



La différenciation érythrocytaire : caractérisation moléculaire de l'engagement cellulaire

Souad Zreika

► To cite this version:

Souad Zreika. La différenciation érythrocytaire : caractérisation moléculaire de l'engagement cellulaire. Biologie cellulaire. Université de Lyon; École doctorale des Sciences et de Technologie (Beyrouth), 2021. Français. <NNT : 2021LYSEN070>. <tel-03576339>

HAL Id: tel-03576339

<https://theses.hal.science/tel-03576339v1>

Submitted on 16 Feb 2022

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.



HAL Authorization



Numéro National de Thèse : 2021LYSEN070

THESE DE DOCTORAT DE L'UNIVERSITE DE LYON

opérée par
l'Ecole Normale Supérieure de Lyon
en cotutelle avec
l'Université Libanaise

Ecole Doctorale N° 340
Biologie Moléculaire Intégrative et Cellulaire

Discipline : Sciences de la vie et de la santé
Spécialité de doctorat : Biologie

Soutenue publiquement le 30/11/2021, par :
Souad ZREIKA

La différenciation érythrocytaire : caractérisation moléculaire de l'engagement cellulaire

Devant le jury composé de :

Commes-Maerten, Thérèse	PU	INSERM Univ. de Montpellier	Rapporteuse
Jaffredo, Thierry	DR	CNRS UPMC	Rapporteur
Gandrillon, Olivier	DR	CNRS ENS de Lyon	Examineur
Bazarbachi, Ali	PR	American Univ. of Beirut	Examineur
El Chamy, Laure	DR	Université St Josep, Beirut	Examinatrice
Giraud, Sandrine	MCU	Université Claude-Bernard Lyon1	Directrice de thèse
El Khatib, Mazen	PU	Université Libanaise	Co-Directeur
Al Safadi, Rim	DR	Université Libanaise	Co-encadrante

Remerciements

Avant tout développement sur cette expérience, il apparaît opportun de commencer ce rapport de thèse par des remerciements, à ceux qui ont eu la gentillesse de faire de cette thèse un moment inoubliable.

Je tiens tout d'abord à adresser mes plus sincères remerciements à Sandrine, qui m'a fait le grand honneur de bien vouloir m'encadrer dans ce travail. Depuis que j'ai commencé ma thèse, Sandrine m'a énormément aidée à avancer sur le plan professionnel mais aussi personnel. Elle a toujours été à mes côtés et m'a soutenue même dans les moments les plus durs. Elle m'a toujours encouragé à être fière d'avoir tenu sans jamais abandonner malgré les nombreux obstacles expérimentaux. Je tenais à la remercier de tout mon cœur, pour sa gentillesse et sa direction avisée, fidèle et exigeante à laquelle cette thèse doit beaucoup.

Je tiens aussi à remercier Olivier, qui m'a accueillie dans son équipe et qui m'a encouragée afin d'accomplir ce travail. Grâce à sa pédagogie et à son encouragement, j'ai énormément appris et ça a été pour moi une très grande opportunité de mener mon travail de recherche au sein de son équipe multidisciplinaire. Enfin, je tiens à le remercier pour sa gentillesse, sa disponibilité permanente et pour ses nombreux conseils.

Je remercie aussi tous les autres membres de l'équipe, plus particulièrement Elodie, pour sa sympathie et pour toute son aide technique durant toute la période de thèse. C'était vraiment très agréable et super enrichissant de travailler avec elle surtout pour les expériences de mon deuxième projet. Merci aussi à Camille, l'épaule sur laquelle j'ai pu toujours compter. J'ai eu beaucoup de chance de partager ces années de thèse avec elle. Merci pour tous ces beaux moments qui sont devenus inoubliables. Elle était si merveilleuse, si fidèle et toujours là pour moi. Merci aussi pour tous les services qu'elle m'a rendus. Un grand Merci à Gérard qui m'a toujours rassurée et m'a donné confiance en moi surtout pendant les moments les plus durs. Je

n'oublierais jamais ses mots, les moments de partage pendant les pauses café et nos échanges enrichissants.

Je remercie également les autres membres de l'équipe SBDM, dont Catherine, pour sa bonne surveillance et sa sympathie. Merci à Margaux, pour sa gentillesse et son aide quand j'avais besoin. Merci à Matteo avec qui le travail a été un vrai plaisir et très enrichissant. Merci également à Elias, toujours agréable, avec qui les discussions étaient toujours intéressantes. Je n'oublie pas de remercier Alexey, Christophe, Frank, Laurent et Vincent pour tous leurs conseils en mathématiques et en statistiques au cours de ma thèse, et les anciens membres de l'équipe, Mathilde et Arnaud pour les échanges scientifiques et tous les beaux moments.

Merci encore à Sébastien et Victor de la plateforme de cytométrie. J'ai pris beaucoup de plaisir en les rencontrant et en passant avec eux beaucoup de temps à trier les cellules mais aussi à échanger scientifiquement et culturellement. Je tiens à remercier également, les étudiants du LBMC et particulièrement Philippe, Léa, Laura et Loïc avec lesquels j'ai passé des soirées inoubliables. Merci aussi à Christophe, j'ai eu grand plaisir à faire sa connaissance et à partager différentes idées et cultures.

J'adresse mes remerciements aussi à mon directeur de thèse au Liban, Mazen et à ma codirectrice Rim, qui m'ont fait l'honneur de bien vouloir m'aider à trouver cette thèse et de suivre mon travail.

Mes vifs remerciements s'adressent aux membres du jury, Pr. Thérèse Commes-Maerten et Dr. Thierry Jaffredo qui ont accepté de lire et d'évaluer mon travail. Merci également au Dr. Laure Chamy et au Pr. Ali Bazarbachi qui ont accepté de participer à mon jury de thèse.

Merci à mes meilleurs amis rencontrés à Lyon, surtout Hana et Sabah pour tous les beaux moments partagés et pour avoir pris soin de moi. Je remercie aussi tous mes voisins et mes copains de Lyon de m'avoir toujours rappelé que la famille n'est jamais loin.

Un grand Merci à Adnan, mon futur mari et la personne avec laquelle je vais passer toute ma vie. Durant ces trois ans, il était toujours un support sans jamais hésiter. Il savait bien comment m'écouter, me consoler et surtout, me rendre heureuse.

Finalement, je veux adresser mes grands remerciements à mes parents, à ma sœur Rim et mon frère Abdul pour l'amour et pour leur grand soutien tout au long de ma thèse. Merci également à la famille de Adnan, et à ma cousine Mona, d'être toujours à mon écoute et pour toutes leurs aides. Vous avez été incroyables !

Résumé

Chaque cellule d'un organisme doit faire des choix définissant son destin, choix entre prolifération, différenciation, mais aussi survie ou mort. Afin de mieux comprendre le processus de différenciation, notre équipe a précédemment réalisé des analyses d'expression génique par RT-qPCR en cellule unique (scRT-qPCR) sur des progéniteurs érythrocytaires primaires de poulet (T2EC). Ces études ont montré une augmentation de la variabilité de l'expression génique pendant les 24 premières heures du processus de différenciation, phase durant laquelle les cellules ont gardé la capacité de s'autorenouveler avant de s'engager irréversiblement en différenciation. De plus, l'application de l'outil WASABI sur les données de scRT-qPCR des T2EC en différenciation a permis d'inférer 364 réseaux de régulation de gènes (GRN) candidats qui sous-tendent ce processus. Dans ce contexte, l'objectif de ce travail de thèse a été de mieux caractériser moléculairement la dynamique de la différenciation des T2EC et de définir le GRN qui décrit au mieux ce processus. Ainsi, dans la première partie de ce travail, nous avons analysé par scRT-qPCR et MARS-seq, le transcriptome des T2EC induites à se différencier pendant 24h puis replacées en milieu d'autorenouvellement (T2EC en réversion). L'analyse des données d'expression par différents outils bioinformatiques, a montré une très forte similarité entre les T2EC en réversion et les T2EC constamment maintenues en état d'autorenouvellement. Néanmoins, les cellules en réversion ont gardé une marque de leur engagement dans le processus de différenciation en activant l'expression de gènes liés à la voie de synthèse de l'hème. Dans la deuxième partie de ce travail, nous avons cherché à générer de nouvelles données d'expression à l'échelle de la cellule unique afin de définir au mieux le GRN sous-jacent au processus de différenciation des T2EC. Dans ce but, nous avons réalisé des expériences de Knock Out (KO) du gène *FNIP1* (Folliculin Interacting Protein 1) dont le positionnement est le plus incertain entre les réseaux obtenus avec l'outil WASABI. A l'échelle de la population cellulaire, ce KO affecte la prolifération des T2EC mais n'a pas d'effet notable sur leur différenciation. A l'échelle de la cellule unique, nous avons pu établir un protocole strict de validation au niveau génomique et transcriptomique du KO et obtenir récemment 2 clones KO pour *FNIP1* en cours d'analyse par scRT-qPCR. En conclusion, ce travail a permis de mieux caractériser la différenciation érythrocytaire et la réversibilité de destin des T2EC qui n'ont pas encore atteint un état stable de différenciation, processus dirigé par le comportement dynamique et stochastique d'un GRN sous-jacent dont la structure reste à définir.

Mots clés : T2EC, différenciation, Réversion, scRT-qPCR, MARS-seq, GRN, KO, *FNIP1*.

Abstract

During development, cells choose the most appropriate fate between proliferation, differentiation, survival or death. To better understand the differentiation process, our team has previously performed single cell RT-qPCR (scRT-qPCR) gene expression analyses on primary chicken erythrocyte progenitors (T2EC). These studies showed an increase in gene expression variability during the first 24 hours of the differentiation process, a phase during which the cells retained the ability to self-renew before irreversibly committing to differentiation. Furthermore, the application of the WASABI tool on scRT-qPCR data of differentiating T2EC allowed to infer 364 candidate gene regulatory networks (GRN) underlying this process. In this context, the aim of this work consists in better molecularly characterizing the dynamics of T2EC differentiation and defining the GRN that best describes this process. In the first part of this work, we analyzed by scRT-qPCR and MARS-seq, the transcriptome of T2EC induced to differentiate for 24 hours and then put back in self-renewal medium (reversed T2EC). The analysis of the expression data by different bioinformatics tools, showed a strong similarity between reversed T2EC and T2EC constantly maintained in self-renewal state. Moreover, reversed cells retained a mark of their involvement in the differentiation process by activating the expression of genes related to the heme synthesis pathway. In the second part of this work, we sought to generate new expression data at the single cell level to better define the GRN underlying the T2EC differentiation process. To this end, we performed Knock Out (KO) experiments of *FNIP1* (Folliculin Interacting Protein 1) gene whose positioning is the most uncertain between the networks obtained with the WASABI tool. At the population level, this KO affects T2EC proliferation but has no significant effect on their differentiation. At the single cell level, we have been able to establish a strict protocol for KO validation at the genomic and transcriptomic level and recently obtained 2 KO clones for *FNIP1* that are being analyzed by scRT-qPCR.

In conclusion, this work has allowed us to better characterize the erythrocytic differentiation and fate reversibility of T2EC that have not yet reached a stable state of differentiation. Those processes are driven by the dynamic and stochastic behavior of an underlying GRN whose structure remains to be defined.

Keywords: T2EC, differentiation, reversion, scRT-qPCR, MARS-seq, GRN, KO, FNIP1.

Table des matières

Table des matières	7
Liste des figures.....	10
Liste des tableaux.....	15
Liste des abréviations	16
1. Introduction bibliographique	19
1.1 Différenciation cellulaire : les deux visions du processus	19
1.1.1 Différenciation cellulaire dans la vision classique	23
1.1.1.1 Un processus linéaire et unidirectionnel	23
1.1.1.2 Différenciation érythrocytaire dans la vision classique	23
1.1.2 Différenciation cellulaire dans la vision moderne	26
1.1.2.1 Stochasticité de l'expression des gènes (SGE).....	27
1.1.2.2 Différentes méthodes d'analyse de la variabilité de l'expression génique.....	29
1.1.2.3 SGE et dynamique de la différenciation cellulaire	30
1.1.2.4 Différenciation érythrocytaire dans la vision moderne	33

1.2 Analyses du transcriptome à l'échelle de la cellule unique.....	35
1.2.1 RNAseq en cellule unique (scRNA-seq).....	36
1.2.1.1 Principe général du RNA-seq	36
1.2.1.2 ScRNA-seq : les différentes étapes.....	37
1.2.1.3 Techniques de scRNA-seq les plus couramment utilisées	48
1.2.1.4 Comparaison des différentes technologies de scRNA-seq.....	52
1.2.2 RT-qPCR en cellule unique	53
1.3 Approches bio-informatiques pour l'analyse des données d'expression obtenues en cellule unique.....	55
1.3.1 Méthodes de réduction de dimensions	55
1.3.2 Approche du Différentiel d'Expression (DE).....	57
1.3.3 Approche Sparse PLS	58
1.3.4 Approche des Réseaux de Régulation des Gènes (GRN)	59
1.3.4.1 Définition des GRN	59
1.3.4.2 Importance et rôles des GRN	60
1.3.4.3 Structure des GRN	61

1.3.4.4 Inférence des GRN.....	62
1.3.4.5 Inférence des GRN par l’algorithme itératif : WASABI	66
1.3.4.6 Stratégie de Design of Experiment (DOE)	68
1.3.4.7 Application du DOE sur les réseaux candidats inférés par Wasabi à partir des données de différenciation érythrocytaire	69
1.4 Protéine d’interaction avec la Folliculine : FNIP1	71
1.4.1 Rôles de FNIP1 dans les processus biologiques	73
1.4.1.1 Régulation du métabolisme	73
1.4.1.2 Apoptose et cycle cellulaire	78
1.4.1.3 Activités de co-chaperonne	78
1.4.1.4 Réponse immunitaire	79
1.4.1.5 Différenciation	79
1.4.1.6 Suppression de tumeurs	80
2. Résultats	81
2.1 Caractérisation moléculaire de la réversion des cellules pré-engagées dans la différenciation érythrocytaire.....	83

2.1.1 Introduction	83
2.1.2 Article	86
2.2 La différenciation érythrocytaire : Du réseau de gènes à l'analyse fonctionnelle	125
2.2.1 Introduction	125
2.2.2 Matériels et Méthodes	127
.....	135
2.2.3 Résultats	138
2.2.3 Conclusions	156
3. Discussion et perspectives	157
Références	163

Liste des figures

Figure 1-Organisation de l'hématopoïèse adulte	22
Figure 2- Schéma de l'érythropoïèse selon la vision classique.....	24
Figure 3- Paysage épigénétique de Waddington	31
Figure 4- Principe général du RNA-seq	37
Figure 5- Grandes étapes du scRNA-seq	38
Figure 6- Différentes approches d'isolement et de capture des cellules uniques pour le scRNA-seq	39
Figure 7- Principe du Template switching	41
Figure 8- Principe de l'IVT	43
Figure 9- Principe des UMI	45
Figure 10-Principe du SPLIT-seq	47
Figure 11- Protocole de scRT-qPCR-Fluidigm	54
Figure 12- Exemple d'analyse de données de scRNA-seq par 3 techniques différentes de réduction de dimensions	57
Figure 13- Principe de prédiction par Sparse PLS	59
Figure 14- Réseau de Régulation de Gènes.....	61

Figure 15- Exemple d'un GRN candidat inféré par WASABI	67
Figure 16- Principe de la stratégie de DO.....	69
Figure 17- Incertitude sur le positionnement des gènes dans les 364 GRN candidats.....	70
Figure 18- Structure du gène <i>FNIP1</i> et de la protéine correspondante	71
Figure 19-Activation de mTORC1 par la voie PI3/AKT/TSC	75
Figure 20- Régulation de la voie de signalisation mTORC1	76
Figure 21-Caractérisation moléculaire de la dynamique de la différenciation érythrocytaire.	82
Figure 22- Stratégie expérimentale suivie pour la caractérisation moléculaire des T2EC en réversion	84
Figure 23- Plasmide pCRISPR-P2A-ARNt	127
Figure 24- Analyse de la délétion de <i>FNIP1</i> au niveau génomique induite par les vecteurs CRISPR-Cas9 double guides.....	134
Figure 25- Analyse de la délétion de <i>FNIP1</i> au niveau transcriptomique induite par les vecteurs CRISPR-Cas9 double guides.....	135
Figure 26-Analyse des délétions/mutations de <i>FNIP1</i> au niveau génomique induite par les vecteurs CRISPR-Cas9 simple guides	136
Figure 27- Analyse de la délétion de <i>FNIP1</i> au niveau transcriptomique induite par les vecteurs CRISPR-cas9 simple guides.....	137

Figure 28- Boxplots représentant la distribution de l'expression de <i>FNIP1</i> dans les données de RT-qPCR précédemment publiées (Richard et al., 2016) acquises sur les progéniteurs érythrocytaires induits à se différencier pendant à 0, 8,24 ,48 et 72h.	138
Figure 29-Les deux vecteurs CRISPR-Cas9 double guides (Vecteurs 1 et 2)	140
Figure 30- Validation de l'efficacité des vecteurs V1 et V2 au niveau génomique	141
Figure 31-Rapport du niveau d'expression de <i>FNIP 1</i> dans les cellules transfectées par le vecteur 1 ou 2 par rapport aux cellules contrôles	142
Figure 32-Mesure de taux de croissance des cellules KO pour <i>FNIP1</i>	143
Figure 33-Mesure du pourcentage de mortalité des cellules KO pour <i>FNIP1</i>	144
Figure 34-Mesure du pourcentage de cellules différenciées après le KO de <i>FNIP1</i>	145
Figure 35-Mesure de la croissance et de la mortalité des cellules KO pour <i>FNIP1</i> lors de la différenciation.....	146
Figure 36-Premier protocole de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 double guides et d'obtention de clones de cellules KO pour l'analyse transcriptomique à l'échelle de la cellule unique.....	148
Figure 37-Profiles attendus après analyse par PCR de la région ciblée par les deux guides et dépôt sur gel des amplifications des clones transfectés avec les vecteurs doubles guides	149

Figure 38-Différentes combinaisons possibles au niveau génomique et transcriptomique suite à la transfection des cellules avec les vecteurs CRISPR-Cas9 double guides.....	150
Figure 39-Protocole final de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 double guides et d’obtention de clones de cellules KO pour l’analyse transcriptomique à l’échelle de la cellule unique	151
Figure 40-Rapport du niveau d’expression de FNIP 1 dans les cellules transfectées par le vecteur 9 par rapport aux cellules contrôles	153
Figure 41- Protocole de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 simple guide et d’obtention de clones de cellules KO pour l’analyse transcriptomique à l’échelle de la cellule unique	155

Liste des tableaux

Tableau 1- Différentes techniques de scRNA-seq 47

Tableau 2 Séquences des oligonucléotides utilisés pour la construction des plasmides CRISPR-

Cas9..... 129

Tableau 3- Séquences des oligonucléotides utilisés pour les analyses en qPCR 132

Liste des abréviations

AMPK : AMP-activated protein kinase

ARACNE : Algorithm for the Reconstruction of Accurate Cellular Networks

Bcl-xL : B-cell lymphoma-extra large

BEC : Blood Endothelial Cells

BFU-E : Bust Forming Unit-Erythroid

BHD : Birt-Hogg-Dubé

CEL seq : Cell Expression by Linear amplification and Sequencing

CFU-E : Colony Forming Unit-Erythroid

CFU-GEMM : Colony Forming unit-Granulocyte, monocyte, macrophage

CK2 : Casein-kinase-2

CLP : Common Lymphoid Precursor

CLR : Context Likelihood of Relatedness

CMP:Common Myeloid Precursor

CNCC : Cranial Neural Crest *Cells*

DoE : Design of Experiment

ERCC : External RNA Controls Consortium

Esrrb : Estrogen Related Receptor Beta

EPO : Erythropoietin

FNIP : Folliculin Interactin Protein

GAP : GTPase-activating proteins

GEF : *Guanine Nucleotide Exchange Factor*

GM-CSF : Granulocyte-macrophage Colony-Stimulating Factor

GM-CSF:Macrophage Colony-Stimulating Factor

GRN : Gene Regulatory Network

HSC : Hematopoietic Stem Cell

IL11 : Interleukine 1

IL13 : Interleukine 3

IVT : In Vitro Transcription

KO : *Knock-out*

LEC : Lymphatic Endothelial Cells

LEC : Lymphatic endothelial cells

MAST : Model-based Analysis of Single Cell Transcriptomics

MARS-seq : Massively parallel single cell RNA-Seq

MEP : Megakaryocyte Erythroid Progenitor

MGG : May-Grünwald-Giemsa

MMLV : Moloney Murine Leukemia Virus

MPP : Mutipotent progenitors

mTORC1 : mammalian Target Of Rapamycin Complex 1

PCA : Principal Component Analysis

PI3K : Phosphoinositide3-Kinase

Prox1 : Prospero Homeobox 1

Rag : Ras related gtpase

RAPTOR : Regulatory-associated protein of mTOR

Rheb : Ras homolog enriched in brain

RNAseq : RNA sequencing

RT-qPCR : Reverse Transcription Polymerase Chain Reaction

SAGE : Serial Analysis of Gene Expression

Sc : Single Cell

SCF : Stem Cell Factor

SEG : Stochastic Gene Expression

sparse PLS : Partial least squares regression

SPLIT-seq : *Split*-Pool Ligation-Based Transcriptome Sequencing

STRT-seq : Single-cell Tagged Reverse Transcription

T2EC : TGF- α /TGF- β -induced erythrocytic cells

TAT : Trans-Activator of Transcription

TFE3 : Transcription factor E3

TPO : Thrombopoietin

TS : Template Switching

TSCC : Tuberous Sclerosis Complex

tSNE : t-distributed Stochastic Neighbor Embedding

UMI : Unique Molecular Identifier

WASABI : Waves Analysis Based Inference

WGCNA : Weighted Gene Co-expression Network Analysis

1. Introduction bibliographique

1.1 Différenciation cellulaire : les deux visions du processus

Le destin cellulaire mais surtout comment une cellule choisit entre différentes destinées est une des questions fondamentales en biologie. Ce choix peut être entre la survie ou la mort, la quiescence ou la prolifération, l'auto-renouvellement ou la différenciation. Au cours des dernières décennies, différents travaux ont cherché à définir un cadre théorique pour décrire les phénomènes impliqués dans la prise de décision cellulaire au cours de la différenciation et à décrire les caractéristiques les plus pertinentes de ce processus.

Deux grandes visions du processus de différenciation ont été largement décrites et peuvent être brièvement résumées comme le suit : a/ la vision d'un processus déterministe de la différenciation, selon laquelle la différenciation cellulaire est considérée comme un processus linéaire, unidirectionnel et irréversible, b/ la vision d'un processus dynamique de la différenciation, selon laquelle chaque cellule prend ses propres décisions définissant ainsi son destin, régies par des programmes complexes de régulation génétique et par un large éventail de signaux externes.

Dans cette partie bibliographique, nous allons nous concentrer principalement sur la compréhension des mécanismes sous-jacents à la prise de décision cellulaire dans la différenciation hématopoïétique et notamment la différenciation érythrocytaire.

L'hématopoïèse est le processus continu et régulé de formation, de prolifération, de différenciation et de maturation de toutes les lignées de cellules sanguines (érythrocytes, mégacaryocytes, granulocytes, macrophages et lymphocytes) qui se produit pendant toute la période du développement embryonnaire et se poursuit tout au long de l'âge adulte (Jagannathan-Bogdan and Zon 2013).

Chez les vertébrés, il existe deux grands types d'hématopoïèse : l'hématopoïèse primitive et l'hématopoïèse définitive (Galloway and Zon 2003).

L'hématopoïèse primitive se produit dans le sac vitellin lors des premiers stades du développement embryonnaire et se poursuit pendant une période relativement courte. L'objectif principal de l'hématopoïèse primitive est de produire des globules rouges pour l'oxygénation des tissus alors que l'embryon est en phase de croissance rapide. Elle implique la formation d'un progéniteur érythroïde qui donne naissance à des érythrocytes et des macrophages primitifs du système respiratoire précoce. Cette étape primitive est toutefois transitoire, et ces progéniteurs érythroïdes non pluripotents n'ont pas de capacité de renouvellement (Paik and Zon 2010).

L'hématopoïèse définitive implique quant à elle les cellules souches hématopoïétiques (HSC, Hematopoietic Stem Cells). Chez les vertébrés, les HSC émergent du plancher de l'aorte dorsale dans une région appelée splanchnopleure para-aortique puis aorte-gonade-mésonephros par une transition endothélio-hématopoïétique dans l'embryon en développement, au moment où se met en place la circulation sanguine (Adamo et al. 2009). Une expérience de marquage grâce à la GFP des cellules endothéliales de l'embryon de poulet a mis en évidence l'origine endothéliale des HSC grâce à la détection du marqueur également dans ces HSC (Jaffredo et al. 2010). D'autres études réalisées sur l'embryon humain ont aussi montré que les cellules endothéliales prélevées dans l'aorte-gonade-mésonephros et cultivées en présence d'une couche stromale se différencient en cellules hématopoïétiques (Oberlin et al. 2002). Ces études ont ainsi confirmé l'origine endothéliale du système hématopoïétique. Une fois formées, les HSC migrent ensuite vers le foie fœtal puis vers la moelle osseuse, le siège des HSC chez l'adulte (Cumano and Godin 2007).

L'hématopoïèse a été historiquement schématisée par un diagramme pyramidal et hiérarchique dans lequel la HSC multipotente est placée au sommet (Till and McCulloch 1980) (Akashi et al. 2000). Selon ce modèle, les HSC perdent petit à petit leur capacité d'auto-renouvellement en s'engageant dans la différenciation sous l'influence de facteurs stimulants. Cette perte de capacité est liée notamment à la formation d'une population appelée MPP (Multipotent progenitors) (Wolff and Humeniuk 2013). Une fois engagés, les MPP génèrent deux types de progéniteurs multipotents : les CLP (Common Lymphoid Precursor) et les CMP (Common Myeloid Precursor) permettant la formation du lignage lymphoïde et du lignage myéloïde, respectivement. Les CLP, sous l'effet de l'IL7 (Interleukine 7), s'orientent vers le lignage

lymphoïde en se différenciant d'abord en précurseurs immatures : les pro- et les pré-lymphoblastes pour donner ensuite les lymphocytes T et B. Par contre, les CMP sous l'effet de l'IL3 (Interleukine 3) et du SCF (Stem Cell Factor), se différencient en MEP (Megakaryocyte–Erythroid Progenitor) ou en GMP (Granulocyte-Monocyte Progenitor) pour donner naissance à des progéniteurs unipotents impliqués dans la différenciation mono-lignage : les CFU (Colony forming units) de type Basophile (-Baso), Eosinophile (-Eo), Mégacaryocytaire (-MK), Granulo-Monocytaire (-GM) et Erythrocytaire (-E) qui, au fur et la mesure des divisions, vont permettre la production des cellules sanguines matures (les granulocytes, monocytes/macrophages et les érythrocytes) qui pourront avec les lymphocytes exercer leurs fonctions en passant dans la circulation sanguine (S. H. Orkin 2000) (Stuart H. Orkin and Zon 2008) (Figure 1).

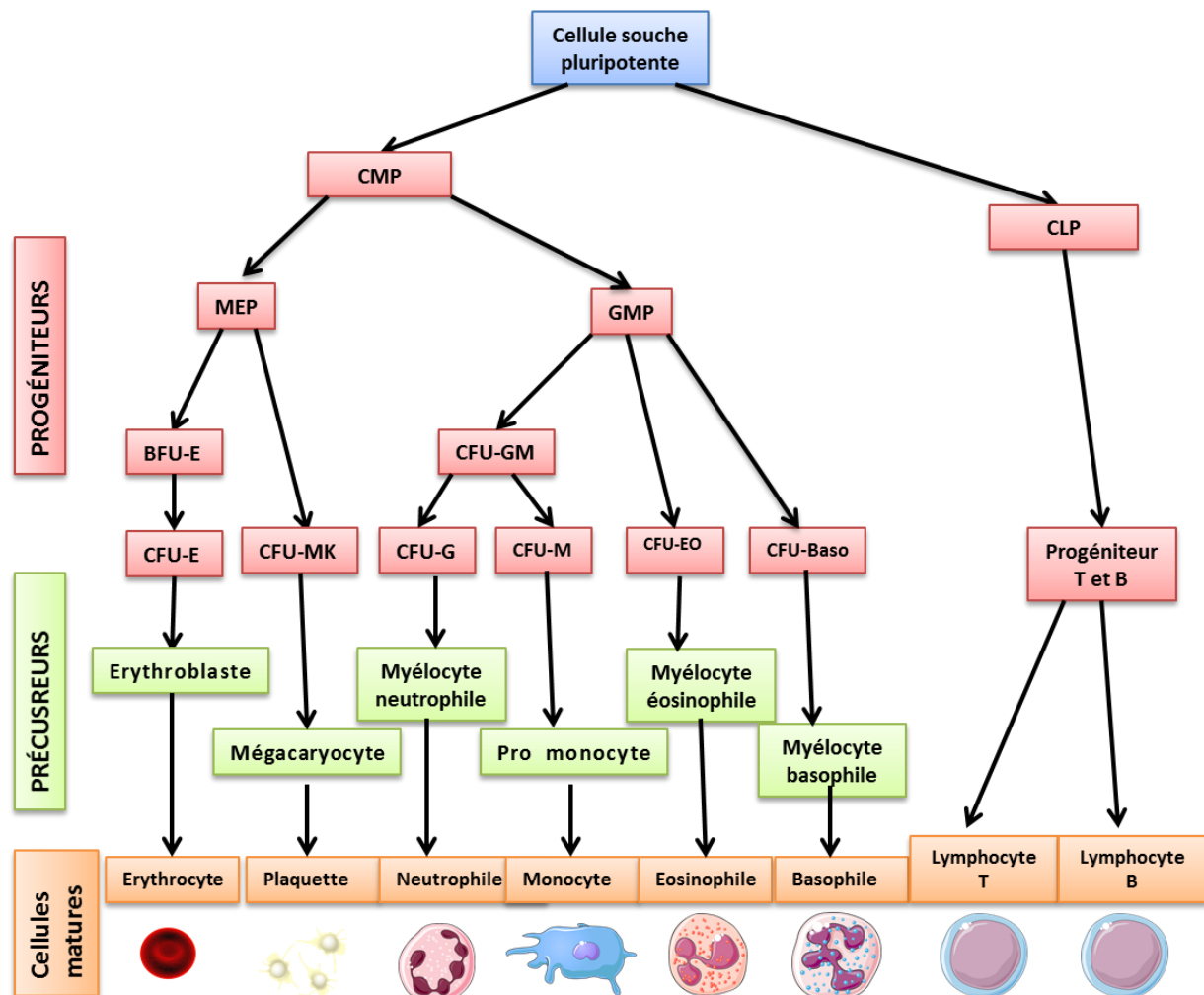


Figure 1-Organisation de l'hématopoïèse adulte

L'hématopoïèse est un processus de division et de maturation qui permet la formation de toutes les cellules sanguines à partir d'une cellule souche hématopoïétique (HSC). La HSC se différencie en deux progéniteurs multipotents : le progéniteur lymphoïde (CLP) et le progéniteur myéloïde (CMP). Ces derniers vont se diviser pour donner des cellules sanguines matures issues des deux lignées lymphoïde et myéloïde. Figure inspirée de (Shahi et al. 2019).

Parmi les différents lignages hématopoïétiques, l'érythropoïèse est le processus qui permet la formation des globules rouges appelés aussi érythrocytes ou hématies, indispensables à l'oxygénation de l'organisme grâce à la présence d'hémoglobine au sein de leur structure.

1.1.1 Différenciation cellulaire dans la vision classique

1.1.1.1 Un processus linéaire et unidirectionnel

La différenciation cellulaire a été largement définie comme un processus unidirectionnel et irréversible permettant la transformation d'une cellule immature en une cellule mature exerçant une fonction cellulaire spécifique.

Afin de comprendre les mécanismes sous-jacents au processus de différenciation cellulaire, de nombreuses études ont été menées à l'échelle de la population cellulaire en réponse à un signal de différenciation. Selon le point de vue déterministe, le phénotype découle directement du génotype et la différenciation cellulaire est schématisée par des diagrammes linéaires selon lesquels une cellule, sous l'action de signaux spécifiques, s'engage dans sa voie de différenciation selon une séquence définie d'états intermédiaires (Wolpert 1994). Ces signaux agissent sur l'activation ou à la désactivation de l'expression de gènes ou de groupes de gènes sélectionnés (Chan and Kyba 2013). Ces gènes précis ont été décrits comme des « régulateurs principaux » ou des « gènes maîtres » qui déterminent un destin cellulaire unique (Chan and Kyba 2013) (Witczak, Sharoff, and Goodyear 2008).

Par ailleurs, il a été décrit que lors de son engagement, la cellule subit des changements fonctionnels et phénotypiques qui ne peuvent pas être facilement révertés (Balázs, van Oudenaarden, and Collins 2011). Lorsque la cellule a choisi son destin et mis en œuvre les changements nécessaires, une cellule ne peut donc généralement plus revenir en arrière (Strohman 1997). Historiquement, cet engagement irréversible a été expliqué comme résultant de processus de condensation de l'ADN par désacétylation des histones (Cheung, Briggs, and Allis 2000) et hyperméthylation des îlots CpGs (Razin and Cedar 1991).

1.1.1.2 Différenciation érythrocytaire dans la vision classique

L'érythropoïèse a été tout d'abord schématisée par un diagramme linéaire selon lequel la HSC s'engage dans la voie de la lignée rouge pour permettre la formation des précurseurs érythroïdes assurant ensuite la production des globules rouges matures. Selon ce modèle,

l'érythropoïèse a été divisée en deux grandes étapes : l'érythropoïèse précoce et l'érythropoïèse terminale (Figure 2).

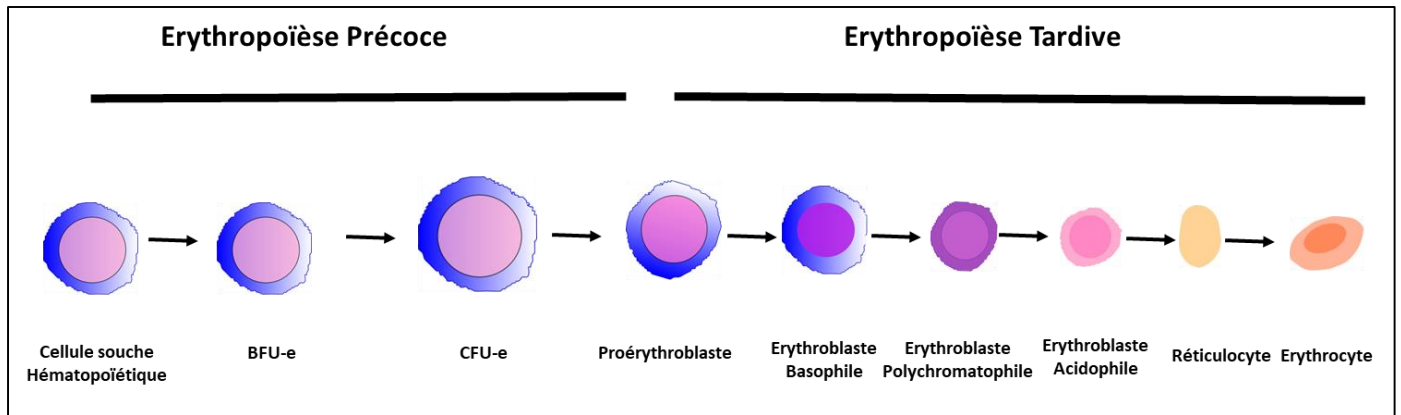


Figure 2- Schéma de l'érythropoïèse selon la vision classique

L'érythropoïèse se divise en deux grandes étapes : l'érythropoïèse précoce et l'érythropoïèse tardive. Les différents stades de la différenciation érythrocytaire sont présentés.

L'érythropoïèse précoce

En réponse à certains facteurs environnementaux, la HSC peut emprunter la voie myéloïde : elle se différencie alors en un progéniteur myéloïde commun ou CMP appelé aussi CFU-GEMM (clony forming unit-Granulocyte, monocyte, macrophage) qui donne naissance à un progéniteur MEP (Megakaryocyte Erythroid Progenitor). Le MEP se différencie ensuite en progéniteur BFU-E (Bust Forming Unit-Erythroid), qui s'engage vers la lignée érythroïde en se différenciant en CFU-E (Colony Forming Unit-Erythroid), le progéniteur érythroïde le plus mature (C. J. Gregory and Eaves 1978) (Figure 2). Les HSC représentent environ 0,05% des cellules médullaires tandis que les BFU-E et CFU-E représentent chacune 0,3% de la population des cellules de la moelle osseuse.

La caractérisation des cellules de l'érythropoïèse précoce se fait grâce à leurs marqueurs membranaires : la HSC exprime à sa surface la molécule CD34 dont l'expression est maintenue jusqu'au stade BFU-E. Au-delà du stade CFU-GEMM, le récepteur SCF (Stem Cell Factor), également appelé CD117 ou c-kit, est présent. À partir du stade BFU-E, le récepteur de l'EPO (Erythropoïétin) est trouvé également à la surface des cellules.

L'érythropoïèse tardive

L'érythropoïèse tardive correspond à la phase ultérieure de la différenciation terminale des progéniteurs érythrocytaires. Cette phase, comprend l'ensemble des modifications morphologiques et fonctionnelles qui aboutissent à la production d'un globule rouge mature.

Dans l'érythropoïèse tardive, le SCF agit en synergie avec l'EPO dans la différenciation des CFU-E (H. Wu et al. 1995). En effet, il a été démontré que le récepteur de l'EPO ne peut pas être activé sans la phosphorylation de son domaine cytosolique médiée par la liaison de SCF à son récepteur KIT suivie d'une activation médiée par l'EPO.

L'érythropoïèse tardive regroupe les stades : proérythroblaste, érythroblaste basophile (stade immature), érythroblaste polychromatophile, érythroblaste acidophile (stade mature). Ces cellules tirent leur nom de la coloration de leur cytoplasme suite à la coloration de May-Grünwald-Giemsa (MGG).

Lors de l'érythropoïèse tardive, les cellules qui se forment d'abord sont basophiles, et deviennent acidophiles grâce à l'augmentation de leur quantité d'hémoglobine cellulaire. Parallèlement, la chromatine se condense, les organites tels que le réticulum endoplasmique ou l'appareil de Golgi sont peu à peu éliminés et la taille cellulaire diminue (Dolznic et al. 1995).

De plus, il a été décrit qu'aux stades terminaux de la différenciation, les érythroblastes acidophiles de certaines espèces perdent leurs noyaux. L'expulsion du noyau dépend de la réorganisation de protéines d'adhésion au niveau de la membrane plasmique et des interactions des érythroblastes avec les macrophages (Lee et al. 2004) (Soni et al. 2006). Parallèlement, il a été mis en évidence que chez ces espèces, les érythrocytes peuvent se débarrasser de leurs

mitochondries par autophagie pour devenir des érythrocytes matures. Ces derniers fabriquent de l'hémoglobine permettant de transporter l'oxygène vers tous les tissus de l'organisme (Stuart H. Orkin and Zon 2008). Cependant, ces caractéristiques de l'érythropoïèse tardive ne s'appliquent pas chez les oiseaux. Le noyau est conservé mais se condense et devient non fonctionnel. Ainsi, toute activité de transcription devient impossible dans ce noyau (Stuart H. Orkin and Zon 2008). Par ailleurs, chez les oiseaux, les mitochondries ne sont pas phagocytées et restent toujours fonctionnelles (Kowalski and Pałyga 2011).

L'expression génique dans la vision classique de la différenciation érythrocytaire

Selon, le modèle déterministe, l'expression de certains gènes apparaît indispensable pour la différenciation érythrocytaire. Par exemple, le facteur de transcription GATA-1 a été reconnu comme le régulateur transcriptionnel principal de l'érythropoïèse, contrôlant tous les aspects de cette différenciation (Ferreira et al. 2005) (Katsumura 2017). De plus, il a été rapporté qu'en présence de l'EPO, GATA-1 induit fortement l'expression du gène codant la protéine anti-apoptotique Bcl-xL (B-cell lymphoma-extra large) afin de promouvoir la survie des cellules érythroïdes (T. Gregory et al. 1999) (Wagner et al. 2000). En plus de GATA1, Tal1 (T cell acute lymphocytic leukemia 1 protein ou SCL) et Klf1 (Erythroid Kruppel-like factor ou EKLF) ont été désignés comme des "maîtres régulateurs" de l'érythropoïèse primitive et définitive dont l'absence a entraîné une anémie sévère et la mort chez les souris (Perkins, Sharpe, and Orkin 1995). En outre, l'analyse des profils d'expression génétique a montré que GATA-1, Tal1 et Klf1 sont tous trois nécessaires à l'expression des gènes codant la β -globine (Hbb) et l' α -globine (Hba) ainsi qu'à l'activation d'un grand nombre d'autres gènes érythroïdes chez les souris. Cela suggère que GATA-1, Tal1 et Klf1 fonctionnent de manière large et synergique pour réguler le programme transcriptionnel érythroïde (Kassouf et al. 2010) (W. Wu et al. 2011).

1.1.2 Différenciation cellulaire dans la vision moderne

1.1.2.1 Stochasticité de l'expression des gènes (SGE)

Toutes les cellules d'un organisme multicellulaire possèdent le même code génétique contenu dans leurs molécules d'ADN. Cependant, elles présentent des phénotypes et des fonctions extrêmement variées : les globules rouges échangent de l'oxygène, les cellules musculaires se contractent, les cellules du système immunitaire reconnaissent les agents pathogènes... De ce fait, une régulation fine de l'expression des gènes est mise en place pour que chaque cellule produise les protéines nécessaires à son fonctionnement et lui permettant de s'adapter à son environnement.

L'expression des gènes, encore appelée expression génique ou expression génétique, désigne l'ensemble des processus biochimiques par lesquels l'information héréditaire stockée dans un gène est lue pour assurer la fabrication de molécules comme les ARN ou les protéines qui auront un rôle actif dans le fonctionnement cellulaire.

Les points de contrôle permettant de réguler l'expression génétique sont situés à différents niveaux de l'échelle cellulaire allant de l'ADN à la protéine. Tout d'abord, le contrôle de la synthèse de l'ARN (ou de la transcription), est médié par des facteurs qui se lient en amont des gènes sur les promoteurs ou qui influencent la liaison d'autres molécules aux éléments cis-régulateurs du promoteur. Parce que de tels événements de liaison sont le résultat de rencontres aléatoires entre molécules, dont certaines sont présentes en petit nombre, les processus biochimiques qui régulent l'initiation de la transcription sont intrinsèquement stochastiques. De plus, les processus en plusieurs étapes qui aboutissent à la dégradation des ARNm, à la synthèse et dégradation des protéines, sont également soumis à un bruit au niveau moléculaire.

D'autres facteurs peuvent également influencer la stochasticité de l'expression des gènes comme l'épigénétique et la dynamique chromatinienne (Singer et al. 2014) (Viñuelas et al. 2013a). En effet, l'ADN est empaqueté dans le noyau de la cellule sous forme de chromatine dont l'unité est le nucléosome. La chromatine à l'état condensé empêche l'accès des facteurs de transcription et des complexes enzymatiques à l'ADN. Pour dépasser cette barrière, et sous

l'influence des facteurs environnementaux, la cellule utilise différentes stratégies dont l'incorporation de variants d'histones au sein de la particule nucléosomale. L'ensemble de ces histones, leur position et leur capacité à être méthylées/déméthylées, à être acétylées/déacétylées, forment un système complexe tout au long du génome qui contribue à la stochasticité de l'expression génique en influençant directement l'expression des gènes (Verdone, Caserta, and Di Mauro 2005).

Les premiers indices de la variabilité aléatoire de l'expression des gènes, appelée aussi improprement « bruit », datent de la fin des années 1950. Des microbiologistes de l'Université de Chicago, travaillant sur le modèle de l'opéron lactose ont montré que des bactéries génétiquement identiques et placées dans des milieux de culture identiques avaient un comportement différent (Novick and Weiner 1957) (Cohn and Horibata 1959); les unes étaient capables de métaboliser le lactose signifiant qu'elles fabriquaient l'enzyme impliquée dans cette réaction ; les autres ne pouvaient pas assurer cette réaction biochimique, l'enzyme étant produite en trop petite quantité ou pas du tout. Cette première expérience a ainsi permis de mettre en évidence que l'expression des gènes était stochastique. Ainsi, la mise en évidence de la SEG a fait émerger une vision novatrice de l'expression des gènes. D'autres travaux plus récents ont ensuite permis de mieux caractériser la stochasticité de l'expression génique dans divers organismes et dans des conditions très variées (Elowitz et al. 2002) (Blake et al. 2003).

Bien qu'elle ait été longtemps négligée (Tian et al. 2014) et considérée comme un bruit plutôt néfaste, de nombreuses études sur la SGE démontrent aujourd'hui son influence bénéfique et nécessaire dans certains processus biologiques (C. Liu 2004) (Dong et al. 2016) (Haghverdi, Buettner, and Theis 2015) (Pina et al. 2015). Par exemple, dans le cas du virus HIV, il a été mis en évidence que la variabilité d'expression du gène TAT (Trans-Activator of Transcription) entraîne une variabilité phénotypique et influence la latence virale (Weinberger et al. 2005). Comme autre exemple, certaines souches de bactéries et de levures utilisent la stochasticité de l'expression génique pour maintenir des sous-populations capables de résister à des traitements antibiotiques ou à des environnements stressants, permettant ainsi de disposer d'un phénotype résistant, tout en gardant une réversibilité rapide lorsque le stress disparaît (Stern et al. 2007) (Thattai and van Oudenaarden 2004).

1.1.2.2 Différentes méthodes d'analyse de la variabilité de l'expression génique

Différents outils mathématiques permettent d'approcher l'hétérogénéité cellulaire. Nous pouvons citer des outils mathématiques comme le coefficient de variation qui permet de mesurer la variation par rapport à la moyenne, le facteur de Fano et la variance normalisée (Vallejos, Marioni, and Richardson 2015) (Munsky, Neuert, and van Oudenaarden 2012) (Viñuelas et al. 2013b). La variabilité intercellulaire peut être aussi approchée par des mesures de distances. Par exemple, la distance Euclidienne qui consiste à déterminer la distance qui sépare deux cellules en fonction du niveau d'expression des gènes et l'approche du centre médian qui consiste à déterminer le centre médian dans l'espace multidimensionnel des gènes et à préciser ainsi, la position, à laquelle la distance à chaque cellule analysée est la plus courte possible (Stumpf et al. 2017). Pour mesurer le bruit transcriptionnel, d'autres chercheurs ont calculé les distances entre les cellules par paire en se basant sur la corrélation de rang des gènes hypervariables exprimés dans les cellules (Mohammed et al. 2017). Bien que l'ensemble de ces méthodes d'analyse permettent de quantifier la variabilité de l'expression génique, elles sont influencées par plusieurs facteurs notamment l'expression moyenne des gènes (Eling, Morgan, and Marioni 2019). Enfin, d'autres métriques de mesure comme l'entropie de Shannon sont également utilisées. L'entropie sert à mesurer la proportion d'aléatoire ou d'hétérogénéité dans un système : la valeur d'entropie diminue quand l'expression d'un gène est homogène entre les cellules alors qu'elle augmente quand l'expression inter-cellulaire de ce gène est hétérogène (Stumpf et al. 2017). Récemment, la puissance de l'entropie pour quantifier la variabilité a été bien décrite (Olivier Gandrillon et al. 2021). En effet, elle se distingue des autres métriques, car elle mesure la variabilité des éléments au sein d'une distribution donnée, approche différente des méthodes classiques basées sur la moyenne.

1.1.2.3 SGE et dynamique de la différenciation cellulaire

Le rôle de la stochasticité de l'expression des gènes dans la différenciation cellulaire a été décrit dans de nombreux modèles expérimentaux.

En utilisant la cytométrie en flux, Chambers et al (2007) ont établi que le facteur de pluripotence Nanog, couplé à la GFP par recombinaison homologue, est exprimé de manière hétérogène dans les cellules souches embryonnaires de souris (Chambers et al. 2007). De plus, il a également été démontré que l'augmentation de la variabilité intercellulaire conduit à une augmentation de la probabilité d'engagement de ces cellules dans un lignage (Moris et al. 2018). De plus, le rôle de la SEG dans la détermination du destin des cellules immunitaires a été aussi mis en évidence. En effet, la SEG contrôle la prise de décision des cellules T CD8 activée entre le statut de la cellule effectrice et le statut de la cellule mémoire (Girel and Crauste 2018).

De nombreuses études ont été menées pour décrire les processus cellulaires, et particulièrement le processus de différenciation, par des modèles stochastiques dynamiques (Huang 2009). Dans ces modèles, les cellules d'une population sont présentées comme des individus indépendants présentant une hétérogénéité quantitative et qualitative importante et la SEG est considérée comme l'une des causes sous-jacentes de l'engagement cellulaire (Hume 2000).

Une des premières théories qui explique la dynamique cellulaire au cours du développement est celle de C. Waddington illustrée dans les années 1957. Dans le paysage de Waddington, une cellule est représentée par une bille qui roule sur une surface (paysage), qui traverse les vallées et cherche son destin (état stable) représenté par les points les plus bas ou les bassins. Avant les points les plus bas, les vallées se ramifient de sorte que chaque bille (ou cellule) peut emprunter différents chemins (Figure 3 - panel A) (Waddington 1957). Lorsque la cellule a pris une décision, son potentiel cellulaire diminue restreignant ainsi ses possibilités de destin (Moris, Pina, and Arias 2016) (Rajagopal and Stanger 2016). De plus, selon Waddington, la forme du paysage épigénétique dépend de l'activité des gènes (Figure 3 - Panel B). Aujourd'hui, il est clairement démontré que les gènes ne fonctionnent pas indépendamment, mais qu'ils sont

liés d'une manière dynamique formant des Réseaux de Régulation de Gènes (GRN) (Huang 2012) (Elowitz et al. 2002).

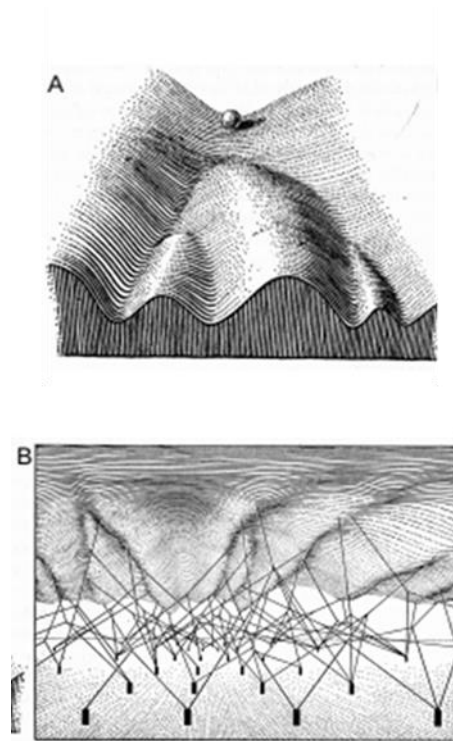


Figure 3- Paysage épigénétique de Waddington

(A) La bille représente une cellule qui se déplace dans le paysage. (B) Les interactions des gènes (bâtonnets noirs) qui forment le paysage de Waddington. Figure tirée de (Waddington 1957).

Ainsi, plusieurs facteurs influencent l'état des cellules dans le paysage de Waddington : un facteur interne qui est représenté par la SEG et des facteurs externes (comme l'environnement et les stimuli extérieurs) qui peuvent modifier les réseaux des gènes et moduler ensuite la topologie du paysage (le nombre et la nature des états stables du paysage) (Huang et al. 2007) (Verd, Crombach, and Jaeger 2014).

Selon la vision de Waddington, une fois que la cellule a atteint son état cellulaire stable, elle ne peut pas retourner en arrière. Toutefois, il est aujourd'hui clairement admis que certaines cellules conservent une certaine plasticité. Dans une approche plus moderne, la détermination du destin cellulaire est donc beaucoup plus large et les cellules peuvent prendre des voies distinctes de différenciation en réponse à différents signaux extracellulaires. En effet, selon cette

vision, les cellules évoluent dans un paysage peuplé de puits (attracteurs) et oscillent sur leur position à cause des variations aléatoires de la quantité des ARNm et des protéines (Baedke 2013). Ainsi, nous pouvons assister à des événements de dédifférenciation ou de réversion cellulaire dans la mesure où une cellule qui a atteint un intermédiaire de différenciation peut revenir à un stade antérieur, naturellement ou sous l'action de signaux appropriés. Conformément à cette vision, nombreuses études ont mis en lumière la plasticité du destin cellulaire. Par exemple, une étude a mis en évidence le phénotype labile et réversible des cellules endothéliales lymphatiques (LEC pour Lymphatic Endothelial Cells) différenciées en montrant que la désactivation du gène *Prox1* (Prospero Homeobox 1) était suffisante pour initier une cascade de reprogrammation menant à la dédifférenciation des LEC en cellules endothéliales plus précoces appelées BEC (Blood Endothelial Cells) (Johnson et al. 2008). En outre, d'autres travaux ont montré que des cellules épithéliales luminales entièrement différenciées pouvaient se dédifférencier en cellules souches basales complètement fonctionnelles après l'ablation des cellules souches des voies respiratoire chez la souris (Tata et al. 2013). D'autres études se sont concentrées sur l'analyse d'une population cellulaire transitoire spécifique aux vertébrés, appelée cellules de la crête neurale crânienne (CNCC, ou Cranial Neural Crest Cells). Bien qu'elles proviennent de l'ectoderme et qu'elles soient capables de générer la plupart du squelette cranio-facial, les CNCC prémigratoires peuvent aussi donner naissance à des types de cellules mésenchymateuses canoniquement associées au lignage mésoderme, comme l'os, le cartilage et les muscles lisses. En effet, l'analyse du transcriptome à l'échelle de la cellule unique des CNCC murins génétiquement marqués, pendant 14 heures de développement, a permis d'identifier une population de précurseurs exprimant transitoirement le facteur de pluripotence Oct4. L'augmentation du taux d'Oct4 confère aux précurseurs du CNCC le potentiel de changer rapidement leur identité transcriptionnelle neuro-épidermique pour réverser et adopter une nouvelle signature mésodermique capable de former les dérivés typiques du mésoderme, comme le mésenchyme (Zalc et al. 2021).

1.1.2.4 Différenciation érythrocytaire dans la vision moderne

Les programmes qui régissent l'hématopoïèse ont été étudiés principalement par l'analyse, à l'échelle de la population, des cellules souches et des progéniteurs hématopoïétiques. Ces études sont basées essentiellement sur l'analyse de l'expression de marqueurs de surface cellulaire et des gènes régulateurs principaux des différents lignages. Cependant, une population supposée homogène de cellules, déterminée par un ensemble bien défini de marqueurs de surface cellulaire par exemple, peut inclure différents types cellulaires fonctionnellement distincts. Par exemple, l'analyse transcriptomique menée à l'échelle de la cellule unique d'une population de progéniteurs myéloïdes de la moelle osseuse de souris a permis d'identifier 18 sous-populations de progéniteurs myéloïdes avec des potentiels de lignage distincts envers les érythrocytes, les mégacaryocytes, les monocytes, les cellules dendritiques, les neutrophiles, les basophiles et les éosinophiles (Paul et al. 2015). Dans cette étude, les chercheurs ont découvert que *Cebpa* et *Cebpe* sont deux facteurs de transcription de décision ayant un rôle clef dans le destin cellulaire des progéniteurs myéloïdes. Des niveaux élevés d'expression de *Cebpa* orientent les progéniteurs myéloïdes vers les lignées des neutrophiles, des monocytes et des basophiles, tandis que des niveaux élevés d'expression de *Cebpe* sont observés principalement dans les sous-populations de neutrophiles et d'éosinophiles (Paul et al. 2015). En outre, l'analyse transcriptomique à l'échelle de la cellule unique de 180 HSC *bcl11a*^{-/-} et *bcl11a*^{+/+} a mis en évidence que les HSC *bcl11a*^{-/-} sont constituées de deux sous-populations distinctes restreintes à la différenciation myéloïde et érythrocytaire (Tsang et al. 2015). Pour citer un autre exemple, l'analyse du profil de 1 656 HSPC uniques a révélé des changements d'expression associés à la différenciation précoce des lymphocytes, des érythrocytes et des granulocytes-macrophages et a permis d'illustrer les HSC comme des cellules qui produisent toutes les cellules sanguines matures *via* la formation de progéniteurs intermédiaires (Nestorowa et al. 2016). Ainsi, cette étude a permis la reconstruction d'un atlas répertoriant différentes trajectoires et qui révèle une véritable variabilité cellulaire (Nestorowa et al. 2016).

D'autres études ont également décrit des modèles dynamiques dans lesquels, l'hétérogénéité intercellulaire est considérée comme l'une des causes sous-jacentes du

processus du développement des lignées hématopoïétiques (Moussy et al. 2017) (Velten et al. 2017). Par exemple, l'analyse à l'échelle de la cellule unique des profils transcriptomiques des cellules CD34+ pendant 4 temps de différenciation a mis en évidence des changements transcriptomiques conduisant à un état d'expression génique multipotent atteint très rapidement au début du premier cycle cellulaire. Cet état instable est suivi par l'émergence de deux groupes cellulaires avec des profils de transcription distincts avant la fin du premier cycle cellulaire. Dans le premier groupe, les cellules présentent une morphologie stable et un profil de transcription cohérent dès la fin du premier cycle cellulaire. Dans le deuxième groupe, les cellules ont une morphologie instable avec un comportement dynamique et hésitant durant plusieurs cycles cellulaires. Grâce à l'identification de cette dernière catégorie de cellules avec ses caractéristiques phénotypiques fluctuantes, cette étude a mis en évidence la cinétique variable et la nature dynamique de l'engagement des cellules hématopoïétiques qui s'opposent à la vision classique de la différenciation historiquement proposé (Moussy et al. 2017).

Par conséquent, tous ces résultats, remettent en question les modèles classiques décrivant une hématopoïèse basée sur la perte progressive du potentiel de différenciation de l'ensemble des cellules.

Dans notre équipe, nous nous intéressons particulièrement à la dynamique du processus de différenciation des progéniteurs érythrocytaires de poulet appelés T2EC (O Gandrillon et al. 1999). Ces cellules présentent l'avantage de constituer un modèle physiologiquement pertinent de cellules primaires phénotypiquement homogènes, sans aucune transformation génétique et qui ne peuvent se différencier que dans un seul lignage, le lignage érythrocytaire, après changement du milieu de culture. Récemment, l'analyse de l'expression génique à l'échelle de la cellule unique des T2EC pendant 6 temps de différenciation a montré une augmentation transitoire de la SEG et des gènes du cycle cellulaire entre 8h et 24h du processus de différenciation (Richard et al. 2016). Selon un point de vue dynamique, cette augmentation peut être expliquée par le fait que les cellules ont besoin d'explorer un espace d'expression génétique plus large pour pouvoir se différencier. En outre, le rôle de la SEG dans le processus de différenciation des T2EC a été également bien décrit. En se basant sur des données de RT-qPCR acquises à l'échelle de la cellule unique et des outils de modélisation, notre équipe a en effet

démontré que des drogues modulant la SGE sont capables de modifier le processus de différenciation des T2EC (Guillemin et al. 2019).

De plus, l'augmentation de la SEG lors de la différenciation des T2EC est accompagnée d'un maintien de la capacité des progéniteurs à retourner dans un état d'auto-renouvellement jusqu'à 24 heures après l'induction de la différenciation ; au-delà de 24h, les cellules sont engagées de manière irréversible dans la différenciation (Richard et al. 2016). Malgré ces preuves de la dynamique du processus de différenciation érythrocytaire, on sait peu de choses sur la façon dont la réversibilité de ces cellules coïncide vraiment avec des changements moléculaires au sein de ces mêmes cellules.

Ainsi, le premier objectif de mon travail de thèse a été de caractériser moléculairement les cellules capables d'une telle réversion, grâce à des analyses transcriptomiques menées à l'échelle de la cellule unique.

1.2 Analyses du transcriptome à l'échelle de la cellule unique

Comme déjà évoqué précédemment, il est aujourd'hui largement admis, que même dans une population de cellules clonale, les cellules ne sont pas identiques et présentent une hétérogénéité quantitative et qualitative significatives. Les études du profil de l'expression génique en cellule unique constituent donc une nécessité afin de mieux comprendre les processus biologiques impliqués dans la prise de décision cellulaire et le rôle de la SGE dans ces processus.

Au cours des dernières années, les avancées technologiques permettant l'étude du transcriptome à l'échelle de la cellule unique et à haut débit ont ouverts des champs de recherche considérables. Grâce à ces technologies, il est possible aujourd'hui d'intégrer la nature stochastique de l'expression génétique dans les interprétations des processus biologiques, et d'obtenir des informations précises et plus justes des différents types cellulaires composant les tissus dans un contexte physiologique et pathologique (Guo et al. 2010). D'ailleurs, pour ne citer qu'un exemple, ces études à l'échelle de la cellule isolée ont des impacts considérables

notamment en cancérologie en permettant une caractérisation précise et complète des cellules souches cancéreuses, des clones métastatiques ainsi que des clones résistants aux médicaments et en permettant ainsi l'utilisation de thérapies plus ciblées et personnalisées (Hannemann et al. 2011).

Différentes méthodes existent pour analyser les ARN à l'échelle de la cellule unique. Nous avons fait le choix de n'en décrire que deux : le RNAseq (RNAsequencing) et la RT-qPCR (Reverse Transcription – Quantitative Polymerase Chain Reaction).

1.2.1 RNAseq en cellule unique (scRNA-seq)

1.2.1.1 Principe général du RNA-seq

Le RNA-seq est une des approches d'analyse du transcriptome les plus récentes qui permet de déterminer la quantité d'ARN présente dans un échantillon en utilisant la technologie de Next Generation Sequencing (NGS) (Y. Chu and Corey 2012). Très simplement, le principe général est le suivant : 1/ une population d'ARN (messenger ou ARN totaux) est convertie en fragments d'ADN ; 2/ ces fragments sans ou après amplification sont ligués avec des adaptateurs de séquençage ; 3/ les fragments d'ADN sont séquencés au niveau d'une seule ou des deux extrémités permettant d'obtenir des lectures de séquences simples ou appariées ; 4/ les séquences obtenues sont ensuite alignées sur un transcriptome de référence et peuvent être classées en trois types : reads exoniques, reads de jonction et reads de côté poly(A) ; 5/ les informations recueillies (nature du gène et nombre de séquences relatives à ce gène) permettent ensuite de générer un profil d'expression pour chaque gène (Z. Wang, Gerstein, and Snyder 2009) (Figure 4).

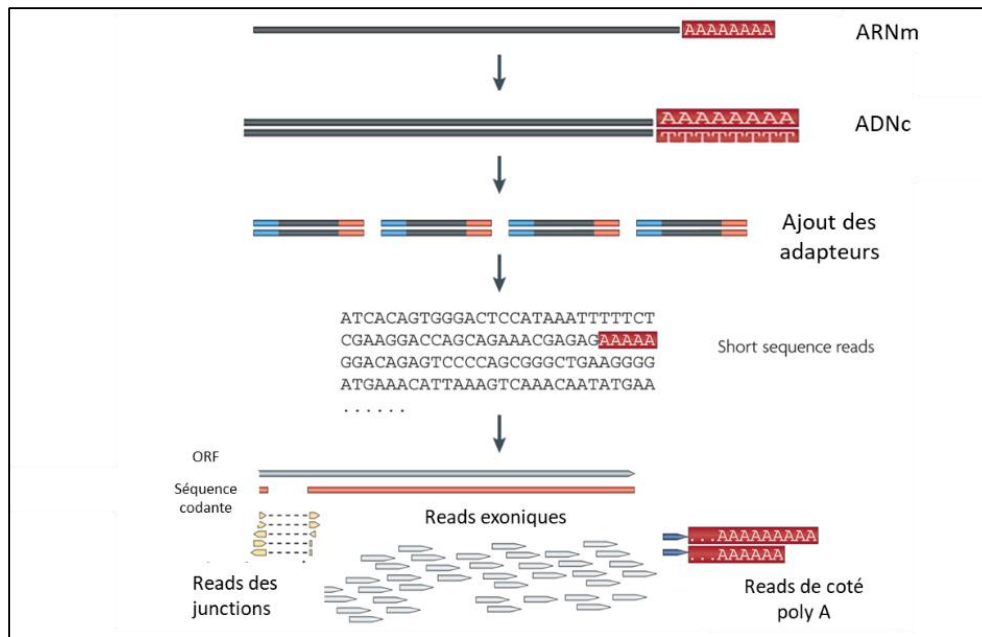


Figure 4- Principe général du RNA-seq

Une population d'ARN (ARNm ici) est convertie en fragments d'ADNc. Des adaptateurs de séquençage (bleus et rouges) sont ensuite ajoutés à chaque fragment d'ADNc et les séquences sont définies par séquençage. Les séquences sont alignées avec une séquence de référence et peuvent être classées en trois types : reads exoniques, reads de jonction et reads de côté poly(A). Figure adaptée de (Z. Wang, Gerstein, and Snyder 2009).

Contrairement au RNA-seq en population qui permet de mesurer l'expression moyenne des transcrits dans une population de cellule, le scRNA-seq permet de mesurer l'expression des transcrits dans chaque cellule d'une population prise individuellement. Dans ce rapport, nous nous focaliserons sur les méthodes mesurant uniquement les ARNm dans les cellules.

1.2.1.2 ScRNA-seq : les différentes étapes

Les approches d'analyse du transcriptome par scRNA-seq comportent plusieurs étapes : l'isolement et la lyse des cellules, la rétrotranscription de leurs ARNm en ADNc, l'amplification des ADNc puis leur séquençage à haut débit. Les données de séquençage obtenues sont ensuite nettoyées, normalisées et analysées grâce à des outils bioinformatiques appropriés (Figure 5).

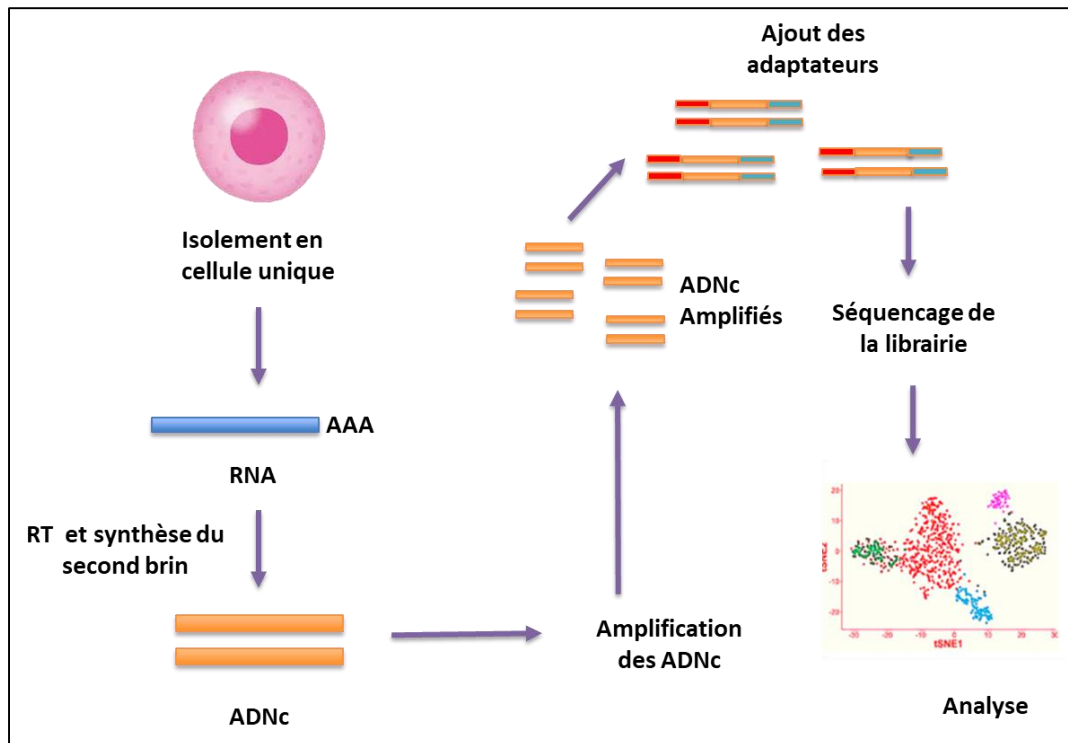


Figure 5- Grandes étapes du scRNA-seq

Les cellules sont tout d'abord isolées et les ARNm sont captés puis rétrotranscrits en ADNc. Les ADNc sont ensuite amplifiés pour permettre l'identification des gènes exprimés. Puis, les adaptateurs de séquençage sont ajoutés et les ADNc sont séquencés. Les résultats du séquençage seront finalement traités avec l'aide de différentes approches bioinformatiques.

Isolement des cellules

L'isolement de cellules uniques constitue la première étape cruciale du scRNA-seq. Afin d'isoler les cellules, plusieurs approches ont été proposées (Figure 6).

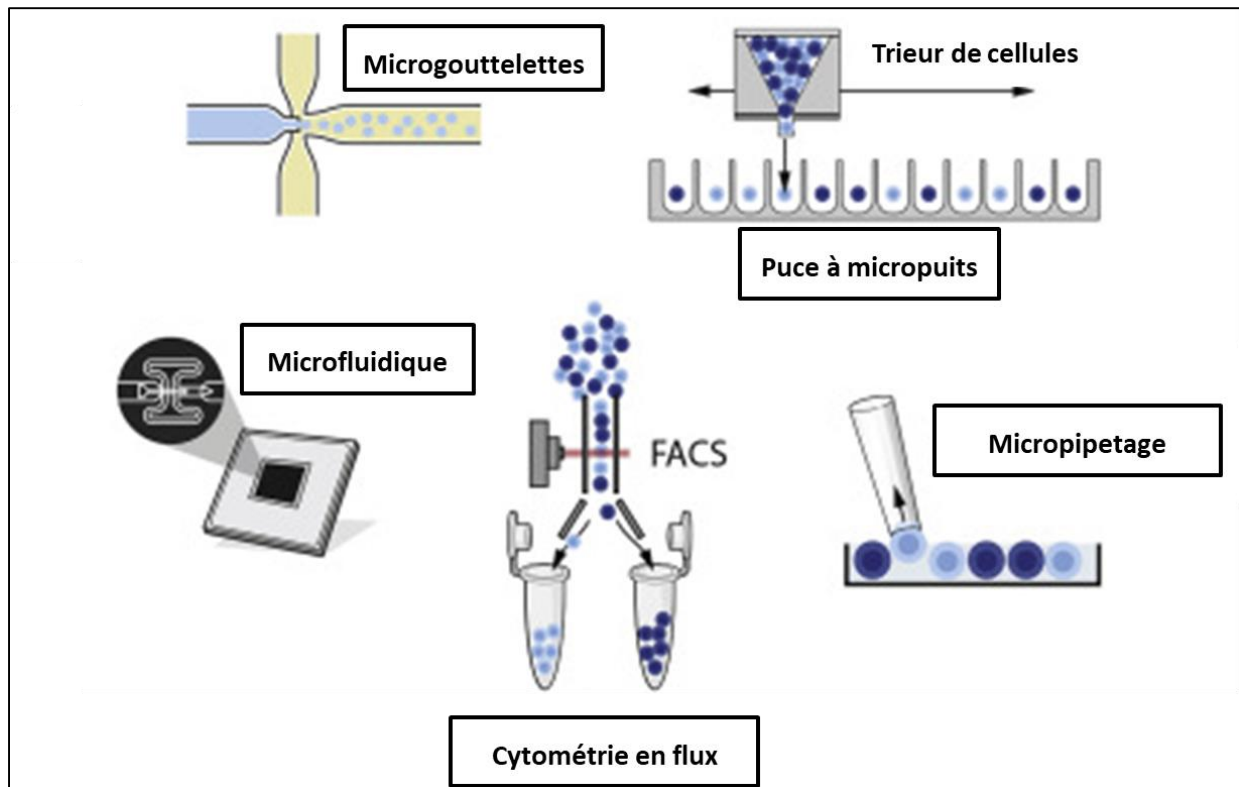


Figure 6- Différentes approches d'isolement et de capture des cellules uniques pour le scRNA-seq

Figure adaptée de (Hedlund and Deng 2018) .

Tout d'abord, le micropipetage (manuel ou automatisé) a été utilisé pour capturer des cellules à partir d'échantillons contenant un petit nombre de cellules. Une fois les cellules isolées, elles peuvent être visualisées au microscope pour s'assurer que seules des cellules uniques de haute qualité sont bien isolées. Cependant, cette approche prend du temps et a un faible débit.

Par ailleurs, la cytométrie en flux ou FACS (Fluorescence Activated Cell Sorting) est communément utilisée quand il s'agit d'isoler un nombre plus grand de cellules dans l'échantillon. La cytométrie en flux permet d'isoler les cellules en fonction de leur morphologie ou en fonction de la présence de marqueurs de surface.

Une autre approche basée sur une technologie en microfluidique est aussi largement employée. Il s'agit du module C1 de Fluidigm qui rend la visualisation des cellules isolées en

microscopie possible et qui permet d'automatiser la lyse unicellulaire, l'extraction des ARN et la synthèse des ADNc pour un maximum de 800 cellules sur une seule puce.

D'autres approches d'isolement de cellules basées sur des microgouttelettes telles que les technologies Drop-seq (Droplet-Sequencing) et 10X Genomics sont également utilisées (Macosko et al. 2015) (Weisenfeld et al. 2017). Ces techniques consistent à encapsuler des cellules uniques avec des billes de capture d'ARNm à code-barres dans des gouttelettes de l'ordre du nanolitre. Le volume de réaction très inférieur requis par ce système par rapport aux chambres microfluidiques standard permet la manipulation et le criblage de milliers à millions de cellules à un coût réduit.

Par ailleurs, les cellules peuvent être simplement isolées dans des micropuits chargés avec des billes magnétiques comme dans la technologie Cyto-Seq (Fan, Fu, and Fodor 2015). Cette méthode utilise une surface contenant jusqu'à 100 000 micropuits. Une suspension de cellules uniques et de billes magnétiques sont déposés à une densité qui permet à une seule cellule et à une seule bille d'occuper chaque micropuit.

En général, les cellules isolées sont lysées immédiatement dans un tampon hypotonique et une capture des ARNm poly(A)+ est effectuée à l'aide d'amorces poly (dT).

Préparation des ADNc

La deuxième étape du scRNA-seq consiste à préparer les ADNc à partir des ARNm. Généralement, une version modifiée de la transcriptase inverse de Moloney murine leukemia virus (MMLV) présentant une faible activité RNase H et une thermostabilité accrue est utilisée pour la synthèse du premier brin (Arezi and Hogrefe 2009). Le second brin peut être généré en utilisant soit la queue poly(A) (Sasagawa et al. 2013) soit par un mécanisme de Template switching (Islam et al. 2011) (Figure 7). Lorsqu'elle atteint l'extrémité 5' de la matrice d'ARN, l'activité transférase terminale de la transcriptase inverse de MMLV ajoute quelques nucléotides supplémentaires (principalement de la désoxycytidine) à l'extrémité 3' du brin d'ADNc neosynthétisé. Une séquence d'oligonucléotide Template Switching (TS), peut alors s'hybrider et

servir de point de départ pour la synthèse du brin complémentaire. L'ADNc résultant contient une séquence universelle à l'extrémité 3' (complémentaire de la séquence TS) permettant une amplification par PCR de l'ADNc sur toute sa longueur.

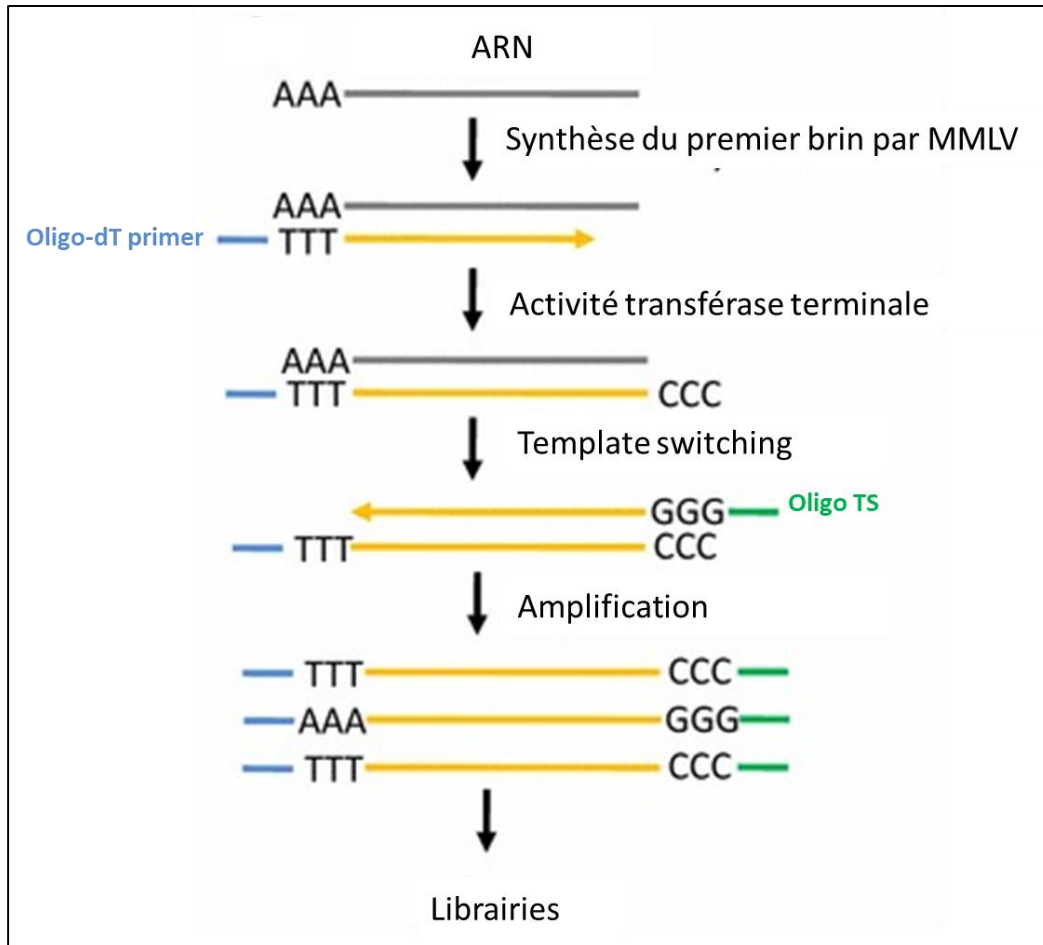


Figure 7- Principe du Template switching

La synthèse du premier brin est effectuée en partant d'un ARNm polyadénylé comme matrice et d'une amorce oligo-dT. En atteignant l'extrémité 5' de la matrice d'ARN, la transcriptase inverse de MMLV introduit des résidus de cytosine. Une amorce TS complémentaire se fixe sur ces derniers. L'oligo TS sert comme un point de départ pour la synthèse du brin complémentaire. Les ADNc obtenus contenant les séquences d'amorces connues aux deux extrémités peuvent être amplifiés et envoyés au séquençage. Figure adaptée de (J. Wang and Song 2017).

Amplification des ADNc

Une fois synthétisés, les ADNc issus de chaque cellule et en toutes petites quantités nécessitent d'être amplifiés. L'amplification est généralement réalisée par PCR classique qui assure une

amplification exponentielle des fragments mais elle peut aussi se faire par transcription *in vitro* (IVT) (Hashimshony et al. 2012).

L'IVT est une technique d'amplification qui utilise l'ARN polymérase T7 et qui permet une amplification linéaire (Figure 8). Lors de la première étape de reverse transcription, une amorce qui contient une séquence poly-T, un code-barre unique, un adaptateur Illumina et une séquence reconnue par la polymérase T7 est utilisée. L'amorce s'hybride sur la queue poly-A de l'ARNm, et le premier et deuxième brins d'ADNc sont synthétisés par la reverse transcriptase. Les différents échantillons reconnus par le code-barre peuvent alors être regroupés. L'amplification des ADNc est ensuite réalisée grâce à la polymérase T7 créant des ARN antisens (ARNa). Ces derniers sont ensuite fragmentés et un autre adaptateur Illumina est ligué aux extrémités 3'. Un autre cycle de reverse transcription permet d'obtenir des fragments d'ADNc double brin. Enfin, les fragments contenant les codes-barres et les deux adaptateurs Illumina, c'est-à-dire les fragments les plus proches de l'extrémité 3' des transcrits, sont sélectionnés pour séquençage.

Bien que l'utilisation de l'IVT assurant une amplification linéaire au lieu de la PCR réduit le biais d'amplification, le protocole IVT nécessite une étape supplémentaire de reverse transcription. Ainsi, Ce protocole rajoute une nouvelle étape de capture des queue poly A des ARNm par les oligomères poly-dT qui est fortement biaisée vers l'extrémité 3' des transcrits (Grün and van Oudenaarden 2015) (Kolodziejczyk et al. 2015a).

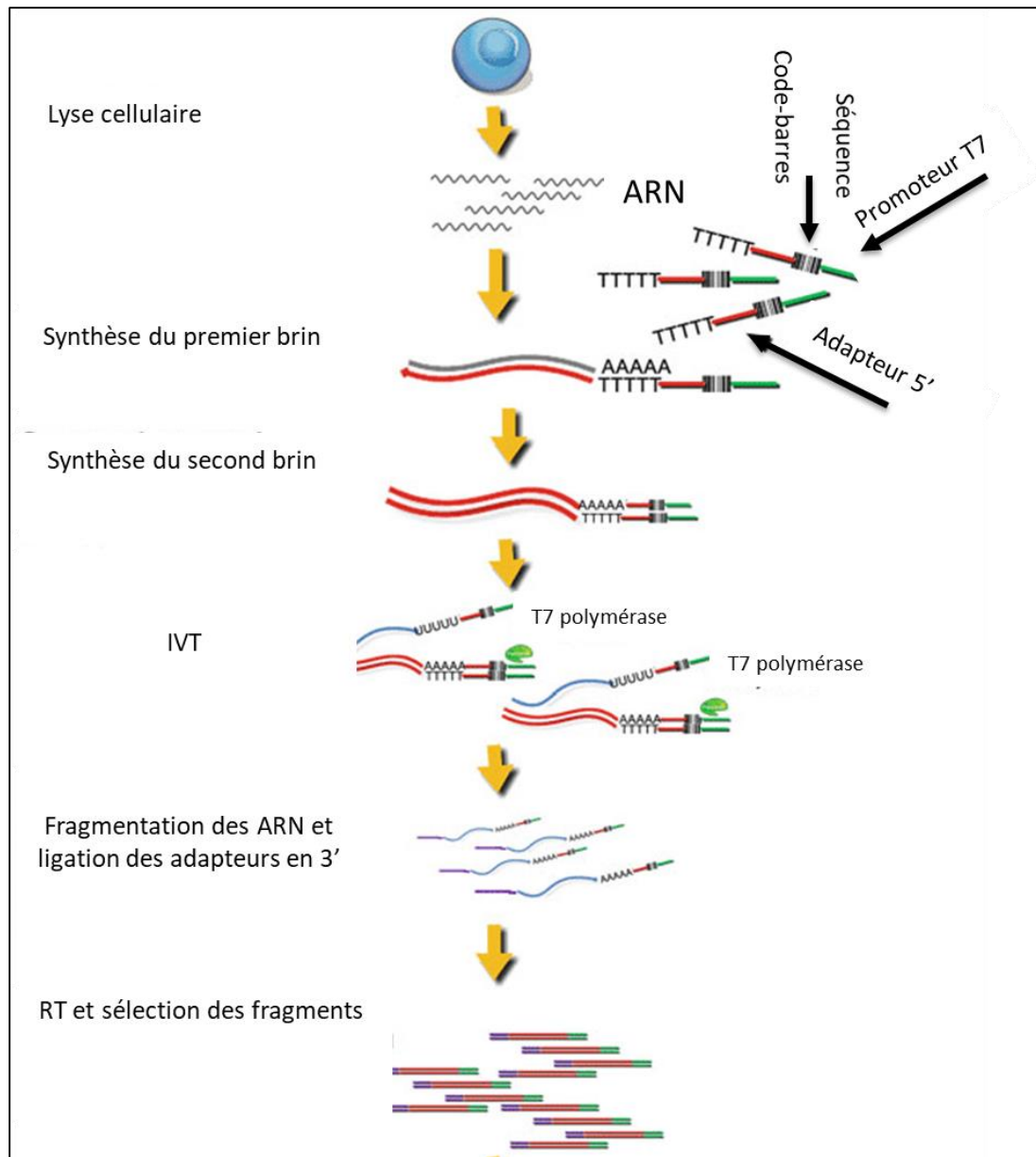


Figure 8- Principe de l'IVT

Après la lyse des cellules, une amorce contenant un promoteur T7, une séquence poly(T) et une séquence de code-barres est utilisée pour la transcription inverse. Après synthèse du second brin des ADNc, les échantillons sont regroupés et la transcription in vitro est effectuée à l'aide de la polymérase T7. Les ARN amplifiés sont ensuite fragmentés, Un autre cycle de reverse transcription permet d'obtenir des fragments d'ADNc double brin. Ces derniers sont purifiés et sélectionnés pour être séquencés. Figure adaptée de (K. Chen, Dong, and Wu 2015)

Utilisation des code-barres cellulaires et moléculaires dans le scRNA-seq

Tang et al ont été les premiers à décrire la technique initiale du scRNAseq utilisée dans leur étude pour caractériser le transcriptome complet d'une cellule d'un blastomère (Tang et al. 2009). Leur approche a permis d'obtenir un profil transcriptomique à partir d'une seule cellule grâce à une profondeur de séquençage très importante. Bien que leur travail fût particulièrement unique et intéressant, il était limité à la caractérisation d'une seule cellule du fait de l'absence de différents code-barres moléculaires permettant de multiplexer les analyses.

La notion d'analyse multiplex est apparue en 2011 (Islam et al. 2011) avec l'innovation de la technique STRT-seq (Single-cell Tagged Reverse Transcription) rendant possible l'analyse de 96 cellules uniques en parallèle. Tout d'abord les cellules ont été isolées, à l'aide d'un sélecteur de cellules semi-automatique, dans une plaque 96 puits contenant du tampon de lyse. Dans chaque puits, les ADNc ont été synthétisés par RT en incorporant un code-barre unique, différent dans chaque puits, à l'extrémité 3' par la technologie de Template switching (Zhu et al. 2001). Une fois étiquetées avec les code-barres, les ADNc de chaque cellule ont été mélangés, ce qui a permis d'améliorer considérablement et efficacement les étapes de préparation des librairies et de séquençage.

Une des plus grandes innovations mises au point pour le RNA-seq est celle des UMI (Unique Molecular Identifier). Les UMI sont des séquences de bases aléatoires, généralement de 5bp de taille, soit 1024 séquences possibles. Le principe consiste en l'inclusion de ces identificateurs moléculaires uniques pendant l'étape de réverse transcription afin de créer une identité distincte pour chaque molécule d'ARNm (Islam et al. 2014). Cela permet alors, d'estimer l'efficacité avec laquelle les molécules d'ARNm de départ, sont échantillonnées, d'identifier le biais d'échantillonnage et, surtout, de corriger les effets du biais d'amplification induits lors de l'étape d'amplification par PCR. Ce dernier point est particulièrement important surtout lorsque de nombreux cycles de PCR sont nécessaires pour obtenir suffisamment de matériel nécessaire à l'étape de séquençage.

Après la PCR, les molécules partageant le même UMI sont supposées être dérivées de la même molécule initiale (Figure 9).

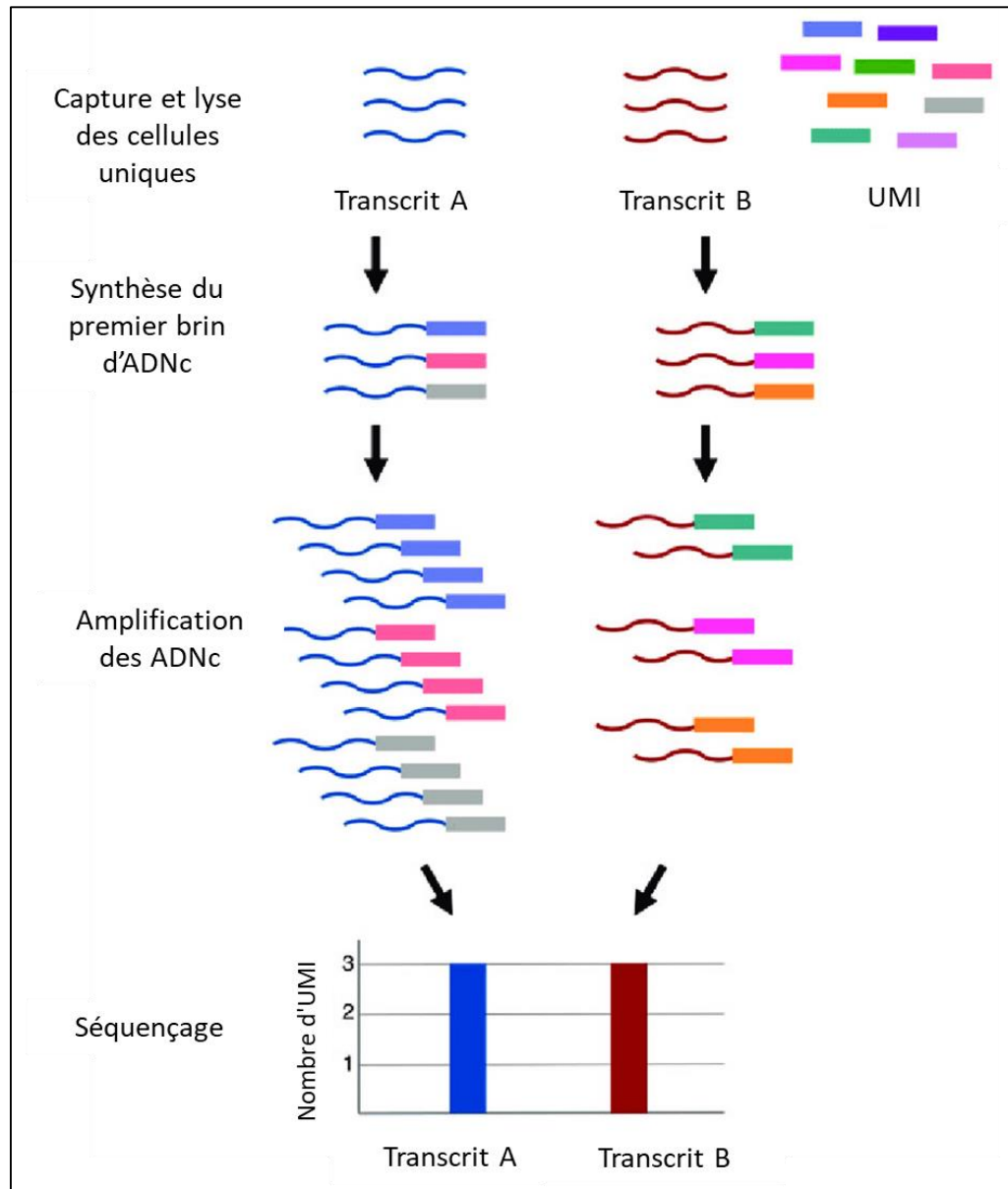


Figure 9- Principe des UMI

Les UMI sont des séquences de bases aléatoires utilisées pour marquer chaque molécule d'ADNc pendant l'étape de réverse transcription, facilitant ainsi l'identification des duplicas de PCR. Le principe d'utilisation des UMI consiste à compter le nombre de séquences UMI uniques associées à chaque transcript au lieu de compter le nombre de séquences qui correspondent à un transcript particulier. Cette stratégie permet d'obtenir une estimation précise de la quantité relative des ARNm dans l'échantillon original. Figure adaptée de (Kolodziejczyk and Lönnberg 2018).

Récemment, une nouvelle technique appelée SPLIT-seq (split-pool ligation-based transcriptome sequencing) a été mise au point (Rosenberg et al. 2018). Cette technologie est basée sur le principe d'ajout de codes-barres à plusieurs étapes permettant d'analyser jusqu'à 100 000 cellules en parallèle.

Dans cette technique, les cellules sont fixées et perméabilisées puis distribuées dans des puits. Dans chaque puits, les ARNm sont rétrotranscrits grâce à une amorce polyT qui contient un code-barre spécifique à chacun des puits. Cette première étape permet de manipuler un grand nombre des cellules et une grande variété d'échantillons en parallèle, minimisant ainsi les effets de lots. Les échantillons de tous les puits sont ensuite regroupés et redistribués dans de nouveaux puits, où un deuxième code-barre spécifique à chaque puits est ajouté par ligation. Cette étape est répétée pour permettre l'ajout sur les ADNc d'un troisième code-barre toujours spécifique à chaque puits. L'étape est encore répétée une quatrième fois, en utilisant cette fois-ci un nouveau code-barre contenant également un UMI. Pour finir, les cellules sont regroupées et divisées en sous-groupes puis des codes-barres de séquençage sont introduits par PCR. Grâce à ces 4 étapes d'indexation, on peut générer presque 20 millions de combinaisons des 4 code-barres, ce qui permet de multiplexer presque 1 million de cellules (Figure 10).

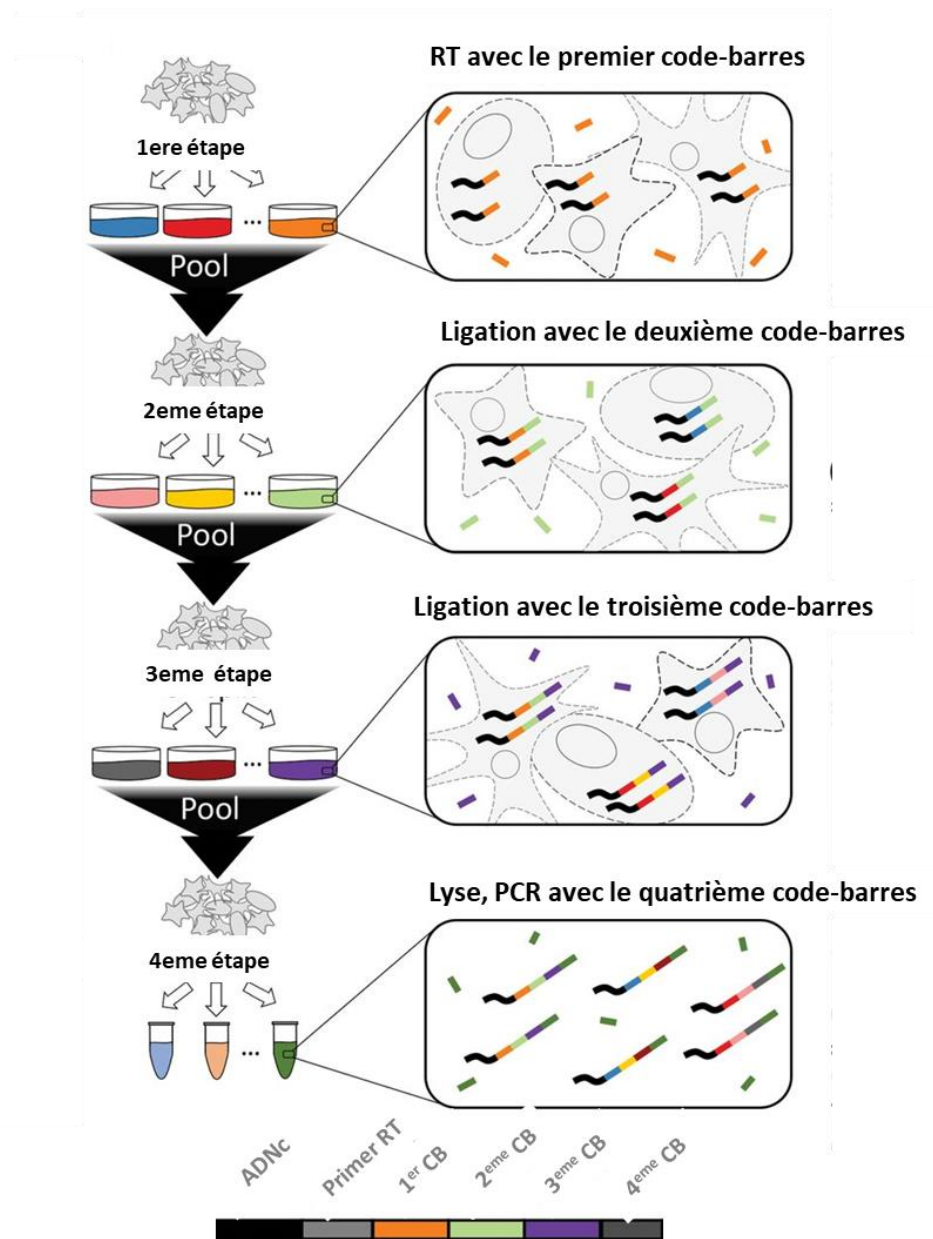


Figure 10-Principe du SPLIT-seq

Le principe du SPLIT-seq repose sur l'ajout de codes-barres à différentes étapes. Dans la première étape, les ARNm sont rétrotranscrits grâce à une amorce polyT qui contient un code-barre spécifique dans chacun des puits. Deux autres codes-barres sont ensuite rajoutés successivement par ligation aux ADNc. Un quatrième code-barre est ajouté aux molécules d'ADNc par PCR pendant la préparation de la bibliothèque de séquençage. Le schéma du bas montre la molécule finale d'ADNc contenant les différents code-barres (CB: Code-Barre). Figure adaptée de (Rosenberg et al. 2018).

Normalisation des données de scRNA seq

Les techniques de scRNA-seq peuvent présenter des limites dans la sensibilité de détection qui dépend de l'efficacité de capture des ARNm, de la RT et de la profondeur de séquençage.

Afin de limiter les biais et de déterminer justement la quantité d'ARN, la normalisation des données de scRNA-seq constitue une étape indispensable. Il existe deux grands types de normalisation : intra-échantillons et inter-échantillons (Vallejos et al. 2017). La normalisation intra-échantillon vise à supprimer les biais spécifiques aux gènes dépendant par exemple de la teneur en GC et de la longueur des transcrits. Afin de limiter les biais inter-échantillons, des molécules synthétiques externes ajoutées en quantité connue appelées ERCC (External RNA Controls Consortium) et les UMI (Stegle, Teichmann, and Marioni 2015) sont utilisés pour normaliser les données (Bacher and Kendzierski 2016).

1.2.1.3 Techniques de scRNA-seq les plus couramment utilisées

Le scRNA-seq est devenu une approche incontournable pour disséquer l'hétérogénéité cellulaire et d'identifier les sous-populations des différents types de cellules.

Différentes méthodes et technologies scRNA-seq ont été mises en place. Une des différences majeures entre les méthodes scRNA-seq est que certaines d'entre elles permettent le séquençage des transcrits sur toute la longueur (ou presque) (comme le Smart-seq et Smart-seq2), tandis que d'autres ne capturent et ne permettent le séquençage que d'une seule extrémité (comme le CEL-seq 1 et 2, MARS-seq, Drop-seq et Chromium 10X).

Dans cette partie, nous allons détailler les 5 techniques de scRNA-seq les plus couramment utilisées. Ces techniques sont présentées dans le tableau 1 ci-dessous.

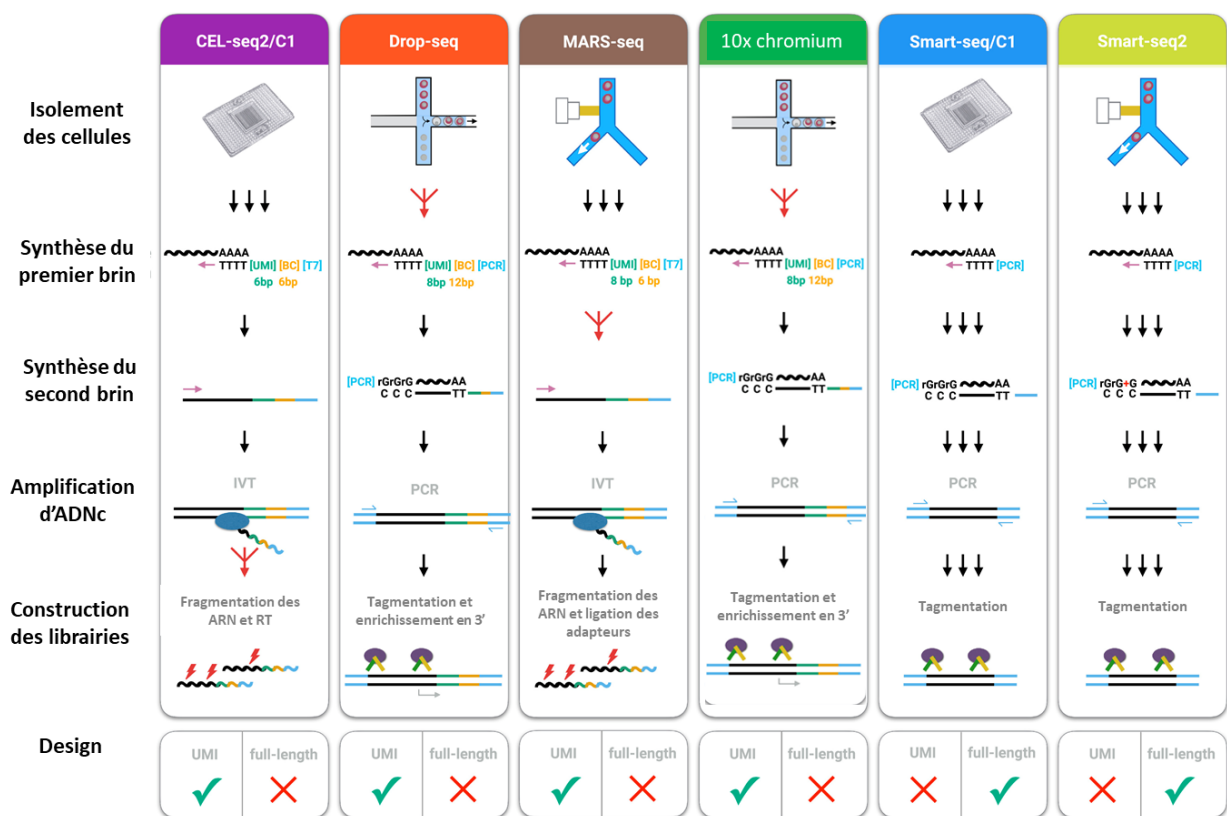


Tableau 1- Les différentes techniques de scRNA-seq.

Les différentes étapes de préparation des librairies pour chaque technique sont présentées. Adapté de (Ziegenhain et al. 2017).

MARS-seq

Le MARS-seq (Massively parallel single cell RNA-Seq) est une des premières techniques qui permet un haut niveau de multiplexage pour l'analyse du transcriptome en cellules uniques. Le protocole consiste à introduire des UMI dans les amorces oligo-dT en plus de barres-codes spécifiques aux cellules et aux plaques. Cette stratégie a permis d'étudier les profils transcriptomiques entre 10^2 et 10^3 cellules simultanément et de détecter entre 10^4 et 10^5 reads par cellule (Jaitin et al. 2014) (Haque et al. 2017). C'est est la technique que nous avons utilisée dans nos études qui seront présentées dans la partie Résultats.

Drop-seq

La méthode Drop-seq utilise un dispositif microfluidique comportant un canal qui contient les cellules uniques à analyser et un autre canal qui contient des billes (Macosko et al. 2015). Chaque bille contient plus de 108 séquences individuelles qui partagent une même amorce PCR, un même code-barre cellulaire, une amorce oligo(dT) mais qui possèdent des identifiants moléculaires uniques (UMI) différents. Le dispositif microfluidique, permet d'encapsuler chaque cellule individuelle avec une bille dans une gouttelette d'eau dans une phase huileuse. Immédiatement après la formation des gouttelettes, les cellules sont lysées et les ARNm libérés, puis hybridés à la surface des billes de microparticules. Les ARNm hybridés peuvent être facilement manipulés et regroupés pour l'étape de réverse transcription, l'amplification par PCR et la préparation de la bibliothèque dans un seul tube.

Le Drop-seq permet d'analyser entre 10^3 et 10^4 cellules simultanément et peut détecter entre 10^4 et 10^5 reads par cellule, réduisant considérablement le coût de l'analyse du transcriptome (Haque et al. 2017).

CEL-seq et CEL-seq2

Le CEL-seq (Cell Expression by Linear amplification and Sequencing) est la première méthode à utiliser à la fois des code-barres cellulaires et l'amplification IVT au lieu de l'amplification PCR (Hashimshony et al. 2012). L'introduction précoce de code-barres spécifiques à chaque cellule offrant la possibilité de multiplexer permet de réduire le coût d'une analyse à grande échelle et de réduire les biais techniques. La deuxième version de la méthode, CEL-seq2, a amélioré la quantification de l'expression des gènes grâce à l'introduction d'UMI dans l'amorce de RT (Hashimshony et al. 2016).

Grâce au protocole CEL-seq, il est possible d'analyser entre 10^2 et 10^3 cellules parallèlement et de détecter entre 10^4 et 10^5 reads dans chaque cellule (Haque et al. 2017).

10X Genomics

La méthode 10X Genomics utilise un dispositif Chromium permettant la quantification des ARNs dans dix mille cellules en une seule expérience. Il peut traiter 8 échantillons différents simultanément grâce aux 8 canaux indépendants présents dans la puce microfluidique. D'abord, les cellules passent dans ces canaux contenant des billes de gel liées chacune à une amorce PCR, une séquence oligo(dT), un UMI, et un code-barre cellulaire. Ensuite, la formation d'une émulsion de billes de gel isole les cellules de sorte que chaque gouttelette pleine contient une cellule unique, une bille de gel à code-barres, des réactifs de lyse cellulaire et de RT. En raison de l'approche de capture aléatoire, l'efficacité de capture des cellules est limitée à 65 % des cellules chargées, ce qui produit un nombre important de gouttelettes "vides" (contre moins de 5 % pour le protocole Drop-seq). La lyse cellulaire commence lorsque les cellules sont encapsulées. Les ADNc produits contiennent donc un UMI et un code-barre cellulaire. Enfin, les ADNc sont regroupés et amplifiés par PCR pour la préparation de la librairie et son séquençage.

Le 10X Genomics permet d'analyser entre 10^3 et 10^4 cellules en parallèle et peut détecter entre 10^4 et 10^5 reads par cellule (Haque et al. 2017).

Smart-seq et Smart-seq 2

Le protocole Smart-seq a été le premier à intégrer la technologie de Template switching et a été développé pour permettre le séquençage des transcrits sur toute la longueur (ou presque). Le protocole Smart-seq a ensuite été optimisé pour aboutir au protocole Smart-seq2 qui est aujourd'hui largement utilisé (Picelli et al., 2014b) (Tirosh et al. 2016). Le Smart-seq2 utilise des séquences d'oligonucléotides Template Switching (TS) modifiés ainsi qu'une polymérase haute-fidélité pour l'amplification. Ces améliorations ont permis d'augmenter de manière significative la couverture intégrale des transcrits et d'améliorer ainsi considérablement le rendement. Le Smart-seq2 permet ainsi d'analyser entre 10^2 et 10^3 cellules et peut détecter un nombre très important de gènes exprimés dans chacune des cellules analysées (10^6 reads /cellule en moyenne) (Haque et al. 2017).

1.2.1.4 Comparaison des différentes technologies de scRNA-seq

La meilleure méthode de scRNA-seq est celle qui permet de quantifier tous les transcrits cellulaires codants et non codants avec une grande efficacité en révélant les changements les plus subtils dans l'expression des gènes. L'une des principales limites des méthodes actuelles est que toutes reposent sur une étape de transcription inverse de l'ARN en ADNc, suivie d'une amplification par PCR ou IVT avant le séquençage. Ces deux approches sont sujettes à des biais qui ne peuvent être évités puisque les technologies actuelles de séquençage n'offrent pas la possibilité de séquencer directement l'ARN en raison aussi d'un nombre trop faible de molécules par cellule.

Des études comparatives ont été faites sur les principales techniques de scRNA-seq dans le but d'évaluer leurs efficacités et leurs performances (Kolodziejczyk et al. 2015b) (Picelli 2017). Dans une étude récente, Ziegenhain et al ont comparé des protocoles différents de scRNA-seq : le Smart-seq2, le CEL-seq2, le MARS-seq et le Drop-seq sur le même échantillon de cellules souches embryonnaires de souris en calibrant le nombre de cellules et la profondeur de séquençage (Ziegenhain et al. 2017). Ainsi, les auteurs ont pu comparer directement la sensibilité, les niveaux de bruit et les coûts des différents protocoles. Grâce à cette étude, ils ont mis en évidence que le Smart-seq2 permet de détecter un plus grand nombre de gènes exprimés que d'autres technologies. En outre, ils ont montré que les techniques de CEL-seq2, Drop-seq, MARS-seq permettent une quantification des niveaux d'ARNm avec moins de bruit d'amplification grâce à l'utilisation des UMI. De plus, le Drop-seq apparaît plus efficace pour la quantification du transcriptome d'un grand nombre de cellules, tandis que le MARS-seq et Smart-seq2 sont plus efficaces lorsqu'il s'agit d'analyser moins de cellules.

D'autre part, par rapport aux protocoles permettant le séquençage que d'une seule extrémité des transcrits, il a été démontré que les méthodes de séquençage des transcrits dans leur intégralité sont plus performantes pour détecter des transcrits résultant d'épissages différents et pour repérer les SNP (Single Nucleotide Polymorphisms) ainsi que les mutations (Ramsköld et al. 2012) (Ziegenhain et al. 2017). Néanmoins, les technologies basées sur les gouttelettes comme le Drop-seq ou le 10X Genomics sont bien adaptées à l'analyse d'un grand

nombre de cellules et donc de tissus complexes comme des échantillons de tumeurs car elles permettent de bien identifier les sous-populations cellulaires tout en fournissant un coût de séquençage par cellule très inférieur à celui des techniques de séquençage des transcrits sur toute sa longueur (G. Chen, Ning, and Shi 2019).

Finalement, il apparaît important de déterminer en amont la méthode de scRNA-seq la plus appropriée pour répondre à la question biologique posée. Ce choix méthodologique doit considérer les avantages de chaque technique, le nombre de cellules à séquencer mais aussi le coût financier.

1.2.2 RT-qPCR en cellule unique

Les RT-qPCR en cellule unique ou sc RT-qPCR sont des méthodes les plus développées pour le profilage de l'ARNm unicellulaire. Ces techniques nécessitent plusieurs étapes expérimentales comme l'isolement des cellules, la préamplification des ARNm et la quantification des ADNc.

Durant mon travail de thèse, j'ai utilisé l'approche de sc-RNAseq de FLuidigm qui est une technique de mesure d'expression génétique par RT-qPCR à haut débit basée sur des réseaux micro fluidiques (Sanchez-Freire et al. 2012) (Figure 11). La première étape de sc-RNAseq de FLuidigm nécessite un isolement des cellules dans des plaques 48 ou 96 puits contenant un tampon de lyse. Une fois les cellules lysées, les ARNm (Ce sont les types d'ARN sur lesquels je me concentre dans mon étude) sont rétrotranscrits puis les ADNc sont préamplifiés. D'une façon intéressante, ce système permet des mesures simultanées d'expression génique par PCR en temps réel dans une seule puce, tout en nécessitant moins de pipetages que lors de l'utilisation d'une plaque classique de 96 puits (Spurgeon, Jones, and Ramakrishnan 2008). En effet, dans la puce nanofluidique contenant des réseaux fluidiques (96.96 DynamicArray IFC (Fluidigm)), 96 puits sont chargés avec 96 couples de primers et les 96 autres sont chargés avec les échantillons d'ADNc. Les primers et les échantillons sont ensuite « pressurisés » dans la puce en utilisant un contrôleur IFC HX (Fluidigm). 9 216 réactions (96 X96) de qPCR en parallèle sont ensuite réalisées dans le Biomark [™] HD (Fluidigm). Finalement, les données brutes sont traitées avec le logiciel

Real-Time PCR Analysis de Fluidigm. De plus, afin de limiter les variabilités techniques, des ARN synthétiques, appelés Spikes, sont ajoutés de manière homogène dans chaque puit et seront ensuite utilisés pour normaliser les données de sc RT-qPCR.

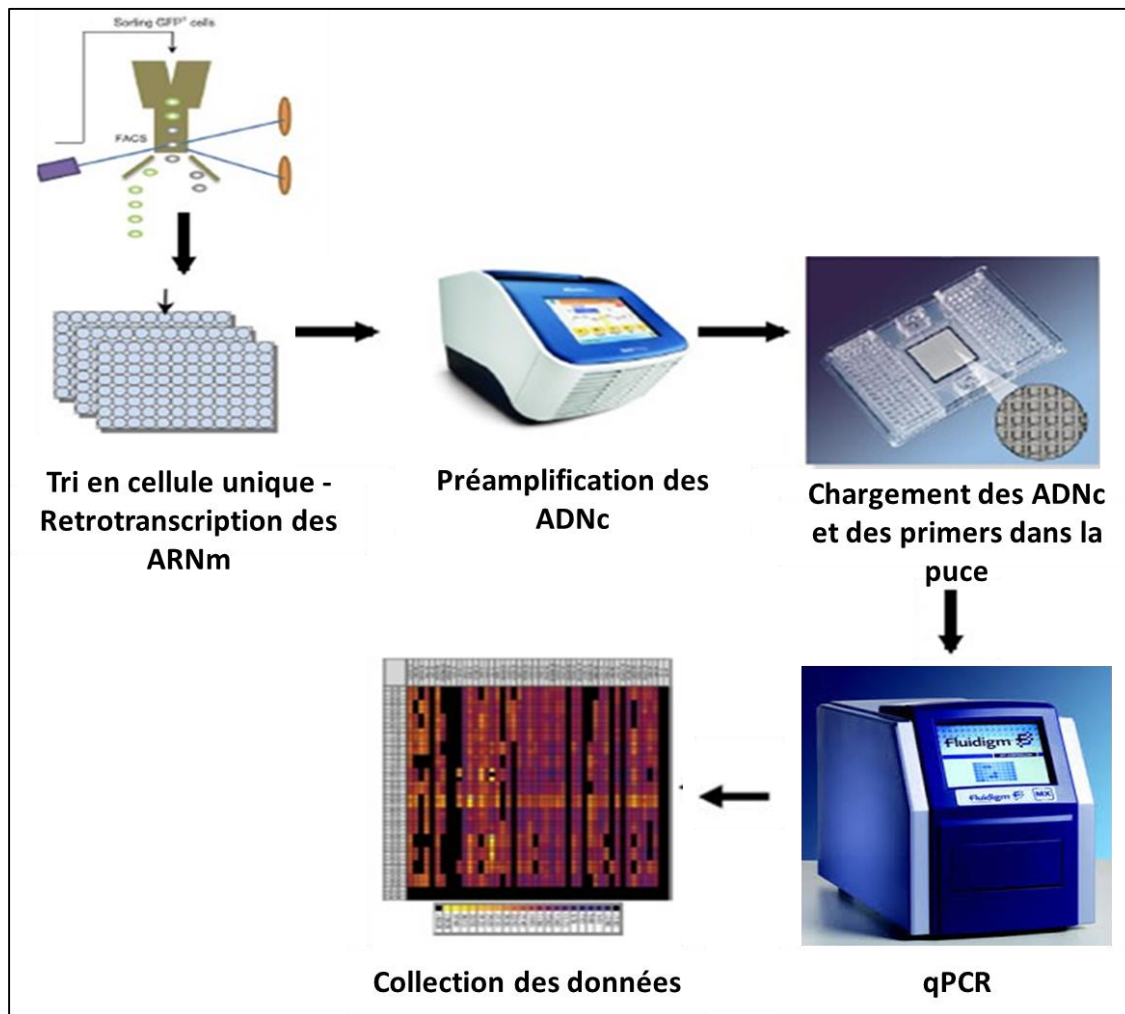


Figure 11- Protocole de scRT-qPCR-Fluidigm

Les cellules sont d'abord isolées dans des plaques contenant du tampon de lyse. Une fois lysées, les ARN des cellules sont rétrotranscrits puis les ADNc préamplifiés. Ces derniers sont ensuite chargés avec les couples de primers dans la puce microfluidique puis quantifiés par qPCR dans le Biomark TM HD (Fluidigm).

1.3 Approches bio-informatiques pour l'analyse des données d'expression obtenues en cellule unique

Après la normalisation des données, différentes méthodes d'analyse bioinformatique peuvent être utilisées afin de donner du sens aux données et permettre la description du système biologique sous-jacent.

1.3.1 Méthodes de réduction de dimensions

L'analyse du transcriptome en cellules uniques génère une matrice d'expression génique représentant la quantité de chaque transcrit (ou le nombre de reads de chaque gène) dans chaque cellule. Afin d'extraire les informations importantes de cette matrice à multiples dimensions (centaines/milliers de gènes et centaines/milliers de cellules), des algorithmes de réduction de dimension spécialisés sont utilisés. Il s'agit plus exactement de méthodes statistiques, dont le rôle est de changer le repère dans lequel les données sont projetées initialement. Ces techniques servent à créer alors un sous espace de faible dimension, compréhensible par tous, qui cherche à conserver le maximum d'informations contenues dans les données initiales.

La PCA pour Principal Components Analysis est une méthode de réduction de dimension linéaire non supervisée largement utilisée. Cette méthode consiste à identifier l'hyperplan le plus proche des données, en créant un sous-espace à n dimensions, puis à projeter les cellules dans ce sous-espace. Dans le cas d'une réduction à 2 dimensions, ce sous-espace est défini par deux vecteurs : la composante principale 1 (PC1) et la composante principale 2 (PC2). La PCA cherche à expliquer le maximum de variabilité dans l'ensemble des données (Lever, Krzywinski, and Altman 2017).

D'autres méthodes non linéaires de réduction de dimension, comme la méthode t-distributed Stochastic Neighbor Embedding (tSNE) ou Uniform Manifold Approximation and Projection (UMAP) sont également utilisées (Maaten and Hinton 2008) (McInnes, Healy, and Melville 2020).

L'approche tSNE trouve des modèles dans les données en se basant sur la similarité entre les cellules. Cette similarité est calculée comme une probabilité conditionnelle que les cellules A et B soient voisines. L'algorithme tente ensuite de minimiser la différence entre ces probabilités conditionnelles ce qui permet d'ajuster la position de chaque cellule dans un espace à deux dimensions.

L'UMAP tente de faire correspondre les cellules à un système de coordonnées global tout en préservant la structure locale des données. Elle possède une performance comparable à la méthode tSNE, mais elle est légèrement meilleure pour la préservation des distances et se distingue par son temps d'exécution plus rapide (Becht et al. 2018). Par rapport à un algorithme de réduction linéaire de dimension, l'UMAP et tSNE permettent une meilleure visualisation des sous-types rares de cellules.

Un exemple d'analyse des mêmes données à l'aide des approches PCA, t-SNE et UMAP est présenté dans la figure 12. Dans cet exemple, les chercheurs ont analysé par ces 3 méthodes, des données de scRNA-seq (10X Genomics) obtenues sur 2700 cellules mononucléaires de sang périphérique. L'approche UMAP a montré une meilleure préservation de la structure locale (distances des cellules au sein d'une grappe) et de la structure globale des données (distances entre grappes), soulignant sa capacité à mieux différencier des populations cellulaires même subtilement différentes.

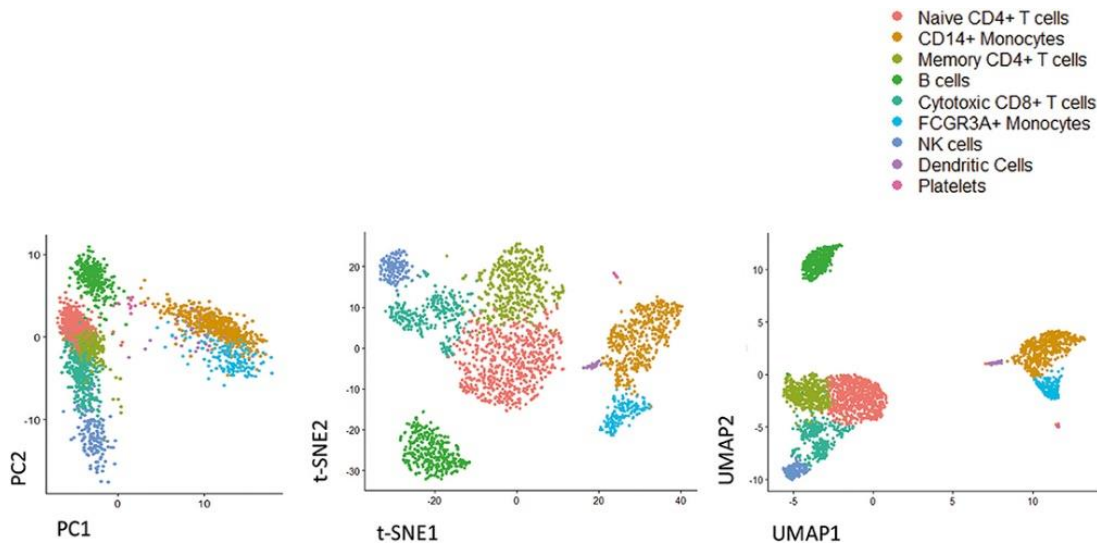


Figure 12- Exemple d'analyse de données de scRNA-seq par 3 techniques différentes de réduction de dimensions

Un jeu de données scRNA-seq (PBMC, 10X Genomics) est analysé par les approches PCA, t-SNE et UMAP. (Gibellini et al. 2020).

1.3.2 Approche du Différentiel d'Expression (DE)

Une question courante posée aux données d'expression est de savoir quels gènes sont exprimés de manière différentielle entre deux conditions expérimentales. Le test du Différentiel d'Expression (DE) peut être très utile pour interpréter la différence biologique entre deux conditions comparées. L'analyse de l'expression différentielle consiste à prendre les données normalisées de séquençage et à définir si, pour un gène donné, une différence observée dans les comptes de lecture est significative, c'est-à-dire si elle est supérieure à ce qui serait attendu du fait d'une variation strictement aléatoire.

Récemment, plusieurs méthodes de DE ont été proposées pour les données de scRNA-seq, comme le MAST (Model-based Analysis of Single Cell Transcriptomics) (Finak et al. 2015), le Census (Qiu et al. 2017) (Qiu et al., 2017), le DEsingle (Miao et al. 2018) et le BCseq (L. Chen and Zheng 2018)(Chen et Zheng, 2018). L'approche MAST est basée sur l'ajustement de modèles linéaires et le test du rapport de vraisemblance. Le Census modélise le nombre des isoformes d'épissage de chaque gène avec un modèle linéaire sous la forme d'une distribution Dirichlet-

multinomiale. Le DEsingle utilise le modèle binomial négatif et estime les « vrais » zéro dans les données. Enfin, le BCseq atténue le bruit technique d'une manière adaptée aux données.

1.3.3 Approche Sparse PLS

Il existe différentes façons d'effectuer une sélection de variables dans de grands ensembles de données. Dans le cas de multiples variables fortement corrélées, la régression partielle des moindres carrés appelée Sparse PLS (Partial least squares regression) (Antoniadis, Lambert-Lacroix, and Leblanc 2003) (Dai, Lieu, and Rocke 2006) semble appropriée pour expliquer la structure de variance/covariance des données en utilisant des combinaisons linéaires des variables originales. La Logit SPLS est une analyse statistique avec une version adaptée de la Sparse PLS (Durif et al. 2018) qui permet dans un premier temps de compresser les données, de chercher une représentation des données dans un espace de dimension plus petite, puis de prédire l'appartenance d'une cellule à une classe sur la base de son profil d'expression génique.

Afin de réaliser la prédiction par Sparse PLS, la matrice d'expression est d'abord décomposée en une matrice représentant les cellules en fonction de la valeur d'expression de chaque gène (variables indépendantes) et en vecteurs représentant chacune les conditions biologiques à comparer. L'ensemble de ces données est ensuite compressé pour former des structures latentes puis l'analyse par Logit-SPLS pour chaque paire de vecteurs est effectuée. Cela permet d'obtenir une maximisation de la covariance entre les composantes de deux ensembles de structures latentes et de calculer les combinaisons linéaires par des vecteurs de poids appelés vecteurs de chargement. Les vecteurs de chargement sont donc associés pour chaque gène (Figure 13).

Cette approche fournit en sortie des listes pondérées de gènes dont les valeurs d'expression caractérisent l'appartenance à une classe.

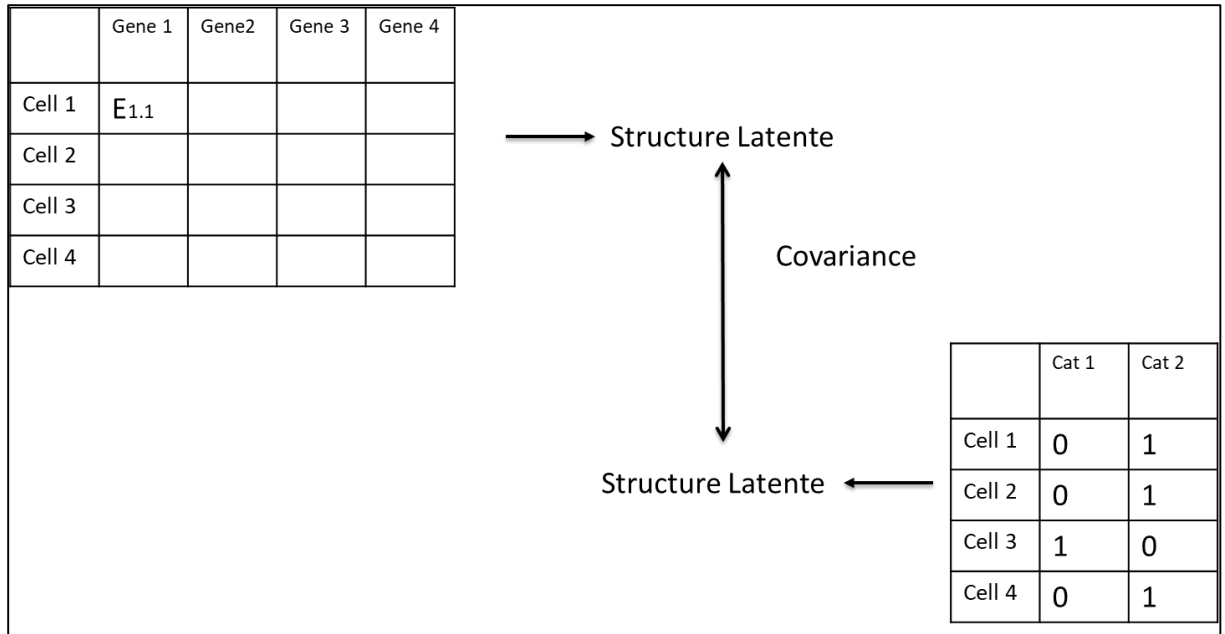


Figure 13- Principe de prédiction par Sparse PLS

La matrice d'expression est d'abord décomposée en matrice représentant les cellules en fonction de leurs valeurs d'expression (matrice à gauche) et en une autre matrice précisant l'appartenance des cellules à chacune des conditions (Cat 1 et Cat 2) (matrice à droite). L'ensemble de ces données est ensuite compressé pour former des structures latentes. Ensuite, l'approche Logit-SPLS maximise la covariance des structures latentes et calcule les vecteurs de chargement pour chaque gène.

1.3.4 Approche des Réseaux de Régulation des Gènes (GRN)

1.3.4.1 Définition des GRN

La prise de décision cellulaire, est un des processus biologiques où la stochasticité joue un rôle important. Cette variabilité pourrait s'apparenter à un phénomène aléatoire contrôlé dépendant à la fois de paramètres biophysiques régissant l'expression des gènes et de la structure des réseaux de gènes sous-jacents (Elowitz et al. 2002) (Coulon, Gandrillon, and Beslon 2010).

L'évolution des méthodes d'analyse de l'expression génique à l'échelle de la cellule unique, a permis de confirmer que les gènes ne fonctionnent pas indépendamment mais bien ensemble. En effet, les gènes influencent l'expression d'autres gènes par un réseau complexe d'interactions moléculaires. Les GRN, représentent les gènes interconnectés dont l'expression

des uns va contrôler l'expression des autres à différents niveaux (transcription, traduction, etc.) grâce à un ensemble d'interactions dynamiques.

Les GRN peuvent être générés grâce à une mesure expérimentale instantanée de l'expression des gènes dans différentes conditions expérimentales, et on parle dans ce cas de réseaux statiques. Le comportement dynamique peut quant à lui être inféré grâce à des mesures dans le temps de l'expression des gènes. Dans ce cas, on parle de GRN dynamiques qui décrivent comment les interactions entre les gènes évoluent dans le temps. Ces GRN dynamiques nécessitent une modélisation afin de retranscrire les observations dynamiques des données (Ferrazzi et al. 2007).

1.3.4.2 Importance et rôles des GRN

Les GRN ont été souvent utilisés pour mieux comprendre les états physiologiques. Par exemple, l'utilisation d'un algorithme appelé ARACNe (Algorithm For The Reconstruction of Accurate Cellular Networks) a permis d'inférer les GRN à partir de données d'expression obtenues sur puces à ARN, notamment sur une population cellulaire de Lymphocytes B humains. ARACNe cherche à analyser d'abord toutes les interactions possibles entre deux gènes puis établit les réseaux en identifiant les interactions directes entre les gènes et en éliminant les gènes avec des relations indirectes. Les résultats de cette étude ont suggéré un réseau hiérarchique, où quelques gènes fortement interconnectés représentent la plupart des interactions dans les LB (Basso et al. 2005).

Plus largement, les GRN ont permis de constituer une base de connaissances pour la pratique médicale et clinique (Barabási, Gulbahce, and Loscalzo 2011). Par exemple, une étude a cherché à identifier les réseaux de co-expression de gènes associés à des locis génétiques connus pour être impliqués dans l'obésité, le diabète et l'athérosclérose. Ces réseaux de co-expression ont identifié tout d'abord les paires de gènes qui sont fonctionnellement liées puis les sous-réseaux qui sont significativement associés à chacune de ces pathologies (Y. Chen et al. 2008).

Ainsi, une bonne connaissance de l'organisation des réseaux de gènes dans les cellules « endommagées » permet de mieux cibler les gènes qui sont directement impliqués dans la maladie. L'utilisation des réseaux de gènes paraît donc très prometteuse pour le design de médicaments ou de thérapies efficaces (Ghosh and Basu 2012). Ces réseaux pourraient donc permettre à terme non seulement de modéliser les maladies, mais de prédire aussi les traitements les plus appropriés tout en réduisant le coût des essais cliniques par gain de temps.

Par ailleurs, l'identification des GRN semble aussi indispensable pour la compréhension de différents programmes de régulation mis en place dans un organisme tels que la réponse immunitaire et les processus de différenciation (Rozenblatt-Rosen et al. 2012) (Yosef et al. 2013).

1.3.4.3 Structure des GRN

Les GRN sont généralement représentés par des réseaux dirigés dans lesquels les nœuds représentent des gènes ou des produits de l'expression des gènes (Figure 14). Certains gènes peuvent réguler l'expression d'autres gènes et ces relations de régulation sont indiquées par des flèches. Ces flèches représentent diverses connexions régulatrices (activation, auto-activation, répression, modulation, etc.) et elles peuvent être différemment représentées pour décrire la force de la relation de régulation (Shimoni et al. 2010) (Figure 14).

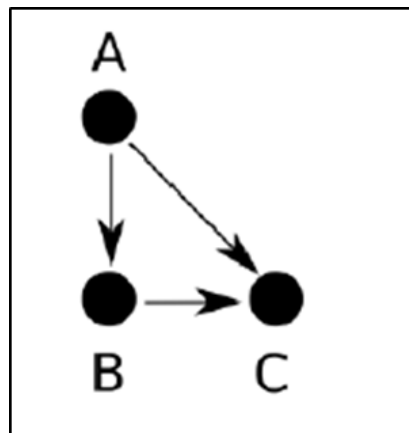


Figure 14- Réseau de Régulation de Gènes

Selon le schéma, le gène A régule à la fois les gènes B et C, tandis que le gène C ne régule aucun des gènes représentés.

1.3.4.4 Inférence des GRN

On peut définir la tâche d'inférence des GRN comme la détermination de la structure des GRN à partir de mesures expérimentales. Ainsi, lors d'une inférence, seulement la partie de la structure du GRN influencée par les conditions dans lesquelles se trouvent les cellules sera présentée.

L'inférence d'un réseau de régulation génique est rendue d'abord possible grâce à la disponibilité de données d'expression génique issues d'analyses sur populations cellulaires ou sur cellules isolées.

Les technologies de mesure du transcriptome sur populations cellulaires mesurent l'expression génique moyenne des cellules. Ainsi, elles permettent par exemple, d'évaluer les changements dans les échantillons prélevés dans différentes conditions. Cependant, elles sont insuffisantes pour étudier les systèmes hétérogènes tels que les tissus complexes et ne permettent pas d'approcher les différences entre les différents types de cellules et leur organisation spatiale. Ainsi, les réseaux construits à partir de données en population ne sont pas capables de décrire les relations entre les gènes dans les différents types de cellules constituant un tissu complexe, comme la peau, par exemple. C'est évidemment les données acquises à l'échelle de la cellule unique qui donnent l'information la plus riche et qui permettent d'établir les GRN les plus représentatifs du fonctionnement cellulaire au sein d'un environnement complexe.

Inférence des GRN à partir des données transcriptomiques obtenues en populations

A la fin des années 1990, grâce à l'utilisation de puces d'expression à ADN, l'inférence des GRN a constitué un défi informatique central en biologie systémique (Schena et al. 1995) (Lockhart et al. 1996). Le développement d'autres méthodes basées sur le séquençage pour la caractérisation du transcriptome comme la technique SAGE (Serial Analysis of Gene Expression)

(Velculescu et al. 1995) a permis aussi d'étudier des systèmes biologiques à un niveau jamais atteint auparavant et d'ouvrir le champ à différentes applications.

Grâce à différentes évolutions technologiques, ces méthodes d'études d'expression des gènes à l'échelle de la population cellulaire associées à des algorithmes de plus en plus performants ont permis de construire des cartes pertinentes de relations possibles de régulations géniques. Les algorithmes d'inférence de réseaux à partir des données d'expression en population sont multiples (Chai et al. 2014). Il existe : les réseaux bayésiens et bayésiens dynamiques (Dojer et al. 2006) (Xuan Vinh et al. 2012), les réseaux de co-expression (Steuer et al. 2002), les réseaux booléens (Akutsu, Miyano, and Kuhara 1999) et les équations différentielles ordinaires (ODE) (Polynikis, Hogan, and di Bernardo 2009) qui peuvent être couplées aux réseaux bayésiens (Bansal et al. 2007).

Inférence bayésienne et bayésienne dynamique

Les réseaux bayésiens sont représentés par des réseaux acycliques où la probabilité d'expression d'un gène dit fils, est décrite comme la probabilité conditionnelle de ses parents (Friedman et al. 2000) (Yu et al. 2004).

L'inférence bayésienne implique deux étapes : 1/la définition des interactions causales entre les différentes composantes du réseau génétique ; 2/ la détermination des probabilités qui traduisent ces interactions. Les réseaux inférés sont ensuite évalués à l'aide d'un score afin de déterminer le modèle le plus fiable.

L'inférence bayésienne permet de représenter intuitivement des GRN et de travailler avec des systèmes pour lesquels les connaissances sont incomplètes, en tenant compte de la présence de bruit dans les données d'expression. Cependant, elle présente certaines limites. Tout d'abord, l'inférence de GRN à partir de séries temporelles est limitée car elle ne permet pas d'étudier l'évolution du système dans le temps. L'inférence bayésienne nécessite aussi de grandes ressources de calcul, limitant ainsi son application à des GRN de petites tailles. Par ailleurs, par sa forme acyclique, le réseau ne peut pas prendre en compte les possibilités de rétro-action et de boucles (Österlund, Bordel, and Nielsen 2015). Cependant, ces deux dernières limitations

peuvent être surmontées grâce à l'utilisation des réseaux bayésiens dynamiques qui permettent la modélisation des interactions cycliques entre les gènes par duplication des nœuds au sein du réseau et par l'ajout d'une dimension temporelle dans le but d'étudier l'évolution du système dans le temps.

Inférence par la théorie de l'information

L'inférence par la théorie de l'information ou inférence par co-expression, est une des méthodes les plus utilisées. Cette approche cherche d'identifier des similitudes et dissemblances entre des paires de gènes et considère que les deux gènes qui partagent la même information sont incontestablement liés par une interaction plus ou moins directe. En effet, si le coefficient de corrélation de deux gènes dépasse un seuil (une valeur minimale) fixé, ces deux gènes sont considérés comme des gènes interagissant (Steuer et al. 2002). Des algorithmes tels que ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) et ses dérivés (Lachmann et al. 2016), CLR (Context Likelihood of Relatedness) (Faith et al. 2007) et WGCNA (Weighted Gene Co-expression Network Analysis) (Langfelder and Horvath 2008) utilisent cette approche.

Cette approche permet d'inférer des GRN de grande taille à partir de peu de données. Cependant, elle ne permet pas d'implémenter des boucles de rétro-contrôle qui jouent un rôle important dans les GRN. De plus, elle ne prend pas en compte l'implication de plusieurs régulateurs dans la régulation d'un gène cible, cas très fréquent dans les cellules.

Inférence booléenne

Les réseaux booléens sont schématisés par des graphes dirigés qui représentent les relations entre les gènes par des fonctions de logique booléennes. Dans cette approche, l'expression de chaque gène est discrétisée pour prendre une valeur qui correspond à son état d'activation (1) ou d'inactivation (0). Ainsi, le but de cette inférence booléenne est d'interpréter les fonctions logiques entre les gènes et leurs régulateurs afin que les données discrétisées puissent être expliquées par le réseau inféré (Silvescu and Honavar 2001).

Cette méthode est facile à implémenter et permet de générer des GRN de grande taille. Elle est de plus adaptée à la simulation de systèmes dynamiques qui permet de représenter les

phénomènes biologiques sous-jacents. Cependant, l'approche booléenne est dépendante de la présence de données temporelles et elle est limitée à l'inférence de GRN avec un faible nombre de gènes. De plus, le choix du seuil de discrétisation est sensible au bruit des données d'expression ce qui impacte l'inférence des réseaux.

Inférence par équations différentielles (ODE)

Les modèles par ODE sont des modèles mécanistes qui représentent d'une façon quantitative et continue les changements dans les niveaux d'expression des gènes (Gardner et al. 2003). Dans ces modèles, l'expression de chaque gène est définie par une fonction qui prend en compte l'expression des autres gènes et les paramètres environnementaux du système. Ainsi, l'intégration de connaissances à priori est indispensable pour définir la structure et les paramètres du modèle.

Les méthodes par ODE permettent de reproduire des comportements observés mais aussi d'être prédictives. Cependant, le besoin de connaissances à priori, la taille limitée des GRN ainsi que les temps de calculs importants constituent des limitations à cette approche.

Inférence des GRN à partir des données transcriptomiques obtenues en cellule unique

Plus récemment, la détermination de la structure des GRN à partir de profils d'expression génique générés par des analyses menées à l'échelle de la cellule unique a constitué un objectif majeur de nombreux travaux (De Smet and Marchal 2010) (Thorne and Stumpf 2012) (Oates, Amos, and Spencer 2014) (Young, Raftery, and Yeung 2014).

Afin d'inférer les GRN à partir des données transcriptomiques obtenues en cellule unique plusieurs approches d'inférence de GRN à partir de données en population ont été adaptées. Les approches d'inférence des GRN à partir des données transcriptomiques obtenues en cellule unique peuvent être classées en quatre catégories : le modèle booléen, l'équation différentielle, la théorie de l'information et le modèle de corrélation sur pseudo-temps. Les 3 premières catégories utilisent le même principe d'inférence de GRN qu'à partir de données en population

mais elles permettent de traiter des données plus complexes issues des analyses en cellule unique afin de décrire les relations entre les gènes à partir de ces données. Comme l'inférence par la théorie d'information, les méthodes de corrélation sur le pseudo-temps, s'appuient sur la corrélation entre les gènes pour représenter la relation par paire entre les gènes mais elles prennent aussi en compte que la relation entre les gènes peut changer au cours du temps (Nguyen et al. 2021).

Malgré les différentes méthodes développées, l'inférence des GRN à partir des données d'analyse en cellules uniques s'accompagne de limitations majeures : la nature stochastique de l'expression des gènes et les sources de variabilité techniques qui peuvent advenir à chaque étape des manipulations. Toutes ces sources de variabilité peuvent entraîner des biais sur le niveau de détection des signaux entre les cellules et entre les gènes pouvant en conséquence influencer l'inférence des GRN (Hicks et al. 2018) (Tung et al. 2017). Néanmoins, ces effets de variabilité peuvent être lissés lorsque le nombre de cellules augmente, ce qui permet d'avoir une vision plus juste grâce à une analyse sur grands effectifs. Aujourd'hui, la plupart des méthodes d'inférence de GRN sont conçues pour analyser des ensembles de données avec un nombre de cellules faibles ou modérés ; le défi d'augmenter le nombre de cellules intégrables dans ces analyses reste donc à être relevé (Nguyen et al. 2021).

1.3.4.5 Inférence des GRN par l'algorithme itératif : WASABI

WASABI (WAVes Analysis Based Inference) ou en français "Inférence basée sur l'analyse des vagues" est un algorithme itératif d'inférence des GRN à partir des données d'expression génique issues de cellules uniques, qui ne considère pas la variabilité comme un bruit, mais plutôt comme une information nécessaire à l'inférence des GRN (Bonnaïffoux et al. 2018).

WASABI est basé sur la logique de trouver des relations de cause à effet entre gènes : en réponse à un stimulus, les gènes d'un réseau s'activent les uns après les autres en cascade ce qui amène à l'identification de vagues d'activation de gènes.

La capacité de WASABI à inférer des GRN a été testée dans un premier temps sur des données *in silico*, puis sur des données de RT-qPCR générées *in vitro*. WASABI permet de prédire

un ensemble de réseaux avec des topologies similaires qui sont capables de reproduire *in silico* nos données. Ces réseaux, topologiquement très proches, sont nommés “réseaux candidats”. Un des réseaux candidats généré à partir des données d’expression obtenues en cellule unique sur des progéniteurs érythrocytaires en différenciation est présenté dans la figure 15.

Il faut noter que WASABI produit en sortie des GRN exécutables, c’est-à-dire des réseaux dont on peut simuler informatiquement le comportement. Il devient alors possible de tester *in silico* l’effet de perturbations, telles que le KO de certains gènes, ouvrant ainsi la voie au Design of Experiment (DOE).

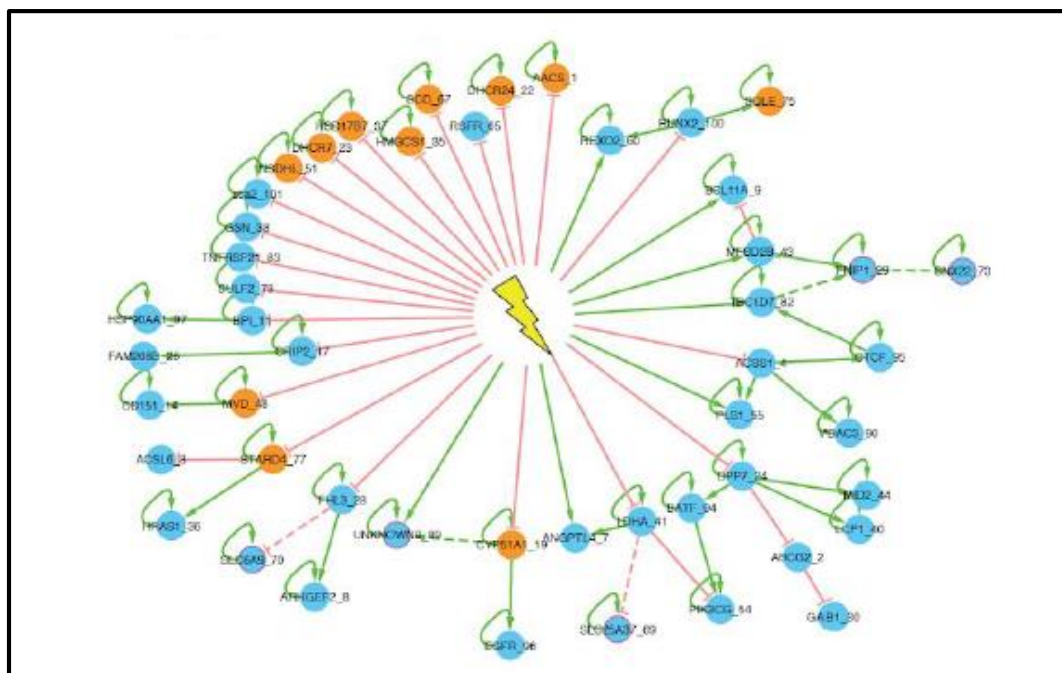


Figure 15- Exemple d’un GRN candidat inféré par WASABI

L’application de WASABI sur les données d’expression obtenues à l’échelle de la cellule unique sur des progéniteurs érythrocytaires en différenciation a permis d’inférer 364 GRN candidats. Un de ces candidats est présenté dans cette figure. Les flèches vertes indiquent les interactions positives tandis que les flèches rouges indiquent les interactions négatives. Les lignes continues représentent les interactions trouvées dans tous les GRN candidats et les lignes en pointillés, les interactions trouvées seulement dans certains GRN candidats. (Bonnafeux et al. 2018)

1.3.4.6 Stratégie de Design of Experiment (DOE)

Afin de réduire le nombre de réseaux candidats, notre équipe a mis en place une stratégie de Design of Experiment (DoE, ou Plan d'Expériences en français). C'est une stratégie qui permet de perturber *in silico* les réseaux proposés par WASABI afin d'identifier la ou les perturbations qui pousseraient les réseaux candidats à se comporter le plus différemment possible. En effet, si le comportement devient différent, il devient alors plus facile d'identifier le meilleur réseau candidat. L'objectif du DoE est donc d'identifier l'expérience de perturbation qui imposerait aux réseaux candidats à se comporter le plus différemment possible.

Les perturbations qui peuvent impacter directement le comportement des GRN peuvent être des knock-down, des knock-out, ou une surexpression de certains gènes. D'autres types de perturbations peuvent aussi impacter indirectement les GRN, par exemple le changement d'un stimulus extérieur.

Après identification *in silico* de l'expérience de perturbation la plus pertinente, il est nécessaire de revenir au niveau expérimental pour la tester. Les expériences menées *in vitro* permettront alors de générer de nouvelles données qui seront intégrées dans WASABI pour permettre de réduire le nombre de GRN candidats. Ce cycle pourra être répété jusqu'à trouver le seul GRN candidat sous-jacent au processus étudié et parfaitement compatible avec les données biologiques (Figure 16).

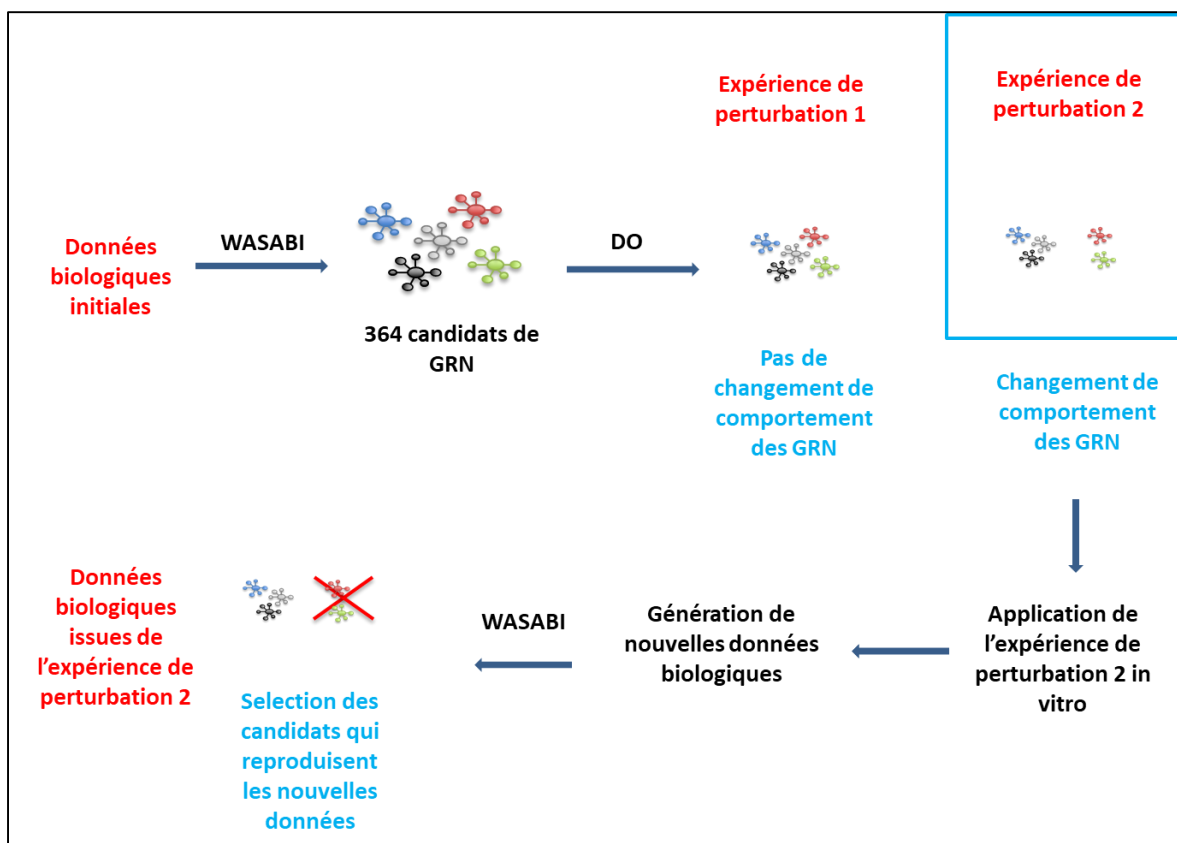


Figure 16- Principe de la stratégie de DOE

Les données biologiques sont analysées par WASABI qui déduit un ensemble de GRN candidats. Dans cette figure seulement 5 candidats sont présentés. Des expériences de perturbations (1 et 2) sont ensuite simulées *in silico*. L'expérience 2 a permis d'augmenter la distance entre les réseaux à la différence de l'expérience 1. C'est donc l'expérience 2 qui sera réalisée *in vitro* et les données biologiques obtenues seront ensuite utilisées afin de réduire le nombre de réseaux candidats.

1.3.4.7 Application du DOE sur les réseaux candidats inférés par Wasabi à partir des données de différenciation érythrocytaire

La capacité de WASABI à inférer les GRN a été testée sur les données de scRT-qPCR obtenues dans l'équipe à partir d'une cinétique de différenciation de progéniteurs érythrocytaires aviaires (T2EC). WASABI a ainsi généré un ensemble de 364 réseaux candidats.

Une stratégie de DOE a ensuite été utilisée afin de trouver la perturbation permettant de maximiser le changement de comportements des réseaux et d'identifier ainsi le réseau le plus pertinent et représentatif du processus de différenciation. Pour chacun des gènes des GRN, des perturbations de KI, KO et KD ont été simulées.

Lorsque j'ai débuté ma thèse, des analyses préliminaires avaient pointé le gène *FNIP1* (Folliculin-interacting protein 1) comme un gène potentiellement pertinent pour discriminer les GRN. De plus, l'analyse des interactions entre les gènes a aussi permis d'identifier *FNIP1* et également *MID2* (Midline 2) comme deux gènes dont les connexions avec leurs cibles potentielles sont les plus incertaines avec 6 % des interactions qui varient entre les 364 GRN candidats (Figure 17). Il apparaissait donc pertinent de tester l'effet d'une inhibition de *FNIP1* ayant un rôle connu dans plusieurs processus cellulaires (métabolisme, apoptose...), sur le comportement de ces GRN afin d'identifier le réseau candidat décrivant au mieux le processus de différenciation des érythrocytes.

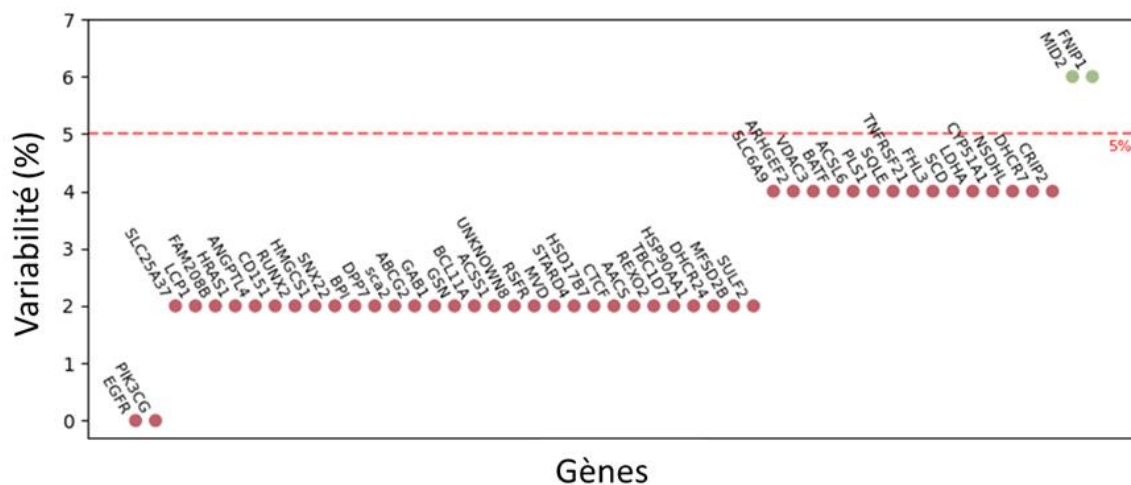


Figure 17- Incertitude sur le positionnement des gènes dans les 364 GRN candidats

Pour chaque gène la variabilité est calculée comme le pourcentage de liens entre un gène et ses gènes cibles qui ne sont pas conservés dans les 364 GRNs inférés par Wasabi. Par exemple, tous les gènes régulés par EGFR sont conservés dans les 364 GRN candidats. A l'inverse, dans le cas de *FNIP-1*, seules 94 % des interactions avec des gènes régulés sont conservées, donc 6 % des interactions varient entre les 364 GRN candidats.

Ainsi, la deuxième partie de mon travail de thèse a donc eu pour objectif d'obtenir des cellules isolées présentant un KO de *FNIP1*, identifié comme un gène très incertain et d'étudier le transcriptome de ces cellules par scRT-qPCR.

Une meilleure connaissance du gène *FNIP1* et de ses fonctions s'est donc révélée nécessaire avant le début des expérimentations. La partie bibliographique suivante sera donc consacrée au gène *FNIP1*.

1.4 Protéine d'interaction avec la Folliculine : FNIP1

La protéine d'interaction avec la Folliculine 1 est codée par un gène comprenant 18 exons représenté dans la figure 18. Cette protéine de 130kDa interagit notamment avec la folliculine (FLCN, Folliculin) et appartient à la famille des FNIP, tout comme FNIP2 (Hasumi et al. 2008) (Takagi et al. 2008). Elle contient un domaine DENN en tant que structure majeure caractéristique des protéines de la famille FNIP dans leurs extrémités C-terminale (Centini et al. 2018) (McBride et al. 2009) (Siggs et al. 2016) (figure 18). La protéine FNIP1 est modérément conservée parmi les espèces (Baba et al. 2006).

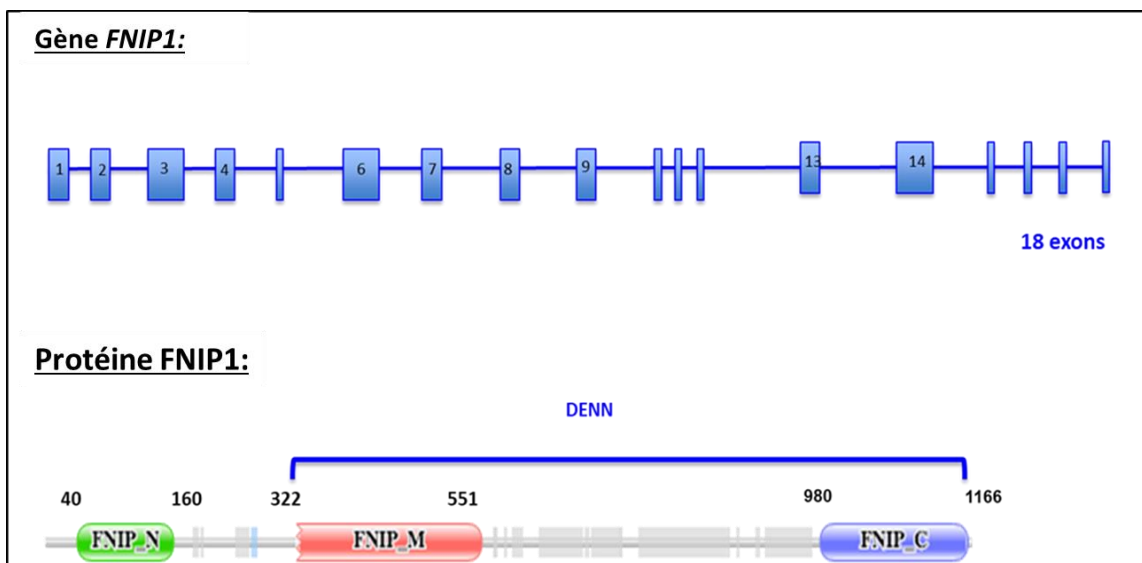


Figure 18- Structure du gène *FNIP1* et de la protéine correspondante

FNIP1 est constitué de 18 exons et code pour la protéine FNIP1 de 1666 AA contenant un domaine DENN dans son extrémité C-terminale.

Les complexes formés par FNIP1

Grâce aux acides aminés 322 à 1166 (domaine DENN), la protéine FNIP1 interagit avec l'extrémité C terminale de la FLCN (Hasumi et al. 2015). Le gène codant FLCN est associé au syndrome de Birt-Hogg-Dubé (BHD), au pneumothorax spontané héréditaire chez l'homme et agit comme un gène suppresseur de tumeur (Nickerson et al. 2002). En effet, des études ont montré qu'une FLCN tronquée en C-terminal est incapable de se lier à FNIP1 ce qui pourrait compromettre la fonction de FLCN et conduire au phénotype Birt-Hogg-Dubé chez l'homme (Schmidt and Linehan 2015a) (Schmidt and Linehan 2015b) (Schmidt 2013). Il s'agit d'un trouble systémique multi-organes rare, résultant d'une mutation autosomique dominante et caractérisé par la présence de fibrofolliculomes, de kystes pulmonaires avec ou sans pneumothorax spontané et de tumeurs rénales (Menko et al. 2009).

Grâce à des expériences de co-immunoprécipitation réalisées dans des cellules HEK293, il a été montré que la protéine FNIP2 était présente dans tous les complexes protéiques avec FNIP1 et FLCN (Hasumi et al. 2008). De plus, il a été observé que le complexe FNIP1 et 2 /FLCN était localisé dans le cytoplasme selon un motif réticulaire (Takagi et al. 2008). Par ailleurs, d'autres études ont montré que les deux protéines FNIP1 et FNIP2 pouvaient fonctionner indépendamment de la présence de la FLCN (Hasumi et al. 2008).

Les profils d'expression de la *FLCN*, *FNIP1* et *FNIP2* dans les tissus humains analysés par RT-qPCR montrent que l'expression de ces gènes est généralement similaire et systématiquement élevée dans certains tissus ou organes comme les muscles, la muqueuse nasale, les glandes salivaires, soutenant que FLCN, FNIP1 et FNIP2 fonctionnent ensemble dans ces organes (Hasumi et al. 2008). D'autre part, l'expression de *FNIP2* apparaît plus élevée que celle de *FNIP1* dans la graisse, le foie et le pancréas, ce qui suggère que *FNIP2* pourrait avoir une fonction spécifique dans ces tissus à fortes fonctions métaboliques (Hasumi et al. 2008).

L'activité du complexe FNIP1 et 2 /FLCN est régulé principalement par la modulation de l'activité de deux protéines clés, mTORC1 (mammalian target of rapamycin complex 1) et l'AMPK (AMP-activated protein kinase). En effet il a été décrit que mTORC1 et AMPK phosphorylent

directement ou indirectement FLCN sur les sérines 302 et 62, respectivement (L. Wang et al. 2010) (Zoncu et al. 2011). Cependant, bien que les effets de ces modifications sur la fonction de FLCN ne sont pas encore bien caractérisés, la phosphorylation sur la ser62 augmente la liaison de FLCN à l'AMPK, tandis que la phosphorylation sur la ser302 a l'effet inverse.

Plus récemment, il a été décrit que FNIP1 était aussi phosphorylée mais par la Casein-kinase-2-(CK2) au niveau des résidus sérine 938, 939, 941 et 948 (Woodford et al. 2016).

1.4.1 Rôles de FNIP1 dans les processus biologiques

Le rôle de FNIP1 a été mis en évidence dans plusieurs processus biologiques comme la régulation du métabolisme via mTOR, le maintien de l'équilibre métabolique des lymphocytes B et des lymphocytes iNKT (Baba et al. 2012) (H. Park et al. 2012) (H. Park et al. 2014) (Iwata et al. 2017), le fonctionnement des muscles squelettiques (Reyes et al. 2015), la fonction rénale (Hasumi et al. 2015) ou encore les processus d'autophagie (Behrends et al. 2010), pour ne citer que les principaux.

1.4.1.1 Régulation du métabolisme

mTORC1

mTORC1 est un complexe protéique qui fonctionne comme un capteur de nutriments/énergie/redox et qui contrôle la synthèse des protéines (Hay and Sonenberg 2004) (D.-H. Kim et al. 2002). Son champ d'action est donc très large puisqu'il agit ainsi sur la croissance et survie des cellules, l'homéostasie cellulaire, la synthèse des protéines, la lipogenèse, le métabolisme du glucose, l'autophagie, la biogenèse des lysosomes...(Ben-Sahra and Manning 2017). Parmi les principaux composants centraux de mTORC1, on peut citer mTOR, le noyau catalytique de mTORC1 et RAPTOR (Regulatory-associated protein of mTOR) (Loewith et al. 2002).

Le niveau d'activation de mTORC1 dépend de l'état nucléotidique de petites GTPases : la protéine Rag GTPase (Ras-Related GTP-binding protein) (Sancak et al. 2008) et la protéine Rheb

GTPase (GTP-bound Ras homolog enriched in brain) (Sato et al. 2009). Alors que les stimuli externes, tels que les facteurs de croissance ou le stress cellulaire, régulent l'activité de mTORC1 par l'intermédiaire des Rheb GTPases, les signaux intracellulaires liés à l'état nutritif et énergétique tels que la disponibilité des acides aminés, régulent quant à eux l'activité des Rag GTPases pour contrôler la localisation lysosomale de mTORC1.

Le principal régulateur de Rheb est le complexe TSCC (Tuberous Sclerosis Complex), composé notamment des deux protéines TSC1 et TSC2 (Tomasoni and Mondino 2011). TSC2 agit comme une protéine d'activation de la GTPase (GAP) de Rheb, en favorisant sa conformation inactive liée au GDP, et agit ainsi comme un régulateur négatif de mTORC1 (Zhang et al. 2003). Généralement, l'activation de la voie de PI3K-Akt par les facteurs de croissance extracellulaires entraînent l'inhibition du complexe TSCC permettant à la protéine Rheb d'activer mTORC1 (S. Park et al. 2010) (Figure 19).

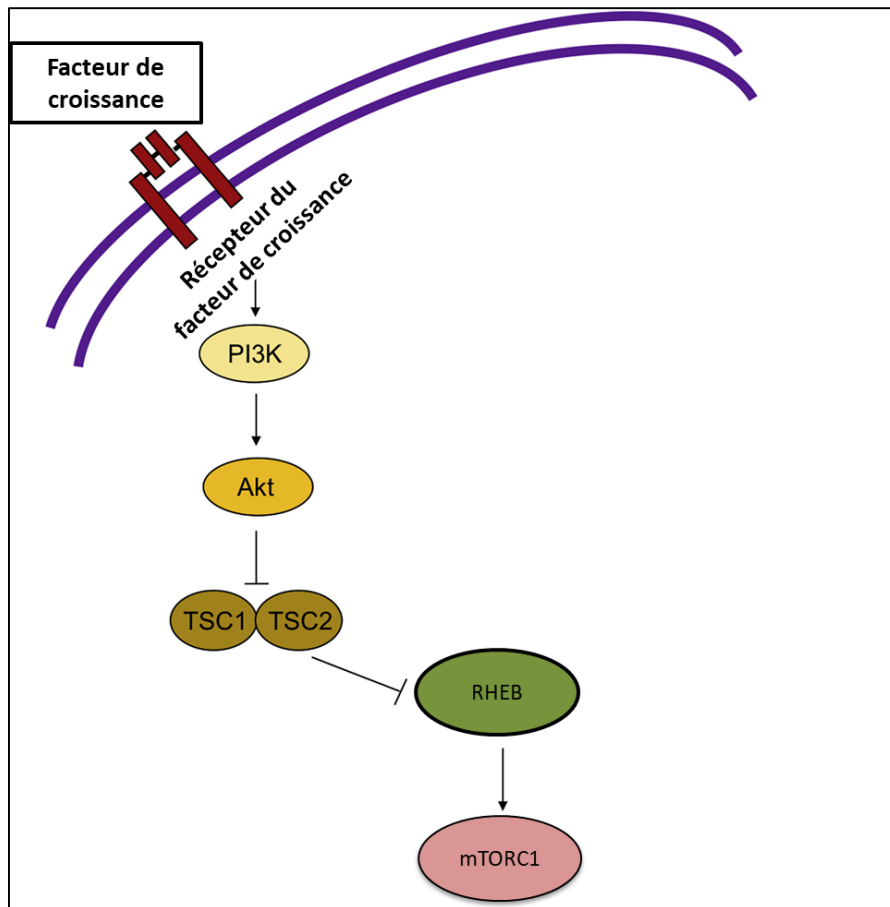


Figure 19-Activation de mTORC1 par la voie PI3/AKT/TSC

L'activation de la voie PI3K-Akt par des facteurs de croissance extracellulaires entraîne l'inhibition du complexe protéique TSC1/TSC2, ce qui conduit à l'activation de mTORC1 par Rheb.

Les signaux énergétiques intracellulaires, particulièrement les acides aminés, activent la voie des Rag GTPase, qui interagissent directement avec mTORC1 pour le recruter au niveau du lysosome (Sancak et al. 2008). Les Rag GTPase sont constituées de RagA ou B liées à RagC ou D. Les nutriments activent le complexe en faisant passer les GTPases Rag dans une conformation active, dans laquelle RagA/B est chargée en GTP et Rag C/D est chargée en GDP (E. Kim et al. 2008). Cette conformation module à son tour mTORC1 et sa distribution lysosomale.

La régulation des Rag GTPase implique plusieurs complexes multi-protéiques agissant comme des GAPs. Parmi ces derniers, GAROR1 a été identifié comme un GAP pour RagA/B favorisant l'état RagA/B-GDP et le complexe de FNIP1 et 2 /FLCN a été identifié comme un GAP

pour RagC /D favorisant l'état RagC/D-GDP et donc l'activation du complexe Rag GTPase (Lawrence et al. 2019) (Dodding, 2017).

Dans des conditions de carence en nutriments, GATOR 1 inactive le complexe Rag GTPase en favorisant l'état RagA-GDP. Parallèlement, AMPK phosphoryle le complexe FNIP1 et 2 /FLCN qui sera recruté dans le lysosome pour interagir avec RagA-GDP et empêcher le recrutement de mTORC1 (Baba et al. 2006). Cependant, lorsque les niveaux de nutriments sont rétablis, le complexe FNIP1 et 2 /FLCN exerce son activité GAP envers RagC, puis il retourne dans le cytosol permettant à RagA-GTP de recruter mTORC1 dans le lysosome afin d'être activé (Dodding, 2017) (Lawrence et al. 2019) (Tsun et al. 2013) (Figure 20).

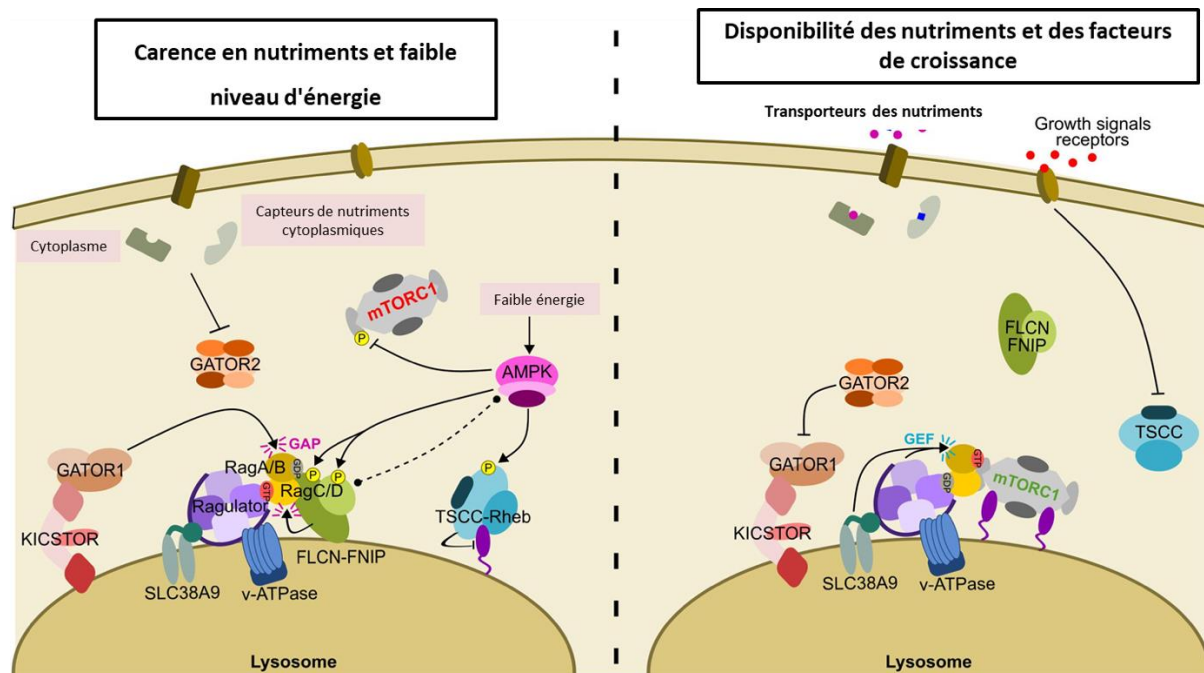


Figure 20- Régulation de la voie de signalisation mTORC1

En l'absence de nutriments, mTORC1 n'est pas recruté vers les lysosomes. Les faibles niveaux d'énergie activent l'AMPK qui phosphoryle TSCC permettant d'inhiber l'activité de mTORC1. Parallèlement, l'AMPK phosphoryle le complexe FNIP1 et 2 /FLCN qui sera recruté dans le lysosome. Au niveau du lysosome, le complexe FNIP1 et 2 /FLCN interagit avec RagA-GDP activé par GATOR1 pour empêcher le recrutement de mTORC1 vers le lysosome. Dans la situation inverse, lorsque des nutriments et les facteurs de croissance sont disponibles, le complexe FNIP1 et 2 /FLCN exerce son activité GAP sur RagC et retourne rapidement vers le cytosol pour recruter mTorc1 qui sera activé au niveau du lysosome (de Martín Garrido and Aylett 2020).

Régulation de la protéine kinase activée par l'AMP (AMPK)

L'AMPK (AMP-activated protein kinase) est une kinase hétérotrimérique formée par un noyau catalytique, la sous-unité α , et deux sous-unités régulatrices β . C'est un acteur essentiel de l'homéostasie énergétique : un régulateur positif du métabolisme catabolique et un régulateur négatif des processus anaboliques dans des conditions de faible énergie.

Le rôle de FNIP1 dans la régulation de l'AMPK a été largement discuté et il a été montré que selon les études et les modèles cellulaires, les effets de FNIP1 sur la régulation de l'AMPK pouvaient être contraires.

Par exemple, une étude récente réalisée sur des souris, a mis en évidence que la délétion constitutive du gène *FNIP1* était suffisante pour entraîner la diminution de l'activation de l'AMPK, concomitante à une hyperactivation métabolique et à la formation de kystes rénaux (Hasumi et al. 2015). De plus, des analyses par western blot réalisées sur des lysats rénaux de souris *Fnip1*^{-/-} ont révélé une diminution du taux de phosphorylation de l'AMPK au niveau de la thréonine 172 par rapport aux souris WT, accompagnée d'une diminution de l'activation de l'AMPK.

Une autre étude a montré que les souris homozygotes déficiente pour *FNIP1* développent une cardiomyopathie caractérisée par une hypertrophie ventriculaire gauche et une accumulation de glycogène. Ces souris présentent une activité élevée de la sous unité $\gamma 2$ de l'AMPK dans le myocarde néonatal déficient en *FNIP1*. Contrairement à l'étude précédente, ces résultats supportent pour FNIP1 un rôle de régulateur négatif de l'AMPK (McBride et al. 2009). D'autres études soutiennent aussi le rôle du complexe FNIP1 et 2 /FLCN comme régulateur négatif de l'AMPK ; il a en effet été montré que la déplétion de *FLCN* ou *FNIP1* entraîne une activation constitutive de l'AMPK dans les cellules de *C. elegans*, les fibroblastes et les Lymphocytes B des souris (Possik et al. 2014) (Siggs et al. 2016).

En outre, la perte de *FNIP1* dans le muscle squelettique strié est associée à une activité accrue de l'AMPK dans une étude (Reyes et al. 2015) mais à une activité réduite dans d'autres (Hasumi et al. 2015) (Centini et al. 2018).

1.4.1.2 Apoptose et cycle cellulaire

Le rôle de *FNIP1* dans la mortalité et le cycle cellulaire est bien décrit. Tout d'abord, le rôle de *FNIP1* dans l'apoptose a été mis en lumière dans le modèle des cellules murines iNKT. En effet, le développement des cellules iNKT *Fnip1*^{-/-} est arrêté à un stade immature ; une diminution du nombre de mitochondries et du taux d'ATP dans ces cellules entraîne une sensibilité accrue à l'apoptose. Ces résultats indiquent donc que *FNIP1* est vital pour maintenir l'équilibre métabolique pendant le développement des cellules iNKT (H. Park et al. 2014).

Il a été aussi rapporté que la perturbation de *FNIP* entraîne un déséquilibre métabolique, qui déclenche un arrêt de la maturation des cellules B au stade des grandes cellules pré-B et une diminution de la viabilité de ces progéniteurs (H. Park et al. 2012).

D'autre part, l'activité de la folliculine, qui fait partie du complexe FNIP1 et 2 /FLCN est liée à la progression dynamique du cycle cellulaire. Plus précisément, la perte de la *FLCN* accélère la progression des cellules UOK257 dans la phase G2/M (Laviolette et al. 2013). De plus, des défauts dans la division cellulaire ont été observés chez les drosophiles déficientes en *FLCN* (J. Liu et al. 2016). Parallèlement, chez le poisson zèbre, la *FLCN* est importante pour le développement du cerveau embryonnaire : en son absence, le nombre de cellules en G1 est augmenté, tandis que le nombre de cellules en phase S-M est réduit (Kenyon et al. 2016).

1.4.1.3 Activités de co-chaperonne

La protéine de choc thermique 90 (Hsp90) est un chaperon moléculaire responsable de la stabilité et de l'activité de nombreux substrats appelés "protéines clientes". Elle est ainsi impliquée dans de nombreux processus cellulaires tels que le repliement des protéines, la réparation de l'ADN, la réponse immunitaire mais semble aussi participer activement à certaines maladies neurodégénératives (Schopf, Biebl, and Buchner 2017). Son activité ATPase est nécessaire pour sa fonction et est généralement modulée par des co-chaperones.

De manière intéressante, la *FLCN* a été identifiée comme un substrat de Hsp90 et FNIP1/2 comme des co-chaperones. La liaison de FNIP1 et 2 à Hsp90 facilite l'interaction de la *FLCN* avec

Hsp90 et augmente la stabilité de la FLCN (Woodford et al. 2016). La fonction de FNIP1 en tant que co-chaperones de Hsp90 est modulée par phosphorylation par la CK2 sur les sérines 938, 939, 941 et 948. Une fois phosphorylée, la protéine FNIP1 se lie à Hsp90 et diminue son activité ATPase. Ensuite, la déphosphorylation de FNIP1 par la protéine phosphatase Ser/Thr PP5 permet l'ajout de O-GlcNAc (O-linked-N-acetylglucosamine) sur sa ser938. Ce processus antagonise la phosphorylation de FNIP1, empêchant son interaction avec Hsp90, et favorise alors l'ubiquitination de la lysine 1119 de FNIP1 et sa dégradation par le protéasome (Sager et al. 2019).

De manière intéressante, les sous-unités AMPK, RAPTOR et mTOR ont été identifiées comme des substrats aussi de Hsp90 (Sager et al. 2018), suggérant que FNIP1 peut aussi réguler ces voies par l'activité de Hsp90.

1.4.1.4 Réponse immunitaire

La FLCN et FNIP1 / 2 ont été récemment associées à l'expression des gènes impliqués dans la réponse à l'interféron dans des cellules rénales et des tumeurs du syndrome BHD (Glykofridis et al. 2021). Ainsi, la délétion de la *FLCN* ou de *FNIP1/2* entraîne une augmentation de l'expression de *STAT2* et de l'expression des gènes de la réponse à l'interféron médiée par *STAT2*. Toutefois, la manière dont FLCN et FNIP régulent l'expression de *STAT2* n'a pas été encore élucidée.

1.4.1.5 Différenciation

FNIP1 est une protéine qui apparaît aussi impliquée dans les processus de différenciation cellulaire. On lui attribue le rôle de point de contrôle métabolique dans le développement des lymphocytes B en agissant comme un commutateur moléculaire permettant la différenciation et la survie des cellules pré-B, le développement des cellules B et le maintien de leur homéostasie métabolique (H. Park et al. 2012).

Le rôle de FNIP1 dans la différenciation musculaire est aussi bien décrit. En effet, il a été démontré que FNIP1 modifie la composition des types de fibres musculaires squelettiques (type I « à contraction lente » et type II « à contraction rapide ») et qu'elle peut réguler la spécification

du type de fibres du muscle squelettique par la voie AMPK / PGC-1 α (Reyes et al. 2015)(J. Liu et al. 2016). La perte de *FNIP1* ou de la *FLCN* chez les souris, induit une augmentation du nombre de fibres musculaires squelettiques de type I en stimulant l'expression de la *myoglobine* dans les cellules musculaires squelettiques *Fnip1*^{-/-}, provoquant ainsi un remplacement des fibres à contraction rapide par des fibres à contraction lente.

De plus, il a été signalé que la *FLCN*, ainsi que ses partenaires d'interaction *FNIP1* et *FNIP2*, régulent la sortie des cellules souches embryonnaire de souris de leur état de pluripotence. Dans ces cellules, le complexe *FLCN-FNIP1-FNIP2* contrôle la différenciation en empêchant la localisation nucléaire de la protéine TFE3 (Transcription factor E3), bloquant ainsi l'activation de l'expression d'*Esrrb* (Estrogen Related Receptor Beta), un régulateur bien connu de pluripotence (Martello et al. 2012). Enfin, il a été décrit que les souris *FNIP1* double KO présentaient des cellules B bloquées dans leur développement entre le stade cellules pro-B et le stade cellules pré-B, au niveau de la rate, mettant encore en lumière le rôle de *FNIP1* dans la différenciation. Par contre, dans cette étude, le nombre de thymocytes et de lymphocytes T périphériques ne semble pas être affecté (Park et al. (2012).

1.4.1.6 Suppression de tumeurs

Le rôle de *FNIP1* en tant que suppresseur de tumeur a été bien décrit dans le rein, une délétion constitutive du gène *FNIP1* ayant notamment pour conséquence la formation de kystes rénaux (Centini et al. 2018). Suite à une double inactivation de *FNIP1* et *FNIP2* dans les reins de souris, une hypertrophie des reins polykystiques a été également observée. De plus, ces souris ont développé un cancer rénal à 24 mois, de manière similaire à ce qui avait été observé dans un modèle de souris hétérozygote knock-out pour la *FLCN*, confortant ainsi le fait que *FNIP1* et *FNIP2* jouent un rôle essentiel dans la suppression des tumeurs rénales, en coopération avec la *FLCN* (Hasumi et al. 2015).

Le gène *FNIP1* semble donc être très important pour les cellules puisque son rôle a été mis en évidence dans de nombreux processus biologiques comme la régulation du métabolisme, la différenciation et la survie cellulaire, dans les processus d'autophagie et de suppression de

tumeurs mais également dans d'autres processus biologiques sans ou avec la formation du complexe FLCN/FNIP.

2. Résultats

Comme présenté dans l'introduction bibliographique, différents modèles suggèrent le dynamisme du processus de la différenciation érythrocytaire. Dans ce contexte, mon travail de thèse a consisté dans un premier temps, à caractériser moléculairement la réversion de progéniteurs aviaires déjà pré-engagés dans le processus de différenciation érythrocytaire. Dans un deuxième temps, l'objectif de mes recherches a été de générer des données d'expression génique à l'échelle de la cellule unique pour permettre de définir au mieux le GRN sous-jacent au processus de différenciation érythrocytaire grâce notamment à la réalisation d'expériences de KO (Figure 21).

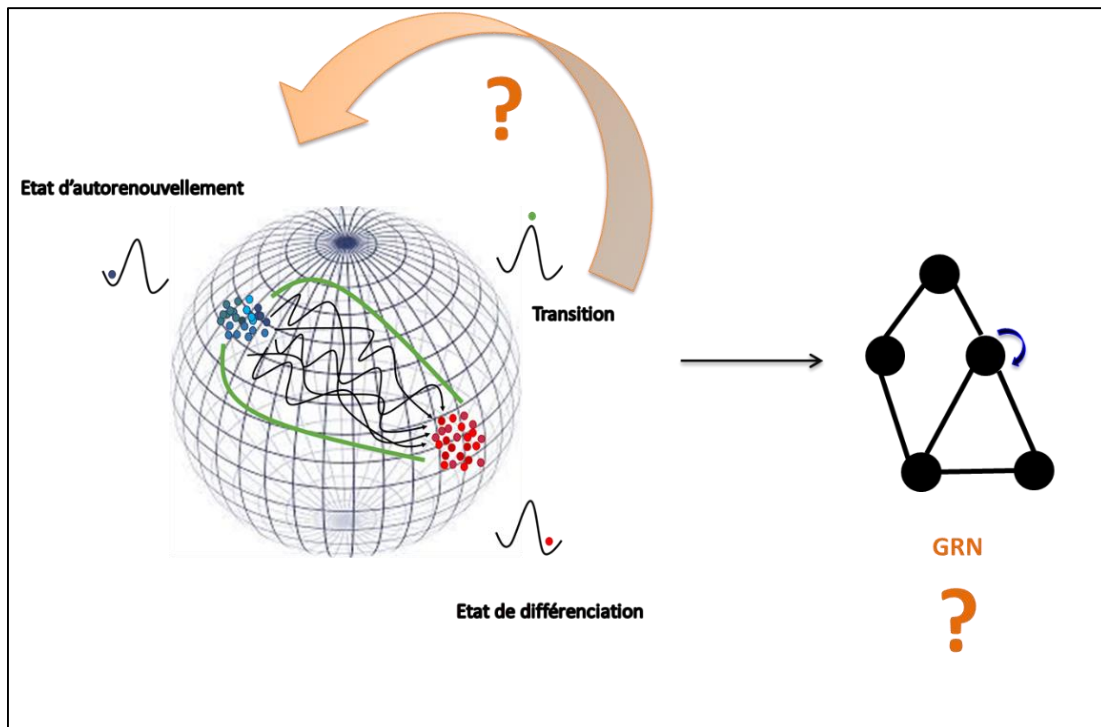


Figure 21-Characterisation moléculaire de la dynamique de la différenciation érythrocytaire

Mon travail de thèse a consisté en 1/ caractériser moléculairement la réversion de progéniteurs pré-engagés dans le processus de différenciation érythrocytaire et en 2/ définir le GRN sous-jacent au processus de différenciation érythrocytaire par approches bioinformatiques et KO.

2.1 Caractérisation moléculaire de la réversion des cellules pré-engagées dans la différenciation érythrocytaire

2.1.1 Introduction

Afin de caractériser moléculairement la réversion des progéniteurs aviaires déjà pré-engagés dans le processus de différenciation érythrocytaire, les T2EC initialement en état d'auto-renouvellement (0H Différenciation) ont été induites à se différencier dans un milieu de culture spécifique pendant 24 h (24H Différenciation) ; puis une partie de ces cellules a été retransférée dans le milieu d'auto-renouvellement (48H Réversion) et l'autre partie a été maintenue en différenciation pendant encore 24h (48H Différenciation) (Figure 22).

L'expression des gènes a été analysée à l'échelle de la cellule unique en utilisant deux approches différentes d'analyse du transcriptome, la scRT-qPCR et le MARS-Seq. Grâce à la technique de scRT-qPCR, nous avons pu analyser l'expression de 83 gènes pertinents étudiés déjà dans l'étude réalisée par Richard et al (Richard et al. 2016a). Parallèlement, l'approche MARS-Seq nous a permis une quantification bien plus large du transcriptome. Nous avons analysé environ 200 cellules par expérience pour chacun des 4 points de temps de différenciation ((0H, 24H, 48H Réversion, 48H Différenciation).

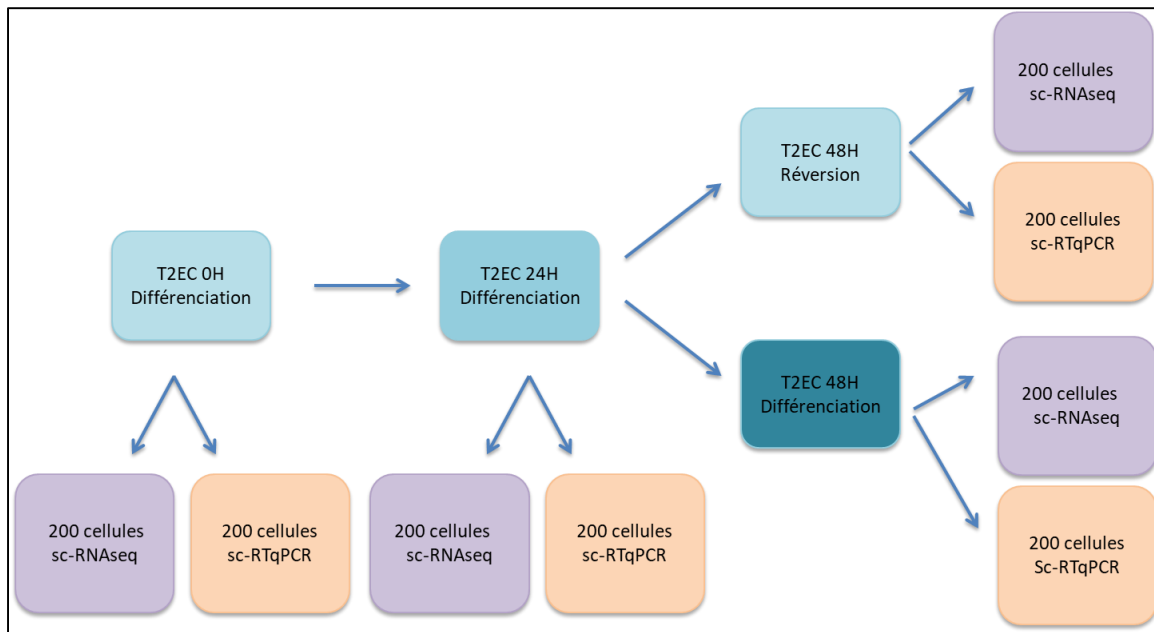


Figure 22- Stratégie expérimentale suivie pour la caractérisation moléculaire des T2EC en réversion

Les T2EC (0H Différenciation) sont induites à se différencier dans un milieu de culture spécifique pendant 24 h (24H Différenciation) puis une partie des cellules est retransférée dans le milieu d'auto-renouvellement (48H Réversion) et l'autre partie est maintenue en différenciation pendant encore 24h (48H Différenciation). Pour chaque condition, 200 cellules sont isolées pour l'analyse du transcriptome par scRT-qPCR et 200 autres cellules sont isolées pour l'analyse du transcriptome par MARS-Seq.

Un test de Pearson a montré une forte corrélation des niveaux d'expression entre notre jeu données de scRT-qPCR et celui du scRT-qPCR publié précédemment dans l'équipe (Richard et al. 2016). De plus, un fort niveau de corrélation des niveaux d'expression entre les deux technologies d'analyse du transcriptome utilisées (scRT-qPCR et MARS-Seq) a été également observé assurant la solidité de nos données.

La projection par UMAP des données de scRT-qPCR et de MARS-seq a montré un chevauchement entre la condition 48H Réversion et les conditions 0H et 24H de Différenciation. D'une manière intéressante, l'analyse de DE des données de MARS-Seq a mis en évidence un seul gène, *RSFR* (Gallus gallus leukocyte ribonuclease A-2) significativement exprimé de manière différentielle (juste au-dessus du seuil) entre les conditions 48H Réversion et 0H. Ce gène est reconnu pour être fortement exprimé dans les cellules précurseurs de la moelle osseuse de poulet (Klenova et al. 1992). Au contraire, plusieurs gènes différentiellement exprimés ont été

détectés en comparant les données 48H Réversion avec les données 24H Différenciation ou 48H Différenciation. Enfin, l'analyse des distances de Wasserstein a mis en évidence des distances plus étroites entre les cellules en réversion et les cellules indifférenciées, en comparaison avec les cellules différenciées. Ces résultats suggèrent donc que les conditions 0H et 48H Réversion sont moléculairement très proches mais néanmoins pas complètement identiques. En effet, l'identification par Sparse PLS des gènes les plus prédictifs pour discriminer la condition 48H réversion et la condition 0H a mis en lumière le gène codant la Bétoglobine. Les cellules 48H Réversion tout en étant redevenues très proches de leur état initial gardent donc une marque de leur engagement dans le processus de différenciation.

Tous ces résultats suggèrent que les cellules induites à se différencier, puis remises en condition d'auto-renouvellement ont retrouvé un profil moléculaire similaire à celui des cellules indifférenciées, bien que l'expression du gène codant la Bétoglobine reste différente entre les deux conditions. Il s'agit bien d'un mécanisme de réversion car les tests de viabilité cellulaire et l'analyse de la distribution gamma pour chaque gène, n'ont montré aucune différence significative entre les conditions 0H Différenciation et 48H Réversion. Ces résultats démontrent donc qu'il n'y a pas une sous-population particulière de cellules coexistant dans la population cellulaire 24H Différenciation qui serait à l'origine des cellules en réversion.

Ainsi, nos travaux ont permis de mieux comprendre la dynamique des cellules avant leur engagement définitif dans la différenciation en utilisant plusieurs approches d'analyse transcriptomique menées à l'échelle de la cellule unique, couplées à différentes analyses bioinformatiques.

2.1.2 Article

L'article présenté ci-après est en cours de finalisation pour une soumission prévue en janvier 2022.

Molecular characterization of pre-commitment cell reversion during erythroid differentiation

Souad Zreika^{1-2*}, Camille Fournaux^{1*}, Elodie Vallin¹, Laurent Modolo¹, Rémi Seraphin, Alice Moussy³, Elias Ventre¹⁻⁴⁻⁵, Matteo Bouvier¹⁻⁶, Olivier Gandrillon^{1-4#}, Sandrine Giraud^{1#}

1- ENS de LYON, CNRS UMR 5239, Laboratory of Biology and Modelling of the Cell, 46 allée d'Italie Site Jacques Monod, F-69007, Lyon, France

2- Azm Center for Research in Biotechnology and its Applications, LBA3B, EDST, Lebanese University, Tripoli 1300, Lebanon

3- Ecole Pratique des Hautes Etudes, PSL Research University, Paris, France

4- Inria Team Dracula, Inria Center Grenoble Rhone-Alpes, Grenoble, France

5- Institut Camille Jordan, CNRS UMR 5208, Université Claude Bernard Lyon 1, 43 Bd du 11 novembre 1918, F-69622 Villeurbanne, France

6- Vidium solutions, F-69007 Lyon, France

* These first authors contributed equally to this work

These last authors contributed equally to this work

Abstract

Many attempts have been made to understand the link between gene expression and cellular decision-making in developing organisms. In the context of dynamical systems theory, one of the most widely used concepts is that of Conrad Waddington's "epigenetic landscape" (Waddington, 1957). According to this concept, a cell is compared to a ball rolling down from the top of a mountain and crosses furrows before stopping in basins or "attractor states" to reach its stable cellular state. However, it has become clear that some cells retain different levels of reversibility, a feature that Waddington did not directly address. Based on the results of a recent study performed on primary chicken erythroid progenitors (T2EC), we tested the hypothesis of reversibility of T2EC cells induced to differentiate for 24h and then transferred to the self-renewal medium. Genes expression was analysed for 0H, 24H and 48H after induction of differentiation and 24H after placing in self-renewal medium the cells induced previously 24h in differentiation, using two different single cell molecular analysis approaches, high-throughput scRT-qPCR and scRNA-seq. Our overall results showed that for the 83 genes and 2000 most variable genes analysed with RT-qPCR and scRNA-seq respectively, cells induced to differentiate and then placed in self-renewal medium are molecularly very close to undifferentiated cells. It seems therefore that an almost complete reversion can be observed molecularly after 24 hours of differentiation.

Introduction

During its life, every cell must integrate various information from its environment to combine it into a coherent biological response. Whether as unicellular organisms living within a population or as part of multicellular organisms, this integration is critical to ensure function and survival. Thus, cellular decision-making has been the subject of numerous studies since it assures the normal functioning of organisms but is also responsible for numerous pathologies when compromised.

Mechanisms governing cell fate decision-making have often been described based upon cells population data leading to the belief that those processes are entirely deterministic. From the deterministic point of view, cells commitment to a "predefined" cell type is triggered through a stereotyped sequence of intermediate states under the influence of specific signals (1).

Single-cell approaches have allowed to change the scale of observation and revealed that an important heterogeneity in genes expression lies within isogenic cell populations (2,3). This heterogeneity in gene expression is critical for the process of decision-making as it provides diversity without the cost of hardwire-encoded fate programs (4,5). Stochasticity in gene expression arises from different causes, such as availability of transcription and translation regulatory proteins, the probabilistic nature of molecular interactions or transcriptional bursts (6). Cell to cell variability is visible at all levels of gene expression, but is being widely studied at the transcriptome level since various molecular biology tools are available for this scale of investigation (7).

Numerous single-cell studies have enabled the development of stochastic models to describe cell fate decision-making from single-cell transcriptomics data. In those models, a probability is assigned to each fate; cells from a population are presented as independent individuals with significant quantitative and qualitative heterogeneity, which results in the generation of different cell types (8–10). In the stochastic model, gene expression variability is seen as one of the underlying causes of cell commitment (11–13).

One of the best-known stochastic view is Conrad Waddington's landscape, which incorporates the non-genetic heterogeneity of cells (14). According to Waddington, the shape of the landscape

is determined by gene activity and cells transitions are modelled as channelling events: a cell, presented as a ball, starts from a mountain top and crosses valleys before reaching stable cell states by occupying basins or attractor states (15). Once a cell makes the definitive decision, its cellular potential decreases and its fate becomes restricted or even irreversible (16–21). In the modern version of this concept, gene expression variability is the support for natural selection and cell state is seen as a vector of probabilities (22–24).

However, it is now clearly accepted that some cells retain plasticity and there are exceptions where a cell thought to have reached an intermediate degree of differentiation can spontaneously return to an earlier, more pluripotent stage (9,17,18,20). In addition, a recent study showed that chicken primary erythroid progenitor cells (T2EC) have retained the capacity to go back to self-renewal state for up to 24 h after the induction of differentiation before they irreversibly engaged in the differentiation process (21).

The purpose of our work was therefore to characterize the cells capable of fate reversion and to question the extent to which this phenomenon actually coincides with molecular changes within these cells. For this, we characterized genes expression at the single cell level during the reversion process of T2EC. We used two different complementary single-cell transcriptomics technologies, scRT-qPCR and scRNAseq. scRT-qPCR allows for highly sensitive quantification but is knowledge-driven and offers information of a limited number of genes while scRNAseq, although less precise for low expression level (25) enables genome-wide quantification without prior knowledge. Furthermore, using two different single-cell technologies allowed us to cross validate our observations and point toward robust conclusions.

We collected cells in self-renewing state (0H) and changed the medium for differentiation medium. After 24H of differentiation, a first part of the cells was recovered (24h Differentiation), a second part of the cells was kept in differentiation medium (48H Differentiation) and the cells left were shifted back in self-renewal medium (48H Reversion) (Figure 1).

Using dimensional reduction tools, we were able to show that reversion cells (48H Reversion) and undifferentiated cells (0H) were much more similar, whereas a separation was clearly visible between cells maintained in differentiation (48H Differentiation) and cells in reversion (48H Reversion). We then performed differential gene expression analysis from scRNA-seq data and

saw no major difference in gene expression between the 48H reversion cells and 0H cells. Moreover, gene pattern distribution of 48H reversion cells showed a shift toward expression pattern distribution of 0H cells. Finally, we identify genes that discriminate 48H reversion cells and 0H cells, performing Partial Least Square analysis (26). We were able to show that only the expression of *HBBA* (*Hemoglobin subunit beta A*) was different between reversion cells and undifferentiated cells showing that reversion cells kept transcriptional traces of their induction to differentiation.

In conclusion, our results show that reversion cells display gene expression pattern similar to undifferentiated cells while retaining traces of their response to differentiation induction, which suggest an almost complete molecular reversion after 24 hours of differentiation induction.

Material and methods

Cellular biology

T2EC were extracted from bone marrow of 19-d-old SPAFAS white leghorn chicken's embryos (INRA, Tours, France). Cells were grown in self-renewal in LM1 medium (α -MEM, 10% Foetal bovine serum (FBS), 1 mM HEPES, 100 nM β -mercaptoethanol, 100 U/ mL penicillin and streptomycin, 5 ng/mL TGF- α , 1 ng/mL TGF- β and 1 mM dexamethasone) as previously described (27).

Differentiation was induced by removing LM1 medium and placing the cells into DM17 medium (α -MEM, 10% foetal bovine serum (FBS), 1 mM Hepes, 100 nM β -mercaptoethanol, 100 U/mL penicillin and streptomycin, 10 ng/mL insulin and 5% anemic chicken serum (ACS; (28)).

Differentiation kinetics was achieved by collecting a sub fraction of the cells at different times after induction of differentiation (0H and 24H). After 24H, DM17 medium was removed and half of the cells was placed back into LM1 medium while the other half was kept in DM17 medium to achieve 48H reversion and 48H differentiation time points respectively.

Cell population mortality was assessed by counting dead and living cells from the different time points and conditions after Trypan blue staining and using a Malassez cell.

Single-cell sorting

Single cells were sorted using a FACS Aria II μ , BD. We collected around 200 cells per time point for each experiment (sc RT-qPCR and scRNA-seq, see below).

Single cell RT-qPCR analysis

All the manipulations related to the high-throughput scRT-qPCR experiments in microfluidics were performed according to the protocol recommended by the Fluidigm company (PN 68000088 K1, p.157-172). All steps from single cell isolation to RTqPCR, genes selection, data generation and cleaning are described in detail in (21). Expression matrix was log1p transformed before subsequent analysis.

Single-cell RNA-seq

Single-cell RNA-seq was performed using an adapted version of MARS-seq protocol (Massively parallel single-cell RNA sequencing) (29). All indicated concentrations correspond to final concentrations.

Individual cells were sorted into 96-well plates containing 4µL of lysis buffer and index RT primers (0,2% Triton (Sigma Aldrich), 0,4 U/µL RNaseOUT (Thermofisher Scientific), 400nM RT_primers (Sigma Aldrich)). Index RT_primers (**Table 1**) contain oligo-dT chain to capture mRNA, a T7 RNA polymerase promoter for further *in vitro* transcription (IVT), unique cell barcodes for subsequent de-multiplexing and unique molecular identifiers (UMIs) for PCR bias deduplication. After cell sorting, plates were immediately centrifuged and frozen on dry ice before storage at -80°C until reverse transcription (RT) was performed. The plates were put at 72°C for 3 minutes for denaturation. 4µL of RT mix were added in each well (2mM dNTP (Thermo scientific), 20mM DTT, 2X First stranded buffer, 5 U/µL Superscript III RT enzyme (Superscript III RT enzyme kit Thermo scientific), 10% (W/V) PEG 8000 (Sigma Aldrich)). ERCC RNA spike-in (Thermo Scientific) was diluted into the RT mix (dilution 1/40). The plates were then transferred into a thermocycler (program: 42°C-2min, 50°C-50min, 85°C-5min, 4°C hold).

After reverse transcription, samples were pooled by plate and Exonuclease I digestion was performed (1µL per sample Exonuclease I NEB), followed by 1,2X AMPure beads purification (Beckman Coulter). Samples were eluted in 17µL of 10mM Tris-HCl, pH=7,5. Second strand cDNA synthesis was performed, 2µL 10x SSS buffer and 1µL SSS enzyme was added to the sample (NebNext mRNA second strand synthesis kit NEB; thermocycler program: 16°C-150min, 65°C-20min, 4°C hold). Resulting double strand cDNA was linearly amplified by IVT overnight (10mM ATP, 10mM GTP, 10mM UTP, 10mM GTP, 1X reaction buffer, 1/10 T7 RNA polymerase mix (HighScribe T7 High Yield RNA synthesis NEB)) at 37°C. IVT product was purified with 1,3X Ampure beads and eluted in 10µL of 10mM Tris-HCl, 0,1mM EDTA. 9µL of amplified RNA were fragmented (1X RNA fragmentation buffer (RNA fragmentation reagents Invitrogen)) at 70°C for 3 min. The fragmentation reaction was stopped with 34µL of STOP mix (0,3X Stop solution (RNA

fragmentation reagents Invitrogen), 0,7X AMPure beads, in TE buffer (Ambion)) and samples were purified. Differing from original MARSseq protocol, the second RT was done with P5N6 primers (Table 1) containing random hexamers and specific barcode to distinguish the different plates (5mM DTT, 500μM dNTP, 10μM P5N6_XXXX, 1X First stranded buffer, 10U/μL Superscript III RT enzyme, 2U/μL RNaseOUT; thermocycler program: 25°C 5min, 55°C 20min, 70°C 15min, 4°C hold). The cDNA were then purified with 1,2x AMPure beads. Illumina primers (Table 1) were added by PCR (0,5 μM Mix primer P5.rd1/P7;Rd2, 1X KAPA Hifi HotStart PCR Mix (Kapa Biosystem); thermocycler program: 95°C 3min, 12 times [98°C 20sec, 57°C 30sec, 72°C 40sec], 72°C 5min, 4°C hold), and PCR products were purified with 0,7x AMPure beads and eluted in 15μL.

Libraries were sequenced on a Next500 sequencer (Illumina) with a custom paired-end protocol to avoid a decrease of sequencing quality on read1 due to high number of T added during polyA reading (130pb on read1 and 20pb on read2). We aimed for a depth of 200 000 raw reads per cell.

Primer Name	5' to 3'
Index_RT_primers (cell BC and UMI)	5'-CGATTGAGGCCGTAATACGACTCACTATAGGGGCGACG TGTGCTCTTCCGATCTXXXXXXNNNNNNNNNTTTTTTTTTTTTTTTTTTTTV-3'
P5N6_XXXX (Plate BC)	5'-CTACACGACGCTCTTCCGATCTXXXXNNNNNNN-3'
P5.Rd1	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCT-3'
P7.Rd2	5'-CAAGCAGAAGACGGCATACGAGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

Table 1: List and sequences of primers used for scRNA-seq libraries construction.

Bio-informatics pipeline

Fastq files were pre-processed through a bio-informatics pipeline developed in the team on the Nextflow platform (30). Briefly, the first step removed Illumina adaptors. The second step de-multiplexed the sequences according to their plate barcodes. Then, all sequences containing at

least 4T following cell barcode and UMI were kept. Using UMItools whitelist, the cell barcodes and UMI were extracted from the reads. The sequences were then mapped on the reference transcriptome (Gallus GallusGRCG6A.95 from Ensembl) and UMI were counted. Finally, a count matrix was generated for each plate.

Data filtering, normalization and analysis

All analysis were carried out using R software (4.0.5;(31)). Matrixes from the eight plates were pooled together. Cells were filtered based on several criteria: reads number, genes number, counts number and ERCC content. For each criteria the cut off values was calculated as follows:

$$\text{mean} - 3 * \text{sd}$$

We selected genes present in at least two cells. Filtered matrix was then normalized using SCTransform from Seurat package (32) and we corrected for batch effect, time effect and sequencing depth effect. Expression matrix was finally log1p transformed.

Variable genes were identified using FindVariableFeatures from Seurat, vst method (33) and we retained the 2000 most variable features. Differentially expressed genes were identified using FindMarkerGenes function from Seurat (33). Analysis was done comparing conditions two by two, genes with log fold change $\geq 0,5$ and adjusted p-value $< 0,05$ were kept as significant. More information on QC filtering is given in supplementary 1.

Statistical analysis

All statistical analysis were performed using R software (4.0.5; (31)).

Dimensionality reduction and visualization was performed using UMAP (34).

Adaptive Sparse PLS for Logistic Regression was performed using the plsgenomics package (26).

Wasserstein distances computation was done using the Transport R package (35), and was accomplished for each genes of the scRNA-seq dataset.

Gini index were calculated using the Ineq R package on Wasserstein distances distributions (36).

Bootstrapping was done using sample_frac function from Dplyr R package(37).

Estimation of multi-stability levels

For estimating the level of multi-stability in the data, we consider that the probability law of each gene can be approximated by a Gamma distribution, or a mixture of Gamma distributions, since they are known to describe accurately continuous single-cell data (38). More precisely, we parametrize the distribution of a gene i by:

$$p_i^m(x) = \sum_{j=1}^m \mu(j) \frac{x^{a_i^j-1} b_i^j e^{-b_i x}}{\Gamma(a_i^j)},$$

where Γ denotes the Gamma function: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Note that only the parameters $(a_i^j)_{j=1,\dots,m}$ depends on the mixture component j : this is related to the distribution arising from the well-established two-states model of gene expression(39), when only the frequency of the transitions between the states of the promoters depends on the gene expression product, as described in(40).

For every condition, we constructed 10 training sets consisting of 80% of the cells in the population (randomly-chosen), and we estimated the parameters $[(a_i^j)_{j=1,\dots,m}, b_i]$ with a MCMC algorithm for the numbers of mixture components $m = 1, 2, 3$ successively. We then consider that the optimal number of components for gene i is the one which minimizes the average BIC score estimated on the 10 corresponding test sets.

Results

Single-cell transcriptomics analysis

We analyzed the expression of 83 relevant genes identified in a previous published study describing T2EC differentiation (21) using scRT-qPCR. We obtained the expression level of those 83 genes in 173, 173, 168, 171 single cells for 0H, 24H, 48H Differentiation and 48H Reversion, respectively.

As a quality control, we performed Pearson's correlation analysis on gene expression values for the 83 genes in the sc-RTqPCR published dataset (21) and our sc-RTqPCR dataset. Correlation value was 0,85 (p-value = $2.2e-16$) between the two experiments.

We also generated scRNA-seq libraries to study fate reversion at a genome wide scale. To do so, we used an adapted version of the MARS-seq protocol ((21)- see material and methods) – the main difference in protocol consist in using random hexamer primers incorporated by reverse transcription instead of ligation. After filtering, we obtained gene expression level in 174, 181, 169, 186 single cells for 0H, 24H, 48H Differentiation and 48H Reversion, respectively.

Using Seurat FindVariableFeatures to identify the most variable genes of the data set, we retained the 2000 most variable genes for our subsequent analysis.

We performed Pearson's correlation analysis on gene expression value between genes common in the sc-RTqPCR dataset and the sc-RNAseq dataset. We retained 74 common genes between the two datasets and the correlation value for those gene expression was 0,73 (p-value = $1,34e-13$), indicating good level of gene expression correlation between the two techniques.

Dimensionality reduction analysis captures a strong similarity between the reversion cells and the undifferentiated cells

We started by questioning if reversion cells (48H reversion) were more similar to one of the other subgroup of differentiating cells. To answer this question, we used UMAP dimensionality reduction method to capture the general tendencies of the 48H Reversion cells on both dataset (FIGURE 2).

First, we projected the 4 conditions on two dimensions for the scRT-qPCR dataset and the scRNA-seq dataset (FIGURE 2: Panel A and B). We focused on the normal differentiation process using the 3 time points of differentiation (0H, 24H and 48H Differentiation) (FIGURE 2: Panel C - H). For both experiments, while the different conditions don't form obvious clusters, there is a visible separation between the conditions 0H and 48H Differentiation (FIGURE 2: Panel G and H). When focusing on 24H differentiation cells, when projected two by two, we see that they tend to overlap with both 0H cells (Figure 2: Panel C and D) and 48H differentiating cells (Figure 2: Panel E and F). This may suggest that the cells at 24H are "exploring" the space of gene expression by a rise in gene expression variability as previously described (13,21).

We then projected the 48H Reversion cells to compare this condition to the other ones (FIGURE 2: panel I - N). When projected the conditions two by two, we see that, the 48H reversion cells separate well from the 48H Differentiation cells (FIGURE 2: M AND N), but are visually not distinguishable from the 0H cells (FIGURE 2: I AND J).

Almost identical results were observed when, instead of plotting cells on the UMAP calculated from the mix of the 4 conditions, we recalculated the UMAPs for each pair of conditions (Supplementary 2).

Those results suggest that at both scale of observation the 48H Reversion cells transcriptome seems more similar to the undifferentiated cells than to any other condition as they are grouped together in the UMAP framework.

48H reversion cells have gene expression pattern similar to 0H cells

We then questioned if 48H Reversion cells had gene expression patterns identical to 0H cells or retained, for some genes, an expression pattern more similar to 24H or 48H differentiation cells. To address this question we performed DE analysis on scRNA-seq data to grasp potential changes at the genome wide scale (FIGURE 3). We compared conditions two by two. As expected, the "normal" erythrocyte differentiation process showed an increase in the expression of hemoglobin related genes during the kinetics (*HBBA*, Hemoglobin Alpha-Locus 1 (*HBA1*), and Hemoglobin Alpha, subunit D (*HBAD*)). On the other hand, 0H cells expressed high level of *LDHA* (Lactate Dehydrogenase A), marker for glycolysis metabolism used by self-renewing cells (41)

and *ID2* (Inhibitor Of DNA Binding 2), coding for a transcription factor involved in differentiation inhibition (42) (FIGURE 3: panel A, B and C).

Interestingly, when comparing 0H with 48H Reversion, we saw only one gene that was significantly differentially expressed just above the threshold (FIGURE 3: panel D), *RSFR* (RNase Super Family Related), that is highly expressed in precursor cells from chicken bone marrow (26). Furthermore, when comparing 48H Reversion with 48H Differentiation, we found hemoglobin related genes up in the 48H Differentiation condition and *LDHA* and *ID2* up in 48H Reversion condition (FIGURE 3: panel E).

We then investigated closer genes expression distribution within the different conditions to see how genes expression pattern would evolve during the reversion process (FIGURE 4). We selected a few genes differentially expressed and which expression increases or decreases during the differentiation process. *HBA1*, *HBBA*, *HBAD* and *FECH* (Ferrochelatase) are involved in hemoglobin and heme pathways and are more expressed by differentiating cells while *LDHA*, *ID2*, *CSTA* (cystatin A1) and *CRIP1* (Cysteine-rich intestinal protein1) are more expressed by proliferating undifferentiated cells. We plotted and compared their distribution between the 4 conditions. For the genes involved in differentiation, we see a shift in the distributions toward a higher level of expression as cells get more differentiated (FIGURE 4: Panel A - D) and we see the opposite shift for genes involved in proliferation (FIGURE 4: Panel E - H). The 48H Reversion cells expression pattern for those genes seem to shift back to patterns closer to the 0H cells. At the time of observation and especially for genes up in differentiation, the 48H Reversion cells expression pattern are not completely similar to those of 0H cells, the offset observed could probably be due to mRNA and proteins half-life and degradation which show that they actually entered the differentiation process before reverting back.

To go further on genes distributions comparison, we computed Wasserstein distances – geometric distances well suited for comparing multimodal distributions - for each 2000 genes of the scRNA-seq dataset between each condition two by two. We then obtain 6 distributions of Wasserstein distances value. Finally, we measured Gini index in each distribution. The Gini index is a measure of statistical dispersion. The higher Gini index, the higher inequality among the

values. We performed 100 bootstraps and compared the Gini values obtained (FIGURE 5). Distribution of Wasserstein distances between 0H cells and 48H Reversion cells had the smallest Gini index among all distributions. This result points toward a closer global transcriptional activity between 48H reversion cells and 0H cells.

48H reversion cells retain molecular traces of a commitment into differentiation

To further characterize the molecular changes we sought to identify predictive genes that discriminate the most the reversion cells and the three differentiation conditions (0H, 24H and 48H Differentiation) using firstly, the sc-RT-qPCR dataset. We performed a statistical analysis through a sparse PLS least squares regression analysis (20) for each pair of temporal conditions (FIGURE 6). Interestingly, when performing sparse PLS analysis on the 0H cells versus the 48H reversion cells, *HBBA* was the only gene that discriminating between the two cell groups. In order to visualize the difference in *HBBA* expression between the 0H and 48H reversion cells, we used the previous UMAP projections of the two conditions and colored the cells according to their *HBBA* expression values (FIGURE 6: panel A). Furthermore, 48H reversion cells have significantly higher *HBBA* expression than 0H cells (Wilcoxon test p-val= 7.689e-16; FIGURE 6: panel D).

Furthermore, when performing sparse PLS analysis on the 48H reversion cells versus the 48H differentiation cells, *HBBA* and *PTPRC* were the two genes most discriminating between the two cell groups (FIGURE 6: panel B and C). *PTPRC* also called *CD45*, codes for an ubiquitous membrane glycoprotein expressed on almost all hematopoietic cells, except for mature erythrocyte (25). Looking closely, 48H Reversion cells have a significantly lower expression of *HBBA* than 48H differentiation cells (Wilcoxon test p-val= < 2.2e-16; FIGURE 6: panel D) and a significantly increased *PTPRC* expression than 48H differentiation cells (Wilcoxon test p-val= < 2.2e-16; FIGURE 6: panel E).

These results confirmed that the 48H reversion cells display a gene expression pattern close to those of 0H cells while still retaining traces of their engagement into the differentiation process by expressing *HBBA* at an intermediate level between differentiated and undifferentiated cells.

Population 24H is not a mix of two population

At that stage, two postulates could be made: 1. Either all cells have engaged into a differentiation process, and do molecularly revert to a self-renewal transcriptional state or 2. At 24 hours of differentiation two subpopulations coexist: one which is still undifferentiated and would give rise to the 48H Reversion cells and a second more differentiated which would lead to the 48H Differentiation population and die in the reversion experiment.

We hypothesized that the existence of two subpopulations at 24 hours should lead to a higher number of modes in the distribution of some genes at that time point. To test this hypothesis, we therefore estimated for each condition the most-likely number of components for the probability distribution of each gene, as assessed through a Gamma mixture (see material and methods). We found no significant difference of components distribution between the 4 populations (FIGURE 7). This confirms that the cells collected from 24H do not show more multi-stability than the other groups, and is thus unlikely to be a mix of two populations.

The alternative hypothesis would also imply that in the 24H population the cells engaged too far in the differentiation process would die a short time after media was changed, while only the undifferentiated ones would survive. We then measured viability rate during the kinetics and found no difference in viability between the conditions and especially between the 24H conditions and the 48H Reversion condition (Supplementary 3).

Finally, the last hypothesis would imply that the reversion cells are simply cells that did not yet enter the differentiation process. It would therefore be at odds with the evidence that the 48H Reversion cells do not exactly overlap with 0H cells, but do retain traces of their engagement into the differentiation process (see upper).

Those results strongly suggest that the 24H cell population is not composed of two coexisting subpopulations of cells.

Discussion

In the present study, we molecularly characterized avian erythrocyte progenitors fate reversion by using two different single cell transcriptomics analysis techniques.

For this, cells were harvested and sorted into single cells at three time points of differentiation (0H, 24H and 48H Differentiation) and one reversion point (48H Reversion) and their single cell transcriptome analyzed. To determine proximity between the groups of cells we performed UMAP projections. The results showed an overlap between the 48H Reversion condition and the 0H cells and on the other hand, a clear separation between the 48H Reversion cells and the 48H Differentiation cells.

The analysis of gene expression distribution patterns of the 48H reversion cells confirmed a switch toward the 0H cells gene expression profile. First, DE analysis of scRNA-seq data showed only one gene significantly differentially expressed between the two conditions. Second, Wasserstein distances analysis showed closer distances between 48H reversion cells and 0H cells than to any other group. Third, sparse-PLS analysis showed that only *HBBA* expression discriminate between the 48H reversion cells and the 0H cells in the scRTqPCR dataset. This suggests that most of the 48H Reversion cells experienced a start in their differentiation process, traces of which can still be found after they were transferred back into the self-renewal medium. Finally, we validated that the 48H reversion cells did not originate from a subpopulation of undifferentiated cells that never entered the differentiation process. First, cells viability tests did not show any significant increase in mortality among any conditions. Second, analysis of gamma distribution for each gene between the different conditions, showed no significant difference between any of the conditions. These results strongly argue against the possible coexistence of two subpopulations of cells in the 24H cell population.

All of our results therefore favor the hypothesis that a vast majority of the 48H reversion cells responded to differentiation induction by modifying their gene expression profile but then returned to the self-renewal state. Of must note that this would not be the sole example of large scale transcriptome changes on (relatively) short time scales (9) . The question as to whether

such large scale transcriptome changes are accompanied by (reversible) large scale epigenetic changes remains a question for future studies.

It has been described in the literature that during cellular decision making, the cell state is maintained by dynamic interactions between positive and negative regulatory molecules (45) within the frame of a Gene Regulatory Network (GRN). These interactions can be repurposed by changing the stoichiometry of ubiquitous and specific regulatory molecules and factors (17,46). In our study, the analysis of genes expression pattern during the reversion process confirmed that the determination of the fate of erythrocyte progenitors is directed by the constraints of the dynamics of the GRN, influenced by signals emitted by the changing conditions of the environment surrounding the cells. In the absence of differentiation signals (or in the presence of self-renewal inducing signals), there is no ratchet in place that would prevent (at least at early stages in our case) the system to return to its original quasi-steady state. This is in excellent agreement with the previous demonstration that there was a duration threshold for some GRN under which the systems can return back to its original state (47). This was proposed to allow cells to discriminate between *bona fide* signals and random noise in their environment, and could represent a physiological system for finely tuning the *in vivo* production of red blood cells while preserving the pool of progenitors.

Taken together, our results point toward a physiological plasticity and reversibility with respect to erythrocyte decision-making. In term of epigenetic landscape, our work implies that instead of a continuous gradient that the cells will roll down as in the classical Waddington's depiction (14), differentiation should be more appropriately described as cells moving from wells to wells, that is from one metastable state (38–40) to an other (FIGURE 8). This view abides by the multistability framework where a complex quasi-potential landscape aims at describing both normal and pathological differentiation processes (10,38). It is important at that stage to remember that Waddington himself was aware that his drawing was but a simplification. Adapting and refining this landscape should not be considered as departing from his views. Such a non-monotonous landscape has been proposed to account for the depiction of regeneration in adult tissues (48), and is consistent with previously proposed dynamical principles of cell fate restriction (16). It is also in good accordance with the recent depiction that “climb uphill on

Waddington's epigenetic landscape" during cranial neural crest cells development (20). It is beyond the scope of this discussion to go into details, but a cell "climbing uphill" should be seen as equivalent as "the landscape bending into a new valley".

We recently proposed a methodology for inferring the GRN underlying T2EC differentiation (49). For that we kept in silico cells under constant differentiation stimulus. It would be of interest to see if the inferred GRN would be able to revert, up to a certain point where no "spontaneous" return is possible (21), to its original state. This would be a very strong constraint to impose and should severely limit the number of putative GRN able to reproduce experimental data.

In conclusion, this work has provided a detailed molecular characterization of the probabilistic nature of erythrocyte cell fate determination, influenced by the constraints of the underlying gene regulatory network dynamics, and driven by environmental influences. It adds one more layer of constraint that should be incorporated into the executable GRN inference scheme.

Figures

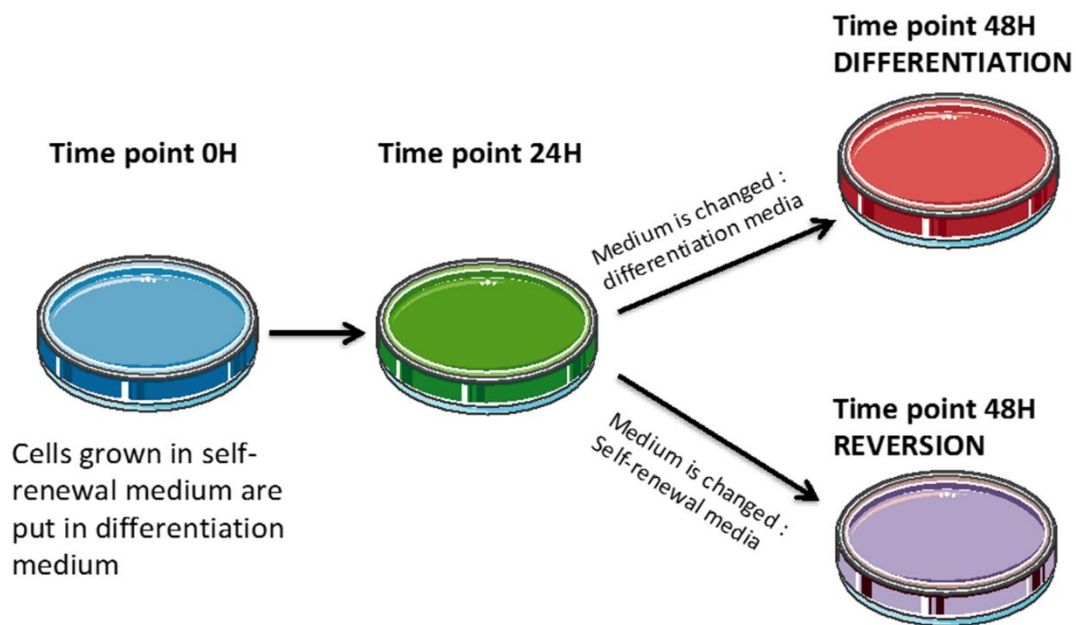


Figure 1: Experimental design.

At 0H, cells grown in self-renewal medium are shifted in differentiation medium. 24H later, cell population is divided in two, half being grown back into self-renewal medium, half being kept in differentiation medium. At each time point, 192 cells are collected for each subsequent experiment: scRT-qPCR and scRNA-seq.

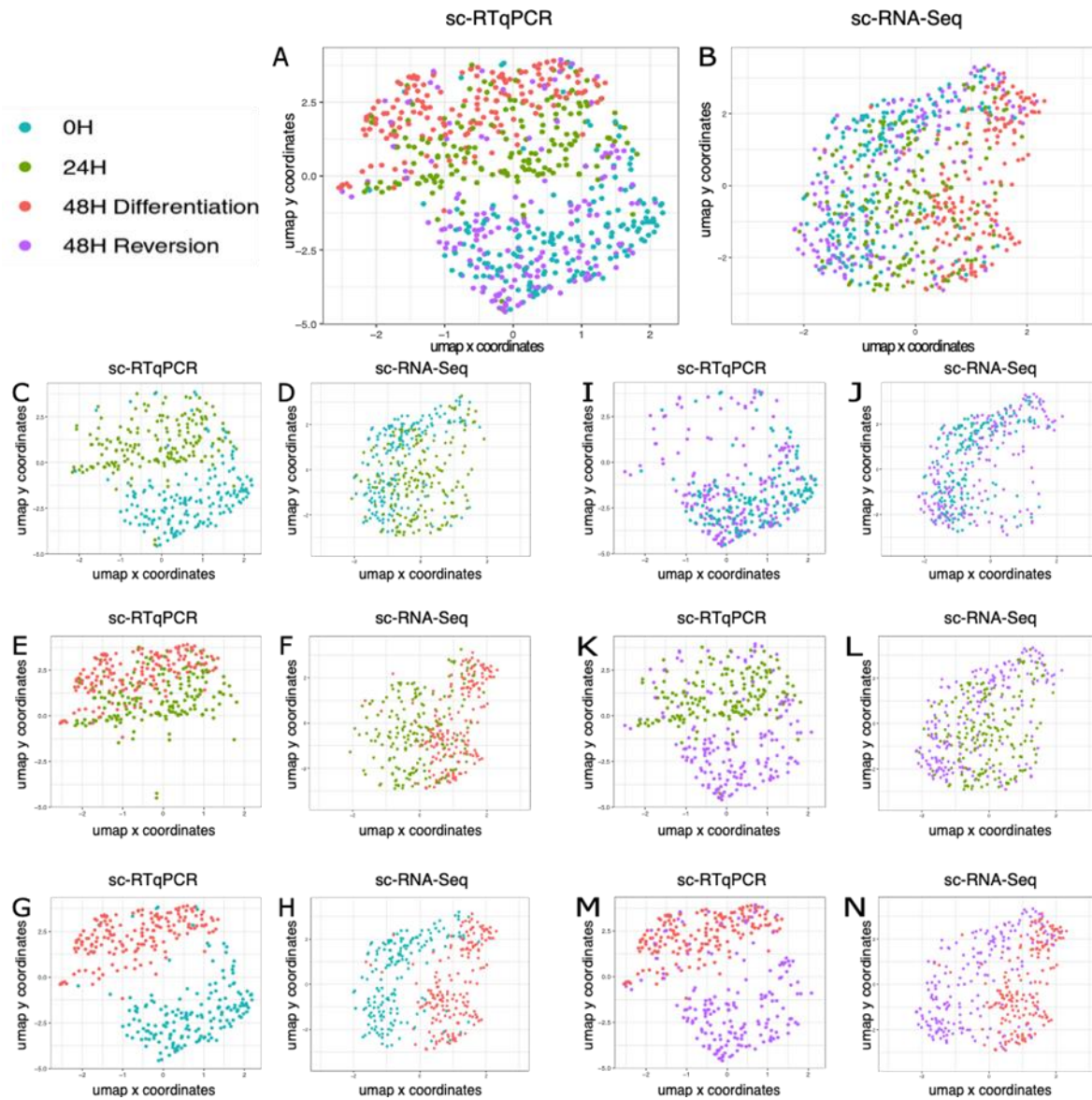


Figure 2: UMAP visualization of scRT-qPCR and scRNA-seq data. All UMAPS are calculated using the 4 biological conditions. 0H cells are displayed in blue, 24H cells in green, 48H Differentiation cells in red and 48H Reversion cells in purple.

Panels A, C, E, G, I, K and M: scRTqPCR data.

Panels B, D, F, H, J, L and N: scRNAseq data.

Panels A and B: all 4 conditions

Panels C and D: condition 0H and 24H

Panels E and F: condition 24H and 48H Differentiation

Panels G and H: condition 0H and 48H Differentiation

Panels I and J: condition 0H and 48H Reversion

Panels K and L: condition 24H and 48H Reversion

Panels M and N: condition 48H Differentiation and 48H Reversion

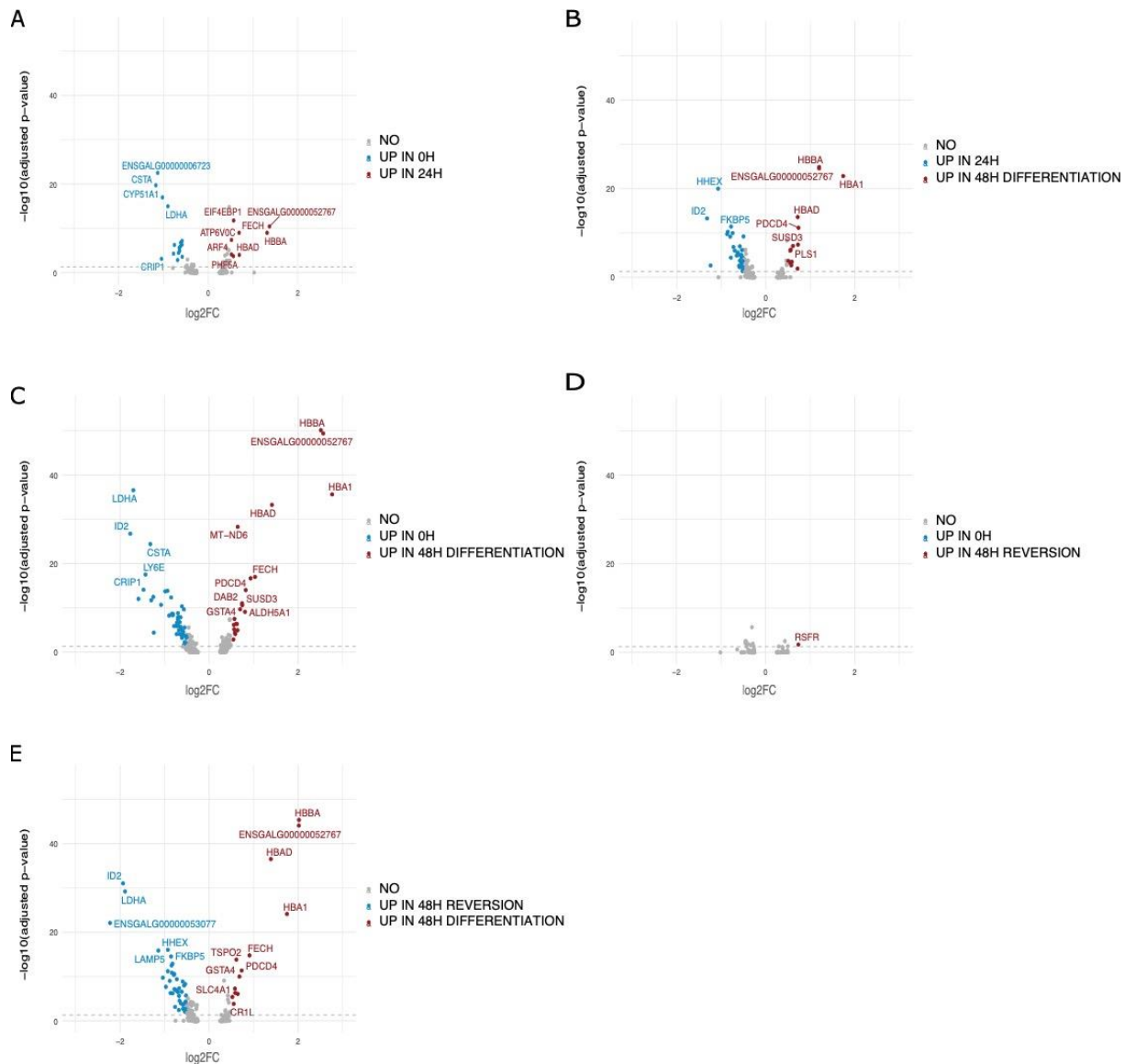


Figure 3: Volcano plot of genes differentially expressed between conditions analyzed two by two. Genes are considered significantly differentially expressed when the fold change is equal or above 0,5 and adjusted p-value is below 0,05 (grey dotted line).

Panel A: Condition 0H and condition 24H. Blue dots represent significantly up-regulated genes in 0H condition and red dots represent significantly up-regulated genes in 24H condition.

Panel B: Condition 24H and condition 48H Differentiation. Blue dots represent significantly up-regulated genes in 24H condition and red dots represent significantly up-regulated genes in 48H Differentiation condition.

Panel C: Condition 0H and condition 48H Differentiation. Blue dots represent significantly up-regulated genes in 0H condition and red dots represent significantly up-regulated genes in 48H Differentiation condition.

Panel D: Condition 0H and condition 48H Reversion. Blue dots represent significantly up-regulated genes in 0H condition and red dots represent significantly up-regulated genes in 48H Reversion condition.

Panel E: Condition 48H Reversion and condition 48H Differentiation. Blue dots represent significantly up-regulated genes in 48H Reversion condition and red dots represent significantly up-regulated genes in 48H Differentiation condition.

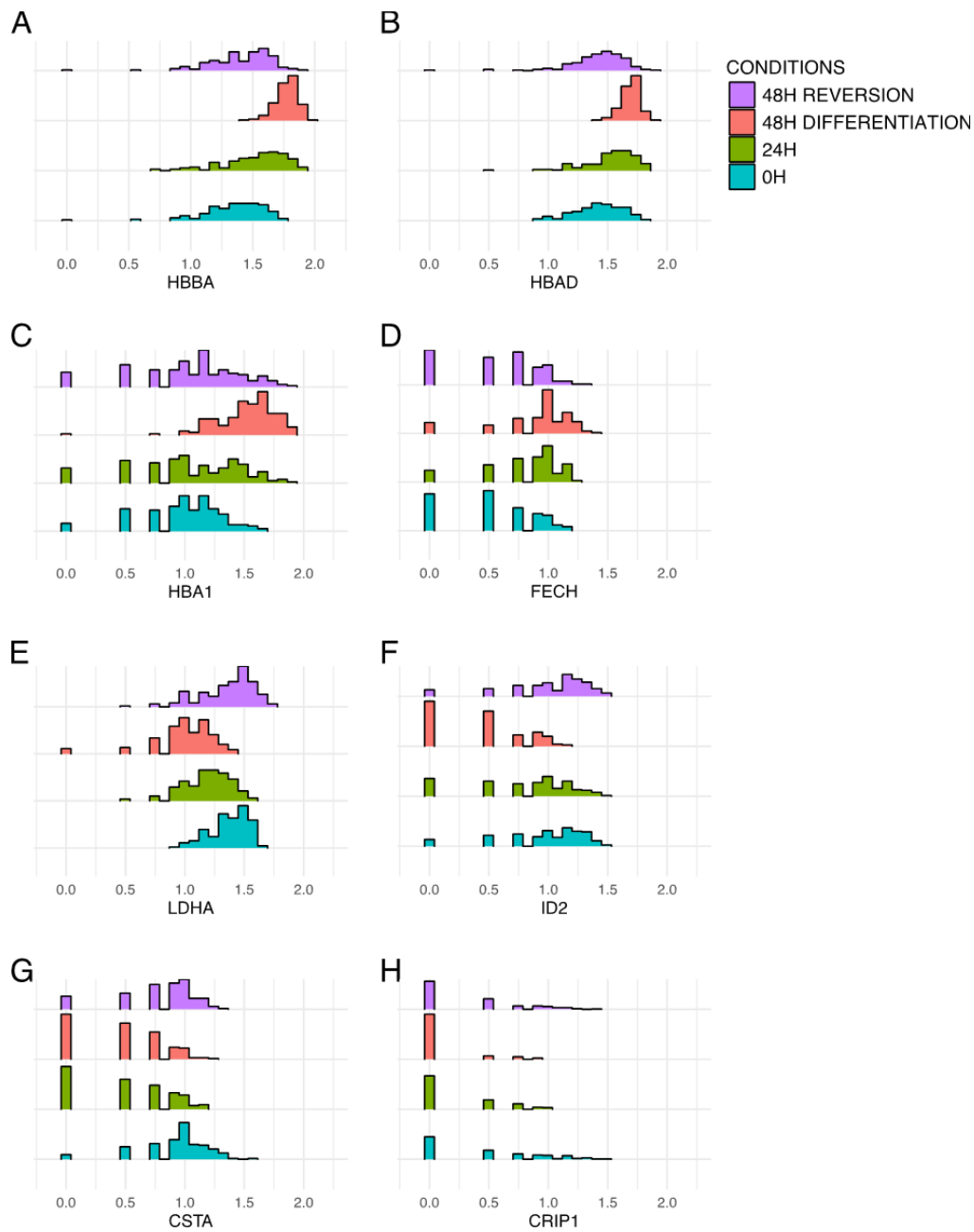


Figure 4: Comparison of gene expression pattern distribution between 48H Reversion cells and cells at four experimental time-points (0H, 24H, 48H Differentiation and 48H Reversion). Y scale represents log1p of genes expression.

Color legend: cells from condition 0H (in blue), 24H after differentiation induction (in green), 48H after differentiation induction (in red), 48H reversion (in purple).

Histograms of gene expression distribution for *HBBA* (Panel A), for *HBAD* (Panel B), for *HBA1* (Panel C), for *FECH* (Panel D), for *LDHA* (Panel E), for *ID2* (Panel F), for *CSTA* (Panel G) and for *CRIP1* (Panel H).

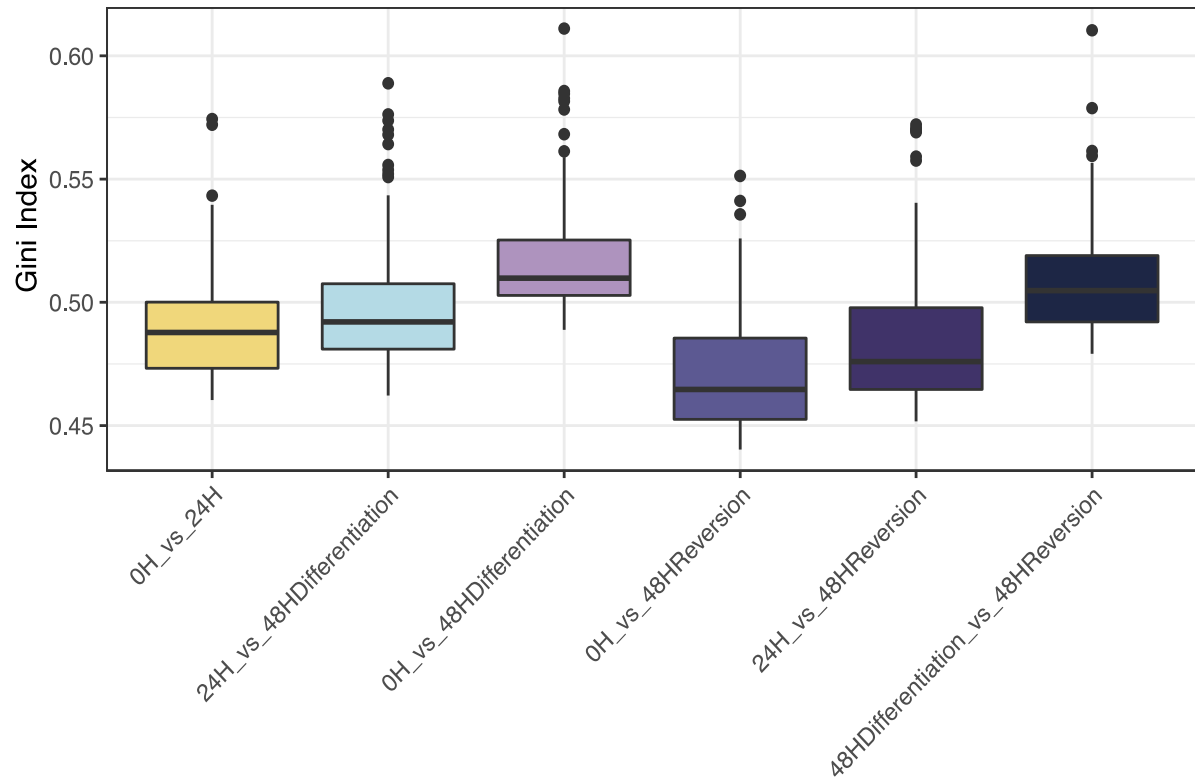


Figure 5: Plot of Gini index values of Wasserstein distances distribution between conditions in pairs computed for each of the 2000 genes from scRNA-seq data bootstrapped 100 times.

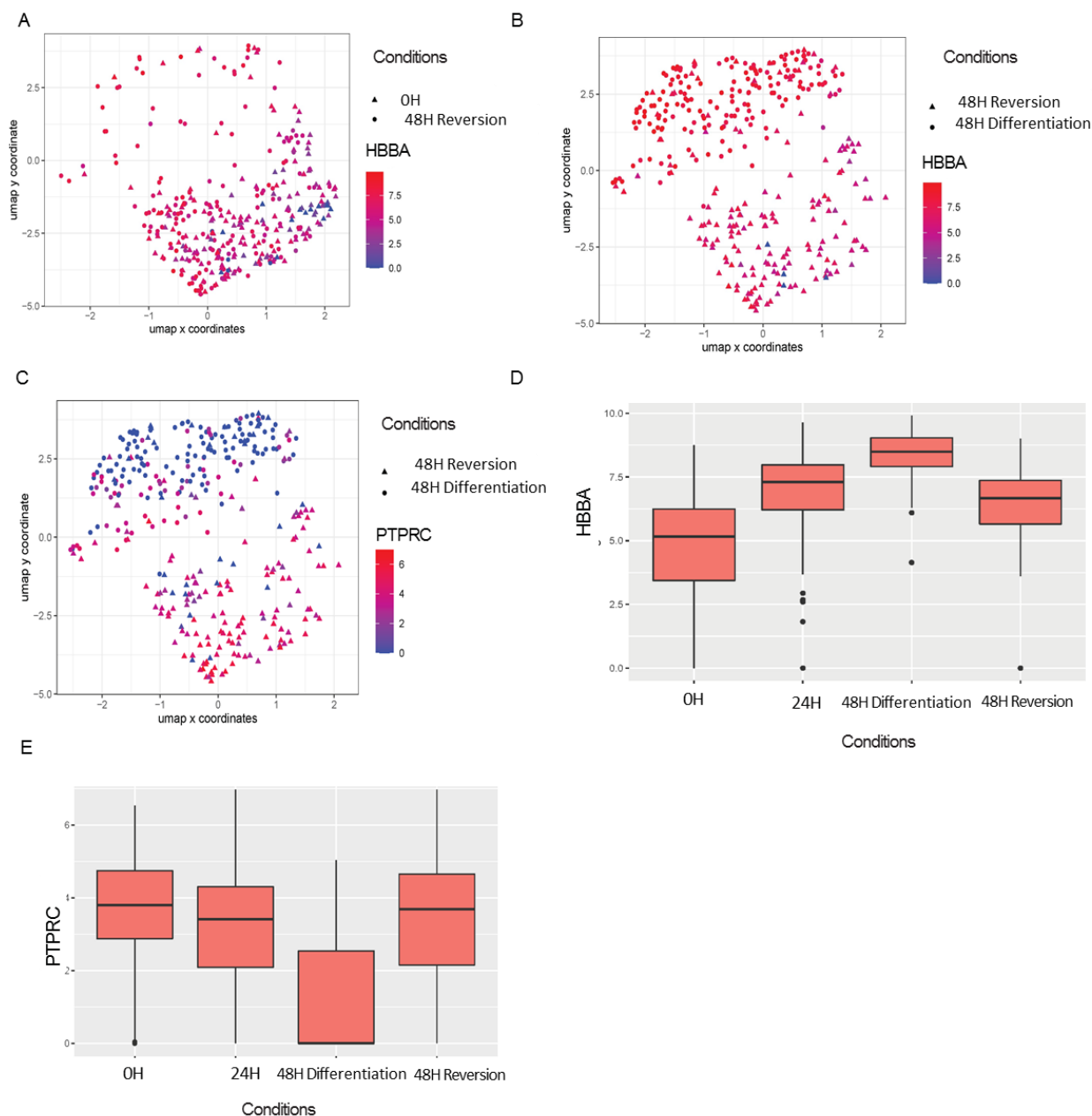


Figure 6: Gene prediction analysis by sparse PLS for 48H Reversion cells vs 0H cells and 48 Differentiation cells.

Panel A: UMAP of RT-qPCR colored based on *HBBA* expression level. Only conditions 0H and 48H Reversion are projected.

Panel B: UMAP of RT-qPCR colored based on *HBBA* expression level. Only condition 48H Reversion and 48H Differentiation are projected.

Panel C: UMAP of RT-qPCR colored based on *PTPRC* expression level. Only condition 48H Reversion and 48H Differentiation are projected.

Panel D: Boxplots of median expression level of *HBBA* in log1p expression in the 4 conditions (0H, 24H, 48H Differentiation and 48H Reversion)

Panel E: Boxplots of median expression level of *PTPRC* in log1p expression in the 4 conditions (0H, 24H, 48H Differentiation and 48H Reversion).

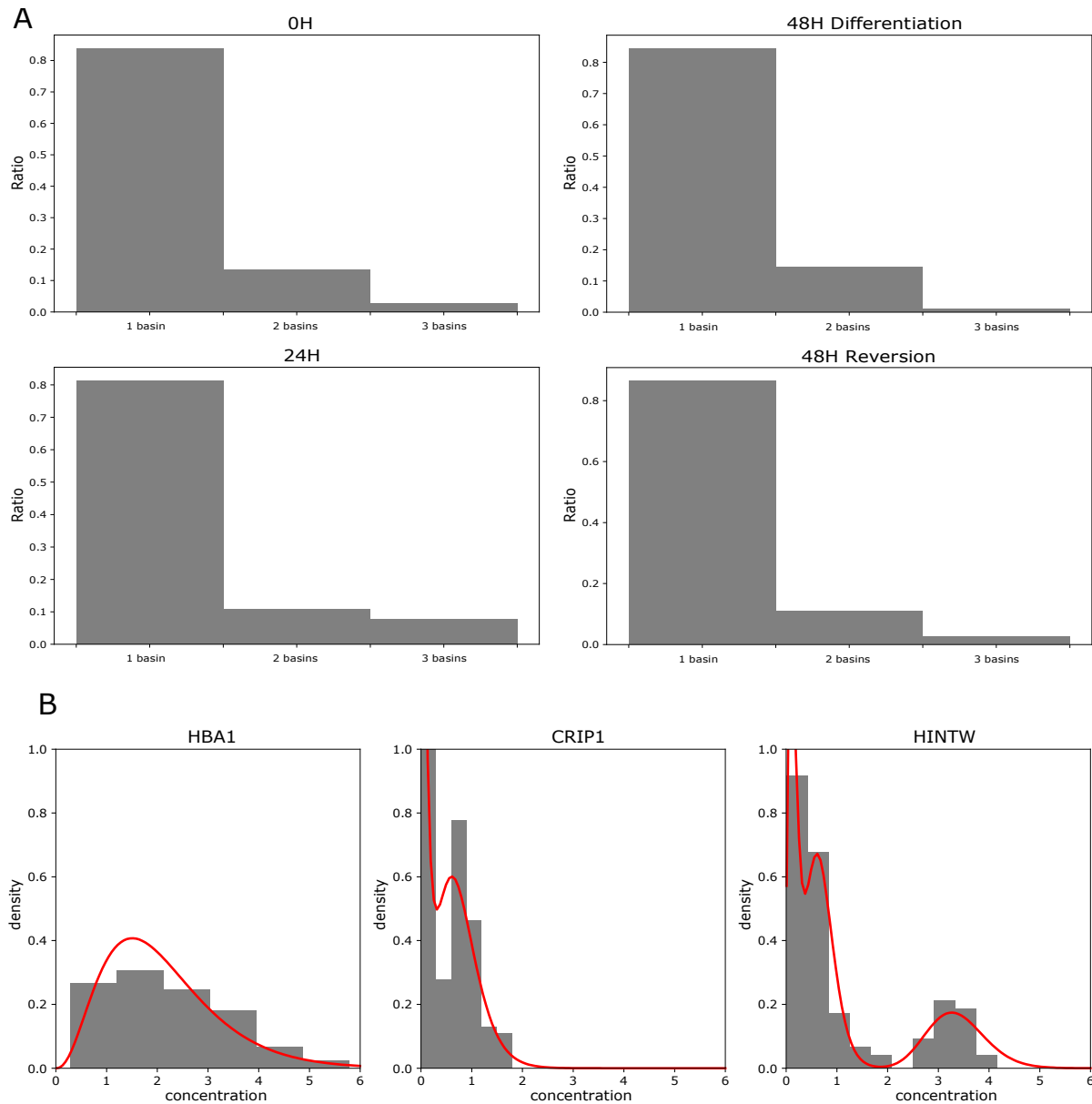


Figure 7: The repartition in the number of basins which have been detected for the 200 most variables genes, characterizing the level of multi-stability which is observed for each condition.

Panel A: Repartition in the number of components for each biological condition.

Panel B: Examples of genes which distribution fits 1 basin (left), 2 basins (middle), 3 basins (right).

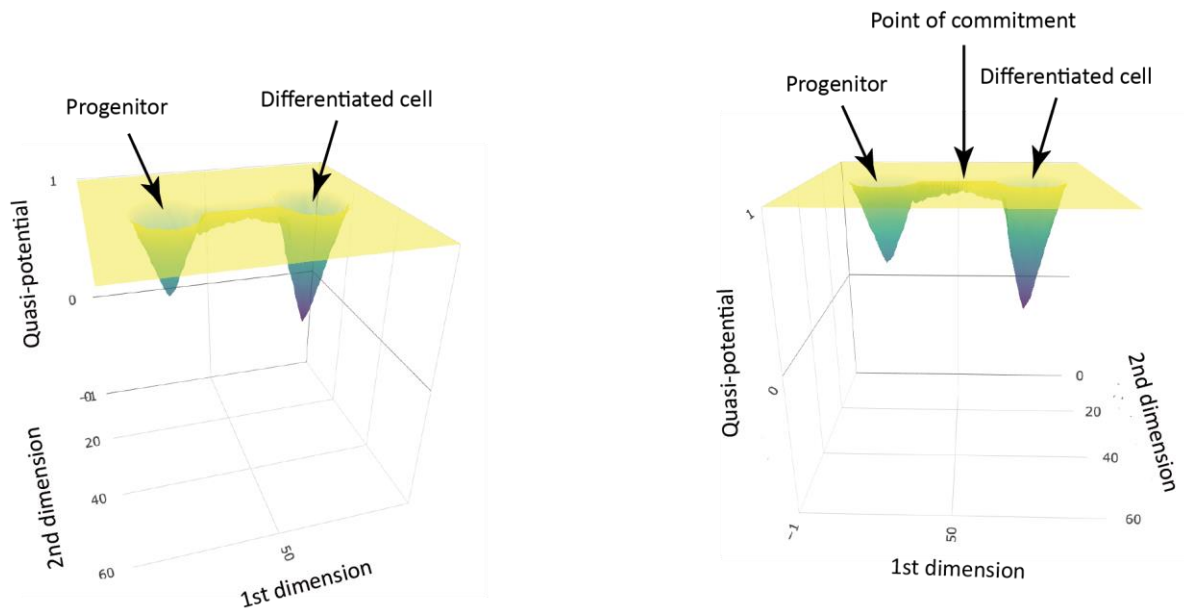
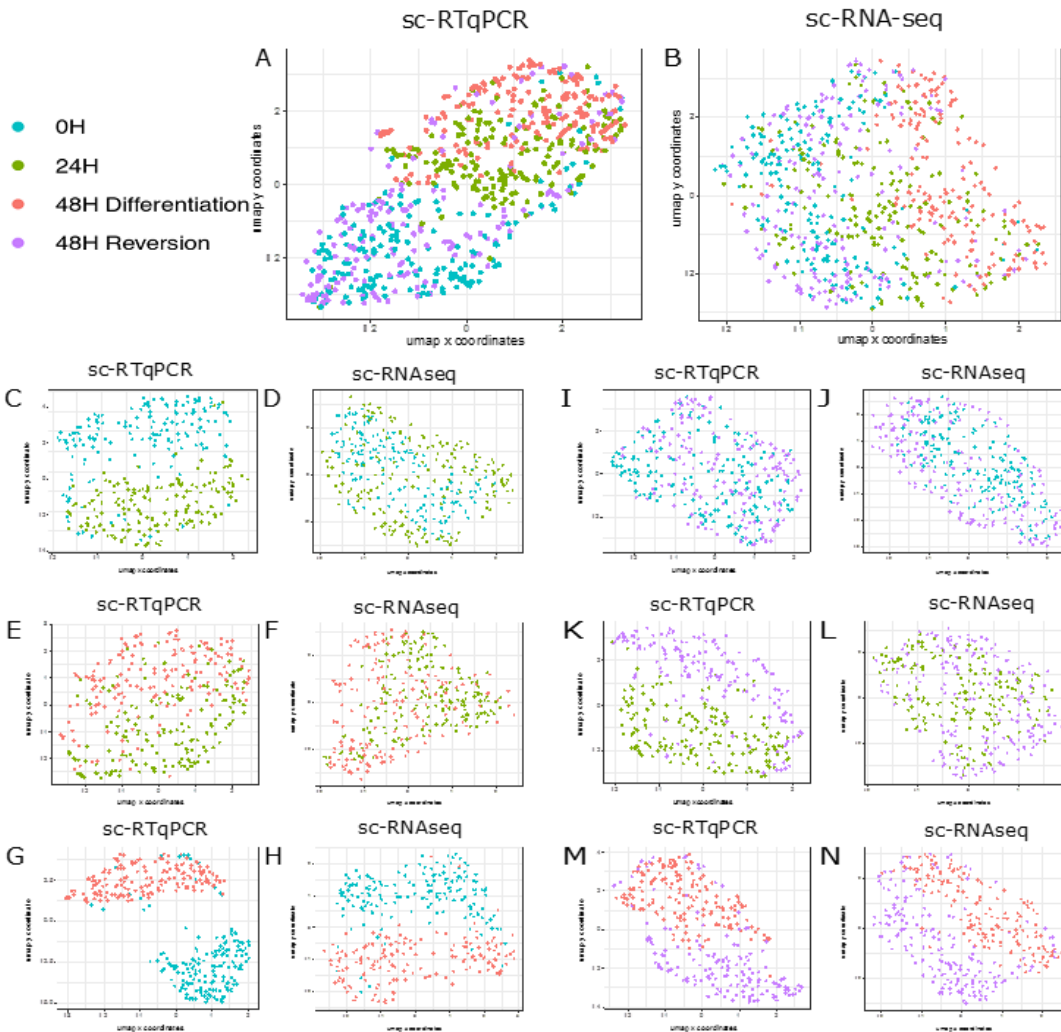


Figure 8: A quasi-potential wells depiction of the erythroid differentiation process. While the cells have not escaped the zone of influence of the progenitor attractor (i.e. when they have not passed the point of commitment, aka the point of no return **(21)**) the removal of the environmental influences results in their relaxing back to their original attractor state. The two images are different angle views.

Supplementary

<u>QC step</u>	<u>Filter value</u>	<u>Number of cells remaining</u>	<u>Number of features remaining</u>
Number of sequencing reads (per cell)	80000 reads	741 cells	/
Percentage of mapped reads (per cell)	46%	740 cells	/
Percent of reads mapped to ERCC (per cell)	<20%	710 cells	/
Number of detected genes (per cell)	233 genes	710 cells	/
Number of UMIs (per cell)	1584 UMIS	710 cells	/
Normalization	/	710 cells	8682 genes
Variable features selection	2000 most variable genes	710 cells	2000 genes

Supplementary 1: Summary table of scRNA-seq data filtering steps and threshold values for each step. Threshold determination is explained in material and methods.



Supplementary 2: UMAP visualization of scRT-qPCR and scRNA-seq data. All UMAPS are recalculated when projecting conditions two by two. 0H cells are displayed in blue, 24H cells in green, 48H Differentiation cells in red and 48H Reversion cells in purple.

Panels A, C, E, G, I, K and M: scRTqPCR data.

Panels B, D, F, H, J, L and N: scRNAseq data.

Panels A and B: all 4 conditions

Panels C and D: condition 0H and 24H

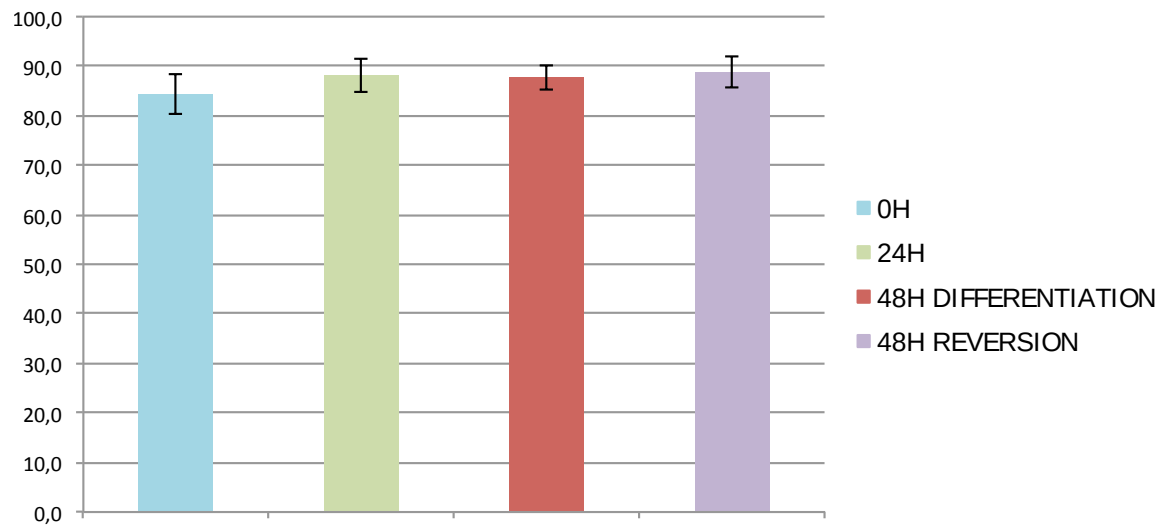
Panels E and F: condition 24H and 48H Differentiation

Panels G and H: condition 0H and 48H Differentiation

Panels I and J: condition 0H and 48H Reversion

Panels K and L: condition 24H and 48H Reversion

Panels M and N: condition 48H Differentiation and 48H Reversion



t-test	0H	24H	48H DIFFERENTIATION	48H REVERSION
24H	0,10469355	/	/	/
48H DIFFERENTIATION	0,11659985	0,78210894	/	/
48H REVERSION	0,06275294	0,80544221	0,565747784	/

Supplementary 3: Histograms of viability rate during reversion and differentiation measured by Trypan blue staining and counting in a Malassez chamber and p-values from the associated t-tests.

Acknowledgements

We gratefully thank all members of SBDM team for very fruitful discussions and particularly Franck Picard and Gerard Benoit for their suggestions and critics of our project. We also thank Ghislain Durif for his great technical support during PLS computation.

References

1. Wolpert L. Do we understand development? *Science*. 1994 Oct 28;266(5185):571–2.
2. Elowitz MB. Stochastic Gene Expression in a Single Cell. *Science*. 2002 Aug 16;297(5584):1183–6.
3. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet*. 2002 May;31(1):69–73.
4. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. Goodell MA, editor. *PLoS Biol*. 2009 Jul 7;7(7):e1000149.
5. Blake WJ, Balázs G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Mol Cell*. 2006 Dec;24(6):853–65.
6. Symmons O, Raj A. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol Cell*. 2016 Jun;62(5):788–802.
7. Kolodziejczyk AA, Lönnberg T. Global and targeted approaches to single-cell transcriptome characterization. *Brief Funct Genomics*. 2018 Jul 1;17(4):209–19.
8. Hume DA. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood*. 2000 Oct 1;96(7):2323–8.
9. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*. 2008 May;453(7194):544–7.
10. Huang S. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*. 2009 May;31(5):546–60.
11. Moris N, Edri S, Seyres D, Kulkarni R, Domingues AF, Balayo T, et al. Histone Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity. *Stem Cells Dayt Ohio*. 2018 Dec;36(12):1828–38.
12. Guillemin A, Duchesne R, Crauste F, Gonin-Giraud S, Gandrillon O. Drugs modulating stochastic gene expression affect the erythroid differentiation process. *PLoS ONE*. 2019 Nov 21;14(11):e0225166.
13. Moussy A, Cosette J, Parmentier R, da Silva C, Corre G, Richard A, et al. Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. Huang S, editor. *PLOS Biol*. 2017 Jul 27;15(7):e2001867.
14. Waddington, C.H. *The Strategy of the Genes* [Internet]. 1st ed. Routledge; 1957. Available from: <https://doi.org/10.4324/9781315765471>
15. Shi J, Teschendorff AE, Chen W, Chen L, Li T. Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. *Brief Bioinform*. 2018;5.
16. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet*. 2016 Nov;17(11):693–703.
17. Baron, Margaret H. Reversibility of the differentiated state in somatic cells. *Curr Opin Cell Biol*. 1993;5(6):1050–6.
18. Johnson NC, Dillard ME, Baluk P, McDonald DM, Harvey NL, Frase SL, et al. Lymphatic endothelial cell identity is reversible and its maintenance requires Prox1 activity. *Genes Dev*. 2008 Nov 17;22(23):3282–91.
19. Rajagopal J, Stanger BZ. Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? *Dev Cell*. 2016 Jan;36(2):133–7.

20. Zalc A, Sinha R, Gulati GS, Wesche DJ, Daszczuk P, Swigut T, et al. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science*. 2021 Feb 5;371(6529):eabb4776.
21. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process. Teichmann SA, editor. *PLOS Biol*. 2016 Dec 27;14(12):e1002585.
22. Ladewig J, Koch P, Brüstle O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol*. 2013 Apr;14(4):225–36.
23. Bendall SC, Nolan GP. From single cells to deep phenotypes in cancer. *Nat Biotechnol*. 2012 Jul;30(7):639–47.
24. Huang J, Ghosh P, Hatfull GF, Hong Y. Successive and Targeted DNA Integrations in the *Drosophila* Genome by Bxb1 and ϕ C31 Integrases. *Genetics*. 2011 Sep 1;189(1):391–5.
25. Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, et al. Cell Fate Decision as High-Dimensional Critical State Transition. *PLoS Biol*. 2016 Dec 27;14(12):e2000640.
26. Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P. A Sparse PLS for Variable Selection when Integrating Omics Data. *Stat Appl Genet Mol Biol* [Internet]. 2008 Jan 18 [cited 2021 Jun 8];7(1). Available from: <https://www.degruyter.com/document/doi/10.2202/1544-6115.1390/html>
27. Gandrillon O, Schmidt U, Beug H, Samarut J. TGF- β cooperates with TGF- α to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *EMBO J*. 1999 May 17;18(10):2764–81.
28. Gandrillon O, Samarut J. Role of the different RAR isoforms in controlling the erythrocytic differentiation sequence. Interference with the v-erbA and p135gag-myb-ets nuclear oncogenes. *Oncogene*. 1998 Feb 5;16(5):563–74.
29. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*. 2014 Feb 14;343(6172):776–9.
30. Nextflow - A DSL for parallel and scalable computational pipelines [Internet]. [cited 2021 Sep 24]. Available from: <https://www.nextflow.io/>
31. Team R. A language and environment for statistical computing. *Computing*. 2006 Jan 1;1.
32. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression [Internet]. *Genomics*; 2019 Mar [cited 2021 Jan 11]. Available from: <http://biorxiv.org/lookup/doi/10.1101/576827>
33. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single cell data [Internet]. *Genomics*; 2018 Nov [cited 2019 Nov 20]. Available from: <http://biorxiv.org/lookup/doi/10.1101/460147>
34. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018 Dec 3;37(1):38–44.
35. Schuhmacher D, Bähre B, Gottschlich C, Hartmann V, Heinemann F, Schmitzer B. transport: Computation of Optimal Transport Plans and Wasserstein Distances. R package version 0.12-2 [Internet]. 2020. Available from: <https://cran.r-project.org/package=transport>
36. Zeileis A. ineq: Measuring Inequality, Concentration, and Poverty [Internet]. 2014.

Available from: <https://CRAN.R-project.org/package=ineq>

37. Hadley Wickham , Romain François , Lionel Henry , Kirill Müller. dplyr: A Grammar of Data Manipulation [Internet]. 2021. Available from: <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>
38. Ventre E, Espinasse T, Bréhier C-E, Calvez V, Lepoutre T, Gandrillon O. Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastability [Internet]. Systems Biology; 2020 Sep [cited 2021 Sep 1]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.09.04.283176>
39. Guillemin A, Stumpf MPH. Noise and the molecular processes underlying cell fate decision-making. Phys Biol. 2021 Jan 7;18(1):011002.
40. Pisco AO, Fouquier d'Hérouël A, Huang S. Conceptual Confusion: The case of Epigenetics [Internet]. Systems Biology; 2016 May [cited 2021 Sep 1]. Available from: <http://biorxiv.org/lookup/doi/10.1101/053009>
41. Richard A, Vallin E, Romestaing C, Roussel D, Gandrillon O, Gonin-Giraud S. Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements. PLOS ONE. 2019 Sep 4;14(9):e0221472.
42. Yokota Y, Mori S, Narumi O, Kitajima K. In vivo function of a differentiation inhibitor, Id2. IUBMB Life. 2001 Apr;51(4):207–14.
43. Klenova EM, Botezato I, Laudet V, Goodwin GH, Wallace JC, Lobanenko VV. Isolation of a cDNA clone encoding the RNASE-superfamily-related gene highly expressed in chicken bone marrow cells. Biochem Biophys Res Commun. 1992 May;185(1):231–9.
44. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. Nat Rev Genet. 2016 Nov;17(11):693–703.
45. Blau HM. Differentiation requires continuous active control. Annu Rev Biochem. 1992;61:1213–30.
46. Johnson NC, Dillard ME, Baluk P, McDonald DM, Harvey NL, Frase SL, et al. Lymphatic endothelial cell identity is reversible and its maintenance requires Prox1 activity. Genes Dev. 2008 Dec 1;22(23):3282–91.
47. Sokolik C, Liu Y, Bauer D, McPherson J, Broeker M, Heimberg G, et al. Transcription Factor Competition Allows Embryonic Stem Cells to Distinguish Authentic Signals from Noise. Cell Syst. 2015 Aug;1(2):117–29.
48. Rajagopal J, Stanger BZ. Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? Dev Cell. 2016 Jan 25;36(2):133–7.
49. Bonnafeux A, Herbach U, Richard A, Guillemin A, Giraud S, Gros P-A, et al. WASABI: a dynamic iterative framework for gene regulatory network inference. bioRxiv. 2018 Apr 4;292128.

2.2 La différenciation érythrocytaire : Du réseau de gènes à l'analyse fonctionnelle

2.2.1 Introduction

Le gène *FNIP1*, présentant des fonctions biologiques nombreuses et diversifiées, a été identifié comme un des gènes les plus incertains, impliqués dans de nombreuses interactions avec d'autres gènes constituant les GRN candidats sous-jacent au processus de différenciation des T2EC.

Dans le but de réduire le nombre de GRN candidats et d'identifier le plus pertinent, nous avons utilisé la technologie d'édition du génome CRISPR-Cas9 pour induire le KO du gène *FNIP1*, et étudier notamment ses conséquences sur l'expression génique.

La technologie CRISPR-Cas9 est l'une des approches les plus puissantes, utilisée pour générer des délétions et/ou des mutations dans les gènes dans différents types cellulaires et notamment les cellules primaires (H. W. Chu et al. 2015) (Brunetti et al. 2018) (Anderson et al. 2017).

Dans une première partie de mon travail, nous avons construit plusieurs vecteurs CRISPR-Cas9 double guides, chacun ciblant des exons différents de *FNIP1* capables de produire une coupure d'ADN visible et facilement identifiable au niveau génomique. Nous avons également construit des vecteurs simple guides suite aux résultats inattendus obtenus avec les vecteurs double guides. Nous avons ensuite évalué l'efficacité de ces vecteurs au niveau génomique et transcriptomique sur des populations cellulaires. En effet, une validation de l'efficacité du KO au niveau génomique constitue une première étape indispensable mais elle doit aussi être accompagnée d'une vérification de la présence d'un décalage du cadre de lecture sur l'ARNm.

A l'échelle de la population cellulaire, nous avons aussi étudié l'effet d'une altération du gène *FNIP1* sur l'auto-renouvellement et la différenciation des T2EC.

Dans une seconde partie de mon travail, nous avons analysé à l'échelle de la cellule unique, l'effet du KO de *FNIP1*, dans le but de mieux caractériser le GRN sous-jacent au processus

de différenciation. Les vecteurs CRISPR-Cas9 n'étant pas efficaces à 100%, une étape de génération de cellules clonales avec délétion de *FNIP1* sur deux allèles (KO 100%) ou portant au moins une délétion sur un seul allèle (KO 50%) était indispensable pour acquérir des données fiables. En effet, une seule cellule primaire possédant le gène ciblé défectueux peut être utilisée pour générer un clone d'un ensemble de cellules KO. Ces clones constituent alors un outil essentiel pour explorer l'hétérogénéité des effets d'un KO de *FNIP1* au sein d'un même clone mais aussi entre les cellules issues de clones différents. Cependant, l'obtention de clones KO à partir de cellules primaires transfectées est techniquement difficile et complexifiée par de multiples facteurs, comme l'efficacité des vecteurs mais aussi la durée de vie des cellules et l'efficacité du clonage (survie et division des cellules après isolement en clone).

L'ensemble des travaux réalisés nous ont permis de décrire un protocole optimisé pour la génération de clones KO de progéniteurs érythroïdes primaires. Dans ce protocole, nous fournissons une stratégie pour la transfection des cellules primaires, le clonage des cellules transfectées et la validation des knock-out au niveau génomique et transcriptomique sur une petite population clonale afin de pouvoir mener ensuite des analyses d'expression génique par scRT-qPCR. Ce protocole pourra être utile aux chercheurs qui souhaiteraient appliquer la technologie CRISPR pour étudier l'effet de certains gènes sur le transcriptome, dans des modèles cellulaires difficiles (cellules primaires) et en se plaçant à l'échelle de la cellule unique.

2.2.2 Matériels et Méthodes

Construction des vecteurs CRISPR simple et double guides

Le plasmide pSpCas9(BB)-2A-GFP (PX458) (Addgene plasmid # 48138 ; <http://n2t.net/addgene:48138> ; RRID : Addgene_48138) qui exprime la Cas9 et la GFP a été utilisé pour les deux stratégies (simple et double guide).

La séquence de clivage T2A (peptide 2A dérivé de thossea asigna virus) initialement présente dans le vecteur pSpCas9(BB)-2A-GFP est remplacée par une séquence P2A (peptide 2A dérivé du teschovirus-1 porcin) qui présente une efficacité de clivage plus importante (Tableau 2). Ce nouveau vecteur est appelé pCRISPR-P2A et sa séquence est validée par séquençage. Ce vecteur est encore modifié par l'ajout d'une séquence codant pour l'ARNt de poulet (tRNA Glu chicken) et d'un nouveau site multiple de restriction. Après validation par séquençage, ce plasmide final a été renommé pCRISPR-P2A-ARNt (Figure 23).

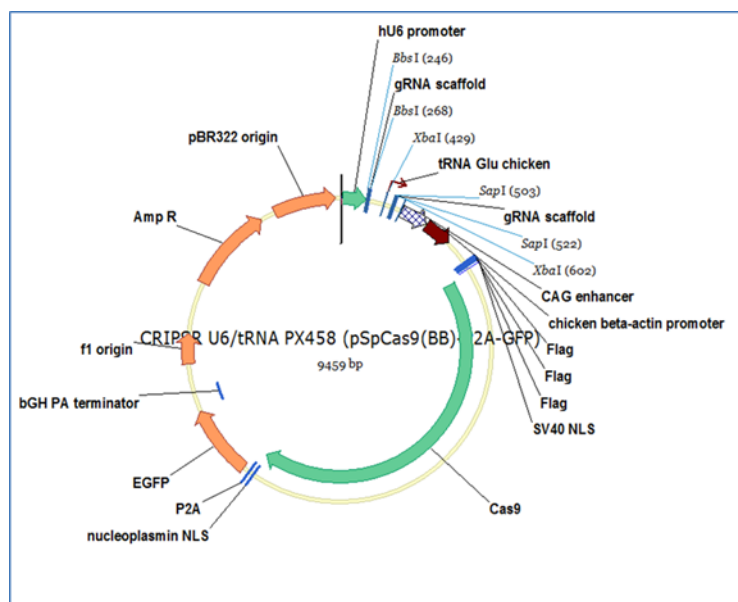


Figure 23- Plasmide pCRISPR-P2A-ARNt

Le plasmide pCRISPR-P2A-ARNt permet l'insertion d'une ou de deux séquences d'ARNg et porte les séquences codant pour la Cas9 et la GFP

Pour les deux types de vecteurs, les ARNg cibles sont définis à l'aide de l'outil crispor tefor (<http://crispor.tefor.net/>). Pour une transcription plus efficace par l'ARN polymérase III, une guanine supplémentaire est ajoutée à l'extrémité 5' de l'ARNg (Tableau 2).

Vecteur simple guide : afin d'insérer l'ARN guide, un site de reconnaissance pour l'enzyme de restriction BbsI est incorporé dans le guide. Ensuite, le vecteur pCRISPR-P2A-ARNt, est digéré avec BbsI et purifié sur gel en utilisant le kit de purification PCR (Qiagen), puis déphosphorylé en utilisant une phosphatase alcaline thermosensible, TSAP (Promega). La séquence du guide présentant des extrémités BbsI est ensuite liguée en aval du promoteur hU6.

Vecteur double guide : la même stratégie que celle décrite ci-dessus pour l'insertion du premier guide en aval du promoteur U6 a été suivie. Pour le second guide, le site de reconnaissance pour l'enzyme Sap1 a été incorporé dans le guide. Le plasmide pCRISPR-P2A-ARNt est donc digéré avec Sap1 puis déphosphorylé et le second guide présentant des extrémités Sap1 est ligué après la séquence codant l'ARNt.

L'efficacité des vecteurs CRISPR simple guide est confirmée en analysant les mutations par amplification par PCR de la zone d'ADNg ciblée, en utilisant le mix SYBR Premix Ex Taq (TaKaRa). Les produits d'amplification sont ensuite ligués dans des vecteurs pCRTM4-TOPO[®] TA grâce au kit TOPO[™] TA Cloning[™] Kit for Sequencing, without competent cells (Invitrogen). Ces vecteurs sont utilisés pour transformer des bactéries ; la zone du gène *FNIP1* ciblée est analysée par séquençage après amplification PCR sur 20 colonies bactériennes.

L'efficacité des vecteurs CRISPR double guides est confirmée en analysant dans les cellules transfectées, la délétion de *FNIP1* après amplication par PCR de la région de l'ADNg ciblée et analyse de la taille des fragments sur gel.

Oligonucléotides	Séquences (5'→3')
P2A—Forward	gccaggcaaaaaagaaaaagGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCA GGCTGGAGACGTGGAGGAGAACCCTGGACCTgtgagcaagggcgaggagct
P2A—Reverse	agctcctcgcccttgctcacAGGTCCAGGGTTCTCCTCCACGTCTCCAGCCTGCTTC AGCAGGCTGAAGTTAGTAGCTCCGCTTCCcttttctttttgcctggc
tRNA—Forward	ctagaTCCCATATGGTCTAGTGGTTAGGATTCTGGTTTTACCCAGGCGG CCCGGGTTCGACTCCCGGTATGGGAATGAAGAGCGCTCTTCAGTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTT
tRNA—Reverse	ctagAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAG CCTTATTTTAACTTGCTATTTCTAGCTCTAAACTGAAGAGCGCTCTTCA TTCCCATACCGGGAGTCGAACCCGGGCCGCTGGGTGAAAACCAGGAA TCCTAACCCTAGACCATATGGGA
FNIP1#V1—double guide— BbsI—Forward	CACC G TCCTCAGATGCCAACATGCT
FNIP1#V1—double guide— SapI—Reverse	GAA CGCACTGGAAGCAGCATCTA
FNIP1#V2—double guide— BBS1— Forward	CACC G AGTTATACCAAACAAGGGAA
FNIP1#V2—double guide— SapI—Reverse	GAA TGTTGCCAGCTGAAGCCTGG
FNIP1#V9—simple guide— BBS1— Forward	CACC G CTTGGGTCGAACTCCGGCAA

Tableau 2 - Séquences des oligonucléotides utilisés pour la construction des plasmides CRISPR-Cas9

Cellules et conditions de culture

Les BMC sont extraites de la moelle osseuse d'embryons de poulet SPAFAS white leghorn (INRA, Tours, France) âgés de 19 jours puis cultivées en incubateur (à 37 °C avec 5% de CO₂) dans du milieu LM1 (α -MEM, sérum fœtal bovin (FBS) à 10%, HEPES 1 mM, 100 nM de β -mercaptoéthanol, 100 U / mL de pénicilline-streptomycine, 5 ng / mL de TGF α , 1 ng / mL de TGF- β , et dexaméthasone 1 mM)). Afin d'obtenir des T2EC, les BMC sont maintenues en état d'auto-renouvellement dans du milieu LM1 pendant 9 jours au moins.

La différenciation des T2EC est induite en retirant le milieu LM1 et en plaçant les cellules dans le milieu DM17 (α -MEM, 10% de sérum bovin fœtal (FBS), 1 mM Hepes, 100 nM β -mercaptoéthanol, 100 U/mL pénicilline et streptomycine, 10 ng/mL insuline et 5% de sérum de poulet anémié (ACS)).

Mesure de la croissance et de la viabilité des cellules à l'échelle de la population cellulaire

Le nombre de T2EC vivantes est mesuré par comptage sur cellule de Mallassez des cellules vivantes après coloration au bleu de Trypan (SIGMA).

Mesure du taux de différenciation des cellules à l'échelle de la population cellulaire

Le niveau de différenciation est évalué en comptant les cellules bleues sur cellule de Mallassez après coloration des cellules à la benzidine (SIGMA) qui colore l'hémoglobine en bleu.

Transfection des cellules

Les cellules sont transfectées par nucléofection grâce au dispositif « NucleofectorTM 2b Device » développé par la société Lonza. Pour chaque transfection, $3 \cdot 10^6$ cellules ont été reprises dans 100 μ l de la solution de nucléofection contenant 6,5 ng de vecteur pCRISPR-P2A-ARNt. Les cellules ont été ensuite introduites dans une cuvette, placées dans le nucleofector et transfectées

avec le programme T16-K562. Après nucléofection, 500µl de milieu de culture RPMI est ajouté sur les cellules. 8 min après cet ajout, les cellules sont transférées dans le milieu d'autorenouveau LM1 sans antibiotique puis cultivées dans les conditions standards de culture.

24h après la transfection, les cellules sont récoltées et remises en suspension dans du milieu LM1. Le tri des cellules exprimant la GFP est effectué en utilisant un cytomètre en flux BD FACS Aria 1. Les cellules transfectées donc GFP positives et présentant une bonne viabilité sont ensuite replacées à l'incubateur dans du milieu de culture neuf.

Génération des clones KO

Afin de générer les clones KO, les cellules GFP+ sont triées 24h après transfection par cytométrie en flux dans des plaques 96 puits contenant du milieu LM1 avant d'être replacées à 37°C pendant une durée maximum de 3 semaines. Pour les clones qui se sont divisés, du milieu de culture neuf est rajouté au moins deux fois par semaine dans les puits.

RT-qPCR à l'échelle de la population cellulaire

L'ARN total est extrait des cellules à l'aide du kit RNeasy Mini (Qiagen) conformément au protocole du fabricant. La reverse transcription est réalisée à l'aide du SuperScript™ III First-Strand Synthesis kit (Thermo Fisher Scientific) selon le protocole du fabricant. Les réactions de qPCR sont réalisées avec le SYBR Premix Ex Taq (TaKaRa) et analysées à l'aide du système de détection CFX connect Real-Time PCR (Bio-Rad). Les amorces pour la qPCR ont été commandées auprès d'Eurogentec (Tableau 3).

Les données de RT-qPCR sont normalisées en fonction de l'expression des gènes standards *HnRNP* et *Eif5* en utilisant la formule (Schmittgen and Livak 2008) :

$$QR_{S1} = \frac{2^{-(Ct_{SrefGi} - Ct_{S1Gi})}}{2^{-(Ct_{SrefGr} - Ct_{S1Gr})}}$$

SrefGi : gène d'intérêt, condition contrôle

S1Gi : gène d'intérêt, condition expérimentale

SrefGr : gène de référence, condition contrôle

S1Gi : gène de référence, condition expérimentale

Nom des primers	Séquences (5' > 3')
<i>hnRNP—Forward</i>	CAGCAACTACTGGAATCAGGGG
<i>hnRNP—Reverse</i>	CCATAACCACTCTGCTGATTGC
<i>eif5—Forward</i>	CTGCAAGACATGTTGGATGG
<i>eif5—Reverse</i>	TCCCTGTACCAGTGTACCA
<i>FNIP1—Forward1</i>	CCCTGGGTGACAGTGAAAGT
<i>FNIP1—Reverse1</i>	TCTTCCAGGCCACATAGGTC
<i>FNIP1—Forward2</i>	CTGCAGCGATGCTCAGGT
<i>FNIP1—Reverse2</i>	AAACTACAGACTCAAAGCTACAGA

<i>FNIP1—Forward3</i>	GATGATGTTTGGCTCTGTGG
<i>FNIP1—Reverse3</i>	CTGTGGTCGGGCTTTAATGT
<i>FNIP1—Forward4</i>	AGTGTGCTTCTAAACAAGTGGA
<i>FNIP1—Reverse4</i>	CCACAGAGCCAAACATCATC
<i>FNIP1—Forward5</i>	CTGCAGCGATGCTCAGGT
<i>FNIP1—Reverse5</i>	TGATATTTTGGGCATTGTTCTTT

Tableau 3- Séquences des oligonucléotides utilisés pour les analyses en qPCR

Analyse des clones KO

7 jours après l'isolement des cellules GFP+, la moitié du volume des puits des plaques (contenant environ 100 cellules) est prélevée pour l'extraction de l'ADNg et des ARN à l'aide du kit Quick-DNA/RNA™ Microprep Plus Kit (Ozyme) conformément au protocole du fabricant.

A/ Stratégie pour les vecteurs double guides

Pour la validation au niveau génomique de la délétion de *FNIP1* par les vecteurs double guides, les ADNg sont extraits à l'aide du kit Quick-DNA/RNA™ Microprep Plus Kit (Ozyme) conformément au protocole du fabricant, amplifiés par PCR de part et d'autre de la zone de coupure puis les amplifications sont déposées sur gel (Figure 24).

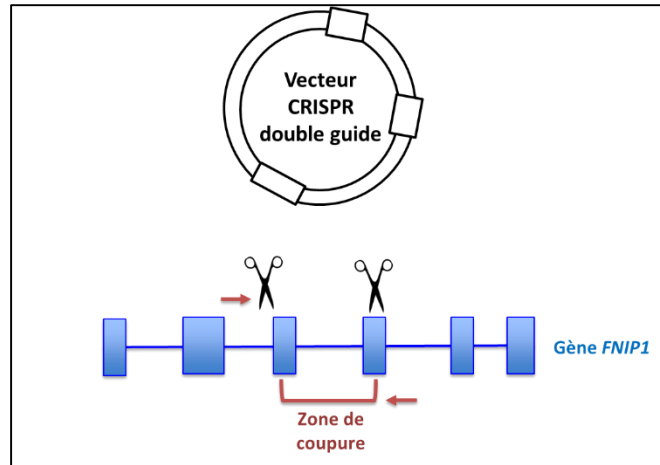


Figure 24- Analyse de la délétion de *FNIP1* au niveau génomique induite par les vecteurs CRISPR-Cas9 double guides

Afin de valider la délétion au niveau génomique, un seul couple de primers est utilisé pour amplifier *FNIP1* à partir d'ADNg de part et d'autre de la zone de coupure ciblée par les deux guides. Le couple de primers est représenté par les flèches rouges.

Pour la validation au niveau transcriptomique de l'effet du KO, les ARN sont extraits à l'aide du kit Quick-DNA/RNA™ Microprep Plus Kit (Ozyme) et la transcription inverse est réalisée à l'aide du kit Sensifast cDNA synthesis (Bioline) selon le protocole du fabricant. Les réactions de qPCR sont ensuite réalisées avec le SYBR Premix Ex Taq (TaKaRa) en utilisant différents couples de primers (se positionnant avant, de part et d'autre et après la coupure) et analysées à l'aide du système de détection CFX connect Real-Time PCR (Bio-Rad) (Figure 25). La délétion avec décalage de phase de lecture au niveau des ARN est validée par séquençage en se focalisant de part et d'autre de la coupure de *FNIP 1*.

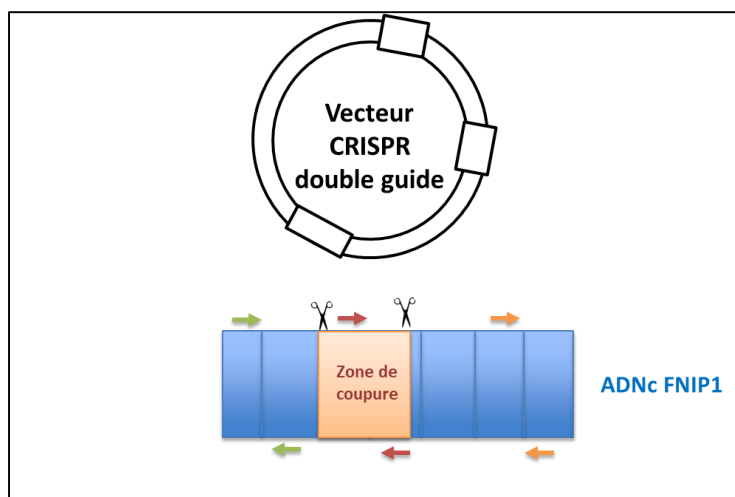


Figure 25- Analyse de la délétion de *FNIP1* au niveau transcriptomique induite par les vecteurs CRISPR-Cas9 double guides

Afin de valider la délétion au niveau transcriptomique, plusieurs couples de primers sont utilisés pour amplifier l'ADNc de *FNIP1* dans différentes régions : avant (flèches vertes), de part et d'autre (flèches rouges) et après (flèches oranges) la zone de coupure ciblée par les deux guides.

B/ Stratégie pour les vecteurs simple guide

Pour la validation au niveau génomique de la délétion de *FNIP1* par les vecteurs simple guides, les ADNg sont extraits à l'aide du kit Quick-DNA/RNA™ Microprep Plus Kit (Ozyme) conformément au protocole du fabricant puis amplifiés de part et d'autre de la zone de coupure en utilisant le SYBR Premix Ex Taq (TaKaRa) (Figure 26). Les fragments amplifiés sont ligués dans le vecteur pCR™4-TOPO® TA en utilisant le kit TOPO™ TA Cloning™ Kit for Sequencing, without competent cells (Invitrogen). Ces vecteurs sont ensuite utilisés pour transformer des bactéries ; la zone du gène *FNIP1* ciblée est analysée par séquençage après amplification PCR sur 20 colonies bactériennes.

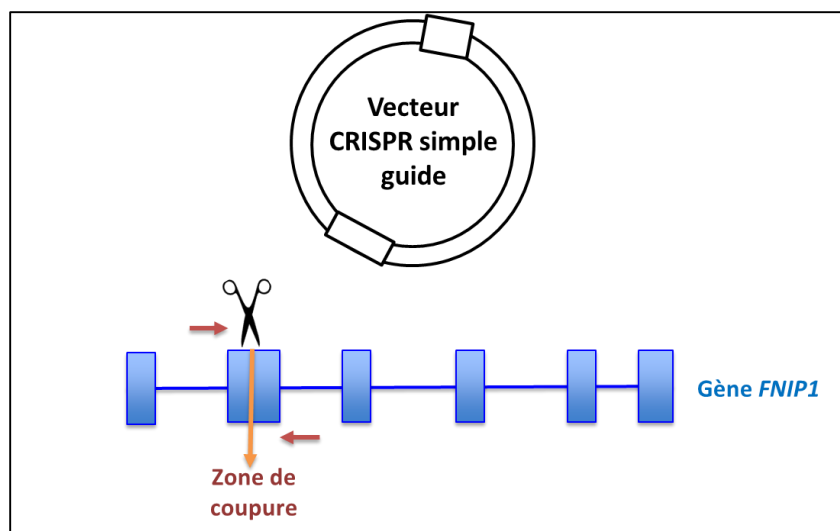


Figure 26-Analyse des délétions/mutations de FNIP1 au niveau génomique induite par les vecteurs CRISPR-Cas9 simple guides

Afin de valider la délétion/mutation au niveau génomique, un seul couple de primers est utilisé pour amplifier *FNIP1* à partir d'ADNg de part et d'autre de la zone de coupure ciblée par le guide. Le couple de primer est représenté par les flèches rouges.

Pour la validation au niveau transcriptomique de l'effet du KO, les ARN sont extraits à l'aide du kit Quick-DNA/RNA™ Microprep Plus Kit (Ozyme) et la reverse transcription est réalisée à l'aide du kit Sensifast cDNA synthesis (Bioline), conformément aux protocoles des fabricants. Ensuite, les réactions de qPCR sont réalisées en utilisant le SYBR Premix Ex Taq (TaKaRa) avec plusieurs couples de primers (situés avant, de part et d'autre et après la zone ciblée) et analysées à l'aide du système de détection CFX connect Real-Time PCR (Bio-Rad) (Figure 27).

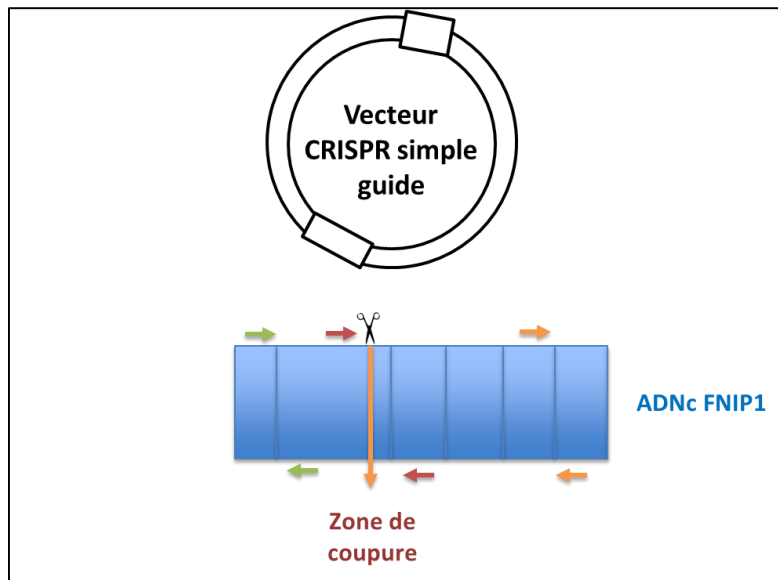


Figure 27- Analyse de la délétion de *FNIP1* au niveau transcriptomique induite par les vecteurs CRISPR-cas9 simple guides

Afin de valider la délétion au niveau transcriptomique, plusieurs couples de primers sont utilisés pour amplifier l'ADNc de *FNIP1* dans différentes régions : avant (flèches vertes), de part et d'autre (flèches rouges) et après (flèches oranges) de la zone de coupure ciblée par le guide.

RT-qPCR en microfluidique à l'échelle de la cellule unique

Les manipulations liées aux expériences de RT-qPCR à haut débit en microfluidique ont été réalisées selon le protocole recommandé par la société Fluidigm (PN 68000088 K1, p.157-172). Toutes les étapes, de l'isolement des cellules à la génération des données par RT-qPCR puis leur nettoyage, sont décrites en détail dans (Richard et al. 2016). Concernant le choix des gènes étudiés, nous avons analysé l'expression de 92 gènes identifiés dans l'étude de Richard et al. 2016.

2.2.3 Résultats

Analyse de l'expression de *FNIP 1* lors de la différenciation des T2EC

Dans un premier temps, nous nous sommes intéressés aux variations des niveaux d'expression de *FNIP 1* au cours du processus de différenciation érythrocytaire. Dans ce but, nous avons analysé les niveaux d'expression de ce gène dans les données de RT-qPCR précédemment publiées et acquises sur les progéniteurs érythrocytaires induits à se différencier pendant 0, 8, 24, 48 et 72h (Richard et al. 2016) (Figure 28).

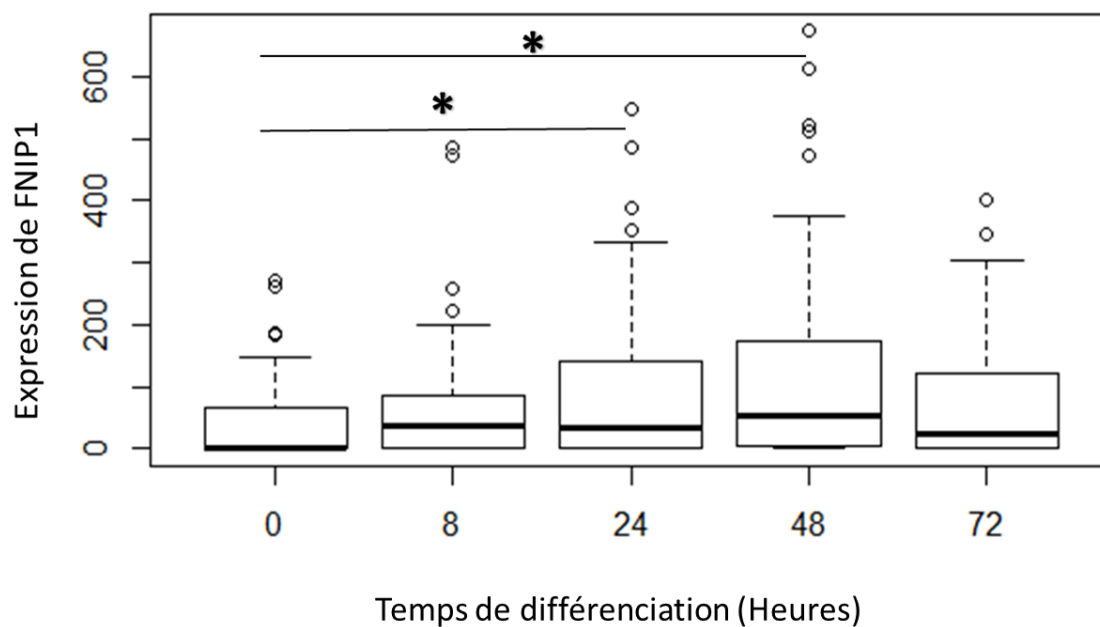


Figure 28- Boxplots représentant la distribution de l'expression de *FNIP1* dans les données de RT-qPCR précédemment publiées (Richard et al., 2016) acquises sur les progéniteurs érythrocytaires induits à se différencier pendant à 0, 8, 24, 48 et 72h.

Les tirets noirs représentent les valeurs médianes d'expression de *FNIP1* dans chacune des conditions. La significativité de la différence entre le nombre de cellules obtenu dans la condition contrôle et après invalidation de *FNIP1* a été calculée à l'aide d'un test wilcoxon (* : valeur p inf. à 0,05).

Nous pouvons constater que le niveau d'expression de *FNIP1* augmente significativement 24h et 48h après l'induction de différenciation, avant de retrouver son niveau initial au-delà de 48h de différenciation. De plus, comme mentionné dans l'introduction, *FNIP1* a été identifié comme un des gènes les plus incertains, impliqués dans de nombreuses interactions avec d'autres gènes, dans les 364 GRN candidats inférés à partir des données de scRT-qPCR obtenues sur les T2EC à 6 temps de différenciation. En conséquent, nous avons décidé de tester la pertinence de *FNIP1* dans la différenciation des T2EC. Dans ce but, nous avons utilisé la technologie d'édition du génome CRISPR-Cas9 afin d'induire le KO de ce gène.

Vecteurs CRISPR-cas9 double guides

A/ Validation de l'efficacité des vecteurs CRISPR-cas9 double guides au niveau génomique

Dans un premier temps, nous avons construit deux vecteurs CRISPR-cas9 double guides (V1 et V2) dont chacun cible des régions différentes du gène. Le vecteur 1 cible les exons 7 et 8 et induit une coupure de 382 pb et le vecteur 2 cible l'exon 6 et l'intron qui le suit induisant une coupure de 98 pb (Figure 29). L'intérêt des vecteurs double guides est de produire une coupure visible sur gel d'électrophorèse après amplification PCR de l'ADN génomique des cellules transfectées.

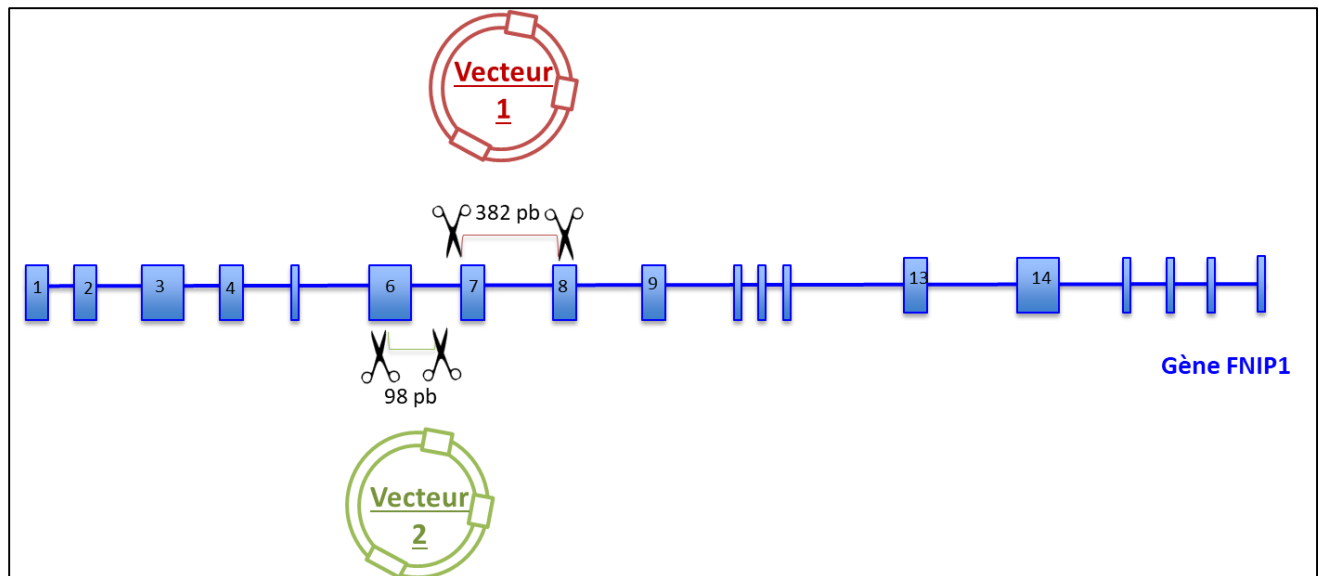


Figure 29-Les deux vecteurs CRISPR-Cas9 double guides (Vecteurs 1 et 2)

Le vecteur 1 cible les exons 7 et 8 du gène *FNIP1* et induit une coupure de 382 pb alors que le vecteur 2 cible l'exon 6 et l'intron qui le suit induisant une coupure de 98 pb.

L'efficacité des vecteurs double guides 1 et 2 est évaluée en population de cellules. Pour cela, les cellules sont transfectées avec l'une ou l'autre des constructions, triées par FACS sur la base de l'expression de la GFP puis l'ADNg est extrait, amplifié de part et d'autre de la coupure et l'amplicon est déposé sur gel. Les résultats présentés dans la Figure 30 montrent que les vecteurs 1 et 2 produisent respectivement les coupures attendues de 382 pb et 98pb. Cependant, la bande d'ADN sauvage de 598 pb est toujours présente indiquant que les deux vecteurs ne sont pas efficaces à 100%.

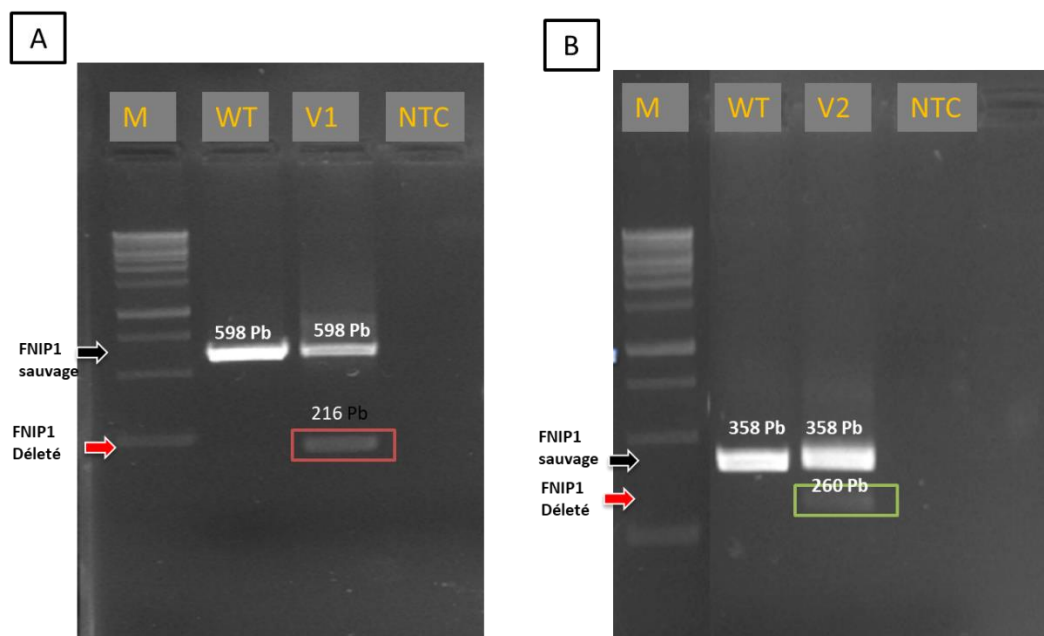


Figure 30- Validation de l'efficacité des vecteurs V1 et V2 au niveau génomique

Les cellules sont transfectées avec le vecteur vide ou les vecteurs CRISPR-Cas 9 double guides puis triées au FACS (Cellules GFP+). L'ADNg des cellules est ensuite extrait, amplifié de part et d'autre de la zone de coupure pour chaque vecteur et l'amplification est analysée sur gel. Panel A : efficacité du vecteur 1 (V1). Panel B : efficacité du vecteur 2 (V2). WT : Vecteur vide. NTC : contrôle négatif de PCR. M : marqueur de taille.

B/ Validation de l'efficacité des vecteurs CRISPR-Cas9 double guides au niveau transcriptomique

Dans le but de quantifier l'effet de la coupure induite par les deux vecteurs 1 et 2 sur le taux d'ARNm de *FNIP1*, nous avons mesuré par RT-qPCR l'expression relative de *FNIP1* dans les populations cellulaires transfectées avec les deux vecteurs par rapport aux cellules contrôles transfectées avec le vecteur vide, et triées sur l'expression de la GFP. Comme le montre la Figure 31, l'expression de *FNIP1* diminue de 48% et de 70% dans les cellules transfectées par le vecteur 1 et 2 respectivement, par rapport aux cellules contrôles. Par conséquent, l'altération génomique de *FNIP1* diminue bien le niveau d'expression de ce gène à l'échelle de la population cellulaire.

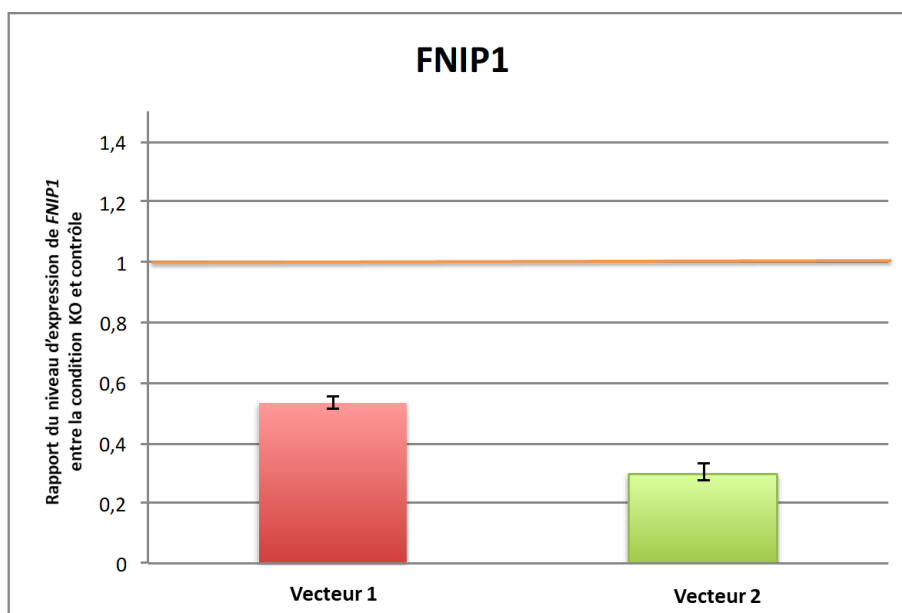


Figure 31-Rapport du niveau d'expression de *FNIP 1* dans les cellules transfectées par le vecteur 1 ou 2 par rapport aux cellules contrôles

Les T2EC ont été transfectées avec les vecteurs 1 ou 2 dirigés contre *FNIP1*. Les données présentées sont la moyenne des rapports +/- l'écart-type de l'expression de *FNIP 1* dans les T2EC transfectées par les vecteurs CRISPR-Cas9 (V1 ou V2) par rapport à la condition contrôle (vecteur vide) à partir de trois expériences indépendantes. Les données ont été au préalable normalisées en fonction de l'expression des gènes standards *hnRNP* et *eif5*.

C/ Etude de l'effet du KO de *FNIP1* sur la prolifération cellulaire et sur la survie des cellules à l'échelle de la population cellulaire

Afin d'étudier l'effet du KO de *FNIP1* sur la prolifération et la survie des cellules, nous avons réalisé des comptages et des colorations au bleu trypan des cellules maintenues en culture dans le milieu de prolifération après transfection avec les vecteurs 1 et 2 et sélection des cellules GFP+ (Figure 32). Les résultats montrent une diminution significative du nombre de cellules 24h après transfection avec les vecteurs 1 et 2 par rapport aux cellules contrôles. Ceci indique, que *FNIP1* semble être important pour la prolifération des cellules.

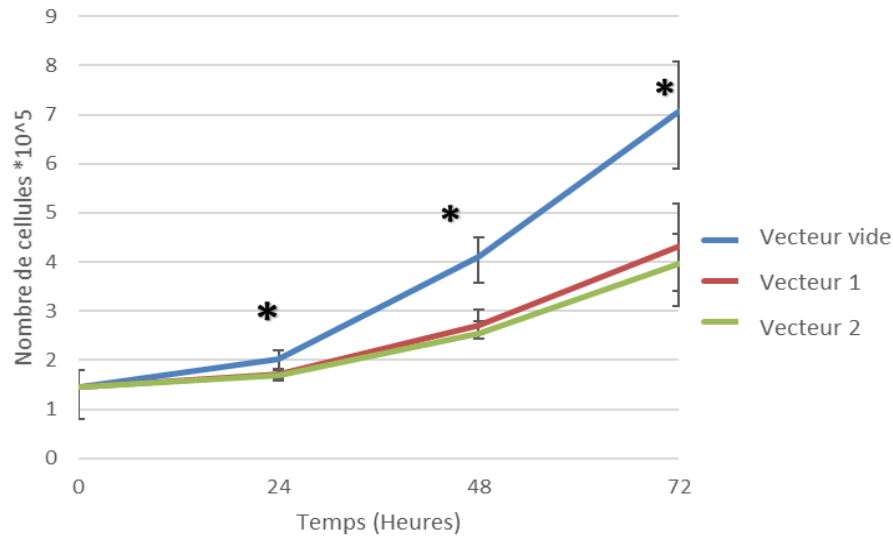


Figure 32-Mesure de taux de croissance des cellules KO pour *FNIP1*

Les T2EC ont été transfectées avec les vecteurs CRISPR 1 et 2 dirigés contre *FNIP1* ou avec le vecteur vide, triées au FACS sur la base de la GFP puis ont été comptées toutes les 24 heures pendant 3 jours. Les courbes orange et verte indiquent la croissance des T2EC transfectées par les deux vecteur V1 et V2 respectivement et la courbe bleue montre la croissance des T2EC transfectées avec le vecteur vide (contrôle). Les données présentées sont la moyenne du nombre de cellules +/- l'écart-type calculée à partir de trois expériences indépendantes. La significativité de la différence entre le nombre de cellules obtenu dans la condition contrôle et après invalidation de *FNIP1* a été calculée à l'aide d'un test t (* : valeur p inf. à 0,05).

Parallèlement, afin de déterminer si la diminution du nombre de cellules dans la condition *FNIP1*-KO était associée à une augmentation de la mortalité des cellules, nous avons réalisé des mesures de la viabilité par coloration des cellules au bleu de Trypan (Figure 33). Les résultats ne montrent aucune différence significative entre le pourcentage de cellules mortes après transfection avec les deux vecteurs CRISPR par rapport à la condition contrôle (vecteur vide), indiquant que le KO de *FNIP1* affecte la prolifération des cellules sans affecter leur mortalité.

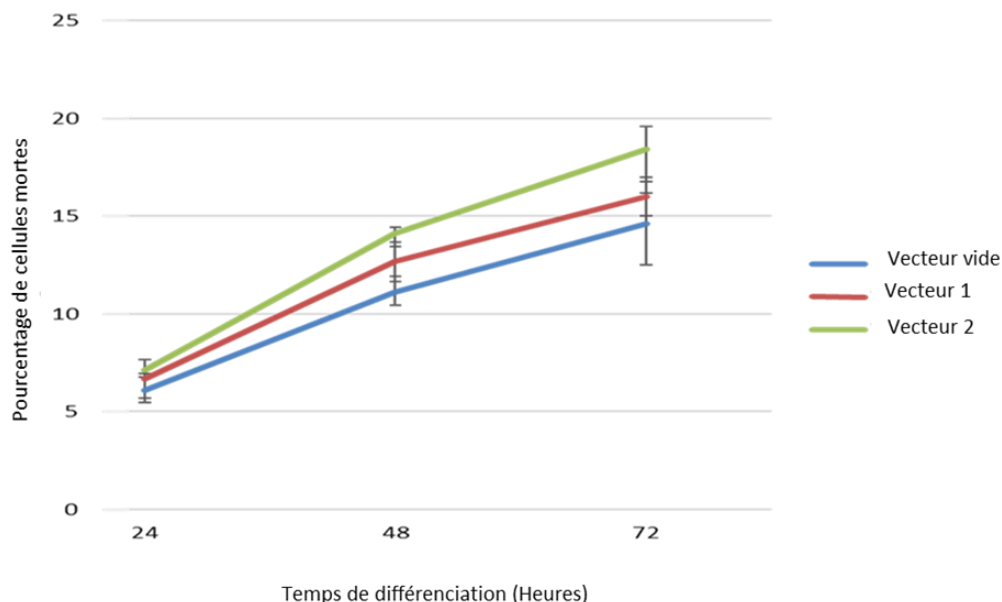


Figure 33-Mesure du pourcentage de mortalité des cellules KO pour *FNIP1*

Les T2EC ont été transfectées avec les vecteur CRISPR 1 et 2 dirigé contre *FNIP1* ou avec le vecteur vide, triées au FACS sur la base de la GFP puis comptées toutes les 24 heures pendant 3 jours. Le pourcentage de cellules mortes est calculé en utilisant la formule suivante : $(\text{Nombre de cellules bleues} \times 100) / \text{Nombre total de cellules}$. Les courbes orange et verte indiquent le pourcentage de mortalité des T2EC après transfection avec le vecteur 1 ou 2 respectivement et la courbe bleue indique le pourcentage de mortalité des T2EC avec le vecteur vide. Les données présentées sont la moyenne du pourcentage de mortalité $\pm$ l'écart-type calculée à partir de trois expériences indépendantes. La significativité de la différence entre la condition contrôle et les conditions KO a été calculée à l'aide d'un test t.

D/ Etude de l'effet du KO de *FNIP1* sur la différenciation à l'échelle de la population cellulaire

Par ailleurs, afin d'étudier l'effet du KO de *FNIP1* sur la différenciation des érythrocytes, nous avons réalisé des comptages après coloration à la benzidine des cellules maintenues dans le milieu de différenciation 48h et 72h post-transfection (Figure 34). Les résultats du comptage des cellules bleues positives à la benzidine n'ont montré aucune différence du niveau de différenciation entre des cellules transfectées par les vecteurs CRISPR-Cas9 ou le vecteur contrôle.

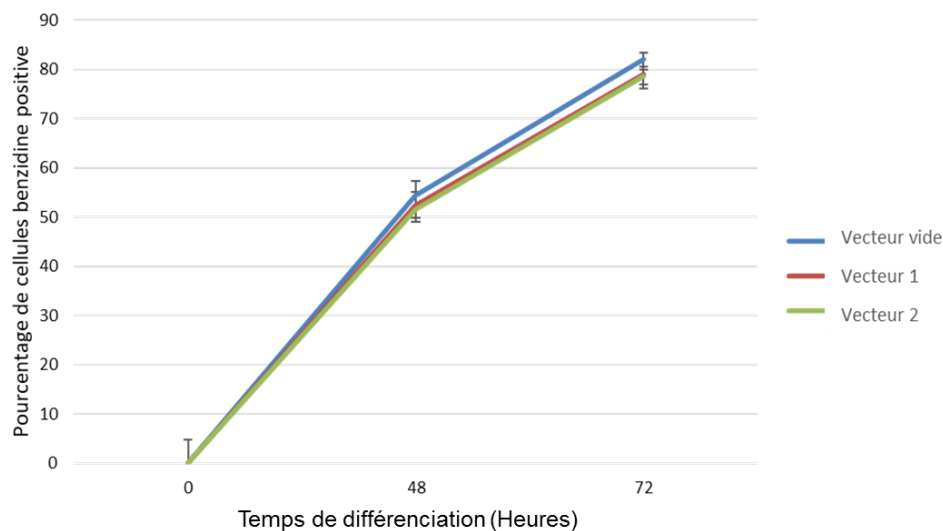


Figure 34-Mesure du pourcentage de cellules différenciées après le KO de *FNIP1*

Les T2EC ont été transfectées avec les vecteurs CRISPR dirigés contre *FNIP1* et avec le vecteur vide puis triées par FACS sur la base de la GFP et mises en culture dans le milieu de différenciation DM17. Les cellules ont été ensuite comptées 48 et 72 heures après l'induction de la différenciation, après coloration à la benzidine. Le pourcentage de cellules positives à la benzidine a été calculé en utilisant la formule suivante : $(\text{Nombre de cellules bleues} \times 100) / \text{Nombre total de cellules}$. Les courbes orange et vertes indiquent le pourcentage des cellules positives à la benzidine après transfection avec le vecteur 1 et 2, respectivement et la courbe bleue indique le pourcentage de cellules positives à la benzidine après transfection avec le vecteur vide. La significativité de la différence entre la condition contrôle et les conditions KO a été calculée à l'aide d'un test t.

De plus, nous avons étudié l'effet du KO de *FNIP1* sur la prolifération et la mortalité des cellules induites à se différencier pendant 24h, 48h et 72h après transfection. Les résultats des comptages des cellules montrent que les cellules délétées pour *FNIP1* induites à se différencier prolifèrent moins rapidement que les cellules contrôles bien qu'aucune différence de mortalité n'ait été observée (Figure 35).

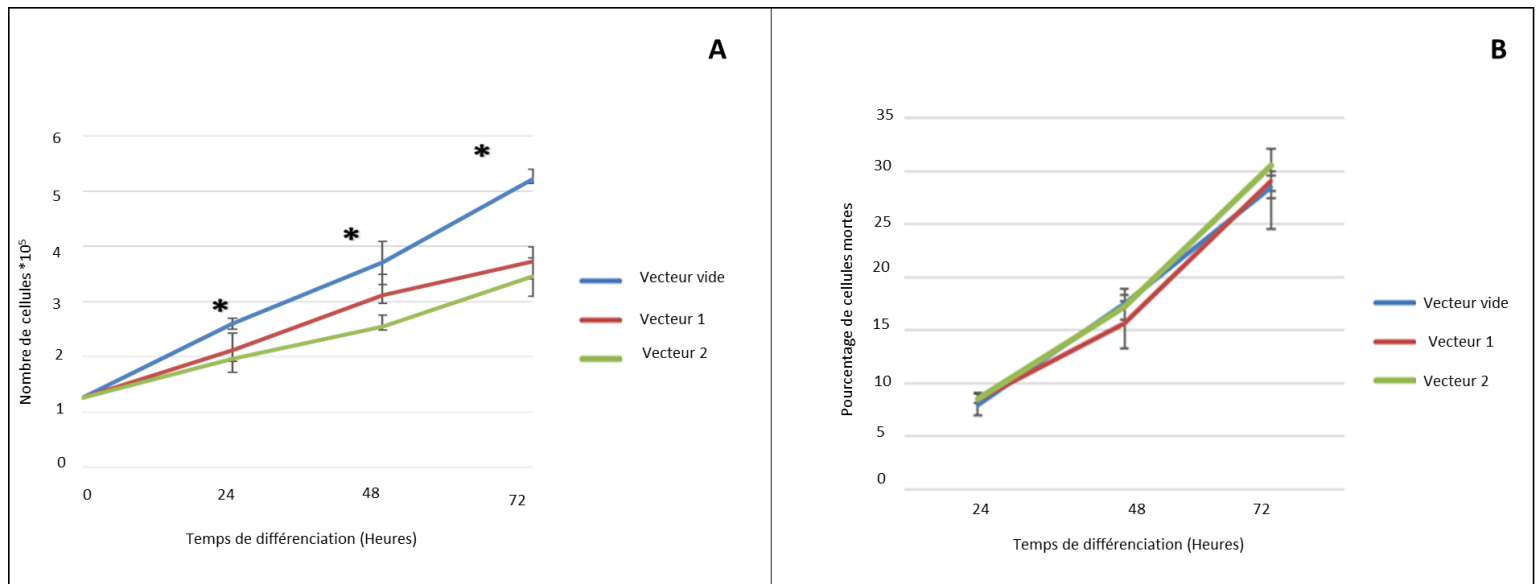


Figure 35-Mesure de la croissance et de la mortalité des cellules KO pour *FNIP1* lors de la différenciation

Les T2EC ont été transfectées avec les vecteurs CRISPR 1 et 2 dirigés contre *FNIP1* et avec le vecteur vide, triées au FACS sur la base de la GFP puis mises en culture dans le milieu de différenciation DM17. Elles ont ensuite été comptées toutes les 24 heures pendant 3 jours. La mortalité a été déterminée après coloration au bleu trypan.

A- Croissance : Les courbes orange et verte présentent la croissance des T2EC après transfection par les deux vecteurs CRISPR1 et 2, respectivement et la courbe bleue indique la croissance des T2EC après transfection par le vecteur vide. Les données présentées sont la moyenne du nombre de cellules +/- l'écart-type calculée à partir de trois expériences indépendantes

B- Viabilité : les courbes orange et verte indiquent le pourcentage de mortalité des T2EC après transfection avec les deux vecteurs CRISPR 1 et 2, respectivement et la courbe bleue indique le pourcentage de mortalité des T2EC transfectées avec le vecteur vide. Les données présentées sont la moyenne du pourcentage de mortalité (Nombre de cellules colorées au bleu trypan*100) / Nombre total de cellules) +/- l'écart-type, calculée à partir de trois expériences indépendantes.

La significativité de la différence entre la condition contrôle et les conditions KO pour *FNIP1* a été calculée à l'aide d'un test t (* : valeur p inf. à 0,05).

En conclusion, l'altération génomique de *FNIP 1* ne semble pas affecter le processus de différenciation des T2EC. Néanmoins, les cellules induites à se différencier se divisent moins rapidement que les cellules contrôles sans qu'une différence de mortalité ne soit observée.

Ces résultats confortent un rôle de *FNIP1* dans la prolifération/division cellulaire.

E/ Etude de l'effet du KO de FNIP1 à l'échelle de la cellule unique

Dans un deuxième temps, nous avons concentré nos efforts sur la caractérisation de l'effet d'un KO de 100% ou de 50% de *FNIP1* à l'échelle de la cellule unique afin de pouvoir identifier le ou les différents GRN affectés par cette perturbation. Malgré sa simplicité apparente, la réalisation d'une expérience de KO d'un gène peut être considérablement influencée par de nombreux paramètres techniques et biologiques : les caractéristiques des cellules (cellules primaires, lignées, durée de vie....) et leur vitesse de division, l'efficacité des vecteurs CRISPR-Cas9, le risque de clivages n'induisant pas de décalage de phase dans l'ARNm et/ou pas de dégradation de l'ARNm... Nous avons donc cherché à optimiser toutes les étapes pour la génération et la caractérisation de clones KO à partir de nos progéniteurs érythrocytaires aviaires primaires dont la durée de vie se situe entre 3 et 4 semaines.

Optimisation du protocole pour l'obtention de clones KO pour FNIP1 par la stratégie CRISPR-Cas 9 double guide et pour la caractérisation de ces clones avant l'analyse de leur transcriptome à l'échelle de la cellule unique

Le premier protocole que nous avons optimisé avait pour objectif de générer des clones à partir de T2EC transfectées avec les vecteurs CRISPR-cas9 double guide et de vérifier l'invalidation du gène *FNIP1* à l'échelle génomique sur une très petite population clonale (100 cellules). Ces analyses réalisées, le reste des cellules du clone d'intérêt devait être utilisé pour les analyses de l'expression génique à l'échelle unicellulaire par RT-qPCR en microfluidique (Figure 36).

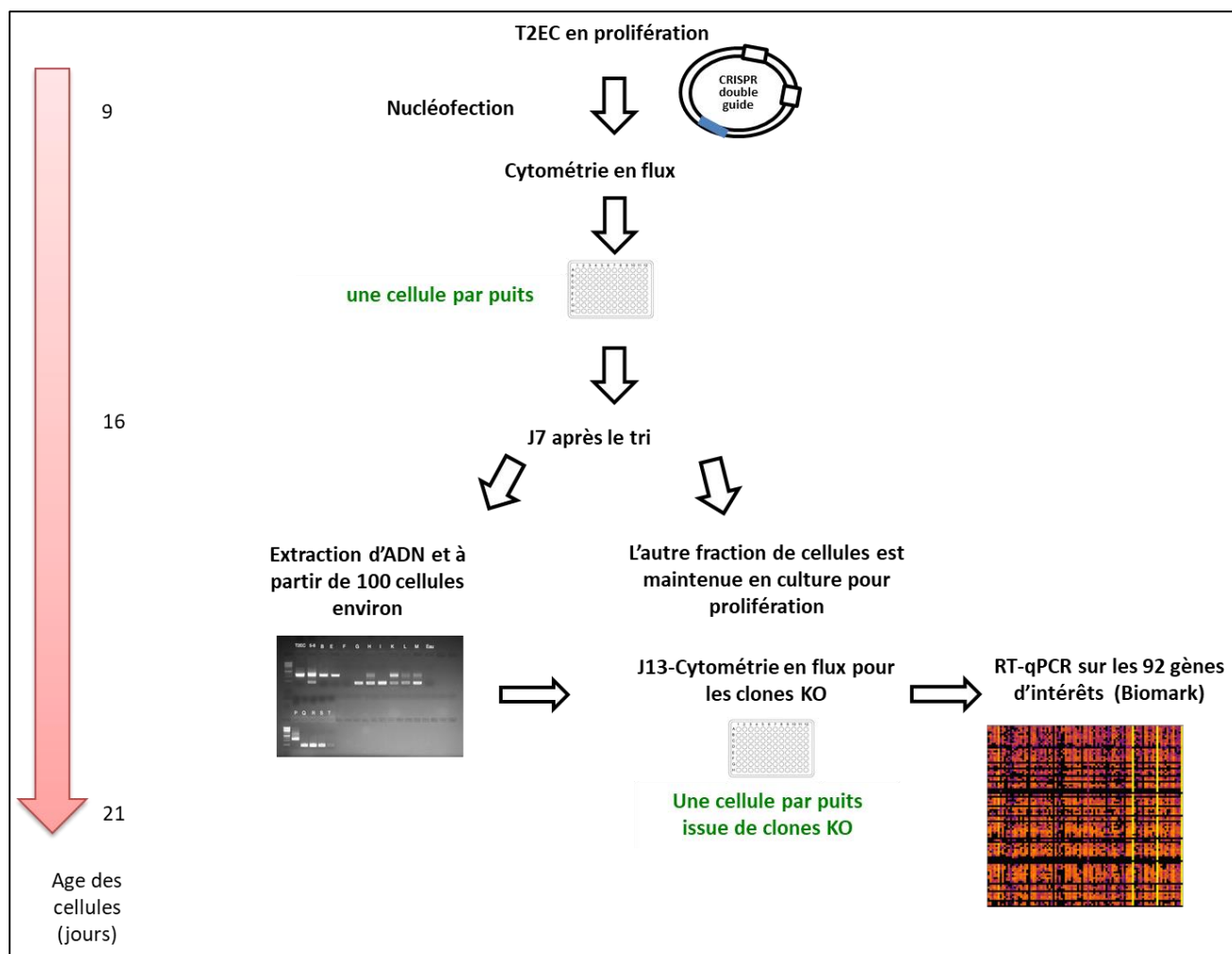


Figure 36-Premier protocole de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 double guides et d'obtention de clones de cellules KO pour l'analyse transcriptomique à l'échelle de la cellule unique

Dans ce protocole, les cellules sont transfectées avec un vecteur CRISPR-Cas9 double guide et isolées par cytométrie en flux dans des plaques 96 puits. 7 jours après le clonage, une partie des cellules (100 cellules / clone) est prélevée et les ADNg extraits pour analyse par PCR. Après identification des clones d'intérêt, la partie restante des cellules de ce clone est triée par FACS afin de déposer une cellule par puits pour réaliser ensuite l'analyse par RT-qPCR.

Par l'approche CRISPR-Cas 9 double guide, nous pouvons distinguer 3 types de clones au niveau génomique : clones délétés sur les deux allèles, clones délétés sur un seul allèle et des clones non délétés avec présence de mutations ponctuelles (sur un ou deux allèles) ou ne présentant aucune mutation (clone sauvage dans ce cas) (Figure 37).

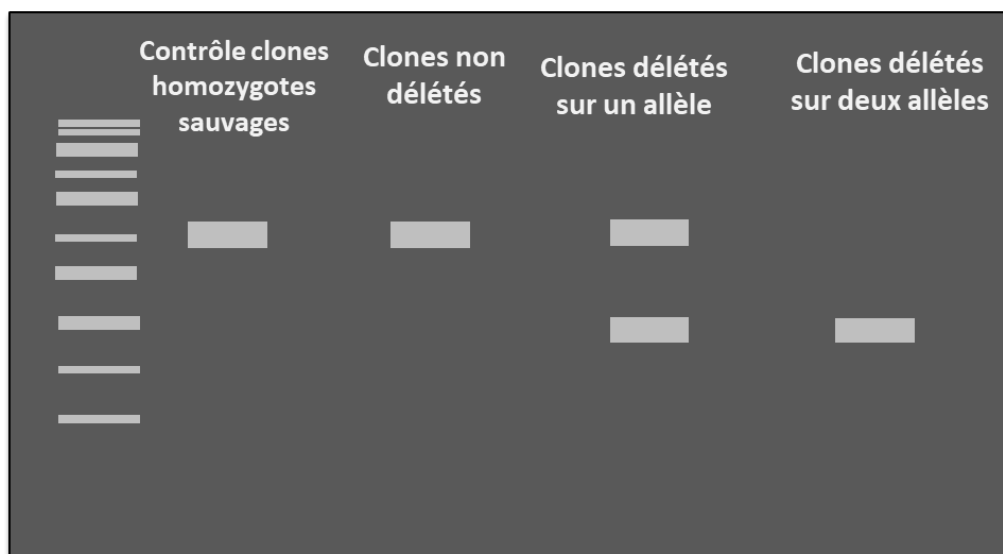


Figure 37-Profiles attendus après analyse par PCR de la région ciblée par les deux guides et dépôt sur gel des amplifications des clones transfectés avec les vecteurs doubles guides

Suite à la transfection des cellules avec les vecteurs double guides, plusieurs types de clones peuvent être générés : des clones délétés sur deux allèles, des clones délétés sur un seul allèle et des clones non délétés qui peuvent présenter ou non des mutations (sur un, deux ou aucun allèle).

Grâce à ce protocole, nous avons réussi à trouver deux clones délétés sur les deux brins au niveau génomique.

D'une manière non attendue, une première analyse par scRT-qPCR sur des cellules isolées à partir de ces clones a montré que l'ARNm de *FNIP1* était toujours présent d'où la possibilité d'avoir toujours une protéine FNIP1 fonctionnelle au moins en partie dans ces cellules. En absence d'anticorps anti-FNIP1 fonctionnels dans notre modèle aviaire, une étape de caractérisation des clones au niveau transcriptomique s'est donc avérée indispensable afin de s'assurer que les clones délétés pour *FNIP1* au niveau génomique présentaient un décalage de phase de lecture sur l'ARNm correspondant, et donc, une vraie perturbation de la fonction du gène *FNIP1*.

Malheureusement, pour chacun des profils génétiques obtenus suite à la transfection des cellules par les vecteurs CRISPR-cas9 double guides, de nombreuses combinaisons sont possibles au niveau transcriptomique. En effet, au-delà de l'action ou pas de CRISPR-Cas sur les deux

allèles, chaque délétion ou mutation peut ou pas induire un décalage du cadre ouvert de phase de lecture (Figure 38).

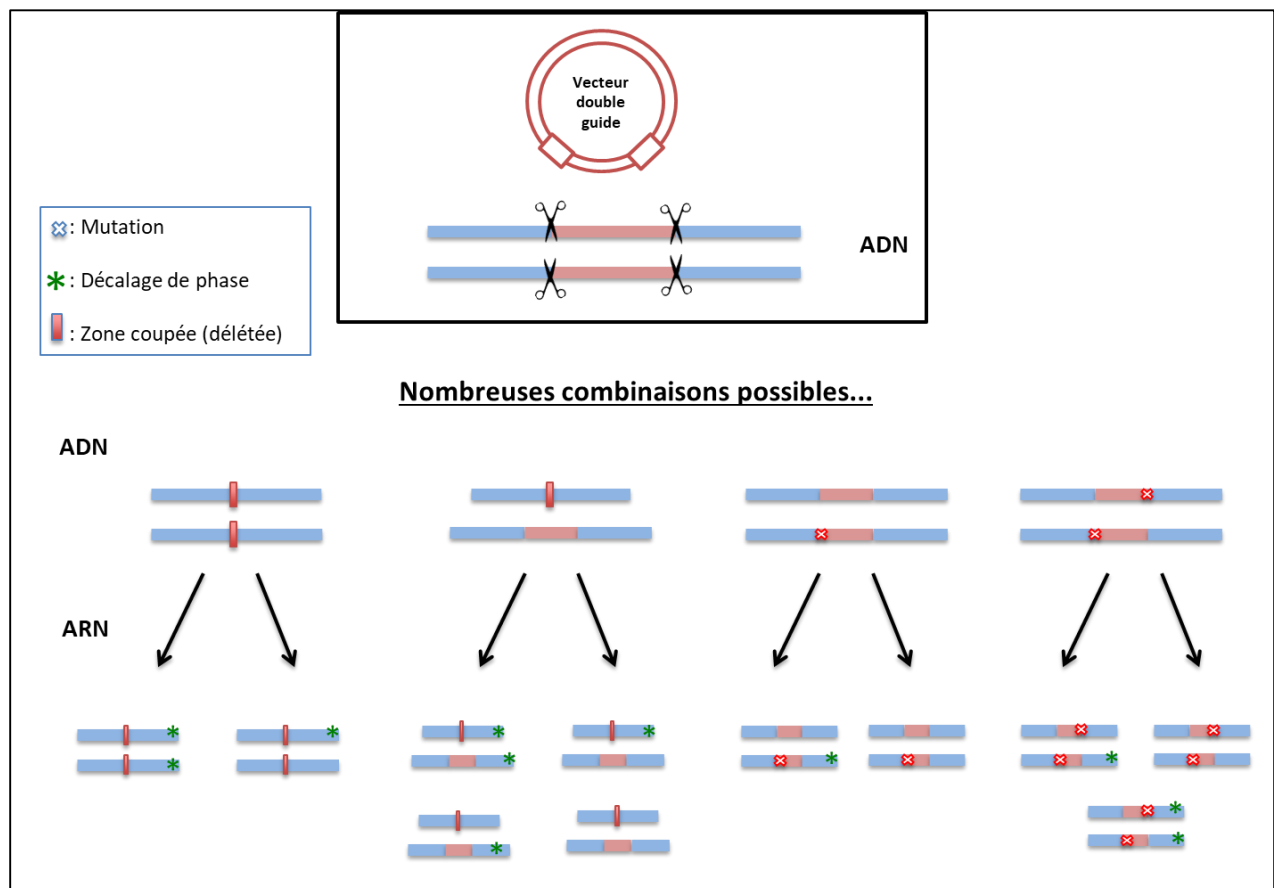


Figure 38-Différentes combinaisons possibles au niveau génomique et transcriptomique suite à la transfection des cellules avec les vecteurs CRISPR-Cas9 double guides.

En conséquence, notre nouvel objectif était de chercher spécifiquement deux types de clones : les clones mutés ou déletés sur les deux brins dont les modifications induisent un double décalage de phase de lecture (KO-100%) ou les clones avec une délétion ou mutation sur un seul brin mais induisant bien un décalage de phase de lecture (KO-50%). Dans ce but, nous avons modifié le premier protocole en ajoutant une étape de validation des clones à l'échelle transcriptomique parallèlement à l'étape de validation à l'échelle génomique. Le protocole final optimisé est présenté dans la Figure 39.

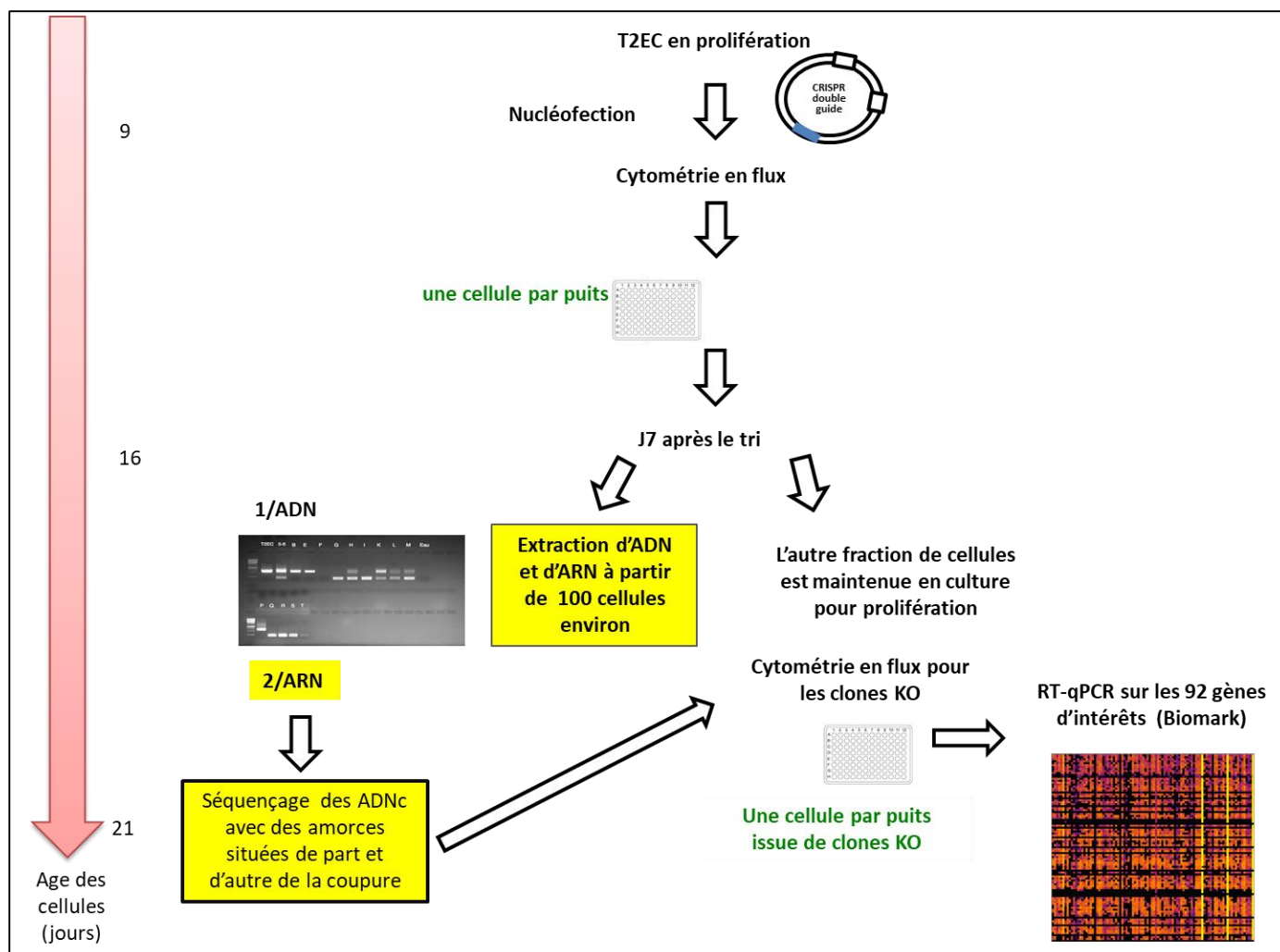


Figure 39-Protocole final de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 double guides et d'obtention de clones de cellules KO pour l'analyse transcriptomique à l'échelle de la cellule unique

Dans ce protocole, les cellules sont transfectées avec un vecteur CRISPR-Cas9 double guide, ensuite isolées par cytométrie en flux dans des plaques 96 puits. 7 jours après le clonage, une partie des cellules (100 cellules / clone) est prélevée et l'ADN et les ARN sont extraits puis analysés par PCR ou RT-PCR respectivement. Après identification des clones d'intérêt, la partie restante des cellules de ce clone est triée par FACS afin de déposer une cellule par puits pour l'analyse ultérieure du transcriptome par RT-qPCR.

F/ Difficultés rencontrées pour la caractérisation au niveau transcriptomique des clones KO pour FNIP1 après transfection par les vecteurs double guides

Alors que l'efficacité de l'édition du gène a pu être facilement vérifiée au niveau génomique, une caractérisation systématique de la présence d'un décalage du cadre de lecture sur l'ARNm a été rendue très difficile malheureusement en grande partie à cause de l'action de

deux guides. Par exemple, dans la majorité des cas obtenus, la lecture du séquençage est rendue impossible à cause des actions différentes des guides sur les deux allèles de *FNIP1* qui génèrent des séquences d'ARNm très différentes. En outre, pour certains clones touchés sur un seul allèle, le premier guide a induit seulement une petite délétion de deux paires de bases et le deuxième, une délétion d'une seule paire de base. Ainsi, le cadre ouvert de lecture de cet allèle a été reconstitué et une protéine FNIP1 de fonction inconnue a été possiblement produite.

Malgré le nombre très important de clones avec des profils génétiques différents testés (60 clones pour le vecteur 1 et 52 clones pour le vecteur 2), aucun clone 100% KO ni 50% KO n'a pu être identifié. En effet, les clones 100% KO semblaient ne pas survivre même si d'autres vérifications sont nécessaires pour affirmer que le KO de *FNIP1* est létal pour les clones. D'autre part, les KO 50% KO auraient dû survivre très probablement mais avec l'action de deux guides, la probabilité d'en trouver un était finalement très faible.

Par conséquent, afin d'augmenter la probabilité d'obtention de clones d'intérêt, nous avons décidé de changer de stratégie et d'utiliser des vecteurs CRISPR-Cas9 avec un seul guide.

Vecteurs CRISPR-Cas9 simple guides

A/ Validation de l'efficacité des vecteurs CRISPR-Cas9 simple guide au niveau génomique et transcriptomique

Plusieurs vecteurs CRISPR-Cas9 simple guide qui coupent dans différentes régions de *FNIP1* ont été construits. Afin de valider l'efficacité de ces vecteurs simple guide, nous avons déterminé le pourcentage de mutations induites sur population de cellules. Pour cela, pour chaque vecteur séparément, l'ADNg des cellules a été extrait après transfection et tri des cellules GFP+, puis la zone ciblée par le guide a été amplifiée par PCR. Les produits d'amplification ont été ligués dans des vecteurs pCR™4-TOPO® TA, utilisés ensuite pour transformer des bactéries. Plusieurs clones bactériens contenant donc un seul exemplaire des différentes séquences générées par l'action du guide, ont été analysés par PCR puis séquençage. Seul le vecteur avec la meilleure efficacité (vecteur 9) a été conservé pour poursuivre les analyses. Ce vecteur coupe

dans l'exon 5 du gène *FNIP1*. Au niveau génomique, le vecteur 9 présente une efficacité de mutations de 36%. La majorité des séquences mutées présentent des délétions de deux ou 3 paires de bases au niveau de la région *FNIP1* ciblée par le guide.

Au niveau transcriptomique, le vecteur 9 induit une diminution de 60% de l'expression de *FNIP1* dans les populations de cellules transfectées en comparaison aux cellules transfectées avec le vecteur vide (Figure 40).

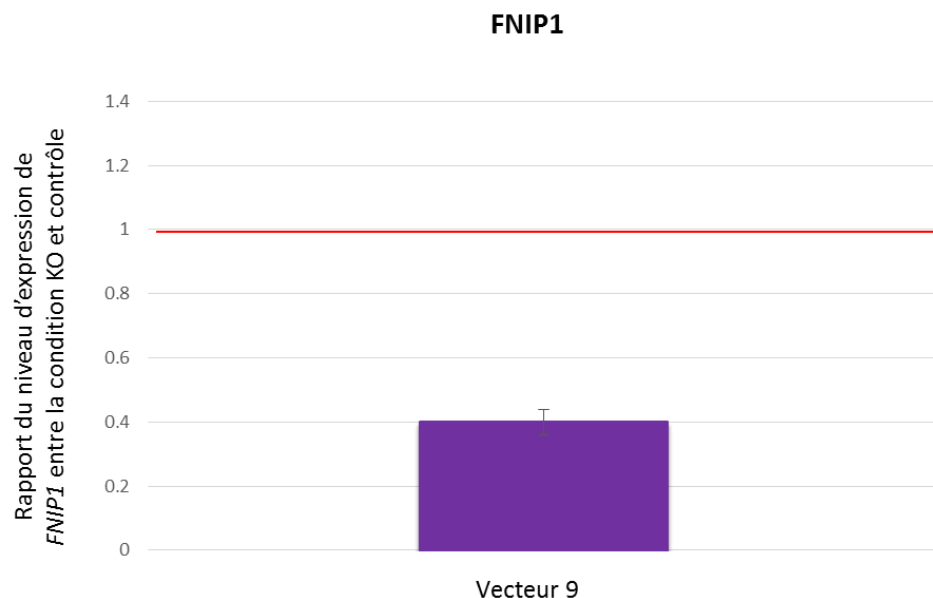


Figure 40-Rapport du niveau d'expression de FNIP 1 dans les cellules transfectées par le vecteur 9 par rapport aux cellules contrôles

Les T2EC ont été transfectées avec le vecteur 9 dirigés contre *FNIP1*. Les données présentées sont la moyenne des rapports +/- l'écart-type de l'expression de *FNIP 1* dans les T2EC transfectées par les vecteurs CRISPR-Cas9 par rapport à la condition contrôle (vecteur vide) à partir de trois expériences indépendantes. Les données ont été au préalable normalisées en fonction de l'expression des gènes standards *hnRNP* et *eif5*.

B/Optimisation du protocole pour l'obtention de clones KO pour FNIP1 par la stratégie CRISPR-Cas 9 simple guide et pour la caractérisation de ces clones avant l'analyse de leur transcriptome à l'échelle de la cellule unique

Puisqu'avec les vecteurs simple guide, seuls des clones avec des profils génomiques « non délétés » peuvent être produits, nous avons travaillé sur un autre protocole qui reproduit exactement les mêmes étapes du premier mais qui consiste à mieux caractériser les mutations générées au niveau génomique. Cette caractérisation nécessite une amplification par PCR au niveau de la région ciblée par le guide, le clonage des fragments amplifiés et leur séquençage (Figure 41). Ainsi, ce protocole permet de générer des clones de cellules primaires à partir de cellules transfectées avec les vecteurs CRISPR-Cas9 simple guide isolées par FACS, de vérifier l'invalidation du gène *FNIP1* à l'échelle génomique mais aussi au niveau transcriptomique en séquençant les ADN et les ADNc, respectivement, d'une petite population clonale (100 cellules). Une fois les clones validés, le reste des cellules du clone d'intérêt est trié par FACS pour permettre l'isolement des cellules et faire les analyses ultérieures de l'expression génique à l'échelle unicellulaire.

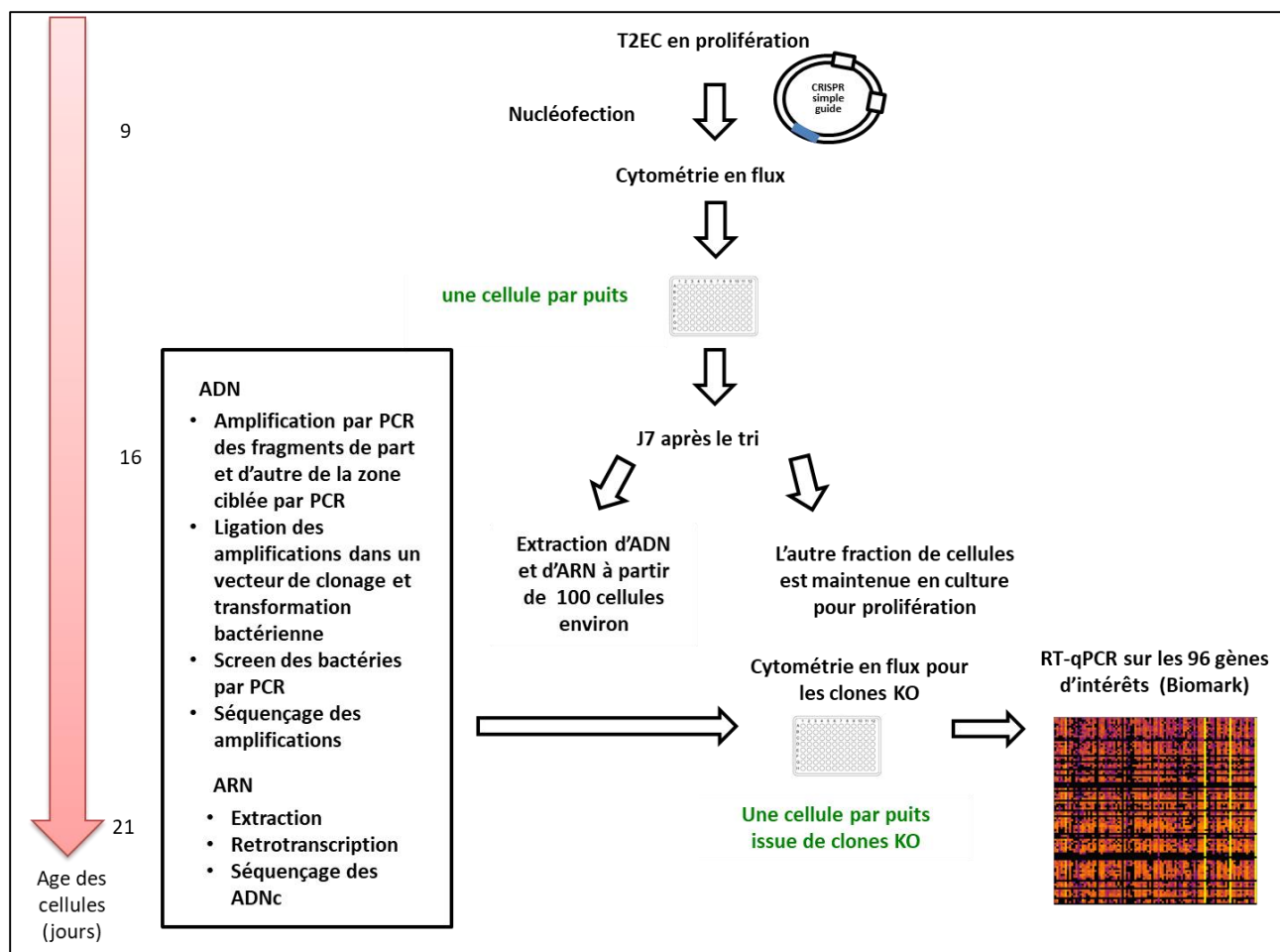


Figure 41- Protocole de caractérisation des cellules transfectées avec les vecteurs CRISPR-Cas9 simple guide et d'obtention de clones de cellules KO pour l'analyse transcriptomique à l'échelle de la cellule unique

Ce protocole consiste à transfecter les cellules avec un vecteur CRISPR-Cas 9 simple guide, puis les isoler par cytométrie en flux dans des plaques 96 puits. 7 jours après le clonage, une partie des cellules (100 cellules / clone) est isolée et l'ADN et les ARN sont extraits puis analysés par PCR et RT-PCR respectivement puis séquençage. Après identification des clones d'intérêt, la partie restante des cellules de ce clone est triée par FACS afin de déposer une cellule par puits pour l'analyse ultérieure en RT-qPCR.

Grâce à ce protocole, nous avons récemment pu obtenir deux clones présentant des mutations avec un décalage de phase sur les deux allèles du gène *FNIP1* (100% KO) et deux autres présentant une mutation avec un décalage de phase sur un seul des deux allèles *FNIP1* (50% KO). L'analyse de ces clones est actuellement en cours.

2.2.3 Conclusions

Nous avons étudié l'effet d'une perturbation de *FNIP1*, le positionnement de ce gène étant le plus incertain dans les réseaux de régulation de gènes (GRN) qui sous-tendent le processus de la différenciation des progéniteurs érythrocytaires. Dans ce cadre, nous avons construit des vecteurs CRISPR-Cas9 simple et double guides afin de produire un KO de *FNIP1*.

Dans un premier temps, nous avons validé les efficacités des vecteurs double guides et étudié les effets du KO de *FNIP1* en population cellulaire. A cette échelle, nos analyses montrent que l'altération de *FNIP1* affecte la prolifération des T2EC en auto-renouvellement et en différenciation suggérant l'importance de *FNIP1* pour la division des cellules et peut-être son rôle dans la progression du cycle cellulaire. Au contraire, l'altération de *FNIP1*, n'a pas d'effet notable ni sur la mortalité ni sur le taux de différenciation des T2EC.

Dans un deuxième temps, nous avons développé une approche expérimentale pour la génération et la validation de clones KO permettant ensuite une analyse de l'expression génique à l'échelle de la cellule unique. Cette approche prend en compte différentes contraintes inhérente notamment à notre modèle cellulaire (cellules primaires qui prolifèrent mieux sous forme de grappes et dont la durée de vie est courte).

A l'échelle de la cellule unique, un KO complet de *FNIP1* semble être très délétère pour les cellules puisqu'il nous a été extrêmement difficile d'en obtenir un. Toutefois, des analyses par RT-qPCR en microfluidique sur deux clones 100% KO obtenus très récemment avec la stratégie CRISPR simple guide sont actuellement en cours. Une fois les analyses terminées, nous pourrons utiliser ces nouvelles données pour l'inférence de GRN afin de réduire le nombre de GRN candidats sous-jacents au processus de différenciation érythrocytaire, initialement proposés (Bonnaïffoux et al. 2018).

3. Discussion et perspectives

La différenciation cellulaire est un processus selon lequel une cellule non spécialisée se divise pour donner naissance à une population de cellules spécialisées qui remplissent une fonction tissulaire bien désignée. Ce processus a été historiquement décrit comme un processus composés d'étapes bien définies ; en effet, les cellules souches s'engageraient dans la différenciation en passant par plusieurs étapes intermédiaires avant de donner naissance à des cellules différenciées, stables et incapables de revenir en arrière (Wolpert 1994).

Plus récemment, le développement des approches expérimentales et bioinformatiques applicables à l'échelle de la cellule unique a permis de révéler l'importance de l'hétérogénéité de l'expression des gènes dans une population de cellules isogéniques (Tanay and Regev 2017) (Choi and Kim 2019) et de décrire le rôle de cette hétérogénéité dans la prise de décision cellulaire notamment au cours de la différenciation (Sagar and Grün 2020). Grâce à l'acquisition de nouvelles données expérimentales en cellules uniques, plusieurs modèles de différenciation dynamique ont été décrits ébranlant ainsi la vision linéaire classiquement admise du processus de différenciation (Ridden et al. 2015)(Mojtahedi et al. 2016)(Richard et al. 2016). La caractérisation de la dynamique du système d'interactions de régulation génique qui sous-tend la prise de décision cellulaire est donc devenu un objectif majeur en recherche ces dernières années (Dunn et al. 2014) (Macneil and Walhout 2011).

Dans le contexte de la différenciation hématopoïétique, plusieurs études ont décrit la dynamique de ce processus (Moussy et al. 2017) (Pina et al. 2012). Dans ces études, la variabilité de l'expression génétique est considérée comme l'une des causes sous-jacentes de l'engagement des cellules hématopoïétiques.

En outre, une étude récente menée dans l'équipe, basée sur l'analyse de l'expression génique par scRT-qPCR et sur des outils de modélisation, a décrit l'influence de la stochasticité sur la détermination du destin des érythrocytes (Guillemin et al. 2019). Cependant, on sait encore peu de choses sur la façon dont les progéniteurs érythrocytaires choisissent et s'engagent dans l'érythropoïèse et sur le réseau moléculaire exact qui sous-tend cet engagement.

Ainsi, au cours de ma thèse, je me suis intéressée à la caractérisation moléculaire du processus dynamique de la différenciation et particulièrement celui des progéniteurs érythrocytaires aviaires appelés T2EC (O Gandrillon et al. 1999). Dans ce but, mes recherches se sont divisées en deux grandes parties.

La première partie a découlé d'un travail antérieur mené dans l'équipe et portant sur l'analyse de l'expression génique par scRT-qPCR dans des T2EC au cours d'une cinétique de différenciation (Richard et al. 2016). Ce travail a mis en évidence une augmentation transitoire de la variabilité intercellulaire entre 8h et 24h de différenciation ; pendant ces 24 premières heures de différenciation, les cellules gardent la capacité de s'auto-renouveler ce qui n'est plus le cas au-delà. Nous avons donc observé un engagement irréversible des T2EC dans la différenciation érythrocytaire lorsque la variabilité intercellulaire diminue.

Dans ce cadre, mon travail a été de caractériser moléculairement les T2EC qui ont gardé la capacité de s'auto-renouveler malgré l'engagement amorcé dans le processus de différenciation ; nous souhaitons plus spécifiquement préciser si les cellules avaient totalement révertées notamment d'un point de vue moléculaire. Nos résultats indiquent que les cellules en réversion ont déjà commencé à répondre au signal d'induction de la différenciation pendant les 24 premières heures en activant l'expression de gènes liés à la voie de synthèse de l'hème. Toutefois, le changement de signaux externes lorsque les cellules ont été replacées en milieu d'auto-renouvellement a permis aux T2EC de retrouver une signature moléculaire proche des cellules d'origines en état d'auto-renouvellement et donc de revenir vers un état indifférencié.

Dans l'ensemble, ces résultats suggèrent une forme de plasticité dynamique et une réversibilité physiologique lorsque les cellules n'ont pas encore atteint un état stable dans le processus de différenciation, état stable qui ne leur permet plus ensuite de revenir en arrière.

Ces données appuient les nouvelles théories de réversibilité du destin (Jopling, Boue, and Belmonte 2011). Dans le contexte du paysage épigénétique, notre travail montre que les cellules en différenciation, ne progressent pas *via* un processus continu comme celui proposé par Waddington (Waddington 1957) ; au contraire, elles évoluent dans un paysage peuplé de puits et oscillent sur leur position à cause des variations aléatoires des niveaux d'expression génique

et de l'influence de l'environnement, en se déplaçant de puits en puits, c'est-à-dire d'un état métastable à un autre (Guillemin and Stumpf 2021) (Ventre et al. 2020).

En accord avec nos résultats, de nombreux travaux ont mis en évidence que lors de la différenciation cellulaire, un certain degré de plasticité phénotypique peut être observé et que des cellules qui atteignent un degré intermédiaire de différenciation peuvent se dé-différencier pour retrouver la signature moléculaire des cellules d'origines. Par exemple, l'analyse par scRNAseq a permis aux chercheurs d'identifier dans les CNCC murins, une population de cellules exprimant transitoirement le facteur de pluripotence Oct4 dont l'augmentation confère à ces cellules le potentiel de réverser pour adopter une nouvelle signature mésodermique capable de former les dérivés typiques du mésoderme, comme le mésenchyme (Zalc et al. 2021). D'autres travaux, menés à l'échelle de la population cellulaire, ont également confirmé la plasticité de destin cellulaire (Johnson et al. 2008) (Mani et al. 2008) (Tata et al. 2013) (Li et al. 2016). En outre, une étude plus récente a mis en évidence que les cellules hématopoïétiques peuvent être induites à se différencier en cellules souches multipotentes sous l'action d'un facteur libéré par les fibroblastes, le M-CSF. En effet, ce facteur induit l'expression des marqueurs de cellules souches multipotentes tels que SSEA-1, SSEA-3, OCT3/4, NANOG, SOX2 (Li et al. 2016). En accord avec nos résultats, ces études menées à l'échelle de la population ont bien décrit la plasticité de destin des cellules. Cependant, ces études ne prennent pas en considération le caractère stochastique de l'expression génétique dans la plasticité du destin cellulaire puisqu'elles considèrent les cellules comme un matériel homogène.

Lors de la prise de décision cellulaire, l'état des cellules est contraint et maintenu par l'ensemble des interactions dynamiques entre les gènes, appelée aussi réseaux de régulation de gènes (Peter and Davidson 2015). En effet, les GRN servent à décrire la façon dont les modèles d'expression génique changent au fil du temps et, permettent ainsi une meilleure compréhension de l'aspect dynamique crucial de la prise de décision cellulaire. Identifier le réseau de régulation de gènes qui influence la dynamique du processus de différenciation et de réversion des T2EC apparaît donc important nécessitant des études plus approfondies et des modélisations mathématiques appropriées.

Le second objectif de mon travail de thèse a donc été de mieux caractériser le GRN sous-jacent au processus de différenciation. L'application de l'outil WASABI aux données de scRT-qPCR acquises précédemment sur une cinétique de différenciation des T2EC a permis d'inférer 364 réseaux de régulation de gènes (GRN) candidats qui sous-tendent ce processus. Afin de discriminer ces différents GRN, nous avons étudié, en utilisant l'approche de CRISPR-Cas9, l'effet du KO de *FNIP1* sur les T2EC, *FNIP1* étant le gène dont le positionnement est le plus incertain dans les réseaux.

A l'échelle de la population cellulaire, un ralentissement de la prolifération cellulaire induit par le KO de *FNIP1* sans que la mortalité ne soit affectée a été observé. Ces résultats suggèrent l'importance de *FNIP1* pour la division des cellules et peut-être un rôle de celui-ci dans la progression du cycle cellulaire, rôle qui reste néanmoins à être expérimentalement démontré. Le rôle connu de *FNIP1* dans la régulation de mTORC 1 pourrait expliquer ces observations (Tsun et al. 2013). En effet, mTORC1 joue un rôle très important dans la régulation de la croissance et de la progression du cycle cellulaire en réponse aux facteurs de croissance, au statut énergétique et au stress (Hsieh et al. 2018). Il ne serait donc pas surprenant que la déficience de *FNIP1*, un des principaux régulateurs de mTORC1 puisse altérer la prolifération et la progression normale des cellules dans les phases du cycle cellulaire.

Par ailleurs, il a été décrit que la perte de *FLCN*, qui fait partie du complexe FNIP /*FLCN*, induit une augmentation du nombre de cellules en phase G1 et une diminution du nombre de cellules en phase S-M dans le cerveau embryonnaire de souris (Kenyon et al. 2016). Parallèlement, des défauts dans la division cellulaire ont été observés chez les drosophiles déficientes en *FLCN* (J. Liu et al. 2016).

Malgré son effet sur la prolifération des T2EC, notre travail a montré que le KO de *FNIP1* n'a aucune influence sur la différenciation de ces cellules. En ce sens, notre modèle cellulaire présente des différences avec ce qui a pu être décrit dans d'autres modèles cellulaires comme les cellules musculaires dans lesquelles le rôle de *FNIP1* dans leur différenciation a été démontré (Reyes et al. 2015)(J. Liu et al. 2016). Il a été aussi décrit que *FLCN*, ainsi que ses partenaires d'interaction *FNIP1* et *FNIP2*, régulent la sortie des cellules souches de la pluripotence

(Betschinger et al. 2013). Enfin, des souris double KO pour *FNIP1* présentent dans la rate un blocage des cellules B entre le stade pro-B et pré-B alors que le nombre de thymocytes et de lymphocytes T périphériques apparaît normal (H. Park et al. 2012).

L'ensemble de ces données supportent donc le rôle de *FNIP1* dans l'auto-renouvellement des T2EC. Cependant, d'autres expériences, comme des expériences d'analyse du cycle cellulaire, doivent être réalisées pour mieux caractériser l'effet du KO sur ces cellules. Parallèlement, des expériences de surexpression de *FNIP1* pour étudier son effet sur la prolifération et différenciation des T2EC sont en cours de laboratoire.

A l'échelle de la cellule unique, la recherche d'un clone KO de *FNIP1* s'est avérée très complexe. En effet, l'obtention de clones KO à partir de cellules primaires est rendue techniquement difficile par de multiples facteurs, notamment la durée de vie limitée des cellules limitant aussi la taille des clones obtenus (mortalité et vieillissement cellulaire accentués aussi par l'étape de transfection), la prolifération réduite des cellules lorsqu'elles sont isolées alors qu'en conditions normales la formation de grappes facilitent leur prolifération et survie. De plus, en accord avec d'autres études (Anderson et al. 2017), nous avons établi que même si l'efficacité du KO peut être vérifiée à l'échelle génomique, il est absolument essentiel de procéder à sa caractérisation à l'échelle transcriptomique pour s'assurer de l'induction d'un décalage du cadre de lecture dans l'ARNm d'autant plus si celui-ci n'est pas dégradé. En effet, une étude assez récente a révélé que la perturbation de l'épissage est une conséquence fréquente des mutations générées par la technologie d'édition de gènes CRISPR-Cas9 pouvant conduire à la production de protéines aberrantes dans les cellules mais finalement dont la fonction si elle existe, reste complètement inconnue (Sharpe and Cooper 2017). Pour citer un exemple, la délétion induite par CRISPR-Cas9 du gène *KRAS* dans une lignée cellulaire d'adénocarcinomes pulmonaires, a bien provoqué la coupure prévue avec la formation d'un codon de terminaison précoce sur l'ARNm. Cependant, la caractérisation des lignées par RNA-seq a révélé une délétion de tout l'exon contenant la zone ciblée et en conséquence la conservation du cadre de lecture du gène *KRAS*. Ces lignées présentent encore des taux détectables de transcrits *KRAS* qui sont traduits en protéines *KRAS* aberrantes dont on ignore le rôle (Popp and Maquat 2016). Dans une autre étude

portant sur le gène codant la β -caténine, le processus CRISPR-Cas9 ciblant l'exon 3, a induit une disparition de cet exon permettant néanmoins la production d'une protéine β -caténine stabilisée et active de manière constitutive, qui est retenue dans le noyau à la place du cytoplasme. Ces résultats montrent clairement que des délétions génomiques peuvent conduire à une disparition d'exon pouvant entraîner l'expression d'une protéine tronquée avec une localisation différente, et vraisemblablement une fonction différente de la protéine normale (Mou et al. 2017). Ainsi, une caractérisation génomique et transcriptomique des clones constitue une étape indispensable dans toute stratégie CRISPR-Cas9.

A ce jour, il a été possible d'obtenir deux clones présentant un KO sur les deux allèles du gène *FNIP1* et deux autres présentant un KO uniquement sur un des deux allèles au niveau génomique. Des expériences de caractérisation de ces clones au niveau transcriptomique sont en cours au laboratoire. Cependant, à l'échelle de la cellule unique, un KO complet de *FNIP1* semble être délétère pour les cellules car nous avons dû analyser un nombre très important de clones avant de pouvoir obtenir ces 2 clones *FNIP1*^{-/-} avec une viabilité suffisante pour poursuivre les analyses. D'autres vérifications sont bien sûr nécessaires pour affirmer que le KO de *FNIP1* pourrait être létal à plus ou moins long terme pour les cellules d'autant plus que nous n'avons pas observé d'augmentation de la mortalité dans les expériences de KO menées à l'échelle de la population cellulaire. Cependant, il a été précédemment montré que les cellules pré-B issues de souris *FNIP1*^{-/-} présentaient une activité caspase accrue par rapport aux cellules normales avec une augmentation du nombre de cellules mortes (Baba et al. 2012). Ces études suggèrent donc que *FNIP1* est nécessaire à la survie de certains types cellulaires. De plus, l'effet d'une altération de *FNIP1* sur la létalité des cellules pré-B a été aussi confirmé par l'introduction d'un transgène anti-apoptotique Bcl2 dans les cellules *FNIP1*^{-/-}. Dans ces études, l'expression de Bcl2 a permis de contourner le blocage des pré-B et de permettre la survie de la population de cellules B (Baba et al., 2012) (Park et al. 2012).

Lorsque les clones de T2EC KO pour *FNIP1* seront complètement caractérisés notamment au niveau transcriptomique, nous pourrons étudier l'expression des gènes d'intérêt identifiés dans Richard et al par la technologie de scRT-qPCR. Nous comparerons les effets du KO de *FNIP1*

sur l'expression de ces gènes en comparaison à des cellules contrôles. Les nouvelles matrices d'expression générées pourront être analysées par WASABI (Bonnaffoux et al. 2018) afin d'inférer les nouveaux GRN et réduire le nombre de GRN candidats sous-jacents au processus de différenciation érythrocytaire.

Afin de réduire encore plus le nombre de GRN candidats, d'autres expériences de perturbations devront probablement être réalisées, comme la surexpression de *FNIP1* ou des expériences de KO du second gène incertain dans les réseaux, le gène *MID2*. En effet, il est bien connu que le GRN décrivant le plus justement un processus doit changer de comportement pour s'adapter à toute perturbation subie par les cellules. Ainsi, d'autres données d'analyse en cellule unique, comme les données acquises à partir des cellules en réversion, qui ont subi des perturbations importantes suite à un changement des signaux extérieurs, pourront être aussi utilisées pour la réduction du nombre de GRN candidats dans notre modèle cellulaire. Ainsi, les données acquises sur les T2EC KO pour le gène *FNIP1* et celles acquises dans mes expériences de réversion pourront être utilisées parallèlement afin de chercher et de sélectionner le meilleur GRN qui décrira au mieux le processus de différenciation des T2EC.

En conclusion, l'ensemble des résultats de cette thèse nous permettent une meilleure compréhension de l'engagement des cellules érythrocytaires dans la différenciation érythrocytaire influencé par les contraintes de la dynamique du système d'interactions de régulation génique sous-jacent.

Références

- Adamo, Luigi, Olaia Naveiras, Pamela L. Wenzel, Shannon McKinney-Freeman, Peter J. Mack, Jorge Gracia-Sancho, Astrid Suchy-Dicey, et al. 2009. "Biomechanical Forces Promote Embryonic Haematopoiesis." *Nature* 459 (7250): 1131–35. <https://doi.org/10.1038/nature08073>.
- Akashi, K., D. Traver, T. Miyamoto, and I. L. Weissman. 2000. "A Clonogenic Common Myeloid Progenitor That Gives Rise to All Myeloid Lineages." *Nature* 404 (6774): 193–97. <https://doi.org/10.1038/35004599>.
- Akutsu, T., S. Miyano, and S. Kuhara. 1999. "Identification of Genetic Networks from a Small Number of Gene Expression Patterns under the Boolean Network Model." *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 17–28. https://doi.org/10.1142/9789814447300_0003.
- Anderson, Jennifer L., Timothy S. Mulligan, Meng-Chieh Shen, Hui Wang, Catherine M. Scahill, Frederick J. Tan, Shao J. Du, Elisabeth M. Busch-Nentwich, and Steven A. Farber. 2017. "MRNA Processing in Mutant Zebrafish Lines Generated by Chemical and CRISPR-Mediated Mutagenesis Produces Unexpected Transcripts That Escape Nonsense-Mediated Decay." *PLoS Genetics* 13 (11): e1007105. <https://doi.org/10.1371/journal.pgen.1007105>.
- Antoniadis, A., S. Lambert-Lacroix, and F. Leblanc. 2003. "Effective Dimension Reduction Methods for Tumor Classification Using Gene Expression Data." *Bioinformatics* 19 (5): 563–70. <https://doi.org/10.1093/bioinformatics/btg062>.
- Arezi, Bahram, and Holly Hogrefe. 2009. "Novel Mutations in Moloney Murine Leukemia Virus Reverse Transcriptase Increase Thermostability through Tighter Binding to Template-Primer." *Nucleic Acids Research* 37 (2): 473–81. <https://doi.org/10.1093/nar/gkn952>.
- Baba, Masaya, Seung-Beom Hong, Nirmala Sharma, Michelle B. Warren, Michael L. Nickerson, Akihiro Iwamatsu, Dominic Esposito, et al. 2006. "Folliculin Encoded by the BHD Gene Interacts with a Binding Protein, FNIP1, and AMPK, and Is Involved in AMPK and MTOR Signaling." *Proceedings of the National Academy of Sciences of the United States of America* 103 (42): 15552–57. <https://doi.org/10.1073/pnas.0603781103>.

- Baba, Masaya, Jonathan R. Keller, Hong-Wei Sun, Wolfgang Resch, Stefan Kuchen, Hyung Chan Suh, Hisashi Hasumi, et al. 2012. "The Folliculin-FNIP1 Pathway Deleted in Human Birt-Hogg-Dubé Syndrome Is Required for Murine B-Cell Development." *Blood* 120 (6): 1254–61. <https://doi.org/10.1182/blood-2012-02-410407>.
- Bacher, Rhonda, and Christina Kendzierski. 2016. "Design and Computational Analysis of Single-Cell RNA-Sequencing Experiments." *Genome Biology* 17 (1): 63. <https://doi.org/10.1186/s13059-016-0927-y>.
- Baedke, Jan. 2013. "The Epigenetic Landscape in the Course of Time: Conrad Hal Waddington's Methodological Impact on the Life Sciences." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 756–73. <https://doi.org/10.1016/j.shpsc.2013.06.001>.
- Balázsi, Gábor, Alexander van Oudenaarden, and James J. Collins. 2011. "Cellular Decision Making and Biological Noise: From Microbes to Mammals." *Cell* 144 (6): 910–25. <https://doi.org/10.1016/j.cell.2011.01.030>.
- Bansal, Mukesh, Vincenzo Belcastro, Alberto Ambesi-Impiombato, and Diego di Bernardo. 2007. "How to Infer Gene Networks from Expression Profiles." *Molecular Systems Biology* 3 (February): 78. <https://doi.org/10.1038/msb4100120>.
- Barabási, Albert-László, Natali Gulbahce, and Joseph Loscalzo. 2011. "Network Medicine: A Network-Based Approach to Human Disease." *Nature Reviews Genetics* 12 (1): 56–68. <https://doi.org/10.1038/nrg2918>.
- Basso, Katia, Adam A. Margolin, Gustavo Stolovitzky, Ulf Klein, Riccardo Dalla-Favera, and Andrea Califano. 2005. "Reverse Engineering of Regulatory Networks in Human B Cells." *Nature Genetics* 37 (4): 382–90. <https://doi.org/10.1038/ng1532>.
- Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Gehroux, and Evan W. Newell. 2018. "Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP." *Nature Biotechnology*, December. <https://doi.org/10.1038/nbt.4314>.

- Behrends, Christian, Mathew E. Sowa, Steven P. Gygi, and J. Wade Harper. 2010. "Network Organization of the Human Autophagy System." *Nature* 466 (7302): 68–76. <https://doi.org/10.1038/nature09204>.
- Ben-Sahra, Issam, and Brendan Manning. 2017. "MTORC1 Signaling and the Metabolic Control of Cell Growth." *Current Opinion in Cell Biology* 45 (April). <https://doi.org/10.1016/j.ceb.2017.02.012>.
- Betschinger, Joerg, Jennifer Nichols, Sabine Dietmann, Philip D. Corrin, Patrick J. Paddison, and Austin Smith. 2013. "Exit from Pluripotency Is Gated by Intracellular Redistribution of the BHLH Transcription Factor Tfe3." *Cell* 153 (2): 335–47. <https://doi.org/10.1016/j.cell.2013.03.012>.
- Blake, William J., Mads KAErn, Charles R. Cantor, and J. J. Collins. 2003. "Noise in Eukaryotic Gene Expression." *Nature* 422 (6932): 633–37. <https://doi.org/10.1038/nature01546>.
- Bonnaïffoux, Arnaud, Ulysse Herbach, Angelique Richard, Anissa Guillemin, Sandrine Giraud, Pierre-Alexis Gros, and Olivier Gandrillon. 2018. "WASABI: A Dynamic Iterative Framework for Gene Regulatory Network Inference." *BioRxiv*, April, 292128. <https://doi.org/10.1101/292128>.
- Brunetti, Lorenzo, Michael C. Gundry, Ayumi Kitano, Daisuke Nakada, and Margaret A. Goodell. 2018. "Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9." *Journal of Visualized Experiments: JoVE*, no. 134 (April). <https://doi.org/10.3791/57278>.
- Centini, Ryan, Mark Tsang, Terri Iwata, Heon Park, Jeffrey Delrow, Daciana Margineantu, Brandon M. Iritani, et al. 2018. "Loss of Fnip1 Alters Kidney Developmental Transcriptional Program and Synergizes with TSC1 Loss to Promote MTORC1 Activation and Renal Cyst Formation." *PLOS ONE* 13 (6): e0197973. <https://doi.org/10.1371/journal.pone.0197973>.
- Chai, Lian En, Swee Kuan Loh, Swee Thing Low, Mohd Saberi Mohamad, Safaai Deris, and Zalmiyah Zakaria. 2014. "A Review on the Computational Approaches for Gene Regulatory Network Construction." *Computers in Biology and Medicine* 48 (May): 55–65. <https://doi.org/10.1016/j.combiomed.2014.02.011>.

- Chambers, Ian, Jose Silva, Douglas Colby, Jennifer Nichols, Bianca Nijmeijer, Morag Robertson, Jan Vrana, Ken Jones, Lars Grotewold, and Austin Smith. 2007. "Nanog Safeguards Pluripotency and Mediates Germline Development." *Nature* 450 (7173): 1230–34. <https://doi.org/10.1038/nature06403>.
- Chan, Sunny Sun-Kin, and Michael Kyba. 2013. "What Is a Master Regulator?" *Journal of Stem Cell Research & Therapy* 3 (May): 114. <https://doi.org/10.4172/2157-7633.1000e114>.
- Chen, Geng, Baitang Ning, and Tielu Shi. 2019. "Single-Cell RNA-Seq Technologies and Related Computational Data Analysis." *Frontiers in Genetics* 10: 317. <https://doi.org/10.3389/fgene.2019.00317>.
- Chen, Kenian, Xiaomin Dong, and Jiaqian Wu. 2015. "The Application of Single-Cell Sequencing in Dynamic Transcriptomes." In *Single Cell Sequencing and Systems Immunology*, edited by Xiangdong Wang, 41–63. Translational Bioinformatics. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9753-5_4.
- Chen, Liang, and Sika Zheng. 2018. "BCseq: Accurate Single Cell RNA-Seq Quantification with Bias Correction." *Nucleic Acids Research* 46 (14): e82. <https://doi.org/10.1093/nar/gky308>.
- Chen, Yanqing, Jun Zhu, Pek Yee Lum, Xia Yang, Shirly Pinto, Douglas J. MacNeil, Chunsheng Zhang, et al. 2008. "Variations in DNA Elucidate Molecular Networks That Cause Disease." *Nature* 452 (7186): 429–35. <https://doi.org/10.1038/nature06757>.
- Cheung, W. L., S. D. Briggs, and C. D. Allis. 2000. "Acetylation and Chromosomal Functions." *Current Opinion in Cell Biology* 12 (3): 326–33. [https://doi.org/10.1016/s0955-0674\(00\)00096-x](https://doi.org/10.1016/s0955-0674(00)00096-x).
- Choi, Yoon Ha, and Jong Kyoung Kim. 2019. "Dissecting Cellular Heterogeneity Using Single-Cell RNA Sequencing." *Molecules and Cells* 42 (3): 189–99. <https://doi.org/10.14348/molcells.2019.2446>.
- Chu, H. W., C. Rios, C. Huang, A. Wesolowska-Andersen, E. G. Burchard, B. P. O'Connor, T. E. Fingerlin, D. Nichols, S. D. Reynolds, and M. A. Seibold. 2015. "CRISPR-Cas9-Mediated Gene Knockout in Primary Human Airway Epithelial Cells Reveals a Proinflammatory Role for MUC18." *Gene Therapy* 22 (10): 822–29. <https://doi.org/10.1038/gt.2015.53>.

- Chu, Yongjun, and David R. Corey. 2012. "RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation." *Nucleic Acid Therapeutics* 22 (4): 271–74. <https://doi.org/10.1089/nat.2012.0367>.
- Cohn, Melvin, and Kengo Horibata. 1959. "ANALYSIS OF THE DIFFERENTIATION AND OF THE HETEROGENEITY WITHIN A POPULATION OF ESCHERICHIA COLI UNDERGOING INDUCED β -GALACTOSIDASE SYNTHESIS1." *Journal of Bacteriology* 78 (5): 613–23.
- Coulon, Antoine, Olivier Gandrillon, and Guillaume Beslon. 2010. "On the Spontaneous Stochastic Dynamics of a Single Gene: Complexity of the Molecular Interplay at the Promoter." *BMC Systems Biology* 4 (January): 2. <https://doi.org/10.1186/1752-0509-4-2>.
- Cumano, Ana, and Isabelle Godin. 2007. "Ontogeny of the Hematopoietic System." *Annual Review of Immunology* 25: 745–85. <https://doi.org/10.1146/annurev.immunol.25.022106.141538>.
- Dai, Jian J., Linh Lieu, and David Rocke. 2006. "Dimension Reduction for Classification with Gene Expression Microarray Data." *Statistical Applications in Genetics and Molecular Biology* 5: Article6. <https://doi.org/10.2202/1544-6115.1147>.
- De Smet, Riet, and Kathleen Marchal. 2010. "Advantages and Limitations of Current Network Inference Methods." *Nature Reviews Microbiology* 8 (10): 717–29. <https://doi.org/10.1038/nrmicro2419>.
- Dojer, Norbert, Anna Gambin, Andrzej Mizera, Bartek Wilczyński, and Jerzy Tiuryn. 2006. "Applying Dynamic Bayesian Networks to Perturbed Gene Expression Data." *BMC Bioinformatics* 7 (1): 249. <https://doi.org/10.1186/1471-2105-7-249>.
- Dolznic, H., P. Bartunek, K. Nasmyth, E. W. Müllner, and H. Beug. 1995. "Terminal Differentiation of Normal Chicken Erythroid Progenitors: Shortening of G1 Correlates with Loss of D-Cyclin/Cdk4 Expression and Altered Cell Size Control." *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research* 6 (11): 1341–52.
- Dong, Hangming, Mengchen Zou, Ayesha Bhatia, Priyamvada Jayaprakash, Florence Hofman, Qilong Ying, Mei Chen, David T. Woodley, and Wei Li. 2016. "Breast Cancer MDA-MB-231 Cells Use Secreted Heat Shock Protein-90alpha (Hsp90 α) to Survive a Hostile Hypoxic

- Environment.” *Scientific Reports* 6 (February): 20605.
<https://doi.org/10.1038/srep20605>.
- Dunn, S-J, G Martello, B Yordanov, S Emmott, and AG Smith. 2014. “Defining an Essential Transcription Factor Program for Naïve Pluripotency.” *Science (New York, N.Y.)* 344 (6188): 1156–60. <https://doi.org/10.1126/science.1248882>.
- Durif, Ghislain, Laurent Modolo, Jakob Michaelsson, Jeff E. Mold, Sophie Lambert-Lacroix, and Franck Picard. 2018. “High Dimensional Classification with Combined Adaptive Sparse PLS and Logistic Regression.” *Bioinformatics (Oxford, England)* 34 (3): 485–93.
<https://doi.org/10.1093/bioinformatics/btx571>.
- Eling, Nils, Michael D. Morgan, and John C. Marioni. 2019. “Challenges in Measuring and Understanding Biological Noise.” *Nature Reviews Genetics* 20 (9): 536–48.
<https://doi.org/10.1038/s41576-019-0130-6>.
- Elowitz, Michael B., Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. 2002. “Stochastic Gene Expression in a Single Cell.” *Science (New York, N.Y.)* 297 (5584): 1183–86.
<https://doi.org/10.1126/science.1070919>.
- Faith, Jeremiah J., Boris Hayete, Joshua T. Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James J. Collins, and Timothy S. Gardner. 2007. “Large-Scale Mapping and Validation of Escherichia Coli Transcriptional Regulation from a Compendium of Expression Profiles.” *PLoS Biology* 5 (1): e8.
<https://doi.org/10.1371/journal.pbio.0050008>.
- Fan, H. Christina, Glenn K. Fu, and Stephen P. A. Fodor. 2015. “Combinatorial Labeling of Single Cells for Gene Expression Cytometry.” *Science* 347 (6222).
<https://doi.org/10.1126/science.1258367>.
- Ferrazzi, Fulvia, Paolo Magni, Lucia Sacchi, Angelo Nuzzo, Uros Petrovic, and Riccardo Bellazzi. 2007. “Inferring Gene Regulatory Networks by Integrating Static and Dynamic Data.” *International Journal of Medical Informatics* 76 Suppl 3 (December): S462-475.
<https://doi.org/10.1016/j.ijmedinf.2007.07.005>.

- Ferreira, Rita, Kinuko Ohneda, Masayuki Yamamoto, and Sjaak Philipsen. 2005. "GATA1 Function, a Paradigm for Transcription Factors in Hematopoiesis." *Molecular and Cellular Biology* 25 (4): 1215–27. <https://doi.org/10.1128/MCB.25.4.1215-1227.2005>.
- Finak, Greg, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, et al. 2015. "MAST: A Flexible Statistical Framework for Assessing Transcriptional Changes and Characterizing Heterogeneity in Single-Cell RNA Sequencing Data." *Genome Biology* 16: 278. <https://doi.org/10.1186/s13059-015-0844-5>.
- Friedman, N., M. Linial, I. Nachman, and D. Pe'er. 2000. "Using Bayesian Networks to Analyze Expression Data." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 7 (3–4): 601–20. <https://doi.org/10.1089/106652700750050961>.
- Galloway, Jenna L., and Leonard I. Zon. 2003. "Ontogeny of Hematopoiesis: Examining the Emergence of Hematopoietic Cells in the Vertebrate Embryo." *Current Topics in Developmental Biology* 53: 139–58. [https://doi.org/10.1016/s0070-2153\(03\)53004-6](https://doi.org/10.1016/s0070-2153(03)53004-6).
- Gandrillon, O, U Schmidt, H Beug, and J Samarut. 1999. "TGF-Beta Cooperates with TGF-Alpha to Induce the Self-Renewal of Normal Erythrocytic Progenitors: Evidence for an Autocrine Mechanism." *The EMBO Journal* 18 (10): 2764–81. <https://doi.org/10.1093/emboj/18.10.2764>.
- Gandrillon, Olivier, Mathilde Gaillard, Thibault Espinasse, Nicolas B. Garnier, Charles Dussiau, Olivier Kosmider, and Pierre Sujobert. 2021. "Entropy as a Measure of Variability and Stemness in Single-Cell Transcriptomics." *Current Opinion in Systems Biology* 27 (September): 100348. <https://doi.org/10.1016/j.coisb.2021.05.009>.
- Gardner, Timothy S., Diego di Bernardo, David Lorenz, and James J. Collins. 2003. "Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling." *Science (New York, N.Y.)* 301 (5629): 102–5. <https://doi.org/10.1126/science.1081900>.
- Ghosh, Sourish, and Anirban Basu. 2012. "Network Medicine in Drug Design: Implications for Neuroinflammation." *Drug Discovery Today* 17 (11–12): 600–607. <https://doi.org/10.1016/j.drudis.2012.01.018>.
- Gibellini, Lara, Sara De Biasi, Camillo Porta, Domenico Lo Tartaro, Roberta Depenni, Giovanni Pellacani, Roberto Sabbatini, and Andrea Cossarizza. 2020. "Single-Cell Approaches to

- Profile the Response to Immune Checkpoint Inhibitors." *Frontiers in Immunology* 11: 490. <https://doi.org/10.3389/fimmu.2020.00490>.
- Girel, Simon, and Fabien Crauste. 2018. "Existence and Stability of Periodic Solutions of an Impulsive Differential Equation and Application to CD8 T-Cell Differentiation." *Journal of Mathematical Biology* 76 (7): 1765–95. <https://doi.org/10.1007/s00285-018-1220-3>.
- Glykofridis, Iris E., Jaco C. Knol, Jesper A. Balk, Denise Westland, Thang V. Pham, Sander R. Piersma, Sinéad M. Loughheed, et al. 2021. "Loss of FLCN-FNIP1/2 Induces a Non-Canonical Interferon Response in Human Renal Tubular Epithelial Cells." *ELife* 10 (January): e61630. <https://doi.org/10.7554/eLife.61630>.
- Gregory, C. J., and A. C. Eaves. 1978. "Three Stages of Erythropoietic Progenitor Cell Differentiation Distinguished by a Number of Physical and Biologic Properties." *Blood* 51 (3): 527–37.
- Gregory, T., C. Yu, A. Ma, S. H. Orkin, G. A. Blobel, and M. J. Weiss. 1999. "GATA-1 and Erythropoietin Cooperate to Promote Erythroid Cell Survival by Regulating Bcl-XL Expression." *Blood* 94 (1): 87–96.
- Grün, Dominic, and Alexander van Oudenaarden. 2015. "Design and Analysis of Single-Cell Sequencing Experiments." *Cell* 163 (4): 799–810. <https://doi.org/10.1016/j.cell.2015.10.039>.
- Guillemin, Anissa, Ronan Duchesne, Fabien Crauste, Sandrine Gonin-Giraud, and Olivier Gandrillon. 2019. "Drugs Modulating Stochastic Gene Expression Affect the Erythroid Differentiation Process." *PLoS ONE* 14 (11): e0225166. <https://doi.org/10.1371/journal.pone.0225166>.
- Guillemin, Anissa, and Michael P. H. Stumpf. 2021. "Noise and the Molecular Processes Underlying Cell Fate Decision-Making." *Physical Biology* 18 (1): 011002. <https://doi.org/10.1088/1478-3975/abc9d1>.
- Guo, X., W. Liu, Y. Pan, P. Ni, J. Ji, L. Guo, J. Zhang, et al. 2010. "Homeobox Gene IRX1 Is a Tumor Suppressor Gene in Gastric Carcinoma." *Oncogene* 29 (27): 3908–20. <https://doi.org/10.1038/onc.2010.143>.

- Haghverdi, Laleh, Florian Buettner, and Fabian J. Theis. 2015. "Diffusion Maps for High-Dimensional Single-Cell Analysis of Differentiation Data." *Bioinformatics (Oxford, England)* 31 (18): 2989–98. <https://doi.org/10.1093/bioinformatics/btv325>.
- Hannemann, Juliane, Sönke Meyer-Staeckling, Dirk Kemming, Iris Alpers, Simon A. Joosse, Heike Pospisil, Stefan Kurtz, et al. 2011. "Quantitative High-Resolution Genomic Analysis of Single Cancer Cells." *PLoS ONE* 6 (11): e26362. <https://doi.org/10.1371/journal.pone.0026362>.
- Haque, Ashraful, Jessica Engel, Sarah A. Teichmann, and Tapio Lönnberg. 2017. "A Practical Guide to Single-Cell RNA-Sequencing for Biomedical Research and Clinical Applications." *Genome Medicine* 9 (1): 75. <https://doi.org/10.1186/s13073-017-0467-4>.
- Hashimshony, Tamar, Naftalie Senderovich, Gal Avital, Agnes Klochender, Yaron de Leeuw, Leon Anavy, Dave Gennert, et al. 2016. "CEL-Seq2: Sensitive Highly-Multiplexed Single-Cell RNA-Seq." *Genome Biology* 17 (1): 77. <https://doi.org/10.1186/s13059-016-0938-8>.
- Hashimshony, Tamar, Florian Wagner, Noa Sher, and Itai Yanai. 2012. "CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification." *Cell Reports* 2 (3): 666–73. <https://doi.org/10.1016/j.celrep.2012.08.003>.
- Hasumi, Hisashi, Masaya Baba, Yukiko Hasumi, Martin Lang, Ying Huang, HyoungBin F. Oh, Masayuki Matsuo, et al. 2015. "Folliculin-Interacting Proteins Fnip1 and Fnip2 Play Critical Roles in Kidney Tumor Suppression in Cooperation with Flcn." *Proceedings of the National Academy of Sciences of the United States of America* 112 (13): E1624–1631. <https://doi.org/10.1073/pnas.1419502112>.
- Hasumi, Hisashi, Masaya Baba, Seung-Beom Hong, Yukiko Hasumi, Ying Huang, Masahiro Yao, Vladimir A. Valera, W. Marston Linehan, and Laura S. Schmidt. 2008. "Identification and Characterization of a Novel Folliculin-Interacting Protein FNIP2." *Gene* 415 (1–2): 60–67. <https://doi.org/10.1016/j.gene.2008.02.022>.
- Hay, Nissim, and Nahum Sonenberg. 2004. "Upstream and Downstream of MTOR." *Genes & Development* 18 (16): 1926–45. <https://doi.org/10.1101/gad.1212704>.

- Hedlund, Eva, and Qiaolin Deng. 2018. "Single-Cell RNA Sequencing: Technical Advancements and Biological Applications." *Molecular Aspects of Medicine* 59 (February): 36–46. <https://doi.org/10.1016/j.mam.2017.07.003>.
- Hicks, Stephanie C., F. William Townes, Mingxiang Teng, and Rafael A. Irizarry. 2018. "Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments." *Biostatistics (Oxford, England)* 19 (4): 562–78. <https://doi.org/10.1093/biostatistics/kxx053>.
- Hsieh, Hui-Ju, Wei Zhang, Shu-Hong Lin, Wen-Hao Yang, Jun-Zhong Wang, Jianfeng Shen, Yiran Zhang, et al. 2018. "Systems Biology Approach Reveals a Link between MTORC1 and G2/M DNA Damage Checkpoint Recovery." *Nature Communications* 9 (1): 3982. <https://doi.org/10.1038/s41467-018-05639-x>.
- Huang, Sui. 2009. "Non-Genetic Heterogeneity of Cells in Development: More than Just Noise." *Development* 136 (23): 3853–62. <https://doi.org/10.1242/dev.035139>.
- . 2012. "The Molecular and Mathematical Basis of Waddington's Epigenetic Landscape: A Framework for Post-Darwinian Biology?" *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 34 (2): 149–57. <https://doi.org/10.1002/bies.201100031>.
- Huang, Sui, Yan-Ping Guo, Gillian May, and Tariq Enver. 2007. "Bifurcation Dynamics in Lineage-Commitment in Bipotent Progenitor Cells." *Developmental Biology* 305 (2): 695–713. <https://doi.org/10.1016/j.ydbio.2007.02.036>.
- Hume, D. A. 2000. "Probability in Transcriptional Regulation and Its Implications for Leukocyte Differentiation and Inducible Gene Expression." *Blood* 96 (7): 2323–28.
- Islam, Saiful, Una Kjällquist, Annalena Moliner, Pawel Zajac, Jian-Bing Fan, Peter Lönnerberg, and Sten Linnarsson. 2011. "Characterization of the Single-Cell Transcriptional Landscape by Highly Multiplex RNA-Seq." *Genome Research* 21 (7): 1160–67. <https://doi.org/10.1101/gr.110882.110>.
- Islam, Saiful, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson. 2014. "Quantitative Single-Cell RNA-Seq with Unique Molecular Identifiers." *Nature Methods* 11 (2): 163–66. <https://doi.org/10.1038/nmeth.2772>.

- Iwata, Terri N., Julita A. Ramírez-Komo, Heon Park, and Brian M. Iritani. 2017. "Control of B Lymphocyte Development and Functions by the MTOR Signaling Pathways." *Cytokine & Growth Factor Reviews* 35 (June): 47–62. <https://doi.org/10.1016/j.cytogfr.2017.04.005>.
- Jaffredo, Thierry, Charlotte Richard, Claire Pouget, Marie-Aimée Teillet, Karine Bollérot, Rodolphe Gautier, and Cécile Drevon. 2010. "Aortic Remodelling during Hemogenesis: Is the Chicken Paradigm Unique?" *The International Journal of Developmental Biology* 54 (6–7): 1045–54. <https://doi.org/10.1387/ijdb.103062tj>.
- Jagannathan-Bogdan, Madhumita, and Leonard I. Zon. 2013. "Hematopoiesis." *Development (Cambridge, England)* 140 (12): 2463–67. <https://doi.org/10.1242/dev.083147>.
- Jaitin, Diego Adhemar, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, Irina Zaretsky, Alexander Mildner, et al. 2014. "Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types." *Science* 343 (6172): 776–79. <https://doi.org/10.1126/science.1247651>.
- Johnson, Nicole C., Miriam E. Dillard, Peter Baluk, Donald M. McDonald, Natasha L. Harvey, Sharon L. Frase, and Guillermo Oliver. 2008. "Lymphatic Endothelial Cell Identity Is Reversible and Its Maintenance Requires Prox1 Activity." *Genes & Development* 22 (23): 3282–91. <https://doi.org/10.1101/gad.1727208>.
- Jopling, Chris, Stephanie Boue, and Juan Carlos Izpisua Belmonte. 2011. "Dedifferentiation, Transdifferentiation and Reprogramming: Three Routes to Regeneration." *Nature Reviews Molecular Cell Biology* 12 (2): 79–89. <https://doi.org/10.1038/nrm3043>.
- Kassouf, Mira T., Jim R. Hughes, Stephen Taylor, Simon J. McGowan, Shamit Soneji, Angela L. Green, Paresh Vyas, and Catherine Porcher. 2010. "Genome-Wide Identification of TAL1's Functional Targets: Insights into Its Mechanisms of Action in Primary Erythroid Cells." *Genome Research* 20 (8): 1064–83. <https://doi.org/10.1101/gr.104935.110>.
- Katsumura, Koichi R., Emery H. Bresnick, and GATA Factor Mechanisms Group. 2017. "The GATA Factor Revolution in Hematology." *Blood* 129 (15): 2092–2102. <https://doi.org/10.1182/blood-2016-09-687871>.
- Kenyon, Emma J., Monique N. H. Luijten, Harmeet Gill, Nan Li, Matthew Rawlings, James C. Bull, Yavor Hadzhiev, Maurice A. M. van Steensel, Eamonn Maher, and Ferenc Mueller. 2016.

- “Expression and Knockdown of Zebrafish Folliculin Suggests Requirement for Embryonic Brain Morphogenesis.” *BMC Developmental Biology* 16 (1): 23. <https://doi.org/10.1186/s12861-016-0119-8>.
- Kim, Do-Hyung, D. D. Sarbassov, Siraj M. Ali, Jessie E. King, Robert R. Latek, Hediye Erdjument-Bromage, Paul Tempst, and David M. Sabatini. 2002. “MTOR Interacts with Raptor to Form a Nutrient-Sensitive Complex That Signals to the Cell Growth Machinery.” *Cell* 110 (2): 163–75. [https://doi.org/10.1016/s0092-8674\(02\)00808-5](https://doi.org/10.1016/s0092-8674(02)00808-5).
- Kim, Eunjung, Pankuri Goraksha-Hicks, Li Li, Thomas P. Neufeld, and Kun-Liang Guan. 2008. “Regulation of TORC1 by Rag GTPases in Nutrient Response.” *Nature Cell Biology* 10 (8): 935–45. <https://doi.org/10.1038/ncb1753>.
- Klenova, E. M., I. Botezato, V. Laudet, G. H. Goodwin, J. C. Wallace, and V. V. Lobanenko. 1992. “Isolation of a cDNA Clone Encoding the RNase-Superfamily-Related Gene Highly Expressed in Chicken Bone Marrow Cells.” *Biochemical and Biophysical Research Communications* 185 (1): 231–39. [https://doi.org/10.1016/s0006-291x\(05\)80980-5](https://doi.org/10.1016/s0006-291x(05)80980-5).
- Kolodziejczyk, Aleksandra A., Jong Kyoung Kim, Valentine Svensson, John C. Marioni, and Sarah A. Teichmann. 2015a. “The Technology and Biology of Single-Cell RNA Sequencing.” *Molecular Cell* 58 (4): 610–20. <https://doi.org/10.1016/j.molcel.2015.04.005>.
- . 2015b. “The Technology and Biology of Single-Cell RNA Sequencing.” *Molecular Cell* 58 (4): 610–20. <https://doi.org/10.1016/j.molcel.2015.04.005>.
- Kolodziejczyk, Aleksandra A., and Tapio Lönnberg. 2018. “Global and Targeted Approaches to Single-Cell Transcriptome Characterization.” *Briefings in Functional Genomics* 17 (4): 209–19. <https://doi.org/10.1093/bfpg/elx025>.
- Kowalski, Andrzej, and Jan Pałyga. 2011. “Chromatin Compaction in Terminally Differentiated Avian Blood Cells: The Role of Linker Histone H5 and Non-Histone Protein MENT.” *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 19 (5): 579–90. <https://doi.org/10.1007/s10577-011-9218-3>.
- Lachmann, Alexander, Federico M. Giorgi, Gonzalo Lopez, and Andrea Califano. 2016. “ARACNe-AP: Gene Network Reverse Engineering through Adaptive Partitioning Inference of

- Mutual Information.” *Bioinformatics (Oxford, England)* 32 (14): 2233–35.
<https://doi.org/10.1093/bioinformatics/btw216>.
- Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics* 9 (1): 559. <https://doi.org/10.1186/1471-2105-9-559>.
- Laviolette, Laura A., Jonas Wilson, Julia Koller, Christopher Neil, Peter Hulick, Tomas Rejtar, Barry Karger, Bin Tean Teh, and Othon Iliopoulos. 2013. “Human Folliculin Delays Cell Cycle Progression through Late S and G2/M-Phases: Effect of Phosphorylation and Tumor Associated Mutations.” *PloS One* 8 (7): e66775.
<https://doi.org/10.1371/journal.pone.0066775>.
- Lawrence, Rosalie E., Simon A. Fromm, Yangxue Fu, Adam L. Yokom, Do Jin Kim, Ashley M. Thelen, Lindsey N. Young, et al. 2019. “Structural Mechanism of a Rag GTPase Activation Checkpoint by the Lysosomal Folliculin Complex.” *Science* 366 (6468): 971–77.
<https://doi.org/10.1126/science.aax0364>.
- Lee, James C.-M., J. Aura Gimm, Annie J. Lo, Mark J. Koury, Sharon W. Krauss, Narla Mohandas, and Joel A. Chasis. 2004. “Mechanism of Protein Sorting during Erythroblast Enucleation: Role of Cytoskeletal Connectivity.” *Blood* 103 (5): 1912–19.
<https://doi.org/10.1182/blood-2003-03-0928>.
- Lever, Jake, Martin Krzywinski, and Naomi Altman. 2017. “Points of Significance: Principal Component Analysis.” *News. Nature Methods*. June 29, 2017.
<https://doi.org/10.1038/nmeth.4346>.
- Liu, Chengjun. 2004. “Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (5): 572–81. <https://doi.org/10.1109/TPAMI.2004.1273927>.
- Liu, Jing, Xijun Liang, Danxia Zhou, Ling Lai, Liwei Xiao, Lin Liu, Tingting Fu, et al. 2016. “Coupling of Mitochondrial Function and Skeletal Muscle Fiber Type by a MiR-499/Fnrip1/AMPK Circuit.” *EMBO Molecular Medicine* 8 (10): 1212–28.
<https://doi.org/10.15252/emmm.201606372>.

- Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, et al. 1996. "Expression Monitoring by Hybridization to High-Density Oligonucleotide Arrays." *Nature Biotechnology* 14 (13): 1675–80. <https://doi.org/10.1038/nbt1296-1675>.
- Loewith, Robbie, Estela Jacinto, Stephan Wullschleger, Anja Lorberg, José L. Crespo, Débora Bonenfant, Wolfgang Oppliger, Paul Jenoe, and Michael N. Hall. 2002. "Two TOR Complexes, Only One of Which Is Rapamycin Sensitive, Have Distinct Roles in Cell Growth Control." *Molecular Cell* 10 (3): 457–68. [https://doi.org/10.1016/s1097-2765\(02\)00636-6](https://doi.org/10.1016/s1097-2765(02)00636-6).
- Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9 (Nov): 2579–2605.
- Macneil, Lesley T., and Albertha J. M. Walhout. 2011. "Gene Regulatory Networks and the Role of Robustness and Stochasticity in the Control of Gene Expression." *Genome Research* 21 (5): 645–57. <https://doi.org/10.1101/gr.097378.109>.
- Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. "Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161 (5): 1202–14. <https://doi.org/10.1016/j.cell.2015.05.002>.
- Martello, Graziano, Toshimi Sugimoto, Evangelia Diamanti, Anagha Joshi, Rebecca Hannah, Satoshi Ohtsuka, Berthold Göttgens, Hitoshi Niwa, and Austin Smith. 2012. "Esrrb Is a Pivotal Target of the Gsk3/Tcf3 Axis Regulating Embryonic Stem Cell Self-Renewal." *Cell Stem Cell* 11 (4): 491–504. <https://doi.org/10.1016/j.stem.2012.06.008>.
- Martín Garrido, Natàlia de, and Christopher H. S. Aylett. 2020. "Nutrient Signaling and Lysosome Positioning Crosstalk Through a Multifunctional Protein, Folliculin." *Frontiers in Cell and Developmental Biology* 8: 108. <https://doi.org/10.3389/fcell.2020.00108>.
- McBride, Andrew, Stephanos Ghilagaber, Andrei Nikolaev, and D. Grahame Hardie. 2009. "The Glycogen-Binding Domain on the AMPK Beta Subunit Allows the Kinase to Act as a Glycogen Sensor." *Cell Metabolism* 9 (1): 23–34. <https://doi.org/10.1016/j.cmet.2008.11.008>.

- McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv:1802.03426 [Cs, Stat]*, September. <http://arxiv.org/abs/1802.03426>.
- Menko, Fred H., Maurice A. M. van Steensel, Sophie Giraud, Lennart Friis-Hansen, Stéphane Richard, Silvana Ungari, Magnus Nordenskjöld, et al. 2009. "Birt-Hogg-Dubé Syndrome: Diagnosis and Management." *The Lancet. Oncology* 10 (12): 1199–1206. [https://doi.org/10.1016/S1470-2045\(09\)70188-3](https://doi.org/10.1016/S1470-2045(09)70188-3).
- Miao, Zhun, Ke Deng, Xiaowo Wang, and Xuegong Zhang. 2018. "DEsingle for Detecting Three Types of Differential Expression in Single-Cell RNA-Seq Data." *Bioinformatics (Oxford, England)* 34 (18): 3223–24. <https://doi.org/10.1093/bioinformatics/bty332>.
- Mohammed, Hisham, Irene Hernando-Herraez, Aurora Savino, Antonio Scialdone, Iain Macaulay, Carla Mulas, Tamir Chandra, et al. 2017. "Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation." *Cell Reports* 20 (5): 1215–28. <https://doi.org/10.1016/j.celrep.2017.07.009>.
- Mojtahedi, Mitra, Alexander Skupin, Joseph Zhou, Ivan G. Castaño, Rebecca Y. Y. Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang. 2016. "Cell Fate Decision as High-Dimensional Critical State Transition." *PLoS Biology* 14 (12): e2000640. <https://doi.org/10.1371/journal.pbio.2000640>.
- Moris, Naomi, Shlomit Edri, Denis Seyres, Rashmi Kulkarni, Ana Filipa Domingues, Tina Balayo, Mattia Frontini, and Cristina Pina. 2018. "Histone Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity." *Stem Cells (Dayton, Ohio)* 36 (12): 1828–38. <https://doi.org/10.1002/stem.2919>.
- Moris, Naomi, Cristina Pina, and Alfonso Martinez Arias. 2016. "Transition States and Cell Fate Decisions in Epigenetic Landscapes." *Nature Reviews Genetics* 17 (11): 693–703. <https://doi.org/10.1038/nrg.2016.98>.
- Mou, Haiwei, Jordan L. Smith, Lingtao Peng, Hao Yin, Jill Moore, Xiao-Ou Zhang, Chun-Qing Song, et al. 2017. "CRISPR/Cas9-Mediated Genome Editing Induces Exon Skipping by Alternative Splicing or Exon Deletion." *Genome Biology* 18 (1): 108. <https://doi.org/10.1186/s13059-017-1237-8>.

- Moussy, Alice, Jérémie Cosette, Romuald Parmentier, Cindy da Silva, Guillaume Corre, Angélique Richard, Olivier Gandrillon, Daniel Stockholm, and András Páldi. 2017. "Integrated Time-Lapse and Single-Cell Transcription Studies Highlight the Variable and Dynamic Nature of Human Hematopoietic Cell Fate Commitment." *PLoS Biology* 15 (7): e2001867. <https://doi.org/10.1371/journal.pbio.2001867>.
- Munsky, Brian, Gregor Neuert, and Alexander van Oudenaarden. 2012. "Using Gene Expression Noise to Understand Gene Regulation." *Science (New York, N.Y.)* 336 (6078): 183–87. <https://doi.org/10.1126/science.1216379>.
- Nestorowa, Sonia, Fiona K. Hamey, Blanca Pijuan Sala, Evangelia Diamanti, Mairi Shepherd, Elisa Laurenti, Nicola K. Wilson, David G. Kent, and Berthold Göttgens. 2016. "A Single-Cell Resolution Map of Mouse Hematopoietic Stem and Progenitor Cell Differentiation." *Blood* 128 (8): e20–31. <https://doi.org/10.1182/blood-2016-05-716480>.
- Nguyen, Hung, Duc Tran, Bang Tran, Bahadir Pehlivan, and Tin Nguyen. 2021. "A Comprehensive Survey of Regulatory Network Inference Methods Using Single Cell RNA Sequencing Data." *Briefings in Bioinformatics* 22 (3): bbaa190. <https://doi.org/10.1093/bib/bbaa190>.
- Nickerson, Michael L., Michelle B. Warren, Jorge R. Toro, Vera Matrosova, Gladys Glenn, Maria L. Turner, Paul Duray, et al. 2002. "Mutations in a Novel Gene Lead to Kidney Tumors, Lung Wall Defects, and Benign Tumors of the Hair Follicle in Patients with the Birt-Hogg-Dubé Syndrome." *Cancer Cell* 2 (2): 157–64. [https://doi.org/10.1016/s1535-6108\(02\)00104-6](https://doi.org/10.1016/s1535-6108(02)00104-6).
- Novick, A., and M. Weiner. 1957. "ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON." *Proceedings of the National Academy of Sciences of the United States of America* 43 (7): 553–66.
- Oates, Chris J., Richard Amos, and Simon E. F. Spencer. 2014. "Quantifying the Multi-Scale Performance of Network Inference Algorithms." *Statistical Applications in Genetics and Molecular Biology* 13 (5): 611–31. <https://doi.org/10.1515/sagmb-2014-0012>.
- Oberlin, Estelle, Manuela Tavian, István Blazsek, and Bruno Péault. 2002. "Blood-Forming Potential of Vascular Endothelium in the Human Embryo." *Development* 129 (17): 4147–57. <https://doi.org/10.1242/dev.129.17.4147>.

- Orkin, S. H. 2000. "Diversification of Haematopoietic Stem Cells to Specific Lineages." *Nature Reviews. Genetics* 1 (1): 57–64. <https://doi.org/10.1038/35049577>.
- Orkin, Stuart H., and Leonard I. Zon. 2008. "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology." *Cell* 132 (4): 631–44. <https://doi.org/10.1016/j.cell.2008.01.025>.
- Österlund, Tobias, Sergio Bordel, and Jens Nielsen. 2015. "Controllability Analysis of Transcriptional Regulatory Networks Reveals Circular Control Patterns among Transcription Factors." *Integrative Biology* 7 (5): 560–68. <https://doi.org/10.1039/c4ib00247d>.
- Paik, Elizabeth J., and Leonard I. Zon. 2010. "Hematopoietic Development in the Zebrafish." *The International Journal of Developmental Biology* 54 (6–7): 1127–37. <https://doi.org/10.1387/ijdb.093042ep>.
- Park, Heon, Karen Staehling, Mark Tsang, Mark W. Appleby, Mary E. Brunkow, Daciana Margineantu, David M. Hockenbery, et al. 2012. "Disruption of Fnip1 Reveals a Metabolic Checkpoint Controlling B Lymphocyte Development." *Immunity* 36 (5): 769–81. <https://doi.org/10.1016/j.immuni.2012.02.019>.
- Park, Heon, Mark Tsang, Brian M. Iritani, and Michael J. Bevan. 2014. "Metabolic Regulator Fnip1 Is Crucial for INKT Lymphocyte Development." *Proceedings of the National Academy of Sciences of the United States of America* 111 (19): 7066–71. <https://doi.org/10.1073/pnas.1406473111>.
- Park, Sophie, Nicolas Chapuis, Jérôme Tamburini, Valérie Bardet, Pascale Cornillet-Lefebvre, Lise Willems, Alexa Green, Patrick Mayeux, Catherine Lacombe, and Didier Bouscary. 2010. "Role of the PI3K/AKT and MTOR Signaling Pathways in Acute Myeloid Leukemia." *Haematologica* 95 (5): 819–28. <https://doi.org/10.3324/haematol.2009.013797>.
- Paul, Franziska, Ya'ara Arkin, Amir Giladi, Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Deborah Winter, et al. 2015. "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors." *Cell* 163 (7): 1663–77. <https://doi.org/10.1016/j.cell.2015.11.013>.

- Perkins, A. C., A. H. Sharpe, and S. H. Orkin. 1995. "Lethal Beta-Thalassaemia in Mice Lacking the Erythroid CACCC-Transcription Factor EKLF." *Nature* 375 (6529): 318–22. <https://doi.org/10.1038/375318a0>.
- Picelli, Simone. 2017. "Single-Cell RNA-Sequencing: The Future of Genome Biology Is Now." *RNA Biology* 14 (5): 637–50. <https://doi.org/10.1080/15476286.2016.1201618>.
- Pina, Cristina, Cristina Fugazza, Alex J. Tipping, John Brown, Shamit Soneji, Jose Teles, Carsten Peterson, and Tariq Enver. 2012. "Inferring Rules of Lineage Commitment in Haematopoiesis." *Nature Cell Biology* 14 (3): 287–94. <https://doi.org/10.1038/ncb2442>.
- Pina, Cristina, José Teles, Cristina Fugazza, Gillian May, Dapeng Wang, Yanping Guo, Shamit Soneji, et al. 2015. "Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis." *Cell Reports* 11 (10): 1503–10. <https://doi.org/10.1016/j.celrep.2015.05.016>.
- Polynikis, A., S. J. Hogan, and M. di Bernardo. 2009. "Comparing Different ODE Modelling Approaches for Gene Regulatory Networks." *Journal of Theoretical Biology* 261 (4): 511–30. <https://doi.org/10.1016/j.jtbi.2009.07.040>.
- Popp, Maximilian W., and Lynne E. Maquat. 2016. "Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine." *Cell* 165 (6): 1319–22. <https://doi.org/10.1016/j.cell.2016.05.053>.
- Possik, Elite, Zahra Jalali, Yann Nouët, Ming Yan, Marie-Claude Gingras, Kathrin Schmeisser, Lorena Panaite, et al. 2014. "Folliculin Regulates Ampk-Dependent Autophagy and Metabolic Stress Survival." *PLoS Genetics* 10 (4): e1004273. <https://doi.org/10.1371/journal.pgen.1004273>.
- Qiu, Xiaojie, Andrew Hill, Jonathan Packer, Dejun Lin, Yi-An Ma, and Cole Trapnell. 2017. "Single-Cell mRNA Quantification and Differential Analysis with Census." *Nature Methods* 14 (3): 309–15. <https://doi.org/10.1038/nmeth.4150>.
- Rajagopal, Jayaraj, and Ben Z. Stanger. 2016. "Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues?" *Developmental Cell* 36 (2): 133–37. <https://doi.org/10.1016/j.devcel.2015.12.021>.

- Ramsköld, Daniel, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R. Faridani, Gregory A. Daniels, et al. 2012. "Full-Length mRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells." *Nature Biotechnology* 30 (8): 777–82. <https://doi.org/10.1038/nbt.2282>.
- Razin, A., and H. Cedar. 1991. "DNA Methylation and Gene Expression." *Microbiological Reviews* 55 (3): 451–58. <https://doi.org/10.1128/mr.55.3.451-458.1991>.
- Reyes, Nicholas L., Glen B. Banks, Mark Tsang, Daciana Margineantu, Haiwei Gu, Danijel Djukovic, Jacky Chan, et al. 2015. "Fnip1 Regulates Skeletal Muscle Fiber Type Specification, Fatigue Resistance, and Susceptibility to Muscular Dystrophy." *Proceedings of the National Academy of Sciences of the United States of America* 112 (2): 424–29. <https://doi.org/10.1073/pnas.1413021112>.
- Richard, Angélique, Loïs Boullu, Ulysse Herbach, Arnaud Bonnafox, Valérie Morin, Elodie Vallin, Anissa Guillemin, et al. 2016a. "Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process." *PLoS Biology* 14 (12): e1002585. <https://doi.org/10.1371/journal.pbio.1002585>.
- . 2016b. "Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process." *PLoS Biology* 14 (12): e1002585. <https://doi.org/10.1371/journal.pbio.1002585>.
- Ridden, Sonya J., Hannah H. Chang, Konstantinos C. Zygalkis, and Ben D. MacArthur. 2015. "Entropy, Ergodicity, and Stem Cell Multipotency." *Physical Review Letters* 115 (20): 208103. <https://doi.org/10.1103/PhysRevLett.115.208103>.
- Rosenberg, Alexander B., Charles M. Roco, Richard A. Muscat, Anna Kuchina, Paul Sample, Zizhen Yao, Lucas T. Graybuck, et al. 2018. "Single-Cell Profiling of the Developing Mouse Brain and Spinal Cord with Split-Pool Barcoding." *Science* 360 (6385): 176–82. <https://doi.org/10.1126/science.aam8999>.
- Rozenblatt-Rosen, Orit, Rahul C. Deo, Megha Padi, Guillaume Adelmant, Michael A. Calderwood, Thomas Rolland, Miranda Grace, et al. 2012. "Interpreting Cancer Genomes Using Systematic Host Network Perturbations by Tumour Virus Proteins." *Nature* 487 (7408): 491–95. <https://doi.org/10.1038/nature11288>.

- Sagar, null, and Dominic Grün. 2020. "Deciphering Cell Fate Decision by Integrated Single-Cell Sequencing Analysis." *Annual Review of Biomedical Data Science* 3 (July): 1–22. <https://doi.org/10.1146/annurev-biodatasci-111419-091750>.
- Sager, Rebecca A., Mark R. Woodford, Sarah J. Backe, Alan M. Makedon, Alexander J. Baker-Williams, Bryanna T. DiGregorio, David R. Loiselle, et al. 2019. "Post-Translational Regulation of FNIP1 Creates a Rheostat for the Molecular Chaperone Hsp90." *Cell Reports* 26 (5): 1344-1356.e5. <https://doi.org/10.1016/j.celrep.2019.01.018>.
- Sager, Rebecca A., Mark R. Woodford, Oleg Shapiro, Mehdi Mollapour, and Gennady Bratslavsky. 2018. "Sporadic Renal Angiomyolipoma in a Patient with Birt-Hogg-Dubé: Chaperones in Pathogenesis." *Oncotarget* 9 (31): 22220–29. <https://doi.org/10.18632/oncotarget.25164>.
- Sancak, Yasemin, Timothy R. Peterson, Yoav D. Shaul, Robert A. Lindquist, Carson C. Thoreen, Liron Bar-Peled, and David M. Sabatini. 2008. "The Rag GTPases Bind Raptor and Mediate Amino Acid Signaling to mTORC1." *Science* 320 (5882): 1496–1501. <https://doi.org/10.1126/science.1157535>.
- Sanchez-Freire, Veronica, Antje D. Ebert, Tomer Kalisky, Stephen R. Quake, and Joseph C. Wu. 2012. "Microfluidic Single Cell Real-Time PCR for Comparative Analysis of Gene Expression Patterns." *Nature Protocols* 7 (5): 10.1038/nprot.2012.021. <https://doi.org/10.1038/nprot.2012.021>.
- Sasagawa, Yohei, Itoshi Nikaido, Tetsutaro Hayashi, Hiroki Danno, Kenichiro D. Uno, Takeshi Imai, and Hiroki R. Ueda. 2013. "Quartz-Seq: A Highly Reproducible and Sensitive Single-Cell RNA Sequencing Method, Reveals Non-Genetic Gene-Expression Heterogeneity." *Genome Biology* 14 (4): 3097. <https://doi.org/10.1186/gb-2013-14-4-r31>.
- Sato, Tatsuhiro, Akio Nakashima, Lea Guo, and Fuyuhiko Tamanoi. 2009. "Specific Activation of mTORC1 by Rheb G-Protein in Vitro Involves Enhanced Recruitment of Its Substrate Protein." *The Journal of Biological Chemistry* 284 (19): 12783–91. <https://doi.org/10.1074/jbc.M809207200>.

- Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. 1995. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." *Science (New York, N.Y.)* 270 (5235): 467–70. <https://doi.org/10.1126/science.270.5235.467>.
- Schmidt, Laura S. 2013. "Birt-Hogg-Dubé Syndrome: From Gene Discovery to Molecularly Targeted Therapies." *Familial Cancer* 12 (3): 357–64. <https://doi.org/10.1007/s10689-012-9574-y>.
- Schmidt, Laura S., and W. Marston Linehan. 2015a. "Clinical Features, Genetics and Potential Therapeutic Approaches for Birt-Hogg-Dubé Syndrome." *Expert Opinion on Orphan Drugs* 3 (1): 15–29. <https://doi.org/10.1517/21678707.2014.987124>.
- . 2015b. "Molecular Genetics and Clinical Features of Birt-Hogg-Dubé Syndrome." *Nature Reviews. Urology* 12 (10): 558–69. <https://doi.org/10.1038/nrurol.2015.206>.
- Schmittgen, Thomas D., and Kenneth J. Livak. 2008. "Analyzing Real-Time PCR Data by the Comparative C(T) Method." *Nature Protocols* 3 (6): 1101–8. <https://doi.org/10.1038/nprot.2008.73>.
- Schopf, Florian H., Maximilian M. Biebl, and Johannes Buchner. 2017. "The HSP90 Chaperone Machinery." *Nature Reviews Molecular Cell Biology* 18 (6): 345–60. <https://doi.org/10.1038/nrm.2017.20>.
- Shahi, Rajendra Bahadur, Sylvia De Brakeleer, Ben Caljon, Ingrid Pauwels, Maryse Bonduelle, Sofie Joris, Christel Fontaine, et al. 2019. "Identification of Candidate Cancer Predisposing Variants by Performing Whole-Exome Sequencing on Index Patients from BRCA1 and BRCA2-Negative Breast Cancer Families." *BMC Cancer* 19 (1): 313. <https://doi.org/10.1186/s12885-019-5494-7>.
- Sharpe, Joshua J., and Thomas A. Cooper. 2017. "Unexpected Consequences: Exon Skipping Caused by CRISPR-Generated Mutations." *Genome Biology* 18 (1): 109. <https://doi.org/10.1186/s13059-017-1240-0>.
- Shimoni, Yishai, Marc Y. Fink, Soon-gang Choi, and Stuart C. Sealfon. 2010. "Plato's Cave Algorithm: Inferring Functional Signaling Networks from Early Gene Expression Shadows." *PLOS Computational Biology* 6 (6): e1000828. <https://doi.org/10.1371/journal.pcbi.1000828>.

- Siggs, Owen M., Alexander Stockenhuber, Mukta Deobagkar-Lele, Katherine R. Bull, Tanya L. Crockford, Bethany L. Kingston, Greg Crawford, et al. 2016. "Mutation of Fnip1 Is Associated with B-Cell Deficiency, Cardiomyopathy, and Elevated AMPK Activity." *Proceedings of the National Academy of Sciences of the United States of America* 113 (26): E3706-3715. <https://doi.org/10.1073/pnas.1607592113>.
- Silvescu, Adrian, and Vasant Honavar. 2001. "Temporal Boolean Network Models of Genetic Networks and Their Inference from Gene Expression Time Series: Proceedings of the Atlantic Symposium on Computational Biology and Genome Information Systems and Technology, GBGIST 2001." Edited by C.H. Wu, P.P. Wang, and J.T.L. Wang. *Proceedings of the Atlantic Symposium on Computational Biology and Genome Information Systems and Technology, CBGIST 2001*, December, 260–65.
- Singer, Zakary S., John Yong, Julia Tischler, Jamie A. Hackett, Alphan Altinok, M. Azim Surani, Long Cai, and Michael B. Elowitz. 2014. "Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells." *Molecular Cell* 55 (2): 319–31. <https://doi.org/10.1016/j.molcel.2014.06.029>.
- Soni, Shivani, Shashi Bala, Babette Gwynn, Kenneth E. Sahr, Luanne L. Peters, and Manjit Hanspal. 2006. "Absence of Erythroblast Macrophage Protein (Emp) Leads to Failure of Erythroblast Nuclear Extrusion." *The Journal of Biological Chemistry* 281 (29): 20181–89. <https://doi.org/10.1074/jbc.M603226200>.
- Spurgeon, Sandra L., Robert C. Jones, and Ramesh Ramakrishnan. 2008. "High Throughput Gene Expression Measurement with Real Time PCR in a Microfluidic Dynamic Array." *PloS One* 3 (2): e1662. <https://doi.org/10.1371/journal.pone.0001662>.
- Stegle, Oliver, Sarah A. Teichmann, and John C. Marioni. 2015. "Computational and Analytical Challenges in Single-Cell Transcriptomics." *Nature Reviews Genetics* 16 (3): 133–45. <https://doi.org/10.1038/nrg3833>.
- Stern, Shay, Tali Dror, Elad Stolovicki, Naama Brenner, and Erez Braun. 2007. "Genome-Wide Transcriptional Plasticity Underlies Cellular Adaptation to Novel Challenge." *Molecular Systems Biology* 3 (April): 106. <https://doi.org/10.1038/msb4100147>.

- Steuer, R., J. Kurths, C. O. Daub, J. Weise, and J. Selbig. 2002. "The Mutual Information: Detecting and Evaluating Dependencies between Variables." *Bioinformatics* 18 (suppl_2): S231–40. https://doi.org/10.1093/bioinformatics/18.suppl_2.S231.
- Strohman, Richard C. 1997. "The Coming Kuhnian Revolution in Biology." *Nature Biotechnology* 15 (3): 194–200. <https://doi.org/10.1038/nbt0397-194>.
- Stumpf, Patrick S., Rosanna C. G. Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann Babbie, Thalia E. Chan, et al. 2017. "Stem Cell Differentiation as a Non-Markov Stochastic Process." *Cell Systems* 5 (3): 268-282.e7. <https://doi.org/10.1016/j.cels.2017.08.009>.
- Takagi, Y., T. Kobayashi, M. Shiono, L. Wang, X. Piao, G. Sun, D. Zhang, et al. 2008. "Interaction of Folliculin (Birt-Hogg-Dubé Gene Product) with a Novel Fnip1-like (FnipL/Fnip2) Protein." *Oncogene* 27 (40): 5339–47. <https://doi.org/10.1038/onc.2008.261>.
- Tanay, Amos, and Aviv Regev. 2017. "Scaling Single-Cell Genomics from Phenomenology to Mechanism." *Nature* 541 (7637): 331–38. <https://doi.org/10.1038/nature21350>.
- Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, et al. 2009. "MRNA-Seq Whole-Transcriptome Analysis of a Single Cell." *Nature Methods* 6 (5): 377–82. <https://doi.org/10.1038/nmeth.1315>.
- Tata, Purushothama Rao, Hongmei Mou, Ana Pardo-Saganta, Rui Zhao, Mythili Prabhu, Brandon M. Law, Vladimir Vinarsky, et al. 2013. "Dedifferentiation of Committed Epithelial Cells into Stem Cells in Vivo." *Nature* 503 (7475): 218–23. <https://doi.org/10.1038/nature12777>.
- Thattai, Mukund, and Alexander van Oudenaarden. 2004. "Stochastic Gene Expression in Fluctuating Environments." *Genetics* 167 (1): 523–30.
- Thorne, Thomas, and Michael P. H. Stumpf. 2012. "Inference of Temporally Varying Bayesian Networks." *Bioinformatics* 28 (24): 3298–3305. <https://doi.org/10.1093/bioinformatics/bts614>.
- Tian, Wen-Liang, Fei He, Xue Fu, Jun-Tang Lin, Ping Tang, Yu-Min Huang, Rong Guo, and Ling Sun. 2014. "High Expression of Heat Shock Protein 90 Alpha and Its Significance in Human

- Acute Leukemia Cells.” *Gene* 542 (2): 122–28.
<https://doi.org/10.1016/j.gene.2014.03.046>.
- Till, J. E., and E. A. McCulloch. 1980. “Hemopoietic Stem Cell Differentiation.” *Biochimica Et Biophysica Acta* 605 (4): 431–59. [https://doi.org/10.1016/0304-419x\(80\)90009-8](https://doi.org/10.1016/0304-419x(80)90009-8).
- Tirosh, Itay, Benjamin Izar, Sanjay M. Prakadan, Marc H. Wadsworth, Daniel Treacy, John J. Trombetta, Asaf Rotem, et al. 2016. “Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq.” *Science (New York, N.Y.)* 352 (6282): 189–96.
<https://doi.org/10.1126/science.aad0501>.
- Tomasoni, Romana, and Anna Mondino. 2011. “The Tuberous Sclerosis Complex: Balancing Proliferation and Survival.” *Biochemical Society Transactions* 39 (2): 466–71.
<https://doi.org/10.1042/BST0390466>.
- Tsang, Jason C. H., Yong Yu, Shannon Burke, Florian Buettner, Cui Wang, Aleksandra A. Kolodziejczyk, Sarah A. Teichmann, Liming Lu, and Pentao Liu. 2015. “Single-Cell Transcriptomic Reconstruction Reveals Cell Cycle and Multi-Lineage Differentiation Defects in Bcl11a-Deficient Hematopoietic Stem Cells.” *Genome Biology* 16 (September): 178. <https://doi.org/10.1186/s13059-015-0739-5>.
- Tsun, Zhi-Yang, Liron Bar-Peled, Lynne Chantranupong, Roberto Zoncu, Tim Wang, Choah Kim, Eric Spooner, and David M. Sabatini. 2013. “The Folliculin Tumor Suppressor Is a GAP for the RagC/D GTPases That Signal Amino Acid Levels to MTORC1.” *Molecular Cell* 52 (4): 495–505. <https://doi.org/10.1016/j.molcel.2013.09.016>.
- Tung, Po-Yuan, John D. Blischak, Chiaowen Joyce Hsiao, David A. Knowles, Jonathan E. Burnett, Jonathan K. Pritchard, and Yoav Gilad. 2017. “Batch Effects and the Effective Design of Single-Cell Gene Expression Studies.” *Scientific Reports* 7 (1): 39921.
<https://doi.org/10.1038/srep39921>.
- Vallejos, Catalina A., John C. Marioni, and Sylvia Richardson. 2015. “BASiCS: Bayesian Analysis of Single-Cell Sequencing Data.” *PLoS Computational Biology* 11 (6): e1004333.
<https://doi.org/10.1371/journal.pcbi.1004333>.

- Vallejos, Catalina A., Davide Risso, Antonio Scialdone, Sandrine Dudoit, and John C. Marioni. 2017. "Normalizing Single-Cell RNA Sequencing Data: Challenges and Opportunities." *Nature Methods* 14 (6): 565–71. <https://doi.org/10.1038/nmeth.4292>.
- Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. "Serial Analysis of Gene Expression." *Science (New York, N.Y.)* 270 (5235): 484–87. <https://doi.org/10.1126/science.270.5235.484>.
- Velten, Lars, Simon F. Haas, Simon Raffel, Sandra Blaszkiewicz, Saiful Islam, Bianca P. Hennig, Christoph Hirche, et al. 2017. "Human Haematopoietic Stem Cell Lineage Commitment Is a Continuous Process." *Nature Cell Biology* 19 (4): 271–81. <https://doi.org/10.1038/ncb3493>.
- Ventre, Elias, Thibault Espinasse, Charles-Edouard Bréhier, Vincent Calvez, Thomas Lepoutre, and Olivier Gandrillon. 2020. "Reduction of a Stochastic Model of Gene Expression: Lagrangian Dynamics Gives Access to Basins of Attraction as Cell Types and Metastability." <https://doi.org/10.1101/2020.09.04.283176>.
- Verd, Berta, Anton Crombach, and Johannes Jaeger. 2014. "Classification of Transient Behaviours in a Time-Dependent Toggle Switch Model." *BMC Systems Biology* 8 (1): 43. <https://doi.org/10.1186/1752-0509-8-43>.
- Verdone, Loredana, Micaela Caserta, and Ernesto Di Mauro. 2005. "Role of Histone Acetylation in the Control of Gene Expression." *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire* 83 (3): 344–53. <https://doi.org/10.1139/o05-041>.
- Viñuelas, José, Gaël Kaneko, Antoine Coulon, Elodie Vallin, Valérie Morin, Camila Mejia-Pous, Jean-Jacques Kupiec, Guillaume Beslon, and Olivier Gandrillon. 2013a. "Quantifying the Contribution of Chromatin Dynamics to Stochastic Gene Expression Reveals Long, Locus-Dependent Periods between Transcriptional Bursts." *BMC Biology* 11 (1): 15. <https://doi.org/10.1186/1741-7007-11-15>.
- . 2013b. "Quantifying the Contribution of Chromatin Dynamics to Stochastic Gene Expression Reveals Long, Locus-Dependent Periods between Transcriptional Bursts." *BMC Biology* 11 (1): 15. <https://doi.org/10.1186/1741-7007-11-15>.

- Waddington, C. H. 1957. "The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology. With an Appendix by H. Kacser." *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology. With an Appendix by H. Kacser.* <https://www.cabdirect.org/cabdirect/abstract/19580101706>.
- Wagner, K. U., E. Claudio, E. B. Rucker, G. Riedlinger, C. Broussard, P. L. Schwartzberg, U. Siebenlist, and L. Hennighausen. 2000. "Conditional Deletion of the Bcl-x Gene from Erythroid Cells Results in Hemolytic Anemia and Profound Splenomegaly." *Development (Cambridge, England)* 127 (22): 4949–58.
- Wang, Jian, and Yuanlin Song. 2017. "Single Cell Sequencing: A Distinct New Field." *Clinical and Translational Medicine* 6 (1): 10. <https://doi.org/10.1186/s40169-017-0139-4>.
- Wang, Lu, Toshiyuki Kobayashi, Xianghua Piao, Masatoshi Shiono, Yumiko Takagi, Reiko Mineki, Hikari Taka, et al. 2010. "Serine 62 Is a Phosphorylation Site in Folliculin, the Birt-Hogg-Dubé Gene Product." *FEBS Letters* 584 (1): 39–43. <https://doi.org/10.1016/j.febslet.2009.11.033>.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews. Genetics* 10 (1): 57–63. <https://doi.org/10.1038/nrg2484>.
- Weinberger, Leor S., John C. Burnett, Jared E. Toettcher, Adam P. Arkin, and David V. Schaffer. 2005. "Stochastic Gene Expression in a Lentiviral Positive-Feedback Loop: HIV-1 Tat Fluctuations Drive Phenotypic Diversity." *Cell* 122 (2): 169–82. <https://doi.org/10.1016/j.cell.2005.06.006>.
- Weisenfeld, Neil I., Vijay Kumar, Preyas Shah, Deanna M. Church, and David B. Jaffe. 2017. "Direct Determination of Diploid Genome Sequences." *Genome Research* 27 (5): 757–67. <https://doi.org/10.1101/gr.214874.116>.
- Witczak, C. A., C. G. Sharoff, and L. J. Goodyear. 2008. "AMP-Activated Protein Kinase in Skeletal Muscle: From Structure and Localization to Its Role as a Master Regulator of Cellular Metabolism." *Cellular and Molecular Life Sciences: CMLS* 65 (23): 3737–55. <https://doi.org/10.1007/s00018-008-8244-6>.

- Wolff, Linda, and Rita Humeniuk. 2013. "Concise Review: Erythroid Versus Myeloid Lineage Commitment: Regulating the Master Regulators." *STEM CELLS* 31 (7): 1237–44. <https://doi.org/10.1002/stem.1379>.
- Wolpert, Lewis. 1994. "Do We Understand Development?" *Science*, October. <https://www.science.org/doi/abs/10.1126/science.7939707>.
- Woodford, Mark R., Diana M. Dunn, Adam R. Blanden, Dante Capriotti, David Loiselle, Chrisostomos Prodromou, Barry Panaretou, et al. 2016. "The FNIP Co-Chaperones Decelerate the Hsp90 Chaperone Cycle and Enhance Drug Binding." *Nature Communications* 7 (June): 12037. <https://doi.org/10.1038/ncomms12037>.
- Wu, H., U. Klingmüller, P. Besmer, and H. F. Lodish. 1995. "Interaction of the Erythropoietin and Stem-Cell-Factor Receptors." *Nature* 377 (6546): 242–46. <https://doi.org/10.1038/377242a0>.
- Wu, Weisheng, Yong Cheng, Cheryl A. Keller, Jason Ernst, Swathi Ashok Kumar, Tejaswini Mishra, Christopher Morrissey, et al. 2011. "Dynamics of the Epigenetic Landscape during Erythroid Differentiation after GATA1 Restoration." *Genome Research* 21 (10): 1659–71. <https://doi.org/10.1101/gr.125088.111>.
- Xuan Vinh, Nguyen, Madhu Chetty, Ross Coppel, and Pramod P. Wangikar. 2012. "Gene Regulatory Network Modeling via Global Optimization of High-Order Dynamic Bayesian Network." *BMC Bioinformatics* 13 (1): 131. <https://doi.org/10.1186/1471-2105-13-131>.
- Yosef, Nir, Alex K. Shalek, Jellert T. Gaublot, Hulin Jin, Youjin Lee, Amit Awasthi, Chuan Wu, et al. 2013. "Dynamic Regulatory Network Controlling T H 17 Cell Differentiation." *Nature* 496 (7446): 461–68. <https://doi.org/10.1038/nature11981>.
- Young, William Chad, Adrian E. Raftery, and Ka Yee Yeung. 2014. "Fast Bayesian Inference for Gene Regulatory Networks Using ScanBMA." *BMC Systems Biology* 8 (1): 47. <https://doi.org/10.1186/1752-0509-8-47>.
- Yu, Jing, V. Anne Smith, Paul P. Wang, Alexander J. Hartemink, and Erich D. Jarvis. 2004. "Advances to Bayesian Network Inference for Generating Causal Networks from Observational Biological Data." *Bioinformatics (Oxford, England)* 20 (18): 3594–3603. <https://doi.org/10.1093/bioinformatics/bth448>.

- Zalc, Antoine, Rahul Sinha, Gunsagar S. Gulati, Daniel J. Wesche, Patrycja Daszczuk, Tomek Swigut, Irving L. Weissman, and Joanna Wysocka. 2021. "Reactivation of the Pluripotency Program Precedes Formation of the Cranial Neural Crest." *Science* 371 (6529). <https://doi.org/10.1126/science.abb4776>.
- Zhang, Yong, Xinsheng Gao, Leslie J. Saucedo, Binggen Ru, Bruce A. Edgar, and DuoJia Pan. 2003. "Rheb Is a Direct Target of the Tuberous Sclerosis Tumour Suppressor Proteins." *Nature Cell Biology* 5 (6): 578–81. <https://doi.org/10.1038/ncb999>.
- Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert. 2001. "Reverse Transcriptase Template Switching: A SMART Approach for Full-Length cDNA Library Construction." *BioTechniques* 30 (4): 892–97. <https://doi.org/10.2144/01304pf02>.
- Ziegenhain, Christoph, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, Martha Smets, Heinrich Leonhardt, Holger Heyn, Ines Hellmann, and Wolfgang Enard. 2017. "Comparative Analysis of Single-Cell RNA Sequencing Methods." *Molecular Cell* 65 (4): 631–643.e4. <https://doi.org/10.1016/j.molcel.2017.01.023>.
- Zoncu, Roberto, Liron Bar-Peled, Alejo Efeyan, Shuyu Wang, Yasemin Sancak, and David M. Sabatini. 2011. "mTORC1 Senses Lysosomal Amino Acids through an Inside-out Mechanism That Requires the Vacuolar H⁽⁺⁾-ATPase." *Science (New York, N.Y.)* 334 (6056): 678–83. <https://doi.org/10.1126/science.1207056>.