



HAL
open science

Etude de l'architecture clonale des leucémies aiguës myéloïdes. Application à la mesure de la maladie résiduelle

Pierre Hirsch

► **To cite this version:**

Pierre Hirsch. Etude de l'architecture clonale des leucémies aiguës myéloïdes. Application à la mesure de la maladie résiduelle. Médecine humaine et pathologie. Université Pierre et Marie Curie - Paris VI, 2016. Français. NNT: 2016PA066088 . tel-01696019

HAL Id: tel-01696019

<https://theses.hal.science/tel-01696019>

Submitted on 30 Jan 2018

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.

Université Pierre et Marie Curie

Ecole doctorale 394 : Physiologie, Physiopathologie et Thérapeutique

Centre de recherche Saint-Antoine UMRS 938. Equipe AMIEH

Etude de l'architecture clonale des leucémies aiguës myéloïdes *Application à la mesure de la maladie résiduelle*

Par Pierre HIRSCH

Thèse de doctorat de Sciences de la Vie

Dirigée par M. François DELHOMMEAU

Présentée et soutenue publiquement le 28 janvier 2016

Devant un jury composé de :

M. Thierry JAFFREDO, Président du jury

Mme Virginie PENARD-LACRONIQUE, Rapporteur

M. Claude PREUDHOMME, Rapporteur

M. Olivier KOSMIDER, Examineur

L'ÉLÈVE

Je connais mes saisons, n'est-ce pas, monsieur ?

LE PROFESSEUR

Mais oui, mademoiselle... ou presque. Mais ça viendra. De toute façon, c'est déjà bien. Vous arriverez à les connaître, toutes vos saisons, les yeux fermés. Comme moi.

L'ÉLÈVE

C'est difficile.

LE PROFESSEUR

Oh, non. Il suffit d'un petit effort, de la bonne volonté, mademoiselle. Vous verrez. Ça viendra, soyez-en sûre.

L'ÉLÈVE

Oh, je voudrais bien, monsieur. J'ai une telle soif de m'instruire. Mes parents aussi désirent que j'approfondisse mes connaissances. Ils veulent que je me spécialise. Ils pensent qu'une simple culture générale, même si elle est solide, ne suffit plus, à notre époque.

LE PROFESSEUR

Vos parents, mademoiselle, ont parfaitement raison. Vous devez pousser vos études. Je m'excuse de vous le dire, mais c'est une chose nécessaire. La vie contemporaine est devenue très complexe.

Eugène Ionesco, La Leçon.

Remerciements

Merci à tous ceux qui m'ont soutenu pendant et avant ce travail, et qui continueront de le souhaiter à me supporter au quotidien.

Merci à Jean-Pierre Marie, Ollivier Legrand et Mohamad Mohty sans qui ce travail n'aurait jamais été commencé.

Merci à Ruoping Tang, pour son soutien, ses conseils et son travail depuis le master 2, et sans qui tout aurait été tellement plus compliqué.

Merci à tous ceux qui ont participé à ce travail, et qui continuent de supporter mon impatience au quotidien, et notamment à Christophe Marzac, Fanny Fava et Florence Lorre qui ont dû recevoir chacun plusieurs milliers de messages à la recherche d'échantillons.

Merci à tous les membres du service d'Hématologie clinique et du laboratoire d'Hématologie biologique de l'hôpital Saint-Antoine et de l'Hôpital Trousseau, pour leur aide précieuse au quotidien.

Merci à Luc Douay et à toute son équipe pour leur assistance technique, leurs conseils, et leur accueil.

Merci à Fawzia Louache et Yanyan Zhang pour leur collaboration fructueuse.

Merci à tous les patients qui ont accepté de participer à ces études.

Merci aux membres du jury.

Merci à François pour son encadrement (juste la bonne dose), sa patience et sa rigueur scientifique (et ses coups de gueule). En espérant que tout continue longtemps.

Merci à ma famille.

Merci à Cécile, pour tout. Merci à Agathe et Clément, pour tout le reste.

Sommaire

Remerciements	2
Sommaire	3
Abréviations	4
Introduction	5
A- Généralités sur l'hématopoïèse	5
A.1- Cellules souches hématopoïétiques.....	6
A.2- Différenciation hématopoïétique.....	6
A.3- Régulation de la différenciation cellulaire	7
A.4- Niche hématopoïétique	9
B- Leucémies aiguës myéloïdes : généralités	10
B.1- Définition	10
B.2- Epidémiologie	11
B.3- Facteurs étiologiques.....	11
B.4- Traitements et résultats.....	11
C- Leucémie aiguës myéloïdes : aspects biologiques et physiopathologie.....	14
C.1- Données classiques et classification des LAM	14
C.2- Données moléculaires récentes	20
C.3- Phylogénie et dynamique clonale dans les LAM	30
C.4- Evénements précoces et marqueurs de cellules souches pré-leucémiques.....	35
C.5- Formes familiales de LAM et prédispositions génétique	38
C.6- Mesure de la maladie résiduelle	40
D- Objectifs et problématiques du travail	46
Résultats : Article 1	49
Article 1 : Données supplémentaires.....	82
Article 1 : Commentaires	113
Résultats : Article 2	118
Article 2 : Commentaires	128
Données complémentaires.....	130
A- Phylogénie clonale des LAM de l'enfant.....	130
B- Mesure de la maladie résiduelle en NGS	134
Discussion et Conclusion	138
Bibliographie	142
Table des illustrations.....	153
Table des tableaux	154

Abréviations

ADN : Acide désoxyribonucléique

CBF : *Core-binding factor*

CSH : Cellule souche hématopoïétique

EPO : Erythropoïétine

FAB : Classification Française, Américaine et Britannique

FISH : Hybridation in situ par fluorescence

FLT3 : *Fms-like protein kinase 3*

FT : Facteur de transcription

GVH : maladie du greffon contre l'hôte

HCLA : Hématopoïèse clonale liée à l'âge

ITD : Duplication interne en tandem

LA : Leucémie aiguë

LAL : Leucémie aiguë lymphoblastique

LAM : Leucémie aiguë myéloïde

LMMC : leucémie myélo-monocytaire chronique

MRD : Maladie résiduelle

OMS : Organisation mondiale de la santé

RC : Rémission complète

SCF : *Stem cell factor*

TPO : Thrombopoïétine

USP : Unité de sang placentaire

VAF : Fréquence allélique des variants

Introduction

A- Généralités sur l'hématopoïèse

L'hématopoïèse est le processus qui permet à partir d'une seule cellule souche hématopoïétique (CSH) la fabrication de toutes les cellules sanguines. Les grands principes du modèle classique de l'hématopoïèse sont exposés très brièvement dans cette partie. L'hématopoïèse a lieu chez l'adulte sain dans la moelle osseuse. Le modèle classique de l'hématopoïèse est pyramidal et hiérarchisé : toutes les cellules médullaires et sanguines dérivent d'un contingent de CSH ; à partir des CSH, les cellules vont se multiplier et se différencier en progéniteurs plus engagés dans la voie lymphoïde ou la voie myéloïde, puis en précurseurs de plus en plus différenciés, jusqu'à aboutir aux éléments matures du sang (figure 1).

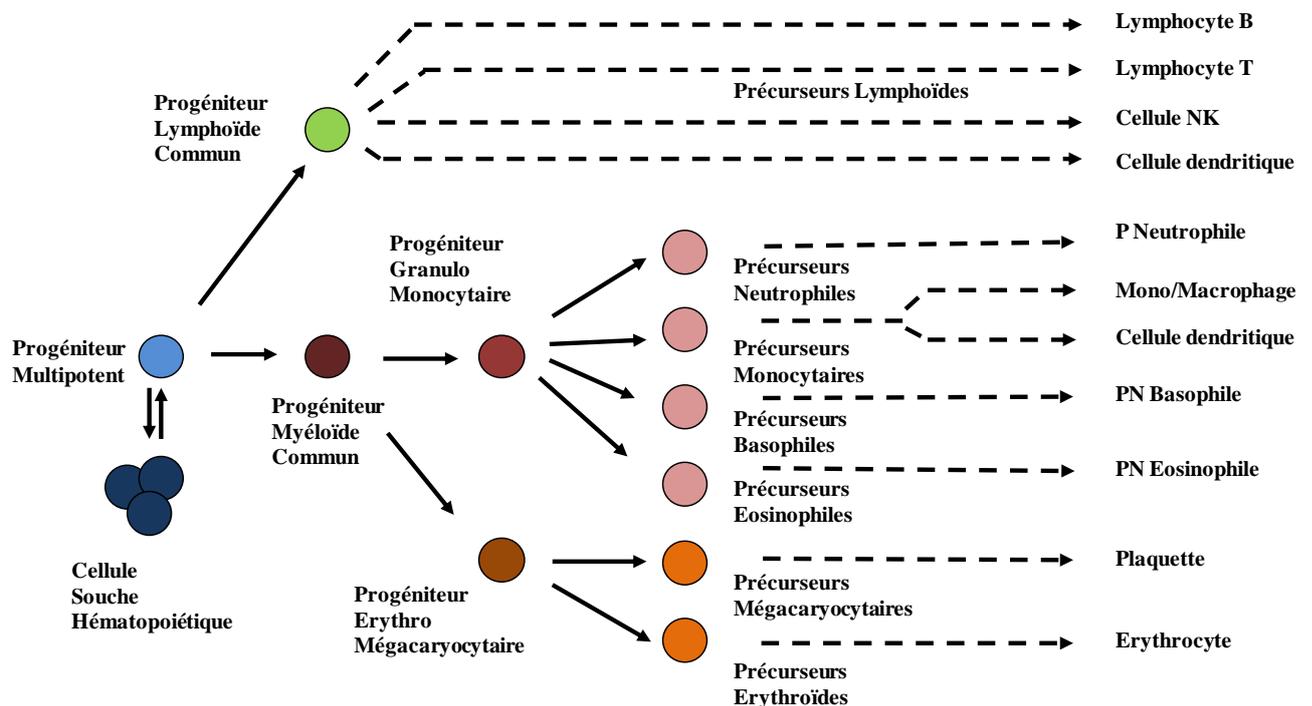


Figure 1 : Organisation schématique de l'hématopoïèse. A partir des CSH, les cellules se différencient en progéniteurs puis en précurseurs de plus en plus engagés, jusqu'à l'obtention des cellules sanguines matures.

A.1- Cellules souches hématopoïétiques

L'existence des CSH a été démontrée par la capacité des cellules médullaires ou spléniques de rongeurs à reconstituer totalement l'hématopoïèse après greffe, chez des sujets ayant subi une irradiation par une dose entraînant des lésions hématopoïétiques irréversibles (aplasie médullaire) ¹. Les CSH sont très rares dans l'organisme, et on estime leur nombre à environ 0.001% des cellules médullaires ². Leurs principales caractéristiques sont une capacité d'auto-renouvellement, qui permet de maintenir leur nombre constant, une capacité de différenciation, qui va aboutir à la production de tous les types cellulaires de l'hématopoïèse, et une capacité de quiescence, permettant une réserve en cas nécessité de production cellulaire accrue. Afin de protéger leur patrimoine génétique, le nombre de division cellulaire des CSH est très faible. On estime que les CSH non quiescentes se divisent une fois tous les 200 à 300 jours chez l'homme alors que 10^{11} à 10^{12} cellules sanguines sont produites chaque jour ³.

Beaucoup de travaux ont cherché à caractériser le phénotype des CSH afin de faciliter leur isolation et leur étude. Des tests fonctionnels de repopulation hématopoïétique *in vivo* après transplantation ont permis l'identification de marqueurs de surface de plus en plus spécifiques. Ainsi il a été d'abord démontré que les CSH sont comprises dans la population $CD34^+CD38^-$ et n'exprimant pas d'autre marqueur d'engagement dans d'autres lignées (lin^- , pour « *lineage negative* ») ⁴. Les capacités d'efflux du Hoechst 33342 ont également été utilisées pour définir une "side population" dans laquelle les CSH sont comprises ⁵. Plus récemment, le phénotype des CSH a été précisé avec l'utilisation d'autres marqueurs comme CD45RA, CD90, CD49, ou d'autres ^{6,7}.

A.2- Différenciation hématopoïétique

La différenciation des cellules médullaires en cellules sanguines à partir des CSH se fait en plusieurs étapes (figure 1).

En premier lieu, les progéniteurs hématopoïétiques dérivent directement des CSH. Ils sont capables de prolifération, et de différenciation en plusieurs types cellulaires plus matures. On compte parmi eux les progéniteurs multipotents, qui sont les plus immatures, et peuvent se différencier en progéniteur myéloïde commun ou en progéniteur lymphoïde, à l'origine des lymphocytes B, T, NK ⁸. Les progéniteurs myéloïdes communs ⁹ peuvent se différencier en

progéniteur érythro-mégacaryocytaire à l'origine des lignées érythroïdes et plaquettaires, ou en progéniteur granulo-monocytaire, à l'origine des lignées granuleuses et monocytaires. Comme pour les CSH, l'expression d'un ensemble de marqueurs de surface permet l'identification par cytométrie en flux de chacun des sous-types de progéniteurs^{7,10}.

A.3- Régulation de la différenciation cellulaire

Le processus de différenciation hématopoïétique est finement régulé, afin de permettre le bon équilibre des populations de cellules matures, et une réaction adaptée en cas de stress (hémorragie, infection ou autre...).

Les cytokines sont les principaux régulateurs extrinsèques de la différenciation hématopoïétique¹¹. Il s'agit de protéines synthétisées au niveau de différents organes, et capables de circuler dans le sang jusqu'aux cellules médullaires. Leur fixation sur des récepteurs membranaires spécifiques va aboutir à l'activation de voies de signalisation et de facteurs de transcriptions (FT) sous-jacents. Parmi ces cytokines, l'érythropoïétine (EPO), synthétisée principalement au niveau rénal, est responsable de la différenciation érythroïde, et la thrombopoïétine (TPO), de synthèse rénale et hépatique, est responsable de la différenciation plaquettaire. Les différenciations granuleuses et monocytaires sont régulées par l'association de nombreuses cytokines, parmi lesquelles l'interleukine 3, l'interleukine 6, le G-CSF, ou le GM-CSF. Certaines cytokines comme le *stem cell factor* (SCF), la TPO ou le FLT3-ligand jouent également un rôle dans les phases très précoces de l'hématopoïèse. L'expression des récepteurs aux cytokines est variable en fonction des stades de différenciation des progéniteurs et des précurseurs hématopoïétiques. Cette régulation permet une réponse différente au niveau cellulaire selon la présence ou l'absence des récepteurs, et selon l'exposition aux différentes cytokines, avec un engagement progressif vers les différentes lignées selon les besoins^{12,13}.

De manière schématique, l'engagement vers les différentes voies de différenciation myéloïde est lié à l'expression d'un FT et à la régulation négative de son antagoniste dans les progéniteurs ou les précurseurs, secondairement à l'activation des récepteurs aux cytokines¹⁴. Les FT vont ensuite se lier à l'ADN, et engager un programme d'expression génique spécifique de la différenciation de chaque lignée¹⁵⁻²⁰. Ainsi par exemple, l'expression du FT *GATA1* va engager les cellules vers une différenciation érythro-mégacaryocytaire, alors que le FT antagoniste *PU.1* va les engager vers la différenciation granuleuse. L'ensemble des FT

impliqués dans l'hématopoïèse se régulent globalement deux à deux, et la présence ou l'absence d'un des FT va être responsable de l'engagement vers une voie de différenciation. L'ensemble des principaux FT impliqués dans l'hématopoïèse myéloïde est résumé dans la figure 2.

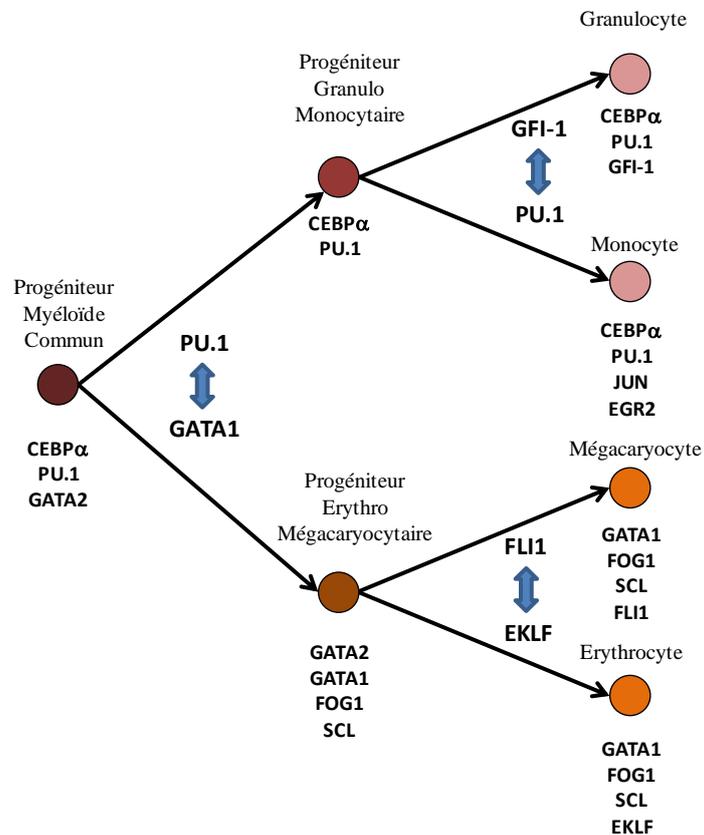


Figure 2 : Facteurs de transcriptions impliqués dans la régulation de l'hématopoïèse myéloïde. Les facteurs de transcription exprimés dans les différentes populations cellulaires sont listés sous chaque type cellulaire. L'équilibre des facteurs régissant l'engagement vers chaque lignée est représenté entre chaque branche.

D'autres systèmes de régulation plus complexes peuvent intervenir dans l'hématopoïèse en plus des mécanismes déjà décrits. On peut citer notamment un contrôle épigénétique par la régulation de la méthylation de l'ADN ou de l'acétylation des histones²¹⁻²³, ou le rôle de certains micro-ARN dans la maintien des CSH, et lors de la différenciation cellulaire²⁴.

A.4- Niche hématopoïétique

La niche hématopoïétique est l'environnement médullaire nécessaire à la survie, au maintien en quiescence et à l'entrée en division des CSH²⁵. Il s'agit d'un environnement constitué de nombreux sous-types cellulaires capables de sécréter, d'acheminer, et de réguler les facteurs extrinsèques responsables de l'activité des CSH. Cette niche est organisée schématiquement en deux compartiments (niche vasculaire à proximité des capillaires artériels ou veineux, et niche endostéale à proximité des ostéoblastes). Les CSH sont majoritairement observées dans la niche vasculaire, à proximité des vaisseaux²⁶. Au sein de la niche, les cellules endothéliales, les cellules souches mésenchymateuses, les ostéoblastes et les macrophages vont jouer un rôle essentiel dans la régulation des CSH, notamment par la sécrétion de SCF²⁷ qui interagit avec les CSH (récepteur KIT), de CXCL12 qui interagit avec CXCR4²⁸, ou d'autres facteurs²⁹. Le système nerveux sympathique est également impliqué dans la mobilisation des CSH en périphérie³⁰. Enfin, la niche participe à la régulation de la différenciation vers des cellules hématopoïétiques plus matures, comme par exemple la différenciation lymphocytaire B³¹.

B- Leucémies aiguës myéloïdes : généralités

B.1- Définition

La définition des leucémies aiguës (LA) dérive de la conception hiérarchique de l'hématopoïèse. Les LA sont en effet considérées dans le modèle classique de la leucémogénèse, comme des maladies clonales des progéniteurs hématopoïétiques, qui vont perdre leur capacité normale de différenciation, et acquérir des capacités anormales de prolifération et de survie. Ces progéniteurs hématopoïétiques anormaux, appelés blastes, vont envahir la moelle osseuse, aux dépens des lignées hématopoïétiques normales, ce qui conduit à un défaut de production des éléments normaux du sang (hématies, plaquettes et leucocytes).

D'après la classification des tumeurs du système hématopoïétique publiée en 2008 par l'organisation mondiale de la santé (OMS)³², le diagnostic de LA est porté, sauf exception, devant une proportion de blastes supérieure à 20 % des cellules nucléées dans la moelle osseuse. La LA sera dite myéloïde (LAM) si ces blastes sont d'origine myéloïde, ce qui peut être affirmé par des caractères cytologiques (notamment par la présence de corps d'Auer dans le cytoplasme des blastes), par des caractères cytochimiques (coloration à la myéloperoxydase ou au noir Soudan), ou par une analyse en cytométrie de flux³³. Il faut cependant noter plusieurs exceptions à cette définition. En effet, quand plus de 80 % des cellules nucléées de la moelle sont de lignée monocytaire, le diagnostic de LA monocytaire (ou monoblastique) pourra être retenu. De même, le diagnostic d'érythroleucémie sera porté si plus de 50 % des cellules nucléées médullaires appartiennent à la lignée érythroïde, et si au moins 20 % des cellules n'appartenant pas à cette lignée sont des myéloblastes. Les LA mégacaryoblastiques, de diagnostic souvent difficile cytologiquement, seront affirmées par la présence d'un nombre de blastes supérieur à 20 %, dont au moins la moitié appartiennent à la lignée mégacaryocytaire, définie par la présence de marqueurs spécifiques en cytométrie de flux (CD41, CD42 ou CD61).

D'autre part, la présence de certaines anomalies cytogénétiques spécifiques dans la population de cellules blastiques (t(15;17)(q22;q12), inv(16)(p13.1q22), ou t(8;21)(q22;q22)), suffit à affirmer le diagnostic de LAM, même si le compte des blastes est inférieur à 20 %.

B.2- Epidémiologie

D'après l'analyse du registre américain des cancers, les LAM représentent environ 25% de toutes les leucémies de l'adulte. Elles restent cependant des maladies très rares, avec une incidence ajustée à l'âge de 3,9 /100 000 habitants/an sur la période de 2007 à 2012, ce qui représente moins de 3 % de la totalité des cancers. Sur la même période, l'incidence est discrètement plus élevée chez les hommes que chez les femmes (4,87/100000 habitants/an contre 3,26 /100000 habitants /an) et augmente avec l'âge. L'âge moyen au diagnostic est de 67 ans aux Etats-Unis ³⁴.

B.3- Facteurs étiologiques

Certains facteurs environnementaux favorisant l'apparition des LAM ont été identifiés, comme l'exposition aux radiations ionisantes (radiothérapie ou autres), ou au benzène. L'utilisation de certains produits de chimiothérapie, et notamment des agents alkylants et des inhibiteurs de la topoisomérase II, peut également favoriser le développement secondaire d'une LAM ³⁵⁻³⁷. Les LAM induites par les traitements sont considérées comme des entités particulières dans la classification de l'OMS ³².

De plus, certaines pathologies hématologiques, comme les syndromes myélodysplasiques, les néoplasmes myéloprolifératifs, ou les aplasies médullaires, peuvent aussi prédisposer à l'apparition de LAM. Enfin, certains syndromes génétiques congénitaux sont associés à une fréquence plus élevée de LAM, comme par exemple le syndrome de Down (trisomie 21), la maladie de Fanconi, ou le syndrome de Li-Fraumeni ³⁸. Les LAM secondaires aux syndromes myélodysplasiques et les LA associées au syndrome de Down sont des entités à part dans la classification de l'OMS ³².

B.4- Traitements et résultats

En l'absence de traitement, les LAM sont d'évolution rapidement défavorable en quelques semaines ou quelques mois, en raison des conséquences des cytopénies (infections sévères et hémorragies).

A l'heure actuelle, les traitements consistent en une ou plusieurs cures de chimiothérapie, associées éventuellement à une intensification thérapeutique de clôture par

greffe de CSH allogéniques. Ces traitements intensifs ne peuvent être proposés que chez des sujets sans comorbidité majeure, en raison de leurs effets secondaires potentiels.

Le plus souvent, le traitement initial, ou chimiothérapie d'induction de la rémission complète (RC), repose sur l'association d'aracytine pendant 7 jours et d'une anthracycline pendant 3 jours (schéma "3+7"), ou sur une variation autour de ce schéma. Il permet d'obtenir une RC, c'est à dire la récupération d'une formule sanguine quasi-normale (plaquette $> 100 \cdot 10^9/L$, PNN $> 1 \cdot 10^9/L$) et d'une moelle normale (richesse normale et absence d'excès de blastes) dans 60-80% des cas environ. Des cures supplémentaires (consolidation de la RC) de chimiothérapie à base d'aracytine sont ensuite proposées ³⁹. Dans les LAM de mauvais pronostic, si un donneur HLA compatible est disponible, et selon l'état général du patient (âge, comorbidités...), une allogreffe de CSH de fin de traitement peut être proposée. Ce dernier traitement est grevé d'une lourde morbidité induite (infections et réaction du greffon contre l'hôte), ainsi que d'une mortalité liée à la procédure de 10 à 20 %, et doit par conséquent être réservé aux seules formes de mauvais pronostic. A l'heure actuelle, les critères pronostiques utilisés pour guider les traitements reposent principalement sur les caractéristiques initiales de la maladie (secondaire ou induite, cytogénétique, biologie moléculaire), et sur la réponse au traitement (obtention de la RC avec une seule cure d'induction, mesure de la maladie résiduelle) ³⁹. Les critères d'indication d'allogreffe de CSH doivent être améliorés afin de réserver cette procédure aux malades qui peuvent réellement en bénéficier.

Malgré l'amélioration de la prise en charge globale des patients, le pronostic des LAM reste très sévère, en particulier chez les sujets âgés. En effet, même si la survie globale s'est améliorée dans les années 2000, par rapport aux années 1980, celle-ci reste d'environ 19,5 % à 5 ans et d'environ 17 % à 10 ans, d'après les analyses du registre américain ³⁴ avec un pronostic toujours très défavorable chez les sujets âgés. L'analyse du registre suédois des cancers montre également une survie globale à 10 ans plus élevée dans les populations plus jeunes, aux environs de 47 % pour les patients de moins de 35 ans et qui diminue rapidement avec l'âge : aux environs de 30 % dans la tranche d'âge de 35 à 54 ans, de moins de 20 % pour la tranche 55-64 ans, et de moins de 5 % pour les patients de plus de 65 ans ⁴⁰.

Le résultat global des traitements laisse à penser qu'une meilleure compréhension des mécanismes de leucémogénèse est indispensable afin de développer de nouvelles stratégies thérapeutiques plus ciblées, et d'améliorer le pronostic global des malades. Une amélioration

des critères de sélection des patients à proposer pour une allogreffe de clôture est également nécessaire.

C- Leucémie aiguës myéloïdes : aspects biologiques et physiopathologie

C.1- Données classiques et classification des LAM

Les LAM sont des maladies très hétérogènes sur le plan biologique, au niveau cytologique, cytogénétique et moléculaire. Plusieurs classifications des LAM ont été proposées. D'abord purement descriptives, elles ont été améliorées par la meilleure connaissance de la physiopathologie et des caractéristiques cytogénétiques et moléculaires des maladies.

C.1.a- Aspects cytologiques

La première classification internationale des LAM ou classification franco-américaine et britannique (FAB) ⁴¹ était purement cytologique, puis a été précisée par la cytométrie en flux dans des cas particuliers. Selon le niveau où se situe le blocage de maturation, les cellules malades seront plus ou moins différenciées dans une lignée particulière, myéloïde (LAM 0 à LAM 3), myélo-monocytaire (LAM 4 et LAM 5), érythroïde (LAM 6) ou mégacaryocytaire (LAM 7). Cette classification est résumée dans le tableau 1.

Tableau 1 : Résumé de la classification FAB des LAM, d'après Bennet et al ⁴¹.

Type de LA	Caractéristiques
LA indifférenciée LAM 0	Absence de signe de différenciation des blastes Présence de marqueurs myéloïdes en immunophénotypage
LAM sans maturation LAM 1	Proportion de cellules granuleuses en cours de maturation inférieure à 10 % dans la moelle osseuse
LAM avec maturation granuleuse LAM 2	Proportion de cellules granuleuses en cours de maturation supérieure à 10 % dans la moelle osseuse Proportion de cellules monocytaires inférieure à 20 % dans la moelle osseuse
LAM promyélocytaire LAM 3	Présence de corps d'Auer en fagots Présence de promyélocytes anormaux
LAM myélo-monocytaire LAM 4	Proportion de cellules granuleuses en cours de maturation supérieure à 10 % dans la moelle osseuse Proportion de cellules monocytaires supérieure à 20 % dans la moelle osseuse
LAM monoblastique LAM 5	Proportion de cellules de lignée monocytaire dans la moelle supérieure à 80 %
Erythroleucémie LAM 6	Proportion de cellules érythroïdes médullaires supérieure à 50 % Proportion de blastes supérieure à 30 % des cellules non érythroïdes
LA à mégacaryocytes LAM 7	50 % des blastes au moins sont de lignée mégacaryocytaire (immunophénotypage)

C.1.b- Anomalies cytogénétiques

Dans les années 1970 et 1980, l'étude du caryotype des cellules blastiques a mis en évidence de nombreuses anomalies chromosomiques. Ces anomalies sont présentes dans 40 à 50 % des LAM environ, les autres LAM présentant un caryotype normal en technique standard (bandes G). La plupart de ces anomalies sont récurrentes, et associées à une valeur pronostique pour la réponse au traitement ⁴²⁻⁴⁵.

Les conséquences fonctionnelles associées à ces anomalies cytogénétiques ont été étudiées pour un certain nombre d'entre elles, mais restent toujours mal connues pour d'autres. Les principales anomalies récurrentes et leur retentissement moléculaire sont résumés dans le tableau 2.

Tableau 2 : Résumé des principales anomalies chromosomiques récurrentes dans les LAM, de leurs conséquences fonctionnelles et de leur valeur pronostique.

Anomalie	Anomalies moléculaires associées	Pronostic associé
t(8;21)(q22;q22)	Transcrit de fusion <i>RUNX1-RUNX1T1</i>	Favorable
inv(16) (p13.1q22) t(16;16) (p13.1q22)	Transcrit de fusion <i>CBFB-MYH11</i>	Favorable
t(15;17) (q22 ; q12)	Transcrit de fusion <i>PML-RARα</i>	Favorable
del(5q) et del(5)	Complexes	Défavorable
del(7q) et del(7)	Complexes	Défavorable
t(6 ;9)(p23 ;q34)	Transcrit de fusion <i>DEK-NUP214</i>	Défavorable
t(9;22)(q34 ;q11)	Transcrit de fusion <i>BCR-ABL</i>	Défavorable
inv(3)(q21q26.2)	Remaniement d' <i>EVII</i>	Défavorable
Translocations impliquant le locus <i>MLL</i> t(11,v)	Perte de fonction de <i>MLL (KMT2A)</i> Histone méthyltransférase	Défavorable ou intermédiaire
del(17p) et del(17)	Perte du locus <i>TP53</i> Anomalie de réparation de l'ADN	Défavorable
del(20q)	Complexes	Intermédiaire
Trisomie 8	Mal connue Augmentation du nombre de copies de <i>c-myc</i>	Intermédiaire

C.1.c- Biologie moléculaire : événements "classiques"

Dans les années 1990-2000, un certain nombre d'anomalies moléculaires récurrentes ont été identifiées dans les LAM, et en particulier en cas de caryotype normal. Leurs conséquences fonctionnelles sont variables. Le plus souvent, une valeur pronostique a été associée à ces mutations.

C.1.c.1- Mutations de la *fms-like* protéine kinase 3

La *fms-like* protéine kinase 3 (FLT3) est une protéine transmembranaire à activité tyrosine kinase ^{46,47} normalement exprimée au niveau des progéniteurs hématopoïétiques immatures. Il s'agit du récepteur au ligand de FLT3 ⁴⁸, pouvant être également exprimé dans les cellules leucémiques ⁴⁹. Son activation conduit à une cascade de phosphorylations intracytoplasmiques, aboutissant à une prolifération cellulaire accrue, et à une résistance à l'apoptose ⁵⁰.

De nombreuses études à partir de lignées cellulaires de LAM ont démontré que la présence de duplications en tandem du domaine juxta-membranaire (*FLT3-ITD*) conduit à une activation constitutive de FLT3 et des voies de prolifération sous-jacentes ⁵¹. Cette anomalie est retrouvée chez 20 à 30 % des patients atteints de LAM ⁵²⁻⁵⁴ et associée à un pronostic défavorable ^{39,52,54}.

D'autres mutations conduisant à une activation constitutive de FLT3 ont été décrites. On compte parmi elles les mutations du domaine tyrosine kinase (*FLT3-TKD*), ou d'autres domaines fonctionnels ^{55,56}.

C.1.c.2- Mutations de la nucléophosmine

La nucléophosmine (NPM1) est une protéine chaperon, de localisation préférentiellement nucléaire, capable de faire la navette entre le cytoplasme et le noyau des cellules. Elle joue un rôle dans la régulation de la transcription, et assure le transport et l'assemblage des particules pré-ribosomales dans le cytoplasme. NPM1 agit également comme contrôle de la voie de la protéine p53 ⁵⁷. Plusieurs types de mutations de *NPM1*, localisées au niveau de l'exon 12 ^{58,59}, ont été identifiés chez les patients atteints de LAM. Elles ont toutes comme conséquence l'accumulation intracytoplasmique de la protéine ⁵⁸, et la perte de sa fonction normale, par défaut de localisation nucléaire. Ces mutations ont également un impact sur les capacités de différenciation cellulaire ^{60,61} et les molécules induisant la différenciation (Acide tout trans rétinolique et trioxyde d'Arsenic) semblent avoir une action en cas de présence de ces mutations ⁶². Elles sont retrouvées dans environ 25 % de l'ensemble des LAM ⁵⁹, et dans 40 à 60 % des LAM à caryotype normal. Le pronostic associée à ces mutations est globalement favorable en l'absence de *FLT3-ITD* ^{39,52,58}.

C.1.c.3- Mutations de la « CAAT/enhancer binding protein alpha »

La CAAT/enhancer binding protein alpha (CEBP α) est un FT, qui joue un rôle essentiel dans la maturation finale de la lignée granulocytaire^{63,64} (figure 2). Sa délétion, dans des modèles murins, conduit à l'absence totale de granuleux matures⁶⁵. Des mutations dans un des trois domaines actifs de *CEBPA* (domaine leucine zipper, ou domaines trans-activateurs TAD1 et TAD2), conduisant à une perte d'activité, ont été décrites chez environ 10 % à 15 % des patients atteints de LAM^{66,67} et dans 8 à 20 % des LAM à cytogénétique normale^{39,52,68}. Elles ont été associées, particulièrement en cas de perte des 2 allèles normaux, à un pronostic favorable^{69,70}.

C.1.d- Classification 2008 de l'OMS

La classification des tumeurs hématopoïétiques établies en 2008 par l'OMS récapitule l'ensemble des données physiopathologiques et pronostiques connues lors de sa publication, au niveau cytologique, cytogénétique et moléculaire. Elle reste la classification de référence à l'heure actuelle³². Cette classification est résumée dans le tableau 3.

Tableau 3 : Résumée de la classification des LAM proposée par l'OMS en 2008.

LAM avec anomalies génétiques récurrentes	
LAM avec translocations ou inversions chromosomiques récurrentes	Translocation (8;21)(q22 ;q22) Inversion (16)(p13.1q22) ou translocation (16;16)(p13.1;q22) Translocation (15;17)(q22;q12) Translocation (9;11)(p22 ;q23) Translocation (6;9)(p23 ;q34) Inversion (3)(q21q26.2) ou translocation (3;3)(q214 ;q26.2) Translocation (1;22)(p13 ;q13)
LAM avec mutations de gènes	Entité provisoire : LAM avec mutation de <i>NPM1</i> Entité provisoire : LAM avec mutation de <i>CEBPA</i> La recherche de <i>FLT3-ITD</i> est recommandée
LAM secondaire aux myélodysplasies	
LAM induite par les traitements (chimiothérapie et radiothérapie)	
LAM sans autre spécificité	LAM avec différenciation minimale LAM sans maturation LAM avec maturation LA myélomonocytaire LA monoblastique et monocytaire LA érythroïde LA à différenciation mégacaryoblastique LA à différenciation basophile Panmyélose aiguë avec myélofibrose
Sarcome myéloïde	
Proliférations myéloïdes en rapport avec un syndrome de Down	Anomalie transitoire de la myélopoïèse LAM associée au syndrome de Down
Leucémie à cellules dendritiques plasmacytoïdes	

L'ensemble des données accumulées ces 10 dernières notamment grâce aux techniques de séquençage à haut débit, et la meilleure compréhension des mécanismes de la leucémogénèse, va aboutir prochainement à des changements dans cette classification.

C.2- Données moléculaires récentes

C.2.a- De nouvelles classes de mutations

Depuis une dizaine d'années, l'essor des techniques de séquençage à haut débit a permis, par l'étude du génome entier ou de l'exome des LAM, la découverte de nombreuses mutations impliquant divers gènes. De nombreuses études fonctionnelles complémentaires ont permis de préciser la fonction des différents gènes impliqués quand elle n'était pas connue, et ont permis d'identifier les rôles prépondérants des régulateurs de l'épigénétique et des composants de la machinerie d'épissage dans les LAM. L'étude exhaustive du génome de 200 cas de LAM *de novo* par le *Cancer Genome Atlas Research Network* propose de regrouper les mutations en 8 classes fonctionnelles d'événements, complétant le spectre génétique déjà décrit ⁷¹. Ces 8 classes de mutations sont résumées dans la figure 3. Toutes les fonctions cellulaires des différents gènes ne sont pas connues, et certains gènes peuvent en fait partager différentes fonctions.

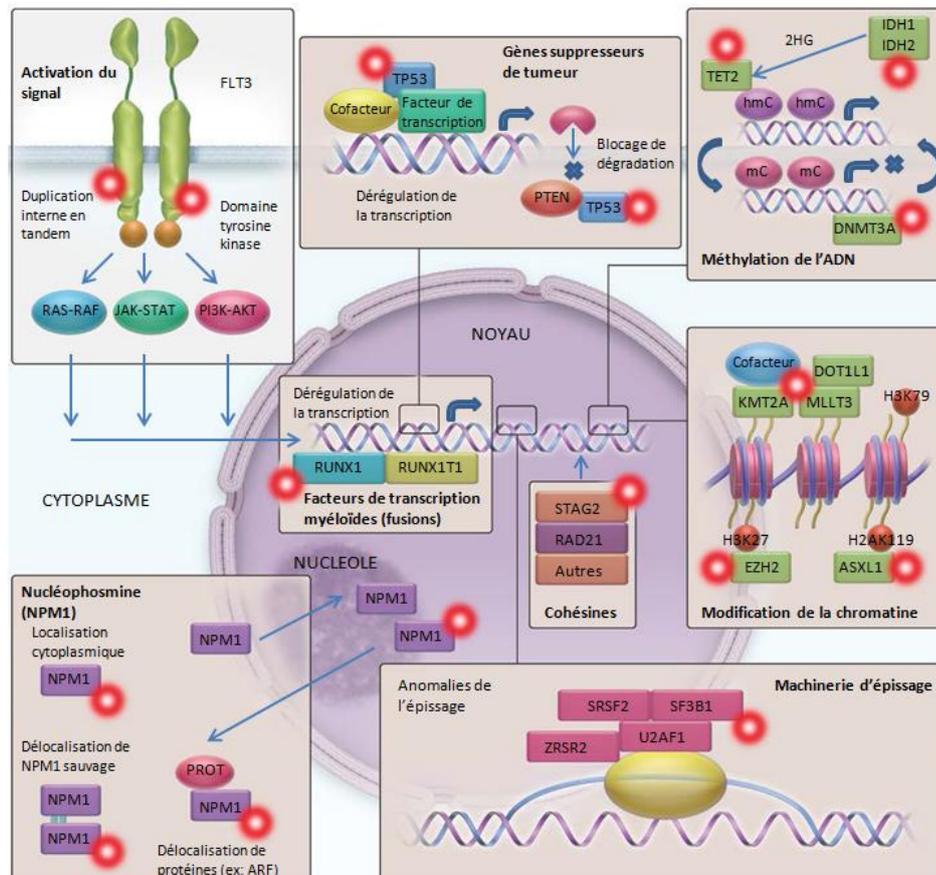


Figure 3: Les 8 classes de mutations fréquemment observées dans les LAM d'après les données du Cancer Genome Atlas. (Adapté d'après Döhner et al, NEJM, 2015 ³⁹)

C.2.a.1- Régulateurs de la méthylation de l'ADN

La dérégulation des mécanismes épigénétiques joue un rôle majeur dans la physiopathologie des LAM, en contrôlant de manière réversible le niveau d'expression de nombreux gènes impliqués dans l'hématopoïèse^{21,72}. La méthylation et les modifications post-méthylation des cytosines au niveau des îlots CpG de l'ADN est un des mécanismes principaux de cette régulation. Dans les LAM, les mutations des régulateurs de la méthylation, comprennent principalement les mutations de la DNA-méthyl-transférase 3a (*DNMT3A*) ou d'autres méthyl-transférases, de *TET2* (ou d'autres membres de la famille TET) et, par voie indirecte, des isocitrate-déshydrogénases (*IDH1* et *IDH2*). Le rôle de ces différents acteurs dans la régulation des méthyl-cytosines est résumé dans la figure 4.

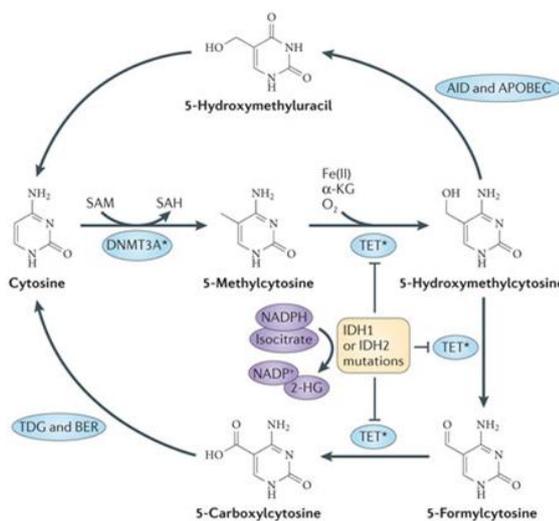


Figure 4 : Mécanismes de régulation de la méthylation de l'ADN et effet des différentes mutations retrouvées dans les LAM sur la méthylation. Adapté d'après Shih et al, Nature Review cancer, 2012⁷².

Mutations de *DNMT3A*

DNMT3A est une méthyl-transférase dont la fonction principale est d'ajouter un groupement méthyle aux cytosines des îlots CpG de l'ADN. Les mutations de *DNMT3A* ont été identifiées par étude du génome entier et sont présentes dans environ 20% des LAM. Elles peuvent affecter une grande partie du gène *DNMT3A*, même si les mutations du résidu R882 sont de loin les plus fréquentes ⁷³. Ces mutations entraînent une diminution des capacités catalytiques de l'enzyme et de ses capacités de liaison à l'ADN ^{74,75}. En cas de mutation, la quantité globale de 5-méthylcytosine (5mC) n'est pas modifiée, mais des profils spécifiques de méthylation et d'expression génique ⁷³ (notamment surexpression de *HOXA* et *HOXB*) ont été mis en évidence ^{76,77}. In vivo dans des modèles murins, la diminution d'expression de *DNMT3A* dans les CSH conduit à une expansion progressive de leur nombre ⁷⁸. Les mutations de *DNMT3A* sont à ce jour un des seuls événement dont le caractère pré-leucémique a été formellement démontré ⁷⁹ (voir paragraphe C.4).

Mutations de *TET2*

La fonction principale de *TET2* est de convertir les 5mC en 5-hydroxyméthylcytosine (5hmC) ^{80,81}. Les 5hmC vont bloquer les interactions entre l'ADN et certaines protéines de liaison à l'ADN méthylé et qui entraînent normalement une diminution d'expression génique ⁸². L'enrichissement en 5hmC dans des sites de régulation de la transcription et notamment sur des sites enhancers peut également être associé à une régulation de la transcription génique ⁸³⁻⁸⁵. L'accumulation de 5hmC peut conduire également à une déméthylation passive de l'ADN, car la méthyl-transférase *DNMT1* ne peut pas reconnaître les 5hmC ⁸⁶, ou à une déméthylation active par la voie AID-APOBEC ⁸⁷. Enfin les enzymes de la famille *TET* peuvent également convertir la 5hmC en 5-formyl-cytosine et en 5-carboxyl-cytosine, qui peuvent elles-mêmes être reconverties en cytosine déméthylées ⁸⁸. Les mutations de *TET2* ont été décrites dans environ 15% des pathologies myéloïdes ⁸⁹ et dans 7 à 20% des LAM. Elles entraînent une perte de l'activité catalytique normale de l'enzyme, ce qui aboutit globalement à une diminution de 5hmC, à une accumulation de 5mC et à une dérégulation du niveau de méthylation de l'ADN ⁹⁰. Dans des modèles murins, la déplétion en *Tet2* dans les CSH conduit à une augmentation des capacités de repopulation hématopoïétiques post greffe, à une augmentation du nombre de CSH et à un biais de différenciation vers la lignée myélo-monocytaire ⁹¹⁻⁹⁵.

Mutations d'*IDH1* et d'*IDH2*

Les isocitrate déshydrogénases (IDH) sont des enzymes impliquées dans le métabolisme énergétique (cycle de Krebs), qui convertissent l'isocitrate en α -ceto-glutarate (α CG). Des mutations dans *IDH1* et *IDH2* ont été identifiées dans plusieurs types de cancers dont les LAM^{96,97} par séquençage du génome entier, avec principalement des mutations des résidus R132 d'*IDH1*, et des résidus R140 ou R172 d'*IDH2*. Ces mutations conduisent à une modification de l'activité catabolique normale des IDH, et induisent la production néomorphique de 2-hydroxyglutarate de type D (2HG)⁹⁸. La déplétion en α CG, qui est un cofacteur indispensable à l'activité des protéines TET, va entraîner des altérations épigénétiques en partie comparables à celles des mutations de *TET2*, et la production de 2HG des anomalies de méthylation des histones, aboutissant à un blocage de différenciation cellulaire^{99,100}. Ces mutations pourraient également jouer sur d'autres voies enzymatiques et métaboliques impliquées dans la leucémogénèse.

C.2.a.2- Modificateurs de la chromatine

Le deuxième grand mécanisme de régulation de l'épigénétique implique les modifications de structure de la chromatine et les modifications des histones. De nombreuses protéines impliquées dans ces modifications peuvent être mutées dans les LAM, et changer la structure de la chromatine, avec pour conséquence des profils particuliers d'expression génique. Dans cette classe, on peut citer notamment les membres du complexe Polycomb PRC2 (notamment *EZH2*), *ASXL1*, *MLL* ou ses partenaires.

Mutations d'*EZH2* et des complexes polycombs

Les complexes polycombs sont des répresseurs transcriptionnels jouant un rôle fondamental dans la différenciation cellulaire. Ils sont organisés en 2 complexes PRC1 et PRC2¹⁰¹. Le complexe PRC2 est constitué des protéines *EZH1* ou *EZH2*, *EED*, *SUZ12* et *RBAP4*. Des mutations de ces différents composants ont été décrites dans les LAM, mais les mutations d'*EZH2* sont largement majoritaires, mêmes si elles restent rares. *EZH2* est une H3K27 méthyl-transférase. Les mutations d'*EZH2* peuvent entraîner un gain de fonction enzymatique observé dans les pathologies lymphoïdes, ou une perte de fonction enzymatique observée dans les pathologies myéloïdes, et changent l'état de méthylation d'H3K27, avec des conséquences précises au niveau cellulaire encore mal connues¹⁰²⁻¹⁰⁴. La diminution isolée

de l'expression d'*EZH2* n'entraîne pas d'altération des CSH ¹⁰⁵, mais des anomalies des lymphocytes B, tandis que la diminution d'expression d'autres membres du complexe PRC2 comme *EED* entraîne une augmentation des capacités d'auto-renouvellement des CSH ¹⁰⁶. La surexpression d'*Ezh2* dans des précurseurs hématopoïétiques peut induire un phénotype de syndrome myéloprolifératif dans des modèles murins ¹⁰⁷. Les mutations du complexe PRC1 n'ont pas été rapportées à ce jour dans les LAM.

Mutations d'*ASXL1*

Les mutations d'*ASXL1* se produisent dans 5 à 17% des LAM ^{39,71}. La fonction précise d'*ASXL1* dans l'hématopoïèse est mal connue. *ASXL1* est capable d'interactions directes avec le complexe PRC2 et est impliqué dans la régulation de la tri-méthylation d'H3K27. Les profils d'expression génique en cas de d'expression d'*ASXL1* mutant dans des cellules hématopoïétiques partagent des caractéristiques communes avec les profils d'expression observés en cas de remaniement de *MLL*, avec notamment des modifications d'expression des gènes *HOX*. Dans des modèles murins, la co-transfection de mutants d'*Asx1l* et de *Nras* dans des cellules hématopoïétiques conduit à un phénotype de syndrome myéloprolifératif d'apparition plus précoce qu'en cas de transfection isolée de *Nras* muté ¹⁰⁸. Après expression d'un transcrite tronqué d'*ASXL1* et en coopération avec *BAP1*, les capacités de prolifération cellulaire in vitro de cellules hématopoïétiques augmentent, avec un biais vers la différenciation myéloïde, et des capacités accrue de formation de CFU-GM en culture ¹⁰⁹. La perte d'expression d'*Asx1l* dans les CSH dans des modèles murins entraîne un phénotype mimant la myélodysplasie, et aboutit à une diminution des capacités d'auto-renouellent des CSH. L'association avec la perte de fonction de *Tet2* restaure les capacités d'auto-renouvellement des CSH, et entraîne également un phénotype de myélodysplasie ¹¹⁰.

Mutations et remaniement de *MLL*

Les anomalies impliquant *MLL* sont très fréquentes dans les LAM ³⁹. Il peut s'agir de translocations chromosomiques impliquant *MLL* et différents partenaires (10-15% des LAM) ou de duplications partielles en tandem (environ 5% des LAM). *MLL* comprend un domaine avec une activité H3K4 histone méthyl-transférase, et peut interagir également avec de très nombreux autres partenaires ¹¹¹. Les translocations impliquant *MLL* entraînent la perte de ce domaine fonctionnel. Les produits de fusion impliquant *MLL* vont de plus être capables de recruter la protéine *DOTL1*, qui comprend des fonctions de méthylation d'H3K79, ce qui

aboutit également à des modifications de structure de la chromatine. Les profils d'expression génique secondaire au remaniement de *MLL* comprennent notamment une dérégulation des gènes *HOX*^{112,113}. Au niveau fonctionnel, l'acquisition de transcrit de fusion impliquant *MLL* dans des progéniteurs hématopoïétiques myéloïdes différenciés entraîne une ré-acquisition de capacité d'auto-renouvellement. L'expression de transcrits impliquant *MLL* dans des progéniteurs CD34⁺ humains dérivés de sang de cordon ombilical va aboutir, après xéno greffe à des souris immunodéprimées, à une hématopoïèse multi-lignées, avec une augmentation des capacités de prolifération par rapport aux CD34⁺ normaux¹¹⁴.

C.2.a.3- Anomalies des facteurs de transcription myéloïdes et fusions les impliquant

De nombreux FT impliqués dans la différenciation myéloïde ou dans la répression de la différenciation vers les autres lignées peuvent être mutés ou remaniés dans les LAM. Les remaniements principaux impliquent le *core-binding factor*, qui comprend les translocations impliquant *RUNX1* et *CBFB*. Beaucoup d'autres FT peuvent potentiellement être mutés, et aboutir à un blocage de la maturation des cellules leucémiques par perte de fonction.

Translocations avec anomalies du core binding factor

Le core-binding factor (CBF) est un FT hétérodimérique constitué de 2 sous-unités *RUNX1* et *CBFB*. A l'état normal, *RUNX1* est capable de liaison à l'ADN. *CBFB* potentialise cette capacité de liaison. Ce complexe de transcription est nécessaire à l'établissement de l'hématopoïèse pendant la vie foetale. La perte isolée d'une des 2 sous unités entraîne une létalité in utero par défaut d'hématopoïèse définitive¹¹⁵⁻¹¹⁸.

La translocation t(8;21)(q22;q22), souvent observée dans les LAM2, entraîne une fusion des gènes *RUNX1* et *RUNX1T1*. L'inversion péricentrique du chromosome 16, inv(16)(p13.1q22), associée aux LAM4 avec anomalies des éosinophiles, entraîne la fusion des gènes *CBFB* et *MYH11*. Ces deux événements vont entraîner des anomalies de fonctionnement du complexe CBF, ainsi que de multiples conséquences au niveau cellulaire¹¹⁹.

Les deux transcrits entraînent d'importantes modifications au niveau épigénétique. En effet, des profils particuliers de méthylation de l'ADN ont été observés en cas de LAM à CBF¹²⁰⁻¹²². *RUNX1-RUNX1T1* est capable de recruter des histone-déacétylases ainsi que *DNMT1*, et d'entraîner la répression transcriptionnelle de plusieurs cibles impliquées dans

l'hématopoïèse ^{123,124}. *MYH11* est également capable d'interaction avec des histone-déacétylases, et peut entraîner une répression transcriptionnelle ^{125,126}.

En dehors de la dérégulation épigénétique, les 2 transcrits peuvent réguler l'expression de nombreux gènes par interaction directe ou indirecte avec leurs cibles. Ainsi, ils vont entraîner une diminution d'expression de FT essentiels de la granulopoïèse comme *PU.1* ¹²⁷, *CEBPA* ¹²⁸, *GATA1* ¹²⁹, et d'une partie des protéines E (E2A, E2-2, HEB). L'expression de plusieurs gènes suppresseurs de tumeur va également être dérégulée par ces transcrits de fusion : *RUNX1-RUNX1T1* réprime la transcription de *p14* (ARF) ¹³⁰ et de *NF1* ¹³¹, et *CBFB-MYH11* celle de *p15* (INK45B) ¹³². La transcription de *BCL2*, anti-apoptotique majeur, est par ailleurs activée par *RUNX1-RUNX1T1* ¹³³. De plus, de nombreux gènes impliqués dans des voies de signalisation sont surexprimés dans les LAM CBF, conduisant à l'activation aberrante du signal, notamment la voie NOTCH (surexpression de *JAG1*) ¹³⁴, la voie *Wnt* ¹³⁵ ou la voie *NF-KB* ¹³⁶. Enfin ces transcrits semblent déréguler l'expression de miRNA impliqués dans l'hématopoïèse (miRNA-223 ou autres) ^{122,137}.

Mutations d'autres facteurs de transcriptions

Des mutations de nombreux autres facteurs de transcription impliqués dans l'hématopoïèse ont été décrites dans les LAM. Les conséquences fonctionnelles de ces mutations ne sont pas toujours connues, mais sont globalement associées à un blocage de maturation dans les cellules mutées. Parmi ces mutations, les plus fréquentes concernent *RUNX1* ^{138,139} et *CEBPA* (cf paragraphe C.1.c.3). Des mutations de *GATA2* ou d'autres facteurs de transcription ont également été décrites ⁷¹.

C.2.a.4- Mutations de *NPM1*

Voir paragraphe C.1.c.2.

C.2.a.5- Mutations de la machinerie d'épissage

Les mutations de la machinerie de l'épissage des ARNm sont fréquentes dans les pathologies myéloïdes, notamment dans les myélodysplasies ¹⁴⁰. Elles sont aussi retrouvées dans environ 14% des LAM. Ces mutations se produisent principalement dans les gènes *U2AF1*, *SRSF2*, et *SF3B1*. Les conséquences fonctionnelles précises de ces mutations sont mal connues, mais il a été observé des défauts d'épissages des transcrits de *TET2* et de

RUNX1 chez certains patients, pouvant laisser supposer une altération de fonction de ces protéines en cas de mutation ^{71,141}. Un défaut d'épissage d'autres cibles impliquées dans la régulation de l'hématopoïèse et la leucémogénèse a été observé, et notamment d' *EZH2* ^{142,143}.

C.2.a.6- Mutations activatrices du signal

Mutations activatrices de tyrosines kinases

En plus des mutations de *FLT3* (voir paragraphe C.1.c.1), des mutations activatrices d'autres tyrosine-kinases ont également été décrites, avec les mêmes conséquences fonctionnelles en aval. Parmi elles, les mutations de *KIT*, récepteur au SCF exprimé dans les cellules hématopoïétiques immatures, sont fréquemment retrouvées dans certains types de LAM (LAM avec remaniement du CBF), et sont associées à une plus grande capacité de prolifération cellulaire ¹⁴⁴. Des mutations d'autre tyrosine ou sérine-thréonine kinases ont été décrites, qui aboutiraient à l'activation constitutive du signal sous-jacent, et à une plus grande prolifération cellulaire ⁷¹.

Mutations impliquant la voie RAS

Les protéines de la famille RAS ont une activité GTPase intrinsèque, et participent au contrôle de la prolifération cellulaire, en activant la cascade des MAP/kinases et la voie PI3K-AKT. Les protéines RAS sont activées, entre autres, par les cascades de signalisation de tyrosine kinase comme *FLT3* et *KIT*. Dans les LAM, des mutations aboutissant à une activation constitutive du signal ont été observées dans les gènes *NRAS* (10 à 20 % des LAM) et *KRAS* ^{71,145}. De même, la perte de régulation de la voie RAS (mutations de *NF1* ou *PTPN11*), peut aboutir aux mêmes conséquences d'activation du signal sous-jacent ¹⁴⁶⁻¹⁴⁸.

C.2.a.7- Mutations du complexe de la cohésine

Le complexe de la cohésine est indispensable à la ségrégation chromosomique durant la méiose et la mitose. Ce complexe est indispensable pour la division cellulaire. Il est constitué par les protéines SMC1A, SMC3, STAG1, STAG3, et RAD21. Dans les cancers en général, les mutations de ce complexe sont associées à une instabilité chromosomique. Ces mutations sont retrouvées dans environ 6 % des LAM, principalement en cas de caryotype normal ¹⁴⁹. Le KO des gènes de la cohésine dans des modèles d'hématopoïèse murine entraîne

une augmentation des capacités de repiquage en série, ainsi qu'un phénotype de syndrome myéloprolifératif in vivo ¹⁵⁰.

C.2.a.8- Mutations de gènes suppresseurs de tumeur

Mutations de *TP53*

L'importance de la protéine p53 dans la carcinogenèse est bien connue. P53 est inactivée dans plus de la moitié des cancers chez l'homme, et la perte germinale d'une des copies de *TP53* (Syndrome de Li-Fraumeni) prédispose à de multiples cancers. *TP53* est mutée dans 5 à 10 % des LAM. Ces mutations sont associées aux LAM induites et à une cytogénétique complexe ^{71,151}. De nombreuses voies métaboliques sont contrôlées par p53. Son activation se produit en réponse à divers stress cellulaires (dommages dans l'ADN, hypoxie, stress oxydatif, déplétion en ribonucléotides, etc...). Elle peut induire un arrêt du cycle cellulaire, l'engagement vers la sénescence et l'apoptose, mais également le recrutement de protéines de réparation de l'ADN et le contrôle de l'autophagie. P53 est également impliquée dans la régulation de l'auto-renouvellement, et dans le maintien des cellules souches ^{152,153}. P53 agit en partie comme un facteur de transcription. La plupart des mutations de *TP53* se produisent dans les domaines de liaison à l'ADN, et induisent une perte de fonction de la protéine, aboutissant à une accumulation d'événements génétiques dans les cellules par absence de déclenchement de signal d'apoptose secondairement à ces lésions.

Autres suppresseurs de tumeur

D'autres gènes suppresseurs de tumeurs peuvent être mutés dans les LAM. Les fonctions de ces gènes peuvent être multiples et peuvent se rattacher aux catégories précédemment décrites. Il peut s'agir de facteurs de transcription comme *WT1* ¹⁵⁴, qui est également impliqué dans la régulation épigénétique ^{155,156}, ou *PHF6* ¹⁵⁷, de phosphatases comme *PTEN* ¹⁵⁸ (également impliquée dans la régulation de p53) ou d'autres protéines impliquées dans la voie p53.

C.2.b- Associations mutationnelles et ontogénèse des LAM

Les techniques récentes de séquençage, en plus de décrire de multiples mutations dans les LAM, ont permis d'étudier les associations des différents événements génétiques entre eux. Ainsi, dans les LAM *de novo*, des associations récurrentes, ou au contraire des exclusions quasi systématiques d'événements ont été rapportées^{39,71}. Ceci suggère une ontogénie particulière de ces LAM, car la probabilité que ces associations récurrentes soient liées au hasard est extrêmement faible. Par exemple, dans de nombreux cas, on retrouve une association entre mutation de *FLT3*, de *NPM1* et de *DNMT3A*, ou une association entre les mutations de *KIT* et la présence de transcrit *MYH11-CBFB* ou *RUNX1-RUNX1T1*, ou entre les mutations de *TP53* et la présence d'un caryotype complexe. Au contraire, il existe une quasi-exclusion entre les transcrits de fusion impliquant *MLL*, *RUNX1-RUNX1T1* et *MYH11-CBFB* et les mutations de *DNMT3A* ou de *NPM1*, ou entre les mutations de *RUNX1* et celles de *TP53*, de *FLT3* et de *NPM1*. Ces profils d'association ou d'exclusion de mutation suggèrent des voies différentes d'ontogénèse et des sous-catégories particulières de LAM.

Dans une étude comparative entre 93 LAM secondaires et 180 LAM *de novo*, il a été démontré que les mutations de la machinerie d'épissage (*SRSF2*, *SF3B1*, *U2AF1*, *ZRSR2*), ou d'*ASXL1*, *EZH2*, *BCOR* et *STAG2* se produisent quasi exclusivement dans les LAM secondaires. De même, les mutations de *NPM1*, les translocations impliquant le CBF et les réarrangements de *MLL* se produisent surtout dans les LAM *de novo*. Les LAM avec mutations de *TP53* (*de novo* ou secondaires) semblent former une troisième entité, comprenant peu d'autres mutations, mais de nombreux événements chromosomiques. Ces résultats suggèrent donc 3 ontogénies distinctes dans les LAM. L'étude complémentaire de 101 cas de LAM induite retrouve également ces 3 ontogénies, suggérant que les LAM induites ne forment pas vraiment une entité séparée, mais qu'elles peuvent se rattacher aux 3 entités précédemment décrites d'un point de vue génétique, avec une émergence des clones leucémiques favorisée par l'exposition aux cytotoxiques ou aux radiations ionisantes¹⁵¹.

Au bilan, de très nombreuses anomalies chromosomiques et génétiques ont été décrites ces dernières années dans les LAM. Leurs conséquences fonctionnelles sont variables et ont été regroupées en 8 catégories. Des profils d'association particuliers entre ces mutations suggèrent l'existence de plusieurs voies d'ontogénie différentes.

C.3- Phylogénie et dynamique clonale dans les LAM

L'identification de ces nombreux types de mutations a conduit à explorer l'architecture clonale globale des LAM. L'étude des variations des mutations entre diagnostic et rechute a permis de comprendre partiellement les mécanismes conduisant à la rechute, et de démontrer dans un certain nombre de cas, l'existence d'un clone fondateur des LAM, dont la persistance pourrait être à l'origine des rechutes. Les études génomiques séquentielles dans le temps comparant chez un même sujet des échantillons de myélodysplasie et de LAM secondaire, ou des échantillons en l'absence de pathologie hématologique et au diagnostic de LAM ont permis d'identifier en partie la chronologie d'apparition de certaines mutations, et à introduire la notion d'hématopoïèse clonale pré-leucémique.

C.3.a- Phylogénie clonale des LAM et étude des rechutes

La phylogénie clonale des LAM de novo et l'évolution clonale à la rechute ont été partiellement décrites après séquençage du génome entier chez 7 patients à caryotype normal et une LA promyélocytaire¹⁵⁹. L'étude des fréquences de variants (VAF) des différentes mutations a permis de mettre en évidence la co-existence au diagnostic de LAM, de multiples sous-clones, dérivant tous d'un même clone fondateur et porteurs d'anomalies communes (figure 5). Ce clone fondateur peut résister au traitement, et être à l'origine de la rechute : soit de nouvelles mutations se produisent après traitement dans le clone fondateur, ce qui conduit à l'émergence d'un nouveau sous-clone dominant en rechute, soit un des sous-clones présent au diagnostic est sélectionné pendant le traitement, et devient majoritaire à la rechute. Dans les 8 cas décrits, le clone fondateur restait détectable en rechute, démontrant sa résistance aux traitements. Dans cette première étude de grande ampleur, les anomalies génétiques détectées en rechute pouvaient appartenir à toutes les classes de mutation.

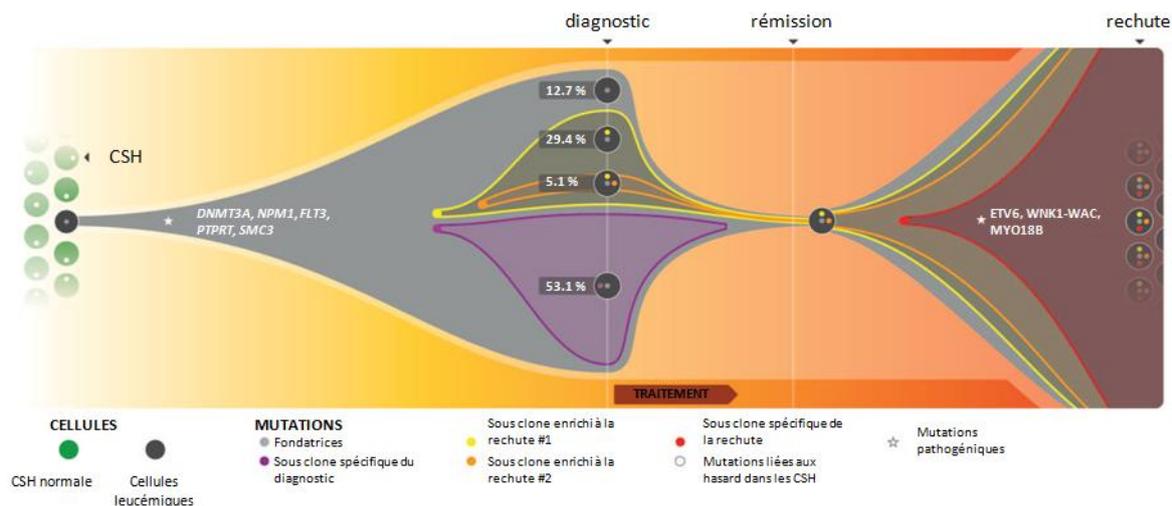


Figure 5 : Exemple d'architecture clonale et d'évolution clonale à la rechute chez un patient atteint de LAM. Différents sous clones dérivant d'un même clone fondateur coexistent au diagnostic de LAM. La persistance de certains de ces sous-clones après traitement est à l'origine de la rechute, après accumulation de nouvelles mutations. (Adapté d'après Ding et al, nature, 2012)

L'étude séquentielle au diagnostic et en rechute de 53 patients porteurs de LAM avec mutation de *NPM1*¹⁶⁰ a montré la perte de cette dernière mutation en rechute dans environ 10 % des cas. Entre les deux temps, une grande variabilité dans les mutations activatrices du signal et dans les mutations d'*IDH1* (gains ou perte), ainsi que le gain possible d'anomalies cytogénétiques non récurrentes ont également été observées. Seules les mutations de *DNMT3A* étaient quasiment toujours stables entre diagnostic et rechute chez ces patients, suggérant que ces mutations sont les plus précoces dans l'histoire clonale tandis que les mutations d'*IDH1* et celles associées à une activation du signal sont plus tardives. L'étude de 13 patients porteurs de LAM avec *FLT3-ITD* au diagnostic et en rechute montre également la grande stabilité des mutations de *DNMT3A*, et leur persistance en RC, suggérant également leur précocité dans la phylogénie¹⁶¹.

Enfin, l'étude couplée au diagnostic, en RC et en rechute de 39 patients dont 11 malades réfractaires aux traitements, tout caryotype confondu, a montré la persistance de certaines anomalies cytogénétiques entre diagnostic et rechute : t(9 ;11)(p22 ;q23) ; t(11 ;19)(q23 ;p13.1) ; inv(9)(p11q13) ; inv(16) ; t(8 ;21) ; +8 ; del(12p12q13). A l'inverse, d'autres anomalies non récurrentes pouvaient varier entre les deux temps, avec souvent un

gain d'anomalie lors de la rechute, au sein d'un clone dérivé du clone présent au diagnostic. Dans cette même étude, des variations de mutations ont été retrouvées entre les 2 temps, avec notamment des pertes du clone porteur de *FLT3-ITD*, tandis que les autres anomalies, et notamment celles impliquant les régulateurs de l'épigénétique, étaient toujours conservées ¹⁶².

C.3.b- Phylogénie clonale des LAM secondaires

La phylogénie clonale des LAM secondaire a été partiellement étudiée, en comparant l'évolution du génome issu de 7 patients atteints de LAM secondaire, au diagnostic de la LAM et pendant le suivi de la myélodysplasie ¹⁶³. De même que pour les LAM de novo, plusieurs sous-clones dérivant d'un même clone initiateur peuvent exister dans les myélodysplasies. L'apparition d'au moins une nouvelle mutation (notamment de *CDH23*, *PTPN11*, *RUNX1* ou *WT1*) dans l'un de ces sous-clones peut conduire à la transformation en LAM. Là encore, le clone fondateur de la myélodysplasie est toujours retrouvé au diagnostic de LAM. (figure 6)

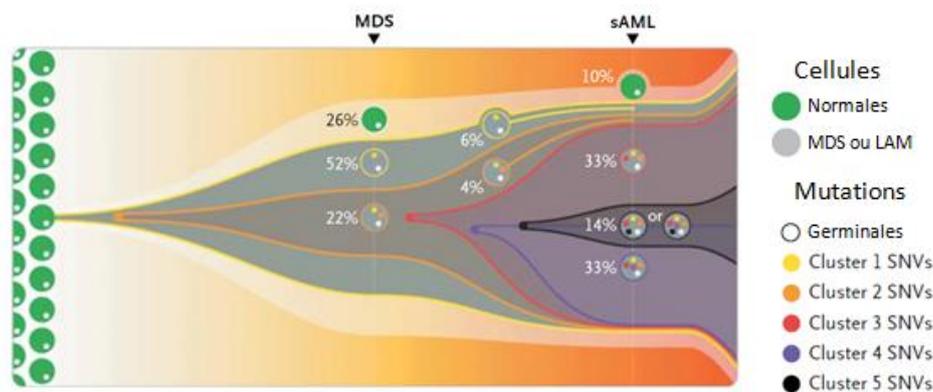


Figure 6 : Exemple d'évolution clonale chez un patient atteint de LAM secondaire à un syndrome myélodysplasique. L'accumulation de nouvelles mutations au sein d'un des sous-clones du syndrome myélodysplasique aboutit à la transformation en LAM secondaire. (adapté d'après Walter et al, NEJM, 2012)

Dans une autre étude séquentielle sur 38 patients entre les stades myélodysplasie et LAM secondaire, il a été démontré globalement une persistance des anomalies détectées au stade myélodysplasie, et l'apparition sous-clonale d'anomalies chromosomiques (trisomie 8 et délétion du 5), ou de mutations activatrices du signal (mutations de *FLT3* ou de la voie *RAS*), lors de la transformation ¹⁶⁴.

Enfin, dans une série de 18 malades, l'acquisition d'au moins une mutation de *FLT3* ou d'une mutation impliquant la voie RAS était associée à la transformation leucémique ¹⁶⁵. À noter dans ces études, une grande stabilité des mutations entre les 2 stades en cas de mutations de *TP53*, suggérant une ontogénie différente dans ces cas particuliers.

C.3.c- Hématopoïèse clonale pré-leucémique

Dans 3 études séparées, incluant chacune 17182, 12380, et 2728 individus, un séquençage de l'exome a été réalisé sur le sang ¹⁶⁶⁻¹⁶⁸. Les individus n'étaient pas sélectionnés par rapport aux antécédents cancéreux dans 2 de ces études, ni à la présence de maladie hématologique. Chez un certain nombre d'entre eux, il a été identifié des mutations de gènes impliqués dans la leucémogénèse, en dehors de toute anomalie morphologique sanguine, ou de toute pathologie hématologique déclarée. Cette hématopoïèse clonale est plus fréquente avec le vieillissement ¹⁶⁶⁻¹⁶⁸. Elle touche environ 1% des sujets de moins de 50 ans, mais plus de 10 % des sujets de plus de 60 ans et environ 20 % des sujets de plus de 90 ans. La présence de cette hématopoïèse clonale liée à l'âge (HCLA) est un facteur de risque de développer un "cancer hématologique" (LA, lymphome, myélome). Ces derniers peuvent se produire plusieurs années après l'apparition de l'HCLA. Dans l'une des 3 séries, 42 % des patients développant une pathologie maligne hématologique lors du suivi avaient une histoire d'HCLA. Dans les quelques cas étudiés, les pathologies myéloïdes secondaires étaient porteuses des anomalies de l'HCLA, suggérant que les mutations retrouvées initialement sont des mutations fondatrices du clone leucémique. L'accumulation secondaire d'événements génétiques finit par aboutir à une LAM, selon le même modèle que la transformation d'une myélodysplasie en LAM (voir C.3.b). Cette accumulation peut prendre plusieurs années, ou plusieurs dizaines d'années, et peut être favorisée par un défaut dans les mécanismes de réparation de l'ADN ¹⁶⁹. Les principaux gènes impliqués dans l'HCLA sont *DNMT3A*, *ASXL1* et *TET2*, et, dans une moindre mesure, *TP53*, *PPM1D*, *JAK2*, *SRSF2*, *SF3B1*, *U2AF1*, *ATM*, *CBL* et *IDH2*. Par ailleurs, la présence de mutations de *TET2* dans le cadre de cette HCLA s'accompagne d'un excès de 5mC et d'une baisse de 5hmC, ainsi que d'une altération du profil de méthylation par rapport aux sujets non porteurs de ces anomalies ¹⁷⁰. Cela suggère que ces mutations, en influant sur la transcription, peuvent entraîner des conditions favorisant la dominance du clone muté et l'acquisition secondaire d'autres mutations.

Suivant le même modèle, il a été démontré que les mutations de *TP53*, qui sont plus fréquentes dans les LAM induites, existent déjà dans un sous-clone minime au diagnostic du

premier cancer, suggérant une mutation ancienne dans une CSH de phénotype normal ¹⁵³. Le traitement par chimiothérapie donne un avantage au clone muté pour *TP53*, comme démontré dans un modèle murin après exposition aux nitrosurés, à cause de la meilleure résistance des CSH mutées pour *TP53* aux traitements (défaut d'apoptose après accumulations de lésions génétiques). L'accumulation secondaire d'anomalies cytogénétiques ou éventuellement de mutations dans d'autres gènes va aboutir à la LAM induite.

La mise en évidence de cette HCLA a conduit à proposer une nouvelle classification des maladies myéloïdes, avec l'introduction de la notion d'hématopoïèse clonale de potentiel indéterminé ¹⁷¹. Cette entité est caractérisée par la présence isolée de mutations récurrentes, en l'absence d'autre anomalie. L'évolution vers des pathologies myéloïdes franche est non systématique, mais plus importante que dans la population générale, avec un taux de transformation d'environ 0.5 à 1 % par an. De même, dans un certains nombre de cas de cytopénies isolées, l'absence d'anomalie morphologique ou cytogénétique ne permettait pas de poser de diagnostic clair. La présence de mutations récurrente affirme le caractère clonal de l'hématopoïèse, et peut donc préciser la définition des cytopénies clonales de signification indéterminée, également associées à un potentiel évolutif vers une maladie myéloïde. Ces deux entités peuvent globalement se comparer dans leur concept aux pics monoclonaux de signification indéterminée (MGUS), qui risquent d'évoluer vers un myélome multiple, et nécessitent probablement une surveillance plus ou moins rapprochée des patients.

Par ailleurs, l'existence d'un mosaïcisme chromosomique clonal chez des sujets sains, avec présence de délétions ou de gains de matériel dans des zones récurrentes plus ou moins larges a également été décrit ^{172,173}. Ce mosaïcisme est plus fréquent chez les sujets plus âgés, et est associé au risque de cancer et de LA. De manière intéressante, parmi les zones remaniées, on retrouve la délétion du chromosome 20q, et des délétions des régions incluant *TET2* et *DNMT3A*. Ce mécanisme pourrait aussi contribuer à la leucémogénèse précoce.

Au bilan, l'étude de la phylogénie clonale des LAM révèle que de multiples sous-clones dérivant d'un même clone initiateur peuvent être observés au diagnostic. La persistance d'un des sous-clones, ou l'accumulation de nouvelles mutations au sein du clone fondateur pourrait conduire aux rechutes. Les événements épigénétiques comme les mutations de *TET2* ou *DNMT3A* semblent être précoces : ils sont fréquemment retrouvés au sein du clone fondateur, et peuvent également se produire dans des CSH normales et donner lieu à une HCLA.

C.4- Événements précoces et marqueurs de cellules souches pré-leucémiques

Secondairement à l'identification de la phylogénie clonale des LAM, et à la mise en évidence de l'HCLA, plusieurs équipes ont cherché à identifier les anomalies les plus précoces au cours de la leucémogénèse, afin de mieux en comprendre les mécanismes précis.

L'étude la plus complète a été faite sur les mutations de *DNMT3A*⁷⁹. Chez des patients au diagnostic de LAM avec mutations de *DNMT3A* (*DNMT3Am*) et mutations de *NPM1* (*NPM1m*), les *DNMT3Am* ont été retrouvées à la fois dans les lymphocytes T et les cellules CD33+, alors que les *NPM1m* ont principalement été retrouvées dans les populations myéloïdes. Cela suggère que les *DNMT3Am* se produisent dans un progéniteur très précoce, ancêtre commun des lignées lymphoïdes et myéloïdes, tandis que les *NPM1m* se produisent dans un précurseur myéloïde plus tardif. Les *DNMT3Am* ont été retrouvées dans la plupart des sous-populations hématologiques, y compris les populations lymphocytaires, et les populations les plus immatures (CSH), tandis que les *NPM1m* n'ont été retrouvées que dans des populations plus engagées dans la différenciation myéloïde (progéniteurs communs granulo-monocytaires, et erythro-mégacaryocytaires). Ceci suggère l'existence de CSH pré-leucémiques, porteuses de *DNMT3Am*, et capables de produire une hématopoïèse mutée multi-lignée. Chez 4 patients double-mutés, l'analyse en RC de cellules CD33+ triées retrouvait seulement les *DNMT3Am*. L'analyse séquentielle de plusieurs échantillons d'un même patient en RC retrouve une ré-augmentation progressive de la proportion de CD33+ avec *DNMT3Am*. Alors que les *NPM1m* n'étaient pas retrouvées en RC précoce, elles étaient de nouveau détectables en rémission tardive, suggérant soit la ré-ascension du clone initial, soit une nouvelle mutation de *NPM1* dans le clone résiduel avec *DNMT3Am*. En rechute, chez 2 des patients, on retrouvait également les *DNMT3Am* dans toutes les sous-populations cellulaires. Les *NPM1m* étaient absente des populations les plus immatures (CSH) en rémission comme en rechute. Enfin, l'étude de la repopulation hématopoïétique en dilution limite, après greffe dans des souris immunodéprimées, d'échantillons de LAM avec *DNMT3Am* montrait des capacités de repopulation leucémique ou non leucémique (lymphoïdes ou myéloïdes) de ces échantillons. Les *DNMT3Am* ont été retrouvées dans une grande proportion des prises non leucémiques multi-lignées, tandis que les *NPM1m* étaient retrouvée seulement en cas de prise myéloïde. Au bilan, ce travail démontre l'existence de CSH pré-leucémiques porteuses de *DNMT3Am*, capables de générer une hématopoïèse clonale multi-lignée, et dans laquelle des mutations peuvent se produire plus tardivement dans

la différenciation pour générer une maladie leucémique vraie. Dans ce même travail, les mutations d'*IDH2* paraissent plus précoces que les mutations de *NPM1*, et pourraient être retrouvées dans des CSH pré-leucémiques, sans que la démonstration formelle n'en soit faite.

Une autre équipe a d'abord séparé, chez 6 patients atteints de LAM avec mutation de *FLT3-ITD*, les CSH leucémiques des CSH résiduelles non leucémiques, en utilisant une stratégie de tri en cytométrie basée sur l'expression de CD47, de CD99 et de TIM3. Après avoir vérifié fonctionnellement que les CSH résiduelles non leucémiques permettaient une repopulation multi-lignées, mais pas de repopulation leucémique après transplantation dans des souris immunodéprimées, les anomalies présentes dans la LAM ont été recherchées dans ces 2 types de CSH. *FLT3-ITD* n'était jamais retrouvée dans les CSH normales résiduelles, et chez un patient, une mutation d'*IDH1* était également absente de cette population, suggérant des événements tardifs dans la phylogénie. A l'inverse, les mutations de *TET2*, *SMC1A*, et *CTCF* ont été retrouvées dans les 2 populations de CSH, suggérant des événements précoces. Chez 2 patients, l'analyse de colonies dérivées des CSH résiduelles après culture en méthylcellulose a permis de démontrer que les mutations de *TET2* étaient le premier événement de la phylogénie. D'autres mutations non récurrentes dans les LAM ont également été retrouvées comme premier événement chez 2 autres patients (*KCTD4* et *SKP2*). Les autres mutations récurrentes présentes dans les 2 types de CSH (*NPM1*, *CTCF*, *SMC1A*) se produisent plus tardivement dans la phylogénie d'après l'étude des colonies ¹⁷⁴.

En utilisant la même stratégie chez 10 patients complémentaires ¹⁷⁵, la même équipe a démontré que les mutations d'*IDH2* et de *DNMT3A* étaient précoces dans la phylogénie. Les mutations d'*ASXL1*, d'*IKZF1* et l'inversion du chromosome(16) étaient détectables dans les CSH résiduelles non leucémiques suggérant des événements précoces, mais l'étude en colonies n'a pas confirmé d'ordre mutationnel précis. Chez ces patients, les mutations impliquant la voie *RAS* ou *NPM1* étaient retrouvées seulement dans les CSH leucémiques, suggérant des événements plus tardifs. Au final, les travaux de cette équipe suggèrent une acquisition séquentielle des anomalies génétiques dans les LAM, selon un modèle récurrent (figure 7).

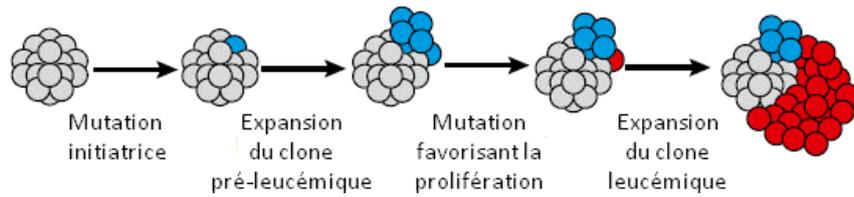


Figure 7 : Modèle de leucémogénèse par acquisition séquentielle de mutation. Une mutation initiatrice (en bleu) se produit dans une CSH donnant naissance à une hématopoïèse pré-leucémique. L'accumulation secondaire de mutations dans le clone pré-leucémique (en rouge) aboutit à la transformation en maladie très proliférative. Adapté d'après Corces-Zimmerman et al, PNAS, 2014)

Dans une autre étude centrée sur les LAM avec t(8;21) et de mutation de *c-KIT*, il a été démontré qu'au diagnostic, dans toutes les cellules $CD34^+CD38^-$ on pouvait détecter le transcrite *AML1-ETO* et les mutations de *c-KIT*. L'analyse de colonies dérivées de la même population cellulaire en RC retrouvait environ 1% de cellules dans lesquelles le transcrite *AML1-ETO* est détectable, tandis que les mutations de *c-KIT* étaient toujours absentes. La présence du transcrite *AML1-ETO* dans une fraction des granulocytes matures et des lymphocytes en RC suggère la nature pré-leucémique précoce de cet événement ¹⁷⁶.

Au bilan, l'ensemble des études à l'échelon clonal a permis de démontrer que les mutations de *DNMT3A* sont des événements pré-leucémiques. Les mutations d'autres régulateurs de l'épigénétique comme *TET2* et *IDH2* sont des événements initiateurs du clone, et peuvent être retrouvés au sein d'une hématopoïèse pré-leucémique. Les transcrits de fusion impliquant le CBF et les mutations d'*ASXL1* semblent également être des événements très précoces dans l'histoire des clones leucémiques.

C.5- Formes familiales de LAM et prédispositions génétique

A partir de l'étude de cas familiaux de LAM, des mutations prédisposant à la maladie ont pu être identifiées. Ces mutations sont retrouvées dans une dizaine de gènes impliqués dans différentes fonctions cellulaires. On compte parmi elles, les mutations classiques de *TP53*, impliquant des anomalies de réparation de l'ADN (syndrome de Li-Fraumeni) ou les pathologies du complexe de la télomérase avec mutations de *TERT* ou de *TERC*³⁹ qui peuvent également prédisposer aux aplasies médullaires (dyskératose congénitale ou maladie apparentées), et aux LAM secondaires par évolution clonale¹⁷⁷.

Certaines mutations de FT impliqués dans la maturation myéloïde, et déjà décrites comme somatique dans les LAM, ont été retrouvées avec une atteinte germinale. Ces mutations vont prédisposer les patients aux LAM. Ainsi, ont été décrites des mutations de *RUNX1*, associées à des thrombopénies familiales, et qui peuvent évoluer en LAM¹⁷⁸. Des mutations de *GATA2*¹⁷⁹, qui prédisposent à des syndromes myélodysplasiques de mauvais pronostic ont également été rapportées. On peut également citer des mutations familiales de *CEBPA*, dans lesquelles la perte du deuxième allèle est un second événement fréquent et parfois isolé au cours de la transformation leucémique^{180,181}.

Récemment, des mutations germinales de *DDX41*, protéine impliquée dans la machinerie d'épissage ont été rapportées. Ces mutations prédisposent à l'acquisition de mutations somatiques du 2^{ème} allèle de *DDX41*, ce qui aboutit à des pathologies myéloïdes. Dans près de 50% des cas aucune autre mutation complémentaire n'a été détectée au cours de la maladie, suggérant un rôle essentiel de *DDX41* dans la leucémogénèse. La perte de fonction de *DDX41 in vitro* et dans des modèles murins aboutit effectivement à une plus grande prolifération cellulaire, et à une plus grande capacité clonogénique des précurseurs hématopoïétiques. Le mécanisme cellulaire précis est mal connu, mais les mutations de *DDX41* sont associées à de nombreuses anomalies de l'épissage (anomalies de longueur des ARNm par rapport à ceux des cellules sans mutation). A noter que ce gène est porté sur le chromosome 5q, et pourrait être impliqué en plus d'autres mécanismes dans les syndromes 5q-¹⁸².

Enfin, récemment, la duplication du locus 14q32.2 impliquant les gènes *GSKIP* et *ATG2B*, dont les fonctions précises sont mal connues, a été décrite comme pouvant prédisposer à de nombreuses pathologies myéloïdes, et majoritairement des néoplasmes

myéloprolifératifs, pouvant se transformer en LAM de très mauvais pronostic ¹⁸³. L'acquisition de mutations somatiques de *TET2*, *IDH1* ou *IDH2* est fréquente au moment de la transformation en LAM, suggérant un mécanisme de leucémogénèse différent du cas précédent.

Bien qu'assez mal connues jusque-là, les mutations germinales impliquées dans les LAM pourraient être très sous-estimées, en raison de temps de latence souvent longs avant d'arriver à une maladie symptomatique. Ce délai de latence, malgré la présence des mutations à toutes les générations, peut expliquer l'absence de maladies dans certaines d'entre elles, ce qui complique les enquêtes familiales. Les explorations génomiques systématiques qui devraient être menées dans les années à venir pourraient révéler une fréquence beaucoup plus élevée d'anomalies germinales que celle attendue d'après les séries historiques.

C.6- Mesure de la maladie résiduelle

L'évaluation du niveau de réponse aux traitements paraît intuitivement un facteur pronostique important pour les patients atteints de LAM, avec un pronostic d'autant meilleur que la réponse est importante. Les techniques de cytologie classique permettent seulement d'affirmer la RC cytologique, soit, par définition le passage d'une moelle complètement infiltrée par les blastes (entre 20 et 100 %) à une infiltration inférieure à 5%, soit au mieux une réduction d'un facteur 100 environ. L'obtention de la RC cytologique après une seule cure est un facteur pronostique bien démontré dans les LAM ^{39,184}.

On estime cependant qu'au diagnostic de LAM, le nombre de cellules malades est aux alentours de 10^{11} . L'évaluation de la RC cytologique seule, même si elle reste essentielle, a donc une puissance très limitée. Plusieurs outils complémentaires peuvent être utilisés et ont été validés pour mieux évaluer la maladie résiduelle après traitement (MRD), c'est à dire la persistance de cellules de la maladie en RC en dessous du seuil de détection cytologique.

C.6.a- Cytogénétique classique et FISH

Des anomalies cytogénétiques sont présentes au diagnostic dans 40 à 50 % des LAM. L'étude du caryotype des cellules médullaires lors de la RC peut donc fournir des informations complémentaires sur le degré de réponse au traitement. La cytogénétique classique a une sensibilité proche de celle de la cytologie (étude de 20 mitoses environ). La persistance d'anomalie cytogénétique malgré un aspect de RC cytologique peut cependant être observée, et a été dans plusieurs études de grande ampleur, globalement associée à un pronostic défavorable, proche de celui de l'absence de réponse cytologique ^{185,186}.

Les techniques de FISH ont une sensibilité supérieure à la cytogénétique classique (étude de 200 noyaux interphasique +/- mitoses), même si elle reste limitée. Dans plusieurs études, la persistance de lésions chromosomiques détectables en FISH ont été associées à un pronostic défavorable, quelles que soient les anomalies étudiées ¹⁸⁷.

L'importance pronostique de la persistance des anomalies cytogénétiques est probablement à préciser en fonction de la place de ces événements dans la phylogénie clonale (occurrence précoce ou tardive dans l'architecture clonale).

C.6.b- Cytométrie en flux

L'étude de la persistance des cellules leucémiques par cytométrie de flux a été largement évaluée par différentes équipes. L'absence de marqueur réellement spécifique aux cellules leucémiques et permettant de les distinguer des cellules normales, ou de CSH pré-leucémiques, peut être une limitation à cette technique. L'utilisation d'une combinaison de marqueurs spécifique à chaque patient peut tout de même permettre l'obtention de sensibilité estimée de 10^{-3} à 10^{-4} environ. Chez certains patients, aucun phénotype spécifique n'est retrouvé, ne permettant donc pas d'assurer de suivi. Ces techniques sont par ailleurs difficilement reproductibles. Pratiquées dans des laboratoires spécialisés et entraînés, et avec des panels de marqueurs rigoureusement sélectionnés, elles sont cependant une des techniques de référence de suivi des malades. La persistance de niveaux élevés de MRD en cytométrie de flux après une ou plusieurs cures de chimiothérapie ou avant greffe de moelle a globalement été associée à un pronostic défavorable chez les sujets jeunes comme chez les sujets âgés^{188,189}.

C.6.c- Biologie moléculaire classique

Les outils classiques de biologie moléculaire, et notamment la PCR quantitative ou la RT-PCR quantitative, permettent la détection de certaines anomalies génétiques avec une sensibilité variant de 10^{-4} à 10^{-5} . Ces techniques sont utilisées en pratique courante pour certains transcrits de fusion liés aux réarrangements chromosomiques (CBF ou *MLL*), ou le suivi de la charge mutationnelle de *NPM1*.

Comme déjà mentionné, de nombreuses translocations chromosomiques sont associées à des transcrits de fusion aberrants et spécifiques des cellules leucémiques ou pré-leucémiques. Ainsi, la t(8;21) correspond au transcrit *RUNX1-RUNX1T1* et l'inv(16) au transcrit *CBFB-MYH11*. Ces deux transcrits sont quantifiables avec une sensibilité d'environ 10^{-4} en RT-PCR. Dans plusieurs études de large ampleur chez les sujets jeunes, le niveau de ces 2 transcrits après induction ou plusieurs cures de chimiothérapie a été associé au pronostic, avec des seuils différents selon le type de transcrit et selon les études^{190,191}. De manière générale, la ré-ascension du niveau de ces transcrits pendant le suivi du malade est corrélé à une rechute cytologique. La détection d'autres transcrits de fusion pour le suivi des LAM est également possible, notamment *PML-RAR α* dans les LA promyélocytaire, *BCR-ABL* dans les t(9;22), les transcrits secondaires au remaniements chromosomiques impliquant

MLL, ou d'autres transcrits comme par exemple DEK-NUP214. Ces 2 dernières cibles, bien que couramment utilisées sont moins bien validées sur le plan pronostique que les précédentes, notamment à cause de la fréquence limitée du nombre de patients, ne permettant pas d'étude de large ampleur.

Les mutations de *NPM1* ont également été utilisées pour le suivi de la MRD. La nature de ces mutations (insertion de 4 pb, avec 3 types de mutations couvrant environ 90% des malades mutés) permet l'utilisation de PCR spécifiques, avec un seuil de sensibilité proche de 10^{-5} . La quantification de l'allèle mutant par rapport à l'allèle sauvage est largement utilisée en pratique quotidienne. Plusieurs études ont démontrées que la diminution de l'allèle mutant en dessous d'un certain seuil (variable selon les études et la nature des traitements proposés aux patients) correspond à un pronostic plus favorable. La ré-ascension de la quantité d'allèle mutant est associée à la survenue de rechute cytologique. La fréquence élevée des mutations de *NPM1* en fait une cible particulièrement intéressante^{192,193}. La perte de ces mutations dans un certain nombre de rechutes, traduisant le caractère relativement tardif de ces mutations dans la phylogénie (10% des cas environ), peut cependant être une limite au suivi des patients par cette technique¹⁶⁰.

L'évaluation de l'expression de *WT1* dans les cellules sanguines ou médullaires est également couramment utilisée pour l'évaluation de la MRD. Le gène *WT1* est en effet très fréquemment surexprimé dans les LAM au diagnostic, toute cytogénétique confondue¹⁹⁴⁻¹⁹⁶. Cette expression est beaucoup plus basse chez les sujets sains ou en RC après traitement. La persistance d'un niveau élevé de *WT1* après traitement a globalement été associé à un moins bon pronostic, et sa ré-ascension lors du suivi à la rechute cytologique^{196,197}. La principale limitation à l'utilisation de ce marqueur est le faible niveau de sensibilité de la technique (maximum 10^{-2} à 10^{-3}).

D'autres cibles ont été étudiées pour le suivi des malades, notamment *FLT3-ITD* en utilisant des démarches spécifiques de patients car les ITD diffèrent de longueur et de position génomique entre chaque patient, et ont été évaluées sur de plus petites séries¹⁹⁸. L'évaluation de la MRD par ces cibles n'est pas pratiquée couramment. Le suivi d'autres cibles par des techniques classiques serait envisageable. Le suivi des gènes fréquemment mutés et dont les mutations sont conservées lors de la rechute comme *RUNX1* ou *CEBPA* ou d'autres pourrait être spécialement intéressant. L'approche obligatoirement spécifique de chaque patient, en

lien avec la grande variabilité de localisation des mutations dans ces différents gènes entre chaque patient, est un frein au développement de ces stratégies en pratique.

A l'heure actuelle, les marqueurs utilisés couramment pour le suivi de la MRD avec les techniques standard de biologie moléculaires (RUNX1-RUNX1T1, CBFβ-MYH11, NPM1, BCR-ABL) couvrent environ 50% des patients (hors LA promyélocytaire). Les patients concernés sont principalement les patients de pronostic favorable (CBF ou à caryotype normal avec mutation de *NPM1* en cas d'absence de *FLT3-ITD*), et qui bénéficient moins de ces stratégies de suivi. En effet, chez ces patients, l'indication d'allogreffe est rarement retenue en première ligne de traitement. Le développement de stratégies pour les patients de pronostic intermédiaire chez qui l'indication d'allogreffe est plus discutable est nécessaire.

C.6.d. Suivi en NGS

L'essor des techniques de séquençage à haut débit a poussé certaines équipes à utiliser ce nouvel outil pour la mesure de la MRD. De nouvelles cibles, comprenant des anomalies récurrentes ou non, pourraient potentiellement être définies. Le degré de couverture plus ou moins étendu des cibles selon les designs utilisés, permet, avec une même technique, de couvrir tout un ensemble de mutations au sein d'un même gène, au contraire des PCR allèle-spécifique ne couvrant qu'un seul type mutationnel.

Ainsi, dans une étude incluant 814 patients, des mutations de *RUNX1* ont été identifiées en NGS chez 211 d'entre eux¹⁹⁹. Les types de mutations étaient très variables entre les patients. Les mutations étaient globalement réparties sur l'ensemble du gène *RUNX1*. Sur 52 rechutes analysées, les mutations étaient stables dans 82% des cas, en faisant une cible potentielle acceptable pour un suivi en MRD. *RUNX1* a été secondairement séquençé chez 103 patients porteurs de mutations au diagnostic et ayant atteint la RC après traitement. Les patients ayant une charge allélique en mutation de *RUNX1* au-dessus de la médiane de la cohorte lors de cette évaluation avaient un pronostic plus défavorable. La sensibilité de la technique est d'environ 0.5 à 0.1%, et la médiane de fréquence de mutation en RC de 3,6%. Cette technique paraît donc intéressante comme outil de mesure de MRD, d'autant plus que les mutations de *RUNX1* sont fréquentes et quasi exclusives des mutations de *NPM1*.

Le NGS a également été utilisé pour le suivi de cibles classiques, mais avec des variations dans la séquence mutée selon les patients. Ainsi, plusieurs équipes ont montré que

la détection de *FLT3-ITD* était possible en NGS avec une meilleure sensibilité que la technique standard (genescan). La persistance de ces anomalies en RC et la ré-ascension de la charge allélique était observée avant la rechute cytologique chez plusieurs patients, suggérant un intérêt potentiel de suivi, notamment pour l'indication de traitements ciblés complémentaires par des inhibiteurs de FLT3^{200,201}. Le manque de stabilité de *FLT3-ITD* à la rechute, avec soit la disparition de l'ITD initiale, soit le développement d'un autre ITD, ou l'émergence d'un ITD minoritaire au diagnostic, et sélectionnée par la pression de traitement, paraît cependant être la principale limitation à l'utilisation de cette cible en pratique courante.

Peu d'étude se sont pour l'instant consacrées à l'évaluation d'autres cibles en NGS ou au suivi combiné de plusieurs cibles pour évaluer la MRD.

C.6.e- Suivi multi-cibles

La multiplication des événements génétiques décrits dans les LAM, et la facilitation de leur détection au diagnostic par séquençage ciblé d'un panel limité de gènes d'intérêt, ou par séquençage de l'exome entier, fait poser la question de l'intérêt d'un suivi conjoint de toutes les cibles, ou de la détermination des cibles les plus intéressantes pour ce suivi.

Dans une étude récente sur 50 cas de LAM et incluant 32 patients de pronostic intermédiaire, un séquençage d'exome ou du génome complet a été pratiqué au diagnostic²⁰². Un reséquencage ciblé de 6 à 15 lésions par patient a été pratiqué à la RC (au bout de 30 jours environ). Avec un seuil de VAF fixé à 2,5% (soit 5 % des cellules en considérant les mutations hétérozygotes), 24 patients avaient au moins un des variants détectable à J30, tandis que l'ensemble des variants n'étaient plus détectés chez 26 patients. Le pronostic global était meilleur en cas d'absence de lésion qu'en cas de persistance des variants (impact sur la survie globale et la survie sans maladie). De manière générale, les principales mutations persistant à J30 concernaient des régulateurs épigénétiques comme *TET2* ou *DNMT3A*, tandis que les mutations aboutissant à une plus grande prolifération étaient en dessous du seuil de détection, suggérant une résistance au traitement du clone pré-leucémique, tandis que le clone leucémique peut y être plus sensible. Cette étude s'intéresse à l'ensemble des marqueurs mais ne montre pas d'analyse fine en fonction de l'architecture clonale, et de l'ordre d'apparition des mutations au sein du clone. Cette dernière analyse permettrait probablement le développement de panels plus standardisés de re-séquencage ciblé, adaptable au suivi de plusieurs patients. Malgré le faible nombre de patients inclus, la différence de pronostic impressionnante

observée entre les 2 groupes, en fixant un seuil de détection relativement élevé, proche du seuil de détection cytologique, suggère que des techniques de détection très précises ne sont pas forcément utiles. La détection, même à un seuil relativement élevé, des anomalies les plus précoces dans la phylogénie pourrait donc être suffisante pour évaluer le pronostic des malades.

Dans une autre étude incluant 31 patients porteurs de mutations de *NPM1*, un suivi conjoint de mutations dans d'autres gènes (*IDH1*, *IDH2*, et *DNMT3A*) a été pratiqué en NGS. Cette étude a démontré une bonne corrélation de la mesure de MRD en utilisant les mutations *d'IDH1/2* et de *NPM1*, avec des taux plus élevés de mutations *d'IDH1/2* résiduelle par rapport à *NPM1*, probablement en raison d'une occurrence précoce dans la phylogénie, et une ré-ascension fréquente de la charge mutationnelle à la rechute. A l'inverse, les mutations de *DNMT3A* étaient souvent détectables à un niveau élevé en RC, à cause à la persistance d'une hématopoïèse pré-leucémique, y compris en cas de RC prolongée sans rechute (mutations retrouvées dans les populations myéloïdes et lymphocytaires) et semblait en faire un mauvais marqueur pour le suivi de l'épisode leucémique²⁰³.

De ces 2 études, on peut penser que l'utilisation des marqueurs très précoces dans la phylogénie n'est pas la plus adaptée au suivi précoce des leucémies. En effet, la persistance de ces mutations en RC ne semble pas forcément associée au risque de rechute précoce. On peut cependant s'interroger sur le risque de nouvelle accumulation d'événements à plus long terme, ce qui aboutirait à la rechute à partir d'un clone pré-leucémique persistant. La disparition des événements précoces, signant la disparition de l'hématopoïèse clonale pré-leucémique, devrait par contre garantir une guérison complète. Cette évaluation en fin de traitement, ou le suivi au long cours de ces anomalies en cas de persistance après traitement pourrait potentiellement être informatif. L'utilisation de marqueurs très tardifs dans la phylogénie paraît de moindre intérêt, car ces marqueurs sont souvent absents à la rechute, même en cas de rechute précoce. L'évaluation d'une stratégie de suivi séquentiel de tous les marqueurs présents au diagnostic des LAM, avec une évaluation fine des meilleures cibles pour un suivi précoce de l'épisode leucémique, et pour un suivi au long terme de l'hématopoïèse pré-leucémique reste à faire.

D- Objectifs et problématiques du travail

Les mécanismes de la leucémogénèse ont été partiellement décrits. Une majorité des LAM semble dériver d'un clone pré-leucémique. L'apparition d'une mutation dans une CSH normale conduit à une hématopoïèse pré-leucémique, où toutes les cellules sont porteuses de la mutation initiatrice. Dans ce clone pré-leucémique, les événements génétiques et chromosomiques vont s'accumuler avec le temps, jusqu'à la transformation en LAM. Plusieurs gènes connus pour être mutés dans les LAM sont fréquemment retrouvés dans l'HCLA, notamment *TET2*, *ASXL1* et *DNMT3A*. Seules les mutations de *DNMT3A* présentes dans environ 20% des LAM ont été formellement démontrées comme pré-leucémiques, c'est à dire à la fois précoces dans la phylogénie clonale, persistantes en rémission et en rechute, et capables d'initier un état pré-leucémique ou leucémique in vivo. Les mutations de *TET2*, et d'autres événements sont très précoces, mais soit leur capacité de persistance après le traitement ou en rechute, soit la démonstration de leur potentiel à conduire à un état pré-leucémique sont moins claires. La phylogénie clonale des LAM porteuses d'autres événements génétiques ou chromosomiques n'a été que partiellement décrite, et il n'est pas certain que l'HCLA soit le mécanisme de leucémogénèse de toutes les LAM.

Un premier objectif de ce travail est l'étude de la phylogénie clonale de l'ensemble des LAM de l'adulte, et l'identification des lésions initiatrices et des lésions pré-leucémiques. Pour cela, nous avons étudié la phylogénie clonale chez 74 patients adultes atteints de LAM « tout-venant », incluant 53 cas consécutifs traités intensivement dans le service d'hématologie de l'hôpital Saint-Antoine (Paris), et reflétant l'ensemble des LAM de l'adulte du point de vue cytogénétique et moléculaire. L'étude à l'échelon clonal (colonies) sur 16 patients au diagnostic, dont 8 patients à caryotype anormal, nous a permis d'établir avec certitude des ordres récurrents d'accumulation des mutations, et d'identifier les événements génétiques les plus précoces. L'étude de l'architecture clonale approchée, déduite de la fréquence des variants ou des anomalies chromosomiques chez le reste des patients confirme ces ordres récurrents dans la majorité des cas. L'étude dans des modèles de xénogreffe des capacités de repopulation hématopoïétique de cellules issues de 38 patients montre des capacités de prise de greffe accrues en cas de présence d'événement épigénétique précoce (impliquant notamment *TET2*, *ASXL1*, *DNMT3A*, *MLL*, ou la del(20q)). Les prises leucémiques et non leucémiques avec une part d'hématopoïèse multi-lignées démontrent les capacités fonctionnelles de ces différents événements à engendrer une hématopoïèse pré-

leucémique. Parmi les patients ne répondant pas à ces hiérarchies récurrentes de mutations, on retrouve la majorité des patients porteurs de prédisposition aux LAM (n=5), ainsi que tous les patients porteurs de mutation de *TP53* (n=7).

L'étude de l'évolution clonale des LAM après traitement n'a été que partiellement décrite, principalement dans des catégories précises de patients (mutations de *FLT3*, ou mutations de *NPM1* ou de *RUNX1*) ou dans des séries non sélectionnées d'effectif limité, ou avec un nombre limité de mutation étudiées.

Un deuxième objectif de notre étude est la caractérisation de l'architecture clonale des LAM en rechute. Pour cela, l'étude des 17 patients ayant rechuté dans la série de 53 cas consécutifs, avec un panel de 122 gènes, a permis d'établir un maintien constant des lésions précoces présentes au diagnostic, tandis que les lésions les plus tardives ne sont que très rarement maintenues, quel que soient le caryotype et le génotype initial des patients. L'étude de 5 cas exceptionnels de rechute après plus de 5 ans, nous permet d'observer la même conservation des événements précoces. Elle permet par ailleurs d'établir une relation directe entre le nombre d'événements variant dans la composition clonale, et le délai jusqu'à la rechute, démontrant le mécanisme darwinien de la sélection clonale dans les LAM.

Les techniques actuelles de suivi moléculaire de la MRD ne permettent le suivi que d'environ 50% des patients, dont une majorité associée à un pronostic favorable. Les techniques de NGS n'ont pour l'instant été utilisées que sur des cibles très limitées. La seule étude multi-cibles adaptée à l'ensemble des événements présents chez un patient s'est intéressée à la disparition de l'ensemble des anomalies présentes au diagnostic, mais n'a pas établi de stratégie claire de suivi par rapport à la phylogénie clonale des LAM des patients. La persistance fréquente des mutations de *DNMT3A* à un haut niveau après chimiothérapie, sans rechute systématique peut faire remettre en doute la pertinence de stratégies incluant les marqueurs d'HCLA pour le suivi des LAM.

Un troisième objectif du travail est de définir de nouvelles stratégies de suivi de la MRD multi-cibles en utilisant le NGS, et de définir les meilleures cibles pour établir ce suivi. L'étude de 20 cas en RC par reséquençage en profondeur, dont une majorité de patients ayant rechuté a permis d'observer la persistance systématique des anomalies les plus précoces de la phylogénie. La persistance de multiples marqueurs à un haut niveau d'expression semble

signer un risque élevé de rechute précoce. Cet intérêt doit être confirmé par l'étude supplémentaire de patients n'ayant pas rechuté.

Résultats : Article 1

Title: Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia

Pierre Hirsch ^{1-4*}, Yanyan Zhang ^{5,6*}, Ruoping Tang ⁴, Virginie Joulin ^{5,6}, H  l  ne Boutroux ^{1-3,7}, Elodie Pronier ⁶, Christophe Marzac ⁸, Dominique Bories ⁹, Fanny Fava ¹, Hayat Mokrani ⁶, Aline Betems ⁶, Florence Lorre ¹⁰, R  mi Favier ⁸, Fr  d  ric F  ger ⁸, Mohamad Mohty ^{1,3}, Luc Douay ^{1,2,8}, Ollivier Legrand ¹⁻⁴, Chryst  le Bilhou-Nabera ^{1-3,8}, Fawzia Louache ^{5,6†} and Fran  ois Delhommeau ^{1-3,8†}

1-Sorbonne Universit  s, UPMC Univ Paris 06, UMR_S 938, CDR Saint-Antoine, F-75012, Paris, France

2-INSERM, UMR_S 938, CDR Saint-Antoine, F-75012, Paris, France ;

3-Sorbonne Universit  s, UPMC Univ Paris 06, GRC n  7, Groupe de Recherche Clinique sur les My  loprolif  rations Aigu  s et Chroniques MYPAC, Paris, France.

4- AP-HP, H  pital St Antoine, Service d'H  matologie clinique et de th  rapie cellulaire, F-75012, Paris, France

5- Institut National de la Sant   et de la Recherche M  dicale (INSERM) UMRS 1170, CNRS GDR 3697 Micronit, 94805 Villejuif, France

6- Institut Gustave Roussy, Univ Paris-Sud, Universit   Paris Saclay, 94805 Villejuif, France

7- AP-HP, H  pital Armand-Trousseau, Department of Pediatric Hematology and Oncology, F-75012, Paris, France

8- AP-HP, H  pital Saint-Antoine & H  pital Armand-Trousseau, Service d'h  matologie biologique, F-75012, Paris, France

9- AP-HP, H  pital Henri Mondor, Unit   d'H  matologie Mol  culaire, F-94010, Cr  teil, France

10- AP-HP, H  pital Saint-Antoine, Laboratoire commun de biologie et g  n  tique mol  culaires, F-75012, Paris, France

* these two authors contributed equally

† these two authors contributed equally

Acute myeloid leukaemia (AML) emerges from haematopoietic stem/progenitor cells (HSPCs) that acquire multiple genomic or chromosomal aberrations, some being considered as initiating lesions. These pre-leukaemic events can be identified through their early occurrence in the clone, their persistence at relapse, and their ability to provide a selective advantage to mutant over normal HSPCs *in vivo*. Here, we show that mutations in *DNMT3A*, *TET2*, *ASXL1*, as well as *MLL* translocations and chromosome 20q deletions are the first driver events in most AMLs. These early lesions persist in remission, and are retained at relapse in clones that variegate increasingly with time from diagnosis to recurrence, thereby re-initiating the leukaemic process at various stages. Cells from patients with these lesions can repopulate bone marrow of xenotransplanted NOD/SCID/IL-2R γ -null (NSG) mice with leukaemic or non leukaemic engraftment, a functional signature of expanding pre-leukaemic stem cells. Based on the nature, order, and association of lesions, we delineate recurrent genetic hierarchies in AML. The most common one involves early mutations in *DNMT3A*, *TET2*, and *ASXL1*, the epigenetic regulators involved in age-related clonal haematopoiesis (ARCH). These events frequently aggregate together or with mutations targeting other epigenetic modifiers in a dominant clone that further acquires mutations in *NPM1* or in haematopoietic transcription factors, and then in proliferation regulators. Among alternative associations of events, infrequent co-mutation of *DNMT3A* and *TP53* can occur in *de novo* AMLs where *DNMT3A* mutations may provide an expansion potential to *TP53* mutant HSPCs. We conclude that initiating lesions, variegation, and selection pressure govern the expansion and adaptive behaviour of the malignant clone, shaping AML in a time dependent manner.

Some of the lesions that occur in AML are believed to result in clonal expansion of mutant HSPCs, leading to ARCH¹⁻⁴, a condition that predisposes to the subsequent acquisition of leukaemic mutations. To figure out how this model of leukaemogenesis fits with the genetic variety of AML, we analysed a series of cases (Supplementary Table 1) with respect to three major hallmarks of such candidate events: 1) their early occurrence in a founding clone⁵⁻⁸, 2) their persistence at relapse^{5,7}, and 3) their ability to initiate leukaemia or pre-leukaemia *in vivo*^{5,8}.

To address the first point, we analysed the clonal composition of 53 consecutive AMLs by routine cytogenetic and PCR techniques, targeted resequencing of 122 genes recurrently

mutated in AML, and exome sequencing in four *MLL*-rearranged cases (Supplementary Tables 2, 3, 4). As multiple lesions can lead to various functional consequences^{9,10}, we classified them into four categories (Extended Data Table 1). The first category comprises lesions disturbing epigenetic regulation: mutations in DNA methylation and chromatin modifiers, *MLL* and core binding factor (CBF) translocations¹¹⁻¹³, and del(20q)¹⁴. The second category includes mutations in splicing factors, transcription factors, and in *NPM1*, all impairing haematopoietic differentiation¹⁵⁻¹⁸. The third category refers to mutations in genes that regulate proliferation and the fourth category to all other events. Fifteen samples, eight from AMLs with normal karyotype, and seven from AMLs with abnormal karyotype, were seeded in methylcellulose cultures to obtain single-cell derived colonies. Individual colonies were then analysed to infer the architecture of the principal clones from the existence of progenitors with none, all, or combinations of lesions previously detected in the bulk material.

We first focused on eight patients with normal karyotype and mutations in *TET2*, *DNMT3A*, or *ASXL1*, the most frequently involved genes in ARCH¹⁻³. In four patients, *TET2* mutations were found in first position, preceding *DNMT3A* mutations in two double-mutant cases. In three other patients without *TET2* mutations, *DNMT3A* mutations were found first (Fig. 1a, b, Extended Data Fig.1a, b, Supplementary Fig.1). The last patient had an *ASXL1* mutation followed by lesions in *EZH2*, *RUNX1* and *FLT3* (Fig. 1b, Supplementary Fig.1). In three cases (UPN2014-001, UPN2014-020, UPN2014-022), we found evidence for branching subclones involving variants in *TET2*, *FLT3*, and *NRAS*, all detected with low allele frequencies in the bulk AML DNA (Supplementary Table 3).

We then analysed the seven AMLs with abnormal karyotypes. Informative polymorphisms and fluorescent in situ hybridization (FISH) were used to detect losses of heterozygosity (LOH) and translocations, respectively. In one case, del(20q) was the founder event, followed by a splicing machinery mutation in *U2AF1*, and a mutation in the tyrosine kinase *RET* (Fig. 1c, e). In a second patient, del(16q) was found after an *U2AF1* mutation, and preceding an *FLT3*-internal tandem duplication (ITD) with LOH. In one patient with monosomy 7, the first event was a mutation in *SF3B1*, followed by one in *PTPN11*, the loss of chromosome 7 and an *FLT3* mutation. In four other patients, *MLL* translocations were identified as first events (Fig. 1d,e, Extended Data Fig. 1c, Supplementary Fig. 2).

Taken together, single cell derived colony analyses revealed that mutations in epigenetic regulators, *MLL* rearrangements, and del(20q) are early events. Conversely, in 13/15 cases, proliferation-associated events, including *FLT3* and RAS pathway mutations, were found as last events.

A second characteristic of pre-leukaemic lesions is that they persist at relapse^{5,7}. To investigate this, we listed the changes in clonal composition between diagnosis and relapse in 22 patients, including five additional patients who had experienced relapse over five years after diagnosis. In all patients but two, from one to 18 genomic or chromosomal changes were detected (Fig. 2a, Extended Data Table.2, Supplementary Fig.3, Supplementary Table 1 and 5). In contrast to epigenetic modifying events, which were mostly unchanged in the relapse clones (38/54 lesions), late proliferative events changed frequently (27/35 events, $p < .0001$, Fisher's exact test) (Fig. 2b). Moreover, all epigenetic lesions identified as first events in our previous colony analyses -*DNMT3A*, *TET2*, *ASXL1* mutations, *MLL* rearrangement and chromosome 20q deletions- were retained at relapse (Fig. 2a, Extended Data Table 2). We also observed that, while the number of retained lesions decreased with time to relapse (4 ± 0.6 before five years versus 1.6 ± 0.7 after five years, $p = .016$, Mann Whitney test), the number of variegating lesions increased in the meantime (2.2 ± 0.3 before five years versus 7.6 ± 2.9 after five years $p = .019$) (Fig. 2c, Extended Data Table 2, Supplementary Fig.3). Of note, four out the five patients with late relapses had multiple changes in clonal composition. In one patient (UPN2015-003), we concluded that a second disease, with fully distinct genotype and karyotype from those of the initial leukaemia, had occurred (Fig. 2c, Extended Data Table.2). In the three other patients, persistent *DNMT3A* mutant clones had lost most mutations, including an *NPM1* mutation (UPN2015-014), an *IDH2* mutation (UPN2015-004), and a *RUNX1* mutation (UPN2015-005). At relapse, these founding clones had re-evolved with new mutations in *NPM1*, *IDH1*, and *RUNX1*, respectively (Fig. 2a, e, Extended Data Table.2). None of the relapse specific mutations was detected by deep sequencing with a 0.001 sensitivity (Supplementary Table 6). In patient UPN2015-014, the analysis of sequential samples showed that the relapse arose after the emergence of an *IDH2* mutant sub-clone which had subsequently acquired a second *NPM1* type A mutation identical to that of the diagnosis clone (Fig. 2e). These results suggest that leukaemia treatments had reset the clones back to their pre-leukaemic states, with evolutionary potentials similar to those of the ancestral clones.

To assess the persistence of relapse reservoirs after treatment, we performed Sanger sequencing, FISH, quantitative RT-PCR or deep sequencing analyses in samples from 12 patients in remission, including 10 who further experienced relapse. Early *DNMT3A*, but also *TET2* and *ASXL1* mutations, as well as *MLL* and del(20q) rearrangements were all detected in these samples, whereas later events were inconstantly present. (Fig. 2e, Extended Data Fig.2, Supplementary Table 6). Taken together, these results show that cells carrying these lesions are frequently not eradicated by the treatment, and are able to re-initiate the disease.

The pre-leukaemic potential of AML lesions can be assessed in xenotransplantation models^{5,8}. We thus injected $5 \cdot 10^6$ T-depleted mononuclear cells from 38 AML blood samples to sublethally irradiated NSG mice to analyse their repopulation capacities with respect to cytogenetic and mutational patterns. Among these samples, 13 led to overt leukaemic engraftment with a clonal composition of human cell population remarkably mirroring the one of injected samples (Fig. 3a, Extended Data Fig.3, Supplementary Table 7, Supplementary Fig.4). Fifteen samples repopulated NSG bone marrow with more than 0.5% of human CD45⁺(hCD45⁺) cells but no overt leukaemia (Fig. 3a,b). Lymphoid or lympho-myeloid repopulation was detected in 13 of these 28 engrafting samples, including samples with *DNMT3A*, *TET2*, *ASXL1* mutations, or *MLL* rearrangements at injection. In total, of the 38 AML samples, most samples with *TET2* mutations (7/8), *DNMT3A* mutations (9/12), *ASXL1* mutations (6/7), and *MLL* fusions (3/4) led to either leukaemic or non leukaemic repopulation (Fig.3a,b, Extended Data Fig.3, Supplementary Fig.4).

As *TET2* mutations were associated with *DNMT3A* or *ASXL1* mutations in 6/7 engrafting samples, and preceded *DNMT3A* mutations in some patients (Fig.1 and Extended Data Fig.1), we next asked whether isolated *TET2* loss could induce pre-leukaemic expansion. CD34⁺ cells from cord blood were transduced with lentiviruses expressing small hairpin RNA designed to knockdown *TET2* (shRNA *TET2*) or scramble control¹⁹ and injected into NSG mice. Four months after injection, *TET2* knocked-down cells showed a 2.6-fold greater repopulation (mean \pm s.e.m: $50 \pm 6\%$ of hCD45⁺ cells in mouse bone marrow) than control cells ($19 \pm 5\%$), with both lymphoid and myeloid potentials (Fig. 3c,d, Extended Data Fig.3, Supplementary Fig.5a). Moreover, *TET2* knocked-down human cells sorted from primary recipients' bone marrow led to lympho-myeloid repopulation 12 weeks after injection into secondary recipients (Supplementary Fig.5b). This result shows that *TET2* depletion is sufficient to improve the multi-lineage repopulation of NSG bone marrow with intact self-

renewing capacity, a signature of pre-leukaemic stem cell function, as previously defined for *DNMT3A* mutations⁵. It has been proposed that the expansion of HSPCs with *DNMT3A* mutations favours subsequent accumulation of additional mutations, driving progression to AML^{5,6}. Altogether, our xenotransplantation results support this model for other putative pre-leukaemic lesions, including *TET2* mutations, *ASXL1* mutations, and *MLL* rearrangements.

By combining our data, we then attempted to reconstruct the clonal phylogeny of all AMLs (Extended data Fig. 4). In 49/72 cases at diagnosis, the first events observed were mutations in epigenetic regulators, CBF translocations, *MLL* rearrangements or chromosome 20q deletions, with a frequent (n=27) early accumulation of epigenetic events in a dominant clone. This first hit was recurrently followed by lesion affecting *NPM1*, transcription factors, or the splicing machinery, and then mutations in signalling pathways (Fig. 4a, Supplementary Table 8 and 9). Depending on initial lesions, distinct genetic hierarchies were noticed. First, 27/72 patients had a genetic hierarchy reminiscent of ARCH, with *DNMT3A*, *TET2*, or *ASXL1* mutations co-occurring with either mutations in *NPM1* or in *RUNX1*, *CEBPA*, and *GATA2* (Fig.4b, Extended data Fig. 5, Supplementary Table 8 and 9). A second group of 14/72 patients had CBF, *MLL* or chromosome 20q rearrangements in first position, but no mutations in *NPM1* or in the aforementioned transcription factors (Fig.4c, Extended Data Fig. 5, Supplementary Table 8 and 9).

In 21/72 patients, neither CBF, *MLL*, 20q rearrangements nor mutations in *NPM1*, *RUNX1*, *GATA2* or *CEBPA* were retrieved (Extended Data Fig.5). Beside cases with germ line variants predisposing to AML - three *DDX41* mutations²⁰, one *CSF3R* mutation, and one 14q32.2 duplication²¹ - (Fig.4d, Extended Data Fig.6), this group comprised all seven patients with *TP53* mutations (Fig.4e, Extended Data Fig.6). Within the latter cases, we found co-occurring *DNMT3A* mutations in three out of four *de novo* AMLs but not in two therapy-related AMLs (t-AML) and one secondary AML. As VAFs indicated that *TP53* and *DNMT3A* mutations occurred in a unique dominant clone (Fig. 4e, Extended Data Fig.6, Supplementary Table 3), we analysed 155 single cell-derived colonies from patient UPN2014-027. We failed in conclusively determining which of the *TP53* p.Y181C or *DNMT3A* L508P mutation occurred first because all 138 mutant colonies had both mutations. This result contrasted with those previously obtained from the seven *DNMT3A* and *TET2* mutant samples, where subsets of colonies with single founding mutations were detected, as a reflection of an expansion step preceding the acquisition of the second event (Fig. 4f, Extended Data Fig.7). *TP53* mutations

play a major role in t-AML, in which the previous exposure to chemotherapy for a first cancer leads to the preferential expansion of pre-existent heterozygous mutant HSPCs²². Such an exposure does not exist in *de novo* AMLs, and our results suggest that co-occurring lesions may be essential to provide an expansion capacity to *TP53* mutant HSPCs. In line with this, among the 39 *TP53* mutations detected in 1,125 individuals with ARCH¹⁻³, nine had co-occurred with other mutations, including two in *DNMT3A* and four in *TET2*.

Collectively, our data identify distinct routes for AML initiation and evolution in adults. The main one involves mutations in epigenetic regulators, such as *DNMT3A*, *TET2*, and *ASXL1*. Lesions in these genes were detected in 68% of individuals with ARCH and haematologic driver mutations¹⁻³ but in smaller proportions of *de novo* non-promyelocytic AML in our series (31/59, 53%) and in The Cancer Genome Atlas study⁹ (66/184, 36%). This suggests that distinct pre-leukaemic expanding lesions initiate other AMLs. First, *del(20q)* may be a *bona fide* ARCH lesion, as it was found in non-tumoral cells from patients with multiple myeloma devoid of myeloid malignancy²³. Second, *RUNX1/RUNX1T1* translocations were detected in Guthrie cards from new-borns who developed AML several years later²⁴, and both CBF and *MLL* translocations promote non-leukaemic repopulation of immunocompromised mice by human HSPCs²⁵⁻²⁷. These data and our results converge to the hypothesis that all the aforementioned lesions lead to HSPC expansion and pre-leukaemic clonal haematopoiesis that requires additional hits for the onset of AML. In expanding HSPCs, the excess in mitoses may increase the absolute number of unrepaired replication errors, thereby favouring the accumulation of subsequent mutations. Depending on the genetic hierarchy of AML - i.e the nature of the initial lesions and the number of additional mutations necessary to trigger a full blown disease -, distinct latencies from the onset of clonal haematopoiesis to AML may be expected (Extended Data Fig.8a). For instance, CBF and *MLL* rearrangements induce both epigenetic and haematopoietic transcription factor deregulation¹¹⁻¹³. This may result in the requirement of less additional lesions, and shorter latency, than in ARCH-derived diseases for the progression to acute leukaemia.

Finally, we established a link between the number of clonal changes and the delay from diagnosis to relapse. Relapse can arise from the persistence of a clone in which a few lesions changed, or in which only the first founding one was retained. In this last situation we observed that some ancestral clones were able to re-evolve in a way similar to the one that built the initial genetic hierarchy, triggering relapse up to ten years after the first AML

diagnosis. In patients who achieve complete remission, this return to a pre-leukaemic condition raises the question whether the clone will progressively re-accumulate new driver lesions or will stay dormant as is seen in the vast majority of individuals with ARCH who never develop any haematopoietic malignancy (Extended Data Fig.8b).

Methods

Samples

BM and blood samples were collected at the time of diagnosis, complete remission (CR) and relapse from AML patients after informed consent in accordance with the declaration of Helsinki. Mononuclear cells were obtained after ficoll separation and cryopreserved at the tumour bank of Saint-Antoine Hospital, Paris. Fresh or frozen samples were used for further cellular and molecular experiments. BM cells were sorted to enrich the samples in leukaemic or non-leukaemic progenitor cells. CD34⁺, CD38⁺, CD34⁻ or CD34⁺CD38⁻ cell sorting strategies were established according to initial leukaemia immuno-phenotype. Blood cells from CR were sorted to isolate the CD3⁺ fraction. Immuno-magnetic cell sorting was performed using CD3, CD34 or CD38 MicroBeads and columns (MACS®, Miltenyi Biotec™) according to the manufacturer's protocols.

Umbilical cord blood samples were collected from healthy new-borns with mothers' consent (Etablissement Français du Sang, Créteil France). CD34⁺ cells were isolated using a dextran/ficoll based procedure followed by immuno-magnetic separation (MACS®, Miltenyi Biotec™).

Targeted sequencing

A panel of 122 genes mutated in AML and myeloid malignancies was designed (supplemental table). Amplicon libraries were obtained from 225 ng of BM DNA, using HaloPlex Target Enrichment System® (Agilent technologies), according to the manufacturer's protocol. Sequencing was performed using a MiSeq® sequencer (Illumina^{INC}) using manufacturer recommendations.

Results were analysed after alignment of the reads using the SureCall® Software version 3.0.1.4 from Agilent Technologies™. BWA MEM algorithm was used for alignment and Illumina^{INC} SNPPEP SNP Caller algorithm was used to identify SNP and indels variants. Minimum allele frequency for variant calling was set at 5% with a minimum local depth at 40. All variants were manually checked using IGV 2.3 software. Variants identified in diagnosis or relapse samples were manually checked in paired relapse or diagnosis samples using IGV software to study clonal evolution between samples and to rule out low frequency variants. For all samples, average depth in target regions was 619 (range 270-1650) and 98.26% (range 93.5-99.5) of target regions were covered by at least 100 reads

Detection of described polymorphism was performed by an in-house software using Ensembl database. Non-described variants of potential interest with VAF >10% were sequenced using the Sanger method in the diagnosis BM sample and in the CD3⁺ fraction from CR or non-sorted CR samples, in order to rule out artifacts and non-somatic variants. All primers are described in Supplementary Information (Supplementary Table 10).

Exome sequencing

Exome sequencing was done after sequence capture, enrichment and elution according to manufacturer's instructions (SureSelect, Agilent) without modification except for library preparation (NEBNext® Ultra kit, New England Biolabs®): 600 ng of each genomic DNA were fragmented by sonication and purified to yield fragments of 150-200 bp. Paired-end adaptor oligonucleotides from the NEBNext® Ultra kit were ligated on repaired A tailed fragments, then purified and enriched by 8 PCR cycles. 1,200 ng of these purified libraries were then hybridized to the SureSelect oligo probe capture library for 72 hours. After hybridization, washing, and elution, the eluted fraction was PCR-amplified with 9 cycles, purified and quantified by QPCR to obtain sufficient DNA template for downstream applications. Each eluted-enriched DNA sample was then sequenced on an Illumina HISEQ 2000 as paired-end 75b reads. Image analysis and base calling was performed using Illumina Real Time Analysis (RTA 1.17.21.3) with default parameters.

The bioinformatic analysis of sequencing data was based on the Illumina pipeline (CASAVA 1.8.2). CASAVA performs alignment of a sequencing run to a reference genome (hg19), calls the SNPs based on the allele calls and read depth, and detects variants (SNPs & Indels). The alignment algorithm used is ELANDv2 (performs multiseed and gapped alignments). Only the positions included in the bait coordinates were conserved. Genetic variation annotation was realized from IntegraGen in-house pipeline, consisting in annotation of genes (RefSeq) and known polymorphisms (dbSNP 132, 1000Genome, EVS) followed by a mutation characterization (exonic, intronic, silent, nonsense...). For each position, the exomic frequencies (homozygous & heterozygous) were determined from all the exomes already sequenced at Integragen, and the exome results provided by 1000G, EVS and HapMap. Minimum average depth on the exome is around 70X. Samples from diagnosis and complete remission were paired-analyzed and manually checked with IGV v2.3 to identify somatic variants. Confirmed somatic variants are shown in supplemental data. full data is available on request.

Methyl cellulose assays

Sorted cells from diagnostic samples were seeded in triplicate at 100 to 300,000 cells per 1 mL culture dish in 2% standard methylcellulose medium (Stem Cells Technologies™) supplemented with 37% fetal calf serum, 12% bovine serum albumin, 1% L-glutamine, 10^{-6} M β -mercaptoethanol, 1 IU/mL of EPO, 50 ng/mL of SCF, 25 ng/mL of FLT3 ligand, 10 ng/mL of IL3, 10 ng/mL of IL6, 10 ng/mL of G-CSF, 5 ng/mL of GM-CSF, 10 ng/mL of TPO, 100 IU/mL penicillin and 100 μ g/mL streptomycin. Colonies were counted after 14 days and picked. When appropriate, after a wash in PBS, individual colonies were splitted into two cell suspensions to perform both genotyping and FISH analyses. The first cell suspension was gently deposited onto 18-well immunofluorescence slides. Slides were dried and fixed for further FISH analyses. The second cell suspension, or whole colonies when no translocation was detected in bulk AML, was stored at -80°C for further genotyping assays.

Single cell derived colony genotyping assays

DNA from individual colonies was prepared from a 50 μ L lysis buffer containing 0.5 μ L tween 20, 18.5 μ g recombinant proteinase K (Thermo scientific™) and H_2O , after incubation at 56°C for 1 hour and 95°C for 15 minutes. Genotyping was performed using custom Taqman® SNP genotyping assays (Life technologies™) with 5 μ L colony DNA, using a 7500 fast real time PCR system (Applied Biosystem™), according to the manufacturer's protocol. For chromosomal deletion investigation, SNP genotyping assays of described SNP of the deleted area were performed, using *MYBL1* p.I624M for del(20q), *FANCA* p.G809D for del(16q) and *CUX1* p.A418T for del(7). Sequences of probes and primers are described in Supplementary Information (Supplementary Table 11).

***FLT3-ITD and HNRNPU* large indel detection**

The mutational status for *FLT3-ITD* and *HNRNPU* at diagnosis and in colonies was determined using high-resolution sizing of fluorescent dye-labelled PCR amplification. Beyond its contribution to mutation detection, high-resolution sizing also allowed to estimate the allelic burden by measuring peak height ratios (mutant/wild-type + mutant) on a fluorescence scale. PCR probes and primers are described in Supplementary Information (Supplementary Table 10).

Cytogenetic and fluorescent in situ hybridization (FISH) analyses

Conventional cytogenetic analysis was performed in diagnosis and relapse samples on R-banding metaphases obtained from 24h unstimulated culture using standard procedures. Karyotypes were interpreted according to the recommendations of the International System for Human Cytogenetic Nomenclature. Chromosomal rearrangements were confirmed by fluorescent in situ hybridization (FISH), using the following probes: *MLL* Breakapart probe (LPH 013), *AML1 (RUNX1)* Breakapart probe (LPH 027), del (5q) probe (LPH 024), del(7q) deletion probe (LPH 025), del(20q) deletion probe (LPH 020), inv(16) probe (LPH 022), fast FISH X, Y an 18 (LPF 002) (all from CytoCELL Ltd, Cambridge UK). NUP98 (11p15) Break probe (Kreatech), Vysis LSI D7S522 (7q31)/CEP7 Dual Colour probe (Abbott) and XL 7q22 / 7q36 deletion (MetaSystems) were also used. Hybridization signals were scored in 176 to 500 nuclei and analysed using CytoLabView BandView/FISHView Analysis (VDS) - 6.0 software (Applied Spectral Imaging Ltd, Israel). The same FISH analyses were performed on colonies, except that fewer nuclei were analysed.

***MLL/AF6 (MLL/MLLT4)* transcript detection and quantification.**

Total RNA was isolated from Ficoll isolated mononuclear cells by use of isothiocyanate guanidium method (Extract All, Eurobio, France). RNA (1 ug) was reverse transcribed using RT-MMLV. cDNA synthesis was performed with the following primers (forward : 5' GAGGATCCTGCCCAAAGAAAAG 3' ; reverse : 5' GGGAGAGGACAGCATTCGC 3'). Sanger sequencing of the PCR product was performed in order to confirm the *MLL* and *AF6* exons implicated into the fusion transcript. The patient follow-up was monitored using real time quantitative PCR of the *MLL-AF6* fusion according to the ELN recommendation (Gabert Leukemia 2003), using the following primers and probe (forward : 5' GTCCAGAGCAGAGCAAACAGAAA 3'; reverse : 5' GAAAATAAAAATCTCATCACTCCATGG 3'; probe : 5' CTCCCCGCCCAAGTATCCCTGTAAA 3') adapted from published data²⁸ and normalized to GUS reference gene expression . Sensitivity reached 10⁻⁵.

***CBFB/MYH11* transcript detection and quantification.**

CBFB/MYH11 transcript detection and quantification were performed as previously described²⁹.

Deep targeted sequencing

A targeted-resequencing panel including all variants detected at diagnosis and relapse in 11 patients was designed. Amplicon libraries were obtained from 57.6 ng of DNA, using HaloPlex HS Target Enrichment System® (Agilent technologies), according to the manufacturer protocol. With this method, after an initial enzymatic digestion, DNA fragments are captured with target probes, and hybridized with two indexes. The first index is sample specific. The second index is a unique random sequence of 10 nucleotides, with a specific sequence for each DNA fragment of the sample. After PCR amplification at the end of library preparation, amplicon families bearing the same indexes can be identified, allowing a better detection of PCR and sequencing errors, and a better sensitivity than standard HaloPlex target enrichment system. Sequencing was performed using a MiSeq® sequencer (Illumina^{INC}) using manufacturer recommendations. Results were analysed after alignment of the reads using the SureCall® Software version 3.0.2.1 from Agilent TechnologiesTM. Variants detected at diagnosis and relapse were manually checked in the CR sample, using IGV software version 2.3 in order to identify low frequency variants. The background error rate at an SNV position was calculated as the ratio of the sum of amplicon families with “non-reference” or “non-mutant” bases to the total number of amplicon families at this position. For all patient-specific samples, the maximum background error rate was defined as the maximum value of all background error rates obtained at target positions. As indels are not subject to sequencing error dependent miscalling, the sensitivity of indel detection was set as $1/(\text{number of amplicon families})$.

Xenograft experiments with primary AML cells and TET2 depleted cord blood CD34+ cells

NOD/SCID/IL-2Rg_c-null (NSG) mice were bred and maintained under specific pathogen free conditions with acidified water (pH 5.3) at the animal facility of Gustave Roussy Institute. Animal experiments were performed in accordance with guidelines established by the Institutional Animal Committee. Peripheral blood mononuclear cells from AML patients were depleted in CD3⁺ cells by RosetteSep human CD3⁺ depletion cocktail (StemCell Technologies) and 5×10^6 cells were i.v. injected to mice 24h after irradiation at 2.5 Gy from a ¹³⁷Cs source³⁰. Mice were analyzed at 8-43 weeks post-injection. Cells from mouse BM were stained with rat anti-mouse CD45 (Biolegend) and mouse anti-human CD45, anti-human CD19, anti-human CD33, and anti-human CD3 antibodies (all from BD Pharmingen) (clones and fluorochromes are indicated in supplementary table 12). Stained cells were analysed on

FACSort or FACSCanto II cytometers (BD Biosciences). The presence of less than 0.5% of human CD45⁺ (hCD45⁺) population was considered as non-engraftment. The presence of >5% of hCD45⁺ cells with major CD33⁺ population (> 75% of the hCD45⁺ cells) was considered as overt AML engraftment. The presence of 0.5 to 5% of hCD45⁺ or the presence of >5% of hCD45⁺ with <75% of CD33⁺ in the hCD45⁺ population was considered as non-overt or non-leukaemic engraftment.

CD34⁺ cells from three to ten distinct cord blood samples were pooled and transduced as previously described with lentiviruses expressing the green fluorescent protein (GFP) and either a short hairpin RNA targeting TET2 (shRNA-TET2, 5'-GGGTAAGCCAAGAAAGAAA-3') or a scramble sequence (shRNA-scramble, 5'-GCCGGCAGCTAGCG (□CGCCAT-3') as control ¹⁹. Twenty four hours after transduction, 2.10⁵ cells were i.v. injected to sublethally irradiated NSG mice. Mice were sacrificed 15 to 17 weeks after injection, and repopulation of mouse bone marrow (femurs and tibias) by human cells was assessed by flow cytometry, using APC-conjugated mouse anti-human CD45, PE-conjugated anti-human CD19, PE-conjugated anti-CD33 (all from BD Pharmingen).

For NGS experiments and for secondary transplantation, bone marrow of repopulated mice was enriched in human cells using a mouse/human chimera isolation kit (Stem Cell Technologies).

Determination of clonal composition

To determine the order of chromosomal and genomic lesions in each AML, we combined the following quantitative results: 1) frequencies of karyotype metaphases harbouring each chromosomal aberration, 2) frequencies of interphasic nuclei with specific translocations, rearrangements, deletions or gains, 3) VAFs from exome or targeted sequencing runs for somatic gene mutations (SNVs and indels), 4) VAFs from exome or targeted sequencing for SNPs in sequenced regions with LOHs or CNVs, 5) Peak height ratios of high-resolution sizing of fluorescent dye-labelled PCR amplification for *FLT3*-ITDs. Except for *FLT3*-ITD peak height ratios, 95% confidence intervals were calculated by using the numbers of analysed nuclei and read depths as sample sizes for cytogenetic and NGS quantifications, respectively. These quantitative values were converted to fractions of cells harbouring the lesions (Variant Cell Fraction, VCF), taking into account LOHs, with or without CNVs, as well as the gender of the patient for lesions on chromosome X. Then, lesions were ordered according to VCFs with correction in case of variant alleles involving genes mapped on

chromosomes with imbalances or LOH. When available, results from colony analyses or quantitative data from later follow-up samples were used to correct or refine the clonal phylogeny.

References

1. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
2. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
3. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
4. Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179–1181 (2012).
5. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
6. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
7. Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2548–2553 (2014).
8. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118 (2012).
9. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
10. Döhner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute Myeloid Leukemia. *N. Engl. J. Med.* **373**, 1136–1152 (2015).
11. Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
12. Li, Y. *et al.* Epigenetic silencing of microRNA-193a contributes to leukemogenesis in t(8;21) acute myeloid leukemia by activating the PTEN/PI3K signal pathway. *Blood* **121**, 499–509 (2013).
13. Alvarez, S. *et al.* DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PloS One* **5**, e12197 (2010).
14. Aziz, A. *et al.* Cooperativity of imprinted genes inactivated by acquired chromosome 20q deletions. *J. Clin. Invest.* **123**, 2169–2182 (2013).
15. El Hajj, H. *et al.* Retinoic acid and arsenic trioxide trigger degradation of mutated NPM1, resulting in apoptosis of AML cells. *Blood* **125**, 3447–3454 (2015).
16. Dolatshad, H. *et al.* Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* **29**, 1092–1103 (2015).
17. Bolli, N. *et al.* Expression of the cytoplasmic NPM1 mutant (NPMc+) causes the expansion of hematopoietic cells in zebrafish. *Blood* **115**, 3329–3340 (2010).
18. Sportoletti, P. *et al.* The human NPM1 mutation A perturbs megakaryopoiesis in a conditional mouse model. *Blood* **121**, 3447–3458 (2013).
19. Pronier, E. *et al.* Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood* **118**, 2551–2555 (2011).

20. Polprasert, C. *et al.* Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* **27**, 658–670 (2015).
21. Saliba, J. *et al.* Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. *Nat. Genet.* **47**, 1131–1140 (2015).
22. Wong, T. N. *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**, 552–555 (2015).
23. White, J. S., Zordan, A., Batzios, C. & Campbell, L. J. Deletion(20q) as the sole abnormality in plasma cell myeloma is not associated with plasma cells as identified by cIg FISH. *Cancer Genet.* **205**, 644–652 (2012).
24. Wiemels, J. L. *et al.* In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* **99**, 3801–3805 (2002).
25. Moriya, K. *et al.* Development of a multi-step leukemogenesis model of MLL-rearranged leukemia using humanized mice. *PloS One* **7**, e37892 (2012).
26. Bäsecke, J. *et al.* AML1/ETO promotes the maintenance of early hematopoietic progenitors in NOD/SCID mice but does not abrogate their lineage specific differentiation. *Leuk. Lymphoma* **46**, 265–272 (2005).
27. Wunderlich, M., Krejci, O., Wei, J. & Mulloy, J. C. Human CD34+ cells expressing the inv(16) fusion protein exhibit a myelomonocytic phenotype with greatly enhanced proliferative ability. *Blood* **108**, 1690–1697 (2006).
28. Scholl, C. *et al.* Development of a real-time RT-PCR assay for the quantification of the most frequent MLL/AF9 fusion types resulting from translocation t(9;11)(p22;q23) in acute myeloid leukemia. *Genes. Chromosomes Cancer* **38**, 274–280 (2003).
29. Gabert, J. *et al.* Standardization and quality control studies of ‘real-time’ quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. *Leukemia* **17**, 2318–2357 (2003).
30. Zhang, Y. *et al.* CXCR4 inhibitors selectively eliminate CXCR4-expressing human acute myeloid leukemia cells in NOG mouse model. *Cell Death Dis.* **3**, e396 (2012).

Acknowledgments

This project was funded by the ARC foundation (N°EML20110602421), by the Région Ile-de-France (N°2012-2-eml-06-UPMC_12016710), by the Association Laurette Fugain (N°J15I409), and by the Institut National du Cancer (INCA) (PH). We would like to thank The Cancer Genome Atlas (TCGA) for providing free access to their somatic mutational data. We would like to thank Simona Lapusan, Françoise Isnard, Anne-Claire Mamez for their help in collecting samples, Christelle Mazurier and Christine Nguyen for their technical advices and assistance, Mélanie Letexier, Jean-Paul Saraiva, and Emmanuel Martin for exome sequencing and analysis, Nicole Casadevall and William Vainchenker for helpful discussions and critical reading of the manuscript.

Authors Contributions

PH performed cell culture, genotyping and NGS experiments, interpreted the results and wrote the manuscript

YZ, FL performed xenograft experiments and wrote the manuscript

VJ performed xenograft experiments

RT designed NGS assays and participated in material collection

HB contributed in cell culture experiments

EP, HM and AB performed shTET2 experiments

CM and DB performed standard molecular analysis

FF and FL contributed in material collection, sample preparation and in standard molecular analyses

RF performed *RUNXI* Sanger sequencing

FF performed flow cytometry analysis of patient samples

MM and OL contributed in follow-up of patients

CBN performed cytogenetic and FISH analyses

FD designed the research, interpreted the results and wrote the manuscript

All authors contributed in manuscript review

Authors information

Authors declare no competing financial interest.

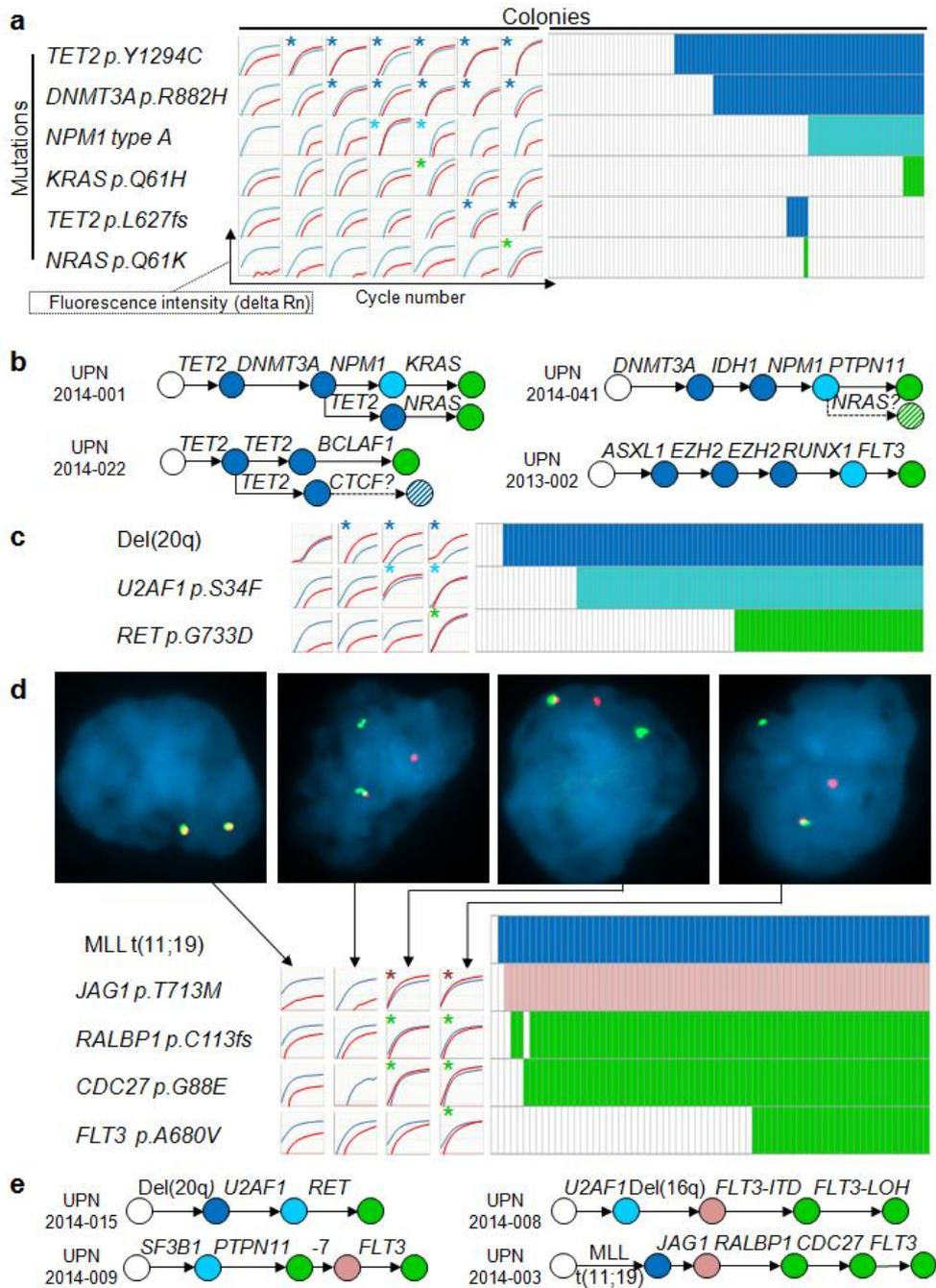


Figure 1 | Single cell-derived colony analysis identifies a recurrent order of acquisition of AML mutations and chromosomal rearrangements. **a**, Left panel : allele specific (AS) PCR amplification plots of mutant (red) and wild type (blue) alleles in seven colonies from patient UPN2014-001. Asterisks indicate mutant allele detection. Right panel: results of AS assays (rows) in 87 colonies (columns). Mutant colonies are indicated by coloured boxes. Blue: lesions in epigenetic regulators; Cyan: lesions involving *NPM1*, transcription or splicing factors; Green: proliferative lesions; Pink: other lesions. **b**, Mutation order in four patients with normal karyotype AML. Hatched circles indicate positions that could not be determined. **c**, Analysis of 85 colonies from patient UPN2014-015 as in **a**. To detect del(20q), an informative single nucleotide polymorphism (SNP)(rs11556379) was used. **d**, Left panel: genotyping analyses, as in **a**, and FISH analyses of 4 colonies from patient UPN2014-003. Dissociations of the red and green signals indicate *MLL* rearrangements. Nuclei were stained with DAPI. Right panel: results of FISH and AS assays, as in **a**. **e**, Order of lesions in four abnormal karyotype AMLs .

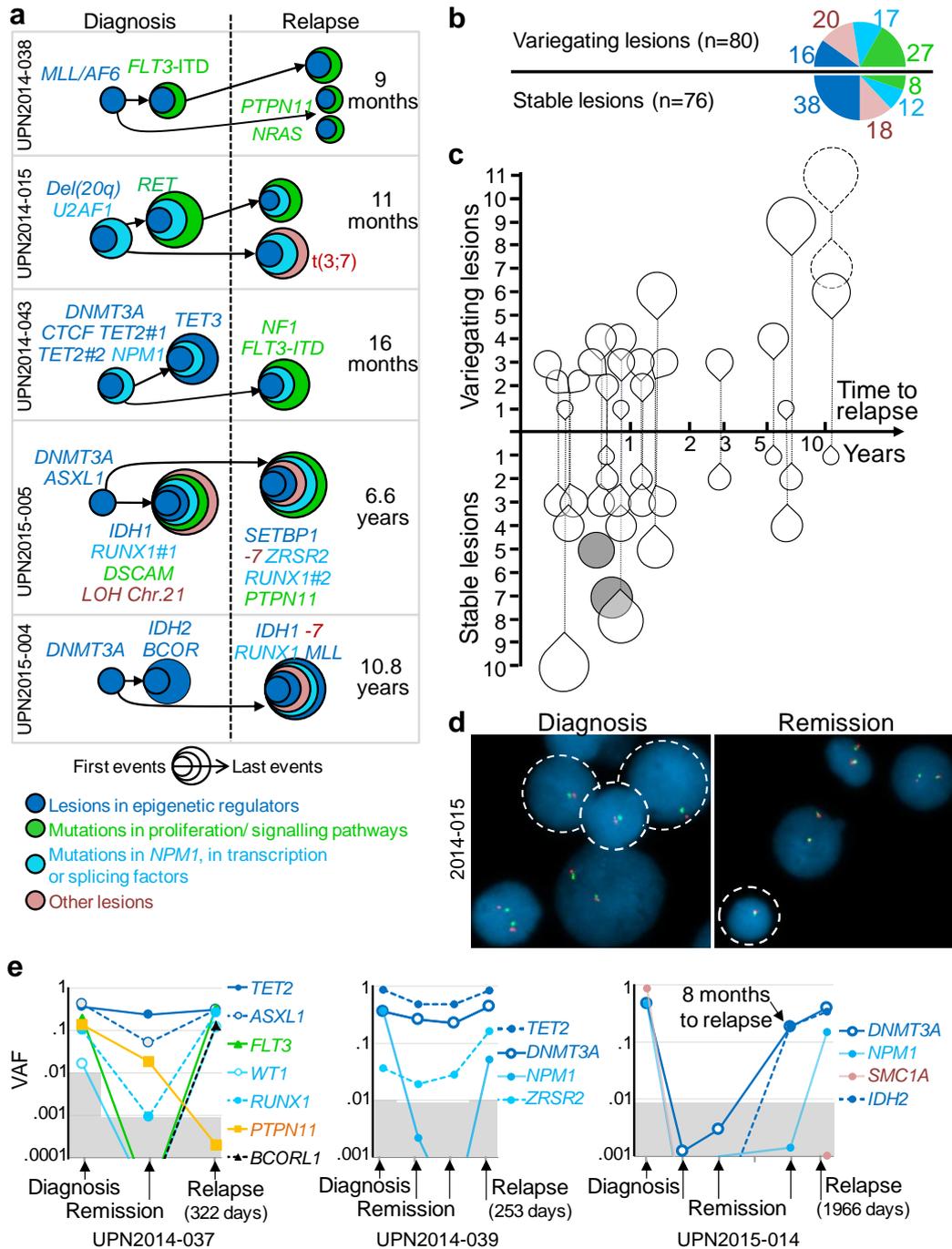


Figure 2 | Retention of early lesions and temporal variegation govern AML relapse. a, Clonal composition of 6 AMLs at diagnosis and relapse. Internally tangent circles represent successive lesions. Delay from diagnosis to relapse is indicated. **b**, Numbers of stable and variating lesions between diagnosis and relapse samples in 22 AMLs. Colours are as in **a**. **c**, Changes in clonal composition as a function of time to relapse. Droplets above and below the x-axis represent the number of variating and stable lesions, respectively. Paired droplets of single patients are connected by vertical lines. Droplets outlined by dashed lines indicate lesions from one patient whose relapse lesions (top droplet) were all different from diagnostic ones (bottom droplet). Grey circles indicate cases with no changes in clonal composition. **d**, FISH analysis of diagnosis and remission samples from a patient with persistence of del(20q). Presence of one instead of two red+green signals indicates del(20q)(outlined nuclei). **e**, Kinetics of variant allele frequencies (VAFs) in three patients. Shaded areas indicate the maximum background error rates of the NGS experiments for single nucleotide variants (SNVs).

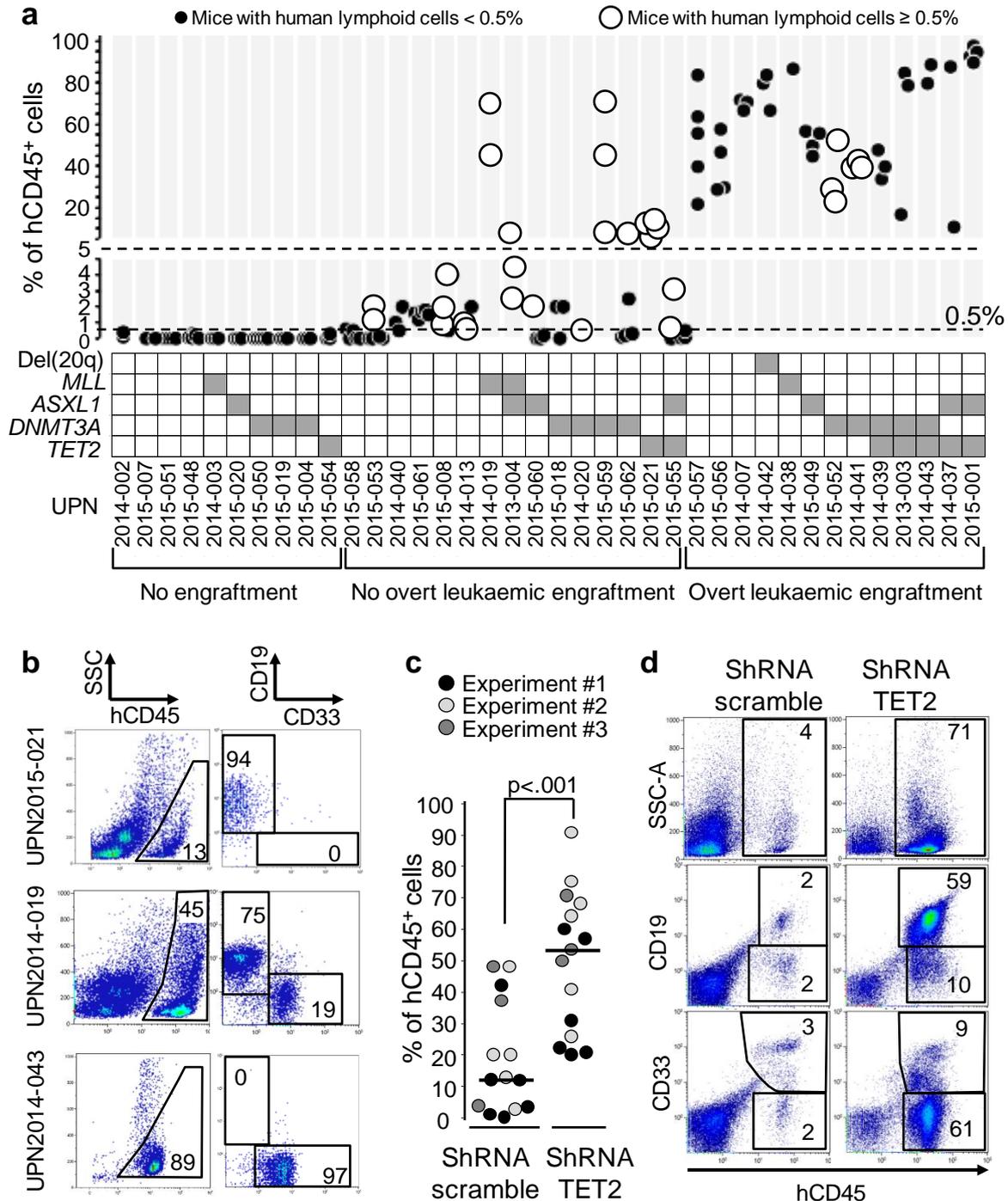


Figure 3 | AMLs with early epigenetic lesions have leukaemic and non-leukaemic repopulation capacities in NSG mice. a, Percentages of human CD45⁺ (hCD45⁺) cells in the bone marrow of NSG mice 8 to 43 weeks after injection of 5.10⁶ mononuclear cells from 38 AML blood samples. The presence of candidate pre-leukaemic lesions in injected cells is indicated in the co-mutation table. **b,** Flow cytometric analysis of NSG bone marrow repopulated with non leukaemic (UPN2014-019, UPN2015-021) and leukaemic (UPN2014-043) cells. **c,** Repopulation of NSG bone marrow by shRNA TET2 or scramble transduced cord blood CD34⁺ cells (Mann Whitney test). **d,** Flow cytometric analysis of the bone marrow from two representative mice transplanted with control (shRNA scramble) and TET2 knocked-down (shRNA TET2) cells.

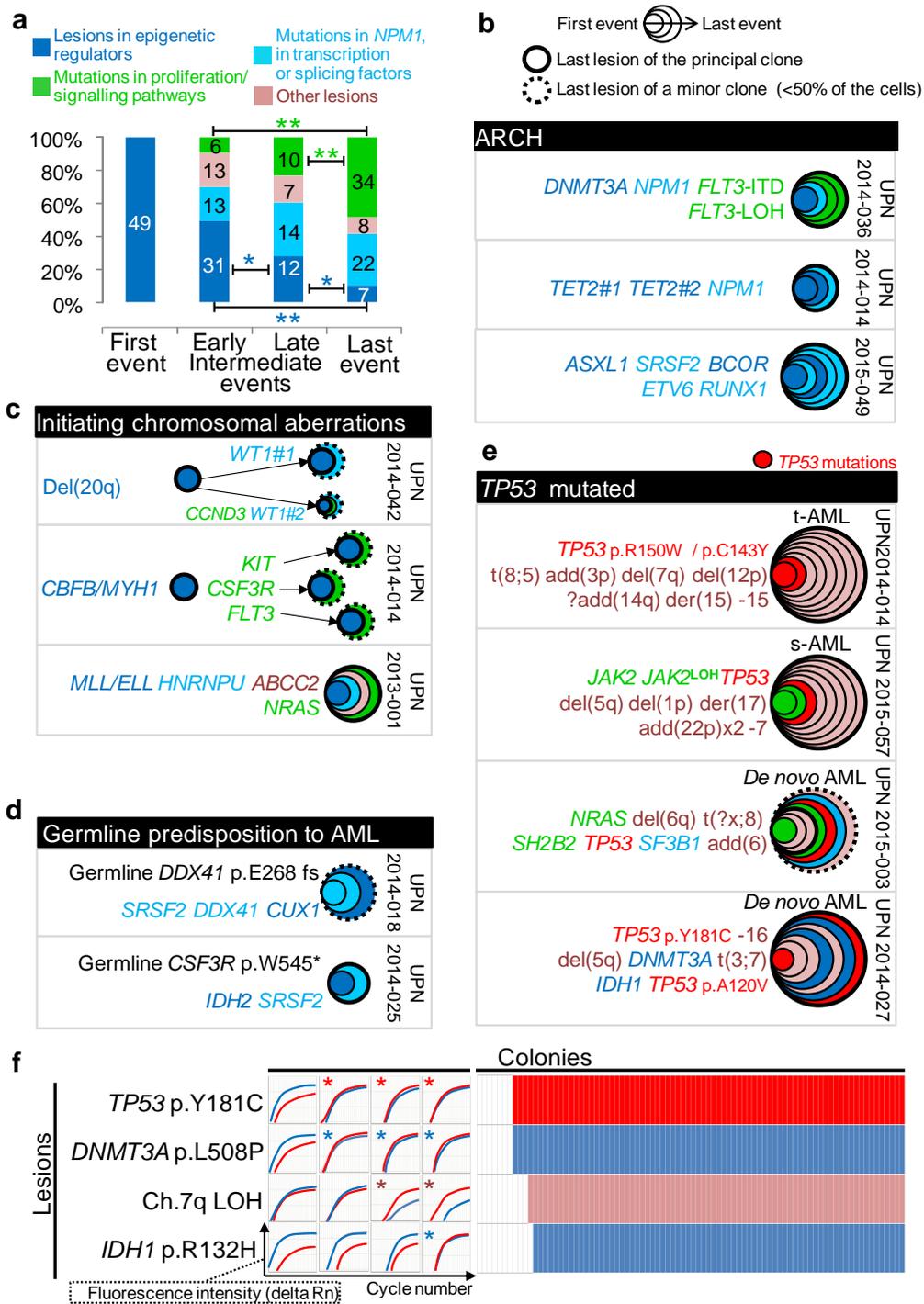


Figure 4 | Founding lesions define distinct genetic hierarchies in AML. **a**, In 49 patients with lesions in epigenetic regulators first, subsequent events were classified as early intermediate (second to median), late intermediate (median+1 to penultimate), and last event. Histograms show the distribution of events at each chronological position. * and ** indicate $p < .05$ and $p < .01$, respectively, Fisher's exact test. **b**, **c**, **d**, **e**, Clonal composition of representative AMLs with distinct genetic hierarchies: ARCH related AMLs (**b**), AMLs with initiating chromosomal abnormalities (**c**), AMLs with germline predisposition (**d**), AMLs with somatic *TP53* mutations (**e**), t-AML: therapy-related AML; s-AML: secondary AML. Internally tangent circles represent successive events. Colours are as in **a**. **f**, Analysis of 84 colonies from patient UPN2014-027 as in Fig. 1a. Colours are as in **a** and **e**.

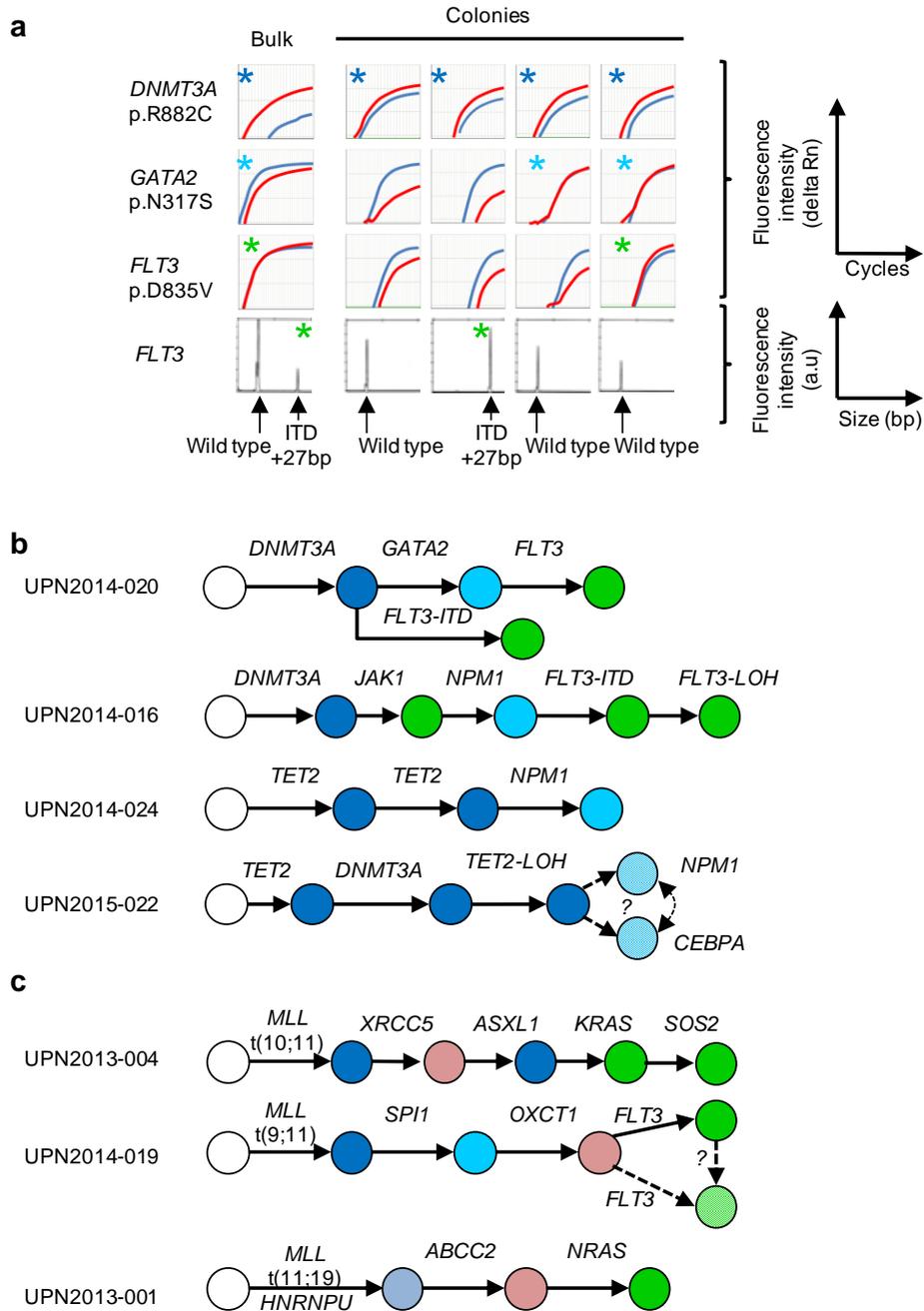
Extended data

Extended Data Table 1 | Classification of chromosomal and genetic events in four categories

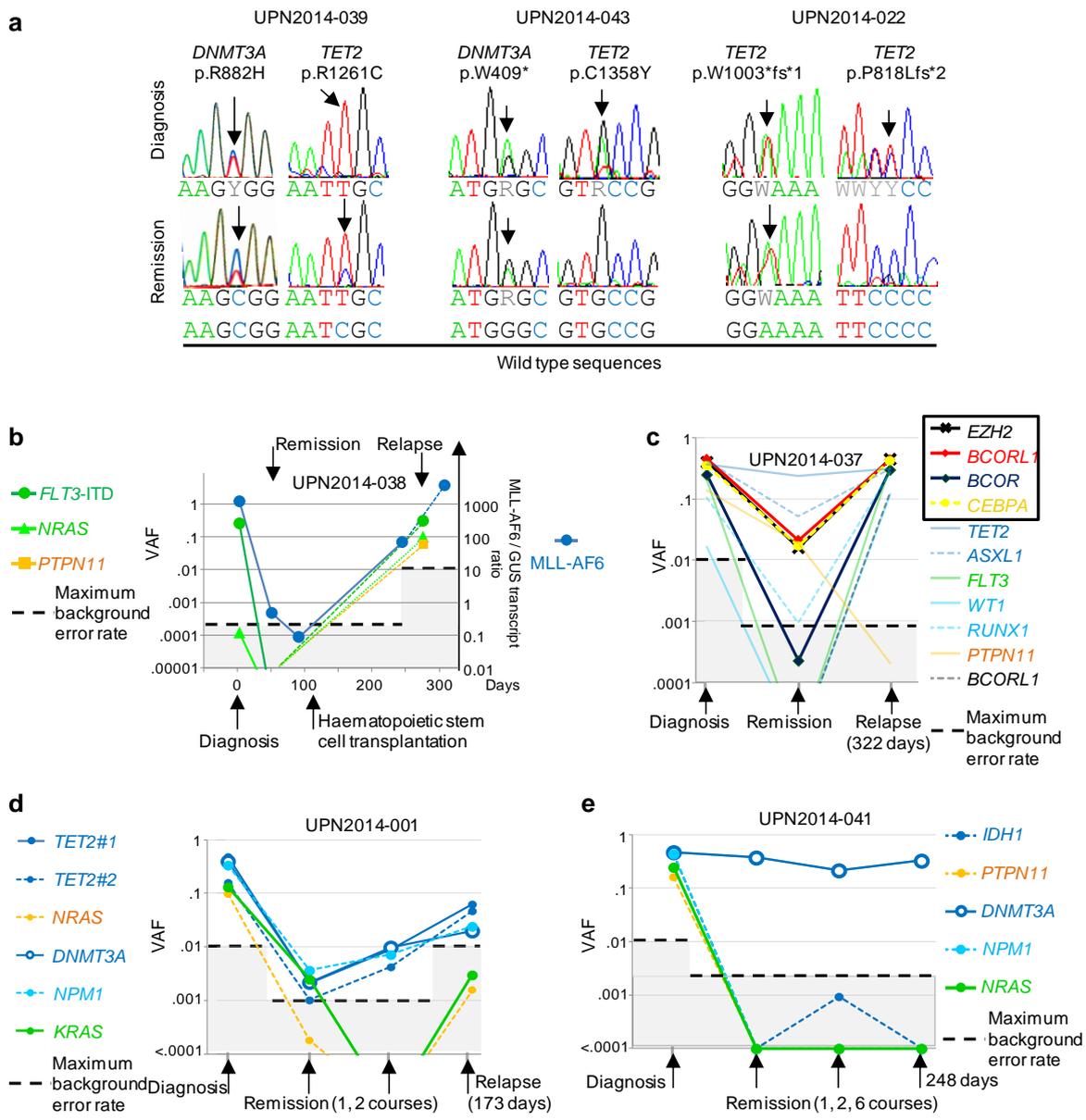
Lesions in epigenetic regulators	Mutations in <i>NPM1</i> , transcription factors, and splicing factors	Proliferative mutations	Other events
<i>ASXL1</i>	<i>CEBPA</i>	<i>BCLAF1</i>	<i>ABCC2</i>
<i>BCOR</i>	<i>DDX41</i>	<i>CBL</i>	<i>ATM</i>
<i>BCORL1</i>	<i>ETV6</i>	<i>CCND3</i>	<i>DDX1</i>
<i>CTCF</i>	<i>GATA2</i>	<i>CDC27</i>	<i>FBXW7</i>
<i>CUX1</i>	<i>HNRNPU</i>	<i>CSF3R</i>	<i>JAG1</i>
<i>DNMT3A</i>	<i>IKZF1</i>	<i>DSCAM</i>	<i>NOTCH1</i>
<i>EZH2</i>	<i>NPM1</i>	<i>FLT3</i>	<i>NUMA1</i>
<i>IDH1</i>	<i>RUNX1</i>	<i>JAK1</i>	<i>OXCT1</i>
<i>IDH2</i>	<i>SF3B1</i>	<i>JAK2</i>	<i>PDS5B</i>
<i>KDM6A</i>	<i>SPI1</i>	<i>KIT</i>	<i>PHF6</i>
<i>MLL</i>	<i>SRSF2</i>	<i>KRAS</i>	<i>RB1</i>
<i>MLL2</i>	<i>U2AF1</i>	<i>NF1</i>	<i>SMC1A</i>
<i>NSD1</i>	<i>WT1</i>	<i>NRAS</i>	<i>STAG2</i>
<i>SETBP1</i>	<i>ZRSR2</i>	<i>PTPN11</i>	<i>TP53</i>
<i>TET2</i>		<i>RALBP1</i>	<i>XRCC5</i>
<i>TET3</i>		<i>RET</i>	Any other cytogenetic abnormality
del(20q) ¹⁴		<i>SH2B2</i>	
t(11;v) involving <i>MLL</i>		<i>SOS2</i>	
inv(16) ¹¹⁻¹³			
t(8;21) ¹¹⁻¹³			

Extended Data Table 2 | Temporal evolution of AML clones at relapse

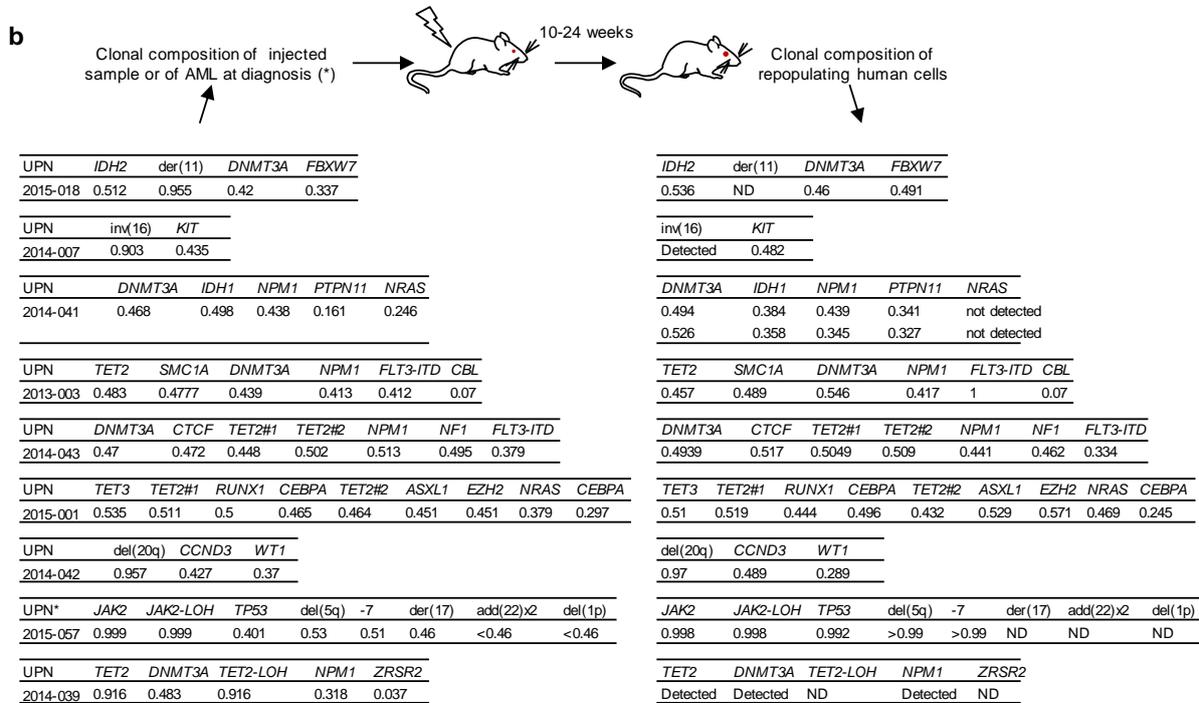
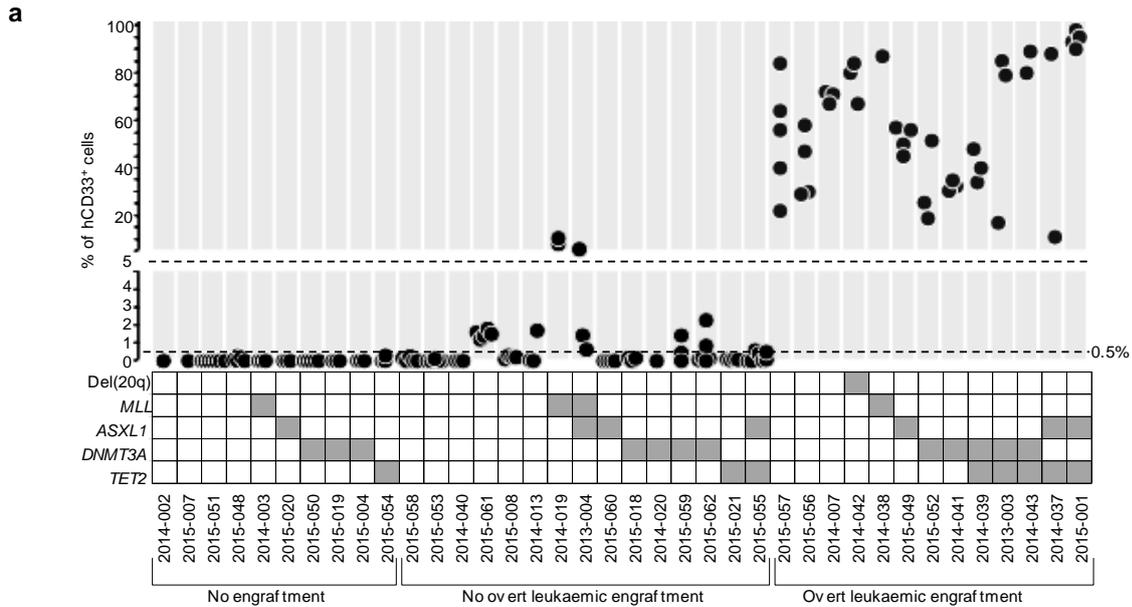
UPN	Delay to relapse (days)	Stable lesions	Gained lesions	Lost lesions	Quantitative changes (>5 fold) or switch in dominant clone
2014-033	162	<i>DNMT3A</i> p.R882C, <i>RUNX1</i> p.S386fs, <i>TET2</i> p.G1137D	<i>FLT3-ITD</i> 24bp	<i>FLT3-ITD</i> 21bp <i>FLT3-ITD</i> 78 bp	
2015-047	164	<i>TP53</i> p.H179R, del(5q), -7, -8, -17, -18, add(11p), add(5q), ?20, add(3q)	add(19q)		
2014-001	173	<i>TET2</i> p.Y1294C, <i>DNMT3A</i> p. R882H, <i>NPM1</i> p.L258fs, <i>TET2</i> p.L627fs			<i>NRAS</i> p.Q61K↘ <i>KRAS</i> p.Q61H↘
2014-017	175	<i>IDH1</i> p.R132C, <i>DNMT3A</i> p.R736C, <i>ATM</i> p.P2029R		<i>KDM6A</i> p.V455fs <i>NPM1</i> p. L258fs	
2014-016	240	<i>DNMT3A</i> p.L773H, <i>JAK1</i> p.L1053V, <i>NPM1</i> p. L258fs, <i>FLT3-ITD</i> 63bp, <i>FLT3-LOH</i>			
2014-039	253	<i>TET2</i> p.R1261C, <i>TET2-LOH</i> , <i>DNMT3A</i> p.R882H	<i>MLL2</i> p.R2734*, Complex Chr.X,4,8 rearrangement		<i>NPM1</i> p.L258fs ↘ <i>ZRSR2</i> p.E133V↗
2014-042	272	del(20q)		<i>WT1</i> p.A153fs	<i>WT1</i> p.P148fs↗ <i>CCND3</i> p.T285fs↗
2014-038	273	t(6;11), <i>FLT3-ITD</i>	<i>NRAS</i> p.Q61K, <i>PTPN11</i> p.A72T		
2014-027	287	<i>TP53</i> p.Y220C, <i>DNMT3A</i> p.L508P, <i>IDH1</i> p.R132H, <i>TP53</i> p.A159V, del(5q), -16, t(3;7)			
2014-029	321	<i>DNMT3A</i> p.R882H, <i>SRSF2</i> p.P95T, <i>ASXL1</i> p.H630fs, <i>IDH1</i> p.R132C	<i>RUNX1</i> p.R319fs, t(1;6)		<i>IDH2</i> p.R140Q↗
2014-037	322	<i>TET2</i> p.A1605S, <i>EZH2</i> p.K629M, <i>BCOR</i> p.R810*, <i>ASXL1</i> p.H630fs, <i>CEBPA</i> p.43fs, <i>BCORL1</i> p.R1196*, <i>RUNX1</i> p.V137fs, <i>FLT3-ITD</i>	<i>BCORL1</i> p.R1420*, t(6;19;16)	<i>PTPN11</i> p.Q510H	<i>WT1</i> p.Y510*↗
2014-015	322	del(20q), <i>U2AF1</i> p.S34F, <i>RET</i> p.G733D	t(1;3)		
2013-006	413	<i>CCND3</i> p.R271fs, <i>WT1</i> p.R141fs	<i>FLT3-ITD</i> 39bp, <i>FLT3-ITD</i> 66bp	<i>FLT3-ITD</i> 129bp	
2014-009	414	<i>PTPN11</i> p.F285S, <i>SF3B1</i> p.K666N, -7	del(7q)	<i>FLT3</i> p.D835Y	
2014-043	491	<i>DNMT3A</i> p.W409*, <i>TET2</i> p.A1779fs, <i>TET2</i> p.C1358Y, <i>CTCF</i> p.R377H, <i>NPM1</i> p.W259fs	<i>NF1</i> p. L300fs <i>FLT3-ITD</i> 45bp	<i>TET3</i> p.R899*	
2013-002	497	<i>ASXL1</i> p.Q373*, <i>EZH2</i> p.C504G, <i>EZH2</i> p. R249*	<i>SETBP1</i> p.G870S, <i>JAK2</i> p.V617F, <i>FLT3-ITD</i> , t(11;21)	<i>FLT3</i> p.D839E, <i>RUNX1</i> p.D66fs	
2014-021	1038	t(8;21), <i>TET2</i> p.Y819fs		<i>FLT3</i> p.Q580p, del(11p)	-7 ↗
2015-014	1966	<i>DNMT3A</i> p.VG567G	<i>IDH2</i> p.R140L, <i>NPM1</i> p.L258fs	<i>NPM1</i> p.L258fs, <i>SMC1A</i> p. R790W	
2015-006	2312	<i>IDH2</i> p.R140Q, <i>NOTCH1</i> p.V1676I, <i>SRSF2</i> p.95T, <i>NPM1</i> p.L258fs		+Y	
2015-005	2411	<i>DNMT3A</i> p.R882H, <i>ASXL1</i> p.Q882*	<i>SETBP1</i> p.D868N, <i>ZRSR2</i> p.V304M, <i>RUNX1</i> p.R174*, <i>PTPN11</i> p.A72V, -7	<i>DSCAM</i> p.E1836*, LOH chr(21) <i>RUNX1</i> p.A33fs, <i>IDH1</i> p.R132C	
2015-003	3910	none	<i>CBL</i> p.L380P, <i>ASXL1</i> p.G643fs, <i>TET3</i> p.P175A, <i>WT1</i> p.R140*, <i>SETBP1</i> p.G870S, <i>JAK1</i> p.L4V, <i>IKZF1</i> p.G103A, <i>FLT3</i> p.D835V, <i>FLT3</i> p.E444Q, del(7p), del(7q),	<i>SF3B1</i> p.-705S, <i>SH2B2</i> p.A392G, <i>TP53</i> p.R273H, <i>NRAS</i> p.G13D, add(6), t(?X;8), del(6q)	
2015-004	3939	<i>DNMT3A</i> p.V716I	<i>MLL</i> p.M2850I, <i>IDH1</i> p.R132H, -7, <i>RUNX1</i> p.R135K	<i>IDH2</i> p.R172K, <i>BCOR</i> p.N575fs	



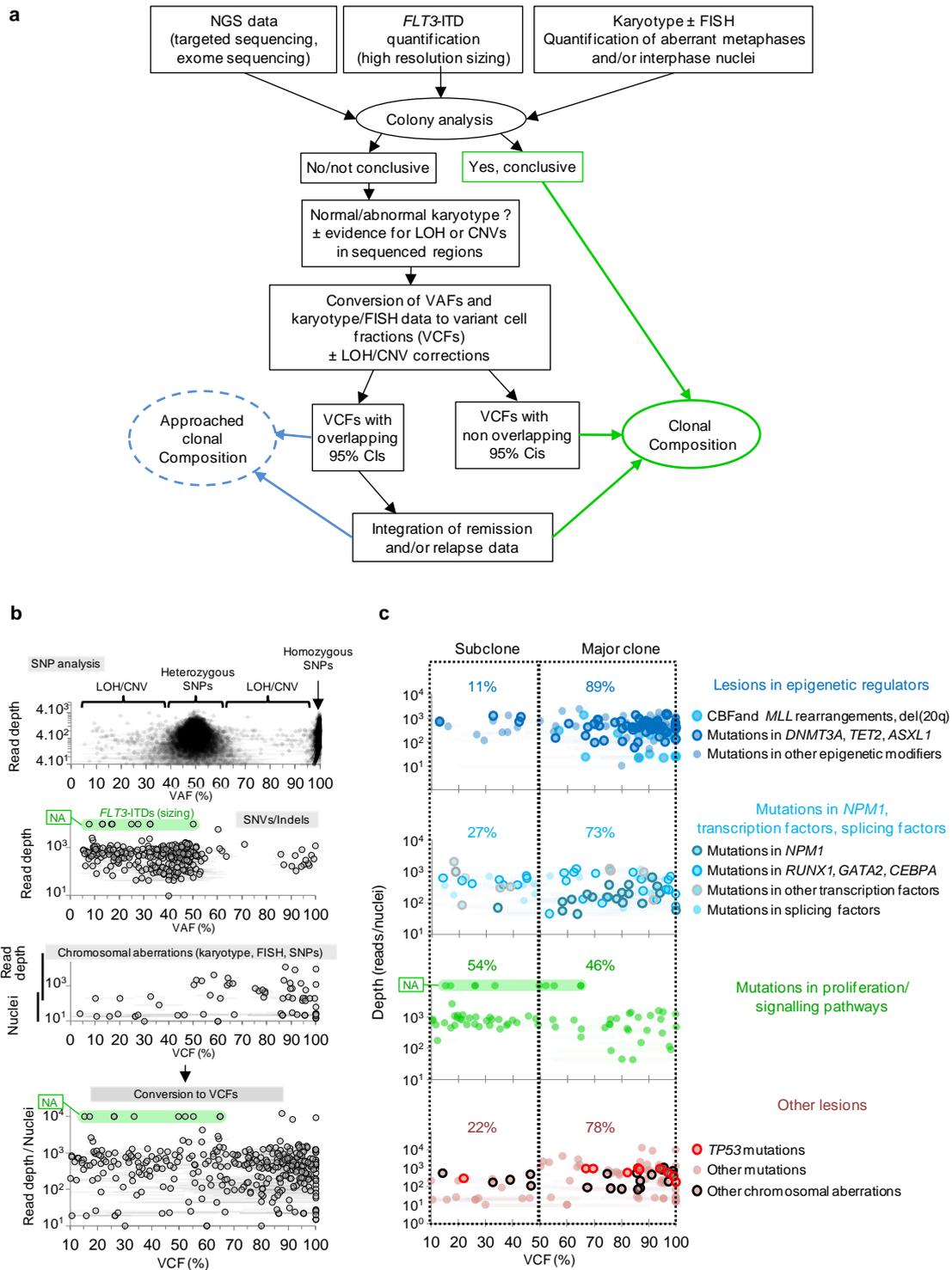
Extended Data Figure 1 | Single-cell derived colonies analyses in seven additional patients confirm a recurrent order of chromosomal and genetic lesions acquisition. a, Allele specific PCR amplification plots, as in figure 1a, and traces of high resolution sizing of fluorescent PCR amplicons of *FLT3* in bulk AML and four colonies from patient UPN2014-020. Length of *FLT3* amplicons is 414 bp when wild-type, and 441 bp in case of ITD. **b**, Mutation order in four patients with normal karyotype AML as in figure 1b. **c**, Mutation and chromosomal translocation order in three patients with *MLL* rearrangements. Colour code is as in Fig.1. Light blue represents co-occurrence of epigenetic and splicing lesions.



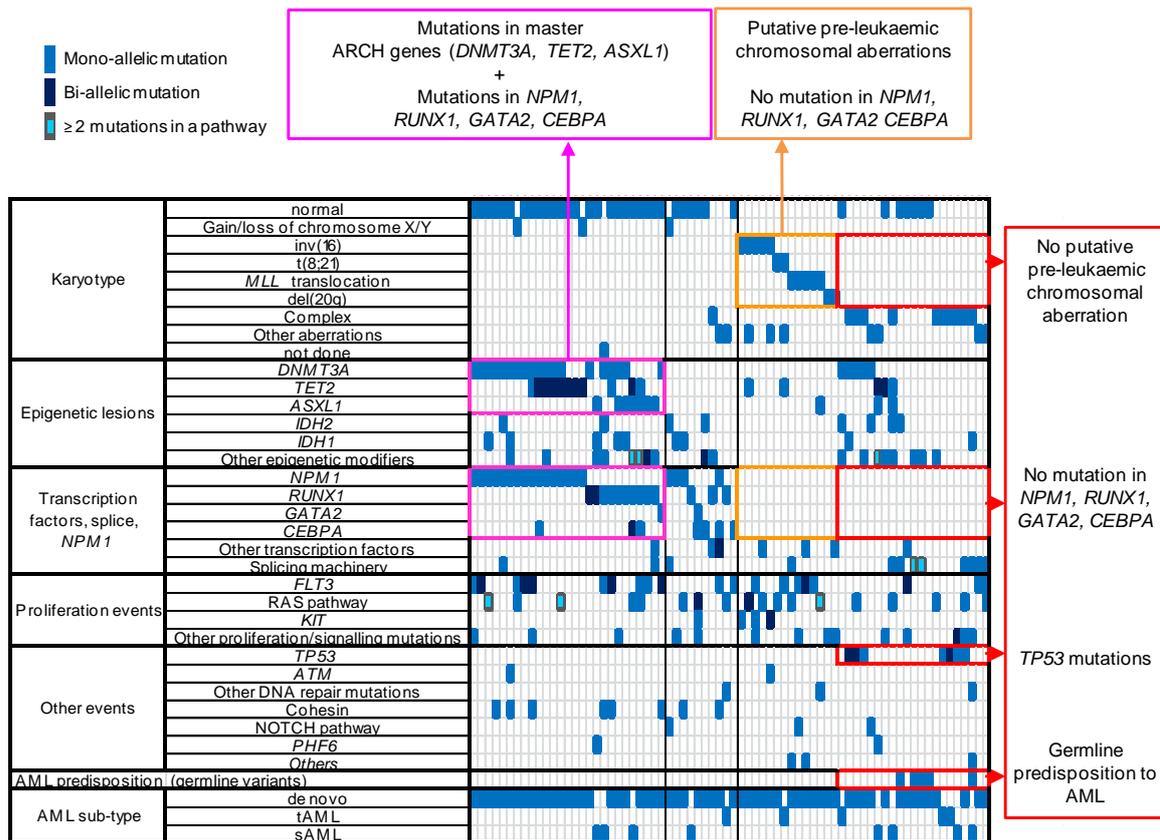
Extended Data Figure 2 | Retention of lesions in remission samples defines reservoirs for persistent clonal haematopoiesis and relapse a, Sequence traces of paired diagnostic and remission samples from three AML patients with *DNMT3A* and *TET2* mutations. Arrows indicate the detection of mutant alleles. **b**, Kinetics of VAFs, determined by targeted or deep sequencing, and MLL-AF6 transcript quantification in patient UPN2014-038. Black dashed lines over shaded areas delimit the maximum background error rate for SNV detection of patient specific mutations by deep sequencing, and the 0.01 limit of detection for the gene panel targeted sequencing. The limit of detection of MLL-AF6 transcript is 0.01. Different time points of follow-up are marked by arrows. **c**, Kinetics of VAFs in patient UPN2014-037, as in **b**. VAFs of *EZH2*, *BCORL1*, *BCOR*, and *CEBPA* variants, not shown in **Fig. 2e**, are highlighted. **d**, **e**, Kinetics of VAFs in patients UPN2014-001 and UPN2014-041, as in **b**.



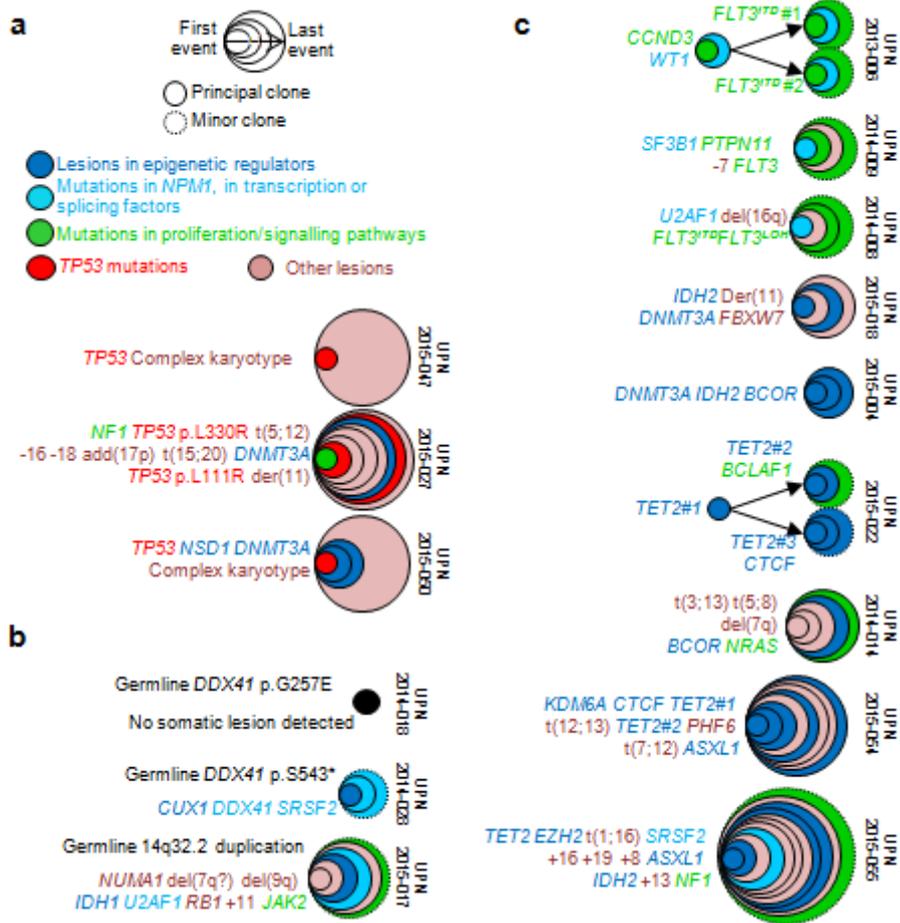
Extended Data Figure 3 | Clonal composition of NSG repopulating leukaemia mirrors that of injected samples. **a**, Percentages of human CD45⁺ CD33⁺ (hCD33⁺) cells in the bone marrow of NSG mice 8 to 43 weeks after injection of 5.10⁶ mononuclear cells from 38 AML blood samples. The co-mutation table is as in Figure 3a. **b**, Clonal composition of injected samples (left panel) and repopulating human cells after engraftment (right panel) in nine patients with overt leukaemic engraftment. VAFs are indicated for mutations. *FLT3-ITD* was detected by high resolution sizing of PCR products. Chromosomal aberrations were quantified using karyotype or interphasic FISH results and conversion of VAFs to cell fractions when informative SNPs were available. In post engraftment sample, Inv(16) was studied using RT-PCR for the CBFβ-MYH11 transcript. *, For patient UPN2015-057, cells from relapse AML were injected, but only cells from diagnosis were available for clonal composition studies.



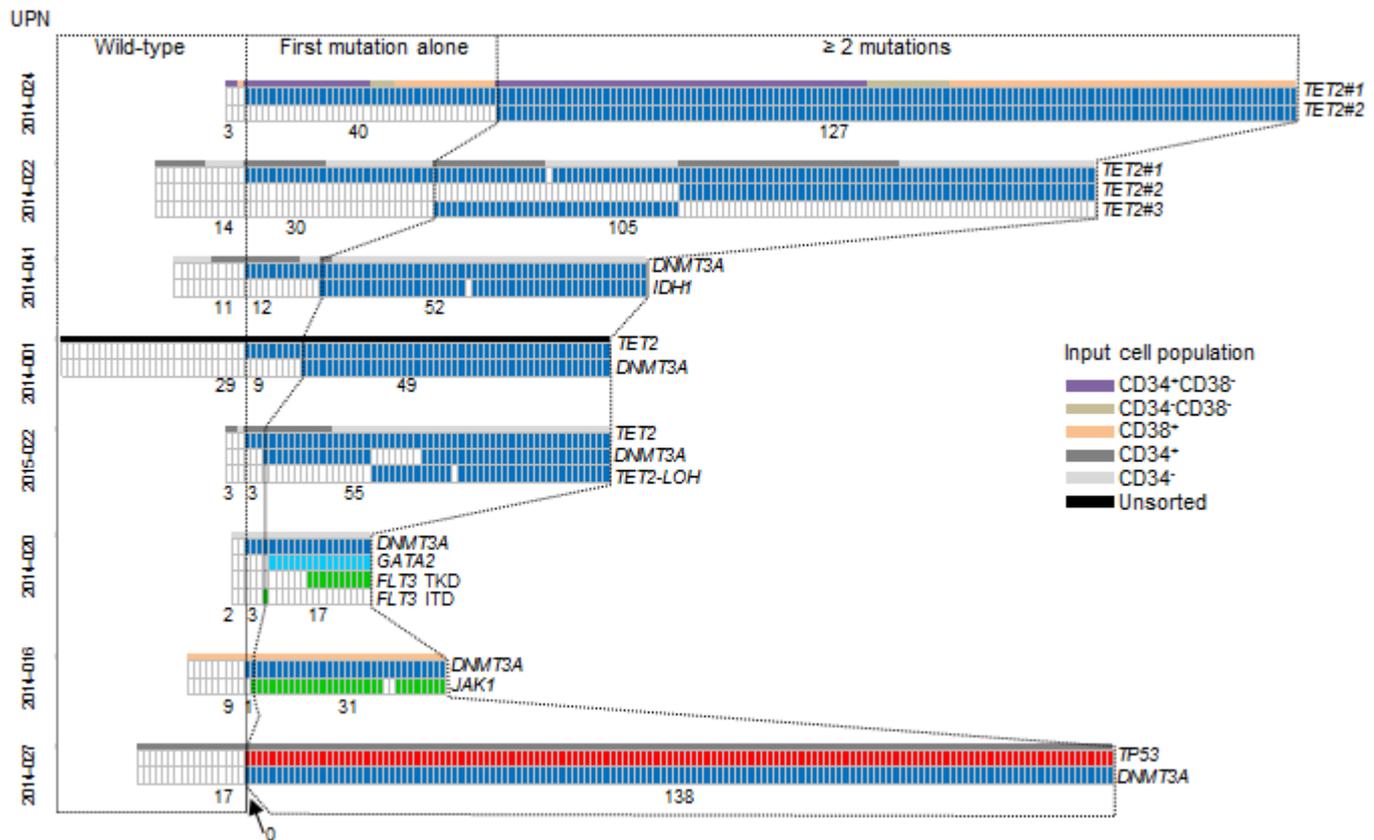
Extended Data Figure 4 | Multiparametric determination of clonal composition. **a**, Flow chart for the determination of the order of events in AML. **b**, Allele frequencies of SNPs, SNVs, and indels are plotted against read depth. Chromosomal aberrations were quantified by karyotypic and FISH analyses. When available, the quantification was inferred from VAFs of SNPs in imbalanced regions (LOH, CNVs). The values are plotted against the number of analysed nuclei (karyotype/FISH) or against the cumulated read depth of the SNPs used for quantification. Quantitative data were then converted to variant cell fractions (VCFs) for all detected chromosomal and genomic lesions above a threshold of 10% of cells. **c**, VCFs according to distinct categories of lesions. Some *FLT3*-ITDs, only detected by the sizing technique, are plotted within green areas. Shaded error bars span 95% confidence intervals.



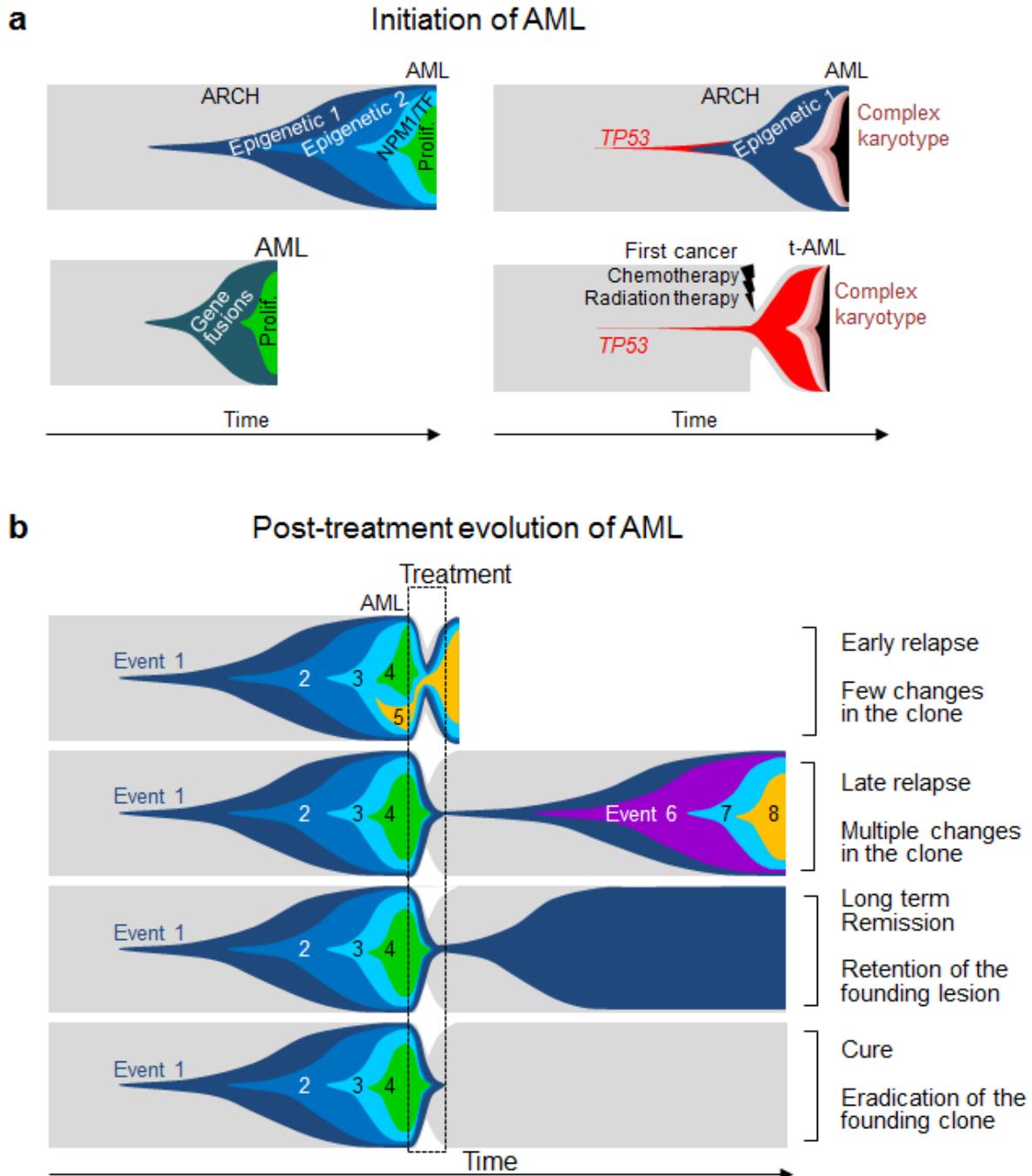
Extended Data Figure 5 | Genetic hierarchy in AML as a reflection of clonal history. Co-mutation table of 72 AMLs at diagnosis. Groups of patients with distinct genetic hierarchies were defined according to the co-occurrence or exclusion of mutations in the three master genes involved in ARCH - *DNMT3A*, *TET2*, *ASXL1* -, mutations in *NPM1* and in haematopoietic transcription factors - *RUNX1*, *GATA2*, *CEBPA* -, and pre-leukaemic chromosomal aberrations - *MLL* and CBF rearrangements, del(20q) -.



Extended Data Figure 6 | Clonal composition of 15/21 AMLs with no putative pre-leukaemic chromosomal aberration and no mutations in *NPM1*, *RUNX1*, *GATA2*, or *CEBPA*. **a**, AMLs with *TP53* mutations. **b**, AMLs with germline predispositions. **c**, Other cases. Internally tangent circles represent successive events. For patients UPN2015-047 and UPN2015-050 the chromosomal aberrations from complex karyotype were merged.



Extended Data Figure 7 | Capture of early clonal expansion steps by single cell derived colony analysis. Panels show the mutational pattern of individual colonies from patients with early initiating lesions as in Fig.1. Empty and colored boxes represent wild-type and mutated colonies, respectively. Cell sorting strategies before culture are indicated above each table. Dotted lines separate wild-type colonies, colonies with one mutation, and colonies with two or more mutations. The number of colonies of each category is indicated below the tables.



Extended Data Figure 8 | The action of time and treatment selection pressure in the initiation and evolution of AML. **a**, Schematic representation of the Initiation and evolution of the malignant clone (coloured areas) at the expense of normal HSPCs (grey area) from the acquisition of the first event to the onset of AML. Left diagrams show AMLs emerging after the occurrence of the first lesions which endow HSPCs with enhanced expansion capacity. Right diagrams depict AMLs with *TP53* mutations. For therapy-related AMLs (t-AML), the occurrence and treatment of a first cancer, followed by HSPC depletion and recovery, are indicated. The nature of each lesion is indicated as follows: Epigenetic 1: mutations in *DNMT3A*, *TET2*, *ASXL1*; Epigenetic 2: Epigenetic 1 or other mutations in epigenetic modifiers; Gene fusions: CBF or *MLL* translocations; *NPM1/TF*: mutations in *NPM1* or haematopoietic transcription factors; Prolif.: mutation in proliferation/signalling pathways; *TP53*: *TP53* mutations. **b**, Post-treatment evolution of AML, according to various outcomes. ARCH: age related clonal haematopoiesis

Article 1 : Données supplémentaires

Supplementary tables

Supplementary Table 1 | Summary of patients' characteristics and clinical history

Supplementary Table 2 | list of the 122 genes included in the targeted re-sequencing panel

Supplementary Table 3 | List of somatic variants detected with the targeted resequencing panel at diagnosis.

Supplementary Table 4 | list of somatic variants detected by exome sequencing

Supplementary Table 5 | List of somatic variants detected with the targeted resequencing panel at relapse.

Supplementary Table 6 | deep sequencing analyses in eleven patients at diagnosis relapse or in complete remission at different steps of follow-up

Supplementary Table 7 | Analysis of NSG bone marrow repopulation by primary AML cells.

Supplementary Table 8 | Order of lesion acquisition in 49 patients with lesions in epigenetic regulators in first position

Supplementary Table 9 | Order of lesion acquisition in 25 patients with no evidence for lesions in epigenetic regulators in first position

Supplementary Table 10 | Primers used for Sanger sequencing and FLT3-ITD detection (5'-3')

Supplementary Table 11 | Primers and probes for allele specific PCR assays

Supplementary Table 12 | Antibodies used for flow cytometry analyses

Supplementary data

Supplementary figure 1 | Summary of single cell derived colony assays in seven patients with normal karyotype

Supplementary figure 2 | Summary of single cell derived colony assays in six patients with abnormal karyotype

Supplementary Figure 3 | Temporal evolution of AML clones at relapse

Supplementary Figure 4 | Leukaemic and non leukaemic repopulation of NSG bone marrow by AML cells

Supplementary Figure 5 | TET2 depletion enhances the repopulation of NSG mouse bone marrow by human cord blood CD34⁺ cells

Supplementary Table 1 | Summary of patients' characteristics and clinical

UPN	characteristics at diagnosis										Treatment				Characteristics at relapse				Last follow-up time from diagnosis (days)	alive at last follow-up	
	Gender	Age	Prior myeloid malignancy	Prior exposure to cytotoxic chemotherapy or radiation therapy	WBC (x 10 ⁹ /L)	FAB	WHO	% Blast cells in BM	Karyotype	FLT3-ITD (stng)	CEBPA mutation (RNA and Sanger)	Intensive chemotherapy	Allogeneic BMT in first CR	Relapse ?	Time to relapse (days)	% Blast cells in BM	Karyotype	FLT3-ITD (stng)			CEBPA mutation (RNA and Sanger)
2013-001	M	31	No	No	17.6	M2	AML-NOS	43	46,XY,t(11;19)(p23;p13)[14] ish(11;19)(p19)(MLL+3MLL-3MLL+)[3]nuc is(MLLx2,5MLL sep-3,MLLx1)[19/200]	WT	WT	yes	no							671	yes
2013-002	M	74	No	No	3.57	M2	AML-NOS	44	46,XY	WT	WT	yes	yes	487	16		MUT (mut=2)	WT	WT	559	yes
2013-003	F	57	No	No	61	M1	AML with NPM1 mutation	81	46,XX	MUT (mut=0)	WT	yes	no							736	yes
2013-004	F	53	No	No	71	M5a	AML-NOS	94	46,XX,del(7)(q10)(q21;p22),der(10)(t(10;11)(q27;p12),q23inv(11)(q13q23),der(1)(t(1)(q11)(p12q23)[19]/46,XX[3]	WT	WT	yes	no	no (died from sepsis in CR1, 108 days after relapse)						108	no
2013-006	F	18	No	No	12.2	M5a	AML-NOS	85	46,XX	MUT	WT	yes	yes (post BMT)	413	93		46,XX	MUT	WT	869	yes
2014-001	M	65	No	No	30	M5b	AML with NPM1 mutation	53	46,XY	WT	WT	yes	no	173	11		46,XY	WT	WT	763	yes
2014-002	F	30	No	No	14.6	M2	AML with t(8;21)	35	46,XX,(8;21)(q22;q22)[19]/46,XX[3]	WT	WT	yes	no							428	yes
2014-003	M	58	No	No	53	M4	AML-NOS	61	46,XY,t(11;19)(p23;p13)[21]	WT	WT	yes	no							743	yes
2014-004	M	74	No	No	3.11	M2	AML with NPM1 mutation	24	46,XY	WT	WT	yes	no							953	yes
2014-006	M	71	No	No	6.7	M1	AML with NPM1 mutation	84	46,XY	WT	WT	yes	no							665	yes
2014-007	M	54	No	No	16.5	M4	AML with inv(16)	49	46,XY,inv(16)(p13;q23)[28]/46,XY[3]	WT	WT	yes	no							702	yes
2014-008	M	56	No	No	29	M5a	AML-NOS	89	46,XY,del(16)(q12)[2]/46,XY[1]	WT	WT	yes	yes (post BMT)	414	9		46,XY,-7,-mar [cp8]/46,XY[8]	WT	WT	606	yes
2014-009	M	38	No	No	3.58	M4	AML-NOS	38	45,XY,-7[9]/46,XY[22]	WT	WT	yes	no							566	yes
2014-010	M	68	Yes (MDS)	No	9.54	M2	sAML	30	46,XY	MUT	WT	yes	no							532	yes
2014-013	F	48	No	No	1.17	M1	AML with NPM1 mutation	62	47,XX,+8[18]/46,XX[1]	MUT (mut=WT=0,33)	WT	yes	yes							559	yes
2014-014	F	56	No	Yes (Chemotherapy and radiation therapy for breast cancer)	8.26	M2	tAML	22	46,XX,add(9)(p11),add(5)(q11-12),der(5)(t(5;9)(q11-12;q21),der(7)(q21;q36),add(9)(q11),del(12)(p11),add(19)(q24),-15,-mar [del(15)(p13q24)]	WT	WT	yes	yes (post BMT)	194	ND (hemodilution)		47,X,add(X)(q21),del(1)(q22;p376),add(9)(p11),add(9)(q11-12),del(7)(q21;q36),add(19)(q24),-15,-21,-3mar [2]/46,XX[2]	WT	WT	231	no
2014-015	M	74	No	No	1.6	M2	AML-NOS	21	46,XY,del(20)(q11q13)[2]/46,XY[7]	WT	WT	yes	no	322	12		46,XY,del(20)(q11q13)[1]/46,XY[2]	WT	WT	559	yes
2014-016	F	47	No	No	117	M5a	AML-NOS	97	46,XX	MUT	WT	yes	no (relapse before BMT)	240	48		46,XX	MUT	WT	564	yes
2014-017	F	63	No	No	16.4	M1	AML with NPM1 mutation	95	46,XX	WT	WT	yes	yes (after relapse)	175	9		46,XX	WT	WT	479	yes
2014-018	F	59	No	No	0.82	M1	AML-NOS	65	46,XX	WT	WT	yes	yes							520	yes
2014-019	F	50	No	No	81.4	M5a	AML with t(8;11)	95	46,XX,(8;11)(p22;q23)[8]/46,XX[15]	WT	WT	yes	no (died from sepsis in CR1, 76 days after relapse)							76	no
2014-020	M	55	No	No	29.7	M2	AML-NOS	34	46,XY	MUT	WT	yes	yes							481	yes
2014-021	F	83	No	Yes (Chemotherapy and radiation therapy for Breast Cancer)	193	M1	tAML	96	46,XX,(18;21)(q22;q22),del(11)(p11p14)[21]/46,XX[1]	WT	WT	yes	no	1038	8		46,XX,(18;21)(q22;q22)[1]/46,XX[23]	WT	WT	1517	yes
2014-022	F	81	No	No	2.6	M5/M7	AML-NOS	44	46,XY	WT	WT	yes	no							419	yes
2014-024	F	72	No	No	71.3	M4	AML with NPM1 mutation	89	46,XY	WT	WT	yes	no							336	yes
2014-025	M	75	No	No	1.62	M2	AML-NOS	24	46,XY	WT	WT	yes	no							376	yes
2014-027	F	72	No	No	4.94	M2	AML-NOS	41	46,XX,add(3)(q21),del(9)(q21),del(7)(q37)(p13-21;q21 or q31),-16 [18]	WT	WT	yes	yes (post BMT)	287	25		45,XX,der(3)(pter->3p21-22:7),del(9)(q21),del(7)(pter->7q21:3q21-22->3qter),-16 [825/46,XX[18]	WT	WT	413	yes

UPN	Gender	Age	Prior myeloid malignancy	Prior exposure to cytotoxic chemotherapy or radiation therapy	WBC (x 10 ⁹ /L)	FAB	WHO	% Blast cells in BM	Characteristics at diagnosis			Treatment			Characteristics at relapse				Last follow-up time from diagnosis (days)	alive at last follow-up		
									Karyotype	FLT3-ITD (sizing)	CEBPA mutation (HRM and sanger)	intensive chemotherapy	Allogeneic BMT in first CR	Relapse ? (donor cell leukemia)	Time to relapse (days)	% Blast cells in BM	Karyotype	FLT3-ITD (sizing)			CEBPA mutation (HRM and sanger)	
2014-028	M	53	No	No	0.95	M2	AML NOS	50	46, XY	WT	WT	yes	no (donor cell leukemia)	763	69	48, XY, +4, der(13)t(12;21)(p11.7; p11.2) [2] 46, XY [3] 46, XY [2]	WT	WT	WT	845	no	
2014-029	M	53	No	No	1.96	M2	AML NOS	37	46, XY	WT	WT	yes	yes	321	17	46, XY, t(15)(p15;q22) [4] 46, XY [2]	WT	WT	WT	415	no	
2014-031	M	65	No	No	3.85	M2	AML NOS	44	48, XY, +10, -21 [1] 48, idem, t(19)(q10) [5] 46, XY [2]	WT	WT	yes	no				WT			381	yes	
2014-032	F	66	No	No	2.38	M2	AML NOS	33	46, XX	WT	WT	yes	no				WT			382	yes	
2014-033	M	45	No	No	70.04	M1	AML NOS	87	46, XY	MUT	MUT	yes	yes	162	6	46, XY	MUT (mut/wt=3%, 2.4bp ITD)	WT	WT	WT	381	yes
2014-034	F	59	No	No	0.46	M1	AML with CEBPA mutation	78	46, XX	WT	MUT	yes	no				WT			285	no	
2014-036	M	44	No	No	39.6	M1	AML with NPM1 mutation	96	46, XY	MUT	MUT	yes	yes	284	61	46, XX, t(6;19)(q27;q11;p13) [3] 46, XX [13]	ND	ND	ND	385	yes	
2014-037	F	77	No	No	44	M5b	AML with CEBPA mutation	62	46, XX	MUT	MUT	yes	yes (post BMT)	322	86	46, XY, t(6;11)(q27;q23) [19] 46, XY [2]	MUT	MUT	MUT	357	no	
2014-038	M	31	No	No	96.4	M5a	AML NOS	88	46, XY [16] 46, XY [24]	MUT	MUT	yes	yes (post BMT)	273	80	48, XY, der(4)t(4;8)(q13;q22-23) +8, der(8)(q23) der(8)(q23) [12] 46, XY [16]	MUT	WT	WT	427	yes	
2014-039	M	66	No	No	50	M5b	AML with NPM1 mutation	74	46, XY	WT	WT	yes	no	253	25	46, XY, t(3;13)(p24;q13), t(5;8)(q14;q23) or t(5;8)(q12;q21), del(7)(q27;q19) [15]	WT	WT	WT	382	no	
2014-040	M	64	No	No	155	M0	AML NOS	89	46, XY, del(20)(q12)[7] 46, XY [20]	WT	WT	yes	no	590	ND	ND	ND	ND	ND	ND	610	yes
2014-041	M	40	No	No	90.7	M5a	AML with NPM1 mutation	70	46, XY	WT	WT	yes	no				WT			812	yes	
2014-042	M	48	No	No	2.3	M2	AML NOS	20	46, XY, del(20)(q12)[7] 46, XY [20]	WT	WT	yes	yes (post BMT)	272	65	FISH (20q12: 20q13.2), del(20q) in 201/267 nuclei	WT	WT	WT	401	no	
2014-043	F	73	No	Yes (radiation therapy for adenocarcinoma of the endometrium)	205	M1	AML NOS	93	46, XX	WT	WT	yes	no	491	90	46, XX	MUT	WT	WT	724	no	
2015-001	M	68	Yes (MDS)	Yes (Radiation therapy for prostate cancer and 5azacytidine for MDS)	8.45	M2	sAML	25	46, XY	WT	MUT	no	no (no CR)				WT			59	no	
2015-002	F	66	No	No	1.2	M4eo	AML with inv(16)	28	46, XX, inv(16)(p13q22) [26] 47, sl, +8 [1] 46, XX, del(6)(q12q24) [5] / 46, XX, -8, +17x8 [7] 10, 7q10 [5] / 46, XX, add(6)(p22) [2] / anomalies variables [3] 46, XX [4]	WT	WT	yes	no				WT			950	yes	
2015-003	F	19	No	No	3.6	M2	AML NOS	51	46, XY, inv(16)(p13q22) [16] 46, XY [8] 47, XY, +Y	WT	WT	yes	yes (CMML-2)	3910	16	46, XX, del(7)(p11) [24] / 46, idem, del(7)(q22qter) [5]	WT	WT	WT	5548	no	
2015-004	F	57	No	Yes (chemotherapy and radiotherapy for breast cancer)	0.6	M1	tAML	76	46, XX	WT	WT	yes	yes (RAEB 2)	3939	14	45, XX, -7 [24]	WT	WT	WT	4358	no	
2015-005	F	56	No	No	1.7	M0	AML NOS	60	46, XX	WT	WT	yes	no	2411	24	46, XX, -7	WT	WT	WT	3062	no	
2015-006	M	55	No	No	259	M1	AML NOS	94	47, XY, +Y	WT	WT	yes	no	2312	93	46, XY	WT	WT	WT	4054	yes	
2015-007	M	45	No	No	117	M4eo	AML with inv(16)	NA	46, XY, inv(16)(p13q22) [16] 46, XY [8] 47, XY, inv(16)(p13q22), +22 [20] 46, XY [1] ish inv(16)(p13)(MYH11+), CBFB+(q22)CBFB+, MYH11+ [5] nuc ish(MYH11x3), (CBFBx3), (MYH11 con CBFBx2) [17] 200] 46, XY [4]	WT	WT	yes	no				WT			389	yes	
2015-008	M	20	No	No	107	M4eo	AML with inv(16)	70	46, XY, 7add or del(7)(q7), del(9)(q12q33) [5] 47, idem, +t(13)(q46, XX) [4]	WT	WT	yes	no (dead at induction)				WT			2	no	
2015-014	M	52	No	No	8.35	M2	AML NOS	35	46, XY, 7add or del(7)(q7), del(9)(q12q33) [5] 47, idem, +t(13)(q46, XX) [4]	WT	WT	yes	yes	1966	25	46, XY	WT	WT	WT	2232	yes	
2015-017	F	70	Yes (MPN)	Yes (hydroxyurea during one year)	2.24	M2	sAML	36		WT	WT	yes	no				WT			235	yes	

UPN	Gender	Age	Characteristics at diagnosis				Treatment				Characteristics at relapse				Last follow-up Time from diagnosis (days)	Last follow-up Time from last diagnosis (days)						
			Prior exposure to cytotoxic chemotherapy or radiation therapy	WBC (x 10 9/L)	FAB	WHO	% Blast cells in BM	Karyotype	FLT3-ITD (sizing)	CEBPA mutation (HRM and Sanger)	Intensive chemotherapy	Allogeneic BMT in first CR	Relapse ?	Time to relapse (days)			% Blast cells in BM	Karyotype	FLT3-ITD (sizing)	CEBPA mutation (HRM and Sanger)		
2015-018	M	50	No	164	M1	AML NOS	97	47, XY, der(11)(qter->q22-23; p14- >q10) [9]/46, XY [1]	WT	yes	no	no (No CR)				WT		83	yes			
2015-019	F	90	No	32.6	NA	sAML	ND	ND	WT	no	no	no	WT			WT		13	no			
2015-020	M	62	No	2.09	M2	AML NOS	ND	46, XX, del(3)(q14) [5], delXY [12]	ND	yes	no	yes	ND			WT	WT	1329	yes			
2015-021	M	88	No	77.9	NA	AML NOS	ND	45, X, Y [20]	WT	no	no	no	WT			WT		19	yes			
2015-022	M	67	No	93.6	M1	AML with NPWT mutation	96	46, XY	WT	yes	no	no	MUT			WT		211	yes			
2015-024	M	60	No	93	M1	AML with NPWT mutation	92	46, XY	WT	yes	no	no	WT			WT		187	yes			
2015-027	M	69	No	60	M5a	AML NOS	75	42-43, XY, der(5)add(5)(p14)(5i12)(q11)(q11), der(11)del(11)(q14q22)(p11)(q23), - 12, -15, -16, add(17)(p11), -18, der(20)(15:20)(q271q173), +2mar [cp21]	WT	no	no	no (died at induction)				WT		4	no			
2015-047	M	63	No	4.09	M1	IA/M	55	44, XY, del(5)(q22q24), -7, -8, -17, -18, 720, +mar1 [?der(7)del(7)(q7)add(7)(p7)add(7)(q7)] ,-mar2 [?der(17)del(17)(p17)add(17)(q25)] ,-mar3 [?der(45, X1, add(3)(q17), del(5)(q22q24)(2), add(1)(q21) [3], -7, 8, add(1)(p13) [2], -17, -18, 720, +mar1, +mar2 [q65]	WT	yes	no	yes	164	73	43-44, XY, add(3)(q17), del(5)(q22q24), -7, -8, add(1)(q13), - 17, -18, add(19)(q13), add(20)(q172), [?der(7)del(7)(q7)add(7)(p7)add(7)(q7)] ,-mar2 [?der(17)del(17)(p17)add(17)(q25)] ,-mar3 [?der(45, X1, add(3)(q17), del(5)(q22q24)(2), add(1)(q21) [3], -7, 8, add(1)(p13) [2], -17, -18, 720, +mar1, +mar2 [q65]	WT		WT		177	yes	
2015-048	M	60	No	90	M4eo	AML with inv(16)	31	46, XY, inv(16)(p13q22) [16]/46, ?idem [3]/46, XY [2]	WT	yes	no	no (died at induction)				WT		33	no			
2015-049	M	65	No	11.4	M0	AML NOS	65	46, XY	WT	yes	no	yes	365	63	46, XY, t(9;22)(q31;q11) [5]/46, XY [13]	ND	ND	590	no			
2015-050	F	68	No	1.33	M2	AML NOS	23	56, XX, +2, add(5)(q22), +6, +8, +8, +9, +11, +add(11)(p12), +13, del(17)(p172), +19, +22 [9/55, X, addX](q27), +2, add(5)(q22), add(5)(q23), +6, +8, +8, +11, +11, +13, +19, +22 [9/57, 52, +addX](q27), +10, -22 [7/46, XX [9]	WT	no	no	no (no CR)				WT			WT		453	no
2015-051	M	42	No	69	M2	AML w/ CEBPA mutation	71	46, XY	WT	yes	yes	no				WT		2009	yes			
2015-052	F	66	No	77.5	M4	AML with NPWT mutation	40	46, XX	WT	yes	no	no (died at induction)				WT		18	no			
2015-053	F	35	No	195	M4	AML w/ CEBPA mutation	85	46, XY	MUT	yes	no	no				WT		2438	yes			
2015-054	F	79	Yes (Chloambucil for CLL; hydroxyurea and azacytidine for CMML)	5.59	M2	sAML	23	46, XX, t(12;13)(q13;q24) [9]/46, idem, t(7;12)(q22;q24) [24]	WT	no	no	no (no CR)				WT		438	no			
2015-055	M	74	Yes (MDS)	13.17	M5a	sAML	55	46, XY, der(11)(q11)(11) [2]/50, idem, +8, +13, +16, +18 [24]	WT	no	no	no (no CR)				WT		55	yes			
2015-056	F	48	No	7.2	M0	IA/M	58	45, XX, del(5)(q21-22)(35-34), -7 [2]/46, XX [5]	WT	yes	no	no (no CR)				WT		128	no			
2015-057	F	60	Yes (MDS)	13.2	M6v	sAML	60	45, XX, del(1)(p24), del(5)(q13q34), -7, der(17)(17)(q27?) [?q27?], add(22)(p11) x2, var [2]/46XX [2]	WT	yes	no	yes	119	61	ND	ND	ND	213	no			
2015-058	M	76	No	64.61	M5a	AML NOS	87	46, XY	MUT	yes	no	yes	431	ND	46, XY	ND	ND	648	no			
2015-059	M	37	No	87.47	M6b	AML with NPWT mutation	90	46, XY	MUT	yes	no	no				WT		305	yes			
2015-060	M	71	Yes (MDS)	2.01	M2	sAML	32	46, XY	WT	yes	no	no (no CR)				WT		105	no			
2015-061	M	56	No	11.18	M2	AML with NPWT mutation	32	46, XY	WT	yes	no	no				WT		1450	yes			
2015-062	F	50	No	8.95	M4	AML with NPWT mutation	70	46, X, addX(p21-23) or ?(X)(p10) [15]/46, XX [7]	WT	yes	no	no				WT		1861	yes			

Supplementary Table 2 | list of the 122 genes included in the targeted re-sequencing panel

<i>ABCB11</i>	<i>DNM2</i>	<i>MAML1</i>	<i>SH2B3</i>
<i>ACSS3</i>	<i>DNMT3A</i>	<i>MAP2K1</i>	<i>SHKBP1</i>
<i>AKAP13</i>	<i>DOK2</i>	<i>MAP2K2</i>	<i>SMC1A</i>
<i>APH1A</i>	<i>DSCAM</i>	<i>MLL</i>	<i>SMC3</i>
<i>ARHGEF2</i>	<i>EED</i>	<i>MLL2</i>	<i>SPI1</i>
<i>ASXL1</i>	<i>EGR1</i>	<i>MPL</i>	<i>SRSF2</i>
<i>ASXL2</i>	<i>ERCC2</i>	<i>MSR1</i>	<i>SSRP1</i>
<i>ATM</i>	<i>ETV6</i>	<i>MYBL2</i>	<i>STAG1</i>
<i>BCLAF1</i>	<i>EZH2</i>	<i>NCOA7</i>	<i>STAG2</i>
<i>BCOR</i>	<i>FBXW7</i>	<i>NCSTN</i>	<i>STAT3</i>
<i>BCORL1</i>	<i>FLT3</i>	<i>NF1</i>	<i>SUZ12</i>
<i>BOC</i>	<i>FOXP1</i>	<i>NMNAT2</i>	<i>TEK</i>
<i>BRAF</i>	<i>FZD1</i>	<i>NOTCH1</i>	<i>TERC</i>
<i>BRPF1</i>	<i>GATA1</i>	<i>NOTCH2</i>	<i>TERT</i>
<i>CBL</i>	<i>GATA2</i>	<i>NPM1</i>	<i>TET2</i>
<i>CBLB</i>	<i>GATAD2B</i>	<i>NRAS</i>	<i>TET3</i>
<i>CCND3</i>	<i>GBP4</i>	<i>NSD1</i>	<i>TLE4</i>
<i>CDKN2A</i>	<i>GDF5</i>	<i>NUMA1</i>	<i>TP53</i>
<i>CDKN2B</i>	<i>GLI1</i>	<i>PBRM1</i>	<i>TP73</i>
<i>CEBPA</i>	<i>HJURP</i>	<i>PDS5B</i>	<i>TYK2</i>
<i>CHEK2</i>	<i>HRAS</i>	<i>PHF6</i>	<i>U2AF1</i>
<i>CSF3R</i>	<i>IDH1</i>	<i>PTEN</i>	<i>WAC</i>
<i>CTCF</i>	<i>IDH2</i>	<i>PTPN11</i>	<i>WT1</i>
<i>CUL3</i>	<i>IKZF1</i>	<i>RAD21</i>	<i>XRCC1</i>
<i>CUX1</i>	<i>ITGAX</i>	<i>RB1</i>	<i>XRCC3</i>
<i>CUX2</i>	<i>JAK1</i>	<i>RBMX</i>	<i>ZRSR2</i>
<i>CXXC4</i>	<i>JAK2</i>	<i>RET</i>	
<i>DAAM2</i>	<i>JAK3</i>	<i>RUNX1</i>	
<i>DAXX</i>	<i>JARID2</i>	<i>SETBP1</i>	
<i>DDX1</i>	<i>KDM6A</i>	<i>SF3A1</i>	
<i>DDX41</i>	<i>KIT</i>	<i>SF3B1</i>	
<i>DIS3</i>	<i>KRAS</i>	<i>SH2B2</i>	

Supplementary Table 3 | List of somatic variants detected with the targeted resequencing panel at diagnosis. The list of all detected variants is in **Supplementary Table 3bis** (Excel file).

patient	gene	chromosome	location	reference base	altered base	frequency	AA	NM
2013-001	NRAS	1	115256528	T	A	0.391	Q61H	NM_002524
2013-002	ASXL1	20	31021118	C	T	0.355	Q373*	NM_015338
2013-002	EZH2	7	148512036	A	C	0.322	C504G	NM_152998
2013-002	EZH2	7	148523591	G	A	0.290	R249*	NM_152998
2013-002	FLT3	13	28592628	A	T	0.088	D839E	NM_004119
2013-002	RUNX1	21	36259211	T	TGTCG	0.316	D66fs	NM_001001890
2013-003	CBL	11	119149290	C	T	0.074	P433L	NM_005188
2013-003	DNMT3A	2	25457242	C	T	0.434	R882H	NM_022552
2013-003	NPM1	5	170837543	C	CTCTG	0.413	L258fs	NM_002520
2013-003	SMC1A	X	53432045	G	A	0.478	R699C	NM_006306
2013-003	TET2	4	106164916	C	T	0.483	R1262W	NM_001127208
2013-004	ASXL1	20	31022572	AGT	A	0.402	K686fs	NM_015338
2013-004	KRAS	12	25398281	C	T	0.511	G13D	NM_004985
2013-006	CCND3	6	41903745	C	CG	0.467	R271fs	NM_001136017
2013-006	WT1	11	32417941	C	CACTTTTT	0.378	R141fs	NM_000378
2014-001	DNMT3A	2	25457242	C	T	0.399	R882H	NM_022552
2014-001	KRAS	12	25380275	T	G	0.132	Q61H	NM_004985
2014-001	NPM1	5	170837543	C	CTCTG	0.337	L258fs	NM_002520
2014-001	NRAS	1	115256530	G	T	0.098	Q61K	NM_002524
2014-001	TET2	4	106156978	CT	C	0.160	L627fs	NM_001127208
2014-001	TET2	4	106180853	A	G	0.484	Y1294C	NM_001127208
2014-002	NRAS	1	115258748	C	T	0.407	G12S	NM_002524
2014-003	FLT3	13	28602329	G	A	0.150	A680V	NM_004119
2014-006	DSCAM	21	41741170	A	G	0.368	S171P	NM_001389
2014-006	IDH1	2	209113113	G	C	0.351	R132G	NM_005896
2014-006	NPM1	5	170837543	C	CTCTG	0.377	L258fs	NM_002520
2014-007	KIT	4	55599320	G	T	0.432	D812Y	NM_001093772
2014-008	U2AF1	21	44524456	G	A	0.464	S34F	NM_001025203
2014-009	FLT3	13	28592642	C	A	0.120	D835Y	NM_004119
2014-009	PTPN11	12	112915455	T	C	0.446	F285S	NM_002834
2014-009	SF3B1	2	198267359	C	A	0.402	K666N	NM_012433
2014-010	RUNX1	21	36231773	C	T	0.489	R177Q	NM_001001890
2014-013	CEBPA	19	33793199	TGCGCGGGGCC	T	0.520	p.A37fs	NM_004364
2014-013	NPM1	5	170837543	C	CTCTG	0.466	L258fs	NM_002520
2014-013	PDS5B	13	33306277	A	AC	0.164	P722fs	NM_015032
2014-013	WT1	11	32456289	G	T	0.193	Y201*	NM_000378
2014-013	WT1	11	32417911	A	ACC	0.119	S152fs	NM_000378
2014-014	TP53	17	7577094	G	A	0.334	R282W	NM_005475
2014-014	TP53	17	7577114	C	T	0.347	C276Y	NM_000546
2014-015	RET	10	43612093	G	A	0.173	G733D	NM_020630
2014-015	U2AF1	21	44524456	G	A	0.268	S34F	NM_001025203
2014-016	DNMT3A	2	25463175	A	T	0.457	L773H	NM_022552
2014-016	JAK1	1	65301882	A	C	0.430	L1053V	NM_002227
2014-016	NPM1	5	170837544	T	C	0.413	L258fs	NM_002520
2014-017	ATM	11	108186629	C	G	0.458	P2029R	NM_000051
2014-017	DNMT3A	2	25463287	G	A	0.461	R736C	NM_022552
2014-017	IDH1	2	209113113	G	A	0.483	R132C	NM_005896
2014-017	KDM6A	X	44920601	TG	T	0.409	V455fs	NM_021140
2014-017	NPM1	5	170837543	C	CTCTG	0.470	L258fs	NM_002520
2014-018	CUX1	7	101758539	AG	A	0.080	E221fs	NM_001202543
2014-018	DDX41	5	176939370	C	T	0.219	R525H	NM_016222
2014-018	SRSF2	17	74732956	G	A	0.262	P96L	NM_001195427
2014-019	FLT3	13	28602376	C	T	0.111	M664I	NM_004119
2014-019	FLT3	13	28610138	G	A	0.120	S451F	NM_004119
2014-019	SP1	11	47377080	G	T	0.376	R171S	NM_003120
2014-020	DNMT3A	2	25457243	G	A	0.464	R882C	NM_022552
2014-020	FLT3	13	28592641	T	A	0.241	D835V	NM_004119
2014-020	FLT3	13	28608259	A	ATATTCATATTCCTCGAAATCAACGTAG	0.093	ITD	NM_004119
2014-020	GATA2	3	128202770	T	C	0.109	N317S	NM_001145661
2014-021	FLT3	13	28608317	T	G	0.475	Q580P	NM_004119
2014-021	TET2	4	106157554	TATAGTCAGACC	T	0.500	Y819fs	NM_001127208
2014-022	BCLAF1	6	136599694	G	A	0.199	R107C	NM_001077440
2014-022	CTCF	16	67650719	C	T	0.063	R14C	NM_001191022
2014-022	TET2	4	106157551	CCT	C	0.278	P818fs	NM_001127208
2014-022	TET2	4	106158106	TG	T	0.426	W1003fs	NM_001127208
2014-022	TET2	4	106196627	C	T	0.064	Q1654*	NM_001127208
2014-024	NPM1	5	170837543	C	CTCTG	0.373	L258fs	NM_002520
2014-024	TET2	4	106164778	C	T	0.388	R1216*	NM_001127208
2014-024	TET2	4	106197401	C	T	0.427	H1912Y	NM_001127208
2014-025	IDH2	15	90631934	C	T	0.475	R140Q	NM_002168
2014-025	SRSF2	17	74732935	CGCGCGGTGGTGTGAGTCCGGGG	C	0.454	PPDSHHSRR95R	NM_001195427
2014-027	DNMT3A	2	25468153	A	G	0.440	L508P	NM_022552
2014-027	IDH1	2	209113112	C	T	0.439	R132H	NM_005896
2014-027	TP53	17	7578190	T	C	0.485	Y220C	NM_000546
2014-027	TP53	17	7578454	G	A	0.410	A159V	NM_000546
2014-028	CUX1	7	101847813	TCCAGGCCAGCAG	T	0.189	V1017fs	NM_001202543
2014-028	DDX41	5	176939370	C	T	0.155	R525H	NM_016222
2014-028	SRSF2	17	74732959	G	A	0.074	P95S	NM_001195427
2014-029	ASXL1	20	31022402	TCACCACTGCCATAGAGAGCGGC	T	0.330	H630fs	NM_015338
2014-029	DNMT3A	2	25457242	C	T	0.351	R882H	NM_022552
2014-029	IDH1	2	209113113	G	A	0.451	R132C	NM_005896
2014-029	IDH2	15	90631934	C	T	0.007	R140Q	NM_002168
2014-029	RUNX1	21	36164838	C	CG	0.014	R319fs	NM_001001890
2014-029	SRSF2	17	74732959	G	T	0.020	P95T	NM_001195427
2014-031	BCOR	X	39932579	G	GT	0.217	H674fs	NM_001123383
2014-031	ETV6	12	11992103	G	GACTTTTCCTT	0.104	D65fs	NM_001987
2014-031	RUNX1	21	36231792	C	T	0.320	D171N	NM_001001890

patient	gene	chromosome	location	reference base	altered base	frequency	AA	NM
2014-032	DNMT3A	2	25457242	C	T	0.337	R882H	NM_022552
2014-032	RUNX1	21	36252876	C	G	0.071	R135S	NM_001001890
2014-032	RUNX1	21	36252937	G	GCC	0.226	A115fs	NM_001001890
2014-033	DNMT3A	2	25457243	G	A	0.430	R882C	NM_022552
2014-033	RUNX1	21	36164638	A	AG	0.352	S386fs	NM_001001890
2014-033	TET2	4	106158509	G	A	0.519	G1137D	NM_001127208
2014-034	BCOR	X	39922191	G	GT	0.370	N1293fs	NM_001123383
2014-034	BCOR	X	39911571	C	A	0.145	V1635L	NM_001123384
2014-034	CEBPA	19	33793174	CGGGGCGG	C	0.124	A47fs	NM_004364
2014-034	IDH2	15	90631839	T	A	0.403	R172W	NM_002168
2014-036	DNMT3A	2	25463576	A	T	0.447	D702E	NM_022552
2014-036	NPM1	5	170837543	C	CTCTG	0.500	L258fs	NM_002520
2014-037	ASXL1	20	31022402	TCACCACTGCCATAGAGAGCGGGC	T	0.420	H630fs	NM_015338
2014-037	BCOR	X	39932171	G	A	0.446	R810*	NM_001123383
2014-037	BCORL1	X	129155104	C	T	0.249	R1196*	NM_021946
2014-037	CEBPA	19	33793191	CG	C	0.350	P43fs	NM_004364
2014-037	EZH2	7	148506458	T	A	0.400	K629M	NM_001203249
2014-037	FLT3	13	28608242	A	AACTCCCATTTGAGATCAT	0.207	ITD	NM_004119
2014-037	PTPN11	12	112926910	G	T	0.137	Q510H	NM_002834
2014-037	RUNX1	21	36252866	G	GTGTC	0.106	V137fs	NM_001001890
2014-037	TET2	4	106196480	G	C	0.401	A1605S	NM_001127208
2014-037	WT1	11	32421575	G	T	0.016	Y110*	NM_001198552
2014-038	FLT3	13	28608251	T	TTGAGATCATATTCATATTCCTC	0.277	ITD	NM_004119
2014-039	DNMT3A	2	25457242	C	T	0.481	R882H	NM_022552
2014-039	NPM1	5	170837543	C	CTCTG	0.318	L258fs	NM_002520
2014-039	TET2	4	106164913	C	T	0.914	R1261C	NM_001127208
2014-040	BCOR	X	39923053	C	A	0.946	E1167*	NM_001123384
2014-040	NRAS	1	115256529	T	C	0.473	Q61R	NM_002524
2014-041	DNMT3A	2	25457243	G	A	0.469	R882C	NM_022552
2014-041	IDH1	2	209113112	C	T	0.440	R132H	NM_005896
2014-041	NPM1	5	170837543	C	CTCTG	0.438	L258fs	NM_002520
2014-041	NRAS	1	115256530	G	T	0.246	Q61K	NM_002524
2014-041	PTPN11	12	112888199	C	T	0.161	A72V	NM_002834
2014-042	CCND3	6	41903698	CATCT	C	0.017	T285fs	NM_001136017
2014-042	WT1	11	32417907	G	GCCGA	0.176	A153fs	NM_000378
2014-042	WT1	11	32417924	C	CGGGGCTG	0.020	P148fs	NM_000378
2014-043	CTCF	16	67654643	G	A	0.530	R377H	NM_006565
2014-043	DNMT3A	2	25469541	C	T	0.545	W409*	NM_022552
2014-043	NPM1	5	170837545	C	CTGTA	0.291	W259fs	NM_002520
2014-043	TET2	4	106190795	G	A	0.475	C1358Y	NM_001127208
2014-043	TET2	4	106197002	G	GC	0.421	A1779fs	NM_001127208
2014-043	TET3	2	74320089	C	T	0.274	R899*	NM_144993
2015-001	ASXL1	20	31022405	C	CG	0.451	H631fs	NM_015338
2015-001	CEBPA	19	33792438	C	C	0.461	A295T	NM_004364
2015-001	CEBPA	19	33793191	G	GC	0.297	A44fs	NM_004364
2015-001	EZH2	7	148506433	A	C	0.442	N637K	NM_001203249
2015-001	NRAS	1	115258747	C	G	0.378	G12A	NM_002524
2015-001	RUNX1	21	36206720	C	CT	0.500	Q237fs	NM_001001890
2015-001	TET2	4	106157506	C	T	0.463	Q803*	NM_001127208
2015-001	TET2	4	106180783	TG	T	0.511	C1271fs	NM_001127208
2015-001	TET3	2	74327616	A	AG	0.535	N1099fs	NM_144993
2015-002	NRAS	1	115256530	G	T	0.325	Q61K	NM_002524
2015-002	NRAS	1	115258747	C	T	0.058	G12D	NM_002524
2015-002	TET2	4	106156095	C	A	0.432	C332*	NM_001127208
2015-003	NRAS	1	115258744	C	T	0.376	G13D	NM_002524
2015-003	SF3B1	2	198266818	C	CTGA	0.099	-705S	NM_012433
2015-003	SH2B2	7	101957773	C	G	0.129	A392G	NM_020979
2015-003	TP53	17	7577120	T	T	0.110	R273H	NM_000546
2015-004	BCOR	X	39932876	C	TG	0.409	N575fs	NM_001123383
2015-004	DNMT3A	2	25463536	C	T	0.466	V716I	NM_022552
2015-004	IDH2	15	90631838	C	T	0.464	R172K	NM_002168
2015-005	ASXL1	20	31023159	C	T	0.302	Q882*	NM_015338
2015-005	DNMT3A	2	25457242	C	T	0.358	R882H	NM_022552
2015-005	DSCAM	21	41414478	C	A	0.564	E1836*	NM_001389
2015-005	IDH1	2	209113113	G	A	0.315	R132C	NM_005896
2015-005	RUNX1	21	36259304	CGTCCGGG	C	0.000	A33fs	NM_001001890
2015-006	IDH2	15	90631934	C	T	0.512	R140Q	NM_002168
2015-006	NOTCH1	9	139397775	C	T	0.474	V1676I	NM_017617
2015-006	NPM1	5	170837543	C	CTCTG	0.402	L258fs	NM_002520
2015-006	SRSF2	17	74732959	G	T	0.443	P95T	NM_001195427
2015-007	CSF3R	1	36932135	C	CG	0.100	L778fs	NM_000760
2015-007	FLT3	13	28602380	T	C	0.052	K663R	NM_004119
2015-007	KIT	4	55589766	GACTTACG	G	0.176	Y418fs	NM_000222
2015-008	ATM	11	108143317	C	T	0.232	L1046F	NM_000051
2015-008	KIT	4	55589321	A	T	0.323	D812V	NM_001093772
2015-008	KIT	4	55589776	TTACGACAGGCTCG	TTACTACGACTGTC	0.052	YDRLV417YDCL	NM_000222
2015-014	DNMT3A	2	25467172	CCCA	C	0.453	VG567G	NM_022552
2015-014	NPM1	5	170837543	C	CTCTG	0.400	L258fs	NM_002520
2015-014	SMC1A	X	53430550	G	A	0.865	R790W	NM_006306
2015-017	IDH1	2	209113113	G	A	0.278	R132C	NM_005896
2015-017	JAK2	9	5073770	G	T	0.106	V617F	NM_004972
2015-017	NUMA1	11	71718270	G	A	0.373	R1810C	NM_006185
2015-017	RB1	13	48878084	CGCCGCCGCT	C	0.285	AAA13-	NM_000321
2015-017	RB1	13	48878091	G	A	0.286	A15T	NM_000321
2015-017	U2AF1	21	44524456	G	A	0.239	S34F	NM_001025203
2015-018	DNMT3A	2	25457242	C	T	0.423	R882H	NM_022552
2015-018	FBXW7	4	153271203	T	G	0.337	E112A	NM_018315
2015-018	IDH2	15	90631934	C	T	0.512	R140Q	NM_002168
2015-019	BCOR	X	39913178	A	AG	0.490	L1612fs	NM_001123383
2015-019	DNMT3A	2	25463302	A	G	0.445	F731L	NM_022552
2015-019	FLT3	13	28592640	A	C	0.273	D835E	NM_004119
2015-019	IDH2	15	90631838	C	T	0.452	R172K	NM_002168
2015-019	RUNX1	21	36259202	A	AGTT	0.294	N69K	NM_001001890
2015-019	STAG2	X	123179197	C	T	0.071	R216*	NM_001042749
2015-021	NPM1	5	170837543	C	CTCTG	0.376	L258fs	NM_002520
2015-021	TET2	4	106190797	C	T	0.955	R1359C	NM_001127208

patient	gene	chromosome	location	reference base	altered base	frequency	AA	NM
2015-022	CEBPA	19	33792553	C	CA	0.267	L256fs	NM_004364
2015-022	DNMT3A	2	25471025	C	CA	0.480	P245fs	NM_022552
2015-022	NPM1	5	170837545	C	CTGCA	0.171	W259fs	NM_002520
2015-022	TET2	4	106158468	TG	T	0.940	V1124fs	NM_001127208
2015-024	NPM1	5	170837543	C	CTCTG	0.391	L258fs	NM_002520
2015-024	TET2	4	106164778	C	T	0.508	R1216*	NM_001127208
2015-024	TET2	4	106196920	TA	T	0.475	K1752fs	NM_001127208
2015-027	DNMT3A	2	25458661	T	C	0.444	N839D	NM_022552
2015-027	NF1	17	29553477	A	AC	0.487	T676fs	NM_000267
2015-027	TP53	17	7576857	A	C	0.47	L330R	NM_000546
2015-027	TP53	17	7579355	A	C	0.429	L111R	NM_000546
2015-047	TP53	17	7578394	T	C	0.432	H179R	NM_000546
2015-048	NRAS	1	115258747	C	T	0.452	G12D	NM_002524
2015-048	SPH	11	47381470	CATG	C	0.441	HM87Q	NM_003120
2015-049	ASXL1	20	31022402	TCACCACTGCCATAGAGAGCGGC	T	0.510	H630fs	NM_015338
2015-049	BCOR	X	39934013	T	TGTAA	0.919	M196fs	NM_001123383
2015-049	ETV6	12	12006463	C	CTGCT	0.459	P144fs	NM_001987
2015-049	RUNX1	21	36206875	G	A	0.364	Q186*	NM_001001890
2015-049	SRSF2	17	74732959	G	GGGC	0.481	R94RR	NM_001195427
2015-050	DNMT3A	2	25457158	G	A	0.616	A910V	NM_022552
2015-050	NSD1	5	176720934	G	C	0.634	G1920R	NM_172349
2015-050	TP53	17	7577114	T	C	0.897	C276Y	NM_000546
2015-051	CCND3	6	41903779	G	A	0.151	Q179*	NM_001136017
2015-051	CEBPA	19	33793203	CG	C	0.440	P39fs	NM_004364
2015-051	GATA2	3	128202767	G	A	0.474	A318V	NM_001145661
2015-051	KIT	4	55599340	T	A	0.061	N818K	NM_001093772
2015-051	NRAS	1	115256529	T	C	0.190	Q61R	NM_002524
2015-051	NRAS	1	115258747	C	T	0.102	G12D	NM_002524
2015-052	DNMT3A	2	25468165	AG	A	0.483	L504fs	NM_022552
2015-052	IDH2	15	90631934	C	T	0.468	R140Q	NM_002168
2015-052	NPM1	5	170837544	T	TCTGC	0.308	L258fs	NM_002520
2015-052	SRSF2	17	74732960	G	C	0.435	P95A	NM_001195427
2015-053	CEBPA	19	33792362	T	TCGG	0.208	-320P	NM_004364
2015-054	ASXL1	20	31024450	C	T	0.355	A1312V	NM_015338
2015-054	CTCF	16	67854615	C	T	0.505	R368C	NM_006565
2015-054	KDM6A	X	44879974	A	T	0.524	K188M	NM_021140
2015-054	PHF6	X	133547940	C	T	0.494	R225*	NM_001015877
2015-054	TET2	4	106156748	GAGAT	G	0.499	D551fs	NM_001127208
2015-054	TET2	4	106156747	C	T	0.501	R550*	NM_001127208
2015-055	ASXL1	20	31022441	A	AG	0.214	G643fs	NM_015338
2015-055	EZH2	7	148511065	G	A	0.516	R557W	NM_001203249
2015-055	IDH2	15	90631838	C	T	0.191	R172K	NM_002168
2015-055	NF1	17	29588751	C	T	0.086	R1513*	NM_000267
2015-055	SRSF2	17	74732935	CGCGGGCTGTGGTGTGAGTCCGGGG	C	0.495	PPDSHHSRR95R	NM_001195427
2015-055	TET2	4	106197287	G	C	0.600	E1874Q	NM_001127208
2015-056	DDX1	2	15737544	A	G	0.490	D70N	NM_004939
2015-056	KRAS	12	25398284	G	C	0.410	G12A	NM_033360
2015-056	RUNX1	21	36231792	T	C	0.460	D198N	NM_001754
2015-057	JAK2	9	5073770	G	T	0.999	V617F	NM_004972
2015-057	TP53	17	7578266	T	A	0.401	I195F	NM_000546
2015-059	DNMT3A	2	25457242	C	T	0.474	R882H	NM_022552
2015-059	FLT3	13	28608262	T	TTCATATTCTCTGAAATCAACGTAG	0.943	E598ATLISENMK	NM_004119
2015-059	NPM1	5	170837545	C	CTGTT	0.468	L258fs	NM_002520
2015-060	ASXL1	20	31023388	C	G	0.163	S958*	NM_015338
2015-060	BCOR	X	39932183	C	CGGTCCCA	0.36	V806fs	NM_001123383
2015-060	IDH1	2	209113113	G	A	0.208	R132C	NM_005896
2015-060	PHF6	X	133559270	TGAA	T	0.467	E337-	NM_001015877
2015-060	RUNX1	21	36252866	G	A	0.174	R139*	NM_001001890
2015-060	RUNX1	21	36259173	C	G	0.224	W106C	NM_001754
2015-061	IDH1	2	209113112	C	T	0.371	R132H	NM_005896
2015-061	NPM1	5	170837543	C	CTCTG	0.291	L258fs	NM_002520
2015-061	NRAS	1	115258747	C	G	0.378	G12A	NM_002524
2015-061	SMC1A	X	53441941	C	T	0.857	R96H	NM_006306
2015-062	DNMT3A	2	25463289	T	C	0.409	Y735C	NM_022552
2015-062	FLT3	13	28602376	C	T	0.188	M664I	NM_004119
2015-062	NPM1	5	170837543	C	CTCTG	0.346	L258fs	NM_002520
2015-062	PTPN11	12	112926885	C	T	0.157	S502L	NM_002834

AA: amino-acid sequence

Supplementary table 4 | list of somatic variants detected by exome sequencing

UPN	chromosome	position	gene	refseq	Protein variation	Variation type	Freq variant in tumor (%)	Depth used	Genotype	Base_ref	rs	EVS Variant Freq (%)	1000G Variant Freq (%)	Sanger sequencing confirmation
2013-001	chr1	115256528	NRAS	NM_002524	Q61H	missense	47	73	A	T				yes
2013-001	chr1	245019780	HNRNPU	NM_004501 NM_031844	E611fs	FRAMESHIFT	22	67	GGGGAAAGCCGG GA				0	yes
2013-001	chr3	38103776	DLEC1	NM_007335 NM_007337	E264K E264K	missense	44	124	A	G	rs116202356	1,63	1	
2013-001	chr3	108117621	MYH15	NM_014981	A1686T	missense	33	33	T	C				
2013-001	chr3	185826264	ETV5	NM_004454		5-UTR	53	53	G	C				
2013-001	chr5	54518098	MCIDAS	NM_001190787	R171H	missense	55	22	T	C				
2013-001	chr6	70098604	BAI3	NM_001704	A1464T	missense	40	135	A	G				
2013-001	chr8	101717195	PABPC1	NM_002568	L593V	missense	17	24	C	G	rs113574896			
2013-001	chr8	101717220	PABPC1	NM_002568	M584I	missense	17	29	T	C	rs112868101			
2013-001	chr10	101572813	ABCC2	NM_000392	V669A	missense	36	63	C	T				yes
2013-001	chr10	135084295	ADAM8	NM_001164490 NM_0011109 NM_001164489	E488K E527K E527K	missense	48	73	T	C		0		
2013-001	chr11	67798356	NDUFS8	NM_002496		5-UTR	23	13	C	A	rs117359991			
2013-001	chr12	53298504	KRT8	NM_002273 NM_001256293 NM_001256282	R88C R88C R116C	missense	25	12	A	G				
2013-001	chr14	24608199	EMC9	NM_016049		3-UTR	52	21	T	C		0		
2013-001	chr14	105415607	AHNAK2	NM_138420	V2061L	missense	20	41	G	C	rs112699389	0,07	1	
2013-001	chr16	31470913	ARMC5	NM_024742 NM_001288767 NM_001105247	A23V A118V A23V	missense	44	46	T	C				
2013-001	chr17	38975149	KRT10	NM_000421	S546R	missense	29	17	T	G				
2013-001	chr17	45234707	CDC27	NM_001256 NM_001114091	L173F L173F	missense	12	40	A	T	rs75353677			
2013-001	chr19	5229617	PTPRS	NM_130854 NM_002850	A732E A745E	missense	21	14	T	G	rs1064295			
2013-001	chrX	135958704	RBMX	NM_002139	P167A	missense	21	14	C	G	rs112089728			
2014-003	chr2	85570767	RETSAT	NM_017750	Y562fs	FRAMESHIFT	13	75		AGAT			0	
2014-003	chr4	79343136	FRAS1	NM_025074 NM_001166133	M1554V M1554V	missense	64	11	G	A				
2014-003	chr4	185615925	CENPU	NM_024629		3-UTR	32	53		CAAGAGT			0	
2014-003	chr4	185615925	PRIMPOL	NM_152683	Q559fs	FRAMESHIFT	32	53		CAAGAGT			0	
2014-003	chr5	837488	ZDHHC11	NM_024786	V298F	missense	77	13	A	C	rs28461719			
2014-003	chr6	167592571	TCP10L2	NM_001145121	A244T	missense	14	55	A	G	rs28690444		4	
2014-003	chr8	49642389	EFCAB1	NM_001142857 NM_024593	S69T S121T	missense	61	46	T	A				
2014-003	chr8	80915262	MRPS28	NM_014018	V123I	missense	40	107	T	C				
2014-003	chr8	145577021	TMEM249	NM_001252404 NM_001252402 NM_001280561	Q200H S232I S232I	missense	33	21	A	C				
2014-003	chr9	77354770	TRPM6	NM_001177310 NM_001177311 NM_017662	V178I V178I V178I	missense	41	71	T	C				
2014-003	chr10	70646119	STOX1	NM_152709 NM_001130161 NM_001130162	D856V D856V D746V	missense	48	173	T	A				
2014-003	chr11	1651615	KRTAP5-5	NM_001001480	Y182C	missense	15	33	G	A	rs74396270			
2014-003	chr11	110050621	RDX	NM_001260493		3-UTR	38	8	T	C				
2014-003	chr13	28602329	FLT3	NM_004119	A680V	missense	15	47	A	G		0,01		yes
2014-003	chr16	58075547	MMP15	NM_002428	T313P	missense	31	13	C	A				
2014-003	chr17	39262166	KRTAP4-9	NM_001146041	V176L	missense	13	39	C	G	rs75622986			
2014-003	chr17	45247397	CDC27	NM_001114091 NM_001256	G88E G88E	missense	42	41	T	C				yes

UPN	chromosome	position	gene	refseq	Protein variation	Variation type	Freq variant in tumor (%)	Depth used	Genotype	Base_ref	rs	EVS Variant Freq (%)	1000G Variant Freq (%)	Sanger sequencing confirmation	
2014-003	chr18	9522393	RALBP1	NM_006788	C313fs	FRAMESHIFT	38	80	T				0	yes	
2014-003	chr19	50412220	IL4I1	NM_001258017 NM_172374 NM_001258018		5-UTR	50	12	C	G					
2014-003	chr19	50412220	NUP62	NM_153718 NM_016553 NM_012346 NM_001193357 NM_153719	T282S T282S T282S T282S T282S	missense	50	12	C	G					
2014-003	chr19	55344235	KIR2DS4	NM_001281971 NM_001281972 NM_012314	I6V I6V I6V	missense	20	41	G	A	rs113410873				
2014-003	chr20	10625880	JAG1	NM_000214	T713M	missense	38	91	A	G				yes	
2014-003	chr22	45404497	PHF21B	NM_001284296		5-UTR	43	7	A	G	rs77409377	3,19	4		
2014-003	chrX	140993864	MAGEC1	NM_005462	F225S	missense	17	23	C	T	rs34836042				
2013-004	chr1	40253868	BMP8B	NM_001720	R97Q	missense	20	15	T	C		0,81	4		
2013-004	chr1	145209119	NOTCH2NL	NM_203458		5-UTR	42	12	C	G					
2013-004	chr1	161072096	PFDN2	NM_012394	M49V	missense	45	130	C	T					
2013-004	chr2	216977841	XRCC5	NM_021141	V42I	missense	43	90	A	G				yes	
2013-004	chr4	85993	ZNF595	NM_001286053 NM_001286052 NM_001286054 NM_182524	P199fs	FRAMESHIFT	25	20		TACAAATGTGAAAA ATGTGGCAAAGCCT TTAATAGGTCCACA TCACTTAGTAAACA TAAGAGAATTC			0		
2013-004	chr4	996204	IDUA	NM_000203	T374P	missense	23	13	C	A					
2013-004	chr4	7735055	SORCS2	NM_020777	A1039P	missense	18	17	C	G					
2013-004	chr7	23871924	STK31	NM_001260505 NM_001260504 NM_032044 NM_031414	T977M T977M T977M T1000M	missense	27	26	T	C	rs55794023	0,09			
2013-004	chr7	99526894	GJC3	NM_181538	Q117L	missense	37	94	A	T					
2013-004	chr7	131241055	PODXL	NM_001018111 NM_005397	S22P S22P	missense	31	13	G	A	rs117510299				
2013-004	chr7	143096444	EPHA1	NM_005232	Q300*	nonsense	83	12	A	G					
2013-004	chr10	55626415	PCDH15	NM_001142765 NM_001142766 NM_001142771 NM_001142773 NM_001142772 NM_001142763 NM_001142767 NM_001142764 NM_033056 NM_001142768 NM_001142769 NM_001142770	K1164R K1235R K1240R K1213R K1235R K1240R K1198R K1235R K1235R K1213R K1247R K1235R	missense	50	88	C	T					
2013-004	chr11	59480952	OR10V1	NM_001005324	Q123*	nonsense	24	63	A	G	rs499037	1,26	5		
2013-004	chr11	65414096	SIPA1	NM_006747 NM_153253	T531P T531P	missense	20	15	C	A					
2013-004	chr11	92714834	MTNR1B	NM_005959	H149Y	missense	48	214	T	C					
2013-004	chr11	117109543	RNF214	NM_001077239 NM_207343	V112M V112M	missense	46	110	A	G					
2013-004	chr12	25398281	KRAS	NM_033360 NM_004985	G13D G13D	missense	40	37	T	C	rs112445441			yes	
2013-004	chr12	47630036	PCED1B	NM_001281429 NM_138371	H397R H397R	missense	48	166	G	A					
2013-004	chr12	57114838	NACA	NM_001113203	T159N	missense	42	117	T	G					
2013-004	chr14	50623814	SOS2	NM_006939	T654A	missense	49	86	C	T				yes	
2013-004	chr16	30485552	ITGAL	NM_001114380 NM_002209	A33T A33T	missense	49	41	A	G					
2013-004	chr16	72821638	ZFXH3	NM_006885 NM_001164766	S3513G S2599G	missense	50	6	C	T			1		
2013-004	chr19	16045064	CYP4F11	NM_001128932 NM_021187	P52L P52L	missense	48	56	A	G					
2013-004	chr20	31022573	ASXL1	NM_015338	K886fs	FRAMESHIFT	43	96		GT			0	yes	

UPN	chromosome	position	gene	refseq	Protein variation	Variation type	Freq variant in tumor (%)	Depth used	Genotype	Base_ref	rs	EVS Variant Freq (%)	1000G Variant Freq (%)	Sanger sequencing confirmation
2014-019	chr3	62478027	CADPS	NM_003716 NM_183394 NM_183393	P941H P951H P911H	missense	41	140	T	G				
2014-019	chr3	195451841	ENSG00000176945	ENST0000032073 6 ENST0000044723 4 NM_001282506 ENST0000043640 8 ENST0000044552 2	G123R G123R G123R G123R G88R	missense	41	17	A	G				
2014-019	chr3	195512143	MUC4	NM_018406	S2103F	missense	26	23	A	G				
2014-019	chr5	41805770	OXCT1	NM_000436	R285Q	missense	40	98	T	C				yes
2014-019	chr6	30994862	MUC22	NM_001198815	G552fs	FRAMESHFT	18	133		GGCTCTGAGA CCTACTATGGCC TCTACCATAGG CCCTGAGACCA CCAAAGTCTCC ACTGCAAGCTC TGAGGTGACCA CAGT CTTTGCTGCAG GCTCTGAGACA ATCAGAGCCTC TACCGTAGGCT CTGAGACCACC ACAGTCTCTAC CACAGGCTCTG AGA CCACCACAGCC TCCATCATGGG CTCTGAGACCA GCACAGATTCT ACCACA			0	
2014-019	chr6	32557536	HLA-DRB1	NM_001243965 NM_002124		5-UTR	16	19	A	G	rs17211071			
2014-019	chr6	36688986	RAB44	NM_001257357	P359A	missense	17	24	G	C				
2014-019	chr7	44161487	POLD2	NM_001256879 NM_006230 NM_001127218	T56P T91P T56P	missense	21	19	G	T				
2014-019	chr9	95618618	ZNF484	NM_001007101 NM_001261459 NM_001261460		5-UTR	22	18	A	C	rs66503633			
2014-019	chr9	95618619	ZNF484	NM_001261460 NM_001007101 NM_001261459		5-UTR	17	18	C	A				
2014-019	chr11	5632402	TRIM6	NM_058166 NM_001198645 NM_001198644 NM_001003818	V433I V258I V258I V461I	missense	37	142	A	G				
2014-019	chr11	47377080	SPI1	NM_003120 NM_001080547	R171S R172S	missense	39	46	T	G				yes
2014-019	chr12	6640736	NCAPD2	NM_014865		3-UTR	50	6	T	C				
2014-019	chr12	22065906	ABCC9	NM_020297 NM_005691	R304H R304H	missense	14	73	T	C				
2014-019	chr19	46242936	ENSG00000237452	ENST0000045705 2	E42K	missense	46	102	A	G				
2014-019	chr19	54976259	CDC42EP5	NM_145057		3-UTR	50	10	A	G	rs113630689	0,02	1	
2014-019	chr21	45959819	KRTAP10-1	NM_198691	T72I	missense	18	17	A	G				
2015-056	chr1	220142184	EPRS	NM_004446.2	P1501fs	FRAMESHFT	42	136	AG	A				
2015-056	chr6	142737163	GPR126	NM_198569.2	I967T	missense	53	65	T	C				
2015-056	chr11	33565657	KIAA1549L	NM_012194.2	G553R	missense	45	380	G	A				
2015-056	chr12	25398284	KRAS	NM_033360.3	G12A	missense	41	87	C	G	rs121913529			
2015-056	chr16	71976871	PKD1L3	NM_181536.1	P1304fs	FRAMESHFT	46	160	TG	T				
2015-056	chr2	28739693	PLB1	NM_153021.4	T25I	missense	44	158	C	T				
2015-056	chr21	36231792	RUNX1	NM_001754.4	D198N	missense	46	274	C	T				
2015-056	chr10	105561071	SH3PXD2A	NM_014631.2	Y41C	missense	36	52	T	C				
2015-056	chr2	138033576	THSD7B	NM_001080427.1	G796V	missense	43	101	G	T				
2015-056	chr2	15737544	DDX1	NM_004939.2	D70N	missense	49	63	G	A				

Supplementary Table 5 | List of somatic variants detected with the targeted resequencing panel at relapse. The list of all detected variants is in **Supplementary Table 5bis (Excel file).**

patient	gene	chromosome	loc	ref base	altered base	frequency	AA	NM
2013-002	ASXL1	20	31021118	C	T	0.398	Q373*	NM_015338
2013-002	EZH2	7	148512036	A	C	0.410	C504G	NM_152998
2013-002	EZH2	7	148523591	G	A	0.449	R249*	NM_152998
2013-002	JAK2	9	5073770	T	T	0.095	V617F	NM_004972
2013-002	SETBP1	18	42531913	G	A	0.065	G870S	NM_015559
2013-006	CCND3	6	41903745	C	CG	0.497	R271fs	NM_001136017
2013-006	WT1	11	32417941	C	CACTTTTT	0.384	R141fs	NM_000378
2014-001	DNMT3A	2	25457242	C	T	0.020	R882H	NM_022552
2014-001	NPM1	5	170837543	C	CTCTG	0.024	L258fs	NM_002520
2014-001	TE12	4	106157554	TATAGTCAGACC	T	0.047	Y819fs	NM_001127208
2014-001	TE12	4	106180853	A	G	0.063	Y1294C	NM_001127208
2014-009	PTPN11	12	112915455	T	C	0.125	F285S	NM_002834
2014-009	SF3B1	2	198267359	C	A	0.063	K666N	NM_012433
2014-015	RET	10	43612093	G	A	0.090	G733D	NM_002630
2014-015	U2AF1	21	44524456	G	A	0.448	S34F	NM_001025203
2014-016	DNMT3A	2	25463175	A	T	0.174	L773H	NM_022552
2014-016	FLT3	13	28622454	TTTCGAGAGAAGG	T	0.112	TFSRK384K	NM_004119
2014-016	JAK1	1	65301882	A	C	0.097	L1053V	NM_002227
2014-016	NPM1	5	170837544	T	TCTGC	0.164	L258fs	NM_002520
2014-017	ATM	11	108186629	C	G	0.114	P2029R	NM_000051
2014-017	DNMT3A	2	25463287	A	A	0.101	R736C	NM_022552
2014-017	IDH1	2	209113113	G	A	0.125	R132C	NM_005896
2014-021	TE12	4	106157554	TATAGTCAGACC	T	0.130	Y819fs	NM_001127208
2014-027	DNMT3A	2	25468153	A	G	0.299	L508P	NM_022552
2014-027	IDH1	2	209113112	C	T	0.381	R132H	NM_005896
2014-027	TP53	17	7578190	T	C	0.336	Y220C	NM_000546
2014-027	TP53	17	7578454	A	A	0.317	A159V	NM_000546
2014-029	ASXL1	20	31022402	TCACCACTGCCATAGAGAGCGCGC	T	0.229	H630fs	NM_015338
2014-029	DNMT3A	2	25457242	C	T	0.237	R882H	NM_022552
2014-029	IDH1	2	209113113	A	A	0.098	R132C	NM_005896
2014-029	IDH2	15	90631934	C	T	0.121	R140Q	NM_002168
2014-029	RUNX1	21	36164838	C	CG	0.104	R319fs	NM_001001890
2014-029	SRSF2	17	74732959	G	T	0.210	P95T	NM_001195427
2014-033	DNMT3A	2	25457243	G	A	0.120	R882C	NM_022552
2014-033	RUNX1	21	36164638	A	AG	0.067	S386fs	NM_001001890
2014-033	TE12	4	106158509	A	A	0.072	G1173D	NM_001127208
2014-037	ASXL1	20	31022402	TCACCACTGCCATAGAGAGCGCGC	T	0.406	H630fs	NM_015338
2014-037	BCOR	X	39932171	G	A	0.387	R810*	NM_001123383
2014-037	BCORL1	X	129155104	C	T	0.311	R1196*	NM_021946
2014-037	BCORL1	X	129162789	C	T	0.094	R1420*	NM_021946
2014-037	CEBPA	19	33793191	CG	C	0.367	P43fs	NM_004364
2014-037	EZH2	7	148506458	T	T	0.548	K629M	NM_001203249
2014-037	FLT3	13	28608242	A	AACTCCCATTGGAGATCAT	0.288	ITD	NM_004119
2014-037	RUNX1	21	36252866	G	GTGTC	0.176	V137fs	NM_001001890
2014-037	TE12	4	106196480	G	T	0.420	A1605S	NM_001127208
2014-037	WT1	11	32421575	G	C	0.131	Y110*	NM_001198552
2014-038	FLT3	13	28608251	T	TTGAGATCATATTCATATTC	0.302	ITD	NM_004119
2014-038	NRAS	1	115256530	G	T	0.997	Q61K	NM_002524
2014-038	PTPN11	12	112888198	G	A	0.057	A72T	NM_002834
2014-038	DNMT3A	2	25457242	C	T	0.449	R882H	NM_022552
2014-039	MLL2	12	49433247	G	A	0.061	R2734*	NM_003482
2014-039	NPM1	5	170837543	C	CTCTG	0.051	L258fs	NM_002520
2014-039	TE12	4	106164913	C	T	0.857	R1261C	NM_001127208
2014-039	ZRSR2	X	15822319	A	T	0.163	E133V	NM_005089
2014-042	CCND3	6	41903698	CATCT	C	0.437	T285fs	NM_001136017
2014-042	WT1	11	32417924	C	CGGGGCTG	0.367	P148fs	NM_000378
2014-043	CTCF	16	67654643	G	A	0.471	R377H	NM_006565
2014-043	DNMT3A	2	25469541	C	T	0.469	W409*	NM_022552
2014-043	NF1	17	29527448	TC	T	0.495	L300fs	NM_000287
2014-043	NPM1	5	170837545	C	CTGTA	0.513	W259fs	NM_002520
2014-043	TE12	4	106197002	G	GC	0.502	A1779fs	NM_001127208
2014-043	TE12	4	106190795	G	A	0.446	C1358Y	NM_001127208
2015-003	ASXL1	20	31022441	A	AG	0.241	G643fs	NM_015338
2015-003	CBL	11	119148919	T	C	0.944	L380P	NM_005188
2015-003	FLT3	13	28592641	T	A	0.073	D835V	NM_004119
2015-003	FLT3	13	28610160	C	G	0.105	E444Q	NM_004119
2015-003	IKZF1	7	50450385	G	C	0.366	G103A	NM_001220766
2015-003	JAK1	1	65349155	G	C	0.449	L4V	NM_002227
2015-003	SETBP1	18	42531913	G	A	0.469	G870S	NM_015559
2015-003	TE12	2	74273972	C	G	0.513	P175A	NM_144993
2015-003	WT1	11	32417947	G	A	0.615	R140*	NM_001198552
2015-004	DNMT3A	2	25463536	C	T	0.382	V716I	NM_022552
2015-004	IDH1	2	209113113	G	T	0.320	R132S	NM_005896
2015-004	MLL	11	118375166	A	A	0.073	M2850I	NM_005933
2015-004	RUNX1	21	36252877	C	T	0.118	R135K	NM_001001890
2015-005	ASXL1	20	31023159	C	T	0.418	Q882*	NM_015338
2015-005	DNMT3A	2	25457242	C	T	0.458	R882H	NM_022552
2015-005	PTPN11	12	112888199	C	T	0.417	A72V	NM_002834
2015-005	SETBP1	18	42531907	G	A	0.407	D868N	NM_015559
2015-005	ZRSR2	X	15838412	G	A	0.447	V304M	NM_005089
2015-006	IDH2	15	90631934	C	T	0.508	R140Q	NM_002168
2015-006	NOTCH1	9	139397775	C	T	0.516	V1676I	NM_017617
2015-006	NPM1	5	170837543	C	CTCTG	0.515	L258fs	NM_002520
2015-006	SRSF2	17	74732959	G	T	0.489	P95T	NM_001195427
2015-014	DNMT3A	2	25467172	CCCA	C	0.327	VG567G	NM_022552
2015-014	IDH2	15	90631934	C	A	0.370	R140L	NM_002168
2015-014	NPM1	5	170837543	C	CTCTG	0.253	L258fs	NM_002520
2015-020	ASXL1	20	31022441	A	AG	0.211	G643fs	NM_015338
2015-020	BCOR	X	39932084	T	TG	0.969	K839fs	NM_001123383
2015-020	FLT3	13	28608255	G	GATCATATTCATATTCCTGAA	0.456	ITD	NM_004119
2015-020	SRSF2	17	74732959	G	T	0.429	P95T	NM_001195427
2015-020	STAG2	X	123195170	G	T	0.980	E505*	NM_001042749
2015-047	TP53	17	7578394	T	C	0.512	H179R	NM_000546
2015-058	RUNX1	21	36206875	G	A	0.482	Q186*	NM_001001890
2015-058	SF3B1	2	198267705	C	T	0.480	E592K	NM_012433
2015-058	SMC1A	X	53432009	G	A	0.973	R711W	NM_006306

Supplementary Table 6 | deep sequencing analyses in eleven patients at diagnosis relapse or in complete remission at different steps of follow-up. Data show the numbers of amplicon families (see methods) with the known altered bases, alternative (non altered/non reference) bases, and the total number of amplicon families (amplicons).

UPN 2014-001						CR after one course				CR after two courses			
Gene	AA	Chr	Pos	Reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
TET2	Y1294C	4	106180853	A	G	25	10	11810	0.0021	165	6	18226	0.009053001
DNMT3A	R882H	2	25457242	C	T	14	1	6395	0.00218921	78	2	7913	0.009857197
NPM1	L258fs	5	170837543	C	CTCTG	11	na	3007	0.003658131	32	na	4460	0.007174888
KRAS	Q61H	12	25380275	T	G	16	6	6446	0.002482159	0	0	11309	0
TET2	L627fs	4	106156978	CT	C	8	na	7816	0.001023541	49	na	11538	0.004246837
NRAS	Q61K	1	115256530	G	T	1	5	5372	<0.001	0	6	9053	0

UPN 2014- 039						diagnosis (deep sequencing)				CR after one course				CR after two courses			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
TET2	R1261C	4	106164913	C	T	5942	5	6689	0.88	3382	3	6998	0.49	2183	26	4548	0.478
NPM1	L258fs	5	170837544	C	CTCTG	1327		3545	0.374	8	na	3586	0.0022	0	na	2296	0
DNMT3A	R882H	2	25457242	C	T	2346	9	6572	0.356	1649	5	6389	0.258	857	5	3799	0.226
ZRSR2	E133V	X	15822319	A	T	21	2	570	0.037	10	1	527	0.019	13	4	468	0.028

UPN 2014-029						diagnosis (deep sequencing)				CR after one course			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	R882H	2	25457242	C	T	2424	1	7092	0.342	1825	1	7249	0.252
IDH2	R140Q	15	90631934	C	T	26	5	10278	0.0025	12	11	9596	0.0012
ASXL1	H630fs	20	31022403	TCACCACTG CCATAGAGA GGCGGC	T	1842	na	7703	0.239	1140	na	6727	0.169
IDH1	R132C	2	209113113	G	A	2800	5	6905	0.405	1939	7	6465	0.3
RUNX1	R319fs	21	36164838	C	CG	159	na	12724	0.0124	1049	na	12910	0.081

UPN 2015-014						diagnosis (deep sequencing)				CR after one course			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	VG378G	2	25467172	CCCA	C	961	na	2038	0.4715	2	na	1630	0.00122
NPM1	L258fs	5	170837544	C	CTCTG	1412	na	2981	0.473	0	na	2102	0
SMC1A	R790W	X	53430550	G	A	2159	5	2447	0.882	0	18	2137	0
IDH2	R140L	15	90631934	C	A	8	10	9052	<0.001	0	8	5377	0

UPN 2015-014						CR after 3 courses				CR 8 months before relapse				relapse (deep sequencing)			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	VG378G	2	25467172	CCCA	C	6	na	2026	0.0029	259	na	1404	0.1844	1772	na	4527	0.3914
NPM1	L258fs	5	170837544	C	CTCTG	4	na	4011	<0.001	4	na	2890	0.001384083	1098	na	7415	0.1480782
SMC1A	R790W	X	53430550	G	A	3	8	3060	<0.001	1	10	2044	<0.001	8	2	7770	0.0010296
IDH2	R140L	15	90631934	C	A	0	7	8306	0	984	7	5156	0.191	5954	14	17704	0.336

UPN 2014-038						diagnosis (deep sequencing)				CR after one course			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
FLT3	ITD	13	28608252	T	TTGAGATC ATATTCA ATTCTC	2998	na	11936	0.251	0	na	9409	0
NRAS	Q61K	1	115256530	G	T	1	2	8375	<0.001	0	0	7089	0
PTPN11	A72T	12	112888198	G	A	0	0	1440	0	0	0	1031	0

UPN 2014-037						CR after one course				
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	
ASXL1	H630fs	20	31022402	TCACCACTGCC ATAGAGGGCG	GC	T	317	na	6014	0.052
BCOR	R810*	X	39932171	G	A	217	15	9854	0.021	
BCORL1	R1196*	X	129155104	C	T	2	7	8976	<0.001	
CEBPA	P43fs	19	33793191	CG	C	117	na	6924	0.016897747	
EZH2	K629M	7	148506458	T	A	38	1	2417	0.016	
FLT3	ITD	13	28608243	A	AACTCCCATTT GAGATCAT	0	na	7859	0	
RUNX1	V137fs	21	36252871	G	GTGC	5	na	5365	0.000931966	
TET2	A1605S	4	106196480	G	T	733	2	3064	0.238	
WT1	Y110*	11	32421575	G	C	0	3	6117	0	
PTPN11	Q510H	12	112926910	G	T	147	4	7898	0.019	
BCORL1	R1420*	X	129162789	C	T	0	8	9016	0	

UPN 2014-017						CR after one course			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	R736C	2	25463287	G	A	133	4	4275	0.031
ATM	P2029R	11	108186629	C	G	80	1	2163	0.037
IDH1	R132C	2	209113113	G	A	149	3	5499	0.027
NPM1	L258fs	5	170837544	C	CTCTG	0	na	3504	0
KDM6A	V455fs	X	44920601	TG	G	0	na	5601	0

UPN 2014-033						CR after one course			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	R882C	2	25457243	G	A	608	5	5344	0.114
RUNX1	S386fs	21	36164639	A	AG	492	na	4193	0.117
TET2	G1137D	4	106158509	G	A	518	4	3749	0.138

UPN 2015-004						diagnosis (deep sequencing)				relapse (deep sequencing)			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
BCOR	N575fs	X	39932877	T	TG	3549	na	7582	0.468	0	na	7110	0
DNMT3A	V716I	2	25463536	C	T	5832	5	13851	0.421	4502	3	12200	0.369
IDH2	R172K	15	90631838	C	T	6182	19	14059	0.44	0	4	12985	0
IDH1	R132S	2	209113113	G	T	3	5	6666	<0.001	1856	2	5785	0.321
RUNX1	R135K	21	36252877	C	T	0	12	7597	0	471	2	6192	0.076

UPN 2015-005						diagnosis (deep sequencing)				relapse (deep sequencing)			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
ASXL1	Q882*	20	31023159	C	T	1498	6	5425	0.276	1720	1	4451	0.386
DNMT3A	R882H	2	25457242	C	T	2483	9	8115	0.305	2502	7	6902	0.36
PTPN11	A72V	12	112888199	C	T	0	2	1178	0	572	2	1234	0.462
RUNX1	A33fs	21	36259304	CGTCCGG G	C	0	na	15996	0	3569	na	11990	0.297
SETBP1	D868N	18	42531907	G	A	2	6	17543	<0.001	5832	6	14149	0.412
ZRSR2	V304M	X	15838412	G	A	0	4	9693	0	3588	5	8063	0.445
RUNX1	R174*	21	36231783	G	A	4986	4	10416	0.478	2	4	12101	<0.001
IDH1	R132C	2	209113113	G	A	2104	3	7562	0.278	0	4	6934	0
DSCAM	E1836*	21	41414478	C	A	4899	1	8024	0.61	0	1	6500	0

UPN 2014-041						CR post one course				CR post two courses				CR post six courses			
Gene	AA	Chr	Pos	reference base	altered base	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency	n altered bases	n alternative bases	amplicons	frequency
DNMT3A	R882C	2	25457243	G	A	3002	14	8014	0.373	994	6	4606	0.215	2083	6	6390	0.326
IDH1	R132H	2	209113112	C	T	0	3	7026	0	3	7	3197	<0.001	0	5	4968	0
PTPN11	A72V	12	112888199	C	T	0	2	1440	0	0	1	600	0	0	0	855	0
NPM1	L258fs	5	170837544	C	CTCTG	0	na	3743	0	0	na	2368	0	0	na	3005	0
NRAS	Q61K	1	115256530	G	T	0	2	7937	0	0	3	3883	0	0	0	6070	0

na: not applicable (for indels, see methods)

Supplementary table 7 | Analysis of NSG bone marrow repopulation by primary AML cells.

UPN	Time from injection to analysis	MOUSE	hCD45%	CD45+ CD33+%	CD45+ CD19+%	CD45+ CD3+%
2014-002	28	mouse 1	0.1	0	0	0
		mouse 2	0.4	0	0	0
2015-007	25	mouse 1	0	0	0	0
2015-051	26	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
		mouse 4	0	0	0	0
		mouse 5	0	0	0	0
		mouse 6	0	0	0	0
		mouse 7	0	0	0	0
2015-058	20	mouse 1	0.6	0.15	0	0.48
		mouse 2	0	0	0	0
		mouse 3	0.5	0.25	0	0.25
		mouse 4	0	0	0	0
		mouse 5	0.03	0	0	0
		mouse 6	0	0	0	0
2015-048	14	mouse 1	0.3	0.29	0	0
		mouse 2	0.06	0.05	0	0
		mouse 3	0.02	0.02	0	0
		mouse 4	0.2	0.2	0	0
		mouse 5	0	0	0	0
2014-003	13	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
2015-020	19	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
2015-050	16	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
		mouse 4	0	0	0	0
		mouse 5	0	0	0	0
2015-019	19	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
2015-004	19	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
2015-054	22	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0.3	0.3	0	0
2015-053	30	mouse 1	0	0	0	0
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
		mouse 4	2	0.12	1.8	0
		mouse 5	0.12	0	0.11	0.01
		mouse 6	1.2	0.12	1.08	0
2014-040	12	mouse 1	1	1	0	0
		mouse 2	0.5	0.5	0	0
		mouse 3	2	2	0	0
2015-061	13	mouse 1	1.6	1.6	0	0
		mouse 2	1.2	1.2	0	0
		mouse 3	1.7	1.39	0.22	0
		mouse 4	1.8	1.8	0	0
		mouse 5	1.5	1.5	0	0
2015-008	20	mouse 1	1	0.09	0.77	0
		mouse 2	2	0.3	1.34	0
		mouse 3	0.5	0.21	0.21	0
		mouse 4	4	0.2	3.36	0
2015-060	21	mouse 1	2	0	0	2
		mouse 2	0	0	0	0
		mouse 3	0	0	0	0
		mouse 4	0.2	0	0	0.2
2014-013	12	mouse 1	0.7	0.04	0.02	0.62
		mouse 2	0.5	0	0	0.5
		mouse 3	2	1.7	0.2	0

UPN	Time from injection to analysis	MOUSE	hCD45%	CD45+ CD33+%	CD45+ CD19+%	CD45+ CD3+%
2015-018	12	mouse 1	2	0.16	0	0
		mouse 2	0	0	0	0
		mouse 3	2	0.16	0	0
2014-020	24	mouse 1	0.5	0	0.5	0
2015-059	29	mouse 1	71	1.42	ND *	8.52
		mouse 2	45	0.45	ND *	7.65
		mouse 3	8	0	ND *	7.2
2015-062	14	mouse 1	0.1	0.06	0.04	0
		mouse 2	0.2	0.17	0.03	0
		mouse 3	2.5	2.28	0.2	0
		mouse 4	0.3	0.12	0.18	0
		mouse 5	7	0.84	0.35	ND **
2015-021	15	mouse 1	11.33	0.09	11.24	0
		mouse 2	5.54	0.05	5.49	0
		mouse 3	12.66	0.1	12.56	0
		mouse 4	8.83	0.07	8.74	0
2015-055	16	mouse 1	3	0	0	3
		mouse 2	0.7	0	0.13	0.61
		mouse 3	0	0	0	0
		mouse 4	0.6	0.6	0	0
	35	mouse 5	0.4	0.4	0	0
		mouse 6	0.03	0.03	0	0
		mouse 7	0.08	0.08	0	0
		mouse 8	0.5	0.5	0	0
2015-057	10	mouse 1	64	64	0	0
		mouse 2	22	22	0	0
		mouse 3	40	40	0	0
		mouse 4	84	84	0	0
		mouse 5	56	56	0	0
2015-056	8	mouse 1	58	58	0	0
		mouse 2	30	30	0	0
		mouse 3	29	29	0	0
		mouse 4	47	47	0	0
2014-007	24	mouse 1	72	72	0	0
		mouse 2	71	71	0	0
		mouse 3	67	67	0	0
2014-042	11	mouse 1	80	80	0	0
		mouse 2	84	84	0	0
		mouse 3	67	67	0	0
2014-038	11	mouse 1	85	85	0	0
		mouse 2	87	87	0	0
2014-019	11	mouse 1	45	8.1	33.75	0
		mouse 2	70	10.5	11.9	47.6
2013-004	20	mouse 1	8	5.84	1.68	0
		mouse 2	2.5	1.43	0.75	0
		mouse 3	4.5	0.63	3.15	0
2015-049	28	mouse 1	50	50	0	0
		mouse 2	45	45	0	0
		mouse 3	57	57	0	0
		mouse 4	56	56	0	0
2014-037	19	mouse 1	88	88	0	0
		mouse 2	11	11	0	0
2015-052	43	mouse 1	28	25.48	0.84	2.52
		mouse 2	23	19.55	0.46	2.53
		mouse 3	52	49.4	2.6	2.6
2014-041	18	mouse 1	39	32.37	2.535	0
		mouse 2	39	30.42	5.46	0
		mouse 3	41	34.85	2.67	0
2014-039	22	mouse 1	48	48	0	0
		mouse 2	34	34	0	0
		mouse 3	40	40	0	0
2013-003	22	mouse 1	17	17	0	0
		mouse 2	85	85	0	0
		mouse 3	79	79	0	0
2014-043	12	mouse 1	80	80	0	0
		mouse 2	89	89	0	0
2015-001	12	mouse 1	93	93	0	0
		mouse 2	98	98	0	0
		mouse 3	95	95	0	0
		mouse 4	90	90	0	0

ND: no data

* CD19+ cells were detected in blood of these mice at week 12

** SSC/FSC/CD45 data indicating a lymphocyte morphology

Supplementary table 8 | Order of lesion acquisition in 49 patients with lesions in epigenetic regulators in first position

UPN	Somatic events										
	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
2014-002	NRAS										
2014-003	MLL	FLI3-ITD									
2015-007	WT	FLI3									
2014-002	DNMT3A	RUNX1/IT1									
2014-025	IDH2	SRSF2									
2014-028	CUX1	IDH4T1									
2014-024	TE12H1	TE12H2	NP1T								
2015-048	NRAS	NRAS G12	SPT1								
2014-042	60k20q	CGI02	WTF2								
2015-014	DNMT3A	SMC1A	NP1T								
2014-034	IDH2	BCOR1	BCOR2								
2015-024	TE12H1	TE12H2	NP1T								
2015-002	nr16	TE12H3	CTCF								
2014-015	60k20q	TE12H3	CTCF								
2014-020	DNMT3A	FLI3-ITD	FLI3-LOH								
2015-052	DNMT3A	IDH2	SRSF2								
2014-036	DNMT3A	NP1T	FLI3								
2015-062	DNMT3A	NP1T	FLI3-LOH								
2014-033	TE12	DNMT3A	RUNX1								
2014-011	MLL	SPI1	OXY1								
2014-019	DNMT3A	IDH1	NP1T								
2013-001	MLL	NRXNP2J	ABC2								
2015-008	nr16	nr22	KIT1F								
2015-008	nr16	nr22	RIT2								
2014-021	TE12	Y	FLI3-LOH								
2014-021	TE12	(8:21)	del11p								
2014-039	TE12	DNMT3A	TE12-LOH								
2015-011	IDH2	60k11	DNMT3A								
2015-059	DNMT3A	FLI3-ITD	FLI3-LOH								
2015-022	TE12	DNMT3A	TE12-LOH								
2014-029	DNMT3A	ASXL1	SRSF2								
2015-006	IDH2	NOTCH1	SRSF2								
2014-017	BCOR	DNMT3A	ATM								
2015-019	DNMT3A	NP1T	NP1T								
2014-019	DNMT3A	NAK1	NP1T								
2014-003	ASXL1	EZH2H1	EZH2H2								
2015-044	SRSF2	BCOR	ETV6								
2015-004	ASXL1	JAG1	RALBP1								
2013-003	TE12	SMC1A	DNMT3A								
2015-005	DNMT3A	ASXL1	IDH1								
2014-037	TE12	EZH2	BCOR								
2015-054	KDM6A	CTCF	TE12H1								
2015-001	TE12	TE12H1	RUNX1								
2015-053	TE12	EZH2	SRSF2								

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.864	0.41										
1.000	0.176	WT, FLI3, and CSFR3 VAFs suggest the existence of 3 parallel sub-clones (85% of the cells).									
0.697	0.097	WT, FLI3, and CSFR3 VAFs suggest the existence of 3 parallel sub-clones (85% of the cells).									
0.338	0.071	An alternative linear clonal architecture with bi-allelic RUNX1 mutation in a single clone is possible.									
0.476	0.455	Alternative clonal architectures are possible.									
0.19	0.156	0.074	An alternative clonal architecture with parallel clones is possible.								
0.428	0.388	0.373	Order of events confirmed by colony analysis.								
0.905	0.452	0.441	Order of events confirmed by colony analysis.								
0.509	0.017	0.051	WTF1 lost at relapse. CCA03 VAF=0.395 > WTF2 VAF=0.327, in a major clone, indicating linear acquisition.								
0.453	0.176	0.4	At relapse, loss of SMC1A mutation, and change in NP1T mutation (antitoxin by an IDH2 mutation in the relapse clone).								
0.403	0.37	0.147									
0.51	0.475	0.391	Order of events confirmed by colony analysis. No colony with CTCF mutation.								
0.426	0.46	0.165	VAFs suggest the existence of 3 parallel sub-clones in the TE12H3 mutant subclone.								
0.693	0.058	NRAS G12 and G61 VAFs, and the percentage of cells with +8 suggest the existence of 3 parallel sub-clones representing 80% of the cells.									
1.000	0.433	0.325	Order of events confirmed by colony analysis.								
0.61	0.296	0.037	A linear alternative clonal architecture is possible.								
0.463	0.093	ND	Order of events confirmed by colony analysis. FLI3-LOH detected by colony analysis.								
0.447	0.5	0.5	FLI3 LOH in 52% of the cells. DNMT3A mutant was placed before NP1T mutant based on the analysis of blood sample at relapse with 19:4q4 DNMT3A mutant reads and 0:107 NP1T mutant reads.								
0.409	0.346	0.61	0.157	An alternative clonal architecture with linear acquisition of FLI3 and PTPN11 mutations is possible.							
0.519	0.431	0.332	0.13	An alternative clonal architecture with linear FLI3-ITD acquisition is possible.							
0.484	0.399	0.163	0.098	Order of events confirmed by colony analysis.							
0.789	0.38	0.4	0.107	Order of events confirmed by colony analysis. No colony was found with FLI3 S451F mutation.							
0.695	0.22	0.36	0.181	An alternative clonal architecture with linear acquisition of FLI3 S451F mutation after NRAS mutation is possible.							
0.488	0.498	0.438	0.181	Order of events confirmed by colony analysis. No NRAS mutant colony was found.							
0.905	0.905	0.325	0.233	Order of events confirmed by colony analysis. NRAS and PTPN11 mutations occurred in distinct subclones (NRAS > PTPN11, non overlapping 95% CI)							
0.959	1.000	0.022	0.378	TE12 LOH in 60% of the cells, indicating that 100% of the cells are TE12 mutant.							
0.5	0.865	0.01	0.945	Karyotype, FISH, and NGS data at relapse indicate the loss of del(11) and FLI3 mutation.							
0.916	0.483	0.916	0.318	Deep sequencing and follow-up analyses indicating the presence of the ZRSR2 mutation in a subclone at diagnosis.							
0.512	0.65	0.42	0.037								
0.474	0.943	0.943	0.456	FLI3 LOH in 94% of the cells.							
0.94	0.462	0.941	0.171	Order of events confirmed by colony analysis, with a complex pattern. CEPBA analysis was not done on colonies.							
0.35	0.33	0.32	0.05	0.014	Changes at relapse indicate that DNMT3A, ASXL1, and SRSF2 mutations preceded IDH1 mutation.						
0.512	0.474	0.443	0.402	0.2	VAFs suggest that IDH1 and IDH2 mutations occurred in distinct clones, and that RUNX1 mutation is in last position.						
0.482	0.462	0.456	0.466	0.294	0.273	An alternative linear clonal architecture is possible.					
0.491	0.453	0.446	0.294	0.273	An alternative linear clonal architecture is possible.						
0.458	0.43	0.413	0.709	0.709	Order of events confirmed by colony analysis. FLI3-LOH in 83% of the cells, detected in colonies.						
0.357	0.322	0.291	0.316	0.088	Order of events confirmed by colony analysis.						
0.975	0.38	0.38	0.42	0.157	Order of events confirmed by colony analysis.						
0.511	0.481	0.919	0.459	0.384	Order of events confirmed by colony analysis.						
0.483	0.478	0.439	0.413	0.412	0.017	Alternative clonal architectures are possible with CEL mutation branching after either SMC1A, DNMT3A, or NP1T mutations.					
0.547	0.502	0.48	0.421	0.291	0.275	Changes at relapse indicating that the TE12 mutation occurred after the NP1T mutation.					
0.359	0.303	0.316	0.607	0.605	0.624	Chr. 21 LOH in 62% of the cells. IDH1, RUNX1, and DSCAM mutations lost at relapse.					
0.409	0.4	0.447	0.424	0.35	0.137	Loss of PTPN11 mutation at relapse VAFs, complete remission, and relapse NGS analyses suggest the proposed clonal architecture. Alternative clonal architectures are possible.					
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

UPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11
0.529	0.506	0.501	1.000	0.5	0.494	0.857	0.356				
0.535	0.511	0.5	0.465	0.464	0.451	0.379	0.297				
0.601	0.517	1.000	0.495	0.923	0.541	0.436	0.215	0.191	0.274	0.006	

Supplementary table 9 | Order of lesion acquisition in 25 patients with no evidence for lesions in epigenetic regulators in first position

JPN	event 1	event 2	event 3	event 4	event 5	event 6	event 7	event 8	event 9	event 10	event 11	Quantification of mutant alleles or cells* no somatic mutation detected
014-004												0.49
014-010												0.324 0.208
015-053												0.377 0.369 0.352
014-006												0.24 0.218 0.08
014-018												0.467 0.378 0.167
013-006												0.275
015-051												0.474 0.44
015-051												0.151 0.081
014-008												0.464 0.955 0.17 0.12
014-009												0.403 0.446 0.325 0.12
015-051												0.857 0.378 0.371 0.291
015-051												0.492 0.588 0.437 0.899 0.413
015-051												1.000 1.000 1.000 0.951 0.474
015-051												0.467 0.224 0.208 0.36 0.174 0.163
015-051												0.319 0.638 0.592 0.268 0.192 0.094
015-051												0.377 0.263 0.263 0.129 0.11 0.09
014-027												0.486 0.857 0.947 0.44 0.958 0.44
015-057												0.999 0.999 0.401 0.53 0.51 0.46
015-017												0.374 0.6 0.6 0.279 0.239 0.196 0.3
014-014												0.335 0.348 NA 0.647 0.583 NA
015-027												0.487 0.47 0.903 0.896 0.895 NA
015-047												0.433 0.782 0.752 0.752 0.641 NA
015-050												0.89 0.635 0.617 NA
015-058												0.973 0.482 0.48 0.324
015-020												0.892 0.969 0.429 0.456 0.211

Subclonal events with possible alternative architectures:	Subclonal events with order of occurrence informed from copy analysis and/or allelic imbalance (non overlapping 95% confidence intervals), and/or remission/relapse data
<ul style="list-style-type: none"> Epigenetic events PM1.F factors Other events 	<ul style="list-style-type: none"> Subclonal events with possible alternative architectures: Relapse sample

Black: VAFs	Blue: VAFs (X linked genes, male patient)	Red: quantification of chromosome aberrations (karyotype, interphase FISH, SNP analysis of NGS data)
		Green: FLT3-ITD quantification (sizing)

Quantification of mutant alleles or cells*
no somatic mutation detected

0.49

0.324 0.208

0.377 0.369 0.352

0.24 0.218 0.08

0.467 0.378 0.167

0.275

0.474 0.44

0.151 0.081

0.464 0.955 0.17 0.12

0.403 0.446 0.325 0.12

0.857 0.378 0.371 0.291

0.492 0.588 0.437 0.899 0.413

1.000 1.000 1.000 0.951 0.474

0.467 0.224 0.208 0.36 0.174 0.163

0.319 0.638 0.592 0.268 0.192 0.094

0.377 0.263 0.263 0.129 0.11 0.09

0.486 0.857 0.947 0.44 0.958 0.44

0.999 0.999 0.401 0.53 0.51 0.46

0.374 0.6 0.6 0.279 0.239 0.196 0.3

0.335 0.348 NA 0.647 0.583 NA

0.487 0.47 0.903 0.896 0.895 NA

0.433 0.782 0.752 0.752 0.641 NA

0.89 0.635 0.617 NA

0.973 0.482 0.48 0.324

0.892 0.969 0.429 0.456 0.211

Supplementary table 10 | Primers used for Sanger sequencing and FLT3-ITD detection (5'-3')

ABCC2_V669A_F	TCAATACCCAACCCCTGCTA
ABCC2_V669A_R	ATTCCGGGAGTCAGAGGCTTT
AKAP13_G1613R_FWD	CCAAAGGGATTTTGTGTCCCT
AKAP13_G1613R_REV	CACCACTGAATGGGTGTCTG
APH1A_P446A_FWD	GGTGGGTTTGAATCTGCACT
APH1A_P446A_REV	ATACAGGGCGAGGACATCAG
ASXL1_delGT_F	AGAGGTCACCACTGCCATAGA
ASXL1_delGT_R	TGGGTATGCTCCCCATTTAG
ASXL1_FWD	ACCCTGGGTGGTTAAAGGTC
ASXL1_FWD	ACCCTGGGTGGTTAAAGGTC
ASXL1_FWD	ACCCTGGGTGGTTAAAGGTC
ASXL1_G1397S_FWD	GGGGTCTCTTAAGGCAAAAT
ASXL1_G1397S_REV	GAGTTGGGAGGGGAGAGAAG
ASXL1_Q373*_FWD	CCTGAAACTGATGGCTGTGA
ASXL1_Q373*_R	CAGATTCCCTTCTGGCTCTGG
ASXL1_Q882STOP_FWD	AATCCTCACCACTGATTGC
ASXL1_Q882STOP_REV	CATTCGATGGGATGGGTATC
ASXL1_REV	CTCTGCCACCTCCCTCATC
ASXL1_REV	CTCTGCCACCTCCCTCATC
ASXL1_REV	CTCTGCCACCTCCCTCATC
ASXL1_Y300C_FWD	GATGAGGGAGGTGGCAGAG
ASXL1_Y300C_R	GAAGGCAGGTCCCTCTCTCCT
ATM_D1853V_FWD	ATATGTCAACGGGGCATGAA
ATM_D1853V_R	GTGTCCAAGACAGCTGGTGA
ATM_F858L_FWD	GTGCCCAGCCTGATTAGGTA
ATM_F858L_FWD	GTGCCCAGCCTGATTAGGTA
ATM_F858L_REV	CTTTGGCTCTCTCCAGGTTT
ATM_F858L_REV	CTTTGGCTCTCTCCAGGTTT
ATM_L1046F_FWD	GAAAACCTTACTTGATTTTCAGGCATC
ATM_L1046F_REV	TGGCCCATTTTGAATAAGGA
ATM_P1054R_FWD	TGCTTGAGGTGAGTTTTTTC
ATM_P1054R_REV	GATTGACTCTGCAGCCAACA
ATM_P2029R_FWD	TTGTTGTTTCCATGTTTTCAGG
ATM_P2029R_R	CCACATTGCTTCGTGTTTCAT
ATM_Q1128R_FWD	ATGTTTCCAGGACACGAAGG
ATM_Q1128R_FWD	ATGTTTCCAGGACACGAAGG
ATM_Q1128R_REV	ATGGGATATTCATAGCAAGCA
ATM_Q1128R_REV	ATGGGATATTCATAGCAAGCA
ATP2C2_M466I_F	CAACAATGCGGTCATCAGAA
ATP2C2_M466I_R	CTCCTCCACCACCAACAGAT
BCLAF1_K162E_FWD	CGGATGTGGCTGATGACTTA
BCLAF1_K162E_REV	CAGATCTCGCCGTCATATAG
BCLAF1_R107C_FWD	ATGGGTACAGAGGAAGGGGTA
BCLAF1_R107C_REV	TGGGGATGAAGAACGAGAAG
BCLAF1_S343P_FWD	GGGATTGGGCAACACATTTA
BCLAF1_S343P_REV	CCCTCCCTTCTCTGATCCT
BCOR_E1167*_FWD	CAACCTGAAGGTGTGCATTG
BCOR_E1167*_REV	TCTTCCGACCAGCTTCTGTT
BCOR_FER_FWD	AAAACAAGGCATTGGACTGG
BCOR_FER_REV	TACTGTGCTTGGCAGGAGTG
BCOR_LET_FWD	ACCGAGTTTCAAGCAACG
BCOR_LET_REV	CCAGGAAACAGACTGCCATT
BCOR_R810STOP_FWD	CCCACCGACAAGAACCCTAAA
BCOR_R810STOP_REV	TCTCTCAGGCGATGAAATC
BCORL1_R1196STOP_FWD	CCAGGAATCCACCAAGAAAA
BCORL1_R1196STOP_REV	TTCCAGAGTGAAGGAGTCAGC
BRPF1_G1058S_FWD	ACTTTCCAGAGGACAGCAG
BRPF1_G1058S_REV	GGCCAGACTCTTCCCTTACC
CBL_L380P_FWD	GGAAACAAGTCTTCACTTTTTCTGT
CBL_L380P_REV	CCGTACCTGCCAGGATGTAA
CBL_R844W_FWD	TGCAGGTGAAGATGCTTTTG
CBL_R844W_REV	TTGGCTTTTCCCTCCTACCT
CCND3_D205H_FWD	GAAGCTGCACTCAGGGAGAG
CCND3_D205H_REV	CCCTTCAGGCTTAGATGTGG
CCND3_FS_FWD	GAAGCTGCACTCAGGGAGAG
CCND3_FS_REV	CCCTTCAGGCTTAGATGTGG
CHEK2_G222STOP_FWD	CCACTGAGAATGCCACTTGA
CHEK2_G222STOP_REV	AGAAACTCCCACCACAGCAC
CSF3R_W547STOP_FWD	GGAAGCCACAAGAAGTCCAA
CSF3R_W547STOP_REV	TGTGGGGAACTGAGGATAG

CTCF_R377H_FWD	ACTTCAGTGCCCCAAGCTA
CTCF_R377H_REV	TCATGTGCCTTTTCAGCTTG
CUX1_FWD	AGAAAGGCCGAGAACCCTTC
CUX1_REV	AGTTCCTCCTCCCTCTCTCG
CUX2_P834L_FWD	TCCTCCTCTGGCTACTCTGG
CUX2_P834L_R	ACGTAGGCCGGGTAGTAGG
DAAM2_H650Q_FWD	GGCACCGTATGGAATGAGAT
DAAM2_H650Q_REV	GAAGGGGAAGATGGAGGAAG
DAAM2_R335Q_FWD	CTCAACCAAGGGTGTGTCCT
DAAM2_R335Q_REV	CACGGGGTCTAGAAAAGTGG
DAAM2_R481Q_FWD	TCCAGAACACATGGAGCTTG
DAAM2_R481Q_REV	TGTAAAGGGGGAGATGCTTG
DAXX_D339H_FWD	AAAAGTGAGGAGGGCGAGAG
DAXX_D339H_R	CAGAGGGTTTGGTTCTTGCT
DDX41_G267E_FWD	GGTTCAGGCAAGACACTGGT
DDX41_G267E_REV	CGGCAGTAGTACTCCAGGATG
DDX41_NIC_FWD	ATGGGTTAGGCCGAAAAG
DDX41_NIC_REV	GCTCTGCAGTCACCTCCAAT
DDX41_R525H_FW	TGACATGCCAGAGGAGATTG
DDX41_R525H_R	ACCCAGGGAACAGCTAAGGT
DDX41_T472S_FWD	TGCTGCAGTAGCAGTGCTTT
DDX41_T472S_REV	AATTGATGACGTGCTGGATG
DIS3_D408N_FWD	CCTCTTGGTTTCAGGGACAC
DIS3_D408N_FWD	TAATGCCTCTTGGTTTCAGG
DIS3_D408N_R	TAATGCTCCAGGGCATCTTT
DIS3_D408N_REV	TAATGCTCCAGGGCATCTTT
DNMT3A_C162Y_FWD	CTCCTCTTTGCATCGGGTAA
DNMT3A_C162Y_REV	ACCTGCACTCCAACCTCCAG
DNMT3A_D513E_FWD	CTCCTGTCTGCCTCTGTCC
DNMT3A_D513E_REV	CAAGGAGGAAGCCTATGTGC
DNMT3A_ex23_F	CTGCCCTCTCTGCCCTTTTCT
DNMT3A_ex23_R	TTGTGTCGCTACCTCAGTTTG
dnmt3a_fwd	GCACCCGACTCTTTTCTG
DNMT3A_G34S_FWD	TACACTGCTGGGATCCACCT
DNMT3A_G34S_REV	TCCTTGCAAGGACATACATCA
DNMT3A_L548H_FWD	GGCTTTCTCTTCCGACCTCT
DNMT3A_L548H_R	ATGAAGCAGCAGTCCAAGGT
dnmt3a_rev	TACCTTGCAAGTTTGGCACA
DNMT3A_V527L_FWD	CTTCTGTCTGCCTCTGTCC
DNMT3A_V527L_REV	CAAGGAGGAAGCCTATGTGC
DNMT3A_W220STOP_FWD	CTCAGAGTCTGGCCTTGAGC
DNMT3A_W220STOP_REV	TGCCCTCATTACCTTCTGG
dnmt3afs_fwd	AGGTGGCCTTGCTAATTCCT
dnmt3afs_rev	CTTGGTGGCATTCTTGCTCC
DOK2_FWD	CTGCCCTGTCCCTCTATGAC
DOK2_REV	GGAGGAGTCCACAGCAGAAG
DSCAM_E1836STOP_FWD	GATGGTGAGTCGTCCGTCTT
DSCAM_E1836STOP_REV	GGAGGTCAGACTGTCCGTGT
DSCAM_K322R_FWD	CGTGATGCACTTGCTTTGTT
DSCAM_K322R_REV	TCCAGGGTTGAGGATTTTAC
DSCAM_S171P_FWD	GTGGTTGGGTTGCTTTTCATT
DSCAM_S171P_REV	ACAAAAAGTCTGGCGTGT
EGR1_P459S_FWD	AAAGTGTGTGGCCTCTTCG
EGR1_P459S_REV	GGGGAACAGAGGAGTACGTG
EZH2_G538STOP_FWD	AAGCACAGTGAACACCAAG
EZH2_G538STOP_REV	TTTGCCCCAGCTAAATCATC
EZH2_K629M_FWD	AGGCAAACCCGGAAGAAGT
EZH2_K629M_REV	GGACTGAAAAGGGAGTTCCA
EZH2_N637K_FWD	AGGCAAACCCGGAAGAAGT
EZH2_N637K_REV	GGACTGAAAAGGGAGTTCCA
EZH2_R249*_FWD	CCTGAATGTACCCCAACAT
EZH2_R249*_R	AGCACTCTCCAAGCTGCTTTA
FANCA_G501S_F	CTGGGACAGGTGTGAGGAGT
FANCA_G501S_R	CCAGGCAGTTCCAGACTAA
FANCA_G809D_F	TCTTCAGTGCCCTGGAAAAAT
FANCA_G809D_R	AACAAGGAATCCCTCGTCCT
FANCA_L63V_F	ACCACGCCCAACCCTTTT
FANCA_L63V_R	TGCCTCCTGAAGTCCAGAAT
FANCA_R73S_F	AGGCCATGTCCGACTCGT
FANCA_R73S_R	GGCGTTCGGGAAAGAAGG

FANCA_T266A_F	TCCTGAAGTGCCGTTTCTTT
FANCA_T266A_R	TTACAGTCTGGGCTGCAGTG
FBXW7_E112A_FWD	TTGGACCATGGTCTGAGGT
FBXW7_E112A_REV	TTCCCGTAATCTCAAATGTG
FBXW7_F86L_FWD	AGCTGGCTTTTGAAATGAA
FBXW7_F86L_REV	CCTCATCTTGTTCCACCAGCA
FBXW7_FWD	ATTTTCCCCTGCAGAATGTG
FBXW7_REV	TGAGTTGCTGTGCTGTTC
FLT3_A680V_FWD	GCTCTGAAAGAGAGGCACTCA
FLT3_A680V_R	CCACTTGGGTTTGAGAGTTCA
FLT3_E444Q_FWD	CCACCTTGGCTTCACAAAGT
FLT3_E444Q_REV	TTCTTCCAGGTCCAAGATGG
FLT3_M664I_FWD	AATGCAGATTGACTCTGAGCTG
FLT3_M664I_R	GGTGATTTTCGTGGAAGTGG
FLT3_Q580P_FWD	TCTGCAGAACTGCCTATTCCCT
FLT3_Q580P_REV	CTGCAGAAACATTTGGCACA
FLT3_S451F_FWD	GCCCAGCCAGTGAGCTTAT
FLT3_S451F_R	GGTCAGAGAGTTTTATGTTCTTCCA
FLT3-ITD_F	FAM-TGGTGTGTCTCTCTTCTTTCATTG
FLT3-ITD_R	GTGCGTTCATCACTTTTCCAA
FOXP1_T2901S_FWD	CCACGCATCCTCTGTGTTAC
FOXP1_T2901S_REV	GTTGGGGTCTGGAGATAG
FZD1_P48S_FWD	GCGAGCTGGGAACCTTGTG
FZD1_P48S_REV	GCTCTGTGCTGCTGAGG
GATA2_N317S_F	CCGGGAGTGTGTCAACTGT
GATA2_N317S_R	AAAGCGTCTGCATTTGAAGG
GATAD2B_R10C_FWD	AACTTCACCATTCCACCTTTTT
GATAD2B_R10C_REV	TTGCCAAATCCTTCCCTTTTG
GCSH_S9L_F	CTCTGCCAGTCAGGCTCT
GCSH_S9L_R	CCAGGGATCCCAACAGAAAT
HJURP_T539I_FWD	CGCAGGCTGAGTTTACCTTC
HJURP_T539I_REV	GTTTTGCTGGGTGACACTGA
HNRPU_INDEL_F	TCCAGCGAAAAGCTGTTGTA
HNRPU_INDEL_R	TTTCAGGAGGCTAAGTCCATA
HPR_H272D_F	CTGGAGGAGGACACCTGGTA
HPR_H272D_R	ACCCATCAGCTTCAAACCAC
IDH1_F	CCATTTGTCTGAAAACCTTGCT
IDH1_R	GCAAAATCACATTATTGCCAAC
IDH2_F	GCTGCAGTGGGACCACTATT
IDH2_R	GTGCCCAGTTCAGTGGAT
IDH2_R261H_FWD	ATGCCATCCAGAAGAAATGG
IDH2_R261H_REV	AGCTGAGCCAAATGCACTCT
IKZF1_G103A_FWD	ATTCACCCAGAGGGCAAC
IKZF1_G103A_REV	GACTCAGGGTTAGCCAGCAA
JAG1_T713M_F	CCACTGGGACTCACACTGAA
JAG1_T713M_R	CCAGGCCAGAGAAATATCA
JAK1_L4V_FWD	CCAACACCAAATGTTTCTGC
JAK1_L4V_REV	CCTCCAGGTTCACTCAGTC
JAK3_V722I_FWD	TCCCCTTTTCATTCCCTCAG
JAK3_V722I_REV	TTCTGTCAAAGTGGGGGTTTC
KDM6A_FS_FWD	CATTTGGCCTCCTCTAACCA
KDM6A_FS_REV	CCTGTAATTTCTGTGGTGTCAAA
KIT_D812Y_FWD	TGGTTTTCTTTTCTCCTCCAA
KIT_D812Y_REV	TGCAGGACTGTCAAGCAGAG
KIT_OLI_FWD	GCTGAGGTTTTCCAGCACTC
KIT_OLI_REV	CCTCTGCTCAGTTCCTGGAC
KIT_V526I_FWD	TGCCAAAGTTTGTGATTCCA
KIT_V526I_REV	GTGGGGAGAAAAGGGAAAAAT
KRAS_G13D_FWD	CTTAAGCGTCGATGGAGGAG
KRAS_G13D_REV	AGAATGGTCCCGCACCAGTAA
KRAS_Q61H_FWD	TGTGTTTCTCCTTCTCAGGA
KRAS_Q61H_R	GCATGGCATTAGCAAAGACTC
MAML1_M759R_FWD	CCCTGGACATGCTTCAGTTT
MAML1_M759R_REV	CCCAGAGAGTTCTGCCATA
MAP2K2_P16L_FWD	GGCTCCCGATCCCCTTAT
MAP2K2_P16L_R	ACCCCTGCCACTCACTC
MLL2_H1525R_FWD	CTTCCACTGTGAATGCCAGA
MLL2_H1525R_REV	TGGAGAACAGAGACTGGAGGA
MLL2_P1157T_FWD	CTCTGGATGGGATTGATGCT
MLL2_P1157T_REV	CGTGGCTCTTCTGTCTTCTC

MLL2_P2557L_FWD	TTCACCTTCCCTCAGGCAGT
MLL2_P2557L_REV	GTGGAAGTTCCTGTGGCTA
MLL2_R1189C_FWD	CTCTGGATGGGATTGATGCT
MLL2_R1189C_REV	CTCTGAGCCAGGAAAATGG
MLL2_R1299C_FWD	TGACTCACTATGTGCGATGC
MLL2_R1299C_REV	CACAAAGCAAGGTGGGAAAG
MLL3_E3662G_FWD	TATAACTGCCCCACCGACTC
MLL3_E3662G_REV	GGTTTCCATGGAGAGCTTGT
MPL_T374A_FWD	CTGCCACTTCAAGTCACGAA
MPL_T374A_REV	GGCACAGGGTCAGATTCAGT
MPL_T481S_FWD	GCGATCTCGCTACCGTTTAC
MPL_T481S_R	CACCAAGGAGATCCAGGCTA
MSR1_D174Y_FWD	GGAACACATGAGCAACATGG
MSR1_D174Y_REV	GTGAGCTGCAAAATCAAGCA
MSR1_R293STOP_FWD	CCCATTTTTCTATGCTTTACAA
MSR1_R293STOP_REV	AGCAATCCTCCCCTACACAT
NCOA7_D219E_FWD	AGATGGGGCAAGAGAGAAT
NCOA7_D219E_REV	TGCCCTCTACATGGACCTTC
NCSTN_A298S_FWD	CTCCTGGGTTGTCTCCATTG
NCSTN_A298S_REV	GGTCACATCAGGTGCCTTTT
NF1_A2315T_FWD	TCTGCACAAAGCCCTCTTTT
NF1_A2315T_REV	AGGAACCTCAAGGCAAAGTT
NF1_FWD	TGAGTTTTAGAGGCTGTTAATTTGC
NF1_REV	TTGCAGCACTTCTGTGTCAGC
NOTCH1_S1588G_FWD	ACTGCAAGGACCCTTCAGC
NOTCH1_S1588G_R	CACGCTTGAAGACCACGTT
NOTCH1_V1671I_FWD	GCAACAGCTCCTTCCACTTC
NOTCH1_V1671I_REV	CGGAACCTCCGTCTCTTTTA
NOTCH1_V1676I_FWD	GTCTGACTGTGGCCTCAT
NOTCH1_V1676I_REV	CCTCGATCTTGTAGGGGATG
NOTCH1_V2536I_FWD	ACAGCTACTCCTCGCCTGTG
NOTCH1_V2536I_REV	TACTTGAAGGCTCCGGAAT
NPM1-F	GTGTTGTGGTTCCTTAACCACAT
NPM1-R (TYPE A)	TCCTCCACTGCCAGACAGAG
NPM1-R (TYPE D)	CTCCACTGCCAGGCAGAG
NRAS_G12S_FWD	GATGTGGCTCGCCAATTAAC
NRAS_G12S_REV	CACTGGGCCTCACCTCTATG
NRAS_G13D_FWD	GGTTTCCAACAGGTTCTTGC
NRAS_G13D_REV	CACTGGGCCTCACCTCTATG
NRAS_Q61H_F	CACCCCAAGGATTCTTACAG
NRAS_Q61H_R	CCCCATAAAGATTCAGAACACA
NSD1_F1203C_FWD	GCCCAGAGCTGGACTCTGTA
NSD1_F1203C_REV	GAATTCCTGGCTCCTTTTCC
NUMA1_N314D_FWD	TGCGTGACAAGAATGAGAGG
NUMA1_N314D_REV	CACCACCTTAACACCCACCT
OXCT1_R285Q_F	TGGGAGAGAGCTCATCTTTTTT
OXCT1_R285Q_R	CATGCCATCCTCAAACCTCAA
PRIMPOL_INDEL_F	CATCACCTAGCAGGCTGTCA
PRIMPOL_INDEL_R	AAAAACAAAAATTGGAAGCAAA
PTPN11_A72V_FWD	ATGGAGCTGTCAACCACATC
PTPN11_A72V_REV	TCTGACACTCAGGGCACAAG
PTPN11_F285S_FWD	TGGACTAGGCTGGGGAGTAA
PTPN11_F285S_REV	TTTCAGGACATGAGGAAGGA
PTPN11_Q510H_FWD	CCTGGCTCTGCAGTTTCTCT
PTPN11_Q510H_REV	CCTGTCTCCTGCTCAAAAG
RALBP1_INDEL_F	TGCGAGACCTTCCAGAGAAT
RALBP1_INDEL_R	CCATGTACCTGCACAGTTGG
RB1_A15T_FWD	GGTTTTTCTCAGGGGACGTT
RB1_A15T_REV	CTGTCTGCTCTGGGTCCT
RET_G733D_F	GGAGGATCCAAAGTGGGAAT
RET_G733D_R	AGCCGCTCTAGAACAGCATT
RUNX1_ex1_FWD	GCTGAAACAGTGACCTGTCTTGG
RUNX1_ex1_REV	AGCTGCCATTTTCATTACAGGCAAGC
RUNX1_ex2_FWD	CTCCAGTGCTAAAAAGTGTAAAG
RUNX1_ex2_REV	CATATACACATCTATGAAGGTGTGTAC
RUNX1_ex3_FWD	AGCTGCTTGTGTAAGATCCG
RUNX1_ex3_REV	GCCTGTCTCCACCACCTCTCTC
RUNX1_ex4_FWD	CAACCTAAAAAGAAATCATTGAA
RUNX1_ex4_REV	CCGAGTTTCTAGGGATTCCA
RUNX1_ex5_FWD	GAAGGGCTGGACAGCATAAA

RUNX1_ex5_REV	GAAAGGTTGAACCCAAGGAA
RUNX1_ex6_fwd	CCCCAGTTTTAGGAAATCCA
RUNX1_ex6_FWD	CCCCAGTTTTAGGAAATCCA
RUNX1_ex6_rev	AGTTGGTCTGGGAAGGTGTG
RUNX1_ex6_REV	AGTTGGTCTGGGAAGGTGTG
RUNX1_ex7_FWD	AACAGAGGCAGATACTGGACT
RUNX1_ex7_REV	CTTCATGCACCTCTAGTCTCCTG
RUNX1_ex8_FWD	CCATCCTCCTAGGCGGTAT
RUNX1_ex8_REV	GGCTGGTCCGGAACAGGAG
RUNX1_R177Q_FWD	GCCACCAACCTCATTCTGT
RUNX1_R177Q_R	GGTTGAACCCAAGGAATCTG
SETBP1_C1287I_FWD	CTTTGAGGTGGACACCCTGT
SETBP1_C1287I_REV	CGTTCCAAAGCCTTCATAGC
SETBP1_G870S_FWD	GAAGCTGTCTCCACCAGAC
SETBP1_G870S_REV	CCAGGGAGCAGAAATCAAAA
SF3A1_N78T_FWD	GTCCAATGGGCAGTGTCT
SF3A1_N78T_REV	TTCCCTTGAACCTCGCTGACCT
SF3A1_P589A_FWD	TCTCATCAGCCCTGAGTCCT
SF3A1_P589A_REV	GAACTCCTCCTCTGGCATGA
SH2B2_A392G_FWD	TTGTCCCCCTTTCTCCTTC
SH2B2_A392G_REV	GTCTCACTTTGGCGGATCAC
SH2B3_L295V_FWD	TACAGCAGACCCAACCTGT
SH2B3_L295V_R	TGCATCTCTGCTTCTGTGCT
SHKBP1_G344S_FWD	ACGGAAGGAAAGGAATGGAA
SHKBP1_G344S_REV	GAGCGGGTTTGTCCAAGT
SMC1A_R699C_FWD	GTCAGCCACCACCTTACCAC
SMC1A_R699C_FWD	GTCAGCCACCACCTTACCAC
SMC1A_R699C_R	ACCTAGGCCAGGAATGTGTG
SMC1A_R699C_REV	ACCTAGGCCAGGAATGTGTG
SOS2_T654A_F	CCACTTTTCGTGGAAGTGGT
SOS2_T654A_R	CTGGTTGGACATATTCCTTGC
SPI1_R171S_FWD	GAGTGTGGGTGCGTTGTGTA
SPI1_R171S_R	CACCAGATGCTGTCCCTTCAT
SRSF2_A62T_FWD	CGCCAGTTGTTACTCAGGT
SRSF2_A62T_REV	CCACCTTGAGGGAGGTCATA
SRSF2_DEL99R_FWD	GACCGCTACACCAAGGAGTC
SRSF2_DEL99R_REV	CCTCAGCCCCGTTTACCT
SRSF2_T132I_FWD	ATGGCATCCATAGCGTCCCT
SRSF2_T132I_REV	CAAGGTGGACAACCTGACCT
SUZ12_N277S_FWD	GCAAAAATGGAAACATGTGG
SUZ12_N277S_REV	TTGTGCAACAAATGCTTTTCC
TEK_I718K_FWD	AAGGGCCTAGAGCCTGAAAC
TEK_I718K_REV	CCTATAGGGCTGCACGGTAA
TERC_E20K_FWD	GCTCCCTTTATAAGCCGACTC
TERC_E20K_REV	GCTGAAAGTCAGCGAGAAAA
TET2_A1605S_FWD	TTCTGGATCCACCAATCCAT
TET2_A1605S_REV	ATATGGGGAGCAGTTGTCCA
TET2_C1382Y_FWD	TGTCATTCCATTTTGTTCCTGG
TET2_C1382Y_REV	GCATGAGCACAGAAGTCCAA
TET2_C332STOP_FWD	TGTGATGCTGATGATGCTGA
TET2_C332STOP_REV	ATTGCTGCTGGAACCTGAAC
TET2_FS_FWD	ACATGCTCCATGAACAACCA
TET2_FS_FWD	CAAAATCAAGCGAGTTCGAG
TET2_FS_REV	GTGTGGTTATGCCACAGCTT
TET2_FS_REV	GGGTCTTGTTCCTGCAAAA
TET2_FWD	ACCACCAATCTGAGCAATC
TET2_H1912Y_FWD	CAATAGGAATCACCCACCA
TET2_H1912Y_REV	AGGCTCCCGTTTCACTTTT
TET2_Q803STOP_FWD	ACTTTTCTCACCCCAAAG
TET2_Q803STOP_REV	TGCACTTGATTTTCATGGTCTG
TET2_R1261C_FWD	GTGAGGCTGCAGTGATTGTG
TET2_R1261C_REV	ACCAAAGATTGGGCTTTCCCT
TET2_REV	TGTAAGCCTCCTTGACACA
TET2_S1203R_FWD	TGCAAGTGACCTTGTTTTG
TET2_S1203R_REV	ATTCCTCAGCGTCTCGGTA
TET2_W1003_FWD	TTAAGGTGGAACCTGGATGC
TET2_W1003_REV	ACTGGCCCTGACATTTCAAC
TET2_Y1294C_FWD	AGGAGAAGTTGCGCCTGTC
TET2_Y1294C_R	CAAACCTCTTTTGGGTCATCC
tet21216stop_fwd	TGCAAGTGACCTTGTTTTG

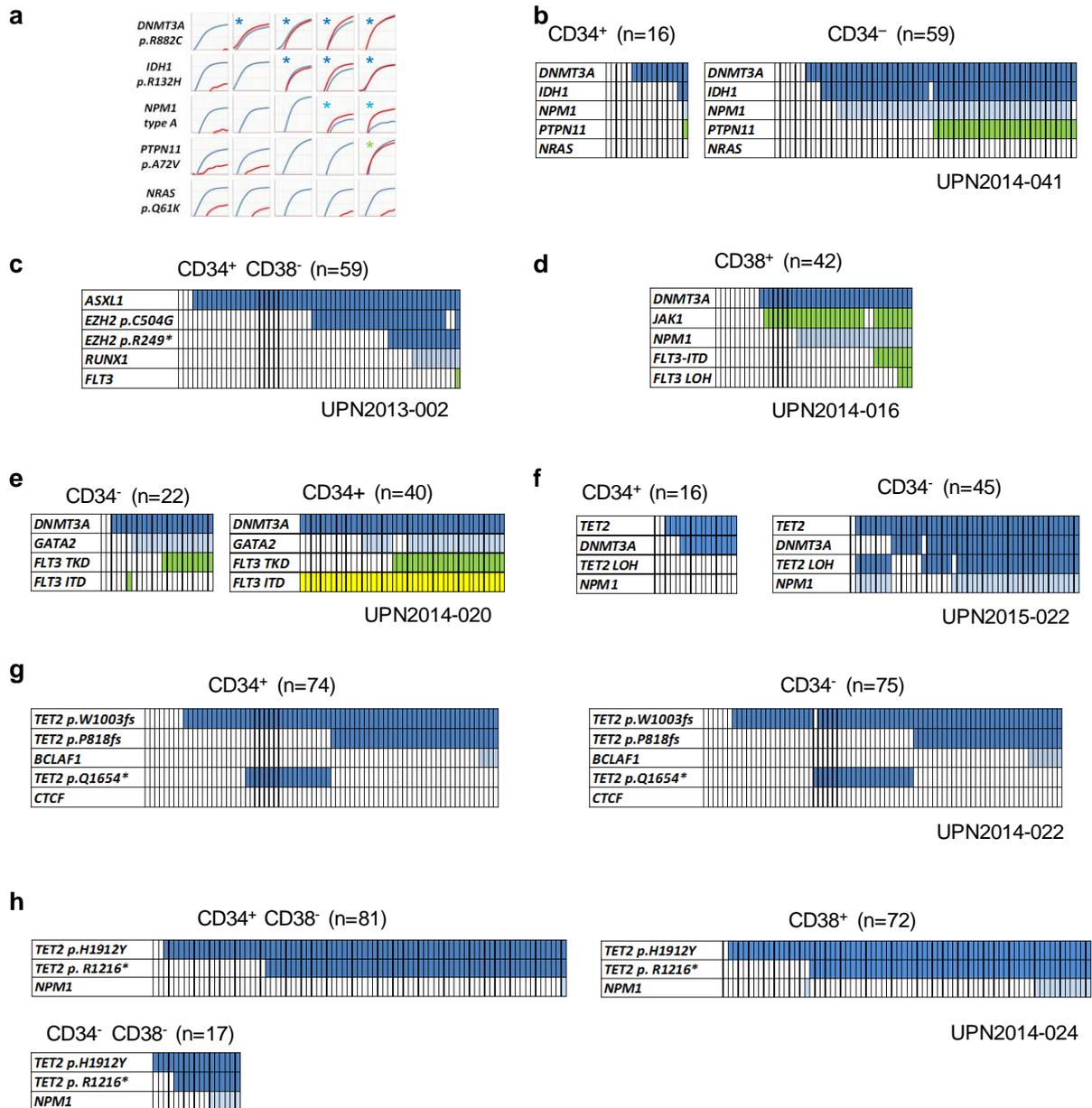
tet21216stop_fwd	ATTTCCTCAGCGTCTCGGTA
tet2fs_fwd	TAGCCACACCCAGCTTTAG
tet2fs_rev	TTCTGGATAAACGCCATGTG
TET3_FWD	CTCTGAAGGGTGGATTGTCC
TET3_P175A_FWD	GAAACCACCCAACTGCAACT
TET3_P175A_REV	GAGAGATCTGGGGCACCTCT
TET3_P512I_FWD	GACAGACCACCCAGGAGAA
TET3_P512I_REV	AAGCCCTTCCAGGACAATCT
TET3_R577Q_FWD	CCTTCTCTTGCCTATTTGC
TET3_R577Q_REV	GGGAGACATGTTTGCTGGTT
TET3_R899STOP_FWD	ATGTGTGACTGCCCTCTCT
TET3_R899STOP_REV	ACTCTCCCACTTCCCACCTT
TET3_REV	ACTTGGAGTGGAGCCATTG
tp53_fwd	CCAACTGGCCAAGACCTG
TP53_FWD	GGGAGTAGATGGAGCCTGGT
TP53_R141H_FWD	CAAGGGTGGTTGGGAGTAGA
TP53_R141H_REV	TTGCGGAGATTCTCTCCCTC
tp53_rev	CTTAACCCCTCCTCCAGAG
TP53_REV	GCTTCTTGTCTGCTTGCTT
TYK2_LOB_FWD	CAGGGAGGGTGAAGTACCTGA
TYK2_LOB_REV	CCGTCAAAGCAGATCTCCAG
U2AF1_S34F_FWD	TGCTGCTGACATATCCATGT
U2AF1_S34F_REV	AGTCGATCACCTGCCCTCACT
WT1_CAST_FWD	GCCGAGGCTAGACCTTCTCT
WT1_CAST_REV	TCCAATCCCTCTCATCACAA
WT1_FWD	GGAGTGTGAATGGGAGTGGT
WT1_PES_FWD	GGAGTGTGAATGGGAGTGGT
WT1_PES_REV	TTATTGCAGCCTGGGTAAGC
WT1_R140STOP_FWD	GGAGTGTGAATGGGAGTGGT
WT1_R140STOP_REV	TTATTGCAGCCTGGGTAAGC
WT1_REV	TTATTGCAGCCTGGGTAAGC
WT1_Y201*_FWD	CGCCTTCACTGTCCACTTTT
WT1_Y201*_REV	CCGGCCTACTTACCCTGATT
XRCC5_V42I_F	GACGTGGGCTTTACCATGAG
XRCC5_V42I_R	AACAGAGGGGCTGGACACTA
ZRSR2_E133V_FWD	GACCCGAAGAAGAGCATCAG
ZRSR2_E133V_REV	ATGGGAAGAACAGCAGAGGA
ZRSR2_V304M_FWD	CAAGCAGCCCTTCTCTGT
ZRSR2_V304M_REV	AGGCAAAAAGCTATGCCTCA

Supplementary table 11 | Primers and probes for allele specific PCR assays

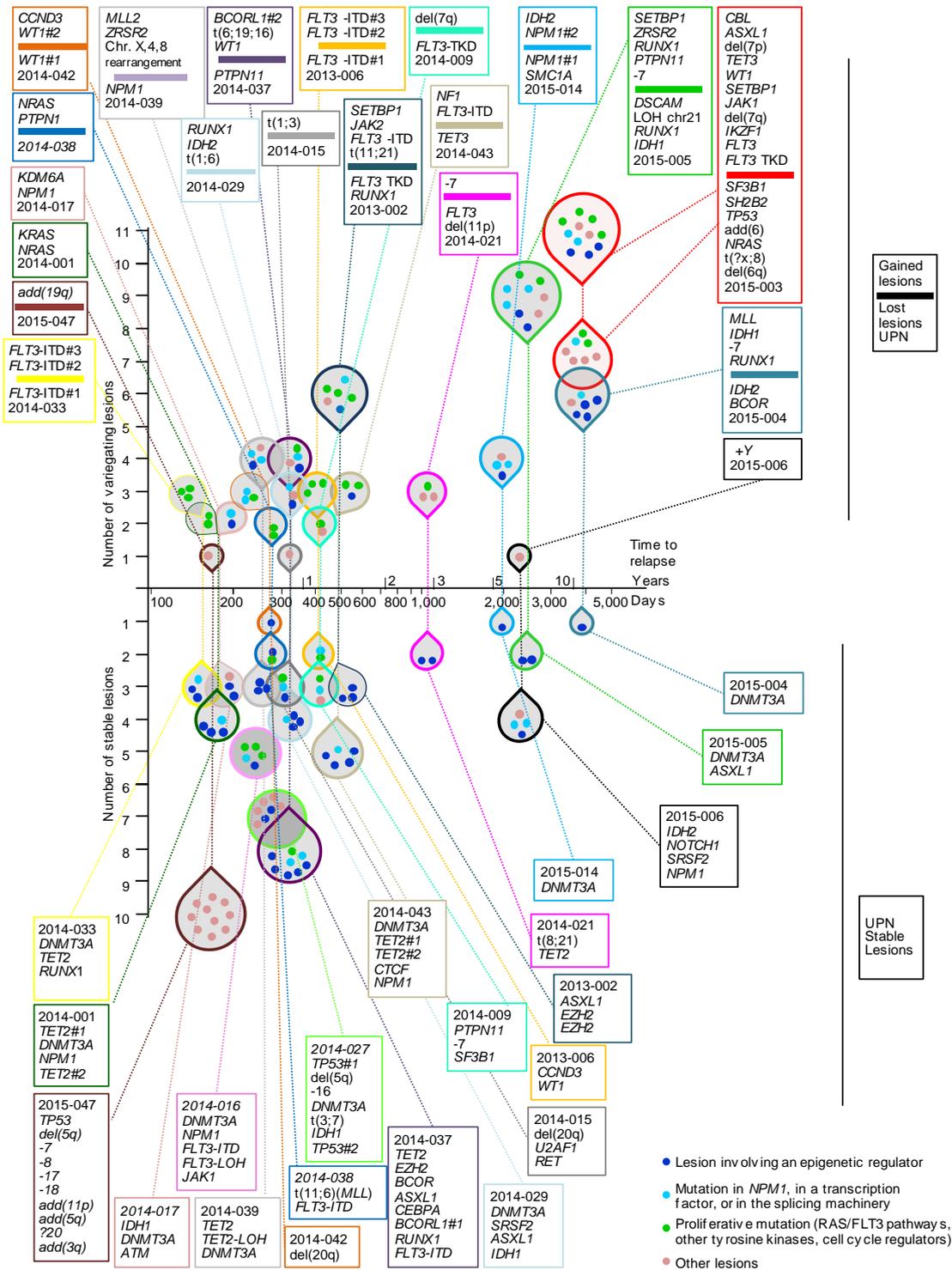
Forward Primer Name	Forward Primer Seq.	Reverse Primer Name	Reverse Primer Seq.	Reporter 1 Name	Reporter 1 Sequence	Reporter 2 Name	Reporter 2 Sequence
ABCC2_1668A_F	GGACATATGCGAGCCGACCTT	ABCC2_1668A_R	ATGGCTGATATACAGGAGGATTC	ABCC2_1668A_V	CAGGCGATATCAGACGC	ABCC2_1668A_M	AGGCGCTATATCCAGCC
ASXL1_FS_F	GGAGGAGCCGACGACGAC	ASXL1_FS_R	GGATATGATGTTTGGCTGTGAG	ASXL1_FS_V	GGATATGATGTTTGGCTGTGAG	ASXL1_FS_M	ATCTGATGATTTTCCAG
ASXL1_G873stop_F	GGTGGGATGCGCAGAGAGAGGT	ASXL1_G873stop_R	GGACGACAGCGAGCACTTTTGT	ASXL1_G873stop_V	GGACGACAGCGAGCACTTTTGT	ASXL1_G873stop_M	CGACGCTACTACTGCAAT
BCLAF1_R107C_F	CTGTGATGATGATGATGATGATGAT	BCLAF1_R107C_R	CTGTGATGATGATGATGATGATGAT	BCLAF1_R107C_V	CTGTGATGATGATGATGATGATGAT	BCLAF1_R107C_M	GTGATGATGATGATGATGAT
CDZ2_G88E_F	GGAGATGATGATGATGATGATGAT	CDZ2_G88E_R	CACTGATGATGATGATGATGATGAT	CDZ2_G88E_V	CACTGATGATGATGATGATGATGAT	CDZ2_G88E_M	AGATTTGATGATGATGATGATGAT
ck1_F	CTCCGATGCTGCTGCTGCA	ck1_R	GATGGGCTGCTGCTGCTGCTGCA	ck1_V	GATGGGCTGCTGCTGCTGCTGCA	ck1_M	AGGCTGCTGCTGCTGCTGCA
DNMT3A_F	GGTGGAGCTGTGGATTTTACC	DNMT3A_R	CTTGGAGCTGTGGATTTTACC	DNMT3A_V	GGTGGAGCTGTGGATTTTACC	DNMT3A_M	AGGCTGCTGCTGCTGCTGCA
dnmt3a_fs	GGTGGAGCTGTGGATTTTACC	dnmt3a_rs	GGTGGAGCTGTGGATTTTACC	dnmt3a_vs	GGTGGAGCTGTGGATTTTACC	dnmt3a_ms	AGGCTGCTGCTGCTGCTGCA
DNMT3A_L548H_F	GGTGGAGCTGTGGATTTTACC	DNMT3A_L548H_R	GGTGGAGCTGTGGATTTTACC	DNMT3A_L548H_V	GGTGGAGCTGTGGATTTTACC	DNMT3A_L548H_M	AGGCTGCTGCTGCTGCTGCA
DNMT3A_R882C_F	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882C_R	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882C_V	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882C_M	AGGCTGCTGCTGCTGCTGCA
DNMT3A_R882H_F	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882H_R	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882H_V	GGTGGAGCTGTGGATTTTACC	DNMT3A_R882H_M	AGGCTGCTGCTGCTGCTGCA
EZH2_R248stop_F	GGTGGAGCTGTGGATTTTACC	EZH2_R248stop_R	GGTGGAGCTGTGGATTTTACC	EZH2_R248stop_V	GGTGGAGCTGTGGATTTTACC	EZH2_R248stop_M	AGGCTGCTGCTGCTGCTGCA
FANCA_G809D_F	GGTGGAGCTGTGGATTTTACC	FANCA_G809D_R	GGTGGAGCTGTGGATTTTACC	FANCA_G809D_V	GGTGGAGCTGTGGATTTTACC	FANCA_G809D_M	AGGCTGCTGCTGCTGCTGCA
FLT3_D835V_F	GGTGGAGCTGTGGATTTTACC	FLT3_D835V_R	GGTGGAGCTGTGGATTTTACC	FLT3_D835V_V	GGTGGAGCTGTGGATTTTACC	FLT3_D835V_M	AGGCTGCTGCTGCTGCTGCA
FLT3_M664L_F	GGTGGAGCTGTGGATTTTACC	FLT3_M664L_R	GGTGGAGCTGTGGATTTTACC	FLT3_M664L_V	GGTGGAGCTGTGGATTTTACC	FLT3_M664L_M	AGGCTGCTGCTGCTGCTGCA
FLT3_S451F_F	GGTGGAGCTGTGGATTTTACC	FLT3_S451F_R	GGTGGAGCTGTGGATTTTACC	FLT3_S451F_V	GGTGGAGCTGTGGATTTTACC	FLT3_S451F_M	AGGCTGCTGCTGCTGCTGCA
fk3_k4_F	GGTGGAGCTGTGGATTTTACC	fk3_k4_R	GGTGGAGCTGTGGATTTTACC	fk3_k4_V	GGTGGAGCTGTGGATTTTACC	fk3_k4_M	AGGCTGCTGCTGCTGCTGCA
GAT42_N817S_F	GGTGGAGCTGTGGATTTTACC	GAT42_N817S_R	GGTGGAGCTGTGGATTTTACC	GAT42_N817S_V	GGTGGAGCTGTGGATTTTACC	GAT42_N817S_M	AGGCTGCTGCTGCTGCTGCA
JAG1_T173M_F	GGTGGAGCTGTGGATTTTACC	JAG1_T173M_R	GGTGGAGCTGTGGATTTTACC	JAG1_T173M_V	GGTGGAGCTGTGGATTTTACC	JAG1_T173M_M	AGGCTGCTGCTGCTGCTGCA
KRAS_G13D_F	GGTGGAGCTGTGGATTTTACC	KRAS_G13D_R	GGTGGAGCTGTGGATTTTACC	KRAS_G13D_V	GGTGGAGCTGTGGATTTTACC	KRAS_G13D_M	AGGCTGCTGCTGCTGCTGCA
KRAS_Q61H_F	GGTGGAGCTGTGGATTTTACC	KRAS_Q61H_R	GGTGGAGCTGTGGATTTTACC	KRAS_Q61H_V	GGTGGAGCTGTGGATTTTACC	KRAS_Q61H_M	AGGCTGCTGCTGCTGCTGCA
MYBL1_R24M_F	GGTGGAGCTGTGGATTTTACC	MYBL1_R24M_R	GGTGGAGCTGTGGATTTTACC	MYBL1_R24M_V	GGTGGAGCTGTGGATTTTACC	MYBL1_R24M_M	AGGCTGCTGCTGCTGCTGCA
npn1_F	GGTGGAGCTGTGGATTTTACC	npn1_R	GGTGGAGCTGTGGATTTTACC	npn1_V	GGTGGAGCTGTGGATTTTACC	npn1_M	AGGCTGCTGCTGCTGCTGCA
NPM1_type_D_F	GGTGGAGCTGTGGATTTTACC	NPM1_type_D_R	GGTGGAGCTGTGGATTTTACC	NPM1_type_D_V	GGTGGAGCTGTGGATTTTACC	NPM1_type_D_M	AGGCTGCTGCTGCTGCTGCA
NPM1_Q61H_F	GGTGGAGCTGTGGATTTTACC	NPM1_Q61H_R	GGTGGAGCTGTGGATTTTACC	NPM1_Q61H_V	GGTGGAGCTGTGGATTTTACC	NPM1_Q61H_M	AGGCTGCTGCTGCTGCTGCA
OXC11_R285Q_F	GGTGGAGCTGTGGATTTTACC	OXC11_R285Q_R	GGTGGAGCTGTGGATTTTACC	OXC11_R285Q_V	GGTGGAGCTGTGGATTTTACC	OXC11_R285Q_M	AGGCTGCTGCTGCTGCTGCA
PTPN11_F	GGTGGAGCTGTGGATTTTACC	PTPN11_R	GGTGGAGCTGTGGATTTTACC	PTPN11_V	GGTGGAGCTGTGGATTTTACC	PTPN11_M	AGGCTGCTGCTGCTGCTGCA
RALBP1_FS_F	GGTGGAGCTGTGGATTTTACC	RALBP1_FS_R	GGTGGAGCTGTGGATTTTACC	RALBP1_FS_V	GGTGGAGCTGTGGATTTTACC	RALBP1_FS_M	AGGCTGCTGCTGCTGCTGCA
RET_G733D_F	GGTGGAGCTGTGGATTTTACC	RET_G733D_R	GGTGGAGCTGTGGATTTTACC	RET_G733D_V	GGTGGAGCTGTGGATTTTACC	RET_G733D_M	AGGCTGCTGCTGCTGCTGCA
RUNX1_D98fs_F	GGTGGAGCTGTGGATTTTACC	RUNX1_D98fs_R	GGTGGAGCTGTGGATTTTACC	RUNX1_D98fs_V	GGTGGAGCTGTGGATTTTACC	RUNX1_D98fs_M	AGGCTGCTGCTGCTGCTGCA
SOS2_T654A_F	GGTGGAGCTGTGGATTTTACC	SOS2_T654A_R	GGTGGAGCTGTGGATTTTACC	SOS2_T654A_V	GGTGGAGCTGTGGATTTTACC	SOS2_T654A_M	AGGCTGCTGCTGCTGCTGCA
TE121003fs_F	GGTGGAGCTGTGGATTTTACC	TE121003fs_R	GGTGGAGCTGTGGATTTTACC	TE121003fs_V	GGTGGAGCTGTGGATTTTACC	TE121003fs_M	AGGCTGCTGCTGCTGCTGCA
SPH1_R171S_F	GGTGGAGCTGTGGATTTTACC	SPH1_R171S_R	GGTGGAGCTGTGGATTTTACC	SPH1_R171S_V	GGTGGAGCTGTGGATTTTACC	SPH1_R171S_M	AGGCTGCTGCTGCTGCTGCA
1b2_js_f	GGTGGAGCTGTGGATTTTACC	1b2_js_r	GGTGGAGCTGTGGATTTTACC	1b2_js_v	GGTGGAGCTGTGGATTTTACC	1b2_js_m	AGGCTGCTGCTGCTGCTGCA
TE12_FS_F	GGTGGAGCTGTGGATTTTACC	TE12_FS_R	GGTGGAGCTGTGGATTTTACC	TE12_FS_V	GGTGGAGCTGTGGATTTTACC	TE12_FS_M	AGGCTGCTGCTGCTGCTGCA
TE12_H1912V_F	GGTGGAGCTGTGGATTTTACC	TE12_H1912V_R	GGTGGAGCTGTGGATTTTACC	TE12_H1912V_V	GGTGGAGCTGTGGATTTTACC	TE12_H1912V_M	AGGCTGCTGCTGCTGCTGCA
TE12_R1216sL_F	GGTGGAGCTGTGGATTTTACC	TE12_R1216sL_R	GGTGGAGCTGTGGATTTTACC	TE12_R1216sL_V	GGTGGAGCTGTGGATTTTACC	TE12_R1216sL_M	AGGCTGCTGCTGCTGCTGCA
TE12_1294c_F	GGTGGAGCTGTGGATTTTACC	TE12_1294c_R	GGTGGAGCTGTGGATTTTACC	TE12_1294c_V	GGTGGAGCTGTGGATTTTACC	TE12_1294c_M	AGGCTGCTGCTGCTGCTGCA
TP53_181c_F	GGTGGAGCTGTGGATTTTACC	TP53_181c_R	GGTGGAGCTGTGGATTTTACC	TP53_181c_V	GGTGGAGCTGTGGATTTTACC	TP53_181c_M	AGGCTGCTGCTGCTGCTGCA
U2AF1_S34F_F	GGTGGAGCTGTGGATTTTACC	U2AF1_S34F_R	GGTGGAGCTGTGGATTTTACC	U2AF1_S34F_V	GGTGGAGCTGTGGATTTTACC	U2AF1_S34F_M	AGGCTGCTGCTGCTGCTGCA
XRCC5_V42I_F	GGTGGAGCTGTGGATTTTACC	XRCC5_V42I_R	GGTGGAGCTGTGGATTTTACC	XRCC5_V42I_V	GGTGGAGCTGTGGATTTTACC	XRCC5_V42I_M	AGGCTGCTGCTGCTGCTGCA
PTPN11_A72V_F	GGTGGAGCTGTGGATTTTACC	PTPN11_A72V_R	GGTGGAGCTGTGGATTTTACC	PTPN11_A72V_V	GGTGGAGCTGTGGATTTTACC	PTPN11_A72V_M	AGGCTGCTGCTGCTGCTGCA
NRAS_Q61K_F	GGTGGAGCTGTGGATTTTACC	NRAS_Q61K_R	GGTGGAGCTGTGGATTTTACC	NRAS_Q61K_V	GGTGGAGCTGTGGATTTTACC	NRAS_Q61K_M	AGGCTGCTGCTGCTGCTGCA
NPM1_1b9a_F	GGTGGAGCTGTGGATTTTACC	NPM1_1b9a_R	GGTGGAGCTGTGGATTTTACC	NPM1_1b9a_V	GGTGGAGCTGTGGATTTTACC	NPM1_1b9a_M	AGGCTGCTGCTGCTGCTGCA
IDH1_R132H_F	GGTGGAGCTGTGGATTTTACC	IDH1_R132H_R	GGTGGAGCTGTGGATTTTACC	IDH1_R132H_V	GGTGGAGCTGTGGATTTTACC	IDH1_R132H_M	AGGCTGCTGCTGCTGCTGCA
RUNX1FS_F	GGTGGAGCTGTGGATTTTACC	RUNX1FS_R	GGTGGAGCTGTGGATTTTACC	RUNX1FS_V	GGTGGAGCTGTGGATTTTACC	RUNX1FS_M	AGGCTGCTGCTGCTGCTGCA
EZH2_C504G_F	GGTGGAGCTGTGGATTTTACC	EZH2_C504G_R	GGTGGAGCTGTGGATTTTACC	EZH2_C504G_V	GGTGGAGCTGTGGATTTTACC	EZH2_C504G_M	AGGCTGCTGCTGCTGCTGCA
SF3B1_K666N_F	GGTGGAGCTGTGGATTTTACC	SF3B1_K666N_R	GGTGGAGCTGTGGATTTTACC	SF3B1_K666N_V	GGTGGAGCTGTGGATTTTACC	SF3B1_K666N_M	AGGCTGCTGCTGCTGCTGCA
IAK1_L1053V_F	GGTGGAGCTGTGGATTTTACC	IAK1_L1053V_R	GGTGGAGCTGTGGATTTTACC	IAK1_L1053V_V	GGTGGAGCTGTGGATTTTACC	IAK1_L1053V_M	AGGCTGCTGCTGCTGCTGCA
FLT3_D839E_F	GGTGGAGCTGTGGATTTTACC	FLT3_D839E_R	GGTGGAGCTGTGGATTTTACC	FLT3_D839E_V	GGTGGAGCTGTGGATTTTACC	FLT3_D839E_M	AGGCTGCTGCTGCTGCTGCA
TE12_Q1654sL_F	GGTGGAGCTGTGGATTTTACC	TE12_Q1654sL_R	GGTGGAGCTGTGGATTTTACC	TE12_Q1654sL_V	GGTGGAGCTGTGGATTTTACC	TE12_Q1654sL_M	AGGCTGCTGCTGCTGCTGCA
CTCF_R14C_F	GGTGGAGCTGTGGATTTTACC	CTCF_R14C_R	GGTGGAGCTGTGGATTTTACC	CTCF_R14C_V	GGTGGAGCTGTGGATTTTACC	CTCF_R14C_M	AGGCTGCTGCTGCTGCTGCA
TE12_L627s_F	GGTGGAGCTGTGGATTTTACC	TE12_L627s_R	GGTGGAGCTGTGGATTTTACC	TE12_L627s_V	GGTGGAGCTGTGGATTTTACC	TE12_L627s_M	AGGCTGCTGCTGCTGCTGCA

Supplementary table 12 | Antibodies used for flow cytometry analyses

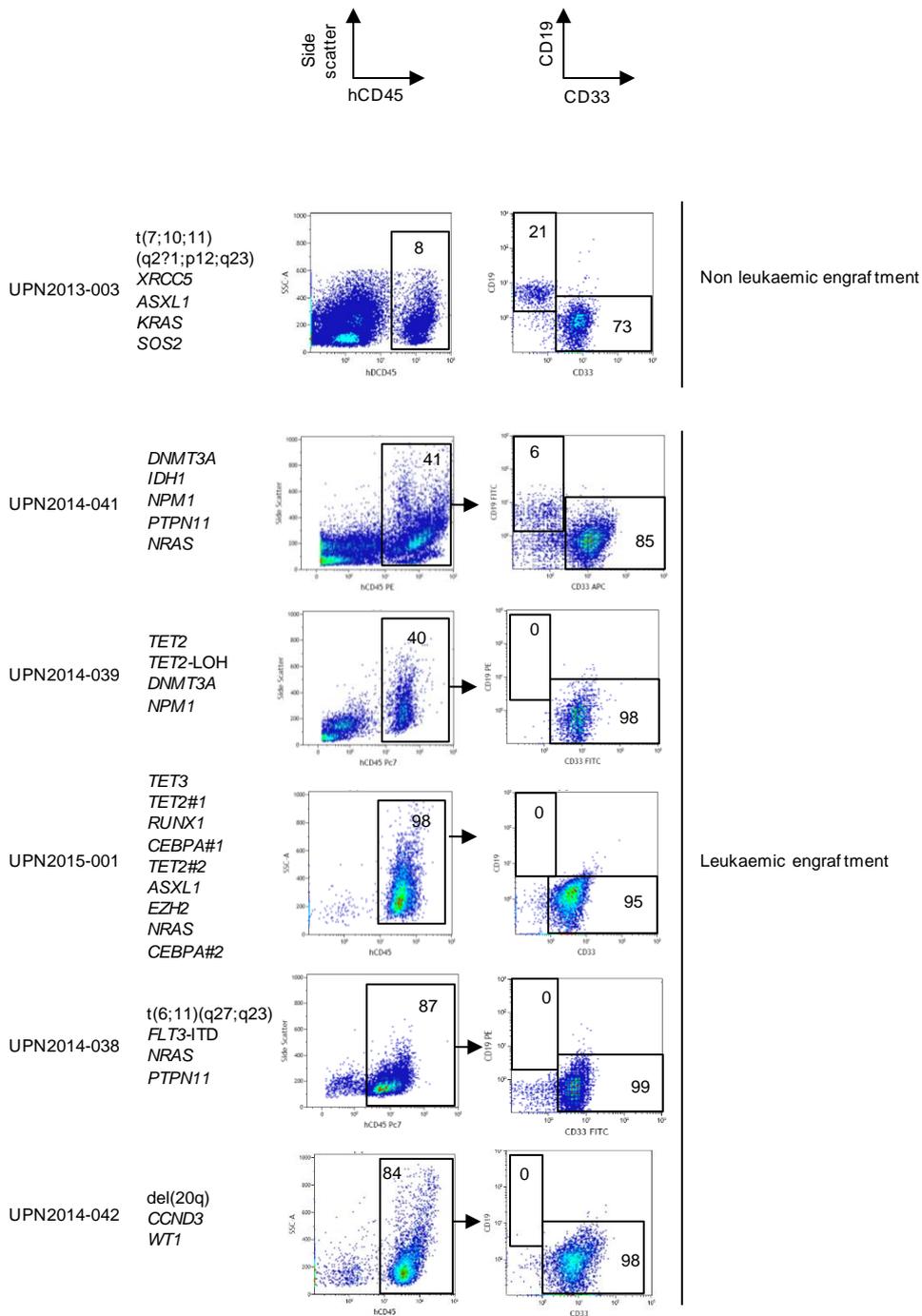
Antibody		Clone
PE-Cy TM 7 Mouse Anti-Human CD45	BD pharmingen	HI30
APC Mouse Anti-Human CD45	BD pharmingen	HI30
PE Mouse Anti-Human CD45	BD pharmingen	HI30
PE Mouse Anti-Human CD33	BD pharmingen	WM53
FITC Mouse Anti-Human CD33	BD pharmingen	HIM3-4
PerCP-Cy TM 5.5 Mouse Anti-Human CD33	BD pharmingen	P67.6
PE Mouse Anti-Human CD19	BD pharmingen	HIB19
FITC Mouse Anti-Human CD19	BD pharmingen	HIB19
PE Mouse Anti-Human CD3	BD pharmingen	UCHT1
FITC Mouse Anti-Human CD3	BD pharmingen	UCHT1
APC Mouse Anti-Human CD3	BD pharmingen	UCHT1
APC Rat anti-mouse CD45	Biolegend	30-F11
PerCP Rat anti-mouse CD45	Biolegend	30-F11
PE Mouse IgG1, κ Isotype Control	BD pharmingen	MOPC-21
PerCP-Cy TM 5.5 Mouse IgG1, κ Isotype Control	BD pharmingen	MOPC-21
PE-Cy TM 7 Mouse IgG1 κ Isotype Control	BD pharmingen	MOPC-21
FITC Mouse IgG1, κ Isotype Control	BD pharmingen	MOPC-21
APC-Cy TM 7 Rat Anti-Mouse CD45	BD pharmingen	30-F11
APC-Cy TM 7 Mouse IgG2b, κ Isotype Control	BD pharmingen	27-35
APC Rat IgG2b, κ Isotype Ctrl	Biolegend	RTK4530
PerCP Rat IgG2b, κ Isotype Ctrl	Biolegend	RTK4530



Supplementary figure 1 | Summary of single cell derived colony assays in seven patients with normal karyotype. a, PCR amplification plots of allele-specific SNP (AS) assays in five colonies (colonies: columns; mutations: rows) from UPN2014-041 as in **Fig.1. b,c,d,e,f,g,h**, Results of AS assays in 7 patients as in figure 1. For each patient, the cell phenotype before culture and the number of analysed colonies are indicated above the table. Yellow box means no data.



Supplementary Figure 3 | Temporal evolution of AML clones at relapse. Clonal composition of 22 AMLs at diagnosis and at relapse. The numbers of stable or variegating lesions are plotted as droplets above or below the time axis, respectively. Except for two cases (patients UPN2014-016 and UPN2014-027) with no changes in clonal composition (circles), coloured droplets of single patients are connected by vertical dashed lines. Droplets and circles are filled with dots that represent individual lesions from the four categories listed in the key, and summarized for each patient in connected boxes as indicated. Stable and variegating lesions are listed in coloured boxes below or above the time axis as indicated.



Supplementary Figure 4 | Leukaemic and non leukaemic repopulation of NSG bone marrow by AML cells. Flow cytometric analysis of NSG bone marrow repopulated with non leukaemic and leukaemic engraftments. Chromosomal aberrations and mutant genes detected in the injected samples are indicated.

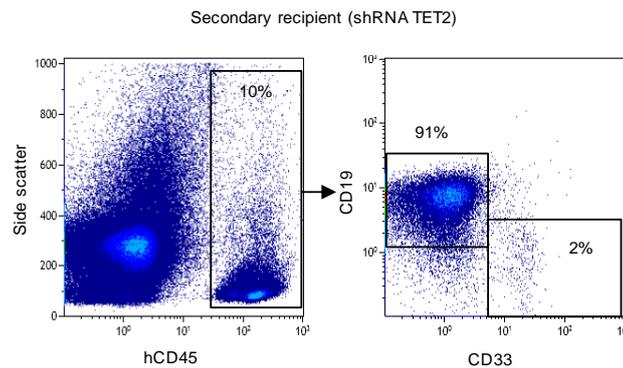
a

Experiment	Condition	Mouse	hCD45+	hCD19+	hCD33+
Exp#1	ShRNA_scramble	#136	1.2	ND	0.2
Exp#1	ShRNA_scramble	#137	12.1	ND	1.8
Exp#1	ShRNA_scramble	#138	3.5	ND	1.2
Exp#1	ShRNA_scramble	#139	0.2	ND	0.0
Exp#1	ShRNA_scramble	#140	42.3	ND	8.4
Exp#1	ShRNA_scramble	#144	12.3	ND	3.1
Exp#2	ShRNA_scramble	#910	20.4	15.4	3.4
Exp#2	ShRNA_scramble	#911	3.1	2.2	1.4
Exp#2	ShRNA_scramble	#918	13.3	6.7	7.4
Exp#2	ShRNA_scramble	#919	20.5	10.3	5.3
Exp#2	ShRNA_scramble	#921	48.5	20.5	16.2
Exp#3	ShRNA_scramble	#334	48.3	40.3	7.1
Exp#3	ShRNA_scramble	#338	37.6	29.9	5.5
Exp#3	ShRNA_scramble	#339	4.2	1.7	3.0
		median	12.8	12.8	3.2
shRNA_Scramble		mean	19.1	15.9	4.6
		SEM	4.8	4.9	1.2

Experiment	Condition	Mouse	hCD45+	hCD19+	hCD33+
Exp#1	ShRNA_TET2	#132	60.0	ND	7.5
Exp#1	ShRNA_TET2	#135	20.8	ND	0.9
Exp#1	ShRNA_TET2	#141	20.3	ND	2.2
Exp#1	ShRNA_TET2	#142	30.8	ND	2.8
Exp#1	ShRNA_TET2	#143	22.4	ND	3.4
Exp#1	ShRNA_TET2	#149	57.0	ND	5.2
Exp#2	ShRNA_TET2	#906	26.2	22.2	2.9
Exp#2	ShRNA_TET2	#908	68.5	55.2	8.9
Exp#2	ShRNA_TET2	#912	41.2	18.9	15.2
Exp#2	ShRNA_TET2	#913	75.5	62.5	8.4
Exp#2	ShRNA_TET2	#917	90.9	29.7	47.5
Exp#2	ShRNA_TET2	#920	64.4	33.1	24.4
Exp#3	ShRNA_TET2	#324	53.7	48.6	5.8
Exp#3	ShRNA_TET2	#337	70.7	58.7	9.2
Exp#3	ShRNA_TET2	#345	50.2	29.8	5.1
		median	53.7	33.1	5.8
shRNA_TET2		mean	50.2	39.8	10.0
		SEM	5.8	5.5	3.1

Mann-Whitney test	hCD45+	hCD19+	hCD33+
shRNA_scramble v s shRNA_TET2 p=	0.0003	0.0071	0.0505

b



Supplementary Figure 5 | TET2 depletion enhances the repopulation of NSG mouse bone marrow by human cord blood CD34⁺ cells. **a**, Percentages of human CD45, CD19, and CD33 positive cells in bone marrow of primary recipient mice 15-17 weeks after injection of cord blood CD34⁺ cells transduced with lentiviruses expressing shRNAs designed against scramble and TET2 sequences. Statistical analysis was performed using the Mann-Whitney test. **b**, Flow cytometric analyses of the bone marrow from a secondary recipient mouse injected with human cells sorted from the bone marrow of mice #324, #337, and #345.

Article 1 : Commentaires

Pour ce travail, la totalité des expériences de culture, de génotypage et de séquençage ont été pratiquées à l'U938 (centre de recherche Saint-Antoine). Les expériences de xénogreffe (patients et sh*TET2*) ont été pratiquées à l'Institut Gustave Roussy, par l'équipe de Fawzia Louache et Yanyan Zhang. Les sh*TET2* ont été développés par Elodie Pronier. Les analyses de génotypage post greffe (patients et sh*TET2*) ont toutes été réalisées à l'U938.

Dans cet article, par l'étude de la phylogénie à l'échelon clonal sur des LAM à caryotype normal et anormal, nous avons d'abord identifié les lésions initiatrices du clone chez 15 patients. Chez les patients à caryotype normal, les lésions initiatrices retrouvées impliquent les gènes *TET2*, *DNMT3A* et *ASXL1*, soit les principales mutations retrouvées dans l'hématopoïèse clonale liée à l'âge. Chez les patients à caryotype anormal, les anomalies récurrentes impliquant *MLL* ainsi que la délétion du chromosome 20q ont été retrouvées comme mutations initiatrices. Chez le patient 2014-009, porteur d'une délétion du chromosome 7, une mutation de *SF3B1* était l'événement initiateur. Aucun événement impliquant l'épigénétique n'a été retrouvé chez ce patient avec l'analyse des 122 gènes. Une analyse complémentaire de l'exome du patient n'a pas permis de retrouver d'autre événement, suggérant que l'anomalie de l'épissage a fondé le clone. La persistance à un très haut niveau de cette anomalie à la RC (VAF à 40 %) tandis que les autres anomalies diminuent après traitement semble confirmer cette hypothèse (voir données complémentaires – évaluation de la MRD). De même, chez le patient 2014-008, porteur d'une atypique délétion du chromosome 16q, une mutation d'*U2AF1* était le premier événement, sans qu'un événement impliquant un régulateur épigénétique ne soit retrouvé avec l'analyse de 122 gènes. La séquence d'événement n'a pas pu être étudiée à l'échelon clonal chez des patients porteurs de remaniement du CBF, à cause d'une pousse très limitée des progéniteurs en méthyl-cellulose, ne permettant pas l'analyse précise de la phylogénie. A noter cependant que des analyses partielles ont pu être menées chez un patient pédiatrique (voir données supplémentaires - phylogénie des LAM pédiatriques) et ont révélé la translocation t(8;21) comme événement initiateur. Les données de la littérature suggèrent également la nature pré-leucémique des anomalies impliquant le CBF.

Quel que soit le caryotype, l'analyse en colonies révèle une séquence similaire d'accumulation des événements génétiques, avec globalement une implication précoce des

régulateurs de l'épigénétique, puis des anomalies de la maturation (*NPM1*, facteurs de transcription, épissage) et, en dernière position, les anomalies des gènes impliqués dans la prolifération cellulaire (*FLT3*, voie *RAS*, autres tyrosines kinase). L'absence d'événement lié aux gènes de prolifération chez certains patients pourrait être liée à la stratégie de reséquencage de 122 gènes ne permettant pas de détecter les anomalies atypiques présentes dans d'autres tyrosine-kinase ou phosphatase, comme il a été décrit dans l'analyse de 200 LAM du Cancer Génome Atlas ⁷¹.

L'analyse des variations clonales entre le diagnostic et la rechute montre que les lésions précoces du clone (architecture obtenue à partir des colonies ou estimée à partir des fréquences de variant en NGS et des fréquences d'anomalie chromosomiques en FISH) sont quasiment toujours conservées entre les 2 temps, tandis que les anomalies les plus tardives sont globalement variables. Dans 3 cas de rechute très tardive (5 à 11 ans après le diagnostic), les seules anomalies conservées étaient les plus précoces de la phylogénie (mutations de *DNMT3A* et *ASXL1*), alors que tous les autres événements avaient changé. Sur l'ensemble des patients étudiés, 2 exceptions sont à noter : le patient 2014-028 qui a déclaré une LA sur cellules du greffon, dont la phylogénie est détaillée dans l'article 2, et le patient 2015-003. Ce dernier a développé secondairement une LMMC avec des caractéristiques génétiques complètement différentes de la LAM initiale. L'analyse d'un échantillon du suivi 4 ans après le traitement de la LAM ne retrouve aucune trace des mutations initiales, mais retrouve l'apparition des premières anomalies du clone présent à la rechute avec des VAF de l'ordre de 15% (*JAK1* p.L4V, *SETBP1* p.G870S, et *TET3* p.P175A) (voir données complémentaires - figure 10). Ces anomalies n'étaient pas retrouvées de manière significative au diagnostic de la LAM en séquençage en profondeur. Des analyses complémentaires (exome) sont programmées pour compléter l'exploration de ce cas et conclure formellement sur la nature secondaire ou non de la LMMC.

La persistance des anomalies précoces entre diagnostic et rechute nous a conduits à étudier les échantillons de RC. L'analyse a d'abord été pratiquée chez 12 patients dont 10 ont rechuté, et comprend un patient avec une rechute très tardive, chez qui du matériel obtenu lors de multiples points du suivi était disponible. Les anomalies les plus précoces de la phylogénie étaient toujours détectables en RC. Les anomalies plus tardives étaient inconstamment retrouvées, bien qu'elles puissent être conservées à la rechute dans un certain nombre de cas. Les anomalies majoritaires à la rechute étaient parfois détectables minoritairement au

diagnostic et en rémission. Cela implique qu'une partie des rechutes est liée à la sélection d'un sous-clone plus résistant au traitement, et déjà préexistant lors du diagnostic (ex : patient 2014-039). Les autres rechutes sont probablement liées à la ré-accumulation d'événements à partir d'un clone pré-leucémique résistant au traitement, pouvant porter un nombre variable d'anomalies génétiques. L'exemple le plus démonstratif de cette dernière situation dans notre étude est le patient 2015-014, qui a rechuté après 5 ans de RC. L'analyse séquentielle par séquençage en profondeur après 1 et 3 cures de chimiothérapie montre la persistance faible mais significative de l'anomalie fondatrice du clone (*DNMT3A* p.VG567G), tandis que les autres anomalies du diagnostic (*NPM1* type A et *SMC1A* p.R790W) ne sont pas retrouvées. Après 4 ans de RC, cette anomalie initiatrice est retrouvée à un haut niveau, et une mutation d'*IDH2* p.R140L non détectée au diagnostic (séquençage en profondeur) est également retrouvée, avec une minime fraction de mutation de *NPM1* type A. A la rechute, 8 mois plus tard, les 3 événements sont retrouvés avec des VAF élevées. Les VAF et la séquence d'analyse suggère une ré-accumulation de mutations à partir du clone pré-leucémique, avec l'apparition d'une nouvelle mutation de *NPM1* de type A à la rechute.

L'analyse des capacités de repopulation hématopoïétique post xénogreffe est un des seuls tests fonctionnels pour déterminer le caractère pré-leucémique des cellules injectées. La repopulation multi-lignées après xénogreffe est considérée comme une preuve fonctionnelle de la nature pré-leucémique des lésions. Dans notre étude, la majorité des échantillons porteurs de translocation de *MLL*, ou de mutations d'*ASXL1*, de *DNMT3A* ou de *TET2* ont une capacité de reconstitution multi-lignées ou de reconstitution leucémique, démontrant les capacités pré-leucémiques d'une partie au moins des cellules injectées. La déplétion isolée en *TET2* par utilisation de shRNA dans des CSH humaines reproduit cet avantage après xénogreffe dans des souris, démontrant formellement le potentiel pré-leucémique des mutations perte de sens de *TET2*. De manière notable, ces souris n'ont pas développé de LAM ni après greffe primaire, ni après greffe secondaire. Les cellules humaines après greffe primaire et secondaire ont été analysées par reséquencage ciblé (panel de 122 gènes). Il n'a été retrouvé aucune mutation de ces gènes, dans aucun des échantillons (sacrifice des individus au bout de 16 semaines). Un temps plus prolongé pourrait être nécessaire pour atteindre un nombre suffisant d'erreurs lors des mitoses des CSH, et accumuler des lésions secondaires dans des gènes entraînant une transformation leucémique. A noter par ailleurs que 2/4 échantillons porteurs d'inv(16) ont conduit à une reconstitution hématopoïétique (multi-lignées ou leucémique). Au bilan l'étude des xénogreffes démontre des capacités

fonctionnelles des CSH pré-leucémique en cas de mutation de *TET2*, *d'ASXL1*, ou de *DNMT3A*, mais aussi en cas de translocation impliquant *MLL*.

L'analyse des profils d'association des lésions génétiques nous a permis de mettre en évidence plusieurs ontogénies différentes dans les LAM. Dans 27/72 cas, nous retrouvons un profil d'évolution après HCLA : les lésions initiatrices sont des mutations de *DNMT3A*, de *TET2* ou *d'ASXL1*, puis des lésions de FT liés à l'hématopoïèse (*CEBPA*, *RUNX1*, *GATA2*) ou de *NPM1*, et le plus souvent des mutations associées à une dérégulation de la prolifération (*FLT3*, *RAS* ou autres). Dans 10/72 cas supplémentaires, le profil mutationnel est similaire, mais les mutations de régulateurs épigénétiques concernent d'autres gènes (*IDH1*, *IDH2* ou autres). La plupart de ces profils sont observés en cas de caryotype normal, de perte de chromosome sexuels, ou d'anomalies de pronostic intermédiaire. Dans 14/72 autres cas, on retrouve comme lésion initiatrice les translocations impliquant *MLL*, les anomalies du CBF ou la délétion 20q. Dans ce groupe, on observe peu de mutations associées, et notamment une quasi-absence de mutation des FT impliqués dans la régulation de l'hématopoïèse ou de *NPM1*. Les anomalies de prolifération sont par contre très fréquentes, et souvent associées de manière sous-clonale. Cela suggère que ces "anomalies chromosomiques pré-leucémiques" ont des conséquences fonctionnelles multiples sur les CSH mutées. En plus de leur donner un avantage sélectif par augmentation des capacités de mitoses, elles aboutissent sans doute à des anomalies de la différenciation dans les cellules dérivées de ces CSH. Le dernier groupe de 21/72 patients regroupe les patients ne rentrant pas dans les catégories précédentes. On y retrouve globalement les LAM à caryotype complexe ou défavorable (délétion du chromosome 7). De manière frappante ce groupe comprend tous les patients porteurs d'anomalies germinales favorisant les LAM, et tous les patients porteurs de mutations de *TP53*, ainsi que la majorité des patients porteurs d'anomalies de la machinerie d'épissage. Ces résultats suggèrent 3 hiérarchies génétiques bien différenciées dans les LAM. L'association fréquente dans notre série de mutations de *DNMT3A* et de *TP53* dans un même clone dans les LAM de novo suggère qu'en l'absence d'avantage sélectif donné par un agent extérieur favorisant sa sélection (chimiothérapie ou radiothérapie), une association à un autre évènement favorisant l'expansion du clone muté *TP53* peut être nécessaire avant qu'il devienne majoritaire et aboutisse à terme à l'état leucémique.

Au bilan, dans ce travail, nous avons identifié plusieurs lésions pré-leucémiques : les mutations de gènes fréquemment décrites dans l'HCLA (*TET2*, *DNMT3A* et *ASXL1*), mais

aussi des anomalies chromosomiques pré-leucémiques (translocations impliquant *MLL* et la délétion 20q). Selon l'événement initial, des hiérarchies mutationnelles récurrentes sont observées, avec des mécanismes différents de leucémogénèse. En cas d'anomalie première de gènes impliqués dans l'HCLA, le deuxième événement touche fréquemment un autre régulateur épigénétique, puis est suivi d'une mutation impliquant des troubles de différenciation cellulaire (*NPM1*, facteurs de transcription) et enfin des anomalies de la prolifération. En cas de translocation récurrente (*MLL* mais aussi anomalies du *CBF*) ou de délétion 20q, très peu d'événements de maturation sont observés, mais fréquemment de multiples événements de prolifération sous-clonaux sont retrouvés. Le nombre plus faible d'événements génétiques nécessaire à la transformation leucémique dans cette catégorie implique probablement un délai plus court entre le premier événement et la transformation leucémique que dans le modèle dérivé de l'HCLA. Enfin, les études en RC des patients ayant rechuté montre la persistance constante des anomalies précoces du clone après traitement. La persistance de la seule première anomalie peut conduire à la rechute 5 à 10 ans plus tard. Cela suggère que la guérison des LA ne peut être affirmée qu'en cas de disparition complète de toutes les anomalies du clone. La persistance d'anomalies fondatrices peut potentiellement aboutir à la rechute des années plus tard. Cependant, dans l'HCLA, de nombreux sujets n'évoluent jamais vers une maladie maligne agressive. Le retour après traitement à une HCLA est donc probablement associé à un risque de rechute, mais ce risque n'est pas clairement quantifié. L'attitude thérapeutique ou le suivi à entreprendre dans cette situation ne sont pour l'instant pas clairement définis.

Résultats : Article 2

Clonal history of a cord blood donor cell leukemia with pre-natal somatic JAK2 V617F mutation

P Hirsch ^{1,4}, AC Mamez ^{1,4}, R Belhocine ⁴, S Lapusan ^{3,4}, R Tang ⁴, L Suner ^{1-3,5}, D Bories ⁶, C Marzac ⁵, F Fava ⁴, O Legrand ^{1,4}, M Mohty ^{1,4}, L Douay ^{1-3,5}, F Delhommeau ^{1-3,5}

1-Sorbonne Universités, UPMC Univ Paris 06, UMR_S 938, CDR Saint-Antoine, F-75012, Paris, France

2-INSERM, UMR_S 938, CDR Saint-Antoine, F-75012, Paris, France ;

3-Sorbonne Universités, UPMC Univ Paris 06, GRC n°7, Groupe de Recherche Clinique sur les Myéloproliférations Aiguës et Chroniques MYPAC, Paris, France.

4- AP-HP, Hôpital St Antoine, Service d'Hématologie clinique et de thérapie cellulaire, F-75012, Paris, France

5- AP-HP, Hôpital Saint-Antoine & Hôpital Armand-Trousseau, Service d'hématologie biologique, F-75012, Paris, France

6- AP-HP, Hôpital Henri Mondor, Unité d'Hématologie Moléculaire, F-94010, Créteil, France

Donor cell derived leukemia is an uncommon complication after cord-blood transplantation, and has been suggested to result from the evolution of a pre-leukemic clone existing in the cord blood¹. Here, we report a case of a cord blood derived leukemia (CBDL) and describe the clonal history of the disease.

A 54 years old man was referred to our institution with deep asthenia and digestive bleeding. He had neither prior medical history nor familial antecedents. Cell blood counts and blood smear examination revealed pancytopenia with haemoglobin at 45 g/L, platelets at $12 \times 10^9/L$, and leukocytes at $1.0 \times 10^9/L$. Bone marrow was infiltrated by 50 % of small and medium sized myeloid blasts, with a $CD34^+ HLADR^+ CD33^+ CD13^+ CD117^+ MPO^+ CD36^- CD7^-$ immunophenotype (Figure 1a, b). Karyotype and standard molecular analyses (*FLT3-ITD*, *NPM1* and *CEBPA* mutation screening) were normal, and the disease was classified as acute myeloid leukemia (AML) with maturation. No sign of extra-medullar involvement was noticed. The initial treatment consisted in chemotherapy (cytarabine, idarubicine, etoposide), and complete remission (CR) was reached after 38 days. After two consolidation courses, a double cord-blood (CB) allogeneic transplantation was realized, with a conditioning regimen consisting of thiotepea, busulfan, fludarabine and anti-thymoglobulin. Chimerism analyses were full donor for one CB unit at day 15 post-transplantation. Despite cyclosporine and mycophenolate mofetil prophylaxis, acute graft versus host disease (GvHD) occurred at day 30. Supplemental immunosuppressive therapy with steroids and methotrexate was begun, and could be tapered after 8 months. The development of chronic GvHD led to cyclosporine and steroids re-introduction. Moreover, a disseminated *Mycobacterium haemophilum* infection was diagnosed at day 60 and was successfully treated.

Twenty months after transplantation, the patient presented with hepato-splenomegaly and nodular skin lesions. Cell blood counts and blood smear examination retrieved anemia, thrombocytopenia, and $24.6 \times 10^9/L$ leukocytes with 40% of large myeloid blasts. The medullar analysis showed 69% of $CD34^- HLADR^- CD33^+ CD13^+ CD117^+ MPO^- CD36^+ CD7^+$ blast infiltration (Figure 1a, b), with a $48,XY,+6,der(13)t(13;?)(p11;?),+21[12]/46,XY[3]$ karyotype. As clinical, morphologic, immunophenotypic, and karyotypic presentations at relapse were different from those at diagnosis, the hypothesis of donor cell leukemia was envisioned. Chimerism analysis confirmed this hypothesis by retrieving 99% of donor cells in blood samples, indicating that the second leukemia arose from donor cells.

To further genotype the two diseases, we analyzed diagnostic and relapse samples by targeted resequencing of a panel of 122 genes recurrently mutated in AML. In the first AML sample,

we found somatic *DDX41* p.R525H, *CUX1* p.V1017fs and *SRSF2* p.P95S mutations. In addition, a heterozygous *DDX41* p.S543* mutation was found in both diagnosis and complete remission (CR) samples, whereas the other mutations were not detected in the remission sample by using Sanger sequencing and high sensitivity targeted sequencing, suggesting that this *DDX41* stop variant was germ-line (Figure 1c, d). Targeted sequencing of the donor cell leukemia revealed that, while the previous variants were all absent, five new mutations (*JAK2* p.V617F, *EZH2* p.G538*, *CHEK2* p.G222*, *IKZF1* p.R143W, and *DSCAM* p.Q803K) had occurred (Figure 1d). Moreover, the analysis of single nucleotide polymorphisms (SNPs) located in the sequenced regions showed that the two leukemias had two distinct haplotypes. Altogether, these sequencing data demonstrated that the second leukemia had occurred in cells from one of the two transplanted CB units.

To understand the clonal composition of the donor cell leukemia, we first attempted to infer the fractions of cells carrying each chromosomal or genomic lesion from variant allele frequencies (VAFs). The percentages of cells with *JAK2*, *EZH2*, *IKZF1*, and *CHEK2* mutations were calculated at 78.5 (95% confidence interval: 70.2-85.6), 83.0 (77.1-88.8), 68.8 (60.2-77.4), and 65.7 (57.1-74.2), respectively. The analysis of several SNPs located on chromosome 6 indicated that around 69.9% (95% confidence interval: 68.6-71.2) of the cells carried the trisomy 6. As *DSCAM* is located on chromosome 21, and in the absence of available markers on chromosome 21, we could calculate neither the fraction of *DSCAM* mutant cells nor the percentage of cells with trisomy 21. In the same way, no marker of the chromosome 13 rearrangement was found.

To refine the clonal architecture, cells from the second leukemia were seeded in methylcellulose supplemented with 37% fetal calf serum, 12% bovine serum albumin, 1% L-glutamine, 10^{-6} M of B-mercaptoethanol, 1IU/mL of EPO, 50 ng/mL of SCF, 25 ng/mL of FLT3 ligand, 10 ng/mL of IL3, 10 ng/mL of IL6, 10 ng/mL of GCSF, 5 ng/mL of GMCSF, 10 ng/mL of TPO, 100 IU/mL penicillin and 100 μ g/mL streptomycin. Individual single cell-derived colonies were picked after 14 days to backtrack the *JAK2*, *EZH2*, *CHEK2*, and *DSCAM* mutations as well as a trisomy 6 marker. Results of 37 colonies were available to infer an order of events, where *JAK2* mutation was found in first position, followed by *EZH2*, *CHEK2*, and *DSCAM* mutations, and then trisomy 6 (Figure 2a). We next performed high sensitivity NGS analysis (Haloplex HS Target enrichment system, Agilent technologies) of the initial CB unit and a medullar sample collected 282 days post-transplantation. With this method, single DNA fragments are tagged with unique random 10-nucleotide indexes before the first PCR amplification step, allowing the confident detection of low frequency variants.

By using this technology, we could not detect chromosomal aberrations. However, based on our colony analysis, and as all abnormal metaphases had concomitant chromosome 6, chromosome 13, and chromosome 21 abnormalities on the karyotype analysis, we assumed that chromosomal events had occurred after the gene mutations. Strikingly, only one variant, *JAK2* p.V617F, was detected in both the initial CB unit and the day 282 post-transplant sample with variant alleles frequencies of 0.003 (59/19380 reads and 8/2397 amplicon families in the CB unit and 87/19758 reads and 12/4352 amplicon families at day 282, respectively). All other variants were found below the maximum background error rate defined as the ratio of the number of amplicon families with non-variant plus non-reference calls to total number of amplicon families (Figure 2b). Altogether, these results indicate that around 0.6% of cells with the *JAK2* p.V617F mutation were present in the transplanted CB unit, and that the donor cell leukemia occurred after the accumulation of the subsequent mutations and chromosomal aberrations in this *JAK2* mutant clone.

AML is the result of the accumulation of multiple genomic and chromosomal lesions in hematopoietic stem and progenitor cells ^{2,3}. In several works, the onset of an age-related clonal hematopoiesis (ARCH) has been shown to be the initial step of many myeloid malignancies, including AML. Altogether, mutations in only a handful of genes, including *JAK2*, were shown to account for 85% of cases with ARCH (*DNMT3A*, 54%, *TET2* 10%, *ASXL1* 9%, *JAK2* 5%, *SF3B1* 4%, *TP53* 4%), most individuals being older than 50 ^{4,5}. In a recent report, a donor cell leukemia with a pre-existing mutation in *DNMT3A* (the most frequently mutated gene in ARCH) in transplanted cells was described. In this case, the donor was the 52 years old recipient's brother, and did not develop any myeloid malignancy ⁶. Here, we demonstrate that donor cell leukemia can also emerge from a clone with pre-natal origins, as the *JAK2* mutation, albeit at very low frequency, was found in the CB unit whereas other mutations were not detected. Other lesions, such as *DNMT3A* mutations ², *ETV6-RUNX1* fusion, and *RUNX1-RUNX1T1* fusion have already been found in cord blood cells ^{7,8}, indicating that pre-leukemic lesions can be acquired in the pre-natal period.

Strikingly, both the clinical presentation, with hepato-splenomegaly, and the genotype of the CBDL, with *JAK2* mutation but also *EZH2*, *CHEK2* ⁹, and *IKZF1* ¹⁰ mutations, were reminiscent of the acute transformation of a myeloproliferative neoplasm. This suggests that the onset and evolution of the donor cell disease could have been favored and accelerated in the recipient context. Defects in the bone marrow micro-environment could lead to the preferential development of a pre-leukemic clone at the expense of normal hematopoiesis ¹¹. In our patient, a germ-line *DDX41* mutation, known to predispose to myeloid neoplasia ¹²,

was found. Thus, one could hypothesize that the *DDX41* mutant bone marrow micro-environment remained partly permissive to the development of AML. In addition, a rare *Mycobacterium* infection was observed, suggesting a severe immunity defect after transplantation in the patient. Defects of the immune system, due to both immunosuppressive therapy and prolonged T-cell immaturity could have impaired anti-tumoral activity.

In conclusion, this case recapitulates the different mechanisms that may contribute to the emergence of a CBDL: the pre-natal origin of the pre-leukemic clone, the subsequent accumulation of additional mutations and chromosomal aberrations, the permissive bone marrow micro-environment of the host, and the defects in anti-tumoral immune surveillance.

References

- 1 Shiozaki H, Yoshinaga K, Kondo T, Imai Y, Shiseki M, Mori N *et al.* Donor cell-derived leukemia after cord blood transplantation and a review of the literature: differences between cord blood and BM as the transplant source. *Bone Marrow Transplant* 2014; **49**: 102–109.
- 2 Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* 2012; **150**: 264–278.
- 3 Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* 2012; **4**: 149ra118.
- 4 Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014; **371**: 2488–2498.
- 5 Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014; **371**: 2477–2487.
- 6 Yasuda T, Ueno T, Fukumura K, Yamato A, Ando M, Yamaguchi H *et al.* Leukemic evolution of donor-derived cells harboring IDH2 and DNMT3A mutations after allogeneic stem cell transplantation. *Leukemia* 2014; **28**: 426–428.
- 7 Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A* 2002; **99**: 8242–8247.
- 8 Wiemels JL, Xiao Z, Buffler PA, Maia AT, Ma X, Dicks BM *et al.* In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* 2002; **99**: 3801–3805.
- 9 Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013; **369**: 2391–2405.
- 10 Jäger R, Gisslinger H, Passamonti F, Rumi E, Berg T, Gisslinger B *et al.* Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. *Leukemia* 2010; **24**: 1290–1298.
- 11 Flynn CM, Kaufman DS. Donor cell leukemia: insight into cancer stem cells and the stem cell niche. *Blood* 2007; **109**: 2688–2692.
- 12 Polprasert C, Schulze I, Sekeres MA, Makishima H, Przychodzen B, Hosono N *et al.* Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* 2015; **27**: 658–670.

Acknowledgments

This project was funded by the ARC foundation (N°EML20110602421), by the Région Ile-de-France (N°2012-2-eml-06-UPMC_12016710), by the Association Laurette Fugain (N°J15I409), and by the Institut National du Cancer (INCA) (PH).

Authors Contributions

PH performed cell culture, genotyping and NGS experiments, interpreted the results and wrote the manuscript

ACM and SL collected clinical data and contributed in follow-up of the patient

RB coordinated HSC transplantation and contributed in material collection

RT designed NGS assays

LS performed cytological and flow cytometry analysis of patient samples

DB and CM performed standard molecular and chimerism analysis

FF contributed in material collection and sample preparation

MM and OL contributed in follow-up of the patient

FD designed the research, interpreted the results and wrote the manuscript

All authors contributed in manuscript review

Authors information

Authors declare no competing financial interest.

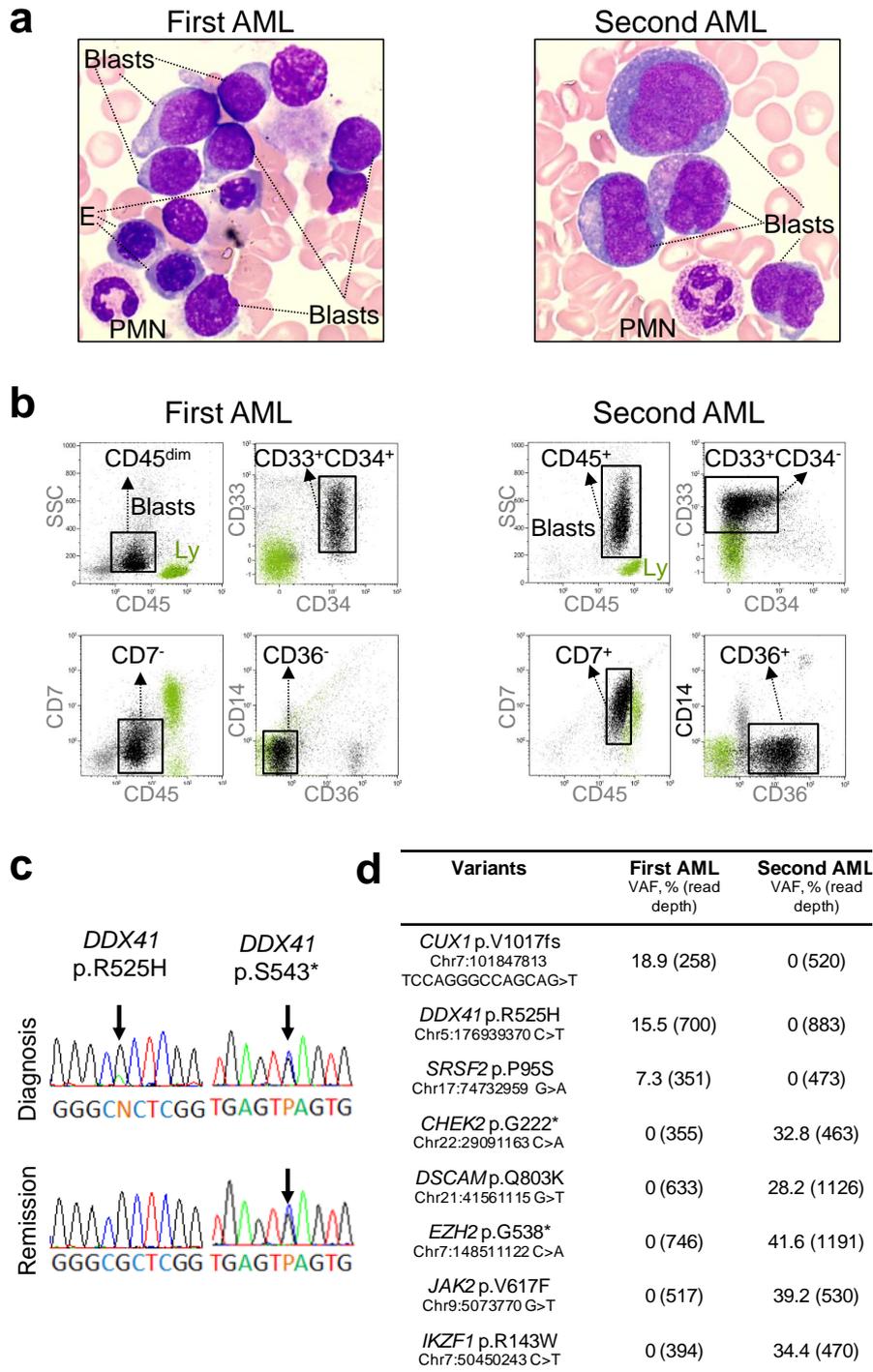


Figure 1: Distinct morphologic, immunophenotypic, and molecular features of initial disease and donor cell leukemia **a.** Bone marrow morphology of the initial leukemia (first AML, left) and the donor cell leukemia (second AML, right). E: erythroblasts; PMN : polymorphonuclear leukocytes **b.** Flow cytometric analysis of the two leukemias. Black and green dots denote blasts and lymphocytes (Ly), respectively. The immunophenotype of the blasts is indicated in each scattergram. **c.** Sanger sequence traces of diagnosis and remission samples from the first leukemia suggesting a somatic p.R525H mutation and a germline p.S543* mutation in *DDX41*. **d.** Variant allele frequencies of the mutations detected by targeted resequencing in the two leukemias.

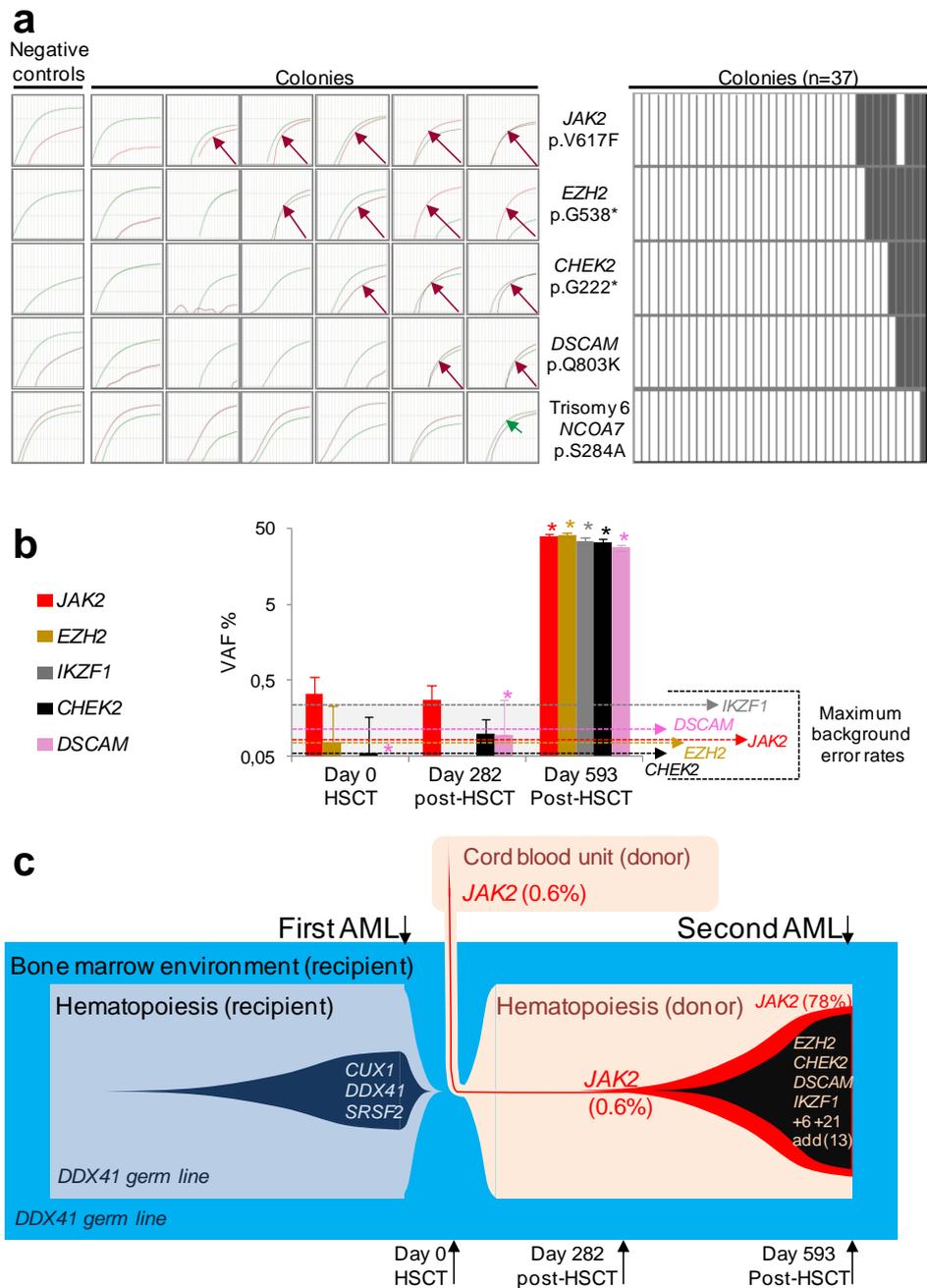


Figure 2 : Clonal history of the cord-blood derived leukemia. a. Analysis of single-cell derived colonies from the cord blood derived leukaemia. Left panel : allele specific PCR amplification plots of wild type and mutant alleles in six colonies and in negative controls. Red arrows indicate mutant allele detection. The green arrow denotes the duplicated allele of *NCOA7*, indicating trisomy 6. Right panel : summary of the single-cell derived colony assays. Columns : colonies; rows: genetic events. Coloured boxes represent colonies with mutations or trisomy 6. **b.** Results of targeted resequencing in the cord blood unit (day 0), at day 282, and in the cord blood derived leukaemia at day 593. Histograms represent VAFs and error bars denote 95% confidence intervals. (*) indicates analyses performed with the 122-gene panel NGS assay. Other analyses were performed with high sensitivity assays. Dotted arrows represent the maximum background error rates for the indicated variants. **c.** Summary of the clonal history of the two leukemias. HSCT : hematopoietic stem cell transplantation.

Article 2 : Commentaires

Ce cas clinique étudie l'architecture clonale d'une LAM développée après greffe de CSH issues de sang placentaire (USP). Cette deuxième leucémie comporte des anomalies cytogénétiques, phénotypiques et moléculaires complètement différentes de la LAM initiale. L'étude des polymorphismes (identiques entre les cellules de la LAM et celles du sang placentaire) confirme qu'elle dérive des CSH de l'USP. La clinique de cette LA évoque une transformation d'un néoplasme myéloprolifératif (splénomégalie majeure et atteinte extra-médullaire). Son génotype est également typique de cette situation, avec une association des mutations dans les gènes "classiques" *JAK2*, *EZH2*, *IKZF1* ou *CHEK2*^{204,205}. Les autres mutations et événements cytogénétiques retrouvés sont moins classiques.

L'étude séquentielle dans le temps par reséquençage ciblé et séquençage en profondeur révèle la chronologie d'apparition des événements, avec un événement précoce (*JAK2* p.V617F), déjà détectable dans les cellules de sang placentaire. L'expansion du clone muté pour *JAK2*, et l'acquisition secondaire en plusieurs mois des autres événements génétiques aboutit au développement de la LAM. L'étude des colonies n'a pas pu intégrer tous les événements par manque de matériel après culture, mais confirme que *JAK2* p.V617F est l'événement fondateur du clone, et que les autres événements sont acquis secondairement. Cette architecture clonale est tout à fait similaire à celle des néoplasmes myéloprolifératifs. Les mutations de *JAK2* sont retrouvées dans l'HCLA, même si leur fréquence est inférieure à celle des mutations épigénétiques. L'apparition d'événements fondateurs de LAM (translocations chromosomiques) in utero a déjà été décrite²⁰⁶, ainsi que la présence de mutations de *DNMT3A*²⁰⁷. L'acquisition secondaire des mutations chez notre patient semble mimer celle de l'HCLA et des néoplasmes myéloprolifératifs, avec un délai de développement beaucoup plus rapide.

Plusieurs facteurs ont pu favoriser cette émergence rapide de la maladie. L'immunodépression du patient a été sévère et prolongée, comme en témoigne l'infection post transplantation à mycobactérie atypique. Elle a été favorisée par la nécessité de traitement immuno-suppresseur lourd et prolongé en raison de la maladie du greffon contre l'hôte (GVH). Un défaut dans le micro-environnement est également très probable chez le patient. D'une part la GVH est connue pour entraîner de telles lésions, et d'autre part les mutations germinales de *DDX41* comme celle retrouvée chez le patient, pourraient potentiellement entraîner des anomalies du micro environnement qui persisteraient après l'allogreffe. L'étude

complémentaire des cellules souches mésenchymateuses (CSM) n'a pas pu être réalisée chez le patient, car les cultures de CSM entreprises lors du deuxième épisode leucémique ont été un échec.

Enfin, l'état clinique du donneur est inconnu (impossibilité de remonter à l'identité du donneur pour des raisons éthiques). Ce dernier est âgé d'une quinzaine d'années à ce jour. Il serait intéressant de savoir s'il a développé une pathologie hématologique comprenant la mutation de *JAK2*, et d'étudier son hématopoïèse à la recherche d'autres événements secondairement acquis. L'étude d'un cas comparable de LAM sur greffon à partir d'un donneur géno-identique a ainsi mené à découvrir, à partir des mêmes anomalies fondatrices présentes lors de l'allogreffe, des séquences différentes d'accumulation d'événement récurrents chez le donneur et chez le receveur ²⁰⁸, sans que le donneur ne développe de pathologie myéloïde.

Au bilan, ce cas clinique montre l'architecture et l'histoire clonale d'une LAM développée à partir de CSH d'USP, avec présence d'une mutation de *JAK2* acquise *in utero*. Le développement de la LAM secondaire selon une séquence classique dans les néoplasmes myéloprolifératifs, s'est fait à partir de cellules mutées présentes déjà dans l'USP. Les défauts de l'immunité et du micro-environnement ont probablement potentialisés l'émergence rapide de cette LAM.

Données complémentaires

A- Phylogénie clonale des LAM de l'enfant

L'étude de la phylogénie clonale a également été pratiquée dans 5 cas de LAM de l'enfant. Chacun des 5 patients sélectionnés a rechuté après le premier traitement. Dans les 5 cas, un séquençage de l'exome a été pratiqué sur des échantillons obtenus au diagnostic, à la RC et à la rechute. Les principales données cliniques et biologiques au diagnostic et en rechute des 5 patients sont résumées dans le tableau 4.

Tableau 4 : Résumé des données cliniques et biologiques de 5 LAM de l'enfant.

N° Patient		Sexe	Age	GB (x10 ⁹ /L)	FAB	Cytogénétique	Biologie moléculaire
2014-046	Diag	F	5,7	70	M4	45, XX, -7, der2	
	Rechute		6,2	41			
2014-044	Diag	M	8,8	11,2	M1	45, X, -Y, t(8;21)(q22,q22)	<i>AML1-ETO</i> <i>KIT p.D816V</i>
	Rechute		9,6	9,6			
2014-045	Diag	F	9,5	11,3	M5a	47, XX, +8, t(9;11)(p22;q23)	<i>MLL-AF9</i>
	Rechute		10,2	14,4		48, XX, +8,+8, t(9;11)(p22;q23)	
2015-009	Diag	F	11	3,9	M2	46,XX	
	Rechute		12,1	7,6			
2014-047	Diag	F	8,6	5,7	M5b	47 XX +8 (14); 46, XX (15)	
	Rechute		10,4	7,2		46, XX	

Les principaux résultats du séquençage des exomes au diagnostic et à la rechute sont résumés dans le tableau 5.

Tableau 5 : Résumé des données de séquençage d'exome pour les 5 patients pédiatriques. Les mutations d'intérêt ont été sélectionnées selon la fonction des gènes et la probabilité des lésions d'être délétères (littérature, conservation des acides aminés et domaine fonctionnel atteint).

Patient	gène	coordonnées	AA	VAF diag	VAF rechute	VAF RC
2014-044	<i>KIT</i>	Chr4:55599321	D816V	40	38	0
	<i>SMC3</i>	Chr10:112361545	L932P	38	34	0
	<i>FZD2</i>	Chr17:42636716	R554C	44	45	0
	<i>ARGHAP21</i>	Chr10:24873823	R1799W	48	40	0
2014-045	<i>EED</i>	Chr11:85988118	R355*	31	45	0
	<i>ELK1</i>	ChrX:47496309	S402T	43	44	0
	<i>GSDMC</i>	Chr8:130789716	R40*	31	54	0
2015-009	<i>DDX11</i>	Chr12:31237978	R186W	19	7	3
	<i>USP9X</i>	ChrX:41091748	V4546I	46	18	1
	<i>NT5E</i>	Chr6:86200349	S445F	43	23	0
	<i>TET2</i>	Chr4:106180775	Site d'épissage	47	0	< 1
	<i>POLR2A</i>	Chr17:7404653	Y699C	38	0	0
	<i>UBE2J1</i>	Chr6:90039499	G286R	42	0	0
2014-047	<i>IDH1</i>	Chr2:209113112	R132H	46	39	0
	<i>CEBPA</i>	Chr19:33792302	G340fs	0	24	0
	<i>KDM6A</i>	ChrX:44942752	R1111fs	0	22	0
2014-046	Pas d'anomalie identifiée					

Pour 3 des patients, des colonies ont été obtenues par culture des cellules du diagnostic après un tri magnétique CD34 et CD38. Chez 2 de ces 3 patients, des cultures complémentaires de LTC-IC ont été pratiquées. Pour les 2 autres patients, les cultures ont été un échec, et trop peu de colonies ont été obtenues pour permettre une analyse de la phylogénie clonale.

certain patients développant dans les suites une LAM ou une LAL portant ces anomalies. Il est donc probable que ces différentes anomalies soient apparues très tôt dans l'enfance, ou lors de la vie fœtale, conduisant à une hématopoïèse clonale, avec une accumulation secondaire de lésions génétiques, selon un modèle identique aux LAM de l'adulte. Ce mécanisme représente probablement une minorité des LAM de l'enfant. En effet, on constate en pédiatrie une quasi-absence des lésions épigénétiques « classiques de l'HCLA » retrouvées dans presque 50% des LAM de l'adulte (*DNMT3A*, *TET2*, *ASXL1* ou autres). En pédiatrie, l'hématopoïèse pré-leucémique favorisant l'expansion clonale des CSH est probablement rare, et seules de rares cas en plus des LAM porteuses de translocations récurrentes notamment celles impliquant le CBF et *MLL*, pourraient répondre à ce modèle (cf patient 2014-047 porteur de mutation d'*IDH1* au diagnostic et en rechute). L'analyse complémentaire de la phylogénie clonale des patients pédiatriques porteurs de translocations impliquant *MLL* est en cours (exome et colonies chez 6 patients complémentaires).

Chez le patient 2015-009, aucune mutation somatique classique n'a été retrouvée. La mutation de *TET2* implique un site d'épissage, et n'est pas un événement précoce dans la phylogénie. Même si une accumulation progressive d'événements génétiques somatiques a été observée chez ce patient, les mécanismes fonctionnels de cette expansion clonale restent à explorer. Chez ce patient, l'événement initiateur implique le gène *DDX11*, une hélicase impliquée dans la cohésion des chromatides et indispensable à la ségrégation chromosomique.

Enfin chez le patient 2014-046, l'absence de lésion somatique identifiable par l'analyse de l'exome suggère des mécanismes différents de leucémogénèse. Des lésions germinales pré-disposantes, non identifiées chez nos patients (absence des mutations classiques décrites dans la littérature) pourraient, par exemple, jouer un rôle dans la régulation du microenvironnement tumoral, dans les mécanismes de réparation de l'ADN (voie p53 ou autre), ou dans d'autres mécanismes de régulation des CSH. Là encore, des analyses à grande échelle sont à prévoir.

B- Mesure de la maladie résiduelle en NGS

Les études par séquençage en profondeur en rémission complète sont directement applicables pour la mesure de la maladie résiduelle chez les patients. Au total, un séquençage en profondeur en RC après une cure d'induction (ou après ré-induction pour 2 patients en échec du premier traitement) a été pratiqué chez un total de 20 patients. Chez plusieurs d'entre eux, des points de suivi plus tardifs ont été également étudiés. Parmi ces 20 patients, 2 n'ont pas rechuté à ce jour (environ 3 ans après l'obtention de la RC), 1 patient a développé une LAM après greffe d'USP (voir article 2) et 1 patient a développé une LMMC secondaire 11 ans après le premier épisode leucémique. Les 16 autres patients ont rechuté dans un délai de 2 ans après l'obtention de la RC. Pour chaque patient, un panel personnalisé de séquençage ciblé comprenant l'intégralité des mutations identifiées au diagnostic et en rechute par le panel de 122 gènes a été utilisé. Pour l'évaluation en RC, les bibliothèques d'amplicons ont été obtenues avec les kits Haloplex HS (voir méthode article 1). Pour chaque expérience et chaque patient, un seuil d'erreur maximal a été défini (voir article 1). Une partie des données ont été présentées dans l'article 1 et 2. Les données complémentaires sont présentées sur les figures 9 et 10.

Cette étude est pour l'instant incomplète car elle intègre très peu de patients en RC prolongée et n'ayant pas rechuté. De premières conclusions sont néanmoins possibles. Tout d'abord, la quasi-totalité des patients peuvent être suivis en MRD avec cette technique. En effet, dans notre étude, sur 84 patients étudiés avec le kit 122 gènes (10 patients adultes supplémentaires non inclus dans l'article 1), aucune cible moléculaire n'a été retrouvée seulement chez 3 d'entre eux (dont 1 patient avec une inversion du chromosome 16, et un patient porteur de translocation impliquant MLL, tous deux évaluables par les techniques classiques de RT-PCR). Le suivi sera probablement particulièrement intéressant chez les patients de pronostic intermédiaire, et devrait pouvoir contribuer à guider les décisions d'intensification thérapeutique dans cette population, quand des études de large ampleur auront été menées.

L'analyse des premiers résultats permet d'ébaucher des stratégies de suivi. L'étude des patients ayant rechuté dans les 2 ans montre, dans les échantillons de RC post induction, une fréquente persistance à un haut niveau de VAF de plusieurs mutations, correspondant à des événements précoces de la phylogénie. Cependant, la persistance isolée du seul événement fondateur (*DNMT3A* chez le patient 2014-041, voir article 1 ; *U2AF1* chez le patient 2014-008, voir figure 10) ne semble pas nécessairement associée à une rechute précoce, mais plutôt

à un risque théorique de rechute tardive après ré-accumulation d'événements dans le clone pré-leucémique (comme chez le patient 20015-014, voir article 1). A l'inverse la persistance d'une majorité d'événements à un haut niveau de VAF signe la résistance au traitement du clone leucémique, ou d'un clone pré-leucémique à haut potentiel de transformation. Cette persistance semble associée à une rechute précoce (patients 2014-033, 2015-035, 2014-017, 2014-029, ou autres, voir figure 9). Par exemple, chez le patient 2014-033, malgré l'obtention de la rémission cytologique, toutes les mutations restent détectables avec une VAF d'environ 10 %, ce qui représente donc plus de 20 % des cellules (mutations hétérozygotes), signant une très mauvaise réponse au traitement. L'étude complémentaire en RC de 45 patients dont 25 n'ont pas rechuté précocement est prévue. Elle devrait pouvoir établir les meilleures stratégies de suivi de MRD, et définir les meilleures cibles pour le suivi précoce (probablement les cibles en position intermédiaire dans la phylogénie), et pour le suivi tardif (probablement les événements fondateurs du clone). L'évaluation du pronostic en fonction du niveau de MRD sera également pratiquée sur les patients, notamment en cas de pronostic intermédiaire (45 patients de cytogénétique intermédiaire, dont 31 sans FLT3-ITD, et 18 sans mutation de *NPM1* ni de FLT3-ITD).

L'étude des RC met par ailleurs en évidence des réponses différentes au traitement selon les différents sous-clones présents au diagnostic. En effet, le clone fondateur résiste le plus souvent complètement au traitement. Des sous-clones minoritaires présents au diagnostic peuvent par ailleurs résister au traitement et même se développer pour devenir majoritaires en rechute (par exemple, clone muté pour *RUNX1* chez le patient 2014-029 ou clone muté pour *JAK2* et *SETBP1* chez le patient 2013-002). L'identification précoce au diagnostic de ces sous-clones et leur suivi pourrait être un élément important pour prévoir le pronostic des malades à court terme (développement des sous-clones dès la première cure de chimiothérapie). Ces résultats ont également été observés très récemment par une autre équipe sur 5 malades, sans étude systématique des rechutes²⁰⁹.

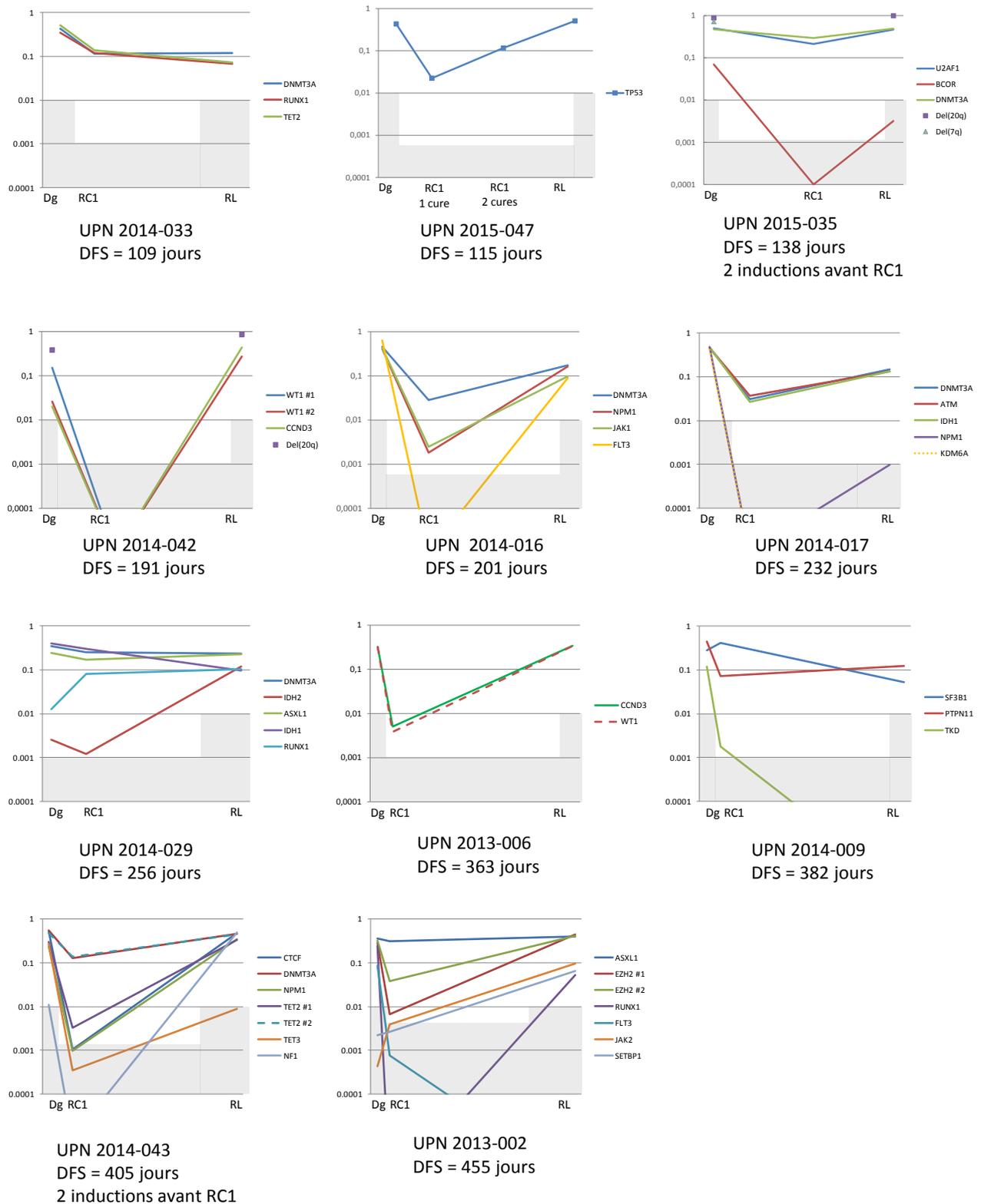


Figure 9 : Cinétique d'évolution des VAF de tous les variants détectés chez 11 patients ayant rechuté. Pour chaque patient et chaque run, un seuil maximal d'erreur a été calculé et représenté en gris. L'évaluation de la RC a été faite par réséquençage en profondeur. La fréquence des anomalies chromosomiques calculée à partir de SNP a été figurée quand elle était évaluable. DFS : survie sans maladie. DG = diagnostic. RC = rémission complète. RL = rechute.

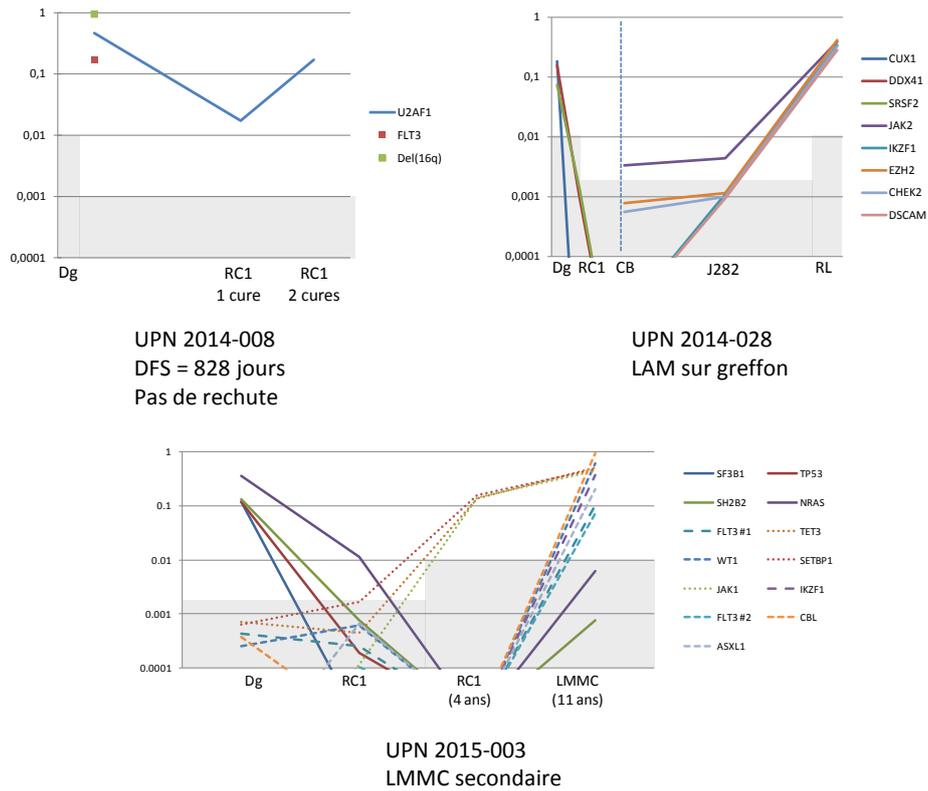


Figure 10 : Cinétique d'évolution des VAF de tous les variants chez 2 patients ayant développé des LAM secondaires, et chez un patient n'ayant pas rechuté. Chez le patient 2014-028 la ligne verticale en pointillés représente l'allogreffe de sang placentaire. Chez le patient 2014-008, une allogreffe a été réalisée un mois après la dernière évaluation, et l'étude de points du suivi post greffe a été programmée. La fréquence allélique de FLT3-ITD chez le patient 2014-008 est mesurée en PCR et la fréquence de del(16q) en technique cytogénétique classique. Ces anomalies ont été étudiées et sont indétectables au 2^{ème} point de suivi.

Discussion et Conclusion

Au bilan, dans ce travail, nous avons pratiqué l'étude de la phylogénie clonale dans 74 cas de LAM de l'adulte et 5 cas de LAM de l'enfant. Chez l'adulte, nous avons mis en évidence plusieurs hiérarchies mutationnelles récurrentes.

Un premier modèle hiérarchique dérive de l'HCLA. A partir d'un clone pré-leucémique comprenant des mutations de régulateurs de l'épigénétique (et notamment *TET2*, *DNMT3A* ou *ASXL1*), des mutations secondaires s'accumulent jusqu'à la transformation leucémique franche. Ces événements secondaires touchent d'abord fréquemment d'autres régulateurs de l'épigénétique, puis des gènes impliqués dans la différenciation cellulaire (*NPM1* et FT hématopoïétiques), puis des gènes impliqués dans la prolifération (*FLT3*, *RAS*, *KIT* ou autres). Ce mécanisme paraît être le plus fréquent chez l'adulte, mais doit être beaucoup plus rare en pédiatrie, car l'hématopoïèse clonale se développe principalement avec le vieillissement. Notons cependant qu'un des enfants étudié présentait des mutations classiques des LAM l'adulte, avec comme événement fondateur une mutation d'*IDH1*, suggérant qu'une partie des LAM de l'enfant peut tout de même dériver d'HCLA. La basse fréquence des mutations épigénétiques chez l'enfant suggère cependant que ce mécanisme de leucémogénèse reste très à la marge. La présence de mutation de *DNMT3A* dans du sang placentaire a déjà été décrite²⁰⁷ et l'existence de mutation de *JAK2* p.V617F dans des CSH issues d'USP (voir article 2) suggère que d'autres mutations pré-leucémiques de l'HCLA peuvent aussi se produire *in utero*. Des mutations d'autres régulateurs épigénétiques pourraient également se produire de la même manière, pour aboutir à une HCLA, et à des LAM après plusieurs années, notamment si des facteurs favorisant la sélection du clone pré-leucémique sont associés (immunodépression, troubles du micro-environnement...).

Une deuxième hiérarchie d'événements génétiques implique des anomalies chromosomiques récurrentes comme événement fondateur, et notamment les translocations impliquant *MLL* ou le CBF. Ces événements sont fréquents chez l'adulte et chez l'enfant. L'accumulation d'un nombre moins élevé d'événements secondaires semble suffisante dans cette situation pour aboutir à une LAM. En effet, ces translocations impliquent à la fois une dérégulation épigénétique et des anomalies de la transcription, car elles impliquent des FT majeurs de l'hématopoïèse. Un délai plus court de l'événement fondateur à la leucémie

pourrait donc être possible, avec la nécessité d'occurrence d'un événement touchant la prolifération pour obtenir une transformation franche (*FLT3*, *KIT*, ou voie RAS principalement). Notons cependant que des récentes études ont retrouvé de fréquentes mutations d'*ASXL2* ou des gènes de la cohésine dans les LAM avec remaniement du CBF²¹⁰, ce qui suggère que ces événements doivent contribuer à conférer un avantage sélectif au clone pré-leucémique. Les anomalies de *MLL* ou du CBF peuvent se développer *in utero*, et ont déjà été observées sur des tests de Guthrie chez des enfants ayant déclaré des LAM des années plus tard²⁰⁶. Ce mécanisme de leucémogénèse dérivant "d'anomalies chromosomiques pré-leucémiques" est observé assez fréquemment. D'autres lésions chromosomiques récurrentes pourraient également s'y rattacher, et notamment les délétions du chromosome 20q qui sont les événements fondateurs du clone chez plusieurs des patients de notre étude.

Enfin, un troisième groupe est constitué par les patients sans anomalie épigénétique ou chromosomique fondatrice, et sans mutation des FT majeurs de l'hématopoïèse ou de *NPM1*. Chez nos patients adultes, nous observons dans ce groupe tous les patients porteurs de mutations de *TP53* (n=7) et d'anomalies germinales favorisant les LAM (n=5). Il est probable que dans les LAM de l'enfant, ce troisième groupe soit le plus important en nombre, avec des anomalies germinales conduisant à un défaut de réparation, ou favorisant par d'autres moyens la leucémogénèse. Ainsi, chez un des enfants étudié aucune mutation somatique n'a été identifiée, et chez un autre, un profil mutationnel très atypique a été retrouvé. Aucune mutation germinale prédisposante connue n'a été retrouvée chez les 5 patients pédiatriques de notre étude. Des analyses complémentaires à large échelle sont nécessaires pour identifier de potentiels événements favorisant. Notons que des anomalies germinales sont assez fréquemment retrouvées chez l'adulte dans notre série (entre 5 et 10 %), et devraient par conséquent être plus systématiquement recherchées, notamment pour assurer une sélection plus sûre des donneurs de CSH pour une éventuelle allogreffe de clôture de la séquence thérapeutique, voir pour assurer un conseil génétique dans l'avenir, quand les conséquences de ces événements seront mieux connues (risque de développer une LAM, pénétrance, *etc...*)

Par ailleurs, de nouvelles stratégies de suivi de MRD pourront être développées à partir de ce travail. La quasi intégralité des patients semble avoir une cible évaluable en MRD. L'étude de 16 patients ayant rechuté précocement montre une persistance quasi-constante à un haut niveau (>1 %) des mutations les plus précoces de la phylogénie, même si les mutations plus tardives ont disparu. Chez un patient ayant rechuté très tardivement, la mutation

fondatrice de *DNMT3A* était détectable au moins à faible niveau après les traitements, et a ré-entraîné l'accumulation de nouvelles lésions jusqu'à la rechute. Chez 1 patient n'ayant pas rechuté après plus de 3 ans de suivi, la mutation fondatrice de *DNMT3A* est restée à très haut niveau, et les autres mutations sont indétectables après de nombreuses séquences de traitement. Un suivi séquentiel en NGS de ce dernier patient serait particulièrement intéressant. Le suivi du seul événement fondateur paraît peu adapté au suivi de l'épisode leucémique. A l'inverse, seule la disparition complète de l'HCLA après les traitements paraît pouvoir affirmer la guérison complète. Cette conclusion reste à vérifier, mais les résultats d'une étude récente de l'évaluation de la MRD multi-cible semble suggérer qu'une diminution importante de toutes les mutations, comprenant les mutations initiatrices est associée à un meilleur pronostic ²⁰². Le risque de redévelopper une LAM en cas de retour à l'HCLA comportant une seule anomalie reste à quantifier, et doit probablement s'approcher du risque de développer une première LAM après HCLA, en l'absence d'émergence précoce d'un sous clone minoritaire sélectionné par les traitements ²⁰⁹. Parmi nos patients, deux ont atteint la rémission moléculaire sur tous les marqueurs au diagnostic, mais ont développé une leucémie secondaire, l'un une LAM après allogreffe et l'autre une LMMC après 11 ans d'évolution. L'étude complémentaire de 45 patients dont plus de la moitié n'a pas rechuté à deux ans de suivi est prévue, afin d'étudier l'évolution des profils moléculaires en RC prolongée et d'identifier si la persistance de certains marqueurs après traitement, et si le niveau de cette persistance, sont associés au risque de rechute et au pronostic global.

Deux obstacles sont à noter pour développer cette stratégie de suivi de MRD en pratique courante. En premier lieu, ces stratégies sont très coûteuses. En effet, un kit pour 48 réactions de séquençage haute sensibilité coûte environ 10000 euros. Pour atteindre des profondeurs de séquençage correctes, il est nécessaire de diminuer le plus possible le nombre de cibles, et on ne peut en pratique inclure qu'une quinzaine ou une vingtaine de patients par kit. Par ailleurs, les délais de fabrication des kits spécifiques sont d'environ 2 mois, ce qui pose un réel problème de délai technique pour l'évaluation des patients. En effet, le délai habituel d'organisation d'une allogreffe étant de 2 à 3 mois après le diagnostic de LAM, l'indication doit être posée le plus rapidement possible, et ne peut pas être retardé, car le patient risque de rechuter en cas de temps d'attente trop prolongé. Une solution serait de développer des kits standardisés incluant seulement les mutations les plus fréquemment retrouvées chez les patients. Cette option permettrait de suivre une proportion moins importante de patient. Une alternative serait de développer des stratégies de suivi multi-cibles

adaptées au patient, en utilisant une approche spécifique des patients en PCR digitale, à développer dès que les résultats du séquençage de la LAM sont connus. Le développement de ces stratégies est prochainement programmé dans l'équipe, après que les meilleurs cibles auront été déterminées par les compléments de l'étude en NGS.

En conclusion, notre étude a mis en évidence plusieurs hiérarchies génétiques dans les LAM, qui dépendent de la nature de l'événement fondateur du clone. Chez les patients ayant rechuté, la persistance du clone pré-leucémique est quasi systématique en rémission. La variation des événements entre diagnostic et rechute dépendant du temps jusqu'à la rechute. Les techniques de NGS permettent l'évaluation de la MRD chez la quasi-totalité des patients, mais les possibilités d'application au quotidien sont limitées pour des raisons pratiques. Le niveau de persistance des différentes anomalies en RC pourrait être associé au pronostic. Des études complémentaires sur des patients n'ayant pas rechuté sont nécessaires pour conclure sur l'impact de la mesure de la maladie résiduelle multi-cibles en NGS, et sur la meilleure stratégie de suivi à développer pour l'ensemble des malades, en évaluant également d'autres techniques comme la PCR digitale sur des cibles personnalisées selon les patients.

Bibliographie

1. Lorenz, E., Congdon, C. & Uphoff, D. Modification of acute irradiation injury in mice and guinea-pigs by bone marrow injections. *Radiology* **58**, 863–877 (1952).
2. Abkowitz, J. L., Catlin, S. N., McCallie, M. T. & Gutterop, P. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood* **100**, 2665–2667 (2002).
3. Shepherd, B. E., Gutterop, P., Lansdorp, P. M. & Abkowitz, J. L. Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. *Exp. Hematol.* **32**, 1040–1050 (2004).
4. Bhatia, M., Wang, J. C., Kapp, U., Bonnet, D. & Dick, J. E. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5320–5325 (1997).
5. Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J. Exp. Med.* **183**, 1797–1806 (1996).
6. Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218–221 (2011).
7. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* (2015). doi:10.1126/science.aab2116
8. Galy, A., Travis, M., Cen, D. & Chen, B. Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity* **3**, 459–473 (1995).
9. Manz, M. G., Miyamoto, T., Akashi, K. & Weissman, I. L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11872–11877 (2002).
10. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120–136 (2012).
11. Metcalf, D. Hematopoietic cytokines. *Blood* **111**, 485–491 (2008).
12. Kondo, M. *et al.* Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines. *Nature* **407**, 383–386 (2000).
13. Winter, S. S., Howard, T. & Ware, R. E. Regulation of expression of the human erythropoietin receptor gene. *Blood Cells. Mol. Dis.* **22**, 214–224; discussion 224a (1996).
14. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
15. Dahl, R., Iyer, S. R., Owens, K. S., Cuylear, D. D. & Simon, M. C. The transcriptional repressor GFI-1 antagonizes PU.1 activity through protein-protein interaction. *J. Biol. Chem.* **282**, 6473–6483 (2007).
16. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J.* **21**, 3770–3781 (2002).
17. Rhodes, J. *et al.* Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev. Cell* **8**, 97–108 (2005).
18. Starck, J. *et al.* Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Mol. Cell. Biol.* **23**, 1390–1402 (2003).
19. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
20. Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014).

21. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).
22. Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell* **5**, 442–449 (2009).
23. Wilting, R. H. *et al.* Overlapping functions of Hdac1 and Hdac2 in cell cycle regulation and haematopoiesis. *EMBO J.* **29**, 2586–2597 (2010).
24. Kluiver, J., Kroesen, B.-J., Poppema, S. & van den Berg, A. The role of microRNAs in normal hematopoiesis and hematopoietic malignancies. *Leukemia* **20**, 1931–1936 (2006).
25. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
26. Kiel, M. J., Radice, G. L. & Morrison, S. J. Lack of evidence that hematopoietic stem cells depend on N-cadherin-mediated adhesion to osteoblasts for their maintenance. *Cell Stem Cell* **1**, 204–217 (2007).
27. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
28. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
29. Zhou, B. O., Ding, L. & Morrison, S. J. Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *eLife* **4**, e05521 (2015).
30. Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. Haematopoietic stem cell release is regulated by circadian oscillations. *Nature* **452**, 442–447 (2008).
31. Zhu, J. *et al.* Osteoblasts support B-lymphocyte commitment and differentiation from hematopoietic stem cells. *Blood* **109**, 3706–3712 (2007).
32. Swerdlow, S.H *et al.* *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition.* (WHO, 2008).
33. Bene, M. C. *et al.* Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* **9**, 1783–1786 (1995).
34. Acute Myeloid Leukemia - SEER Stat Fact Sheets. at <http://seer.cancer.gov/statfacts/html/amyl.html>
35. Morton, L. M. *et al.* Evolving risk of therapy-related acute myeloid leukemia following cancer chemotherapy among adults in the United States, 1975-2008. *Blood* **121**, 2996–3004 (2013).
36. Nakanishi, M., Tanaka, K., Shintani, T., Takahashi, T. & Kamada, N. Chromosomal instability in acute myelocytic leukemia and myelodysplastic syndrome patients among atomic bomb survivors. *J. Radiat. Res. (Tokyo)* **40**, 159–167 (1999).
37. Travis, L. B. *et al.* Hematopoietic malignancies and related disorders among benzene-exposed workers in China. *Leuk. Lymphoma* **14**, 91–102 (1994).
38. Deschler, B. & Lübbert, M. Acute myeloid leukemia: epidemiology and etiology. *Cancer* **107**, 2099–2107 (2006).
39. Döhner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute Myeloid Leukemia. *N. Engl. J. Med.* **373**, 1136–1152 (2015).
40. Pulte, D., Gondos, A. & Brenner, H. Improvements in survival of adults diagnosed with acute myeloblastic leukemia in the early 21st century. *Haematologica* **93**, 594–600 (2008).
41. Bennett, J. M. *et al.* Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br. J. Haematol.* **33**, 451–458 (1976).

42. Bloomfield, C. D., Goldman, A., Hassfeld, D. & de la Chapelle, A. Fourth International Workshop on Chromosomes in Leukemia 1982: Clinical significance of chromosomal abnormalities in acute nonlymphoblastic leukemia. *Cancer Genet. Cytogenet.* **11**, 332–350 (1984).
43. Byrd, J. C. *et al.* Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* **100**, 4325–4336 (2002).
44. Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322–2333 (1998).
45. Slovak, M. L. *et al.* Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood* **96**, 4075–4083 (2000).
46. Rosnet, O., Mattei, M. G., Marchetto, S. & Birnbaum, D. Isolation and chromosomal localization of a novel FMS-like tyrosine kinase gene. *Genomics* **9**, 380–385 (1991).
47. Lyman, S. D. *et al.* Characterization of the protein encoded by the flt3 (flk2) receptor-like tyrosine kinase gene. *Oncogene* **8**, 815–822 (1993).
48. Lyman, S. D. *et al.* Cloning of the human homologue of the murine flt3 ligand: a growth factor for early hematopoietic progenitor cells. *Blood* **83**, 2795–2801 (1994).
49. Rosnet, O. *et al.* Human FLT3/FLK2 receptor tyrosine kinase is expressed at the surface of normal and malignant hematopoietic cells. *Leukemia* **10**, 238–248 (1996).
50. Lisovsky, M. *et al.* Flt3 ligand stimulates proliferation and inhibits apoptosis of acute myeloid leukemia cells: regulation of Bcl-2 and Bax. *Blood* **88**, 3987–3997 (1996).
51. Hayakawa, F. *et al.* Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines. *Oncogene* **19**, 624–631 (2000).
52. Schlenk, R. F. *et al.* Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N. Engl. J. Med.* **358**, 1909–1918 (2008).
53. Nakao, M. *et al.* Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911–1918 (1996).
54. Kottaridis, P. D. *et al.* The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood* **98**, 1752–1759 (2001).
55. Whitman, S. P. *et al.* FLT3 D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking FLT3 internal tandem duplications. *Blood* **111**, 1552–1559 (2008).
56. Fröhling, S. *et al.* Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* **12**, 501–513 (2007).
57. Borer, R. A., Lehner, C. F., Eppenberger, H. M. & Nigg, E. A. Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell* **56**, 379–390 (1989).
58. Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* **352**, 254–266 (2005).
59. Suzuki, T. *et al.* Clinical characteristics and prognostic implications of NPM1 mutations in acute myeloid leukemia. *Blood* **106**, 2854–2861 (2005).

60. Bolli, N. *et al.* Expression of the cytoplasmic NPM1 mutant (NPMc+) causes the expansion of hematopoietic cells in zebrafish. *Blood* **115**, 3329–3340 (2010).
61. Sportoletti, P. *et al.* The human NPM1 mutation A perturbs megakaryopoiesis in a conditional mouse model. *Blood* **121**, 3447–3458 (2013).
62. El Hajj, H. *et al.* Retinoic acid and arsenic trioxide trigger degradation of mutated NPM1, resulting in apoptosis of AML cells. *Blood* **125**, 3447–3454 (2015).
63. Lekstrom-Himes, J. & Xanthopoulos, K. G. Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J. Biol. Chem.* **273**, 28545–28548 (1998).
64. Tenen, D. G., Hromas, R., Licht, J. D. & Zhang, D. E. Transcription factors, normal myeloid development, and leukemia. *Blood* **90**, 489–519 (1997).
65. Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 569–574 (1997).
66. Pabst, T. *et al.* Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat. Genet.* **27**, 263–270 (2001).
67. Preudhomme, C. *et al.* Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). *Blood* **100**, 2717–2723 (2002).
68. Fröhling, S. *et al.* CEBPA mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **22**, 624–633 (2004).
69. Green, C. L. *et al.* Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 2739–2747 (2010).
70. Wouters, B. J. *et al.* Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* **113**, 3088–3091 (2009).
71. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
72. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12**, 599–612 (2012).
73. Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
74. Gowher, H. *et al.* Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *J. Mol. Biol.* **357**, 928–941 (2006).
75. Yamashita, Y. *et al.* Array-based genomic resequencing of human leukemia. *Oncogene* **29**, 3723–3731 (2010).
76. Thol, F. *et al.* Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 2889–2896 (2011).
77. Yan, X.-J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* **43**, 309–315 (2011).
78. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2012).
79. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).

80. Pronier, E. *et al.* Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood* **118**, 2551–2555 (2011).
81. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
82. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100–4108 (2004).
83. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
84. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
85. Hon, G. C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* **56**, 286–297 (2014).
86. Valinluck, V. & Sowers, L. C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).
87. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
88. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
89. Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
90. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
91. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
92. Li, Z. *et al.* Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509–4518 (2011).
93. Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14566–14571 (2011).
94. Shide, K. *et al.* TET2 is essential for survival and hematopoietic stem cell homeostasis. *Leukemia* **26**, 2216–2223 (2012).
95. Quivoron, C. *et al.* TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25–38 (2011).
96. Marcucci, G. *et al.* IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 2348–2355 (2010).
97. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
98. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
99. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).

100. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
101. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
102. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).
103. Yu, J. *et al.* A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res.* **67**, 10657–10663 (2007).
104. Tanaka, S. *et al.* Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. *Blood* **120**, 1107–1117 (2012).
105. Su, I.-H. *et al.* Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* **4**, 124–131 (2003).
106. Lessard, J. *et al.* Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev.* **13**, 2691–2703 (1999).
107. Herrera-Merchan, A. *et al.* Ectopic expression of the histone methyltransferase Ezh2 in haematopoietic stem cells causes myeloproliferative disease. *Nat. Commun.* **3**, 623 (2012).
108. Abdel-Wahab, O. *et al.* ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180–193 (2012).
109. Balasubramani, A. *et al.* Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat. Commun.* **6**, 7307 (2015).
110. Abdel-Wahab, O. *et al.* Deletion of *Asx11* results in myelodysplasia and severe developmental defects in vivo. *J. Exp. Med.* **210**, 2641–2659 (2013).
111. Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer* **7**, 823–833 (2007).
112. Bernt, K. M. *et al.* MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).
113. Deshpande, A. J. *et al.* Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood* **121**, 2533–2541 (2013).
114. Moriya, K. *et al.* Development of a multi-step leukemogenesis model of MLL-rearranged leukemia using humanized mice. *PloS One* **7**, e37892 (2012).
115. Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321–330 (1996).
116. Wang, Q. *et al.* Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 3444–3449 (1996).
117. Wang, Q. *et al.* The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell* **87**, 697–708 (1996).
118. Sasaki, K. *et al.* Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12359–12363 (1996).
119. Goyama, S. & Mulloy, J. C. Molecular pathogenesis of core binding factor leukemia: current knowledge and future prospects. *Int. J. Hematol.* **94**, 126–133 (2011).
120. Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
121. Alvarez, S. *et al.* DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PloS One* **5**, e12197 (2010).

122. Li, Y. *et al.* Epigenetic silencing of microRNA-193a contributes to leukemogenesis in t(8;21) acute myeloid leukemia by activating the PTEN/PI3K signal pathway. *Blood* **121**, 499–509 (2013).
123. Amann, J. M. *et al.* ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol. Cell. Biol.* **21**, 6470–6483 (2001).
124. Wang, J., Hoshino, T., Redner, R. L., Kajigaya, S. & Liu, J. M. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10860–10865 (1998).
125. Durst, K. L., Lutterbach, B., Kummalue, T., Friedman, A. D. & Hiebert, S. W. The inv(16) fusion protein associates with corepressors via a smooth muscle myosin heavy-chain domain. *Mol. Cell. Biol.* **23**, 607–619 (2003).
126. Lutterbach, B., Hou, Y., Durst, K. L. & Hiebert, S. W. The inv(16) encodes an acute myeloid leukemia 1 transcriptional corepressor. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12822–12827 (1999).
127. Vangala, R. K. *et al.* The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood* **101**, 270–277 (2003).
128. Pabst, T. *et al.* AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat. Med.* **7**, 444–451 (2001).
129. Choi, Y., Elagib, K. E., Delehanty, L. L. & Goldfarb, A. N. Erythroid inhibition by the leukemic fusion AML1-ETO is associated with impaired acetylation of the major erythroid transcription factor GATA-1. *Cancer Res.* **66**, 2990–2996 (2006).
130. Linggi, B. *et al.* The t(8;21) fusion protein, AML1 ETO, specifically represses the transcription of the p14(ARF) tumor suppressor in acute myeloid leukemia. *Nat. Med.* **8**, 743–750 (2002).
131. Yang, G. *et al.* Transcriptional repression of the Neurofibromatosis-1 tumor suppressor by the t(8;21) fusion protein. *Mol. Cell. Biol.* **25**, 5869–5879 (2005).
132. Markus, J. *et al.* Methylation-independent silencing of the tumor suppressor INK4b (p15) by CBFbeta-SMMHC in acute myelogenous leukemia with inv(16). *Cancer Res.* **67**, 992–1000 (2007).
133. Klampfer, L., Zhang, J., Zelenetz, A. O., Uchida, H. & Nimer, S. D. The AML1/ETO fusion protein activates transcription of BCL-2. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14059–14064 (1996).
134. Alcalay, M. *et al.* Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *J. Clin. Invest.* **112**, 1751–1761 (2003).
135. Müller-Tidow, C. *et al.* Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. *Mol. Cell. Biol.* **24**, 2890–2904 (2004).
136. Sun, X. *et al.* Comparative analysis of genes regulated in acute myelomonocytic leukemia with and without inv(16)(p13q22) using microarray techniques, real-time PCR, immunohistochemistry, and flow cytometry immunophenotyping. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **20**, 811–820 (2007).
137. Fazi, F. *et al.* Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. *Cancer Cell* **12**, 457–466 (2007).
138. Gaidzik, V. I. *et al.* RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 1364–1372 (2011).
139. Mendler, J. H. *et al.* RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **30**, 3109–3118 (2012).

140. Damm, F. *et al.* Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* **119**, 3211–3218 (2012).
141. Makishima, H. *et al.* Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* **119**, 3203–3210 (2012).
142. Kim, E. *et al.* SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* **27**, 617–630 (2015).
143. Zhang, J. *et al.* Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4726–4734 (2015).
144. Boissel, N. *et al.* Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia* **20**, 965–970 (2006).
145. Bos, J. L. *et al.* Amino-acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia. *Nature* **315**, 726–730 (1985).
146. Bentires-Alj, M. *et al.* Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. *Cancer Res.* **64**, 8816–8820 (2004).
147. Tartaglia, M. *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **34**, 148–150 (2003).
148. Zhao, Z. *et al.* Cooperative loss of RAS feedback regulation drives myeloid leukemogenesis. *Nat. Genet.* **47**, 539–543 (2015).
149. Thol, F. *et al.* Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* **123**, 914–920 (2014).
150. Mullenders, J. *et al.* Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med.* **212**, 1833–1850 (2015).
151. Lindsley, R. C. *et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* **125**, 1367–1376 (2015).
152. Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science* **336**, 1440–1444 (2012).
153. Wong, T. N. *et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**, 552–555 (2015).
154. King-Underwood, L. & Pritchard-Jones, K. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. *Blood* **91**, 2961–2968 (1998).
155. Rampal, R. *et al.* DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep.* **9**, 1841–1855 (2014).
156. Wang, Y. *et al.* WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Mol. Cell* **57**, 662–673 (2015).
157. Van Vlierberghe, P. *et al.* PHF6 mutations in adult acute myeloid leukemia. *Leukemia* **25**, 130–134 (2011).
158. Liu, T. C. *et al.* Mutation analysis of PTEN/MMAC1 in acute myeloid leukemia. *Am. J. Hematol.* **63**, 170–175 (2000).
159. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
160. Krönke, J. *et al.* Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. *Blood* **122**, 100–108 (2013).
161. Garg, M. *et al.* Profiling of somatic mutations of acute myeloid leukemia, FLT3-ITD subgroup at diagnosis and relapse. *Blood* (2015). doi:10.1182/blood-2015-05-646240
162. Parkin, B. *et al.* Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia. *Blood* **121**, 369–377 (2013).

163. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098 (2012).
164. Meggendorfer, M. *et al.* Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. *Haematologica* (2015). doi:10.3324/haematol.2015.127985
165. Takahashi, K. *et al.* Dynamic acquisition of FLT3 or RAS alterations drive a subset of patients with lower risk MDS to secondary AML. *Leukemia* **27**, 2081–2083 (2013).
166. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
167. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
168. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
169. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
170. Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179–1181 (2012).
171. Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).
172. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, 651–658 (2012).
173. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–650 (2012).
174. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118 (2012).
175. Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2548–2553 (2014).
176. Shima, T. *et al.* The ordered acquisition of Class II and Class I mutations directs formation of human t(8;21) acute myelogenous leukemia stem cell. *Exp. Hematol.* **42**, 955–965.e1–5 (2014).
177. Calado, R. T. & Young, N. S. Telomere diseases. *N. Engl. J. Med.* **361**, 2353–2365 (2009).
178. Song, W. J. *et al.* Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175 (1999).
179. Hahn, C. N. *et al.* Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **43**, 1012–1017 (2011).
180. Pabst, T., Eyholzer, M., Haefliger, S., Schardt, J. & Mueller, B. U. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **26**, 5088–5093 (2008).
181. Tawana, K. *et al.* Disease evolution and outcomes in familial AML with germline CEBPA mutations. *Blood* **126**, 1214–1223 (2015).
182. Polprasert, C. *et al.* Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* **27**, 658–670 (2015).
183. Saliba, J. *et al.* Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. *Nat. Genet.* **47**, 1131–1140 (2015).
184. Chen, X. *et al.* Relation of clinical response and minimal residual disease and their prognostic impact on outcome in acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 1258–1264 (2015).

185. Chen, Y. *et al.* Persistence of cytogenetic abnormalities at complete remission after induction in patients with acute myeloid leukemia: prognostic significance and the potential role of allogeneic stem-cell transplantation. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 2507–2513 (2011).
186. Marcucci, G. *et al.* Abnormal cytogenetics at date of morphologic complete remission predicts short overall and disease-free survival, and higher relapse rate in adult acute myeloid leukemia: results from cancer and leukemia group B study 8461. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **22**, 2410–2418 (2004).
187. Hirsch, P. *et al.* Interest of cytogenetic and FISH evaluation for prognosis evaluation in 198 patients with acute myeloid leukemia in first complete remission in a single institution. *Leuk. Res.* **38**, 907–912 (2014).
188. Freeman, S. D. *et al.* Prognostic relevance of treatment response measured by flow cytometric residual disease detection in older patients with acute myeloid leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **31**, 4123–4131 (2013).
189. Terwijn, M. *et al.* High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **31**, 3889–3897 (2013).
190. Jourdan, E. *et al.* Prospective evaluation of gene mutations and minimal residual disease in patients with core binding factor acute myeloid leukemia. *Blood* **121**, 2213–2223 (2013).
191. Yin, J. A. L. *et al.* Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. *Blood* **120**, 2826–2835 (2012).
192. Gorello, P. *et al.* Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (NPM1) gene mutations. *Leukemia* **20**, 1103–1108 (2006).
193. Krönke, J. *et al.* Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 2709–2716 (2011).
194. Inoue, K. *et al.* Long-term follow-up of minimal residual disease in leukemia patients by monitoring WT1 (Wilms tumor gene) expression levels. *Blood* **88**, 2267–2278 (1996).
195. Inoue, K. *et al.* WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. *Blood* **84**, 3071–3079 (1994).
196. Cilloni, D. *et al.* Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized WT1 assay to enhance risk stratification in acute myeloid leukemia: a European LeukemiaNet study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **27**, 5195–5201 (2009).
197. Ogawa, H. *et al.* The usefulness of monitoring WT1 gene transcripts for the prediction and management of relapse following allogeneic stem cell transplantation in acute type leukemia. *Blood* **101**, 1698–1704 (2003).
198. Abdelhamid, E. *et al.* Minimal residual disease monitoring based on FLT3 internal tandem duplication in adult acute myeloid leukemia. *Leuk. Res.* **36**, 316–323 (2012).
199. Kohlmann, A. *et al.* Monitoring of residual disease by next-generation deep-sequencing of RUNX1 mutations can identify acute myeloid leukemia patients with resistant disease. *Leukemia* **28**, 129–137 (2014).
200. Zuffa, E. *et al.* Revealing very small FLT3 ITD mutated clones by ultra-deep sequencing analysis has important clinical implications in AML patients. *Oncotarget* **6**, 31284–31294 (2015).

201. Bibault, J.-E. *et al.* Next-generation sequencing of FLT3 internal tandem duplications for minimal residual disease monitoring in acute myeloid leukemia. *Oncotarget* **6**, 22812–22821 (2015).
202. Klco, J. M. *et al.* Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *JAMA* **314**, 811–822 (2015).
203. Debarri, H. *et al.* IDH1/2 but not DNMT3A mutations are suitable targets for minimal residual disease monitoring in acute myeloid leukemia patients: a study by the acute leukemia french association. *Oncotarget* (2015). doi:10.18632/oncotarget.5645
204. Jäger, R. *et al.* Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. *Leukemia* **24**, 1290–1298 (2010).
205. Nangalia, J. *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N. Engl. J. Med.* **369**, 2391–2405 (2013).
206. Wiemels, J. L. *et al.* In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* **99**, 3801–3805 (2002).
207. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
208. Yasuda, T. *et al.* Leukemic evolution of donor-derived cells harboring IDH2 and DNMT3A mutations after allogeneic stem cell transplantation. *Leukemia* **28**, 426–428 (2014).
209. Wong, T. N. *et al.* Rapid expansion of pre-existing non-leukemic hematopoietic clones frequently follows induction therapy for de novo AML. *Blood* (2015). doi:10.1182/blood-2015-10-677021
210. Micol, J.-B. *et al.* Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations. *Blood* **124**, 1445–1449 (2014).

Table des illustrations

Figure 1 : Organisation schématique de l'hématopoïèse.....	5
Figure 2 : Facteurs de transcriptions impliqués dans la régulation de l'hématopoïèse myéloïde.....	8
Figure 3: Les 8 classes de mutations fréquemment observées dans les LAM d'après les données du Cancer Genome Atlas.....	20
Figure 4 : Mécanismes de régulation de la méthylation de l'ADN et effet des différentes mutations retrouvées dans les LAM sur la méthylation.....	21
Figure 5 : Exemple d'architecture clonale et d'évolution clonale à la rechute chez un patient atteint de LAM.....	31
Figure 6 : Exemple d'évolution clonale chez un patient atteint de LAM secondaire à un syndrome myélodysplasique.....	32
Figure 7 : Modèle de leucémogénèse par acquisition séquentielle de mutation.....	37
Figure 8 : Résumé du génotypage des colonies et des LTC-IC et de la phylogénie clonale déduite chez 3 patients pédiatriques.....	132
Figure 9 : Cinétique d'évolution des VAF de tous les variants détectés chez 11 patients ayant rechuté.....	136
Figure 10 : Cinétique d'évolution des VAF de tous les variants chez 2 patients ayant développé des LAM secondaires, et chez un patient n'ayant pas rechuté.....	137

Table des tableaux

Tableau 1 : Résumé de la classification FAB des LAM.....	15
Tableau 2 : Résumé des principales anomalies chromosomiques récurrentes dans les LAM, de leurs conséquences fonctionnelles et de leur valeur pronostique.....	16
Tableau 3 : Résumée de la classification des LAM proposée par l'OMS en 2008.....	19
Tableau 4 : Résumé des données cliniques et biologiques de 5 LAM de l'enfant.....	130
Tableau 5 : Résumé des données de séquençage d'exome pour les 5 patients pédiatriques.....	131

Résumé :

Les leucémies aiguës myéloïdes (LAM) dérivent de progéniteurs hématopoïétiques dans lesquels se sont accumulés des événements chromosomiques ou génétiques conduisant à leur transformation. Parmi elles les mutations de *DNMT3A* ont été démontrées comme pré-leucémiques. Dans ce travail, en établissant la hiérarchie clonale de multiples lésions récurrentes dans les LAM, nous démontrons que les événements impliquant les régulateurs de l'épigénétique comme *DNMT3A*, *TET2*, *ASXL1*, les réarrangements impliquant *MLL* ou le chromosome 20q sont les premiers événements dans le clone. A l'inverse, les mutations régulant la prolifération (voie *RAS*, *FLT3* ou autres tyrosines kinase) se produisent tardivement. Les événements précoces sont quasi constamment détectables en rémission complète. Par l'étude de patients en rechute, nous montrons que les clones qui persistent après traitement servent de réservoir pour les rechutes, avec une variation clonale qui augmente avec la durée de la rémission. Après xénogreffe à des souris NOD/SCID/IL2-R γ -null, les échantillons de patients porteurs de lésions épigénétiques précoces sont capables de repopulation hématopoïétique leucémique ou non leucémique, ce qui est la signature fonctionnelle des événements pré-leucémiques. Cette hiérarchie génétique est observée chez la majorité des patients. Cependant dans 1/3 des cas, il n'est pas retrouvé d'événement épigénétique précoce. Chez ces patients, on observe notamment des prédispositions germinales aux LAM, ou des mutations acquises de *TP53*. Chez 3 patients avec des LAM de novo, mais pas en cas de LAM induite ou secondaire, les mutations de *TP53* sont associées à des mutations de *DNMT3A* dans un même clone. Cela suggère que les mutations de *DNMT3A* et de *TP53* sont complémentaires pour obtenir une dominance clonale pré leucémique. En conclusion, dans la majorité des LAM, des hiérarchie clonales récurrentes, initiées par des lésions pré-leucémiques variées vont promouvoir l'expansion ou la variation clonale, jusqu'à l'émergence de la maladie. Ces données sont essentielles pour le développement de nouvelles stratégies personnalisées de suivi de la maladie résiduelle.

Mots clés : Leucémie aiguë myéloïde, architecture clonale, lésion pré-leucémiques, maladie résiduelle

Clonal architecture of acute myeloid leukaemias and consequences for minimal residual disease evaluation

Abstract :

Acute myeloid leukaemia (AML) emerge from haematopoietic stem/progenitor cells that acquire genomic or chromosomal aberrations, some being considered as pre-leukemic lesions. Here, by establishing the chronological hierarchy of multiple driver lesions in AML, we show that most events affecting epigenetic regulators - *DNMT3A*, *TET2*, *ASXL1* mutations, as well as *MLL* and chromosome 20q rearrangements - are the first lesions in the clone. In contrast to late mutations involving signalling pathways such as *FLT3* or *RAS*, these early lesions are frequently detectable in complete remission sample. By studying late relapses, we show that persistent clones behave as long-term relapse reservoirs, and variegate increasingly with delay to relapse. Cells from patients with early epigenetic defects can repopulate bone marrow of xenotransplanted NOD/SCID/IL-2R γ -null (NSG) mice with leukemic or non leukemic engraftment, a functional signature of pre-leukemic events. This genetic hierarchy is observed in most patients but in one third of them lesions in epigenetic regulators are not the first events. Some of these patients have genetic predisposition, or somatic mutations in *TP53*. In three de novo AMLs, but not in three secondary or therapy-related AMLs, these *TP53* mutations were accompanied by *DNMT3A* mutations in a unique clone. This suggests that *DNMT3A* mutations may complement *TP53* mutations to generate a dominant pre-leukemic clone. We conclude that in most AMLs recurrent genetic hierarchies, initiated by distinct pre-leukemic lesions, promote progressive clonal expansion or variegation, leading to the development of a full blown disease. These data are essential for the development of new personalised strategies for MRD evaluation.

Keywords: Acute myeloid leukaemia, clonal architecture, pre-leukemic lesions, minimal residual disease