

Université de Montréal

**Identification de déterminants impliqués dans la
différenciation des cellules souches embryonnaires**

par
Simon Fortier

Institut de Recherche en Immunologie et Cancérologie (IRIC)
Faculté de Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Doctorat
en Biologie Moléculaire
option Biologie des systèmes

Décembre 2014

© Simon Fortier, 2014

RÉSUMÉ

Les cellules souches ont attiré l'attention du public ces dernières années, grâce non-seulement à leur utilisation comme thérapies visant à s'attaquer à certains types de cancers, mais aussi en relation avec leur potentiel dans le domaine de la médecine régénérative.

Il est établi que le destin cellulaire des cellules souches embryonnaires (ESC) est régulé de façon intensive par un groupe de facteur clés agissant sur leur pluripotence. Il est néanmoins envisageable que certains déterminants influençant l'auto-renouvellement et la différenciation de ces cellules soient toujours inconnus. Afin de tester cette hypothèse, nous avons généré, en utilisant une méthode par infections virales, une collection de ESC contenant des délétions chromosomales chevauchantes que nous avons baptisée DeIES (Deletion in ES cells). Cette librairie contient plus de 1000 clones indépendants dont les régions délétées couvrent environ 25% du génome murin. À l'aide de cette ressource, nous avons conduit un criblage de formation de corps embryoïdes (EB), démontrant que plusieurs clones délétés avaient un phénotype de différenciation anormal. Nos études de complémentation sur un groupe de clones ont par la suite permis l'identification de *Rps14* - un gène codant pour une protéine ribosomale (RP) comme étant haploinsuffisant pour la formation de EB. Dans un deuxième temps, l'analyse approfondie des résultats de notre crible a permis d'identifier un groupe de gènes codants pour des RP qui semblent essentiels pour la différenciation des ESC, mais dispensables pour leur auto-renouvellement. De manière intéressante, les phénotypes anormaux de formation en EB les plus marqués sont associés à des délétions de RP qui se retrouvent au site de sortie des ARN messagers (ARNm) du ribosome, soit *Rps5*, *Rps14* et *Rps28*. Étonnamment, alors qu'un débalancement des RP conduit généralement à une réponse de type p53, l'haploinsuffisance de ces trois gènes ne peut être renversée par une simple réduction des niveaux d'expression de ce gène suppresseur de tumeurs. Finalement, nos études de profilage polysomal et de séquençage à haut-débit montrent une signature spécifique de gènes liés au mésoderme chez un clone hétérozygote pour *Rps5*, suggérant ainsi une explication au phénotype de différenciation p53-indépendant identifié chez ces ESC.

Nos travaux rapportent donc la création d'une ressource intéressante de génomique fonctionnelle qui a permis de mettre à jour le rôle essentiel que jouent les RP dans le processus de formation de EB. Nos résultats permettent aussi de documenter une réponse p53-indépendante suite à un débalancement de RP dans un contexte opposant l'auto-renouvellement et la différenciation des ESC.

Mots-clés : Cellules souches embryonnaires, différenciation, auto-renouvellement, corps embryoïdes, protéines ribosomale.

ABSTRACT

Stem cells have captured public's attention in the last years, thanks to their involvement in cancer therapies and also their huge theoretical potential in the regenerative medicine field. In order to translate this new technology to the clinic, a better understanding of their regulatory mechanisms is still needed.

It is well established that mouse embryonic stem cell (ESC) fate is highly regulated by core pluripotency factors. However, it is conceivable that novel self-renewal or differentiation regulators are not yet described. To investigate this possibility, we used a viral-based approach to generate a collection of ESC with nested chromosomal deletions called DelES (Deletion in ES cells). This library contains more than a thousand independent ESC clones highly enriched in chromosomal deletions which together cover ~25% of the mouse genome. Using this resource, we conducted an embryoid body (EB) differentiation screen and showed that several clones were having an abnormal EB formation phenotype. Complementation studies later identified *Rps14*-a ribosomal protein (RP) coding gene- as a novel haploinsufficient gene in EB formation from undifferentiated ESC. Further analyses of our screen results showed a strong bias for a subset of small subunit ribosomal protein genes which are critical for ESC differentiation but not for their self-renewal activity. Interestingly, the most severe differentiation phenotypes were found with ribosomal proteins associated to the ribosome's mRNA exit site, namely *Rps5*, *Rps14* and *Rps28*. While RP gene imbalance often leads to a p53 response that can be corrected by p53 suppression, ESC clones with decreased expression of mRNA exit site RP genes were surprisingly insensitive to p53 reduction, but were rescued by BAC or cDNA complementation, thus confirming the causative nature of these genes in the ESC phenotype. Finally, polysomal profiling and RNA-Seq studies showed that *Rps5* deleted ESC exhibit an abnormal mesodermal gene signature.

Together, our work presents a highly valuable resource for functional genomic studies in ESC and also highlights a novel p53-independent role linked to RP gene imbalance. Our results shed light on the relevance of these subunits for the developmental transition of ESC from a pluripotent to a differentiated state.

Keywords : Embryonic stem cells, differentiation, self-renewal, embryoid body, ribosomal proteins

Table des matières

RÉSUMÉ	iii
ABSTRACT	iv
Table des matières	v
Liste des figures	xi
Liste des tableaux	xiii
Liste des abréviations	xiv
Remerciements	xvii
CHAPITRE 1: INTRODUCTION	1
1.1 Cellules souches et destin cellulaire	2
1.1.1 Cellules souches embryonnaires	2
1.1.1.1 Développement et origines des ESC	3
1.1.1.2 Facteurs régulant la pluripotence des ESC	4
1.1.1.2.1 Voie de signalisation LIF/STAT3	4
1.1.1.2.2 Voie de signalisation BMP/Smad	5
1.1.1.2.3 Voie de signalisation Wnt/Tcf3	5
1.1.1.2.4 Oct3/4	5
1.1.1.2.5 Sox2	6
1.1.1.2.6 Nanog	6
1.1.1.3 Facteurs régulant la différenciation des ESC	6
1.1.1.3.1 Voie ectodermique	7
1.1.1.3.2 Voie endodermique	8
1.1.1.3.3 Voie mésodermique	8
1.1.1.3.4 Différenciation in vitro des ESC en corps embryoides	8
1.1.2 Cellules souches pluripotentes induites et promesses cliniques	9
1.1.2.1 Reprogrammation cellulaire	9

1.1.3 ESC et génomique fonctionnelle.....	11
1.1.3.1 Criblage gain de fonction	12
1.1.3.2 Criblage perte de fonction	12
1.1.3.2.1 Ciblage spécifique de gènes	13
1.1.3.2.2 Interférence à l'ARN	13
1.1.3.2.3 Génération de bris double brins: Nucléases à doigts de zinc, à domaines TALE et Cas9.	14
1.1.3.2.4 Mutations aléatoires induites par des agents chimiques	15
1.1.3.2.5 Ingénierie chromosomale; système Cre-loxP	16
1.1.3.2.6 Cellules souches embryonnaires haploïdes	16
1.2 Mécanismes et activité traductionnelle	18
1.2.1 Biogénèse des ribosomes	18
1.2.1.1 Composantes ribosomales	19
1.2.1.1.1 ARN ribosomaux	19
1.2.1.1.2 Protéines ribosomales	20
1.2.1.1.3 Autres composantes nucléolaires.....	21
1.2.1.2 Structures ribosomales	21
1.2.1.2.1 Le centre de transfert peptidique (PTC)	22
1.2.1.2.2 Site de l'ARNt de type E	22
1.2.1.2.3 Le tunnel à ARNm	22
1.2.1.3 Initiation de la traduction	23
1.2.2 Stress ribosomal	24
1.2.3 Activité traductionnelle et différenciation cellulaire	26
1.2.3.1 Analyse expérimentale de l'activité traductionnelle.....	27
1.3 Mise en contexte et objectifs des travaux	29
CHAPITRE 2: Genome-wide interrogation of mammalian stem cell fate determinants by nested chromosome deletions	30

2.1 Contribution des auteurs	31
2.2 Summary	32
2.3 Introduction	33
2.3.1 Mammalian genomes and ESC characteristics	33
2.3.2 Networks regulating ESC fate	33
2.3.3 Functional genomics in mammalian stem cells.....	34
2.4 Results.....	36
2.4.1 Generation and molecular characterization of DeIES resource	36
2.4.2 Genomic coverage of DeIES	37
2.4.3 Organization of DeIES collection in a public database	38
2.4.4 Quality control of DeIES clones.....	38
2.4.5 Functional screen of DeIES: in vitro EB formation	39
2.4.6 Complementation studies of DeIES family 9.....	40
2.4.7 Candidate gene evaluation for DeIES family 9	42
2.5 Discussion	44
2.6 Conclusion	46
2.7 Figures.....	47
2.8 Methods.....	57
2.8.1 Generation of DeIES collection.....	57
2.8.2 High-throughput functional assays.....	57
2.8.3 Embryoid body formation screening.....	57
2.8.4 DNA analyses.....	58
2.8.5 Mapping of proviral integration sites.....	58
2.8.6 Construction of database	59
2.8.7 BAC engineering.....	59
2.8.8 BAC complementation.....	59
2.8.9 Generation of chimeric mice	60

2.8.10 FISH analysis	60
2.9 Acknowledgements	61
2.10 References	62
2.11 Supplemental Figures and Tables	68
2.12 Supplemental Experimental Procedures	77
2.12.1 DeIES library generation	77
2.12.2 Cryopreservation of cells	78
2.12.3 Functional assays	78
2.12.3.1 Alkaline phosphatase detection	78
2.12.3.2 Flow cytometry analyses	78
2.12.3.3 Colony staining	79
2.12.4 PCR studies	79
2.12.4.1 Real-time quantitative PCR (Q-PCR) screening of presumptive trisomies	79
2.12.4.2 Q-PCR detection of hygromycin resistance gene	79
2.12.4.3 Mapping of proviral integrations	80
2.12.4.4 Q-RT-PCR	80
2.12.5 DeIES database functionalities	81
2.12.6 SelectaBAC modification strategy	81
2.13 Supplemental References	83
CHAPITRE 3: Haploinsufficiency screen highlights two distinct groups of ribosomal protein genes essential for embryonic stem cell fate	84
3.1 Contribution des auteurs	85
3.2 Abstract	86
3.3 Introduction	87
3.4 Results	88
3.4.1 Gene Ontology Term Analysis Identifies RP Genes as Prime Candidates for EB Formation.	88

3.4.2 Physiological Consequence of RP Haploinsufficiency	88
3.4.3 RP Genes Are Critical for EB Formation: Complementation Studies	89
3.4.4 p53-Independent Ribosomal Stress in a Subset of RP-Deleted Clones	89
3.4.5 Expression of Mesoderm Genes Is Impaired in <i>Rps5^{Δ/wt}</i> Mutants	90
3.5 Discussion	92
3.6 Figures.....	94
3.7 Materials and Methods.....	100
3.7.1 ESC Maintenance.....	100
3.7.2 EB Formation.....	100
3.7.3 Ribosome Profiling.....	100
3.7.4 GO Analysis	100
3.7.5 RNAseq.....	100
3.8 Acknowledgements	102
3.9 References.....	103
3.10 Supplemental Figures and Tables.....	106
3.11 Supplemental Materials and Methods.....	119
3.11.1 ESC Transfections	119
3.11.2 EB Formation	119
3.11.3 Ribosome Profiling.....	119
3.11.4 GO Analysis	120
3.11.5 Assessment of p53 Expression Levels.....	120
3.11.6 BrdU Pulse-Chase Assay.....	120
3.11.7 Annexin V/Propidium Iodide Staining.....	120
3.11.8 qRT-PCR.....	120
3.12 Supplemental References.....	121
Chapitre 4: Discussion et perspectives	122
4.1 Choix d'un crible fonctionnel de DeES	122

4.1.1 Formation de corps embryoides	122
4.1.2 Établissement d'un criblage spécifique à l'hématopoïèse.....	123
4.1.3 Utilisations de DeES et modèles alternatifs	124
4.1.4 Limite de la méthode.....	126
4.2 Déterminants régissant le destin cellulaire des ESC	127
4.2.1 Voie de signalisation p53-indépendante et protéines ribosomales	127
4.2.2 Rôle spécialisé de certaines protéines ribosomales.....	128
4.2.3 Rmrp.....	131
4.2.4 Autres déterminants identifiés dans le criblage	131
4.3 Conclusion	134
BIBLIOGRAPHIE	135

Liste des figures

Figure 1.1: Développement embryonnaire.....	3
Figure 1.2: Différenciation <i>in vitro</i> des ESC.....	7
Figure 1.3: Structures neuronales.....	7
Figure 1.4: Différenciation des ESC en corps embryoïdes (EB).	9
Figure 1.5: Cellules souches pluripotentes induites.....	10
Figure 1.6: Biogénèse des ribosomes.....	19
Figure 1.7: Principales structures du ribosome eucaryote.	23
Figure 1.8: Mécanisme du stress ribosomal.....	26
Figure 1.9: Évaluation de l'activité traductionnelle.....	28
Figure 2.1: Creation of DeIES resource.	47
Figure 2.2: Molecular characterization of DeIES clones.	49
Figure 2.3: Genomic coverage of DeIES.	50
Figure 2.4: Quality control of DeIES clones.....	51
Figure 2.5: Embryoid body formation screen.	52
Figure 2.6: Rescue of differentiation defects of tertiary clones from DeIES family 9.	53
Figure 2.7: Candidate gene evaluation of DeIES family 9.	55
Figure S2.1: Graphical representation of Ki67 values for each puro ^S tertiary clones grouped in normalized plate sets based on proliferation rate.....	68
Figure S2.2: DeIES interface.	69
Figure S2.3: BAC engineering for DeIES complementation.....	71
Figure 3.1: The EB-formation phenotype correlates with RP gene deletions.....	94
Figure 3.2: Complementation experiments reveal a site-specific role for RP genes.....	95
Figure 3.3: <i>Rps5</i> and <i>Rps28</i> are essential for EB formation.	96
Figure 3.4: p53-independent role of exit-site specific RP deletions in EB formation.	97
Figure 3.5: Polyribosomal AUC is increased upon EB formation.....	98

Figure 3.6: RNAseq analysis of ribosomal fractions.	99
Figure S3.1: Correlation between deletion size and EB numbers.....	106
Figure S3.2: Increased cell death upon EB formation.	107
Figure S3.3: BAC complementation.....	108
Figure S3.4: p53 status and evaluation of knockdown efficiency.....	109
Figure S3.5: Polyribosome profiles of undifferentiated Δ RP-ESCs.	110
Figure S3.6: Differential expression analysis of wild-type vs. <i>Rps5^{Δ/wt}</i> deleted clones.	111
Figure S3.7: Enrichment graphs of next-generation sequencing experiments.....	112
Figure 4.1: Détection de précurseurs hématopoïétiques lors du criblage de formation de EB...	124
Figure 4.2: Modèle de stress nucléolaire et ribosomal.....	129
Figure 4.3 Sévérité des phénotypes de formation de EB.	130

Liste des tableaux

Tableau 1.1: Présentation des principales ribosomopathies humaines.....	25
Tableau S2.1: Global statistics of DeIES library generation.....	72
Tableau S2.2: Molecular characterization details of DeIES clones used in functional assays.	73
Tableau S2.3: Summary of deletions distribution.	73
Tableau S2.4: Known or predicted haploinsufficient/imprinted candidate genes present in the vicinity of DeIES anchor loci that are not associated with chromosomal deletions.	74
Tableau S2.5: Genomic content of DeIES mapped deletions.	75
Tableau S2.6: Embryoid body screen raw data.	75
Tableau S2.7: List of oligonucleotides used in all reported assays.....	76
Tableau S3.1: Analysis of GO-term clustering from the EB-formation screen.	114
Tableau S3.2: DeIES families used in validation experiments.	115
Tableau S3.3: List of significant ($P_{adj} < 0.1$) genes obtained from DESeq analysis.....	116
Tableau S3.4: Analysis of GO-term clustering of polyribosomal fractions.	117
Tableau S3.5: Primers and probes used for qRT-PCR studies.	118
Tableau 4.1: Termes d'ontologie génique les plus significatifs des gènes affectant la formation de EB.....	132

Liste des abréviations

ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
ARNr	Acide ribonucléique ribosomal
ARNt	Acide ribonucléique de transfert
BAC	Chromosome artificiel bactérien
BMP	Protéines morphogénétiques osseuses
CHH	Hypoplasie cartilage-cheveux
CSH	Cellules souches hématopoïétiques
DBA	Anémie de type Diamond-Blackfan
DelES	Librairie de clones délétés (<u>De</u> letions in <u>E</u> mbryonic <u>S</u> tem cells)
DFC	Composante dense fibrillaire
EB	Corps embryoïdes
EC	Cellules embryonnaires de carcinome
eIF	Facteur d'initiation eucaryotique
ENU	N-nitroso-N-éthylurée
ESC	Cellules souches embryonnaires
FC	Centre fibrillaire
FGF	Facteur de croissance fibroblastique
G418	Généticine
GC	Composante granulaire
GO	Ontologie génique
GPCR	Récepteur couplé à une protéine G
ICM	Masse cellulaire interne
iPSC	Cellules souches pluripotentes induites
IRES	Site d'entrée interne du ribosome
LIF	Facteur inhibiteur de leucémie
LIFR	Récepteur du facteur inhibiteur de leucémie
LMC	Leucémie myéloïde chronique
lncARN	Longs ARN non-codant
MDS	Syndrome myélodysplasique

microARN	Petit ARN interférant endogène
ncARN	Acide ribonucléique non-codant
NPM1	Nucléophosmine
PIC	Complexe pré-initiation
Poll,II,III	Polymérase à ARN de type I, II et III
Pré-ARNr	Précurseurs d'ARN ribosomiaux
RISC	Complexe d'interférence induit par l'ARN
RP	Protéines ribosomales
Rpl	Protéine ribosomale de la grande sous-unité
Rps	Protéine ribosomale de la petite sous-unité
SBDS	Syndrome Swachman-Diamond
SHH	Voie de signalisation Sonic Hedgehog
shARN	Petit ARN en épingle à cheveux
siARN	Petit ARN interférant
snoARN	Petits ARN nucléolaires
snoRNP	Petits complexes ribonucléoprotéiques
T	Gène brachyury
TALE	Famille d'homéoprotéines avec une boucle de 3 acides aminés
TALLEN	Nucléases à domaines TALE
TCS	Syndrome de Treacher-Collins
TGF	Facteurs de croissance transformants
VEGF	Facteur de croissance vésiculaire endothélial
ZFN	Nucléases à doigts de zinc

À ma famille et mes amis, ma plus grande richesse

Remerciements

Je voudrais en premier lieu remercier mon directeur de thèse, Guy Sauvageau, pour avoir su me donner les justes doses de soutien lorsque nécessaire et de liberté lorsque souhaitée. Il faut aussi souligner sa grande disponibilité, et ce malgré un horaire surchargé. À mes collègues et amis du laboratoire, ce fut pour moi une chance incroyable de pouvoir vous fréquenter sur une base quotidienne, dans le contexte du travail comme à l'extérieur. Je ne peux passer sous silence l'apport indispensable des spécialistes des différentes plateformes technologiques de l'Institut, pour y avoir facilité mon séjour. Aussi, à toutes les personnes impliquées de près ou de loin dans le Défi Persévérance, je suis fier d'avoir pu soutenir cette cause et d'avoir contribué à la campagne de financement depuis ses débuts.

Je voudrais aussi remercier les membres du jury, dont les commentaires et recommandations auront certainement permis d'amener cette thèse à un niveau supérieur.

Sur un plan plus personnel, je serai bref, car ils le savent probablement tous déjà. À ma copine Shanti, ma famille immédiate et éloignée, mes amis et maintenant à mon fils Hector, vous êtes ce qu'il y a de plus important à mes yeux. Simplement merci d'avoir su m'accepter et m'endurer tout au long de ce parcours du combattant, au périple duquel mon humeur devenait probablement de plus en plus massacrate. Mes plus sincères excuses.

CHAPITRE 1: INTRODUCTION

Ce chapitre premier a pour objectif principal de mettre en place les notions qui seront nécessaires à la compréhension de cette thèse, et ainsi d'introduire le lecteur à l'univers fascinant des cellules souches. Dans une première section, les caractéristiques moléculaires et fonctionnelles des cellules souches embryonnaires (ESC) seront discutées plus en détail, de même que l'utilité de la recherche fondamentale employant ces dernières comme modèle expérimental. Par la suite, la seconde section de cette introduction fournira un bref survol, non-exhaustif, de la machinerie et de l'activité traductionnelle permettant la production de protéines à partir de séquences d'ARN messagers (ARNm). Ensemble, ces sections mettront la table aux résultats présentés respectivement dans les deux chapitres 2 et 3.

1.1 Cellules souches et destin cellulaire

Les cellules souches sont caractérisées par deux propriétés fondamentales; leur capacité d'auto-renouvellement de même que leur potentiel de différenciation. Ces deux concepts sont intimement liés. Lors de la division cellulaire, une cellule souche donnera naissance à deux cellules filles qui pourront soit posséder des propriétés identiques à celles de la cellule mère (auto-renouvellement), soit se commettre dans une voie de différenciation qui limitera ainsi leur potentiel comparativement à celui de la cellule progénitrice. Au point de vue de la nomenclature, des cellules souches dites totipotentes sont à même de générer tous les types de cellules germinales d'un organisme, soit des cellules de l'endoderme, de l'ectoderme et du mésoderme, en plus de pouvoir produire des cellules trophoblastiques (placenta), seule capacité que des cellules souches dites pluripotentes n'ont pas en commun avec ces dernières. En contrepartie, les cellules souches multipotentes sont caractérisées par un stade de différenciation plus avancé et sont ainsi plus limitées quant aux types cellulaires qu'elles peuvent générer. Dans cette classe, mentionnons par exemple les cellules souches hématopoïétiques (CSH), qui seront à l'origine de la production des diverses cellules sanguines spécialisées (érythrocytes, monocytes, granulocytes, etc). Comme on peut se l'imaginer, l'intégrité fonctionnelle des mécanismes moléculaires régissant ces processus de différenciation est essentielle au bon fonctionnement d'un organisme vivant.

1.1.1 Cellules souches embryonnaires

C'est au milieu des années 1970, grâce à des recherches portant sur le processus de développement tumoral, qu'un groupe de chercheurs ont remarqués que certaines cellules à l'origine des tumeurs (cellules embryonnaires de carcinome (EC)) partageaient d'importantes caractéristiques avec les cellules participant au développement embryonnaire¹. D'ailleurs, pendant de nombreuses années et encore de nos jours, bien que de moins en moins pratiqué², l'un des essais ultime permettant de juger le potentiel d'une cellule souche embryonnaire était la possibilité de ces cellules à former des tératocarcinomes, comprenant différents types cellulaires, lorsque transplantées *in vivo* chez un hôte³. C'est au tournant des années 1980 que les travaux de recherches de Martin, Evans et Kaufman ont conduit à l'identification chez la souris des cellules souches embryonnaires (ESC)^{4,5}. La contribution de ces cellules pluripotentes au développement embryonnaire a ensuite été démontrée lorsqu'elles ont été réinjectées dans un blastocyste receveur^{6,7}. Ces travaux, combinés à ceux portant sur la recombinaison homologue⁸ (qui décrivent l'introduction de gènes, marqueurs ou mutations dans le génome des ESC), sont à la base d'un tout nouveau domaine de recherche duquel découleront d'innombrables

découvertes dans les décennies suivantes⁹. En effet, non seulement l'avènement des techniques de génie génétique a permis l'élaboration de modèles expérimentaux répliquant des maladies humaines complexes, mais l'étude des mécanismes moléculaires régissant la pluripotence et la différenciation des ESC sera la pierre angulaire de ce que nous connaissons aujourd'hui sous l'appellation de médecine régénérative¹⁰.

1.1.1.1 Développement et origines des ESC

L'embryogénèse chez le modèle murin se déroule sur une durée d'environ 19 jours (Figure 1.1). Au tout début du processus, trois jours sont nécessaires pour passer du stade de zygote (cellule fécondée) à celui de blastocyste¹¹. À ce moment, ce dernier est constitué d'environ 32 cellules, soit celles formant le trophoctoderme (précurseur du placenta) et celles formant la masse cellulaire interne (ICM), d'où originent les ESC, elle-même composée de cellules endodermiques primitives et de cellules de l'épiblaste^{12,13}. L'implantation du blastocyste dans la paroi utérine marque une étape importante de l'embryogénèse: les processus de différenciation subséquents conduiront

au développement de cellules des trois lignées germinales, soit l'endoderme, le mésoderme et l'ectoderme. Au début du processus de gastrulation, les ESC pluripotentes de l'épiblaste vont former une structure appelée ligne primitive (primitive streak), qui est caractérisée par l'expression des gènes *Brachyury*¹⁴ et *Mixl1*¹⁵. De cette structure émergeront les cellules précurseurs du mésoderme (cellules provenant de la partie postérieure de la ligne primitive, exprimant le gène *Flk-1*¹⁶) ou de l'endoderme (cellules provenant de la partie antérieure de la ligne primitive, exprimant le gène *Foxa2*¹⁷) nécessaire à la poursuite de l'embryogénèse^{18,19}. Les cellules provenant de

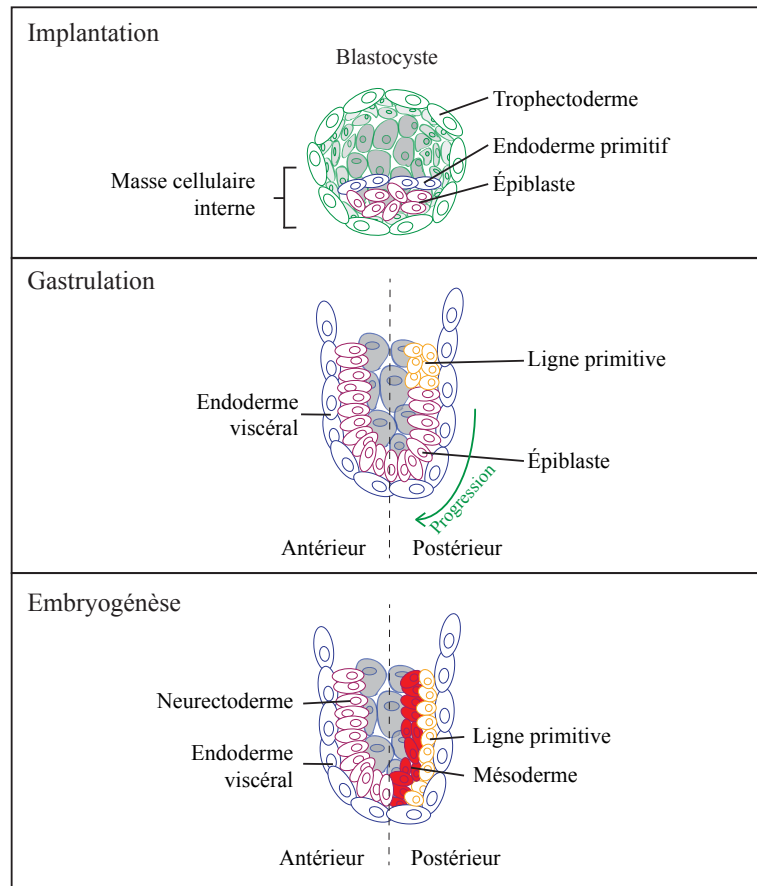


Figure 1.1: Développement embryonnaire. Étapes principales du développement *in vivo* de l'embryon.

l'épiblaste n'ayant pas transité par la ligne primitive donneront naissance aux cellules ectodermiques. La régulation spatiale et temporelle de l'embryogénèse suggèrent fortement que l'expression génique et la signalisation moléculaire dans l'environnement immédiat des cellules impliquées est primordiale et non aléatoire^{12,20,21}.

1.1.1.2 Facteurs régulant la pluripotence des ESC

Lors du développement embryonnaire, les ESC pluripotentes retrouvées dans l'ICM ne sont présentes que de façon transitoire. Ainsi, du moment où les ESC ont été isolées^{4,5}, il devint impératif de mettre à jour les mécanismes moléculaires contrôlant autant leur auto-renouvellement que ceux régissant leur différenciation, dans le but d'exploiter leur plein potentiel *in vitro*. Au stade du blastocyste, le facteur inhibiteur de leucémie (LIF), une cytokine de la famille des interleukines-6 qui a comme principale fonction d'inhiber la différenciation cellulaire, est produit de façon transitoire par les cellules du trophoblaste, tandis que son récepteur associé (LIFR) est fortement exprimé dans celles de l'ICM. Afin de reproduire cet environnement *in vitro*, les ESC dérivées de l'ICM doivent ainsi recevoir du LIF d'une source exogène, en plus de nécessiter la présence de protéines morphogénétiques osseuses (BMP)²², pour être maintenues et conserver leurs propriétés intrinsèques²³. Plus récemment, l'étude des facteurs régulant la pluripotence des ESC a permis d'établir des conditions optimales pour leur culture *in vitro*, délaissant l'utilisation du sérum en combinaison avec l'utilisation d'inhibiteurs spécifiques de voies de signalisation cruciales pour la pluripotence des ESC. La transmission de ces signaux via la voie des cytokines, de même que l'influence d'importants facteurs de transcription seront discutés dans les prochains paragraphes.

1.1.1.2.1 Voie de signalisation LIF/STAT3

La liaison du LIF au LIFR induit l'hétérodimérisation de ce dernier avec la sous-unité responsable de la propagation du signal, gp130, et active les voies de signalisations intracellulaires JAK/STAT et MAPK^{22,24} propre à cette famille. Cette transmission intracellulaire favorise la phosphorylation de STAT3, qui sera ainsi transloqué au noyau pour induire entre autres la transcription de Sox2²⁵, de même que ses cibles transcriptionnelles Oct3/4^{26,27} et Nanog²⁸. Ces gènes ont été identifiés comme essentiels au maintien des propriétés des ESC, jouant un rôle de premier plan dans l'activation de gènes spécifiques à l'auto-renouvellement et dans l'inhibition de gènes propres à la différenciation des ESC²⁹⁻³¹. Ces trois facteurs de transcriptions centraux de la pluripotence seront d'ailleurs étudiés plus en détails dans les lignes qui suivent.

1.1.1.2.2 Voie de signalisation BMP/Smad

Les BMPs représentent une des deux sous-famille de cytokines de la famille des facteurs de croissance transformants β (TGF- β), l'autre étant celle comprenant les TGF- β et activines³². Ces deux sous-familles de molécules diffèrent de par leur structure et leur récepteur cible (type I ou II), de même que par les voies de signalisation intracellulaire qu'elles activent. La liaison du ligand à son récepteur conduit à la phosphorylation intracellulaire du récepteur de type I et par le fait même à l'activation de sa fonction kinase. Les principaux substrats de ce récepteur activé sont les protéines Smads, plus spécifiquement Smad2 et Smad3 dans le cas d'une activation par la sous-famille TGF- β et Smad1, Smad5 et Smad8 suite à la liaison d'une cytokine de la sous-famille des BMPs. Une fois phosphorylées, les protéines Smad sont transloquées dans le noyau où, chez les ESC, elles influencent principalement la transcription des gènes inhibiteurs de la différenciation²².

1.1.1.2.3 Voie de signalisation Wnt/Tcf3

Dans la voie de signalisation canonique, les protéines de la famille Wnt se lient à un récepteur couplé à une protéine G (GPCR) de la famille Frizzled, conduisant ainsi à une accumulation de la β -caténine cytoplasmique. Une fois transloquée au noyau, cette molécule agit comme co-activateur transcriptionnel de facteurs de transcription de la famille des TCF/LEF³³. Chez les ESC, le facteur Tcf3 est fortement exprimé et partage plusieurs sites de liaisons à l'ADN avec les principaux gènes associés à la pluripotence, suggérant ainsi un rôle important dans la régulation de l'auto-renouvellement^{34,35}.

1.1.1.2.4 Oct3/4

Ce facteur de transcription, déjà fortement exprimé dans l'oocyte lors de l'embryogénèse, est le produit du gène *Pou5f1* et joue un rôle de premier plan dans la régulation du destin cellulaire des ESC en réprimant la différenciation. En effet, il a été démontré qu'autant une diminution qu'une augmentation de l'expression de ce gène avaient pour effet de pousser les ESC vers une différenciation trophoctodermique ou endodermique, respectivement^{26,36}. Des sites de liaison à l'ADN spécifiques à *Oct3/4*, situés dans des régions activatrices de nombreux gènes activateurs ou répresseurs de la différenciation, ont été identifiés chez les ESC^{37,38}, illustrant ainsi l'importance des niveaux d'expression d'*Oct3/4* dans le maintien des capacités pluripotentes des ESC.

1.1.1.2.5 Sox2

Les gènes Sox font partis de la famille des gènes contenant une séquence HMG (High Mobility Group-Box). Des expériences d'interférence à l'ARN dirigées contre le gène *Sox2* ont montré qu'une diminution de son expression résultait en une différenciation en cellules trophoctodermiques³⁹, un résultat similaire à celui obtenu lors d'une inhibition du gène *Pou5f1*. De façon intéressante, la présence des sites de liaison à l'ADN propres à *Sox2* à proximité de ceux spécifique à *Oct3/4* a démontré que ces deux facteurs de transcription agissent fréquemment comme co-activateurs et lient tous deux des promoteurs de gènes impliqués dans la régulation de la pluripotence des ESC³⁸. Toutefois, les effets d'une réduction du niveau d'expression de *Sox2* sur la différenciation cellulaire peuvent être renversés en surexprimant le gène *Oct3/4*, démontrant du même coup l'effet stabilisateur de *Sox2* sur le niveau d'expression de *Oct3/4* et le fait que la présence de *Sox2* aux promoteurs Oct - Sox spécifiques est dispensable au maintien de la pluripotence⁴⁰.

1.1.1.2.6 Nanog

En compagnie de *Oct3/4* et *Sox2*, la protéine *Nanog* est le troisième membre du trio de facteurs de transcriptions considérés comme le circuit transcriptionnel principal régissant la pluripotence des ESC en formant un boucle d'auto-régulation³⁸. La surexpression de *Nanog* favorise l'auto-renouvellement des ESC suivant un mécanisme LIF/Stat3 indépendant^{41,42} et il a été démontré que sa présence est indispensable à l'établissement et au maintien d'un état pluripotent d'origine²⁸. De façon similaire à ce que l'on peut observer en amont des séquences codantes des gènes *Oct3/4* et *Sox2*, des sites de liaison à l'ADN spécifiques à *Nanog* sont situés dans les régions promotrices de gènes régulant la pluripotence et la différenciation des ESC, où il n'est pas rare d'ailleurs d'observer la liaison de ces trois facteurs de transcription^{34,43}.

1.1.1.3 Facteurs régulant la différenciation des ESC

Comme décrit précédemment, les ESC n'étant présentes que de façon transitoire dans le processus *in vivo* d'embryogénèse, l'établissement de conditions assurant le maintien de leur auto-renouvellement *in vitro* était donc capital dans l'optique d'étudier par la suite les mécanismes régulant leur différenciation¹². Dans les paragraphes qui suivent, les principales voies de signalisation impliquées cette fois dans la différenciation des ESC vers les lignées germinales seront approfondies, de même que la transmission de ces signaux au niveau transcriptionnel (Figure 1.2).

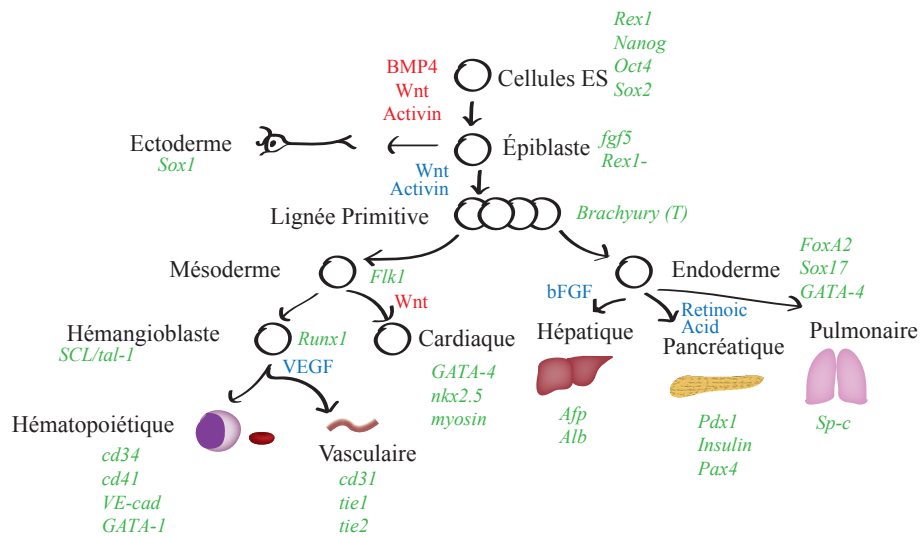


Figure 1.2: Différenciation *in vitro* des ESC.

Différenciation des ESC vers les lignées germinales. Les gènes caractéristiques de chaque étape sont identifiés (vert). Les molécules activatrices (bleues) et inhibitrices (rouges) sont aussi identifiées.

1.1.1.3.1 Voie ectodermique

La différenciation des ESC en cellules des lignées neuronale et de la peau est considérée comme étant celle par défaut, puisqu'en absence d'une activation des voies de signalisation BMP, Wnt et activine/Nodal, nécessaires à l'induction de la lignée primitive, l'apparition de tissus dérivés du neuroectoderme est spontanée (Figure 1.3). Ces cellules expriment le gène *Sox1* et leur différenciation en neurones est stimulée par le facteur de croissance FGF⁴⁴.

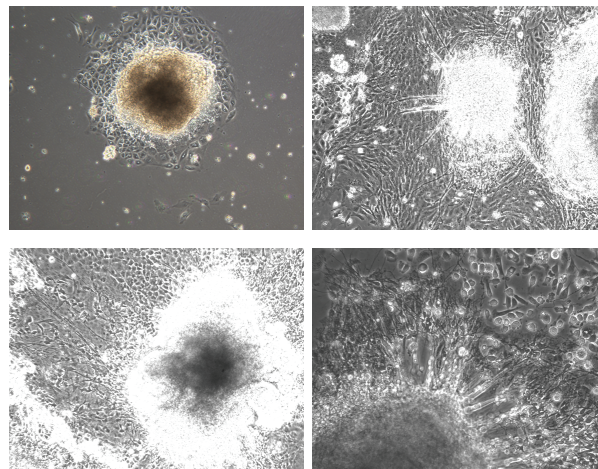


Figure 1.3: Structures neuronales.

Déprivées de LIF et BMP, les ESC, en présence d'acide rétinoïque, initient une différenciation ectodermique.

1.1.1.3.2 Voie endodermique

La stimulation des voies de signalisation Wnt et Nodal permet non seulement d'inhiber la différenciation en tissus ectodermiques, mais du même coup elle permet la formation de la lignée primitive, caractérisée par des cellules exprimant le gène *Brachyury* (T). De ces précurseurs seront dérivés les lignées propres au mésoderme et à l'endoderme. Des études menées sur ces cellules au potentiel mésendodermique ont permis d'établir que celles exprimant fortement le gène *Foxa2*, lorsque stimulées de façon soutenue par l'activine, donneraient naissance à des tissus de l'endoderme définitif¹⁷. Par la suite, la différenciation en tissus hépatiques et pancréatiques est favorisée soit par l'activation au FGF⁴⁵ ou par l'action de l'acide ascorbique combinée à l'inhibition de la voie Sonic Hedgehog (SHH)⁴⁶, respectivement.

1.1.1.3.3 Voie mésodermique

Les cellules appartenant à la lignée primitive exprimant faiblement le marqueur *Foxa2* et exposées à une activation de la voie BMP produiront des précurseurs mésodermiques exprimant le gène marqueur *Flk1*¹⁶. À ce stade, l'inhibition de la voie Wnt/ β -caténine conduit à l'apparition de précurseurs de la lignée cardiaque tandis que sa stimulation donnera naissance à un précurseur commun des lignées hématopoïétiques et endothéliales, l'hémangioblaste⁴⁷. Par la suite, l'ajout de VEGF et Wnt semblent essentiels pour générer une population hématopoïétique primitive¹⁶, récapitulant ainsi la première vague de l'hématopoïèse survenant dans le sac vitellin. À ce jour, très peu d'études rapportent la dérivation de cellules hématopoïétiques définitives ayant un potentiel reconstitution élevé^{12,48,49}, d'où la nécessité de poursuivre les études dans ce domaine.

1.1.1.3.4 Différenciation *in vitro* des ESC en corps embryoides

La caractérisation d'un modèle de différenciation récapitulant *in vitro* l'embryogénèse à l'aide de ESC fut rapportée par Doetschman et al., en 1985⁵⁰. Suite à la privation des signaux de pluripotence requis pour leur auto-renouvellement (LIF et BMP), les ESC initieront un programme de différenciation en formant des agrégats de cellules appelés corps embryoides (EB) (Figure 1.4). La prolifération des cellules conduira ensuite à l'établissement d'une structure tri-dimensionnelle s'apparentant à celle des premiers stades de l'embryogénèse *in vivo*, où les cellules internes et externes du blastocyste sont en contact avec des environnements signalétiques différents. Parmi les types cellulaires pouvant être retrouvés lors de ce processus de différenciation, les précurseurs hématopoïétiques sont rapidement générés et représentatifs de l'hématopoïèse normale, ce qui rend ce modèle très intéressant pour l'étude de la différenciation des ESC vers ces cellules sanguines spécialisées⁵¹.

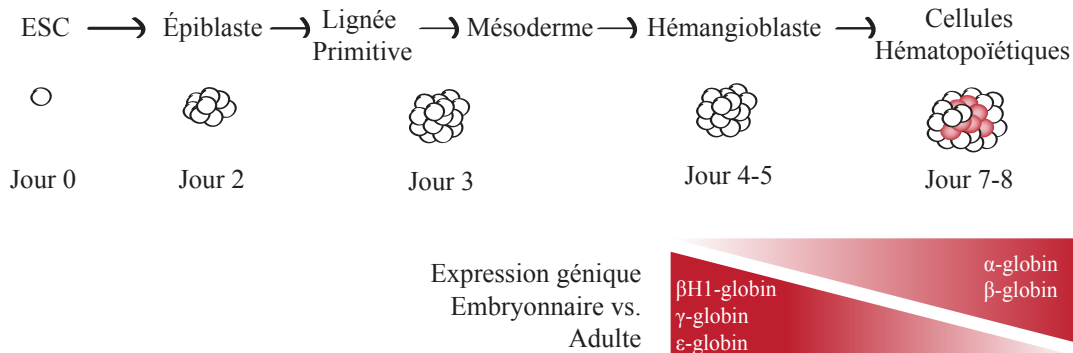


Figure 1.4: Différenciation des ESC en corps embryoides (EB).

En absence de LIF et BMP, les ESC amorcent un processus de différenciation en formant tout d'abord des agrégats tridimensionnels, à l'intérieur desquels des précurseurs hématopoïétiques seront identifiables après ~7 jours. La différenciation *in vitro* des ESC en EB est représentative des étapes de différenciation retrouvées lors du développement *in vivo*.

1.1.2 Cellules souches pluripotentes induites et promesses cliniques

Puisqu'elles possèdent la capacité de se différencier en tous les types cellulaires, les cellules souches sont vues comme ayant un potentiel énorme en médecine régénérative⁵². Deux types d'application sont envisagées; les utiliser pour remplacer un tissu endommagé en participant directement à sa régénération, ou encore en tant que vecteur pouvant transporter de potentielles molécules thérapeutiques vers ces mêmes tissus⁵³. Toutefois, la disponibilité et la compatibilité de ces dernières avec le système immunitaire du patient receveur font en sorte que ces utilisations sont techniquement très difficilement envisageables. Au cours des dernières années, d'importantes percées scientifiques, découlant directement des connaissances acquises sur l'auto-renouvellement et la différenciation des ESC, ont ouvert la porte à un tout nouveau domaine de la recherche sur les cellules souches; la reprogrammation de cellules somatiques adultes à un stade pluripotent. Ce «retour en arrière» dans le processus normal de développement pourrait ultimement permettre de palier aux problèmes d'immuno-compatibilité et de quantité de cellules nécessaires au traitement de diverses pathologies, que ce soit dans des cas de cancer nécessitant une transplantation (e.g. les cancers du système hématopoïétique), des blessures de la moëlle épinière ou encore d'insuffisance cardiaque, par exemple. (Figure 1.5).

1.1.2.1 Reprogrammation cellulaire

Briggs et King furent les premiers à démontrer, au début des années 1950, que le transfert d'un noyau adulte dans un oeuf énucléé était suffisant pour générer des descendants viables chez le modèle *Rana pipiens*⁵⁴. Par la suite, les travaux du chercheur anglais John B. Gurdon, principalement chez le modèle *Xenopus*, ont permis notamment d'établir les bases moléculaires de la technique de

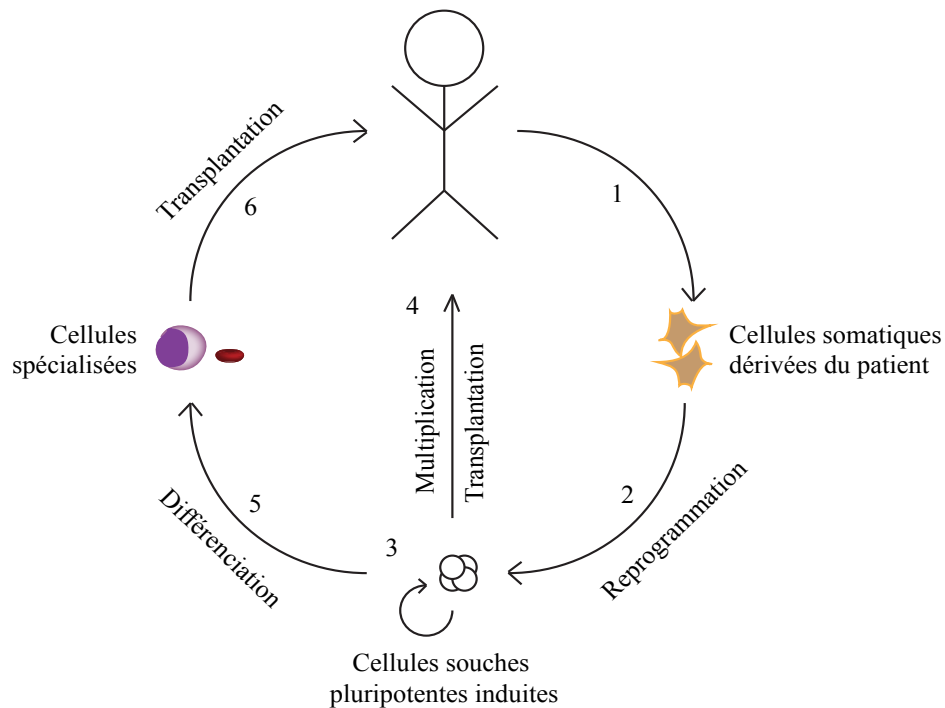


Figure 1.5: Cellules souches pluripotentes induites.

Schématisation des procédés visant à reprogrammer des cellules somatiques adultes d'un individu dans le but d'utiliser les tissus obtenus à des fins thérapeutiques chez cette même personne, évitant ainsi les problèmes d'histocompatibilité et les risques de rejets.

transfert nucléaire, où l'ADN d'intérêt d'une cellule somatique est substitué à celui d'origine dans un oeuf non-fécondé⁵⁵. De façon remarquable, ses recherches ont permis de mettre en évidence que la différenciation n'affectait pas le bagage génétique des cellules engagées dans ce processus, mais induisait plutôt un changement dans la régulation de l'expression de facteurs essentiels au développement⁵⁶. Globalement, ces travaux ont permis de mettre en lumière l'importance lors de la différenciation des molécules présentes dans le cytoplasme, ouvrant ainsi la voie à la découverte de facteurs ayant la capacité d'induire la reprogrammation nucléaire de cellules adultes différenciées^{57,58}. C'est au tournant des années 2000, en appliquant les connaissances acquises sur le transfert nucléaire, que la première démonstration d'un processus de reprogrammation fonctionnelle chez des mammifères fut rapportée⁵⁹ (le clonage de la brebis Dolly). Quelques années plus tard, il fut démontré que la fusion de cellules somatiques avec des ESC était suffisante pour produire une reprogrammation cellulaire^{60,61}. Ces avancées suggèrent une fois de plus que des facteurs essentiels à la pluripotence étaient présents chez les ESC ou encore dans l'ovocyte, conférant ainsi aux cellules somatiques une plasticité permettant un retour à un état de différenciation antérieur. C'est en 2006 qu'un groupe a démontré qu'une

combinaison de 4 facteurs était suffisante pour induire une «dédiﬀérentiation» de cellules somatiques^{62,63}. En effet, la coexpression des facteurs de transcription Oct3/4, Sox2, c-Myc et Klf4 fut suffisante pour reprogrammer des cellules ﬁbroblastiques murines à un niveau pluripotent, d'où leur nom de cellules souches pluripotentes induites (iPSC). Depuis cette découverte, des équipes s'attardent à raffiner les techniques de reprogrammation, puisque le cocktail original de Yamanaka implique la surexpression d'un puissant oncogène (c-Myc), ce qui pourrait s'avérer problématique dans l'optique d'utiliser ces cellules sur des patients humains. Une fois les cellules somatiques reprogrammées, des protocoles de différenciation vers les types cellulaires souhaités doivent être établis. Récemment, des travaux rapportent l'interconversion de cellules spécialisées, qui sont reprogrammées, par exemple, à partir de ﬁbroblastes en surexprimant des facteurs de transcriptions spéciﬁques à un autre type de cellule somatique^{64,65}, court-circuitant ainsi l'étape d'obtenir une cellule pluripotente.

Le potentiel lié à la reprogrammation cellulaire et à la dérivation de cellules partageant les caractéristiques des ESC à partir de tissus de patients permet d'envisager de multiples applications en recherche, que ce soit pour l'étude des mécanismes moléculaires spéciﬁques à diverses maladies ou encore à la mise sur pieds de criblages visant la découverte de nouvelles drogues^{31,66}. Ce volet de la recherche fondamentale étant encore tout jeune, il est plus que jamais impératif d'approfondir nos connaissances sur les mécanismes moléculaires impliqués dans les processus de pluripotence et de différenciation des cellules souches.

1.1.3 ESC et génomique fonctionnelle

Tel que décrit dans les paragraphes précédents, la caractérisation des propriétés intrinsèques des ESC a permis de développer un domaine de la recherche qui voit à l'utilisation de ces dernières en tant qu'outils permettant de répondre à diverses questions biologiques fondamentales. En effet, la possibilité d'introduire de façon exogène des modiﬁcations génétiques, mutations ou délétions à même le génome des ESC en fait un outil aujourd'hui indispensable afin d'établir des modèles *in vivo* servant aux études de génétique inverse, où l'on étudie les phénotypes découlant d'une modiﬁcation à un ou des gènes. Aussi, le développement de nouvelles techniques permettant d'obtenir des iPSC passe inévitablement par une compréhension poussée des mécanismes moléculaires ayant une incidence sur les propriétés caractéristiques d'auto-renouvellement et de différenciation des ESC. À cette fin, différentes approches expérimentales ont été développées au courant des dernières années pour permettre d'explorer d'une manière fonctionnelle le génome des ESC. Les méthodes permettant de mettre en valeur la fonction d'un gène peuvent être divisées en deux groupes: les approches par gain de fonction (surexpression)

et les criblages par perte de fonction (sous-expression). Au final, comme il sera décrit dans les prochains paragraphes, la nature des questions pouvant être répondues à l'aide de ces différentes techniques, de même que leurs avantages et inconvénients, justifie le choix du modèle expérimental à utiliser.

1.1.3.1 Criblage gain de fonction

Ce type de criblage nécessite normalement de créer une librairie d'ADN complémentaires (ADNc), qui une fois surexprimés dans les cellules cibles conduiront potentiellement à l'identification de facteurs donnant un avantage à ces dernières. Cette technique est relativement complexe, nécessitant entre autre le sous-clonage des ADNc dans un vecteur approprié au type cellulaire à infecter. Les criblages gain de fonction favorisent l'identification de gènes avantageant fortement les cellules suite à leur surexpression, au détriment de gènes candidats ayant un impact négatif ou de faible intensité. Ainsi, les gènes identifiés par une telle approche sont souvent en amont de voies de signalisation essentielles. Par exemple, le rôle déterminant de la protéine Nanog en tant que facteur central de la pluripotence des ESC a été mis à jour en surexprimant une librairie de 10^5 plasmides chez des ESC cultivées en absence de cytokines et des facteurs permettant leur auto-renouvellement. De ce fait, en surexprimant de façon importante des ADNs étrangers, Chambers et al. ont pu identifier un gène, *Nanog*, comme un acteur central agissant sur le maintien de colonies de ESC indifférenciées⁴¹. Toutefois, les niveaux d'expression élevés générés par cette méthode, qui sont loin de représenter les niveaux physiologiques normaux, de même que l'utilisation d'antibiotiques de sélection afin d'isoler des clones d'intérêt doivent être vu comme de potentiels inconvénients à cette approche. Pritsker et al. ont pu démontré qu'en employant toujours une approche par surexpression, mais en comparant plutôt à l'aide de biopuces (microarray) les gènes différentiellement exprimés par rapport à une condition contrôle, il était possible d'identifier plus de candidats jouant un rôle important dans la pluripotence des ESC⁶⁷. Il est donc dorénavent plus probable, grâce au développement de nouvelles technologies facilitant entre autre le séquençage des clones d'intérêt et rendant les cribles à haut-débit plus accessible, d'obtenir des résultats probants avec une telle approche.

1.1.3.2 Criblage perte de fonction

À l'opposé d'une stratégie de surexpression, il y a bien entendu celle où l'on tente d'altérer de façon négative l'expression de certains gènes (mutagénèse insertionnelle) ou régions génomiques (mutagénèse aléatoire) pour observer les effets encourus sur les cellules cibles.

1.1.3.2.1 Ciblage spécifique de gènes

Comme leur nom l'indique, ces méthodes de criblage visent des éléments génomiques déterminés, par exemple à petite échelle une famille de gènes, et sont à l'opposé d'une approche d'intégration aléatoire. Deux groupes ont rapporté la création de ressources permettant de cibler systématiquement des séquences d'ADN spécifiques.

Dans un premier temps, la ressource Micer⁶⁸ est une collection d'environ 150,000 vecteurs permettant l'intégration d'une séquence étrangère d'ADN par recombinaison homologe, générant ainsi des mutations qui provoquent un décalage du cadre de lecture normal ou encore des irrégularités au niveau de l'épissage. À ce jour, ~94,000 clones sont cartographiés correctement et disponibles à la communauté scientifique (www.sanger.ac.uk), pouvant ainsi être employés dans un contexte de disruption génique, de réarrangements chromosomiaux (délétions, inversions) et même pour induire des mutations bi-alléliques. Parmi les avantages d'une telle approche, il faut noter la spécificité du site d'intégration et de la modification introduite. Néanmoins, la charge de travail nécessaire pour développer les vecteurs d'intégration est importante et cette méthode permet l'étude de régions restreintes.

La seconde initiative est une ressource provenant de l'IGTC (International Gene-Trap Consortium⁶⁹) qui contient environ 125,000 lignées cellulaires d'ESC, cette fois ciblées à l'aide d'une méthode appelée «gene-trap»; en utilisant un vecteur d'intégration contenant un signal d'arrêt prématuré pour la traduction, la production de la protéine est ainsi altérée et le site d'intégration du vecteur «rapporté» par le marqueur utilisé (e.g. β -galactosidase). Cette ressource permet une couverture d'environ 40% des gènes codants du génome des ESC murines et ont l'avantage de conduire à l'obtention de mutations invalidantes. En étant disponibles directement dans les ESC (www.genetrap.org), les clones de cette librairie permettent donc d'évaluer rapidement *in vivo* le rôle des gènes d'intérêts. Bien que cette méthode soit relativement simple, le design des vecteurs d'intégration utilisés peut mener à des sites d'intégration préférentiels (e.g. haute activité transcriptionnelle) et cette technique est spécifiquement utile pour étudier la fonction des séquences codantes.

1.1.3.2.2 Interférence à l'ARN

Dans le but de diminuer l'expression de gènes spécifiques, l'interférence à l'ARN est une technique couramment utilisée, en utilisant des petits ARN interférant (siARN) ou des petits ARN en épingle à cheveux (short-hairpin RNA, shARN)⁷⁰. Ces deux approches font ultimement appel au même mécanisme cellulaire, soit celui d'une dégradation de l'ARN complémentaire

aux séquences introduites par le complexe RISC (RNA-Induced Silencing Complex)⁷¹. Les différences majeures entre les siARN et les shARN se situent dans leur structure initiale. Les siARN sont introduits dans les cellules cibles sous la forme d'un simple brin d'ARN, immédiatement accessible au complexe RISC. Bien que cette particularité soit un avantage, deux inconvénients majeurs sont associés à l'utilisation de siARN: leur effet transitoire, dû à l'instabilité des composés d'ARN, de même que leur haut taux d'effets non-spécifiques. En contrepartie, les shARN sont plutôt utilisés sous la forme d'une structure en épingle à cheveux d'ARN double brin, répliquant la présentation des microARN retrouvés naturellement dans les cellules. La maturation intracellulaire des shARN implique l'enzyme nucléaire Drosha, qui produira un pré-shARN exportable vers le cytoplasme. Ce dernier sera ensuite clivé par Dicer avant d'être pris en charge sous la forme d'un petit ARN simple brin par le complexe RISC. Les avantages de cette approche sont la persistance de l'effet d'inhibition (intégration du vecteur à l'ADN de la cellule hôte) et le plus bas taux d'effets non-ciblés. Cependant, l'utilisation de virus pour introduire les shARN dans les cellules cibles, peut potentiellement causer des problèmes pour l'utilisation subséquente de cette approche à des fins de thérapies.

D'un point de vue pratique, des approches utilisant des siARN et des shARN ont été utilisées chez les ESC humaines et murines, pour identifier des facteurs nécessaires à leur pluripotence. Chia et al. ont ainsi identifié et caractérisé, à partir d'une librairie de 21,121 siARN, le gène *PRDM14* comme un facteur essentiel à la pluripotence des ESC humaines⁷². D'autres exemples existent dans la littérature démontrant l'efficacité de siARN pour identifier des déterminants propres aux ESC murines^{73,74}. L'utilité des shARN à révéler des gènes essentiels à l'auto-renouvellement des ESC murines a aussi été démontrée. Des ARN interférant dirigés contre 65 gènes codant pour des facteurs de transcription et de liaison à l'ADN³⁹ ou encore ciblant 929 gènes codant pour des kinases et phosphatases⁷⁵ ont été employés avec succès pour révéler les fonctions de pluripotence propre à *Esrrb*, *Tbx3* et *Tcl1*, de même que celles de *Nme6* et *Nme7*, respectivement.

1.1.3.2.3 Génération de bris double brins: Nucléases à doigts de zinc, à domaines TALE et Cas9.

Ces nouvelles technologies permettant d'éditer le génome sont similaires en ce point qu'elles requièrent l'activité d'une endonucléase (FokI dans les deux premiers cas, Cas9 pour le système CRISPR-Cas9) pour créer des coupures doubles brins dans l'ADN, mais différent quant à leur mode de reconnaissance des séquences ciblées. Dans les cas des nucléases à doigts de zinc (ZFN), les domaines Cys2His2 peuvent reconnaître de façon spécifique des séquences d'environ 3 paires de base. En groupant plusieurs de ces domaines, il est possible d'augmenter la spécificité des ZFN envers la cible d'intérêt⁷⁶. Pour ce qui est des TALENS (nucléases à

domaines TALE), ce sont plutôt des domaines de liaison à l'ADN hautement conservés provenant des homéoprotéines TALEs, chaque domaine de 33-35 acides aminés comportant deux résidus hypervariables permettant la reconnaissance spécifique d'une base d'une séquence d'ADN. Ainsi, en combinant plusieurs de ces domaines, il est possible de générer un outil hautement spécifique à l'ADN, et ce de manière plus rapide et facile que pour l'élaboration de ZFNs⁷⁷. Finalement, la technique CRISPR-Cas9 fait plutôt appel à une séquence guide d'ARN complémentaire à la séquence d'ADN ciblée, qui sera ensuite dégradé sous l'action de la nucléase Cas9. Cette méthode est encore plus prometteuse que les deux premières, étant donné la nature du domaine d'interaction qui doit être généré (ARN-ADN versus complexe protéique-ADN pour les deux méthodes précédentes)^{78,79}. Néanmoins, la liaison des domaines d'interaction spécifiques à l'ADN par l'une de ces trois méthodes, une fois couplé à l'activité d'une nucléase, conduira à la formation de bris à l'ADN qui seront réparés par la machinerie cellulaire, soit par un processus de liaison d'extrémités non-homologues ou par recombinaison homologue, altérant ainsi le cadre de lecture original et menant à une disruption protéique⁸⁰. Au final, ces méthodes peuvent être vues comme une alternative à une approche par interférence à l'ARN décrites précédemment, plus efficaces et adaptables à grande échelle.

L'utilisation de ces techniques étant plus récentes, les rapports démontrant leur utilité à identifier des facteurs propres aux ESC sont peu nombreux. Toutefois, il est à prévoir que les travaux exploitant ces nouvelles technologies se multiplieront au cours des prochaines années⁸¹. À titre d'exemple, Koike-Yusa et al. ont généré récemment une librairie d'environ 88,000 RNA guides couvrant au bas mot 19,000 gènes codants du génome de ESC murines. Cette ressource leur a permis d'obtenir rapidement et avec une très grande efficacité des mutations bi-alléliques qu'ils ont par la suite utilisées pour faire la démonstration du potentiel unique de la technique CRISPR-Cas9 en criblant pour des gènes influençant la sensibilité des ESC à certaines toxines⁸².

1.1.3.2.4 Mutations aléatoires induites par des agents chimiques

Afin d'induire de façon aléatoire des modifications génomiques, une des premières méthodes développée fait appel à un agent mutagène, e.g. l'ENU, ou N-nitroso-N-éthylurée, qui induit des mutations à l'ADN de l'organisme exposé. Cette méthode a bien sûr l'avantage de ne pas cibler spécifiquement certains éléments, en plus de générer des modifications stables qui peuvent être transmises au niveau germinale et donc utilisables pour des études *in vivo*⁸³. Au nombre des inconvénients liés à cette approche, il faut principalement noter la possibilité d'obtenir des mutations multiples au sein du même génome de même que la faible taille des modifications (e.g. mutations ponctuelles pour l'ENU) induites négatifs de cette approche, rendant dangereuse

l'identification des sites de mutations. C'est la raison principale expliquant la prévalence des approches dites dirigées employées par plusieurs équipes de recherche au cours de la dernière décennie, bien qu'encore une fois le développement des techniques de séquençage à haut-débit pourrait sensiblement améliorer la capacité à cartographier les mutations.

1.1.3.2.5 Ingénierie chromosomale; système Cre-loxP

Établi au milieu des années 1990, le système Cre-loxP fait, comme son nom l'indique, appel à la capacité de l'enzyme Cre d'induire la recombinaison de séquences loxP et ainsi permettre la modification de séquence de nucléotides. En effet, dépendamment de l'orientation et de l'emplacement des séquences loxP introduites, cette approche permet de générer des délétions (même orientation), des inversions (orientation opposée) de même que des translocations (chromosomes non-homologues). L'utilisation de la méthode Cre-loxP est maintenant couramment utilisée pour inactiver de façon définitive ou conditionnelle des gènes d'intérêt chez les ESC dans le but de développer des modèles d'études *in vivo*, en plus de servir à la génération de larges délétions chromosomales. En effet, Ramirez-Solis et al. ont en premier décrit l'utilisation d'une approche, faisant appel à l'intégration par recombinaison homologue chez les ESC de deux sites loxP distants de 90kb à ~3Mb du chromosome 11, conduisant à l'isolation des clones recombinants qui comportent des délétions chromosomales⁸⁴. Le même groupe rapportait quelques années plus tard l'obtention d'une efficacité de recombinaison d'environ 10% pour des sites loxP séparés de quelques Kb seulement à des distances aussi grandes que 10Mb⁸⁵. La complexité des étapes successives de recombinaison homologue liées à cette approche, nécessaires pour livrer les sites loxP, a mené au développement de méthodes d'intégration alternatives, que ce soit par l'utilisation d'une infection virale pour établir le deuxième site loxP⁸⁶ ou encore en adaptant une méthode faisant entièrement appel aux rétrovirus et à leur capacité d'intégration aléatoire pour implanter les deux sites loxP dans le génome des ESC⁸⁷. De ce fait, la possibilité de générer de façon stable des modifications génomiques chez les ESC, ainsi que d'étudier des régions codantes et non-codantes grâce à des délétions de tailles variables sont des avantages certains de cette approche.

1.1.3.2.6 Cellules souches embryonnaires haploïdes

Une autre approche intéressante permettant de s'attarder à des mutations homozygotes chez les ESC a été développée au cours des dernières années. En effet, des modèles de ESC haploïdes ont été générés⁸⁸⁻⁹⁰ et a servi à mettre à jour des facteurs nécessaires à la transition entre pluripotence et différenciation⁹¹. Il fut toutefois intrigant de constater qu'une telle lignée puisse être établie,

considérant le fait que l'haploinsuffisance de certains gènes dans le processus de différenciation des ESC a déjà été démontrée (e.g. *Rps14*^{92,93}). Bien que les travaux publiés jusqu'à présent font état d'une contribution de ces cellules à la lignée germinale (e.g. une contribution de 2% de cellules haploïdes sont retrouvées à E6.5 au plus tard, indétectable par la suite⁸⁹), il faut souligner qu'une diploïdisation rapide est aussi observée lorsque ces cellules sont engagées dans une voie de différenciation, aussi bien dans des contextes *in vitro* qu'*in vivo*. D'ailleurs, de multiples rondes de purification par cytométrie en flux sont nécessaires afin de maintenir cette population cellulaire, témoignant de l'instabilité de cet état. Néanmoins, en comparaison avec les approches mentionnées préalablement, les ESC haploïdes ont l'avantage de pouvoir révéler rapidement des phénotypes récessifs suite à des altérations (mutations, délétions, interférence) modifiant l'allèle restante, comparativement aux autres approches qui sont somme toute moins directes pour générer des modifications bi-alléliques.

1.2 Mécanismes et activité traductionnelle

Le terme traduction fait référence au processus global par lequel les ARNm sont décodés par les ribosomes pour résulter en la production de protéines. À ce jour, l'analyse transcriptionnelle des facteurs régulant l'activité et l'auto-renouvellement des cellules souches est abondamment documentée. Il a en outre été démontré que l'élongation des transcrits est l'étape limitante chez les ESC humaines⁹⁴ et que la production protéique de la majorité des gènes était limitée par les niveaux d'expression des transcrits. Dans ce contexte, la régulation au niveau de la traduction et des modifications post-traductionnelles a historiquement été considérée comme un mécanisme ayant un rôle de deuxième ordre sur l'expression protéique⁹⁵. Toutefois, les mécanismes de régulations post-transcriptionnels agissant au niveau de la pluripotence des cellules souches sont de plus en plus décrits dans la littérature, levant ainsi le voile sur des acteurs moléculaires jusqu'ici non-caractérisés⁹⁶.

1.2.1 Biogénèse des ribosomes

Les ribosomes se retrouvent autant chez les procaryotes et les eucaryotes, et sont les machines moléculaires essentielles à la traduction des séquences d'ARNm en protéines. La majorité de leur assemblage se déroule dans le nucléole, avant qu'ils ne soient importés dans le cytoplasme où la traduction des ARNm en protéines a lieu. Le nucléole est compartimentalisé en fonction des activités moléculaires qui s'y tiennent. Ainsi, on retrouve des molécules associées à la maturation des ARN ribosomiaux (ARNr) dans les régions du centre fibrillaire (FC) et de la composante dense fibrillaire (DFC), tandis que les sous-unités ribosomiales nouvellement formées et leurs facteurs associés sont concentrés dans la région de la composante granulaire (GC) du nucléole⁹⁷. De façon temporelle, les étapes nucléolaires menant à l'obtention de ribosomes cytoplasmiques matures peuvent être résumées ainsi: la transcription des précurseurs d'ARN ribosomiaux (pré-ARNr) à partir de séquences d'ADN ribosomal (ADNr), l'association de facteurs d'assemblage, de protéines ribosomiales (importées du cytoplasme) et de petits complexes ribonucléoprotéiques (snoRNP) essentiels à la maturation des pré-ARNr et finalement le processus d'exportation des sous-unités ribosomiales⁹⁸ (Figure 1.6).

Bien qu'ils soient conservés dans les différents règnes, les ribosomes matures ont un poids moléculaire de 1.35KDa supérieur chez les eucaryotes comparativement aux ribosomes bactériens, reflétant ainsi leurs fonctions additionnelles spécifiques. En comparaison, les données cristallographiques ont permis d'identifier 350KDa de rRNA additionnels chez le modèle eucaryote, 800KDa de nouvelles protéines et 200KDa de nouveaux domaines eucaryotes-

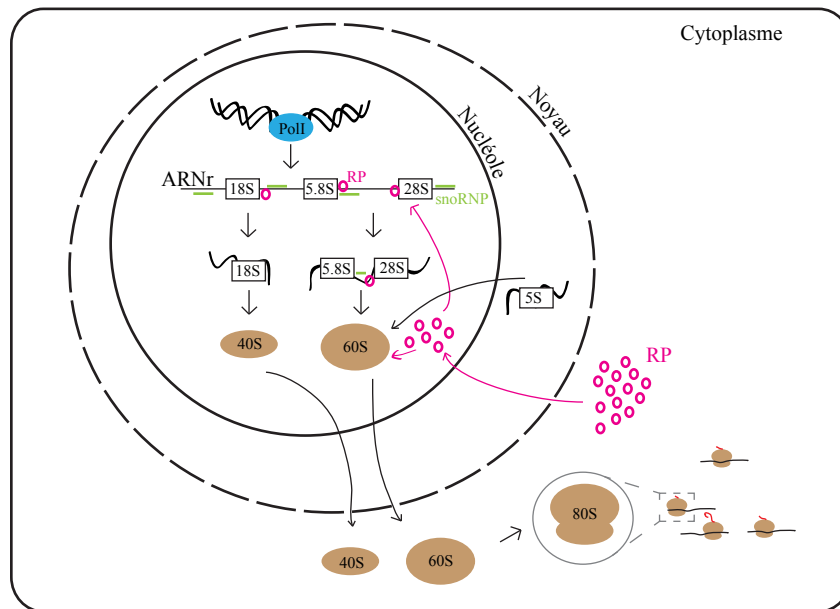


Figure 1.6: Biogénèse des ribosomes.

Principales étapes et localisation intracellulaire du processus de formation des ribosomes. Les ARN ribosomiaux mûrent dans le nucléole, à l'exception du fragment 5S qui est importé dans le nucléole après sa synthèse nucléaire. Les sous-unités ribosomiales sont assemblées dans le nucléole avant d'être exportées dans le cytoplasme où elles se joignent sur les ARN messagers à traduire. Les protéines ribosomiales de même que les petits complexes ribonucléoprotéiques participent activement à la maturation des ARNr.

spécifiques ajoutés à des protéines conservées entre les deux espèces⁹⁹. Dans les prochaines sections, les composantes nécessaires à la formation d'un ribosome fonctionnel seront tout d'abord décrites, avant de s'attarder plus spécifiquement à leur structure.

1.2.1.1 Composantes ribosomiales

Les ribosomes eucaryotes matures (80S) sont composés de deux sous-unités principales; la grande sous-unité ribosomiale 60S, formée des ARN ribosomiaux 28S, 5.8S et 5S et de 46 protéines ribosomiales (Rpl, ribosomal proteins of the large subunit), et la petite sous-unité ribosomiale 40S, formée de l'ARNr 18S et de 33 protéines ribosomiales (Rps, ribosomal proteins of the small subunit)¹⁰⁰.

1.2.1.1.1 ARN ribosomiaux

La transcription de l'ADN ribosomal afin de générer les ARNr est dépendante de l'activité de l'ARN polymérase I (PolI) (18S, 25S et 5.8S) ou de l'ARN polymérase III (5S). Les ARN ribosomiaux forment la partie catalytique des ribosomes et sont chargés non seulement

de la reconnaissance de séquences complémentaires d'ARN, mais aussi du recrutement des protéines ribosomales et autres complexes protéiques nécessaires à la traduction. Les 4 ARNr constituant le ribosome sont tous issus de pré-ARNr qui sont clivés à plusieurs reprises lors de la maturation nucléolaire, nécessitant par le fait même un nombre important de facteurs associés; endo et exo-ribonucléases, hélicases à ARN, facteurs d'assemblage et autres petits complexes ribonucléoprotéiques nucléolaires¹⁰¹. Tafforeau et al. ont récemment identifié plus de 250 facteurs impliqués dans le processus de clivage des ARNr humains, et de façon étonnante 74 facteurs spécifiques qui ne sont pas retrouvés chez les levures¹⁰². De par la multitude d'étapes devant conduire à la formation essentielle d'ARNr matures, plusieurs drogues couramment utilisées pour le traitement de divers cancers ont pour cibles des molécules impliquées dans ce processus de maturation¹⁰³. En règle générale, puisque les ARNr sont la pièce maîtresse de la biogénèse des ribosomes, une perturbation de leur synthèse cause un démantèlement rapide du nucléole, pouvant mener à une perturbation du cycle cellulaire et induire l'apoptose¹⁰⁴. Ces mécanismes de régulation sont décrits plus en détails à la section 1.2.2.

1.2.1.1.2 Protéines ribosomales

Contrairement aux ARNr, la transcription de l'ADN des protéines ribosomales est assurée par l'ARN polymérase II (PolII) et la traduction des ARNm générés est assurée dans le cytoplasme. Lorsqu'il fut démontré, au tournant des années 2000, que l'ARNr était l'unité catalytique des ribosomes, les protéines ribosomales étaient alors considérées comme ayant un rôle plutôt structural, en facilitant notamment le repliement des séquences d'acides aminés¹⁰⁵. Toutefois, des études plus récentes ont mis à jour les rôles extra-ribosomaux de certaines protéines ribosomales, suggérant ainsi que ces dernières pourraient en plus avoir un impact sur des processus indépendants de leur fonction structurale au sein du ribosome¹⁰⁶⁻¹⁰⁸. Pour ne citer que quelques exemples, mentionnons notamment le rôle de certaines protéines ribosomales dans; le processus de maturation des ARNr¹⁰⁹⁻¹¹¹, l'exportation des sous-unités ribosomales du noyau vers le cytoplasme^{112,113} et la liaison à certains facteurs d'assemblage essentiels à la maturation des ribosomes (e.g. Rio2 et Nob1)¹¹⁴.

Lorsque l'on analyse les niveaux d'expression génique de divers types de cellules eucaryotes, les transcrits codant pour les protéines ribosomales sont parmi les plus abondants. Ainsi, au niveau de leur régulation, il a été démontré que cette situation mène à une dégradation rapide au protéasome des protéines produites en excès^{115,116}. Les niveaux d'expression de ces dernières doivent en effet être maintenus de façon stricte, de manière à retrouver non seulement des quantités équimolaires de chacun des constituants protéiques du ribosome mature¹¹⁷, mais aussi

de conserver les proportions protéines/ARNr¹¹⁸. Il n'est ainsi pas surprenant que des mutations hétérozygotes, de même que l'haploinsuffisance, de plusieurs protéines ribosomales ont été répertoriées dans diverses pathologies humaines (section 1.2.2). Quelques unes d'entre elles seront décrites dans les sections suivantes.

1.2.1.1.3 Autres composantes nucléolaires

Toujours localisés dans le nucléole, un nombre important de facteurs d'assemblages et de complexes ribonucléoprotéiques (plus de 75) y jouent divers rôles, étant entre autre impliqués dans la maturation des ARNr, l'assemblage et l'exportation des sous-unités ribosomales vers le cytoplasme en plus de participer activement aux cascades de signalisation des anomalies perturbant l'homéostasie nucléolaire.

Les modifications et la maturation des ARNr sont ainsi coordonnées par les snoRNP, des complexes composés de petits ARN nucléolaires (snoARN), qui assurent la reconnaissance des sites d'action spécifiques de par leur complémentarité avec les ARNr, et de protéines, qui elles catalysent les réactions au niveau des groupements 2'-O-ribose des molécules d'ARNr^{101,119}. Plus spécifiquement, la fibrilarine (Nop1p) est l'enzyme principale des complexes RNP impliqués dans les modifications de méthylation (complexes qui comprennent aussi Nop56p, Nop58p et Snu13p), tandis que la dyskérine (Cbf5p) est l'unité catalytique de ceux responsables de la pseudouridylation (comprenant aussi Gar1p, Nhp2p et Nop10p) du même groupement¹²⁰. Évidemment, d'autres facteurs jouent un rôle important dans ce processus de maturation. Par exemple, les hélicases Fal1, Rok1 et Rrp3, de même que Dbp3, Dbp6 et Mak5, entre autres, possèdent des fonctions les liant aux snoRNA et aux sous-unités ribosomales 40S et 60S, respectivement^{121,122}. Encore une fois, le nombre important de facteurs associés aux ribosomes vient complexifier l'identification de rôles spécifiques pour ces derniers. Le développement de nouvelles techniques expérimentales, permettant notamment l'identification de complexes protéines-ARNr¹²³ et la reconstitution *in vitro* de la maturation des sous-unités ribosomales¹²⁴, devrait permettre dans un avenir rapproché de faire des avancées considérables en ce qui concerne les fonctions moléculaires d'acteurs essentiels à la maturation des ribosomes.

1.2.1.2 Structures ribosomales

Chacune des sous-unités ribosomales est caractérisée par des régions d'intérêt aux niveaux fonctionnel et structurel, qui confèrent aux ribosomes leur activité traductionnelle. Les données de cristallographie décrivant la structure du ribosome eucaryote étant récentes^{99,125,126}, les propriétés spécifiques de ce dernier, comparativement au modèle procaryote qui a été plus

longuement caractérisé, sont maintenant étudiées de façon extensives. De manière intéressante, les premières études comparatives ont démontré que peu de différences entre eucaryotes et procaryotes sont observables dans les parties internes du ribosome par rapport aux modifications notées en surface, suggérant ainsi que ces régions internes fortement conservées sont essentielles à l'activité de traduction⁹⁹. Dans les prochaines sections, les structures critiques assurant le fonctionnement du ribosome eucaryote seront décrits.

1.2.1.2.1 Le centre de transfert peptidique (PTC)

C'est la région catalytique du ribosome, comprenant les sites A (aminoacyl) et P (peptidyl) de fixation des ARN de transfert (ARNt), où sont formés les nouveaux liens peptidiques. Le PTC est localisé sur la sous-unité 60S et fait parti d'une structure conservée appelée protubérance centrale (Figure 1.7). Plusieurs inhibiteurs de la synthèse protéique ciblent cette région pour empêcher la formation de protéines *de novo*. Outre la réaction enzymatique à l'origine des nouveaux liens peptidiques, les plus récentes études cristallographiques démontrent que les interactions entre ARNr-protéines ribosomales-ARNm au site P sont aussi importantes afin d'induire les réarrangements structuraux nécessaires à la progression du ribosome sur l'ARNm¹²⁷. Les études soulignant l'importance de cette région viennent compléter celles de Kurkcuoglu et al., qui avaient déjà démontré que le positionnement de l'ARNm au «centre de décodage» était crucial afin d'assurer la fidélité de la traduction¹²⁸.

1.2.1.2.2 Site de l'ARNt de type E

Situé majoritairement sur la sous-unité 60S, cette fois dans une région structurale nommée protubérance L1, on retrouve le site E (Exit) de fixation des ARNt (Figure 1.7). Cette région est aussi adjacente au tunnel de sortie des protéines de même qu'à celui de l'ARNm, localisés sur les sous-unités 60S et 40S, respectivement. Les études de modélisation ont permis d'établir la spécificité pour le site E de certains inhibiteurs de la traduction couramment employés, tel la cycloheximine, qui compétitionnent directement avec l'ARNt de type E pour l'occupation de ce domaine ribosomal¹²⁹. Cette région du ribosome, bien qu'étant une des moins mobiles¹²⁸, a été identifiée comme essentielle à la translocation¹³⁰ et à la maintenance d'un cadre de lecture adéquat¹³¹.

1.2.1.2.3 Le tunnel à ARNm

Située principalement sur la petite sous-unité ribosomale 40S, le tunnel de l'ARNm a pour fonction de guider le ribosome le long du fragment d'acides aminés à traduire. Cette région est

une de celles démontrant le plus d'éléments eucaryotes spécifiques, en comparaison avec le modèle procaryote. Alors que la région d'entrée du tunnel est importante pour l'initiation de la traduction au site A, en conférant à cette région, via des domaines d'interaction entre protéines ribosomales et ARNm, une activité hélicase permettant l'admission de l'ARNm¹²⁶, le site de sortie de l'ARN messenger (à proximité du site E, dans une région appelé la plateforme de la petite sous-unité) est grandement responsable de la fidélité du processus traductionnel¹²⁸. De façon intéressante, la structure du ribosome eucaryote au tunnel de l'ARNm a permis d'identifier une région riche en structures secondaires s'apparentant à celle spécifique à la séquence Shine-Dalgarno (initiation de la traduction) chez les procaryotes¹²⁵, suggérant ainsi un rôle pour cette région dans la reconnaissance du codon initial. Appuyant cette hypothèse, le contact de ce site avec un site d'entrée interne du ribosome (IRES) a par la suite été rapporté⁹⁹.

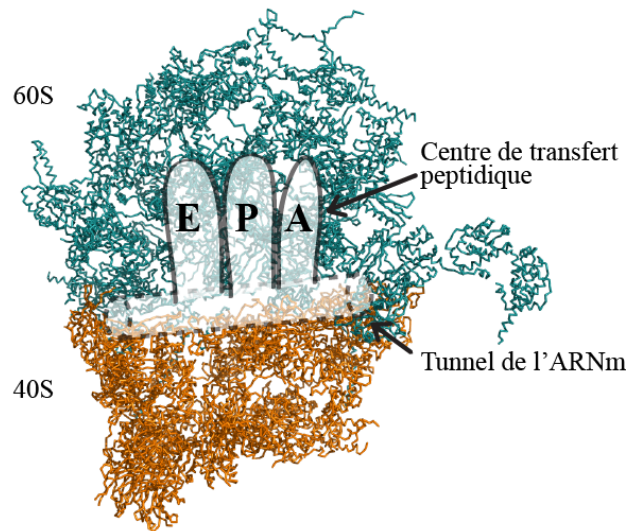


Figure 1.7: Principales structures du ribosome eucaryote.

Schématisme du ribosome eucaryote mature 80S, formé des sous-unités 60S (turquoise) et 40S (orange). Les sites de liaison des ARNt sont identifiés (E, P et A), de même que le tunnel de l'ARNm. Le tunnel de sortie des nouvelles protéines est situé tout juste au dessus du site E, mais n'est pas visible sur cette représentation. Structure réalisée à partir des modèles PDB3J3A-3J3B du ribosome humain.

1.2.1.3 Initiation de la traduction

Comme mentionné précédemment, les sous-unités ribosomales formées au niveau du nucléole sont exportées séparément vers le cytoplasme, où l'initiation de la traduction aura lieu. À ce stade, de nombreux facteurs d'initiation (eIF) viendront contribuer à la formation du ribosome 80S qui assurera la traduction de l'ARNm.

De façon plus spécifique, un complexe de pré-initiation (PIC) 43S est généré, formé par la petite sous-unité 40S, le facteur d'initiation eIF2 lié à un ARN de transfert (ARNt) reconnaissant le

codon de départ (Met-ARNt) en plus des facteurs associés eIF1, eIF3 et eIF5. Le PIC se lie ensuite à l'ARNm préalablement identifié (dans le cas d'une traduction CAP-dépendante) par la présence des facteurs de la famille eIF4 à l'extrémité 5' de la séquence de nucléotides. Ce complexe inspectera ensuite la séquence d'ARNm à la recherche d'un codon de départ compatible avec l'ARNt chargé au site P. Une fois identifié, les facteurs d'initiation seront relâchés, ce qui permettra à la grande sous-unité de se joindre et amorcer la synthèse protéique¹³²⁻¹³⁴.

1.2.2 Stress ribosomal

De par la multitude de facteurs requis uniquement pour procéder à la biogénèse des ribosomes, il est facile de concevoir que la régulation moléculaire au niveau du nucléole est primordiale à la poursuite normale des activités cellulaires. À prime abords considéré uniquement comme le site principal de génération des ribosomes, les données les plus récentes tendent à démontrer que le nucléole possède d'autres fonctions essentielles à la cellule, par exemple la survie cellulaire, la réparation de l'ADN et la régulation du cycle cellulaire^{135,136}.

Il est proposé qu'un des mécanismes principaux expliquant les multiples fonctions des nucléoles est en fait leur capacité à séquestrer des facteurs impliqués dans des voies de signalisation importantes. Les évènements moléculaires ayant des effets négatifs sur les nucléoles sont regroupés sous les termes de stress nucléolaire ou stress ribosomal.

En induisant des dommages à l'ADN et en inhibant de façon spécifique l'activité de la polymérase I, Rubbi et Milner ont démontré que le stress nucléolaire menait à un arrêt du cycle cellulaire déclenché par une surexpression de p53¹³⁷. Ce mécanisme identifiait donc le nucléole comme un important senseur du stress cellulaire pouvant induire une réponse p53-dépendante. Depuis ce temps, plusieurs facteurs nucléolaires ont été documentés dans la littérature¹⁰³. Un de ceux-ci est la protéine suppresseur de tumeur Arf (p19^{Arf} chez la souris et p14^{ARF} chez l'humain), surexprimée en contexte oncogénique, qui a la capacité d'activer p53 en liant le domaine acide central de MDM2, inhibant ainsi l'activité ubiquitine ligase de MDM2 qui conduit normalement à la dégradation de p53¹³⁸. Dans un contexte cellulaire normal, Arf est séquestré dans le nucléole de part son interaction avec une autre protéine fortement nucléolaire, la nucléophosmine (NPM1). Cette dernière est principalement impliquée dans le processus de maturation des pré-ARNr, mais possède aussi la capacité à faire la navette entre le nucléole et le noyau cellulaire, où, bien que son rôle exact fasse toujours l'objet de recherches actives, elle régule l'expression, entre autres, de Myc, NFkB et p53^{97,139}.

Outre ces facteurs qui peuvent médier une activation de la signalisation p53-dépendante, un

débalancement au niveau de l'expression des protéines ribosomales affecte aussi la biogénèse des ribosomes. Le terme ribosomopathies fait référence aux pathologies humaines associées à un stress ribosomal^{106,140-142} (Tableau 1.1).

Tableau 1.1: Présentation des principales ribosomopathies humaines.

Maladie	Gènes impliqués	Présentation clinique	Progression vers cancer	Références
Anémie Diamond-Blackfan	RPS7, RPS10, RPS17, RPS19, RPS24, RPS26, RPL5, RPL11, RPL35A	Anémie macrocytique, réticulocytopénie, malformations	MDS, AML, adénocarcinome du colon, ostéosarcome, cancer génital	143-146
Dyskeratosis congenita liée au Chromosome X	DKC1	Hyperpigmentation de la peau, dystrophie des ongles, disfonction de la moëlle osseuse	AML, tumeurs tête et cou	147
Syndrome 5q-Syndrome	RPS14	Anémie macrocytique	AML	148,149
Shwachman–Diamond	SBDS	Insuffisance pancréatique, hémato-poïèse anormale, malformations	MDS, AML	150,151
Hypoplasie cartilage-cheveux	RMRP	Dysplasie squelettique, cheveux fins et rares, dysfonction immunitaire, anémie macrocytique and lymphopénie	Lymphome non-Hodgkinien , carcinome baso-cellulaire	152-154
Syndrome Treacher-Collins	TCOF1	Malformations cranofaciales		155-158

Une des premières maladies humaines associée à une dérégulation d'une protéine ribosomale est l'anémie de Diamond-Blackfan (DBA), pour laquelle une mutation récurrente du gène *Rps19* a été identifiée^{144,159}. Plusieurs exemples de pathologies liées à des mutations chez des gènes codants pour des protéines impliquées dans la biogénèse des ribosomes sont présents dans la littérature: le gène *Tcof1* est impliqué dans le syndrome de Treacher-Collins (TCS)¹⁶⁰, *Rps14* dans le syndrome myélodysplasique 5q-¹⁴⁸, *SBDS* dans le syndrome Swachman-Diamond¹⁵⁰. Ces syndromes sont généralement caractérisés par des anomalies au niveau hématologique et squelettique, tandis qu'au niveau moléculaire l'activation de p53 semble être une signature classique de ces conditions. Il a été démontré que lors d'un stress ribosomal, les protéines ribosomales Rps3, Rps7, Rps14, Rpl5, Rpl11, Rpl23^{97,161,162}, qui ne sont plus associées aux sous-unités ribosomales en maturation, ont la capacité de lier le domaine central de MDM2 et ainsi d'inhiber ses fonctions d'ubiquitination. Cette activation de p53 et des voies de signalisation en aval de cette protéine, potentiellement indépendantes de Arf¹⁶³, semblent être commune aux ribosomopathies, puisque qu'une diminution des niveaux d'expression de p53 parvient à renverser les phénotypes anormaux associés à ces syndromes^{140,164,165} (Figure 1.8).

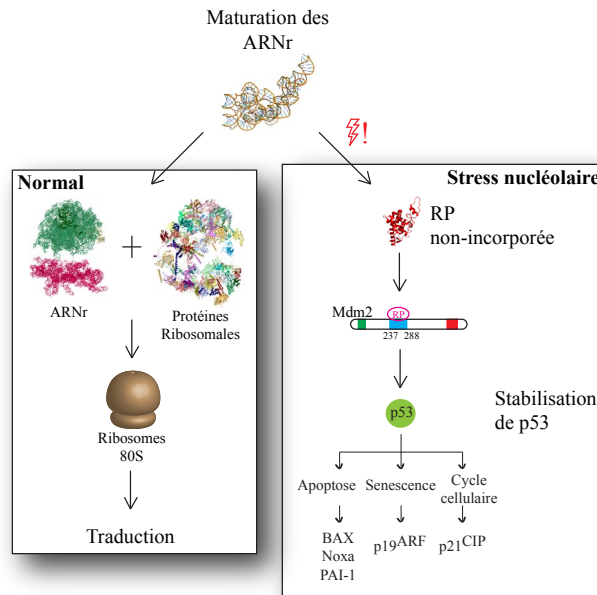


Figure 1.8: Mécanisme du stress ribosomal.

Dans des conditions normales, la maturation des ARNr conduit à la formation de ribosomes fonctionnels pouvant assurer la traduction. En situation de stress ribosomal, certaines protéines ribosomales non-associées aux sous-unités ont la capacité à lier la partie centrale de Mdm2. Les fonctions d'ubiquitination de cette dernière contre p53 ne sont plus assurées, résultant en l'activation de voies de signalisation p53-dépendantes.

Dernièrement, certaines études font mention de mécanismes p53-indépendants suite à une induction du stress ribosomal. Cette littérature toute récente, qui jette un regard nouveau sur les mécanismes moléculaires relatifs aux perturbations nucléolaires, sera abordée plus en détail dans la discussion du présent travail, puisque nos travaux de recherche présentés dans les chapitres qui suivent viennent documenter cet aspect du stress ribosomal.

1.2.3 Activité traductionnelle et différenciation cellulaire

En ayant un impact direct sur la biogénèse des ribosomes, l'induction d'un stress nucléolaire peut ainsi affecter la production de protéines et potentiellement la survie cellulaire. Dans un contexte pro-différenciation, où plusieurs facteurs sont requis afin de soutenir le développement de types cellulaires spécifiques, il a été démontré que la synthèse protéique en générale était plus élevée⁹⁵. Ces données laissaient présager qu'un stress ribosomal pouvait avoir un impact considérable sur le protéome de cellules pluripotentes s'engageant dans une voie de différenciation. Historiquement, et de façon plus évidente ces dernières années avec l'avènement des techniques de séquençage à haut-débit, les recherches dans ce domaine se sont plutôt attardées aux niveaux d'expression des transcrits nécessaires à la mise en place des mécanismes moléculaires liés à la

différenciation plutôt qu'aux modulations traductionnelles. Outre les contraintes technologiques favorisant les études transcriptionnelles au détriment des approches protéomiques, ce phénomène s'explique aussi par des résultats démontrant que l'élongation des transcrits est l'étape limitante chez les ESC humaines⁹⁴ et que la production de protéines était plutôt limitée par l'abondance des ARNm⁹⁵. Récemment, l'adaptation de techniques déjà existantes et le développement de nouvelles méthodes expérimentales ont permis de documenter la régulation de processus de différenciation cellulaire d'un point de vue traductionnel.

1.2.3.1 Analyse expérimentale de l'activité traductionnelle

Dans le but de faire un lien entre les niveaux d'expression génique et la production de protéines, certaines méthodes expérimentales permettant l'évaluation de l'activité traductionnelle ont été développées. La technique du profilage polysomal fût ainsi développée afin de déterminer le rapport entre l'abondance des transcrits et l'activité ribosomale (le nombre de ribosomes par transcrit) (Figure 1.9, panneau inférieur gauche). De façon concrète, les ribosomes en cours de traduction sont fixés aux ARNm, avant que ces derniers ne soient physiquement séparés sur un gradient de sucrose en fonction de leur poids moléculaire, qui lui est directement proportionnel au nombre de ribosomes présents. Ainsi, il est possible de déterminer si un ARNm est plus ou moins traduit suite à une perturbation cellulaire, en comparant les niveaux totaux d'expression génique et la quantité de ribosomes assurant sa traduction¹⁶⁶. Une variante de cette approche récemment développée implique plutôt l'analyse des séquences d'ARNm protégés par les ribosomes, le profilage ribosomal (Figure 1.9, panneau inférieur droit)¹⁶⁷. Le principal avantage de cette technique est qu'elle ne requiert pas la séparation des complexes ribosomes-ARNm selon leur poids moléculaire, réduisant de façon considérable les limitations techniques et la difficulté dans l'interprétation des résultats dû à un manque de résolution lors du fractionnement. Toutefois, l'information relative à la multiplicité des ribosomes sur un même transcrit y est perdue, puisqu'il est impossible de faire la distinction entre une variation du nombre de transcrits et une modulation du nombre de ribosomes par transcrit pour expliquer les différences au niveau de l'expression d'empreintes ribosomales¹⁶⁸.

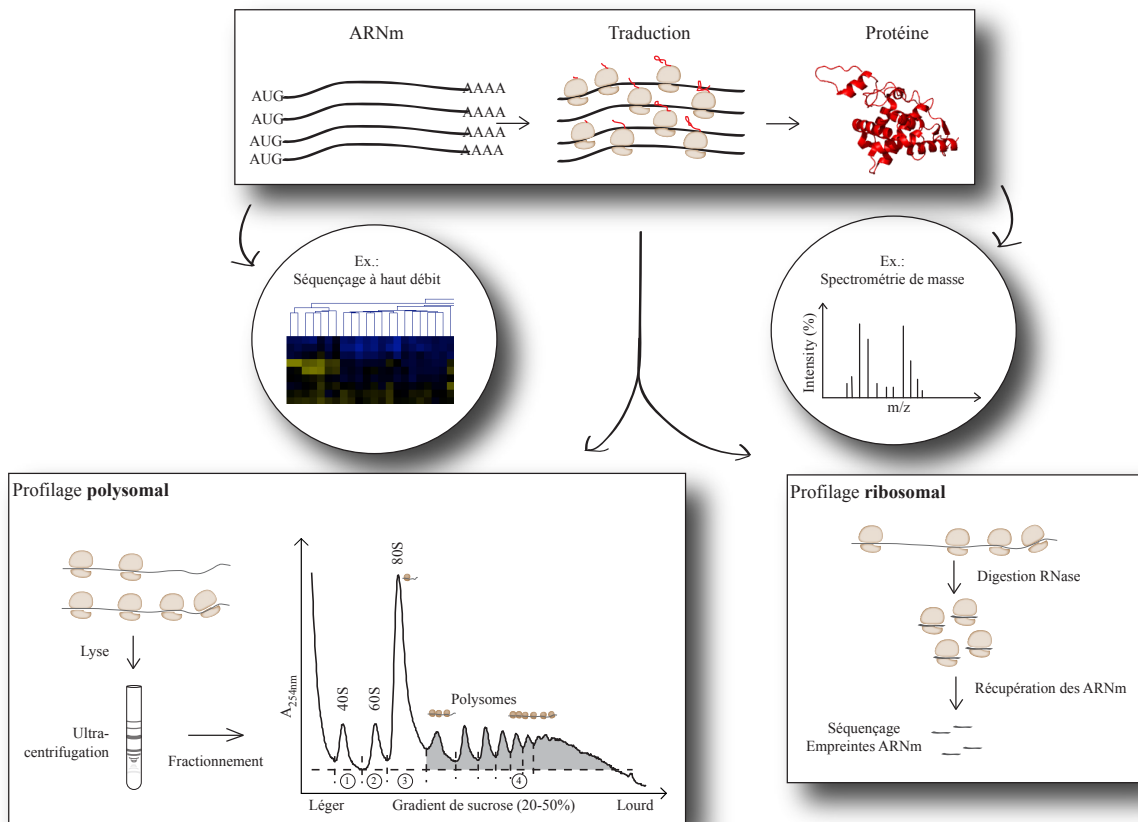


Figure 1.9: Évaluation de l'activité traductionnelle.

À la différence de méthodes conventionnelles s'intéressant aux niveaux d'expression génique (ARNm et séquençage, à titre exemple) et à l'expression protéique (protéines et spectrométrie de masse, à titre d'exemple), les techniques de profilage polysomal ou ribosomal donnent un aperçu de la dynamique traductionnelle.

Grâce au profilage ribosomal, Ingolia et al.¹⁶⁹ ont démontré une diminution de l'ordre de 3 à 4 fois des fragments d'ARNm codants pour des protéines ribosomales protégés par les ribosomes lors de la différenciation de ESC en EB, bien que les niveaux totaux d'ARNm soient similaires. Ces résultats sont intéressants puisqu'il s'agissait d'une première démonstration que la différenciation des ESC ne semblait pas dépendante du nombre de ribosomes présents, suggérant que la traduction d'espèces spécifiques d'ARNm était affectée lors d'une stress ribosomal. Allant dans le même sens, Bywater et al. ont démontré qu'une perturbation nucléolaire induite par l'inhibition de la l'ARN polymérase I n'avait pas d'impact sur le nombre total de ribosomes¹⁷⁰. À la lumière de ces résultats, les études portant sur la régulation de l'activité traductionnelle de même que des analyses protéomiques complémentaires s'avèrent nécessaires pour interpréter de façon adéquate les observations faites au niveau de l'expression génique¹⁶⁸.

1.3 Mise en contexte et objectifs des travaux

L'isolation de lignées de cellules souches embryonnaires au milieu des années 70 fut non seulement d'une importance capitale dans notre compréhension du développement embryonnaire, mais elle contribua au développement d'outils indispensables pour l'étude des gènes impliqués dans le fonctionnement cellulaire. Que ce soit par rapport à des processus de développement normaux ou encore par comparaison à des modèles de maladies, l'utilisation du potentiel à se différencier des ESC fut inestimable pour la recherche en génomique fonctionnelle.

Dans le cas des ESC, le réseau de facteurs de transcription essentiels au maintien de leur pluripotence a été identifié et est étudié depuis déjà de nombreuses années. Toutefois, l'identité de certains joueurs importants pour ces processus d'auto-renouvellement et de différenciation est potentiellement toujours inconnue. C'est pourquoi nous avons entrepris de cribler, en utilisant une méthode récemment établie dans le laboratoire et basée sur l'intégration aléatoire de rétrovirus⁸⁷, le génome de ESC murines dans le but d'identifier de nouveaux déterminants essentiels aux ESC. Lors de mon arrivée au laboratoire, une librairie de clones de ESC comportant des délétions hétérozygotes était en fin de production. Mes objectifs à court et long terme étaient les suivants. Dans un premier temps, caractériser et rapporter la création de cette ressource. Par la suite, confirmer la fonctionnalité des clones et démontrer qu'ils pourraient être utilisés afin d'identifier de nouveaux déterminants régulant leur pluripotence et leur différenciation. Les travaux portant sur ces deux premiers aspects sont présentés au Chapitre 2 de cette thèse. Ensuite, l'objectif prioritaire de mes recherches fut d'établir un protocole de criblage permettant d'évaluer l'impact des délétions chromosomales hétérozygotes sur la capacité des ESC à se différencier et ainsi identifier de nouveaux éléments essentiels à leur fonctionnement. Les résultats de ces travaux, portant sur l'identification des certaines protéines ribosomales et leur rôle dans la pluripotence des ESC sont rapportés dans le Chapitre 3..

CHAPITRE 2: Genome-wide interrogation of mammalian stem cell fate determinants by nested chromosome deletions

Simon Fortier^{1,µ}, Mélanie Bilodeau^{1,µ}, Tara MacRae¹, Jean-Philippe Laverdure¹, Valeria Azcoitia¹, Simon Girard¹, Jalila Chagraoui¹, Nancy Ringuette¹, Josée Hébert^{1,2}, Jana Kros¹, Nadine Mayotte¹ and Guy Sauvageau^{1,2,*}

¹Molecular Genetics of Stem Cells Laboratory, Institute of Research in Immunology and Cancer (IRIC); University of Montreal; Montreal, Quebec, H3C 3J7; Canada.

²Division of Hematology and Leukemia Cell Bank of Quebec (BCLQ); Maisonneuve-Rosemont Hospital; Montreal, Quebec, H1T 2M4; Canada.

*Correspondence: [Guy Sauvageau](#)

^µ Co-first authors

Publié dans la revue *PLoS Genetics*, 2010:

PLoS Genet 6(12): e1001241. doi:10.1371/journal.pgen.1001241

2.1 Contribution des auteurs

Conceived and designed the experiments: SF MB TM JC NR GS.

Performed the experiments: SF MB TM VA SG JC NR JH JK NM GS.

Analyzed the data: SF MB TM JPL VA SG JC NR JH GS.

Wrote the paper: SF MB TM JPL GS.

De façon plus détaillée, ma contribution à cet article se décrit comme suit:

J'ai réalisé toutes les expériences de complémentations. La partie «création de la librairie» venait tout juste d'être complétée à mon arrivée au laboratoire. J'ai réalisé toutes les figures et analysé les résultats contenus dans l'article, incluant ceux d'écrivant la génération de la librairie. J'ai écrit le manuscrit, sous la supervision de Guy Sauvageau, avec la collaboration de Mélanie Bilodeau (sections 2.4.1 et 2.8), Tara MacRae (sections 2.8 et 2.12.6) et Jean-Philippe Laverdure (section 2.4.3). Les autres co-auteurs ont été impliqués au point de vue expérimental lors de la génération des clones de la librairie, que ce soit en support pour la culture des ESC (VA, JK, NR, JC), la production des cellules nourricières et des virus (NM) et la cartographie des délétions (SG, JH).

2.2 Summary

Understanding the function of important DNA elements in mammalian stem cell genomes would be enhanced by the availability of deletion collections in which segmental haploidies are precisely characterized. Using a modified Cre-*loxP* based system, we now report the creation and characterization of a collection of ~1300 independent embryonic stem cell (ESC) clones enriched for nested chromosomal deletions. Mapping experiments indicate that this collection spans over 25% of the mouse genome with good representative coverage of protein-coding genes, regulatory RNAs and other non-coding sequences. This collection of clones was screened for *in vitro* defects in differentiation of ESC into embryoid bodies (EB). Several putative novel haploinsufficient regions, critical for EB development, were identified. Functional characterization of one of these regions, through BAC complementation, identified the ribosomal gene *Rps14* as a novel haploinsufficient determinant of embryoid body formation. This new library of chromosomal deletions in ESC (DeIES: <http://bioinfo.irc.ca/deles>) will serve as a unique resource for elucidation of novel protein-coding and non-coding regulators of ESC activity.

2.3 Introduction

2.3.1 Mammalian genomes and ESC characteristics

Mouse ESCs, derived from the inner cell mass of the blastocyst [1,2], are a lineage of choice to perform functional genomic studies for several reasons. First, ESCs constitute a sustained source of starting material since they indefinitely self-renew symmetrically in defined culture conditions, generating two functionally identical daughter cells per division [3]. Second, pluripotent ESCs enable the study of most developmental processes *in vivo* or *in vitro*, owing to their capacity to make all somatic cell types, including germ cells [4,5]. Third, ESCs can model various aspects of tumorigenesis. Undifferentiated ESCs are characterized by the absence of a robust G₁/S cell cycle checkpoint [6], a feature frequently observed in tumor cells [7]. Moreover, ESCs are tumorigenic when ectopically implanted [1,2]. Lastly, the ESCs genome is easily modifiable with various mutagenesis techniques. Because ESCs and induced pluripotent stem cells (iPS) are valuable resources for modeling human diseases *in vitro* and *in vivo* as well as a potential source for cell replacement therapy, major efforts are ongoing to decipher the molecular determinants regulating the cardinal features pertaining to these cells, such as self-renewal, pluripotency, multilineage differentiation and tumorigenic potential.

ESCs are capable of being maintained undifferentiated *in vitro* in the presence of LIF and BMP signaling [8]. Upon removal of self-renewal signals (e.g. LIF), ESCs will differentiate *in vitro* into aggregated structures called “embryoid bodies” or “EB”. ESC differentiation into EB occurs in an ordered manner, with the generation of derivatives from the 3 germ layers [9]. This feature of *in vitro* ESC differentiation seems to recapitulate, in a spatiotemporal manner, several of the differentiation processes observed *in vivo* (i.e., normal embryonic development [10]). Moreover, ESC differentiation into endoderm, mesoderm, and ectoderm is highly regulated and correlates with expression of a panel of specific markers, which can be used to characterize the extent of the differentiation process at the molecular level [11].

2.3.2 Networks regulating ESC fate

Although several proteins involved in signaling, transcriptional regulation and chromatin modification are implicated in ESC activity, we still do not understand all genetic hierarchies dictating ESC fate [12-15]. Recent studies have also documented a function for non-coding RNAs such as microRNAs and lincRNAs in ESC behavior [16,17].

Aside from these large classes of determinants, sequence comparison analyses suggest that

other elements of the mammalian genome might be regulating biological functions, including ESC behavior. Among these elements are 480 segments of >200 bp termed “ultraconserved elements”, characterized by 100% sequence conservation (higher degree of conservation than protein-coding regions) between human, mouse, and rat genome [18]. Of these “ultraconserved” elements, more than 50% show no evidence of transcription, while others overlap with protein coding genes [18]. These sequences are enriched for homeodomain-binding modules, which is intriguing considering the important role of homeodomain transcription factors in ESC pluripotency and developmental processes [19]. Finally, although evolutionary conserved sequences may pinpoint functionally important genomic regions, other crucial elements may lack evolutionary constraints [20].

2.3.3 Functional genomics in mammalian stem cells

Several large-scale functional genomics initiatives are currently ongoing to understand the molecular bases of embryonic stem cells. These include single gene inactivation (or alleles generation) using diverse strategies such as chemically-induced point mutations [21], gene/exon trapping (e.g., the international gene trap consortium: www.genetrap.org) and homologous recombination (The comprehensive knockout mouse project consortium: [22]). A repository for KOMP now exists (www.komp.org) in which 8500 genes are being targeted (several in conditional alleles) within relatively short periods of time. This repository contains several available lines from other initiatives. As a result, in mouse, most protein-coding genes will be deleted and available, many of them as conditional alleles, within the coming years. While these collections represent an outstanding resource for the community, they nonetheless leave a significant proportion of the “functional genome” unexplored. Moreover they fail to examine synthetic interaction between gene neighbours.

To complement existing resources that explore functional elements in the mammalian stem cell genomes, we have applied our recently developed retroviral tools to create a collection of ESC with nested chromosomal deletions [23]. Here we document the generation of DeLES (Deletion in ES cells) library, which contains more than a thousand independent ESC clones highly enriched in chromosomal deletions and representing a large coverage of the mouse genome. Evidence is provided to demonstrate that a large proportion of these clones are competent in functional assays. A complementary method was also optimized to introduce, at high efficiency, a series of selectable marker genes in the backbone of BACs in order to rescue the inability of selected ESC clones to form embryoid bodies *in vitro*. This first validation allowed the identification of a novel gene essential for EB formation. In addition, a database was created (<http://bioinfo.irc.ca/>

[deles](#)) to assist in sample management and to compile and interpret all genetic and phenotypic data related to the collection of clones. This database will facilitate the search for genomic regions regulating the ESC activity and the further design of rescue experiments. The library of ESC clones described herein thus has the added potential of deciphering novel determinants involved in ESC activity. On that basis, it is highly complementary to other international functional genomics initiatives.

2.4 Results

2.4.1 Generation and molecular characterization of DeIES resource

In order to generate a library of ESC clones containing nested chromosomal deletions (DeIES library), we used a retroviral-based method that exploits Cre-*LoxP* technology as described [23] (summarized in Figure 2.1A). Assisted by robotic cell culture manipulation, we upscaled the previously described procedure to generate the DeIES collection (Figure 2.1B, see also Section 2.12 and Tableau S2.1). Statistics about the various groups of clones in DeIES (primary, secondary and tertiary clones) and the types of chromosome rearrangements (e.g., nested chromosomal deletions) are detailed in Figure 2.1C. A total of 4929 G418^R tertiary clones (i.e., ESC clones harboring recombination events are selected with geneticin) originating from 156 anchor sites (i.e., families) were isolated (Figure 2.1C and Tableau S2.1). Of these, 33.8% (n= 1670) were sensitive to puromycin (puro^S) of which 78.3% (n=1307) were cryopreserved in 96 well plates. Previous work has shown an expected 80% chromosomal deletion rate in puro^S clones [23].

When further characterized for proviral integration patterns by Southern blot analyses, we found that these 1307 independent clones harbored in fact 512 distinct chromosomal rearrangements (referred to as sub-families, not shown). Moreover, each family, characterized by a common anchor site, presented an average of 5.39 (range 1 to 20) distinct chromosomal rearrangements (Tableau S2.2). So far 423 deletions of which 294 are unique have been mapped by inverse or ligation mediated PCR (Figure 2.2A), representing 25.4% of the mouse genome (Figure 2.2B). The genomic coverage varies by chromosome, with no identified deletions on chromosomes 19, X or Y; limited coverage of chromosomes 8 and 13 (8.7% and 4.2%, respectively); and approximately 50% coverage of chromosomes 6 and 18 (Figure 2.2B). On average, there is approximately 23% genome coverage per autosome. Deletion sizes range from 736 bp to 100.79 Mb, with a median of 1.61 Mb (average size: 4.95 Mb) (Figure 2.2C), and vary according to the chromosome (Tableau S2.3). Chromosomes 1, 8 and 16 are characterized by many small deletions, while chromosomes 18 and 14 have a few large deletions (DeIES database, <http://bioinfo.iric.ca/deles>).

As depicted in Figure 2.1A and detailed earlier [23], deletions are typically characterized by G418R clones which have lost the hygromycin and the puromycin genes. Interestingly, we found 29 families in which none of the G418R clones had lost hygromycin and/or puromycin resistance genes. One possibility that could explain this observation is that the anchor virus may have integrated in the vicinity of haplolethal loci. The Table S4 provides a list of genes present

in the vicinity of 9 independent anchor loci that are not associated with chromosomal deletions (e.g. 9 families which had a minimum of 15 G418R tertiary clones but no puroS hygromycin clones). A literature search was performed to identify candidate genes in these regions which are known or predicted to be haploinsufficient or imprinted (in red). On average, ~1 haploinsufficient/imprinted candidate gene was identified per 1.61 Mbp window (median deletion size) starting from each directional anchor site (Tableau S2.4). These candidate genes, alone or in combination with other genes or non-coding elements within these regions, could potentially regulate essential cellular functions or specific characteristics of ESCs and thus cannot be maintained in a heterozygous state.

Taken together, these results reveal that close to 300 independent deletions exhibiting a genome-wide distribution have been confirmed in the DeIES collection.

2.4.2 Genomic coverage of DeIES

To evaluate the content of DeIES genomic coverage, genes included in currently mapped deletions were classified according to their gene ontology (GO) terms. Gene ontology analysis of molecular functions of the 7083 mapped deleted genes revealed similar percentages in each category to that obtained for all annotated MGI genes (with known functions) (data not shown). When genes were grouped by molecular function, the most abundant group was genes with signal transduction activity, followed by transcriptional regulation and nucleic acid binding activities (data not shown).

Distribution of some key genomic elements covered by mapped deletions, such as protein-coding genes, CpG islands, miRNA, ultraconserved elements, lincRNA, LINE/SINE elements, cancer-related genes and large deletions associated with cancers was evaluated (Figure 2.3, Tableau S2.5). For this analysis, elements found in all of DeIES's mapped deletions were compared to publicly available genome-wide entries for each category. Percentages represent ratios between the number of observed elements (found in DeIES mapped deletions) and the number of reported entries (assuming random distribution of elements), based on the current genome coverage of DeIES (25.4%). Interestingly, mapped deletions cover close to 100% of each category: genes (7083/7348), CpG islands (4265/3515), miRNA (128/139), lincRNA (470/540), ultraconserved elements (241/277), LINE/SINE (648571/602886), cancer related genes (108/104) and large cancer-associate deletions (5/7). Thus, several categories of protein-coding and non-coding sequences are well represented in deletions that are currently mapped. Moreover, clustering of specific genomic determinants has been reported [18,24-26]. As expected, several clusters

of protein coding and non-coding elements were deleted in DeES clones (highlighted in red, Tableau S2.5). Deletions of entire clusters represent another strength of DeES, as it presents the opportunity to analyze synthetic interactions between family members and to study possible functional redundancies.

2.4.3 Organization of DeES collection in a public database

In order to facilitate access to the DeES collection and all clone-specific information, a database accessible through a web interface offering data mining tools was constructed (Figure S2.2, <http://bioinfo.irc.ca/deles>). A detailed explanation of the content of the interface can be found in Figure S2.2 and in Section 2.12.

Taken as a whole, the DeES database allows for the management of biological material (*Plate* tab) and facilitates the search for ESC clones through phenotypic or genetic annotations (*Selection* tab). The results of the search are directly linked to the complete data sets (*Families* tab, Figure S2.2). Phenotypic information can rapidly be associated to a graphical representation (*Screen* tab) of the mapped deletion including the implicated genomic features and BACs available for complementation studies (e.g., family 9).

2.4.4 Quality control of DeES clones

Primary and tertiary clones in the DeES collection are distributed and frozen in 96 well plates. Localization of each clone within plates can be directly visualized online under the plate collection tab (<http://bioinfo.irc.ca/deles>). Colored well images indicate the presence of suspected undesired chromosomal anomalies and rearrangements (e.g., detection of hygromycin gene). Master plates containing DeES collection were thawed once to rule out microbial contamination and to determine the proportion of clones which proliferate normally (e.g. high proportion of Ki67⁺ cells) or those which express high levels of alkaline phosphatase activity, typically associated with undifferentiated ESC.

Figure 2.4A shows that nine percent of the puro^S hygro^r tertiary clones showed low alkaline phosphatase activity (scores <3; families with clusters of clones with low AP activity are identified as r in Figure 2.4), suggesting that their pluripotent capabilities were impaired. Specific genomic features covered by the nested deletions in these cells may be responsible for maintaining the pluripotent state of ESCs. This possibility will be investigated separately as part of a screen which will include multilineage differentiation assays and additional markers of pluripotency, such as Oct4.

Figure 2.4B shows that 14% of puro^S hygro^r tertiary clones presented low levels (< 60%) of Ki67⁺ cells. Clusters of clones presenting altered proliferation status (Ki67⁺ < 60%) were observed in 17 families (identified as ψ , Figure 2.4). The quantification of Ki67 expression was highly correlated with observed cell proliferation rate measured by flow cytometry using calibrated beads (data not shown) and batch collection of clones based on cell density or expansion (i.e. clones collected in first batch “A” expand faster than those in last batch “D” which were the slowest to expand; see Figure S2.1 and Section 2.12). This observation was also validated by a cell density assay which estimated ESC colony number one day after plating (data not shown). Unfortunately, clones with very low proportion of Ki67⁺ cells are easily lost upon freeze thaw procedures and are difficult to maintain in the collection (S.F., personal observation).

Overall, the vast majority of clones in the DeES collection express high levels of alkaline phosphatase; they proliferate normally; they support freeze-thaw procedures and they appear free of microbial contaminant. Clones can be recovered individually (all frozen in 96 well plates and individually) from several freeze thaw cycles. This suggests that DeES clones might be amenable to different functional screening procedures and that they can be validated separately. However, clones with low Ki67 activity are difficult to maintain.

2.4.5 Functional screen of DeES: in vitro EB formation

Using control (parental) R1 ESCs, we observed a strong correlation between the number of EB generated in culture and the number of ESCs plated (Figure 2.5A). This value was reliable when cell density was above 5% (S.F., Figure 2.5A and data not shown). This observation was exploited to develop a functional screen in which each clone from DeES was individually seeded in two 96 well plates, one with LIF for estimation of ESC colony numbers (seeding density in Figure 2.5A) and the other without LIF for EB differentiation (Figure 2.5B), aiming to identify minimal deleted regions that cause a block in normal EB development. Three criteria were used to identify clones and families with EB formation anomalies: 1) clones were only considered if cell density was above 5% (45% of clones); 2) EB formation was considered abnormal in a tertiary clone if EB number was below one fifth of that detected in the corresponding primary clone (16.4% of clones); 3) families with EB phenotype were selected only if all clones with deletions exceeding that of the minimal deleted region also show the phenotype (see Figure 2.5C for selected vs rejected families). The high percentage of clones that were eliminated based on the first criteria reflects the wide distribution of cells (e.g., less than 1% to over 10%) recovered after freeze-thaw process (see also Methods for methylene blue staining). Using criteria described above, 15.6% of the families (n=14) were considered as potentially interesting for the future identification of

EB formation determinants (Figure 2.5D and <http://bioinfo.irc.ca/deles> for details of each clones in the selected families. See also Tableau S2.6 for primary screen data).

Several tertiary clones from five randomly selected families (9*, 5061*, 5035, 5214, 5238; *=families with phenotype) were tested for *in vitro* EB formation using standard assays in 60 mm dish in duplicate experiments. Corresponding primary clones were included in these validation experiments. These studies showed a concordance rate of 78% (28 out of 36 tested) between tertiary clones tested in validation experiments and the results obtained in the primary screen. In total, 3 of the 5 tested families were validated including two (9, 5061) which show putative phenotype associated with a minimal deletion region, located on chromosome 11 and 18 for family 5061 and 9, respectively (see <http://bioinfo.irc.ca/deles>) From this assay, it thus appears that the primary screen underestimates the frequency of families which include clones with EB differentiation phenotype. Consistent with this, our 15.6% hit rate is below that previously observed in our pilot studies of nine families (33% of families showed EB differentiation phenotype [23]).

2.4.6 Complementation studies of DeES family 9

DeES family 9 was chosen as the prototype for complementation studies since clones of this family harboring large deletions are unable to differentiate into embryoid bodies. Of importance for the complementation studies described below, the frequency of EB formation in clones containing large deletions was lower than 1 in 5000 (i.e. 1 EB for 5000 cell plated). The minimal region responsible for the abnormal phenotype (e.g., red line in Figure 2.6A) was mapped between the breakpoints of tertiary clones 9-35 (736 bp deletion, normal *in vitro* differentiation) and 9-37 (4.3 Mbp deletion) (Figure 2.6A). This minimal deleted region contains 30 known protein-coding genes and can be covered by 20 contiguous bacterial artificial chromosomes BACs (Figure 2.6A). These 20 BACs were modified for selection in ESC using a strategy that we adapted from existing recombineering systems to allow for the introduction of selectable marker gene into the chloramphenicol resistance sequence present in the backbone of the different BACs [27]. Details about this method, called SelectaBAC, are provided in Section 2.12 and in Figure S2.3.

Two independent tertiary clones with EB formation phenotype, 9-37 and 9-18, were transfected with each of the 20 BAC constructs separately and assayed for embryoid body formation. Interestingly, none of the BAC but one -RP23-143E19- led to a complete rescue of the differentiation defect of clone 9-37, and partial rescue of clone 9-18 which contains a larger

deletion (Figure 2.6B).

To validate the presence of transfected BAC in complemented ESCs, metaphase fluorescence in-situ hybridization (FISH) was conducted using two differentially labeled probes: the first being RP23-143E19 itself and the second a control BAC which maps adjacent to the deleted region (Figure 2.6C, *lower-left panel*). Transfected clones were compared with untransfected tertiary controls and R1 ESCs. As expected, normal R1 ESCs had two pairs of closely localized signals, corresponding to the intact mitotic chromosomes 18 (Figure 2.6C). Haploid deletions were confirmed in tertiary clones 9-37 and 9-18, which had only one pair of RP23-143E19 signals, closely-localized to one of the two pairs of control BAC signals. When clones 9-37 and 9-18 were transfected with the BAC of interest, 86% and 69 % of cells counted displayed a pattern consistent with stable integration of the BAC (Figure 2.6C, *upper and lower-right panels*). This pattern corresponds to two pairs of RP23-143E19 signals, one on chromosome 18 identified by the control BAC signals and another on a different chromosome (not identified by the control BAC signals). Twelve percent of transfected 9-37 clones had larger and more intense red signals (RP23-143E19), which potentially indicates multiple integrations within the same chromosomal region (data not shown).

Control primary 9 clone, tertiary clones 9-18 and 9-37 and BAC-complemented tertiary clones 9-18 and 9-37 were injected separately into blastocysts or aggregated with CD1 morulas to evaluate their contribution to developing embryo. Mouse embryos, at E9.5 and E14.5 were analyzed for the presence of the neomycin gene (A1 provirus or A1-S1 recombined proviruses) by PCR, whereas the level of chimerism in newborns was estimated by coat color variation (Figure 2.6D). As previously reported for the clone 9-18 [23], the unmodified tertiary ESC clone 9-37 also failed to contribute to tissue chimerism in early embryos or newborn mice (Figure 2.6D). In contrast, primary clone 9, used as a positive control, contributed to tissue chimerism of 17 out of 55 mice analyzed (Figure 2.6D). RP23-143E19-transfected clone 9-37 also contributed to embryogenesis with tissue chimerism in 36% and 18% of E14.5 and E9.5 embryos, respectively, and in 4 of 12 newborn mice (Figure 2.6D). RP23-143E19 transfected clone 9-18 also produced chimeric embryos with a frequency of 50% at E14.5. Thus far, all chimeras (embryos and newborn) appear phenotypically normal.

Confirmation that BAC-transfected ESCs contributed to the chimeric embryos was obtained using Southern blot analyses performed with gDNA extracted from fetal liver cells (Figure 2.6E). For newborn and adult mice, percentage of tissue chimerism was estimated at 80-95% and 10-35% for derivatives of primary clone 9 and BAC-complemented clone 9-37, respectively

(Figure 2.6F).

2.4.7 Candidate gene evaluation for DeES family 9

A more detailed analysis of BAC RP23-143E19 reveals the presence of four protein-coding genes: *Ndst1*, *Tcof1*, *Rps14* and *Cd74* (Figure 2.7A). Q-RT-PCR analyses were performed to assess the expression level of these genes in family 9 ESCs and EBs, with or without BAC RP23-143E19 complementation. All four genes analyzed are expressed in both control ESCs and EBs (primary 9) with delta CT values ranging between 0.5 (*Rps14*, highest expression) to 14.4 (*Cd74*, lowest expression) (data not shown). Expression levels of all 4 genes in tertiary clone ESCs were about half that observed in primary clone 9 (compare red bars, tertiary clones to black bars, primary clone in Figure 2.7B). Upon BAC transfection, expression levels of all four genes became either comparable to -or exceeded- that found in the primary clone (compare blue with black bars for undifferentiated ESCs and pink with green bars for EBs in Figure 2.7B).

To gain insight on the contribution of selected elements present on BAC RP23-143E19 to the observed phenotype, seven distinct deletions were generated (Figure 2.7A). These included deletion of all 4 protein-coding genes separately, the intergenic region between *Ndst1* and *Rps14*, a distal promoter to *Ndst1* and a lincRNA close to *Rps14*. Six out of the 7 constructs complemented the EB formation defect observed in clone 9-37 to levels comparable to control primary 9 clone. These 6 constructs complemented clone 9-18 to levels equivalent to those detected with the unmodified BAC (data not shown). Interestingly, the BAC containing the small deletion (3.89Kb) corresponding to the *Rps14* gene did not rescue the phenotype observed in clone 9-37 (Figure 2.7C) and clone 9-18 (data not shown), indicating that this genomic region is haploinsufficient for EB formation.

Following this observation, *Rps14* cDNA expression vector was introduced in 9-37 and 9-18 ESC clones co-transfected with the BAC RP23-143E19 construct no. 5 (lacking the *Rps14* gene, Figure 2.7A), to verify the possibility that a hidden genetic element located within this small region that includes *Rps14* was responsible for the EB formation phenotype. Results from these experiments indicated that 3 out of 8 clones isolated from 9-37 doubly transfected cells showed full complementation and 1 out of 4 clones from 9-18 cells was partially rescued (Figure 2.7D). Expression analyses of these complemented clones revealed that all 4 protein-coding genes (*Ndst1*, *Tcof1*, *Cd74* and *Rps14*) were expressed at endogenous levels when compared to the primary clone (data not shown). These results thus strongly suggest that *Rps14* is haploinsufficient for EB formation.

We then transfected *Rps14* cDNA alone in 9-18 (data not shown) and 9-37 tertiary clones to test if this gene is the sole element responsible for the abnormal phenotype. Interestingly, analyses of several transfected clones showed no complementation with *Rps14* cDNA (Figure 2.7E), raising the possibility that another genetic element is necessary for complementation of DeIES family 9. Nevertheless, these experiments show that *Rps14* is not sufficient, but required, to complement the EB formation defect found in DeIES family 9.

2.5 Discussion

The DeES library described herein contains a large collection of independent ESC clones harboring chromosomal deletions generated by a retroviral-based Cre-*loxP* system. Clonal analysis revealed the presence of several independent deletions in this library of which a significant proportion (close to 300) has been mapped to various locations on mouse autosomes. With a median deletion size of 1.61 megabase pairs, deletions cover more than 25% of the haploid murine genome. Overall, they include 7083 genes, of which one is reported to be haploinsufficient (*Pml*), seven are predicted to be imprinted and 108 are associated with cancer. In addition, 128 miRNAs, 241 ultraconserved elements, 470 lincRNAs and 648571 SINE/LINE elements are covered by the deleted regions (Figure 2.3, Tableau S2.5). The first screening test for this library consisted in analyzing EB formation upon LIF deprivation. Fourteen families exhibited an abnormal phenotype. Pertinent information of each clone in the collection is included in a public database which was created to centralize information of this expanding resource. This paper also includes the description of an optimized BAC engineering strategy which greatly facilitates complementation studies. This approach was thoroughly validated by the successful rescue of the *in vitro* differentiation defect of one family of clones, identifying *Rps14* as a novel haploinsufficient gene in EB formation from undifferentiated ESC.

To date, few genes, when disrupted, have generated overt *in vitro* differentiation defects in ESCs. For example, ESCs that are homozygous null for *smad4* will generate much fewer cystic embryoid bodies *in vitro* [28]. B-catenin activity also appears to play a critical role in ESC differentiation since *Apc* homozygous null ESCs fail to form neuroectodermal derivatives, although they form (visceral and parietal) endoderm [29]. EB formation from *ext1*^{-/-} (a tumor suppressor gene coding for an enzyme required for the biosynthesis of heparin sulfate) ESCs is also defective with a total absence of cavity formation, similar to the phenotype observed with *smad4* mutant cells [30]. Surprisingly, all of these genes are known for their tumor-suppressive potential in various human cancers: *smad4*, *apc* and *ext1* in pancreatic, colon and bone cancers, respectively. Although any generalization from these findings would be premature, it is tempting to speculate that additional tumor suppressor genes will be uncovered from functional screens for genes that interfere with EB formation.

Recent studies would indicate that the ribosomal protein (RP) coding gene *Rps14* identified in this study may also perform tumor suppressor functions in a preleukemia syndrome in humans [31,32]. Indeed, this gene has been implicated in the human 5q- syndrome (now called myelodysplastic syndrome with isolated del(5q)), characterized by anemia and other

cytopenias, as well as other ribosomal protein coding genes that have been linked to several blood differentiation disorders in human [32]. A p53-dependent mechanism was identified lately has being responsible of a mouse model of the 5q- syndrome. This model includes a deletion of the *Rps14* gene [33].

Ribosomes are made of large (60S) and small (40S) subunits each preassembled in the nucleolus. The large subunit is made of 3 RNA species (5S, 5.8S and 28S) and about 49 different RPs. The small subunit is made of a single 18S RNA molecule and approximately 33 RPs [34]. Recent studies suggest that a common mechanism involving defective ribosome biogenesis (also called “nucleolar stress”) may underlie the cell differentiation defect found in several human blood cell differentiation disorders, now collectively called “ribosomopathies”[35]. It was recently shown that a RPS6 partial deficiency triggers a p53 response through RPL11 interaction with MDM2, providing insights on molecular effectors of the “ribosomal stress” response [36]. Based on our findings with *Rps14*, future experiments will aim to better characterize this RP-MDM2-P53 axis in ES cells differentiation and to identify other genetic elements that contribute to our EB formation phenotype.

The assessment of haploinsufficient synthetic interactions between several contiguous genomic determinants, transcribed or not, is a feature that distinguishes chromosomal deletion engineering from other genome-wide reverse genetic approaches. Distinctively based on a precisely localizable pair of retroviral vectors, DeIES is a resource complementary to other large scale methodologies that allow the creation of nested chromosomal deletions, such as: MICER [37] which is based on *loxP*-containing insertional targeting vectors, or DelBank [38] which is based on irradiation-induced deletions of integrated cassettes.

DeIES clones offer many possibilities for reverse genetics since they can be used to conduct *in vitro* differentiation assays, to generate mutant animal models (chimeric, heterozygous, and homozygous animals) [39] and potentially to create teratoma/teratocarcinoma following heterotopic grafts into syngenic mice. Moreover, engineered ESCs could be used to study dominant and recessive (*in vivo*) chromosomal deletions associated with human conditions with complex phenotypic profiles that cannot be modeled using gene-specific mutagenesis approaches, such as certain cancers (4 syntenic regions covered by DeIES clones). For this purpose, the correlation between chromosomal deletion size and germ line transmission rate will need further investigation.

2.6 Conclusion

In conclusion, DeIES is a new resource that offers a library of ESC deletion clones, a BAC complementation system (SelectaBAC) and a comprehensive database. DeIES benefits from precisely localizable *loxP*-containing retroviral vectors which accelerate the generation of segmental haploidy, and is complementary to other functional genomics resources. Its usefulness for uncovering ESC fate determinants was demonstrated herein with the identification of *Rps14* as a novel haploinsufficient gene for EB formation and early embryonic development. DeIES is designed to assess the roles of adjacent coding and non-coding sequences in the mammalian genome, as well as their genetic interactions. Our current efforts are to extend the coverage of mapped deletions in DeIES clones and to conduct additional functional screens, such as cell cycle analysis, pluripotency assessment and hematopoietic differentiation, to enrich our publicly available resource.

2.7 Figures

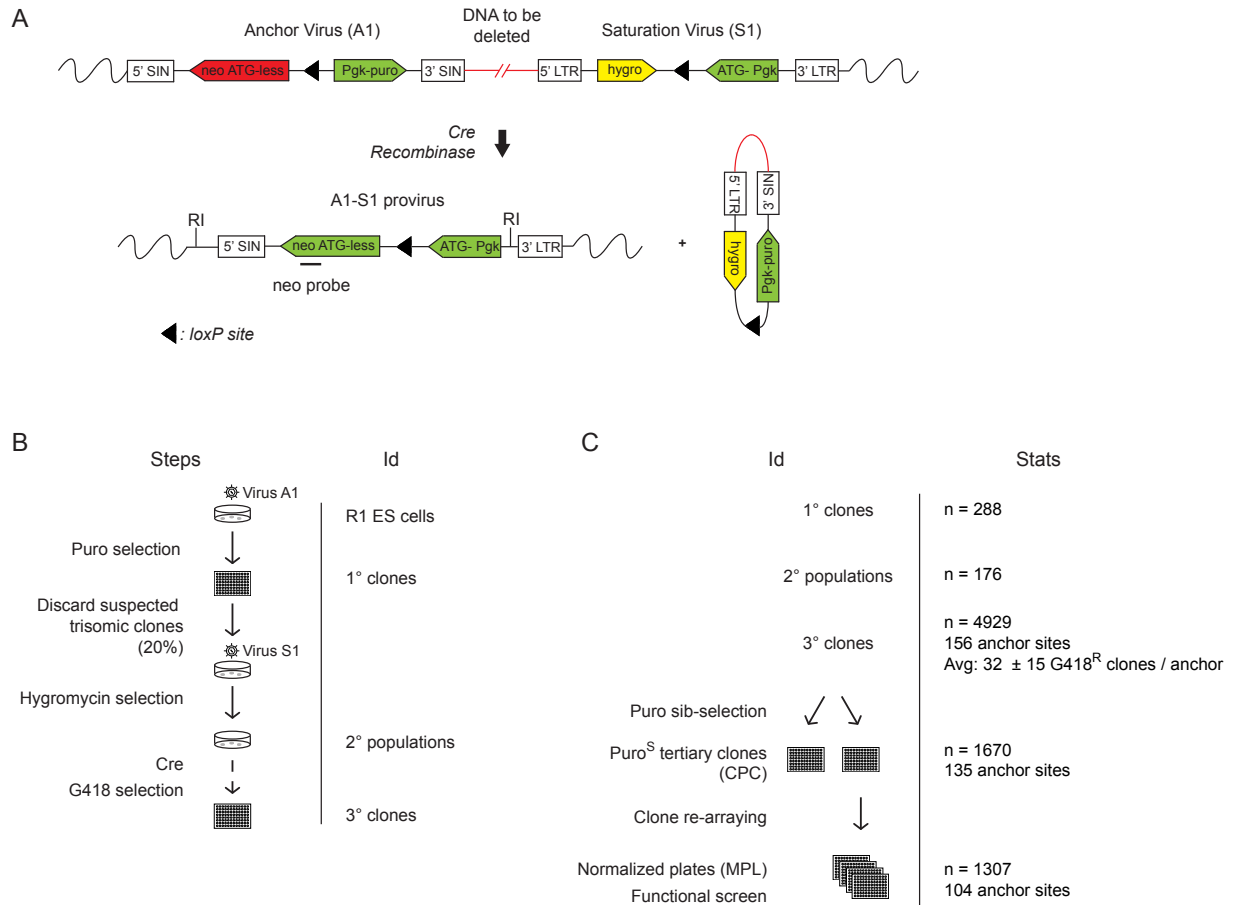


Figure 2.1: Creation of DeES resource.

(A) The engineering methodology is comprised of two compatible retroviruses each containing a loxP sequence. The anchor virus (A1) includes a functional puromycin (puro) gene and a truncated neomycin (neo ATG-less) gene. The saturation virus (S1) includes a hygromycin (hygro) gene and a Pkg-ATG (murine phosphoglycerate kinase) promoter cassette that, following Cre recombination, drives expression of a functional neomycin gene. ESC clones harboring recombination events are selected with geneticin (G418). Chromosomal deletions lead to the concomitant loss of the puromycin and hygromycin genes. LTR, long terminal repeat; SIN, long terminal repeat containing a deletion in the U3 region. (B) Three steps were followed to generate the ESC clones in DeES: i) isolation of puromycin-resistant (puro^R) clones containing a single A1 virus integration, referred to as primary clones; ii) infection of individual primary clones with a low titer S1 virus to generate puro^R and hygromycin resistant (hygro^R) polyclonal populations of secondary clones; iii) transient Cre expression and isolation of G418-resistant

(G418R) recombinant tertiary clones. The group of tertiary clones originating from a specific primary clone and secondary clone population is referred to as a family. (C) Summary of statistics relative to cryopreserved clones. CPC, Cherry-picked clones; MPL, master plate.

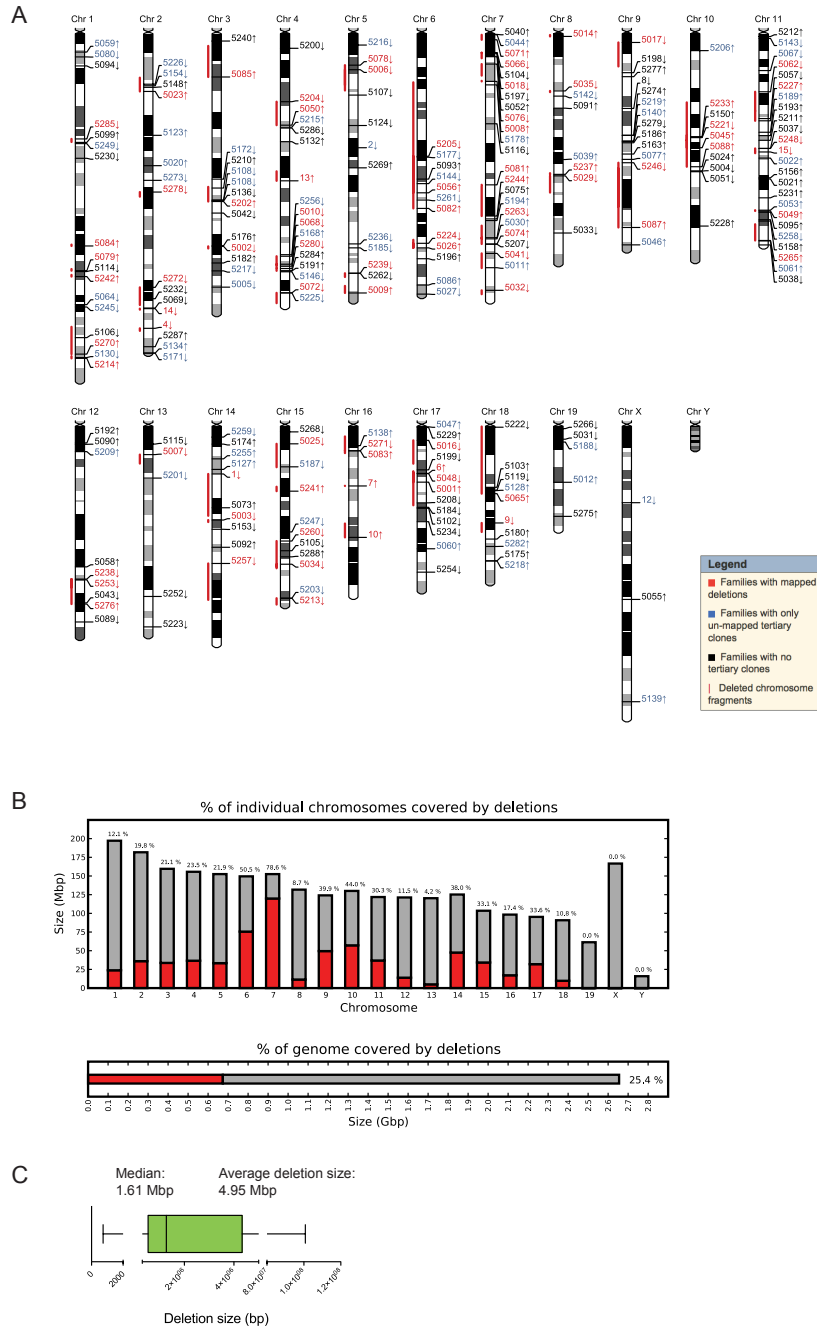


Figure 2.2: Molecular characterization of DeIES clones.

(A) Karyoview representation of anchor virus integration site and coverage of the longest deletion (red bar) mapped for each family of clones. The arrow indicates the predicted deletion orientation relative to the specific anchor site. (B) Genome-wide coverage of deletions mapped in DeIES clones. Haploid deletions are mapped to most autosomes, covering more than 25% of the murine genome. (C) Box plot representing the size distribution associated to 95% of mapped deletions. Average and median deletion sizes are indicated.

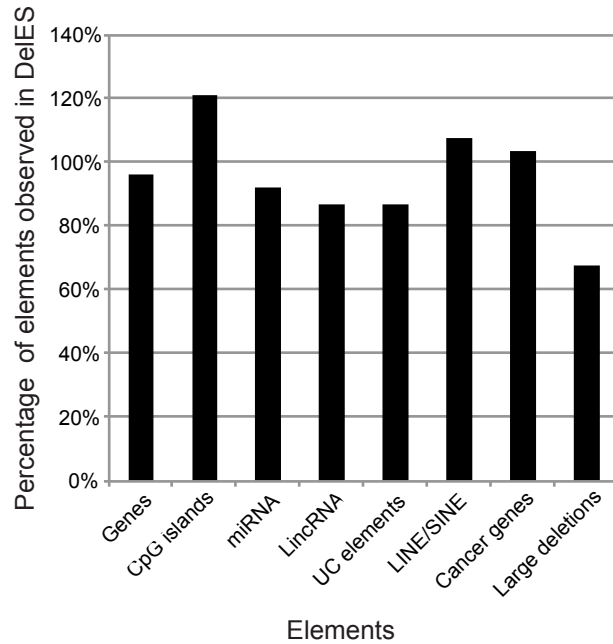


Figure 2.3: Genomic coverage of DeIES.

According to a hypothetical random distribution of elements throughout the murine genome, the indicated features mapped in DeIES deletions were proportionally represented. Numbers used for the calculation were extracted from the following sources: MGI [49] for genes, miRBase13 [43] for miRNAs, Ensembl for CpG islands and LINE/SINE [44], the Wellcome Trust Sanger Institute website [50] for cancer related genes and large deletions linked to cancer, and a previous report [18] for ultraconserved (UC) elements.

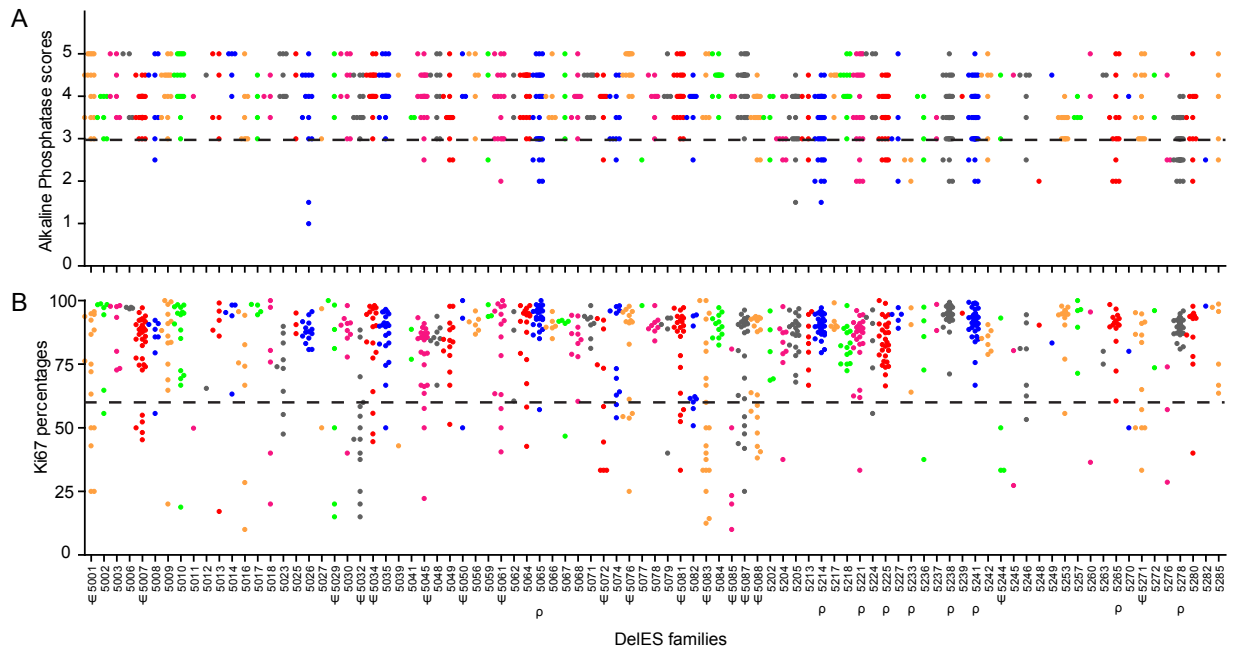


Figure 2.4: Quality control of DeIES clones.

1080 puroS hygro- tertiary clones, grouped in 89 independent families, were analyzed for alkaline phosphatase (AP) staining (self-renewal/pluripotency) and expression of Ki67 (proliferation). (A) AP scores, ranging from 1 to 5 (1 = weak staining; 5 = strong staining), represent the mean values of one experiment performed with two independent cell densities. Approximately 9% of clones (9 families with more than two clones) demonstrated an abnormal phenotype for this assay (dashed bar represents the abnormal phenotype cutoff, arbitrarily set at 3). (B) Ki67 values are expressed as the percentage of Ki67-positive cells, gated on viable cell populations by flow cytometry. Approximately 14% of the clones (17 families with more than 2 clones) presented a proliferation defect (dashed bar represents the abnormal proliferation cutoff, stringently set at 60% Ki67-expressing cells). Families with at least 3 clones harboring an abnormal phenotype are marked by the following symbols: ρ or ψ, for AP or Ki67 category, respectively.

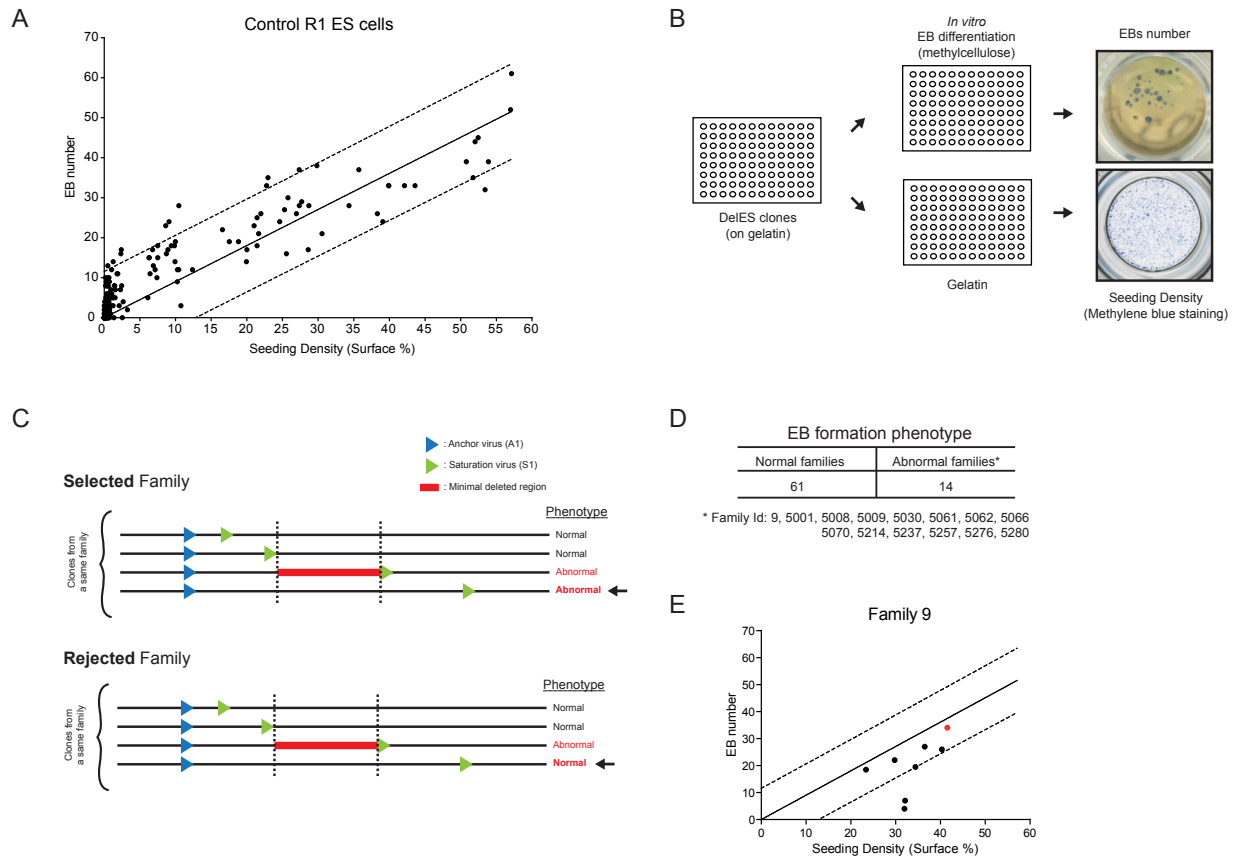


Figure 2.5: Embryoid body formation screen.

(A) Correlation of R1 control ESC seeding density versus embryoid body formation. Linear regression curve from control cells is shown \pm 1 SD. (B) Screen setup. Normalized master plates were thawed on irradiated feeder cells (MEF), before being seeded on gelatinized plates prior to differentiation. After trypsinization, cells were plated in parallel, at equal volumes, in a gelatinized plate and in a 96 well plate containing semi-solid differentiation media. Cells in gelatinized plates were stained 24 hours after seeding with methylene blue (seeding density) and EB number was assessed 8 days after seeding. (C) Phenotype identification of selected vs rejected families using criteria described in text. (D) Phenotype identification in DelES families. (E) Distribution of family 9 clones in the screen setup described in B. Linear regression curve from control cells is shown \pm 1 SD. Outliers correspond to abnormal clones 9-18 and 9-37 which are used in complementation studies (Figures 2.6-2.7). Red dot corresponds to primary clone 9.

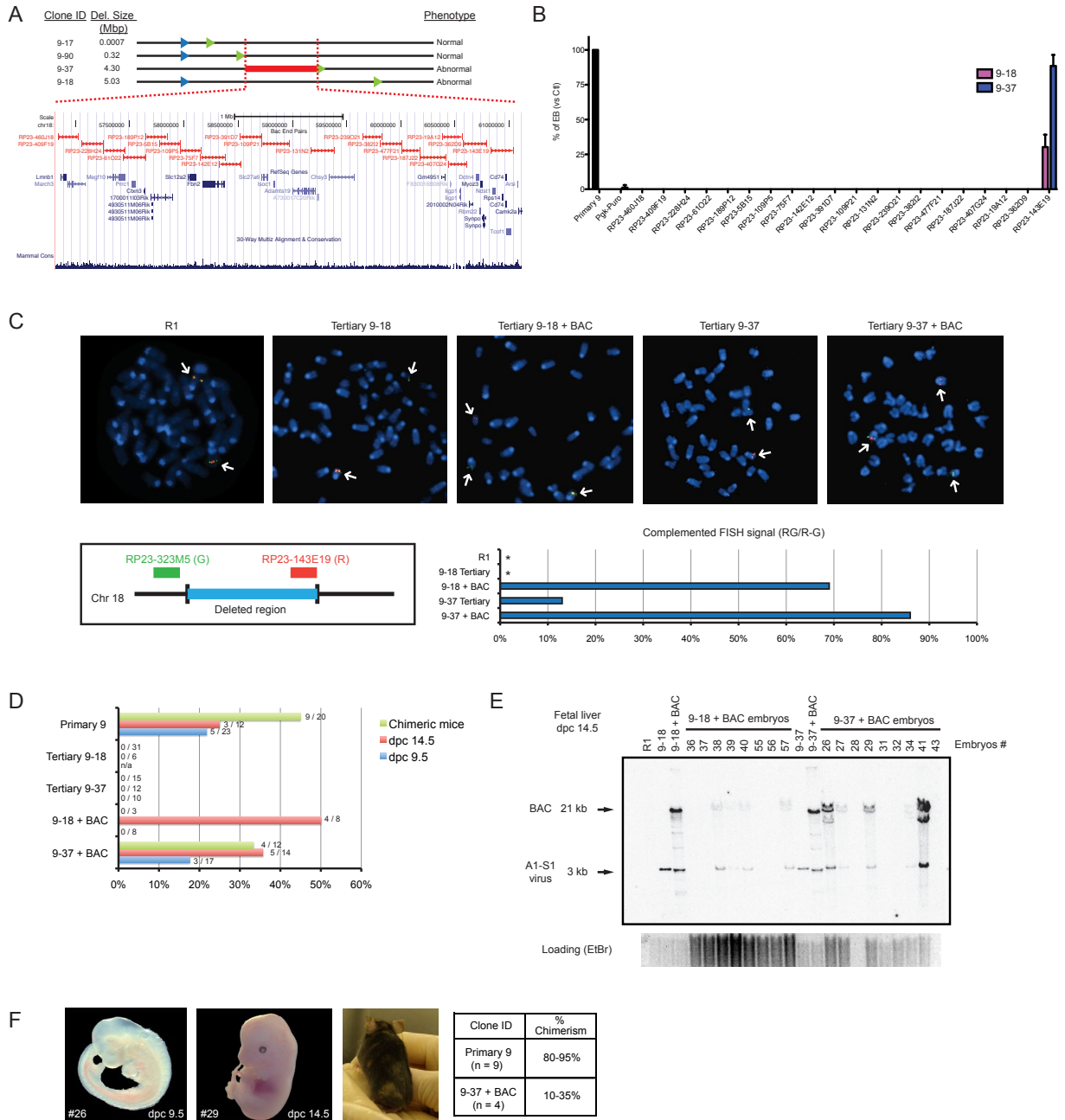


Figure 2.6: Rescue of differentiation defects of tertiary clones from DeIES family 9.

(A) Genomic representation of minimal deleted region leading to a differentiation phenotype in family 9 tertiary clones. Schema adapted from UCSC Genome Browser (B) Embryoid body differentiation assay following stable transfections of modified BACs covering the region of interest. Only BAC RP23-143E19 transfection could rescue the differentiation defect in tertiary clones 9-37 (complete) and 9-18 (partial). Bars represent average EB number (transfected tertiary clones normalized to control primary clone) from four independent transfections (n=4 tests per experiment). Complete and partial complementation leads to at least 300X and 100X

EB over background EB number, respectively. (C) FISH analysis of BAC-transfected clones. Upper panel: representative metaphase FISH images for selected clones. Lower left panel: FISH assay design, using two BACs as probes. RP23-323M5 is the control probe, which hybridizes outside of the deleted region, and is labeled with a green fluorescent marker (G). The BAC of interest, RP23-143E19 is labeled with a red fluorescent marker (R). Lower right panel: Complemented FISH signal reported as the proportion of cells which display a yellow (Red + Green: RG) signal combined with Red and Green (R-G) signals over the total number of cells analyzed (n=100 interphase cells and n=10 metaphase cells). *, no RG+(R-G) cell observed. (D) In vivo contribution of BAC-transfected and control clones to chimeric animals was evaluated by PCR for E9.5 and E14.5 embryos and by coat color chimerism for adult mice. Values are represented as proportion of neomycin positive embryos over total number of embryos analyzed and the proportion of adult chimeric mice over total number of newborns. Data related to non-transfected tertiary clone 9-18 has been previously reported [23]. (E) Southern blot analysis of DNA extracted from chimeric dpc14.5 fetal livers. EcoRV restriction digest combined with a neo probe revealed a 3Kb and a 21Kb fragment, corresponding to the primary anchor virus and modified BAC (EcoRV restriction sites in both insert and backbone), respectively. (F) Pictures of representative embryos, chimeric mouse and table of average chimerism estimated by coat color determination (newborn and adult mice). Dpc, days post coitum.

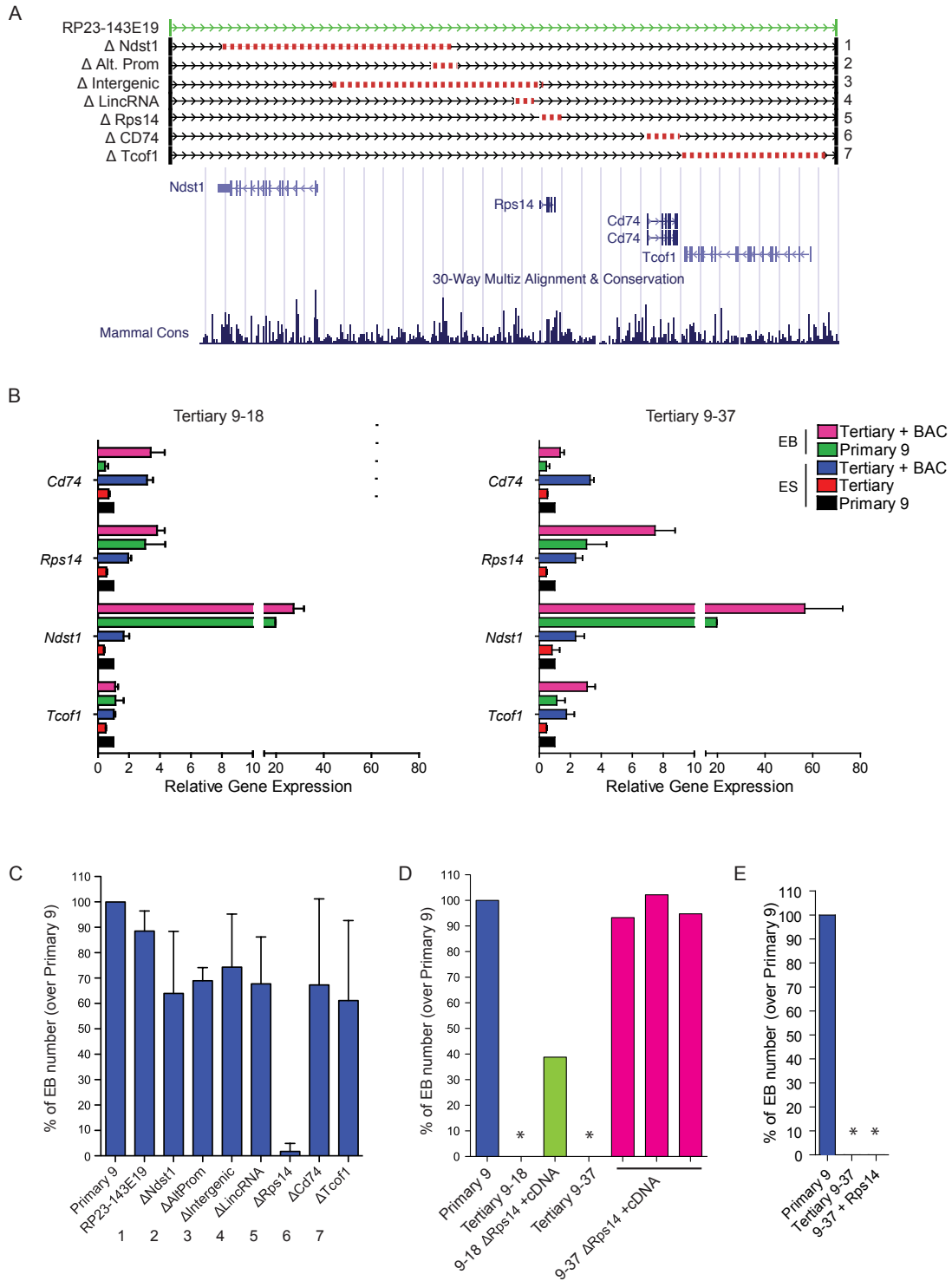


Figure 2.7: Candidate gene evaluation of DelES family 9.

(A) Genomic representation corresponding to the coverage of BAC RP23-143E19 and 7 deletion constructs. Schema adapted from UCSC Genome Browser (B) Expression analysis of candidate genes included in BAC RP23-143E19. Q-RT-PCR results are represented as

relative expression to that observed in family 9 primary clone (n=4 independent experiments in triplicates). Average ΔCt per gene, for family 9 primary clone normalized with GAPDH expression levels, in ESCs and EBs, respectively; Ndst1: 10.85 and 6.55; Tcof1: 5.82 and 5.84; Rps14: 0.51 and 0.95; Cd74: 13.09 and 14.39. (C) Relative EB number derived from clone 9-37 cells previously transfected with BAC RP23-143E19 or each of the 7 deletion constructs presented in “A”. Results are compared to those obtained with identical experiments performed with the corresponding primary clone 9 (n=4 independent experiments, in duplicate; values for primary clone 9 adjusted to 100%). (D) EB number following co-transfection of Rps14 deletion-containing BAC plus Rps14 cDNA. Values are represented as EB number of 3 transfected 9-37 clones and 1 transfected 9-18 clone over primary 9 clone levels. No EB was observed in cultures initiated with cells from tertiary 9-18 and 9-37 clones (*). (E) EB number following transfection of Rps14 cDNA in tertiary clone 9-37. No EB was observed in this condition (*, n=11 clones analyzed).

2.8 Methods

2.8.1 Generation of DelES collection

Viral producer cell lines and infection of target cells were conducted as described [23]. Reagents used for Cre-*loxP* recombination (A1 and S1 retroviruses, and pCX-Cre constructs) were described previously [23]. Briefly, following R1 ESCs [40] infection with anchor virus A1, approximately 288 puromycin resistant primary clones were isolated. Q-PCR assays were performed on genomic DNA to discard primary clones containing presumptive trisomies. Five million primary clone cells (one clone at a time) were infected with the saturation virus S1. Following hygromycin selection, 10^7 cells from these secondary populations (secondary population are derived from a single primary clone) were electroporated with 25 ug of supercoiled pCX-Cre and selected with G418, as described previously [23]. Up to 44 G418^R tertiary clones were isolated per electroporated secondary population and maintained in 96-well plates (labeled TER0xxx). ESCs maintained in 96-well plates were either dissociated manually or with a Biomek FX robot (Beckman Coulter) enclosed in a biosafety cabinet. G418 resistant tertiary clones sensitive to puromycin (puro^S) were arrayed together in 96-well plates (labeled CPC0xxx). “Normalized” 96-well plates were also generated with puro^S clones presenting similar proliferation rate for use in functional assays. ESCs were cryopreserved at each stage of DelES collection generation. A detailed description of the methods used for the generation of DelES collection can be found in Section 2.12.

2.8.2 High-throughput functional assays

Detailed descriptions of the high-throughput assays performed with puro^S clones arrayed in normalized plates are provided in Section 2.12. The Alkaline Phosphatase Kit (Chemicon) was used according to the manufacturer’s protocol. ESCs immunostained with a PE-conjugated mouse anti-human Ki67 monoclonal antibody (dilution 1:100, BD Biosciences) were analyzed by flow cytometry. Cell counts were performed by flow cytometry using TruCOUNT reference beads (BD Biosciences). Cell densities were evaluated by methylene blue staining of ESC colonies.

2.8.3 Embryoid body formation screening

Gelatin-plated clones were seeded in 96 well plates (Sarstedt) containing a semi-solid differentiation media and in parallel on a new gelatinized plate (NUNC). EBs were counted

following 8 days of differentiation, while colonies on gelatinized plates were stained with methylene blue twenty-four hours after seeding. Automated quantification of the methylene blue stained area was used to evaluate the cell input that produced the corresponding EB number (Metamorph software). Criteria were established to determine families with clones presenting abnormal EB formation phenotypes, i.e. insufficient EB number or disaggregation. The first was to exclude tertiary clones with low cell input values (methylene blue < 5%, n = 722 clones, 55.2%), which could be the result of a defect in proliferation or cell adhesion, or simply a technical issue. The number of EBs obtained for each tertiary clone was compared to that of the corresponding primary clone. Based on values obtained from larger format experiments, a tertiary clone was called abnormal when it formed less than 20% EBs relative to its primary clone (n = 96 out of 585). Our final criterion in identifying an abnormal family was to verify a correlation between decreased EB formation with a larger deletion size (where mapping was available).

2.8.4 DNA analyses

Genomic DNA (gDNA) from primary clones was extracted with DNeasy 96 Blood & Tissue Kit (Qiagen protocol) and used for Q-PCR screening of presumptive trisomies and mapping of proviral integration sites (see below and Section 2.12). Genomic DNA from primary clones, all tertiary clones (labeled TER0xxx), and puro^S tertiary clones (labeled CPC0xxx or MPL0xxx) were extracted using DNAzol (Invitrogen) by centrifugation in V-bottom 96-well plates. Southern blot analyses were performed as previously described [23]. To verify single integration of anchor virus, primary clone gDNA was digested with *BglIII-BamHI* restriction enzymes and detection performed with a neomycin probe. Southern blot analyses with tertiary clone gDNA (*EcoRI* restriction digest), were either performed with a neomycin probe to assess clonal diversity of rearrangements (e.g. clone classification into sub-family) or with a hygromycin probe to confirm the loss of hygromycin resistance gene. The presence/absence of hygromycin gene was also monitored by Q-PCR assays (Section 2.12).

2.8.5 Mapping of proviral integration sites

Integration sites of the anchor virus were mapped in primary clones by I-PCR or LM-PCR. Saturation virus integration sites were mapped in tertiary clones by LM-PCR. The I-PCR approach was previously described [23]. The LM-PCR strategy, which relies on specific oligonucleotides described in Tableau S2.7, was adapted from a published protocol [41] summarized in Section 2.12. DNA sequences corresponding to proviral integration sites were mapped using the BLAT

alignment tool of the UCSC Genome Browser (<http://genome.ucsc.edu/>, NCBI mouse Build 37) [42].

2.8.6 Construction of database

Biological material tracking and database construction. Frozen cells, DNA, RNA, and maintenance plates were identified with bar codes and specific labeling (details provided in Section 2.12). A PostgreSQL database was set up in order to maintain a centralized repository of the biological sample's storage locations, as well as to accumulate various types of results and annotations. A web front-end running on a Webware for Python application server was also developed to enable a user-friendly access to the majority of the data contained in the database. Numerous visualization, data-mining, and sample management tools are still under development to provide a flexible interface to query the annotations and to manage access to the biological samples. Genome annotations presented in some of DeIES' visualization tools as well as the gene searching capabilities rely on locally deployed instances of mirBASE [43] (<http://microrna.sanger.ac.uk/>) and Ensembl [44] (<http://www.ensembl.org/index.html>). Published LincRNA [16] and Ultraconserved Elements [18] annotations were also inserted into the DeIES database to facilitate integration. More functionalities of DeIES database are described in Section 2.12.

2.8.7 BAC engineering

BACs from the RP23 library (pBACe3.6 vector [45]) were obtained from the BACPAC Resource Center (Children's Hospital Oakland Research Institute, Oakland, California) and maintained in their original host strain DH10B in the presence 12 mg/ml chloramphenicol (unmodified BACs) or 25mg/ml kanamycin (retrofitted BACs). SelectaBAC retrofitting strategy, adapted from published protocols [27,46,47], is described in the Section 2.12.

2.8.8 BAC complementation

ESCs maintained on a feeder layer in 12-well plates were transfected with 2ug of circular BAC DNA using Lipofectamine 2000 Reagent (Invitrogen), according to manufacturer's protocol. Selection was started 48h later, with the following concentration of drugs maintained for at least 5 days: 1.5 ug/ml puromycin (Sigma), or 150 ug/ml hygromycin (Roche), or 15 ug/ml blasticidin (Sigma), or 30 ug/ml zeocin (Invitrogen). ESC differentiation in embryoid bodies was performed in a LIF-deprived semi-solid media, as described [11]. Genomic DNA from BAC transfected clones was isolated using DNAzol, according to the manufacturer's instructions

(Invitrogen). Southern blot detection of transfected BAC DNA was performed using *EcoRV* digestion and a probe specific to the neomycin gene, as described [48]. Total cellular RNA was isolated from BAC transfected clones (undifferentiated ESCs or embryoid bodies) with Trizol (Invitrogen), according to the manufacturer's instructions. Quantitative RT-PCR assays were performed according to standard protocols described in Section 2.12.

2.8.9 Generation of chimeric mice

Mouse chimeras were generated by the transgenic facility of IRIC. ESCs [40] corresponding to primary clone no.9, tertiary clones 9-18 and 9-37 (with *in vitro* phenotype) and BAC-transfected 9-18 and 9-37 clones (rescued *in vitro* phenotype), were injected into C57BL/6 blastocysts or aggregated with CD1 morulas. Tertiary clone 9-18 results showed in Figure 2.6 are already published [23]. ESCs contribution to chimeric embryos (at E9.5 and E14.5) was evaluated by PCR using genomic DNA extracted with a standard protocol (lysis with of 100mM NaCl, 10mM Tris pH 8.0, 25mM EDTA pH 8.0, 0.5% SDS, and 2.5 ug/ml Proteinase K followed by phenol-chloroform extraction and ethanol precipitation) or Sigma REDExtract-N-Amp Tissue PCR kit (primers specific to the neomycin gene are described in the Tableau S2.7). Southern blot analysis was performed as previously described [48] with a neomycin probe and *EcoRV*-digested genomic DNA isolated from E14.5 fetal livers using DNAzol (Invitrogen). ESC contribution to adult mice was determined by evaluation of the coat color chimerism.

2.8.10 FISH analysis

BACs RP23-143E19 and RP23-323M5 were labeled with Spectrum Orange and Green fluorochromes, respectively, via Nick Translation (Abbott Molecular Cat. No. 32-801300). BACs have been tested both separately and together on mouse control cells from Leukemia Cell Bank of Quebec. A minimum of one hundred interphase nuclei and ten metaphases were evaluated per sample and results are given as signal distribution percentages.

2.9 Acknowledgements

The authors thank Jean Duchaine from IRIC HTS platform, Véronique Paradis and Sébastien Harton from IRIC transgenesis platform, Mélanie Fréchette and Edlie St-Hilaire who assisted with mice and MEF production, Christian Charbonneau from IRIC imagery platform, Danièle Gagné from IRIC flow cytometry platform, Annie Bilodeau for FISH experiments (BCLQ), Dr. Pierre Chagnon and Raphaëlle Lambert from IRIC genomic platform for their expertise with quantitative RT-PCR and Dr Janet Rossant for scientific comments.

2.10 References

1. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292: 154-156.
2. Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78: 7634-7638.
3. Smith AG (2001) Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* 17: 435-462.
4. Rossant J (2008) Stem cells and early lineage development. *Cell* 132: 527-531.
5. Murry CE, Keller G (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 132: 661-680.
6. Boheler KR (2009) Stem cell pluripotency: a cellular trait that depends on transcription factors, chromatin state and a checkpoint deficient cell cycle. *J Cell Physiol* 221: 10-17.
7. Malumbres M, Barbacid M (2001) To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer* 1: 222-231.
8. Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 115: 281-292.
9. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* 87: 27-45.
10. Shen MM, Leder P (1992) Leukemia inhibitory factor is expressed by the preimplantation uterus and selectively blocks primitive ectoderm formation in vitro. *Proc Natl Acad Sci U S A* 89: 8240-8244.
11. Keller G, Kennedy M, Papayannopoulou T, Wiles MV (1993) Hematopoietic commitment

- during embryonic stem cell differentiation in culture. *Mol Cell Biol* 13: 473-486.
12. Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132: 567-582.
 13. Silva J, Smith A (2008) Capturing pluripotency. *Cell* 132: 532-536.
 14. Macarthur BD, Ma'ayan A, Lemischka IR (2009) Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* 10: 672-681.
 15. Kim J, Chu J, Shen X, Wang J, Orkin SH (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132: 1049-1061.
 16. Guttman M, Amit I, Garber M, French C, Lin MF, et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458: 223-227.
 17. Gangaraju VK, Lin H (2009) MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* 10: 116-125.
 18. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, et al. (2004) Ultraconserved elements in the human genome. *Science* 304: 1321-1325.
 19. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.
 20. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799-816.
 21. Chen Y, Yee D, Dains K, Chatterjee A, Cavalcoli J, et al. (2000) Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat Genet* 24: 314-317.
 22. Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, et al. (2004) The knockout mouse project. *Nat Genet* 36: 921-924.

23. Bilodeau M, Girard S, Hebert J, Sauvageau G (2007) A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Methods* 4: 263-268.
24. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294: 853-858.
25. Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet* 31: 180-183.
26. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.
27. Wang Z, Engler P, Longacre A, Storb U (2001) An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res* 11: 137-142.
28. Sirard C, de la Pompa JL, Elia A, Itie A, Mirtsos C, et al. (1998) The tumor suppressor gene *Smad4/Dpc4* is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev* 12: 107-119.
29. Kielman MF, Rindapaa M, Gaspar C, van Poppel N, Breukel C, et al. (2002) *Apc* modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nat Genet* 32: 594-605.
30. Lin X, Wei G, Shi Z, Dryer L, Esko JD, et al. (2000) Disruption of gastrulation and heparan sulfate biosynthesis in *EXT1*-deficient mice. *Dev Biol* 224: 299-311.
31. Ebert BL, Pretz J, Bosco J, Chang CY, Tamayo P, et al. (2008) Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* 451: 335-339.
32. Narla A, Ebert BL (2010) Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115: 3196-3205.
33. Barlow JL, Drynan LF, Hewett DR, Holmes LR, Lorenzo-Abalde S, et al. (2010) A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human

5q- syndrome. *Nat Med* 16: 59-66.

34. Doudna JA, Rath VL (2002) Structure and function of the eukaryotic ribosome: the next frontier. *Cell* 109: 153-156.
35. Robledo S, Idol RA, Crimmins DL, Ladenson JH, Mason PJ, et al. (2008) The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* 14: 1918-1929.
36. Fumagalli S, Di Cara A, Neb-Gulati A, Natt F, Schwemberger S, et al. (2009) Absence of nucleolar disruption after impairment of 40S ribosome biogenesis reveals an rpL11-translation-dependent mechanism of p53 induction. *Nat Cell Biol* 11: 501-508.
37. Adams DJ, Biggs PJ, Cox T, Davies R, van der Weyden L, et al. (2004) Mutagenic insertion and chromosome engineering resource (MICER). *Nat Genet* 36: 867-871.
38. Goodwin NC, Ishida Y, Hartford S, Wnek C, Bergstrom RA, et al. (2001) DelBank: a mouse ES-cell resource for generating deletions. *NatGenet* 28: 310-311.
39. Bilodeau M, MacRae T, Gaboury L, Laverdure J, Hardy M, et al. (2009) Analysis of blood stem cell activity and cystatin gene expression in a mouse model presenting a chromosomal deletion encompassing *Csta* and *Stfa211*. *PloS One*.
40. Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC (1993) Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *ProcNatlAcadSciUSA* 90: 8424-8428.
41. Mikkers H, Allen J, Knipscheer P, Romeijn L, Hart A, et al. (2002) High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nat Genet* 32: 153-159.
42. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12: 656-664.
43. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36: D154-158.

44. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, et al. (2009) Ensembl 2009. *Nucleic Acids Res* 37: D690-697.
45. Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, et al. (1999) A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics* 58: 250-253.
46. Gong S, Yang XW, Li C, Heintz N (2002) Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res* 12: 1992-1998.
47. Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, et al. (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* 73: 56-65.
48. Kroon E, Kros J, Thorsteinsdottir U, Baban S, Buchberg AM, et al. (1998) *Hoxa9* transforms primary bone marrow cells through specific collaboration with *Meis1a* but not *Pbx1b*. *Embo J* 17: 3714-3725.
49. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 36: D724-728.
50. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177-183.
51. Dang VT, Kassahn KS, Marcos AE, Ragan MA (2008) Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet* 16: 1350-1357.
52. Luedi PP, Hartemink AJ, Jirtle RL (2005) Genome-wide prediction of imprinted murine genes. *Genome Res* 15: 875-884.
53. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 36: D724-728.
54. McKusick VA (1998) *Mendelian Inheritance in Man: A Catalog of Human Genes and*

Genetic Disorders. Baltimore: Johns Hopkins University Press.

2.11 Supplemental Figures and Tables

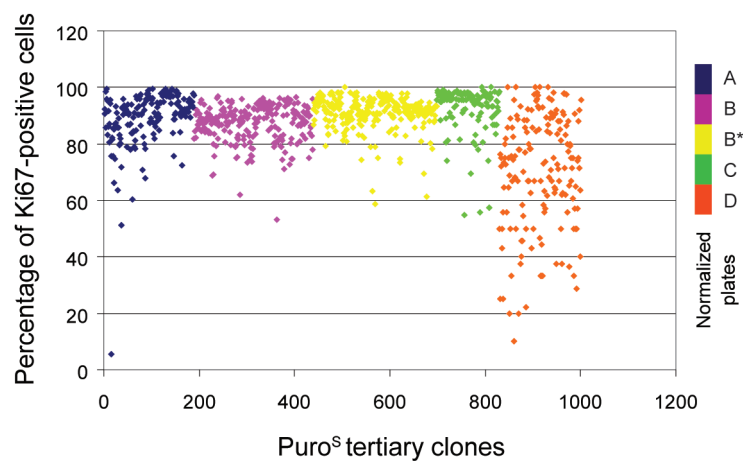


Figure S2.1: Graphical representation of Ki67 values for each puro^S tertiary clones grouped in normalized plate sets based on proliferation rate.

PuroS tertiary ESC clones presenting similar proliferation rate were arrayed together in 96-well plates: five normalized plate sets were generated (A, B, B*, C, and D) based on the timing of harvest (A=earliest collection, D, latest). Ki67 expressing cells were quantified by flow cytometry. Most tertiary clones presenting <60% Ki67+ cells had a slow proliferation rate (e.g. arrayed in plate set D) compared to other clones.

Available languages:
English
French

Family ID number

DeIES navigation tabs: Home, Plates, Families, Screen, Selection, Statistics

View Selector: Karyoview, Families

Color Legend:

- Current family
- Families with mapped deletions
- Families with only un-mapped tertiary clones
- Families with no tertiary clones

Search for Specific gene Family Clone: Search Database

MGI genes

microRNA

Ultraconserved elements

lincRNA

Mapped tertiary clones deletion details

Track Display Options:

- MGI Genes
- BAC End Pairs
- microRNA
- lincRNA
- UC elements

Tertiary Clones - Actual deletion size

Sub-families

DeIES: engineered Deletions in Embryonic Stemcells

Details of family: S082.12

Primary Clone Details:

- Detection of host by southern blot: Present (8.65 Kb)
- Location of primary integration: Chr. 9: 100,333,697
- View in Genome Browser: GDB

Graphical Overview of Family's Genomic Content

Family deletion details:

CloneID	Deletion size	Phenotype	Nb MGI Genes
S082.11	8,911,782 bp	Phenotype: Normal	38
S082.17	13,305,523 bp	Phenotype: Normal	107
S082.12	17,886,224 bp	Phenotype: Normal	185
S082.26	35,117,820 bp	Phenotype: Diff. homozyg.	204
S082.29	70,237,795 bp	Phenotype: Diff. Prof.	274

Sub-family A: North Southern Blot Size: 4.95 Kb

Batch	CloneID	Integration	Score	Southern	Phenotype	Comment	Density	KbP	MS	AK_PQ ₂
1*	S082.12	Undetermined	-	Present	Normal	-	1000/10	01.00	2	4

Sub-family B: North Southern Blot Size: 7.05 Kb

Batch	CloneID	Integration	Score	Southern	Phenotype	Comment	Density	KbP	MS	AK_PQ ₂
A	S082.14	gPCR	1.00	Present	Normal	-	6700/10	03.00	2	2

Sub-family C: North Southern Blot Size: 7.34 Kb

Batch	CloneID	Integration	Score	Southern	Phenotype	Comment	Density	KbP	MS	AK_PQ ₂
1*	S082.24	Undetermined	-	Present	Normal	-	2107/10	02.00	2	4
1*	S082.41	Present	2.00	Present	Normal	-	2080/10	03.00	2	5

Sub-family D: North Southern Blot Size: 10.90 Kb

Batch	CloneID	Integration	Score	Southern	Phenotype	Comment	Density	KbP	MS	AK_PQ ₂
1*	S082.22	Present	1.00	Present	Normal	-	1002/10	01.70	2	4

©IC: Institute for Research in Immunology and Cancer
20 July 2009 - 15:07 | [Home](#) | [Help](#) | [Feedback](#)

Figure S2.2: DeIES interface.

Overview of major functionalities of DeIES web database, available online at: www.bioinfo.irc.ca/deles. The Families tab is depicted here as an example. It is divided into 2 sub-sections: Karyoview and Families. The Karyoview sub-section provides a graphical representation of the mapped primary virus insertion sites as well as the orientation of the deletions for a given family

(color-coded icons, as indicated). Deleted chromosome portions are displayed as red lines to the left of the ideograms. Especially noteworthy is the graphical representation of the genomic context of the deletions. Currently available and customizable tracks are: MGI genes, RP23 library BACs, miRBASE miRNAs, lincRNAs and ultraconserved elements. The Families subsection presents most of the accumulated genetic and phenotypic observations related to puroS tertiary clones. Raw screen data are also accessible from this tab. For more DeIES functionalities, see Supplemental Experimental Procedures.

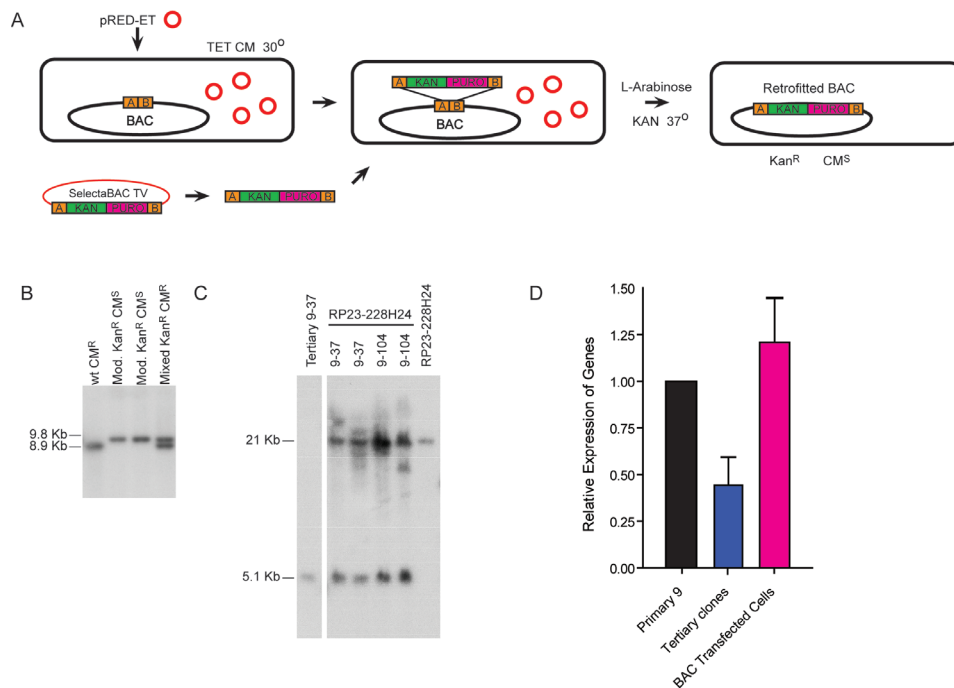


Figure S2.3: BAC engineering for DelES complementation.

(A) The SelectaBAC retrofitting strategy was optimized to introduce a targeting vector (TV) containing a eukaryote (puromycin is depicted) and a prokaryote (kanamycin; Kan) resistance gene into the chloramphenicol (CM) gene of the BAC vector. This protocol relies on the inducible expression (addition of L-arabinose and temperature shift) of λ phage proteins which mediate homologous recombination events between the homology arms of the targeting cassette (identified as A and B) and the BAC vector. Bacteria containing the retrofitted BAC are resistant to kanamycin (Kan^R) and sensitive to chloramphenicol (CM^S). (B) Southern blot performed with BAC DNA extracted from bacteria. EcoRI restriction digest combined with an external probe hybridizing to the *SacB* gene of the BAC vector was used to detect proper BAC modification. Fragments of 8.9 kb and 9.8 kb were observed for the unmodified (wt CM^R) and the modified BACs (Mod. Kan^R CM^S), respectively. Both fragments were observed with BAC DNA extracted from mixed bacterial colonies (contain both modified and unmodified BAC) (C) Southern blot performed with genomic DNA extracted from ESCs stably transfected with a modified BAC. NheI restriction digest (NheI sites in both BAC insert and vector) combined with a neo probe revealed a 5.1Kb and a 21Kb fragment, corresponding to the integration site of the primary anchor virus and the modified BAC, respectively. (D) Combined relative expression of *Lmnb1*, *Iigp1*, *Isoc1* and *Slc12a2*, in tertiary clone 9-37 following BAC transfections. Values are relative to family 9 primary clone expression levels with standard error (\pm SE), representative of 2 independent experiments performed in duplicate reactions.

Tableau S2.1: Global statistics of DelES library generation.

Primary clones	
Total number (mapped)	300 (238)
No signs of trisomy (chr 1,8,11,14)	230
Generated tertiary clones	167
Generated Puro ^S clones	135
Generated Hygro ⁻ clones (qPCR score >= 3)	94
Secondary populations	
Total number	176
Tertiary Clones	
Total number	4990
Puromycin sensitive clones	1670
MPL clones	1307
Hygro ⁻ clones	1018
Nb of sub-families (based on southern)	618
Nb of sub-families With Hygro ⁻ members	355
Nb of subfamilies with a mapped member	266
Mapped clones	423
Mapped unique deletions	294

Puro^S, Puromycin sensitive; Hygro⁻, Hygromycin negative; MPL, master plate clones; Nb, number.

Tableau S2.2: Molecular characterization details of DeIES clones used in functional assays.

Listing of all DeIES clones included in the functional screen grouped in families according to anchor site, and by subfamilies according to Southern blot fragment size.

Due to size constraints, this table is not included in the present document.

File is available as an Excel file from PLoS Genetics OpenAccess at:

doi:10.1371/journal.pgen.1001241.s006

Tableau S2.3: Summary of deletions distribution.

Chromosome	Number of unique deletions	Statistics	
		Average size (bp)	Median size (bp)
chr1	24	1,315,206	749,709
chr2	31	2,192,129	1,348,099
chr3	12	4,635,361	1,916,165
chr4	34	2,641,852	1,882,398
chr5	15	4,493,260	1,211,535
chr6	22	9,719,444	3,133,375
chr7	25	7,338,688	2,099,453
chr8	10	1,976,226	664,597
chr9	19	11,392,100	13,059,580
chr10	17	9,343,643	1,540,903
chr11	18	3,727,251	2,162,374
chr12	8	5,831,859	2,393,053
chr13	3	3,566,830	2,893,092
chr14	7	7,561,018	1,800,012
chr15	17	3,285,026	2,245,474
chr16	10	2,513,349	1,008,103
chr17	11	6,271,243	5,670,907
chr18	10	1,863,542	171,747
chr19	0	0	0
chrX	0	0	0
chrY	0	0	0
Total Genome	293	4,951,151	1,610,759

Tableau S2.4: Known or predicted haploinsufficient/imprinted candidate genes present in the vicinity of DelES anchor loci that are not associated with chromosomal deletions.

Family	Chromosome	Insertion Position	G418 ^R tertiary clones (Max of 44)	Puro ^R tertiary clones	Number of G418 ^R Puro ^R tertiary clones	Hygro ^R tertiary clones	Deleted Genes*
5005	chr3	144,272,281	16	2	0	0	2410004B18Rik, A1747448, Bc110, Bxdc5, C1ca1, C1ca2, C1ca3, C1ca4, C1ca5, C1ca6, Cc124a1, C1bs, Cyr61 , Ddah1 , EG622139, Gng5, Lpar3, Mcoln2, Mcoln3, Odj2, Sh3glb1, Spata1, Ssx2lp, Syde2, Wdr63, Znhh16
5046	chr9	120,675,709	44	0	0	0	5630454E08Rik, 6330176C04Rik, Acaa1a, Acaa1b, Acv2b, Axud1, Ccr6, Ctdsp1, Cx3cr1, Dlec1, EG546166, Efl1b, Endoglin, Entpd3, Gorasp1, Iiga9, Mobb, Myd88, Myrip, Oxsr1, Pcd1, Rpl14, Rpsa, Scn10a, Scn11a, Sorba, Slc22a13, Slc22a14, Slc25a38, Ttc21a, Villi, Wdr48, Xirp1, Xylb
5053	chr11	101,315,037	44	3	0	0	1110033F04Rik, 1110036O03Rik, 1110054F19Rik, 1300010M03Rik, 2300003K06Rik, 2300006N05Rik, 2310043L02Rik, 4733401H21Rik, A030010K20Rik, Aarsd1, Ach, Acc2, Acc3, Atp6v0a1, Becn1, Ccnd5f6, Ccr10, Cnp, Cntd1, Cntrap1 , Coasy, Dhx58, Dnajc7, E1f1, Ezr1, Fkbp10, Gaps, Gas1, Ghdc, Hap1, Hprt, Hsd17b1, Hspb9, Ifi35, Jup , Kat5a, Kchh4, Kih10, Kih11, Krt13, Krt14, Krt15, Krt16, Krt17, Krt19, Krt31, Krt32, Krt33a, Krt33b, Krt34, Krt35, Krt36, Krt39, Krt40, Krt42, Krt9, Krtap1-3, Krtap1-4, Krtap1-7, Krtap2-4, Krtap3-1, Krtap3-2, Krtap3-3, Krtap3-1-2, Krtap4-16, Krtap4-7, Krtap4-8, Krtap4-9, Krtap9-1, Krtap9-5, Mlx, Naglu, Nkiras2, Nf5c3l, Plekhh3 , Psmc3lp, Psme3, Prf1, Rab5c, Ramp2, Rpl27, Rundc1, Slat3, Stats5a, Stat5b, Ttc25, Tubg1, Tubg2, Vps25, Wnk4, Zfp385c
5201	chr13	31,773,949	22	9	0	0	D930007J09Rik, Foxc1 , Grnd5, Mylk4, Serpinb1ps1, Serpinb1a, Serpinb1b, Serpinb1c, Serpinb6b, Serpinb9, Serpinb9b, Serpinb9c, Serpinb9d, Serpinb9e, Serpinb9f, Serpinb9g, Wmp1
5206	chr10	12,270,785	15	1	0	0	Epm2a, Fbxo30, Grrm1, Shprh, Utrn
5219	chr9	59,571,683	22	0	0	0	2410076I21Rik, 4933407119Rik, 6030419C18Rik, Adpgk , Arid3b, Arh11, Bbs4, Bruno6l6, Ccdc33, Cd276, Cyp11a1, Gramd2, Hcrn4, Hexa, Islr, Islr2, Lox1, Neo1, Nptn, Parp6, Pkm2, Pmi , Sema7a, Stom1, Stra6, Tbc1d21, Tmem202, Ubi7
5247	chr15	61,822,445	28	0	0	0	Fam49b, Gsdmc1, Gsdmc2, Gsdmc3, Gsdmc4, Gsdmd1, PV11
5255	chr14	21,034,019	20	1	0	0	1000394441, 100040671, 100041306, 1810069B07Rik, 2700060E02Rik, EG432825, Gng2, Kcnk5, Nid2, Nkiras1, Rpl15, Ube2e1, Ube2e2
5261	chr6	95,665,810	22	0	0	0	1700123L14Rik, A130022J15Rik, A16ip5, Fam19a1, Fam19a4, Fmdd4b, Lmod3, Suclg2, Tmf1, Uba3

*Presented are hypothetical deleted genes, based on median size. Known or predicted haploinsufficient/imprinted genes are in red

Tableau S2.5: Genomic content of DeES mapped deletions.

List of current DeES mapped families and elements included in the family largest deletion. Clusters (genes, miRNA, ultraconserved elements, lincRNA) are in represented in red and have been identified by manually looking at all mapped families. Number of genes with associated mouse phenotypes has been extracted from MGI (Bult et al., 2008). OMIM entries are based on human syntenic regions (McKusick, 1998).

Due to size constraints, this table is not included in the present document.

File is available as an Excel file from PLoS Genetics OpenAccess at:

doi:10.1371/journal.pgen.1001241.s008

Tableau S2.6: Embryoid body screen raw data.

Table presenting raw data obtained for primary and tertiary clones EB formation screening. EB number were counted manually and seeding densities were obtained by an automated evaluation of the methylene blue stained area.

Due to size constraints, this table is not included in the present document.

File is available as an Excel file from PLoS Genetics OpenAccess at:

doi:10.1371/journal.pgen.1001241.s009

Tableau S2.7: List of oligonucleotides used in all reported assays.

Adapter and PCR oligonucleotides used for mapping proviral integration sites by LM-PCR.

Adapter oligonucleotides ^a		
NlaIII linker	GTAATACGACTCACTATAGGGCTCCGGCTTAAGGGACCATG	Phos-GTCCCTTAAGCGGAG-C3spacer
Tsp5091 linker	GTAATACGACTCACTATAGGGCTCCGGCTTAAGGGAC	Phos-AATTGTCCCTTAAGCGGAG-C3spacer-3'
Oligonucleotides ^b used for PCR amplification of A1 anchor virus integration site (5'SIN LTR found in primary clones or tertiary clones)		
First round	GTAATACGACTCACTATAGGGCTCCG (anneals in adapter)	CTGGAAACATCTGATGGTTCTCTAGCTTG (anneals over SIN deletion junction)
Second round	AGGGCTCCGCTTAAGGGAC (nested adapter primer)	CTCTAGCTTGCCAAACCTACAGGT (nested primer on SIN deletion junction)
Third round	AGGGCTCCGCTTAAGGGAC (nested adapter primer)	CCAAACCTACAGGTGGGGTC
Oligonucleotides ^b used for PCR amplification of A1 anchor virus integration site (3'SIN LTR found in primary clones)		
First round	ACCTGTAGGTTGGCAAGCTAGAG (anneals over SIN deletion junction)	GTAATACGACTCACTATAGGGCTCCG (anneals in adapter)
Second round	CAAGCTAGAGAACCATCAGATGTTCCAG (nested primer on SIN deletion junction)	AGGGCTCCGCTTAAGGGAC (nested adapter primer)
Third round	CCAAACCTACAGGTGGGGTC	AGGGCTCCGCTTAAGGGAC (nested adapter primer)
Oligonucleotides ^b used for PCR amplification of S1 saturation virus integration site (3'LTR found in tertiary clones)		
First round	CAGGATATCTGTGGTAAGCAGTTCCTGCC (anneals in LTR)	GTAATACGACTCACTATAGGGCTCCG (anneals in adapter)
Second round	GCTCAGGGCCAAAGACAGATGG (nested LTR primer)	AGGGCTCCGCTTAAGGGAC (nested adapter primer)
Third round	AGTCTCCGATGAGCTGCG (nested LTR primer)	AGGGCTCCGCTTAAGGGAC (nested adapter primer)

^aAdapter and ^bPCR amplification oligonucleotides were purchased from Integrated DNA Technologies and Invitrogen, respectively.

Q-PCR assays employed to detect presumptive trisomies of chromosome 1, 8, 11, and 14

Chromosomal band	Gene	Accession number	Primer_A	Primer_B	Universal Probe Library probe number ^a
1qB	<i>Actr1b</i>	NM_146107	ATGCAGCCAAGATCAGAGC	TGAAAGAGAGTGGGGCAAAC	21
3qA1 ^b	<i>E2f5</i>	NM_007892	CCTCCAGTGACCACATT CAGT	TGAACTGGAGCCTGCTGTAA	55
8qA1.1	<i>Lamp1</i>	AK004637	GGCATCTGGCTGGGTACA	GGAAAGTGGCAGCTCACG	99
8qA4	<i>Msr1</i>	LD4274	GTGCTAGTGAAGCCCATGA	CCAGCGATCATCAGATTG	31
11qA1	<i>Ramp3</i>	NM_019511	GGCTCGTTCCTAGTTTCT	TCAGGACTAGAAATGGGTCAGG	53
11qD	<i>Hoxb4</i>	NM_010459	CTCTCGACCGCTACACT	GGTAGCGAATTGTAGTAAACTCC	62
14qA1	<i>Pxk</i>	NM_145458	CATTAAACCAGATAAGGGTTGTC	AATGTTGGCCCTCCTCTAC	102

^aUniversal Probe Library, UPL; Roche; ^bChromosome 3 used as calibrator

Primer and probe sets^a used for Q-PCR detection of hygromycin and neomycin resistance genes.

Gene	Forward primer	Reverse primer	Internal probe
Neomycin	GCGCCCGTTCTTTTG	CCTCGTCTGCAGTTCATTCA	TET-CAAGACCAGCTGTCCGGTGCC-BHQ
Hygromycin	CAGCGAGAGCCTGACCTATTG	GCAGGCTTGGCAACGTGACA	FAM-ATCTCCCGCTGCACAGG-BHQ

^aOligonucleotides were purchased from Integrated DNA technologies.

Primer and probe sets used for RT-Q-PCR studies.

Gene	Forward Primer	Reverse Primer	Universal Probe Library probe number ^a
<i>Lmb1</i>	GGGAAGTTTATCGCTTGAAGA	ATCTCCCAGCCTCCATT	15
<i>Isoc1</i>	GGAGTGTCTGTGTTGTTGGAG	ACGATGTGAACCTCGATGC	1
<i>Igp1</i>	AAGCACAATTTCCACCAG	GCCGAAATAATGAAGAAATCG	80
<i>Slc12a2</i>	TGGGAGAAGGTGCACAATAC	TGTTTGGCTTATACGACCA	66
<i>Ndst1</i>	GCTGGACAAGATCATCATGG	AAGTGTGTATGAGATGCTGATAG	99
<i>Tco1</i>	GGACTCAGACAGTGAAGACGA	GGACCTTTCCAGAGGCTTT	20
<i>Cd74</i>	GCCCTAGAGACCAGAAAGG	TGGTACAGGAAGTAAGCAGTGG	21
<i>Rps14</i>	CAGAGGTCAAGGAGTTGG	AGAGCTCTGAGGGGTGACTG	85

^aUniversal Probe Library, UPL; Roche

2.12 Supplemental Experimental Procedures

2.12.1 DelES library generation

R1 ESCs [1] were infected with low titer of virus A1 (virus-containing supernatant diluted 1:500) and 288 primary ESC clones were isolated following 6-9 days of puromycin selection (1.5 ug/ml, Sigma) (Figure 2.1C and Tableau S2.1). Clonal analysis performed by Southern Blot using genomic DNA extracted from primary clones revealed multiple insertions of the A1 virus at low frequency (10 out of 235 analyzed genomic DNA samples). Prior to the second infection (retrovirus S1), primary ESC clones were screened for potential trisomies, which were frequently observed in these culture conditions (e.g. chromosome 1, 8, 11, and 14) (see real-time quantitative PCR section below). From these analyses, 20.1% of primary clones (58 out of 288 analyzed) were rejected (Figure 2.1B and Supplemental Table 1). Among primary clones that appear free of presumptive trisomy, 176 were independently infected with the S1 saturation virus (virus-containing supernatant diluted 1:12), and then selected with hygromycin (150 ug/ml, Roche) for 7 days, generating the secondary populations used for Cre-electroporation (Figure 2.1C and Tableau S2.1). Following 7-9 days of G418 selection (300 ug/ml, Invitrogen), 4929 tertiary clones related to 156 anchor sites were isolated (Figure 2.1C and Tableau S2.1). No G418^R tertiary clones could be derived from 20 secondary clone populations, mainly for technical reasons. Puromycin sib-selection was conducted for each tertiary clone. An average of 11 ± 10 (range: 0 to 42 clones) puromycin-sensitive (puro^S) tertiary clones were collected per family in 96-well plates (n=1670 Cherry-picked clones; CPC clones) (Figure 2.1C, Tableau S2.1). From the 156 families of G418^R tertiary clones, 21 did not contain puro^S tertiary clones (puro^S tertiary clones are related to 135 anchor site) (Figure 2.1C).

A large proportion of puro^S tertiary ESC clones (n=1307), containing chromosomal rearrangements anchored to 104 independent loci, were used to conduct preliminary functional screens (Figure 2.1C and Tableau S2.1). Puro^S tertiary ESC clones presenting similar proliferation rate were arrayed together in 96-well plates, in order to get homogenous cell densities for functional studies. Five normalized plate sets were generated (A, B, B*, C, and D) based on the timing of harvest (A=earliest collection, D, latest) (see Supplemental Table 2 for normalized plate set attribution for each tertiary clone). A robotic cell culture system assisted the expansion of cells in normalized plates and the seeding of the following functional assays: alkaline phosphatase staining, flow cytometry analysis, and colony staining (see *functional assays* sections below). The last column of each plate was left empty in order to add control samples (2 wells each): R1 wild-type ESCs, and tertiary clones 9-35, 9-18, and 9-104. Cells from normalized plates

were frozen in 96-wells plates (labeled MPL0xxx) and genomic DNA extracted with DNAzol, according to manufacturer's instructions (Invitrogen).

2.12.2 Cryopreservation of cells

At each stage of the pipeline, cells were cryopreserved in liquid nitrogen for archiving of the library (long-term viability confirmed after 3 years of storage). Primary clones and secondary clone populations were individually frozen in 1ml-format CryoTubes™ (NUNC). Puro^S tertiary (CPC) clones, maintained in 96-well plates, were individually frozen in cryotubes labeled with a 2D bar code (CryoBank™, NUNC). All tertiary clones (plates labeled TER0xxx) and isolated puro^S tertiary clones (plates labeled CPC0xxx) were frozen in 96-well polypropylene plates (Costar, Fisher Scientific) sealed with rubber mats (Fisher Scientific).

2.12.3 Functional assays

2.12.3.1 Alkaline phosphatase detection

Prior to alkaline phosphatase detection, ESCs were maintained on a feeder layers in presence of LIF, as previously described [2]. After dissociating the cells from the normalized plates, the following proportions of cells were used to seed the phosphatase alkaline assay plates: 2%, 4%, and 8% (2%, 6%, and 18% for normalized plate set A only). Following three days of culture, alkaline phosphatase detection was performed with two of these plates, according to the manufacturer's instructions (Alkaline Phosphatase Detection Kit, Chemicon). A score from 1 (differentiated) to 5 (undifferentiated) was manually assigned to each clone using an inverted microscope.

2.12.3.2 Flow cytometry analyses

Cells from tertiary clones, maintained on 96-well gelatinized plates in presence of LIF, were dissociated. Fifty percent of the cells from each well were counted using TruCOUNT reference beads (BD Biosciences) or stained with a PE-conjugated mouse anti-human Ki67 monoclonal antibody (BD Biosciences) and analyzed by flow cytometry. Debris and dead cells were excluded from the analysis using forward and orthogonal light scatter criterias. All the samples were processed using an analytical flow cytometer (BD LSR II, BD Biosciences, San Jose, CA, USA) and data analyzed with FlowJo software (Tree Star Inc., Asland, OR, USA).

2.12.3.3 Colony staining

For each tertiary clone present on a 96-well normalized plate, ~15% of the cells dissociated in each well were used to seed a gelatinized 96-well plate. Cells were maintained in presence of LIF for one day. Media was removed and staining was directly performed in the plates with 100ul of methylene blue solution per well (0.3% methylene blue in methanol, Sigma). After 10 minutes incubation at room temperature, plates were washed in water and dried. Scoring was done manually using an inverted microscope (score: 0-no colonie to 5-high cell density) or by automated microscopy (colony number, average size of colonies, and surface of the well covered with cells) using Northern Eclipse software (EMPIX Imaging Inc.).

2.12.4 PCR studies

2.12.4.1 Real-time quantitative PCR (Q-PCR) screening of presumptive trisomies

Gene copy number was determined using primer and probe sets from Universal ProbeLibrary (Exiqon TaqMan probes, Roche Diagnostics) described in Supplemental Table 6. Gene copy number was evaluated for one gene per chromosome (e.g. chromosome 1, 8, 11, and 14) and normalized to a gene present on chromosome 3, since this chromosome is not frequently involved in mouse ESC trisomy [2,3]. PCR reactions for 384-well plate formats were performed using 2 µl of DNA sample (50 ng), 5 µl of the TaqMan PCR Master Mix (Applied Biosystems, CA), 2 µM of each primer and 1 µM of the Universal TaqMan probe in a total volume of 10 µl. The ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) was used to detect the amplification level and was programmed to an initial step of 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. All reactions were run in triplicate and the average values were used for quantification. A standard curve was generated for each assay (absolute quantification). Normalization was done according to chromosome 3 assay. A ratio of 1.5 (test versus chromosome 3 control) was indicative of a potential trisomy.

2.12.4.2 Q-PCR detection of hygromycin resistance gene

For detection of the hygromycin resistance gene in gDNA of puro^S tertiary clones, duplex Q-PCR reactions were performed with differentially labeled TaqMan probes specific to the hygromycin and neomycin (neo, reference) resistance genes. Threshold cycle (Ct) values for hygro (Ct_{hygro}) and neo (Ct_{neo}) were compared for each clone ($\Delta Ct = Ct_{hygro} - Ct_{neo}$) and represented by a score from 1 to 5: score 1, $\Delta Ct \leq -2$; score 2, $\Delta Ct = -2$ to 2; score 3, $\Delta Ct = 2$ to 5; score 4, $\Delta Ct = 5$ to 10; and score 5, $\Delta Ct > 10$. Puro^S tertiary clones associated to scores 1-3 or 4-5 were considered

positive or negative for hygromycin gene, respectively.

2.12.4.3 Mapping of proviral integrations

Integration sites of the anchor virus were mapped in primary clones by I-PCR or LM-PCR. Saturation virus integration sites were mapped in tertiary clones by LM-PCR. The I-PCR approach was previously described [2]. The LM-PCR strategy, which relies on specific oligonucleotides described in Supplemental Table 6, was adapted from a published protocol summarized here [4]. Briefly, 1 ug of gDNA was linearized with *Nla*III or *Tsp*5091 (New England Biolabs) and these restriction enzymes were subsequently heat inactivated. Splinkerette adapters were made by annealing 50mM of complementary oligonucleotides together (5 minutes thermal denaturation followed by gradual cooling). Seventy-five picomoles of splinkerette adapter were ligated to 250ng linearized DNA with 20U T4 DNA ligase (Invitrogen) in a final volume of 20ml at 16°C overnight followed by heat inactivation. Five microliters of ligation product was used as a template in the first round of PCR, while the second and third rounds of PCR amplified 2ul of previous PCR product diluted 1/50. PCR reactions contained 200mM each dNTPs, 400mM each primer, 2mM MgCl₂, 1X reaction buffer, and 2,5U of Taq DNA polymerase (Invitrogen), in a total volume of 50ml. PCR was performed using a GeneAmp 9700 thermocycler (Perkin Elmer) with the following program: 94°C for 5min; 94°C for 30sec, 56°C for 30sec, 72°C for 60sec (25 cycles); 72°C for 5min; 4°C hold. Third round PCR DNA products were excised from 2% agarose gel, purified with Qiaquick gel extraction kit (Qiagen), and directly sequenced with AGTCCTCCGATAGACTGCG (3'LTR) or CCAAACCTACAGGTGGGGTC (5'LTR) primer. Vector sequence contaminations were removed with Cross_match (Laboratory of Phil Green, University of Washington, <http://www.phrap.org/>) and sequences mapped using the BLAT alignment tool of the UCSC Genome Browser (<http://genome.ucsc.edu/>, NCBI mouse Build 37) [5].

2.12.4.4 Q-RT-PCR

Two micrograms of ESC or EB RNA was reverse transcribed using Superscript II Reverse Transcriptase and random hexamers, according to the manufacturer's guidelines (Invitrogen). Primer and probe sets for target genes (Exiqon probes, Roche Diagnostics) were designed using the Roche Universal ProbeLibrary assay design software (Advanced primer3 settings) (Tableau S2.6). Gene expression assays were tested for maximum efficiency by standard curve analysis (slope = 3.1 - 3.6) (Applied Biosystems). Reference gene assays (*Gapdh* and *β-actin*) were purchased from ABI (20X primer-probe mix, VIC labeled). Triplicate reactions were

prepared in 384-well plates (5-25 ng of cDNA per reaction) and monitored with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) (see *Real-time quantitative PCR (Q-PCR) screening of presumptive trisomies* section for PCR mix and amplification parameters). Relative quantification of the target genes was calculated with respect to the average level of two calibrator genes, *Gapdh* and β -*actin*, and normalized against the corresponding mock-transfected primary clone, using the delta-delta Ct method.

2.12.5 DeLES database functionalities

The *Home* tab provides information on the database, supporting grants and Figures detailing the experimental procedures. The *Plates* tab (restricted access) provides information on all the biological material (Cells, DNA, RNA) and their storage location. The *Families* tab has been described in Figure S2.2. The *Screen* tab presents a chromosome view of scored phenotypes and complementation information when available. The *Selection* tab is the main data-mining tool of the web interface and allows users to select DeLES families based on a series of predefined parameters such as the presence of clones presenting specific phenotypes, deletions spanning microRNAs, and others. Finally, the *Statistics* tab provides an overview of the percentage of the genome and individual chromosomes covered by the mapped deletions. Some specific information on the library such as the number of primary clones or the number of unique mapped deletions is also present. The most common points of entry of the database are the Karoyview, the Selection tool and the Search box. The Search box, present on most pages of the site, allows users to search the contents of DeLES by inputting keywords. Users can initiate a search using clone or plate names or input a gene symbol and find any deletions covering said gene. Since DeLES uses a local installation of Ensembl to manage the annotation tracks of the *Families* tab's genome browser version, users can search DeLES using any gene symbol or synonym known to Ensembl Genome Browser [6] (<http://www.ensembl.org/index.html>).

2.12.6 SelectaBAC modification strategy

Two homology arms specific to the chloramphenicol resistance gene were generated by PCR and subcloned in pDrive vector (Qiagen), flanking the bacterial kanamycin resistance gene from the vector and an eukaryote resistance marker gene generated by PCR (puromycin, or hygromycin, or blasticidin, or zeocin) regulated by a *Pgk* promoter and a SV40 mRNA polyadenylation signal. The engineered cassettes were then sub-cloned into the pLD53 vector by *Sst*I and *Mlu*I (pDrive)/*Asc*I (pLD53) digestion, thus creating the pSelectaBAC-Puro, -Hygro, -Blast, -Zeo retrofitting vectors, which were maintained in Pir2 E.coli (Invitrogen). The linear targeting

cassettes were excised from the plasmid backbone by *Bam*HI + *Sst*I digestion and gel purified for retrofitting. DH10B E.coli containing BACs of interest were made electro-competent by washing fresh mid-log culture twice in ice-cold water and were transformed with the temperature sensitive pRedET plasmid (GeneBridges) and maintained at 30°C with 3mg/ml tetracycline, thus rendering them recombination proficient. The presence of pRedET in the BAC clones was confirmed by plasmid mini-prep and restriction digests. BAC clones containing pRedET were grown at 30°C to mid-log, recombination proteins were induced in with 0.33% L-arabinose at 37°C for 45 minutes. The induced cells were made electro-competent as before and transformed with the linear targeting cassette. Retrofitted BACs were selected on kanamycin (25mg/ml) plates at 37 °C. Preliminary identification of kanamycin resistant bacterial clones containing modified BACs was performed by submitting each clone to chloramphenicol and ampicillin sib-selection in order to exclude clones containing unmodified BACs (chloramphenicol resistance) and/or persistent targeting vector (ampicillin resistance), respectively. As expected, none of the colonies exhibited resistance to ampicillin, confirming that residual pLD53-based targeting vector was unable to replicate in the BAC host strain. Approximately 65% of kanamycin resistant colonies also exhibited resistance to chloramphenicol (1/10 to 5/10 colonies checked, n=42 experiments). Dual resistant bacterial clones were in fact mixed colonies, possibly due to late recombination events, since single kanamycin or chloramphenicol resistant sub-clones can be isolated from them. Homologous recombination events were confirmed by PCR using primers spanning the homology region and/or by Southern blot using an *Eco*RI enzymatic digestion and an external probe hybridizing to the *Sac*B gene of the BAC vector. Southern blot analysis demonstrated homologous recombination events in all of the kanamycin resistant colonies tested. The absence of non-homologous recombination events was confirmed by comparative DNA fingerprinting of retrofitted and unmodified BACs (*Eco*RI restriction digests). BAC DNA for restriction digests, PCR and Southern blots was prepared with the modified alkaline lysis mini-prep protocol described by the BACPAC Resource Center (<http://bacpac.chori.org/>). BAC DNA for transfection into ESCs was prepared by modified alkaline lysis followed by purification with NucleoBond AX 500 Tips (Clontech), as described by the University of Michigan Transgenic Animal Model Core [7].

2.13 Supplemental References

1. Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC (1993) Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci USA* 90: 8424-8428.
2. Bilodeau M, Girard S, Hebert J, Sauvageau G (2007) A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Methods* 4: 263-268.
3. Liu X, Wu H, Loring J, Hormuzdi S, Distèche CM, et al. (1997) Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Dev Dyn* 209: 85-91.
4. Mikkers H, Allen J, Knipscheer P, Romeijn L, Hart A, et al. (2002) High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nat Genet* 32: 153-159.
5. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12: 656-664.
6. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, et al. (2009) Ensembl 2009. *Nucleic Acids Res* 37: D690-697.
7. Van Keuren ML, Gavrilina GB, Filipiak WE, Zeidler MG, Saunders TL (2009) Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res* 18: 769-785.

CHAPITRE 3: Haploinsufficiency screen highlights two distinct groups of ribosomal protein genes essential for embryonic stem cell fate

Authors: Simon Fortier^a, Tara MacRae^a, Mélanie Bilodeau^{a,b}, Tobias Sargeant^{a,c}, Guy Sauvageau^{a*}

Affiliations:

^a Molecular Genetics of Stem Cells Laboratory, Institute for Research in Immunology and Cancer (IRIC), University of Montreal, Montreal, Quebec, Canada.

^b Program in Developmental and Stem Cell Biology, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, Ontario, M5G 0A4, Canada.

^c Division of Molecular Medicine, The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia. Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3010, Australia.

*Correspondence to: Guy Sauvageau

Publié dans la revue PNAS.

Proceedings of the National Academy of Sciences, 2015 vol.112 (7) pp.2127-2132;

Published ahead of print February 2, 2015;

doi:10.1073/pnas.1418845112

3.1 Contribution des auteurs

Conceived and designed the experiments: SF MB GS.

Performed the experiments: SF TM.

Analyzed the data: SF TS GS.

Wrote the paper: SF TM GS.

De façon plus détaillée, ma contribution à cet article se décrit comme suit:

J'ai conçu et réalisé toutes les expériences, avec l'aide de Tara MacRae lorsque nécessaire (qRT-PCR, modification des BACs et fractionnements polysomaux). J'ai analysé tous les résultats (avec l'aide de TS pour l'analyse de termes GO, section 3.4.1) et produit toutes les figures. J'ai aussi écrit le manuscrit en entier, sous la supervision de Guy Sauvageau. Mélanie Bilodeau est l'instigatrice du projet et elle a aussi lu et commenté le manuscrit.

3.2 Abstract

In a functional genomics screen of mouse embryonic stem cells (ESCs) with nested hemizygous chromosomal deletions, we reveal that ribosomal protein (RP) genes are the most significant haploinsufficient determinants for embryoid body (EB) formation. Hemizygosity for three RP genes (*Rps5*, *Rps14*, or *Rps28*), distinguished by the proximity of their corresponding protein to the ribosome's mRNA exit site, is associated with the most profound phenotype. This EB phenotype was fully rescued by BAC or cDNA complementation but not by the reduction of p53 levels, although such reduction was effective with most other RP-deleted clones corresponding to non-mRNA exit-site proteins. RNA-sequencing studies further revealed that undifferentiated ESCs hemizygous for *Rps5* showed reduced expression levels of several mesoderm-specific genes as compared with wild-type counterparts. Together, these results reveal that RP gene dosage limits the differentiation, not the self-renewal, of mouse ESCs. They also highlight two separate mechanisms underlying this process, one of which is p53 independent.

3.3 Introduction

Embryonic stem cell (ESC) fate is intricately controlled at the transcriptional level by factors such as Oct4 (Pou5f1), Sox2 and Nanog (1–3), Dax1, Rex1, and SalI (reviewed in ref. 4), and the Mediator complex (3, 5, 6). Although much progress has been made on the identification of the transcriptional circuitry that defines pluripotency, the regulation of mRNA translation and posttranslational protein modifications currently are considered in this context only as fine-tuning to balance protein production and activity (7).

ESC self-renewal can be assessed readily by their ability to give rise to undifferentiated progeny in a clonal fashion. Similarly, in the absence of leukemia inhibitory factor (LIF) and bone morphogenetic protein signaling, ESC pluripotent potential can be investigated in vitro by initiating differentiation of the three germ layers using an embryoid body (EB) formation assay (8, 9).

To identify novel regulators of ESC fate, we developed a retrovirus-based methodology that randomly generates hemizygous chromosomal deletions with increasing size (nested) (Fig. 3.1A and detailed in ref. 10). Using this method, we reported the generation of a library of 1,307 ESC clones covering more than 25% of the mouse genome (the DeIES library) along with an improved methodology for BAC recombineering and complementation (11). Phenotypical characteristics and a genetic map of all clones in our DeIES collection are available at www.bioinfo.irc.ca/deles.

Here, we report the integrative analysis linking the physical mapping to phenotype for all 1,307 clones in our library, documenting that genes involved in ribosome biogenesis and function are key players in ESC fate. We tested this hypothesis extensively in functional studies and identified a p53-independent response to ribosomal stress for mRNA exit-site ribosomal proteins (RPs).

3.4 Results

3.4.1 Gene Ontology Term Analysis Identifies RP Genes as Prime Candidates for EB Formation.

Using our collection of 1,307 previously described ESC clones, we first tested the contribution to overall fitness of the hemizygous deletion of gene sets extracted from the Gene Ontology (GO) database to determine underlying biological processes with the most reproducible and extreme effect on embryoid body (EB) development. GO term clustering ($n = 7,083$ genes out of 1,307 clones) showed that deletion of genes linked with the small ribosomal subunit ($n = 10$ genes) strongly correlated with a negative EB-formation outcome. Consistent with this finding, other categories such as rRNA processing ($n = 27$ genes) and cytosolic small ribosomal subunit ($n = 12$ genes) also are found in the 10 most significant categories of genes negatively affecting EB formation ($P < 0.0005$; red dots in Fig. 3.1B and Tableau S3.1). Within our collection, RP gene deletions are present in 16 families (Fig. 3.1C and Tableau S3.2), representing members of both the small (Rps, $n = 12$) and large (Rpl, $n = 12$) ribosome subunits (Fig. 3.1D). Because the DeLES collection is made up of clones with hemizygous deletions, the abnormal EB-formation phenotype shared by this group of clones suggests that although RP gene dosage has no clear impact on self-renewing ESCs, it appears to be crucial for ESC differentiation.

To confirm this observation, we randomly selected 112 clones from DeLES and validated their phenotype in standard in vitro assays (60-mm dish format). These validation experiments revealed that clones with the hemizygous deletion of at least one RP gene invariably exhibit defects in EB formation (Fig. 3.1E). Fig. 3.1F shows that the median relative EB number is 0.48 for non-RP-deletion clones versus 0.055 for RP-deletion clones ($P < 0.0001$). Importantly, no significant correlation was found between EB formation and chromosome deletion size (Fig. S3.1).

3.4.2 Physiological Consequence of RP Haploinsufficiency

Under self-renewal conditions, wild-type and RP-hemizygous ESCs (hereafter called “ Δ RP-ESCs”) could be maintained undifferentiated and were undistinguishable based on colony morphology and proliferation and apoptosis rates. Importantly, although cell-cycle progression was not affected (Fig. S3.2A), we detected an increased tendency of Δ RP-ESCs to undergo cell death following LIF removal (Fig. S3.2 B and C).

3.4.3 RP Genes Are Critical for EB Formation: Complementation Studies

To confirm the requirement of RP genes for EB formation, we performed BAC complementation as described (11) and using the experimental design detailed in Fig. 3.2A. A total of 17 BACs were modified to complement clones from 11 different families, each containing one distinct RP gene deletion, i.e., *Rps5*, *Rps12*, *Rps14*, *Rps16*, *Rps19*, *Rps28*, *Rpl12*, *Rpl22*, *Rpl27a*, *Rpl29*, and *Rpl39*. Of these 11 families, five (*Rps5*, *Rps14*, *Rps28*, *Rpl12*, and *Rpl27a*) were functionally rescued (indicated by asterisks in Fig. 3.2B and Fig. S3.3). Projection of these RP proteins onto a 3D ribosomal structure revealed a consistent localization around the mRNA exit site for the three identified Rps proteins (Fig. 3.2C, Right). Interestingly, these three clones, which showed the strongest EB formation defects (no background and early disaggregation), also exhibited the highest level of complementation (Fig. 3.2B), suggesting that RPs located in this region may be critical for EB formation.

Additional studies were performed to confirm that RP genes— not other elements— are responsible for this complementation. To do so, we focused our studies on the two best complemented families, namely those in which *Rps5* and *Rps28* are deleted [families 5066 and 5001 in Fig. 3.2B; *Rps14* haploinsufficiency for EB formation already has been described with the characterization of DeLES library (11)]. Family 5066 (*Rps5* heterozygote) was complemented only by the two BACs that included the *Rps5* gene (BAC coverage is indicated by red lines in Fig. 3.3A; complementation results are shown in Fig. 3.3B). Importantly, a modified version of BAC RP23-389P15 engineered to lack the coding region of *Rps5* and all other BACs lacking this gene failed to complement (Fig. 3.3B). In line with these results, transfection of a *Rps5* cDNA expression vector effectively rescued EB formation to levels observed in primary clones (those containing no deletion but only an anchor virus; Fig. 3.3B). Similar data were obtained for clone 5001 (which has a large hemizygous deletion including the *Rps28* gene), in which BAC complementation was dependent on the presence of *Rps28*, and cDNA complementation was sufficient to rescue EB formation to levels seen in undeleted cells (Fig. 3.3C-D).

3.4.4 p53-Independent Ribosomal Stress in a Subset of RP-Deleted Clones

RP imbalance affects ribosome biogenesis, leading to human diseases commonly referred to as “ribosomopathies” (reviewed in ref. 12). These syndromes, often causing hematological and skeletal anomalies, are characterized by ribosomal or nucleolar stress, a condition most frequently associated with p53 activation. Mechanistically, a subset of free RPs from destabilized ribosomal subunits physically associates with mouse double minute 2 (Mdm2), thus preventing

p53 degradation (summarized in Fig. 3.4A) (13). This p53 response has been hypothesized to represent the cornerstone of ribosomopathies, given that many symptoms of these conditions can be corrected by a simple attenuation of p53 activity (12, 14, 15).

Using an shRNA vector against p53, we first noted a twofold increase in EB numbers in wild-type R1 ESCs compared with controls (black bar in Fig. 3.4B, Right), likely indicating that p53 normally limits EB formation. Reduction of p53 levels in “non– exit-site” *Rpl22*^{Δ/wt} and *Rpl27*^{Δ/wt} ESC clones produced a complete rescue of EB formation (Fig. 3.4B, Right). In contrast, treatment with p53 shRNA was unable to complement “exit-site” clones hemizygous for *Rps5*^{Δ/wt}, *Rps14*^{Δ/wt}, or *Rps28*^{Δ/wt} (Fig. 3.4B, Right), despite evidence of ribosomal stress (reduced 40S subunit abundance and/or increased p53 protein levels) (Fig. S3.4A) and successful p53 protein reduction upon shRNA expression (Fig. S3.4 B and C). However, p53 levels were not significantly higher in *Rps5*^{Δ/wt} than in control cells, although polysome profiles clearly showed decreased levels of the 40S subunit. Interestingly, although not localized at the mRNA exit site, *Rpl12*^{Δ/wt} ESCs were not rescued by reduction in p53 levels.

These studies thus reveal two different types of complementation, one that involves attenuation of p53 activity and one that does not (see Fig. S3. 4D for all clones studied herein). Indeed, p53 shRNA treatment and BAC complementation were comparable in rescuing EB formation in clones hemizygous for non–exit-site RP genes (Fig. 3.4C, Right). In contrast, none of the three mRNA exit-site RP-deletion clones could be rescued by p53 shRNA (blue triangles in in Fig. 3.4C), although BAC complementation (green squares in Fig. 3.4C) or cDNA efficiently rescued these clones. To gain further insights into the nature of this phenotype, we isolated three subclones from the p53 shRNA-transduced *Rps5*^{Δ/wt} ESC clone and reintroduced *Rps5* cDNA separately in each clone. Again, and confirming the results shown in Fig. 3.4C, cDNA transfection was very effective even when p53 levels were experimentally decreased (Fig. 3.4D).

3.4.5 Expression of Mesoderm Genes Is Impaired in *Rps5*^{Δ/wt} Mutants

Next, we analyzed the polyribosome profiles of wild-type ESCs in self-renewing and differentiating conditions. Using area under the curve (AUC) measurement (Fig. S3.5A) (7), we observed a 2.5-fold increase in polyribosome fractions with the onset of EB formation. (Ribosome profiles are shown in Fig. 3.5A, and a summary is given in Fig. 3.5B.) In contrast to this polyribosome response, the 80S fraction was constant between the two states (Fig. 3.5).

Although self-renewing Δ RP-ESCs showed nearly normal polysome profiles, their polyribosome response was almost absent upon LIF removal (Fig. 3.5A; quantification in Fig. 3.5B). As

expected of a ribosomal stress condition, the small and large ribosome subunits, and hence the mature 80S ribosome fractions, often were perturbed in these cells (Fig. S3.5B).

To gain further insights into the impact of this failure in polyribosome response, we performed RNA sequencing (RNAseq) using mRNA isolated from different ribosome fractions from wild-type and *Rps5*^{Δ/wt} cells, in both self-renewing (ESCs) and differentiation onset (EBs, 48h) conditions (Fig. 3.6A), before the appearance of disaggregation to avoid potential contamination by cell death signals. Analyses conducted from all EB fractions (total mRNA and monosomal or polysomal fractions) showed a significant enrichment of genes linked to mesodermal lineage specification in wild-type clones as compared with clones with *Rps5* hemizygous deletion (Fig. 3.6B; see also the network representation in Fig. S3.6A, the comprehensive list of all differentially expressed genes in Tableau S3.3, and GO term clustering in Tableau S3.4). Interestingly, similar results were obtained when we compared the read counts of undifferentiated samples (see the representative DESeq graph in Fig. S3.6B), suggesting that a defect in the expression of mesodermal-specific genes already is established in self-renewing *Rps5*^{Δ/wt} ESCs, even before LIF removal and EB formation.

We next compared differentially expressed genes from mono- somal and polysomal fractions of the *Rps5* hemizygous clone which undergo EB formation. We found that the expression level of genes deleted in these cells, namely *Rps5*, *Zfp110*, and *Zscan22*, was reduced by approximately twofold in all fractions examined (Fig. 3.6C; see Fig. S3.7 A–C for fraction-specific enrichment curves). As expected, the expression level of genes outside the deleted region, *Zfp329* and *Zfp324*, was unchanged (Fig. 3.6C). We also found that very few genes are enriched in wild-type and *Rps5*^{Δ/wt} clones. Apart from genes related to the previously identified mesodermal differentiation network, no significant differentially expressed candidate could be observed in monosomal fractions. However, in polysomal fractions, two noncoding genes, RNA component of mitochondrial RNA processing (*Rmrp*) and ribonuclease P RNA component H1 (*Rpph1*), are the most enriched genes, having reads per kilobase of transcript per million reads mapped (RPKM) ratios of 209.4 and 45.2, respectively (Fig. 3.6C). Reinforcing this observation, we observed that ESCs transduced with *Rps5* expression vector (Fig. 3.3B) showed increased levels of *Rmrp* and *Rpph1* by RNAseq [38.2- and 10.4-fold, respectively; see Fig. 3.6D and Tableau S3.5 for confirmation by quantitative RT-PCR (qRT-PCR) studies].

3.5 Discussion

The idea that RP-coding genes could have more than a merely structural function has evolved recently. Several RP-coding genes were shown to be essential in a species- and tissue-specific fashion, and mutations of these genes have been identified as genetic markers of human malignancies (12, 16). In the present study, we report a critical role for RP genes in EB formation from mouse ESCs. Of interest, this approach allowed us to highlight fundamental differences by which self-renewing pluripotent ESCs are less sensitive to nucleolar stress (RP hemizygous) than their differentiation-engaged progenies. Complementation experiments revealed two subsets of RPs regulating EB formation, one of which acts independently of p53. Of interest, these RPs are localized to the mRNA exit site, and their reduction conferred the most severe phenotypes.

Rmrp and *Rpph1* were identified as depleted in the *Rps5*^{Δ/wt} polysomal fraction. Of interest, these two genes are the most positively correlated ones (based on Pearson's correlation coefficient derived from expression values) in all publicly available RNAseq datasets analyzed to date (Fig. S3.7D). Although currently there is no known link between these two noncoding transcripts, their high degree of correlation suggests their participation in a network that might be involved in mesodermal differentiation, given that *Rmrp* mutations in humans give rise to dysplasia-related syndromes such as cartilage hair hypoplasia syndrome (17). Future experiments will be needed to determine the nature of this *Rmrp*–*Rpph1* coexpression and their regulation by *Rps5*.

The transition of ESCs to EBs is not a lineage-specific assay and thus allows the generation of the three germ-layer derivatives (8, 9). Therefore, it is difficult to reconcile the “mesodermal signature” observed in *Rps5*^{Δ/wt} ESCs and the severe EB phenotype observed in these clones. Previous studies have shown hematopoietic defects, but not defects in EB formation, upon deletion of mesodermal genes (18–20). However, it is possible that the combined deficiencies of several mesodermal genes as observed in *Rps5*^{Δ/wt} ESCs contribute to this severe EB phenotype. Further studies are needed to clarify these observations, but the early onset of this phenotype makes it difficult to set a directed differentiation approach, because those techniques usually require EB formation before cytokine induction.

The ongoing cell differentiation occurring within EBs is a highly demanding metabolic process and most likely triggers an important increase in protein synthesis, which could be impossible to achieve in RP hemizygous cells. Along these lines, we observed that ribosomal loading (AUC of polyribosomal profiles; Fig. 3.5) was higher in differentiating than in self-renewing wild-type ESCs. Sampath et al. (7) also observed this “polyribosome response” in EBs and further demonstrated an enhanced protein production in these cells compared with ESCs.

It is well known that RP gene imbalance activates p53 signaling. qRT-PCR analysis of a panel of p53 target genes showed no significant expression changes in the context of RP hemizygous deletions. Imbalance in RP-coding genes also affects other cellular functions, such as transcription (21, 22). In line with this finding, our results document a p53-independent function for at least a subset of RP genes, those associated with the ribosome mRNA exit site.

The essential role of RP-coding genes reported here highlights fundamental differences between self-renewing ESCs and their differentiated counterparts, as well as a unique p53-independent role in this transition step for the mRNA exit-site *Rps5*, *Rps14*, and *Rps28* genes. To our knowledge, this is the first time that a p53-independent response to nucleolar stress has been identified in a model that assesses self-renewal and EB formation from pluripotent ESCs.

3.6 Figures

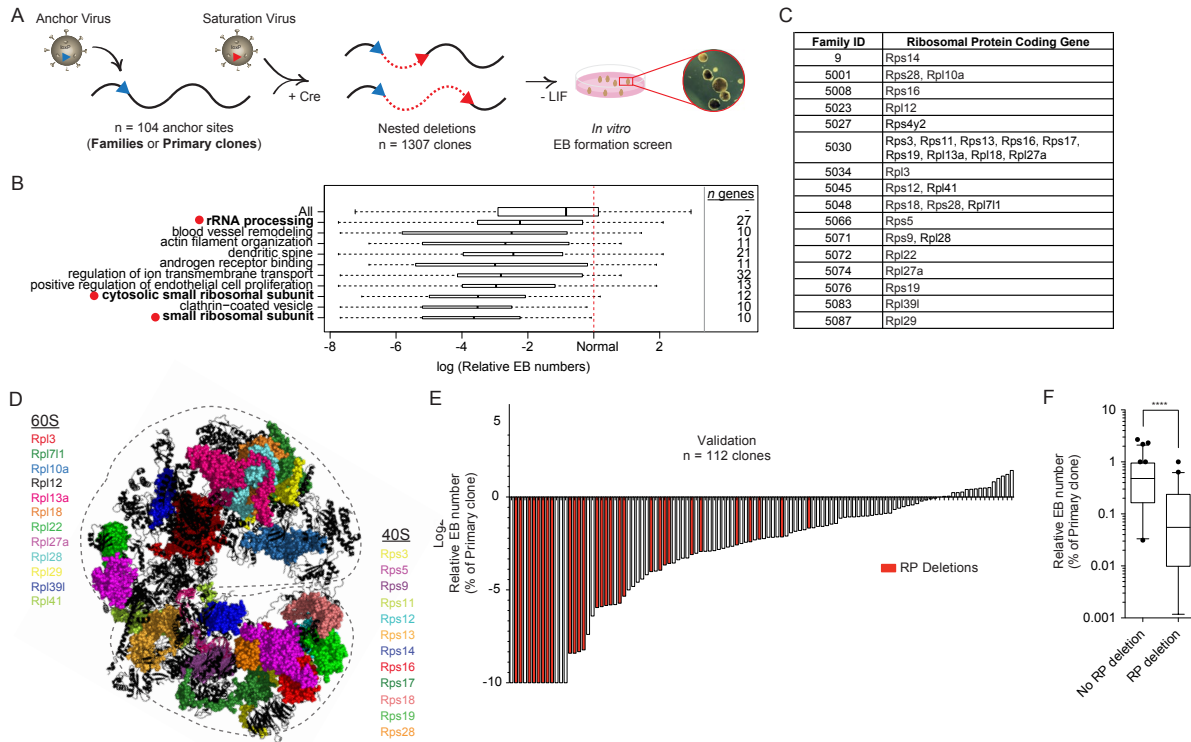


Figure 3.1: The EB-formation phenotype correlates with RP gene deletions.

(A) Schematic overview of DeES library generation. For more details, see ref. 11. (B) GO analysis of deleted regions with respect to EB-formation phenotypes obtained in our high-throughput screen. The 10 GO terms present in at least 20 deletions with the most significant ($P < 0.0005$) negative effect on EB formation are shown. Bar thickness indicates the square root of the number of clones that hit a given GO term. See Tableau S3.1 for complete analysis of GO-term clustering. Relative EB numbers are calculated by dividing EB numbers of hemizygous clones by those of primary (wild-type) clones. (C) Rps and Rpl genes deleted in the first 100 families of the DeES library. (D) Location of deleted RP-coding genes highlighted on the 80S eukaryotic ribosome 3D structure (PDB ID codes 3U5E, 3U5F, and 3U5G) (26). (E) Waterfall-plot representation of the validation experiment in which 112 randomly selected DeES ESC clones from 35 unique anchor points were tested for the EB-formation phenotype in 60-mm dishes. Bars with values of -10 represent absolute EB numbers of 0 and should be considered as a negative infinite. Red shading indicates clones with RP gene deletions. (F) Box plot of EB-formation validation experiments; error bars indicate 95% confidence intervals. **** $P < 0.0001$.

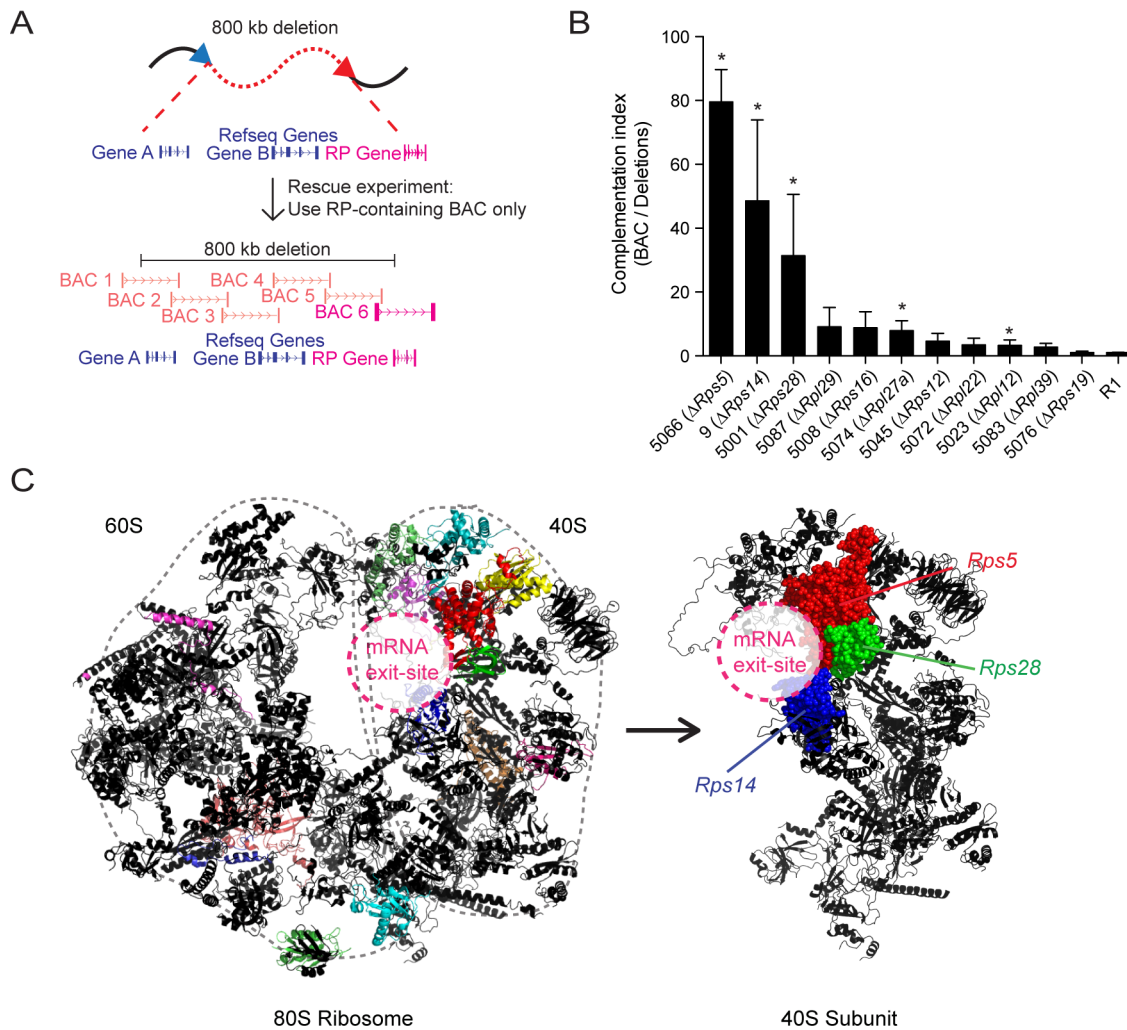


Figure 3.2: Complementation experiments reveal a site-specific role for RP genes.

(A) Design of BAC complementation experiments. The backbone of a BAC vector containing the affected RP gene was modified using puromycin-, hygromycin-, or zeocin-resistance genes and was transfected into DeIES clones of interest lacking one copy of the corresponding RP gene. (B) Complementation experiments of RP-deleted clones. Clones with a single RP gene deletion are shown. (See Fig. S3.3 for all values in the bar graph.) The complementation index was obtained by dividing relative EB numbers (based on primary clone values) of BAC-transfected clones by those of mock-transfected clones. Error bars show SEM of at least two independent experiments. (C) The RP-deleted clones that showed the best BAC complementation (farthest left columns in B) are colocalized at the mRNA exit site of the 40S ribosomal subunit (PDB ID codes 3U5E, 3U5F, and 3U5G) (26, 27).

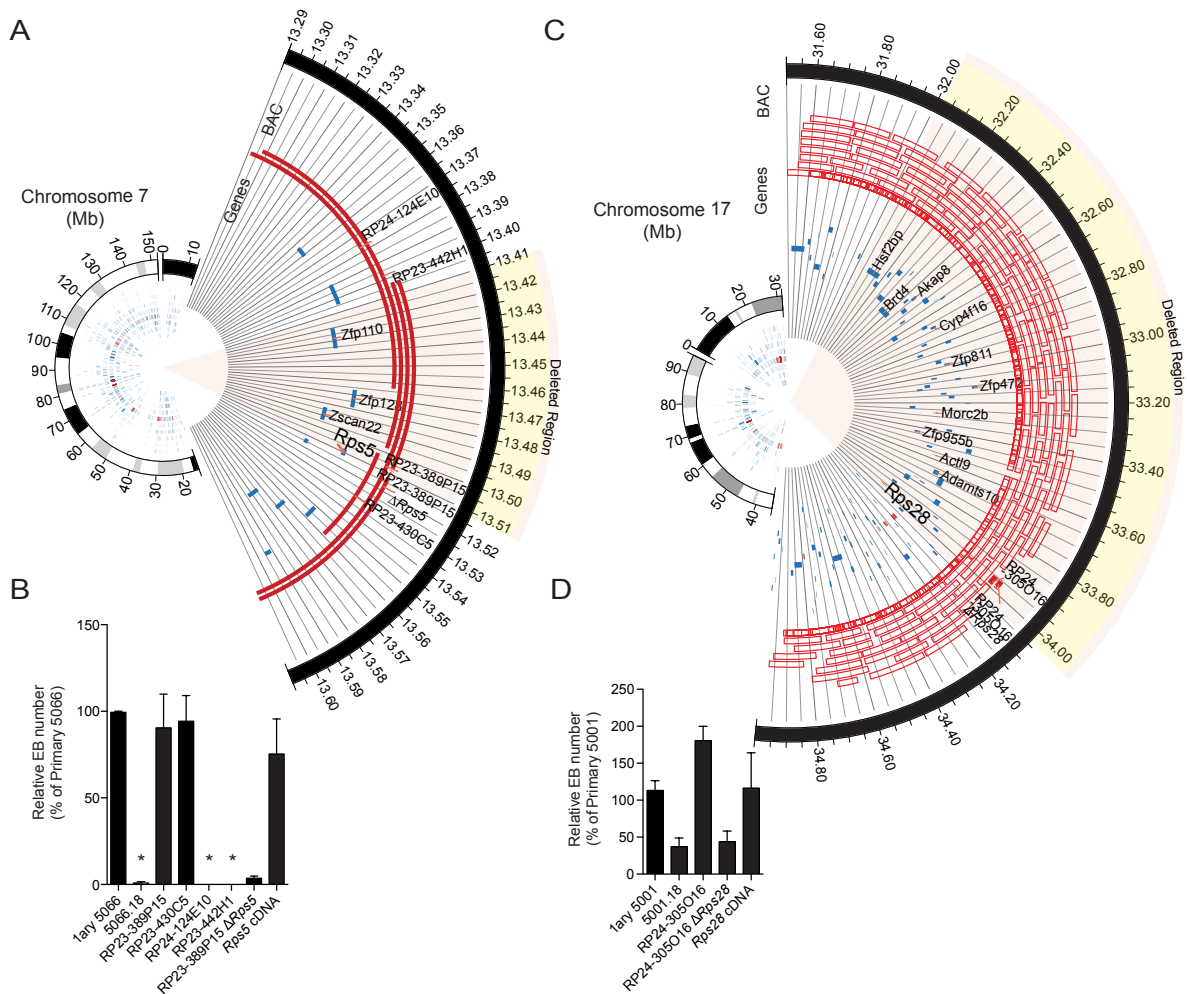


Figure 3.3: *Rps5* and *Rps28* are essential for EB formation.

(A) Circos (28) plots representing the genomic context of clone 5066.18 (*Rps5*^{Δ/wt}) ~90-kb deletion (mm9 mouse assembly). (B) Relative EB numbers (expressed as percentage of primary clone numbers) following transfection of BAC DNA identified in A. Error bars represent the SEM for at least two independent experiments performed in duplicate. Asterisks denote null values. (C) Circos plots representing the genomic context of clone 5001.18 (*Rps28*^{Δ/wt}) ~2Mb deletion (mm9 mouse assembly). Genes have been partially annotated to simplify visualization. (D) Relative EB numbers (expressed as percentage of primary clone numbers) following transfection of BAC DNA identified in C. Error bars represent the SEM for at least two independent experiments performed in duplicate.

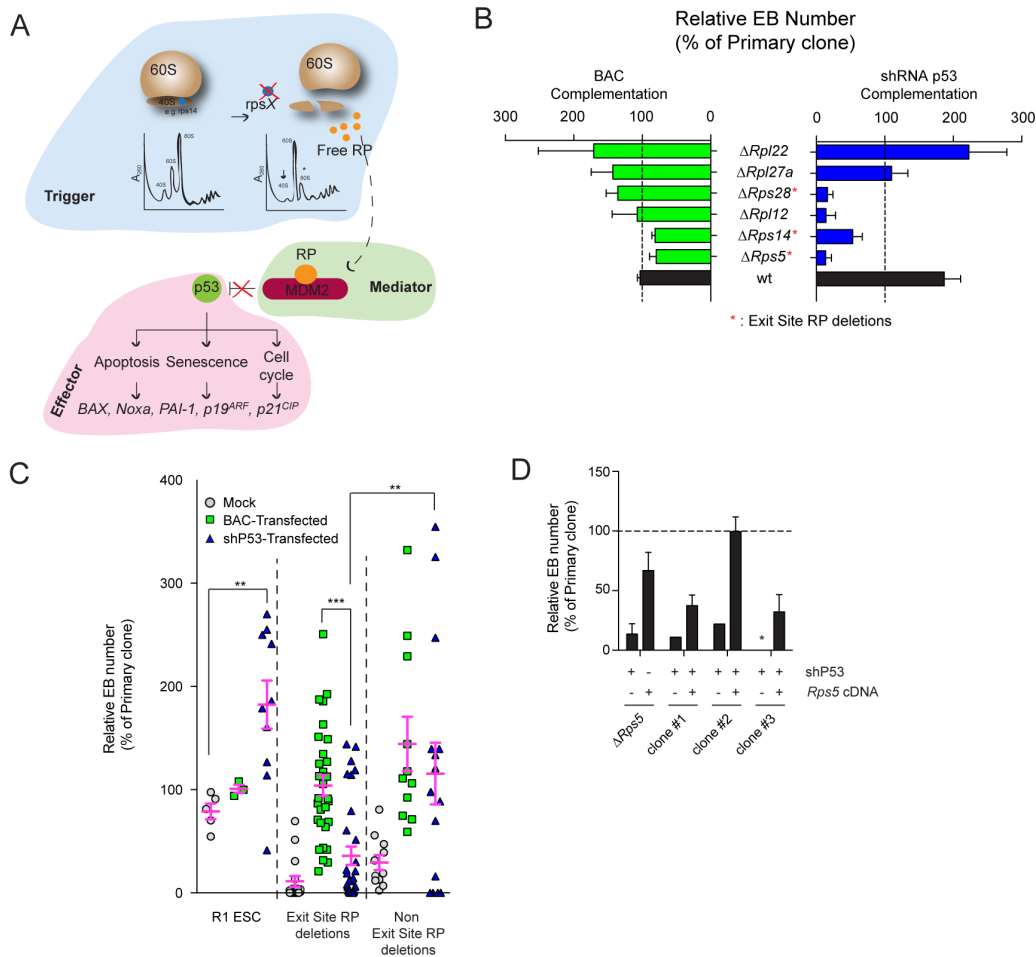


Figure 3.4: p53-independent role of exit-site specific RP deletions in EB formation.

(A) Schematic representation of ribosomal (nucleolar) stress pathway. A RP imbalance can lead to formation of specific RP-Mdm2 complexes, resulting in p53 stabilization and downstream signaling. (B) Relative EB numbers (expressed as percentage of Primary clone numbers) following BAC or shRNA p53 transfection. Results from clones where BAC complementation was at least 50% are shown, see Fig. S3.4C for results of all clones. Error bars show SEM of at least n=2 (BAC transfections) and n=4 (shp53 transfections) independent experiments performed in duplicate (C) Comparison of BAC vs shRNA p53 complementation in wild-type R1 ESC, exit-site specific deletions (n=3 clones) and other RP deletions (n=3 clones). Pink bars represent mean values +/- SEM. p53 knockdown levels can be found in Fig. S3.4A-B (D) Reintroduction of *Rps5* cDNA following unsuccessful shRNA p53 rescue experiments. Three clones refractory for shRNA p53 complementation where subsequently transfected with the *Rps5* cDNA to test for EB formation capability and p53-independency. Error bars show SEM of at least n=2 independent experiments performed in duplicates.

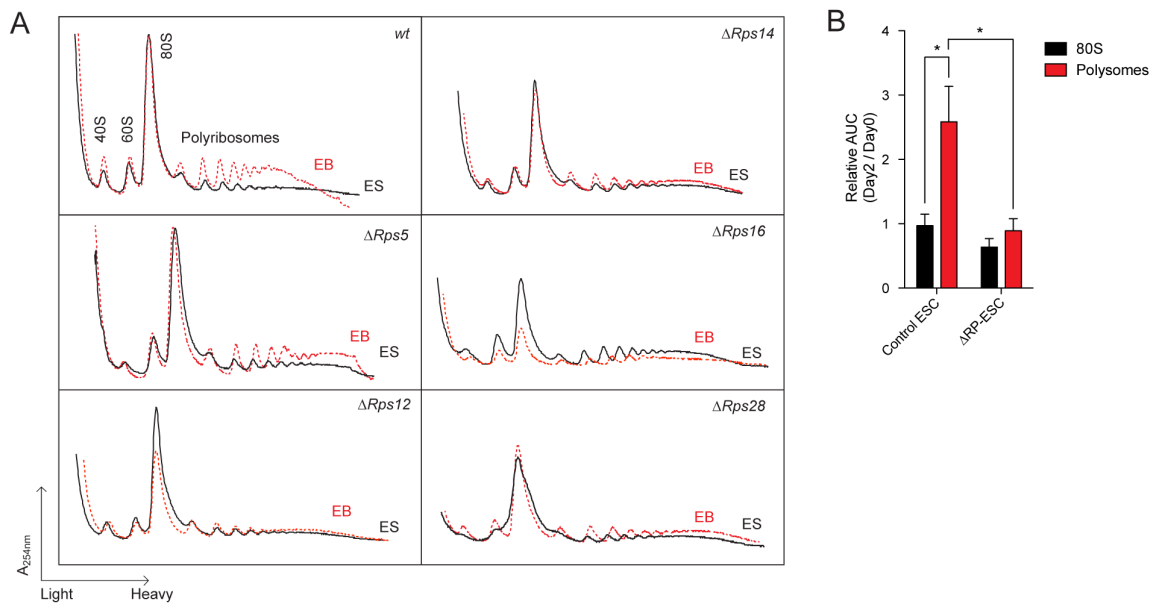


Figure 3.5: Polysomal AUC is increased upon EB formation.

(A) Polysome profiles obtained from wild-type (upper left) and Δ RP-ESC (ES, black lines) and embryoid bodies (EB, red dashed lines). Profiles are representative of at least two independent experiments. For a full description of the experimental setup, see Materials and Methods section. (B) Quantification of Area Under the Curve (AUC) of polysomal profiles. Values represent relative area units (day 2 / day 0). Error bars show SEM values of $n=6$ (wild-type (wt), various ESC lines) and $n=7$ (RP deletions). *, $p<0.05$. See Fig. S3.5A and Materials and Methods for a detailed AUC quantification methodology.

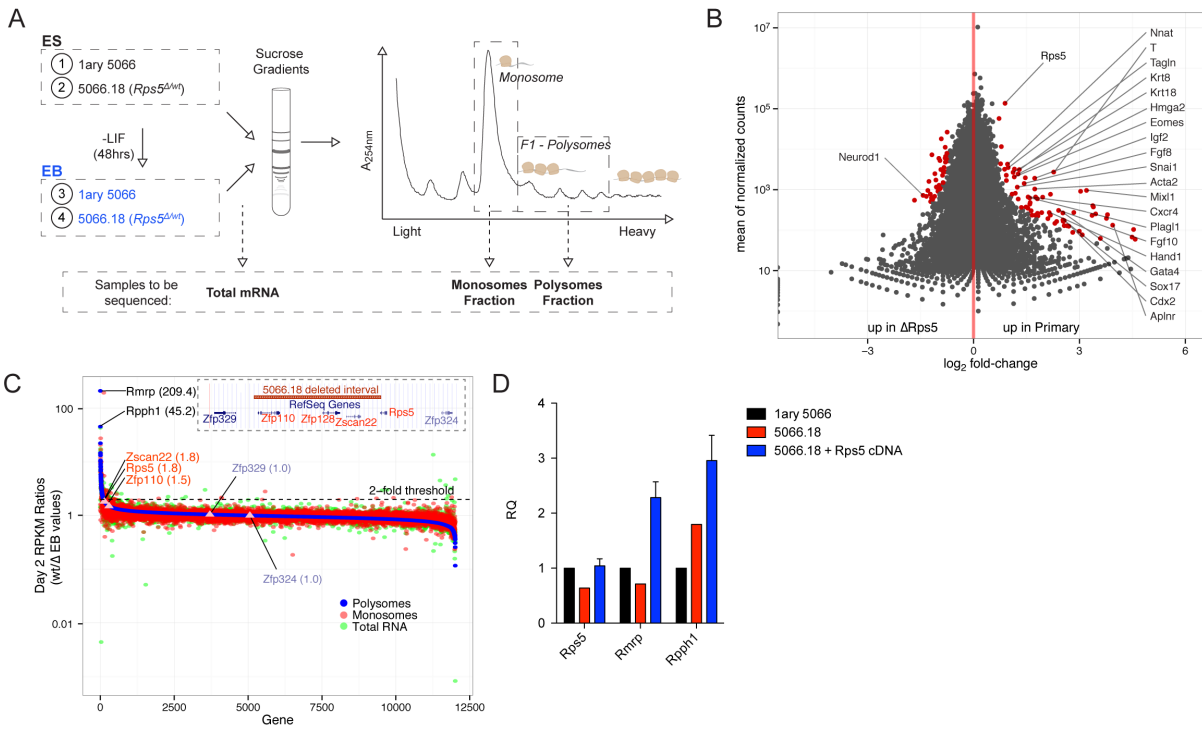


Figure 3.6: RNaseq analysis of ribosomal fractions.

(A) Experimental design. Wild-type (1ary 5066) and *Rps5*^{Δ/wt} lysates (undifferentiated ESC and EBs) were fractionated on sucrose gradients and RNA was extracted from monosome and polysome fractions. 12 samples were hybridized to 3 lanes of an Illumina array and subjected to next-generation sequencing. (B) Differential expression (DESeq package (31)) analysis obtained from EB polysomal fractions. Genes that are differentially expressed ($p_{adj} < 0.1$) are shown in red. Genes found in mesodermal differentiation network are labeled (Fig. S3.6). Full data table of the analysis is presented in Tableau S3.3. (C) Enrichment curve showing RPKM ratios (wild-type / *Rps5*^{Δ/wt} EB) of genes found in polysome (blue track), monosome (red track) and total RNA (green track). Genes showing the biggest differential expression based on RPKM values of polysomal fractions are labeled in black. Genes present in the deletion (red) and those located just outside of the deletion (blue) are also identified. (D) qRT-PCR assessment of relative *Rmp* and *Rpph1* expression (RQ) in wild-type, *Rps5*^{Δ/wt} and *Rps5* cDNA complemented ESC. Error bars are representative of $n = 8$ independent cDNA clones and are expressed as mean \pm SEM. RNA-Seq data are deposited on GEO: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=gzsxqqsgzngfroz&acc=GSE57648>

3.7 Materials and Methods

3.7.1 ESC Maintenance

Male R1 ES cells (23) were grown on irradiated mouse embryonic fibroblasts or on gelatin-coated dishes in DMEM medium (high glucose with L-glutamine and pyruvate) (Invitrogen) containing 15% (vol/vol) FCS (Invitrogen), 1.5×10^{-4} M α -monothioglycerol (Sigma), and 1×10^{-4} M nonessential amino acids (Invitrogen), supplemented with 1,000 U/mL of LIF-conditioned medium from transfected COS cells. The generation of the DeES library and clone maintenance have been described previously (11).

3.7.2 EB Formation

ESC differentiation into EBs was performed in LIF-deprived semisolid medium, as previously described (8). See SI Materials and Methods for more details.

3.7.3 Ribosome Profiling

Sucrose gradient velocity sedimentation was used to isolate polysomal fractions, as described (24). See SI Materials and Methods for more details.

3.7.4 GO Analysis

To determine common mechanisms and pathways underlying the measured EB-formation phenotypes, GO terms were tested for a significant association with EB formation. See SI Materials and Methods for a complete description.

3.7.5 RNAseq

RNAseq was performed on RNA samples purified from sucrose gradient fractions of 1.5 mL. Briefly, read count data were generated using the htseq-count module and were mapped to the mm9 .gtf file from the University of California, Santa Cruz. Reads mapping to introns were excluded from the analysis. The full methodology can be provided on request. Transcriptome sequencing was done as described for our previously reported T-ALL collection (25). Four samples were hybridized per lane of an Illumina HiSeq2000 (coverage of ~87–113 million reads per sample). RNAseq data have been deposited in the Gene Expression Omnibus database under accession no. GSE57648.

Transfections, flow cytometry, and complementary methods are described in SI Materials and Methods.

3.8 Acknowledgements

The authors thank Tatiana Traboulsi, who helped with qRT-PCR studies and cDNA generation as a summer student in our laboratory; Geneviève Boucher and Jean-Philippe Laverdure for RNAseq data processing and help with bioinformatic analyses; Joseph Tcherkezian for his expertise with polysome experiments; Raphaëlle Lambert and Marianne Arteau from IRIC genomic platform for their expertise with quantitative RT-PCR and RNAseq; and Dr Bernhard Lehnertz for scientific comments. This work was supported by the Canadian Institute of Health Research (CIHR) (Grant no. MOP 231107). G.S. holds a Canada Research Chair on molecular genetics of stem cells; Salary of S.F. was partially paid by a studentship from Canadian Institute for Health Research. TS contribution was made possible through Victorian State Government Operational Infrastructure Support, Australian Government NHMRC IRIISS and an NHMRC Program Grant (1016647).

3.9 References

1. Chambers I et al. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113:643–655.
2. Ivanova N et al. (2006) Dissecting self-renewal in stem cells with RNA interference. *Nature* 442:533–538.
3. Young RA (2011) Control of the Embryonic Stem Cell State. *Cell* 144:940–954.
4. Orkin SH, Hochedlinger K (2011) Chromatin connections to pluripotency and cellular reprogramming. *Cell* 145:835–850.
5. Kagey MH et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467:430–435.
6. Whyte WA et al. (2013) Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* 153:307–319.
7. Sampath P et al. (2008) A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* 2:448–460.
8. Keller G, Kennedy M, Papayannopoulou T, Wiles MV (1993) Hematopoietic commitment during embryonic stem cell differentiation in culture. *Molecular and Cellular Biology* 13:473–486.
9. Keller GM (1995) In vitro differentiation of embryonic stem cells. *Current Opinion in Cell Biology* 7:862–869.
10. Bilodeau M, Girard S, Hébert J, Sauvageau G (2007) A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Meth* 4:263–268.
11. Fortier S et al. (2010) Genome-wide interrogation of Mammalian stem cell fate determinants by nested chromosome deletions. *PLoS Genet* 6:e1001241.
12. Teng T, Thomas G, Mercer CA (2013) Growth control and ribosomopathies. *Current Opinion in Genetics & Development*:1–9.

13. Zhou X, Liao J-M, Liao W-J, Lu H (2012) Scission of the p53-MDM2 Loop by Ribosomal Proteins. *Genes & Cancer* 3:298–310.
14. Fumagalli S et al. (2009) Absence of nucleolar disruption after impairment of 40S ribosome biogenesis reveals an rpL11-translation-dependent mechanism of p53 induction. *Nature Cell Biology* 11:501–508.
15. McGowan KA et al. (2011) Reduced ribosomal protein gene dosage and p53 activation in low-risk myelodysplastic syndrome. *Blood* 118:3622–3633.
16. Narla A, Ebert BL (2010) Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115:3196–3205.
17. Ridanpää M et al. (2001) Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* 104:195–203.
18. Keller G (2005) Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes & Development* 19:1129–1155.
19. Lacaud G, Kouskoff V, Trumble A, Schwantz S, Keller G (2004) Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. *Blood* 103:886–889.
20. Fehling HJ et al. (2003) Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* 130:4217–4227.
21. Lindström MS (2009) Emerging functions of ribosomal proteins in gene-specific transcription and translation. *Biochemical and Biophysical Research Communications* 379:167–170.
22. Provost E et al. (2012) Ribosomal biogenesis genes play an essential and p53-independent role in zebrafish pancreas development. *Development* 139:3232–3241.
23. Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC (1993) Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci USA* 90:8424–8428.
24. Cargnello M et al. (2012) Phosphorylation of the eukaryotic translation initiation factor 4E-transporter (4E-T) by c-Jun N-terminal kinase promotes stress-dependent P-body

assembly. *Molecular and Cellular Biology* 32:4572–4584.

25. Simon C et al. (2012) A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes & Development* 26:651–656.
26. Ben-Shem A et al. (2011) The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* 334:1524–1529.
27. Ben-Shem A, Jenner L, Yusupova G, Yusupov M (2010) Crystal Structure of the Eukaryotic Ribosome. *Science* 330:1203–1209.
28. Krzywinski M et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.
29. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.

3.10 Supplemental Figures and Tables

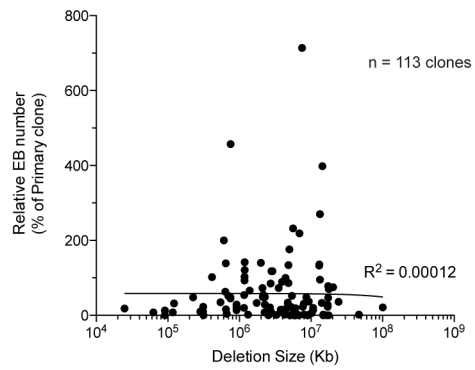


Figure S3.1: Correlation between deletion size and EB numbers.

Graphical representation of EB numbers compared with deletion size. The correlation coefficient shows there is no correlation between the two factors.

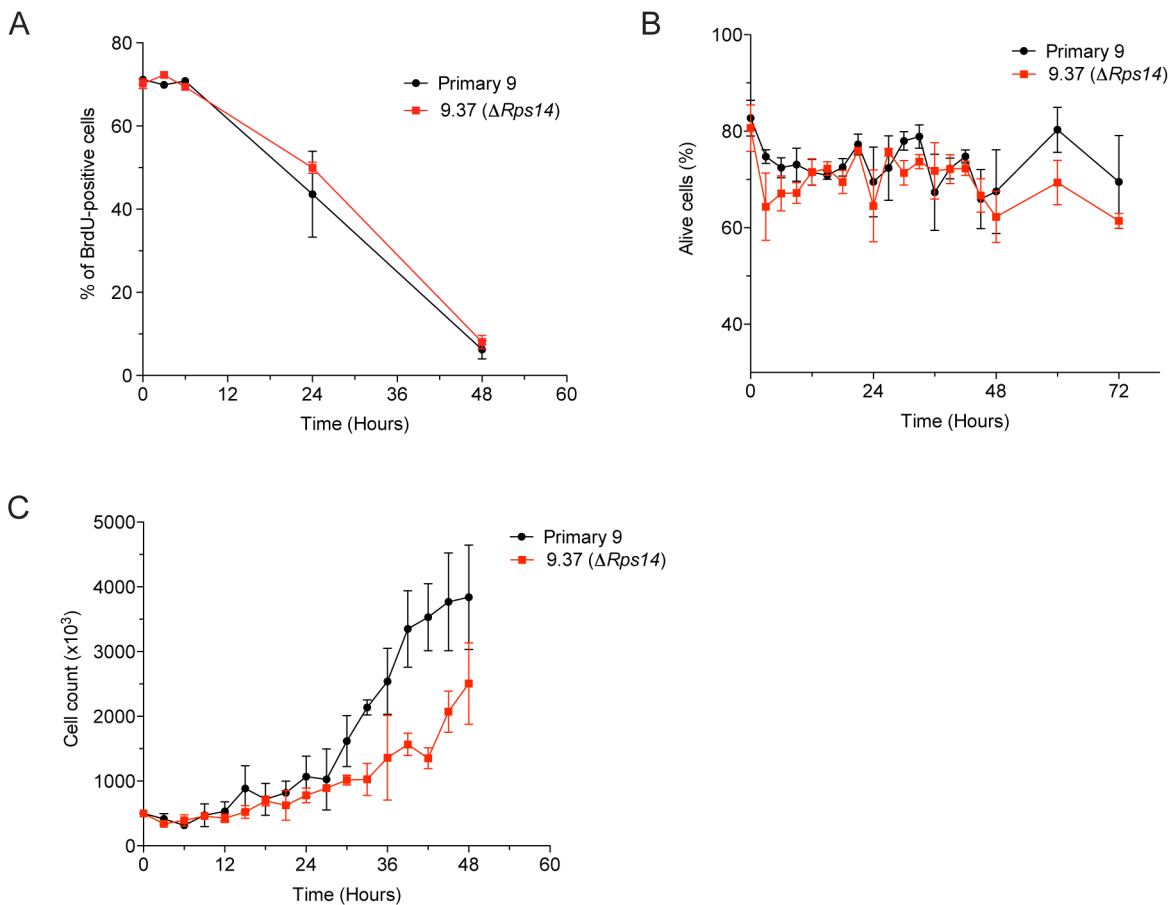


Figure S3.2: Increased cell death upon EB formation.

(A) BrdU pulse-chase experiments on self-renewing wild-type and $Rps14^{\Delta/wt}$ clones. The percentage of BrdU-positive cells is an indication of cell-cycle progression and was assessed 0, 3, 6, 24, and 48 h after the initial BrdU pulse. Error bars show SEM values of two independent experiments. (B) Cell viability experiments. Annexin V and propidium iodide staining was performed on differentiating (EB) wild-type and $Rps14^{\Delta/wt}$ clones. Viability was assessed every 3 h for the first 72 h of differentiation. Error bars show SD values of three biological replicates. (C) Total cell counts obtained before cell viability staining for the first 48 h of differentiation. Error bars show SD values of three biological replicates.

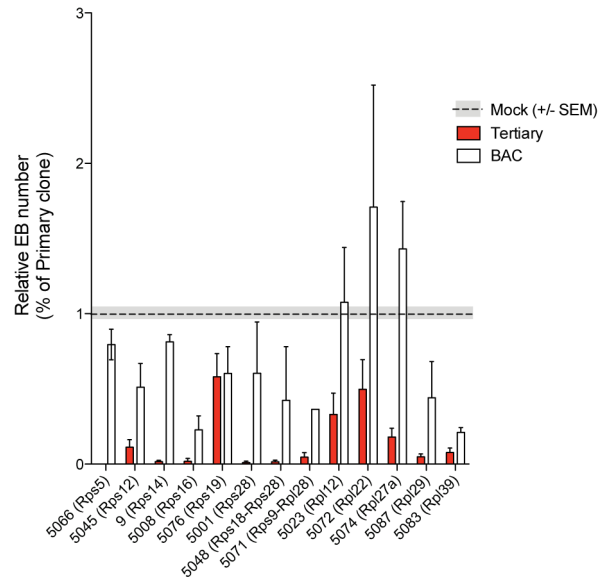


Figure S3.3: BAC complementation.

Complementation experiments using Δ RP-ESCs. Relative EB numbers (based on primary clone values) obtained from mock- transfected (red bars) and BAC-transfected (white bars) clones are represented. Error bars show the SEM of at least two independent experiments.

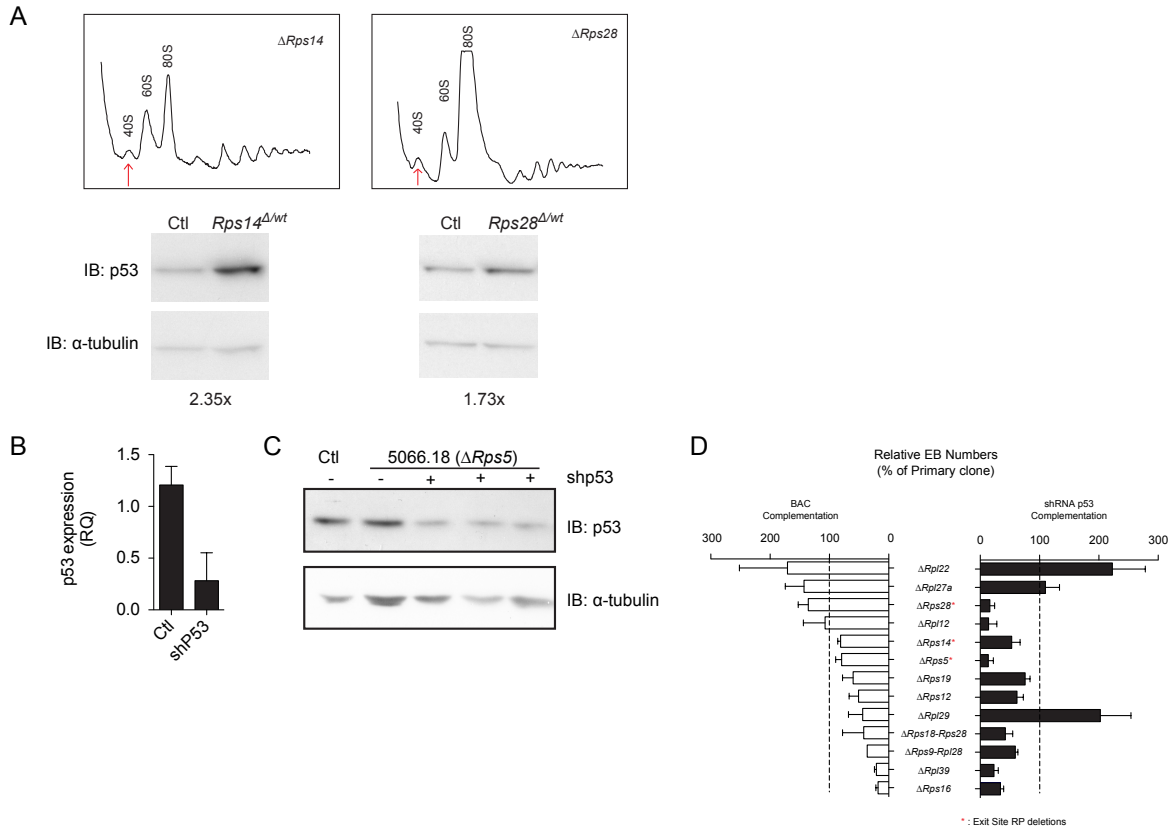


Figure S3.4: p53 status and evaluation of knockdown efficiency.

(A) Polysome profiles and p53 expression levels of $Rps14^{\Delta/wt}$ and $Rps28^{\Delta/wt}$ clones. The reduced 40S peak is indicated by red arrows. Western blot quantifications were made by normalizing p53 levels on α -tubulin with ImageJ software. (B) qRT-PCR assessment of relative p53 expression (RQ) of mock-transfected (Ctl, n = 3) and sh-p53-transfected (n = 10) cells from family 5066. The ΔC_t values range from 5.88 (Ctl) to 9.87 (sh-p53 with ~92% knockdown). (C) Western blot analysis of p53 protein levels in mock- and sh-p53-transfected clones (n = 3 independent transfections) from family 5066. α -Tubulin is used as loading control. IB, immunoblot. (D) Relative EB numbers (expressed as a percentage of primary clone numbers) following BAC or shRNA p53 transfection. Error bars show the SEM of at least two (BAC transfections) or four (shp53 transfections) independent experiments performed in duplicate.

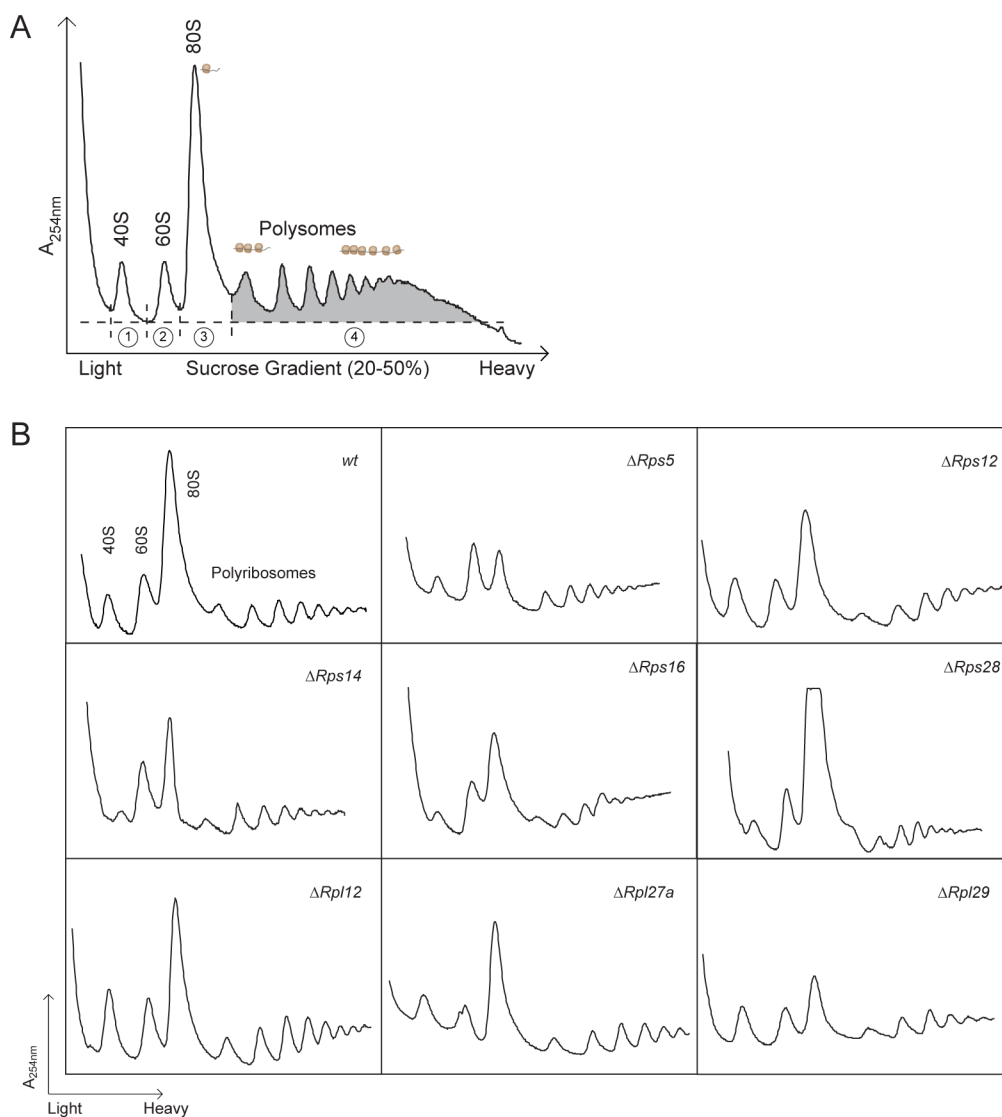


Figure S3.5: Polyribosome profiles of undifferentiated Δ RP-ESCs.

(A) AUC evaluation. A horizontal line has been drawn from baseline levels, and peaks were delimited manually (dashed lines). For example, the highlighted gray region represents the polysomal AUC to be calculated. The 40S, 60S, 80S, and polysomal regions are labeled 1 to 4, respectively. Quantitation was done using ImageJ image analysis software. (B) Ribosomal profiles of all tested *Rp*-deleted clones (*Rps* and *Rpl* genes) in self-renewal conditions (undifferentiated). Profiles shown are representative of at least two independent experiments. For a full description of the experimental setup, see SI Materials and Methods.

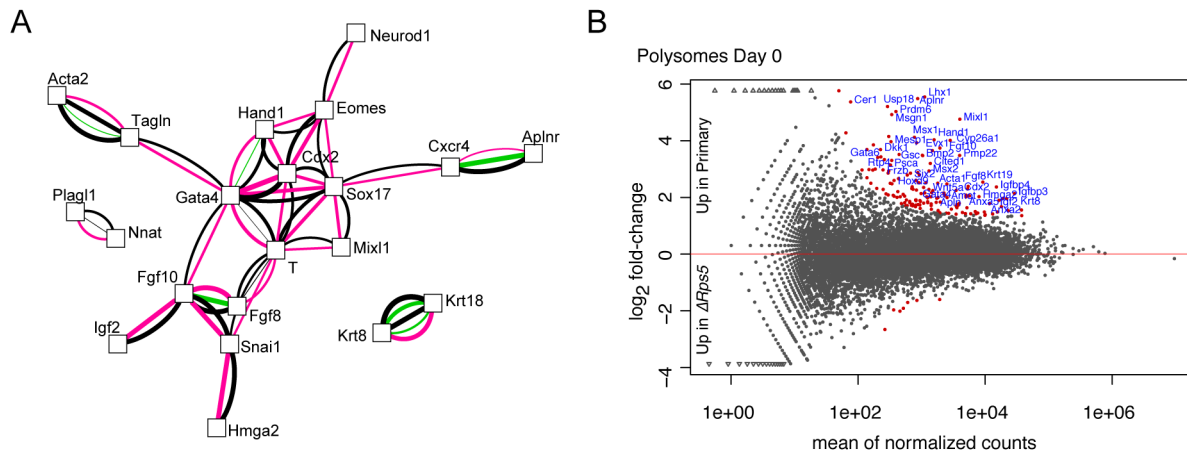


Figure S3.6: Differential expression analysis of wild-type vs. *Rps54/wt* deleted clones.

(A) Molecular network, built in String (1), of genes identified from DESeq analysis of EB polysomal fractions. Genes with an adjusted P value (P_{adj}) < 0.1 have been used, and unlinked nodes have been removed. Edges (colored connections) represent predicted functional links, as established by String. (B) DESeq analysis obtained from ESC polysomal fractions. Genes that are differentially expressed (P_{adj} < 0.1) are shown in red. Genes with a P_{adj} < 0.01 are labeled and are enriched for mesodermal differentiation-related genes.

1. Franceschini A, et al. (2013) STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41(Database issue):D808–D815.

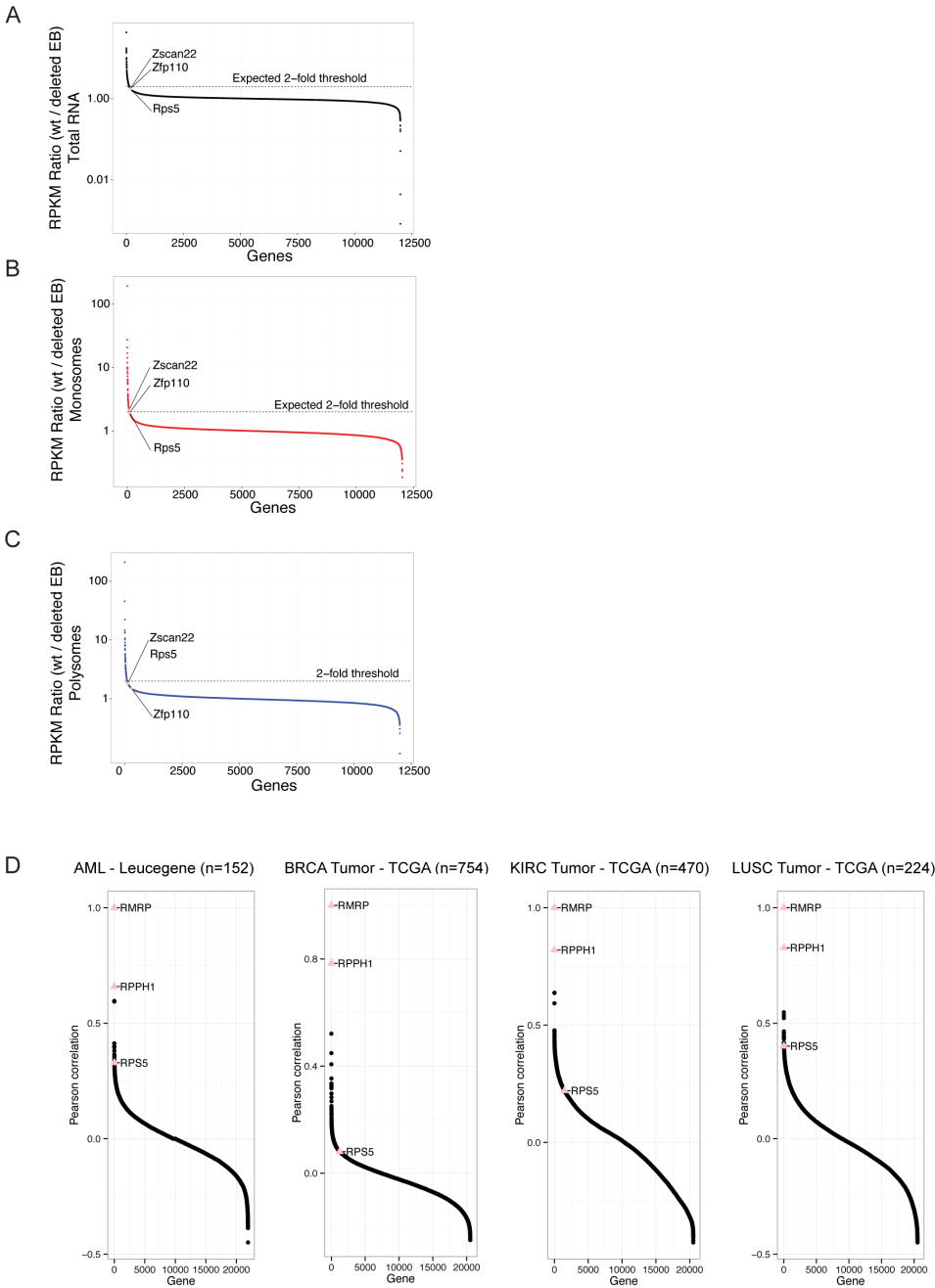


Figure S3.7: Enrichment graphs of next-generation sequencing experiments.

Enrichment graphs of next-generation sequencing experiments. (A) Enrichment curve showing RPKM ratios (wild-type/*Rps5*^{Δ/wt} EBs) of genes found in total RNA extracts. (B) Enrichment curve showing RPKM ratios (wild-type/*Rps5*^{Δ/wt} EBs) of genes found in monosomal RNA extracts. (C) Enrichment curve showing RPKM ratios (wild-type/*Rps5*^{Δ/wt} EBs) of genes found in polysomal RNA extracts. Genes with RPKM ≥ 1 are shown (n = 12,009 genes). Genes included in

the deleted interval are identified on each graph (*Zfp128* RPKM <1). The dashed lines represent the theoretical twofold threshold expected for hemizygous deletions. (D) Pearson's correlation coefficient curve of log₂-transformed expression values for RMRP in human cancer samples obtained from the The Cancer Genome Atlas (1) collection and the Institute for Research in Immunology and Cancer Leucégène Project.

1. Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216):1061–1068.

Tableau S3.1: Analysis of GO-term clustering from the EB-formation screen.

The dataset shows the complete GO term clustering analysis file performed on 320 deletion clones with available mapping information. See SI Materials and Methods for complete methodology used for clustering.

Due to size constraints (rows>7000), this table is not complete in the present document.

A preview of the table is shown.

File is available as an Excel file online at <http://www.pnas.org/content/early/2015/01/29/1418845112.abstract?tab=ds>

go_term	n_del_pos	n_del_neg	n_fam_pos	log_eb_pos	log_eb_neg	T	p	go_desc	deleted_genes
rps	43	277	18	-2.78	-1.25	-4.273	0.000025517		Rpl12,Rpl18,Rpl22,Rpl28,Rpl29,Rpl3,Rpl38,Rpl8,Rps11,Rps12,Rps14,Rps16,Rps17,Rps18,Rps19,Rps27,Rps28,Rps3,Rps5,Rps9
GO:0000002	12	308	5	-0.876	-1.478	0.913	0.3619811	mitochondrial genome maintenance	AK3,Mef2a,Mpv17,Mrpl17
GO:0000010	4	316	1	-2.131	-1.447	-0.605	0.546312	trans-hexaprenyltransferase acti	Prss2
GO:0000012	14	306	5	-1.505	-1.453	-0.083	0.9336019	single strand break repair	Aplf,Tdp1,Tpcc2,Xrcc1
GO:0000014	1	319	1	-5.094	-1.444	-1.629	0.1042797	single-stranded DNA specific endonuclease	Erc4
GO:0000018	1	319	1	0.348	-1.461	0.805	0.4212631	regulation of DNA recombination	Smarcad1
GO:0000028	5	315	4	-3.225	-1.428	-1.764	0.0753369	ribosomal small subunit assembly	Rps14,Rps17
GO:0000030	8	312	5	-2.011	-1.441	-0.709	0.4790456	mannosyltransferase activity	Pigm,Plgv,Pomt1,Pomt2
GO:0000033	4	316	3	-3.773	-1.426	-2.091	0.0373515	alpha-1,3-mannosyltransferase activiti	Alg2,Alg8

Tableau S3.2: DeIES families used in validation experiments.

Identification of all clones used in validation experiments are grouped by anchor points (families).

The presence of a RP gene deletion is indicated in the last column.

Family ID	Validated clone	RP Gene Deleted?
9	9.18	Rps14
	9.37	Rps14
5001	5001.02	Rps28, Rpl10a
	5001.06	Rps28
	5001.18	Rps28
	5001.30	Rps28
	5002.25	-
5002	5002.31	-
	5002.37	-
	5006.03	-
5006	5006.13	-
	5006.28	-
	5007.04	-
5007	5007.06	-
	5007.09	-
	5008.31	Rps16
5008	5008.20	-
	5008.38	Rps16
	5008.42	Rps16
	5010.14	-
5010	5010.19	-
	5010.22	-
	5016.22	-
5016	5016.25	-
	5016.30	-
	5017.04	-
5017	5017.05	-
	5017.11	-
	5023.12	Rpl12
5023	5023.21	-
	5023.30	Rpl12
	5026.09	-
5026	5026.17	-
	5026.18	-
	5029.01	-
5029	5029.15	-
	5029.38	-
	5030.07	-
5030	5030.16	-
	5030.24	-
	5030.28	Rps3, Rps11, Rps13, Rps16, Rps17, Rps19, Rpl13a, Rpl18, Rps27a
	5032.09	-
5032	5032.16	-
	5032.23	-
	5034.02	Rpl3
5034	5034.11	Rpl3
	5035.01	-
	5035.10	-
5035	5035.30	-
	5045.10	-
	5045.16	Rps12
5048	5048.31	Rps18, Rps28, Rpl71
	5048.34	Rps18, Rps28
	5056.02	-
5056	5056.11	-
	5056.33	-
	5066.18	Rps5
5067	5067.09	-
	5067.30	-
	5067.36	-
	5068.04	-
5068	5068.05	-
	5068.09	-
	5068.22	-
	5071.15	-
5071	5071.21	Rps9, Rpl28
	5072.15	Rpl22
5072	5072.19	-
	5074.18	Rpl27a
5074	5074.26	Rpl27a
	5074.31	Rpl27a
	5076.07	Rps19
5076	5076.39	Rps19
	5078.03	-
	5078.05	-
5078	5078.09	-
	5079.26	-
	5079.29	-
5079	5079.32	-
	5081.06	-
	5081.39	-
5081	5082.12	-
	5082.29	-
	5082.31	-
5083	5083.14	-
	5083.26	Rpl39l
	5084.28	-
5084	5084.33	-
	5085.15	-
	5085.16	-
5085	5087.27	-
	5087.42	-
	5087.44	Rpl29
	5088.10	-
5088	5088.14	-
	5088.32	-

Tableau S3.3: List of significant ($P_{adj} < 0.1$) genes obtained from DESeq analysis.

The dataset shows the list of genes that are differentially expressed in wild-type and *Rps5*^{Δ/wt} clones in the EB polysomal fraction. Only significant ($P_{adj} < 0.1$) genes are shown.

Due to size constraints (rows>300), this table is not complete in the present document.

A preview of the table is shown.

File is available as an Excel file online at <http://www.pnas.org/content/early/2015/01/29/1418845112.abstract?tab=ds>

Column1	id	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldCha	pval	padj
17533	Rps5	136085.90	95307.44	176864.35	1.86	0.89	1.35E-06	7.11E-04
12874	Mtl	57322.24	43281.46	71363.01	1.65	0.72	9.85E-05	2.40E-02
4061	Ceng1	26425.07	33126.63	19723.51	0.60	-0.75	6.69E-05	1.84E-02
20690	Trp53imp1	18050.84	23587.67	12514.00	0.53	-0.91	1.51E-06	7.72E-04
17448	Rpl22l1	11575.69	8491.48	14659.89	1.73	0.79	4.51E-05	1.41E-02
16615	Ptp4a3	11168.44	14200.54	8136.35	0.57	-0.80	3.30E-05	1.07E-02
3516	Btg2	7959.26	9825.41	6093.10	0.62	-0.69	4.70E-04	8.25E-02
17841	Sdc4	7913.87	9940.95	5886.79	0.59	-0.76	1.28E-04	2.92E-02
22221	Zfp365	7277.39	10097.54	4457.24	0.44	-1.18	3.77E-09	4.10E-06
115	1500009L16	7236.56	8961.27	5511.86	0.62	-0.70	4.11E-04	7.43E-02
15655	Phlda3	6631.67	8411.28	4852.07	0.58	-0.79	7.28E-05	1.95E-02
801	2900011O08	5998.12	7631.16	4365.08	0.57	-0.81	6.51E-05	1.82E-02
2214	Ak4	5942.00	7892.79	3991.20	0.51	-0.98	1.20E-06	6.94E-04
7032	Fam81a	5720.51	7211.03	4229.98	0.59	-0.77	1.45E-04	3.11E-02
11817	Mep1b	5517.18	6943.67	4090.69	0.59	-0.76	1.72E-04	3.56E-02
3426	Bnip3	5238.27	6907.39	3569.15	0.52	-0.95	3.32E-06	1.61E-03
15884	Plekhg1	5121.12	6385.08	3857.15	0.60	-0.73	3.78E-04	7.10E-02
6308	Eda2r	4784.82	6073.80	3495.84	0.58	-0.80	1.11E-04	2.65E-02
17403	Romo1	4362.16	3230.26	5494.05	1.70	0.77	2.31E-04	4.63E-02
13521	Nnat	4275.47	2900.84	5650.10	1.95	0.96	4.30E-06	1.93E-03
8103	Gml3247	3545.52	4545.09	2545.95	0.56	-0.84	9.12E-05	2.31E-02
19626	Tagln	3517.41	2367.07	4667.75	1.97	0.98	4.89E-06	2.10E-03
15786	Pkdcc	3231.45	2000.41	4462.48	2.23	1.16	1.06E-07	7.49E-05
3026	Axin2	3199.25	2254.40	4144.10	1.84	0.88	5.08E-05	1.51E-02
21871	Wnt7b	3134.63	4108.72	2160.55	0.53	-0.93	2.05E-05	7.03E-03
10836	Rrt8	3128.54	1811.35	4445.72	2.45	1.30	3.58E-09	4.10E-06
21763	Wdr35	2819.76	3636.07	2003.46	0.55	-0.86	9.90E-05	2.40E-02
6379	Egln3	2769.70	3512.89	2026.50	0.58	-0.79	3.34E-04	6.42E-02
19566	T	2724.25	935.75	4512.75	4.82	2.27	1.17E-22	7.17E-19
10801	Krt18	2679.89	1690.09	3669.68	2.17	1.12	5.73E-07	3.78E-04
176	1700007K13	2477.35	3375.40	1579.30	0.47	-1.10	1.33E-06	7.11E-04
9650	Hmga2	2445.40	1438.96	3451.85	2.40	1.26	3.05E-08	2.81E-05
18032	Serpinh1	2272.69	1566.91	2978.48	1.90	0.93	5.25E-05	1.54E-02
13267	Ndufa4l2	2260.20	2928.53	1591.87	0.54	-0.88	1.25E-04	2.88E-02
6560	Eomes	2065.01	1281.41	2848.61	2.22	1.15	8.72E-07	5.54E-04
19005	Sp5	1966.99	1045.56	2888.41	2.76	1.47	7.93E-10	1.13E-06
10001	Igf2	1918.06	888.01	2948.11	3.32	1.73	8.03E-13	2.17E-09
9998	Igdcc4	1731.33	2344.16	1118.50	0.48	-1.07	1.07E-05	4.10E-03
22579	Zscan22	1683.96	1159.19	2208.72	1.91	0.93	1.30E-04	2.92E-02
5485	D630023F18	1598.34	2257.27	939.41	0.42	-1.26	3.36E-07	2.30E-04
16943	Rb1	1498.28	1933.57	1062.99	0.55	-0.86	5.35E-04	9.13E-02
7258	Egf8	1420.14	767.70	2072.58	2.70	1.43	2.09E-08	2.03E-05
12632	Mmm2	1284.26	1703.45	865.06	0.51	-0.98	1.62E-04	3.40E-02
674	2410076I21E	1161.48	1567.86	755.09	0.48	-1.05	7.59E-05	2.00E-02
18749	Snai1	1154.87	687.49	1622.24	2.36	1.24	3.72E-06	1.72E-03
1764	Aard	1144.44	1541.13	747.76	0.49	-1.04	9.52E-05	2.37E-02
20694	Trp73	1066.32	1402.68	729.96	0.52	-0.94	5.29E-04	9.11E-02
1957	Acta2	1040.22	481.24	1599.20	3.32	1.73	5.79E-10	8.90E-07
12559	Mixl1	958.06	188.11	1728.02	9.19	3.20	1.89E-25	3.49E-21
5636	Ddit4l	955.38	1357.80	552.97	0.41	-1.30	4.59E-06	2.01E-03
5302	Cyp26a1	911.47	198.61	1624.34	8.18	3.03	8.49E-23	7.17E-19
1931	Acp5	908.90	1214.57	603.24	0.50	-1.01	3.93E-04	7.32E-02
5059	Cryab	814.75	498.43	1131.07	2.27	1.18	6.21E-05	1.76E-02
17227	Rhox6	775.81	1086.62	464.99	0.43	-1.22	4.57E-05	1.41E-02
4454	Chac1	775.62	495.57	1055.66	2.13	1.09	2.64E-04	5.19E-02
2358	Amot	770.40	278.82	1261.98	4.53	2.18	2.51E-12	5.80E-09
13353	Neurod1	721.92	1053.20	390.64	0.37	-1.43	3.70E-06	1.72E-03
7401	Foxj1	684.23	933.84	434.62	0.47	-1.10	4.01E-04	7.32E-02
13334	Nell2	683.11	978.72	387.50	0.40	-1.34	2.06E-05	7.03E-03
5254	Cxcr4	660.23	337.06	983.40	2.92	1.54	1.28E-06	7.11E-04
9315	H19	658.85	300.78	1016.91	3.38	1.76	4.65E-08	3.73E-05
9132	Greb1	653.13	267.36	1038.91	3.89	1.96	1.71E-09	2.25E-06
15835	Plagl1	623.78	311.28	936.27	3.01	1.59	1.07E-06	6.34E-04
17230	Rhox9	623.06	867.00	379.12	0.44	-1.19	2.14E-04	4.33E-02
13794	Nupr1	595.15	264.49	925.80	3.50	1.81	5.97E-08	4.58E-05
12876	Mt3	586.55	344.70	828.40	2.40	1.26	1.22E-04	2.85E-02
18698	Smc1b	579.69	800.16	359.22	0.45	-1.16	4.65E-04	8.24E-02

Tableau S3.4: Analysis of GO-term clustering of polyribosomal fractions.

Complete GO term clustering analysis based on processes performed on 12,009 genes (RPKM ≥ 1) from next-generation sequencing of total, monosomal, and polysomal RNA extracted from sucrose gradient fractions. See SI Materials and Methods for complete methodology used for clustering.

Due to size constraints (rows > 150), this table is not complete in the present document.

A preview of the table is shown.

File is available as an Excel file online at <http://www.pnas.org/content/early/2015/01/29/1418845112.abstract?tab=ds>

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	ln2	ln3	Genes
GO:0060429	epithelium develop	9.36E-07	5.90E-05	87.41	11800	30	18	4	[Fgf10 - fibroblast growth factor 10, Msx1 - msh homeobox 1, Bmp2 - bone morphogenetic protein 2, Lhx1 - lim homeobox protein 1]
GO:0023019	signal transduction i	4.09E-12	9.73E-10	81.94	11800	12	84	7	[Msx2 - msh homeobox 2, Gsc - goosecoid homeobox, Foxa2 - forkhead box a2, Mesp1 - mesoderm posterior 1, Gata4 - gata binding protein 4]
GO:0030916	otic vesicle formatio	8.53E-07	5.44E-05	53.03	11800	5	178	4	[Fgf10 - fibroblast growth factor 10, Fgf8 - fibroblast growth factor 8, Fgf3 - fibroblast growth factor 3, Hesx1 - homeobox gene expressed in es cells]
GO:0009952	anterior/posterior pa	4.26E-09	5.72E-07	40.02	11800	86	24	7	[Cdx2 - caudal type homeobox 2, Msgr1 - mesogenin 1, Msx1 - msh homeobox 1, Bmp2 - bone morphogenetic protein 2, Cyp26a1 - cytochrome p450 family 26 subfamily a, polypeptide 1, Lhx1 - lim homeobox protein 1, Gata4 - gata binding protein 4]
GO:0060317	cardiac epithelial to	3.57E-08	3.34E-06	38.02	11800	14	133	6	[Msx1 - msh homeobox 1, Msx2 - msh homeobox 2, Fgf8 - fibroblast growth factor 8, Shal1 - snail homolog 1 (drosophila), Bmp2 - bone morphogenetic protein 2, Tgfb2 - transforming growth factor, beta 2]
GO:0042060	wound healing	5.72E-11	1.13E-08	25.01	11800	39	121	10	[Fgf10 - fibroblast growth factor 10, Serpine 1 - serine (or cysteine) peptidase inhibitor, clade e, member 1, Wnt6a - wingless-related mmtv integration site 5a, Cd44 - cd44 antigen, Msx2 - msh homeobox 2, Col1a1 - collagen, type i, alpha 1, Cx3c1 - chemokine (c-x3-c motif) ligand 1, Alox15 - arachidonate 15-lipoxygenase, Tgfb2 - transforming growth factor, beta 2, Nog - noggin]
GO:0001837	epithelial to mesenc	6.14E-10	9.59E-08	24.95	11800	32	133	9	[Eomes - eomesodermin homolog (xenopus laevis), Msx1 - msh homeobox 1, Msx2 - msh homeobox 2, Fgf8 - fibroblast growth factor 8, Shal1 - snail homolog 1 (drosophila), Bmp2 - bone morphogenetic protein 2, Tgfb2 - transforming growth factor, beta 2, Hmga2 - high mobility group at hook 2, Nog - noggin]
GO:0003007	heart morphogenesis	1.02E-07	8.79E-06	24.94	11800	36	92	7	[Msx1 - msh homeobox 1, Fgf8 - fibroblast growth factor 8, Mesp1 - mesoderm posterior 1, Gata4 - gata binding protein 4, Tgfb2 - transforming growth factor, beta 2, Hand1 - heart and neural crest derivatives expressed transcript 1, Col1a1 - collagen, type v, alpha 1]

Tableau S3.5: Primers and probes used for qRT-PCR studies.

The table shows the sequences used for qRT-PCR studies. Unless otherwise mentioned, probes from Roche Universal Probe Library were used.

Primers	Sequence
Rmrp_fwd	gctctgaaggcctgtttct
Rmrp_rev	cttcttggcgggctaacagt
Rmrp_HybProbe	cttatccttcgcctagggg
Rpph1_fwd	ggtgagtcccagagagcag
Rpph1_rev	cagccattgaactgcact
Rpph1_HybProbe	agcttggaacagactcacgg
Rps5_fwd	cactgcgtcgagtgaatcag
Rps5_rev	gctcatctgcaaggcactc
Tp53_fwd	acgcttctccgaagactgg
Tp53_rev	agggagctcgaggctgata

3.11 Supplemental Materials and Methods

3.11.1 ESC Transfections

ESCs maintained on a feeder layer in 12-well plates were transfected with 2 μg of circular BAC DNA or p53 shRNA lentiviral plasmids (Open Biosystems) using Lipofectamine 2000 Reagent (Invitrogen), according to the manufacturer's protocol. Selection began after 48 h, with the following concentration of drugs maintained for at least 5 d: 1.5 $\mu\text{g}/\text{mL}$ puromycin (Sigma), 150 $\mu\text{g}/\text{mL}$ hygromycin (Roche), 15 $\mu\text{g}/\text{mL}$ blasticidin (Sigma), or 30 $\mu\text{g}/\text{mL}$ zeocin (Invitrogen). Genomic DNA from BAC-transfected clones was isolated using DNAzol (Invitrogen), according to the manufacturer's instructions. Southern blot detection of transfected BAC DNA was performed using EcoRV digestion and a probe specific to the neomycin gene, as described (1). Total cellular RNA was isolated from BAC-transfected clones (undifferentiated ESCs or EBs) with TRIzol (Invitrogen), according to the manufacturer's instructions.

3.11.2 EB Formation

ESCs were cultured for at least one passage on gelatin-coated dishes before being seeded at a concentration of 1.25×10^3 cells/mL in bacterial-grade 60-mm dishes without LIF [Iscove's Modified Dulbecco's Medium; Sigma], 15% heat-inactivated FCS (Invitrogen), 5% serum-free and protein-free medium for hybridoma culture (Invitrogen), 2×10^{-3} M L-glutamine (Invitrogen), 50 $\mu\text{g}/\text{mL}$ ascorbic acid (Sigma), and 3×10^{-4} M α -monothioglycerol]. EBs were counted at day 8 after LIF removal. For most experiments, EB numbers are expressed as relative values based on primary clone numbers.

3.11.3 Ribosome Profiling

Ten minutes before collection, 100 $\mu\text{g}/\text{mL}$ cycloheximide was added to the culture medium. Cells were washed in ice-cold PBS supplemented with 100 $\mu\text{g}/\text{mL}$ cycloheximide and were collected in polysome lysis buffer [15 mM Tris (pH 7.4), 250 mM NaCl, 15 mM MgCl_2 , 1% Triton X-100, 100 $\mu\text{g}/\text{mL}$ cycloheximide, 1 mM DTT, 400 U/mL RNaseOut (Invitrogen), and protease inhibitors]. Samples were centrifuged at $10,000 \times g$ for 10 min at 4 $^\circ\text{C}$. The resulting supernatant was layered on a 20–50% linear sucrose gradient and centrifuged in a Beckman SW41Ti rotor at 35,000 rpm for 3 h at 4 $^\circ\text{C}$. After centrifugation, the A254 was monitored continuously and recorded using a Gradient Station IP instrument (Biocomp) attached to a UV-MII spectrophotometer (GE Healthcare).

3.11.4 GO Analysis

Each clone was annotated with a set of GO terms representing the union of the set of GO terms ascribed to all of the genes present in the deleted region of the clone. For each GO term assigned to at least one clone, a two-sided t-test was performed comparing the log EB density (normalized EB numbers) for clones annotated with that GO term against the log EB density for all other clones. GO terms then were ranked based on the significance of their effect on EB formation. The GOrilla application (2) also was used to determine GO enrichments from ranked gene lists.

3.11.5 Assessment of p53 Expression Levels

After shRNA selection, gene expression was assessed by qRT-PCR using a Roche Light Cycler 480 with Roche Universal ProbeLibrary (UPL) assays or TaqMan assays (endogenous control genes). For Trp53 (NM_011640) detection, UPL probe no. 25 and the primer pair acgcttctccgaagactgg + agggagctcgaggctgata were used. Reference gene assays (Gapdh and β -actin) were purchased from ABI (20 \times primer-probe mix, VIC-labeled). Protein levels were detected from whole-cell extracts (Laemmli buffer) by Western blotting using p53 (Cell Signaling Technology, catalog no. 2524) and α -tubulin (Cell Signaling Technology, catalog no. 2144) antibodies.

3.11.6 BrdU Pulse-Chase Assay

Cells were labeled with BrdU for 15 min and then were collected at various time points before being analyzed for fluorescence by flow cytometry using a BD Canto II cytometer (BD Bioscience) and BD FACSDiva 4.1 software.

3.11.7 Annexin V/Propidium Iodide Staining

ESCs or EBs were dissociated and counted before staining with Alexa350-annexin V (Invitrogen Molecular Probes) and propidium iodide (50 μ g/mL) in accordance with the manufacturer's instructions. Fluorescence was evaluated by flow cytometry using a BD Canto II cytometer (BD Bioscience), and data were analyzed with BD FACSDiva 4.1 software.

3.11.8 qRT-PCR

Relative quantification of the target genes was calculated as described previously. Primers used for PCR reactions are provided in Tableau S3.5.

3.12 Supplemental References

1. Kroon E, et al. (1998) Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J* 17(13):3714–3725.2.
2. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.

Chapitre 4: Discussion et perspectives

Malgré l'avènement de nouvelles technologies permettant l'interrogation du génome, la stratégie expérimentale développée pour créer DeIES est toujours d'actualité. À l'époque où l'utilisation de rétrovirus fut optimisée dans notre laboratoire pour créer, par un processus de recombinaison, des délétions chromosomales chez les ESC, cette méthodologie était complémentaire à d'autres efforts internationaux conduits parallèlement aux nôtres et ayant des objectifs similaires, soit l'identification de facteurs régulant la pluripotence des ESC. Toutefois, notre approche proposait des intégrations rétrovirales aléatoires, permettant du coup de façon avantageuse l'obtention de délétions aussi bien dans des régions codantes et non-codantes du génome murin, en plus de potentiellement permettre l'identification d'interactions synthétiques essentielles (plusieurs gènes ou éléments qui interagissent). Aussi, la génération de délétions hétérozygotes permettrait d'établir des cribles pouvant conduire à l'identification de facteurs haploinsuffisants. En rétrospective, et tel que détaillé dans le Chapitre 2 de cette thèse, notre stratégie a permis de générer 1307 clones comportant des délétions chromosomales de tailles variant entre ~700 bp et ~100 Mb, couvrant dans les proportions attendues autant des éléments codants que non-codants du génome murin. Au total, avec une couverture génomique d'environ 25% (423 clones présentement cartographiés sur une possibilité de 1307), nous avons démontré que la librairie DeIES est utilisable pour identifier de potentiels éléments contribuant à la différenciation des ESC.

4.1 Choix d'un crible fonctionnel de DeIES

À l'origine de la création de DeIES, notre objectif principal était d'établir une collection de clones de ESC qui serait disponible pour l'ensemble de la communauté scientifique sous forme de ressource. Par la suite, notre but était d'utiliser cette librairie afin d'identifier des marqueurs spécifiques de la différenciation hématopoïétique.

4.1.1 Formation de corps embryoides

Lors de la mise au point de la technique d'infection virale qui serait subséquentement utilisée pour créer notre librairie, des familles de ESC prototypes comportant des délétions aléatoires ont été générées. Afin d'évaluer l'impact de ces délétions sur les clones produits, ces derniers ont été employés dans un essai de différenciation hématopoïétique. Ces travaux de mise au point ont montré que certaines des délétions générées perturbaient la formation même des EB dans les premiers jours de différenciation et non pas seulement un type cellulaire spécifique devant

apparaître plus tard au cours du processus⁸⁷. Ce phénotype anormal, facilement identifiable et qui laissait une fenêtre plus qu'intéressante pour les expériences de complémentation subséquentes, a réaffirmé que cette approche était valable dans l'objectif d'identifier de nouveaux éléments affectant la pluripotence des ESC. L'adaptation de notre protocole de différenciation en milieu semi-solide (méthylcellulose) devant permettre un criblage à grande échelle (plaques de 96 puits) était aussi envisageable, bien que techniquement le développement d'une méthode fiable et reproductible représentait un défi. Finalement, tel que présenté au chapitre 2, cet essai à grande échelle des 1307 clones de notre collection a permis d'avoir des données sur la formation de EB pour 45% des ESC. Ce nombre peut sembler petit si l'on considère l'ensemble des clones de la librairie. Toutefois, c'est une des limites que comporte l'utilisation de cette collection dans un criblage. En effet, les clones sont stockés dans des plaques de 96 puits devant servir à cribler de façon efficace la totalité de la librairie. De ce fait, il est évident que suite au processus de congélation/décongélation de chaque plaque, l'utilisateur de la librairie doit composer avec des confluences variables à l'intérieur d'une même plaque. Ce faisant, puisque la confluence des ESC détermine de façon critique le maintien de leur pluripotence, nous avons décidé que lorsque 50% des clones contenus dans une même plaque seraient suffisamment confluent, le passage de la plaque entière serait effectué. De cette façon, nous nous assurons de conserver les clones qui semblaient en bon état au lieu de mettre l'emphase sur ceux qui proliféraient moins bien ou qui avaient mal résistés à la congélation.

4.1.2 Établissement d'un criblage spécifique à l'hématopoïèse

Comme mentionné précédemment, la formation *in vitro* de EB permet d'observer le développement de précurseurs hématopoïétiques¹⁷¹. Lors du criblage de la librairie, un marquage à l'aide de la benzidine a été effectué sur les EB de jour 8 afin de révéler l'existence de cellules hématopoïétiques primitives (résultats non-publiés). En présence d'hémoglobine, l'oxydation de ce composé organique produit une coloration bleue distinctive. La facilité et la rapidité de ce marquage, qui peut être effectué directement en méthylcellulose, ont motivé notre décision d'utiliser la benzidine afin d'avoir un aperçu de la capacité des cellules à former adéquatement des précurseurs hématopoïétiques. À la lumière de nos résultats, plusieurs familles de clones ont été identifiées comme ayant un phénotype anormal de différenciation hématopoïétique (% de EB positifs pour la benzidine < 30%, Figure 4.1). Ces résultats sont bien entendus préliminaires, notre objectif actuel étant d'adapter un essai de validation plus précis¹⁷², qui comportera une différenciation spécifique des ESC vers la voie hématopoïétique, suivie de l'identification par qRT-PCR de gènes marqueurs des différents stades de maturation des cellules progénitrices de même que la caractérisation par cytométrie en flux des populations présentes. Cet essai sera dans

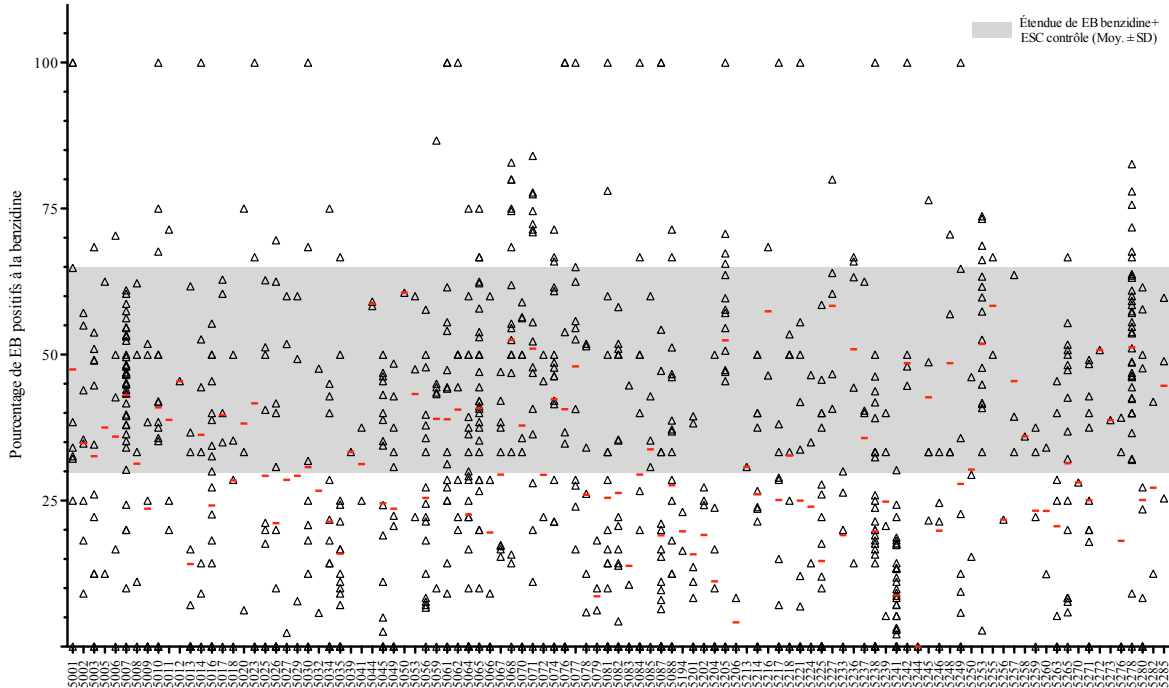


Figure 4.1: Détection de précurseurs hématopoïétiques lors du criblage de formation de EB.

Un marquage à la benzidine a été effectué afin d'évaluer la présence de précurseurs hématopoïétiques lors de la formation de EB. Le pourcentage de EB positifs à la coloration est exprimé pour chaque clone des familles de la librairie.

un premier temps utilisé pour valider les candidats découlant du marquage à la benzidine, avec en tête d'en adapter les paramètres pour permettre de potentiellement utilisé une méthodologie similaire pour tester à grande échelle sur la librairie le rôle des délétions chromosomales hétérozygotes sur le développement de la lignée hématopoïétique.

4.1.3 Utilisations de DeIES et modèles alternatifs

À ce jour, le criblage primaire effectué sur les clones de DeIES, en plus de celui portant sur la spécification hématopoïétique qui est en cours d'optimisation (section 4.1.2), avaient pour but principal, en plus de faire la démonstration du bon fonctionnement de notre stratégie, d'identifier des déterminants essentiels à la différenciation des ESC en EB et ultimement leur implication au niveau de l'hématopoïèse. On peut toutefois imaginer d'autres utilisations pour cette collection de clones.

Par exemple, d'un point de vue technique, il a été démontré que l'utilisation de concentrations élevées de généticine (G418) produit une perte d'hétérozygoté, puisque la résistance à cette

drogue est dépendante des niveaux d'expression du gène de la néomycine¹⁷³. Ainsi, bien que nous n'ayons pas démontré la faisabilité de cette approche à partir de clones de DeIES, il est envisageable que des délétions homozygotes puissent être obtenues de façon relativement facile à partir des intégrations hétérozygotes, puisque la variante du gène de la néomycine permettant cette transformation a été utilisée.

Aussi, une utilisation additionnelle pour DeIES serait l'étude de délétions couvrant des régions non-codantes du génome, dans le but d'identifier des éléments s'y retrouvant qui pourraient jouer un rôle de régulation important au niveau cellulaire. Des études rapportant l'identification d'ARN non-codants ayant des fonctionnalités essentielles au sein de différents types cellulaires sont de plus en plus relevées dans la littérature. En effet, que ce soit sous l'action de microARN ou encore de longs ARN non-codant (lncARN), ces recherches lèvent le voile sur une autre dimension de la régulation transcriptionnelle, post-transcriptionnelle et épigénétique de processus cellulaires variés comme la pluripotence, la différenciation, le développement oncogénique ou, à l'opposé, des activités suppresseur de tumeurs^{70,174}. L'interrogation des régions non-codantes pourrait ainsi permettre d'identifier des rôles moléculaires présentement non-appréciés propres à ces éléments. En ce sens, des résultats préliminaires non-publiés permettent de croire que deux microARN (mir-143 et mir-145) situés en aval de la région minimale délétée dans la famille 9 (clone 9.18, voir section 2.4.6) sont responsables du phénotype additionnel que comporte cette délétions plus large par rapport au clone 9.37. De façon intéressante, le microARN-145 a été identifié comme affectant la différenciation en érythrocytes et en mégacaryocytes dans le contexte du syndrome 5q-¹⁷⁵, alors que la diminution du niveau d'expression de ces deux microARNs provoquait un blocage de la différenciation conduisant au développement de leucémies¹⁷⁶. Bien entendu, des expériences supplémentaires sont nécessaires pour démontrer le rôle de mir-143/mir-145 dans la différenciation des ESC, mais ces résultats préliminaires illustrent bien le potentiel de DeIES pour permettre l'identification d'éléments localisés dans des régions non-codantes.

De la même façon que les utilisations de la librairie peuvent être multiples, l'adaptation de la méthodologie utilisée pour créer DeIES, soit l'utilisation d'intégrations virales aléatoires, pourrait s'avérer fort utile afin de répondre à d'autres types de questions dans divers contextes expérimentaux.

Par exemple, une alternative intéressante pour obtenir des délétions homozygotes serait d'employer notre méthodologie d'infection virale, mais cette fois-ci dans une lignée de ESC haploïdes¹⁷³ (section 1.1.3.2.6). Ce modèle permettrait ainsi d'évaluer l'influence des

modifications chromosomales sur l'apparition de phénotypes récessifs dans un contexte de pluripotency et de différenciation. Dans ce même ordre d'idée, l'infection avec notre combinaison de virus d'une lignée cellulaire haploïde de leucémie myéloïde chronique (LMC) humaine (lignée KBM-7) sera sous peu mise en branle dans le laboratoire. Combinée à un criblage dans le but d'identifier des petites molécules chimiques qui induisent la mort cellulaire, l'obtention de larges délétions, homozygotes cette fois, pourrait permettre de réduire considérablement la charge de travail nécessaire à l'identification des voies moléculaires ciblées par de potentiels thérapeutiques. En effet, on peut s'imaginer qu'un clone comportant une délétion d'un gène ciblé par une molécule X aura un avantage important sur la population clonale en général. De par leur forte représentation par rapport aux clones toujours affectés par les traitements, ces délétions pourraient être facilement cartographiables à l'aide de techniques de séquençage à haut-débit. Bien que plusieurs paramètres expérimentaux restent à déterminer, la mise au point de cette approche pourrait s'avérer d'une grande utilité afin de déterminer les mécanismes d'actions de nouvelles molécules thérapeutiques.

Ces utilisations alternatives exploitant l'intégration aléatoire des virus et la génération de délétions chromosomales démontrent bien l'énorme potentiel de cette technique dans le but de réaliser des expériences de génomique fonctionnelle.

4.1.4 Limite de la méthode

Malgré les nombreux usages envisageables d'une telle méthodologie, une des limites de cette ressource est la complexité liée à faire la démonstration qu'un élément identifié dans le contexte d'une large délétion chromosomale est le seul responsable d'un phénotype en particulier. Évidemment, dans le cas des clones de DelES, des questions relatives à l'intégrité fonctionnelle des cellules étudiées doivent être adressées avant de pouvoir conclure hors de tout doute à l'identification d'un nouvel élément essentiel au maintien des propriétés intrinsèques des ESC. En ce sens, une décennie après le développement de notre stratégie, de nouvelles techniques permettant d'éditer le génome ont vu le jour et pourraient s'avérer complémentaires à notre approche. Ainsi, l'utilisation de techniques dirigées permettant de diminuer l'expression génique, telle que celle par CRISPR-Cas9 (section 1.1.3.2.3), pourrait être envisageable afin de sortir du contexte des grandes délétions chromosomales. En générant chez des ESC sauvages une délétion d'un gène candidat identifié préalablement dans un clone de DelES, il serait ainsi possible de minimiser les impacts négatifs potentiels imputables à d'autres éléments délétés. Une telle approche permettrait aussi de ne pas reposer uniquement sur les études de complémentation pour valider le caractère essentiel des facteurs identifiés, puisque ces expériences, qui sont

dépendantes de la surexpression efficace des éléments souhaités, représentent souvent un défi au point de vue technique.

4.2 Déterminants régissant le destin cellulaire des ESC

L'évaluation fonctionnelle des clones de la librairie DeES a mené à l'identification de gènes codants pour des protéines ribosomales comme étant essentiels à la formation de EB. Le regroupement par termes d'ontologie génique (GO) de gènes délétés liés à un développement anormal de EB a en effet permis d'identifier ce groupe de gènes comme celui affectant de façon la plus significative ($p < 6 \times 10^{-7}$) le processus de différenciation des ESC. Nous avons ainsi choisi ces candidats pour la suite de nos études, considérant aussi la littérature grandissante à l'époque relevant l'implication probable des RP dans le développement de troubles hématopoïétiques humains. En effet, parallèlement à la publication de la librairie DeES et à la première démonstration d'une complémentation fonctionnelle en réintroduisant la séquence codant pour le gène *Rps14*, Ebert et al. démontraient que ce gène était en fait un marqueur d'un syndrome myélodysplasique (MDS), la pré-leucémie 5q-¹⁴⁸. La littérature faisant référence au stress nucléolaire induit par un déséquilibre du niveau d'expression des protéines ribosomales étant relativement pauvre il y a à peine quelques années, cette catégorie de gènes se révélait ainsi comme très prometteuse pour des études plus en profondeur des mécanismes moléculaires liant l'expression des protéines ribosomales et la différenciation cellulaire.

4.2.1 Voie de signalisation p53-indépendante et protéines ribosomales

Étant donné que le premier modèle murin de ribosomopathies (*Rps19*^{-/-}) démontrait un phénotype embryonnaire léthal¹⁷⁷ et phénocopiait les caractéristiques cellulaires de l'anémie Diamond-Blackfan tels que vus chez l'humain¹⁴⁴, les études subséquentes s'attardèrent majoritairement aux effets sur le système hématopoïétiques et à la capacité à corriger spécifiquement ces déficiences. De nombreuses études pointaient vers l'implication d'une réponse p53-dépendante suite à de telles perturbations qui mènent fréquemment au développement de maladies^{110,178-180} (voir aussi Chapitre 1.2.2).

Récemment, un groupe a pu démontrer qu'une réduction des niveaux d'expression de p53 n'arrivait qu'à renverser partiellement les anomalies vues dans un modèle de DBA et soulevait ainsi pour la première fois l'hypothèse qu'un processus p53-indépendant était responsable d'un phénotype anormal de formation des érythrocytes¹⁸¹. Hors, au cours des dernières années, les études rapportant des effets moléculaires indépendants d'une réponse p53 suite à un

débalancement de l'expression de protéines ribosomales se multiplie, en ligne avec nos plus récentes observations sur le rôle de *Rps5* dans la formation de corps embryoides (Rpl3¹⁸², Rpl5¹⁴⁵, Sbsd¹⁸³). De façon intéressante, comme nos résultats le démontrent (non publiés), le simple fait de diminuer les niveaux d'expression de p53 dans des ESC sauvages conduit à une augmentation de l'ordre de trois fois du nombre de EB obtenus, ce qui concorde avec les observations qu'une augmentation globale et non-spécifique de la prolifération et la différenciation des cellules pourrait être responsable des améliorations résultantes d'une diminution de p53¹⁴⁵. Dans un même ordre d'idée, des études récentes ont aussi démontré que l'ajout de L-Leucine à la diète d'animaux modélisant les syndromes humains DBA et 5q- pouvait améliorer la condition de ces derniers^{149,184}. La L-leucine stimulant la production de protéines, ces résultats suggèrent qu'un déficit à ce niveau pourrait être l'une des causes de ces maladies. Toutefois, une étude toute récente démontre l'importance d'avoir un équilibre au niveau de la synthèse protéique chez les cellules souches hématopoïétiques, rapportant qu'un taux trop élevé de protéines conduit à des effets négatifs sur la capacité de ces cellules à repeupler la moelle osseuse d'animaux compromis¹⁸⁵. De façon intéressante, ces résultats servent d'hypothèse pouvant expliquer le fait que nous n'avons jamais pu observer une augmentation marquée du niveau d'expression de toutes les protéines ribosomales testées lors des expériences de complémentation. Seulement des clones exprimant un niveau près de celui endogène ont pu être obtenus après sélection.

Toutes ces nouvelles données soutiennent notre hypothèse qu'une modulation de l'expression d'un groupe restreint de gènes/protéines est potentiellement responsable des effets p53-indépendent relevés dans le cadre de nos travaux sur les RP.

4.2.2 Rôle spécialisé de certaines protéines ribosomales

Nos plus récents résultats rapportés au Chapitre 3 de la présente thèse ont permis d'identifier plusieurs gènes jouant un rôle dans le processus de différenciation des ESC au niveau du mésoderme et de la lignée primitive qui étaient sous-représentés dans les clones hétérozygotes pour *Rps5*. Le processus de différenciation *in vitro* des ESC en EB répliquant de façon temporelle celui observé *in vivo*, il semble cohérent, voire même attendu, qu'une signature anormale impliquant des gènes spécifiant le mésoderme soit retrouvée chez des cellules *Rps5*^{Δ/wt} n'étant pas capable d'initier adéquatement cette première étape de la différenciation. Le résultat inattendu est plutôt venu de l'identification d'une signature similaire, mais cette fois provenant des ESC encore indifférenciées, maintenues dans un milieu d'auto-renouvellement. Cette observation suggère ainsi que les causes expliquant l'incapacité des cellules *Rps5*^{Δ/wt} à se différencier sont déjà présentes dans leur état indifférencié, bien que ces clones semblent à

plusieurs égards normaux. Ainsi, d'un point de vue pratique, ce résultat vient donc simplifier les expérimentations à venir, puisque l'identification d'un régulateur plus général devrait être possible dans des conditions d'auto-renouvellement, sans avoir à passer par une initiation de la différenciation *in vitro*.

Dans cette ligne de pensée, certaines études ont démontré un effet sur la traduction de certains gènes spécifiques suite à un stress ribosomal, par exemple celle de plusieurs gènes Hox via un processus impliquant Rpl38^{186,187} et de façon plus générale la traduction des ARNm 5'-TOP dans un modèle d'haploinsuffisance du gène *Rpl11*¹⁶⁴. Il s'agit là d'un concept récent qui suggère l'existence de ribosomes «spécialisés»¹⁸⁸, rendant ainsi possible de concevoir que certaines protéines ribosomales peuvent liées des ARNm spécifiques permettant donc la production de protéines requises à des moments précis de la différenciation cellulaire.

On peut donc imaginer que certaines des protéines ribosomales identifiées lors de nos recherches, ou encore des protéines ribosomales libres suite à un déséquilibre, puissent lier de façon spécifique certains gènes mésodermaux, ou encore un régulateur en amont de ceux-ci, et qu'ainsi cette liaison engendre des fluctuations d'expressions incompatibles avec la différenciation des ESC (Figure 4.2).

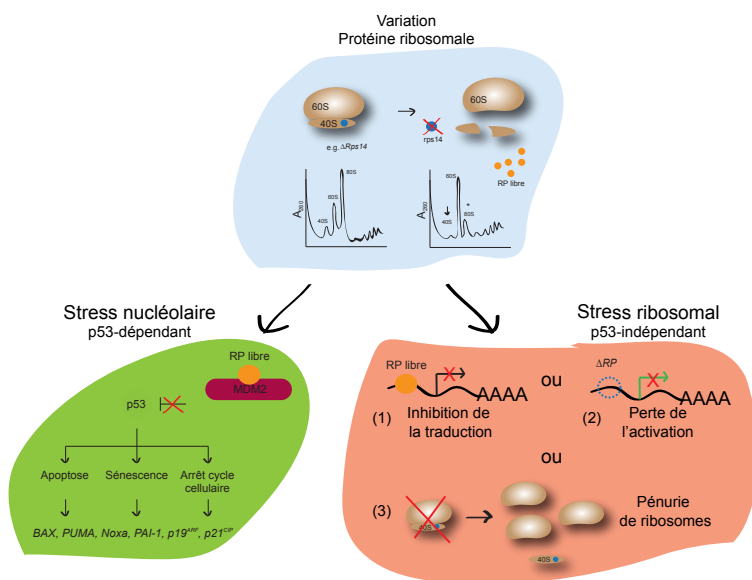


Figure 4.2: Modèle de stress nucléolaire et ribosomal.

La littérature fait déjà état d'une signalisation p53-dépendante (zone verte) suite à une perturbation de la biogénèse des ribosomes (zone bleue). À la lumière de nos résultats relatifs à des délétions hémizygotiques pour *Rps5*, *Rps14* et *Rps28*, une voie alternative, p53-indépendante, est possiblement impliquée aussi lors de ce processus (zone orangée). Comme pistes d'explication, on peut imaginer divers scénarios possibles: (1) une protéine ribosomale libre suite au stress se lie à des ARNm spécifiques et inhibe leur traduction; (2) une protéine ribosomale déletée ne peut plus lier efficacement certains ARNm pour activer leur traduction; (3) le déséquilibre des sous-unités ribosomales mène à un nombre insuffisant de ribosomes devant assurer la traduction.

En plus, à la lumière de nos observations rapportant une corrélation entre la sévérité du phénotype de différenciation et la proximité des protéines ribosomales par rapport au site de sortie des ARNm (Figure 4.3), il est aussi possible d’imaginer que des interactions impliquant globalement les protéines ribosomales de cette région avec l’ARNm jouent un rôle de premier plan dans la régulation traductionnelle.

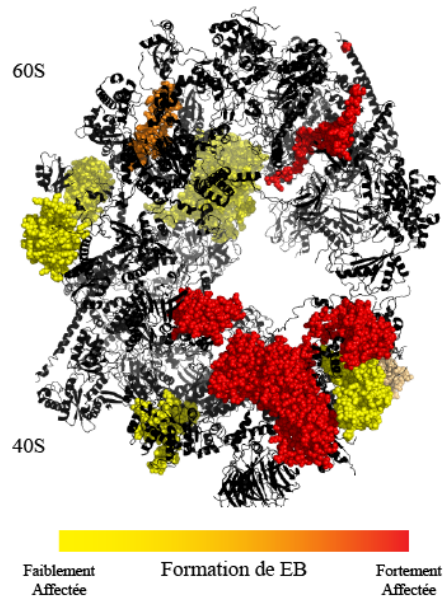


Figure 4.3 Sévérité des phénotypes de formation de EB.

Les protéines ribosomales correspondant aux gènes délétés chez des clones de DeIES sont représentées sur la structure du ribosomes selon une échelle de couleur témoignant de l’intensité du phénotype de différenciation.

Bien entendu, de nombreuses expériences devront être conduites, en combinant aux approches génomiques rapportées des méthodes protéomiques, permettre ainsi d’approcher la question en tenant compte, possiblement de façon dynamique, des niveaux de protéines en lien avec l’expression de certains ARNm. Aussi, il est important de noter que pour l’instant, nos observations quant à l’importance des protéines ribosomales présentes au site de sortie des ARNm en rapport avec les phénotypes de différenciation p53-indépendent reposent sur les études conduites sur *Rps5* et sont extrapolées aux gènes *Rps14* et *Rps28* qui semblent produire un phénotype similaire. Il sera bien entendu intéressant de s’attarder à la signature moléculaire des autres RP retrouvées dans cette région, en comparaison à celles qui sont localisées ailleurs dans le ribosome. Ces expériences supplémentaires sont primordiales à plus forte raison encore dans le cas des études de profilage polysomal, présentées somme toute de façon succincte dans le Chapitre 3. Afin de pouvoir généraliser l’impact des modulations de l’expression des protéines ribosomales sur l’activité de traduction (comparaison des fractions monosomes et polysomes),

il sera en effet important d'obtenir des données de séquençage pour un plus grand nombre de clones afin d'améliorer notre seuil de détection d'expression différentielle. Pour l'instant, il s'agit d'une des limites de notre analyse.

4.2.3 Rmrp

Identifié lors de nos expériences de profilage ribosomal dans les EB sauvages versus ceux hétérozygotes pour *Rps5*, il sera intéressant de poursuivre dans le futur nos investigations sur l'ARN non-codant Rmrp. En effet, cet ARN, constitué de 267 nucléotides, est un petit ARN nucléolaire (snoARN) pour lequel des mutations sont connues dans le cas du syndrome humain de l'hypoplasie cartilage-cheveux (Cartilage-Hair Hypoplasia, CHH)¹⁵². Le complexe ribonucléoprotéique dont fait parti cet ARN non-codant joue un rôle dans la maturation des pre-ARN ribosomiaux dans le nucléole¹⁸⁹. Toutefois, de façon surprenante, Rmrp a aussi le potentiel de former un complexe avec la sous-unité catalytique de la télomérase, la reverse transcriptase TERT, et ainsi participer à la génération d'ARN double brins pouvant servir, suite à leur clivage dans le complexe DICER, à un processus d'interférence à l'ARN¹⁹⁰. Le fait que l'expression de cet ARN soit diminuée suite à une perturbation de *Rps5* permet d'émettre l'hypothèse selon laquelle certains gènes normalement réprimés en contexte normal ne le sont plus de façon efficace et sont responsables de l'apparition d'un phénotype anormal. À la lumière de ces données, il serait intéressant de s'attarder plus longuement aux gènes ayant subi une surexpression suite à la délétion de *Rps5*, bien que les analyses réalisées jusqu'à maintenant ne montrent qu'un groupe très restreint de gènes remplissant ces critères.

4.2.4 Autres déterminants identifiés dans le criblage

Tel que mentionné précédemment, notre choix d'étudier le rôle des RP dans la différenciation des ESC fut motivé par l'enrichissement considérable de ces gènes chez des clones ayant des phénotypes anormaux. Toutefois, en s'attardant aux autres catégories GO les plus significativement affectées par des délétions (Tableau 4.1), le rôle d'autres groupes de gènes affectant la pluripotence des ESC mériterait potentiellement d'être investigué.

Tableau 4.1: Termes d'ontologie génique les plus significatifs des gènes affectant la formation de EB

Identité GO	Description	p	Gènes délétés
GO:0015935	Petite sous-unité ribosomale	0.0000006	Dap3, Gnb2l1, Mrps12, Mrps2, Rps16, Rps18, Rps28, Rps3, Rps5, Rps9
GO:0030136	Vésicules tapissées de clathrine	0.0000021	Aqp2, Clint1, Clta, Clvs2, Gga2, Hax1, Hip1, Picalm, Ston2, Vamp3
GO:0022627	Petite sous-unité ribosomale cytosolique	0.0000026	Rps11, Rps12, Rps14, Rps16, Rps17, Rps18, Rps19, Rps27, Rps28, Rps3, Rps5, Rps9
GO:0001938	Régulation positive de la prolifération des cellules endothéliales	0.0000861	Acvr11, Arg1, Ccl24, Ccl26, Ecm1, Fgfr3, Flt4, Mtor, Nr4a1, Pdgfb, Prkd2, Rptor, Vegfa
GO:0034765	Régulation du transport membranaire des ions	0.0001826	Cacna1a, Cacna1c, Cacna1e, Cacna1i, Cacna1s, Cacna2d2, Cacna2d4, Cacng2, Cacng3, Clic1, Clic4, Clic5, Hcn3, Kcna7, Kcnab2, Kcnc1, Kcnc3, Kcnc3, Kcnh3, Kcnip1, Kcnj11, Kcnj16, Kcnj2, Kcnj4, Kcnj9, Kcnk13, Kcnk9, Scn1b, Scn8a, Shkbp1, Tpen1, Trpm4
GO:0050681	Attachement au récepteur androgénique	0.0002008	Calr, Daxx, Ep300, Kdm3a, Pkn1, Prkcb, Prmt2, Rnf6, Tcf21, Trim68, Wipi1
GO:0043197	Épines dendritiques	0.0002947	Accn2, Apbb1, Atp1a2, Cald1, Canx, Chrna7, Fbxo2, Gabbr1, Gipcl, Grid2ip, Grin2a, Grm5, Lama2, Myh10, Mylk2, P2rx6, Shank1, Slc8a2, Strn4, Synpo, Tmem90b
GO:0051480	Homéostasie du calcium cytosolique	0.0003318	Cdh23, F2, Pde6a, Tgfa, Trpc1
GO:0007015	Organisation des filaments d'actine	0.0003537	Arhgap17, Arhgef2, Crhr2, Cryaa, Enah, Inpp11, Nck1, Pcdh15, Ppargc1b, Racl, Srf
GO:0001974	Remodelage des vaisseaux sanguins	0.0003693	Adralb, Atg5, Axl, Bax, Cbs, Eln, Foxs1, Hoxa3, Nos3, Vegfa

Par exemple, plusieurs gènes regroupés sous l'ontologie génique «Régulation positive de la prolifération des cellules épithéliales» font partie de la voie de signalisation mTor. Étant donné que deux kinases (S6K1 et S6K2) responsables de la phosphorylation de protéines impliquées dans la biogénèse des ribosomes sont des régulateurs traductionnels qu'on retrouve dans la cascade de signalisation mTORC1 (rapamycine dépendante)¹⁹¹, il serait justifié de comparer les effecteurs moléculaires médiant les effets négatifs de ces clones en comparaison à ceux obtenus dans le contexte de délétions de RP. De façon intéressante, des expériences menées sur des cellules déplétées de Rps6 ou Rps23 ont montré que l'effet inhibiteur de la rapamycine sur la traduction de certains mRNA spécifiques n'était pas aussi prononcé que chez les cellules contrôles¹⁶⁴, suggérant ainsi que des fonctions moléculaires distinctes propres aux deux groupes de gènes identifiés lors du criblage sont potentiellement impliquées dans la formation de EB.

Un autre terme d'ontologie génique associé à un phénotype anormal de différenciation qui sort du lot est celui regroupant des gènes codants pour des protéines impliquées dans la formation de vésicules tapissées de clathrine. Ces invaginations, que l'on peut retrouver à la surface cellulaire, ont comme fonction principale l'endocytose de molécules et récepteurs membranaires. Elles ont été identifiées comme jouant un rôle important dans la prolifération et la différenciation de certain types de cellules souches en régulant, entre autre, l'internalisation des facteurs de la voie de signalisation Delta/Notch^{192,193}. Il serait donc pertinent de s'attarder aux délétions hétérozygotes de gènes impliqués dans la formation de ces vésicules dans un contexte de

différenciation et de pluripotence chez les ESC.

Finalement, certains gènes délétés affectant de manière significative la formation de EB codent pour des protéines faisant partie du complexe Mediator, un coactivateur transcriptionnel retrouvé à de nombreux promoteurs des cellules eucaryotes¹⁹⁴. Plusieurs membres de ce complexe ont été identifiés comme étant nécessaires au maintien des marqueurs spécifiques des ESC, dont *Med17*⁷², que l'on retrouve aussi au sein d'une famille de clones (5017) ayant un phénotype de différenciation anormal. De plus, une délétion à *Med18* (famille 5068) produit un effet similaire dans notre contexte expérimental. Bien que l'implication de ces gènes dans le maintien des propriétés caractéristiques des ESC soit déjà connue, il est très intéressant de noter qu'au moins deux d'entre eux sont potentiellement haploinsuffisant pour la différenciation des ESC.

4.3 Conclusion

Les travaux de recherches présentés dans cette thèse ont permis de mettre à la disposition de la communauté scientifique un outil permettant, par une approche de génomique fonctionnelle, l'identification des facteurs essentiels à l'auto-renouvellement et à la différenciation des ESC. En utilisant cette ressource, nous avons dans un premier temps démontré que les clones comportant des délétions chromosomales hétérozygotes pouvaient être utilisés pour mettre à jour des déterminants qui influencent le destin cellulaire des ESC.

Ensuite, en criblant cette librairie pour déterminer le potentiel de ces cellules à former des EB, nous avons identifié les gènes codants pour des protéines ribosomales comme étant essentiels à la différenciation des ESC. Des études plus approfondies portant sur les effets d'une modulation de l'expression de certaines RP ont permis de relever un processus p53-indépendant en réponse à un stress nucléolaire, ce qui vient ainsi documenter un mécanisme de réponse encore peu apprécié dans la littérature.

Finalement, les recherches à venir devront s'attarder à décrire cette voie de signalisation, en plus de tenter de faire un lien entre l'expression génique et celle au niveau des protéines en fonction de la régulation de l'activité traductionnelle. De meilleures connaissances au niveau de cette dynamique moléculaire permettront potentiellement de raffiner les méthodes de différenciation des ESC ou iPSC.

BIBLIOGRAPHIE

1. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1441–1445 (1975).
2. Buta, C. *et al.* Reconsidering pluripotency tests: do we still need teratoma assays? *Stem Cell Res* **11**, 552–562 (2013).
3. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150–7160 (2004).
4. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
5. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634–7638 (1981).
6. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255–256 (1984).
7. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323**, 445–448 (1986).
8. Doetschman, T. *et al.* Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576–578 (1987).
9. Evans, M. Discovering pluripotency: 30 years of mouse embryonic stem cells. *Nature Publishing Group* **12**, 680–686 (2011).
10. Ben-David, U., Kopper, O. & Benvenisty, N. Expanding the boundaries of embryonic stem cells. *Cell Stem Cell* **10**, 666–677 (2012).
11. Kaufman, M. H. *The Atlas of Mouse Development*. (1992).
12. Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661–680 (2008).
13. Tam, P. P. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3–25 (1997).
14. Kispert, A. & Herrmann, B. G. Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Developmental Biology* **161**, 179–193 (1994).
15. Hart, A. H. *et al.* Mixl1 is required for axial mesendoderm morphogenesis and patterning

- in the murine embryo. *Development* **129**, 3597–3608 (2002).
16. Nostro, M. C., Cheng, X., Keller, G. M. & Gadue, P. Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. *Cell Stem Cell* **2**, 60–71 (2008).
 17. Gadue, P., Huber, T. L., Paddison, P. J. & Keller, G. M. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16806–16811 (2006).
 18. Kinder, S. J. *et al.* The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development* **128**, 3623–3634 (2001).
 19. Filosa, S. *et al.* Goosecoid and HNF-3beta genetically interact to regulate neural tube patterning during mouse embryogenesis. *Development* **124**, 2843–2854 (1997).
 20. Kinder, S. J. *et al.* The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691–4701 (1999).
 21. Gadue, P., Huber, T. L., Nostro, M. C., Kattman, S. & Keller, G. M. Germ layer induction from embryonic stem cells. *Experimental Hematology* **33**, 955–964 (2005).
 22. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
 23. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).
 24. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & Development* **12**, 2048–2060 (1998).
 25. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development* **17**, 126–140 (2003).
 26. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372–376 (2000).
 27. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
 28. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).
 29. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling

- pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118–122 (2009).
30. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
 31. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
 32. Shi, Y. & Massagué, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700 (2003).
 33. Cadigan, K. M. Wnt signaling--20 years and counting. in *Trends Genet.* **18**, 340–342 (2002).
 34. Young, R. A. Control of the Embryonic Stem Cell State. *Cell* **144**, 940–954 (2011).
 35. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes & Development* **22**, 746–755 (2008).
 36. Niwa, H. How is pluripotency determined and maintained? *Development* **134**, 635–646 (2007).
 37. Pesce, M. & Schöler, H. R. Oct-4: gatekeeper in the beginnings of mammalian development. *Stem Cells* **19**, 271–278 (2001).
 38. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
 39. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).
 40. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology* **9**, 625–635 (2007).
 41. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
 42. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
 43. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol* **4**, a008128 (2012).
 44. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture.* *Nat. Biotechnol.* **21**, 183–186 (2003).
 45. Gouon-Evans, V. *et al.* BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat. Biotechnol.* **24**, 1402–1411 (2006).
 46. D'Amour, K. A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat. Biotechnol.* **24**, 1392–1401 (2006).

47. Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
48. Wang, Y., Yates, F., Naveiras, O., Ernst, P. & Daley, G. Q. Embryonic stem cell-derived hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 19081–19086 (2005).
49. Burt, R. K. *et al.* Embryonic stem cells as an alternate marrow donor source: engraftment without graft-versus-host disease. *J. Exp. Med.* **199**, 895–904 (2004).
50. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27–45 (1985).
51. Wiles, M. & Keller, G. Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Development* **111**, 259–267 (1991).
52. Doulatov, S. & Daley, G. Q. A Stem Cell Perspective on Cellular Engineering. *Science* **342**, 700–702 (2013).
53. Daley, G. Q. & Scadden, D. T. Prospects for stem cell-based therapy. *Cell* **132**, 544–548 (2008).
54. Briggs, R. & King, T. J. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc. Natl. Acad. Sci. U.S.A.* **38**, 455–463 (1952).
55. Gurdon, J. B., ELSDALE, T. R. & FISCHBERG, M. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**, 64–65 (1958).
56. Gurdon, J. B. & Uehlinger, V. 'Fertile' intestine nuclei. *Nature* **210**, 1240–1241 (1966).
57. Gurdon, J. B. & Byrne, J. A. The first half-century of nuclear transplantation. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8048–8052 (2003).
58. Jullien, J., Pasque, V., Halley-Stott, R. P., Miyamoto, K. & Gurdon, J. B. Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? *Nature Publishing Group* **12**, 453–459 (2011).
59. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
60. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Current Biology* **11**, 1553–1558 (2001).
61. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).
62. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic

- and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
63. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
 64. Pfaff, N. & Cantz, T. From skin to blood: a new member joins the iClub. *Cell Stem Cell* **13**, 131–133 (2013).
 65. Lucas, D. & Frenette, P. S. Stem cells: Reprogramming finds its niche. *Nature* **511**, 301–302 (2014).
 66. Rubin, L. L. Stem cells and drug discovery: the beginning of a new era? *Cell* **132**, 549–552 (2008).
 67. Pritsker, M., Ford, N. R., Jenq, H. T. & Lemischka, I. R. Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6946–6951 (2006).
 68. Adams, D. J. *et al.* Mutagenic insertion and chromosome engineering resource (MICER). *Nat Genet* **36**, 867–871 (2004).
 69. Skarnes, W. C. *et al.* A public gene trap resource for mouse functional genomics. *Nat Genet* **36**, 543–544 (2004).
 70. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nature Publishing Group* (2014). doi:10.1038/nrg3722
 71. Macrae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
 72. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
 73. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes & Development* **23**, 837–848 (2009).
 74. Fazio, T. G., Huff, J. T. & Panning, B. An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. *Cell* **134**, 162–174 (2008).
 75. Wang, C.-H. *et al.* A shRNA functional screen reveals Nme6 and Nme7 are crucial for embryonic stem cell renewal. *Stem Cells* **30**, 2199–2211 (2012).
 76. Pietro Genovese *et al.* Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235–240 (2015).
 77. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49–55 (2013).
 78. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat Meth* **10**, 957–963 (2013).
 79. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of

- bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
80. Baker, M. Gene-editing nucleases. *Nat Meth* **9**, 23–26 (2012).
81. Platt, R. J. *et al.* CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* 1–16 (2014). doi:10.1016/j.cell.2014.09.014
82. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
83. Chen, Y. *et al.* Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat Genet* **24**, 314–317 (2000).
84. Ramírez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
85. Yu, Y. & Bradley, A. Engineering chromosomal rearrangements in mice. *Nat. Rev. Genet.* **2**, 780–790 (2001).
86. Su, H., Wang, X. & Bradley, A. Nested chromosomal deletions induced with retroviral vectors in mice. *Nat Genet* **24**, 92–95 (2000).
87. Bilodeau, M., Girard, S., Hébert, J. & Sauvageau, G. A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Meth* **4**, 263–268 (2007).
88. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* 1–5 (2011). doi:10.1038/nature10448
89. Yang, H. *et al.* Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell* **149**, 605–617 (2012).
90. Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* (2012). doi:10.1038/nature11435
91. Leeb, M., Dietmann, S., Paramor, M., Niwa, H. & Smith, A. Short Article. *Stem Cell* **14**, 385–393 (2014).
92. Dutt, S. *et al.* Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood* **117**, 2567–2576 (2011).
93. Fortier, S. *et al.* Genome-wide interrogation of Mammalian stem cell fate determinants by nested chromosome deletions. *PLoS Genet.* **6**, e1001241 (2010).
94. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
95. Sampath, P. *et al.* A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* **2**, 448–460 (2008).

96. Ye, J. & Blelloch, R. Regulation of Pluripotency by RNA Binding Proteins. *Stem Cell* **15**, 271–280 (2014).
97. James, A., Wang, Y., Raje, H., Rosby, R. & DiMario, P. Nucleolar stress with and without p53. *Nucleus* **5**, (2014).
98. Olson, M. O. J. & Dundr, M. The moving parts of the nucleolus. *Histochem. Cell Biol.* **123**, 203–216 (2005).
99. Ben-Shem, A. *et al.* The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* **334**, 1524–1529 (2011).
100. Jenner, L. *et al.* Crystal structure of the 80S yeast ribosome. *Curr. Opin. Struct. Biol.* **22**, 759–767 (2012).
101. Fromont-Racine, M., Senger, B., Saveanu, C. & Fasiolo, F. Ribosome assembly in eukaryotes. *Gene* **313**, 17–42 (2003).
102. Tafforeau, L. *et al.* The Complexity of Human Ribosome Biogenesis Revealed by Systematic Nucleolar Screening of Pre-rRNA Processing Factors. *Molecular Cell* **51**, 539–551 (2013).
103. Hein, N., Hannan, K. M., George, A. J., Sanij, E. & Hannan, R. D. The nucleolus: an emerging target for cancer therapy. *Trends Mol Med* (2013). doi:10.1016/j.molmed.2013.07.005
104. Zhang, Q., Shalaby, N. A. & Buszczak, M. Changes in rRNA Transcription Influence Proliferation and Cell Fate Within a Stem Cell Lineage. *Science* **343**, 298–301 (2014).
105. Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411–418 (2003).
106. Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115**, 3196–3205 (2010).
107. Warner, J. R. & McIntosh, K. B. How Common Are Extraribosomal Functions of Ribosomal Proteins? *Molecular Cell* **34**, 3–11 (2009).
108. Wool, I. G. Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21**, 164–165 (1996).
109. O’Donohue, M.-F., Choesmel, V., Faubladiere, M., Fichant, G. & Gleizes, P.-E. Functional dichotomy of ribosomal proteins during the synthesis of mammalian 40S ribosomal subunits. *The Journal of Cell Biology* **190**, 853–866 (2010).
110. Robledo, S. *et al.* The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* **14**, 1918–1929 (2008).
111. Horos, R. *et al.* Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood*

- 119**, 262–272 (2012).
112. Karbstein, K. Quality control mechanisms during ribosome maturation. *Trends Cell Biol.* **23**, 242–250 (2013).
 113. Ferreira-Cerca, S., Pöll, G., Gleizes, P.-E., Tschochner, H. & Milkereit, P. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Molecular Cell* **20**, 263–275 (2005).
 114. Campbell, M. G. & Karbstein, K. Protein-protein interactions within late pre-40S ribosomes. *PLoS ONE* **6**, e16194 (2011).
 115. Lam, Y. W., Lamond, A. I., Mann, M. & Andersen, J. S. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Current Biology* **17**, 749–760 (2007).
 116. Lindström, M. S. Emerging functions of ribosomal proteins in gene-specific transcription and translation. *Biochemical and Biophysical Research Communications* **379**, 167–170 (2009).
 117. Perry, R. P. Balanced production of ribosomal proteins. *Gene* **401**, 1–3 (2007).
 118. Donati, G. *et al.* The balance between rRNA and ribosomal protein synthesis up- and downregulates the tumour suppressor p53 in mammalian cells. *Oncogene* **30**, 3274–3288 (2011).
 119. Kressler, D. *et al.* Synchronizing Nuclear Import of Ribosomal Proteins with Ribosome Assembly. *Science* **338**, 666–671 (2012).
 120. Bachellerie, J. P., Cavaillé, J. & Hüttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775–790 (2002).
 121. Bleichert, F. & Baserga, S. J. The long unwinding road of RNA helicases. *Molecular Cell* **27**, 339–352 (2007).
 122. Kressler, D., Hurt, E. & Bassler, J. Driving ribosome assembly. *Biochim. Biophys. Acta* **1803**, 673–683 (2010).
 123. Granneman, S., Kudla, G., Petfalski, E. & Tollervy, D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences* **106**, 9613–9618 (2009).
 124. Ulbrich, C. *et al.* Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell* **138**, 911–922 (2009).
 125. Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal Structure of the Eukaryotic Ribosome. *Science* **330**, 1203–1209 (2010).
 126. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730–

- 736 (2011).
127. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* 1–7 (2013). doi:10.1038/nature12355
 128. Kurkcuoglu, O., Doruker, P., Sen, T. Z., Kloczkowski, A. & Jernigan, R. L. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Phys Biol* **5**, 046005 (2008).
 129. Garreau de Loubresse, N. *et al.* Structural basis for the inhibition of the eukaryotic ribosome. *Nature* (2014). doi:10.1038/nature13737
 130. Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391–394 (2006).
 131. Devaraj, A., Shoji, S., Holbrook, E. D. & Fredrick, K. A role for the 30S subunit E site in maintenance of the translational reading frame. *RNA* **15**, 255–265 (2009).
 132. Voigts-Hoffmann, F., Klinge, S. & Ban, N. Structural insights into eukaryotic ribosomes and the initiation of translation. *Curr. Opin. Struct. Biol.* (2012). doi:10.1016/j.sbi.2012.07.010
 133. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113–127 (2010).
 134. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
 135. Boulon, S., Westman, B. J., Hutten, S., Boisvert, F.-M. & Lamond, A. I. The Nucleolus under Stress. *Molecular Cell* **40**, 216–227 (2010).
 136. Boisvert, F.-M., van Koningsbruggen, S., Navascués, J. & Lamond, A. I. The multifunctional nucleolus. *Nat Rev Mol Cell Biol* **8**, 574–585 (2007).
 137. Rubbi, C. P. & Milner, J. Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. *The EMBO Journal* **22**, 6068–6077 (2003).
 138. Llanos, S., Clark, P. A., Rowe, J. & Peters, G. Stabilization of p53 by p14ARF without relocation of MDM2 to the nucleolus. *Nature Cell Biology* **3**, 445–452 (2001).
 139. Colombo, E., Alcalay, M. & Pelicci, P. G. Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases. *Oncogene* **30**, 2595–2609 (2011).
 140. Teng, T., Thomas, G. & Mercer, C. A. Growth control and ribosomopathies. *Current Opinion in Genetics & Development* 1–9 (2013). doi:10.1016/j.gde.2013.02.001
 141. McCann, K. L. & Baserga, S. J. Genetics. Mysterious ribosomopathies. *Science* **341**, 849–850 (2013).
 142. Liu, J. M. & Ellis, S. R. Ribosomes and marrow failure: coincidental association or

- molecular paradigm? *Blood* **107**, 4583–4588 (2006).
143. Zhang, Y. *et al.* Defects of protein production in erythroid cells revealed in a zebrafish Diamond-Blackfan anemia model for mutation in RPS19. *Cell Death Dis* **5**, e1352 (2014).
 144. Flygare, J. *et al.* Deficiency of ribosomal protein S19 in CD34+ cells generated by siRNA blocks erythroid development and mimics defects seen in Diamond-Blackfan anemia. *Blood* **105**, 4627–4634 (2005).
 145. Singh, S. A. *et al.* p53-Independent Cell Cycle and Erythroid Differentiation Defects in Murine Embryonic Stem Cells Haploinsufficient for Diamond Blackfan Anemia-Proteins: RPS19 versus RPL5. *PLoS ONE* **9**, e89098 (2014).
 146. Garçon, L. *et al.* Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients. *Blood* (2013). doi:10.1182/blood-2013-01-478321
 147. Pereboom, T. C., van Weele, L. J., Bondt, A. & MacInnes, A. W. A zebrafish model of dyskeratosis congenita reveals hematopoietic stem cell formation failure resulting from ribosomal protein-mediated p53 stabilization. *Blood* **118**, 5458–5465 (2011).
 148. Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).
 149. Yip, B. H. *et al.* Effects of L-leucine in 5q- syndrome and other RPS14-deficient erythroblasts. **26**, 2154–2158 (2012).
 150. Wong, C. C., Traynor, D., Basse, N., Kay, R. R. & Warren, A. J. Defective ribosome assembly in Shwachman-Diamond syndrome. *Blood* **118**, 4305–4312 (2011).
 151. Burwick, N., Coats, S. A., Nakamura, T. & Shimamura, A. Impaired ribosomal subunit association in Shwachman-Diamond syndrome. *Blood* (2012). doi:10.1182/blood-2012-04-420166
 152. Ridanpää, M. *et al.* Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* **104**, 195–203 (2001).
 153. Bonafé, L., Schmitt, K., Eich, G., Giedion, A. & Superti-Furga, A. RMRP gene sequence analysis confirms a cartilage-hair hypoplasia variant with only skeletal manifestations and reveals a high density of single-nucleotide polymorphisms. *Clin. Genet.* **61**, 146–151 (2002).
 154. Thiel, C. T. & Rauch, A. The molecular basis of the cartilage-hair hypoplasia-anauxetic dysplasia spectrum. *Best Pract. Res. Clin. Endocrinol. Metab.* **25**, 131–142 (2011).
 155. Jones, N. C. *et al.* Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nature Medicine* **14**, 125–133 (2008).

156. The Treacher Collins Syndrome Collaborative Group *et al.* Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome. *Nat Genet* **12**, 130–136 (1996).
157. Trainor, P. A., Dixon, J. & Dixon, M. J. Treacher Collins syndrome: etiology, pathogenesis and prevention. *Eur J Hum Genet* **17**, 275–283 (2008).
158. Ruggero, D. & Pandolfi, P. P. Does the ribosome translate cancer? *Nat. Rev. Cancer* **3**, 179–192 (2003).
159. Draptchinskaia, N. *et al.* The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet* **21**, 169–175 (1999).
160. J, D., P, T. & MJ, D. Treacher Collins syndrome. *Orthod Craniofacial Res* 1–8 (2007).
161. Bursac, S. *et al.* Mutual protection of ribosomal proteins L5 and L11 from degradation is essential for p53 activation upon ribosomal biogenesis stress. *Proceedings of the National Academy of Sciences* (2012). doi:10.1073/pnas.1218535109
162. Zhou, X., Hao, Q., Liao, J., Zhang, Q. & Lu, H. Ribosomal protein S14 unties the MDM2-p53 loop upon ribosomal stress. *Oncogene* (2012). doi:10.1038/onc.2012.63
163. Deisenroth, C. & Zhang, Y. The Ribosomal Protein-Mdm2-p53 Pathway and Energy Metabolism: Bridging the Gap between Feast and Famine. *Genes & Cancer* **2**, 392–403 (2011).
164. Fumagalli, S. *et al.* Absence of nucleolar disruption after impairment of 40S ribosome biogenesis reveals an rpL11-translation-dependent mechanism of p53 induction. *Nature Cell Biology* **11**, 501–508 (2009).
165. McGowan, K. A. *et al.* Reduced ribosomal protein gene dosage and p53 activation in low-risk myelodysplastic syndrome. *Blood* **118**, 3622–3633 (2011).
166. Arava, Y. *et al.* Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3889–3894 (2003).
167. Ingolia, N. T., Ghaemmighami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
168. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Publishing Group* (2014). doi:10.1038/nrg3645
169. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).
170. Bywater, M. J. *et al.* Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer Cell* **22**, 51–65 (2012).

171. Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M. V. Hematopoietic commitment during embryonic stem cell differentiation in culture. *Molecular and Cellular Biology* **13**, 473–486 (1993).
172. Sturgeon, C. M. *et al.* Primitive erythropoiesis is regulated by miR-126 via nonhematopoietic Vcam-1+ cells. *Developmental Cell* **23**, 45–57 (2012).
173. Lefebvre, L., Dionne, N., Karaskova, J., Squire, J. A. & Nagy, A. Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nat Genet* **27**, 257–258 (2001).
174. Huarte, M. & Rinn, J. L. Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19**, R152–61 (2010).
175. Kumar, M. S. *et al.* Coordinate loss of a microRNA and protein-coding gene cooperate in the pathogenesis of 5q- syndrome. *Blood* **118**, 4666–4673 (2011).
176. Batliner, J., Buehrer, E., Fey, M. F. & Tschan, M. P. Inhibition of the miR-143/145 cluster attenuated neutrophil differentiation of APL cells. *Leukemia Research* 1–4 (2011). doi:10.1016/j.leukres.2011.10.006
177. Matsson, H. *et al.* Targeted disruption of the ribosomal protein S19 gene is lethal prior to implantation. *Molecular and Cellular Biology* **24**, 4032–4037 (2004).
178. Deisenroth, C. & Zhang, Y. Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. *Oncogene* **29**, 4253–4260 (2010).
179. Zhang, Y. & Lu, H. Signaling to p53: Ribosomal Proteins Find Their Way. *Cancer Cell* **16**, 369–377 (2009).
180. Danilova, N., Sakamoto, K. M. & Lin, S. Ribosomal protein S19 deficiency in zebrafish leads to developmental abnormalities and defective erythropoiesis through activation of p53 protein family. *Blood* **112**, 5228–5237 (2008).
181. Torihara, H. *et al.* Erythropoiesis failure due to RPS19 deficiency is independent of an activated Tp53 response in a zebrafish model of Diamond-Blackfan anaemia. *Br J Haematol* **152**, 648–654 (2011).
182. Russo, A. *et al.* Human rpL3 induces G₁/S arrest or apoptosis by modulating p21 (waf1/cip1) levels in a p53-independent manner. **12**, 76–87 (2013).
183. Provost, E. *et al.* Ribosomal biogenesis genes play an essential and p53-independent role in zebrafish pancreas development. *Development* **139**, 3232–3241 (2012).
184. Jaako, P. *et al.* Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. *Blood* (2012). doi:10.1182/blood-2012-05-431437
185. Signer, R. A. J., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* 1–16 (2014). doi:10.1038/

nature13035

186. Kondrashov, N. *et al.* Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning. *Cell* **145**, 383–397 (2011).
187. Xue, S. *et al.* RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature* (2014). doi:10.1038/nature14010
188. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol* **13**, 355–369 (2012).
189. Topper, J. N. & Clayton, D. A. Characterization of human MRP/Th RNA and its nuclear gene: full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP. *Nucleic Acids Research* **18**, 793–799 (1990).
190. Maida, Y. *et al.* An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **461**, 230–235 (2009).
191. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149**, 274–293 (2012).
192. Furthauer, M. & Gonzalez-Gaitan, M. Endocytosis, asymmetric cell division, stem cells and cancer: Unus pro omnibus, omnes pro uno. *Molecular Oncology* **3**, 339–353 (2009).
193. Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006).
194. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).