



HAL
open science

Mécanismes développementaux orchestrant la différenciation des cellules souches pluripotentes induites en cellules souches hématopoïétiques

Olivier Piau

► **To cite this version:**

Olivier Piau. Mécanismes développementaux orchestrant la différenciation des cellules souches pluripotentes induites en cellules souches hématopoïétiques. Biochimie, Biologie Moléculaire. Sorbonne Université, 2023. Français. NNT : 2023SORUS082 . tel-04125236

HAL Id: tel-04125236

<https://theses.hal.science/tel-04125236v1>

Submitted on 12 Jun 2023

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

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



Sorbonne Université

Ecole Doctorale Complexité du vivant (ED515)

Laboratoire de biologie de développement (CNRS UMR7622, Inserm U1156) - Migration et différenciation des cellules souches hématopoïétiques

Centre de Recherche Saint-Antoine (UMRS938) - Développement hématopoïétique et leucémique: maladies et thérapie cellulaire

Mécanismes développementaux orchestrant la différenciation des cellules souches pluripotentes induites en cellules souches hématopoïétiques

Olivier Piau

21/04/2023

Thèse de biologie cellulaire et du développement

Directeurs de thèse : Dr. Thierry Jaffredo et Pr. Hélène Lapillonne

Co-encadrante de thèse : Laurence Guyonneau-Harmand

Membres du jury :

Pr. Bruno Fève : Président du jury

Dr. Estelle Duprez : Rapportrice

Dr. Laurent David : Rapporteur

Dr. Laure-Emmanuelle Zaragosi : Examinatrice

Pr. Georges Lacaud : Examineur

Dr. Thierry Jaffredo : Directeur de thèse

Pr. Hélène Lapillonne : Directrice de thèse

Dr. Laurence Guyonneau-Harmand : Co-encadrante de thèse

Résumé

Les cellules souches hématopoïétiques sont à l'origine de la production de l'ensemble des cellules sanguines dans notre corps. Ces cellules souches sont générées pendant une phase courte du développement au niveau de l'aorte de l'embryon. Cette production aortique est unique et les cellules souches hématopoïétiques produites au cours de cette période sont les fondatrices du système hématopoïétique définitif. Malgré quelques avancées notables, les capacités d'amplification des cellules souches hématopoïétiques restent encore insuffisantes. Quant à la production *ex vivo* de cellules souches hématopoïétiques, il n'existe pas encore de protocole utilisable pour des applications cliniques. Mon projet de thèse porte sur l'analyse d'un protocole de différenciation de cellules souches pluripotentes induites humaines en cellules souches hématopoïétiques. Il est basé sur la formation de corps embryoides et leur maturation en une seule étape de 17 jours de culture en utilisant une combinaison nouvelle et spécifique de cytokines et de facteurs de croissance sans stroma et sans transgènes. Ce protocole produit des cellules souches hématopoïétiques capables de prises de greffe sériées dans des souris immunodéprimées irradiées. Grâce à des analyses transcriptomiques en cellules uniques réalisées à différents jours du protocole de différenciation, j'ai pu caractériser les différents types cellulaires produits. A l'issue des 17 jours de différenciation, les corps embryoides produisent des cellules des lignages mésodermiques, endodermiques et ectodermiques, ainsi que des cellules présentant un profil extra-embryonnaire dont la présence a pu être confirmée par immunofluorescence. La comparaison avec des jeux de données publics d'aortes embryonnaires humaines séquencées au moment de la production des cellules souches hématopoïétiques a montré la présence de cellules endothéliales présentant un profil similaire entre les cellules aortiques et celles de nos corps embryoides. De plus, des cellules présentant une signature transcriptomique proche des cellules souches hématopoïétiques embryonnaires humaines ont pu être identifiées. Ainsi, notre protocole semble récapituler la genèse des cellules souches hématopoïétiques dans l'aorte à partir d'une transition endothélio-hématopoïétique similaire à celle qui se produit *in vivo* dans l'embryon humain. Enfin, l'analyse transcriptomique en cellule unique des moelles de souris greffées a pu montrer que les cellules humaines injectées sont capables de récapituler l'ensemble des lignages hématopoïétiques. Mon travail de thèse apporte donc une meilleure compréhension de la production *ex vivo* des cellules souches hématopoïétiques à partir de cellules pluripotentes induites, tout en fournissant des éclairages sur les mécanismes développementaux orchestrant leur apparition *in vivo*.

« Le plus important, c'est de bien connaître sa table des fyffes »

Daniel Skrzypczak

Remerciements

Je souhaiterais, en premier lieu, remercier mes encadrants de thèse : Thierry, Laurence et Hélène. Sans les retours nombreux et avisés de Thierry, une partie des résultats de cette thèse n'auraient probablement jamais vu le jour. Malgré ton emploi du temps souvent chargé, tu as toujours su trouver un moment pour me conseiller ou m'aiguiller dans mes recherches. Un immense merci à Laurence pour son soutien, ses conseils et ses chouquettes. Tu as réussi l'exploit de faire vivre ce projet tant en supportant la bande de stagiaires devenus thésards du 11^{ème} étage. Si ce projet existe et survit, c'est grâce à toi et j'espère que tu arriveras à le mener à bout. Enfin, il convient de remercier l'apport d'Hélène, notamment pour le projet RESCUE où elle nous a été d'un grand secours.

Je tiens ensuite à remercier les membres du jury d'avoir accepté d'évaluer mon travail : Bruno Fève pour avoir su libérer son emploi du temps pour présider le jury, Laurent David et Estelle Duprez d'être rapporteurs pour ma thèse, et enfin Laure-Emmanuelle Zaragosi et Georges Lacaud qui ont accepté le rôle d'examineur.

Une grande partie de ce projet n'aurait été possible sans Alain Chapel, qui a su trouver les sources de financement et mettre à notre disposition les ressources de l'IRSN pour les expériences de greffe de nos cellules souches hématopoïétiques chez la souris ou le cochon nain. La contribution de Laurence Petit, pour les analyses cytométriques, sur l'avancée de ce projet a également été indéniable.

L'aide de Pierre Charbord dans la caractérisation du mésenchyme sous-aortique a été considérable. Sans son expertise, une grande partie de nos données n'aurait pu être caractérisée.

Je remercie François Delhommeau pour son accueil chaleureux au sein du laboratoire de Saint-Antoine et pour les interactions scientifiques et culinaires qui ont toujours été passionnantes.

Cette thèse n'aurait pas été la même sans Mathias, qui a été mon duo de choc du stage de Master jusqu'à la fin de cette thèse. Tels Dupont et Dupond (j'étais Dupond avec un D, il était Dupont avec un T), nous avons su braver toutes les épreuves qui se sont présentées à nous et compléter la collection complète des blagues Carambars entre deux repas diététiques au restaurant Istanbul.

Je n'oublie pas les deux autres membres de l'agence tous risques du Master : Marilou et Emeline. Le sens de l'humour sans faille de Marilou aura été un atout indéniable à notre cohésion et le bureau de Saint-Antoine n'aurait été qu'emplein de tristesse sans les décorations d'Emeline.

Les autres membres de la 1108 de Saint-Antoine, présents et passés, ne saurait être oubliés : Irène pour sa prime d'environnement, Séverine pour sa formation à la culture cellulaire, Caroline pour sa bonne humeur au quotidien et Ladan pour ses monologues poétiques avec son ordinateur.

Tout le reste de l'équipe de Saint-Antoine mérite également d'être remercié : Nawa, Laetia, Romuald, Andràs, Daniel (et ses expériences de caléfaction), Pierre, Yaël, Mira, Thi My Anh, Hélène (Guermouche), Jean-Alain, Aude, Manon, Thomas, Clara, Julie et Nina.

Le laboratoire de Jussieu et ses thésards n'est pas sans reste. A kind thank to Olivera who was the perfect office colleague with all of the passioning talks and the Serbian political podcasts. Going to your wedding in Požarevac was an amazing experience and working with a Marvel fan was a blast. Giovanna, the conference in Edimburg was a lot of fun and the quality of the Italian restaurants you recommend is to be acknowledged. Mayssa, je garde un souvenir impérissable de ton tajin.

Un grand merci à Laurent pour les multiples discussions scientifiques, à Clémence pour son caractère pétillant qui apporte chaque jour du soleil dans le laboratoire, à Pierre-Yves pour son aide avec le cryostat et les immunofluorescences, à Charles pour l'organisation des journal clubs et le travail sur le neuro-mésenchyme et à Isabelle pour les discussions des pauses thés.

A big shout-out to Shoko for all the scRNAseq discussions and the help for some of my analysis even if they aren't included in this manuscript. Un grand merci à Jin Liu pour m'avoir convié aux réunions du groupe de bioinformatique de single-cell de Curie. A tremendous thank to Jason for the organisation of the Curie single-cell bioinformatics meetings, the discussions in Edimburgh, and the wise advises when I needed to analyze the mouse bone marrow dataset.

Il convient également de remercier Clara Münger et Timo Kholer pour les échanges fructueux lors du congrès joint des sociétés françaises et japonaises de biologie du développement. Sans leur expertise, nous n'aurions jamais envisagé la possibilité que nos corps embryoïdes puissent produire des cellules semblables à l'amnios.

Je remercie également Nicolas David et Denis Thieffry qui ont accepté de faire partie de mon comité de suivi de thèse après avoir joué un rôle majeur dans ma formation, à l'école Polytechnique pour l'un et à l'Ecole Normale Supérieure pour l'autre. Le cours de Nicolas David et d'Alice Meunier sur la biologie du développement et les cellules souches a probablement été l'un des cours favori de ma scolarité, tant de par son contenu que de par la qualité pédagogique de ses enseignants.

L'équipe du département Mathématiques du Palais de la Découverte et les doctorants que j'ai pu croiser méritent également d'être remerciés. J'ai passé des moments formidables avec Robin, Guillaume, Noëlle, Valentin, Marjolaine, Coline et Arnaud qui m'ont permis de renouer avec la

médiation scientifique à mon plus grand bonheur. Une pensée pour Jade également, avec qui nous avons bataillé dur pour être payer de notre mission par Sorbonne Université.

Enfin, une petite note s'impose pour les organisateurs et les fidèles de l'IDPA qui ont su faire vivre la barre Cassan : Simon, Judith, Agnès, Théo, Agathe, Samar, Cuong, Pénélope, Jonas, Baptiste et tant d'autres. Vous avez fait un travail formidable et avez su créer une cohésion et une dynamique incroyable au sein des étudiants de l'IBPS.

Ces remerciements ne seraient complets sans évoquer ceux qui m'ont donné l'énergie nécessaire pour mener à bien cette thèse : l'izakaya Jinchuan Shokudo et son poulet karaage, le restaurant mexicain l'Adelita et son ceviche, le restaurant chinois Heng Long et ses brioches fourrées, le restaurant Istanbul et ses kebabs et enfin les kouign-amanns de chez Cyril Lignac.

Table des matières

Résumé.....	2
Remerciements.....	4
Table des illustrations.....	10
Introduction et état de l'art.....	11
Les cellules souches hématopoïétiques : fonctionnalité et perspectives.....	12
La cellule souche hématopoïétique.....	12
Caractérisation phénotypique et fonctionnelle des cellules souches hématopoïétiques.....	14
Formation des cellules souches hématopoïétiques.....	15
Vagues d'hématopoïèse dans le sac vitellin.....	15
Vague d'hématopoïèse intra-embryonnaire.....	16
Maturation et migration dans la moelle osseuse.....	17
Utilisation thérapeutique des cellules souches hématopoïétiques.....	18
Amplification des cellules souches hématopoïétiques.....	19
Production <i>in vitro</i> de cellules souches hématopoïétiques.....	20
Modélisation du développement embryonnaire <i>in vitro</i>	24
Développement embryonnaire précoce de l'embryon humain.....	24
Modèles tridimensionnels <i>in vitro</i> du développement embryonnaire.....	26
Modélisation de l'embryon.....	26
Modélisation des annexes extra-embryonnaires et cellules souches trophoblastiques.....	27
Blastoïdes et embryons synthétiques.....	29
Analyse du séquençage des ARNs en cellule unique (scRNAseq).....	31
Extraction des échantillons.....	31
La séparation en plaques.....	31
La séparation en gouttelettes.....	31
Avantages et inconvénients.....	32
Contrôles qualités et sélection des gènes pour les analyses.....	33
Normalisation.....	35
Motivation et normalisation par taille des bibliothèques.....	35

Normalisation par groupement et déconvolution des cellules.....	36
Normalisation par transformation stabilisant la variance	37
Réduction de la dimensionnalité	39
Réduction de la taille des données	40
Visualisation des données.....	41
Clustering et labellisation des cellules.....	43
Clustering.....	43
Annotation manuelle des identités cellulaires	44
Stratégies d'annotation automatique	44
Intégration de multiples jeux de données.....	45
Trajectoires de différenciation	46
Apport du scRNAseq à la compréhension de la formation des cellules souches hématopoïétiques.....	47
Résultats	50
Différenciation de cellules souches pluripotentes induites humaines en cellules souches hématopoïétiques	50
Manuscrit soumis pour publication : <i>Generation of transgene-free hematopoietic stem cells from human induced pluripotent stem cells</i>	50
Analyses complémentaires du protocole de différenciation.....	110
Précisions sur les populations détectées en scRNAseq lors du protocole de différenciation en corps embryoides.....	110
Influence des sources de protéine sur les types cellulaires produits	116
Nouvelles approches bioinformatiques explorées lors des analyses	119
Utilisation de Gene Ontology à l'échelle de la cellule unique	119
Séparation des cellules murines et humaines dans les moelles de souris.....	120
Analyse d'un protocole d'amplification sériée des cellules souches hématopoïétiques	123
Evaluation du potentiel des cellules souches hématopoïétiques au cours de la culture	124
Analyse scRNAseq des cellules produites	125
Analyse du microenvironnement murin de l'AGM.....	130
Discussion et perspectives	163

Bibliographie..... 168

Table des illustrations

Figure 1 : Modèle classique de l'hématopoïèse	12
Figure 2 : Architectures classiques et revisitées du système hématopoïétique	14
Figure 3 : Formation des tissus humains et implantation de l'embryon entre le 7ème et le 11ème jour du développement	25
Figure 4 : Embryon humain au début de la gastrulation	26
Figure 5 : Les modèles tridimensionnels du développement humain	30
Figure 6 : Représentation schématique des étapes du prétraitement et de l'analyse d'un jeu de données scRNAseq	33
Figure 7 : Conclusions des analyses scRNAseq de l'endothélium hémogénique entre CS10 et CS15 par Zeng et al.	48
Figure 8 : Schéma des étapes principales de la formation et de la maturation des cellules souches hématopoïétiques au cours du développement de l'embryon	49
Figure 9 : Caractérisation des populations extra-embryonnaires dans les corps embryoides	112
Figure 10 : Marquage par immunofluorescence des populations trophoblastiques dans les corps embryoides	113
Figure 11 : Evolution des populations dans les corps embryoides au cours du protocole de différenciation	115
Figure 12 : Influence de la source protéique sur les populations produites dans les corps embryoides	117
Figure 13 : Validation par immunofluorescence de l'influence des sources de plasma sur les populations cellulaires produites	118
Figure 14 : Stratégie d'analyse par les Gene Ontology à l'échelle de la cellule unique	120
Figure 15 : Séparation des cellules humaines et murines dans l'analyse scRNAseq des moelles greffées	122
Figure 16 : Protocole d'amplification des cellules souches hématopoïétiques et caractérisation des cellules obtenues	124
Figure 17 : Caractérisation fonctionnelle des cellules après amplification	125
Figure 18 : Sélection des cellules souches hématopoïétiques	127
Figure 19 : Modifications transcriptomiques des cellules souches hématopoïétiques lors du protocole d'amplification	129

Introduction et état de l'art

De tout temps, l'Homme s'est intéressé aux cellules souches hématopoïétiques^a. Si, au cours de l'histoire, le sang a pu avoir une symbolique forte dans de nombreuses civilisations, la formulation et la conceptualisation de la cellule souche hématopoïétique ne se sont faites que bien plus tard, suite à deux découvertes scientifiques indépendantes réalisées en 1945 (Weissman & Shizuru, 2008). La première fut la constatation du partage de groupes sanguins entre jumeaux dans les élevages de bétail (Owen, 1945), ce qui implique une contribution embryonnaire au système sanguin adulte. La seconde fut la découverte des effets de l'irradiation des bombes atomiques de Hiroshima et Nagasaki. Les scientifiques avaient alors constaté que le système hématopoïétique était l'un des premiers organes affectés et qu'une protection de la rate (Jacobson et al., 1950), ou une injection de cellules de la moëlle osseuse (Lorenz et al., 1951), avaient un effet protecteur. Ces deux résultats amenèrent ensuite à la découverte des cellules souches hématopoïétiques dans les années 60 (Till & McCulloch, 1961).

Depuis, les cellules souches hématopoïétiques sont un modèle d'étude tant sur le plan fondamental que thérapeutique. De nombreux patients souffrant de désordres de l'hématopoïèse, la compréhension des mécanismes régulant la production et la différenciation des cellules souches hématopoïétiques est essentielle à l'établissement de stratégies thérapeutiques. Les travaux décrits dans cette thèse s'inscrivent dans ce processus de compréhension et d'amélioration de protocoles à visées thérapeutiques et de leurs apports sur le plan fondamental. La majorité de ces analyses ont été réalisées en se basant sur la technologie du séquençage des ARNs en cellule unique (scRNAseq).

Dans une première partie, cet état de l'art présentera les cellules souches hématopoïétiques et les protocoles de production et d'amplification dont l'étude constitue le corps de cette thèse. Dans une seconde partie, quelques éléments de biologie du développement seront rappelés afin d'introduire les différents modèles existant pour reproduire le développement embryonnaire *in vitro*. Enfin, dans une dernière partie, l'analyse des données produites par les technologies de séquençage des ARNs en cellule unique sera détaillée et sera complétée par une description des avancées que cette technique a permis dans la compréhension des cellules souches hématopoïétiques.

^a En effet, la présence de peinture rouge dans la grotte de Niaux suggère l'importance accordée au sang par ses habitants (Danic & Lefrère, 2010) (et donc indirectement aux cellules souches hématopoïétiques, même si le concept n'était pas encore formulé à l'époque).

Les cellules souches hématopoïétiques : fonctionnalité et perspectives

La cellule souche hématopoïétique

L'hématopoïèse est le processus physiologique couvrant la production et le renouvellement de l'ensemble des cellules sanguines. Le renouvellement permanent de ces cellules qui représenterait environ 90% de la production de nouvelles cellules du corps humain n'est possible que grâce à une petite sous-population des cellules : les cellules souches hématopoïétiques (Cosgrove et al., 2021).

Les cellules souches hématopoïétiques se caractérisent par deux propriétés fondamentales : l'auto-renouvellement et la capacité de différenciation en l'ensemble des types cellulaires sanguins. Le terme est apparu à la fin du XIX^{ème} siècle pour désigner les cellules de la moëlle osseuse ayant un double potentiel de différenciation myéloïde et érythroïde (Pappenheim, 1896). La capacité d'auto-renouvellement de ces cellules, quant à elle, ne sera démontrée qu'avec les travaux de Becker, et de Till et McCulloch dans les années 60 (Becker et al., 1963; Till & McCulloch, 1961).

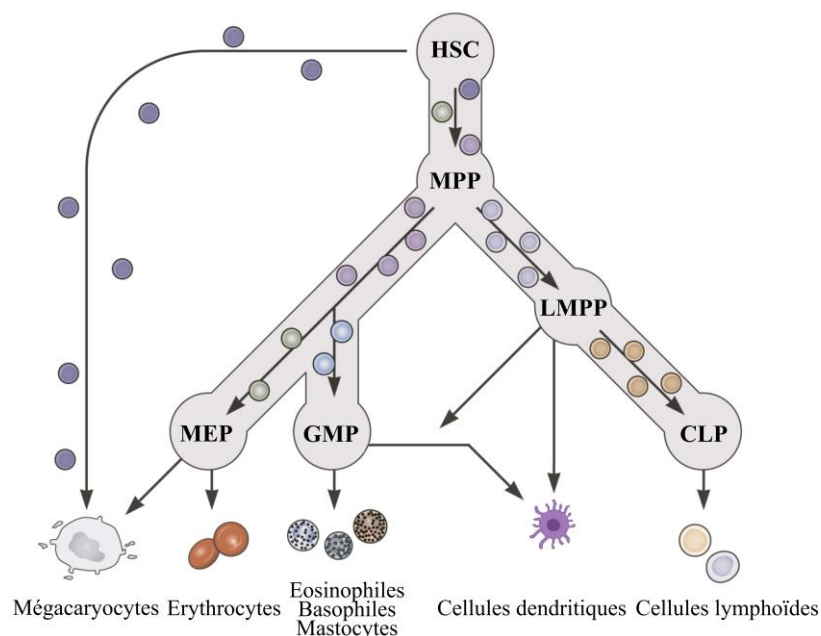


Figure 1 : *Modèle classique de l'hématopoïèse*

HSC : cellule souche hématopoïétique, MPP : cellule progénitrice multipotente, LMPP : cellule progénitrice myélo-lymphoïde, MEP : cellule progénitrice érythro-mégacaryocytaire, GMP : cellule progénitrice granuleuse, CLP : cellule progénitrice lymphoïde

Adapté de Psaila et al. (Psaila & Mead, 2019) et de Laurenti et al. (Laurenti & Göttgens, 2018)

Les cellules souches hématopoïétiques assurent le renouvellement des cellules sanguines en se différenciant en cellules progénitrices multipotentes douées d'une forte capacité d'amplification. Ces progéniteurs multipotents se différencient ensuite en progéniteurs plus spécialisés qui vont donner naissance aux cellules plus matures. On distingue généralement les progéniteurs lymphoïdes, responsables de la production des lymphocytes B et T, des progéniteurs myéloïdes qui sont à l'origine des cellules érythrocytaires, des plaquettes, des granulocytes et des monocytes (Figure 1). Ces progéniteurs multipotents ne possédant pas la propriété d'auto-renouvellement propre aux cellules souches hématopoïétiques, ils finissent par s'épuiser.

Cette vision hiérarchique de l'hématopoïèse, héritée des résultats des premiers tests clonogéniques et des premières expériences d'irradiation a été affinée avec l'apparition de nouvelles techniques comme les expériences de suivi du lignage des cellules (Busch et al., 2015) et la transcriptomique à l'échelle de la cellule unique (Laurenti & Göttgens, 2018). Ces dernières ont pu en effet montrer que les cellules se différenciaient de manière continue, et non pas par palier (Velten et al., 2017). De plus, dans des conditions physiologiques, les cellules progénitrices sont les principales contributrices de la production hématopoïétique quotidienne (Busch et al., 2015), les cellules souches hématopoïétiques ne se différenciant pour former ces progéniteurs (Sawai et al., 2016) qu'une fois toutes les 40 semaines en moyenne (Catlin et al., 2011). Ainsi, bien que seules les cellules souches hématopoïétiques soient nécessaires au maintien de l'hématopoïèse sur le long terme ou à la réponse aux stress, elles sont la majorité du temps en quiescence dans l'organisme adulte (Wilson et al., 2008).

La division des progéniteurs multipotents en progéniteurs communs myéloïdes et lymphoïdes, bien qu'ayant été le consensus pendant de nombreuses années, a été remise en question par des expériences de suivi des cellules à l'échelle unique qui ont pu montrer que les progéniteurs myéloïdes étaient en réalité déjà biaisés vers les voies érythrocytaires ou monocytaires (Perié et al., 2015) et qu'une majorité des cellules souches hématopoïétiques ne se différenciaient qu'en cellules mégacaryocytaires (Rodriguez-Fraticelli et al., 2018). De plus, il a été montré que les schémas de différenciation chez l'Homme évoluent au cours du développement, ce qui aboutit à l'architecture de l'hématopoïèse présentée en Figure 2.

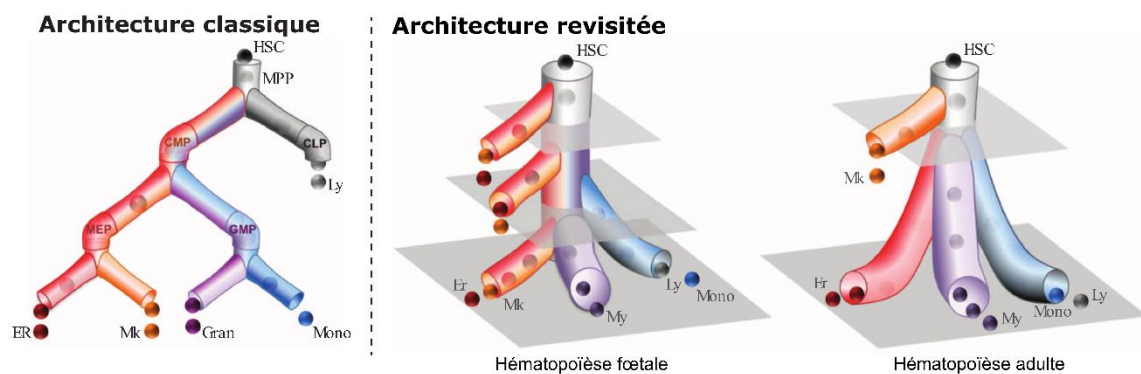


Figure 2 : Architectures classiques et revisitées du système hématopoïétique

HSC : cellules souches hématopoïétiques, MPP : cellules progénitrices multipotentes, CMP : cellules progénitrices myéloïdes, CLP : cellules progénitrices lymphoïdes, MEP : cellules progénitrices érythro-mégacaryocytaires, GMP : cellules progénitrices monocytaires-granuleuses, ER : cellules érythroïdes, Mk : mégacaryocytes, Gran : granulocytes, Mono : monocytes, Ly : lymphocytes.

Adapté de Notta et al. (Notta et al., 2016)

Caractérisation phénotypique et fonctionnelle des cellules souches hématopoïétiques

Malgré de nombreuses recherches sur les cellules souches hématopoïétiques, il s'avère extrêmement difficile de différencier ces dernières des progéniteurs multipotents les plus précoces. Plusieurs stratégies d'enrichissement, basées sur l'expression de marqueurs de surface suivi de tri cellulaire, permettent néanmoins de sélectionner une fraction plus ou moins pure contenant les cellules souches hématopoïétiques. Une stratégie classique, chez l'Homme, consiste à sélectionner les cellules CD34⁺ (Civin et al., 1984), CD38⁻ (Huang & Terstappen, 1994), CD90⁺ (Baum et al., 1992), CD45RA⁻ (Lansdorp et al., 1990), CD49f⁺ (Notta et al., 2011), Lin⁻ (Majeti et al., 2007). Une étude récente (Anjos-Afonso et al., 2022) montre que, pour le sang de cordon, la sélection des cellules arborant le phénotype CD34⁺ EPCR⁺ serait potentiellement plus pertinente car la population sélectionnée serait plus homogène. Malheureusement, ces stratégies ne sont pas parfaites et les cellules sélectionnées contiennent toujours une proportion plus ou moins importante de cellules progénitrices. Ainsi, la seule stratégie de caractérisation des cellules souches hématopoïétiques existant à l'heure actuelle est basée sur des tests de greffes chez des souris irradiées similaires aux expériences menées par Till et McCulloh (Dick et al., 1991). Les cellules candidates sont greffées dans des souris irradiées. Après quelques semaines, on vérifie la présence de cellules hématopoïétiques humaines dans la moelle des souris greffées pour attester du caractère souche des cellules injectées. Les progéniteurs multipotents sont considérés comme épuisés après 8 semaines (Morrison et al., 1997; Wagers & Weissman, 2006), mais il est recommandé d'attendre au moins 12 semaines pour vérifier une reconstitution hématopoïétique

à court terme (Cheng et al., 2013) et 16 semaines minimum pour une capacité de reconstitution hématopoïétique à long terme (Arcangeli et al., 2020; Purton & Scadden, 2007). Les premières greffes de cellules souches hématopoïétiques humaines furent réalisées dans les années 80 sur des souris immunodéficientes SCID (*severe combined immunodeficiency*) (McCune et al., 1988). Afin de pallier au faible taux de greffe et à la nécessité d'injecter des cytokines humaines à la souris tout au long de la période de greffe, le modèle NOD/SCID (*non-obese diabetic/severe combined immunodeficiency*) fut mis au point à la fin des années 90 (Shultz et al., 1995). Des progrès furent encore faits dans les années 2000 avec l'apparition des souris NSG (Shultz et al., 2005) (NOD/SCID gamma mouse) qui représentent la souche la plus utilisée à l'heure actuelle (Radtke et al., 2020). Le test de référence afin de vérifier la capacité de repopulation des cellules comporte ensuite une greffe secondaire à partir des cellules issues des moelles des souris primaires (Arcangeli et al., 2020; Broxmeyer et al., 2005; Wilkinson et al., 2020). Cette greffe secondaire permet d'attester de la capacité d'auto-renouvellement des cellules présentes dans les moelles des récipients primaires. Il a été montré avec les cellules souches hématopoïétiques murines qu'il est possible de réaliser 4 à 5 transplantations sériées avant l'épuisement du caractère souche des cellules (Harrison, 1975) ; cela est probablement dû au stress cellulaire engendré par la procédure de transplantation (Harrison et al., 1978).

Formation des cellules souches hématopoïétiques

La formation de cellules hématopoïétiques dans l'embryon des vertébrés se fait en plusieurs vagues. Deux vagues de production de progéniteurs multipotents ont lieu dans le sac vitellin ; puis une troisième vague, productrice de cellules souches hématopoïétiques, a lieu dans l'aorte.

Vagues d'hématopoïèse dans le sac vitellin

Les différentes vagues d'hématopoïèse dans le sac vitellin ont été finement caractérisées dans le système murin, notamment par les travaux de James Palis (Palis et al., 2010). Dans l'embryon de souris, les premières traces d'hématopoïèse sont détectables au 7^{ème} jour du développement de l'embryon (E7.25), peu après la gastrulation. Cette première vague, de E7.25 à E8.5, est dite primitive car les cellules érythrocytaires produites présentent une taille plus importante que les érythrocytes adultes et expriment la globine β H1, spécifique du développement embryonnaire (Kingsley et al., 2006; Palis, 2014). Des cellules macrophagiques ou mégacaryocytaires sont également produites lors de cette première vague (Moore & Metcalf, 1970; Tober et al., 2007). Ces cellules sont produites par des progéniteurs unipotents qui se différencient *in-situ* avant de quitter le sac vitellin (Palis, 2008). Une seconde vague hématopoïétique, présente également dans

le sac vitellin, a lieu quelques jours plus tard (E8.5) et produit dans un premier temps des progéniteurs érythro-myéloïdes (M. J. Chen et al., 2011; McGrath et al., 2015), puis des progéniteurs myélo-lymphoïdes à partir de E9 (Yoshimoto et al., 2011, 2012). Cette vague s'étend ensuite à l'allantoïde et au reste de l'embryon. Cette seconde vague est une vague d'hématopoïèse dite définitive puisque les érythrocytes produits expriment des globines adultes telles que la globine β (Palis et al., 1999).

Chez l'Homme, les premières traces d'hématopoïèse détectables dans l'embryon sont visibles à partir de 16 jours, au stade de Carnegie 7 (CS7), dans le sac vitellin (Bloom & Bartelmez, 1940). Lors de cette première vague, seuls des érythrocytes immatures, des macrophages et des mégacaryocytes sont produits (Fukuda, 1973; Luckett, 1978). Ces érythrocytes primitifs sont visibles dans la cavité cardiaque à partir du 21^{ème} jour (CS10), ce qui marque le début de la circulation sanguine dans l'embryon (Tavian et al., 1999). La présence d'une seconde vague, similaire à celle ayant lieu chez la souris à partir de E8.5, est encore sujette à débat (Ivanovs et al., 2017; Julien et al., 2016). Bien que la présence de progéniteurs érythro-myéloïdes soit détectable dans le sac vitellin à partir du CS13 (Migliaccio et al., 1986), il n'a pas pu être déterminé si ces derniers sont originaires du sac vitellin ou de la vague intra-embryonnaire ayant lieu au même moment. Contrairement à la souris, l'hématopoïèse extra-embryonnaire ne produit pas de cellules lymphoïdes (Tavian et al., 2001).

Chez la souris, ces deux vagues d'hématopoïèse extra-embryonnaire sont suffisantes pour permettre à l'embryon de se développer jusqu'à la naissance (Wolber et al., 2002). Cependant, elles n'ont qu'un rôle transitoire dans l'hématopoïèse, aussi bien chez la souris que chez l'Homme, car elles ne produisent pas de cellules souches hématopoïétiques (Ivanovs et al., 2011).

Vague d'hématopoïèse intra-embryonnaire

La troisième vague hématopoïétique a lieu dans l'embryon, dans une région appelée Aorte-Gonade-Mésonephros (AGM), plus précisément dans la partie ventrale de l'aorte dorsale chez l'embryon humain (Tavian et al., 1996; Yvernogeu et al., 2023). Si l'hypothèse de l'émergence de cellules hématopoïétiques dans l'aorte peut être tracée jusqu'aux débuts du XX^{ème} siècle (Jordan, 1917), la démonstration de l'origine aortique des cellules souches hématopoïétique ne s'est faite que bien plus tard, en 1975, chez l'oiseau, à l'aide de systèmes de greffe caille-poulet (Dieterlen-Lievre, 1975). Ces résultats ont ensuite pu être également démontrés chez l'amphibien (Ciau-Uitz et al., 2000; Turpen et al., 1981), la souris (Cumano et al., 1996; Medvinsky & Dzierzak, 1996) puis chez l'Homme (Tavian et al., 2001).

Lors de cette troisième vague, une partie des cellules endothéliales du plancher de l'aorte dorsale se transdifférencient en cellules hématopoïétiques selon un processus appelé transition

endothélio-hématopoïétique (Jaffredo et al., 1998; Kissa & Herbomel, 2010; Zovein et al., 2008) (parfois abrégée en EHT pour *endothelial-to-hematopoietic transition*). Elles forment alors des clusters de cellules hématopoïétiques le long de la paroi dans la lumière de l'aorte qui vont ensuite donner naissance aux cellules souches hématopoïétiques. Ce processus s'opère lors d'une fenêtre assez réduite du développement, entre les stades CS 13 et 17 chez l'embryon humain (soit entre 27 et 42 jours après fécondation) (Tavian et al., 1996, 1999).

Cette transition endothélio-hématopoïétique est notamment régulée par le facteur de transcription RUNX1 (T. North et al., 1999) qui est responsable du changement d'identité des cellules endothéliales. Il a en effet été montré que ce dernier est nécessaire pour l'apparition des clusters intra-aortiques mais pas pour la maturation des cellules souches hématopoïétiques (M. J. Chen et al., 2009), bien que son expression continue à être détectée. Au moment de leur naissance dans l'aorte, les cellules souches hématopoïétiques humaines arborent un phénotype particulier (Ivanovs et al., 2014) (CD34⁺, CD144⁺, CD45⁺, CD117⁺, CD90⁺, CD105⁺, RUNX1⁺, CD38⁻) avec la persistance de certains marqueurs endothéliaux dont l'expression s'atténue progressivement au cours du développement.

Il semblerait que la transition endothélio-hématopoïétique ne soit cependant pas uniquement restreinte à l'aorte puisque cette dernière a également pu être détectée dans d'autres organes tels que l'endocarde (Nakano et al., 2013), la tête (Li et al., 2012) ou encore les reins (Hwang et al., 2021).

Maturation et migration dans la moelle osseuse

Après leur naissance dans l'aorte, vers la sixième semaine du développement de l'embryon, les cellules souches hématopoïétiques migrent dans le foie fœtal (Tavian et al., 1999) où elles vont maturer et être amplifiées. Durant cette période du développement embryonnaire, la contribution des cellules souches hématopoïétiques à l'hématopoïèse de l'embryon est encore négligeable et est majoritairement assurée par des progéniteurs multipotents (Rybtsov et al., 2016) et qui sont eux aussi amplifiés dans le foie fœtal (Ganuza et al., 2022; Yokomizo et al., 2022). Les cellules souches migrent ensuite dans la moelle osseuse (Charbord et al., 1996) où elles vont progressivement adopter leur fonction de pilier de l'hématopoïèse.

Récemment, une quatrième vague a été identifiée dans la moelle osseuse en formation de l'oiseau et la souris. Les cellules endothéliales de la moelle osseuse sont en effet capables de réaliser une transition endothélio-hématopoïétique similaire à celle observée chez l'embryon et produisent des cellules souches hématopoïétiques capables de prise de greffe (Yvernogeu et al., 2019).

Utilisation thérapeutique des cellules souches hématopoïétiques

Les applications thérapeutiques de cellules souches hématopoïétiques peuvent se diviser en deux catégories :

- Les thérapies géniques qui consistent à greffer à un patient des cellules génétiquement modifiées (préférentiellement autologues) afin de corriger des gènes défectueux.
- Les thérapies cellulaires qui consistent à greffer à un patient des cellules souches hématopoïétiques autologues ou hétérologues pour régénérer une hématopoïèse défailante.

Dans le cas de la thérapie génique, la méthode la plus courante consiste à extraire les cellules du patient afin de réaliser une modification *ex vivo* des cellules souches hématopoïétiques. Une fois les cellules modifiées, elles sont réinjectées au patient. Ceci est possible car les cellules souches hématopoïétiques sont facilement accessibles, soit par prélèvement dans la moelle osseuse, soit par mobilisation à partir du sang périphérique avec des injections du facteur de croissance hématopoïétique spécifique de la lignée granulocytaire (G-CSF pour *Granulocyte Colony-Stimulating factor*) (Panch et al., 2017). Le recours à une modification *ex vivo* par rapport à une approche *in vivo* présente l'avantage d'éviter de modifier d'autres types cellulaires que celui d'intérêt. Le premier essai clinique de thérapie génique chez l'Homme fut effectué en 1990 en modifiant des lymphocytes T d'un enfant touché par des immunodéficiences combinées sévères par déficit en adénosine désaminase (Blaese et al., 1995). En 2000, l'équipe d'Alain Fischer réalisa une thérapie génique en modifiant les cellules souches hématopoïétiques de patients atteints de déficit immunitaire combiné sévère lié à l'X (Cavazzana-Calvo et al., 2000). Malgré ces avancées, un défi majeur pour l'utilisation de cette technique est de maintenir les cellules souches hématopoïétiques en culture et de les amplifier durant l'étape de modification génétique tout en ne perdant pas leur caractère souche.

La thérapie cellulaire est inspirée des travaux de 1951 (Lorenz et al., 1951) montrant l'utilité de la greffe de cellules de la moelle osseuse chez des souris et des cochons d'Inde afin de réduire le taux de mortalité après irradiation. Les premiers essais de greffe de cellules de la moelle osseuse chez l'Homme ont eu lieu en 1957 (Thomas et al., 1957) mais n'ont malheureusement abouti qu'à une greffe à court terme. Le premier succès thérapeutique a eu lieu avec la première greffe hétérologue suite à l'accident de l'institut des sciences nucléaires de Vinča (Mathé, 1960) en 1958 où une prise de greffe temporaire a permis la survie des patients le temps que leur propre moelle se reconstitue. La première greffe à long-terme fut une greffe géno-identique publiée en 1961 avec la transplantation d'un patient atteint d'aplasie médullaire avec les cellules de son frère jumeau

(Robins & Noyes, 1961). Quelques années plus tard, en 1965, la première greffe hétérologue à long terme permit de soigner des patients atteints de leucémie (Mathé et al., 1965). La méconnaissance des systèmes d'histocompatibilité a cependant posé problème jusqu'aux années 90 (P. Martin & Aulagner, 2009), puisque beaucoup de patients décèdent de maladies du greffon contre l'hôte (ou GVH pour *Graft-versus-Host disease*) ou sont atteints de GVH chronique comme le patient de 1965. La découverte du rôle crucial des antigènes HLA (Morishima, 2002) dans la tolérance de la greffe a permis de grandement améliorer le taux de succès des greffes de moelle osseuse. Désormais, la greffe de moelle osseuse est utilisée dans le cas de leucémies (Mathé et al., 1965) ou d'autres maladies rares telles l'aplasie médullaire (Robins & Noyes, 1961), les myélodysplasies (Witte et al., 1990) ou encore l'anémie de Fanconi (Gluckman et al., 1980). La plupart des greffes sont réalisées de manière hétérologue en utilisant un donneur compatible. Lorsqu'aucun donneur n'est disponible dans la famille du patient, le registre international des donneurs est consulté. Cependant, plusieurs milliers de patients n'ont pas de donneurs compatibles chaque année (World Donor Marrow Association, 2021). Une solution proposée à cette pénurie est la production *in vitro* de cellules souches hématopoïétiques à partir de cellules souches (Blau & Daley, 2019). La production de cellules souches hématopoïétiques *in vitro* est également vue comme une opportunité de modéliser les différents types de leucémies afin de trouver de nouvelles cibles thérapeutiques (Deslauriers et al., 2021).

Amplification des cellules souches hématopoïétiques

Les premières approches visant à amplifier en culture sans stroma les cellules souches hématopoïétiques ont cherché à utiliser des combinaisons de cytokines présentes dans la moelle osseuse. En 1997, deux protocoles permettant de quadrupler le nombre de cellules souches hématopoïétiques humaines furent mis au point (Bhatia et al., 1997; Conneally et al., 1997). Ces derniers se basent sur l'emploi de Stem Cell Factor (SCF), d'interleukine 3 (IL-3), d'interleukine 6 (IL-6) et de G-CSF. Quelques années plus tard, deux protocoles (Kobari et al., 2000; Ueda et al., 2000) ont proposé l'utilisation de la thrombopoïétine pour obtenir une amplification des cellules initiatrices de culture à long terme (LTC-IC) pouvant atteindre un facteur 47 (Kobari et al., 2000). Plus récemment (Himburg et al., 2010), il a été montré que la pleiotrophine permet d'amplifier les cellules souches hématopoïétiques *in vitro* et *in vivo*. Afin d'améliorer ces rendements et d'augmenter le temps de culture pendant lesquelles il est possible de conserver des cellules souches hématopoïétiques (de 4 à 14 jours pour les protocoles décrits précédemment), la piste des petites molécules a été explorée. La première petite molécule ayant fait l'objet d'un essai clinique est le tétraéthylène-pentamine (de Lima et al., 2008), un chélateur du cuivre favorisant la

prolifération des populations repeuplant la moelle osseuse (T. Peled et al., 2004). Une seconde molécule fut identifiée en s'inspirant de travaux de 1974 (Fehér & Gidàli, 1974) montrant l'impact de la prostaglandine E2 (PGE2) dans la prolifération des cellules souches hématopoïétiques. Un dérivé plus stable de la PGE2, le 16,16-diméthyl-PGE2 (dmPGE2) permet de doubler le nombre de cellules souches hématopoïétiques prenant la greffe à long terme (T. E. North et al., 2007) et un essai clinique a pu démontrer qu'il permettait d'accélérer la prise de greffe à partir de sangs de cordons (Cutler et al., 2013). Suite à un criblage de 100 000 molécules, la StemReginin 1 (Boitano et al., 2010) (SR1) fut identifiée comme permettant d'obtenir une amplification d'un facteur 17 des cellules souches hématopoïétiques prenant la greffe à long-terme. Un essai clinique (Wagner et al., 2016) a montré que l'utilisation de SR-1 pour l'amplification de sangs de cordons permettait une meilleure prise de greffe chez les patients. Une quatrième molécule, identifiée en 2012, la nicotinamide (T. Peled et al., 2012) a également montré de bonnes capacités d'amplification des cellules souches hématopoïétiques et une prise de greffe améliorée chez les patients (Horwitz et al., 2014). Enfin, en 2014, une nouvelle molécule prometteuse fut identifiée : l'UM171 (Fares et al., 2014). Cette dernière est capable d'augmenter d'un facteur 100 les cellules souches permettant une prise à long terme en 21 jours et a également fait l'objet d'un essai clinique de phase I/II (Cohen et al., 2020) où elle a montré de meilleurs résultats que la nicotinamide et la SR1.

Bien qu'initialement développées pour l'amplification de sangs de cordons, la PGE2 et l'UM171 sont désormais étudiées dans le cadre de la thérapie génique (Zonari et al., 2017) afin de pallier à la perte progressive des capacités de prise de greffe des cellules souches hématopoïétiques suite à la mise en culture nécessaire aux modifications géniques (Glimm et al., 2000; Kallinikou et al., 2012; Larochelle et al., 2012).

Production *in vitro* de cellules souches hématopoïétiques

La source privilégiée pour la production *in vitro* de cellules souches hématopoïétiques est la cellule souche pluripotente. Ces dernières sont caractérisées par deux propriétés : la capacité d'auto-renouvellement et la capacité de différenciation dans l'ensemble des types cellulaires produits par l'embryon (pluripotence). La capacité d'auto-renouvellement permet d'amplifier ces cellules en culture aisément avant de les différencier vers le type cellulaire d'intérêt ; dans notre cas, les cellules souches hématopoïétiques.

Le terme cellule souche vient de l'allemand *Stammzelle* et servait initialement à décrire le premier organisme unicellulaire à l'origine de la vie (Haeckel, 1879). Suite à des expériences d'agrégation d'embryons et de blastocystes chimériques, il fut montré que les cellules de la masse cellulaire

interne du blastocyste sont pluripotentes (Brinster, 1974; Dewey et al., 1977; Gardner, 1968). Des protocoles de culture permettant d'assurer leur auto-renouvellement furent progressivement développés et aboutirent à l'obtention des premières cellules souches embryonnaires murines en 1981 (Evans & Kaufman, 1981; G. R. Martin, 1981) et humaines en 1998 (Thomson et al., 1998). En 2006, une nouvelle source de cellules souches pluripotentes fut mise au point par Shinya Yamanaka en reprogrammant des fibroblastes murins en cellules souches pluripotentes par induction de 4 transgènes : Oct4, Sox2, Klf1 et Myc (Takahashi & Yamanaka, 2006). Ce protocole fut ensuite adapté pour la reprogrammation de cellules humaines un an plus tard (Takahashi et al., 2007). La même année (J. Yu et al., 2007), un protocole alternatif, utilisant les gènes OCT4, SOX2, NANOG et LIN28 fut publié par l'équipe de James Thomson. Ces cellules issues de la reprogrammation de cellules adultes sont dénommées cellules souches pluripotentes induites, par opposition aux cellules souches embryonnaires dérivées de la masse cellulaire interne d'embryons.

Le premier protocole de différenciation des cellules souches pluripotentes humaines en cellules hématopoïétiques a été publié en 2001 (Kaufman et al., 2001) avec la production de colonies hématopoïétiques capables de former des cellules érythroïdes, myéloïdes et mégacaryocytaires à partir de cellules souches pluripotentes embryonnaires. D'autres protocoles ont ensuite été développés pour les différents lignages hématopoïétiques en utilisant des cellules souches embryonnaires, ou des cellules souches pluripotentes induites pour les lignées érythroïdes (Lapillonne et al., 2010; Lu et al., 2010), lymphoïdes (Carpenter et al., 2011; Lei et al., 2009) ou mégacaryocytaires (Feng et al., 2014; Takayama et al., 2008). La production de cellules présentant un phénotype proche des cellules souches hématopoïétiques (Vodyanik et al., 2005; Weisel et al., 2006) ou des progéniteurs du sac vitellin (Zambidis et al., 2005) a été atteint assez tôt en utilisant des cocultures avec des cellules stromales. Néanmoins, les cellules produites par ces protocoles ne sont pas capables de coloniser les moelles de souris (Vodyanik et al., 2005). L'expression de certains gènes clés du développement n'étant pas respectée, certains protocoles ont essayé d'induire l'expression des programmes nécessaires par ajout de cytokines et de facteurs de croissance. On peut par exemple citer BMP4 (Chadwick et al., 2003; Kennedy et al., 2007; Lengerke et al., 2009), FGF2 (Pick et al., 2007), APLN (Q. C. Yu et al., 2012), l'activine A (Cerdan et al., 2012) et VEGFA (Kennedy et al., 2007). Malheureusement, aucune de ces combinaisons ne s'est avérée être suffisante pour obtenir une greffe à long terme.

Une piste intéressante a été proposée avec le gène HOXB4, qui est connu pour son implication dans l'auto-renouvellement des cellules souches hématopoïétiques (Sauvageau et al., 1995). En effet, l'induction de HOXB4 dans des cellules souches embryonnaires murines permet la production de cellules souches hématopoïétiques capables d'assurer une greffe à long terme

(Kyba et al., 2002). Cependant, l'induction de HOXB4 dans des cellules humaines ne présente pas un effet similaire sur la prise de greffe (Wang et al., 2005) même si une amélioration de l'engagement vers la voie hématopoïétique peut être observée (Bowles et al., 2006). Le cluster des gènes HOXA étant également dérégulé chez l'humain (Wang et al., 2005), il est possible que la famille de gènes HOXA joue un rôle plus important chez l'humain, ce qui expliquerait la différence observée par rapport aux cellules murines (Ding et al., 2023). Une étude publiée en 2012 a montré qu'un blocage de la différenciation pouvait être relié à la sous-expression de certains micro-ARNs (Risueño et al., 2012), posant des questions sur un éventuel blocage dû à des cicatrices épigénétiques de la reprogrammation des cellules souches.

Ces doutes ont été levés avec les premiers protocoles ayant réussi à obtenir une prise de greffe (Amabile et al., 2013; Suzuki et al., 2013). Ces deux protocoles consistent à l'injection de cellules souches pluripotentes embryonnaires ou induites accompagnées de cellules souches stromales OP9 dans des souris afin d'induire la formation de tératomes. Il est ensuite possible d'extraire de ces tératomes des cellules souches hématopoïétiques fonctionnelles après 8 semaines. Ces protocoles ont ainsi montré qu'il n'existait pas de barrière intrinsèque à la différenciation en cellules souches hématopoïétiques, mais que les signaux du micro-environnement jouaient un rôle crucial.

Dans le même temps, en 2013, d'autres laboratoires ont exploré l'induction de multiples facteurs de transcription. Le laboratoire de Georges Daley a ainsi réussi à obtenir une prise de greffe à court terme (Doulatov et al., 2013) en induisant l'expression de 5 facteurs de transcription (ERG, HOXA9, RORA, SOX4 et MYB) dans des cellules souches hématopoïétiques dérivées de cellules souches pluripotentes. En 2017, le protocole fut amélioré pour obtenir une greffe multi-lignage à long terme (Sugimura et al., 2017) grâce à la surexpression de 7 facteurs de transcription (ERG, HOXA5, HOXA9, HOXA10, LCOR, RUNX1 et SPI1) dans des cellules CD34⁺ dérivées de cellules souches pluripotentes induites. La présence de 3 gènes de la famille HOXA confirme le rôle clé de cette famille dans la production de cellules souches hématopoïétiques compétentes chez l'Homme. D'autres laboratoires ont ensuite tenté de réduire le nombre de facteurs de transcription nécessaires. En 2018, le laboratoire de Yuet Wai Kan a proposé une reprogrammation n'utilisant que la protéine de fusion MLL-AF4 (Tan et al., 2018). Bien que les cellules produites soient capables de prise de greffe à long terme, la présence de transformations à caractère leucémique dans les cellules greffées limite les applications de ce protocole. Une reprogrammation partant de cellules endothéliales adultes, et n'utilisant que 4 facteurs : FOSB, GFI1, RUNX1 et SPI1 (Sandler et al., 2014), permet une greffe chez des récipients primaires et secondaires mais ne réussit cependant pas à apporter une réelle reconstitution multi-lignage.

Ainsi, la production de cellules souches à partir de cellules souches à des fins thérapeutiques reste encore un défi à relever. Des résultats encourageants ont pu être présentés en 2022 au congrès annuel de l'ISEH (*International Society of Experimental Hematology*) par Elizabeth Ng et Andrew Elefanty avec un protocole sans transgène ou co-culture permettant d'obtenir une reconstitution multi-lignage dans les souris primaires. Néanmoins, la capacité de prise de greffe de ces cellules dans des récepteurs secondaires n'a pas encore été atteinte (Ditadi & Sturgeon, 2023).

Modélisation du développement embryonnaire *in vitro*

Développement embryonnaire précoce de l'embryon humain

La première étape de ségrégation des feuillets dans l'embryon humain a lieu au stade blastula vers le 4^{ème} jour du développement de l'embryon (Shahbazi, 2020). Les cellules situées sur la périphérie extérieure forment le trophoctoderme qui va donner naissance au futur placenta, tandis que les cellules situées à l'intérieur forment la masse cellulaire interne qui va donner naissance aux différents feuillets embryonnaires et au sac vitellin.

Peu après l'implantation de l'embryon, les cellules souches du trophoctoderme vont proliférer afin d'envahir la caduque basilaire maternelle (Figure 3) en formant un syncytium primordial et en se différenciant en cytotrophoblastes (Knöfler et al., 2019). La fonction principale du syncytium primordial durant les deux premières semaines de l'embryon est de permettre son implantation dans la caduque basilaire en digérant l'endomètre et ses capillaires afin de former des lacunes trophoblastiques dans lesquelles va se déverser le sang maternel, permettant ainsi les échanges de nutriments et un apport accru en oxygène (Jaremek et al., 2021). Pendant ce temps, les cellules du cytotrophoblaste vont proliférer et former les villosités primaires du placenta qui vont progressivement recouvrir l'embryon et s'infiltrer dans le syncytium primordial. Le sommet des villosités placentaires va se différencier en trophoblaste extra-villeux qui va envahir la caduque basilaire maternelle pour remodeler ses vaisseaux sanguins (Velicky et al., 2018), tandis que la périphérie des villosités primaires va fusionner afin de former le syncytium définitif qui va assurer les transferts entre la mère et le fœtus : le syncytiotrophoblaste (Evain-Brion & Malassine, 2003).

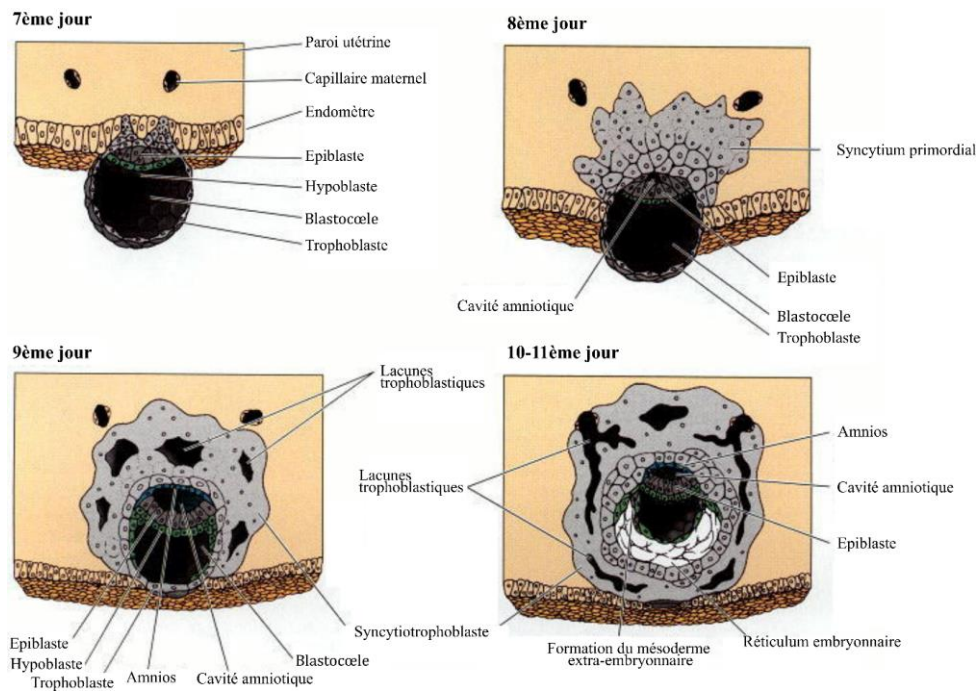


Figure 3 : Formation des tissus humains et implantation de l'embryon entre le 7ème et le 11ème jour du développement

Adapté de Gilbert et al. (Gilbert, 2000)

Dans le même temps, la seconde étape de ségrégation des feuillet de l'embryon va se produire avec la formation de l'hypoblaste et de l'épiblaste. L'hypoblaste, aussi appelé endoderme primitif, contribue à formation du sac vitellin et aux cellules mésodermiques extra-embryonnaires (Spencer Chapman et al., 2021) qui vont coloniser le placenta pour participer à la formation des villosités placentaires secondaires et donner naissance aux vaisseaux sanguins qui assurent le transport des éléments nutritifs de la mère à l'embryon (Enders & King, 1988). L'épiblaste est initialement composé de cellules pluripotentes, mais les cellules épiblastiques au contact du cytotrophoblaste vont adopter une morphologie épithéliale et former l'amnios (Shahbazi et al., 2017a) au cours du 10^{ème} jour du développement. La recherche sur les embryons humains en culture étant limitée aux 14 premiers jours du développement, la compréhension des étapes suivantes du développement provient surtout de collections d'embryons telles que la collection de Carnegie et d'expériences de modélisation *in vitro* du développement embryonnaire à partir de cellules souches (Ghimire et al., 2021). L'épiblaste prend une forme de disque épithélial (Hertig et al., 1956) et commence sa gastrulation après le 14^{ème} jour par la formation du sillon primitif. Le long de ce sillon primitif, les cellules épiblastiques réalisent une transition épithélio-mésenchymateuse et génèrent le mésoderme et l'endoderme alors qu'elles s'invaginent dans ce sillon primitif (Tam & Behringer, 1997). Les précurseurs de l'endoderme franchissent le sillon primitif les premières afin de s'intercaler avec les cellules de l'hypoblaste pour former l'endoderme définitif, puis sont suivies par les précurseurs du mésoderme qui vont s'infiltrer sous

l'épiblaste (Figure 4). La gastrulation se termine 21 jours après fécondation, les cellules ne s'étant pas invaginées dans le sillon primitif forment enfin le dernier feuillet embryonnaire : l'ectoderme.

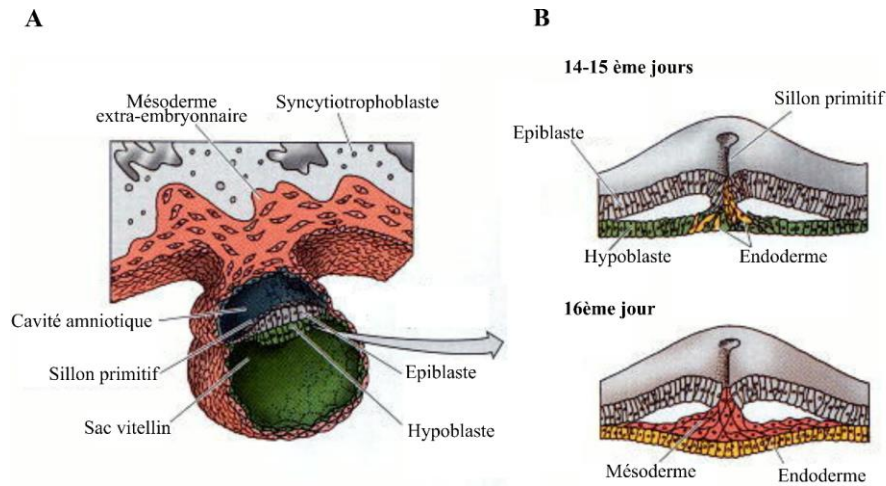


Figure 4 : **Embryon humain au début de la gastrulation**

(A) Embryon humain et ses connexions utérines au 15^{ème} jour du développement
 (B) Structure de l'embryon humain entre les 14^{ème} et 16^{ème} jours du développement
 Adapté de Gilbert et al. (Gilbert, 2000)

Modèles tridimensionnels *in vitro* du développement embryonnaire

En s'inspirant d'expériences déjà réalisées à partir de cellules souches embryonnaires murines, la mise en place de protocoles visant à agréger des cellules souches en trois dimensions afin d'obtenir une différenciation en trois feuillets embryonnaires s'est développée à partir des années 2000 (Itskovitz-Eldor et al., 2000). Cette technique peut être utilisée afin de développer des structures embryonnaires à différentes périodes critiques du développement (Weatherbee et al., 2021).

Modélisation de l'embryon

L'utilisation de structures tridimensionnelles, au contraire des cultures en deux dimensions souvent utilisées afin d'induire la différenciation vers un type cellulaire précis (Robinton & Daley, 2012), permettent l'apparition de schémas de différenciation beaucoup plus complexes (Rowe & Daley, 2019). Cependant, en l'absence de synchronisation entre les différents types cellulaires, il est impossible de modéliser un stade précis du développement embryonnaire. Les corps embryoïdes adaptent alors un profil similaire aux tératomes avec une différenciation vers différents types cellulaires des trois feuillets primordiaux (Flamier et al., 2017; Sheridan et al., 2012). Des protocoles ont progressivement été développés afin d'obtenir l'ensemble des types

cellulaires présents dans un organe donné dans des structures appelées organoïdes. On peut notamment citer les organoïdes intestinaux (Sato et al., 2009), rénaux (Xia et al., 2013), cardiaques (Ronaldson-Bouchard et al., 2018), pulmonaires (Jacob et al., 2017; McCauley et al., 2017) ou cérébraux (Lancaster et al., 2013; Mariani et al., 2012) qui réussissent à reproduire l'organisation spatiale des organes qu'ils modélisent et offrent ainsi une opportunité unique de modéliser l'impact de mutations génétiques sur le développement des organes.

Une grande avancée pour la modélisation du développement embryonnaire a été publiée en 2020, par le laboratoire d'Alfonso Martinez Arias (Moris et al., 2020), qui a réussi à obtenir une structure reproduisant l'élongation axiale de l'embryon entre les stades de Carnegie 8 et 9, à l'exception notable des structures neurales antérieures et des organes extra-embryonnaires. Cette structure récapitulant le développement de l'embryon au moment de la gastrulation a été dénommée gastruloïde et offre de nouvelles perspectives pour modéliser et comprendre le développement embryonnaire humain. Plus récemment, un laboratoire a développé un autre modèle récapitulant la somitogenèse, publié en 2022 (Sanaki-Matsumiya et al., 2022). On peut alors observer la formation progressive de structures similaires aux somites de l'embryon alors que le corps embryoïde s'allonge. Ces somitoïdes offrent ainsi la possibilité d'étudier les maladies liées aux cycles de segmentation de l'embryon telles que le syndrome de Jarcho-Levin ou la scoliose congénitale.

Ces modèles, bien que récapitulant des étapes de l'ontogenèse post-gastrulation, ne récapitulent pas les étapes pré-gastrulation, ni les interactions avec les cellules extra-embryonnaires qui peuvent jouer un rôle crucial lors du développement (Shahbazi, 2020). De plus, il semblerait qu'ils ne soient capables de récapituler qu'une fenêtre précise du développement. En effet les tissus générés ne parviennent généralement pas à continuer leur différenciation selon le processus prévu dans le cas du développement de l'embryon.

Modélisation des annexes extra-embryonnaires et cellules souches trophoblastiques

Les cellules souches pluripotentes humaines possèdent la capacité de se différencier en cellules présentant un phénotype proche des cellules des feuilletts extra-embryonnaires. Les premiers protocoles ayant mis en évidence ce potentiel utilisaient le BMP4 pour obtenir des cellules semblables au syncytiotrophoblaste ou au trophoblaste extra-villeux (R.-H. Xu et al., 2002). Cependant, l'identité exacte des cellules formées porte à débat car les cellules produites exhibent parfois un profil similaire aux cellules de l'amnios (Bernardo et al., 2011; Ohgushi et al., 2022).

Une autre piste intéressante pour la modélisation des populations extra-embryonnaires est l'utilisation de cellules souches trophoblastiques (parfois abrégées TSC pour *Trophoblastic Stem Cells*). De manière similaire aux cellules souches embryonnaires qui peuvent régénérer l'ensemble

des tissus de l'embryon, les cellules souches trophoblastiques sont capables de générer les différents types cellulaires présents dans le trophoblaste. Si l'isolation de cellules souches trophoblastiques a été possible dès 1998 chez la souris (Tanaka et al., 1998), l'isolation de cellules souches trophoblastiques humaines n'a été réalisée qu'en 2018 (Okae et al., 2018). De fait de la difficulté d'obtenir des échantillons, plusieurs laboratoires se sont intéressés à la différenciation de cellules souches pluripotentes en cellules souches trophoblastiques. Cette dernière fut atteinte aussi bien pour les cellules souches embryonnaires qu'induites en 2020 (Castel et al., 2020; Cinkornpumin et al., 2020). La différenciation à partir de cellules souches induites ouvre plusieurs perspectives thérapeutiques, notamment avec l'investigation de la contribution des mutations génétiques dans le développement du placenta et dans l'apparition de maladies telles que la prééclampsie ou les choriocarcinomes.

D'autres laboratoires se sont intéressés à modéliser l'amnios. Cette structure d'origine épiblastique apparaissant après l'implantation de l'embryon, il est relativement difficile d'étudier sa formation *in vivo*, d'autant plus que les protocoles actuels de culture *ex vivo* d'embryon ne parviennent pas à récapituler sa formation (Shahbazi et al., 2017b). En cultivant des cellules souches embryonnaires sur du matrigel et en utilisant une induction avec du BMP4, il est possible d'entraîner la formation d'une cavité similaire à la cavité pro-amniotique mais sans différenciation amniotique. L'utilisation de gels mous (Shao et al., 2017) et de dispositifs microfluidiques (Y. Zheng et al., 2019) a permis de grandement améliorer ces systèmes de différenciation avec l'apparition de cellules similaires à celle de l'amnios à une extrémité de la cavité et de cellules épiblastiques de l'autre. Cependant, la difficulté d'accès à des échantillons humains est un frein à la comparaison de ces structures avec leurs homologues *in vivo*. Pour résoudre ce problème, certains protocoles utilisent des cellules souches pluripotentes provenant de primates non-humains pour lesquels un atlas transcriptomique a été publié (Bergmann et al., 2022). Ainsi, ils ont réussi à générer des sphéroïdes épiblastiques et amniotiques récapitulant les stades de Carnegie 5 et 6 (Munger et al., 2022).

Le mésoderme extra-embryonnaire est également l'objet d'investigations. Récemment, le laboratoire de Vincent Pasqué (Pham et al., 2022) a réussi à produire ces cellules mésodermes extra-embryonnaires à partir de cellules souches pluripotentes. Bien que plusieurs résultats précédents, basés sur l'analyse d'embryons par microscopie électronique (Enders & King, 1988), des analyses transcriptomiques (Nakamura et al., 2016) ou de suivi de lignage (Spencer Chapman et al., 2021) aient amenés à supposer que ces cellules sont dérivées de l'hypoblaste, le système de culture développé ne produit pas de cellules de l'endoderme primitif. Il serait donc possible que ces cellules soient produites par l'épiblaste, comme suggéré par Patrick Lockett en 1978 (Lockett,

1978; Ross & Boroviak, 2020). L'origine embryonnaire du mésoderme extra-embryonnaire reste donc encore une question ouverte.

Blastoïdes et embryons synthétiques

L'unique modèle, publié à l'heure actuelle, arrivant à récapituler fidèlement les feuillets embryonnaires et extra-embryonnaires à partir de cellules souches humaines est celui des blastoïdes. En utilisant des cellules souches pluripotentes embryonnaires ou induites, plusieurs laboratoires sont parvenus en 2021 à produire des structures similaires au stade blastula de l'embryon (Yanagida et al., 2021; L. Yu et al., 2021). En 2022, le laboratoire de Nicolas Rivron (Kagawa et al., 2022) a pu montrer qu'il est possible de générer ces blastoïdes de manière à ce qu'ils récapitulent l'ordre d'apparition et la temporalité des différents types cellulaires *in vivo* et réagissent aux signaux implantatoires sécrétés par, des cellules de l'endomètre stimulées.

Une autre avancée remarquable a été réalisée en 2022 chez la souris simultanément par deux laboratoires à partir d'agrégats de cellules souches embryonnaires, de cellules souches extra-embryonnaires endodermiques et de cellules souches trophoblastiques (Amadei et al., 2022; Tarazi et al., 2022). Ces derniers, mis en culture avec des systèmes disposants de systèmes d'échanges de gaz, sont capables de produire des embryons synthétiques récapitulant jusqu'aux 8 premiers jours du développement murin, au début de la neurulation et de la morphogenèse. Le raffinement des conditions de culture et la perspective adaptation de ces protocoles aux cellules humaines pourraient permettre une bien meilleure compréhension du développement embryonnaire. Cependant, l'apparition de modèles toujours plus proche de l'embryon humain pose des questions éthiques qui ne sont pas encore encadrées par les lignes de conduites actuelles puisque la limite actuelle des 14 jours n'est pas adaptée à ces systèmes de différenciation (Rivron et al., 2018; Weatherbee et al., 2021).

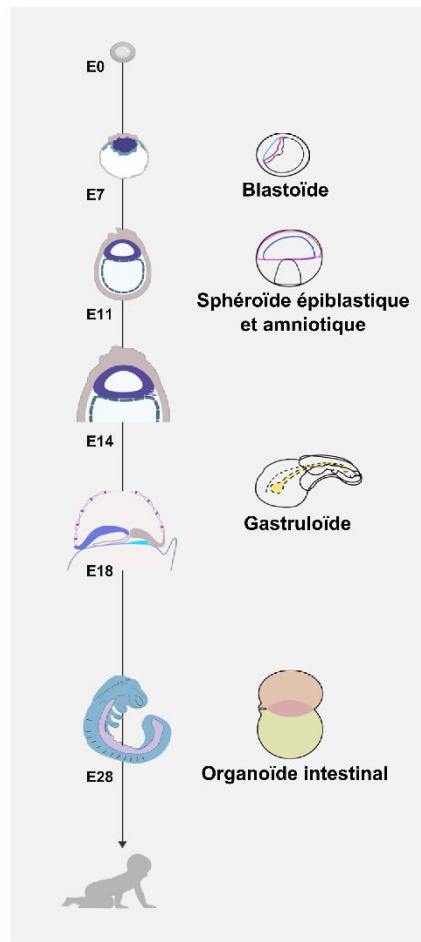


Figure 5 : *Les modèles tridimensionnels du développement humain*
 Adapté de Zhai et al. (Zhai et al., 2022)

Les avancées dans ces différents modèles ouvrent de nouvelles perspectives quant à la compréhension de l'hématopoïèse embryonnaire humaine. Du fait de la faible accessibilité des échantillons, le développement de méthodes d'analyses *in vitro* est nécessaire. Il est supposé que les cellules mésodermiques extra-embryonnaires sont à l'origine de la première vague d'hématopoïèse dans le sac vitellin (Ross & Boroviak, 2020). Ces cellules étant encore relativement peu caractérisées, leur dérivation *in vitro* offre donc l'opportunité d'étudier plus finement l'hématopoïèse primitive. De même, le développement de modèles plus tardifs, fidèles au développement embryonnaire, comme les gastruloïdes ou les somitoïdes, pourrait fournir un outil précieux pour l'analyse de la mise en place de l'aorte dorsale et de sa niche en fournissant des modèles plus complets que les systèmes de différenciation orientée développés jusqu'à maintenant.

Analyse du séquençage des ARNs en cellule unique (scRNAseq)

Extraction des échantillons

Les protocoles permettant d'isoler les ARNs de cellules uniques se divisent en deux catégories : les méthodes utilisant une séparation en plaques (*plate-based methods*) et les méthodes utilisant une séparation en gouttes (*droplet-based methods*). Ces méthodes se réalisent sur des cellules en suspension et supposent donc une étape de dissociation des cellules adaptée dans le cas de l'étude de tissus ou de structures multicellulaires.

La séparation en plaques

Les méthodes basées sur le tri en plaques visent à séparer les cellules dans des puits séparés afin que chaque cellule puisse être traitée indépendamment. Les protocoles les plus modernes utilisent généralement la cytométrie en flux afin de pouvoir automatiser le dépôt des cellules. Le protocole de séparation le plus utilisé à l'heure actuelle est le protocole SmartSeq dont la troisième version (SmartSeq-3) a été publiée en 2020 (Hagemann-Jensen et al., 2020). Une fois chaque cellule isolée, la paroi de chaque cellule est lysée, puis les ARNs sont amplifiés par rétrotranscription et réaction en chaîne par polymérase (souvent abrégée en rt-PCR pour *reverse transcription Polymerase Chain Reaction*). Une des améliorations de la méthode SmartSeq3 par rapport à l'itération précédente est l'ajout d'identifiants moléculaires uniques (ou UMI pour *Unique Molecular Identifier*) aux primers des oligonucléotides. Ces derniers sont des séquences nucléotidiques uniques à chaque primer et permettent de différencier les copies issues de l'amplification lors de l'étape de PCR des différentes copies de l'ARN présentes dans la cellule séquencée.

La séparation en gouttelettes

La séparation en gouttelette vise à isoler chaque cellule dans une goutte de quelques nanolitres. Ceci est possible grâce aux développements de systèmes microfluidiques qui peuvent isoler les cellules dans des gouttes non miscibles avec l'eau, formant ainsi une émulsion. La méthode la plus répandue à l'heure actuelle est la méthode Chromium (G. X. Y. Zheng et al., 2017). Avec cette méthode, chaque cellule est emprisonnée avec une bille comprenant des oligonucléotides se terminant par une répétition de thymine (queue poly-T). Ainsi, après l'étape de lyse de la cellule, les ARNs polyadénylés sont capturés par la bille. Chacun de ces oligonucléotides comprend également une séquence spécifique à chaque bille pour pouvoir différencier à posteriori à quelle cellule l'ARN séquencé appartient, ainsi qu'un UMI.

Il est à noter que le système microfluidique isole des billes sans aucune cellule. Ces dernières peuvent alors attraper des ARNs ambiants qui génèrent ensuite des artefacts dans les résultats du séquençage. Ces dernières doivent donc être supprimées du jeu de données lors de l'analyse. Dans le cas de l'utilisation de la suite CellRanger pour l'alignement des séquences, un filtre est automatiquement appliqué afin de les supprimer. Dans le cas contraire, l'algorithme EmptyDrops (A. T. L. Lun et al., 2019) permet de filtrer ces gouttes vides.

Une fois la librairie constituée avec la méthode de son choix, les échantillons peuvent être séquencés. Les séquences sont ensuite alignées sur le génome. Les pipelines les plus utilisés sont la pipeline STAR (Dobin et al., 2013) pour le SmarSeq3, ou la pipeline propriétaire CellRanger, fournie par 10XGenomics (G. X. Y. Zheng et al., 2017), pour le Chromium.

Avantages et inconvénients

La mise au point des techniques basées sur l'isolation en gouttes a permis une avancée considérable dans le nombre de cellules analysables. En effet, il est possible de traiter jusqu'à une dizaine de milliers de cellules en un seul échantillon avec le Chromium, là où le SmartSeq ne permet de s'intéresser qu'à des échantillons de l'ordre de la centaine de cellules. Néanmoins, les méthodes de séparation en plaques présentent certains avantages qui font qu'elles sont toujours utilisées pour certaines applications :

- Le nombre moyen de gènes détectés par cellules est plus élevé, ce qui rend ces techniques plus efficaces pour détecter des gènes faiblement exprimés (Ziegenhain et al., 2017)
- Le gène entier est séquencé (et non pas uniquement l'extrémité terminale 3' ou 5' comme pour le Chromium). Cela permet de s'intéresser à l'expression des différentes isoformes des gènes (Griffiths et al., 2018).
- La méthode de prélèvement permet un meilleur traçage de l'origine des cellules puisqu'il est possible de connaître avec précision les puits dans lesquels chaque partie de l'échantillon a été déposé, notamment dans le cas de dissection de différentes sous-parties d'un embryon (Tyser et al., 2021).

Ces deux techniques souffrent néanmoins de limitations inhérentes au séquençage de ARNs. Au-delà des limites des seuils de détection, il existe une différence significative entre les données transcriptomiques et la composition protéique de la cellule puisque la transcription des ARNs se produit par salves (Raj et al., 2006) alors que le niveau d'expression des protéines a tendance à adopter un profil plus lissé du fait de leur stabilité plus importante. Une absence de détection des ARNs n'est donc pas synonyme d'absence de protéine, notamment pour les protéines membranaires dont le temps de résidence à la membrane peut être relativement élevé.

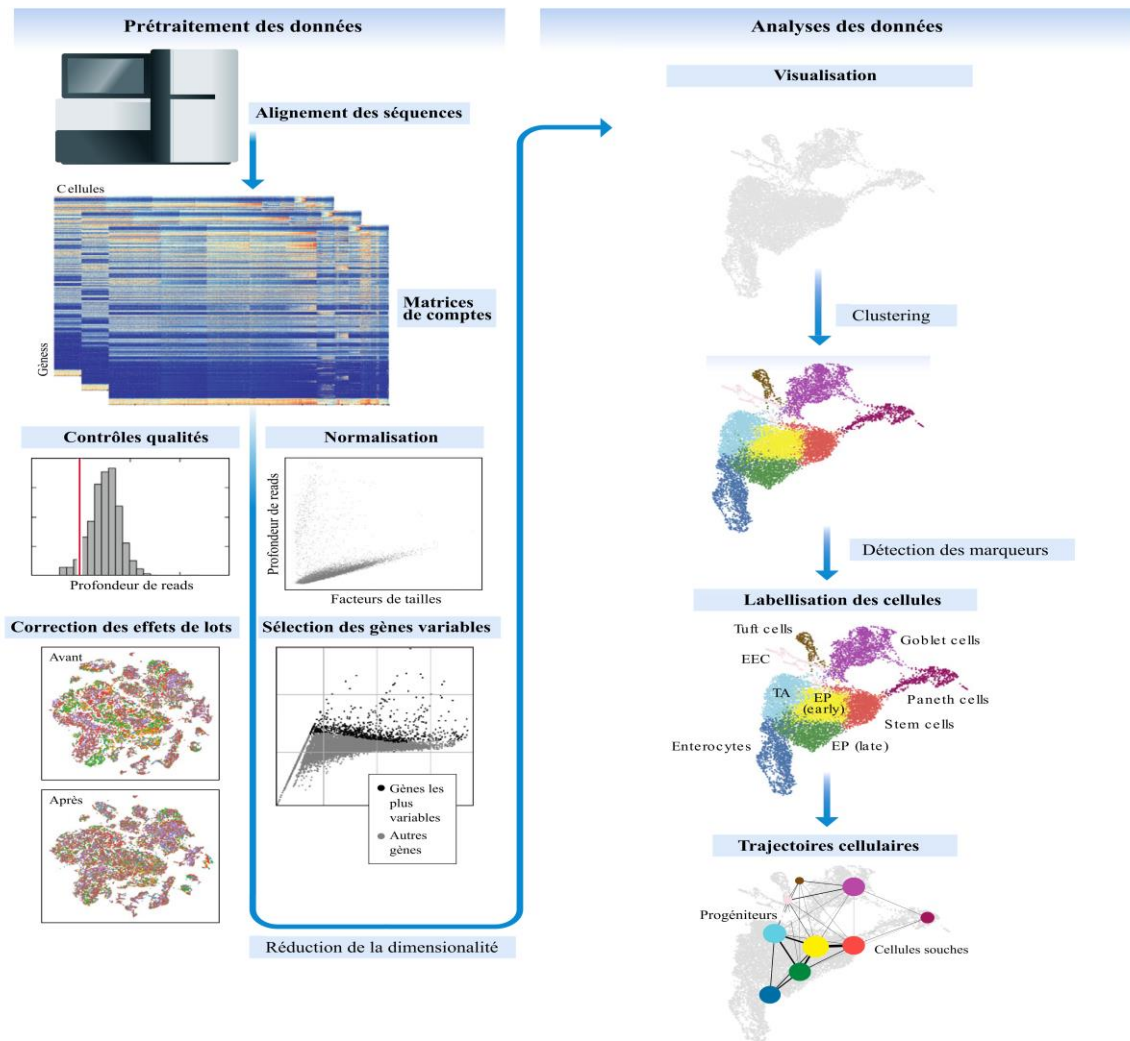


Figure 6 : Représentation schématique des étapes du prétraitement et de l'analyse d'un jeu de données scRNAseq
 Adapté de Luecken et al. (Luecken & Theis, 2019)

Contrôles qualités et sélection des gènes pour les analyses

La première étape de toute analyse de données de scRNAseq est le contrôle de la qualité des échantillons. Pour ce paragraphe, et les suivants, nous nous concentrerons sur le traitement d'échantillons à base d'UMI tels ceux produits par la technologie Chromium de 10X Genomics puisqu'il s'agit du protocole qui est utilisé pour les jeux de données dont les analyses sont présentées dans la partie « Résultats ».

Un premier contrôle de la qualité des données, relativement répandu, consiste à visualiser le nombre d'UMIs par cellule (i.e. le nombre d'ARNs différents capturés par cellule), ainsi que le

nombre de gènes différents détectés par cellule (Griffiths et al., 2018; Luecken & Theis, 2019). On s'attend généralement à une distribution unimodale avec éventuellement un second pic pour les faibles valeurs qui correspond généralement à des cellules mortes. Il est également possible de filtrer les cellules exprimant un nombre anormalement élevé de gènes et avec un grand nombre d'UMI car elles correspondent souvent à des doublets. Des filtres supplémentaires, basés sur des modèles statistiques plus développés, peuvent également être appliqués pour retirer ces doublets (McGinnis et al., 2019; Wolock et al., 2019).

Une autre procédure courante est de s'intéresser à la proportion d'ARNs mitochondriaux détectés. Les cellules présentant une forte proportion d'ARNs mitochondriaux et un faible nombre d'UMI et de gènes détectés sont généralement assimilées à des cellules mortes dont la membrane perforée a laissé s'échapper les ARNs contenus dans le cytoplasme (Ilicic et al., 2016). Une forte proportion d'ARNs mitochondriaux peut être liée à des processus respiratoires et ne suffit donc pas à exclure les cellules.

Il est généralement recommandé d'appliquer des filtres peu stricts (Germain et al., 2020) et d'affiner ces derniers si certaines des cellules retenues se montrent délétères pour la suite des analyses. Cela permet d'éviter d'éliminer des populations ayant des profils peu conventionnels de l'analyse avant même d'avoir pu vérifier leur identité.

Une seconde étape est la sélection des gènes utilisée pour la suite de l'analyse. Afin de limiter la taille du jeu de données et d'accélérer le temps d'exécution des algorithmes utilisés par la suite, on retire du jeu de données les gènes exprimés dans très peu de cellules (le plus souvent moins de 10). Ensuite, on raffine cette sélection en établissant une liste des gènes les plus variables (généralement abrégés en HVG pour *Highly Variable Genes*). L'idée est que les gènes exprimés mais ne variant pas ne permettent pas de discriminer les cellules et sont donc peu utiles pour les analyses de groupement ou d'identification des types cellulaires.

La méthode la plus couramment utilisée consiste à sélectionner entre 1000 et 5000 gènes les plus variables pour la suite des analyses. Cette sélection doit cependant s'effectuer en prenant en compte la dépendance entre la variance d'un gène et son niveau moyen d'expression. Ne sélectionner que les gènes présentant la plus grande variance entraînerait un biais pour les gènes les plus exprimés. Les approches les plus courantes utilisent donc des modèles reposant sur des lois de Poisson (Amezquita et al., 2020) ou des lois binomiales (Hafemeister & Satija, 2019) afin d'estimer une variance corrigée pour le niveau moyen d'expression. D'autres solutions alternatives ont également été proposées comme la sélection en utilisant la déviance (Townes et al., 2019). La plupart des analyses réalisées par la suite sont généralement assez robustes au nombre de gènes sélectionnés lors de cette étape (Klein et al., 2015), il est donc recommandé de

sélectionner un nombre élevé de gènes (Luecken & Theis, 2019). Les différences observées entre méthodes diminuant avec le nombre de gènes sélectionnés (Germain et al., 2020), la méthode de sélection exacte de ces gènes, bien que pouvant influencer sur les analyses, ne semble pas avoir d'impact majeur sur les résultats finaux.

Normalisation

Motivation et normalisation par taille des librairies

Une fois les contrôles qualités effectués, une étape cruciale du traitement des données est la normalisation de ces dernières. Les motivations derrière cette étape sont les suivantes :

- Une grande partie des variations observées peut être dû à des effets techniques (taux de capture des ARNs par les billes, efficacité de l'étape de PCR, ...)
- La variance des données est corrélée au niveau d'expression des cellules
- Les différences des niveaux d'expression d'un gène entre cellules sont souvent moins informatives que les différences de ratio d'expression (*fold-change*)

Les premières approches proposées pour traiter les deux premiers points ont été inspirées de celles utilisées pour le RNAseq en utilisant des facteurs de taille (*size factors*). L'approche initiale consiste à diviser le niveau d'expression de chaque gène par le nombre d'UMI détectés dans la cellule, ce qui correspond à la taille de la librairie que représente chaque cellule. On multiplie généralement ce résultat par un million afin d'obtenir le nombre théorique d'ARN que l'on détecterait pour ce gène si l'on séquençait exactement un million d'ARN pour cette cellule (dénommé CPM pour *Count per million*) :

$$CPM = \frac{\text{Expression du gène}}{\text{Nombre d'UMI détectés dans la cellule}} * 10^6$$

Pour s'affranchir du troisième point, une transformation logarithmique est appliquée. Du fait de la propriété de morphisme du logarithme, regarder une différence entre les données transformées revient à regarder le ratio des données non transformées. Ainsi, les distances euclidiennes qui seront exploitées par la suite seront proportionnelles au *fold-change* entre les cellules et non plus aux différences de niveau d'expression. Cependant, la divergence de la fonction logarithme en 0 n'est pas compatible avec les gènes dont le niveau de détection est égal à 0 avant normalisation. Pour pallier ce problème, on ajoute une valeur constante, appelée pseudo-compte au ratio avant la transformation. Le pseudo-compte est le plus souvent fixé à 1 car il permet d'obtenir une valeur

normalisée de 0 pour un gène dont le niveau d'expression serait nul. On obtient donc la procédure de normalisation suivante :

$$\text{Expression normalisée du gène} = \log_2 \left(1 + \frac{\text{Expression du gène}}{\text{Nombre d'UMI détectés dans la cellule}} * 10^6 \right)$$

Dans certains cas, une valeur différente de 1 peut être utilisée pour le pseudo-compte, si l'on désire n'avoir aucune valeur nulle dans la suite de nos analyses (ce qui est le cas de l'algorithme LandSCENT (W. Chen et al., 2019) qui se base sur des ratios d'expression normalisée entre gènes) ou si l'on désire l'utiliser comme paramètre d'ajustement de la normalisation (A. Lun, 2018).

Un autre avantage de la transformation logarithmique réside dans le fait qu'elle va rapprocher le profil d'expression des gènes d'une distribution suivant la loi normale pour lesquels un certain nombre d'algorithmes utilisés dans la suite de nos analyses ont été optimisés (Luecken & Theis, 2019).

Néanmoins, cette approche comporte plusieurs défauts :

- La correction de la dépendance moyenne-variance n'est que partielle
- Les données de scRNAseq sont souvent hétéroscédastiques, i.e. la variance des données varient en fonction de certaines sous-populations cellulaires, et bénéficieraient d'une méthode prenant en compte cette complexité
- La normalisation par la taille de la librairie suppose que la majorité des gènes ne sont pas différentiellement exprimés, ce qui n'est généralement pas le cas dans les jeux de données comportant plusieurs types cellulaires distincts

Pour remédier à ces défauts, plusieurs méthodes ont été développées, soit en ajustant les méthodes visant à déterminer le facteur de normalisation, soit en cherchant à modéliser la distribution théorique de l'expression des gènes pour appliquer dans transformations non logarithmiques visant à stabiliser la variance.

Normalisation par groupement et déconvolution des cellules

La méthode *scran* (A. T. L. Lun et al., 2016) se base également sur des facteurs de taille tout en cherchant à limiter l'impact des valeurs nulles dues à la non capture d'un ARN et à s'affranchir de l'hypothèse selon laquelle la majorité des gènes ne sont pas différentiellement exprimés entre les cellules du jeu de données.

Pour limiter l'impact des zéros dus aux limites des taux de capture pour les gènes peu exprimés, chaque cellule est groupée avec les cellules présentant un niveau d'expression proche. L'expression de chaque gène dans la pseudo-cellule est ensuite divisée par la somme des

expressions de ce gène dans toutes les cellules. On obtient de cette manière, pour un gène non différentiellement exprimé, une valeur proportionnelle aux biais d'origine technique de la pseudo-cellule et indépendante de la valeur d'expression initiale du gène. Ce score est calculé pour chaque gène dans chaque pseudo-cellule et la valeur médiane du score de l'ensemble des gènes de chaque pseudo-cellule est gardée comme facteur de taille de la pseudo-cellule. On peut ensuite obtenir un facteur pour chaque cellule individuelle par déconvolution en résolvant un système linéaire d'équations. On obtient ainsi un facteur de taille moins sensible au bruit car son estimation a été « lissée » par le recours aux pseudo-cellules.

La méthode décrite ci-dessus ne fonctionne cependant que si la majorité des gènes ne sont pas différentiellement exprimés. En effet, le facteur estimé pour la pseudo-cellule n'est indépendant de l'expression du gène que si ce dernier n'est pas différentiellement exprimé. L'utilisation de la médiane sur l'ensemble des gènes permet d'avoir un certain pourcentage de gènes différentiellement exprimés mais ne suffit pas dans le cas de jeux de données très hétérogènes. Pour circonvenir à ce problème, un clustering hiérarchique est réalisé afin de grouper ensemble les cellules ayant des profils d'expression suffisamment similaires. Chaque cluster est traité de manière isolée, puis les clusters sont normalisés par un facteur propre à chaque cluster afin d'obtenir des échantillons comparables. De cette manière, on peut s'affranchir de l'hypothèse selon laquelle la majorité des gènes ne sont pas différentiellement exprimés pour la normalisation.

Cette méthode de normalisation a représenté une nette amélioration par rapport aux techniques précédentes comme montré lors d'une analyse comparative (Vieth et al., 2019) et a donc été pendant considéré pendant plusieurs années comme la méthode de référence à utiliser (Luecken & Theis, 2019).

Normalisation par transformation stabilisant la variance

Le choix du modèle statistique approprié pour les jeux de données scRNAseq a fortement été soumis à débat. Suite à la constatation que les données générées par les techniques n'utilisant pas d'UMI présentaient un nombre beaucoup plus élevé de zéros que ce qui était attendu (Kharchenko et al., 2014), plusieurs algorithmes ont été développés en utilisant des distributions avec des excès en zéro (*zero-inflated distributions*) (Risso et al., 2018). Cette distribution excessive en zéro a souvent été associée au terme *dropouts* qui est censé désigner l'absence de détection d'ARN pour des gènes qui sont réellement exprimés. Bien que le terme ait parfois été utilisé pour désigner tous les zéros d'ordre technique, ces derniers sont à différencier des valeurs nulles observées pour des gènes peu exprimés du fait des limites de détection de la technique, puisqu'il est attendu qu'un gène peu exprimé puisse ne pas être capturé ou qu'il ne soit pas suffisamment amplifié lors de l'étape de PCR pour être détecté lors du séquençage. Lors de l'essor des techniques basées sur les

UMI, les méthodes développées à partir de distributions présentant des excès en zéro ont continuées à être appliquées.

En parallèle, d'autres techniques se sont développées et ont plutôt opté pour des modélisations basées sur des lois de Poisson (T. H. Kim et al., 2020; Townes et al., 2019) ou binomiales négatives sans excès de zéro (Grün et al., 2014; Hafemeister & Satija, 2019). Néanmoins, l'analyse de jeux de données issus de populations pures triées indique que les méthodes basées sur l'utilisation d'UMI ne produisent pas de données présentant des excès en zéro (Svensson, 2020; Townes et al., 2019). Il semblerait donc que les modèles basés sur des distributions avec excès en zéros ne soient pas adaptés aux protocoles les plus récents qui disposent des UMIs. Cependant, il a été remarqué que différents types cellulaires au sein d'un même échantillon peuvent présenter des distributions avec des paramètres différents et dont la superposition peut donner une distribution globale similaire à une loi présentant des excès en zéro (T. H. Kim et al., 2020).

Une des méthodes les plus populaires pour la normalisation des données est la normalisation *sctransform* (Hafemeister & Satija, 2019) du package *Seurat*. L'idée sous-jacente de cette normalisation est de modéliser la distribution théorique des gènes afin de pouvoir appliquer une transformation qualifiée de stabilisatrice de la variance, qui supprime la corrélation entre la variance d'un gène et son niveau d'expression. Le fonctionnement de cet algorithme est le suivant :

- L'expression de chaque gène est modélisée par une loi binomiale négative en se basant sur le niveau d'expression dans l'ensemble des cellules
- Les paramètres des modèles obtenues sont ensuite lissés en les remplaçant par une estimation par régression du noyau afin d'obtenir une « valeur moyenne » des paramètres attendus en fonction du niveau d'expression. Cette étape de régularisation permet de s'affranchir des problèmes de surapprentissage du modèle.
- On applique pour chaque gène dans chaque cellule le modèle régularisé et on compare le niveau d'expression réel du gène avec le modèle. On garde ensuite le résidu de Pearson (i.e. la différence entre le niveau d'expression réel et le modèle, normalisée par la variance du modèle).

On obtient ainsi une valeur normalisée réduite pour la suite de l'analyse. Si cette valeur est positive, elle indique que le niveau d'expression du gène dans cette cellule est supérieur à ce qui serait attendu par rapport à l'expression moyenne du gène dans le jeu de données et en tenant compte du nombre d'ARNs exprimés par la cellule.

Cette méthode, bien que se basant sur la distribution théorique des gènes dans les cellules, comporte quelques défauts. Afin d'éviter certains problèmes pouvant surgir lors de l'analyse de jeux de données présentant une distribution peu conventionnelle, il faut avoir à l'esprit que :

- La modélisation de l'expression des gènes par une loi binomiale négative suppose que la proportion des cellules dans lesquelles un gène est différentiellement exprimé n'est pas trop importante. L'étape de régularisation relaxe cependant cette hypothèse en autorisant qu'une partie des gènes ne suivent pas la distribution théorique (Hafemeister & Satija, 2019).
- Le passage d'une matrice creuse à une matrice dense augmente considérablement la taille des données, et les temps de calculs (A. Lun, 2020) (c'est pour cette raison que seule l'expression des gènes les plus variables est calculée par défaut).
- L'estimation pour les gènes faiblement exprimés est surdispersée (Lause et al., 2021)
- Les valeurs normalisées ne sont pas aussi interprétables que les données ayant subi une transformation logarithmique et ne donnent pas de bons résultats pour l'analyse des gènes différentiellement exprimés (Crowell et al., 2020).

Malgré ces défauts, *sctransform* semble mieux performer que ses concurrents (Germain et al., 2020) pour une grande partie des analyses, notamment le clustering et la réduction de la dimensionnalité. Pour pallier certains des défauts évoqués plus haut, et en s'appuyant sur les résultats d'une étude en 2021 (Lause et al., 2021), une amélioration de l'algorithme a été publiée par les auteurs en 2022 (Choudhary & Satija, 2022). Cette nouvelle version améliore l'estimation des modèles, exclut de la régularisation les gènes présentant une faible expression ou ayant une variance trop faible ; et empêche les gènes très peu exprimés d'atteindre de grandes valeurs. D'après les résultats montrés par les auteurs, ces améliorations semblent donner de meilleurs résultats pour l'analyse des gènes différentiellement exprimés.

Réduction de la dimensionnalité

Une autre étape cruciale dans l'analyse des données de scRNAseq est la réduction de la dimensionnalité. Son but est de représenter les données dans un espace en plus petite dimension, tout en essayant de garder les informations dont nous avons besoin pour la suite des analyses. Par défaut, la dimension des données de scRNAseq est le nombre de gènes dans le génome de l'individu étudié. La sélection des gènes les plus variables évoquée plus haut est une forme de réduction de la dimensionnalité. Nous évoquerons ici deux types de réduction de la

dimensionnalité : celles visant à réduire la taille des données pour les analyses futures et celles utilisées pour la visualisation des données.

Réduction de la taille des données

L'hypothèse sous-jacente à l'utilisation d'un algorithme de réduction de la dimensionnalité pour réduire la taille des données est que l'information biologique contenue dans les données initiales peut être représentée sans réelles pertes dans un espace de dimension bien inférieure aux 30000 dimensions de départ. Dans le cadre des données de scRNAseq, cette hypothèse semble pour l'instant être totalement justifiée. L'algorithme le plus utilisé pour la réduction des données en scRNAseq est l'analyse en composantes principales (Hotelling, 1933) (abrégée par PCA pour *Principal Component Analysis*). La PCA consiste à réaliser un changement de repère orthonormé de manière à ce que la variance expliquée par chaque dimension soit maximale. Ainsi, la première dimension de la PCA sera celle qui comportera la plus grande variance possible étant donné le jeu de données ; la deuxième composante sera celle comportant la plus grande variance si l'on exclue la première composante du jeu de données ; idem pour la troisième si l'on exclut les deux premières, etc. Le procédé mathématique permettant de calculer ces composantes consiste à prendre le changement de repère qui permet de diagonaliser la matrice de covariance du jeu de données. Ensuite, on peut ne garder que les premières dimensions de la PCA qui vont représenter la majorité de la variance du jeu de données et supprimer les autres que l'on assimile à du bruit ou à des effets négligeables. Il est à noter que les dimensions de la PCA ne représentent pas des gènes en particulier mais des combinaisons de gènes.

Il existe encore un débat dans la communauté concernant la pertinence d'une étape de réduction des données (*scaling* en anglais) avant l'application de la PCA. La réduction des données consiste à retrancher à chaque gène sa valeur moyenne puis à le diviser par sa variance. On obtient ainsi une valeur moyenne nulle et une variance équivalente sur l'ensemble des gènes analysés. L'utilisation de données réduites revient donc à donner une importance égale à tous les gènes dans la PCA, alors qu'utiliser des valeurs non réduites donne plus de poids aux gènes présentant une variance plus forte. Certains considèrent que cette étape retire une partie du signal biologique pour la suite de l'analyse et ne devrait donc pas être appliquée (A. Lun, 2019b). Une étude comparative a cependant montré qu'elle semble avoir une influence positive sur l'efficacité du clustering dans la majorité des cas (Germain et al., 2020). Il est à noter que les données en sortie de *sctransform* sont des valeurs réduites par construction (Hafemeister & Satija, 2019) et que l'utilisateur est donc forcé d'utiliser des données sous forme réduite pour la PCA.

Parmi les autres méthodes classiques de l'analyse de données qui sont utilisées pour la réduction de dimension du scRNAseq, on peut citer l'analyse en composantes indépendantes (Hyvarinen, 1999) (ICA pour *independent component analysis*) qui cherche à décomposer les données selon

des dimensions indépendantes sur le plan statistique et dont chacune serait porteuse d'un signal biologique différent, et la factorisation en matrices non-négatives (Lee & Seung, 1999; Paatero et al., 1991) (NNMF pour *non-negative matrix factorization*). L'ICA est parfois utilisée car elle peut permettre de s'affranchir aisément de certains effets de lots (Nazarov et al., 2019). La NNMF a l'avantage de regrouper naturellement les gènes de manière aisément interprétables et une variante de cette dernière est notamment utilisée par l'algorithme PopAlign (S. Chen et al., 2020).

D'autres méthodes ont été développées spécifiquement pour réduire les données de scRNAseq. Parmi celles-ci, on peut citer GLM-PCA (Generalized Linear Model Principal Component Analysis). La projection opérée par la PCA peut-être caractérisée comme la maximisation d'une fonction de vraisemblance d'un modèle Gaussien. GLM-PCA réalise la même opération en utilisant un modèle basé sur une loi binomiale négative ou une loi de Poisson (la seconde version est préférée pour des raisons de stabilité). Comme la projection prend en compte la distribution statistique sous-jacente aux données de scRNAseq, l'étape de normalisation n'est pas nécessaire. Il existe certaines similitudes entre GLM-PCA et la normalisation *sctransform* puisque le modèle sous-jacent à *sctransform* est équivalent à une approximation au premier ordre de la GLM-PCA (Lause et al., 2021).

Visualisation des données

La réduction de la dimensionnalité peut également servir à obtenir une représentation en deux ou trois dimensions à des fins de visualisation. Ces dernières sont particulièrement utiles afin de se représenter le jeu de données et d'observer l'expression individuelles de gènes. Les algorithmes de visualisation les plus utilisés sont les algorithmes t-SNE et UMAP.

L'algorithme t-SNE (van der Maaten & Hinton, 2008) (*t-distributed Stochastic Neighbor Embedding*) est une technique visant à conserver les notions de voisinage entre les points. Le fonctionnement est le suivant (on suppose ici que l'on veuille projeter en deux dimensions) :

- Les distances entre les points dans le jeu de données initiales est converti en une distribution de probabilité gaussienne qui modélise la similarité entre les points
- Les cellules sont réparties en deux dimensions de manière aléatoire.
- On convertit les distances de l'espace en deux dimensions en une distribution de probabilité suivant la loi de Student avec un degré de liberté (la distribution gaussienne n'est pas utilisée ici car utiliser une distribution avec une queue plus lourde est plus adéquate afin de mieux ségréger les populations).
- On modifie la position des points dans l'espace en deux dimensions en faisant converger les distributions de probabilité. Pour cela, on réalise une descente de gradient

stochastique qui vise à minimiser la divergence de Kullback-Leibler entre les deux distributions.

On obtient ainsi une carte en deux dimensions où les points proches dans le jeu de données initial sont également proches sur la carte et réciproquement. Néanmoins, du fait de l'utilisation de la divergence de Kullback-Leibler, les moyennes et grandes distances ne sont pas conservées. Le t-SNE a été rapidement adopté car il donne une représentation en deux dimensions où les différents types cellulaires sont généralement bien ségrégués.

L'algorithme UMAP (McInnes et al., 2018) (*Uniform Manifold Approximation and Projection*) utilise une approche similaire :

- On établit un graphe orienté des plus proches voisins dans le jeu de données d'origine où chaque arête possède un poids similaire à la probabilité gaussienne utilisée pour le t-SNE
- La représentation en deux ou trois dimensions est initialisée avec le *Laplacian Eigenmap* (Belkin & Niyogi, 2003) (traduisible par espace propre du laplacien) et l'on calcule un graphe des plus proches voisins de manière similaire à l'étape précédente.
- Chaque point de la représentation en deux ou trois dimensions est ensuite ajusté par descente de gradient visant à maximiser l'entropie croisée entre les arêtes des deux graphes.

Les deux algorithmes peuvent sembler très similaires, mais certaines différences, en apparence subtiles, ont une influence considérable sur le résultat final. Le recours à l'entropie croisée au lieu de la divergence de Kullback-Liebler améliore grandement la préservation des moyennes et grandes distances et l'initialisation non aléatoire améliore la convergence de l'algorithme. Un aspect intéressant de ces améliorations est qu'elles s'appuient sur des fondements théoriques basés sur la théorie de la variété et la topologie, qui sont détaillés dans le papier princeps (McInnes et al., 2018).

L'algorithme UMAP s'est rapidement démocratisé en raison de ses meilleurs résultats sur la conservation des distances, sa capacité à mieux récapituler les trajectoires de différenciation (là où le t-SNE peut avoir un effet patchwork) et sa vitesse d'exécution (Becht et al., 2019). Cependant, il semblerait qu'une partie de ses différences puisse être compensée en changeant le paramétrage de l'algorithme t-SNE en modifiant l'initialisation de la représentation en deux dimensions pour les premières dimensions de la PCA (Kobak & Linderman, 2021) et en augmentant la perplexité (Kobak & Berens, 2019). En effet, la perplexité définit l'étalement de la distribution gaussienne utilisée par l'algorithme t-SNE. Augmenter la perplexité donne donc un poids plus important aux distances entre points plus éloignés et permet donc une meilleure estimation des distances.

Néanmoins, l'augmentation de la perplexité rallonge les temps de calculs (Oskolkov, 2019) et peut ralentir fortement la convergence, ce qui donne une représentation similaire aux dimensions de la PCA utilisées pour l'initialisation dans des cas extrêmes (Oskolkov, 2020a, 2020b).

L'utilisation de la PCA, suivie d'une étape non linéaire telle que le t-SNE ou l'UMAP est parfois critiquée car la préservation des voisinages (Cooley et al., 2019) et des distances (Chari et al., 2021) serait beaucoup moins bonne que ce que les utilisateurs de ces algorithmes peuvent penser et ne représenterait pas assez fidèlement les données d'origine. Malheureusement, il n'existe pas encore d'alternative corrigeant ces défauts et qui soit adaptée pour de l'analyse exploratoire. Il convient cependant de rester vigilant sur les limites des algorithmes utilisés et sur l'interprétation qui pourrait en découler. Les représentations en faible dimensions ne sont que des approximations du jeu de données étudié et ne représentent majoritairement que des distances déformées. Elles sont un outil précieux pour l'analyse de données de scRNAseq ; mais, comme tout outil, l'utilisateur se doit de l'utiliser en ayant conscience de ses limitations.

Clustering et labellisation des cellules

L'étape suivante dans l'analyse des données consiste souvent en l'identification des populations présentes. Deux approches complémentaires sont généralement utilisées : le partitionnement des données (connu également sous le nom de *clustering*) et l'annotation automatique. Le partitionnement des données vise à regrouper ensemble les cellules exhibant un profil similaire qui vont ensuite être annotées manuellement, alors que l'annotation automatique vise à comparer notre jeu de données avec des bases de référence pour connaître l'identité des cellules.

Clustering

Les premières approches de clustering utilisées pour le traitement de données de scRNAseq se sont servies de la méthode de partitionnement en k -moyennes (ou *k-means*) (Buettner et al., 2015; T. Kim et al., 2019; Kiselev et al., 2017). Cette approche classique (Macqueen, 1965) vise à regrouper les cellules en un nombre prédéterminé k de clusters. Pour ce faire, k cellules, qui seront dénommées centroïdes, sont tirés au hasard dans le jeu de données et un numéro de cluster est attribué à chacun de ces centroïdes. Ensuite, chaque cellule est attribuée au cluster du centroïde le plus proche. Une fois les clusters constitués, on réitère l'algorithme en choisissant pour centroïde la valeur moyenne de chaque cluster. Cette opération est répétée jusqu'à ce que l'algorithme converge, c'est-à-dire jusqu'à ce que la position des centroïdes ne varie plus d'une itération à l'autre. Les clusters finaux conservés sont ceux obtenus lors de la dernière itération de l'algorithme.

La seconde approche développée est celle des détections de communauté. Ces dernières se basent sur le graphe des plus proches voisins. Plutôt que de chercher à grouper les cellules ensemble en se basant sur leur niveau d'expression, ces algorithmes utilisent les relations de voisinage entre cellules. Un avantage considérable de ces algorithmes est leur vitesse d'exécution car ils n'utilisent que le graphe des plus proches voisins qui est de taille bien plus réduite que le jeu de données utilisé pour un algorithme comme le *k-means* (Luecken & Theis, 2019). Le package *Scran* utilise l'algorithme *walktrap* (Pons & Latapy, 2005) qui se base sur des marches aléatoires afin de détecter les communautés. Le package *Seurat*, lui, utilise la méthode de Louvain (Blondel et al., 2008) qui cherche à optimiser la modularité des communautés formées. La modularité mesure la densité d'arêtes comprise à l'intérieur des communautés par rapport aux nombres d'arêtes partagés entre deux communautés. Ainsi, une modularité élevée correspond à des communautés dont les cellules sont très connectées entre elles sur le graphe des plus proches voisins.

Les études comparatives publiées en 2018 (Duò et al., 2018; Freytag et al., 2018) ont montré que la méthode de Louvain était la méthode de clustering la plus efficace pour les jeux de données de scRNAseq. Cette méthode est désormais la méthode de clustering par défaut de plusieurs packages, dont *Seurat* et *Scanpy*. Une amélioration de la méthode de Louvain, nommée la méthode de Leyde (Traag et al., 2019), a également été implémentée. Néanmoins, les implémentations de cette dernière étant plus demandeuses en capacité de calcul, elle ne peut être utilisée pour les grands jeux de données.

Annotation manuelle des identités cellulaires

Une fois les cellules clustérisées, la labellisation des identités se fait en utilisant les gènes différentiellement exprimés entre les clusters. Dans le cas de l'identification de marqueurs, des tests statistiques simples tels que le t-test, le test des rangs signés de Wilcoxon ou le TF-IDF (Young, 2020) sont généralement utilisés pour identifier les gènes surexprimés dans chaque cluster. Les gènes les plus significatifs peuvent ensuite être comparés à diverses bases de données. Les tests comparatifs (Soneson & Robinson, 2018; Vieth et al., 2019) réalisés semblent indiquer que l'algorithme le plus performant est *limma-trend* (Law et al., 2014), un algorithme initialement développé pour le RNAseq.

Stratégies d'annotation automatique

L'annotation automatique se base sur l'assignation d'identités cellulaires à chacune des cellules du jeu de données. Le fonctionnement de ces algorithmes se base sur la comparaison des cellules à annoter avec une base de données et assigne à chaque cellule l'identité correspondant au type cellulaire le plus proche disponible dans la base de données.

Un des premiers algorithmes ayant été créé dans ce but, *SingleR* (Aran et al., 2019), utilise les coefficients de corrélation de Spearman de l'expression des gènes variables entre le jeu de données de référence et les cellules à annoter. Bien que le package possède plusieurs bases de données de référence, il est possible pour l'utilisateur de définir de nouvelles bases de données à partir de ses propres données. Une approche similaire a été déployée en 2020 par le laboratoire de Guoji Guo sur un atlas cellulaire humain de plus de 700 000 cellules couvrant des organes fœtaux et adultes (Han et al., 2020), fournissant ainsi une des bases de données les plus complètes existant à ce jour pour l'Homme.

D'autres approches algorithmiques pour réaliser les comparaisons ont ensuite été développées, notamment en utilisant des réseaux neuronaux (Ma & Pellegrini, 2019), des modèles linéaires généralisés (Pliner et al., 2019) ou encore l'apprentissage par transfert (Lotfollahi et al., 2022). Une initiative récente, dénommée Azimuth (Hao et al., 2021), portée par le laboratoire de Rahul Satija et un projet d'atlas biomoléculaire des instituts américains de la santé (HuBMAP), produit des atlas de différents organes humains et murins qu'il est possible d'utiliser pour annoter ses données en utilisant un algorithme se basant sur l'analyse en composantes principales supervisées (Barshan et al., 2011).

Une limitation importante de ces stratégies repose sur la base de données utilisée pour construire les types cellulaires de référence. Il n'est en effet possible que de détecter les types cellulaires présents dans la base de données de référence. De plus, dans le cas où le type cellulaire de la cellule observé n'est pas présent dans le jeu de données de référence, ces algorithmes peuvent amener à de mauvaises annotations. Il incombe donc à l'utilisateur d'être vigilant sur les bases de données utilisées et, a fortiori, sur les scores de confiance attribués aux identités afin de détecter d'éventuelles mauvaises attributions. Enfin, il est généralement recommandé de vérifier manuellement l'expression de gènes marqueurs présents dans la bibliographie.

Intégration de multiples jeux de données

Un grand défi du traitement des jeux de données de scRNAseq est l'analyse simultanée de plusieurs jeux de données. La grande sensibilité de cette technique aux protocoles d'obtention des jeux de données ou aux conditions expérimentales entraîne souvent des variations importantes qualifiées d'effets de lots (*batch-effect* en anglais). Afin de pouvoir comparer les données provenant de plusieurs échantillons, plusieurs techniques dites d'intégration ont été mises au point. L'idée sous-jacente à ces dernières est d'appliquer une réduction de la dimensionnalité afin d'obtenir un espace latent dans lequel les effets de lots ne sont pas présents tout en gardant le

signal biologique que l'on souhaite étudier. Un grand nombre de méthodes différentes ont été développées, certaines s'appliquant directement sur les matrices de données brutes, d'autres sur les données après réduction de la dimensionnalité ou encore sur les graphes des plus proches voisins.

Une approche pionnière fut proposée par le laboratoire de John Marioni. Elle consiste à trouver des voisins les plus proches partagés (MNN en anglais pour *Mutual Nearest Neighbors*) (Haghverdi et al., 2018). Pour se faire, on calcule pour chaque cellule ses plus proches voisins parmi les cellules des autres jeux de données. Si deux cellules sont chacune voisines l'une de l'autre de cette manière, elles forment alors des voisins les plus proches partagés. Les différences observées entre ces cellules sont considérées comme étant purement dues aux effets de lots et sont compensées afin d'obtenir les données corrigées. Cette méthode a ensuite été adaptée pour être utilisée sur les données après application de la PCA (A. Lun, 2019a), d'une analyse canonique des corrélations (Stuart et al., 2019), ou encore de l'UMAP (Polański et al., 2020). D'autres méthodes présentant des résultats équivalents (Tran et al., 2020) furent également développées en utilisant la factorisation en matrices non-négatives intégrative (Welch et al., 2019) ou une correction itérative des données issues de la PCA par clustering (Korsunsky et al., 2019).

L'essor des méthodes basées sur l'apprentissage profond a permis le développement d'algorithmes plus performants comme *scVI* (Lopez et al., 2018), dont l'efficacité a été mise en exergue lors d'une analyse comparative multi-échelle réalisée par le laboratoire de Fabian Theis (Luecken et al., 2022). *ScVI*, et son extension *scANVI* (C. Xu et al., 2021), utilisent des réseaux de neurones afin d'estimer les paramètres des modèles d'expression des gènes et peuvent ainsi découpler les signaux biologiques des variations inter-échantillons.

Trajectoires de différenciation

La majeure partie des processus développementaux étant des composés de transitions continues d'un état à l'autre, des méthodes ont été développées afin de modéliser les trajectoires de différenciation et d'étudier l'expression des gènes le long de ces dernières. Inspirées par le succès de l'algorithme Monocle (Trapnell et al., 2014), les premières méthodes développées se sont basées sur l'établissement de trajectoires minimisant les variations transcriptionnelles entre cellules voisines par le calcul d'arbres couvrants minimaux calculés à partir de diverses méthodes de réduction de la dimensionnalité. Une comparaison exhaustive des différentes méthodes, publiée en 2019 (Saelens et al., 2019), a montré qu'il n'existait pas de méthode optimale pour le calcul des trajectoires puisque les performances des algorithmes dépendent du type de trajectoire

que l'on cherche à modéliser (différenciation linéaire, branchée comme dans l'hématopoïèse, ou même cyclique). Deux méthodes semblent néanmoins tirer le plus souvent leur épingle du jeu : *Slingshot* (Street et al., 2018) et *PAGA* (Wolf et al., 2019). Si *Slingshot* est également basé sur l'utilisation d'arbres couvrants minimaux, la méthode *PAGA* a une approche plus originale puisqu'elle produit un graphe abstrait qui peut être vu comme un arbre des plus proches voisins représenté à l'échelle des clusters. Chaque nœud du graphe de *PAGA* est un cluster du jeu de données, et ses derniers sont reliés par une arête si et seulement si une fraction suffisante des cellules de chaque cluster sont voisines l'une de l'autre dans le graphe des plus proches voisins.

Un des défauts des approches mentionnées ci-dessus consiste en la nécessité pour l'utilisateur de fournir le point de départ des trajectoires pour orienter les trajectoires et ordonner les cellules, ce qui nécessite une connaissance préalable du système étudié. Une approche permettant de résoudre ce problème est l'utilisation de la vélocité des ARNs. Cette technique se base sur les ratios d'ARNs épissés et non-épissés pour savoir si un gène est en train d'être induit ou réprimé dans les cellules. Ainsi, il est possible d'estimer la dynamique d'évolution des cellules et de former des trajectoires orientées sans intervention de l'utilisateur. Initialement publiée en 2018 (la Manno et al., 2018), cette méthode a ensuite été améliorée en 2020 (Bergen et al., 2020) et a fondé beaucoup d'espoirs sur son potentiel à fournir des trajectoires fiables pour des systèmes où l'orientation de ces dernières n'étaient pas encore connue. Malheureusement, il a depuis été montré que les modèles mathématiques utilisés pour l'estimation de la vélocité sont mis en défaut dans un certain nombre de systèmes biologiques dont les trajectoires sont connues (Bergen et al., 2021). C'est notamment le cas de l'hématopoïèse embryonnaire (Barile et al., 2021), au cours de laquelle certains gènes voient leur transcription augmenter brutalement lors de la différenciation érythroïde, ce qui brouille les estimations de trajectoires.

Apport du scRNAseq à la compréhension de la formation des cellules souches hématopoïétiques

L'essor du scRNAseq a rapidement entraîné son utilisation dans le but de comprendre l'ontogenèse du système hématopoïétique. Il a ainsi été possible de réaliser en 2019 un atlas du développement embryonnaire de la souris de la gastrulation au début de l'organogenèse (Pijuan-Sala et al., 2019). En utilisant des embryons chimériques, les auteurs ont ainsi pu étudier le rôle du gène *Tal1* dans le développement de l'hématopoïèse murine et sa contribution aux différents lignages de l'embryon. Malheureusement, de telles approches sont impossibles dans l'embryon

humain et le faible nombre d'échantillons disponibles limite drastiquement la résolution que nous avons actuellement sur le déroulé des différentes vagues de l'hématopoïèse humaine.

Le premier jeu de données d'aorte embryonnaire a été publié en 2019 et s'est intéressé à l'apparition des premières cellules souches hématopoïétiques dans l'embryon (Zeng et al., 2019). En utilisant des jeux de données provenant d'embryons de CS10 à CS14, les auteurs ont pu caractériser deux types d'endothéliums hémogéniques : un endothélium hémogénique présentant un phénotype artériel présent à partir du CS12, et un endothélium hémogénique plus précoce, ne présentant pas de phénotype artériel, présent au CS10 (Figure 7). En 2020, un second article s'est intéressé aux populations endothéliales et hématopoïétiques de l'aorte fœtale au moment de la transition endothélio-hématopoïétique et a réalisé une analyse scRNAseq de l'AGM d'un embryon au CS16 (Crosse et al., 2020). Ils ont ainsi pu identifier deux populations endothéliales : une présentant un profil aortique avec l'expression de gène GJA5 et une présentant un profil veineux avec l'expression du marqueur APLNR. Les auteurs n'ont pas réussi à identifier le phénotype exact de l'endothélium hémogénique à CS16 à partir de ce jeu de données, même si les analyses réalisées avec *PAGA* semblent indiquer un phénotype veineux.

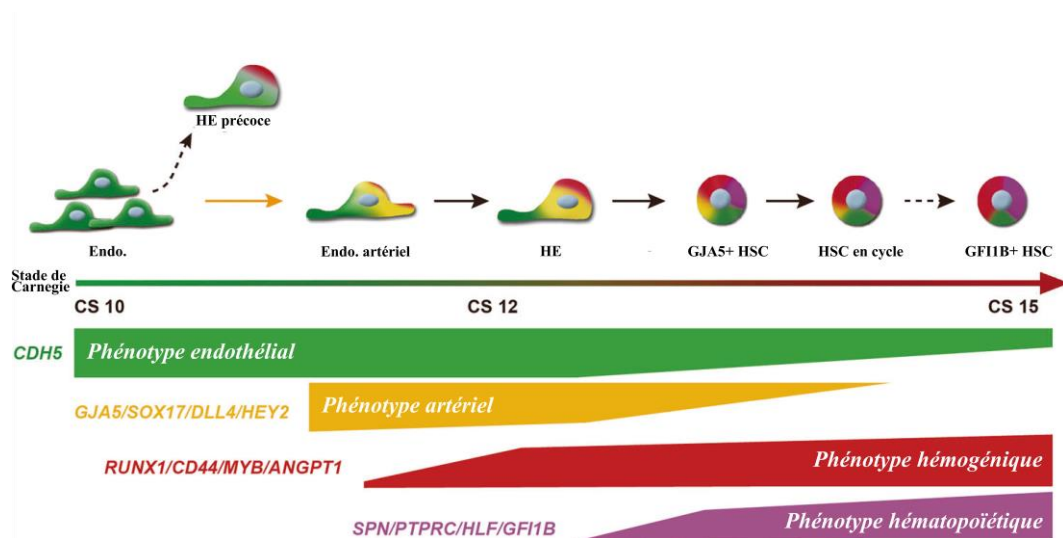


Figure 7 : Conclusions des analyses scRNAseq de l'endothélium hémogénique entre CS10 et CS15 par Zeng et al. HE : endothélium hémogénique, Endo. : endothélium, HSC : cellule souche hématopoïétique Adapté de Zeng et al. (Zeng et al., 2019)

Dans le même temps, deux autres jeux de données apportant un éclairage sur le développement hématopoïétique humain ont été publiés. Le premier est une analyse d'un embryon entier à CS7 (Tyser et al., 2021) et le second est une analyse des cellules hématopoïétiques du foie fœtal réalisées entre les semaines 7 et 17 post-conception (Popescu et al., 2019). Ces deux jeux de

données fournissent des points de comparaison essentiels afin de pouvoir comparer l'évolution transcriptomique des cellules souches hématopoïétiques et des progéniteurs multipotents au cours du développement. Il est à noter que l'hématopoïèse définitive identifiée au CS7 semble suggérer que l'hématopoïèse se développe plus rapidement chez l'humain que chez la souris relativement au rythme de développement du reste de l'embryon.

Ces jeux de données aortiques et hépatiques ont ensuite été intégrés avec de nouveaux jeux de données réalisés au CS14-15 dans une étude publiée en 2022 (Calvanese et al., 2022). Cette dernière a permis de définir une signature de six gènes (RUNX1, HOXA9, MLLT3, MECOM, HLF et SPINK2) qui permet d'identifier les cellules souches hématopoïétiques embryonnaires. De plus, les auteurs ont pu établir des listes de gènes permettant de caractériser le profil transcriptomique de l'endothélium hémogénique et des cellules souches hématopoïétiques aux différents stades du développement de l'embryon. Un schéma récapitulatif de leur analyse est présenté en Figure 8.

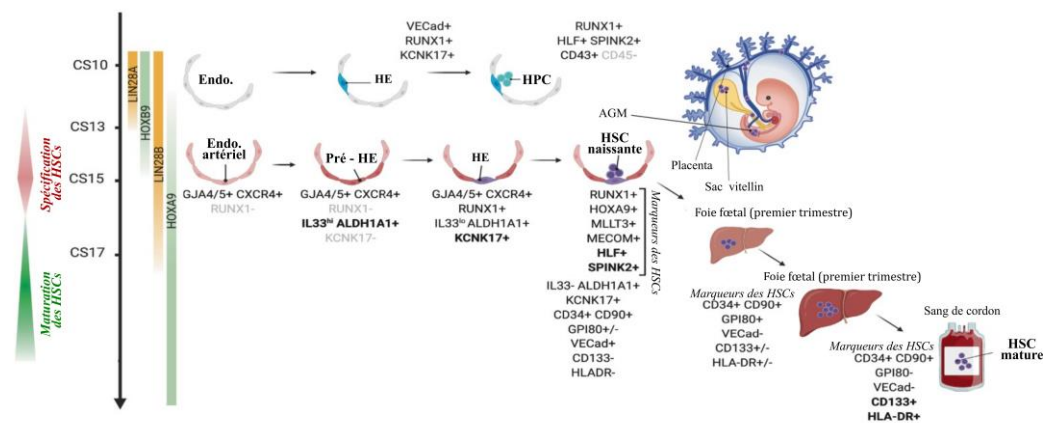


Figure 8 : Schéma des étapes principales de la formation et de la maturation des cellules souches hématopoïétiques au cours du développement de l'embryon

HSC : cellule souche hématopoïétique, HPC : cellule progénitrice hématopoïétique AGM : aorte-gonades-mésonephros, Endo. : endothélium, HE : endothélium hémogénique, Pré-HE : endothélium pré-hémogénique
Adapté de Calvanese et al. (Calvanese et al., 2022)

Résultats

Différenciation de cellules souches pluripotentes induites humaines en cellules souches hématopoïétiques

Manuscrit soumis pour publication : *Generation of transgene-free hematopoietic stem cells from human induced pluripotent stem cells*

Ce manuscrit présente un protocole de différenciation de cellules souches pluripotentes induites en cellules souches hématopoïétiques sans utilisation de transgènes ou de cocultures avec des cellules stromales. Conçu dans l'optique d'une mise à l'échelle dans le but d'applications cliniques, il est composé d'une unique étape de différenciation en corps embryoides de 17 jours à l'aide de cytokines et de facteurs de croissances présents tout au long de la culture. Les cellules souches hématopoïétiques produites par protocole présentent un profil transcriptomique similaire à celui des cellules souches hématopoïétiques produites dans l'aorte embryonnaire et sont entièrement fonctionnelles puisqu'elles sont en capacité de réaliser une reconstitution multi-lignage lors de greffes sériées dans des souris immunodéficientes irradiées.

Ma contribution à ce papier réside essentiellement dans les analyses bioinformatiques ayant permis de caractériser les cellules produites en culture et les cellules humaines présentes dans les moelles des souris greffées. J'ai également réalisé les expériences d'immunohistochimie qui ont permis d'étudier la distribution spatiale des différents types cellulaires dans les corps embryoides.

Title: Generation of transgene-free hematopoietic stem cells from human induced pluripotent stem cells.

Olivier Piau^{1,2,4§}, Mathias Brunet-Manquat^{1,2,4§}, Bruno L'Homme⁵, Laurence Petit³, Brigitte Birebent⁴, Christine Linard⁵, Laetitia Moeckes⁶, Thomas Zuliani⁶, H  l  ne Lapillonne^{1,2,7}, Marc Benderitter⁵, Luc Douay⁷, Alain Chapel^{1,2,5§}, Laurence Guyonneau-Harmand^{1,2,4§*} and Thierry Jaffredo^{3§*}

¹ Sorbonne Universit  , UMR_S938 CDR Saint-Antoine, Prolif  ration et Diff  renciation des Cellules Souches, F-75012 Paris, France.

² INSERM, UMR_S938, Prolif  ration et Diff  renciation des Cellules Souches, F-75012 Paris, France.

³ Sorbonne Universit  , CNRS, Inserm U1156, Institut de Biologie Paris Seine, Laboratoire de Biologie du D  veloppement/UMR7622, 9 Quai St-Bernard, 75005 Paris, France.

⁴ EFS Ile de France, Unit   d'Ing  nierie et de Th  rapie Cellulaire, Cr  teil, F-94017, France.

⁵ Laboratoire de radiobiologie des expositions m  dicales (LRMed), Institut de Radioprotection et de S  ret   Nucl  aire (IRSN), F-92262 Fontenay-aux-Roses, France.

⁶ Etablissement Fran  ais du Sang – Atlantic Bio GMP – 2, rue Aronnax 44800 Saint-Herblain, France.

⁷ AP-HP, H  pital St Antoine/Trousseau, Service d'H  matologie Biologique, F- 75012 Paris, France.

*Correspondence to: laurence.harmand@upmc.fr; thierry.jaffredo@upmc.fr

§Co-first authors

  Co-senior authors

SUMMARY: 147 words

Hematopoietic stem cells (HSCs) are the rare cells responsible for the lifelong curative effects of hematopoietic cell transplantation. The demand for clinical-grade HSCs has increased significantly in recent decades, leading to major difficulties in treating patients. A promising but not yet achieved goal is the generation of HSCs from pluripotent stem cells. Here, we have obtained vector- and stroma-free transplantable HSCs by differentiating human induced pluripotent stem cells (hiPSCs) using an original one-step culture system. After injection into immunocompromised mice, cells derived from hiPSCs settle in the bone marrow and form a robust multilineage, hematopoietic population that can be serially transplanted. Single cell RNA sequencing shows that this repopulating activity is due to a hematopoietic population that is transcriptionally similar to human embryonic aorta-derived HSCs. Overall, our results demonstrate the generation of HSCs from hiPSC and will help identify key regulators of HSC production during human ontogeny.

INTRODUCTION

Despite intensive efforts over the last decades, vector-free, transplantable human hematopoietic stem cells (HSCs) have not yet been successfully generated. In contrast, many blood cell types useful for human therapies have been efficiently generated from human embryonic stem cells (hESCs) or induced pluripotent stem cells (hiPSCs), including erythroid cells¹⁻⁴, B^{5,6} and T cells⁷⁻¹⁰, myeloid cells¹¹, megakaryocytes^{12,13}, or hematopoietic progenitor cells¹⁴. In recent years, many studies have aimed to derive from pluripotent stem cells a population of CD34+CD45+ cells with a phenotype approaching adult HSCs¹⁵⁻¹⁹, the cells used in clinical transplantation. These include reprogramming of adult cells by forced expression of transcription factors in fibroblasts²⁰ or endothelial cells²¹. However, these CD34+CD45+ populations showed mainly poor transplantation potential in immunocompromised mice, undermining therapeutic use. The only reports of successful generation of bona fide HSCs were obtained by redirecting hiPSC-derived or primary human endothelial cells with transcription factors²²⁻²⁴. However, the introduction of plasmids encoding oncogenes or the use of feeder cells precludes the use of these reprogrammed cells for clinical applications.

Interestingly, transplantable human HSCs can be obtained in teratomas, albeit with low frequency²⁵⁻²⁷. While not directly applicable to human therapy, this demonstrates that generation of HSCs from human pluripotent stem cells is possible and that cell-cell interactions play a key role in promoting self-renewal and multilineage repopulation activity²⁶. Teratomas recapitulate some aspects of embryonic development. Although this has not been specifically investigated, the presence of HSCs argues for mechanisms reminiscent of early hematopoietic development. Indeed, the first transplantable HSCs arise from a specialized population of endothelial cells, hemogenic endothelial cells, formed early during embryonic development in the aorta-gonad-mesonephros (AGM) region^{28,29}, through an endothelial-hematopoietic transition (EHT)^{30,31}. Some of these steps are likely recapitulated in embryoid body (EB) cultures that mimic events during hematopoietic development. In particular, the formation of the hemogenic endothelium and budding of hematopoietic cells are well documented³²⁻³⁴.

Here, we describe a vector-free and stroma-free system for generating from hiPSCs human HSCs capable of robust, long-term multipotent reconstitution and self-renewal *in vivo*. This protocol uses EB formation to generate a cell population resembling nascent HSCs found in the human embryonic AGM. Overall, our results demonstrate that the generation of feeder- and vector-free human HSCs is now feasible and is achieved by producing cells that more closely resemble the original HSCs that emerge during ontogeny.

RESULTS

Establishment of the culture conditions

The optimal culture conditions to generate HSCs were established using a Design of Experiment approach (DOE). DOE is a statistical method for improving process performance by determining the optimal combination of input factors to optimize the output parameters, using the concept of desirability³⁵. We first performed a literature review on *in vitro* HSC differentiation, amplification and maintenance (Figure 1, Table S1). Based on this, we selected an ensemble of 10 cytokines and growth factors that have been reported to be important at different stages of mesoderm and hematopoietic specification³⁶⁻⁴¹ that we applied to two different hiPSC lines, FD136-25 and IMR90-16. These cytokines and growth factors are (i) bone morphogenetic protein 4 (BMP-4), vascular endothelial growth factor (VEGF), and insulin-like growth factor-1 (IGF-1) for mesoderm commitment and (ii) FMS-like tyrosine kinase ligand (FLT3-L), stem cell factor (SCF), interleukin-1 (IL1), interleukin-3 (IL3), interleukin-6 (IL6), granulocyte colony-stimulating factor (G-CSF), and thrombopoietin (TPO) for hematopoietic commitment. The time frame of the protocol was set to 17 days (D17), a growth time used in Lengerke and colleagues⁴⁰. In this case, we defined a concentration range for each of the cytokines as an input variable (Table S1). As output variables, we chose cell amplification, the frequency of long-term culture-initiating cells (LTC-ICs) and the expression of several important surface markers associated with populations amenable to long-term transplantation, i.e., CD34, CD38, CD71, CD109, HLA-DR, CD90, CD33, CD13, CD117 (Table S2). The software defined sixteen different combinations (Table S1), which were tested by measuring baseline variables (Table S3). These results were then subjected to regression analysis, and the response was plotted. The predictive model generated by the algorithm identified three efficient combinations of cytokines. These three combinations were then used to differentiate two hiPSC lines for a culture period of 17 days. Each of them triggered cell proliferation, low hematopoietic (CD34⁺CD45⁺ and/or CD117⁺CD90⁺) commitment, LTC-IC formation and graft in NOD-SCID primary recipients (Figure 1 and Table S4). Flow cytometric analysis of cell populations grown under the three conditions was performed using a panel of endothelial and hematopoietic-specific markers (Table S5). From this analysis, we found that Combination A showed the best ratio of immature (CD34⁺, CD43⁺, and CD117⁺) to mature hematopoietic markers (CD45⁺); it also had the highest rate of human CD45⁺ cells in NOD

/SCID-LtSz-scid/scid (NOD /SCID) mice. Therefore, combination A was selected for further use (Figure 1). While in standard protocols for hiPSC differentiation CD34⁺CD45⁺ hematopoietic cells appeared to burst from embryoid bodies (EBs) at D10 to D14^{18,22,40,42}, EBs in combination A remained compact and spherical at D17 without bursting, indicating a delay in differentiation (Figure 2A). These culture conditions were reproducibly applied with success to four additional individual hiPSC lines obtained with different reprogramming protocols (Figure 2A; Table S6, Material and Methods), demonstrating the robustness of the method. The results describing the individual capacities of the five hiPSC lines to graft in NSG mice are displayed in Figure S1A.

Long-term, multilineage repopulating activity and phenotypic analysis of the hematopoietic populations.

The hematopoietic repopulation potential of D17 EB cells was tested by transplanting retro-orbitally 4.10^5 dissociated cells for 20 weeks into sub lethally irradiated (2.4-3.5 Gy) 8-week-old NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) mice (Figure 2A). The efficiency of reconstitution was systematically compared with CD34⁺ cord blood (CB) cells. The presence of human hematopoietic cells in the mouse bone marrow (BM) was monitored using flow cytometry, functional hematopoietic assays (T-cell activation and hemoglobin switch), cytological assays, RT-qPCR, and single-cell RNA sequencing (scRNA-Seq) approaches (Table S6). Primary NSG recipients (Figure 2C) transplanted with D17 cells derived from one of the five hiPSC lines (Figure S1A and Table S6) showed human hematopoietic cell reconstitution (59/60), quantified by surface expression of human (h) CD45 (Figure 2B, C), hCD34, hCD43 in the mouse BMs. The hiPSC lines tested showed an overall similar reconstitution efficiency with $13.3 \pm 1.5\%$ hCD45⁺, $7.2 \pm 1.0\%$ CD43⁺, and $4.1 \pm 0.6\%$ CD34⁺ cells in total mouse BM mononucleated cells. This is more than 130-fold the threshold of 0.1% of hCD45⁺ that is usually considered positive for human hematopoietic transplantation in NSG mice⁴³. To test the robustness of the protocol, Pci-CAU hiPSCs were differentiated by an independent laboratory (Atlantic Bio GMP Platform, Nantes; France), inoculated into irradiated NSG mice, and observed over an extended period of time (Figure S1B-C). These cells showed robust engraftment and multilineage differentiation, similar to the cells produced in the original laboratory (Figure S1D). Overall, these numbers were comparable to those of CD34⁺ CB cells

transplanted under identical conditions (Figure S1E), confirming the ability of our cells to repopulate the BM and the reproducibility of the protocol.

The mouse BMs harbor robust human myeloid, B- and T-lymphoid cells (hCD33⁺, hCD19⁺ and hCD3⁺) and hCD235a⁺ erythroid progenitors/precursors 20 weeks after transplantation (Figure 2D and S2A) ⁴⁴⁻⁴⁶. Figure S2B-F shows a representative cytometric analysis of BMs isolated from NSG primary recipients injected with saline or transplanted with D17 EB cells or CD34⁺ CB cells. The D17 EB cells (Figure S2C-D) show a similar multilineage pattern as the CD34⁺ CB cells (Figure S2E, F, G). Peripheralization of the transplanted cells was confirmed by the presence of hCD45⁺ cells in the blood, which showed a multilineage pattern (Figure S3A, B) which resembled the one observed in the peripheral blood of CD34⁺ CB transplanted recipients (Figure S3C). In the thymus, hCD3⁺ cells showed TCR $\alpha\beta$ expression greater than 60% and TCR $\gamma\delta$ expression less than 30% (Figure S3D), a pattern consistent with CD34⁺ CB cells (Figure S3E) and indicative of the presence of mature T cells ⁴⁷. Similarly, hCD45⁺ cells in the spleen were characterized by T cell expression of hTCR β and B lymphopoiesis by expression of hCD19 (Figure S3F), consistent with CD34⁺ CB grafts (Figure S3G).

In order to get an unbiased view of the human cells present in the mice BM after the graft, we performed 10X Genomics scRNA-Seq on the BM of two representative primary mice 20 weeks after transplantation. As shown in Figure 2E and S4A-D, the mouse BM contains human erythroid, B, T, plasmacytoid dendritic, monocytic, and MAST cells that were present in both recipients (the strategy to distinguish human from mouse cells is extensively described in the material and method section). Comparison with FACS analysis showed that the proportion of the different cell types was consistent between the two individuals using both techniques (Figures 2F, G and S4B-G). Human erythroid cells expressed fetal and adult hemoglobin subunits, indicating mature erythropoiesis (Figure S4H).

To substantiate the functionality of the transplanted cells, we performed a human-selective clonogenic assay using BM cells isolated from primary recipients. We obtained an overall frequency of 17.5 ± 4.3 colonies from a total of 10^4 BM cells (Figure 3A), which were divided into CFU-GEMM, BFU-E, and CFU-GM colonies (Figure 3B). Cytological analysis of colonies revealed the presence of mature macrophages (Figure 3C1), histiomonocytes (Figure 3C2), and erythroblasts (Figure 3C3). We also analyzed the ability of human erythroid

progenitor cells isolated from mouse BM to undergo hemoglobin switching *in vivo* (n=3, 30 recipients). RT-qPCR analysis of total BM cells and erythroid colonies *in vitro* revealed high levels of human β - (39.51 \pm 4.95% and 28.91 \pm 2.43%, respectively) and γ - (57.49 \pm 3.95% and 68.70 \pm 5.55%, respectively) globin gene expression, whereas ϵ -globin was dramatically reduced to 3.0 \pm 1.2% and 2.39 \pm 1.59% of total globin content, respectively (Figure 3D), consistent with the scRNA-seq analysis (Figure S4H). This is a hallmark of definitive erythropoiesis in the primary recipients and comparable to the globin contents obtained during erythroid differentiation of CD34+ CB cells (Figure 3D) except for a small amount of epsilon globin, which is present in colonies derived from BM cells. In addition, we analyzed the globin content of several BFU-E from primary recipients (n=3, 15 recipients) and found different patterns of β -, γ -, and ϵ -globins from clone to clone (Figure S5).

Regarding the T lineage, peripheral blood hCD3+ cells displayed high TCR $\alpha\beta$ levels (Figure 3E) and low TCR $\gamma\delta$ levels, similar to thymic hCD3+ cells (Figure S4B), indicating normal T lymphopoiesis. To test the *ex vivo* expansion potential of thymic hCD3 T cells, cells were stimulated by hCD3 and hCD28, and CFSE-labeled (n=3, 30 recipients). Cell numbers increased over the course of 5 days (Figure 3F) thereby demonstrating T cells functionality. B and T cells were also confirmed by q-PCR of CD19 (B lymphocyte surface antigen B4) and TCF7 (transcription factor 7), respectively, in BM, thymus, and spleen (Figure S6A). A hematopoietic progenitor cell assay revealed the presence of granulocytes and erythroblasts (Figure S6B). A TaqMan[®] Array Human Hematopoiesis RT-qPCR disclosed the presence of the canonical blood cell markers in the CD45+ hematopoietic fraction (Figure S6C and Table S4). To estimate the number of functionally engrafting human HSCs, we performed a limiting dilution assay in primary NSG recipients and estimated that 1 D17 cell out of 15700 is a fully SCID-repopulating cell (Figure 3G, Table S7-S8).

To confirm the presence of human self-renewing HSCs, secondary transplants (n=40, Table 6) were performed (Figure 2A). For each primary recipient, one secondary recipient was engrafted with half of the flushed BM. Twenty weeks after transplantation, hCD45+ cells accounted for 12.2 \pm 0.8% of mononucleated BM cells, 8.1 \pm 0.6% of CD43+, and 4.8 \pm 0.6% of CD34+ cells, indicating a sustained reconstitution capacity (Figure 2H). In 40 of 40 secondary mice receiving BM cells (Figure S7A, S8A-D), hematopoietic reconstitution was comparable to the pattern obtained with secondary CD34+ CB cell transplantation (Figure 2I, S7B, S8E-F). The

cloning efficiency of the human colony-forming cells was $5.5 \pm 3.1\%$ in 10^4 total mouse BM cells, indicating a robust and long-lasting self-renewal potential (Figure 3A, right). The cells generated human erythroid progenitors that had high levels of β - (36.61 \pm 5.86%) and γ -globin (61.39 \pm 4.86%), whereas ϵ -globin was drastically reduced to 2.1 \pm 1.1% of total globin (Figure 3D), as analyzed by RT-qPCR, a hallmark of definitive erythropoiesis.

A successful three-round serial-transplant assay was performed in NOD-SCID mice. All three tertiary-recipient mice exhibited multilineage engraftment at 12 weeks (Figure S9 and Table 6). This proves that D17-EB-derived human HSCs display robust multilineage and self-renewal potential *in vivo*.

Functional, molecular and phenotypic identification of the EB cells.

To gain further insight into our differentiation protocol, we transcriptionally profiled EBs at 4 different time points, ie, D13, D15, D17, and D19 of culture, using 10X Genomics scRNA-Seq (n=2 biological replicates per time point for a total of 26027 cells). UMAP visualization of the merged datasets identified of 27 different cell types distributed among three main UMAP branches corresponding to the three germ layers, i.e., the ectoderm, mesoderm, and endoderm, and a fourth branch containing cells of the trophoblast lineage (Figure 4A and Figure S10). To better characterize the events leading to HSC development, we focused on the endothelial and hematopoietic cell clusters (Figure 4A, top left). As a first approach, we compared this cell subset with a published scRNA-Seq dataset from the human aorta gonad mesonephros region at Carnegie stage (CS) 16, at the time of HSC development⁴⁸, (Figure 4B). Strikingly, both data sets show two types of vascular endothelial cells, an arterial-like CD34⁺, CDH5⁺, HEY2⁺, and GJA5⁺ cluster and a venous-like CD34⁺, CDH5⁺, NR2F2⁺, and APLNR⁺ cluster (Figure 4C, D), suggesting that EB culture produces similar populations to the one present in human embryo aorta, when HSCs arise from the hemogenic endothelium.

To further explore this hypothesis, we examined the pattern of expression of MYB and RUNX1 by RT-qPCR and compared it with the profile obtained by scRNA-Seq (Figure 5A) from day 13 to 17. An increase in their expression after day 15 was observed by both methods. Flow cytometric analysis of several endothelial markers (CD34, CDH5) showed a decrease in endothelial cells after day 13 and the gradual appearance of the PTPRC⁺ population between D15 and 17 (Figure 5B). We also studied the endothelial and hematopoietic capacity of EB cells

at D15, 16, and 17 by specific assays *in vitro* (Figure S11). D15 EB cells showed strong endothelial potential, as evidenced by the formation of pseudo microtubules (Figure S11A)⁴⁹, but lacked hematopoietic capacity, as they were unable to form clonogenic colonies and had a very low frequency of LTC-ICs (Figure S11B-D). In contrast, D17 cells lost their endothelial potential but increased their higher hematopoietic potential (Figure S11A-D). This data suggest D17-EB derived cells have the highest HSC potential, corresponding to the onset of hematopoiesis. Immunostaining of sections of D17 EBs showed that endothelial CDH5⁺CD31⁺ cells and NG2⁺ pericytes are distributed throughout the periphery of the EBs and are absent in the center (Figure 5D, E, F, G). The interior of this endothelium-rich outer region is bordered by a ring of SMA⁺ cells (Figure 5G). Interestingly, clusters of RUNX1⁺ endothelial cells and hematopoietic cells were also identified (Figure 5F). This suggests that the periphery of EB contains a dense vascular network in which there is likely a transition from endothelial cells to hematopoietic cells, confirming the scRNA-Seq results.

Finally, we decided to identify the subpopulation of HSCs produced. We projected the recently published six-gene signature and HSC score⁵⁰, which distinguishes human embryonic HSCs from hematopoietic progenitor cells, onto the scRNA-Seq hematopoietic clusters (Figure 4F). We identified several RUNX1⁺ HOXA9⁺ MLLT3⁺ MECOM⁺ HLF⁺ SPINK2⁺ cells clustered into a candidate HSC subpopulation (Figure 4E). Refined comparisons with additional markers of human HSC ontogeny suggest that these candidate HSCs exhibit an intermediate phenotype between CS10 early AGM HSCs⁵¹ and CS14-15 HSCs⁵⁰, with some rather early markers such as LIN28, GAD1, and FGF23 maintained alongside more mature markers such as HOXA9, PTPRC, and STAT5A (Table S8). Overall, EBs appear to recapitulate AGM formation and give rise to HSCs with an early to intermediate AGM phenotype.

DISCUSSION

We successfully produced HSCs capable of long-term transplantation and self-renewal from hiPSCs using a novel combination of cytokines and growth factors, defined by DOE. This was achieved by a one-step, GMP-compliant procedure without the use of reprogramming vectors.

We used DOE to finely tune a combination of cytokines and growth factor concentrations that enabled the formation of HSCs. The protocol is extremely robust, as HSCs with similar reconstitution potential, even approaching the reconstitution potential of CD34+ CB cells, could be produced from five different hiPSC lines, regardless of the method by which they were reprogrammed. A scaling-up of the protocol was successfully assessed since GMP-compliant batches of D17 cells have now been produced with equal efficiency in an advanced therapy medicinal product platform (Atlantic Bio GMP Platform, Nantes; France). In conclusion, both the robustness and the reproducibility of this protocol meet clinical purpose requirements.

According to scRNA-Seq analysis, we generate about 4.6% hematopoietic cells at day 17 and around 1 in 10 000 EB cells carries the six-gene signature of embryonic HSCs⁵⁰. This is fully consistent with our limiting dilution experiments showing that 1 in 15 700 EB cells is an HSC. This highly efficient engraftment can be compared with experiments performed on human AGM HSCs⁵². The latter showed that although each AGM contains only 1 or 2 HSCs, these cells exhibit a high repopulation capacity. Similarly, our HSCs show a very robust reconstitution potential in multiple recipients with the full spectrum of hematopoietic lineages in BM, circulating blood, thymus, and spleen. In addition, no signs of leukemia or teratomas were observed in any of the transplanted mice.

The main issue with *de novo* generation of HSCs is clonal reconstitution and the presence of multipotent progenitor cells instead of true HSCs. The latter is of particular concern because NSG mice, which are known to exhibit a B lymphoid bias⁵³⁻⁵⁵. Our primary D17 recipients show variable grafting patterns, both in terms of hCD45 expression and composition of the different hematopoietic cell lineages, and the secondary recipients show grafting patterns different than their primary counterparts. This argues against clonal reconstitution. In addition, individual BFU-E from hematopoietic colony-forming cell assays

performed within a single transplanted mice BM showed different ratios of β -, γ -, and ϵ -globins. This suggests that erythropoiesis is maintained by multiple progenitor cells, which in turn implies that the graft is not maintained by a clonal mechanism.

ScRNA-Seq analysis of the D17 EBS reveals 2 major endothelial clusters: an arterial-like GJA5⁺ cluster and a venous-like APLNR⁺ cluster. These profiles are essentially the same as those previously described for the human CS16 embryo⁴⁸. Interestingly, Calvanese and colleagues⁵⁰ describe the hemogenic endothelium as GJA5⁺ in the aorta of the human CS14 embryo. In contrast, some cells of our APLNR⁺ endothelial cluster colocalize with the hematopoietic cells on the UMAP, suggesting that the HSCs obtained by our protocol are likely derived from the APLNR⁺ hemogenic endothelium. These results are consistent with other findings describing a possible switch in the identity of the hemogenic endothelium during development from a venous to an arterial phenotype between CS10 and CS13⁵¹. Since EB-derived HSCs identified thanks to the six-gene embryonic HSC signature⁵⁰ exhibit an intermediate phenotype between CS10 and CS14, it seems highly plausible that the hemogenic endothelium from which they arise also display an earlier phenotype than CS14. Overall, our protocol likely recapitulates the emergence of HSCs from the AGM between CS10 and CS14 of the human embryo.

Thus, the plasticity of the hemogenic endothelium phenotype argues for an important role of the microenvironment in HSC production. This is reminiscent of the protocols that have succeeded in generating functional HSCs from teratomas^{25; 26}. These protocols have demonstrated that a variety of cell types from all germ layers can provide a rich microenvironment²⁷ for the generation of HSCs suitable for long-term transplantation. Indeed, HSC formation in the AGM is driven by signals from a variety of sources, including non-mesodermal tissues such as the sympathetic nervous system⁵⁶. The diversity of communications between cell types in teratomas likely reflects these signals, which are sent from different tissues during development and are critical for HSC to be produced. Similarly, EBs contain cells from the ectoderm, endoderm, and mesoderm. We hypothesize that this ensemble of cell types forms a rich, complex microenvironment that is critical for the formation of transplantable HSCs. This may explain why our protocol succeeds in generating HSCs with long-term transplantability, whereas most protocols using EBs only report cells with short-term transplantability.

In summary, we report the successful and robust generation of multipotent long-term HSCs from hiPSCs that support hematopoietic reconstitution and self-renewal *in vivo* with a defined one-step protocol. Such a transgene- and feeder-free tool could pave the way for promising future clinical applications and the discovery of new therapeutic targets by providing a new model for deciphering the emergence of hereditary blood malignancies. Our system is also a real step towards a better understanding of developing hematopoiesis in the early human embryo, as it can overcome the ethical constraints associated with human embryo research.

EXPERIMENTAL PROCEDURES

hiPSC amplification

The study was conducted using six different hiPSC lines: the FD136-25 (skin primary fibroblasts) reprogrammed with lentiviral vectors and Thomson's combination (endogenous expression of Oct4, Sox2, Nanog and Lin28), IMR90-16 (fetal lung fibroblasts) reprogrammed with lentiviral vectors and Thomson's combination (endogenous expression of Oct4, Sox2, Nanog and Lin28) ², the Pci-1429 and Pci-CAU lines (peripheral blood mononuclear cells-Phenocell) reprogrammed with episomes (Sox2, Oct4, KLF, cMyc); the LAM001-005 and LAM002-002 lines (skin primary fibroblasts), reprogrammed with sendai virus (Oct3/4, Nanog, KLF and cMyc) by IPSCDTC core facility, Nantes, France. hiPSCs were maintained on matrigel (Corning) in TESRplus medium (Stemcell Technologies) and the cells were passaged 1:6 onto freshly coated plates every 3 to 5 days using standard clump passaging with TRYple select (Invitrogen) or ReLeSR (Stemcell Technologies).

ABG ATMP platform worked with Pci-CAU line and used vitronectin matrix (Gibco™), IPSbrew medium (Miltenyi Biotech) and ReLeSR (Stemcell Technologies) for passaging.

EB differentiation

EBs were cultured in liquid medium IMDM supplemented with stabilized glutamine (SAFC), 330 µg/mL holo-human transferrin (Scipac), 10 µg/mL recombinant human insulin (Incelligent SG; CellGen), 2 IU/mL heparin Choay, and 5% solvent/detergent virus-inactivated AB plasma (Etablissement Français du Sang). This plasma is used in a clinical trial registered at <http://www.clinicaltrials.gov> as NCT0929266.

The medium was supplemented with stem cell factor (24 ng/mL), thrombopoietin (21 ng/mL), FLT3 ligand (21 ng/mL), recombinant human bone morphogenetic protein 4 (194 ng/mL), recombinant human vascular endothelial growth factor (200 ng/mL), interleukin-3 (50 ng/mL), interleukin-6 (50 ng/mL), interleukin-1 (5 ng/mL), granulocyte colony-stimulating factor (100 ng/mL) and insulin growth factor (1 ng/mL) (GMP grade when available otherwise Premium grade, Peprotech) . The medium was replenished every other day until 19 days at 37°C in a humidified 5% CO₂ atmosphere. EBs were dissociated in 0.1% collagenase B (Roche Diagnostics, Laval, QC, Canada) for 1 hour at 37°C then 0.25% trypsin for 3 min and disrupted by gentle mechanical agitation

Colony assays

At the indicated times, 10⁵ dissociated EBs or 3.10⁴ cells from xenotransplanted recipient BM were plated into 3 mL of complete methylcellulose medium in the presence of SCF, IL-3, EPO and GM-CSF (PeproTech, Neuilly-sur-Seine, France). As G-CSF also stimulates mouse progenitors, it was replaced by granulocyte-macrophage colony-stimulating factor (GM-CSF). Aliquots (1 mL) of the mix were distributed into one 30 mm dish twice and maintained in a humidified chamber for 14 days. Colony-forming Cells (CFC) were scored on day 14.

Long-term culture-initiating cell assays

Long-term culture-initiating cell (LTC-IC) assays were performed as described previously (35), 15–100,000 cells/well on day 17 for the EBs and on day 0 for the control CD34+. Absolute LTC-IC counts corresponded to the cell concentrations, yielding 37% negative wells using Poisson statistics.

Pseudo-microtubules

For Pseudo-microtubules formation, cells were transferred onto growth factor reduced Matrigel (Corning) and culture in EGM2 medium (Lonza).

Flow cytometry

Staining of hiPSCs or dissociated EBs or mouse tissue cells was performed with $2-5 \cdot 10^5$ cells in 100 μ L staining buffer (PBS containing 2% FBS and FcR Blocking Reagent) with 5:100 dilution of each antibody, for 20 min at room temperature in the dark. Viability staining was performed using either DAPI or the Zombie aqua fixable viability kit (Biolegend). Data were collected using a Becton Dickinson Canto II cytometer or a MACS Quant VyB or a MACS Quant 10 (Miltenyi).

Three labs analyzed the data, the following list of antibodies is not exhaustive.

Beckman Coulter antibodies (Brea, USA)					
Stemness markers		Lymphoid markers		Isotypes	
CD90-FITC	IM1893U	CD3-FITC	A07747	IgG1 FITC	A07795
CD109-Pe	A08933	CD4-FITC	A07750	IgG1 PE	A07796
CD34-PE	IM2648U	CD8-PE	A07757	IgG1 ECD	A07797
CD34-APC	IM2472	CD19-FITC	A07768	IgG1 PC5.5	A62833
CD117-PC5.5	A66333	Megakaryocyte marker		IgG1 PC7	737662
CD38-PC7	A54189	CD41-FITC	IM00649	IgG1 PB	A74764
Hematopoietic markers		Myeloid markers		IgG1 APC A750	A79393

CD43-FITC	IM3264U	CD14-FITC	IM0645U	IgG1 APC	IM2475
CD43-PE	A32560	CD15-PE	IM1954U	Endothelial markers	
CD45-KRO	A96416	CD13-ECD	A33097	CD146-PE	A07483
CD45-PC5.5	A62835	Differentiation markers		CD309-PE	A64615
Erythroid markers		HLA-DR-PB	A74781	CD31-PE	IM02409
CD235a-FITC	IM2212	CD33	A70200		

Clinisciences antibodies (Montrouge, France)			
anti-SSEA4-PE	NBP2-27383	anti-SSEA1-PE	NBP2-27378

Design of experiment

Optimal response surface modeling was applied to determine the optimal concentrations of the 10 selected cytokines. A two-factorial experimental design was chosen with a range of 10 independent variables. The following cytokines were selected on the basis of personal experience and data from the literature: SCF (20 to 300 ng/mL), IGF-1 (5 to 50 ng/mL), TPO (20 to 100 ng/mL), IL1 (5 to 50 ng/mL), IL3 (5 to 50 ng/mL), IL6 (5 to 50 ng/mL), G-CSF (10 to 100 ng/mL), BMP4 (20 to 200 ng/mL), VEGF (20 to 200 ng/mL) and FLT3 (20 to 300 ng/mL) (Table S1). Human iPSCs were cultured for 17 days using 16 different combinations of cytokines. In each case, the cells were phenotyped and their amplification rate and their ability to generate LTC-ICs were determined (table S2). Statistical design, regression analysis of the

results and plotting of the response surface were performed using the software JMP-8 (SAS Institute Inc., Cary, USA). The software generated an algorithm which identified three efficient combinations of cytokines (table S3).

Real-Time Quantitative PCR

After isolation of total mRNA with an RNA minikit (Qiagen, Courtaboeuf, France), the mRNA integrity was checked by analysis on a Bioanalyzer 2100 (Agilent Technologies, Massy, France). Complementary DNAs were constructed by reverse transcription with Superscript (Life Technologies, Carlsbad, USA). PCR assays were performed using a TaqMan PCR master mix (Life Technologies) and specific primers (Applied BioSystems, Carlsbad, USA) for selected genes (see table below), together with a sequence detection system (QuantStudio™ 12K Flex Real-Time PCR System, Life Technologies). In each sample the fluorescent PCR signal of each target gene was normalized to the fluorescent signal of the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

Hs00174333_m1	Human CD19	Hs00193527_m1	Human C-MYB	Hs00231079_m1	Human RUNX1
Hs01556515_m1	Human TCF7	Hs02758991_g1	Human GAPDH	Hs00758889_s1	Human HBB
Taqman array #4418803		Human hematopoiesis		Hs00361131_g1	Human HBG
Hs00362216_m1		Human HBE			

Mouse transplantation

NOD/SCID-LtSz-scid/scid (NOD/SCID) and NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) (Charles River, L'Abresle, France) were housed in the IRSN animal care facility. All experiments and procedures were performed in compliance with the French Ministry of Agriculture regulations for animal experimentation and approved by the local ethics committee (APAFIS approval number # 17559-2018111613396032v2 (2024)).

Mice, 5-8 weeks old and raised under sterile conditions, were sublethally irradiated with 2.4-3.5 Grays from a ¹³⁷Cs source (2.115 Gy/min) 24 h before cell injection. To ensure consistency between experiments, only male mice were used. Prior to transplantation, the mice were temporarily sedated with an intraperitoneal injection of ketamine and xylazine. Cells (4 x10⁵ per mouse for D17 cells) were transplanted by retro-orbital injection in a volume of 100 µL using a 28.5-gauge insulin needle. A total of 251 mice were used in this study.

For the engraftment potential of the D17 cells on the five different hiPSC lines:

188 NSG mice were used as follows: 113 primary recipients, 53 secondary recipients and 22 as negative controls.

63 NOD-SCID mice were used as follows: 35 primary recipients and 16 secondary recipients, 3 tertiary recipients and 9 as negative controls (see Table S6).

Limiting dilution.

For the limiting dilution assay (LDA), 33 NSG mice were used, i.e., 30 primary recipients and 3 controls. The number of human HSCs was determined by LDA. Increasing doses of d17 cells (1000, 5000, 10000, 50000, 100000) were retro-orbitally injected into sublethally irradiated NSG recipient mice. Twenty weeks after transplantation, the percentage of hCD45+ cells was analyzed by flow cytometry. HSC frequency was calculated and recorded using ELDA software ⁵⁷ (bioinf.wehi.edu.au/software/elda/).

Assessment of human cell engraftment

Mice were sacrificed at week 12, 18 or 20. Femurs, tibias, blood, liver, spleen and thymus were removed. Single cell suspensions were prepared by standard flushing and aliquots containing 5×10^5 cells were stained in a total volume of 100 μ L staining buffer cf flow cytometry chapter. Non-injected mouse BM was used as a control for non-specific staining.

Data were acquired on a BD Canto II cytometer/ MACSQuant 10 and MACSQuant VYB (Miltenyi).

T-cell maturity and functionality assay

The presence of TCR $\alpha\beta$ and TCR $\gamma\delta$ in peripheral blood was assessed by flow cytometry using the following human markers: hCD3 clone UCHT1 (positive gating), TCR $\alpha\beta$ clone IP26A and TCR $\gamma\delta$ clone IMMU510 (all from Beckman Coulter antibodies, Brea, USA).

Thymus and spleen cells were isolated, CFSE labelled and seeded in cell culture media complemented or not with hCD3 and hCD28 (Beckman Coulter both 1mg/ml). After 5 days, cells were harvested and stained with anti-hCD3 clone UCHT1 and analysed on a BD Canto II cytometer. FlowJo analysis software was used to gate on CD3⁺T-cells and generate the overlaid histogram plots.

Cell preparation for sequencing

PCi-CAU EB cell suspensions at different times were loaded on a Chromium single cell instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell GEMs and prepared libraries according to the manufacturer's recommendations and performed Illumina Sequencing. Experiments were done in duplicates.

Separation strategy for human and mouse cells in the scRNA-Seq of engrafted mice bone marrow

As a control, the human (SRR7881423) and mouse (SRR6835854) public BM datasets were aligned to mm10 and grch38 using CellRanger 6.0.0. Only cells that met CellRanger's filter threshold and had at least 200 features in grch38 or mm10 were selected for further analysis. All cells were then pooled and analyzed using the standard Seurat SCTransform pipeline for grch38 data: SCTransform (default parameters), RunPCA, RunUMAP (first 40 PCA dimensions, n.neighbors=100).

As shown in Figure S12A-B, the ratio between the number of features detected in grch38 and mm10 assemblies was sufficient to clearly distinguish all human cells from mouse cells. In addition, the human and mouse datasets are separated in the final UMAP plot (Figure S12E). In the mouse transplanted bone marrow dataset, although two populations could be separated based on the feature ratio of grch38/mm10, some cells had an intermediate ratio (Figure S12C-D). Because no cell type in the control exhibited such a phenotype and these intermediate cells were scattered across the different UMAP populations (Figure S12F), we considered them duplicates and filtered them out, resulting in the plot shown in Figure S12G. The human cells were those that had a ratio greater than 2.5 for mouse L1 and 2.85 for mouse L5. The mouse cells were those that had a ratio of less than 0.55 for mouse L1 and 0.47 for mouse L5. The total number of cells filtered (8%) was consistent with the expected number of duplicates using 10X Chromium technology.

ScRNA-Seq analysis of the embryoid bodies

Each individual sample was filtered to obtain only genes expressed in at least 3 cells and cells that expressed between 200 and 8000 genes. Due to a higher percentage of dead cells, only cells with more than 1300 genes were selected for one of the D15 datasets (J15-TJ3). Cells expressing more than 10% mitochondrial RNA were filtered out. The merged dataset was then normalized using SCTransform with default parameters. UMAP visualization and Leiden clustering with a resolution of 1.2 were then performed for the first 82 PCA components, the latter generated with the parameter `approx= FALSE` in Seurat. To identify cell type, differentially expressed genes were used with the Wilcoxon assay of Seurat, the `tf-idf` of SoupX, and the `scHCL` in addition to manual examination of marker genes to confirm cell identity.

To more accurately identify the cell type of hematopoietic cells, cells with a `UMAP_1` coordinate less than -7 and a `UMAP_2` coordinate greater than -1 were extracted and renormalized using SCTransform with default parameters. Leiden clustering was then performed at a resolution of 1.3 on the first 35 PCA dimensions. Cell type identification followed the same procedure as described for the entire dataset.

ScRNA-Seq analysis of mouse bone marrow

Both samples were aligned to mm10 and grch38 using CellRanger 6.1.2. For each sample, we selected only the cells expressing at least 200 genes in both genome assemblies. We identified the mouse cells from the human cells as described in the section "Separation strategy for human and mouse cells in scRNA-Seq" and filtered out the duplicates. Cells expressing less than 5000 genes on grch38 and 6000 genes on mm10 were retained. Cells expressing more than 10% of human mitochondrial RNA or 4% of mouse mitochondrial RNA were filtered out.

The merged dataset from the two mice was then normalized using SCTransform with default parameters. UMAP visualization (n.neighbors=100) and Louvain clustering at 0.6 resolution were then performed for the first 38 PCA components. The same process was used for the mouse L5 and human cell plots, except for the number of PCA components selected, which varied: 39 for mouse L5 and 40 for human cells. Cell type identification was performed for the merged dataset of the two bone marrow cells with differentially expressed genes using SoupX tf-idf gene lists, scHCL and scMCA mapping, and manual examination of marker genes.

Immunostaining of the embryoid bodies

For immunofluorescence analysis (IF), EBs were fixed with 3.7% paraformaldehyde for 15 minutes at room temperature and sectioned using a Leica cryostat CM3050. Samples were then incubated for 60 minutes at room temperature with IF buffer [phosphate-buffered saline (PBS), 0.2% Triton, 5% FBS], which also served as a blocking solution. Samples were incubated with the primary antibodies overnight at 4 °C. The following antibodies were used: anti-alpha SMA (1:500; Sigma-Aldrich A2547), anti-NG2 (1:400, Cell Signaling E3B3G), anti-VECAD (1:400, Cell Signaling D87F2), anti-CD31 (1:500, Cell Signaling 89C2) and anti-RUNX1 (1:100, Abcam ab 92336). Incubation with the secondary antibodies (1:500, Invitrogen A-21121, Jackson 711-605-152) was performed for 1 hour at room temperature. Finally, 4,6-diamidino-2-phenylindole (DAPI) nuclear staining was performed (1:500 for 5 minutes). Immunofluorescence images were acquired using a Leica DMI 6000 B microscope. Images were processed using ImageJ software.

ACCESSION NUMBERS

Sequencing data supporting the findings of this study have been deposited at the Gene Expression Omnibus (GEO) under accession code GSE224081.

AUTHOR CONTRIBUTIONS

L.G.H., OP, and MBM performed the experiments with assistance from LP, BL, BB, and CL. OP performed the single-cell transcriptome analysis. LP replicated the FACS analysis independently. H.L., L.D., and A.C. helped with the design of the study. M.B. helped design the animal study. A.C., T.J., and L.G.H. monitored the study. L.G.H. and T.J. designed the study, analyzed the data, interpreted the experiments, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank S. Viville (IGBMC, Strasbourg, France) and L. David (ISPCDC, Nantes, France) for providing iPS lines, Rima Haddad for scientific support and Sophie Gournet for excellent drawing assistance. This work benefited from equipment and services from the iGenSeq core facility, at ICM. We are indebted to F. Letourneur (Plateforme génomique Institut Cochin Paris) and to Y. Marie (Plateforme Génomique ICM) for help in single-cell sequencing and demultiplexing. Part of the bioinformatics analyses was performed on the Core Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). This research was funded by the ANR-13-ASTR-009 “GISPSI” and ANR-19-ASMA-0005 “RESCUE” from the Direction Générale de l’Armement via the ASTRID/ANR program for AC, the ANR-20-CE17-0032-01 “HEMPLUS”, the FRM grant EQU201903007867 for TJ and the Etablissement Français du Sang via the APR 2013 for LGH.

A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

FIGURE LEGENDS

Figure 1: Design of Experiment step by step.

Based on a literature review, an ensemble of cytokines and growth factors was identified (input variables) and a range of input concentrations was defined for each of them. Output variables such as cell surface markers (Table S2), cell proliferation, and LTC-IC frequency were defined. Based on these variables, 16 experimental designs were created that differed in their individual concentrations of growth factors and cytokines (Table S1). Baseline variable data from the 16 different experimental designs were collected. Based on these results, mathematical modeling of the interactions between the input and output variables was calculated. Three final combinations were proposed and tested on two different iPS cell lines, i.e FD136-25 and IMR90-16. One combination was selected for further use. Data are presented as mean +/- SEM percent of positive cells.

See also supplemental experimental procedures and Tables S1, S2, and S3.

Figure 2: *In vivo* engraftment of D17 EB cells in immunocompromised NSG mice.

(A) hiPSCs were differentiated in liquid culture supplemented with combination A of growth factors and cytokines for a total period of 17 days. EB cells were dissociated and inoculated into primary irradiated NSG recipients over a total period of 20 weeks. BM cells from primary recipients were collected and inoculated into secondary irradiated NSG recipients over a further period of 20 weeks in total.

(B) Flow cytometric analysis of human and mouse CD45⁺ cell expression in BM of a representative primary recipient.

(C) Multilineage analysis by FACS on NSG mice BM, 20 weeks post-transplantation, on hCD34⁺hCD43⁺ and hCD45⁺ HSPCs into primary recipients. Data are mean +/- SEM. Each circle/square/triangle represents a single animal.

(D) Distribution of T, B, Pro-erythrocyte and monocyte lineages in representative individual primary recipients analyzed by flow cytometry. Mice 1-6, a-e and L1-L5 are from one independent experiment analysed at one week interval i.e., a-e and L1-L5 were analysed at the same time by two independent laboratories.

(E) UMAP representation of single-cell RNA-Seq from NSG mouse L5 BM transplanted with D17 EB cells for 20 weeks. Mouse and human cell clusters are identified.

(F-G) Bone marrow of NSG mouse L5 primary recipient analyzed by FACS at 20 weeks for human (F) erythroid cells (235a⁺ and 235a⁺CD71⁺), (G) B cells (IgM⁺CD10⁺), neutrophils (CD33⁺CD14⁺), and T cells (CD4⁺CD3⁺).

(H-I) Human hematopoietic lineage in secondary recipients, 20 weeks post-transplantation. (H) FACS analysis on hCD34⁺hCD43⁺ and hCD45⁺ HSPCs into secondary recipients. (I) Distribution of T, B, Pro-erythrocyte and monocyte lineages in representative individual primary recipients analyzed by flow cytometry. Data are mean +/- SEM.

Figure 3: Functional assessment of the grafted cells

(A) Clonogenic hematopoietic assays on BM cells isolated from primary or secondary recipients. Frequency of CFU-GM, BFU-E, and CFU-GEMM colonies.

(B) Representative colonies of CFU-GEMM (1), CFU-E (2) and CFU-GM (3) from primary or secondary BM cells.

(C) Cytospins. May Grünwald-Giemsa staining of cells isolated from clonogenic assays in primary or secondary recipients. Mature macrophages (1), histio monocytes (2), myelocytes (2), and erythroblasts (3). B: bar=50µm; C: bar= 10µm

(D) Human globin expression from CB CD34⁺HSPCs erythroid culture, from BM primary and secondary recipients and BFU-E from BM primary recipients. Data are mean +/- SEM.

(E) Maturation of human T cells. Peripheral blood stained with antibodies against hTCR αβ and hTCR γδ, cells are first gated for hCD3 and analyzed for TCR expression.

(F) Human T cell functionality. The entire thymic population is labeled with CFSE at D0 and

analyzed for hCD3+ expression (green). At D5, the unstimulated population is red, while the stimulated parent population is blue.

(G) Confidence intervals of $1/(\text{stem cell frequency})$ were calculated by ELDA (<http://bioinf.wehi.edu.au/software/elda/>) according to the Poisson distribution.

Figure 4: Transcriptomic analysis of the EB cells

(A) UMAP embeddings of cells in EBs at D13, 15, 17, and 19. EVT: Extravillous Trophoblast. MEP: Megakaryocyte-Erythroid Progenitors. ECM: Extracellular Matrix. AGM: Aorta-Gonad-Mesonephros.

(B) Expression of genes in Crosse et al., 2020 characterizing endothelial populations in the AGM applied to the EB dataset

(C) UMAP embedding of cells in the data set of Crosse et al., 2020. HSPC: Hematopoietic stem and progenitor cells. PGC: Primordial germ cells. Hem: Hematopoietic cells. CL8: Cluster 8 described in Crosse et al., 2020.

(D) Expression of genes described in Crosse et al., 2020 characterizing endothelial populations in the AGM applied to Crosse et al. data set.

(E) Expression of the six-gene molecular HSC signature of Calvanese et al., 2022 in the hematopoietic population of the embryoid body dataset

(F) Close-up of the UMAP embedding on the hematopoietic cell populations in the EB dataset framed in A. MEP: Megakaryocyte-Erythroid Progenitors. HSC: Hematopoietic Stem Cells. HE: Hemogenic Endothelium. Erythro: Erythroid cells

Figure 5: Molecular characterization of the cells in the embryoid bodies

(A-C) Time-course detection of molecular markers of endothelial and hematopoietic cell populations between scRNA-Seq and q-PCR datasets comparison (A), in FACS (B) and scRNA-Seq (C).

(D-G) Immunostaining on transverse sections of D17 EBs showing spatial organization of endothelium and peri-endothelial populations. (D) Endothelial cells (CDH5). (E, F) cells expressing PECAM1 and the transcription factor RUNX1. (G) Pericytes (NG2⁺) and smooth muscle cells (αSMA⁺).

D, E, F, G: Bars=100 microns.

SUPPLEMENTARY FIGURES

Supplementary Figure 1 is related to Figure 1. *In vivo* engraftment of D17 human EB cells derived from different hiPSC lines in primary NSG mice.

(A) D17 EB cells from five different hiPSC lines i.e., FD136-25 (n=10), PC1429 (n=10), PCi-CAU (n=26), LAM01.005 (n=6), and LAM02.002 (n=7) in NSG primary recipients 20 weeks after transplantation of 4×10^5 D17 EB cells. FACS results are expressed as % of hCD34+, hCD43+, and hCD45+.

(B-C) D17 EB cells from PCi-CAU, prepared by the Atlantic Bio GMP platform (Nantes, France), at 8 weeks (n=5) (B) and 12 weeks (n=5) (C) post grafts. FACS results are expressed as % of hCD34+, hCD43+, and hCD45+.

(D) Kinetics of engraftment at 8, 12, 20 and 32 weeks post graft of D17 EB cells from Pci-CAU in NSG mice. Results are % of hCD45+ cells.

(E) NSG primary recipient after transplantation with 4×10^5 CD34+ CB cells (n=13). FACS results are expressed as % of hCD34+, hCD43+, and hCD45+.

Supplementary Figure 2 is related to Figure 2. Multilineage engraftment. Flow cytometric analysis of BM. Primary recipients at 20 weeks.

(A) Human hematopoietic cell lineages in primary recipients transplanted with 4×10^5 D17 EB cells (n=30) showing Myeloid (M), B, T and Pro-erythroid cells (Pro-E). hCD45+ population was analyzed for expression of hCD33 (M), hCD19 (B), and hCD3 (T), whereas hCD235a (Pro-E) was analyzed on whole BM cells. Data are mean +/- SEM.

(B) Representative control primary recipient (n=20)

(C-D) Representative primary recipient transplanted with PCi-CAU. (n=26).

(C) Flow cytometric analysis for myeloid (hCD14/hCD15) and lymphoid (hCD19/hCD20, SSC-A/hCD3 and hCD4/hCD8) lineages. hCD45+ gated BM cells.

(D) Flow cytometric analysis of the erythroid lineage. Analyzes were performed on whole BM cells.

(E) Human hematopoietic cell lineages in primary recipients transplanted with 4×10^5 CD34+ CB cells (n=13).

(F) Representative NSG recipient, hCD45+ gated BM cells.

G) Flow cytometric analysis of the erythroid lineage. Analyzes were performed on whole BM cells.

Supplementary Figure 3 is related to Figure 2. Multilineage engraftment. Flow cytometric analysis of NSG peripheral blood, thymus and spleen. Primary recipients at 20 weeks.

Myeloid (CD11B/CD14) and lymphoid (CD19/IgM and CD4/CD8) lineages in peripheral blood. Analyzes were performed on human CD45+ cells, except for non-transplanted recipients.

(A) Representative control (n=20).

(B) Representative NSG transplanted with D17 EB cells (n=26).

C) Representative NSG transplanted with CD34+ CB cells (n=13).

(D-E) Representative NSG Thymus. hTCR alpha/beta and gamma/delta flow cytometric analysis.

(D) D17 EB cells-transplanted NSG (n=20).

E) CD34+ CB cells-transplanted NSG (n=13). The CD3-gated thymus cells had a large amount of TCR-alpha-beta cells, very similar to the amount found on CD34+ CB cells.

(F-G) Representative NSG spleen. hTCR beta and hCD19 flow cytometric analysis.

(F) D17 EB cells-transplanted NSG (n=26).

(G) CD34+ CB cells-transplanted NSG (n=13).

The hCD45+-gated splenocytes exhibited a B- or T-lymphoid phenotype, with similar ratios to CD34+ CB cells.

Supplementary Figure 4 is related to Figure 2. ScRNA-seq analysis of the BM from two representative mice engrafted with D17 EB cells 20 weeks after transplant

(A, C) UMAP embedding of a merged data set from two representative transplanted mice. BM data set colored by cell type (A) and by sample (B).

(B, D) UMAP embedding of a merged dataset of human cells in the two engrafted BM datasets colored by cell type (C) and by sample (D).

(E) Stacked histogram of relative proportion of major hematopoietic cell populations detected in the transplanted mouse BM. Myeloid (M), Pro-erythroid cells (Pro-E)

(F-G) Bone marrow chimerism of primary mouse L5 transplanted with D17 EB cells at 20 weeks. (F) Erythroid cells (CD235a/CD71) from a total of BM cells. (G) B cells (IgM⁺/CD10⁺), neutrophils (CD33⁺/CD14⁺), and T cells (CD4⁺/CD3⁺).

(H) Expression of globin genes in human early erythroid cells in transplanted mouse bone marrow datasets.

Supplementary Figure 5 is related to Figure 2. Globin RT-qPCR analysis from BFU-E at the single colony level.

(A) Individual BFU-E colonies from BM cell cultures of primary recipients (approximately 18% of total colonies).

(B) CD34⁺ human CB cells. Shown is the ratio of beta (blue), gamma (red) and epsilon (green) globin. CD34 CB cells give rise to colonies with a homogeneous ratio of beta- and gamma-globin, whereas BM cells give rise to a more diverse globin composition with a weak expression of epsilon-globin.

Supplementary Figure 6 is related to Figure 2. Molecular characterization of the grafted cells.

(A) qRT-PCR analysis for hCD19 (B lymphocyte surface antigen B4, B cell-specific marker) and hTCF7, a T-cell-specific marker, was performed on thymus, spleen, and bone marrow. For each gene, delta Ct is the mean \pm SEM of 3 independent experiments.

(B) Cytospins. May Grünwald-Giemsa staining of cells isolated from clonogenic assays in primary and secondary recipients. Scale bars represent 10 μ m.

(C) Representative qRT-PCR results obtained with a TaqMan[®] Array Human Hematopoiesis 96-well plate.

Supplementary Figure 7 is related to Figure 2. *In vivo* engraftment of D17 human EB cells derived from different hiPSC lines in secondary NSG mice by FACS analysis.

(A) D17 EBs from five different hiPSC lines i.e., FD136-25 (n=10), PC1429 (n=10), PCi-CAU (n=10), LAM01.005 (n=4), and LAM02.002 (n=6) in NSG secondary recipients 20 weeks after transplantation of 7x10⁶ D17 EBs. Results are expressed as % of CD34⁺, CD43⁺, and CD45⁺ human cells.

(B) NSG secondary recipient 20 weeks after transplantation with 7x10⁶ CD34⁺ CB cells (n=13).

Supplementary Figure 8 is related to Figure 2. Multilineage engraftment. Flow cytometric analysis of BM. Secondary recipients at 20 weeks.

Flow cytometric analysis of myeloid (CD14/CD15), lymphoid (CD19/CD20, SSC-A/CD3 and CD4/CD8) lineages and erythroid (CD235a) lineages. Cells were gated on hCD45⁺ expression

for CD33, CD19, and CD3, while CD235a was analyzed on total BM cells.

(A) Representative control, secondary recipient (n=9)

(B-C) Representative secondary recipient transplanted with 7.10^6 whole BM from primary D17 EB recipients. (n=10).

(B) Flow cytometric analysis for myeloid (hCD14/hCD15) and lymphoid (hCD19/hCD20, SSC-A/hCD3 and hCD4/hCD8) lineages. hCD45+ gated BM cells.

(C) Flow cytometric analysis of the erythroid lineage. Analyzes were performed on whole BM cells.

(D) Distribution of human hematopoietic lineages in secondary recipients transplanted with 7.10^6 CD34+ CB cells from primary recipient whole bone marrow (n=13). Cells were gated using hCD45+ expression for CD33, CD19, and CD3, while CD235a was analyzed on total BM cells. Data are mean +/- SEM.

(E-F) Representative NSG mouse.

(E) hCD45+ gated BM cells.

(F) Erythroid lineage. Analyzes were performed on whole BM cells.

Supplementary Figure 9 is related to Figure 2. Hematopoietic engraftment in primary, secondary and tertiary NOD-SCID recipients.

(A) Primary NOD-SCID recipient inoculated with 4.10^5 D17 EB cells (n=20).

(B) Secondary NOD-SCID recipients inoculated with 7.10^6 BM cells from primary recipients (n=16).

(C) Tertiary NOD-SCID recipient inoculated with 7.10^6 BM cells from secondary recipients (n=3).

Supplementary Figure 10 is related to Figure 4: Identification and evolution of the cell populations detected in the embryoid bodies by scRNA-Seq at D13, 15, 17 and 19.

Repartition of the cells from each timepoint on the UMAP embedding of the embryoid bodies. EVT: Extravillous Trophoblast. MEP: Megakaryocyte-Erythroid Progenitors. ECM: Extracellular Matrix. AGM: Aorta-Gonad-Mesonephros.

Supplementary Figure 11 is related to Figure 4. Endothelial and hematopoietic potential of D0 to D19 EBs *ex vivo*.

EBs were assessed for their ability to form endothelial or hematopoietic colonies using approved functional assays.

(A) Tubule formation in Matrigel and flow cytometry with Ac LDL uptake and hCD31 as a marker. The highest EC differentiation was found between D7 and D13. Scale bars represent 100µm.

(B) Colony forming units from D0 to D19 EB cells. CFUs peaked at D17, but the number of hematopoietic cells remained low, indicating weak hematopoietic differentiation. Scale bars represent 200µm

(C) Histograms showing the frequency of CFUs from D0 to D19.

(D) LTC-IC assays were performed with D13, D15, and D17 cells. Confidence intervals of 1/ (LTC-IC cell frequency) were calculated using ELDA (<http://bioinf.wehi.edu.au/software/elda/>) according to Poisson distribution.

Supplementary Figure 12 is related to Figure 4: Separation strategy for human and mouse cells in scRNA-Seq

(A, C) Scatter plot of the number of genes detected in each cell on the grch38 and the mm10 assembly in the control dataset (A) and the engrafted mouse dataset (C)

(B, D) Violin Plot of the ratio of genes detected with the grch38 assembly and the mm10 assembly in the control dataset (B) and the engrafted mouse dataset (D)

(E, F) UMAP embedding of the cells profiled by scRNA-Seq in the control dataset (E) and the engrafted mouse dataset (F)

(G) UMAP embeddings of the cells in the engrafted mouse dataset before filtering out doublets.

REFERENCES

1. Lu, S.J., Feng, Q., Park, J.S., and Lanza, R. (2010). Directed differentiation of red blood cells from human embryonic stem cells. *Methods Mol Biol* 636, 105-121. 10.1007/978-1-60761-691-7_7.
2. Lapillonne, H., Kobari, L., Mazurier, C., Tropel, P., Giarratana, M.C., Zanella-Cleon, I., Kiger, L., Wattenhofer-Donze, M., Puccio, H., Hebert, N., et al. (2010). Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. *Haematologica* 95, 1651-1659. 10.3324/haematol.2010.023556.

3. Hirose, S., Takayama, N., Nakamura, S., Nagasawa, K., Ochi, K., Hirata, S., Yamazaki, S., Yamaguchi, T., Otsu, M., Sano, S., et al. (2013). Immortalization of erythroblasts by c-MYC and BCL-XL enables large-scale erythrocyte production from human pluripotent stem cells. *Stem Cell Reports* *1*, 499-508. 10.1016/j.stemcr.2013.10.010.
4. Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., Tani, K., and Nakamura, Y. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS One* *8*, e59890. 10.1371/journal.pone.0059890.
5. Carpenter, L., Malladi, R., Yang, C.T., French, A., Pilkington, K.J., Forsey, R.W., Sloane-Stanley, J., Silk, K.M., Davies, T.J., Fairchild, P.J., et al. (2011). Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood* *117*, 4008-4011. 10.1182/blood-2010-08-299941.
6. French, A., Yang, C.T., Taylor, S., Watt, S.M., and Carpenter, L. (2015). Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. *Stem Cells Dev* *24*, 1082-1095. 10.1089/scd.2014.0318.
7. Galic, Z., Kitchen, S.G., Kacena, A., Subramanian, A., Burke, B., Cortado, R., and Zack, J.A. (2006). T lineage differentiation from human embryonic stem cells. *Proc Natl Acad Sci U S A* *103*, 11742-11747. 10.1073/pnas.0604244103.
8. Nishimura, T., Kaneko, S., Kawana-Tachikawa, A., Tajima, Y., Goto, H., Zhu, D., Nakayama-Hosoya, K., Iriguchi, S., Uemura, Y., Shimizu, T., et al. (2013). Generation of rejuvenated antigen-specific T cells by reprogramming to pluripotency and redifferentiation. *Cell Stem Cell* *12*, 114-126. 10.1016/j.stem.2012.11.002.
9. Themeli, M., Kloss, C.C., Ciriello, G., Fedorov, V.D., Perna, F., Gonen, M., and Sadelain, M. (2013). Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy. *Nat Biotechnol* *31*, 928-933. 10.1038/nbt.2678.
10. Vizcardo, R., Masuda, K., Yamada, D., Ikawa, T., Shimizu, K., Fujii, S., Koseki, H., and Kawamoto, H. (2013). Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8 (+) T cells. *Cell Stem Cell* *12*, 31-36. 10.1016/j.stem.2012.12.006.
11. Haruta, M., Tomita, Y., Yuno, A., Matsumura, K., Ikeda, T., Takamatsu, K., Haga, E., Koba, C., Nishimura, Y., and Senju, S. (2013). TAP-deficient human iPSC cell-derived myeloid cell lines as unlimited cell source for dendritic cell-like antigen-presenting cells. *Gene Ther* *20*, 504-513. 10.1038/gt.2012.59.
12. Takayama, N., Nishikii, H., Usui, J., Tsukui, H., Sawaguchi, A., Hiroyama, T., Eto, K., and Nakauchi, H. (2008). Generation of functional platelets from human embryonic stem cells in vitro via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors. *Blood* *111*, 5298-5306. 10.1182/blood-2007-10-117622.
13. Nakamura, S., Sugimoto, N., and Eto, K. (2021). Development of platelet replacement therapy using human induced pluripotent stem cells. *Dev Growth Differ* *63*, 178-186. 10.1111/dgd.12711.
14. Flippe, L., Gaignerie, A., Serazin, C., Baron, O., Saulquin, X., Anegon, I., David, L., and Guillonnet, C. (2022). Generation of CD34 (+)CD43 (+) Hematopoietic Progenitors to Induce Thymocytes from Human Pluripotent Stem Cells. *Cells* *11*. 10.3390/cells11244046.
15. Wang, Y., Yates, F., Naveiras, O., Ernst, P., and Daley, G.Q. (2005). Embryonic stem cell-derived hematopoietic stem cells. *Proc Natl Acad Sci U S A* *102*, 19081-19086.
16. Narayan, A.D., Chase, J.L., Lewis, R.L., Tian, X., Kaufman, D.S., Thomson, J.A., and Zanjani, E.D. (2006). Human embryonic stem cell-derived hematopoietic cells are capable of engrafting primary as well as secondary fetal sheep recipients. *Blood* *107*, 2180-2183. 10.1182/blood-2005-05-1922.
17. Ledran, M.H., Krassowska, A., Armstrong, L., Dimmick, I., Renstrom, J., Lang, R., Yung, S., Santibanez-Coref, M., Dzierzak, E., Stojkovic, M., et al. (2008). Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches. *Cell Stem Cell* *3*, 85-98.

18. Risueno, R.M., Sachlos, E., Lee, J.H., Lee, J.B., Hong, S.H., Szabo, E., and Bhatia, M. (2012). Inability of human induced pluripotent stem cell-hematopoietic derivatives to downregulate microRNAs in vivo reveals a block in xenograft hematopoietic regeneration. *Stem Cells* 30, 131-139. 10.1002/stem.1684.
19. Sandler, V.M., Lailier, N., and Bouhassira, E.E. (2011). Reprogramming of embryonic human fibroblasts into fetal hematopoietic progenitors by fusion with human fetal liver CD34+ cells. *PLoS One* 6, e18265. 10.1371/journal.pone.0018265.
20. Szabo, E., Rampalli, S., Risueno, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* 468, 521-526. 10.1038/nature09591.
21. Sandler, V.M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J.M., Scandura, J.M., and Rafii, S. (2014). Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature* 511, 312-318. 10.1038/nature13547.
22. Doulatov, S., Vo, L.T., Chou, S.S., Kim, P.G., Arora, N., Li, H., Hadland, B.K., Bernstein, I.D., Collins, J.J., Zon, L.I., and Daley, G.Q. (2013). Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* 13, 459-470. 10.1016/j.stem.2013.09.002.
23. Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.F., Goettel, J.A., Serrao, E., Rowe, R.G., Malleshaiah, M., et al. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* 545, 432-438. 10.1038/nature22370.
24. Lis, R., Karrasch, C.C., Poulos, M.G., Kunar, B., Redmond, D., Duran, J.G.B., Badwe, C.R., Schachterle, W., Ginsberg, M., Xiang, J., et al. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature* 545, 439-445. 10.1038/nature22326.
25. Amabile, G., Welner, R.S., Nombela-Arrieta, C., D'Alise, A.M., Di Ruscio, A., Ebralidze, A.K., Kraysberg, Y., Ye, M., Kocher, O., Neuberger, D.S., et al. (2013). In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. *Blood* 121, 1255-1264. 10.1182/blood-2012-06-434407.
26. Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., and Nakauchi, H. (2013). Generation of engraftable hematopoietic stem cells from induced pluripotent stem cells by way of teratoma formation. *Mol Ther* 21, 1424-1431. 10.1038/mt.2013.71.
27. McDonald, D., Wu, Y., Dailamy, A., Tat, J., Parekh, U., Zhao, D., Hu, M., Tipps, A., Zhang, K., and Mali, P. (2020). Defining the Teratoma as a Model for Multi-lineage Human Development. *Cell* 183, 1402-1419 e1418. 10.1016/j.cell.2020.10.018.
28. Dzierzak, E., and Bigas, A. (2018). Blood Development: Hematopoietic Stem Cell Dependence and Independence. *Cell Stem Cell* 22, 639-651. 10.1016/j.stem.2018.04.015.
29. Canu, G., and Ruhrberg, C. (2021). First blood: the endothelial origins of hematopoietic progenitors. *Angiogenesis* 24, 199-211. 10.1007/s10456-021-09783-9.
30. Ottersbach, K. (2019). Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans* 47, 591-601. 10.1042/BST20180320.
31. Lange, L., Morgan, M., and Schambach, A. (2021). The hemogenic endothelium: a critical source for the generation of PSC-derived hematopoietic stem and progenitor cells. *Cell Mol Life Sci* 78, 4143-4160. 10.1007/s00018-021-03777-y.
32. Sturgeon, C.M., Ditadi, A., Clarke, R.L., and Keller, G. (2013). Defining the path to hematopoietic stem cells. *Nat Biotechnol* 31, 416-418. 10.1038/nbt.2571.
33. Slukvin, I. (2013). Hematopoietic specification from human pluripotent stem cells: current advances and challenges toward de novo generation of hematopoietic stem cells. *Blood* 122, 4035-4046. 10.1182/blood-2013-07-474825.
34. Vo, L.T., and Daley, G.Q. (2015). De novo generation of HSCs from somatic and pluripotent stem cell sources. *Blood* 125, 2641-2648. 10.1182/blood-2014-10-570234.
35. Arboretti, R., Ceccato, R., Pegoraro, L., Salmaso, L., Housmekerides, C., Spadoni, L., Pierangelo, E., Quaggia, S., Tveit, C., and Vianello, S. (2022). Machine learning and design of

- experiments with an application to product innovation in the chemical industry. *J Appl Stat* **49**, 2674-2699. 10.1080/02664763.2021.1907840.
36. Huang, S., and Terstappen, L.W. (1994). Lymphoid and myeloid differentiation of single human CD34+, HLA-DR+, CD38- hematopoietic stem cells. *Blood* **83**, 1515-1526.
 37. Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., and Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* **102**, 906-915.
 38. Lam, A.C., Li, K., Zhang, X.B., Li, C.K., Fok, T.F., Chang, A.M., James, A.E., Tsang, K.S., and Yuen, P.M. (2001). Preclinical ex vivo expansion of cord blood hematopoietic stem and progenitor cells: duration of culture; the media, serum supplements, and growth factors used; and engraftment in NOD/SCID mice. *Transfusion* **41**, 1567-1576. 10.1046/j.1537-2995.2001.41121567.x.
 39. Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, I. (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* **105**, 617-626.
 40. Lengerke, C., Grauer, M., Niebuhr, N.I., Riedt, T., Kanz, L., Park, I.H., and Daley, G.Q. (2009). Hematopoietic development from human induced pluripotent stem cells. *Ann N Y Acad Sci* **1176**, 219-227. 10.1111/j.1749-6632.2009.04606.x.
 41. Watari, K., Lansdorp, P.M., Dragowska, W., Mayani, H., and Schrader, J.W. (1994). Expression of interleukin-1 beta gene in candidate human hematopoietic stem cells. *Blood* **84**, 36-43.
 42. Ditadi, A., Sturgeon, C.M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A.D., Azzola, L., Ng, E.S., Stanley, E.G., French, D.L., et al. (2015). Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat Cell Biol* **17**, 580-591. 10.1038/ncb3161.
 43. Dick, J.E., Pflumio, F., and Lapidot, T. (1991). Mouse models for human hematopoiesis. *Semin Immunol* **3**, 367-378.
 44. Tourino, C., Pflumio, F., Novault, S., Masse, A., Guiller, M., Bonnet, M.L., Valteau-Couanet, D., Hartmann, O., Vainchenker, W., Beaujean, F., et al. (2001). Efficient ex vivo expansion of NOD/SCID-repopulating cells with lympho-myeloid potential in hematopoietic grafts of children with solid tumors. *Hematol J* **2**, 108-116. 10.1038/sj/thj/6200083.
 45. Vodyanik, M.A., Thomson, J.A., and Slukvin, I. (2006). Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. *Blood* **108**, 2095-2105. 10.1182/blood-2006-02-003327.
 46. Uenishi, G., Theisen, D., Lee, J.H., Kumar, A., Raymond, M., Vodyanik, M., Swanson, S., Stewart, R., Thomson, J., and Slukvin, I. (2014). Tenascin C promotes hematoendothelial development and T lymphoid commitment from human pluripotent stem cells in chemically defined conditions. *Stem Cell Reports* **3**, 1073-1084. 10.1016/j.stemcr.2014.09.014.
 47. Shultz, L.D., Lyons, B.L., Burzenski, L.M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S.D., King, M., Mangada, J., et al. (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hematopoietic stem cells. *J Immunol* **174**, 6477-6489. 10.4049/jimmunol.174.10.6477.
 48. Crosse, E.I., Gordon-Keylock, S., Rybtsov, S., Binagui-Casas, A., Felchle, H., Nnadi, N.C., Kirschner, K., Chandra, T., Tamagno, S., Webb, D.J., et al. (2020). Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human Hematopoietic Stem Cell Development. *Cell Stem Cell* **27**, 822-839 e828. 10.1016/j.stem.2020.08.004.
 49. Smadja, D.M., Cornet, A., Emmerich, J., Aiach, M., and Gaussem, P. (2007). Endothelial progenitor cells: characterization, in vitro expansion, and prospects for autologous cell therapy. *Cell Biol Toxicol* **23**, 223-239. 10.1007/s10565-007-0177-6.
 50. Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E.S., Ekstrand, S., Aguade-Gorgorio, J., Vavilina, A., Lefaudeux, D., et al. (2022). Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature*. 10.1038/s41586-022-04571-x.

51. Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., Bian, L., et al. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res* 29, 881-894. 10.1038/s41422-019-0228-6.
52. Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L., and Medvinsky, A. (2011). Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J Exp Med* 208, 2417-2427. jem.20111688 [pii]
10.1084/jem.20111688.
53. Rongvaux, A., Takizawa, H., Strowig, T., Willinger, T., Eynon, E.E., Flavell, R.A., and Manz, M.G. (2013). Human hemato-lymphoid system mice: current use and future potential for medicine. *Annu Rev Immunol* 31, 635-674. 10.1146/annurev-immunol-032712-095921.
54. Sippel, T.R., Radtke, S., Olsen, T.M., Kiem, H.P., and Rongvaux, A. (2019). Human hematopoietic stem cell maintenance and myeloid cell development in next-generation humanized mouse models. *Blood Adv* 3, 268-274. 10.1182/bloodadvances.2018023887.
55. Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., Ueyama, Y., Koyanagi, Y., Sugamura, K., Tsuji, K., et al. (2002). NOD/SCID/gamma (c) (null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* 100, 3175-3182. 10.1182/blood-2001-12-0207.
56. Fitch, S.R., Kimber, G.M., Wilson, N.K., Parker, A., Mirshekar-Syahkal, B., Gottgens, B., Medvinsky, A., Dzierzak, E., and Ottersbach, K. (2012). Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis. *Cell Stem Cell* 11, 554-566. S1934-5909 (12)00417-1 [pii]
10.1016/j.stem.2012.07.002.
57. Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* 347, 70-78. 10.1016/j.jim.2009.06.008.

LITERATURE MINING



STATISTICAL STEPS



RESULTS



Figure 1

INPUT VARIABLES

Concentrations of growth factors involved in mesoderm commitment

BMP4, VEGF

Concentrations of cytokines involved in hematopoietic commitment

SCF, TPO, FLT3, IL3, IL6, IL1, GCSF, IGF

OUTPUT VARIABLES

Cell amplification

Cell characterization by flow cytometry

CD34, CD117, CD43, CD45, CD13, CD33, CD38, CD90, CD109, HLA-DR

Long Term Culture-Initiating Cell (LTC-IC) frequency

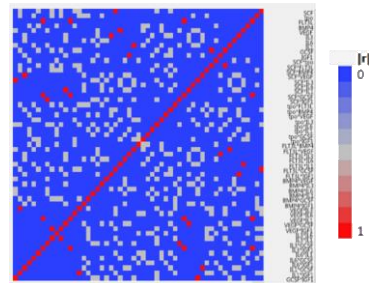
CUSTOM DESIGN

16 experiments displaying statistically different concentrations of cytokines and growth factors

Cell amplification : 16 results

Flow cytometry : 144 results

LTC-IC frequency : 16 results

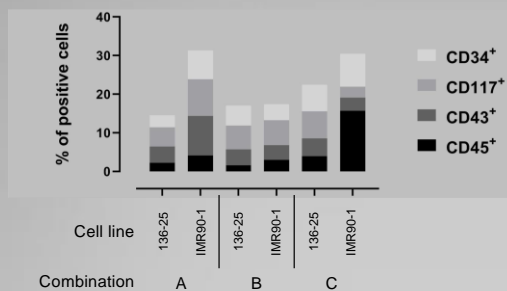


Mathematical modeling of the input/output variable interactions

Three combinations of cytokines were selected to obtain highly amplifying cells harboring hematopoietic commitment
CD34, CD117, CD43, CD45

Flow cytometric analysis of D17 Ebs

Tested on two different hiPSC lines, FD136-25 and IMR90



NOD SCID primary recipients

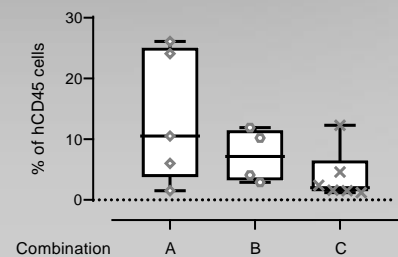


Figure 2

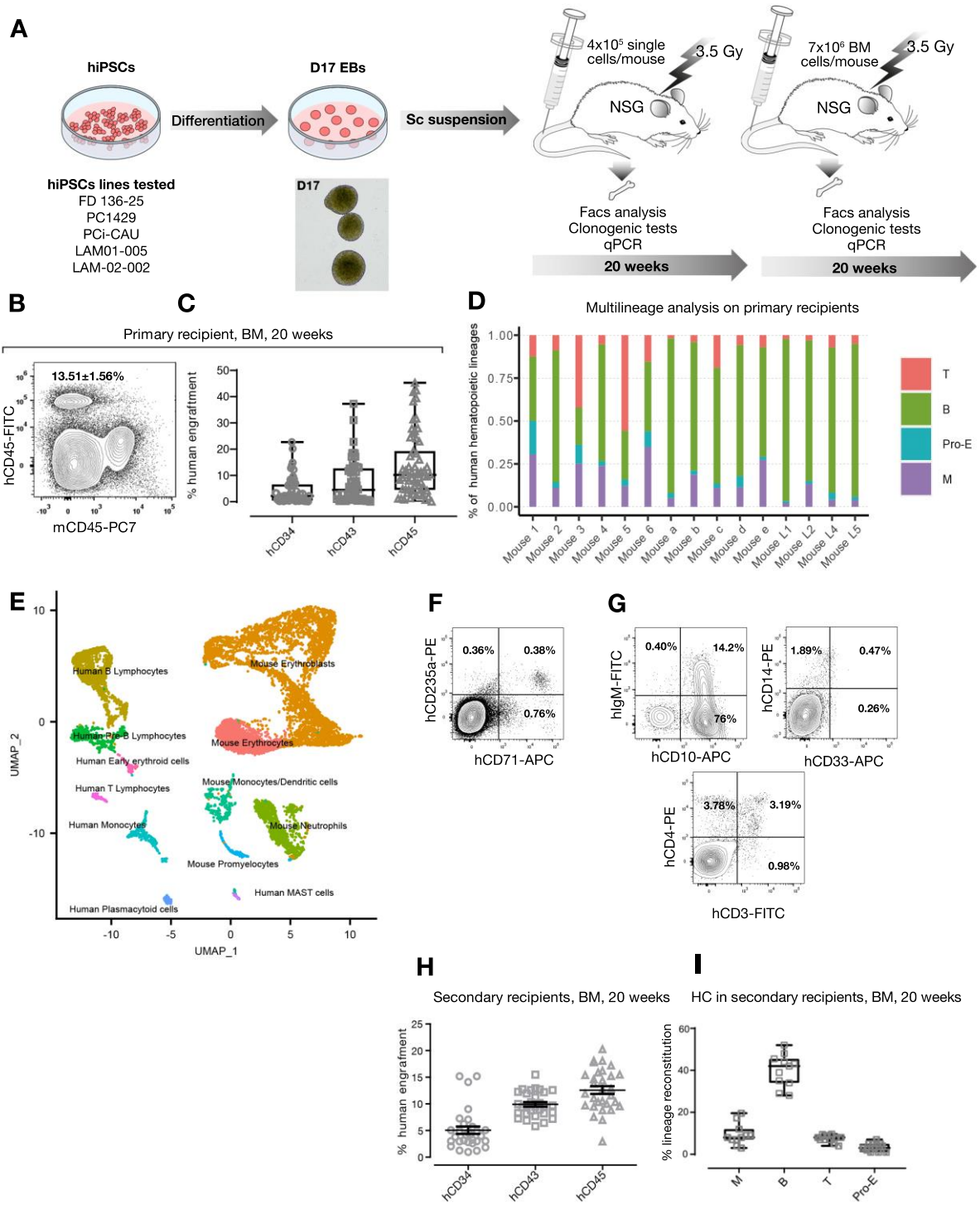


Figure 3

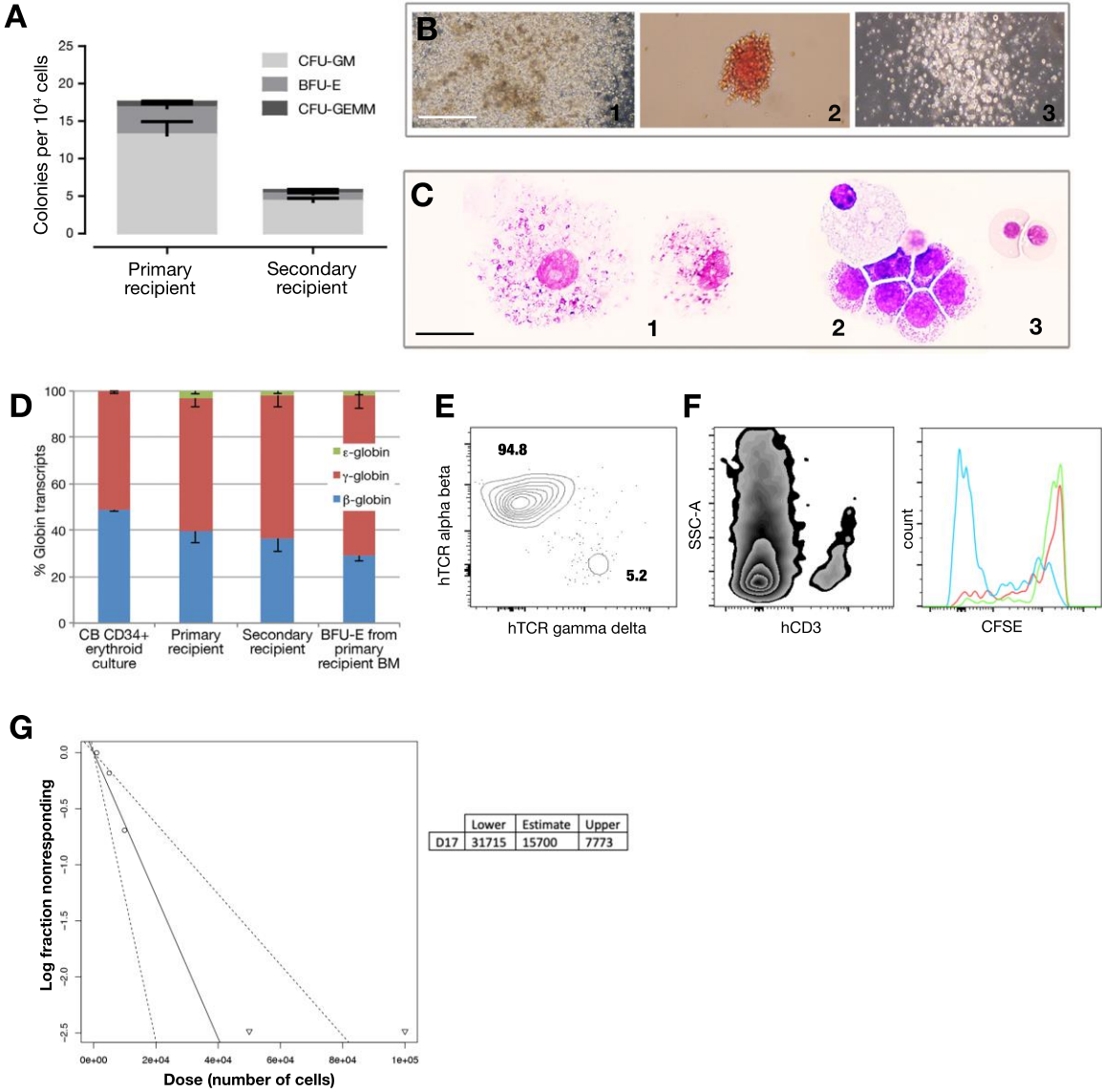


Figure 4

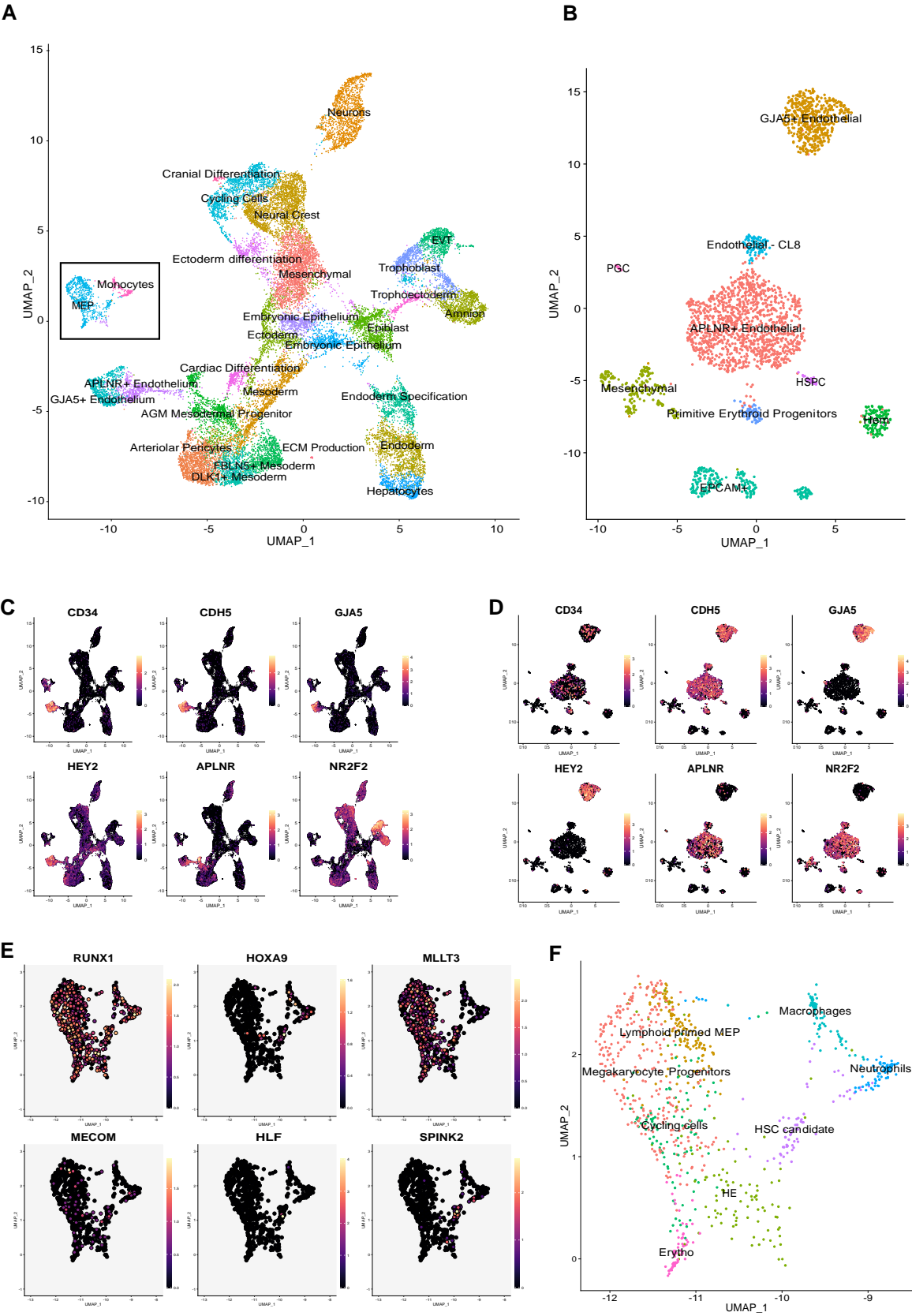


Figure 5

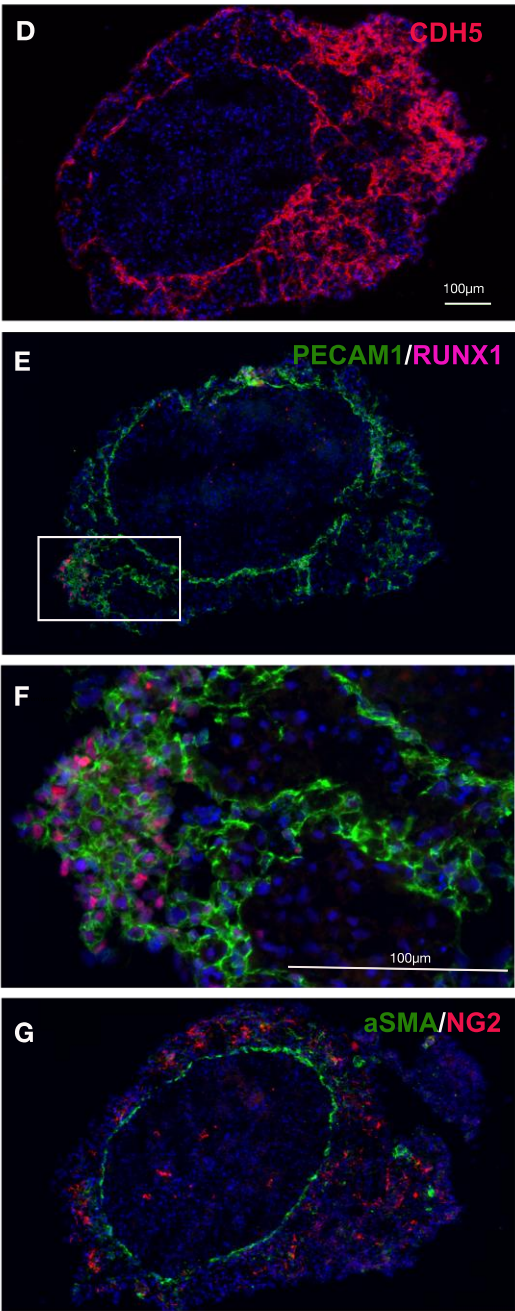
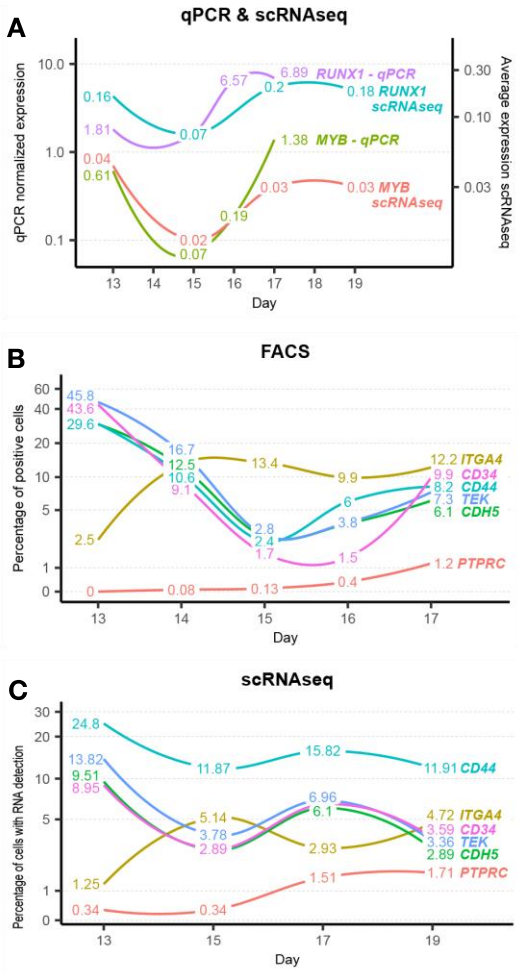


Figure S1

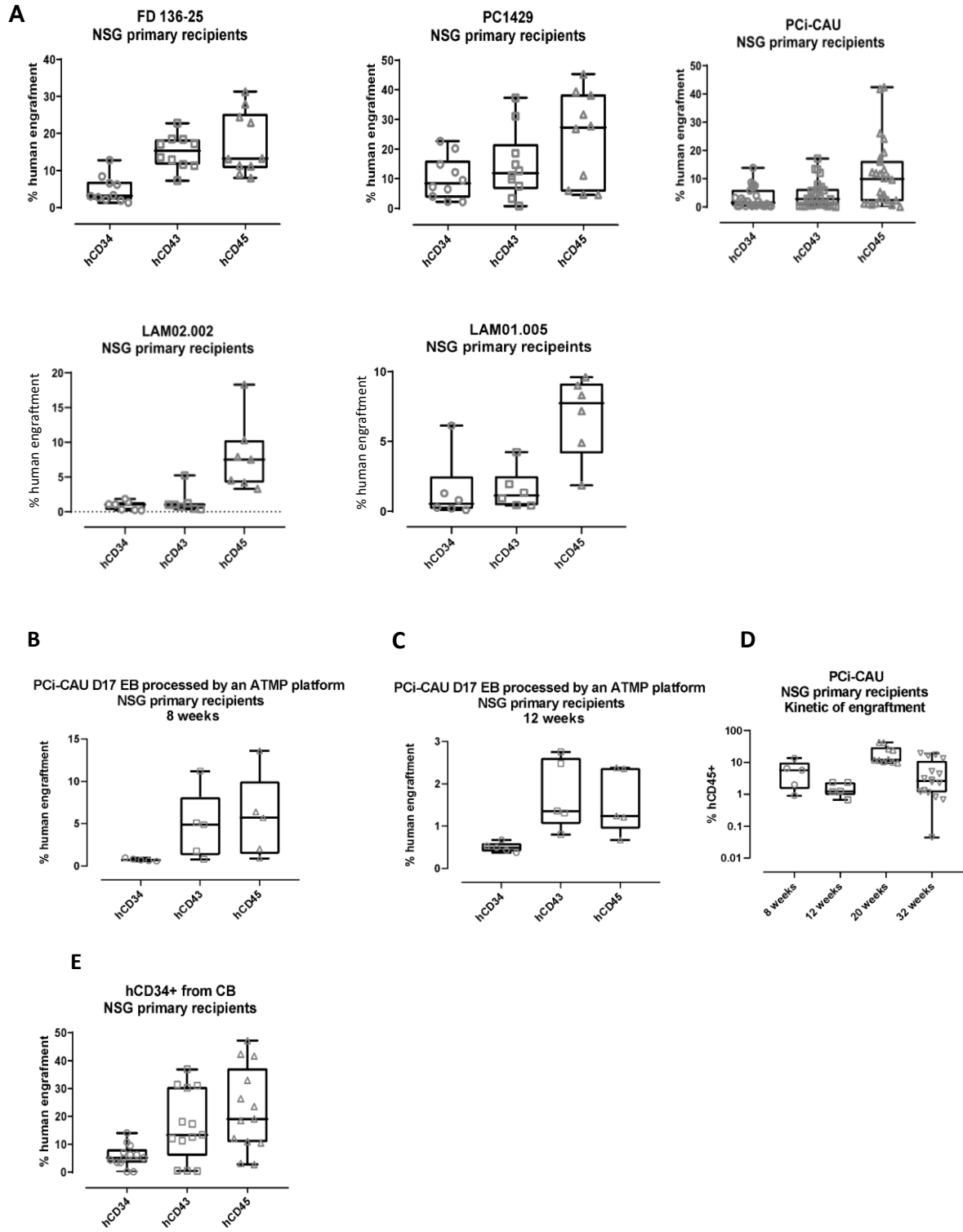


Figure S2

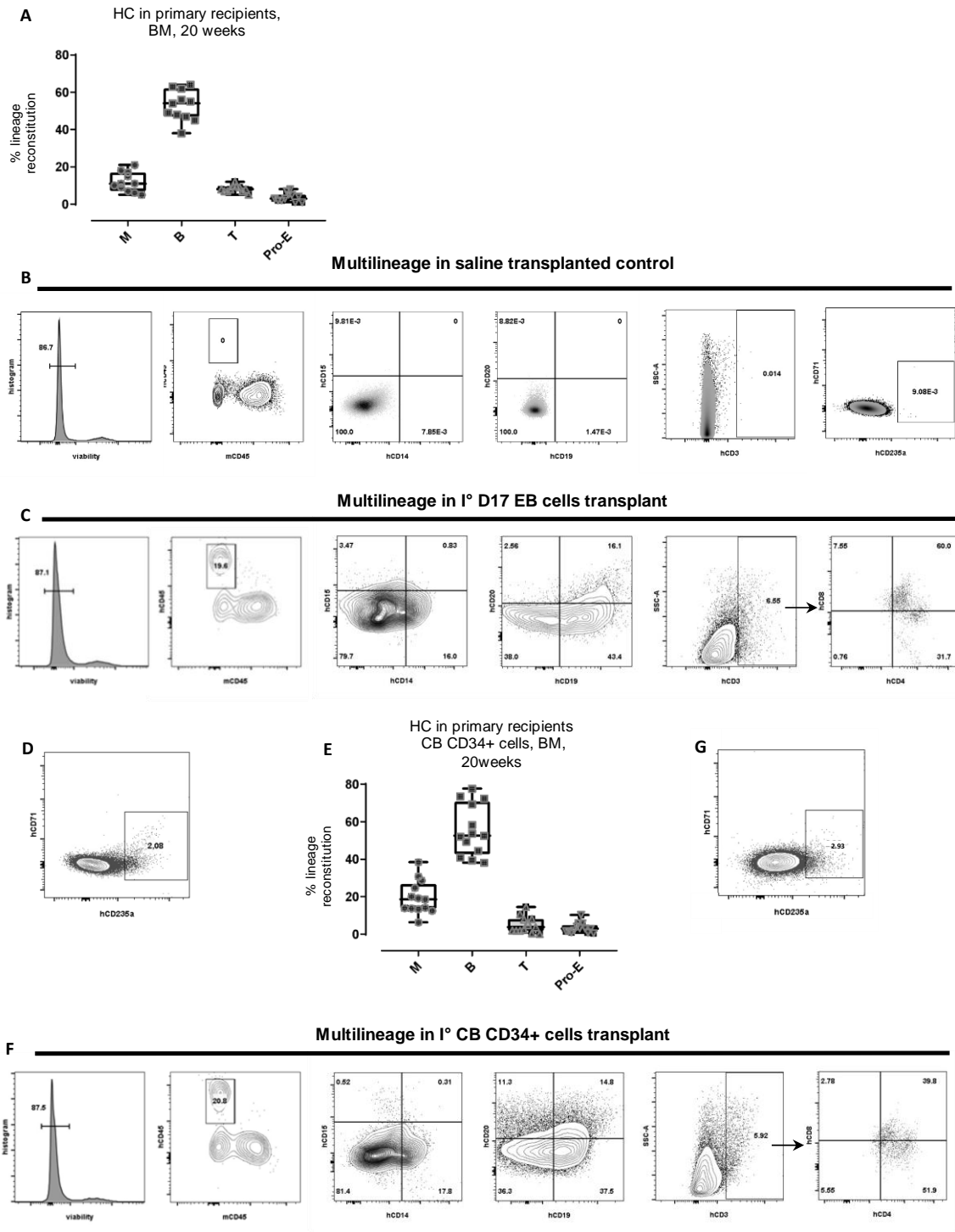
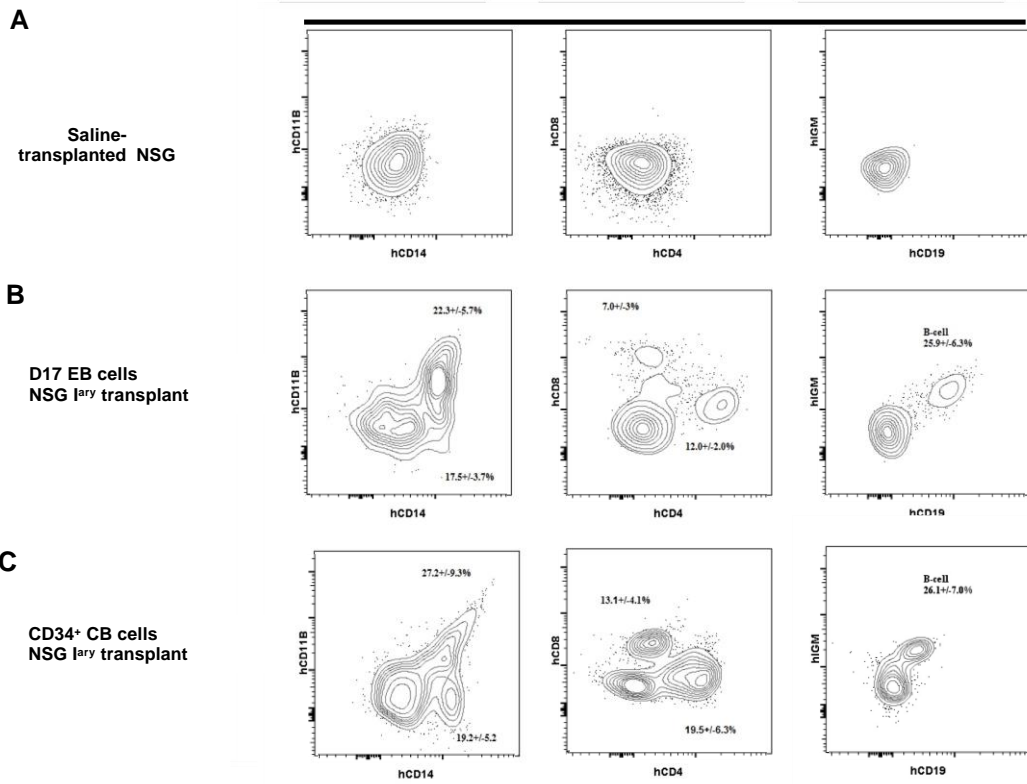


Figure S3

Peripheral blood. Multilineage analysis



Multilineage analysis

D17 EB cells I^{ary} transplant CD34+ CB cells I^{ary} transplant

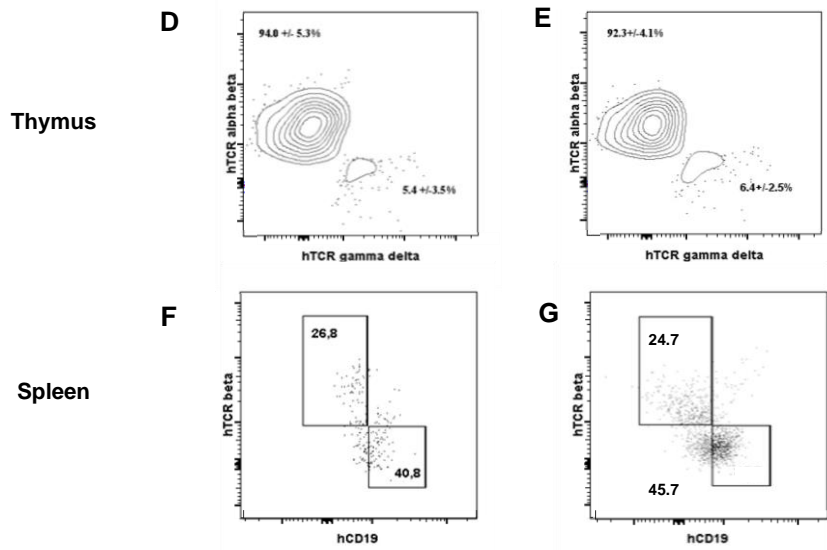


Figure S4

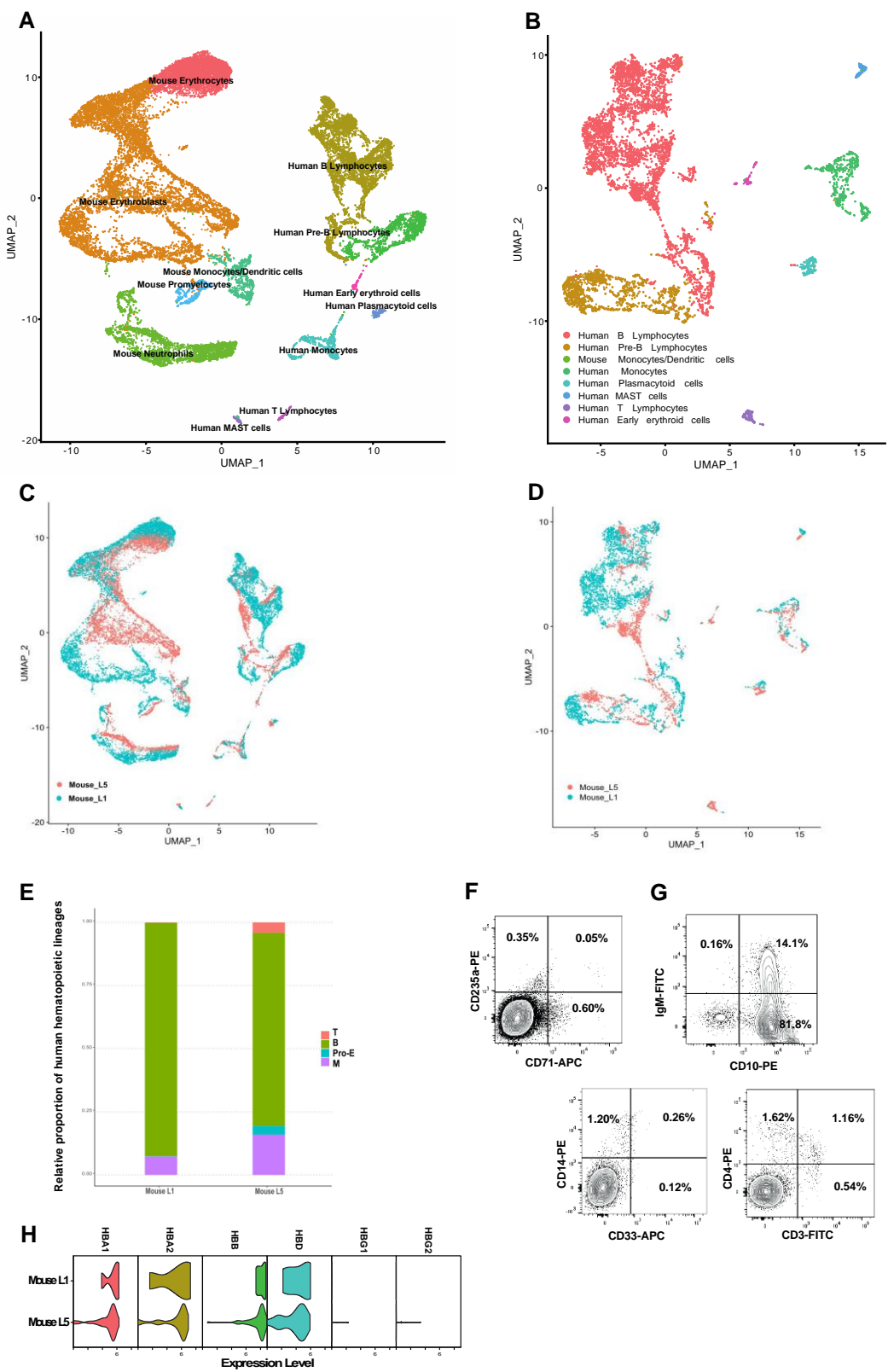
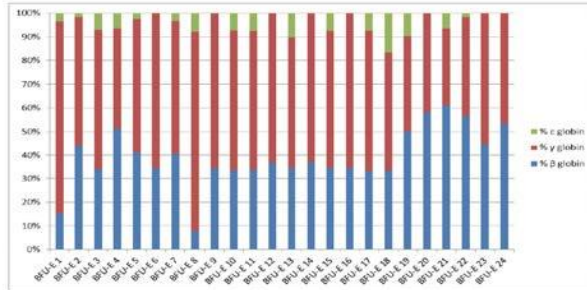


Figure S5

Globin q-PCR analysis at single BFU-E colony level

A

Human BFU-E from
BM primary recipients



B

Human BFU-E from
CD34⁺ CB cells

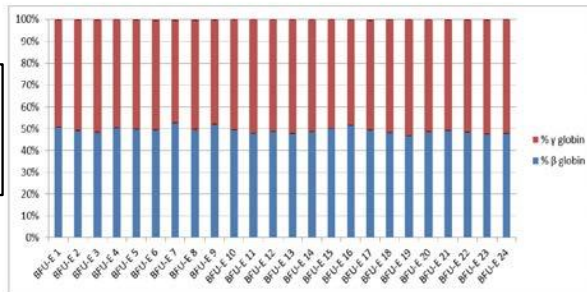


Figure S6

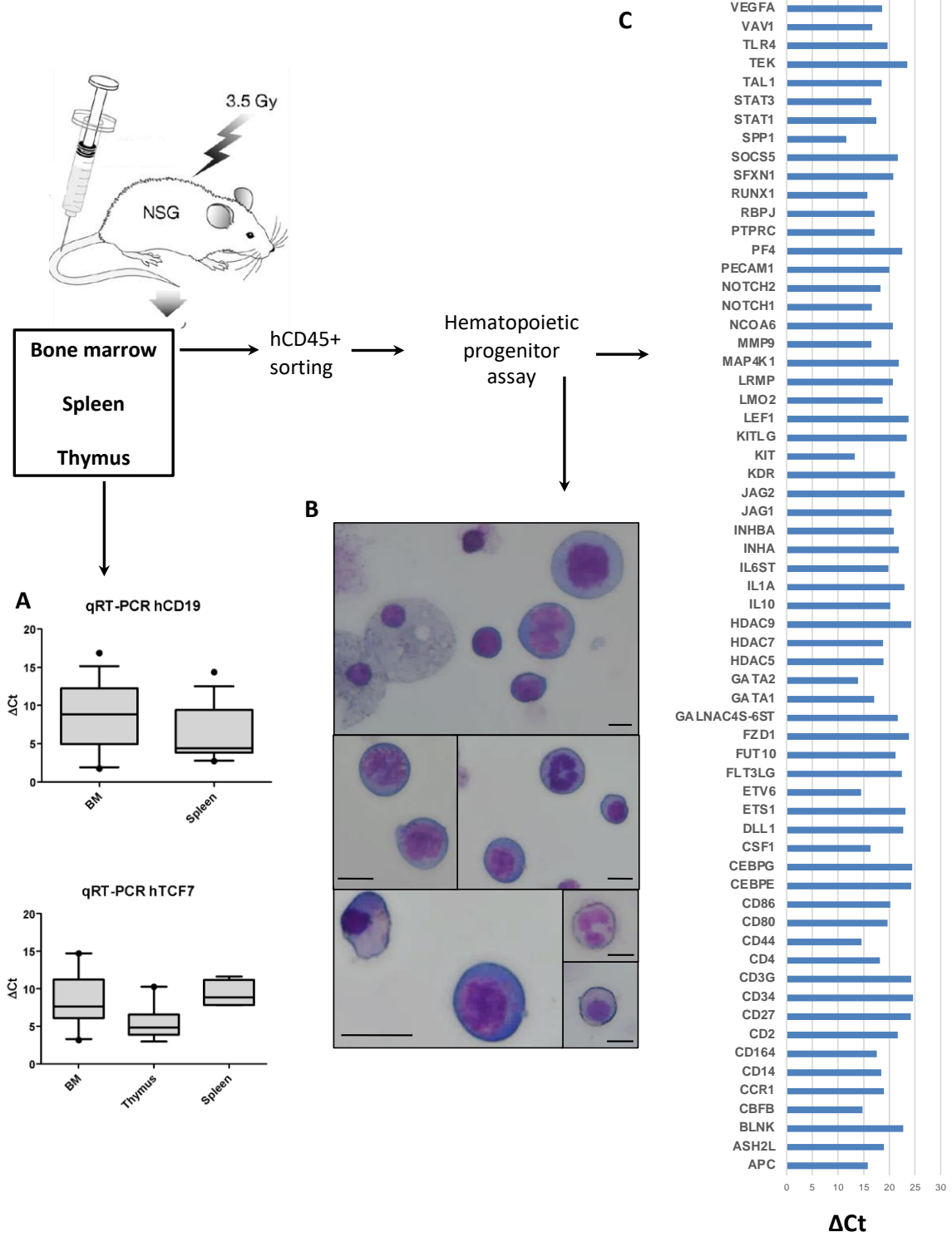


Figure S7

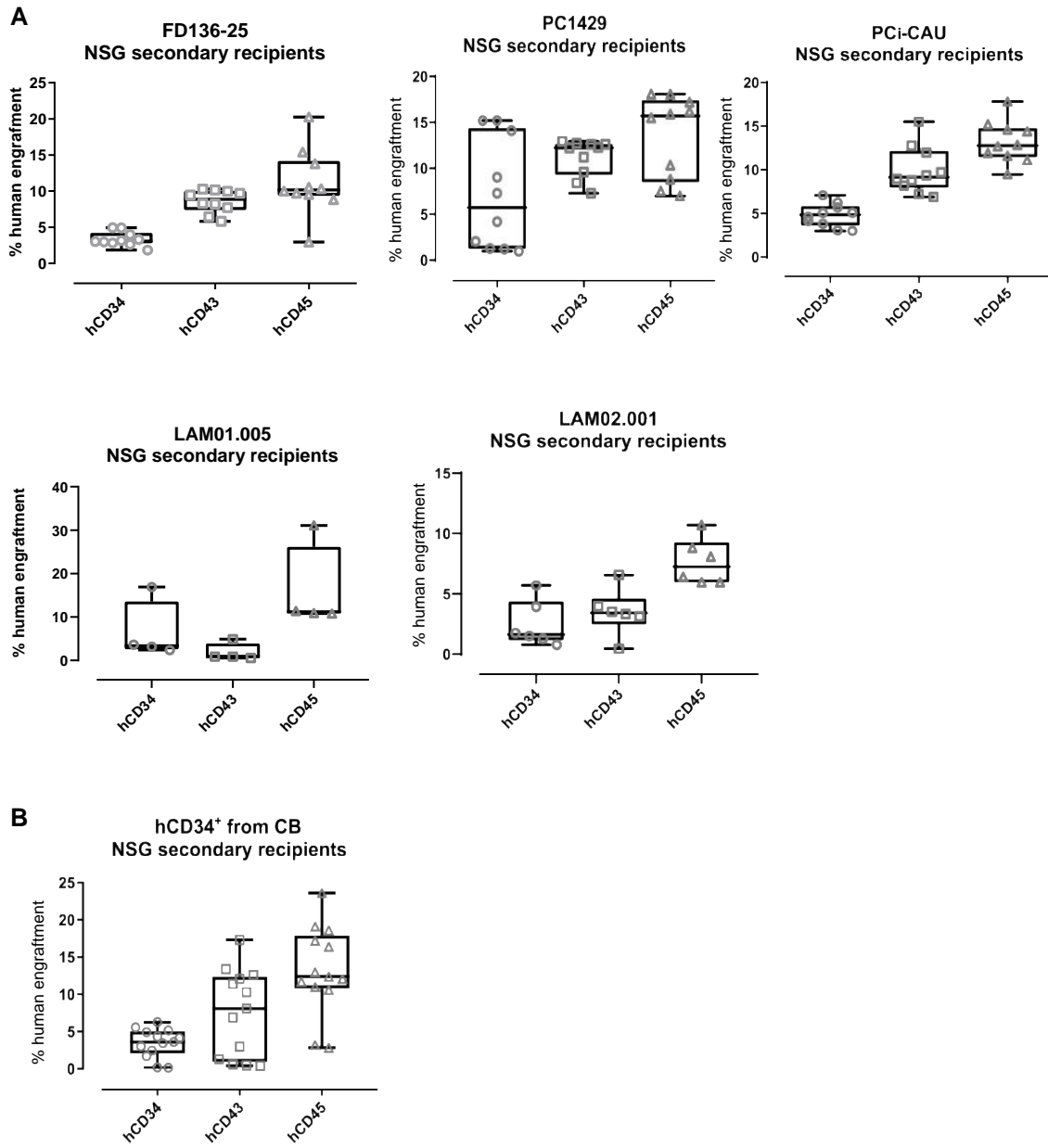


Figure S8

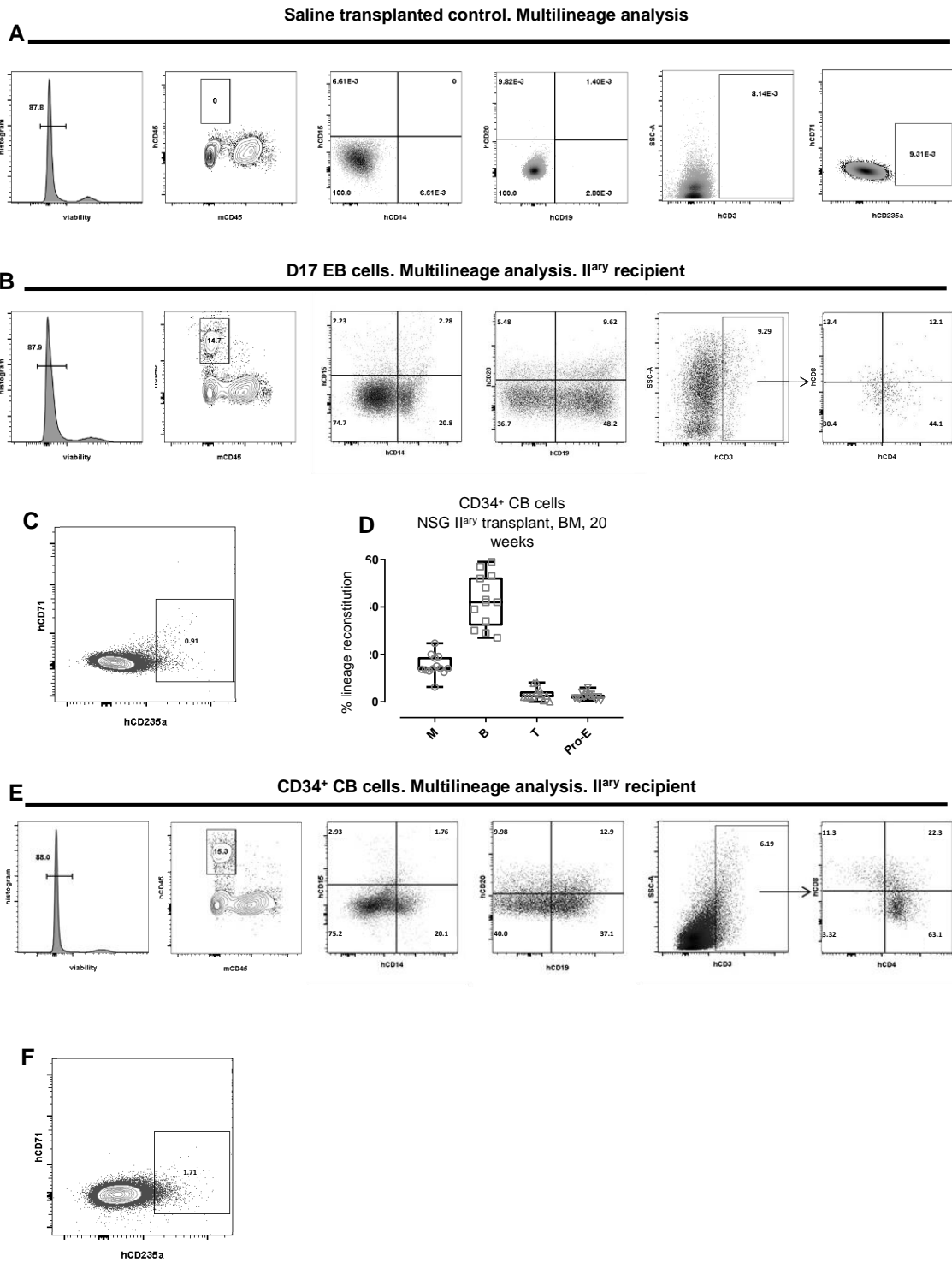


Figure S9

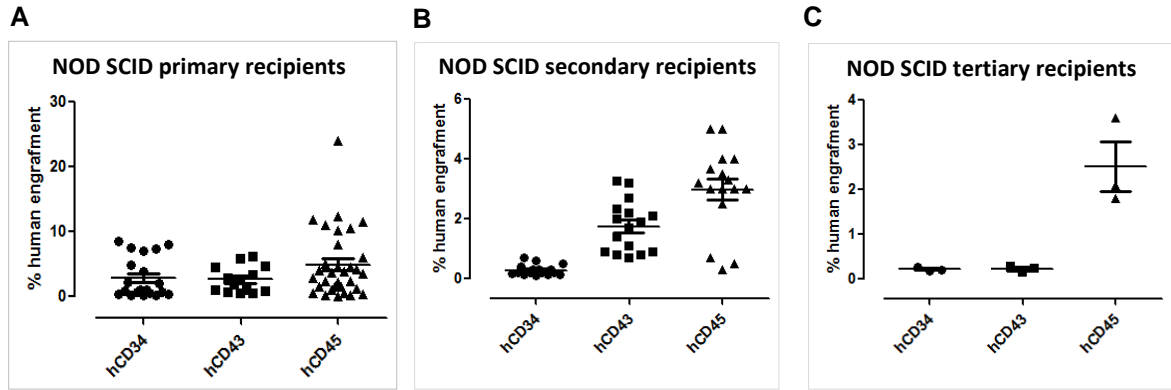


Figure S10

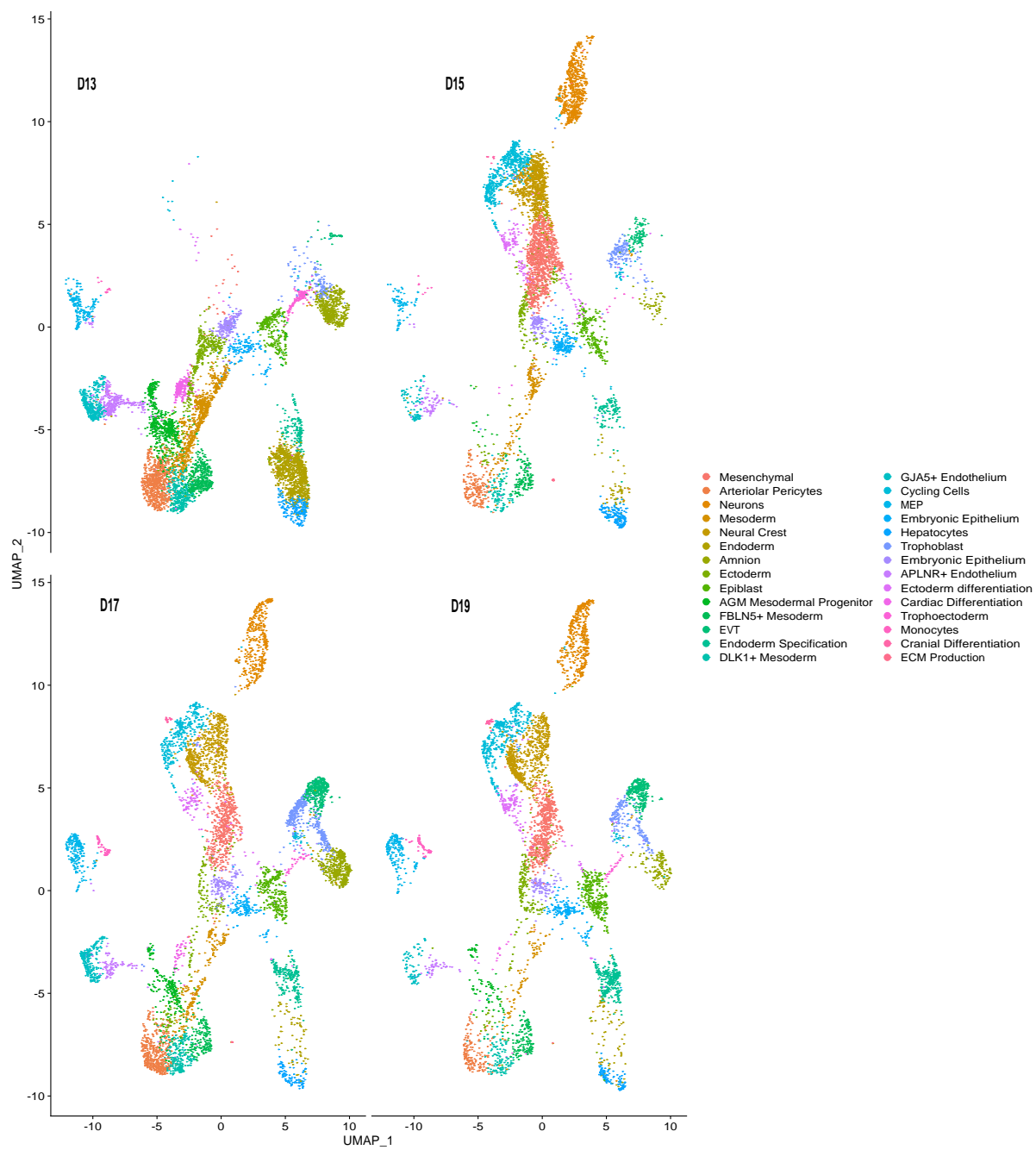
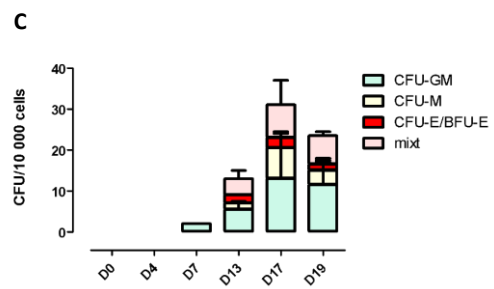
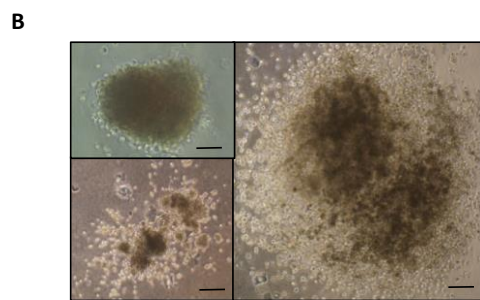
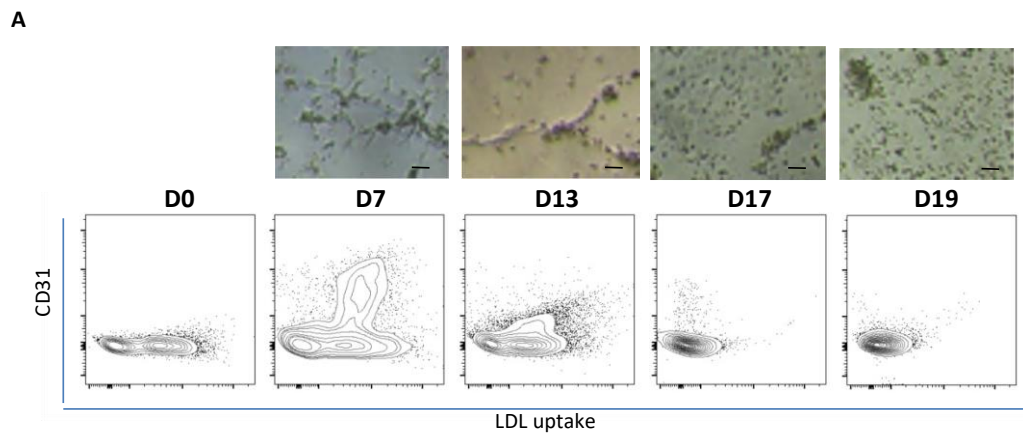
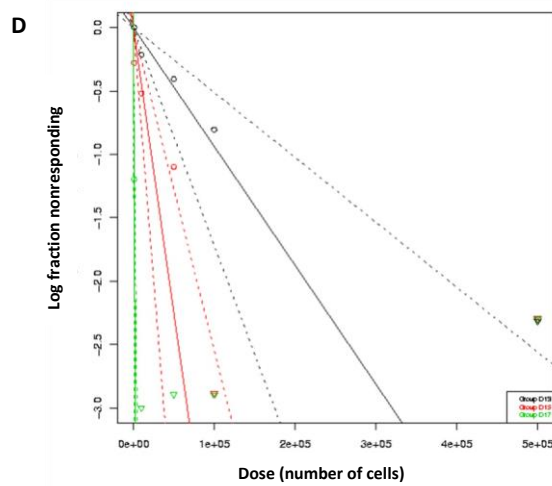


Figure S11

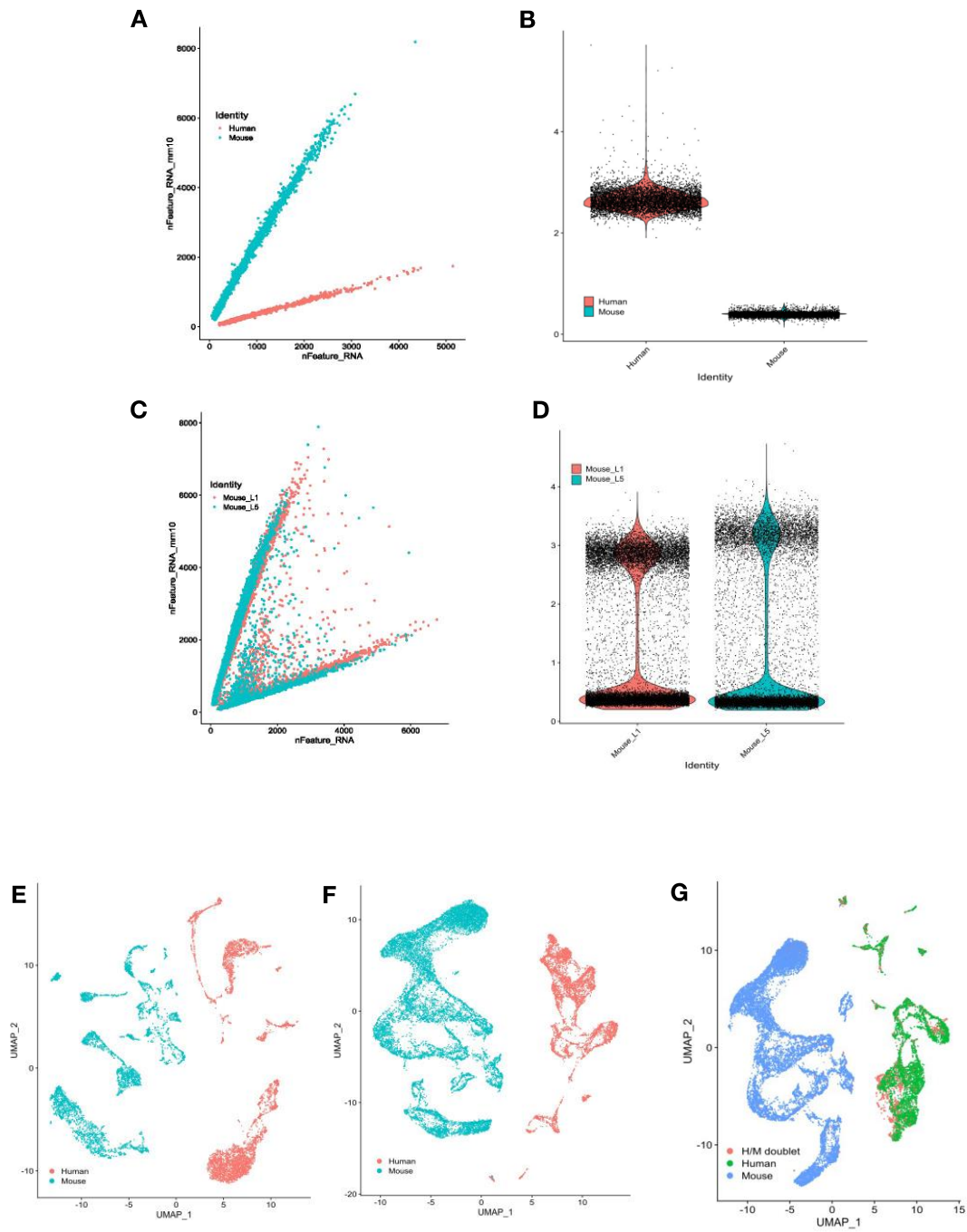


	D0	D7	D13	D17	D19
CFU/10 000 seeded cells	0	0.6 +/- 0.7	13 +/- 2.1	30.2 +/- 5.7	24.3 +/- 3.6
CFU/10 000 seeded CD34+ cells	0	1.8 +/- 2.1	82.7 +/- 13.4	203.8 +/- 38.4	198.2 +/- 29.4



Group	Lower	Estimate	Upper
D13	195415	106723	58285
D15	39318	22258	12600
D17	1448	831	476

Figure S12



SUPPLEMENTARY TABLES

Table S1

Factor (ng/mL) Combination N°	SCF	TPO	FLT3	BMP4	VEGF	IL3	IL6	IL1	GCSF	IGF1
1	300	100	20	20	200	5	50	50	10	50
2	300	20	300	200	200	50	50	50	10	5
3	20	20	300	20	200	5	5	5	10	5
4	300	20	20	200	20	5	50	5	100	5
5	20	100	20	20	200	50	50	50	100	5
6	20	20	20	200	200	5	5	50	100	5
7	20	100	300	20	20	5	50	5	10	5
8	300	20	20	20	200	50	5	5	10	50
9	300	100	20	20	20	50	5	5	100	5
10	20	20	20	200	20	50	50	5	10	50
11	300	100	300	200	20	50	5	50	10	5
12	300	100	300	200	200	5	5	5	100	50
13	20	20	300	20	20	50	5	50	100	50
14	20	100	300	200	200	50	50	5	100	50
15	300	20	300	20	20	5	50	50	100	50
16	20	100	20	200	20	5	5	50	10	50

Table S1. Cytokine combinations and cytokine concentrations tested for their ability to promote hematopoietic differentiation.

Table S2

Cell phenotypes	References
CD34+ CD109+	CD109 is expressed on a subpopulation of CD34+ cells enriched in hematopoietic stem and progenitor cells. DOI: 10.1016/s0301-472x(99)00071-5
CD34+ CD38 ^{low/-}	DO Sequential Generations of Hematopoietic Colonies Derived From Single Nonlineage-Committed CD34+CD38- Progenitor Cells. DOI: 10.1182/blood.V77.6.1218.1218
CD34+ HLA-DR+ CD38-	Lymphoid and myeloid differentiation of single human CD34+, HLA-DR+, CD38-hematopoietic stem cells. DOI: 10.1182/blood.V83.6.1515.1515
CD34+ CD133+	Large-scale isolation of CD133+ progenitor cells from G-CSF mobilized peripheral blood stem cells. DOI: 10.1038/sj.bmt.1703792 AC133, a Novel Marker for Human Hematopoietic Stem and Progenitor Cells. DOI: 10.1182/blood.V90.12.5002
CD34+ CD90+	A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. DOI: 10.1126/scitranslmed.aan1145 Enrichment of Human Hematopoietic Stem Cell Activity in the CD34+Thy-1+Lin- Subpopulation From Mobilized Peripheral Blood. DOI: 10.1182/blood.V85.2.368.368
CD34+CD38-CD117 ^{low}	Human primitive hematopoietic progenitor cells are more enriched in KIT ^{low} cells than in KIT ^{high} cells DOI: 10.1182/blood.V82.11.3283.3283
CD34+CD38+CD117+HLA-DR+CD13+CD33+	The number of CD34+CD38+CD117+HLA-DR+CD13+CD33+ cells indicates post-chemotherapy hematopoietic recovery in patients with acute myeloid leukemia. DOI : 10.1371/journal.pone.0180624

Table S2. References used to plot the variables for the DOE approach.

Table S3

Combination N	% of cells positive for:									LTC-ICs at D16	Amplification on 10 ⁶ Cells obtained /10 ⁶ iPSCs initially seeded
	CD90	CD109	CD13	CD117	CD38	HLA-DR	CD45	CD34	CD33		
1	2.39	0.1	0.07	0.44	0.1	0.06	0.41	0.94	0.1	11/1040000	3.01875
2	1.11	0.55	0.55	2.34	0.65	0.38	1.92	2.02	0.15	4/640000	1.7325
3	1.22	0.24	0.42	1.12	0.44	0.18	0.69	0.96	0.13	6/1390000	5.355
4	0.82	1.87	3.09	3.38	1.48	0.66	3.71	4.61	0.97	4/1340000	3.81675
5	1.77	1.97	2.66	5.24	2.9	1.01	3.58	5.48	2.2	6/1440000	6.23175
6	16.89	1.51	0.69	1.4	0.59	0.83	6.29	10.08	0.59	1/100000	0.462
7	3.44	0.28	0.15	1.05	0.31	0.24	0.51	1.14	0.23	1/1040000	3.78
8	0.92	1.54	2.71	3.47	1.39	0.62	2.62	4.27	0.63	7/1340000	3.92175
9	1.47	0.23	0.22	0.77	0.22	0.14	0.53	1.56	0.16	0	3.07125
10	5.22	0.48	0.48	0.68	0.75	0.2	3.99	1.48	0.14	0	1.6485
11	0.91	0.75	0.28	1.5	0.16	0.23	1.96	0.51	0.03	0	0.504
12	1.28	1.27	0.39	6.34	0.83	0.89	4.98	3.95	0.14	0	0.1785
13	1.78	0.34	0.17	0.88	0.19	0.11	0.35	1.45	0.31	3/90000	2.73
14	4.43	2.01	2.33	3.76	0.99	0.71	3.71	12.88	0.47	4/70000	0.336
15	1.91	0.45	0.39	0.82	0.19	0.2	0.8	1.8	0.28	9/1040000	3.07125
16	3.71	0.58	0.8	1.34	0.3	0.21	1.02	2.43	0.07	2/640000	1.87

Table S3. Matrix based on the DOE using the 16 combinations of cytokines and growth factors in terms of cell phenotype, cell amplification and ability to generate LTC-ICs.

Table S4

	SCF	TPO	FLT3	BMP4	VEGF	IL3	IL6	IL1	GCSF	IGF1
	(Concentrations in ng/mL)									
Combination A	24	21	21	194	200	50	50	5	100	5
Combination B	25	27	22	198	196	5	50	6	100	5
Combination C	22	20	300	22	200	50	50	5	100	50

Table S4. Matrix based on the three best combinations of cytokines and growth factors.

Table S5

Combination	Cells	% of CD34+	% of CD43+	% of CD45+	% of CD117+
A	FD136-25	3,11	4,25	2,24	4,97
	IMR90-16	7,49	10,24	4,17	9,47
B	FD136-25	5,14	4,16	1,6	6,15
	IMR90-16	4,15	3,83	3,02	6,42
C	FD136-25	6,89	4,58	3,99	7,02
	IMR90-16	8,53	3,42	15,72	2,82

Table S5 Percentages of CD34, CD43, CD45 and CD117 obtained with the three culture combinations.

Table S6**DOE**

N°	iPS	cytometry	qRT-PCR	functional ex vivo assays	functional in vivo assays	First recipient
1	FD136	Yes	Yes	No	No	No
2	IMR90	Yes	Yes	Yes	No	No
3	FD136	Yes	Yes	No	No	15 (NOD SCID)

Kinetic experiments

N°	iPS	cytometry	qRT-PCR	Immunofluorescence	scRNA seq	functional ex vivo assays
4	FD136	Yes	Yes	No	No	No
5	PC1429	Yes	Yes	No	No	Yes
6	PCi-CAU	Yes	Yes	No	No	Yes
7	PC1429	Yes	Yes	No	No	Yes
8	PCi-CAU	Yes	Yes	No	Yes	Yes
9	PCi-CAU	Yes	Yes	No	Yes	Yes
10	PC1429	Yes	Yes	Yes	No	No
11	PCi-CAU	Yes	Yes	Yes	No	No

Experiments with NOD SCID

N°	iPS	Cytometry qRT-PCR hematopoietic assays	First recipients	Cytometry qRT-PCR hematopoietic assays	Second recipients	Cytometry qRT-PCR hematopoietic assays	Third recipients	Cytometry qRT-PCR hematopoietic assays
12	FD136	Yes	4	Yes	0	No	0	No
13	FD136	Yes	8	Yes	8	Yes	0	No
14	FD136	Yes	8	Yes	8	Yes	3	Yes

Experiments with NSG

N°	iPS	Cytometry qRT-PCR hematopoietic assays	First recipient	Cytometry qRT-PCR hematopoietic assays	Second recipient	Cytometry qRT-PCR hematopoietic assays	scRNA seq
15	FD136	Yes	4	Yes	4	Yes	No
	PC1429	Yes	3	Yes	3	Yes	No
	PCi-CAU	Yes	3	Yes	3	Yes	No
	CT -	No	2	Yes	2	Yes	No
16	FD136	Yes	3	Yes	3	Yes	No
	PC1429	Yes	4	Yes	4	Yes	No
	PCi-CAU	Yes	3	Yes	3	Yes	No
	CT -	No	2	Yes	2	Yes	No
17	FD136	Yes	3	Yes	3	Yes	No
	PC1429	Yes	3	Yes	3	Yes	No
	PCi-CAU	Yes	4	Yes	4	Yes	No
	CT -	No	2	Yes	2	Yes	No
18	PC1429	Yes	30	Yes	0	No	No

LDA	CT -	No	3	Yes	0	No	No
19	LAM01.005	Yes	6	Yes	4	Yes	No
	LAM02.002	Yes	7	Yes	6	Yes	No
	CT -	No	2	Yes	2	Yes	No
20	PCi-CAU	Yes	16	Yes	0	No	No
	CT -	No	2	Yes	0	No	No
21 AB G	PCi-CAU	Yes	10	Yes	0	No	No
	CT -	No	2	Yes	0	Yes	No

Cord blood sorted CD34+

N°	cells	Cytometry qRT-PCR hematopoietic assays	First recipients	Cytometry qRT-PCR hematopoietic assays	Second recipients	Cytometry qRT-PCR hematopoietic assays
22	CD34+	Yes	4	Yes	4	Yes
23	CD34+	Yes	5	Yes	5	Yes
24	CD34+	Yes	4	Yes	4	Yes

Table S6 Table summarizing the different in vitro and in vivo experiments

Table S7

Number of cells	Number of recipients	Number of mice with >1% human cell chimerism
100 000	6	6
50 000	6	6
10 000	6	3
5 000	6	1
1 000	6	0

	Lower	Estimate	Upper
D17	31715	15700	7773

Table S7 Limiting dilution assay of D17 cells during primary transplantation.

Graded doses of one hundred thousand to one thousand D17 cells were transplanted into irradiated NSG mice, and the percentage of human CD45+ cells in the BM was analyzed 20 weeks after transplantation (n=6 mice per group). Confidence intervals of 1/ (stem cell frequency) were calculated using ELDA (<http://bioinf.wehi.edu.au/software/elda/>) according to the Poisson distribution. The limiting cord blood dilution assay published by Guo et al (Nat Med. 2018 Mar;24 (3):360-367. doi: 10.1038/nm.4477) must be used as a reference

Gene	Endothelial (EB)		Hematopoietic cells (EB)		Early AGM (CS10)
	APLNR+	GJA5+	HE	HSC candidate	HSC
RUNX1	High	High	High	High	High
HOXA9	High	High	Low	Low	High
MLLT3	High	High	Low	High	High
PTPRC	Low	Low	Low	High	High
SPN	High	High	High	High	High
THY1	High	High	High	Low	High
VNN2	Low	Low	Low	High	High
HLF	Low	High	Low	High	High
SPINK2	High	Low	Low	High	High
GFI1	High	Low	High	High	High
SELP	High	High	High	High	High
STAT5A	High	High	High	High	High
ITGA4	High	High	Low	High	High
SVOPL	Low	High	Low	High	High
EMCN	High	High	Low	Low	High
ACE	High	High	High	High	High
PROCR	High	High	High	High	High
HOXB9	High	High	High	High	High
LIN28B	High	High	High	High	High
CSF1R	Low	Low	Low	High	High
IL3RA	High	High	High	High	High
HLA-DRA	Low	Low	Low	Low	High
SELL	High	High	Low	High	High
MSI2	High	High	Low	High	High
HEMGN	Low	Low	Low	Low	High
PROM1	High	High	Low	Low	High
CDH5	High	High	High	High	High
IGFBP2	High	High	High	High	High
MEIS2	High	High	High	High	High
NRP2	High	High	High	High	High
SOX17	High	High	High	High	High
GJA5	High	High	High	High	High
IL33	Low	High	Low	Low	High
DKK1	Low	Low	Low	Low	High
AGTR2	High	High	Low	Low	High
ALDH1A1	High	High	Low	Low	High
CD44	High	High	High	High	High
KCNK17	Low	High	High	High	High
BCL11A	High	High	High	High	High
LIN28A	High	High	High	High	High
GAD1	High	High	High	High	High
FGF23	High	High	High	High	High

Table S8: Expression of the markers of the HSc ontogeny described in Calvanese et al., 2022 in the endothelial and hematopoietic cells of the EBs.

Analyses complémentaires du protocole de différenciation

Précisions sur les populations détectées en scRNAseq lors du protocole de différenciation en corps embryoides

Caractérisation des populations extra-embryonnaires

La présence de populations présentant un profil similaire extra-embryonnaire est fortement inhabituelle pour un système de différenciation en corps embryoides. Si la présence de BMP4 dans le milieu de culture est un facteur commun avec les protocoles produisant des cellules extra-embryonnaires à partir de cellules souches, la présence de telles cellules dans des corps embryoides au côté de cellules des trois feuillets embryonnaires ne semble pas encore avoir été reportée dans la littérature. Les tératomes issus de cellules souches pluripotentes injectées dans des souris, dont le profil de différenciation anarchique peut être rapproché des types cellulaires non synchronisés dans les corps embryoides, ne présentent pas de cellules trophoblastiques (McDonald et al., 2020).

Pour caractériser nos cellules, nous nous sommes basés sur l'atlas transcriptionnel publié par le laboratoire de Laurent David (Castel et al., 2020) pour les cellules de type trophoblastique, et sur les résultats d'Ogushi et al. (Ohgushi et al., 2022) pour les populations de type amniotique (Figure 9). On peut observer l'expression de TFAP2C et CDX2 au début de la branche des populations extra-embryonnaires, qui sont caractéristiques du trophoctoderme. Dans la branche de gauche, les gènes trophoblastiques GATA2, NR2F2 et LRP2 sont détectés, avec une différenciation en cytotrophoblaste extra-villeux marquée par l'expression de HLA-G. Enfin, une petite population, regroupée seule en haut à droite sur l'UMAP mais trop petite pour avoir été partitionnée par le clustering utilisé, exprime CGB7, un gène caractéristique du syncytiotrophoblaste. Pour vérifier ces identités, nous avons utilisé des modules de gènes marqueurs des populations trophoblastiques (Castel et al., 2020) qui ont confirmé l'identité de chacune de ces populations (Figure 9B). La branche de droite est VTCN1+HAND1+GABRP+, qui sont exprimés par les cellules de l'amnios décrites dans Ohgushi et al. (Ohgushi et al., 2022), mais pas par les populations trophoblastiques.

Afin de confirmer la présence des populations identifiées, nous avons réalisé des coupes des corps embryoides au cours des 17 jours de différenciation. Ces coupes ont été analysées par immunofluorescence grâce à des anticorps spécifiques des populations trophoblastiques (Figure 10). Pour identifier les cellules du trophoblaste extra-villeux, nous avons marqué l'expression de la protéine de surface HLA-G (Figure 10B). Nous avons utilisé l'expression de NR2F2 et CGB pour

marquer l'expression des cellules du syncytiotrophoblaste (Figure 10C, D). Il est possible de détecter des cellules NR2F2⁺CGB⁺ au 13^{ème} jour du protocole de différenciation, ainsi que des cellules HLA-G⁺ au 17^{ème} jour du protocole de différenciation, ce qui confirme les identités identifiées par les analyses de scRNAseq. Il semblerait que la population de type syncytiotrophoblaste soit de taille plus importante que la population identifiée par scRNAseq, mais il est probable que la majorité de la structure syncytiale n'ait pas été séquencée par le Chromium, puisque sa grande taille n'est pas compatible avec une encapsulation en gouttelettes.

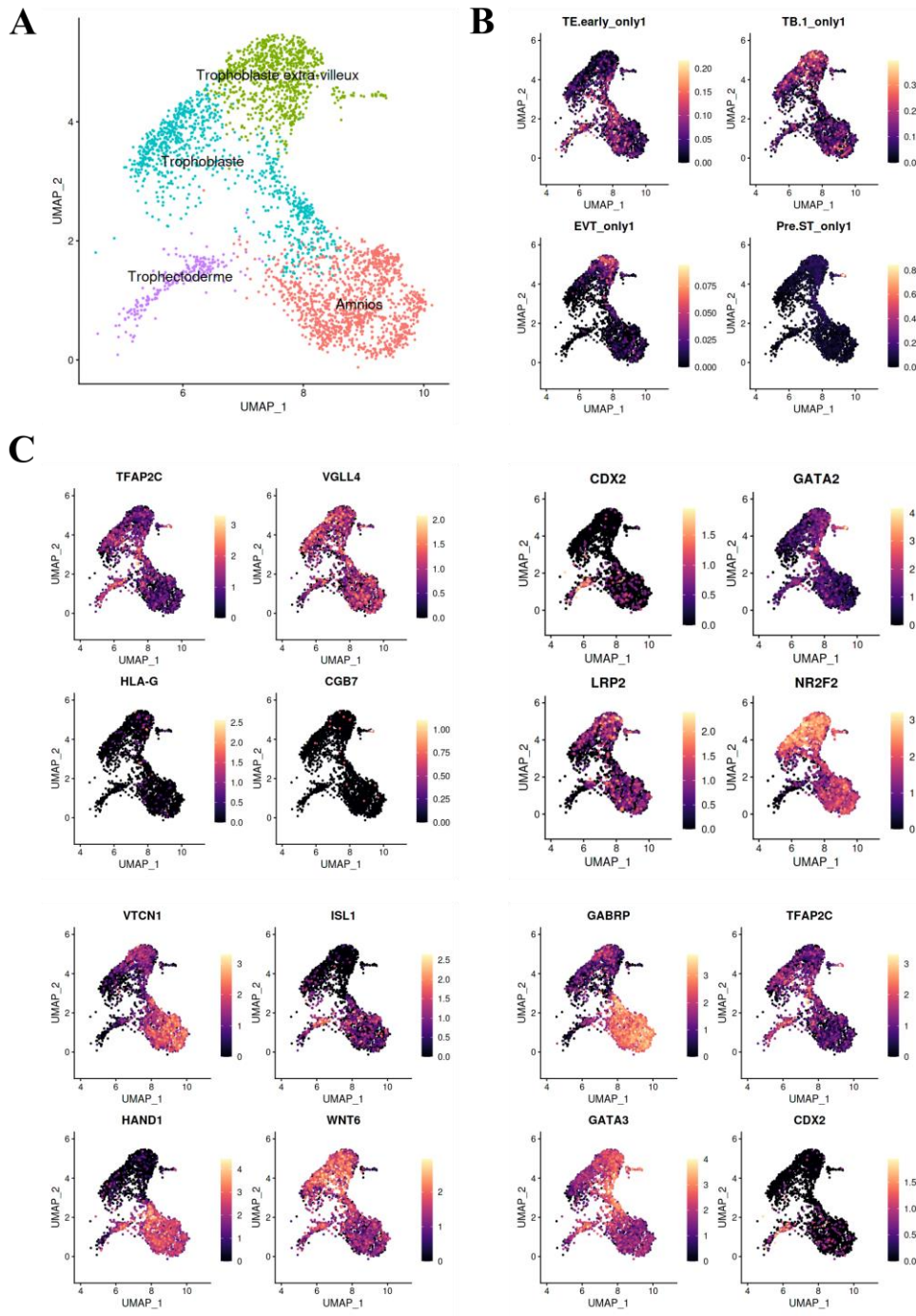


Figure 9 : Caractérisation des populations extra-embryonnaires dans les corps embryoides

A : Zoom sur les populations extra-embryonnaires de la carte cellulaire des populations présentes entre les jours 13 et 19 dans les corps embryoides.

B : Expression des modules de gènes décrits dans Castel et al. (Castel et al., 2020) dans les corps embryoides. Seuls les gènes présents dans un seul module ont été utilisés pour calculer le score d'expression.

C : Expression des gènes décrits dans Castel et al. (Castel et al., 2020) et Ohgushi et al. (Ohgushi et al., 2022) dans les corps embryoides.

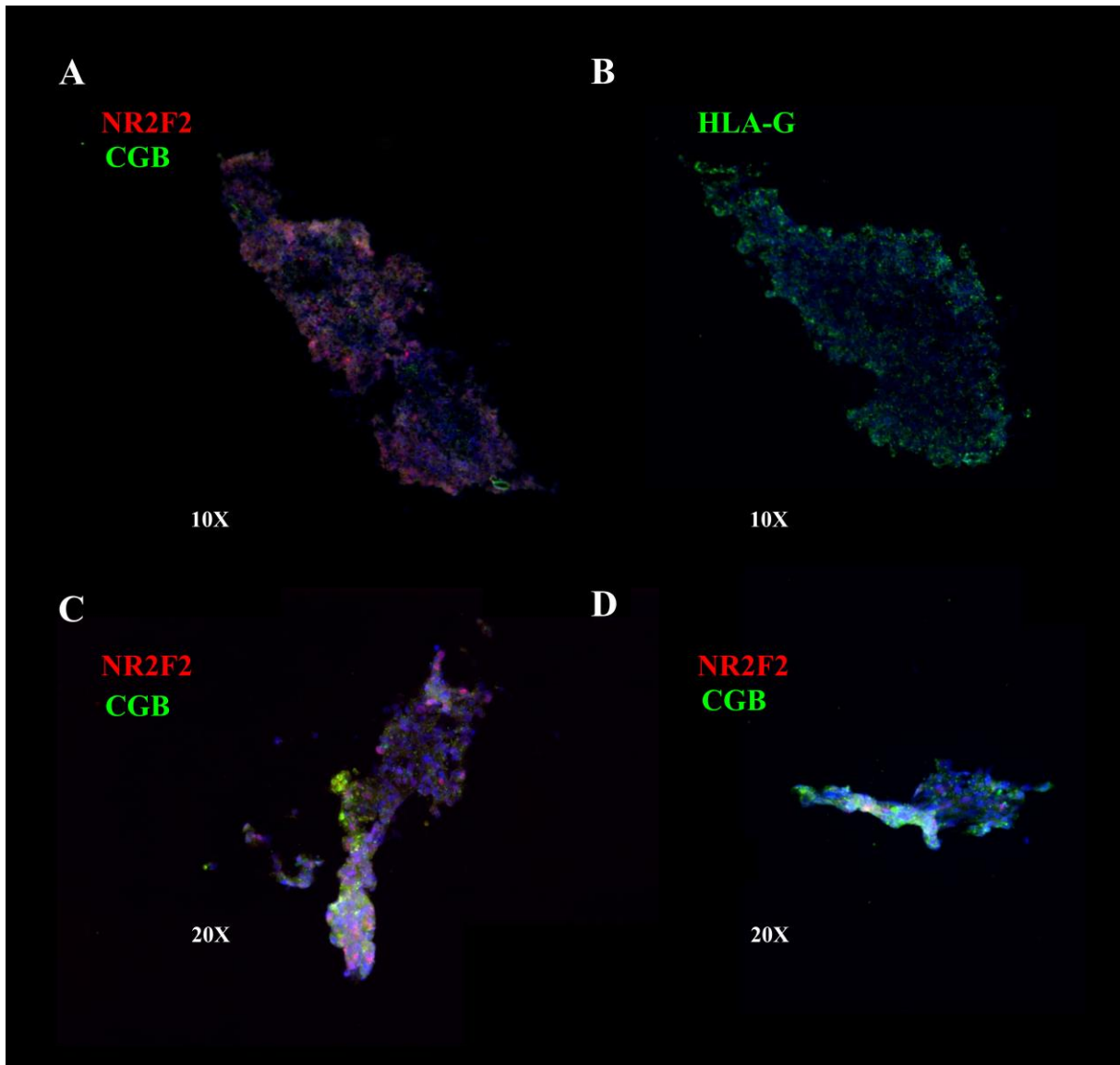


Figure 10 : Marquage par immunofluorescence des populations trophoblastiques dans les corps embryoïdes

A : Expression de NR2F2 et CGB dans un corps embryoïde au 17^{ème} jour de culture

B : Expression de HLA-G dans un corps embryoïde au 17^{ème} jour de culture

C-D : Expression de NR2F2 et CGB dans des corps embryoïdes au 13^{ème} jour de culture

Temporalité de l'apparition des différentes populations

L'apparition des différentes populations décrites dans l'analyse des corps embryoïdes entre les jours 13 et 19 ne se fait pas de manière homogène. En s'appuyant sur des jeux de données scRNAseq réalisés aux 4^{ème} et 7^{ème} jours de différenciation (Figure 11A, C), il est possible d'ordonner l'apparition des différents feuillet. Au 4^{ème} jour de différenciation, les cellules sont majoritairement indifférenciées, il est cependant possible d'identifier le début de la différenciation endodermique.

Au 7^{ème} jour, les populations endodermiques et mésodermiques sont identifiables. La fraction mésodermique est alors composée de cellules stromales et endothéliales. Les populations endothéliales identifiées sont encore relativement précoces et ne présentent pas de phénotype artériel puisqu'elles expriment encore le récepteur à l'apeline (APLNR). La différenciation vers les populations extra-embryonnaires semble s'amorcer puisqu'il est possible d'observer une population CDX2⁺ positive bourgeonnant des populations ayant un profil épiblastique.

Au 13^{ème} jour, les populations endothéliales au profil artériel et les cellules hématopoïétiques commencent à apparaître. La fraction mésenchymateuse du mésoderme est plus diversifiée, notamment avec l'apparition des péricytes. La population endothéliale colocalisée avec les cellules hématopoïétiques, labellisée comme endothélium hémogénique dans le manuscrit de l'article retranscrit plus haut, est majoritairement détectée à ce temps de culture. Les populations aux profils amniotiques et syncytiotrophoblastiques sont également apparues.

Au 15^{ème} jour, on peut voir l'apparition des cellules aux profils ectodermiques et cytotrophoblastique extra-villeux. Les cellules syncytiotrophoblastiques, en revanche, ne sont plus détectées. La fraction endodermique se réduit avec l'apparition des cellules les plus matures exprimant des gènes caractéristiques de la lignée hépatocytaire.

Les 17^{ème} et 19^{ème} jours présentent des populations similaires au 15^{ème} jour. Une analyse des cellules présentes après 24 jours de culture (Figure 11C) montre cependant que la majorité des types cellulaires ont disparu, puisque seuls subsistent des cellules des lignages mésodermiques et endodermiques.

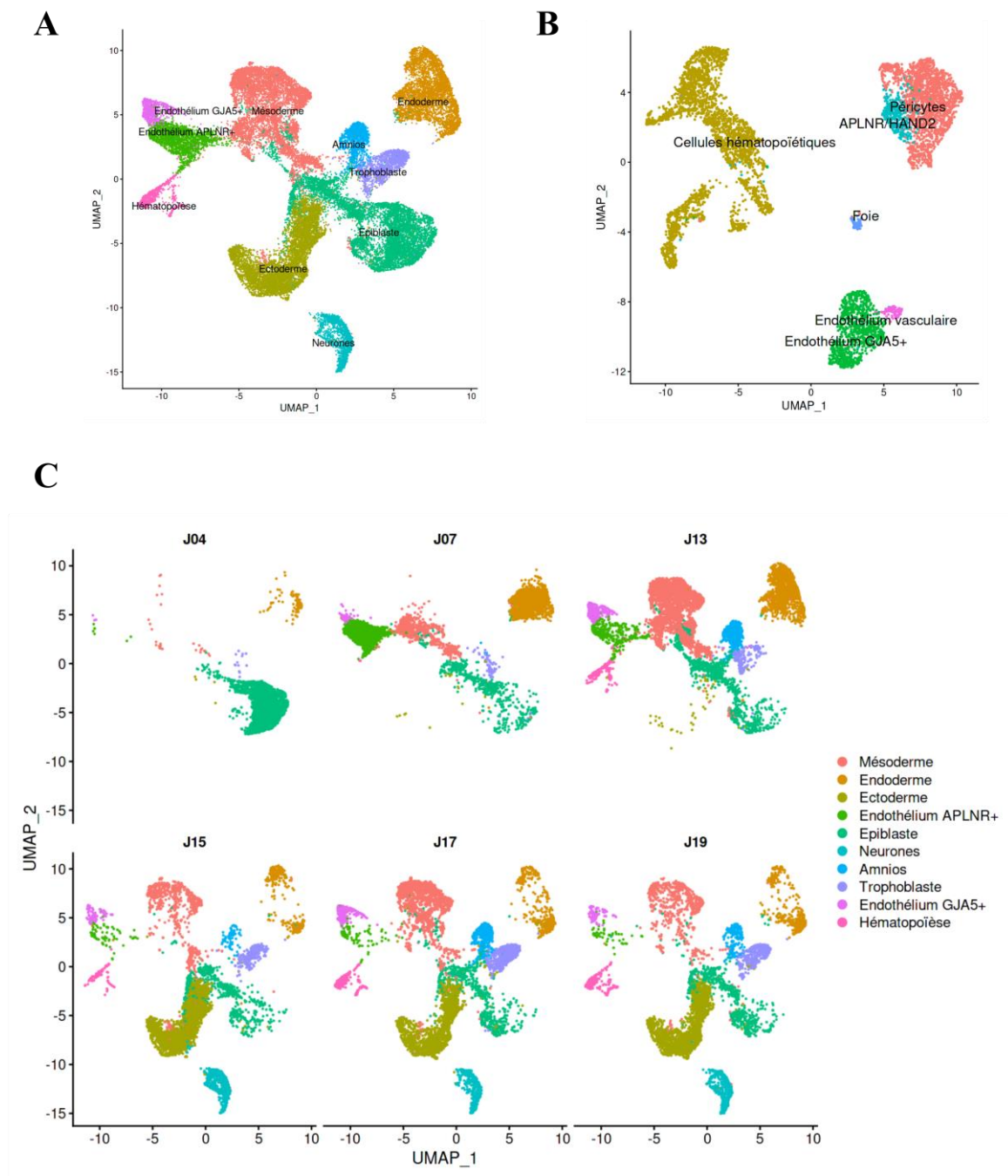


Figure 11 : Evolution des populations dans les corps embryoides au cours du protocole de différenciation

A : Populations détectées dans les corps embryoides entre les 4^{ème} et 19^{ème} jours de culture

B : Populations détectées dans les corps embryoides après 24 jours de culture

C : Populations détectées dans les corps embryoides entre les 4^{ème} et 19^{ème} jours de culture, affichées par jour de culture

Influence des sources de protéine sur les types cellulaires produits

Notre protocole de différenciation utilise comme source de protéine un plasma produit par l'établissement français du sang (EFS). Ce plasma n'étant pas disponible commercialement pour les laboratoires de recherche ne travaillant pas avec l'EFS, nous avons voulu comparer les résultats de notre protocole avec d'autres sources de protéine. Nous avons ainsi testé le plasma produit par l'entreprise Octapharma. L'utilisation de ce plasma a un effet fortement délétère sur notre protocole de culture, caractérisé par une formation diminuée du nombre de corps embryoides formés et une forte mortalité (Figure 12A). Nous avons donc effectué des expériences de scRNAseq à 17 jours de différenciation à l'aide du plasma EFS, du plasma Octapharma et de l'albumine de sérum bovin (BSA pour *bovine serum albumin*) afin de comparer l'impact des variations de source de protéine sur les populations cellulaires produites.

Comme identifié précédemment, les corps embryoides produits avec le plasma EFS sont composés de cellules des trois feuillets embryonnaires et de cellules extra-embryonnaires après 17 jours de différenciation (Figure 12B). En revanche, les corps embryoides cultivés pendant 17 jours avec le plasma Octapharma ou la BSA ne produisent que des cellules mésodermiques (mésenchymateuses, endothéliales ou hématopoïétiques) et endodermiques (foie fœtal).

Pour confirmer ces résultats et vérifier que l'absence de populations extra-embryonnaires au 17^{ème} jour n'est pas due à une accélération de la différenciation des corps embryoides, nous avons réalisé des marquages par immunofluorescence avec NR2F2 et CGB aux jours 7, 13 et 17 des corps embryoides cultivés avec les deux sources de plasma (Figure 13). Aucune trace de différenciation trophoblastique n'est observée dans les corps embryoides cultivés avec le plasma Octapharma aux trois temps examinés (Figure 13A). En revanche, l'utilisation d'un marquage CD144, CD34, CD45 permet de montrer l'apparition des populations endothéliales et hématopoïétiques dans les deux conditions (Figure 13B).

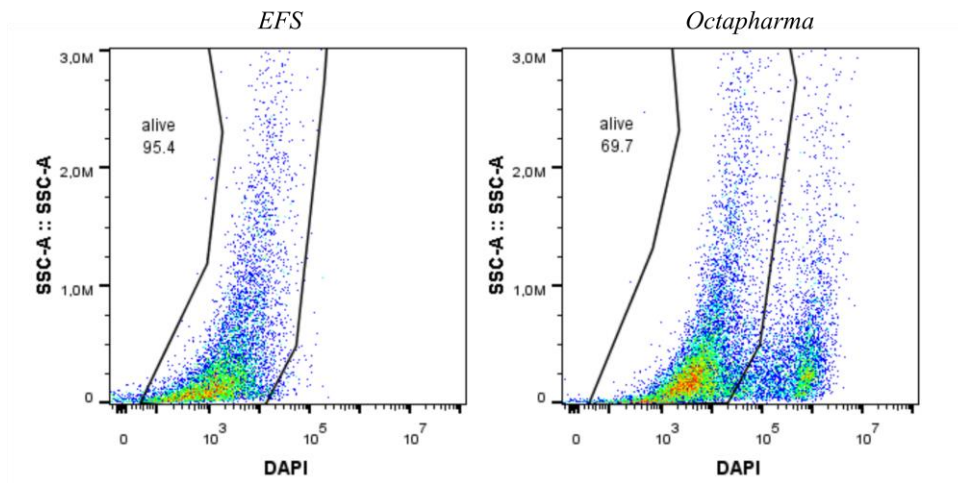
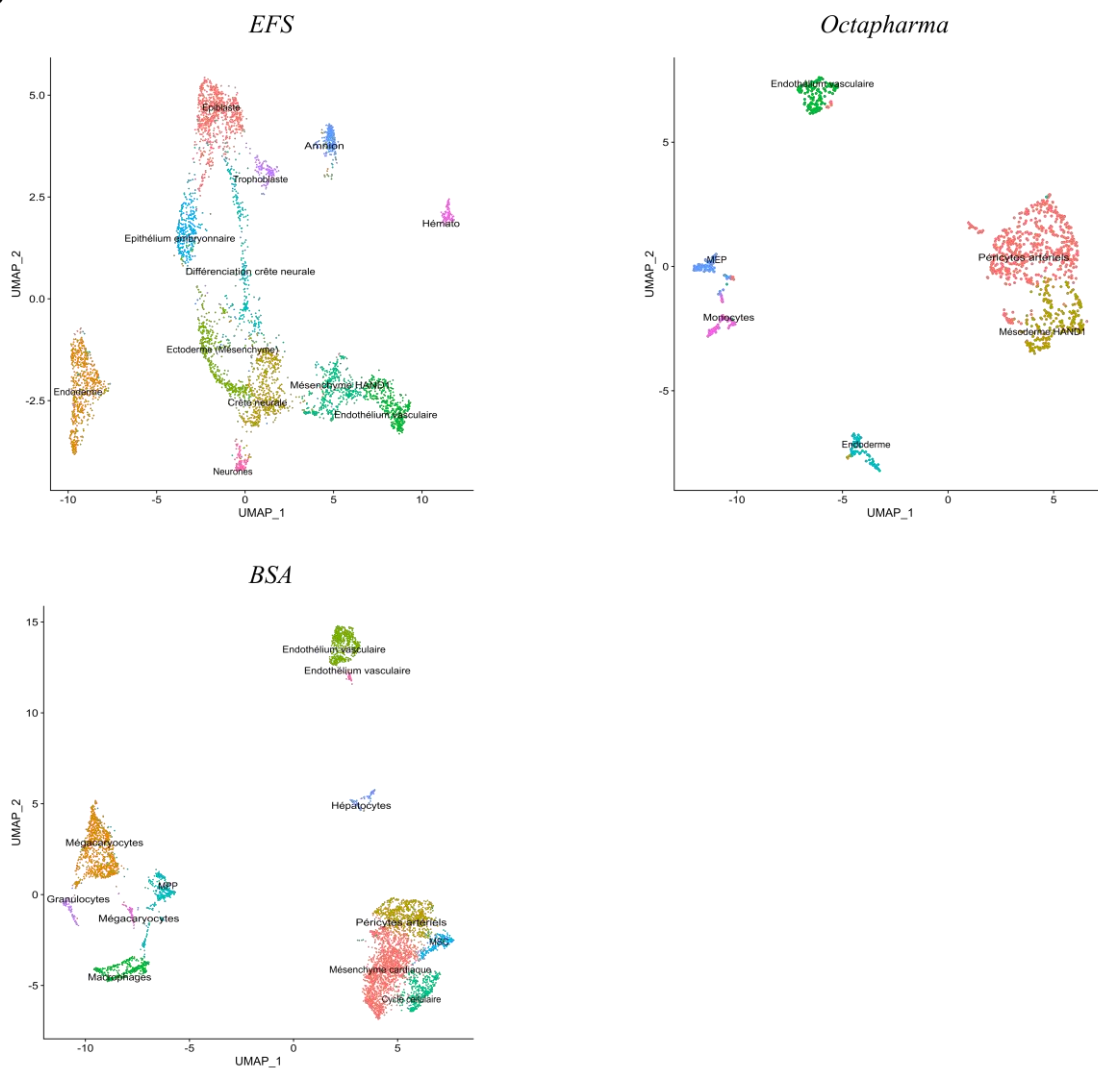
A**B**

Figure 12 : Influence de la source protéique sur les populations produites dans les corps embryoides
 A : Détection des cellules mortes par marquage au DAPI en cytométrie en flux dans des corps embryoides de 17 jours cultivés avec les plasmas EFS et Octapharma.
 B : Carte cellulaire des analyses scRNAseq réalisées sur des corps embryoides après 17 jours de culture avec du plasma EFS, Octapharma, ou de la BSA.

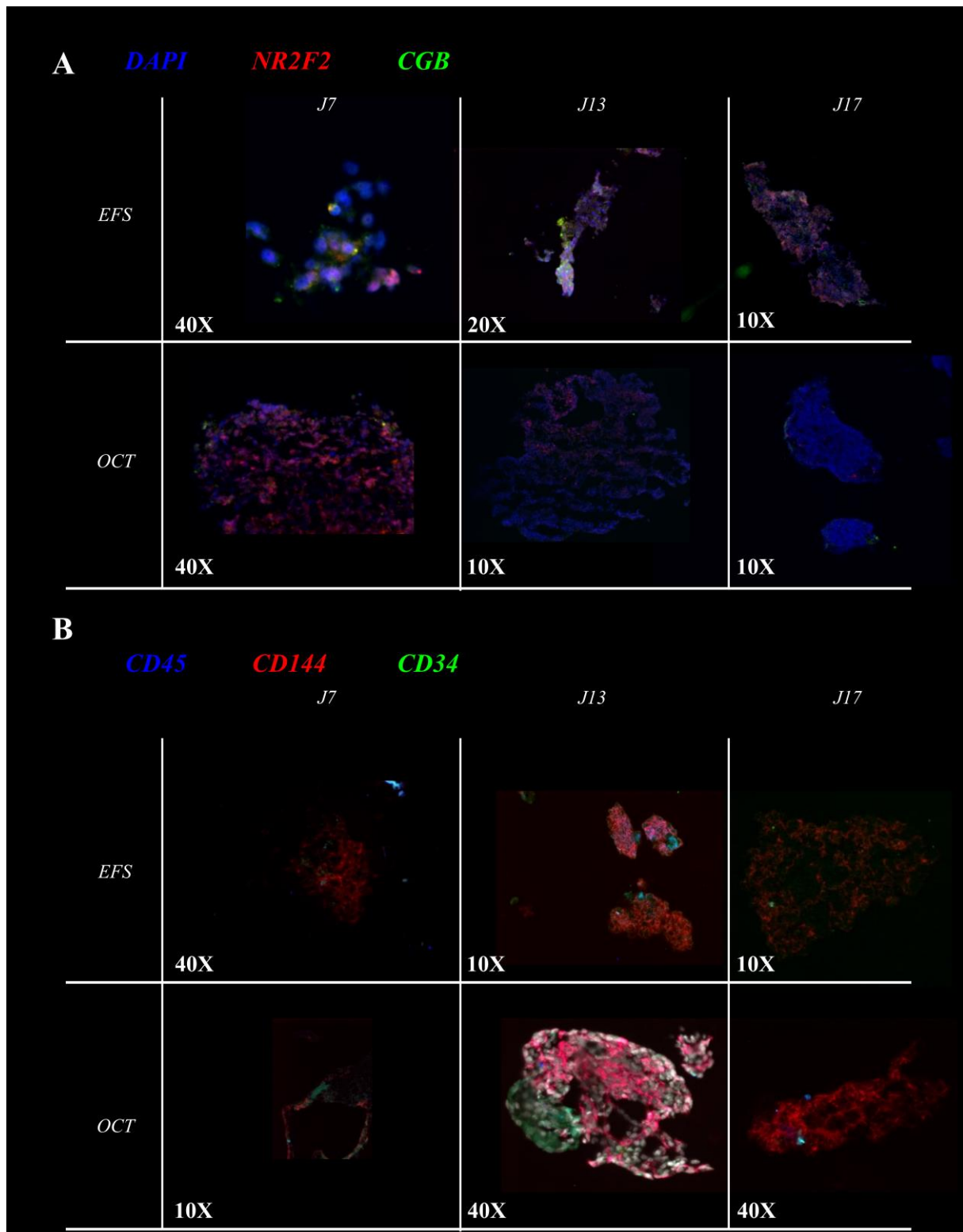


Figure 13 : Validation par immunofluorescence de l'influence des sources de plasma sur les populations cellulaires produites

A : Expression de NR2F2 et CBG dans des corps embryoides cultivés avec les plasmas EFS et Octapharma pendant 7,13 et 17 jours

B : Expression de CD45, CD144 et CD34 dans des corps embryoides cultivés avec les plasmas EFS et Octapharma pendant 7,13 et 17 jours

Nouvelles approches bioinformatiques explorées lors des analyses

Utilisation de Gene Ontology à l'échelle de la cellule unique

Une des étapes particulièrement fastidieuse de l'analyse des données de scRNAseq est l'annotation des types cellulaires. Lors de l'analyse des données sur les corps embryoïdes, les bases de données permettant une annotation automatique étaient encore peu nombreuses et ne couvraient pas forcément tous les types cellulaires possibles, ce qui demande de vérifier les identités cellulaires obtenues à l'aide de marqueurs qu'il faut extraire de la littérature. L'utilisation de stratégies d'annotation manuelle à base des gènes différentiellement exprimés dans les clusters peut également s'avérer laborieuse car les différenciations *in vitro* produisent des types cellulaires peu matures dont le profil transcriptomique est moins bien caractérisé que pour les types cellulaires adultes. Il peut cependant être assez facile pour un expert de regarder les profils d'expressions de certains gènes caractéristiques de voies de différenciation dans son domaine d'expertise ou de repérer certains gènes clés dans une liste de gènes différentiellement exprimés pour estimer rapidement l'identité des cellules. Dans le cas de système de différenciation aussi complexe que nos corps embryoïdes qui produisent des cellules des trois feuilletts embryonnaires et des cellules extra-embryonnaires, peu de personnes disposent d'une expertise suffisante.

Afin de pouvoir accélérer ce processus d'annotation, je me suis intéressé à la possibilité d'utiliser une base de données tel que la *Gene Ontology* à l'échelle des cellules uniques pour faciliter ces identifications (Figure 14). L'approche développée a été la suivante : donner la liste des gènes détectés dans chaque cellule à un algorithme d'enrichissement en sets de gènes (GSEA) pour obtenir une valeur p (*p-value*) mesurant la probabilité d'expression de chaque module de gène dans chaque cellule. Cette *p-value* est transformée en score par transformation logarithmique pour obtenir une matrice de modules de gènes par cellule. On peut ensuite utiliser cette matrice pour voir quelles modules de gènes sont surreprésentés dans les clusters ou pour visualiser leur expression dans le jeu de données.

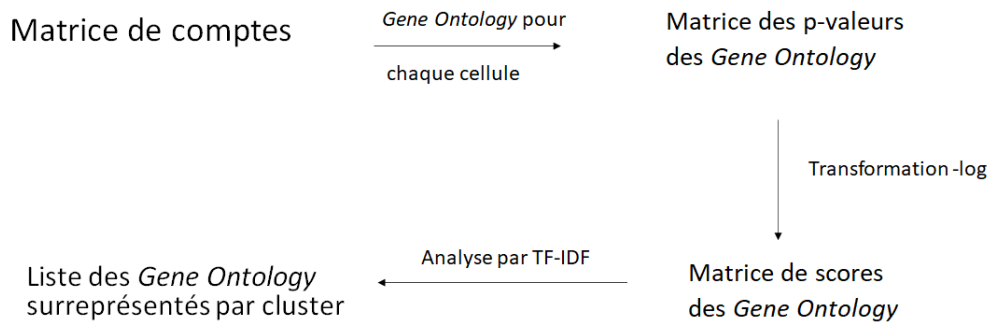


Figure 14 : Stratégie d'analyse par les Gene Ontology à l'échelle de la cellule unique

Dans notre cas, cette approche a été utile pour identifier la présence de cellules extra-embryonnaires car nous éprouvions beaucoup de doutes sur les résultats des annotations automatiques pour ces cellules. Bien qu'elle soit efficace pour identifier les voies de différenciation dans plusieurs lignages, elle est fortement limitée par l'exhaustivité de la base de données utilisée. Ainsi, les modules de gènes obtenus ressortant comme différenciellement exprimés pour les cellules épiblastiques ou similaires aux étapes précoces du développement de l'embryon ne permettent pas de déterminer l'identité des cellules. Une version beaucoup plus élaborée de cette méthode, se basant sur des réseaux de neurones apprenant à détecter l'expression des modules de gènes, a depuis été mise au point par le laboratoire de Fabian Theis et a été publiée en 2023 (Lotfollahi et al., 2023).

Séparation des cellules murines et humaines dans les moelles de souris

Lors de l'analyse des moelles de souris greffées par scRNAseq, il a été nécessaire de pouvoir séparer les cellules humaines des cellules murines dans le jeu de données obtenu. En effet, aucun tri n'avait été effectué sur les cellules de la moelle avant séquençage afin de capturer la totalité de l'hétérogénéité des populations cellulaires présentes. Les génomes humains et murins étant proches, les séquences de beaucoup d'ARNs murins s'alignent sur les séquences de leurs homologues humains et réciproquement. Bien que les génomes murins et humains possèdent plusieurs différences, comme la spécificité du gène HLA-DR à l'Homme, la nécessité que le gène soit exprimé et détecté rend difficile l'utilisation de gènes marqueurs pour ségréger de manière fiable les cellules humaines et murines.

On peut faire l'hypothèse que le nombre de gènes détectés lors de l'alignement des séquences sur le génome d'une autre espèce est plus faible que dans le cas de l'alignement sur le génome correspondant à la bonne espèce. Cependant, du fait de la variabilité naturellement présente dans le nombre de gènes ou d'UMI détectés, aucune ségrégation des populations nette n'apparaît si l'on s'intéresse au nombre de gènes détectés par cellule en réalisant un alignement sur le génome

humain. Pour circonvenir à ce problème, nous avons aligné les échantillons sur les génomes humains et murins et avons comparé pour chaque cellule le nombre de gènes détectés sur chaque génome (Figure 15A). On peut alors voir apparaître une ségrégation de deux populations, une population ayant plus de gènes humains détectés que de gènes murins et une autre ayant plus de gènes murins détectés. On peut également observer la présence de quelques cellules présentant un profil d'expression équilibré, qui peuvent correspondre à des types cellulaires particuliers ou à des doublets composés d'une cellule humaine et d'une cellule murine.

Afin de disposer d'un contrôle pour vérifier la fiabilité de cette méthode de ségrégation, nous avons pris deux jeux de données publiques, un de moelle humaine et un de moelle murine, que nous avons fusionné afin de vérifier si notre méthode sépare bien les populations par espèce d'origine. L'utilisation du ratio entre le nombre de gènes humains et murins détectés permet bien de séparer les cellules des deux espèces et aucun type cellulaire ne semble présenter le profil intermédiaire observé dans les moelles des souris greffées (Figure 15B, C), ce qui suggère que ces cellules sont des doublets. Nous avons donc choisi de retirer ces cellules de nos jeux de données pour le reste de nos analyses.

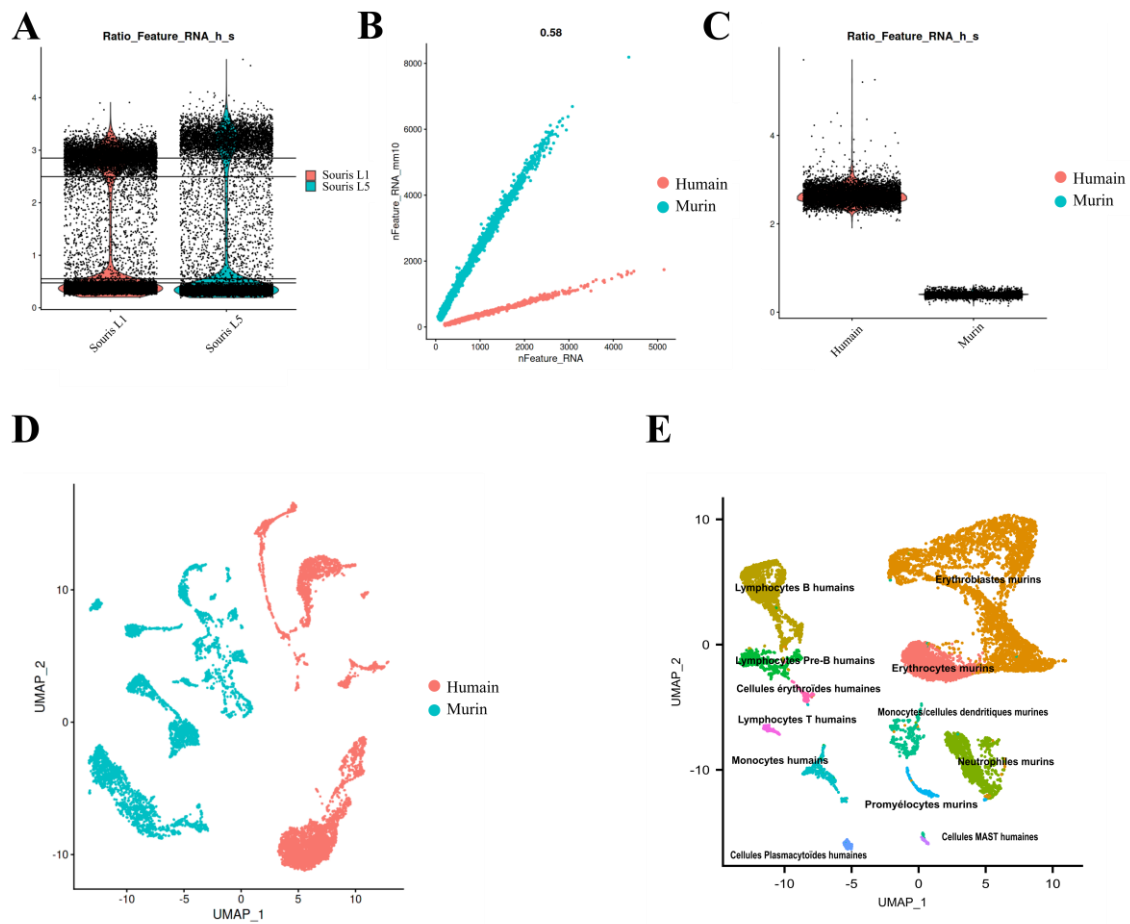


Figure 15 : Séparation des cellules humaines et murines dans l'analyse scRNAseq des moelles greffées

A : Expression du ratio du nombre de gènes humains détectés par rapport au nombre de gènes murins dans les moelles greffées

B : Nombre de gènes humains et murins détectés dans le jeu de données contrôle

C : Expression du ratio du nombre de gènes humains détectés par rapport au nombre de gènes murins dans le jeu de données contrôle

D : Carte cellulaire du jeu de données contrôle

E : Carte cellulaire des moelles de souris greffées

Analyse d'un protocole d'amplification sériée des cellules souches hématopoïétiques

Dans le cadre d'une montée en échelle du protocole de différenciation de cellules souches pluripotentes en cellules souches hématopoïétiques avec injection dans modèle gros animal (le cochon nain d'Aix-la-Chapelle) (Plotzki et al., 2016), il nous a été nécessaire d'établir un contrôle positif. Ainsi, notre laboratoire a été amené à amplifier des cellules souches hématopoïétiques de sangs de cordons humains, qui représentent le test de référence pour les greffes de cellules souches hématopoïétiques. En effet, la plupart des sangs de cordons à notre disposition ne contenaient pas assez de cellules souches hématopoïétiques pour un système porcin. En s'inspirant d'un protocole d'amplification élaboré au sein du laboratoire pour réaliser des greffes de cellules souches hématopoïétiques dans des primates non-humains, une nouvelle méthode de culture à base de cytokines (SCF, TPO, FLT3-L, IL3, IL6, GM-CSF) a été mise au point. Bien que l'amplification obtenue au bout de 10 jours de culture ait été suffisante, nous nous sommes intéressés à la possibilité d'améliorer le rendement de cette production par une amplification réalisée de manière séquentielle. Avec cette nouvelle méthode, les cellules CD34+ obtenues après les 10 jours d'amplification sont à nouveau mises en culture pendant 10 jours dans le but d'être amplifiées de manière répétée (Figure 16). Afin de pouvoir mesurer et quantifier le potentiel des cellules souches hématopoïétiques ayant été soumises à ces amplification sériées, les cellules obtenues à chaque étape sont caractérisées par cytométrie en flux, par des tests clonogéniques, par la capacité à effectuer une différenciation érythroïde, par greffe dans des souris immunodéficientes irradiées et enfin par scRNAseq (Figure 16). Ma contribution à ce projet a consisté en l'analyse des jeux de données scRNAseq effectuées aux jours 0, 10 et 20 du protocole.

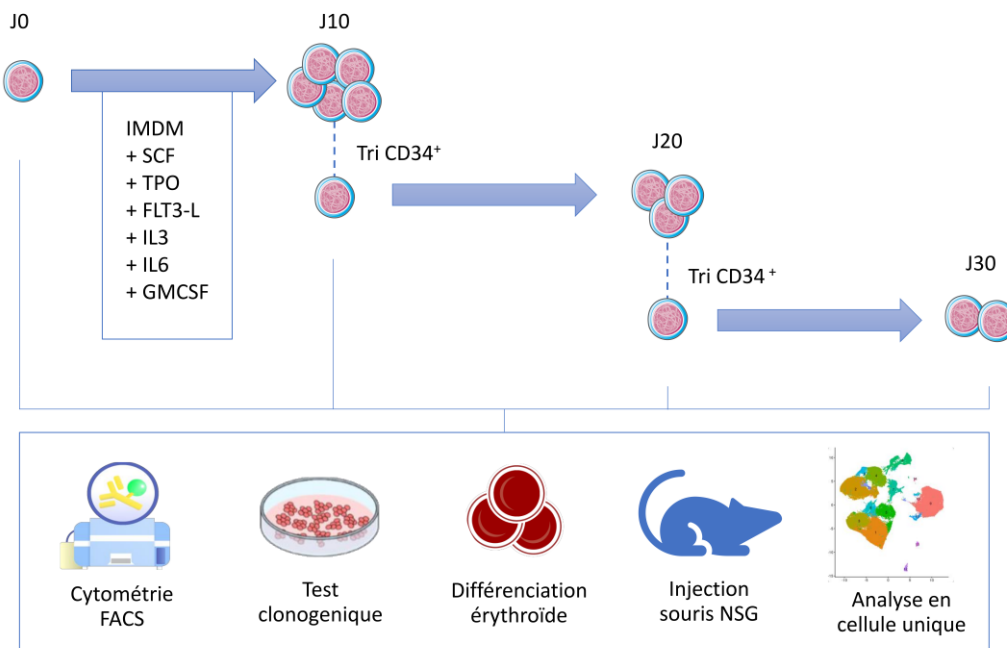


Figure 16 : Protocole d'amplification des cellules souches hématopoïétiques et caractérisation des cellules obtenues

Réalisée par Mathias Brunet-Manquat

Evaluation du potentiel des cellules souches hématopoïétiques au cours de la culture

La quantification du nombre de cellules CD34⁺ obtenues par cytométrie en flux montre une amplification d'un facteur 15 lors des 10 premiers jours de culture, suivie d'une amplification d'un facteur 4 entre le 10^{ème} et le 20^{ème} jour de culture (Figure 17A), permettant ainsi une amplification d'un facteur 60 par rapport aux cellules de départ (Figure 17B). Malheureusement, l'amplification entre le 20^{ème} et le 30^{ème} jour de culture est très faible et ne permet pas d'obtenir une différence significative dans la quantité de cellules générées.

Cette baisse de la capacité d'amplification au cours de la culture est corroborée par les tests clonogéniques, puisque le nombre de colonies CFUGM/G généré par cellule baisse avec le temps de culture (Figure 17C). On peut également observer une baisse de la capacité d'énucléation des cellules générées en protocole de maturation et différenciation érythroïde (Figure 17D).

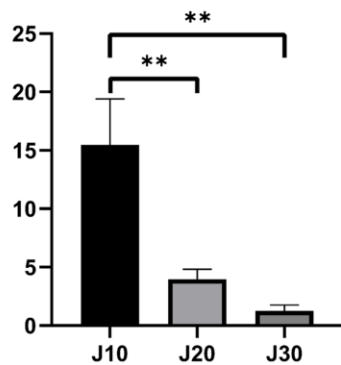
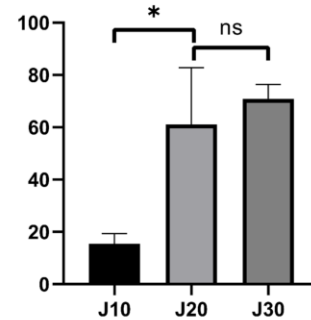
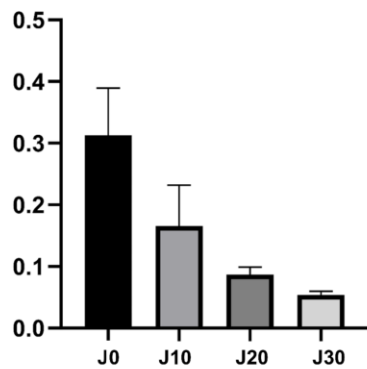
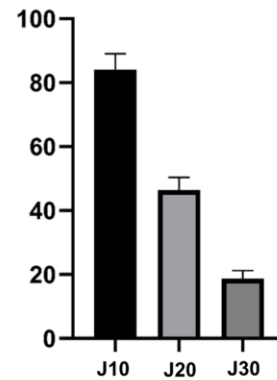
A**B****C****D**

Figure 17 : Caractérisation fonctionnelle des cellules après amplification

A : Facteur d'amplification des cellules CD34⁺ à chaque itération du protocole

B : Facteur d'amplification cumulée des cellules CD34⁺ du protocole d'amplification à chaque itération du protocole

C : Nombre de CFUGM/G produit par cellule CD34⁺ à chaque itération du protocole

D : Pourcentage d'énucléation des globules rouges produits en différenciation érythroïde à chaque itération du protocole

Analyse scRNAseq des cellules produites

Les tests clonogéniques et les tests de différenciation érythroïde ont montré une baisse du caractère souche des cellules CD34⁺ avec le protocole de culture. L'objectif des analyses scRNAseq réalisées ici visent à identifier quels changements sont observés dans les cellules cultivées afin de pouvoir raffiner le protocole de culture.

Pour ce faire, les jeux de données des cellules CD34⁺ obtenues aux jours 0, 10 et 20 avec trois sangs de cordons ont été agrégés (Figure 18A). On peut observer une ségrégation forte des populations par jour de culture, et l'apparition d'une ségrégation des échantillons par sang de cordon après le

10^{ème} jour de culture. Un des échantillons de sangs de cordon au 20^{ème} jour de culture a malheureusement connu des problèmes techniques lors de sa préparation et n'a donc pas pu être inclus dans l'analyse.

Afin d'identifier les populations possédant le plus fort caractère souche, nous nous sommes basé sur les modules de gènes publiés par le laboratoire de Lars Steinmetz (Velten et al., 2017). Nous avons pu détecter une expression forte du module HLF-ZFP36L2, présent habituellement au sommet de la hiérarchie de différenciation hématopoïétique, dans une sous-population de chaque échantillon (Figure 18B). Néanmoins, la procédure classique de clustering des échantillons s'avèrent peu efficace pour extraire ces populations et les scores d'expression des modules de gènes présentent une expression trop stochastique pour pouvoir sélectionner de manière fiable les populations désirées. Une approche alternative a donc été développée afin d'avoir un contrôle plus fin sur les cellules sélectionnées.

Pour les cellules séquencées après 10 ou 20 jours de culture, nous avons réalisé une trajectoire de différenciation à l'aide de Monocle3 (Figure 18C). Cette dernière nous a permis de filtrer les cellules sur leur position dans la trajectoire, offrant ainsi un contrôle plus fin des cellules sélectionnées (Figure 18E). Nous avons ainsi pu isoler 1614 cellules au 10^{ème} jour de différenciation et 1559 au 20^{ème} jour de différenciation.

L'expression des modules de gènes dans les cellules du jour 0 (non cultivées) est plus disparate que pour les cellules cultivées. Une approche basée sur la modélisation de trajectoire n'est donc pas adaptée dans ce cas. Pour résoudre ce problème, nous avons réalisé un clustering réalisé sur un graphe des plus proches voisins construit en utilisant seulement les gènes présents dans les modules de gènes les plus souches décrits dans Velten et al. (Velten et al., 2017). Ainsi, nous avons pu obtenir un cluster constitué majoritairement de cellules présentant un niveau élevé d'expression du module HLF-ZFP36L2 (Figure 18D) que nous avons utilisé pour la suite des analyses.

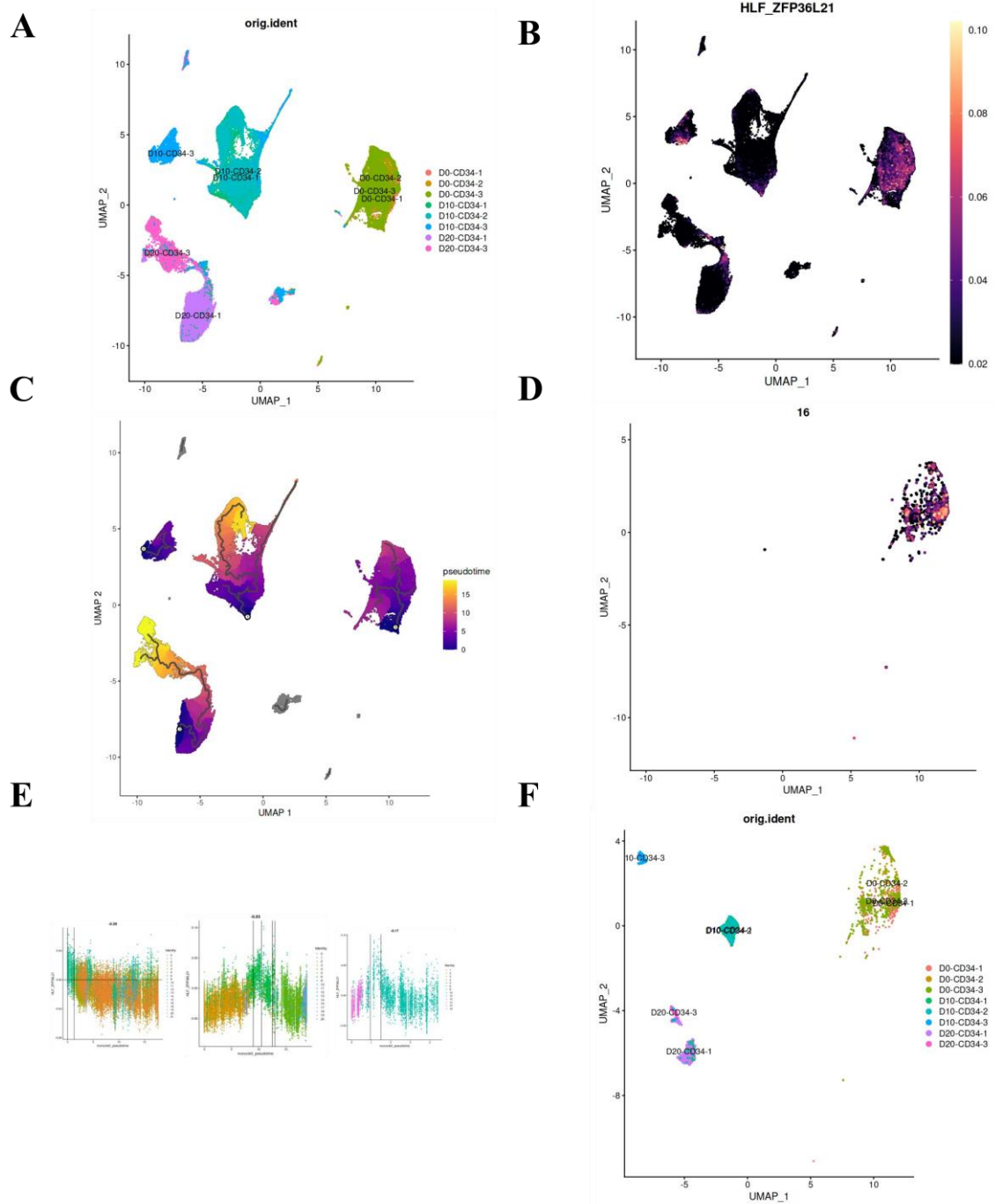


Figure 18 : Sélection des cellules souches hématopoïétiques

A : Carte cellulaire des échantillons de sangs e cordons séquencés après tri des cellules CD34⁺, colorée par échantillon

B : Expression du module HLF-ZFP36L2 dans les cellules séquencées

C : Trajectoires cellulaires réalisées avec Monocle3 afin d'extraire les populations souches

D : Cluster des cellules non cultivés (J0) sélectionné sur la base de l'expression du module HLF-ZFP36L2, coloré par niveau d'expression du module HLF-ZFP36L2.

E : Sélection des populations les plus souches en utilisant les trajectoires fournies par Monocle3, les cellules de chaque module sont ordonnées par pseudotime

F : Projection des populations souches sélectionnées sur la carte cellulaire des échantillons

Une fois les cellules les plus souches sélectionnées (Figure 18F), nous avons pu comparer l'expression des gènes exprimés entre les cellules non cultivées et les cellules cultivées au 10^{ème} jour afin d'évaluer les effets de la culture *ex vivo*, mais également les différences entre les 10^{ème} et 20^{ème} jours de culture afin d'identifier les changements transcriptomiques sous-jacents à la perte progressive de potentiel des cellules cultivées. On peut observer la surexpression de gènes liés à l'inflammation et le stress oxydatif après 10 jours de culture. Ces cellules expriment l'antigène de surface CD48 qui est documenté comme étant absent des cellules souches hématopoïétiques quiescentes murines et nécessaire à la prolifération, ce qui est cohérent avec la mise en culture et l'utilisation d'un protocole d'amplification. Il est également possible d'observer une surexpression d'IL18 au bout de 10 jours de culture. IL18 a été reporté comme favorisant la quiescence des cellules souches hématopoïétiques (Howard et al., 2021), ce qui suggère la mise en place d'une balance entre les voies de quiescence et d'amplification en réaction au protocole.

Entre le 10^{ème} et le 20^{ème} jour de culture, on peut observer la surexpression de gènes liés à la réponse aux corticostéroïdes et aux glucocorticoïdes. Une augmentation de l'expression de gènes liés à ces voies suite à la mise en culture de cellules souches hématopoïétique a été documenté dans un article publié en janvier 2023 (Dong et al., 2023) qui a défini une liste de gènes liés à la perte du caractère souche suite à la mise en culture (CT-SRGs pour *culture-time-stemness related genes*). Nous avons également observé une surexpression de FOSB et JUNB entre le 10^{ème} et le 20^{ème} jour de culture, ce qui suggère une possible évolution de la régulation médiée par le facteur de transcription AP-1 lors de la seconde phase d'amplification (Figure 19A). Cependant, si une augmentation des gènes CT-SRG est bien observée suite aux 10 premiers jours de culture (Figure 19B), le niveau d'expression des gènes du CT-SRG reste stable les 10 jours suivants (Figure 19C). Il semble donc que la réponse identifiée entre les jours 10 et 20 du protocole d'amplification corresponde à une réponse non caractérisée par les CT-SRG, bien que faisant également parti de la réponse aux corticostéroïdes et aux glucocorticoïdes.

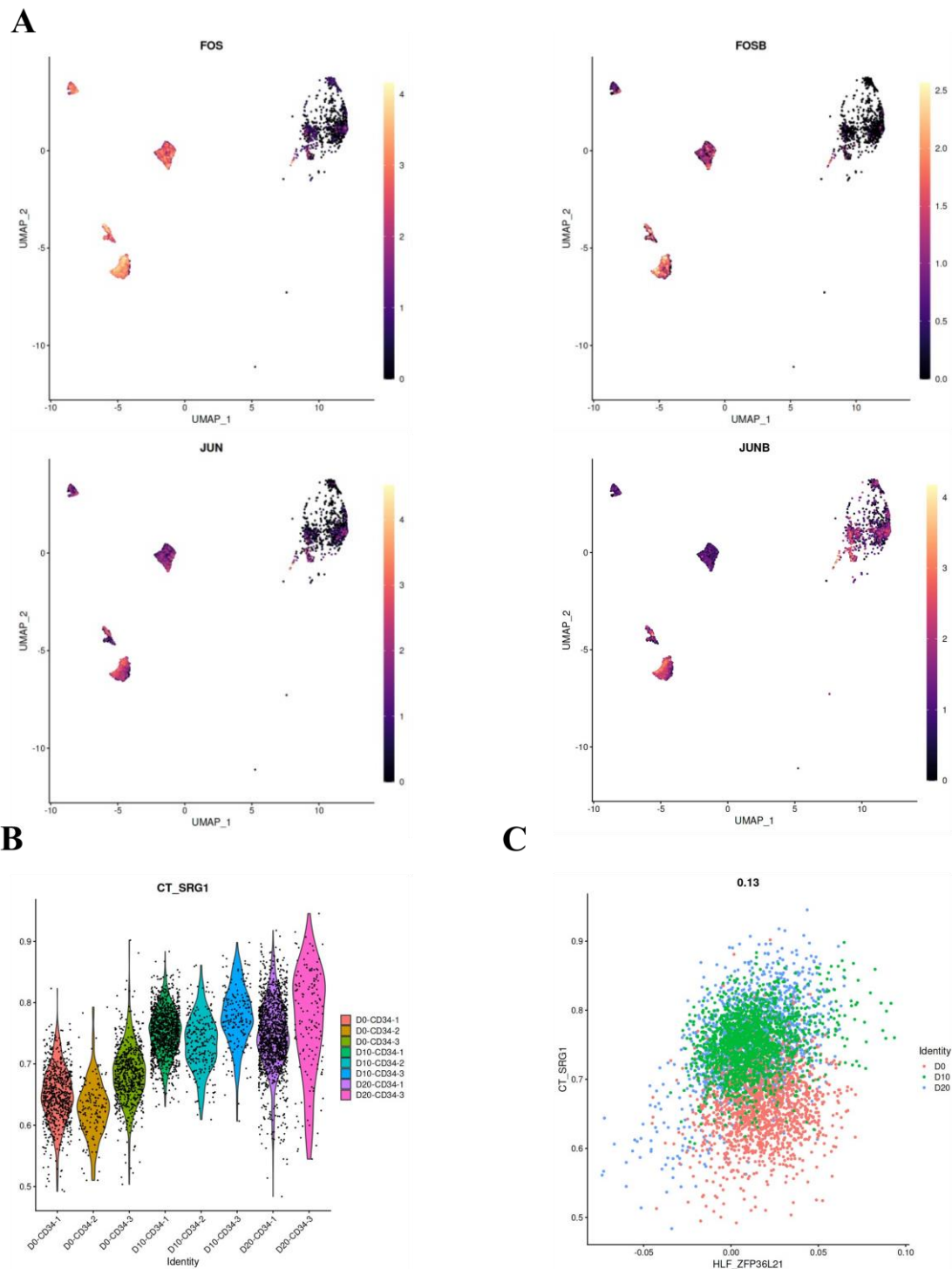


Figure 19 : **Modifications transcriptomiques des cellules souches hématopoïétiques lors du protocole d'amplification**

A : Expression de FOS, FOSB, JUN et JUNB dans les cellules souches hématopoïétiques

B : Expression du module des CT-SRG en fonction du temps de culture

C : Expression du module CT-SRG et du module HLF-ZFP36L2 dans les cellules souches hématopoïétiques

Analyse du microenvironnement murin de l'AGM

Ce manuscrit décrit une caractérisation des micro-environnements ventraux et dorsaux de l'aorte embryonnaire à E11.5. Par des approches transcriptomiques globales ou à l'échelle de la cellule unique, il a ainsi été possible de mettre en évidence une population stromale exprimant des gènes caractéristiques de fonctions neurales. Le rôle de cette population, qui ne provient pas des crêtes neurales, a ensuite été validée fonctionnellement par l'étude de l'importance de la décorine et la ténacine, deux protéines extra-cellulaires, dans la formation des cellules souches hématopoïétiques. Ma contribution à ce projet a consisté en l'analyse d'un jeu de données scRNAseq d'AGM de souris séquencée à E11.5 qui a permis de confirmer l'existence de cette population neuro-mésenchymateuse initialement identifiée par des analyses de RNAseq.

A neuro-mesenchymal cell subpopulation characterizes the subaortic hematopoietic stem cell niche

Pierre-Yves Canto^{1*}, Olivera Miladinovic^{1*}, Claire Pouget², Olivier Piau^{1,3}, Nevenka Radic¹, Priscilla Freschu¹, Alexandre Megherbi¹, Sébastien Jacques⁴, Estelle Hirsinger¹, Audrey Geeverding⁵, Sylvie Dufour⁶, Michèle Souyri⁷, Hervé Isambert⁸, David Traver², Thierry Jaffredo^{1**}, Pierre Charbord^{1**} and Charles Durand^{1**}

¹ Sorbonne Université, CNRS, Inserm U1156, Institut de Biologie Paris Seine, Laboratoire de Biologie du Développement/UMR7622, 9 Quai St-Bernard, 75005 Paris, France.

² Department of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92093-0380, USA.

³ Sorbonne Université, Inserm, UMR-S 938, Centre de Recherche Saint-Antoine-Team Proliferation and Differentiation of Stem Cells, Institut Universitaire de Cancérologie, F-75012 Paris, France.

⁴ Plateforme de génomique, Université de Paris, Institut Cochin, Inserm, CNRS, F-75014 Paris, France.

⁵ Service de microscopie électronique, Fr3631, Institut de Biologie Paris Seine, Sorbonne Université, CNRS.

⁶ Univ Paris Est Creteil, Inserm, IMRB, F94010 Creteil, France.

⁷ Université de Paris, Inserm UMR 1131, Institut de Recherche Saint Louis, Hôpital Saint Louis, 1 Avenue Claude Vellefaux, 75010 Paris, France.

⁸ Institut Curie, PSL Research University, CNRS UMR168, Paris, France.

* Co-authors

** Co-senior authors

Correspondence to Charles Durand: charles.durand@sorbonne-universite.fr
<https://orcid.org/0000-0002-1970-0467>

Short running title: Molecular identity of the HSC subaortic niche

Abstract

The first hematopoietic stem and progenitor cells (HSPCs) emerge in the embryonic dorsal aorta of midgestation mouse embryo, but the precise nature of the supportive microenvironment remains largely unexplored. To address this question, we designed a laser microdissection-based approach to isolate the dorsal and ventral tissues of the aorta at three distinct developmental stages harboring HSPCs. Bulk and single cell transcriptomics combined with computational analyses disclosed a large mesenchymal cell population in the E11.5 subaortic tissue including a subset expressing interconnected genes implicated in adhesive and neuronal functions. *In vivo* lineage tracing showed that this subset does not derive from neural crest cells. Histological and functional studies validated our strategy and revealed that Decorin, and Tenascin C, two components of the extracellular matrix (ECM), are essential for HSPC development. Together, our study identifies a yet unrecognized subaortic mesenchymal cell population and underscores the critical role of the ECM on embryonic HSPCs.

Key points

A subaortic cell subset, not derived from neural crest cells, exhibits adhesive and neuronal molecular features.

Decorin and Tenascin C regulate the HSPC development in the dorsal aorta.

Introduction

The first adult-type hematopoietic stem cells (HSCs) are generated in the Aorta-Gonad-Mesonephros (AGM) region of midgestation mouse embryo^{1,2}. Intra-aortic hematopoietic clusters (IAHCs), containing hematopoietic stem and progenitor cells (HSPCs) closely attached to the ventral side of the dorsal aorta³⁻⁵, originate from hemogenic endothelial cells through a process of endothelial-to-hematopoietic transition (EHT)⁵⁻¹². In the mouse, IAHCs are also found on the dorsal side of the aorta¹³, however HSC activity is preferentially detected in the ventral aortic domain^{13,14}. Moreover, preventing the migration of the subaortic mesenchyme in the embryo was shown to inhibit IAHC production¹⁵. Collectively, these data indicate that the aortic hematopoietic activity is polarized along the dorsoventral axis and that the ventral domain constitutes a supportive HSPC microenvironment. Recent studies have highlighted the heterogeneity of the AGM subaortic niche in terms of phenotype, gene expression and lineage history¹⁶⁻¹⁹. Molecularly, extrinsic factors including cytokines^{14,20,21}, signaling pathways^{14,22-28} and shear stress mediators, proinflammatory and sympathetic nervous systems²⁹⁻³⁵ participate in HSPC regulation. The subaortic extracellular matrix (ECM)-embedded mesenchyme being known to provide critical HSPC niche factors^{14,23,25,36}, it appears essential to define its molecular identity, which remains a technical and biological challenge since the ventral domain contains diverse cell types including endothelial cells, macrophages and neural crest cells^{31-35,37,38}.

We previously reported a systems biology approach for extracting the molecular core of HSPC niches^{39, 40}. We present here an original strategy combining precise laser capture, bulk and single cell transcriptomics combined with computational analyses (including correlation- and information-based gene network analyses) and functional validation to capture the molecular identity of the E11.5 AGM mesenchyme. The subaortic mesenchyme exhibited a unique molecular signature with strongly interconnected neuronal and adhesion-related genes. Lineage tracing experiments indicated that this singular cell population was not derived from neural crest cells. Using the zebrafish model, we functionally validated Decorin and Tenascin C as critical AGM HSPC regulators. Altogether, this study provides an unexpected insight onto the molecular identity of the AGM HSPC niche.

Methods

Mouse and human embryos

C57BL/6 mice were bred at Janvier (Le Genest, France) and maintained in the animal facility of the Laboratory of Developmental Biology according to institutional guidelines. Mouse models used in this study were the Tg (PLAT-cre)¹¹⁶Sdu⁴¹, referred as (HtPA)-Cre, and Cre reporter mice B6.Cg-Gt (ROSA)²⁶Sortm9 (CAG-tdTomato)Hze/J⁴² for tdTomato. Human embryonic tissues were obtained following voluntary, spontaneous or therapeutic abortions. Informed consent for the use of the embryo in research was obtained from the patient, and tissue collection and use were performed according to the guidelines and with the approval of the French National Ethics Committee.

Tissue collection for transcriptomics

In a first step, we used the Arcturus Laser Capture Microdissection System to isolate from mouse embryos the ventral and dorsal aortic tissues surrounding the dorsal aorta. The dorsal tissue localized in between the notochord and the roof of the aorta and the ventral tissue underneath the aorta were isolated from the same embryos. In a second step, AGMs were dissected from E11.5 embryos, digested with collagenase and used for single cell transcriptomics.

Microarray, single cell screening and data analysis

cDNA from microdissected tissues was hybridized to GeneChip® MouseGene 2.0 ST (Affymetrix). Normalization and functional analyses of the datasets were performed using Partek Genomics Suite software. E11.5 AGMs were dissected and submitted to scRNAseq using 10X Genomics Chromium next GEM single cell 3' kit version 3.1. Gene networks were obtained using the Weighed Gene Correlation Network Analysis (WGCNA)⁴³ and multivariate information-based inductive causation (miic) algorithms^{44,45}.

Zebrafish studies

Zebrafish were maintained according to the 'University of California, San Diego, International Animal Care' and 'User Committee and aquatic animal guidelines of Sorbonne Université'. Wild-type or transgenic zygotes were injected with a morpholino solution and incubated at 28.5°C until they reached the stage of interest. Embryos were collected, staged, fixed, and processed for *in situ* hybridization as previously described⁴⁶.

Statistics

Statistical analysis was performed using GraphPad Prism software. Unpaired t or Mann and Whitney tests were used to evaluate statistical significance. Data were considered statistically significant for $p < 0.05$.

Detailed protocols and computational analyses are in supplemental Methods.

Data sharing statement

The microarray data from this study have been deposited in the Gene Expression Omnibus database (accession number GSE159592).

Results

The dorsal and ventral aortic tissues exhibit dynamic and distinct molecular signatures

We used a laser microdissection approach to isolate the ventral (VT) and dorsal (DT) aortic tissues from E10.5, E11.5 and E12.5 mouse embryos (Figure 1A, Supplemental Table 1 and Video 1). The strategy is schematized in Figure 1 and developed below.

ANOVA with a low stringency p-value ($p < 0.06$) was used to extract the lists of differentially expressed genes (DEGs) between VT and DT. There were 2704, 1838 and 1873 DEGs at E10.5, E11.5 and E12.5, respectively. VT and DT were characterized by different gene ontology (GO) profiles (Figure 2A). VT profiles (top panels), either peaked at E11.5 (e.g., 'ECM', 'cell junction', 'synapse'), or declined from E10.5 to E11.5 (e.g., 'transcription', 'cell cycle'). DT profiles (bottom panels) either peaked at E12.5 (e.g., 'developmental protein', 'transcription', 'cell cycle'), or declined from E10.5 onwards ('synapse', 'cytokine activity', 'cell junction'). The contrast of VT vs. DT was confirmed using Gene Set Enrichment Analysis⁴⁷ (GSEA, Figures 2B and C). The gene sets 'Complement', 'IL6 Jak3 signaling' and 'Interferon α response' were significantly enriched in VT, in agreement with recent studies demonstrating the role of proinflammation on AGM HSCs³²⁻³⁵. Moreover, the 'Epithelial-to-Mesenchymal Transition' gene set was also

significantly enriched in VT, including numerous highly ranked genes coding for ECM proteins such as *Dcn* and *Tnc*.

Gene networks including neuronal- and mesenchymal-related genes characterize the dorsal and ventral tissues at E11.5

Our further study focused on E11.5 since the expression profiles were most contrasted at this time point, and it is known that HSC numbers increase in the AGM between E11 and E12⁴⁸.

To enrich the database and eventually disclose the mesenchyme-specific genes, we merged our E11.5 transcriptomes with two other expression sets. The first one corresponded to AGM stromal lines with differing capacities to support HSPCs (AM14-1C4, AM30-3F4, UG26-1B6 and UG26-3B5)^{39,49}. Using Principal Component Analysis (PCA), VT and DT were not discriminated using the entire transcriptome (Supplemental Figure 1A, top panel), but was evidenced after selection of the 3635 DEGs (gene set 1) (Supplemental Figure 1A, bottom panel). The second one corresponded to flow cytometry AGM sorted CD144^{pos}CD45^{neg} endothelial cells (EC), CD34^{low}Kit^{low}CD45^{pos} differentiated hematopoietic cells (HC), and CD34^{high}Kit^{high}CD45^{pos} HSPCs. Similar to the first expression set, the different cell populations were not segregated by PCA using the entire transcriptome (Supplemental Figure 1B, top panel) but segregated after selection of 3231 DEGs (gene set 2). Importantly, DT discriminated from VT according to PC1, and EC from HC according to PC2 (Supplemental Figure 1B, bottom panel). To refine the mesenchyme-specific gene list, we intersected gene sets 1 and 2 providing a list of common 2401 genes listed in Supplemental Table 2. PCA (Figure 3A) still enabled discriminating DT from VT according to PC1 (34,2% of the variance) and EC from according to PC2 HC (19,1% of the variance).

To find out which genes were specific to each of the four discriminated cell and tissue populations (VT, DT, EC, HC) and to reveal their relationships, Weighed Gene Correlation Network Analysis (WGCNA) (Supplemental Table 2 and Supplemental Figure 1C) combined to PCA was applied to the gene set 3⁴³. PC1 and PC2 were used to quantify the external traits (DT-to-VT and EC-to-HC contrasts), and the merged dynamic tree cut

method was used to identify the modules (Figure 3B). Eigenvector correlations to PC1 indicated that two modules, designated blue and yellow, corresponded to VT and another two, brown and turquoise, to DT (Figure 3C). Likewise, correlation to PC2 trait enabled identifying one major cyan module for EC and two modules, magenta and salmon, for HC (Figure 3C). Gene analysis confirmed the endothelial and hematopoietic identities of the populations corresponding to the EC and HC modules, respectively (Supplemental Table 2). Taken together these results indicate that our *in-silico* strategy has allowed excluding the contribution of genes expressed by endothelial and hematopoietic cells (not polarized along the dorsal to ventral axis) from those expressed by DT and VT tissues.

The DT and VT modules had very distinctive features. The brown DT module included 151 genes (Supplemental Table 2) with genes implicated in neurogenesis and/or the cell cycle. Many transcription factors essential for neuronal development and differentiation such as *Sox2*, *Pou3f3*, *Pou3f2*, and *Npas3* were among the most connected genes. The turquoise module included 297 genes (Supplemental Table 2) many of which were implicated in chondrogenesis and synapse formation. The transcription factors *Pax1*, *Pax9* expressed in pre-chondrocyte condensations, *Gli1* required for skull skeleton, vertebral column and cranio-facial development, the ECM cartilage collagen components *Col2a1* and *Col9a1* and the cartilage proteoglycan *Acan*, were among the most connected genes.

The blue VT module included 295 genes (Supplemental Table 2) implicated in either adhesion-related ('cell adhesion', 'basement membrane', 'ECM', 'collagen', 'homophilic cell adhesion', 'focal adhesion') or neuronal-related processes ('synapse', 'ion channel', 'GABA-A receptor activity', 'Neuroactive ligand-receptor', 'Neurotrophin signaling') (Figure 3D). Among the most connected genes were the ECM components *Dcn*, *Tnc* and *Col25a1*, the neurotrophin-3, glutamate and netrin receptors *Ntrk3*, *Grid1* and *Unc5c*, and molecules implicated in synapse functioning and/or neuronal development, *Clstn2*, *Amph*, *Syt1*, *Cntn4*, *Ntf4* (Figures 3E and 3F). The yellow VT module with 288 genes (Supplemental Table 2) was less well defined since including poorly characterized or predicted genes (olfactory receptors, Riken, Gm genes). However, among the most connected genes were the ECM component *Lama2*, the ECM-organizing protein *Fras1* and *Cnn1* implicated in the regulation of smooth muscle contraction.

These data suggested the presence of several cell populations within the dorsal and ventral tissues expressing different genes organized in distinct networks. To find out the identities of those cell populations and how they would integrate within the AGM region we performed single cell transcriptomic analysis on the entire AGM.

Single cell analysis discloses a large mesenchymal cell population including dorsal and ventral subpopulations expressing the different gene networks

After Quality Control (Supplemental Figure 1D) we retained 15154 out of 15862 cells, with 2698 genes detected on average. After normalization and dimension reduction, the Louvain algorithm allowed defining clusters projected on Uniform Manifold Approximation and Projection (UMAP) embedding (Figure 4A). Thirteen of the 22 clusters were packed together forming a larger cluster, clearly discriminated from the 9 others that presented as single entities.

The large cluster consisted in mesenchymal cells as indicated by the widespread expression of the interstitial collagens *Col1a1*, *Col1a2* and *Col3a1*, the lysyl oxidase *Loxl1* essential for collagen crosslinking, the microfibril component *Fbn2*, the ECM-degrading metalloproteases *Mmp16* and *Bmp1*, the proteoglycan *Vcan*, several molecules implicated in matrix assembly (*Ccdc80*), matrix organization (*Fras1*) and cell-collagen interactions (*Tgfb1*), the receptors for the growth factors of mesenchymal cells (*Pdgfra* and *Pdgfrb*), the hematopoietic chemokine expressed by stromal cells *Cxcl12*, and molecules implicated in epithelial-mesenchymal transition, the transcription factor *Snai2* and the growth factor *Ptn* (Figure 4B, Supplemental Figure 2A). Of note, we had found in a previous work that some of these genes (*Ccdc80*, *Snai2* and *Tgfb1*) were regulators of HSPC development in the AGM.

The next step was to ascertain in which of the cell subpopulations the gene networks described above were expressed (Figure 4C). The projection of the genes of the networks onto the cluster map indicated that the blue module genes were expressed in two clusters, C6 and C7, within the mesenchymal population. Since according to PCA and WGCNA the blue module was characteristic of the ventral tissue, it may be concluded that the C6 and C7 represented subaortic subpopulations of mesenchymal cells. Similarly, the VT yellow

module genes were projected mainly onto the C4 cluster that consisted therefore in another subset of mesenchymal cells ventrally located (there was also a minor contribution to C8). The projection of the DT turquoise module genes showed that the genes were expressed for most in the C0 cluster, albeit with a minor contribution in C2, indicating consequently that C0 and C2 cells were also mesenchymal subsets, but with a different localization from the previous ones since present in the dorsal tissue. Finally, the brown module genes were expressed in a non-mesenchymal population, the C13 peripheral cluster located also within the dorsal tissue (Figure 4C).

Analysis of the distribution of individual genes (Supplemental Figure 2B) confirmed that cells in C0 underwent chondrocytic differentiation, since expressing not only key transcription factors for pre-chondrocytes and chondrocytes such as *Pax1* and *Sox9*, but also cartilage-specific collagens and proteoglycans such as *Col2a1* and *Acan*. Of note, cells in C0 also expressed *Runx2*, the osteoblast-specific transcription factor, but none of the other transcription factors and bone-specific proteins (such as *Sp7*, *Bglap* or *Spp1*). Individual gene analysis also confirmed that cells in C4 followed a vascular smooth muscle pathway up to the contractile stage since expressing the transcription factor *Myocd*, smooth muscle actin *Acta2*, and smooth muscle myosin heavy chain *Myh11*; these cells also expressed the proteoglycan *Cspg4* characteristic of pericytes. Further analysis of individual genes in the non-mesenchymal populations confirmed that the C13 cells expressed neuronal genes such as the neuronal intermediate filament *Ina* essential for neuron morphogenesis, and the synaptic protein *Rims2*. A few genes expressed in C13 were also expressed in C11, such as the transcription factor *Pou3f2*.

Analysis of individual genes of the C6 and C7 clusters corresponding to the blue module, which exhibited such a singular molecular identity with both neuronal and adhesive genes, confirmed the expression of genes of both categories. Examples are shown on Supplemental Figure 2C: the ECM component *Tnc*, the neurotrophin-3 receptor *Ntrk3*, *Syt1*, the calcium sensor triggering neurotransmitter release at the synapse, and the glutamate receptor *Grid1*. Moreover, the analysis aimed at answering three questions: i) were C6 and C7 representative of a large subset since sharing most of the genes; ii) if so, what was the percentage of cells expressing both neuronal and mesenchymal genes? iii) were some mesenchymal genes more expressed in this cell subpopulation as compared

to the others? Out of a hundred genes expressed in C6 71% were also expressed in C7 at about the same intensity level, establishing therefore that the C6 and C7 clusters may be considered as a unique cell subpopulation (the two-clusters definition by the Louvain algorithm was likely related to the remaining 29% of genes expressed for most only in C7 cells). By extracting two lists of genes representative of the neural and mesenchymal lineages from the blue module (Supplemental Table 3), we used gene set scoring to show that 63.9% of cells in clusters 6 and 7 were positive for both gene subsets, confirming their neuro-mesenchymal signature (Figure 4D). As can be seen already on that figure the mesenchymal gene profile appeared distinctive, since expressed not only in C6 and C7 but also in C1 (including genes implicated in metabolic pathways). Searching for individual genes expressing such profile we found that it corresponded to a several ECM molecules, Decorin (*Dcn*), a member of the small leucine-rich proteoglycan family with essential role in collagen fibril assembly, and collagen family members, *Col4a5*, the major component of basement membranes, and two chains of Collagen VI (*Col6a3* and *Col6a2*), acting as a cell-binding protein (Supplemental Figure 2D).

Finally, we used the single cell study to get more insight on the organization of genes characterizing the neuro-mesenchyme and identify direct gene-to-gene interactions or potential causal links. We applied the multivariate information-based inductive causation (miic) algorithm ^{44,45} on the blue gene module. To yield significant results the miic algorithm requires a large number of observations as compared to that of variables. We restricted therefore the analysis to the central core of the module including 98 genes and we considered all cells (N = 12516) belonging to the large mesenchymal cluster. After filtering out low confidence links, 89 nodes defining direct high-confidence paths were retained (Figure 4E). The resulting pruned network (as compared to that provided by WGCNA) included four major hubs (with degree ≥ 17), *Unc5c*, *Col1a1*, the neural cell adhesion molecule *Ncam1*, and the transcription factor *Zmiz1*, and some minor ones such as the dopa decarboxylase *Ddc*, primarily expressed by C7 cells as were all of its 7 links (including *Tnc*, the channel interacting protein *Kcnip4*, the ion transport ATPase *Nkain2* and the metalloproteinase *Adam23*), establishing therefore a relationship between the spatial distribution of genes and the graphical representation of their interaction network (Figure 4E, right panel). Importantly, the links between *Dcn* and *Col1a1* and *Ncam1* and *Ntrk3* (via *Tnc*) were confirmed by the analysis of protein-protein interactions using

STRING⁵⁰. Finally, the miic algorithm uncovered a number of 'v-structures' (e.g., the nuclear receptor *Nr1h5* --> *Tnc* <-- *Ncam1*) suggesting cause-and-effect interactions to be investigated in future works.

Analysis of the remaining clusters using first-ranked genes indicated that they included genes implicated in metabolic pathways (C1, C5), cell cycle (C2), and developing tissues (C3, C9, C10, C11) (Table in Figure 4A). The same approach for the non-mesenchymal populations indicated that C15 was made of endothelial cells and C16 of innate immunity cells (genes belonging to networks cyan, salmon and magenta were projected on these clusters). There were also minor populations of red cells, cells involved in hemostasis, kidney and liver cells, sarcomeric muscle and pluripotent cells (Table in Figure 4A).

Taken together our findings reveal the heterogeneity of the subaortic mesenchyme that includes a cell subpopulation expressing genes organized in a network in which cell adhesion genes are directly linked to neuronal ones.

Subaortic neuro-mesenchymal cells do not derive from neural crest cells and strongly express Decorin (Dcn), Tenascin C (Tnc) and the Neurotrophin-3 growth factor receptor (Ntrk3)

Single cell analysis revealed that the neural crest cell marker *Sox10* was expressed in C14 whereas the late markers of the sympathoadrenal compartment, *Phox2a* and *Th*, and the neuronal marker *Tubb3* were expressed in C13 (Figure 5A), confirming a clear segregation between the subaortic neuro-mesenchyme and neural/neural crest cell clusters. Moreover, to examine whether this C6/C7 cell subpopulation may derive from neural crest cells, we performed lineage tracing and immunodetection experiments for three molecules expressed in this subset and in the corresponding blue module: *Dcn*⁵¹ and *Tnc*⁵², encoding for two extracellular matrix genes and *Ntrk3* encoding a neurotrophin receptor⁵³. We took advantage of a Human tissue Plasminogen Activator (HtPA)-Cre mice labelling neural crest cells and their derivatives *in vivo*²⁹⁶. After crossing those mice with GT Rosa Tomato mice, E11.5 embryos were collected and analyzed for the expression of Tomato, *Dcn*, *Tnc* and *Ntrk3*. Tomato⁺ cells were detected as expected on the dorsal, lateral and ventral sides of the aorta (Figures 5 B-D). There was little, if any,

overlap between Tomato expression and that of Dcn, Tnc or Ntrk3 (Figures 5 B-D). Dcn was indeed detected at E11.5 around the notochord and in subaortic tissue, but not in aortic endothelial, vascular smooth muscle cells, or macrophages as shown by the absence of overlapping staining between Dcn and CD31, α -smooth muscle actin (α -SMA) or Allograft inflammatory factor 1 (Iba1) (Figures 5F, 5M and Supplemental Figure 3A). Dcn was expressed in a dynamic manner along the antero-posterior axis with strong expression in cells and associated ECM in large areas of the middle part of the AGM known to be enriched for HSC activity in mouse and human embryos^{54,55} (Supplemental Figure 3E-H). Tnc was also detected in a dynamic manner and found also highly expressed in cells and associated ECM in large areas of the subaortic tissue (Figures 5C, 5G and Supplemental Video 2). In keeping with a previous report⁵⁶, similar distribution of Dcn and Tnc were observed in human embryos at 30 days of gestation (Figures 5H and I). Ntrk3 was present not only as expected in the dorsal root ganglion cells but also in extensive areas in cells around and underneath the dorsal aorta (Figures 5D and 5J-L and Supplemental Video 3). Immunostaining for Tyrosine hydroxylase (Th), an enzyme involved in catecholamine biosynthesis, revealed discrete positive cells in the dorsal and ventral domains of the aorta, corresponding to sympathetic nerve cells (Figure 5K). Immunostaining for Tuj1, a neuron-specific class III beta-tubulin expressed in neurons of the peripheral and central nervous system, revealed a dense array of Tuj1⁺ cells on the lateral sides, and a looser cell distribution underneath the aorta (Figures 5E, 5L, 5M and Supplemental Figure B-D). Importantly, the distributions of Dcn, Tnc and Ntrk3 staining were quite similar in large areas of the subaortic tissue, and clearly distinct from that of Th⁺ cells and Tuj1⁺ cells (Figures 5J-M and Supplemental Figures 3B-D). Finally, transmission electron microscopy revealed the presence of transverse and longitudinal ECM fibers in the intercellular space between subaortic cells at E11.5 (Figures 5N and 5O).

Thus, in line with our single cell transcriptomics, these results suggest that the subaortic mesenchymal subset expressing Dcn, Tnc and Ntrk3 were not derived from neural crest cells and were clearly distinct from neuronal Tuj1⁺ cells. They also highlighted the presence of a dense network of ECM molecules underneath the dorsal aorta.

Decorin and Tenascin C are essential for HSPC development

These observations prompted us to further investigate the roles of *Dcn* and *Tnc* on AGM HSPCs. We performed knockdown experiments in zebrafish embryos using antisense morpholinos (MO). The formation of HSPCs in the dorsal aorta was monitored by whole mount *in situ* hybridization (WISH) for the expression of the HSPC markers *runx1* and *cmyb*. The morphants *dcn* injected with start MO exhibited strong hematopoietic defects in the dorsal aorta, as shown by significant reduced expression of *runx1* and *cmyb* in morphants as compared to respective controls at 30 and 36 hours post-fertilization (hpf), respectively (Figures 6A-F). Similar phenotypes were observed with the injection of a splice MO (Supplemental Figures 4A and B). Since adult *dcn*^{-/-} zebrafishes (*dcn*^{sa20247} line), harboring a point mutation that induces a stop codon in the *dcn* coding sequence were viable and fertile, we crossed *dcn*^{-/-} females with *dcn*^{-/-} males. Resulting embryos exhibited a significant decrease in the numbers of *runx1* and *cmyb* positive cells as compared to *dcn*^{+/+} embryos from wild-type sibling crosses (Figures 6G-L), thus confirming the phenotypes of the *dcn* morphants.

Morphants for *tnc* also exhibited a decrease in the numbers of *runx1* and *cmyb*- expressing cells in the dorsal aorta as compared to respective controls (Supplemental Figures 4A and B). The reduced expression of the two HSC genes, *runx1* and *gata2b*, was also confirmed in *tnc* and *dcn* morphants as compared to uninjected embryos by qRT-PCR (Supplemental Figure 4A, right). Confocal microscopy using *kdlr:mCherry*; *gata2b:UAS*; *GAL4:eGFP* transgenic animals showed a strong reduction in the number of definitive HSPCs in the dorsal aorta in *dcn* and *tnc* morphants (Supplemental Figure 4C). These effects occurred in absence of any vascular defects as shown by confocal imaging using *flt1:dtomato*; *flt4:yfp* transgenic lines reporting vessel development (Supplemental Figure 4D). Collectively, these results indicate that Dcn and Tnc are critical HSPC regulators.

Discussion

In this work we have uncovered the cell populations of dorsal (DT) and ventral (VT) aortic tissues at the critical period of HSPC formation in the mouse embryo. We focused on E11.5 as the time of HSC number increase in the AGM, and showing peak activity of a number of VT genes. By combining DT and VT transcriptomes with that of sorted hematopoietic and

endothelial populations from age-matched AGM, we could exclude the specific contributions of the endothelial and hematopoietic tissues and focus on the other, yet less understood tissue components, including the aortic mesenchyme. One major finding of our study was the identification within a large population of mesenchymal cells of a subpopulation (that we called ‘neuromesenchyme’) in which neuronal genes were directly linked to genes implicated in cell adhesion. Moreover, we showed that the HSPC-supportive niche activity of the subaortic tissue was related to two ECM components, Tenascin C and Decorin, which validated our approach.

The functional association between the hematopoietic and neuronal systems has already been documented in the embryonic context. For example, it has been reported that *Gata3* controls the numbers of AGM HSCs through the production of catecholamines, showing a strong influence of the sympathetic nervous system on AGM HSCs ³¹. A transcriptomic analysis in the human embryo at the time of HSC development has highlighted a neuronal molecular signature associated with the ventrolateral domain of the dorsal aorta ⁵⁷. In zebrafish, trunk neural crest cells were shown to interact with precursors of aortic HSPCs and to contribute to their supportive microenvironment ⁵⁸. In mice, subaortic mesenchymal cells may have a dual origin (mesodermal and neural crest-derived) ^{16,17} with *Pdgfra*⁺ mesodermal stromal cells being replaced by neural crest-derived cells at E13.5 coinciding with the loss of EHT ¹⁶. Remarkably, in our present work we identified at E11.5 a mesenchymal cell subpopulation, not derived from neural crest cells, containing both neuronal genes, implicated in synapse formation, membrane channel activity, and neurotrophin signaling, and mesenchymal genes, in particular some distinctive ECM components.

To evaluate the roles of Decorin and Tenascin C, we performed loss-of-function experiments in zebrafish embryos, which revealed that both proteins are necessary for HSPC formation and suggested an important role for ECM on AGM hematopoiesis. Interestingly, in BM HSPC-supportive stromal cells *Dcn* is not up- but down-regulated ⁴⁰, which is in agreement with its reported role as HSPC negative regulator in adult mice ⁵⁹. These data underscore the context-dependency of some of the mediators, activators in the embryonic niche but inhibitors in the adult one. Decorin and Tenascin C could exert their effects on AGM HSPCs as structural and/or signaling molecules by interacting with other

ECM proteins (e.g., fibronectins, collagens, periostin) and/or binding to cell surface receptors and soluble factors. For example, it has been reported that Tenascin C interacts with Wnt3a, the Transforming growth factor β and the Vascular endothelial growth factor⁵². As a consequence, Decorin and Tenascin C may play a critical role on cell-to-cell communication by regulating the half-life, local concentration and availability of critical growth factors, cytokines and morphogens that may modify the properties of the overlaying endothelial cell layer and /or act beyond this layer on HSPCs.

In conclusion, at E11.5 we have identified a unique mesenchymal cell population harboring both an adhesive and neuronal molecular identity within the subaortic tissue. Our data also highlight the role of specific ECM components as part of the HSPC-supportive niche in the embryo.

Author contributions

PC, TJ and CD designed research, analyzed the data and wrote the manuscript. PYC, OM and CP performed experiments and analyzed the data. DT and CD designed research on zebrafish and analyzed the data. NR, PF, AM and EH performed experiments and analyzed the data. SJ and AG provided support and expertise on bulk transcriptomics and transmission electron microscopy, respectively. OP and HI analyzed data. SD and MS provided materials. TJ provided financial support.

Acknowledgments

This work was supported by grants of the Fondation pour la Recherche Médicale (DEQ20100318258 and EQU201903007867), Centre National de la Recherche Scientifique (CNRS), Institut national de la santé et de la recherche médicale (Inserm) and Sorbonne Université. We thank Pascal Legendre for critical reading of the manuscript, and Monara Angelim and Pascal Legendre (CNRS UMR8246) for providing us with the anti-Tuj1 and -Iba1 antibodies. We thank the imaging (Institut de la Vision and Institut de Biologie Paris-Seine) facilities for their expertise and advices. We thank Michael Trichet

and Alexis Canette (Service de microscopie électronique, Institut de Biologie Paris Seine, Sorbonne Université, CNRS) for their support and advices on the transmission electron microscopy experiments. We thank Sophie Gournet (CNRS UMR7622) for drawing assistance.

Figure legends

Figure 1. A laser microdissection and systems biology strategy.

(A) Tissue sample collection for transcriptomics. In a first step, dorsal and ventral surrounding aortic tissues were isolated by laser microdissection at E10.5, E11.5 and E12.5 (solid lines) and utilized for bulk transcriptomics. In a second step, E11.5 AGMs were dissected (dash line) and proceeded for single cell transcriptomics.

(B) Flow chart of the study. Green color indicates methods except for transcriptomics in red.

Ao: dorsal aorta; CV: cardinal vein; Nc: notochord. Scale bar for Figure 1A is 100 μm .

Figure 2. The dorsal and ventral tissues surrounding the dorsal aorta exhibit dynamic and distinct molecular signatures.

(A) GO analysis of DT and VT tissues at E10.5, E11.5 and E12.5.

(B) GSEA showing 4 gene sets significantly enriched in VT at E11.5. NES: normalized enrichment score, FDR: false discovery rate.

(C) Rank metric score of the 23 first genes from the 'Epithelial_Mesenchymal_Transition' gene set enriched in VT at E11.5.

Figure 3. Gene networks including neuronal- and mesenchymal-related genes characterize the dorsal and ventral tissues at E11.5.

(A) PCA (using 2423 genes and 15 biological samples) highlighting the different clusters. The score plot of PC1 vs. PC2 is shown.

(B) Module detection using WGCNA. Modules named by colors correspond to squares along the diagonal of the topological overlap matrix plot shown below the clustering tree. DT and VT modules are highlighted.

(C) Module eigenvectors corresponding to the DT to VT (PC1 trait) and to the EC to HC contrasts (PC2 trait).

(D) Significant GO categories corresponding to genes in VT blue module.

(E) Connectivity plot of genes belonging to the VT blue module for the 119 most connected genes (out of a total of 295 genes).

(F) Connectivity bar plot of the Top-30 genes.

Figure 4. Single cell analysis discloses a large mesenchymal cell population including dorsal and ventral subpopulations expressing the different gene networks.

(A) UMAP embedding of the Louvain clusters corresponding to E11.5 AGM single cell analysis.

(B) Expression pattern of genes defining the mesenchymal cell population.

(C) Distribution of genes of the turquoise, yellow, blue and brown WGCNA modules among the different clusters.

(D) Distribution of genes of the blue WGCNA module and the extracted neuronal and mesenchymal subsets among the different clusters.

(E) Miic plot of the central core genes of the blue module. On the right is shown the expression pattern for *Ddc* (encircled on the miic plot) and of three of its connected genes.

Figure 5. Subaortic neuro-mesenchymal cells do not derive from neural crest cells and strongly express Decorin (Dcn), Tenascin C (Tnc) and the NT-3 growth factor receptor (Ntrk3).

(A) Expression pattern of *Sox10*, *Phox2a*, *Th* and *Tubb3*.

(B-D) Immunodetection for Dcn (A), Tnc (B) and Ntrk3 (C) on E11.5 sections of HtPA-Cre; R26R-tomato mouse embryos.

(E) Immunodetection for Tuj1 on E11.5 mouse embryo sections.

(F) Immunodetection for Dcn and CD31 on E11.5 mouse embryo sections.

(G) Immunodetection for Tnc and CD31 on E11.5 mouse embryo sections.

(H-I) Immunodetection for Dcn (G), Tnc (H), and CD105 on 30 days human embryo sections.

(J) Immunodetection for Ntrk3 and CD31 on E11.5 mouse embryo sections.

(K) Immunodetection for Ntrk3 and Th on E11.5 mouse embryo sections.

(L) Immunodetection for Ntrk3, Iba1 and Tuj1 on E11.5 mouse embryo sections.

(M) Immunodetection for Dcn, Iba1 and Tuj1 on E11.5 mouse embryo sections.

(N-O) Transmission electron microscopy of E11.5 subaortic tissue. N is a magnification of the square indicated in Figure 5M.

Scale bar for Figures 5A, B, C, D, E, F, I, J, K and Land O is 100 μ m. Scale bar for Figures 5G and H is 50 μ m. Scale bars for Figures 5 N and O are 10 μ m and 200 nm, respectively.

Nc: notochord, Ao: dorsal aorta.

Figure 6. Decorin and Tenascin C are essential for HSPC development.

(A-B) WISH for *runx1* in 30 hpf *dcn* morphants and their respective uninjected controls.

(C) Numbers of *runx1*⁺ cells in the dorsal aorta of control and *dcn* morphant embryos.

(D-E) WISH for *cmyb* in 36 hpf *dcn* morphants and their respective uninjected controls.

(F) Numbers of *cmyb*⁺ cells in the dorsal aorta of control and *dcn* morphant embryos.

(G-H) WISH for *runx1* in 30 hpf *dcn*^{+/+} and *dcn*^{-/-} embryos.

(I) Numbers of *runx1*⁺ cells in the dorsal aorta of *dcn*^{+/+} and *dcn*^{-/-} embryos.

(J and K) WISH for *cmyb* in 36 hpf *dcn*^{+/+} and *dcn*^{-/-} embryos.

(L) Numbers of *cmyb*⁺ cells in the dorsal aorta of *dcn*^{+/+} and *dcn*^{-/-} embryos.

**** indicates p value < 0.0001 using either two-tailed t-test or Mann and Whitney test.
24 to 57 embryos were analyzed for each condition from 1 to 3 independent experiments.

Conflict-of-interest disclosure

The authors declare no competing interests.

References

1. Medvinsky A, Dzierzak E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*. 1996;86 (6):897-906.
2. Müller AM, Medvinsky A, Strouboulis J, Grosveld F, Dzierzak E. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*. 1994;1 (4):291-301.
3. Tavian M, Coulombel L, Luton D, Clemente HS, Dieterlen-Lièvre F, Péault B. Aorta-associated CD34+ hematopoietic cells in the early human embryo. *Blood*. 1996;87 (1):67-72.
4. Yokomizo T, Dzierzak E. Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development*. 2010;137 (21):3651-3661.
5. Jaffredo T, Gautier R, Eichmann A, Dieterlen-Lièvre F. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*. 1998;125 (22):4575-4583.
6. Oberlin E, Tavian M, Blazsek I, Péault B. Blood-forming potential of vascular endothelium in the human embryo. *Development*. 2002;129 (17):4147-4157.
7. de Bruijn MF, Ma X, Robin C, Ottersbach K, Sanchez MJ, Dzierzak E. Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity*. 2002;16 (5):673-683.
8. Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*. 2009;457 (7231):887-891.
9. Zovein AC, Hofmann JJ, Lynch M, et al. Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell*. 2008;3 (6):625-636.
10. Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DY, Traver D. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*. 2010;464 (7285):108-111.
11. Boisset JC, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*. 2010;464 (7285):116-120.
12. Kissa K, Herbomel P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*. 2010;464 (7285):112-115.
13. Taoudi S, Medvinsky A. Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. *Proc Natl Acad Sci U S A*. 2007;104 (22):9399-9403.
14. Souilhoul C, Gonneau C, Lendinez JG, et al. Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. *Nat Commun*. 2016;7:10784.
15. Richard C, Drevon C, Canto PY, et al. Endothelio-mesenchymal interaction controls runx1 expression and modulates the notch pathway to initiate aortic hematopoiesis. *Dev Cell*. 2013;24 (6):600-611.
16. Chandrakanthan V, Rorimpandey P, Zanini F, et al. Mesoderm-derived PDGFRA (+) cells regulate the emergence of hematopoietic stem cells in the dorsal aorta. *Nat Cell Biol*. 2022;24 (8):1211-1225.
17. Kapeni C, Nitsche L, Kilpatrick AM, et al. p57Kip2 regulates embryonic blood stem cells by controlling sympathoadrenal progenitor expansion. *Blood*. 2022;140 (5):464-477.
18. Sa da Bandeira D, Kilpatrick AM, Marques M, et al. PDGFRbeta (+) cells play a dual role as hematopoietic precursors and niche cells during mouse ontogeny. *Cell Rep*. 2022;40 (3):111114.
19. Fadlullah MZH, Neo WH, Lie ALM, et al. Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE. *Blood*. 2022;139 (3):343-356.
20. Robin C, Ottersbach K, Durand C, et al. An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells. *Dev Cell*. 2006;11 (2):171-180.
21. Petit-Cocault L, Volle-Challier C, Fleury M, Péault B, Souyri M. Dual role of Mpl receptor during the establishment of definitive hematopoiesis. *Development*. 2007;134 (16):3031-3040.
22. Marshall CJ, Kinnon C, Thrasher AJ. Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. *Blood*. 2000;96 (4):1591-1593.
23. Durand C, Robin C, Bollerot K, Baron MH, Ottersbach K, Dzierzak E. Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. *Proc Natl Acad Sci U S A*. 2007;104 (52):20838-20843.
24. Peeters M, Ottersbach K, Bollerot K, et al. Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development. *Development*. 2009;136 (15):2613-2621.
25. Wilkinson RN, Pouget C, Gering M, et al. Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta. *Dev Cell*. 2009;16 (6):909-916.
26. Robert-Moreno A, Espinosa L, de la Pompa JL, Bigas A. RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of intra-embryonic hematopoietic cells. *Development*. 2005;132 (5):1117-1126.
27. Pouget C, Peterkin T, Simões FC, Lee Y, Traver D, Patient R. FGF signalling restricts haematopoietic stem cell specification via modulation of the BMP pathway. *Nat Commun*. 2014;5:5588.

28. Lee Y, Manegold JE, Kim AD, et al. FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. *Nat Commun.* 2014;5:5583.
29. Adamo L, Naveiras O, Wenzel PL, et al. Biomechanical forces promote embryonic haematopoiesis. *Nature.* 2009;459 (7250):1131-1135.
30. North TE, Goessling W, Peeters M, et al. Hematopoietic stem cell development is dependent on blood flow. *Cell.* 2009;137 (4):736-748.
31. Fitch SR, Kimber GM, Wilson NK, et al. Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis. *Cell Stem Cell.* 2012;11 (4):554-566.
32. Espín-Palazón R, Stachura DL, Campbell CA, et al. Proinflammatory signaling regulates hematopoietic stem cell emergence. *Cell.* 2014;159 (5):1070-1085.
33. Li Y, Esain V, Teng L, et al. Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. *Genes Dev.* 2014;28 (23):2597-2612.
34. Mariani SA, Li Z, Rice S, et al. Pro-inflammatory Aorta-Associated Macrophages Are Involved in Embryonic Development of Hematopoietic Stem Cells. *Immunity.* 2019;50 (6):1439-1452.e1435.
35. Sawamiphak S, Kontarakis Z, Stainier DY. Interferon gamma signaling positively regulates hematopoietic stem cell emergence. *Dev Cell.* 2014;31 (5):640-653.
36. Yvernogeu L, Klaus A, Maas J, et al. Multispecies RNA tomography reveals regulators of hematopoietic stem cell birth in the embryonic aorta. *Blood.* 2020.
37. Travnickova J, Tran Chau V, Julien E, et al. Primitive macrophages control HSPC mobilization and definitive haematopoiesis. *Nat Commun.* 2015;6:6227.
38. Takahashi Y, Sipp D, Enomoto H. Tissue interactions in neural crest cell development and disease. *Science.* 2013;341 (6148):860-863.
39. Charbord P, Pouget C, Binder H, et al. A systems biology approach for defining the molecular framework of the hematopoietic stem cell niche. *Cell Stem Cell.* 2014;15 (3):376-391.
40. Desterke C, Petit L, Sella N, et al. Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations. *iScience.* 2020;23 (6):101222.
41. Pietri T, Eder O, Blanche M, Thiery JP, Dufour S. The human tissue plasminogen activator-Cre mouse: a new tool for targeting specifically neural crest cells and their derivatives in vivo. *Dev Biol.* 2003;259 (1):176-187.
42. Madisen L, Zwingman TA, Sunkin SM, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci.* 2010;13 (1):133-140.
43. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
44. Verny L, Sella N, Affeldt S, Singh PP, Isambert H. Learning causal networks with latent variables from multivariate information in genomic data. *PLoS Comput Biol.* 2017;13 (10):e1005662.
45. Sella N, Verny L, Uguzzoni G, Affeldt S, Isambert H. MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data. *Bioinformatics.* 2018;34 (13):2311-2313.
46. Thisse C, Thisse B, Schilling TF, Postlethwait JH. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development.* 1993;119 (4):1203-1215.
47. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102 (43):15545-15550.
48. Kumaravelu P, Hook L, Morrison AM, et al. Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development.* 2002;129 (21):4891-4899.
49. Oostendorp RA, Harvey KN, Kusadasi N, et al. Stromal cell lines from mouse aorta-gonads-mesonephros subregions are potent supporters of hematopoietic stem cell activity. *Blood.* 2002;99 (4):1183-1189.
50. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47 (D1):D607-d613.
51. Zhang W, Ge Y, Cheng Q, Zhang Q, Fang L, Zheng J. Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment. *Oncotarget.* 2018;9 (4):5480-5491.
52. Midwood KS, Chiquet M, Tucker RP, Orend G. Tenascin-C at a glance. *J Cell Sci.* 2016;129 (23):4321-4327.
53. Bartkowska K, Turlejski K, Djavadian RL. Neurotrophins and their receptors in early development of the mammalian nervous system. *Acta Neurobiol Exp (Wars).* 2010;70 (4):454-467.
54. Mascarenhas MI, Parker A, Dzierzak E, Ottersbach K. Identification of novel regulators of hematopoietic stem cell development through refinement of stem cell localization and expression profiling. *Blood.* 2009;114 (21):4645-4653.
55. Easterbrook J, Rybtsov S, Gordon-Keylock S, et al. Analysis of the Spatiotemporal Development of Hematopoietic Stem and Progenitor Cells in the Early Human Embryo. *Stem Cell Reports.* 2019;12 (5):1056-1068.

56. Marshall CJ, Moore RL, Thorogood P, Brickell PM, Kinnon C, Thrasher AJ. Detailed characterization of the human aorta-gonad-mesonephros region reveals morphological polarity resembling a hematopoietic stromal layer. *Dev Dyn*. 1999;215 (2):139-147.
57. Crosse EI, Gordon-Keylock S, Rybtsov S, et al. Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human Hematopoietic Stem Cell Development. *Cell Stem Cell*. 2020;27 (5):822-839.e828.
58. Damm EW, Clements WK. Pdgf signalling guides neural crest contribution to the haematopoietic stem cell specification niche. *Nat Cell Biol*. 2017;19 (5):457-467.
59. Ichii M, Frank MB, Iozzo RV, Kincade PW. The canonical Wnt pathway shapes niches supportive of hematopoietic stem/progenitor cells. *Blood*. 2012;119 (7):1683-1692.

Figure 01

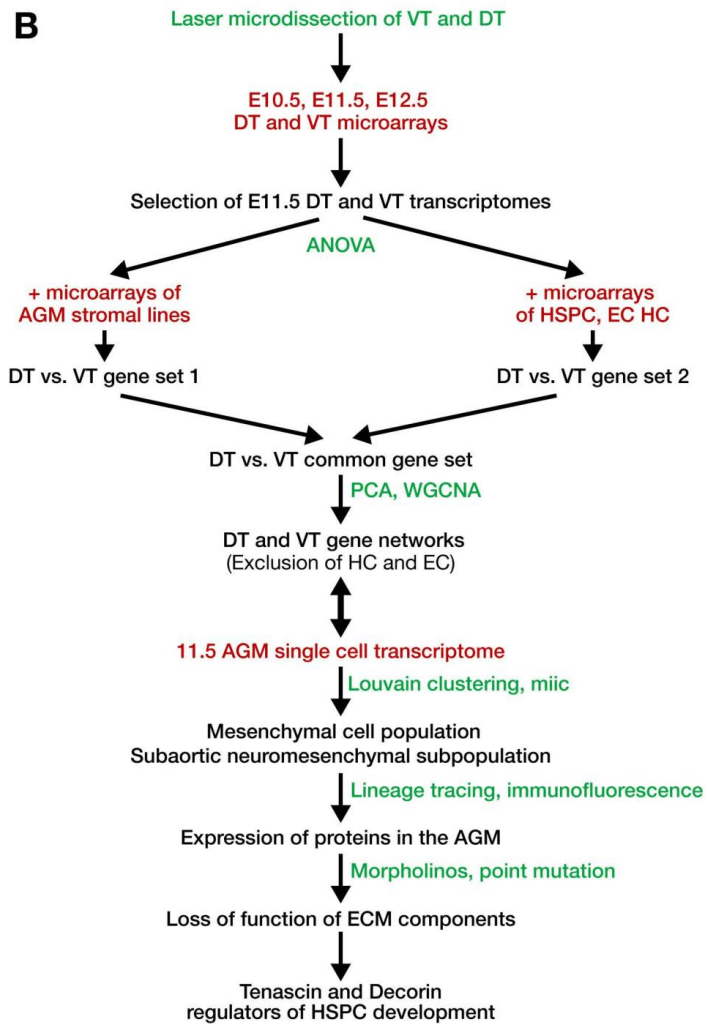
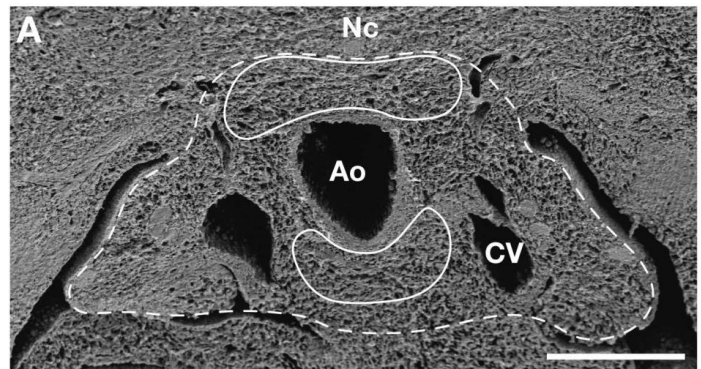


Figure 02

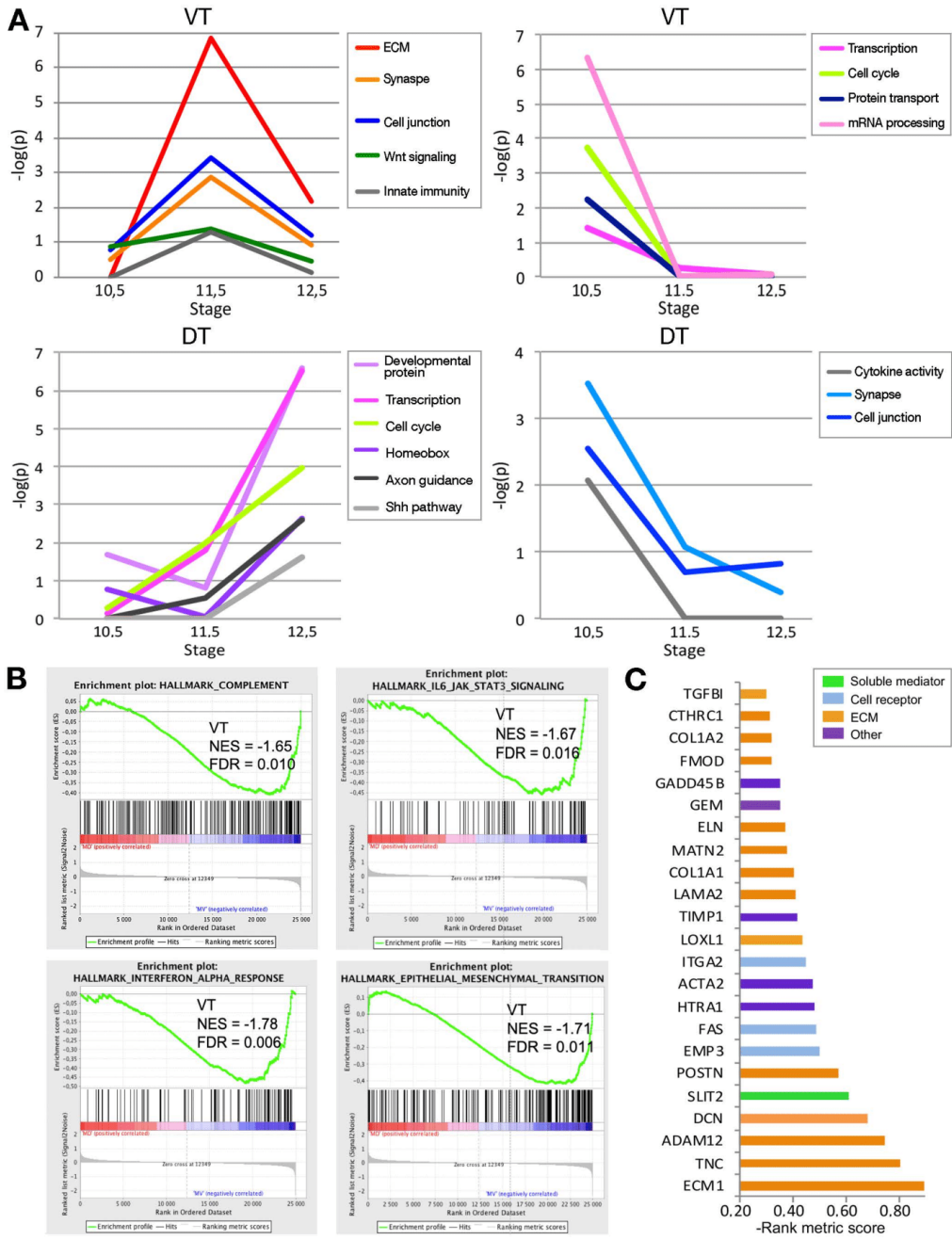


Figure 03

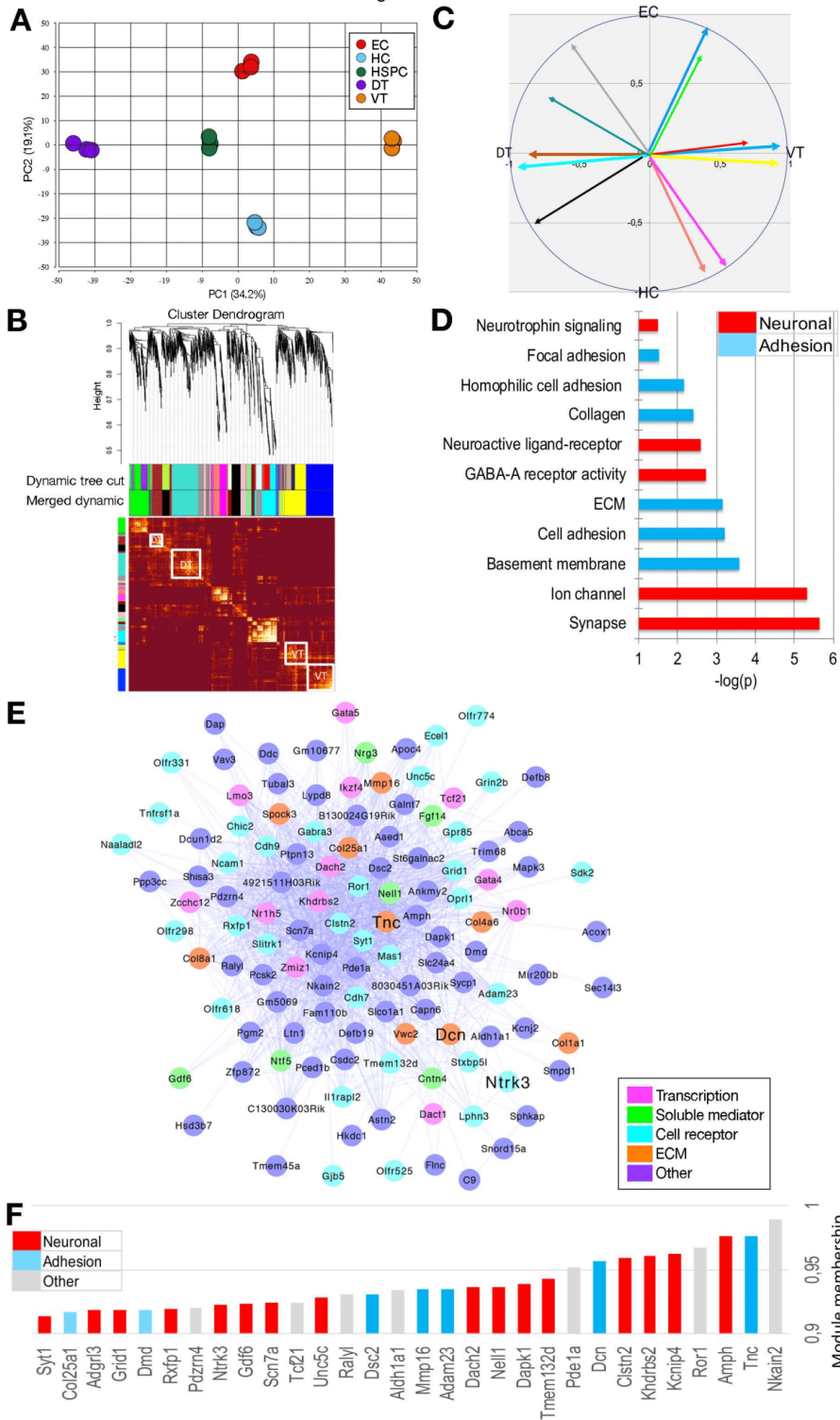


Figure 04

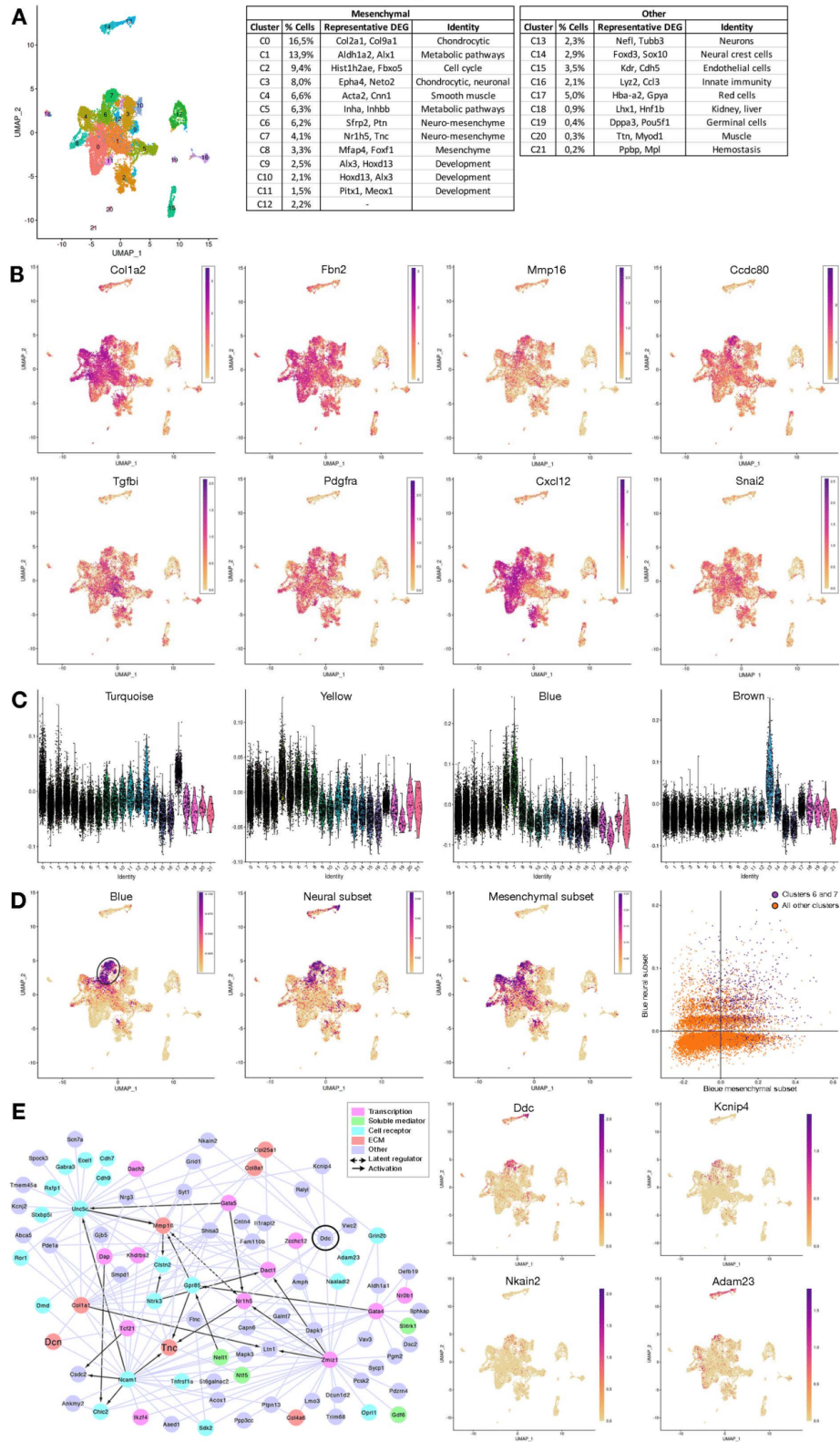


Figure 05

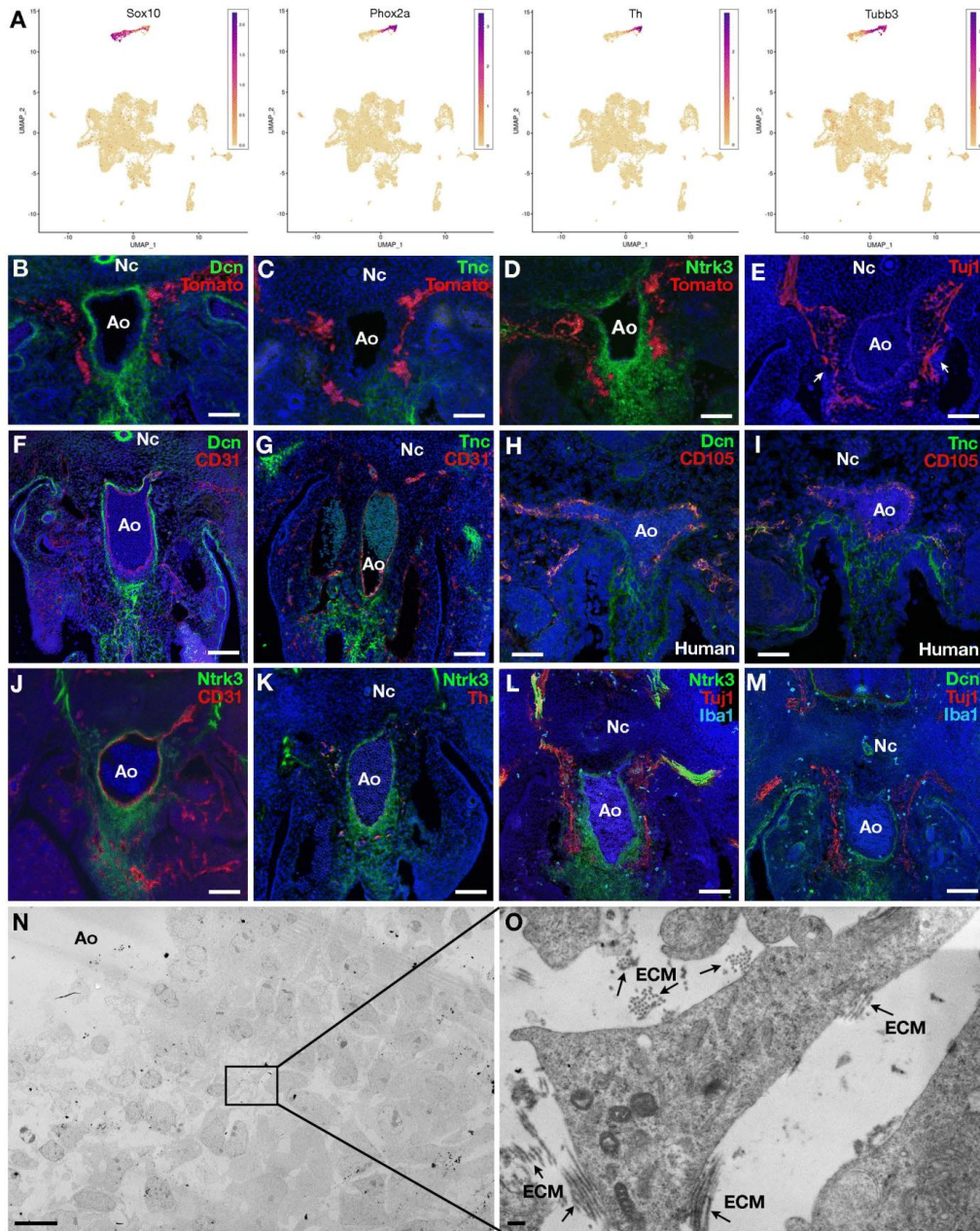


Figure 06

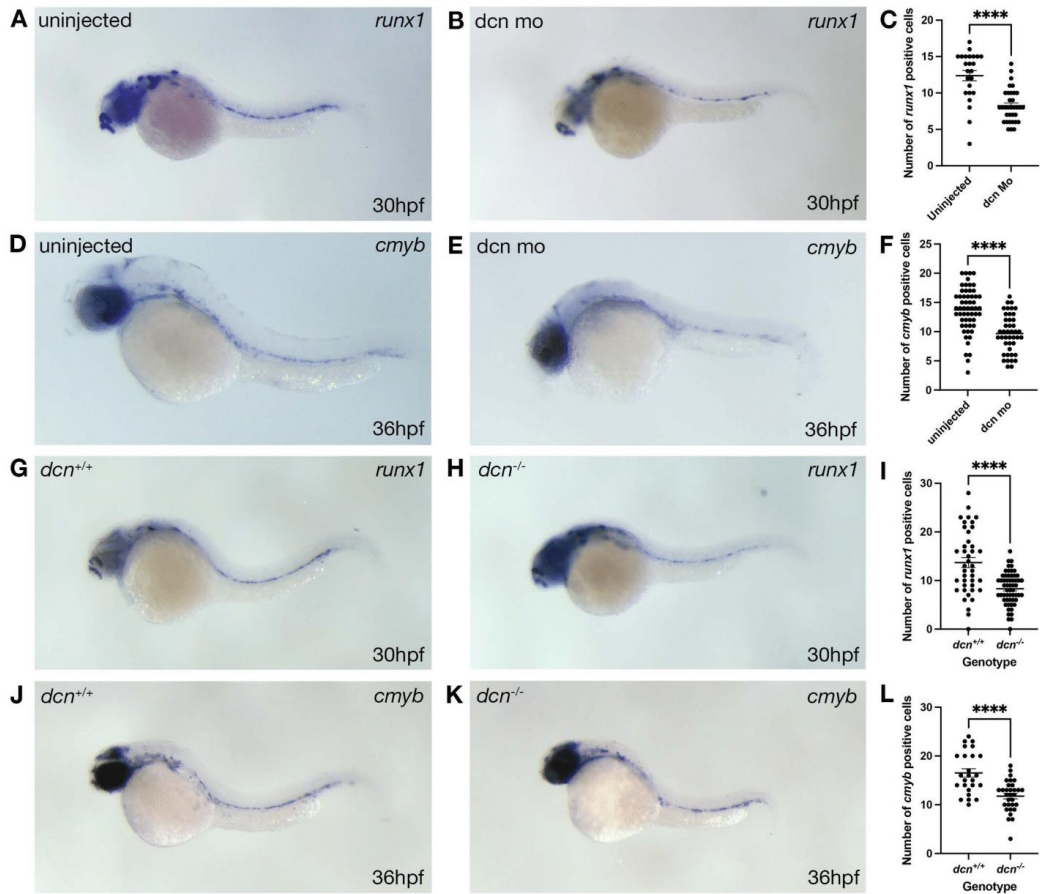


Figure S1

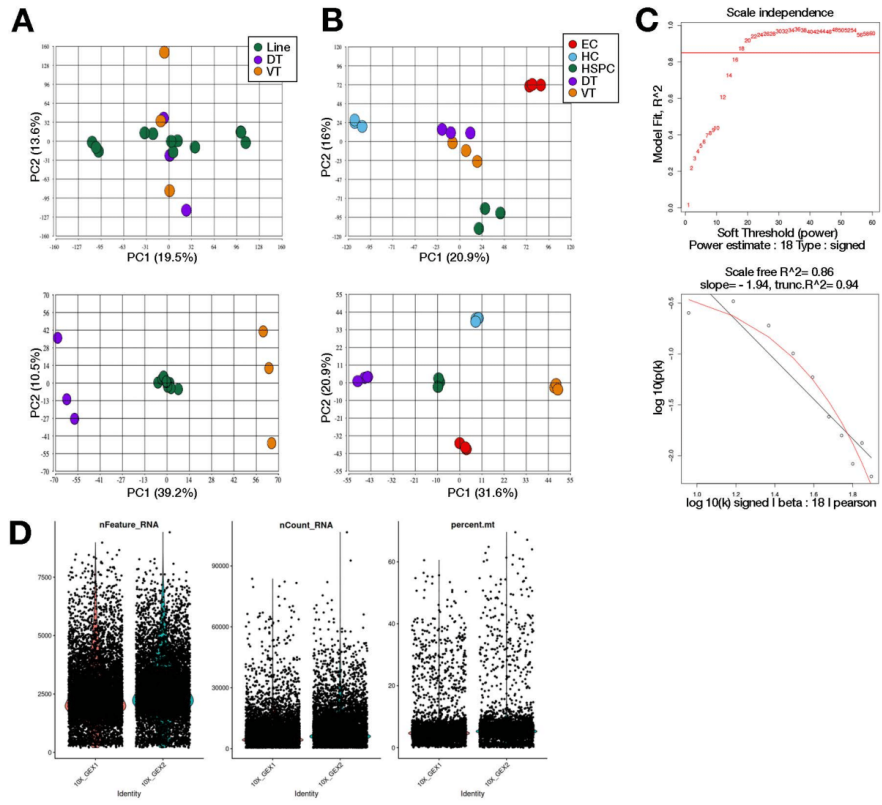


Figure S2

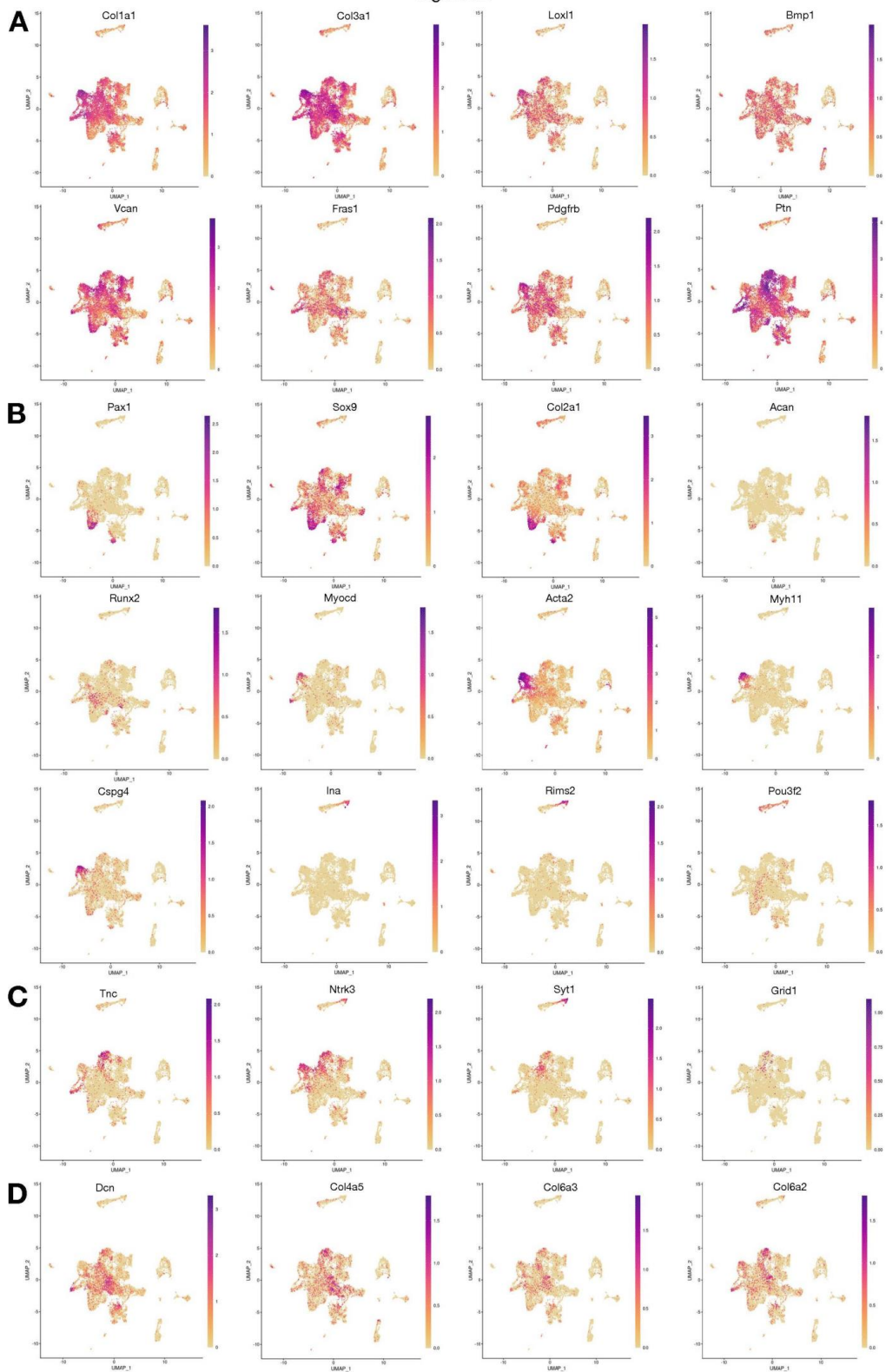


Figure S3

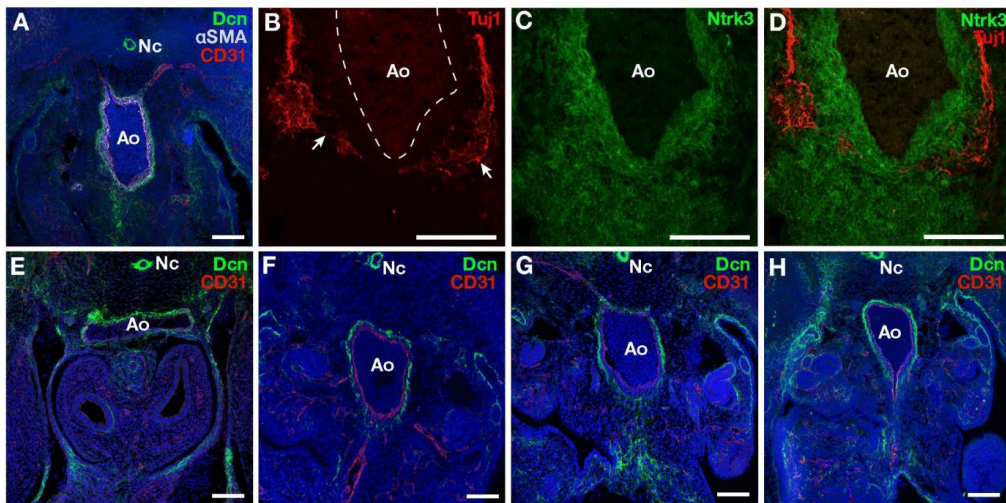
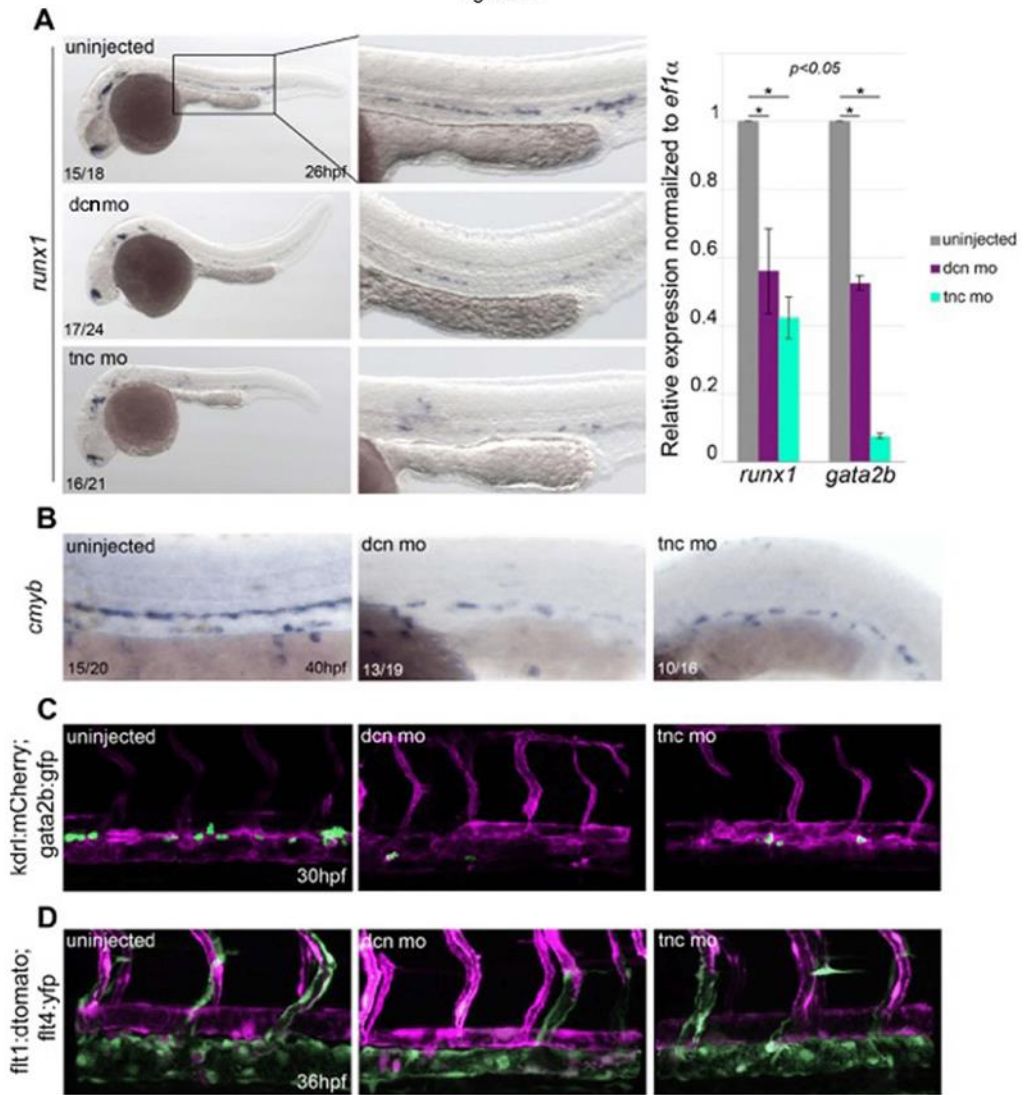


Figure S4



Discussion et perspectives

Ce travail de thèse a permis de caractériser le premier protocole de différenciation *in vitro* de cellules souches pluripotentes induites en cellules souches hématopoïétiques sans utilisation de transgènes ou de cocultures. Bien que ce dernier ait initialement été conçu pour une différenciation dirigée, nous avons pu montrer que les corps embryoïdes formés produisent des cellules des trois feuillets embryonnaires, ainsi que des cellules extra-embryonnaires. La présence d'une telle diversité cellulaire est inhabituelle, aussi bien dans les corps embryoïdes (Han et al., 2018) que dans les tératomes (McDonald et al., 2020). Même s'il existe des protocoles de différenciation des cellules souches embryonnaires ou induites vers les différents feuillets ou vers des types cellulaires extra-embryonnaires, la coexistence de ces différents éléments est peu courante. A notre connaissance, le seul système *in vitro* de différenciation cellulaire ayant réussi une différenciation simultanée vers des cellules embryonnaire et extra-embryonnaires humaines est le blastoïde (Kagawa et al., 2022), mais les types cellulaires représentés sont beaucoup moins matures que dans notre cas. Une différenciation vers les trois feuillets embryonnaires et la lignée trophoblastique a également été rapportée dans les tératomes réalisés à partir de cellules souches reprogrammées à un stade proche de l'embryon humain à 8 cellules (8CLC pour *8-cell-like-cells*) (Mazid et al., 2022). Les gènes de totipotence exprimés lors de cette reprogrammation ne sont pas détectés au 4^{ème} jour du protocole de différenciation des corps embryoïdes, qui correspond au jeu de données le plus précoce dont nous disposons pour nos analyses de scRNAseq. Cependant, la faible granulosité temporelle dont nous disposons n'exclut pas que notre milieu de culture induise un passage fugace par cet état. En effet, l'analyse scRNAseq d'un protocole de différenciation de cellules souches en cellules mésodermiques extra-embryonnaires publiée en 2022 montre un passage par le stade 8CLC (Pham et al., 2022) ; il est donc possible que ce dernier soit un point de passage privilégié pour la différenciation vers des populations extra-embryonnaires.

La présence de ces différentes populations dans les corps embryoïdes semble être liée à un facteur encore inconnu présent dans la source de protéine utilisée puisque le changement de la source de plasma induit la disparition des populations ectodermiques et extra-embryonnaires. L'identification de ce facteur pourrait être une piste intéressante pour la mise au point de protocoles de différenciation complexes incorporant simultanément des cellules intra et extra-embryonnaires. Nous avons donc commencé une analyse protéomique comparative des plasmas EFS et Octapharma afin d'identifier ce facteur.

Un point non élucidé par nos analyses est la séquence d'apparition des cellules du syncytiotrophoblaste et de l'amnios dans nos corps embryoïdes. Si cette dernière est simultanée

dans le protocole d'Ohgushi et al. (Ohgushi et al., 2022), l'absence de jeux de données intermédiaires entre les jours 7 et 13 ne nous permet pas de vérifier si un phénomène similaire a lieu dans les corps embryoïdes. En effet, l'amnios se formant à partir des cellules épiblastiques au contact du syncytiotrophoblaste, il est possible que cette interaction soit reproduite dans les corps embryoïdes. Nous avons prévu d'étudier ce point par des études en immunofluorescence de coupes des corps embryoïdes entre les jours 7 et 13 du protocole de différenciation.

Les cellules souches hématopoïétiques produites par notre protocole ont un phénotype intermédiaire entre celles détectées au CS10 et au CS14 dans l'embryon humain. Bien qu'elles soient très peu nombreuses (environ 1/10 000 dans le jeu de données scRNAseq et 1/16 000 d'après les expériences de dilution limite), elles permettent d'assurer une prise de greffe multi-lignage robuste chez des récipients primaires, secondaires et tertiaires. Cette efficacité peut-être rapprochée des expériences de greffes de cellules souches hématopoïétiques issues de l'aorte fœtale réalisées par le laboratoire d'Alexandre Medvinsky (Ivanovs et al., 2011, 2014). Malgré le fait que l'aorte fœtale ne contienne qu'une à trois cellules souches hématopoïétiques, ces dernières possèdent également un fort potentiel de reconstitution hématopoïétique.

Dans nos corps embryoïdes, les cellules souches hématopoïétiques semblent émerger d'un endothélium hémogénique non artériel APLNR⁺, puisque certaines cellules endothéliales dans le cluster de l'endothélium APLNR⁺ sont colocalisées avec les cellules hématopoïétiques sur projection UMAP. De plus, l'expression du facteur de transcription RUNX1 n'est détectée que dans l'endothélium APLNR⁺. Ainsi, bien que les cellules souches hématopoïétiques n'aient été détectées dans l'embryon humain qu'à partir du CS13 (Ivanovs et al., 2011; Tavian et al., 1999), où l'endothélium hémogénique a été caractérisé comme artériel, il semblerait que le phénotype artériel ne soit pas nécessaire à l'endothélium hémogénique pour la production de cellules souches hématopoïétiques via une transition endothélio-hématopoïétique. Il serait donc possible que l'endothélium hémogénique précoce mis en évidence au CS10 (Zeng et al., 2019) soit compétent pour la formation de cellules souches hématopoïétiques.

La question de la réalisation de ce potentiel *in vivo* reste ouverte du fait du faible nombre d'échantillons dont nous disposons, ainsi que du manque de techniques permettant d'étudier finement et de caractériser *in vitro* les cellules souches hématopoïétiques fonctionnelles. La signature transcriptionnelle de six gènes des cellules souches hématopoïétiques embryonnaires décrite dans Calvanese et al. (Calvanese et al., 2022), bien qu'ayant permis de restreindre fortement les cellules candidates sur le plan transcriptomique, n'a pas encore été validée *in vivo*. On peut donc s'interroger sur le caractère nécessaire ou suffisant de cette signature. Des cellules détectées au CS7 (Tyser et al., 2021) lors de la première vague d'hématopoïèse définitive dans le sac vitellin ou au CS10 qui expriment cinq des six gènes de cette signature (le gène HOXA9 n'est

pas ou très faiblement détecté) pourraient représenter de potentiels candidats. Les expériences réalisées par le laboratoire d'Alexander Medvinsky en 2011 (Ivanovs et al., 2011) n'ont détecté de cellules souches hématopoïétiques qu'à partir du CS14, mais la fonctionnalité de ces dernières a été testée par injection intraveineuse dans la queue des souris, ce qui suppose une capacité de d'autoguidage et de nidation dans la moelle osseuse médiée par CXCR4 (A. Peled et al., 1999) et PROCR (Gur-Cohen et al., 2015). Les données récoltées par le laboratoire d'Hanna Mikkola (Calvanese et al., 2022) montrent une expression plus faible de PROCR avant le CS13, il est donc possible que les éventuelles cellules souches hématopoïétiques fonctionnelles ne soient pas en capacité de coloniser la moelle suite à une injection intraveineuse. Une méthode permettant de tester cette hypothèse serait donc de reproduire ces expériences en procédant à des greffes intra-fémorales afin de vérifier si les cellules produites par ces endothéliums hémogéniques précoces sont en capacité de repeupler les moelles osseuses.

Dans le cas où les cellules produites à ces stades plus précoces (entre le CS7 et le CS12) ne contiendraient pas de cellules souches hématopoïétiques, cela signifierait que les différences entre les cellules produites au CS10 et dans nos corps embryoides marquent une réelle différence de fonctionnalité. Certaines de nos cellules exprimant HOXA9 (qui est absent au CS10), l'expression de ce dernier pourrait être un marqueur du potentiel des cellules souches hématopoïétique, ce qui serait cohérent avec sa présence dans la liste des marqueurs de transcription utilisé par Georges Daley (Doulatov et al., 2013; Sugimura et al., 2017) pour produire ses cellules souches hématopoïétiques ou les travaux d'Andrew Elefanty (Ng et al., 2016).

La plasticité du phénotype de l'endothélium hémogénique mis en évidence par ces différents résultats semble indiquer un rôle majeur des signaux extrinsèques à l'endothélium hémogénique dans la naissance des cellules souches hématopoïétiques. Bien que la présence de systèmes de régulation intrinsèques ne soit pas à exclure, la possibilité d'une transition endothélio-hématopoïétique depuis un endothélium précoce ou présentant un phénotype artériel, couplé au faible nombre de cellules souches hématopoïétiques générées dans l'embryon, nous amène à penser que les signaux du micro-environnement sont le facteur décisif de la production des cellules souches hématopoïétiques. En effet, outre notre protocole, les seuls protocoles de différenciation ayant réussi à produire des cellules souches hématopoïétiques à partir de cellules souches pluripotentes sont les protocoles utilisant la formation de tératomes (Amabile et al., 2013; Suzuki et al., 2013). Le point commun entre ces protocoles est une très grande diversité cellulaire composée de plusieurs feuilletts embryonnaires. Ces microenvironnements très riches sont plus à mêmes de récapituler les communications cellulaires complexes ayant lieu dans l'aorte fœtale et ainsi de donner les signaux nécessaires à la production et la mise en condition des cellules souches hématopoïétiques. La plupart des autres protocoles visant à produire des cellules

souches hématopoïétiques induisent une différenciation mésodermique, puis endothéliale et enfin hématopoïétique. Bien que cette stratégie de différenciation guidée par paliers ait porté ses fruits pour la génération de types cellulaires différenciées matures, il est probable que les cellules souches hématopoïétiques, de par le caractère instable de leur indifférenciation régulée par des signaux complexes du micro-environnement, ne peuvent que difficilement être générées via cette procédure. Il est beaucoup plus difficile de supplanter ou de mimer la multitude de signaux générés dans l'embryon, dont nous n'avons pas encore une connaissance parfaite, que de faire générer ces derniers par d'autres cellules. Dans notre cas, nous avons ainsi pu identifier grâce aux analyses scRNAseq des dialogues entre les différents feuillet et les cellules endothéliales ou hématopoïétiques. Par exemple, il est possible de détecter l'expression du gène TH dans les populations neuronales des corps embryoïdes. Ce dernier régule la production des catécholamines qui sont sécrétées par le système sympathique dans l'AGM et contribuent à la formation des cellules souches hématopoïétiques (Fitch et al., 2012). Les cellules épiblastiques et extra-embryonnaires sont responsables de la production d'ostéopontine (SPP1) qui joue un rôle crucial dans la prolifération des cellules souches hématopoïétiques (Nilsson et al., 2005) et les cellules endodermiques produisent l'IGF1 qui a été identifié comme étant un des acteurs clefs de la niche du foie fœtal (Gao et al., 2022).

L'importance de ses différents facteurs de régulation qui ont été identifiés grâce aux analyses scRNAseq réalisées sur les AGM embryonnaires ou sur les corps embryoïdes devront faire l'objet d'une validation fonctionnelle. Bien que beaucoup de ces expériences soient réalisables dans le modèle murin ou aviaire, le modèle des corps embryoïdes offre désormais la possibilité de les étudier dans un modèle humain et d'explorer le domaine des différences inter-espèces. Si les jeux de données d'aorte embryonnaire humaine permettent déjà d'explorer quels types cellulaires sont conservées et quels signaux sont spécifiques à chaque espèce, les corps embryoïdes pourraient offrir un système de validation à ces interactions.

Les analyses scRNAseq portent la promesse de déchiffrer les mécanismes moléculaires du développement à une échelle jamais atteinte auparavant. Néanmoins, au-delà des limitations techniques et théoriques de cette technologie, une de ses limites majeures réside dans l'exploitation des résultats. La quantité d'information pertinente contenue dans un jeu de données est colossale mais n'est malheureusement que rarement exploitée à son plein potentiel, soit par manque d'expertise, de techniques d'analyses ou de temps. Le traitement des jeux de données scRNAseq a ainsi malheureusement tendance à se restreindre à une caractérisation plus ou moins détaillée des populations présentes suivie d'analyses plus précises afin de répondre à une ou deux hypothèses de travail, laissant de côté une grande partie de l'information totale. Dans le cas du jeu de données d'aorte embryonnaire murine, notre analyse s'est focalisée sur la population sous-

aortique au profil neuro-mésenchymateux ; mais d'autres sous-populations stromales de l'aorte foétale restent à caractériser. Par exemple, si l'on s'intéresse à l'expression du gène RUNX1, qui est également exprimé dans le mésenchyme sous-aortique chez la souris (Fadlullah et al., 2022), on peut identifier ce dernier dans les péricytes et une population mésenchymateuse exprimant des gènes associés à un profil gonadique (Inha, Inhbb, Amhr2, Gata4). Cette population au profil inattendue semble également être présente chez l'humain puisqu'il est possible de détecter une population stromale AMHR2⁺GATA4⁺ au CS15 et semble donc être un candidat intéressant pour des investigations futures. En dehors de ces deux populations, plusieurs clusters mériteraient sûrement notre attention afin de caractériser pleinement les populations sous-aortiques et leurs interactions. Ainsi, bien que de plus en plus d'atlas soient réalisés chaque année et nous donnent accès à une cartographie transcriptionnelle du développement embryonnaire, la majorité des informations qu'ils renferment reste encore à exploiter.

Sur le plan thérapeutique, notre protocole de différenciation fait l'objet d'une montée en échelle dans des conditions compatibles avec les bonnes pratiques de fabrication (ou GMP pour *Good Manufacturing Practices*) par la plateforme Atlantic Bio GMP de l'EFS. Ces cellules serviront à la preuve de concept de la prise de greffe hématopoïétique dans un modèle gros animal. Avec la réussite de cette dernière étape, nous démontrerons que notre protocole est en capacité de produire des cellules souches hématopoïétiques en quantité suffisante pour une application clinique.

Bibliographie

- Amabile, G., Welner, R. S., Nombela-Arrieta, C., D'Alise, A. M., di Ruscio, A., Ebralidze, A. K., Kraytsberg, Y., Ye, M., Kocher, O., Neuberg, D. S., Khrapko, K., Silberstein, L. E., & Tenen, D. G. (2013). In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. *Blood*, *121* (8), 1255–1264. <https://doi.org/10.1182/blood-2012-06-434407>
- Amadei, G., Handford, C. E., Qiu, C., de Jonghe, J., Greenfeld, H., Tran, M., Martin, B. K., Chen, D.-Y., Aguilera-Castrejon, A., Hanna, J. H., Elowitz, M. B., Hollfelder, F., Shendure, J., Glover, D. M., & Zernicka-Goetz, M. (2022). Embryo model completes gastrulation to neurulation and organogenesis. *Nature*, *610* (7930), 143–153. <https://doi.org/10.1038/s41586-022-05246-3>
- Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Sonesson, C., Waldron, L., Pagès, H., Smith, M. L., Huber, W., Morgan, M., Gottardo, R., & Hicks, S. C. (2020). Orchestrating single-cell analysis with Bioconductor. *Nature Methods*, *17* (2), 137–145. <https://doi.org/10.1038/s41592-019-0654-x>
- Anjos-Afonso, F., Buettner, F., Mian, S. A., Rhys, H., Perez-Lloret, J., Garcia-Albornoz, M., Rastogi, N., Ariza-McNaughton, L., & Bonnet, D. (2022). Single cell analyses identify a highly regenerative and homogenous human CD34+ hematopoietic stem cell population. *Nature Communications*, *13* (1), 1–13. <https://doi.org/10.1038/s41467-022-29675-w>
- Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R. P., Wolters, P. J., Abate, A. R., Butte, A. J., & Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, *20* (2), 163–172. <https://doi.org/10.1038/s41590-018-0276-y>
- Arcangeli, M.-L., Brault, P., Bourhis, J.-H., Kuhnowskie, F., Henry, E., Barroca, V., Koscielny, S., Pflumio, F., & Amsellem, S. (2020). Combined G-CSF and Plerixafor enhance hematopoietic recovery of CD34+ cells from poor mobilizer patients in NSG mice. *Experimental Hematology*, *86*, 15-20.e2. <https://doi.org/10.1016/j.exphem.2020.05.006>
- Barile, M., Imaz-Rosshandler, I., Inzani, I., Ghazanfar, S., Nichols, J., Marioni, J. C., Guibentif, C., & Göttgens, B. (2021). Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. *Genome Biology*, *22* (1), 1–22. <https://doi.org/10.1186/s13059-021-02414-y>
- Barshan, E., Ghodsi, A., Azimifar, Z., & Zolghadri Jahromi, M. (2011). Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, *44* (7), 1357–1371. <https://doi.org/10.1016/j.patcog.2010.12.015>
- Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M., & Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. *Proceedings of the National Academy of Sciences*, *89* (7), 2804–2808. <https://doi.org/10.1073/pnas.89.7.2804>

- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, *37* (1), 38–44. <https://doi.org/10.1038/nbt.4314>
- Becker, A. J., McCulloch, E. A., & Till, J. E. (1963). Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. *Nature*, *197* (4866), 452–454. <https://doi.org/10.1038/197452a0>
- Belkin, M., & Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, *15* (6), 1373–1396. <https://doi.org/10.1162/089976603321780317>
- Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-020-0591-3>
- Bergen, V., Soldatov, R. A., Kharchenko, P. v, & Theis, F. J. (2021). RNA velocity—current challenges and future perspectives. *Molecular Systems Biology*, *17* (8). <https://doi.org/10.15252/msb.202110282>
- Bergmann, S., Penfold, C. A., Slatery, E., Siriwardena, D., Drummer, C., Clark, S., Strawbridge, S. E., Kishimoto, K., Vickers, A., Tewary, M., Kohler, T. N., Hollfelder, F., Reik, W., Sasaki, E., Behr, R., & Boroviak, T. E. (2022). Spatial profiling of early primate gastrulation in utero. *Nature*, *609* (7925), 136–143. <https://doi.org/10.1038/s41586-022-04953-1>
- Bernardo, A. S., Faial, T., Gardner, L., Niakan, K. K., Ortmann, D., Senner, C. E., Callery, E. M., Trotter, M. W., Hemberger, M., Smith, J. C., Bardwell, L., Moffett, A., & Pedersen, R. A. (2011). Brachyuri and Cdx2 Mediate BMP-Induced Differentiation of Human and Mouse Pluripotent Stem Cells into Embryonic and Extraembryonic Lineages. *Cell Stem Cell*, *9* (2), 144–155. <https://doi.org/10.1016/j.stem.2011.06.015>
- Bhatia, M., Bonnet, D., Kapp, U., Wang, J. C. Y., Murdoch, B., & Dick, J. E. (1997). Quantitative Analysis Reveals Expansion of Human Hematopoietic Repopulating Cells After Short-term Ex Vivo Culture. *Journal of Experimental Medicine*, *186* (4), 619–624. <https://doi.org/10.1084/jem.186.4.619>
- Blaese, R. M., Culver, K. W., Miller, A. D., Carter, C. S., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P., Greenblatt, J. J., Rosenberg, S. A., Klein, H., Berger, M., Mullen, C. A., Ramsey, W. J., Muul, L., Morgan, R. A., & Anderson, W. F. (1995). T Lymphocyte-Directed Gene Therapy for ADA- SCID: Initial Trial Results After 4 Years. *Science*, *270* (5235), 475–480. <https://doi.org/10.1126/science.270.5235.475>
- Blau, H. M., & Daley, G. Q. (2019). Stem cells in the treatment of disease. *New England Journal of Medicine*, *380* (18), 1748–1760. <https://doi.org/10.1056/NEJMra1716145>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). *Fast unfolding of communities in large networks*. <https://doi.org/10.1088/1742-5468/2008/10/P10008>

- Bloom, W., & Bartelmez, G. W. (1940). Hematopoiesis in young human embryos. *American Journal of Anatomy*, 67 (1), 21–53. <https://doi.org/10.1002/aja.1000670103>
- Boitano, A. E., Wang, J., Romeo, R., Bouchez, L. C., Parker, A. E., Sutton, S. E., Walker, J. R., Flaveny, C. A., Perdew, G. H., Denison, M. S., Schultz, P. G., & Cooke, M. P. (2010). Aryl Hydrocarbon Receptor Antagonists Promote the Expansion of Human Hematopoietic Stem Cells. *Science*, 329 (5997), 1345–1348. <https://doi.org/10.1126/science.1191536>
- Bowles, K. M., Vallier, L., Smith, J. R., Alexander, M. R. J., & Pedersen, R. A. (2006). HOXB4 Overexpression Promotes Hematopoietic Development by Human Embryonic Stem Cells. *STEM CELLS*, 24 (5), 1359–1369. <https://doi.org/10.1634/stemcells.2005-0210>
- Brinster, R. L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *Journal of Experimental Medicine*, 140 (4), 1049–1056. <https://doi.org/10.1084/jem.140.4.1049>
- Broxmeyer, H. E., Orschell, C. M., Clapp, D. W., Hangoc, G., Cooper, S., Plett, P. A., Liles, W. C., Li, X., Graham-Evans, B., Campbell, T. B., Calandra, G., Bridger, G., Dale, D. C., & Srour, E. F. (2005). Rapid mobilization of murine and human hematopoietic stem and progenitor cells with AMD3100, a CXCR4 antagonist. *Journal of Experimental Medicine*, 201 (8), 1307–1318. <https://doi.org/10.1084/jem.20041385>
- Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., & Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, 33 (2), 155–160. <https://doi.org/10.1038/nbt.3102>
- Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., & Rodewald, H.-R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 518 (7540), 542–546. <https://doi.org/10.1038/nature14242>
- Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E. S., Ekstrand, S., Aguadé-Gorgorió, J., Vavilina, A., Lefaudeux, D., Nadel, B., Li, J. Y., Wang, Y., Lee, L. K., Ardehali, R., Iruela-Arispe, M. L., Pellegrini, M., Stanley, E. G., Elefanty, A. G., ... Mikkola, H. K. A. (2022). Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature*, 604 (7906), 534–540. <https://doi.org/10.1038/s41586-022-04571-x>
- Carpenter, L., Malladi, R., Yang, C.-T., French, A., Pilkington, K. J., Forsey, R. W., Sloane-Stanley, J., Silk, K. M., Davies, T. J., Fairchild, P. J., Enver, T., Watt, S. M., & Dunn, W. (2011). Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. <https://doi.org/10.1182/blood-2010-08>
- Castel, G., Meistermann, D., Bretin, B., Firmin, J., Blin, J., Loubersac, S., Bruneau, A., Chevolleau, S., Kilens, S., Chariou, C., Gaignerie, A., Francheteau, Q., Kagawa, H., Charpentier, E., Flippe, L., François--Campion, V., Haider, S., Dietrich, B., Knöfler, M., ... David, L. (2020). Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. *Cell Reports*, 33 (8). <https://doi.org/10.1016/j.celrep.2020.108419>

- Catlin, S. N., Busque, L., Gale, R. E., Gutter, P., & Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells in vivo. *Blood*, *117* (17), 4460–4466. <https://doi.org/10.1182/blood-2010-08-303537>
- Cavazzana-Calvo, M., Hacein-Bey, S., Basile, G. de saint, Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.-L., Bousso, P., Deist, F. le, & Fischer, A. (2000). Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science*, *288* (5466), 669–672. <https://doi.org/10.1126/science.288.5466.669>
- Cerdan, C., McIntyre, B. A. S., Mechael, R., Levadoux-Martin, M., Yang, J., Lee, J. B., & Bhatia, M. (2012). Activin A Promotes Hematopoietic Fated Mesoderm Development Through Upregulation of Brachyury in Human Embryonic Stem Cells. *Stem Cells and Development*, *21* (15), 2866–2877. <https://doi.org/10.1089/scd.2012.0053>
- Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A., & Bhatia, M. (2003). Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood*, *102* (3), 906–915. <https://doi.org/10.1182/blood-2003-03-0832>
- Charbord, P., Tavian, M., Humeau, L., & Péault, B. (1996). Early ontogeny of the human marrow from long bones: an immunohistochemical study of hematopoiesis and its microenvironment. *Blood*, *87* (10), 4109–4119. <http://www.ncbi.nlm.nih.gov/pubmed/8639768>
- Chari, T., Banerjee, J., & Pachter, L. (2021). The Specious Art of Single-Cell Genomics. *BioRxiv*, 1–23. <https://doi.org/10.1101/2021.08.25.457696>
- Chen, M. J., Li, Y., De Obaldia, M. E., Yang, Q., Yzaguirre, A. D., Yamada-Inagawa, T., Vink, C. S., Bhandoola, A., Dzierzak, E., & Speck, N. A. (2011). Erythroid/Myeloid Progenitors and Hematopoietic Stem Cells Originate from Distinct Populations of Endothelial Cells. *Cell Stem Cell*, *9* (6), 541–552. <https://doi.org/10.1016/j.stem.2011.10.003>
- Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E., & Speck, N. A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*, *457* (7231), 887–891. <https://doi.org/10.1038/nature07619>
- Chen, S., Rivaud, P., Park, J. H., Tsou, T., Charles, E., Haliburton, J. R., Pichiorri, F., & Thomson, M. (2020). Dissecting heterogeneous cell populations across drug and disease conditions with PopAlign. *Proceedings of the National Academy of Sciences*, *117* (46), 28784–28794. <https://doi.org/10.1073/pnas.2005990117>
- Chen, W., Morabito, S. J., Kessenbrock, K., Enver, T., Meyer, K. B., & Teschendorff, A. E. (2019). Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. *Communications Biology*, *2* (1). <https://doi.org/10.1038/s42003-019-0554-8>

- Cheng, H., Liang, P. H., & Cheng, T. (2013). Mouse Hematopoietic Stem Cell Transplantation. In *Methods in Molecular Biology* (Vol. 976, pp. 25–35). Humana Press Inc. https://doi.org/10.1007/978-1-62703-317-6_3
- Choudhary, S., & Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biology*, *23* (1), 1–20. <https://doi.org/10.1186/s13059-021-02584-9>
- Ciau-Uitz, A., Walmsley, M., & Patient, R. (2000). Distinct Origins of Adult and Embryonic Blood in Xenopus. *Cell*, *102* (6), 787–796. [https://doi.org/10.1016/S0092-8674\(00\)00067-2](https://doi.org/10.1016/S0092-8674(00)00067-2)
- Cinkornpumin, J. K., Kwon, S. Y., Guo, Y., Hossain, I., Sirois, J., Russett, C. S., Tseng, H.-W., Okae, H., Arima, T., Duchaine, T. F., Liu, W., & Pastor, W. A. (2020). Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports*, *15* (1), 198–213. <https://doi.org/10.1016/j.stemcr.2020.06.003>
- Civin, C. I., Strauss, L. C., Brovall, C., Fackler, M. J., Schwartz, J. F., & Shaper, J. H. (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *The Journal of Immunology*, *133* (1), 157–165. <https://doi.org/10.4049/jimmunol.133.1.157>
- Cohen, S., Roy, J., Lachance, S., Delisle, J.-S., Marinier, A., Busque, L., Roy, D.-C., Barabé, F., Ahmad, I., Bambace, N., Bernard, L., Kiss, T., Bouchard, P., Caudrelier, P., Landais, S., Larochelle, F., Chagraoui, J., Lehnertz, B., Corneau, S., ... Sauvageau, G. (2020). Hematopoietic stem cell transplantation using single UM171-expanded cord blood: a single-arm, phase 1–2 safety and feasibility study. *The Lancet Haematology*, *7* (2), e134–e145. [https://doi.org/10.1016/S2352-3026\(19\)30202-9](https://doi.org/10.1016/S2352-3026(19)30202-9)
- Conneally, E., Cashman, J., Petzer, A., & Eaves, C. (1997). Expansion in vitro of transplantable human cord blood stem cells demonstrated using a quantitative assay of their lympho-myeloid repopulating activity in nonobese diabetic scid/scid mice. *Proceedings of the National Academy of Sciences*, *94* (18), 9836–9841. <https://doi.org/10.1073/pnas.94.18.9836>
- Cooley, S. M., Hamilton, T., Deeds, E. J., & Ray, J. C. J. (2019). A novel metric reveals previously unrecognized distortion in dimensionality reduction of scRNA-Seq data. *BioRxiv*, 1–36. <https://doi.org/10.1101/689851>
- Cosgrove, J., Hustin, L. S. P., de Boer, R. J., & Perié, L. (2021). Hematopoiesis in numbers. In *Trends in Immunology* (Vol. 42, Issue 12, pp. 1100–1112). Elsevier Ltd. <https://doi.org/10.1016/j.it.2021.10.006>
- Crosse, E. I., Gordon-Keylock, S., Rybtsov, S., Binagui-Casas, A., Felchle, H., Nnadi, N. C., Kirschner, K., Chandra, T., Tamagno, S., Webb, D. J., Rossi, F., Anderson, R. A., & Medvinsky, A. (2020). Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human Hematopoietic Stem Cell Development. *Cell Stem Cell*, *27* (5), 822–839.e8. <https://doi.org/10.1016/j.stem.2020.08.004>

- Crowell, H. L., Sonesson, C., Germain, P. L., Calini, D., Collin, L., Raposo, C., Malhotra, D., & Robinson, M. D. (2020). muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications*, *11* (1). <https://doi.org/10.1038/s41467-020-19894-4>
- Cumano, A., Dieterlen-Lievre, F., & Godin, I. (1996). Lymphoid Potential, Probed before Circulation in Mouse, Is Restricted to Caudal Intraembryonic Splanchnopleura. *Cell*, *86* (6), 907–916. [https://doi.org/10.1016/S0092-8674\(00\)80166-X](https://doi.org/10.1016/S0092-8674(00)80166-X)
- Cutler, C., Multani, P., Robbins, D., Kim, H. T., Le, T., Hoggatt, J., Pelus, L. M., Desponts, C., Chen, Y.-B., Rezner, B., Armand, P., Koreth, J., Glotzbecker, B., Ho, V. T., Alyea, E., Isom, M., Kao, G., Armant, M., Silberstein, L., ... Shoemaker, D. D. (2013). Prostaglandin-modulated umbilical cord blood hematopoietic stem cell transplantation. *Blood*, *122* (17), 3074–3081. <https://doi.org/10.1182/blood-2013-05-503177>
- Danic, B., & Lefrère, J.-J. (2010). Le sang dans l'art, l'art dans le sang. *Transfusion Clinique et Biologique*, *17* (5–6), 382–385. <https://doi.org/10.1016/j.tracli.2010.09.152>
- de Lima, M., McMannis, J., Gee, A., Komanduri, K., Couriel, D., Andersson, B. S., Hosing, C., Khouri, I., Jones, R., Champlin, R., Karandish, S., Sadeghi, T., Peled, T., Grynszan, F., Daniely, Y., Nagler, A., & Shpall, E. J. (2008). Transplantation of ex vivo expanded cord blood cells using the copper chelator tetraethylenepentamine: a phase I/II clinical trial. *Bone Marrow Transplantation*, *41* (9), 771–778. <https://doi.org/10.1038/sj.bmt.1705979>
- Deslauriers, A. G., Kotini, A. G., & Papapetrou, E. P. (2021). Modeling Leukemia Stem Cells with Patient-Derived Induced Pluripotent Stem Cells. In *Methods in Molecular Biology* (Vol. 2185, pp. 411–422). Humana Press Inc. https://doi.org/10.1007/978-1-0716-0810-4_26
- Dewey, M. J., Martin, D. W., Martinf, G. R., & Mintz, B. (1977). *Cell Biology Mosaic mice with teratocarcinoma-derived mutant cells deficient in hypoxanthine phosphoribosyltransferase (Lesch-Nyhan disease/cell selection/trisomy/allophenic mice/blastocyst injection)* (Vol. 74, Issue 12). <https://www.pnas.org>
- Dick, J. E., Lapidot, T., & Pflumio, F. (1991). Transplantation of Normal and Leukemic Human Bone Marrow into Immune-Deficient Mice: Development of Animal Models for Human Hematopoiesis. *Immunological Reviews*, *124* (1), 25–43. <https://doi.org/10.1111/j.1600-065X.1991.tb00614.x>
- Dieterlen-Lievre, F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. *Journal of Embryology and Experimental Morphology*, *33* (3), 607–619. <http://www.ncbi.nlm.nih.gov/pubmed/1176862>
- Ding, J., Li, Y., & Larochelle, A. (2023). De Novo Generation of Human Hematopoietic Stem Cells from Pluripotent Stem Cells for Cellular Therapy. *Cells*, *12* (2), 321. <https://doi.org/10.3390/cells12020321>
- Ditadi, A., & Sturgeon, C. M. (2023). Back to the future: lessons from development drive innovation of human pluripotent stem cell therapies. *Experimental Hematology*, *117*, 9–14. <https://doi.org/10.1016/j.exphem.2022.11.003>

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29* (1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong, G., Xu, X., Li, Y., Ouyang, W., Zhao, W., Gu, Y., Li, J., Liu, T., Zeng, X., Zou, H., Wang, S., Chen, Y., Liu, S., Sun, H., & Liu, C. (2023). Stemness-related genes revealed by single-cell profiling of naïve and stimulated human CD34 + cells from CB and mPB. *Clinical and Translational Medicine*, *13* (1). <https://doi.org/10.1002/ctm2.1175>
- Doulatov, S., Vo, L. T., Chou, S. S., Kim, P. G., Arora, N., Li, H., Hadland, B. K., Bernstein, I. D., Collins, J. J., Zon, L. I., & Daley, G. Q. (2013). Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell*, *13* (4), 459–470. <https://doi.org/10.1016/j.stem.2013.09.002>
- Duò, A., Robinson, M. D., & Soneson, C. (2018). A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research*, *7*, 1141. <https://doi.org/10.12688/f1000research.15666.1>
- Enders, A. C., & King, B. F. (1988). Formation and differentiation of extraembryonic mesoderm in the rhesus monkey. *American Journal of Anatomy*, *181* (4), 327–340. <https://doi.org/10.1002/aja.1001810402>
- Evain-Brion, D., & Malassine, A. (2003). Human placenta as an endocrine organ. *Growth Hormone and IGF Research*, *13* (SUPPL. A). [https://doi.org/10.1016/S1096-6374\(03\)00053-4](https://doi.org/10.1016/S1096-6374(03)00053-4)
- Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, *292* (5819), 154–156. <https://doi.org/10.1038/292154a0>
- Fadlullah, M. Z. H., Neo, W. H., Lie-a-ling, M., Thambyrajah, R., Patel, R., Mevel, R., Aksoy, I., do Khoa, N., Savatier, P., Fontenille, L., Baker, S. M., Rattray, M., Kouskoff, V., & Lacaud, G. (2022). Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE. *Blood*, *139* (3), 343–356. <https://doi.org/10.1182/blood.2020007885>
- Fares, I., Chagraoui, J., Gareau, Y., Gingras, S., Ruel, R., Mayotte, N., Csaszar, E., Knapp, D. J. H. F., Miller, P., Ngom, M., Imren, S., Roy, D.-C., Watts, K. L., Kiem, H.-P., Herrington, R., Iscove, N. N., Humphries, R. K., Eaves, C. J., Cohen, S., ... Sauvageau, G. (2014). Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science*, *345* (6203), 1509–1512. <https://doi.org/10.1126/science.1256337>
- Fehér, I., & Gidàli, J. (1974). Prostaglandin E2 as stimulator of haemopoietic stem cell proliferation. *Nature*, *247* (5442), 550–551. <https://doi.org/10.1038/247550a0>
- Feng, Q., Shabrani, N., Thon, J. N., Huo, H., Thiel, A., Machlus, K. R., Kim, K., Brooks, J., Li, F., Luo, C., Kimbrel, E. A., Wang, J., Kim, K. S., Italiano, J., Cho, J., Lu, S. J., & Lanza, R. (2014). Scalable generation of universal platelets from human induced pluripotent stem cells. *Stem Cell Reports*, *3* (5), 817–831. <https://doi.org/10.1016/j.stemcr.2014.09.010>

- Fitch, S. R., Kimber, G. M., Wilson, N. K., Parker, A., Mirshekar-Syahkal, B., Göttgens, B., Medvinsky, A., Dzierzak, E., & Ottersbach, K. (2012). Signaling from the Sympathetic Nervous System Regulates Hematopoietic Stem Cell Emergence during Embryogenesis. *Cell Stem Cell*, 11 (4), 554–566. <https://doi.org/10.1016/j.stem.2012.07.002>
- Flamier, A., Singh, S., & Rasmussen, T. P. (2017). A standardized human embryoid body platform for the detection and analysis of teratogens. *PLOS ONE*, 12 (2), e0171101. <https://doi.org/10.1371/journal.pone.0171101>
- Freytag, S., Tian, L., Lönnstedt, I., Ng, M., & Bahlo, M. (2018). Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Research*, 7, 1297. <https://doi.org/10.12688/f1000research.15809.1>
- Fukuda, T. (1973). Fetal hemopoiesis. I. Electron microscopic studies on human yolk sac hemopoiesis. *Virchows Archiv. B, Cell Pathology*, 14 (3), 197–213. <http://www.ncbi.nlm.nih.gov/pubmed/4203889>
- Ganuza, M., Hall, T., Myers, J., Nevitt, C., Sánchez-Lanzas, R., Chabot, A., Ding, J., Kooienga, E., Caprio, C., Finkelstein, D., Kang, G., Obeng, E., & McKinney-Freeman, S. (2022). Murine foetal liver supports limited detectable expansion of life-long haematopoietic progenitors. *Nature Cell Biology*, 24 (10), 1475–1486. <https://doi.org/10.1038/s41556-022-00999-5>
- Gao, S., Shi, Q., Zhang, Y., Liang, G., Kang, Z., Huang, B., Ma, D., Wang, L., Jiao, J., Fang, X., Xu, C.-R., Liu, L., Xu, X., Göttgens, B., Li, C., & Liu, F. (2022). Identification of HSC/MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Research*, 32 (1), 38–53. <https://doi.org/10.1038/s41422-021-00540-7>
- Gardner, R. L. (1968). Mouse Chimaeras obtained by the Injection of Cells into the Blastocyst. *Nature*, 220 (5167), 596–597. <https://doi.org/10.1038/220596a0>
- Germain, P. L., Sonrel, A., & Robinson, M. D. (2020). PipeComp, a general framework for the evaluation of computational pipelines, reveals performant single cell RNA-seq preprocessing tools. *Genome Biology*, 21 (1). <https://doi.org/10.1186/s13059-020-02136-7>
- Ghimire, S., Mantziou, V., Moris, N., & Martinez Arias, A. (2021). Human gastrulation: The embryo and its models. *Developmental Biology*, 474, 100–108. <https://doi.org/10.1016/j.ydbio.2021.01.006>
- Gilbert, S. F. (2000). *Developmental Biology* (6th Edition). Sinauer Associates.
- Glimm, H., Oh, I.-H., & Eaves, C. J. (2000). Human hematopoietic stem cells stimulated to proliferate in vitro lose engraftment potential during their S/G2/M transit and do not reenter G0. *Blood*, 96 (13), 4185–4193. <https://doi.org/10.1182/blood.V96.13.4185>
- Gluckman, E., Devergie, A., Schaison, G., Bussel, A., Berger, R., Sohler, J., & Bernard, J. (1980). Bone Marrow Transplantation in Fanconi Anaemia. *British Journal of Haematology*, 45 (4), 557–564. <https://doi.org/10.1111/j.1365-2141.1980.tb07178.x>

- Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Molecular Systems Biology*, *14* (4). <https://doi.org/10.15252/msb.20178046>
- Grün, D., Kester, L., & van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nature Methods*, *11* (6), 637–640. <https://doi.org/10.1038/nmeth.2930>
- Gur-Cohen, S., Itkin, T., Chakrabarty, S., Graf, C., Kollet, O., Ludin, A., Golan, K., Kalinkovich, A., Ledergor, G., Wong, E., Niemeyer, E., Porat, Z., Erez, A., Sagi, I., Esmon, C. T., Ruf, W., & Lapidot, T. (2015). PAR1 signaling regulates the retention and recruitment of EPCR-expressing bone marrow hematopoietic stem cells. *Nature Medicine*, *21* (11), 1307–1317. <https://doi.org/10.1038/nm.3960>
- Haeckel, E. (1879). *Natürliche Schöpfungsgeschichte*. DE GRUYTER. <https://doi.org/10.1515/9783111464503>
- Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, *20* (1). <https://doi.org/10.1186/s13059-019-1874-1>
- Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G. J., Larsson, A. J. M., Faridani, O. R., & Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology*, *38* (6), 708–714. <https://doi.org/10.1038/s41587-020-0497-0>
- Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, *36* (5), 421–427. <https://doi.org/10.1038/nbt.4091>
- Han, X., Chen, H., Huang, D., Chen, H., Fei, L., Cheng, C., Huang, H., Yuan, G. C., & Guo, G. (2018). Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biology*, *19* (1), 1–19. <https://doi.org/10.1186/s13059-018-1426-0>
- Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., Zhou, Y., Ye, F., Jiang, M., Wu, J., Xiao, Y., Jia, X., Zhang, T., Ma, X., Zhang, Q., ... Guo, G. (2020). Construction of a human cell landscape at single-cell level. *Nature*, *581* (7808), 303–309. <https://doi.org/10.1038/s41586-020-2157-4>
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184* (13), 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Harrison, D. E. (1975). Normal Function of Transplanted Marrow Cell Lines from Aged Mice. *Journal of Gerontology*, *30* (3), 279–285. <https://doi.org/10.1093/geronj/30.3.279>

- Harrison, D. E., Astle, C. M., & Delaittre, J. A. (1978). Loss of proliferative capacity in immunohemopoietic stem cells caused by serial transplantation rather than aging. *Journal of Experimental Medicine*, *147* (5), 1526–1531. <https://doi.org/10.1084/jem.147.5.1526>
- Hertig, A. T., Rock, J., & Adams, E. C. (1956). A description of 34 human ova within the first 17 days of development. *American Journal of Anatomy*, *98* (3), 435–493. <https://doi.org/10.1002/aja.1000980306>
- Himburg, H. A., Muramoto, G. G., Daher, P., Meadows, S. K., Russell, J. L., Doan, P., Chi, J.-T., Salter, A. B., Lento, W. E., Reya, T., Chao, N. J., & Chute, J. P. (2010). Pleiotrophin regulates the expansion and regeneration of hematopoietic stem cells. *Nature Medicine*, *16* (4), 475–482. <https://doi.org/10.1038/nm.2119>
- Horwitz, M. E., Chao, N. J., Rizzieri, D. A., Long, G. D., Sullivan, K. M., Gasparetto, C., Chute, J. P., Morris, A., McDonald, C., Waters-Pick, B., Stiff, P., Wease, S., Peled, A., Snyder, D., Cohen, E. G., Shoham, H., Landau, E., Friend, E., Peleg, I., ... Peled, T. (2014). Umbilical cord blood expansion with nicotinamide provides long-term multilineage engraftment. *Journal of Clinical Investigation*, *124* (7), 3121–3128. <https://doi.org/10.1172/JCI74556>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24* (6), 417–441. <https://doi.org/10.1037/h0071325>
- Howard, J. E., Smith, J. N. P., Fredman, G., & MacNamara, K. C. (2021). IL-18R-mediated HSC quiescence and MLKL-dependent cell death limit hematopoiesis during infection-induced shock. *Stem Cell Reports*, *16* (12), 2887–2899. <https://doi.org/10.1016/j.stemcr.2021.10.011>
- Huang, S., & Terstappen, L. (1994). Lymphoid and myeloid differentiation of single human CD34+, HLA-DR+, CD38- hematopoietic stem cells. *Blood*, *83* (6), 1515–1526. <https://doi.org/10.1182/blood.V83.6.1515.1515>
- Hwang, J. W., Desterke, C., Loisel-Duwattez, J., Griscelli, F., Bennaceur-Griscelli, A., & Turhan, A. G. (2021). Detection of Hematopoietic Stem Cell Transcriptome in Human Fetal Kidneys and Kidney Organoids Derived From Human Induced Pluripotent Stem Cells. *Frontiers in Cell and Developmental Biology*, *9*. <https://doi.org/10.3389/fcell.2021.668833>
- Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10* (3), 626–634. <https://doi.org/10.1109/72.761722>
- Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology*, *17* (1), 29. <https://doi.org/10.1186/s13059-016-0888-1>
- Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., & Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Molecular Medicine (Cambridge, Mass.)*, *6* (2), 88–95. <http://www.ncbi.nlm.nih.gov/pubmed/10859025>

- Ivanovs, A., Rybtsov, S., Anderson, R. A., Turner, M. L., & Medvinsky, A. (2014). Identification of the niche and phenotype of the first human hematopoietic stem cells. *Stem Cell Reports*, 2 (4), 449–456. <https://doi.org/10.1016/j.stemcr.2014.02.004>
- Ivanovs, A., Rybtsov, S., Ng, E. S., Stanley, E. G., Elefanty, A. G., & Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. *Development*, 144 (13), 2323–2337. <https://doi.org/10.1242/dev.134866>
- Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R. A., Turner, M. L., & Medvinsky, A. (2011). Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *Journal of Experimental Medicine*, 208 (12), 2417–2427. <https://doi.org/10.1084/jem.20111688>
- Jacob, A., Morley, M., Hawkins, F., McCauley, K. B., Jean, J. C., Heins, H., Na, C. L., Weaver, T. E., Vedaie, M., Hurley, K., Hinds, A., Russo, S. J., Kook, S., Zacharias, W., Ochs, M., Traber, K., Quinton, L. J., Crane, A., Davis, B. R., ... Kotton, D. N. (2017). Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell*, 21 (4), 472–488.e10. <https://doi.org/10.1016/j.stem.2017.08.014>
- Jacobson, L. O., Simmons, E. L., Marks, E. K., Robson, M. J., Bethard, W. F., & Gaston, E. O. (1950). The role of the spleen in radiation injury and recovery. *The Journal of Laboratory and Clinical Medicine*, 35 (5), 746–770.
- Jaffredo, T., Gautier, R., Eichmann, A., & Dieterlen-Lievre, F. (1998). Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development*, 125 (22), 4575–4583. <https://doi.org/10.1242/dev.125.22.4575>
- Jaremek, A., Jeyarajah, M. J., Jaju Bhattad, G., & Renaud, S. J. (2021). Omics Approaches to Study Formation and Function of Human Placental Syncytiotrophoblast. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.674162>
- Jordan, H. E. (1917). Aortic Cell Clusters in Vertebrate Embryos. *Proceedings of the National Academy of Sciences*, 3 (3), 149–156. <https://doi.org/10.1073/pnas.3.3.149>
- Julien, E., el Omar, R., & Tavian, M. (2016). Origin of the hematopoietic system in the human embryo. *FEBS Letters*, 590 (22), 3987–4001. <https://doi.org/10.1002/1873-3468.12389>
- Kagawa, H., Javali, A., Khoei, H. H., Sommer, T. M., Sestini, G., Novatchkova, M., Scholte op Reimer, Y., Castel, G., Bruneau, A., Maenhoudt, N., Lammers, J., Loubersac, S., Freour, T., Vankelecom, H., David, L., & Rivron, N. (2022). Human blastoids model blastocyst development and implantation. *Nature*, 601 (7894), 600–605. <https://doi.org/10.1038/s41586-021-04267-8>
- Kallinikou, K., Anjos-Afonso, F., Blundell, M. P., Ings, S. J., Watts, M. J., Thrasher, A. J., Linch, D. C., Bonnet, D., & Yong, K. L. (2012). Engraftment defect of cytokine-cultured adult human mobilized CD34+ cells is related to reduced adhesion to bone marrow niche elements. *British Journal of Haematology*, 158 (6), 778–787. <https://doi.org/10.1111/j.1365-2141.2012.09219.x>

- Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R., & Thomson, J. A. (2001). Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proceedings of the National Academy of Sciences*, *98* (19), 10716–10721. <https://doi.org/10.1073/pnas.191362598>
- Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S., & Keller, G. (2007). Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood*, *109* (7), 2679–2687. <https://doi.org/10.1182/blood-2006-09-047704>
- Kharchenko, P. v., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, *11* (7), 740–742. <https://doi.org/10.1038/nmeth.2967>
- Kim, T., Chen, I. R., Lin, Y., Wang, A. Y.-Y., Yang, J. Y. H., & Yang, P. (2019). Impact of similarity metrics on single-cell RNA-seq data clustering. *Briefings in Bioinformatics*, *20* (6), 2316–2326. <https://doi.org/10.1093/bib/bby076>
- Kim, T. H., Zhou, X., & Chen, M. (2020). Demystifying “drop-outs” in single-cell UMI data. *Genome Biology*, *21* (1), 196. <https://doi.org/10.1186/s13059-020-02096-y>
- Kingsley, P. D., Malik, J., Emerson, R. L., Bushnell, T. P., McGrath, K. E., Bloedorn, L. A., Bulger, M., & Palis, J. (2006). “Maturation” globin switching in primary primitive erythroid cells. *Blood*, *107* (4), 1665–1672. <https://doi.org/10.1182/blood-2005-08-3097>
- Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R., & Hemberg, M. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods*, *14* (5), 483–486. <https://doi.org/10.1038/nmeth.4236>
- Kissa, K., & Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, *464* (7285), 112–115. <https://doi.org/10.1038/nature08761>
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., & Kirschner, M. W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*, *161* (5), 1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>
- Knöfler, M., Haider, S., Saleh, L., Pollheimer, J., Gamage, T. K. J. B., & James, J. (2019). Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cellular and Molecular Life Sciences*, *76* (18), 3479–3496. <https://doi.org/10.1007/s00018-019-03104-6>
- Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, *10* (1), 5416. <https://doi.org/10.1038/s41467-019-13056-x>
- Kobak, D., & Linderman, G. C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nature Biotechnology*, *39* (2), 156–157. <https://doi.org/10.1038/s41587-020-00809-z>
- Kobari, L., Pflumio, F., Giarratana, M.-C., Li, X., Titeux, M., Izac, B., Leteurtre, F., Coulombel, L., & Douay, L. (2000). In vitro and in vivo evidence for the long-term multilineage (myeloid, B, NK, and T)

- reconstitution capacity of ex vivo expanded human CD34+ cord blood cells. *Experimental Hematology*, 28 (12), 1470–1480. [https://doi.org/10.1016/S0301-472X\(00\)00557-9](https://doi.org/10.1016/S0301-472X(00)00557-9)
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, 16 (12), 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>
- Kyba, M., Perlingeiro, R. C. R., & Daley, G. Q. (2002). HoxB4 Confers Definitive Lymphoid-Myeloid Engraftment Potential on Embryonic Stem Cell and Yolk Sac Hematopoietic Progenitors. *Cell*, 109 (1), 29–37. [https://doi.org/10.1016/S0092-8674\(02\)00680-3](https://doi.org/10.1016/S0092-8674(02)00680-3)
- la Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., ... Kharchenko, P. v. (2018). RNA velocity of single cells. *Nature*, 560 (7719), 494–498. <https://doi.org/10.1038/s41586-018-0414-6>
- Lancaster, M. A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., & Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, 501 (7467), 373–379. <https://doi.org/10.1038/nature12517>
- Lansdorp, P. M., Sutherland, H. J., & Eaves, C. J. (1990). Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *Journal of Experimental Medicine*, 172 (1), 363–366. <https://doi.org/10.1084/jem.172.1.363>
- Lapillonne, H., Kobari, L., Mazurier, C., Tropel, P., Giarratana, M.-C., Zanella-Cleon, I., Kiger, L., Wattenhofer-Donze, M., Puccio, H., Hebert, N., Francina, A., Andreu, G., Viville, S., & Douay, L. (2010). Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. *Haematologica*, 95 (10), 1651–1659. <https://doi.org/10.3324/haematol.2010.023556>
- Larochelle, A., Gillette, J. M., Desmond, R., Ichwan, B., Cantilena, A., Cerf, A., Barrett, A. J., Wayne, A. S., Lippincott-Schwartz, J., & Dunbar, C. E. (2012). Bone marrow homing and engraftment of human hematopoietic stem and progenitor cells is mediated by a polarized membrane domain. *Blood*, 119 (8), 1848–1855. <https://doi.org/10.1182/blood-2011-08-371583>
- Laurenti, E., & Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature*, 553 (7689), 418–426. <https://doi.org/10.1038/nature25022>
- Lause, J., Berens, P., & Kobak, D. (2021). Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *Genome Biology*, 22 (1). <https://doi.org/10.1186/s13059-021-02451-7>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15 (2), R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401 (6755), 788–791. <https://doi.org/10.1038/44565>

- Lei, F., Haque, R., Weiler, L., Vrana, K. E., & Song, J. (2009). T lineage differentiation from induced pluripotent stem cells. *Cellular Immunology*, *260* (1), 1–5. <https://doi.org/10.1016/j.cellimm.2009.09.005>
- Lengerke, C., Grauer, M., Niebuhr, N. I., Riedt, T., Kanz, L., Park, I.-H., & Daley, G. Q. (2009). Hematopoietic Development from Human Induced Pluripotent Stem Cells. *Annals of the New York Academy of Sciences*, *1176* (1), 219–227. <https://doi.org/10.1111/j.1749-6632.2009.04606.x>
- Li, Z., Lan, Y., He, W., Chen, D., Wang, J., Zhou, F., Wang, Y., Sun, H., Chen, X., Xu, C., Li, S., Pang, Y., Zhang, G., Yang, L., Zhu, L., Fan, M., Shang, A., Ju, Z., Luo, L., ... Liu, B. (2012). Mouse Embryonic Head as a Site for Hematopoietic Stem Cell Development. *Cell Stem Cell*, *11* (5), 663–675. <https://doi.org/10.1016/j.stem.2012.07.004>
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nature Methods*, *15* (12), 1053–1058. <https://doi.org/10.1038/s41592-018-0229-2>
- Lorenz, E., Upiioff, D., Reiu, T. R., & Ton, E. S. (1951). Modification of Irradiation Injury in Mice and Guinea Pigs by Bone Marrow Injections. *JNCI: Journal of the National Cancer Institute*. <https://doi.org/10.1093/jnci/12.1.197>
- Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., Rybakov, S., Misharin, A. v., & Theis, F. J. (2022). Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology*, *40* (1), 121–130. <https://doi.org/10.1038/s41587-021-01001-7>
- Lotfollahi, M., Rybakov, S., Hrovatin, K., Hedyeh-zadeh, S., Talavera-López, C., Misharin, A. v., & Theis, F. J. (2023). Biologically informed deep learning to query gene programs in single-cell atlases. *Nature Cell Biology*. <https://doi.org/10.1038/s41556-022-01072-x>
- Lu, S.-J., Feng, Q., Park, J. S., & Lanza, R. (2010). Directed Differentiation of Red Blood Cells from Human Embryonic Stem Cells. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 636, pp. 105–121). https://doi.org/10.1007/978-1-60761-691-7_7
- Luckett, W. P. (1978). Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. *American Journal of Anatomy*, *152* (1), 59–97. <https://doi.org/10.1002/aja.1001520106>
- Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & Theis, F. J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, *19* (1), 41–50. <https://doi.org/10.1038/s41592-021-01336-8>
- Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*. <https://doi.org/10.15252/msb.20188746>
- Lun, A. (2018). Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. *BioRxiv*. <https://doi.org/10.1101/404962>

- Lun, A. (2019a). *A description of the theory behind the fastMNN algorithm.*
<https://marionilab.github.io/FurtherMNN2018/theory/description.html>
- Lun, A. (2019b, March 17). *Should we standardize the log-counts?*
<https://tla.github.io/SingleCellThoughts/general/standardization.html>
- Lun, A. (2020, January 20). *What transformation should we use?*
<https://tla.github.io/SingleCellThoughts/general/transformation.html>
- Lun, A. T. L., Bach, K., & Marioni, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology*, 17 (1), 1–14. <https://doi.org/10.1186/s13059-016-0947-7>
- Lun, A. T. L., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T., & Marioni, J. C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology*, 20 (1), 63. <https://doi.org/10.1186/s13059-019-1662-y>
- Ma, F., & Pellegrini, M. (2019). ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics*, July, 1–6. <https://doi.org/10.1093/bioinformatics/btz592>
- Macqueen, J. (1965). Some Methods for Classification and Analysis of Multivariate Observations. *Proc. 5th Berkeley Symposium on Math., Stat., and Prob*, 281.
- Majeti, R., Park, C. Y., & Weissman, I. L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell*, 1 (6), 635–645. <https://doi.org/10.1016/j.stem.2007.10.001>
- Mariani, J., Simonini, M. V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A. M., Horvath, T. L., & Vaccarino, F. M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, 109 (31), 12770–12775. <https://doi.org/10.1073/pnas.1202944109>
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78 (12), 7634–7638. <https://doi.org/10.1073/pnas.78.12.7634>
- Martin, P., & Aulagner, P. G. (2009). La greffe de cellules souches hématopoïétiques. *Actualités Pharmaceutiques Hospitalières*, 5 (20), 16–28. [https://doi.org/10.1016/S1769-7344\(09\)70205-7](https://doi.org/10.1016/S1769-7344(09)70205-7)
- Mathé, G. (1960). *Transfusion et greffe de moelle osseuse homologue chez l'homme* (Vol. 8, pp. 395–412). <https://doi.org/10.1159/000386704>
- Mathé, G., Amiel, J. L., Schwarzenberg, L., Cattan, A., & Schneider, M. (1965). Adoptive immunotherapy of acute leukemia: experimental and clinical results. *Cancer Research*, 25 (9), 1525–1531. <http://www.ncbi.nlm.nih.gov/pubmed/5323965>

- Mazid, Md. A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., Liu, H., Fu, L., Yang, Y., Ibañez, D. P., Lai, J., Wei, X., An, J., Guo, P., Yuan, Y., ... Esteban, M. A. (2022). Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature*, *605* (7909), 315–324. <https://doi.org/10.1038/s41586-022-04625-0>
- McCauley, K. B., Hawkins, F., Serra, M., Thomas, D. C., Jacob, A., & Kotton, D. N. (2017). Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell*, *20* (6), 844–857.e6. <https://doi.org/10.1016/j.stem.2017.03.001>
- McCune, J. M., Namikawa, R., Kaneshima, H., Shultz, L. D., Lieberman, M., & Weissman, I. L. (1988). The SCID-hu Mouse: Murine Model for the Analysis of Human Hematolymphoid Differentiation and Function. *Science*, *241* (4873), 1632–1639. <https://doi.org/10.1126/science.2971269>
- McDonald, D., Wu, Y., Dailamy, A., Tat, J., Parekh, U., Zhao, D., Hu, M., Tipps, A., Zhang, K., & Mali, P. (2020). Defining the Teratoma as a Model for Multi-lineage Human Development. *Cell*, *183* (5), 1402–1419.e18. <https://doi.org/10.1016/j.cell.2020.10.018>
- McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*, *8* (4), 329–337.e4. <https://doi.org/10.1016/j.cels.2019.03.003>
- McGrath, K. E., Frame, J. M., Fegan, K. H., Bowen, J. R., Conway, S. J., Catherman, S. C., Kingsley, P. D., Koniski, A. D., & Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports*, *11* (12), 1892–1904. <https://doi.org/10.1016/j.celrep.2015.05.036>
- McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. <http://arxiv.org/abs/1802.03426>
- Medvinsky, A., & Dzierzak, E. (1996). Definitive Hematopoiesis Is Autonomously Initiated by the AGM Region. *Cell*, *86* (6), 897–906. [https://doi.org/10.1016/S0092-8674\(00\)80165-8](https://doi.org/10.1016/S0092-8674(00)80165-8)
- Migliaccio, G., Migliaccio, A. R., Petti, S., Mavilio, F., Russo, G., Lazzaro, D., Testa, U., Marinucci, M., & Peschle, C. (1986). Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac---liver transition. *Journal of Clinical Investigation*, *78* (1), 51–60. <https://doi.org/10.1172/JCI112572>
- Moore, M. A. S., & Metcalf, D. (1970). Ontogeny of the Haemopoietic System: Yolk Sac Origin of In Vivo and In Vitro Colony Forming Cells in the Developing Mouse Embryo. *British Journal of Haematology*, *18* (3), 279–296. <https://doi.org/10.1111/j.1365-2141.1970.tb01443.x>
- Moris, N., Anlas, K., van den Brink, S. C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., & Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. *Nature*, *582* (7812), 410–415. <https://doi.org/10.1038/s41586-020-2383-9>

- Morishima, Y. (2002). The clinical significance of human leukocyte antigen (HLA) allele compatibility in patients receiving a marrow transplant from serologically HLA-A, HLA-B, and HLA-DR matched unrelated donors. *Blood*, *99* (11), 4200–4206. <https://doi.org/10.1182/blood.V99.11.4200>
- Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., & Weissman, I. L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. *Development*, *124* (10), 1929–1939. <https://doi.org/10.1242/dev.124.10.1929>
- Munger, C., Kohler, T. N., Slatery, E., Ellermann, A. L., Bergmann, S., Penfold, C. A., Ampartzidis, I., Chen, Y., Hollfelder, F., & Boroviak, T. E. (2022). Microgel culture and spatial identity mapping elucidate the signalling requirements for primate epiblast and amnion formation. *Development*, *149* (20). <https://doi.org/10.1242/dev.200263>
- Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T., & Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, *537* (7618), 57–62. <https://doi.org/10.1038/nature19096>
- Nakano, H., Liu, X., Arshi, A., Nakashima, Y., van Handel, B., Sasidharan, R., Harmon, A. W., Shin, J. H., Schwartz, R. J., Conway, S. J., Harvey, R. P., Pashmforoush, M., Mikkola, H. K. A., & Nakano, A. (2013). Haemogenic endocardium contributes to transient definitive haematopoiesis. *Nature Communications*, *4*, 1–10. <https://doi.org/10.1038/ncomms2569>
- Nazarov, P. v., Wienecke-Baldacchino, A. K., Zinovyev, A., Czerwińska, U., Muller, A., Nashan, D., Dittmar, G., Azuaje, F., & Kreis, S. (2019). Deconvolution of transcriptomes and miRNomes by independent component analysis provides insights into biological processes and clinical outcomes of melanoma patients. *BMC Medical Genomics*, *12* (1), 132. <https://doi.org/10.1186/s12920-019-0578-4>
- Ng, E. S., Azzola, L., Bruveris, F. F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V. J., Yu, Q. C., Maksimovic, J., Liebscher, S., Januar, V., Zhang, Z., Williams, B., Conscience, A., Durnall, J., Jackson, S., Costa, M., Elliott, D., ... Elefanty, A. G. (2016). Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nature Biotechnology*, *34* (11), 1168–1179. <https://doi.org/10.1038/nbt.3702>
- Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertocello, I., Bendall, L. J., Simmons, P. J., & Haylock, D. N. (2005). Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood*, *106* (4), 1232–1239. <https://doi.org/10.1182/blood-2004-11-4422>
- North, T. E., Goessling, W., Walkley, C. R., Lengerke, C., Kopani, K. R., Lord, A. M., Weber, G. J., Bowman, T. v., Jang, I.-H., Grosser, T., FitzGerald, G. A., Daley, G. Q., Orkin, S. H., & Zon, L. I. (2007). Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature*, *447* (7147), 1007–1011. <https://doi.org/10.1038/nature05883>

- North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-Padilla, M., & Speck, N. A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development*, *126* (11), 2563–2575. <https://doi.org/10.1242/dev.126.11.2563>
- Notta, F., Doulatov, S., Laurenti, E., Poepl, A., Jurisica, I., & Dick, J. E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science*, *333* (6039), 218–221. <https://doi.org/10.1126/science.1201219>
- Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y., & Dick, J. E. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, *351* (6269). <https://doi.org/10.1126/science.aab2116>
- Ohgushi, M., Taniyama, N., Vandenbon, A., & Eiraku, M. (2022). Delamination of trophoblast-like syncytia from the amniotic ectodermal analogue in human primed embryonic stem cell-based differentiation model. *Cell Reports*, *39* (12), 110973. <https://doi.org/10.1016/j.celrep.2022.110973>
- Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H., & Arima, T. (2018). Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell*, *22* (1), 50-63.e6. <https://doi.org/10.1016/j.stem.2017.11.004>
- Oskolkov, N. (2019, December 31). *Why UMAP is Superior over tSNE*. <https://towardsdatascience.com/why-umap-is-superior-over-tsne-faa039c28e99>
- Oskolkov, N. (2020a, March 4). *tSNE vs. UMAP: Global Structure*. <https://towardsdatascience.com/tsne-vs-umap-global-structure-4d8045acba17>
- Oskolkov, N. (2020b, April 7). *tSNE Degrades to PCA*. <https://towardsdatascience.com/tsne-degrades-to-pca-d4abf9ef51d3>
- Owen, R. D. (1945). Immunogenic consequences of vascular anastomoses between bovine twins. *Science*, *102* (2651), 400–401. <https://doi.org/10.1126/science.102.2651.400>
- Paatero, P., Tapper, U., Aalto, P., & Kulmala, M. (1991). Matrix factorization methods for analysing diffusion battery data. *Journal of Aerosol Science*, *22*, S273–S276. [https://doi.org/10.1016/S0021-8502\(05\)80089-8](https://doi.org/10.1016/S0021-8502(05)80089-8)
- Palis, J. (2008). Ontogeny of erythropoiesis. *Current Opinion in Hematology*, *15* (3), 155–161. <https://doi.org/10.1097/MOH.0b013e3282f97ae1>
- Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. *Frontiers in Physiology*, *5*. <https://doi.org/10.3389/fphys.2014.00003>
- Palis, J., Malik, J., Mcgrath, K. E., & Kingsley, P. D. (2010). Primitive erythropoiesis in the mammalian embryo. *The International Journal of Developmental Biology*, *54* (6–7), 1011–1018. <https://doi.org/10.1387/ijdb.093056jp>

- Palis, J., Robertson, S., Kennedy, M., Wall, C., & Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*, *126* (22), 5073–5084. <https://doi.org/10.1242/dev.126.22.5073>
- Panch, S. R., Szymanski, J., Savani, B. N., & Stroncek, D. F. (2017). Sources of Hematopoietic Stem and Progenitor Cells and Methods to Optimize Yields for Clinical Cell Therapy. *Biology of Blood and Marrow Transplantation*, *23* (8), 1241–1249. <https://doi.org/10.1016/j.bbmt.2017.05.003>
- Pappenheim, A. (1896). Ueber Entwicklung und Ausbildung der Erythroblasten. *Archiv Für Pathologische Anatomie Und Physiologie Und Für Klinische Medicin*, *145* (3), 587–643. <https://doi.org/10.1007/BF01969901>
- Peled, A., Petit, I., Kollet, O., Magid, M., Ponomaryov, T., Byk, T., Nagler, A., Ben-Hur, H., Many, A., Shultz, L., Lider, O., Alon, R., Zipori, D., & Lapidot, T. (1999). Dependence of Human Stem Cell Engraftment and Repopulation of NOD/SCID Mice on CXCR4. *Science*, *283* (5403), 845–848. <https://doi.org/10.1126/science.283.5403.845>
- Peled, T., Landau, E., Mandel, J., Glukhman, E., Goudsmid, N. R., Nagler, A., & Fibach, E. (2004). Linear polyamine copper chelator tetraethylenepentamine augments long-term ex vivo expansion of cord blood-derived CD34+ cells and increases their engraftment potential in NOD/SCID mice. *Experimental Hematology*, *32* (6), 547–555. <https://doi.org/10.1016/j.exphem.2004.03.002>
- Peled, T., Shoham, H., Aschengrau, D., Yackoubov, D., Frei, G., Rosenheimer G, N., Lerrer, B., Cohen, H. Y., Nagler, A., Fibach, E., & Peled, A. (2012). Nicotinamide, a SIRT1 inhibitor, inhibits differentiation and facilitates expansion of hematopoietic progenitor cells with enhanced bone marrow homing and engraftment. *Experimental Hematology*, *40* (4), 342-355.e1. <https://doi.org/10.1016/j.exphem.2011.12.005>
- Perié, L., Duffy, K. R., Kok, L., de Boer, R. J., & Schumacher, T. N. (2015). The Branching Point in Erythro-Myeloid Differentiation. *Cell*, *163* (7), 1655–1662. <https://doi.org/10.1016/j.cell.2015.11.059>
- Pham, T. X. A., Panda, A., Kagawa, H., To, S. K., Ertekin, C., Georgolopoulos, G., van Knippenberg, S. S. F. A., Allsop, R. N., Bruneau, A., Chui, J. S. H., Vanheer, L., Janiszewski, A., Chappell, J., Oberhuemer, M., Tchinda, R. S., Talon, I., Rhodeer, S., Rossant, J., Lluís, F., ... Pasque, V. (2022). Modeling human extraembryonic mesoderm cells using naive pluripotent stem cells. *Cell Stem Cell*, *29* (9), 1346-1365.e10. <https://doi.org/10.1016/j.stem.2022.08.001>
- Pick, M., Azzola, L., Mossman, A., Stanley, E. G., & Elefanty, A. G. (2007). Differentiation of Human Embryonic Stem Cells in Serum-Free Medium Reveals Distinct Roles for Bone Morphogenetic Protein 4, Vascular Endothelial Growth Factor, Stem Cell Factor, and Fibroblast Growth Factor 2 in Hematopoiesis. *Stem Cells*, *25* (9), 2206–2214. <https://doi.org/10.1634/stemcells.2006-0713>
- Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L., Reik, W., Srinivas, S., Simons, B. D., Nichols, J., Marioni, J. C., &

- Göttgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, *566* (7745), 490–495. <https://doi.org/10.1038/s41586-019-0933-9>
- Pliner, H. A., Shendure, J., & Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. *Nature Methods*, *16* (10), 983–986. <https://doi.org/10.1038/s41592-019-0535-3>
- Plotzki, E., Heinrichs, G., Kubíčková, B., Ulrich, R. G., & Denner, J. (2016). Microbiological characterization of a newly established pig breed, Aachen Minipigs. *Xenotransplantation*, *23* (2), 159–167. <https://doi.org/10.1111/xen.12233>
- Polański, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., & Park, J. E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics (Oxford, England)*, *36* (3), 964–965. <https://doi.org/10.1093/bioinformatics/btz625>
- Pons, P., & Latapy, M. (2005). *Computing communities in large networks using random walks (long version)*. <http://arxiv.org/abs/physics/0512106>
- Popescu, D. M., Botting, R. A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E. F., Polanski, K., Goh, I., Efremova, M., Acres, M., Maunder, D., Vegh, P., Gitton, Y., Park, J. E., Vento-Tormo, R., Miao, Z., Dixon, D., Rowell, R., ... Haniffa, M. (2019). Decoding human fetal liver haematopoiesis. In *Nature* (Vol. 574, Issue 7778). <https://doi.org/10.1038/s41586-019-1652-y>
- Psaila, B., & Mead, A. J. (2019). Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. *Blood*, *133* (13), 1427–1435. <https://doi.org/10.1182/blood-2018-11-835371>
- Purton, L. E., & Scadden, D. T. (2007). Limiting Factors in Murine Hematopoietic Stem Cell Assays. *Cell Stem Cell*, *1* (3), 263–270. <https://doi.org/10.1016/j.stem.2007.08.016>
- Radtke, S., Humbert, O., & Kiem, H.-P. (2020). Mouse models in hematopoietic stem cell gene therapy and genome editing. *Biochemical Pharmacology*, *174*, 113692. <https://doi.org/10.1016/j.bcp.2019.113692>
- Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology*, *4* (10), e309. <https://doi.org/10.1371/journal.pbio.0040309>
- Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S., & Vert, J. P. (2018). A general and flexible method for signal extraction from single-cell RNA-seq data. *Nature Communications*, *9* (1). <https://doi.org/10.1038/s41467-017-02554-5>
- Risueño, R. M., Sachlos, E., Lee, J.-H., Lee, J. B., Hong, S.-H., Szabo, E., & Bhatia, M. (2012). Inability of Human Induced Pluripotent Stem Cell-Hematopoietic Derivatives to Downregulate MicroRNAs In Vivo Reveals a Block in Xenograft Hematopoietic Regeneration. *Stem Cells*, *30* (2), 131–139. <https://doi.org/10.1002/stem.1684>

- Rivron, N., Pera, M., Rossant, J., Martinez Arias, A., Zernicka-Goetz, M., Fu, J., van den Brink, S., Bredenoord, A., Dondorp, W., de Wert, G., Hyun, I., Munsie, M., & Isasi, R. (2018). Debate ethics of embryo models from stem cells. *Nature*, *564* (7735), 183–185. <https://doi.org/10.1038/d41586-018-07663-9>
- Robins, M. M., & Noyes, W. D. (1961). Aplastic Anemia Treated with Bone-Marrow Transfusion from an Identical Twin. *New England Journal of Medicine*, *265* (20), 974–979. <https://doi.org/10.1056/NEJM196111162652002>
- Robinton, D. A., & Daley, G. Q. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature*, *481* (7381), 295–305. <https://doi.org/10.1038/nature10761>
- Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., & Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, *553* (7687), 212–216. <https://doi.org/10.1038/nature25168>
- Ronaldson-Bouchard, K., Ma, S. P., Yeager, K., Chen, T., Song, L., Sirabella, D., Morikawa, K., Teles, D., Yazawa, M., & Vunjak-Novakovic, G. (2018). Advanced maturation of human cardiac tissue grown from pluripotent stem cells. *Nature*, *556* (7700), 239–243. <https://doi.org/10.1038/s41586-018-0016-3>
- Ross, C., & Boroviak, T. E. (2020). Origin and function of the yolk sac in primate embryogenesis. *Nature Communications*, *11* (1), 3760. <https://doi.org/10.1038/s41467-020-17575-w>
- Rowe, R. G., & Daley, G. Q. (2019). Induced pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Genetics*, *20* (7), 377–388. <https://doi.org/10.1038/s41576-019-0100-z>
- Rybtsov, S., Ivanovs, A., Zhao, S., & Medvinsky, A. (2016). Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. *Development*, *143* (8), 1284–1289. <https://doi.org/10.1242/dev.131193>
- Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, *37* (5), 547–554. <https://doi.org/10.1038/s41587-019-0071-9>
- Sanaki-Matsumiya, M., Matsuda, M., Gritti, N., Nakaki, F., Sharpe, J., Trivedi, V., & Ebisuya, M. (2022). Periodic formation of epithelial somites from human pluripotent stem cells. *Nature Communications*, *13* (1), 2325. <https://doi.org/10.1038/s41467-022-29967-1>
- Sandler, V. M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J. M., Scandura, J. M., & Rafii, S. (2014). Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature*, *511* (7509), 312–318. <https://doi.org/10.1038/nature13547>
- Sato, T., Vries, R. G., Snippert, H. J., van de Wetering, M., Barker, N., Stange, D. E., van Es, J. H., Abo, A., Kujala, P., Peters, P. J., & Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*, *459* (7244), 262–265. <https://doi.org/10.1038/nature07935>
- Sauvageau, G., Thorsteinsdottir, U., Eaves, C. J., Lawrence, H. J., Largman, C., Lansdorp, P. M., & Humphries, R. K. (1995). Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more

- primitive populations in vitro and in vivo. *Genes & Development*, 9 (14), 1753–1765. <https://doi.org/10.1101/gad.9.14.1753>
- Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D. J. H. F., Lavin, Y., Lau, C. M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., Mirny, L. A., Merad, M., Eaves, C. J., & Reizis, B. (2016). Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. *Immunity*, 45 (3), 597–609. <https://doi.org/10.1016/j.immuni.2016.08.007>
- Shahbazi, M. N. (2020). Mechanisms of human embryo development: from cell fate to tissue shape and back. *Development*, 147 (14). <https://doi.org/10.1242/dev.190629>
- Shahbazi, M. N., Scialdone, A., Skorupska, N., Weberling, A., Recher, G., Zhu, M., Jedrusik, A., Devito, L. G., Noli, L., Macaulay, I. C., Buecker, C., Khalaf, Y., Ilic, D., Voet, T., Marioni, J. C., & Zernicka-Goetz, M. (2017a). Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature*, 552 (7684), 239–243. <https://doi.org/10.1038/nature24675>
- Shahbazi, M. N., Scialdone, A., Skorupska, N., Weberling, A., Recher, G., Zhu, M., Jedrusik, A., Devito, L. G., Noli, L., Macaulay, I. C., Buecker, C., Khalaf, Y., Ilic, D., Voet, T., Marioni, J. C., & Zernicka-Goetz, M. (2017b). Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature*, 552 (7684), 239–243. <https://doi.org/10.1038/nature24675>
- Shao, Y., Taniguchi, K., Townshend, R. F., Miki, T., Gumucio, D. L., & Fu, J. (2017). A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nature Communications*, 8 (1), 208. <https://doi.org/10.1038/s41467-017-00236-w>
- Sheridan, S. D., Surampudi, V., & Rao, R. R. (2012). Analysis of Embryoid Bodies Derived from Human Induced Pluripotent Stem Cells as a Means to Assess Pluripotency. *Stem Cells International*, 2012, 1–9. <https://doi.org/10.1155/2012/738910>
- Shultz, L. D., Lyons, B. L., Burzenski, L. M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S. D., King, M., Mangada, J., Greiner, D. L., & Handgretinger, R. (2005). Human Lymphoid and Myeloid Cell Development in NOD/LtSz- scid IL2R γ null Mice Engrafted with Mobilized Human Hemopoietic Stem Cells. *The Journal of Immunology*, 174 (10), 6477–6489. <https://doi.org/10.4049/jimmunol.174.10.6477>
- Shultz, L. D., Schweitzer, P. A., Christianson, S. W., Gott, B., Schweitzer, I. B., Tennent, B., McKenna, S., Mobraaten, L., Rajan, T. v, & Greiner, D. L. (1995). Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. *The Journal of Immunology*, 154 (1), 180–191. <https://doi.org/10.4049/jimmunol.154.1.180>
- Soneson, C., & Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods*, 15 (4), 255–261. <https://doi.org/10.1038/nmeth.4612>
- Spencer Chapman, M., Ranzoni, A. M., Myers, B., Williams, N., Coorens, T. H. H., Mitchell, E., Butler, T., Dawson, K. J., Hooks, Y., Moore, L., Nangalia, J., Robinson, P. S., Yoshida, K., Hook, E., Campbell, P. J., & Cvejic, A.

- (2021). Lineage tracing of human development through somatic mutations. *Nature*, May 2020. <https://doi.org/10.1038/s41586-021-03548-6>
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19 (1). <https://doi.org/10.1186/s12864-018-4772-0>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177 (7), 1888-1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>
- Sugimura, R., Jha, D. K., Han, A., Soria-Valles, C., da Rocha, E. L., Lu, Y. F., Goettel, J. A., Serrao, E., Rowe, R. G., Malleshaiah, M., Wong, I., Sousa, P., Zhu, T. N., Ditadi, A., Keller, G., Engelman, A. N., Snapper, S. B., Doulatov, S., & Daley, G. Q. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*. <https://doi.org/10.1038/nature22370>
- Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., & Nakauchi, H. (2013). Generation of engraftable hematopoietic stem cells from induced pluripotent stem cells by way of teratoma formation. *Molecular Therapy*, 21 (7), 1424–1431. <https://doi.org/10.1038/mt.2013.71>
- Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. *Nature Biotechnology*, 38 (2), 147–150. <https://doi.org/10.1038/s41587-019-0379-5>
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131 (5), 861–872. <https://doi.org/10.1016/j.cell.2007.11.019>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126 (4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Takayama, N., Nishikii, H., Usui, J., Tsukui, H., Sawaguchi, A., Hiroyama, T., Eto, K., & Nakauchi, H. (2008). Generation of functional platelets from human embryonic stem cells in vitro via ES-sacs, VEGF-promoted structures that concentrate hematopoietic progenitors. *Blood*, 111 (11), 5298–5306. <https://doi.org/10.1182/blood-2007-10-117622>
- Tam, P. P. L., & Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development*, 68 (1–2), 3–25. [https://doi.org/10.1016/S0925-4773\(97\)00123-8](https://doi.org/10.1016/S0925-4773(97)00123-8)
- Tan, Y.-T., Ye, L., Xie, F., Beyer, A. I., Muench, M. O., Wang, J., Chen, Z., Liu, H., Chen, S.-J., & Kan, Y. W. (2018). 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 (9), 2180–2185. <https://doi.org/10.1073/pnas.1718446115>

- Tanaka, S., Kunath, T., Hadjantonakis, A.-K., Nagy, A., & Rossant, J. (1998). Promotion of Trophoblast Stem Cell Proliferation by FGF4. *Science*, 282 (5396), 2072–2075. <https://doi.org/10.1126/science.282.5396.2072>
- Tarazi, S., Aguilera-Castrejon, A., Joubran, C., Ghanem, N., Ashouokhi, S., Roncato, F., Wildschutz, E., Haddad, M., Oldak, B., Gomez-Cesar, E., Livnat, N., Viukov, S., Lokshtanov, D., Naveh-Tassa, S., Rose, M., Hanna, S., Raanan, C., Brenner, O., Kedmi, M., ... Hanna, J. H. (2022). Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. *Cell*, 185 (18), 3290-3306.e25. <https://doi.org/10.1016/j.cell.2022.07.028>
- Tavian, M., Coulombel, L., Luton, D., Clemente, H. S. H., Dieterlen-Lievre, F., Peault, B., Dieterlen-Lièvre, F., & Péault, B. (1996). Aorta-associated CD34+ hematopoietic cells in the early human embryo. *Blood*, 87 (1), 67–72. <https://doi.org/10.1182/blood.V87.1.67.bloodjournal87167>
- Tavian, M., Hallais, M. F., & Péault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development*, 126 (4), 793–803.
- Tavian, M., Robin, C., Coulombel, L., & Péault, B. (2001). The Human Embryo, but Not Its Yolk Sac, Generates Lympho-Myeloid Stem Cells. *Immunity*, 15 (3), 487–495. [https://doi.org/10.1016/S1074-7613\(01\)00193-5](https://doi.org/10.1016/S1074-7613(01)00193-5)
- Thomas, E. D., Lochte, H. L., Lu, W. C., & Ferrebee, J. W. (1957). Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. *New England Journal of Medicine*, 257 (11), 491–496. <https://doi.org/10.1056/NEJM195709122571102>
- Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*, 282 (5391), 1145–1147. <https://doi.org/10.1126/science.282.5391.1145>
- Till, J. E., & McCulloch, E. A. (1961). A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiation Research*, 14 (2), 213. <https://doi.org/10.2307/3570892>
- Tober, J., Koniski, A., McGrath, K. E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K. K. L., Waugh, R., & Palis, J. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood*, 109 (4), 1433–1441. <https://doi.org/10.1182/blood-2006-06-031898>
- Townes, F. W., Hicks, S. C., Aryee, M. J., & Irizarry, R. A. (2019). Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biology*, 20 (1), 1–16. <https://doi.org/10.1186/s13059-019-1861-6>
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9 (1). <https://doi.org/10.1038/s41598-019-41695-z>

- Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., & Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology*. <https://doi.org/10.1186/s13059-019-1850-9>
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32* (4), 381–386. <https://doi.org/10.1038/nbt.2859>
- Turpen, J. B., Knudson, C. M., & Hoefen, P. S. (1981). The early ontogeny of hematopoietic cells studied by grafting cytogenetically labeled tissue anlagen: Localization of a prospective stem cell compartment. *Developmental Biology*, *85* (1), 99–112. [https://doi.org/10.1016/0012-1606\(81\)90239-6](https://doi.org/10.1016/0012-1606(81)90239-6)
- Tyser, R. C. V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., & Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature*, *600* (7888), 285–289. <https://doi.org/10.1038/s41586-021-04158-y>
- Ueda, T., Tsuji, K., Yoshino, H., Ebihara, Y., Yagasaki, H., Hisakawa, H., Mitsui, T., Manabe, A., Tanaka, R., Kobayashi, K., Ito, M., Yasukawa, K., & Nakahata, T. (2000). Expansion of human NOD/SCID-repopulating cells by stem cell factor, Flk2/Flt3 ligand, thrombopoietin, IL-6, and soluble IL-6 receptor. *Journal of Clinical Investigation*, *105* (7), 1013–1021. <https://doi.org/10.1172/JCI8583>
- van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. In *Journal of Machine Learning Research* (Vol. 9).
- Velicky, P., Meinhardt, G., Plessl, K., Vondra, S., Weiss, T., Haslinger, P., Lendl, T., Aumayr, K., Mairhofer, M., Zhu, X., Schütz, B., Hannibal, R. L., Lindau, R., Weil, B., Ernerudh, J., Neesen, J., Egger, G., Mikula, M., Röhr, C., ... Pollheimer, J. (2018). Genome amplification and cellular senescence are hallmarks of human placenta development. *PLOS Genetics*, *14* (10), e1007698. <https://doi.org/10.1371/journal.pgen.1007698>
- Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W. K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A. G., & Steinmetz, L. M. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology*, *19* (4), 271–281. <https://doi.org/10.1038/ncb3493>
- Vieth, B., Parekh, S., Ziegenhain, C., Enard, W., & Hellmann, I. (2019). A systematic evaluation of single cell RNA-seq analysis pipelines. *Nature Communications*, *10* (1), 4667. <https://doi.org/10.1038/s41467-019-12266-7>
- Vodyanik, M. A., Bork, J. A., Thomson, J. A., & Slukvin, I. I. (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood*, *105* (2), 617–626. <https://doi.org/10.1182/blood-2004-04-1649>

- Wagers, A. J., & Weissman, I. L. (2006). Differential Expression of $\alpha 2$ Integrin Separates Long-Term and Short-Term Reconstituting Lin⁻/loThy1.1loc-kit⁺ Sca-1⁺ Hematopoietic Stem Cells. *Stem Cells*, *24* (4), 1087–1094. <https://doi.org/10.1634/stemcells.2005-0396>
- Wagner, J. E., Brunstein, C. G., Boitano, A. E., DeFor, T. E., McKenna, D., Sumstad, D., Blazar, B. R., Tolar, J., Le, C., Jones, J., Cooke, M. P., & Bleul, C. C. (2016). Phase I/II Trial of StemRegenin-1 Expanded Umbilical Cord Blood Hematopoietic Stem Cells Supports Testing as a Stand-Alone Graft. *Cell Stem Cell*, *18* (1), 144–155. <https://doi.org/10.1016/j.stem.2015.10.004>
- Wang, L., Menendez, P., Shojaei, F., Li, L., Mazurier, F., Dick, J. E., Cerdan, C., Levac, K., & Bhatia, M. (2005). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. *Journal of Experimental Medicine*. <https://doi.org/10.1084/jem.20041888>
- Weatherbee, B. A. T., Cui, T., & Zernicka-Goetz, M. (2021). Modeling human embryo development with embryonic and extra-embryonic stem cells. *Developmental Biology*, *474*, 91–99. <https://doi.org/10.1016/j.ydbio.2020.12.010>
- Weisel, K. C., Gao, Y., Shieh, J.-H., & Moore, M. A. S. (2006). Stromal cell lines from the aorta-gonadomesonephros region are potent supporters of murine and human hematopoiesis. *Experimental Hematology*, *34* (11), 1505–1516. <https://doi.org/10.1016/j.exphem.2006.06.013>
- Weissman, I. L., & Shizuru, J. A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood*, *112* (9), 3543–3553. <https://doi.org/10.1182/blood-2008-08-078220>
- Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., & Macosko, E. Z. (2019). Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell*, *177* (7), 1873–1887.e17. <https://doi.org/10.1016/j.cell.2019.05.006>
- Wilkinson, A. C., Igarashi, K. J., & Nakauchi, H. (2020). Haematopoietic stem cell self-renewal in vivo and ex vivo. *Nature Reviews Genetics*, *21* (9), 541–554. <https://doi.org/10.1038/s41576-020-0241-0>
- Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lió, P., MacDonald, H. R., & Trumpp, A. (2008). Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell*, *135* (6), 1118–1129. <https://doi.org/10.1016/j.cell.2008.10.048>
- Witte, T. de, Zwaan, F., Hermans, J., Vernant, J., Kolb, H., Vossen, J., Lönnqvist, B., Beelen, D., Ferrant, A., Gmür, J., Yin, J. L., Troussard, X., Cahn, J., Lint, M. van, & Gratwohl, A. (1990). Allogeneic bone marrow transplantation for secondary leukaemia and myelodysplastic syndrome: a survey by the Leukaemia Working Party of the European Bone Marrow Transplantation Group (EBMTG). *British Journal of Haematology*, *74* (2), 151–155. <https://doi.org/10.1111/j.1365-2141.1990.tb02558.x>

- Wolber, F. M., Leonard, E., Michael, S., Orschell-Traycoff, C. M., Yoder, M. C., & Srour, E. F. (2002). Roles of spleen and liver in development of the murine hematopoietic system. *Experimental Hematology*, *30* (9), 1010–1019. [https://doi.org/10.1016/S0301-472X\(02\)00881-0](https://doi.org/10.1016/S0301-472X(02)00881-0)
- Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., & Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, *20* (1), 1–9. <https://doi.org/10.1186/s13059-019-1663-x>
- Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*, *8* (4), 281–291.e9. <https://doi.org/10.1016/j.cels.2018.11.005>
- World Donor Marrow Association. (2021). *WMDA global trends report 2021*. <https://wmda.info/wp-content/uploads/2022/05/21042022-GTR-2021-Search-Match-Service-1.pdf>
- Xia, Y., Nivet, E., Sancho-Martinez, I., Gallegos, T., Suzuki, K., Okamura, D., Wu, M.-Z., Dubova, I., Esteban, C. R., Montserrat, N., Campistol, J. M., & Belmonte, J. C. I. (2013). Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells. *Nature Cell Biology*, *15* (12), 1507–1515. <https://doi.org/10.1038/ncb2872>
- Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M. I., & Yosef, N. (2021). Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular Systems Biology*, *17* (1). <https://doi.org/10.15252/msb.20209620>
- Xu, R.-H., Chen, X., Li, D. S., Li, R., Addicks, G. C., Glennon, C., Zwaka, T. P., & Thomson, J. A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature Biotechnology*, *20* (12), 1261–1264. <https://doi.org/10.1038/nbt761>
- Yanagida, A., Spindlow, D., Nichols, J., Dattani, A., Smith, A., & Guo, G. (2021). Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell*, *28* (6), 1016–1022.e4. <https://doi.org/10.1016/j.stem.2021.04.031>
- Yokomizo, T., Ideue, T., Morino-Koga, S., Tham, C. Y., Sato, T., Takeda, N., Kubota, Y., Kurokawa, M., Komatsu, N., Ogawa, M., Araki, K., Osato, M., & Suda, T. (2022). Independent origins of fetal liver haematopoietic stem and progenitor cells. *Nature*, *609* (7928), 779–784. <https://doi.org/10.1038/s41586-022-05203-0>
- Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M. J., Porayette, P., Shelley, W. C., Conway, S. J., Dorshkind, K., & Yoder, M. C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proceedings of the National Academy of Sciences*, *108* (4), 1468–1473. <https://doi.org/10.1073/pnas.1015841108>

- Yoshimoto, M., Porayette, P., Glosson, N. L., Conway, S. J., Carlesso, N., Cardoso, A. A., Kaplan, M. H., & Yoder, M. C. (2012). Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood*, *119* (24), 5706–5714. <https://doi.org/10.1182/blood-2011-12-397489>
- Young, M. D. (2020). *Single cell differential expression testing*. <https://constantamateur.github.io/2020-04-10-scDE/>
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, *318* (5858), 1917–1920. <https://doi.org/10.1126/science.1151526>
- Yu, L., Wei, Y., Duan, J., Schmitz, D. A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G. C., & Wu, J. (2021). Blastocyst-like structures generated from human pluripotent stem cells. *Nature*, *591* (7851), 620–626. <https://doi.org/10.1038/s41586-021-03356-y>
- Yu, Q. C., Hirst, C. E., Costa, M., Ng, E. S., Schiesser, J. v., Gertow, K., Stanley, E. G., & Elefanty, A. G. (2012). APELIN promotes hematopoiesis from human embryonic stem cells. *Blood*, *119* (26), 6243–6254. <https://doi.org/10.1182/blood-2011-12-396093>
- Yvernogeu, L., Dainese, G., & Jaffredo, T. (2023). Dorsal aorta polarization and haematopoietic stem cell emergence. *Development*, *150* (1). <https://doi.org/10.1242/dev.201173>
- Yvernogeu, L., Gautier, R., Petit, L., Khoury, H., Relaix, F., Ribes, V., Sang, H., Charbord, P., Souyri, M., Robin, C., & Jaffredo, T. (2019). In vivo generation of haematopoietic stem/progenitor cells from bone marrow-derived haemogenic endothelium. *Nature Cell Biology*, *D*. <https://doi.org/10.1038/s41556-019-0410-6>
- Zambidis, E. T., Peault, B., Park, T. S., Bunz, F., & Civin, C. I. (2005). Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. *Blood*, *106* (3), 860–870. <https://doi.org/10.1182/blood-2004-11-4522>
- Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., Bian, L., Lan, Y., & Liu, B. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Research*, *29* (11), 881–894. <https://doi.org/10.1038/s41422-019-0228-6>
- Zhai, J., Xiao, Z., Wang, Y., & Wang, H. (2022). Human embryonic development: from peri-implantation to gastrulation. *Trends in Cell Biology*, *32* (1), 18–29. <https://doi.org/10.1016/j.tcb.2021.07.008>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8* (1), 14049. <https://doi.org/10.1038/ncomms14049>

- Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S. N., Li, Z., Muncie, J. M., Lakins, J. N., Weaver, V. M., Gumucio, D. L., & Fu, J. (2019). Controlled modelling of human epiblast and amnion development using stem cells. *Nature*, *573* (7774), 421–425. <https://doi.org/10.1038/s41586-019-1535-2>
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., & Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*, *65* (4), 631-643.e4. <https://doi.org/10.1016/j.molcel.2017.01.023>
- Zonari, E., Desantis, G., Petrillo, C., Boccalatte, F. E., Lidonnici, M. R., Kajaste-Rudnitski, A., Aiuti, A., Ferrari, G., Naldini, L., & Gentner, B. (2017). Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy. *Stem Cell Reports*, *8* (4), 977–990. <https://doi.org/10.1016/j.stemcr.2017.02.010>
- Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C., Tallquist, M. D., & Iruela-Arispe, M. L. (2008). Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. *Cell Stem Cell*, *3* (6), 625–636. <https://doi.org/10.1016/j.stem.2008.09.018>