

Différenciation des précurseurs hématopoïétiques chez l'embryon

Nathalie Lehmann¹, Mariam Sissoko¹

Enseignants référents: **Hervé Isambert²**, **Louis Verny²**, **Nadir Sella²**

Résumé

Le but de ce projet de Master 2 de bioinformatique est de reconstruire le réseau de régulation régissant les expressions des facteurs de transcription clés pour la différenciation des précurseurs hématopoïétiques chez l'embryon. Dans cet objectif, nous avons d'abord procédé à un filtrage des données afin de garder les gènes d'intérêt, puis analysé celles-ci selon deux approches différentes: reconstruction de réseau via l'algorithme polynomial PC[1] (*Peter-Clark*) et établissement d'un arbre binaire par clustering hiérarchique non supervisé. Enfin, en comparant nos résultats avec ceux présents dans la littérature scientifique, nous avons fait état d'un modèle graphique simplifié expliquant les mécanismes impliquant la différenciation des cellules primitives en deux lignées distinctes : hématopoïétique et endothéliale. Ainsi, nous avons pu visualiser le rôle central de certains facteurs de transcription (FT), dont Tal1, l'*erythroid differentiation factor* connu pour son implication dans différents cas de leucémies, ou Runx1 pour induire les précurseurs des cellules épithéliales. Cette reconstruction de réseau a été effectuée à partir de données analysées par *single cell RNA-seq* puis binarisées.

Mots-clés

Réseaux de régulation – Facteurs de transcription – Hématopoïèse

¹ Master 2 Bioinformatique et Modélisation, Université Paris 6, France

² Institut Curie, France

Table des matières

Introduction	1
1 Données et méthodes	2
1.1 Dataset	2
1.2 Filtre des données	2
1.3 Gènes d'intérêt	2
1.4 Reconstruction de réseaux	3
2 Résultats et discussion	4
2.1 Vue d'ensemble : clustering hiérarchique	5
2.2 Vue d'ensemble : réseau obtenu par PC	5
2.3 Tri par lignée - fonctionnel	5
2.4 Tri par stade embryonnaire - temporel	6
Conclusion	7
Références	7

Introduction

Au cours du développement de l'embryon des Vertébrés, tous les tissus hématopoïétiques successivement

actifs (foie, thymus, rate et moelle osseuse) sont colonisés par des cellules souches hématopoïétiques (CSH) d'origine extrinsèque. Le sac vitellin (SV) constitue l'unique exception à cette règle, puisque des CSH s'y développent in situ. Il a été observé que le SV constitue le premier site d'hématopoïèse de l'embryon[2] : c'est le lieu d'apparition des premières cellules sanguines propres à l'embryon. Cependant, de la lignée primitive à l'origine de celles-ci, émerge aussi les premières cellules endothéliales (constituant la paroi interne des vaisseaux sanguins). Dans ces conditions, quels sont les facteurs de transcription suffisants et/ou nécessaires pour induire cette différenciation de la lignée primitive ?

Reconstruire le réseau de régulation contrôlant cette différenciation pourrait permet de mieux appréhender les mécanismes de l'hématopoïèse primitive et de la formation des tissus sanguins. Or l'origine de certaines leucémies (ie l'anémie de Fanconi[3]) reste encore difficile à déterminer, et l'établissement de tels réseaux pourrait alors favoriser la compréhension et l'établissement

de protocoles expérimentaux mieux ciblés.

Il est important de spécifier que ce projet s'appuie largement sur l'article de Moignard et al., *Decoding the regulatory network of early blood development from single-cell expression measurements*[4]. En effet, les données utilisées pour réaliser ce projet sont similaires à celles utilisées par les auteurs de l'article sus-nommé, et la démarche globale de reconstruction de réseau est relativement semblable, bien qu'allégrement simplifiée.

1. Données et méthodes

La reconstruction des voies moléculaires contrôlant le développement embryonnaire des organes est entravée par le manque de méthodes adaptées pour l'étude de phénomènes extrêmement précis, d'autant plus si le matériel disponible est limitant. Les techniques traditionnelles telles que le RNA-seq se révèlent alors insuffisantes. La stratégie que les auteurs de l'article de référence ont choisi est particulièrement pertinente car il s'agit d'une approche combinant le séquençage *single cell* et les analyses computationnelles de reconstruction de réseaux à partir des graphes des états de transition. En effet, le séquençage *single cell* (sc) permet une analyse transcriptomique à l'échelle d'une seule cellule. Grâce au sc-RNAseq, il devient possible d'estimer l'hétérogénéité intra-tumorale, mais aussi d'étudier des stades embryonnaires précoces, ou encore de retracer les lignées cellulaires au cours du développement (source : [bioinfo-fr](http://bioinfo-fr.net))¹.

1.1 Dataset

Le dataset proposé rassemble les données d'expression binarisées de différents gènes pouvant soit être des facteurs de transcription (33 gènes), soit d'autres gènes marqueurs (spatiaux ou temporels - 9 gènes), ou encore des gènes servant de contrôle (*housekeepers* - 4 gènes). Chaque ligne du dataset correspond au profil d'expression d'une cellule, analysée par sc-RNAseq. Les colonnes, quant à elles, correspondent aux gènes dont l'expression a été quantifiée. Les données étant binaires, le '1' représente un gène exprimé dans la condition correspondante, le '0' un gène non exprimé. Au total, cela fait donc 46 gènes analysés dans 3934 cellules issues d'embryons de souris, prélevées à quatre stades différents du développement embryonnaire précoce. Celles-ci deviendront éventuellement des cellules sanguines (en rouge

sur la figure 1) ou endothéliales (en violet). Comme indiqué sur la figure 1, cinq populations sont analysées :

- E7.0 (*primitive streak*, PS),
- E7.5 (*neural plate*, NP)
- E7.75 (*head fold*, HF).
- E8.25, cellules GFP+ (*four somite*, 4SG) cellules sanguines potentielles
- E8.25, cellules Flk1+GFP (4SFG) cellules endothéliales potentielles.

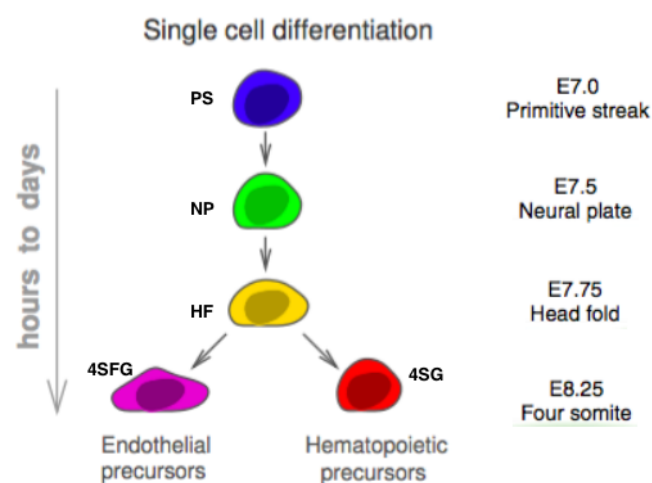


FIGURE 1. Processus de différenciation de la lignée primitive (PS) en 2 lignées distinctes : endothéliale (4SFG) et hématopoïétique (4SG)

Les étapes critiques du développement hématopoïétique murin sont recensées figure 17 en annexe.

1.2 Filtre des données

Afin d'obtenir un set de données non biaisées, nous avons choisi d'appliquer un filtre afin d'éliminer les gènes exprimés dans 100% des cas (codé en Python, on ôte du dataset les colonnes où il n'existe que des '1'). Les gènes qui disparaissent alors sont référencés, et leur fonction en tant que *housekeepers* a été vérifiée via le site de la NCBI² ou via le site Gene Cards³ : *Eif2b1*, *Mrpl19*, *Polr2a*, *Ubc*.

1.3 Gènes d'intérêt

Une fois le filtre appliqué, il nous a fallu procéder à différents tris des données.

2. <https://www.ncbi.nlm.nih.gov/gene>

3. Dans toute la suite du document, les fonctions des gènes ont été identifiées via le site Gene Cards.

1. <http://bioinfo-fr.net>

A 3-steps algorithm

REQUIRE Conditional independence information among all variables in V , and an ordering order(V) on the variables

- 1 Find a **skeleton** and **separation sets**
- 2 **Orient unshielded triples** in the skeleton based on the separation sets
- 3 **Orient** as many of the **remaining undirected edges** as possible by repeated application of rules $R1 - R3$

RETURN Best case \rightarrow *DAG* (usually a *PDAG*) and separation sets

FIGURE 2. Algorithme PC (Spirtes, Glymour, Scheines (1993))

1.3.1 Tri par facteur de transcription

Les 42 gènes restant dans notre dataset n'étant pas tous impliqués directement dans la différenciation cellulaire, nous avons effectué un premier tri où ne sont conservés que les 33 facteurs de transcription. Les gènes marqueurs qui ne se trouvent plus dans le dataset codent pour les protéines suivantes : les protéines d'adhésion cellulaire calcio-dépendantes **Cdh1** et **Cdh5**, une sous-unité du facteur d'initiation de la traduction **Egfl7**, la globine embryonnaire **Hbb-bH1**, les récepteurs **Itga2b** (ou CD41), **Kit**, **Procr** et **Kdr** (ce dernier étant un récepteur de la VEGF - *Vascular endothelial growth factor*), et enfin **Pecam1**, une protéine de surface impliquée dans les jonctions inter-cellulaires des cellules endothéliales. En fonction de leur rôle et de leur localisation, certains de ces gènes jouent un rôle clé pour la différenciation des précurseurs hématopoïétiques, et cela peut être facilement visualisé sur les réseaux qui ont été reconstruits. Ainsi, dans chacun des tris décrits ci-dessous, a été conservé un dataset avec les 42 gènes, et un autre avec les 33 facteurs de transcription afin de permettre d'analyser de façon spécifique toutes les interactions potentielles.

1.3.2 Tri par stade embryonnaire

Afin d'étudier les relations au niveau temporel, les données ont été séparées par type cellulaire présent dans le dataset (PS, NP, HF, 4SFG, 4SG).

1.3.3 Tri par lignée

Enfin, une séparation des données au niveau fonctionnel a été effectuée. Nous avons alors un dataset pour les gènes préférentiellement exprimés dans la lignée primitive (en bleu sur la figure 4), un autre dont les gènes sont davantage associés à l'hématopoïèse (en rouge) et

le dernier pour les gènes impliqués dans la formation de l'endothélium (en rose/violet).

1.4 Reconstruction de réseaux

Un réseau est défini comme un ensemble de points appelés noeuds connectés entre-eux par des liens, ces derniers pouvant être orientés ou non. On appelle degré d'un noeud le nombre de liens que celui-ci établit avec ses voisins. Dans le cas d'un réseau orienté, le degré regroupe les liens entrants et sortants. Dans notre étude, les réseaux ont été établis en définissant les gènes pour noeuds et les interactions de régulation (activation ou inhibition) comme liens (les données étant des niveaux d'expression binarisés comme décrit ci-dessus). Nous avons choisi d'utiliser deux types d'approches différentes.

1.4.1 Algorithme PC

Une première démarche pour reconstruire les réseaux est d'appliquer l'algorithme PC[1] (Peter-Clark). Il s'agit d'un algorithme polynomial pour l'inférence de l'architecture des réseaux. Pour cela, deux choix s'offraient à nous : utiliser le package *pcalg* disponible sur R, ou bien le [Miic Web Server](https://miic.curie.fr)⁴ (*Multivariate Information based Inductive Causation*), outil développé par l'équipe enseignante. La robustesse de l'outil, la maîtrise facile de l'interface et la diversité des paramètres modifiables pour la manipulation des données ont vite orienté notre choix pour l'utilisation de ce dernier. Nous avons notamment fait usage de l'interface [Cytoscape](http://cytoscape.org)⁵ accessible via le Miic Web Server. Miic a pour but de reconstruire des réseaux de causalité, non-causalité, ou mixte, entre les variables du dataset qui lui est soumis. Il permet de

4. <https://miic.curie.fr>

5. <http://cytoscape.org>

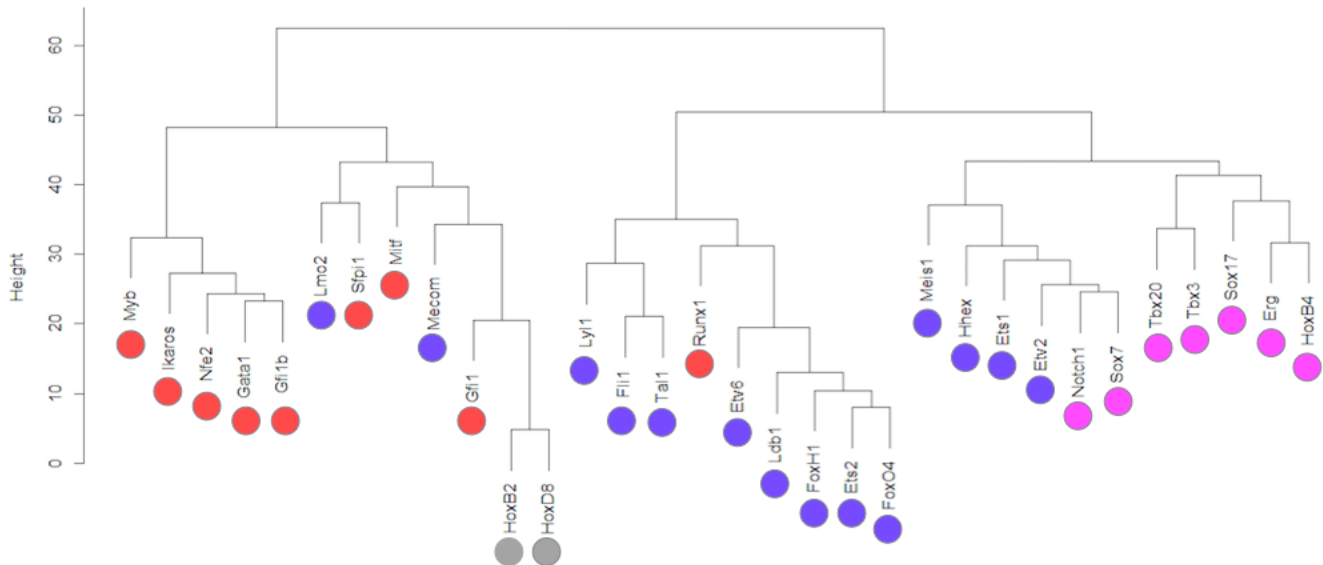


FIGURE 3. Arbre obtenu par clustering hiérarchique non supervisé - 33 gènes

reconstruire des graphes acycliques dirigés (DAG).

Parmi les paramètres par défaut (ceux que nous avons utilisés), on peut noter que les variables du réseaux sont considérées comme indépendantes, même pour des conditions expérimentales identiques. Aussi, le réseau est reconstruit par maximum de vraisemblance normalisé (pour des analyses futures, on pourrait faire varier ce critère de complexité, notamment en utilisant la reconstruction basée sur les informations bayésiennes). Enfin, les effets de causes latentes sur les relations entre les noeuds ne sont pas mesurées par défaut mais pourrait également être un critère intéressant pour une analyse plus fine du réseau.

Les différentes étapes de l'algorithme sont détaillées figure 2. Brièvement, la première étape consiste en la reconstruction de l'architecture du réseau. Les directions des arcs appartenant aux V-structures sont ensuite déterminées. Enfin, quand cela est possible, la direction des arcs restants est calculée en tenant compte du principe d'acyclicité.

1.4.2 Clustering hiérarchique

L'objectif principal des méthodes de classification automatique est de répartir les éléments d'un ensemble en groupes, c'est-à-dire d'établir une partition de cet ensemble. A cette partition vient s'ajouter un critère de hiérarchie de parties, qui permettent alors de former un arbre binaire, appelé le dendrogramme⁶. L'algorithme de clustering hiérarchique non supervisé est disponible

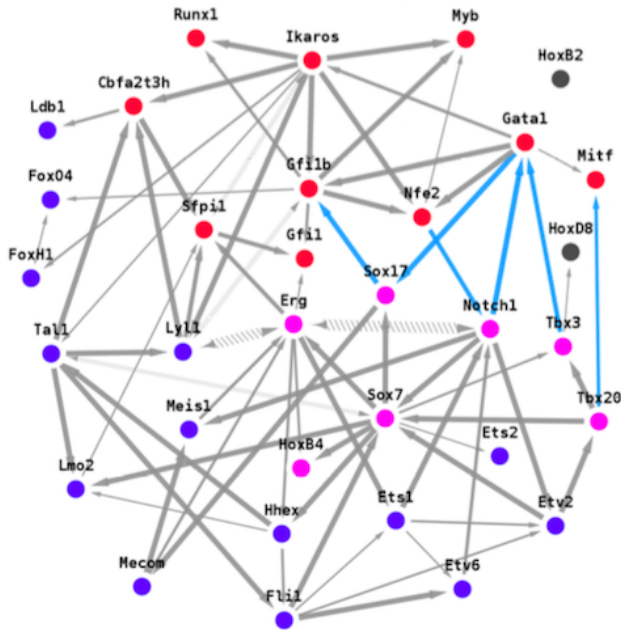
via la fonction *hclust*⁷ de R. Celle-ci prend pour *input* une matrice de distance. Nous avons donc préalablement constitué cette matrice à partir des données à analyser (distance euclidienne calculée entre chaque couple de données via la fonction *dist* de R). Le dendrogramme permet de visualiser simplement les regroupements de gènes par profil d'expression, et ainsi d'établir d'éventuelles catégories fonctionnelles.

2. Résultats et discussion

Face à la diversité des paramètres qui peuvent être modifiés via MIIC (sur l'interface Cytoscape), nous avons choisi de ne nous focaliser que sur un unique paramètre : le seuil de confiance. Ainsi nous avons construit, pour chaque set de données généré, de 2 à 10 réseaux différents, les premiers réseaux ayant un seuil de confiance élevé (par rapport à l'étendue de celui-ci) et les derniers ayant un seuil de confiance plus bas. Le nombre de réseau obtenu est fonction du nombre de gènes compris dans le dataset. Tous les autres paramètres par défaut sont restés inchangés par souci de compréhension et de clarté. Ne sont représentés ici que les réseaux qui nous ont semblé les plus pertinents et intelligibles. Notons tout de même que pour certains réseaux, il nous a fallu faire un choix entre précision et lisibilité du réseau. C'est sur ce dernier critère que nous nous sommes concentrées afin de pouvoir obtenir des données interprétables. Il faut prendre en compte que plus le seuil de confiance est élevé, moins les relations sont nombreuses, donc des

6. <http://pbil.univ-lyon1.fr/R/pdf/stage7.pdf>

7. <https://cran.r-project.org/web/packages/cluster/cluster.pdf>



Verny et al. submitted

FIGURE 4. Résultats issus de Verny et al. (les couleurs correspondent aux lignées décrites dans la figure 1).

gènes disparaissent du réseau ainsi formé. Cependant, ces relations restantes sont d'autant plus fiables.

2.1 Vue d'ensemble : clustering hiérarchique

Le dendrogramme obtenu avec *hclust*, à partir des 33 facteurs de transcription, est visualisable sur la figure 3. On peut remarquer que deux branches principales apparaissent, dont l'une présente une grande majorité (89%) des gènes impliqués préférentiellement dans la lignée hématopoïétique (en rouge). La seconde branche principale diverge elle-même en deux embranchements principaux. L'un forme un cluster constitué à 90% des gènes de la lignée primitive (en bleu). L'autre forme un cluster qui se répartit en deux plus petits clusters dont l'un contient uniquement des gènes précurseurs de l'endothélium (en rose/violet).

Cette répartition est donc relativement bien définie, bien que l'on retrouve des gènes de la lignée primitive dans presque tous les clusters. Le déroulement asynchrone de l'hématopoïèse dont l'article de Moignard et al. fait mention (figure 15) pourrait expliquer cette répartition : "le fait d'obtenir des cellules issues de différents stades embryonnaires dans les mêmes clusters suggère que la maturation des cellules du mésoderme précoce est asynchrone, avec notamment des cellules issues de différents stades embryonnaires qui présentent

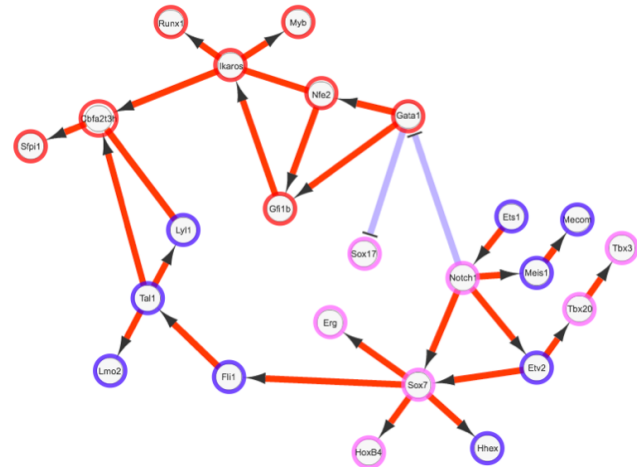


FIGURE 5. Compromis obtenu après différents tests sur les facteurs de transcription uniquement (les couleurs correspondent aux lignées décrites dans la figure 1).

un profil d'expression identique".

2.2 Vue d'ensemble : réseau obtenu par PC

Les réseaux obtenus sur l'ensemble des 42 gènes ou des 33 facteurs de transcription (FT) sont difficilement lisibles. Le réseau final a donc été choisi en faisant des compromis (via la modification du seuil de confiance), à partir du jeu de données des FT, afin d'être comparable au réseau présenté lors du cours de RESYS, figure 4. Nos résultats sont présentés figure 5. On y retrouve les mêmes *hubs* (noeuds dont le degré est élevé) : **Ikaros**, **Sox7**, **Tal1** et **Notch1**. Cependant, le *hub* formé par **Erg** a disparu dans notre réseau. Les autres noeuds présentent des degrés plus faibles, ce qui tend à émettre l'hypothèse d'une répartition non aléatoire. En effet, le nombre moyen de connexions entre les noeuds suit une répartition trop étendue pour être aléatoire. La distribution de degré pourrait donc suivre une loi de puissance. Le réseau représentant les relations qui régissent la différenciation des précurseurs hématopoïétiques chez l'embryon est donc un réseau de type *scale free*[5].

2.3 Tri par lignée - fonctionnel

Nous avons sélectionné trois graphes (un par lignée) via Miic : sur les figures 6, 7 et 8 (en annexe) vous pourrez observer les *correlation plot* obtenus (graphes basés sur le réseau, où les couleurs des liens sont échelonnées en fonction de leur coefficient de corrélation partielle). Donc ce qui est en bleu représente une inhibition (régulation négative) et ce qui est en rouge indique une activation (régulation positive).

Figure 6 (lignée hématopoïétique), le FT **Runx1** ne présente aucun lien sortant (le degré sortant d_N^{out} du noeud N est nul). Ceci semble cohérent avec le rôle majeur que joue **Runx1** dans la différenciation des cellules hématopoïétiques. En effet, ce gène est connu pour son implication dans le développement normal de l'hématopoïèse (des translocations chromosomiques se répercutant sur ce gène ont précisément été associées avec différents types de leucémies, dont la leucémie aigue lymphoblastique chez l'enfant[6]). C'est d'ailleurs celui-ci que les auteurs de l'article de référence utilisent comme marqueur pour les cellules 4SG par opposition aux cellules 4SFG au stade *four somite*.

Figure 7 (lignée endothéliale), on observe qu'il existe une régulation positive (non orientée) entre le FT **Sox7**, impliqué dans la différenciation endothéliale, et le FT central de la différenciation endothéliale **Erg**. De plus, **Sox7** semble avoir un rôle central dans la régulation transcriptionnelle de la lignée endothéliale : ceci est parfaitement cohérent sachant que **Sox7** est un répresseur de **Runx1** (facteur de transcription clé de la différenciation hématopoïétique)[4].

Figure 8 (lignée primitive) (en annexe), on retrouve certaines relations identiques à celles visibles figure 4 : **Fli1** en tant qu'activateur principal de **Etv6**, **Lyl1** comme cible de **Tal1**. Par contre certains liens ne sont pas en corrélation avec les données de la figure 4, tels que l'inhibition de **Meis1** par **Lyl1** ou le lien entre **Fli1** et **Tal1** qui est orienté dans la direction opposée.

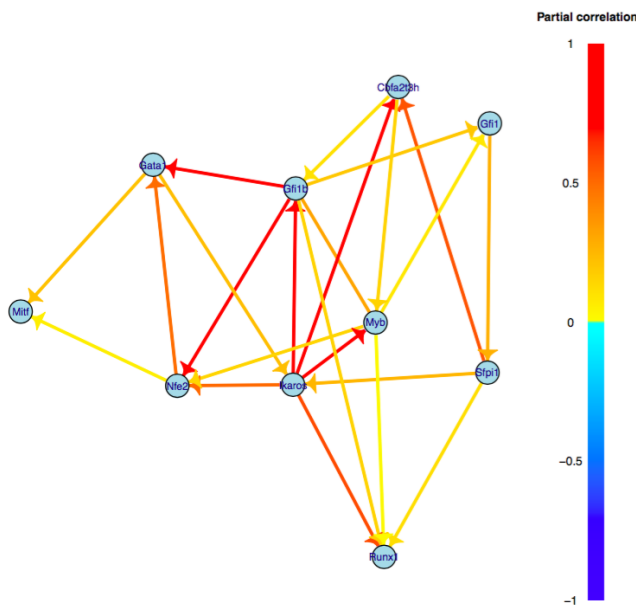


FIGURE 6. Lignée hématopoïétique

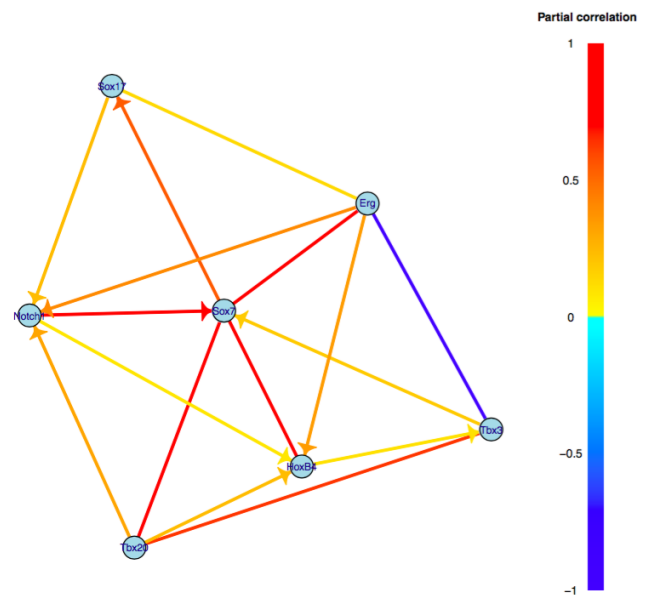


FIGURE 7. Lignée endothéliale

2.4 Tri par stade embryonnaire - temporel

Par souci de clarté ces figures ont été placées en annexe. De façon générale, il faut souligner l'importance des gènes marqueurs en tant qu'inducteurs des différents stades embryonnaires. Si l'on observe par exemple les différences entre les graphes 11 et 12, où dans un cas le réseau a été reconstruit avec l'ensemble des 42 gènes, et dans l'autre, avec les 33 FT uniquement, le rôle central des cadhérines **Cdh1** et **Cdh5** est indéniable.

Au cours des premiers stades (surtout PS et un peu NP), on peut observer le rôle central du FT **Tal1**. Au niveau du réseau, il constitue un *hub*. Or son nom est très équivoque, **Tal1** signifiant *T cell acute lymphocytic leukemia 1* (ou encore *erythroid differentiation factor*). Il s'agit donc justement d'un FT connu pour son rôle dans la différenciation hématopoïétique et son implication dans différents cas de leucémies.

Le stade 4SG est particulièrement marqué par l'inhibition des FT majoritaires des stades primitifs, dont **Ikaros** qui est un régulateur important pour les décisions du destin cellulaire (spécification, détermination et différenciation) pendant l'hématopoïèse. On pourrait donc supposer qu'à ce stade, le destin cellulaire est déjà établi, et la cellule engagée dans cette voie ne pourra pas en changer.

De même, au stade 4SFG on peut observer le rôle majeur de **Sox7**, un autre FT impliqué dans la régulation du développement embryonnaire et dans la détermination du destin cellulaire. Ce stade fait aussi état du rôle crucial du FT **Erg** (qui appartient à la famille des ETS ou

erythroblast transformation-specific) qui est précisément l'un des régulateurs clés de l'angiogenèse. L'action de ce FT se fait d'ailleurs via l'intervention des cadhérines **Cdh1** et **Cdh5** (visibles sur les réseaux issus des data-sets de 42 gènes), ce qui pourrait expliquer leur situation constituant des *hubs*, comme décrit ci-dessus[7].

Conclusion

D'une manière générale, on peut observer que le caractère asynchrone du développement des précurseurs hématopoïétiques soulève la difficulté de recréer un réseau causal. Aussi **Runx1** est absent des graphes représentatifs de population 4SG (où il est attendu) : ceci évoquerait une certaine autonomie de ce FT indispensable au bon déroulement de l'hématopoïèse. Une autre hypothèse possible est qu'il se trouve dans une boucle de régulation mais l'algorithme appliqué ne permet justement pas de modéliser des cycles de régulation (nous avons reconstruit des DAG ou *graphes acycliques dirigés*). Si une boucle de rétrocontrôle existait, elle ne pourrait pas être visualisée par ce type de réseau. Enfin, les *hubs* qui obtenus figure 5 font justement partie des FT indispensables pour le bon déroulement de l'hématopoïèse cités dans la littérature[8] (un schéma récapitulatif est disponible figure 16 [9]).

Quant aux améliorations, nous pourrions prolonger cette étude en modélisant des réseaux via d'autres outils de reconstruction (par exemple Aracne ou constraint-based). Cependant, de par l'extrême complexité des interactions et de par leur caractère quasi instantané, il est très difficile qu'une modélisation seule puisse mettre en lumière tous les liens inter-moléculaires ainsi que leur nature. Il faudrait donc qu'une recherche bibliographique et expérimentale plus vaste vienne compléter cette modélisation.

Références

- [1] Spirtes and al. 5.4.1 :82, (2000).
- [2] Cumano and al. Hématopoïèse intra-embryonnaire chez la souris : Emergence et caractérisation de cellules souches hématopoïétiques pendant le développement : aspects fondamentaux et cliniques. *Comptes rendus des séances de la Société de biologie et de ses filiales*, 189(4) :617–627, 1995.
- [3] Sahar Messouadi Anass Es-Seddiki, Anass Ayyad and Rim Amrani. Fanconi anemia : report of a new case. *Pan Afr Med J.*, 20(92), 2015.

- [4] Moignard and al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnology*, 33 :269–276, 2015.
- [5] Oliver Hein Michael Schwind Wolfgang König. Scale-free networks - the impact of fat tailed degree distribution on diffusion and communication processes. *Wirtschaftsinformatik*, 48(4) :267–275, 2006.
- [6] Bhojwani D and al. Etv6-runx1-positive childhood acute lymphoblastic leukemia : improved outcome with contemporary therapy. *Leukemia*, 26(2) :265–270, 2012.
- [7] Amsellem V Gebhardt F Sahnun K Haskard DO Dejana E Mason JC Randi AM. Birdsey GM1, Dryden NH. Transcription factor *erg* regulates angiogenesis and endothelial apoptosis through *ve-cadherin*. *Blood*, 111(7) :3498–3506, 2008.
- [8] Karen K. Hirschi Kathrina L. Marcelo, Lauren C. Goldie. Regulation of endothelial cell differentiation and specification. *Circulation Research*, 112 :1272–1287, 2013.
- [9] Shu Huang Ya Zhou Kohichiro Tsuji Bo Chen, Bin Mao and Feng Ma. Human embryonic stem cell-derived primitive and definitive hematopoiesis. *INTECH*, chap. 4, 2014.
- [10] Gaudin and Cumano. Les cellules souches hématopoïétiques : une double origine embryonnaire ? *Med Sci (Paris)*, 23(8-9) :681–684, 2007.

Annexes

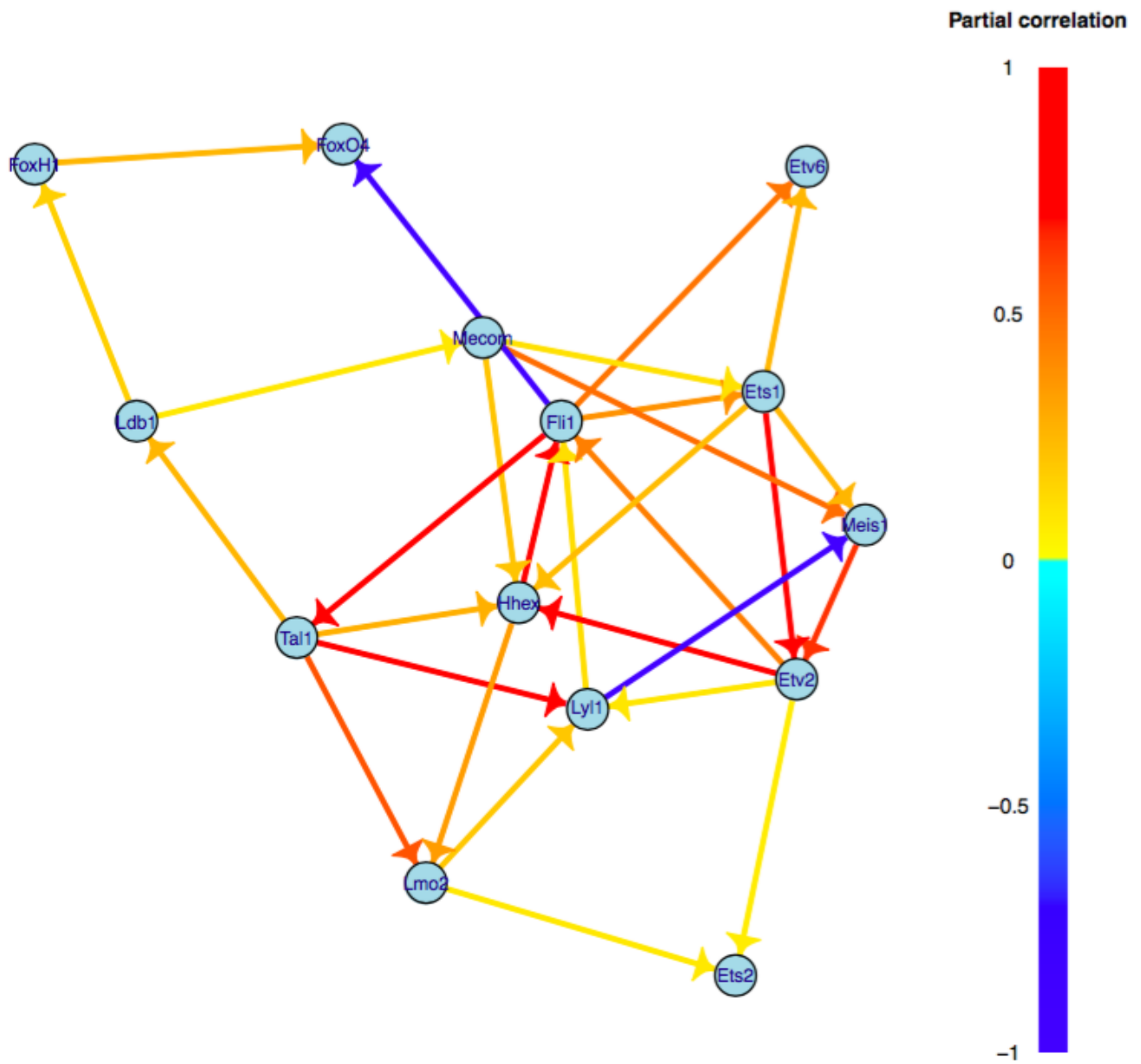


FIGURE 8. Lignée primitive

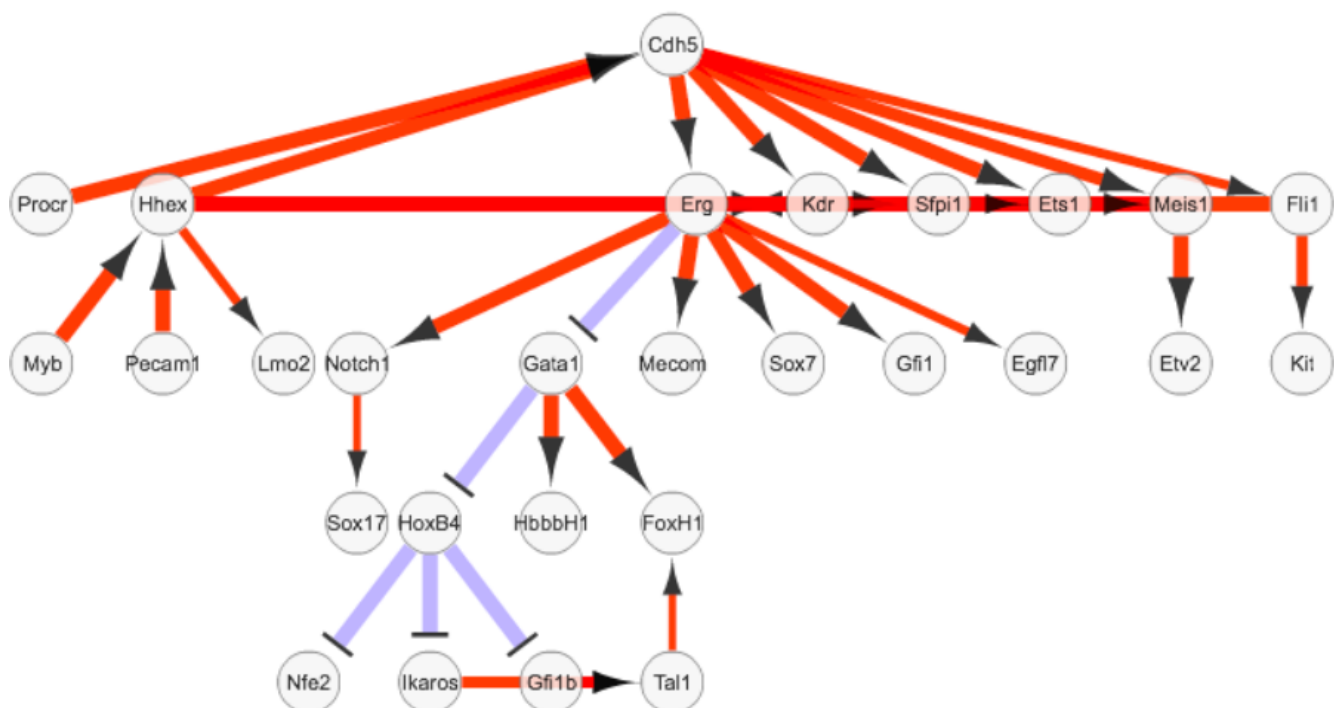


FIGURE 9. Population des cellules 4SG (à partir des 42 gènes). Le gène marqueur Cdh5 agit comme régulateur. Cascade d'inhibition Erg - Gata1 - HoxB4 - Ikaros

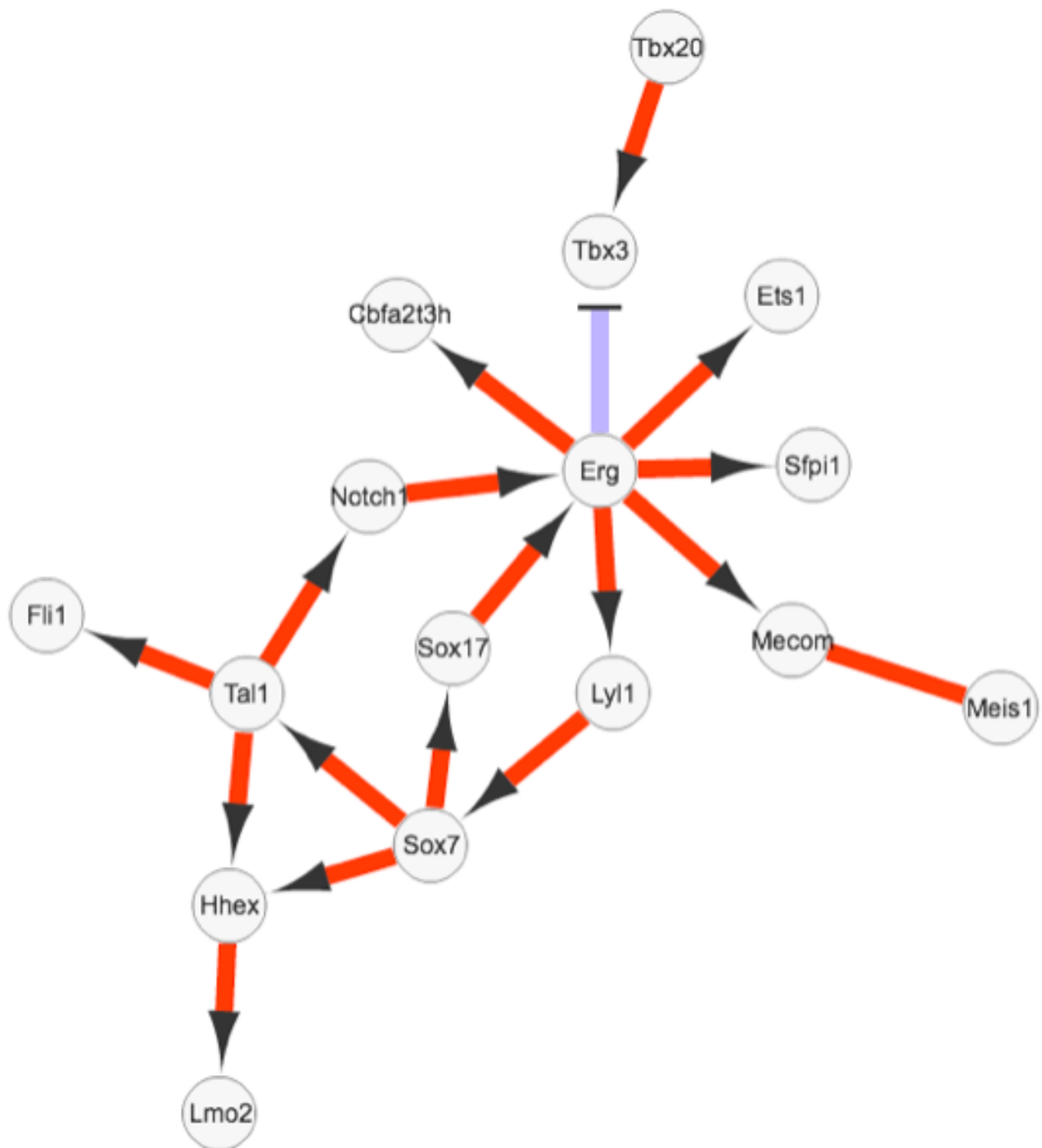


FIGURE 10. Population des cellules 4SFG (à partir des FT). On peut remarquer le rôle central du FT Erg.

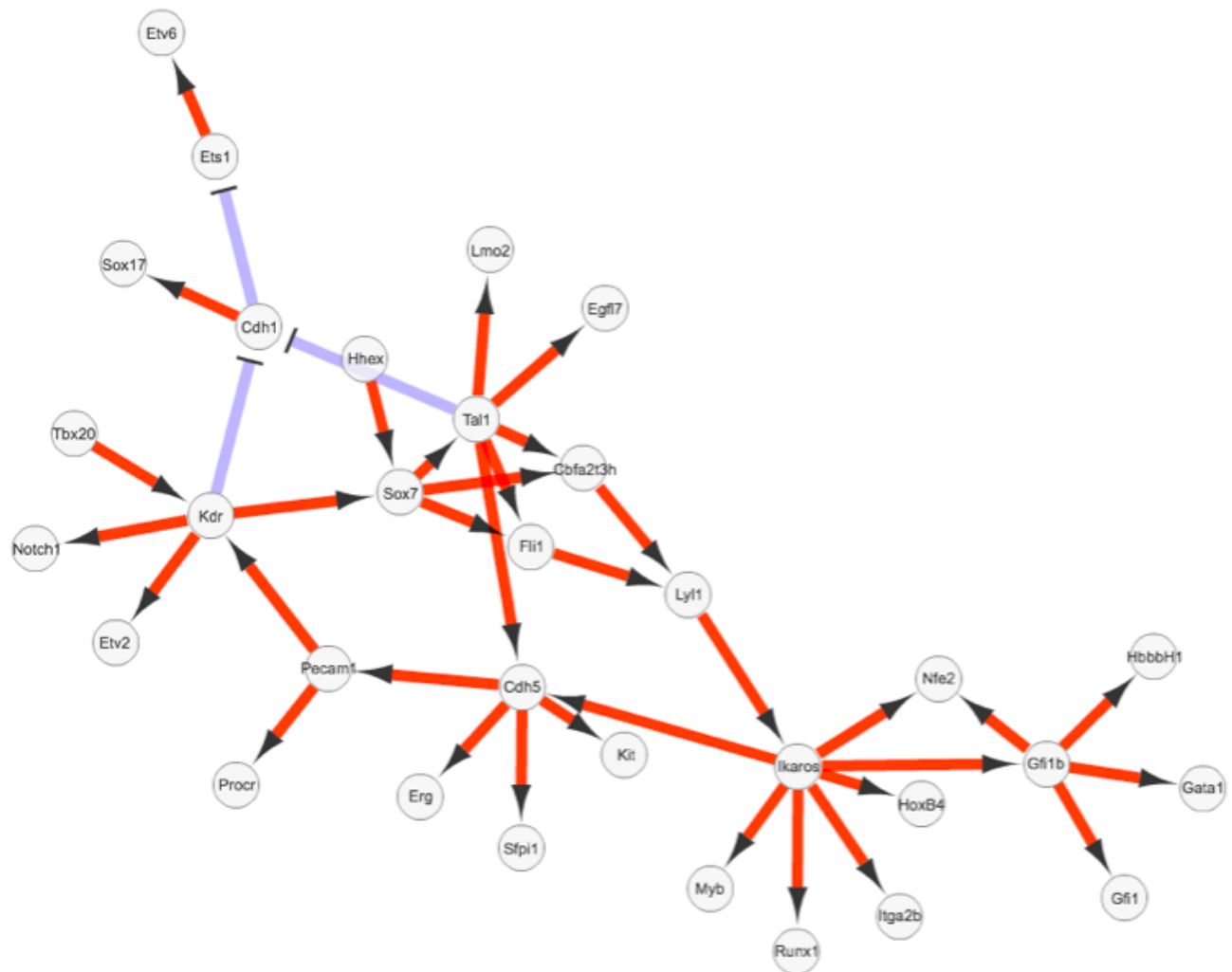


FIGURE 11. Population des cellules NP (à partir des 42 gènes). Notons l'inhibition de l'inhibiteur de Ets1 et activateur de Sox17 (Cdh1). Rôle central de Kdr, Tal1, Cdh5, Ikaros, Gfi1b

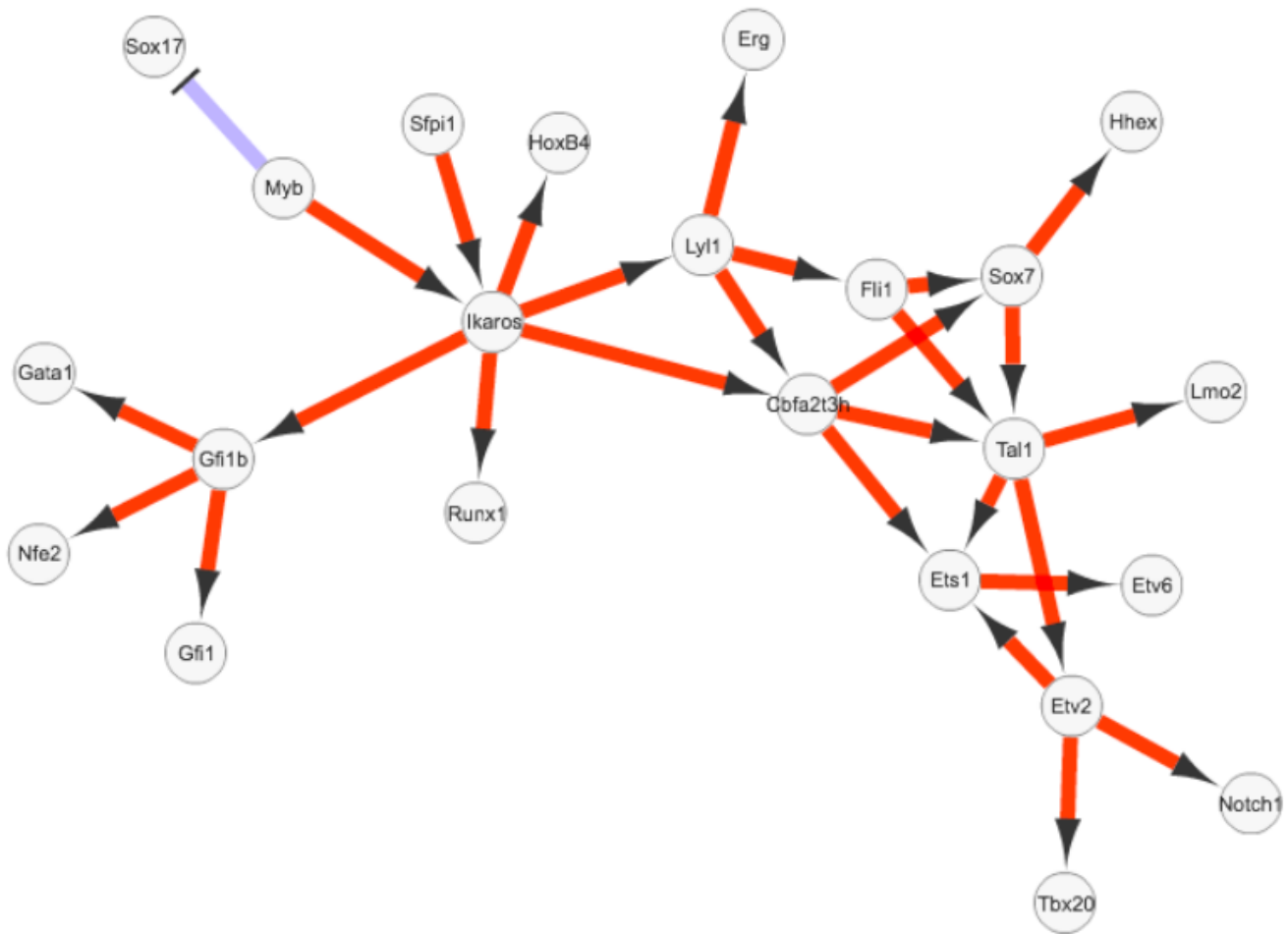


FIGURE 12. Population des cellules NP (à partir des 33 gènes) : perte des interactions liées aux gènes marqueurs Cdh1 et Cdh5 par rapport au graphe avec 42 gènes, fondamentales pour la différenciation - le graphique précédent permet donc une visualisation plus précise des interactions au stade NP

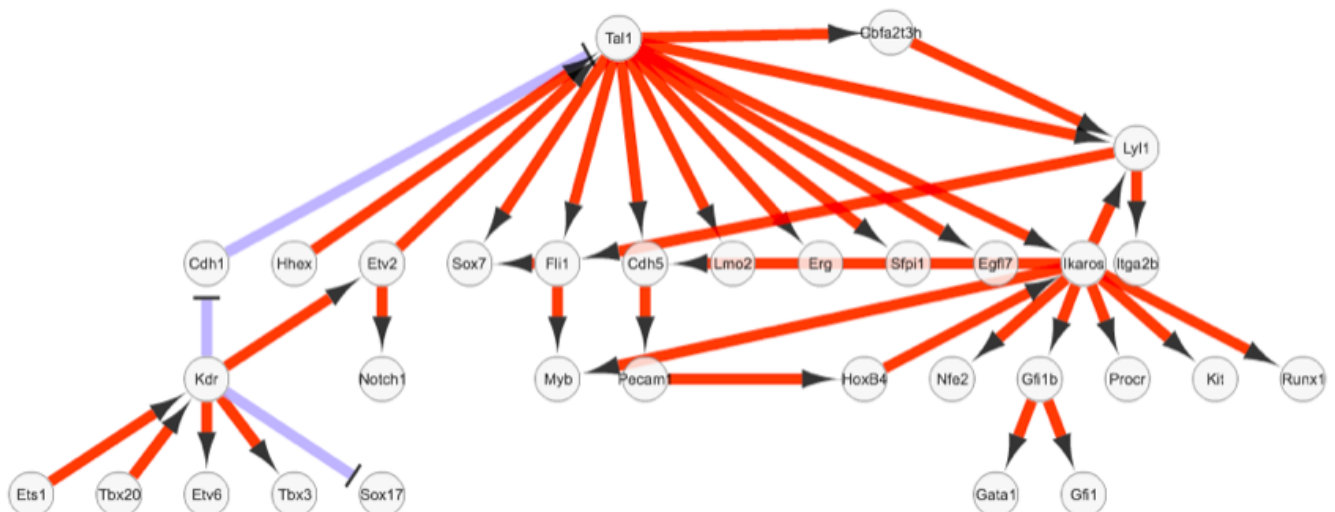


FIGURE 13. Population des cellules PS (à partir des 42 gènes). Kdr agit comme régulateur négatif du gène marqueur Cdh1 qui lui-même agit négativement sur Tal1. Rôle central de Kdr, Tal1 et Ikaros. Inhibition de Sox17.

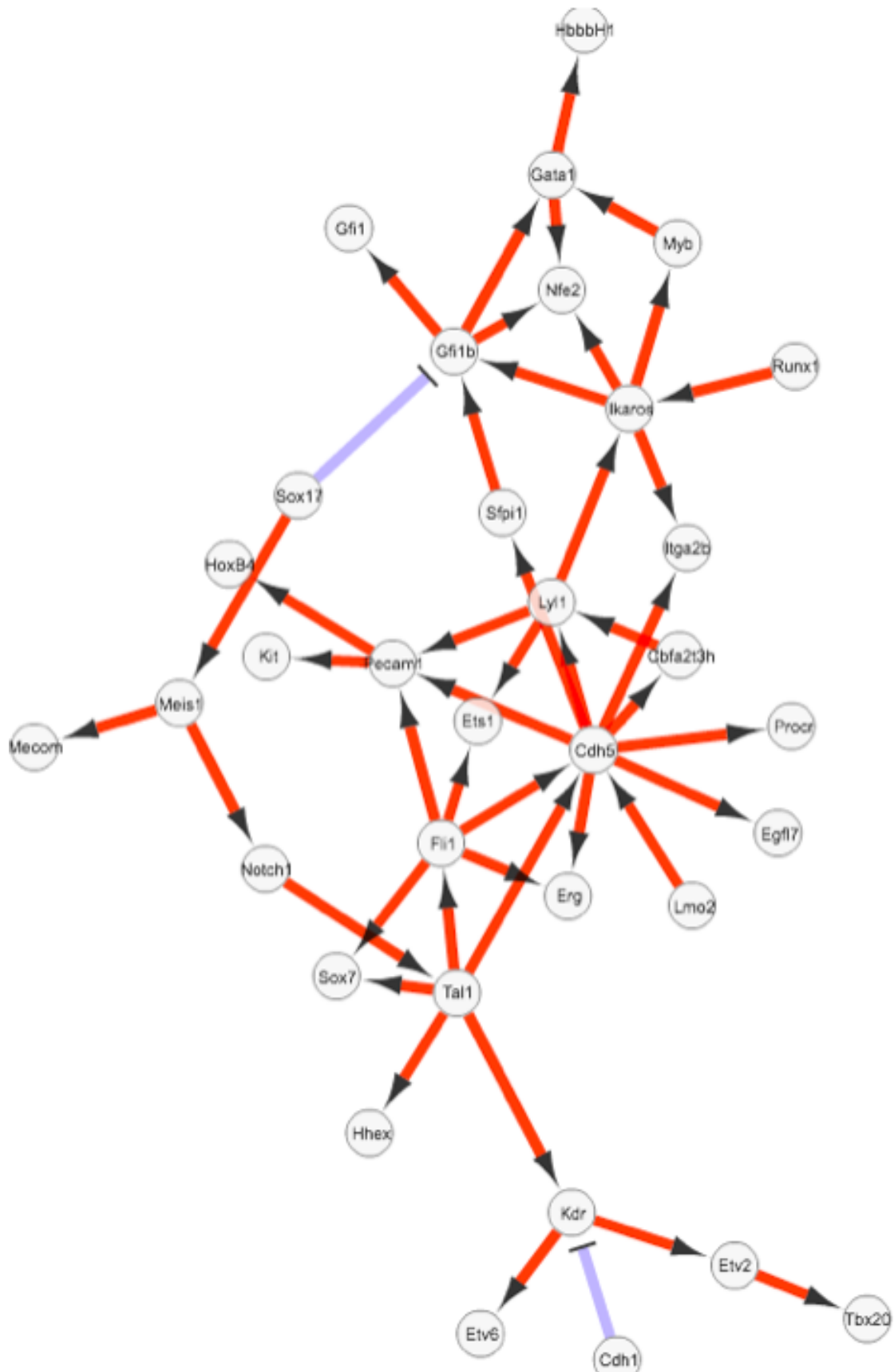


FIGURE 14. Population des cellules HF (à partir des 42 gènes). Notons l'inhibition de Kdr par Cdh1 et de Gfi1b par Sox17. Rôle central de Tal1, Cdh1, Fli1, Lyl1, Ikaros.

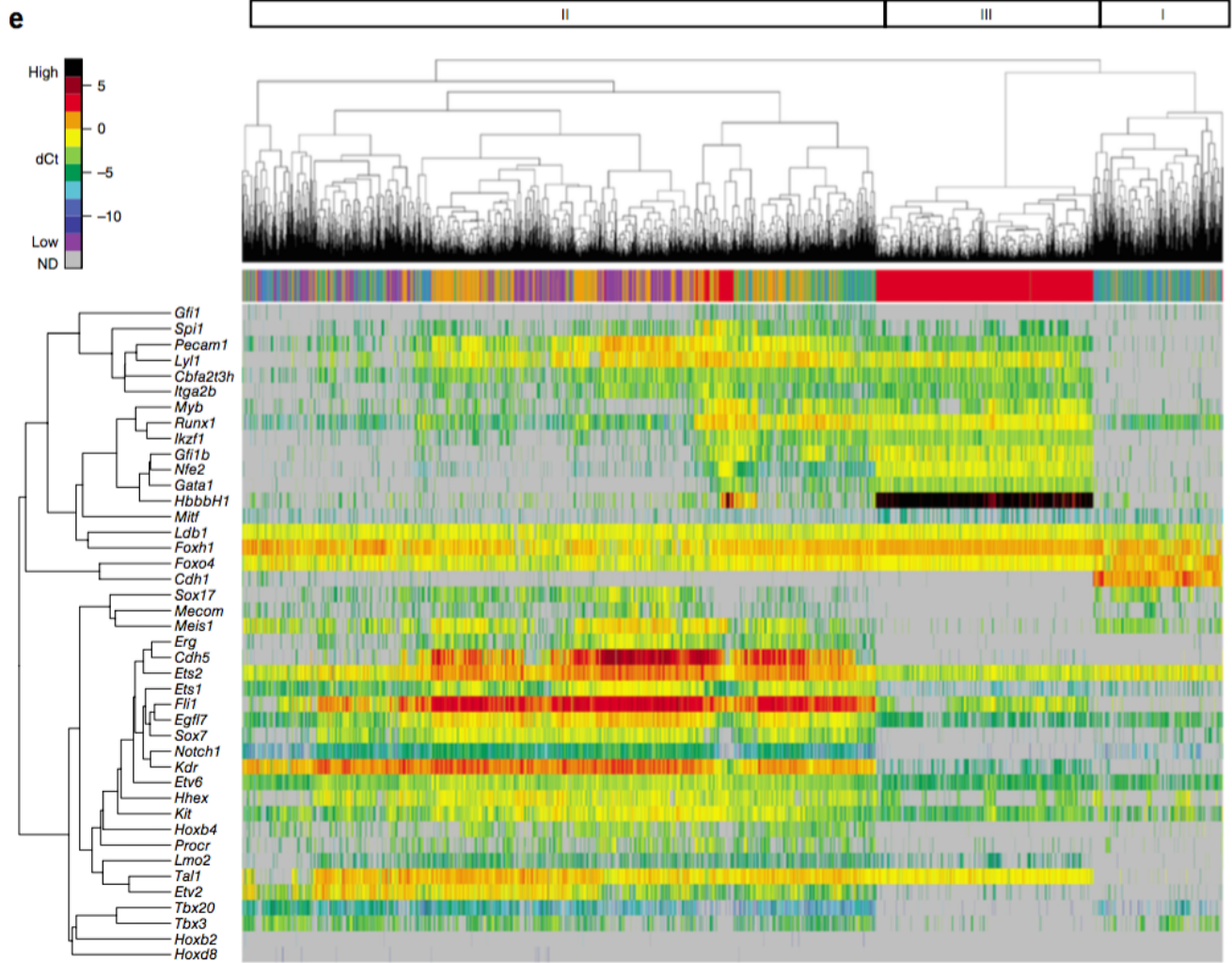


FIGURE 15. Résultats du clustering hiérarchique non supervisé issus de *Moignard et al.* à partir de l'analyse par *single cell*. On peut observer le niveau d'expression pour chaque gène dans toutes les cellules. Les colonnes représentent les cellules et les lignes les gènes. Les couleurs correspondent au stade embryonnaire d'où chaque cellule a été extraite.

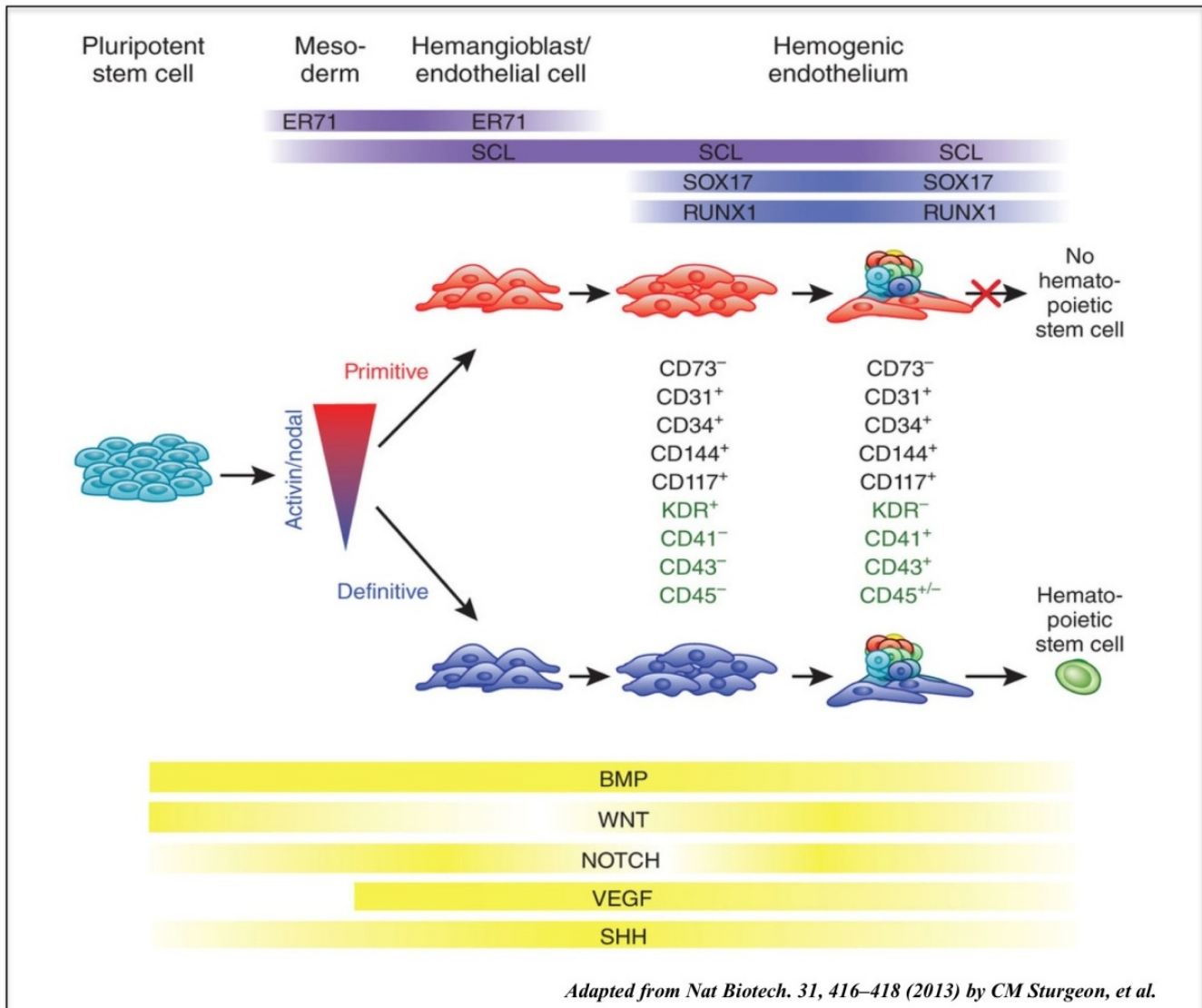


FIGURE 16. Développement de l'hémangioblaste (précurseur commun des cellules endothéliales et hématopoïétiques) et de l'épithélium hémogénique à partir de cellules pluripotentes[9]

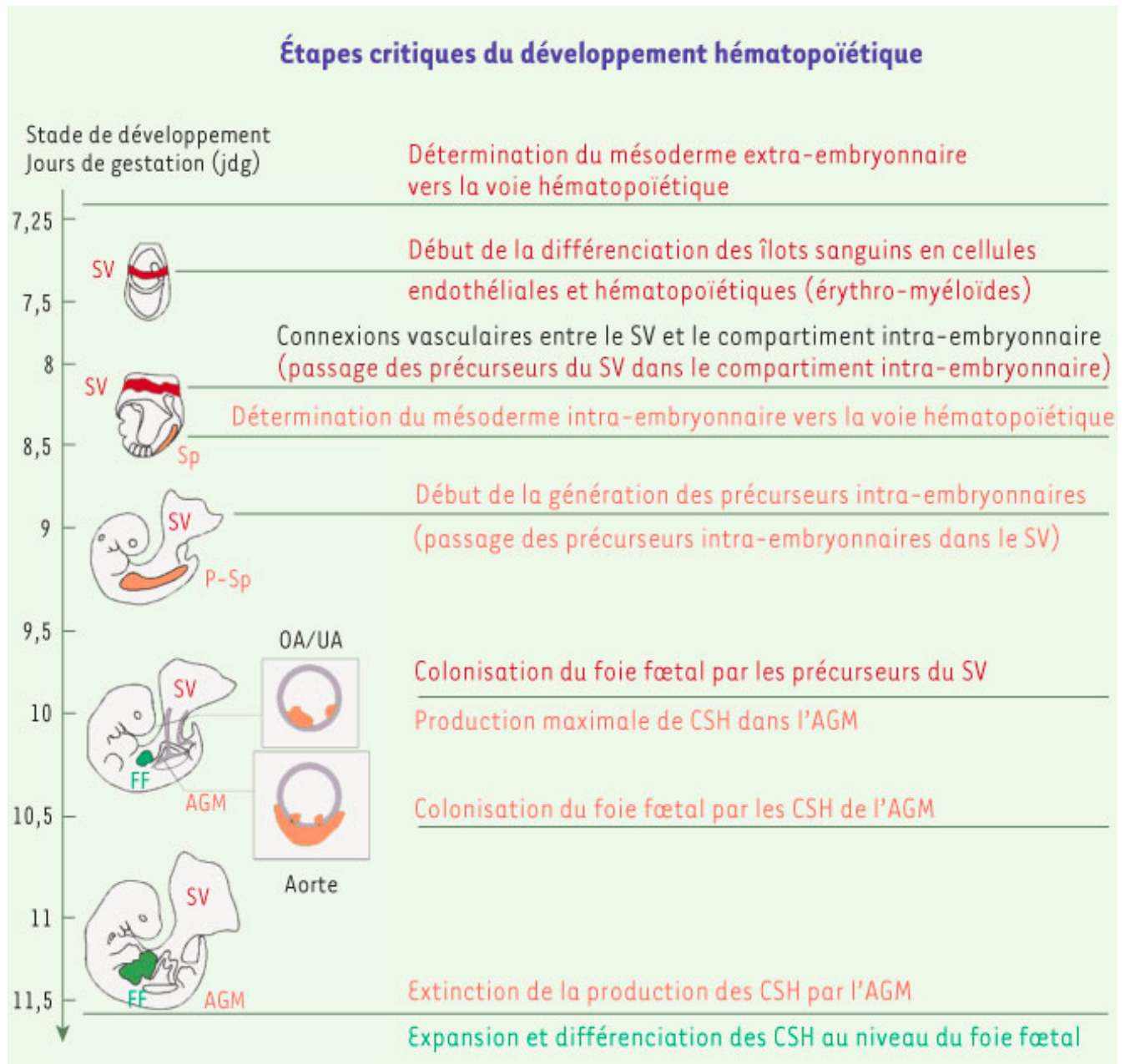


FIGURE 17. Détails du développement précoce de *Mus musculus*[10] du stade E7 à E11.5, dans le compartiment extra-embryonnaire (en rouge) et intra-embryonnaire (en jaune). En encart figurent les sites impliqués dans la génération des CSH, c'est-à-dire l'aorte et sa partie ventrale (et les artères omphalomésentérique (OA) et ombilicale (UA)). AGM : aorte-gonades-mésonephros ; FF : foie fœtal ; P-Sp : splanchnopleure para-aortique ; Sp : splanchnopleure ; SV : sac vitellin