



HAL
open science

Analyse de la variabilité de l'expression génique et du métabolisme glycolytique au cours du processus de différenciation érythrocytaire : de l'analyse à grande échelle aux questions mécanistiques

Angélique Richard

► To cite this version:

Angélique Richard. Analyse de la variabilité de l'expression génique et du métabolisme glycolytique au cours du processus de différenciation érythrocytaire : de l'analyse à grande échelle aux questions mécanistiques. Biologie cellulaire. Université de Lyon, 2018. Français. NNT : 2018LYSE1058 . tel-02092005v1

HAL Id: tel-02092005

<https://theses.hal.science/tel-02092005v1>

Submitted on 7 Apr 2019 (v1), last revised 8 Apr 2019 (v2)

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.



N° d'ordre NNT : 2018LYSE1058

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED340
Biologie Moléculaire Intégrative et Cellulaire

Spécialité de doctorat :
Biologie

Soutenue publiquement le 06/04/2018, par :
Angélique Richard

**Analyse de la variabilité de l'expression
génique et du métabolisme glycolytique
au cours du processus de différenciation
érythrocytaire : De l'analyse à grande
échelle aux questions mécanistiques**

Devant le jury composé de :

PALDI Andràs, DE, EPHE-PSL

Président

CREPIEUX Pascale, DR, INRA

Rapporteuse

PERONNET Frédérique, DR, IBPS

Rapporteuse

GANDRILLON Olivier, DR, LBMC

Examinateur

GONIN-GIRAUD Sandrine, MCU, LBMC

Directrice de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Président du Conseil Académique

Vice-président du Conseil d'Administration

Vice-président du Conseil Formation et Vie Universitaire

Vice-président de la Commission Recherche

Directrice Générale des Services

M. le Professeur Frédéric FLEURY

M. le Professeur Hamda BEN HADID

M. le Professeur Didier REVEL

M. le Professeur Philippe CHEVALIER

M. Fabrice VALLÉE

Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de la Réadaptation

Département de formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur G.RODE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C. VINCIGUERRA

Directeur : M. X. PERROT

Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

UFR Sciences et Techniques des Activités Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Ecole Supérieure du Professorat et de l'Education

Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI

Directeur : M. le Professeur F. THEVENARD

Directeur : Mme C. FELIX

Directeur : M. Hassan HAMMOURI

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur G. TOMANOV

Directeur : M. le Professeur H. BEN HADID

Directeur : M. le Professeur J-C PLENET

Directeur : M. Y.VANPOULLE

Directeur : M. B. GUIDERDONI

Directeur : M. le Professeur E.PERRIN

Directeur : M. G. PIGNAULT

Directeur : M. le Professeur C. VITON

Directeur : M. le Professeur A. MOUGNIOTTE

Directeur : M. N. LEBOISNE

Remerciements

Il s'agit probablement pour moi de la partie la plus difficile à écrire, pour l'émotion qu'elle génère et pour trouver les mots qui exprimeront avec justesse mes sentiments de reconnaissance. Ces années de thèse ont été une belle aventure, de celles que l'on n'oublie pas, et je souhaiterais donc remercier toutes les personnes grâce à qui j'ai pu vivre et écrire cette histoire.

Je souhaiterais tout d'abord remercier Sandrine, pour tout ce qu'elle m'a apporté, tant du point de vue personnel que du point de vue scientifique. Depuis le jour où nous avons commencé à travailler ensemble elle m'a sans cesse encouragée et soutenue, ce qui m'a permis de toujours rester motivée malgré les épreuves de la thèse. Mais elle m'a surtout aidée à prendre progressivement confiance en moi et à croire en mes capacités, ce qui n'était vraiment pas gagné. Grâce à Sandrine, j'ai réussi à m'épanouir scientifiquement et à atteindre aujourd'hui ce que je n'aurais jamais imaginé être possible avant. Elle a toujours attaché de l'importance à ma réussite mais sans en oublier mon bien être, et je lui en suis extrêmement reconnaissante. Elle a vraiment représenté un pilier pour moi pendant ces années de thèse, et j'espère très sincèrement que la fin de cette collaboration ne séparera pas nos chemins. Je tenais donc à la remercier de tout coeur à travers ces quelques lignes, qui sont bien peu par rapport à ma véritable reconnaissance, pour la superbe encadrante, compétente et juste, qu'elle a été, mais aussi pour toutes ses qualités humaines, et surtout son incroyable gentillesse.

Je souhaiterais également remercier Olivier, tout d'abord pour avoir accepté de m'accueillir dans son équipe en thèse, et pour m'avoir permis de démarrer une nouvelle aventure à leurs côtés. Grâce à Olivier, je n'ai pas juste fait une thèse, j'ai découvert un domaine qui m'était nouveau, qui m'a permis d'élargir mon éventail de connaissances et de compétences scientifiques. En effet, il m'a donné l'opportunité de travailler sur une thématique difficile, mais qu'il a su rendre accessible, et surtout passionnante et motivante. Je tiens donc particulièrement à le remercier car grâce à lui, j'ai osé plonger sans même hésiter dans un monde qui était éloigné de mes connaissances, et qui contre toute attente est devenu aujourd'hui mon univers scientifique. Cela a vraiment été une chance de travailler à ses côtés, et je pèse mes mots, car le challenge était de taille mais grâce à son aide, qui m'a été très précieuse, j'ai aujourd'hui réussi ce défi, qui me permet de prendre progressivement confiance en mes capacités. Enfin, je tiens aussi à le remercier pour toutes ses qualités humaines, qui font qu'au final que l'on soit 5 ou 12 dans son équipe, l'ambiance est toujours aussi bonne.

Je ne peux pas ne pas remercier aussi chacun des membres de l'équipe SBDM qui ont tous apporté leur petite contribution à cette superbe aventure. Je tiens premièrement à tous les remercier pour rendre le quotidien au laboratoire si agréable, car grâce à eux les journées ne paraissent plus aussi longues. J'ai eu la chance de faire partie d'une équipe joyeuse et qui s'entraide. Un petit coup de blues, levée du mauvais pied? Dans l'équipe ce n'est pas un problème, il y a toujours une personne disponible pour boire un café, refaire le monde, et finalement retourner travailler l'esprit plus léger. La bonne ambiance au sein de l'équipe installe un climat humain confortable, qui motive même les jours où l'on aurait préféré rester caché sous la couette. C'est donc pour toutes ces raisons que je tiens à remercier chacun d'entre eux individuellement.

Je souhaite tout d'abord remercier particulièrement Elodie, qui comme Sandrine et Olivier, est à mes côtés depuis mes débuts dans l'équipe. Elle m'a apporté une aide technique précieuse tout au long de ma thèse, et surtout dernièrement dans le projet métabolisme. Mais c'est surtout pour être elle même que je souhaite la remercier. J'ai découvert une très belle personne, dont la gentillesse et la bonne humeur naturelles font que c'est un bonheur de travailler à ses côtés au quotidien.

Comment ne pas remercier Anissa aussi, sur qui j'ai pu compter tout au long de ma thèse. Je ne compte plus le nombre de petits services qu'elle m'a rendu juste par sympathie. Mais je retiens surtout qu'elle a toujours été présente, dans les moments joyeux comme difficiles de ma thèse. Toujours droite et franche envers moi, elle a gagné ma confiance, et je ne l'ai jamais regretté. Je la remercie donc très sincèrement pour ces qualités, mais aussi pour tous les bons moments passés ensemble en dehors du laboratoire, et qui ont parfois redonné une note bien plus positive aux fins de semaines un peu moroses.

Je remercie ensuite notre Maman MacGyver du laboratoire, Catherine, pour toute l'aide qu'elle nous apporte au quotidien. Elle veille sur notre sécurité, notre santé et notre bien être. À tout problème, elle trouve toujours une solution, elle m'impressionne vraiment. Donc Catherine, je ne peux que te remercier d'être toi.

Il est également important pour moi de remercier Gérard, de qui je n'oublie pas les paroles toujours encourageantes et réconfortantes par rapport à mon travail. Même si ces paroles peuvent paraître bien peu à dire, elle m'ont toujours touchée, mais surtout rassurée et elles m'ont permis de reprendre confiance en moi lorsque parfois (souvent) j'en avais besoin.

Je remercie aussi Ronan, avec qui l'on passe toujours de bons moments, et qui a souvent gentiment accepté de m'aider s'il le pouvait.

De même je remercie Rebecca avec qui je trouve toujours une bonne raison de rire, et avec qui j'ai passé quelques bons moments en dehors du laboratoire, que je n'oublie pas.

Merci aussi à Arnaud et Ulysse, pour tous leurs conseils en statistiques au cours de ma thèse, et nos nombreuses discussions, toujours très intéressantes.

Et enfin, je n'oublie pas Geneviève, qui entretient notre veille scientifique dans l'équipe et qui m'a permis de lire de nombreuses publications en lien avec mes projets de thèse, à côté desquelles je serais sûrement passée si elle ne me les avait pas envoyés.

J'ai également à coeur de remercier certains anciens membres de l'équipe BM2A, qui ont aujourd'hui volé vers d'autres horizons. Je garde en mémoire Loïs, pour toute l'aide qu'il m'a apportée dans mes débuts sous R et avec les statistiques. Je le remercie vraiment pour toute la patience dont il a fait preuve, et tout le temps qu'il a passé à m'expliquer et me réexpliquer les outils statistiques qu'il utilisait. Je n'ai jamais autant progressé en mathématiques qu'à ses côtés. Et je n'oublie pas non plus Valérie, qui m'a en grande partie formée en Master, et qui m'a ensuite apporté son aide au début de ma thèse.

Par ailleurs, je tiens à remercier Andràs Paldi et Gaël Yvert pour avoir accepté de faire partie de mon comité de suivi de thèse et pour avoir parfaitement joué leur rôle. À chaque réunion, sans exception, ils ont permis de faire avancer mes travaux de thèse grâce à leurs remarques pertinentes, et ont toujours veillé à m'encourager.

Je tiens aussi à remercier particulièrement Andràs Paldi et tous les membres de son équipe, pour les nombreuses discussions très intéressantes et constructives que nous avons eu à propos de nos travaux respectifs et des projets en collaboration. Il est également important pour moi de souligner qu'il a toujours été très agréable de les voir et de passer quelques heures avec eux, autant scientifiquement qu'humainement.

Je remercie aussi Caroline Romestaing et Damien Roussel qui m'ont permis de réaliser toutes les expériences de respiration cellulaire. De plus, leurs conseils d'experts sur les travaux réalisés dans la deuxième partie de ma thèse m'ont été très précieux. Mais je tenais surtout à les remercier pour leur accueil qui a toujours été très chaleureux à chaque fois que je leur ai rendu visite.

Je remercie également Pierre Sujobert et Adriana Plesa pour l'enthousiasme dont ils ont fait preuve pour travailler avec nous dans le cadre de leur étude sur les myéلودysplasies. Ils m'ont permis de donner une toute nouvelle couleur à mes travaux, qui s'étendent maintenant à un cadre pathologique. Je les remercie donc pour m'avoir donné la possibilité de découvrir un autre milieu, qu'est celui de la clinique, ainsi que pour leur engouement contagieux pour ce projet à chacune de nos réunions, qui sont toujours très agréables.

Enfin, je souhaiterais remercier toutes les personnes du LBMC/CIRI/RDP qui nous ont chaleureusement accueillis lorsque nous sommes arrivés au LBMC. Et je remercie particulièrement toutes les personnes aux côtés de qui je travaille tous les jours et qui rendent la vie au laboratoire si agréable, je pense notamment à mes chers colocataires de la pièce de culture cellulaire. Grâce à toutes ces personnes je ne regrette jamais de m'être levée le matin pour aller au laboratoire. Par ailleurs, je reste admirative de la solidarité qui s'est créée entre les doctorants (anciens, présents et futurs) et je les remercie pour m'avoir permis de faire partie de cette grande cohésion. Grâce à eux, je ne me suis jamais sentie isolée et j'ai découvert de très belles personnes. Je ne vais malheureusement pas tous vous nommer, car si je le faisais mes remerciements occuperaient la moitié de mon manuscrit, mais chacun d'entre vous êtes uniques et j'espère vraiment que nos départs respectifs, qui se font de plus en plus fréquents, ne marqueront pas la fin de notre amitié.

J'aimerais à présent remercier toutes les personnes qui n'ont pas directement contribué à mes travaux de thèse, mais qui ont rendu cette aventure unique, et grâce à qui j'en suis ici aujourd'hui.

À commencer par mon petit noyau, qui ont fait chavirer ma vie. Je ne les nomme pas car je sais pertinemment qu'ils sauront se reconnaître sans hésiter une seconde. Les larmes me montent aux yeux lorsqu'il s'agit de leur dire à quel point ils comptent pour moi et à quel point je les remercie d'avoir croisé un jour mon chemin. Si je ne les avais pas rencontrés, l'idée de faire une thèse ne m'aurait même pas effleuré. Ils sont tellement incroyables et attachants que pour ne pas risquer de les perdre, j'ai donné tout ce que je pouvais pour pouvoir les suivre en master. Mais, étant tous brillants et intelligents, ils m'ont obligée à me surpasser pour, non seulement, les suivre, mais surtout atteindre leur niveau. Ainsi, sans même avoir eu le temps de m'en rendre compte, ils m'ont malgré eux propulsé en thèse. Je ne les remercierai jamais assez pour tout ce qu'ils ont fait pour moi et pour m'avoir permis d'en être là aujourd'hui. Ils ont cru en moi, lorsque j'en étais incapable, ils m'ont aidée à gravir les échelons jusqu'à la thèse, et m'ont ensuite soutenue. S'enthousiasmant de chaque petite réussite et demeurant optimistes malgré les échecs, j'ai pu compter sur eux pour m'encourager ou me remonter le moral tout au long de ma thèse, qui n'aurait pas été aussi supportable s'ils

n'avaient pas été là. Je les remercie donc d'avoir non seulement contribué à cette histoire, mais surtout de faire partie de ma vie aujourd'hui.

Parmi ces amis, j'aimerais cependant adresser une mention plus spéciale à Nico, avec qui tout a commencé et grâce à qui j'ai connu toutes ces personnes, aussi exceptionnelles que lui. Il m'a appris à prendre confiance en moi, à me surpasser et à me battre pour ce que je voulais. Je le remercie pour tout ce qu'il a supporté en étant à mes côtés pendant ces années de thèse, surtout à la fin, sans jamais se plaindre. Je suis aujourd'hui émue par la sincérité, la gentillesse et l'affection qu'il me témoigne encore chaque jour. Malgré toutes ces épreuves, il n'a cessé de me soutenir, de s'occuper de moi, de recharger mes batteries lorsque j'en avais besoin, et il a fait preuve de beaucoup de force et de patience pour me relever dans les moments difficiles. Nico, je te remercie d'avoir cru en moi et de n'avoir jamais cessé de le faire.

D'autre part, je souhaiterais remercier ma meilleure amie, alias "ma copaindre", pour son soutien permanent au cours de ma thèse, et pour toute sa bienveillance envers moi au quotidien. Bien qu'il devienne à présent difficile de compter les années, je suis heureuse que notre amitié, elle, ne s'use pas (contrairement à nous... Et oui ma copaindre, il faut dire ce qui est. Entre tes chats, mes tisanes, nos bouillottes et nos petits plaid, à juste 30 ans, on est mal parties ...). Je lui suis reconnaissante d'avoir toujours supporté mes choix sans jamais me juger et d'avoir toujours répondu présente à mes appels de détresse. Elle est aujourd'hui un repère pour moi, aussi important que les membres de ma famille, et elle représente un véritable pilier dans ma vie. Ma copaindre je te remercie d'être entrée dans ma vie et d'en faire encore partie depuis toutes ces années.

Par ailleurs, je remercie tous les amis qui ont donné du rythme à cette thèse, autant au sens figuré qu'au sens réel du terme.

Monika et Olivier, je suis extrêmement fière d'être la marraine de votre fils, merci de m'avoir fait suffisamment confiance pour me confier cette responsabilité. Je vous remercie aussi pour tous les supers moments passés ensemble, qui m'ont à chaque fois permis de me vider la tête et de repartir plus légère. Mes salseros adorés, Aurélie et Fred, je vous remercie pour toutes ces soirées salsas, et autres, pendant lesquelles je me suis défoulée et pendant lesquelles j'ai relâché la pression, grâce à vous.

Mes super diététiciennes, Chloé et Barbara, merci pour toutes nos petites sorties culinaires, qui me réchauffent le coeur à chaque fois, ainsi que pour tous vos encouragements et félicitations par rapport à mon parcours en thèse, vous êtes tellement adorables.

Max, merci pour toutes ces rigolades avant et pendant la thèse, que je n'oublie pas. Aurel, Toto, et JB, qui êtes progressivement devenus mes amis aussi, je vous remercie pour tous les bons moments, aux frissons garantis, que l'on a passés ensemble à Lyon comme à Paris.

Val, je ne sais pas si je dois te remercier pour cela, mais merci quand même pour toutes ces péripéties dont tu nous régales à chaque fois et dont on continue de rire encore aujourd'hui. En tout cas je garde de très très bons souvenirs de Belle île en mer et de Papillorama, et j'espère sincèrement que nous referons d'aussi belles balades avec toi. Enfin, comme la thèse est pleine de surprise, j'ai eu la chance de rencontrer un petit groupe de personnes à la fin de ma thèse, à qui je me suis rapidement attachée. Je dirais qu'ils sont ma bouffée d'air internationale, car avec eux je n'ai pas besoin de partir parcourir le monde, je le découvre à travers eux. Je les remercie donc pour tous ces moments à leurs côtés où j'ai eu l'impression de faire le tour du monde en quelques heures. También quiero añadir un comentario especial por Yeraldinne, en quien encuentro una muy buena amiga. Estoy muy feliz de conocerla y espero que eso es el principio de una linda amistad.

Evidemment, j'ai gardé le meilleur pour la fin. Ainsi, pour conclure en beauté je remercie toute ma famille, qui m'a supportée tout au long de ma thèse sans relâche. Merci pour tout l'intérêt et la curiosité dont vous avez sans cesse fait preuve par rapport à mes travaux de thèse. Vous m'avez toujours montré que vous croyiez en moi, ce qui m'a donné la force de me battre, car il n'y a pas de meilleurs encouragements que ceux de sa propre famille. Votre soutien et vos nombreux compliments m'ont toujours énormément émue, et ont représenté un vrai moteur dans l'avancée de ma thèse. Je vous remercie donc d'avoir été mes premiers fans au cours de cette thèse et de m'avoir soutenue jusqu'au bout.

Je souhaite cependant remercier plus spécialement mes parents et mon petit frère, bien que je n'ai pas de mots pour exprimer à quel point je vous aime et je vous suis reconnaissante. Vous êtes ma plus grande motivation et la raison pour laquelle j'ai eu envie de me surpasser chaque jour. J'ai face à moi trois battants, qui m'ont appris et démontré que dans la vie il ne faut jamais baissé les bras. C'est en suivant leurs exemples qu'en partant pourtant de bien bas je suis arrivée à présent aux portes du diplôme de docteur en Biologie. Ainsi, voir leur yeux briller de fierté aujourd'hui est au final mon plus beau succès, et ce qui me rend la plus heureuse. Leurs nombreuses félicitations bien avant l'heure, leurs petits mots réconfortants, leurs conseils lorsque j'étais perdue et leurs encouragements m'ont donné la force et l'envie d'avancer malgré les difficultés et de donner toujours plus. Je suis donc plus qu'heureuse de pouvoir aujourd'hui leur offrir et leur dédier, à tous les trois, cet aboutissement de ma vie, qui pour moi représente beaucoup, mais qui pour eux, je le sais, représente une montagne.

Table des matières

Table des matières	8
Table des figures	11
1 Introduction	12
1.1 Etudier la stochasticité de l’expression des gènes (SEG) et le métabolisme pour mieux comprendre le processus de différenciation	12
1.2 Le processus de différenciation	13
1.2.1 La vision classique du processus de différenciation cellulaire	13
1.2.2 Caractéristiques du processus de différenciation érythrocytaire	14
1.2.3 La différenciation cellulaire à l’ère de la biologie systémique	16
1.3 Stochasticité d’expression génique (SEG) et différenciation cellulaire .	19
1.3.1 La SEG	19
1.3.1.1 Définition, origines et rôle de la SEG	19
1.3.1.1.1 La SEG, un hasard contraint	19
1.3.1.1.2 Causes de la SEG	22
1.3.1.1.3 La SEG dans les processus biologiques .	25
1.3.1.2 Quels outils pour mesurer quantitativement la stochasticité?	27
1.3.1.2.1 Etat des lieux des différentes méthodes d’analyse de la variabilité de l’expression génique	27
1.3.1.2.2 L’entropie de Shannon	31
1.3.2 La SEG, une force motrice de l’engagement en différenciation ?	34

1.4	Métabolisme et différenciation cellulaire	37
1.4.1	Présentation du métabolisme cellulaire eucaryote	37
1.4.1.1	Introduction au métabolisme	37
1.4.1.2	Co-facteurs métaboliques et réservoirs énergétiques	38
1.4.2	Le métabolisme du glucose	39
1.4.2.1	Catabolisme et anabolisme du glucose	39
1.4.2.2	Les voies énergétiques de la glycolyse	42
1.4.2.2.1	La glycolyse anaérobie	42
1.4.2.2.2	La phosphorylation oxydative	46
1.4.3	Carrefours et convergences des différentes voies métaboliques	50
1.4.4	Le switch métabolique, moteur de la différenciation?	52
1.5	Différenciation cellulaire, SEG et métabolisme glycolytique	56
2	Analyse de la variabilité de l'expression génique en cellule unique au cours du processus de différenciation	60
2.1	Article 1 : Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process	60
2.1.1	Principaux résultats de l'article 1	60
2.1.2	Principales conclusions de l'article 1	63
2.1.3	Article 1	64
2.2	Article 2 : Automated cell cycle and cell size measurements for single-cell gene expression studies	100
2.2.1	Principaux résultats de l'article 2	100
2.2.2	Principales conclusions de l'article 2	101
2.2.3	Article 2	102
2.3	Article 3 : Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment	110
2.3.1	Principaux résultats de l'article 3	110
2.3.2	Principales conclusions de l'article 3	112

2.3.3	Article 3	113
3	Rôle de LDHA et du métabolisme du glucose dans l’auto-renouvellement et la différenciation de progéniteurs érythrocytaires aviaires	137
3.1	Article 4 : Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements	138
3.1.1	Principaux résultats de l’article 4	138
3.1.2	Principales conclusions de l’article 4	139
3.1.3	Article 4	140
4	Discussion et perspectives	168
4.1	Mise en évidence expérimentale d’un scénario commun à différents processus de différenciation	168
4.2	Mieux comprendre le rôle de la variabilité de l’expression génique dans la différenciation	173
4.3	Remaniements du métabolisme glycolytique au cours de la différenciation érythrocytaire	176
4.4	Implications d’autres voies métaboliques dans la différenciation érythrocytaire	178
4.5	Intégration du pic énergétique dans la chronologie des évènements moléculaires identifiés au cours de la différenciation	180
4.6	Communications entre le réseau de régulation de gènes et le réseau métabolique	182

Table des figures

1	Représentation schématique de l'érythropoïèse selon la vision classique des processus de différenciation	15
2	Vision systémique du processus de différenciation	18
3	Distribution de l'expression d'un gène rapporteur dans une population de cellule clonale	21
4	Deuxième loi de la thermodynamique et entropie de Shannon	32
5	Représentation des cellules au cours du processus de différenciation érythrocytaire dans l'espace multidimensionnel des gènes	36
6	Étapes majeures de la glycolyse	40
7	Voies énergétiques du pyruvate	42
8	La glycolyse anaérobie	44
9	Le cycle de Cori	46
10	Principales étapes du cycle de krebs	47
11	La phosphorylation oxydative	48
12	Représentation des intersections et convergences des grandes voies métaboliques	52
13	Vision globale des destinées du pyruvate	53
14	Implications possibles du remaniement métabolique intégré dans le paysage épigénétique de Waddington au cours de la différenciation	58
15	Orchestration des événements moléculaires au cours de la différenciation érythrocytaire	169
16	Superposition des événements métaboliques et d'expression génique au cours de la différenciation érythrocytaire	181
17	Causalité circulaire entre le métabolisme, l'épigénétique et l'expression génique .	186

1 Introduction

1.1 Etudier la stochasticité de l'expression des gènes (SEG) et le métabolisme pour mieux comprendre le processus de différenciation

À l'échelle de l'organisme il est nécessaire de maintenir un équilibre entre les cellules s'auto-renouvelant, afin de générer ou régénérer les tissus vitaux, et les cellules matures fonctionnelles, pour assurer le maintien en vie de l'organisme. Des perturbations de cet équilibre peuvent avoir des conséquences pathologiques très fortes, comme le développement de cancers [1], ou de malformations congénitales [2]. Afin de maintenir cette balance, la capacité d'auto-renouvellement et de différenciation des cellules souches doit être finement régulée, conduisant ainsi à l'une des questions fondamentales, comment une cellule prend-elle la décision de se différencier ? De nombreuses études ont tenté de disséquer les mécanismes sous-jacents à la prise de décision cellulaire, apportant ainsi déjà quelques éléments de réponse à cette problématique. Cependant ces résultats demeurent encore insuffisants pour comprendre entièrement et maîtriser ce processus. L'étude de la stochasticité de l'expression génique (SEG), en fort accroissement ces dernières années, a permis de dévoiler de nouveaux éléments et ainsi d'apporter de nouvelles pièces au puzzle. Comme il sera expliqué dans les sections suivantes, l'effet de la variabilité de l'expression génique inter-cellulaire est de plus en plus mise en évidence dans les processus biologiques, notamment les prises de décision cellulaire telles que dans le processus de différenciation. D'autre part, outre le développement des nouvelles technologies de biologie moléculaire, des études récentes remettent le métabolisme, une branche ancienne de la biologie, au cœur des processus biologiques. Aujourd'hui, ces deux approches semblent converger et se développer conjointement afin de mieux comprendre le fonctionnement des processus biologiques. En effet, la simple observation de profils métaboliques différents entre les cellules capables d'auto-renouvellement et les cellules différenciées, conduisent à penser que les remaniements métaboliques sont impliqués dans les processus de différenciation, tout comme la SEG. Il est alors possible de se demander lequel de la SEG ou de métabolisme, est réellement la force motrice permettant aux cellules de quitter leur

état d'auto-renouvellement et de basculer dans un processus de différenciation. Cependant, comme nous le verrons au fil de cette introduction, tous deux semblent représenter des forces motrices du processus de différenciation. Ainsi, bien qu'ils soient considérés comme différents en terme de domaine d'étude, ils sont liés intrinsèquement, notamment dans le cadre de la prise de décision cellulaire.

1.2 Le processus de différenciation

1.2.1 La vision classique du processus de différenciation cellulaire

Les processus de différenciation ont lieu au cours du développement embryonnaire, dans le but de générer un organisme multi-cellulaire, ainsi que pendant la vie post-natale pour régénérer les tissus adultes. La différenciation peut être simplement décrite comme le processus permettant à une cellule immature de se spécialiser en une cellule adulte fonctionnelle. Les cellules immatures sont capables de se diviser symétriquement ou asymétriquement. La division asymétrique aboutira à deux cellules immatures filles, il s'agit alors d'auto-renouvellement, ou à deux cellules plus différenciées [3]. Quant à la division asymétrique, elle donnera lieu à une cellule immature et une cellule plus différenciée [4]. Les mécanismes moléculaires sous-jacents à ces divisions sont étudiés depuis plusieurs années afin de comprendre de façon plus précise comment les cellules prennent la décision de se différencier. Cependant les instruments expérimentaux et les outils de biologie moléculaire ont représenté une limite à l'avancée des connaissances scientifiques.

Le processus de différenciation a donc d'abord été étudié à l'échelle du tissu cellulaire, supposant que toutes les cellules fonctionnent à l'identique et participent de façon homogène au mécanisme physiologique final. Les processus de différenciation ont alors été décrits selon des schémas très précis, linéaires et hiérarchiques, sur la base de moyennes, générées à partir de populations de cellules. D'un point de vue tout à fait déterministe, il a été décrit que ces transformations successives étaient déclenchées par des stimuli extérieurs et s'accompagnaient d'un changement programmé du pattern d'expression génique. Plus exactement, il a été suggéré que l'engagement des cellules en différenciation était gouverné par des gènes nommés les gènes maîtres, déclenchant toutes les voies de régulation en cascade nécessaires

à la différenciation des cellules [5]. Ce paradigme, très discuté aujourd'hui, suggérait ainsi que pour chaque phénotype correspondait un génotype très précis dit "programmé" par les gènes maîtres [6, 7]; la plupart de ces gènes étant des facteurs de transcription ou des gènes suppresseurs de tumeurs [8, 9]. Ainsi, de façon schématique, les processus de différenciation cellulaire ont tout d'abord été définis tels des diagrammes hiérarchiques, linéaires ou branchés, dans lesquels chaque flèche unidirectionnelle représente une division conduisant à un nouveau phénotype plus différencié et caractérisé par un ou plusieurs gènes maîtres.

1.2.2 Caractéristiques du processus de différenciation érythrocytaire

Dans ce manuscrit, nous nous intéresserons particulièrement au processus de différenciation hématopoïétique, et plus spécifiquement à la branche de l'érythropoïèse. L'hématopoïèse permet de fournir à l'organisme toutes les cellules du tissu sanguin. De façon simplifiée, le schéma classique de la différenciation hématopoïétique est représenté par un diagramme hiérarchique branché, dans lequel une cellule souche hématopoïétique peut tout d'abord fournir un progéniteur lymphoïde ou myéloïde dont les capacités d'auto-renouvellement sont diminuées. Ces progéniteurs peuvent ensuite se différencier progressivement en cellules matures spécialisées. Les érythrocytes sont produits à partir de progéniteurs myéloïdes pouvant également générer des mastocytes, basophiles, éosinophiles, neutrophiles, mégacaryocytes et monocytes. Ainsi le processus de différenciation érythrocytaire, tel qu'il a été dépeint auparavant, peut s'illustrer par une succession d'étapes spécialisant progressivement les cellules filles dont le potentiel de différenciation se restreint un peu plus à chaque division asymétrique ou symétrique (donnant lieu à deux cellules un peu plus différenciées), sans retour possible en arrière. Différents stades cellulaires ont été décrits et nommés au cours de l'érythropoïèse, principalement caractérisés par l'expression de gènes maîtres particuliers (figure 1).

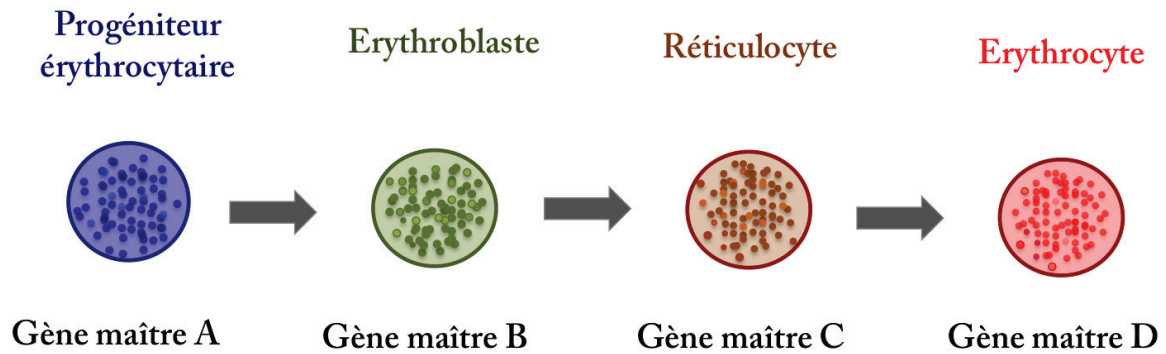


FIGURE 1 – Représentation schématique de l'érythropoïèse selon la vision classique des processus de différenciation

Ce schéma est un exemple de la vision classique de la différenciation reposant sur l'analyse de données acquises à l'échelle des populations. L'érythropoïèse, comme illustrée ici, est donc décrite comme un enchainement d'évènements de différenciation linéaire dans lequel toutes les cellules agissent de la même façon, comme un ensemble coordonné et homogène (représenté sur le schéma par l'encerclement des cellules). Il a ainsi été décrit que ce processus est programmé par le génotype et que chaque étape de la différenciation est gouvernée par quelques gènes qualifiés de gènes maîtres. Ainsi à chaque stade de différenciation correspond un phénotype, représenté sur ce schéma par la couleur, et déterminé par un gène maître (par exemple le gène maître B pour l'érythroblaste).

En outre, différents paliers ont également été définis parmi les progéniteurs (stades CFU-GEMM, BFU-e, CFU-e). Ces stades précoces de la différenciation érythrocytaire permettent notamment de distinguer les cellules progénitrices dont les possibilités de différenciation sont encore multiples, des cellules dont le potentiel de différenciation est restreint au lignage érythrocytaire. D'un point de vue phénotypique, l'érythropoïèse se caractérise notamment par une diminution de la taille des cellules et une expulsion du noyau au cours des derniers stades du processus de différenciation [10, 11]. De plus, lors des toutes dernières divisions précédents l'érythrocyte mature, les mitochondries sont éliminées, notamment *via* un mécanisme d'autophagie [12]. En effet, la fonction principale d'un érythrocyte est de transporter l'oxygène, qui se fixe sur les molécules d'hémoglobine, pour l'acheminer jusqu'aux différents tissus et organes à travers le système sanguin. Il ne représente alors plus qu'une enveloppe contenant une grosse quantité d'hémoglobine lui permettant d'assurer sa fonction de transporteur de l'oxygène. Cependant, ces caractéristiques ne s'appliquent pas aux érythrocytes des oiseaux,

dont le noyau n'est pas expulsé et pour lesquels les mitochondries ne sont pas phagocytées. En effet, les érythrocytes aviaires conservent leur noyau, qui se condensent de façon importante, empêchant ainsi les gènes de s'exprimer, de la même façon que si le noyau avait été expulsé [13]. De plus leurs mitochondries ne sont pas éliminées et demeurent fonctionnelles [14]. Par ailleurs, le régulateur principal de l'érythropoïèse connu est l'hormone érythropoïétine (Epo), dont son récepteur est EpoR [15]. De plus, l'érythropoïèse s'accompagne d'une diminution de l'expression de gènes impliqués dans le maintien en auto-renouvellement des progéniteurs, comme *sca-2* [16] et *OSC* [17]. À l'inverse, les gènes codant pour les chaînes de l'hémoglobine, tels que la *betaglobin* voient leur expression augmenter au cours de ce processus de différenciation [18], conduisant ainsi à une accumulation progressive d'hémoglobine.

1.2.3 La différenciation cellulaire à l'ère de la biologie systémique

L'évolution des méthodes d'analyse de l'expression des gènes, devenues haut débit, ont tout d'abord permis de démontrer que les gènes ne fonctionnaient pas de façon indépendante, mais bien comme un ensemble. Ces expériences ont également mis en évidence la complexité des relations entre gènes sous-jacente aux phénotypes, conduisant à nuancer le concept de programme génétique. Les processus cellulaires, tels que la différenciation, ne sont donc plus décrits par quelques gènes indépendants, tels que les gènes maîtres comme il était entendu dans le passé, mais plutôt comme un réseau dynamique de régulation de gènes [19, 20]. Par la suite, la démonstration expérimentale de l'existence de la SEG par Michael B. Elowitz, citée en amont [21], conjointement aux expériences de J. M. Levsky publiées au même moment, ont éveillé les esprits scientifiques, premièrement, sur l'importance des analyses à l'échelle de la cellule unique et deuxièmement sur l'existence d'une SEG dans les processus biologiques. Ces publications signent ainsi le début d'une nouvelle ère de la Biologie, celle de la cellule unique et de la biologie systémique. Ce champ d'étude évolue aujourd'hui dans un environnement interdisciplinaire reposant sur la collaboration de chercheurs qui proviennent de différents univers scientifiques tels que des biologistes, physiciens, mathématiciens, ou encore philosophes. Ces interactions ont pour objectif de rassembler toutes les compétences nécessaires à la compréhension des fonctionnements cellulaires, tels des systèmes complexes [22].

Dans un tel contexte, un phénotype peut être perçu comme l'état stationnaire d'un réseau

dynamique de gènes connectés, s'influençant les-uns-les-autres. Ce réseau est dynamique pour deux raisons fortement liées. Premièrement, la SEG diffuse dans le réseau, les interactions et régulations géniques ne sont donc pas complètement statiques même en état stationnaire. Par exemple, dans une population de cellules s'auto-renouvelant, le réseau se maintient dans le même état au cours du temps, mais l'expression d'un gène mesuré aux temps t et $t+1$ sera pourtant quelque peu différente [23]. Deuxièmement, lors d'un processus de différenciation, le taux d'expression de certains gènes va fortement varier. Le réseau est donc dynamique au cours du temps. Il est important de souligner que dans un contexte tel que la différenciation cellulaire le réseau ne change que par le niveau d'expression des gènes, et non pas dans sa structure [19]. Autrement dit, lors du processus de différenciation, les noeuds du réseau, représentés par les gènes, et les relations entre les gènes (activation ou inhibition) restent inchangés, mais le taux d'expression de ces gènes évolue de façon coordonnée, en fonction des influences que les uns ont envers les autres (figure 2).

Par ailleurs, nous pouvons noter que la dimension haut débit des nouvelles technologies permettant d'analyser le contenu moléculaire de cellules uniques a permis de mieux appréhender ces relations entre gènes et ainsi de modéliser des réseaux de régulation de gènes (RRG) dynamiques de plus en plus proches de la réalité biologique. En effet, l'abondance des données générées par les méthodes "omics" a permis d'alimenter et de faire évoluer ces différents modèles de réseaux de gènes. Cependant il existe encore de nombreuses boîtes noires que les outils mathématiques et expérimentaux d'aujourd'hui ne permettent pas encore de révéler.

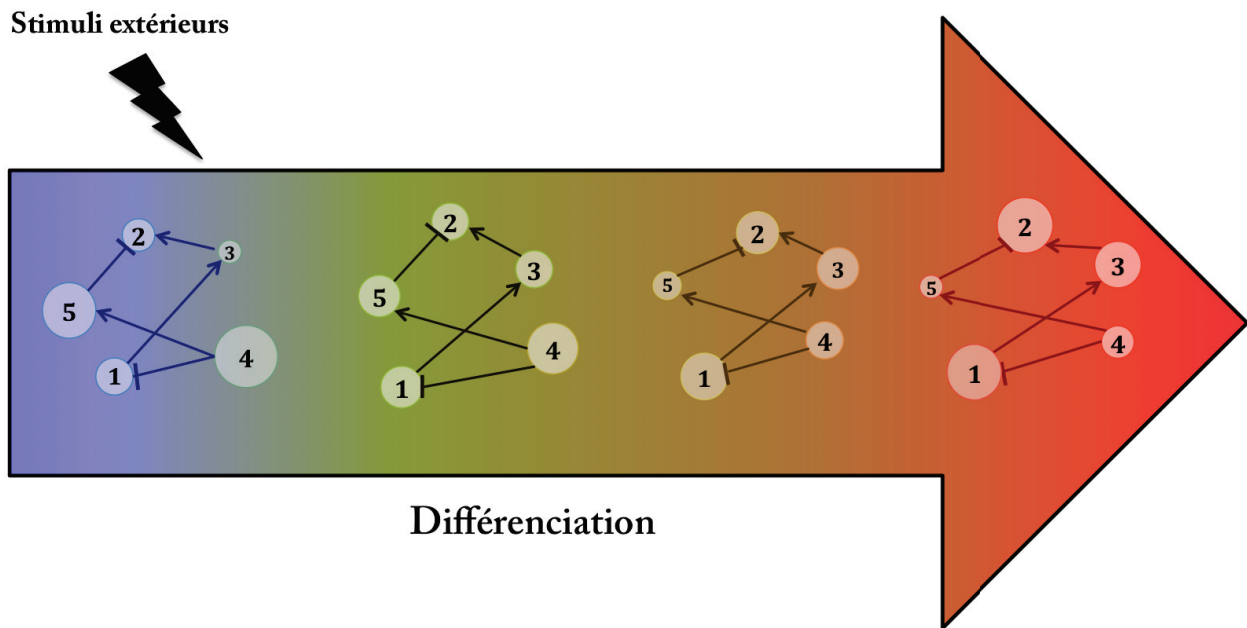


FIGURE 2 – Vision systémique du processus de différenciation

Ce schéma illustre l'évolution du réseau de régulation de gènes au cours du processus de différenciation, déclenché par des stimuli de l'environnement extra-cellulaire. Dans cet exemple, les gènes 1 à 5 représentent les noeuds du réseau, les flèches indiquent leurs relations, et la grosseur de leur encerclement décrit leur taux d'expression. Au cours de la différenciation la structure du réseau, correspondant aux gènes et à leurs relations ne varient pas. En revanche l'état du réseau, quant à lui, évolue au cours de la différenciation de part la variation de l'expression des gènes. Par exemple, dans le premier réseau illustré, qui pourrait représenter l'état immature, le gène 4 est très fortement exprimé et active l'expression du gène 5. Au cours du processus de différenciation, l'expression du gène 4 diminue progressivement, entraînant ainsi la diminution progressive de l'expression du gène 5.

Ainsi, la différenciation cellulaire, ne peut se résumer à une succession continue de transformations phénotypiques gouvernées par l'expression indépendante de quelques gènes hiérarchiquement "supérieurs" aux autres (figure 1). En effet, la convergence des nouvelles approches systémiques et de l'abondance des données d'expression de gènes à différentes échelles ont conduit à modifier cette vision aux traits déterministes, et à envisager la différenciation plutôt comme un processus stochastique gouverné par un RRG dynamique (figure 2). De cette nouvelle définition, différents modèles de différenciation ont alors émergés [24, 25, 26].

1.3 Stochasticité d'expression génique (SEG) et différenciation cellulaire

1.3.1 La SEG

1.3.1.1 Définition, origines et rôle de la SEG

1.3.1.1.1 La SEG, un hasard contraint

Avant même de s'intéresser à la stochasticité de l'expression des gènes (SEG), il est nécessaire d'énoncer ce qui se cache derrière le terme "expression" lui-même. Ce seul terme, d'utilisation courante et aux sens multiples dans le langage non-scientifique, définit en réalité un mécanisme multi-étapes et multi-échelles très complexe. Au cours de l'Histoire, la Biologie a avancé au gré des pensées/théories de scientifiques passionnés, mais s'est sans cesse heurtée aux contraintes technologiques [27]. A l'étude des populations, de l'anatomie puis de la physiologie, a succédé la biologie cellulaire, et la biologie moléculaire, grâce au développement des technologies et à leur évolution dans l'observation de l'infiniment petit. Je reviendrai dans les prochains paragraphes sur ces technologies récentes, qui continuent aujourd'hui encore de progresser. La biologie moléculaire marque, quant à elle, un tournant dans l'Histoire de la Biologie, en offrant enfin un accès aux "rouages" internes de l'unité qui forme et fait fonctionner un organisme, la cellule. De cette époque est née une définition forte de l'expression des gènes, le dogme central, proposé pour la première fois par Francis Crick en 1958. Ce dogme, basé sur les observations de l'époque, veut qu'un gène transmette une information dans un ARNm, qui à son tour la transmettra dans une protéine sous forme de code transmis par leurs séquences nucléotidiques, *via* les processus respectifs de transcription et de traduction. Cette formulation suggère ainsi, à l'image d'un circuit linéaire, qu'une information diffuse d'un gène à une protéine de façon séquentielle [28], laissant comprendre que tout phénotype résulterait des instructions directement transmises par les gènes aux protéines. Il s'agit ici d'une vision tout à fait déterministe dans laquelle les événements se succèdent de façon ordonnée, et dans laquelle une information transmise en amont renvoie à un résultat précis, telle la loi de cause à effet. Cette représentation de l'expression des gènes a ainsi conduit progressivement à l'idée de programme génétique qui est publiée pour la première fois en 1961 [29]. Ainsi, cette vision

déterministe de l'expression des gènes a rapidement été acceptée pour l'époque et est devenue un socle de connaissance de la génétique qui s'est transmis sur plusieurs générations, et sur lequel beaucoup de biologistes se sont reposés.

Dès les années 1980, le dogme central et la notion de programme génétique ont été controversés par certains chercheurs proposant une conception stochastique de l'expression des gènes [30]. Des indices de l'existence de stochasticité dans l'expression des gènes se sont progressivement infiltrés dans certains résultats expérimentaux d'analyse d'expression génique, dans lesquels la variabilité inter-cellulaire était omniprésente. En effet, il suffit par exemple d'observer la distribution de la fluorescence d'une protéine codée par un gène rapporteur, dans une population cellulaire clonale : la largeur de la courbe, représentant la variabilité du taux de protéines entre les cellules, est un reflet même de la SEG (figure 3). Dans un premier temps cette variabilité a majoritairement été qualifiée de nocive pour les processus étudiés et nombre de biologistes ont ardemment essayé de la réduire. Ainsi, la SEG a fait l'objet de diverses études, théoriques et expérimentales, mais a longtemps été débattue avant de faire la première page du journal *Science* en 2002, grâce aux expériences de Michael B. Elowitz et collègues qui disposaient d'outils moléculaires et instrumentaux plus appropriés que leurs confrères 30 ans plus tôt [21]. Dans l'article publié par Michael B. Elowitz, deux gènes rapporteurs fluorescents ont été introduits dans des bactéries issues d'une même colonie. L'observation de la fluorescence des bactéries a alors révélé que les protéines codées par les gènes rapporteurs n'étaient pas exprimées en quantité identique dans toutes les cellules, possédant pourtant le même fond génétique. Cette expérience est donc l'une des premières à démontrer clairement que l'expression des gènes est stochastique. Ainsi, la mise en évidence de la SEG a introduit une vision novatrice de l'expression des gènes.

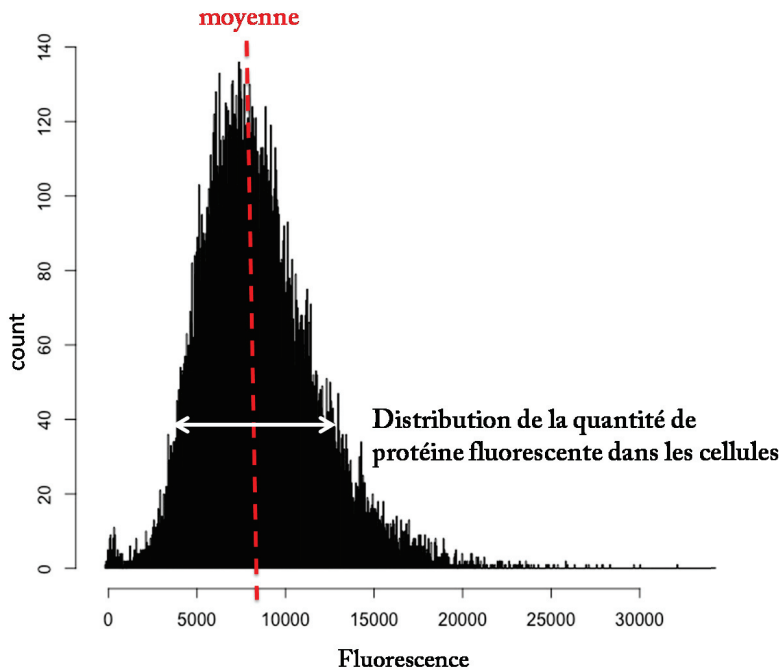


FIGURE 3 – Distribution de l’expression d’un gène rapporteur dans une population cellulaire clonale

Ce schéma représente une courbe décrivant la distribution d’une protéine fluorescente codée par un gène rapporteur, mesurée par cytométrie en flux dans un échantillon d’une population cellulaire d’origine clonale. Le taux d’expression de la protéine est représenté sur l’axe horizontal, alors que le nombre de cellule est représenté sur l’axe vertical. La dispersion autour de la moyenne, comme indiquée par la double flèche blanche, est le reflet de la SEG. Toutes les cellules n’expriment pas la protéine fluorescente au même taux, certaines cellules l’expriment fortement (côté droit de l’histogramme), alors que d’autres cellules l’expriment plus faiblement (côté gauche de la courbe).

Pour décrire cette nouvelle conception, le terme "hasard" est couramment substitué au mot "stochasticité", laissant concevoir que l’expression des gènes est soumise à un phénomène aléatoire. Cependant, il est difficile d’envisager dans un organisme multicellulaire parfaitement structuré, qu’un tissu ou organe puisse exercer une fonction très précise, tel que le transport de l’oxygène, si le comportement des cellules qui le composent dépend d’une expression totalement aléatoire et chaotique des gènes. Il convient alors de préciser que le terme "hasard" fait référence à une expression probabiliste, et non pas anarchique, des gènes. Il est alors possible de parler de hasard contraint ou de variabilité contrôlée, où l’état chromatinien pourrait par exemple représenter une forme de contrainte dynamique de cette stochasti-

cit  [31, 32, 33]. De fa on imag e et simplifi e, nous pouvons ainsi comparer la SEG   un jeu de d s. Les contraintes du d  sont li es   son architecture, compos e de six faces planes parfaitement  gales, offrant   chaque face exactement le m me taux de probabilit  de $1/6$ d' tre la face qui sera au dessus du d  lors d'un tir. De m me, chaque g ne dans une cellule aurait une probabilit  de $1/x$, o  x peut prendre plusieurs valeurs, d' tre exprim    un instant t . Une vision dite d terministe de l'expression des g nes repr senterait alors un exemple extr me de cette repr sentation probabiliste et s'apparenterait   un syst me binaire o  chacun des g nes poss derait une probabilit  uniquement  gale   1 ou 0 de s'exprimer.

Enfin, il est important de noter qu'une telle d finition de la SEG d'apparence conceptuelle simple cache en r alit  un m canisme extr mement plus complexe, de part le caract re notamment dynamique, mais  galement multi- chelles et multi- tapes, de l'expression des g nes.

1.3.1.1.2 Causes de la SEG

L'expression des g nes est un processus stochastique et dynamique dont les m canismes sous-jacents sont encore mal connus. Les  v nements induisant le caract re probabiliste de l'expression g nique sont probablement nombreux et restent en partie   identifier. Tout d'abord la probabilit  de l'expression des g nes peut provenir de la nature m me des proc d s de production des ARN et des prot ines. En effet, il a  t  largement d montr  computationnellement et exp rimentalement que la transcription et la traduction sont des processus au caract re pulsatile. Le random telegraph, propos  en 2005, mod lise le m canisme d'expression g nique en le r sumant   deux  tats, "on" et "off", entre lesquels le promoteur oscille al atoirement, conduisant ainsi   l'activation (on) ou l'inhibition (off) du g ne [34, 35, 36]. Le switch al atoire entre ces deux conditions donne alors lieu   des pics de production d'ARN, commun ment appel s "burst". Ces pics sont g n r s de fa on stochastique et permettent de produire un nombre al atoire de mol cules ; ils sont ainsi d crits selon deux param tres, leur fr quence et leur taille. L'observation du comportement de mol cules en direct dans des cellules vivantes individuelles a permis d'apporter une preuve directe du fonctionnement pulsatile al atoire et dynamique de la transcription [36]. De m me de nouvelles techniques d'imagerie ont contribu    montrer que la traduction, comme la transcription, s'effectue de

façon pulsatile et aléatoire [37, 38]. Ainsi la production d'ARNm et de protéines sont des procédés dont l'occurrence et l'abondance sont stochastiques, participant de fait à la SEG et aux variations d'expression génique observées entre les cellules d'une population possédant le même fond génétique. Nous pouvons alors nous interroger à propos des mécanismes conduisant au phénomène de burst. Un grand nombre d'évènements moléculaires semblent pouvoir participer à la génération de pics de transcription et de traduction.

La transcription étant un mécanisme soumis à de nombreuses régulations, divers éléments pouvant être à l'origine des bursts transcriptionnels ont été suggérés. Le procédé par lequel un facteur de transcription se fixe à sa région promotrice est l'un des plus récemment étudiés grâce à l'émergence d'outils adaptés à sa caractérisation. En effet, les techniques d'imagerie de molécule unique ont permis d'observer et de mesurer la fréquence de ce mécanisme dans une cellule vivante [39]. Il semblerait cependant que la cinétique d'assemblage-désassemblage des facteurs de transcription soit bien plus rapide que l'occurrence des pics de transcription [40]. En revanche, d'autres études démontrent que les facteurs de transcription évoluent selon une méthode de recherche essai-erreur, impliquant un grand nombre de fixations non-spécifiques, avant de se fixer spécifiquement à leur région cible [41]. La fixation spécifique d'un facteur de transcription sur son promoteur serait donc espacée de plusieurs interactions non-spécifiques [42]. Il est alors possible que malgré la rapidité de la cinétique de fixation du facteur de transcription à l'ADN, peu de ces contacts protéine-ADN soient spécifiques et conduisent à la production d'ARNm. De façon plus générale, les récentes études menées à l'échelle de la molécule unique suggèrent que les caractéristiques biochimiques, structurales et dynamiques [43] des interactions ADN-protéines ainsi que la concentration en facteur de transcription, souvent peu nombreux [44, 45], participeraient au phénomène de bursts transcriptionnels et, de fait, à la SEG [46, 47]. Néanmoins, de part la nouveauté des technologies utilisées ces résultats sont encore pionniers dans le domaine. La dynamique de fixation des facteurs de transcription aux régions promotrices devra donc être mieux caractérisée pour démontrer son implication directe dans la génération de bursts. D'autre part, il a été montré que le surenroulement de l'ADN, crée par les ADN et ARN polymérase lors de l'ouverture de la double hélice, serait probablement l'une des principales sources des bursts transcriptionnels [48, 49]. De plus, la durée et la fréquence d'arrêt de l'ARN polymérase au cours du

processus d'élongation de la transcription ont également été proposés mais demeurent encore peu étudiés, dû aux contraintes techniques [50]. Seulement récemment, une étude a démontré que les interactions entre les ARN polymérases sur la séquence d'ADN, pourraient être à l'origine de ces bursts de transcription [51]. Parmi les mécanismes ayant aussi retenus l'attention, le positionnement des nucléosomes [52], les modifications épigénétiques [53, 54, 55], la position des gènes sur les chromosomes [56], la configuration et le remodelage de la chromatine [57, 58] peuvent participer à la génération de bursts. Par ces exemples, il semble évident que l'environnement épigénétique et l'organisation spatiale de la chromatine jouent un rôle important dans la variabilité de l'expression génique. Il est alors possible d'envisager que ces évènements puissent intervenir notamment au niveau du contrôle de l'inhérente stochasticité du processus d'expression génique.

Par ailleurs, l'ensemble des réactions biochimiques se produisant dans les processus de régulation de l'expression génique, tel que la transcription et la traduction, peuvent par nature participer à la SEG. En effet, les molécules impliquées dans ces réactions sont des entités discrètes qui dans de petits systèmes tels qu'une cellule peuvent être en nombre très limité. Ainsi la fluctuation de quelques unités pour une molécule abondante n'aura que peu de conséquences, en revanche une telle variation entraînera des effets significatifs pour une molécule peu représentée, telle que certains facteurs de transcription. Le mécanisme de traduction, par exemple, nécessite le recrutement d'un complexe protéique ainsi que d'unités ribosomiques autour de l'ARNm. D'un point de vue biochimique, les interactions entre tous ces éléments moléculaires, dont la concentration diffère entre les cellules et varie de façon dynamique, sont probabilistes. En effet, le faible nombre d'ARNm ainsi que les potentielles rencontres des protéines du complexe ou des ribosomes avec d'autres entités cellulaires, peuvent ralentir la dynamique de rassemblement et conduire à une production stochastique des protéines. Le processus de traduction présente donc un caractère aléatoire pouvant être impliqué dans la SEG [27, 59].

Enfin, les réactions moléculaires impliquées dans la régulation de la machinerie transcriptionnelle font intervenir des enzymes ou encore des métabolites dont l'expression et l'activité dépendent de régulations métaboliques. L'Acétyl CoA est par exemple un co-facteur essentiel impliqué dans la modification des histones, régulateurs essentiels du niveau de compaction

de la chromatine. Il est produit par les mitochondries dont le nombre et l'activité varie. Il est ainsi probable que des fluctuations de la concentration en Acétyl CoA dans la cellule puisse représenter une source de la SEG [60]. Le lien entre métabolisme et SEG étant de plus en plus étudié, il sera développé plus en détail dans la dernière section de l'introduction.

1.3.1.1.3 La SEG dans les processus biologiques

Il est aujourd'hui clairement démontré que les cellules d'une population, même clonale, présentent des différences d'expression génique. L'émergence des nouvelles technologies à l'échelle de la cellule unique et de la molécule unique ont permis de quantifier et de caractériser plus en détail cette variabilité manifeste. Bien qu'elle ait été longtemps négligée [61], de nombreuses études sur la stochasticité ont aujourd'hui conduit à la reconnaissance de son influence bénéfique et nécessaire dans les processus biologiques [62]. Elle a par exemple été mise en évidence dans l'activation de la réponse immunitaire [63, 64], la reprogrammation cellulaire [65], le développement [66, 67] ou encore les processus de différenciation [55, 68]. Il a par exemple été montré que la stochasticité de l'expression du gène *spineless* dans les cellules photoréceptrices de l'oeil de drosophile, était à l'origine de la mosaïque ommatidienne de la rétine, requise pour la vision en couleur. En effet, l'expression stochastique de ce gène conduit à l'expression aléatoire des protéines opsines Rh3 et Rh4, responsables de la perception de la lumière. La protéine Rh3 détecte des longueurs d'onde courtes alors que Rh4 détecte plutôt des longueurs d'onde longues. Les cellules photoréceptrices exprimeront alors l'un ou l'autre type de protéines grâce au caractère probabiliste de l'expression de *spineless* conduisant ainsi à des ommatidies dites mosaïques.

Malgré sa mise en évidence dans différents processus biologiques, le rôle de la SEG est encore mal caractérisé et les études visant à mieux le comprendre ont cruellement souffert du manque d'outils moléculaires permettant de la manipuler expérimentalement. Ainsi, en 2014, un criblage de 1600 composés bioactifs a permis d'identifier plus de 80 molécules réduisant ou amplifiant la variabilité de l'expression de gènes rapporteurs en aval du promoteur LTR du virus HIV [69]. L'effet de certains composés a été testé sur la réactivation du virus en période de latence. L'un des plus intéressants résultats de ces travaux démontre que les composés amplifiant le bruit fonctionnent en synergie avec des activateurs transcriptionnels

de la réactivation virale (TNF ou Prostratin). En effet, la combinaison de ces inducteurs avec des molécules amplifiant la variabilité génique conduit à une amélioration de la réactivation du virus HIV latent. Ces travaux suggèrent donc que la variabilité de l'expression des gènes d'HIV joue un rôle dans la capacité du virus à se réactiver lors de période de latence. D'autre part, l'identification de ces composés associée aux rares molécules déjà connues et utilisées, telles que la 5-Azacytidine et la trichostatine A [32], sont des outils essentiels pour comprendre le rôle de la stochasticité dans les processus biologiques.

Néanmoins, pour palier au manque d'outils moléculaires permettant de moduler la variabilité de l'expression génique, divers modèles mathématiques ont été développés et exploités [70]. Pour illustrer ce propos, la modélisation de la réponse à un agent cytotoxique, TRAIL, a été combinée à la quantification en cellule unique de protéines impliquées dans l'apoptose, afin d'étudier le rôle de la stochasticité dans ce mécanisme [71, 72]. Ces études apportent une évidence supplémentaire de l'omniprésence de la variabilité dans les processus biologiques et de son importance dans les comportements cellulaires à l'échelle des populations. Il semblerait en effet que les cellules ne présentent pas toutes le même potentiel face à l'apoptose, certaines bénéficieraient d'un temps de résistance plus important. Cette hétérogénéité dans la réponse cellulaire, en présence d'un élément inducteur d'apoptose, serait en partie liée au caractère probabiliste du taux d'expression de certaines protéines apoptotiques. Dans le contexte des cancers, il semble évident que cette variabilité pourrait notamment apporter un avantage à certaines cellules face aux traitements chimio-thérapeutique. Les travaux sur la stochasticité dans les processus apoptotiques permettraient donc entre autre de mieux comprendre la résistance cellulaire face aux chimio-thérapies, et d'ouvrir la voie à de nouvelles pistes dans les traitements anti-cancéreux. En outre, le phénomène probabiliste de l'expression des gènes permettant à certaines cellules de mieux résister face à un inducteur de l'apoptose, laisse à penser que la SEG et la plasticité cellulaire pourraient être étroitement liées. Cette hypothèse a en effet fait l'objet de plusieurs études menées chez la levure [73], les bactéries [74] et les plantes [75], suggérant que la variabilité, notamment de l'expression des gènes de réponse au stress, confère une certaine plasticité cellulaire, avantageuse pour la population. Pour citer un exemple, il semblerait qu'elle puisse être à l'origine de la diversité de la réponse bactérienne observée dans la résistance aux antibiotiques [76, 77].

Enfin, le rôle de cette variabilité biologique a particulièrement été étudié dans le cadre de la prise de décision cellulaire. Beaucoup d'études défendent l'hypothèse selon laquelle une augmentation de la stochasticité serait nécessaire pour permettre à un système, tel qu'une cellule, de passer d'un état à un autre [24, 26, 25, 78]. La notion d'état renvoie ici à une condition phénotypique caractérisée par un profil d'expression génique, certes stochastique, mais stable. Dans le cas de la bactérie *Subtilis*, par exemple, il s'agit du passage de l'état végétatif à l'état compétent dans lequel les bactéries sont aptes à intégrer de l'ADN exogène [79]. D'après le concept exposé ci-dessus, une forte augmentation de la SEG permettrait d'induire la transition des bactéries vers l'état de compétence. L'acquisition de cet état repose principalement sur une boucle de rétrocontrôle positive du facteur de transcription *comK*, activant des gènes impliqués dans l'assimilation d'ADN extra-cellulaire [79]. Dans une population bactérienne le pourcentage d'individus développant cette capacité est restreint. Une diminution de la stochasticité de l'expression de *comK* tout en maintenant son taux d'expression moyen, conduit à une diminution drastique du nombre de bactéries compétentes [80]. Ces expériences démontrent donc bien que la SEG est nécessaire pour permettre aux bactéries de passer de leur état non-compétent à l'état compétent. Le concept exposé brièvement à travers cet exemple a particulièrement été étudié dans les processus de différenciation cellulaire, et sera plus amplement décrit dans les prochaines sections.

Ainsi, comme l'ont illustré les quelques cas détaillés dans cette section, la SEG peut jouer des rôles différents, à l'échelle individuelle au niveau de la régulation de l'expression génique (comme améliorer l'efficacité des activateurs transcriptionnels) [69, 81], et à l'échelle des populations en créant une variabilité inter-cellulaire pouvant être bénéfique pour le groupe (comme la survie ou la fonctionnalité de la population) [77, 80]. Il est donc évident que la SEG est présente et actrice dans un très grand nombre de processus biologiques.

1.3.1.2 Quels outils pour mesurer quantitativement la stochasticité ?

1.3.1.2.1 Etat des lieux des différentes méthodes d'analyse de la variabilité de l'expression génique

Les techniques de biologie moléculaire ont connu un essor remarquable au cours des quinze dernières années notamment dans le domaine des études à l'échelle de la cellule unique. Il

est possible de mesurer en cellule unique le taux de quelques protéines en parallèle depuis longtemps grâce à la cytométrie de flux. Cependant, il était jusqu’alors impossible de mesurer le taux d’ARNm dans une seule cellule à cause de la faible quantité de matériel biologique et du manque de technologies adaptées. Les techniques d’analyse transcriptionnelle telles que la qPCR et le RNASeq ont alors évolué vers des outils haut débit et multiplexes, permettant d’interroger le transcriptome de centaines de cellules de façon individuelle et simultanée [82, 83]. Parallèlement, les méthodes de génomique et d’épigénomique ont également connu un progrès important dans le domaine de la cellule unique. Nous possédons aujourd’hui pas moins de dix techniques différentes permettant d’étudier les régulations épigénétiques de cellules individuelles [84]. Parmi les plus connues, nous pouvons noter le ChiP-seq pour les interactions ADN-protéine, la DNase-seq et l’ATAC-seq en ce qui concerne la structure chromatidienne, et le HiC pour l’organisation tri-dimensionnelle des chromosomes [84] et références incluses). Par ailleurs, les contraintes techniques inhérentes à l’analyse des protéines n’ont pas permis une évolution aussi rapide dans le domaine de l’analyse du protéome à l’échelle de la cellule unique. Cependant, le développement récent de nouvelles méthodes, telles que le CITE-Seq semblent prometteuses et permettent déjà aujourd’hui d’accéder à une grande partie du compartiment protéique d’une seule cellule [85]. Par ailleurs, l’ensemble de ces technologies récentes, tout en permettant d’accéder à une échelle bien plus fine qu’est la cellule individuelle, introduit de nouveaux enjeux et questionnements concernant la maîtrise de la variabilité technique, déjà tant redoutée par le passé.

Il est important de souligner que l’essor des nouvelles technologies à l’échelle de la cellule unique a offert un terrain plus que favorable à l’étude de la SEG. En effet, ces approches permettent de quantifier la variabilité inter-cellulaire de l’expression génique, ouvrant néanmoins la voie à de nombreuses interrogations théoriques et techniques. Les variations non biologiques représentent l’un des premiers obstacles auquel les analyses de variabilité se sont confrontées. Nous savons que reproductibilité et significativité représentent deux points essentiels de la biologie expérimentale car toute expérience est soumise à des variations techniques et humaines sur lesquelles nous n’avons que peu de contrôle. Malgré l’évolution des technologies et du matériel expérimental, la variabilité générée par les aléas techniques perdurent. L’interprétation des expériences effectuées sur populations de cellules s’accommodent

de ces variations, notamment en multipliant les réplicats et en testant la significativité des différences au moyen de tests statistiques classiques. En revanche, cette démarche ne peut s'appliquer aux analyses d'expression génique à l'échelle de la cellule individuelle, où chaque cellule représente à la fois un réplicat technique et un réplicat biologique. De façon générale, la plupart des techniques de transcriptomique haut débit en cellule unique nécessitent plusieurs étapes de préparation expérimentale et une étape finale en puce microfluidique. Par conséquent, les variations finales mesurées incluent de manière évidente les variations techniques, qu'il est donc essentiel d'identifier. Ainsi, l'un des enjeux des analyses de la SEG réside dans le développement de méthodes permettant d'identifier le bruit expérimental et de l'extraire de la variabilité totale mesurée. En outre, le deuxième challenge de ce type d'analyses repose sur la nécessité d'identifier, mais aussi de créer, des métriques adaptées pour mesurer cette hétérogénéité biologique. À ce jour, nombre de stratégies et d'outils mathématiques ont été proposés, dont le coefficient de variation (ou variance normalisée) reste l'approche la plus classique et la plus basique mais ne permettant de mesurer qu'une variation par rapport à la moyenne.

Ainsi, à l'heure où les nouvelles approches de biologie moléculaire en cellule unique progressent et se multiplient, les outils mathématiques et statistiques sont plus que jamais sollicités. Une des stratégies répondant simultanément aux deux enjeux mentionnés ci-dessus consiste à intégrer dans un même algorithme l'identification du bruit technique et l'inférence de la "véritable" variabilité biologique. Il s'agit par exemple du modèle bayésien BASiC (Bayesian Analysis of Single-Cell Sequencing data), permettant de modéliser et séparer simultanément les variations expérimentales et l'hétérogénéité inter-cellulaire de l'expression des gènes, à partir de données de RNASeq obtenues en cellule unique (scRNASeq) [86, 87]. Le bruit technique est estimé à partir de données d'expression d'ARN exogènes, les spikes-in, introduits artificiellement en quantités égales dans chaque lysat cellulaire. Ainsi, l'algorithme déduit la variabilité liée aux fluctuations expérimentales des données de spikes-in, et infère l'hétérogénéité biologique des données d'expression de gènes endogènes selon un modèle poissonien.

Outre l'utilisation de modèles probabilistes, la variabilité entre cellules peut également être approchée par des mesures de distances. La plus courante est la distance Euclidienne,

particulièrement utilisée dans les méthodes de clustering comme l'analyse en composante principale. Cette statistique permet de calculer la distance entre deux objets dans un espace multidimensionnel. Autrement dit, dans notre contexte, cette mesure tend à déterminer quelle distance sépare deux cellules, dans un espace où chaque gène correspond à une dimension. Il devient alors possible de comparer la dispersion des cellules de deux populations différentes, pour le même set de gènes, et de statuer sur la variabilité inter-cellulaire dans le contexte étudié. D'autre part, une autre estimation de la dispersion a récemment été proposée pour mesurer cette hétérogénéité, il s'agit du centre médian [88]. Cette métrique consiste à déterminer dans l'espace multidimensionnel des gènes le centre médian, autrement dit la position, à laquelle la distance à chaque cellule analysée est la plus courte possible. Ainsi la dispersion de chaque cellule correspond à sa distance au centre médian, et la dispersion de la population cellulaire qu'elles composent est égale à la somme de ces distances. Enfin, de nouvelles métriques empruntées à d'autres champs d'études, telle que l'entropie, commencent progressivement à être utilisées et feront l'objet de la prochaine sous-partie (1.1.1.2.2. L'entropie de shannon).

En somme, différents outils mathématiques et computationnels ont été suggérés, et ne cessent encore d'émerger au rythme des progressions techniques. Beaucoup de ces stratégies visent à capturer et caractériser au mieux la variabilité de l'expression génique, cependant toutes ne répondent pas aux mêmes demandes. L'une des approches les plus utilisées est l'inférence de cette variabilité dans les modèles dynamiques d'expression de gène [59, 89]. L'inférence de l'hétérogénéité génique permet ainsi d'obtenir des modèles d'expression de gène et de réseaux de régulation d'expression génique (RRG) plus proches de la réalité biologique et par conséquent de mieux comprendre les interactions entre gènes [90, 91]. Cependant, étudier le rôle de l'hétérogénéité dans les processus biologiques dans un contexte tissulaire général, nécessite de pouvoir mesurer la variabilité produite par le système dynamique. Dans ce cadre, il semble préférable d'utiliser un outil renvoyant une mesure statistique de la variabilité, sous forme d'indice numérique, tel que les mesures de distances ou les indices de dispersion. Chacune de ces statistiques comporte ses avantages et aussi ses limites, mais aucune ne semble encore optimale. De fait, la ou les métriques idéales pour mesurer la variabilité biologique n'ont pas encore été clairement identifiées et demeurent à explorer. Dans cet objectif, des

métriques empruntées à d'autres champs d'études, telle que l'entropie, commencent progressivement à émerger et feront l'objet de la section suivante.

1.3.1.2.2 L'entropie de Shannon

La formule et l'interprétation de l'entropie varient en fonction des domaines et des contextes, menant à l'existence aujourd'hui de plusieurs types d'entropie. Sa plus ancienne définition est issue du domaine de la thermodynamique [92], la décrivant comme une mesure de la quantité de "désordre" dans un système. Ainsi, la valeur d'entropie est d'autant plus grande que les entités du système sont désordonnées. Le second principe de la thermodynamique énonce que le degré de désordre dans un système ne peut qu'augmenter [93]. Ce second principe sous-entend donc que dans l'univers les systèmes isolés évoluent spontanément vers l'arrangement qui a la plus grande entropie. Afin d'illustrer ce principe, l'analogie avec le désordre créé par l'humain est souvent de mise. En effet, avec le temps qui passe une personne tendra spontanément à engendrer du désordre dans une pièce (par exemple, un stylo posé sur une table, des piles de feuilles s'éparpillant en divers endroits, une tasse de café oubliée). Remettre de l'ordre dans cette pièce nécessitera en contrepartie un effort intentionnel d'organisation et un apport d'énergie, c'est à dire une libération de chaleur. Une autre illustration de ce concept est la répartition des molécules dans un contenant (figure 4A). Si toutes les molécules sont rassemblées en un endroit du récipient, l'entropie mesurée est faible car les molécules sont localisées en un espace précis. Avec le temps, ces éléments moléculaires vont ensuite se dissiper dans le récipient et se répartir de façon aléatoire et désordonnée. Dans ce cas l'entropie sera grande et le système thermodynamique aura atteint son état d'équilibre, donc un état stable. Certains systèmes, cependant, semblent défier ce second principe de la thermodynamique. Par conséquent, l'interprétation de la mesure d'entropie a été adaptée en fonction des contextes.

Elle a premièrement été reprise dans le domaine de l'informatique, par Claude Shannon [94]. Celui-ci adapta son interprétation dans le cadre de la théorie de l'information. Dans ce contexte l'entropie mesure la quantité d'information délivrée par le système. Ainsi plus le message fourni est uniforme et certain, plus l'entropie est minimale. Par exemple, une répétition du chiffre 1 piochés dans un intervalle de 1 à 9 produira une entropie minimale (figure

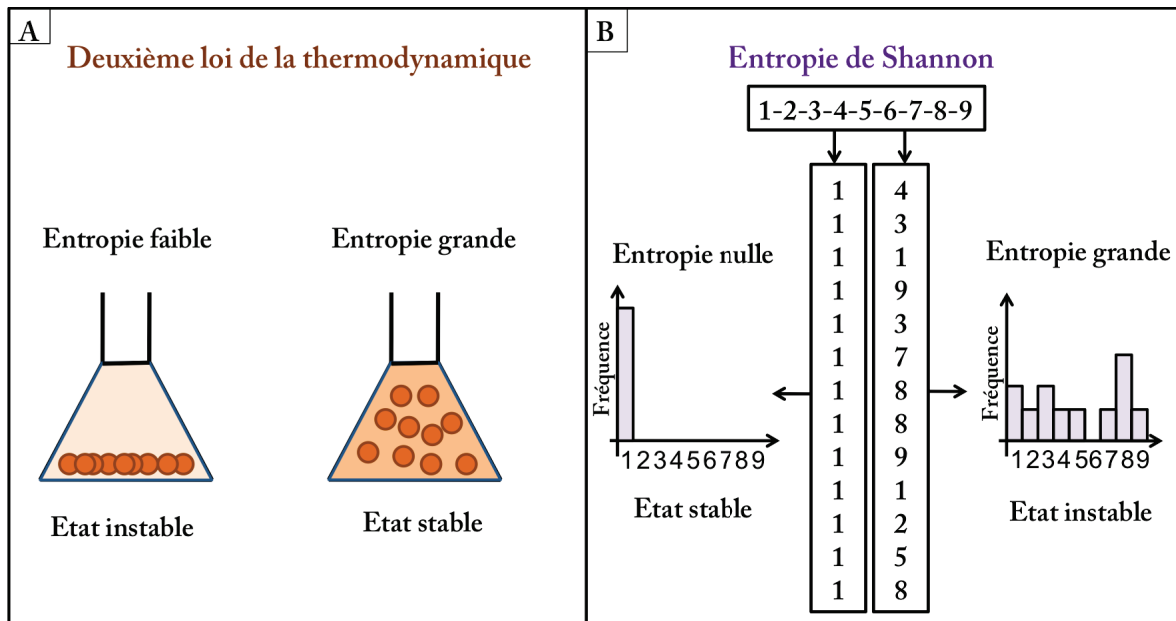


FIGURE 4 – Deuxième loi de la thermodynamique et entropie de Shannon

(A) Illustration de la deuxième loi de la thermodynamique suggérant que tout système évolue spontanément vers le désordre. L'entropie thermodynamique mesure donc le degré de désordre d'un système, dans lequel chaos est synonyme de stabilité. Contrairement à l'entropie de Shannon (B), énoncée dans le cadre de la théorie de l'information, pour laquelle le caractère aléatoire et variable de l'information transmise signifie que le système est instable et imprédictible mais informatif.

4B). Il s'agit ici du cas déterministe dans lequel il est possible de prédire que le prochain chiffre sera un 1, par conséquent l'entropie est nulle. À l'inverse, plus le système fournit de l'information incertaine, comme une succession aléatoire des chiffres de 1 à 9, plus l'entropie est grande. Dans ce cas on ne peut prédire quel chiffre sera pioché, l'entropie est donc maximale. Ainsi, le contexte dans lequel l'entropie est mesurée et son interprétation représentent deux éléments essentiels de son usage. La formulation utilisée en thermodynamique et par Claude Shannon a été utilisée plus tardivement en Biologie pour quantifier le degré de variabilité dans les systèmes cellulaires.

Le Jensen-Shannon-divergence (JSD) est par exemple une métrique inspirée de l'entropie de Shannon qui a été appliquée pour mesurer la variabilité génique de différentes populations cellulaires dans le cadre de la reprogrammation. Cette formule permet de quantifier à quel point des distributions de probabilité sont distinctes. Dans le contexte de la reprogrammation

cellulaire, chaque cellule a ainsi été transformée en un vecteur de probabilités, où chacune des probabilités correspond au niveau d'expression d'un gène normalisé par rapport à 1, sur lequel une mesure de JSD a été appliquée. Cette mesure du JSD renvoie donc un degré de dissemblance entre les cellules de la population étudiée, sur la base de leur niveau d'expression de gènes [65]. L'entropie de Shannon, elle-même, a ensuite été utilisée comme mesure de l'hétérogénéité inter-cellulaire. Contrairement à un système thermodynamique isolé, les cellules tendent à générer de l'ordre, par exemple pour leur survie ou leur croissance qui impliquent des chaînes de réaction ordonnées (métabolisme), et défient donc le second principe de la thermodynamique. Une nouvelle interprétation de l'entropie adaptée aux systèmes cellulaires a ainsi été introduite : la valeur d'entropie est d'autant plus grande que la variabilité de l'expression génique entre les cellules est importante [95, 88]. Comme pour l'entropie de Shannon, une entropie maximale est synonyme d'incertitude et donc d'instabilité du système, alors qu'un faible degré d'entropie se traduit par un système stable générant une information claire. Il est ainsi possible de spéculer sur l'état dans lequel se trouve une population de cellules par rapport à une autre. À la différence de certaines statistiques exposées dans la section précédente, cette mesure ne permet pas d'identifier des sous-groupes cellulaires, dans la mesure où elle ne calcule pas une distance entre des entités d'une population mais bel et bien un degré de variabilité.

1.3.2 La SEG, une force motrice de l'engagement en différenciation ?

La vision systémique du processus de différenciation a ouvert la voie à différentes configurations possibles reposant toutes sur l'expression des gènes et dans lesquelles la SEG doit augmenter pour permettre aux cellules de s'engager sur un chemin de différenciation [25, 24, 22]. Mes travaux de thèse s'appuient principalement sur la représentation proposée par Sui Huang [26], s'inspirant du paysage épigénétique conceptualisé par Conrad Hal Waddington [96]. Dans cette représentation, les cellules peuvent être représentées par des points dans un espace à multiples dimensions, dans lequel chaque dimension correspond à un gène (figure 5). Les cellules humaines se déplacent par exemple dans un espace à 30000 dimensions. Le niveau d'expression des gènes détermine ainsi la position des cellules. Par conséquent, si l'expression d'un gène change dans une cellule, celle-ci se déplacera dans l'espace multi-dimensionnel. La représentation d'un ensemble de cellules souches ou progénitrices en état d'auto-renouvellement dans cet espace, consiste donc en un nuage de points, correspondant à un profil génique donné (figure 5). À cet instant, les cellules se situent dans un état dit "attracteur", dans lequel leur expression génique est, certes stochastique mais, stable. Pour se déplacer dans l'espace multi-dimensionnel jusqu'à un autre état attracteur (un état cellulaire différencié, par exemple), il serait ainsi nécessaire que l'expression génique des cellules varie fortement. Chaque cellule suit alors une trajectoire différente jusqu'au nouvel état attracteur, l'état différencié, où la variabilité inter-cellulaire est réduite, et correspondant à un profil d'expression génique différent. L'expression des gènes doit fortement varier pour permettre aux cellules de se déplacer dans l'espace multi-dimensionnel, elles entrent par conséquent dans un état d'instabilité (le long de la trajectoire) avant de converger vers un nouvel état stable. D'autre part, les variations de l'expression des gènes étant stochastiques, chaque cellule emprunte un chemin différent, générant ainsi une augmentation significative de la variabilité inter-cellulaire. Cependant ces variations ne sont pas totalement incontrôlées, puisqu'elles dépendent du RRG ; par conséquent les trajectoires empruntables sont délimitées par les contraintes du réseau. Ainsi dans le cas où les cellules auront par exemple le choix entre plusieurs lignages, comme dans l'hématopoïèse, elles s'engageront dans l'une des trajectoires possibles, "dessinées" par les contraintes du RRG, mais emprunteront individuellement des chemins différents.

D'autre part, ce modèle suggère que l'engagement en différenciation s'accompagne d'une forte augmentation de la stochasticité inter-cellulaire [97, 26]. Cette variabilité pourrait alors jouer un rôle moteur dans l'engagement en masse des cellules dans la différenciation. Il a d'ailleurs été proposé depuis déjà quelques années que la SEG pourrait en effet être une force motrice de l'engagement en différenciation et être à l'origine de la diversité cellulaire générée lors du développement embryonnaire [24]. Ainsi, bien qu'il ait été suggéré dans plusieurs modèles, le rôle de la SEG dans la différenciation demeure reste encore à démontrer formellement. Comprendre le rôle de cette variabilité permettrait pourtant de mieux caractériser et contrôler les processus de différenciation, ce qui représente un véritable enjeu notamment dans le traitement des pathologie résultant d'un défaut de la différenciation, comme les myélodysplasies blastiques [98].

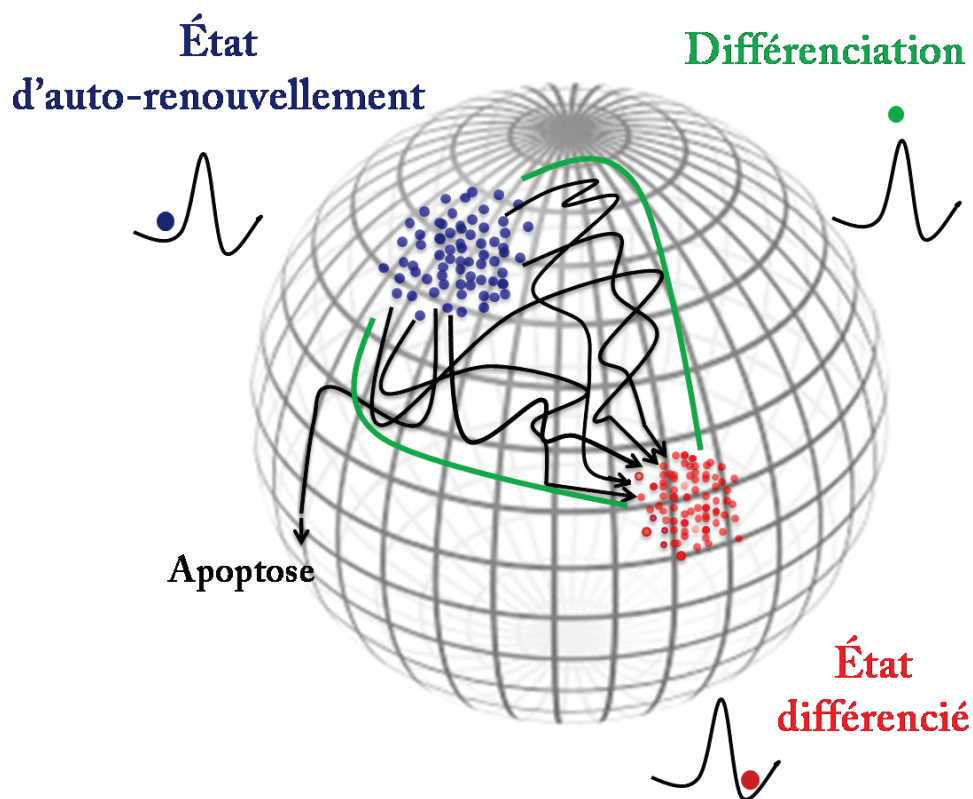


FIGURE 5 – Représentation des cellules au cours du processus de différenciation érythrocytaire dans l'espace multidimensionnel des gènes

La sphère représente l'espace multidimensionnel des gènes. L'œil humain ne pouvant percevoir plus de trois dimensions, il est impossible de représenter réellement l'ensemble des dimensions définissant cet espace, qui est ici illustré par la sphère. Les cellules sont représentées par des points localisés sous forme de nuage en deux endroits de l'espace multidimensionnel. Le nuage bleu représente des cellules en état d'auto-renouvellement et le nuage rouge représente des cellules après différenciation. Un schéma type du paysage de Waddington est également représenté à côté de ces deux états correspondant à des vallées du paysage, autrement dit à des bassins d'attraction. Dans l'espace multidimensionnel des gènes, le mouvement d'une cellule correspond à une variation de l'expression de ses gènes. En état d'auto-renouvellement, la SEG entraîne le déplacement des cellules autour de l'état attracteur bleu. Or une forte augmentation de cette SEG permettrait aux cellules de quitter cet état stable, pour s'engager sur le chemin de la différenciation, représenté par les flèches noires. Chaque cellule peut emprunter un chemin différent, cependant leur trajectoire est bornée par les contraintes épigénétiques représentées par les lignes vertes. Certaines cellules peuvent s'échapper de cette trajectoire, pour entrer en apoptose, par exemple. Cette importante augmentation de la SEG permettrait ainsi aux cellules de franchir le sommet séparant les deux bassins attracteurs, et d'atteindre un autre état stable, l'état différencié, représenté sur ce schéma en rouge. Ce nouvel état attracteur est localisé dans un autre endroit de l'espace multidimensionnel, il correspond ainsi à un profil d'expression génique différent.

1.4 Métabolisme et différenciation cellulaire

1.4.1 Présentation du métabolisme cellulaire eucaryote

1.4.1.1 Introduction au métabolisme

Le métabolisme est un système régulé et régulateur qui peut se décrire comme un ensemble de transformations moléculaires successives, dans lesquelles les produits de réactions transformantes deviennent à leur tour les substrats de prochaines réactions. Le métabolisme se divise en deux catégories : (1) le catabolisme, équivalent à la dégradation graduelle de molécules complexes provenant de l'environnement extracellulaire, et à leur transformation en molécules simples utilisables par la cellule ; (2) l'anabolisme, équivalant à la synthèse de grosses molécules essentielles pour les cellules, à partir de petites molécules cellulaires. Catabolisme et anabolisme font intervenir respectivement des réactions d'oxydation et de réduction catalysées par des enzymes. Le catabolisme a pour objectif de produire de l'énergie cellulaire *via* l'oxydation progressive de molécules organiques. De façon réciproque, l'anabolisme consomme de l'énergie afin d'alimenter les réactions de biosynthèses successives conduisant à la production de grandes molécules à partir de petits éléments issus du catabolisme. Parmi les voies cataboliques, nous pouvons ainsi distinguer la glycolyse, la lipolyse et la protéolyse dégradant respectivement les glucides, les lipides, et les protéines. Les voies anaboliques sont donc, quant à elles, des voies de synthèse moléculaire, telles que la glucogenèse et la néoglucogenèse en ce qui concerne le glucose. Ainsi, toutes ces transformations métaboliques partagent le même objectif, qu'est l'apport de ressources nécessaires aux cellules pour leur permettre de se maintenir en vie et d'exercer leurs fonctions.

Dans un organisme, les macromolécules à l'origine des voies de dégradation sont d'abord soumises à digestion enzymatique dans les conduits digestifs, et entrent ensuite dans les cellules sous forme de composés intermédiaires. Les produits de digestion des glucides, lipides et protéines alimentaires sont respectivement des polysaccharides, tels que le glucose ou le fructose, des triglycérides ou autres dérivés de lipides (cholestérol) et des acides aminés. Pour la plupart des cellules, les polysaccharides représentent la source principale de carburant énergétique [99].

1.4.1.2 Co-facteurs métaboliques et réservoirs énergétiques

Chacune des grandes voies métaboliques conduisent à la production ou à la consommation d'énergie. La majorité de cette énergie se matérialise sous forme moléculaire. Il existe différents transporteurs de cette énergie dont les principaux sont l'ATP, dans lequel l'énergie est véhiculée sous forme de liaisons chimiques, ainsi que le NADH et le NADPH, transportant des électrons fortement chargés en énergie. L'ATP et le NADH sont notamment des co-facteurs des réactions cataboliques, alors que le NADPH intervient préférentiellement dans les voies anaboliques. Ces molécules transportent temporairement l'énergie produite par les voies métaboliques, et interviennent ensuite sous forme de substrats, de co-facteurs enzymatiques, ou encore de précurseurs, pour apporter l'énergie ou les groupements chimiques nécessaires à de nombreuses autres réactions cellulaires.

L'ATP est le transporteur d'énergie le plus utilisé dans les cellules. Son hydrolyse permet de libérer l'énergie qu'il transporte et de l'investir dans la réaction chimique à laquelle il est couplé. Par exemple, pour fonctionner, les transports membranaires dits "actifs", nécessitent d'être couplés à une charge énergétique apportée par l'ATP. Son hydrolyse aboutit également à la libération de phosphate inorganique et d'une molécule d'ADP qui pourra à nouveau être phosphorylée pour former de l'ATP. La dégradation de cette molécule énergétique en fait également un précurseur de la synthèse des séquences nucléiques. En effet, l'ATP est constitué d'adénine, qui est une base azotée entrant dans la composition des nucléotides [100].

Par ailleurs, l'énergie cellulaire est aussi véhiculée *via* les transporteurs d'électrons le NADH et le NADPH. Comme précisé en amont, ces deux molécules n'interviennent pas dans les mêmes voies métaboliques. Le NADH et le NADPH sont respectivement produits lors de réactions d'oxydation à partir de NAD^+ , ou de réduction à partir de NADP^+ . En outre, les cellules maintiennent un ratio NAD^+/NADH plus élevé que le ratio $\text{NADP}^+/\text{NADPH}$ [101].

Le couple NAD^+/NADH joue un rôle essentiel dans la régulation de la glycolyse. Le NAD^+ intervient comme co-facteur d'une étape des étapes successives de la glycolyse, dans laquelle il est réduit sous forme de NADH (figure 6). Pour entretenir la glycolyse, il est donc nécessaire que le NADH soit ré-oxydé en NAD^+ dans le cytosol. De plus, le NAD^+ est indispensable au fonctionnement de la phosphorylation oxydative qui est un pilier du métabolisme cellulaire. Le stock de molécules de NAD^+ cytosolique peut être renouvelé

par des mécanismes impliquant deux voies énergétiques différentes, la navette du glycérol-3-phosphate *via* la chaîne respiratoire mitochondriale, ou l'oxydation du pyruvate en lactate dans la voie de la glycolyse anaérobie. La chaîne respiratoire ainsi que la glycolyse anaérobie seront décrites plus en détail dans la section *Voies énergétiques de la glycolyse*.

Le NADPH, quant à lui, est utilisé comme agent réducteur des réactions de biosynthèse. Il intervient par exemple dans la synthèse du cholestérol ou des acides gras. Il est régénéré à partir de NADP⁺ dans le cytosol, *via* la voie des pentoses phosphates, correspondant à la conversion du glucose-6-phosphate en ribulose-5-phosphate, par une succession de réaction en série produisant deux molécules de NADPH.

1.4.2 Le métabolisme du glucose

1.4.2.1 Catabolisme et anabolisme du glucose

Le glucose, qui est l'intermédiaire glucidique le plus fréquent, est oxydé en pyruvate dans le cytosol des cellules *via* une série de dix réactions en chaînes, la glycolyse, conduisant chacune à un sucre intermédiaire différent et impliquant une enzyme différente [102] (figure 6).

Cette voie de dégradation, autrement dit la glycolyse, permet de fournir des substrats aux autres voies métaboliques, de réguler l'état redox cellulaire et de produire de l'énergie. Sa vitesse est donc soumise à régulation au niveau de trois étapes limitantes, en fonction des besoins énergétiques [103]. Le premier niveau de contrôle a lieu dès la première étape de la glycolyse, commune à la glycolyse et à la glycogénèse (figures 6 et 14). Il s'agit d'une réaction irréversible consommant de l'ATP pour produire du glucose-6-phosphate par phosphorylation de la molécule de glucose. Elle nécessite l'intervention des enzymes hexokinases (HK1 et HK2), dont les défauts d'expression, notamment pour HK2, sont impliqués dans un grand nombre de cancers [104, 105]. Les transporteurs membranaires du glucose, GLUT (GLUT1 à GLUT 5 en fonction du type cellulaire), ne peuvent prendre en charge le glucose phosphorylé, l'empêchant ainsi de s'échapper de la cellule. De plus, le glucose-6-phosphate bloqué dans la cellule peut participer à une boucle de rétrocontrôle négatif de la glycolyse [106]. En effet, en quantité trop importante dans la cellule, cet intermédiaire se fixe aux hexokinases et inhibe ainsi leur activité enzymatique [106]. La troisième réaction de la glycolyse représente

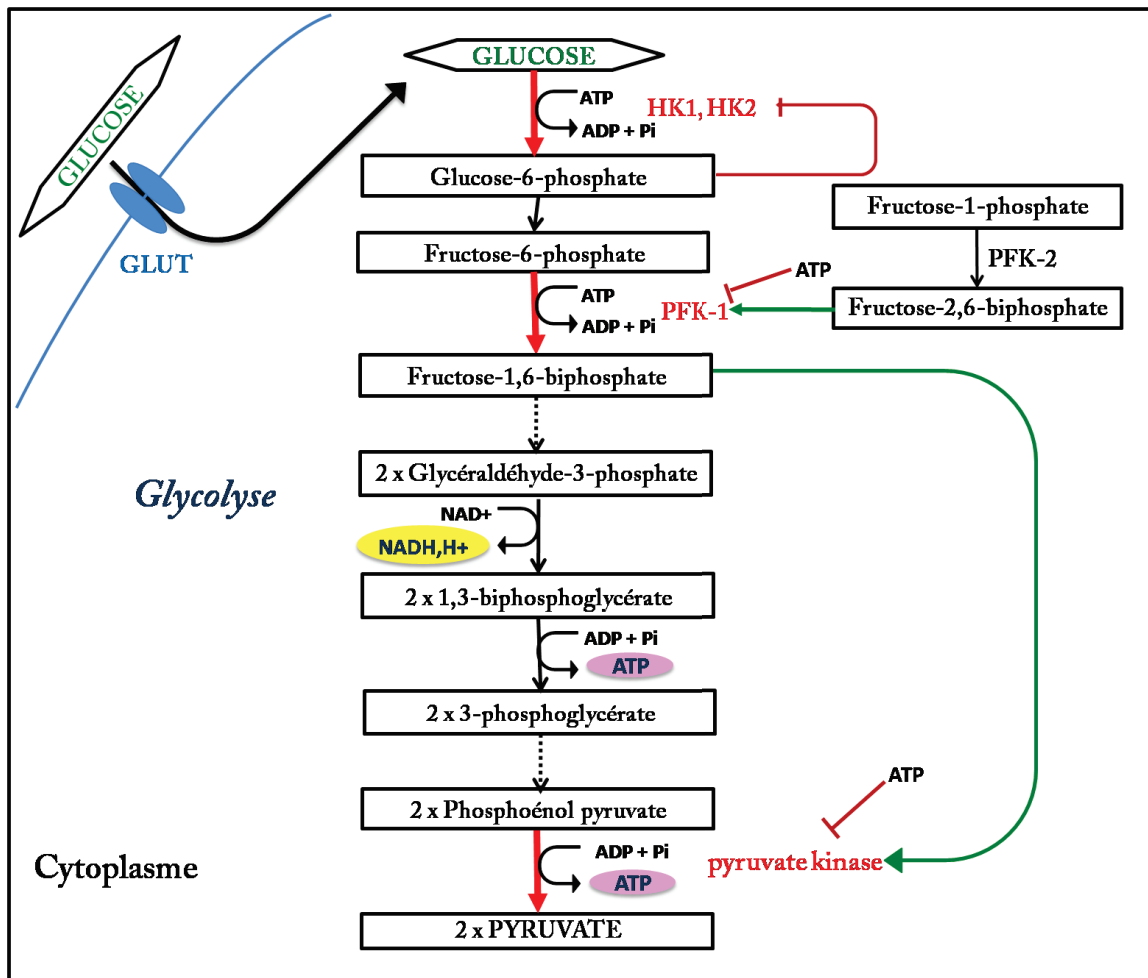


FIGURE 6 – Étapes majeures de la glycolyse

Représentation des principales étape du catabolisme du glucose, depuis son entrée dans la cellule grâce aux transporteurs GLUT, jusqu'à la production de pyruvate. Les trois étapes régulatrices de la glycolyse sont identifiées par des flèche rouges. Les flèches vertes représentent des activations, les lignes rouges correspondent à des inhibitions et les flèches en pointillés sont des étapes intermédiaires non représentées sur ce schéma. Seules les enzymes impliquées dans les trois étapes régulatrices ont été citées sur cette illustration.

le deuxième point de contrôle. Contrairement à la première étape régulatrice, elle est unique à la glycolyse et engage le glucose de façon irréversible dans cette voie. Cette réaction, catalysée par la phosphofructokinase-1 (PFK-1), transforme par phosphorylation couplée à une consommation d'ATP, le fructose-6-phosphate en fructose-1,6-biphosphate [107]. La PFK-1 est une enzyme clef de la glycolyse dont l'activité est régulée par la charge énergétique de la cellule et le fructose-2,6-bisphosphate [108]. En effet, si la charge énergétique cellulaire

est élevée, il n'est plus nécessaire pour la cellule d'oxyder du glucose pour produire plus d'énergie. Dans ce cas, la PFK-1 peut être inhibée par l'ATP s'il est en quantité trop importante, qui freinera ainsi le flux de la glycolyse. D'autre part, elle peut être activée par la molécule de fructose-2,6-bisphosphate, produite à partir de fructose-1-phosphate grâce à la PFK-2 [108, 109], elle même régulée par la protéine kinase A (figure 6) [110]. Par ailleurs, l'énergie investie sous forme d'ATP dans ces deux étapes de contrôle, est recouvrée dans les dernières étapes de la glycolyse. Il est également intéressant de noter en terme d'énergie que la conversion du glycéraldéhyde-3-phosphate en 1,3-biphosphoglycérate est couplée à une consommation de NAD^+ , qui devra donc être renouvelée pour maintenir le flux de la glycolyse (figure 6). Enfin, la troisième étape limitante correspond à la dernière réaction de la glycolyse, et contrôle ainsi la sortie de la voie. Cette réaction est en effet couplée à une production d'ATP permettant de compenser les pertes associées aux deux premières étapes régulatrices. Elle consiste à transformer le phosphoenolpyruvate en pyruvate grâce à l'enzyme pyruvate kinase, régulée par la molécule de fructose-1,6-biphosphate, formée en amont dans la glycolyse (figure 6) [111]. Le pyruvate produit peut ensuite rejoindre deux voies de production d'énergie différentes, la phosphorylation oxydative ou la glycolyse anaérobie (figure 7), qui seront décrites dans les prochaines sections.

Ainsi, ces trois étapes régulatrices ont pour fonction de fluidifier les réactions en chaînes de la glycolyse et d'éviter l'accumulation de composés intermédiaires. Enfin, la glycolyse représente un processus important et central dans la production d'énergie. Par ailleurs, à l'échelle de l'organisme, le glucose en excès est phosphorylé puis stocké sous forme de glycogène, particulièrement dans les cellules hépatiques et musculaires. Contrairement aux hépatocytes, les cellules musculaires dégradent le glycogène en glucose-6-phosphate pour renouveler leur propre énergie lors d'efforts physiques. Les cellules hépatiques, quant à elles, représentent le siège de la glycogénèse, glycogénolyse et néoglucogénèse. En effet ces cellules stockent les polysaccharides sous forme de glycogène qu'elles dégradent par glycogénolyse en glucose-6-phosphate, qui rejoint la voie de la glycolyse. Grâce à l'expression de la glucose-6-phosphatase, les hépatocytes sont les seules cellules capables de régénérer du glucose à partir de sa forme phosphorylée. En situation d'hypoglycémique les cellules hépatiques dégradent d'abord le glycogène, pour libérer du glucose dans le sang. En cas extrêmes, elles sont également capable

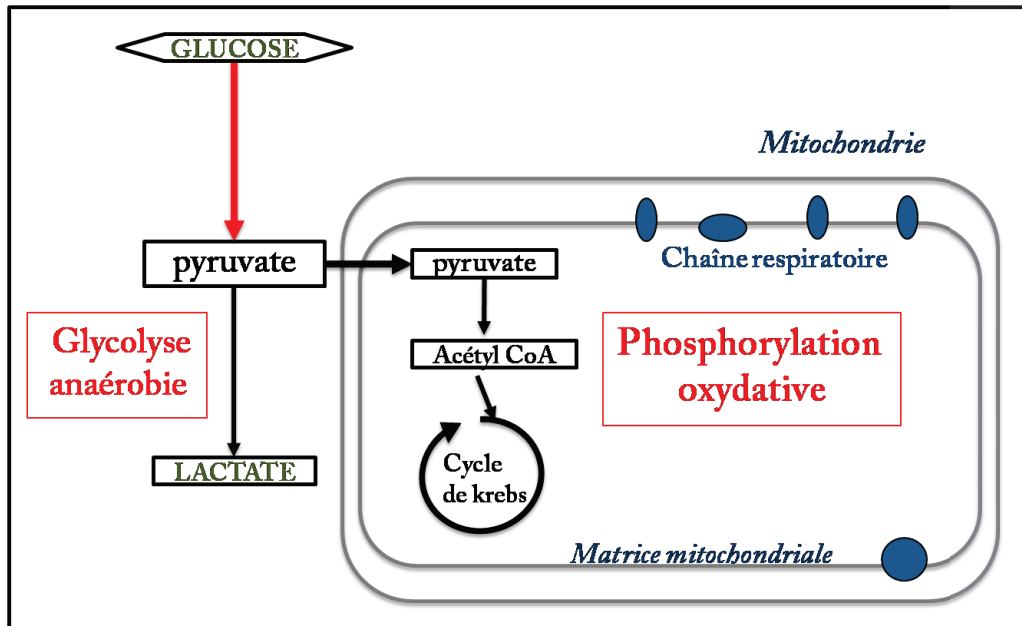


FIGURE 7 – Voies énergétiques du pyruvate

Cette représentation représente les voies énergétiques dans lesquelles le pyruvate peut être impliqué. Il peut en effet entrer dans la mitochondrie, être converti en acétyl CoA puis alimenter le cycle de krebs qui participe à la phosphorylation oxydative. Il peut également être transformé en lactate dans le cytosol et participer à la glycolyse anaérobie.

de produire du glucose à partir de composés non glucidiques *via* la néoglucogenèse.

1.4.2.2 Les voies énergétiques de la glycolyse

1.4.2.2.1 La glycolyse anaérobie

Le dioxygène est important pour le maintien vital et fonctionnel d'un grand nombre de cellules. Cependant, certains types cellulaires, étant par exemple localisées dans des environnements hypoxiques, tels que les cellules souches hématopoïétiques (HSC) dont la niche se situe dans la moelle osseuse [112], ne peuvent utiliser l'oxygène pour assurer leur survie et leurs fonctions. Dans ces situations, le pyruvate généré par la glycolyse emprunte donc préférentiellement la voie énergétique dite anaérobie, au lieu d'être transporté dans les mitochondries pour participer à la phosphorylation oxydative nécessitant du dioxygène (figure 7) [113]. D'autre part, dans les cellules cancéreuses particulièrement, mais également dans des contextes physiologiques tels que la réponse immunitaire, la voie anaérobie est mise en

place malgré l'accès au dioxygène [114]. Dans ces conditions la glycolyse est plutôt qualifiée de glycolyse aérobie [114]. Ce type de glycolyse a été mise en évidence par Otto Warburg [115]. Otto Warburg constata qu'en contexte cancéreux, la consommation de glucose des cellules était plus accrue. Ces observations lui ont alors permis de montrer que le métabolisme glycolytique des cellules cancéreuses est différent de celui des tissus normaux et qu'il repose sur un processus comparable à la fermentation lactique [115].

Toutefois, pour ne pas induire de confusion dans ce manuscrit nous utiliserons toujours le terme "anaérobie" lorsque nous discuterons de la voie glycolytique aboutissant à la production de lactate et dont la finalité n'est pas la phosphorylation oxydative.

La glycolyse anaérobie permet de renouveler le NAD^+ afin de maintenir le flux glycolytique, et de produire de l'énergie utilisable. Néanmoins elle ne produit que deux molécules d'ATP, autrement dit seize fois moins que la phosphorylation oxydative. D'autre part, elle est beaucoup plus rapide et demande par conséquent un apport en glucose bien plus important [116, 117]. La glycolyse anaérobie se résume à une seule réaction cytosolique catalysée par la lactate déshydrogénase A (LDHA), catalysant la conversion du pyruvate en acide lactique (lactate) (figure 8). L'enzyme LDHA fait partie de la famille des lactate déshydrogénases (LDH) constituée de cinq isoenzymes et codée par quatre gènes différents. Il s'agit des gènes *LDHA*, *LDHB*, *LDHC*, majoritairement exprimés chez les vertébrés, et *LDHD*, notamment exprimé dans les bactéries [117]. Les plus communs, *LDHA* et *LDHB* codent respectivement pour les chaînes M et H s'associant pour former les cinq isomères [118]. LDHA (appelé aussi LDH5) est ainsi constituée de quatre chaînes M alors que LDHB (appelé aussi LDH1) se compose de quatre chaînes H [117]. Ces deux LDH possèdent les mêmes sites actifs mais présentent des affinités différentes pour le pyruvate et le lactate. LDHA converti préférentiellement le pyruvate en lactate, et le NADH en NAD^+ , alors que LDHB catalyse préférentiellement la réaction inverse.

Le rôle de LDHA a été principalement étudié dans l'effet Warburg des cellules cancéreuses, ainsi que dans les fonctions des cellules souches hématopoïétiques et neuronales [119, 120, 121, 122]. Il a notamment été suggéré que l'enzyme LDHA est essentielle au maintien des cellules souches ou cancéreuses dans un état indifférencié et prolifératif [123, 124, 125]. Le gène *LDHA* est régulé par plusieurs facteurs de transcription, dont $\text{HIF1-}\alpha$, *cMyc*, *FoxM1* et

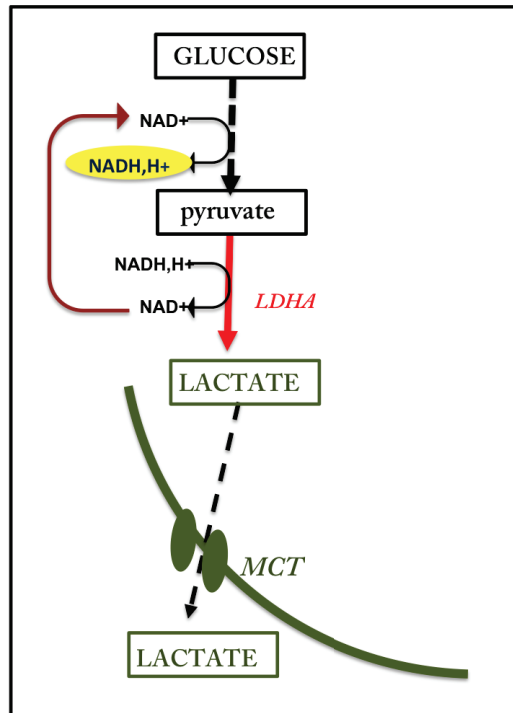


FIGURE 8 – La glycolyse anaérobie

Le glucose est catabolisé en pyruvate *via* la glycolyse. Le pyruvate est ensuite converti dans le cytosol en lactate par l'enzyme LDHA. Cette étape permet alors de régénérer du NAD⁺ pour alimenter la glycolyse. Le lactate peut ensuite être relargué dans le milieu extracellulaire par les transporteurs MCT.

KLF4 [117].

HIF1 est un facteur induit par l'hypoxie. Ainsi, en absence de dioxygène, il active l'expression de gènes glycolytiques, tels que *LDHA*, pour subvenir aux besoins énergétiques des cellules *via* la glycolyse anaérobie [126]. Ce facteur de transcription a particulièrement été étudié dans le cadre de la glycolyse anaérobie des cellules cancéreuses. Dans l'effet Warburg, la sous-unité HIF1- α est stabilisée par différents facteurs, dont des cibles de mTOR [117]. L'activation d'HIF1 est associée à l'augmentation de l'expression de facteurs glycolytiques, telles que les hexokinases [127], le transporteur du glucose GLUT1 [128], et LDHA comme il a été mentionné en amont [129]. De plus, HIF1 est également capable d'induire la sur-expression de la pyruvate déshydrogénase kinase, inhibant le complexe PDH qui permet au pyruvate d'entrer dans le cycle de krebs [129]. De cette façon, l'expression d'HIF1 empêche le pyruvate d'alimenter la phosphorylation oxydative *via* le cycle de krebs, et promeut la

glycolyse anaérobie. En effet, il a par exemple été montré dans les LT-HSC (Long-Term Hematopoietic Stem Cells), qu'HIF1- α régule le métabolisme anaérobie sur lequel la survie et les fonctions de ces cellules reposent [130].

Par ailleurs, l'augmentation de l'expression de LDHA en conditions anaérobiques (ou aérobiques) s'accompagne de l'augmentation de l'expression des transporteurs du lactate, appartenant à la famille des transporteurs monocarboxyles (MCT) [131, 132]. Parmi les membres de cette famille, MCT1 est particulièrement impliqué dans l'entrée, mais aussi, la sortie du lactate de la cellule. En revanche, MCT4 intervient préférentiellement dans l'évacuation du lactate, notamment dans les tissus glycolytiques (anaérobisme) et les tumeurs [133]. Il a été montré que l'hypoxie régule l'expression des MCT4 au travers de l'expression d'HIF1- α [133]. Par conséquent, MCT4 est sur-exprimé en conditions hypoxiques, et permet ainsi aux cellules de rejeter le lactate dans l'espace extra-cellulaire. Enfin, il semblerait que LDHA soit impliquée dans une boucle de rétro-contrôle avec HIF1 [134]. En faveur de cette hypothèse, il a par exemple été suggéré que le lactate puisse être un régulateur de l'expression d'HIF1 grâce aux transporteurs MCT1 [135]. Ainsi HIF1 et LDHA représentent des médiateurs importants du métabolisme anaérobie et de l'effet Warburg. Par ailleurs, le lactate n'est pas qu'un simple déchet de la glycolyse anaérobie, il semble participer à divers processus biologiques *via* ses transporteurs MCT et son récepteur GPR81 [136]. Il est par exemple impliqué dans la réponse anti-inflammatoire, la protection et la plasticité neuronale, et la croissance tumorale [136]. L'homéostasie du lactate est donc régulée à l'échelle de l'organisme, particulièrement par les hépatocytes. Lorsqu'il est relâché dans l'espace extracellulaire, il est notamment capturé par ces cellules, dans lesquelles il est utilisé comme substrat de la néoglucogenèse. En cas de besoins (baisse de la glycémie, par exemple) les hépatocytes peuvent alors rejeter le glucose produit à partir de lactate, dans la circulation sanguine pour alimenter les autres cellules. Ce mécanisme reliant la glycolyse et la néoglucogenèse tout en maintenant l'homéostasie lactique, est nommé cycle de Cori, en l'honneur du couple de chercheur britannique ayant reçu un prix nobel pour l'avoir mis en évidence, notamment entre les cellules musculaires et les hépatocytes [137, 138]. Outre les cellules musculaires, les érythrocytes participent également à ce cycle grâce aux transporteurs MCT1, leur permettant de relâcher le lactate dans la circulation sanguine [139]. Le glucose produit à partir de ce lactate est distribué en retour par les

hépatocytes pourra alors être intégré par les globules rouges, comme les cellules musculaires en cas d'activité physique intense par exemple (figure 9).

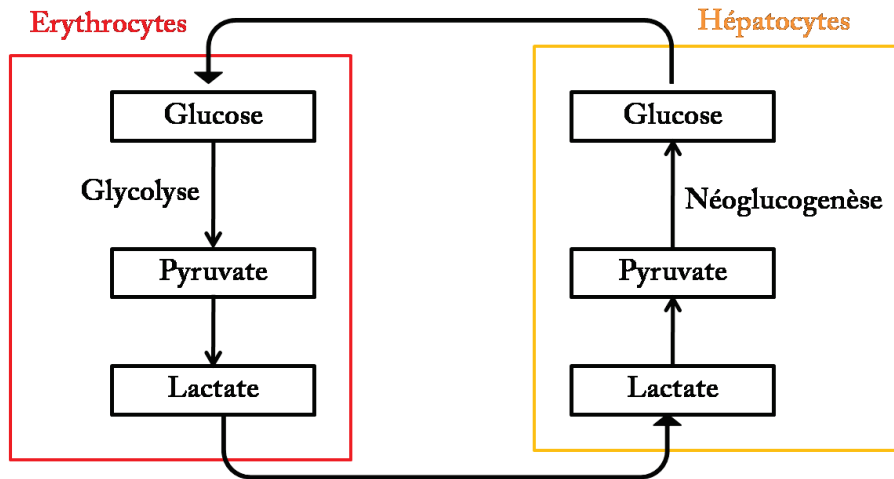


FIGURE 9 – Le cycle de Cori

Est représenté ici le cycle de Cori entre les cellules hépatiques et les cellules musculaires. Le lactate des cellules musculaires est rejeté dans la circulation sanguine et récupéré par les hépatocytes pour être métabolisé en glucose de nouveau *via* la néoglucogenèse.

1.4.2.2.2 La phosphorylation oxydative

De façon générale, en présence de dioxygène, le pyruvate produit à l'issue de la glycolyse entre majoritairement dans les mitochondries dans lesquelles il peut participer à la phosphorylation oxydative, après être converti en acétyl CoA (figure 7). La phosphorylation oxydative correspond à une succession de réactions mitochondriales, exploitant l'énergie produite par l'oxydation complète de métabolites cellulaires, tels que l'acétyl CoA. Elle a pour principaux objectifs de synthétiser de l'ATP et de régénérer du NAD⁺. Cependant avant de décrire plus en détail le processus de phosphorylation oxydative, il est nécessaire d'apporter quelques précisions quant au cycle de krebs, également nommé cycle de l'acide citrique. L'oxydation cytosolique d'une molécule de glucose aboutit à deux molécules de pyruvate qui peuvent être investies dans plusieurs voies glycolytiques. En présence d'oxygène le pyruvate rejoint préférentiellement la voie de la phosphorylation oxydative, *via* le cycle de krebs. Dans cette situation le pyruvate est transporté dans la mitochondrie par le complexe mitochondrial MPC1/MPC2 [140], comme mentionné dans la section précédente, dans laquelle il pourra être

converti en acétyl CoA et ainsi intégrer le cycle de krebs. Parmi les huit étapes composant ce cycle, quatre réactions d'oxydation sont couplées à la génération de NADH ou de FADH₂, un autre transporteur d'électrons (figure 10). L'énergie portée par ces deux molécules servira alors à alimenter la chaîne respiratoire, correspondant à une chaîne de transport d'électrons (CTE), couplée avec une phosphorylation de l'ADP en ATP.

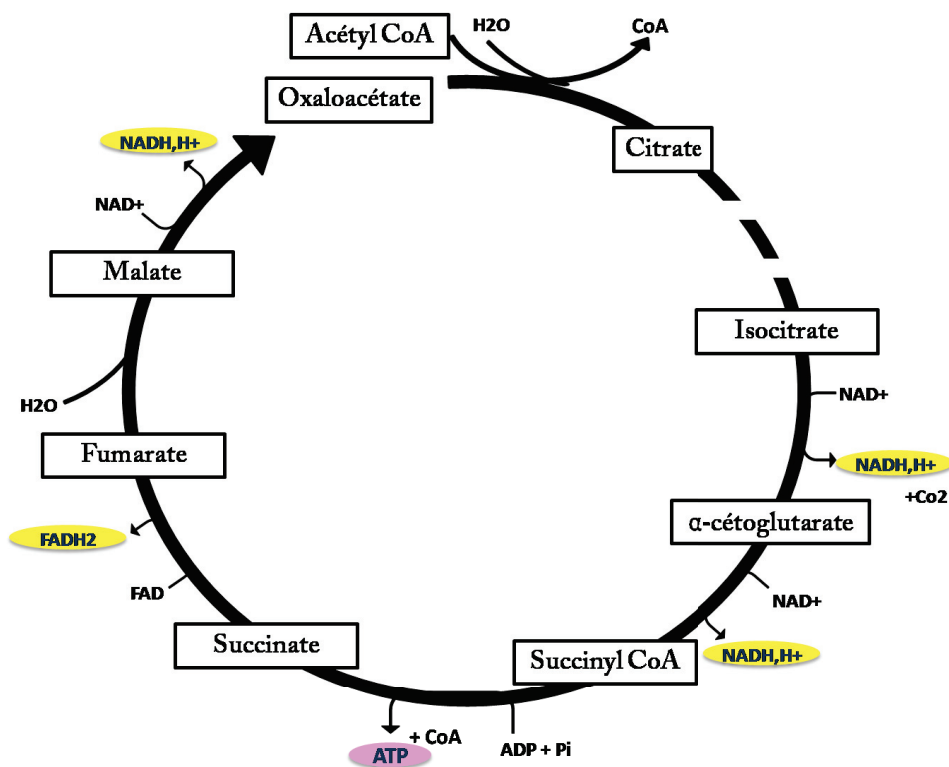


FIGURE 10 – Principales étapes du cycle de krebs

L'acétyl CoA entre dans le cycle de krebs en même temps que l'oxaloacétate pour former le citrate. Seules les étapes couplées à la production d'énergie, sous forme de co-facteurs ou d'ATP, ont ensuite été représentées. Les étapes intermédiaires sont indiquées par des coupures dans la flèche continue donnant la direction du cycle.

La CTE prend place dans l'espace intermembranaire mitochondrial. Il s'agit d'une chaîne de quatre complexes protéiques transmembranaires successifs, fonctionnant tels des pompes à protons H⁺, à l'exception du complexe II (figure 11) [141]. Comme son nom l'indique, cette chaîne correspond au transfert d'électrons en série de molécules donneuses, telles que le NADH, sur des molécules receveuses, *via* des réactions d'oxydo-réduction catalysées par différentes enzymes [141].

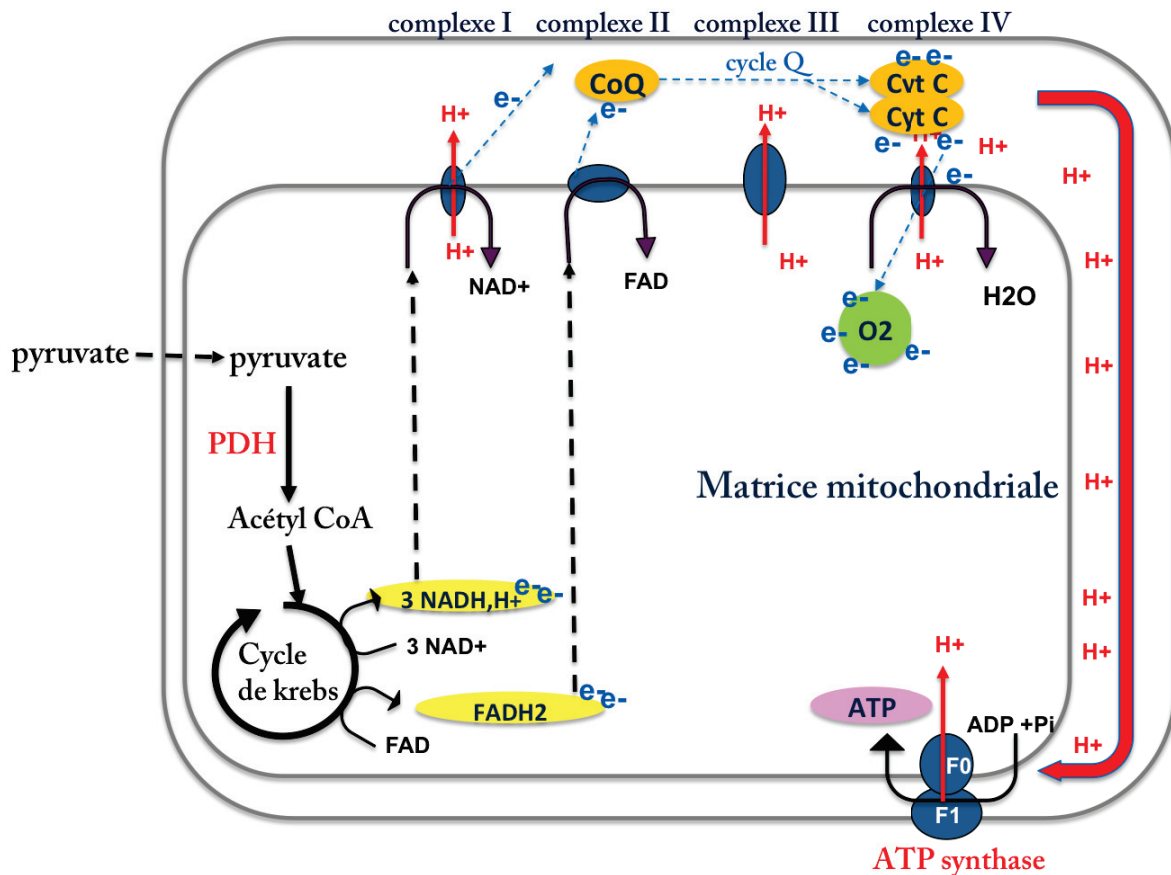


FIGURE 11 – La phosphorylation oxydative

Cette représentation schématique correspond à la dégradation du glucose, depuis son entrée dans la glycolyse jusqu'à sa combustion complète par la phosphorylation oxydative. Le NADH et le FADH₂ produits par le cycle de krebs rejoignent la chaîne respiratoire composée des quatre complexes pour transférer leurs électrons à la molécule ubiquinone (CoQ), qui transmet alors au travers du cycle Q les électrons au cytochrome C, qui les transfère sur l'accepteur final, le dioxygène. Le pompage des protons H⁺ par les complexes I, III, et IV à chacune de ces étapes permet de produire une force protons-motrice, représentée par la flèche rouge. Ce gradient de proton permet d'activer l'ATP synthase pour produire de l'ATP dans la mitochondrie. Les flèches pointillées correspondent à des étapes intermédiaires non développées sur cette représentation.

Premièrement, le NADH et le FADH₂, synthétisés au cours du cycle de krebs, diffusent jusqu'à la CTE. Le NADH est oxydé au niveau du complexe I par la NADH-déshydrogénase, et le FADH₂ est oxydé par l'enzyme succinate-déshydrogénase du complexe II. Ces réactions permettent de régénérer du NAD⁺ et du FAD, les rendant à nouveau disponibles pour recevoir l'énergie générée par le cycle de krebs. Deuxièmement, les électrons libérés par l'oxy-

dation du NADH et du FADH₂ sont transférés sur une molécule receveuse, l'ubiquinone, aussi nommée coenzyme Q (CoQ). Cette dernière, est alors réduite en ubiquinol, et transporte les électrons jusqu'au complexe III. La sous-unité ubiquinol-cytochrome c oxidoreductase du complexe III [142] entraîne ensuite le transfert d'électrons de l'ubiquinol aux molécules de cytochrome c, *via* le cycle Q [143]. De façon globale, ce cycle permet de transférer un par un les électrons, issus du NADH et du FADH₂, sur les molécules de cytochrome c au moyen d'un complexe fer-souffre. Enfin, les cytochromes c transportent les électrons jusqu'au complexe de la cytochrome c oxidase (complexe IV), dans lequel ils sont transférés sur l'accepteur final d'électrons, le dioxygène.

Une partie de l'énergie libérée par ce flux d'électrons est utilisée parallèlement par les complexes I, III et IV pour pomper des protons H⁺ de la matrice mitochondriale vers l'espace intermembranaire. Le dioxygène ayant reçu les électrons du cytochrome c, peut alors s'associer à quatre protons H⁺, apportés par l'action de pompage du complexe IV, pour former deux molécules d'eau. Par ailleurs, le déplacement des protons H⁺ à travers la membrane crée un gradient de voltage, aussi appelé potentiel de membrane, et un gradient de concentration de protons. L'association de ces deux flux génère un gradient électrochimique de protons exerçant une force proton-motrice. Ce potentiel membranaire permet ainsi d'induire la production d'ATP par un cinquième complexe protéique transmembranaire, le complexe de l'ATP synthase, aussi nommé F₀/F₁ [144, 145]. Le complexe de l'ATP synthase est constitué principalement de deux sous unités, la sous unité F₁ ATPase, représentant le corps catalytique, et la sous-unité F₀ correspondant au transporteur de protons H⁺. Ainsi, les protons pompés par les complexes I, III et IV diffusent à travers le flux électrochimique jusqu'à l'ATP synthase, qui entraîne leur retour dans la matrice mitochondriale. La diminution de la quantité de protons dans l'espace intermembranaire affaiblit par conséquent le potentiel de membrane, qui sera à nouveau augmenté par l'entrée de H⁺ lors de l'oxydation d'autres molécules de NADH et FADH₂. D'autre part, l'ATP synthase catalyse la formation d'ATP à partir d'ADP et de phosphate inorganique, dans la matrice mitochondriale. Enfin, le gradient électrochimique permet de coupler la CTE à la synthèse d'ATP, définissant ainsi le terme de phosphorylation oxydative. L'oxydation complète d'une seule molécule de glucose par ce mécanisme, produit une trentaine de molécules d'ATP.

Les connaissances du mécanisme de phosphorylation oxydative résultent de l'utilisation de différentes substances inhibitrices du transport d'électrons. Parmi ces inhibiteurs nous pouvons citer, la Roténone Amytal agissant au niveau du complexe I [146], l'Antimycine A intervenant dans le cycle Q du complexe III [147], et l'Oligomycin A inhibant l'ATP synthase [148]. Certaines substances, telles que le FCCP, sont des agents découplants. Ces substances, liposolubles, agissent comme des transporteurs de protons H^+ . Cependant de part leur nature liposoluble elles diffusent dans la membrane mitochondriale et exportent les H^+ en dehors de l'espace intermembranaire. Par conséquent, la force proton-motrice se dissipe et l'ATP ne peut plus être produit. Ainsi, en présence de ces agents découplants la synthèse d'ATP est arrêtée mais les cellules continuent de consommer du dioxygène. Ces inhibiteurs de la CTE permettent ainsi de rendre compte du fonctionnement de la phosphorylation oxydative.

1.4.3 Carrefours et convergences des différentes voies métaboliques

Avant de s'intéresser aux connexions entre les grandes voies métaboliques, il convient de présenter de façon synthétique les métabolisme des lipides et des protéines. Tout comme la glycolyse, la dégradation des lipides permet de produire de l'énergie. Dans la plupart des cellules ces macromolécules représentent la deuxième source d'énergie après les glucides. Les triglycérides entrent dans les cellules, dans lesquelles ils sont hydrolysés, fournissant ainsi trois molécules d'acides gras et du glycérol. Les acides gras sont oxydés dans la matrice mitochondriale et/ou les peroxisomes pour former de l'acétyl Coenzyme A (acétyl CoA). Cette molécule rejoint ensuite la voie de la phosphorylation oxydative pour produire de l'énergie. Le glycérol, quant à lui, est convertit en glycéraldéhyde-3-phosphate et s'intègre ainsi dans la glycolyse comme sucre intermédiaire. D'autre part, il est important de noter que le métabolisme lipidique joue un rôle essentiel dans le maintien de l'intégrité cellulaire, car les lipides sont les composants principaux des membranes cellulaires.

Quant aux protéines, elles possèdent des fonctions variées, elles peuvent être des composants structuraux, des enzymes, des transporteurs, des médiateurs impliqués dans les communications intercellulaires ou encore des anticorps. Du fait de leur rôle essentiels, les protéines sont peu utilisées à des fins énergétiques. Le catabolisme des acides aminés aboutit à la

production de molécules intermédiaires de la glycolyse et du cycle de krebs, telles que le pyruvate, l'acétyl-CoA, l' α -cétoglutarate, le fumarate, le succinyl-CoA ou l'oxaloacétate. Il permet également de produire de nouveaux acides aminés à partir d'autres acides aminés abondants, tels que la glutamine.

Les voies métaboliques du glucose, des lipides et des protéines convergent et s'interceptent au niveau de trois molécules clefs : le glucose-6-phosphate, l'acétyl CoA, et le pyruvate (figure 12).

Le glucose-6-phosphate intervient dans la glycolyse pour produire de l'énergie et régénérer le NAD⁺, la glycogénèse pour stocker l'excès de glucose intracellulaire, et la voie des pentoses phosphates pour régénérer le NADPH.

L'acétyl CoA, étant un point de convergence énergétique important, représente une molécule clef essentielle. En effet, la lipolyse, la glycolyse et le catabolisme des acides aminés aboutissent toutes *in fine* à la production d'acétyl CoA qui s'engage dans le cycle de krebs et produit de l'énergie *via* la phosphorylation oxydative (figure 11). Par ailleurs, l'acétyl CoA participe à la biosynthèse des lipides dans le cytosol. Il peut, par exemple, s'engager dans la synthèse des stérols ou des phospholipides pour constituer les membranes cellulaires, ou être retransformé en acides gras pour former des triglycérides, si l'énergie apportée par les substrats glycolytiques est suffisante.

Le pyruvate, issu de la glycolyse ou de la transamination de l'alanine, est quant à lui situé à l'intersection de nombreuses voies métaboliques. Il peut intégrer les voies anaboliques de la néoglucogénèse et de la biosynthèse des acides aminés, ainsi que les voies énergétiques que sont la glycolyse anaérobie et la phosphorylation oxydative (figure 13). En participant à la phosphorylation oxydative *via* sa conversion en acétyl CoA, le pyruvate permet de régénérer du NAD⁺ et de produire de l'énergie. Cette réaction est donc très finement régulée par le complexe multi-enzymes pyruvate déshydrogénase (PDH) [149], dans le but de maintenir l'homéostasie du glucose et la balance énergétique [150] (figure 11).

Ainsi, toutes les voies métaboliques (dont seulement les principales sont mentionnées dans ce manuscrit) s'interceptent au niveau de métabolites essentiels représentant les noeuds d'un vaste réseau métabolique régulé par un large panel d'enzymes [151].

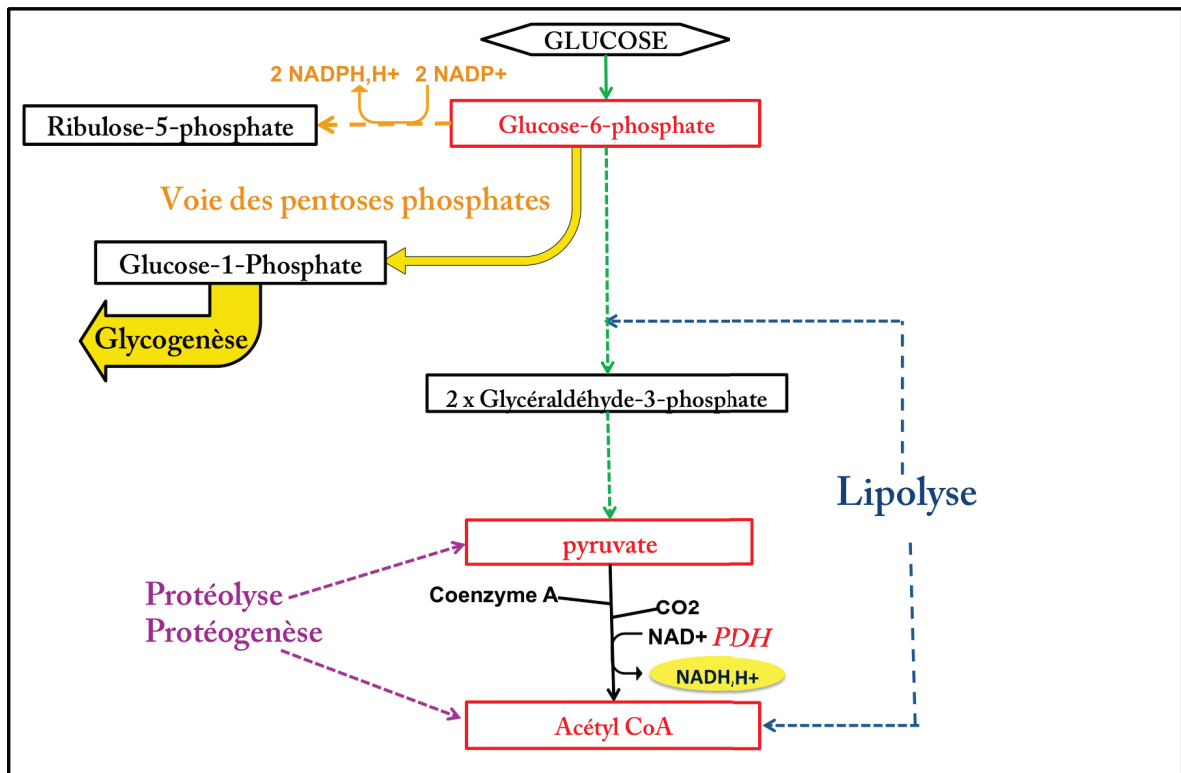


FIGURE 12 – Représentation des intersections et convergences des grandes voies métaboliques

Cette figure montre les principaux carrefours entre le métabolisme du glucose (en vert et jaune), des lipides (en bleu) et des acides aminés (en violet). Les molécules clés de ces intersections sont représentées en rouge. Les flèches en pointillées correspondent à des étapes intermédiaires non représentées sur ce schéma.

1.4.4 Le switch métabolique, moteur de la différenciation ?

Les cellules souches ont la capacité de s'auto-renouveler ou de se différencier. Chacune de ces deux activités nécessite de l'énergie, par conséquent les stocks en molécules énergétiques doivent être maintenus. Différents types de cellules souches et progénitrices, comme les ESC (cellules souches embryonnaires) [152], les MSC (cellules souches mésenchymateuses) [153] ou les HPSC (cellules souches et progénitrices hématopoïétiques) [154], reposent préférentiellement sur un métabolisme glycolytique anaérobie pour générer l'énergie nécessaire à leurs fonctions. La capacité d'auto-renouvellement de ces cellules est associée à une importante activité anabolique, consommatrice d'énergie. Un des principaux rôles de la glycolyse anaérobie dans ces cellules, pourrait donc consister à fournir, de part sa vitesse, suffisam-

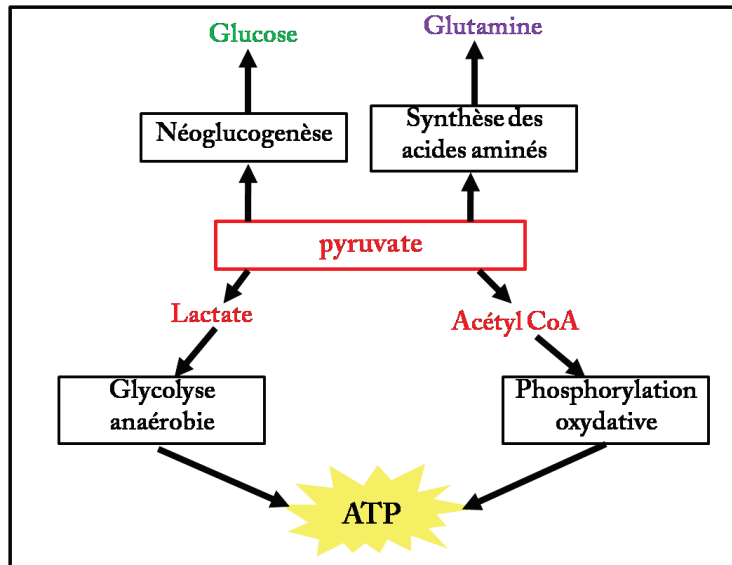


FIGURE 13 – Vision globale des destinées du pyruvate

Le pyruvate peut entrer dans la néoglucogénèse pour former du glucose, participer à la biosynthèse des acides aminés, notamment de la glutamine, ainsi qu'intégrer deux principales voies énergétiques, la glycolyse anaérobie et la phosphorylation oxydative.

ment d'intermédiaires glycolytiques carbonés pour supporter leurs activités de biosynthèse, nécessaire à leur croissance et divisions [155]. En revanche, la production d'énergie dans les cellules différenciées dépend davantage de la phosphorylation oxydative, moins rapide que la glycolyse anaérobie, mais dont le rendement énergétique est plus important [116, 141]. Néanmoins, quelques exceptions ont été relevées, dans lesquelles des cellules matures peuvent de nouveau majoritairement utiliser la voie de la glycolyse anaérobie pour répondre à leurs besoins énergétiques. Ces exceptions sont toutes liées à des processus particuliers, tel que le développement de la mémoire de la réponse immunitaire innée (qualifiée de "trained immunity") dans lequel les monocytes adoptent un métabolisme glycolytique aérobie après avoir été stimulés par la présence de microorganismes [114]. Les cellules musculaires, sont également capables d'osciller entre différents métabolismes énergétiques [156]. En cas d'effort physique court et très intense, la glycolyse anaérobie augmentera fortement pour assurer leurs fonctions, mais la chaîne respiratoire continuera tout de même de fonctionner à un taux plus faible. En effet, il est important de noter que toute voie énergétique n'en annule pas totalement une autre, toutes les voies fonctionnent en combinaison mais participent aux besoins

énergétiques cellulaires à des taux différents en fonction des demandes de l'organisme ou des apports en nutriments [156].

Bien que le métabolisme énergétique de certains types cellulaires matures varie pour répondre à leurs besoins fonctionnels, un switch métabolique de la glycolyse anaérobie vers la phosphorylation oxydative semble essentiel au processus de différenciation. Ainsi, contrairement aux cellules immatures, le métabolisme énergétique des cellules différenciées matures peut majoritairement reposer sur la phosphorylation oxydative, ou varier pour répondre à différentes activités cellulaires. Le processus de différenciation permettant d'atteindre cet état différencié fonctionnel, semble quant à lui s'accompagner d'un remaniement métabolique essentiel, correspondant à une transition de la glycolyse anaérobie vers la phosphorylation oxydative [157, 121, 158]. Il a été suggéré que ce switch métabolique pourrait jouer un rôle moteur dans la différenciation. Une analyse de l'expression différentielle des gènes, au cours de l'érythropoïèse, a par exemple montré que le profil d'expression transcriptionnel de progéniteurs érythrocytaires reflète une forte activité glycolytique de type anaérobie. De plus, à l'image des cellules cancéreuses, ce profil glycolytique est associé à un important taux de prolifération et de production de lactate cytosolique [159]. Or, l'expression d'un groupe de gènes communément impliqués dans le métabolisme glycolytique diminue de façon importante au cours de la différenciation [159]. Il a ainsi été proposé, qu'au cours de l'érythropoïèse et notamment lors de l'entrée des progéniteurs dans la circulation sanguine, leur métabolisme énergétique pourrait évoluer vers un métabolisme reposant davantage sur la phosphorylation oxydative [160]. Par ailleurs, une analyse moléculaire de progéniteurs neuronaux et de neurones a permis de dessiner un modèle transcriptionnel du switch métabolique, pouvant avoir lieu entre l'état d'auto-renouvellement et l'état de différenciation [121]. Pour subvenir à leurs besoins énergétiques, les neurones dépendent notamment de la phosphorylation oxydative, alors que les progéniteurs neuronaux reposent préférentiellement sur la glycolyse anaérobie. Les outils moléculaires utilisés dans cette étude sont variés, et ont permis de mesurer au cours de la différenciation neuronales le taux de certains métabolites comme le pyruvate, mais aussi le niveau d'expression des gènes du métabolisme par RNAseq. Les résultats de cette étude montrent que la différenciation des progéniteurs neuronaux en neurones s'accompagne d'une forte diminution de l'expression de LDHA, HK2, GLUT 1 et GLUT3, ainsi que de la sub-

stitution de l'isoforme 2 (PKM2), par l'isoforme 1 (PKM1) de la pyruvate kinase [121]. La diminution de ces enzymes glycolytiques au cours de la différenciation neuronale suggère donc le passage d'un métabolisme aérobie vers un métabolisme reposant sur la phosphorylation oxydative. Un tel switch métabolique a également été mis en évidence dans d'autres processus de différenciation, tels que la maturation des cryptes intestinales [158] ou encore la reprogrammation des iPSC [157]. Il est également intéressant de noter que la co-expression de LDHA et HK2 ne permet pas aux neurones de retrouver leur immaturité, l'expression de ces enzymes induisant leur apoptose [161, 121]. En effet, dans les cellules différenciées, la forte diminution des transporteurs du glucose ne permet pas de fournir la quantité de substrat suffisante pour alimenter la glycolyse anaérobie, dont le flux est augmenté par la présence d'HK2 et LDHA [121]. D'autre part, contrairement à l'inhibition des enzymes LDHA et HK2, la stimulation de la phosphorylation oxydative et de la biogenèse mitochondriale n'entraîne pas la mort cellulaire, mais semble plutôt induire la différenciation [121].

Ces résultats laissent suggérer que le métabolisme glycolytique pourrait jouer un rôle régulateur dans la balance entre l'état indifférencié et l'état différencié, comme il a été proposé dans le cas des cellules souches hématopoïétiques [162]. Par ailleurs, le switch métabolique s'effectue dès les premières heures de la différenciation, et pourrait donc être à l'origine de l'engagement des cellules dans ce processus [163].

Enfin, des études se sont intéressées aux mécanismes sous-jacents du switch métabolique. L'acétyl CoA, qui est le point de convergence de toutes les voies métaboliques, est un substrat important dans l'acétylation des histones. En conditions d'auto-renouvellement, cette molécule n'est pas utilisée par le cycle de krebs et la phosphorylation oxydative pour produire de l'énergie. Dans ce cas, elle est donc disponible pour l'acétylation des histones. Or, l'inhibition de l'acétyl CoA dans les cellules souches pluripotentes, entraîne une diminution de l'acétylation et induit de la différenciation spontanée [164]. Il semblerait ainsi qu'au cours de la différenciation, l'activation de la phosphorylation oxydative augmente la consommation de l'acétyl CoA cellulaire qui devient alors indisponible pour acétyler les histones. Cette diminution de l'acétylation s'accompagne alors, dès les premières heures de la différenciation, d'une diminution des marqueurs de pluripotence et de l'induction spontanée des cellules souches en différenciation [164]. Ce même scénario a également été proposé dans le cadre de l'activation

de la réponse immunitaire [165]. De plus, l'antimycine A, inhibiteur de la CTE au niveau du complexe III, semble bloquer la différenciation des cellules souches embryonnaires [147]. Il a effectivement été décrit pour ces cellules que leur immaturité se caractérise par une glycolyse anaérobie, remplacée après différenciation par la phosphorylation oxydative. Une étude a alors montré qu'en présence d'antimycine A la production de lactate augmente et l'expression des gènes impliqués dans la différenciation diminue. En contrepartie lorsque l'antimycine A est retirée, l'expression de ces gènes augmente mais l'expression des gènes de la pluripotence ne varie pas. Il semblerait donc qu'en empêchant la phosphorylation oxydative de fonctionner, l'antimycine A empêche les cellules de sortir de leur état de pluripotence, sans affecter la survie [147]. Ces résultats sont soutenus par différentes études démontrant également, sur d'autres modèles de cellules souches ou progénitrices, que le blocage de la CTE compromet la capacité des cellules immatures à se différencier [166, 167].

L'ensemble de ces résultats renforce ainsi l'hypothèse de plus en plus probable, selon laquelle le switch métabolique, de la glycolyse anaérobie vers la phosphorylation oxydative, est une force motrice de l'engagement en différenciation [125].

1.5 Différenciation cellulaire, SEG et métabolisme glycolytique

L'activité du réseau métabolique est représenté par les concentrations en métabolites ainsi que par les vitesses et flux de transformations de ces éléments en autres métabolites. Le réseau semble principalement régulé par le RRG, l'épigénome et les facteurs environnementaux, tels que la concentration en nutriments. Les enzymes agissent comme des régulateurs du métabolisme en contrôlant les flux et vitesses des voies cataboliques ou anaboliques, en fonction des besoins cellulaires [107]. Or, comme les autres protéines cellulaires, leur taux d'expression, c'est à dire la quantité d'enzyme présente dans la cellule, dépend de l'état du RRG, suggérant que le métabolisme est en effet sous contrôle du réseau d'expression génique. Cependant, l'activité de ces enzymes peut ensuite être régulée par les métabolites, eux mêmes, comme dans le cas de la régulation de l'enzyme PFK-1, mentionnée dans les sections précédentes [109]. Il semblerait alors que le réseau métabolique puisse dans une certaine mesure s'auto-réguler pour maintenir l'homéostasie cellulaire. Enfin, de plus en plus d'études démontrent que les métabolites sont également des substrats permettant de générer

des modifications au niveau de la chromatine et de l'ADN. Ces modifications peuvent être de différents types, telles que des phosphorylations, ubiquitinylation, succinylations, ou acétylations, pouvant être générées ou supprimées par des enzymes dépendant de substrats ou co-facteurs dérivés des voies métaboliques. Il a par exemple été montré que la concentration du transporteur d'électron NAD^+ est proportionnelle à la cinétique enzymatique de plusieurs histones déacétylases [168]. Se met alors en place une compétition entre les enzymes épigénétiques et les métabolites nécessitant ces substrats [168]. D'autre part, il est observé dans certains cancers que des mutations dans les gènes codants pour des enzymes métaboliques, entraînent des modifications des marques épigénétiques, car elles déclenchent une accumulation de métabolites utilisant des substrats communs avec des enzymes épigénétiques [169]. Le surplus anormal de ces métabolites séquestre ainsi les substrats pour leur propre activité, les rendant indisponibles pour les enzymes épigénétiques qui ne peuvent plus assurer leurs fonctions.

Par ailleurs, dans une vision systémique, les modifications épigénétiques réguleraient la variabilité de l'expression génique [170], en imposant des contraintes trajectorielles dans l'espace multi-dimensionnel des gènes pour transiter d'un état cellulaire à un autre. Selon le paysage épigénétique de Waddington, des vallées représentent les états cellulaires, étant séparées par des barrières à franchir, pour passer d'un état cellulaire à un autre [96]. Ainsi, les substrats métaboliques, de part leur disponibilité, peuvent influencer l'activité des enzymes épigénétiques et créer dans le paysage épigénétique de nouvelles vallées, correspondant à un nouvel état cellulaire (figure 14) [168].

Parmi les substrats métaboliques dont dépendent les enzymes modifiant l'épigénome, il est important de citer le NAD^+ et l'acétyl CoA qui est impliqué dans l'acétylation des histones, processus critique pour la maintenance des cellules souches en auto-renouvellement [171, 164]. Les médiateurs de cette modification épigénétique sont les histones acétyltransférases (HATs) dont le rôle est de transférer le groupe acétyl de l'acétyl CoA sur les lysines des histones, et de catalyser leur acétylation. Ainsi l'activité des HATs est sensible aux fluctuations de concentration cellulaire de l'acétyl CoA [171]. Comme cité en amont, le processus de différenciation s'accompagne d'une diminution de l'acétylation des histones, de part l'indisponibilité de l'acétyl CoA, recruté pour le cycle de krebs et la phosphorylation oxydative [164]. Ces résultats

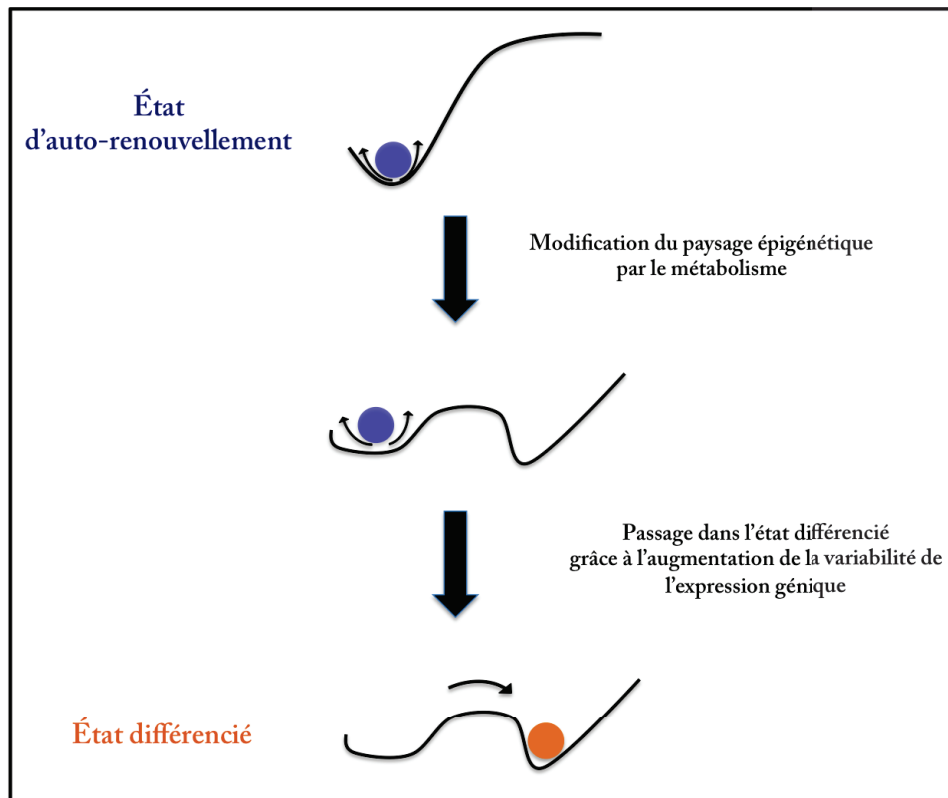


FIGURE 14 – Implications possibles du remaniement métabolique intégré dans le paysage épigénétique de Waddington au cours de la différenciation

Le paysage épigénétique de Waddington est représenté par les courbes noires. La cellule est représentée par un cercle de couleur. Le remaniement métabolique pourrait permettre d'abaisser les barrières épigénétiques et ainsi d'augmenter la variabilité de l'expression des gènes, représentés par l'abaissement de la courbe noire et l'agitation de la cellule dans sa vallée initiale (l'état d'auto-renouvellement). De plus, le remaniement métabolique conduirait à un remodelage épigénétique représenté par la création d'une nouvelle vallée. La variabilité de l'expression des gènes pourrait alors permettre à la cellule de se déplacer jusqu'à cette nouvelle vallée (l'état différencié)

suggèrent ainsi que le switch métabolique est impliqué dans la régulation du paysage épigénétique. Par ailleurs l'activité d'histones déacétylases, les sirtuins (SIRT), mais également des PARPs (Poly-ADP-ribose polymerase), dépendent du niveau de NAD^+ cellulaire, les reliant ainsi au métabolisme. Le NAD^+ , quant à lui, est un co-facteur enzymatique régulé par les voies métaboliques, notamment la glycolyse anaérobie en état d'auto-renouvellement, mais est également une molécule de signalisation. Les sirtuins sont capables de détecter le ratio NAD^+/NADH et de transformer les changements de niveau de ce ratio en signal [172]. SIRT1

est par exemple impliquée dans la maintenance de l'état pluripotent des cellules souches embryonnaires *via* la répression de plusieurs gènes du développement [173]. Son rôle a également été suggéré dans le processus de différenciation des cellules neuronales [174]. D'autre part, la régénération du NAD⁺ est essentielle au maintien de l'état indifférencié [175]. Il a été montré que la reprogrammation des iPSC et la maintenance de ces cellules et des ESC, sont améliorées par une supplémentation en précurseurs du NAD⁺, et altérées lorsque le NAD est dégradé [175]. Il a de plus été décrit qu'au cours du processus de différenciation le niveau de NAD⁺ tendait à diminuer, tout comme le niveau d'activité des sirtuins, et notamment de SIRT1 [176, 177]. Ainsi, il a été suggéré que le switch métabolique et les changements du paysage épigénétique conjointement joueraient un rôle central dans les processus de différenciation cellulaire [178].

Dans une vision systémique du processus de différenciation, il est ainsi important de porter un regard critique sur le lien entre les changements du réseau métabolique, les changements épigénétiques et la variabilité de l'expression génique. Cependant, des données restent encore à acquérir concernant la cinétique et la synchronisation de ces événements à l'échelle des cellules individuelles, au cours de la différenciation.

Dans cette thèse, nous nous intéresserons donc à la variabilité de l'expression génique et aux remaniements métaboliques au cours de la différenciation érythrocytaire. À la lumière de nos résultats et de la littérature, nous discuterons alors du lien entre l'expression génique et le métabolisme, et de leur rôle probablement synergique dans le processus de différenciation.

2 Analyse de la variabilité de l'expression génique en cellule unique au cours du processus de différenciation

2.1 Article 1 : Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process

Comme présenté dans l'introduction bibliographique, différents modèles théoriques suggèrent en effet que le processus de différenciation s'accompagne d'une augmentation de la variabilité de l'expression génique [24, 25, 26].

Dans ce contexte, la première partie de mon doctorat a consisté à analyser expérimentalement la variabilité de l'expression génique au cours du processus de différenciation érythrocytaire, grâce à des technologies récentes permettant d'accéder au contenu moléculaire de plusieurs cellules, individuellement. Cette analyse a fait l'objet d'une publication dans le journal PLoS Biology (article 1) [95].

2.1.1 Principaux résultats de l'article 1

Cette étude avait pour objectif d'analyser la variabilité de l'expression des gènes à l'échelle de la cellule unique, au cours du processus de différenciation érythrocytaire.

Dans un premier temps, l'expression de 110 gènes, sélectionnés sur la base d'une analyse RNASeq préalablement réalisée dans l'équipe, a été mesurée par RTqPCR dans des populations de progéniteurs érythrocytaires, les T2EC [179], en état d'auto-renouvellement (0h), ou induites en différenciation depuis 8h, 24h, 48h et 72h. Une analyse en composante principale (ACP) a permis de séparer les populations de T2EC, qui se regroupent en fonction de leurs temps respectifs de différenciation, et s'ordonnent le long de la première composante principale (figure 1A dans l'article 1). Cette analyse suggèrent ainsi que le processus de différenciation est la principale source de variabilité à l'échelle de la population cellulaire et que les gènes sélectionnés sont pertinents pour analyser le processus de différenciation.

Afin d'analyser dans un second temps les variations de l'expression génique à l'échelle de la cellule unique dans un deuxième temps, nous avons utilisé une approche reposant sur un

système de microfluidique, permettant de mesurer simultanément l'expression de 96 gènes dans 96 cellules individuelles par RTqPCR. Nous avons ainsi restreint notre analyse à 92 gènes afin de pouvoir inclure quatre contrôles.

L'expression de ces 92 gènes a été analysée dans des T2EC individuelles isolées à 0h, 8h, 24h, 33h, 48h et 72h du processus de différenciation. Contrairement aux résultats obtenus à l'échelle de la population, l'ACP ne permet pas ici de séparer clairement les T2EC en fonction de leur progression dans la différenciation (figure 2A de l'article 1). Le processus de différenciation ne semble donc pas être la première source de variabilité à l'échelle de la cellule unique. Une ACP réalisée uniquement aux points extrêmes, 0h et 72h, permet en revanche de séparer plus nettement les T2EC à 0h des T2EC à 72h de différenciation (figure 3A de l'article 1). Parmi les 92 gènes analysés, *LDHA*, qui est l'un des principaux gènes permettant de distinguer les T2EC en auto-renouvellement des T2EC à 72h de différenciation, a particulièrement retenu notre attention. Une étude plus approfondie de l'expression de *LDHA* au cours du processus de différenciation, a donné lieu à la deuxième partie de mon doctorat, et sera donc décrite dans la prochaine partie de ce manuscrit.

D'autre part, pour chaque temps de différenciation étudié, nous avons calculé un taux de corrélation entre tous les gènes, deux à deux (figure 5B de l'article 1). Pour chaque temps de différenciation, nous avons alors identifié trois noeuds principaux, correspondant aux gènes présentant le plus grand nombre de corrélations significatives (figure 5C de l'article 1). Il est intéressant de noter que l'identité de ces principaux noeuds varient au cours de la différenciation, et qu'aux temps 8h et 24h *LDHA* est l'un d'entre eux. D'autre part, nous avons constaté que le nombre de corrélations significatives global chute à 8h, avant d'augmenter à nouveau progressivement jusqu'à 72h (figure 5A de l'article 1).

Afin de tester l'hypothèse selon laquelle l'engagement en différenciation s'accompagne d'une forte augmentation de la variabilité de l'expression génique, nous avons utilisé la mesure de l'entropie. Ainsi, pour chaque temps de différenciation, nous avons calculé une valeur d'entropie par gène et comparé la distribution de ces valeurs au cours de la différenciation (figure 8 de l'article 1). Nos résultats montrent que l'entropie, augmente significativement à 8h, reste stable jusqu'à 24h, et diminue progressivement jusqu'à 72h, suggérant ainsi que la différenciation érythrocytaire s'accompagne d'un pic de la variabilité d'expression génique

intercellulaire à 8h-24h du processus.

Nous nous sommes alors intéressés aux causes potentielles de l'augmentation de cette variabilité, telles que le cycle et la taille cellulaires, ainsi que l'asynchronie des cellules dans le processus de différenciation. En ce qui concerne le cycle cellulaire, nous avons montré que la répartition des cellules dans les phases du cycle cellulaire n'est pas significativement différente entre 0h et 8h de différenciation (figure 9B de l'article 1). Le cycle cellulaire ne semble donc pas être à l'origine de l'augmentation de la variabilité génique à 8h-24h de différenciation. Dans un deuxième temps, la taille des cellules a été approchée par le paramètre FSC, mesuré par cytométrie en flux, montrant que la variabilité de la taille des cellules augmente de façon importante à 48h de différenciation, autrement dit, en aval de l'augmentation de l'entropie (figure 10B de l'article 1). L'hétérogénéité dans la taille des cellules ne semble donc pas être une cause, mais plutôt une conséquence de l'augmentation de la variabilité de l'expression génique intercellulaire. Une étude menée dans l'équipe afin de répondre de façon plus approfondie aux interrogations concernant l'influence de la taille des cellules et du cycle cellulaire sur la variabilité de l'expression génique sera présentée dans la prochaine section. Différentes approches *in silico*, nous ont ensuite permis d'appréhender l'effet de l'asynchronie des cellules dans le processus de différenciation sur la variabilité d'expression génique. Les résultats obtenus par les différents algorithmes suggèrent que l'asynchronie des cellules n'est pas une source de l'augmentation de la variabilité d'expression génique, observée à 8-24h de différenciation.

Les résultats de l'analyse de l'expression génique en cellule unique, ont permis de mettre en lumière les points 8h et 24h du processus de différenciation. Nous avons alors émis l'hypothèse selon laquelle 24h pourrait être le point de non-retour dans le processus de différenciation. Afin de tester cette hypothèse, les T2EC ont été induites à se différencier pendant 24h ou 48h, puis remises dans le milieu d'auto-renouvellement. Les résultats de cette expérience montrent qu'après 24h de différenciation, les cellules sont encore capables de s'auto-renouveler, alors qu'après 48h, elles ne prolifèrent plus (figure 10A de l'article 1). Le point de non-retour de l'engagement des T2EC dans le processus de différenciation semble donc se situer entre 24h et 48h.

2.1.2 Principales conclusions de l'article 1

L'ensemble des résultats de cette analyse a permis de montrer que l'engagement en différenciation des progéniteurs érythrocytaires est précédé d'un pic de variabilité de l'expression génique intercellulaire ainsi que d'une diminution des corrélations entre gènes. Ces observations semblent donc en accord avec les nouveaux modèles théoriques, suggérant que le processus de différenciation s'accompagne d'une augmentation de la variabilité de l'expression des gènes. Ainsi, les résultats obtenus dans cette première partie de mon doctorat, ont permis de proposer une nouvelle vision du processus de différenciation, reposant sur le comportement d'un réseau de gènes dynamique sous-jacent, et non pas sur des événements moléculaires exécutés de façon identique par toutes les cellules de la population, comme cela avait été suggéré auparavant.

2.1.3 Article 1

Publié le 27 Décembre 2016 dans le journal PLoS Biology

URL : <https://doi.org/10.1371/journal.pbio.1002585>

RESEARCH ARTICLE

Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process

Angélique Richard¹, Loïs Boullu^{2,3,4}, Ulysse Herbach^{1,2,3}, Arnaud Bonnafoux^{1,2,5}, Valérie Morin⁶, Elodie Vallin¹, Anissa Guillemin¹, Nan Papili Gao^{7,8}, Rudiyanto Gunawan^{7,8}, Jérémie Cosette⁹, Ophélie Arnaud¹⁰, Jean-Jacques Kupiec¹¹, Thibault Espinasse³, Sandrine Gonin-Giraud¹, Olivier Gandrillon^{1,2}*



CrossMark
click for updates

1 Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratory of Biology and Modelling of the Cell, 46 allée d'Italie Site Jacques Monod, F-69007, Lyon, France, **2** Inria Team Dracula, Inria Center Grenoble Rhône-Alpes, France, **3** Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan 43 blvd du 11 novembre 1918, F-69622 Villeurbanne-Cedex, France, **4** Département de Mathématiques et de statistiques de l'Université de Montréal, Pavillon André-Aisenstadt, 2920, chemin de la Tour, Montréal (Québec) H3T 1J4 Canada, **5** The CoSMo company, 5 passage du Vercors – 69007 LYON – France, **6** Univ Lyon, Univ Claude Bernard, CNRS UMR 5310 - INSERM U1217, Institut NeuroMyoGène, F-69622 Villeurbanne-Cedex, France, **7** Institute for Chemical and Bioengineering, ETH Zurich, Zurich, Switzerland, **8** Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Genopode, 1015 Lausanne Switzerland, **9** Genethon – Institut National de la Santé et de la Recherche Médicale – INSERM, Université d'Evry-Val-d'Essonne – 1 rue de l'internationale 91000 Evry, France, **10** RIKEN - Center for Life Science Technologies (Division of Genomic Technologies)—CLST (DGT), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan, **11** INSERM, Centre Cavallès, Ecole Normale Supérieure, F-75005 Paris, France

* These authors contributed equally to this work.

* Olivier.Gandrillon@ens-lyon.fr

OPEN ACCESS

Citation: Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. (2016) Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process. *PLoS Biol* 14(12): e1002585. doi:10.1371/journal.pbio.1002585

Academic Editor: Sarah A. Teichmann, EMBL-European Bioinformatics Institute & Wellcome Trust Sanger Institute, UNITED KINGDOM

Received: June 6, 2016

Accepted: September 22, 2016

Published: December 27, 2016

Copyright: © 2016 Richard et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The eight RNA-seq libraries (raw sequences, 4 conditions, 2 libraries per condition: paired-end libraries) have been deposited at SRA and are available at: <http://www.ncbi.nlm.nih.gov/sra/SRP076011>. The resulting counting table (matrix_2793.txt), the list of the resulting 424 differentially expressed genes (424_differential_genes.txt), the raw Fluidigm data for cell populations (Pop_1361941161.csv and Pop_VM117_AR2_03-03-14.csv), the raw Fluidigm

Abstract

In some recent studies, a view emerged that stochastic dynamics governing the switching of cells from one differentiation state to another could be characterized by a peak in gene expression variability at the point of fate commitment. We have tested this hypothesis at the single-cell level by analyzing primary chicken erythroid progenitors through their differentiation process and measuring the expression of selected genes at six sequential time-points after induction of differentiation. In contrast to population-based expression data, single-cell gene expression data revealed a high cell-to-cell variability, which was masked by averaging. We were able to show that the correlation network was a very dynamical entity and that a subgroup of genes tend to follow the predictions from the dynamical network biomarker (DNB) theory. In addition, we also identified a small group of functionally related genes encoding proteins involved in sterol synthesis that could act as the initial drivers of the differentiation. In order to assess quantitatively the cell-to-cell variability in gene expression and its evolution in time, we used Shannon entropy as a measure of the heterogeneity. Entropy values showed a significant increase in the first 8 h of the differentiation process, reaching a peak between 8 and 24 h, before decreasing to significantly lower values. Moreover, we observed that the previous point of maximum entropy

data for single cells (0 to 8 hours kinetics: Single_AR78_1_to_2.csv; 0 to 72 kinetics: Single_AR85_1_to_6.csv) the actinomycin D experiment (export_RNA_deg_exp_Diff_0h.csv; export_RNA_deg_exp_Diff_24h.csv; export_RNA_deg_exp_Diff_72h.csv), as well as data for Figures are available at osf.io/k2q5b (DOI [10.17605/OSF.IO/K2Q5B](https://doi.org/10.17605/OSF.IO/K2Q5B)).

Funding: This work was supported by funding from the Institut Rhônalpin des Systèmes Complexes (IXXI), La Ligue contre le Cancer (comité de Haute-Savoie) and by two grants from the French agency ANR (Stochagene; ANR 2011 BSV6 014 01 and ICEBERG; ANR-IABI-3096). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: CV, coefficient of variation; DNB, dynamical network biomarker; HCA, hierarchical cluster analysis; LDHA, Lactate dehydrogenase A; OSC, oxidosqualene cyclase; PC1, first principal component; PCA, principal component analysis; RT-qPCR, reverse transcription quantitative PCR; SNE, Stochastic Neighbor Embedding.

precedes two paramount key points: an irreversible commitment to differentiation between 24 and 48 h followed by a significant increase in cell size variability at 48 h. In conclusion, when analyzed at the single cell level, the differentiation process looks very different from its classical population average view. New observables (like entropy) can be computed, the behavior of which is fully compatible with the idea that differentiation is not a “simple” program that all cells execute identically but results from the dynamical behavior of the underlying molecular network.

Author Summary

The differentiation process has classically been seen as a stereotyped program leading from one progenitor toward a functional cell. This vision was based upon cell population-based analyses averaged over millions of cells. However, new methods have recently emerged that allow interrogation of the molecular content at the single-cell level, challenging this view with a new model suggesting that cell-to-cell gene expression stochasticity could play a key role in differentiation. We took advantage of a physiologically relevant avian cellular model to analyze the expression level of 92 genes in individual cells collected at several time-points during differentiation. We first observed that the process analyzed at the single-cell level is very different and much less well ordered than the population-based average view. Furthermore, we showed that cell-to-cell variability in gene expression peaks transiently before strongly decreasing. This rise in variability precedes two key events: an irreversible commitment to differentiation, followed by a significant increase in cell size variability. Altogether, our results support the idea that differentiation is not a “simple” series of well-ordered molecular events executed identically by all cells in a population but likely results from dynamical behavior of the underlying molecular network.

Introduction

The classical view of a linear differentiation process driven by the sequential activation of master regulators [1] has been increasingly challenged in the last few years both by experimental findings and theoretical considerations.

Thanks to the recent development in single-cell profiling technologies, researchers are now able to investigate qualitatively and quantitatively the cell-to-cell variability in gene expression in more detail. In this context, several experimental studies at single-cell level involving the regulation of self-renewal and differentiation processes in embryonic stem cells [2–8] and the generation of induced pluripotent stem cells [9] have shown that gene expression variability might be involved in cell differentiation. To support this claim, recent researches on hematopoietic stem cells highlighted the role of molecular heterogeneity in differentiation [10, 11]. Further evidence was also obtained during an *ex vivo* differentiation process [12], and in the generation of cells of the immune system [13–18].

The overt cell-to-cell variability is deeply rooted in the inherent stochasticity of the gene expression process [19–23]. Numerous explanations have been put forward regarding the molecular and cellular sources for such variability (see [24] and references therein). Some of those causes involve biophysical processes (e.g., the random partitioning during mitosis, as

discussed in [25]), whereas others are more related to biochemical regulation (e.g., the dynamical functioning of the intracellular network [26] or the chromatin dynamics [27]).

At least three models of cell differentiation based on stochastic gene expression have been proposed, in which a peak in the gene expression variability is expected to occur. In the first model, stochastic gene expression is the driving force of cell differentiation that generates cell type diversity, on which a selective constraint is then exerted [28]. In the second model, noise in gene expression causes bifurcations in the dynamics of gene regulatory networks [21]. In the third model, cell differentiation is viewed as a dynamical process in which differentiating cells are thought of as particles moving around in a state space [29, 30]. This formal space can be used to display gene expression patterns. Hence, when some parameters that describe gene regulatory interactions change, the cell particle “moves” in the state space. In this view, discrete identified cell states (e.g., self-renewing, differentiated) correspond to different regions of this space that could be seen as different attractor states. The transition process between attractors therefore first requires the exit from the original state that may be fueled by an increase in gene expression stochasticity [31]. Regardless of the differences between these models, they all assume that the differentiation process is represented by cell trajectories leading from one state to another through a phase of biased random walk in gene expression. This phase is followed by stabilization (convergence) toward a particular pattern of gene expression corresponding to a stable attractor state, the differentiated final state, in which noisy fluctuations of gene expression is minimized by the stabilizing effect of the attractor. Therefore, changes in the extent of cell-cell variability could be a new observable metric to characterize the cell differentiation process.

The purpose of the present study was then to assess whether gene expression variability changes during the differentiation process, as suggested by the above-quoted models, and whether such variation concurs with any physiological cellular change. We investigated the extent of gene expression variability at the single-cell level, both before and during the cell differentiation process. To do this, we analyzed the differentiation process of T2EC, which is an original cellular system consisting of non-genetically modified avian erythrocytic progenitor cells grown from a primary culture [32]. These cells can be maintained *ex vivo* in a self-renewal state under a combination of growth factors (TGF- α , TGF- β , and dexamethasone) and can also be induced to differentiate exclusively toward erythrocytes by changing the combination of the external factors present in the medium. The primary cause for differentiation is therefore known and relies upon change in the information carried by the extracellular environment. The differentiation process in those cells has been previously analyzed at the population level [33–35].

We first selected a pool of 110 relevant genes on the basis of RNA-Seq analysis performed on populations of T2EC in self-renewal state or induced to differentiate for 48 h. Multivariate statistical analysis of the data allowed us to select 92 genes for further analysis. We then performed high-throughput reverse transcription followed by reverse transcription quantitative PCR (RT-qPCR) of the 92 selected genes on single-cells collected at six time-points of differentiation. Several dimensionality reduction algorithms were used to visualize trends in the datasets. In agreement with the above hypothesis, cell heterogeneity, as measured by entropy, significantly increased during the first hours of the differentiation process and reached a maximal value at 8 to 24 h before decreasing toward the end of the process. The peak in entropy preceded an increase in cell size variability at 48 h. These observations suggested that 24 h is a crucial turning point in the erythrocytic differentiation process, which was experimentally verified by showing that T2EC committed irreversibly to the differentiation process between 24 h and 48 h.

Results

Identification of Differentially Expressed Genes Between Self-Renewing and Differentiating Progenitors

In order to identify a pool of genes potentially relevant in the differentiation process, we analyzed the transcriptome of self-renewing and differentiating primary chicken erythrocytic progenitor cells (T2EC) using RNA-Seq. We sequenced two independent libraries from self-renewing T2EC and two independent libraries from T2EC induced to differentiate for 48 h. For each condition, we first verified that read counts between replicates were reproducible (S3A and S3B Fig). We then identified 424 significantly differentially expressed genes (p -value < 0.05, S3C Fig). Gene ontology analysis using the DAVID database [60] revealed a clear over-representation of genes involved in sterol biosynthesis in this list (not shown). This finding was in line with our previous analysis showing that the oxydosterol cyclase (OSC), which is involved in cholesterol synthesis, is required to maintain self-renewal in T2EC [35]. However, no other over-represented function emerged from the present analysis.

Identification of Genes Relevant to Analyze the Erythrocytic Differentiation Process

To identify a smaller subset of relevant genes for further analysis by RT-qPCR using the Fluidigm array (see below), we tested 56 down-regulated and 77 up-regulated genes among the above 424 genes differentially expressed in self-renewing versus differentiating cells, which had the smallest set of p -values. We also included 32 non-regulated genes, selected among the most invariant ones. We then measured the expression of these 165 genes first using RNA from bulk cell populations taken at five time-points during differentiation (0, 8, 24, 48, and 72 h). Based on qPCR primer efficiency, 55 genes were removed (see Materials and Methods), which left a total of 110 genes for the subsequent analysis.

A principal component analysis (PCA) on the bulk gene expression levels (Fig 1A) showed a clear separation of the time-point 0 h (self-renewal) from the differentiation time-points. Samples along the differentiation process were well ordered according to the first principal component (PC1). PC1 explained 56.2% of the data variability suggesting that the differentiation process is the main source of variability at the population level for the selected genes.

We also performed a hierarchical cluster analysis (HCA), which again showed a clear arrangement of the samples according to their position along the differentiation process (Fig 1B). We further noticed that the gene expression patterns at 0, 8, and 24 h time-points were more similar to each other, while those at 48 h and 72 h time-points were also more similar to each other.

Thus, the 110 selected genes allowed us to clearly distinguish cell populations according to their progression along the differentiation sequence, indicating that they were relevant for analyzing this process. However, since the single-cell measurement technology used in this study could only accommodate 92 genes (not including two spikes and two repeats for the *RPL22L1* gene), we further refined our gene choice by performing a K-means clustering on the above data. The algorithm grouped genes based on their expression profile, and identified seven different gene clusters with respect to expression kinetics (S4 Fig).

The patterns mainly showed decreasing or increasing gene expressions during the differentiation process, while one cluster displayed a more complex dynamic (cluster 4). The latter was composed of genes whose expression decreased during the first 8 h, then increased and stabilized between 24 h and 48 h, before decreasing again until 72 h. Interestingly, all genes belonging to this cluster were linked by their involvement in sterol biosynthesis, reinforcing the

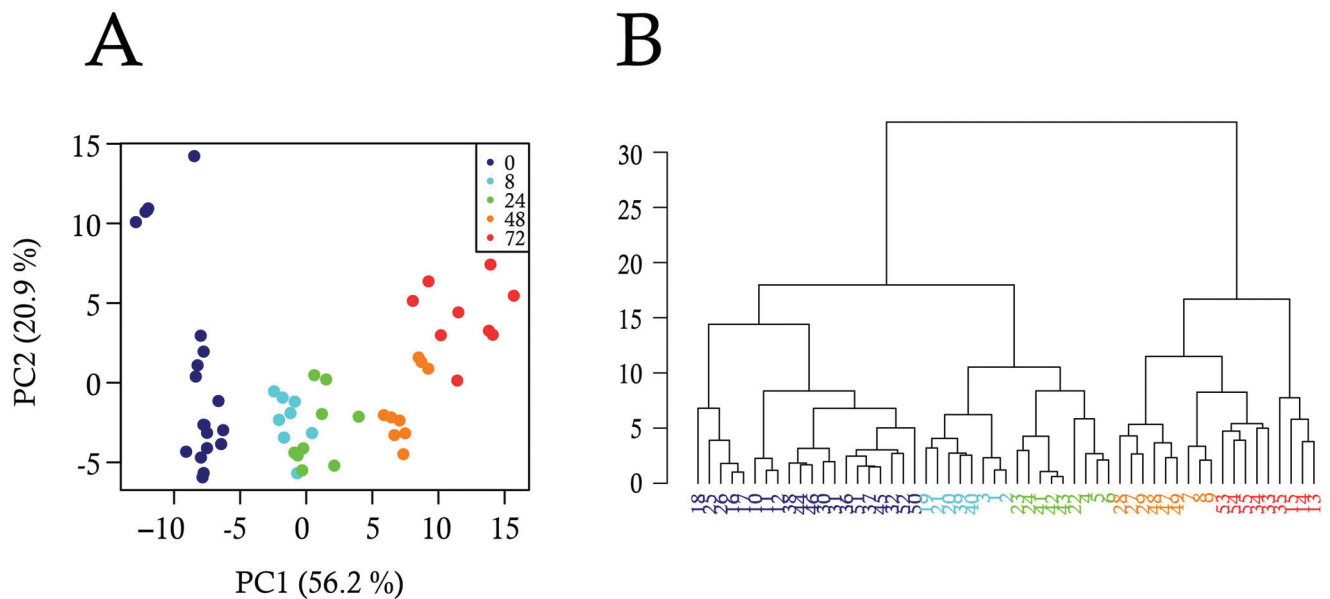


Fig 1. Analysis of bulk-cell gene expression during the differentiation process. Gene expression data were produced by RT-qPCR in triplicate from three independent T2EC populations collected at five differentiation time-points (0 h, 8 h, 24 h, 48 h, 72 h). The expression level of 110 genes (18 invariants, 50 down-regulated and 42 up-regulated) was analyzed by two different multivariate statistical methods: (A) Principal component analysis (PCA), and (B) Dendrogram resulting from hierarchical cluster analysis (HCA). The dots in (A) and leaves in (B) indicate the different cell populations and the colors indicate the differentiation time-points at which they were collected.

doi:10.1371/journal.pbio.1002585.g001

previously noted role of this pathway in erythroid differentiation. Based on the result of K-means clustering, we selected around thirteen genes per group to represent each cluster equally. This left us with 92 genes for further analysis (S1 Table).

We then used STRING database to search for known connections among these genes. The result confirmed the existence of a strongly connected subnetwork associated with sterol synthesis (S5B Fig). Moreover, this analysis also revealed the presence of another highly connected subnetwork mostly composed of genes involved in signaling cascades and two transcription factors (BATF and RUNX2). Those two main networks are linked by the gene *HSP90AA1* which encodes the molecular chaperone *HSP90alpha*. Its activity is not only involved in stress response but also in many different molecular and biological processes because of its important interactome. *HSP90alpha* represents 1%–2% of total cellular protein in unstressed cells. Interestingly, *HSP90alpha* level is up-regulated and correlated with poor disease prognosis in leukemia [61]. *HSP90alpha* has also been shown to be involved in the survival of cancer cells in hypoxic conditions [62].

Cell-to-Cell Heterogeneity Blurred Cell Differentiation Process

We measured the expression level of the selected 92 genes by single-cell RT-qPCR using 96 cells isolated from the most informative time-points of the differentiation sequence. Based upon preliminary experiments, we decided to analyze cells from six time-points during differentiation. After data cleaning (see Materials and Methods), we obtained the expression level of 90 genes in 55, 73, 72, 70, 68, and 51 single cells from 0, 8, 24, 33, 48, and 72 h of differentiation, respectively.

One should note that the variability we observed at the single-cell level originates from two types of sources: biological sources and experimental sources. We therefore tested the

technical reproducibility of different RT-qPCR steps liable to generate such experimental noise (see [Materials and Methods](#)). As expected, reverse transcription (RT) was the main source of experimental variability, since pre-amplification and qPCR steps brought negligible amount of variability ([S1 Fig](#)). Moreover, using external RNA spikes controls whose Cq value depends only on the experimental procedure, we noted that technical variability was negligible compared to the biological variability (see [Materials and Methods](#)). Quality control (see [Materials and Methods](#)) led to the elimination of 2 genes, letting us with 90 genes for subsequent analysis.

We first used PCA on the single-cell expression of these 90 genes ([Fig 2A](#)). In contrast to the whole-population data, the single-cell data did not immediately demarcate into well-separated clusters. The differentiation process was most apparent by looking at the second principal component (PC2), which explained 9.9% of the variability in the dataset. Hence, unlike in the population-averaged data, the differentiation process did not represent the main source of variability at the single-cell level.

The application of HCA further confirmed that the classification became more complex for single-cell data ([Fig 2B](#)). Contrary to bulk analysis, individual cells from the same time-point

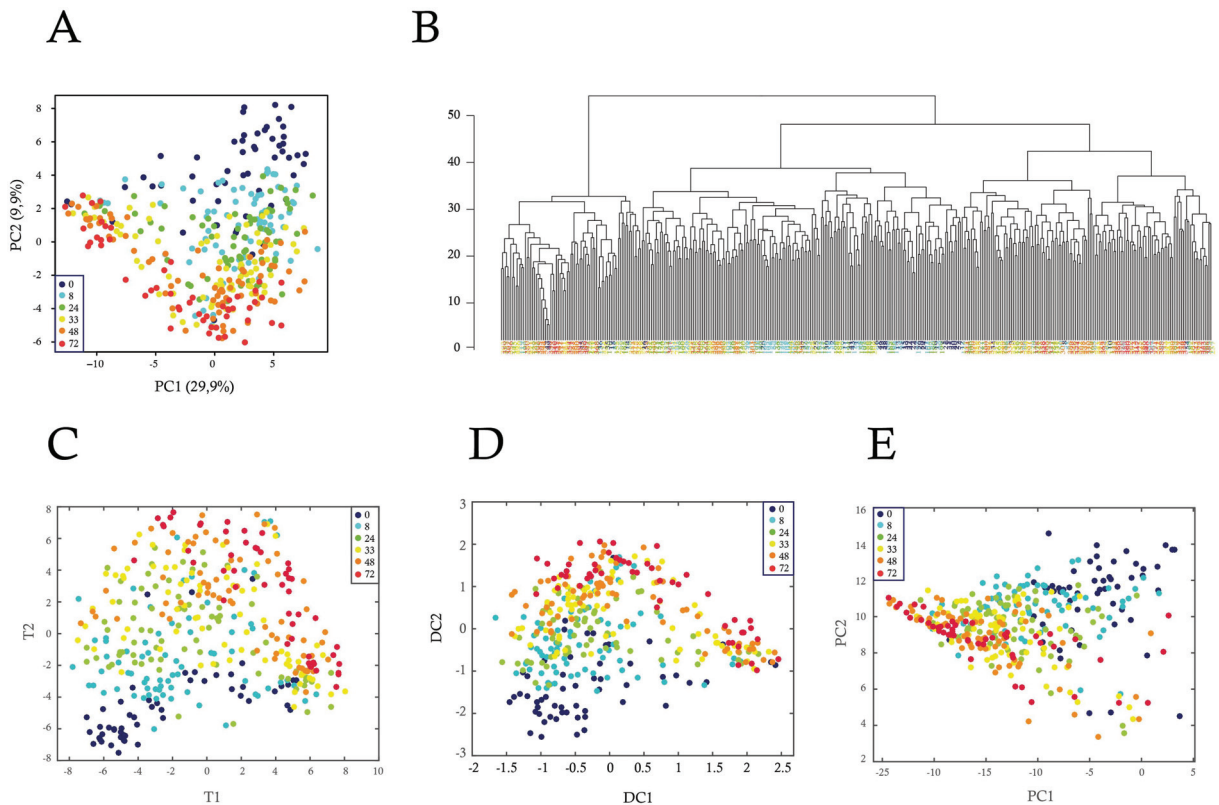


Fig 2. Analysis of single-cell gene expression during the differentiation process. Gene expression data were produced by RT-qPCR from individual T2EC collected at six differentiation time-points (0, 8, 24, 33, 48, and 72 h). The expression of 90 genes was analyzed in single-cells by five different multivariate statistical methods: (A) Principal component analysis (PCA), (B) Hierarchical cluster analysis (HCA), (C) t-SNE, (D) Diffusion map, and (E) kernel PCA. The dots in (A, C, D, and E) and leaves in (B) indicate the single-cells, and the colors indicate the differentiation time-points at which they were collected. t-SNE analysis was performed using the following parameters: initial_dims = 30; perplexity = 60. Diffusion map was run using the following parameters: no_dims = 4, t = 1, and sigma = 1000. Kernel PCA was run with a parameter for computing the “poly” and “gaussian” kernel of 0.1. Only the first two dimensions are plotted.

doi:10.1371/journal.pbio.1002585.g002

were not necessarily more similar to each other than to cells from neighboring time-points. Consequently, the clustering of individual cells into groups became complicated. The picture of cell differentiation process that emerged from the single-cell analysis thus far was more complex than the one obtained from the population level analysis. This difference between single-cell and population-level analysis arises from the unraveling of cell-to-cell heterogeneity in the single-cell data, which could have been hidden by the averaging effect of the population (see below).

PCA is a linear method for dimensionality reduction of single-cell data. In view of non-linear relationships of cell states in state space, recently nonlinear techniques like t-SNE [55] or diffusion maps [63] have been applied in single-cell data analysis. t-SNE is a variation of Stochastic Neighbor Embedding deemed capable of capturing more local structures than classical PCA, while also revealing global structure such as the presence of clusters at several scales. Diffusion maps use a non-linear distance metric (referred to as diffusion distance), which is deemed conceptually relevant in view of noisy diffusion-like dynamics during differentiation [63]. We therefore applied these algorithms on our datasets, as well as another non-linear version of PCA, called Kernel PCA [64], not previously applied to single-cell gene expression data (Fig 2C to 2E). The general conclusions obtained by PCA did not appreciably change when using these non-linear dimensionality reduction techniques. There was again an obvious trend reflecting the differentiation process, as well as a significant amount of intermingling of cells from different time-points.

Single-Cell Data Embed Population Information and Reveal New Discriminating Genes Involved in the Differentiation Process

In order to assess to what extent the differentiation process was still visible in the single-cell data, we performed PCA on datasets from the two extreme time-points, 0 and 72 h (Fig 3A). The result showed a clear separation of both time-points with only a few cells intermingled. We also performed HCA on datasets from the same time-points (Fig 3B). Again, the segregation of the cells was still not perfect, but cells were not as mixed as before. Here, there exist two clusters of self-renewing and differentiating cells. When compared to the analysis of the entire time series, the separation between cells from the two extreme time-points looked clearer. Therefore, the analysis of single-cell data confirmed that part of the information present in the single-cell data is linked to the differentiation process.

The idea that shared information was present in single-cell and population-based data was reinforced by the analysis of the correlation matrices within and between the two datasets (S6 Fig). It was apparent that (1) the global intensity of the correlations was higher with population-based data and (2) there existed a co-structure between the two datasets. At the population level, we showed that the set of genes selected was relevant to analyze the differentiation process (Fig 1). The cross-correlation analysis strengthened this view and demonstrated that when looking at the single-cell scale, the information held by these genes was not totally erased by cell-to-cell variability.

We then looked at the genes that contributed the most to the PCA outcome (Fig 3C). Among the genes that discriminate the most self-renewing cells, one could highlight *LDHA* (Lactate dehydrogenase A), *CRIP2*, and *Sca2*. *Sca2* is a gene that we previously have shown to be associated with the self-renewal of erythroid progenitors [34]. *LDHA* is less expected and will be discussed below. Among the genes that contributed the most to discriminating differentiated cells, one could highlight *RHPN2* and *betaglobin*. Since betaglobin is a part of hemoglobin, the most abundant protein in erythrocytes, it was expected to be associated with differentiating cells.

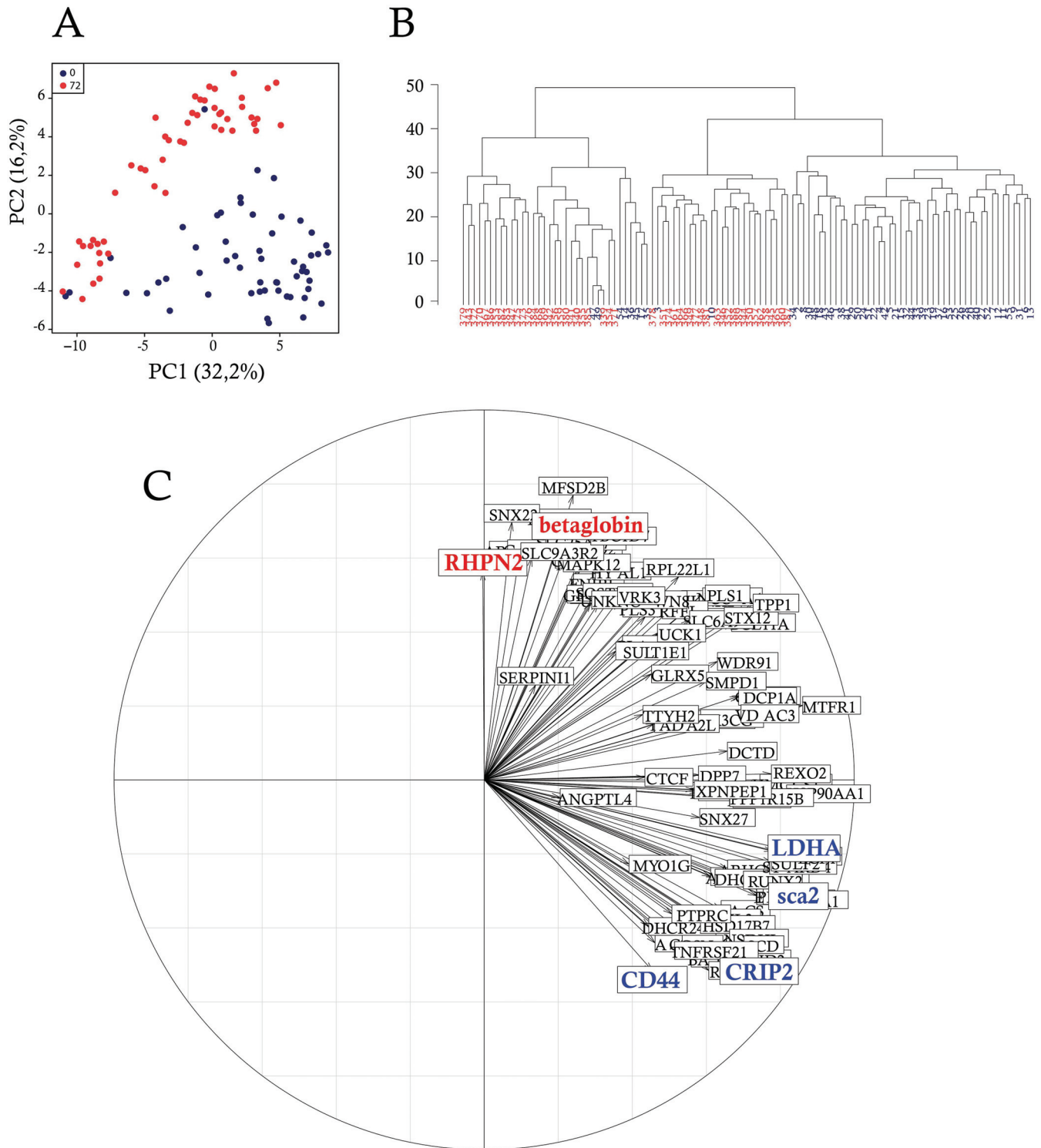


Fig 3. Gene expression-based discrimination between self-renewing and differentiating individual cells. Single-cell gene expression data were analyzed considering only self-renewing cells and cells induced to differentiate since 72 h. (A) Principal component analysis (PCA); (B) Hierarchical cluster analysis (HCA) was used to sort single-cells picked up at 0 h and 72 h of the differentiation process according to similarity measurement; (C) Two-dimensional representation of the contribution of each variable (gene) to the inertia. The direction of the arrows displays the contribution of that variable to the underlying component. The colored genes highlight genes of interest and genes that contributed the most to the PCA outcome, associated with self-renewal (blue) and the erythroid differentiation process (red).

doi:10.1371/journal.pbio.1002585.g003

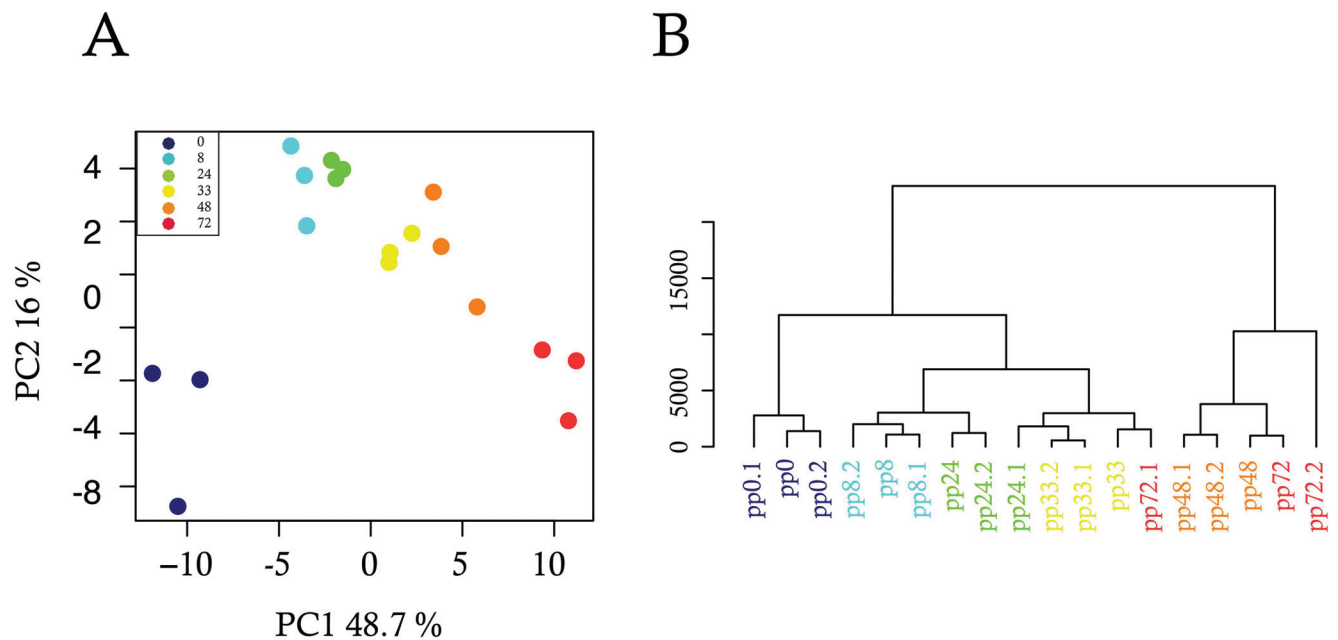


Fig 4. Analysis of single-cell data averaged over pseudo-populations. We separated single-cells into three pseudo-populations with around one-third of single cells for each time-point. We then calculated the average gene expression over each pseudo-population, and analyzed the resulting averaged data using multivariate statistical methods. (A) Principal component analysis (PCA); (B) Hierarchical cluster analysis (HCA).

doi:10.1371/journal.pbio.1002585.g004

Single-Cell Data Averaging Recapitulates Results from Population-Level Analysis

Given that the analysis of single-cell gene expression did not produce a clear separation of the temporal stages, in contrast to whole populations, we hypothesized that by averaging over a population of individual cells, we should be able to reproduce the bulk results. For this purpose, we generated three pseudo-populations (sub-populations) of about one-third of cells from the single-cell data and computed their average gene expressions for each time-point. By performing PCA on the mean gene expressions of these pseudo-populations, we noticed that the averaged data showed more organization and, importantly, that the differentiation progression materialized along the PC1 dimension (Fig 4A).

The PCA result of the pseudo-population therefore looked much more like the population than the single-cell results. Similarly, HCA generated a clustering that was not quite as clear as the analysis of bulk RNA data, but much better than the single-cell analysis (Fig 4B). The HCA results showed for example similarities between gene expressions from time-points 48 and 72 h. Together the pseudo-population analysis obtained by statistical averaging of single-cell data mostly recapitulated, albeit not entirely, the population-based results, suggesting that the clear-cut classification of bulk-cell-based data is due to the (physical) averaging effect in populations, in line with a previous account [65].

The Correlation Networks are Very Dynamical Entities

Single-cell data offers access to the patterns of the relationship of genes with respect to both their marginal (S7 Fig), as well as their full joint distribution (not shown). This provides us with a new observable that we used to characterize the progression of the differentiation process in finer details.

For each time-point, we computed a correlation matrix to evaluate how correlated the expression of any pair of genes was, across all cells at a given time. Since data were log-normally distributed, we employed the Spearman correlation coefficient. We then calculated the significance of the correlation and used a *p*-value below 0.05 as a cutoff. Two genes (the nodes of a graph) that exhibited a significant correlation were connected by an edge. Finally, we sub-sampled 85% of the cells for 10,000 iterations, so as to obtain robust correlation networks that will not depend upon the sampling process. We then constructed a gene correlation network for each time-point. Although both positive and negative correlations were computed, negative correlations proved much less robust and were eliminated by the sub-sampling process, in which we only kept significant correlations that appeared in all of the 10,000 subsampling.

As shown in (Fig 5A), the density of the resulting networks (number of significant correlations) was clearly varying along the differentiation process.

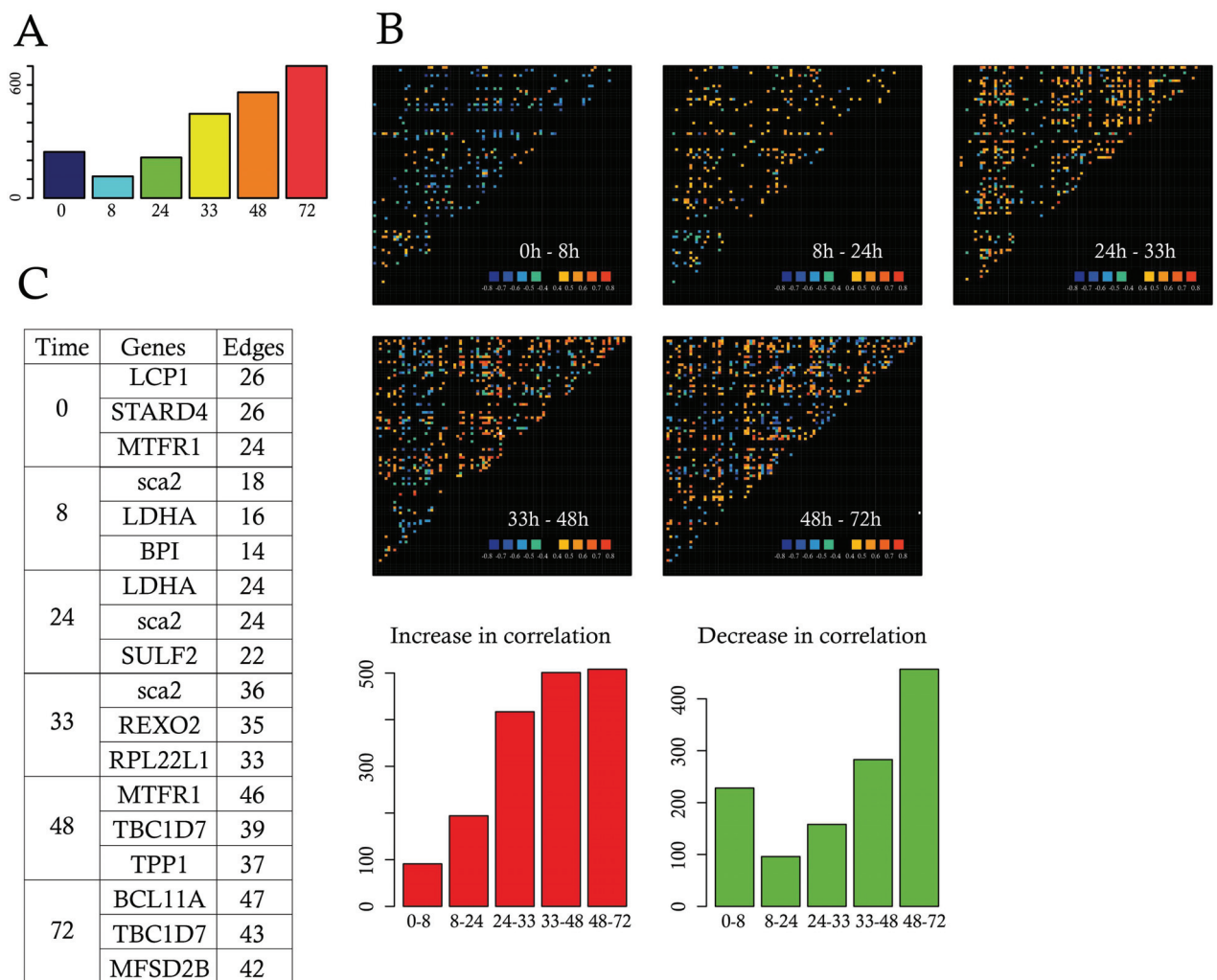


Fig 5. Gene expression correlations. (A) Shown is the number of significant correlations, between any pair of genes, surviving 10,000 sub-sampling iterations, per time-point; (B) Correlation variations between two consecutive time-points using the color code bar shown at the bottom right of the panels. Cold colors (blue and green) indicate decreasing genes correlations and hot colors (from yellow to red) stand for increasing gene correlations between the time-points considered. Intermediary variations (between -0.4 and +0.4) as displayed in black. The bottom left red barplot indicates the number of increasing correlations, whereas the green barplot shows the number of decreasing correlations between each pair of consecutive time-points; (C) The three genes that displayed the highest number of edges at each time-point were listed in the table, as well as the number of edges connecting those genes. Data for this figure (A and B) can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g005

One observed a sudden drop in the number of correlations by 8 h that then steadily increased to reach a maximum value at 72 h much higher than the initial value. Interestingly, this global behavior resulted from both an increase and a decrease in gene-to-gene correlation values (Fig 5B). Even between 48 and 72 h, some gene pair correlation decreased while the overall net balance resulted in a global increase.

This fast-changing density of the networks was also accompanied by a progressive change in the identity of the most highly correlated nodes (Fig 5C). Both *Sca2* and *LDHA* that were previously identified by the PCA also appeared as prominent among the correlation network from 8 to 24 h, while later time-points were characterized by the appearance of other genes as *TBC1D7* and *BCL11A*.

One should note that such correlation networks are to be seen as resulting from the behavior of the underlying mechanistic gene interaction networks, but can not be taken per se as a faithful representation of such dynamical interaction networks.

Evidence for the DNB Theory

Contrary to previous accounts [12, 66], we observed a global decrease in the correlation intensity between 0 and 8 h. Nevertheless, we noticed that some gene pairs showed an increased correlation coefficient. We therefore reasoned that those genes could represent a putative dynamical network biomarker (DNB), a subgroup of genes involved in the critical transition phase of a dynamical system [51]. To qualify for a DNB, three conditions have to be fulfilled: (1) the coefficient of variation (CV) of each variable in the DNB should increase, (2) the correlation (PCCin) within the DNB should increase, and (3) the correlation (PCCout) between the DNB and outside genes should decrease. All three conditions can be simultaneously quantified using the I score (see Materials and Methods). We therefore first selected a group of 12 genes by a two-stage process: (1) we first selected all of the genes that participated in at least one pair that showed an increased correlation of at least 0.5 between 0 and 8 h and (2) among those genes, we selected the genes that showed an increase in their CV value between 0 and 8 h. We then computed the I score of that group of genes at each time-point (Fig 6).

Although PCCin slightly decreased with time, this group of genes nevertheless might still qualify for a DNB since they matched two out of the three criteria used to identify DNBs. Their I value first sharply increased before returning to lower values. This rise is mostly due to a sharp decrease in PCCout between 0 and 8 h, accompanied by a more modest increase in CV. As mentioned, the internal correlation value PCCin decreased, and therefore was not driving the I value. One must note that we computed a Pearson correlation coefficient as advocated [51]. We also tried a Spearman correlation value, which showed a slightly different behavior with a modest increase in PCCin between 8 and 24 h and continued to increase steadily up to 72 h, not affecting the global surge in I value (not shown).

The Initial Driver Genes belong to the Sterol Synthesis Pathway

Since we observed major changes after 8 h of differentiation, one asked how early changes in gene expression could be detected. For this we performed a second single-cell kinetic experiment, where we obtained the expression level of 90 genes in 48, 48, 39, and 41 single cells from 0, 2, 4, and 8 h of differentiation, respectively.

We then defined the first wave of response as genes that showed a significant difference between 0 and 2 h. Two genes satisfied this criterion (Fig 7), establishing that the transcriptional response to the medium change was a very fast process, but concerned only a very limited number of genes.

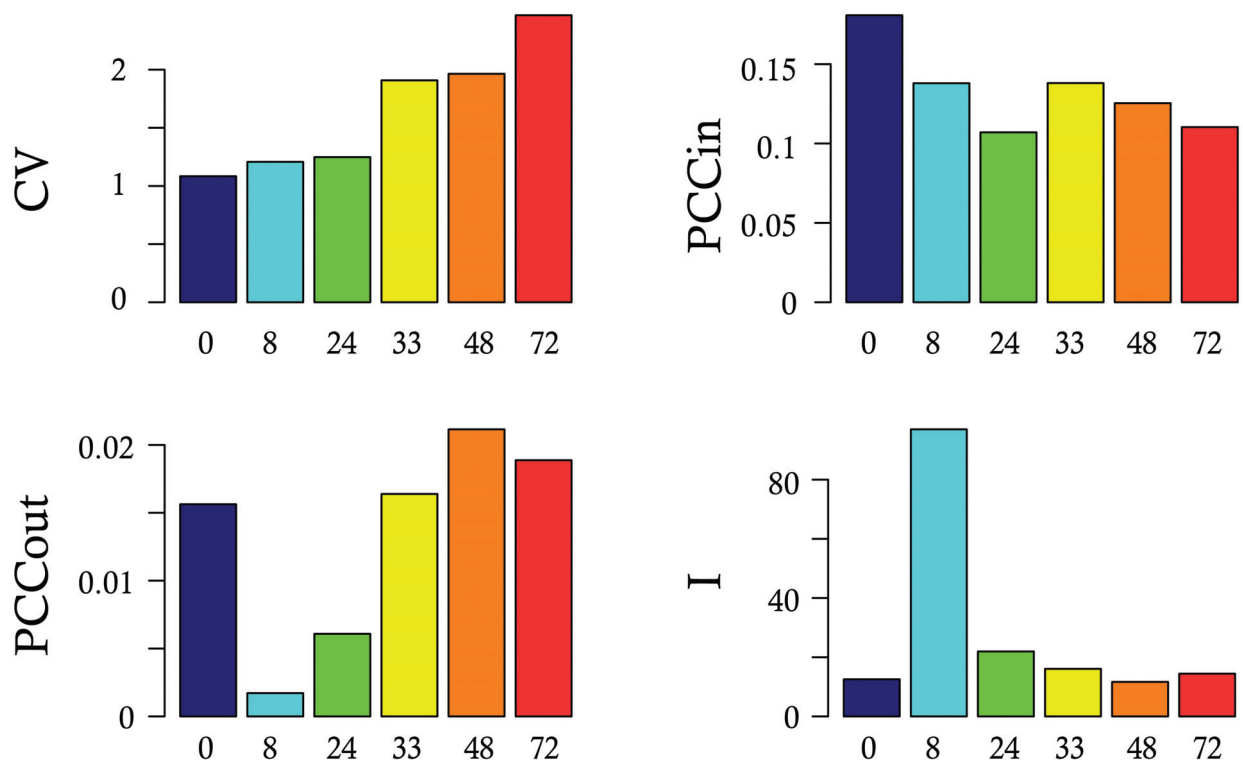


Fig 6. Identification of a dynamical network biomarker. Shown is the behavior of a subset composed of 12 genes fitting the following criteria: increase in their standard deviation and participation to increasing correlations, between 0h and 8h. For this subset, we plotted the mean coefficient of variation (CV), the mean of the correlation between any pair of genes belonging to the subset (PCCin), the mean of the correlation between any one gene of the subset and any one gene outside of the subset (PCCout) and the resulting I-scores, at each time-point. The DNB group included the following genes: *ACSS1*, *ALAS1*, *BATF*, *BPI*, *CD151*, *CRIP2*, *DCP1A*, *EMB*, *FHL3*, *HSP90AA1*, *LCP1*, *MTFR1*. Data for this figure can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g006

The second wave was defined as genes not belonging to wave 1 and showing a significant difference between 2 and 4 h of the response. Five genes satisfied this criterion (Fig 7). It was remarkable that six out of the seven genes from waves 1 and 2 belonged to the same functional group, that is the group of genes associated with sterol synthesis. This proved to be highly statistically significant ($p = 1.8 \times 10^{-6}$). We therefore can propose that the sterol synthesis pathway could act as one of the drivers of the changes that will update the internal network from the changes in external conditions. This would be in line with our previous demonstration for the role of cholesterol synthesis in the decision making process in our cells [35].

A Surge in Cell-to-Cell Variability

A critical novel opportunity provided by single-cell analysis is to study cell-to-cell variability of gene expression as an observable per se and also to add new insight to characterize the temporal progression of differentiation. The question as to what may be the best metrics for quantifying gene expression variability is still open. An aggregated measure called the Jensen-Shannon divergence has been proposed previously as a measure for gene expression noise [9]. One of the main drawbacks of this metric is that it was not possible to assess whether or not the differences observed were statistically significant. We therefore decided to use a simpler Shannon measure of the heterogeneity among the cells for their gene expression profile (see [Materials and Methods](#) and [S2 Fig](#)). Such a measure provided a distribution of entropy values per gene

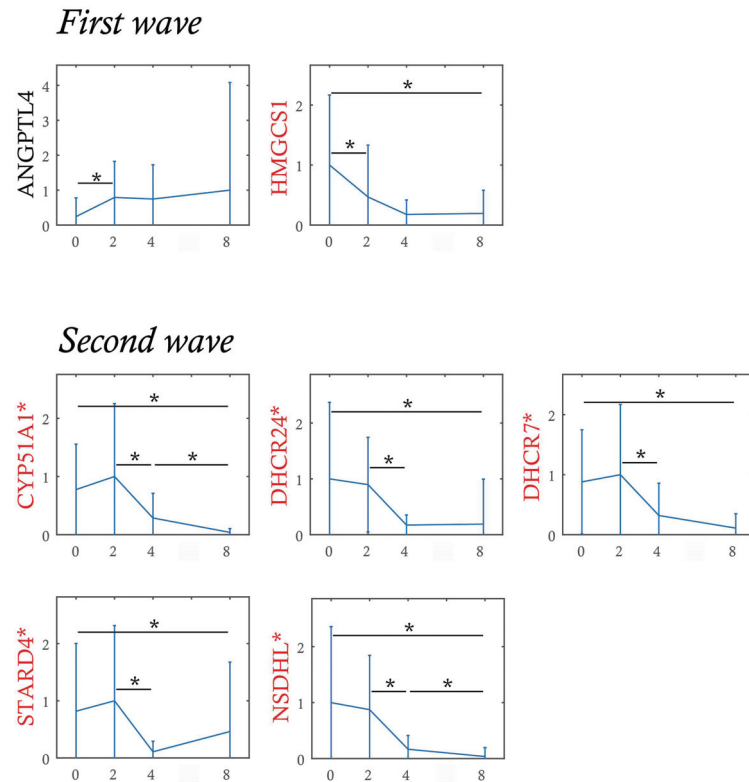


Fig 7. Initial expression waves analysis. Genes are sorted according to the time of the first significant expression variation. The first wave corresponds to genes with a significant variation detected during 0 h and 2 h. The second wave corresponds to genes with a significant variation detected during 2 h and 4 h but without significant variation detected earlier. Genes labeled in red belong to the group of genes associated with sterol synthesis. Significant variations (-*) are detected by non-parametric Mann-Whitney test (p -value < 0.05) if the test is positive in more than 90% of 1,000 bootstrap samples. Genes prefixed by * have a significant variation between 0 h and 8 h detected in both experiments (0 to 72 h, as well as 0 to 8 h). The probability of having 6 genes over 7 (in the first and second waves) belonging to the 10 sterol cluster genes among all 90 genes is estimated to $p = 1.8 \times 10^{-6}$ with the hypergeometric probability density function. Data for this figure can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g007

per time-point, allowing to perform statistical tests. We observed that this entropy increased gradually along the differentiation process, reaching its maximal value at 8 to 24 h, before declining toward 72 h (Fig 8A).

Such an increase of entropy between 0 and 8h resulted from a global increase of each gene entropy, except for a few (Fig 8B). The observed rise in entropy value was highly significant as early as 8 h when compared to 0 h of differentiation. Furthermore, decrease in entropy also became significant between 24 and 33 h of differentiation (Fig 8C). Consequently, since entropy can be defined as a measure of the disorder of a system, this result suggested that a maximal heterogeneity was achieved at 8–24 h of the differentiation process in the expression of our 90 genes, before significantly decreasing to a much lower level of heterogeneity.

Potential Explanation for the Rise in Variability

Different potential causes can be envisioned to explain this increase in entropy, including cell size and cell-cycle stage variations, asynchrony in the differentiation process, and more dynamical causes.

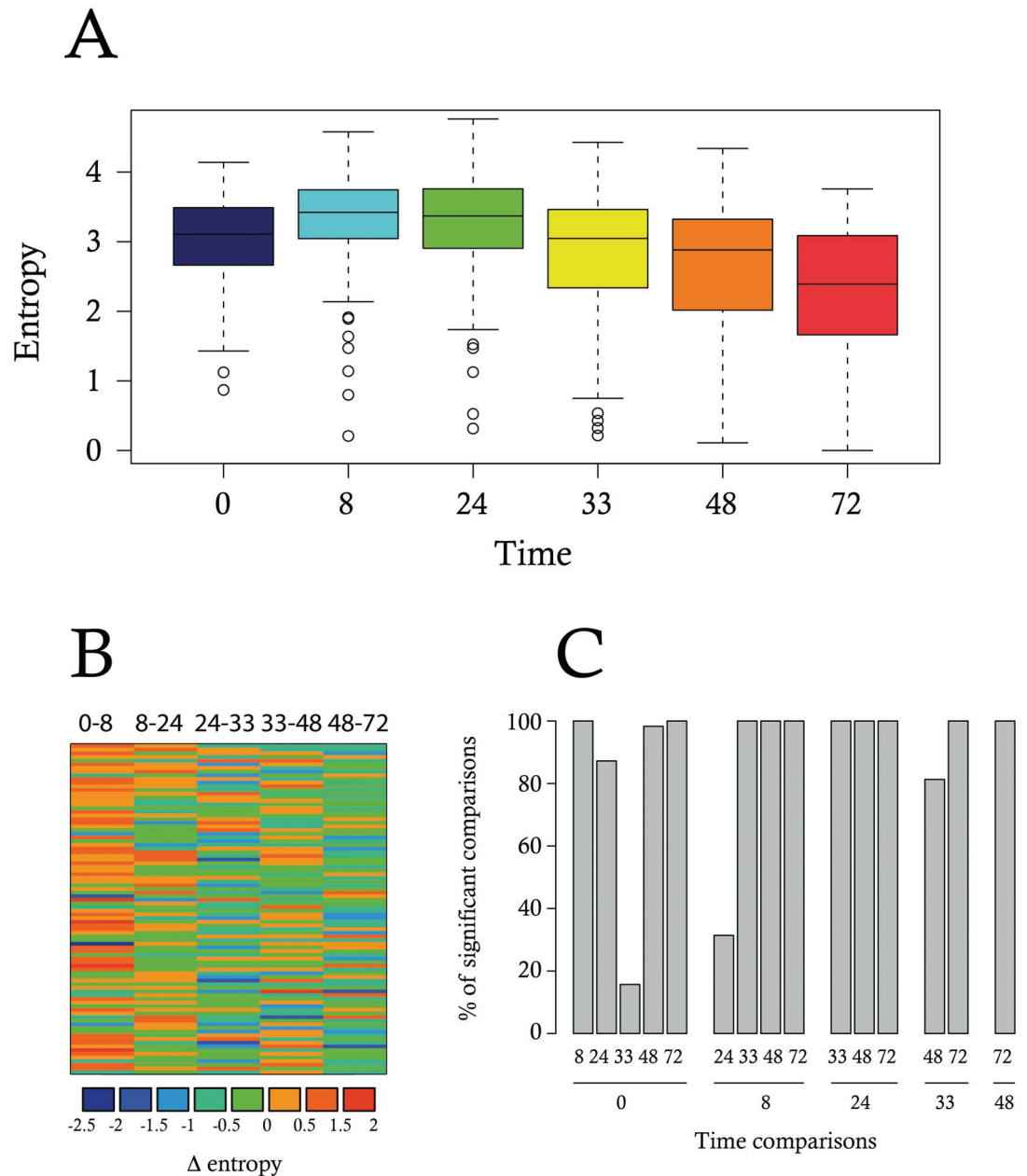


Fig 8. Cell-to-cell heterogeneity measurement using Shannon entropy. (A) A Shannon entropy was calculated for each time-point for each gene. Boxplots represent the distribution of the entropy values; (B) Gene entropy variation: for each gene (i.e., lines), we represented the difference between entropy values at two consecutive time-points (Δ -entropy) using a color gradient code. Negative and null delta entropies (i.e., for a given time-point, the entropy value for these genes decreased or does not change, compared to the earlier time-point) are colored in blue and green. Positive delta entropies are colored in orange or red; (C) We assessed the significance of the differences between any pair of time-point through a Wilcoxon test. The robustness of the result was assessed by performing subsampling. The barplot shows the results as the percentage of 1,000 iterations for which a significant difference (p -value < 0.05) was detected. Data for this figure can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g008

As suggested in some previous works, cell size and cell-cycle stage variations could influence gene expression, and become confounding factors [67–69]. Nevertheless, variability due to variations in cell cycle has been shown to be quantitatively negligible in erythroid precursors [70]. We also added in our gene list the CTCF gene, known to be cell-cycle regulated in chicken cells [71]. Almost no correlation was detected between this gene and any of the 91 other genes (Fig 9A) demonstrating that our gene list contained virtually no other cell-cycle-regulated gene. Furthermore, we assessed whether or not the repartition of our cells within the different phases of the cell cycle could have been modified at a time where entropy was peaking. No significant difference in cell cycle repartition could be seen at 8 h of differentiation (Fig 9B). Altogether, those results demonstrate that a potential effect of cell cycle variation would only marginally explain our data. Regarding cell size, it is important to note that in our system the peak in gene expression variability at 8–24 h occurs at a time where cell size is not affected (Fig 10B). If anything, we observed a slight increase in cell size, which could be responsible for a decrease, and not an increase, in noise [72].

We then assessed a potential effect of asynchrony in the differentiation process. For this, we first employed the following algorithms: SCUBA [52], WANDERLUST [53] and TSCAN [54] to reorder the cells according to the calculated pseudotimes. However, SCUBA led to a cell re-ordering that was highly inconsistent with the actual time-points, where all self-renewing cells (time 0 h) were placed in the middle of the SCUBA order (not shown). WANDERLUST and TSCAN produced a more reasonable cell ordering. However, the trajectories of the gene expression profiles following this ordering were quite erratic (not shown). Nevertheless, the entropy of sub-populations of cells, grouped according to either their WANDERLUST pseudotimes or TSCAN clusters, showed the same rise-then-fall profile as with the original single cell data (Fig 9C and 9D).

In theory, these algorithms are supposed to reconstruct a posteriori the “hidden” order along the differentiation pathway. Within this frame, the behavior of entropy in re-ordered cells tends to support the idea that asynchrony in the differentiation process is not the leading cause of our observed increase in entropy.

However the intrinsic burstiness of the gene expression process [24, 73–75] might cause some issues in the use of cell re-ordering algorithms. We therefore examined this question by using a more formal approach. We reasoned that a modeling strategy might be useful in establishing the role asynchrony might play, especially since forcing a synchronous differentiation is not accessible *in vitro*, but can be done *in silico*. We used a two-state model of gene expression [27, 39–41, 56], for which we could learn the parameters from the data (see [Materials and Methods](#)). In the synchronous case, we obtained a variation in entropy resembling the one we calculated from the data (Fig 9E). The introduction of asynchrony induced a flatter time profile of the entropy (Fig 9F).

This finding did not, however, prove that our cells are synchronously differentiating, but only demonstrated the effect of asynchrony: in the background of bursty gene transcriptional process, asynchrony will tend to smoothen (and not augment) the entropy of the system. Therefore the observed surge in entropy can not be attributed to the asynchrony of the process.

The rise-and-fall of entropy in our data is in line was examined in a different setting, namely a reprogramming process [58]. The authors stated, “The initial transcriptional response is relatively homogeneous,” offering the opportunity to examine the entropy time profile in such a homogeneous process. Our analysis of this dataset produced a similar behavior for entropy which significantly increased initially, before returning to lower values (S8 Fig).

Altogether our analysis is compatible with the notion that the rise and fall in entropy is the consequence of the dynamical behavior of the underlying gene regulatory network.

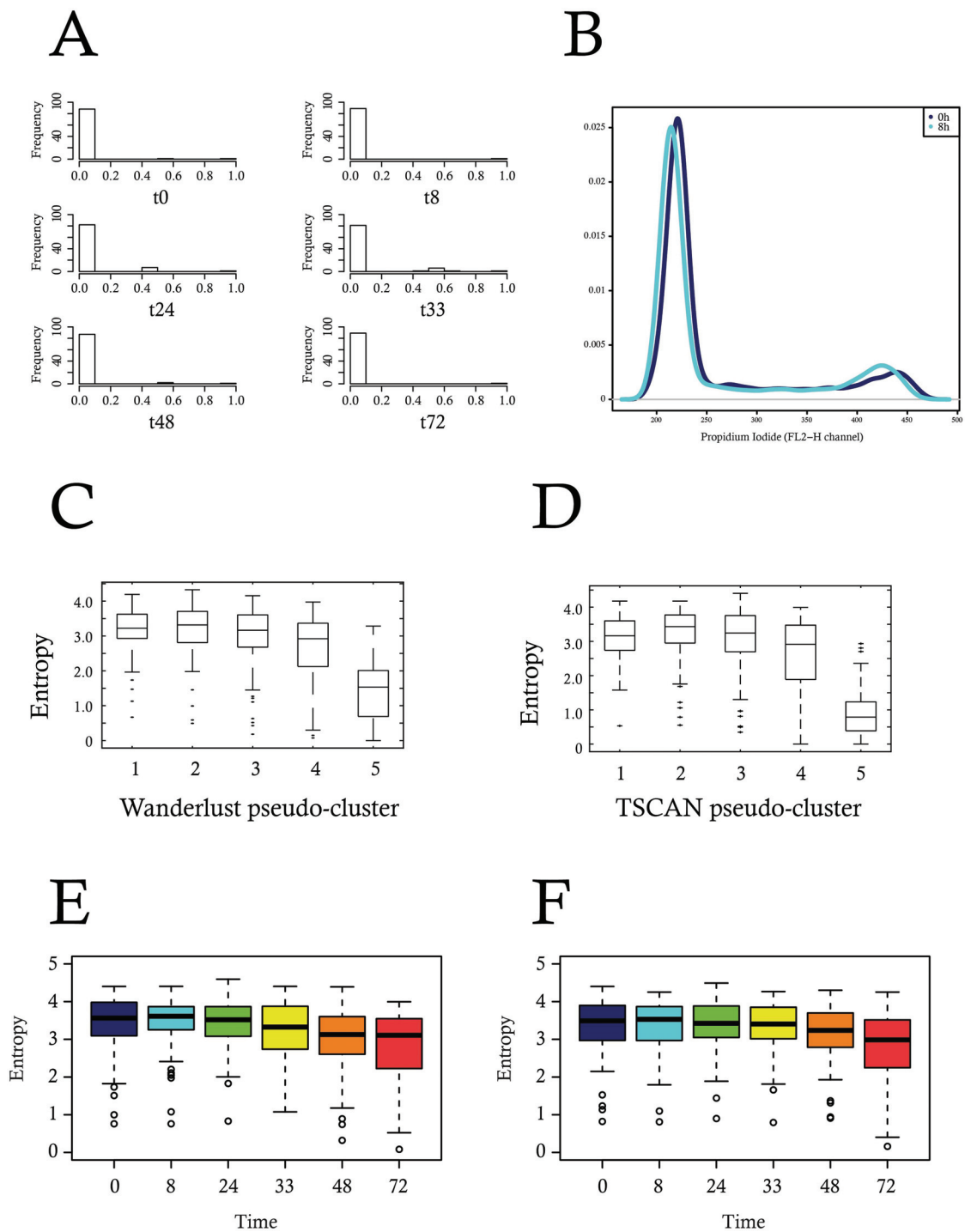


Fig 9. Exploration of potential confounding factors. (A) Correlation of the CTCF gene with the rest of the 91 genes, at all six time-points. (B) FACS analysis of the cell cycle repartition at 0 and 8 h of differentiation. The difference between the two distributions was found not to be statistically significant ($p = 0.18$ using a Wilcoxon test). (C and D): calculation of the entropy content per cluster of cells re-organized using either WANDERLUST (C) or TSCAN algorithm (D). (E and F) In silico comparison of the effect of a synchronous versus an asynchronous differentiation process on the evolution of entropy. Data for this figure (C to F) can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g009

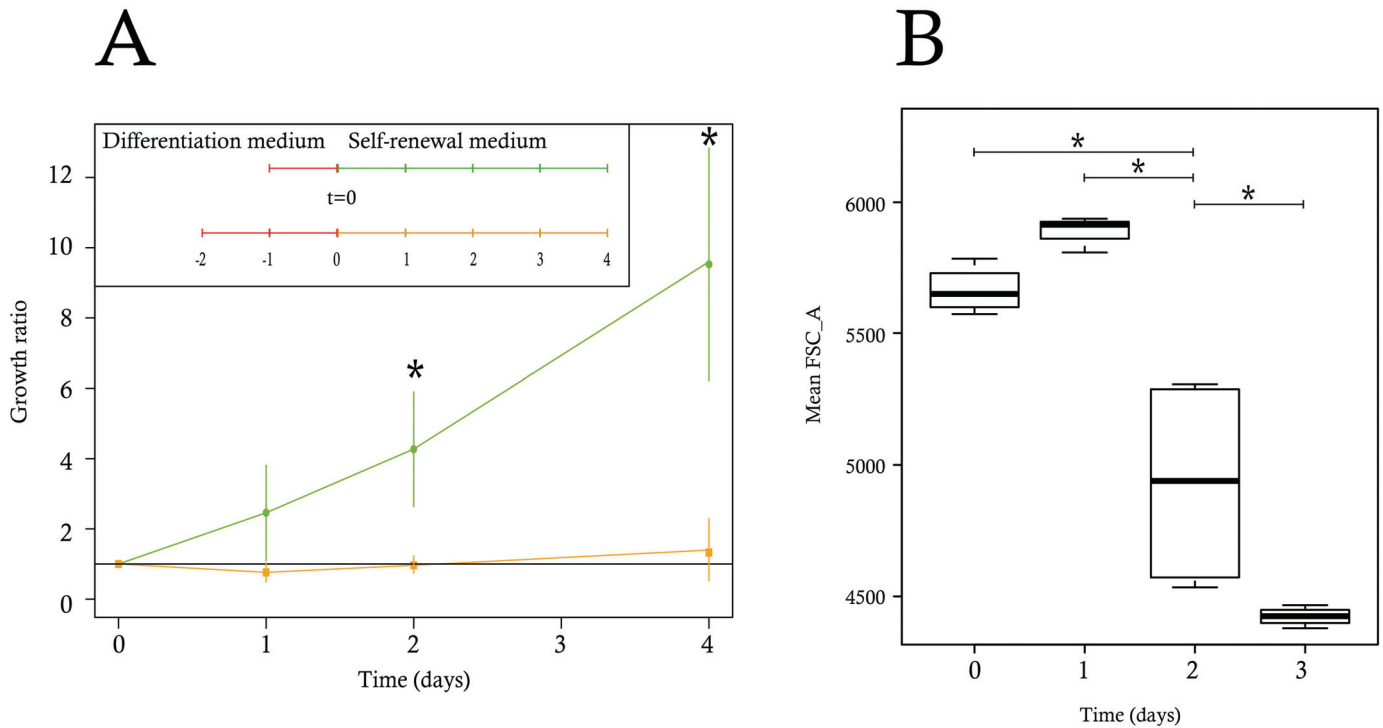


Fig 10. Evolution of physiological differentiation parameters. (A) T2EC were induced to differentiate for 24 and 48 h and subsequently seeded back in self-renewal conditions. Cells were then counted every day for 5 d. The green curve represents the growth of cells induced to differentiate for 24 h and the orange curve indicates the growth of cells induced to differentiate for 48 h. The data shown are the mean \pm standard deviation calculated on the basis of three independent experiments for the time-points 72 h and 96 h and four experiments for all other time-points. The growth ratio was computed as the cell number divided by the total cells at day 0. The significance of the difference between growth ratios at 24 h and 48 h was calculated using a Wilcoxon test. (B) The boxplots of the mean size observed were based on four independent experiments, each using 50,000 cells, using FSC_A as a proxy for cell size. All of the variances were compared by pairs using the F test and the * indicates when the variances were significantly different. Data for this figure can be found at osf.io/k2q5b.

doi:10.1371/journal.pbio.1002585.g010

The Point of No Return in T2EC Differentiation is Located between 24 h and 48 h

The above analysis of single-cell transcript profiles displays the following pattern:

1. A decrease in correlation value is observed between 0 and 8 h, and then correlation increases between 24 and 72 h.
2. An increase in I score value is observed between 0 and 8 h, then a return to its initial value at about 33 h, before continuing to decrease gradually.
3. A surge in entropy is significant at 8–24 h, and significantly decreases between 24 and 72 h.

Altogether, those results point toward the 8 and 24 h time-points as being a possible decision point, hence, a “point-of-no-return” in the differentiation process, beyond which cells are irreversibly committed toward erythrocytic differentiation. Consequently, we hypothesized that committed cells would be unable to revert back to a self-renewal process after 24 h of differentiation. To test this hypothesis we induced T2EC to differentiate for 24 h or 48 h, after which cells were transferred back into the self-renewal medium, in order to determine whether or not cells could revert back to the undifferentiated state after they had received differentiation signals for a given period of time. We observed that T2EC induced to differentiate for 24

h were still able to self-renew upon change of medium, while cells induced for 48 h could not do so (Fig 10A).

T2EC induced for 48 h seemed to stay in a quiescent state until they died. We therefore concluded that the physiological point of no return is located between 24 h and 48 h of our differentiation process, as suggested by our *in silico* analysis. Finally we determined whether cell size, a phenotypic integrated variable that has historically been used to monitor erythroid maturation [76, 77] would manifest the behavior of the underlying molecular network with respect to cell-cell variability. We therefore assessed cell size variation during the differentiation process. As expected [32], mean cell size started to decrease during differentiation to reach a minimum by 72 h (Fig 9B). Interestingly, cell size variability significantly peaked at 48 h before dropping precipitously by 72 h. Thus the high variability of gene expression observed at 24 h preceded a significant peak in cell size variability 1 d later.

Discussion

In the present work we assessed, using single-cell RT-qPCR, the temporal changes of gene expression in individual cells in a population of cells undergoing differentiation. For this, we used a physiologically relevant cellular system, which presents three main advantages: (i) those cells are primary, non-transformed cells; (ii) they do not show any tendency to spontaneous differentiation; and (iii) they can only differentiate along the erythrocytic lineage, excluding heterogeneity arising from coexistence of cells differentiating along different lineages.

To quantitatively assess the role of gene expression variability, we first defined a subset of genes relevant for analyzing the differentiation process. At the level of whole-population analysis this gene subset allowed a clear distinction among differentiation time-points. However, when assessed at the single-cell level, our analyses revealed a much higher cellular heterogeneity. Despite this heterogeneity, the selected genes were still effective in separating the two most extreme time-points in T2EC differentiation, confirming that information associated with the differentiation process is embodied in the gene expression data at the single-cell level. From the dataset that we generated at the single-cell level, two main results could be obtained: (i) regarding the biology of the erythroid differentiation, we identified previously unidentified genes as being important components of the self-renewal and differentiation of erythroid progenitors, and (ii) on a larger perspective, our results fully supported a dynamical view where differentiation can be seen as a critical phase transition driven by stochasticity.

Identification of new genes involved in the erythroid differentiation process

One question deals with the possible identification of important genes that can be seen as “drivers” of the process. At least three list of genes were generated during the course of this work that may qualify:

1. the “early drivers,” genes identified in the wave analysis;
2. the genes qualifying for the DNB, and
3. the most densely connected genes in the correlation graph;

Restricting only to the most densely correlated genes at 0 and 8 h (since the two other lists were validated on those time-points), one observed a partial overlap between the three lists (S9 Fig), with no gene being common to all three lists. One possible explanation is simply that the three lists were obtained through different approaches, not supposed to identify the same set of genes. This result nevertheless suggests that although all of those genes might be functionally

important for the differentiation process, they might be involved in the global response at different levels. The early drivers might be more important for informing the whole network at early time points, whereas the two other genes sets might be involved in a more global reconfiguration of the network at later time-points. In any case those gene lists are to be seen as traces resulting from the behavior of the underlying dynamical network, and should not be mistaken for the dynamical network itself. It would therefore be of utmost importance to be able to correctly infer such a network. We are actively pursuing this goal in our group.

We discuss below possible functions of some of those genes, a full discussion for all genes being out of the scope of the present paper.

As previously mentioned, *Sca2* is a gene which we have previously shown to be associated with the self-renewal of erythroid progenitors [34].

LDHA encodes an enzyme that catalyzes the conversion of pyruvate to lactate, and has been involved in the Warburg effect (or anaerobic glycolysis), which is the propensity of cancer cells to take up glucose avidly and convert it to lactate [78]. Furthermore, deletion of *LDHA* has been shown to significantly inhibit the function of both hematopoietic stem and progenitor cells during murine hematopoiesis [79].

Since *LDHA* expression is under the control of HIF1 α transcription factor [79], it could be involved in the response of immature erythroid progenitors to anemia. Those cells have to show a significant amount of self-renewal for recovering from a strong anemia, implying low oxygen condition [80]. It makes perfect sense that in this case the metabolism of self-renewing progenitors would rely upon an anaerobic pathway.

Moreover, HIF1 α has also been shown to be an upstream regulator of *HSP90alpha* secretion in cancer cells in a protective way against the hypoxic tumoral environment [81]. Therefore, our results are in line with other findings showing that anaerobic glycolysis is favored in hypoxic conditions, such as the bone marrow environment, and required for stem cell maintenance [82]. Otherwise, since *LDHA* and *HSP90alpha* form part of the lists of potentially important genes between 0 and 8 h, our finding suggests that erythroid differentiation might be accompanied by a change from anaerobic glycolysis toward mitochondrial oxidative phosphorylation, as recently proposed [83].

Finally, our analysis highlighted the importance of the sterol synthesis pathway in the self renewal process since:

1. Among genes identified by RNAseq whose expression changed significantly, we found different genes associated to the sterol synthesis, such as *HMGCS1*, *CYP51A1*, *DHCR24*, *DHCR7*, *STARD4*, and *NSDHL* (S4 Fig);
2. The expression of those genes decreased promptly after the change of the external conditions, i.e the induction of the differentiation (Fig 7);
3. *STARD4* was both an early driver and one of the genes that displayed the highest number of edges at 0 h (Fig 5C). It has recently been demonstrated that *STARD4* expression could be used as poor prognosis gene in a six genes signature that defines aggressive subtypes in adult acute lymphoblastic leukemia [84].

These observations support the importance of sterol synthesis in the maintenance of cellular self renewal state and the necessity of a decrease of some sterol associated genes expression to allow the differentiation. The question as to why this group of genes act as the early sensors of change in environmental conditions remains elusive. In line with our previous results [35], one could hypothesize that cholesterol synthesis is a barrier toward differentiation/apoptosis that has to be lowered for differentiation to proceed.

A functional role for the surge in gene expression during critical transition?

On a more global perspective, the importance of cell-to-cell heterogeneity as a “biological observable” at the single-cell level, even among cells classified as belonging to the same “cell type” [85], is increasingly recognized [86]. But to what extent and when is such heterogeneity functionally important? Most single-cell transcript profile analyses of cell populations have so far focused mostly on computational descriptive analysis to identify clusters, and temporal progression, or to test dimensionality reduction and visualization tools, but less so to test a biological hypothesis. Here we used the single-cell granularity of gene expression analysis to test the long-standing hypothesis that stochastic cell-cell variability is not simply the byproduct of molecular noise but that such randomness of cell state plays a key role in differentiation [28]. In this Darwinian view, differentiation starts with an unstable gene expression pattern, generating cell type diversity. Therefore, one testable prediction was that an increase in gene expression heterogeneity should be observed during the critical phase of cell differentiation whenever the irreversible decision to commit is made.

Our main contribution is a demonstration that the increase in molecular variability precedes critical functional variations in cellular parameters, most importantly including the commitment status of the cells. Taken together, the timing of three observables achieved at single-cell resolution provides a coherent picture of a temporal structure of differentiation that would be invisible to traditional whole-population averaging techniques: (i) the surge in cell-to-cell variability of gene expression patterns of individual cells at 8–24 h; (ii) a sudden drop in the overall correlation, concomitant with the emergence of a DNB; and (iii) followed by the phenotypic marker of differentiation, the decrease of cell size, for which variability peaks at 48 h.

An important question is the relevance of that peak in variability. We demonstrated experimentally that no cell was able to return to a self-renewal state after 48 h in a differentiation medium. A similar timing for point-of-no return has previously been suggested in FDPC-mix cells [87]. Such an irreversible commitment to differentiation preceded by a highly significant increase in cell-to-cell variability is consistent with the explanation that cells differentiate by passing through two phases [87]: a first phase in which the self-renewing state is destabilized and primed by perturbation of their extracellular environment, followed by a second phase of a stochastic commitment to differentiation.

These observables (emergence of a DNB, drop in correlation, significant increase in entropy, surge in cellular parameters variations) jointly suggest a critical state transition, a class of dynamical behaviors that has been proposed to explain the qualitative, almost discrete and noise-driven “switching” into a new cell state as embodied by differentiation [88]. This conceptual framework naturally explains the irreversibility of fate commitment [89]. Indeed the maximum of the above three observables coincided with the functionally demonstrated point-of-no return to the self-renewal state in T2EC differentiation process, which was located between 24 and 48 h.

From a more biological perspective, we can view differentiation induction as a process of adaptation in which the cell’s internal molecular network, adapted for growth in self-renewal conditions, has to adjust to the new external conditions when differentiation is induced by the change in external conditions. For example, in yeast, it has been shown that a nonspecific transcriptional response reflecting the natural plasticity of the regulatory network supports adaptation of cells to novel challenges [90]. The underlying mechanisms are yet to be discovered, but one would expect global mechanisms to be involved. Modifications of the chromatin dynamics [27] under the possible control of metabolic changes [91] are obvious candidates for such a role. Fluctuation in important transcription factor level has

also been proposed to be involved [92]. The surge of non-specific variability would allow exploration of new regions in the gene expression space. Preventing such an increase in variability has been associated to trapping cells in an undifferentiated state [93]. This increase would lead to a reconfiguration of the gene expression network into a state which is compatible with differentiation conditions and which is robust and consistent with a new attractor state in the network [29]. Then the decrease of molecular variability might reflect the implementation of the fully differentiated phenotype as cells settle down in the next stable state.

In this study, we exploited the wealth of information available in single-cell data by highlighting the critical molecular changes occurring along the differentiation sequence. First, the initial gene expression waves might represent a very early signal that happens between 0 and 8 h, followed by a pre-transition warning signal revealed by the DNB analysis, concomitant with the drop in gene correlations and the rise in cell-to-cell variability. Such a pattern are thought to reflect the underlying dynamical molecular mechanisms that drives the evolution of cells through the differentiation process. The first signals could be seen as an adaptative response to environmental changes, as suggested above, whereas the last warning signal, before irreversible commitment, could be seen as the point of cell decision making. At that stage it is hard to really be sure that the DNB genes actually drives the critical transition, but at the very least they represent a clear signal that our cells are experiencing such a transition. Until 24 h, at least, cells would still be able to functionally respond to self-renewal signals. This implies that at that stage the state of the network would be compatible with both a differentiation and a self-renewal process. One of the remaining challenging questions is what makes the cell takes the irreversible decision to differentiate at a point when the system seems to be totally disorganized. We strongly believe that this will be an emerging properties from the behavior of dynamical high-dimensional molecular network.

While the current study offers a single-cell resolution view on gene expression, it does so only through snapshots at strategically selected time-points. In the future it would therefore be of great importance to obtain a continuous measurement of the underlying gene expression network in order to explain the state changes in individual cells and to reconstruct the entire trajectory of each cell in gene expression state space. This information would expose the actual process of diversification that leads to the maximal heterogeneity marking the point of no return of differentiation.

NOTE ADDED IN PROOF: During the submission of this manuscript we became aware of the work of Mojtahedi, et al., 2016 (doi: [10.1371/journal.pbio.2000640](https://doi.org/10.1371/journal.pbio.2000640)) which arrived at a similar conclusion, and we cite that work in our discussion.

Materials and Methods

Cells and Culture Conditions

T2EC were extracted from bone marrow of 19-d-old SPAFAS white leghorn chickens embryos (INRA, Tours, France). These primary cells were maintained 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 [32]. T2EC were induced to differentiate by removing the LM1 medium and placing cells into the 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)). Differentiation kinetics were obtained by collecting cells at different times after the induction in differentiation.

Cell Population Growth Measurement

Cell population growth was evaluated by counting living cells using a Malassez cell and Trypan blue staining.

Propidium Iodide Staining

T2EC in self-renewal medium and T2EC induced to differentiate during 8 h were incubated for 30 min on ice with 100% cold ethanol, and then 30 min at 37°C with 1 mg/mL RNase A (Invitrogen). Propidium Iodide (SIGMA) was added at 50 µg/mL 2 min prior to analysis and fluorescence was measured with the BD FACS Calibur 4-color flow cytometer, using the FL-2 channel. Data files were then extracted and analyzed using the bioconductor flowCore package.

T2EC Collection by Flow Cytometry

T2EC were collected individually in a 96-well plate using a flow cytometer (FACS ARIA I). Each individual cell was immediately gathered into a lysis buffer (Vilo [Invitrogen], 6U SUPERase-In [Ambion], 2.5% NP40 [ThermoScientific]), containing also Arraycontrol RNA spikes (Ambion). After collection, single-cells were immediately frozen on dry ice and stored at -80°C.

Total RNA Extraction

Cell cultures were centrifuged and washed with 1X phosphate-buffered saline (PBS). Total RNA were extracted and purified using the RNeasy Plus Mini kit (Qiagen). Then, RNA were treated with DNase (Ambion) and quantified using the Nanodrop 2000 spectrophotometer (ThermoScientific).

RNA-Seq Libraries Preparation

RNA-Seq libraries were prepared according to Illumina technology, using NEBNext mRNA library Prep Master Mix Set kit (New England Biolabs). Libraries were performed according to manufacturer's protocol. mRNA were purified using NEBNext Oligo d(T)25 magnetic beads and fragmented into 200 nucleotides RNA fragments by heating at 94°C for 5 min, in the presence of RNA fragmentation Reaction Buffer. Fragmented mRNA were cleaned using RNeasy MinElute Spin Columns (Qiagen). Double strand cDNA were obtained by two-step RNA reverse transcription (RT) with random primers and purified using Magnetic Agencourt AMPure XP beads. To produce blunt ends, purified cDNA were incubated with NEBNext End Repair reaction buffer and NEBNext End Repair enzyme mix for 30 min at 20°C. cDNA were purified again using Agencourt AMPure XP beads, and dA-tail were added to these cDNA fragments by incubating them with NEBNext dA-Tailing reaction buffer and klenow fragment for 30 min at 37°C. After purification of the dA-tailed DNA, illumina adaptators were ligated to cDNA in the presence of NEBNext quick ligation reaction buffer, quick T4 DNA ligase, and USER enzyme. After size selection, purified adaptor-ligated cDNA were enriched by PCR with NEBNext High-fidelity 2X PCR Master mix, universal PCR primers and Index primers, and using thermal cycling conditions recommended by manufacturer's procedure. Finally, enriched cDNA were purified and sequenced by the Genoscope institute (Evry, France).

RNA-Seq Library Analysis

Sequencing files were loaded onto Galaxy (<https://usegalaxy.org/>). Quality was checked using FastQC. Groomed sequences were aligned on the galGal4 version of the chicken genome,

using TopHat [36]. The resulting .BAM files were transformed into .SAM files using SAM Tools. The gene counts table was generated using HTSeq [37] and the chr_M_Gallus_gallus.Galgal4.72.gtf annotated genome version. Differential gene expression was computed using EdgeR and plotted with the `plotSmear` function [38].

High-Throughput Microfluidic-based RT-qPCR

Every experiment related to high-throughput microfluidic-based RT-qPCR was performed according to Fluidigm's protocol (PN 68000088 K1, p.157–172) and recommendations.

Reverse transcription of isolated bulk-cell RNA and single-cell RNA.

- *Isolated bulk-cell RNA*

Fifty nanograms of extracted bulk-cell RNA were reverse-transcribed using the Superscript III First-Strand Synthesis SuperMix for qRT-PCR kit (Invitrogen). The reverse transcription step and RNase H treatments were performed according to manufacturer's instructions. Reverse transcription was performed during 30 min at 50°C, followed by 5 min at 80°C, and RNase H treatment was run at 37°C during 20 min. Finally, cDNA were stored at -20°C.

- *Single-cell RNA*

Single-cell lysates were thawed on ice and denatured for 1.5 min at 65°C. RNA were reverse-transcribed in presence of SuperScript III Reverse Transcriptase enzyme, from the SuperScript VILO cDNA Synthesis kit (Invitrogen), and T4 gene 32 protein (New England Biolabs) to improve reverse transcription efficiency. The reaction thermal cycling conditions were 5 min at 25°C, 30 min at 50°C, 25 min at 55°C, 5 min at 60°C and 10 min at 70°C.

Specific target amplification of cDNA. Primers were designed using the Ensembl database (http://www.ensembl.org/Gallus_Gallus/Info/Index/) and Primer3Plus software (<http://www.bioinformatics.nl/primer3plus/>). For information about the primers sequences used, please contact the authors.

The cDNA pre-amplification was performed using the TaqMan PreAmpMaster (Applied Biosystems) mixed with all primer pairs of the genes of interest (Sigma-Aldrich), diluted at 500 M. For single-cell cDNA pre-amplification, this reaction mix was also composed of 0.5 M pH8 EDTA. The thermal cycling program used for single-cell cDNA is 10 min of enzyme activation at 95°C, followed by 22 cycles at 96°C for 5 s and 60°C for 4 min. For bulk-cell cDNA, the enzyme activation step was followed by 14 cycles at 95°C for 15 s and 60°C for 4 min.

Exonuclease treatment. Exonuclease I (*E. coli*, New England BioLabs) was used on pre-amplified cDNA to eliminate single-strand DNA. The treatment was performed at 37°C during 30 min and then the enzyme was inactivated at 80°C during 15 min. For bulk-cell, cDNA were diluted in TE (10 mM pH8 Tris, 1 mM EDTA). For single-cell, cDNA were diluted in low EDTA TE buffer (10 mM pH8 Tris, 100 nM EDTA). All samples were then stored at -20°C.

RT-qPCR: data generation. Pre-amplified cDNA were mixed with Sso Fast EvaGreen Supermix With Low ROX (Bio-Rad) and DNA binding dye sample loading reagent (Fluidigm). Primer pairs of the genes of interest were diluted at 5 μM with the Assay Loading Reagent (Fluidigm) and low EDTA buffer. First, the 96.96 DynamicArray IFC chip (Fluidigm) was primed. Then, prepared cDNA and primer pairs were loaded in the inlets of this device.

To avoid chip-linked variability, when analyzing single-cell data we were careful to represent every time-point in each of the four microfluidic-based chip analyzed.

The prime step and transfer of cDNA samples and primers from the inlets into the chip were performed using the IFC Controller HX (BioMark HD system). The chip was analyzed

using the BioMark HD reader according to the GE 96 × 96 PCR + Melt v2.pcl program, thanks to the data collection software. Then, raw data were analyzed with the Fluidigm Real-Time PCR Analysis software.

Positive exogenous controls (RNA spikes) were used to validate the RT-qPCR experiment as recommended by Fluidigm Company. We also used the RNA spikes to normalize the data (see below). To determine qPCR efficiency of every primer pairs used, serial dilution scales of bulk-cell cDNA were performed. PCR efficiencies were calculated as follows: $E = 10^{-1/\text{slope}}$. Primer pairs presenting PCR efficiency less than 80% or more than 120% were removed from subsequent analyses.

RT-qPCR: low-level data analysis. First, a manual examination was performed regarding data quality. RTqPCR data were exported from the BioMark HD data collection software. On every microfluidic-based chip, each gene was controlled in a qualitative manner in order to keep only reliable and good quality data. For this we manually edited the data files by adding a new column named “DELETED.” Numbers “0” or “1” were appended in this column according to various criteria. Quality control was based both upon amplification and melting curves examination. For one given gene all the melting curves had to be centered on a unique melting temperature. When a given melting curve peak shifted to a higher or lower T_m , “1” was added into the DELETED column for this amplification. Moreover, data displaying a double peak were also considered unreliable and annotated with a “1.” Finally, “noisy” amplification curves departing from the smooth classical sigmoidal shape were also tagged as “1.” We allowed the quantification cycle (Cq) to be as high as 30. For a higher number of cycles, the machine returned a value of 999, meaning that there were not enough molecules to be detected. After this quality control, Cq values of data tagged as “1” were replaced with UD (for “undefined”) in the raw data file, since they would not be taken into account in later analysis. Then the new table underwent an automatic formatting consisting in a second multiple-criteria cleaning process using an in-house R script. Cq values were converted into (approximately) absolute numbers of molecules according to the following steps. First, we selected cells with at least one valid spike measurement (i.e., whose Cq is different from UD and 999). Then, we normalized the raw value $\widehat{Cq}_{i,j}$ for cell i and RNA j according to the cell mean spike value \overline{Cq}_i (or the only available spike if one is invalid), with the global mean spike value \overline{Cq}_0 as reference. That is, the normalized value $Cq_{i,j}$ for cell i and RNA j is defined by

$$Cq_{i,j} = \widehat{Cq}_{i,j} - (\overline{Cq}_i - \overline{Cq}_0).$$

After removing cells with abnormally important amount of genes with low expression (high $Cq_{i,j}$ values, suggesting the absence of a cell in the well), the numbers of mRNA molecules were estimated, considering the following: a maximum Cq equal to 30 as the measurement of 1 molecule in the well after 22 cycles of pre-amplification, a dilution factor corresponding to 1 cell extract diluted in 96 wells, and a sampling of 1/45 for PCR measurement. Thus the number $m_{i,j}$ of RNA j molecules in cell i is given by

$$m_{i,j} = 96 \times 45 \times 2^{30-22-Cq_{i,j}}.$$

We consistently set $m_{i,j} = 0$ when $Cq_{i,j} = 999$, and $m_{i,j} = \text{UD}$ when $Cq_{i,j} = \text{UD}$.

Replacing missing values. Since some statistical tools (like PCA) do not support missing values, the UDs had to be replaced with some *appropriate* numerical values, i.e., that do not change the data distribution, nor introduce any artificial correlation.

To this end, we calibrated the marginal distribution of each gene at each time-point using the 3-parameter Poisson-Beta family, which corresponds to the stationary distribution of the

widely-used “two-state” model of gene expression [39–41]. As emphasized in [41], it can be obtained as the mixture distribution $\mathcal{D}(a, b, c)$ of X resulting from the hierarchical model

$$\begin{cases} Z \sim \text{Beta}(a, b) \\ X \sim \mathcal{P}(cZ) \end{cases}$$

where a , b , and c are positive. Thus for each time-point t and each gene j , we estimated the parameters a_j^t , b_j^t and c_j^t by taking the absolute value of the moment-based estimators proposed in [39]. Note that these slightly modified estimators are also convergent since the parameters are assumed to be positive. This estimation was only performed for genes with at least 20 valid cells and conducted to delete genes with too many UDs. This led us to delete two genes, resulting in a total of 90 genes analysed. The data was fitted very well in practice, making it relevant to simply replace the UDs with independent samples from the corresponding distributions $\mathcal{D}(a_j^t, b_j^t, c_j^t)$. Considering the actual inferred parameter regime (large values of c , meaning that the numbers of molecules span a high range) and the continuous nature of our data, we actually ignored the Poisson step and sampled from $c_j^t \text{Beta}(a_j^t, b_j^t) \approx \mathcal{D}(a_j^t, b_j^t, c_j^t)$.

Obviously, such artificially generated values should not be seen as data, but they ensure that the dimension-reduction algorithms perform at their best and compute relevant projection axes (e.g., the main two axes for a PCA). We checked that indeed consistent PCA outputs were generated from different UD replacement operations (not shown).

Technical Reproducibility

Since RT-qPCR experimental procedure introduces unavoidable technical noise, we decided to explore which steps were the main sources of this variability (S1 Fig). We first assessed the reproducibility of the cDNA pre-amplification step by amplifying four cDNA samples from the same RT before analyzing it by qPCR. Gene expression levels differences between pre-amplification replicates were found to be negligible (S1A and S1B Fig). We then checked the RT-qPCR amplification step by analyzing the *RPL22L1* gene three times per chip. Expression levels between *RPL22L1* triplicates were quantitatively extremely similar (S1C to S1E Fig), confirming that amplification brings a negligible amount of variability as previously shown [42, 43]. We also tested the experimental variability induced by the RT reaction. We observed significant gene expression level differences between three RT from the same sample (S1A and S1F Fig), contrary to replicates from other critical steps. Indeed, it has been demonstrated and discussed that the RT reaction is the main source of technical noise, since it introduces biases through priming efficiency, RNA integrity and secondary structures and reverse transcriptase dynamic range [42, 44, 45]. In order to estimate the amount of variation introduced in our experiments by this step, we used external RNA spikes. The variation affecting those spikes spanned 5.8 Cqs (mean of $C_{q_{\max}} - C_{q_{\min}}$ across the spikes) whereas the variability affecting the genes spanned a much larger region of 22.9 Cqs (mean of $C_{q_{\max}} - C_{q_{\min}}$ across the genes), showing that the biological variability was much larger than the variability introduced by the RT step.

Statistical Analysis

Software. Most of the statistical analyses were performed using R [46]. The k -means clustering was performed using the `stats` R library. PCAs were performed using the `ade4` package [47]. All PCAs were centered (mean subtraction) and normalized (dividing by the standard deviation). All PCAs were displayed according to PC1 and PC2, which are the first and second axis of the PCA respectively. Hierarchical cluster analysis was performed applying

the `R` `hclust` function, using the complete linkage method on Euclidean distances. Dendrograms were built and plotted using the `dendextend` `R` library. Correlation analysis was performed using `rcorr` from the `Hmisc` `R` library. The p -value was corrected for multiple testing using the Bonferroni method [48]. Networks were computed using Cytoscape [49]. Cross-correlation analysis was performed using the `matcor` function from the `CCA` `R` library. Normality of the distributions was tested using the `shapiro.test` function. The variances were compared using the `F` test with the `var.test` function. Wilcoxon test was performed using the `wilcox.test` function. t-SNE and diffusion maps were computed using the Matlab Toolbox for Dimensionality Reduction (<http://lvdmaaten.github.io/drtoolbox/>). The t-SNE analysis was performed on a normalized version of the data, using `zscore` function. Kernel PCA was computed using the Matlab `kPCA` script [50] applying polynomial with fractional power 0.1. All linear analysis methods (PCA, HCA and correlation analysis) were performed after applying the transformation $m \mapsto \ln(m + 1)$ to the data, which gives access to the more linear C_q structure. All non-linear analysis methods (t-SNE, diffusion maps and Kernel PCA) were performed using untransformed m values.

I score calculation. The I score was calculated as previously described in [51] as follows: among the $n = 90$ studied genes, we defined a subset D containing n_D genes. We then defined the I score as:

$$I = CV \frac{PCC_{in}}{PCC_{out}}$$

with

$$CV = \frac{1}{n_D} \sum_{i \in D} CV_i, \quad PCC_{in} = \frac{1}{n_D^2} \sum_{i,j \in D} C_{ij}, \quad PCC_{out} = \frac{1}{n_D(n - n_D)} \sum_{\substack{i \in D \\ j \notin D}} C_{ij}$$

where CV_i is the coefficient of variation of gene i and C_{ij} stands for Pearson's correlation coefficient between genes i and j .

Wave analysis. One thousand boot-strap expression matrices were generated from genes RNA counts distribution for each time-point (0, 2, 4, and 8 h). New expression matrices were generated by uniform sampling of cells, which correspond to matrix lines, using the `randsample` Matlab command with replacement. For each time-point combination, a Mann-Whitney U test was performed using the `ranksum` Matlab command to detect significant variation. Wave membership was based on time variations. By definition a gene belongs to the wave at time T if there is at least one variation detected between time T and a previous time-point and if the gene does not belong to a previous wave. Only genes identified in a wave that displayed a significant variation in more than 90% of boot-strap samples were kept in this wave.

Estimation of entropy. We estimated the Shannon entropy of each gene j at each time-point t as follows: we computed basic histograms of the genes with $N = N_c/2$ bins, where N_c is the number of cells, which provided the probabilities $p_{j,k}^t$ of each class k . Finally, the entropies were defined by

$$E_j^t = - \sum_{k=1}^N p_{j,k}^t \log_2(p_{j,k}^t).$$

When all cells express the same amount of a given gene, this gene's entropy will be null. On the contrary, the maximum value of entropy will result from the most variable gene expression level (S2 Fig).

Re-ordering algorithms. We performed the pseudotemporal ordering of cells using three different algorithms: SCUBA [52], WANDERLUST [53] and TSCAN [54]. SCUBA is a two-step cell-ordering algorithm, in which one first reduces the data dimensionality by using t-SNE [55] and then determines the principal curve in the low-dimensional projection. We applied SCUBA by reducing the data into 2-D using tSNE (perplexity = 30) and by adopting k-segments algorithm (maximal number of segments = 8) as the option for the principal curve analysis. Since the differentiation path estimated by SCUBA was undirected, we set *LDHA* as the anchor-gene/marker to define the beginning and the end of pseudotime.

In contrast, WANDERLUST is a non-branching trajectory detection algorithm [53]. The method estimates the pseudotimes by representing each single-cell as a node in an ensemble of k-nearest-neighbor graph, followed by assigning a trajectory for each graph. This trajectory is defined by connecting cells with similar gene expressions through the shortest path. To reinforce this path assembly, a set of cells is randomly chosen as waypoints. The final cell ordering corresponds to the average trajectories over the ensemble of graphs. Here, we adopted the cosine similarity distance function for the trajectory detection, in which the single cell with the maximum *LDHA* expression was used as the initial node. Each cell's pseudotime has a value normalized between 0 and 1, reflecting its position along the differentiation path. For the entropy calculation, we grouped the cells into five pseudo-clusters, by collecting cells within five evenly spaced pseudotime window between 0 and 1 (e.g., pseudo-cluster 1 contained cells with pseudotime between 0 and 0.2, pseudo-cluster 2 contained cells with pseudotime between 0.2 and 0.4, and so on).

Finally, TSCAN is a cluster-based minimum spanning tree ordering algorithm [54]. The algorithm begins with clustering cells according to the similarity in their gene expressions, and continues with building the minimum spanning tree (MST) connecting the centroids of these clusters. The pseudotime is calculated by projecting each single cell to the MST edges. The algorithm also implements a preprocessing step involving gene clustering and dimensional reduction in order to alleviate the effect of drop-out events [54]. The preprocessing of our data produced 36 gene clusters, on which we employed the independent component analysis (ICA) to obtain a 2-D projection. Finally, we applied TSCAN using five cell clusters to generate the cell pseudotimes.

We computed the entropy for each cluster of cells following the procedure described above.

In silico simulations of mRNA level for single cells. In silico results were generated using the two-state model of gene expression [27, 39–41, 56]. We first inferred a set of model parameters (K_{on} , K_{off} , S_0 , D_0) specific to each gene and depending on time. For that we used an inference method based on moment analysis [39] from our single cell expression matrix allowing to estimate three of these parameters (K_{on} , K_{off} and S_0). To estimate D_0 (mRNA degradation rate) we used population data of mRNA decay kinetic using actinomycin D-treated T2EC (osf.io/k2q5b). To simulate mRNA level we used the Gillespie algorithm [57]. In order to validate this modeling approach, we simulated for a given gene its mRNA evolution for 100 cells and extracted its distribution among cells at different time-points (0, 8, 24, 33, 48, and 72 h). We then compared in vitro and in silico distributions with a non-parametric Mann-Whitney U test. In silico measurements reproduced qualitatively the evolution of mean and distributions measured in vitro (not shown).

In silico simulations of the differentiation process. In order to stabilize the model before differentiation start, we ran the simulation for 100 h (model time) with constant parameters (value corresponding to 0 h). In silico differentiation was induced by a change in parameters values to now impose the parameters deduced from the in vitro data at different time-points. At each time step we computed parameters value with a linear interpolation between the two nearest time-points. For example at simulation time 4 h parameters

values correspond to the mean value between 0 and 8 h. We simulated 100 cells at each time-point. In order to study the impact of asynchronous differentiation, we compared two situations:

1. All cells had their parameters changed simultaneously, corresponding to a synchronous differentiation.
2. We randomly chose for each cell a time lag from a uniform distribution between 0 and 24 h. Then during the simulation, parameters started to change at $t = 0 \text{ h} + \text{time lag}$. This corresponded to an asynchronous differentiation.

We then used the same metrics for analyzing those *in silico* distributions as those used for analyzing the *in vitro* data.

scRNA-seq data analysis. Counting table from [58] was downloaded from the following URL: <http://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE67310>. The original (Log₂ [FKPM]) data were transformed into FKPM data for analysis using the BPglm algorithm [59]. Running the algorithm with an FDR value of less than 0.00005 and using the Bonferroni correction method for multiple testing led us to a list of 776 differentially expressed genes, on which entropy was computed. Statistical significance was computed using the Wilcoxon non parametric test.

Supporting Information

S1 Fig. Reproducibility of the pre-amplification and RT-qPCR amplification steps. (A) the protocol used for assessing variation sources; (B) variations induced by four independent pre-amplifications when assessing the level of expression of the OSC gene; (C–E) variations induced by the PCR amplification step. The *RPL22L1* gene expression was analyzed three times per single-cell. Shown is the correlation between those three RT-qPCR replicates. The corresponding correlation coefficients are plotted on the graphs. The slopes of the linear regression lines are 0.99 for all three comparisons; (F) variations induced by three independent reverse-transcriptions when assessing the level of expression of the OSC gene. (PDF)

S2 Fig. Schematic description of the entropy value. On the left are shown gene expression values that are transformed into probabilities (*p*) to observe a given expression level in a cell population. The upper case illustrates the deterministic case where all cells do express the same expression level, resulting in a probability of 1 of observing such a level. This results in a null entropy (see [Materials and Methods](#) for the calculation). The lower case illustrates the other extreme case, where all the cells have different expression level, resulting in a much higher entropy. (PDF)

S3 Fig. Scatter and MA plots showing the reproducibility of read counts between replicates and the differential expression during the differentiation process. (A,B) Relationship between biological replicates of two independent RNA-Seq experiments: self-renewing T2EC (left panel) and T2EC induced to differentiate for 48 h (right panel). For each condition, the *x*-axis represents the read counts of the first biological experiment, whereas read counts of the second biological replicate are given on the *y*-axis. Each dot corresponds to the expression level of one gene. (C) Comparative analysis of RNA-Seq data generated from two independent libraries of T2EC in self-renewing state and T2EC induced to differentiate for 48 h. The *x*-axis shows the expression level of each gene (transcript raw counts divided by the library size and multiplied by 1 million, averaged between the two independent libraries)

while the fold change (self-renewal versus differentiation) appears in the y -axis. Red-colored dots highlights genes that are significantly differentially expressed (p -value < 0.05).
(PDF)

S4 Fig. Identification of common patterns of expression during the differentiation process using K-means clustering. K-means clustering was used to separate the 110 selected genes into seven clusters regarding the expression profiles along the differentiation process. Starting models of gene expression pattern, corresponding to the centroid of each cluster, are represented in the first graph (starting cluster). We identified seven patterns of gene expressions with increasing, decreasing and one complex (cluster 4) dynamic profiles. The final centroid was recalculated after gene allotment, and might slightly differ from the starting one.
(PDF)

S5 Fig. Representation of the 92 selected genes. (A) On the basis of RNA-Seq data and k-means analysis (S4 Fig), the 92 genes selected for the single-cell analysis (S1 Table) can be separated into three types: up-regulated (red circles), invariant (green circles), and down-regulated genes (blue circles) at 48 h of the differentiation process. For each gene (x -axis) the fold-change (FC) between the self-renewal state and the differentiation state at 48 h (Diff/SR) was plotted along the y -axis. (B) Representation of known connections among the 92 genes selected according to the STRING database (<http://string.embl.de/>). Each edge between two genes corresponds to a known association between those genes. The densely connected component at the center of the network graph is composed of genes involved in sterol biosynthesis. A cluster of gene encoding proteins involved in signal transduction is apparent on the top right part of the network.
(PDF)

S6 Fig. Cross-correlation analysis between the gene expression value in populations and in single cells. The correlation matrix is divided into four smaller matrices: the correlation matrix of each dataset (populations: top-left panel; single-cells: bottom-right panel) and the correlation matrix between the two datasets (top-right and bottom-left panels, showing the same values). The values of the correlations are color-coded according to the scale given below. Correlation are calculated for each gene either across populations samples or across single cells.
(PDF)

S7 Fig. Distributions of the expression values for three genes up-, down-, and non-regulated during the differentiation process. The histograms show the expression distribution of three genes among single cells at 0 and 72 h differentiation time-points. The gene expression levels (m value) are shown on the x -axis, the number of cells (count) is represented on the y -axis.
(PDF)

S8 Fig. Variation of entropy during a reprogramming process. We computed differential gene expression between 0 and 2 d using the scRNA-seq data from [58]. We then computed an entropy value per time-point for the 776 resulting genes. Statistical significance was computed using a Wilcoxon test.
(PDF)

S9 Fig. Overlapping genes between DNBs, early drivers and correlation network nodes at 0–8 h of differentiation. The Venn diagram shows the overlap of the three lists of genes obtained from the initial expression waves analysis (green), the correlation networks (pink), and the DNB theory (blue). The common genes between these lists were searched at 0 and 8 h

when all three analyses have been performed (early driver genes were only identified between 0 and 8 h).

(PDF)

S1 Table. Supplementary Table 1. Shown is the complete list of the 92 genes we analyzed, together with their expression value in the four RNA-Seq libraries (SR_1 and SR_2 being the two independent libraries made using self-renewing cells and Diff_1 and Diff_2 being two independent libraries made from cells differentiated for 48 h) and the group of variation at 48 h to which they belong (up-, down-, or non-regulated).

(CSV)

Acknowledgments

We would like to thank the following colleagues for helpful advice and discussions throughout this work: Vincent Lacroix (LBBE/UCBL), Marie Sémon (IGFL/ENSL), Gaël Yvert (LBMC/ENSL), Didier Auboeuf (LBMC/ENSL), Isabelle Durand (CLB), David Cox (CLB), Nicole Dalla-Venezia (CLB), Jordan C. Moore (Fluidigm), Frédéric Moret (INMG/UCBL), Julien Falk (INMG/UCBL), and all the members of the Stochagène and Iceberg projects. We thank the Génoscope, and especially Carole Dossat, for their invaluable help in sequencing of the RNAseq libraries. We would like to thank Gérard Benoit (LBMC/ENSL), Marieke von Lindern (Sanquin, Amsterdam), and Sui Huang (ICSB, Seattle) for critical reading of the manuscript. We thank Geneviève Fourel for pointing our attention to the [58] paper.

Author Contributions

Conceptualization: OG UH AB AR SGG JJK NPG RG JC.

Formal analysis: AR LB UH AB NPG RG TE OG.

Funding acquisition: OG SGG.

Investigation: AR VM EV AG OA.

Supervision: OG SGG.

Writing – original draft: AR SGG OG.

Writing – review & editing: OG AR SGG JJK LB JC RG NPG.

References

1. Wolff L, Humeniuk R. Concise review: erythroid versus myeloid lineage commitment: regulating the master regulators. *Stem Cells*. 2013; 31(7):1237–44. doi: [10.1002/stem.1379](https://doi.org/10.1002/stem.1379) PMID: [23559316](https://pubmed.ncbi.nlm.nih.gov/23559316/)
2. Torres-Padilla ME, Chambers I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development*. 2014; 141(11):2173–81. doi: [10.1242/dev.102624](https://doi.org/10.1242/dev.102624) PMID: [24866112](https://pubmed.ncbi.nlm.nih.gov/24866112/)
3. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell*. 2014; 55(2):319–31. doi: [10.1016/j.molcel.2014.06.029](https://doi.org/10.1016/j.molcel.2014.06.029) PMID: [25038413](https://pubmed.ncbi.nlm.nih.gov/25038413/)
4. Luo Y, Lim CL, Nichols J, Martinez-Arias A, Wernisch L. Cell signalling regulates dynamics of Nanog distribution in embryonic stem cell populations. *J R Soc Interface*. 2012; doi: [10.1098/rsif.2012.0525](https://doi.org/10.1098/rsif.2012.0525)
5. Chickarmane V, Olariu V, Peterson C. Probing the role of stochasticity in a model of the embryonic stem cell: heterogeneous gene expression and reprogramming efficiency. *BMC Syst Biol*. 2012; 6:98. doi: [10.1186/1752-0509-6-98](https://doi.org/10.1186/1752-0509-6-98) PMID: [22889237](https://pubmed.ncbi.nlm.nih.gov/22889237/)

6. Sturrock M, Hellander A, Matzavinos A, Chaplain MA. Spatial stochastic modelling of the Hes1 gene regulatory network: intrinsic noise can explain heterogeneity in embryonic stem cell differentiation. *J R Soc Interface*. 2013; 10(80):20120988. doi: [10.1098/rsif.2012.0988](https://doi.org/10.1098/rsif.2012.0988) PMID: [23325756](https://pubmed.ncbi.nlm.nih.gov/23325756/)
7. Ochiai H, Sugawara T, Sakuma T, Yamamoto T. Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci Rep*. 2014; 4:7125. doi: [10.1038/srep07125](https://doi.org/10.1038/srep07125) PMID: [25410303](https://pubmed.ncbi.nlm.nih.gov/25410303/)
8. Wu J, Tzanakakis ES. Deconstructing stem cell population heterogeneity: Single-cell analysis and modeling approaches. *Biotechnol Adv*. 2013; 31:1047–1062. doi: [10.1016/j.biotechadv.2013.09.001](https://doi.org/10.1016/j.biotechadv.2013.09.001) PMID: [24035899](https://pubmed.ncbi.nlm.nih.gov/24035899/)
9. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*. 2012; 150(6):1209–22. doi: [10.1016/j.cell.2012.08.023](https://doi.org/10.1016/j.cell.2012.08.023) PMID: [22980981](https://pubmed.ncbi.nlm.nih.gov/22980981/)
10. Haas S, Hansson J, Klimmeck D, Loeffler D, Velten L, Uckelmann H, et al. Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell*. 2015; doi: [10.1016/j.stem.2015.07.007](https://doi.org/10.1016/j.stem.2015.07.007) PMID: [26299573](https://pubmed.ncbi.nlm.nih.gov/26299573/)
11. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J, et al. Inferring rules of lineage commitment in haematopoiesis. *Nat Cell Biol*. 2012; 14(3):287–94. doi: [10.1038/ncb2442](https://doi.org/10.1038/ncb2442) PMID: [22344032](https://pubmed.ncbi.nlm.nih.gov/22344032/)
12. Kouno T, de Hoon M, Mar JC, Tomaru Y, Kawano M, Carninci P, et al. Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome Biol*. 2013; 14(10):R118. doi: [10.1186/gb-2013-14-10-r118](https://doi.org/10.1186/gb-2013-14-10-r118) PMID: [24156252](https://pubmed.ncbi.nlm.nih.gov/24156252/)
13. Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G. Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science*. 2008; 321(5892):1081–4. doi: [10.1126/science.1158013](https://doi.org/10.1126/science.1158013) PMID: [18719282](https://pubmed.ncbi.nlm.nih.gov/18719282/)
14. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 2013; 498(7453):236–40. doi: [10.1038/nature12172](https://doi.org/10.1038/nature12172) PMID: [23685454](https://pubmed.ncbi.nlm.nih.gov/23685454/)
15. Arsenio J, Kakaradov B, Metz PJ, Kim SH, Yeo GW, Chang JT. Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat Immunol*. 2014; 15(4):365–72. doi: [10.1038/ni.2842](https://doi.org/10.1038/ni.2842) PMID: [24584088](https://pubmed.ncbi.nlm.nih.gov/24584088/)
16. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. 2015; 33(2):155–60. doi: [10.1038/nbt.3102](https://doi.org/10.1038/nbt.3102) PMID: [25599176](https://pubmed.ncbi.nlm.nih.gov/25599176/)
17. Helmstetter C, Flossdorf M, Peine M, Kupz A, Zhu J, Hegazy AN, et al. Individual T helper cells have a quantitative cytokine memory. *Immunity*. 2015; 42(1):108–22. doi: [10.1016/j.immuni.2014.12.018](https://doi.org/10.1016/j.immuni.2014.12.018) PMID: [25607461](https://pubmed.ncbi.nlm.nih.gov/25607461/)
18. Lu Y, Xue Q, Eisele MR, Sulistijo ES, Brower K, Han L, et al. Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proc Natl Acad Sci U S A*. 2015; 112(7):E607–15. doi: [10.1073/pnas.1416756112](https://doi.org/10.1073/pnas.1416756112) PMID: [25646488](https://pubmed.ncbi.nlm.nih.gov/25646488/)
19. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002; 297(5584):1183–1186. doi: [10.1126/science.1070919](https://doi.org/10.1126/science.1070919) PMID: [12183631](https://pubmed.ncbi.nlm.nih.gov/12183631/)
20. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011; 332(6028):472–4. doi: [10.1126/science.1198817](https://doi.org/10.1126/science.1198817) PMID: [21415320](https://pubmed.ncbi.nlm.nih.gov/21415320/)
21. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*. 2005; 6(6):451–64. doi: [10.1038/nrg1615](https://doi.org/10.1038/nrg1615) PMID: [15883588](https://pubmed.ncbi.nlm.nih.gov/15883588/)
22. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods*. 2008; 5(10):877–9. doi: [10.1038/nmeth.1253](https://doi.org/10.1038/nmeth.1253) PMID: [18806792](https://pubmed.ncbi.nlm.nih.gov/18806792/)
23. Lestas I, Vinnicombe G, Paulsson J. Fundamental limits on the suppression of molecular fluctuations. *Nature*. 2010; 467(7312):174–8. doi: [10.1038/nature09333](https://doi.org/10.1038/nature09333) PMID: [20829788](https://pubmed.ncbi.nlm.nih.gov/20829788/)
24. Viñuelas J, Kaneko G, Coulon A, Beslon G, Gandrillon O. Toward experimental manipulation of stochasticity in gene expression. *Progress in Biophysics and Molecular Biology*. 2012; 110:44–53. doi: [10.1016/j.pbiomolbio.2012.04.010](https://doi.org/10.1016/j.pbiomolbio.2012.04.010) PMID: [22609563](https://pubmed.ncbi.nlm.nih.gov/22609563/)
25. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet*. 2011; 43(2):95–100. doi: [10.1038/ng.729](https://doi.org/10.1038/ng.729) PMID: [21186354](https://pubmed.ncbi.nlm.nih.gov/21186354/)
26. Cagatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Suel GM. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*. 2009; 139(3):512–22. doi: [10.1016/j.cell.2009.07.046](https://doi.org/10.1016/j.cell.2009.07.046) PMID: [19853288](https://pubmed.ncbi.nlm.nih.gov/19853288/)

27. Viñuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, et al. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biology*. 2013; 11:15. doi: [10.1186/1741-7007-11-15](https://doi.org/10.1186/1741-7007-11-15) PMID: [23442824](https://pubmed.ncbi.nlm.nih.gov/23442824/)
28. Kupiec JJ. A Darwinian theory for the origin of cellular differentiation. *Mol Gen Genet*. 1997; 255(2):201–8. doi: [10.1007/s004380050490](https://doi.org/10.1007/s004380050490) PMID: [9236778](https://pubmed.ncbi.nlm.nih.gov/9236778/)
29. Huang S. Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Philosophical transactions Series A, Mathematical, physical, and engineering sciences*. 2011; 366:2247–59.
30. Yvert G. 'Particle genetics': treating every cell as unique. *Trends Genet*. 2014; 30(2):49–56. doi: [10.1016/j.tig.2013.11.002](https://doi.org/10.1016/j.tig.2013.11.002) PMID: [24315431](https://pubmed.ncbi.nlm.nih.gov/24315431/)
31. Rebhahn JA, Deng N, Sharma G, Livingstone AM, Huang S, Mosmann TR. An animated landscape representation of CD4(+) T-cell differentiation, variability, and plasticity: Insights into the behavior of populations versus cells. *Eur J Immunol*. 2014; 44(8):2216–29. doi: [10.1002/eji.201444645](https://doi.org/10.1002/eji.201444645) PMID: [24945794](https://pubmed.ncbi.nlm.nih.gov/24945794/)
32. Gandrillon O, Schmidt U, Beug H, Samarut J. TGF-beta cooperates with TGF-alpha to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *Embo J*. 1999; 18(10):2764–2781. doi: [10.1093/emboj/18.10.2764](https://doi.org/10.1093/emboj/18.10.2764) PMID: [10329623](https://pubmed.ncbi.nlm.nih.gov/10329623/)
33. Damiola F, Keime C, Gonin-Giraud S, Dazy S, Gandrillon O. Global transcription analysis of immature avian erythrocytic progenitors: from self-renewal to differentiation. *Oncogene*. 2004; 23:7628–7643. doi: [10.1038/sj.onc.1208061](https://doi.org/10.1038/sj.onc.1208061) PMID: [15378009](https://pubmed.ncbi.nlm.nih.gov/15378009/)
34. Bresson C, Gandrillon O, Gonin-Giraud S. sca2: a new gene involved in the self-renewal of erythroid progenitors. *Cell Proliferation*. 2008; 41:726–738. doi: [10.1111/j.1365-2184.2008.00554.x](https://doi.org/10.1111/j.1365-2184.2008.00554.x)
35. Mejia-Pous C, Damiola F, Gandrillon O. Cholesterol synthesis-related enzyme oxidosqualene cyclase is required to maintain self-renewal in primary erythroid progenitors. *Cell Prolif*. 2011; 44(5):441–52. doi: [10.1111/j.1365-2184.2011.00771.x](https://doi.org/10.1111/j.1365-2184.2011.00771.x) PMID: [21951287](https://pubmed.ncbi.nlm.nih.gov/21951287/)
36. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7(3):562–78. doi: [10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016) PMID: [22383036](https://pubmed.ncbi.nlm.nih.gov/22383036/)
37. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2014; doi: [10.1093/bioinformatics/btu638](https://doi.org/10.1093/bioinformatics/btu638) PMID: [25260700](https://pubmed.ncbi.nlm.nih.gov/25260700/)
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. doi: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) PMID: [19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/)
39. Peccoud J, Ycart B. Markovian Modelling of Gene Product Synthesis. *Theoretical population biology*. 1995; 48:222–234. doi: [10.1006/tpbi.1995.1027](https://doi.org/10.1006/tpbi.1995.1027)
40. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci U S A*. 2008; 105(45):17256–61. doi: [10.1073/pnas.0803850105](https://doi.org/10.1073/pnas.0803850105) PMID: [18988743](https://pubmed.ncbi.nlm.nih.gov/18988743/)
41. Kim JK, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol*. 2013; 14(1):R7. doi: [10.1186/gb-2013-14-1-r7](https://doi.org/10.1186/gb-2013-14-1-r7) PMID: [23360624](https://pubmed.ncbi.nlm.nih.gov/23360624/)
42. Stahlberg A. Comparison of reverse transcriptases in gene expression analysis. *clin Chem*. 2004; 50:1678–80. doi: [10.1373/clinchem.2004.035469](https://doi.org/10.1373/clinchem.2004.035469) PMID: [15331507](https://pubmed.ncbi.nlm.nih.gov/15331507/)
43. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001; 29(9):e45. doi: [10.1093/nar/29.9.e45](https://doi.org/10.1093/nar/29.9.e45) PMID: [11328886](https://pubmed.ncbi.nlm.nih.gov/11328886/)
44. Brooks EM, Sheflin LG, Spaulding SW. Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR. *BioTechniques*. 1995; 19:806–815. PMID: [8588921](https://pubmed.ncbi.nlm.nih.gov/8588921/)
45. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol*. 2002; 29:23–39. doi: [10.1677/jme.0.0290023](https://doi.org/10.1677/jme.0.0290023) PMID: [12200227](https://pubmed.ncbi.nlm.nih.gov/12200227/)
46. Team RDC. R: A language and environment for statistical computing. Vienna, Austria ISBN 3-900051-07-0, URL <http://www.R-project.org>. 2008;.
47. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*. 2007; 22:1–20.
48. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995; 310(6973):170. doi: [10.1136/bmj.310.6973.170](https://doi.org/10.1136/bmj.310.6973.170) PMID: [7833759](https://pubmed.ncbi.nlm.nih.gov/7833759/)
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13(11):2498–504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303) PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)

50. Wang Q. Kernel Principal Component Analysis and its Applications in Face Recognition and Active Shape Models.; 2012.
51. Liu R, Chen P, Aihara K, Chen L. Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Sci Rep.* 2015; 5:17501. doi: [10.1038/srep17501](https://doi.org/10.1038/srep17501) PMID: [26647650](https://pubmed.ncbi.nlm.nih.gov/26647650/)
52. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc Natl Acad Sci U S A.* 2014; 111(52):E5643–50. doi: [10.1073/pnas.1408993111](https://doi.org/10.1073/pnas.1408993111) PMID: [25512504](https://pubmed.ncbi.nlm.nih.gov/25512504/)
53. Bendall SC, Davis KL, Amir el AD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell.* 2014; 157(3):714–25. doi: [10.1016/j.cell.2014.04.005](https://doi.org/10.1016/j.cell.2014.04.005) PMID: [24766814](https://pubmed.ncbi.nlm.nih.gov/24766814/)
54. Ji Z, Ji H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 2016; doi: [10.1093/nar/gkw430](https://doi.org/10.1093/nar/gkw430)
55. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research.* 2008; 9:2579–2605.
56. Paulsson J. Models of stochastic gene expression. *Phys Life Rev.* 2005; 2:157–175. doi: [10.1016/j.plrev.2005.03.003](https://doi.org/10.1016/j.plrev.2005.03.003)
57. Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics.* 1976; 22(4):403–434. doi: [10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3)
58. Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature.* 2016; 534(7607):391–5. doi: [10.1038/nature18323](https://doi.org/10.1038/nature18323) PMID: [27281220](https://pubmed.ncbi.nlm.nih.gov/27281220/)
59. Vu TN, Wills QF, Kalari KR, Niu N, Wang L, Rantalainen M, et al. Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics.* 2016; doi: [10.1093/bioinformatics/btw202](https://doi.org/10.1093/bioinformatics/btw202) PMID: [27153638](https://pubmed.ncbi.nlm.nih.gov/27153638/)
60. Dennis J G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003; 4(5):P3. doi: [10.1186/gb-2003-4-5-p3](https://doi.org/10.1186/gb-2003-4-5-p3)
61. Tian WL, He F, Fu X, Lin JT, Tang P, Huang YM, et al. High expression of heat shock protein 90 alpha and its significance in human acute leukemia cells. *Gene.* 2014; 542(2):122–8. doi: [10.1016/j.gene.2014.03.046](https://doi.org/10.1016/j.gene.2014.03.046) PMID: [24680776](https://pubmed.ncbi.nlm.nih.gov/24680776/)
62. Dong H, Zou M, Bhatia A, Jayaprakash P, Hofman F, Ying Q, et al. Breast Cancer MDA-MB-231 Cells Use Secreted Heat Shock Protein-90alpha (Hsp90alpha) to Survive a Hostile Hypoxic Environment. *Sci Rep.* 2016; 6:20605. doi: [10.1038/srep20605](https://doi.org/10.1038/srep20605) PMID: [26846992](https://pubmed.ncbi.nlm.nih.gov/26846992/)
63. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics.* 2015; doi: [10.1093/bioinformatics/btv325](https://doi.org/10.1093/bioinformatics/btv325)
64. Liu C. Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition. *IEEE transactions on pattern analysis and machine intelligence.* 2004; 26:572–581. doi: [10.1109/TPAMI.2004.1273927](https://doi.org/10.1109/TPAMI.2004.1273927) PMID: [15460279](https://pubmed.ncbi.nlm.nih.gov/15460279/)
65. Piras V, Selvarajoo K. Reduction of gene expression variability from single cells to populations follows simple statistical laws. *Genomics.* 2014; PMID: [25554103](https://pubmed.ncbi.nlm.nih.gov/25554103/)
66. Pina C, Teles J, Fugazza C, May G, Wang D, Guo Y, et al. Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. *Cell Rep.* 2015; 11(10):1503–10. doi: [10.1016/j.celrep.2015.05.016](https://doi.org/10.1016/j.celrep.2015.05.016) PMID: [26051941](https://pubmed.ncbi.nlm.nih.gov/26051941/)
67. Singh A. Cell-Cycle Control of Developmentally Regulated Transcription Factors Accounts for Heterogeneity in Human Pluripotent Cells. *Stem cell reports.* 2013; 1:532–44. doi: [10.1016/j.stemcr.2013.10.009](https://doi.org/10.1016/j.stemcr.2013.10.009) PMID: [24371808](https://pubmed.ncbi.nlm.nih.gov/24371808/)
68. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A.* 2002; 99(20):12795–800. doi: [10.1073/pnas.162041399](https://doi.org/10.1073/pnas.162041399) PMID: [12237400](https://pubmed.ncbi.nlm.nih.gov/12237400/)
69. Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, et al. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell.* 2015; 58(2):339–52. doi: [10.1016/j.molcel.2015.03.005](https://doi.org/10.1016/j.molcel.2015.03.005) PMID: [25866248](https://pubmed.ncbi.nlm.nih.gov/25866248/)
70. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature.* 2008; 453(7194):544–7. doi: [10.1038/nature06965](https://doi.org/10.1038/nature06965) PMID: [18497826](https://pubmed.ncbi.nlm.nih.gov/18497826/)

71. Klenova EM, Fagerlie S, Filippova GN, Kretzner L, Goodwin GH, Loring G, et al. Characterization of the chicken CTCF genomic locus, and initial study of the cell cycle-regulated promoter of the gene. *J Biol Chem*. 1998; 273(41):26571–9. doi: [10.1074/jbc.273.41.26571](https://doi.org/10.1074/jbc.273.41.26571) PMID: [9756895](https://pubmed.ncbi.nlm.nih.gov/9756895/)
72. Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. Tunability and noise dependence in differentiation dynamics. *Science*. 2007; 315(5819):1716–9. doi: [10.1126/science.1137455](https://doi.org/10.1126/science.1137455) PMID: [17379809](https://pubmed.ncbi.nlm.nih.gov/17379809/)
73. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010; 467(7312):167–73. doi: [10.1038/nature09326](https://doi.org/10.1038/nature09326) PMID: [20829787](https://pubmed.ncbi.nlm.nih.gov/20829787/)
74. Larson DR, Singer RH, Zenklusen D. A single molecule view of gene expression. *Trends Cell Biol*. 2009; 19(11):630–7. doi: [10.1016/j.tcb.2009.08.008](https://doi.org/10.1016/j.tcb.2009.08.008) PMID: [19819144](https://pubmed.ncbi.nlm.nih.gov/19819144/)
75. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription Factors Modulate c-Fos Transcriptional Bursts. *Cell Rep*. 2014; 8(1):75–83. doi: [10.1016/j.celrep.2014.05.053](https://doi.org/10.1016/j.celrep.2014.05.053) PMID: [24981864](https://pubmed.ncbi.nlm.nih.gov/24981864/)
76. Dolznig H, Bartunek P, Nasmyth K, Müllner EW, Beug H. Terminal differentiation of normal erythroid progenitors: shortening of G1 correlates with loss of D-cyclin/cdk4 expression and altered cell size control. *Cell Growth Differ*. 1995; 6:1341–1352. PMID: [8562472](https://pubmed.ncbi.nlm.nih.gov/8562472/)
77. von Lindern M, Deiner EM, Dolznig H, Parren-Van Amelsvoort M, Hayman MJ, Mullner EW, et al. Leukemic transformation of normal murine erythroid progenitors: v- and c-ErbB act through signaling pathways activated by the EpoR and c-Kit in stress erythropoiesis. *Oncogene*. 2001; 20(28):3651–3664. doi: [10.1038/sj.onc.1204494](https://doi.org/10.1038/sj.onc.1204494) PMID: [11439328](https://pubmed.ncbi.nlm.nih.gov/11439328/)
78. Le A, Cooper CR, Gouw AM, Dinavahi R, Maitra A, Deck LM, et al. Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc Natl Acad Sci U S A*. 2010; 107(5):2037–42. doi: [10.1073/pnas.0914433107](https://doi.org/10.1073/pnas.0914433107) PMID: [20133848](https://pubmed.ncbi.nlm.nih.gov/20133848/)
79. Wang YH, Israelsen WJ, Lee D, Yu VW, Jeanson NT, Clish CB, et al. Cell-state-specific metabolic dependency in hematopoiesis and leukemogenesis. *Cell*. 2014; 158(6):1309–23. doi: [10.1016/j.cell.2014.07.048](https://doi.org/10.1016/j.cell.2014.07.048) PMID: [25215489](https://pubmed.ncbi.nlm.nih.gov/25215489/)
80. Crauste F, Pujo-Menjouet L, Genieys S, Molina C, Gandrillon O. Adding Self-Renewal in Committed Erythroid Progenitors Improves the Biological Relevance of a Mathematical Model of Erythropoiesis. *J Theor Biol*. 2008; 250:322–338. doi: [10.1016/j.jtbi.2007.09.041](https://doi.org/10.1016/j.jtbi.2007.09.041) PMID: [17997418](https://pubmed.ncbi.nlm.nih.gov/17997418/)
81. Sahu D, Zhao Z, Tsen F, Cheng CF, Park R, Situ AJ, et al. A potentially common peptide target in secreted heat shock protein-90alpha for hypoxia-inducible factor-1alpha-positive tumors. *Mol Biol Cell*. 2012; 23(4):602–13. doi: [10.1091/mbc.E11-06-0575](https://doi.org/10.1091/mbc.E11-06-0575) PMID: [22190738](https://pubmed.ncbi.nlm.nih.gov/22190738/)
82. Takubo K, Nagamatsu G, Kobayashi CI, Nakamura-Ishizu A, Kobayashi H, Ikeda E, et al. Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. *Cell Stem Cell*. 2013; 12(1):49–61. doi: [10.1016/j.stem.2012.10.011](https://doi.org/10.1016/j.stem.2012.10.011) PMID: [23290136](https://pubmed.ncbi.nlm.nih.gov/23290136/)
83. Oburoglu L, Romano M, Taylor N, Kinet S. Metabolic regulation of hematopoietic stem cell commitment and erythroid differentiation. *Curr Opin Hematol*. 2016; 23(3):198–205. doi: [10.1097/MOH.000000000000234](https://doi.org/10.1097/MOH.000000000000234) PMID: [26871253](https://pubmed.ncbi.nlm.nih.gov/26871253/)
84. Wang J, Mi JQ, Debernardi A, Vitte AL, Emadali A, Meyer JA, et al. A six gene expression signature defines aggressive subtypes and predicts outcome in childhood and adult acute lymphoblastic leukemia. *Oncotarget*. 2015; 6(18):16527–42. doi: [10.18632/oncotarget.4113](https://doi.org/10.18632/oncotarget.4113) PMID: [26001296](https://pubmed.ncbi.nlm.nih.gov/26001296/)
85. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. 2015; doi: [10.1038/nrg3833](https://doi.org/10.1038/nrg3833) PMID: [25628217](https://pubmed.ncbi.nlm.nih.gov/25628217/)
86. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. *Development*. 2009; 136(23):3853–62. doi: [10.1242/dev.035139](https://doi.org/10.1242/dev.035139) PMID: [19906852](https://pubmed.ncbi.nlm.nih.gov/19906852/)
87. Huang S, Guo YP, May G, Enver T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol*. 2007; 305(2):695–713. doi: [10.1016/j.ydbio.2007.02.036](https://doi.org/10.1016/j.ydbio.2007.02.036) PMID: [17412320](https://pubmed.ncbi.nlm.nih.gov/17412320/)
88. 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; 14(12):e2000640. doi: [10.1371/journal.pbio.2000640](https://doi.org/10.1371/journal.pbio.2000640)
89. Chen P, Liu R, Chen L, Aihara K. Identifying critical differentiation state of MCF-7 cells for breast cancer by dynamical network biomarkers. *Front Genet*. 2015; 6:252. doi: [10.3389/fgene.2015.00252](https://doi.org/10.3389/fgene.2015.00252) PMID: [26284108](https://pubmed.ncbi.nlm.nih.gov/26284108/)
90. Stern S, Dror T, Stolovicki E, Brenner N, Braun E. Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol Syst Biol*. 2007; 3:106. doi: [10.1038/msb4100147](https://doi.org/10.1038/msb4100147) PMID: [17453047](https://pubmed.ncbi.nlm.nih.gov/17453047/)
91. Paldi A. What makes the cell differentiate? *Prog Biophys Mol Biol*. 2012; 110(1):41–3. doi: [10.1016/j.pbiomolbio.2012.04.003](https://doi.org/10.1016/j.pbiomolbio.2012.04.003) PMID: [22543273](https://pubmed.ncbi.nlm.nih.gov/22543273/)

92. Pelaez N, Gavaldà-Mirallès A, Wang B, Navarro HT, Gudjonson H, Rebay I, et al. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *Elife*. 2015; 4. doi: [10.7554/eLife.08924](https://doi.org/10.7554/eLife.08924) PMID: [26583752](https://pubmed.ncbi.nlm.nih.gov/26583752/)
93. Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*. 2014; 516(7529):56–61. doi: [10.1038/nature13920](https://doi.org/10.1038/nature13920) PMID: [25471879](https://pubmed.ncbi.nlm.nih.gov/25471879/)

2.2 Article 2 : Automated cell cycle and cell size measurements for single-cell gene expression studies

Au cours de la première partie de ma thèse nous nous sommes interrogés sur les causes, autres que moléculaires, de la variabilité de l'expression génique. La taille et le cycle cellulaires faisant partie des causes suggérées [74, 180, 181], nous avons alors étudié plus en détail l'effet de ces deux paramètres sur la variabilité de l'expression génique. Ainsi, en collaboration avec la plateforme ProfilXpert et le plateau d'imagerie CIQLE, j'ai mis au point les expériences de RTqPCR en cellule unique avec le système C1 de Fluidigm, associé à l'imagerie des cellules analysées. À son arrivée au laboratoire j'ai alors formé Anissa Guillemin, doctorante dans notre équipe, à cette technologie. Anissa a ensuite développé une méthode de marquage pouvant se coupler à ces analyses moléculaires en cellule unique, afin de pouvoir déterminer à quel point la taille et le cycle cellulaires influencent la variabilité de l'expression génique.

Les résultats de cette étude ont récemment été publiés dans le journal BMC Research (article 2) [182]. Ces travaux font suite et sont en adéquation avec les résultats que j'ai obtenus au cours de la première partie de ma thèse.

2.2.1 Principaux résultats de l'article 2

L'objectif de cette étude consiste à proposer une nouvelle méthode permettant de coupler marquages cellulaires et analyse de l'expression génique en cellule unique, afin d'appréhender l'influence de l'hétérogénéité de la taille des cellules et de leur position dans le cycle cellulaire, sur la variabilité de l'expression génique.

Le système C1 de Fluidigm utilise une puce microfluidique possédant 96 pièges, destinés à capturer 96 cellules individuelles pouvant être observées au microscope. Deux marqueurs peu cyto-toxiques, le CFSE et le Hoechst, ont été utilisés afin de mieux visualiser et délimiter les membranes. En effet le CFSE permet de marquer les protéines intracellulaires et de surface, et le Hoechst est un marqueur nucléaire (figure 1 de l'article 2). Après marquage, les cellules ont été chargées dans la puce, et chaque piège a été visualisé au microscope à fluorescence et photographié. La taille des cellules et des noyaux ainsi que l'intensité de fluorescence du Hoechst ont ensuite été mesurées et analysées à partir de ces prises d'images. Les résultats

démontrent que la taille des cellules n'est pas une bonne approximation de leur position dans le cycle cellulaire, contrairement à la taille du noyau, qui corrèle avec l'intensité de fluorescence du Hoechst (figure 2 de l'article 2). Ce double marquage permet donc de mesurer la taille des cellules et leur position dans le cycle cellulaire, de façon indépendante.

Afin de s'assurer que le marquage n'a aucun effet sur l'expression génique, 92 gènes ont été analysés par RTqPCR en système C1 sur des populations de T2EC non marquées et doublement marquées au CFSE et Hoechst. À l'échelle des populations, les résultats montrent que le marquage ne semble pas avoir d'effet sur l'expression des gènes (figure 3a de l'article 2). Cette expérience a ensuite été réalisée à l'échelle de la cellule unique. Une ACP réalisée sur ces données montre que les deux populations de cellules (marquées et non marquées) se chevauchent, suggérant que le double-marquage n'a également pas d'incidence sur l'expression des gènes en cellule unique.

Suite à ces vérifications expérimentales, les effets de la taille et du cycle cellulaire sur la variabilité de l'expression génique ont été testés. Les taux de corrélation entre le nombre d'ARNm pour chaque gène et la taille des cellules ou leur position dans le cycle cellulaire, ont donc été calculés. Les résultats de cette analyse montrent que l'expression d'aucun gène ne corrèle avec la taille ou avec la position dans le cycle cellulaire, suggérant ainsi qu'aucun de ces deux paramètres n'influence les variations de l'expression des gènes.

2.2.2 Principales conclusions de l'article 2

Cette nouvelle méthode couplant analyses moléculaires et imagerie cellulaire, permet d'appréhender des questions importantes soulevées lors des analyses de l'expression génique en cellule unique. En effet, il a été suggéré que l'hétérogénéité de la physiologie cellulaire pourrait influencer la variabilité intercellulaire de l'expression génique. Une telle approche expérimentale permet ainsi de mesurer l'expression des gènes en cellules uniques tout en s'assurant que la variabilité mesurée est principalement moléculaire, et non dû à des variations de taille et de position dans le cycle cellulaire. Ces travaux confirment et renforcent donc les résultats obtenus dans l'article 1, suggérant que le cycle cellulaire et la taille des cellules ne sont pas des sources de la variabilité de l'expression génique.

2.2.3 Article 2

Publié le 1er Février 2018 dans le journal BMC Research

URL : <https://doi.org/10.1186/s13104-018-3195-y>

RESEARCH NOTE

Open Access



Automated cell cycle and cell size measurements for single-cell gene expression studies

Anissa Guillemain^{1*} , Angélique Richard¹, Sandrine Gonin-Giraud¹ and Olivier Gandrillon^{1,2}

Abstract

Objectives: Recent rise of single-cell studies revealed the importance of understanding the role of cell-to-cell variability, especially at the transcriptomic level. One of the numerous sources of cell-to-cell variation in gene expression is the heterogeneity in cell proliferation state. In order to identify how cell cycle and cell size influences gene expression variability at the single-cell level, we provide an universal and automatic toxic-free label method, compatible with single-cell high-throughput RT-qPCR. The method consists of isolating cells after a double-stained, analyzing their morphological parameters and performing a transcriptomic analysis on the same identified cells.

Results: This led to an unbiased gene expression analysis and could be also used for improving single-cell tracking and imaging when combined with cell isolation. As an application for this technique, we showed that cell-to-cell variability in chicken erythroid progenitors was negligibly influenced by cell size nor cell cycle.

Keywords: Cell size, Cell cycle, Gene expression, Single-cell transcriptomic

Introduction

It has been known for decades that isogenic cells can differ from each other in their molecular composition [1, 2]. The refinement of molecular techniques together with computational approaches has recently allowed to get a quantitative view on this cell-to-cell variability. This strongly highlighted the importance of understanding the causes in such variations, leading to a recent turning point in single-cell studies [3–6].

A leading source of cell-to-cell variability can be attributed to stochastic gene expression [7–9]. Numerous factors contribute to cell-to-cell variability such as reactions involving a low-copy number of molecules especially during transcription processes [5, 7, 10, 11] or differences in the internal states of a cell population such as cellular age or cell cycle stage. In literature, we can find

contradictory results regarding the influence of cell cycle and cell size on gene expression. Some studies argued that both of these morphological parameters affect gene expression variation [7, 12–17] whereas, others support that this impact is negligible [18–21].

Stochastic gene expression takes various biological meaning [22–24]. In a cell fate context, stochastic gene expression could drive cells into the differentiation process [25]. It has been shown that during the erythroid differentiation process, we can observe an increase in cell-to-cell variability among genes expression that may participate to the decision making process within differentiation [21].

Together, these information highlight the importance to precisely identify the sources of gene expression variability involved in these phenomena in order to understand their role, and to discard potential confounding effects.

Cell cycle variability can be identified and suppressed by fluorescent-labeling of cell cycle-specific genes, however this method requires genetical alteration of the investigated cells [26]. Other studies, based on computational approach, deconvolute the cell cycle variables in

*Correspondence: anissa.guillemain@ens-lyon.fr

¹ Laboratoire de biologie et modélisation de la cellule. LBMC-Ecole Normale Supérieure-Lyon, Université Claude Bernard Lyon 1, Institut National de la Santé et de la Recherche Médicale: U1210-Ecole Normale Supérieure de Lyon, Centre National de la Recherche Scientifique: UMR5239, 46 Allée d'Italie, 69007 Lyon, France
Full list of author information is available at the end of the article



order to normalize their single-cell gene expression data. Most of them use cell cycle marker genes to train algorithms that can predict cell cycle stage of individual cells [14, 27, 28]. However, these genes have different function or timing according to cell type, even in a same organism [29].

In this article, we propose a more direct approach that consists in measuring morphological parameters in a high-throughput single-cell RT-qPCR study. Using a non-cytotoxic double-staining technique we measured automatically cell cycle phase and cell size of every single-cell isolated from T2EC, a primary chicken erythroid progenitor cells [30]. We demonstrated that the labelling had no detectable effects at the single-cell transcriptomic level in those primary progenitors, suggesting that this technique could be an useful tool for molecular single-cell based studies.

We finally showed that in our cellular system neither cell size nor cell cycle state could be deemed responsible for the cell-to-cell variation we observed, ruling out their potential confounding effects.

Main text

Methods

Cell culture

T2EC were extracted from bone marrow of 19 days-old SPAFAS white leghorn chickens embryos (INRA, Tours, France). The composition of the culture medium has been previously described [21].

Double-staining

Cells were incubated in their initial medium for 30 min with CFSE (5-(and 6)-carboxyfluorescein diacetate succinimidyl ester, Life Tech.) at 5 μ M and Hoechst 33342 (Life Tech.) at 5 μ g/mL at 37 °C in a tube protected from light. After 2 washings in phosphate-buffered saline (PBS, Life Tech.), cells were loaded in the C1 system (Fluidigm).

RT-qPCR at population level

Cell culture were washed with PBS 4 h after the double-staining. Total RNA was extracted using RNeasy Mini Kit (Qiagen).

Reverse transcription assays were performed using the Superscript III First-Strand Synthesis System (Invitrogen) for 500 ng of total RNA.

Real-time PCR was performed with SYBR Green PCR Kit (ClonTech) in the CFX96 real-time PCR system (Bio-rad). Specific primers were used to quantify the expression of genes [21].

RT-qPCR at single-cell level

- From cell isolation to pre-amplification Cells were diluted with C1 cell suspension reagent (Fluidigm) at

a concentration of 4×10^5 cells/mL. This step was followed by a cell filtration in a cellular sieve (50 μ m). Cells were loaded in the C1 IFC (5–10 μ m trap size, Fluidigm). The C1 system performed the cell isolation and pictures were taken with 2 different lasers (UV laser providing excitation at \sim 350 nm and another at \sim 488 nm) using a PALM-STORM NIKON Microscope (CIQLE). Then, the microplate was back in the C1 system where lysis, reverse-transcription and pre-amplification was performed. Primers have been previously described [21]. cDNA were loaded in a classic 96 well plate and conserved at -20 °C until the RT-qPCR.

- Biomark real-time PCR quantification of cDNA were performed using EvaGreen following the Fluidigm user guide available on their website. Each condition was loaded in parallel in the same microfluidic-based chip to avoid chip-to-chip technical variability. An IFC Controller HX performed the load of cDNA samples and primers from the inlets into the chip. The Biomark HD analyzed the chip according to the GE 96 \times 96 PCR + Melt v2.pcl program. RNA spikes were used as positive control to validate the RT-qPCR experiment. From this outlet, the analysis software generated cycle of quantification values (C_q) for each reaction.

ImageJ analysis

Each image corresponding at each lasers used were analyzed following a previously described procedure [31]. We visually confirmed the capture for each well and extracted automatically morphological information using ImageJ. After checking that all cells were detected by the software, we run the measurement of cell area (CFSE), nucleus area and intensity (Hoechst). The cell-volume (2) was then calculated from area measurements (1) using these following formulae:

$$r = \sqrt{\frac{S}{\pi}} \quad (1)$$

$$V = \frac{4}{3} \times \pi \times r^3 \quad (2)$$

with r the radius of cell, S the area and V the cell volume in μm^3 .

Analysis of gene expression

For population RT-qPCR analysis, ratios of gene expression variation between conditions were calculated following this following formulae [32]. Because of its low variability between all conditions, HnRNP was used as referential gene in these analyses.

For single-cell RT-qPCR, raw C_q data was then computed using R [33] via a specific script that was previously described [21]. Some genes were excluded from analyses due to the quality control during the RTqPCR. The output file comprising absolute values of mRNA was used as a template for all following analysis. Statistical non-parametric tests were performed: correlations between gene expression and cell morphological parameters were performed using spearman tests. Wilcoxon tests were used to compare gene expression between stained and unstained conditions. Each time, Bonferroni correction was applied to p-values for the use of multiple tests.

PCA

PCAs were performed using ade4 package [34]. PCA was centered (mean subtraction) and normalized (dividing by the standard deviation). PCA was displayed according to PC1 and PC2, which are the first and second axis of the PCA respectively.

Results

Cellular morphological automatic measuring

We choose the two low toxic fluorescent dyes, CFSE and Hoechst 33342 that stably incorporates into cells. In this study, CFSE was used as a cell area marker in tandem with Hoechst 33342 [35] as a nuclear marker. The use of two different lasers allowed revealing each staining (Fig. 1a, b) merged in Fig. 1c. It allowed us to automatically measure morphological cell parameters and inferred volumes.

We can observe that the cell volume is very poorly correlated with the nucleus volume (Fig. 2a). Therefore cell size by itself does not seem to be a good proxy for determining cell cycle position probably because it integrated other unknown parameters. Both cell and nucleus volume density distributions confirm that cell size spans a

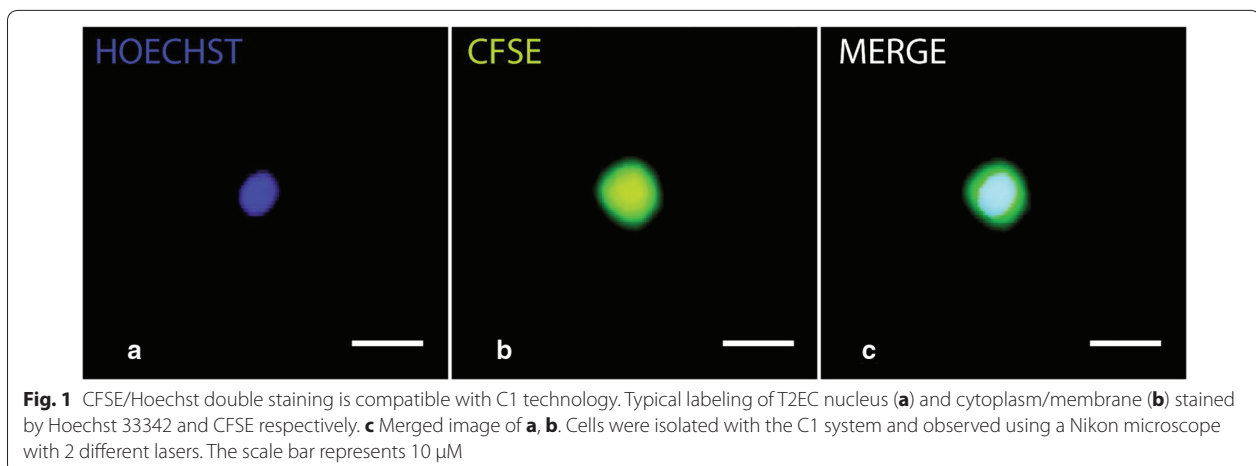
much larger range than the nucleus size which displays the classical $2n/4n$ distribution (Fig. 2b). Nuclear-volume was clearly more correlated with Hoechst fluorescence intensity than cell-volume (Fig. 2a, c). The nucleus volume can therefore be considered as a good indicator for the position of the cell in the cell cycle. Furthermore it should be noted that volume is a purely geometrical object that is not influenced by the laser bleaching, as Hoechst fluorescence intensity parameter.

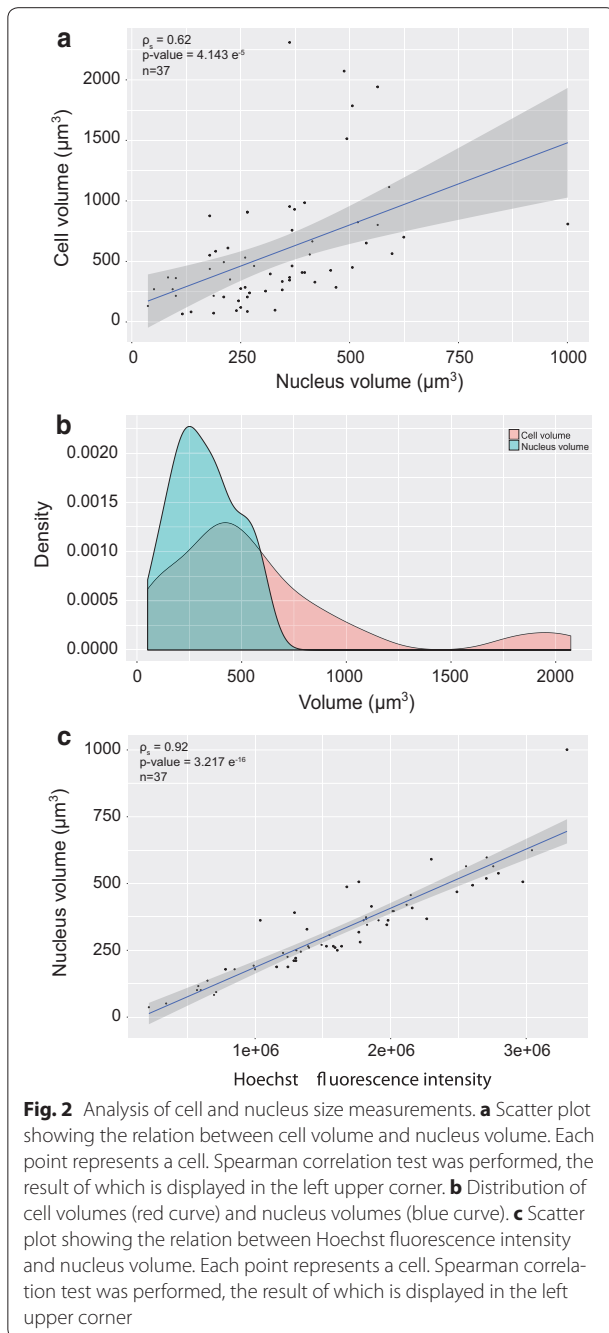
We therefore described a double-staining procedure compatible with microscopy associated at the C1 system to measure, for each cell, their size and cell cycle state independently.

Staining effect

First, we assessed the influence of the double-staining procedure on gene expression at the population level by performing RT-qPCR on 5 selected genes known to be involved in erythroid differentiation or metabolism. The relative value of these gene expressions did not change significantly compared to unstained cells (Fig. 3a). These results suggested that cell and nucleus staining had no major influence on T2EC mean gene expression.

We then needed to discard possible modifications visible only at the individual-cell level. Therefore we performed high-throughput RT-qPCR on single cells using 77 genes that cover various functions as metabolism, differentiation process and proliferation [21]. We compared 30 single stained cells and 47 single unstained cells in the same microchip. Data was analyzed using a PCA-based dimensionality reduction algorithm (Fig. 3b) as well as Wilcoxon signed-rank tests (see Additional file 1: Table S1). The PCA does not show any separation between both conditions (PC1 and PC2 explained less than 12% of the variability), and the statistical analysis shows that no gene was significantly varying between the two





conditions. These results therefore show that the staining did not affect the expression of these 77 genes in T2EC even when examined at the single-cell level.

Finally as an application example for our double-staining approach, we investigated the influence of cell cycle and cell size on cell-to-cell variability among our gene expressions using the coupling of labeling and gene expression measurements at the single-cell level.

Cell morphological impact on T2EC gene expression

For each single cell, we measured the size, the position in the cell cycle and the mRNA amount. Among 69 genes analyzed (retained in this study for technical quality control), none presented a significant spearman correlation between its expression level among single cell volumes or cell cycle: all p-values were above the 5% threshold. These results confirmed that neither cell size nor the position in cell cycle were relevant parameters in explaining the cell-to-cell variations observed for 69 genes examined. This information is important for stochastic single-cell-based gene expression analysis, for which these morphological parameters can be excluded of the potential sources of variability between cells.

Conclusion

We performed a non-cytotoxic CFSE/Hoechst double-staining compatible with the C1 system. This approach allowed automatic identification and measure of morphological parameters. It can be used to measure the influence of cell cycle and cell size on single-cell gene expression analysis without any potential misleading cell state effects induced by cell cycle synchronization methods. It could be also represent an alternative method to avoid artificial cell sorting according to their size or their cell cycle phase, which could be interesting for low amount of cells. This is equivalent to the recently described technique using flow cytometry [36], but applicable in the C1 system. As an alternative, it has recently been described that predefined gene combination could be used a posteriori [37]. Unfortunately, the best combinations seems to be cell type dependent, making it potentially limited [38].

We then used the Biomark system to perform gene expression quantification. We showed that the double staining did not impact gene expression in our cells. Moreover, by measuring the influence of cell cycle and cell size on the expression level of 69 genes, our results support our previous claim that cell cycle and cell size have a negligible influence on gene expression variability in certain settings [21]. This is in line with the recent demonstration that the cell cycle explains only a very small amount (5–17%) of gene expression variability [18, 20].

Limitations

In this study, the main limitation was the optimization of cell capture in the C1 microchip. We obtained a maximum of 65% of capture whereas with other cells, this percent raise up to 95%. Numerous parameters were involved and have to be optimized in order to obtain more individual cells per microchip.

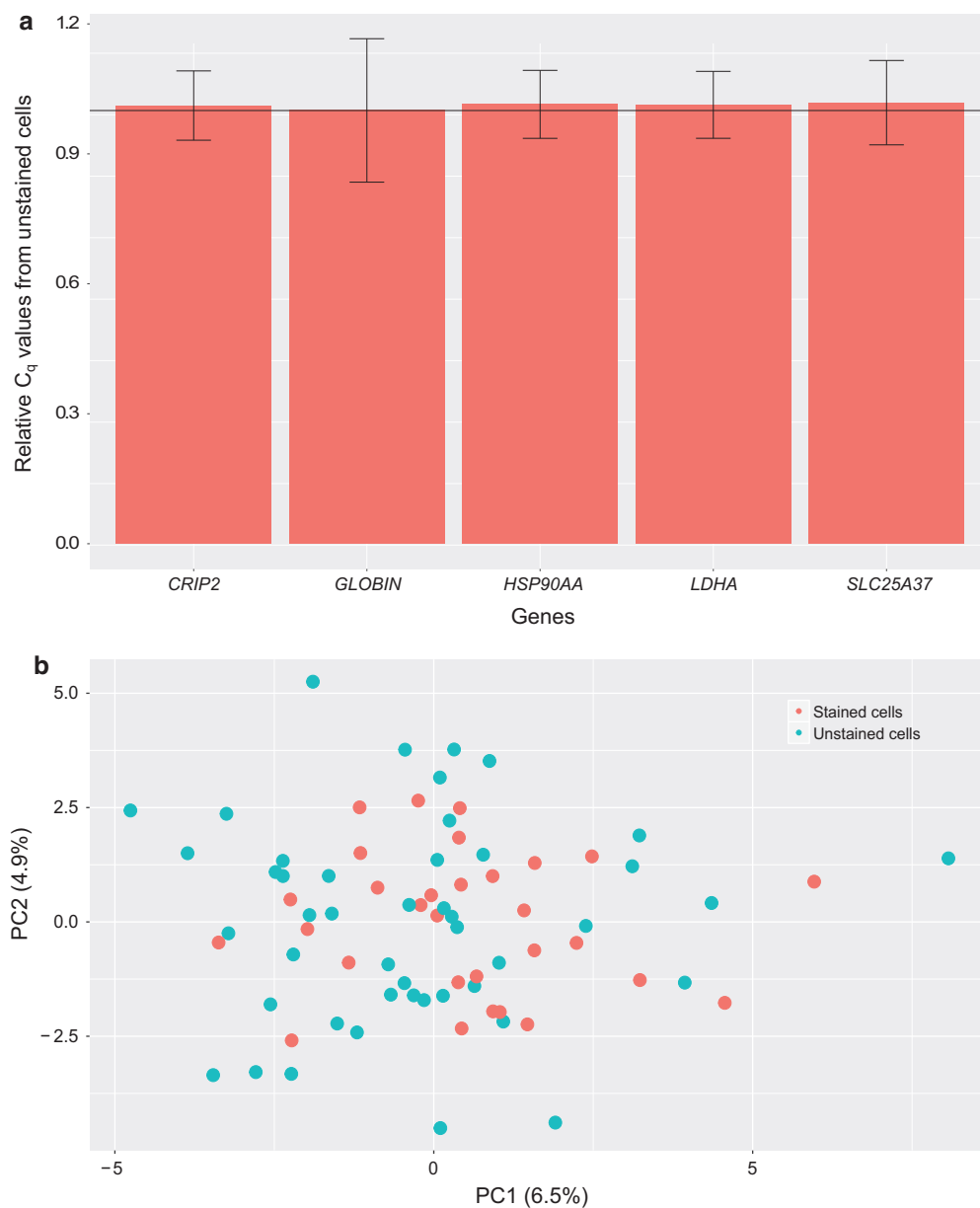


Fig. 3 Analysis of the influence of the staining procedure on gene expression. **a** Real-time PCR gene expression analysis of stained and unstained cells. Total RNA was extracted from T2EC cells stained or not. Reverse transcription and real-time PCR analyses, with specific primers [21], were performed to quantify the amount of GLOBIN (β -GLOBIN), SLC (SLC25A37), HSP (HSP90AA1), CRIP2 and LDHA mRNA (C_q for cycle of quantification). The fold variations represented here correspond to the ratio of mRNA of staining cells compared to unstained cells. The black line corresponds to the null variation between the two conditions. The vertical bars represent the standard error of the mean value ($n = 3$). **b** Principal Component Analysis of single cell expression data acquired on stained or unstained cells. Projection of 77 T2EC single-cell stained or not onto PC1 and PC2 results in a cloud of points without any clear separation. Percentages shown are the percentage of variance explained by each component

Additional file

Additional file 1: Table S1. Statistical analysis of gene expression according to the stained or the unstained condition. Statistical tests (Wilcoxon) were performed for each gene between their expression in stained and unstained condition. A Bonferroni correction was applied in p-values for multiple tests.

Abbreviations

(RT-q)PCR: reverse transcription quantitative polymerase chain reaction; T2EC: TGF- α /TGF- β -induced erythrocytic cells; (c)DNA: complementary deoxyribonucleic acid; (m)RNA: messenger ribonucleic acid; CFSE: carboxyfluorescein diacetate succinimidyl ester; PBS: phosphate-buffered saline; HnRNP: heterogeneous nuclear ribonucleoprotein; GE: gene expression; PCA: Principal Component Analysis; PC1: first principal component; PC2: second principal component.

Authors' contributions

AG designed and performed experiments and drafted the manuscript. AR designed and performed experiments. OG and SGG supervised the study. All the authors reviewed the manuscript. All authors read and approved the final manuscript.

Author details

¹ Laboratoire de biologie et modélisation de la cellule. LBMC-Ecole Normale Supérieure-Lyon, Université Claude Bernard Lyon 1, Institut National de la Santé et de la Recherche Médicale: U1210-Ecole Normale Supérieure de Lyon, Centre National de la Recherche Scientifique: UMR5239, 46 Allée d'Italie, 69007 Lyon, France. ² Inria Dracula, 69100 Villeurbanne, France.

Acknowledgements

We thank ProfileXpert and "Plateforme d'imagerie CIQLE" (Genomique & Microgenomique Unit, Université Lyon 1, SFR Santé LYON-EST, UCBL-INSERM US 7-CNRS UMS 3453) for the use of C1 system and imagery.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The data set regarding the simultaneous measurement of gene expression, cell volume and nucleus volume is available at: <https://osf.io/8x5n7/>.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by funding from the French agency ANR (ICEBERG; ANR-IABI-3096) and La Ligue Contre le Cancer (Comite de Haute Savoie).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 November 2017 Accepted: 23 January 2018

Published online: 01 February 2018

References

1. Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*. 1961;190:372–3.
2. Spudich JL, Koshland DE. Non-genetic individuality: chance in the single cell. *Nature*. 1976;262(5568):467–71.

3. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006;4(10):309.
4. Coulon A, Gandrillon O, Beslon G. On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst Biol*. 2010;4(4):2.
5. Gandrillon O, Kolesnik-Antoine D, Kupiec JJ, Beslon G. Chance at the heart of the cell. *Prog Biophys Mol Biol*. 2012;110:1–4.
6. Junker JP, van Oudenaarden A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell*. 2014;157(1):8–11.
7. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA*. 2002;99(20):12795–800.
8. Elowitz MB, Levine AJ, Siggia ED. Stochastic gene expression in a single cell. *Science*. 2002;297(5584):1183–6.
9. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science*. 2005;309(5743):2010–3.
10. Morelli MJ, Allen RJ, Wolde PR. Effects of macromolecular crowding on genetic networks. *Biophys J*. 2011;101(12):2882–91.
11. Komorowski M, Miekisz J, Stumpf MP. Decomposing noise in biochemical signaling systems highlights the role of protein degradation. *Biophys J*. 2013;104(8):1783–93.
12. Singh A. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem Cell Rep*. 2013;1:532–44.
13. Padovan-Merhar O, Nair GP, Biaisch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A. Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms. *Mol Cell*. 2015;58(2):339–52. <https://doi.org/10.1016/j.molcel.2015.03.005>,p40.
14. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. 2015;33(2):155–60. <https://doi.org/10.1038/nbt.3102>,p40.
15. Kimmerling RJ, Lee Szeto G, Li JW, Genshaft AS, Kazer SW, Payer KR, de Riba Borrajo J, Blainey PC, Irvine DJ, Shalek AK, Manalis SR. A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat Commun*. 2016;7:10220.
16. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006;441(7095):840–6.
17. Loewer A, Lahav G. We are all individuals: causes and consequences of non-genetic heterogeneity in mammalian cells. *Curr Opin Genet Dev*. 2011;21(6):753–8.
18. McDavid A, Dennis L, Danaher P, Finak G, Krouse M, Wang A, Webster P, Beechem J, Gottardo R. Modeling bi-modality improves characterization of cell cycle on gene expression in single cells. *PLoS Comput Biol*. 2014;10(7):1003696.
19. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*. 2008;453(7194):544–7.
20. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161(5):1187–201. <https://doi.org/10.1016/j.cell.2015.04.044>,p40.
21. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, Guillemin A, Papili Gao N, Gunawan R, Cosette J, Arnaud O, Kupiec JJ, Espinasse T, Gonin-Giraud S, Gandrillon O. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol*. 2016;14(12):1002585.
22. Thattai M, van Oudenaarden A. Stochastic gene expression in fluctuating environments. *Genetics*. 2004;167(1):523–30. 0016-6731 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
23. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity. *Cell*. 2005;122(2):169–82.
24. Cagatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Suel GM. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*. 2009;139(3):512–22. <https://doi.org/10.1016/j.cell.2009.07.046>,p39.

25. Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, Giuliani A, Huang S. Cell fate-decision as high-dimensional critical state transition. *BioRxiv*. 2016. <https://doi.org/10.1101/041541>.
26. Sakaue-Sawano A, Miyawaki A. Visualizing spatiotemporal dynamics of multicellular cell-cycle progressions with fucci technology. *Cold Spring Harb Protoc*. 2014;132(3):487–498. <https://doi.org/10.1016/j.ccell.2007.12.033>.
27. Scialdone A, Natarajan KN, Saraiva LR, Proserpio V, Teichmann SA, Stegle O, Marioni JC, Buettner F. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*. 2015;85:54–61. <https://doi.org/10.1016/j.jymeth.2015.06.021.p40>.
28. Vallejos CA, Marioni JC, Richardson S. Basics: Bayesian analysis of single-cell sequencing data. *PLoS Comput Biol*. 2015;11(6):1004333. <https://doi.org/10.1371/journal.pcbi.1004333.p40>.
29. Bruce A, Alexander J, Julian L, Martin R, Keith R, Peter W. *Molecular biology of the cell*. 4th edn. 2002.
30. Gandrillon O, Schmidt U, Beug H, Samarut J. TGF-beta cooperates with TGF-alpha to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *Embo J*. 1999;18(10):2764–81. 0261-4189 Journal Article
31. Grishagin IV. Automatic cell counting with ImageJ. *Anal Biochem*. 2015;473:63–5.
32. Bresson C, Keime C, Faure C, Letrillard Y, Barbado M, Sanfilippo S, Benhra N, Gandrillon O, Gonin-Giraud S. Large-scale analysis by SAGE reveals new mechanisms of v-erbA oncogene action. *BMC Genomics*. 2007;8:390.
33. Team RDC. R: a language and environment for statistical computing. Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>. 2008.
34. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw*. 2007;22:1–20.
35. Johansson L, Sohn D, Thorberg SO, Jackson DM, Kelder D, Larsson LG, Renyi L, Ross SB, Wallsten C, Eriksson H, Hu PS, Jerning E, Mohell N, Westlind-Danielsson A. The pharmacological characterization of a novel selective 5-hydroxytryptamine1a receptor antagonist, NAD-299. *J Pharmacol Exp Ther*. 1997;283(1):216–25.
36. Blasi T, Hennig H, Summers HD, Theis FJ, Cerveira J, Patterson JO, Davies D, Filby A, Carpenter AE, Rees P. Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nat Commun*. 2016;7:10256.
37. Liu Z, Lou H, Xie K, Wang H, Chen N, Aparicio OM, Zhang MQ, Jiang R, Chen T. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat Commun*. 2017;8(1):22.
38. Dolatabadi S, Candia J, Akrap N, Vannas C, Tesan Tomic T, Losert W, Landberg G, Aman P, Stahlberg A. Cell cycle and cell size dependent gene expression reveals distinct subpopulations at single-cell level. *Front Genet*. 2017;8:1.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



2.3 Article 3 : Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment

Notre équipe collabore depuis plusieurs années avec l'équipe d'Andr s Paldi, avec laquelle nous avons r guli rement des discussions scientifiques   propos de nos travaux respectifs. Les recherches r alis es dans nos deux  quipes, partagent en effet un int r t commun pour la biologie des syst mes et l' tude des processus de prise de d cision cellulaire, notamment dans le cadre de la diff renciation. Ayant acquis de l'expertise sur la technique de RTqPCR en cellule unique au cours de mon Master 2 et de ma th se, j'ai donc  t  en mesure de former Alice Moussy, doctorante dans l' quipe d'Andr s Paldi,   cette technologie. Elle a ainsi pu g n rer dans le courant de sa th se des donn es d'expression de g nes en cellule unique sur une cin tique de diff renciation multilignage de cellules souches h matopo i tiques. Ses travaux ont  t  r cemment publi s dans un article dont je suis co-auteur, paru dans le journal PLoS Biology (article3) [183]. Les r sultats de cette publication sont en accord avec mes travaux de th se, et apportent des  l ments suppl mentaires en faveur d'une vision stochastique des processus de diff renciation.

2.3.1 Principaux r sultats de l'article 3

Cet article a pour objectif d'analyser la dynamique de l'engagement en diff renciation des cellules dans un choix multilignage, par des approches transcriptionnelles, et morphologiques par time-lapse. Cette  tude a  t  men e dans des cellules h matopo i tiques CD34+ issues de sang de cordon pouvant  tre induites en diff renciation par stimulation *via* des cytokines.

L'expression de 90 g nes a tout d'abord  t  analys e par RTqPCR dans des cellules uniques CD34+ isol es   0h (non stimul es), ainsi qu'  24h, 48h et 72h apr s stimulation. L'analyse de ces donn es par l'outil de r duction de dimension t-SNE [184] montre que la stimulation des CD34+ induit dans un premier temps une r ponse d coordonn e, caract ris e par une population cellulaire tr s h t rog ne (figure 1B de l'article 3) et un faible nombre de corr lations entre l'expression des g nes  tudi s (figure 1D de l'article 3). Cet  tat tr s variable est suivi d s 48h de l'apparition dans la population, de deux patterns d'expression

transcriptionnels différents. Ces profils correspondent pour l'un à un profil similaire aux progéniteurs hématopoïétiques myéloïdes (CMP-like), et pour l'autre à un profil proche de celui des cellules CD34+. La distinction entre ces deux patterns d'expression est d'autant plus renforcée à 72h, où plus de la moitié des cellules présentent un profil CMP-like. Ces résultats suggèrent ainsi qu'une prise de décision a lieu dès les premières divisions du processus de différenciation, marquant à 24h un état de transition dit "multiprimed" dans lequel des gènes associés à différents lignages sont exprimés simultanément.

Des cellules individuelles CD34+ ont ensuite été suivies par time-lapse afin d'apporter une dimension dynamique à la cinétique d'expression génique n'offrant que des aperçus temporels figés du processus de différenciation. Une analyse des comportements morphologiques cellulaires au cours de la différenciation, montre premièrement que le délai de la première division est plus long, que les divisions suivantes. Deuxièmement, les cellules semblent adopter deux morphologies, une forme ronde et une forme polarisée, entre lesquelles elles oscillent. Les cellules CD34+ non stimulées fluctuent très rarement entre ces deux morphologies, elles présentent plutôt une forme polarisée (vidéos S1-S3 de l'article 3). Le phénomène de fluctuation entre les deux formes est observé bien plus fréquemment dès la deuxième ou la troisième division du processus de différenciation. Une partie de la population garde ensuite ce phénotype oscillatoire, et la majorité adopte progressivement la morphologie ronde. Ces résultats suggèrent ainsi que l'état fluctuant des cellules représente un phénotype intermédiaire au comportement "hésitant", entre la forme polarisée adoptée davantage par les cellules multipotentes et la forme ronde des progéniteurs myéloïdes.

Au cours de la différenciation des cellules CD34+, l'analyse d'expression génique en cellules uniques a permis d'identifier deux groupes distincts aux profils différents, alors que l'analyse morphologique en time-lapse a révélé trois différents phénotypes dynamiques. Les cellules présentant un pattern d'expression génique semblable aux cellules CD34+, expriment le gène CD133, alors que les cellules au profil plutôt CMP-like ne l'expriment pas ou très faiblement. L'expression de CD133 a donc été utilisée pour isoler trois groupes de cellules, exprimant fortement (CD133high), moyennement (CD133medium) ou faiblement (CD133low) le gène CD133 après 48h de différenciation. L'analyse morphologique en time-lapse de chacun de ces compartiments cellulaires a alors démontré que les cellules CD133high et CD133low

présentent majoritairement des phénotypes respectivement, polarisés et ronds, alors que la plupart des cellules CD133 medium fluctuent fréquemment entre les deux morphologies, qui apparaissent de façon équilibrée (figure 4B de l'article 3). D'autre part, l'expression des gènes sélectionnés pour les précédentes analyses, ont été analysés à l'échelle de la cellule unique dans ces trois mêmes compartiments cellulaires. Les cellules CD133medium présentent un profil d'expression tout à fait différent des cellules CD133high et CD133low, s'approchant respectivement de celui des cellules CD34+ et des cellules CMP-like (figure 5A de l'article 3). Ainsi l'analyse morphologique suggèrent que les cellules rondes dérivent des cellules polarisées, après avoir traversé une phase d'"hésitation", caractérisée par une fluctuation fréquente entre le deux morphologies. D'autres part, l'analyse d'expression génique a permis de mettre en évidence le lien direct entre le profil transcriptionnel et la morphologie cellulaire. Enfin le phénotype oscillatoire des cellules CD133medium est associé à un profil d'expression génique qui lui est propre, et qui est différent des cellules polarisées et des cellules rondes. Cette observation suggère, d'une part, que cet état intermédiaire n'est pas qu'une simple transition d'un état cellulaire vers un autre, et confirme, d'autre part, le lien entre expression génique et morphologie.

2.3.2 Principales conclusions de l'article 3

La combinaison d'analyses de microscopie en time-lapse et d'expression génique à l'échelle de la cellule unique, a mis en évidence le caractère dynamique et variable de la prise de décision cellulaire des cellules hématopoïétiques, en terme d'expression génique mais également de morphologie cellulaire. De plus, l'ensemble de ces résultats a permis d'identifier une nouvelle catégorie de cellules caractérisée par un phénotype oscillatoire. Un scénario dynamique a alors été proposé dans lequel l'état stochastique dit "multilignage primed" est suivi d'une phase d'hésitation de durée variable correspondant à des transcriptomes fluctuants, avant d'atteindre un nouvel état stable dans lequel les cellules sont engagées dans un lignage spécifique.

2.3.3 Article 3

Publié le 27 Juillet 2017 dans le journal PLoS Biology

URL : <https://doi.org/10.1371/journal.pbio.2001867>

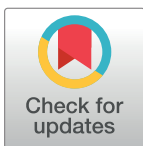
RESEARCH ARTICLE

Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment

Alice Moussy^{1,2}, Jérémie Cosette², Romuald Parmentier², Cindy da Silva¹, Guillaume Corre², Angélique Richard³, Olivier Gandrillon³, Daniel Stockholm¹, Andrés Páldi^{1*}

1 Ecole Pratique des Hautes Etudes, PSL Research University, UMRS 951, INSERM, Univ-Evry, Evry, France, **2** Genethon, Evry, France, **3** Laboratoire de Biologie et de Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, Université de Lyon, Lyon, France

* andras.paldi@ephe.sorbonne.fr



 OPEN ACCESS

Citation: Moussy A, Cosette J, Parmentier R, da Silva C, Corre G, Richard A, et al. (2017) Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLoS Biol* 15 (7): e2001867. <https://doi.org/10.1371/journal.pbio.2001867>

Academic Editor: Sui Huang, Institute for Systems Biology, United States of America

Received: December 21, 2016

Accepted: June 23, 2017

Published: July 27, 2017

Copyright: © 2017 Moussy et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The quantitative RT-PCR, time-lapse and cytometry data presented in the paper and used to generate the graphical representations are available in the "Supporting information" as S1-S3 Data files respectively. Raw Q-RCR data and original time-lapse files are available on request.

Funding: Ecole Pratique des Hautes Etudes www.ephe.fr (grant number 11REC/BIMO). Received by AP. The funder had no role in study design, data

Abstract

Individual cells take lineage commitment decisions in a way that is not necessarily uniform. We address this issue by characterising transcriptional changes in cord blood-derived CD34+ cells at the single-cell level and integrating data with cell division history and morphological changes determined by time-lapse microscopy. We show that major transcriptional changes leading to a multilineage-primed gene expression state occur very rapidly during the first cell cycle. One of the 2 stable lineage-primed patterns emerges gradually in each cell with variable timing. Some cells reach a stable morphology and molecular phenotype by the end of the first cell cycle and transmit it clonally. Others fluctuate between the 2 phenotypes over several cell cycles. Our analysis highlights the dynamic nature and variable timing of cell fate commitment in hematopoietic cells, links the gene expression pattern to cell morphology, and identifies a new category of cells with fluctuating phenotypic characteristics, demonstrating the complexity of the fate decision process (which is different from a simple binary switch between 2 options, as it is usually envisioned).

Author summary

Hematopoietic stem cells are classically defined as a specific category of cells at the top of the hierarchy that can differentiate all blood cell types following step-by-step the instructions of a deterministic program. We have analysed this process, and our findings support a much more dynamic view than previously described. We apply time-lapse microscopy coupled to single-cell molecular analyses in human hematopoietic stem cells and find that fate decision is not a unique, programmed event but a process of spontaneous variation and selective stabilisation reminiscent of trial-error processes. We show that each cell explores (at its own pace and independently of cell division) many different possibilities before reaching a stable combination of genes to be expressed. Our results suggest,

collection and analysis, decision to publish, or preparation of the manuscript. Agence Nationale de la Recherche www.anr.fr (grant number BSV6 014 02 « Stochagene »). Received by AP and OG. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Genethon www.genethon.fr. Received by AP and DS. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: CLP, common lymphoid progenitor; CMP, common myeloid progenitor; CTV, CellTrace Violet; HDAC, histone deacetylase; HSPC, hematopoietic stem and progenitor cell; MPP, multipotent progenitor; PCA, principal component analysis; qRT-PCR, quantitative reverse transcription polymerase chain reaction; t-SNE, t-distributed stochastic neighbour embedding; VPA, valproic acid.

therefore, that multipotency seems to be more like a transitory state than a feature of a specific cell category.

Introduction

Hematopoietic stem and progenitor cells (HSPCs) give rise to all the cellular components of blood. The major stages of differentiation and the key genes participating in this process are now well characterised [1]. According to the classical view, haematopoiesis is a hierarchically organised process of successive fate commitments, in which differentiation potential is progressively restricted in an orderly way over cell divisions. There are several variants of the model [2–6]. In all cases, the first fate decision is a binary choice taken by multipotent progenitors (MPPs), which leads to 2 different committed progenitors (for the purpose of simplicity, these progenitors are designated here as common myeloid progenitors [CMP] and common lymphoid progenitors [CLP]). In molecular terms, the choice is believed to be the result of the strictly regulated activation of master regulator genes and their underlying transcriptional network [7]. However, the strict hierarchical logic of classical models has recently been challenged by a number of *in vivo* and *in vitro* studies [8–10]. Single-cell gene expression studies have revealed a much higher heterogeneity of cell subtypes than can be detected using a combination of surface markers [11]. It is not surprising that the number of identifiable cell types increases with the resolution of the detection method. Although correct cell type classification is a key step in understanding the cell fate decision issue, it cannot reveal the dynamic features of the fate commitment process and leaves a number of unanswered questions. Do different phenotypic forms represent different cell types or different stages of the same process? How does the transition between the forms occur? How long does it take?

Until recently, fully deterministic explanations were predominant, but recent studies have suggested other alternatives. Two different possibilities have been put forward. According to the first, the commitment process starts with the sporadic, independent activation of genes within the same cell. The simultaneous stochastic expression of regulatory genes specifying different lineages creates a multiprimed intermediate state that enables these cells to choose 1 of the lineages [12–16]. A coherent lineage-specific expression profile would then emerge from this metastable state. According to the second, commitment is preceded by transcriptome fluctuations between different lineage-biased states [17–19]. Surprisingly, the time scale of transformations related to the cellular fate decision process remains largely unexplored. The transcriptome of the same cell can be analysed only once, because the cell is destroyed by RNA extraction. Therefore, indirect approaches are required to identify trends and patterns in time series.

We addressed the issue of the dynamics and the time scale of the commitment process by integrating single-cell quantitative reverse transcription polymerase chain reaction (qRT-PCR), cell division history, and morphological changes determined by time-lapse analysis. Contrary to the common strategy consisting of isolating defined cell subpopulations with the help of specific surface markers and characterising their gene expression profiles at the single-cell level [20], we used an alternative approach. Individual cells were randomly isolated from the heterogeneous cord blood CD34+ cell fraction at different time points after cytokine stimulation, and their gene expression profiles were determined using single-cell qRT-PCR. The data provided a series of snapshots, showing the actual statistical distribution of single-cell gene expression patterns across the whole population. The structure of the population at the successive time points was revealed by unsupervised classification of the expression profiles

according to their similarity using multiparametric approaches. The progression of the fate commitment process was deduced from the evolution of the population structure. At the same time, using time-lapse microscopy, we tracked randomly isolated individual CD34⁺ cells and their progeny for several days after cytokine stimulation. We recorded the division history and the morphological changes of each cell in the clones. The population structure was deduced on the basis of the statistical analysis of these observations. The efficiency of the time-lapse approach in investigating cell fate decisions has recently been shown [21]. To reinforce this approach, the time-lapse and gene expression data were integrated into a coherent scenario. This was done by using CD133 protein expression levels to isolate cells with 1 transcription profile or the other and directly record their dynamic phenotype, thereby providing a direct link between dynamic phenotype and transcription profile.

Altogether, our results revealed that fate decision is a dynamic, fluctuating process that is more complex than a simple binary switch between 2 options, as it is usually envisioned.

Results

Single-cell gene expression

The transcriptional profile of individual cord blood CD34⁺ cells was determined at 0, 24, 48, and 72 h after the beginning of cytokine stimulation (Fig 1A). Single-cell qRT-PCR was used to quantify the mRNA levels of 90 different genes. A set of 32 genes was selected for their known function in the early differentiation of hematopoietic cells and was expected to inform on the functional state of the cells (see S1 Table). A second set of 54 genes was chosen randomly from a list of genes known to be expressed in the hematopoietic lineage [22,23]. These genes provided an assessment of the overall transcriptional activity of the genome. Five additional genes were added to the list for their role in maintaining the pluripotent state in embryonic stem cells. A heat map of all data and a violin plot of the expression profile of each gene at the 4 separate time points are shown in S1 Fig and S2 Fig. The normalized single-cell quantitative gene expression data obtained for the different time points were merged into a single database and screened for subpopulations by k-means clustering. The number of statistically distinguishable groups was inferred using gap statistics [24]. The groups were visualised on heat maps and on a 2D plot using t-distributed stochastic neighbour embedding (t-SNE) [25]. Although every cell had a unique gene expression pattern, this approach enabled us to clearly identify subgroups of cells in the population on the basis of the statistical similarity of their gene expression patterns (Fig 1B).

Nonstimulated CD34⁺ cells isolated from cord blood represented the $t = 0$ h time point. A heat map of the single-cell transcriptional profiles of genes contributing significantly to the identification of subgroups (S1 Fig) showed that this population of cells was heterogeneous. Several genes reported to play a role in self-renewal, quiescence, and other stem cell functions (CD71, CD133, CXCR4, GATA2, and FLT3) were expressed sporadically and at variable levels in a fraction of cells. Genuine pluripotent stem cell genes were also expressed at low levels in a fraction of cells (NANOG, OCT4, KLF4). Nevertheless, no correlation was found between these genes (Fig 1D), and the statistical analysis did not reveal distinguishable expression patterns that could define cell types. The only detectable differences were donor-associated and probably reflected differences related to the processing of individual blood samples. Donor-specific differences disappeared at later stages.

The gene expression profile 24 h after the onset of cytokine stimulation was found to be fundamentally different to $t = 0$ h cells. Almost every cell responded to cytokine stimulation by increasing transcript levels and generating a unique gene expression pattern (Fig 1C). When represented on the 2D t-SNE (Fig 1B) and principal component analysis (PCA) plots (S3 Fig),

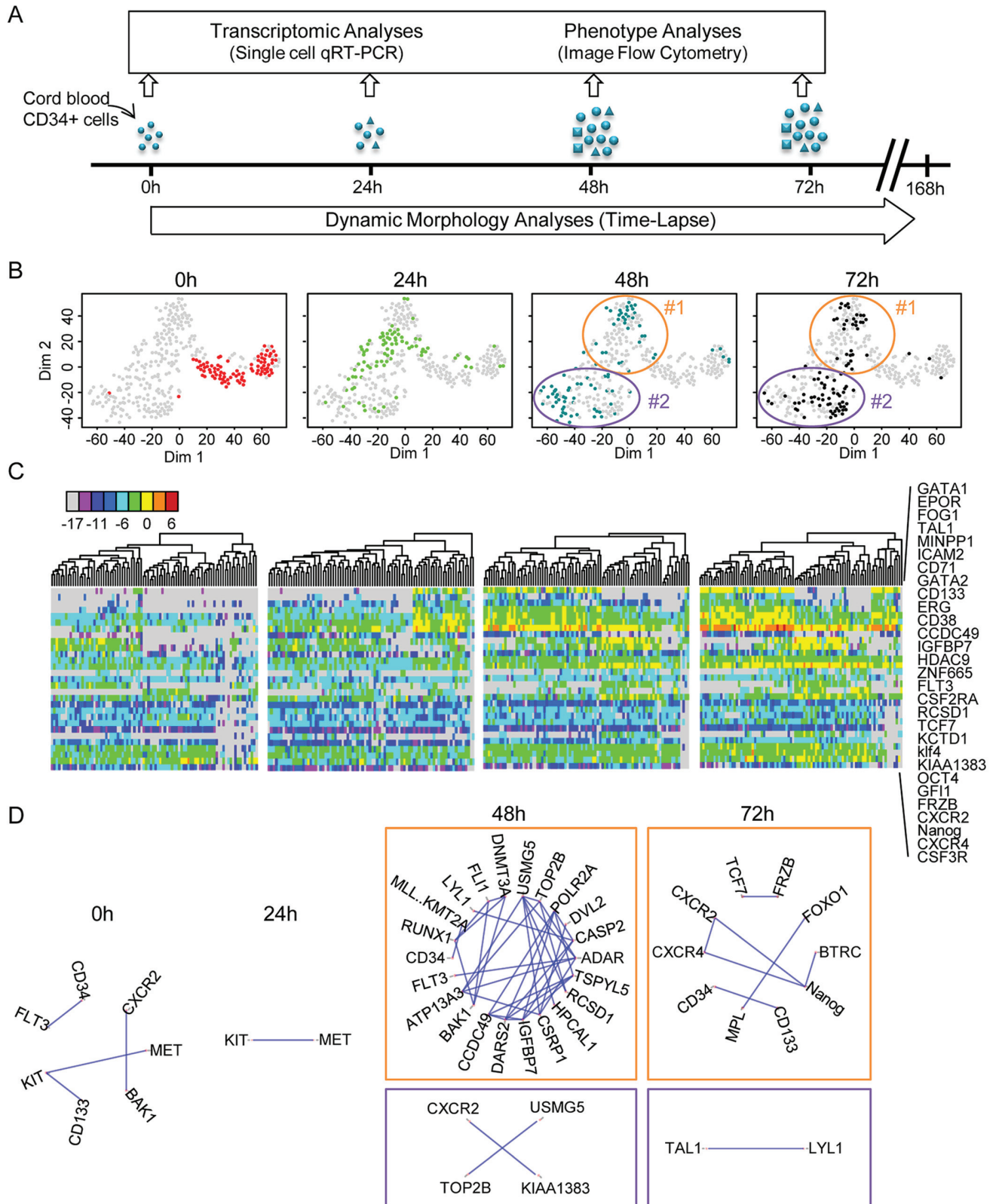


Fig 1. Transcriptional profile of cord blood-derived CD34+ cells at t = 0 h, t = 24 h, t = 48 h, and t = 72 h after the beginning of the experiment. (A) CD34+ cells were isolated from human cord blood and cultured in serum-free medium with early acting cytokines. Single-cell quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to analyse single-cell transcription at 0 h, 24 h, 48 h and 72 h. At the same time, individual clones were continuously monitored using time-lapse microscopy. (B) t-distributed stochastic neighbour embedding (t-SNE) map of single-cell transcription data. The 4 panels show analysis of the same data set, with each point representing a single cell highlighted in different colours depending on the given time point. The 2 clusters identified by gap statistics at t = 48 h and t = 72 h are surrounded by an ellipse and numbered #1 and #2 for multipotent and common myeloid progenitor (CMP)-like cells. Note the rapid evolution of the expression profiles. (Underlying data can be found in [S1 Data](#).) (C) A heat map representation of the expression levels of a subset of genes that strongly contributed to the differentiation of the different groups (as detected by principal component analysis [PCA]; see [S2 Fig](#)) and cluster analysis of expression profiles at the different time points show the rapid evolution of gene expression. The list of the genes used (shown on the right) includes well-known genes acting during hematopoietic differentiation but also many randomly selected genes. The colour code for expression levels is indicated below. (Underlying data can be found in [S1 Data](#).) (D) Pairwise correlations between the genes analysed using single-cell quantitative reverse transcription polymerase chain reaction (qRT-PCR). Only the gene pairs with a Pearson correlation coefficient higher than 0.8 are indicated for each time point. The 2 clusters identified at t = 48 h and t = 72 h are represented separately. Note the transient increase of the average correlation in cluster #2 at t = 48 h, indicating a state transition. (Underlying data can be found in [S1 Data](#).)

<https://doi.org/10.1371/journal.pbio.2001867.g001>

the cells formed a single but dispersed cluster, well separated from the t = 0 h cells. In a fraction of cells, moderate to high transcription of previously nonexpressed hematopoietic regulator genes was observed in addition to that already seen at t = 0 h. For example, the expression of GATA1, GATA2, PU1, CD71, FOG1, CD133, or EPOR increased or was more frequent than at t = 0 h. In some cells, all these genes were expressed simultaneously. Nevertheless, no distinct subpopulations could be identified at the resolution of our approach. The pairwise correlation coefficients between genes remained low ([Fig 1D](#)). It is therefore likely that the patterns observed at 24 h resulted from essentially uncoordinated up-regulation of gene transcription and led to a highly heterogeneous cell population. This is a transition state reminiscent of the reported multilineage primed state with simultaneous expression of lineage-affiliated genes specifying alternative cell fates [[12,15](#)].

The first signs of coordinated differential gene expression appeared at t = 48 h after cytokine stimulation. At this stage, 2 distinct gene expression patterns emerged from the highly variable background of earlier stages. The 2 clusters are clearly distinguishable on the t-SNE plot ([Fig 1B](#)) and identified by gap statistics. They are also easily seen on the heat map representing gene expression levels ([Fig 1C](#)). Cluster #2 comprised cells with simultaneous expression of genes characteristic of CMPs such as GATA1 and EPOR [[7](#)]. The expression profile of the cells in cluster #1 was characterised by the strong expression of genes reported for multipotent cells (CD133, GFI1, KLF4, or FLT3) and the lack of expression of GATA1 and EPOR. Although this pattern is reminiscent of a self-renewing, multipotent profile, it is difficult to determine the exact identity of these cells at the level of resolution used in our study [[26](#)]. Typical genes for pluripotent stem cells like NANOG and OCT4 were expressed at moderate levels in many cells from both clusters ([Fig 1C](#), [S1 Fig](#) and [S2 Fig](#)). Randomly selected genes were also good predictors for the 2 groups of cells. Only a small fraction of cells could not be classified in 1 of the 2 main clusters at t = 48 h ([Fig 1B](#)). The tendency observed at t = 48 h was further reinforced by t = 72 h. The cells in cluster #2 with CMP-like profiles represented more than half of all cells ([Fig 1B and 1C](#)). We observed a strong but transient increase in the number of highly correlated genes in this group ([Fig 1D](#)). Such an increase in the overall gene-to-gene correlation is a typical hallmark of imminent state transition in these cells [[27–29](#)]. Indirectly, this suggested that the cluster #1 profile was more in continuity with the previous profile observed at t = 24 h and that the cluster #2 profile at t = 48 h represented a transition to a new pattern.

Taken together, these single-cell gene expression observations revealed that the cell fate decision process in cytokine-stimulated CD34+ cord blood cells occurred during the first 2 d. Initially, each cell responds to cytokine stimulation with an uncoordinated change in gene expression, which is followed by the emergence of 2 distinct gene expression patterns reminiscent of the 2 known major types of hematopoietic progenitor cells. Although indications of

this second change may appear as early as 24 h after stimulation, the 2 distinct gene expression patterns are clearly distinguishable at 48 h and consolidated by 72 h. By this stage, almost every cell seems to have adopted 1 profile or the other.

Time-lapse tracking studies

In order to integrate the gene expression snapshots into a dynamic scenario, we made time-lapse records of individual CD34⁺ cells under *in vitro* conditions identical to those in the single-cell gene expression studies. We imaged individual cells in microwells at a rate of 60 frames per hour for 7 d (Fig 2A). Using a semiautomatic image analysis approach, we established individual clonal pedigrees and recorded cell cycle durations and major morphological changes. As shown in Fig 2B and 2C, the pedigrees of individual clones were highly variable but shared some general features. Some clones produced only a few cells during the observation period, while others proliferated faster and produced up to 30–40 siblings. We focused our attention on the first 3 generations. As reported for cells cultured in early acting cytokines [30], the first cell cycle was exceptionally long in all clones. The division of the founder cell occurred between 35 h and 80 h after the start of culturing, with the median cell cycle length being 58 h (Fig 2). We questioned whether the culturing of isolated cells in microwells, in which direct contact with the other cells was not possible, influenced cell cycle length. To measure the division rate in a population context, the cells cultured together were labelled using CellTrace Violet (CTV). The results (S4 Fig) showed that the cells had similar division profiles regardless of whether they were cultured individually or in population. The unusually long first cell cycle was particularly important when interpreting results. It implied that the transition from the initial to the multilineage primed transcription profile followed by 1 of the 2 types of progenitor-like profiles observed at 24 h and 48 h after CD34⁺ cell stimulation occurred during the life of the founder cell, before the first mitosis.

Previous studies have demonstrated that there is a connection between cell morphology and the differentiation potential of CD34⁺ cells. Two major morphological forms have been described in the CD34⁺ cord blood cell fraction. Polarised cells are capable of active motion with the help of lamellipodia and possess, on their opposite end, large protrusions called uropods. These cells have been found to retain primitive self-renewing and stem cell functions [31,32]. The second morphological type is round. These cells have been considered as already engaged in differentiation [31,32].

Time-lapse records revealed that the 2 cell morphologies were not permanent; most cells were able to switch between forms several times during the cell cycle. After recovering from the stress of isolation and manipulation, founder cells acquired polarised morphologies within a few hours, developing uropods and starting to move actively (see S1–S3 Movies). During the first cell cycle, cells mostly conserved the polarised form, and switches between the 2 morphologies were rare. As indicated above, the first cell division occurred (on average) at 58 h, and the average lengths of subsequent cell cycles were around 20 h to 22 h. The daughter (second generation) and granddaughter (third generation) cells were able to switch between the 2 morphologies at a much higher frequency compared to the founder cells. In order to quantify these events, we manually tracked each cell and recorded each switch. Representative profiles are presented in S5 Fig.

In order to compare quantitatively the dynamic phenotype of cells, we calculated 3 parameters based on their dynamic profiles. The first parameter was calculated as a ratio of the time a given cell spent in a round shape compared to the time spent in a polarised shape. This parameter was close to 0 for stable polarised cells and 1 for stable round cells. Intermediate values correspond to the fraction of time cells spent in round shape. The second parameter was the

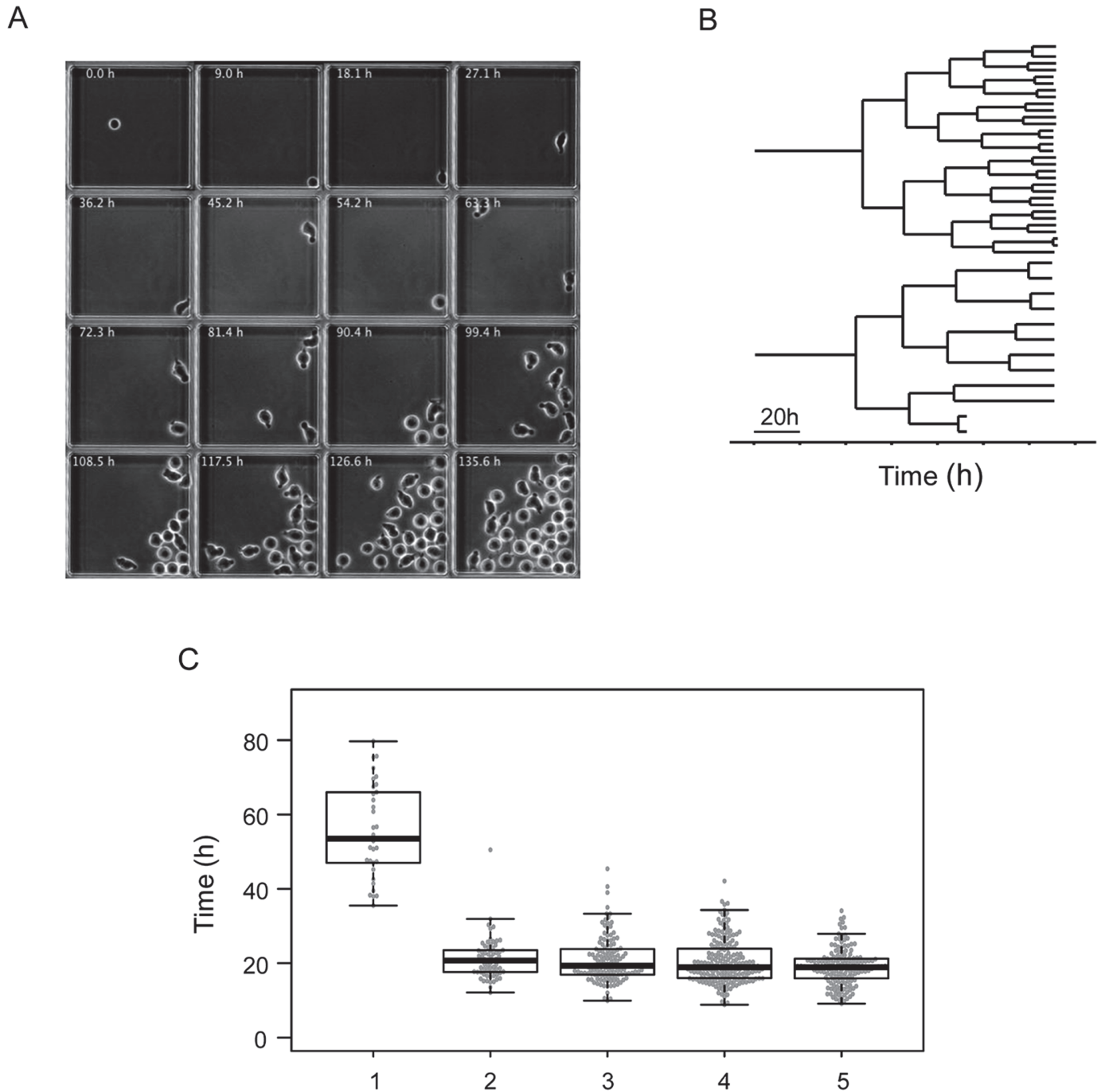


Fig 2. Time-lapse tracking of individual clones. (A) These frames, extracted from a representative time-lapse record, show a microwell containing a single founder cell, which divides to produce a clone. Each individual cell was tracked, and their morphological characteristics were recorded. (B) Two representative lineage pedigrees obtained from the time-lapse record. The strong difference in clone size observed at the end of the record is established gradually after the third cell division. (C) Box plot representation of cell cycle lengths obtained from the time-lapse records of every clone. Note the long first cell cycle. Subsequent cell cycles have comparable lengths, with a slight tendency to become shorter. (Underlying data can be found in [S2 Data](#).)

<https://doi.org/10.1371/journal.pbio.2001867.g002>

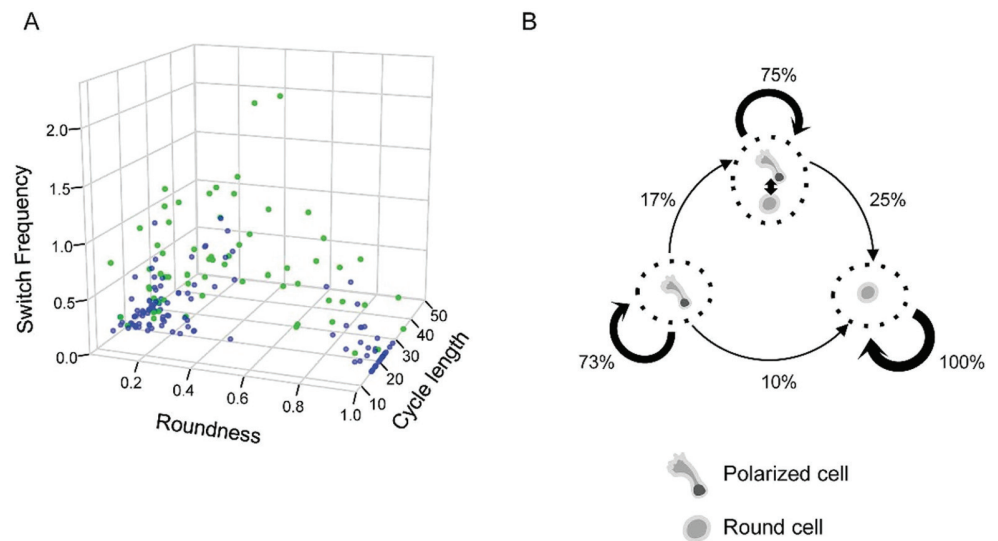


Fig 3. Quantitative analysis of dynamic phenotypes as determined by time-lapse data. (A) Association between the morphology, switch frequency, cell cycle length, and the type of cell divisions of second- and third-generation cells. Each point represents a single cell. Siblings with different dynamic behaviour and morphology (in green) are usually characterised by high switch frequencies. Siblings with similar dynamic behaviour and morphologies are shown in blue. The morphology is given as a ratio of time spent in round/polarised shape by a cell during the cell cycle. Switch frequency is given in number of morphological transformations per hour. Cell cycle length is in hours. (B) Dynamic phenotype change during the first 2 cell divisions as determined on the basis of time-lapse records. Three different dynamic phenotypes were identified: stable polarised, frequent switchers, and stable round. Cells tended to transmit dynamic phenotypes to daughter cells during cell division. Polarised and frequent switchers produced round cells, and frequent switchers were always produced by polarised mothers. Phenotypic change is not associated with asymmetric division; it can occur at any time in the cell cycle. Since round cells always produce round daughters, the whole process is biased and the proportion of this phenotype increases. (Underlying data can be found in [S2 Data](#).)

<https://doi.org/10.1371/journal.pbio.2001867.g003>

frequency of morphological switches during the cell cycle. This parameter expressed the cell's ability to maintain a stable morphology. The third parameter was the cell cycle length. When cells were represented as individual points in the space determined by the 3 parameters, we identified 3 major categories ([Fig 3A](#)). The first category included cells with mainly polarised shapes; the second category was composed of cells with predominantly round shapes; the cells in the third category switched shape frequently, generally fluctuating between both morphologies ([Fig 3A](#)).

When sister cell pairs were examined, it became obvious that many displayed very similar dynamic phenotypes. In some cases, periods of stable morphology and switching events coincided almost perfectly ([S5 Fig](#)). In other cases, the 2 sister cells behaved differently. In the most extreme cases, 1 sister cell adopted a stable round form and the other a stable polarised form immediately after division.

We calculated the frequency with which a cell with a given dynamic phenotype was produced by a mother cell with similar or dissimilar phenotype ([Fig 3B](#)). Maternal cells clearly tended to transmit the dynamic phenotype to daughter cells. We also observed the regularity with which phenotype changes occurred in daughter cells. Polarised cells were systematically produced by polarised cells. At lower probabilities, both polarised and fluctuating cells could produce stable round phenotype cells. Round cells always gave rise to round siblings ([Fig 3B](#)). Following these simple rules, the cumulative outcome of the process was the gradual increase of round cells in the population. Cells with fluctuating morphologies appeared to be an intermediate form between polarised and round cells. Since 25% of daughter cells conserved this

phenotype, the fluctuating intermediate cells persisted in the population. On static snapshots, however, this category remained undetectable: only polarised and round cells were observed. A polarised form was considered to be a feature of multipotent cells and the round form a committed myeloid progenitor phenotype [31,32].

Coupling the molecular and cellular scales

The dynamically fluctuating behaviour we have described here for the first time represents a transition between the 2 states and reflects a 'hesitant', incomplete fate-determination process. Since we detected only 2 major transcription profiles but observed 3 different dynamic behaviours, it is possible that 'hesitant' cells are not characterised by a clearly distinct transcription pattern. Morphology fluctuations may be accompanied by fluctuations in the transcript or protein levels of at least some key genes.

To test this assumption, we took advantage of the observation that the gene coding for the CD133 cell surface protein was expressed preferentially in 1 of the 2 transcription patterns detected at 48 h (Fig 1B and 1C). Previous reports have established that CD133 protein is typically present in cells with polarised forms and accumulates in the uropod [31–33]. We confirmed this using image cytometry and immunohistochemistry on fixed cells (S6 Fig). Cells expressing high levels of CD133 were mostly polarised, while those with low levels of CD133 were round (S6 Fig). This observation explicitly established a direct link between the cell morphology and the transcription patterns detected by single-cell qRT-PCR.

We therefore used the CD133 protein as a proxy for the isolation of a cell fraction enriched in either polarised or round cells and recorded their dynamic phenotype. The 'high', 'medium', and 'low/negative' CD133-expressing cell fractions were isolated 48 h after cytokine stimulation, put in culture, and tracked using time-lapse microscopy for an additional 48 h (Fig 4A). The fraction of the time the cell spent in round or in polarised shape, the switch frequency, and the division asymmetry of the tracked cells were quantified. The 'high' CD133-expressing cells and their progeny reproduced the 3 types of cells observed previously but in different proportions (Fig 4B). Most of the cells displayed stable polarised morphologies or were frequent switchers; only a few cells displayed stable round morphologies (Fig 4B). By contrast, the 'low/negative' cells produced either stable round progeny or cells with fluctuating morphologies (Fig 4B). The 'medium' CD133-expressing cells had a higher switch frequency, and both shapes were represented in a more equilibrated manner (Fig 4B).

These observations confirmed the idea that cells with stable round shapes were derived from cells with polarised shapes and high CD133 levels following a period in which they had a fluctuating phenotype. The fluctuations occurred in a bi-stable manner; the cells switched from 1 morphology to another and back rapidly without stable intermediate states (see S1–S3 Movies). The process of transformation did not correlate with the cell cycle; some cells reached a stable morphology rapidly, while others fluctuated over several cycles. The process was accompanied by a gradual decrease in CD133 protein levels in cells. We found no evidence that asymmetric divisions played a direct role in this process.

Next, we isolated individual cells with high, medium, and low/negative CD133 protein levels using a cell sorter and performed single-cell qRT-PCR analysis using the same gene panel used previously on the unsorted population. Fig 5A shows the t-SNE representation of the single-cell gene expression profiles. The heat map representation of the full set of gene expression results is shown in S7 Fig together with the PCA analysis. The 'high' and 'low' CD133 cells displayed different transcription profiles similar to clusters #1 and #2, respectively, found in cells of the unsorted population at $t = 48$ h (Fig 1). The cell fraction isolated on the basis of intermediate CD133 levels contained a large number of cells with intermediate transcription profiles,

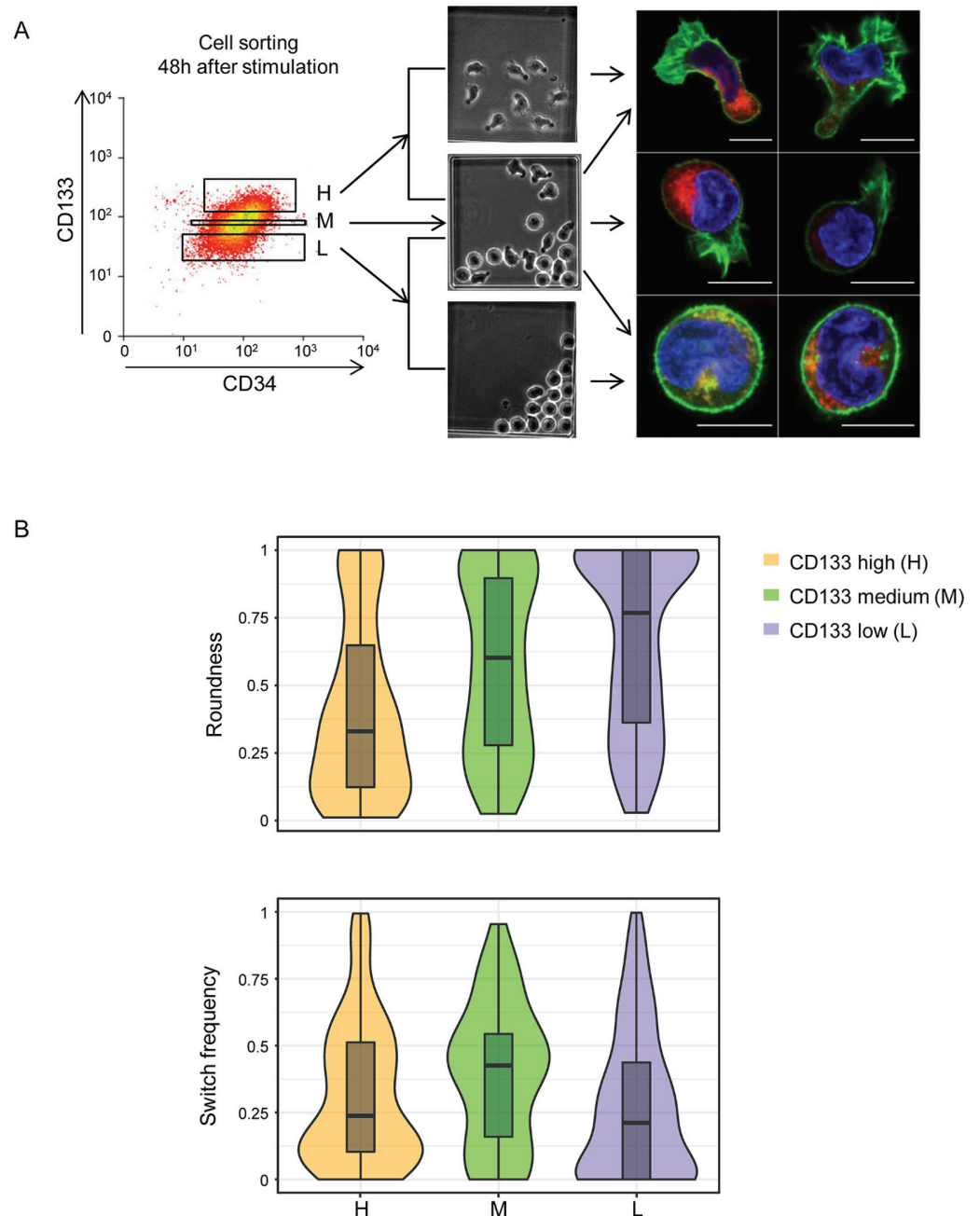


Fig 4. Isolation and time-lapse analysis of 'high', 'medium', and 'low' CD133-expressing cells.

(A) Cell-sorting strategy to isolate cells on the basis of the CD133 surface protein level. The sorted cells were cultured individually and tracked by time-lapse microscopy. They produced cells with a polarised, round, or fluctuating dynamic phenotype (illustrated by the middle panel). Examples of cells with different morphologies are shown on the right, as detected by confocal microscopy. Red: CD133 protein. Green: actin filaments detected by phalloidine. Blue: DNA. Note the preferential localisation of the CD133 protein in the uropods of polarised cells. Actin is concentrated in lamellipodia or evenly distributed in the periphery of round cells. (B) Quantitative evaluation of cell morphology and switch frequency. Distribution of the 'roundness' parameter (upper panel) indicates a gradual increase of the proportion of round cells between the 'high' and 'low' fraction. Distribution of the switch frequency as switch/h of sorted 'high', 'medium', and 'low' CD133 cells is shown in the lower panel. Note that the switch frequency is the highest in 'medium' CD133 cells. (Underlying data can be found in [S2 Data](#).)

<https://doi.org/10.1371/journal.pbio.2001867.g004>

again supporting their dynamic transitory phenotype. When expression pattern of individual genes is examined (for example, CXCR4, CXCR2, DVL2, FOXO3, NANOG, ZNF665, or TSPYL5 [S8 Fig]), some of them displayed a very different distribution in the 'medium CD133' compared to the 'high' and 'low' CD133 cells. This further demonstrates that the 'medium' CD133 cells are more than simple intermediates between the 'high' and 'low' cells; they have their own dynamic transcription profile. One can conjecture that if the cell shape correlates with the CD133 level that correlates with the transcription profile, then those cells that change shape must also change transcription profile.

PU1 and GATA1 are well-known transcription factor-coding genes that play an important role in the specification of granulocytic–monocytic and erythroid–megakaryocytic cells [34]. It has been proposed that PU1 and GATA1 can cross-inhibit each other's activity and generate a bi-stable switch between the 2 lineages [35], but more recent observations challenged this model [21]. Our analysis showed that 'high' CD133 cells express only PU1; 'medium' CD133 cells express PU1 only or coexpress the 2 genes; and 'low' CD133 cells express PU1 only, coexpress the 2 genes, or express GATA1 only (Fig 5B). This observation places the 'medium' CD133 cells as a possible intermediate between the cells expressing only PU1 and cells committed to different pathways, without providing evidence either for or against a direct competition between them.

Single-cell transcription profile of the multipotent stage

In order to determine which of the observed phenotypes correspond to the multipotent stage, we took advantage of recent observations demonstrating that the inhibition of histone deacetylase (HDAC) activity with a pharmacological agent resulted in a substantial increase in their incidence in the CD34+ cord blood population [13,36,37]. We anticipated that this would increase the proportion of cells with transcription profiles typical of the multipotent phenotype. Since valproic acid (VPA) was shown to be the most efficient [36], we used this agent to treat CD34+ cord blood cells stimulated by cytokines as above, before sampling transcription profiles. The increase of the CD90 marker (as analysed by flow cytometry) confirmed that the VPA effect was already visible after 24 h and gradually grew stronger during subsequent steps (Fig 6A and S9 Fig). The expression of CD34 and CD38 markers remained unchanged (S9 Fig). Although we did not analyse the *in vivo* potential of these cells, based on previous reports, we considered them enriched for bona fide multipotent cells. We performed single-cell qRT-PCR at 0 h, 24 h, 48 h, and 72 h after the start of the experiment, as in control cells. At all 4 time points, cell populations were very heterogeneous. At each time point, the cells displayed a unique transcription profile (Fig 6B), and no identifiable transcription patterns appeared during the 72 h of the experiment, despite slight profile evolutions. Overall, transcription patterns in individual cells were reminiscent of the uncoordinated multilineage primed profile detected in control cells at 24 h, but the 2 groups clustered separately on t-SNE maps (Fig 6C). Since the cells did not divide during the first 48 h, the increase observed in the multipotent cell fraction could not result from the selective proliferation of an initially small subpopulation of cells. Instead, this occurred because cells already present in the population changed the expression of many genes in response to the VPA.

Discussion

In this study, we aimed to identify the initial stages of fate commitment in the CD34+ cell fraction of human cord blood and determine the typical time scale for these events. Without cytokine stimulation, CD34+ cells remain quiescent and die after a few days in culture. Early acting cytokines allow these cells to survive, become metabolically active, and enter the cell cycle [38] without showing overt signs of differentiation during the first few days. This creates ideal conditions for studying early events. Our experimental design combined continuous time-lapse observations

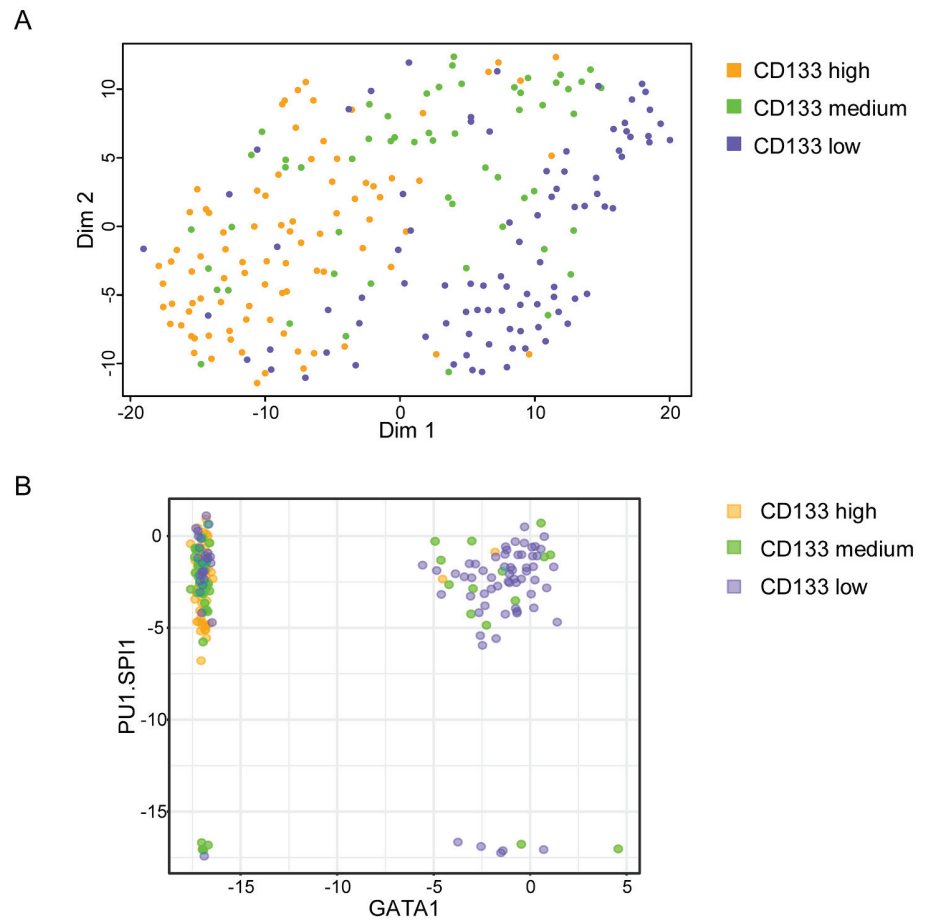


Fig 5. Single-cell gene expression in 'high', 'medium', and 'low' CD133 cells. (A) t-stochastic neighbour embedding (t-SNE) map of single-cell transcriptional data. Each point represents a single cell highlighted in a different colour for 'high', 'medium', and 'low' CD133 cells. 'High' and 'low' cells are in separated clusters corresponding to cluster #1 and #2 in Fig 1B. 'Medium' CD133 cells are distributed in and between these 2 clusters, indicating their intermediate character. (B) Scatter plot representation of PU1 and GATA1 expression in individual cells of the 'high', 'medium', and 'low' CD133 fraction. Note that GATA1 is not expressed in 'high' cells. Coexpression of the 2 genes is observed only in some 'medium' and 'low' cells. (Underlying data can be found in S1 Data.)

<https://doi.org/10.1371/journal.pbio.2001867.g005>

with snapshots of high-resolution single-cell transcriptome analysis. The data can be integrated in a dynamic fate-decision process scenario. Fate decision is necessarily accompanied by a change in the gene expression pattern. This is a multistep process. First, upon stimulation, cells rapidly reach the multiprimed state, which is characterised by a promiscuous gene expression pattern and predominantly polarised morphologies. This is an unstable phase, and 2 distinct transcription profiles start to emerge before the end of the first cell cycle. The process by which cells relax from a multiprimed to a more stable state is continuous and of variable length. Some cells reach stable morphology and a coherent, lineage-affiliated transcription profile by the end of the first cell cycle, which they then transmit to daughter cells. Other cells divide into unstable daughter cells with dynamic, 'hesitant' behaviour. This behaviour is characterised by fluctuations between polarised, actively moving amoeboid and round morphologies over several cell cycles, suggesting that instability can be transmitted mitotically. Although we have no formal evidence that the

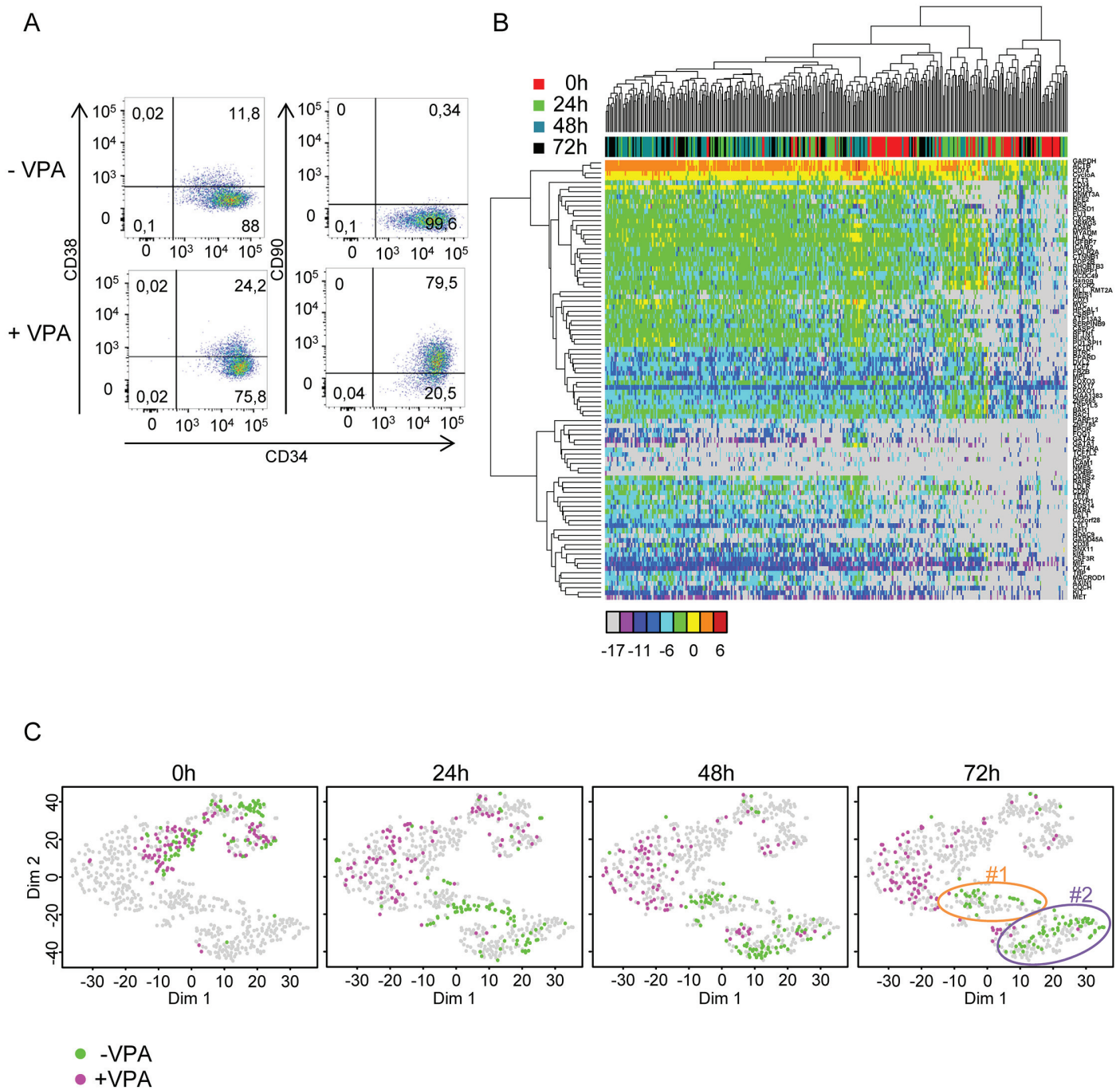


Fig 6. Transcriptional profile of cord blood-derived CD34⁺ cells treated with valproic acid (VPA) at t = 0 h, t = 24 h, t = 48 h, and t = 72 h after the beginning of the experiment as compared to untreated, normal control cells. (A) A cytometric analysis of the effect of VPA on cord blood CD34⁺ cells shows an increase in the CD90 protein in most cells, while the CD34 and CD38 markers remain essentially unchanged. (B) Heat map representation of the expression levels of 90 genes as determined by single-cell quantitative reverse transcription polymerase chain reaction (qRT-PCR) in VPA-treated cells at t = 0 h, t = 24 h, t = 48 h, and t = 72 h. The colour codes for the time points of cells are indicated on the right; the colour codes for expression levels are indicated below the heat map. Note the high heterogeneity and lack of clear clustering of the expression patterns. (C) t-distributed stochastic neighbour embedding (t-SNE) plot representation of transcription data obtained for VPA-treated cells compared to untreated normal cells (data for these cells are the same as in Fig 1). The gene expression data obtained in the 2 experiments were mapped together. Each point represents a single cell, and the cells at t = 0 h, t = 24 h, t = 48 h, and t = 72 h are highlighted separately in the 4 panels. The colour codes for VPA-treated (+VPA) and VPA-untreated (-VPA) are indicated below the panels. Clusters #1 and #2, identified at t = 48 h and t = 72 h in -VPA cells (see Fig 1), are indicated on the t = 72 h panel. Note the clear separation

of the +VPA and -VPA cells at every time point except $t = 24$ h. Note also that +VPA cells do not contribute to clusters #1 and #2, indicating that they do not acquire expression profiles typical of these cells. (Underlying data can be found in [S1 Data](#).)

<https://doi.org/10.1371/journal.pbio.2001867.g006>

transcriptome of these cells also fluctuates, 2 observations suggest that this could be the case. First, we only found 2 established transcription profiles that correspond to polarised and round morphologies with 'high' and 'low' CD133 protein levels ([Fig 1B](#)). However, we observe 3 dynamic phenotypes, 1 of which is fluctuating. Second, cells isolated on the basis of having 'medium' CD133 protein levels represent a transition between the stable polarised and round morphology. We propose a dynamic scenario in which the initial stochastic multilineage primed state is followed by a period of relaxation and uncertain 'hesitant' phase of variable length with fluctuating transcriptomes [[15,17](#)] before a stable lineage committed state is reached. In addition, this scenario is also in accordance with the recent proposal that there exists a fraction of low-primed, undifferentiated cells called 'CLOUD'-HSPCs in which that can fluctuate without passing through fixed, discrete states [[10](#)].

Increased stochastic variation in gene expression may be responsible for the rapid shift away from the initial quiescent state and lead to the uncommitted multilineage primed state [[10,12,15,16](#)]. Cell division is not required for this process; it occurs during the first cell cycle following stimulation. Cells on the path toward the new phenotype represent the committed state. The critical moment in this process is the transition between the 2 phenotypes, when the old gene network has broken down but the new network is not yet assembled. We consider that cells with fluctuating morphologies represent this transition state. The rapidity of the transition may be dependent on the time required for the new gene expression network to settle into a stable state. Since phenotypic stability of a cell lineage largely depends on the frequency of transcription initiation and the stability of the resulting mRNAs and proteins [[39,40](#)], the observed 'hesitant' phenotype might be the consequence of stochastic fluctuations due to rapid mRNA and protein turnover. The consolidation of the chromatin structure appears to be an essential element in this process, because, as shown in single-cell transcription studies, the HDAC inhibitor VPA delays the transition and blocks cells in a promiscuous gene expression pattern typical of a multilineage primed state. Indeed, HDAC inhibitors have been shown to increase gene expression stochasticity by increasing chromatin acetylation [[41](#)].

In summary, in this study, we identified the earliest phases of fate commitment in human cord blood CD34+ cells and assigned a time scale to this process. We demonstrated that the rapid initiation of the process occurs within a single cell cycle and is followed by a dynamic transition state of variable length that may span several cell cycles. Since experimental conditions were constant, the changes observed are likely to reflect cell-intrinsic processes, whereas the convergence toward a similar endpoint may reflect the constraints imposed by these conditions. From this perspective, fate decision appears to be a process of spontaneous variation/selective stabilisation reminiscent of trial-error learning, in which each cell explores many different possibilities at its own pace by expressing a large variety of genes before finding a stable combination corresponding to the actual environment. This is in remarkable agreement with earlier theoretical predictions and experimental work [[42–46](#)]. At least 3 independent theoretical models predicted the existence of an initial fluctuating phase during differentiation. According to the first theory, cell differentiation is a variation/selection process analogous to evolution [[42,43](#)]. Variations are created by stochastic fluctuations of gene expression, and some patterns are selectively stabilised through interactions with the environment and neighbouring cells. Another approach envisions cell phenotype as an attractor state in the parameter space defined by the gene expression network [[47](#)]. Differentiation is seen as a transition from 1 attractor to another and governed by the stochastic dynamics and self-organisation of the

gene network. Finally, a dynamic system view of differentiation was independently proposed by Kaneko [48]. A common theme of these approaches is the prediction that differentiating cells must necessarily go through a dynamically fluctuating phase with oscillating gene expression. Several recent studies have reported on the existence of gene expression fluctuations during the critical state transitory phase of the differentiation process [15,29,46,49–51]. Our study goes a step further by demonstrating that the cellular phenotype also fluctuates during the critical transitory phase.

Materials and methods

Ethics statement

Human umbilical cord blood (UCB) was collected from placentas and/or umbilical cords obtained from Etablissement Français du Sang (EFS), Saint Louis Hospital, France or from Centre Hospitalier Sud Francilien, Evry, France in accordance with international ethical principles and French national law (bioethics law n° 2011–814) under declaration N° DC-201-1655 to the French Ministry of Research and Higher Studies.

Human sample and cell culture

Human CD34+ cells were isolated from the mononuclear fraction of UCB samples using the autoMACSpro (Miltenyi Biotec, Paris, France) immunomagnetic cell separation system. They were then cryopreserved in Cryostor (StemCell, Paris, France) and stored in liquid nitrogen or used directly without freezing.

Cells were cultured at 37°C in a humidified atmosphere containing 5% CO₂ in a 24-well plate in X-VIVO (Lonza) supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco, Thermo Scientific), 50 ng/ml h-FLT3, 25 ng/ml h-SCF, 25 ng/ml h-TPO, and 10 ng/ml h-IL3 (Miltenyi Biotec, Paris, France) final concentration. VPA (Sigma Aldrich) was used at a final concentration of 1.25 mM.

Single-cell qRT-PCR

Single-cell qRT-PCR was carried out using the BioMark HD System (Fluidigm). Deltagenes assays (Life Technologies) were used at a final concentration of 500 nM for each of the 96 assays. Individual cells were sorted directly into a reverse transcription RT mix solution and spikes (Life Technologies) in a 96-well plate. RNA was denatured and reverse-transcribed. Twenty cycles of preamplification of 96 specific cDNA were performed by denaturing the cDNA at 96°C for 5 seconds, followed by annealing and extension at 60°C for 4 min. Unincorporated primers were cleaned up by Exonuclease I, and the preamplified products were diluted 5-fold. Amplification was performed with Evagreen supermix with low ROX (Bio-Rad) and inventoried DeltaGenes assays in 96.96 Dynamic Arrays on a BioMark HD System (Fluidigm). Cycle threshold (Ct) values were calculated from the system's software (BioMark Real-Time PCR Analysis, Fluidigm).

Single-cell data normalisation

Ct values obtained from the BioMark HD System (Fluidigm) were normalised with the help of 2 externally added controls (spike 1 and spike 4, Life Technologies) according to a set of rules provided below. For each gene, inconsistent readings or 'Failed' quality control readings were removed. Cells with failed or inconsistent detection of spikes were removed. Expression values were calculated by subtracting the gene Ct value from the geometric average of Ct values from

spike 1 and spike 4 in the corresponding cell. An arbitrary differential cycle threshold (dCt) value of -17 was assigned for all the genes with a dCt value less than -17 .

Single-cell qRT-PCR data analysis

Analyses of qRT-PCR single-cell data were done with R software (R Core Team [2016]. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>) using Heatmap3 [52], factomineR [53], k-means, and ggplot2 packages [54]. Correlation calculations were performed using custom R scripts. t-SNE and gap statistics calculations were performed as described by Grun et al. [55]

Confocal microscopy

Images were obtained with a spectral confocal Leica SP8 scanning microscope (Leica Microsystems, Germany). $5 \cdot 10^4$ cells were cultured in a 48-well plate in 200 μL prestimulation medium. After 72 h, 100 μL of 3% glutaraldehyde was added to the cell-containing well (1% final) for 15 min. Cells were washed twice with PBS 1X and incubated 2 h with 2 mg/mL NaBH_4 at room temperature. Fc γ receptors were saturated with Gamma Immune (Sigma Aldrich) for 5 min at 4°C (1:2 dilution). The cells were permeabilised with the fix/perm kit (BD-Biosciences), labeled for 20 min at 4°C with a 1:10 dilution of the mouse anti-human CD133-APC antibody (clone Ac133, Miltenyi Biotec), a 1:1,000 dilution of phalloidin-Tetramethylrhodamine B isothiocyanate (Sigma Aldrich) and stained with DAPI.

The images were acquired using a 63X PL APO CS2 1.40 NA oil immersion objective (Leica Microsystems, Germany). DAPI was excited with a 405-nm laser, TRITC with a 552-nm laser, and APC with a 635-nm laser. Finally, images were processed with a contrast enhancement algorithm (histogram equalisation) and a home-designed background subtraction algorithm.

Microgrid cell culture

A polydimethylsiloxane (PDMS) microgrid array (Microsurfaces, Australia) of 1,024 microwells (125- μm width, 60- μm depth) was placed in a specialised culture dish divided into 4 parts (Hi-Q4, Ibidi, Germany). Each part of the dish was filled with cell culture medium. A suspension of 5×10^3 cells per case was added at a concentration likely to lead to a high number of wells with a single cell.

Time-lapse microscopy

The time-lapse microscopy protocol was previously described [56]. Time-lapse acquisitions were performed with the Biostation IM time-lapse microscope (Nikon Instruments, Europe). Twenty field positions were recorded covering 4 microwells each. Images were acquired every minute for 2 d to 7 d using a 20X magnitude phase contrast objective. Only microwells containing a single cell were considered in the analyses.

Image analyses

Images were analysed using ImageJ 1.47g 64-bits software (Rasband, W.S., ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997–2014). Cell tracking was performed manually using the ImageJ TrackMate plugin. The morphologies of first, second, and third generation cells were analysed semiautomatically with Fiji (ImageJ 1.50e). A cell counter plugin was used to identify the moment when the cell switches from a round to a polarised morphology.

Analysis of the time-lapse records

Analyses of time-lapse data were performed using R software. Cell lineage representations, cycle length, roundness, and switch frequency were calculated with custom R-made scripts. Euclidean distances of the last 3 parameters (cycle length, roundness, and switch frequency) between the 2 sister cells were calculated. Cells were classified into 2 groups using the k-means algorithm: with similar or divergent dynamic phenotypes. Box plot representation combined with individual points was calculated with the beeswarm package (Aron Eklund [2016]. beeswarm: The Bee Swarm Plot, an Alternative to Stripchart. R package version 0.2.3. <https://CRAN.R-project.org/package=beeswarm>). The ggplot2 package was used to represent the roundness and switch frequency of cells sorted on CD133 protein.

Proliferation assay

CD34⁺ cells were labeled with 2.5 μM of CTV (Life technologies) at $t = 0$ h and analysed using flow cytometry (LSRII-BD biosciences, France) after 24 h, 48 h, and 72 h with ModFit LT software as described previously by Neildez et al. [57]

Image flow cytometry

Image flow cytometry analysis was performed using Image Stream MKII (Amnis, Proteigen, Merk Millipore). $5 \cdot 10^4$ cells were cultured in a 48-well plate in 200 μL prestimulation medium. After 72 h, 100 μL of 3% glutaraldehyde was added to the cell-containing well (1% final) for 15 min. Glutaraldehyde offers good preservation of cell shape. Cells were washed twice with PBS 1X and incubated 2 h with 2 mg/mL NaBH_4 at room temperature. Fc γ receptors were saturated with Gamma Immune (Sigma Aldrich) for 5 min at 4°C (1:2 dilution). Cells were labeled for 20 min at 4°C with a 1:10 dilution of mouse anti-human CD133-APC antibody (clone AC133, Miltenyi Biotec). Cells were then suspended in PBS and analysed with the image flow cytometer. Bright Field and APC channels were recorded (Bright Field: 745-nm laser; APC: 642-nm laser) with the 40X magnitude objective. Analyses of image stream data were performed with the IDEAS software (Amnis, Proteigen, Merk Millipore).

Cell sorting

The CD34⁺CD133^{high}, CD34⁺CD133^{medium} and CD34⁺CD133^{low/neg} cells were sorted at $t = 48$ h. Prior to labeling, Fc receptors were saturated with Gamma Immune (Sigma Aldrich). The CD34⁺ cells were labeled with CD34-PE (Miltenyi Biotec), CD45-APC-H7 (Beckman Coulter) and CD133-APC (clone AC133, Miltenyi Biotec) antibodies and 7-Aminoactinomycin D (Sigma Aldrich). Isotype controls were used for the gating strategy. Cells were purified using a MoFlo Astrios cell sorter (Beckman Coulter, France) and analysed with Kaluza software.

Flow cytometric analysis

The CD34⁺ cells were labeled using the following cell-surface markers: CD34-PE (Miltenyi Biotec), CD38-Pacific Blue (Beckman Coulter), and CD90-APC-Cy7 (Beckman Coulter) antibodies and 7-AAD marker (Sigma Aldrich). Isotype controls were used for gating strategies. Cells were analysed at 72 h after prestimulation by flow cytometry (LSRII-BD biosciences, France) and analysed with FlowJo (v10.1) software.

Supporting information

S1 Fig. Full set of gene expression data obtained using single-cell qRT-PCR in cord-blood CD34⁺ cells cultured in vivo with early-acting cytokines. Extended heat map of the

transcriptional profiles of cord blood-derived CD34⁺ cells at $t = 0\text{h}$, $t = 24\text{h}$, $t = 48\text{h}$ and $t = 72\text{h}$ after the beginning of the experiment. The color codes for the time points of cells are indicated on the right, the color code for expression level are indicated below the heat-map. Note the tendency of cells with the same time-points to cluster. (Underlying data can be found in [S1 Data](#)).

(TIF)

S2 Fig. Violin plot representation of individual gene expression levels at the four time points. (Underlying data can be found in [S1 Data](#)).

(TIF)

S3 Fig. Principal component analysis of single-cell expression profiles. A. 2D PCA plot. Each point represents a single cell and the different time-points are coloured differently. Color codes are in the box to the right of the plot. B. Contribution of individual genes to principal component 1 and 2. Only the 40 highest contributions are indicated. (Underlying data can be found in [S1 Data](#)).

(TIF)

S4 Fig. Analysis of cell division rates. A. The number of cells at $t = 24\text{h}$, $t = 48\text{h}$ and $t = 72\text{h}$ as observed by time-lapse microscopy. The cells of different generations are color coded in the histogram. Note that none of the cells has divided after 24 hours and only 11 of the 32 cells underwent one division after 48 hours. At $t = 72\text{h}$, three of the founder cells have not undergone division. (Underlying data can be found in [S2 Data](#)) B. Cell division analysis using Cell Trace Violet labelling. Cells were labelled at $t = 0\text{h}$ (not shown) and analyzed using flow cytometry at $t = 24\text{h}$, $t = 48\text{h}$ and $t = 72\text{h}$. When divided, the average fluorescence intensity of the two daughter cells is reduced by half compared to the maternal cell. Therefore, the peak on the right represents the parental generation. The number of the peaks to the left indicates the number of cell generations in the culture and the size of the peaks is indicative of the number of cells in each generation. Note that after 24h no cell division is detected and after 72h a fraction of undivided cells can still be detected. Most of the cells underwent one or two divisions. Overall, the profile is very similar to that detected by time lapse. (Underlying data can be found in [S3 Data](#)).

(TIF)

S5 Fig. Representations of morphological profiles of cells in three representative clones.

Each horizontal box in the three panels represents the morphology of an individual cell. The cell morphology—polarized or round—is shown with a horizontal line, the length of which is proportional to the time spent in the corresponding form. Vertical lines show the transitions between forms. The length of the horizontal lines is proportional to duration of the cell cycle and the time scale in hours is the same for each cell. The founder cell is numbered Cell_1, the two daughter cells Cell_11 and Cell_12 and granddaughter cell pairs as Cell_111, Cell_112 and Cell_121, Cell_122 respectively. In clone number 1 the polarized founder cell gives rise to frequent switcher daughters and granddaughters. Note the striking similarity of the time profiles for the morphological switches that can be observed in sister cells. In clone number 2 the polarized founder cell gives rise to stable polarized siblings. In clone number 3 the founder cell and its progeny are round. The two daughter cells switch to polarized shape for short periods. Note again the striking similarity of the sister cells' switch profiles. (Underlying data can be found in [S2 Data](#)).

(TIF)

S6 Fig. Cell morphology and CD133 localisation. Image-based cytometry analysis shows correlation between the CD133 protein expression level and cell morphology at $t = 72\text{h}$. The middle plot shows the CD133 protein density detected in glutaraldehyde-fixed cells.

Representative examples of the morphologies of “high” (upper frame) and “low” (lower frame) expressing cells are shown on the left and right respectively.
(TIF)

S7 Fig. The full set of the gene expression data obtained on “high,” “medium,” and “low” CD133 expressing individual cells. A. Heat-map representation of the expression levels of 90 genes as determined by single-cell qRT-PCR. Color codes for the “high”, “medium” and “low” fractions are indicated on the right, and the color codes for expression levels are indicated below the heat-map. Note the intermediate expression pattern of the “medium” cells. B. Principal component analysis of the single-cell gene expression data shown on the panel A. “Medium” cells are intermediate. (Underlying data can be found in [S1 Data](#)).
(TIF)

S8 Fig. Violin plot representation of individual gene expression levels in the “high,” “medium,” and “low” CD133 cells. The color code is identical to that on [S7 Fig](#). (Underlying data can be found in [S1 Data](#)).
(TIF)

S9 Fig. Cytometry analysis of the effects of valproic acid on CD34+ cells. The histogram in the left panel indicates the proportion of CD34+/CD38- cells in VPA+ and VPA- cell cultures at different time points. Note that there is no substantial difference between the two. The right panel indicates the proportion of CD34+/CD90+ cells in the same cultures. Note the increasing proportion of CD34+/CD90+ cells in VPA+ culture. This rapid increase cannot be explained by the selective proliferation of the CD90+ cells and is the result of the de novo synthesis of the CD90 protein, because as indicated in [Fig 2](#), and [S4 Fig](#), cells do not divide before 72h. (Underlying data can be found in [S3 Data](#)).
(TIF)

S1 Table. List of genes analyzed and primer sequences used for single-cell qRT-PCR amplification.
(XLSX)

S1 Data. RTqPCR normalized.
(XLSX)

S2 Data. Timelapse.
(XLSX)

S3 Data. Cytometry.
(XLSX)

S1 Movie. Time-lapse video of a cell clone with cells conserving polarized morphologies. The video has been accelerated to 5 frames per second.
(MOV)

S2 Movie. Time-lapse video of a cell clone with cells conserving round morphologies. The video has been accelerated to 5 frames per second.
(MOV)

S3 Movie. Time-lapse video of a cell clone with cells changing morphology at high frequency (dynamic phenotype of frequent switchers). Only a period between 61 and 81 h is shown. Note that individual snapshots taken at different moments may show a population composed of only polarized, only round or cells with mixed morphology. The video is the original speed.
(MOV)

Acknowledgments

The authors are grateful to Jean-Jacques Kupiec and François Delhommeau for helpful discussions, to Peggy Sanatine and Simon Jimenez (Imaging-Cytometry Platform of Genethon) for the precious help with cell sorting, to Nadège Brunel-Meunier for her help in the use of Bio-mark HD System (FluExGen Platform of Saint Antoine Hospital [Paris, France]), and to Rhonda Campbell for the language revision of the manuscript. The authors are also grateful to the mothers and the staff of the Centre Hospitalier Sud Francilien (Evry, France) for UCB samples.

Author Contributions

Conceptualization: Olivier Gandrillon, Daniel Stockholm, András Páldi.

Data curation: Alice Moussy, Jérémie Cosette, Romuald Parmentier, Cindy da Silva, Daniel Stockholm, András Páldi.

Formal analysis: Alice Moussy, Olivier Gandrillon, Daniel Stockholm, András Páldi.

Funding acquisition: Olivier Gandrillon, András Páldi.

Investigation: Alice Moussy, Jérémie Cosette, Romuald Parmentier, Cindy da Silva, Angélique Richard.

Methodology: Alice Moussy, Jérémie Cosette, Romuald Parmentier, Guillaume Corre, Angélique Richard.

Software: Jérémie Cosette, Guillaume Corre.

Supervision: András Páldi.

Validation: Alice Moussy.

Visualization: Alice Moussy, Jérémie Cosette, Daniel Stockholm, András Páldi.

Writing – original draft: András Páldi.

Writing – review & editing: Alice Moussy, Jérémie Cosette, Angélique Richard, Olivier Gandrillon, Daniel Stockholm, András Páldi.

References

1. Kawamoto H, Katsura Y. A new paradigm for hematopoietic cell lineages: revision of the classical concept of the myeloid-lymphoid dichotomy. *Trends in immunology*. 2009; 30(5):193–200. <https://doi.org/10.1016/j.it.2009.03.001> PMID: 19356980.
2. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 2000; 404(6774):193–7. <https://doi.org/10.1038/35004599> PMID: 10724173.
3. Iwasaki H, Somoza C, Shigematsu H, Duprez EA, Iwasaki-Arai J, Mizuno S, et al. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood*. 2005; 106(5):1590–600. <https://doi.org/10.1182/blood-2005-03-0860> PMID: 15914556; PubMed Central PMCID: PMC1895212.
4. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*. 1997; 91(5):661–72. PMID: 9393859.
5. Nutt SL, Metcalf D, D'Amico A, Polli M, Wu L. Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *The Journal of experimental medicine*. 2005; 201(2):221–31. <https://doi.org/10.1084/jem.20041535> PMID: 15657291; PubMed Central PMCID: PMC2212785.
6. Terszowski G, Waskow C, Conradt P, Lenze D, Koenigsmann J, Carstanjen D, et al. Prospective isolation and global gene expression analysis of the erythrocyte colony-forming unit (CFU-E). *Blood*. 2005; 105(5):1937–45. <https://doi.org/10.1182/blood-2004-09-3459> PMID: 15522951.

7. Wolff L, Humeniuk R. Concise review: erythroid versus myeloid lineage commitment: regulating the master regulators. *Stem cells*. 2013; 31(7):1237–44. <https://doi.org/10.1002/stem.1379> PMID: [23559316](https://pubmed.ncbi.nlm.nih.gov/23559316/).
8. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016; 351(6269):aab2116. <https://doi.org/10.1126/science.aab2116> PMID: [26541609](https://pubmed.ncbi.nlm.nih.gov/26541609/); PubMed Central PMCID: PMC4816201.
9. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho YJ, et al. Clonal dynamics of native haematopoiesis. *Nature*. 2014; 514(7522):322–7. <https://doi.org/10.1038/nature13824> PMID: [25296256](https://pubmed.ncbi.nlm.nih.gov/25296256/); PubMed Central PMCID: PMC4408613.
10. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nature cell biology*. 2017; 19(4):271–81. <https://doi.org/10.1038/ncb3493> PMID: [28319093](https://pubmed.ncbi.nlm.nih.gov/28319093/).
11. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 2015; 163(7):1663–77. <https://doi.org/10.1016/j.cell.2015.11.013> PMID: [26627738](https://pubmed.ncbi.nlm.nih.gov/26627738/).
12. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, et al. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes & development*. 1997; 11(6):774–85. PMID: [9087431](https://pubmed.ncbi.nlm.nih.gov/9087431/).
13. Milhem M, Mahmud N, Lavelle D, Araki H, DeSimone J, Sauntharajah Y, et al. Modification of hematopoietic stem cell fate by 5aza 2' deoxycytidine and trichostatin A. *Blood*. 2004; 103(11):4102–10. <https://doi.org/10.1182/blood-2003-07-2431> PMID: [14976039](https://pubmed.ncbi.nlm.nih.gov/14976039/).
14. Nimmo RA, May GE, Enver T. Primed and ready: understanding lineage commitment through single cell analysis. *Trends in cell biology*. 2015; 25(8):459–67. <https://doi.org/10.1016/j.tcb.2015.04.004> PMID: [26004869](https://pubmed.ncbi.nlm.nih.gov/26004869/).
15. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J, et al. Inferring rules of lineage commitment in haematopoiesis. *Nature cell biology*. 2012; 14(3):287–94. <https://doi.org/10.1038/ncb2442> PMID: [22344032](https://pubmed.ncbi.nlm.nih.gov/22344032/).
16. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*. 2016. <https://doi.org/10.1038/nature19348> PMID: [27580035](https://pubmed.ncbi.nlm.nih.gov/27580035/).
17. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*. 2008; 453(7194):544–7. <https://doi.org/10.1038/nature06965> PMID: [18497826](https://pubmed.ncbi.nlm.nih.gov/18497826/).
18. Dooner GJ, Colvin GA, Dooner MS, Johnson KW, Quesenberry PJ. Gene expression fluctuations in murine hematopoietic stem cells with cell cycle progression. *Journal of cellular physiology*. 2008; 214(3):786–95. <https://doi.org/10.1002/jcp.21273> PMID: [17894410](https://pubmed.ncbi.nlm.nih.gov/17894410/); PubMed Central PMCID: PMC4286177.
19. Koga S, Yamaguchi N, Abe T, Minegishi M, Tsuchiya S, Yamamoto M, et al. Cell-cycle-dependent oscillation of GATA2 expression in hematopoietic cells. *Blood*. 2007; 109(10):4200–8. <https://doi.org/10.1182/blood-2006-08-044149> PMID: [17255359](https://pubmed.ncbi.nlm.nih.gov/17255359/).
20. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell stem cell*. 2015; 16(6):712–24. <https://doi.org/10.1016/j.stem.2015.04.004> PMID: [26004780](https://pubmed.ncbi.nlm.nih.gov/26004780/); PubMed Central PMCID: PMC4460190.
21. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. 2016; 535(7611):299–302. <https://doi.org/10.1038/nature18320> PMID: [27411635](https://pubmed.ncbi.nlm.nih.gov/27411635/).
22. Cattoglio C, Pellin D, Rizzi E, Maruggi G, Corti G, Miselli F, et al. High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood*. 2010; 116(25):5507–17. <https://doi.org/10.1182/blood-2010-05-283523> PMID: [20864581](https://pubmed.ncbi.nlm.nih.gov/20864581/).
23. Livak KJ, Wills QF, Tipping AJ, Datta K, Mittal R, Goldson AJ, et al. Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. *Methods*. 2013; 59(1):71–9. <https://doi.org/10.1016/j.ymeth.2012.10.004> PMID: [23079396](https://pubmed.ncbi.nlm.nih.gov/23079396/); PubMed Central PMCID: PMC3562442.
24. Tibshirani R, Guenther W, Hastie T. Estimating the number of clusters in a data set via gap statistics. *Journal of Royal Statistical Society B*. 2001; 63:411–23.
25. van der Maaten L, Hinton G. Data representation using t-SNE. *Journal of Machine Learning Research*. 2008; 9:2679–05.
26. Moignard V, Macaulay IC, Swiers G, Buettner F, Schutte J, Calero-Nieto FJ, et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene

- expression analysis. *Nature cell biology*. 2013; 15(4):363–72. <https://doi.org/10.1038/ncb2709> PMID: [23524953](https://pubmed.ncbi.nlm.nih.gov/23524953/); PubMed Central PMCID: PMC3796878.
27. Bargaje R, Trachana K, Shelton MN, McGinnis CS, Zhou JX, Chadick C, et al. Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114(9):2271–6. <https://doi.org/10.1073/pnas.1621412114> PMID: [28167799](https://pubmed.ncbi.nlm.nih.gov/28167799/); PubMed Central PMCID: PMC5338498.
 28. Giuliani A. Statistical mechanics of gene expression networks: Increasing connectivity as a response to stressful condition. *Advances in Systems Biology*. 2014; 3(1):3–6.
 29. Mojtahedi M, Skupin A, Zhou J, Castano IG, Leong-Quong RY, Chang H, et al. Cell Fate Decision as High-Dimensional Critical State Transition. *PLoS Biol*. 2016; 14(12):e2000640. <https://doi.org/10.1371/journal.pbio.2000640> PMID: [28027308](https://pubmed.ncbi.nlm.nih.gov/28027308/); PubMed Central PMCID: PMC5189937.
 30. Brummendorf TH, Dragowska W, Zijlmans J, Thornbury G, Lansdorp PM. Asymmetric cell divisions sustain long-term hematopoiesis from single-sorted human fetal liver cells. *The Journal of experimental medicine*. 1998; 188(6):1117–24. PMID: [9743530](https://pubmed.ncbi.nlm.nih.gov/9743530/); PubMed Central PMCID: PMC2212541.
 31. Gorgens A, Ludwig AK, Mollmann M, Krawczyk A, Durig J, Hanenberg H, et al. Multipotent hematopoietic progenitors divide asymmetrically to create progenitors of the lymphomyeloid and erythromyeloid lineages. *Stem cell reports*. 2014; 3(6):1058–72. <https://doi.org/10.1016/j.stemcr.2014.09.016> PMID: [25448068](https://pubmed.ncbi.nlm.nih.gov/25448068/); PubMed Central PMCID: PMC4263999.
 32. Wagner W, Ansorge A, Wirkner U, Eckstein V, Schwager C, Blake J, et al. Molecular evidence for stem cell function of the slow-dividing fraction among human hematopoietic progenitor cells by genome-wide analysis. *Blood*. 2004; 104(3):675–86. <https://doi.org/10.1182/blood-2003-10-3423> PMID: [15090461](https://pubmed.ncbi.nlm.nih.gov/15090461/).
 33. Bauer N, Wilsch-Brauninger M, Karbanova J, Fonseca AV, Strauss D, Freund D, et al. Haematopoietic stem cell differentiation promotes the release of prominin-1/CD133-containing membrane vesicles—a role of the endocytic-exocytic pathway. *EMBO molecular medicine*. 2011; 3(7):398–409. <https://doi.org/10.1002/emmm.201100147> PMID: [21591261](https://pubmed.ncbi.nlm.nih.gov/21591261/); PubMed Central PMCID: PMC3210830.
 34. Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, et al. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*. 1991; 349(6306):257–60. <https://doi.org/10.1038/349257a0> PMID: [1987478](https://pubmed.ncbi.nlm.nih.gov/1987478/).
 35. Graf T, Enver T. Forcing cells to change lineages. *Nature*. 2009; 462(7273):587–94. <https://doi.org/10.1038/nature08533> PMID: [19956253](https://pubmed.ncbi.nlm.nih.gov/19956253/).
 36. Chaurasia P, Gajzer DC, Schaniel C, D'Souza S, Hoffman R. Epigenetic reprogramming induces the expansion of cord blood stem cells. *The Journal of clinical investigation*. 2014; 124(6):2378–95. <https://doi.org/10.1172/JCI70313> PMID: [24762436](https://pubmed.ncbi.nlm.nih.gov/24762436/); PubMed Central PMCID: PMC4038563.
 37. De Felice L, Tatarelli C, Mascolo MG, Gregorj C, Agostini F, Fiorini R, et al. Histone deacetylase inhibitor valproic acid enhances the cytokine-induced expansion of human hematopoietic stem cells. *Cancer research*. 2005; 65(4):1505–13. <https://doi.org/10.1158/0008-5472.CAN-04-3063> PMID: [15735039](https://pubmed.ncbi.nlm.nih.gov/15735039/).
 38. Borge OJ, Ramsfjell V, Cui L, Jacobsen SE. Ability of early acting cytokines to directly promote survival and suppress apoptosis of human primitive CD34+CD38- bone marrow cells with multilineage potential at the single-cell level: key role of thrombopoietin. *Blood*. 1997; 90(6):2282–92. PMID: [9310479](https://pubmed.ncbi.nlm.nih.gov/9310479/).
 39. Corre G, Stockholm D, Arnaud O, Kaneko G, Vinuelas J, Yamagata Y, et al. Stochastic fluctuations and distributed control of gene expression impact cellular memory. *PLoS ONE*. 2014; 9(12):e115574. <https://doi.org/10.1371/journal.pone.0115574> PMID: [25531401](https://pubmed.ncbi.nlm.nih.gov/25531401/); PubMed Central PMCID: PMC4274012.
 40. Schwanhausser B, Wolf J, Selbach M, Busse D. Synthesis and degradation jointly determine the responsiveness of the cellular proteome. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 2013; 35(7):597–601. <https://doi.org/10.1002/bies.201300017> PMID: [23696377](https://pubmed.ncbi.nlm.nih.gov/23696377/).
 41. Vinuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, et al. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC biology*. 2013; 11:15. <https://doi.org/10.1186/1741-7007-11-15> PMID: [23442824](https://pubmed.ncbi.nlm.nih.gov/23442824/); PubMed Central PMCID: PMC3635915.
 42. Kupiec JJ. A chance-selection model for cell differentiation. *Cell death and differentiation*. 1996; 3(4):385–90. PMID: [17180108](https://pubmed.ncbi.nlm.nih.gov/17180108/).
 43. Kupiec JJ. A Darwinian theory for the origin of cellular differentiation. *Molecular & general genetics: MGG*. 1997; 255(2):201–8. PMID: [9236778](https://pubmed.ncbi.nlm.nih.gov/9236778/).
 44. Paldi A. Stochastic gene expression during cell differentiation: order from disorder? *Cellular and molecular life sciences: CMLS*. 2003; 60(9):1775–8. <https://doi.org/10.1007/s00018-003-23147-z> PMID: [14523542](https://pubmed.ncbi.nlm.nih.gov/14523542/).
 45. Paldi A. What makes the cell differentiate? *Progress in biophysics and molecular biology*. 2012; 110(1):41–3. <https://doi.org/10.1016/j.pbiomolbio.2012.04.003> PMID: [22543273](https://pubmed.ncbi.nlm.nih.gov/22543273/).

46. 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. *PLoS Biol.* 2016;(14):e1002585. doi.org/10.1371/journal.pbio.1002585.
47. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. *Development.* 2009; 136(23):3853–62. <https://doi.org/10.1242/dev.035139> PMID: 19906852; PubMed Central PMCID: PMC2778736.
48. Furusawa C, Kaneko K. A dynamical-systems view of stem cell biology. *Science.* 2012; 338(6104):215–7. <https://doi.org/10.1126/science.1224311> PMID: 23066073.
49. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell.* 2012; 150(6):1209–22. <https://doi.org/10.1016/j.cell.2012.08.023> PMID: 22980981; PubMed Central PMCID: PMC3457656.
50. Hayashi K, Lopes SM, Tang F, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell stem cell.* 2008; 3(4):391–401. <https://doi.org/10.1016/j.stem.2008.07.027> PMID: 18940731; PubMed Central PMCID: PMC3847852.
51. Kobayashi T, Mizuno H, Imayoshi I, Furusawa C, Shirahige K, Kageyama R. The cyclic gene *Hes1* contributes to diverse differentiation responses of embryonic stem cells. *Genes & development.* 2009; 23(16):1870–5. <https://doi.org/10.1101/gad.1823109> PMID: 19684110; PubMed Central PMCID: PMC2725939.
52. Zhao S, Guo Y, Sheng Q, Shyr Y. Heatmap3: An Improved Heatmap Package. <https://CRANR-project.org/package=heatmap3>. 2015.
53. Le S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software.* 2008; 25:1–18.
54. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. 2009.
55. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature.* 2015; 525(7568):251–5. <https://doi.org/10.1038/nature14966> PMID: 26287467.
56. Cosette J, Moussy A, Onodi F, Auffret-Cariou A, Neildez-Nguyen TM, Paldi A, et al. Single Cell Dynamics Causes Pareto-Like Effect in Stimulated T Cell Populations. *Scientific reports.* 2015; 5:17756. <https://doi.org/10.1038/srep17756> PMID: 26648396; PubMed Central PMCID: PMC4673432.
57. Neildez-Nguyen TM, Bigot J, Da Rocha S, Corre G, Boisgerault F, Paldi A, et al. Hypoxic culture conditions enhance the generation of regulatory T cells. *Immunology.* 2014; 144:431–43. <https://doi.org/10.1111/imm.12388> PMID: 25243909; PubMed Central PMCID: PMC4557680.

3 Rôle de LDHA et du métabolisme du glucose dans l'auto-renouvellement et la différenciation de progéniteurs érythrocytaires aviaires

Mes premiers travaux de thèse ont permis de mettre en lumière le gène LDHA, dont l'expression diminue au niveau transcriptionnel, au cours de la différenciation érythrocytaires. À l'échelle de la cellule unique, LDHA représente également l'un des gènes permettant le plus de distinguer les cellules en état d'auto-renouvellement des cellules différenciées. D'autre part, LDHA fait partie des trois gènes étant corrélés au plus grand nombre d'autres gènes, aux points de différenciation critiques 8h et 24h [95]. Nous avons donc souhaité approfondir nos recherches quant à l'implication de LDHA dans l'auto-renouvellement et la différenciation des progéniteurs érythrocytaires, ce qui nous conduit à nous intéresser de façon plus générale au métabolisme glycolytique. Nous avons alors émis l'hypothèse selon laquelle la différenciation érythrocytaire pourrait s'accompagner d'une transition métabolique de la glycolyse anaérobie vers l'OXPPOS. Par ailleurs, comme mentionné dans l'introduction de ce manuscrit, la combinaison de ces travaux sur le métabolisme, avec le travail effectué sur la variabilité de l'expression des gènes prend tout son sens, notamment au regard des connaissances actuelles à propos du lien épigénétique rapprochant métabolisme et expression génique.

Cette seconde partie de ma thèse fait l'objet de travaux encore en cours de préparation. Les résultats d'ores et déjà obtenus seront présentés sous la forme d'un article scientifique pas encore finalisé. Il est donc important de noter qu'il s'agit encore d'une étude incomplète dans laquelle certaines expériences devront être répétées et quelques données restent encore à acquérir, afin d'obtenir des conclusions robustes.

3.1 Article 4 : Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements

3.1.1 Principaux résultats de l'article 4

Cette étude a pour objectif d'analyser les variations du métabolisme glycolytique au cours de la différenciation érythrocytaire, ainsi que de comprendre le rôle de LDHA dans l'auto-renouvellement des progéniteurs érythrocytaires.

Dans un premier temps, nous avons analysé le niveau d'expression de LDHA et de huit autres protéines impliquées dans la glycolyse ou dans la phosphorylation oxydative (OXPHOS), dans des populations de progéniteurs érythrocytaires, les T2EC [179], en état d'auto-renouvellement (0h), ou induites en différenciation. Cette première analyse moléculaire suggère qu'au cours de la différenciation érythrocytaire, la glycolyse anaérobie pourrait être réduite. En effet, après une augmentation du niveau protéique de LDHA à 12h de différenciation, celui-ci chute drastiquement jusqu'à 72h. Parallèlement la phosphorylation oxydative semblerait augmenter dans les premières heures de la différenciation, puis se stabiliser, comme suggéré par les fluctuations du niveau des protéines de la chaîne respiratoire.

Afin de confirmer cette hypothèse nous avons ensuite comparé la production de lactate et le potentiel de membrane mitochondrial (MMP) de T2EC en état d'auto-renouvellement et induites à se différencier. Nos résultats suggèrent que la production cellulaire de lactate est plus importante en état d'auto-renouvellement et que le MMP augmente au cours de la différenciation. Par ailleurs, nous avons également mesuré la respiration cellulaire de T2EC en milieu d'auto-renouvellement et en milieu de différenciation. Cette analyse a mis en évidence un pic de la respiration cellulaire à 24h de différenciation, majoritairement allouée à la production d'ATP. Ce pic s'accompagne également d'une augmentation de la capacité oxydative cellulaire, qui quant à elle reste stable, comme précédemment suggéré par l'analyse moléculaire de protéines de la chaîne respiratoire. L'ensemble de ces résultats confortent ainsi l'hypothèse selon laquelle la différenciation érythrocytaire pourrait s'accompagner d'un changement de statut métabolique, passant de la glycolyse anaérobie vers l'OXPHOS. D'autre part, le point 24h, préalablement montré comme étant critique au niveau de l'expression des

gènes, pour l'engagement des cellules en différenciation [95], est à nouveau mis en lumière par cette nouvelle analyse. Une augmentation des besoins énergétiques semble donc nécessaire à ce moment crucial de la différenciation érythrocytaire.

Enfin, nous avons étudié les conséquences de l'inhibition de LDHA sur l'auto-renouvellement des T2EC. Nos résultats suggèrent que LDHA est nécessaire au maintien en auto-renouvellement des progéniteurs érythrocytaires. Nous avons alors proposé l'hypothèse selon laquelle la diminution de l'expression de LDHA serait nécessaire pour permettre aux cellules de s'échapper de leur état d'auto-renouvellement, impliquant une diminution de la glycolyse anaérobie, et d'entrer en différenciation. Afin d'apporter un premier élément de réponse, nous avons analysé l'effet de l'inhibition de LDHA sur le MMP dans des T2EC en auto-renouvellement. L'inhibition de LDHA entraîne une augmentation du MMP, suggérant ainsi que LDHA pourrait être impliquée dans le changement de l'état métabolique des T2EC. Cependant, afin de pouvoir tester de façon plus complète cette hypothèse, un vecteur de surexpression de LDHA est en cours de construction dans le laboratoire. Prochainement, nous pourrons ainsi surexprimer LDHA dans des T2EC induites en différenciation, afin d'appréhender leur capacité à se différencier en présence de LDHA et de signaux externes de différenciation.

3.1.2 Principales conclusions de l'article 4

Nos premiers résultats suggèrent que le processus de différenciation érythrocytaire s'accompagne d'un changement métabolique correspondant au passage de la glycolyse anaérobie, vers une production d'énergie aérobie, reposant sur la phosphorylation oxydative. De plus, nous avons mis en évidence un pic énergétique à 24h de différenciation, parallèlement à une augmentation de la capacité oxydative cellulaire et de l'expression des enzymes glycolytiques LDHA et de PKM, et des protéines des complexes de la chaîne respiratoire. Enfin nos résultats suggèrent que LDHA pourrait jouer un rôle important dans le maintien des progéniteurs érythrocytaires en auto-renouvellement, et pourrait être à l'origine de modifications du métabolisme.

3.1.3 Article 4

En cours de préparation

Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements

Angélique Richard¹, Elodie Vallin¹, Caroline Romestaing², Damien Roussel², Olivier Gandrillon^{1,3}, Sandrine Gonin-Giraud¹

1 Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratory of Biology and Modelling of the Cell, 46 allée d'Italie Site Jacques Monod, F-69007, Lyon, France

2 Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, UMR 5023 CNRS, Université de Lyon, France

3 Inria Team Dracula, Inria Center Grenoble Rhône-Alpes, France

Abstract

It has been suggested that a switch from anaerobic glycolytic metabolism toward mitochondrial oxidative phosphorylation (OXPHOS) could be a driving force during stem cell differentiation.

Based upon our initial results regarding the drop in LDHA at the mRNA level during the differentiation of chicken erythroid progenitors we decided to investigate the possibility that such a switch might also be operating in those cells.

We first analyzed the level of 9 enzymes, including LDHA, involved either in glycolysis or OXPHOS, in self-renewing progenitor cells and at different time-points of their differentiation process. Our results support the hypothesis that erythroid differentiation might be accompanied by an enhancement of the respiratory chain and the anaerobic glycolysis activities at 12h, followed by a strong decline of the glycolytic pathway and a stabilization of OXPHOS until 48h, at a time when OXPHOS might slightly decrease.

To confirm that OXPHOS might be increased and anaerobic glycolysis decreased during erythroid differentiation, we measured lactate concentration and mitochondrial membrane potential (MMP) of self-renewing and differentiating cells. Our findings shows that at 12h-24h of differentiation, a surge of energy is needed, which could be fueled jointly by anaerobic glycolysis and OXPHOS. Then the energy demand comes back to normal and might be supplied by OXPHOS instead of anaerobic glycolysis. These results brought new elements in support of the hypothesis suggesting that erythroid differentiation is associated with a metabolic switch from anaerobic glycolysis to OXPHOS.

Finally, we also assessed LDHA role in erythroid progenitors self-renewal and the metabolic states changes. Inhibitory experiments, using two different strategies, showed that LDHA activity might influence the metabolic status of erythroid progenitors, and plays an important role in the maintenance of their self-renewal state.

We propose that epigenetics could be the link between the concomitant gene expression and metabolism rearrangements which might represent a circular causal driving force pushing the cells out of their self-renewing state and into their new differentiated state.

Introduction

Metabolism is a biological process mostly involved in energy consumption, production and distribution, which is essential for cell functions and survival.

An emerging theme is the possible causal involvement of metabolic changes during a differentiation process. In particular, anaerobic glycolysis was shown to be characteristic of self-renewing stem cells and of different types of progenitors, such as neuronal and hematopoietic progenitor cells [1, 2] and it has been demonstrated that stem cells tend to switch from anaerobic glycolytic metabolism toward mitochondrial oxidative phosphorylation (OXPHOS) while they differentiate [3, 4, 5]. A decline of the anaerobic glycolysis signature during differentiation was also identified in other cell types, and is generally followed by an enhancement of OXPHOS that takes over energy production. It was suggested that a balance between glycolysis and OXPHOS metabolism could somehow guide the choice between self-renewal and differentiation in stem cells fate [6]. It has also been demonstrated that preventing the shutting down the glycolytic pathway was detrimental for neuronal differentiation [4].

In light of these new insights we decided to explore the hypothesis that erythroid progenitor cells could undergo a metabolic switch from anaerobic glycolysis toward OXPHOS while they differentiate.

Erythroid progenitors depend on anaerobic glycolysis and present a Warburg-like profile as they display a high proliferation rate and produce an abundant amount of lactate [7]. However, few is known about glucose metabolism behavior during erythroid differentiation, when compared to erythroid progenitors self-renewal. Mature mammals erythrocytes were shown to rely upon lactate dehydrogenase to fulfill their functions, depending therefore upon the anaerobic pathway [8, 9]. This is in agreement with the fact that mammals erythroid cells totally eliminate their mitochondria at the end of their maturation [10]. Therefore, it is not possible for mammal erythrocytes to produce ATP through mitochondrial OXPHOS anymore. However it is different for avian erythroid cells, that keep their mitochondria until the very last mature state [11].

Recent results from high-throughput transcriptome and proteome analysis highlighted that key actors of the anaerobic metabolism such as LDHA decreased while erythroid progenitors differentiate both in chicken [12] mice [13] and human [14]. These new findings suggested that erythroid progenitors glycolytic metabolism could escape from its anaerobic state during the differentiation process, and might switch toward OXPHOS.

In the present study, we explored different parameters of the glycolytic metabolism during the first three days of primary avian erythroid progenitor cells differentiation. We first analyzed the protein level of a few key enzymes involved either in glycolysis or OXPHOS pathway, at six time-points of the differentiation process. To confirm our results at the physiological level we then compared lactate production and mitochondrial membrane potential (MMP), that partially reflects OXPHOS activity, of self-renewing and differentiating cells. We further dissected OXPHOS mechanisms by analyzing cell respiration response to different respiratory chain inhibitors. LDHA role in erythroid progenitors self-renewal was assayed using known molecular inhibitors and a the CRISPR/Cas9 knock-out approach. Finally, to investigate whether LDHA downregulation could be responsible for the metabolic state changes observed during erythroid differentiation, we measured MMP of self-renewing cells following

inhibition of LDHA activity. Our findings support our hypothesis that erythroid maturation could be accompanied with a switch from anaerobic glycolysis toward OXPHOS.

Materials and methods

Cell culture conditions and treatments

T2EC were grown as previously described [15]. For LDHA inhibitors treatment, cells were cultured at $1,25 \cdot 10^5$ cells/ml in fresh culture media containing either Galloflavin ($50 \mu\text{M}$) (SIGMA), FX11 ($30 \mu\text{M}$) (SIGMA) or DMSO. Thereafter, cells were counted every 24h, or harvested for lactate and ISX analysis.

Cell growth and differentiation measurements

Cell growth was evaluated by counting living cells using Trypan blue staining. Cell differentiation was assessed using benzidine staining (SIGMA, [16]).

Cell transfection and sorting

Cells were transfected by electroporation using the Neon transfection system 100ul kit (Thermo fisher) according to manufacturers recommendations. For each condition $10 \cdot 10^6$ cells were washed once with 1X PBS. Cell pellets were resuspended in T transfection buffer supplied in the kit and $10 \mu\text{g}$ of pCRISPR-P2A-tRNA (see subsection 3.8 for a description) was added. Cells were loaded into dedicated tips using the Neon pipette and electroporated (1500V 20ms 3 pulses) using the Neon transfection system device. After electroporation cells were immediately transferred in pre-warmed post-transfection medium (made with RPMI 1640 instead of αMEM , and antibiotics removed). Following 3h of incubation in the post-transfection medium, cells were pelleted by centrifugation, resuspended in fresh medium and grown in standard culture conditions.

After 16h, transfected cells were harvested and resuspended in RPMI 1640 with 2% FCS. Sorting was performed using FACS Aria 1 flow cytometer. Non-transfected cells were used as a negative control to discriminate between positive and negative GFP cells. Thus living GFP-expressing cells were collected, pelleted by centrifugation and grown back in LM1 or DM17 fresh media.

ImageStreamX analysis

Sample preparation

Cells were pelleted by centrifugation, resuspended in fresh media and splitted in half. One part was incubated in culture conditions for 10 min with FCCP ($10 \mu\text{gM}$), whereas the other one was incubated without any addition. FCCP (carbonyl cyanide-p-trifluoromethoxyphenylhydrazone) (ABCAM) was used as a negative control for mitochondrial membrane potential (MMP) measurements.

A staining solution was prepared with MitoTracker Green FM (50 nM) (BD Horizon), TMRE (ultra pure) mitochondria dye (4 nM) (Enzo) and Hoescht 33342 ($2.5 \mu\text{g/ml}$) (REF) diluted in fresh media. Following incubation, cells were harvested, resuspended in the staining solution and incubated again for 20 min in culture conditions. Following incubation cells were washed three times with 1X PBS. To discriminate between live and dead cells, they

were finally incubated for 15 min at 4°C in the dark with Fixable Viability Stain 660 (FVS 660; BD Horizon) at 1:2000 dilution. Cells were washed twice and resuspended in 1X PBS before measurements with ImageStreamX Mark II (Amnis). Single staining were performed in parallel as compensation samples.

Images acquisition

Stained cells were loaded within ISX device. During acquisition, cells were gated according to the area and aspect ratio features. For each condition, ten thousand cells were recorded with 60x objective at the rate of 60 cell/s. Lasers 405, 488, 561, 642, 785 nm were used to analyze respectively DNA, mitochondria, mitochondrial membrane potential, dead cells, and cell granularity.

Image analysis

Data generated from ISX were analysed with the dedicated IDEAS software. First, image compensation was performed using single stained samples. Then cells were more precisely gated for focused cells using bright-field and gradient RMS feature, for singlets using area and Hoechst intensity features, and for living cells using intensity feature of FVS 660 (negative cells were gated). Then, cell size, cellular mitochondrial content and MMP were assessed from bright-field images, green mitotracker intensity, and TMRE intensity, respectively, using cell perimeter and intensity features. Those data were then exported and analyzed with R software [17].

Cell respiration measurements

Oxygen consumption rates were measured using Oroboros oxygraph (Oroboros Instruments), composed of two analysis chambers. Four respirometers were available and two conditions per apparatus could be tested, therefore eight conditions per experiment were performed. To avoid technical variability when comparing kinetic time points, conditions replicates were loaded randomly in different respirometers. One million cells per chamber were loaded at a final density of 5.10^5 cells/ml in their culture media. During measurements cells were maintained at 37°C under constant stirring within the chambers.

Routine respiration was first measured after respiration stabilization in the chambers. To estimate proton leak, the ATP synthase inhibitor Oligomycin (1,25 $\mu\text{g/ml}$) was added in the chambers. Then, a subsequent titration from 2.5 μgM to 10 μgM of FCCP was performed to assess maximal respiratory capacity. To evaluate mitochondrial contribution to total cellular oxygen consumption, mitochondrial respiratory chain was impaired by inhibiting cytochrome b-c1 complex (complex III) using Antimycin A (22.5 μgM). Finally, to measure cytochrome C oxidase complex (complex IV) capacity, Ascorbate was added at a final concentration of 2,5mM followed by a subsequent titration of TMPD (N,N,N',N'-tetramethyl-p-phenylenediamine) from 125 μM to 750 μM .

Measurement of extra-cellular lactate concentration

Cellular supernatants were loaded in a 10 kDa ultracentrifugation filter (Amicon ; Merk Millipore) to eliminate lactate dehydrogenase enzymes, as suggested by manufacturers protocol. Soluble fractions were then assessed using the Lactate assay kit (SIGMA) according to manufacturers procedure, and samples absorbance was measured at 570 nm using the Multiskan GO plate reader (Thermo Scientific). Background absorbance was subtracted from samples reading and lactate amount was determined using a standard dilution curve supplied by the Lactate assay kit. Lactate concentration was then calculated as recommended by manufacturers protocol. To avoid bias in lactate concentration measurements linked to differences in cell culture density, supernatants were diluted regarding cell concentrations.

Proteomic data generation

Proteomic data were generated using by the Plateforme de Protéomique de l'Université Paris Descartes (3P5), under the supervision of Dr Patrick Mayeux, using a label-free quantification (LFQ) method [14]

CRISPR plasmid construction

T2A substitution by P2A

It has been shown in various species that P2A (2A peptide derived from porcine teschovirus-1) presents a higher cleavage efficiency than T2A (2A peptide derived from Thoseaasigna virus) [18], which was initially present in pSpCas9(BB)-2A-GFP (PX458) vector (Addgene). Thus we decided to substitute T2A for P2A to enhance vector efficiency.

Oligonucleotides encoding P2A and homologous sequences were purchased from Eurogentec (table 1). P2A was substituted for T2A using the PCR Geiser method [19]. P2A oligonucleotides were used as mega primers for PCR on PX458 vector.

Multiplex CRISPR system construction

pCRISPR-P2A vector was modified to produce a multiplex CRISPR/Cas9 system to allow expression of two RNA guides (gRNA) with the same vector. Therefore we added a tRNA and a new restriction multiple site to insert a second gRNA in pCRISPR-P2A vector. Oligonucleotides encoding chicken tRNA Glu and RNA scaffold were purchased from Eurogentec, annealed and then cloned into XbaI-digested pCRISPR-P2A vector. After validation by sequencing the final plasmid was named pCRISPR-P2A-tRNA.

LDHA gRNA cloning

Two RNA guides (crLDHA#1 and crLDHA#2) against LDHA sequence were designed using CRISPR design tool (<http://crispr.mit.edu>) to target the first exons of the sequence and purchased. The first guide was cloned after hU6 promoter into BbsI-digested pCRISPR-P2A-tRNA vector. The second guide (crLDHA#2) was cloned after chicken tRNA Glu into SapI-digested pCRISPR-P2A-tRNA vector.

Oligonucleotides	Sequences
P2A-Forward	gccaggcaaaaaagaaaaagGGAAGCGGAGCTACTAACTTCAGCCTGCTG AAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTgtgagcaag ggcgaggagct
P2A-Reverse	agtcctcgcaccttgctcacAGGTCCAGGGTTCTCCTCCACGTCTCCAGCCT GCTTCAGCAGGCTGAAGTTAGTAGCTCCGCTTCCctttttcttttt tgctggc
tRNA-Forward	ctagaTCCCATATGGTCTAGTGGTTAGGATTCCTGGTTTTTCACCC AGGCGGCCCGGGTTCGACTCCCGGTATGGGAATGAAGAGCGCTCT TCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT TATCAACTTGAAAAAGTGGCACCCAGTCCGGTGCTTTT
tRNA-Reverse	ctagAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGGA CTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTGAAGAGC GCTCTTCATTTCCCATACCGGGAGTCGAACCCGGGCCGCTGGGTG AAAACCAGGAATCCTAACCACTAGACCATATGGGA
crLDHA#1- BbsI-Forward	caccgTGAAGATATTGACGTTGCGT
crLDHA#1- BbsI-Reverse	aaacACGCAACGTCAATATCTTCac
crLDHA#2- SapI-Forward	gaaTGCAGCCACTACCGATAACA
crLDHA#2- SapI-Reverse	aacTGTTATCGGTAGTGGCTGCA

Table 1: Oligonucleotides sequences used for CRISPR plasmid construction

The efficiency of our CRISPR vector was confirmed by analyzing LDHA deletion by PCR on transfected cells (data not shown).

Statistical analysis

Every statistical analyses were performed using R software [17]. The normality of each distribution was checked in order to compute the appropriate tests.

A paired t-test was performed to compare mitochondrial content between the different time-points.

For each respiration parameter assessed, the mean oxygen consumption rate was calculated, among five experimental replicates analyzed in duplicates. A paired t-tests were performed to compare distributions means.

The significance of the difference between growth ratios was determined using a Wilcoxon test.

Results

Variations of the level of glycolytic enzymes during erythroid differentiation

We first analyzed protein levels of glycolytic enzymes, in self-renewing and in differentiating primary chicken erythroid progenitor cells (T2EC). In order to dissect energetic glycolytic metabolism we selected important proteins from the following categories : (1) proteins involved in glycolysis, (2) proteins involved in oxidative phosphorylation (OXPHOS), and (3) proteins involved in anaerobic glycolysis (Fig 1A).

Glycolysis is finely regulated at three key rate-limiting steps catalysed by Hexokinases (HK), Phosphofruktokinase (PFKP) and Pyruvate kinase (PKM) [20, 21, 22]. Thus we decided to analyse the expression of these enzymes including two isoforms, HK1 and HK2, of hexokinase (Fig 1B). Globally HK1, PKM and PFKP expression was sharply decreased (more than 70%) during the differentiation process, whereas HK2 expression decreased to a lesser extent (less than 50%). However PKM protein level increased slightly (around 10%) at 12h. The important decline of the expression level of these proteins suggested that self-renewing T2EC displayed a high glycolytic flow, that was strongly reduced during differentiation.

We also analyzed the protein level of the single enzyme involved in anaerobic glycolysis, LDHA (Fig 1C). LDHA expression first increased (20%) at 12h, and then dropped massively (more than 80%) until 72h. Such an important decrease of LDHA protein expression implies a decrease of pyruvate conversion into lactate and suggest that anaerobic glycolysis might be reduced during erythroid differentiation.

Glycolysis end-product is pyruvate, that can enter in the anaerobic pathway, or the aerobic pathway through OXPHOS, to produce ATP. Briefly, pyruvate must be converted to acetyl CoA within mitochondrias, and enter in the tricarboxylic acid (TCA) cycle to generate electron transporters, NADH,H⁺ and FADH₂, in order to feed respiratory chains coupled with ATP production, i.e. OXPHOS (Fig 1A). Thus, we included in OXPHOS genes category, pyruvate dehydrogenase complex subunits, PDHA1 and PDHB, that catalyse pyruvate conversion into acetyl CoA [23, 24]. Moreover we also selected cytochrome c electron carrier (CYC) involved in respiratory chain [25], and the ubiquinol-cytochrome c reductase (UQCRFS1), a key component of complex III of the respiratory chain [26] (Fig 1D). PDHA1 expression do not show any clear tendency during differentiation, whereas PDHB expression first increased slightly (20%) at 12h and then declined slowly. Expression of respiratory chains components UQCRFS1 and CYC displayed an interesting profile. They first show a 40% raise at 12-24h of differentiation and then returned progressively to the protein level displayed in self-renewing T2EC. Increased expression of PDH complex subunit PDHB suggested that more pyruvate molecules might be recruited in the mitochondria to be converted into acetyl CoA at 12h of the differentiation process. Moreover the increase of UQCRFS1 and CYC proteins also suggest that chain respiratory activity might be boosted at that stage of the differentiation process. It is important to note here that during erythroid differentiation, gene expression decreases progressively because of the nucleus condensation [14]. The slight decrease of the respiratory chain proteins after 24h thus, could be due to this general decline, which did not mean necessarily that OXPHOS capacity is reduced. PDHA1 level remained stable during the differentiation, despite the general gene expression decrease, suggesting

that it is particularly needed.

Overall, during T2EC differentiation, LDHA and PKM protein levels increased at 12h, and then decreased significantly. The other glycolysis-related enzymes decreased gradually from the self-renewal state (0h) up to 72h. OXPHOS-related enzymes increased at 12-24h, whereas PDC complex subunits remained most likely stable over differentiation. One can therefore suggest that erythroid differentiation might be accompanied by an enhancement of the respiratory chains and the anaerobic glycolysis activities at 12h, followed by a strong decline of the glycolytic pathway and a stabilization of OXPHOS until 48h, at a time when OXPHOS might slightly decrease.

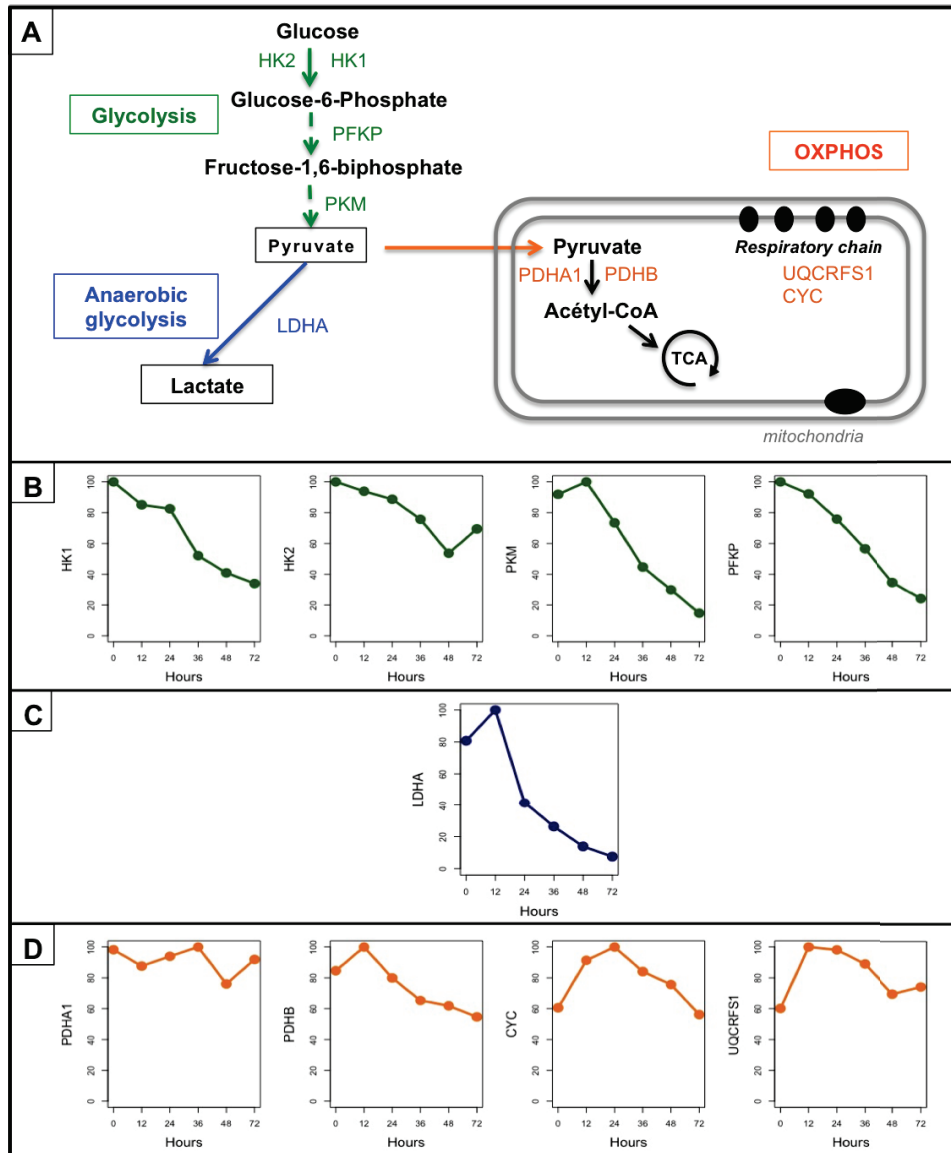


Figure 1: Variations in expression of relevant proteins for glycolytic metabolism analysis

(A) Proteins were selected from three different categories : key enzymes involved in glycolysis regulation (Glycolysis), LDHA alone (anaerobic glycolysis), and OXPHOS-related enzymes (OXPHOS). (B-D) Protein quantification was performed using mass spectrometry (MS/MS) upon three independent self-renewing T2EC populations and three independent T2EC differentiation kinetics (T2EC induced to differentiate for 12, 24, 36, 48 and 72h). Mean protein levels were normalized as follows : for each enzyme the maximal value was considered equal to 100, and all other values were proportionally converted. Normalized protein levels are represented on the y-axis whereas x-axis correspond to differentiation kinetic time-points (hours). LDHA: Lactate dehydrogenase A, HK1: Hexokinase isoform 1, HK2: Hexokinase isoform 2, PFKP : Phosphofructokinase Platelet, PKM: Pyruvate Kinase Muscle, PDHA1: Pyruvate Dehydrogenase E1 Alpha 1 Subunit; PDHB: Pyruvate Dehydrogenase E1 Beta Subunit, UQCRFS1: Ubiquinol-Cytochrome C Reductase, Rieske Iron-Sulfur Polypeptide 1, CYC: Cytochrome C

Energetic glycolytic metabolism changes during erythroid differentiation

We first compared the capacity of self-renewing and differentiating T2EC to produce lactate. We observed that the differentiation was accompanied by a decrease in the ability of our cells to produce lactate (Fig 2). This result is in line with observed decrease in LDHA protein expression, suggesting that anaerobic glycolysis was indeed reduced during the differentiation process

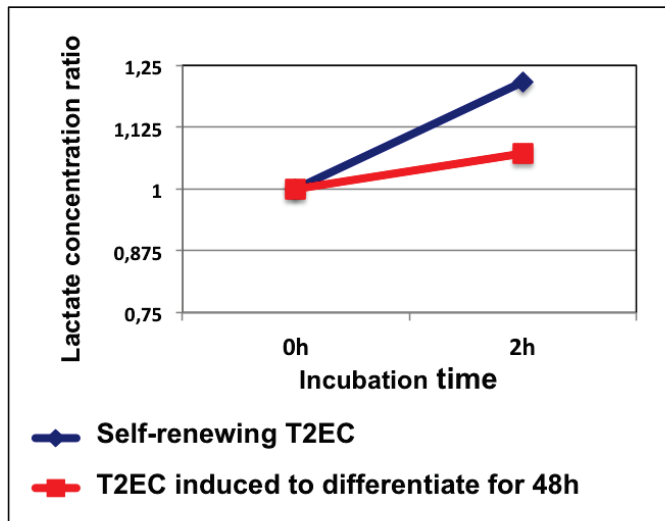


Figure 2: Lactate production by self-renewing and differentiating cells

Extra-cellular lactate concentration in self-renewing cells and cells induced to differentiate for 48h was assessed following incubation during 2h in fresh media. The red curve corresponds to lactate production of T2EC induced to differentiate for 48 h, the blue curve represents lactate production of self-renewing T2EC. Lactate concentration ratio was calculated as extra-cellular lactate concentration divided by lactate concentration in fresh media (0h). Shown is the result from one experiment.

We then wanted to evaluate whether glycolytic metabolism might switch toward OXPHOS during the erythroid differentiation process. For this we decided to use TMRE staining as a measure of mitochondrial activity. Since we know that cell size strongly decreases during differentiation [12] and that the overall amount of mitochondria has been previously shown to correlate with cell size [27], we first assessed the overall amount of mitochondria using a green mitotracker. As expected mitotracker intensity showed a strong decrease at 24h of the differentiation (Fig 3). Thus to account for cellular mitochondrial content variations, we normalized TMRE intensity with the green mitotracker intensity. We then assumed that the mitochondrial membrane potential (MMP) corresponds to the ratio of TMRE upon mitotracker intensity.

Finally we used the FCCP uncoupler as a control for TMRE staining. As expected, following FCCP treatment TMRE intensity was strongly reduced, whereas mitotracker intensity

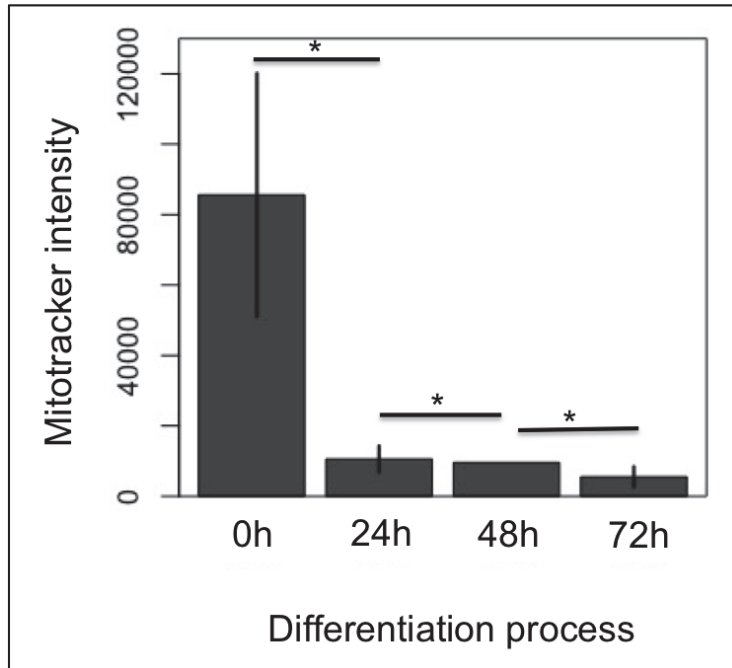


Figure 3: Mitotracker intensity evolution during the erythroid differentiation process

Shown is the distribution of the mitotracker intensity at each time-point of the differentiation process. Data were obtained from one experiment. 3814 cells were analyzed for 0h, 4781 cells for 24h, 5163 cells for 48h and 4219 cells for 72h.

was maintained, confirming that mitotracker is not dependent from MMP (Fig 4A).

We finally plotted MMP and analyzed its behavior during the differentiation process (Fig 4C). We noted that MMP increased significantly during erythroid differentiation. Such an increase is a strong indication of an enhanced OXPHOS, and thus tend to support the hypothesis that T2EC switched from anaerobic glycolysis toward OXPHOS during their differentiation process.

In order to study OXPHOS activity into finer details, we assessed oxygen consumption in self-renewing and differentiating T2EC (Fig 5). We first observed that routine cell respiration was significantly increased in T2EC induced to differentiate for 24h, and then decreased at 48h and 72h to reach a similar level than self-renewing T2EC respiration rate (Fig 5A).

Different calculations were then performed to better decipher the respiration mechanisms. First, to focus only on mitochondrial oxygen consumption, respiration measured after antimycin A addition was subtracted from all respiration measurements. We noted that the remaining respiration rate following antimycin A treatment was very low compared to routine respiration, suggesting that mitochondria participated mostly to oxygen consumption (data not shown).

After oligomycin addition, we evaluated oxygen consumption resulting from proton leak. Oxygen consumption dedicated to ATP production was thus estimated by removing respiration due to proton leak from routine respiration (Fig 5B). Similarly to routine respiration

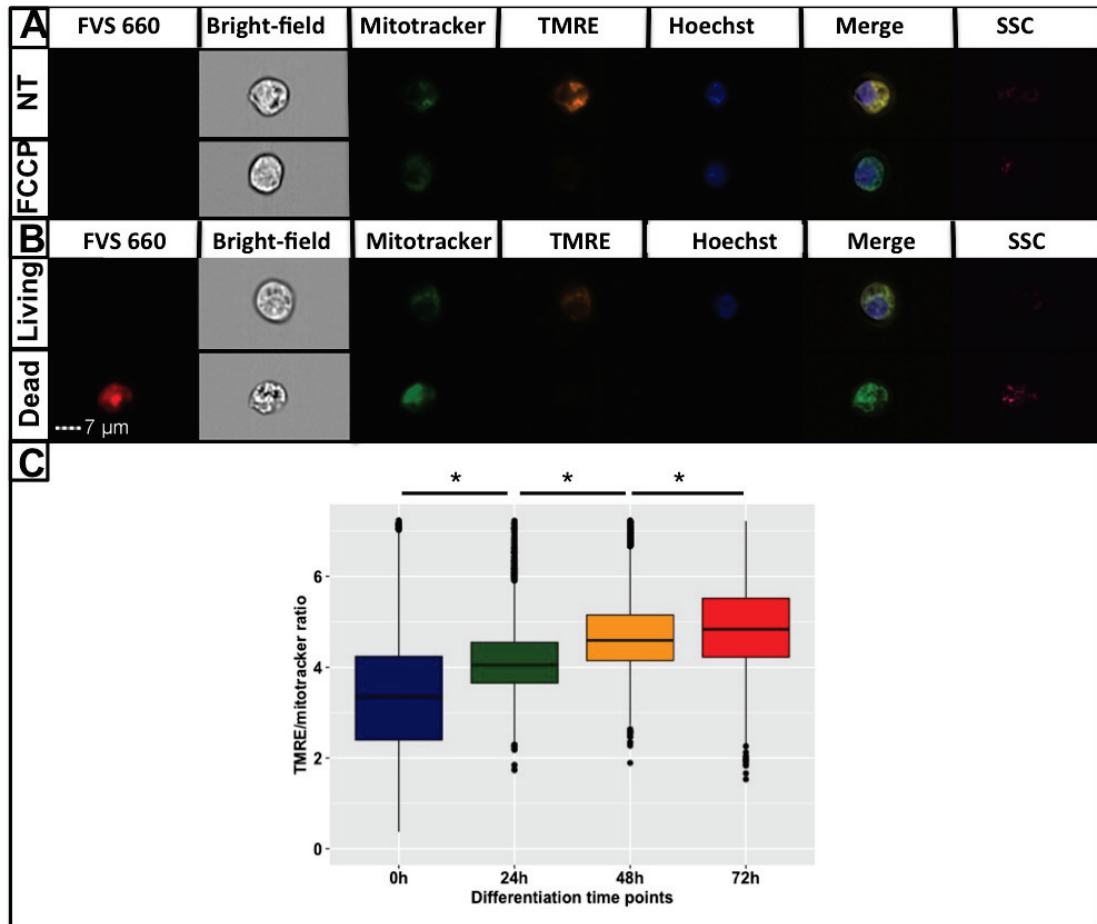


Figure 4: Mitochondrial membrane potential (MMP) evolution during erythroid differentiation

MMP was assessed using TMRE dye and normalized with green mitotracker intensity. Nucleus was stained using Hoechst and cell granularity (SSC) is represented in pink. Living cells were identified using the live/dead FVS 660 dye. (A) The FCCP uncoupler was used as a negative control of MMP staining. In the presence of FCCP, only TMRE staining disappeared, reflecting the specific response of TMRE to the dissipation of mitochondrial membrane potential by uncoupling, and mitotracker independence from MMP. It also confirmed that the TMRE concentration used is a nonquenching dose. NT: No treatment. (B) Shown is an example of living and dead cells. (C) Normalized TMRE distributions. Outliers are not shown. A paired t-test was applied to assess whether distributions were significantly different (p -value < 0.05). Data were obtained from one experiment. 3814 cells were analyzed for 0h, 4781 cells for 24h, 5163 cells for 48h and 4219 cells for 72h.

profile along differentiation, the oxygen consumption devoted to ATP production increased at 24h and then decreased significantly to come back to the initial respiration rate at 72h (Fig 5B). Therefore, the surge observed in routine respiration at 24h might be due to an increase of cellular energy needs. Moreover, as shown upstream, cellular mitochondrial content decreased during the differentiation process as soon as 24h (Fig 3) and LDHA protein expression was drastically reduced at 24h of the differentiation process (Fig 1C), suggesting that the increased cellular respiration at that stage did not rely upon an increase in mitochondrial content, but mostly on an increase of the mitochondrial OXPHOS activity.

We also assessed maximal respiration capacity by subtracting routine respiration from the maximal respiration rate obtained by uncoupling OXPHOS with FCCP (Fig 5C). We observed no significant variations of respiration capacity until 72h, when it decreased significantly.

Finally, activity of cytochrome C oxidase, reflecting the oxidative capacity of cells, was estimated by removing the maximal respiration rate obtained following TMPD addition from respiration measured after ascorbate addition (Fig 5D). Cell oxidative capacity was significantly increased at 24h and remained stable until 72h, suggesting that differentiating T2EC displayed a better capacity to perform OXPHOS than self-renewing T2EC, confirming MMP measurements.

Taken together these results strongly suggested that during erythroid differentiation, energetic demand increased at 24h, which was satisfied by a boost of OXPHOS, and that cell metabolism switches from anaerobic glycolysis toward OXPHOS during erythroid differentiation.

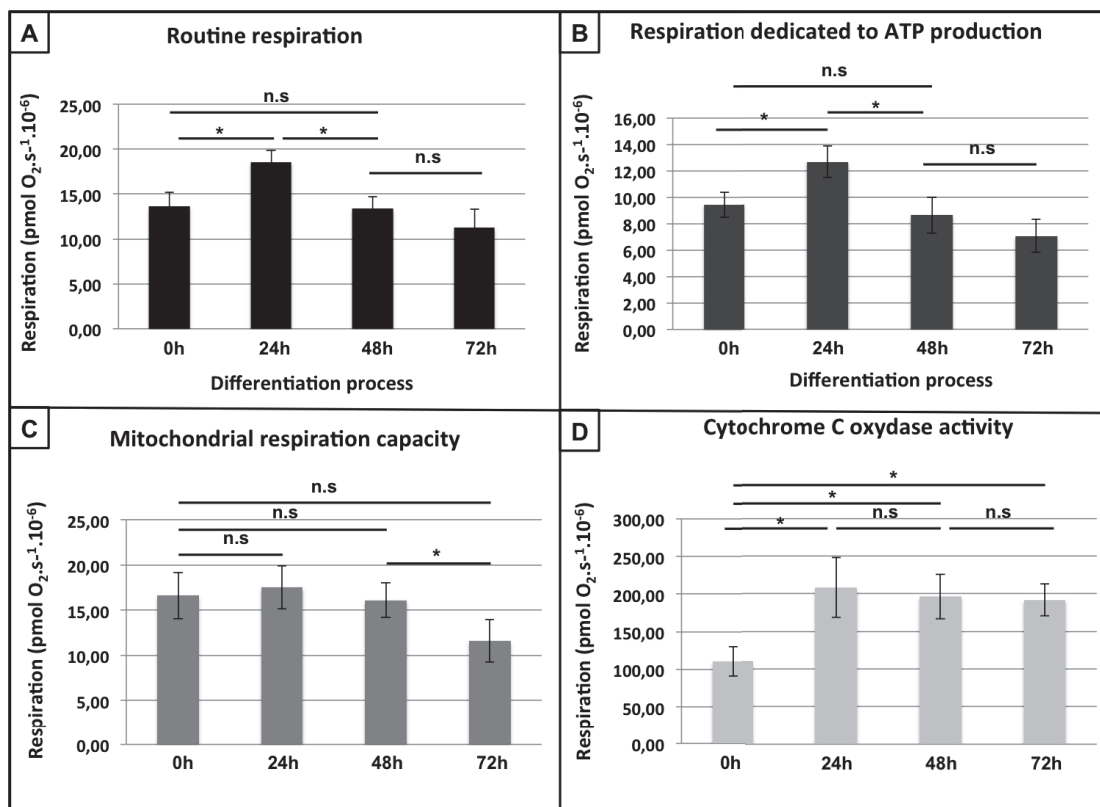


Figure 5: Dissection of cellular respiration during the differentiation process

Different respiration parameters were assessed in self-renewing T2EC (0h) and T2EC induced to differentiate for 24, 48 and 72. Routine respiration (A) corresponds to oxygen consumption without any addition. Respiration dedicated to ATP production (B) was evaluated by substracting respiration after oligomycin treatment from routine respiration. Substraction of routine respiration from maximal uncoupled respiration by FCCP reflected mitochondrial respiration capacity (C). Cytochrome C oxydase activity (D) was assessed by substracting respiration following ascorbate addition from maximal TMPD-related respiration. Bars represents means +/- S.E.M. for five independent experiments performed in duplicates. The significance of the differences (p-value < 0.05) between each time-point was computed using a paired t-test.

LDHA might be involved in glycolytic changes during erythroid differentiation

We showed that LDHA protein expression decreased strongly during the differentiation process. This might suggest that a decrease of LDHA activity is necessary for cells to commit to differentiation. Moreover, since LDHA is the most important enzyme of the anaerobic glycolytic pathway, this event might be involved in the metabolic switch observed during the erythroid differentiation process. To assess whether LDHA activity could influence the metabolic status of T2EC, we inhibited LDHA activity using Galloflavin and FX11, two known LDHA inhibitors [28, 29, 30].

We first determined non-cytotoxic Galloflavin (50 μM) and FX11 (30 μM) concentrations by titration (data not shown). Then, we assayed Galloflavin and FX11 efficiency by measuring extra-cellular lactate concentration after treating T2EC for 24h with these inhibitors. As expected, Lactate production was diminished following LDHA inhibitors treatments (Fig 6A).

We then incubated T2EC with Galloflavin or FX11, for 24, 48 and 72h and measured MMP using ISX (Fig 6B). Following inhibitor treatment MMP increased across time, mostly after FX11 treatment. Consequently, the decrease of LDHA activity was accompanied with an increase of mitochondria activity, suggesting that LDHA might indeed be involved in metabolic status changes. These results highlighted therefore a link between the rapid drop of LDHA protein level and the switch toward OXPHOS observed during the erythroid differentiation process.

Maintenance of erythroid progenitors self-renewal required LDHA expression

The important decrease of LDHA expression during the erythroid differentiation process suggested that LDHA might be important to maintain self-renewal of T2EC. We therefore inhibited LDHA activity in self-renewing T2EC (Fig 7A). Either FX11 or Galloflavin treatment slowed down significantly cell growth, suggesting that LDHA activity is important for T2EC self-renewal. We also incubated T2EC with Galloflavin or FX11 and induced them to differentiate at the same time, to assess whether reducing more strenuously LDHA activity had consequences on T2EC differentiation (Fig 7A). The results showed that a further inhibition of LDHA activity during the differentiation process did not influence erythroid differentiation.

To confirm these results we used the CRISPR-cas9 technology to inhibit *LDHA* expression, and evaluated cell growth and differentiation rate. As observed with Galloflavin and FX11 treatments, the inhibition of *LDHA* expression with the CRISPR vector, tended to slow down cell growth compared to cells transfected with an empty vector (Fig 7B). We therefore confirmed the involvement of LDHA in the self-renewal of erythroid progenitors.

Regarding its role in differentiation, we transfected T2EC with the CRISPR vector or an empty vector and induced these cells to differentiate (Fig 7B). We showed that the inhibition of *LDHA* expression did not affect T2EC differentiation, which was consistent with the fact that LDHA expression is already sharply decreased under such condition, and confirmed the experiments performed with LDHA molecular inhibitors (Galloflavin and FX11).

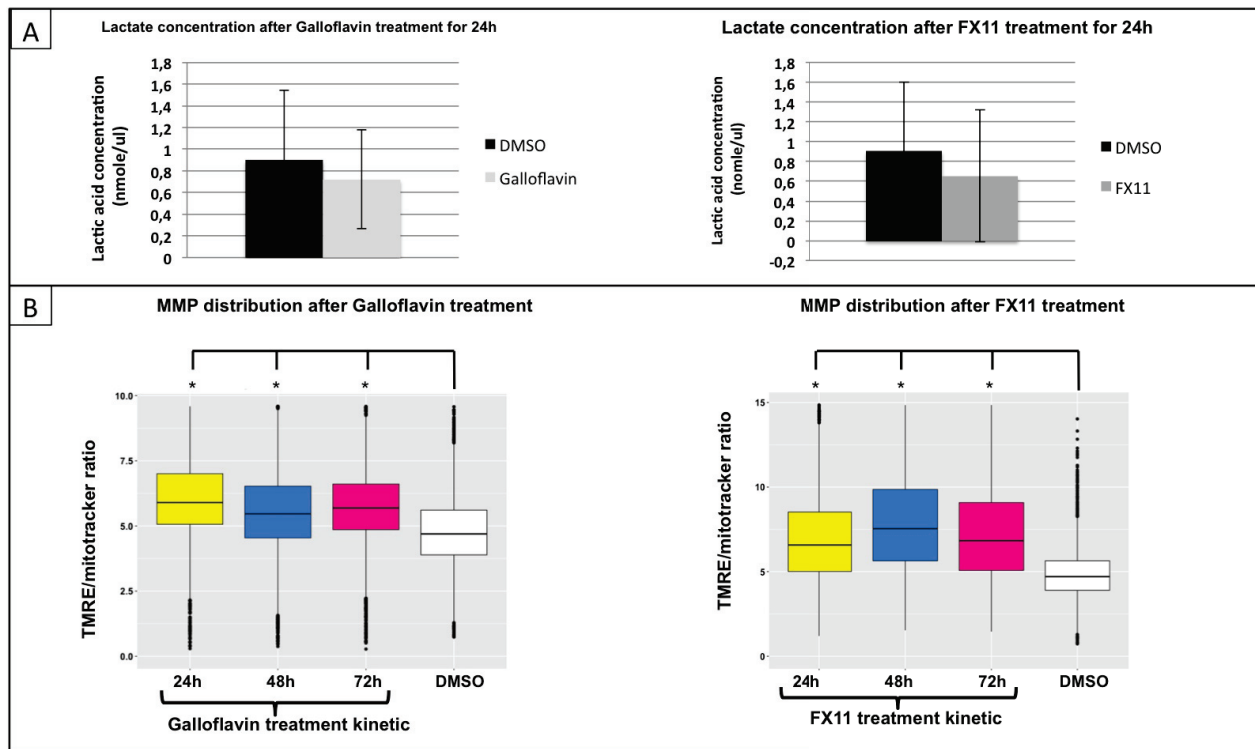


Figure 6: Mitochondrial membrane potential changes following LDHA inhibition T2EC were incubated with Galloflavin and FX11 for 24h, 48h, and 72h. (A) Lactate concentration was measured in T2EC media following incubation with Galloflavin or FX11 for 24h to confirm inhibitors efficiency. Bars represents mean +/- S.D. for two independent experiments assayed in duplicates. (B) MMP was assessed on living cells (FVS 660 dye) using TMRE dye, normalized by dividing TMRE intensity by mitotracker intensity. FCCP uncoupler was used as a negative control for MMP staining. (C) Shown are the normalized TMRE distributions. Outliers not shown. A t-test was applied to assess whether distributions were significantly different (p -value < 0.05). Data were obtained from one experiment. For FX11 treatment, 4290 cells were analyzed for 0h, 5588 cells for 24h, 5450 cells for 48h, and 3876 cells for 72h. For Galloflavin treatment, 4290 cells were analyzed for 0h, 4661 cells for 24h, 4335 cells for 48h, and 4693 cells for 72h.

Since the inhibition of LDHA activity or quantity did not affect the differentiation process, it suggests that the slight increase of its expression at 12h (Fig 1C) is not essential for cells to engage into differentiation.

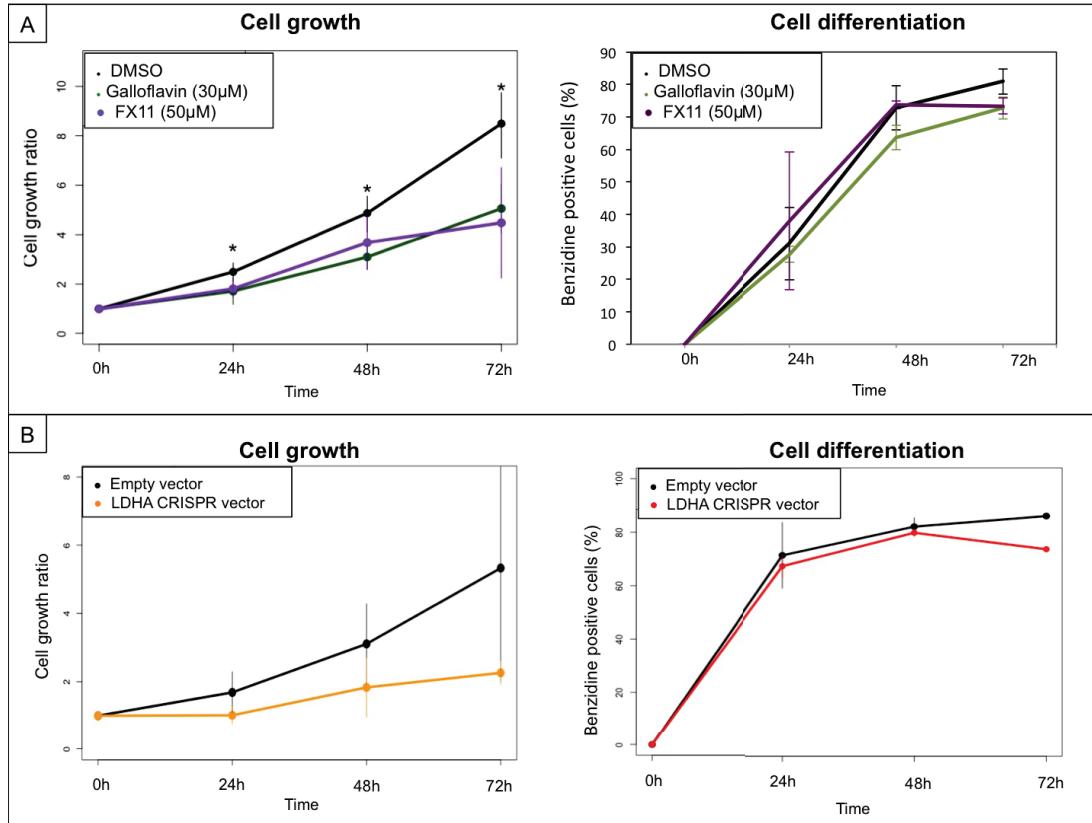


Figure 7: Consequences of LDHA inhibition on cell growth and differentiation
 (A) T2EC were incubated with Galloflavin or FX11 and were counted every 24 hours for 3 days (left panel). Results are the mean \pm S.D. calculated from four independent experiments for FX11, and five independent experiments for Galloflavin. T2EC were induced to differentiate and treated with Galloflavin or FX11. The differentiation rate was assessed every 24h using benzidine (right panel). Results are the mean \pm S.D. calculated from two independent experiments for each inhibitor. (B) T2EC were transfected with a CRISPR vector directed against *LDHA* or an empty vector, and were counted every 24 hours for 3 days (left panel). The orange curve indicates the growth of T2EC with the CRISPR vector and the black curve indicates the growth of T2EC with the empty vector. The data shown are the mean \pm S.D. calculated from three independent experiments. T2EC were transfected with the CRISPR or empty vector and induced to differentiate for 3 days. The data shown are the mean \pm S.D. calculated from two independent experiments for time-points 0h and 24h. Only one experiment was performed for the time-point 72h. Growth ratios were calculated as the cell number divided by the total cells at day 0. The significance of the difference (p-value $< 0,05$) between control growth ratios (black cruves) and LDHA inhibitors were calculated using a Wilcoxon test.

Discussion

This study aimed to determine how glycolytic metabolism changes during the erythroid differentiation process, and which role LDHA might play in erythroid progenitors metabolic status. To do so we used a multi-parametric approach including molecular analysis, metabolic parameter measures and cellular physiological insights.

To obtain a glimpse of the changes that might occur in the glycolytic metabolism during erythroid differentiation, we defined a subset of nine enzymes, involved either in glycolysis, OXPHOS or the anaerobic pathway. The protein profile of these enzymes suggested that the glycolysis flow was increased at 12h and then slowed down until 72h. Therefore, besides the increase at 12h, the anaerobic glycolysis seemed to be impaired during erythroid differentiation. On the other hand, respiratory chain activity, as measured by the protein level of chain respiratory complex subunits, seemed to increase at 12h-24h, before returning to its initial level. Lactate concentration and mitochondrial membrane potential (MMP) assessments reinforced our molecular analysis, demonstrating that self-renewing progenitors produced more lactate than differentiating cells, and that MMP increased progressively during differentiation. Concomitant with these results, cell respiration rate peaked at 24h, mostly due to ATP production, and most importantly, cells oxidative capacity was intensified in differentiating cells.

All these results suggested that (i) self-renewing progenitor rely upon anaerobic glycolysis through LDHA, (ii) during the first phase of differentiation, one observe a surge in energy demand, and (iii) erythroid differentiation is accompanied by a metabolic switch from anaerobic glycolysis to OXPHOS.

Importance of better understanding LDHA role in erythropoiesis

LDHA is essential to maintain glycolysis as it regenerates the NAD⁺ cofactors required to run glycolysis. The co-inhibition of LDHA and HK2 together can induce the death of cancer cells that depend upon anaerobic glycolysis [31]. However, the inhibition of LDHA only rather impair cell growth and tumorigenesis in the context of cancer [32, 33]. Similarly LDHA inhibition in human progenitor cells reduces cell growth, in accordance with our findings [2].

It is tempting to ask whether the reverse experiment of LDHA over-expression could affect the differentiation process. It has been shown that the constitutive co-expression of LDHA and HK2 during neuronal differentiation resulted in neuronal death [4]. Shutting down the glycolytic pathway seemed therefore essential for neuronal differentiation. Whether a similar effect could be observed in erythroid progenitors is the subject of ongoing work in our group.

More metabolism than glucose metabolism

In the present work we focused on glucose metabolism. However, this pathway is part of a broader network where glucose metabolism is highly connected with lipid and amino acid metabolisms. Interestingly, glutamine metabolism was shown to participate to erythroid differentiation [34]. Sterol metabolism has also been highlighted in erythroid cells. It was shown in our team, that Oxydosterol Cyclase (OSC), which is involved in cholesterol biosynthesis, is needed to maintain the self-renewal state of erythroid progenitors [35]. We have also

observed that sterol-related genes were early drivers of the differentiation process, as their expression changed as soon as 2h after the induction of differentiation [12]. Therefore, the orchestration of glucose and sterol metabolisms during erythroid differentiation could be a promising topic to study, that would bring new insights regarding metabolism implication in erythroid differentiation.

The generation of time-scaled metabolomic data along differentiation should be useful to acquire a more global point of view of the metabolic network and better explain energetic fluctuation and its repercussions, particularly regarding OXPHOS, the point of convergence of many metabolic pathways.

May glucose metabolism switch drive erythroid differentiation?

There is increasing evidence from the literature that cells could adapt their metabolism in accordance with their environment and energy needs [36]. Hematopoietic progenitor cells and long-term hematopoietic stem cells, for instance, are dependent upon anaerobic glycolysis and switch toward mitochondrial oxidation while they differentiate [37, 38]. On the other hand mature naïve T cells, use OXPHOS to produce energy, and shift back to the anaerobic pathway once they are activated [39]. Monocytes are also able to switch from the anaerobic glycolysis toward OXPHOS, and to switch back then to the glycolytic pathway, when they faced a microorganism [36].

In a recent review [7] the authors suggested that a switch from anaerobic glycolysis to OXPHOS might happen during erythroid differentiation. Such a view is fully supported by our results. Moreover, it is also interesting to note that a recent proteomic analysis performed on human erythroid differentiation, showed expression profiles for glucose metabolism-related enzymes (HK1, PKM, PFKP and LDHA) that are very similar to those observed in our avian erythroid cells [14].

One key question is then to what extent this metabolism change might be necessary for erythroid cells to differentiate, and if it might drive cells into the differentiation process, as it has been proposed for other cell types [40, 4, 41].

Besides the global protein level decline of enzymes involved in glycolysis and the anaerobic pathway, our analysis highlighted that some of these key enzymes slightly increased at 12h before dropping sharply, compared to OXPHOS-related enzymes that increased at 12h, stabilized until 24h and went back to their initial level. As observed by *Oburoglu and colleagues*, at some point of the erythroid differentiation process, energy demand increased, and anaerobic glycolysis as well as OXPHOS were enhanced [34]. Moreover a peak of cell respiration was reported during erythroid differentiation at the high proliferation phase [42]. It has been shown that erythroid differentiation initiation indeed was characterized by an initial increase of the proliferation rate [43, 15]. Thus, the initiation of erythroid differentiation could be associated with an increase of the proliferation rates and of the energetic demand, involving both anaerobic and aerobic pathways.

One is left with the view that at the very beginning of the differentiation process, more energy is needed and that both anaerobic and aerobic pathways would similarly participate. One could then ask if this could participate in the generation of an uncertainty phase [44] both at the metabolic (this study) and gene expression [12] level.

There are many known reasons why gene expression and metabolic variations might be correlatively involved for cell decision making in the context of differentiation. The most obvious link is through epigenetic modifications. Indeed, metabolites such as acetyl CoA generated from pyruvate and feeding OXPHOS, represent a substrate for histone acetylase [45]. On the other hand, the nicotinamide adenine dinucleotide (NAD), regenerated by various metabolic pathways, is involved in histone deacetylation. It was shown for instance that NAD concentration was proportional to the enzymatic kinetic of various histone deacetylases, including sirtuins and also other HDAC (histone deacetylases) [46]. Sirtuins are NAD-dependent histone-deacetylases that are intrinsically linked to the cellular energetic state through their dependence upon NAD [47]. Furthermore, various complexes involved in chromatin modification and remodelling are ATP consuming [48], which could participate in the peak of energy demand that we observed at the beginning of the differentiation process. Such a global remodelling could be an important aspect of the commitment process [12].

Therefore epigenetics could be the link between the concomitant gene expression and metabolism rearrangements [46, 49] which might represent a circular causal driving force pushing the cells out of their self-renewing state and into their new differentiated state.

Bibliography

- [1] Do-Yeon Kim, Inmoo Rhee, and Jihye Paik. Metabolic circuits in neural stem cells. *Cellular and Molecular Life Sciences*, 71(21):4221–4241, Nov 2014.
- [2] William J. Israelsen, Dongjun Lee, Vionnie W.C. Yu, Nathaniel T. Jeanson, Clary B. Clish, Lewis C. Cantley, Matthew G. Vander Heiden, and David T. Scadden. Differential dependence on aerobic glycolysis in normal and malignant hematopoietic stem and progenitor cells to sustain daughter cell production. *Blood*, 122(21):793–793, 2013.
- [3] Laura C. Shum, Noelle S. White, Bradley N. Mills, Karen L. de Mesy Bentley, and Roman A. Eliseev. Energy metabolism in mesenchymal stem cells during osteogenic differentiation. *Stem Cells and Development*, 25(2):114–122, October 2015.
- [4] Xinde Zheng, Leah Boyer, Mingji Jin, Jerome Mertens, Yongsung Kim, Li Ma, Li Ma, Michael Hamm, Fred H. Gage, and Tony Hunter. Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *eLife*, 5:e13374, 2016.
- [5] Imelda T Sandoval, Richard Glenn C Delacruz, Braden N Miller, Shauna Hill, Kristofor A Olson, Ana E Gabriel, Kevin Boyd, Christeena Satterfield, Holly Van Remmen, Jared Rutter, and David A Jones. A metabolic switch controls intestinal differentiation downstream of *Adenomatous polyposis coli* (*apc*). *eLife*, 6:e22706, apr 2017.
- [6] Folmes Clifford D. L. and Terzic Andre. Metabolic determinants of embryonic development and stem cell fate. *Reproduction, Fertility and Development*, 27(1):82–88, December 2014.
- [7] M. H. Baron, J. Isern, and S. T. Fraser. The embryonic origins of erythropoiesis in mammals. *Blood*, 119(21):4828–37, 2012.
- [8] Q. Xue and E. S. Yeung. Variability of intracellular lactate dehydrogenase isoenzymes in single human erythrocytes. *Anal Chem*, 66(7):1175–8, 1994.
- [9] S. S. Zail and A. K. Vandenhoeck. Lactate-dehydrogenase isoenzymes of human erythrocyte-membranes. *Clinica Chimica Acta*, 79(1):15–19, 1977.
- [10] H. Sandoval, P. Thiagarajan, S. K. Dasgupta, A. Schumacher, J. T. Prchal, M. Chen, and J. Wang. Essential role for nix in autophagic maturation of erythroid cells. *Nature*, 454(7201):232–U66, 2008.
- [11] A. Stier, P. Bize, Q. Schull, J. Zoll, F. Singh, B. Geny, F. Gros, C. Royer, S. Massemin, and F. Criscuolo. Avian erythrocytes have functional mitochondria, opening novel perspectives for birds as animal models in the study of ageing. *Frontiers in Zoology*, 10, 2013.
- [12] A. Richard, L. Boullu, U. Herbach, A. Bonnafoux, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J. J. Kupiec, T. Espinasse, S. Gonin-Giraud, and O. Gandrillon. Single-cell-based analysis highlights a surge in cell-to-cell

- molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol*, 14(12):e1002585, 2016.
- [13] J. Isern, Z. He, S. T. Fraser, S. Nowotschin, A. Ferrer-Vaquer, R. Moore, A. K. Hadjantonakis, V. Schulz, D. Tuck, P. G. Gallagher, and M. H. Baron. Single-lineage transcriptome analysis reveals key regulatory pathways in primitive erythroid progenitors in the mouse embryo. *Blood*, 117(18):4924–34, 2011.
- [14] E. F. Gautier, S. Ducamp, M. Leduc, V. Salnot, F. Guillonneau, M. Dussiot, J. Hale, M. C. Giarratana, A. Raimbault, L. Douay, C. Lacombe, N. Mohandas, F. Verdier, Y. Zermati, and P. Mayeux. Comprehensive proteomic analysis of human erythropoiesis. *Cell Rep*, 16(5):1470–84, 2016.
- [15] Olivier Gandrillon, Uwe Schmidt, Hartmut Beug, and Jacques Samarut. Tgf- β cooperates with tgf- α to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *The EMBO Journal*, 18(10):2764–2781, 1999.
- [16] O. Gandrillon and J. Samarut. Role of the different rar isoforms in controlling the erythrocytic differentiation sequence. interference with the v-erba and p135gag-myb-ets nuclear oncogenes. *Oncogene*, 16(5):563–74, 1998.
- [17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [18] J. H. Kim, S. R. Lee, L. H. Li, H. J. Park, J. H. Park, K. Y. Lee, M. K. Kim, B. A. Shin, and S. Y. Choi. High cleavage efficiency of a 2a peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*, 6(4):e18556, 2011.
- [19] M. Geiser, R. Cebe, D. Drewello, and R. Schmitz. Integration of pcr fragments at any specific site within cloning vectors without the use of restriction enzymes and dna ligase. *Biotechniques*, 31(1):88–+, 2001.
- [20] J. E. Wilson. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *Journal of Experimental Biology*, 206(12):2049–2057, 2003.
- [21] E.C. Cheung Mor, I. and K.H. Vousden. ontrol of glycolysis through regulation of pfk1: old friends and recent additions. *Cold Spring Harb. Symp. Quant. Biol.*, 76:211–216, 2011.
- [22] M. S. Jurica, A. Mesecar, P. J. Heath, W. X. Shi, T. Nowak, and B. L. Stoddard. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure*, 6(2):195–210, 1998.
- [23] M. S. Patel and T. E. Roche. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *FASEB J*, 4(14):3224–33, 1990.
- [24] M. C. Sugden and M. J. Holness. Recent advances in mechanisms regulating glucose oxidation at the level of the pyruvate dehydrogenase complex by pdks. *American Journal of Physiology-Endocrinology and Metabolism*, 284(5):E855–E862, 2003.

- [25] S. K. Chan and E. Margoliash. Amino acid sequence of chicken heart cytochrome c. *J Biol Chem*, 241(2):507–15, 1966.
- [26] M. Nishikimi, Y. Hosokawa, H. Toda, H. Suzuki, and T. Ozawa. The primary structure of human rieske iron-sulfur protein of mitochondrial cytochrome bc1 complex deduced from cDNA analysis. *Biochemistry International*, 20(1):155–160, 1990.
- [27] I. G. Johnston, B. Gaal, R. P. das Neves, T. Enver, F. J. Iborra, and N. S. Jones. Mitochondrial variability as a source of extrinsic cellular noise. *Plos Computational Biology*, 8(3), 2012.
- [28] M. Manerba, M. Vettraino, L. Fiume, G. Di Stefano, A. Sartini, E. Giacomini, R. Buonfiglio, M. Roberti, and M. Recanatini. Galloflavin (cas 568-80-9): a novel inhibitor of lactate dehydrogenase. *ChemMedChem*, 7(2):311–7, 2012.
- [29] Lorraine M. Deck, Robert E. Royer, Brian B. Chamblee, Valerie M. Hernandez, Richard R. Malone, Jose E. Torres, Lucy A. Hunsaker, Robert C. Piper, Michael T. Makler, and David L. Vander Jagt. Selective inhibitors of human lactate dehydrogenases and lactate dehydrogenase from the malarial parasite *Plasmodium falciparum*. *Journal of Medicinal Chemistry*, 41(20):3879–3887, 1998.
- [30] A. Le, C. R. Cooper, A. M. Gouw, R. Dinavahi, A. Maitra, L. M. Deck, R. E. Royer, D. L. Vander Jagt, G. L. Semenza, and C. V. Dang. Inhibition of lactate dehydrogenase induces oxidative stress and inhibits tumor progression. *Proc Natl Acad Sci U S A*, 107(5):2037–42, 2010.
- [31] L. Galluzzi, O. Kepp, M. G. Vander Heiden, and G. Kroemer. Metabolic targets for cancer therapy (vol 12, pg 829, 2013). *Nature Reviews Drug Discovery*, 12(12):965–965, 2013.
- [32] Yongjie Zhang, Xia Zhang, Xi Wang, Lu Gan, Guanzhen Yu, Ying Chen, Ke Liu, Ping Li, Jun Pan, Jiejun Wang, and Shukui Qin. Inhibition of ldh-a by lentivirus-mediated small interfering RNA suppresses intestinal-type gastric cancer tumorigenicity through the downregulation of oct4. *Cancer Letters*, 321(1):45 – 54, 2012.
- [33] Eric J. Rellinger, Brian T. Craig, Alexandra L. Alvarez, Haley L. Dusek, Kwang W. Kim, Jingbo Qiao, and Dai H. Chung. Fx11 inhibits aerobic glycolysis and growth of neuroblastoma cells. *Surgery*, 161(3):747 – 752, 2017.
- [34] L. Oburoglu, S. Tardito, V. Fritz, S. C. de Barros, P. Merida, M. Craveiro, J. Mamede, G. Cretenet, C. Mongellaz, X. An, D. Klysz, J. Touhami, M. Boyer-Clavel, J. L. Battini, V. Dardalhon, V. S. Zimmermann, N. Mohandas, E. Gottlieb, M. Sitbon, S. Kinet, and N. Taylor. Glucose and glutamine metabolism regulate human hematopoietic stem cell lineage specification. *Cell Stem Cell*, 15(2):169–84, 2014.
- [35] C. Mejia-Pous, F. Damiola, and O. Gandrillon. Cholesterol synthesis-related enzyme oxidosqualene cyclase is required to maintain self-renewal in primary erythroid progenitors. *Cell Proliferation*, 44(5):441–452, 2011.

- [36] W. Jones and K. Bianchi. Aerobic glycolysis: beyond proliferation. *Front Immunol*, 6:227, 2015.
- [37] Y. H. Wang, W. J. Israelsen, D. J. Lee, V. W. C. Yu, N. T. Jeanson, C. B. Clish, L. C. Cantley, M. G. V. Heiden, and D. T. Scadden. Cell-state-specific metabolic dependency in hematopoiesis and leukemogenesis. *Cell*, 158(6):1309–1323, 2014.
- [38] N. Shyh-Chang and H. H. Ng. The metabolic programming of stem cells. *Genes Development*, 31(4):336–346, 2017.
- [39] K. A. Frauwirth and C. B. Thompson. Regulation of t lymphocyte metabolism. *Journal of Immunology*, 172(8):4661–4665, 2004.
- [40] A. D. Panopoulos, O. Yanes, S. Ruiz, Y. S. Kida, D. Diep, R. Tautenhahn, A. Herrerias, E. M. Batchelder, N. Plongthongkum, M. Lutz, W. T. Berggren, K. Zhang, R. M. Evans, G. Siuzdak, and J. C. Izpisua Belmonte. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res*, 22(1):168–77, 2012.
- [41] M. J. Rodriguez-Colman, M. Schewe, M. Meerlo, E. Stigter, J. Gerrits, M. Pras-Raves, A. Sacchetti, M. Hornsveld, K. C. Oost, H. J. Snippert, N. Verhoeven-Duif, R. Fodde, and B. M. T. Burgering. Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature*, 543(7645):424–+, 2017.
- [42] S. M. Browne, H. Daud, W. G. Murphy, and M. Al-Rubeai. Measuring dissolved oxygen to track erythroid differentiation of hematopoietic progenitor cells in culture. *Journal of Biotechnology*, 187:135–138, 2014.
- [43] EW Müllner, H Dolznig, and H Beug. Cell cycle regulation and erythroid differentiation. *Curr Top Microbiol Immunol*, pages 212:175–94, 1996.
- [44] A. Moussy, J. Cosette, R. Parmentier, C. da Silva, G. Corre, A. Richard, O. Gandrillon, D. Stockholm, and A. Pldi. Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *Plos Biology*, in the press, 2017.
- [45] A. Moussaieff, M. Rouleau, D. Kitsberg, M. Cohen, G. Levy, D. Barasch, A. Nemirovski, S. Shen-Orr, I. Laevsky, M. Amit, D. Bomze, B. Elena-Herrmann, T. Scherf, M. Nissim-Rafinia, S. Kempa, J. Itskovitz-Eldor, E. Meshorer, D. Aberdam, and Y. Nahmias. Glycolysis-mediated changes in acetyl-coa and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab*, 21(3):392–402, 2015.
- [46] M. A. Reid, Z. Dai, and J. W. Locasale. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat Cell Biol*, 19(11):1298–1306, 2017.
- [47] S. Imai, F. B. Johnson, R. A. Marciniak, M. McVey, P. U. Park, and L. Guarente. Sir2: an nad-dependent histone deacetylase that connects chromatin silencing, metabolism, and aging. *Cold Spring Harb Symp Quant Biol*, 65:297–302, 2000.

- [48] S. K. Hota and B. G. Bruneau. Atp-dependent chromatin remodeling during mammalian development. *Development*, 143(16):2882–97, 2016.
- [49] C. D. L. Folmes and A. Terzic. Energy metabolism in the acquisition and maintenance of stemness. *Seminars in Cell Developmental Biology*, 52:68–75, 2016.

4 Discussion et perspectives

4.1 Mise en évidence expérimentale d'un scénario commun à différents processus de différenciation

Au début de ma thèse, je me suis intéressée à la variabilité de l'expression génique dans la différenciation de progéniteurs érythrocytaires aviaires primaires, à l'échelle de la cellule unique. Nous avons tout d'abord sélectionné par un process multi-étapes, un groupe de 92 gènes significativement sur-exprimés, sous-exprimés ou invariants, entre des cellules en auto-renouvellement et des cellules induites à se différencier pendant 48h. L'expression des ces 92 gènes a ensuite été analysée par RT-qPCR dans des cellules isolées à différents temps de différenciation. Pour chaque temps de cinétique étudié, nous avons inféré un réseau de corrélation de gènes, dont on a pu noter que la densité variait selon le temps de différenciation. Nous avons alors observé une chute significative du nombre de gènes corrélés à 8 et 24h de différenciation, par rapport au temps 0h. Par ailleurs, l'utilisation de nouvelles métriques, telle que l'entropie, nous ont permis d'évaluer la variabilité d'expression génique intercellulaire. Nous avons ainsi pu montrer que la variabilité de l'expression des gènes atteint un niveau maximal à 8-24h de différenciation, simultanément à une chute du nombre de corrélations entre gènes. Ces deux évènements précèdent l'engagement irréversible des cellules dans le processus de différenciation érythrocytaire identifié entre 24h et 48h (figure 15). Ces résultats ont ainsi apporté un soutien expérimental à l'hypothèse théorique soulevée précédemment par plusieurs auteurs [24, 25, 26], suggérant qu'une augmentation de la variabilité de l'expression génique serait essentielle à la prise de décision cellulaire, dans le choix entre auto-renouvellement et différenciation.

Depuis la publication de notre article, au moins trois études ont conforté nos résultats [185, 186, 88].

Premièrement, les résultats obtenus par *Stefan Semrau et al.* montrent une très grande similarité avec les nôtres. Il est notamment démontré dans cette analyse qu'une augmentation de la variabilité de l'expression génique, marquant la sortie de l'état de pluripotence, précède l'engagement progressif mais non linéaire, des cellules dans un lignage spécifique [186]. Cette étude porte plus exactement sur la différenciation des cellules souches embryonnaires

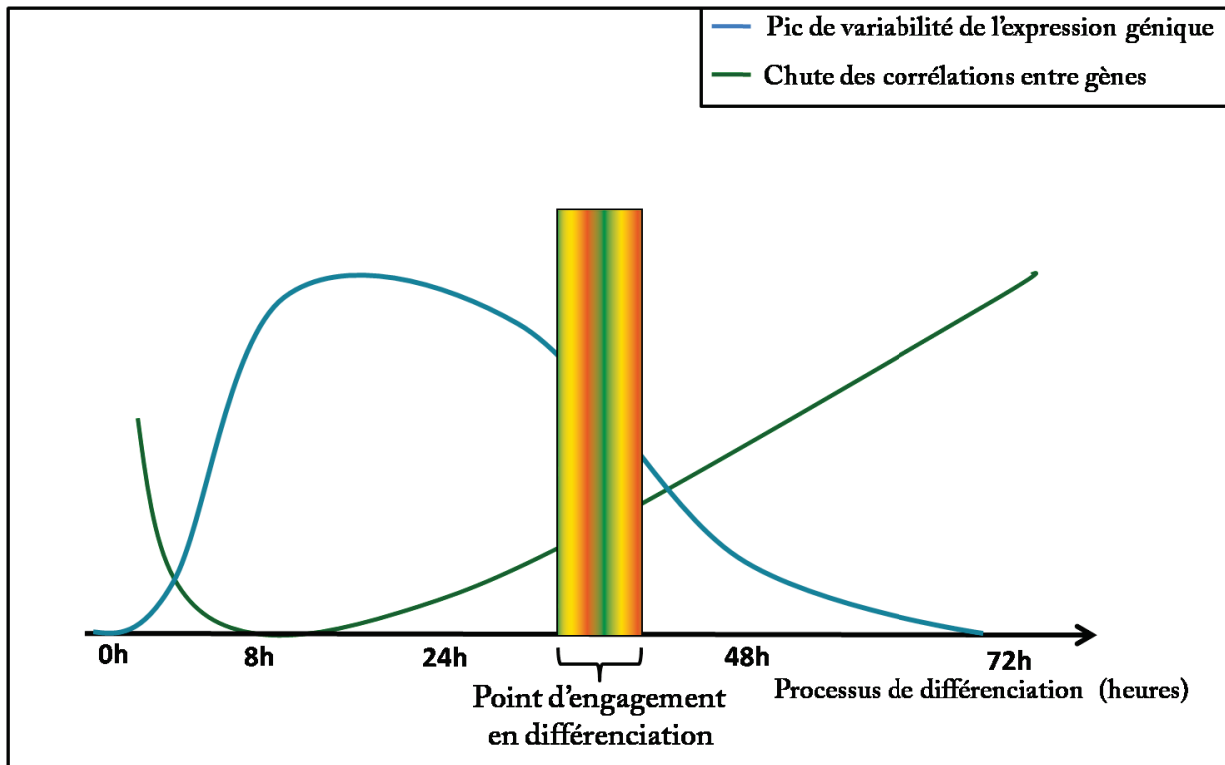


FIGURE 15 – Orchestration des événements moléculaires au cours de la différenciation érythrocytaire

Ce schéma représente la conclusion principale de notre analyse de l'expression génique en cellules uniques au cours de la différenciation. Un pic de la variabilité inter-cellulaire de l'expression des gènes, parallèlement à une chute des corrélations entre gènes, précèdent le point d'engagement irréversible des cellules dans la différenciation érythrocytaire.

de souris (mESC) en ectoderme et endoderme extra-embryonnaires (XEN), par traitement à l'acide rétinoïque. Le transcriptome de mESC a été analysé au cours de la différenciation, à l'échelle de la cellule unique par SCRBS-Seq (Single Cell RNA Barcoding and Sequencing method) [187]. L'analyse de ces données par t-SNE a permis de confirmer que les cellules commençaient à diverger en terme de profil d'expression génique après 12h. Tout comme nous l'avons démontré dans notre système, ils ont alors pu mettre en évidence une augmentation de la variabilité de l'expression génique, mesurée par fano factor (variance divisée par la moyenne), à partir de 12h de différenciation. Cependant dans leur système cellulaire cette variabilité continue ensuite d'augmenter jusqu'à 72h et diminue à 96h, où deux profils d'expression génique sont bien distincts et robustes. Ce résultat suggère alors que l'augmentation

de la variabilité, conditionne le passage d'un état stable, l'état pluripotent, vers un autre état stable, l'état différencié. De plus, une analyse par ACP (analyse en composant principales) leur a permis de montrer que l'augmentation de la variabilité d'expression génique à 12h, était suivie d'une bifurcation entre deux profils d'expression géniques à partir de 24h. Bien que ces deux profils se confondent et expriment des marqueurs de chacun des deux lignages attendus, entre 24h et 48h, ils divergent ensuite graduellement et sont consolidés à la fin de la différenciation (96h). L'augmentation de la variabilité d'expression génique permettrait donc d'ouvrir la fenêtre des possibles, générant ainsi une phase instable dans laquelle les cellules ne sont plus dans leur état de pluripotence, mais n'ont pas atteint non plus l'état différencié. Enfin, il a pu être observé qu'après 48h de différenciation, l'expression de facteurs de transcriptions très spécifiques du lignage endoderme pour un cluster, et du lignage ectoderme extra-embryonnaire, pour l'autre cluster augmentent fortement. Ainsi, bien que le modèle de différenciation et l'outils de mesure de la variabilité soient différents, les résultats obtenus dans cette étude confortent nos travaux.

Par ailleurs les travaux de *Patrick S. Stumpf et al.*, ont également mis en évidence une augmentation de la variabilité de l'expression génique au cours de la différenciation [88]. Cette étude s'intéresse à la différenciation de deux lignées de cellules ESC murines (R1 et E14) en progéniteurs neuronaux (NPC). L'expression d'un groupe de gènes préalablement sélectionnés a été analysé à l'échelle de la cellule unique par RT-qPCR, au cours de la différenciation. L'analyse de ces données avec divers outils statistiques, comme l'ACP ou les k-means, a permis de mettre en évidence trois stades cellulaires au cours de la différenciation neuronale : l'état initial pluripotent, l'état épiblastique pluripotent et le stade NPC. Les auteurs ont alors émis l'hypothèse que la variabilité de l'expression génique pourrait augmenter entre chaque état cellulaire. Or, cette variabilité, tout d'abord évaluée par une mesure multivariée de la dispersion (mesure de la variabilité d'expression de chaque gène, prenant en compte les covariances entre gènes) [88], augmente significativement, entre l'état initial, pluripotent, des cellules mESC, et l'état le plus différencié, autrement dit le stade NPC, pour les deux lignées cellulaires. À la différence de notre analyse, la variabilité mesurée ici ne montre pas un pic transitoire de la variabilité de l'expression génique mais plutôt une augmentation graduelle. Ces résultats ont ensuite été vérifiés par l'utilisation d'une autre métrique, l'entropie de Shan-

non, que nous avons nous même utilisée. La variabilité de l'expression génique, mesurée par l'entropie de Shannon, montre alors une augmentation immédiatement après l'induction en différenciation, dans les deux lignées de cellules ESC. En accord avec nos résultats, la valeur d'entropie augmente de façon transitoire jusqu'à 24h de différenciation et rediminue ensuite pour la lignée E14. En revanche, pour la lignée R1, l'entropie augmente graduellement jusqu'à la fin de la différenciation. Bien que la raison de cette différence ne soit pas clairement identifiée, cette distinction suggère que le contexte génétique initial pourrait influencer la cinétique de la différenciation en cellules NPC. En revanche, quelque soit le contexte génétique initial ou l'outil de mesure de variabilité utilisé, l'initiation de la différenciation est inévitablement caractérisée par une augmentation de la variabilité de l'expression génique, marquant la sortie des cellules de leur état de pluripotence, déstabilisé en réponse à des stimuli extra-cellulaires. Ainsi, suite aux mêmes stimuli de différenciation toutes les cellules ne convergent pas simultanément vers un état intermédiaire définit (épiblaste), comme les résultats à l'échelle de la population menait à penser. Chaque cellule arbore plutôt sa propre dynamique et sa propre trajectoire au cours du processus de différenciation. Cette hypothèse est d'autant plus confirmée dans l'article de *Alice Moussy et. al.*, à laquelle j'ai participé, par l'analyse morphologique en time-lapse couplée à l'expression de l'antigène CD133 dans les cellules hématopoïétiques CD34+ [88]. Ces cellules ont été isolées après 48h de différenciation, en fonction de leur expression de CD133 pouvant se séparer en trois catégories : CD133high, CD133medium, CD133low. Cet antigène est très peu exprimé par la lignée myéloïde, émergeant de la différenciation des cellules CD34+ ; il est en revanche fortement exprimé par les cellules CD34+ pluripotentes. Les cellules CD133medium se trouvent donc dans un état intermédiaire entre ces deux états stables, et présentent un comportement dynamique dit "hésitant" qui leur est propre, autant du point de vue phénotypique que transcriptionnel. En effet le profil transcriptionnel de ces cellules n'a pas de signature particulière et est très différent des cellules CD133high et CD133low. De plus les cellules CD133medium oscillent de façon fréquente entre deux morphologies, caractéristiques des cellules CD34+ pluripotentes et des progéniteurs myéloïdes. Ces résultats suggèrent ainsi qu'après stimulation, les CD34+ pourraient entrer dans une phase d'"hésitation" instable, de durée variable, caractérisé par un profil d'expression génique particulier et une morphologie stable. Ainsi, l'ensemble de ces

résultats suggèrent que le processus de différenciation, défini comme le passage des cellules d'un état stable vers un nouvel état stable, est intrinsèquement stochastique.

Enfin, les travaux de *Mitra Mojtahedi et al.*, portant sur un autre modèle de différenciation, conduisent aux mêmes conclusions proposées dans notre étude, grâce à l'utilisation de méthodes d'analyses et d'outils statistiques différents. Nos résultats convergeant dans la même direction, nous avons alors fait le choix, avec les auteurs de cette étude, de publier conjointement nos deux articles [185, 95]. Le modèle de différenciation utilisé dans l'analyse de *Mitra Mojtahedi et al.* correspond à une lignée de précurseurs hématopoïétiques multipotents murins (EML), pouvant s'engager dans le lignage érythrocytaire en présence d'érythropoïétine (Epo), ou dans le lignage myéloïde sous la combinaison de granulocyte macrophage colony-stimulating factor (GM-CSF) et d'interleukine 3 (IL3) [185]. L'analyse de données d'expression génique mesurée à l'échelle de la cellule unique, au cours du processus de différenciation, leur a notamment permis de mettre en lumière une transition stochastique cruciale entre l'état de multipotence et l'état différencié. Cette transition est caractérisée par une augmentation de la variabilité de l'expression génique, mesurée par le coefficient de variation (CV) et le taux de corrélation entre chaque paire de cellule ($R(\text{cell } k, \text{cell } l)$), simultanément à une augmentation de la corrélation entre les gènes ($R(g_i, g_j)$). De façon intéressante, cette analyse utilise l'indice I_c , correspondant au ratio des corrélations entre gènes sur les corrélations entre cellules afin d'identifier une transition critique dans l'expression des gènes, pouvant marquer par exemple un événement d'engagement en différenciation. Dans les données de *Mitra Mojtahedi et al.* l'indice I_c augmente après trois jours de différenciation, ce qui correspond à la période où les cellules commencent à se séparer en deux profils transcriptionnels spécifiques des lignages érythroïde et myéloïde. L'augmentation de cet indice signe alors une transition cruciale au troisième jour de différenciation des précurseurs hématopoïétiques multipotents, précédant l'engagement des cellules dans des lignages bien spécifiques. Notre analyse de l'expression génique en cellule unique au cours du processus de différenciation érythrocytaire nous avait également permis d'identifier une telle phase critique. En effet, l'indice I_c décrit par *Mitra Mojtahedi et al.* est très similaire au DNB (dynamical network biomarker) calculé dans notre étude [95]. Le DNB a été créé dans le but d'identifier des signaux précoces annonçant une transition critique dans l'expression des gènes [188]. L'indice

I du DNB, rend compte de trois paramètres : le coefficient de variation, la moyenne des corrélations entre chaque paire de gènes faisant partie du DNB, et la moyenne des corrélations entre chaque gène du DNB et les gènes n'en faisant pas partie. Nous avons alors démontré que l'indice I du DNB augmentait fortement à 8h de différenciation, avant de rediminuer. Ce signal apparaît donc simultanément à l'augmentation de la variabilité de l'expression génique et de la chute des corrélations entre gènes (contrairement aux résultats obtenus dans l'étude de *Mitra Mojtahedi et. al*, montrant une augmentation des corrélations.), suggérant que les cellules traversent entre 8h et 24h une phase critique instable, avant de s'engager de façon irréversible dans la différenciation.

Ainsi, bien que l'utilisation d'outils statistiques et de méthodes d'analyses distincts et variés, des résultats très similaires aux nôtres ont été obtenus à partir de modèles de différenciation de différents types cellulaires et différentes espèces. Les conclusions et hypothèses mises en avant dans nos travaux semblent donc pouvoir être généralisées à plusieurs processus de différenciation. L'ensemble de nos résultats et des études citées tout au long de cette sous-partie, pourrait ainsi permettre d'établir un scénario commun à tous les processus de différenciation. Un tel scénario pourrait alors s'organiser de la phase suivante : des stimuli extra-cellulaires déclencheraient l'augmentation de la variabilité de l'expression génique, permettant aux cellules de s'"échapper" de leur état d'auto-renouvellement et d'explorer le champs des possibles dans l'espace de l'expression génique. Cette phase instable pourrait alors être qualifiée d'état d'"hésitation", précédant l'apparition de profils d'expression génique s'approchant de l'état différencié, et l'engagement irréversible des cellules dans un lignage particulier.

4.2 Mieux comprendre le rôle de la variabilité de l'expression génique dans la différenciation

La majorité des études s'intéressant à la variabilité de l'expression des gènes au cours de la différenciation, ne permet aujourd'hui que d'apporter une observation expérimentale de l'augmentation de cette variabilité. Ainsi, ces résultats, seuls, ne permettent pas de montrer que la variabilité de l'expression génique joue un véritable rôle fonctionnel dans l'engagement

des cellules en différenciation. Déterminer le lien causal de l'augmentation de la variabilité sur le processus de différenciation représente une perspective intéressante de mes premiers travaux de thèse. Cependant, étudier expérimentalement le rôle de la variabilité de l'expression des gènes dans les processus biologiques, nécessite tout d'abord d'être capable de la moduler. Les travaux de *Roy Dar et al.* ayant consisté à tester l'effet de 1600 drogues thérapeutiques sur la variabilité d'expression génique, a permis d'identifier quelques composés ayant la capacité de modifier la SEG sans influencer significativement le taux d'expression des gènes [69]. La disponibilité de tels composés et quelques échanges avec cette équipe ont alors permis de débiter un nouveau projet dans l'équipe, actuellement mené par Anissa Guillemin. Ce projet consiste ainsi à déterminer le rôle fonctionnel de la stochasticité de l'expression des gènes (SEG) dans le processus de différenciation érythrocytaire. Au cours de cette étude, Anissa a pu montrer que des drogues diminuant la SEG, diminuent la différenciation, alors qu'une drogue augmentant la SEG, augmente la différenciation des progéniteurs érythrocytaires. Ces résultats viennent donc conforter les conclusions précédemment obtenues dans la première partie de ma thèse, et permettent d'établir un premier lien fonctionnel entre la variabilité de l'expression génique et le processus de différenciation.

Par ailleurs, dans un contexte physiologique normal, de plus en plus de travaux, dont ceux réalisés au cours de ma thèse et par Anissa Guillemin, suggèrent que la variabilité de l'expression génique participe au choix entre auto-renouvellement et différenciation. Or, il est essentiel à l'échelle de l'organisme, que la balance entre auto-renouvellement et différenciation soit entretenue. En effet, le stock de cellules souches doit être maintenu afin de régénérer des cellules matures et fonctionnelles. De plus, le contrôle du taux de cellules matures est aussi essentiel pour que l'organisme puisse remplir pleinement ses fonctions. Ainsi, un dérèglement de cette balance, entre cellules immatures et cellules matures, engendrera un défaut du tissu concerné et affectera le fonctionnement de l'organisme, pouvant contribuer au développement de certaines pathologies. La variabilité de l'expression génique jouant probablement un rôle dans la maintenance de l'équilibre entre auto-renouvellement et différenciation, il pourrait alors être intéressant de la mesurer dans un contexte pathologique dans lequel la différenciation est altérée. En effet, analyser cette variabilité dans un contexte pathologique, permettrait, d'une part, de mieux caractériser son rôle dans le processus de différenciation,

et d'autre part, de comprendre son implication dans le développement des pathologies liées aux défauts de différenciation. Cette perspective a récemment été initiée, en collaboration avec Pierre Sujobert et Adriana Plesa, du Laboratoire d'Hématologie du Centre hospitalier Lyon Sud, dans le cadre de leur travail de recherche sur les syndromes myélodysplasiques.

Enfin, il a été suggéré que la variabilité de l'expression génique pourrait apporter une flexibilité aux populations cellulaires, telle qu'elles puissent s'adapter à de nouveaux environnements ou de nouveaux signaux, comme cela a pu être observé dans la résistance de certaines bactéries aux antibiotiques [189]. Cette variabilité pourrait donc être le reflet de la capacité des cellules à pouvoir moduler leur réseau de régulation de gène de façon à trouver une solution, parmi de nombreuses possibilités, leur permettant de survivre face à un environnement hostile. Cette théorie a notamment été mise en avant dans l'analyse menée par Elez Braun en 2015 [190]. Il est montré dans cette étude qu'une population clonale de levures est capable de reconstruire tout son réseau de régulation de gènes face à une situation à laquelle elle n'a jamais fait face auparavant, et à laquelle elle n'aurait pas pu être naturellement confrontée. D'autre part, toutes les cellules répondent de façon différente à cette situation, et chaque population clonale présente des comportements distincts, bien qu'elles soient exposées aux mêmes conditions. Ce résultat suggère donc que l'état leur permettant de survivre face à cette situation inconnue, ne serait pas "pré-enregistré". Il serait alors nécessaire de lever les contraintes épigénétiques afin d'explorer l'espace des gènes à la recherche de cet état. D'autre part, ces résultats soutiennent l'hypothèse selon laquelle les processus biologiques évoluent entre robustesse et flexibilité. En effet, dans l'espace des gènes, un cadre semblerait nécessaire pour empêcher les cellules de s'échapper de leur état vital et fonctionnel. La variabilité de l'expression génique, quant à elle, permettrait aux cellules d'en échapper, pour faire face autant aux situations totalement inconnues, qu'à de simples changements environnementaux physiologiques. Les contraintes imposant l'organisation de l'expression des gènes, telle qu'elle existe dans un système stable, est représentées par l'épigénome et pourrait alors être héréditaires. En revanche, lors de la levée de ces contraintes en réponse à des signaux environnementaux, la probabilité que deux cellules empruntent le même chemin dans l'espace des gènes, pourrait être nulle. Ce scénario peut alors être transposé dans tout autre processus biologique, tel que la différenciation, comme le mettent notamment en lumière nos résultats.

En effet, en réponse aux facteurs extra-cellulaires de différenciation, une partie des contraintes épigénétiques pourrait être levée, offrant ainsi la capacité aux cellules d'explorer de façon stochastique toutes les possibilités s'offrant à elles. Cette phase exploratoire pourrait donc être associée à l'augmentation de la variabilité de l'expression génique au début du processus de différenciation. Pendant cette période, chaque cellule pourrait alors, de façon figurée, démarrer une marche aléatoire dans l'espace des gènes, les menant ainsi à emprunter leur propre chemin, tel qu'illustré par la figure 5 de l'introduction de ce manuscrit. Cette traversée stochastique pourrait donc créer un état d'instabilité, dans lequel l'expression des gènes perdrait de sa coordination, se reflétant donc par une diminution de la corrélation entre l'expression de différents gènes. Dans notre système cellulaire, en conditions de culture contrôlées, chacun des chemins pouvant être empruntés pourront alors aboutir à l'état d'auto-renouvellement, à l'engagement en différenciation ou à la mort. Dans d'autres systèmes de différenciation, les contraintes épigénétiques seront différentes et pourront par exemple ouvrir dans l'espace des possibles, différentes voies conduisant à la différenciation en divers lignages [186]. La variabilité de l'expression génique semble donc apporter la possibilité aux cellules de se différencier, à travers une phase transitoire d'exploration de l'espace des gènes, sans laquelle les cellules ne pourrait s'échapper de leur état initial.

4.3 Remaniements du métabolisme glycolytique au cours de la différenciation érythrocytaire

Le rôle de LDHA dans l'auto-renouvellement des progéniteurs a été suggéré dans la première partie de ma thèse. À l'échelle de la cellule unique, LDHA permet de distinguer les cellules en auto-renouvellement des cellules en différenciation, et fait partie des principaux nœuds des réseaux de corrélation inférés aux points 8 et 24h, points critiques du processus de différenciation (chute des corrélations entre gènes et pic de variabilité de l'expression génique). LDHA semble donc particulièrement importante pour les progéniteurs érythrocytaires, mais également à l'initiation du processus de différenciation. Il s'agit d'une enzyme clef de la glycolyse anaérobie, impliquée dans la prolifération cellulaire et l'effet Warburg [191, 192]. Ayant de plus constaté dans notre première étude, que l'expression transcriptionnelle de LDHA

diminuait fortement au cours de la différenciation, nous avons souhaité tout d'abord tester l'hypothèse selon laquelle les progéniteurs érythrocytaires pourraient dépendre de la glycolyse anaérobie, et permuter vers la voie aérobie de la phosphorylation oxydative (OXPHOS), au cours de leur processus de différenciation. Un tel changement de métabolisme est de plus en plus suggéré dans les processus de différenciation, et a déjà été mis en évidence dans les cellules souches hématopoïétiques (HSC) et neuronales [119, 121]. Dans la seconde partie de ma thèse, nous avons tout d'abord montré que le niveau protéique de divers enzymes glycolytiques clés, dont LDHA, diminuait fortement au cours de la différenciation, contrairement au niveau protéique de composants des complexes de la chaîne respiratoire. L'expression de LDHA et de PKM montrent cependant une augmentation ponctuelle de leur expression à 12h de différenciation. Nous avons complété ces résultats par des mesures du métabolisme, telles que la concentration de lactate et le potentiel de membrane mitochondrial (MMP). Nos résultats suggèrent qu'au cours de la différenciation érythrocytaire, les cellules produiraient moins de lactate et que l'activité de l'OXPHOS serait augmentée. Afin de consolider ces résultats nous avons ensuite mesuré et disséqué l'évolution de la respiration cellulaire au cours de la différenciation. Les résultats obtenus ont permis, d'une part, de valider les conclusions de nos précédentes expériences, et d'autre part, de mettre en évidence une augmentation de la production énergétique à 24h de différenciation. L'augmentation ponctuelle du niveau protéique de LDHA et PKM à 12h, a alors permis de proposer un scénario dans lequel l'augmentation de la production énergétique à 24h pourrait être assurée par la glycolyse et l'OXPHOS conjointement [160]. Ainsi, bien que certains de ces résultats demandent à être consolidés, l'ensemble des conclusions que nous avons pu établir soutiennent notre hypothèse suggérant que la différenciation érythrocytaire s'accompagne d'un switch métabolique, de la glycolyse anaérobie vers l'OXPHOS.

D'autre part, comme démontré par nos résultats, ainsi que dans un modèle humain [193], la différenciation érythrocytaire semble s'accompagner d'une augmentation ponctuelle des besoins énergétiques. Comme suggéré dans l'analyse de *Baron et al.*, la soudaine augmentation de la production d'énergie pourrait répondre à une demande accrue d'ATP, nécessitant l'intervention de la glycolyse anaérobie comme de l'OXPHOS [160]. Cette hypothèse expliquerait ainsi le pic d'expression de ces deux enzymes clés de la glycolyse anaérobie à 12h,

précédant l'augmentation de la respiration cellulaire, observée à 24h. En effet, la cinétique que nous avons choisie de réaliser ne comprend pas le point à 12h, cependant il est possible que la respiration cellulaire puisse déjà être augmentée dès 12h. Ainsi, afin de compléter notre analyse et de déterminer si la glycolyse anaérobie et l'OXPPOS participent simultanément au pic de production d'énergie, il pourrait être intéressant de mesurer à 12h le taux de production de lactate ainsi que la respiration cellulaire.

Enfin nous nous sommes intéressés aux rôles de LDHA dans l'auto-renouvellement et la différenciation des progéniteurs érythrocytaires. Les résultats d'ores et déjà obtenus, confirment que LDHA joue bien un rôle dans le maintien en auto-renouvellement des progéniteurs. De plus, une première expérience nous a permis de suggérer que l'inhibition de LDHA pourrait être nécessaire au passage métabolique de la glycolyse anaérobie vers l'OXPPOS. Cependant, afin de démontrer de façon plus directe que l'inhibition de LDHA est essentielle au remaniement métabolique, et que celui-ci est nécessaire à la différenciation érythrocytaire, nous construisons actuellement un vecteur de surexpression de LDHA. Nous pourrions ainsi maintenir l'expression de LDHA dans les progéniteurs érythrocytaires soumis aux signaux de différenciation, afin de déterminer si la diminution de son expression est essentielle à l'engagement des cellules en différenciation.

4.4 Implications d'autres voies métaboliques dans la différenciation érythrocytaire

Les premiers travaux de ma thèse ont également mis en lumière l'importance des gènes de la biosynthèse des stérols dans le processus de différenciation érythrocytaire. Il a été montré dans l'équipe, que cette voie métabolique est particulièrement importante dans le maintien des progéniteurs en auto-renouvellement [17]. Toutefois, le métabolisme des stérols reste peu caractérisé dans la différenciation érythrocytaire, et il est également mal connu, de façon générale, dans les autres processus de différenciation eucaryotes. Nos résultats démontrent pourtant qu'un certain nombre de gènes impliqués dans la synthèse de stérols pourraient jouer un rôle important dans la différenciation érythrocytaire, et particulièrement dès les premières heures du processus. En effet, parmi tous les gènes différentiellement exprimés

entre les progéniteurs en auto-renouvellement et les progéniteurs induits en différenciation depuis 48h, une dizaine de gènes impliqués dans la synthèse des stérols a été retenue pour former le groupe de 92 gènes étudiés en cellule unique. Ces gènes n'ont pas été initialement sélectionnés pour leur fonction, mais parce qu'ils faisaient partie des gènes les plus différentiellement exprimés entre les deux conditions. Lorsque ces gènes ont été analysés à l'échelle de la population, ils ont pu être regroupés en fonction de leurs profils d'expression, très similaires, au cours du processus de différenciation. De plus leur profil d'expression arbore une dynamique très particulière au cours du processus de différenciation. Il se différencie de celui des autres groupes de gènes analysés par ses importantes fluctuations, contrairement aux autres profils, plutôt monotones. Ce profil présente tout d'abord une chute de l'expression génique à 8h, suivie d'une augmentation à 24h et d'une stabilisation jusqu'à 48h, où l'expression génique chute à nouveau. Cette observation suggère donc que le métabolisme des stérols semble lui aussi subir un important remaniement au cours du processus de différenciation. Une analyse de l'expression des 92 gènes en cellule unique, dans des progéniteurs érythrocytaires en auto-renouvellement et après 2h, 4h et 8h de différenciation, a révélé que parmi tous les gènes étudiés, seulement 7 montraient des variations d'expression significatives dès 2h et 4h. Or, parmi ces 7 gènes, 6 sont impliqués dans la synthèse des stérols, suggérant que le métabolisme des stérols pourrait ainsi être l'un des premiers paramètres métaboliques à être impliqué dans l'initiation de la différenciation. Cette voie métabolique pourrait donc éventuellement détecter des changements de l'environnement extra-cellulaire, et répondre immédiatement à ces perturbations en influençant le réseau métabolique. Il pourrait être alors intéressant d'étudier plus en détails l'évolution du métabolisme des stérols, parallèlement au métabolisme du glucose, dans une cinétique précoce (dès 2h de différenciation) et plus rapprochée du processus de différenciation. Les connexions directes entre le métabolisme des stérols et du glucose, pourront également être analysées, afin de mieux caractériser le comportement du réseau métabolique dès l'initiation du processus de différenciation.

D'autre part, il a été mis en évidence que le métabolisme de la glutamine jouait un rôle important dans l'engagement des HSC dans le lignage érythrocytaire [193]. En effet, en présence d'érythropoïétine les cellules souches se différencient en érythrocytes, mais si l'entrée de la glutamine dans les cellules est bloquée ou la dégradation de la glutamine stoppée, les

HSC ne s'engagent plus dans la lignée érythrocytaire mais se spécialisent préférentiellement dans le lignage myélomonocytaire [162]. De plus, il a été suggéré que les nutriments présents dans le milieu extracellulaire pourraient modifier le statut métabolique des cellules souches et influencer ainsi leur choix de lignage [194, 195]. Ainsi, le métabolisme semble jouer un rôle essentiel dans les processus de prise de décision des cellules, notamment lorsqu'elles doivent choisir entre se maintenir dans leur état d'auto-renouvellement ou s'engager en différenciation [125]. Il serait alors intéressant de réaliser des analyses du métabolome au cours du processus de différenciation, afin de disséquer le fonctionnement du réseau métabolique sous-jacent, et de mieux caractériser ce processus. De plus, ces analyses pourraient permettre à terme d'identifier de nouvelles cibles thérapeutiques dans le traitement de pathologies résultant de défauts de la différenciation [196]. Une analyse du métabolome réalisée dans le cadre de l'effet Warburg, a par exemple permis de modéliser les métabolismes du glucose et de la glutamine, afin de mieux comprendre le réarrangement métabolique, de l'OXPHOS vers la glycolyse anaérobie, caractéristique des cellules cancéreuses [197]. Enfin, dans le cadre de notre étude, analyser le métabolome au cours de la différenciation érythrocytaire nous permettrait, d'une part, de mieux caractériser le métabolisme de ce lignage hématopoïétique, et d'autre part, d'apporter de nouveaux éléments de réponse concernant le rôle du métabolisme des stérols à l'initiation de la différenciation et son éventuelle implication dans le remaniement métabolique observé au cours de la différenciation.

4.5 Intégration du pic énergétique dans la chronologie des événements moléculaires identifiés au cours de la différenciation

Nos résultats ainsi que les travaux de Leal Oburoglu et. al. démontrent qu'un pic énergétique a lieu à l'initiation du processus de différenciation [193]. Or, nous avons constaté que cet événement métabolique concordait, en terme de cinétique, avec les événements critiques de l'expression génique, précédemment mis en évidence dans la première partie de ma thèse (figure 16).

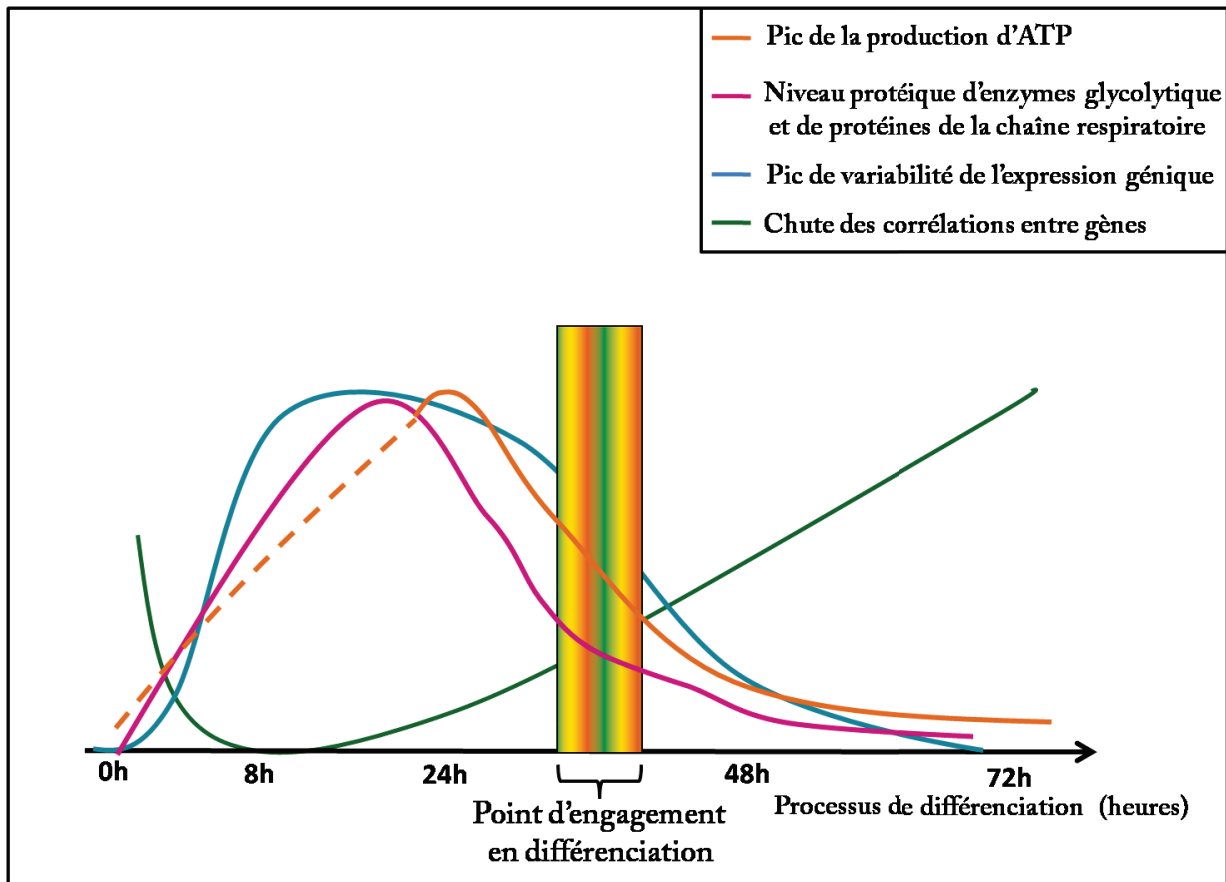


FIGURE 16 – Superposition des événements métaboliques et d'expression génique au cours de la différenciation érythrocytaire

Ce schéma représente les principales conclusions de notre analyse métabolique au cours du processus de différenciation érythrocytaire, vu de façon superposée aux résultats obtenus dans le cadre de l'analyse de l'expression des gènes en cellule unique. Le niveau protéique de deux enzymes glycolytiques clés, et de deux constituants des complexes de la chaîne respiratoire, augmente à 12h, suivit d'un pic de la consommation de dioxygène allouée à la production d'ATP. Les pointillés représentent l'évolution de la courbe pour laquelle nous n'avons pas de données.

En effet, le pic énergétique, l'augmentation de la variabilité de l'expression génique et la chute des corrélations entre gènes ont lieu exactement dans la même fenêtre de temps, située entre 8h et 24h, précédant l'engagement irréversible des cellules en différenciation. Bien que les deux analyses aient été réalisées de façon indépendante, l'orchestration de ces événements au cours du processus de différenciation fait sens. L'augmentation de la variabilité de l'expression génique, pourrait être en partie due à un remaniement de l'épigénome

ayant pour conséquence la levée temporaire de certaines contraintes sur l'expression des gènes [96, 168]. Des modifications épigénétiques ont en effet été mises en évidence dans les processus de différenciation [198, 199, 200]. Or, il a été montré que ces modifications épigénétiques, et notamment les remaniements de la chromatine, pouvaient être coûteux en terme d'énergie [201]. L'augmentation de la variabilité de l'expression génique et de la production d'énergie au même moment de la cinétique de différenciation, pourrait alors être expliqué par ce processus de remodelage du paysage épigénétique. Ainsi, lors de cette phase de remaniements épigénétiques, perturbant le réseau de régulation de gènes, les cellules pourraient se trouver dans un état transitoire instable, où l'ancien réseau est déstabilisé et le nouveau réseau pas encore reformé. Cet état instable peut alors se refléter à différents niveaux moléculaires et cellulaires. Il est mis en évidence, d'une part, par l'augmentation de la variabilité de l'expression des gènes et la diminution des corrélations entre gènes. D'autre part, il se reflète dans le statut métabolique. En effet, le passage du statut anaérobique vers l'état aérobie au cours de la différenciation, traverse une phase pendant laquelle les deux voies semblent actives. Enfin, l'état d'instabilité créée par les remaniements du paysage épigénétique, peut se refléter phénotypiquement, comme Alice Moussy l'a montré dans ses travaux [183]. En effet, les cellules hématopoïétiques CD34+, qui suite à leur induction en différenciation, adoptent un phénotype intermédiaire dit "hésitant", fluctuant entre deux formes, correspondant à l'état plus différencié de progéniteur myéloïde ou à l'état multipotent [183]. Ainsi toutes les transitions observées au cours du processus de différenciation, qu'elles concernent le métabolisme, l'expression des gènes ou le phénotype, entrent dans un processus global correspondant au passage d'un stade cellulaire à un autre, au travers d'une phase instable, générant de la variabilité et élargissant le choix des possibles, comme expliqué dans la section 4.1.

4.6 Communications entre le réseau de régulation de gènes et le réseau métabolique

L'ensemble des travaux présentés dans ce manuscrit met en évidence de façon indépendante, que sous l'influence de signaux extérieurs de différenciation, l'expression des gènes et le métabolisme traversent parallèlement une phase particulière instable, avant d'atteindre un

nouvel état, pouvant être caractérisé par un profil d'expression génique et un statut métabolique différents. Bien qu'il ait été suggéré à plusieurs reprises que la variabilité de l'expression génique et le remaniement métabolique puissent être des forces motrices du processus de différenciation [185, 88, 186, 157, 121, 158], il n'est pas encore clairement établi si l'une ou l'autre de ces deux forces joue un rôle leader, ou si elles travaillent de concert. Nous avons discuté dans le paragraphe 4.5 de la concordance temporelle des événements métaboliques avec les événements moléculaire de l'expression génique. Nous avons alors suggéré qu'ils participaient à un processus global par lequel la cellule change d'état. Or, la transition de l'état cellulaire initial, au nouvel état, pourrait impliquer un remodelage de l'environnement chromatinien et des marques épigénétiques. Je me propose donc à présent d'ouvrir la discussion sur les connexions reliant ces événements entre eux pour mieux comprendre l'influence mutuelle des réseaux entre eux.

Le réseau de régulation de gènes est connu pour jouer un rôle très important dans les processus biologiques, et a souvent été identifié comme étant le principal responsable de tous les événements impliquant une prise de décision cellulaire. Dans la vision proposée par Conrad Waddington en 1957 [96], et modernisée au fil des années [202], l'augmentation de la variabilité de l'expression génique, permettant de sortir de l'état attracteur initial, serait nécessaire pour permettre ensuite aux cellules de choisir entre s'auto-renouveler ou se différencier. Les états d'auto-renouvellement et de différenciation seraient alors limités par le réseau de régulation de gènes sous le contrôle de l'épigénome, et chaque cellule pourrait présenter sa propre probabilité d'entrer dans l'une ou l'autre de ces voies. Une telle description de la différenciation suggère que ce processus est particulièrement soumis au comportement du réseau de régulation de gène. Toutefois, la force motrice entraînant de telles modifications du réseau de régulation de gènes n'est pas encore clairement défini. Comme nous l'avons mentionné, l'épigénome pourrait jouer un rôle important dans la transition stochastique, entre l'état d'auto-renouvellement et l'état différencié. Réciproquement, l'épigénome peut à son tour être influencé par le métabolisme et l'expression génique. En effet, des modifications de l'épigénome peuvent résulter de fluctuations dans la disponibilité en substrats métaboliques. Premièrement, nous savons aujourd'hui que les enzymes modifiant l'organisation de la chromatine, utilisent l'énergie libérée par l'hydrolyse de l'ATP, pour induire des modifications de

la conformation chromatinienne [203]. Ainsi, des variations de la quantité d'ATP intercellulaire pourrait avoir des conséquences sur l'organisation de la chromatine, et ainsi jouer un rôle dans les processus de différenciation cellulaire [201]. A la lumière de ces connaissances, nos résultats prennent donc un sens tout particulier. En effet, le pic de respiration cellulaire alloué à la production d'ATP, qui est observé au début du processus de différenciation érythrocytaire, pourrait permettre aux enzymes ATP-dépendantes d'induire des modifications de la chromatine, résultant en une augmentation de la variabilité de l'expression génique. D'autre part, des métabolites comme l'acétyl CoA, provenant de la dégradation du glucose et des lipides, sont utilisés comme substrats par des enzymes régulant la compaction de la chromatine. Il s'agit par exemple des histone acétyltransférases, qui transfèrent le groupe acétyl de l'acétyl CoA sur les histones [164]. Cette acétylation des histones réduit alors leur interaction avec l'ADN et résulte en une structure plus relâchée de la chromatine. Ainsi la disponibilité en acétyl CoA est également une source de modifications épigénétiques. d'autre part, ce métabolite est également utilisé par différentes voies de synthèse et de dégradation. Par conséquent, un défaut de régulation du métabolisme peut entraîner un surplus de sa consommation par les voies métaboliques, le rendant indisponible pour l'acétylation des histones. Un tel phénomène aurait alors des répercussions sur la régulation de l'expression des gènes, pouvant conduire au développement de pathologies [169]. Enfin, contrairement à l'acétyl CoA, le NAD⁺, est impliqué, quant à lui, dans la déacétylation des histones. Les histones déacétylases appelées sirtuins, par exemple, utilisent l'énergie délivrée par l'hydrolyse du NAD pour déacétyler les histones, entraînant ainsi une condensation de la chromatine [172, 204]. Il a notamment été montré que SIRT1, un membre de la famille des sirtuins, serait impliquée dans la régulation du processus de différenciation des lymphocytes T, de part son rôle d'intermédiaire entre le métabolisme et l'expression génique [104]. Ainsi, bien que le réseau de régulation de gènes soit à l'origine de nombreux processus cellulaires, il semble pouvoir agir sous le contrôle du réseau métabolique, par l'intermédiaire de l'épigénome (figure 17).

En revanche, le métabolisme est également sous le contrôle du réseau de régulation de gènes (figure 17). En effet, certains facteurs de transcription, étant principalement reconnus pour leur capacité à influencer l'expression d'un grand nombre de gènes, peuvent contrôler de grandes voies métaboliques. Les facteurs de transcription SREBP et HIF1- α , par

exemple, sont respectivement impliqués dans la régulation des métabolismes des lipides et du glucose [205, 206]. SREBP régule l'expression de gènes impliqués dans la biosynthèse des stérols et HIF1- α contrôle l'expression de gènes codant pour des enzymes clefs de la glycolyse et pour *LDHA*. Ces protéines pourraient représenter des noeuds particulièrement importants du réseau de régulation de gènes, car ils sont aux carrefours de différentes voies cellulaires. En effet, SREBP comme HIF1- α , interviennent non seulement en amont de grandes voies métaboliques, mais sont également elles-mêmes régulées par d'autres gènes, tels que *mTOR* [207, 208]. mTOR fait partie de la voie de signalisation PI3K-akt-mTOR, aujourd'hui bien caractérisée. Cette voie est particulièrement connue pour jouer un rôle important dans le maintien des cellules en auto-renouvellement, de part son influence sur un grand nombre de gènes, impliqués dans des paramètres cellulaires différents, tels que le métabolisme [209]. Il a par exemple été montré qu'Akt favorise le maintien et le recrutement à la surface membranaire des transporteurs membranaires du glucose, les GLUT, et active l'expression de l'hexokinase 1 (HK1) [210]. Cependant, seule, l'accumulation de GLUT ne permet pas de promouvoir la glycolyse. En effet, en activant l'expression d'HIF1- α , Akt permet d'activer l'expression de plusieurs gènes cibles d'HIF1- α , impliqués dans la glycolyse [208]. La voie PI3K-akt-mTOR favorise donc la glycolyse anaérobie et permet par conséquent de retenir les cellules en état d'auto-renouvellement.

En condition d'auto-renouvellement, les cellules se divisent à plusieurs reprises et nécessitent alors de produire de nouveaux constituants structuraux, comme des membranes. Les stérols étant d'importants composants membranaires, il semble alors essentiel que leur synthèse soit augmentée en état d'auto-renouvellement. mTOR est un activateur de SREBP, il favorise donc la voie anabolique des stérols, permettant aux cellules de proliférer [207]. Le facteur SREBP permet donc de coupler la croissance cellulaire avec le métabolisme lipidique [205]. Ces exemples illustrent ainsi le contrôle du métabolisme par l'expression génique. En revanche, il a été démontré en conditions pathologiques, que LDHA pourrait activer l'expression d'HIF1- α , maintenant par conséquent la glycolyse anaérobie et retenant les cellules dans leur état d'auto-renouvellement [134]. Ces boucles de rétro-contrôle entre facteurs de transcription et enzymes du métabolisme ne font alors que renforcer la causalité circulaire semblant exister entre le réseau de régulation de gènes et le réseau métabolique (figure17).

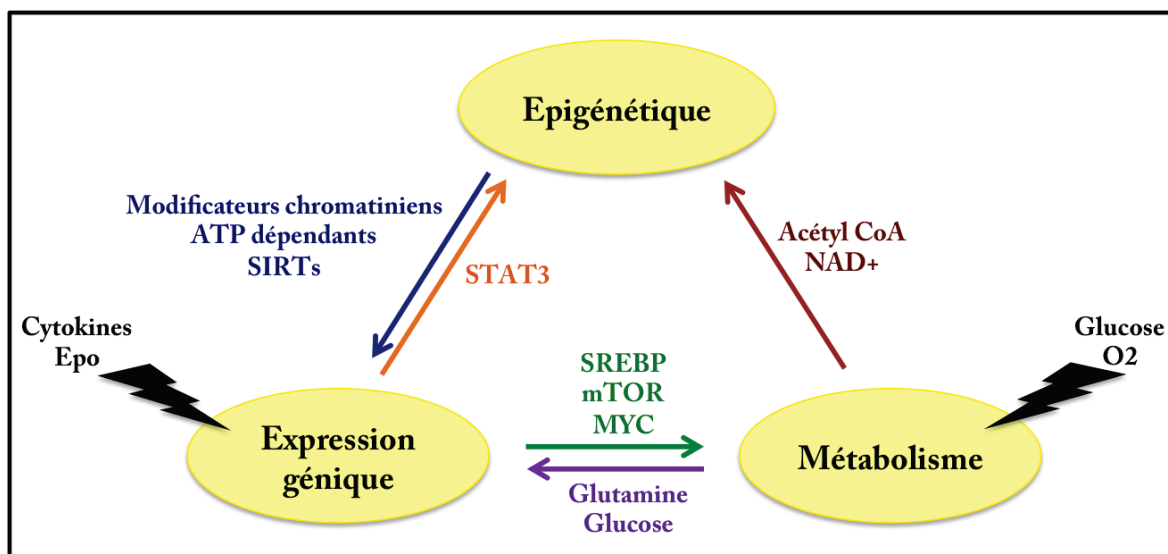


FIGURE 17 – Causalité circulaire entre le métabolisme, l'épigénétique et l'expression génique

Le métabolisme, l'épigénome et l'expression génique s'influencent mutuellement de façon circulaire. Les flèches représentent des actions causales. Des exemples concrets ont été ajoutés pour illustrer chacune de ces connexions.

Comme illustré au cours des deux paragraphes précédents, les fluctuations du métabolisme entraînent des changements dans l'expression génique en modifiant l'état de l'épigénome ; alors qu'inversement, le réseau de régulation de gènes peut activer l'expression de facteur de transcription dont les cibles sont des enzymes clés de grandes voies métaboliques. Le réseau de régulation de gènes et le réseau métabolique s'influencent donc mutuellement, et il semble difficile de définir lequel d'entre eux pourrait représenter le véritable moteur de la différenciation. Cette relation circulaire a été mise en évidence et discutée dans un contexte tumoral [211]. L'oncogène Myc et les voies métaboliques du glucose et de la glutamine jouent chacun un rôle important dans le développement et la maintenance de l'état tumoral. Myc régule l'expression d'un grand nombre de gènes, dont des gènes impliqués dans le métabolisme du glucose et de la glutamine. Il régule par exemple l'expression du transporteur membranaire du glucose, GLUT1, et l'enzyme glutaminase (GLS) catalysant la glutaminolyse. Inversement, le glucose et la glutamine sont par exemple utilisés comme substrats de protéines induisant des modifications post-transcriptionnelles, dont MYC est la cible. Par conséquent un manque de glucose ou de glutamine peut entraîner une diminution de l'expression de *myc* [211]. Métabolisme et expression génique pourraient donc agir ensemble, en

fonction des signaux provenant de l'environnement extra-cellulaire (nutriments, cytokines, O₂, ou autre), pour permettre aux cellules de choisir entre la voie de l'auto-renouvellement, ou la ou les voie(s) de différenciation.

Bibliographie

- [1] A. S. Sperling, C. J. Gibson, and B. L. Ebert. The genetics of myelodysplastic syndrome : from clonal haematopoiesis to secondary leukaemia. *Nature Reviews Cancer*, 17(5), 2017.
- [2] M. S. Chung. Congenital differences of the upper extremity : classification and treatment principles. *Clin Orthop Surg*, 3(3) :172–7, 2011.
- [3] J. Yang, M. V. Plikus, and N. L. Komarova. The role of symmetric stem cell divisions in tissue homeostasis. *PLoS Comput Biol*, 11(12) :e1004629, 2015.
- [4] J. A. Knoblich. Mechanisms of asymmetric stem cell division. *Cell*, 132(4) :583–97, 2008.
- [5] L. Robb, N. J. Elwood, A. G. Elefanty, F. Kontgen, R. L. Li, L. D. Barnett, and C. G. Begley. The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. *Embo Journal*, 15(16) :4123–4129, 1996.
- [6] S. J. Szabo, S. T. Kim, G. L. Costa, X. K. Zhang, C. G. Fathman, and L. H. Glimcher. A novel transcription factor, t-bet, directs th1 lineage commitment. *Cell*, 100(6) :655–669, 2000.
- [7] A. K. Panigrahi and D. Pati. Higher-order orchestration of hematopoiesis : is cohesin a new player? *Exp Hematol*, 40(12) :967–73, 2012.
- [8] J. Medvedovic, A. Ebert, H. Tagoh, and M. Busslinger. Pax5 : a master regulator of b cell development and leukemogenesis. *Adv Immunol*, 111 :179–206, 2011.
- [9] S. S. Chan and M. Kyba. What is a master regulator? *J Stem Cell Res Ther*, 3, 2013.
- [10] E. Skutelsky and D. Danon. Expulsion of erythroid nucleus and its phagocytosis. *Anatomical Record*, 173(1) :123–+, 1972.
- [11] K. M. Moritz, G. B. Lim, and E. M. Wintour. Developmental regulation of erythropoietin and erythropoiesis. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 273(6) :R1829–R1844, 1997.
- [12] M. Mortensen, D. J. P. Ferguson, and A. K. Simon. Mitochondrial clearance by autophagy in developing erythrocytes clearly important, but just how much so? *Cell Cycle*, 9(10) :1901–1906, 2010.
- [13] A. Kowalski and J. Palyga. Chromatin compaction in terminally differentiated avian blood cells : the role of linker histone h5 and non-histone protein ment. *Chromosome Research*, 19(5) :579–590, 2011.

- [14] A. Stier, P. Bize, Q. Schull, J. Zoll, F. Singh, B. Geny, F. Gros, C. Royer, S. Massemin, and F. Criscuolo. Avian erythrocytes have functional mitochondria, opening novel perspectives for birds as animal models in the study of ageing. *Frontiers in Zoology*, 10, 2013.
- [15] C Lacombe and P Mayeux. Biology of erythropoietin. *Haematologica*, 83(8) :724–732, 1998.
- [16] C. Bresson-Mazet, O. Gandrillon, and S. Gonin-Giraud. Stem cell antigen 2 : a new gene involved in the self-renewal of erythroid progenitors. *Cell Proliferation*, 41(5) :726–738, 2008.
- [17] C. Mejia-Pous, F. Damiola, and O. Gandrillon. Cholesterol synthesis-related enzyme oxidosqualene cyclase is required to maintain self-renewal in primary erythroid progenitors. *Cell Proliferation*, 44(5) :441–452, 2011.
- [18] M. H. Baron and T. Maniatis. Regulated expression of human alpha-globin and beta-globin genes in transient heterokaryons. *Molecular and Cellular Biology*, 11(3) :1239–1247, 1991.
- [19] S. Huang. The molecular and mathematical basis of waddington’s epigenetic landscape : a framework for post-darwinian biology? *Bioessays*, 34(2) :149–57, 2012.
- [20] J. Ladewig, P. Koch, and O. Brustle. Leveling waddington : the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol*, 14(4) :225–36, 2013.
- [21] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584) :1183–1186, 2002. p39.
- [22] G. Yvert. ‘particle genetics’ : treating every cell as unique. *Trends Genet*, 30(2) :49–56, 2014. P39.
- [23] A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, and A. Regev. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498(7453) :236–40, 2013. p40.
- [24] J. J. Kupiec. A darwinian theory for the origin of cellular differentiation. *Molecular and General Genetics MGG*, 255(2) :201–208, Jun 1997.
- [25] Mads Kærn, Timothy C. Elston, William J. Blake, and James J. Collins. Stochasticity in gene expression : from theories to phenotypes, Jun 1997.
- [26] Sui Huang. Systems biology of stem cells : three useful perspectives to help overcome the paradigm of linear pathways. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 366(1575) :2247–2259, 2011.

- [27] D. S. Goodsell. Inside a living cell. *Trends in Biochemical Sciences*, 16(6) :203–206, 1991.
- [28] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258) :561–3, 1970. Crick, F
Historical Article England Nature Nature. 1970 Aug 8;227(5258) :561-3.
- [29] F. Jacob and J. Monod. On regulation of gene activity. *Cold Spring Harbor Symposia on Quantitative Biology*, 26 :193–, 1961.
- [30] O. G. Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol*, 71(4) :587–603, 1978. p39.
- [31] L. Weinberger, Y. Voichek, I. Tirosh, G. Hornung, I. Amit, and N. Barkai. Expression noise and acetylation profiles distinguish hdac functions. *Molecular Cell*, 47(2) :193–202, 2012.
- [32] J. Viñuelas, G. Kaneko, A. Coulon, E. Vallin, V. Morin, C. Mejia-Pous, J-J. Kupiec, G. Beslon, and O. Gandrillon. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biology*, 11(1) :15, Feb 2013.
- [33] E. Pujadas and A. P. Feinberg. Regulated noise in the epigenetic landscape of development and disease. *Cell*, 148(6) :1123–31, 2012.
- [34] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6) :1025–36, 2005. p39.
- [35] J. M. Pedraza and J. Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861) :339–43, 2008. p39.
- [36] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332(6028) :472–4, 2011. p39.
- [37] B. Wu, C. Eliscovich, Y. J. Yoon, and R. H. Singer. Translation dynamics of single mrnas in live cells and neurons. *Science*, 352(6292) :1430–1435, 2016.
- [38] X. Pichon, A. Bastide, A. Safieddine, R. Chouaib, A. Samacoits, E. Basyuk, M. Peter, F. Mueller, and E. Bertrand. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J Cell Biol*, 214(6) :769–81, 2016.
- [39] J. C. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie. Single-molecule imaging of transcription factor binding to dna in live mammalian cells. *Nat Methods*, 10(5) :421–6, 2013.

- [40] Leonardo A. Sepúlveda, Heng Xu, Jing Zhang, Mengyu Wang, and Ido Golding. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science*, 351(6278) :1218–1222, 2016.
- [41] J. J. Chen, Z. J. Zhang, L. Li, B. C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 156(6) :1274–1285, 2014.
- [42] M. D. White, J. F. Angiolini, Y. D. Alvarez, G. Kaur, Z. W. Zhao, E. Mocskos, L. Bruno, S. Bissiere, V. Levi, and N. Plachta. Long-lived binding of sox2 to dna predicts cell fate in the four-cell mouse embryo. *Cell*, 165(1) :75–87, 2016.
- [43] M. Dadiani, D. van Dijk, B. Segal, Y. Field, G. Ben-Artzi, T. Raveh-Sadka, M. Levo, I. Karpow, A. Weinberger, and E. Segal. Two dna-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. *Genome Res*, 23(6) :966–76, 2013. p39.
- [44] N. Pelaez, A. Gavaldà-Miralles, B. Wang, H. T. Navarro, H. Gudjonson, I. Rebay, A. R. Dinner, A. K. Katsaggelos, L. A. Amaral, and R. W. Carthew. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *Elife*, 4, 2015. Pelaez, Nicolas Gavaldà-Miralles, Arnau Wang, Bao Navarro, Heliodoro Tejedor Gudjonson, Herman Rebay, Ilaria Dinner, Aaron R Katsaggelos, Aggelos K Amaral, Luis A N Carthew, Richard W eng P50GM081892/GM/NIGMS NIH HHS/ R01 GM077581/GM/NIGMS NIH HHS/ R01 GM080372/GM/NIGMS NIH HHS/ R01GM077581/GM/NIGMS NIH HHS/ R01GM80372/GM/NIGMS NIH HHS/ Howard Hughes Medical Institute/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, Non-P.H.S. England 2015/11/20 06 :00 Elife. 2015 Nov 19;4. pii : e08924. doi : 10.7554/eLife.08924.
- [45] L. M. Octavio, K. Gedeon, and N. Maheshri. Epigenetic and conventional regulation is distributed among activators of flo11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet*, 5(10) :e1000673, 2009.
- [46] A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq. Transcription factors modulate c-fos transcriptional bursts. *Cell Rep*, 8(1) :75–83, 2014. p39.
- [47] Peng Dong and Zhe Liu. Shaping development by stochasticity and dynamics in gene regulation. *Open Biology*, 7(5), 2017.

- [48] S. S. Chong, C. Y. Chen, H. Ge, and X. S. Xie. Mechanism of transcriptional bursting in bacteria. *Biophysical Journal*, 106(2) :484a–485a, 2014.
- [49] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo. Stochastic model of supercoiling-dependent transcription. *Phys. Rev. Lett.*, 117 :018101, Jun 2016.
- [50] T. Rajala, A. Hakkinen, S. Healy, O. Yli-Harja, and A. S. Ribeiro. Effects of transcriptional pausing on gene expression dynamics. *PLoS Comput Biol*, 6(3) :e1000704, 2010.
- [51] K. Fujita, M. Iwaki, and T. Yanagida. Transcriptional bursting is intrinsically caused by interplay between rna polymerases on dna. *Nature Communications*, 7, 2016.
- [52] C. R. Brown, C. Mao, E. Falkovskaia, M. S. Jurica, and H. Boeger. Linking stochastic fluctuations in chromatin structure and gene expression. *PLoS Biol*, 11(8) :e1001621, 2013. p39.
- [53] J. Viñuelas, G. Kaneko, A. Coulon, G. Beslon, and O. Gandrillon. Toward experimental manipulation of stochasticity in gene expression. *Progress in Biophysics Molecular Biology*, 110 :44–53, 2012.
- [54] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci U S A*, 109(43) :17454–9, 2012. Dar, Roy D Razooky, Brandon S Singh, Abhyudai Trimeloni, Thomas V McCollum, James M Cox, Chris D Simpson, Michael L Weinberger, Leor S OD006677/OD/NIH HHS/United States P50 GM085764/GM/NIGMS NIH HHS/United States P50GM081879/GM/NIGMS NIH HHS/United States Research Support, N.I.H., Extramural Research Support, U.S. Gov’t, Non-P.H.S. United States Proceedings of the National Academy of Sciences of the United States of America Proc Natl Acad Sci U S A. 2012 Oct 23;109(43) :17454-9. doi : 10.1073/pnas.1213530109. Epub 2012 Oct 11.
- [55] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz. Dynamic heterogeneity and dna methylation in embryonic stem cells. *Mol Cell*, 55(2) :319–31, 2014. p39.
- [56] G. Chalancon, C. N. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, and M. M. Babu. Interplay between gene expression noise and regulatory network architecture. *Trends Genet*, 28 :221–232, 2012. p39.

- [57] E. Splinter and W. de Laat. The complex transcription regulatory landscape of our genome : control in three dimensions. *EMBO J*, 30(21) :4345–55, 2011.
- [58] A. Raj, S. A. Rifkin, E. Andersen, and A. van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283) :913–8, 2010. p39.
- [59] J. Paulsson. Models of stochastic gene expression. *Phys. Life Rev.*, 2 :157–175, 2005. p39.
- [60] R. Guantes, J. Diaz-Colunga, and F. J. Iborra. Mitochondria and the non-genetic origins of cell-to-cell variability : More is different. *Bioessays*, 38(1) :64–76, 2016.
- [61] J.J Kupiec. A probabilistic theory for cell differentiation, embryonic mortality and dna c-value paradox. *Speculations in Science and Technology*, 6(5) :471–478, 1983.
- [62] S. Huang. Non-genetic heterogeneity of cells in development : more than just noise. *Development*, 136(23) :3853–62, 2009. p39.
- [63] O. Feinerman, J. Veiga, J. R. Dorfman, R. N. Germain, and G. Altan-Bonnet. Variability and robustness in t cell activation from regulated heterogeneity in protein levels. *Science*, 321(5892) :1081–4, 2008. p39 p37.
- [64] R. A. Kellogg and S. Tay. Noise facilitates transcriptional control under dynamic inputs. *Cell*, 160(3) :381–92, 2015. p39.
- [65] Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. van Oudenaarden, and R. Jaenisch. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6) :1209–22, 2012. p39, p40.
- [66] M. F. Wernet, E. O. Mazzoni, A. Celik, D. M. Duncan, I. Duncan, and C. Desplan. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, 440(7081) :174–80, 2006.
- [67] J. Rossant and P. P. Tam. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development*, 136(5) :701–13, 2009. 0950-1991 (Print) Journal Article Research Support, Non-U.S. Gov’t Review.
- [68] F. Paul, Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. B. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay, and I. Amit. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*, 163(7) :1663–1677, 2015.

- [69] Roy D. Dar, Nina N. Hosmane, Michelle R. Arkin, Robert F. Siliciano, and Leor S. Weinberger. Screening for noise in gene expression identifies drug synergies. *Science*, 344(6190) :1392–1396, 2014.
- [70] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad. Cellular noise regulons underlie fluctuations in *saccharomyces cerevisiae*. *Mol Cell*, 45(4) :483–93, 2012. p39.
- [71] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger. Non-genetic origins of cell-to-cell variability in trail-induced apoptosis. *Nature*, 459(7245) :428–32, 2009. p39.
- [72] F. Bertaux, S. Stoma, D. Drasdo, and G. Batt. Modeling dynamics of cell-to-cell variability in trail-induced apoptosis explains fractional killing and predicts reversible resistance. *PLoS Comput Biol*, 10(10) :e1003893, 2014. Bertaux, Francois Stoma, Szymon Drasdo, Dirk Batt, Gregory eng Research Support, Non-U.S. Gov’t 2014/10/24 06 :00 PLoS Comput Biol. 2014 Oct 23 ;10(10) :e1003893. doi : 10.1371/journal.pcbi.1003893. eCollection 2014 Oct.
- [73] B. Lehner. Conflict between noise and plasticity in yeast. *PLoS Genet*, 6(11) :e1001185, 2010.
- [74] AM. Singh. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem cell reports*, 1 :532–44, 2013.
- [75] K. Abley, J. C. W. Locke, and H. M. O. Leyser. Developmental mechanisms underlying variable, invariant and plastic phenotypes. *Annals of Botany*, 117(5) :733–748, 2016.
- [76] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690) :1622–5, 2004. 1095-9203 (Electronic) Journal Article Research Support, Non-U.S. Gov’t Research Support, U.S. Gov’t, P.H.S.
- [77] D. Fraser and M. Kaern. A chance at survival : gene expression noise and phenotypic diversification strategies. *Mol Microbiol*, 71(6) :1333–40, 2009. 1365-2958 (Electronic) Journal Article Research Support, Non-U.S. Gov’t Review.
- [78] J. A. Rebhahn, N. Deng, G. Sharma, A. M. Livingstone, S. Huang, and T. R. Mosmann. An animated landscape representation of cd4(+) t-cell differentiation, variability, and plasticity : Insights into the behavior of populations versus cells. *Eur J Immunol*, 44(8) :2216–29, 2014. Rebhahn, Jonathan A Deng, Nan Sharma, Gaurav Livingstone, Alexandra M Huang, Sui Mosmann, Tim R eng R24 AI054953/AI/NIAID NIH HHS/ Germany 2014/06/20 06 :00 Eur J Immunol. 2014 Aug ;44(8) :2216-29. doi : 10.1002/eji.201444645.
- [79] G. M. Suel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083) :545–550, 2006.

- [80] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat Genet*, 31(1) :69–73, 2002.
- [81] B. Munsky, G. Neuert, and A. van Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336(6078) :183–7, 2012. p39.
- [82] L. Warren, D. Bryder, I. L. Weissman, and S. R. Quake. Transcription factor profiling in individual hematopoietic progenitors by digital rt-pcr. *Proc Natl Acad Sci U S A*, 103(47) :17807–12, 2006.
- [83] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann. The technology and biology of single-cell rna sequencing. *Mol Cell*, 58(4) :610–20, 2015.
- [84] S. J. Clark, H. J. Lee, S. A. Smallwood, G. Kelsey, and W. Reik. Single-cell epigenomics : powerful new methods for understanding gene regulation and cell identity. *Genome Biology*, 17, 2016.
- [85] M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, and P. Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*, 14(9) :865–868, 2017.
- [86] C. A. Vallejos, J. C. Marioni, and S. Richardson. Basics : Bayesian analysis of single-cell sequencing data. *PLoS Comput Biol*, 11(6) :e1004333, 2015.
- [87] S. A. ; Hagai T. Miragaia, R. J. ; Teichmann. Single-cell insights into transcriptomic diversity in immunity. *Current Opinion in Systems Biology*, 5 :63–71, 2017.
- [88] Patrick S. Stumpf, Rosanna C.G. Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann Babbie, Thalia E. Chan, Michael P.H. Stumpf, Colin P. Please, Sam D. Howison, Fumio Arai, and Ben D. MacArthur. Stem cell differentiation as a non-markov stochastic proces. *Cell Systems*, 5(3) :268–282.e7, 2017.
- [89] A. Llamosi, A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt. What population reveals about individual cell identity : Single-cell parameter estimation of models of gene expression in yeast. *PLoS Comput Biol*, 12(2) :e1004706, 2016.
- [90] T. E. Chan, M. P. H. Stumpf, and A. C. Babbie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst*, 5(3) :251–267 e3, 2017.
- [91] U. Herbach, A. Bonnaïfoux, T. Espinasse, and O. Gandrillon. Inferring gene regulatory networks from single-cell data : a mechanistic approach. *Bmc Systems Biology*, 11, 2017.

- [92] K. Sharp and F. Matschinsky. Translation of ludwig boltzmann's paper "on the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium" *sitzungsberichte der kaiserlichen akademie der wissenschaften. mathematisch-naturwissen classe. abt. ii, lxxvi 1877*, pp 373-435 (wien. ber. 1877, 76 :373-435). reprinted in *wiss. abhandlungen*, vol. ii, reprint 42, p. 164-223, barth, leipzig, 1909. *Entropy*, 17(4) :1971–2009, 2015.
- [93] P. W. Atkins. The second law : Energy, chaos and form. *Scientific American Book*, 1994.
- [94] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423,623–656, 1948.
- [95] A. Richard, L. Boullu, U. Herbach, A. Bonnafoux, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J. J. Kupiec, T. Espinasse, S. Gonin-Giraud, and O. Gandrillon. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol*, 14(12) :e1002585, 2016.
- [96] C. H. Waddington. The strategy of the genes. *London, UK : Allen and Unwin*, 1957.
- [97] Thomas Heams. *Approche endodarwinienne de la variabilité intercellulaire de l'expression génétique*. PhD thesis, 2004. Thèse de doctorat dirigée par Kupiec, Jean-Jacques Biologie. Différenciation Paris, Institut national d'agronomie de Paris Grignon 2004.
- [98] A. Raza and N. Galili. The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. *Nat Rev Cancer*, 12(12) :849–59, 2012.
- [99] E. Jequier. Carbohydrates as a source of energy. *American Journal of Clinical Nutrition*, 59(3) :682s–685s, 1994.
- [100] B. S. Khakh and G. Burnstock. The double life of atp. *Sci Am*, 301(6) :84–90, 92, 2009.
- [101] Weihai Ying. Nad^+/nadh and $\text{nadp}^+/\text{nadph}$ in cellular functions and cell death : Regulation and biological consequences. *Antioxidants Redox Signaling*, 10(2) :179, 206, 2007.
- [102] N. Kresge, R. D. Simoni, and R. L. Hill. Otto fritz meyerhof and the elucidation of the glycolytic pathway. *J Biol Chem*, 280(4) :e3, 2005.
- [103] JM Berg, JL Tymoczko, and L Stryer. The glycolytic pathway is tightly controlled. *Biochemistry. 5th edition. New York : W H Freeman*, Section 16.2, 2002.

- [104] Y. Wang, Y. Bi, C. Chen, X. Li, Y. Li, Z. Zhang, J. Wang, Y. Lu, Q. Yu, H. Su, H. Yang, and G. Liu. Histone deacetylase sirt1 negatively regulates the differentiation of interleukin-9-producing cd4+ t cells. *Immunity*, 44(6) :1337–1349, 2016.
- [105] L. E. Botzer, S. Maman, O. Sagi-Assif, T. Meshel, I. Nevo, I. Yron, and I. P. Witz. Hexokinase 2 is a determinant of neuroblastoma metastasis. *British Journal of Cancer*, 114(7) :759–766, 2016.
- [106] Alexander E Aleshin, Chenbo Zeng, Hans D Bartunik, Herbert J Fromm, and Richard B Honzatko. Regulation of hexokinase i : crystal structure of recombinant human brain hexokinase complexed with glucose and phosphate11edited by i. a. wilson. *Journal of Molecular Biology*, 282(2) :345 – 357, 1998.
- [107] E.C. Cheung Mor, I. and K.H. Vousden. ontrol of glycolysis through regulation of pfk1 : old friends and recent additions. *Cold Spring Harb. Symp. Quant. Biol.*, 76 :211–216, 2011.
- [108] H. G. Hers and E. Vanschaftingen. Fructose 2,6-bisphosphate 2 years after its discovery. *Biochemical Journal*, 206(1) :1–12, 1982.
- [109] L. Hue and M. H. Rider. Role of fructose 2,6-bisphosphate in the control of glycolysis in mammalian-tissues. *Biochemical Journal*, 245(2) :313–324, 1987.
- [110] M. H. Rider, L. Bertrand, D. Vertommen, P. A. Michels, G. G. Rousseau, and L. Hue. 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase : Head-to-head with a bifunctional enzyme that controls glycolysis. *Biochemical Journal*, 381 :561–579, 2004.
- [111] M. S. Jurica, A. Mesecar, P. J. Heath, W. X. Shi, T. Nowak, and B. L. Stoddard. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure*, 6(2) :195–210, 1998.
- [112] P. Eliasson and J. I. Jonsson. The hematopoietic stem cell niche : Low in oxygen but a nice place to be. *Journal of Cellular Physiology*, 222(1) :17–22, 2010.
- [113] G. C. Brown. Control of respiration and atp synthesis in mammalian mitochondria and cells. *Biochemical Journal*, 284 :1–13, 1992.
- [114] W. Jones and K. Bianchi. Aerobic glycolysis : beyond proliferation. *Front Immunol*, 6 :227, 2015.
- [115] O. Warburg, F. Wind, and E. Negelein. The metabolism of tumors in the body. *Journal of General Physiology*, 8(6) :519–530, 1927.
- [116] A L Lehninger, D L Nelson, and M M Cox. Principles of biochemistry second edition. *Worth ; New York*, 1993.

- [117] C. Valvona, L. Donovan, H. L. Fillmore, and G. J. Pilkington. Inhibition of lactate dehydrogenase a (ldha) using sodium oxamate leads to metabolic changes, decreased growth and migration in medulloblastoma cell lines. *Neuro-Oncology*, 17 :10–10, 2015.
- [118] C. L. Markert, J. B. Shaklee, and G. S. Whitt. Evolution of a gene. *Science*, 189(4197) :102–114, 1975.
- [119] Y. H. Wang, W. J. Israelsen, D. J. Lee, V. W. C. Yu, N. T. Jeanson, C. B. Clish, L. C. Cantley, M. G. V. Heiden, and D. T. Scadden. Cell-state-specific metabolic dependency in hematopoiesis and leukemogenesis. *Cell*, 158(6) :1309–1323, 2014.
- [120] S. Daniele, C. Giacomelli, E. Zappelli, C. Granchi, M. L. Trincavelli, F. Minutolo, and C. Martini. Lactate dehydrogenase-a inhibition induces human glioblastoma multiforme stem cell differentiation and death. *Scientific Reports*, 5, 2015.
- [121] Xinde Zheng, Leah Boyer, Mingji Jin, Jerome Mertens, Yongsung Kim, Li Ma, Li Ma, Michael Hamm, Fred H. Gage, and Tony Hunter. Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *eLife*, 5 :e13374, 2016.
- [122] J. Y. An, Y. Zhang, J. J. He, Z. L. Zang, Z. Zhou, X. D. Pei, X. Zheng, W. H. Zhang, H. Yang, and S. Li. Lactate dehydrogenase a promotes the invasion and proliferation of pituitary adenoma. *Scientific Reports*, 7, 2017.
- [123] Yongjie Zhang, Xia Zhang, Xi Wang, Lu Gan, Guanzhen Yu, Ying Chen, Ke Liu, Ping Li, Jun Pan, Jiejun Wang, and Shukui Qin. Inhibition of ldh-a by lentivirus-mediated small interfering rna suppresses intestinal-type gastric cancer tumorigenicity through the downregulation of oct4. *Cancer Letters*, 321(1) :45 – 54, 2012.
- [124] Eric J. Rellinger, Brian T. Craig, Alexandra L. Alvarez, Haley L. Dusek, Kwang W. Kim, Jingbo Qiao, and Dai H. Chung. Fx11 inhibits aerobic glycolysis and growth of neuroblastoma cells. *Surgery*, 161(3) :747 – 752, 2017.
- [125] Gregory M. Kelly and Mohamed I Gatie. Mechanisms regulating stemness and differentiation in embryonal carcinoma cells. *Stem Cells International*, page 20, 2017.
- [126] John D. Firth, Benjamin L. Ebert, and Peter J. Ratcliffe. Hypoxic regulation of lactate dehydrogenase a : Interaction between hypoxia-inducible factor 1 and camp response elements. 270(36) :21021–21027, 1995.

- [127] G. L. Semenza. Hif-1 : upstream and downstream of cancer metabolism. *Current Opinion in Genetics Development*, 20(1) :51–56, 2010.
- [128] J. W. Chen, N. Madamanchi, N. R. Madamanchi, T. T. Trier, and M. J. Keherly. Lamp-1 is upregulated in human glioblastoma cell lines induced to undergo apoptosis. *J Biomed Sci*, 8(4) :365–74, 2001. p20.
- [129] Jung whan Kim, Irina Tchernyshyov, Gregg L. Semenza, and Chi V. Dang. Hif-1-mediated expression of pyruvate dehydrogenase kinase : A metabolic switch required for cellular adaptation to hypoxia. *Cell Metabolism*, 3(3) :177 – 185, 2006.
- [130] T. Simsek, F. Kocabas, J. K. Zheng, R. J. DeBerardinis, A. I. Mahmoud, E. N. Olson, J. W. Schneider, C. C. Zhang, and H. A. Sadek. The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell*, 7(3) :380–390, 2010.
- [131] I. Koukourakis Michael, Giatromanolaki Alexandra, Bougioukas Georgios, and Sivridis Efthimios. Lung cancer : An organized cellular and metabolic domain. *Cancer Biology & Therapy*, 6(9) :1472–1475, 2007.
- [132] Alexandra Giatromanolaki, Michael I Koukourakis, Anastasios Koutsopoulos, Savvas Mendrinos, and Efthimios Sivridis. The metabolic interactions between tumor cells and tumor-associated stroma (tas) in prostatic cancer. *Cancer Biol Ther*, 13(13) :1284–1289, 2012.
- [133] Andrew P. Halestrap. The monocarboxylate transporter family—structure and functional characterization. *IUBMB Life*, 64(1) :1–9, 2012.
- [134] Chunzhu Tang and Wei Liu. Ldha is a feedback activator of hypoxia inducible factor 1-alpha in ovarian cancer. *Int J Clin Exp Pathol*, 9(10) :10437–10443, 2016.
- [135] P. Sonveaux, T. Copetti, C. J. De Saedeleer, F. Vegran, J. Verrax, K. M. Kennedy, E. J. Moon, S. Dhup, P. Danhier, F. Frerart, B. Gallez, A. Ribeiro, C. Michiels, M. W. Dewhirst, and O. Feron. Targeting the lactate transporter mct1 in endothelial cells inhibits lactate-induced hif-1 activation and tumor angiogenesis. *Plos One*, 7(3), 2012.
- [136] S. R. Sun, H. Li, J. H. Chen, and Q. Qian. Lactic acid : No longer an inert and end-product of glycolysis. *Physiology*, 32(6) :453–463, 2017.
- [137] C. F. Cori, G. T. Cori, and K. W. Buchwald. The mechanism of epinephrine action. *American Journal of Physiology-Legacy Content*, 93(1) :273–283, 1930.

- [138] H. E. Himwich, Y. D. Koskoff, and L. H. Nahum. Studies in carbohydrate metabolism. i. a glucose-lactic acid cycle involving muscle and liver. *Journal of Biological Chemistry*, 85(2) :571–584, 1930.
- [139] C. K. Garcia, X. Li, J. Luna, and U. Francke. Cdna cloning of the human monocarboxylate transporter-1 and chromosomal localization of the slc16a1 locus to 1p13.2-p12. *Genomics*, 23(2) :500–503, 1994.
- [140] A. S. Divakaruni and A. N. Murphy. A mitochondrial mystery, solved. *Science*, 337(6090) :41–43, 2012.
- [141] S. Papa, P. L. Martino, G. Capitanio, A. Gaballo, D. De Rasmio, A. Signorile, and V. Petruzzella. The oxidative phosphorylation system in mammalian mitochondria. *Advances in Mitochondrial Medicine*, 942 :3–37, 2012.
- [142] L. Yu, S. Yang, Y. Yin, X. Cen, F. Zhou, D. Xia, and C. A. Yu. Chapter 25 analysis of electron transfer and superoxide generation in the cytochrome bc1 complex. *Methods Enzymol*, 456 :459–73, 2009.
- [143] Peter Mitchell. The protonmotive q cycle : A general formulation. *FEBS Letters*, 59(2) :137–139, 1975.
- [144] E. Racker. A mitochondrial factor conferring oligomycin sensitivity on soluble mitochondrial atpase. *Biochem Biophys Res Commun*, 10 :435–9, 1963.
- [145] E. Racker. A reconstituted system of oxidative phosphorylation. *Biochemical and Biophysical Research Communications*, 14(1) :75–, 1963.
- [146] Werner J.H. Koopman, Leo G.J. Nijtmans, Cindy E.J. Dieteren, Peggy Roestenberg, Federica Valsecchi, Jan A.M. Smeitink, and Peter H.G.M. Willems. Mammalian mitochondrial complex i : Biogenesis, regulation, and reactive oxygen species generation. *Biochemical and Biophysical Research Communications*, 12(12) :1431–1470, 2009.
- [147] S. Varum, O. Momcilovic, C. Castro, A. Ben-Yehudah, J. Ramalho-Santos, and C. S. Navara. Enhancement of human embryonic stem cell pluripotency through inhibition of the mitochondrial respiratory chain. *Stem Cell Research*, 3(2-3) :142–156, 2009.
- [148] H. A. Lardy, D. Johnson, and W. C. McMurray. Antibiotics as tools for metabolic studies .1. a survey of toxic antibiotics in respiratory, phosphorylative and glycolytic systems. *Archives of Biochemistry and Biophysics*, 78(2) :587–597, 1958.

- [149] M. S. Patel and T. E. Roche. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *FASEB J*, 4(14) :3224–33, 1990.
- [150] A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev, and B. E. Bernstein. Single-cell rna-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, 344(6190) :1396–401, 2014. p40.
- [151] V. Hatzimanikatis, C. H. Li, J. A. Ionita, and L. J. Broadbelt. Metabolic networks : enzyme function and metabolite structure. *Current Opinion in Structural Biology*, 14(3) :300–306, 2004.
- [152] T. Ezashi, P. Das, and R. M. Roberts. Low o₂ tensions and the prevention of differentiation of hes cells. *Proc Natl Acad Sci U S A*, 102(13) :4783–8, 2005.
- [153] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6) :1106–17, 2008. p30.
- [154] F. Kocabas, L. Xie, J. Xie, Z. Yu, R. J. DeBerardinis, W. Kimura, S. Thet, A. F. Elshamy, H. Abouellail, S. Muralidhar, X. Liu, C. Chen, H. A. Sadek, C. C. Zhang, and J. Zheng. Hypoxic metabolism in human hematopoietic stem cells. *Cell Biosci*, 5 :39, 2015.
- [155] S. Y. Lunt and M. G. Vander Heiden. Aerobic glycolysis : meeting the metabolic requirements of cell proliferation. *Annu Rev Cell Dev Biol*, 27 :441–64, 2011.
- [156] J. S. Baker, M. C. McCormick, and R. A. Robergs. Interaction among skeletal muscle metabolic energy systems during intense exercise. *J Nutr Metab*, 2010 :905612, 2010.
- [157] A. D. Panopoulos, O. Yanes, S. Ruiz, Y. S. Kida, D. Diep, R. Tautenhahn, A. Herrerias, E. M. Batchelder, N. Plongthongkum, M. Lutz, W. T. Berggren, K. Zhang, R. M. Evans, G. Siuzdak, and J. C. Izpisua Belmonte. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res*, 22(1) :168–77, 2012.
- [158] M. J. Rodriguez-Colman, M. Schewe, M. Meerlo, E. Stigter, J. Gerrits, M. Pras-Raves, A. Sacchetti, M. Hornsveld, K. C. Oost, H. J. Snippert, N. Verhoeven-Duif, R. Fodde, and B. M. T.

- Burgering. Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature*, 543(7645) :424–+, 2017.
- [159] J. Isern, Z. He, S. T. Fraser, S. Nowotschin, A. Ferrer-Vaquer, R. Moore, A. K. Hadjantonakis, V. Schulz, D. Tuck, P. G. Gallagher, and M. H. Baron. Single-lineage transcriptome analysis reveals key regulatory pathways in primitive erythroid progenitors in the mouse embryo. *Blood*, 117(18) :4924–34, 2011.
- [160] M. H. Baron, J. Isern, and S. T. Fraser. The embryonic origins of erythropoiesis in mammals. *Blood*, 119(21) :4828–37, 2012.
- [161] L. Galluzzi, O. Kepp, M. G. Vander Heiden, and G. Kroemer. Metabolic targets for cancer therapy. *Nature Reviews Drug Discovery*, 12(11) :829–846, 2013.
- [162] Metabolic regulation of hematopoietic stem cell commitment and erythroid differentiation. *Current Opinion in Hematology*, 23(3), 2016.
- [163] A. Moussaieff, N. M. Kogan, and D. Aberdam. Concise review : Energy metabolites : Key mediators of the epigenetic state of pluripotency. *Stem Cells*, 33(8) :2374–2380, 2015.
- [164] A. Moussaieff, M. Rouleau, D. Kitsberg, M. Cohen, G. Levy, D. Barasch, A. Nemirovski, S. Shen-Orr, I. Laevsky, M. Amit, D. Bomze, B. Elena-Herrmann, T. Scherf, M. Nissim-Rafinia, S. Kempa, J. Itskovitz-Eldor, E. Meshorer, D. Aberdam, and Y. Nahmias. Glycolysis-mediated changes in acetyl-coa and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab*, 21(3) :392–402, 2015.
- [165] M. Peng, N. Yin, S. Chhangawala, K. Xu, C. S. Leslie, and M. O. Li. Aerobic glycolysis promotes t helper 1 cell differentiation through an epigenetic mechanism. *Science*, 354(6311) :481–484, 2016.
- [166] Susan Chung, Petras P Dzeja, Randolph S Faustino, Carmen Perez-Terzic, Atta Behfar, and Andre Terzic. Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. *Nature Clinical Practice Cardiovascular Medicine*, 4 :S60, 2007.
- [167] S. Mandal, A. G. Lindgren, A. S. Srivastava, A. T. Clark, and U. Banerjee. Mitochondrial function controls proliferation and early differentiation potential of embryonic stem cells. *Stem Cells*, 29(3) :486–495, 2011.
- [168] M. A. Reid, Z. Dai, and J. W. Locasale. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat Cell Biol*, 19(11) :1298–1306, 2017.

- [169] Mengtao Xiao, Hui Yang, Wei Xu, Shenghong Ma, Huaipeng Lin, Honguang Zhu, Lixia Liu, Ying Liu, Chen Yang, Yanhui Xu, Shimin Zhao, Dan Ye, Yue Xiong, and Kun-Liang Guan. Inhibition of -kg-dependent histone and dna demethylases by fumarate and succinate that are accumulated in mutations of fh and sdh tumor suppressors. *Genes Dev*, 26(12) :1326–1338, 2012.
- [170] G. J. Guo, L. Pinello, X. P. Han, S. J. Lai, L. Shen, T. W. Lin, K. Y. Zou, G. C. Yuan, and S. H. Orkin. Serum-based culture conditions provoke gene expression variability in mouse embryonic stem cells as revealed by single-cell analysis. *Cell Reports*, 14(4) :956–965, 2016.
- [171] C. Lu and C. B. Thompson. Metabolic regulation of epigenetics. *Cell Metabolism*, 16(1) :9–17, 2012.
- [172] S. Imai, F. B. Johnson, R. A. Marciniak, M. McVey, P. U. Park, and L. Guarente. Sir2 : an nad-dependent histone deacetylase that connects chromatin silencing, metabolism, and aging. *Cold Spring Harb Symp Quant Biol*, 65 :297–302, 2000.
- [173] V. Calvanese, E. Lara, B. Suarez-Alvarez, R. Abu Dawud, M. Vazquez-Chantada, M. L. Martinez-Chantar, N. Embade, P. Lopez-Nieva, A. Horrillo, A. Hmadcha, B. Soria, D. Piazzolla, D. Herranz, M. Serrano, J. M. Mato, P. W. Andrews, C. Lopez-Larrea, M. Esteller, and M. F. Fraga. Sirtuin 1 regulation of developmental genes during differentiation of stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(31) :13736–13741, 2010.
- [174] Timour Prozorovski, Ulf Schulze-Topphoff, Robert Glumm, Jan Baumgart, Friederike Schröter, Olaf Ninnemann, Elise Siegert, Ivo Bendix, Oliver Brüstle, Robert Nitsch, Frauke Zipp, and Orhan Aktas. Sirt1 contributes critically to the redox-dependent fate of neural progenitors. *Nature Cell Biology*, 10 :385, 2008.
- [175] M. J. Son, M. Y. Son, B. Seol, M. J. Kim, C. H. Yoo, M. K. Han, and Y. S. Cho. Nicotinamide overcomes pluripotency deficits and reprogramming barriers. *Stem Cells*, 31(6) :1121–1135, 2013.
- [176] J. G. Ryall, T. Cliff, S. Dalton, and V. Sartorelli. Metabolic reprogramming of stem cell epigenetics. *Cell Stem Cell*, 17(6) :651–662, 2015.
- [177] Jinbeom Heo, Jisun Lim, Seungun Lee, Jaeho Jeong, Hyunsook Kang, YongHwan Kim, Jeong Wook Kang, Hwan Yeul Yu, Eui Man Jeong, Kyunggon Kim, Magda Kucia, Sabine J. Waigel, Wolfgang Zacharias, Yinlu Chen, In-Gyu Kim, Mariusz Z. Ratajczak, and Dong-

- Myung Shin. Sirt1 regulates dna methylation and differentiation potential of embryonic stem cells by antagonizing dnmt3l. *Cell Reports*, 18(8) :1930 – 1945, 2017.
- [178] C. D. L. Folmes, P. P. Dzeja, T. J. Nelson, and A. Terzic. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell*, 11(5) :596–606, 2012.
- [179] Olivier Gandrillon, Uwe Schmidt, Hartmut Beug, and Jacques Samarut. Tgf- cooperates with tgf- to induce the self-renewal of normal erythrocytic progenitors : evidence for an autocrine mechanism. *The EMBO Journal*, 18(10) :2764–2781, 1999.
- [180] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A*, 99(20) :12795–800, 2002. p39.
- [181] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj. Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms. *Mol Cell*, 58(2) :339–52, 2015. p40.
- [182] Anissa Guillemin, Angélique Richard, Sandrine Gonin-Giraud, and Olivier Gandrillon. Automated cell cycle and cell size measurements for single-cell gene expression studies. *BMC Research Notes*, 11(1) :92, Feb 2018.
- [183] Alice Moussy, Jérémie Cosette, Romuald Parmentier, Cindy da Silva, Guillaume Corre, Angélique Richard, Olivier Gandrillon, Daniel Stockholm, and András Páldi. Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLOS Biology*, 15(7) :1–23, 07 2017.
- [184] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9 :2579–2605, 2008.
- [185] Mitra Mojtahedi, Alexander Skupin, Joseph Zhou, Ivan G. Castaño, Rebecca Y. Y. Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang. Cell fate decision as high-dimensional critical state transition. *PLOS Biology*, 14(12) :1–28, 12 2016.
- [186] Stefan Semrau, Johanna E. Goldmann, Magali Soumillon, Tarjei S. Mikkelsen, Rudolf Jaenisch, and Alexander van Oudenaarden. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature Communications*, 8(1) :1096, 2017.

- [187] Magali Soumillon, Davide Cacchiarelli, Stefan Semrau, Alexander van Oudenaarden, and Tarjei S Mikkelsen. Characterization of directed differentiation by high-throughput single-cell rna-seq. *bioRxiv*, 2014.
- [188] R. Liu, P. Chen, K. Aihara, and L. Chen. Identifying early-warning signals of critical transitions with strong noise by dynamical network markers. *Sci Rep*, 5 :17501, 2015. Liu, Rui Chen, Pei Aihara, Kazuyuki Chen, Luonan eng Research Support, Non-U.S. Gov't England 2015/12/10 06 :00 Sci Rep. 2015 Dec 9;5 :17501. doi : 10.1038/srep17501.
- [189] I. El Meouche, Y. Siu, and M. J. Dunlop. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Scientific Reports*, 6 :19538, 2016.
- [190] Erez Braun. The unforeseen challenge : from genotype-to-phenotype in cell populations. *Reports on Progress in Physics*, 78(3) :036602, 2015.
- [191] Carlotta Granchi, Daniele Fancelli, and Filippo Minutolo. An update on therapeutic opportunities offered by cancer glycolytic metabolism. *Bioorganic Medicinal Chemistry Letters*, 24(21) :4915 – 4925, 2014.
- [192] Cara J. Valvona, Helen L. Fillmore, Peter B. Nunn, and Geoffrey J. Pilkington. The regulation and function of lactate dehydrogenase a : Therapeutic potential in brain tumor. *Brain Pathology*, 26(1) :3–17, 2016.
- [193] L. Oburoglu, S. Tardito, V. Fritz, S. C. de Barros, P. Merida, M. Craveiro, J. Mamede, G. Cretenet, C. Mongellaz, X. An, D. Klysz, J. Touhami, M. Boyer-Clavel, J. L. Battini, V. Dardalhon, V. S. Zimmermann, N. Mohandas, E. Gottlieb, M. Sitbon, S. Kinet, and N. Taylor. Glucose and glutamine metabolism regulate human hematopoietic stem cell lineage specification. *Cell Stem Cell*, 15(2) :169–84, 2014.
- [194] Andras Paldi. What makes the cell differentiate? *Progress in Biophysics and Molecular Biology*, 110(1) :41 – 43, 2012. Special Issue : Chance at the heart of the cell.
- [195] Hui Zhang, Mehmet G. Badur, Ajit S. Divakaruni, Seth J. Parker, Christian Jäger, Karsten Hiller, Anne N. Murphy, and Christian M. Metallo. Distinct metabolic states can support self-renewal and lipogenesis in human pluripotent stem cells under different culture conditions. *Cell Reports*, 16(6) :1536–1547, 2016.

- [196] Daiki Karigane and Keiyo Takubo. Metabolic regulation of hematopoietic and leukemic stem/progenitor cells under homeostatic and stress conditions. *International Journal of Hematology*, 106(1) :18–26, Jul 2017.
- [197] C. Damiani, R. Colombo, D. Gaglio, F. Mastroianni, D. Pescini, H. V. Westerhoff, G. Mauri, M. Vanoni, and L. Alberghina. A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth : The warburq effect. *PLoS Computational Biology*, 13(9) :1–29, 09 2017.
- [198] I. L. de la Serna, Y. Ohkawa, and A. N. Imbalzano. Chromatin remodelling in mammalian differentiation : lessons from atp-dependent remodellers. *Nature Reviews Genetics*, 7 :461.
- [199] Berry Juliandi, Masahiko Abematsu, and Kinichi Nakashima. Chromatin remodeling in neural stem cell differentiation. *Current Opinion in Neurobiology*, 20(4) :408 – 415, 2010. Signalling mechanisms.
- [200] Shuying He, Saima Limi, Rebecca S. McGreal, Qing Xie, Lisa A. Brennan, Wanda Lee Kantorow, Juraj Kokavec, Romit Majumdar, Harry Hou, Winfried Edelmann, Wei Liu, Ruth Ashery-Padan, Jiri Zavadil, Marc Kantorow, Arthur I. Skoultchi, Tomas Stopka, and Ales Cvekl. Chromatin remodeling enzyme snf2h regulates embryonic lens differentiation and de-nucleation. *Development*, 143(11) :1937–1947, 2016.
- [201] S. K. Hota and B. G. Bruneau. Atp-dependent chromatin remodeling during mammalian development. *Development*, 143(16) :2882–97, 2016.
- [202] J. Huang, X. Miao, W. Jin, P. Couble, K. Mita, Y. Zhang, W. Liu, L Zhuang, Y. Shen, C. Keime, O. Gandrillon, P. Brouilly, J. Briolay, G. Zhao, and Y. Huang. Serial analysis of gene expression (sage) in the silkworm, bombyx mori. *Genomics*, 86 :233–241, 2005.
- [203] Craig L Peterson. Atp-dependent chromatin remodeling : going mobile. *FEBS Letters*, 476(1-2) :68–72, 2000.
- [204] Anthony A Sauve and Dou Yeon Youn. Sirtuins : Nad⁺-dependent deacetylase mechanism and regulation. *Current Opinion in Chemical Biology*, 16(5) :535 – 543, 2012. Mechanisms• Aesthetics• Molecular imaging.
- [205] H. Shimano and R. Sato. Srebp-regulated lipid metabolism : convergent physiology — divergent pathophysiology. *Nature Reviews Endocrinology*, 13 :710, 2017.
- [206] Simone Fulda and Klaus-Michael Debatin. Hif-1-regulated glucose metabolism : A key to apoptosis resistance? *Cell Cycle*, 6(7) :790–792, 2007. PMID : 17404504.

- [207] T. Porstmann, C. R. Santos, B. Griffiths, M. Cully, M. Wu, S. Leevers, J. R. Griffiths, Y. Chung, and A. Schulze. Srebp activity is regulated by mtorc1 and contributes to akt-dependent cell growth. *Cell Metab*, 8(3-3) :224–236, 2008.
- [208] S. Cheng, J. Quintin, R. A. Cramer, K. M. Shepardson, S. Saeed, V. Kumar, Evangelos J. Giamarellos-Bourboulis, Joost H. A. Martens, Nagesha Appukudige Rao, Ali Aghajani-refah, Ganesh R. Manjeri, Yang Li, Daniela C. Ifrim, Rob J. W. Arts, Brian M. J. W. van der Veer, Peter M. T. Deen, Colin Logie, Luke A. O’Neill, P. Willems, F. L. van de Veerdonk, J. W. M. van der Meer, A. Ng, L. A. B. Joosten, C. Wijmenga, H. G. Stunnenberg, R. J. Xavier, and M. G. Netea. mtor- and hif-1-mediated aerobic glycolysis as metabolic basis for trained immunity. 345(6204), 2014.
- [209] Jason S. L. Yu and Wei Cui. Proliferation, survival and metabolism : the role of pi3k/akt/mtor signalling in pluripotency and cell fate determination. *Development*, 143(17) :3050–3060, 2016.
- [210] J. C. Rathmell, C. J. Fox, D. R. Plas, P. S. Hammerman, R. M. Cinalli, and C. B. Thompson. Akt-directed glucose metabolism can prevent bax conformation change and promote growth factor-independent survival. *Mol Cell Biol*, 23(20) :7315–7328, 2003.
- [211] Francesca R Dejure and Martin Eilers. Myc and tumor metabolism : chicken and egg. *The EMBO Journal*, 36(23) :3409–3420, 2017.