Université Paris Cité

Ecole doctorale Hématologie, Oncogenèse et Biothérapies (ED561)

Laboratoire INSERM U1170, Institut Gustave Roussy

Développement et caractérisation d'un modèle humain de leucémie pédiatrique ETO2-GLIS2 à partir de cellules souches pluripotentes induites

Par Fabien Boudia

Thèse de doctorat d'Hématologie

Dirigée par Dr. Thomas Mercher

Présentée et soutenue publiquement le 7 décembre 2022

Devant un jury composé de :

Thierry Jaffredo Directeur de recherche, Sorbonne Université	Rapporteur
Isabelle Janoueix-Lerosey Directrice de recherche, Université PSL	Rapportrice
Michaela Fontenay PU-PH, Université Paris Cité	Examinatrice
Erika Brunet Directrice de recherche, Université Paris Cité	Examinatrice
Eric Delabesse, PU, Université Toulouse 3	Examinateur
Thomas Mercher Directeur de recherche, Université Paris Sud	Directeur de Thèse

Remerciements

Je tiens à remercier le Dr. Thierry Jaffredo et le Dr. Isabelle Janoueix-Lerosey pour avoir accepté d'évaluer cette thèse ainsi que le Dr. Erika Brunet, le Pr. Michaela Fontenay et le Pr. Eric Delabesse pour avoir accepté de faire partie de mon jury de thèse.

Je remercie Thomas Mercher de m'avoir accueilli dans son équipe dès le master et d'avoir poursuivi durant cette thèse. Ta capacité de travail et ton esprit scientifique m'ont beaucoup impressionné et j'ai essayé de m'en inspirer pour progresser. Merci de m'avoir fait confiance pour développer un nouvel outil que sont les iPSC dans le labo. Merci également d'avoir tout le temps eu une attitude positive quand je te présentais des résultats ou une manip que je considérais comme ayant moyennement fonctionné mais où tu voyais le verre à moitié plein.

Merci à tous les membres passés et présents de l'équipe. Le fait de savoir que j'allais vous retrouver au labo m'a vraiment aidé à me lever certains matins. Un énorme merci à Zakia. Travailler avec toi fut vraiment une superbe expérience. Sur le plan professionnel, c'est toi qui m'as appris à maniper et je n'ai jamais rencontré quelqu'un d'aussi dédiée à son travail. Sur le plan humain, j'ai apprécié chacune de nos conversations autant sur des sujets légers que sur des thèmes plus profonds. Merci à ma colloc, Marie, pour ta bonne humeur et ton aide pour les manips. J'ai apprécié te « titiller » sur certains sujets car j'aime beaucoup la manière passionnée que tu as de défendre tes opinions. Accroche-toi, je suis sûr que tu feras une superbe thèse et que tu deviendras un médecin génial (j'ai un bon feeling). Mes illustres prédécesseuses, Cécile qui m'a montré qu'il est possible de rigoler non-stop de 9h à 20h tout en produisant un travail de qualité. Merci d'avoir répondu à toutes mes questions (certaines plusieurs fois) avec pédagogie et patience. Anouchka, pour m'avoir accepté « chez toi » durant les premières années de ma thèse. Merci de m'avoir forcé la main pour qu'on parle (et ce n'était pas facile !), j'ai beaucoup apprécié nos discussions et je suis content qu'elles aient continué après la fin de ta thèse. Merci à mon poto Brahim, ces centaines (milliers ?) d'heures passées à cultiver les iPSC étaient moins relou quand tu étais là. J'ai vraiment apprécié nos conversations, aussi bien les discussions scientifiques très pointues que nous avons parfois eu, que les conversations sur la politique nationale et internationale (et un pays en particulier si tu vois ce que je veux dire DZ DZ DZ). Et je n'oublie pas que c'est grâce à ton travail méticuleux et méthodique que les différenciations ont fonctionné ! Merci à Melanight, ma compatriote du 91, pour la bonne humeur constante. J'ai beaucoup aimé nos pauses goûter où on se raconte le moindre petit ragot. Merci à Maroussia, grâce à toi je ne suis plus le seul jeune de l'équipe. Merci à Elie pour les analyses bio-info et les blagues (douteuses). Merci à Julie,

Muriel, Melchior, Aurélie, Alexandre, Laetitia pour tous les conseils et les discussions tout au long de cette thèse.

Merci à tous les membres de l'unité avec qui j'ai eu l'occasion de travailler ou simplement de discuter : Amina, Laure, Paule, Amaury, Seb, Michael et tous les autres...

Je tiens aussi à remercier toutes les personnes qui travaillent sur les iPSC dans les unités d'hématologie à l'IGR et avec qui j'ai passé du temps en culture. D'abord Hana Raslova qui a accepté que de partager la pièce de culture de son équipe et qui a participé à tous les comités de suivi durant ma thèse. Alessandro, qui m'a impressionné par son esprit scientifique et son fanatisme pour les iPSC. Marie Cambot, qui m'a encadré durant mon master 2 et m'a tout appris sur la culture et la différenciation des iPSC, toujours dans la bonne humeur. Et un encouragement pour Nasrine, la jeune génération des iPSC.

Je remercie également nos collaborateurs du CEA, Françoise et Véronica. Et un mot particulier pour Klaudia, j'ai beaucoup apprécié travailler avec toi, tu es bosseuse et rigoureuse, ça va payer !

Merci également à Erika Brunet et Loélia Babin qui m'ont permis d'obtenir le premier clone iPSC avec l'inversion. Sans vous, les travaux de cette thèse n'auraient pas pu avoir lieu.

Je tiens également à remercier les membres des différentes plateformes de l'IGR. En particulier je remercie, Yann, Cyril & Philipe, toujours arrangeants pour trouver un créneau de tri, même à la dernière minute. Également un grand merci à Laure qui a pris soin de mes petites souris de façon très attentionnée.

Enfin un grand merci à ma maman pour les week-ends où je suis rentré manger ses bons petits plats et qui m'ont permis de faire une pause dans cette course interminable qu'est la thèse.

(Et merci à maman et à Marie d'avoir corrigé les nombreuses fautes d'orthographe de cette thèse).

Table des matières

LISTE DES ABREVIATIONS 6

INDEX DES FIGURES 9

INTRODUCTION 10

I) HEMATOPOÏESE NORMALE 11

1) ARCHITECTURE DE L'HEMATOPOÏESE 11

a) Cellules souches hématopoïétiques et modèles de différenciation 11

b) Mégacaryopoïèse 14

c) Érythropoïèse et autres lignées myéloïdes 17

2) DEVELOPPEMENT DU SYSTEME HEMATOPOÏETIQUE 18

a) Rôle de l'endothélium dans l'émergence des cellules hématopoïétiques 18

c) Hématopoïèse pro-définitive 22

d) Hématopoïèse définitive 23

3) REGULATION DE L'HEMATOPOÏESE 25

a) Microenvironnement 25

b) Voies de signalisation 27

i) Interaction cellule-cellule : exemple de la voie NOTCH 27

ii) Cytokines : exemple de la thrombopoïétine 29

c) Transcription 30

d) Régulation épigénétique 35

i) Modifications des histones 37

ii) Méthylation de l'ADN 38

iii) Organisation tridimensionnelle de la chromatine 39

e) Intégration des différents niveaux de régulation transcriptionnelle 41

4) DIFFERENCES HEMATOPOÏESE *IN UTERO* VS ADULTE 42

II) HEMATOPOÏESE PATHOLOGIQUE 44

1) LEUCEMIES AIGUËS MYELOÏDES PEDIATRIQUES 44

2) GENETIQUE ET CONSEQUENCES FONCTIONNELLES DES LAM PEDIATRIQUES 45

a) Types d'altérations 46

b) Description des mutations et conséquences fonctionnelles 47

c) Évolution clonale et Coopération oncogénique 48

d) Cellule d'origine de la leucémie 50

e) Différences leucémies pédiatriques et leucémies de l'adulte 52

3) LAM7 PEDIATRIQUES 55

a) LAM7 associées aux trisomies 21 constitutives 55

b) LAM7 de novo 57

c) ETO2-GLIS2 60

III) LES CELLULES PLURIPOTENTES POUR L'ETUDE DE L'HEMATOPOÏESE HUMAINE 63

1) MODELES D'ETUDE EN HEMATOLOGIE 63

2) CONCEPT DE PLURIPOTENCE 65

3) REPROGRAMMATION 67

4) IPSC EN HEMATOLOGIE 69

a) Protocoles de différenciation hématopoïétique 69

b) Modélisation des hémopathies malignes 73

RESULTATS 78

CONTEXTE ET OBJECTIFS 79

ARTICLE 1 84

ARTICLE 2 99

DISCUSSION 131

ANNEXES 143

BIBLIOGRAPHIE 161

Liste des abréviations

A

AGM : Aorte-Gonade-Mésonéphros

ATAC-seq : Test d'accessibilité de la chromatine par la transposase suivi de séquençage

C

ChIP-seq : Immunoprécipitation de la chromatine suivie de séquençage

D

DNMT : ADN methyltransférase

DS-LAM7 : Leucémie myéloïde aiguë associée au syndrome de Down

E

E : Jour du développement embryonnaire chez la souris

EB : Corps embryoïdes

EHT : Transition endothélio-hématopoiétique

EMP : Progéniteur érythro-myéloïde

Ery : Érythrocyte

Ery-P : Progéniteurs érythroïdes

F

FAB : Classification franco-américano-britannique

FISH : Hybridation fluorescente in situ

FLT3-ITD : Duplication interne en tandem de FLT3

H

HAT : Histones acétyltransférases

HDAC : Histones désacétylases

HE : Endothélium hémogénique

hESC : Cellules souches embryonnaires humaines

HMT : Histones méthyltransférases

HSC : Cellules souches hématopoïétiques

I

IAHC : Clusters hématopoïétiques intra-aortique

IL : Interleukine

iPSC : Cellules souches pluripotentes induites

J

JAK : Janus kinase

L

LAM : Leucémie aiguë myéloïde

LAM7 : Leucémie aiguë mégacaryoblastique

LAL : Leucémie aiguë lymphoïde

LMP : Progéniteur lympho-myéloïde

M

MCI : Masse cellulaire interne

mEpiSC : Cellules souches épiblastiques murines

mESC : Cellules souches embryonnaires murines

MEP : Progéniteur érythro-mégacaryocytaire

Mk : Mégacaryocyte

Mk-P : Progéniteur mégacaryocytaire

MPP : Progéniteurs hématopoïétiques multipotents

MSC : Cellules souches mésenchymateuses

N

NHR : Nervy homology regions

NK : Cellules natural killer

NSG : Souris immunodéficientes NOD.Cg-Prkdc^{scid}IL2rg^{tm1Wjl}/SzJ

O

OSK : Facteurs de reprogrammation OCT4, SOX2, KLF4

OSKM : Facteurs de reprogrammation OCT4, SOX2, KLF4 et c-MYC

P

PDX : Xénogreffes dérivées de cellules de patients

Pol II : ARN Polymérase II

S

scRNA-seq : Séquençage d'ARN à l'échelle de la cellule unique

SMD : Syndrome myélodysplasique

SMP : Syndrome myéloprolifératif

T

T21 : Trisomie 21

TAD : Domaines d'association topologiques

TMD : Désordre myéloïde transitoire

TF : Facteur de transcription

TPO : Thrombopoïétine

V

VAF : Fréquence d'allèle variant

W

WPC : Semaine post-conception

Y

YS : Sac vitellin

Index des figures

Figure 1 : Modèles de la différenciation hématopoïétique. 13
Figure 2 : Maturation mégacaryocytaire. 15
Figure 3 : Hétérogénéité des mégacaryocytes. 16
Figure 4 : Transition endothélio-hématopoiétique. 20
Figure 5 : Hématopoïèse primitive. 21
Figure 6 : Hématopoïèse pré-définitive. 23
Figure 7 : Hématopoïèse définitive. 24
Figure 8 : Niche hématopoïétique des HSC. 26
Figure 9: Voie de signalisation NOTCH. 28
Figure 10 : Activation de la voie de signalisation STAT par la TPO. 30
Figure 11 : Régulation de la transcription. 32
Figure 12 : Complexe heptade. 34
Figure 13 : Organisation de la chromatine. 36
Figure 14 : Impact des modifications d'histones sur la transcription. 38
Figure 15 : Organisation 3D de la chromatine. 40
Figure 16 : Différences entre HSC fœtales et adultes. 43
Figure 17 : Altérations génétiques dans les LAM pédiatriques. 49
Figure 18 : Hypothèse sur l'origine cellulaire des LAM. 52
Figure 19 : Répartition des sous types de LAM selon l'âge. 53
Figure 20 : Modèle d'évolution oncogénique de la DS-LAM7. 56
Figure 21 : Génétique des LAM7 de novo. 58
Figure 22 : Fusion ETO2-GLIS2. 60
Figure 23 : Association entre phénotype et âge chez les patients ETO2-GLIS2. 62
Figure 24 : Pluripotence naïve vs primée. 66
Figure 25 : Différenciation hématopoïétique des iPSC. 70
Figure 26 : Reprogrammation de cellules de patient au stade iPSC. 74
Figure 27 : Reconstruction de la progression clonale grâce à la reprogrammation. 76

Introduction

I) Hématopoïèse Normale

$3{,}3.10^{11}$ cellules sont renouvelées chaque jour dans le corps humain. Près de 90% de ces cellules sont des cellules du sang (*1*). Ces cellules sanguines jouent un rôle dans des processus variés tels que le transport de l'oxygène pour les érythrocytes, la coagulation pour les plaquettes ou encore la défense immunitaire contre les infections pour les leucocytes. Le processus physiologique qui permet la production constante de ces cellules est appelé hématopoïèse et se déroule dans la moelle osseuse.

1) Architecture de l'hématopoïèse

L'hématopoïèse est un processus hiérarchisé dans lequel des cellules souches vont progressivement se spécialiser pour donner des cellules matures fonctionnelles. La compréhension de l'organisation hiérarchique et de la façon dont les cellules souches aboutissent aux cellules matures est essentielle dans le cadre de l'étude des pathologies hématologiques. Les cellules hématopoïétiques appartiennent à différents lignages incluant le lignage mégacaryocytaire, le lignage érythroïde, les autres lignages myéloïdes et les cellules lymphoïdes (qui ne seront pas détaillées dans cette thèse).

a) Cellules souches hématopoïétiques et modèles de différenciation

Dans les années 1960, à travers des expériences de greffe de cellules dans des souris dont la moelle osseuse a été déplétée par irradiation, Till et McCulloch ont démontré l'existence d'un type de cellule capable de reconstituer l'intégralité des cellules hématopoïétiques (*2*). La découverte de ces cellules, appelées cellules souches hématopoïétiques (HSC), a ouvert la voie à l'étude des propriétés permettant le maintien d'un système hématopoïétique fonctionnel tout au long de la vie d'un individu.

Les HSC représentent environ 0,006% des cellules de la moelle osseuse et sont estimées entre $2{,}5.10^{3}$ – $1{,}3.10^{6}$ chez l'homme (*3*). Elles sont caractérisées par 2 propriétés fondamentales que partagent toutes les cellules souches : l'autorenouvellement et la différenciation. L'autorenouvellement est la capacité pour les HSC de se diviser pour donner 2 cellules filles dont au moins une conservera les fonctionnalités de la cellule mère ce qui permet de

maintenir un nombre constant de cellules. Les HSC peuvent également se différencier en tous les types de cellules sanguines : elles sont capables de perdre leurs propriétés de cellules immatures et se spécialiser pour donner de nouveaux types cellulaires. Les HSC les plus immatures sont en majorité quiescentes (*4*) et présentent une activité métabolique réduite (*5*, *6*).

Historiquement, la cytométrie en flux est la technique de référence pour l'étude de la différenciation hématopoïétique. Ainsi, différentes populations ont été définies sur la base des récepteurs présents à la surface de leur membrane plasmique. Ces combinaisons de marqueurs changent régulièrement mais le consensus actuel pour définir les HSC chez l'homme est le phénotype suivant : $Lin^-CD34^+CD38^-CD45RA^-CD90^+CD49f^+$ (*7*).

Le modèle de différentiation « classique » se base notamment sur les travaux du laboratoire de Irving Weissman à la fin des années 1990. Ce modèle propose l'image d'un arbre avec au sommet les HSC, capables de s'auto-renouveler et de se différencier, et à la base les cellules matures. Entre ces différents états, les cellules transitent par différents progéniteurs intermédiaires. Ces progéniteurs représentent des stades clairement définis avec une potentialité précise qui se restreint à chaque étape (*8*, *9*) (Figure 1A).

Ce modèle a été affiné au fil des années avec la définition de nouveaux progéniteurs intermédiaires (*10*, *11*). L'émergence des nouvelles techniques de séquençage et de suivi au niveau clonal des cellules a permis un changement de paradigme. Différentes études utilisant le séquençage d'ARN sur cellule unique (scRNA-seq) montrent que les HSC ne représentent pas un pool homogène mais sont en fait constituées de différentes cellules ayant des potentialités variées (*12*–*17*). Ces études montrent également que bien que certaines HSC soient oligopotentes, la plupart présentent un biais de différentiation conduisant à un engagement précoce vers un seul lignage. Ces différentes observations basées sur le transcriptome, sont également valables pour la conformation de la chromatine (*18*) et sont confirmées expérimentalement par des études de reconstitution de la moelle à partir de cellule unique dans des souris (*19*–*21*) ainsi que par l'utilisation de différents systèmes de code-barres ou de tags permettant de suivre la différenciation des cellules (*17*, *22*–*25*). Dans l'ensemble, toutes les études récentes s'accordent sur le fait qu'il n'existe pas d'états discrets de progéniteurs bien définis mais que la différenciation entre les HSC et les cellules matures se fait sous forme de continuum (Figure 1B).

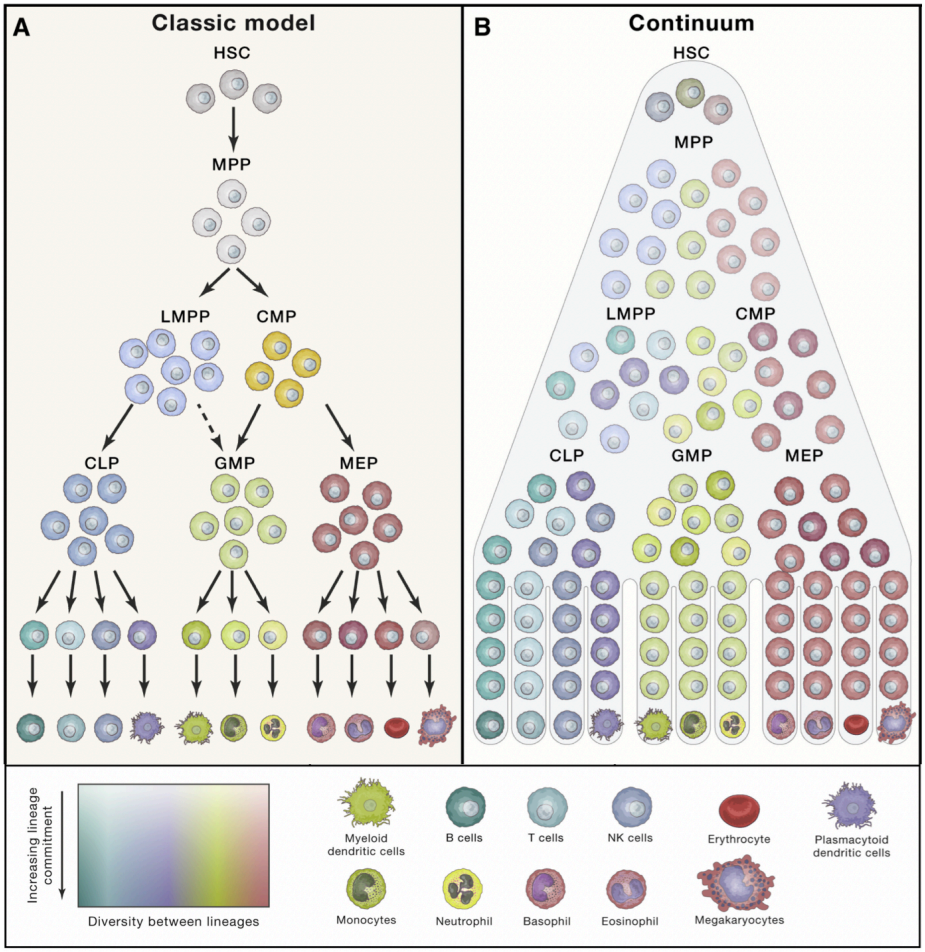


Figure 1 : Modèles de la différenciation hématopoïétique.

(**A**) Vision traditionnelle de l'hématopoïèse selon le modèle de "l'arbre". La différenciation se déroule via une succession d'états discrets à la potentialité de plus en plus restreinte. (**B**) Vision récente de l'hématopoïèse comme un continuum. La différenciation se déroule selon un continuum dans lequel le destin des cellules est spécifié précocement. Adapté de (*26*)

Enfin, comme pour l'étude de nombreux processus biologiques, il est important de noter que l'action de l'expérimentateur influence la réponse biologique obtenue. Ainsi, une observation en condition expérimentale n'est pas toujours un reflet de ce qu'il se passe « au repos ». Par exemple, les études de suivi de différenciation réalisées dans un contexte de greffes après irradiation rapportent une proportion plus importante de HSC oligopotentes capables de reconstituer plusieurs lignages. A l'inverse, les études réalisées dans un contexte natif sans perturbation montrent que la grande majorité des HSC sont unipotentes (*23*, *25*). Ainsi la potentialité des HSC n'est pas une propriété figée des cellules mais plutôt un état qui s'adapte en réponse aux différents stress.

b) Mégacaryopoïèse

Les mégacaryocytes (Mk), qui représentent 0,05-0,1% des cellules de la moelle, sont des cellules beaucoup plus grandes que la moyenne (50-100 µM) dont la fonction principale est la production des plaquettes sanguines. Ces plaquettes, libérées dans le sang interviennent dans divers processus tels que l'hémostase, la cicatrisation ou encore l'inflammation.

L'engagement dans le lignage mégacaryocytaire des HSC afin de produire des progéniteurs mégacaryocytaires (Mk-P) peut être suivi par l'acquisition très précoce du marqueur CD41 (*27*) ainsi que du CD61 puis dans les stades plus avancés le CD42. Le CD41 et CD61 sont 2 sous-unités qui vont s'associer pour former une intégrine qui va, avec le CD42, former un complexe glycoprotéique qui permet aux plaquettes de s'agréger ainsi que d'adhérer à l'endothélium vasculaire. Les Mk, présentent une forte proximité avec les HSC dont ils partagent plusieurs caractéristiques comme des récepteurs de surface (vWF, CD41, MPL, CD150) des régulateurs transcriptionnels (RUNX1, FLI-1, GATA1/2) ou des voies de signalisation (Notch) (*28*). A l'homéostasie, jusqu'à la moitié des HSC présentent un biais vers la différenciation mégacaryocytaire bien qu'elles conservent la capacité à se différencier vers tous les autres lignages (*24*, *25*). En réponse à une inflammation (*29*), à certains néoplasmes (*30*) et même lors du vieillissement normal (*27*) ces HSC se multiplient afin d'augmenter le nombre de Mk-P.

Les Mk-P sont des cellules diploïdes qui prolifèrent et vont ensuite subir une série de changements au niveau cellulaire pour donner les Mk. Tout d'abord, les Mk vont passer par une succession de cycles cellulaires. La mitose de ces cycles, appelée endomitose est particulière car il n'y a ni séparation des cellules (défaut de cytokinèse) ni séparation des noyaux (défaut de karyokinèse). Ce phénomène entraine une polyploïdisation ainsi qu'une augmentation de la masse cytoplasmique des Mk qui ont ainsi une ploïdie pouvant aller de 4 jusqu'à 128N (*31*). Un système membranaire élaboré, le système de démarcation membranaire, va ensuite se mettre en place à partir du Golgi, puis des réorganisations du cytosquelette vont induire des protrusions membranaires (*32*) . Ces protrusions vont traverser la barrière endothéliale pour rejoindre les vaisseaux où, sous l'effet de la circulation sanguine, elles seront fragmentées en proplaquettes (*33*) (Figure 2). Quelques publications suggèrent

qu'un mécanisme alternatif via un « bourgeonnement » du cytoplasme sans protrusions membranaires représenterait la majorité de la production des plaquettes (*34*, *35*) mais ces résultats sont sujet à débat (*36*).

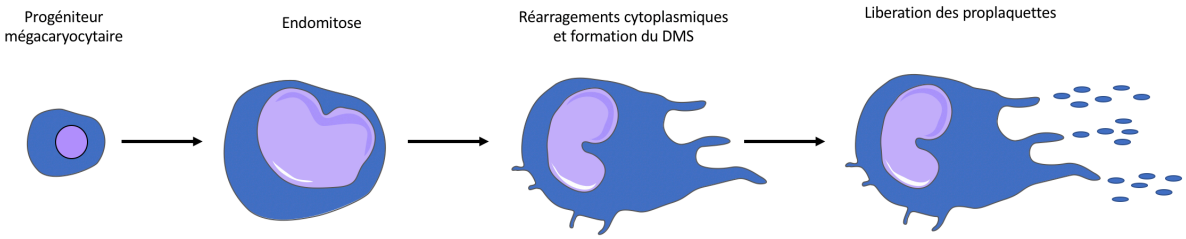


Figure 2 : Maturation mégacaryocytaire.
Maturation du progéniteur mégacaryocytaire qui va subir successivement des étapes de polyploïdisation par endomitose, puis des réarrangements cytoplasmiques et membranaires qui permettent la formation du système de démarcation membranaire (DMS) ainsi que des protrusions. Enfin les protrusions cytoplasmiques vont se fragmenter pour relarguer les proplaquettes dans la circulation sanguine

Dans la vision « traditionnelle » les Mk résident exclusivement dans la moelle osseuse, cependant une étude par Lefrancais *et al.*, dans laquelle les auteurs ont suivi des souris par microscopie photonique, montre une circulation de Mk de la moelle vers les poumons où ils libèreraient jusqu'à 50% du totale des plaquettes présentes dans la circulation sanguines (*37*). Plusieurs articles récents utilisant le scRNA-seq montrent que les Mk sont en réalité hétérogènes avec des populations aux transcriptomes distincts qui reflètent des fonctions différentes (*38–40*). Selon les études le nombre de populations mégacaryocytaires varie mais on retrouve toujours 4 types principaux (Figure 3). Un premier groupe correspond à la fonction « canonique » des mégacaryocytes c'est-à-dire à la production de proplaquettes. Le second groupe correspond à des mégacaryocytes qui participent à la niche des HSC notamment via la sécrétion de molécules permettant la différentiation des ostéoclastes et des fibroblastes ainsi que de facteurs de croissance. Les Mk du troisième groupe présentent un profil immunitaire. Ces Mk expriment un profil transcriptionnel intermédiaire entre les Mk classiques et les cellules immunitaires avec un enrichissement des marqueurs immuns et des gènes associés à l'inflammation. Ils participent à la réponse immunitaire notamment par la présentation d'antigène ce qui permet l'activation des lymphocytes T CD4+ (*41*). Enfin le dernier groupe est constitué de Mk « immatures » avec une faible ploïdie et qui cyclent activement et constituerait un intermédiaire de différenciation servant de pool pour donner les 3 autres

types. Il est intéressant de noter que ces différents profils de Mk pourraient résulter de différence au niveau de leur microenvironnement. En effet, les Mk immatures ainsi que ceux servant de support aux HSC sont localisés dans la moelle osseuse alors que les Mk immuns sont situés au niveau des poumons. Cette spécialisation immunitaire n'est pas due au contact de pathogène car on la retrouve également au stade fœtal où les poumons sont dans un environnement stérile (*42*). Elle pourrait résulter de signaux sécrétés par les cellules pulmonaires comme par exemple l'IL-33 (*41*).

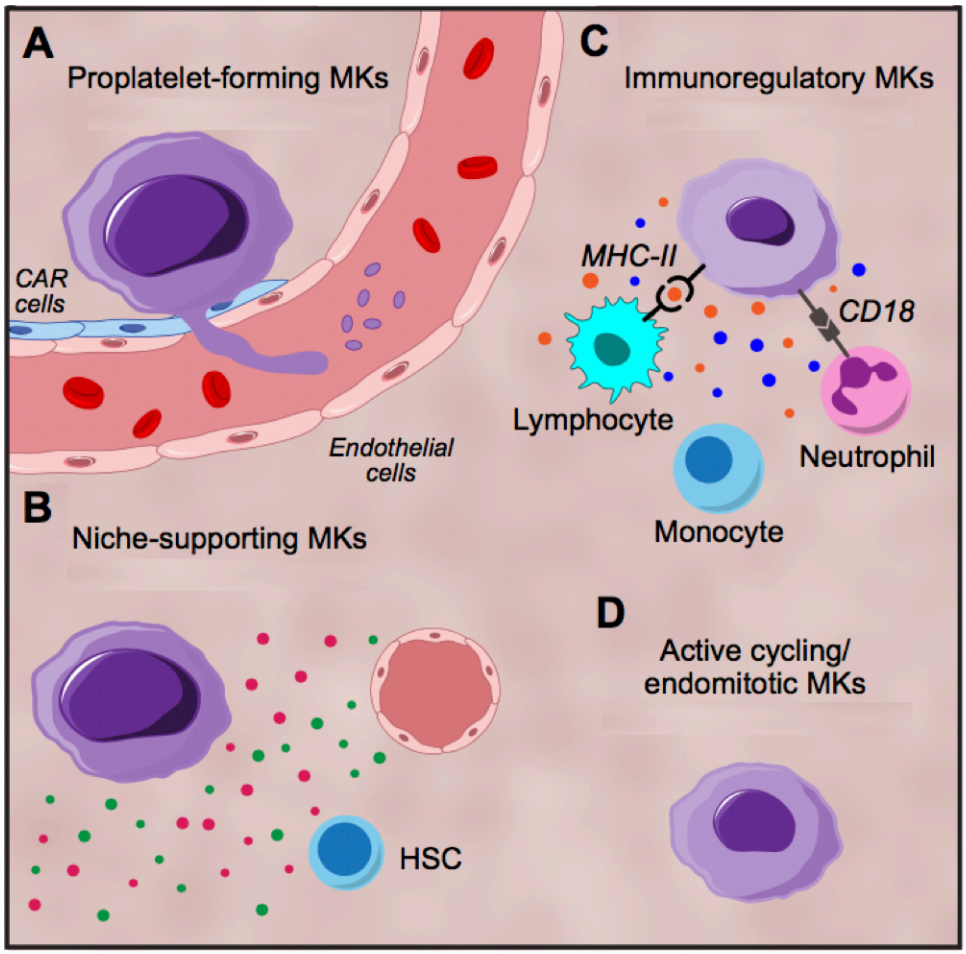


Figure 3 : Hétérogénéité des mégacaryocytes.
Les mégacaryocytes (MKs) peuvent être séparés en 4 groupes sur la base de leur fonction et de leur transcriptome. (**A**) les MKs plaquettogènes, situé à proximité des vaisseaux sanguins. (**B**) Les MKs qui participent à la niche des cellules souches hématopoïétiques (HSC) via le relargage de médiateurs chimiques. (**C**) Les MKs "immunitaires" qui participent à la présentation antigénique et peuvent interagir avec les neutrophiles. (**D**) Les MKs cyclant immatures qui sont suggérés comme réservoir des 3 précédentes populations. Adapté de (*43*)

c) Érythropoïèse et autres lignées myéloïdes

L'érythropoïèse est le processus qui permet la production des érythrocytes (Ery) ou globules rouges. Les progéniteurs érythroïdes (Ery-P) vont se différencier dans la moelle osseuse pour aboutir à des Ery qui sont des cellules anucléées dont le cytosol comprend des molécules d'hémoglobine permettant le transport de l'oxygène. Sur la base de critères morphologiques, on peut caractériser plusieurs états dans cette différenciation. On peut la suivre grâce au marqueurs CD36 et CD71 dans les stades précoces alors que le CD235a marque les Ery en fin de différenciation (*44*).

Il est intéressant de noter que les lignages érythroïde et mégacaryocytaire sont proches et partagent plusieurs régulateurs transcriptionnels comme GATA1 ou TAL1. Historiquement, Les Mk et les Ery ont été décrit comme dérivant d'un progéniteur érythro-mégacaryocytaire commun (MEP). Bien que ce type de progéniteur bipotent existe à l'état d'une population, définie sur la base de marqueurs de surface, son existence au niveau d'une cellule individuelle est questionnée. En effet, chez l'adulte, les études les plus récentes révèlent que la plupart des Mk sont produit directement depuis les HSC et que les MEP produisent en majorité des Ery-P même si certain gardent la capacité de donner également des Mk-P. La place du MEP dans la hiérarchie hématopoïétique est donc une question ouverte (*45*).

Il existe une confusion dans la littérature sur le terme « myéloïde » qui peut désigner, par opposition aux cellules lymphoïdes, l'ensemble des cellules engagées en différenciation qui ne sont pas des lymphocytes B, T ou NK. Cette définition inclut donc les érythrocytes et les mégacaryocytes. Strict au sensu, les cellules appartenant au lignage myéloïde sont uniquement les granulocytes (éosinophiles, neutrophiles et basophiles), les monocytes, les macrophages et les cellules dendritiques. Ces cellules myéloïdes jouent un rôle majeur dans les processus d'inflammation, l'immunité innée et l'immunité adaptative. Différents marqueurs peuvent être utilisés pour suivre la différenciation myéloïde par cytométrie en flux comme le CD15 pour les granulocytes, le CD14 pour les monocytes et les macrophages, le CD11c pour les cellules dendritiques ou encore le marqueur pan-myéloïde CD11b.

2) Développement du système hématopoïétique

Le système sanguin se met en place durant le développement embryonnaire à travers plusieurs vagues successives qui permettent la production de différentes cellules. Le consensus actuel sépare l'émergence hématopoïétique en 3 vagues successives qui se chevauchent partiellement : la vague primitive, la vague pro-définitive et la vague définitive. Les cellules hématopoïétiques appartenant à ces différentes vagues ont toutes une origine endothéliale et sont formées via un processus appelé transition endothélio-hématopoiétique. De par la difficulté à accéder au stade embryonnaire chez l'homme, la plupart des études sur le développement hématopoïétique ont été réalisées chez des organismes modèles comme le poisson zèbre, la caille, le poulet ou la souris.

a) Rôle de l'endothélium dans l'émergence des cellules hématopoïétiques

L'origine des cellules hématopoïétiques est une question de longue date. En 1932 Murray propose que les lignages hématopoïétique et endothélial dérivent d'une même masse cellulaire issue du mésoderme de la ligne primitive qu'il nomme hémangioblaste (*46*). Dans les années 1990, le laboratoire de Gordon Keller identifie, par l'étude d'un système de différenciation *in vitro* de cellules souches embryonnaires, un précurseur mésodermal clonal capable de donner à la fois des cellules endothéliales et des cellules hématopoïétiques (*47*). Cependant une étude plus récente utilisant un système de suivi à l'état unicellulaire des cellules de la ligne primitive *in vivo* chez la souris montre que la majorité des clones génèrent soit des cellules hématopoïétiques soit des cellules endothéliales mais très rarement les deux lignages (*48*). Ces résultats suggèrent que le destin hématopoïétique primitif ou endothélial est spécifié précocement avant la gastrulation. Une autre étude utilisant des cellules souches embryonnaires humaines confirme ces résultats (*49*). L'existence de l'hémangioblaste, bien que de moins en moins soutenue par des preuves expérimentales, est toujours sujette à débat. Une des possibilités est que l'hémangioblaste, à l'instar de l'oligopotence des HSC, soit un « état de compétence » possible pour les cellules mais qui n'est jamais atteint *in vivo* car les cellules sont poussées par le microenvironnement à une différenciation vers un seul lignage.

L'observation que les HSC apparaissent en étroite association avec l'endothélium vasculaire a conduit à une hypothèse alternative selon laquelle les cellules endothéliales peuvent présenter une capacité hémogénique et ainsi générer les cellules hématopoïétiques. En utilisant un système de lipoprotéines couplées à un colorant pour marquer les cellules endothéliales avant l'émergence des cellules hématopoïétiques, Jaffredo *et al* ont observé que le marquage est transmis aux cellules hématopoïétiques émergentes démontrant ainsi l'existence de l'endothélium hémogénique (*50*). Plus récemment, plusieurs études utilisant l'imagerie en temps-réel ont confirmé cette observation chez le poisson zèbre (*51*, *52*) et chez la souris (*53*).

Ainsi les cellules endothéliales qui possèdent la capacité de donner des cellules hématopoïétiques forment l'endothélium hémogénique (HE). Ces cellules sont capables de subir des changements morphologiques indépendamment des divisions cellulaires (*54*). Elles vont rompre les jonctions qui les unissent aux cellules endothéliales adjacentes, ce qui va induire leur circularisation et leur bourgeonnement dans la lumière du vaisseau (*51*), et commencer à exprimer un programme hématopoïétique. Ces cellules vont ainsi former des clusters appelés clusters hématopoïétiques intra-aortique (IAHC) qui vont ensuite permettre la libération des cellules hématopoïétiques dans la circulation. Ce processus est appelé transition endothélio-hématopoiétique (EHT) (Figure 5). Initialement décrit dans le contexte de l'aorte dorsale, il est maintenant établi que ce processus survient dans plusieurs autres localisations telles que le sac vitellin (*55*) ou le placenta (*56*), et la présence de IAHC a même été décrite au niveau de sites non-hématopoïétiques tel que la tête, le cœur ou les somites (*57*). Principalement étudié dans le cadre de l'émergence des HSC durant l'hématopoïèse définitive, ce processus de EHT est également à l'origine de la formation des progéniteurs hématopoïétiques des vagues précédentes.

La régulation de ce processus est complexe et encore incomplètement comprise mais il est clair que le facteur de transcription RUNX1 joue un rôle activateur essentiel (*58*). La voie de signalisation Notch est également importante car elle permet de spécifier l'identité artérielle plutôt que veineuse des cellules endothéliales (*59*, *60*). En effet, bien que les cellules myéloïdes puissent émerger de ces 2 types de HE, les cellules lymphoïdes ainsi que les HSC sont restreintes au HE artériel (*61*).

De plus, le HE est transitoire et restreint au développement embryonnaire (*62*) même si une émergence hématopoïétique limitée a été décrite chez la souris au niveau de la moelle du fœtus en fin de gestation et chez le jeune adulte (*63*).

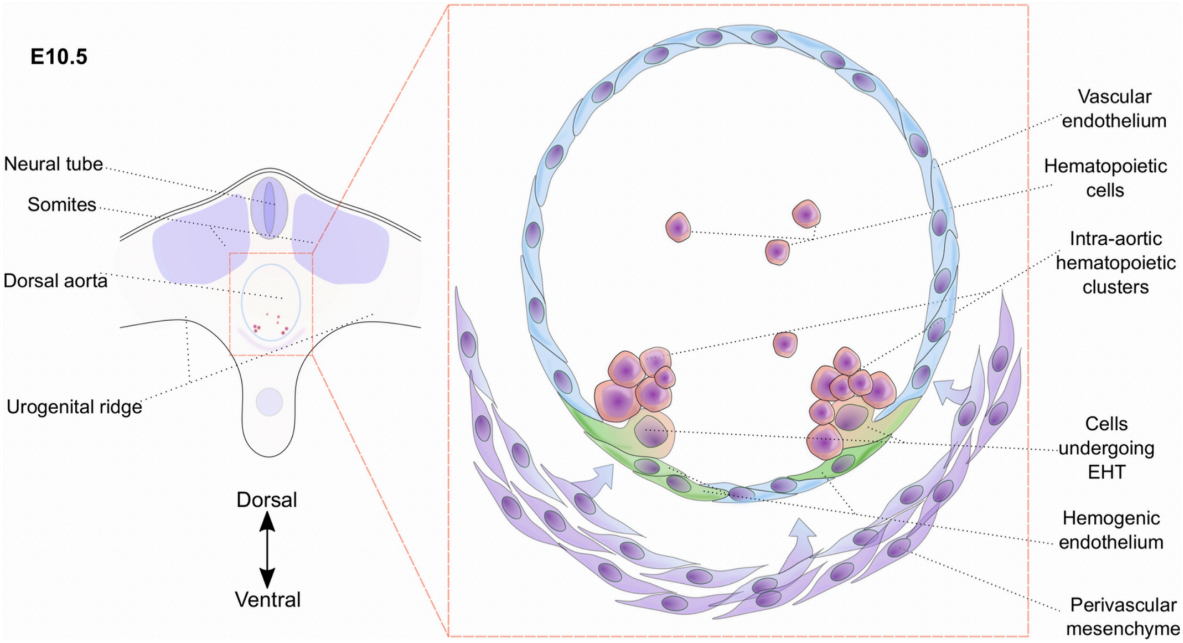


Figure 4 : Transition endothélio-hématopoiétique.
Représentation schématique de l'émergence des cellules hématopoïétiques au niveau de l'aorte dorsale chez la souris : Certaine cellules endothéliales possédant un potentiel hémogénique se transdifférencient pour former des clusters qui bourgeonnent dans la lumière de l'aorte. Ces clusters vont ensuite libérer des cellules hématopoïétiques. Issue de (*64*)

b) Hématopoïèse primitive

Les cellules hématopoïétiques apparaissent d'abord au cours d'une première vague, nommée hématopoïèse primitive, qui débute entre le jour embryonnaire (E) 7.0-7.5 chez la souris et les semaines 2-3 post-conception (WPC) chez l'humain (Figure 5). Elle permet l'émergence d'érythrocytes primitifs, de mégacaryocytes ainsi que de macrophages (*65*, *66*) et se déroule de façon extra-embryonnaire dans le sac vitellin (YS).

Les érythrocytes primitifs émergent de structures situées dans le YS appelées ilots sanguins et qui sont constituées de l'agrégation de cellules mésodermales et de précurseurs hématopoïétiques (*67*). Ces érythrocytes sont nucléés et de plus grande taille que les érythrocytes adultes. Ils sont également caractérisés par l'expression de globines embryonnaires (*68*). Ces cellules permettent l'oxygénation durant tout le développement

fœtal et sont encore détectable à la naissance (*69*). L'origine de ces érythrocytes est encore peu claire, l'hypothèse historique est qu'ils émergent directement du mésoderme après la gastrulation mais des études récentes questionnent cette hypothèse suggérant plutôt une formation via un équivalent de HE dans le YS mais de façon RUNX1 indépendant (*49*).

Les Mk apparaissent dans le YS à E7.5 chez la souris (*70*) et 4 WPC chez l'humain (*38*). Ils sont diploïdes, plus petits que les Mk adultes mais néanmoins capables de produire des plaquettes. Comme pour les érythrocytes, leur origine n'est pas clairement établie.

Les macrophages sont également générés dans le YS a E7.25 (*65*) sans intermédiaire monocyte (*71*). Les macrophages vont ensuite coloniser tous les tissus dont le cerveau (*72*). Ils seront plus tard remplacés par des macrophages dérivés de monocyte sauf pour les cellules microgliales du cerveau qui perdurent durant toute la vie (*73*). Leur absence complète dans des cultures de cellules ES Runx1$^{-/-}$, suggèrent que les macrophages sont générés via EHT (*74*).

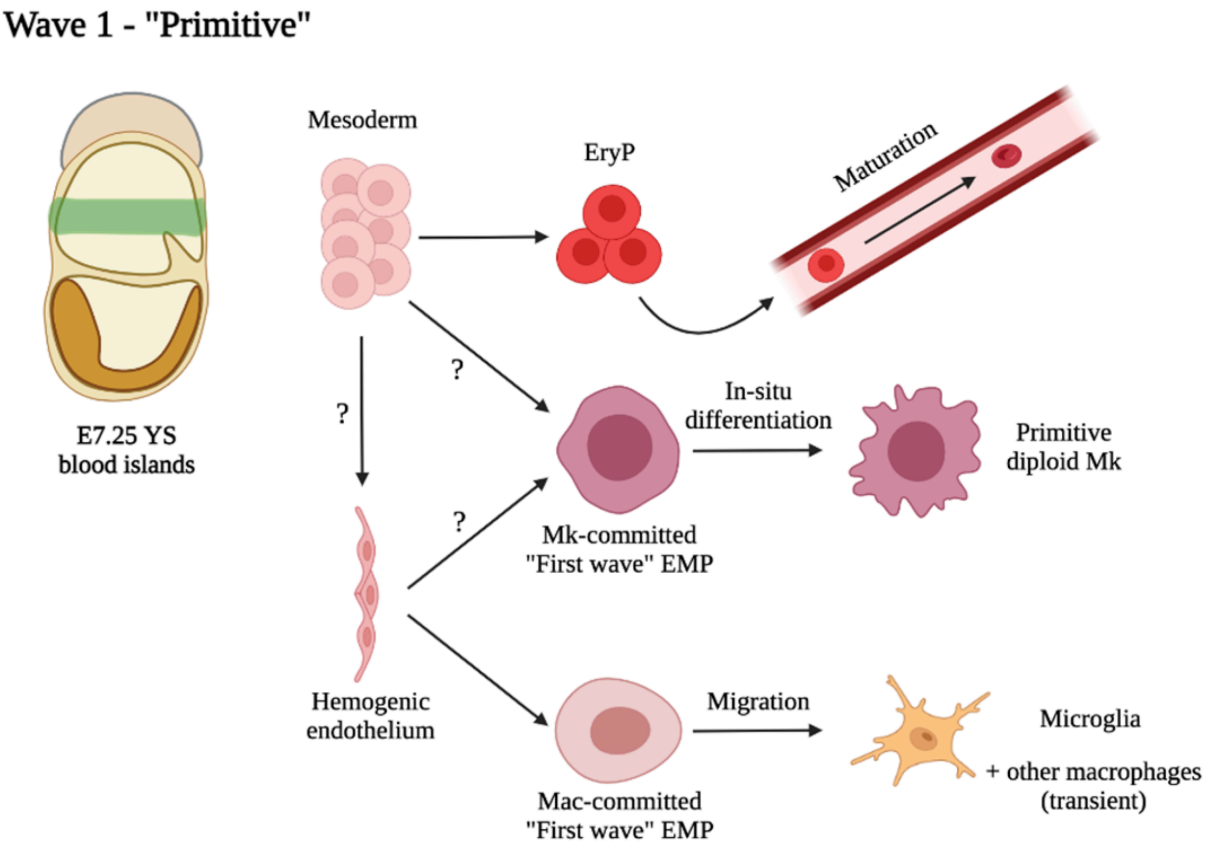


Figure 5 : Hématopoïèse primitive.
Représentation schématique de la vague primitive de l'hématopoïèse. Les points d'interrogation soulignent la connaissance incomplète de l'origine de certaines cellules. Issue de (*75*)

c) Hématopoïèse pro-définitive

La seconde vague, appelée pro-définitive ou pré-définitive/transitoire prend place au E8.25 chez la souris (*76*) et 4 WPC chez l'homme. Elle débute au niveau du HE du YS puis se poursuit dans l'embryon notamment au niveau du foie fœtal. Cette vague voit l'apparition de deux populations de progéniteurs multipotents : le progéniteur érythro-myéloïde (EMP) (*65*) et le progéniteur lympho-myéloïde (LMP) (*77*).

Les EMP correspondent à une population qui présente des potentiels érythroïde, mégacaryocytaire et myéloïde (granulocytes et macrophages) (*78*). Les EMP représentent une population hétérogène dans laquelle les cellules individuelles s'engagent vers différents lignages (*79*). Après leur génération dans le YS, ils vont coloniser le foie fœtal où ils se multiplient puis se différencient pour donner des cellules matures fonctionnelles (*80*, *81*). Les EMP sont notamment la source des macrophages résidents (via un intermédiaire monocytique) qui persistent jusqu'à l'âge adulte (*82*). Enfin, la génération de cellules Natural Killer (NK) à partir des EMP a également été rapporté (*83*). En cytométrie en flux, ces cellules sont caractérisées par la combinaison de marqueur suivante : $CD117^{+}CD31^{+}CD41^{+}CD16/32^{+}Sca^{-}$ (*78*).

Les LMP sont générés dans le YS et participent activement à la génération de progéniteurs myéloïdes dans le foie fœtal (*77*). Ils vont ensuite coloniser le thymus (*84*). Cette seconde vague permet également l'émergence de progéniteurs lymphocytaires T (*85*) et B (*86*) de façon HSC-indépendante à partir de différentes localisations (à la fois extra- et intra-embryonnaire). Le lien entre ces cellules et les LMP demande de plus amples clarifications car il n'est pour l'heure pas établi si les cellules B et T dérivent en totalité des LMP ou sont générées via EHT directement à partir du HE (*87*) (Figure 6).

Un point important est que les différents progéniteurs de cette vague, générés de façon HSC-indépendant, sont suffisants pour la maintenance de l'hématopoïèse et la survie de l'embryon jusqu'à la naissance (*81*).

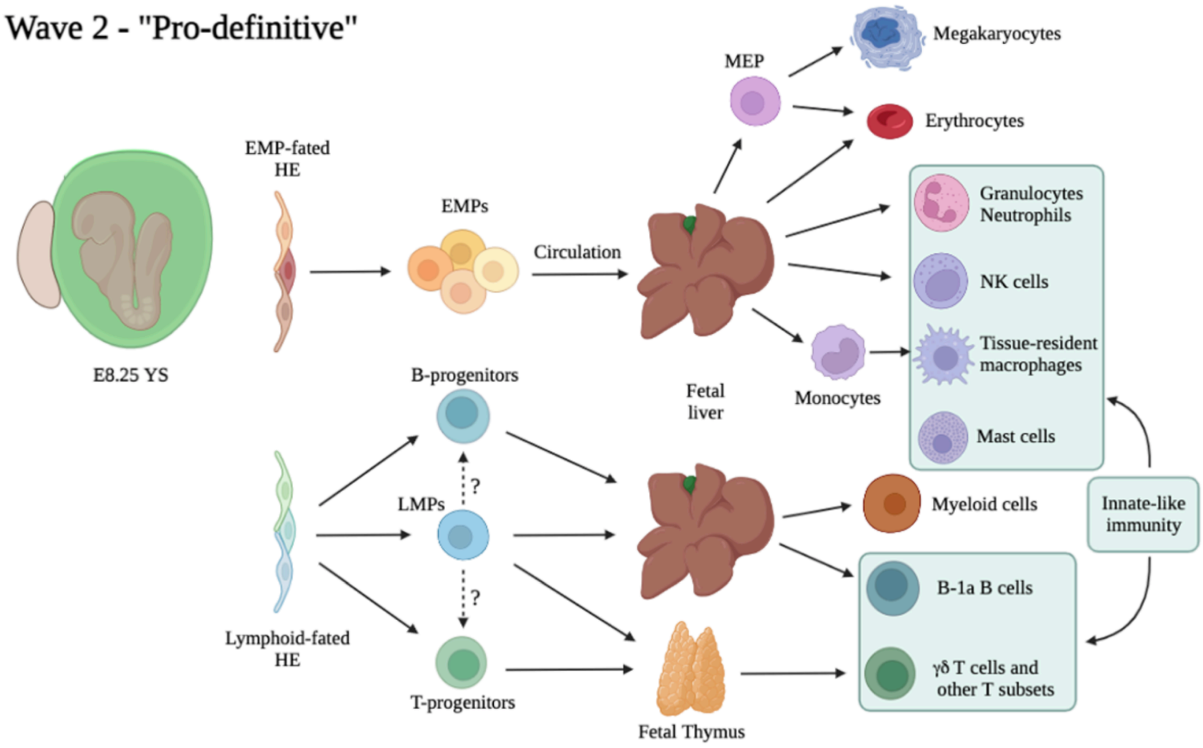


Figure 6 : Hématopoïèse pré-définitive.
Représentation schématique de la seconde vague pré-définitive de l'hématopoïèse. Les lignes en pointillé et les points d'interrogation indiquent les relations incertaines entre les progéniteurs. Issue de (*75*)

d) Hématopoïèse définitive

Enfin la 3ème vague, exclusivement intra-embryonnaire, dite définitive est caractérisée par l'émergence des premières HSC. Ces HSC apparaissent à E10.5 chez la souris (*88*) et au 32ème jour de développement chez l'homme (*89*). C'est dans la région aorte-gonade-mésonéphros (AGM) que sont générées les premières HSC (*90*) de façon autonome (*91*).

Le HE localisé dans la paroi des aortes dorsale, vitelline et ombilicale (*92*) va initier une EHT qui va aboutir au bourgeonnement des IAHC dans la lumière du vaisseau. Ces IAHC sont constitués de précurseurs des HSC (pre-HSC) (*93*) divisés en pro-HSC, pre-HSC de type I et pre-HSC de type II selon leur degré de maturation et leur capacité de reconstitution multilignage après greffe (*94*). Bien que d'un point de vue transcriptionnel les parois dorsale et ventrale de l'aorte dorsale se ressemblent beaucoup (*95*), la majorité des HSC sont produites à partir de la paroi ventrale (*96*). Cela est possiblement due au rôle instructif du plancher mésenchymateux sous aortique mais reste sujet à débat (*97*). Il convient également de noter que, alors que les IAHC comprennent environ 600 cellules Kit^+ dans l'aorte dorsale (*55*) seulement 40-60 pre-HSC au maximum sont détectables (*98*). Cela signifie que la majorité des

IAHC produisent des progéniteurs hématopoïétiques autres que des HSC. Cette hypothèse est appuyée par une étude de scRNA-seq qui montre que les IAHC produisent 2 « sous-vagues », une pour les HSC et une autre, majoritaire, pour des progéniteurs avec un potentiel lympho-myéloïde (*99*).

Les HSC vont ensuite migrer vers le foie fœtal (*100*) où elles vont fortement proliférer, permettant d'amplifier le pool de cellules souches, puis en fin de gestation vers la moelle osseuse où elles demeurent pour le restant de la vie (*101*) (Figure 7).

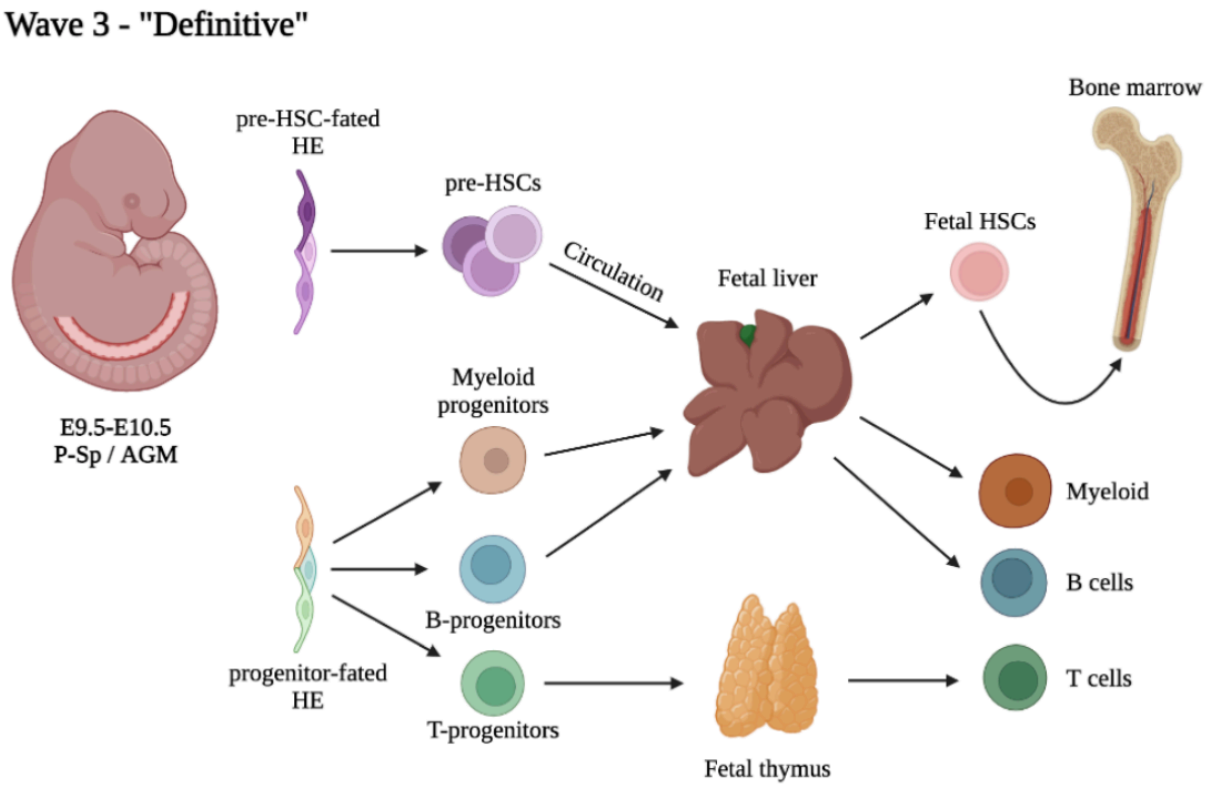


Figure 7 : Hématopoïèse définitive.
Représentation schématique de la 3ème vague définitive de l'hématopoïèse. Issue de (*75*)

Il est important de souligner que cette classification est probablement une simplification de la réalité et ne doit pas être vue comme une séparation stricte en 3 vagues car chacune comporte des progéniteurs appartenant aux mêmes lignages, qui peuvent émerger à des localisations similaires et pendant des timings qui se chevauchent.

3) Régulation de l'hématopoïèse

Au cours de l'hématopoïèse, les processus de prolifération, de quiescence ou de différenciation d'une cellule sont finement régulés. Ils dépendent de facteurs intrinsèques, propre à la cellule, de facteurs extrinsèques à la cellule et de l'interaction entre les deux.
Ainsi, des facteurs épigénétiques contrôlant la conformation de la chromatine, vont permettre l'accès à l'ADN à un type de protéine, les facteurs de transcription qui sont capables d'induire l'expression des gènes. L'activité de ces facteurs de transcription ainsi que celle des facteurs épigénétiques est souvent le dernier maillon de voies de signalisation comprenant plusieurs intermédiaires. Ces voies de signalisation sont activées en réponse à des stimuli extérieurs au niveau de récepteurs à la surface membranaire provenant du microenvironnement des cellules.

a) Microenvironnement

Le microenvironnement ou la « niche » d'une cellule correspond aux tissus et aux cellules à proximité. Cette niche régule une cellule de façon directe, via des contacts cellules-cellules et par les contraintes physiques imposées par la structure du tissu ou par la sécrétion de diverses molécules solubles. Pour les cellules hématopoïétiques, la niche principale est située dans la moelle osseuse. Des modèles de souris transgéniques permettant l'ablation de certains types cellulaires, la déplétion de facteurs solubles ou le suivi des cellules, couplés aux approches de microscopies contribuent à caractériser les différents constituants de la niche hématopoïétique.

Ainsi pour les HSC, la niche est constituée de cellules souches mésenchymateuses (MSC) qui ont des capacités de support, d'ostéoblastes qui ne régulent pas uniquement les HSC mais également des progéniteurs plus engagés et d'adipocytes qui sont des régulateurs négatifs du pool de HSC. Les nerfs sympathiques et les cellules de Schwann participent également à la régulation des HSC. Les cellules endothéliales, qui forment la paroi des vaisseaux ont aussi un rôle important via la sécrétion de plusieurs facteurs solubles tel que le ligand Notch ou les cytokines SCF et CXCL12. Les cellules périvasculaires, les péricytes sont divisées en 2 populations selon qu'elles expriment le récepteur à la leptine (LepR+) (la majorité des cellules LepR+ expriment en plus des niveaux élevés de CXCL12 et sont donc appelées cellules CAR

(CXCL12 abundant reticular)) ou le gène NG2 (NG2+). Enfin les cellules hématopoïétiques elles-mêmes font partie de la niche. Par exemple, une partie des HSC est en étroite association avec les Mk qui contrôlent leur quiescence via la sécrétion de CXCL-4 et TGF-B. Les macrophages participent à la régulation des ostéoblastes et des MSC et permettent la rétention des HSC dans la moelle via la molécule d'adhésion VCAM1. Certaines cellules T promeuvent la survie des HSC grâce à la sécrétion de cytokines, notamment l'interleukine 10 (*101*, *102*) (Figure 8).

Il convient de noter que cette niche n'est pas uniforme, on note des différences selon la localisation dans l'os qui impacte la régulation des HSC. La moelle est généralement divisée en 2 niches : la niche artériolaire qui comporte les péricytes NG2+ et les cellules nerveuses où les HSC ont un biais de différenciation lymphoïde. La niche sinusoïdale comportant les péricytes LepR+ et les Mk où les HSC expriment le récepteur von Willebrand (vwf+), sont plus quiescentes, et ont un biais de différenciation Myéloïde-Mk (*103*).

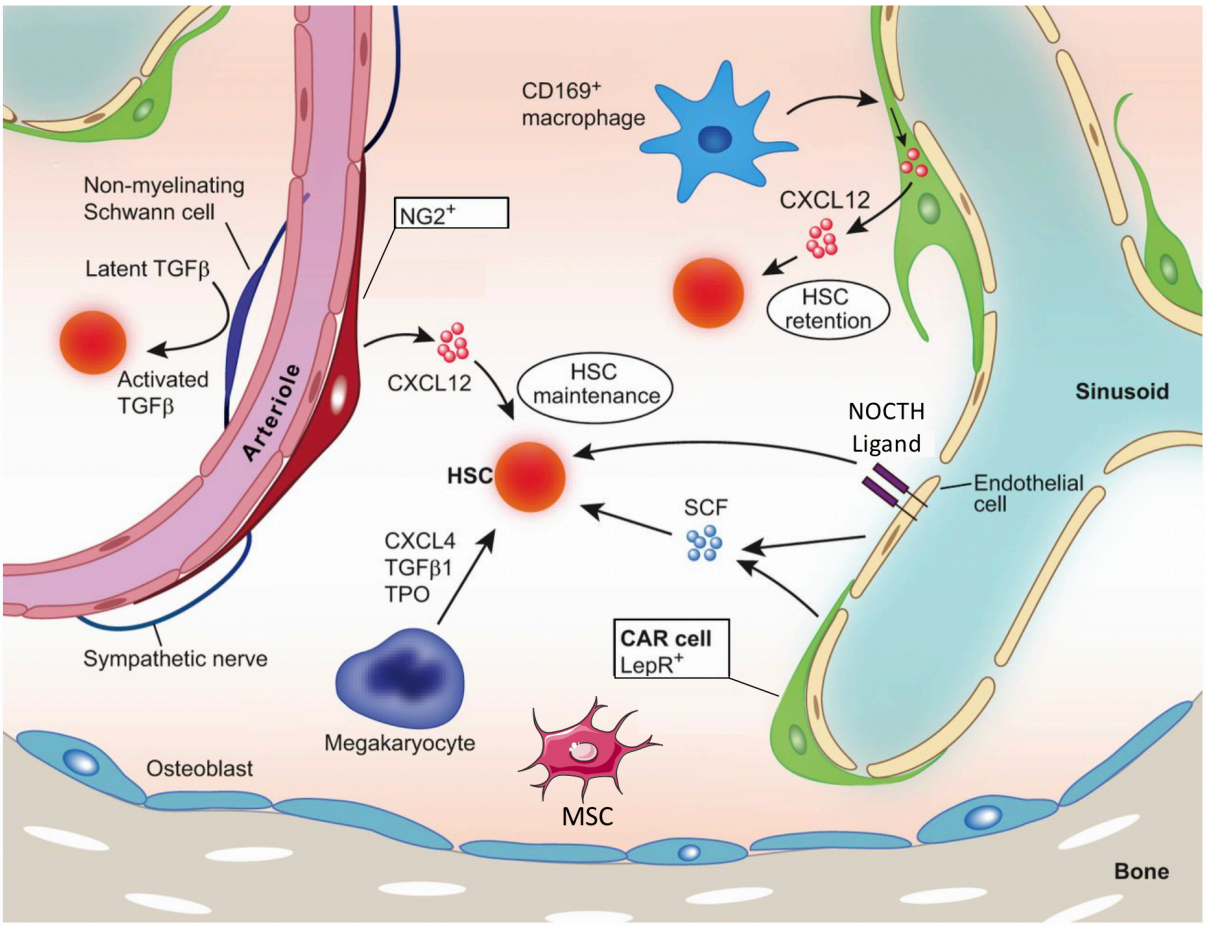


Figure 8 : Niche hématopoïétique des HSC.
Représentation schématique de la niche des HSC, adapté de (*102*)

Concernant les Mk, la régulation par le microenvironnement est également importante. Dans la moelle, les Mk sont en contact avec les neutrophiles et peuvent, via un processus appelé empéripolèse, absorber temporairement ces derniers afin de réaliser un échange de matériel membranaire qui accélère la maturation. Les Mk vont également récupérer certains acides gras libérés par les adipocytes afin de faciliter leur maturation. De plus, la sécrétion de la cytokine CXCL4 par les Mk permet une régulation paracrine qui inhibe la prolifération et induit une maturation (*104*).

Comme pour les cellules hématopoïétiques, le développement des techniques de scRNA-seq ces dernières années révèle une hétérogénéité importante au sein de chaque type cellulaire constituant la niche (*105*).

b) Voies de signalisation

Les contacts avec les cellules du microenvironnement ainsi que les différentes molécules solubles qu'elles sécrètent vont permettre la régulation des cellules hématopoïétiques grâce à l'activation de voies de signalisation.

A leur surface, les cellules possèdent différentes protéines membranaires, appelés récepteurs. Lorsque ces récepteurs sont activés, par une interaction avec leur ligand, cela entraine une cascade de signalisation intracellulaire impliquant plusieurs intermédiaires et qui aboutit généralement à la modulation de l'expression génique. Ces ligands peuvent être eux même des protéines membranaires, dans ce cas l'activation se fait par un contact cellule-cellule, ou bien des facteurs solubles, comme des cytokines. Ces facteurs solubles peuvent être sécrétés par la cellule elle-même (action autocrine), par les cellules voisines (action paracrine) ou bien par des cellules à distance (action endocrine).

i) Interaction cellule-cellule : exemple de la voie NOTCH

Chez les mammifères, il existe 4 récepteurs transmembranaires pour la voie Notch : NOCTH1, NOCTH2, NOTCH3 et NOTCH4. Ces récepteurs sont activés par une interaction avec un des 5 ligands, Delta-like (DLL)1, DLL3, DLL4, Jagged1 ou Jagged2. Ces ligands sont également des protéines transmembranaires, la voie Notch est donc activée dans une cellule par un contact physique direct avec une autre cellule. Lorsque le ligand se lie au récepteur, ce dernier subit

plusieurs étapes de clivage protéolytique qui aboutissent à la libération du domaine intracellulaire de Notch (NCID) qui va migrer dans le noyau où il va recruter des coactivateurs comme MAML puis se lier au régulateur transcriptionnel RBPJ qui va induire l'expression de ces gènes cibles (Figure 9). En absence de NCID, RBPJ est complexé avec des coinhibiteurs qui l'empêche d'activer la transcription (*106*). Les cibles principales de la voie Notch sont les gènes des familles Hey et Hes, importants dans le développement embryonnaire, mais également certains gènes impliqués dans le cycle et la prolifération cellulaire comme *CCND1* ou *MYC* (*107*).

La voie Notch joue un rôle dans l'émergence des HSC durant le développement embryonnaire. D'abord, Notch va promouvoir l'identité artérielle plutôt que veineuse des cellules endothéliales nécessaire pour l'émergence du HE (*59*). Ensuite, la balance entre l'activation forte de la voie par DLL4-NOTCH1 ou plus modérée par Jagged1-NOTCH1 détermine l'identité respectivement endothéliale ou hémogénique, avec l'induction de gènes cibles comme *GATA2* ou *FOXC2*, des cellules du HE (*108*). Pour les HSC, la voie Notch participe au maintien de l'état indifférencié ainsi qu'à l'auto-renouvellement (*109*). Notch est également impliqué dans la différenciation des lymphocytes T (*110*) et dans la différenciation érythro-mégacaryocytaire (*111*, *112*). L'effet de maintien de l'état souche ou pro-différenciant de Notch pourrait dépendre du type de ligand qui active la voie (*113*).

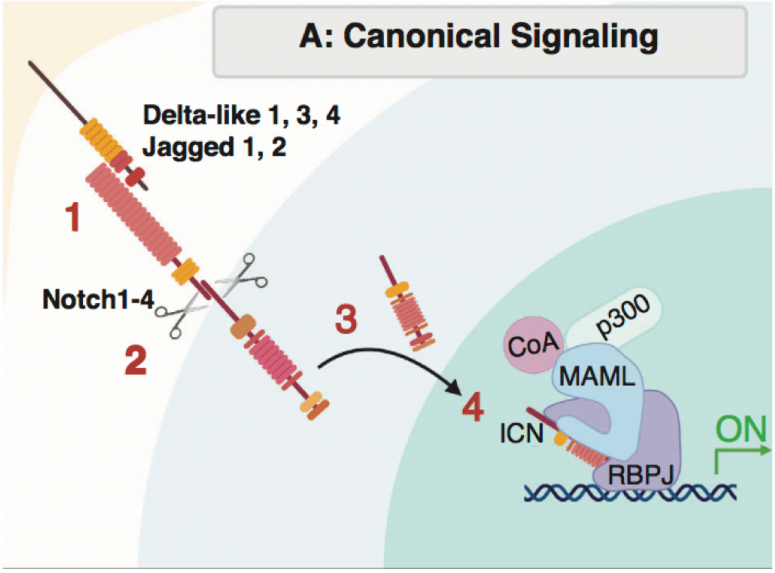


Figure 9: Voie de signalisation NOTCH.

L'interaction ligand-récepteur **(1)** entraine des clivages protéolytiques **(2)** qui aboutissent au relargage du domaine intracellulaire de NOCTH (NCI) qui migre ensuite dans le noyau **(3)** où il va interagir avec le facteur de transcription RBPJ et recruter des coactivateurs pour induire l'expression des gènes cibles **(4)**. Adapté de (*110*)

ii) Cytokines : exemple de la thrombopoïétine

De très nombreuses cytokines participent à la régulation des cellules hématopoïétiques avec des rôles dans les processus de prolifération, de survie et de différenciation cellulaire. Par exemple, plusieurs cytokines jouent un rôle dans la mégacaryopoïèse comme l'interleukine (IL)-3, l'IL-6, l'IL-9, l'IL-11, le LIF, le GM-CSF, le SCF ou encore le FLT3-L (*114*).

Cependant, la cytokine majeure de la différenciation mégacaryocytaire est la thrombopoïétine (TPO) (*115*). La TPO est produite en continue par le foie (*116*). Le récepteur à la TPO, MPL, qui n'est exprimé que sur les HSC, les Mk et les plaquettes, est une protéine transmembranaire qui ne possède pas d'activité propre mais est associée par sa partie cytoplasmique avec des Janus tyrosine kinases (JAK). A l'état basal MPL est sous forme de monomère (ou plus rarement de dimère inactif). Lorsque la TPO se lie sur sa partie extracellulaire, MPL se dimérise et change de conformation ce qui induit une transphosphorylation des résidus tyrosines des protéines JAK associées. Une fois phosphorylées, les protéines JAK vont induire le recrutement et l'activation de plusieurs cascades biochimiques en aval tel que MAPK/ERK, PI3K/AKT ou encore STAT3 et STAT5 (*117*) (Figure 10). Ces différentes voies de signalisation ont de multiples rôles dans les processus de survie et de prolifération. Cette signalisation est soumise à un rétrocontrôle négatif car la stimulation à la TPO induit l'activation de plusieurs protéines qui régulent négativement l'activité de la TPO comme la kinase LYN (*118*). De plus, la fixation de la TPO sur MPL entraine une internalisation et une dégradation par le protéasome. Le taux de TPO circulant est ainsi régulé par les Mk et les plaquettes qui vont l'internaliser conduisant à une corrélation inverse entre taux de TPO et masse de Mk/plaquettes permettant de contrôler le niveau de TPO selon le besoin (*119*).
L'activation de la voie TPO est impliquée dans la prolifération et l'auto-renouvellement des HSC (*120*). De plus, la TPO non seulement induit la différenciation des HSC mais aussi les biaise vers le lignage mégacaryocytaire en augmentant la signalisation mitochondriale (*121*).

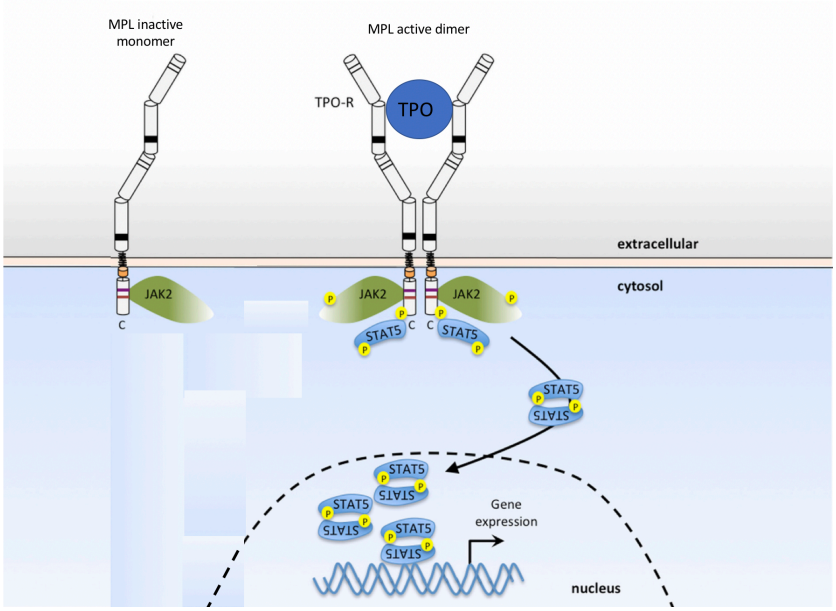


Figure 10 : Activation de la voie de signalisation STAT par la TPO.

En absence de TPO le récepteur est sous forme de monomère inactif. La fixation de la TPO permet la dimérisation et l'autophosphorylation du récepteur. Cela entraine l'activation des protéines STAT5 qui sont relocalisées dans le noyau pour activer les gènes cibles. Adapté de (*122*)

c) Transcription

La transcription est un processus central dans la régulation de la biologie des cellules impliquant plusieurs niveaux de contrôle de l'expression génique.

Chez les eucaryotes, c'est l'ARN polymérase II (Pol II), dont le recrutement et l'activité sont contrôlés par différents facteurs, qui est responsable de la transcription des gènes. Pour initier cette transcription, la Pol II a besoin de 6 cofacteurs, les éléments généraux de transcription, TFIIA,B,D,E,F&H. Cette initiation débute au niveau d'un promoteur. Le promoteur est une courte séquence d'ADN situé immédiatement en amont du gène et qui comporte plusieurs éléments essentiels pour la transcription. Un de ces éléments, présent dans le promoteur des gènes finement régulés, est une séquence appelée « boite TATA ». Cette séquence permet la fixation de TFIID, et de la Pol II avec qui il est complexé, via sa sous unité TBP à l'ADN. Les facteurs TFIIA et TFIIB vont ensuite venir stabiliser cette fixation, puis les facteurs TFIIE et TFIIH s'ajoutent pour former le complexe d'initiation qui va pouvoir « ouvrir » la double hélice d'ADN et débuter la transcription (*123*).

Les facteurs de transcription (TF) sont des protéines qui peuvent se fixer à l'ADN au niveau de séquences nucléotidiques spécifiques mesurant en moyenne de 4 à 6 nucléotides. Ces « motifs » sont différents pour chaque TF mais peuvent être proche voir quasi-identiques au sein d'une même famille protéique. Ces motifs sont présents partout dans le génome mais en

particulier au niveau de modules régulateurs de la transcription. Les promoteurs contiennent ainsi des sites de fixation pour plusieurs TF différents, souvent groupés en « clusters ». Lorsqu'un TF se lie sur un promoteur, il va entrainer le recrutement de la Pol II ce qui induit l'expression du gène. Ce recrutement fait intervenir plusieurs partenaires, un des principaux étant un complexe protéique, le complexe Médiateur, qui fait le lien entre les TF et la machinerie de transcription. A l'inverse, Les TF peuvent également avoir des effets de répression sur l'activité transcriptionnelle d'un gène en recrutant des partenaires protéiques différents qui vont bloquer la polymérase. Un promoteur peut contenir des sites de fixation pour de nombreux TF différents qui peuvent avoir des actions synergiques ou au contraire antagonistes. De plus, un même TF peut avoir un rôle activateur de la transcription pour certains gènes et répresseur pour d'autres. Selon le contexte cellulaire, un TF peut se fixer sur des sites différents et donc réguler différents gènes. Enfin, il est important de noter que les TF agissent souvent en interaction avec des cofacteurs et/ou au sein de complexes multiprotéiques ce qui peut impacter la liaison d'un TF à l'ADN en renforçant sa spécificité ou en permettant la liaison sur des sites sur lesquels il ne peut se lier seul (*124*).

En plus des promoteurs, les enhancers constituent un autre type de régions régulatrices. Ces séquences, beaucoup plus étendues que les promoteurs, peuvent être situées à distance du gène qu'elles régulent, jusqu'à plusieurs centaines de kilobases (Cela est possible en raison de la conformation 3D de l'ADN qui fait que 2 séquences éloignées d'un point de vue « linéaire » peuvent en réalité être en contact du fait de repliements de la molécule d'ADN) (Figure 11). Les enhancers comportent de nombreux sites de fixation pour des TF qui modulent la transcription. Un enhancer peut contrôler un ou plusieurs gènes. A l'inverse, un gène peut être contrôlé par un ou plusieurs enhancers.

Les TF peuvent induire la transcription de gènes codant d'autre TF ou bien contrôler leur propre activation. Cela entraine la formation de réseaux génétiques dont la complexité limite leurs études expérimentalement. Pour pallier à cela, diverses méthodes bio-informatiques se sont développées dans les dernières années afin de mieux caractériser ces réseaux (*125*).

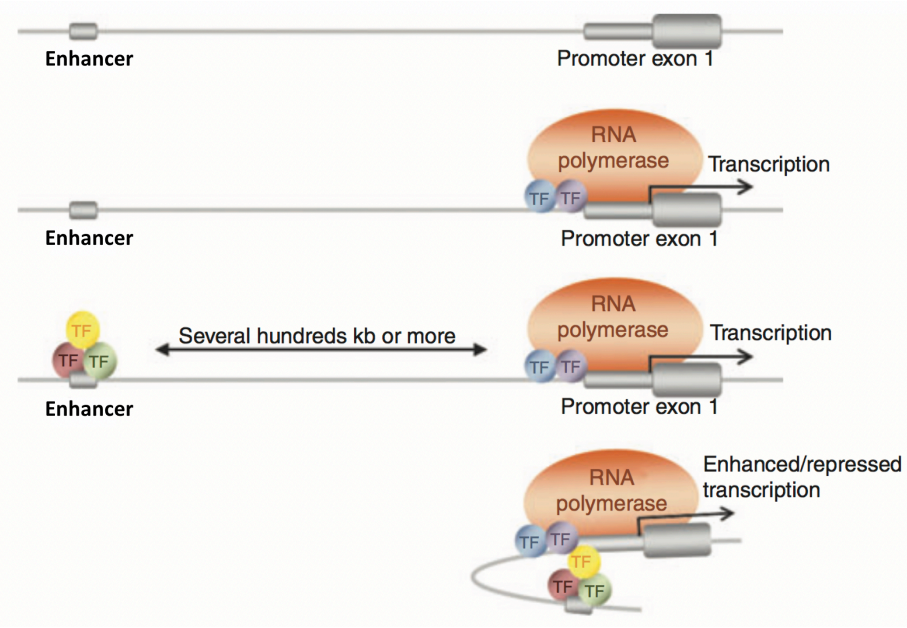


Figure 11 : Régulation de la transcription.
Représentation schématique de la régulation de la transcription d'un gène par des facteurs de transcription. Adapté de (*126*)

L'impact des TF sur les processus de différentiation est étudié depuis de nombreuses années dans le système hématopoïétique et leur importance fonctionnelle au cours du développement et de la différenciation a été démontrée à l'aide de modèles de surexpression ou d'inactivation (*127*).

Dans les HSC, en plus des TF essentiels pour l'autorenouvellement, la quiescence et la survie, des TF clés de différents lignages semblent être co-exprimés à un faible niveau. Ceci permettrait de « primer » les HSC vers la différenciation. Lorsque les cellules s'engagent dans cette dernière, le niveau des TF du lignage associé augmente de façon progressive au cours de la différenciation alors que celui des TF d'autres lignages diminue (*128*, *129*). Le changement dans les abondances relatives des TF est un moteur majeur des processus de différenciation car ces derniers agissent de façon dose dépendante. Par exemple, GATA1 à faible niveau induit la différenciation éosinophile alors qu'à des niveaux plus élevés il induit la différenciation érythro-mégacaryocytaire ou encore GATA2 dont l'haplo-insuffisance induit des défauts dans l'hématopoïèse (*130*). Bien que ces changements soient graduels, l'effet peut être « switch » car il existe des seuils dans la quantité des TF qui déterminent leur action (*129*). De plus, la surexpression d'un facteur de transcription peut reprogrammer une cellule engagée dans un lignage vers un autre lignage, soulignant le rôle clé des TF (*128*).

Plusieurs études montrent que des facteurs importants pour les HSC comme TAL1 ou LYL1 possèdent au niveau de leur promoteur des motifs de liaison pour les TF des familles GATA et ETS. Durant l'hématopoïèse fœtale, ces motifs sont liés respectivement par GATA2 d'une part et FLI1 et ERG d'autre part (*131*). La présence de ces mêmes motifs ainsi que de motifs E-box (permettant la liaison du TF TAL1) est également retrouvé dans les enhancers qui contrôlent les gènes de ces familles révélant une régulation combinatoire complexe des HSC par les TF (*132*). Afin de cartographier cette combinatoire, le groupe de Berthold Göttgens a réalisé des expériences d'immunoprécipitation de la chromatine (ChIP-seq) sur 10 TF dont les motifs sont retrouvés au niveau de gènes importants pour les cellules souches et les progéniteurs. L'analyse des données obtenues montre que certains TF co-localisent de façon significative et des expériences de co-immunoprécipitations ont validé une interaction physique entre ces facteurs. Ainsi cette étude démontre l'existence d'un complexe protéique, le complexe heptade, constitué de 7 TF : TAL1, LYL1, LMO2, GATA2, RUNX1, ERG et FLI1 (*133*). Ce complexe permet le maintien des HSC en inhibant l'expression des gènes de la différenciation. Il contrôle également l'engagement dans la différenciation mégacaryocytaire en se fixant et en induisant un faible « priming » transcriptionnel des gènes impliqués dans la différenciation. Lorsque la différenciation débute, le complexe heptade reste fixé sur ces gènes cibles mais un échange se produit entre le TF GATA2 qui est remplacé par le TF GATA1 qui va induire l'expression des gènes mégacaryocytaires (*134*) (Figure 12).

Des expériences de ChIP-seq sur des cellules primaires humaines par Tijssen *et al* pour les TF GATA2, RUNX1, FLI1 et TAL1 révèlent que ces facteurs, importants pour les HSC, sont toujours actifs dans les Mk confirmant ainsi la proximité entre ces deux lignages (*135*). De plus, l'utilisation de modèles murins permettant une invalidation conditionnelle a permis l'identification de plusieurs TF importants pour la mégacaryopoïèse comme ERG pour l'engagement mégacaryocytaire (*136*) ou MEIS1, EVI1 et NFE2 pour les étapes terminales de la maturation (*126*).

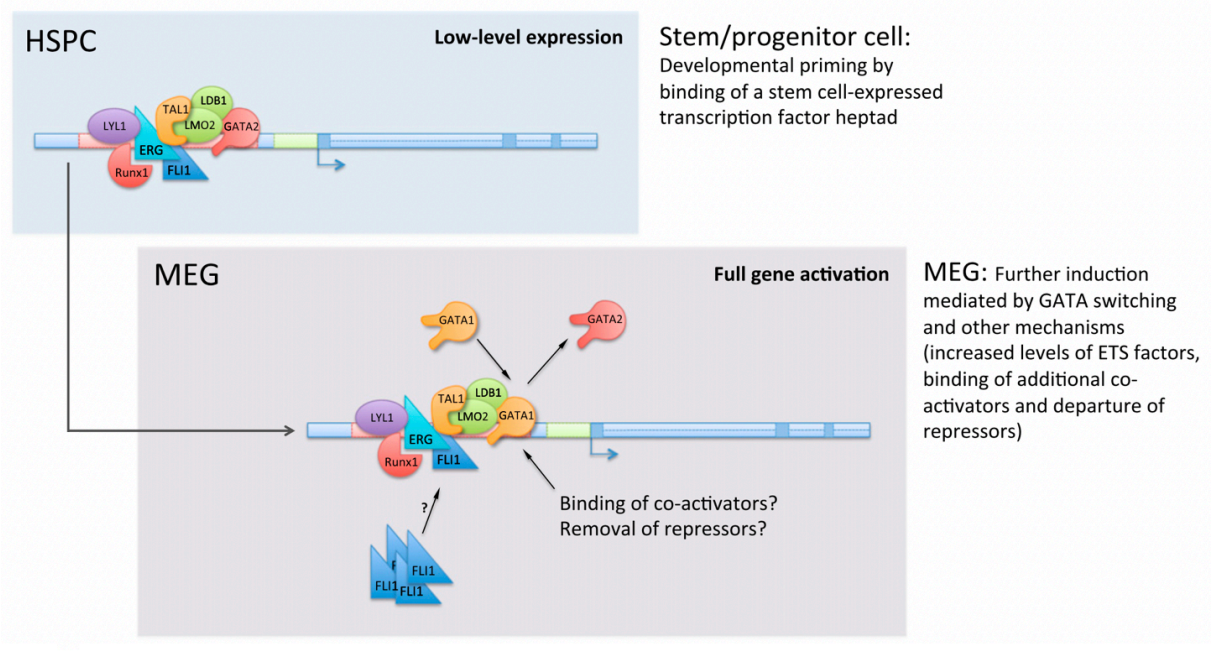


Figure 12 : Complexe heptade.
Dans les cellules souches, les facteurs de transcription du complexe heptade se fixent et "prime" les gènes associés aux lignages. La différenciation des cellules entraine une activation transcriptionnelle des gènes "primés". Cette activation est médiée par le switch GATA1-GATA2 ainsi que par le recrutement de cofacteurs. HSPC = cellules souches et progéniteurs hématopoïétiques ; MEG = mégacaryocytes. D'après (*134*)

Les TF peuvent également s'antagoniser l'un l'autre comme par exemple PU.1 qui induit la différenciation myéloïde et bloque la différenciation érythro-mégacaryocytaire et GATA1 qui à l'effet inverse. De par leur proximité, les lignages érythroïde et mégacaryocytaire partagent beaucoup de régulateurs. Un des points de bascule majeur entre l'engagement vers l'un ou l'autre de ces lignages est l'équilibre entre la quantité des facteurs FLI1 et KLF1. Confirmé au niveau protéique, cette balance se traduit par une augmentation du niveau de FLI1 et une diminution de celui de KLF1 dans la mégacaryopoïèse et inversement dans l'érythropoïèse (*128*). Cette balance est contrôlée par un des facteurs de transcription clé de la mégacaryopoïèse : GATA1.

Chez les mammifères, il existe 6 protéines de la famille GATA dont trois, GATA1, GATA2 et GATA3 sont exprimés dans les cellules hématopoïétiques. GATA1 est exprimé dans plusieurs types de cellules hématopoïétiques comme les éosinophiles, les basophiles, les cellules dendritiques ou encore mastocytes (*137*). GATA1 possède un domaine de liaison à l'ADN comprenant 2 doigts de zinc capables de se lier sur le motif consensus (A/T)GATA(A/G) et dans sa partie N-terminale un domaine transactivateur permettant d'induire l'expression de ces gènes cibles. GATA1 interagit avec de très nombreux partenaires, l'un des plus important

étant le cofacteur FOG1 (*138*). FOG1 et GATA1 interagissent via leurs domaines en doigt de zinc respectifs. Cette interaction permet le recrutement de plusieurs protéines et complexes qui vont réguler l'activité de GATA1 comme des répresseurs transcriptionnels ou des complexes de remodelage de la chromatine (*137*). Enfin, l'activité de GATA1 est régulée par plusieurs modifications post-traductionnelles comme des phosphorylations, des sumoylations (*139*) ou des acétylations. Par exemple, l'acétylation de GATA1 par CREBBP/EP300 participe à sa fixation sur ces cibles (*140*).

GATA1 joue un rôle essentiel dans la différenciation érythro-mégacaryocytaire comme l'ont montré plusieurs études d'inactivation chez la souris. Pendant la différenciation, GATA2 qui est exprimé dans les HSC va être remplacé dans environ 30% de ses sites de liaisons par GATA1 au cours d'un processus appelé « GATA-switch ». GATA1 est une cible de GATA2 qui va induire son expression et à l'inverse être réprimé par celui-ci induisant ainsi une boucle de contrôle favorisant ce remplacement (*135*, *141*). Cependant, GATA2 reste présent dans les Mk contrairement aux érythrocytes où son expression est totalement réprimée (*141*).

GATA1 joue un rôle différent dans la différenciation érythroïde où sont absence conduit à la diminution de cellule par l'entré en apoptose (*142*), et dans les mégacaryocytes où elle entraine un blocage de la différenciation terminale qui aboutit à une expansion de progéniteurs immatures (*143*). Cet effet variable de GATA1 pourrait dépendre de son niveau d'expression car une diminution du niveau de GATA1 n'impacte que modérément les érythrocytes mais bloque la différenciation mégacaryocytaire (*144*, *145*). L'étude détaillée des Mk sans GATA1 montre qu'ils présentent de nombreuses anomalies comme une faible ploïdie, pas de formation du système de démarcation membranaire, peu de plaquettes et une faible activation des gènes associés à la mégacaryopoïèse démontrant ainsi le rôle central de GATA1 dans ce lignage (*146*).

d) Régulation épigénétique

Les facteurs de transcriptions sont les acteurs majeurs de la régulation de l'identité des cellules. Cependant, pour exercer leur fonction, ils doivent se lier sur l'ADN. La liaison d'un facteur sur un motif va dépendre de l'accessibilité de ce dernier.

Dans le noyau, l'ADN est compacté, enroulé autour d'un octamère de protéines formés des histones H2A, H2B, H3 et H4. Cette structure est appelée nucléosome et forme une unité de

147 paires de base dont l'enchainement, et la structuration grâce à l'histone H1, permet la formation de nucléofilaments qui constituent la chromatine. Cette dernière est traditionnellement divisée en 2 états, l'hétérochromatine qui présente une forte condensation des nucléosomes et l'euchromatine plus lâche (Figure 13).

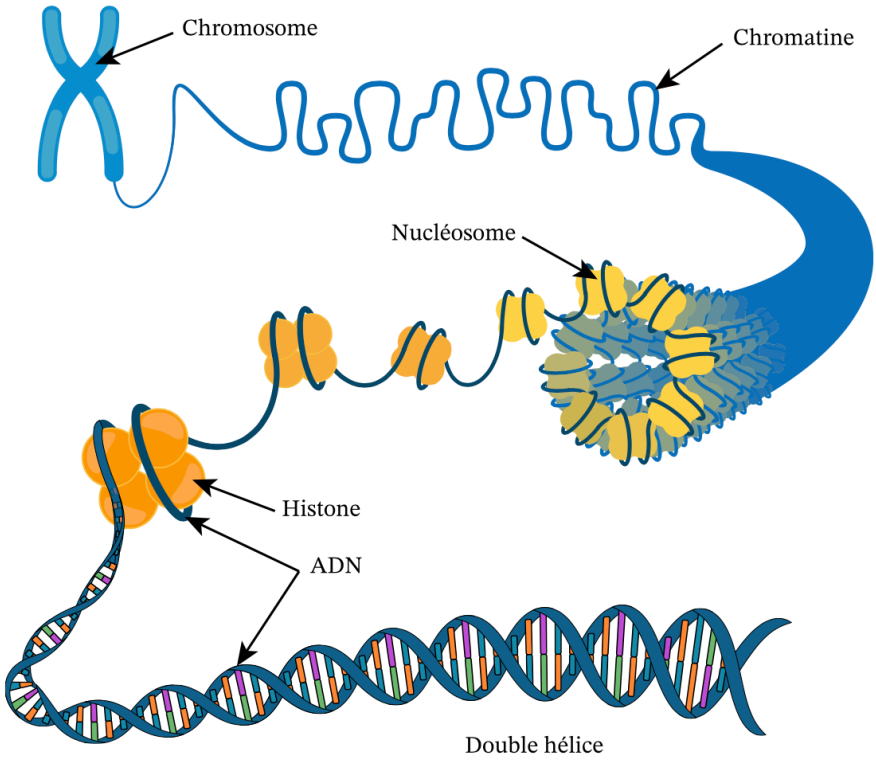


Figure 13 : Organisation de la chromatine.
Représentation des différents niveaux de compaction de l'ADN. Illustration issue de nagwa.com

La compaction de la chromatine dépend de l'enchainement des nucléosomes. Des complexes de remodelage de la chromatine comme SWI/SNF sont capables de déplacer les nucléosomes le long de l'ADN ou même de les éjecter créant ainsi des zones pauvres en nucléosome ou l'ADN est libre pour la fixation de TF, notamment au niveau des promoteurs ce qui facilite l'accès aux TF pour induire la transcription (*147*). Chez la souris, la délétion de *Arid1a*, le composant principal du complexe SWI/SNF affecte à la fois la différenciation myéloïde et lymphoïde (*148*). Celle de *Baf200*, un autre membre du complexe, induit une anémie due à un blocage de l'érythropoïèse et diminue l'expansion des HSC durant le développement dans le foie fœtal (*149*).

Des modifications épigénétiques peuvent également impacter la structure de la chromatine et ainsi l'accessibilité des TF et donc l'expression génique. Ces modifications sont réversibles et peuvent être transmises aux cellules filles lors de la division cellulaire.

i) Modifications des histones

Les histones peuvent être la cible de plusieurs modifications post-traductionnelles comme l'acétylation, la méthylation, la sumoylation ou encore l'ubiquitination (Figure 14).

L'acétylation des histones résulte du transfert d'un groupement acétyle de l'acétyl-CoA sur un résidu lysine de l'histone. Cette réaction entraine une dispersion de la chromatine qui devient plus accessible aux TF. Ainsi, l'acétylation de la lysine 27 de l'histone H3 (H3K27ac) est associée aux enhancers actifs et participe à l'ouverture de la chromatine par le recrutement de protéines à bromodomaine, comme BRD4, qui reconnaissent cette marque. Cette réaction est catalysée par les histones acétyltransférases (HAT) (*150*). Plusieurs de ces facteurs, comme CREBBP/P300, jouent un rôle important dans la différenciation et l'auto-renouvellement des HSC, comme le montre les défauts de reconstitution hématopoïétique des souris K.O. (*151*). A l'inverse, les histones désacétylases (HDAC) induisent une compaction de l'ADN et une répression transcriptionnelle.

Les histones peuvent être mono- di- ou tri-méthylés par des histones méthyltransférases (HMT). Selon le résidu lysine ciblé, cette méthylation peut être associé à l'activation transcriptionnelle comme pour les résidus H3K36me3 ou H3K79me3 ou à la répression comme pour H3K9me3 ou H4K20me3 (*152*).
La méthylation de H3K4me3 est un marqueur des promoteurs actifs. Dans les HSC, elle est médiée par MLL ce qui participe à la prolifération des cellules (*153*). A l'inverse, le complexe PRC2 via ses sous unité EZH2, EED et SUZ12 augmente la marque répressive H3K27me3 et maintient ainsi l'état indifférencié des HSC en inhibant leur entrée en différenciation (*154*).
Il convient de noter que certains promoteurs peuvent co-exprimer les marques H3K4me1/2 et H3K27me3 ce qui les maintient dans un état « poised » permettant une activation rapide si nécessaire (*152*). Toutes ces marques sont réversibles par l'action de protéines histones déméthylases tel que KDM5A ou LSD1.

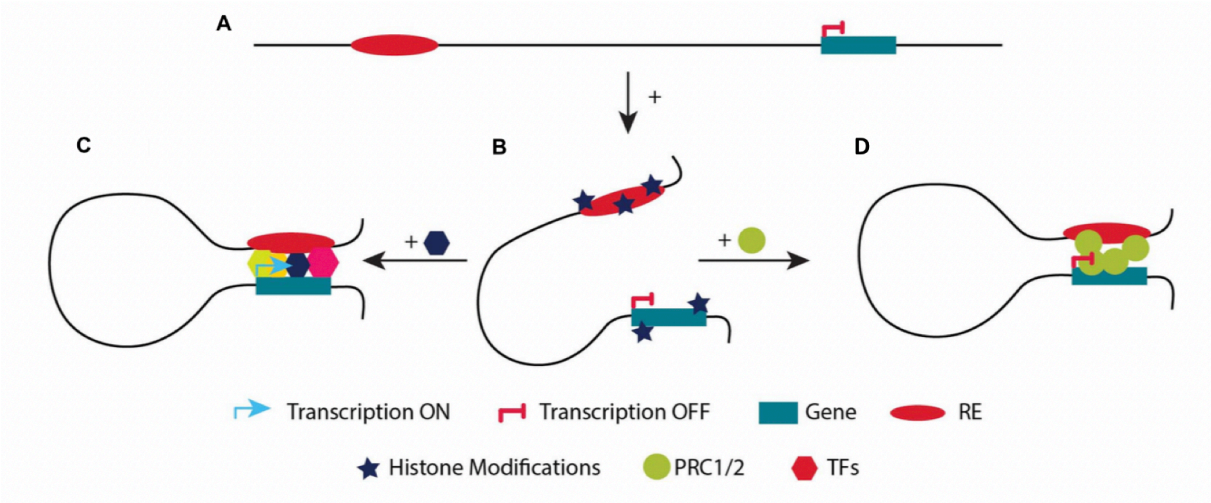


Figure 14 : Impact des modifications d'histones sur la transcription.
(**A**) Distance linéaire entre un élément régulateur (RE) et le gène qu'il contrôle. (**B**) L'ajout de marques répressives ou activatrices va permettre l'activation du gène suite à la fixation de facteur de transcription (**C**) ou bien sa répression via le recrutement des complexes répresseurs PRC1/2 (**D**). Adapté de (*155*)

ii) Méthylation de l'ADN

La méthylation de l'ADN consiste en l'ajout d'un groupement méthyl -CH_3 sur le carbone en position 5 d'une cytosine (5-mC). Cette méthylation survient au niveau de régions riches en dinucléotides cytosine-guanine appelées ilots CpG. Ces ilots sont, en conditions normales, méthylés pour plus de 70% d'entre eux. La plupart des promoteurs contiennent des ilots CpG dont le statut de méthylation est généralement corrélé à l'activité transcriptionnelle. Les gènes réprimés sont associés à une forte méthylation au niveau de leur promoteur tandis que l'hypométhylation est associée à une chromatine ouverte, accessible pour les TF et donc une transcription active (*156*).

Les cartographies à l'échelle du génome montrent que, au cours du développement ou de la différenciation cellulaire, les changements dans les profils de méthylation sont restreints et localisés principalement dans des enhancers au niveau de sites de fixation de TF (*157*, *158*). Ces changements au niveau des sites de liaisons des TF suggèrent que la méthylation permet de « guider » la différenciation en évitant des activations aberrantes de TF (*159*). Par exemple, on observe une hypométhylation des promoteurs des facteurs associés à la différenciation myéloïde durant celle-ci alors qu'ils sont au contraire hyperméthylés durant la différenciation lymphocytaire. On peut ainsi établir des profils de méthylation permettant de caractériser les différentes populations au cours de la différenciation hématopoïétique (*158*).

Les enzymes qui catalysent l'ajout d'un groupement méthyl sont des DNA méthyltransférases (DNMT). Il en existe 3 chez les mammifères, DNMT1 qui maintient la méthylation lors des divisons cellulaires et DNMT3A et DNMT3B qui permettent son établissement *de novo*. Dnmt3a et Dmnt3b présentent une redondance fonctionnelle et leurs K.O. chez la souris induit une forte augmentation des capacités d'auto-renouvellement ainsi que du nombre de HSC (*160*).

TET1, TET2 et TET3 permettent, à l'inverse, la déméthylation de l'ADN via l'oxydation du groupement méthyl ce qui aboutit à une hydroxyméthylcytosine (5-hmC). L'inactivation de Tet2 chez la souris aboutit à un blocage de la différenciation ainsi qu'à une augmentation du nombre de HSC (*161*). Le fait que les K.O. des enzymes qui méthylent et déméthylent l'ADN aboutissent au même phénotype suggèrent que ces 2 processus coopèrent pour réguler finement la transcription ce qui est confirmé dans des modèles de double K.O. (*162*).

iii) Organisation tridimensionnelle de la chromatine

La conformation 3D de l'ADN est essentielle dans la régulation de l'expression génique. En effet, le repliement du génome permet d'augmenter ou diminuer la proximité d'éléments régulateurs de leurs gènes cibles. Ainsi les interactions promoteur-enhancer changent durant la différenciation et sont caractéristiques des différents types cellulaires (*163*).

Le développement de techniques de capture de la conformation de la chromatine (3C) et ses dérivées (4C, 5C, Hi-C) a permis de caractériser l'architecture chromatinienne et de la diviser en plusieurs niveaux allant des territoires chromosomiques, aux compartiments chromatiniens A et B (définis par l'absence d'interaction chromatinienne et qui correspondent à l'euchromatine et l'hétérochromatine respectivement) aux domaines d'association topologique (TAD) jusqu'aux contacts promoteur-enhancer (*164*).

Les TAD forment des territoires génomiques de 500-1000kb isolés les uns des autres. Cette isolation correspond à une absence de contacts chromatiniens au-delà des frontières du TAD. En conséquence, les éléments régulateurs ne peuvent accéder qu'aux promoteurs au sein du même TAD. La formation des TAD est très conservée au sein des différentes espèces et types cellulaires. La structuration des TAD repose sur les protéines CTCF et le complexe cohésine (Figure 15).

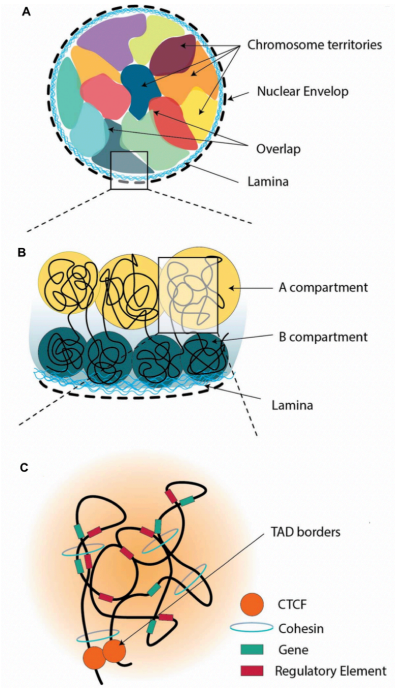


Figure 15 : Organisation 3D de la chromatine.

(A) Représentation schématique des chromosomes en territoires dans le noyau. (B) Organisation de la chromatine en compartiment A et B ; le compartiment B étant associé à la lamina. (C) Formation des TAD par "extrusion des boucles". Les protéines CTCF marquent les frontières des TAD au sein desquels des boucles permettent les interactions gène-élément régulateurs. Adapté de (*155*)

CTCF est une protéine de liaison à l'ADN fortement enrichie aux frontières des TAD et sa déplétion empêche leur formation (*165*). Les frontières des TAD sont formées par l'interaction de 2 protéines CTCF en orientation inverse (*166*). Durant la transition fœtale/adulte des HSC, les frontières des TAD sont renforcées et de nouveaux contacts enhancers-promoteurs sont formés (*167*).

Les cohésines constituent une famille comprenant les protéines : RAD21, SMC1, SMC3, STAG1 et STAG2. Les cohésines jouent un rôle essentiel dans les interactions enhancers-promoteurs au sein d'un TAD. En détails, les cohésines forment un anneau au travers duquel l'ADN va s'extruder jusqu'à la rencontre d'une protéine CTCF qui va ancrer l'anneau et bloquer la conformation de l'ADN. Ceci permet à des enhancers d'interagir avec des séquences promotrices situées à de grandes distances (*166*). Ainsi l'utilisation de souris K.O. pour Stag2 montre un changement de l'accessibilité de la chromatine associé à une répression de l'expression de certains facteurs ce qui empêche l'engagement en différenciation (*168*).

e) Intégration des différents niveaux de régulation transcriptionnelle

Ces différents niveaux de régulation transcriptionnelle agissent ensemble afin de contrôler finement l'expression des gènes durant la différenciation hématopoïétique. En effet, le remodelage de la chromatine, les modifications des histones, la méthylation de l'ADN et l'organisation 3D du génome modulent l'accessibilité de la chromatine, notamment au niveau des éléments régulateurs de la transcription comme les enhancers et les promoteurs, qui contrôlent la fixation des TF.

Le développement de techniques comme le test d'accessibilité de la chromatine par la transposase (ATAC-seq) ont permis de caractériser le paysage chromatinien. De façon intéressante, l'identification des types cellulaires est plus précise sur la base de leurs ouvertures chromatiniennes que de leurs transcriptomes (*169*).

L'intégration de plusieurs techniques (ChIP-seq, ATAC-seq, RNA-seq) permettant d'avoir un panorama comprenant à la fois le transcriptome, l'ouverture chromatinienne et la fixation des TF révèle que la différenciation induit une réorganisation graduelle de la chromatine dans laquelle des enhancers sont établis ou perdus avant que le changement du niveau d'ARN de leurs cibles ne soit détectable. Ainsi, les HSC possèdent un nombre important d'enhancers actifs ou « poised ». Durant la différenciation, les enhancers des lignages associés vont rester actifs et d'autres vont être établis *de novo*. Au sein d'un lignage, le profil des enhancers des progéniteurs ressemble plus à celui des cellules matures de ce lignage qu'à celui des progéniteurs d'autres lignages suggérant que ces changements s'opèrent précocement dans la différenciation (*170*). Dans une étude récente utilisant l'ATAC-seq au niveau de la cellule unique, Buenrostro *et al* confirment ces résultats. Ils observent que les HSC sont « primées » avec une chromatine ouverte au niveau de loci associés à tous les lignages. Au fur et à mesure que les HSC se différencient, la chromatine change avec une plus forte ouverture des loci associés au lignage dans lequel elles s'engagent et une fermeture des autres. Le rôle de cette ouverture de la chromatine dans l'accessibilité aux TF qui contrôlent la différenciation est souligné par la surreprésentation de motif HOX dans les HSC. A l'inverse, durant les différenciations érythro-mégacaryocytaire, myéloïde et lymphoïde, les loci ouverts sont enrichis respectivement en motifs GATA1, CEBP/A et ID3, 3 des principaux facteurs associés à ces lignages (*18*).

Enfin, Il convient de noter que ce n'est pas forcément la permissivité de la chromatine qui détermine l'action des TF mais parfois l'inverse. En effet, certains TF possèdent une activité dite « pionnière » qui leurs permet de se fixer sur une chromatine condensée et de l'ouvrir via le recrutement de partenaires (comme le complexe SWI/SNF) ce qui permet ensuite la fixation d'autres TF. Certains facteurs importants de l'hématopoïèse comme PU.1 où des membres de la famille GATA présentent cette activité pionnière (*171*).

4) Différences hématopoïèse *in utero* vs adulte

Il existe plusieurs différences importantes entre l'hématopoïèse chez l'adulte et les différentes vagues embryonnaires.

Tout d'abord, durant le développement plusieurs localisations et niches hématopoïétiques se succèdent, les principales étant le sac vitellin, l'AGM et le foie fœtal. Chez l'adulte, l'hématopoïèse se déroule exclusivement dans la moelle osseuse. Ce changement de microenvironnement impacte les cellules hématopoïétiques. Dans le foie fœtal, les HSC sont prolifératives et possèdent un fort potentiel d'auto-renouvellement (*172*) grâce à la sécrétion de cytokines comme le SCF, la TPO, l'IGF2, l'ANGPT3 ou le CXCL12 par les hépatocytes (*173*). Dans la moelle osseuse, les ostéoblastes secrètent de l'ostéopontine qui permet de réguler la quiescence des HSC (*174*) qui adoptent un profil adulte environ 3 semaines après la naissance (*172*).

Durant cette transition fœtale-adulte, les HSC vont également subir d'importants changements transcriptionnels et épigénétiques (*175*). Li *et al* ont montré que ces changements surviennent de façon non coordonnée dans chaque HSC sous l'influence de cytokines inflammatoires et notamment l'interféron de type I (*176*). Ainsi, on assiste à une bascule entre les régulateurs clés des HSC fœtales que sont SOX17 (*177*) et l'axe LIN28b-LET7-HMGA2 (*178*) en faveur de GFI-1 (*179*), TEL/ETV6 (*180*), BMI1 (*181*) et CEBPA (*182*). On a également un changement des gènes qui contrôlent l'organisation chromatinienne et notamment les interactions enhancer-promoteur avec TCF3 et MAFB dans les HSC fœtales contre NR4A1 et GATA3 dans les HSC adultes (*167*). Enfin, l'architecture de l'hématopoïèse varie également selon le stade de développement, ainsi les premières HSC dans l'AGM et le foie fœtal ont un fort biais érythroïde (*183*) alors que les HSC de la moelle fœtale génèrent plus de progéniteurs lymphocytaires ou granulocytaires (*184*) (Figure 16).

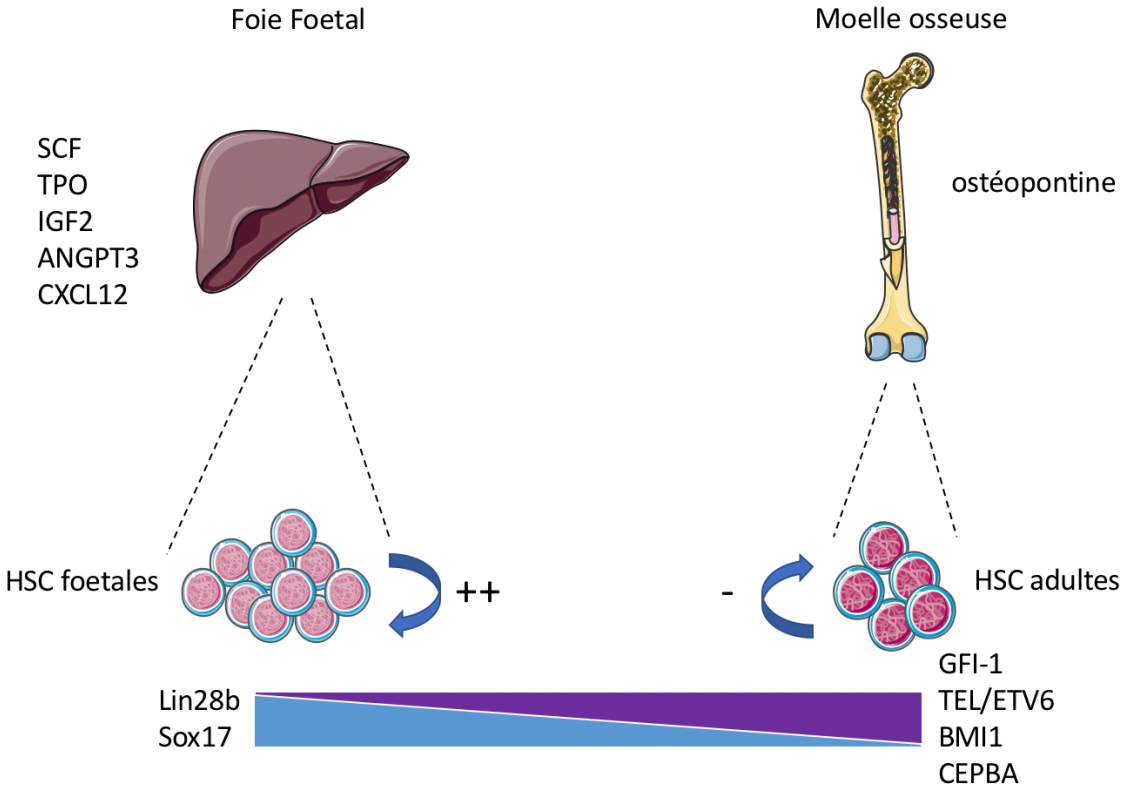


Figure 16 : Différences entre HSC fœtales et adultes.
Les changements dans les régulateurs transcriptionnels ainsi que dans les niches, qui secrètent des facteurs différents, induisent des différences dans les caractéristiques des HSC selon leur ontologie.

En dehors des HSC, les cellules des différents lignages présentent également des différences ontologiques. Ainsi les Mk-P fœtaux sont hyperprolifératifs et produisent plus de Mk. Ces Mk fœtaux sont moins polyploïdes (en moyenne 8N), plus petits et produisent moins de proplaquettes que les Mk adultes (*185*). Ces changements sont dus à une sensibilité différente à la TPO (*186*) ainsi qu'à des niveaux d'expression différents de facteurs importants dans la mégacaryopoïèse tels que GATA1 (*186*) ou LIN28b (*187*). Les érythrocytes subissent un changement dans le type de globines qu'ils expriment entre le stade embryonnaire (HbE, $\zeta 2\varepsilon 2$), fœtal (HbF, $\alpha 2\gamma 2$) et adulte (HbA, $\alpha 2\beta 2$) (*68*). Enfin, certains types de cellules ne sont produites que durant le développement comme les lymphocytes B de type B1-B (*188*) ou les cellules microgliales (*73*).

Ces différences entre l'hématopoïèse durant le développement *in utéro* et chez l'adulte peuvent être la source de différences entre les leucémies pédiatriques du jeune enfant et les leucémies chez l'adulte (discuté plus loin dans cette thèse).

II) Hématopoïèse Pathologique

On appelle hémopathies malignes l'ensemble des pathologies cancéreuses affectant les cellules du sang. Elles peuvent être associées aux différentes lignées lymphoïdes ou myéloïdes.

On distingue plusieurs types d'hémopathies malignes myéloïdes comme :

- Les syndromes myéloprolifératifs (SMP), caractérisés par une accumulation anormale de cellules d'un lignage qui présentent des caractéristiques de cellules matures.
- Les syndromes myélodysplasiques (SMD) qui présentent des altérations de la différenciation aboutissant à un défaut en cellules matures sans pour autant qu'il y est d'accumulation de progéniteurs.
- Les leucémies aiguës myéloïdes (LAM), qui seront plus spécifiquement détaillées ici, sont caractérisées à la fois par une prolifération incontrôlée des progéniteurs hématopoïétiques et un blocage de la différenciation. Cela engendre un défaut dans l'hématopoïèse normale qui peut se traduire par une baisse du nombre de cellules matures fonctionnelles des autres lignages pouvant conduire à une anémie, une neutropénie ou thrombopénie. Elles représentent des sous-types d'hémopathies généralement agressives qui peuvent survenir *de novo* ou faire suite à l'évolution d'un SMP/SMD.

1) Leucémies aiguës myéloïdes pédiatriques

Historiquement, Les LAM ont été classées selon la classification Franco-Américano-Britannique (FAB) sur des critères morphologiques associés au lignage ou au stade de différenciation (*189*)

Classification	Statut
LAM0	Leucémies aiguës avec différenciation minimale
LAM1	Leucémies aiguës sans maturation
LAM2	Leucémies aiguës avec maturation
LAM3	Leucémies aiguës promyélocytaires
LAM4	Leucémies aiguës myélomonocytaires
LAM5	Leucémies aiguës monocytaires
LAM6	Leucémies aiguës érythroïdes
LAM7	Leucémies aiguës mégacaryoblastiques

L'avancement dans la caractérisation des LAM a conduit au développement par l'organisation mondiale de la santé d'une nouvelle classification prenant en compte les altérations immunophénotypiques et génétiques dont la dernière révision date de cette année (*190*).

Les leucémies sont dites pédiatriques si elles surviennent chez des enfants âgés de moins de 15 ans. Selon l'âge au diagnostic, les leucémies pédiatriques peuvent être divisées en leucémies infantiles (patients de moins de 2 ans) et leucémies de l'enfant (patients entre 2 et 15 ans). Les LAM représentent 20% de l'ensemble des leucémies pédiatriques, le reste étant des leucémies aiguës lymphoblastiques (LAL), mais sont majoritaires dans les leucémies infantiles chez les très jeunes enfants (*191*).

2) Génétique et conséquences fonctionnelles des LAM pédiatriques

Les cellules sont soumises à l'apparition d'anomalies dans leur matériel génétique en réponse à des stress externes ou à des erreurs de réplication. Ces mécanismes sont importants car ils génèrent de la diversité génétique permettant dans certains cas aux cellules, et donc aux organismes, de s'adapter à des changements du milieu et sont à la base de la théorie de l'évolution.
Lorsqu'une mutation survient dans une cellule, elle peut être délétère ce qui va entrainer la mort de la cellule. Une mutation peut également être silencieuse c'est-à-dire n'avoir aucun effet sur la cellule. Enfin une mutation peut conférer un avantage sélectif par rapport aux autres cellules. Ce dernier cas peut conduire au développement de cancers.

Dans le cadre du système hématopoïétique, les mutations apportent généralement un avantage aux progéniteurs conduisant à l'augmentation de leur prolifération et/ou au blocage de leur différenciation et peuvent impacter n'importe quel processus biologique de la cellule. Les mutations des LAM sont généralement somatiques (c'est-à-dire acquises durant la vie d'un individu) et restreintes aux cellules hématopoïétiques. Cependant, il existe des mutations germinales conduisant à des syndromes prédisposant au développement de LAM.

a) Types d'altérations

Les techniques cytogénétiques permettant l'observation des chromosomes, comme le caryotype ou l'hybridation fluorescente in situ (FISH), permettent de caractériser des anomalies impactant la structure des chromosomes.

Ces anomalies peuvent modifier le nombre de chromosome avec des pertes ou des gains. C'est par exemple le cas de la trisomie 8, qui consiste en la présence d'un chromosome 8 surnuméraire, retrouvé dans environ 10% des LAM pédiatriques (*192*). La trisomie 21 est également fréquemment retrouvée chez les patients LAM soit comme anomalie additionnelle restreinte aux blastes, soit de façon germinale, prédisposant au développement leucémique. En effet, contrairement au risque de tumeur solide qui est diminué, les enfants trisomiques 21 ont un risque plus élevé de développer une leucémie (*193*).

Ces gains et pertes peuvent également être limités à une portion du chromosome. Ainsi, le gène codant la tyrosine kinase FLT3 est la cible de duplication interne en tandem (FLT3-ITD) qui est l'anomalie la plus fréquente chez les patients LAM pédiatriques. A l'inverse la perte d'une région d'un chromosome est également possible comme par exemple la délétion du bras long du chromosome 9 del(9q) (*192*).

Enfin des translocations peuvent aboutir à échanges de matériel génétique entre plusieurs chromosomes, la plus fréquente étant la translocation t(8;21)(q22;q22) qui fusionne une partie du chromosome 21 avec une partie du chromosome 8 aboutissant à la création d'une protéine chimérique de fusion RUNX1-ETO (également appelée RUNX1-RUNX1T1 ou AML1-ETO) (*192*). Ces échanges peuvent également survenir entre fragments d'un même chromosome, on parle alors d'inversion (ou de délétion) chromosomique telle que l'inversion du chromosome 16 conduisant à la fusion ETO2-GLIS2 qui sera abordée en détail (c.f. infra).

Le second type d'anomalies génétiques présentes dans les LAM pédiatriques sont les mutations ponctuelles, c'est-à-dire des mutations qui n'affectent qu'un petit nombre voir un seul nucléotide. Ces mutations peuvent conduire à une délétion (perte de nucléotides), une insertion (gain de nucléotides) ou une substitution (remplacement d'un nucléotide par un autre). Si elles surviennent dans une région régulatrice, cela peut impacter l'expression d'un gène. Dans une région codante, elles peuvent induire un changement dans la séquence en

acides aminés de la protéine (mutation faux-sens) ou l'apparition d'un codon stop qui aboutira à l'expression d'une protéine tronquée (mutation non-sens).

b) Description des mutations et conséquences fonctionnelles

L'avancé des techniques de séquençage à haut-débit et la mise en place de groupes collaborateurs entre plusieurs équipes permettant d'étudier de larges cohortes de patient (comme le Children's Oncology Group – National Cancer Institute TARGET AML aux USA) ont permis de mieux caractériser les mutations associées aux LAM pédiatriques.
Ainsi des mutations pouvant impacter des facteurs de transcription (ex : GATA2, RUNX1, CEPBA, WT1, ELF1), des régulateurs épigénétiques (ex : MLL, SETD2, ASXL2), des cohésines (ex : STAG2, SMC3), des voies de signalisations (ex: KIT, FLT3, NRAS, KRAS), l'épissage des ARN (ex : MBNL1) ou encore le transport nucléaire (ex : NUP98) ont été identifiées dans les LAM pédiatriques (*192*).
Les mutations de ces différents gènes peuvent avoir des effets et des conséquences fonctionnelles variables, les principales sont détaillées ci-dessous.

Les mutations activatrices conduisent à un gain de fonction se traduisant par l'activation anormale d'une protéine. C'est le cas des mutations FLT3-ITD qui permettent à ce récepteur à activité tyrosine kinase de se dimériser en l'absence de son ligand. Cette dimérisation entraine une phosphorylation et donc une activation constitutive du domaine tyrosine kinase qui permet l'activation de voies de signalisations participant à la leucémogenèse (*194*). D'autres mutations, comme N/K-RAS, bloquent l'activité GTPase ce qui empêche l'hydrolyse du GTP et maintien le récepteur constitutivement actif (*195*).

Les mutations inactivatrices conduisent à une perte de la fonction de la protéine mutée. Dans les cancers, certains gènes connus pour avoir une activité dite « suppresseur de tumeurs » comme TP53, qui participe au maintien de l'intégrité du génome, sont des cibles fréquentes de mutations inactivatrices. Ces mutations peuvent induire une haplo-insuffisance lorsqu'elles touchent un seul allèle ou une perte de fonction totale lorsque les 2 allèles sont mutés. Dans le cadre des LAM pédiatriques, SPRED1 qui est un régulateur négatif de la voie RAS/MAPK présente un niveau d'expression faible et des mutations hétérozygotes induisant une perte

de fonction sont décrites chez certains patients (*196*). Il convient de noter qu'un gène peut avoir un rôle de suppresseur de tumeur dans certains cas alors que dans d'autres contextes il peut être oncogène. C'est le cas par exemple de GATA2 dont des mutations inactivatrices ont été décrites comme agissant en synergie avec une perte d'un chromosome 7 chez des patients pédiatriques atteint de SMD ou de LAM (*197*) mais qui peut également participer à la transformation leucémique en ayant un rôle oncogénique (*198*).

Les mutations néomorphes ou créatrices de fonctions résultent d'un changement en acides aminés qui conduit une protéine à modifier sa fonction ou en acquérir une nouvelle. Dans les LAM pédiatriques, les mutations IDH1 ou IDH2 sont trouvées chez 10% des patients (*192*). Les protéines mutées peuvent ainsi convertir l'α-cétoglutarate (α-KG) en oncométabolite 2-hydroxyglutarate (2-HG). Ce 2-HG va rentrer en compétition avec l'α-KG et réduire l'activité des enzymes qui en sont dépendantes comme la méthyltransférase TET2. Ainsi les mutations des enzymes métaboliques IDH1/2 leur confèrent un rôle dans la régulation de la méthylation de l'ADN (*199*).

Certaines mutations peuvent également induire un changement dans le niveau d'expression d'un gène sans modifier sa séquence. Ces mutations affectent généralement des éléments régulateurs comme des promoteurs ou des enhancers. C'est par exemple le cas de l'inversion du chromosome 3 inv(3)(q21;q26) qui va relocaliser un enhancer distal de GATA2 à proximité du gène EVI1 et ainsi induire sa surexpression ce qui va conduire au développement de la leucémie (*200*).

c) Évolution clonale et Coopération oncogénique

Le développement des techniques de séquençage à haut débit a permis de grandes avancées dans la compréhension de l'architecture des leucémies. Il est possible de déterminer la fréquence d'allèle variant (VAF) d'une mutation, c'est-à-dire le nombre de cellules au sein de la population tumorale qui possède cette mutation et ainsi reconstruire la trajectoire évolutive de la leucémie. Ces analyses ont permis de conforter le modèle d'évolution clonale des leucémies dans lequel les cellules vont d'abord subir une ou plusieurs mutations qui leurs confère un avantage sélectif conduisant à une amplification clonale sans pour autant induire

une pathologie. Ensuite cette population clonale va acquérir un ou plusieurs évènements génétiques additionnels conduisant à la transformation leucémique.

Le séquençage à l'échelle de la cellule unique révèle que les populations leucémiques sont généralement composées de plusieurs clones ayant divergés suite à l'acquisition de mutations variées à différents temps de la leucémie. Cette diversité clonale est le résultat de pressions de sélections qui s'exercent sur les cellules leucémiques, d'abord en compétition avec les cellules hématopoïétiques normales pour les ressources du microenvironnement puis les unes avec les autres lorsque la maladie est plus avancée. Au cours du vieillissement les cellules hématopoïétiques accumulent un certain nombre de mutations conduisant à une réduction de la diversité du pool de HSC et donc à une hématopoïèse clonale formant un état pré-leucémique (*201*). Cette théorie d'évolution clonale est donc particulièrement adaptée aux LAM adultes mais reste valable pour les LAM pédiatriques. Une des pressions de sélection majeure que peuvent subir les leucémies sont les traitements par chimiothérapies qui peuvent induire des changements dans l'architecture clonale avec l'émergence de clones résistants qui deviennent dominants et « drive » la pathologie. De façon intéressante, le fait que plusieurs sous-clones résistants ayant des mutations différentes mais du même gène puissent émerger en réponse au traitement suggère qu'il s'agit bien d'un phénomène de sélection darwinienne et non d'une simple dérive liée à la réduction du pool de cellules leucémiques (*202*).

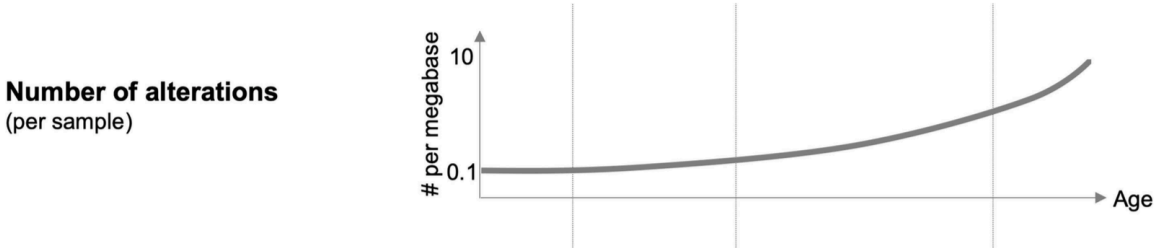


Figure 17 : Altérations génétiques dans les LAM pédiatriques.
Représentation schématique du nombre d'altérations génétiques dans les LAM selon l'âge du patient. Issue de (*203*)

Les LAM, et particulièrement les LAM pédiatriques, présentent un taux de mutation très faible en comparaison des autres types de cancer suggérant que dans certains cas un seul évènement génétique pourrait conduire à la leucémogenèse (Figure 17) (*192*). Cependant, dans la plupart des cas on retrouve plusieurs mutations au sein d'une même cellule

leucémique. La co-occurrence récurrente de certaines mutations à des fréquences supérieures à celles attendues par le simple hasard, suggèrent qu'elles coopèrent pour participer au processus leucémique.

Ainsi, dans les LAM pédiatriques, la mutation FLT3-ITD est fréquemment associée à des mutations impactant les gènes WT1 et NPM1 ou à la translocation chromosomique NUP98-NSD1 (*192*, *204*). Des expériences de greffe de progéniteurs hématopoïétiques murins surexprimant Flt3-ITD ou Nup98-Nsd1 montrent que, individuellement, ces mutations ne conduisent pas au développement leucémique. En revanche leur co-expression induit une leucémie extrêmement agressive (*205*). De manière similaire, l'utilisation de souris transgéniques portant les mutations Flt3-ITD et $Wt1^{R394W}$ donne lieu au développement de pathologies plus agressives que chacune des mutations seules (*206*).

Un autre exemple « modèle » pour l'étude de la coopération oncogénique et de l'évolution clonale dans la leucémogenèse est la leucémie aiguë mégacaryoblastique associée au syndrome de Down. Cette LAM pédiatrique, qui survient chez les enfants atteint d'une trisomie 21 constitutive, résulte de l'acquisition de mutation de GATA1 ainsi que de mutations additionnelles. Le développement de cette leucémie ainsi que caractéristiques moléculaires et cellulaires associées seront détaillées par la suite (c.f. infra).

Enfin il convient de noter que toutes les mutations ne sont pas d'égales importances. Les mutations participant au développement et la maintenance leucémique sont dites « driver » alors que les autres sont qualifiées de « passagères ».

d) Cellule d'origine de la leucémie

Les différences dans les caractéristiques phénotypiques et la réponse au traitement entre les leucémies pourrait être due en partie à une différence dans leur origine cellulaire, c'est à dire dans le type de cellule initialement muté qui conduit à la transformation. Cette hypothèse est soutenue par des expériences des greffes de différentes populations hématopoïétiques dans des modèles murins. Ainsi bien que l'oncogène MLL-AF9, retrouvé dans les LAM pédiatriques, puisse induire une LAM aussi bien à partir de HSC que de progéniteurs engagés en différenciation, l'agressivité et la résistance aux traitements sont plus grandes dans le cas de la transformation de HSC par rapport à des progéniteurs myéloïdes engagés. De plus des

différences transcriptionnelles et de méthylation de l'ADN sont également observées selon l'origine de la leucémie (*207*).

Plusieurs études portant principalement sur des oncogènes de fusion issus de translocation chromosomique ont cherché à déterminer la cellule (ou la population cellulaire) à l'origine de la leucémie. Ainsi la greffe de différentes populations purifiées montre que la transduction rétrovirale des fusions MLL-AF9, MLL-ENL ou MOZ-TIF2 est capable de transformer aussi bien des HSC que différents progéniteurs myéloïdes (*208*, *209*). D'autres altérations comme la fusion RUNX1-ETO sont retrouvées dans des cellules de tous les lignages hématopoïétiques chez les patients et sa surexpression conduit à une augmentation du pool de cellules immatures suggérant une activité oncogénique au niveau du compartiment souche (*210*). De plus, certains oncogènes comme la fusion BCR-ABL conduisent au développement d'une leucémie lorsqu'ils sont exprimés dans des HSC mais sont incapables de transformer des cellules plus engagées (*209*).

Une hypothèse pouvant expliquer ces différences est que certaines LAM, notamment chez l'adulte, nécessitent plusieurs mutations pour le développement leucémique. Or, l'intervalle entre l'acquisition des mutations pré-leucémiques et leucémiques chez l'homme peut être de plusieurs années. Seules les HSC ont une capacité d'auto-renouvellement et une durée de vie suffisante pour permettre une expansion clonale pouvant conduire à l'acquisition de mutations secondaires. A l'inverse, des oncogènes forts associés aux LAM pédiatriques, comme les fusions MLL, pourraient être suffisants pour la transformation et ainsi survenir dans n'importe quel type de progéniteur (*208*) (Figure 18).

Le développement des techniques de séquençage single-cell semble prometteur dans l'identification de l'origine des leucémies. Ainsi, dans un article récent, les auteurs comparent plusieurs types de leucémies à des cellules hématopoïétiques de référence. Ils observent que les cellules leucémiques ne forment pas des populations séparées mais sont connectées par une trajectoire développementale. Ils postulent que la caractérisation de cette trajectoire pourrait permettre d'identifier la cellule qui a initié la leucémie (*211*).

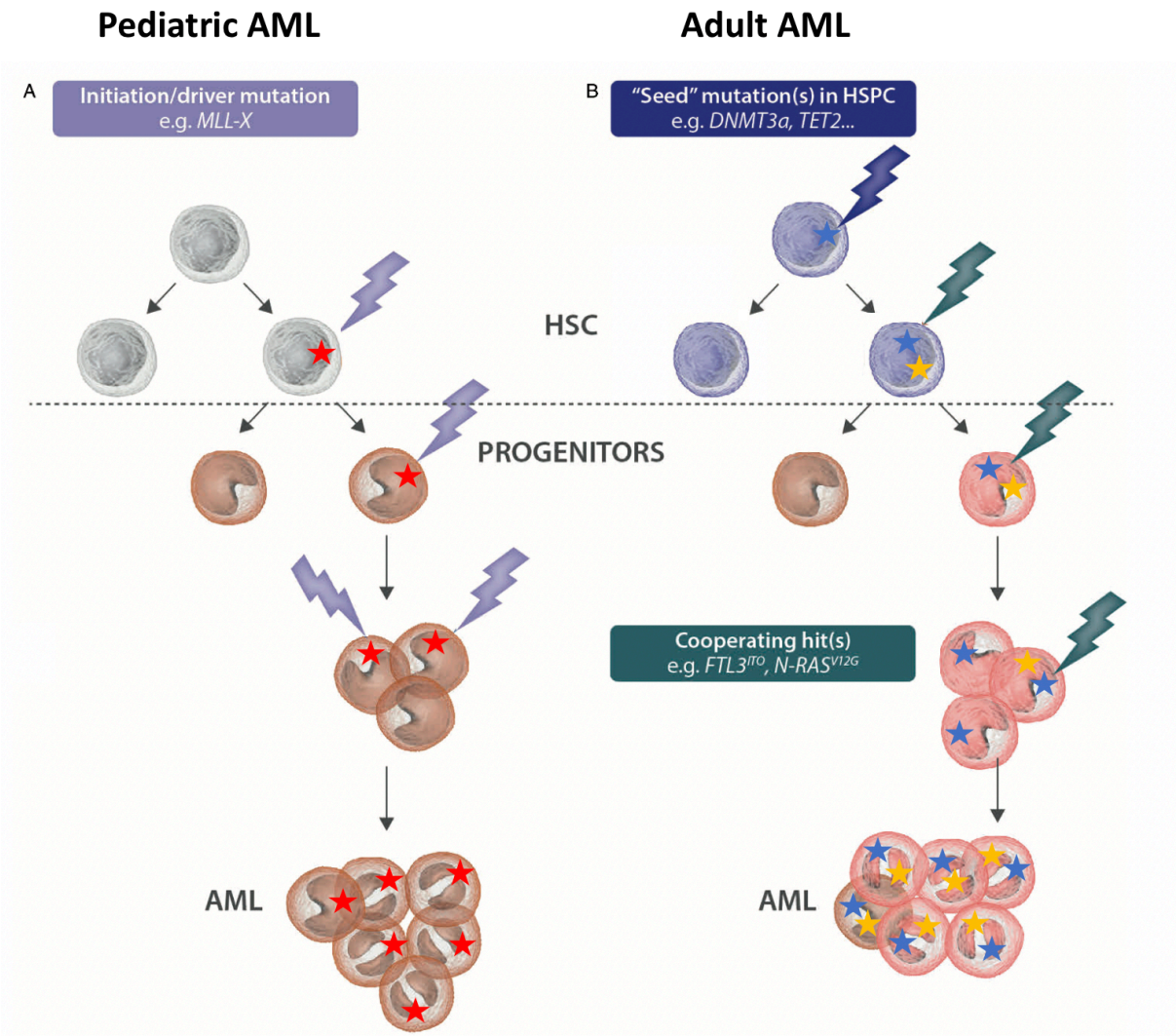


Figure 18 : Hypothèse sur l'origine cellulaire des LAM.
(**A**) Les fusions retrouvées dans les LAM pédiatriques pourraient transformer indifféremment des HSC ou des cellules plus engagées. (**B**) Dans les LAM adultes, une première mutation conférant un avantage sélectif survient dans les HSC. Les mutations coopératrices nécessaire pour le développement de la leucémie peuvent ensuite apparaitre dans les HSC ou dans les progéniteurs plus engages. AML = leucémie aiguë myéloïde. Adapté de (*208*)

e) Différences leucémies pédiatriques et leucémies de l'adulte

Plusieurs différences existent entre Les LAM pédiatriques, divisées en LAM infantiles (patient < 2 ans) et LAM de l'enfant (2 à 15 ans), et les LAM de l'adulte. Ces différences sont visibles dans la prévalence, l'agressivité et la réponse au traitement de ces leucémies.

La répartition des différents types de LAM varie grandement selon l'âge des patients. Ainsi, les LAM infantiles comportent une majorité de LAM5 alors que les LAM0/M1/M2 sont très peu représentées. Ces proportions sont inversées chez l'adulte. De plus, Les LAM

mégacaryoblastiques (M7) constituent une proportion importante des LAM chez les enfants de moins de 2 ans et sont quasiment inexistantes dans les autres classes d'âges (*212–214*) (Figure 19).

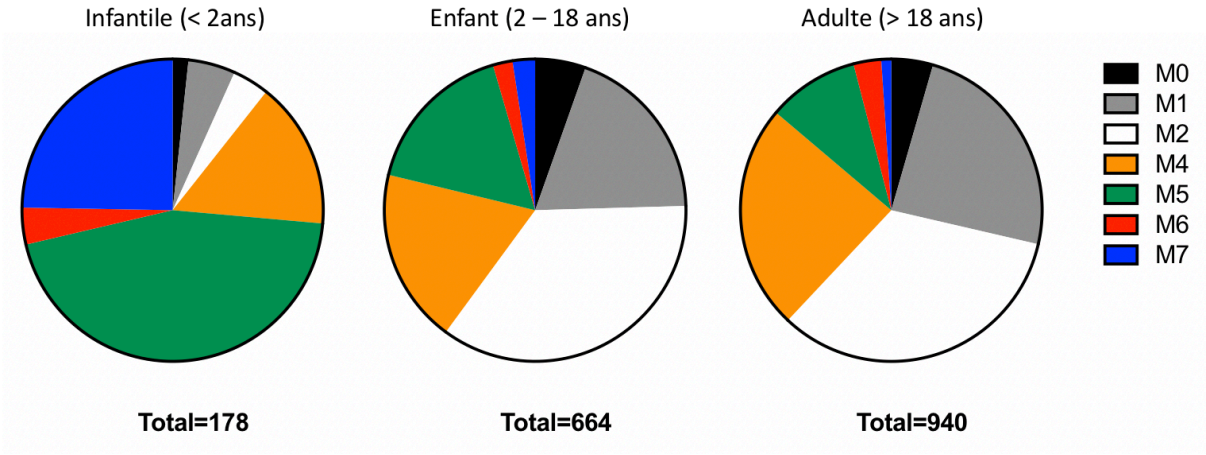


Figure 19 : Répartition des sous types de LAM selon l'âge.
D'après (*212–214*)

Les altérations génétiques sont également différentes entre les classes d'âge. Ainsi, les LAM pédiatriques sont fréquemment associées à un caryotype anormal, conduisant à des translocations chromosomiques, mais présentent en revanche un nombre réduit de mutations additionnelles. A l'inverse, dans les LAM de l'adulte les fusions chromosomiques sont plus rares mais le nombre de mutation est plus élevé.
Certaines mutations sont ainsi fortement associées aux LAM infantiles (fusions MLL, ETO2 et MNX1), de l'enfant (fusions RUNX1, CBFB, RARA) ou de l'adulte (mutations IDH1/2, NPM1, DNMT3A et TP53) (*192*).

L'origine *in utero* de certaines leucémies pédiatriques a été établie par l'observation de plusieurs paires de jumeaux monozygotes développant tous les deux la même leucémie et présentant la même altération génétique (*215*, *216*). Cette hypothèse est confirmée par la détection de mutation via des analyses rétrospectives de prélèvements sanguins effectués à la naissance sur des nouveau-nés ayant, par la suite, développés des leucémies pédiatriques. Ainsi, dans le cadre de LAL pédiatriques, une origine *in utéro* des fusions ETV6-RUNX1 (TEL-AML1), MLL-AF4, BCR-ABL, TCF3-PBX1 est clairement établie (*217*, *218*). De la même manière,

la présence de mutation à la naissance ainsi que des cas de jumeaux malades ont été décrit pour les LAM pédiatriques confirmant une origine prénatale (*215*, *219*).

Plusieurs études démontrent l'importance du stade ontogénique dans la transformation leucémique. Ainsi, la surexpression de Erg dans des modèles murins conduit au développement de leucémies lymphoïdes ou érythro-mégacaryocytaires selon l'origine respectivement adulte ou fœtale des cellules (*220*). De la même manière, l'expression des oncogènes de fusion MLL-AF9 chez la souris et NUP98-HOXA9 chez l'homme, dans des cellules adultes conduit exclusivement au développement de LAM. En revanche leur expression dans des cellules néonatales conduit à des leucémies pouvant présenter un phénotype myéloïde, lymphoïde ou mixte (*221*, *222*). De plus, même lorsqu'une LAM NUP98-HOXA9 se développe à partir de cellules néonatales, elle présente des différences transcriptionnelles avec les LAM adultes (*222*).
L'expression de FLT3-ITD chez la souris transforme efficacement des HSC adultes mais ne conduit pas au développement de leucémies dans des HSC fœtales ou néo-natales. Les HSC deviennent permissives à la transformation à partir de 3 semaines après la naissance, période à laquelle s'opère la bascule entre hématopoïèse fœtale et adulte des HSC suggérant qu'un contexte adulte est requis (*223*). Cette hypothèse est renforcée par le fait que l'interféron, qui contrôle en partie ce changement, peut sensibiliser des HSC fœtales à la transformation par FLT3-ITD (*176*). MLL-ENL transforme efficacement les HSC fœtales mais très mal les cellules adultes. De façon intéressante, les HSC néonatales conduisent à une leucémie encore plus agressive. La surexpression de Lin28b, un régulateur majeur de l'hématopoïèse fœtale, réduit cette efficacité. Cela suggère l'existence d'une fenêtre optimale pour la transformation leucémique entre la diminution de Lin28b à la naissance et l'acquisition des caractéristiques adultes des HSC (*224*).

Une des explications de ces observations pourrait être que les cellules cibles de certains oncogènes ne soient présentes que durant certaines étapes du développement. C'est par exemple le cas pour la fusion ETV6-RUNX1 qui cible un progéniteur issu de la vague primitive de l'hématopoïèse, inexistant chez l'adulte (*225*). Une autre cause de ces différences est le microenvironnement. Dans une étude récente, Rowe *et al* montrent que la greffe de cellules MLL-AF9 ou MLL-ENL dans des souris qui viennent de naître conduit à des leucémies mixtes

myéloïde/lymphoïde alors que ces mêmes cellules induisent des LAM pures dans des souris adultes. Ils démontrent que cela est dû à des différences de composition du microenvironnement (*226*). Ces résultats sont cohérents avec une autre étude indiquant que les LAM NUP98-HOXA9 adultes sont capables de remodeler leur microenvironnement à l'inverse des LAM NUP98-HOXA9 pédiatriques (*222*). Cependant, les causes exactes des différences entre LAM pédiatriques et adultes ne sont pas encore clairement établies.

3) LAM7 pédiatriques

Les leucémies aiguës mégacaryoblastiques (LAM7) sont des leucémies historiquement décrites comme impactant le lignage mégacaryocytaire. Les blastes présentent en effet une morphologie évoquant des progéniteurs mégacaryocytaires avec des cellules de grandes tailles présentant des irrégularités cytoplasmiques et pouvant exprimer à leur surface les marqueurs membranaires mégacaryocytaires CD41, CD61 et CD42. Chez l'adulte, les LAM7 sont extrêmement rares et généralement associées à une leucémie secondaire après traitement chimiothérapique. L'immense majorité des LAM7 survient chez de jeunes enfants de moins de 2 ans où elles sont associées à un pronostic défavorable (*227*).
Les LAM7 sont généralement divisées en 2 catégories selon qu'elles surviennent suite à l'évolution d'un syndrome pré-leucémique chez des enfants trisomique 21, on parle alors de LAM7 associée au syndrome de Down, ou sans prédisposition identifiée à ce jour et sont alors dites LAM7 *de novo*.

a) LAM7 associées aux trisomies 21 constitutives

Les leucémies mégacaryocytaires associées au syndrome de Down (DS-LAM7) représentent un modèle d'acquisition séquentiel de mutations conduisant à la leucémogenèse (Figure 20). Ces leucémies impactent exclusivement les enfants atteints du syndrome de Down c'est-à-dire ayant une trisomie 21 (T21) constitutive qui présentent un risque 500 fois supérieur à la normal. A la naissance 10 à 30% de ces enfants vont présenter une hémopathie appelée désordre myéloïde transitoire (TMD). Ce TMD est due à des mutations du régulateur central des différenciations érythrocytaire et mégacaryocytaire, GATA1. Ces mutations surviennent presque exclusivement dans l'exon 2 de GATA1 conduisant à l'apparition d'un codon stop

prématuré. La protéine ainsi formée, GATA1s (pour GATA1 short), est une forme tronquée conservant la capacité de se lier à l'ADN mais ayant perdu son domaine de transactivation. Ce TMD, qui se traduit par une hyper-prolifération des mégacaryoblastes, va dans la majorité des cas régresser spontanément. Cependant dans 10 à 20% des cas il va évoluer vers une leucémie suite à l'acquisition de mutations secondaires pouvant impacter des voies de signalisation (particulièrement l'axe JAK-STAT), des cohésines ou des régulateurs épigénétiques (notamment le complexe PRC2) (*228*, *229*).

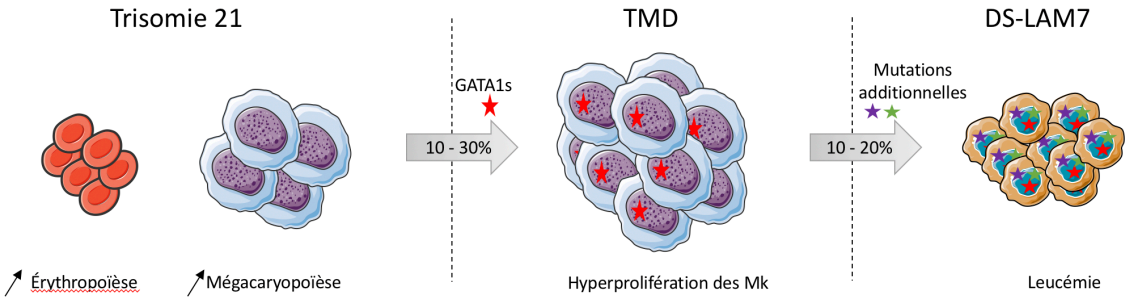


Figure 20 : Modèle d'évolution oncogénique de la DS-LAM7.
La trisomie 21 induit une augmentation du compartiment érythro-mégacaryocytaire. 10 à 30% des nouveau-nées vont acquérir une mutation GATA1s qui conduit à une hyper-prolifération pré-leucémique des mégacaryoblastes. 10 à 20% de ces enfants vont ensuite acquérir des mutations additionnelles conduisant à l'état leucémique.

Le rôle de la T21 dans le développement de la leucémie n'est pas clairement établi. Plusieurs gènes situés sur le chromosome 21 tel que RUNX1, ERG ou ETS2 sont des facteurs de transcription ayant un rôle dans la mégacaryopoïèse et décrit comme étant impliqués dans plusieurs LAM. Leur niveau d'expression n'est que faiblement augmenté, en accord avec un effet dose lié à la présence d'un 3ème chromosome mais pourrait permettre une surexpression de GATA1s (*230*). Plusieurs micro-ARN localisés sur le chromosome 21 participent également au processus de transformation en réprimant des gènes liés à la différenciation normale (*231*). La T21 impacte également la méthylation de l'ADN de plusieurs gènes impliqués dans la mégacaryopoïèse (*232*) et pourrait faciliter l'acquisition des mutations GATA1s (*233*). Il est intéressant de noter que l'effet de la T21 dépend du stade de développement. En effet, dans une hématopoïèse primitive, la T21 induit uniquement un biais vers la différenciation érythroïde (*234*). En revanche, dans le foie fœtal cet effet s'accompagne d'une augmentation de la mégacaryopoïèse ainsi qu'un auto-renouvellement accru des HSC (*235*).

La mutation GATA1s conduit à un défaut de l'érythropoïèse associé à une forte augmentation de la mégacaryopoïèse liée à des changements dans la fixation à l'ADN de GATA1s possiblement due à des changements dans les interactions protéiques par rapport à la forme sauvage (*236*).

Dans un récent article, le groupe de Eric Lechman a utilisé des cellules primaires humaines de foie fœtal disomiques ou trisomiques modifiées par CRIPSR-Cas9 pour introduire diverses mutations afin de caractériser le contexte cellulaire participant à la transformation leucémique. Ils montrent ainsi que la T21 est nécessaire pour le développement pré-leucémique initié par GATA1s mais est dispensable pour la progression vers une leucémie suite l'ajout de mutations additionnelles. L'importance du stade ontogénique est également soulignée par l'obtention de leucémie uniquement à partir de progéniteurs $CD34^+$ issus de foie fœtal ou de sang de cordon et non de cellules adultes. De plus bien que le double mutant GATA1s + STAG2 K.O. (une cohésine souvent mutée dans les DS-LAM7) puisse induire une leucémie à partir de n'importe quel type de progéniteur, ces deux mutations seules conduisent à une prise de greffe uniquement dans des HSC dont ils concluent qu'elles représentent les cellules à l'origine des DS-LAM7 (*237*).

b) LAM7 de novo

Dans la majorité des cas, Les LAM7 *de novo* présentent des remaniements chromosomiques conduisant à l'expression d'oncogène de fusion. Les analyses cytogénétiques et les avancées dans le domaine du séquençage ont permis d'identifier de nombreuses fusions génétiques dans les LAM7 (Figure 21). Selon le type de fusions, elles peuvent être associées à diverses mutations additionnelles. Les principales fusions et les mécanismes moléculaires associés sont présentés ci-dessous.

Molecular subgroups	Frequencies	Putative drivers
ETO2–GLIS2	18–27%	ETO2–GLIS2
HOX rearranged (HOXr)	15%	EP300–HOXA7
		NIPBL–HOXA9
		GATA2–HOXA9
		HOXA9–ANGPT1
		GATA2–HOXA10
		HOXA10–AS–CD164
		HOXA11–BZW2
		PLEK–HOXA11–AS
		C8orf76–HOXA11–AS
		EWSR1–HOXB8
		NIPBL–HOXB9
		BMP2K–HOXD10
NUP98–KDM5A	9–15%	NUP98–KDM5A
DS-like[b]	9–11%	Trisomy 21 + GATA1s
MLL rearranged (MLLr)	7–17%	MLL–AF4
		MLL–AF9
		MLL–AF10
OTT–MAL	5–10%	OTT–MAL
Others	Four cases	TLS–ERG
	One case	MN1–FLI1
	One case	BCR–ABL1
	One case	MAP2K2–AF10
	13%	Unknown

Figure 21 : Génétique des LAM7 de novo.

Fréquence des différents sous-groupes génétiques de LAM7 *de novo* et listes des fusions identifiées. D'après (*238*)

La translocation t(1;22)(p13;q13) est la première à avoir été identifiée. Retrouvée dans 10% des patients, elle conduit à la fusion OTT-MAL qui implique la quasi-totalité des 2 protéines. OTT (aussi appelé RBM15) est un régulateur transcriptionnel, qui est recruté par RBPJ et participe à l'inhibition des cibles de la voie Notch en absence de stimulation (*239*, *240*). MAL (aussi appelé MKL1) est un coactivateur transcriptionnel de SRF. A la fois MAL et SRF sont important dans la mégacaryopoïèse et leur inactivation chez la souris entraine une thrombopénie liée à des défauts dans la maturation terminale (*241*). Un modèle murin de knock-in de la fusion montre une activation anormale des cibles de la voie Notch. Cependant, le développement de LAM7 dans ce modèle est peu pénétrant (moins de 10% des animaux) suggérant la nécessité soit de mutations additionnelles, soit d'un modèle différent (*240*). Le fait que les patients OTT-MAL soient encore plus jeunes que les autres LAM7 et la présence de cette translocation chez des jumeaux monozygotes suggère une origine *in utéro* (*242*).

La protéine NUP98 est une nucléoporine participant à la formation des pores nucléaires mais possédant également une activité transcriptionnelle. NUP98 est fusionnée à de nombreux

gènes dans les LAM. Dans le cadre des LAM7 son partenaire principal est l'histone déméthylase KDM5A (aussi nommé JARID1A) retrouvée dans 10% des cas (*243*). Le processus de leucémogenèse implique le recrutement de l'acétyltransférase CREBBP/P300 ainsi que d'un large complexe protéique, permettant le dépôt de la marque H3K4me3. Cela conduit à une activation transcriptionnelle, les cibles principales étant les gènes HOX (*243*, *244*). L'expression lentivirale de NUP98-KDM5A dans des progéniteurs humains conduit au développement de différents types de leucémies et seulement dans un faible nombre de cas à des LAM7 contrairement à ce qui retrouvé chez les patients (*245*). Cette différence pourrait s'expliquer par la nécessité de mutations additionnelles. Par exemple RB1, un partenaire de GATA1, est muté dans la grande majorité des patients NUP98-KDM5A LAM7 (89%) (*238*).

Les réarrangements du gène codant pour l'histone méthyltransférase MLL (aussi appelée KMT2A) conduisent à des fusions avec plus de 90 protéines différentes dans les différentes leucémies. Dans les LAM7 *de novo*, les partenaires retrouvés sont AF4, AF10 et AF9 (*246*). MLL transforme les cellules via son activité propre mais également grâce à des partenaires comme DOT1L, une autre histone méthyltransférase qui régule positivement des gènes critiques pour la leucémogenèse. La fusion entraine une expression aberrante des gènes HOXA et peut également agir directement sur la Pol II en recrutant un complexe de super élongation (*247*). Plusieurs modèles murins montrent la capacité transformante des fusions MLL mais la majorité ne présente pas un phénotype mégacaryoblastique de LAM7 (*248*).

Les gènes HOX qui sont dérégulés dans le cadre des fusions NUP98-KDM5A et des réarrangements MLL peuvent eux-mêmes être impliqués dans des fusions dont le potentiel leucémique a été démontré pour certaines (*249*).

Parmi les LAM7 *de novo* ne présentant pas de fusions chromosomiques, certaines sont qualifiées de « DS-like » car présentant un profil moléculaire très similaire aux DS-LAM7. En effet, ces patients présentent dans tous les cas une T21, des mutations GATA1 et le même spectre de mutations additionnelles que dans les DS-LAM7. De plus, le fait que le pronostic de ces leucémies DS-like soit identique à celui des DS-LAM7 suggère que les mécanismes d'actions sont très similaires. Une différence notable est que la plupart de ces patients semblent présenter une T21 acquise, restreinte au compartiment hématopoïétique. De façon

intéressante, une T21 acquise est également retrouvée fréquemment dans tous les types de LAM7 (*238*, *250*).

c) ETO2-GLIS2

L'altération la plus courante dans les LAM7 pédiatriques *de novo* est la fusion ETO2-GLIS2, retrouvée dans 20-25% des cas et associée à un mauvais pronostic (*250–253*).

Les 2 gènes impliqués dans la fusion ETO2-GLIS2 sont localisés en orientation inverse aux extrémités du chromosome 16. L'inversion du chromosome 16 peut conduire à différents transcrits de fusion entre les exons 10, 11 ou 12 de ETO2 et les exons 1, 2 ou 3 de GLIS2. La fusion la plus commune est entre l'exon 11 de ETO2 et le 3 de GLIS2 (*252–254*) (Figure 22).

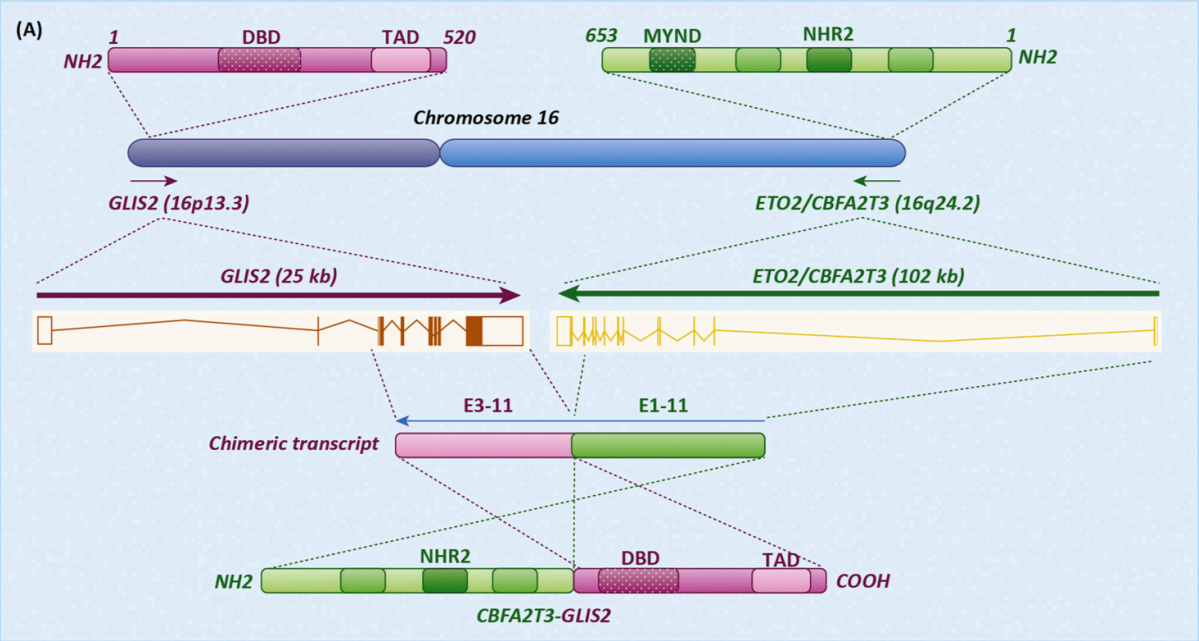


Figure 22 : Fusion ETO2-GLIS2.
Représentation schématique de l'inversion du chromosome 16 qui conduit à la fusion de la partie 5' de ETO2 en phase avec la partie 3' de GLIS2. La fusion la plus fréquente entre exon 11 de ETO2 et exon 3 de GLIS2 est représentée. Cette fusion conserve le domaine NHR2 de ETO2 et les domaines DBD et TAD de GLIS2. DBD = domaine de liaison à l'ADN ; TAD = domaine de transactivation. D'après (*255*)

Très peu de mutations additionnelles sont associées à cette fusion, les plus notables étant des anomalies de caryotype comme la T21, retrouvée chez 40-50% des patients (*250*, *251*, *253*). Cependant, dans une étude récente sur une plus grande cohorte, la fréquence de T21 est de seulement 11% chez les patients EG et une trisomie du chromosome 3 est observée chez 20%

des patients (*254*). Des mutations des voies JAK sont également rapportées chez 20% des patients (*250*, *253*). Enfin, une étude rapporte chez certains patients la présence d'un transcrit chimérique DHH-RHEBL1, sans qu'il n'y ai de fusion des deux gènes, contiguës sur le génome (*256*).

ETO2 (aussi appelée CBFA2T3 ou MTG16) est un cofacteur transcriptionnel. ETO2 contient 4 « nervy homology regions » (NHR) qui permettent son oligomérisation et l'interaction avec d'autre régulateurs. Exprimé dans la plupart des cellules hématopoïétiques, ETO2 interagit avec les protéines du complexe heptade (*257*) et participe au maintien des HSC (*258*) ainsi qu'à la répression des étapes terminales des différenciations érythroïde et mégacaryocytaire (*257*). L'implication de ETO2 dans différentes LAM a également été rapportée (*192*, *259*).
GLIS2 est un facteur de transcription de la famille GLI-similar 1-3 (GLIS1-3) qui présente une forte homologie avec les facteurs GLI impliqués dans la voie de signalisation Hedgehog. GLIS2 comporte un domaine de liaison à l'ADN constitué de 5 motifs en doigts de zinc, un domaine de transactivation en C-terminal et un domaine répresseur potentiel en N-terminal (*255*). Plusieurs études rapportent que GLIS2 n'est pas exprimé dans le système hématopoïétique. Cependant son inactivation par shRNA réduit la capacité de repopulation des HSC (*260*) et GLIS2 est impliqué dans le mécanisme de leucémogenèse de la fusion MOZ-TIF (*261*) suggérant un potentiel rôle dans l'hématopoïèse. Les GLIS participent également aux processus de différenciation, ainsi GLIS1/3 facilitent la reprogrammation de cellules somatiques vers les stades pluripotents alors qu'à l'inverse, GLIS2 inhibe ce processus (*262*).

La surexpression de la fusion ou de chacun des partenaires seuls dans des cellules murines ou humaines montre que, bien que GLIS2 et ETO2-GLIS2 induisent un phénotype mégacaryocytaire, seule la fusion permet l'augmentation des capacités d'auto-renouvellement (*263*). Des analyses de RNA-seq révèlent que la fusion induit un programme transcriptionnel distinct de celui de GLIS2 ou ETO2 avec une activation des voies Hedgehog, JAK-STAT, Hippo, TNF ainsi que de BMP2/4 (*252*–*254*). De façon intéressante, parmi les facteurs de transcription dérégulés on trouve la balance ETS/GATA. Ainsi ERG, un régulateur important de l'autorenouvellement des progéniteurs mégacaryocytaires, est fortement up-régulé et à l'inverse GATA1 qui permet la différenciation terminale est complètement down-régulé (*263*). Des analyses de ChIP-seq révèlent que la fusion est capable de se lier sur les loci

« canoniques » de ETO2 mais également à de nouvelles localisations. Ces nouvelles localisations sont souvent associées à des super-enhancers et sont co-fixées par la fusion et par ERG. De plus l'inactivation de ERG bloque l'activation des gènes cibles (ex : KIT) soulignant son rôle dans le mécanisme d'action de la fusion (*263*).

L'observation que les patients ETO2-GLIS2 LAM7 sont beaucoup plus jeunes que les patients ETO2-GLIS2 d'autres types de LAM a conduit notre équipe à étudier le phénotype induit par la fusion en fonction de l'ontogénie et de la hiérarchie cellulaire. L'utilisation d'un modèle murin inductible à la doxycycline montre que des cellules issues de foie fœtal induisent une leucémie avec une latence très courte et un phénotype mégacaryocytaire. A l'inverse des cellules issues de moelle adulte induisent une leucémie avec une latence prolongée et un phénotype myéloïde. De plus, dans le foie fœtal à la fois les HSC et les progéniteurs plus engagés (progéniteurs multipotents [MPP] 2/3) induisent un phénotype exclusivement mégacaryocytaire. A l'inverse, avec de la moelle adulte, seules les HSC conservent ce phénotype purement mégacaryocytaire, les MPP conduisant à un phénotype myéloïde. Des analyses de scRNA-seq et d'ATAC-seq montrent que ces différences sont associées à la fois à des niveaux basaux différents selon l'ontogénie entre les facteurs Gata1 et Erg d'une part et Cebpa et Spi1 d'autre part, ainsi qu'à une dérégulation transcriptionnelle due à la fusion plus forte dans les cellules fœtales (*264*) (Figure 23).

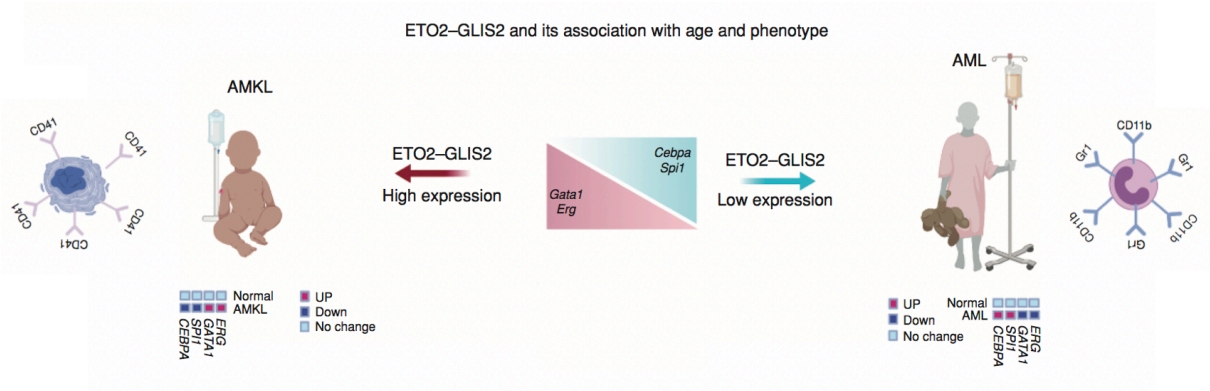


Figure 23 : Association entre phénotype et âge chez les patients ETO2-GLIS2.
Chez les très jeunes patients, la fusion ETO2-GLIS2 induit un phénotype mégacaryocytaire associé à des niveaux plus élevés de GATA1 et ERG. Les patients plus âges présentent un phénotype myéloïde associé à des niveaux plus élevés de CEBPA et SPI1. Adapté de (*265*)

III) Les cellules Pluripotentes pour l'étude de l'hématopoïèse humaine

Depuis leur découverte il y a une quinzaine d'années, l'utilisation des cellules souches pluripotentes induites est en pleine expansion. En effets, ces cellules présentent de nombreux avantages pouvant pallier à certaines des limitations des modèles traditionnellement utilisés en recherche biomédicale. Les cellules sanguines sont parmi les premières pour lesquelles des protocoles de différentiation d'iPSC ont été développés et de nombreuses équipes ont tiré parti de ce nouvel outil afin de modéliser les hémopathies malignes chez l'homme.

1) Modèles d'étude en hématologie

Parmi les premiers outils développés pour l'étude des hémopathies humaines on trouve les lignées cellulaires. Ces lignées cellulaires sont établies à partir de culture *in vitro* d'échantillons primaires de patients. Elles représentent un moyen simple d'étudier certains mécanismes associés à une pathologie et nécessitant un grand nombre de cellules homogènes (ex : analyses biochimiques, test de réponses à des molécules chimiques). Cependant, elles sont issues d'un processus de sélection pour leur capacité à pousser *in vitro*, en présence de stimuli exogènes et sans les stimuli issus de la niche pouvant conduire à l'acquisition de mutations sans relevance pour la pathologie. Ainsi, ces lignées peuvent parfois être éloignées de la situation initiale des cellules de patient, qui, dans la très grande majorité des cas se différencient ou meurent rapidement lorsqu'elles sont mises en culture. De plus l'accumulation des passages cellulaires peut entrainer une instabilité génomique. Ainsi, le profil d'expression des gènes et les mécanismes biologiques associés ne sont souvent pas représentatifs de la maladie vue chez le patient (*266*).

La nécessité de prendre en considération l'environnement physiologique des cellules a conduit au développement de modèles animaux *in vivo*. La facilité d'élevage et de reproduction ainsi que la possibilité de travailler sur un fond génétique homogène a fait de la souris un modèle majeur. Il est ainsi possible de générer des lignées transgéniques afin d'étudier l'impact d'une ou plusieurs mutations. Ces mutations peuvent être constitutives ou conditionnelles, présentes dans toutes les cellules ou restreintes à un tissu. De plus des

expériences impossibles chez l'homme pour des raisons techniques ou éthiques sont possibles chez la souris permettant l'étude de certains processus comme le développement.
Cependant, plusieurs différences importantes existent entre la souris et l'homme incluant la structure du génome, la durée de vie et la permissivité à la transformation cancéreuse. Par exemple, bien que le profil de méthylation de l'ADN soit partiellement conservé entre rongeurs et humains, des différences importantes existent (*267*). De plus, l'organisation chromosomique diffère entre les 2 espèces. Ainsi la région homologue au chromosome 21 humain est repartie entre les chromosomes 10, 16 et 17 chez la souris compliquant l'étude de son impact. Dans le cadre des leucémies pédiatriques, les modèles murins ne récapitulent pas toujours les pathologies observées chez les patients. Par exemple, la surexpression de l'oncogène pédiatrique MLL-AF4 induit des LAL chez l'homme alors qu'elle ne produit que des LAM chez la souris (*268*). De même, bien que des modèles murins pour plusieurs fusions retrouvées dans les LAM7 conduisent au développement de leucémies, seules les fusions MN1-FLI1 et ETO2-GLIS2 induisent des leucémies mégacaryoblastiques similaires à celles retrouvées chez les patients et de façon très dépendante du modèle utilisé (*203*).

L'utilisation de cellules primaires humaines peut pallier à certains de ces problèmes mais ces cellules peuvent s'avérer impossibles à cultiver *in vitro* sur des périodes longues nécessaires à l'observation de la transformation leucémique. De plus, l'accès à de nombreux types cellulaires (ex : cellules fœtales, cellules de patients avec mutations rares) est limité, voire impossible dans certains cas. Le développement de souris immunodéficientes NOD.Cg-Prkdc^{scid}IL2rg^{tm1Wjl}/SzJ (NSG) puis de souris humanisés (exprimant certaines cytokines humaines) représente une avancée majeure (*269*). Ces souris privées de lymphocytes B, T et NK permettent la prise de greffe de cellules hématopoïétiques humaines. Cela a conduit au développement de xénogreffes dérivées de cellules de patient (PDX). Ces PDX permettent d'amplifier les cellules de patients et ainsi d'étudier la biologie des pathologies associées. Cependant ces modèles s'avèrent difficiles à générer et toutes les cellules de patient ne parviennent pas à prendre en greffe dans des souris NSG (*270*). De plus ces PDX représentent le stade terminal de la pathologie car issue d'échantillons de patient et ne permettent pas d'étudier les stades plus précoces d'initiation de la leucémie.

Une alternative intéressante pour pallier à ces différent problèmes, permettant de suivre des processus dynamiques et d'avoir un accès récurent et contrôlé à des cellules humaines, est l'utilisation de cellules souches pluripotentes.

2) Concept de pluripotence

La pluripotence est définie comme la capacité pour une cellule à s'auto-renouveler indéfiniment et à se différencier vers les 3 feuillets embryonnaires (ectoderme, mésoderme et endoderme) et tous leurs dérivés tissulaires. Durant le développement, les cellules pluripotentes sont retrouvées dans la masse cellulaire interne (MCI) de l'embryon et ne participent pas à la formation des annexes extra-embryonnaires (*271*). Cette pluripotence est conservée jusqu'à la gastrulation lorsque les premières cellules somatiques et germinales émergent de l'épiblaste. Cet état de pluripotence peut être capturé et maintenue indéfiniment (en théorie) *in vitro* par l'isolation de cellules de la MCI durant une courte fenêtre entre le stade blastocyste et la gastrulation. Chez la souris, des cellules souches embryonnaires (mESC) ont ainsi pu être isolées à partir de l'épiblaste préimplantatoire (*272*). Le maintien de ces mESC requiert des conditions particulières comme la co-culture sur des cellules nourricières (« feeder cells ») ou l'utilisation de composés chimiques comme le LIF (un activateur de STAT3) et les « 2i » qui sont des inhibiteurs de GSK3 et MEK (CHIR99021 et PD0325901 respectivement) (*273*). Les mESC ainsi cultivées conservent leur pluripotence comme le démontre leur capacité à s'intégrer dans un blastocyste et à contribuer à la génération de souriceaux « chimères » viables. Ces mESC sont qualifiées de « naïves » car elles peuvent être dirigées indifféremment vers n'importe quel destin cellulaire (*271*).

D'autre type de cellules ES ont été isolées chez la souris à partir de l'épiblaste à des stades plus tardifs (post-implantation) (*274*, *275*). Ces cellules, appelées cellules souches épiblastiques (mEpiSC), conservent le potentiel pour se différencier vers les 3 feuillets embryonnaires mais ne sont plus capables de s'intégrer dans un blastocyste et de participer à la formation de chimères. De plus, ces mEpiSC ne peuvent survivre dans les conditions de cultures utilisées pour les mESC (*273*). Il existe donc une différence claire de potentialité entre les mESC et les mEpiSC qui sont qualifiées de mESC « primées ». Ces changements sont reliés à des différences au niveau du métabolisme (*276*), de la chromatine et du transcriptome (*277*) (Figure 24). Les mEpiSC dépendent d'un métabolisme qui repose fortement sur la glycolyse

anaérobie. A l'inverse, les mESC naïves présentent un métabolisme s'appuyant sur la phosphorylation oxydative dans la mitochondrie. Ces différences peuvent impacter l'épigénome car de nombreux régulateurs comme les HDAC ou les DMNT nécessitent des cofacteurs qui sont des produits de cette phosphorylation oxydative (*276*). Ainsi les mESC naïves présentent un ADN hypométhylé et une chromatine « relaxée » avec par exemple l'expression des 2 chromosomes X. A l'inverse dans les mEpiSC, beaucoup des gènes présentent un promoteur « poised » avec des marques activatrices et répressives et seulement un chromosome X actif (*277*). Cette état « primé » pourrait permettre de sensibiliser les cellules aux signaux de différenciation.

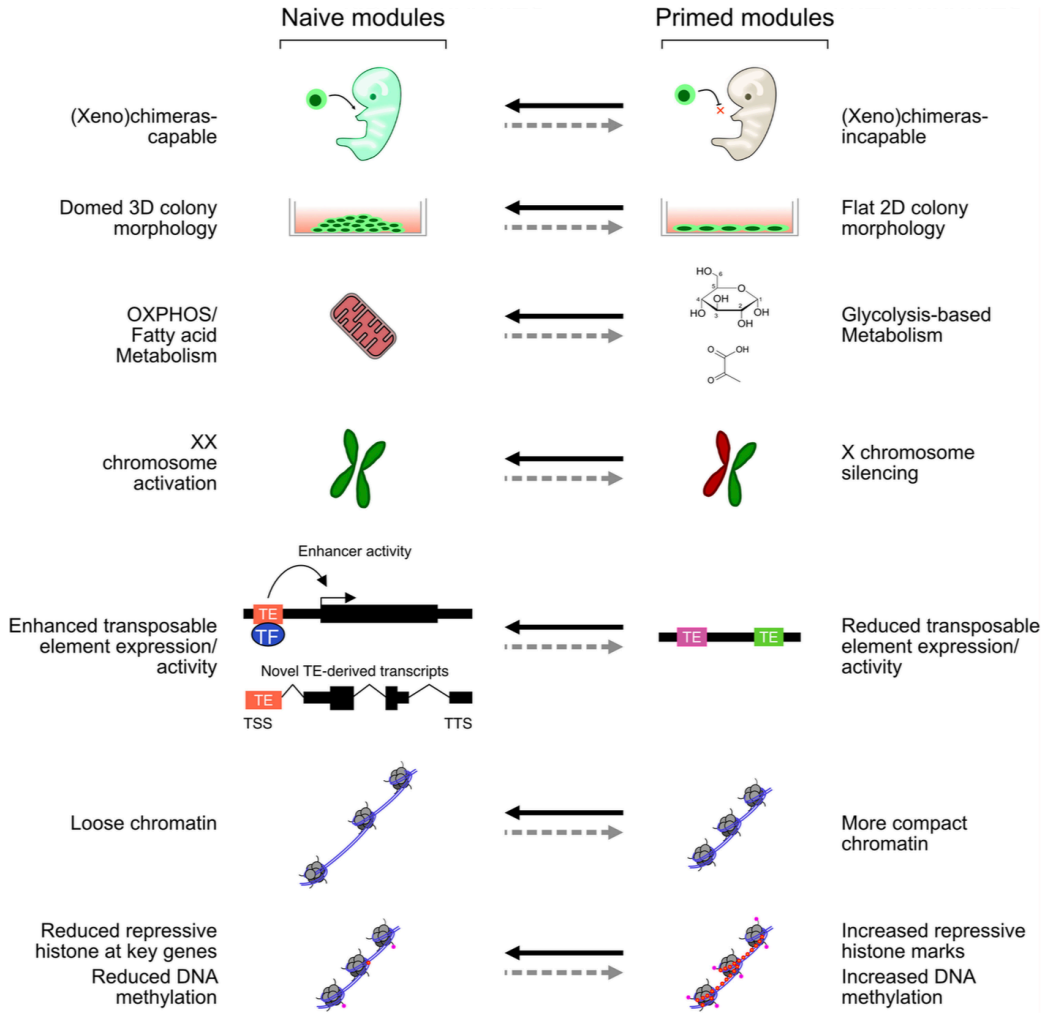


Figure 24 : Pluripotence naïve vs primée.
Principales différences entre des cellules pluripotentes naïves ou primées. Issue de (*277*)

Chez l'homme, des cellules souches embryonnaires (hESC) ont également été isolées (*278*). Le réseau transcriptionnel responsable de la pluripotence est largement conservé entre mESC et hESC avec notamment les régulateurs majeurs que sont OCT4, SOX2 et NANOG. La moitié des gènes fixés par ces facteurs sont activés alors que l'autre moitié, associée à l'engagement vers des lignages spécifiques, est réprimée. De façon intéressante, les promoteurs des gènes réprimés présentent un profil bivalent avec les marques H3K4me3 et H3K27me3 rappelant les mEpiESC (*277*). De plus, les hESC ne peuvent pas être cultivées dans les mêmes conditions que les mESC. Le LIF par exemple, n'a que peu d'effet dû à une plus faible activité STAT3 que chez la souris. Les conditions de cultures des hESC sont plus proches de celles des mEpiESC avec notamment une dépendance à la signalisation FGF et à la voie TGF-B/Activin/Nodal et ses effecteurs en aval SMA2/3 qui permettent une up-régulation de NANOG (*279*). Ainsi les hESC représentent un état de pluripotence « primé » plutôt que naïf. Plusieurs groupes ont concentré leurs efforts pour convertir des hESC primés vers un état naïf (*280*) mais des différences transcriptionnelles avec les mESC subsistent et le seul test fonctionnel pouvant valider cet état, l'intégration à un blastocyste et la création de chimère, est impossible chez l'homme pour des raisons éthiques (*271*).

3) Reprogrammation

Dans le cadre des cellules souches pluripotentes, La reprogrammation est le fait pour une cellule somatique de « remonter la pente » de la différenciation et de réacquérir les caractéristiques de pluripotence. Historiquement, plusieurs stratégies ont été développées pour générer des cellules pluripotentes. Ainsi il est possible de reprogrammer une cellule en la fusionnant avec une cellule déjà pluripotente. Le résultat est une cellule hybride tétraploïde qui présente des caractéristiques de pluripotence mais avec une forte instabilité (*281*). Une autre méthode, développé par John Gurdon à la fin des années 1950, consiste à transférer le noyau d'une cellule somatique dans le cytoplasme d'une cellule pluripotente (un ovocyte) (*282*). Cela permet une reprogrammation du noyau au stade pluripotent (voir même totipotent) démontrant ainsi que le développement et la différenciation sont des processus liés à des changements épigénétiques réversibles mais que le génome d'une cellule conserve une potentialité de pluripotence.

Actuellement, l'approche la plus utilisée pour la génération de cellules pluripotentes repose sur les travaux de Yamanaka et Takahashi. En 2006, dans un article pionnier, ils démontrent que la surexpression de certains facteurs de transcription, sélectionnés car spécifiquement exprimés dans les cellules pluripotentes ou impliqués dans la biologie des cellules ES, permet la conversion de fibroblastes vers des cellules pluripotentes conduisant ainsi à la découverte des cellules souches pluripotentes induites (iPSC). Via une stratégie soustractive minutieuse, ils ont ensuite pu identifier la combinaison minimale de facteurs nécessaires pour la génération d'iPSC à partir de cellules somatiques murines. Ce cocktail minimal de reprogrammation est composé des 4 facteurs suivants : Oct4, Sox2, Klf4 et c-Myc (OSKM) (*283*). Cette même combinaison s'est avérée capable de générer des iPSC à partir de cellules humaines (*284*). Depuis ces articles fondateurs, d'autres combinaisons de facteurs de transcription et de miRNA ont été décrites comme permettant la reprogrammation (*285*). Toutes reposent sur le même principe de réactivation des circuits fondamentaux de la pluripotence. Le processus de reprogrammation consiste en un continuum avec de nombreux états intermédiaires et seule une faible proportion de cellules parvient au stade iPSC. On assiste à une diminution graduelle des facteurs associés aux lignages et une augmentation des facteurs de pluripotence. D'un point de vue mécanistique, les facteurs OCT4, SOX2 et KLF4 (OSK) vont réorganiser l'activité des enhancers. OSK agissent comme des facteurs de transcription pionniers et vont se fixer sur les enhancers associés à la pluripotence et induire l'ouverture de la chromatine. A l'inverse, ils sont capables de recruter les facteurs associés aux lignages à des sites spécifiques les empêchant ainsi de se lier sur leurs cibles canoniques. Cette action, combinée à celle du complexe répresseur PRC2 va venir éteindre les enhancers associés à l'état différencié. c-MYC jouerait un rôle différent en favorisant la reprogrammation via les divisions cellulaires (*286*, *287*). D'un point de vue de la pluripotence, les iPSC murines ressemblent aux mESC naïves alors que les iPSC humaines sont équivalentes aux hESC et mEpiSC.

Initialement utilisés, l'expression de OSKM par des rétrovirus conduit à une reprogrammation très inefficace (entre 0,01 et 0,1 %) et peut induire un développement tumoral selon le site d'insertion dans le génome. Des systèmes plus efficaces et plus sûrs ont depuis été développés comme des virus non intégratifs, des ARNm ou des anticorps activateurs en remplacement de certains facteurs (*285*, *288*). Enfin, un rapport très récent apporte une preuve de concept de

reprogrammation de cellules humaines en iPSC en utilisant uniquement des composés chimiques sans expression de facteurs exogènes dans la cellule, représentant ainsi une piste intéressante dans le cadre d'un développement à grande échelle. (*289*).

Une question liée à la reprogrammation est le degré de mutagenèse associé à ce processus. Bien que des études préliminaires rapportaient la survenue d'anomalies chromosomiques durant les premières étapes de culture, une étude plus récente basée sur le séquençage, montre que très peu de mutations surviennent *de novo* durant le processus de reprogrammation. En effet, la plupart des variants retrouvés dans les iPSC sont déjà présents dans la population d'origine qui, pour certains tissus, présente une hétérogénéité génétique (*290*).

Au-delà de la génétique, une mémoire épigénétique de la cellule d'origine pourrait être conservée par les iPSC et constituerait une source de variabilité importante selon l'origine des iPSC. Bien que certaines études montrent en effet que la différenciation peut être facilitée vers le lignage d'origine des iPSC, ces différences disparaissent après un certain nombre de passages en culture (*291*). Ainsi, les données actuelles suggèrent que la variabilité entre les iPSC est principalement due à des différences interindividuelles et que les différences liées au type cellulaire d'origine pour un même individu sont faibles (*292*, *293*).

4) iPSC en hématologie

a) Protocoles de différenciation hématopoïétique

L'utilisation des iPSC pour la modélisation et l'étude de l'hématopoïèse normale et pathologique nécessite des protocoles de différenciation fiables et robustes. Les iPSC sont théoriquement capables de se différencier vers tous les lignages et il n'existe donc aucune raison pour laquelle elles ne pourraient pas générer toutes les cellules hématopoïétiques y compris les cellules souches. De nombreux laboratoires se sont appuyés sur les connaissances acquises sur la régulation du développement hématopoïétique *in vivo* pour développer des protocoles de différenciation *in vitro* des iPSC. Il existe 3 approches pour induire la différenciation hématopoïétique (Figure 25) (*294*, *295*):

- La formation de corps embryoïdes (EB) dans lesquels les cellules forment des structures 3D qui permettent des interaction cellules-cellules, mimant l'organisation spatiale de l'embryon.

- La co-culture avec des cellules stromales murines, comme les cellules OP9, qui fournissent des contacts cellulaires ainsi des facteurs qui permettent la prolifération et la différenciation cellulaire.

- La culture monocouche en 2D sur une boîte avec un revêtement de type matrice extracellulaire qui, avec l'ajout de cytokines, permet de d'obtenir une différenciation sans cellules exogènes.

Plusieurs protocoles de différenciation ont ainsi été développés, que l'on peut classer en 2 catégories : les protocoles de différenciation directe d'une part et la programmation cellulaire d'autre part.

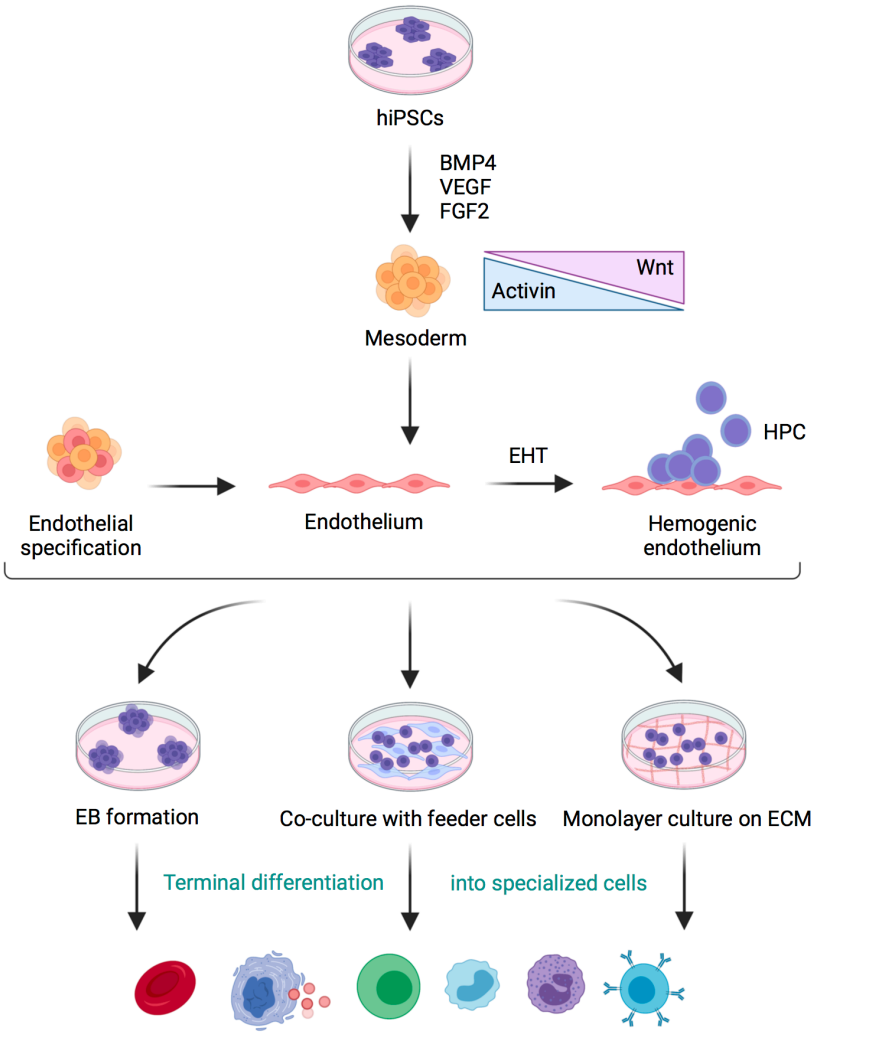


Figure 25 : Différenciation hématopoïétique des iPSC.
Représentation schématique des 3 techniques de différenciation hématopoïétique : les corps embryoïdes (EB), la co-culture sur cellules stromales et la culture en monocouches sur matrice extracellulaire. Le type d'hématopoïèse induit peut-être spécifier en modulant les voies Activin et Wnt au niveau du mésoderme. Issu de (*295*)

La différenciation directe repose sur 2 axiomes : l'activation de voies de signalisation de façon extrinsèque sans modification génétique est suffisante pour induire la différenciation complète des iPSC jusqu'aux cellules matures finales, et la différenciation *in vitro* récapitule correctement ce qui se produit durant le développement *in vivo*.

Les protocoles de différenciation peuvent se décomposer en 2 étapes. D'abord l'induction du mésoderme via des cytokines comme le BMP4, le VEGF ou le bFGF. Ensuite la spécification des cellules hématopoïétiques à partir de l'endothélium hémogénique et l'utilisation de cytokines permettant la différenciation hématopoïétique (*296*). Selon les protocoles, il existe des variations dans la combinatoire des cytokines ainsi que dans les concentrations utilisées.

Tous les types cellulaires hématopoïétiques matures ont ainsi pu être générés à partir d'iPSC. De nombreux protocoles existent, permettant la génération d'érythrocytes, de mégacaryocytes, de monocytes/macrophages, de granulocytes et de cellules dendritiques (*297*). La génération de cellules lymphoïdes s'avère plus complexe mais des protocoles robustes existent désormais pour la génération de lymphocytes T (*298*) et de cellules NK (*299*). Quelques articles rapportent également la possibilité d'obtenir des cellules B (*300*). Il convient de noter que les cellules ainsi générées présentent des caractéristiques de cellules hématopoïétiques embryonnaires plutôt que de cellules adultes. Par exemple, les érythrocytes ne présentent pas une énucléation complète et expriment les globines embryonnaires et fœtales, et les mégacaryocytes sont de petite taille, peu polyploïdes et transcriptionnellement proches des mégacaryocytes fœtaux (*297*, *301*).

Le consensus est que les différents protocoles permettent de récapituler une hématopoïèse de type primitive et/ou pré-définitive mais pas une hématopoïèse définitive. Ceci est renforcé par le fait que, bien que l'on puisse obtenir des progéniteurs ayant une potentialité rappelant les EMP ou les LMPP à partir d'iPSC, aucun protocole actuel ne permet l'obtention de HSC ayant une capacité de reconstitution multi-lignage à long-terme qui est le marqueur de l'hématopoïèse définitive. L'obtention de HSC à partir d'iPSC est cependant possible. En effet, 2 équipes sont parvenues, en parallèle, à isoler des HSC à partir de tératomes générés en injectant des iPSC indifférenciées dans des souris immunodéprimées (*302*, *303*). L'absence de HSC dans les différenciations *in vitro* est donc probablement due à l'utilisation de protocoles inadaptés.

La potentialité des cellules hématopoïétiques est spécifiée dès le stade endothélium hémogénique. Des travaux du laboratoire de Gordon Keller montrent qu'une stimulation du

mésoderme peut permettre de favoriser une hématopoïèse définitive au dépend de l'hématopoïèse primitive via une inhibition de la signalisation Activin/Nodal et une activation de la voie Wnt (Figure 25) (*304*).

La récapitulation exacte du processus de développement *in vivo* via les méthodes de différenciation directe est complexe. L'autre technique, la programmation cellulaire, propose d'utiliser une combinaison d'évènements génétiques et épigénétiques qui seraient suffisants pour forcer la différenciation des iPSC en cellules hématopoïétiques. Cette méthode (« forward programming » en anglais) est conceptuellement similaire à la reprogrammation de cellules somatiques en cellules pluripotentes mais en sens inverse. Cette approche présente l'avantage d'être plus directe et moins sujette à des variations liées aux protocoles de différenciation. Plusieurs facteurs de transcription connus comme ayant un rôle majeur dans la spécification et la différenciation hématopoïétique ont été décrits comme candidats au rôle de facteur programmateur. La surexpression d'un facteur seul n'est pas suffisante pour induire la différenciation hématopoïétique, qui repose sur une combinatoire (*305*).

Ainsi la surexpression de GATA2 et ETV2 dans des iPSC humaines induit la formation de progéniteurs $CD43^+$ avec une potentialité pan-myéloïde. En revanche la surexpression de GATA2 avec TAL1 conduit à des progéniteurs avec une potentialité restreinte aux lignages érythro-mégacaryocytaire (*306*). En 2017, Sugimura *et al* sont parvenus à générer des HSC à partir de cellules pluripotentes humaines. Pour cela ils ont initié une différenciation directe des iPSC en progéniteurs hématopoïétiques $CD34^+$ qu'ils ont ensuite transduit avec différents facteurs de transcription. Ainsi, ils montrent que la surexpression de la combinaison de facteurs ERG, HOXA5, HOXA9, HOXA10, LCOR, RUNX1, et SPI1 permet la génération de HSC avec un potentiel de reconstitution de la moelle osseuse à long terme (*307*). De façon intéressante, un article rapporte également la génération de HSC à partir d'iPSC suite à l'expression d'un unique facteur, la fusion MLL-AF4. Cependant bien que ce facteur permette la production de HSC capables de reconstituer la moelle de souris, ces dernières présentent un risque accru de développer des leucémies à long-terme dû au caractère oncogénique de MLL-AF4 (*308*).

La programmation cellulaire est donc prometteuse mais certains obstacles demeurent. Selon le vecteur employé, les facteurs de transcription peuvent avoir une expression en continue à l'inverse de leur action dynamique durant la différenciation normale. Cette expression peut

induire des problèmes d'autant que beaucoup de facteurs employés sont des oncogènes connus.

b) Modélisation des hémopathies malignes

Depuis leur découverte, les iPSC sont utilisées dans le cadre de l'étude de nombreuses pathologies hématologiques et particulièrement dans les hémopathies malignes car elles présentent de nombreux avantages.

Premièrement, il est possible d'étudier l'effet d'une/plusieurs mutations dans un contexte génétique humain, voir directement dans celui du patient pour les iPSC reprogrammées de patient. Cela est important car, comme mentionné précédemment, les modèles animaux sont parfois inadaptés. C'est par exemple le cas pour les syndromes plaquettaires familiaux, qui prédisposent aux LAM. Ainsi, alors que les mutations hétérozygotes de RUNX1 retrouvées chez les patients ne produisent pas de phénotype observable chez la souris, dans des modèles iPSC elles conduisent à un défaut de maturation des mégacaryocytes (*309*).
Deuxièmement, Les iPSC au stade pluripotent peuvent théoriquement pousser indéfiniment procurant une source illimitée de matériels cellulaires qui permet des expériences contrôlées et reproductibles. De plus, ces iPSC ne sont pas dépendantes d'oncogènes pour leur prolifération et ainsi ne présentent pas le risque de développer des altérations non physiologiques durant leur culture contrairement aux lignées cancéreuses « classiques ». En effet, des études combinant RNA-seq et analyse de la méthylation de l'ADN indiquent que la reprogrammation efface le programme cancéreux au niveau transcriptionnel et épigénétique. Ainsi, dans les iPSC, c'est l'état pluripotent qui prend le dessus et le programme leucémique ne sera ré-exprimé que lorsque les cellules seront différenciées vers le lignage hématopoïétique, et vers ce lignage uniquement (*310*).
Enfin, l'un des intérêts principaux des iPSC est la possibilité de générer des contrôles isogéniques, c'est-à-dire des contrôles ayant strictement le même fond génétique à l'exception de la mutation d'intérêt (Figure 26). Cela permet d'étudier spécifiquement l'effet d'une mutation sans avoir le « bruit de fond » lié à la variabilité interindividuelle.

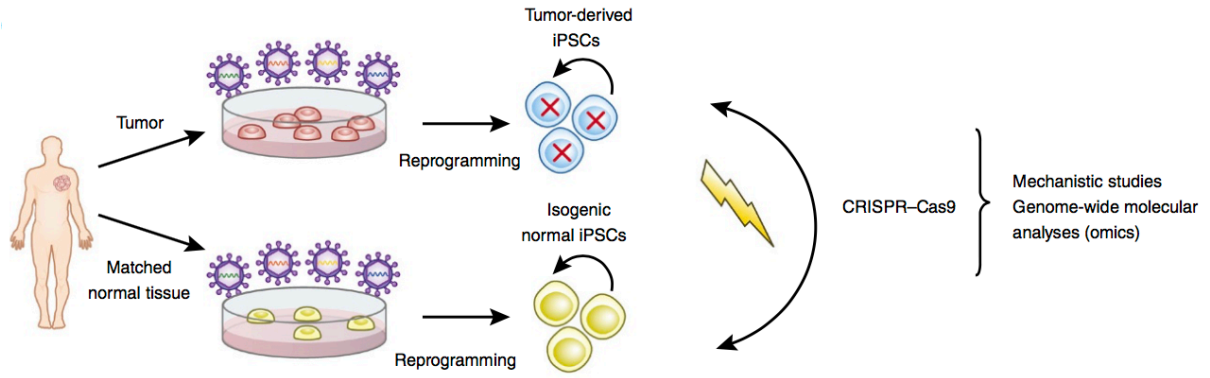


Figure 26 : Reprogrammation de cellules de patient au stade iPSC.
Représentation schématique du processus de reprogrammation d'iPSC à partir d'un prélèvement de patient. Un contrôle isogénique peut être généré en reprogrammant des cellules saines du même patient. Les cellules peuvent ensuite être utilisés pour différentes analyses mécanistiques. Issue de (*311*)

Les modèles iPSC utilisés pour étudier les hémopathies peuvent avoir deux origines : la reprogrammation de cellules de patients ou la modification génétique d'iPSC sauvages.

L'obtention d'iPSC à partir de cellules de patients atteints d'hémopathies se fait classiquement par reprogrammation de cellules $CD34^+$ à partir de prélèvement sanguin. Cela permet d'obtenir des cellules humaines avec l'exact contexte génétique du patient. De plus, il est ainsi possible d'établir des modèles pour des pathologies pour lesquelles la mutation driver est inconnue ou pour les caryotypes complexes, difficiles à reproduire autrement. C'est par exemple le cas pour la délétion d'une partie du chromosome 7 [del(7q)] fréquemment retrouvée dans les SMD. Les régions homologues au chromosome 7 humain sont présentes sur quatre chromosomes différents chez la souris. En conséquence, Kotini *et al*, ont établi un modèle de reprogrammation de cellules de patient en iPSC. Via des approches de criblage génétiques et fonctionnels ils ont ensuite pu établir la région chromosomique minimale puis les gènes candidats dont l'haplo-insuffisance confère le phénotype pathologique (*312*).
En 2017, deux groupes ont rapporté pour la première fois la génération d'iPSC à partir de cellules de LAM. Le groupe de Ravindra Majeti a ainsi montré qu'au stade pluripotent, les iPSC issues de LAM sont identiques aux iPSC contrôles mais que lorsqu'on les différencie vers le lignage hématopoïétique, elles réacquièrent un phénotype, un profil transcriptionnel et une sensibilité aux drogues similaires aux cellules de patient (*310*). Le groupe de Eirini Papapetrou a reprogrammé des iPSC à partir de patients à différents stades de la maladie obtenant ainsi des clones iPSC correspondant à l'état normal, pré-leucémique, SMD et LAM. Ce même groupe a également montré que lorsqu'on différencie ces iPSC elles présentent un phénotype dont la

gravité corrèle avec le stade de la pathologie dont elles sont issues. De plus en corrigeant/ajoutant des mutations par CRISPR il est possible de moduler la progression leucémique (*313*). Un des points majeurs de ces articles est qu'ils constituent les 2 seuls rapports à ce jour montrant le développement d'une leucémie dans des souris immunodéficientes à partir de cellules différenciées d'iPSC.

De façon plus globale, la reprogrammation des échantillons de patients offre la possibilité de reconstruire l'histoire mutationnelle de la pathologie. En effet, les techniques classiques pour étudier l'hétérogénéité clonale des leucémies reposent sur le séquençage du génome des patients pour inférer l'ordre d'acquisition des mutations sur la base des VAF. Ainsi, pour deux mutations ayant les mêmes VAF il est impossible de déterminer l'ordre, et certaines mutations avec une faible VAF peuvent ne pas être détectées. Les technologies de séquençage sur cellule unique sont une solution plus directe pour déterminer la composition clonale des leucémies. Cependant ces techniques ne permettent pas à la fois l'identification moléculaire et la réalisation de tests fonctionnels sur une même cellule rendant difficile la corrélation entre acquisition des mutations et phénotype. La reprogrammation en iPSC est une alternative intéressante puisque ce processus est clonal, permettant ainsi d'obtenir des clones qui peuvent présenter différents profils mutationnels et ainsi reconstruire l'histoire évolutive durant la progression leucémique (Figure 27). De plus, une fois établis, ces clones iPSC peuvent être différenciés et servir dans des tests fonctionnels afin de déterminer la contribution de chaque mutations (*314*).

Cependant un des problèmes majeurs de cette approche est que les cellules cancéreuses et les LAM en particulier sont réfractaires à la reprogrammation (*315*, *316*). Cela pourrait être dû au fait que les LAM sont souvent porteuses de mutations des facteurs épigénétiques qui peuvent interférer avec la reprogrammation. De plus l'incapacité pour de nombreuses LAM à pousser *ex vivo* impacte l'efficacité de reprogrammation car les divisions cellulaires sont un élément essentiel de ce processus. En conséquence, à ce jour un nombre très faible de lignées iPSC ont été générées à partir de LAM.

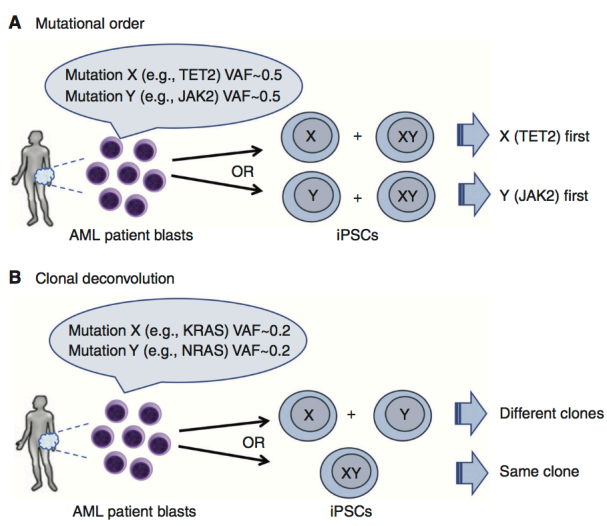


Figure 27 : Reconstruction de la progression clonale grâce à la reprogrammation.

La reprogrammation clonale peut permettre de reconstruire l'histoire évolutive de la leucémie dans des cas où le séquençage n'est pas informatif. Cela peut permettre de déterminer l'ordre d'acquisition de mutations ayant des fréquences d'allèle variant (VAF) similaires (**A**) ou de déconvoluer la composition clonale d'un échantillon comportant plusieurs mutations ayant des VAF faibles qui ne permettent pas de déterminer si elles sont présentes dans le même ou dans différents clones (**B**). Issue de (*317*)

Pour surmonter cet obstacle, une stratégie alternative consiste à utiliser des iPSC sauvages issues de donneurs sains et à les modifier génétiquement afin d'introduire la ou les mutations d'intérêt. En effet, différentes études montrent que, à l'instar des lignées cellulaires et contrairement aux échantillons primaires, les iPSC sont facilement modifiables par les techniques d'édition génétique comme CRISPR-Cas9. Ainsi, dans une autre étude du groupe d'Eirini Papapetrou, les auteurs ont modifié par CRISPR-Cas9 des iPSC normales pour introduire de façon séquentielle les mutations ASXL1, SRFS2 et NRAS (ou FLT3-L) afin d'obtenir des modèles isogéniques correspondant aux stades pré-leucémique, SMD et LAM. Via une analyse moléculaire des profils transcriptionnels et chromatiniens des cellules hématopoïétiques issues de ces différentes iPSC, ils ont ainsi caractérisé les changements associés à ces différentes étapes. Cela leur a permis d'identifier une liste de gènes activés dès l'état SMD et toujours exprimés au stade LAM. Alors que le traitement de référence, la cytarabine, ne fonctionne que sur les LAM, l'utilisation d'un inhibiteur de l'un de ces gènes (ATP6V0A2) est efficace à la fois contre les SMD et les LAM permettant ainsi potentiellement de traiter plus précocement les patients (*318*).

Dans le cadre des leucémies pédiatriques, les iPSC représentent un outil intéressant. Ainsi, de nombreux modèles iPSC de leucémie myélomonocytaire juvénile se sont développés. En effet, cette leucémie qui survient chez de très jeunes enfants s'est avérée difficile à modéliser du fait de l'absence de blocage de différentiation qui rend quasiment impossible la dérivation de lignées cellulaires (*319*). L'étude des LAL-B infantiles ETV6-RUNX1, dont l'origine *in utero* est

prouvée, constitue un autre exemple pertinent d'utilisation des iPSC. En effet, Böiers *et al* ont développé un modèle de cellules iPSC modifiées par CRISPR-Cas9 pour introduire la fusion ETV6-RUNX1 et, en utilisant un protocole de différenciation qui récapitule l'hématopoïèse pré-définitive, ils ont montré que la cible de la fusion est un progéniteur IL-7R^{+} bipotent, présentant un profil d'expression mixte lymphoïde et myéloïde, présent uniquement durant le développement. Lorsque la fusion ETV6-RUNX1 est exprimée, ces progéniteurs ne peuvent plus se différencier en cellules pro-B et s'accumulent (*225*). Enfin, les iPSC sont très utilisées pour étudier les DS-LAM7 car cette leucémie survient durant le développement fœtal et est associée à une T21, difficile à modéliser chez la souris. Dans ce cadre, Banno *et al*, ont généré plus de 40 clones iPSC différents afin de disséquer les contributions de la T21 et de la mutation GATA1s durant la phase TMD pré-leucémique (*230*).

Malgré les nombreux avantages à l'utilisation des iPSC, plusieurs limites demeurent. Comme mentionné précédemment, les leucémies s'avèrent particulièrement réfractaires à la reprogrammation. Bien que très récemment un rapport de reprogrammation d'une LAL-T, une LAM5 et une LAM7 a été publié (*320*), la diversité des lignées iPSC issues de leucémie de patient reste extrêmement faible et il est nécessaire d'améliorer les protocoles afin d'augmenter le spectre de reprogrammation. Une autre limitation est le type de protocoles de différenciation hématopoïétique utilisés. Actuellement, aucun ne permet d'obtenir tous les types cellulaires et notamment des HSC. Ainsi, il est probable que certain modèles d'iPSC ne permettent pas d'obtenir le type cellulaire et/ou le stade développemental nécessaire pour que la mutation exerce pleinement son activité oncogénique. Enfin, bien que plus physiologique que de lignées cellulaires cancéreuses, les iPSC restent un système *in vitro* et souffrent donc des limitations qui y sont liées, à savoir l'absence de microenvironnement et de contact avec les autres types cellulaires (bien que la différenciation en EB apporte en partie ces éléments).

Pour conclure, les iPSC représentent un modèle robuste pour l'étude des hémopathies chez l'homme mais nécessitent encore des améliorations. Les avancées récentes concernant le développement d'organoïdes, d'organe sur puce et de systèmes combinant cela avec la micro-fluidique sont des pistes prometteuses dans l'optique de pouvoir modéliser totalement *in vitro* les interactions d'un organisme entier (*321*).

Résultats

Contexte et objectifs

Les LAM7 *de novo* constituent un sous-type rare de leucémies pédiatriques qui surviennent chez de très jeunes enfants et sont souvent causées par des anomalies chromosomiques qui conduisent à l'expression d'oncogènes de fusions.

Suite à l'identification de la fusion ETO2-GLIS2, résultant de l'inversion du chromosome 16 (*252*, *253*), il a été montré qu'elle est la plus fréquente des anomalies génétiques retrouvées dans les LAM7 *de novo* et est associée au plus mauvais pronostic (*250*, *251*). Il est donc urgent de mieux comprendre le processus de développement oncogénique et les mécanismes moléculaires associés afin de développer des traitements adaptés.

Plusieurs travaux portant sur ETO2-GLIS2 ont été réalisés au laboratoire. Tout d'abord, en s'appuyant principalement sur des lignées cellulaires, notre équipe a étudié les changements transcriptionnels ainsi que la fixation à l'ADN de la fusion dans des cellules humaines (*263*).
Parmi les dérégulations transcriptionnelles associées, on trouve le facteur de transcription ERG qui est fortement surexprimé par la fusion et participe activement au mécanisme d'action de la fusion. En particulier, ERG partage de nombreux sites de fixation sur l'ADN communs avec la fusion ETO2-GLIS2. Cependant les lignées cellulaires ne représentent pas un système idéal pour l'étude de processus dynamiques comme la transformation oncogénique ou la différenciation (c.f. introduction).
Par la suite, le développement d'un modèle murin d'expression inductible de la fusion ETO2-GLIS2 a permis de montrer le rôle du contexte ontogénique (cellules fœtales vs cellules adultes) ainsi que de la hiérarchie hématopoïétique (HSC vs progéniteurs) dans la transformation leucémique (*264*). Ce modèle murin indique que les cellules souches hématopoïétiques fœtales sont beaucoup plus permissives au développement de leucémies agressives par la fusion ETO2-GLIS2 que les cellules adultes.

Ainsi au début de ma thèse, nous disposions soit de lignées cellulaires humaines, représentant un modèle limité de stade final de transformation adapté à la pousse des cellules *in vitro*, soit d'un modèle robuste de transformation leucémique in *viv*o mais uniquement chez la souris.

Un modèle reproduisant la leucémogenèse à partir de cellules humaines n'était pas disponible mais est indispensable pour permettre des comparaisons directes (ex : accessibilité de la chromatine et dérégulations transcriptionnelles) dans un contexte génomique identique à celui des cellules de patients.

Mon projet de thèse est ainsi constitué de 2 grandes parties. Premièrement, j'ai pour objectif de développer un modèle humain de transformation leucémique par la fusion ETO2-GLIS2. Pour cela, j'ai utilisé 2 stratégies différentes à partir de cellules souches pluripotentes induites humaines : un premier modèle d'expression transgénique de la fusion sous contrôle du promoteur de CD43, généré par recombinaison homologue via des nucléases à doigts de zinc et un second modèle d'inversion chromosomique via CRISPR-Cas9. Le deuxième objectif de mon projet de thèse consiste à utiliser ces modèles pour caractériser les mécanismes moléculaires de transformation leucémique par ETO2-GLIS2 chez l'homme. Pour cela, j'ai caractérisé les effets *in vitro* et *in vivo* induit par la fusion lors de la différenciation hématopoïétique de ces iPSC. J'ai ensuite procédé à des analyses de séquençage à haut débit afin d'étudier les impacts au niveau transcriptionnel et chromatinien.

Article 1 : The Pediatric Acute Leukemia Fusion Oncogene ETO2-GLIS2 Increases Self-Renewal and Alters Differentiation in a Human Induced Pluripotent Stem Cells-Derived Model

Dans ce travail, réalisé au début de ma thèse en collaboration avec un autre doctorant (Salvatore Nicola Bertuccio) et une ingénieure (Marie Cambot), nous avons développé un premier modèle d'expression de la fusion ETO2-GLIS2 dans des iPSC issues d'un donneur sain. Pour cela, nous avons introduit une construction permettant l'expression de ETO2-GLIS2 fusionnée à la GFP sous le contrôle du promoteur CD43 au locus AAVS1, présentant une chromatine constitutivement ouverte et ne codant aucune protéine (appelé « safe harbor » en anglais). Cela a été réalisé par recombinaison homologue via des nucléases à doigts de zinc. Le marqueur CD43 étant pan-hématopoïétique, l'utilisation de ce promoteur avait pour objectif une expression de la fusion dans toutes les cellules hématopoïétiques. A partir des clones iPSC contenant la fusion ETO2-GLIS2 sous le contrôle du promoteur CD43 (pCD43-EG),

j'ai réalisé des différenciations hématopoïétiques *in vitro* qui ont été caractérisées par immunophénotypage par cytométrie en flux et des tests de clonogénicité et d'auto-renouvellement en culture semi-solide de méthylcellulose. Ainsi, j'ai démontré que dans ces cellules humaines, la fusion induit un biais de la différenciation vers le lignage mégacaryocytaire et une augmentation de l'auto-renouvellement sans pour autant permettre le développement de leucémie lorsqu'on les greffe dans des souris immunodéficientes. En outre, une population particulière de cellules $CD41^{low}42^{low}$, qui n'existe pas dans la différenciation issue des clones iPSC contrôles, est détectée dans les cellules ETO2-GLIS2. J'ai ensuite réalisé des analyses de RNA-seq qui montrent que les cellules issues d'iPSC ETO2-GLIS2 présentent un profil transcriptomique enrichi en expression de gènes signature des cellules leucémiques de patients ETO2-GLIS2. De plus, la population $CD41^{low}42^{low}$ présente le profil transcriptionnel le plus proche des cellules de patients. Ce travail est détaillé dans l'article 1, publié dans le journal Hemasphere.

Article 2: ETO2-GLIS2 induces changes in transcription factor balance between early and late stages of leukemia in a human pluripotent stem cells model

Dans ce second travail, nous sommes partis de l'observation d'une absence de développement leucémique *in vivo* à partir du premier modèle et j'ai visé le développement d'un nouveau modèle, plus fidèle, qui permettrait d'étudier toutes les étapes de leucémogenèse. Je suis parti de l'hypothèse que le niveau et profil d'expression de la fusion contrôlé par le promoteur CD43 n'était pas approprié et j'ai donc recréé de façon fidèle l'altération génétique retrouvée chez les patients. Pour cela, j'ai modifié des iPSC afin d'induire l'inversion du chromosome 16 par l'approche CRISPR-Cas9. J'ai réalisé cette expérience sur des iPSC contrôles disomiques ainsi que sur des iPSC trisomiques pour le chromosome 21 car cette anomalie est retrouvée chez certains patients. De cette façon, j'ai pu obtenir et valider 3 clones présentant l'inversion du chromosome 16 et la fusion ETO2-GLIS2. J'ai ensuite différencié ces iPSC en cellules hématopoïétiques. Comme avec le précédant modèle, j'ai observé un fort biais mégacaryocytaire induit par la fusion ainsi que la présence de populations anormales retrouvées chez les patients. Cependant, une différence majeure avec ces cellules issues d'iPSC modifiées par CRISPR est qu'elles prennent en greffe dans des souris

immunodéficientes et conduisent à une leucémie agressive présentant un phénotype semblable à celui des patients. De façon inattendue, alors que les cellules ETO2-GLIS2 + T21, présentent une clonogénicité plus importante dans les cultures *in vitro*, elles conduisent à un développement leucémique de façon beaucoup moins efficace que les cellules disomiques.

Ayant obtenu un modèle permettant une transformation leucémique efficace dans des cellules humaines, j'ai ensuite réalisé des analyses d'expression et d'accessibilité de la chromatine afin de comprendre les mécanismes moléculaires impliqués à différents stades du processus de leucémogenèse. Dans ce but, j'ai étudié trois types de cellules issues des modèles iPSC : des cellules contrôles *in vitro*, des cellules ETO2-GLIS2 *in vitro* (représentant un stade précoce de la pathologie) et des cellules ETO2-GLIS2 issues de souris leucémiques (représentant un stade avancé de la pathologie). Des cellules issues de leucémies de patients ETO2-GLIS2 ont également été utilisées afin de permettre des comparaisons. Les analyses de scRNA-seq révèlent que les blastes leucémiques sont associés à différents lignages myéloïdes et s'accumulent particulièrement dans un intervalle de stades de différenciation assez large allant des progéniteurs immatures jusqu'aux cellules mégacaryocytaires. Par comparaison avec une signature transcriptomique des patients ETO2-GLIS2 issus d'une large cohorte de patients atteint de LAM pédiatriques, j'ai confirmé que les cellules issues de notre modèle ressemblent fortement aux cellules de patients. Ensuite, les analyses d'ATAC-seq ont mis en évidence qu'un fort réarrangement de la chromatine se produit dans les cellules ETO2-GLIS2 durant la transition *in vitro* vers *in vivo*. Spécifiquement, nous observons une forte augmentation de l'accessibilité des motifs des facteurs de transcription de la famille ETS et une perte de ceux de la famille GATA. De façon intéressante, j'ai également identifié une nouvelle famille de facteur de transcription à homéodomaine (famille DLX), dont les analyses de motifs indiquent une augmentation d'activité dès les étapes précoces *in vitro*. Les analyses fonctionnelles sont en cours pour définir le rôle de DLX3, l'un des facteurs de cette famille pour lequel les analyses préliminaires suggèrent une contribution à la transformation leucémique.

Résultats additionnels

J'ai également pris part à une étude menée au sein de notre équipe par un post-doctorant (Brahim Arkoun), publiée dans *Journal of Clinical Investigation,* dans laquelle je suis 3ème auteur. Cette étude s'attachait à disséquer les mécanismes cellulaires et moléculaires liés aux LAM7 associées au syndrome de down (DS-LAM7; cf. introduction). Le développement, par une approche de CRISPR-Cas9 d'une vingtaine de clones iPSC isogéniques présentant différentes combinaisons de mutations retrouvées dans les DS-LAM7 a permis de montrer que la mutation GATA1s conduit à une forte prolifération des progéniteurs mégacaryocytaires associée à des défauts majeurs dans la différenciation terminale avec des anomalies dans la maturation membranaire et la formation des proplaquettes. Ces défauts, accentués par une coopération avec la mutation additionnelle de la cohésine SMC3, s'expliquent au niveau moléculaire par une diminution de l'accessibilité et de l'expression des gènes impliqués dans ces processus et notamment du facteur de transcription NFE2 et à l'inverse une suractivation de la voie MYC.

Dans cette étude j'ai participé aux différenciations des iPSC en cellules hématopoïétiques et j'ai réalisé les expériences d'ATAC-seq et participé à l'analyse des données d'ATAC-seq et de RNA-seq.

Article 1

HemaSphere
Powered by EHA

EHA EUROPEAN HEMATOLOGY ASSOCIATION

Letter
OPEN ACCESS

The Pediatric Acute Leukemia Fusion Oncogene ETO2-GLIS2 Increases Self-Renewal and Alters Differentiation in a Human Induced Pluripotent Stem Cells-Derived Model

Salvatore Nicola Bertuccio[1,2], Fabien Boudia[2,3], Marie Cambot[2,4], Cécile K. Lopez[2,3], Larissa Lordier[5], Alessandro Donada[2], Elie Robert[2], Cécile Thirant[2], Zakia Aid[2], Salvatore Serravalle[1], Annalisa Astolfi[6], Valentina Indio[6], Franco Locatelli[7], Andrea Pession[1], William Vainchenker[2,4,5,8], Riccardo Masetti[1], Hana Raslova[2,8,9], Thomas Mercher[2,3,8,9]

Correspondence: Thomas Mercher (e-mail: thomas.mercher@inserm.fr).

Salvatore Nicola Bertuccio, Fabien Boudia and Marie Cambot are Co-first authors.
Andrea Pession, Riccardo Masetti, Hana Raslova, and Thomas Mercher are Co-senior authors.
This work was supported by the Institut National du Cancer (PLBIO-2018-169 to TM), PAIR-Pédiatrie/CONECT-AML (COllaborative Network for Children and Teenagers with Acute Myeloblastic Leukemia: INCa-ARC-LIGUE_11905 and Association Laurette Fugain), Fondation pour la Recherche Médicale (to ZA: FRM-ING20150532273, CKL), Cancéropôle Ile de France (to CKL and Emergence 2015), Associazione Italiana Ricerca sul Cancro to RM (MFAG2016, Id.19117, to Riccardo Masetti), Associazione Italiana Ricerca sul Cancro (to FL). HR, TM and WV are supported by INCA 2016-1-PLBIO-04-IGR-1. WV and MC are supported by the GRex (ANR, investissement d'Avenir). TM and HR are Equipe Labellisée LIGUE principal investigators.
The authors declare no conflicts of interest.
RNAseq data obtained in this study were deposited into EBI - Array-Express under the accession number E-MTAB-7967. Previously published datasets were also re-analyzed: E-MTAB-4367 and GEO: GSE4119.
Supplemental Digital Content is available for this article.
[1]Department of Pediatrics, "Lalla Seràgnoli", Hematology-Oncology Unit, Sant'Orsola-Malpighi Hospital, University of Bologna, Via Massarenti 11, 40137, Bologna, Italy.
[2]INSERM U1170, Gustave Roussy, Villejuif, France
[3]Université Paris Diderot, Paris, France
[4]Institut National de Transfusion Sanguine INTS, Paris, France
[5]Gustave Roussy, IPSC platform, Villejuif, France
[6]"Giorgio Prodi" Cancer Research Center, University of Bologna, Bologna and Department of Biomedical and Specialty Surgical Sciences, University of Ferrara, Ferrara, Italy
[7]Department of Pediatrics, Sapienza, University of Rome, Italy
[8]Université Paris-Sud, Orsay, France
[9]Equipe Labellisée Ligue Contre le Cancer, France.
Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
HemaSphere (2020) 4:1(e319)
Received: 5 August 2019 / Received in final form: 8 October 2019 / Accepted: 24 October 2019
Citation: Bertuccio SN, Boudia F, Cambot M, Lopez CK, Lordier L, Donada A, Robert E, Thirant C, Aid Z, Serravalle S, Astolfi A, Indio V, Locatelli F, Pession A, Vainchenker W, Masetti R, Raslova H, Mercher T. The pediatric acute leukemia fusion oncogene ETO2-GLIS2 increases self-renewal and alters differentiation in a human induced pluripotent stem cells-derived model. *HemaSphere*, 2020;4:1. http://dx.doi.org/10.1097/HS9.0000000000000319

The ETO2-GLIS2 fusion is associated with young and aggressive pediatric acute megakaryoblastic leukemia (AMKL) and is thought to arise *in utero*. To investigate ETO2-GLIS2 consequences in human hematopoietic cells and overcome the difficulty to access appropriate fetal stages, we engineered induced pluripotent stem cells (iPSC) to obtain hematopoietic-specific ETO2-GLIS2 expression and analyzed *in vitro* hematopoietic differentiation of IPSC toward the megakaryocytic lineage, which presents common features with human hematopoiesis during early development. As compared to controls, ETO2-GLIS2 expression induced an increased proportion of $CD41^{+}42^{+}$ megakaryocytes at several differentiation timepoints and generated a $CD41^{low}42^{low}$ population that is absent in controls. In addition, ETO2-GLIS2 enhanced proliferation and self-renewal capacities in methylcellulose assays. To understand the molecular consequences of ETO2-GLIS2 expression in human progenitors, we next performed RNA sequencing on flow purified $CD41^{+}42^{+}$ megakaryocytes and the abnormal $CD41^{low}42^{low}$ population. Compared to wild-type $CD41^{+}42^{+}$ profiles, ETO2-GLIS2-expressing $CD41^{+}42^{+}$ profiles were enriched for several ETO2-GLIS2 AMKL patients blasts signatures and an ETO2-GLIS2-dependent enhancer-associated genes signature. Importantly, these signatures were even more enriched in the ETO2-GLIS2-expressing $CD41^{low}42^{low}$ suggesting that this abnormal population more closely mimic the leukemic blasts found in patients. Of note, expression of *ERG* and *GATA1*, two master hematopoietic transcription factors, were not deregulated to the same extent as in patients' blasts. Together, this human ETO2-GLIS2 iPSC model recapitulates differentiation alterations, increased self-renewal and transcriptional signatures observed in human AMKL and should therefore represent an interesting platform to perform future molecular and preclinical investigations.

De novo pediatric acute megakaryoblastic leukemia (AMKL) is characterized by fusion oncogenes.[1,2] Murine transgenic models have been reported for some fusions, including OTT-MAL and MN1-FLI1.[3,4] The *CBFA2T3-GLIS2 (ETO2-GLIS2)* fusion, present in 20% to 30% of de novo pediatric AMKL, is associated

with a low number of additional genetic alterations and a dismal prognosis.[2,5–8] It induces a deregulation in the balance between hematopoietic master regulators, including a higher expression of ERG and GATA3 associated with a drastic reduction in GATA1 activity in human AMKL cells.[9] However, the consequences of ETO2-GLIS2 expression in normal human hematopoietic cells and its leukemogenic potential is unknown to date. In several instances, *de novo* AMKL patients are diagnosed few weeks/months after birth and in twins,[10] suggesting that the fusions were generated in a hematopoietic stem or progenitor cell during early fetal development. Based on the observation that normal megakaryocyte differentiation from human embryonic stem cells resemble early ontogenic stages,[11] we reasoned that hematopoietic differentiation derived from human induced pluripotent stem cells (iPSC) represents a developmentally-relevant context to model pediatric-specific AMKL fusion oncogenes.

To this aim, we used normal iPSC derived from healthy hematopoietic progenitors[12] to perform a targeted knock-in at the AAVS1 locus, through zinc-finger nucleases-mediated recombination, to introduce a GFP-tagged *ETO2-GLIS2* fusion gene under the transcriptional control of the pan-hematopoietic human CD43 promoter (Fig. 1A). Importantly, this allows to assess ETO2-GLIS2 function independently of iPSC line-to-line variances. We identified 23 clones with homozygous integration (Supplemental Fig. 1A and supplemental Table 1, Supplemental Digital Content 1 and 2, http://links.lww.com/HS/A54, http://links.lww.com/HS/A55). Two clones were selected for further studies, validated for a stable karyotype (Supplemental Fig. 1B, Supplemental Digital Content 1, http://links.lww.com/HS/A54) and pluripotency (Supplemental Fig. 1C,D, Supplemental Digital Content 1, http://links.lww.com/HS/A54) and were thereafter compared to the parental control iPSC line. Functional expression of ETO2-GLIS2 upon hematopoietic differentiation of iPSC was validated through quantitative RT-PCR of *ETO2-GLIS2* (Fig. 1B), through ETO2-GLIS2 protein detection (assessed by GFP in the nucleus) and through higher expression of the known fusion target gene *GATA3* in hematopoietic progenitors at day 15 of differentiation, as compared to control iPSC-derived cells (Supplemental Fig. 1E,F, Supplemental Digital Content 1, http://links.lww.com/HS/A54).

Using an updated version of the hematopoietic differentiation protocol of iPSC described by Chou et al[12,13] (Supplemental Table 2, Supplemental Digital Content 2, http://links.lww.com/HS/A55), we assessed the consequences of ETO2-GLIS2 expression on megakaryocytic differentiation. From control iPSC clones, hematopoietic differentiation is observed by the detection of CD43, CD34, CD41, and CD61 followed by the acquisition of CD42 surface markers (Fig. 1C, 1D and Supplemental Fig. 2A, Supplemental Digital Content 1, http://links.lww.com/HS/A54). $CD41^+CD61^+CD42^-$ represent progenitors while $CD41^+CD61^+CD42^+$ represent maturing megakaryocytes. From *ETO2-GLIS2* iPSC clones, a higher fraction of megakaryocytic cells was reproducibly obtained with an increased percentage of $CD41^+CD42^+$ megakaryocytes as early as day13 and sustained at day 15 and day 18 of differentiation (Fig. 1C and E). Of note, *ETO2-GLIS2*$^+$ iPSC clones maintained a significant production of megakaryocytes over a longer period of time up to day 26, while the number of megakaryocytes generated from control iPSC clones declined after day 22 (data not shown). The *ETO2-GLIS2* iPSC clones also showed an aberrant population characterized by low expression of CD41 and CD42 (named $CD41^{low}CD42^{low}$) (Fig. 1C–E). Similarly to CD41, CD61 surface expression was significantly lower in both the *ETO2-GLIS2*-expressing $CD41^+CD42^+$ and, to a higher extent, the $CD41^{low}CD42^{low}$ populations, as compared to control cells (Fig. 1D and Supplemental Fig. 2B,C, Supplemental Digital Content 1, http://links.lww.com/HS/A54). These results indicated that *ETO2-GLIS2* expression in human iPSC-derived hematopoietic cells induced differentiation alterations.

We then assessed the proliferation and self-renewal properties of *ETO2-GLIS2*-expressing progenitors. Firstly, we sorted 50,000 $CD43^+$ hematopoietic cells from day 15 of differentiation into liquid culture. The output number of cells was counted after 4 days of culture and showed an 8-fold increase in the number of *ETO2*-GLIS2-expressing cells as compared to control cells (Fig. 1F and G). At that stage, morphological studies indicated clear proplatelets formation from controls megakaryocytes while *ETO2-GLIS2*-expressing megakaryocytes were larger and presenting with reduced signs of proplatelets formation (Fig. 1H). Secondly, day15 $CD43^+$ hematopoietic cells were sorted into semi-solid medium to perform colony-forming assays scored after 7 to 10 days of culture and weekly serially-replated thereafter (Fig. 1I). While control cells formed colonies only in the first plates, *ETO2-GLIS2*-expressing cells formed larger colonies in the first plates (Fig. 1J) and could consistently form colonies up to the third replating, but not further (Fig. 1I). Importantly, flow cytometry analyses of *ETO2-GLIS2*-expressing cells after replating revealed an important proportion of GPA^+ and $CD43^-$ cells (Supplemental Figure 2D, Supplemental Digital Content 1, http://links.lww.com/HS/A54), supporting the idea of an erythroid differentiation. Of note, injection of day 15 or day 18 $CD43^+$ cells from either control or *ETO2*-GLIS2 clones into irradiated immunodeficient NSG mice did not result in significant engraftment with a follow-up of over 9 months (data not shown) indicating that iPSC-derived progenitors obtained using this differentiation protocol are not maintained in vivo in these conditions. Together, these results showed that *ETO2-GLIS2* expression enhances proliferation and self-renewal of iPSC-derived human hematopoietic progenitors in vitro but no long-term in vivo engraftment could be detected in conditioned recipients.

To characterize the molecular consequences of ETO2-GLIS2 expression in iPSC-derived cells, we performed a transcriptome analysis on control and the two ETO2-GLIS2 clones at day18 of differentiation. From the control, the $CD41^+CD42^-$ and $CD41^+CD42^+$ cell populations were sorted to obtain normal immature progenitors and maturing megakaryocytes. From *ETO2-GLIS2*-expressing cells, $CD41^+CD42^+$ and the aberrant $CD41^{low}CD42^{low}$ were analyzed (Fig. 2A). As shown by tSNE and PCA plots, replicate conditions and cells from the two clones clustered together and apart from control cells (Fig.2B, Supplemental Fig. 3A, Supplemental Digital Content 1, http://links.lww.com/HS/A54). Gene set enrichment analyses revealed that ETO2-GLIS2$^+$ AMKL patient-derived gene expression signatures[7,14] and ETO2-GLIS2-dependent super enhancer-associated gene signatures[9] were enriched in ETO2-GLIS2-expressing cells (Fig. 2C, Supplemental Fig. 3B, Supplemental Digital Content 1, http://links.lww.com/HS/A54). Notably, ETO2-GLIS2-expressing $CD41^{low}CD42^{low}$ cells showed a significantly higher enrichment of these signatures as compared to ETO2-GLIS2-expressing $CD41^+CD42^+$ (Fig. 2C and D and Supplemental Table 3, Supplemental Digital Content 2, http://links.lww.com/HS/A55) and a significant loss of mature megakaryocyte features assessed by the xCell digital cellular deconvolution method[15] or by gene ontology associated with a higher expression of *ETO2-GLIS2* (Supplemental Fig. 3C,D,E,

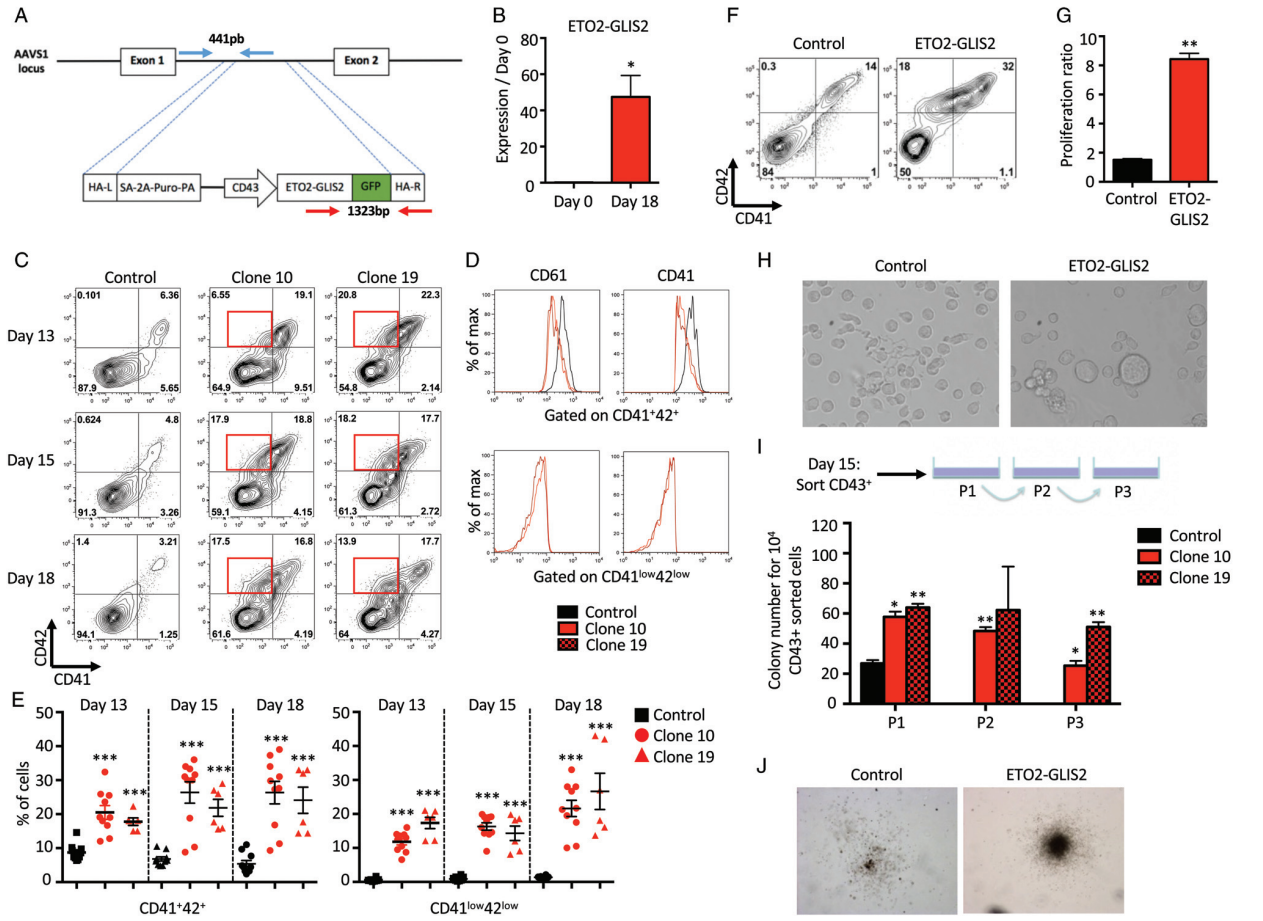


Figure 1. ETO2-GLIS2 induces self-renewal and differentiation alterations in an iPSC model. A. Schematic representation of the gene editing strategy. The constitutively active AAVS1 "safe harbor" locus is represented above the targeting construct. The cDNA expression cassette driving expression of *ETO2-GLIS2* cDNA tagged with GFP under the CD43 promoter was inserted by zinc finger-mediated homologous recombination into the AAVS1 intron. HA, homologous arms left (L) and right (R); SA-2APuro-PA, puromycin drug resistance cassette. The position of oligonucleotides used for the genotyping of iPSC clones by PCR are represented by arrows and the expected sizes of the PCR product for the wild-type allele (blue) and the ETO2-GLIS2 recombined (red) locus are indicated. B. Quantitative RT-PCR analysis of *ETO2-GLIS2* expression in iPSC and iPSC-derived day 18 hematopoietic cells from clone 10 relative to *HPRT*. Histograms represent means+/-SD (n = 3 obtained from one differentiation experiment). C. Flow cytometry analysis of CD41 and CD42 megakaryocytic markers in control and clones expressing *ETO2-GLIS2* fusion gene (clone 10 & 19) at day 13, 15 and 18 of differentiation. Red squares highlight the $CD41^{low}42^{low}$ population. D. Intensity of the CD41 and CD61 markers gating on the $CD41^{+}42^{+}$ populations from control, clone 10 & 19 at day 18 of differentiation (top panel) or on the $CD41^{low}42^{low}$ populations from clone 10 & 19 at day 18 of differentiation (bottom panel). E. Percentage of the different populations showed in C. Histograms represent means +/-SD of independent differentiation experiments (Control: n = 10, Clone 10: n = 10, Clone 19: n = 6). F. Flow cytometry analysis of CD41 and CD42 on $CD43^{+}$ hematopoietic cells derived from iPSC at day 15 and further cultured 4 days in liquid culture. G. Proliferation ratio of $CD43^{+}$ hematopoietic cells after 4 days in liquid culture supplemented with Flt3L 10ng/ul, G-CSF 20ng/ul, IL3 10ng/ul, IL6 10ng/ul, SCF 25ng/ul, TPO 10ng/ul and GM-CSF 10ng/ul. Histograms represent means +/-SD (n = 2 obtained from 2 independent differentiation experiments). H. Morphology of $CD43^{+}$ hematopoietic cells after 4 days of liquid culture. I. Upper panel: schematic representation of the experimental design used to test the self-renewal capacity of 10^4 $CD43^{+}$ sorted cells at day 15 in semi-solid medium. Lower panel: number of colonies as means ± SD (n = 2 obtained from 2 independent differentiation experiments). Cells were replated every 7 days for 4 weeks. P1: first plates; P2: 2^{ary} plates; P3: 3^{ary} plates. J. Representative colonies for control and ETO2-GLIS2 at P1. Statistical significance is indicated as *p*-value (paired Student's *t* test): $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Supplemental Digital Content 1, http://links.lww.com/HS/A54). These data indicate that ETO2-GLIS2-expressing $CD41^{low}CD42^{low}$ cells present prominent transcriptional alterations that closely recapitulate AMKL patient leukemic cells signature. Expression of ETS and GATA factors are known to control proliferation of megakaryocytes obtained from iPSC[16] and high ERG and low GATA1 activities are important for ETO2-GLIS2-transformed human blasts.[9] Therefore, we quantified *ERG* and *GATA1* expression in iPSC-derived hematopoietic cells at different timepoints (day 15, 18, and 22), in AMKL patients or AMKL patients-derived cell lines (eg MO7e) and in a previously described model of ETO2-GLIS2 ectopic expression in HEL cells.[9] *ERG* presented a low basal expression without significant change in $ETO2\text{-}GLIS2^{+}$ as compared to control cells (Supplemental Fig. 4A, Supplemental Digital Content 1, http://links.lww.com/HS/A54) and its expression was significantly lower in hematopoietic cells collected from day 22 $ETO2\text{-}GLIS2^{+}$ iPSC-derived cells as compared to ETO2-GLIS2-expressing AMKL cells or HEL cells (Fig. 2E). Although GATA1 was moderately repressed by ETO2-GLIS2 at day 15, its expression surprisingly increased in $ETO2\text{-}GLIS2^{+}$ cells at day 18 (Supplemental Fig. 4A, Supplemental Digital Content 1, http://links.lww.com/HS/A54) and was significantly higher in day 22 $ETO2\text{-}GLIS2^{+}$ iPSC-derived cells compared to ETO2-GLIS2-

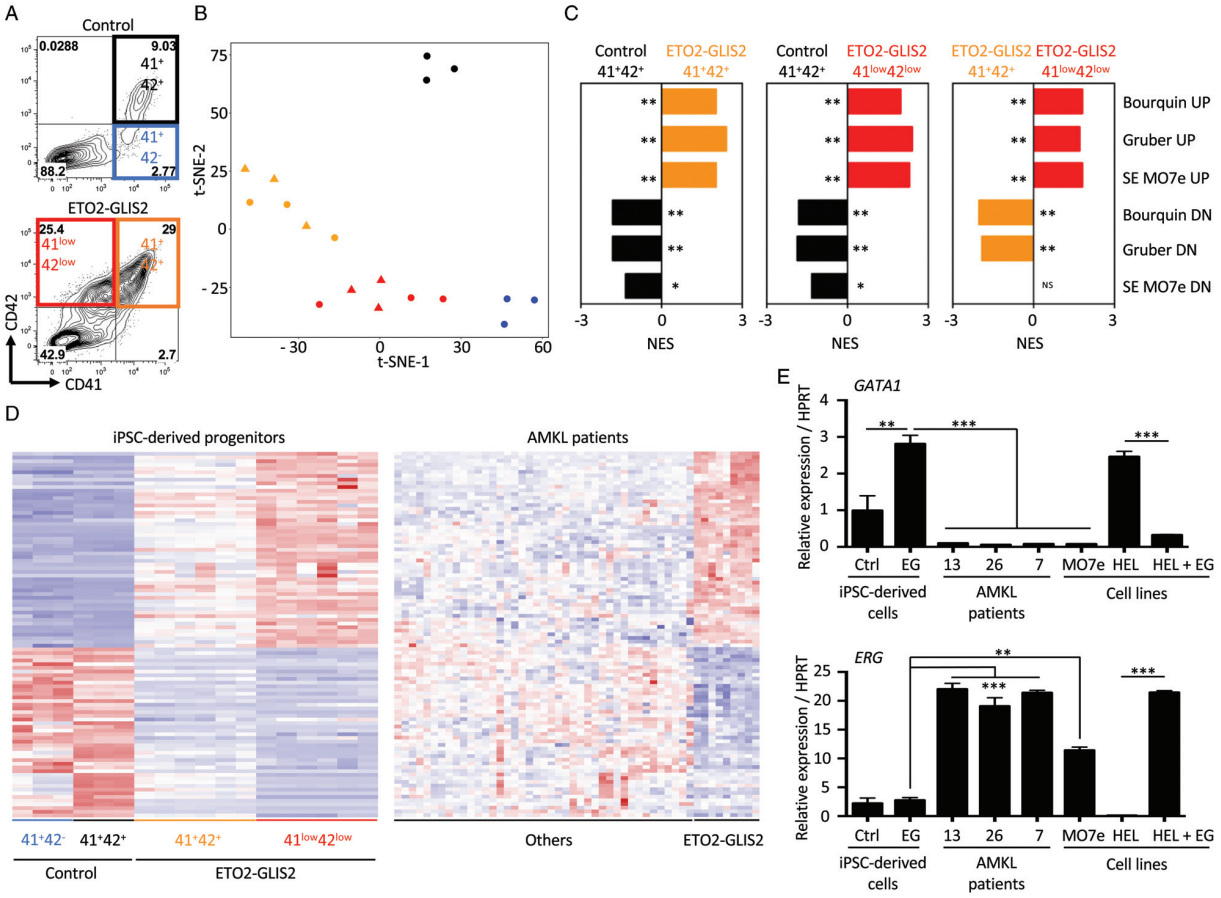


Figure 2. ETO2-GLIS2 expression in an iPSC model recapitulates molecular features of human AMKL. A. Flow cytometry plots highlighting the different sorted populations used for RNA sequencing from control and ETO2-GLIS2 (#10 and 19) clones. The color code used throughout the figure is indicated. B. t-SNE clustering of the different populations based on RNAseq data. Blue circles: CTRL $CD41^{+}CD42^{-}$, Black circles: CTRL $CD41^{+}CD42^{+}$, Orange: ETO2-GLIS2 $CD41^{+}42^{+}$ (circles: clone #10, triangles: clone #19), Red: ETO2-GLIS2 $CD41^{low}42^{low}$ (circles: clone #10, triangles: clone #19). C. Gene Set Enrichment Analysis comparing indicated paired samples for enrichment in different ETO2-GLIS2-related signatures including signature specific for ETO2-GLIS2^{+} AMKL leukemic cells as compared to other AMKL[7,14] and ETO2-GLIS2-dependent super enhancer-associated gene signatures (SE). UP: genes upregulated in ETO2-GLIS2^{+} AMKL, DOWN: genes downregulated in ETO2-GLIS2^{+} AMKL. *FDR < 0.25, ** FDR < 0.05. D. Heatmaps of the genes commonly deregulated by the ETO2-GLIS2 fusion in iPSC-derived cells (left) and in AMKL patients leukemic cells (right).[14] The threshold used to establish the signature was set to fold-change > 2. E. Quantitative RT-PCR quantification of *GATA1* and *ERG* expression in iPSC-derived cells collected at day 22 (Ctrl and EG), in 3 ETO2-GLIS2^{+} AMKL patient leukemic cells (AMKL #13, 26 & 7), in the ETO2-GLIS2^{+} AMKL-patient derived MO7e cell line and in HEL cells (no ETO2-GLIS2 expression) and HEL cells transduced with ETO2-GLIS2 expressing vector.[9] Histograms represent means+/-SD (n = 3 from independent differentiation experiments for IPSC-derived cells and from technical replicates for AMKL patient samples and cell lines). Ctrl: control; EG: ETO2-GLIS2. Statistical significance is indicated as *p*-value (Student's *t* test): $^{***}p < 0.001$ $^{**}p < 0.01$.

expressing AMKL cells or HEL cells (Fig. 2E). We then sorted $CD41^{low}CD42^{low}$ cells from day 22 ETO2-GLIS2^{+} and showed that they present a higher *ERG* and lower *GATA1* expression compared to ETO2-GLIS2^{+} $CD41^{+}CD42^{+}$. However, none of these two populations reached the expression levels observed in the ETO2-GLIS2^{+} MO7e cell line (Supplemental Fig. 4B, Supplemental Digital Content 1, http://links.lww.com/HS/A54). These results demonstrated that ETO2-GLIS2-expressing human iPSC-derived hematopoietic progenitors obtained using this differentiation protocol globally recapitulate the transcriptional program observed in ETO2-GLIS2 AMKL patient leukemic blasts, but lack sustained deregulation of at least the 2 key transcription factors *ERG* and *GATA1*.

Here, we report the first model of ETO2-GLIS2 expression in human hematopoietic progenitors obtained from the precise gene editing of healthy iPSC. Our data demonstrate that ETO2-GLIS2 impairs normal megakaryocytic differentiation, increases self-renewal and globally recapitulates the transcriptional program observed in human AMKL leukemic blasts. Importantly, we identified in ETO2-GLIS2^{+} iPSC-derived hematopoietic cells relevant molecular differences associated with their differences in self-renewal and long-term growth properties as compared to human AMKL patient leukemic blasts. Indeed, *ERG* expression is not strongly induced in ETO2-GLIS2^{+} iPSC-derived progenitors and remains at a low level compared to its expression in ETO2-GLIS2 AMKL patient leukemic cells. In addition, the lack of sustained repression of *GATA1* is consistent with the capacity of these cells to undergo erythroid differentiation upon serial replating and suggest that ETO2-GLIS2 does not fully block erythroid differentiation in these conditions. Therefore, we propose that the transient increase in self-renewal and altered differentiation in this human iPSC model of ETO2-GLIS2 expression results, in part, from an insufficient deregulation of the ERG/GATA1 activities. Of note, these characteristics could

be independent of ETO2-GLIS2 and may result from the type of hematopoiesis obtained with in vitro differentiation of iPSC to date. Indeed, it is generally accepted that current in vitro differentiation protocols do not reach an adult definitive hematopoiesis and it is still debated whether they recapitulate a primitive or transient definitive (EMP-like) hematopoiesis.[17] Whether the strong expression of CD43 in the primitive streak[18] may explain the results obtained here and would warrant the use of a different strategy to control ETO2-GLIS2 expression or whether a different differentiation protocol will be required to obtained definite hematopoiesis and full-blown transformation should be interesting areas of investigation. Together, this modeling approach in iPSC represents an important tool to understand the cellular and molecular bases for the transition between normal to malignant hematopoiesis, to develop more faithful models of childhood leukemia using human cells and to perform preclinical drug testing for the development of novel therapeutic strategies.

Acknowledgments

We are grateful to Mitchell Weiss for providing the AAVS targeting vectors, to Olivier Bernard and Eric Solary for facilitating this collaborative effort, to Esteve Noguera and Daniel Gautheret for bioinformatics support and Paule Zanardo for excellent administrative assistance.

References

1. Mercher T, Coniat MB, Monni R, et al. Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proc Natl Acad Sci U S A.* 2001; 98:5776–5779.
2. de Rooij JDE, Branstetter C, Ma J, et al. Pediatric non-Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nat Genet.* 2017;49:451–456.
3. Mercher T, Raffel GD, Moore SA, et al. The OTT-MAL fusion oncogene activates RBPJ-mediated transcription and induces acute megakaryoblastic leukemia in a knockin mouse model. *J Clin Invest.* 2009; 119:852–864.
4. Wenge DV, Felipe-Fumero E, Angenendt L, et al. MN1-Fli1 oncofusion transforms murine hematopoietic progenitor cells into acute megakaryoblastic leukemia cells. *Oncogenesis.* 2015;4:e179.
5. Thiollier C, Lopez CK, Gerby B, et al. Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J Exp Med.* 2012;209:2017–2031.
6. Masetti R, Pigazzi M, Togni M, et al. CBFA2T3-GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype. *Blood.* 2013;121:3469–3472.
7. Gruber TA, Larson Gedman A, Zhang J, et al. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell.* 2012;22:683–697.
8. Lopez CK, Malinge S, Gaudry M, et al. Pediatric acute megakaryoblastic leukemia: multitasking fusion proteins and oncogenic cooperations. *Trends Cancer.* 2017;3:631–642.
9. Thirant C, Ignacimouttou C, Lopez CK, et al. ETO2-GLIS2 hijacks transcriptional complexes to drive cellular identity and self-renewal in pediatric acute megakaryoblastic leukemia. *Cancer Cell.* 2017;31:452–465.
10. Ng KC, Tan AM, Chong YY, et al. Congenital acute megakaryoblastic leukemia (M7) with chromosomal t(1;22)(p13;q13) translocation in a set of identical twins. *J Pediatr Hematol Oncol.* 1999;21:428–430.
11. Bluteau O, Langlois T, Rivera-Munoz P, et al. Developmental changes in human megakaryopoiesis. *J Thromb Haemost JTH.* 2013;11:1730–1741.
12. Donada A, Balayn N, Sliwa D, et al. Disrupted filamin A/(IIb(3 interaction induces macrothrombocytopenia by increasing RhoA activity. *Blood.* 2019;133:1778–1788.
13. Chou ST, Byrska-Bishop M, Tober JM, et al. Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc Natl Acad Sci U S A.* 2012;109:17573–17578.
14. Bourquin J-P, Subramanian A, Langebrake C, et al. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proc Natl Acad Sci U S A.* 2006;103:3339–3344.
15. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18:220.
16. Moreau T, Evans AL, Vasquez L, et al. Large-scale production of megakaryocytes from human pluripotent stem cells by chemically defined forward programming. *Nat Commun.* 2016;7:11208.
17. Lacaud G, Kouskoff V. Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis. *Exp Hematol.* 2017;49: 19–24.
18. Sturgeon CM, Ditadi A, Awong G, et al. Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat Biotechnol.* 2014;32:554–561.

SUPPLEMENTAL INFORMATION

-Supplemental Figure & Table legends

-Methods

SUPPLEMENTAL FIGURES & TABLE LEGENDS

Supplemental Figure 1

A- PCR screening for the endogenous or ETO2-GLIS2-targeted AAVS1 locus. The expected sizes for the PCR products obtained from amplification of the wild-type allele and the ETO2-GLIS2 recombined allele (represented Figure 1A) are indicated.

B- Karyotype for control and #10 iPSC clones.

C- Quantitative RT-PCR on RNA from embryoid body (EB) formed after 24 days of differentiation. Expression is represented as a ratio between day 24 EBs and undifferentiated iPSC cells. Genes characteristic of the undifferentiated state (*NANOG*, *OCT4*) and of the three embryonic layers, mesoderm (*GATA4*), endoderm (*HAND1*) and ectoderm (*NeuroD1*) were used.

D- Histological section of teratoma formed after injection of iPSC into NSG mice.

E- Confocal fluorescence microscopy images obtained at day 15 of differentiation (left panel: control, right panel: ETO2-GLIS2).

F- Quantitative RT-PCR at day 15 of differentiation for *GATA3* expression. Histograms represent means+/-SD (n=3 obtained from one differentiation experiment). Statistical significance is indicated as p values (Student's t-test): **p-value<0.01

Supplemental Figure 2

A- Flow cytometry analysis of CD43 and CD34 hematopoietic markers in control and clones expressing *ETO2-GLIS2* fusion gene (clone 10 & 19) at day 13, 15 and 18 of differentiation.

B- Mean fluorescent intensity of CD41 and CD61 surface markers in the indicated populations. Mean+/-SD (n=3 independent differentiation experiments) are represented. Statistical significance is indicated as *p*-value (Student's t-test): ***p<0.001 **p<0.01.

C- Intensity of the CD42 marker gating on the $CD41^{+}42^{+}$ populations from control, clone 10 & 19 at day 18 of differentiation (up-left panel) or on the $CD41^{low}42^{low}$ populations from clone 10 & 19 at day 18 of differentiation (up-right panel). Mean fluorescent intensity in the indicated populations. Mean+/-SD (n=3 independent differentiation experiments) are represented. Statistical significance is indicated as *p*-value (Student's t-test): ***p<0.001 **p<0.01.

D- Flow cytometry analysis of CD43 and GPA markers for control and ETO2-GLIS2 cells at different stages of methylcellulose culture serial replating. P1: first plates; P2: 2^{ary} plates; P3: 3^{ary} plates.

Supplemental Figure 3

A- Principal component analysis of the samples used for RNAseq. Color code is the same than in Figure 2. Blue: CTRL $CD41^{+}CD42^{-}$, Black: CTRL $CD41^{+}CD42^{+}$, Orange: ETO2-GLIS2 $CD41^{+}42^{+}$, Red: ETO2-GLIS2 $CD41^{low}42^{low}$.

B- Gene Set Enrichment Analysis plots comparing ETO2-GLIS2 $CD41^{low}42^{low}$ versus ETO2-GLIS2 $CD41^{+}42^{+}$ comparison with the indicated lists at the top of each panel. False discovery rate (FDR) and Normalized Enrichment Score (NES) are indicated.

C- ETO2-GLIS2 expression in the different populations for control ($CD41^{+}42^{-}$ & $CD41^{+}42^{+}$) and ETO2-GLIS2 clones ($CD41^{+}42^{+}$ & $CD41^{low}42^{low}$) from RNAseq. Histograms represent means+/-SD (n=3 independent differentiation experiments).

D- Top 10 Gene Ontology (GO) biological process categories enriched in the up- or down-regulated signatures of the ETO2-GLIS2 $CD41^{low}42^{low}$ population as compared to control and ETO2-GLIS2 $CD41^{+}42^{+}$ cells. The scale represents the combined score calculated based on the z-score enrichment and the p-value as indicated by the Enrichr online software.

E- Digital cellular heterogeneity portrait of each transcriptome using the xCell method (14).

Supplemental Figure 4

A- Quantitative RT-PCR expression of *ERG* and *GATA1* in hematopoietic cells obtained at day 15, 18 and 22 from control and ETO2-GLIS2 #10 iPSC clone normalized to *HPRT*. Histograms represent means+/-SD (n=3 obtained from one differentiation experiment).

B- Quantitative RT-PCR expression of *ERG* and *GATA1* normalized to *HPRT* in purified hematopoietic cells obtained at day 22 from control iPSC ($CD41^{+}42^{+}$ population) and ETO2-GLIS2 iPSC clones 10 and 19 ($CD41^{+}42^{+}$ and $CD41^{low}42^{low}$ populations) as compared to HEL (no ETO2-GLIS2 expression) and MO7e (ETO2-GLIS2+ patient-derived) cell lines. Histograms represent means+/-SD (duplicate from two independent experiments).

Statistical significance is indicated as *p*-value (Student's t-test): ***p<0.001 **p<0.01. *p<0.05.

Supplemental Table 1

List of primers used in this study.

Supplemental Table 2

Growth factors used for the hematopoietic differentiation of iPSC.

Supplemental Table 3

Table of counts (RPKM) for the genes deregulated in common in both iPSC and AMKL leukemic blasts.

METHODS

Cloning of the *ETO2-GLIS2* targeting construct

ETO2-GLIS2-GFP was cloned under the control of the human CD43 promoter into the AAVS1-SA-2A-puro-pA a kind gift of M. Weiss (St. Jude Children's Research Hospital, Memphis, TN, USA) through in-fusion strategy following manufacturer's instructions (In-Fusion® HD Cloning Kit, Takara Bio, USA) and transformation into XL10-Gold Ultracompetent cells (Agilent). The other two vectors encoding the two Zing Finger proteins specific for the AAVS1 locus (left and right arms) were also obtained from M. Weiss.

iPSC culture and editing

iPSC were derived from healthy donor CD34+ cells as described previously (11). iPSC were cultured in Essential 8 medium (Thermofisher) and in vitronectin (Thermofisher) coated wells. iPSC were passed every 3-4 days in DPBS/0.5mM EDTA Ultrapure (Thermofisher). A Mycoplasma screening was routinely performed, according to the manufacturer instructions (Sigma, Saint-Quentin Fallavier, France).

To obtain iPSC presenting pCD43-*ETO2-GLIS2-GFP* integration at the AAVS1 locus, wild-type iPSC were treated with Triple 1x (Thermofisher) to make single cells and $0.8x10^6$ single cells were electroporated with the *ETO2-GLIS2* donor construct in association with the vectors encoding for Zing Finger, using Nucleofector 2B and Stem cell kit (Lonza). Electoroporated iPSC were plated on fresh vitronectin-coated plates in presence of 10µM of Y27632 (Miltenyi Biotec). Three days later the electroporated iPSC were selected with puromycin (0.5 µg/mL) for 48h. After 1 week, 23 individuals clones were picked, expanded individually and screened by PCR for correct integration at the AAVS1 locus. Two pairs of oligonucleotides were used to detect the correct integration. One was used for the recombined allele (expected PCR product size: 1323 bp) and another one was used for the endogenous allele (expected PCR product size: 441bp). Homozygous clones were defined as positive for the recombined allele PCR product and negative for the endogenous allele PCR product.

Characterization of iPSC with *ETO2-GLIS2* Knock-in

Wild-type iPSC and iPSC from clone 10 with ETO2-GLIS2 knock-in were dissociated in single cells through Triple 1x treatment. $1x10^6$ cells were plated in Nunclon™ Sphera (ThermoFisher) to induce embryoid bodies development. Embryoid bodies (EB) were grown in Essential 6 Medium (ThermoFisher) replacing medium every 3 days. RNA from embryoid bodies were extracted after 24 days and the expression of markers from the three layers upon EB differentiation were tested for iPSC CTRL and iPSC clone 10. The list of all primers used for PCR and RT-PCR is reported in Supplementary Table 1.

Hematopoietic differentiation of iPSC

Three clones with ETO2-GLIS2 and CTRL iPSC were induced to differentiate into hematological lineage using 2D co-culture system on matrix adapted from (12). Briefly 3-4 colonies of each iPSC were picked and put into Geltrex (Thermofisher) coated wells in Essential 8 medium. For the first 2 days, the cells are grown in StemPro34 (SP34) medium supplemented with 1X penicillin/streptomycin (PS, GIBCO), glutamine (2 mM), ascorbic acid (50 µg/ml, Sigma) monothioglycerol (MTG, 15 mM, Sigma), BMP4 (5 ng/mL) and VEGF (50 ng/ml) and CHIR (930 ng/mL). The medium was replenished and supplemented until day 18 with a sequential cytokine cocktail, as shown in Table 2. Hematopoietic differentiation of iPSC was then assessed at day 13, 15 and 18 using flow cytometry analyses for hematopoietic progenitors and megakaryocytic markers. All flow cytometry antibodies were purchased from BD Biosciences (CD34-PECy7: #560710, CD43-APC: #560198, CD41-V450: #561425, CD42-PE: #561854, CD61-APC: #564174, CD235a-PE: #561051). To test the proliferation properties of hematopoietic progenitors, CD43+ sorted cells were cultured in SP34 medium supplemented with Flt3L 10ng/ul, G-CSF 20ng/ul, IL3 10ng/ul, IL6 10ng/ul, SCF 25ng/ul, TPO 10ng/ul and GM-CSF 10ng/ul for 4 days. All data were analyzed with FACSCantoII or Fortessa with FACS Diva software (BD Biosciences).

Methylcellulose culture

To test self-renewal properties CD43+ hematopoietic cells derived from iPSC differentiation were sorted by flow cytometry and cultured in methylcellulose (H4230, StemCell Technologies) enriched with Flt3L 10ng/ml, G-CSF 20ng/ml, IL3 10ng/ml, IL6 10ng/ml, SCF 25ng/ml, TPO 10ng/ml, GM-CSF 10ng/ml. All cytokines used were purchased from Peprotech, UK. Cells were replated every 7 days for 4 weeks.

DNA/RNA extraction and RT-PCR

DNA and RNA were isolated using the RNeasy Mini and All prep DNA/RNA Micro Kit (Qiagen) and quantified by NanoDrop (ThermoScientific). Reverse transcription was done with SuperScript II from Invitrogen. Quantitative-PCR for ETO2-GLIS2 was performed using SYBR Select master mix (Applied Biosystems) on 7500HT Fast Real-Time PCR System (Applied Biosystems) following manufacturer's recommendations. All Primers used were reported in Supplementary Table 3. *GATA3* and *ERG* expression was tested with TaqMan probe (Applied Biosystem). *GATA3*: Hs00231122, *ERG*: Hs01554629.

Transcriptome sequencing & bioinformatics analyses

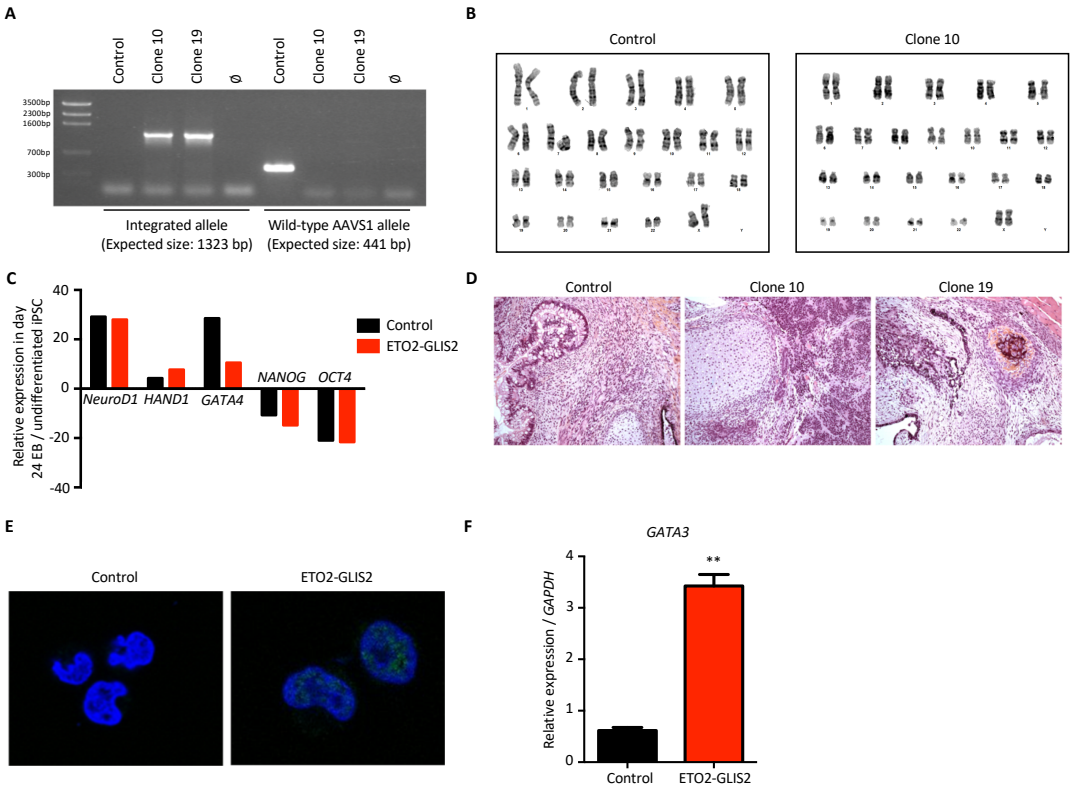
Whole transcriptome sequencing (WTS) was performed at the "Giorgio Prodi" Cancer Research Center using Illumina Nextseq500 platform. cDNA libraries were synthesized from 250 ng total RNA using the TruSeq Stranded mRNA kit (Illumina) following manufacturers' instructions. Briefly, poly-A

containing mRNA molecules were reverse transcribed to double stranded cDNA fragments that were then adenylated at 3' ends and ligated to single-index adapters. PCR-enriched libraries were then quantified by Quant-It picogreen assay (Thermo-Fisher) and sized with the High Sensitivity kit on the 2100 Bioanalyzer (Agilent Technologies). Sequencing was performed at 2x80bp using Illumina Sequencing by synthesis (SBS) technology, yielding an average of 45.68 million reads per sample.

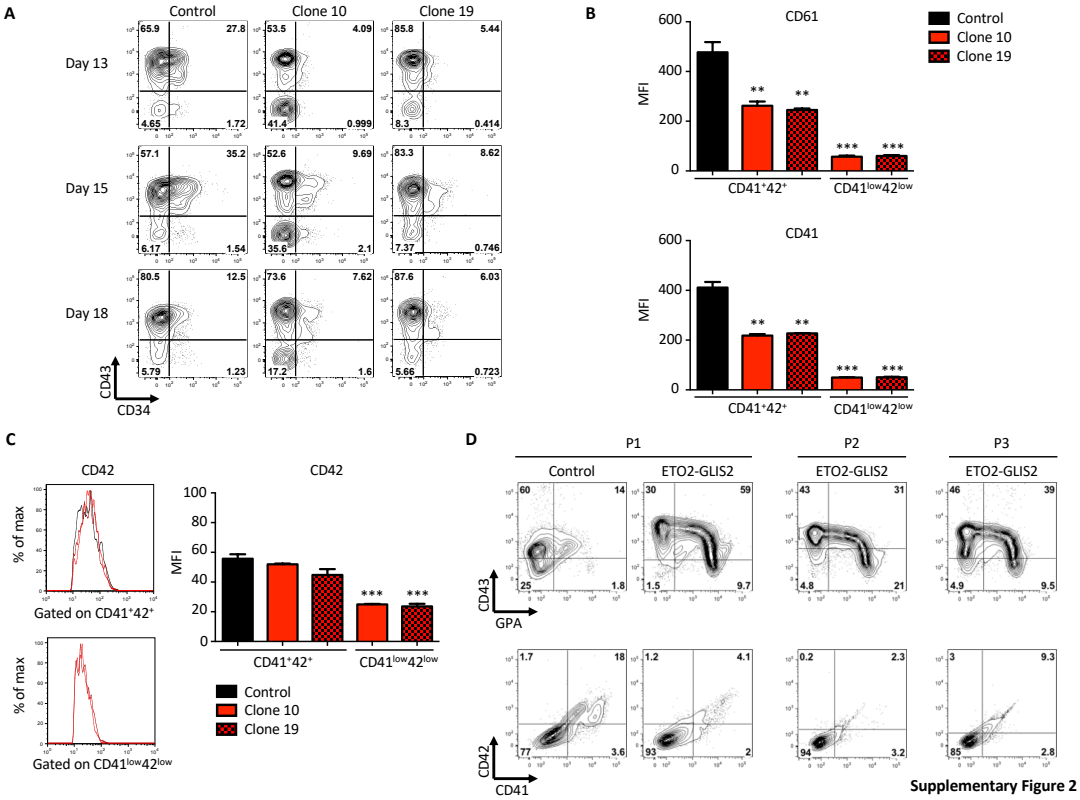
Analysis were performed on R (3.5.1 version) using heatmap3 (1.1.6), Rtsne (0.15) and ggplot2 (3.1.1) packages. Gene set Enrichment Analysis were perform using GSEA software from the Broad Institute. Lineages enrichment analyses were perform using the xCell software (14). Gene Ontology was performed using online gene ontology resource (http://geneontology.org)

Statistical analyses

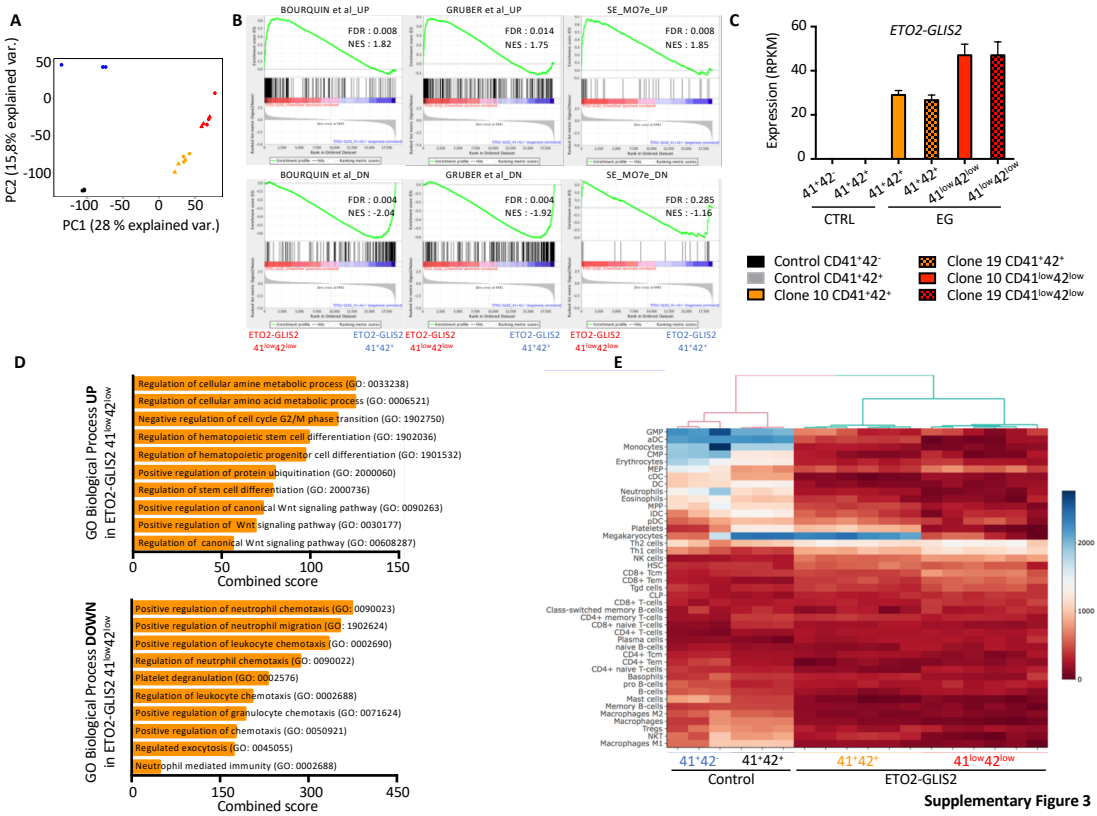
Statistical analyses were performed using the GraphPad Prism software (version Prism 6-2) except otherwise mentioned.



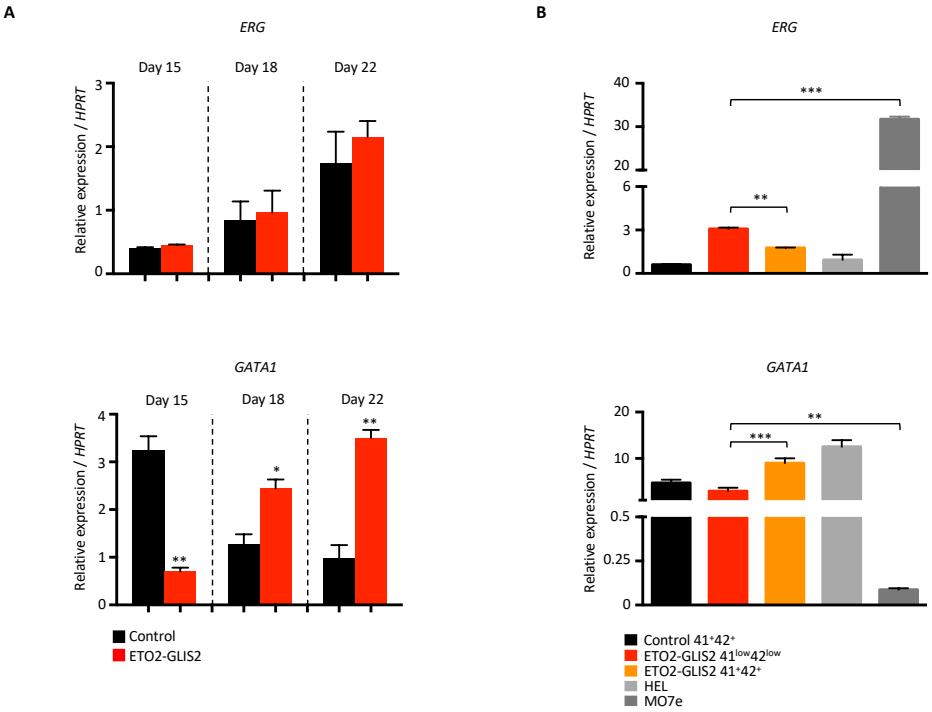
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

Article 2

ETO2-GLIS2 induces changes in transcription factor balance between early and late stages of leukemia in a human pluripotent stem cells model

INTRODUCTION

De novo acute megakaryoblastic leukemia (AMKL) represent a subtype of acute myeloid leukemia (AML) impacting megakaryocytic lineage mainly occurring in young children. AMKL often result from chromosomal alterations leading to fusion oncogenes. The ETO2-GLIS2 (EG) fusion resulting from chromosome 16 inversion inv(16)(p13.3;q24.2) is one of them, found in 20-25% of patients, and is associated with a very young age at diagnosis and poor prognosis (*250–253*).

ETO2 is a transcriptional cofactor express in most of hematopoietic cells. ETO2 interact with the heptad complex and participate to HSC maintenance and repression of the terminal erythroid and megakaryocytic differentiations (*257*, *258*). GLIS2 is a transcription factor from the GLI-similar family, which present a strong homology with GLI factors involved in hedgehog pathways, and possess DNA-binding domain constituted of contain 5 zinc fingers motifs (*255*). GLIS2 is supposedly not express in hematopoietic system although a report of impairment of HSC repopulation ability after GLIS knock-down suggest a potential role in hematopoiesis (*259*).

To date, molecular analyses on the transcriptional consequences of the fusion has been primarily obtained in human cell lines and patient derivate xenograft (PDX) models derived from patients with EG leukemia. RNA-seq analysis in HEL cells that overexpress EG reveal a strong transcriptomic deregulation with notably an upregulation of ERG and downregulation of GATA1 factors. ChIP-seq analysis reveal that the fusion is able to bind ETO2 canonical loci but also at new loci, many of which are share with ERG (*263*).

In addition, a recent murine transgenic model for EG expression has provided evidence that the fusion can efficiently leads to leukemia development and that leukemogenesis is depend on ontogeny. Indeed, murine EG fetal cells lead to shorter-latency and more aggressive leukemia with a megakaryoblastic phenotype compared to adult cells that more rarely lead to leukemia with rather myeloid phenotypes. Furthermore, although the fusion is able to transform a large range of immature progenitors from HSC to MPP4, the leukemia latency and

phenotype vary depending on the targeted cells (*264*). These cell-context differences suggest that elements such as chromatin state of the cell in which the fusion arise play an important role in EG leukemia development. However, observations in murine models cannot be fully transposed to human as differences exist in genome and epigenome between species (*267*). For instance, one of the rare additional mutation found in EG, trisomy 21 (T21), is difficult to study in mouse due to syntenyc differences resulting in human chromosome 21 being split in 3 different chromosomes in mouse.

Together, while these models provide insight into the biology of EG leukemia, developing faithful model of leukemia initiation and progression within a human genome context is required to better understand the molecular mechanisms underlying this aggressive leukemia subgroup.

However, access to human primary normal fetal hematopoietic cells is a challenge and the identification of surrogate source of cells would represent opportunities to systematically assess leukemogenesis using different pediatric leukemia oncogenes. One option to overcome these limitations is the use of human induced pluripotent stem cells (iPSC). Indeed, the system allow to monitor dynamic processes such as hematopoietic differentiation in human cells and contrary to primary samples, to obtain reproducible and large number of cells. Furthermore, iPSC can be precisely edited to obtain the mutation of interest and compared to isogenic control (*311*).

We have previously developed a model in which normal iPSC have been engineered by zinc finger nucleases to introduce the EG fusion at the AAVS1 locus under the control of CD43 promoter (pCD43-EG). Using this model, we have shown that the fusion induced hematopoietic differentiation and transcriptomic alterations consistent with features found in patients. However, hematopoietic cells obtained from this pCD43-EG did not lead to leukemia development upon injection into immunodeficient recipient mice (*322*).

As only rare models of leukemogenesis starting from iPSC have been reported to date (*310*, *313*), improving this iPSC-based modeling approach to more faithfully reproduce human disease development is of wide relevance. Therefore, to better mimic the situation found in EG human leukemia, we decided to precisely recreate the chromosome 16 inversion to obtain

the exact same genetic context than in patients, including expression of the fusion under the control of the endogenous ETO2 promoter.

RESULTS

CRISPR engineering reproduces the ETO2-GLIS2 chromosomal inversion in iPSC

To recapitulate the chromosome 16 inversion, we engineered normal iPSC from healthy donor through CRISPR-Cas9 technology. For this, iPSC were electroporated with Cas9 protein and sgRNA targeting intron 11 of ETO2 and intron 2 of GLIS2 to induce DNA double strand breaks at the location of the most common fusion breakpoints (i.e. fusion between exon 11 of ETO2 and exon 3 of GLIS2 representing 80% of the EG patients (*254*)). Single-cell were then plated to obtained colonies that were screened by PCR for the inv(16)(p13.3;q24.2) (Fig 1A). As, T21 is a frequently acquired alteration in AMKL in general and in EG AMKL specifically (Fig S1A) (*323*), we also edited an iPSC line derived from a donor with T21. Dilution PCR on bulk iPSC after editing indicated a low frequency of EG fusion, estimated between 1/500 and 1/1000 cells (Fig 1B). Single colonies replating led to the identification of three independent clones with the EG fusion (Fig 1C), incuding two clones from disomic iPSC (EG Cl1 and EG Cl2) and one clone from trisomic iPSC (EG Cl3, hereafter named T21 EG). Matching isogenic control clones were also obtained. Of note, the inversions are reciprocal as confirmed by detection of the reciprocal GLIS2-ETO2 fusion and of only one copy of wild-type ETO2 and GLIS2 alleles (Fig 1E). Consistent with the normal ETO2 gene expression profile, quantitative expression analysis confirmed that the *ETO2-GLIS2* transcript is not expressed at the pluripotent iPSC stage whereas it is detected in hematopoietic cells obtained from iPSC differentiation *in vitro* (Fig 1F).

Together, we engineered three independent human iPSC clones that faithfully reproduce the genetic alteration found in patients and which controls expression of the EG fusion from the endogenous ETO2 locus.

ETO2-GLIS2 alters *in vitro* hematopoietic differentiation of human iPSC

In order to investigate the impact of EG on hematopoietic differentiation in human cells, we differentiated iPSC toward hematopoietic lineage using a 2D feeder-free protocol relying on

sequential addition of cytokines and harvested the hematopoietic cells at day 13 (Fig S2A). Flow cytometry analysis revealed that whereas control (CTRL) cells contained a balanced proportion of megakaryocytic, erythroid and myeloid cells, both EG clones presented a strong increase of megakaryocytic cells (Fig 2A, 2B & 2C). Furthermore, abnormal populations that did not exist in CTRL were present in EG conditions. Indeed, EG cells presented a $CD41^{+}117^{+}$ population also found in EG AMKL patients. Besides, in addition to the maturing $CD41^{+}42^{+}$ megakaryocytes, a $CD41^{low}42^{low}$ population was observed (Fig 2D & 2E). CD56 was also strongly expressed in EG cells (Fig 2D & 2E) as observed in patients. Globally, the same phenotypes were also observed in cells derived from the T21 EG clone (Fig S2). However, notable differences included an increase in erythroid cells in T21 EG vs disomiques EG, in line with previous reports of enhanced erythroid lineage development in T21 context (*234*, *324*) (Fig S2B, S2C & S2D). Also, a small proportion of T21 EG cells co-expressed the megakaryocytic CD42b and erythroid CD235a markers (Fig S2E).

At the molecular level, quantitative PCR for two known targets of the fusion showed an upregulation of *ERG* and downregulation of *GATA1* (Fig 2F), consistent with previous results and a hematopoietic differentiation blockage induced by EG. We then tested self-renewal potential of these cells using colony forming assays in semi-solid methylcellulose cultures. Compared to CTRL, EG cells presented an increased self-renewal with a clonogenic potential in secondary plates. However, no colonies were obtained in the third plates (Fig 2G). Interestingly, T21 EG cells, but not T21 CTRL, led to a strong increase in clonogenic activity at passage 1, suggesting a potential cooperation mechanism between EG and trisomy 21 (Fig S2H). However, T21 EG clonogenic activity was lower than EG cells in secondary plates and similarly lacked clonogenic activity in third plates.

Altogether, these data show that EG expression in cells derived from *in vitro* hematopoietic differentiation of iPSC leads to an amplification of megakaryoblastic features, the generation of abnormal phenotypes and an increase self-renewal potential consistent with the hematopoietic differentiation alteration found in EG AMKL patients.

iPSC-derived ETO2-GLIS2 hematopoietic cells lead to a fully penetrant leukemia with AMKL features

To assess the leukemic potential of these cells, we injected 1.10^{6} iPSC-derived hematopoietic cells at day 13 of differentiation from either CTRL or EG clones into NOD.*scid.Il2Rγc*^{null}-SGM3

(NSG-S, expressing human SCF, GM-CSF and IL3 cytokines). All mice engrafted with EG cells (20/20 mice) developed a lethal disease with a mean latency of 125 days while no disease was observed from CTRL cells (Fig 3A). Leukemic mice presented with variable splenomegaly (Fig S3A) and an AMKL phenotype with expression of CD117 and CD42b markers (Fig 3B) consistent with *in vitro* and patient phenotypes (Fig S3B). Histological analysis confirmed expression of other EG blast markers, including CD44 and CD56, in both bone marrow and spleen of leukemic mice (Fig 3C). Notably, injection of leukemic cells from primary recipients led to similar disease development in secondary recipients (Fig 3A). Importantly, we confirmed that cells obtained from our previous model of transgenic iPSC with EG fusion expression under the CD43 promoter (*322*) did not engraft in NSG-S mice (Fig S3C), emphasizing the importance of a faithful genetic reproduction for *in vivo* disease development. Strikingly, T21 EG cells presented a much weaker *in vivo* leukemic potential with a lower penetrance (only 3/11) and a longer latency (198 days vs 125 days; $p < 0.001$) (Fig S3D). To assess whether human cytokine stimulation was essential for *in vivo* leukemogenesis, we then performed the same engraftment procedure in regular NSG recipients. Although, a longer latency was observed in the NSG context (189 days), all recipients eventually developed fatal leukemia (Fig S3E) presenting similar features as in NSG-S recipients, indicating that human SCF, GM-CSF and IL3 cytokine expression accelerates leukemogenesis but is not essential in this model.

Together, these data demonstrate that hematopoietic cells derived from this EG iPSC-based model efficiently leads to leukemia development and is thus a suitable system to study the process of EG leukemogenesis starting from normal human hematopoietic cells.

ETO2-GLIS2 leukemic cells show a phenotypic plasticity

Although the phenotype of leukemic mice is very consistent for most surface markers (Fig S3F), a significant variability in the expression of CD41 and CD45 was observed, reflecting the variability observed in EG patients. Indeed, most of the *in vivo* leukemia presented with a $CD41^{+}45^{-}$ phenotype but a significant proportion of cells also presented with a $CD41^{+}45^{+}$ phenotype while some others were exclusively $CD41^{-}45^{+}$ (Fig 3D). The proportion of each phenotype was variable in the different recipients (Fig 3E). In order to test whether the $CD41^{+}45^{-}$ and the $CD41^{-}45^{+}$ populations are independently maintained or have a hierarchical relationship (i.e. one population could give rise to the other), we engrafted recipients with either $CD41^{+}45^{-}$ or $CD41^{-}45^{+}$ cells from differentiation day 13 iPSC-derived hematopoietic cells

(Fig 3F). No difference in disease latency was observed between the two groups (Fig 3F). In addition, phenotypic characterization of the leukemia showed that each donor cell population could give rise to leukemia presenting both $CD41^{+}45^{-}$ and $CD41^{-}45^{+}$ cells (Fig S3G).
These data indicate that EG is able to generate different type of leukemic phenotypes and suggest that these cells present is a certain degree of phenotypic plasticity.

ETO2-GLIS2 iPSC-based model presents changes in hematopoietic hierarchy similar to patient

We then took advantage of this model that allows to access both early and late stage of the disease development to performed molecular analyses including single-cell RNA sequencing (scRNA-seq) and Assay for Transposase-Accessible Chromatin (ATAC-seq). We used four conditions: day 13 *in vitro* iPSC-derived control (***in vitro* CTRL**) and EG (***in vitro* EG**) cells, after *bona fide* leukemia development in NSG-S mice (***in vivo* EG**) and on patient samples (**Patients**) (Fig 4A). This allowed us to obtain the full spectrum of early and late consequences of EG expression during leukemogenesis. From 4276 to 10127 cells were obtained for scRNA-seq analysis after filtering. We then integrated cells from all conditions and performed UMAP clustering which led to the identification of 15 clusters (Fig 4B). Clusters were labeled based on the associated gene expression (Fig 4B & 4C) and cell cycle status was inferred using Seurat (Fig S4A). We then obtained the projection of each individual condition onto the UMAP of the integration (Fig S4B) and the proportion of the different clusters (Fig S4C). Two clusters (Cluster 0 and Cluster 1) were particularly associated with EG transformation and contained an increasing number of cells from *in vitro* CTRL, to *in vitro* EG and *in vivo* EG and represented more than half of the cells in patients (Fig 2D). Cluster 1 was associated with the megakaryocyte lineage. Cluster 0 could not be labelled precisely but showed features of multipotent progenitors. Interestingly, cluster 14 was virtually absent in *in vitro* CTRL and weak in *in vitro* EG but increased in *in vivo* EG and patient (Fig 4E). This cluster does not express any lineage associated genes but rather show a strong increase in ribosomal genes and a decrease in mitochondrial associated genes (Fig S4D). This abnormal profile of expression was previously reported for AML (*325*). Next, using a transcriptomic signature associated with EG patient leukemic cells (within a large cohort of pediatric AML (*192*), we computed a score for this signature in each cell. This EG gene expression signature was mostly enriched in Mk and immature clusters (Fig 4F). Besides, *in vitro* CTRL cells were not positive for the signature

except for a few megakaryocytic cells, whereas several *in vitro* EG cells in megakaryocyte and immature clusters are positive and virtually all *in vivo* EG and patient cells were enriched for the signature (Fig 4F).

These data suggest that the molecular features of EG expressing cells can be associated with a wide variety of cell populations ranging from immature to megakaryocytic cells and that these populations represent the majority of leukemic blasts in *in vivo* models and patients. Importantly, these data further show that the iPSC-derived leukemia model is closely reproducing the patient situation.

ETO2-GLIS2 induces progressive changes in the transcription factors accessibility landscape

To obtain insights into the consequences of EG on the chromatin landscape, we performed ATAC-seq analysis on $CD41^+$ flow sorted cells from the same conditions as scRNA-seq (Fig 4A). Every sample passed quality controls for sequencing depth and peak calling (Fig S5A and S5B). Principal component analysis (PCA) showed a consistent clustering of replicates with a higher variability for patients that may in part result from the genetic background differences (Fig S3C). Besides, *in vitro* EG samples were closer to *in vitro* CTRL than to *in vivo* EG and patients reflecting differences between initiation and late stages of the disease. Many chromatin changes were observed between CTRL and EG (Fig 5A). Of note, chromatin opening is globally higher *in vivo* and in patient samples than *in vitro* for EG specific peaks (Fig 5B). Moreover, when comparing the number of differential-to-CTRL peaks, we observed a massive chromatin reorganization during *in vitro* to *in vivo* transition with 637 gained and 2019 lost peaks for *in vitro* EG versus 20423 and 19486 gained and 9730 and 9771 lost peaks for *in vivo* EG and patients respectively (Fig 5C). We next looked for the motifs under these differential-to-CTRL peaks in order to find which transcription factors may underlie early and later leukemic stages. *In vitro*, we found that EG induced a gain in accessibility to motifs from transcription factors of the RUNX and Homeobox families and a loss of accessibility to motifs associated with terminal megakaryocytic maturation transcription factors such as NFE2 (Fig 5D). *In vivo*, these RUNX and Homeobox motifs remained enriched in gained peaks but additional motifs became significantly enriched including numerous ETS associated motifs (e.g. ERG, FLI1, ETS1 gained in EG conditions) while other were significantly lost (e.g. GATA motifs) (Fig 5D). Therefore, the chromatin reorganization that occur during *in vitro* to *in vivo* transition leading to leukemia seems to implicate a change in chromatin accessibility between ETS and GATA transcription

factors whereas RUNX and Homeobox factors activities could contribute to early initiation steps (Fig 5E). This idea was strengthened by the observation that when comparing directly *in vitro* EG to patients, additional ETS and GATA motifs were also gained and lost, respectively, while Homeobox motifs accessibility was not significantly changed (Fig S5D). Here again, ATAC-seq data showed a striking similarity between cells from the iPSC-derived leukemia and patient samples (Fig 5C, 5D, 5E & S5E).

Together, the data strongly suggest that EG induces chromatin reorganization involving RUNX and Homeobox activities at early steps of leukemogenesis while an unbalance in ETS and GATA activities is mostly visible in a later stage during progression to full-blown leukemia.

ETO2-GLIS2 leukemia show high expression of the DLX3 homeodomain

Homer analyses suggested that the Homeobox motif accessibility gains are associated with the DLX gene family rather than with the HOXA or HOXB genes families often involved in human AML. DLX is a family of 6 members (DLX1 to DLX6) located in tandem on the genome. They have been involved in craniofacial and testis development. In human, they are mostly associated to bone development and more specifically to osteoblastic differentiation (*326*, *327*). Rare studies investigated the role of DLX factors in normal hematopoiesis, suggesting a role for DLX1/2 in NK cells maturation and a role for DLX4 in megakaryocyte differentiation in mouse (*328*). DLX factors have also been linked to cancers such as glioblastoma, breast cancer and ovarian cancer. While DLX5 has been suggested to be implicated in T-cell lymphoma (*329*), little is known about the role of DLX in malignant hematopoiesis.

To further determine if these factors present functional relevance beyond motif enrichment, we quantified expression of *DLX* factors in iPSC-derived *in vitro* CTRL and EG hematopoietic cells (Fig 6A). *DLX3* showed the most striking difference with a ~100-fold higher expression in EG than in CTRL (Fig 6A). *DLX1* and *DLX4* showed a moderate (3-fold and 2-fold respectively) upregulation compared to CTRL while DLX2 and DLX5 were not significantly different. High *DLX3* expression was also observed using RNA-seq data from our previous works in two other EG model (Fig S6A & S6B). We then confirmed high *DLX3* expression by quantitative PCR on cells from iPSC-derived *in vivo* leukemia and EG leukemic patients (Fig 6B). High *DLX3* was also observed human leukemia cell lines that expressed the EG fusion (MO7e and WSU-AML) but not in the non-EG leukemia HEL cell line (Fig 6C). Notably, *DLX4* expression was also more highly expressed in these cellular contexts but to a lower extent than *DLX3* (Fig S6C).

Importantly, using an independent cohort of pediatric AML (*192*), we found that *DLX3* was significantly more highly expressed in EG as compared to all other pediatric AML molecular subgroups. While *DLX2* and DLX4 were also more highly expressed in EG patients they showed a lower differential expression (Fig 6D). These data nominated *DLX3* as the most consistently and strongly overexpressed *DLX* gene in EG cells (Fig 6A, S6A, S6B & 6D) and high DLX3 protein expression was validated in EG cell lines and patient samples whereas it is undetectable in other AML samples (Fig 6E). As additional evidence of the direct link between EG and high DLX expression, we observed the presence of fusion binding sites at the promoter region of *DLX1*, *DLX2*, *DLX3* and *DLX4* (Fig 6F) in our previously published ChIP-seq data in which control HEL cells were compared to HEL cells overexpressing the EG fusion. No binding sites were detected at *DLX5* and *DLX6* loci in this model (Fig S6D).

Together, these data suggest that EG lead to high expression of both DLX3 and SP7, a gene regulatory module important for the control of bone formation transcriptional programs.

DLX3 is essential for EG cells proliferation

To assess the role of DLX3 in EG leukemic cells, we then performed a *DLX3* knock-down through an inducible shRNA approach using two shRNA that efficiently reduced *DLX3* expression (Fig 7A). M07e and WSU-AML cell lines expressing the fusion were transduced with the most efficient shRNA to characterize cell proliferation. Compared to shRenilla, proliferation of cells transduced with shRNA against *DLX3* was strongly inhibited (Fig 7B). These data indicate that DLX3 activity is essential to maintain proliferation of human EG cell lines in *in vitro* cultures.

In order to test whether DLX3 bears intrinsic transformation properties, we finally ectopically expressed DLX3 in normal hematopoietic progenitors. As current protocols do not allow the generation of all hematopoietic cell types (*295*), we used murine cells. Thus, bone marrow lineage-negative cells from wild-type mice were transduced with retroviruses encoding either control (GFP) or DLX3 cDNA and were grown in liquid cultures. No significant difference in the absolute number of cells was noted during the 18 days of culture, although a trend in favor of DLX3 is observed (Fig 7C). However, the percentage of cherry-positive cells strongly increased in DLX3-transduced cells suggesting a competitive advantage (Fig 7C). Furthermore, an increase of $CD41^{+}$ megakaryocytic cells was observed at early time point as well as the presence of a population expressing both megakaryocytic (CD41) and myeloid (CD11b)

markers, absent in the control condition. At a later time point, all control cells were exhausted with only low proliferation mastocytes remaining whereas DLX3 cells still present megakaryocytic $CD41^+$, myeloid CD11b and double positives $CD41^+11b^+$ cells (Fig 7D).
Together, these results support that DLX3 is essential for EG cells proliferation and contribute to the aberrant self-renewal properties of immature/Mk progenitor cells.

DISCUSSION

In this study, we have developed an iPSC-based model for EG pediatric AMKL. Our model demonstrates that EG induces differentiation alterations *in vitro* and lead to leukemia *in vivo*. This model provided several insights into pediatric leukemia development. Firstly, it demonstrates that the genetic engineering approach to reproduce oncogene expression is crucial for precise modeling. Secondly, it provided a reproducible tool to identify the cellular path from normal to leukemic transformation and showed that EG can lead to accumulation of a wide array of immature to megakaryocytic populations exhibiting some phenotypic plasticity. Finally, transcriptomic and epigenetic analyses of early and later stage of leukemogenesis confirmed that the balance between ETS and GATA transcription factors is crucial for full-blown in vivo leukemia development and identified a previous unrecognized transcriptomic program associated with leukemia initiation.

As our previous pCD43-EG iPSC model was unable to induce to leukemia (*322*) and as oncogene expression level can impact the phenotype in a murine EG model (*264*), we have recreated the chromosome 16 inversion seen in patient using CRISPR-Cas9 strategy.
Our data demonstrate that the expression level and profile is critical to obtain leukemia development with the EG fusion oncogene and highlight the importance of faithfully modeling the genetic event associated with pediatric leukemia. Furthermore, these observations are also relevant for others AML as demonstrate for the MLL-AF9 oncogene. Indeed, using either retroviral transduction or knock-in at endogenous locus, Chen et al, demonstrated striking differences. Thus, in mouse GMP, retroviral MLL-AF9 expression leads to a higher fusion oncogene level, associated to increased proliferation and clonogenic capacities *in vitro*, and leads to efficient leukemic transformation *in vivo*. On the other hand, MLL-AF9 knock-in

expression, under MLL promoter, induces moderate increase in proliferation and self-renewal *in vitro*, and is not able to transform GMP cells (*330*).

Through phenotypic and single-cell transcriptomes characterizations at different stages of leukemogenesis, we observed that EG leads to a strong megakaryocytic bias during *in vitro* hematopoietic differentiation in this iPSC model. This is highlighted by the accumulation of abnormal populations presenting variable expression of megakaryocytic markers (e.g. $CD41^{low}42^{low}$ and $CD41^{+}117^{+}$) also found in patients. However, EG does not abrogate accumulation of abnormal cells close to other lineages as demonstrated by the presence of leukemic cells in other clusters such as cluster 9 or 6 associated to pro-erythrocytes and monocytes/macrophages respectively. Interestingly, a clear phenotypic variability was observed for the CD41 and CD45 surface markers and transplantation of purified populations supported the existence of a phenotypic plasticity in leukemic blasts rather than a hierarchical relationship associated with differentiation processes. Notably, such plasticity has been observed in other types of pediatric cancers (*331*, *332*). Importantly, these data are fully consistent with the observation that, in addition to AMKL (AML-M7), EG is also recurrently associated with other AML subtypes with phenotypes ranging from immature AML-M0 to more myeloid AML-M4/5 (*333*).

Interestingly, T21, often found in EG leukemic cells, do not change the phenotype except for a strong proliferation increase but without self-renewal and enhanced erythroid differentiation. This later observation is puzzling as it is in contradiction with down regulation of erythroid genes such as KLF1 induced by the fusion as well as with the anemia often found in patients. However, an unexpected finding is that T21 seems not to cooperate with EG to induce leukemia and on the contrary impairs leukemic potential. While other analyses will be required to define whether this observation is associated with a different potency of the T21 iPSC clone that we used or the order of acquisition of genetic alterations, these data are consistent with a recent observation that T21 was not required for DS-AMKL leukemogenesis (*237*). Together with the rare occurrence of additional mutations in EG leukemia in patients, our model support that T21 is not required for EG leukemogenesis and that EG is sufficient to efficiently induce a penetrant disease development upon expression in a ontogenically-relevant hematopoietic cell.

In fact, our molecular data provides further observations that the EG induces pleiotropic changes during transformation that involves a massive but progressive rewiring of the chromatin landscape. Indeed, we found that in the very early steps, RUNX and Homeobox factors show higher activity through increased accessibility to their binding motifs whereas in the late steps the more massive chromatin reorganization is associated with strong increase in ETS factors accessibility combined with lower accessibility to GATA motifs. These data confirm previous data showing that the ETS/GATA balance is critical in EG leukemic cells (*263*). As an important novel aspect, this model allowed to study the early steps that arise shortly after the expression of the EG fusion in a normal hematopoietic progenitor. This identified a novel transcription regulatory network active in EG leukemia. Using several public and from our lab RNA-seq and ChIP-seq datasets we find that a transcriptional program normally associated with bone formation is expressed in EG cells. This program involves the transcription factors DLX3 and DLX4 as well as their co-factor SP7 that are strongly upregulated. Direct binding of the EG fusion close to their gene promoter strongly suggest that they are direct targets. This program seems critical for EG action as we demonstrate that DLX3 inhibition strongly impairs EG cells proliferation and that overexpression of DLX3 in wild-type progenitors confers an immature/Mk phenotype. These findings are in line with a previous study linking DLX4 to megakaryocyte differentiation in mouse (*328*).

Together, this work shows the interest of developing pediatric leukemia models that both reproduce faithfully the genetic alterations found in patient and target developmentally-relevant human hematopoietic. This iPSC modeling approach identified novel molecular regulator of leukemogenesis and now represent a reproducible platform to perform additional molecular and screening studies with the perspective to develop novel therapeutic strategies.

FIGURES LEGENDS

Figure 1: Engineer the inv(16)(p13.3q24.4) and endogenous ETO2-GLIS2 fusion in human iPSC.

(**A**) Schematic representation of the experimental design to reproduce the ETO2-GLIS2-encoding inv(16)(p13.3q24.3) in normal human iPSC using the CRISPR/Cas9 gene editing approach. Yellow stars indicate the location of gRNA to generate the breakpoints found in patients. (**B**) Assessment of inv(16)(p13.3q24.4) frequency by PCR on serial dilutions of DNA from bulk transfected cells. 1000 cells roughly correspond to 6ng of DNA. Forward primer is located on *ETO2* intron 11 and reverse primer is on *GLIS2* intron 2. (**C**) Validation PCR on DNA from the *ETO2-GLIS2* (EG) fusion clones. (**D**) Sequence traces of the fusion points. (**E**) Fluorescent in situ hybridization on control and EG clones using ETO2 (red) and GLIS2 (green) probes. Chromosome inversion is highlighted by a yellow spot. (**F**) Quantitative RT-PCR for *ETO2-GLIS2* mRNA expression in control and EG clones in undifferentiated iPSC (that do not express endogenous *ETO2*) or day-13 hematopoietic cells, normalised on *RPLP0* expression. Mean +/- SD is indicated (n = 3).

Figure 2: ETO2-GLIS2 alters human iPSC hematopoietic differentiation *in vitro*.

(**A**) Representative flow cytometry analyses of cells obtained from CTRL, ETO2-GLIS2 clone 1 (EG Cl1) and clone 2 (EG Cl2) at day 13 of differentiation for the CD42b, CD235a and CD11b markers (gated on $CD43^+$ hematopoietic cells). (**B**) Percentage of cells in the gate shown in **A**. Mean +/- SD is indicated (CTRL: n = 16, EG Cl1: n = 13, EG Cl2: n = 12). (**C**) Average proportion of megakaryocytic (Mk: $CD43^+CD42b^+$), erythroid (Ery: $CD43^+CD235a^+$), myeloid (Myelo: $CD43^+CD11b^+$) and other cells as shown in **B** (**D**) Representative flow cytometry analyses of CTRL, EG Cl1 and EG Cl2 at day 13 for combination of markers frequently found in ETO2-GLIS2+ AMKL patients (gated on $CD43^+$ hematopoietic cells). (**E**) Percentage of cells in the gates shown in **D**. Mean +/- SD is indicated (CTRL: n = 16, EG Cl1: n = 13, EG Cl2: n = 12). (**F**) Quantitative RT-PCR for *GATA1* and *ERG* expression in control and EG at day 13, normalised on *RPLP0* expression. Mean +/- SD is indicated (CTRL: n = 3, EG: n = 3). (**G**) Clonogenic activity in methylcellulose assays. 10 000 day 13 hematopoietic cells from CTRL, EG Cl1 or EG CL2 were plated then counted and replated every 10 days. Mean +/- SD is indicated (CTRL: n = 4, EG Cl1:

n = 4, EG Cl2: n = 4). Statistical significance was determined using 1-tailed Mann-Whitney U test. *P < 0.05; **P < 0.01; ***P < 0.001.

Figure 3: iPSC-derived ETO2-GLIS2 cells induce leukemia in vivo.

(**A**) Survival curves of NSG-S mice injected with 0.7-1M day 13 iPSC-derived hematopoietic CTRL (n=7, black line) or EG (EG Cl1: n = 15, EG: Cl2 n = 5, green line) cells. Secondary recipients of EG cells (n=7) are also represented (dotted green line). (**B**). Flow cytometry analyses of bone marrow and spleen cells from a representative leukemic mouse at time of sacrifice. (**C**) HES coloration, CD44 and CD56 staining of bone marrow and spleen from leukemic mouse. (**D**) Flow cytometry analyses of the CD41/CD45 phenotypic variability in leukemic mice. A representative example of the 3 different phenotypes is shown. (**E**) Proportion of $CD41^{+}CD45^{-}$, $CD41^{-}CD45^{+}$ and CD41+45+ cells in individual leukemic mouse. (**F**) Left panel: schematic representation of the experimental design to assess leukemia-propagating activity of the two populations. Right panel: survival curves for recipients injected with 1.10^{6} flow sorted day 13 $CD41^{+}CD45^{-}$ or $CD41^{-}CD45^{+}$ cells. ($CD41^{+}CD45^{-}$: n = 5, $CD41^{-}CD45^{+}$ n = 4). Statistical significance was determined using log-rank Mantel-Cox test. *P < 0.05; **P < 0.01; ***P < 0.001.

Figure 4: Single cell analyses reveal pleiotropic changes in hematopoietic differentiation.

(**A**) Schematic representation of the 4 conditions used for molecular analysis (ATAC-seq & scRNA-seq): control (***in vitro* CTRL**, n = 2) and EG (***in vitro* EG**, n = 2) $CD43^{+}$ iPSC-derived hematopoietic cells at day 13 of differentiation, iPSC-derived EG leukemic mice (***in vivo* EG**, n = 2) & patient samples (**Patient**, n = 1) (**B**) UMAP representation of the integration of cells from all conditions and cluster identification. Clusters were defined using the Louvain algorithm and labeled based on their differentially expressed genes. (**C**) Projection of the expression level of the indicated gene on the UMAP of the integration. (**D**) Bar plots of the proportion of cells in the cluster 0 (top) and cluster 1 (bottom). (**E**) Bar plots of the proportion of cells in the cluster 14. (**F**) Projection of an ETO2-GLIS2 patient signature on the UMAP of the integration (left) or on individual UMAP (right). The signature has been defined using differential gene expression between ETO2-GLIS2 and other pediatric AML in a published dataset (*192*). Cells presenting a significant enrichment of the signature are shown in red. Blue indicates significant depletion of the signature. Grey represents lack of significance.

Figure 5: ETO2-GLIS2 reshapes chromatin accessibility to several transcription factors.

(**A**) Heatmap of hierarchical clustering of ATAC-seq signals in *in vitro* CTRL, *in vitro* EG, *in vivo* EG and Patients. Windows of 3kb centered on open chromatin peaks enriched in CTRL compare to Patients (CTRL peaks) or in Patients compare to CTRL (EG peaks) are shown. (**B**) Intensity profiles of CTRL-specific (lost in EG) or EG-specific (gained in EG) peaks shown in **A**. (**C**) Number of differentials peaks (significantly enriched or lost) in each EG conditions (*in vitro*, *in vivo* and patients) compared to CTRL. (**D**) Dot-plot representation of motifs under the differentials ATAC-seq peaks described in **C** ranked according to their p.value. Dots are colored according to the transcription factor family to which they belong. (**E**) Bar plot of the percentage of RUNX, Homeobox, ETS, GATA & NF-E2 motifs in differential peaks for each EG conditions as compared to CTRL.

Figure 6: ETO2-GLIS2 control DLX3 expression.

(**A**) Quantitative RT-PCR for genes of the *DLX* family in CTRL and EG iPSC-derived day 13 hematopoietic cells, normalised on *RPLP0* expression. Mean +/- SD is indicated (Ctrl: n = 3, EG: n = 3). (**B**) Quantitative RT-PCR for *DLX3* expression in CTRL and EG clones at day-13 hematopoietic cells, iPSC-derived EG leukemic mice and 3 different EG patients. Normalised on *RPLP0* expression. Mean +/- SD is indicated (n = 3 for all conditions). (**C**) Quantitative RT-PCR for *DLX3* expression in one non-EG cell line (HEL) and two EG cell lines (M07e and WSU-AML), normalised on *RPLP0* expression. Mean +/- SD is indicated (HEL: n = 3, M07e: n = 3, WSU-AML: n = 3). Statistical significance was determined using Student t test. (**D**) *DLX1-5* expression in published pediatric AML dataset (*192*). Two ETO2-GLIS2 patient subgroups are compared to other pediatric AML subgroups. Adjusted p-values was computed with Benjamini-Hochberg method. (**E**) Western-Blot showing DLX3 protein level in cell lines (left) with EG expression (M07e and WSU-AML) or without (HEL) and in cells from patient-derived xenograft models of pediatric AML patients (right) with (Patient #3 and #4) and without (Patient #5 and #6) the EG fusion. (**F**) Track depicting ChIP-seq for GFP in HEL-GFP and HEL-EG-GFP cells (*263*) at *DLX1* and *DLX2* locus (top) or *DLX3* and *DLX4* locus (bottom). (**G**) Heatmap showing the expression level of HOXA and HOXB gene families in pediatric AML (same dataset as in **D**). (**H**) Quantitative RT-PCR for *SP7* expression in CTRL and EG clones at day-13 hematopoietic cells, iPSC-derived EG leukemic mice and 3 different EG patients. Normalised on *RPLP0* expression. Mean +/- SD is indicated (n = 3 for all conditions). (**I**) *SP7* expression in

pediatric AML (same dataset as in **D**). (**J**) Track depicting ChIP-seq for GFP in HEL-GFP and HEL-EG-GFP cells (same dataset as in **F**) at SP7 locus. Statistical significance with the indicated tests: * <0.05, **<0.01, ***<0.001, ns >0.05.

Figure 7: DLX3 functional assays in ETO2-GLIS2 cells.

(**A**) Quantitative PCR for *DLX3* expression in M07e cells with shRNA against Renilla or DLX3. (**B**) Number of M07e and WSU-AML GFP positive cells transduced with shRNA against Renilla or *DLX3*. GFP positive cells were sorted and plated at 0.5 M/mL. (**C**) Absolute number (top) and percentage of cherry positives (bottom) cells transduced with cDNA CTRL or cDNA *DLX3*. (**D**) Immunophenotypes of CTRL or

DLX3 transduced cells for the Gr1, CD11b and CD41 markers at day 7 and 18 of *in vitro* culture (gated on cherry positive cells).

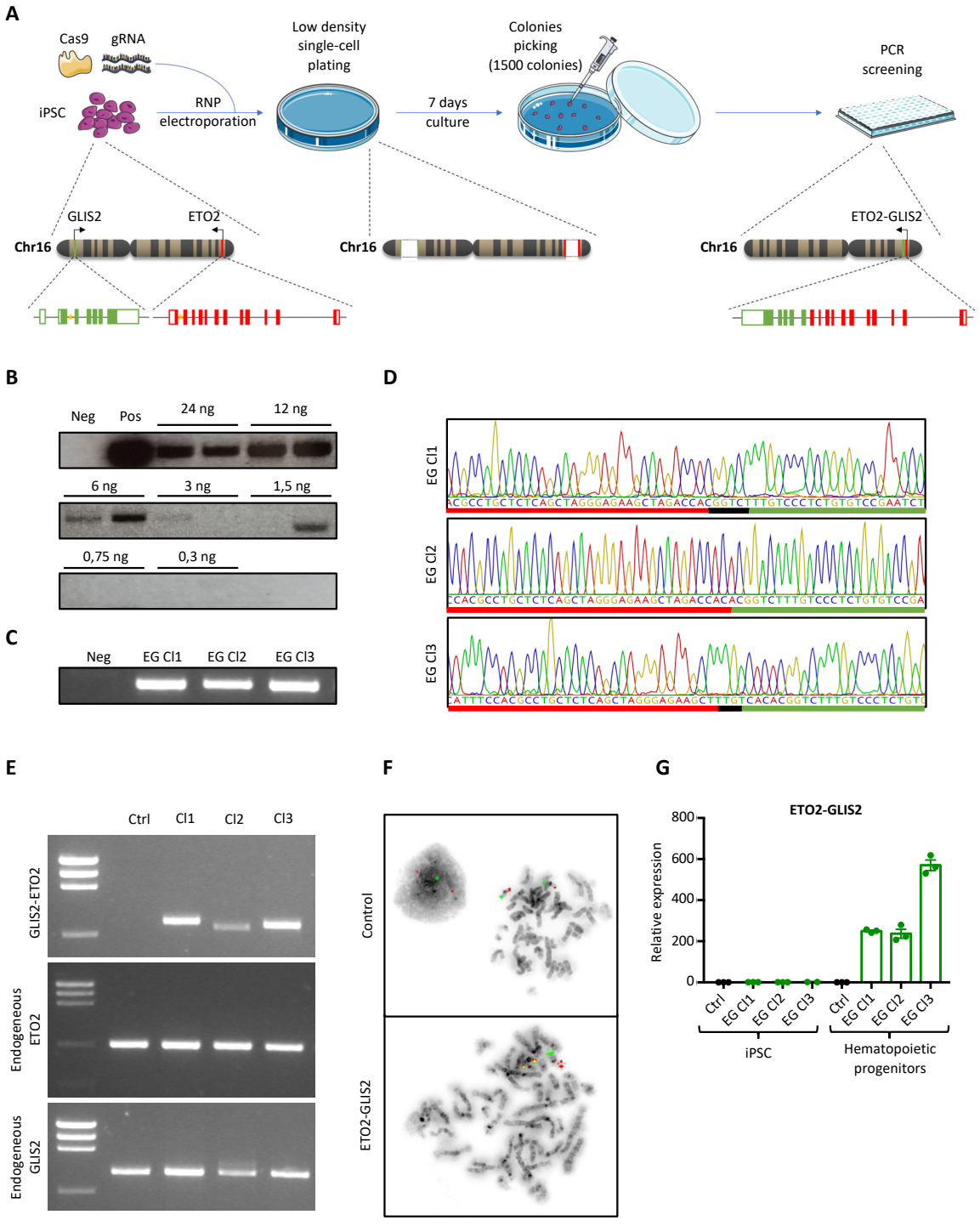

A
Cas9
gRNA
iPSC
RNP
electroporation
Low density
single-cell
plating
7 days
culture
Colonies
picking
(1500 colonies)
PCR
screening
GLIS2
ETO2
Chr16
Chr16
ETO2-GLIS2
Chr16
B
Neg
Pos
24 ng
12 ng
6 ng
3 ng
1,5 ng
0,75 ng
0,3 ng
C
Neg
EG Cl1
EG Cl2
EG Cl3
D
EG Cl1
EG Cl2
EG Cl3
E
Ctrl
Cl1
Cl2
Cl3
GLIS2-ETO2
Endogeneous
ETO2
Endogeneous
GLIS2
F
Control
ETO2-GLIS2
G
ETO2-GLIS2
Relative expression
800
600
400
200
0
Ctrl
EG Cl1
EG Cl2
EG Cl3
Ctrl
EG Cl1
EG Cl2
EG Cl3
iPSC
Hematopoietic
progenitors

Figure 1

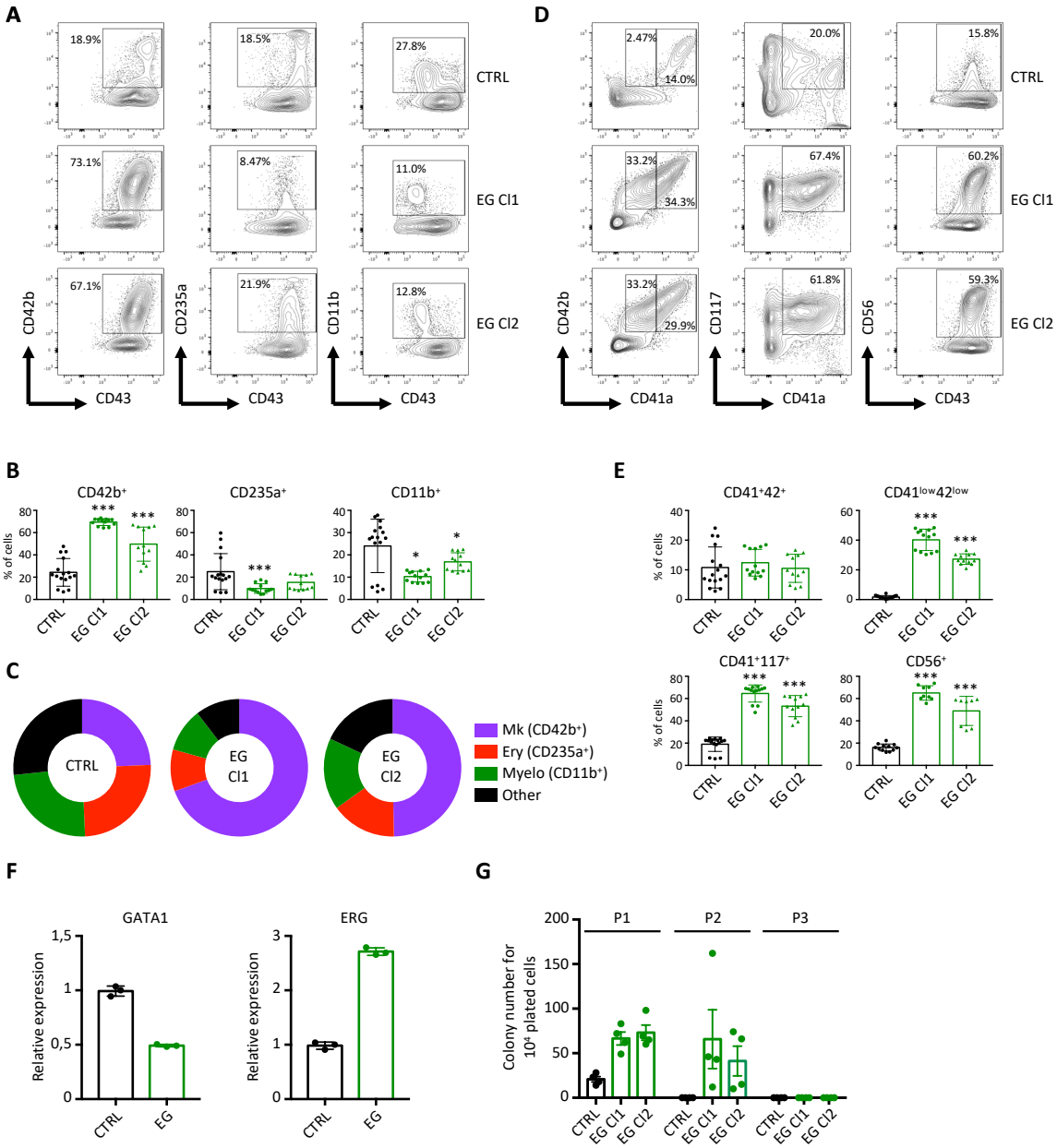

A
18.9%
18.5%
27.8%
CTRL
73.1%
8.47%
11.0%
EG Cl1
67.1%
21.9%
12.8%
EG Cl2
CD42b
CD43
CD235a
CD43
CD11b
CD43
D
2.47%
14.0%
20.0%
15.8%
CTRL
33.2%
34.3%
67.4%
60.2%
EG Cl1
33.2%
29.9%
61.8%
59.3%
EG Cl2
CD42b
CD41a
CD117
CD41a
CD56
CD43
B
CD42b⁺
CD235a⁺
CD11b⁺
% of cells
CTRL
EG Cl1
EG Cl2
C
CTRL
EG Cl1
EG Cl2
Mk (CD42b⁺)
Ery (CD235a⁺)
Myelo (CD11b⁺)
Other
E
CD41⁺42⁺
CD41ˡᵒʷ42ˡᵒʷ
CD41⁺117⁺
CD56⁺
% of cells
F
GATA1
ERG
Relative expression
CTRL
EG
G
P1
P2
P3
Colony number for 10⁴ plated cells

Figure 2

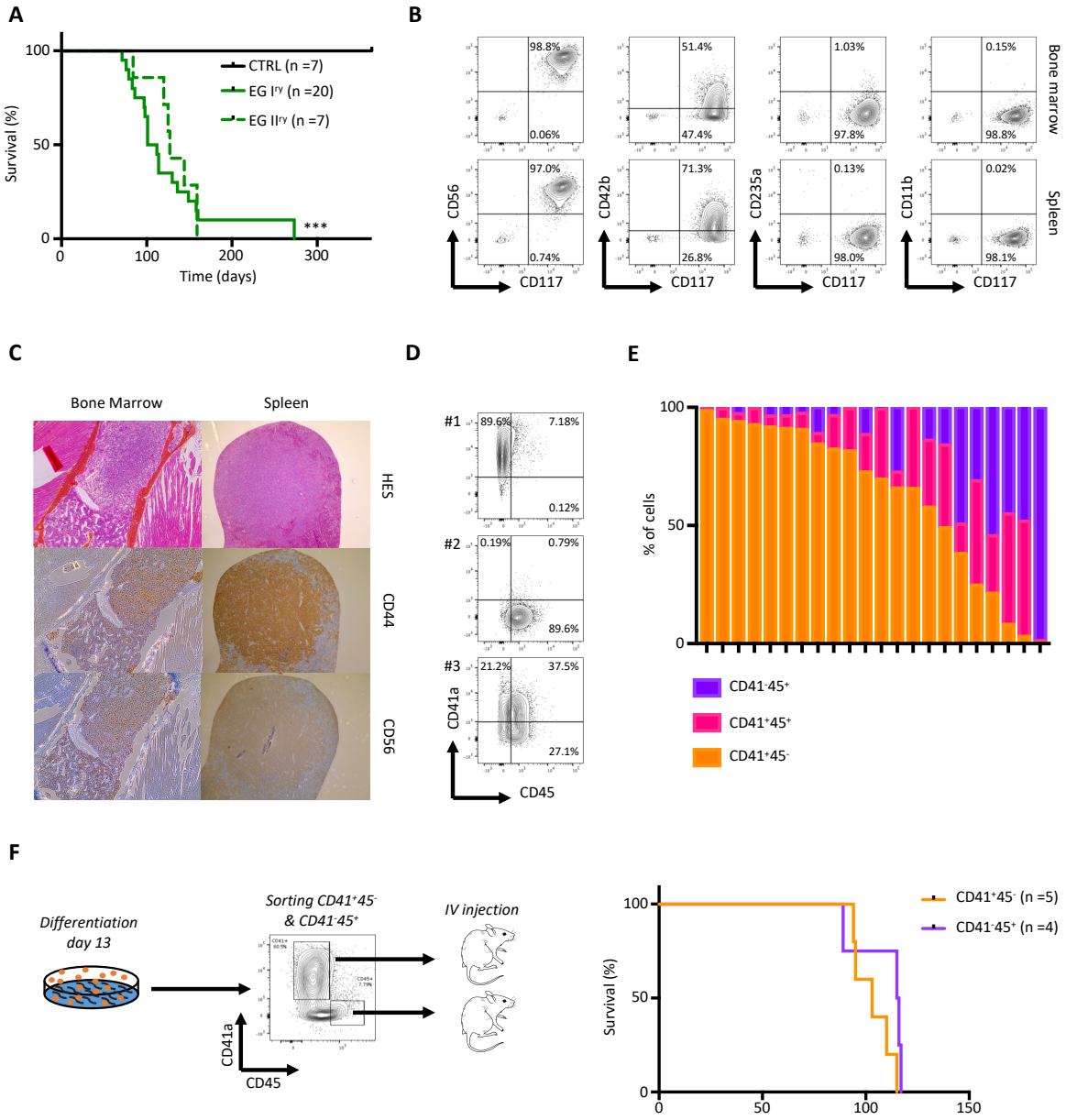

A
Survival (%)
Time (days)
CTRL (n =7)
EG I^IV (n =20)
EG II^IV (n =7)
***
B
Bone marrow
Spleen
CD56
CD42b
CD235a
CD11b
CD117
98.8%
0.06%
97.0%
0.74%
51.4%
47.4%
71.3%
26.8%
1.03%
97.8%
0.13%
98.0%
0.15%
98.8%
0.02%
98.1%
C
Bone Marrow
Spleen
HES
CD44
CD56
D
#1
#2
#3
CD41a
CD45
89.6%
7.18%
0.12%
0.19%
0.79%
89.6%
21.2%
37.5%
27.1%
E
% of cells
CD41⁻45⁺
CD41⁺45⁺
CD41⁺45⁻
F
Differentiation
day 13
Sorting CD41⁺45⁻
& CD41⁻45⁺
IV injection
CD41a
CD45
Survival (%)
CD41⁺45⁻ (n =5)
CD41⁻45⁺ (n =4)

Figure 3

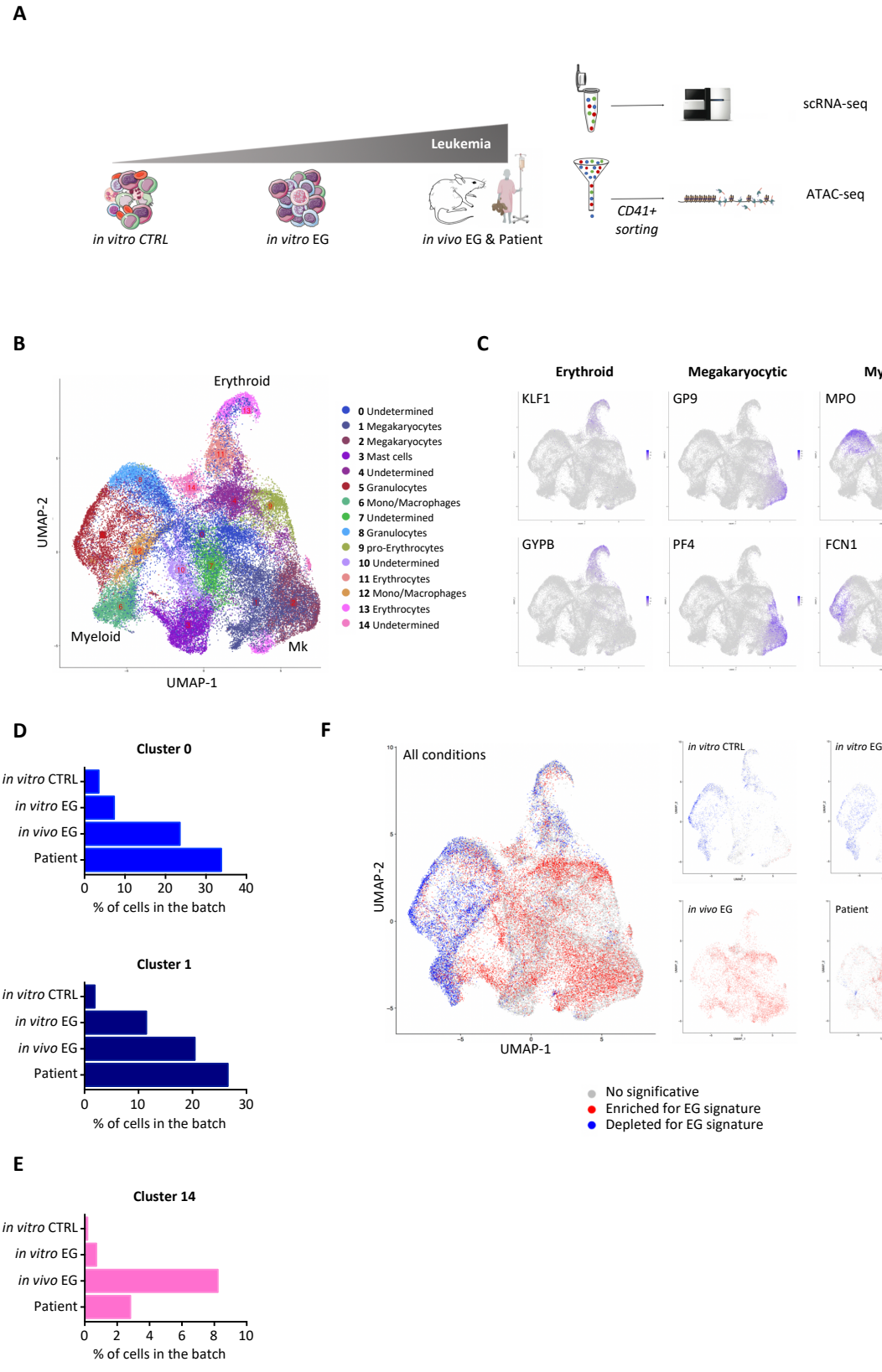

A
Leukemia
in vitro CTRL
in vitro EG
in vivo EG & Patient
scRNA-seq
CD41+ sorting
ATAC-seq
B
Erythroid
Myeloid
Mk
UMAP-2
UMAP-1
0 Undetermined
1 Megakaryocytes
2 Megakaryocytes
3 Mast cells
4 Undetermined
5 Granulocytes
6 Mono/Macrophages
7 Undetermined
8 Granulocytes
9 pro-Erythrocytes
10 Undetermined
11 Erythrocytes
12 Mono/Macrophages
13 Erythrocytes
14 Undetermined
C
Erythroid
Megakaryocytic
Myeloid
KLF1
GP9
MPO
GYPB
PF4
FCN1
D
Cluster 0
in vitro CTRL
in vitro EG
in vivo EG
Patient
0 10 20 30 40
% of cells in the batch
Cluster 1
0 10 20 30
% of cells in the batch
E
Cluster 14
0 2 4 6 8 10
% of cells in the batch
F
All conditions
UMAP-2
UMAP-1
in vitro CTRL
in vitro EG
in vivo EG
Patient
No significative
Enriched for EG signature
Depleted for EG signature

Figure 4

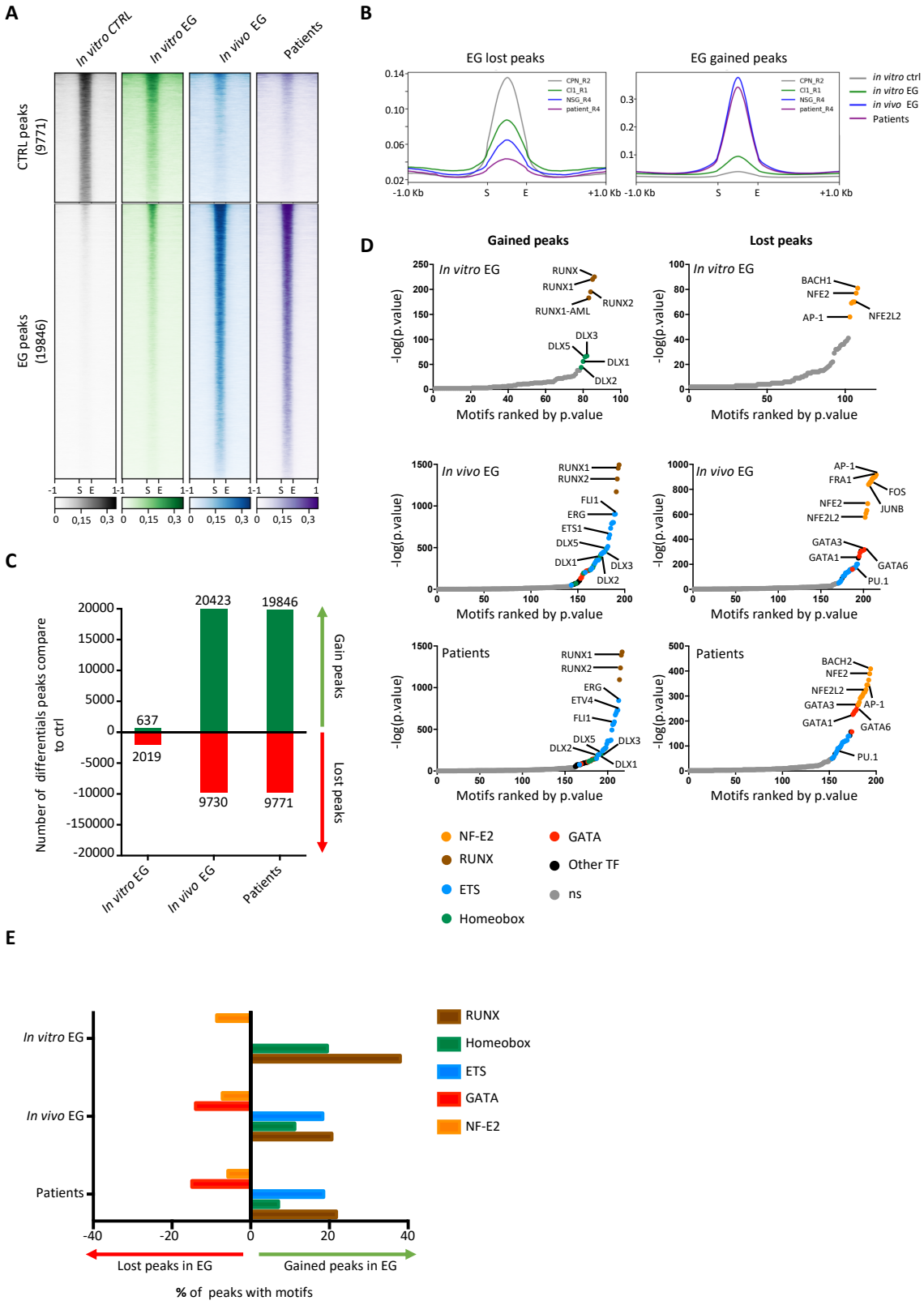

A
In vitro CTRL
In vitro EG
In vivo EG
Patients
CTRL peaks (9771)
EG peaks (19846)
B
EG lost peaks
EG gained peaks
in vitro ctrl
in vitro EG
in vivo EG
Patients
C
Number of differentials peaks compare to ctrl
20423
19846
637
2019
9730
9771
Gain peaks
Lost peaks
D
Gained peaks
Lost peaks
Motifs ranked by p.value
-log(p.value)
NF-E2
RUNX
ETS
Homeobox
GATA
Other TF
ns
E
RUNX
Homeobox
ETS
GATA
NF-E2
Lost peaks in EG
Gained peaks in EG
% of peaks with motifs

Figure 5

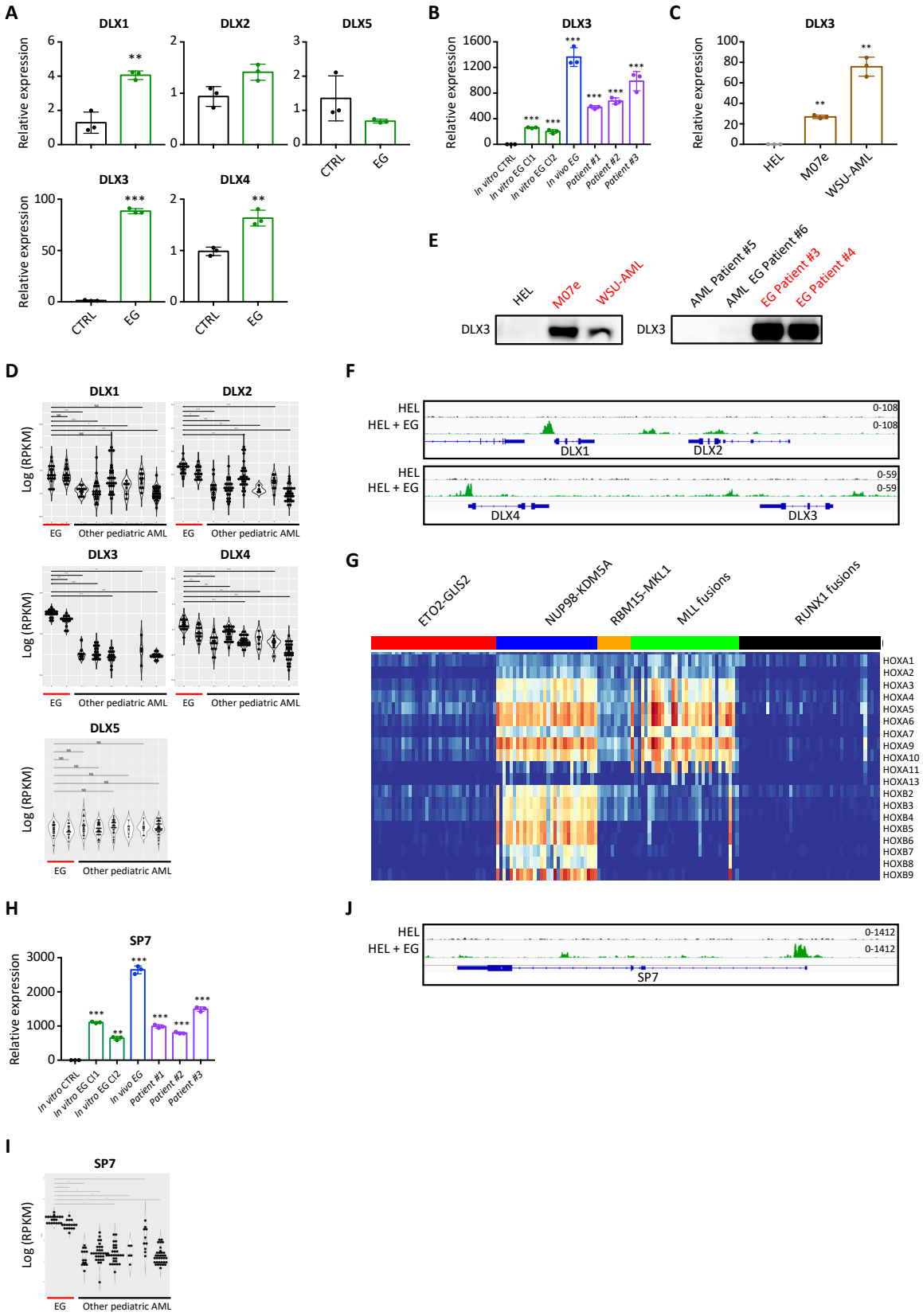

A
DLX1
DLX2
DLX5
DLX3
DLX4
Relative expression
CTRL
EG
B
DLX3
In vitro CTRL
In vitro EG Cl1
In vitro EG Cl2
In vivo EG
Patient #1
Patient #2
Patient #3
C
DLX3
HEL
M07e
WSU-AML
E
HEL
M07e
WSU-AML
AML Patient #5
AML EG Patient #6
EG Patient #3
EG Patient #4
DLX3
D
DLX1
DLX2
DLX3
DLX4
DLX5
Log (RPKM)
EG
Other pediatric AML
F
HEL
HEL + EG
DLX1
DLX2
DLX4
DLX3
G
ETO2-GLIS2
NUP98-KDM5A
RBM15-MKL1
MLL fusions
RUNX1 fusions
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
HOXB2
HOXB3
HOXB4
HOXB5
HOXB6
HOXB7
HOXB8
HOXB9
H
SP7
I
SP7
J
SP7

Figure 6

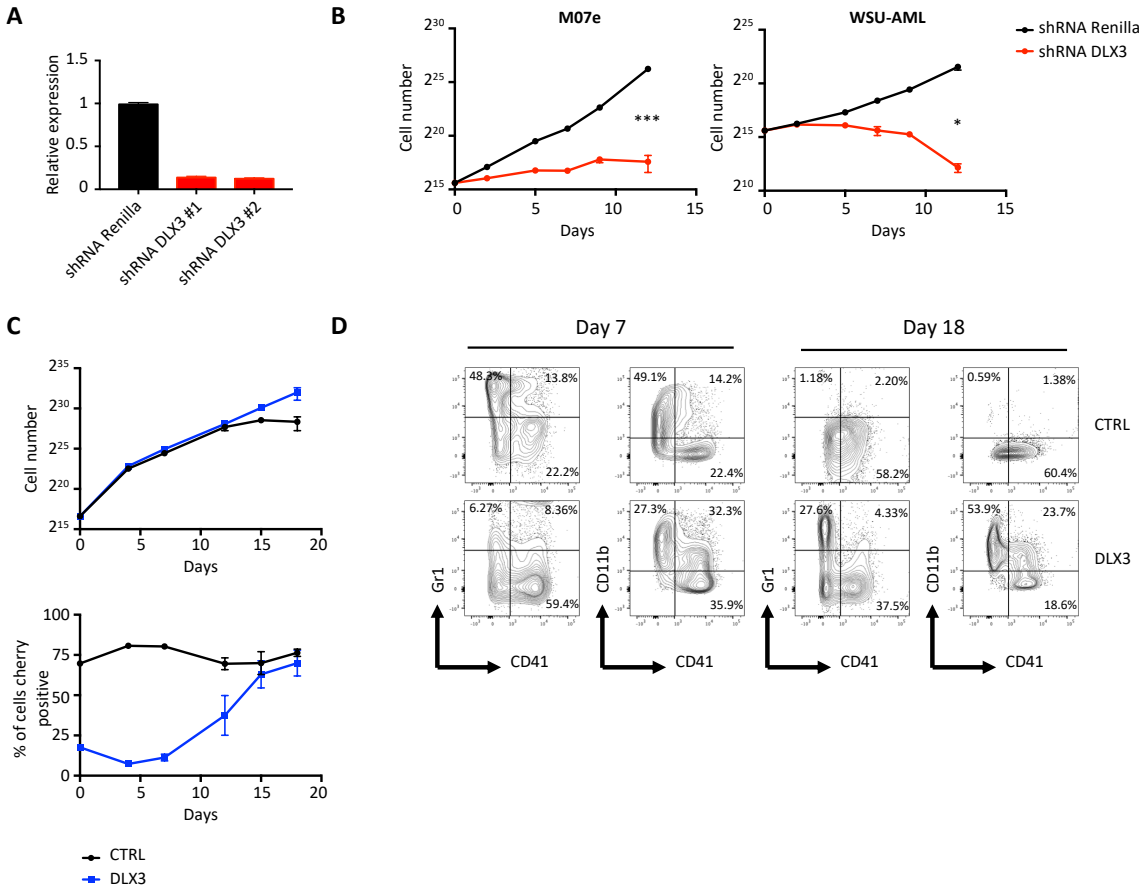

A
Relative expression
1.5
1
0.5
0
shRNA Renilla
shRNA DLX3 #1
shRNA DLX3 #2
B
M07e
WSU-AML
shRNA Renilla
shRNA DLX3
Cell number
Days
***
*
C
Cell number
Days
% of cells cherry positive
CTRL
DLX3
D
Day 7
Day 18
48.3%
13.8%
22.2%
49.1%
14.2%
22.4%
1.18%
2.20%
58.2%
0.59%
1.38%
60.4%
6.27%
8.36%
59.4%
27.3%
32.3%
35.9%
27.6%
4.33%
37.5%
53.9%
23.7%
18.6%
Gr1
CD11b
CD41
CTRL
DLX3

Figure 7

SUPPLEMENTAL FIGURES LEGENDS

Figure S1: Trisomy 21 in ETO2-GLIS2 patient.

(**A**) Proportion of ETO2-GLIS2 patients presenting with a trisomy 21 in leukemic cells. Three cohorts from the literature are represented (*250*, *251*, *254*).

Figure S2: ETO2-GLIS2 hematopoietic differentiation *in vitro* in trisomy 21 context.

(**A**) Schematic overview of iPSC hematopoietic differentiation protocol. (**B**) Representative flow cytometry analyses of cells obtained from trisomy 21 control (T21 CTRL) and trisomy 21 ETO2-GLIS2 (T21 EG) clones at day 13 of differentiation for the CD42b, CD235a and CD11b markers (gated on $CD43^+$ hematopoietic cells). (**C**) Percentage of cells for the markers showed in **B**. Mean +/- SD is indicated (T21 CTRL: n = 8, T21 EG: n = 8). (**D**) Average proportion of megakaryocytic (Mk: $CD43^+CD42b^+$), erythroid (Ery: $CD43^+CD235a^+$), myeloid (Myelo: $CD43^+CD11b^+$) cells as shown in **C**. (**E**) FACS plot of the CD235a erythroid marker for EG Cl1, EG Cl2 and T21 EG (gated on $CD42b^+$ megakaryocytic cells). (**F**) Representative flow cytometry analyses of T21 CTRL, and T21 EG at day 13 for combination of markers frequently found in ETO2-GLIS2+ AMKL patients (gated on $CD43^+$ hematopoietic cells). (**G**) Percentage of cells in the gates shown in **F**. Mean +/- SD is indicated (T21 CTRL: n = 8, T21 EG: n = 8). (**H**) Clonogenic activity in methylcellulose assays. 10 000 day 13 hematopoietic cells from T21 CTRL and T21 EG were plated then counted and replated every 10 days. Colonies from CTRL, EG Cl1 and EG Cl2 shown in **Fig2G** are also represented. Mean +/- SD is indicated (T21 CTRL: n = 4, T21 EG: n = 4). Statistical significance was determined using 1-tailed Mann-Whitney U test. *P < 0.05; **P < 0.01; ***P < 0.001.

Figure S3: Genetic and environmental context in ETO2-GLIS2 leukemic potential.

(**A**) Spleen weights at sacrifice of recipients engrafted with CTRL of EG iPSC-derived day 13 hematopoietic cells. (**B**) Phenotype at time of sacrifice of bone marrow and spleen cells from a representative mouse injected with cells derived from one ETO2-GLIS2 patient. (**C**) Survival curves of NSG-S mice injected with 0.7-1M day 13 iPSC-derived hematopoietic cells from CRISPR engineered ETO2-GLIS2 iPSC (inv16 EG, n = 20) or promoter CD43-ETO2-GLIS2 iPSC (pCD43-EG, n = 5) (*322*). (**D**) Survival curves of NSG-S mice injected with 0.7-1M day 13 iPSC-derived hematopoietic cells from Cl1 EG and Cl2 EG (D21 EG, n = 20) or T21 EG (n = 11). (**E**)

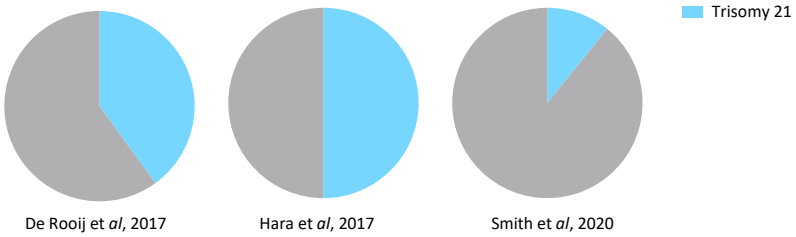
Survival curves of NSG-S (n = 20) or NSG (n = 6) mice injected with 0.7-1M day 13 iPSC-derived EG cells. (**F**) Flow cytometry analysis for CD117, CD56, CD42b, CD235a and CD11b markers for three mice presenting a variable CD41/45 immunophenotype. (**G**) Flow cytometry of the CD41/CD45 phenotypic variability resulting from $CD41^{+}45^{-}$ or $CD41^{-}45^{+}$ cell engraftment. Statistical significance was determined using log-rank Mantel-Cox test. *P < 0.05; **P < 0.01; ***P < 0.001.

Figure S4: Single cell RNA-seq characterisation of iPSC-derived and patient ETO2-GLIS2 cells.
(**A**) UMAP of the integration with cells colored according to the predicted cell-cycle stage (Seurat method). (**B**) Projection of cells from each data set (in red) on the UMAP of the integration. (**C**) Bar plots of the cell proportion in clusters for each condition. (**D**) Projection of the percentage of mitochondrial (left) and ribosomal (right) genes on the UMAP of the integration.

Figure S5: ATAC-seq analysis of the ETO2-GLIS2 cells.
(**A**) Number of reads after filtration for each replicate used for ATAC-seq. (**B**) Number of peaks for each replicate. Peak calling was performed using MACS2. (**C**) Principal Component Analyses (PCA) of ATAC-seq data. (**D** and **E**) Dot-plot representation of motifs under the differentials ATAC-seq peaks between Patients and *in vitro* EG (**D**) or Patients and in vivo EG (**E**). Dots are colored according to the transcription factor family to which they belong.

Figure S6: DLX factors in EG cells.
(**A**) *DLX1-6* expression in $CD41^{+}42^{-}$ CTRL, $CD41^{+}42^{+}$ CTRL, $CD41^{low}42^{low}$ $CD43^{prom}$EG and $CD41^{+}42^{+}$ $CD43^{prom}$EG cells (*322*). (B) *DLX1-6* mRNA expression in HEL and HEL EG cells (*263*). (**C**) Quantitative PCR for *DLX4* expression in CTRL and EG clones at day-13 hematopoietic cells, iPSC-derived EG leukemic mice and 3 different EG patients. Normalised on *RPLP0* expression. Mean +/- SD is indicated (n = 3 for all conditions). (**D**) Track depicting ChIP-seq for GFP in HEL-GFP and HEL-EG-GFP cells (same dataset as in **Fig 2F**) at *DLX5* and *DLX6* locus. (**E**) DLX3, HOXA1 and HOXB2 motifs (from JASPAR database). (**F**) *MEIS1* and *MEIS2* mRNA expression in pediatric AML (same dataset as in **Fig 2D**). Statistical significance was determined using DESeq2 differential expression analysis test. *P < 0.05; **P < 0.01; ***P < 0.001.

A



De Rooij et *al*, 2017 Hara et *al*, 2017 Smith et *al*, 2020

Figure S1

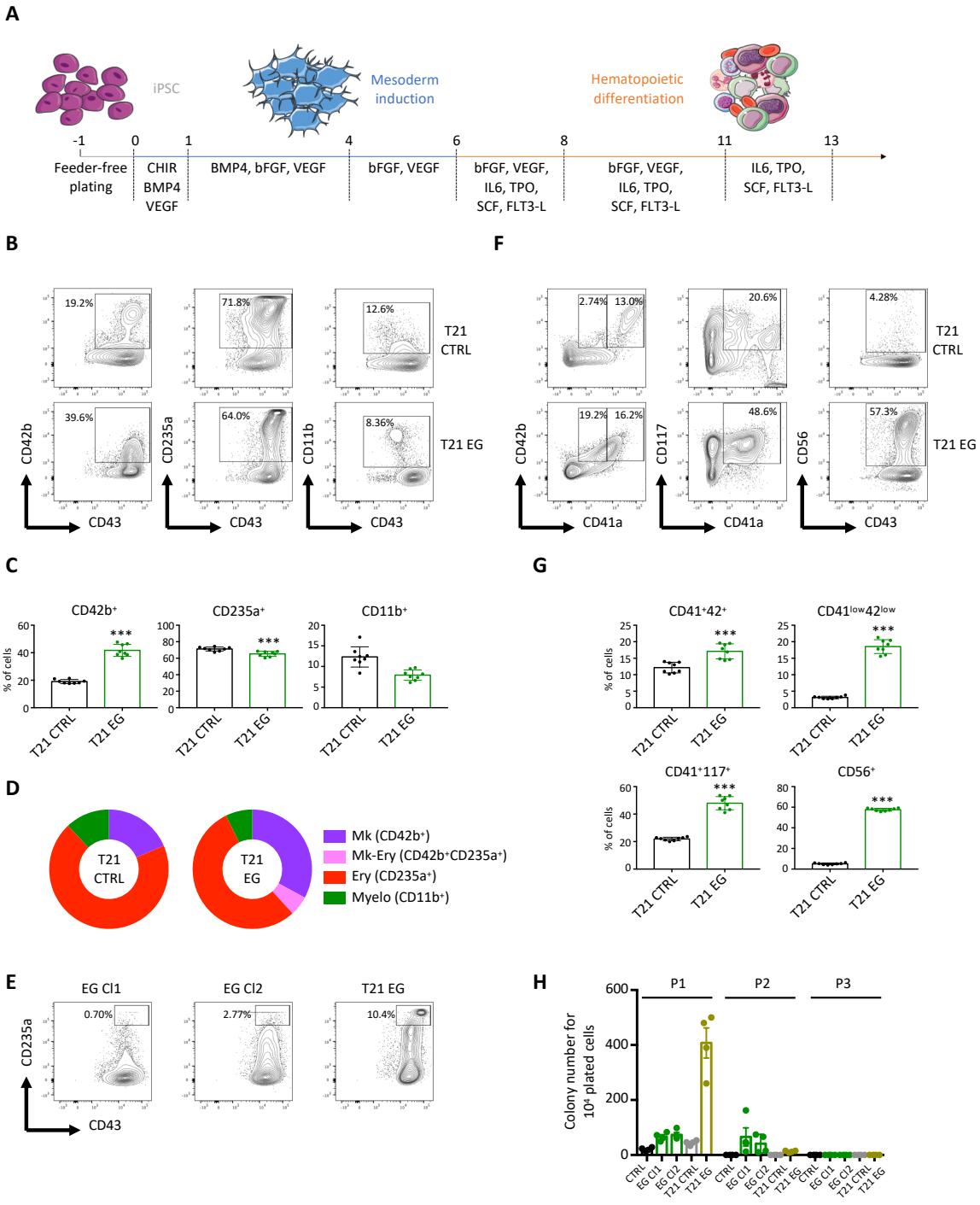

A
iPSC
Mesoderm induction
Hematopoietic differentiation
-1
0
1
4
6
8
11
13
Feeder-free plating
CHIR BMP4 VEGF
BMP4, bFGF, VEGF
bFGF, VEGF
bFGF, VEGF, IL6, TPO, SCF, FLT3-L
bFGF, VEGF, IL6, TPO, SCF, FLT3-L
IL6, TPO, SCF, FLT3-L
B
19.2%
71.8%
12.6%
39.6%
64.0%
8.36%
T21 CTRL
T21 EG
CD42b
CD235a
CD11b
CD43
F
2.74%
13.0%
20.6%
4.28%
19.2%
16.2%
48.6%
57.3%
CD42b
CD117
CD56
CD41a
CD43
C
CD42b⁺
CD235a⁺
CD11b⁺
% of cells
***
T21 CTRL
T21 EG
G
CD41⁺42⁺
CD41ˡᵒʷ42ˡᵒʷ
CD41⁺117⁺
CD56⁺
D
T21 CTRL
T21 EG
Mk (CD42b⁺)
Mk-Ery (CD42b⁺CD235a⁺)
Ery (CD235a⁺)
Myelo (CD11b⁺)
E
EG Cl1
EG Cl2
T21 EG
0.70%
2.77%
10.4%
CD235a
CD43
H
P1
P2
P3
Colony number for 10⁴ plated cells
CTRL
EG Cl1
EG Cl2
T21 CTRL
T21 EG

Figure S2

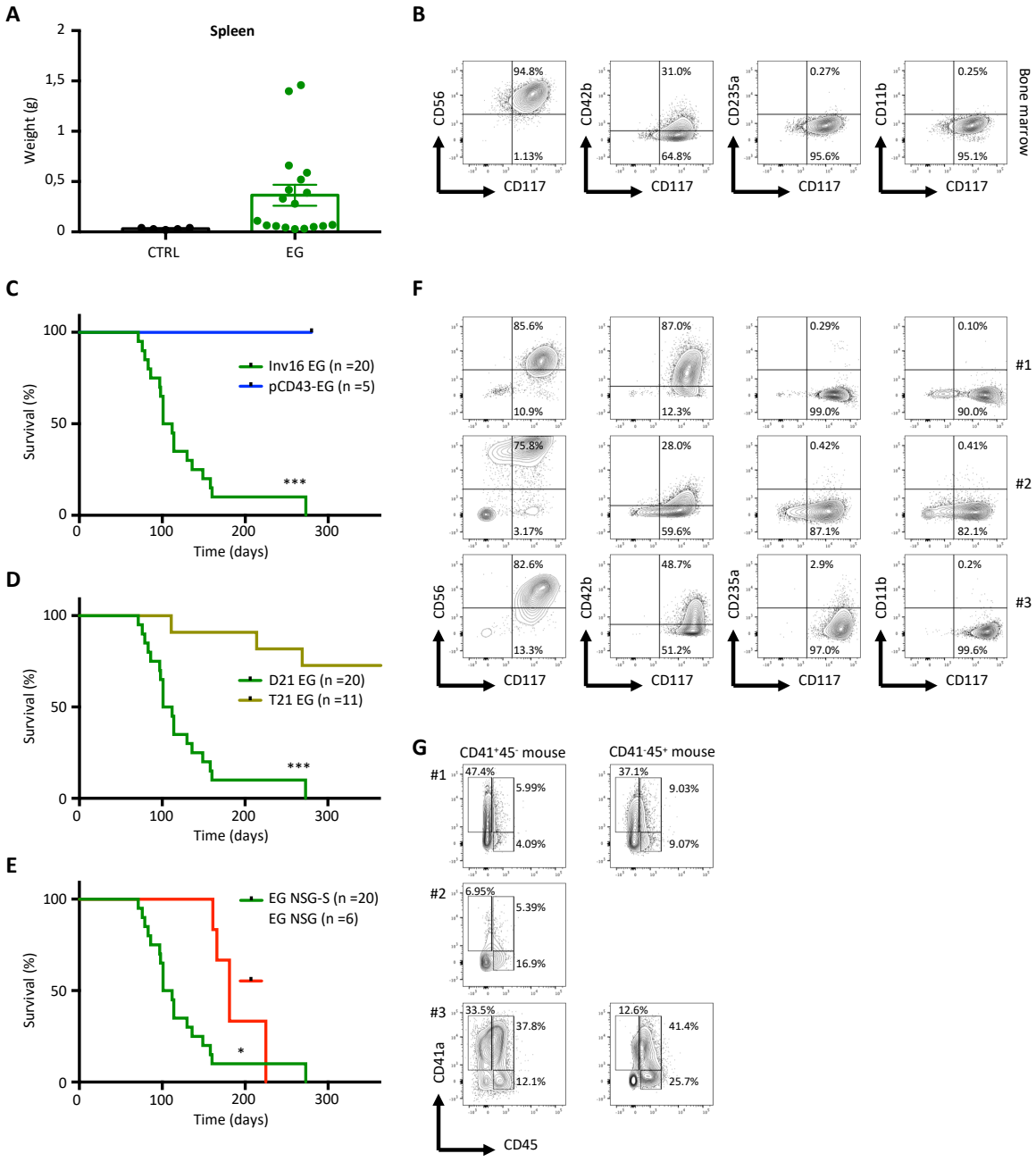

A
Spleen
Weight (g)
2
1,5
1
0,5
0
CTRL
EG
B
CD56
CD42b
CD235a
CD11b
CD117
94.8%
1.13%
31.0%
64.8%
0.27%
95.6%
0.25%
95.1%
Bone marrow
C
Survival (%)
Time (days)
Inv16 EG (n =20)
pCD43-EG (n =5)
***
D
D21 EG (n =20)
T21 EG (n =11)
***
E
EG NSG-S (n =20)
EG NSG (n =6)
*
F
85.6%
10.9%
87.0%
12.3%
0.29%
99.0%
0.10%
90.0%
#1
75.8%
3.17%
28.0%
59.6%
0.42%
87.1%
0.41%
82.1%
#2
82.6%
13.3%
48.7%
51.2%
2.9%
97.0%
0.2%
99.6%
#3
G
CD41+45- mouse
CD41-45+ mouse
#1
47.4%
5.99%
4.09%
37.1%
9.03%
9.07%
#2
6.95%
5.39%
16.9%
#3
33.5%
37.8%
12.1%
12.6%
41.4%
25.7%
CD41a
CD45

Figure S3

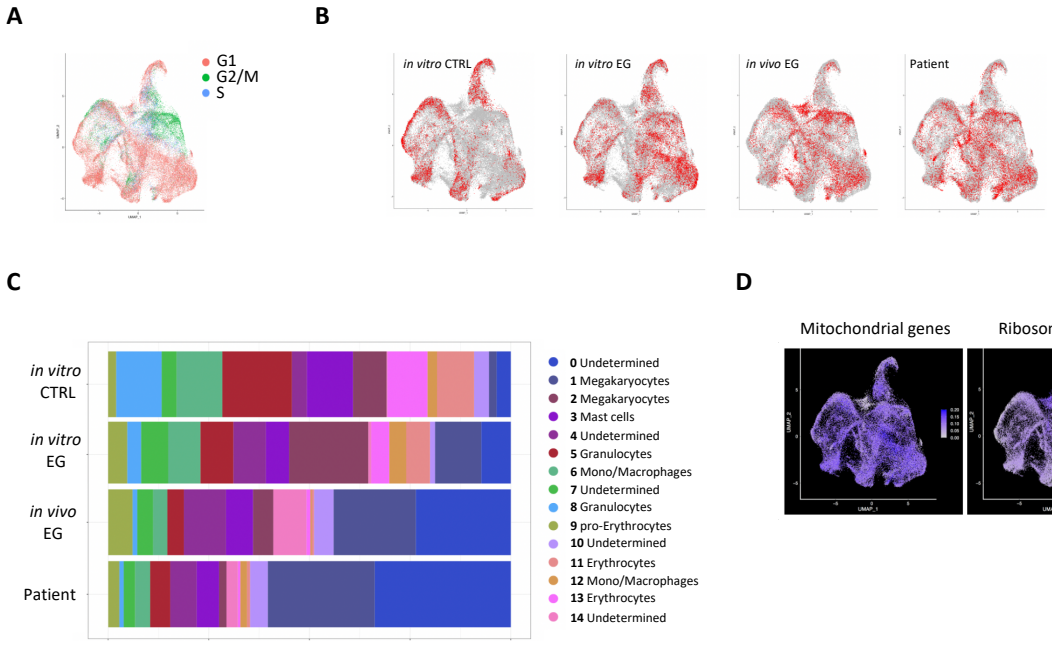

A
G1
G2/M
S
B
in vitro CTRL
in vitro EG
in vivo EG
Patient
C
in vitro CTRL
in vitro EG
in vivo EG
Patient
0 Undetermined
1 Megakaryocytes
2 Megakaryocytes
3 Mast cells
4 Undetermined
5 Granulocytes
6 Mono/Macrophages
7 Undetermined
8 Granulocytes
9 pro-Erythrocytes
10 Undetermined
11 Erythrocytes
12 Mono/Macrophages
13 Erythrocytes
14 Undetermined
D
Mitochondrial genes
Ribosomal genes

Figure S4

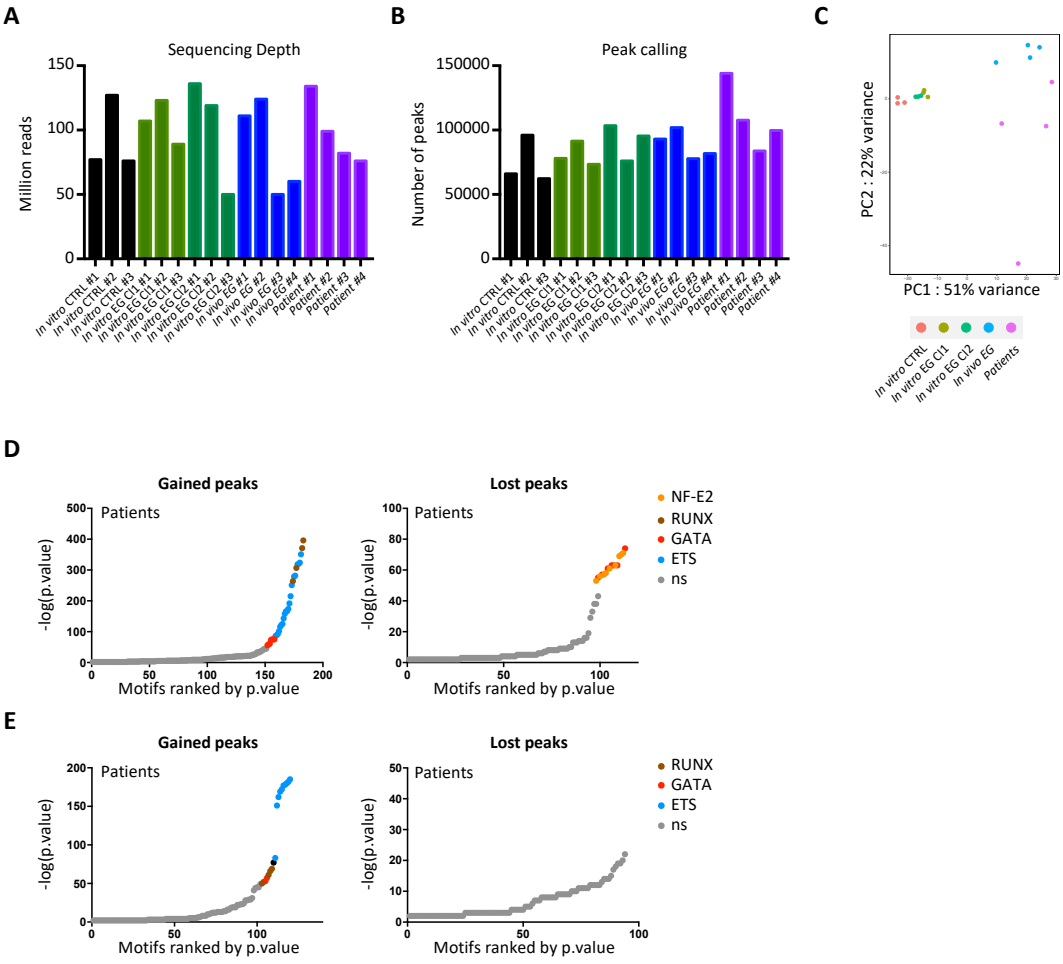

A
Sequencing Depth
Million reads
150
100
50
0
In vitro CTRL #1
In vitro CTRL #2
In vitro CTRL #3
In vitro EG Cl1 #1
In vitro EG Cl1 #2
In vitro EG Cl1 #3
In vitro EG Cl2 #1
In vitro EG Cl2 #2
In vitro EG Cl2 #3
In vivo EG #1
In vivo EG #2
In vivo EG #3
In vivo EG #4
Patient #1
Patient #2
Patient #3
Patient #4
B
Peak calling
Number of peaks
150000
100000
50000
0
C
PC2 : 22% variance
PC1 : 51% variance
In vitro CTRL
In vitro EG Cl1
In vitro EG Cl2
In vivo EG
Patients
D
Gained peaks
Patients
-log(p.value)
Motifs ranked by p.value
Lost peaks
NF-E2
RUNX
GATA
ETS
ns
E
Gained peaks
Lost peaks
RUNX
GATA
ETS
ns

Figure S5

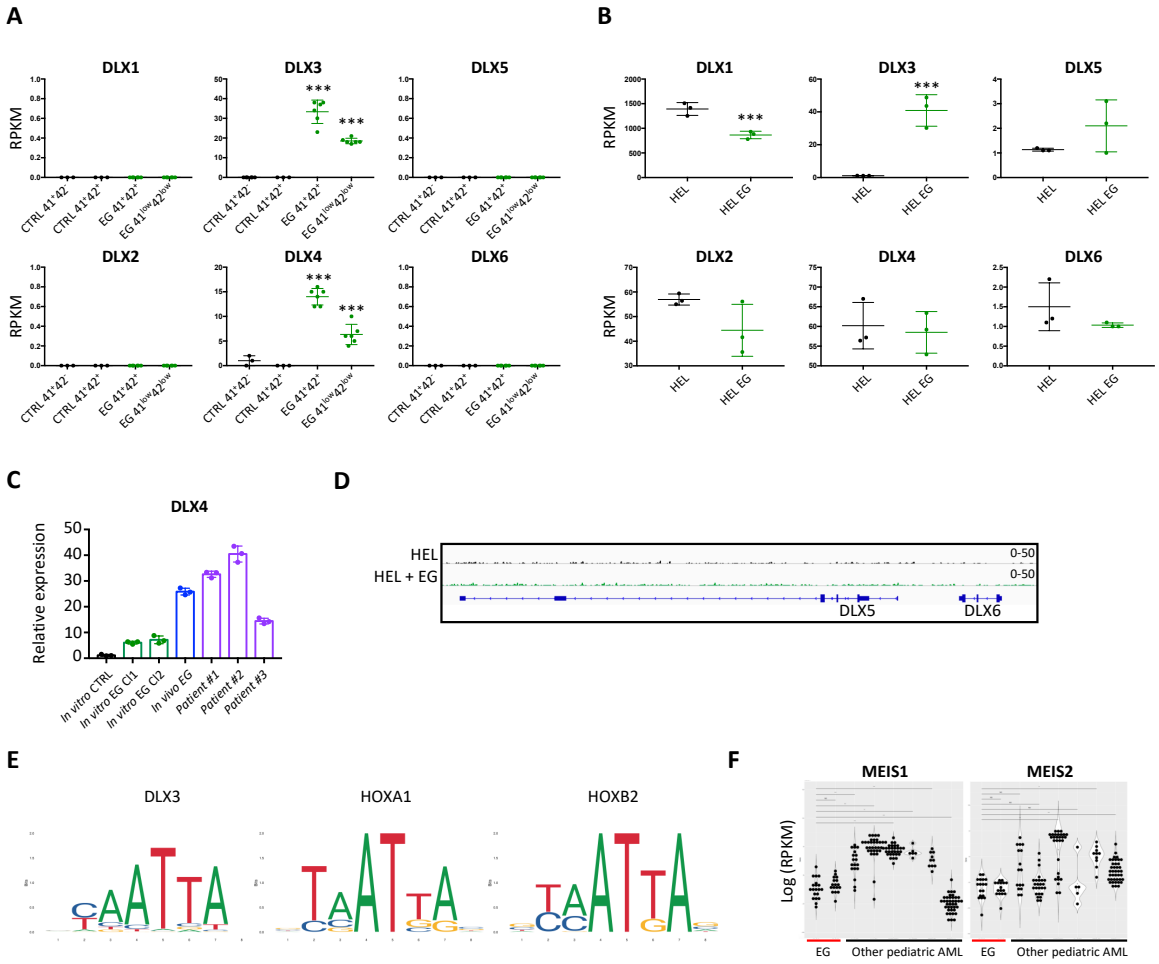

A
DLX1
DLX3
DLX5
DLX2
DLX4
DLX6
RPKM
CTRL 41+42+
CTRL 41+42-
EG 41+42+
EG 41low42low
***
B
DLX1
DLX3
DLX5
DLX2
DLX4
DLX6
RPKM
HEL
HEL EG
***
C
DLX4
Relative expression
In vitro CTRL
In vitro EG Cl1
In vitro EG Cl2
In vivo EG
Patient #1
Patient #2
Patient #3
D
HEL
HEL + EG
0-50
0-50
DLX5
DLX6
E
DLX3
HOXA1
HOXB2
F
MEIS1
MEIS2
Log (RPKM)
EG
Other pediatric AML

Figure S6

Discussion

Au cours de ma thèse, j'ai développé de nouveaux modèles humains, à l'aide de cellules pluripotentes induites, pour étudier les mécanismes de leucémogenèse induits par la fusion ETO2-GLIS2. J'ai montré que les cellules hématopoïétiques exprimant ETO2-GLIS2 issues de ces iPSC présentent des altérations de la différenciation et qu'elles conduisent à des leucémies phénotypiquement similaires aux patients. Par la suite, j'ai utilisé ces modèles afin de disséquer la dynamique moléculaire de la transformation leucémique par la fusion ETO2-GLIS2 et mettre en évidence une progression dans les facteurs impliqués entre les étapes précoces et tardives de la leucémogenèse.

Les résultats obtenus fournissent des observations permettant de progresser sur les méthodes de modélisation des cancers pédiatriques et sur la caractérisation de processus de transformation incluant le rôle de l'ontogénie, du microenvironnement, des mutations coopératrices et des altérations épigénétiques/transcriptionnelles dans la leucémogenèse induite par la fusion ETO2-GLIS2.

1) De l'importance d'avoir un modèle fidèle

Durant mes travaux de thèse, j'ai généré deux modèles différents pour la fusion ETO2-GLIS2. Le premier permet l'expression de la fusion au locus AAVS1 sous contrôle du promoteur CD43 (pCD43-EG) alors que le second a été généré par inversion du chromosome 16 par CRISPR-Cas9 (inv16 EG). Bien que ces deux modèles conduisent à des altérations de la différenciation hématopoïétique *in vitro*, seules les cellules inv16 EG induisent une leucémie *in vivo* lorsqu'elles sont greffées dans des souris immunodéficientes. Ceci est un point majeur car jusqu'à aujourd'hui, seules deux études rapportent le développement de leucémie à partir d'iPSC, les deux pour des fusions MLL (*310*, *313*). De plus cela soulève la question de la raison pour laquelle deux modèles issus d'iPSC provenant du même donneur (et donc ayant le même contexte génétique), présentent cette différence essentielle.

Plusieurs hypothèses peuvent être formulées. Premièrement, le niveau d'expression est probablement différent entre les cellules inv16 EG où la fusion est régulée par le promoteur de ETO2 et par les éléments régulateurs à proximité, et les cellules pCD43-EG (cela reste à confirmer par des expériences de PCR quantitative). Ce niveau d'expression pourrait être

déterminant dans le phénotype et les capacités transformantes de la fusion. Dans le cadre des LAM, il a été montré que des niveaux d'expression différents de l'oncogène de fusion MLL-AF9 conduisent à des changements importants de phénotype. En effet, lorsque MLL-AF9 est surexprimé via une transduction lentivirale dans des progéniteurs myéloïdes murins, cela conduit à une augmentation de leurs capacités clonogéniques et prolifératives ainsi qu'au développement de leucémie *in vivo*. A l'inverse, le Knock-In de cette fusion au locus endogène induit un niveau d'expression beaucoup plus faible, associé à de faibles capacités prolifératives et clonogéniques et une absence de transformation leucémique *in vivo* (*330*). Pour ETO2-GLIS2, nous avons précédemment montré *in vitro* chez la souris qu'un niveau faible de fusion maintient un phénotype myéloïde dans les cellules alors qu'un niveau plus élevé induit un phénotype mégacaryocytaire. Ces résultats sont en accord avec le niveau d'expression du transcrit de fusion qui est jusqu'à 1000 fois plus élevé chez les patients ETO2-GLIS2 LAM7 (phénotype mégacaryocytaire) que chez les patients ETO2-GLIS2 LAM5 (phénotype myéloïde) (*264*). Le niveau d'expression est donc déterminant pour le phénotype induit par la fusion.

Une autre possibilité est que le profil d'expression de la fusion, lié au contexte cellulaire, est différent entre les 2 modèles. En effet, ETO2 est exprimé précocement dans la différenciation hématopoïétique, dans les HSC et les progéniteurs. En revanche, certaines études suggèrent que le CD43 n'est pas exprimé dans des cellules souches mais seulement dans des progéniteurs plus engagés (*334*). Bien que de « vraies » HSC fonctionnelles ne soient pas générées à partir d'iPSC par le protocole de différenciation utilisé, cela reflète une différence dans la cinétique d'expression entre ETO2 et CD43 qui pourrait indiquer que la cible préférentielle de la fusion n'est pas atteinte dans le modèle pCD43-EG.

Un autre point particulièrement intéressant concerne l'ontogénie de cette leucémie. En effet, grâce à notre modèle murin inductible à la doxycycline, l'équipe a précédemment montré que les cellules fœtales murines sont beaucoup plus susceptibles à la transformation que les cellules adultes (*264*). L'utilisation d'iPSC nous offre la possibilité de tester cette observation chez l'homme. En effet, la différenciation des iPSC récapitule une hématopoïèse pré-définitive de type embryonnaire. De plus, dans le cadre d'une collaboration avec l'équipe de Françoise Pflumio, nous avons pu tester le pouvoir transformant de la fusion ETO2-GLIS2 dans des cellules issues de foie fœtal et de sang de cordon ombilical humains. Pour cela, ces cellules

ont été transduites avec un lentivirus permettant l'expression d'un vecteur contrôle ou de la fusion ETO2-GLIS2 puis greffées dans des souris immunodéficientes. Ces expériences ont révélé que dans des contextes embryonnaire (iPSC) et fœtal humain, la fusion ETO2-GLIS2 est capable d'induire efficacement une leucémie. En revanche dans un contexte ontologique plus avancé, tel que le sang de cordon ombilical, la leucémogenèse est extrêmement inefficace (données non montrées de nos collaboratrices).
Ainsi, il semble exister une « barrière » développementale durant la transition fœtale/adulte qui désensibilise les cellules hématopoïétiques à la transformation par la fusion ETO2-GLIS2. Ceci pourrait s'expliquer par les différences transcriptionnelles et les réorganisations de la chromatine, qui modifient les réseaux régulateurs, et qui surviennent durant cette période (c.f. introduction).

Dans tous les cas, ces résultats soulignent l'importance d'utiliser des modèles les plus proches possibles de la situation des patients, à la fois du point de vue génétique mais aussi au niveau du contexte ontogénique, particulièrement dans le cadre de pathologies pédiatriques.

2) Importance du microenvironnement dans la transformation leucémique

Les cellules ETO2-GLIS2 dérivées d'iPSC présentent des différences dans leurs propriétés cellulaires et leurs caractéristiques moléculaires *in vitro* et *in vivo*. En effet, durant la différenciation *in vitro*, bien que les cellules ETO2-GLIS2 présentent un fort biais mégacaryocytaire, elles conservent en partie des capacités de différenciation terminale vers les lignages érythroïde et myéloïde comme le montre l'expression des marqueurs CD235a et CD11b respectivement. De plus, les tests de clonogénicité en méthylcellulose révèlent une augmentation de l'autorenouvellement mais de façon très limitée avec un repiquage des colonies pour 1 passage supplémentaire seulement par rapport au contrôle. A l'inverse, bien que certaines cellules ETO2-GLIS2 issues des souris leucémiques présentent encore un profil transcriptionnel associé aux lignages érythroïde ou myéloïde comme le révèlent les analyses de scRNA-seq, la maturation terminale de ces lignages est bloquée comme l'indiquent les analyses de cytométrie en flux qui montrent que les cellules n'expriment plus les marqueurs CD235a et CD11b. De plus, le fait que ces cellules conduisent au développement leucémique

lors de greffes secondaires révèle qu'elles présentent de fortes capacités d'auto-renouvellement. Ainsi les changements qui s'opèrent entre les cellules *in vitro* présentant un blocage incomplet de la maturation terminale et de faibles capacités d'auto-renouvellement, et les cellules *in vivo* qui présentent toutes les caractéristiques d'une leucémie, se déroulent dans la souris receveuse, soulevant la question du rôle de l'environnement cellulaire dans ce processus. Plusieurs mécanismes pourraient être impliqués.

Le rôle des cytokines dans la différenciation et les processus leucémiques est important. Cependant, bien qu'une diminution de la latence de la pathologie soit observée lorsque les souris receveuses secrètent des cytokines humaines (SCF, GM-CSF et IL3), une leucémie avec la même pénétrance se déclare également dans des souris non humanisées. Ainsi ces cytokines, importantes pour les différentes cellules myéloïdes, ne semblent pas essentielles dans le processus de leucémogenèse. La plupart des cytokines sont spécifiques d'une espèce mais certaines sont interspécifiques. Il est donc possible que des cytokines, autres que celles précédemment mentionnées, sécrétées par les cellules du microenvironnement murin, croisent la barrière des espèces et participent à la leucémogenèse. Le microenvironnement pourrait également être important au travers des contacts physiques qu'il procure aux cellules ETO2-GLIS2. En effet, les propriétés géométriques, la rigidité de la matrice et les forces mécaniques peuvent impacter l'hématopoïèse (*335*).

Une possibilité est que les cellules ETO2-GLIS2 elles-mêmes sécrètent des facteurs ayant une action auto/paracrine qui participerait au processus de transformation leucémique. En effet, les LAM sont capables de sécréter, ou d'induire la sécrétion par les cellules du microenvironnement, de nombreuses molécules, nécessaires à leur survie et leur prolifération (*336*). Les cellules ETO2-GLIS2 expriment un niveau élevé de *BMP2* (*252*, *253*). Les protéines BMP sont membres de la superfamille du TGF-beta et jouent de nombreux rôles dans les processus d'inflammation, de survie, de prolifération et de mort cellulaire. Ainsi, une étude récente montre que certaines LAM sont capables de remodeler leur niche afin de faciliter le développement leucémique via une sécrétion de BMP2. Il serait donc intéressant d'étudier le « sécrétome » des cellules ETO2-GLIS2, par exemple en dosant les peptides présents dans le plasma sanguin des patients et des souris utilisées pour les PDX.

Enfin une autre observation, déjà rapportée dans une précédente publication (*252*) et confirmée durant ma thèse, renforce un rôle du microenvironnement dans le développement leucémique. En effet, les souris immunodéficientes utilisées dans les expériences de xénogreffes ou de greffes de cellules issues d'iPSC développent de façon quasi systématique une paralysie des membres postérieurs. Cette paralysie, qui n'est pas observée pour d'autre type de LAM, s'explique par le fait que les cellules ETO2-GLIS2 viennent se loger préférentiellement dans les os de la colonne vertébrale, plutôt que dans d'autres os comme le fémur ou le tibia, comprimant ainsi les nerfs. Cette observation est intéressante car les os de la colonne sont des os plats contrairement au fémur et au tibia qui sont des os longs et des différences dans la composition et le potentiel de prise de greffe dans ces deux types d'os existent (*337*), suggérant un microenvironnement qui supporte de façon différente la leucémogenèse ETO2-GLIS2.

3) Mutations additionnelles

Les LAM7 ETO2-GLIS2 ne présentent que peu de mutations additionnelles identifiées à ce jour. Cependant, une altération chromosomique récurrente est la trisomie 21, retrouvée en proportion variable selon les cohortes rapportées, jusqu'à 50% des patients (*250*, *251*, *254*). La présence de la T21 et la répression de GATA1 rappelle des éléments de la DS-LAM7 suggérant que certains mécanismes sous-jacents pourraient être partagés entre ces deux types de LAM7. De plus, la présence d'un troisième chromosome 21 est cohérente avec la surexpression du gène ERG, porté par ce chromosome, ainsi qu'avec une potentielle implication de RUNX1, dont les motifs sont retrouvés en ATAC-seq, également localisé sur le chromosome 21.

La modélisation de cette anomalie en plus de ETO2-GLIS2 dans des iPSC conduit à un phénotype légèrement différent des cellules disomiques. En effet, les T21 EG présentent un biais érythroïde important, cohérent avec l'effet déjà connu de la trisomie (*234*) et surtout une forte augmentation de la prolifération *in vitro*. Cependant, de façon surprenante, ces cellules conduisent au développement leucémique de façon beaucoup moins efficace que les cellules disomiques.

Une possible explication est liée au fond génétique. En effet, les cellules disomiques et trisomiques 21 utilisées pour générer les iPSC ETO2-GLIS2 ne proviennent pas du même donneur et ne peuvent donc pas être strictement comparées. Ainsi, il est possible que le donneur T21 utilisé présente des anomalies dans son génome qui inhibe la prise de greffe leucémique. De plus, je n'ai pu dériver qu'un seul clone T21 EG. Il est donc possible qu'un effet lié à une dérive clonale empêche une leucémogenèse efficace. Il convient donc de générer de nouveaux clones à partir de donneurs différents afin de valider ou non cet effet inhibiteur sur la transformation liée à la T21.

Une autre possibilité est que le biais érythroïde induit par la T21 entre en conflit avec l'effet de la fusion. En effet, cette augmentation de l'érythropoïèse conduit à une augmentation d'expression de facteurs associés à ce lignage, comme GATA1. Or, l'équilibre entre l'augmentation de ERG et la répression de GATA1 est essentiel pour la transformation leucémique (c.f. infra). Ainsi il est possible que dans ce système iPSC, la présence de la T21 induise un contexte défavorable pour la leucémogenèse. Cette hypothèse n'est pas incompatible avec une potentielle coopération chez les patients car il a été montré que selon le stade développement, l'impact de la T21 sur la différenciation n'est pas le même. Ainsi dans une hématopoïèse pré-définitive à partir d'iPSC, la T21 induit un biais érythroïde qui pourrait inhiber la transformation (*234*). A l'inverse, dans une hématopoïèse fœtale où la T21 est rapportée comme augmentant à la fois l'érythropoïèse et la mégacaryopoïèse (*235*), elle pourrait contribuer à la transformation leucémique.

Il est intéressant de noter qu'une autre étude, dans laquelle les auteurs modélisent des DS-LAM7 dans des cellules de foie fœtal humain, montre également une prise de greffe plus faible à partir de cellules trisomiques pour le chromosome 21 que à partir de cellules disomiques (*237*).

A l'heure actuelle, aucune étude de séquençage du génome complet rapportant l'ordre d'acquisition des mutations entre la fusion ETO2-GLIS2 et la T21 n'a été publiée. Comme tous les patients EG ne présentent pas de T21 et que la fusion est suffisante pour induire seule la leucémogenèse, il semblerait logique que la T21 constitue une mutation secondaire, acquise dans certains cas seulement. Cependant, au moins un cas de patient présentant une T21 constitutive et ayant acquis secondairement la fusion ETO2-GLIS2 a été publié (*323*). Ainsi, la T21 pourrait dans certains cas, à l'instar des DS-LAM7 où elle est nécessaire pour l'acquisition

de la mutation GATA1s mais dispensable pour la progression leucémique ultérieure, « faciliter » l'acquisition de la fusion ETO2-GLIS2 en induisant une forte prolifération cellulaire.

4) Caractérisation des réseaux de régulation transcriptionnelle impliqués dans la leucémogenèse par ETO2-GLIS2

L'une des forces du modèle iPSC est la possibilité d'étudier les changements à différents stades du développement leucémiques (*318*). Ainsi, j'ai réalisé des expériences de sc-RNAseq et d'ATAC-seq afin d'étudier les changements transcriptionnels et chromatiniens qui surviennent durant les étapes précoces de la leucémogenèse *in vitro* et lorsque la maladie est pleinement développée *in vivo*.

a) Intérêt du scRNA-seq pour l'étude des LAM

Les travaux précédents de l'équipe chez la souris montrent que, *in vitro*, les HSC donnent un phénotype mégacaryocytaire (plus ou moins pur selon que l'on se place dans un contexte fœtal ou adulte) et que plus on « descend » dans la hiérarchie hématopoïétique vers les MPP4, plus le phénotype devient myéloïde. Ces données suggèrent que les cellules cibles de la fusion ETO2-GLIS2 dans le cadre des leucémies de phénotype mégacaryocytaire LAM7 sont les HSC (*264*).

Dans le modèle iPSC, les blastes leucémiques issus des souris greffées à partir des cellules provenant des différenciations hématopoïétiques, présentent une certaine plasticité dans l'expression des marqueurs CD41 et CD45. Cela suggère que, chez l'homme, la gamme de progéniteurs cibles de la fusion est plus large que chez la souris, allant possiblement de cellules très immatures jusqu'à des progéniteurs mégacaryocytaires plus engagés. Cette hypothèse est soutenue par les résultats de scRNA-seq qui confirment que plusieurs populations leucémiques, présentant des profils d'expression différents, existent dans les souris immunodéficientes et chez les patients. De plus, cette observation est en phase avec d'autres publications montrant que les leucémies peuvent parfois contenir des cellules formant un continuum le long de la différenciation plutôt qu'une population bloquée à un stade spécifique (*211*).

Les données de scRNA-seq ont également permis de vérifier que les cellules ETO2-GLIS2 issues du modèle iPSC présentent bien un transcriptome similaire à celui des patients. Pour cela, nous avons développé un outils bio-informatique qui permet, pour une liste de gènes donnée, de calculer un score qui permet de dire si une cellule est significativement enrichie pour l'expression de ces gènes ou non. Ainsi, nous avons défini une liste de gènes qui constitue la signature transcriptomique des LAM ETO2-GLIS2 en regardant les gènes différentiellement exprimés entre les patients ETO2-GLIS2 et les autres patients LAM pédiatriques dans une large cohorte (*192*).

Un autre aspect intéressant qui ressort de l'analyse de scRNA-seq est la présence d'un cluster (ou plusieurs si on modifie les paramètres de clustering afin d'augmenter davantage la séparation des populations) dans lequel on ne trouve presque exclusivement que des cellules ETO2-GLIS2 issues des NSG ou des patients. Cela est particulièrement intriguant car, en visualisation UMAP, ce cluster se localise « entre » les branches normales de la différenciation et n'expriment aucun gène associé aux cellules hématopoïétiques. En effet, ce cluster est caractérisé par un très faible taux de gènes mitochondriaux et à l'inverse une forte expression de gènes liés au métabolisme ribosomal. Ce profil transcriptionnel a déjà été identifié dans une analyse de scRNA-seq d'échantillons de patients LAM et associé à des LAM immatures (*325*) et une augmentation des gènes ribosomaux est également détectée dans les stades leucémiques plus avancés dans un autre modèle iPSC de LAM (*318*).

b) Evolution epigénétique (Epimutation) au cours de la leucémogenèse

Nous avions précédemment montré que, dans des lignées cellulaires et des cellules de patients, les facteurs de transcription ERG et GATA1 jouent un rôle central dans l'action de ETO2-GLIS2, leur expression étant augmentée et diminuée respectivement (*263*). Mes résultats confirment l'importance de ces facteurs et l'analyse de motifs pourrait indiquer une implication d'autres membres des deux familles ETS et GATA. Cependant, de façon surprenante, ces changements ne surviennent pas immédiatement mais seulement lors de la transition qui va aboutir à une leucémie pleinement développée. Ainsi, l'augmentation de l'activité des facteurs ETS et la diminution des facteurs GATA, semblent être les éléments clés

permettant d'obtenir une leucémie ETO2-GLIS2. Ceci est cohérent avec le rôle des facteurs de ces familles dans la différenciation mégacaryocytaire. En effet, ERG est important pour les progéniteurs mégacaryocytaires immatures tandis que GATA1 participe à la différenciation terminale.

Une des découvertes surprenantes de ces expériences d'ATAC-seq est la présence de deux autres familles de facteurs de transcription, actives durant les étapes précoces de la leucémogenèse. La première est la famille RUNX. RUNX1 est un régulateur majeur de l'hématopoïèse impliqué dans le développement hématopoïétique, la biologie des cellules souches et dans de nombreux lignages (*338*). RUNX1 a été rapporté comme muté dans des LAL, des SMD, dans les syndromes plaquettaires familiaux de prédisposition leucémique et dans des LAM. De plus, plusieurs fusions oncogéniques impliquant RUNX1 sont associées aux leucémies, comme les fusions RUNX1-ETO ou ETV6-RUNX1. Des dérégulations de RUNX1 sans mutation sont également fréquentes dans plusieurs leucémies comme par exemple les LAM associées aux mutations FLT3-ITD ou certaines fusions MLL, et sont associées à un mauvais pronostic (*339*). Dans le cadre des LAM7, une prépublication récente rapporte une dérégulation entre les différentes isoformes de RUNX1 qui synergise avec les mutations perte de fonction GATA1s afin d'induire la leucémogenèse dans les DS-LAM7 (*340*). Le rôle de RUNX1 dans les leucémies ETO2-GLIS2 n'avait jusqu'à présent pas été étudié, cependant, comme l'expression de GATA1 est fortement diminué dans les LAM7 ETO2-GLIS2, un mécanisme de dérégulation de RUNX1 similaire à celui des DS-LAM7 pourrait être impliqué.

La seconde famille de facteurs impliquée dans l'initiation de la leucémogenèse est la famille des DLX. Les gènes DLX sont des facteurs de transcription à homéodomaine. Dans l'hématopoïèse, les rôles des gènes homéotiques HOXA et HOXB en conditions normales, pour la régulation des HSC et la différenciation, et pathologiques avec des mutations et des dérégulations dans de nombreuses LAM sont étudiés depuis longtemps (*341*). A l'inverse, très peu de choses sont connues à propos du rôle des facteurs DLX. En effet, très peu d'articles, et aucun chez l'homme, ne rapporte de rôle des DLX dans l'hématopoïèse normale ou dans les leucémies. La famille DLX comprenant 6 membres, j'ai pu confirmer que, DLX4, mais surtout DLX3 sont fortement surexprimées dans les cellules ETO2-GLIS2. De façon intéressante, contrairement à d'autres facteurs comme ERG ou GATA1 qui sont impliqués dans différentes

LAM, DLX3 est très spécifique des leucémies ETO2-GLIS2 car il n'est pas exprimé dans les autres types de leucémies pédiatriques.

DLX3 participe normalement au développement du tissu osseux (*326*) où il interagit notamment avec le cofacteur SP7 (*327*). De façon intéressante, SP7 est également fortement et spécifiquement surexprimé dans les cellules ETO2-GLIS2. Ainsi, ETO2-GLIS2 pourrait « détourner » ce programme de développement osseux impliquant les DLX3/4 et SP7 et l'activé dans les cellules hématopoïétiques durant les étapes précoces de la leucémogenèse. Cette hypothèse est renforcée par le fait que des motifs RUNX sont retrouvés dans les cellules ETO2-GLIS2 ainsi que par les niveaux élevés de certains facteurs BMP. En effet, les facteurs RUNX2 et BMP2 sont important pour le développement osseux.

Les conséquences de cette activation aberrante reste encore à déterminer. Nous sommes actuellement en train d'étudier les gènes associés aux pics comportant les motifs DLX afin d'identifier les processus dans lesquels ils sont impliqués. De plus il serait pertinent de réaliser des expériences de ChIP-seq (ou de ces dérivés plus récent, CUT&RUN et CUT&TAG) afin de déterminer les sites de fixation de DLX3 et DLX4. Cela permettrait de trancher entre 2 hypothèses : Soit DLX3/4 se lient sur les mêmes sites que la fusion et renforcent ainsi son action, soit ils se lient sur des sites propres contrôlant une partie du programme transcriptionnel de la fusion.

Un autre point intéressant est la manière dont la fusion ETO2-GLIS2 est capable d'induire ce réarrangement de la chromatine lors de la transition *in vivo*. En effet, l'organisation 3D de la chromatine dépend entre autres des protéines CTCF qui se lient sur le génome à des sites précis et servent de barrières physiques afin de délimiter les TAD, qui sont des zones de co-régulation sur le génome. De façon intéressante, GLIS2 présente un motif de liaison à l'ADN très similaire à celui de CTCF. Ainsi, il est possible que ETO2-GLIS2 puisse entrer en compétition avec CTCF et l'empêcher de se lier a certaines de ses localisations entrainant un changement dans les frontières des TAD et donc une réorganisation du génome. Des expériences de capture de la chromatine sont nécessaires afin de valider cette hypothèse.

L’ensemble de ces résultats pourrait suggérer un modèle dans lequel la progression leucémique ne se déroule pas par une accumulation de mutations additionnelles mais via une succession d’évènement épigénétiques induits par la fusion ETO2-GLIS2. La capacité de ETO2-GLIS2 d’induire ou non une leucémie serait ainsi dépendante du contexte chromatinien de la cellule dans laquelle la fusion survient. Les différences épigénétiques survenant durant le développement, expliquerait ainsi la permissivité plus ou moins importante des cellules à la transformation leucémique.

Annexes

JCI The Journal of Clinical Investigation

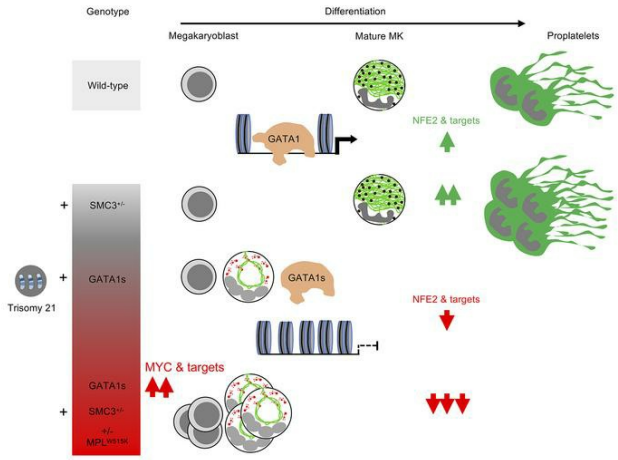
Stepwise GATA1 and SMC3 mutations alter megakaryocyte differentiation in a Down syndrome leukemia model

Brahim Arkoun, … , Thomas Mercher, William Vainchenker

J Clin Invest. 2022;132(14):e156290. https://doi.org/10.1172/JCI156290.

Research Article Hematology Oncology

Graphical abstract



Find the latest version:

https://jci.me/156290/pdf

Stepwise GATA1 and SMC3 mutations alter megakaryocyte differentiation in a Down syndrome leukemia model

Brahim Arkoun,[1,2,3] Elie Robert,[3] Fabien Boudia,[3] Stefania Mazzi,[1] Virginie Dufour,[4] Aurélie Siret,[3] Yasmine Mammasse,[4] Zakia Aid,[3] Matthieu Vieira,[1] Imanci Aygun,[1] Marine Aglave,[5] Marie Cambot,[4] Rachel Petermann,[4] Sylvie Souquere,[6] Philippe Rameau,[7] Cyril Catelain,[7] Romain Diot,[8] Gérard Tachdjian,[8] Olivier Hermine,[2,9] Nathalie Droin,[1,10] Najet Debili,[1] Isabelle Plo,[1,2] Sébastien Malinge,[3,11] Eric Soler,[2,12] Hana Raslova,[1] Thomas Mercher,[3] and William Vainchenker[1,2]

[1]INSERM, UMR1287, Gustave Roussy, Université Paris–Saclay, Equipe Labellisée Ligue Nationale Contre le Cancer, Villejuif, France. [2]Laboratory of Excellence GReX, Université de Paris, Paris, France. [3]INSERM, UMR1170, Gustave Roussy, Université Paris–Saclay, Equipe Labellisée Ligue Nationale Contre le Cancer, OPALE Carnot Institute, PEDIAC consortium, Villejuif, France. [4]Institut National de la Transfusion Sanguine (INTS), Paris, France. [5]Gustave Roussy, Plateforme de Bioinformatique, [6]Gustave Roussy, Université Paris–Saclay, and [7]Gustave Roussy, Plateforme Imagerie et Cytométrie, Université Paris–Saclay, UMS AMMICA, INSERM US23, CNRS UMS 3655, Villejuif, France. [8]Assistance Publique-Hôpitaux de Paris, Service d'Histologie, Embryologie et Cytogénétique, Université Paris–Saclay, Hôpital Antoine Béclère, Clamart, France. [9]INSERM U1163/CNRS ERL8254–Laboratory of Cellular and Molecular Mechanisms of Hematological Disorders and Therapeutic Implications, Institut Imagine, Paris, France. [10]Gustave Roussy, Plateforme de Génomique, Université Paris–Saclay, UMS AMMICA, INSERM US23, CNRS UMS 3655, Villejuif, France. [11]Telethon Kids Cancer Centre, Telethon Kids Institute, University of Western Australia, Perth, Western Australia, Australia. [12]Institut de Génétique Moléculaire de Montpellier (IGMM), University of Montpellier, CNRS, Montpellier, France.

Acute megakaryoblastic leukemia of Down syndrome (DS-AMKL) is a model of clonal evolution from a preleukemic transient myeloproliferative disorder requiring both a trisomy 21 (T21) and a *GATA1*s mutation to a leukemia driven by additional driver mutations. We modeled the megakaryocyte differentiation defect through stepwise gene editing of *GATA1s*, *SMC3*$^{+/-}$, and *MPL*^{W515K}, providing 20 different T21 or disomy 21 (D21) induced pluripotent stem cell (iPSC) clones. *GATA1s* profoundly reshaped iPSC-derived hematopoietic architecture with gradual myeloid-to-megakaryocyte shift and megakaryocyte differentiation alteration upon addition of *SMC3* and *MPL* mutations. Transcriptional, chromatin accessibility, and GATA1-binding data showed alteration of essential megakaryocyte differentiation genes, including *NFE2* downregulation that was associated with loss of GATA1s binding and functionally involved in megakaryocyte differentiation blockage. T21 enhanced the proliferative phenotype, reproducing the cellular and molecular abnormalities of DS-AMKL. Our study provides an array of human cell–based models revealing individual contributions of different mutations to DS-AMKL differentiation blockage, a major determinant of leukemic progression.

Introduction

Acute megakaryoblastic leukemias (AMKLs) are rare disorders that are more frequent in children (1, 2). Children with Down syndrome (DS) who have constitutive trisomy 21 (T21) have an increased risk of developing AMKL (3). While children with DS-AMKL have unique phenotypic and clinical features, this condition also represents a model of multistep leukemogenesis with key characteristics found in many leukemias (4).

DS-AMKL is always preceded by a clonal preleukemic state called transient myeloproliferative disorder (TMD) driven by *GATA1* mutations, mostly in exon 2 and resulting in the expression of an N-terminal deleted shorter protein, called GATA1s (5, 6). Usually, TMD spontaneously regresses, but 10% to 20% of these children will develop DS-AMKL within 5 years (7). This leukemia originates from the initial *GATA1s* clone with 2 or more acquired additional mutations (8, 9), including loss-of-function mutations in cohesin complex genes (e.g., *STAG2*, *RAD21*, *SMC1A/3*) and a mutation in a gene encoding a signaling molecule leading to RAS or JAK/STAT pathway constitutive activation (8, 9). *EZH2* or *SUZ12*, 2 members of PRC2, are frequently mutated in addition to the other mutations.

How *GATA1s* and these secondary mutations cooperate to induce AMKL and why this leukemia development requires a T21 and a fetal hematopoiesis remain unclear and need modeling approaches. Mouse models have been limited by only minor defects induced by Gata1s in late megakaryocyte (MK) differentiation (10–13). Thus, it was crucial to develop human models of DS-AMKL. An approach consisting of introducing *GATA1s* in induced pluripotent stem cells (iPSCs) or deriving iPSCs from T21 individuals with and without TMD has been developed, but was mostly focused on the erythroid cells (14–17). A more recent approach has been to edit the *GATA1s* and *STAG2* mutations in primary human CD34$^+$ cells, including T21 CD34$^+$ fetal liver cells, followed by xenotransplantation (18, 19). The latter approach was focused on hematopoiet-

▶ Related Commentary: https://doi.org/10.1172/JCI161659

Authorship note: TM and WV contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.
Copyright: © 2022, Arkoun et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.
Submitted: November 1, 2021; **Accepted:** May 13, 2022; **Published:** July 15, 2022.
Reference information: *J Clin Invest*. 2022;132(14):e156290.
https://doi.org/10.1172/JCI156290.

ic stem and progenitor cells (HSCP) and has clearly unveiled that the cooperation between *GATA1s* and *STAG2* mutations induces self-renewal and progression toward AMKL.

In contrast to erythroid differentiation (20, 21), the precise role of GATA1s on human MK differentiation is incompletely understood (17). It appears essential to better understanding how GATA1s precisely cooperates with T21 and additional mutations to induce MK differentiation blockage and proliferation, two key features of the leukemic process. Based on the consequences of cohesin mutations on HSPC (22–24), we hypothesized that *SMC3* haploinsufficiency may modify the chromatin occupancy of GATA1s and other transcription factors regulating late MK differentiation. We modeled the effects of different mutations implicated in the progression of TMD to AMKL on human hematopoiesis and more particularly on the MK differentiation. For this purpose, we sequentially introduced *GATA1s*, a gain-of-function *MPL* mutation (MPL^{W515K}) (25, 26) and a heterozygous inactivation of *SMC3* ($SMC3^{+/-}$), using CRISPR-Cas9 technology in T21 iPSC. Moreover, to understand the role of T21, we contrasted disomic 21 (D21) with T21 iPSCs. This approach provides insights at the cellular and molecular levels on how GATA1s alters MK differentiation alone or in cooperation with SMC3 deficiency to deeply enhance differentiation alterations and increase proliferation.

Results

Sequential editing generates a panel of iPSC lines to investigate the different steps of DS-AMKL progression. To recapitulate DS-AMKL progression, we sequentially introduced by CRISPR/Cas9-mediated engineering *GATA1s* mutations (G), followed by an MPL^{W515K} mutation (M) and *SMC3* mutations (S), into the parental T iPSCs (Figure 1A).

First, we generated 2 Trisomy 21 GATA1s (TG) iPSC clones carrying *GATA1s* truncating mutations and confirmed that T MKs expressed both full-length (FL) and short GATA1 isoforms, while TG mutant MKs expressed only GATA1s (Supplemental Figure 1, A, B, and R; supplemental material available online with this article; https://doi.org/10.1172/JCI156290DS1). Thereafter, the MPL^{W515K} gain-of-function mutation was introduced in a TG clone to generate 4 Trisomy 21 GATA1s $MPL^{W515K/W515K}$ (TGM) clones (Supplemental Figure 1, C–F and S). To assess the functionality of the MPL^{W515K} mutation, we quantified the MK differentiation (% of $CD41^+CD42^+$ cells) from $MPL^{WT/WT}$ and $MPL^{W515K/W515K}$ iPSCs (Supplemental Figure 1, G and H). The MK differentiation obtained from the TGM in the presence of Stem Cell Factor (SCF) was Thrombopoietin (TPO) independent and abolished by ruxolitinib, a JAK1/JAK2 inhibitor, demonstrating that the MPL^{W515K} mutant was functional.

To assess the role of *SMC3* haploinsufficiency, we edited *SMC3* in the heterozygous state in T, TG, and TGM and we obtained 3 clones in each genetic context (TS, TGS, TGMS) (Supplemental Figure 1, I–K). *SMC3* haploinsufficiency was confirmed by Western blots (Supplemental Figure 1, L and M). Notably, intracellular flow cytometry analysis indicated a similar increase in GATA1 protein levels in MKs from TG and TGMS as compared with T (Supplemental Figure 1, N and O).

To understand the role of T21, we derived an isogenic D *GATA1s* MPL^{W515K} (DGM) clone by a spontaneous loss of a chromosome 21 (Supplemental Figure 1V). Then we generated 2 isogenic DGMS clones (Supplemental Figure 1, T and V). MKs from DGMS and TGMS presented similar levels of GATA1 protein (Supplemental Figure 1, P and Q).

All iPSC clones were genetically characterized by karyotypic analyses (Supplemental Figure 1, U and V), fluorescence in situ hybridization (Supplemental Figure 1W), and Sanger sequencing (no off-targets in CRISPR/Cas9-edited iPSC clones; Supplemental Table 1). We obtained a panel of 20 different iPSC clones, presenting functionally validated mutations in *GATA1*, *MPL*, and *SMC3* that allowed us to study the contribution of each mutation, including T21, and those mutations' cooperation in the progression to DS-AMKL.

GATA1s and $SMC3^{+/-}$ alter hematopoietic progenitors and MK yield. To investigate the hematopoietic differentiation of the different iPSC lines, we used a 2D cell culture with a sequential addition of cytokines (Figure 1B) and $CD34^+CD43^+$ hematopoietic progenitor cells (HPC) purified from the supernatant at day 13 of culture. While T, TS clones generated erythroid and myeloid colonies, TG clones showed increased myeloid, but abrogated erythroid, colony formation (Supplemental Figure 2, A and B), confirming previous data (16). The addition of $MPL^{W515K/W515K}$ and/or $SMC3^{+/-}$ had no significant effect (Supplemental Figure 2, A and B).

The MK colony-forming potential, the size of MK colonies, and the total number of generated $CD41^+CD42^+$ MKs were significantly increased in T21 only when *GATA1s* and $SMC3^{+/-}$ were combined (Figure 1, C–E). T and TS showed typical micro-MK (Figure 1F), characteristic for the iPSC-derived MKs (iMKs). However, all genotypes presenting GATA1s showed the presence of large-sized basophilic dysplastic multilobulated MKs (Figure 1F). The predominant classes of ploidy were the 2N and 4N, as observed for MK derived from WT human iPSCs or embryonic stem (ES) cells (27). GATA1s increased the mean percentage of more than 4N MKs in synergy with S, but not M (Figure 1, F and G). These data support the idea that GATA1s cooperates with *SMC3* haploinsufficiency in T21 to alter MK differentiation, including increased clonogenic potential and proliferation of MK progenitors with accumulation at an immature stage and increased size and ploidy in some immature MKs.

Effects of $SMC3^{+/-}$ and GATA1s on MK maturation. We then assessed cell-surface marker expression. While a fraction of $CD41^+$ cells were CD42 negative, as revealed by an anti-CD42a antibody, the vast majority expressed the CD42 from all genotypes (Figure 2, A and B), including an independent D21 iPSC derived from a healthy individual (Supplemental Figure 3, A and B). This phenotype matches what is observed in the MK differentiation of iPSCs or primary $CD34^+$ cells. However, MFI was decreased by *GATA1s* independently of the associated mutations, suggesting that GATA1s impaired terminal MK maturation (Figure 2, A and B). The $CD41^+CD42^+$ cell fractions that maintained CD34 expression were 30% and 10% in D21 and T21, respectively (Supplemental Figure 3C). CD34 expression was increased by GATA1s in T21 (TG) to 45% and further by *SMC3* haploinsufficiency to 55% (Figure 2, A and C). TMD and DS-AMKL megakaryoblasts are characterized by the expression of CD34, CD41, and CD42 antigens, but also other markers, such as KIT and CD7 (28, 29). While only a small fraction of D21 or T21 $CD41^+$ cells expresses KIT (Supplemental Figure 3A), it is detected on over half of the *GATA1s* $CD41^+$

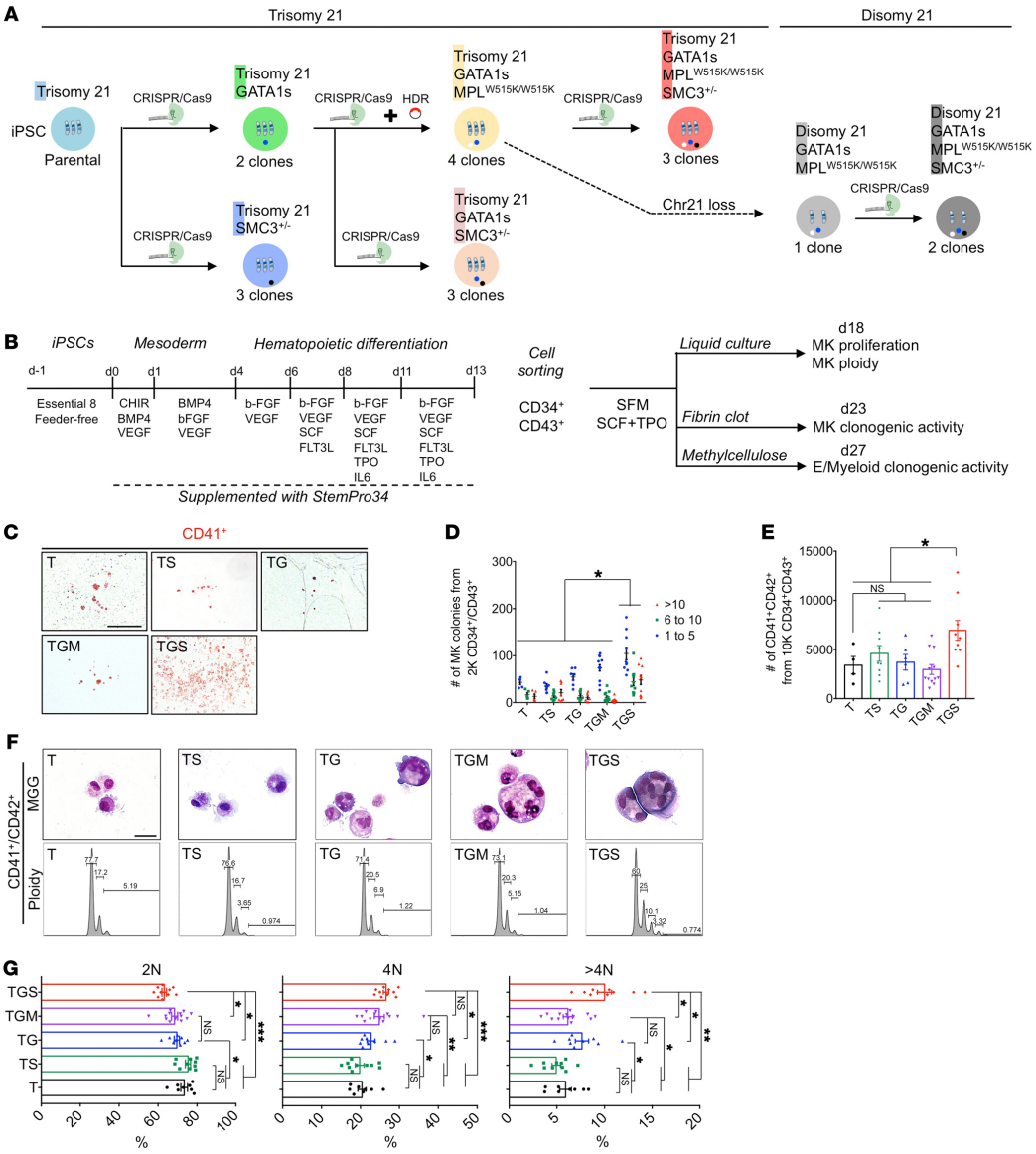


Figure 1. *GATA1s* cooperates with *SMC3*$^{+/-}$ to increase MK clonogenic potential, proliferation, and polyploidization. (**A**) Schematic overview of the iPSC clones generated by a stepwise introduction of *GATA1s*, *MPL*^{W515K}, and *SMC3*$^{+/-}$, using CRISPR/Cas9. Bold letters highlight the abbreviation used hereafter. The number of clones obtained for each genotype and subjected to MK differentiation studies is indicated. The dashed arrow indicates the isogenic D iPSC clone harboring *GATA1s* and *MPL*^{W515K} mutations randomly obtained through loss of 1 chromosome 21. (**B**) Schematic overview of the hematopoietic differentiation method used and the subsequent MK phenotypic characterization. (**C**) Representative images of CFU-MK colonies. Scale bar: 500 μm. (**D**) Histogram of the number of CD41$^+$ colonies obtained from 2000 CD34$^+$CD43$^+$ in fibrin clot assay. (**E**) Histogram of the number of CD41$^+$CD42$^+$ MKs obtained from 10,000 CD34$^+$CD43$^+$ in liquid culture assay. Data are represented as mean ± SEM; n = 3–4. The number of clones tested per genotype was as follows: T/parental = 1; TS = 3; TG = 2; TGM = 4; TGS = 3. (**F**) May-Grünwald-Giemsa (MGG) staining (**E**, upper panels) and ploidy plots (**E**, lower panels) of iMK according to the indicated genotypes. T and TS CD41$^+$CD42$^+$ showed typically mature micro-MKs with acidophilic cytoplasm, while TG, TGM, or TGS showed large polyploid immature MKs with basophilic cytoplasm. Scale bar: 50 μm. (**G**) Histograms of the percentages of 2N, 4N, and >4N of iMKs. Data are represented as mean ± SEM; n = 3 to 8. The number of clones tested per genotype was as follows: T/parental = 1; TS = 2; TG = 2; TGM = 4; TGS = 3. Statistical significance was determined using 1-tailed Mann-Whitney U test. *P < 0.05; **P < 0.01; ***P < 0.001.

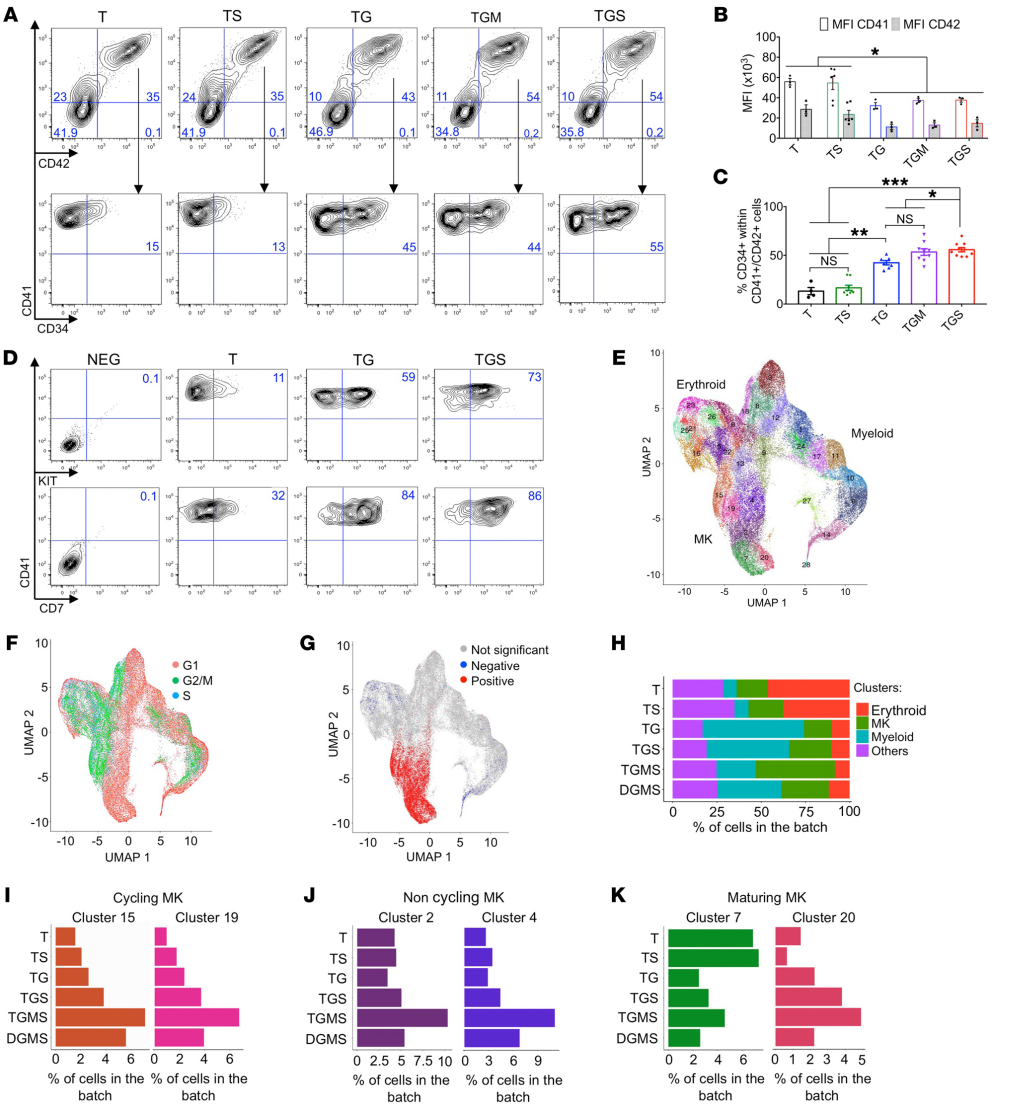


Figure 2. Assessment of iMK differentiation alterations. (**A**–**D**) Immunophenotypes of iMKs for the CD34, KIT, CD7, CD41, and CD42 markers found in DS-AMKL patient blasts. (**A**) Contour plots showing the expression of CD34, CD41, and CD42 markers. (**B**) Histogram shows the MFI of CD41 and CD42. (**C**) Histogram shows the percentage of CD34⁺CD41⁺ per total CD41⁺ population. Data in **B** and **C** are represented as mean ± SEM; n = 3–4. The number of clones tested per genotype was as follows: T/parental = 1; TS = 2; TG = 2; TGM = 3; TGS = 3. Statistical significance was determined using 1-tailed Mann-Whitney's U test: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. (**D**) Representative contour plots of KIT and CD7 expression in the iMK population from 2 independent experiments. (**E** and **F**) scRNA-Seq of CD43⁺ iPSC-derived hematopoietic cells at day 13 of culture. (**E**) UMAP integration of cells from all conditions. Clusters were defined using the Louvain algorithm and numbered and labeled with unique colors. (**F**) UMAP integration with cells colored according to the predicted cell-cycle stage (Seurat method). (**G**) UMAP integration with cells colored according to the enrichment in a MK signature. Red, cells are significantly enriched for the signature; blue, cells are significantly depleted for the signature; gray, no significant enrichment. (**H**) Bar plot shows the proportion of cells in the indicated hematopoietic lineages for each condition. (**I**) Bar plots of the proportion of cells in the 2 clusters of cycling MKs. (**J**) Bar plots of the proportion of cells in the 2 clusters of noncycling MKs. (**K**) Bar plots of the proportion of cells in the 2 clusters of maturing MKs. Cluster 7 represents normal maturing MKs. Cluster 20 represents an abnormal MK population.

cells and further increased by *SMC3* haploinsufficiency (Figure 2D). Similar results were observed for CD7 (Supplemental Figure 3A and Figure 2D). These results show that the GATA1s-induced blockage of MK differentiation at the megakaryoblast stage is worsened by *SMC3* haploinsufficiency.

To obtain a precise picture of the altered differentiation, we performed single-cell RNA-Seq (scRNA-Seq) for D, T, TS, TG, TGS, TGMS, and DGMS at an early differentiation time point (day 13). From 6800 to 14,001 cells with similar characteristics were analyzed per condition (Supplemental Figure 4A). UMAP clustering (https://umap-learn.readthedocs.io/en/latest/clustering.html) obtained by pooling cells from all the conditions provided 28 clusters (Figure 2E) for which cell-cycle status was inferred using Seurat (Figure 2F and Supplemental Figure 5A) and cells of the different clusters showed expression of MK, erythroid, and myeloid differentiation markers (Figure 2G, Supplemental Figure 4, B–D, and Supplemental Figure 5B). To identify all cells belonging to the MK, erythroid, and myeloid lineages, we computed scores for lineage-specific gene lists (Supplemental Figure 4E). T21 differentiation was clearly biased toward the erythroid lineage in comparison with D21 (Supplemental Figure 5C). The highest percentage of MK cells was observed in TGMS (Figure 2H). By quantifying the clusters along MK differentiation, we observed that the cycling, noncycling, and maturing MKs were similar between D21 and T21 (Supplemental Figure 5, D–F). Interestingly, a gradual increase for the cycling MKs (clusters 15 and 19) was observed in TS, TG, TGS, and TGMS as compared with T, with the highest percentage in TGMS (Figure 2I). Noncycling MK clusters (clusters 2 and 4) were more specifically amplified in TGMS as compared with T, TS, TG, and TGS (Figure 2J). Strikingly, 2 clusters were associated with the highest expression of maturing MK markers (clusters 7 and 20) (Figure 2K). Cluster 7 represented the vast majority of cells in T and TS, while cluster 20 was gradually increased in TG, TGS, and TGMS, representing approximately 50% of maturing MKs in TGMS.

Together, these data show that GATA1s and *SMC3*$^{+/-}$ cooperated together and with T21 through (a) an increase in immature cycling MKs and (b) an altered MK differentiation with the predominance of a cluster of maturing MKs that represents a minority among D21 or T21 MKs. MPL^{W515K} markedly increased MK output at the expense of the myeloid lineage.

GATA1s impairs organelle development and proplatelet formation. To better understand the cellular defect in MK differentiation, we performed staining of vWF and CD63, respective markers of α-granules and lysosomes, both deriving from multivesicular bodies (Figure 3A). In contrast with T and TS MKs, TG MKs showed the presence of abnormally large sized α-granules, as attested by the vWF labeling (Figure 3A). Colabeling with the anti-CD63 antibody was as rare as in control MKs (10% of total MK), attesting that these large granules were not multivesicular bodies (Figure 3A). The anti-CD41 antibody was used to highlight demarcation membrane system (DMS) development (Figure 3B; adapted from ref. 30). In T and TS MKs, CD41 was localized at the cell surface and well distributed in the cytoplasm (Figure 3C) in more than 90% of the MKs. In contrast, in the presence of *GATA1s*, CD41 was detected in the center of the cell between the nuclear lobes and only very weakly at the cell surface in the great majority of the cells (Figure 3C). Complete *Z*-stack analysis showed that this intracytoplasmic labeling was in continuity with the cell surface with usually only a single connection demonstrating the presence of a rudimental DMS. This abnormality was even increased by *SMC3*$^{+/-}$ with a more polarized labeling (Figure 3C). Ultrastructural studies confirmed that the *GATA1s* mutation was associated with the presence of very large cells (over 40 μm diameter) with a large multilobulated uncondensed nucleus containing a majority of euchromatin and large atypical granules (Figure 3D). Only a pre-DMS could be detected, usually localized between the nuclear lobes, a defect that was increased by *SMC3*$^{+/-}$.

Then, we investigated the effects of GATA1s and the other mutations on proplatelet (PPT) formation. T and TS MKs were able to form PPTs (Figure 3E). Unexpectedly, *SMC3*$^{+/-}$ increased the percentage of PPT-forming MKs, suggesting that, in the context of GATA1^{WT}, it favors MK maturation (Figure 3F). In contrast, MKs from all *GATA1s* conditions, including with *SMC3*$^{+/-}$, barely formed PPT (Figure 3, E and F).

Combined GATA1s and SMC3$^{+/-}$ *induce proliferative and DS-AMKL signatures*. We then performed transcriptomic analyses. To study a pure population of MKs, we excluded cells positive for the CD33 myeloid marker and sorted CD41$^+$CD42$^+$CD33$^-$ cells using 2 consecutive sorting steps to achieve a purity greater than 98%, confirmed by flow cytometry reanalysis and cytology (Supplemental Figure 6, A–C). By principal component analysis (PCA), a clear clustering of samples according to GATA1s or *SMC3*$^{+/-}$ status was found (Supplemental Figure 6D) with a large number of deregulated genes (DEGs) in TG versus T and TS, respectively (adjusted $P < 0.05$; Supplemental Figure 6E). MPL^{W515K} did not influence the clustering and had little impact on gene expression (Supplemental Figure 6, D and E). In contrast, TGS importantly increased the number of DEGs when compared with T, TS, and TG, respectively (Supplemental Figure 6E).

The genes commonly upregulated by TGS versus T, TS, or TG were enriched in signatures associated with cell-cycle activity and DNA replication (Supplemental Figure 6F), including MYC signatures (Figure 4, A and B) and MYC target genes (e.g. *PAICS*, *MYBP1A*, *BOP1*, *ATIC*, and *CDK4*) (Figure 4C). Importantly, Gene Set Enrichment Analyses (GSEAs) indicated that the gene signature characterizing DS-AMKL versus DS-TMD blast patients (31) was also enriched in the TGS versus TG MKs (Figure 4D). Accordingly, the gene signature of DS-TMD versus DS-AMKL patients was enriched in TG versus TGS MKs (Figure 4D). These results support the idea that *SMC3* haploinsufficiency acquisition, in the context of T21 and GATA1s, leads to increased proliferation properties and reproduces the molecular progression observed from TMD to DS-AMKL in patients.

SMC3$^{+/-}$ *cooperates with GATA1s to impair GATA1 target and MK maturation gene expression*. We subsequently analyzed GATA1 target genes involved in MK maturation. Gene Ontology and GSEA analyses indicated that GATA1 target genes (32) are downregulated in TG versus T (Figure 4, E and F). *SMC3*$^{+/-}$ further decreased expression of GATA targets when associated with GATA1s (Figure 4, E and G), but had no significant effect on GATA1 targets in a GATA1^{WT} context (Figure 4F). Genes involved in DMS development, such as *GP1BA* and *PACSIN2* (Figure 4, H and I), and genes encoding the RAB family proteins controlling membrane traf-

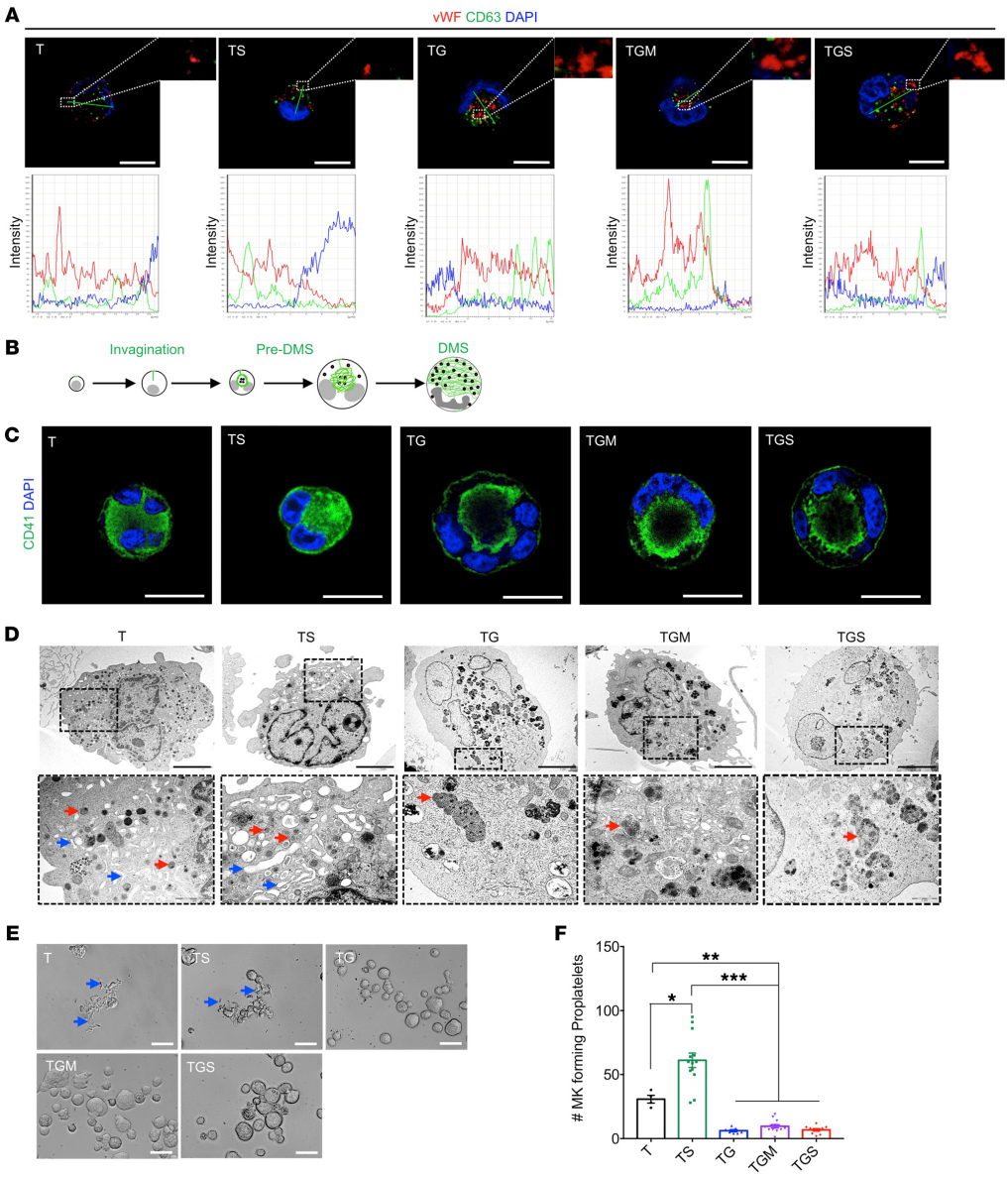


Figure 3. Assessment of MK maturation and platelet formation. (**A**) Confocal analyses of vWF and CD63 expression in MK (left panel) and colocalization analyses (right panel). Scale bars: 50 μm. (**B**) Schematic overview of the normal development steps of DMS. (**C**) Confocal analysis of CD41 marker distribution in iMKs. Scale bars: 50 μm. (**D**) Ultrastructural characterization of the iMK according to the compared genotypes. A representative MK is shown for each condition (upper panels), a part of which (dotted square) is enlarged (lower panels). More than 80% of T and TS MKs showed maturity characteristics associated with condensed nuclei, a well-developed DMS (blue arrows), and the presence of normal α-granules (red arrows). More than 90% of TG, TGM, and TGS MKs showed blockages in maturation characterized by the presence of uncondensed nuclei majorly composed of euchromatin and the absence of DMS formation, with endosomes presenting an abnormal accumulation of granules (red arrows). Scale bars: 5 μm. (**E**) Representative microphotographs of $CD41^+CD42^+$ iMKs under PPT formation assay. PPT-forming MKs are highlighted with blue arrows in the T and TS conditions. Scale bars: 50 μm. (**F**) Histogram of the number of PPT-forming MKs in the compared conditions. Data are represented as mean ± SEM; $n = 3–4$. The number of clones tested per genotype was as follows: T/parental = 1; TS = 3; TG = 2; TGM = 4; TGS = 3. Statistical significance was determined using 1-tailed Mann-Whitney's U test. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

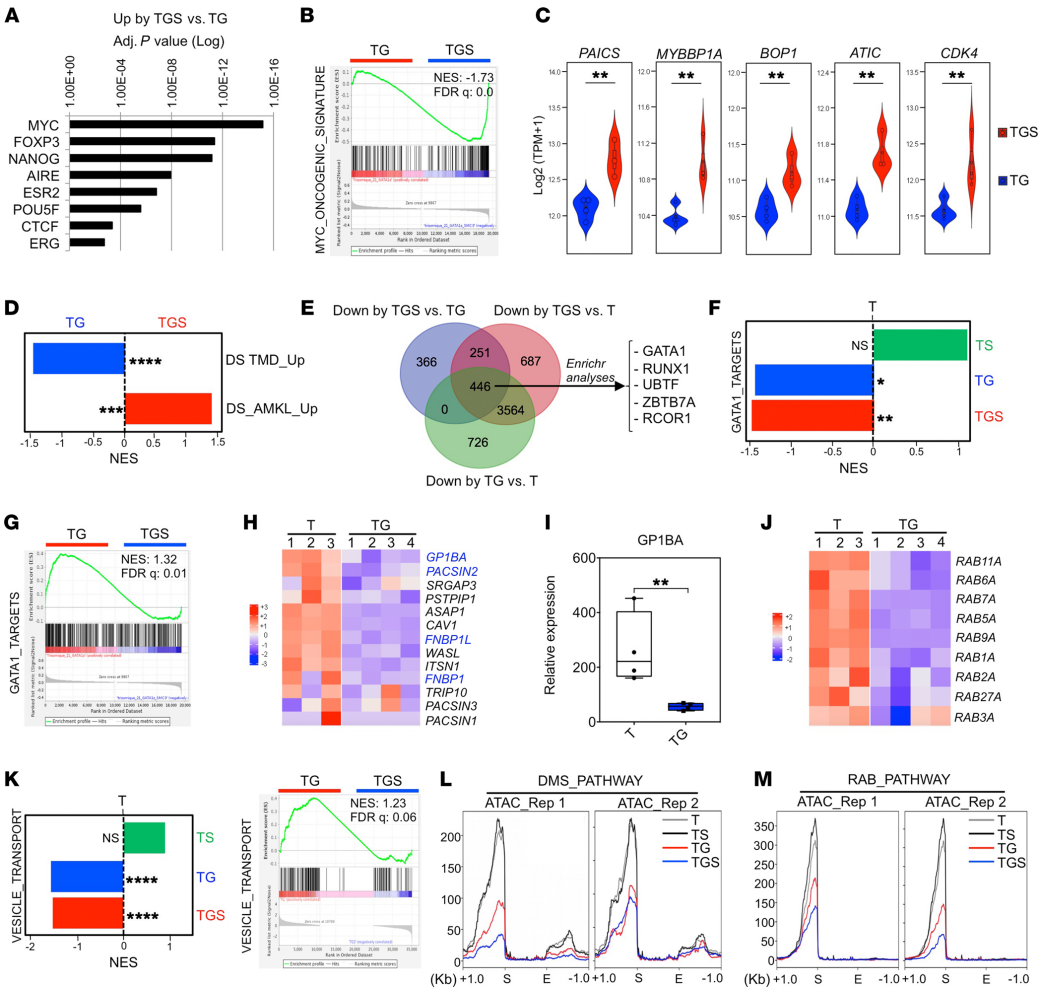


Figure 4. *GATA1s* cooperates with *SMC3*$^{+/-}$ to acquire DS-AMKL features in iMK. (**A**) Transcriptional signature (transcription factor protein-protein interactions) of the upregulated genes in TGS versus TG iMKs. (**B**) GSEA of the MYC oncogenic signature in TGS versus TG. (**C**) Violin plots showing expression (GSEA) of the MYC target genes in TGS versus TG. $^{**}P < 0.01$. (**D**) GSEA of the genes upregulated in DS-TMD compared with DS-AMKL or the genes upregulated in DS-AMKL compared with DS-TMD (31) patients in TGS versus TG iMKs. ***FDR < 0.01; ****FDR < 0.001. (**E**) Venn diagram showing the overlap of downregulated genes in the indicated comparisons. The top transcriptional signatures from the commonly downregulated genes in the compared conditions are annotated with an adjusted $P < 0.000001$ (Encode and ChEA Consensus TFs from CHIP-X). (**F**) GSEA for *GATA1* gene targets (32) in TS versus T, TG versus T, and TGS versus T iMK. NES, normalized enrichment score. *FDR < 0.25; **FDR < 0.05. (**G**) GSEA for *GATA1* gene targets (32) in TGS versus TG iMKs. (**H**) Heatmap showing the expression of genes encoding the F-BAR domain-containing proteins in T versus TG iMKs. Genes known to be direct targets of GATA1 (32) are marked in blue. (**I**) RT-qPCR analysis of *GP1BA* relative expression in T versus TG iMKs. $^{**}P < 0.01$. (**J**) Heatmap showing the expression of the RAB family genes in T versus TG iMKs. (**K**) GSEA for vesicle-mediated transport (GO term GO:0016192) in T versus TG or TG versus TGS iMKs. ****FDR < 0.001. (**L**) Profile plots of ATAC signal on DMS genes comparing T, TS, TG, and TGS iMKs. The start (S) and the end (E) of the genes were plotted across a 1 kb flanking window; the *y* axis indicates depth per million mapped reads. Left panel: replicate (Rep) 1; right panel: replicate 2. (**M**) Profile plots of ATAC signal on RAB family genes comparing T, TS, TG, and TGS iMKs.

ficking and granule movement were downregulated in TG versus T MKs (Figure 4J). In addition, TG MKs were depleted in genes involved in vesicle transport compared with T MKs, and addition of *SMC3*$^{+/-}$ further depleted this signature (Figure 4K).

In view of these marked differences between GATA1 and GATA1s, we hypothesized that GATA1 and GATA1s may modify chromatin accessibility at key differentiation gene loci. Addition of *GATA1s* (TG), but not *SMC3*$^{+/-}$ (TS), significantly reduced the

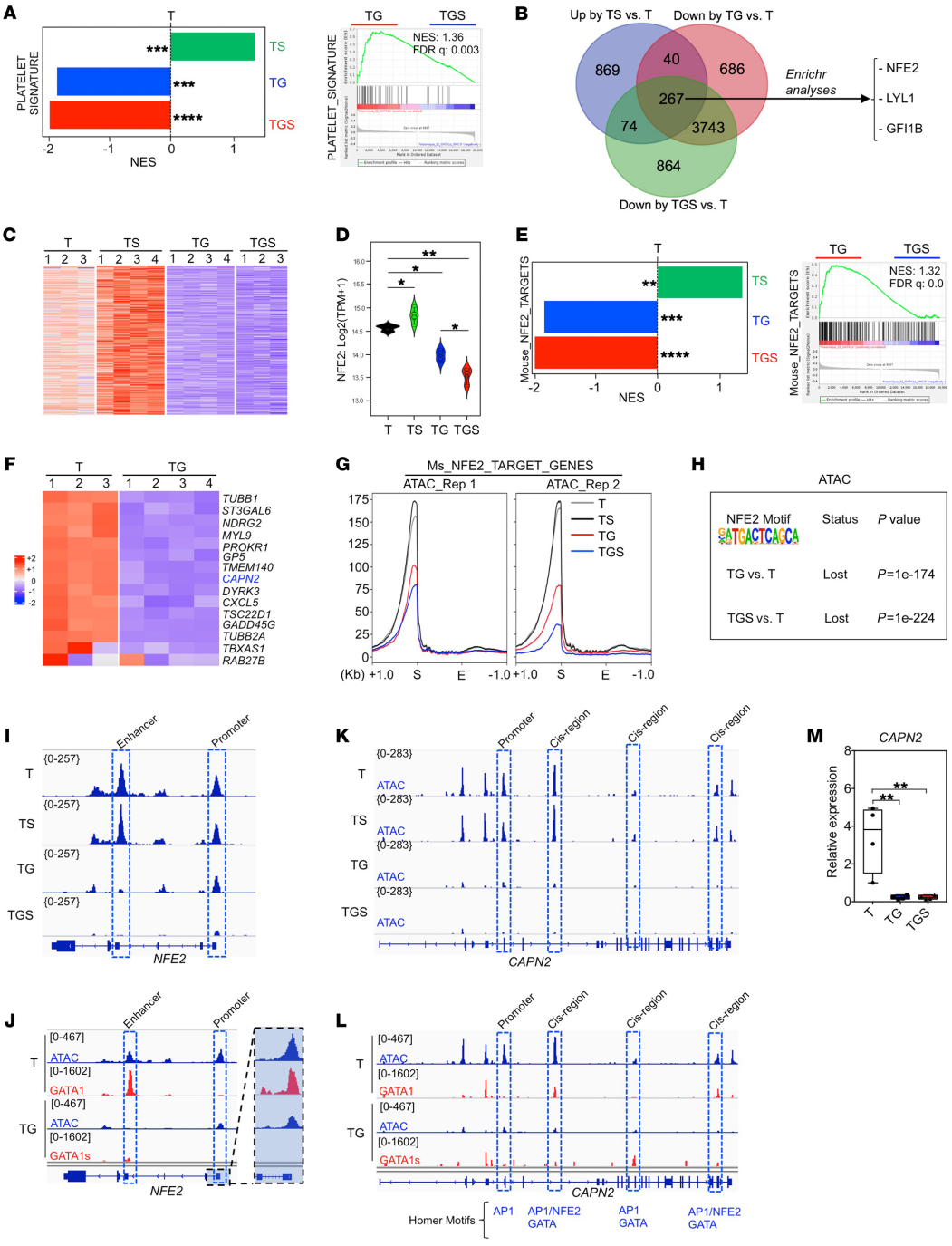

A
T
TS
TG
TGS
PLATELET SIGNATURE
NES
TG
TGS
NES: 1.36
FDR q: 0.003
PLATELET_SIGNATURE
B
Up by TS vs. T
Down by TG vs. T
869
40
686
267
74
3743
864
Down by TGS vs. T
Enrichr analyses
- NFE2
- LYL1
- GFI1B
C
T
TS
TG
TGS
D
NFE2: Log2(TPM+1)
T TS TG TGS
E
T
TS
TG
TGS
Mouse_NFE2_TARGETS
NES
TG
TGS
NES: 1.32
FDR q: 0.0
F
T
TG
TUBB1
ST3GAL6
NDRG2
MYL9
PROKR1
GP5
TMEM140
CAPN2
DYRK3
CXCL5
TSC22D1
GADD45G
TUBB2A
TBXAS1
RAB27B
G
Ms_NFE2_TARGET_GENES
ATAC_Rep 1
ATAC_Rep 2
T
TS
TG
TGS
(Kb) +1.0 S E -1.0 +1.0 S E -1.0
H
ATAC
NFE2 Motif
Status
P value
TG vs. T
Lost
P=1e-174
TGS vs. T
Lost
P=1e-224
I
Enhancer
Promoter
T
TS
TG
TGS
NFE2
K
Promoter
Cis-region
Cis-region
Cis-region
ATAC
CAPN2
M
CAPN2
Relative expression
T
TG
TGS
J
Enhancer
Promoter
ATAC
GATA1
GATA1s
NFE2
L
Promoter
Cis-region
Cis-region
Cis-region
CAPN2
Homer Motifs
AP1
AP1/NFE2 GATA
AP1 GATA
AP1/NFE2 GATA

Figure 5. *GATA1s* cooperates with *SMC3*$^{+/-}$ to downregulate the NFE2 transcriptional program. (**A**) GSEA for platelet genes (47) in the indicated comparisons. ***FDR < 0.01; ****FDR < 0.001. (**B**) Venn diagram showing the overlap of up- and downregulated genes in the indicated comparisons. The top transcriptional signatures from the commonly downregulated genes in TG or TGS and upregulated in TS compared with T are annotated with an adjusted $P < 0.0001$ (Enrichr submissions TF-gene cooccurrence). (**C**) Heatmap showing the expression of the 267 genes that are common in all comparisons (up by TS versus T, down by TG versus T, down by TGS versus T) shown in the Venn diagram in **B**. (**D**) Violin plot showing *NFE2* expression in the indicated conditions. *$P < 0.05$ and *$P < 0.01$. (**E**) GSEA for NFE2 target genes (39) in TS versus T, TG versus T or TGS versus T (left panel), and TGS versus TG (right panel) iMKs. **FDR < 0.05; ***FDR < 0.01; ****FDR < 0.001. (**F**) Heatmap showing the expression of NFE2 target genes that are downregulated in TG compared with T. *CAPN2* gene marked in blue is also known as a direct target of GATA1. (**G**) Profile plots of ATAC signal from NFE2 target genes (Zang et al., 2016) comparing T, TS, TG, and TGS iMKs. The start (S) and the end (E) of the genes were plotted across a 1 kb flanking window; the *y* axis indicates depth per million mapped reads. Left: replicate 1; right, replicate 2. (**H**) Relative loss of the NFE2 motif in ATAC-Seq data from the indicated comparisons. (**I**) ATAC-Seq profile on *NFE2* in T, TS, TG, and TGS. (**J**) Track depicting CUT&TAG-Seq for GATA1 and GATA1s and ATAC-Seq on the *NFE2* locus in T or TG iMKs. (**K**) ATAC-Seq profile on *CAPN2* in T, TS, TG, and TGS. (**L**) Track depicting CUT&TAG-Seq for GATA1 and GATA1s and ATAC-Seq on the *CAPN2* locus in T or TG iMKs. (**M**) RT-qPCR analysis of *CAPN2* expression in T, TS, TG, and TGS iMK.

global intensity of ATAC-Seq signals compared with T (Supplemental Figure 7A). Combination of *GATA1s* and *SMC3*$^{+/-}$ showed further reduction in peak intensity (Supplemental Figure 7A). Lower accessibilities in TG versus T and in TGS versus T were seen in promoter, intergenic, and intronic regions, while the gained sites were mainly intergenic and intronic (Supplemental Figure 7B). The peak distribution was different in GATA1^{WT} (TS versus T) (Supplemental Figure 7B). ATAC profiles on DMS and RAB pathway genes showed a decreased accessibility at their transcription starting sites in TG and even more in TGS versus T (Figure 4, L and M). Together, combined RNA-Seq and ATAC-Seq profiles show that GATA1s in T MKs represses GATA1 targets as well as genes involved in organelle formation, a phenotype that was further enhanced by addition of *SMC3*$^{+/-}$, supporting a model of cooperation leading to a failure of the GATA1-driven transcriptional program.

Molecular signatures of progression identify inhibition of the NFE2 program. Since *SMC3* haploinsufficiency-induced PPT formation in GATA1^{WT} was abrogated in a GATA1s context (Figure 3, E and F), we searched for molecular signatures enriched in TS but depleted in TG and TGS contexts. This included platelet gene signatures (Figure 5A). Enrichment analysis of the 267 common genes suggested that the NFE2-regulated program was inhibited (Figure 5, B and C). Notably, while NFE2 is a well-characterized target of GATA1 in terminal MK differentiation (33), it is not deregulated by Gata1s in mouse MKs (34). Here, *NFE2* expression was profoundly downregulated in a stepwise manner (Figure 5D). GSEA and heatmap analysis showed a depletion in NFE2 targets in TG versus T MKs (Figure 5, E and F) and TGS versus TG MKs (Figure 5E). In contrast, *SMC3*$^{+/-}$ in a GATA1^{WT} (TS) showed an increased expression of both *NFE2* and NFE2 target genes (Figure 5, D and E), supporting the enhanced PPT formation.

ATAC-Seq profiles over TSS of NFE2 target genes showed a gradually lower accessibility in TG and TGS compared with either T or TS MKs (Figure 5G). DNA-binding motif analyses indicated that GATA1s led to a relative loss of chromatin accessibility at NFE2 ($P = 1 \times 10^{-174}$, Figure 5H) as well as of other AP1 family and GATA motifs (Supplemental Figure 7C) and gain in chromatin accessibility at RUNX and ETS motifs (Supplemental Figure 7D) in both TG and TGS (Supplemental Figure 7, E and F). The ETS and RUNX motifs as well as AP1 family motifs were further lost in TGS compared with TG (Supplemental Figure 7G), while no motif was significantly gained (Supplemental Figure 7H).

Importantly, there was a good correlation between *NFE2* gene expression and its chromatin accessibility with a marked decreased accessibility at its promoter as well as cis-regulatory enhancer by GATA1s (Figure 5I). The CUT&Tag approach (35) showed a direct binding of GATA1 at both the *NFE2* promoter and enhancer (T21 context) with a strong reduction in GATA1s binding (TG context; Figure 5J). Notably, a similar observation was made for *CAPN2* (Figure 5, K–M), which was previously demonstrated to be altered by Gata1s in mouse MK (12). Our integrative approach revealed that *NFE2* is a direct target of GATA1 in human MKs, and that GATA1s plus SMC3 mutation leads to a stepwise reduction in *NFE2* expression associated with a global inhibition of the NFE2-mediated transcriptional program.

NFE2 is involved in the MK development defect induced by GATA1s. To assess the role of *NFE2* downregulation by GATA1s, we transduced empty or *NFE2*-encoding lentiviral vector into T and TG iPSC-derived hematopoietic cells. Transduced CD34$^+$CD43$^+$ HPCs were sorted according to GFP expression and induced into MK differentiation (Figure 6A). Ectopic *NFE2* was confirmed by quantitative reverse-transcriptase PCR (RT-qPCR) in T and TG (Figure 6B). Known NFE2 target genes that were strongly downregulated upon *GATA1s* acquisition were significantly upregulated in *NFE2*-transduced compared with empty vector–transduced TG MKs (Figure 6C). *NFE2* expression in TG cells completely restored levels of *CAPN2*, *TBXAS1*, and *RAB27B* and partially restored levels of *TUBB1*.

NFE2 expression did not influence the clonogenic potential of either MK or HPC progenitors in T or TG (Figure 6D, Supplemental Figure 8, A and B), supporting the idea that NFE2 does not regulate MK progenitor proliferation. Also, *NFE2* reexpression had no effect on the ploidy (Figure 6E and Supplemental Figure 8C) or the percentages of CD41$^+$CD42$^+$ and CD34$^+$ cells (Supplemental Figure 8, D–F). However, NFE2 expression significantly decreased the proportion of large megakaryoblasts with abnormal α-granules and a pre-DMS, the maturation signature defect of TG MKs (Figure 6, F and G). Furthermore, ectopic NFE2 induced expression of β_1-tubulin located along the cytoplasmic membrane in 30% of TG MKs (Figure 6H) and partly rescued PPT formation by TG MK (Figure 6, I and J).

Contribution of the T21 to the cooperation between GATA1s and SMC3$^{+/-}$. To characterize the contribution of T21 in the phenotype described above, we generated an isogenic DGM clone and further introduced the SMC3$^{+/-}$ mutation to obtain DGMS clones (Figure 1A and Supplemental Figure 1, T and V).

The percentage of CD41$^+$CD42$^+$ MKs generated in isogenic D and T conditions was close and above 50% (Supplemental Figure 9A and Figure 7A). Unexpectedly, the percentage of CD34$^+$CD41$^+$ cells

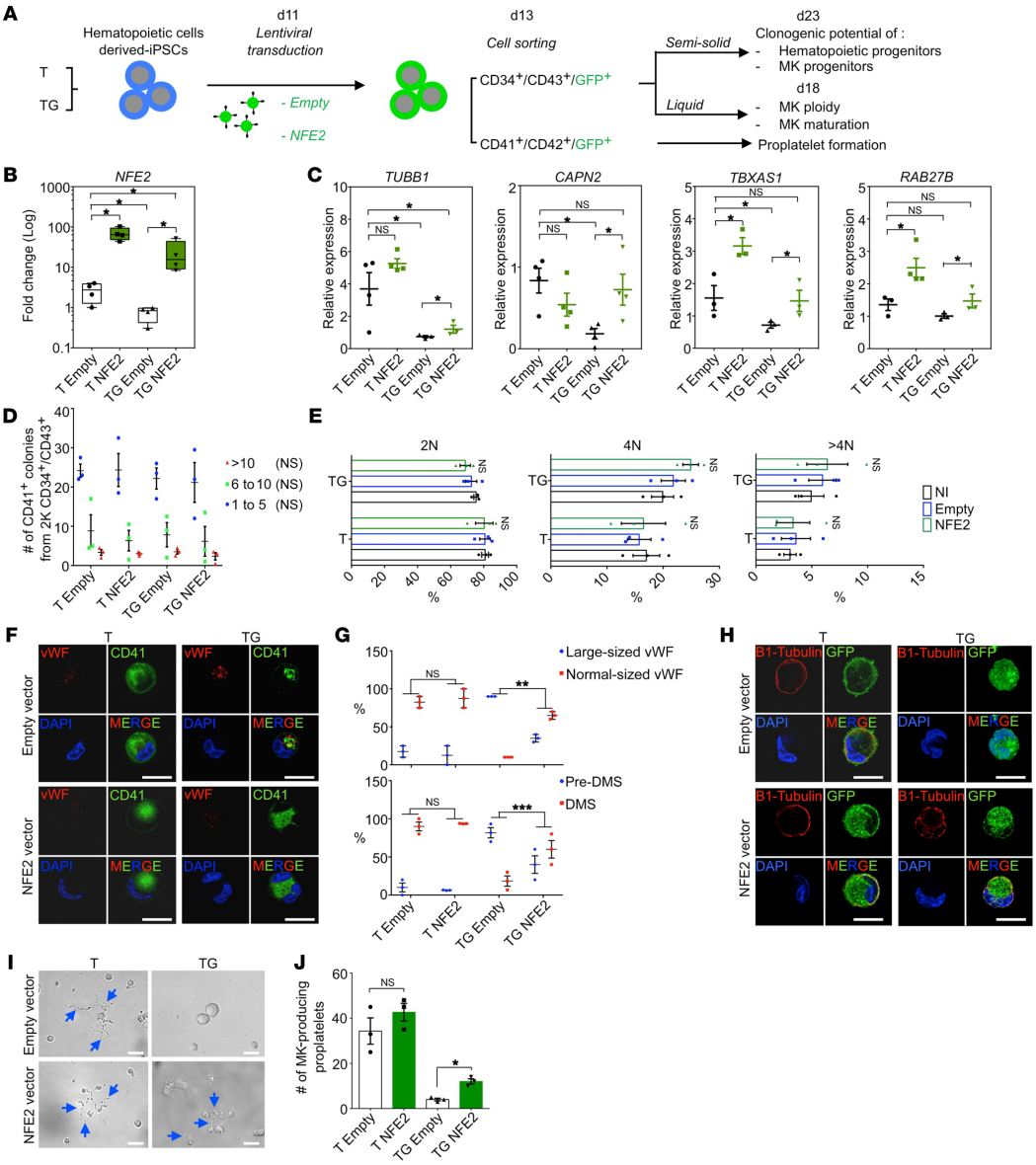


Figure 6. NFE2 overexpression in TG MK partially rescues maturation and PPT defects. (**A**) Schematic overview of NFE2 overexpression strategy in T and TG iMKs and of subsequent phenotypic studies. (**B**) Validation by RT-qPCR of NFE2 overexpression in T and TG iMKs. (**C**) RT-qPCR analyses of the known NFE2 target gene expression with empty or NFE2 lentiviral vectors. (**D**) Effect of NFE2 overexpression on the clonogenic potential of T and TG iMKs. (**E**) Effect of NFE2 overexpression on the ploidization of T and TG iMKs. (**F**) Confocal analyses of NFE2 overexpressing MK in T and TG for CD41 and vWF expression. Scale bars: 50 μm. (**G**) Histograms of the percentages of MKs with normal or large-sized vWF (upper panel) and the percentages of MKs with pre-DMS or DMS (lower panel) according to the absence (empty) or presence of NFE2 lentiviral vector. Quantifications were performed on 20 MKs from 3 independent experiments. (**H**) Confocal analyses of β_1-tubulin expression in T- and TG-derived MKs according to the absence (empty) or presence of NFE2 lentiviral vector. Scale bars: 50 μm. (**I**) Representative microphotographs of $CD41^+CD42^+$ MKs under PPT formation assay. Note the presence of PPT-forming MKs (blue arrows) in the absence or presence of NFE2 lentivector for the T, while in TG, the presence of PPT-forming MK was observed only in the presence of NFE2 lentivector (blue arrows). Scale bars: 50 μm. (**J**) Histogram of the number of PPT-forming MK shown in **I** according to the compared genotypes. Data are represented as mean ± SEM; $n = 3$. Statistical significance was determined using 1-tailed Mann-Whitney's U test. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

was significantly higher in isogenic clones, nearly 80% of the CD41$^+$ cells reproducing the differences previously observed between T21 and D21 MKs (Supplemental Figure 3C, Supplemental Figure 9A, and Figure 7B). This was correlated with a higher chromatin accessibility at the CD34 promoter in TG and TGMS versus T, which was even higher in DGMS (Figure 7C). CUT&Tag for GATA1 revealed a higher binding of GATA1s compared with GATA1 at a cis-intronic region of *CD34* (Figure 7D). At the cellular level, the CD34$^-$ fraction of CD41$^+$CD42$^+$ MKs in TG or DGMS corresponded to small immature blast cells while the CD34$^+$ fraction included the large polyploid megakaryoblasts (Figure 7E). Therefore, we hypothesized that CD34 expression is aberrantly upregulated upon differentiation in GATA1s MKs. To verify this hypothesis, CD34$^-$CD41$^+$CD42$^+$ MKs from T, TG, and DGMS were sorted and seeded with SCF and TPO to investigate phenotypes after 2 days of culture (Figure 7F). While T MKs remained CD34$^-$CD41$^+$CD42$^+$, some TG MKs, and even more so DGMS MKs, reacquired CD34 expression (Figure 7F). These results support the idea that aberrant reacquisition of CD34 expression is controlled by GATA1s in both T and D MKs, with a higher expression in the latter context. Notably, there was no difference in the ploidy between T and D, with a similar increase in the mean percentage of more than 4N MKs by *SMC3*$^{+/-}$ (Supplemental Figure 9B). MK morphological abnormalities, including α-granules and DMS, were identical at confocal and electron microscopy for DGMS and TGMS MKs (Supplemental Figure 9, C and D). PPT formation was similarly abrogated in DGM and DGMS versus TGM and TGMS (Supplemental Figure 9, E and F). The lack of significant enrichment of platelet signature gene lists further supported the proximity between DGMS and TGMS MK maturation defects (Supplemental Figure 9G). Therefore, the data show that T21 is not directly involved in the MK maturation defect induced by GATA1s and its cooperation with *SMC3*$^{+/-}$.

We subsequently studied whether T21 may affect HPC proliferation. DGM and TGM CD34$^+$CD43$^+$ cells showed similar MK clonogenic potential (Figure 7, G and H). In contrast, TGMS HPCs had higher MK clonogenic potential than their DGMS counterparts, with a higher proliferative capacity highlighted by an increased number of MK colonies composed of more than 10 cells and by an increased MK yield in liquid cultures (Figure 7, G–I). Of note, some rare very large TGMS colonies composed of more than 1000 MKs could be observed.

Differential gene expression analyses revealed 495 up- and 289 downregulated genes in TGMS versus DGMS (Supplemental Figure 6, G and H). More specifically, GSEA showed enrichment of a MYC signature in both TGS and TGMS versus DGMS MKs as well as TGMS versus TGM MKs (Figure 7J), indicating that *SMC3*$^{+/-}$ cooperated with T21 and *GATA1s* to enhance the MYC-induced proliferation program. Importantly, GSEA with a DS-AMKL gene list showed signatures significantly enriched in TGMS versus DGMS iMKs (Figure 7K). Among the genes enriched in TGMS versus DGMS and in TG versus DGMS MKs was MYB (Supplemental Figure 9, H and I), a regulator of myeloid leukemogenesis (36, 37). scRNA-Seq between DGMS and TGMS showed that T21 increased MK output as well as all MK-related clusters, including the abnormal cluster 20 (Figure 2, I–K). Together, the results indicate that T21 contributes to the molecular signature of DS-AMKL and cooperates with GATA1s and *SMC3*$^{+/-}$ in part through positive regulation of a MYC and MYB proliferative program.

Pharmacological inhibition of MYC reduces megakaryoblast expansion and enhances their differentiation. To investigate whether MYC inhibition can target the proliferating megakaryoblasts, hematopoietic progenitors from T, TG, TGMS, and DGMS were treated with JQ-1 inhibitor in a dose-dependent manner for 5 days (Supplemental Figure 10, A–E). JQ-1 decreased the percentage of CD41$^+$CD42$^+$ cells, while it increased the percentage of CD34$^+$ cell fraction within T, TG, TGMS, and DGMS MK populations as compared with the nontreated conditions (Figure 8, A–C). Notably, even very low concentration of JQ-1 (50 nM) dramatically decreased the absolute number of TG but also TGMS or DGMS megakaryoblasts (Figure 8D). In parallel, JQ-1 restored the expression levels of CD41 and CD42 in TG, TGMS, and even more in DGMS compared with the respective nontreated conditions (Figure 8, E and F). To precisely determine the impact of JQ-1 inhibitor on proliferation and differentiation, we evaluated the expression of CD41 and CD42 according to the number of divisions using a violet dye cell tracker after 2 days of treatment (Figure 8G). JQ-1 significantly decreased the percentage of total cells, but also the percentage of CD41$^+$CD42$^+$ cells at division 4 in TGMS and DGMS as compared with the nontreated conditions (Figure 8, H and I). Importantly, JQ-1 induced a dramatic decrease in the absolute number of TGMS MK cells from division 2 until division 4 and even more marked in TGMS than DGMS (Figure 8J). Furthermore, JQ-1 did not yet influence the MFI of CD42 at the day 2 early time point whatever the number of divisions. However, it increased the MFI of CD41 through division 2 in TGMS at a level close to that observed in the nontreated DGMS (Figure 8, K and L). These results show that JQ1 has a much more rapid and significant effect in TGMS than in DGMS.

Together, these results strongly suggest that the MYC transcriptional program plays a key role in acquiring the leukemic features of DS megakaryoblasts and can be pharmacologically targeted.

Discussion

We modeled the defect in MK differentiation observed in human DS-AMKL through stepwise introduction of *GATA1* (*GATA1s*), MPL (*MPL*^{W515K}), and *SMC3* haploinsufficiency (*SMC3*$^{+/-}$) mutations in T21 iPSCs to characterize the contribution of individual alterations and how they cooperate to alter hematopoiesis and impose MK differentiation blockage. We showed that GATA1s impairs MK differentiation and that *SMC3*$^{+/-}$ enhances this phenotype by synergistically acting with GATA1s to induce an even more profound failure of the GATA1-dependent MK differentiation program. These differentiation alterations were independent of T21, supporting the recent observation that T21 is not mandatory for inducing DS-AMKL by the combined *GATA1* and *STAG2* mutations (19). However, T21 cooperates with *SMC3*$^{+/-}$ to increase proliferation. Also, *MPL*-activating mutation further enhanced MK output, including through induction of growth factor independence, but did not affect differentiation, as seen with *FLT3* mutation in another iPSC model of AML (38).

These data uncovered the importance of using human cells for the development of DS-AMKL models. Indeed, murine Gata1s models revealed increased proliferation of yolk sac and fetal liver MKs with minimal differentiation alteration and the

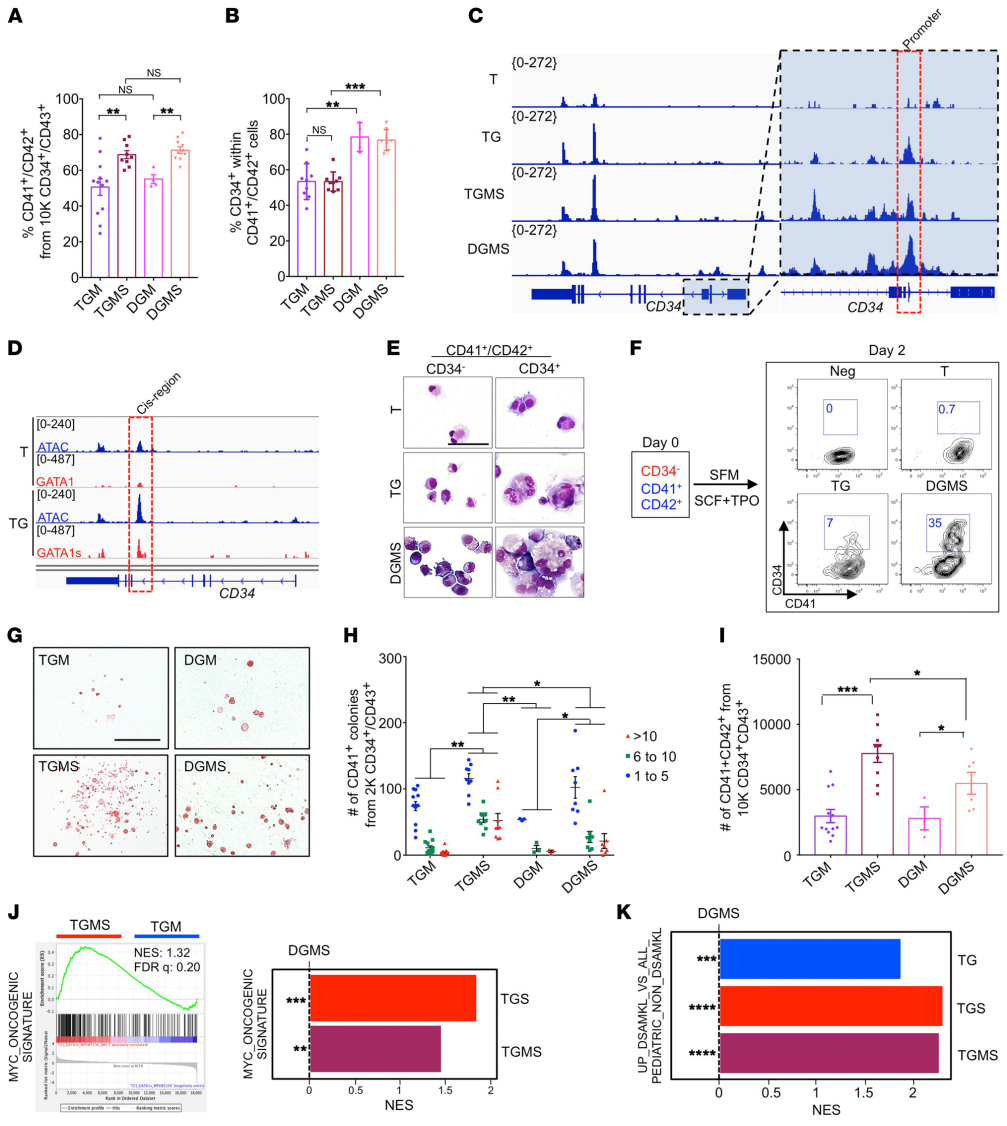


Figure 7. T21 acts in synergy with GATA1s and *SMC3*$^{+/-}$ to increase MK proliferation and to acquire DS-AMKL features. (**A**) Histogram of the mean percentage of CD41$^+$CD42$^+$ cells. (**B**) Percentage of CD34$^+$CD41$^+$ within the CD41$^+$ MK population. Data are represented as mean ± SEM; n = 3 to 4. The number of clones tested per genotype was as follows: TGM = 4; TGMS = 3; DGM = 1; DGMS = 2. (**C**) ATAC-Seq profile on *CD34* in T, TG, TGMS, and DGMS. (**D**) Track depicting CUT&TAG-Seq for GATA1 and GATA1s and ATAC-Seq at the *CD34* locus in T or TG iMKs. (**E**) May-Grünwald-Giemsa coloration of CD34$^-$ or CD34$^+$ fractions within the CD41$^+$CD42$^+$ cell population in T, TG, and DGMS. Scale bar: 50 μm. (**F**) Contour plots show CD34 expression in T, TG, and DGMS iMKs after 2 days of culture. Negative control (Neg CTL): MKs incubated without CD34 antibody. (**G**) Representative microphotographs of CFU-MK colonies in fibrin clot assay for TGM, TGMS, DGM, and DGMS. Scale bar: 500 μm. (**H**) Mean number of CFU-MK colonies obtained from 2000 CD34$^+$CD43$^+$ in fibrin clot assay. (**I**) Mean number of CD41$^+$CD42$^+$ MKs from 10,000 CD34$^+$CD43$^+$ in liquid cultures. (**J**) GSEA for the *MYC* oncogenic signature comparing indicated conditions. (**K**) GSEA for the upregulated genes in DS-AMKL versus non-DS AMKL pediatric patients in TGS versus DGMS or TGMS versus DGMS iMKs. Statistical significance was determined using 1-tailed Mann-Whitney U test. *P < 0.05; **P < 0.01; ***P < 0.001.

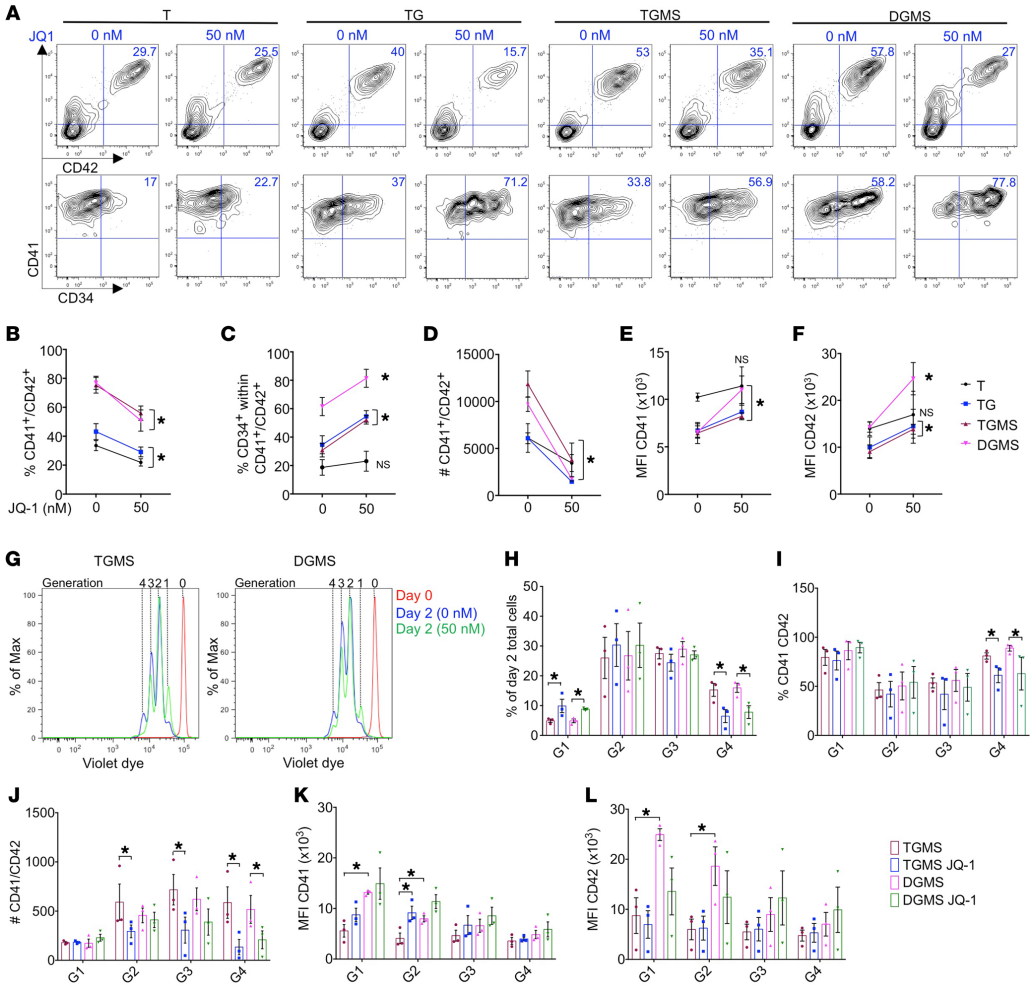


Figure 8. Assessment of MYC inhibition on iMK proliferation and differentiation. (**A**–**F**) Immunophenotypes of iMKs for the CD34, CD41, and CD42 markers found after JQ1 treatment. (**A**) Contour plots showing the expression of CD34, CD41, and CD42 markers. (**B**) Before-and-after graph shows the percentage of $CD41^+$ $CD42^+$. (**C**) Before-and-after graph shows the percentage of $CD34^+$ cells within the total $CD41^+CD42^+$ cell population. (**D**) Before-and-after graph shows the absolute number of CD41 and CD42. (**E**) Before-and-after graph shows the MFI of CD41. (**F**) Before-and-after graph shows the MFI of CD42. Data in **B** through **F** are represented as mean ± SEM; $n = 3$. Conditions with or without JQ-1 are compared, and statistical significance was determined using 1-tailed Mann-Whitney's U test. $^*P < 0.05$. (**G**) Representative flow cytometry plots of a violet dye experiment at day 2 of culture in the TGMS versus DGMS. Red, day 0 of culture; blue, dimethyl sulfoxide control; green, JQ-1. The number of generations is indicated at the top of the plot. (**H**) Histogram of the percentage of day 2 total cells per generation (G) in the indicated conditions. (**I**) Histogram shows the percentage of $CD41^+CD42^+$ per generation in the indicated conditions. (**J**) Histogram shows the absolute number of $CD41^+CD42^+$ in the indicated conditions per generation. (**K**) Histogram shows the MFI of CD41 per generation in the indicated conditions. (**L**) Histogram shows the MFI of CD42 per generation in the indicated conditions. Data are represented as mean ± SEM; $n = 3$. Conditions with or without JQ-1 are compared, and statistical significance was determined using the Kruskal-Wallis test. $^*P < 0.05$.

generation of mature PPT-forming MKs (10). In contrast, the present study showed that GATA1s markedly affects the MK differentiation of both D21 and T21 human iPSCs, inducing the emergence of megakaryoblasts with major impairments in DMS, α-granules, and platelet formation.

Molecularly, numerous GATA1 target genes involved in MK differentiation were downregulated in GATA1s MKs (e.g., *GPIBA*, *PACSIN2*, *RAB* genes and *TUBB1*, respectively, involved in DMS, granule, and PPT formation). Previously, only 2 GATA1 targets (*GPIBA* and *CAPN2*) were reported as downregulated in *Gata1s* murine

models (34, 12). We found that NFE2 and its transcriptional program were profoundly inhibited in human iMKs, which is consistent with the essential role of NFE2 in the regulation of late MK differentiation (33, 39, 40). This resulted from a lower binding of GATA1s to *NFE2* regulatory elements and from further reduction of chromatin accessibility at *NFE2* and NFE2 target loci upon addition of *SMC3*$^{+/-}$ with a concomitant gain in accessibility at ETS motifs (e.g., FLI1 and ERG). Interestingly, *SMC3*$^{+/-}$ has the opposite effect in a GATA1^{WT} background by promoting MK differentiation with an increased PPT formation associated with an enhanced expression of genes involved in MK terminal differentiation, including an NFE2 signature. Therefore, this reveals a functional interplay between cohesin and GATA1 that is dependent of the chromatin organization shaped by GATA1 or GATA1s. The positive effect of *SMC3*$^{+/-}$ on GATA1^{WT} MK maturation was unexpected, as it was reported that a cohesin defect blocks terminal erythroid and myeloid differentiation. However, the precise effect of cohesin mutations on MK maturation was not previously studied, apart from the report of small hypoploid MKs in *Stag2*-deficient mice (24). Together with the functional rescue of PPT formation upon NFE2 ectopic expression, these results emphasized the key role of NFE2 in GATA1s-induced MK differentiation blockage that is further enhanced upon cooperation between *GATA1s* and *SMC3*$^{+/-}$. Thus, the MK differentiation defect induced by GATA1s results from the altered expression of (a) genes directly regulated by GATA1 and (b) genes indirectly regulated through NFE2. Although murine models of Gata1s in combination with T21 and cohesin mutation have not been reported to date, the differential impact of GATA1s on *NFE2* expression in humans and mice may represent a molecular basis for incomplete DS-AMKL modeling in mice.

The predominant effect of endogenous *Gata1s* on mouse hematopoiesis is the induction of hyperproliferation of MK progenitors (10). In this study, we did not observe a significant increase in the number and proliferation of MK progenitors nor on the output of MKs defined as CD41^{+}CD42^{+} cells upon engineering of GATA1s. This is consistent with the effect of *Gata1s* on murine ES cells (13) and the recent observation showing that the increased MK output generated by GATA1s is due to a shift in HSC differentiation toward the MK lineage rather than to a direct effect on MK progenitors (18, 19). However, our results showed a marked synergy between *SMC3*$^{+/-}$ and GATA1s on both proliferation (e.g., increased MK progenitor number and size of MK colonies) and ploidization. This dual effect could be explained by the enrichment in a MYC program when combining T21, GATA1s, and SMC3$^{+/-}$, also seen in other models of *Smc3* haploinsufficiency (41). Indeed, a requirement for a transient MYC activity was reported for MK ploidization (42, 43), while its sustained activity is associated with impaired maturation (42). Notably, the *SMC3*$^{+/-}$-dependent gain in RUNX1/ETS binding site accessibility in T21 MKs may synergize with the previously reported loss of interaction between RB1 and GATA1s (44). An increased level of MYC and of MYC activity in DS-AMKL in comparison with other pediatric AMKL has been previously reported (45), and recently it has been reported that it is related to an excess of RUNX1A isoform (46). Here, we showed that low-dose BET inhibitor treatment inhibits both proliferation and survival of megakaryoblasts. Our data suggest that the MYC pathway is a major and targetable pathway in DS-AMKL and are consistent with a recent study showing that MYC inhibition induces apoptosis of DS-AMKL blasts (46).

While allowing reproducible cellular and molecular analyses from the same clone, the current protocol used to obtain iPSC-derived hematopoiesis does not recapitulate a definitive hematopoiesis with apparent lack of true HSC. This precludes efficient engraftment into immunodeficient mice and clearly represents a limitation in our current model. Indeed, *SMC3* haploinsufficiency did not induce a strong in vitro self-renewal of HSPC nor the engraftment into NSG immunodeficient mice (data not shown) as opposed to what is seen with human primary fetal *GATA1s STAG-2*^{KO} HSPCs (19). Successful engraftment of hematopoietic cells derived from AML or high-grade MDS-iPSC or using an approach similar to ours showed that the sequential introduction of mutations involved in AML (*ASXL1*, *SRSF2*, *NRAS*) allowed low burden in vivo engraftment, but without the capacity of serial transplantations (38). Whether bona fide leukemic engraftment of an iPSC model of DS-AMKL will require more drastic consequences on HSPC self-renewal than *SMC3*$^{+/-}$, such as mutation of *STAG2* presenting only one transcribed allele located on the active X chromosome or a fourth alteration (e.g. targeting the PRC2 complex) (8, 9) or different culture or recipient conditions, remains to be determined. This aspect clearly represents a limitation in our modelization. Nonetheless, such a stepwise introduction of driver mutations of DS-AMKL reproduces the MK leukemic differentiation blockage, sheds light on its molecular mechanism, and paves the way for future studies on other leukemogenesis processes necessary for inducing self-renewal.

In summary, the present in-depth cellular and molecular characterizations of the iPSC-based model contribute to better understanding the progression from TMD to AMKL seen in children with Down syndrome. Specifically, we demonstrated that the low expression of *NFE2* is critical for the induction of MK dysplasia by GATA1s, uncovering a molecular basis of the cooperation among cohesin mutation, GATA1s, and T21 in the MK differentiation defect of DS-AMKL.

Methods

Further information can be found in the Supplemental Methods.

iPSC generation and expansion. The trisomic 21 iPSC line (T21.2; 47XY, +21) was a gift from Stella Chou and Mitchell Weiss. It was derived as previously described (14) through reprogramming of fetal stromal cells after transduction with pMXs-based retroviral supernatant encoding human OCT4, SOX2, KLF4, or MYC. iPSCs were maintained in Essential 8 or Essential 8 Flex Medium (Gibco; Thermo Fisher Scientific) on plates coated with N-truncated human recombinant vitronectin (Gibco; Thermo Fisher Scientific). Cell passages were performed mechanically or by using the StemPro Accutase Solution (Gibco; Thermo Fisher Scientific). A mycoplasma screening was routinely performed, according to the manufacturer's instructions (MilliporeSigma). A list of manufacturers with the catalog number of each product is provided in Supplemental Table 2.

iPSC genome editing. In order to obtain *GATA1s* mutation, oligo encoding sgRNA targeting exon 2 of the gene was inserted into pX330 GFP plasmid (Addgene), while for *SMC3*$^{+/-}$ mutation, sgRNA targeting exon 9 was inserted into lenti-CRISPR V2 GFP plasmid (provided in house). For each experiment, 5 μg of GFP-expressing vector was electroporated into 1 × 10^{6} iPSCs with Human Stem Cell Nucleofector Kit 1 (Lonza) following the manufacturer's instructions. Seventy-two hours

later, the GFP-expressing cells were sorted by FACS and seeded at low density. From 9 to 11 days after seeding, 10 or 40 individual clones were picked manually for *GATA1s* or *SMC3*$^{+/-}$, respectively, where two-thirds of each colony were reserved for genomic DNA extraction. For *GATA1s*, DNA was amplified by PCR using the primers listed in Supplemental Table 3. *GATA1* PCR amplicons were submitted to Sanger sequencing (Eurofins), and 2 clones harboring a frame shift with a premature stop codon in exon 2 were selected. For *SMC3*$^{+/-}$ clone genotyping, 500 ng of purified PCR amplicons were incubated with BanII restriction enzyme (NEB) for 2 hours at 37°C. The genotype was determined thereafter by 2% agarose gel migration. To identify *SMC3* indels of heterozygous mutated clones, the nondigested band was submitted to Sanger sequencing (Eurofins).

For *MPL*$^{W515K/W515K}$ knockin, an sgRNA with a PAM corresponding to W515 of the *MPL* gene was selected and cloned into the lenti-CRISPR V2 GFP. The 5′ and 3′ homology arms of 700 bp of *MPL* surrounding a PGK cherry selection cassette were cloned into pUC57 vector (GenScript). Both CRISPR and donor vectors were cotransfected at a ratio of 2 μg:6 μg, respectively, into 3×10^6 iPSCs using the Human Stem Cell Nucleofector Kit 1 (Lonza). Six days later, cherry-positive expression selection by FACS was performed for clonal expansion. The genotypes of 12 cherry-positive clones were determined by PCR screening of genomic DNA and validated by Sanger sequencing (Eurofins). To delete the PGK-cherry cassette, correctly targeted clones were transfected with flippase-expressing plasmid, and 48 hours later, cells without cherry expression were sorted and expanded. PGK cherry cassette excision was confirmed by PCR.

The guide RNA design and off-target identification were determined using CRISPOR software (http://crispor.tefor.net). Potential indels were screened considering 1 to 4 mismatches by PCR/sequencing. For each edited gene, off-target loci containing a CFD score greater than 0.1 were selected. Off-target regions of each sgRNA used have been amplified by PCR and verified by Sanger sequencing (Eurofins). Karyotyping analyses of all targeted clones were performed by G-banding in a cytogenetics facility (Hôpital Antoine Béclère). All oligos of the different sgRNAs targeting GATA1, SMC3, and MPL genes and those used for PCR amplification are listed in Supplemental Table 3.

Hematopoietic differentiation from iPSC clones. Clumps from iPSC colonies were seeded on Geltrex-coated plates in Essential 8 medium at day −1. The starting cell concentration was at a 20% confluency range that corresponded to approximately 500 to 1000 single cells from 1 iPSC clone per experiment. At day 0, StemPro-34 SFM (Gibco; Thermo Fisher Scientific) xeno-free medium was used, supplemented with 1% penicillin/streptomycin (v/v; Gibco; Thermo Fisher Scientific), 1% L-glutamine (v/v; Gibco), 0.04 mg/mL of 1-thioglycerol (MilliporeSigma), and 50 mg/mL of ascorbic acid (MilliporeSigma). This medium was retained until day 13 and supplemented with different small molecules, cytokines, and growth factors in a time-dependent manner, as described in Supplemental Table 4. At day 13, CD34$^+$CD43$^+$ HPC-like cells of the supernatant were sorted (see Supplemental Methods) and seeded to undergo MK differentiation in a serum-free medium containing 25 ng/ml hSCF (Biovitrum) and 25 ng/ml hTPO (Kirin Brewery). Functional validation experiments of the *MPL*$^{W515K/W515K}$ were performed by adding or not adding 1 μM ruxolitinib (Euromedex). All incubations were performed at 37°C, 5% CO_2. A list of manufacturers with the catalog number of each product is provided in Supplemental Table 2.

Statistics. Statistical analyses were performed with GraphPad Prism software 70.a. Comparisons between different groups were performed using a 1-tailed unpaired Mann-Whitney *U* test or 1-way ANOVA Kruskal-Wallis test. Data are represented as mean ± SEM. For all analyses, $P < 0.05$ was considered statistically significant.

Data access. RNA-Seq, scRNA-Seq, ATAC-Seq, and CUT&Tag raw data were deposited in ArrayExpress (E-MTAB-11032, E-MTAB-11033, E-MTAB11036, E-MTAB-11044).

Author contributions

BA, HR, TM, and WV designed the study. BA, FB, S Mazzi, VD, AS, YM, ZA, MV, IA, MC, N Debili, and N Droin performed gene editing and cellular and molecular analyses. SS performed electron microscopy. RP, PR, CC, RD, GT, OH, IP, S Malinge, and ES contributed essential reagents and expertise. ER, FB, and MA performed the bioinformatics analyses with help from BA, FB, TM, and WV. BA, TM, and WV drafted the manuscript. All authors approved the final manuscript.

Acknowledgments

We thank Stella Chou (Children's Hospital of Philadelphia) and Mitchell Weiss (St. Jude Children's Research Hospital) for providing the T21 iPSCs (14). We also thank Heike Pahl (University of Freiburg Medical Center) for providing the NFE2 vector (pLeGO iG NFE2). This work was supported by Institut National Du Cancer (2016-1-PL BIO-04-1), Ligue Contre le Cancer (Equipe Labelisée: TM, WV, HR, IP), and Sites de Recherche Intégrée sur le Cancer–SOCRATE (SIRIC-SOCRATE) (INCa-DGOS-INSERM 12551). TM is a member of the OPALE Carnot Institute and of the PEDIAC Consortium (INCA_15670, www.programme-pediac.com).

Address correspondence to: William Vainchenker or Thomas Mercher, Gustave Roussy Cancer Campus, INSERM UMR1287, 39 rue Camille Desmoulins, 94800 Villejuif, France. Phone: 33.1.42.11.53.63; Email: william.vainchenker@gustaveroussy.fr (WV); Phone: 33.1.42.11.44.83; Email: thomas.mercher@inserm.fr (TM).

1. Lopez CK, et al. Pediatric acute megakaryoblastic leukemia: multitasking fusion proteins and oncogenic cooperations. *Trends Cancer*. 2017;3(9):631–642.
2. McNulty M, Crispino JD. Acute megakaryocytic leukemia. *Cold Spring Harb Perspect Med*. 2020;10(2):a034884.
3. Roberts I, et al. The impact of trisomy 21 on foetal haematopoiesis. *Blood Cells Mol Dis*. 2013;51(4):277–281.
4. Garnett C, et al. GATA1 and cooperating mutations in myeloid leukaemia of Down syndrome. *IUBMB Life*. 2020;72(1):119–130.
5. Wechsler J, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet*. 2002;32(1):148–152.
6. Roberts I, et al. GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. *Blood*. 2013;122(24):3908–3917.
7. Bhatnagar N, et al. Transient abnormal myelopoiesis and AML in down syndrome: an update. *Curr Hematol Malig Rep*. 2016;11(5):333–341.
8. Yoshida K, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet*. 2013;45(11):1293–1299.
9. Labuhn M, et al. Mechanisms of progression of myeloid preleukemia to transformed myeloid leukemia in children with Down syndrome. *Cancer Cell*. 2019;36(2):123–138.
10. Li Z, et al. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet*. 2005;37(6):613–619.

11. Malinge S, et al. Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *J Clin Invest*. 2012;122(3):948–962.
12. Elagib KE, et al. Calpain 2 activation of P-TEFb drives megakaryocyte morphogenesis and is disrupted by leukemogenic GATA1 mutation. *Dev Cell*. 2013;27(6):607–620.
13. Juban G, et al. Oncogenic Gata1 causes stage-specific megakaryocyte differentiation delay. *Haematologica*. 2021;106(4):1106–1119.
14. Chou ST, et al. Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc Natl Acad Sci U S A*. 2012;109(43):17573–17578.
15. Maclean GA, et al. Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. *Proc Natl Acad Sci U S A*. 2012;109(43):17567–17572.
16. Byrska-Bishop M, et al. Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus. *J Clin Invest*. 2015;125(3):993–1005.
17. Banno K, et al. Systematic cellular disease models reveal synergistic interaction of Trisomy 21 and GATA1 mutations in hematopoietic abnormalities. *Cell Rep*. 2016;15(6):1228–1241.
18. Wagenblast E, et al. Functional profiling of single CRISPR/Cas9-edited human long-term hematopoietic stem cells. *Nat Commun*. 2019;10(1):4730.
19. Wagenblast E, et al. Mapping the cellular origin and early evolution of leukemia in Down syndrome. *Science*. 2021;373(6551):eabf6202.
20. Ling T, et al. Chromatin occupancy and epigenetic analysis reveal new insights into the function of the GATA1 N terminus in erythropoiesis. *Blood*. 2019;134(19):1619–1631.
21. Chlon TM, et al. Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression. *Haematologica*. 2015;100(5):575–584.
22. Mazumdar C, et al. Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell*. 2015;17(6):675–688.
23. Sasca D, et al. Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood*. 2019;134(24):2195–2208.
24. Viny AD, et al. Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell*. 2019;25(5):682–696.
25. Pardanani AD, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood*. 2006;108(10):3472–3476.
26. Defour JP, et al. Oncogenic activation of MPL/thrombopoietin receptor by 17 mutations at W515: implications for myeloproliferative neoplasms. *Leukemia*. 2016;30(5):1214–1216.
27. Bluteau O, et al. Developmental changes in human megakaryopoiesis. *J Thromb Haemost*. 2013;11(9):1730–1741.
28. Langebrake C, et al. Immunophenotype of Down syndrome acute myeloid leukemia and transient myeloproliferative disease differs significantly from other diseases with morphologically identical or similar blasts. *Klin Padiatr*. 2005;217(3):126–134.
29. Wang L, et al. Acute megakaryoblastic leukemia associated with trisomy 21 demonstrates a distinct immunophenotype. *Cytometry B Clin Cytom*. 2015;88(4):244–252.
30. Eckly A, et al. Biogenesis of the demarcation membrane system(DMS) in megakaryocytes. *Blood*. 2014;123(6):921–930.
31. Bourquin JP, et al. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proc Natl Acad Sci U S A*. 2006;103(9):3339–3344.
32. Tijssen MR, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell*. 2011;20(5):597–609.
33. Shivdasani RA, et al. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*. 1995;81(5):695–704.
34. Muntean AG, Crispino JD. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood*. 2005;106(4):1223–1231.
35. Kaya-Okur HS, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*. 2019;10(1):1930.
36. Zuber J, et al. An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes Dev*. 2011;25(15):1628–1640.
37. Takao S, et al. Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *Elife*. 2021;10:e65905.
38. Wang T, et al. Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets. *Cell Stem Cell*. 2021;28(6):1074–1089.
39. Zang C, et al. NF-E2, FLI1 and RUNX1 collaborate at areas of dynamic chromatin to activate transcription in mature mouse megakaryocytes. *Sci Rep*. 2016;6:30255.
40. Potts KS, et al. Membrane budding is a major mechanism of in vivo platelet biogenesis. *J Exp Med*. 2020;217(9):e20191206.
41. Rivas MA, et al. Smc3 dosage regulates B cell transit through germinal centers and restricts their malignant transformation. *Nat Immunol*. 2021;22(2):240–253.
42. Takayama N, et al. Transient activation of c-MYC expression is critical for efficient platelet generation from human induced pluripotent stem cells. *J Exp Med*. 2010;207(13):2817–2830.
43. Muñoz-Alonso MJ, et al. MYC accelerates p21CIP-induced megakaryocytic differentiation involving early mitosis arrest in leukemia cells. *J Cell Physiol*. 2012;227(5):2069–2078.
44. Kadri Z, et al. Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS Biol*. 2009;7(6):e1000123.
45. Klusmann JH, et al. Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis. *Genes Dev*. 2010; 24(15):1659–1672.
46. Gialesaki S, et al. RUNX1 isoform disequilibrium in the development of trisomy 21 associated myeloid leukemia [preprint]. https://doi.org/10.1101/2022.03.07.483334. Posted on bioRxiv March 7, 2022.
47. Gnatenko DV, et al. Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood*. 2003;101(6):2285–2293.

Bibliographie

1. R. Sender, R. Milo, The distribution of cellular turnover in the human body. *Nat Med*. 27, 45–48 (2021).

2. J. E. Till, E. A. McCULLOCH, A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res*. 14, 213–222 (1961).

3. J. Cosgrove, L. S. P. Hustin, R. J. de Boer, L. Perié, Hematopoiesis in numbers. *Trends in Immunology*. 42, 1100–1112 (2021).

4. A. Nakamura-Ishizu, H. Takizawa, T. Suda, The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development*. 141, 4656–4666 (2014).

5. R. A. J. Signer, J. A. Magee, A. Salic, S. J. Morrison, Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature*. 509, 49–54 (2014).

6. M. Hu, J. Wang, Mitochondrial metabolism and the maintenance of hematopoietic stem cell quiescence: *Current Opinion in Hematology*. 26, 228–234 (2019).

7. F. Notta, S. Doulatov, E. Laurenti, A. Poeppl, I. Jurisica, J. E. Dick, Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science*. 333, 218–221 (2011).

8. M. Kondo, I. L. Weissman, K. Akashi, Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell*. 91, 661–672 (1997).

9. K. Akashi, D. Traver, T. Miyamoto, I. L. Weissman, A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 404, 193–197 (2000).

10. N. Cabezas-Wallscheid, D. Klimmeck, J. Hansson, D. B. Lipka, A. Reyes, Q. Wang, D. Weichenhan, A. Lier, L. von Paleske, S. Renders, P. Wünsche, P. Zeisberger, D. Brocks, L. Gu, C. Herrmann, S. Haas, M. A. G. Essers, B. Brors, R. Eils, W. Huber, M. D. Milsom, C. Plass, J. Krijgsveld, A. Trumpp, Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. *Cell Stem Cell*. 15, 507–522 (2014).

11. P. Sommerkamp, M. C. Romero-Mulero, A. Narr, L. Ladel, L. Hustin, K. Schönberger, S. Renders, S. Altamura, P. Zeisberger, K. Jäcklein, D. Klimmeck, A. Rodriguez-Fraticelli, F. D. Camargo, L. Perié, A. Trumpp, N. Cabezas-Wallscheid, Mouse multipotent progenitor 5 cells are located at the interphase between hematopoietic stem and progenitor cells. *Blood*. 137, 3218–3224 (2021).

12. 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, I. Amit, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 163, 1663–1677 (2015).

13. L. Velten, S. F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B. P. Hennig, C. Hirche, C. Lutz, E. C. Buss, D. Nowak, T. Boch, W.-K. Hofmann, A. D. Ho, W. Huber, A. Trumpp, M. A. G. Essers, L. M. Steinmetz, Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 19, 271–281 (2017).

14. D. Karamitros, B. Stoilova, Z. Aboukhalil, F. Hamey, A. Reinisch, M. Samitsch, L. Quek, G. Otto, E. Repapi, J. Doondeea, B. Usukhbayar, J. Calvo, S. Taylor, N. Goardon, E. Six, F. Pflumio, C. Porcher, R. Majeti, B. Göttgens, P. Vyas, Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol*. 19, 85–97 (2018).

15. B. K. Tusi, S. L. Wolock, C. Weinreb, Y. Hwang, D. Hidalgo, R. Zilionis, A. Waisman, J. R. Huh, A. M. Klein, M. Socolovsky, Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*. 555, 54–60 (2018).

16. R. Drissen, S. Thongjuea, K. Theilgaard-Mönch, C. Nerlov, Identification of two distinct pathways of human myelopoiesis. *Sci. Immunol.* 4, eaau7148 (2019).

17. C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, A. M. Klein, Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*. 367, eaaw3381 (2020).

18. J. D. Buenrostro, M. R. Corces, C. A. Lareau, B. Wu, A. N. Schep, M. J. Aryee, R. Majeti, H. Y. Chang, W. J. Greenleaf, Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*. 173, 1535-1548.e16 (2018).

19. C. E. Müller-Sieburg, R. H. Cho, M. Thoman, B. Adkins, H. B. Sieburg, Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood*. 100, 1302–1309 (2002).

20. B. Dykstra, D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S.-J. Lee, R. Brinkman, C. Eaves, Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo. *Cell Stem Cell*. 1, 218–229 (2007).

21. R. Yamamoto, Y. Morita, J. Ooehara, S. Hamanaka, M. Onodera, K. L. Rudolph, H. Ema, H. Nakauchi, Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells. *Cell*. 154, 1112–1126 (2013).

22. L. Perié, K. R. Duffy, L. Kok, R. J. de Boer, T. N. Schumacher, The Branching Point in Erythro-Myeloid Differentiation. *Cell*. 163, 1655–1662 (2015).

23. K. Busch, K. Klapproth, M. Barile, M. Flossdorf, T. Holland-Letz, S. M. Schlenner, M. Reth, T. Höfer, H.-R. Rodewald, Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*. 518, 542–546 (2015).

24. J. Carrelha, Y. Meng, L. M. Kettyle, T. C. Luis, R. Norfo, V. Alcolea, H. Boukarabila, F. Grasso, A. Gambardella, A. Grover, K. Högstrand, A. M. Lord, A. Sanjuan-Pla, P. S.

Woll, C. Nerlov, S. E. W. Jacobsen, Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*. 554, 106–111 (2018).

25. A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein, F. D. Camargo, Clonal analysis of lineage fate in native haematopoiesis. *Nature*. 553, 212–216 (2018).

26. L. A. Liggett, V. G. Sankaran, Unraveling Hematopoiesis through the Lens of Genomics. *Cell*. 182, 1384–1400 (2020).

27. C. Gekas, T. Graf, CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. *Blood*. 121, 4463–4472 (2013).

28. H. Huang, A. B. Cantor, Common features of megakaryocytes and hematopoietic stem cells: What's the connection? *J. Cell. Biochem.* 107, 857–864 (2009).

29. S. Haas, J. Hansson, D. Klimmeck, D. Loeffler, L. Velten, H. Uckelmann, S. Wurzer, Á. M. Prendergast, A. Schnell, K. Hexel, R. Santarella-Mellwig, S. Blaszkiewicz, A. Kuck, H. Geiger, M. D. Milsom, L. M. Steinmetz, T. Schroeder, A. Trumpp, J. Krijgsveld, M. A. G. Essers, Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell*. 17, 422–434 (2015).

30. J. Tong, T. Sun, S. Ma, Y. Zhao, M. Ju, Y. Gao, P. Zhu, P. Tan, R. Fu, A. Zhang, D. Wang, D. Wang, Z. Xiao, J. Zhou, R. Yang, S. J. Loughran, J. Li, A. R. Green, E. H. Bresnick, D. Wang, T. Cheng, L. Zhang, L. Shi, Hematopoietic Stem Cell Heterogeneity Is Linked to the Initiation and Therapeutic Response of Myeloproliferative Neoplasms. *Cell Stem Cell*. 28, 502-513.e6 (2021).

31. S. Mazzi, L. Lordier, N. Debili, H. Raslova, W. Vainchenker, Megakaryocyte and polyploidization. *Experimental Hematology*. 57, 1–13 (2018).

32. D. Ghalloussi, A. Dhenge, W. Bergmeier, New insights into cytoskeletal remodeling during platelet production. *J Thromb Haemost*. 17, 1430–1439 (2019).

33. T. Junt, H. Schulze, Z. Chen, S. Massberg, T. Goerge, A. Krueger, D. D. Wagner, T. Graf, J. E. Italiano, R. A. Shivdasani, U. H. von Andrian, Dynamic Visualization of Thrombopoiesis Within Bone Marrow. *Science*. 317, 1767–1770 (2007).

34. K. S. Potts, A. Farley, C. A. Dawson, J. Rimes, C. Biben, C. de Graaf, M. A. Potts, O. J. Stonehouse, A. Carmagnac, P. Gangatirkar, E. C. Josefsson, C. Anttila, D. Amann-Zalcenstein, S. Naik, W. S. Alexander, D. J. Hilton, E. D. Hawkins, S. Taoudi, Membrane budding is a major mechanism of in vivo platelet biogenesis. *Journal of Experimental Medicine*. 217, e20191206 (2020).

35. S. Nishimura, M. Nagasaki, S. Kunishima, A. Sawaguchi, A. Sakata, H. Sakaguchi, T. Ohmori, I. Manabe, J. E. Italiano, T. Ryu, N. Takayama, I. Komuro, T. Kadowaki, K. Eto, R. Nagai, IL-1α induces thrombopoiesis through megakaryocyte rupture in response to acute platelet needs. *J. Cell Biol.* 209, 453–466 (2015).

36. **J. E. Italiano, M. Bender, G. Merrill-Skoloff, C. Ghevaert, B. Nieswandt, R. Flaumenhaft, Microvesicles, but not platelets, bud off from mouse bone marrow megakaryocytes. *Blood*. 138, 1998–2001 (2021).**

37. **E. Lefrançais, G. Ortiz-Muñoz, A. Caudrillier, B. Mallavia, F. Liu, D. M. Sayah, E. E. Thornton, M. B. Headley, T. David, S. R. Coughlin, M. F. Krummel, A. D. Leavitt, E. Passegué, M. R. Looney, The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. *Nature*. 544, 105–109 (2017).**

38. **H. Wang, Decoding Human Megakaryocyte Development. *Cell Stem Cell*. 28, 1–15 (2021).**

39. **S. Sun, C. Jin, J. Si, Y. Lei, K. Chen, Y. Cui, Z. Liu, J. Liu, M. Zhao, X. Zhang, F. Tang, M. T. Rondina, Y. Li, Q. Wang, Single-cell analysis of ploidy and the transcriptome reveals functional and spatial divergency in murine megakaryopoiesis. *Blood*. 138, 1211–1224 (2021).**

40. **C. Liu, D. Wu, M. Xia, M. Li, Z. Sun, B. Shen, Y. Liu, E. Jiang, H. Wang, P. Su, L. Shi, Z. Xiao, X. Zhu, W. Zhou, Q. Wang, X. Gao, T. Cheng, J. Zhou, Characterization of Cellular Heterogeneity and an Immune Subpopulation of Human Megakaryocytes. *Adv. Sci.* 8, 2100921 (2021).**

41. **D. N. Pariser, Z. T. Hilt, S. K. Ture, S. K. Blick-Nitko, M. R. Looney, S. J. Cleary, E. Roman-Pagan, J. Saunders, S. N. Georas, J. Veazey, F. Madere, L. T. Santos, A. Arne, N. P. T. Huynh, A. C. Livada, S. M. Guerrero-Martin, C. Lyons, K. A. Metcalf-Pate, K. E. McGrath, J. Palis, C. N. Morrell, Lung megakaryocytes are immune modulatory cells. *Journal of Clinical Investigation*. 131, e137377 (2021).**

42. **A. K. Yeung, C. Villacorta-Martin, S. Hon, J. R. Rock, G. J. Murphy, Lung megakaryocytes display distinct transcriptional and phenotypic properties. *Blood Advances*. 4, 6204–6217 (2020).**

43. **J. Tilburg, I. C. Becker, J. E. Italiano, *Blood*, in press, doi:10.1182/blood.2020009302.**

44. **A. L. Caulier, V. G. Sankaran, *Blood*, in press, doi:10.1182/blood.2021011044.**

45. **N. Kwon, E. N. Thompson, M. Y. Mayday, V. Scanlon, Y.-C. Lu, D. S. Krause, Current understanding of human megakaryocytic-erythroid progenitors and their fate determinants. *Current Opinion in Hematology*. 28, 28–35 (2021).**

46. **P. D. F. Murray, The development *in vitro* of the blood of the early chick embryo. *Proc. R. Soc. Lond. B.* 111, 497–521 (1932).**

47. **K. Choi, M. Kennedy, A. Kazarov, J. C. Papadimitriou, G. Keller, A common precursor for hematopoietic and endothelial cells. *Development*. 125, 725–732 (1998).**

48. **L. Padrón-Barthe, S. Temiño, C. Villa del Campo, L. Carramolino, J. Isern, M. Torres, Clonal analysis identifies hemogenic endothelium as the source of the blood-endothelial common lineage in the mouse embryo. *Blood*. 124, 2523–2532 (2014).**

49. M. Stefanska, K. Batta, R. Patel, M. Florkowska, V. Kouskoff, G. Lacaud, Primitive erythrocytes are generated from hemogenic endothelial cells. *Sci Rep*. 7, 6401 (2017).

50. T. Jaffredo, R. Gautier, A. Eichmann, F. Dieterlen-Lievre, Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*. 125, 4575–4583 (1998).

51. K. Kissa, P. Herbomel, Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*. 464, 112–115 (2010).

52. J. Y. Bertrand, N. C. Chi, B. Santoso, S. Teng, D. Y. R. Stainier, D. Traver, Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*. 464, 108–111 (2010).

53. J.-C. Boisset, W. van Cappellen, C. Andrieu-Soler, N. Galjart, E. Dzierzak, C. Robin, In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*. 464, 116–120 (2010).

54. H. M. Eilken, S.-I. Nishikawa, T. Schroeder, Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature*. 457, 896–900 (2009).

55. T. Yokomizo, E. Dzierzak, Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development*. 137, 3651–3661 (2010).

56. K. E. Rhodes, C. Gekas, Y. Wang, C. T. Lux, C. S. Francis, D. N. Chan, S. Conway, S. H. Orkin, M. C. Yoder, H. K. A. Mikkola, The Emergence of Hematopoietic Stem Cells Is Initiated in the Placental Vasculature in the Absence of Circulation. *Cell Stem Cell*. 2, 252–263 (2008).

57. A. D. Yzaguirre, N. A. Speck, Insights into blood cell formation from hemogenic endothelium in lesser-known anatomic sites: 3D Analysis of Embryonic Hematopoiesis. *Dev. Dyn.* 245, 1011–1028 (2016).

58. K. Ottersbach, Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochemical Society Transactions*. 47, 591–601 (2019).

59. G. I. Uenishi, H. S. Jung, A. Kumar, M. A. Park, B. K. Hadland, E. McLeod, M. Raymond, O. Moskvin, C. E. Zimmerman, D. J. Theisen, S. Swanson, O. J. Tamplin, L. I. Zon, J. A. Thomson, I. D. Bernstein, I. I. Slukvin, NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat Commun*. 9, 1828 (2018).

60. F. Bonkhofer, R. Rispoli, P. Pinheiro, M. Krecsmarik, J. Schneider-Swales, I. H. C. Tsang, M. de Bruijn, R. Monteiro, T. Peterkin, R. Patient, Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. *Nat Commun*. 10, 3577 (2019).

61. **M. A. Park, A. Kumar, H. S. Jung, G. Uenishi, O. V. Moskvin, J. A. Thomson, I. I. Slukvin, Activation of the Arterial Program Drives Development of Definitive Hemogenic Endothelium with Lymphoid Potential. *Cell Reports*. 23, 2467–2481 (2018).**

62. **C. Pouget, R. Gautier, M.-A. Teillet, T. Jaffredo, Somite-derived cells replace ventral aortic hemangioblasts and provide aortic smooth muscle cells of the trunk. *Development*. 133, 1013–1022 (2006).**

63. **L. Yvernogeau, R. Gautier, L. Petit, H. Khoury, F. Relaix, V. Ribes, H. Sang, P. Charbord, M. Souyri, C. Robin, T. Jaffredo, In vivo generation of haematopoietic stem/progenitor cells from bone marrow-derived haemogenic endothelium. *Nat Cell Biol*. 21, 1334–1345 (2019).**

64. **L. Lange, M. Morgan, A. Schambach, The hemogenic endothelium: a critical source for the generation of PSC-derived hematopoietic stem and progenitor cells. *Cell. Mol. Life Sci.* 78, 4143–4160 (2021).**

65. **J. Palis, Development of hematopoietic progenitors. *Development*. 126, 5073–84 (1999).**

66. **J. Palis, Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS Lett*. 590, 3965–3974 (2016).**

67. **M. H. Baron, Concise Review: Early Embryonic Erythropoiesis: Not so Primitive After All. *Stem Cells*. 31, 849–856 (2013).**

68. **J. Palis, Primitive and definitive erythropoiesis in mammals. *Front. Physiol.* 5 (2014), doi:10.3389/fphys.2014.00003.**

69. **P. D. Kingsley, J. Malik, K. A. Fantauzzo, J. Palis, Yolk sac–derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood*. 104, 19–25 (2004).**

70. **J. Tober, A. Koniski, K. E. McGrath, R. Vemishetti, R. Emerson, K. K. L. de Mesy-Bentley, R. Waugh, J. Palis, The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood*. 109, 1433–1441 (2007).**

71. **J. Y. Bertrand, A. Jalil, M. Klaine, S. Jung, A. Cumano, I. Godin, Three pathways to mature macrophages in the early mouse yolk sac. *Blood*. 106, 3004–3011 (2005).**

72. **D. A. Ovchinnikov, Macrophages in the embryo and beyond: Much more than just giant phagocytes. *genesis*. 46, 447–462 (2008).**

73. **G. Hoeffel, J. Chen, Y. Lavin, D. Low, F. F. Almeida, P. See, A. E. Beaudin, J. Lum, I. Low, E. C. Forsberg, M. Poidinger, F. Zolezzi, A. Larbi, L. G. Ng, J. K. Y. Chan, M. Greter, B. Becher, I. M. Samokhvalov, M. Merad, F. Ginhoux, C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. *Immunity*. 42, 665–678 (2015).**

74. **G. Lacaud, Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. *Blood*. 100, 458–466 (2002).**

75. **C. Barone, R. Orsenigo, R. Meneveri, S. Brunelli, E. Azzoni, One Size Does Not Fit All: Heterogeneity in Developmental Hematopoiesis. *Cells*, 22 (2022).**

76. **J. Palis, R. J. Chan, A. Koniski, R. Patel, M. Starr, M. C. Yoder, Spatial and temporal emergence of high proliferative potential hematopoietic precursors during murine embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4528–4533 (2001).**

77. **C. Böiers, J. Carrelha, M. Lutteropp, S. Luc, J. C. A. Green, E. Azzoni, P. S. Woll, A. J. Mead, A. Hultquist, G. Swiers, E. G. Perdiguero, I. C. Macaulay, L. Melchiori, T. C. Luis, S. Kharazi, T. Bouriez-Jones, Q. Deng, A. Pontén, D. Atkinson, C. T. Jensen, E. Sitnicka, F. Geissmann, I. Godin, R. Sandberg, M. F. T. R. de Bruijn, S. E. W. Jacobsen, Lymphomyeloid Contribution of an Immune-Restricted Progenitor Emerging Prior to Definitive Hematopoietic Stem Cells. *Cell Stem Cell*. 13, 535–548 (2013).**

78. **K. E. McGrath, J. M. Frame, K. H. Fegan, J. R. Bowen, S. J. Conway, S. C. Catherman, P. D. Kingsley, A. D. Koniski, J. Palis, Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports*. 11, 1892–1904 (2015).**

79. **L. Iturri, L. Freyer, A. Biton, P. Dardenne, Y. Lallemand, E. Gomez Perdiguero, Megakaryocyte production is sustained by direct differentiation from erythromyeloid progenitors in the yolk sac until midgestation. *Immunity*. 54, 1433-1446.e5 (2021).**

80. **E. Azzoni, V. Frontera, K. E. McGrath, J. Harman, J. Carrelha, C. Nerlov, J. Palis, S. E. W. Jacobsen, M. F. Bruijn, Kit ligand has a critical role in mouse yolk sac and aorta–gonad–mesonephros hematopoiesis. *EMBO Rep*. 19 (2018), doi:10.15252/embr.201745477.**

81. **F. Soares-da-Silva, L. Freyer, R. Elsaid, O. Burlen-Defranoux, L. Iturri, O. Sismeiro, P. Pinto-do-Ó, E. Gomez-Perdiguero, A. Cumano, Yolk sac, but not hematopoietic stem cell–derived progenitors, sustain erythropoiesis throughout murine embryonic life. *Journal of Experimental Medicine*. 218, e20201729 (2021).**

82. **E. Gomez Perdiguero, K. Klapproth, C. Schulz, K. Busch, E. Azzoni, L. Crozet, H. Garner, C. Trouillet, M. F. de Bruijn, F. Geissmann, H.-R. Rodewald, Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*. 518, 547–551 (2015).**

83. **C. Dege, K. H. Fegan, J. P. Creamer, M. M. Berrien-Elliott, S. A. Luff, D. Kim, J. A. Wagner, P. D. Kingsley, K. E. McGrath, T. A. Fehniger, J. Palis, C. M. Sturgeon, Potently Cytotoxic Natural Killer Cells Initially Emerge from Erythro-Myeloid Progenitors during Mammalian Development. *Developmental Cell*. 53, 229-239.e7 (2020).**

84. **T. C. Luis, S. Luc, T. Mizukami, H. Boukarabila, S. Thongjuea, P. S. Woll, E. Azzoni, A. Giustacchini, M. Lutteropp, T. Bouriez-Jones, H. Vaidya, A. J. Mead, D. Atkinson, C. Böiers, J. Carrelha, I. C. Macaulay, R. Patient, F. Geissmann, C. Nerlov, R. Sandberg, M. F. T. R. de Bruijn, C. C. Blackburn, I. Godin, S. E. W. Jacobsen, Initial seeding of the**

embryonic thymus by immune-restricted lympho-myeloid progenitors. *Nat Immunol*. 17, 1424–1435 (2016).

85. M. Yoshimoto, P. Porayette, N. L. Glosson, S. J. Conway, N. Carlesso, A. A. Cardoso, M. H. Kaplan, M. C. Yoder, Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood*. 119, 5706–5714 (2012).

86. M. Yoshimoto, E. Montecino-Rodriguez, M. J. Ferkowicz, P. Porayette, W. C. Shelley, S. J. Conway, K. Dorshkind, M. C. Yoder, Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1468–1473 (2011).

87. B. Hadland, M. Yoshimoto, Many layers of embryonic hematopoiesis: new insights into B-cell ontogeny and the origin of hematopoietic stem cells. *Experimental Hematology*. 60, 1–9 (2018).

88. A. M. Müller, A. Medvinsky, J. Strouboulis, F. Grosveld, E. Dzierzakt, Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*. 1, 291–301 (1994).

89. A. Ivanovs, S. Rybtsov, L. Welch, R. A. Anderson, M. L. Turner, A. Medvinsky, Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *Journal of Experimental Medicine*. 208, 2417–2427 (2011).

90. A. L. Medvinsky, N. L. Samoylina, A. M. Müller, E. A. Dzierzak, An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature*. 364, 64–67 (1993).

91. A. Medvinsky, E. Dzierzak, Definitive Hematopoiesis Is Autonomously Initiated by the AGM Region. *Cell*. 86, 897–906 (1996).

92. M. F. T. R. de Bruijn, Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *The EMBO Journal*. 19, 2465–2474 (2000).

93. J.-C. Boisset, T. Clapes, A. Klaus, N. Papazian, J. Onderwater, M. Mommaas-Kienhuis, T. Cupedo, C. Robin, Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta. *Blood*. 125, 465–469 (2015).

94. F. Zhou, X. Li, W. Wang, P. Zhu, J. Zhou, W. He, M. Ding, F. Xiong, X. Zheng, Z. Li, Y. Ni, X. Mu, L. Wen, T. Cheng, Y. Lan, W. Yuan, F. Tang, B. Liu, Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*. 533, 487–492 (2016).

95. C. S. Baron, L. Kester, A. Klaus, J.-C. Boisset, R. Thambyrajah, L. Yvernogeau, V. Kouskoff, G. Lacaud, A. van Oudenaarden, C. Robin, Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nature Communications*. 9 (2018), doi:10.1038/s41467-018-04893-3.

96. S. Taoudi, A. Medvinsky, Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9399–9403 (2007).

97. **A. Medvinsky, S. Rybtsov, S. Taoudi, Embryonic origin of the adult hematopoietic system: advances and questions. *Development*. 138, 1017–1031 (2011).**

98. **S. Rybtsov, A. Ivanovs, S. Zhao, A. Medvinsky, Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. *Development*. 143, 1284–1289 (2016).**

99. **Q. Zhu, P. Gao, J. Tober, L. Bennett, C. Chen, Y. Uzun, Y. Li, E. D. Howell, M. Mumau, W. Yu, B. He, N. A. Speck, K. Tan, Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood*. 136, 12 (2020).**

100. **H. Ema, H. Nakauchi, Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood*. 95, 2284–2288 (2000).**

101. **S. Pinho, P. S. Frenette, Haematopoietic stem cell activity and interactions with the niche. *Nat Rev Mol Cell Biol*. 20, 303–320 (2019).**

102. **X. Gao, C. Xu, N. Asada, P. S. Frenette, The hematopoietic stem cell niche: from embryo to adult. *Development*. 145, dev139691 (2018).**

103. **S. Pinho, T. Marchand, E. Yang, Q. Wei, C. Nerlov, P. S. Frenette, Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. *Developmental Cell*. 44, 634-641.e4 (2018).**

104. **A. P. Stone, T. F. Nascimento, M. N. Barrachina, The bone marrow niche from the inside out: how megakaryocytes are shaped by and shape hematopoiesis. *Blood*. 139, 483–491 (2022).**

105. **O. Kandarakov, A. Belyavsky, E. Semenova, Bone Marrow Niches of Hematopoietic Stem and Progenitor Cells. *IJMS*. 23, 4462 (2022).**

106. **K. Hori, A. Sen, S. Artavanis-Tsakonas, *Journal of Cell Science*, in press, doi:10.1242/jcs.127308.**

107. **Y. Ge, J. Wang, H. Zhang, J. Li, M. Ye, X. Jin, Fate of hematopoietic stem cells determined by Notch1 signaling (Review). *Exp Ther Med*. 23, 170 (2021).**

108. **L. Gama-Norton, E. Ferrando, C. Ruiz-Herguido, Z. Liu, J. Guiu, A. B. M. M. K. Islam, S.-U. Lee, M. Yan, C. J. Guidos, N. López-Bigas, T. Maeda, L. Espinosa, R. Kopan, A. Bigas, Notch signal strength controls cell fate in the haemogenic endothelium. *Nat Commun*. 6, 8510 (2015).**

109. **K. V. Pajcini, N. A. Speck, W. S. Pear, Notch signaling in mammalian hematopoietic stem cells. *Leukemia*. 25, 1525–1532 (2011).**

110. **A. Vanderbeck, I. Maillard, Notch signaling at the crossroads of innate and adaptive immunity. *J Leukoc Biol*. 109, 535–548 (2021).**

111. T. Mercher, M. G. Cornejo, C. Sears, T. Kindler, S. A. Moore, I. Maillard, W. S. Pear, J. C. Aster, D. G. Gilliland, Notch Signaling Specifies Megakaryocyte Development from Hematopoietic Stem Cells. *Cell Stem Cell*. 3, 314–326 (2008).

112. P. Oh, C. Lobry, J. Gao, A. Tikhonova, E. Loizou, J. Manent, B. van Handel, S. Ibrahim, J. Greve, H. Mikkola, S. Artavanis-Tsakonas, I. Aifantis, In Vivo Mapping of Notch Pathway Activity in Normal and Stress Hematopoiesis. *Cell Stem Cell*. 13, 190–204 (2013).

113. I. Van de Walle, G. De Smet, M. Gärtner, M. De Smedt, E. Waegemans, B. Vandekerckhove, G. Leclercq, J. Plum, J. C. Aster, I. D. Bernstein, C. J. Guidos, B. Kyewski, T. Taghon, Jagged2 acts as a Delta-like Notch ligand during early hematopoietic cell fate decisions. *Blood*. 117, 4449–4459 (2011).

114. K. Behrens, W. S. Alexander, Cytokine control of megakaryopoiesis. *Growth Factors*. 36, 89–103 (2018).

115. F. J. de Sauvage, K. Carver-Moore, S. M. Luoh, A. Ryan, M. Dowd, D. L. Eaton, M. W. Moore, Physiological regulation of early and late stages of megakaryocytopoiesis by thrombopoietin. *Journal of Experimental Medicine*. 183, 651–656 (1996).

116. M. Decker, J. Leslie, Q. Liu, L. Ding, Hepatic thrombopoietin is required for bone marrow hematopoietic stem cell maintenance. *Science*. 360, 106–110 (2018).

117. I. S. Hitchcock, K. Kaushansky, Thrombopoietin from beginning to end. *Br J Haematol*. 165, 259–268 (2014).

118. B. J. Lannutti, J. G. Drachman, Lyn tyrosine kinase regulates thrombopoietin-induced proliferation of hematopoietic cell lines and primary megakaryocytic progenitors. *Blood*. 103, 3736–3743 (2004).

119. I. S. Hitchcock, M. M. Chen, J. R. King, K. Kaushansky, YRRL motifs in the cytoplasmic domain of the thrombopoietin receptor regulate receptor internalization and degradation. *Blood*. 112, 2222–2231 (2008).

120. H. Yoshihara, F. Arai, K. Hosokawa, T. Hagiwara, K. Takubo, Y. Nakamura, Y. Gomei, H. Iwasaki, S. Matsuoka, K. Miyamoto, H. Miyazaki, T. Takahashi, T. Suda, Thrombopoietin/MPL Signaling Regulates Hematopoietic Stem Cell Quiescence and Interaction with the Osteoblastic Niche. *Cell Stem Cell*. 1, 685–697 (2007).

121. A. Nakamura-Ishizu, T. Matsumura, P. S. Stumpf, T. Umemoto, H. Takizawa, Y. Takihara, A. O'Neil, A. B. B. A. Majeed, B. D. MacArthur, T. Suda, Thrombopoietin Metabolically Primes Hematopoietic Stem Cells to Megakaryocyte-Lineage Differentiation. *Cell Reports*. 25, 1772-1785.e6 (2018).

122. L. N. Varghese, J.-P. Defour, C. Pecquet, S. N. Constantinescu, The Thrombopoietin Receptor: Structural Basis of Traffic and Activation by Ligand, Mutations, Agonists, and Mutated Calreticulin. *Front. Endocrinol.* 8 (2017), doi:10.3389/fendo.2017.00059.

123. S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*. 16, 129–143 (2015).

124. K. M. Lelli, M. Slattery, R. S. Mann, Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annu. Rev. Genet.* 46, 43–68 (2012).

125. B. Göttgens, Regulatory network control of blood stem cells. *Blood*. 125, 2614–2620 (2015).

126. M. R. Tijssen, C. Ghevaert, Transcription factors in late megakaryopoiesis and related platelet disorders. *J Thromb Haemost*. 11, 593–604 (2013).

127. S. H. Orkin, L. I. Zon, Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*. 132, 631–644 (2008).

128. C. G. Palii, Q. Cheng, M. A. Gillespie, P. Shannon, M. Mazurczyk, G. Napolitani, N. D. Price, J. A. Ranish, E. Morrissey, D. R. Higgs, M. Brand, Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell*. 24, 812-820.e5 (2019).

129. M. Brand, E. Morrissey, Single-cell fate decisions of bipotential hematopoietic progenitors. *Current Opinion in Hematology*. 27, 232–240 (2020).

130. B. Edginton-White, C. Bonifer, The transcriptional regulation of normal and malignant blood cell development. *The FEBS Journal*. 289, 1240–1255 (2022).

131. W. Y. I. Chan, G. A. Follows, G. Lacaud, J. E. Pimanda, J.-R. Landry, S. Kinston, K. Knezevic, S. Piltz, I. J. Donaldson, L. Gambardella, F. Sablitzky, A. R. Green, V. Kouskoff, B. Göttgens, The paralogous hematopoietic regulators Lyl1 and Scl are coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl–/– phenotype. *Blood*. 109, 1908–1916 (2007).

132. J. E. Pimanda, K. Ottersbach, K. Knezevic, S. Kinston, W. Y. I. Chan, N. K. Wilson, J.-R. Landry, A. D. Wood, A. Kolb-Kokocinski, A. R. Green, D. Tannahill, G. Lacaud, V. Kouskoff, B. Göttgens, Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17692–17697 (2007).

133. N. K. Wilson, S. D. Foster, X. Wang, K. Knezevic, J. Schütte, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand, E. Dzierzak, J. E. Pimanda, M. F. T. R. de Bruijn, B. Göttgens, Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell*. 7, 532–544 (2010).

134. M. Pimkin, A. V. Kossenkov, T. Mishra, C. S. Morrissey, W. Wu, C. A. Keller, G. A. Blobel, D. Lee, M. A. Beer, R. C. Hardison, M. J. Weiss, Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Research*. 24, 1932–1944 (2014).

135. M. R. Tijssen, A. Cvejic, A. Joshi, R. L. Hannah, R. Ferreira, A. Forrai, D. C. Bellissimo, S. H. Oram, P. A. Smethurst, N. K. Wilson, X. Wang, K. Ottersbach, D. L. Stemple, A. R. Green, W. H. Ouwehand, B. Göttgens, Genome-wide Analysis of Simultaneous GATA1/2, RUNX1, FLI1, and SCL Binding in Megakaryocytes Identifies Hematopoietic Regulators. *Developmental Cell*. 20, 597–609 (2011).

136. S. J. Loughran, E. A. Kruse, D. F. Hacking, C. A. de Graaf, C. D. Hyland, T. A. Willson, K. J. Henley, S. Ellis, A. K. Voss, D. Metcalf, D. J. Hilton, W. S. Alexander, B. T. Kile, The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol*. 9, 810–819 (2008).

137. K. R. Katsumura, E. H. Bresnick, the GATA Factor Mechanisms Group, The GATA factor revolution in hematology. *Blood*. 129, 2092–2102 (2017).

138. A. P. Tsang, J. E. Visvader, C. A. Turner, Y. Fujiwara, C. Yu, M. J. Weiss, M. Crossley, S. H. Orkin, FOG, a Multitype Zinc Finger Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic Differentiation. *Cell*. 90, 109–119 (1997).

139. H.-Y. Lee, K. D. Johnson, T. Fujiwara, M. E. Boyer, S.-I. Kim, E. H. Bresnick, Controlling Hematopoiesis through Sumoylation-Dependent Regulation of a GATA Factor. *Molecular Cell*. 36, 984–995 (2009).

140. J. M. Lamonica, C. R. Vakoc, G. A. Blobel, Acetylation of GATA-1 is required for chromatin occupancy. *Blood*. 108, 3736–3738 (2006).

141. L. C. Doré, T. M. Chlon, C. D. Brown, K. P. White, J. D. Crispino, Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. *Blood*. 119, 3724–3733 (2012).

142. M. J. Weiss, S. H. Orkin, Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9623–9627 (1995).

143. R. A. Shivdasani, Y. Fujiwara, M. A. McDevitt, S. H. Orkin, A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J*. 16, 3965–3973 (1997).

144. M. A. McDevitt, R. A. Shivdasani, Y. Fujiwara, H. Yang, S. H. Orkin, A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6781–6785 (1997).

145. E. Shin, J.-G. Jeong, H. Chung, H. Jung, C. Park, S. R. Yoon, T.-D. Kim, S. J. Lee, I. Choi, J.-Y. Noh, The Gata1 murine megakaryocyte–erythroid progenitor cells expand robustly and alter differentiation potential. *Biochemical and Biophysical Research Communications*. 528, 46–53 (2020).

146. P. Vyas, K. Ault, C. W. Jackson, S. H. Orkin, R. A. Shivdasani, Consequences of GATA-1 Deficiency in Megakaryocytes and Platelets. *Blood*. 93, 2867–2875 (1999).

147. C. R. Clapier, J. Iwasa, B. R. Cairns, C. L. Peterson, Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*. 18, 407–422 (2017).

148. L. Han, V. Madan, A. Mayakonda, P. Dakle, T. W. Woon, P. Shyamsunder, H. B. M. Nordin, Z. Cao, J. Sundaresan, I. Lei, Z. Wang, H. P. Koeffler, Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. *Leukemia*. 33, 2291–2305 (2019).

149. L. Liu, X. Wan, P. Zhou, X. Zhou, W. Zhang, X. Hui, X. Yuan, X. Ding, R. Zhu, G. Meng, H. Xiao, F. Ma, H. Huang, X. Song, B. Zhou, S. Xiong, Y. Zhang, The chromatin remodeling subunit Baf200 promotes normal hematopoiesis and inhibits leukemogenesis. *J Hematol Oncol*. 11, 27 (2018).

150. X.-J. Sun, N. Man, Y. Tan, S. D. Nimer, L. Wang, The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis. *Front. Oncol.* 5 (2015), doi:10.3389/fonc.2015.00108.

151. W.-I. Chan, R. L. Hannah, M. A. Dawson, C. Pridans, D. Foster, A. Joshi, B. Göttgens, J. M. Van Deursen, B. J. P. Huntly, The Transcriptional Coactivator Cbp Regulates Self-Renewal and Differentiation in Adult Hematopoietic Stem Cells. *Mol Cell Biol*. 31, 5046–5060 (2011).

152. S. Sharma, G. Gurudutta, Epigenetic Regulation of Hematopoietic Stem Cells. *IJSC*. 9, 36–43 (2016).

153. A. V. Krivtsov, S. A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer*. 7, 823–833 (2007).

154. L. Herviou, G. Cavalli, G. Cartron, B. Klein, J. Moreaux, EZH2 in normal hematopoiesis and hematological malignancies. *Oncotarget*. 7, 2284–2296 (2016).

155. I. Boltsis, F. Grosveld, G. Giraud, P. Kolovos, Chromatin Conformation in Development and Disease. *Front. Cell Dev. Biol.* 9, 723859 (2021).

156. J. Langstein, M. D. Milsom, D. B. Lipka, Impact of DNA methylation programming on normal and pre-leukemic hematopoiesis. *Seminars in Cancer Biology*. 51, 89–100 (2018).

157. M. J. Ziller, H. Gu, F. Müller, J. Donaghey, L. T.-Y. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, A. Gnirke, A. Meissner, Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 500, 477–481 (2013).

158. M. Farlik, F. Halbritter, F. Müller, F. A. Choudry, P. Ebert, J. Klughammer, S. Farrow, A. Santoro, V. Ciaurro, A. Mathur, R. Uppal, H. G. Stunnenberg, W. H. Ouwehand, E.

Laurenti, T. Lengauer, M. Frontini, C. Bock, DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell*. 19, 808–822 (2016).

159. C. Bock, I. Beerman, W.-H. Lien, Z. D. Smith, H. Gu, P. Boyle, A. Gnirke, E. Fuchs, D. J. Rossi, A. Meissner, DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Molecular Cell*. 47, 633–647 (2012).

160. G. A. Challen, D. Sun, A. Mayle, M. Jeong, M. Luo, B. Rodriguez, C. Mallaney, H. Celik, L. Yang, Z. Xia, S. Cullen, J. Berg, Y. Zheng, G. J. Darlington, W. Li, M. A. Goodell, Dnmt3a and Dnmt3b Have Overlapping and Distinct Functions in Hematopoietic Stem Cells. *Cell Stem Cell*. 15, 350–364 (2014).

161. K. Moran-Crusio, L. Reavie, A. Shih, O. Abdel-Wahab, D. Ndiaye-Lobry, C. Lobry, M. E. Figueroa, A. Vasanthakumar, J. Patel, X. Zhao, F. Perna, S. Pandey, J. Madzo, C. Song, Q. Dai, C. He, S. Ibrahim, M. Beran, J. Zavadil, S. D. Nimer, A. Melnick, L. A. Godley, I. Aifantis, R. L. Levine, Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*. 20, 11–24 (2011).

162. X. Zhang, J. Su, M. Jeong, M. Ko, Y. Huang, H. J. Park, A. Guzman, Y. Lei, Y.-H. Huang, A. Rao, W. Li, M. A. Goodell, DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet*. 48, 1014–1023 (2016).

163. B. M. Javierre, O. S. Burren, S. P. Wilder, R. Kreuzhuber, S. M. Hill, S. Sewitz, J. Cairns, S. W. Wingett, C. Várnai, M. J. Thiecke, F. Burden, S. Farrow, A. J. Cutler, K. Rehnström, K. Downes, L. Grassi, M. Kostadima, P. Freire-Pritchett, F. Wang, H. G. Stunnenberg, J. A. Todd, D. R. Zerbino, O. Stegle, W. H. Ouwehand, M. Frontini, C. Wallace, M. Spivakov, P. Fraser, J. H. Martens, B. Kim, N. Sharifi, E. M. Janssen-Megens, M.-L. Yaspo, M. Linser, A. Kovacsovics, L. Clarke, D. Richardson, A. Datta, P. Flicek, Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 167, 1369-1384.e19 (2016).

164. A. Kloetgen, P. Thandapani, A. Tsirigos, I. Aifantis, 3D Chromosomal Landscapes in Hematopoiesis and Immunity. *Trends in Immunology*. 40, 809–824 (2019).

165. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 169, 930-944.e22 (2017).

166. J. A. Beagan, J. E. Phillips-Cremins, On the existence and functionality of topologically associating domains. *Nat Genet*. 52, 8–16 (2020).

167. C. Chen, W. Yu, J. Tober, P. Gao, B. He, K. Lee, T. Trieu, G. A. Blobel, N. A. Speck, K. Tan, Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem Cells. *Cell Reports*. 29, 4200-4211.e7 (2019).

168. A. D. Viny, R. L. Bowman, Y. Liu, V.-P. Lavallée, S. E. Eisman, W. Xiao, B. H. Durham, A. Navitski, J. Park, S. Braunstein, B. Alija, A. Karzai, I. S. Csete, M. Witkin, E. Azizi, T.

Baslan, C. J. Ott, D. Pe'er, J. Dekker, R. Koche, R. L. Levine, Cohesin Members Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and Differentiation. *Cell Stem Cell*. 25, 682-696.e8 (2019).

169. M. R. Corces, J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti, H. Y. Chang, Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 48, 1193–1203 (2016).

170. D. Lara-Astiaso, A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Keren-Shaul, A. Mildner, D. Winter, S. Jung, N. Friedman, I. Amit, Chromatin state dynamics during blood formation. *Science*. 345, 943–949 (2014).

171. K. S. Zaret, Pioneer Transcription Factors Initiating Gene Network Changes. *Annu. Rev. Genet.* 54, 367–385 (2020).

172. M. B. Bowie, D. G. Kent, B. Dykstra, K. D. McKnight, L. McCaffrey, P. A. Hoodless, C. J. Eaves, Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5878–5882 (2007).

173. S. Chou, H. F. Lodish, Fetal liver hepatic progenitors are supportive stromal cells for hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7799–7804 (2010).

174. H. Cao, B. Cao, C. K. Heazlewood, M. Domingues, X. Sun, E. Debele, N. E. McGregor, N. A. Sims, S. Y. Heazlewood, S. K. Nilsson, Osteopontin is An Important Regulative Component of the Fetal Bone Marrow Hematopoietic Stem Cell Niche. *Cells*. 8, 985 (2019).

175. S. McKinney-Freeman, P. Cahan, H. Li, S. A. Lacadie, H.-T. Huang, M. Curran, S. Loewer, O. Naveiras, K. L. Kathrein, M. Konantz, E. M. Langdon, C. Lengerke, L. I. Zon, J. J. Collins, G. Q. Daley, The Transcriptional Landscape of Hematopoietic Stem Cell Ontogeny. *Cell Stem Cell*. 11, 701–714 (2012).

176. Y. Li, W. Kong, W. Yang, R. M. Patel, E. B. Casey, T. Okeyo-Owuor, J. M. White, S. N. Porter, S. A. Morris, J. A. Magee, Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth. *Cell Stem Cell*. 27, 732-747.e7 (2020).

177. I. Kim, T. L. Saunders, S. J. Morrison, Sox17 Dependence Distinguishes the Transcriptional Regulation of Fetal from Adult Hematopoietic Stem Cells. *Cell*. 130, 470–483 (2007).

178. M. R. Copley, S. Babovic, C. Benz, D. J. H. F. Knapp, P. A. Beer, D. G. Kent, S. Wohrer, D. Q. Treloar, C. Day, K. Rowe, H. Mader, F. Kuchenbauer, R. K. Humphries, C. J. Eaves, The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nature Cell Biology*. 15, 916–925 (2013).

179. H. Hock, M. J. Hamblen, H. M. Rooke, J. W. Schindler, S. Saleque, Y. Fujiwara, S. H. Orkin, Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature*. 431, 1002–1007 (2004).

180. H. Hock, E. Meade, S. Medeiros, J. W. Schindler, P. J. M. Valk, Y. Fujiwara, S. H. Orkin, Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* 18, 2336–2341 (2004).

181. I. Park, D. Qian, M. Kiel, M. W. Becker, M. Pihalja, I. L. Weissman, S. J. Morrison, M. F. Clarke, Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*. 423, 302–305 (2003).

182. M. Ye, H. Zhang, G. Amabile, H. Yang, P. B. Staber, P. Zhang, E. Levantini, M. Alberich-Jordà, J. Zhang, A. Kawasaki, D. G. Tenen, C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. *Nat Cell Biol*. 15, 385–394 (2013).

183. D.-M. Popescu, R. A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh, Y. Gitton, J.-E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald, J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova, F. Greig, M. D. Young, K. B. Meyer, S. Lisgo, J. Bacardit, A. Fuller, B. Millar, B. Innes, S. Lindsay, M. J. T. Stubbington, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Filby, P. Carey, A.-C. Villani, A. Roy, A. Regev, A. Chédotal, I. Roberts, B. Göttgens, S. Behjati, E. Laurenti, S. A. Teichmann, M. Haniffa, Decoding human fetal liver haematopoiesis. *Nature*. 574, 365–371 (2019).

184. L. Jardine, S. Webb, I. Goh, M. Quiroga Londoño, G. Reynolds, M. Mather, B. Olabi, E. Stephenson, R. A. Botting, D. Horsfall, J. Engelbert, D. Maunder, N. Mende, C. Murnane, E. Dann, J. McGrath, H. King, I. Kucinski, R. Queen, C. D. Carey, C. Shrubsole, E. Poyner, M. Acres, C. Jones, T. Ness, R. Coulthard, N. Elliott, S. O'Byrne, M. L. R. Haltalli, J. E. Lawrence, S. Lisgo, P. Balogh, K. B. Meyer, E. Prigmore, K. Ambridge, M. S. Jain, M. Efremova, K. Pickard, T. Creasey, J. Bacardit, D. Henderson, J. Coxhead, A. Filby, R. Hussain, D. Dixon, D. McDonald, D.-M. Popescu, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Regev, S. Behjati, E. Laurenti, N. K. Wilson, A. Roy, B. Göttgens, I. Roberts, S. A. Teichmann, M. Haniffa, Blood and immune development in human fetal bone marrow and Down syndrome. *Nature*. 598, 327–331 (2021).

185. P. Davenport, Z.-J. Liu, M. Sola-Visner, Fetal vs adult megakaryopoiesis. *Blood*. 139, 3233–3244 (2022).

186. Z.-J. Liu, J. Italiano, F. Ferrer-Marin, R. Gutti, M. Bailey, B. Poterjoy, L. Rimsza, M. Sola-Visner, Developmental differences in megakaryocytopoiesis are associated with up-regulated TPO signaling through mTOR and elevated GATA-1 levels in neonatal megakaryocytes. *Blood*. 117, 4106–4117 (2011).

187. M. C. Stolla, S. C. Catherman, P. D. Kingsley, R. G. Rowe, A. D. Koniski, K. Fegan, L. Vit, K. E. McGrath, G. Q. Daley, J. Palis, Lin28b regulates age-dependent differences in murine platelet function. *Blood Advances*. 3, 72–82 (2019).

188. X. Xu, M. Deobagkar-Lele, K. R. Bull, T. L. Crockford, A. J. Mead, A. P. Cribbs, D. Sims, C. Anzilotti, R. J. Cornall, An ontogenetic switch drives the positive and negative selection of B cells. *Proc. Natl. Acad. Sci. U.S.A.* 117, 3718–3727 (2020).

189. J. M. Bennett, D. Catovsky, M.-T. Daniel, G. Flandrin, D. A. G. Galton, H. R. Gralnick, C. Sultan, Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *Br J Haematol*. 33, 451–458 (1976).

190. J. D. Khoury, E. Solary, O. Abla, Y. Akkari, R. Alaggio, J. F. Apperley, R. Bejar, E. Berti, L. Busque, J. K. C. Chan, W. Chen, X. Chen, W.-J. Chng, J. K. Choi, I. Colmenero, S. E. Coupland, N. C. P. Cross, D. De Jong, M. T. Elghetany, E. Takahashi, J.-F. Emile, J. Ferry, L. Fogelstrand, M. Fontenay, U. Germing, S. Gujral, T. Haferlach, C. Harrison, J. C. Hodge, S. Hu, J. H. Jansen, R. Kanagal-Shamanna, H. M. Kantarjian, C. P. Kratz, X.-Q. Li, M. S. Lim, K. Loeb, S. Loghavi, A. Marcogliese, S. Meshinchi, P. Michaels, K. N. Naresh, Y. Natkunam, R. Nejati, G. Ott, E. Padron, K. P. Patel, N. Patkar, J. Picarsic, U. Platzbecker, I. Roberts, A. Schuh, W. Sewell, R. Siebert, P. Tembhare, J. Tyner, S. Verstovsek, W. Wang, B. Wood, W. Xiao, C. Yeung, A. Hochhaus, The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* (2022), doi:10.1038/s41375-022-01613-1.

191. C. Calvo, O. Fenneteau, G. Leverger, A. Petit, A. Baruchel, F. Méchinaud, Infant Acute Myeloid Leukemia: A Unique Clinical and Biological Entity. *Cancers*. 13, 777 (2021).

192. H. Bolouri, J. E. Farrar, T. Triche, R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvil, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard, S. Meshinchi, The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med*. 24, 103–112 (2018).

193. I. Roberts, D. O'Connor, A. Roy, G. Cowan, P. Vyas, The impact of trisomy 21 on foetal haematopoiesis. *Blood Cells, Molecules, and Diseases*. 51, 277–281 (2013).

194. C. E. Annesley, P. Brown, The Biology and Targeting of FLT3 in Pediatric Leukemia. *Front. Oncol.* 4 (2014), doi:10.3389/fonc.2014.00263.

195. Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, RAS oncogenes: weaving a tumorigenic web. *Nat Rev Cancer*. 11, 761–774 (2011).

196. E. Pasmant, B. Gilbert-Dussardier, A. Petit, B. de Laval, A. Luscan, A. Gruber, H. Lapillonne, C. Deswarte, P. Goussard, I. Laurendeau, B. Uzan, F. Pflumio, F. Brizard, P. Vabres, I. Naguibvena, S. Fasola, F. Millot, F. Porteu, D. Vidaud, J. Landman-Parker, P. Ballerini, SPRED1, a RAS MAPK pathway inhibitor that causes Legius syndrome, is a

tumour suppressor downregulated in paediatric acute myeloblastic leukaemia. *Oncogene*. 34, 631–638 (2015).

197. M. Yoshida, K. Tanase-Nakao, H. Shima, R. Shirai, K. Yoshida, T. Osumi, T. Deguchi, M. Mori, Y. Arakawa, M. Takagi, T. Miyamura, K. Sakaguchi, H. Toyoda, H. Ishida, N. Sakata, T. Imamura, Y. Kawahara, A. Morimoto, T. Koike, H. Yagasaki, S. Ito, D. Tomizawa, N. Kiyokawa, S. Narumi, M. Kato, Prevalence of germline *GATA2* and *SAMD9/9L* variants in paediatric haematological disorders with monosomy 7. *Br. J. Haematol.* 191, 835–843 (2020).

198. S.-J. Zhang, L.-Y. Ma, Q.-H. Huang, G. Li, B.-W. Gu, X.-D. Gao, J.-Y. Shi, Y.-Y. Wang, L. Gao, X. Cai, R.-B. Ren, J. Zhu, Z. Chen, S.-J. Chen, Gain-of-function mutation of *GATA-2* in acute myeloid transformation of chronic myeloid leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2076–2081 (2008).

199. C. D. DiNardo, F. Ravandi, S. Agresta, M. Konopleva, K. Takahashi, T. Kadia, M. Routbort, K. P. Patel, Mark Brandt, S. Pierce, G. Garcia-Manero, J. Cortes, H. Kantarjian, Characteristics, clinical outcome, and prognostic significance of IDH mutations in AML: IDH mutations in AML. *Am. J. Hematol.* 90, 732–736 (2015).

200. H. Yamazaki, M. Suzuki, A. Otsuki, R. Shimizu, E. H. Bresnick, J. D. Engel, M. Yamamoto, A Remote GATA2 Hematopoietic Enhancer Drives Leukemogenesis in inv(3)(q21;q26) by Activating EVI1 Expression. *Cancer Cell*. 25, 415–427 (2014).

201. L. Busque, M. Buscarlet, L. Mollica, R. L. Levine, Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells*. 36, 1287–1294 (2018).

202. A. A. Ferrando, C. López-Otín, Clonal evolution in leukemia. *Nat Med*. 23, 1135–1145 (2017).

203. T. Mercher, J. Schwaller, Pediatric Acute Myeloid Leukemia (AML): From Genes to Models Toward Targeted Therapeutic Intervention. *Front. Pediatr.* 7, 401 (2019).

204. D. R. Buelow, S. B. Pounds, Y. Wang, L. Shi, Y. Li, D. Finkelstein, S. Shurtleff, G. Neale, H. Inaba, R. C. Ribeiro, R. Palumbo, D. Garrison, S. J. Orwick, J. S. Blachly, K. Kroll, J. C. Byrd, T. A. Gruber, J. E. Rubnitz, S. D. Baker, Uncovering the Genomic Landscape in Newly Diagnosed and Relapsed Pediatric Cytogenetically Normal *FLT 3*- ITD AML. *Clin Transl Sci*. 12, 641–647 (2019).

205. A. Thanasopoulou, A. Tzankov, J. Schwaller, Potent co-operation between the NUP98-NSD1 fusion and the FLT3-ITD mutation in acute myeloid leukemia induction. *Haematologica*. 99, 1465–1471 (2014).

206. C. E. Annesley, C. Rabik, A. S. Duffield, R. E. Rau, D. Magoon, L. Li, V. Huff, D. Small, D. M. Loeb, P. Brown, Knock-in of the *Wt1* R394W mutation causes MDS and cooperates with *Flt3/ITD* to drive aggressive myeloid neoplasms in mice. *Oncotarget*. 9, 35313–35326 (2018).

207. A. V. Krivtsov, M. E. Figueroa, A. U. Sinha, M. C. Stubbs, Z. Feng, P. J. M. Valk, R. Delwel, K. Döhner, L. Bullinger, A. L. Kung, A. M. Melnick, S. A. Armstrong, Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia*. 27, 852–860 (2013).

208. J. N. Fisher, N. Kalleda, V. Stavropoulou, J. Schwaller, The Impact of the Cellular Origin in Acute Myeloid Leukemia: Learning From Mouse Models. *HemaSphere*. 3, e152 (2019).

209. B. J. P. Huntly, H. Shigematsu, K. Deguchi, B. H. Lee, S. Mizuno, N. Duclos, R. Rowan, S. Amaral, D. Curley, I. R. Williams, K. Akashi, D. G. Gilliland, MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *CANCER CELL*, 10 (2004).

210. T. Miyamoto, I. L. Weissman, K. Akashi, AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7521–7526 (2000).

211. P. Qin, Y. Pang, W. Hou, R. Fu, Y. Zhang, X. Wang, G. Meng, Q. Liu, X. Zhu, N. Hong, T. Cheng, W. Jin, Integrated decoding hematopoiesis and leukemogenesis using single-cell sequencing and its medical implication. *Cell Discov*. 7, 2 (2021).

212. Y. Hara, N. Shiba, G. Yamato, K. Ohki, K. Tabuchi, M. Sotomatsu, D. Tomizawa, A. Kinoshita, H. Arakawa, A. M. Saito, N. Kiyokawa, A. Tawa, K. Horibe, T. Taga, S. Adachi, T. Taki, Y. Hayashi, Patients aged less than 3 years with acute myeloid leukaemia characterize a molecularly and clinically distinct subgroup. *Br J Haematol*. 188, 528–539 (2020).

213. S. Blais, H. Boutroux, M. Pasquet, T. Leblanc, O. Fenneteau, V. Gandemer, Y. Bertrand, S. Ducassou, G. Michel, B. Nelken, A. Petit, W. Cuccuini, E. Gouache, M.-A. Renaut, A. Baruchel, H. Lapillonne, G. Leverger, Is Acute Myeloblastic Leukemia in Children Under 2 Years of Age a Specific Entity? A Report from the FRENCH ELAM02 Study Group. *HemaSphere*. 3, e316 (2019).

214. F. R. Appelbaum, H. Gundacker, D. R. Head, M. L. Slovak, C. L. Willman, J. E. Godwin, J. E. Anderson, S. H. Petersdorf, Age and acute myeloid leukemia. *Blood*. 107, 3481–3485 (2006).

215. M. F. Greaves, A. T. Maia, J. L. Wiemels, A. M. Ford, Leukemia in twins: lessons in natural history. *Blood*. 102, 2321–2333 (2003).

216. C. M. Bateman, S. M. Colman, T. Chaplin, B. D. Young, T. O. Eden, M. Bhakta, E. J. Gratias, E. R. van Wering, G. Cazzaniga, C. J. Harrison, R. Hain, P. Ancliff, A. M. Ford, L. Kearney, M. Greaves, Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood*. 115, 3553–3558 (2010).

217. P. Kosik, M. Skorvaga, M. Durdik, L. Jakl, E. Nikitina, E. Markova, K. Kozics, E. Horvathova, I. Belyaev, Low numbers of pre-leukemic fusion genes are frequently

present in umbilical cord blood without affecting DNA damage response. *Oncotarget*. 8, 35824–35834 (2017).

218. D. Hein, K. Dreisig, M. Metzler, S. Izraeli, K. Schmiegelow, A. Borkhardt, U. Fischer, The preleukemic TCF3-PBX1 gene fusion can be generated in utero and is present in ≈0.6% of healthy newborns. *Blood*. 134, 1355–1358 (2019).

219. R. F. Abu-Arja, J. L. Dargart, R. P. S. Bajwa, S. B. Kahwash, J. J. Auletta, H. G. Rangarajan, Monozygotic twins diagnosed simultaneously with RAM immunophenotype acute myeloid leukemia. *Pediatr Transplant*. 22, e13291 (2018).

220. C. L. Carmichael, D. Metcalf, K. J. Henley, E. A. Kruse, L. Di Rago, S. Mifsud, W. S. Alexander, B. T. Kile, Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15437–15442 (2012).

221. S. J. Horton, J. Jaques, C. Woolthuis, J. van Dijk, M. Mesuraca, G. Huls, G. Morrone, E. Vellenga, J. J. Schuringa, MLL–AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia*. 27, 1116–1126 (2013).

222. S. Chaudhury, C. O'Connor, A. Cañete, J. Bittencourt-Silvestre, E. Sarrou, Á. Prendergast, J. Choi, P. Johnston, C. A. Wells, B. Gibson, K. Keeshan, Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias. *Nat Commun*. 9, 5280 (2018).

223. S. N. Porter, A. S. Cluster, W. Yang, K. A. Busken, R. M. Patel, J. Ryoo, J. A. Magee, Fetal and neonatal hematopoietic progenitors are functionally and transcriptionally resistant to Flt3-ITD mutations. *eLife*. 5, e18882 (2016).

224. T. Okeyo-Owuor, Y. Li, R. M. Patel, W. Yang, E. B. Casey, A. S. Cluster, S. N. Porter, D. Bryder, J. A. Magee, The efficiency of murine MLL-ENL–driven leukemia initiation changes with age and peaks during neonatal development. *Blood Advances*. 3, 2388–2399 (2019).

225. C. Böiers, S. E. Richardson, E. Laycock, A. Zriwil, V. A. Turati, J. Brown, J. P. Wray, D. Wang, C. James, J. Herrero, E. Sitnicka, S. Karlsson, A. J. H. Smith, S. E. W. Jacobsen, T. Enver, A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1. *Developmental Cell*. 44, 362-377.e7 (2018).

226. R. G. Rowe, E. Lummertz da Rocha, P. Sousa, P. Missios, M. Morse, W. Marion, A. Yermalovich, J. Barragan, R. Mathieu, D. K. Jha, M. D. Fleming, T. E. North, G. Q. Daley, The developmental stage of the hematopoietic niche regulates lineage in MLL-rearranged leukemia. *Journal of Experimental Medicine*. 216, 527–538 (2019).

227. H. Inaba, Y. Zhou, O. Abla, S. Adachi, A. Auvrignon, H. B. Beverloo, E. de Bont, T.-T. Chang, U. Creutzig, M. Dworzak, S. Elitzur, A. Fynn, E. Forestier, H. Hasle, D.-C. Liang, V. Lee, F. Locatelli, R. Masetti, B. De Moerloose, D. Reinhardt, L. Rodriguez, N. Van

Roy, S. Shen, T. Taga, D. Tomizawa, A. E. J. Yeoh, M. Zimmermann, S. C. Raimondi, Heterogeneous cytogenetic subgroups and outcomes in childhood acute megakaryoblastic leukemia: a retrospective international study. *Blood*. 126, 1575–1584 (2015).

228. M. McNulty, J. D. Crispino, Acute Megakaryocytic Leukemia. *Cold Spring Harb Perspect Med*. 10, a034884 (2020).

229. J. Grimm, D. Heckl, J.-H. Klusmann, Molecular Mechanisms of the Genetic Predisposition to Acute Megakaryoblastic Leukemia in Infants With Down Syndrome. *Front. Oncol.* 11, 636633 (2021).

230. K. Banno, S. Omori, K. Hirata, N. Nawa, N. Nakagawa, K. Nishimura, M. Ohtaka, M. Nakanishi, T. Sakuma, T. Yamamoto, T. Toki, E. Ito, T. Yamamoto, C. Kokubu, J. Takeda, H. Taniguchi, H. Arahori, K. Wada, Y. Kitabatake, K. Ozono, Systematic Cellular Disease Models Reveal Synergistic Interaction of Trisomy 21 and GATA1 Mutations in Hematopoietic Abnormalities. *Cell Reports*. 15, 1228–1241 (2016).

231. O. Alejo-Valle, K. Weigert, R. Bhayadia, M. Ng, H. Issa, C. Beyer, S. Emmrich, K. Schuschel, C. Ihling, A. Sinz, M. Zimmermann, C. Wickenhauser, M. Flasinski, E. Regenyi, M. Labuhn, D. Reinhardt, M.-L. Yaspo, D. Heckl, J.-H. Klusmann, The megakaryocytic transcription factor ARID3A suppresses leukemia pathogenesis, 15.

232. I. S. Muskens, S. Li, T. Jackson, N. Elliot, H. M. Hansen, S. S. Myint, P. Pandey, J. M. Schraw, R. Roy, J. Anguiano, K. Goudevenou, K. D. Siegmund, P. J. Lupo, M. F. T. R. de Bruijn, K. M. Walsh, P. Vyas, X. Ma, A. Roy, I. Roberts, J. L. Wiemels, A. J. de Smith, The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis. *Nat Commun*. 12, 821 (2021).

233. K. A. L. Hasaart, F. Manders, M.-L. van der Hoorn, M. Verheul, T. Poplonski, E. Kuijk, S. M. C. de Sousa Lopes, R. van Boxtel, Mutation accumulation and developmental lineages in normal and Down syndrome human fetal haematopoiesis. *Sci Rep*. 10, 12991 (2020).

234. S. T. Chou, M. Byrska-Bishop, J. M. Tober, Y. Yao, D. VanDorn, J. B. Opalinska, J. A. Mills, J. K. Choi, N. A. Speck, P. Gadue, R. C. Hardison, R. L. Nemiroff, D. L. French, M. J. Weiss, Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*. 109, 17573–17578 (2012).

235. A. Roy, G. Cowan, A. J. Mead, S. Filippi, G. Bohn, A. Chaidos, O. Tunstall, J. K. Y. Chan, M. Choolani, P. Bennett, S. Kumar, D. Atkinson, J. Wyatt-Ashmead, M. Hu, M. P. H. Stumpf, K. Goudevenou, D. O'Connor, S. T. Chou, M. J. Weiss, A. Karadimitris, S. E. Jacobsen, P. Vyas, I. Roberts, Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17579–17584 (2012).

236. M. Byrska-Bishop, D. VanDorn, A. E. Campbell, M. Betensky, P. R. Arca, Y. Yao, P. Gadue, F. F. Costa, R. L. Nemiroff, G. A. Blobel, D. L. French, R. C. Hardison, M. J. Weiss,

S. T. Chou, Pluripotent stem cells reveal erythroid-specific activities of the GATA1 N-terminus. *J. Clin. Invest.* 125, 993–1005 (2015).

237. E. Wagenblast, J. Araújo, O. I. Gan, S. K. Cutting, A. Murison, G. Krivdova, M. Azkanaz, J. L. McLeod, S. A. Smith, B. A. Gratton, S. A. Marhon, M. Gabra, J. J. F. Medeiros, S. Manteghi, J. Chen, M. Chan-Seng-Yue, L. Garcia-Prat, L. Salmena, D. D. De Carvalho, S. Abelson, M. Abdelhaleem, K. Chong, M. Roifman, P. Shannon, J. C. Y. Wang, J. K. Hitzler, D. Chitayat, J. E. Dick, E. R. Lechman, Mapping the cellular origin and early evolution of leukemia in Down syndrome. *Science*. 373, eabf6202 (2021).

238. C. K. Lopez, S. Malinge, M. Gaudry, O. A. Bernard, T. Mercher, Pediatric Acute Megakaryoblastic Leukemia: Multitasking Fusion Proteins and Oncogenic Cooperations. *Trends in Cancer*. 3, 631–642 (2017).

239. X. Ma, M. J. Renda, L. Wang, E. Cheng, C. Niu, S. W. Morris, A. S. Chi, D. S. Krause, Rbm15 Modulates Notch-Induced Transcriptional Activation and Affects Myeloid Differentiation. *Mol Cell Biol*. 27, 3056–3064 (2007).

240. T. Mercher, G. D. Raffel, S. A. Moore, M. G. Cornejo, D. Baudry-Bluteau, N. Cagnard, J. L. Jesneck, Y. Pikman, D. Cullen, I. R. Williams, K. Akashi, H. Shigematsu, J.-P. Bourquin, M. Giovannini, W. Vainchenker, R. L. Levine, B. H. Lee, O. A. Bernard, D. G. Gilliland, The OTT-MAL fusion oncogene activates RBPJ-mediated transcription and induces acute megakaryoblastic leukemia in a knockin mouse model. *Journal of Clinical Investigation* (2009), doi:10.1172/JCI35901.

241. C. Ragu, S. Boukour, G. Elain, O. Wagner-Ballon, H. Raslova, N. Debili, E. N. Olson, D. Daegelen, W. Vainchenker, O. A. Bernard, V. Penard-Lacronique, The serum response factor (SRF)/megakaryocytic acute leukemia (MAL) network participates in megakaryocyte development. *Leukemia*. 24, 1227–1230 (2010).

242. K.-C. Ng, A.-M. Tan, Y.-Y. Chong, L.-C. Lau, J. Lou, Congenital Acute Megakaryoblastic Leukemia (M7) With Chromosomal t(l;22)(p13;q13) Translocation in a Set of Identical Twins: *Journal of Pediatric Hematology/Oncology*. 21, 428–430 (1999).

243. J. D. E. de Rooij, I. H. I. M. Hollink, S. T. C. J. M. Arentsen-Peters, J. F. van Galen, H. Berna Beverloo, A. Baruchel, J. Trka, D. Reinhardt, E. Sonneveld, M. Zimmermann, T. A. Alonzo, R. Pieters, S. Meshinchi, M. M. van den Heuvel-Eibrink, C. M. Zwaan, NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. *Leukemia*. 27, 2280–2288 (2013).

244. T. M. Franks, A. McCloskey, M. N. Shokhirev, C. Benner, A. Rathore, M. W. Hetzer, Nup98 recruits the Wdr82–Set1A/COMPASS complex to promoters to regulate H3K4 trimethylation in hematopoietic progenitor cells. *Genes Dev.* 31, 2222–2234 (2017).

245. S. Cardin, M. Bilodeau, M. Roussy, L. Aubert, T. Milan, L. Jouan, A. Rouette, L. Laramée, P. Gendron, J. Duchaine, H. Decaluwe, J.-F. Spinella, S. Mourad, F. Couture, D. Sinnett, É. Haddad, J.-R. Landry, J. Ma, R. K. Humphries, P. P. Roux, J. Hébert, T. A. Gruber, B. T. Wilhelm, S. Cellot, Human models of NUP98-KDM5A megakaryocytic

leukemia in mice contribute to uncovering new biomarkers and therapeutic vulnerabilities. *Blood Advances*. 3, 3307–3321 (2019).

246. C. Meyer, T. Burmeister, D. Gröger, G. Tsaur, L. Fechina, A. Renneville, R. Sutton, N. C. Venn, M. Emerenciano, M. S. Pombo-de-Oliveira, C. Barbieri Blunck, B. Almeida Lopes, J. Zuna, J. Trka, P. Ballerini, H. Lapillonne, M. De Braekeleer, G. Cazzaniga, L. Corral Abascal, V. H. J. van der Velden, E. Delabesse, T. S. Park, S. H. Oh, M. L. M. Silva, T. Lund-Aho, V. Juvonen, A. S. Moore, O. Heidenreich, J. Vormoor, E. Zerkalenkova, Y. Olshanskaya, C. Bueno, P. Menendez, A. Teigler-Schlegel, U. zur Stadt, J. Lentes, G. Göhring, A. Kustanovich, O. Aleinikova, B. W. Schäfer, S. Kubetzko, H. O. Madsen, B. Gruhn, X. Duarte, P. Gameiro, E. Lippert, A. Bidet, J. M. Cayuela, E. Clappier, C. N. Alonso, C. M. Zwaan, M. M. van den Heuvel-Eibrink, S. Izraeli, L. Trakhtenbrot, P. Archer, J. Hancock, A. Möricke, J. Alten, M. Schrappe, M. Stanulla, S. Strehl, A. Attarbaschi, M. Dworzak, O. A. Haas, R. Panzer-Grümayer, L. Sedék, T. Szczepański, A. Caye, L. Suarez, H. Cavé, R. Marschalek, The MLL recombinome of acute leukemias in 2017. *Leukemia*. 32, 273–284 (2018).

247. A. V. Krivtsov, T. Hoshii, S. A. Armstrong, Mixed-Lineage Leukemia Fusions and Chromatin in Leukemia. *Cold Spring Harb Perspect Med*. 7, a026658 (2017).

248. T. A. Milne, Mouse models of MLL leukemia: recapitulating the human disease. *Blood*. 129, 2217–2223 (2017).

249. J. Dang, S. Nance, J. Ma, J. Cheng, M. P. Walsh, P. Vogel, J. Easton, G. Song, M. Rusch, A. L. Gedman, C. Koss, J. R. Downing, T. A. Gruber, AMKL chimeric transcription factors are potent inducers of leukemia. *Leukemia*. 31, 2228–2234 (2017).

250. J. D. E. de Rooij, C. Branstetter, J. Ma, Y. Li, M. P. Walsh, J. Cheng, A. Obulkasim, J. Dang, J. Easton, L. J. Verboon, H. L. Mulder, M. Zimmermann, C. Koss, P. Gupta, M. Edmonson, M. Rusch, J. Y. S. Lim, K. Reinhardt, M. Pigazzi, G. Song, A. E. J. Yeoh, L.-Y. Shih, D.-C. Liang, S. Halene, D. S. Krause, J. Zhang, J. R. Downing, F. Locatelli, D. Reinhardt, M. M. van den Heuvel-Eibrink, C. M. Zwaan, M. Fornerod, T. A. Gruber, Pediatric non–Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nature Genetics*. 49, 451–456 (2017).

251. Y. Hara, N. Shiba, K. Ohki, K. Tabuchi, G. Yamato, M. Park, D. Tomizawa, A. Kinoshita, A. Shimada, H. Arakawa, A. M. Saito, N. Kiyokawa, A. Tawa, K. Horibe, T. Taga, S. Adachi, T. Taki, Y. Hayashi, Prognostic impact of specific molecular profiles in pediatric acute megakaryoblastic leukemia in non-Down syndrome: HARA et al. *Genes Chromosomes Cancer*. 56, 394–404 (2017).

252. C. Thiollier, C. K. Lopez, B. Gerby, C. Ignacimouttou, S. Poglio, Y. Duffourd, J. Guégan, P. Rivera-Munoz, O. Bluteau, V. Mabialah, M. Diop, Q. Wen, A. Petit, A.-L. Bauchet, D. Reinhardt, B. Bornhauser, D. Gautheret, Y. Lecluse, J. Landman-Parker, I. Radford, W. Vainchenker, N. Dastugue, S. de Botton, P. Dessen, J.-P. Bourquin, J. D. Crispino, P. Ballerini, O. A. Bernard, F. Pflumio, T. Mercher, Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *The Journal of Experimental Medicine*. 209, 2017–2031 (2012).

253. T. A. Gruber, A. Larson Gedman, J. Zhang, C. S. Koss, S. Marada, H. Q. Ta, S.-C. Chen, X. Su, S. K. Ogden, J. Dang, G. Wu, V. Gupta, A. K. Andersson, S. Pounds, L. Shi, J. Easton, M. I. Barbato, H. L. Mulder, J. Manne, J. Wang, M. Rusch, S. Ranade, R. Ganti, M. Parker, J. Ma, I. Radtke, L. Ding, G. Cazzaniga, A. Biondi, S. M. Kornblau, F. Ravandi, H. Kantarjian, S. D. Nimer, K. Döhner, H. Döhner, T. J. Ley, P. Ballerini, S. Shurtleff, D. Tomizawa, S. Adachi, Y. Hayashi, A. Tawa, L.-Y. Shih, D.-C. Liang, J. E. Rubnitz, C.-H. Pui, E. R. Mardis, R. K. Wilson, J. R. Downing, An Inv(16)(p13.3q24.3)-Encoded CBFA2T3-GLIS2 Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell*. 22, 683–697 (2012).

254. J. L. Smith, R. E. Ries, T. Hylkema, T. A. Alonzo, R. B. Gerbing, M. T. Santaguida, L. Eidenschink Brodersen, L. Pardo, C. L. Cummings, K. R. Loeb, Q. Le, S. Imren, A. R. Leonti, A. S. Gamis, R. Aplenc, E. A. Kolb, J. E. Farrar, T. J. Triche, C. Nguyen, D. Meerzaman, M. R. Loken, V. G. Oehler, H. Bolouri, S. Meshinchi, Comprehensive Transcriptome Profiling of Cryptic *CBFA2T3–GLIS2* Fusion–Positive AML Defines Novel Therapeutic Options: A COG and TARGET Pediatric AML Study. *Clin Cancer Res*. 26, 726–737 (2020).

255. A. M. Jetten, Emerging Roles of GLI-Similar Krüppel-like Zinc Finger Transcription Factors in Leukemia and Other Cancers. *Trends in Cancer*. 5, 547–557 (2019).

256. R. Masetti, M. Togni, A. Astolfi, M. Pigazzi, E. Manara, V. Indio, C. Rizzari, S. Rutella, G. Basso, A. Pession, F. Locatelli, *DHH-RHEBL1* fusion transcript: a novel recurrent feature in the new landscape of pediatric *CBFA2T3-GLIS2* -positive acute myeloid leukemia. *Oncotarget*. 4, 1712–1720 (2013).

257. I. Hamlett, J. Draper, J. Strouboulis, F. Iborra, C. Porcher, P. Vyas, Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. *Blood*. 112, 2738–2749 (2008).

258. M. A. Fischer, I. Moreno-Miralles, A. Hunt, B. J. Chyla, S. W. Hiebert, Myeloid translocation gene 16 is required for maintenance of haematopoietic stem cell quiescence: Mtg16 maintains haematopoietic stem cell quiescence. *The EMBO Journal*. 31, 1494–1505 (2012).

259. A. Athanasiadou, E. Stalika, V. Sidi, M. Papaioannou, M. Gaitatzi, A. Anagnostopoulos, RUNX1–MTG16 fusion gene in *de novo* acute myeloblastic leukemia with t(16;21)(q24;q22). *Leukemia & Lymphoma*. 52, 145–147 (2011).

260. P. Holmfeldt, M. Ganuza, H. Marathe, B. He, T. Hall, G. Kang, J. Moen, J. Pardieck, A. C. Saulsberry, A. Cico, L. Gaut, D. McGoldrick, D. Finkelstein, K. Tan, S. McKinney-Freeman, Functional screen identifies regulators of murine hematopoietic stem cell repopulation. *The Journal of Experimental Medicine*. 213, 433–449 (2016).

261. H. Shima, E. Takamatsu-Ichihara, M. Shino, K. Yamagata, T. Katsumoto, Y. Aikawa, S. Fujita, H. Koseki, I. Kitabayashi, Ring1A and Ring1B inhibit expression of Glis2 to maintain murine MOZ-TIF2 AML stem cells. *Blood*. 131, 1833–1845 (2018).

262. S.-Y. Lee, H. Bin Noh, H.-T. Kim, K.-I. Lee, D.-Y. Hwang, Glis family proteins are differentially implicated in the cellular reprogramming of human somatic cells. *Oncotarget*. 8 (2017), doi:10.18632/oncotarget.20334.

263. C. Thirant, C. Ignacimouttou, C. K. Lopez, M. Diop, L. Le Mouël, C. Thiollier, A. Siret, P. Dessen, Z. Aid, J. Rivière, P. Rameau, C. Lefebvre, M. Khaled, G. Leverger, P. Ballerini, A. Petit, H. Raslova, C. L. Carmichael, B. T. Kile, E. Soler, J. D. Crispino, C. Wichmann, F. Pflumio, J. Schwaller, W. Vainchenker, C. Lobry, N. Droin, O. A. Bernard, S. Malinge, T. Mercher, ETO2-GLIS2 Hijacks Transcriptional Complexes to Drive Cellular Identity and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell*. 31, 452–465 (2017).

264. C. K. Lopez, E. Noguera, V. Stavropoulou, E. Robert, Z. Aid, P. Ballerini, C. Bilhou-Nabera, H. Lapillonne, F. Boudia, C. Thirant, A. Fagnan, M.-L. Arcangeli, S. J. Kinston, M. Diop, B. Job, Y. Lecluse, E. Brunet, L. Babin, J. L. Villeval, E. Delabesse, A. H. F. M. Peters, W. Vainchenker, M. Gaudry, R. Masetti, F. Locatelli, S. Malinge, C. Nerlov, N. Droin, C. Lobry, I. Godin, O. A. Bernard, B. Göttgens, A. Petit, F. Pflumio, J. Schwaller, T. Mercher, Ontogenic Changes in Hematopoietic Hierarchy Determine Pediatric Specificity and Disease Phenotype in Fusion Oncogene–Driven Myeloid Leukemia. *Cancer Discov*. 9, 1736–1753 (2019).

265. D. C. Hernandez, P. Vyas, Oncogenic Drivers and Development, 4.

266. T. Milan, H. Canaj, C. Villeneuve, A. Ghosh, F. Barabé, S. Cellot, B. T. Wilhelm, Pediatric leukemia: Moving toward more accurate models. *Experimental Hematology*. 74, 1–12 (2019).

267. J. Zhou, R. L. Sears, X. Xing, B. Zhang, D. Li, N. B. Rockweiler, H. S. Jang, M. N. K. Choudhary, H. J. Lee, R. F. Lowdon, J. Arand, B. Tabers, C. C. Gu, T. J. Cicero, T. Wang, Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. *BMC Genomics*. 18, 724 (2017).

268. S. Lin, R. T. Luo, A. Ptasinska, J. Kerry, S. A. Assi, M. Wunderlich, T. Imamura, J. J. Kaberlein, A. Rayes, M. J. Althoff, J. Anastasi, M. M. O'Brien, A. R. Meetei, T. A. Milne, C. Bonifer, J. C. Mulloy, M. J. Thirman, Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell*. 30, 737–749 (2016).

269. P. De La Rochere, S. Guil-Luna, D. Decaudin, G. Azar, S. S. Sidhu, E. Piaggio, Humanized Mice for the Study of Immuno-Oncology. *Trends in Immunology*. 39, 748–763 (2018).

270. K. Wang, M. Sanchez-Martin, X. Wang, K. M. Knapp, R. Koche, L. Vu, M. K. Nahas, J. He, M. Hadler, E. M. Stein, M. S. Tallman, A. L. Donahue, G. M. Frampton, D. Lipson, S. Roels, P. J. Stephens, E. M. Sanford, T. Brennan, G. A. Otto, R. Yelensky, V. A. Miller, M. G. Kharas, R. L. Levine, A. Ferrando, S. A. Armstrong, A. V. Krivtsov, Patient-derived xenotransplants can recapitulate the genetic driver landscape of acute leukemias. *Leukemia*. 31, 151–158 (2017).

271. A. De Los Angeles, F. Ferrari, R. Xi, Y. Fujiwara, N. Benvenisty, H. Deng, K. Hochedlinger, R. Jaenisch, S. Lee, H. G. Leitch, M. W. Lensch, E. Lujan, D. Pei, J. Rossant, M. Wernig, P. J. Park, G. Q. Daley, Hallmarks of pluripotency. *Nature*. 525, 469–478 (2015).

272. M. J. Evans, M. H. Kaufman, Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 292, 154–156 (1981).

273. Q.-L. Ying, A. Smith, The Art of Capturing Pluripotency: Creating the Right Culture. *Stem Cell Reports*. 8, 1457–1464 (2017).

274. P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, R. D. G. McKay, New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 448, 196–199 (2007).

275. I. G. M. Brons, L. E. Smithers, M. W. B. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, L. Vallier, Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 448, 191–195 (2007).

276. D. D. Ehnes, A. M. Hussein, C. B. Ware, J. Mathieu, H. Ruohola-Baker, Combinatorial metabolism drives the naive to primed pluripotent chromatin landscape. *Experimental Cell Research*. 389, 111913 (2020).

277. L. Sun, X. Fu, G. Ma, A. P. Hutchins, Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions. *Front. Cell Dev. Biol.* 9, 637309 (2021).

278. J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*. 282, 1145–1147 (1998).

279. D. James, A. J. Levine, D. Besser, A. Hemmati-Brivanlou, TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*. 132, 1273–1282 (2005).

280. C. Dong, L. A. Fischer, T. W. Theunissen, Recent insights into the naïve state of human pluripotency and its applications. *Experimental Cell Research*. 385, 111645 (2019).

281. C. A. Cowan, J. Atienza, D. A. Melton, K. Eggan, Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science*. 309, 1369–1373 (2005).

282. J. B. Gurdon, T. R. Elsdale, M. Fischberg, Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. *Nature*. 182, 64–65 (1958).

283. K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. 126, 663–676 (2006).

284. K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*. 131, 861–872 (2007).

285. I. K. Rony, A. Baten, J. A. Bloomfield, M. E. Islam, M. M. Billah, K. D. Islam, Inducing pluripotency *in vitro* : recent advances and highlights in induced pluripotent stem cells generation and pluripotency reprogramming. *Cell Prolif.* 48, 140–156 (2015).

286. K. Hochedlinger, R. Jaenisch, Induced Pluripotency and Epigenetic Reprogramming. *Cold Spring Harb Perspect Biol*. 7, a019448 (2015).

287. W. Deng, E. C. Jacobson, A. J. Collier, K. Plath, The transcription factor code in iPSC reprogramming. *Current Opinion in Genetics & Development*. 70, 89–96 (2021).

288. J. W. Blanchard, J. Xie, N. El-Mecharrafie, S. Gross, S. Lee, R. A. Lerner, K. K. Baldwin, Replacing reprogramming factors with antibodies selected from combinatorial antibody libraries. *Nat Biotechnol*. 35, 960–968 (2017).

289. J. Guan, G. Wang, J. Wang, Z. Zhang, Y. Fu, L. Cheng, G. Meng, Y. Lyu, J. Zhu, Y. Li, Y. Wang, S. Liuyang, B. Liu, Z. Yang, H. He, X. Zhong, Q. Chen, X. Zhang, S. Sun, W. Lai, Y. Shi, L. Liu, L. Wang, C. Li, S. Lu, H. Deng, Chemical reprogramming of human somatic cells to pluripotent stem cells. *Nature*. 605, 325–331 (2022).

290. E. M. Kwon, J. P. Connelly, N. F. Hansen, F. X. Donovan, T. Winkler, B. W. Davis, H. Alkadi, S. C. Chandrasekharappa, C. E. Dunbar, J. C. Mullikin, P. Liu, iPSCs and fibroblast subclones from the same fibroblast population contain comparable levels of sequence variations. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1964–1969 (2017).

291. K. Kim, A. Doi, B. Wen, K. Ng, R. Zhao, P. Cahan, J. Kim, M. J. Aryee, H. Ji, L. I. R. Ehrlich, A. Yabuuchi, A. Takeuchi, K. C. Cunniff, H. Hongguang, S. Mckinney-Freeman, O. Naveiras, T. J. Yoon, R. A. Irizarry, N. Jung, J. Seita, J. Hanna, P. Murakami, R. Jaenisch, R. Weissleder, S. H. Orkin, I. L. Weissman, A. P. Feinberg, G. Q. Daley, Epigenetic memory in induced pluripotent stem cells. *Nature*. 467, 285–290 (2010).

292. C. K. Burrows, N. E. Banovich, B. J. Pavlovic, K. Patterson, I. Gallego Romero, J. K. Pritchard, Y. Gilad, Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. *PLoS Genet*. 12, e1005793 (2016).

293. A. Kyttälä, R. Moraghebi, C. Valensisi, J. Kettunen, C. Andrus, K. K. Pasumarthy, M. Nakanishi, K. Nishimura, M. Ohtaka, J. Weltner, B. Van Handel, O. Parkkonen, J. Sinisalo, A. Jalanko, R. D. Hawkins, N.-B. Woods, T. Otonkoski, R. Trokovic, Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports*. 6, 200–212 (2016).

294. A. Kumar, S. S. D'Souza, A. S. Thakur, Understanding the Journey of Human Hematopoietic Stem Cell Development. *Stem Cells International*. 2019, 1–13 (2019).

295. P. Pratumkaew, S. Issaragrisil, S. Luanpitpong, Induced Pluripotent Stem Cells as a Tool for Modeling Hematologic Disorders and as a Potential Source for Cell-Based Therapies. *Cells*. 10, 3250 (2021).

296. M. Ackermann, S. Liebhaber, J. Klusmann, N. Lachmann, Lost in translation: pluripotent stem cell-derived hematopoiesis. *EMBO Mol Med*. 7, 1388–1402 (2015).

297. M. Hansen, M. von Lindern, E. van den Akker, E. Varga, Human-induced pluripotent stem cell-derived blood products: state of the art and future directions. *FEBS Lett*. 593, 3288–3303 (2019).

298. A. Montel-Hagen, G. M. Crooks, From pluripotent stem cells to T cells. *Experimental Hematology*. 71, 24–31 (2019).

299. D. A. Knorr, Z. Ni, D. Hermanson, M. K. Hexum, L. Bendzick, L. J. N. Cooper, D. A. Lee, D. S. Kaufman, Clinical-Scale Derivation of Natural Killer Cells From Human Pluripotent Stem Cells for Cancer Therapy. *Stem Cells Translational Medicine*. 2, 274–283 (2013).

300. Y. Lin, M. Kobayashi, N. Azevedo Portilho, A. Mishra, H. Gao, Y. Liu, P. Wenzel, B. Davis, M. C. Yoder, M. Yoshimoto, Long-Term Engraftment of ESC-Derived B-1 Progenitor Cells Supports HSC-Independent Lymphopoiesis. *Stem Cell Reports*. 12, 572–583 (2019).

301. O. Bluteau, T. Langlois, P. Rivera-Munoz, F. Favale, P. Rameau, G. Meurice, P. Dessen, E. Solary, H. Raslova, T. Mercher, N. Debili, W. Vainchenker, Developmental changes in human megakaryopoiesis. *Journal of Thrombosis and Haemostasis*. 11, 1730–1741 (2013).

302. G. Amabile, R. S. Welner, C. Nombela-Arrieta, A. M. D'Alise, A. Di Ruscio, A. K. Ebralidze, Y. Kraytsberg, M. Ye, O. Kocher, D. S. Neuberg, K. Khrapko, L. E. Silberstein, D. G. Tenen, In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. *Blood*. 121, 1255–1264 (2013).

303. N. Suzuki, S. Yamazaki, T. Yamaguchi, M. Okabe, H. Masaki, S. Takaki, M. Otsu, H. Nakauchi, Generation of Engraftable Hematopoietic Stem Cells From Induced Pluripotent Stem Cells by Way of Teratoma Formation. *Molecular Therapy*. 21, 1424–1431 (2013).

304. C. M. Sturgeon, A. Ditadi, G. Awong, M. Kennedy, G. Keller, Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat Biotechnol*. 32, 554–561 (2014).

305. J. G. Barcia Durán, R. Lis, S. Rafii, Haematopoietic stem cell reprogramming and the hope for a universal blood product. *FEBS Lett*. 593, 3253–3265 (2019).

306. I. Elcheva, V. Brok-Volchanskaya, A. Kumar, P. Liu, J.-H. Lee, L. Tong, M. Vodyanik, S. Swanson, R. Stewart, M. Kyba, E. Yakubov, J. Cooke, J. A. Thomson, I. Slukvin, Direct

induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. *Nat Commun*. 5, 4372 (2014).

307. R. Sugimura, D. K. Jha, A. Han, C. Soria-Valles, E. L. da Rocha, Y.-F. Lu, J. A. Goettel, E. Serrao, R. G. Rowe, M. Malleshaiah, I. Wong, P. Sousa, T. N. Zhu, A. Ditadi, G. Keller, A. N. Engelman, S. B. Snapper, S. Doulatov, G. Q. Daley, Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*. 545, 432–438 (2017).

308. Y.-T. Tan, L. Ye, F. Xie, A. I. Beyer, M. O. Muench, J. Wang, Z. Chen, H. Liu, S.-J. Chen, Y. W. Kan, Respecifying human iPSC-derived blood cells into highly engraftable hematopoietic stem and progenitor cells with a single factor. *Proceedings of the National Academy of Sciences*. 115, 2180–2185 (2018).

309. M. Sakurai, H. Kunimoto, N. Watanabe, Y. Fukuchi, S. Yuasa, S. Yamazaki, T. Nishimura, K. Sadahira, K. Fukuda, H. Okano, H. Nakauchi, Y. Morita, I. Matsumura, K. Kudo, E. Ito, Y. Ebihara, K. Tsuji, Y. Harada, H. Harada, S. Okamoto, H. Nakajima, Impaired hematopoietic differentiation of RUNX1-mutated induced pluripotent stem cells derived from FPD/AML patients. *Leukemia*. 28, 2344–2354 (2014).

310. M. P. Chao, A. J. Gentles, S. Chatterjee, F. Lan, A. Reinisch, M. R. Corces, S. Xavy, J. Shen, D. Haag, S. Chanda, R. Sinha, R. M. Morganti, T. Nishimura, M. Ameen, H. Wu, M. Wernig, J. C. Wu, R. Majeti, Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. *Cell Stem Cell*. 20, 329-344.e7 (2017).

311. E. P. Papapetrou, Modeling Leukemia with Human Induced Pluripotent Stem Cells. *Cold Spring Harb Perspect Med*. 9, a034868 (2019).

312. A. G. Kotini, C.-J. Chang, I. Boussaad, J. J. Delrow, E. K. Dolezal, A. B. Nagulapally, F. Perna, G. A. Fishbein, V. M. Klimek, R. D. Hawkins, D. Huangfu, C. E. Murry, T. Graubert, S. D. Nimer, E. P. Papapetrou, Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. *Nat Biotechnol*. 33, 646–655 (2015).

313. A. G. Kotini, C.-J. Chang, A. Chow, H. Yuan, T.-C. Ho, T. Wang, S. Vora, A. Solovyov, C. Husser, M. Olszewska, J. Teruya-Feldstein, D. Perumal, V. M. Klimek, A. Spyridonidis, R. K. Rampal, L. Silverman, E. P. Reddy, E. Papaemmanuil, S. Parekh, B. D. Greenbaum, C. S. Leslie, M. G. Kharas, E. P. Papapetrou, Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. *Cell Stem Cell*. 20, 315-328.e7 (2017).

314. J. Hsu, A. Reilly, B. J. Hayes, C. A. Clough, E. Q. Konnick, B. Torok-Storb, S. Gulsuner, D. Wu, P. S. Becker, S. B. Keel, J. L. Abkowitz, S. Doulatov, Reprogramming identifies functionally distinct stages of clonal evolution in myelodysplastic syndromes. *Blood*. 134, 186–198 (2019).

315. J.-H. Lee, K. R. Salci, J. C. Reid, L. Orlando, B. Tanasijevic, Z. Shapovalova, M. Bhatia, Brief Report: Human Acute Myeloid Leukemia Reprogramming to Pluripotency Is a Rare Event and Selects for Patient Hematopoietic Cells Devoid of Leukemic

Mutations: Human AML Reprogramming Selects for Normal Genomes. *Stem Cells*. 35, 2095–2102 (2017).

316. T. Nilsson, A. Waraky, A. Östlund, S. Li, A. Staffas, J. Asp, L. Fogelstrand, J. Abrahamsson, L. Palmqvist, An induced pluripotent stem cell t(7;12)(q36;p13) acute myeloid leukemia model shows high expression of MNX1 and a block in differentiation of the erythroid and megakaryocytic lineages. *Intl Journal of Cancer*. 151, 770–782 (2022).

317. E. P. Papapetrou, Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat Med*. 22, 1392–1401 (2016).

318. T. Wang, A. R. Pine, A. G. Kotini, H. Yuan, L. Zamparo, D. T. Starczynowski, C. Leslie, E. P. Papapetrou, Sequential CRISPR gene editing in human iPSCs charts the clonal evolution of myeloid leukemia and identifies early disease targets. *Cell Stem Cell*, S1934590921000114 (2021).

319. Z. Wehbe, F. Ghanjati, C. Flotho, Induced Pluripotent Stem Cells to Model Juvenile Myelomonocytic Leukemia: New Perspectives for Preclinical Research. *Cells*. 10, 2335 (2021).

320. T. Li, Y. Zhang, Y. Li, X. Wang, W. Bao, J. Huang, Y. Ma, S. Li, S. wang, Y. Yang, Y. Liu, Y. Gao, H. Feng, Y. Li, Modeling leukemia with pediatric acute leukemia patient-derived iPSCs. *Stem Cell Research*. 54, 102404 (2021).

321. A. Sharma, S. Sances, M. J. Workman, C. N. Svendsen, Multi-lineage Human iPSC-Derived Platforms for Disease Modeling and Drug Discovery. *Cell Stem Cell*. 26, 309–329 (2020).

322. S. N. Bertuccio, F. Boudia, M. Cambot, C. K. Lopez, L. Lordier, A. Donada, E. Robert, C. Thirant, Z. Aid, S. Serravalle, A. Astolfi, V. Indio, F. Locatelli, A. Pession, W. Vainchenker, R. Masetti, H. Raslova, T. Mercher, The Pediatric Acute Leukemia Fusion Oncogene ETO2-GLIS2 Increases Self-Renewal and Alters Differentiation in a Human Induced Pluripotent Stem Cells-Derived Model. *HemaSphere*. 4, e319 (2020).

323. K. Amano, N. Takasugi, Y. Kubota, Y. Mitani, M. Sekiguchi, K. Watanabe, J. Fujimura, A. Oka, J. Takita, M. Hiwatari, *CBFA2T3-GLIS2* -positive acute megakaryoblastic leukemia in a patient with Down syndrome. *Pediatr Blood Cancer*. 67 (2020), doi:10.1002/pbc.28055.

324. G. A. MacLean, T. F. Menne, G. Guo, D. J. Sanchez, I.-H. Park, G. Q. Daley, S. H. Orkin, Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. *Proceedings of the National Academy of Sciences*. 109, 17567–17572 (2012).

325. J. Wu, Y. Xiao, J. Sun, H. Sun, H. Chen, Y. Zhu, H. Fu, C. Yu, W. E., S. Lai, L. Ma, J. Li, L. Fei, M. Jiang, J. Wang, F. Ye, R. Wang, Z. Zhou, G. Zhang, T. Zhang, Q. Ding, Z. Wang, S. Hao, L. Liu, W. Zheng, J. He, W. Huang, Y. Wang, J. Xie, T. Li, T. Cheng, X. Han, H.

Huang, G. Guo, A single-cell survey of cellular hierarchy in acute myeloid leukemia. *J Hematol Oncol*. 13, 128 (2020).

326. J. Isaac, J. Erthal, J. Gordon, O. Duverger, H.-W. Sun, A. C. Lichtler, G. S. Stein, J. B. Lian, M. I. Morasso, DLX3 regulates bone mass by targeting genes supporting osteoblast differentiation and mineral homeostasis in vivo. *Cell Death Differ*. 21, 1365–1376 (2014).

327. N. Artigas, B. Gámez, M. Cubillos-Rojas, C. Sánchez-de Diego, J. A. Valer, G. Pons, J. L. Rosa, F. Ventura, p53 inhibits SP7/Osterix activity in the transcriptional program of osteoblast differentiation. *Cell Death Differ*. 24, 2022–2031 (2017).

328. B. Q. Trinh, N. Barengo, S. B. Kim, J.-S. Lee, P. A. Zweidler-McKay, H. Naora, *Journal of Cell Science*, in press, doi:10.1242/jcs.168187.

329. Y. Tan, J. R. Testa, DLX Genes: Roles in Development and Cancer. *Cancers*. 13, 3005 (2021).

330. W. Chen, A. R. Kumar, W. A. Hudson, Q. Li, B. Wu, R. A. Staggs, E. A. Lund, T. N. Sam, J. H. Kersey, Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. *Cancer Cell*. 13, 432–440 (2008).

331. M. Gautier, C. Thirant, O. Delattre, I. Janoueix-Lerosey, Plasticity in Neuroblastoma Cell Identity Defines a Noradrenergic-to-Mesenchymal Transition (NMT). *Cancers*. 13, 2904 (2021).

332. G.-A. Franzetti, K. Laud-Duval, W. van der Ent, A. Brisac, M. Irondelle, S. Aubert, U. Dirksen, C. Bouvier, G. de Pinieux, E. Snaar-Jagalska, P. Chavrier, O. Delattre, Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene*. 36, 3505–3514 (2017).

333. R. Masetti, M. Pigazzi, M. Togni, A. Astolfi, V. Indio, E. Manara, R. Casadio, A. Pession, G. Basso, F. Locatelli, CBFA2T3-GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype. *Blood*. 121, 3469–3472 (2013).

334. Y. Zhu, T. Wang, J. Gu, K. Huang, T. Zhang, Z. Zhang, H. Liu, J. Tang, Y. Mai, Y. Zhang, Y. Li, Y. Feng, B. Kang, J. Li, Y. Shan, Q. Chen, J. Zhang, B. Long, J. Wang, M. Gao, D. Zhang, M. Zhou, X. Zhong, J. Chen, D. Pei, J. Nie, B. Liu, G. Pan, Characterization and generation of human definitive multipotent hematopoietic stem/progenitor cells. *Cell Discov*. 6, 89 (2020).

335. P. Zhang, C. Zhang, J. Li, J. Han, X. Liu, H. Yang, The physical microenvironment of hematopoietic stem cells and its emerging roles in engineering applications. *Stem Cell Res Ther*. 10, 327 (2019).

336. E. C. Kaser, The role of various interleukins in acute myeloid leukemia. *Medical Oncology*, 6 (2021).

337. M. Yoshimoto, T. Shinohara, T. Heike, M. Shiota, M. Kanatsu-Shinohara, T. Nakahata, Direct visualization of transplanted hematopoietic cellreconstitution in intact mouse organs indicates the presence of a niche. *Experimental Hematology*. 31, 733–740 (2003).

338. M. de Bruijn, E. Dzierzak, Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood*. 129, 2061–2069 (2017).

339. R. Sood, Y. Kamikubo, P. Liu, Role of RUNX1 in hematological malignancies. *Blood*. 129, 2070–2082 (2017).

340. S. Gialesaki, D. Bräuer-Hartmann, R. Bhayadia, O. Alejo-Valle, H. Issa, E. Regenyi, C. Beyer, M. Labuhn, C. Ihling, A. Sinz, M. Glaß, S. Hüttelmaier, S. Matzk, L. Schmid, F. J. Strüwe, S.-K. Kadel, D. Reinhardt, M.-L. Yaspo, D. Heckl, J.-H. Klusmann, "*RUNX1* isoform disequilibrium in the development of trisomy 21 associated myeloid leukemia" (preprint, Cancer Biology, 2022), , doi:10.1101/2022.03.07.483334.

341. M. Abuhantash, E. M. Collins, A. Thompson, Role of the HOXA cluster in HSC emergence and blood cancer. *Biochemical Society Transactions*. 49, 1817–1827 (2021).

Titre : Développement et caractérisation d'un modèle humain de leucémie pédiatrique ETO2-GLIS2 à partir de cellules souches pluripotentes induites

Mots clés : leucémie, cancers pédiatriques, cellules souches pluripotentes induites, modélisation, transcription

Résumé : Les leucémies aiguës représentent la première cause de mortalité par cancer chez les enfants. Il s'agit de cancers liquides survenant dans des cellules du sang qui ont acquis des mutations génétiques. Nos travaux portent sur les leucémies aiguës mégacaryoblastiques, diagnostiquées principalement chez les jeunes enfants et associés à un mauvais pronostic. Ces leucémies sont fréquemment causées par des anomalies chromosomiques qui induisent la fusion de deux gènes. ETO2-GLIS2 (EG) est l'une de ces fusions qui est associée à une forme de leucémie de pronostic particulièrement défavorable. Le très jeune âge des patients nous a conduit à émettre l'hypothèse que la leucémie survient durant le développement fœtal. L'étude du développement fœtal humain présente plusieurs difficultés qui ont conduit jusqu'ici à recourir à des modèles de souris qui ne récapitulent pas complètement le développement humain. Les cellules souches pluripotentes induites humaines peuvent être cultivées pour reproduire le développement vers tous les types cellulaires de l'organisme. De plus, les récentes avancées technologiques permettent d'une part, de recréer précisément les anomalies génétiques retrouvées dans les leucémies et d'autre part de caractériser leurs conséquences par des analyses moléculaires et bio-informatiques à l'échelle de chaque cellule. Mon projet de thèse consiste à développer un modèle qui permettrait d'étudier la fusion EG dans des cellules humaines. Pour cela, j'ai utilisé des cellules souches pluripotentes humaines qui peuvent être cultivées *in vitro* et donner différents types de cellules de l'organisme. J'ai modifié ces cellules pour reproduire précisément la fusion retrouvée chez les patients. Ces cellules ont ensuite été différenciées en cellules sanguines pour étudier l'impact de EG. Les cellules EG présentent ainsi plusieurs propriétés de cellules leucémiques incluant leur morphologie et leur capacité à proliférer. Ces cellules sont également capables d'induire le développement chez la souris de leucémies similaires à celles des patients. Enfin, l'analyse de données issues des techniques de séquençage à haut débit permet de caractériser la manière dont EG modifie l'expression des gènes et l'organisation de la chromatine. Ces données montrent la validité de ce modèle qui peut permettre de caractériser les mécanismes d'action de cette fusion et, à terme, de tester de nouvelles approches thérapeutiques.

Title: Development and characterization of a human ETO2-GLIS2 pediatric leukemia model from induced pluripotent stem cells

Key words: leukemia, pediatric cancers, model, pluripotent stem cells, transcription

Abstract: Acute leukemia is the leading cause of death in children. Leukemia are liquid cancers occurring in blood cells that acquire genetic mutations. Acute megakaryoblastic leukemia are diagnosed mainly in very young children and associated with a poor prognosis. These types of leukemia are frequently caused by chromosomal abnormalities that induce the fusion of two genes. ETO2-GLIS2 (EG) is one of these fusions that is associated with a form of leukemia with a particularly poor prognosis. The very young age of the patients led us to hypothesize that leukemia occurs during fetal development. Study of human fetal development is hampered by several difficulties which had led to the use of mouse models that does not fully recapitulate human development. Human pluripotent stem cells can be cultured in order to reproduce development toward all cell types of the organism. Moreover, recent technological advances allow to precisely recreate the genetic abnormalities found in leukemias as well as to characterize their consequences by molecular and bioinformatics analyzes at the single-cell level. My thesis project consists in developing a model that would allow us to study the EG fusion in human cells. For this, I used pluripotent stem cells that can be cultured *in vitro* and give rise to different cell types of the body. I have modified these cells to precisely replicate the fusion found in patients. These cells were then differentiated into blood cells to study the impact of the fusion. EG cells exhibit several characteristics of leukemic cells, including their morphology and ability to proliferate. Furthermore, these cells are also able to induce the development in mice of leukemia resembling those in patients. Finally, the analysis of high-throughput sequencing data allows to characterize the way ETO2-GLIS2 modifies the expression of other genes and the chromatin organization. These data validate the model which allow the characterization of functional mechanisms of the fusion, and ultimately, test new therapeutic approaches