A1	AGAGACGGGCCAGAGCATCCC	GGCAGATCTCTGCAGAAGCGGAGC
SPRR3	TCTGCACAGCAGGTCCAGCATC	AGGCTGGCTGGGTTGTTTCACC
TFF3	GGAGTACGTGGGCCTGTCTGC	AGCCCCGGTTGTTGCACTCC

