

Université de Montréal

**Development of a retroviral strategy that efficiently creates  
nested chromosomal deletions in mouse embryonic stem cells  
and its exploitation for functional genomics**

par  
Mélanie Bilodeau

Programmes de biologie moléculaire, Université de Montréal  
Faculté des études supérieures

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de doctorat  
en biologie moléculaire

Avril, 2007

© Mélanie Bilodeau, 2007



QH

506

U54

2007

v.015

## AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

## NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée:

**Development of a retroviral strategy that efficiently creates nested chromosomal deletions in mouse embryonic stem cells and its exploitation for functional genomics**

présentée par :  
Mélanie Bilodeau

a été évaluée par un jury composé des personnes suivantes :

Dr Trang Hoang, présidente-rapporteuse  
Dr Guy Sauvageau, directeur de recherche  
Dr Daniel Dufort, membre du jury  
Dr Janet Rossant, examinatrice externe  
Dr Éric Milot, examinateur externe  
Dr Sylvain Meloche, représentant du doyen de la FES

## Résumé

Les délétions chromosomiques chevauchantes sont un outil exploratoire exceptionnel afin d'annoter fonctionnellement le génome de la souris, puisqu'elles permettent de sonder autant les régions codantes que non-codantes. Toutefois jusqu'à maintenant, la création de délétions chromosomales, cartographiables précisément, était laborieuse ce qui limitait leur application à grande échelle. Les travaux présentés dans cette thèse proposent une nouvelle alternative pour créer des délétions chromosomiques à l'échelle du génome entier, dans un délai raisonnable, et applicable tant aux cellules primaires qu'aux lignées cellulaires. Ce système repose sur deux rétrovirus complémentaires insérant des séquences *loxP* dans le génome, qui servent de substrats pour la recombinaison site spécifique catalysée par l'enzyme Cre. La première section de cette thèse (chapitre 2) décrit cette stratégie et le développement de vecteurs rétroviraux compétents pour produire des délétions chromosomales haploïdes dans les cellules souches embryonnaires murines. Ces cellules pluripotentes mutantes ont révélé trois régions haploinsuffisantes requises pour leur différenciation *in vitro* et leur contribution *in vivo* aux souris chimères. Ces expériences validaient les fondements de l'approche. La deuxième section de cette thèse (chapitre 3) rapporte l'exploitation à grande échelle de cette nouvelle méthodologie. Une bibliothèque de plus de 1200 clones de cellules souches embryonnaires, contenant potentiellement des délétions chevauchantes localisées dans le génome entier, a été générée. Ces cellules ont été exploitées lors d'essais fonctionnels. Les résultats préliminaires révèlent plusieurs régions haploinsuffisantes qui seront validées prochainement. Les constructions rétrovirales, ainsi que les lignées de cellules souches mutantes et leurs annotations fonctionnelles, seront accessibles à la communauté scientifique au courant de l'année.

**Mots-clés** : délétions chromosomiques, rétrovirus, Cre-*loxP*, cellules souches embryonnaires, génomique fonctionnelle

## Abstract

Engineered nested chromosomal deletions are a valuable tool to explore the mouse genome functionalities, because they allow the examination of both protein-coding and non protein-coding regions. Up to now however, the generation of precisely localizable chromosomal deletions was laborious, precluding large-scale applications. The work presented in this thesis brings on a new alternative method to create genome-wide chromosomal deletions within a reasonable timeframe and applicable to both primary cells and to cell lines. This system relies on the creation of two compatible retroviruses delivering *loxP* sequences in the genome, the substrates required to perform Cre-induced site-specific recombination. The first section of this thesis (chapter 2) describes the strategy and the development of optimal retroviral vectors that were created to produce haploid chromosomal deletions in mouse embryonic stem cells. These engineered pluripotent cells revealed three haploinsufficient regions required for their proper *in vitro* differentiation and *in vivo* contribution to chimeric mice. These experiments validated the principles of this approach. The second section (chapter 3) provides the first large-scale exploitation of this new methodology. This involved the creation of a library of more than 1200 embryonic stem cell clones containing potential nested chromosomal deletions, localized throughout the mouse genome. The embryonic stem cell clones were used to perform functional screens and preliminary results uncovered numerous haploinsufficient regions that will be validated shortly. The retroviral constructs, the engineered embryonic stem cell lines and their related functional annotations will be accessible to the scientific community within the coming year.

**Keywords:** chromosomal deletions, retroviruses, Cre-*loxP*, embryonic stem cells, functional genomics

## Table of content

Résumé.....	iii
Abstract.....	iv
List of abbreviations .....	xii
Remerciements.....	xvii
<b>Chapter 1 INTRODUCTION AND LITERATURE OVERVIEW .....</b>	<b>1</b>
<b>PART I: PRESENTATION OF RESEARCH OBJECTIVES.....</b>	<b>2</b>
1.1 The genomic content.....	2
1.2 Selection of an experimental model.....	3
1.3 Functional genomic approaches applied to ESCs .....	4
1.3.1 Gene targeting.....	4
1.3.2 Gene trap screens .....	5
1.3.3 shRNA screens .....	5
1.3.4 Other insertional mutagenesis screens .....	6
1.3.5 Chemical screens .....	6
1.3.6 Mutagenesis with oligonucleotides.....	8
1.3.7 Gain-of-function screens.....	8
1.3.8 Irradiation-based screens .....	9
1.3.9 Cre- <i>loxP</i> technology based screens .....	9
1.4 The proposed approach.....	10
<b>PART II: INTRODUCTION TO EMBRYONIC STEM CELLS .....</b>	<b>12</b>
1.5 The origin of ESC .....	12
1.6 Self-renewal and pluripotency of ESCs .....	14
1.6.1 Update for mechanisms underlying ESC self-renewal and pluripotency .....	14
1.7 Differentiation of ESC .....	17
1.7.1 Contribution of ESCs to mice .....	17
1.7.2 <i>In vitro</i> differentiation of ESCs.....	19
1.7.2.1 <i>In vitro</i> differentiation methods .....	19
1.7.2.2 Additional considerations regarding ESC <i>in vitro</i> differentiation.....	22
1.7.2.3 Successfully derived lineages .....	22
1.7.2.4 Comparison of ESC <i>in vitro</i> versus <i>in vivo</i> hematopoietic differentiation ...	23
1.7.3 Generation of teratomas and teratocarcinomas.....	24
1.7.4 General mechanisms underlying ESC differentiation.....	26
1.8 ARTICLE: Cell cycle checkpoints, telomeres and chromosome maintenance in mouse embryonic stem cells.....	28
1.8.1 Author contributions .....	28
1.8.2 Introduction.....	29

1.8.3 Cell cycle regulation differs between ESCs and differentiated cells .....	29
1.8.3.1 G <sub>1</sub> /S transition .....	30
1.8.3.2 S phase .....	33
1.8.3.3 G <sub>2</sub> /M transition.....	33
1.8.3.4 The M phase.....	34
1.8.3.5 Is ESC cycle regulation an artifact of cell culture conditions? .....	34
1.8.4 Telomere maintenance differs between ESCs and their differentiated progenies...	36
1.8.5 Genetic and chromosome anomalies in ESCs: comparison with somatic cells...	38
1.8.6 Concluding remarks .....	41
1.8.7 Methods.....	41
1.8.8 Acknowledgments.....	42
PART III: INTRODUCTION TO RETROVIRUSES.....	43
1.8.9 Structural characteristics.....	43
1.8.10 Retroviral life cycle.....	44
1.8.11 Properties of retroviral vectors.....	45
1.8.11.1 Structure of a basic retroviral vector.....	45
1.8.11.2 Embryonic stem cell viral vectors .....	47
1.8.12 Properties of the packaging cell line.....	47
1.8.12.1 Principles of a packaging cell line .....	47
1.8.12.2 Tropism and pseudotyping.....	48
1.8.13 Transduceable cells .....	48
1.8.14 Helper viruses, satellite viruses, and satellite RNA .....	49
1.8.15 Retroviral integration .....	50
1.8.15.1 Preferential sites of retroviral integration .....	50
1.8.15.2 Molecular mechanism of retroviral integration .....	51
1.8.15.3 Determination of retroviral integration sites.....	51
AIM OF THE THESIS .....	53
<i>Chapter 2</i> APPLICATION OF A NEW RETROVIRAL SYSTEM TO CREATE CHROMOSOMAL DELETIONS IN ESCs .....	54
ARTICLE: A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells .....	55
2.1 Author contributions .....	56
2.2 Abstract .....	56
2.3 Introduction.....	56
2.4 Results.....	57
2.4.1 Selection of anchor and saturation proviruses .....	57
2.4.2 Evaluation of chromosomal deletions .....	60



2.4.3 Interchromosomal recombination events.....	61
2.4.4 <i>In vitro</i> and <i>in vivo</i> differentiation of recombined clones.....	62
2.4.5 Discussion and Conclusions .....	66
2.5 Methods.....	67
2.6 Acknowledgments.....	68
2.7 Supplementary Figures .....	69
2.8 Supplementary Tables.....	76
2.9 Supplementary Methods .....	79
<b>Chapter 3 CREATION OF A LIBRARY OF ENGINEERED ESC CLONES SUITED FOR FUNCTIONAL ASSAYS.....</b>	<b>83</b>
ARTICLE.....	84
DELES: a new library of nested chromosomal DEletions in mouse ES cells suited for functional screens.....	84
3.1 Author contributions .....	85
3.2 Abstract .....	85
3.3 Introduction.....	86
3.4 Results.....	87
3.4.1 Generation of a chromosomal deletion library in ESC clones.....	87
3.4.2 Presentation of a functional screen performed with puromycin <sup>S</sup> tertiary clones.....	95
3.4.3 Global analyses of functional screens.....	99
3.4.4 Specific analyses of functional screens.....	100
3.5 Discussion and conclusions .....	108
3.6 Methods.....	110
3.7 Acknowledgments.....	113
<b>Chapter 4 DISCUSSION AND PERSPECTIVES .....</b>	<b>114</b>
4.1 Haploinsufficiency and imprinting .....	115
4.2 Pending optimizations .....	116
4.2.1 Complementation approaches.....	116
4.2.1.1 Identification of a minimal interval correlating with an abnormal phenotype.....	116
4.2.1.2 Characterization of deleted segments .....	117
4.2.1.3 Re-introduction of deleted DNA.....	117
4.2.1.4 Mapped regions correlating with differentiation anomalies .....	119
4.2.1.5 Characterization of deletions related to family no.9.....	120
4.2.2 Toward a recessive screen.....	130
4.2.3 Detection of ESC-derived progenies <i>in situ</i> .....	131
4.3 Potential applications of the system.....	132

4.4 Thesis conclusion.....	133
REFERENCES (FOR CHAPTERS AND APPENDIXES).....	134
APPENDIXES .....	xviii
APPENDIX I: Article .....	xviii
Uncovering stemness .....	xviii
Author contributions .....	xviii
Abstract.....	xix
News & Views .....	xix
APPENDIX II: TOWARD THE DESIGN OF A SUCESSFUL RETROVIRAL SYSTEM .....	xxiii
ARTICLE: Shunning pitfalls in retroviral vectors design for functional genomics .....	xxiv
Author contributions .....	xxiv
Abstract.....	xxv
Introduction.....	xxv
Results.....	xxvii
Rearranged proviruses .....	xxvii
Transmission of undesired retroviral-like particle .....	xxix
Undesired neomycin expression .....	xxxiii
Discussion and Conclusions .....	xxxv
Methods.....	xxxv
Acknowledgments.....	xxxvii
Supplementary Figure.....	xxxviii
Supplementary Methods .....	xl
APPENDIX III: Table.....	xli
Table presenting the genomic features of speculative 3 Mb-deletions anchored to virus A1 retroviral integration sites determined by I-PCR (part 1 of 4) .....	xli
Part 2 of 4 .....	xlii
Part 3 of 4 .....	xliii
Part 4 of 4 .....	xliv

## List of tables

Table I • Advantages and disadvantages of ESC functional genomic approaches .....	11
Table II • Frequency of genetic anomalies detected in mouse ESC clones.....	40
Table III • Characteristics of independent deletions confirmed by I-PCR and aCGH .....	61
Table IV • Chimera analysis .....	66
Table V • Summary of the Cre-mediated recombination around 11 randomly chosen loci .....	76
Table VI • <i>In vitro</i> differentiation of primary and tertiary clones carrying deletions.....	78
Table VII • Summary of G418 <sup>R</sup> and G418 <sup>R</sup> puro <sup>S</sup> tertiary clone generation. (Part 1 of 4) ..	91
Table VIII • Genomic features found in a region spanning a virtual 3-Mb deletion anchored to the virus A1 retroviral integration site related to family no. 5276.....	109
Table IX • Q-PCR assays employed to detect chromosome 1, 8, 11, and 14 trisomies. ....	112
Table X • Function and expression of RefSeq genes found in family no.9 minimal interval. ....	125
Table XI • Candidate haploid deletions that could be tested for loss of heterozygosity. ...	130
Table XII • Potential applications of retroviral-based Cre- <i>loxP</i> recombination.....	133

## List of figures

Figure 1-1 Early mouse development during the stage of preimplantation. ....	13
Figure 1-2 Generation of chimeric mice using micro-injection or aggregation .....	18
Figure 1-3 The removal of LIF allows the differentiation of ESCs in EBs <i>in vitro</i> . ....	20
Figure 1-4 Effect of seeding density and culture conditions on EBs differentiation. ....	21
Figure 1-5 ESC differentiation in mesodermal and hematopoietic lineages on OP9 stromal layer. ....	22
Figure 1-6 G <sub>1</sub> /S transition and G <sub>1</sub> DNA damage checkpoint in somatic cells. ....	30
Figure 1-7 Cell cycle regulation of undifferentiated and differentiated ESCs. ....	36
Figure 1-8 General structure of a simple replication-competent retrovirus. ....	44
Figure 1-9 Structures of retroviral vector plasmids, proviruses, and transcripts. ....	46
Figure 1-10 I-PCR allows the determination of retroviral integration sites. ....	52
Figure 2-1 Cre-induced chromosomal rearrangements in mouse ESCs. ....	58
Figure 2-2 <i>In vitro</i> and <i>in vivo</i> differentiation of ESC clones with deletions. ....	64
Figure 2-3 Generation of retroviral vectors. ....	69
Figure 2-4 Cre-induced recombination between integrated proviruses. ....	70
Figure 2-5 Display showing confirmed chromosomal deletions in ESCs. ....	72
Figure 2-6 Evaluation of interchromosomal recombination events. ....	73
Figure 2-7 Full-length gels and blots. ....	75
Figure 3-1 Schematic representation of tertiary clone generation. ....	89
Figure 3-2 Primary clone retroviral integration sites located by I-PCR. ....	90
Figure 3-3 Schematic representation of functional assay approaches. ....	96
Figure 3-4 Presentation of cell counts and performed functional screens. ....	97
Figure 3-5 Ki67 global analyses. ....	100
Figure 3-6 DELES database functionalities. ....	101
Figure 3-7 Puro <sup>S</sup> tertiary clones from family no. 5077 present normal phenotypes. ....	104
Figure 3-8 Puro <sup>S</sup> tertiary clones from family no.5278 present normal phenotypes. ....	105
Figure 3-9 Certain puro <sup>S</sup> tertiary clones from family no.5032 present abnormal phenotypes. ....	106
Figure 3-10 Certain puro <sup>S</sup> tertiary clones from family no.5276 present abnormal phenotypes. ....	107
Figure 4-1 Determination of a candidate region associated with an abnormal phenotype. ....	117
Figure 4-2 SKY analysis of rare EB cells derived from tertiary clone 9-104. ....	118
Figure 4-3 Minimal intervals represented for family no.9. ....	120

Figure 4-4 Mouse mutant alleles and mapped phenotypes for family no. 9 minimal interval. .... 123

Figure 4-5 BACs and cDNAs selected to cover clone 9-104 deleted segment. .... 129

Figure 0-1 Candidate transcription factors that determine stem cell identity .....xxii

Figure 0-2 Retroviral constructs were designed to mediate *Cre-loxP* recombination. ....xxvi

Figure 0-3 Rearranged proviruses and transmission of a retroviral-like particle. ....xxx

Figure 0-4 Retroviruses associated with unexpected neomycin resistance. ....xxxiv

Figure 0-5 Successful retroviruses and efficient *Cre-loxP* recombination. ....xxxviii

## List of abbreviations

<b>Abbreviation</b>	<b>Signification</b>
aCGH	Array based comparative genomic hybridization
Akt	Thymoma viral proto-oncogene
APC	Anaphase-promoting complex
Apc2	Adenomatosis polyposis coli 2
Apc11	Adenomatosis polyposis coli 11
Aprt	Adenine phosphoribosyltransferase
ASLV	Avian sarcoma-leukosis virus
ATM	Ataxia telangiectasia mutated
att	Attachment sites
BAF	Barrier-to-autointegration factor
Bmi1	B lymphoma Mo-MLV insertion region 1
BMP	Bone morphogenic protein
BRCA1	Breast cancer 1
bp	Base pairs
BrdUrd	5-bromouridine
Bub3	Budding uninhibited by benzimidazoles 3 homolog
Cdc20	Cell division cycle 20 homolog
cDNA	Complementary deoxyribonucleic acid
CDK	Cyclin dependent kinase
Cdx2	Caudal-related homeobox 2
Chk1	Checkpoint kinase 1 homolog
c-myc	Myc proto-oncogene protein
CpG	C and G nucleotides linked by a phosphodiester bond
DKO	Double knockout
DMD	Differentially methylated domains
DNA	Deoxyribonucleic acid
Dpc	Days postcoitum
Dppa3	Developmental pluripotency-associated 3
Dppa4	Developmental pluripotency-associated 4
E	Embryonic day
EB	Embryoid body
<i>Eed</i>	Embryonic ectoderm development

EMS	Ethylmethanesulphonate
ENU	<i>N</i> -ethyl- <i>N</i> -nitrosourea
Env	Envelop protein
Eras	Embryonic stem cell-expressed Ras
ES	Embryonic stem (cells)
ESC	Embryonic stem cell
Esrrb	Estrogen-receptor-related receptor beta
EUCOMM	European Conditional Mouse Mutagenesis Program
Ezh2	Histone methyltransferase enhancer of zeste homologue 2
E4f1	E4F transcription factor 1
Fgf4	Fibroblast growth factor 4
FIAU	1-(2-deoxy-2-fluoro- $\beta$ -D-arabinofuranaosyl)-5-iodouracil
Gag	Group-specific-antigen protein
G <sub>0</sub>	Gap 0 phase of the cell cycle
G <sub>1</sub>	Gap 1 phase of the cell cycle
G <sub>2</sub>	Gap 2 phase of the cell cycle
G418	Geneticin
G418 <sup>R</sup>	Geneticin resistant
IGTC	International Gene Trap Consortium
iPS-MEFs	Pluripotent stem cells derived from mouse embryonic fibroblasts
iPS-TTFs	Pluripotent stem cells derived from adult tail-tip fibroblasts
HAT	Hypoxanthine Aminopterin Thymidine
HIV-1	Human immunodeficiency virus
HSC	Hematopoietic stem cell
<i>HOX</i>	<i>Homeobox</i>
Hoxb4	Homeobox protein Hox-B4
<i>Hprt1</i>	<i>Hypoxanthine phosphoribosyl transferase</i> gene
hESC	Human embryonic stem cells
Hygro	Hygromycin
Hygro <sup>R</sup>	Hygromycin resistant
Hygro <sup>S</sup>	Hygromycin sensitive
Hygro <sup>S+R</sup>	Hygromycin sensitive and resistant
IKMC	International Knockout Mouse Consortium
I-PCR	Inverse-polymerase chain reaction
Kb	kilobase pairs
Klf4	Kruppel-like factor 4

KOMP	KnockOut Mouse Project
LAP2 $\alpha$	Lamina-associated polypeptide 2 $\alpha$
LEDGF	Lens-epithelium-derived growth factor
LIF	Leukemia inhibitory factor
<i>LoxP</i>	<i>Locus of crossover x in P1</i>
LTR	Long terminal repeats
M	Mitosis phase of the cell cycle
Mad2	Mitotic arrest deficient 2
Mad3	Mitotic arrest deficient 3
March3	Membrane-associated ring finger (C3HC4) 3
Mb	Megabase pairs
MEFs	Mouse embryonic fibroblasts
mESC	Murine embryonic stem cells
MESV	Murine embryonic stem cell virus
miRNA	Micro ribonucleic acid
MLV	Murine leukemia virus
Mo-MLV	Moloney murine leukemia virus
Mre11	Meiotic recombination 11
mRNA	Messenger ribonucleic acid
MSCV	Murine stem cell virus
Nanog	Homeobox transcription factor Nanog
Nbs1	Nijmegen breakage syndrome 1
NorCOMM	North American Conditional Mouse Mutagenesis Project
<i>NS</i>	<i>Nucleostemin</i>
Oct4	Octamer-binding transcription factor 4
pA	Polyadenylation signal
PBS	Transfer ribonucleic acid-binding site
PcG	Polycomb group genes
PCMV	PCC4-cell-passaged myeloproliferative sarcoma virus
PCR	Polymerase chain reaction
PIC	Preintegration complex
PI3K	Phosphatidylinositol-3-OH kinase
Pol	Polymerase
PPT	Polypurine tract
<i>Pro</i>	<i>Protease</i>
Puro	Puromycin



Puro <sup>R</sup>	Puromycin resistant
Puro <sup>S</sup>	Puromycin sensitive
Puro <sup>S+R</sup>	Puromycin sensitive and resistant
Q-PCR	Real-time quantitative polymerase chain reaction
Rad50	DNA repair protein RAD50
Ring1A	Ring finger protein 1
Ring1B	Ring finger protein 2
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
Rtel	Regulation of telomere length
RT-PCR	Reverse transcriptase-polymerase chain reaction
S	Synthesis phase of cell cycle
shRNA	Small hairpin ribonucleic acid
SKY	Spectral karyotyping
Smad	MAD homolog protein
Sox2	SRY-box2
Stat3	Signal transducer and activator of transcription 3
Suz12	Suppressor of zeste 12
Tbx3	T-box protein 3
<i>Tcl1</i>	<i>T-cell leukemia/lymphoma 1</i>
<i>Terc</i>	<i>Telomerase RNA component</i>
Tert	Telomerase reverse transcriptase
TIGM	Texas Institute for Genomic Medicine
Tk	Herpes simplex thymidine kinase
TKO	Triple knockout
TRF1	Telomeric repeat binding factor 1
TRF2	Telomeric repeat binding factor 2
TRIM5 $\alpha$	Tripartite motif-containing 5
tRNA	Transfer ribonucleic acid
TTFs	Adult tail-tip fibroblasts
UTR	Untranslated region
VSV-G	Vesicular Stomatitis Virus G
Wnt	Wingless-related MMTV integration site

*Aux incroyables,*

## Remerciements

J'aimerais d'abord remercier mon superviseur Dr Guy Sauvageau pour son appui inconditionnel tout au long de mes études graduées. Tant son audace que son optimisme auront été des éléments indispensables à la réussite de mon parcours. J'ai bénéficié d'une formation exceptionnelle au sein de son laboratoire, autant conceptuellement que techniquement. Je remercie également mes anciens collègues et mes compatriotes actuels du laboratoire; pour leur aide, leurs encouragements et leur amitié. Je soulignerai la contribution de plusieurs dans le corps de cette thèse. Aussi, je remercie le Dr Richard Martin, ancien étudiant du laboratoire de Dr Trang Hoang, qui m'a initié à la culture des cellules souches embryonnaires. De plus, j'ai profité de l'expertise de plusieurs personnes compétentes que vous découvrirez au fil de la thèse, œuvrant dans divers Services de l'Institut de recherche en immunologie et en oncologie (Montréal), de l'Institut de recherches cliniques de Montréal, de la Banque de cellules leucémiques du Québec (Montréal) et du Roswell Park Cancer Institute (Buffalo). Je suis reconnaissante envers toutes les agences de financement qui ont contribué à ma formation soit par des bourses d'étude ou des fonds de recherche alloués à mon projet : les Fonds de Recherches en Santé du Québec, les Instituts de Recherche en Santé du Canada, Génome Québec, le Réseau de Recherche en Transgénèse du Québec et les Programmes de Biologie moléculaire de l'Université de Montréal. Finalement, je tiens à remercier ma famille, mon conjoint et mes amis pour leur réconfort, leur humour et leur jugement.

# *Chapter 1* INTRODUCTION AND LITERATURE OVERVIEW

Chapter 1 is divided in three sections: the presentation of research objectives (Part I), introduction to embryonic stem cells (Part II), and introduction to retroviruses (Part III). Part II contains two manuscripts related to embryonic stem cell biology: one concerning self-renewal and pluripotency (published News & Views, **Appendix I**) and one reviewing selected genetic characteristics (review in preparation). Author contributions to manuscripts are described in the respective sections.

## **PART I: PRESENTATION OF RESEARCH OBJECTIVES**

When I joined the laboratory, the field of genomics was effervescent. International collaborations were underway for the sequencing and the assembly of diverse genomes including human, mouse, and other model organisms. The field of functional genomics was also rapidly developing. Indeed, a combination of approaches was implemented aiming to link biological functions to sequence information. We decided to venture in this effort by elaborating a methodology that would be complementary to others that were being developed at that time. Over the years, the genomic knowledge evolved and new functional approaches were designed. Still, the methodology described in this thesis subsisted to this active period of time and positioned itself favorably among other expertise.

### **1.1 The genomic content**

The initial analysis of the human genome sequence revealed striking observations. More than 50 % of the human genome consists of repeat sequences, often referred as “junk” DNA, which include: transposable elements, processed pseudogenes (retroposed copies of cellular genes), simple sequence repeats, segmental duplications, and blocks of tandemly repeated sequences<sup>1</sup>. In fact, coding exons and transcript untranslated regions constitute only 1.2 % and 0.7% of the human genome, respectively<sup>2</sup>. Both for human and mouse, an average of 20 000-25 000 protein-coding genes are predicted (excluding non protein-coding RNA)<sup>2</sup>, a number regularly updated with the completion of genome sequencing combined with new computational and experimental data. Ninety-nine percent of mouse genes have homologues in the human genome; 96% of which are found in syntenic regions<sup>3</sup>. Ninety percent of mouse and human genomes present conserved synteny along with 40% of alignment<sup>3</sup>. Several conserved sequences consist of ancestral repeats<sup>3</sup>. However, comparisons between the genetic material of organisms such as mouse, human and dog suggested that 2.5-5% of the mammalian genome has been under evolutionary selection, thus possibly sustaining biological functions<sup>3-5</sup>. These evolutionary conserved elements are thought to represent protein-coding genes, untranslated region of protein-coding genes, regulatory elements, non protein-coding RNA, and chromosomal structural elements<sup>3</sup>. Fifty percent of the highly conserved non-coding elements cluster in ~200 gene-poor regions<sup>5</sup>. Most of the few genes found in these regions establish or maintain cellular identity (transcription factors involved in differentiation and development, axon guidance receptors)<sup>5</sup>. Many of these non-coding elements could regulate gene expression by diverse mechanisms, including long-range epigenetic silencing or higher order genome organization<sup>5,6</sup>. Biological functions and interconnections between

most of these elements still need to be assessed. Obviously, many conserved elements will be acting cooperatively through physical interactions to sustain biological functions. However, it is also expected that some will cooperate functionally toward physiological functions, without physical interactions. This concept, well established in yeast, is referred as synthetic genetic interactions<sup>7</sup>.

Mammalian genome sequencing and comparative sequence analyses highlight the variable distribution of certain genomic features such as genes, transposable elements, GC content, recombination rate, etc<sup>1,3,5</sup>. For example, the most repeat-poor region in the human genome is the *HOX* gene clusters<sup>1</sup>. Additional conserved repeat-poor regions were identified in mouse and human<sup>3</sup>. These repeat-poor regions are potential sites of elevated gene regulation<sup>3</sup>. Another example of non random distribution is the high frequency of segmental duplications, derived from *trans*-chromosomal recombination in pericentromeric and subtelomeric regions<sup>1</sup>. Recombination rates seem higher in distal regions<sup>1</sup>. According to the genome comparison of different species, synteny block breaks seem to correlate with GC content and might be hot spots of recombination, an hypothesis waiting to be addressed<sup>5</sup>.

Finally, an emerging concept is that some regulatory elements demonstrate conservation, not primarily at the level of DNA sequence, but at the level of epigenetic marks such as histone modifications<sup>6</sup>, which can be missed by sequence comparison analysis<sup>6</sup>. Taking together, these observations suggest that the functional genomics remains largely unexplored.

## 1.2 Selection of an experimental model

The mouse is an advantageous model to gain insights into human biology and disease. The mouse was already used in our laboratory to study normal hematopoiesis and leukemia. At that time, we wanted to perform a functional screen *in vitro*, paving the way for an analysis *in vivo*, to identify hematopoietic stem cell regulators.

Mouse embryonic stem cells (ESCs) became our selected model for several reasons. ESCs can be maintained *in vitro* for extended period of time, usually without compromising their euploid karyotype. Also, procedures were already established to differentiate them *in vitro* in multiple cell types, particularly into mesoderm derivatives such as hematopoietic lineages. In addition, they could be used *in vivo* to produce mouse chimeras when re-introduced into embryos or be employed to generate teratocarcinomas when injected subcutaneously into

syngeneic mice. Finally, their genome was accessible and modifiable, as illustrated by a growing number of mutagenesis strategies applied to these cells. Therefore, ESCs combined both the *in vitro* and *in vivo* differentiation potential in addition to the mutagenesis suitability, resulting in an ideal lineage for functional genomics. In fact, they were already prized for such approaches, as described in the upcoming section.

## 1.3 Functional genomic approaches applied to ESCs

We needed to select a functional genomic strategy that we would apply to ESCs. Some technologies were already optimized at that time, most of them improved over the years and new ones appeared. Different advantages and disadvantages could be recognized for these methodologies. As a preamble, both past and present contexts will be presented to underscore the relevance of the selected approach (next section).

### 1.3.1 Gene targeting

Gene targeting is a methodology that relies on homologous recombination to introduce a modification in a selected region. Typically, a vector containing a selection marker gene flanked by two homology arms is used to abolish the function of a gene, usually by removing the first coding exon. Removal of the selection marker gene is recommended using Cre-*loxP* or Flp-*frt* technologies to prevent unspecific effects<sup>8</sup>. Cre or Flp are site-specific recombinases that catalyze the recombination between two *loxP* or *frt* sequences, respectively. Because various mutations can cause embryonic lethality, precluding analysis later during development and adulthood, conditional gene targeting approaches were designed, again relying on Cre-*loxP* or Flp-*frt* technologies. For this purpose, gene inactivation is regulated in a spatio-temporal manner according to the tissue-specific expression or induction of the recombinases.

Gene targeting approaches were well established at the time this project was initiated but remained time consuming. It was laborious to get information about selected loci, to obtain fragments of DNA corresponding to the targeted regions (e.g., physical maps and BAC contigs were largely unavailable), and to create targeting vectors. Today the picture is completely different: the mouse genome sequence is freely available, mapped libraries of BACs and already-made targeting vectors<sup>9</sup> are obtainable, new engineering approaches allow easier plasmid and/or BAC modifications<sup>10</sup>, etc. Even with these improvements, gene targeting is still laborious mostly because of the work involving the identification of the

proper ESC clones that bear the desired modification(s). However, among the advantages of the methodology are the known and precise location of alterations and the accessibility to almost any region, transcribed or not. So far, ~4000 genes have been targeted in the mouse, with or without a conditional approach<sup>11</sup>. This number is expected to increase shortly with the targeting of 18 500 additional genes by an international effort (IKMC: International Knockout Mouse Consortium<sup>12</sup>) conducted by KOMP (KnockOut Mouse Project), EUCOMM (European Conditional Mouse Mutagenesis Program), and NorCOMM (North American Conditional Mouse Mutagenesis Project)<sup>11</sup>.

### **1.3.2 Gene trap screens**

Gene trap screens are currently the companion of gene targeting with the aim of inactivating every gene in the mouse genome<sup>12</sup>. Different trapping vectors have been generated over the years, based either on plasmids or retroviruses. The principle behind the trapping strategy is to catch a complementary genomic feature which is missing for the expression of a selection gene found in the trap vector: promoter, polyadenylation signal (pA), etc. Depending on the type of vector used, different trapping biases are observed. For example, promoter traps rely on actively transcribed regions while pA traps do not, some retroviral vectors show preferential integration site (discussed later in section III) sometimes resulting in hypomorphic rather than null alleles. Some trap vectors are quite sophisticated, allowing conditional knockdown of gene expression<sup>13</sup>. The International Gene Trap Consortium (IGTC) manages at least 45 000 ESC lines, with integration covering ~40% of known mouse genes<sup>14</sup>. The Texas Institute for Genomic Medicine (TIGM) is currently generating a gene trap library of  $\geq 350\ 000$  C57BL/6 ESC clones, expected to cover ~13 000 genes to completion this year<sup>12</sup>. This methodology is popular because of its simplicity. Integration sites are mapped by different methods such as plasmid rescue or inverse-PCR (discussed later in section III).

### **1.3.3 shRNA screens**

shRNA-based screens, employing small hairpin RNAs to suppress gene expression, were emerging at the time this project was initiated and are now commonly used in ESCs. Elegant vectors are based on lentiviruses coding both for a shRNA and the corresponding inducible target gene<sup>15</sup>. Even if non specific off-target effects and/or partial rather than complete suppression of the gene of interest are frequent, the methodology is relatively efficacious. With the availability of lentivirus-based shRNA libraries<sup>16</sup> or microarrays of



concentrated lentiviruses spotted on glass slides<sup>17</sup>, this methodology should be increasingly used in ESCs. For the moment, RNA interference approaches target protein-coding and non protein-coding transcripts, but cannot target untranscribed regions. However, it is now suspected that microRNAs participate in undefined ways to processes such as methylation and heterochromatization<sup>18</sup> and maybe one day, these functions will be exploited in ESCs.

### 1.3.4 Other insertional mutagenesis screens

Gene targeting, gene trap, and shRNA screens can be viewed as insertional mutagenesis because they rely on vectors integrating in the genome. Additional insertional mutagenesis tools are used in mouse models such as replication-competent retroviruses or retrotransposons, mainly to find proto-oncogenes or tumor suppressor genes<sup>19</sup>. Although these methodologies could be adapted to ESCs, there are other alternatives that appear more advantageous. For example, replication-incompetent retroviral gene trap vector equipped with a reporter gene can be both mutagenic and be exploited to detect the expression profile of the trapped gene. In the case of DNA transposons, such as Sleeping Beauty, they act through a cut and paste (excision and integration) mechanism induced by a transposase. Unfortunately, the integrants are subjected to remobilization, leaving behind hardly detectable mutagenic footprints. In addition, these transposons have a tendency to jump in their neighborhood rather than randomly in the genome<sup>19</sup>.

### 1.3.5 Chemical screens

Chemical screens can be classified in two broad categories: one relying on chemicals as mutagens and the other on compound libraries that alter the function of ESC without necessarily affecting their genome.

Chemical mutagens such as *N*-ethyl-*N*-nitrosourea (ENU) or ethylmethanesulphonate (EMS) were used in ESC to create single base substitutions or alterations in mRNA splicing, transcription, or stability<sup>20,21</sup>. The frameshift mutagen ICR191 was also used to induce the addition of guanine stretches<sup>21</sup>. According to the loss-of-function experiments evaluated for the selectable hemizygote locus *hypoxanthine phosphoribosyl transferase* gene (*Hprt1*), mutation frequency is in the range of 1 per 1000 to 1 per 1200 cells depending on the conditions tested<sup>20,21</sup>, implying multiple mutations in each genome. Because of this amount of subtle mutations that do not contain a landmark for identification, it is necessary to create chimeric mice and proceed through breeding to first dilute the mutation load and then to isolate

candidate gene(s) by positional cloning (germ-line transmission is achievable). Different phenotypes have been observed in mice and important genes identified using this system<sup>21-24</sup>. Several genome-wide ENU-or EMS-based screens for dominant and recessive mutations are currently ongoing. Major efforts include those conducted by the British, German, Australian, American, Canadian, and Japanese groups<sup>25</sup>. Transient expression of Bloom in ESC can stimulate homologous recombination between sister chromatids or homologous chromosomes, allowing the recovery of bi-allelic mutations<sup>26</sup>. This strategy was used in ESC to study a precise pathway (glycosylphosphatidylinositol-anchor biosynthesis) and defects were complemented by candidate genes (cDNA transfection)<sup>26</sup>. For clones not successfully complemented with known genes, other methods must be used to identify the mutated gene (if it is a gene) or a companion in the same pathway (for example: cDNA library). However, if a combination of determinants is necessary to rescue the phenotypic anomaly, it is difficult to achieve at the genome-wide level. Focusing on particular chromosomal regions can be an advantageous strategy to use with chemical mutagenesis. Chemical mutagenesis used in combination with heterozygote chromosomal deletions (see beneath for the methodologies to create deletions) or heterozygote chromosome balancers (chromosome containing an inversion suppressing chromosomal recombination for this region, a dominant visible marker, and a recessive gene inducing lethality in a homozygous state<sup>25</sup>), simplifies the breeding scheme. The mutations caused by chemicals are limited in size and in type according to the mutagen used (for example, single-base substitution involving AT base pairs predominate for ENU<sup>20</sup>). Importantly, ENU-induced mutations are not bias for any region of the genome. Moreover, ENU-mutagenized collection of ESC clones can be screened for mutations in selected genes, allowing the recovery of allelic series<sup>27</sup>. These ESCs can be reintroduced in developing embryos to create mouse chimeras and these specific mutations can be transmitted in the germ-line<sup>27</sup>. This procedure allows the *in vivo* functional evaluation of precise protein domains<sup>27</sup>.

Screening using small molecule libraries is an emerging application in the ESC field, which is expected to be employed more extensively in the near future. For example, a library was used to identify compounds able to maintain ESC self-renewal/pluripotency without the use of serum, LIF, or feeders<sup>28</sup>. The approach was powerful because the team used a pluripotency reporter gene (*Oct4-GFP*) in ESCs and plated them in 384-well plates<sup>28</sup>. They characterized one of these compounds (SC1: pluripotin), which allows the maintenance of ESC in a minimal media without compromising their ability to differentiate *in vitro* and *in vivo* in chimeric mice<sup>28</sup>. Contribution of ESCs to the gonads of these chimeric mice was proven<sup>28</sup>, although the proper functionality of the gametes was not assessed by germ-line

transmission. Importantly, by immobilizing the compound to an affinity matrix and using mass spectrometry, this group identified two cellular targets of their small molecule (Erk1 and RasGAP)<sup>28</sup>. They further showed that the combined inhibition of both proteins was necessary to maintain self-renewal/pluripotency in the conditions used<sup>28</sup>.

### 1.3.6 Mutagenesis with oligonucleotides

The use of single-stranded DNA oligonucleotides to permanently modify 1-4 targeted nucleotides in ESC is a recently developed application. Oligonucleotides can be obtained faster than gene targeting vectors not already made. However, both methodologies require the same amount of work to isolate ESC clones and assess the proper targeting. The frequency of oligonucleotide-based targeting is estimated to be 0.25-1.5 per 10<sup>6</sup> cells, as tested on a limited number of loci<sup>29</sup>. Problematically, this methodology is suppressed by DNA mismatch repair mechanisms, thus requiring the transient suppression of proteins such as Msh2<sup>29</sup>. As a consequence of repressing transiently mismatch repair mechanisms, increased frequency of spontaneous mutations is observed on reporter genes<sup>29</sup>. The distribution of these bystander mutations is unknown, but expected to be lost during mouse breeding since the targeted alteration can be germ-line transmitted<sup>29</sup>. However, this pitfall should be taken into consideration when designing *in vitro* screens. Because the mutation is targeted, as opposed to chemical screens, *in vitro* screens could probably be achieved with the use of independent targeted clones.

### 1.3.7 Gain-of-function screens

Gain-of-function screens have been applied successfully to ESC. A strategy using episomal transduced cDNA library (derived from ESC) identified Nanog, an homeodomain protein allowing the self-renewal of ESC without leukemia inhibitory factor (LIF)<sup>30</sup>. Microarray analyses can also be combined to this type of screen<sup>31</sup>. In a way, gain-of-function screens are attractive because of their simplicity and their rapidity. However, to apply the methodology to a genome-wide level, libraries are disadvantageous. For example in home-made cDNA libraries, very long cDNAs are under-represented, more abundant transcripts are over-represented, some cDNAs are incomplete, etc. To circumvent these drawbacks, one could think about using a BAC library. However, some genes are so big that they are not covered by a single BAC, such as *dystrophin* (2,2 megabase pairs, 79 exons) (<http://genome.ucsc.edu/>, Mouse Build 36, 2006)<sup>32</sup>. Fortunately, vast libraries can be purchased, arrayed in multi-well format (example BACs libraries) or spotted in a book (example: Riken cDNA

library) allowing for a better attempt at normalization of each products. Aside from unwanted effects caused by the ectopic/over- expression (toxicity, non physiological expression levels, abnormal cellular localization, etc.), many studies proved that valuable candidate factors can be isolated with this strategy.

### 1.3.8 Irradiation-based screens

Deletions can be produced in ESCs engineered to express the *herpes simplex thymidine kinase (tk)* gene, by physical irradiation and negative selection (drug FIAU)<sup>33</sup>. Although anchor sites (*tk*) introduced in various genomic regions can be identified by plasmid rescue experiments, the mapping of each deletion is difficult since it requires PCR analysis of numerous simple sequence length polymorphism markers<sup>33</sup>. The possibility of unidentified genetic lesions also complicates the interpretation of results generated with this approach.

### 1.3.9 Cre-*loxP* technology based screens

The Cre-*loxP* technology has been applied by several groups to create large deletions, but this system is also appropriate to create translocations, inversions, and duplications. To produce a deletion in ESC, two regions on the same chromosome are successively targeted by homologous recombination using distinct vectors, each carrying a *loxP* site<sup>34</sup>. Subsequently, the transient expression of Cre leads to the excision of DNA between the integrated *loxP* sequences. To isolate ESC recombinants, two nonfunctional halves of a selection marker gene are inserted in complementary targeting vectors. ESCs are selected in media containing the proper drug(s). Although different combination of marker genes have been developed over the years<sup>35</sup>, the functional reconstitution of *Hprt1* is more widespread, presumably because it was the first reported system<sup>34</sup>. *Hprt1*-deficient ESCs need to be employed in this case.

A variation of the *Hprt1* reconstitution method was also elaborated, where the first *loxP* site was anchored by homologous recombination and the second, delivered by a retroviral vector, avoiding one step of gene targeting<sup>36</sup>. Another strategy to omit one round of homologous recombination consists in targeting one *loxP* site, followed by the co-introduction of a *loxP*-containing plasmid together with the *Cre*, and finally by selecting recombination events by negative selection toward the *tk* gene incorporated inside the vectors<sup>37</sup>.

## 1.4 The proposed approach

Several reasons drove the design of a screen based on chromosomal deletions: several contiguous determinants could be interrogated at the same time, both protein-coding and non protein-coding regions could be screened, potential synthetic interactions could be observed, the alleles were permanently deleted and not only silenced, and the primary work could be done *in vitro*. Irradiation methodology was not a possibility because the mapping of deletions was not precise and like chemical screens, subject to bystander mutations. A Cre-*loxP* strategy was favored but the laborious step of homologous recombination was repelling. We wanted to screen many regions on a genome-wide scale within a reasonable timeframe. In order to overcome these impediments and to bring on an additional tool for functional genomics, two complementary retroviruses were created, each containing a *loxP* site and capable of rapidly generating deletions in mammalian cells following the addition of Cre. **Table I** recapitulates the advantages and disadvantages of the methodologies presented in the previous section in addition to the retroviral-based method which I developed in our laboratory. The remaining sections of the introduction will focus on embryonic stem cells and on retrovirology.

**Table I • Advantages and disadvantages of ESC functional genomic approaches**

Methodologies	Advantages	Disadvantages
Gene targeting	Precise location, precise modification, no bias although better frequencies for some regions	Labor intensive, time consuming
Gene trap	Simple, rapid, localization easy	Integration bias depending on vectors
RNA interference	Simple, rapid	Possible non specific off-target effects, variable degree of suppression
Chemical (mutagens)	No localization bias reported, hypomorphic, hypermorphic, loss-of-function alleles possible (allelic series)	Multiple mutations, very subtle mutations, laborious identification
Chemical (compounds)	Simple, rapid	Limited by the library of compounds, dependent on the concentration of compounds, the target(s) might be difficult to identify
Oligonucleotides	Precise location, simple	Possible bystander mutations, characterization of ESC time-consuming
Overexpression	Simple, rapid	Depends on libraries' coverage, might be prone to unspecific effect (toxicity, non physiological expression levels, etc.)
Irradiation (deletions)	Simple, rapid, can involved large chromosomal segments	Localization of endpoints difficult, possible bystander genetic alterations
<i>Cre-loxP</i> (both <i>loxP</i> targeted or one targeted + one introduced by retroviral gene transfer)	Various rearrangements possible (deletions, inversions, translocation, etc.), precise location, can involved large chromosomal segments	A least one round of laborious and time-consuming gene targeting, some bias might be observed with retroviruses
Our approach: <i>Cre-loxP</i> (both <i>loxP</i> introduced by retroviral gene transfer)	Various rearrangements possible (deletions, inversions, translocation, etc.), precise location, can involved large chromosomal segments, avoid one round of gene targeting	Some bias might be observed with retroviruses

## PART II: INTRODUCTION TO EMBRYONIC STEM CELLS

The *in vitro* derivation from blastocysts of the first mouse embryonic stem cell (ESC) lines was reported in 1981<sup>38,39</sup>. These lineages were holding great promises in developmental biology because of their ability to differentiate into complex tissues *in vivo*, to form embryoid bodies (EBs) in suspension culture, and to produce teratocarcinomas when injected subcutaneously in syngeneic mice<sup>38,39</sup>. As opposed to teratocarcinoma cells, ESCs have a normal karyotype and contribute more successfully to the germ-line of chimeric mice<sup>38,40</sup>. The focus of this section is to review the origin and some of the cardinal features of mouse ESCs: self-renewal and pluripotency, differentiation, and particular genetic properties.

### 1.5 The origin of ESC

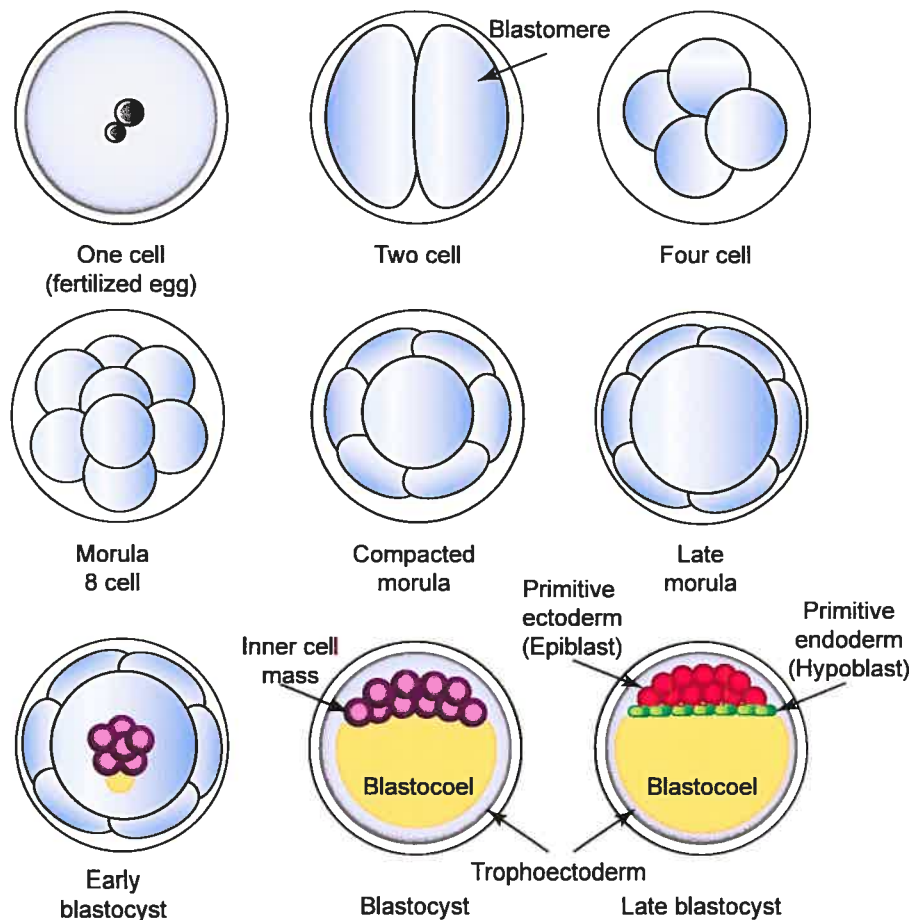
The protocols currently used to derive ESC lines are similar to those established more than 20 years ago<sup>41</sup>. Blastocyst stage embryos or isolated inner cell masses (**Figure 1-1**) are plated on mouse embryonic fibroblast (MEFs) in tissue culture media<sup>41</sup>. Following several days of culture, the masses are dissociated and replated again on MEFs to generate various differentiated and undifferentiated lineages<sup>41</sup>. Colonies with undifferentiated morphology are individually isolated and are expanded to generate ESC lines<sup>41</sup>. Most ESC lines are 40XY because in XX ESCs both X chromosomes are active, an unstable state (in fact one of the X is frequently lost) that correlates with global reduction of DNA methylation which is not favorable for the maintenance of these cells<sup>41,42</sup>.

An ongoing debate is the tissue of origin of ESC, the existence of an *in vivo* counterpart, and the possibility of being an artifact lineage generated from an adaptation to culture environment<sup>43</sup>. Cells from both the inner cell mass and from the primitive ectoderm (**Figure 1-1**), a tissue derived from the inner cell mass, can give rise to ESC lines<sup>43</sup>. However, since not all the cells contained in these tissues can generate ESC lines, ESCs could possibly emerge subsequently from another cell type, such as early germ cells<sup>43</sup>. ESCs are not equivalent to inner cell mass cells because they contribute weakly to extraembryonic endoderm lineages (derivative of the primitive endoderm) *in vivo* (**Figure 1-1**)<sup>43</sup>. A founder population of cells emerges from the primitive ectoderm (epiblast) soon before gastrulation, and passes through the primitive streak to give rise to many structures of the extraembryonic mesoderm and to germ cells, a process involving dominant local and inductive signals, which might be reproducible *in vitro*<sup>43</sup>. ESCs might be related to this founder population<sup>43</sup>. In fact, mouse primordial germ cells can generate ESC-like colonies that can be maintained

for extended period of time in culture and can contribute to chimeras and germ-line transmission<sup>41</sup>. To complicate the story further, under particular cell culture conditions, ESCs can be differentiated into primitive ectoderm-like cells, which can be differentiated *in vitro*, but are unable to contribute to chimeric mice<sup>44</sup>. As expected, ESC and all the potential parental lineages share several marker genes (*Oct4*, *Nanog*, *Dppa3*, etc.)<sup>43</sup>, but none of them demonstrate *in vivo* the permanently high proliferation index of ESC observed in culture. Therefore, this sustained proliferation rate might be the result of cell culture conditions or of the isolation of transient cells with this intrinsic property or more possibly, the artificial combination of both. Fortunately, when ESCs leave the *in vitro* environment to return *in vivo* following re-introduction in the mouse embryo, they respond normally to developmental instructions and therefore, do not correspond to transformed cells. However, if they are not re-introduced in the proper environment, they create teratomas (or teratocarcinomas) instead of contributing adequately to tissues in place.

**Figure 1-1 Early mouse development during the stage of preimplantation.**

Adapted from Ralston, A. & Rossant, J., 2005<sup>45</sup>.





## 1.6 Self-renewal and pluripotency of ESCs

Self-renewal and pluripotency are key characteristics that define ESCs and are typically discussed together. Self-renewal is a mechanism that allows the generation of daughter cell(s) with the same characteristics as the parental cell. ESCs are thought to generate two identical daughter cells per division, a process referred to as symmetrical self-renewal, as well as to preserve immortality<sup>41</sup>. This state strictly relies on well-defined culture conditions. Pluripotency refers to the *in vivo* differentiation potential of clonal ESC to contribute to all lineages derived from the three primary germ layers (ectoderm, mesoderm, endoderm, including the gametes) as well as the extraembryonic mesoderm<sup>46</sup>. Because ESCs contribute weakly to extraembryonic endoderm and trophoblast lineages, they are considered pluripotent rather than totipotent such as the fertilized egg or the blastomeres. Specific culture conditions allow the preservation of the pluripotency. Ironically, to characterize this property, ESCs need to lose it, concomitantly with their identity, through *in vivo* and/or *in vitro* differentiation. The evaluation of ESC pluripotency *in vivo* is the most robust assay to observe both the contribution of ESC to all expected lineages and the proper functionalities of these progenies. However, this experimentation is expensive. Other assays, although not as complete, give reasonable insights into the pluripotency of ESCs. The generation of teratomas or the *in vitro* differentiation in selected media allows the observation of representative lineages from the three primary germ layers. More details about *in vivo* and *in vitro* differentiation of ESC will be presented in the next section.

What are the factors regulating ESC identity (self-renewal and pluripotency)? A manuscript (News & Views) was written by Mélanie Bilodeau and Guy Sauvageau in 2006, presenting a general overview of the field and two approaches used by independent groups to find regulators of self-renewal and pluripotency (**Appendix I**). An update of this area of research will follow.

### 1.6.1 Update for mechanisms underlying ESC self-renewal and pluripotency

It is important to highlight that both cell extrinsic and intrinsic mechanisms governing ESC self-renewal and pluripotency imply not only positive regulation, but also repression of differentiation and maybe apoptosis. Sometimes, a single factor can act both as a positive and a negative regulator. Cell signaling cascades initiated extrinsically are necessarily linked to cell intrinsic parts. In addition, epigenetic characteristics such as DNA methylation and

histone modifications seem involved in ESC self-renewal and pluripotency regulation and their roles should be more extensively defined in the coming years.

BMP4 is a cell extrinsic factor acting as a ligand to a cell intrinsic signaling cascade [BMP receptor-Smad(s)-Id(s)] that suppresses neural determination<sup>47</sup>. Similarly, LIF is a cell extrinsic ligand of a cell intrinsic signaling cascade [LIFR-gp130-Stat3-target genes] suspected to inhibit non-neuronal differentiation rather than promoting stem cell survival<sup>47</sup>. Oct4, Sox2, and Nanog are three core transcription factors that positively regulate ESC specific genes, but also bind non-expressed tissue-specific transcription factors<sup>48</sup>. Oct4 and Nanog are specific to pluripotent cells, but Sox2 is not<sup>49</sup>. Critical levels of Oct4 are required to maintain ESC in an undifferentiated state: repression of Oct4 conveys to loss of pluripotency and formation of trophectoderm, while the overexpression of Oct4 induces differentiation in primitive endoderm and mesoderm<sup>50</sup>. Nanog positively maintains ESC self-renewal in the absence of LIF<sup>30</sup> and is thought to suppress differentiation. Nanog is down-regulated during differentiation<sup>30</sup>, inhibits neuroectodermal differentiation when ectopically expressed<sup>30</sup>, and Nanog-deficient ESCs produce endoderm (possibly primitive)<sup>47</sup>. Oct4 and Cdx2 transcription factors reciprocally inhibit each other functions for the determination of pluripotent cells (Oct4 functions expressed, Cdx2 functions repressed) and trophectoderm (Cdx2 function expressed, Oct4 function repressed)<sup>51</sup>. Nanog, Gata4, and Gata6 might be regulating a balance between the pluripotent state and differentiation in primitive endoderm. The loss of Nanog or the ectopic expression of Gata4 or Gata6 induces ESC to differentiate into primitive endoderm<sup>49</sup>. In addition, Esrrb, Tbx3, Tcl1, and Dppa4 also control a set of target genes by activation and repression<sup>48</sup>. Oct4, Sox2, Nanog, Esrrb, Tbx3, Tcl1, and Dppa4 also possibly share some target genes<sup>48</sup>.

The epigenetic level of regulation is expected to be complex and is just starting to be elucidated. In the case of DNA methylation for example, although neither *Oct4* or *Nanog* genes present annotated CpG islands, their respective promoter present cytosine methylation, correlating with their expression (low level of methylation correlating with expression)<sup>6</sup>. Methylation is thought to induce repression by preventing the binding of some proteins to DNA (such as transcription factors) and/or by binding methyl-CpG binding proteins that interact with histone deacetylases<sup>6</sup>.

ESCs present bivalent domains containing the dual repressive (lysine 27 of histone H3 [H3K27] tri-methylation) and activating (lysine 4 of histone H3 [H3K4] tri-methylation) histone marks<sup>49</sup>. These bivalent domains correspond to highly conserved non

coding elements<sup>49</sup>. Several of them are co-occupied by Oct4-Sox2-Nanog and are found in proximity of some (but not all) developmentally important genes silenced in ESC, but activated during differentiation<sup>49</sup>. During differentiation, the bivalent domains presenting activating and repressive histone marks are resolved: expressed genes are associated with H3K4 tri-methylation, turned off genes are associated with H3K27 tri-methylation, while the weakly induced genes keep both signatures<sup>49</sup>. It is hypothesized that bivalent domains may silence developmental genes in ESCs, but also keep them poised for activation during differentiation<sup>52</sup>.

Silencing could be mediated in part by Polycomb group gene (PcG) complexes. Two of the four known PcG complexes are important in ESC, mainly PRC1 and PRC2<sup>48</sup>. Methylation of H3K27 is induced by the PRC2 complex, which includes *eed* (embryonic ectoderm development), *Suz12* (suppressor of zeste 12), and *Ezh2* (Histone methyltransferase enhancer of zeste homologue 2)<sup>48</sup>. H3K27 methylation is a binding site for the PRC1 complex involving *Ring1A*, *Ring1B*, and *Bmi1*<sup>48</sup>. The precise roles of PcG complexes and associated histone modifications in ESC are not completely understood, but likely interfere with nucleosome dynamics and transcription initiation<sup>48</sup>. *Eed*- and *Ring1B*- deficient ESCs present derepressed transcriptional regulators of development<sup>48</sup>. The recruitment of PRC2 complex to targeted loci may be mediated in part by Oct4, Sox2, and Nanog<sup>48</sup>.

The mechanisms allowing ESCs to remain undifferentiated and to survive in cell culture conditions are not well defined. Is it the same factors that keep them self-renewing in an undifferentiated state that prevent their apoptosis? In standard culture conditions (presence of LIF and BMP), few ESC undergo apoptosis<sup>53</sup>. Are these culture conditions compatible with ESC apoptotic death? The answer is not obvious, because removal of LIF and/or BMP changes the fate of ESC. Alternatively, ESCs might have a cell intrinsic machinery preventing their death by apoptosis.

At first sight, deficiency in the gene *Zfx* uncouples ESC self-renewal properties (apparently lost) from their pluripotential properties (apparently maintained), while in fact, several cells are lost by apoptosis<sup>53</sup>. *Zfx*, located on the X chromosome, encodes a zinc finger protein containing a DNA-binding and a transactivation domains<sup>53</sup>. ESCs (XY) deficient in *Zfx* present abnormal morphology and are defective in proliferation because they die from apoptosis, an effect highlighted in serum-free condition in presence of LIF and BMP4<sup>53</sup>. Strangely, *Zfx*-deficient male embryos (germ-line deletion) develop normally until E9.5 before dying of uncharacterized extraembryonic tissue anomalies<sup>53</sup>. When *Zfx*-deficient ESCs

are induced to differentiate in embryoid bodies, in teratomas, or in chimeric mice, they do so roughly normally (except that they fail to contribute to thymus and bone marrow of chimeric mice)<sup>53</sup>. Zfx overexpression in ESC correlates with massive cell death in presence of LIF, with abnormal EBs formation (absence of LIF), and with failure to contribute to chimeric mice<sup>53</sup>. At the molecular level, Zfx deficiency in ESC leads to the up-regulation of stress-induced genes<sup>53</sup>. Also, Zfx binds to the promoters of Tbx3 and Tcl1<sup>53</sup>. Overexpression or deficiency of Zfx increases or reduces the expression of these genes, respectively<sup>53</sup>. Because at least two suspected equilibriums regulate the pluripotency and the extraembryonic tissue differentiation (Oct4-Cdx2 in the case of the trophoectoderm and Nanog-Gata4-Gata6 for the primitive endoderm), the morphology of Zfx-null ESC colonies is altered and Zfx-deficient mice die from extraembryonic defects, it would be interesting to investigate whether dying cells are in fact differentiated and hardly maintained in ESC-defined culture conditions.

## 1.7 Differentiation of ESC

### 1.7.1 Contribution of ESCs to mice

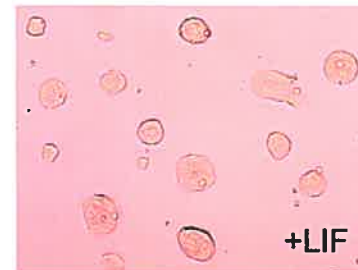
Micro-injection of ESCs in the blastocoel cavity of mouse blastocyst-stage embryo, followed by transfer to pseudopregnant female, was the first methodology developed to generate chimeric mice with ESCs<sup>54</sup> and is still currently used today (**Figure 1-2**). When ESC are injected into blastocysts, they efficiently colonize the tissues that form the fetus and the extraembryonic mesoderm<sup>46</sup>. Moreover, these cells contribute very inefficiently to extraembryonic endoderm and trophoectoderm formation<sup>46</sup>. Initial studies showed that groups of 10-15 ESCs or a single ESC could contribute to chimeric mice, although the percentage of chimerism was systematically lower for the latter<sup>46</sup>. Today, it is thought that possibly 1 or 2 or occasionally 3 ESC(s) contribute to the chimeric mice<sup>55</sup>. Interestingly, single ESC selected according to their large (~15 $\mu$ m) or small (~10 $\mu$ m) size demonstrate no difference in their contribution potential<sup>46</sup>.

ESC micro-injection requires expensive equipments, is time-consuming, and necessitates a serious training<sup>56</sup>. Consequently, the aggregation method was elaborated as an alternative (**Figure 1-2**). For this technique, ESC clumps are cultivated overnight in proximity of morula-stage embryos (undergoing or just completed compaction, with the zona pellucida removed) in little depressions<sup>56</sup>. The following day, aggregated embryos are transferred to pseudopregnant foster mothers<sup>56</sup>. Competent chimeras for germ-line transmission of ESC-derived gametes are generated efficiently with both the micro-injection

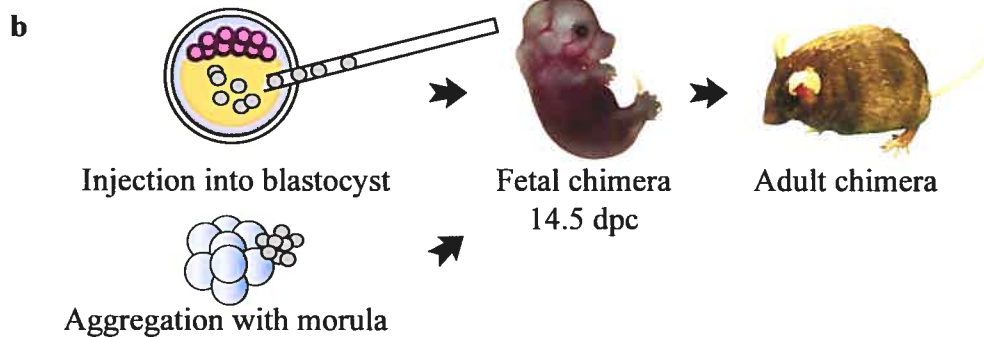
and the aggregation methods<sup>56</sup>. In the aggregation method, random-bred (like CD1) morulas are used advantageously, since the outbred female mice generate more embryos following superovulation compare to inbred strains<sup>56</sup>. However, the use of inbred blastocysts (like C57BL/6) is more efficient for micro-injection<sup>56</sup>.

**Figure 1-2 Generation of chimeric mice a using micro-injection or aggregation**

(a) LIF and BMP signaling maintain ESCs undifferentiated. (b) When reintroduced into mouse blastocyst or aggregated with a morula, ESCs contribute to every tissue of the chimeras.



Undifferentiated ESCs



A variation of the aggregation technique is to use tetraploid morula stage embryos. Electrofusion is performed on blastomeres of a two-cell stage embryo (diploid), creating the tetraploid cell that is further maintained until the morula stage<sup>57</sup>. Alternatively, ESC can be injected in the blastocoel cavity of a tetraploid blastocyst<sup>58</sup>. Using this set-up, ESCs contribute to the fetus and extraembryonic mesoderm, while the tetraploid cells are generally restricted to the trophectoderm and the extraembryonic endoderm<sup>57</sup>. During early embryogenesis, the exact moment where tetraploid cells are out-competed by ESC-derived progenies is not known. Although not labeled autonomously, tetraploid cells have been noticed in all the analyzed chimeric embryos during the gastrulation stage (E6.5-7.5), presenting variable contribution (3-80%) to embryonic derivatives of the three primitive germ layers<sup>59</sup>. When labeled autonomously, tetraploid cells were shown to contribute sporadically to  $\leq 1\%$  of cells in a chimeric fetus (E10), and sometimes to cluster in the hindgut endothelium, the aortic musculature, and the branchial arch vasculature<sup>59</sup>. Importantly, F<sub>1</sub> hybrid ESC lines are crucial for tetraploid complementation assays because ESC derived from inbred embryo engender neonates that die shortly after birth with respiratory distress<sup>58</sup>. The molecular

basis underlying the correlation between the limited genetic heterogeneity and the respiratory defect is unknown, but can be bypassed by laser-assisted injection of ESCs in 8-cell stage diploid morulas (method described beneath)<sup>58,60</sup>. However, tetraploid complementation assay is achievable with ESC lines derived from two related mouse substrains, such as the R1 ESC line (derived from a cross between two 129 mouse substrains)<sup>61</sup>. Newborn animals derived from this methodology are also referred to as F0 mice because they are derived (almost) completely from the ESCs, including their gametes, thus bypassing a step of mouse breeding necessary for traditional chimeras to obtain germ-line transmission.

An exciting method recently developed to create F0 mice consists in laser-assisted injection of ESCs in 8-cell stage morula<sup>60</sup>. Similar to tetraploid complementation, almost entirely ESC-derived chimeras are obtained, even with lower contamination from the host cells ( $\leq 0.1\%$  instead of  $\leq 2\%$ )<sup>60</sup>. Total germ-line transmission (100%) is observed most of the time because of gender conversion<sup>60</sup>. The most important point is that either inbred or hybrid ESCs and either inbred or outbred host embryos can be used without presenting F0 mice with obvious abnormalities<sup>60</sup>. It is fascinating that inbred ESCs, laser-injected in 8-cell stage morula, can generate F0 mice free of respiratory distress while injection in tetraploid blastocyst frequently fails to generate normal mice<sup>62</sup>. Impressive images show the contribution of injected ESCs to the totality of the inner cell mass with this technique, while injection in the blastocyst results in a mixture of ESC and host derived cells<sup>60</sup>. However with both techniques, ESCs fail to contribute notably to the extraembryonic endoderm<sup>60</sup>, reinforcing the idea that ESCs might be more related to primitive ectoderm than the inner cell mass.

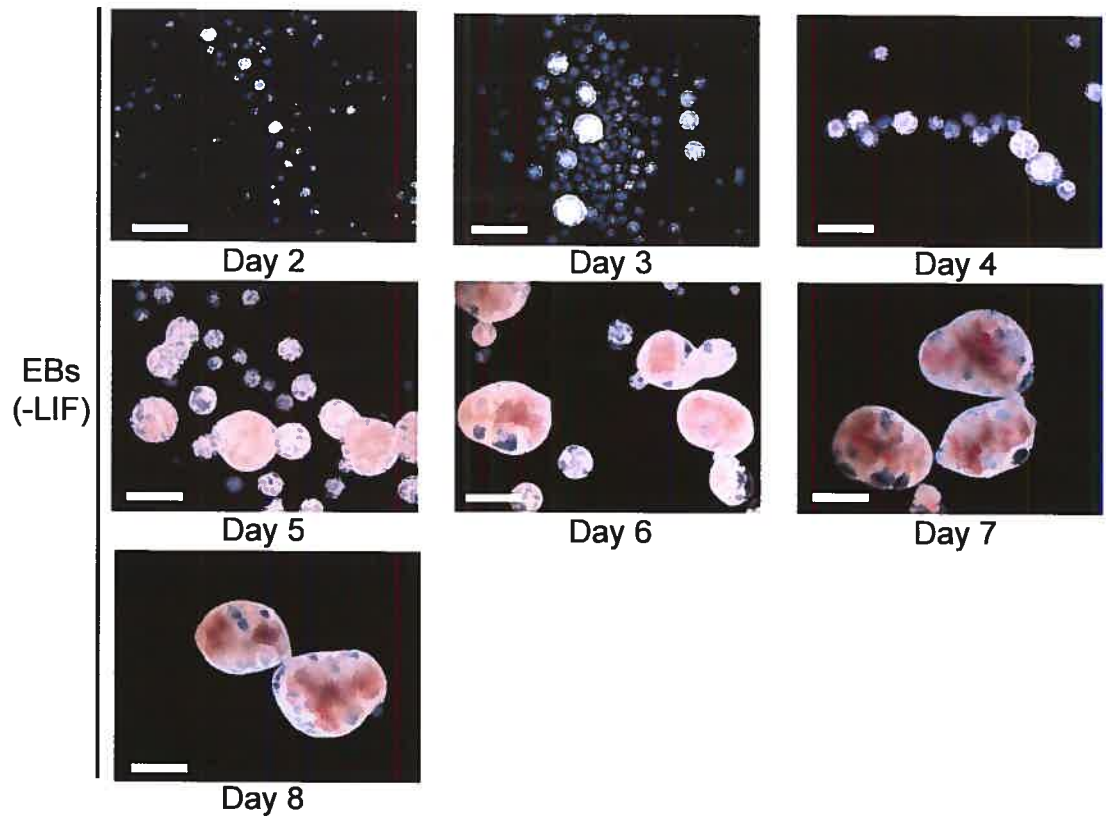
## **1.7.2 *In vitro* differentiation of ESCs**

### **1.7.2.1 *In vitro* differentiation methods**

*In vitro*, three methods are usually used to induce ESC differentiation in absence of LIF. The first one is to grow ESC in liquid or semi-solid differentiation media to generate three-dimensional aggregates called embryoid bodies (EBs) (**Figure 1-3**). The EBs differentiation allows complex developmental programs to occur, mediated by numerous cell-cell interactions<sup>44</sup>. This complexity can be problematic when trying to understand the differentiation into particular lineages, which relies on the proper development of other lineages.

**Figure 1-3** The removal of LIF allows the differentiation of ESCs in EBs *in vitro*.

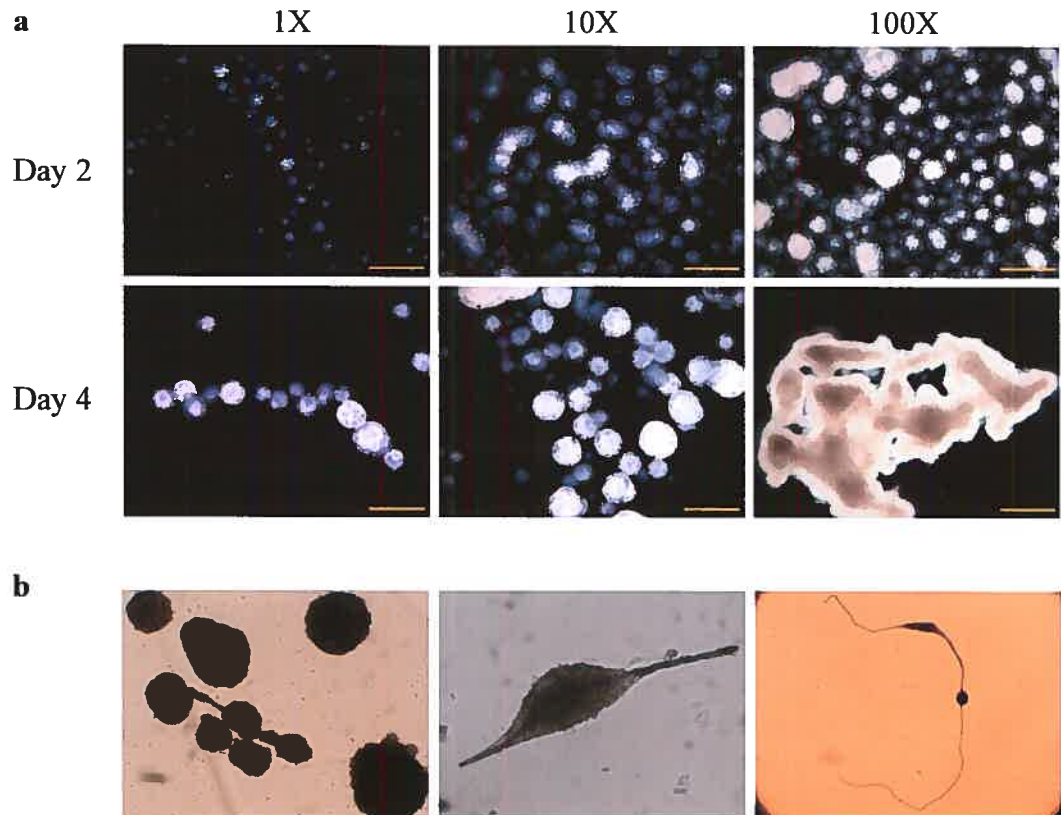
Scale bar: 500 microns.



To add to the fact that each serum lot provides an undefined blend of extracellular factors, the culture media likely becomes rapidly conditioned because the seeding density makes a significant difference in the differentiation profile of ESCs (**Figure 1-4**). Depending on culture conditions, EBs can present fascinating shapes that can be misinterpreted as phenotypic anomalies. For examples, if debris are present in the culture media, EBs have the tendency to wrap around it or, when plated at high density, to fuse into deformed aggregates (**Figure 1-4**).

**Figure 1-4 Effect of seeding density and culture conditions on EBs differentiation.**

(a) The EB differentiation profile is affected by the seeding density. At high density, EBs fuse in bizarre aggregates. Scale bar: 500 microns. (b) EBs' fascinating shapes when large debris are found in the culture media.

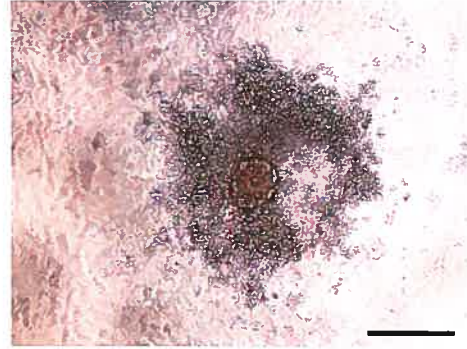


In addition, ESCs can be differentiated over a stromal cell layer such as OP9 cells<sup>63</sup> (**Figure 1-5**). OP9 cells were derived from the calvaria of a newborn mouse deficient in the M-CSF gene<sup>63</sup>. Coculture of ESCs with these stromal cells allow mesodermal and lymphohematopoietic differentiation without the addition of growth factors (but in the presence of serum)<sup>63</sup>. This might be a simpler way to obtain particular cell types because differentiation occurs in a monolayer interacting with the stromal cells. This type of differentiation is also influenced by the seeding density. For particular purposes such as expression studies, stromal cells need to be separated from ESC-derived progenies. Finally, ESC can be differentiated straight in monolayer or on extracellular matrix, with particular media.



**Figure 1-5 ESC differentiation in mesodermal and hematopoietic lineages on OP9 stromal layer.**

Scale bar: 250 microns.



### 1.7.2.2 Additional considerations regarding ESC *in vitro* differentiation

Although ESC differentiation is a remarkable tool, there are several pitfalls. ESC *in vitro* differentiation is highly modulated by cell culture conditions. Surprisingly, *in vitro* differentiation allows ESCs to participate in particular lineages such as the extraembryonic endoderm<sup>64</sup> while they are inefficient to do so *in vivo*. Additionally, when ESCs are genetically modified (for example: suppression of Oct4), they form trophoectoderm (tested in presence of LIF), a phenomenon called dedifferentiation<sup>50</sup>. In addition, because of their property to fuse at low frequency with other cells, ESC can be the unsuspected cause of another process called transdifferentiation (change in cell fate)<sup>65</sup>. Finally, although the *in vitro* differentiation of ESC can be temporally representative of early embryogenesis as discussed below, it occurs without an organization such as axis formation.

For all these reasons, three characteristics were established to conclude that an ESC *in vitro* differentiation model in a particular lineage is relevant<sup>44</sup>. First, the system must be efficient and reproducible<sup>44</sup>. Second, the system should recapitulate the developmental program observed *in vivo*<sup>44</sup>. And finally, the differentiated cells should be functional in culture and when transplanted in animal models<sup>44</sup>.

### 1.7.2.3 Successfully derived lineages

Primary germ layer induction during ESC differentiation shares pathways that are found in embryogenesis: bone morphogenic protein (BMP) and other transforming growth factor- $\beta$  (Nodal/Activin) signalings, Wnt signaling, and fibroblast growth factor (Fgf) signaling<sup>66</sup>.

*In vitro* differentiation produces representative lineages from the mesoderm (hematopoietic, vascular, cardiac, skeletal muscle, osteogenic, chondrogenic, adipogenic), the

endoderm (visceral endoderm, pancreatic islets, hepatocytes, thyrocytes, lung, and intestinal cells), the ectoderm (neural, inner ear progenitors, melanocytes, and keratinocytes) and germ cells (oocyte and male germ cells)<sup>44,67,68</sup>. However, not all lineages are efficiently derived, possibly because of the limited knowledge of culture regiment (growth factors, gradients, cell-cell interactions, etc.) necessary to imitate *in vivo* conditions. Differentiation can be single or multi-step, short-term or for an extended period of time, with various physical supports (liquid or semi-solid environment, particular coating, etc.), with or without serum, with or without growth factors, with or without co-cultivation, etc. One must be cautious about all these details when comparing various studies. Fortunately and surprisingly, many published systems seem to reasonably fulfill the three criterias mentioned previously. One of them is hematopoietic differentiation and will be discussed in more detail.

#### 1.7.2.4 Comparison of ESC *in vitro* versus *in vivo* hematopoietic differentiation

During embryogenesis, hematopoiesis is initiated in the yolk sac around E6.5, as observed by the appearance of blood islands, which are clusters of primitive erythrocytes surrounded by endothelial cells<sup>44</sup>. Primitive erythrocytes are large and nucleated cells, expressing the embryonic form of hemoglobin and are produced specifically in the yolk sac for a narrow period of time<sup>44</sup>. This stage is defined as primitive hematopoiesis<sup>44</sup>. All the other blood lineages (myeloid, lymphoid, and smaller definitive erythrocytes expressing the fetal-adult forms of hemoglobin) are referred to as definitive hematopoiesis<sup>44</sup>. The yolk sac contributes to definitive hematopoiesis by producing macrophages, definitive erythrocytes, and mast cells<sup>44</sup>, but fails to form lymphocytes or hematopoietic stem cells (HSCs) according to assays of long-term reconstitution in irradiated adult mice. The first site able to produce myeloid cells, lymphoid cells, definitive erythrocytes, and HSCs is the intraembryonic para-aortic splanchnopleuro (P-Sp) region, the presumptive territory of the aorta, gonads, and mesonephros (AGM) region<sup>44</sup>. The yolk sac could possibly produce HSCs that cannot be detected by reconstitution of adult recipients because of homing inaptitude and/or the lack of a maturation step, that might be reproduced *in vitro* with AGM stromal cells cocultivation or *in vivo* by injecting the cells in liver of newborn mice<sup>69</sup>.

As assessed by gene expression pattern, expression of cell surface markers, determination of clonal progenitor cells, and gene targeting studies of selected factors thought to act *in vivo*, the hematopoiesis found in differentiating EBs correlates temporally with yolk sac hematopoiesis<sup>44</sup>. The primitive erythroid lineage is transient and the first to appear, followed by macrophages, definitive erythroid cells, and mast cells, in the same

order as in the yolk sac<sup>44</sup>. In EBs, up to 5% of cells can represent clonable hematopoietic progenitors<sup>44</sup>. Lymphoid cells can be observed following extended cell culture period<sup>44</sup>. Both primitive and definitive hematopoiesis can be generated from differentiating EBs, however, primitive hematopoiesis is more abundant<sup>69</sup>. Many studies suggest that HSCs could be generated during ESC differentiation, but usually they present unsatisfying lympho-myeloid long-term engraftment in adult recipients, raising the concern again that the assay might not be appropriate for detecting the potential embryonic HSCs<sup>69</sup>. Possibly, the culture conditions cannot perfectly mimic embryogenesis<sup>69</sup>. When ESC are engineered to express Hoxb4 during differentiation, multilineage engraftments in primary and secondary recipients can be achieved, suggesting the involvement of ESC-derived HSCs<sup>69</sup>. However, since the lymphoid compartment is not as efficiently reconstituted as the myeloid compartment, these ESC-derived HSCs might not be completely similar to those derived from fetal liver or bone marrow<sup>44</sup>.

The hemangioblast, a clonable progenitor, first isolated during ESC differentiation is able to give rise to hematopoietic, endothelial, and vascular smooth muscle cells<sup>44</sup>. Although this progenitor was recently isolated from the posterior primitive streak of the mouse embryo<sup>44</sup>, it was not found in the yolk sac, suggesting its transient existence and rapid commitment to differentiation before reaching the presumptive territory of blood islands<sup>69</sup>. The hemangioblasts (blast colony forming cells) are detected in EBs prior to the apparition of hematopoiesis.

Although the hematopoiesis in early developing EBs (day 1-8) seems to reasonably follow a clean profile, correlating with the yolk sac hematopoiesis, the correspondence *in vitro-in vivo* is not as obvious following an extended cell culture period.

### **1.7.3 Generation of teratomas and teratocarcinomas**

When ESCs are injected subcutaneously into syngeneic mice, they create teratomas containing a disorganized mixture of differentiated cells derived from the three primary germ layers<sup>47</sup>. As opposed to these benign tumors, teratocarcinomas contain both differentiated and undifferentiated cells, allowing the transplantation of tumor cells into secondary recipients (malignant tumors)<sup>47</sup>. This ESC paradox, being oncogenic in a particular environment but totally suited for normal development when introduced in the mouse embryo, is fascinating and is drawing attention to the role of the environment and/or epigenetics in tumorigenesis. Lessons from early embryogenesis will probably improve our understanding of tumorigenesis.

Once more, cell extrinsic and cell intrinsic, positive and negative regulators, must be acting together on a perilous balance. Some factors must promote the proliferation, the characteristics, and the survival of tumor cells while others constraint these properties.

ERas is an example of a teratoma promoting factor. *ERas* is a Ras-like gene (small G protein constitutively active, because it is predominantly bound to GTP) expressed in undifferentiated ESC but not in differentiated ESC (treated with retinoic acid) or adult somatic cells<sup>70</sup>. *ERas*-null ESCs express the pluripotent marker gene (Oct4), are morphologically normal, proliferate slowly (particularly without MEFs) without defects in cell cycle properties, present reduced tumorigenicity by teratoma formation, but are germ-line transmitted in mouse without causing anomalies or infertility (note that *ERas* is located on the X chromosome and the ESC used are XY male)<sup>70</sup>. It was also shown that ERas interacts with phosphatidylinositol-3-OH kinase (PI3K) and mediates its effect through Akt (Akt phosphorylation is diminished in *ERas*-deficient cells and the ectopic expression of Akt rescues the proliferation and the teratoma formation defects)<sup>70</sup>. This was the first report suggesting that ESC tumorigenic properties could be uncoupled from their recognized self-renewal/pluripotency. However, mechanistically, the story is incomplete. First, the diminution of *ERras* KO ESCs' proliferation is not explained by the cell cycle profile, or by differentiation (morphology normal and expression of Oct4), suggesting a possible involvement of apoptosis. Cell death particularly needs to be addressed because *ERas* cDNA can only partially rescue the proliferation and teratoma formation defects, while the use of the *Akt* cDNA gives better yield.

So far, most genes reported to be implicated in teratomas (teratocarcinomas) formation also seem to regulate normal biological functions that are revealed during *in vitro* or *in vivo* differentiation. For example, *Pten* (tumor suppressor)-deficient ESCs generate teratocarcinomas faster and of larger size<sup>71</sup>. These are predominantly made of undifferentiated and neuronal cells, as opposed to wild type and *Pten*<sup>+/-</sup> ESC-derived teratocarcinomas that contain more differentiated cell types<sup>71</sup>. EBs generated from *Pten*<sup>-/-</sup> ESC are disorganized, the formation of the three primary germ layers is abnormal, and the natural process of cavitation, involving apoptosis, is not observed<sup>71</sup>. *Pten*<sup>-/-</sup> ESCs proliferate normally and show no difference in cell cycle properties, but they fail to contribute to chimeric mice<sup>71</sup>. *Pten*<sup>+/-</sup> ESC do contribute to chimeric mice and to the generation of heterozygote mice, but these animals present hyperplasia/dysplasia in some tissues (prostate, skin, colon)<sup>71</sup>. Moreover, they also develop tumors<sup>71</sup>. The intercross of *Pten*<sup>+/-</sup> mice demonstrates that *Pten*<sup>-/-</sup> embryos die *in utero* prior to E7.5<sup>71</sup>.

Alpha 5 ( $\alpha_5$ ) integrin is another protein that constrains teratocarcinoma growth. Integrins are heterodimeric ( $\alpha$  and  $\beta$  subunits) transmembrane glycoproteins acting as receptors binding to the extracellular matrix (fibronectin, laminin, collagen, etc.)<sup>72</sup>. They link the extracellular matrix to the cytoskeleton and signal transduction pathways<sup>72</sup>.  $\alpha_5$  and  $\beta_1$  integrin subunits dimerize to form a receptor that bind fibronectin<sup>73</sup>. Fetuses deficient in  $\alpha_5$  integrin die around E10-11, presenting defects in the posterior trunk and in the intraembryonic and extraembryonic vasculature<sup>72</sup>.  $\alpha_5$  integrin<sup>-/-</sup> ESCs form teratocarcinomas that are 8-times smaller than wild type or  $\alpha_5$  integrin<sup>+/-</sup> ESCs<sup>73</sup>. In comparison to controls,  $\alpha_5$  integrin<sup>-/-</sup> teratocarcinomas present a smaller undifferentiated compartment, which demonstrates a reduced proliferation and an increased apoptosis, and fewer ESC-derived vessels (<5%)<sup>73</sup>. In fact, although these tumors present derivatives from the three primary germ layers, the poor vasculature is mainly host-derived and of a smaller size than controls<sup>73</sup>. The extracellular matrix is also disorganized in these teratocarcinomas.  $\alpha_5$  integrin<sup>-/-</sup> ESCs were also induced to differentiate in attached EBs<sup>73</sup>.  $\alpha_5$  integrin<sup>-/-</sup> EBs present delays in growth, attachment and vasculature organization, but fibronectin deposition seems normal<sup>73</sup>. In summary, the analysis of embryogenesis, teratocarcinoma formation, and EB differentiation from  $\alpha_5$  integrin-deficient cells converges to a defect in vasculature network that might be mediated by anomalies in the remodeling of the extracellular matrix or in perivascular cells, etc<sup>73</sup>.

In conclusion, ERas, Pten, and  $\alpha_5$  integrin are just a few examples among a growing group of regulators that highlight the complex networks sustaining normal biological states, but also pathological conditions and tumorigenesis.

#### 1.7.4 General mechanisms underlying ESC differentiation

Again, both cell extrinsic and intrinsic mechanisms govern differentiation processes but also repress the self-renewal /pluripotency state. Of course, broad families of regulators are controlling *in vitro* and *in vivo* differentiation. This circuitry differs from lineage to lineage. Emerging regulators are microRNAs (miRNAs) and factors that control epigenetic modifications.

*Dgcr8*, a RNA-binding protein, acts with the RNase III enzyme Droscha in the processing of long primary miRNAs<sup>74</sup>. This protein is required for miRNA, but not for ribosomal RNA processing<sup>74</sup>. *Dgcr8*-deficient ESCs present abnormal differentiation characterized by an abnormal EB morphology and the expression of some differentiation marker genes, but with a failure to downregulate pluripotency marker genes<sup>74</sup>. These cells are still able to produce

ESC colonies<sup>74</sup>. Dicer is the protein implicated in the processing of long double-stranded RNA or miRNA precursor into mature effector RNA molecules<sup>18</sup>. Dicer-deficient ESCs form abnormal EBs, presenting little differentiation and a partial suppression of pluripotency marker gene (*Oct4*)<sup>18</sup>. These cells also fail to induce teratocarcinoma or to contribute to chimeric mice<sup>18</sup>. Together, these studies suggest that miRNAs play a role in silencing regulators of the undifferentiated state, but the mechanisms are not yet clearly understood. The miRNA silencing effects could be mediated by transcriptional gene silencing, by post-transcriptional gene silencing, by blocking translation, or by changes in the status of methylation and heterochromatin formation<sup>18</sup>.

Evidences support a role for NuRD, a nucleosome remodeling and histone deacetylation complex, during ESC differentiation. Mbd3, a methyl-CpG binding domain protein, is a component of the NuRD complex<sup>75</sup>. Mbd3<sup>-/-</sup> ESCs, engineered by gene targeting, present an abrogation of NuRD complex formation and a slower proliferation rate, but the proper expression of pluripotency marker genes such as *Oct4*, *Nanog*, and *Sox2*<sup>75</sup>. However, their differentiation in EBs is aberrant<sup>75</sup>. Mbd3<sup>-/-</sup> EBs express pluripotency-associated genes (*Oct4*, *Nanog*, *Rex1*), fail to activate some differentiation marker genes (for example: *Brachyury* or *Gata-6* are not activated but *Fgf5* is activated), express trophoectoderm marker genes, and finally, have no increase in apoptosis<sup>75</sup>. In fact, alkaline phosphatase positive ESC-like colonies (marker of undifferentiated state) can be derived from long-term, LIF-deprived, Mbd3<sup>-/-</sup> EBs' culture<sup>75</sup>. Mbd3<sup>-/-</sup> ESCs fail to contribute properly to chimeric mice following aggregation with morulas and show an abnormal distribution in embryos (E7.5)<sup>75</sup>. The presence of chimerism correlates with different anomalies in embryos<sup>75</sup>. Otherwise, mutant cells are constrained to extraembryonic tissues when embryos are normal (low chimerism)<sup>75</sup>.

In summary, microRNAs, nucleosome remodeling and histone deacetylation complexes such as NuRD are necessary for silencing the pluripotent state as well as to progress through differentiation. The assessment of potential interactions between these factors will be of major importance.

## 1.8 ARTICLE: Cell cycle checkpoints, telomeres and chromosome maintenance in mouse embryonic stem cells

Mélanie Bilodeau<sup>1</sup>, Josée Hébert<sup>2,3</sup> and Guy Sauvageau<sup>1,2,3\*</sup>

<sup>1</sup>Laboratory of Molecular Genetics of Stem Cells, Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montréal, Québec, Canada, H2W 1R7. <sup>2</sup>Department of Medicine, Montréal, Québec, Canada, H3C 3J7. <sup>3</sup>Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, H1T 2M2.

Review in preparation for *Cell Cycle*

\*Correspondence: Guy Sauvageau, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada, H3C 3J7. Email : [REDACTED]

### 1.8.1 Author contributions

Mélanie Bilodeau wrote the manuscript, prepared the figure and table under Guy Sauvageau guidance. **Table II** summarizes the work of many people. Josée Hébert performed the cytogenetic analyses. Array-based comparative genomic hybridization (aCGH) and the real-time quantitative PCR (Q-PCR) were performed by the services mentioned in the Acknowledgments section. Mélanie Bilodeau, Véronique Paradis, Nadine Fradet, Amélie Fredette, Simon Girard, Valeria Azcoitia and Jana Krosi generated the ESC clones.

## 1.8.2 Introduction

This review will visit selected genetic properties of mouse embryonic stem cells (ESCs), highlighting the differences observed when ESC are maintained in an undifferentiated or differentiated state. The focus will be on cell cycle regulation, telomeres, and chromosome maintenance.

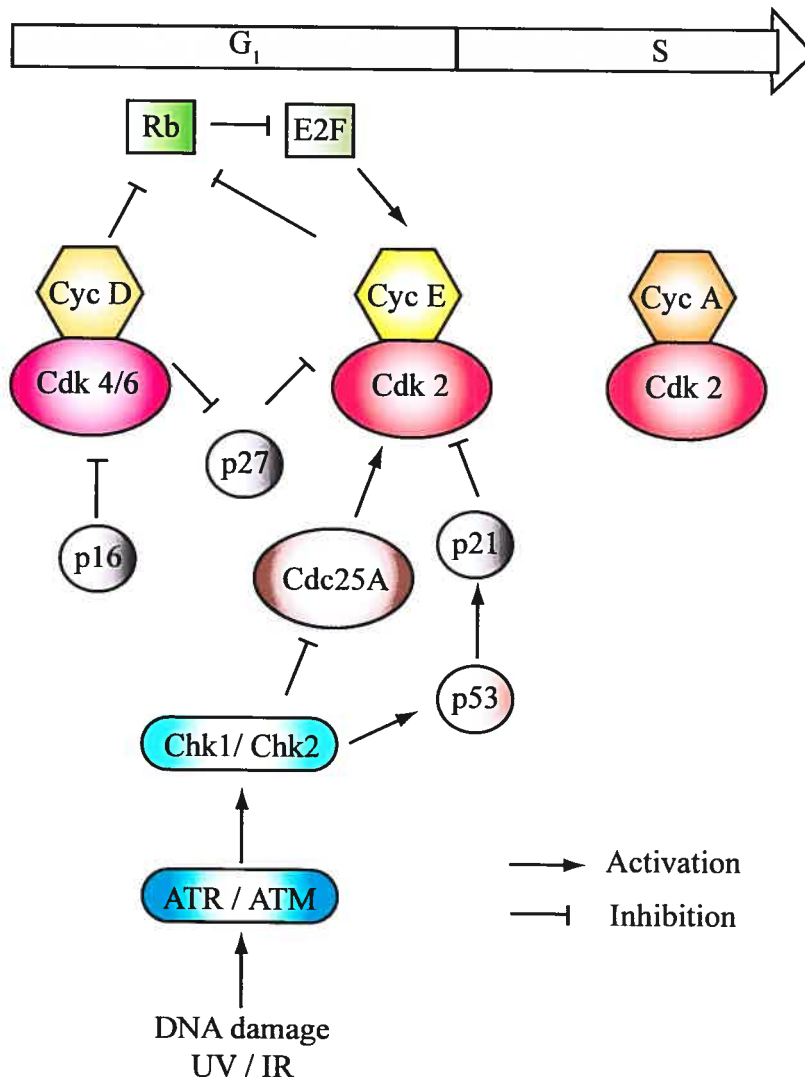
## 1.8.3 Cell cycle regulation differs between ESCs and differentiated cells

Cell cycle is typically subdivided in four phases:  $G_1$ , S,  $G_2$ , and M. DNA synthesis occurs in S phase. Cellular components and genetic material are partitioned between two daughter cells in M phase<sup>76</sup>.  $G_1$  and  $G_2$  (gaps) phases prepare cells to S and M phases, respectively<sup>76</sup>. The restriction point (R) subdivides early and late  $G_1$  phase. Early  $G_1$  is a mitogen-dependent phase as opposed to the late  $G_1$  phase. Cell growth (increased in cell size and proteins level) and division are coordinated<sup>76</sup>. Proper cell cycle progression is regulated by different checkpoints. Cells can be non-dividing in  $G_0$  quiescent stage, poised to cell cycle re-entry upon extracellular signaling, provided they are not terminally differentiated or senescent<sup>76</sup>.

At the molecular level, cell cycle is regulated by cyclin-dependent kinases (CDKs) that form heterodimeric complexes with cyclins<sup>76</sup>. In early  $G_1$  phase, Cdk4 and Cdk6 associate with D-type cyclins (D1, D2, and D3)<sup>76</sup> (**Figure 1-6**). In late  $G_1$  phase, Cdk2 associates with E-type cyclins (E1 and E2)<sup>76</sup> (**Figure 1-6**). These complexes phosphorylate retinoblastoma family proteins (Rb, p107, and p130), releasing E2F transcription factors<sup>76</sup> (**Figure 1-6**). Consequently, E2F responsive genes allow  $G_1/S$  transition<sup>76</sup>. In S phase, Cdk2 associates with A-type cyclins (A1 and A2)<sup>76</sup> (**Figure 1-6**). CDKs are regulated by two classes of CDK kinase inhibitors<sup>76</sup>. The INK4 family (p16, p15, p18, and p19) inhibits Cdk4 and Cdk6 activity<sup>76</sup>. The KIP family (p21, p27, and p57) inhibits Cdk2<sup>76</sup>.



**Figure 1-6  $G_1/S$  transition and  $G_1$  DNA damage checkpoint in somatic cells.**



Cdks (cyclin-dependent kinases) and cyclins (Cyc) heterodimeric complexes sequentially phosphorylate Rb (p107 and p130) in the  $G_1$  phase of the cell cycle, releasing E2F transcription factors and allowing entry in S phase. DNA damage, caused by UV or irradiation during  $G_1$  phase, induces cascades that inhibit Cdk2/Cyclin E complexes, preventing entry in S phase. P16, p27, and p21 are Cdk inhibitors.

### 1.8.3.1 $G_1/S$ transition

In a media supplemented with leukemia inhibitory factor (LIF) and a source of Bone Morphogenic Proteins (BMPs), ESCs are mainly maintained in an undifferentiated state. They self-renew very rapidly (e.g. replication time < 10 hours) presenting a short  $G_1$  phase (~1.5 hour) and a large proportion of the population is in the S phase of the cell cycle (e.g. > 60 %, ~10 % of cells in  $G_1$ )<sup>77-79</sup>. The S phase is the longest stage of their cell cycle (approximately 7 hours<sup>80</sup>).

In undifferentiated cells, the  $G_1$ -associated D-type cyclins (cyclins D1, D2, D3) are not or weakly expressed<sup>79</sup>. Cdk4 is present but with low kinase activity and therefore, no

sensitivity to p16<sup>ink4a</sup> inhibition is observed<sup>79,81</sup>. Cell cycle is possibly regulated by the G<sub>1</sub> phase-associated cyclin D3/Cdk6 (refractory to p16<sup>ink4a</sup> inhibition)<sup>82</sup>, cyclin E/Cdk2, and S phase-associated cyclin A/Cdk2<sup>83</sup>. The formal attestation has yet to come.

ESCs share many properties with cells from the inner cell mass of the blastocyst, although they cannot be considered as equivalent. In this regard, it is surprising that many of the cyclins and CDKs seem dispensable for proliferation in early embryogenesis. Embryonic development can proceed until E14.5 before *Cdk4*<sup>-/-</sup> *Cdk6*<sup>-/-</sup> double knockout (DKO) fetuses die due to hematological defects<sup>84</sup>. *Cdk2*<sup>-/-</sup> *Cdk4*<sup>-/-</sup> DKO fetuses succumb to heart defects around E15<sup>85</sup>. *Cdk2*<sup>-/-</sup> *Cdk6*<sup>-/-</sup> DKO mice are viable, although sterile<sup>84</sup>. Triple knockout (TKO) *Cyclin D1*<sup>-/-</sup> *D2*<sup>-/-</sup> *D3*<sup>-/-</sup> mice develop relatively normally until E13.5, but die prior to E17.5 from severe anemia and cardiac defects<sup>86</sup>. DKO *Cyclin E1*<sup>-/-</sup> *E2*<sup>-/-</sup> ESCs proliferate normally<sup>87</sup>. DKO *Cyclin E1*<sup>-/-</sup> *E2*<sup>-/-</sup> fetuses present cardiac and megakaryocyte defects, nevertheless some animals reach birth using tetraploid complementation to rescue the placental defect<sup>87</sup>. ESCs derived from these various compound mutant embryos (in addition to DKO *Cyclin E1*<sup>-/-</sup> *E2*<sup>-/-</sup> ESCs) should represent a good resource to investigate the functional redundancy between CDKs and Cyclins and their susceptibility to CDK inhibitors throughout cell cycle progression.

CDKs' activities are likely involved because the ectopic expression of p27<sup>Kip1</sup> induces cell cycle arrest in G<sub>1</sub><sup>79</sup>. However, the possibility of exit in G<sub>0</sub> stage or apoptosis was not formally excluded<sup>79</sup>. Whatever the relevant cyclin-CDK combination(s) implicated in this regulation, its major role is not to phosphorylate one of the classical pocket protein because TKO *p107*<sup>-/-</sup> *p130*<sup>-/-</sup>, *pRb*<sup>-/-</sup> ESCs proliferate normally<sup>88</sup>.

In addition, ESCs might lack a G<sub>1</sub> checkpoint upon DNA damage (**Figure 1-6**) for two reasons. One, p53-mediated response is partially ineffective because this protein fails to efficiently translocate to the nucleus and to induce p21 expression<sup>89</sup>. Furthermore, Chk2 is sequestered to centrosomes and fails to phosphorylate Cdc25A, preventing its degradation<sup>90</sup>. ESCs do not respond to DNA damage by arresting in G<sub>1</sub><sup>90,91</sup>. They rather trigger p53-independent apoptosis<sup>89</sup>, arrest in G<sub>2</sub> phase<sup>90,92</sup>, or differentiate by the p53-mediated suppression of *Nanog* (a regulator of ESC self-renewal and pluripotency), possibly creating progenies more prone to DNA damage response<sup>92</sup>. In brief, the most striking feature of G<sub>1</sub> phase regulation in undifferentiated ESCs is the lack of controls known to operate in somatic cells.

Nucleostemin (NS) has been identified in mammals as a nucleolar protein which seems to regulate G<sub>1</sub>/S progression in embryonic cells<sup>93</sup>. *NS*<sup>-/-</sup> embryos die around E4.0<sup>93</sup>.

Blastocysts present failure to enter in S phase<sup>93</sup>. However they do not show signs of cell death or terminal differentiation and their cell nuclei are normal<sup>93</sup>. The mechanism behind this effect is unknown, but is p53-independent<sup>93</sup>. Derivation of *NS*<sup>-/-</sup> ESCs is unsuccessful<sup>93</sup>. Nucleostemin is found in the nucleoli and possibly regulates rRNA processing and ribosome assembly. Since cell growth in G<sub>1</sub> phase depends on mRNA and protein synthesis, altering these processes might arrest or delay progression in this phase. Presumably, ESCs need to spend a minimum of time in G<sub>1</sub> to reach the necessary size to perform symmetric cell division.

*Dgcr8* is a RNA-binding protein that assist Drosha in the processing of microRNAs<sup>74</sup>. *Dgcr8*-deficient ESCs present an extended doubling time and an accumulation in G<sub>1</sub> (22% of the cells rather than 14% in the controls), without obvious sign of differentiation or apoptosis<sup>74</sup>. *Dgcr8* is necessary for the processing of most long primary miRNAs (if not all), but not for ribosomal RNA processing<sup>74</sup>. More studies will be needed to rule out the possibility of cells resting in a G<sub>0</sub> stage, a stage not yet described in ESC. Interestingly, these ESCs also show an EB differentiation defect<sup>74</sup>.

Once ESCs are induced to differentiate upon LIF (leukemia inhibitory factor) removal, the length of the G<sub>1</sub> phase increases<sup>79</sup>. D-type cyclins are up-regulated and associate with Cdk4 in complexes presenting kinase activity<sup>79</sup>. More cyclin E/Cdk2 complexes with kinase activity are detected in the initial phase of *in vitro* differentiation<sup>79</sup>. Cells become sensitive to p16<sup>ink4a</sup> and p27<sup>Kip1</sup> inhibition<sup>79</sup>. p27<sup>Kip1</sup> expression increases and is detected in complexes with Cdk4<sup>79</sup>. p53 protein level is reduced, but its transcriptional activity is increased, inducing the expression of *p21* while repressing *Nanog*<sup>92</sup>. This improved activity of p53 is mediated in part by the phosphorylation of Ser 315, a potential CDKs substrate<sup>92</sup>. Differentiated cells possess a DNA damage checkpoint induced by p53, leading to cell cycle arrest (mostly in G<sub>1</sub>)<sup>89</sup> or to apoptosis. Compound *p107*<sup>-/-</sup>, *p130*<sup>-/-</sup>, *Rb*<sup>-/-</sup> ESCs present a limited differentiation capacity inside teratocarcinomas while the size of the proliferating compartment in these tumors is increased by 15-fold compared to control cells<sup>94</sup>. The induction of differentiation switches the regulation from a pocket-proteins-independent to a pocket-proteins-dependent mode, but it remains unclear whether this effect is mediated through cell cycle controls or through differentiation functions independent of the cell cycle, or both.

### 1.8.3.2 S phase

So far, there are no differences reported between ESC and differentiated cells in the S phase. During differentiation, the proportion of cells found in S phase is reduced<sup>95</sup>, likely as a consequence of an increase in  $G_1$ <sup>79</sup>, changing the ratio of cells in all the other phases.

Three checkpoints are known to delay the progression of cells through the S phase in response to genotoxic stresses<sup>96</sup>. They share some components, in addition to the property of being independent of p53: replication, S-M, and intra-S-phase checkpoints<sup>96</sup>. The replication checkpoint is activated when the replication fork is stalled because of stress such as depletion of deoxyribonucleotides, inhibition of DNA polymerase by chemical or physical constraint<sup>96</sup>. The exact pathway regulating the S-M checkpoint is not known in mammals, but studies in yeast suggest that it is mediated by the same sensors of the replication checkpoint<sup>96</sup>. The intra-S-phase checkpoint is activated upon double-stranded break induced outside of the active replicons<sup>96</sup> (for example: caused by irradiation). Readers interested in the pathways involved are referred to a recently published review<sup>96</sup>.

Presumably undifferentiated ESCs do have an intra-S-phase DNA damage checkpoint. Wild-type and  $p53^{-/-}$  ESCs treated with ultraviolet irradiation accumulate in S phase and present a temporally delayed progression through the phase<sup>97</sup>. RAD50 is part of a complex (NBS1-MRE11-RAD50) thought to detect double-stranded breaks<sup>96</sup>. Rad50 knockdown is lethal for ESCs and early embryos (around E6.5)<sup>98</sup>. E6.5  $Rad50^{-/-}$  embryos present a decreased proliferation (assess by BrdUrd incorporation) and no change in apoptotic status<sup>98</sup>. Chk1, a checkpoint kinase thought to be involved in intra-S-phase checkpoint, is required for the proliferation of inner cell mass cells and the generation of ESCs<sup>96,99</sup>. As early as E3.5,  $Chk1^{-/-}$  blastocysts present cells with abnormal nucleus and increased apoptosis<sup>99</sup>.

### 1.8.3.3 $G_2/M$ transition

The length of the  $G_2/M$  transition in undifferentiated ESCs is roughly one hour<sup>80</sup>. The decatenation checkpoint in  $G_2$  phase, induced when chromosomes are entangled, retards the entry of cells in mitosis, preventing aneuploidy in daughter cells<sup>100</sup>. Suspected regulators of this pathway are: ATR, Polo-like kinase 1, BRCA1, and Werner's syndrome helicase<sup>100</sup>. This checkpoint is less efficient in mouse ESCs than in primary embryonic fibroblasts (MEFs)<sup>100</sup>. However when ESCs are induced to differentiate, the decatenation checkpoint efficiency improves to a level similar to the one observed in MEFs<sup>100</sup>. The molecular basis of this observation is unknown.

### 1.8.3.4 The M phase

The anaphase-promoting complex (APC) is a major regulator of M phase progression. APC, a multisubunit ubiquitin ligase complex, contains at least 13 subunits and three adaptor proteins including Cdc20<sup>101</sup>. Apc2 and Apc11 are the two catalytic subunits<sup>102</sup>. APC complex ubiquitinates Securin, an inhibitor of Separase which inactivates cohesin complexes (by cleaving a subunit)<sup>102</sup>. Cohesin complexes insure sister chromatid cohesion, possibly by forming rings around DNA<sup>102,103</sup>. This cascade allows progression through anaphase. In addition, APC allows the exit from mitosis by targeting the mitotic cyclins for degradation<sup>102</sup>. The functions of APC complexes-adapters are regulated in different cell cycle phases by Cyclin A/Cdk2, Mad2 and others<sup>101</sup>. APC is important in early embryogenesis since no *Apc2*<sup>-/-</sup> embryos are found at E6.5, whereas heterozygous embryos reach birth at expected frequencies<sup>102</sup>.

The spindle assembly checkpoint inhibits the transition in anaphase, mediated by APC, and arrests cells in mitosis until the chromosome kinetochores are properly attached to spindle microtubules<sup>104</sup>. Many proteins are implicated in this checkpoint<sup>105</sup>. When kinetochores are not attached to spindle microtubules, Mad2, Mad3, Bub3, and Cdc20 form a complex inhibiting Cdc20<sup>104</sup>.

So far no report suggests differences in ESC mitosis versus differentiated cells, in agreement with observations seen during early embryogenesis. *Mad2*<sup>-/-</sup> embryos die around E6.5-7.5 with extensive apoptosis with another proportion suspected to do so prior to implantation<sup>106</sup>. These embryos present abnormal segregation of one or few chromosomes<sup>106</sup>. *Mad2*<sup>-/-</sup> blastocysts maintained *in vitro* are insensitive to nocodazole treatment (microtubules inhibitor)<sup>106</sup>. They can be maintained up to E6.5, after which their inner cell mass degenerates<sup>106</sup>. The E4F protein is an additional candidate to regulate M phase in embryonic cells. As opposed to controls, E3.5 *E4F*<sup>-/-</sup> blastocysts fail to hatch from the zona pellucida and to form an inner cell mass outgrowth<sup>107</sup>. Blastocyst *E4F*<sup>-/-</sup> cells are blocked in the prometaphase stage and present an activated spindle checkpoint<sup>107</sup>.

### 1.8.3.5 Is ESC cycle regulation an artifact of cell culture conditions?

Different factors influence the properties of ESC in culture: the amount of LIF extrinsically introduced, the serum constituents which vary from batch to batch, the presence of a feeder layer (MEFs), the type of matrix used when cells are cultivated without a feeder

layer, the cell density, the karyotype of the cells (e.g., proliferation of trisomy 8 cells), etc. Some of these factors likely modify the cycling properties of ESCs and this possibility should be taken into consideration when comparing different studies on the subject. ESCs can reasonably be used as a model to better understand many cell cycle properties that are both shared with the highly proliferative embryonic cells (embryogenesis ~E3.5-E6.5), but strangely also with teratocarcinoma cells.

In an undifferentiated state, no evidence suggests that ESCs present a cell cycle restriction point dependent on external mitogenic stimuli. Serum deprivation does not arrest ESCs, only a small alteration in cell phase distribution is observed with a greater proportion of cells in  $G_0$ - $G_1$  to the detriment of the S phase ( $G_2$ -M phases are unchanged)<sup>108</sup>. Consequently, when starved cells are re-exposed to serum, they show no sign of synchronization<sup>108</sup>. These observations suggest that serum components are dispensable for inducing or sustaining ESCs cycling properties, at least for a short period (38hrs<sup>108</sup>). In fibroblasts, mitogen signaling allows the progression through the restriction point by down-regulating p27<sup>Kip1</sup><sup>109</sup>. ESCs express only weak levels of CDK inhibitors (p16<sup>ink4a</sup>, p27<sup>Kip1</sup> and p21<sup>Cip1</sup>)<sup>79,82</sup>. Perhaps that ESCs expanding in very tight colonies go through  $G_1$  rapidly because they produce their own mitogenic signal (on a technical point of view, ESCs seem to proliferate faster when the cell density is higher). Therefore, undifferentiated ESCs might produce a factor acting in an autocrine-paracrine or in a cell-cell interaction manner. Alternatively, ESCs don't need mitogenic signal and are permanently in a cycling mode.

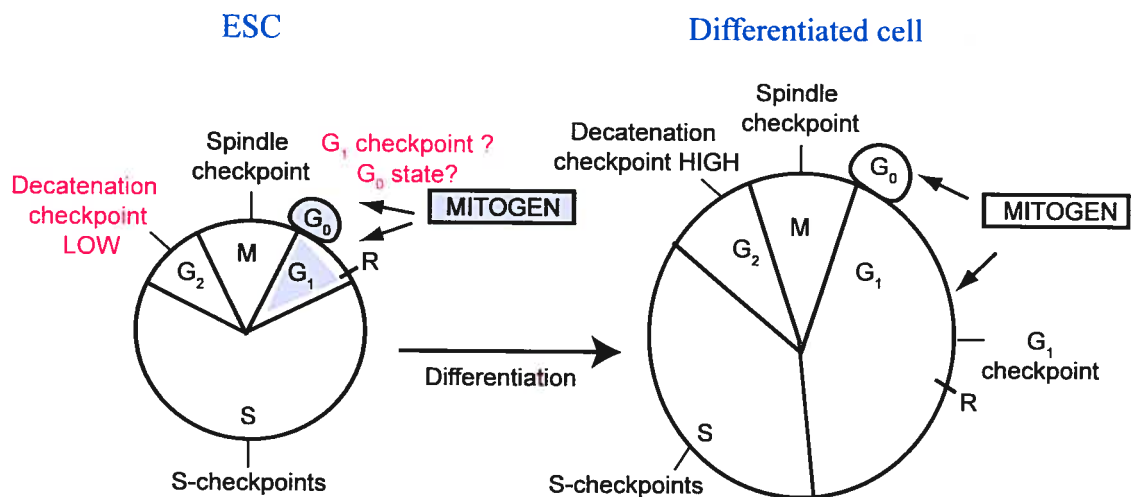
The mechanism that seems to slow down both the ESC and the embryonic cell cycles goes hand in hand with the commitment to differentiate, which is induced in part by changes in external stimuli. In the case of ESCs, differentiation is induced by the removal of LIF or BMPs from the serum. Can an ESC display a long cell cycle time as seen in somatic stem cells or in more differentiated cells? The potential link between the cell cycle time and the (pluri)potency is intriguing but, has not been unequivocally addressed.

Finally, ESC cycle properties present similarities with cells found in early embryos and from that perspective, they are not artifacts. However, the permanent cycling of ESCs in culture is an artifact, because the similar condition of embryonic cells during embryogenesis is temporary, after what, the proliferation is highly regulated and limited during the adult life. Thereafter, the closest entities behaving like ESCs in term of non-exhaustible proliferation... are tumor cells.

In conclusion, undifferentiated ESCs present a cell cycle regulation distinct from differentiated cells (**Figure 1-7**). Some steps still need clarification (**Figure 1-7**).

**Figure 1-7 Cell cycle regulation of undifferentiated and differentiated ESCs.**

ESC cycle is more rapid than that determined in differentiated cells, mainly because of a shorter  $G_1$  phase which is distinct between the two cell types; differences in CDKs, cyclins, and other regulators such as p16. ESCs are frequently recognized as lacking the  $G_1$  DNA damage checkpoint or lacking the entire  $G_1$  phase, however, they may possess a minimal regulation of this phase, possibly at least at the level of cell size. It is not clear if extracellular stimuli are required before reaching the restriction point (R), where cell cycle progression becomes independent of external stimuli, nor if they can be in a  $G_0$  quiescent state given the culture conditions used. In the  $G_2$  phase, the decatenation checkpoint is less efficient (LOW) in ESCs compare to somatic cells (HIGH). So far, no differences have been proposed concerning the regulation of the S and M phases in ESC versus differentiated cells.



**1.8.4 Telomere maintenance differs between ESCs and their differentiated progenies**

The ends of chromosomes are protected by telomeres, a stretch of short G-rich repeat sequences terminated by a single-stranded overhang forming a T loop<sup>95,110</sup>. Telomere size varies from species to species and displays significant differences between subspecies, such as in the mouse<sup>95,110</sup>. Telomeric repeats are bound by factors including TRF1 and TRF2, in addition to nucleosome arrays presenting histone modifications characteristic of

heterochromatin<sup>110</sup>. Telomeric DNA is progressively loss because of incomplete replication, degradation, oxidative stress, and other mechanisms<sup>95</sup>. Cellular proliferation is inhibited by a replicative senescence checkpoint when telomeres become too short<sup>95,111</sup>. Telomerase, a complex formed by a reverse transcriptase (*Tert*), a RNA component (*Terc*), and other proteins, adds telomeric repeats to maintain telomeres<sup>110</sup>. In general, telomerase is expressed in germ and stem cells, in cells that undergo rapid expansion such as lymphocytes and keratinocytes, frequently in tumor cells, but rarely in somatic cells<sup>110,111</sup>. Telomerase expression in cells remains incompletely understood but, is in part regulated by diverse genetics and epigenetics mechanisms<sup>110</sup>.

Undifferentiated ESCs possess telomerase activity<sup>112</sup>. *Terc*-deficient ESCs proliferate for up to ~300 divisions similarly to control<sup>112</sup>. These cells then demonstrate impaired proliferation from 300-450 divisions before they stop dividing (around 450 divisions)<sup>112</sup>. Correlating with the increasing number of divisions, telomere size decreases in *Terc*<sup>-/-</sup> ESCs<sup>112</sup>. Moreover, metaphase spreads demonstrate an increased number of aneuploidies and end-end chromosome fusions in *Terc*<sup>-/-</sup> ESCs, likely causing the proliferation arrest<sup>112</sup>. In two independent *Terc*<sup>-/-</sup> ESC lines, rare subpopulations of cells resolve the proliferation arrest, without re-expressing telomerase, while the controls did not change in proliferation status for two years<sup>113</sup>. One of these ESC line continued to present diminution of telomere sizes while the second line presented a stabilization, characterized by the addition of a fragment containing both telomeric and non telomeric sequences on most chromosomes (85%)<sup>113</sup>. Notably, following more than 650 divisions, both *Terc*<sup>-/-</sup> lineages contained cells with abnormal karyotype, with 68-100% of chromosomes fused<sup>113</sup>. This study shows that ESCs can proliferate extensively with massive cytogenetic anomalies, either in maintaining their telomeres in a telomerase-independent fashion or by another unclear mechanism.

Upon ESC differentiation, telomerase activity is reduced<sup>111</sup>. The kinetics of telomerase expression during differentiation presents differences according to cell culture conditions and has not been evaluated convincingly<sup>111</sup>. A study correlates *mTert* expression, using RT-PCR and a *mTert*-GFP reporter randomly integrated in the genome, with telomerase activity assessed by a TRAP-ELISA assay during embryoid bodies (EBs) differentiation<sup>111</sup>. It would have been informative during this changing process, to correlate the expression of telomerase with a marker of undifferentiated ESC (for example Oct4) to get a better idea of the characteristics of cells presenting telomerase activities in this heterogeneous population. Nevertheless, at day 4 of differentiation, cells dissociated from EBs could be classified in three subpopulations according to the GFP expression (*mTert* promoter) agreeing with the



TRAP-ELISA assay : high, intermediate, and undetectable telomerase activity<sup>111</sup>.

Mechanisms that allow telomere maintenance are different between ESC and differentiated cells. For example, *Rtel* (Regulator of telomere length) is a gene encoding a helicase-like protein, identified as a candidate for the control of telomere sizes between subspecies of mice (distal part of chromosome 2q)<sup>95</sup>. Mouse embryos deficient in *Rtel* die around E10-11.5 from defects in tissues presenting rapid proliferation: nervous system, heart, vasculature, and extraembryonic tissues<sup>95</sup>. In ESC, *Rtel* is found in the nucleus, but does not colocalize with telomeres<sup>95</sup>. *Rtel*<sup>-/-</sup> ESCs hardly recover from thawing, their telomeres are smaller than control ESCs (68% smaller), but they proliferate normally and have a normal karyotype<sup>95</sup>. However, upon differentiation, the picture is completely different. *Rtel*<sup>-/-</sup> EBs are significantly less numerous and smaller than controls<sup>95</sup>. They have less cells in S phase (4% rather than 34% for EBs day3), have an increased cell death index, and have numerous chromosome anomalies (end-end fusions, chromatid gap, cluster of joined chromosomes, etc.)<sup>95</sup>. *Rtel* regulates both the integrity of genomic and telomeric DNA in differentiated cells by an uncertain mechanism that might involve the resolution of secondary G-rich chromatin structure<sup>95</sup>. The phenotype of *Rtel*<sup>-/-</sup> embryos is not dependent on a functional p53 gene<sup>95</sup>.

Interestingly in undifferentiated ESC, *Rif1* promoter, a highly expressed regulator of telomere length, is co-occupied by pluripotency associated transcription factors Oct4, Sox2, and Nanog<sup>49</sup>.

### **1.8.5 Genetic and chromosome anomalies in ESCs: comparison with somatic cells.**

It was reported, from the limited analyses of two loci (*Aprt* and *Hprt1*), that the frequency of spontaneous mutations in ESCs is significantly lower than in somatic cells<sup>114</sup>. Moreover, the type of mutations seen in undifferentiated ESCs differs from somatic cells<sup>114</sup>. Looking at cells heterozygous for the *Aprt* locus (and hemizygous for the *Hprt1* locus because located on the X chromosome), a team noticed that the loss of heterozygosity (~80% of spontaneous mutation) in MEFs is caused by mitotic recombination while in ESCs it is mostly due to chromosome loss and reduplication (non-disjunction, 57% of these events)<sup>114</sup>. Since these results concerned only one autosomal locus (*Aprt*), located on mouse chromosome 8, caution should be taken before generalizing because chromosome 8 trisomy is commonly found in ESC<sup>115</sup>. Moreover, this anomaly gives a proliferative advantage to the ESC carriers that rapidly take over the culture to the detriment of euploid cells<sup>115</sup>.

Nonetheless, this phenomenon has been suspected and used before by many researchers. Fifteen years ago, ESCs targeted by a vector containing a *Pgk*-neomycin resistance gene and maintained in increasing concentration of geneticin (G418) were discovered to present homozygous alterations, while preserving an euploid karyotype<sup>116</sup>. Initially tested on 4 loci, the occurrence was estimated to happen at a rate of  $1.3 \times 10^{-5}$  per cell per generation<sup>116</sup>. Since then, this time-saving approach has been used numerous times, for different loci. From the analysis of 6 loci targeted in such a way, distributed on four mouse chromosomes (2, 5, 10, 17), it was noticed that the regions involved in the loss of heterozygosity were extensive and not only localized around anchor sites<sup>117</sup>. In fact, with the use of a limited number of polymorphic DNA markers, gene conversion was discarded as the cause to the benefit of mitotic recombination or chromosome loss combined with duplication<sup>117</sup>. No matter which of the two last hypothesis is right (maybe both are right), it results in large chromosomal regions of uniparental disomy<sup>117</sup>.

Trisomies of chromosome 8 and 11, the loss of chromosome Y (2% of ESC clones), and other genetic anomalies have been detected in mouse ESC lines<sup>62,115</sup>. Accordingly, we detected by cytogenetic and array-based comparative genomic hybridization (aCGH) analyses, chromosome 1, 12, and 14 trisomies in addition to previously reported anomalies, either in R1 ESC<sup>61</sup> clones containing a single proviral integration (primary clones, experiments A-C), or an engineered chromosomal deletion (tertiary clones) (**Table II**). These independent experiments revealed inconsistency between genetic anomaly frequencies (**Table II**). Most of the primary clones generated in experiment C demonstrated complex karyotypic anomalies, although they were submitted to less genetic manipulations and *in vitro* passages than tertiary clones (**Table II**). Half of the primary clones generated in experiment B were trisomic (**Table II**). These analyses highlighted the preponderance of specific trisomies among others cytogenetic anomalies, possibly due to events that occurred during *in vitro* culture. For this, we designed a real-time quantitative PCR (Q-PCR) strategy to rapidly screen for four different trisomies (chromosome 1, 8, 11, and 14), according to our own experience and to the work of others (**Table II**). These assays were conducted with genomic DNA extracted from 282 primary clones (**Table II**, experiments D-F). Trisomies were suspected for 10-22% of these primary clones, depending on the experiment (**Table II**, experiments D-F). Chromosome 1, 8, 11, or 14 potential trisomies were identified in each of these experiments (**Table II**, experiments D-F). Occasionally, some trisomies were slightly more represented (e.g. trisomy chromosome 14, or 8 and 11 in experiment D or E, respectively) (**Table II**).

**Table II • Frequency of genetic anomalies detected in mouse ESC clones.**

	No. of clones analyzed	Number of clones with or without anomalies (percentage of clone analyzed)							
		No. of clones free of anomalies	Single Trisomy 1	Single Trisomy 8	Single Trisomy 11	Single Trisomy 12	Single Trisomy 14	Combination of Trisomies	Other anomalies <sup>d</sup>
<b>Cytogenetic or aCGH analysis</b>									
Primary clones <sup>a</sup> (A)	9	7 (78%)	1 (11%)	0	0	0	0	0	1 (11%)
Tertiary clones <sup>b</sup>	14	10 (72%)	1 (7%)	1 (7%)	0	0	0	0	2 (14%)
Primary clones (B)	8	3 (38%)	0	1 (12%)	1 <sup>e</sup> (12%)	1 (12%)	1 <sup>f</sup> (12%)	0	1 (12%)
Primary clones (C)	11	1 (9%)	0	1 (9%)	0	0	0	2 <sup>g</sup> (18%)	7 (64%)
<b>Q-PCR<sup>c</sup></b>									
Primary clones (D)	96	(<90%)	1 (1%)	3 (3%)	1 (1%)		5 (5%)	0	
Primary clones (E)	91	(<78%)	2 (2%)	10 (11%)	7 (8%)		3 (3%)	0	
Primary clones (F)	95	(<80%)	3 (3%)	5 (5%)	3 (3%)		4 (4%)	5 (5%)	

<sup>a</sup>Primary clones contain one integration of a replication-incompetent retrovirus (retrovirus A1) and were generated in independent experiments (A-F). <sup>b</sup>Tertiary clones contain a chromosomal deletion induced by a retroviral-based Cre-*loxP* system. The anomalies observed were not present in parental clones from which they were derived (primary clones) and were not suspected to be induced by the rearrangements, however, they could be preserved as a compensatory event. <sup>c</sup>The Q-PCR approach was employed to pinpoint potential trisomies of chromosomes 1, 8, 11, 14 but without assessment for other anomalies. <sup>d</sup>Other anomalies include single or multiple losses of chromosome, gain or loss of genomic segments, but exclude only the loss of chromosome Y. <sup>e</sup>Also loss of chromosome 10 and presence of a marker chromosome. <sup>f</sup>Also presence of a marker chromosome. <sup>g</sup>One clone with trisomies of chromosome 8 and 14 and another clone with trisomies of chromosome 1, 8, 19 combined with additional anomalies. No., number; aCGH, array-based comparative genomic hybridization; Q-PCR, real-time quantitative polymerase chain reaction.

Actually, it is not clear how these mutations or aneuploidies are generated in ESCs. Some of them are probably overrepresented or underrepresented, because they are advantageous (for example, lead to increased proliferation<sup>115</sup>) or detrimental to the ESCs, respectively. The reduced efficiency of the decatenation checkpoint for entangled chromosomes, observed in ESCs, is one hypothetical cause<sup>100</sup>. Possibly the culture conditions also play a role (serum batch, culture in presence or absence of MEFs, regular maintenance depending on the experimenter, etc.).

An interesting observation was made concerning human ESCs (hESCs). hESCs maintained by manual passaging, a technique using a pasteur pipette to break the colonies in clumps of 10-100 cells, can preserve a stable karyotype for more than 100 passages<sup>200</sup>.

However, if they are dissociated with a non enzymatic (cell dissociation buffer) or an enzymatic (collagenase followed by trypsin) method, they acquire genetic anomalies in 25 passages or less (two hESC lines tested)<sup>200</sup>. These anomalies usually correspond to trisomies of chromosome 12 and 17, but also of 14, and sometimes an additional chromosome X is observed<sup>200</sup>. It would be of interest to look for the presence and the synteny of known or candidate regulators of self-renewal located on mouse and human chromosomes associated with ESC trisomies.

### **1.8.6 Concluding remarks**

With the increasing interest regarding the therapeutic potential and the biological properties of pluripotent stem cells (ESCs, germ cells, etc.) and somatic stem cells (hematopoietic, neuronal, etc.), long-lasting dogmas should be reconsidered about cell cycle checkpoints and the incidence of chromosomal instability in these cells. From our perspective and experience on these points, a lot of information found in the literature or intuitively taught concerns somatic cells and excludes ESCs.

### **1.8.7 Methods**

**Table II** summarizes unpublished and published information related to genetic anomalies detected in mouse ESC clones. Primary and tertiary ESC clones were generated as described in the manuscript presented in Chapter 2<sup>118</sup>. Cytogenetic (spectral karyotyping=SKY) and aCGH analyses will be detailed in Chapter 2; and Q-PCR assays in Chapter 3.

### **1.8.8 Acknowledgments**

We thank Véronique Paradis, Nadine Fradet, Amélie Fredette, Simon Girard, Valeria Azcoitia and Jana Krosi for the generation of the ESC clones; Claude Rondeau and Pauline Lussier for cytogenetic technical assistance; Pierre Chagnon (IRIC Genomics Core Facility) and Tara MacRae for the Q-PCR experiments; John Cowell, Michael R. Rossi and Devin McQuaid for the aCGH service (Roswell Park Cancer Institute, Buffalo); Mélanie Fréchette and Nadine Mayotte for technical support and Andras Nagy for providing the R1 ESCs (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto). This work was mostly supported by a grant from Génome Québec to Guy Sauvageau and in part by a grant from the Réseau de Recherche en Transgénèse du Québec to Josée Hébert. Mélanie Bilodeau is a recipient of a Canadian Institutes of Health Research (CIHR) studentship and Guy Sauvageau is a recipient of a Canada Research Chair in molecular genetics of stem cells and a scholar of the Leukemia Lymphoma Society of America.

## PART III: INTRODUCTION TO RETROVIRUSES

Retroviral gene transfer is a commonly used methodology to deliver DNA to target cells. This approach, apparently predictable and simple, hides in fact the amazing and complex biology of retroviruses. Because of this paradox, structural characteristics and the natural life cycle of typical retroviruses will be reviewed in this section. Then retroviral vectors as well as packaging and transduceable cells will be discussed. Considerations regarding the use of retroviruses and packaging cell lines will be highlighted and further discussed in **Appendix II**. Finally, retroviral integration will be introduced because it is central to the work presented in this thesis.

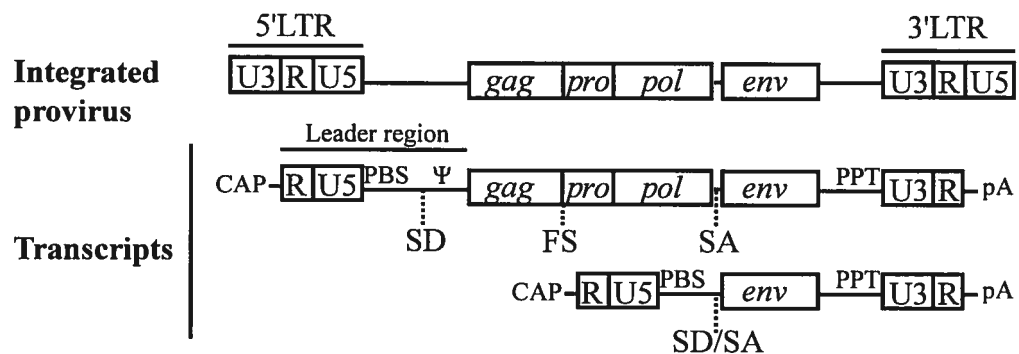
### 1.8.9 Structural characteristics

Retroviruses are provided with an external envelope and a diploid genome consisting of RNA. In a sophisticated process, this genome is reverse-transcribed in double-stranded DNA and ultimately integrated into a host cell genome. The viral RNA contains at least four coding domains: *gag*, *pol*, *env*, and *pro*<sup>119</sup> (**Figure 1-8**). *Gag* encodes matrix proteins, capsid proteins, and nucleoproteins<sup>119</sup>. *Pol* encodes enzymes: the reverse transcriptase and the integrase<sup>119</sup>. *Env* and *Pro* are responsible for the viral envelope proteins and the protease, respectively<sup>119</sup>. Simple retroviruses possess a genome organized as described above, whereas complex retroviruses have additional coding domains<sup>119</sup>.

The size of a replication-competent retrovirus genome is in the range of 7-12 kilobase pairs<sup>119</sup>. By partially or totally removing the viral sequence from the replication domains mentioned previously (e.g. *gag*, *pol*, *env* and *pro*), space is created to accept other coding sequences that can be transduced by the virus, given that the lost functions are provided *in trans* by an external replication-competent retrovirus (e.g. helper virus) or integrated inside a packaging cell line (in the form of a provirus or integrated plasmids). The first tactic is used naturally by oncogenic viruses while the second has been developed by researchers to create a new gene transfer tool.

### Figure 1-8 General structure of a simple replication-competent retrovirus.

In the integrated provirus, transcription is directed by the enhancer/promoter in the U3 region of the 5'LTR (long terminal repeats). *Gag*, *pro*, *pol*, and *env* encode essential proteins necessary for virion assembly and maturation (e.g. packaging functions). Env proteins are translated from a spliced transcript. Gag proteins and Gag-pro-pol precursor polyproteins (produced by frameshift or translational read-through) are translated from full length transcripts. Full length transcripts also serve as viral genomic RNA and are encapsidated because of their encapsidation sequence ( $\psi$ ). CAP, RNA 7-methylguanosine cap; PBS, transfer RNA-binding site; PPT, polypurine tract; SD, splice donor site; SA, splice acceptor site; FS, frameshift site; pA, polyadenylation signal. Adapted from Coffin JM., et al., 1997<sup>119</sup>.



#### 1.8.10 Retroviral life cycle

Glycoproteins at the surface of retrovirus mediate the attachment to receptors located on the plasma membrane, which precede the viral entry (nucleoproteins core) by membrane fusion either at the cell surface or in internalized endosomes<sup>120</sup>. This interaction is specific and determines the retrovirus tropism. Once in the cytoplasm, viral core uncoating occurs by a uncertain mechanism and the reverse transcriptase allows the transcription of the single-stranded, positive-sense, viral genomic RNA in double-stranded DNA<sup>119,121</sup>. The components of the reverse transcriptase complex vary according to the virus studied, but in the case of Moloney murine leukemia virus (Mo-MLV) includes the viral genome, the reverse transcriptase, the integrase, and the capsid proteins<sup>121</sup>. The reverse transcription process involves sequences located at the extremity of the viral RNA: the transfer RNA-binding site (PBS), the polypurine tract (PPT), and the LTRs (long terminal repeats containing three regions: U3-R-U5)<sup>119</sup> (Figure 1-8). Once the double-stranded DNA is formed, the protein complex is referred to as the preintegration complex (PIC)<sup>121</sup>. In the case of Mo-MLV, it includes the viral DNA, the viral capsid proteins, the viral integrase, and possibly

additional cellular proteins<sup>121</sup>. PICs need to reach the nucleus probably by using the cellular cytoskeleton for displacement<sup>121</sup>. PICs are large complexes, larger than the nuclear pores, excluding that they enter the nucleus by a passive mechanism unless they enter after the nuclear membrane breakdowns at mitosis<sup>121</sup>. The viral DNA gets integrated permanently into the host genome. The integration process will be described in an upcoming section (1.8.15). An integrated retrovirus is referred as a provirus. This entity is transcribed by host RNA polymerase II and replicated as part of the cellular genome<sup>120</sup>. Transcription is regulated by the cellular machinery interacting with the promoter in the LTR<sup>119</sup>. Both spliced and unspliced mRNA (and full length viral genomic RNA) are formed (**Figure 1-8**) and exported from the nucleus<sup>120</sup>. Translation of viral proteins occurs in the cytoplasm and is mediated by cellular ribosomes<sup>119</sup>. Viral proteins and full length RNA assemble together at the cell periphery, virions are released by budding of the plasma membrane, and subjected to maturation induced by viral and cellular proteases<sup>119</sup>. At least for Human immunodeficiency virus (HIV-1), different host-derived proteins are encapsidated in the virions<sup>120</sup>. Some might be trapped in a random manner, but others seem specifically recruited<sup>120</sup>. These are suspected to play a role during the viral life cycle<sup>120</sup>.

## 1.8.11 Properties of retroviral vectors

### 1.8.11.1 Structure of a basic retroviral vector

Some elements, referred as *cis*-acting viral elements, are necessary to conduct the retrovirus life cycle and must be preserved in the vector design. In the LTRs, they include: the viral promoter and the polyadenylation signal (U3 and R regions, respectively) for generation of the full-length viral transcript, direct repeated regions (R) for transfer during DNA synthesis, and partially inverted repeats (attachment sites corresponding to U3 and U5 terminal sequences) for the integration<sup>119,122</sup> (**Figure 1-9a**). Outside of the LTRs, the PBS and PPT sites are required for DNA synthesis while the packaging ( $\psi$ ) and dimerization signals are necessary for encapsidation of RNA into virions (**Figure 1-9a**)<sup>119,122</sup>. Since the viral genome consists of RNA, it is primordial to avoid introducing a polyadenylation signal (pA) in forward orientation.

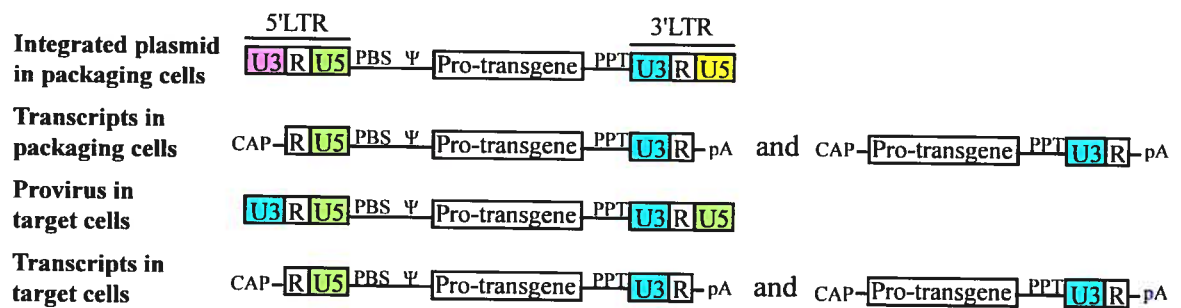
To express multiple proteins, different strategies have been used: alternative splicing, internal promoters combined to viral promoter in the LTR, internal ribosomal entry sites (IRES: allows cap-independent translation<sup>120</sup>) or fusion proteins<sup>119</sup>.



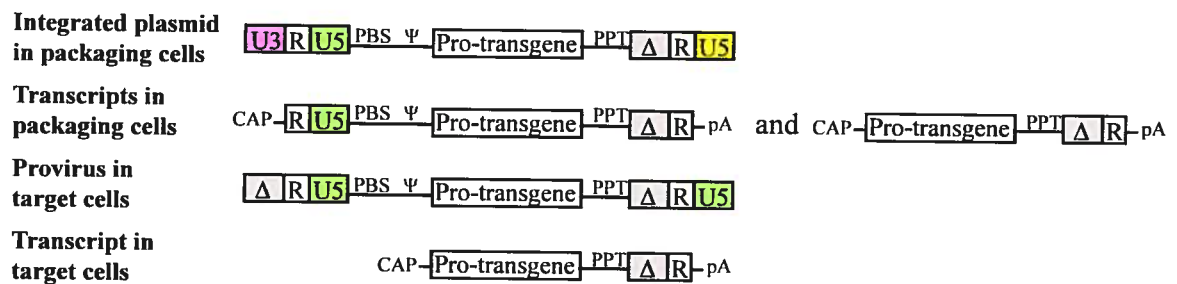
**Figure 1-9 Structures of retroviral vector plasmids, proviruses, and transcripts.**

(a) The vector can carry a transgene, driven optionally by an internal promoter (Pro). In the packaging cells, transcription is directed by the enhancer/promoter in the U3 region of the 5'LTR (U3-5'LTR, pink) and by the internal promoter. In the cytoplasm of the target cells, during reverse transcription, the U5-5'LTR (green) becomes the template for the U5-3'LTR and the U3-3'LTR (blue) is the one for U3-5'LTR. Consequently, the provirus in target cells contains two identical U5 (green) and U3 (blue) regions. LTR, long terminal repeats; PBS, transfer RNA-binding site;  $\Psi$ , packaging signal; PPT, polypurine tract; CAP, RNA 7-methylguanosine cap; pA, polyadenylation signal. (b) For SIN vector, the enhancer/promoter is mutated in the U3-3'LTR ( $\Delta$  gray) while preserving it in the U3-5'LTR (pink). Following reverse transcription, the mutated U3 is found in both LTRs, abolishing transcription. Transgene is expressed from an internal promoter.

**a Replication-incompetent retroviral vector**



**b SIN replication-incompetent retroviral vector**



Following integration, under specific circumstance, the inactivation of the viral promoter can be desirable (e.g. to reduce effects from the viral enhancers/promoters, transcription from the 3' LTR, etc.). This can be achieved by creating a self-inactivating (SIN) vector which relies on the process of reverse transcription (Figure 1-9b). During the first round of viral replication, both U3 regions are derived from the sequence found in the 3' LTR

and both U5 regions from the 5' LTR, respectively<sup>119</sup>. A popular approach to produce a self-inactivating vector is to delete the enhancer/promoter in the U3 region of the 3' LTR (in the plasmid). In the original plasmid construct, a functional enhancer/promoter in the U3 region of the 5' LTR is preserved in order to direct the viral transcription in the packaging cells. This is necessary for virions assembly and release in the culture media. However in the target cells, during the process of reverse transcription, the mutated enhancer/promoter present in the 3' LTR serves as a template for the one in the 5' LTR and as a result, viral transcription is abolished (or decreased) in both LTRs of the provirus. In this situation, expression of transgene(s) is achieved by the use of internal promoters. Other considerations regarding the design of viral constructs will be discussed in **Appendix II**.

#### **1.8.11.2 Embryonic stem cell viral vectors**

Retroviral expression occurs in limited or broad variety of cell types, depending on positive and negative interactions between cellular proteins and the viral enhancer/promoter region (U3 region) or the 5' untranslated leader sequence<sup>123</sup> (**Figure 1-8**). Epigenetic modifiers are also involved in the regulation of viral expression<sup>124,125</sup>. ESCs are refractory to the expression of Mo-MLV and associated vectors, therefore, an alternative retrovirus was generated: Murine embryonic stem cell virus (MESV)<sup>123</sup>. MESV combines changes in two elements that prevent Mo-MLV expression in ESCs<sup>123,126</sup>. The first change was to use the LTR of PCMV (PCC4-cell-passaged myeloproliferative sarcoma virus) which presents seven point mutations in its enhancer region compare to Mo-MLV, further refined to a single essential point mutation required for expression in ESCs<sup>126</sup>. Secondly, the 5' untranslated region (UTR) leader sequence was exchanged for the one of another virus (del-587rev virus) to delete an inhibitory region present in Mo-MLV<sup>123</sup>. Soon after, the MESV served as a template to create the Murine stem cell virus (MSCV) featuring an extended packaging signal, all replication genes replaced by a neomycin gene driven by an internal promoter (*Pgk-1*) downstream of a multiple cloning site to subclone genes of interest<sup>127</sup>.

### **1.8.12 Properties of the packaging cell line**

#### **1.8.12.1 Principles of a packaging cell line**

The role of the packaging cell line is to assist the production of virions by providing the complementary replicative functions that were deleted from the viral genome. For many applications, the engineered retroviruses are replication-defective and precautions are taken in

the packaging cell line to preserve this state and to prevent the transmission of the replicative functions. To reduce the risk of generating replication-competent retroviruses, which is thought to occur by homologous recombination, the packaging functions are integrated separately in the genome of the packaging cells (e.g. *gag-pol* and *env* separated on two plasmids) and depleted as much as possible of other retrovirus-related sequences (e.g. packaging signal, LTR, etc.)<sup>128,129</sup>. Although the principles are simple, not all cell line can become an efficient packaging cell line. Some cell line produces factors that interfere with the viral cell cycle<sup>122</sup>. Some proteins like APOBEC members, evolutionary conserved in vertebrates<sup>130</sup>, are possibly among these factors. Human APOBEC3G is a cytosine-deaminase protein that is incorporated inside HIV-1 virions, causing detrimental hypermutations during the process of reverse transcription<sup>130</sup>. HIV-1 virions deficient in Vif (virion infectivity factor) do not propagate in APOBEC3G expressing cells because one role of Vif is to target APOBEC3G for proteasome degradation (thus reducing its incorporation inside virions)<sup>130</sup>. Additional mechanisms governing the viral restriction by APOBEC family members are suspected since inhibition is observed with retroviruses, other viruses and retrotransposons<sup>131</sup>.

### 1.8.12.2 Tropism and pseudotyping

The host range (tropism) of a retrovirus depends on the envelop proteins (encoded by *env*) which recognize specific receptors at the plasma membrane surface of target cells. Using diverse packaging cell lines, it is possible to substitute the *env* coding domain corresponding to the vector viral type by the *env* corresponding to another retrovirus, a practice called pseudotyping. This strategy allows to change the tropism and/or to increase the stability of the envelop proteins (e.g., VSV-G pseudotyping), property that can be advantageous for experimentations.

### 1.8.13 Transduceable cells

Even if murine retroviruses can enter in target cells, they require cell division in order to access the nucleus, thus restricting the range of transduceable cells<sup>119</sup>. Lentiviruses (for example HIV-1) can infect both dividing and non-dividing cells because the PIC can actively go through the nuclear pore by a poorly understood process<sup>121</sup>. Other retroviruses present an intermediate ability to cross the nuclear membrane<sup>132</sup>. No matter the strategy used by retroviruses to reach the nucleus, it is thought that both viral and cellular proteins are necessary for the process<sup>121</sup>. In some case like HIV-1, it seems to involve additional non-protein elements like the central PPT (viral DNA sequence) and a cellular tRNA<sup>121</sup>.

Additionally, various mechanisms restrict the viral cell cycle, which are probably different according to the target cell involved. For example, the cellular TRIM5 $\alpha$  protein inhibits virus replication by acting at the level of uncoating<sup>120</sup>. In fact, hundreds of cellular genes are suspected to be modulators of retroviral infection<sup>120</sup> in addition to host-encoded miRNAs<sup>133</sup>.

#### **1.8.14 Helper viruses, satellite viruses, and satellite RNA**

Packaging cell lines are living factories for designated retroviruses, but at low frequency, they also produce other undesirable entities. These products can be formed by multiple ways: recombination or read-through transcript at the cellular level; and at the viral level, recombination during the reverse transcription process involving foreign viral or cellular RNA co-encapsidated with the viral genomic RNA, etc.<sup>134-137</sup>.

One of these undesired products is a helper virus. The helper virus is replication-competent and allows the encapsidation of replication-defective retroviruses by providing the expression of the missing replication functions. This type of virus is rarely detected from up-to-date packaging cell lines because the packaging functions are split and retroviral sequences are avoided as much as possible. Nevertheless, different testing methods exist for helper viruses. These are based on the mobilization and scoring of a replication-defective retrovirus (e.g. expressing a selectable marker gene, a fluorescent gene, or the bacterial  $\beta$ -galactosidase gene, etc.). Detection of helper viruses is routinely performed in many laboratories.

Satellite virus and satellite RNA are replication-defective-like retroviruses, with the distinction that the former encodes the *env* proteins while the latter does not. As they can be propagated in combination with a helper virus, they can also be propagated by a packaging cell line. Testing for such viral products is not obvious; particularly for satellite RNAs since they do not contain PCR-detectable replication functions. They stay imperceptible unless specific testing is designed for each of them. Unfortunately, each of them is a long undefined list. There are situations where an abnormal phenotype observed in target cells is suspected to be caused only partially by the retrovirus of interest or totally by something else. The following, non exclusive, doubtful situations are examples: the abnormal phenotype is observed at low frequency or the provirus of interest is not present or constantly rearranged or not expressed. Again, a mobilization experiment can be performed in the target cells, hoping that they will behave as suitable packaging cell lines when transduced with packaging functions. The goal is to verify if “the abnormal phenotype” is transmissible to other target

cells. If this is the case, the intact provirus of interest must be found. Otherwise, concerns must be raised. Human packaging cells were generated to replace their murine counterparts, hoping that they contain fewer endogenous retroviral genomes and express fewer viral-like RNAs, preventing or reducing the generation of unwanted viral products<sup>129</sup> (before genomes' sequencing). However, from our work and the one of others, human packaging cell lines are also able to transmit retroviral-like particles (see **Appendix II**).

Most of these products are possibly inoffensive, while others could behave as oncogenic viruses. In addition, they all share the property of integrating in the host genome, risking insertional mutagenesis at low frequency. These possibilities should be evaluated while analyzing results implicating the use of retroviral vectors-packaging cell lines and their potential application for gene therapy. Nevertheless, even with these low probabilities, retroviruses are an invaluable tool for many research areas.

## **1.8.15 Retroviral integration**

### **1.8.15.1 Preferential sites of retroviral integration**

It is thought that viral DNA needs to be anchored to cellular chromatin prior to the integration and that this interaction probably occurs with the help of cellular proteins that differ according to the virus type<sup>120,121</sup>. Suspected cellular mediators are: the lens-epithelium-derived growth factor (LEDGF), emerin (nuclear envelope protein), the barrier-to-autointegration factor (BAF), the lamina-associated polypeptide 2 $\alpha$  (LAP2 $\alpha$ ), and other unknowns<sup>120,121</sup>. In addition, possible chromatin features modulate the accessibility of the PIC to host genomic DNA<sup>138</sup>. For example, centromeric heterochromatin seems a region disfavored for integration<sup>138</sup>.

Integration site preferences vary according to the retrovirus and the target cells studied and do not seem to be very sequence-specific, although some conserved palindromic sequences can sometimes be observed<sup>139</sup>. From genome-wide annotation of retroviral integration sites in human cells, it was observed that preferential integration (not exclusive) was in active transcription units for HIV-1 and near transcription start sites and CpG islands for MLV<sup>138</sup>. Avian sarcoma-leukosis virus (ASLV) shows random integration<sup>138</sup>. From retroviral integration site surveys, using viral chimeric molecules where the HIV-1 coding domains of *integrase* and *gag* were replaced by MLV related sequences, it appears that the integrase plays a dominant role for the preferential integration<sup>139</sup>. The gag-encoded proteins

might also participate in this bias by an unclear mechanism<sup>139</sup>.

#### **1.8.15.2 Molecular mechanism of retroviral integration**

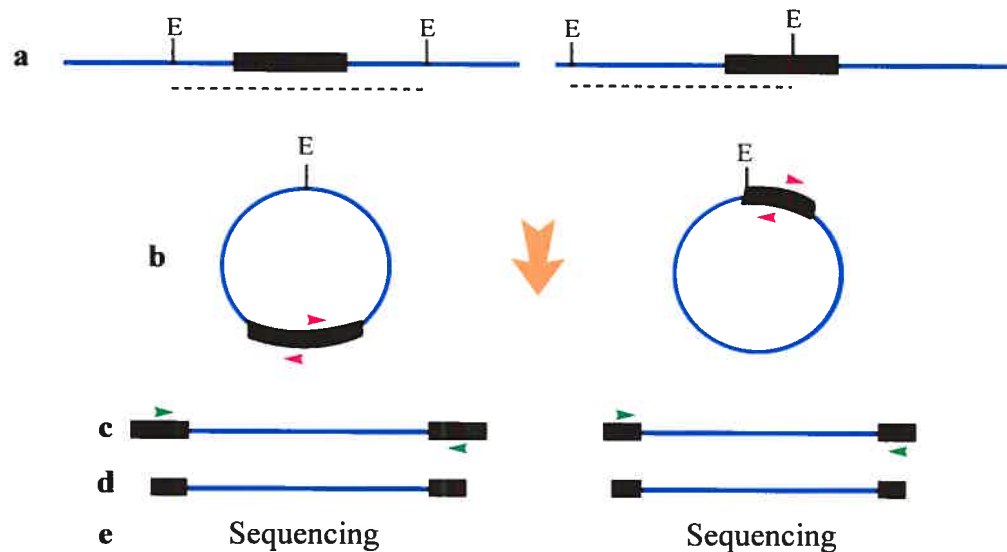
Concerning the integration at the molecular level, the viral integrase cleaves and binds attachment (att) sites in the extremities of LTRs<sup>122</sup>. In addition, the enzyme removes two nucleotides at the 3' ends of viral DNA prior to ligating them 4-5 bases away from 5' cut-ends of host genome, creating two small gaps because of unpairing<sup>119,138</sup>. Cellular DNA repair machinery fills the gaps while creating target sequence repeats on each side of the provirus (4 or 5 bp for MLV or HIV-1, respectively)<sup>122,139</sup>.

#### **1.8.15.3 Determination of retroviral integration sites**





Various methodologies allow the determination of retroviral integration sites. In the case of recombinant retroviruses, some approaches use plasmid rescue. For example, a bacterial origin of replication combined to a selection marker gene like ampicillin can be introduced in the vector backbone<sup>33</sup>. Extracted genomic DNA is linearized by an enzyme that cuts once in the vector but many times in the genome<sup>33</sup>. Following circularization, the plasmid can be recovered in bacteria, containing both a part of the vector and host flanking genomic sequence<sup>33</sup>. Sequencing and mapping allow the determination of retroviral integration sites. Another method is based on the complementation of a truncated kanamycin resistance gene present in a plasmid, by adding a part of a neomycin gene incorporated into genomic DNA by a vector, again isolating a piece of flanking DNA<sup>140</sup>. For the determination of wild-type or recombinant retroviruses integration sites, many PCR techniques are available<sup>141</sup>. One of them is the inverse-PCR (I-PCR) and since it is the methodology used in our experiments, a description can be found in the following figure (**Figure 1-10**).

**Figure 1-10 I-PCR allows the determination of retroviral integration sites.**

(a) The genomic DNA (blue) containing a provirus is digested with an enzyme that either cut outside of the provirus or once in the provirus. (b) The cleaved DNA is circularized in conditions that improve intramolecular ligation (low DNA concentration, large volume of ligation)<sup>141</sup>. (c) A set of PCR primers is designed to recognize known vector sequences and are directed outward in order to amplify flanking DNA. (d) A second round of PCR, with nested primers improves the specificity of the product recovered. (e) The PCR product can be sent directly for sequencing following purification or subcloned first in a plasmid.



**Legend**

	Viral vector		PCR primer
	Enzyme restriction site		Nested PCR primer

## AIM OF THE THESIS

The primary objective of this thesis is to identify stem cell fate regulators using a Cre-*loxP* recombination system delivered by retroviral gene transfer in mouse ESCs.

The first part of the introduction described the context and the reasons that drove the design of a screen based on chromosomal deletions, with the novelty of relying on compatible retroviruses to introduce the *loxP* sites in the genome (see **Table I** for a summary). The second part of the introduction presented the advantages of mouse ESCs for both *in vitro* and *in vivo* studies, highlighting the relevance of this experimental model for functional genomics. The last part of the introduction reviewed the principles behind retroviral gene transfer.

Appendix II and Chapter 2 describe the retroviral system optimization and the validation of chromosomal engineering in ESCs, respectively. Chapter 3 presents the generation of a library of ESC clones containing deletion and its exploitation in genome-wide functional screens. Chapter 4 exposes perspectives and potential applications of the methodology.



## ***Chapter 2* APPLICATION OF A NEW RETROVIRAL SYSTEM TO CREATE CHROMOSOMAL DELETIONS IN ESCs**

Chapter 2 is a published article describing the optimization of a new retroviral-based system to create chromosomal deletions in mouse ESCs. As opposed to other *Cre-loxP* strategies previously employed to create deletions in ESCs, the proposed methodology has the notable advantage of not requiring homologous recombination to deliver *loxP* sites. The technique was studied on eleven genomic loci and several nested deletions were obtained. Three regions altering the capacity of ESC to differentiate *in vitro* were discovered. *In vitro* observations correlated with *in vivo* analyses performed by generating chimeric mice with engineered ESC clones. The work presented in Chapter 2 created the building blocks of Chapter 3 (generation of a library of ESC clones containing deletion and preliminary functional screening). The novel system described in this chapter is in fact a tool that could be used in other interesting cell lines, such as tumorigenic cell lines. Potential applications of this methodology will be exposed in Chapter 4. Mélanie Bilodeau did most of the work regarding Chapter 2, helped by collaborators as mentioned in the following sections: Author contributions and Acknowledgments.

**ARTICLE: A retroviral strategy that efficiently creates  
chromosomal deletions in mammalian cells**

**Mélanie Bilodeau<sup>1</sup>, Simon Girard<sup>1</sup>, Josée Hébert<sup>2,3</sup> and  
Guy Sauvageau<sup>1,2,3\*</sup>**

<sup>1</sup>Laboratory of Molecular Genetics of Stem Cells, Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montréal, Québec, Canada, H2W 1R7. <sup>2</sup>Department of Medicine, Montréal, Québec, Canada, H3C 3J7. <sup>3</sup>Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, H1T 2M2.

Article published in *Nature Methods* 4, 263-8 (2007)

Reprinted by permission from Macmillan Publishers Ltd: NATURE METHODS  
[*Nature Methods* 4, 263-8 (2007)], copyright 2007.

\*Correspondence: Guy Sauvageau, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada, H3C 3J7. Email : XXXXXXXXXX

## 2.1 Author contributions

Mélanie Bilodeau performed all the experiments and the analyses described herein, except for I-PCR and two Southern blots (**Supplementary Figure 2-4g,h**) (Simon Girard), SKY and FISH analyses (Josée Hébert). aCGH experiments and chimeras production were conducted by the services mentioned below (see **Supplementary Methods** section). Mélanie Bilodeau wrote the manuscript, prepared all the figures, and performed the experiments under Guy Sauvageau guidance.

## 2.2 Abstract

Chromosomal deletions, as a genetic tool for functional genomics, remains underexploited in vertebrate stem cells mostly because currently available methods are too labor intensive. To address this, we have developed and validated a set of complementary retroviruses that creates a wide range of nested chromosomal deletions. When applied to mouse embryonic stem cells (ESCs), this retrovirus-based method generated deletions ranging from 6 kb to 23 Mb (average 2.9 Mb), with an efficiency of 64% for drug -selected clones. Importantly, several engineered ESC clones, mostly those with large deletions, showed major alteration in cell fate. In comparison to other methods that have also exploited retroviruses for chromosomal engineering, this modified strategy is more efficient and versatile because it bypasses the need for homologous recombination and thus can be exploited for rapid and extensive functional screens in embryonic and adult stem cells.

## 2.3 Introduction

Capitalizing on the reliability of *Cre/loxP*-based recombination, a group previously reported the generation of nested chromosomal deletions in mouse ESCs by sequentially delivering two *loxP* sequences into the genome, followed by Cre-mediated excision of the chromosomal region between the *loxP* sites. In this approach, the first *loxP* sequence was introduced into a particular locus of choice by homologous recombination using a targeting vector which included a non-functional “split” *Hprt1* cassette<sup>36</sup>. The second *loxP* and complementary *Hprt1* sequences were delivered using retroviral gene transfer<sup>36</sup>. Cre-mediated recombinants were selected in HAT medium following reconstitution of the functional *Hprt1* mini-gene<sup>36</sup>. While this method significantly improved our ability to generate high-resolution sets of nested deletions around a targeted locus, its extension to

several other loci remained labor-intensive and precluded large-scale functional screens in ESCs. Moreover, by its nature, this method was limited to cells permissive to homologous recombination thus excluding most mammalian cells. Here we sought to overcome these limitations and elected to develop a strategy that would strictly rely on the use of replication-defective retroviruses while exploiting the *Cre-loxP* recombination system and reconstitution of a functional neomycin (neo) cassette for selection of recombination events. The first *loxP* sequence is delivered using a vector that we refer to as the anchor virus, and the second by a saturating virus (**Figure 2-1a,b**).

## 2.4 Results

### 2.4.1 Selection of anchor and saturation proviruses

From a series of 10 different retroviral constructs consisting of 5 anchor (no. A1-A5: **Supplementary Figure 2-3**) and 5 saturation viruses (no. S1-S5: **Supplementary Figure 2-3**), we selected viruses A1 and S1 based on the criteria listed in **Supplementary Figure 2-3**. As depicted in **Figure 2-1a**, chromosomal deletions are expected to have occurred in geneticin resistant (G418<sup>R</sup>) clones that have lost both puromycin (puro) and hygromycin (hygro) resistance genes. For a more detailed description of the approach and of the vectors tested, readers are referred to the thesis' **Appendix II**.

Using retroviral preparations adjusted to provide gene transfer to mouse R1 ESCs < 1%, we first confirmed that most clones infected with virus A1 and selected on puromycin (thereafter called primary clones) had a single integrated provirus (data not shown). Eleven randomly selected puromycin resistant (puro<sup>R</sup>) clones were expanded and infected, in 1 to 4 independent experiments (A to D) per each clone, with low titer S1 virus to generate series of hygromycin-resistant (hygro<sup>R</sup>) populations with a complexity of ~20,000 independent secondary clones (**Supplementary Table V**). Following *Cre* electroporation, we observed G418<sup>R</sup> recombinants (thereafter called tertiary clones) for each of the 11 primary clones analyzed with an average frequency of  $2.5 \pm 2.2 \times 10^{-4}$  (details in **Supplementary Table V**). Expected *Cre*-induced rearrangements between the integrated A1 and S1 proviruses were verified by Southern blot analyses for several tertiary clones derived from each of the 11 families (see representative in **Figure 2-1c** and **Supplementary Figure 2-4**). No spontaneous G418 resistance was ever observed in the absence of *Cre* expression. Confirmation of a productive rearrangement leading to the expression of the neomycin gene was obtained for a series of tertiary clones where a single messenger RNA of ~1.0 kb was detected (**Figure**

**2-1d).** Fragments corresponding to the recombination junctions (*Pgk-loxP-neo*) were also PCR-amplified from representative G418<sup>R</sup> tertiary clones and sequenced, confirming the expected breakpoint (n= 5 clones selected in 4 families, not shown).

**Figure 2-1 Cre-induced chromosomal rearrangements in mouse ESCs.**

(a) Representation of the recombination between A1 and S1 proviruses. Following Cre transfection, the coupling of the *Pgk*-ATG in S1 to the neomycin (ATGless neo) gene in A1 allows the selection of recombinants. Deletions are identified by the concomitant losses of puromycin (*puro*) and hygromycin (*hygro*) resistance genes. Symbols are detailed in **Supplementary Figure 2-3**. (b) Cartoon of nested deletions sharing the same endpoint (virus A1). (c) Southern blot analysis of DNA isolated from selected clones documents the integrity of provirus A1 (3.4 kb) and S1 (2.9 kb) and their successful recombination (A1-S1, 3.0 kb). Clonal diversity is shown in the 2 bottom panels. 1<sup>o</sup>, 2<sup>o</sup> and 3<sup>o</sup>: primary, secondary and tertiary clones, respectively. S, sensitive; R, resistant. Unmodified blots are presented in **Supplementary Figure 2-4**. (d) Neomycin resistance gene expression in presence (lanes 3-15 and 17) or absence (lanes 2 and 16) of Cre treatment. Note: the 1-kb transcript is indicated by an asterisk in **a**. (e) CGHAnalyzer<sup>142</sup> representation (middle) linked to the chromosomal localizations of confirmed deletions (Ensembl Karyoview<sup>143</sup>, left) for selected clones in family 9. Red lines represent nested deletions for the indicated clones. Spectral karyotyping for selected clones (right). Scale bars: 10  $\mu$ m. ▲, A1 anchor site; S1 proviral integration site of selected deletions confirmed by I-PCR (▲) or aCGH (▲). (f) Size distribution of confirmed deletions in indicated families. Dots and strokes represent independent deletion and average deletion size per family, respectively.



## 2.4.2 Evaluation of chromosomal deletions

Among the different chromosomal rearrangements obtained, deletions were screened by testing for the concomitant loss of puromycin ( $\text{puro}^{\text{S}}$ ) and hygromycin ( $\text{hygro}^{\text{S}}$ ) (see **Supplementary Table V** for frequencies of  $\text{puro}^{\text{S}}$  and  $\text{hygro}^{\text{S}}$  clones) which occurred in 9 of the 11 families. Inverse-PCR (I-PCR), array-based comparative genomic hybridization (aCGH), and spectral karyotyping (SKY) were employed to confirm deletions in several tertiary clones from 8 of these 9 different families and to assess the genomic integrity of the altered ESCs **Figure 2-1e**, **Table III** and **Supplementary Figure 2-5**). The size distribution of deleted DNA fragments varied according to the family studied (**Figure 2-1f** and **Table III**), ranging between 6 kb to 23 Mb, with an average of  $2.9 \pm 5.2$  Mb. Interestingly, as noticed with tertiary clones derived from family 9, deletion sizes did not follow a normal distribution since they either ranged in the scale of kilobase pairs (6 to 317 kb,  $n=8$  independent deletions) or megabase pairs (4.2 to 5.0 Mb,  $n=3$  independent deletions) (**Figure 2-1f** and **Table III**). On average, I-PCR-confirmed deletions included  $21 \pm 46$  genes,  $15 \pm 28$  CpG islands,  $1689 \pm 3283$  spliced expressed sequence tags (ESTs) and  $0 \pm 1$  microRNA (**Table III** and data not shown). The frequency of clones with deletion that were free of other rearrangement was 0.71 (**Table III**) thus indicating that the frequency of valuable deletions in a pool of  $\text{G418}^{\text{R}}$  colonies was 0.15 ( $0.26_{\text{frequency of } \text{puro}^{\text{S}} \text{ clones}} \times 0.9_{\text{frequency of } \text{puro}^{\text{S}} \text{ hygro}^{\text{S}} \text{ clones}} \times 0.9_{\text{frequency of deletions confirmed out of the } \text{puro}^{\text{S}} \text{ hygro}^{\text{S}} \text{ clones with independent rearrangements}} \times 0.71_{\text{frequency of deletions without other rearrangement, confirmed by aCGH/SKY}}$ ). This frequency of deletions is probably an underestimation since we excluded from the analysis a subgroup of clones that showed ambiguous sensitivity to puromycin or hygromycin ( $\text{puro}^{\text{S+R}} \text{ hygro}^{\text{S+R}}$ ). DNA analyses suggested that these  $\text{puro}^{\text{S+R}} \text{ hygro}^{\text{S+R}}$  cells represented at best a minor fraction of our  $\text{G418}^{\text{R}}$  colonies since, most of the time, we could not detect a signal to these genes in the selected clones (**Supplementary Figure 2-4**).

It was thus possible to engineer large chromosomal deletions for most of the regions tested in our study. Two anchor sites in primary clones no.12 and no.15, respectively located on chromosome X and 11, were not permissive for deletion. This might suggest the proximity of a haplolethal determinant for ESCs or the presence of physical constraints, such as chromatin structure, preventing recombination between *loxP* sites oriented for deletions.

**Table III • Characteristics of independent deletions confirmed by I-PCR and aCGH**

Tertiary clone id <sup>a</sup>	Data source	Chromosome	Start coordinate	End coordinate	Size of deletions (kb)	No. of Refseq genes	No. of spliced ESTs	No. of CpG islands	aCGH confirmation	Genomic anomaly <sup>b</sup>
1-03	I-PCR	14	22165099	23710534	1545	17	1080	14	(+)	(-)
1-13	I-PCR	14	22165099	44937742	22773	206	14400	126	(+)	(-)
4-2	I-PCR	2	167486681	168900222	1414	19	1819	30	n.d.	n.d.
6-36	I-PCR	17	26955839	27622141	666	13	1272	14	(+)	(+) <sup>d,e</sup>
7-30	I-PCR	16	35918443	36011960	94	3	155	1	(+) <sup>e</sup>	(-)
9-31	I-PCR	18	57155985	57162362	6	0	0	0	n.d.	n.d.
9-107	I-PCR	18	57155985	57166502	10	0	0	0	n.d.	n.d.
9-71	I-PCR	18	57155985	57174937	19	0	0	0	n.d.	n.d.
9-17	I-PCR	18	57155985	57174941	19	0	0	0	n.d.	(-) <sup>d</sup>
9-68	I-PCR	18	57155985	57175174	19	0	0	0	n.d.	n.d.
9-29	I-PCR	18	57155985	57177486	22	0	0	0	n.d.	n.d.
9-35	I-PCR	18	57155985	57179077	23	0	0	0	(+) <sup>e</sup>	(-)
9-90	I-PCR	18	57155985	57473132	317	2	114	4	(+) <sup>e</sup>	(+) <sup>f</sup>
9-104	I-PCR	18	57155985	61338307	4182	20	3289	18	n.d.	(+) <sup>g</sup>
9-37	I-PCR	18	57155985	61468765	4313	21	3419	20	(+)	(-) <sup>d</sup>
9-18	I-PCR	18	57155985	62204954	5049	32	4726	27	(+)	(-)
10-18	I-PCR	16	59749084	65165857	5417	12	481	10	(+)	(-)
10-21	I-PCR	16	57307345	65165857	7858	51	1371	22	(+)	(+) <sup>h</sup>
13-34	aCGH	4	78266064	82222600	3956	14	792	10	(+)	(-)
14-16	I-PCR	2	156503387	157071542	568	9	869	8	(+) <sup>e</sup>	(-)
average					2914	21	1689	15		(-) frequency: 0.71
SD					5244	46	3283	28		

Mapping and deletion analyses were done using the UCSC Genome Browser (<http://genome.ucsc.edu/>, NCBI mouse Build 33).

<sup>a</sup>Tertiary clones are labeled according to their family number (same integration of virus A1), followed by a specific id number. If more than one clone presented a redundant rearrangement within the same group infected with virus S1, only one is reported for clarity. <sup>b</sup>Anomaly that was not present in the primary clone from which the tertiary clone was derived, as determined by aCGH or SKY. (-), no anomaly; (+), additional anomaly. <sup>c</sup>The deletion is not observed, in agreement with the resolution of aCGH. <sup>d</sup>Normal except for the loss of chromosome Y. <sup>e</sup>Amplification of chromosome 1. <sup>f</sup>Amplification of chromosome 8. <sup>g</sup>Many chromosomes were lost according to SKY. <sup>h</sup>Amplification on chromosome 14. Id, identification; kb, kilobase pairs; no., number; aCGH, array-based comparative genomic hybridization; SKY, spectral karyotyping; I-PCR, inverse-PCR; n.d., not determined.

### 2.4.3 Interchromosomal recombination events

Interchromosomal events are expected to give single loss of puromycin or hygromycin, or conservation of both resistance genes, but not their concomitant losses<sup>144</sup>. In the course of the aCGH and SKY analyses, we unexpectedly observed 2 inter-chromosomal rearrangements from a group of 22 puro<sup>S</sup> hygro<sup>S</sup> clones (believed to represent deletions), one of which is a confirmed translocation (clone 14-27: t(2;16)) (**Supplementary Figure 2-6**). A possibility that could account for this phenomenon is the loss of a chromosome (for example the loss



of the recombined chromosome bearing the puromycin and hygromycin resistance genes originating from recombination in  $G_1$  or in  $G_2$  with Z-segregation<sup>145</sup>), accompanied by the duplication of the homologous chromosome<sup>114</sup> (**Supplementary Figure 2-6**).

From a subgroup of 190 tertiary clones selected for further analyses, eight showed sensitivity to either puromycin or hygromycin and represented seven independent rearrangements as assessed by clonal analysis of proviral integration. Of these seven clones, two contained productive (that is, confirmed by I-PCR analysis) transchromosomal rearrangements (that is., 2-03 and 14-32, **Supplementary Figure 2-6**). Interestingly, the frequency of single loss of puromycin varied according to the family, showing highest values for family 4 and 14 where the anchor virus was located close to the telomeric ends of the chromosome 2 (**Supplementary Table V**).

Together, these results suggest that recombination events in *trans* occur at higher frequency for particular loci, but also at low frequency in the puro<sup>S</sup> hygro<sup>S</sup> clones, further highlighting the importance of complementary analyses (e.g., SKY) for these types of studies.

#### **2.4.4 *In vitro* and *in vivo* differentiation of recombined clones**

To gain insights into the potential of our approach to generate clones that can be utilized in a functional screen *in vitro*, 43 tertiary clones from 9 families were selected to cover a wide range of deletion sizes (from 6 kb to 23 Mb) and differentiated into embryonic bodies (EBs) for identification of phenotypic anomalies (**Supplementary Table VI** and selected examples in **Figure 2-2a,b**). One third (3/9) of the families studied contained clones which showed major differentiation anomalies, representing 11% (5/43) of our sample size (**Supplementary Table VI**). Clones in families' no. 1 and no. 9 are particularly interesting since they cover a wide range of deletions (**Supplementary Table VI**) and only clones with larger deletions show phenotypic anomalies. For example, clone 1-03 included a 1.5 Mb deletion and differentiated normally while clone 1-13, with a ~23 Mb deletion, failed to differentiate *in vitro*. The correlation between deletion size and phenotype is more striking for family 9 where all 8 clones having less than a 318 kb deletion show normal *in vitro* differentiation while 3 of the 3 clones with greater than 4.1 Mb deletions failed to differentiate. Reinforced by the observation that most of these clones lacked additional DNA rearrangement as assessed by aCGH and SKY analyses (**Table III**), these results argue against other genetic events being responsible for these phenotypes. Additional evidence to support this argument

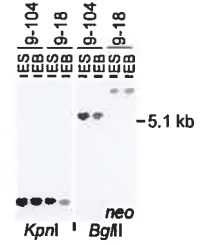
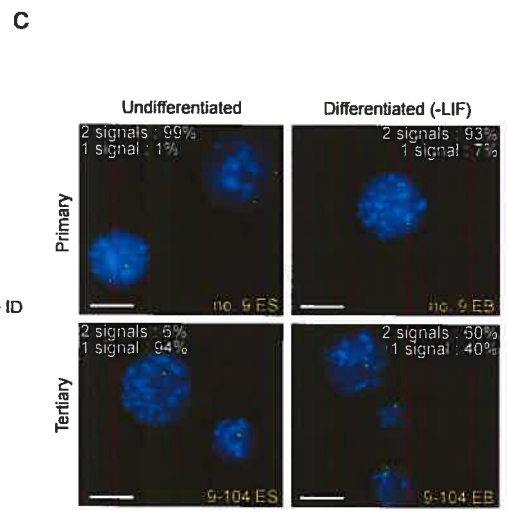
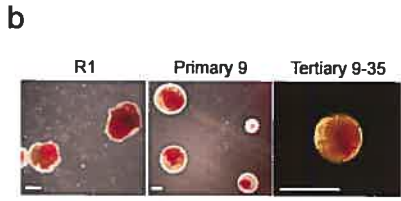
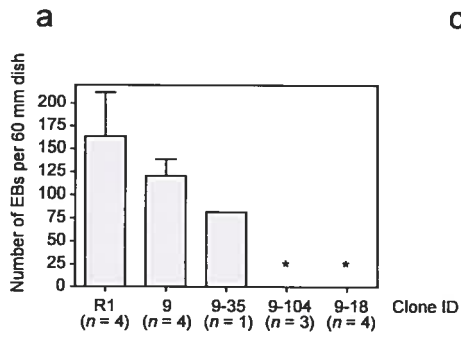
includes the interesting observation that most of the clones with differentiation anomalies (e.g., 9-104) show EB formation at low frequency. In this clone 9-104, we conducted FISH analysis using a BAC probe corresponding to the deleted region of chromosome 18. This analysis revealed that 94% and 6% of undifferentiated 9-104 cells showed one versus two signals, respectively. However, among the rare differentiated cells derived from this clone, 60% displayed 2 signals (see **Figure 2-2c** for example). Clonal analysis of DNA extracted from these rare differentiated cells confirmed their origin from clone 9-104 (see Southern blot in **Figure 2-2c**), ruling out possible contaminants as an explanation for this complementation. This low frequency of revertants is consistent with the chromosomal instability observed in ESCs<sup>115</sup> and confirmed by the extensive aCGH and SKY analyses reported herein. Most importantly, this observation documents the low frequency of spontaneous revertant thus strengthening the argument that differentiation is dependent on the presence of the deleted fragment.

Consistent with the *in vitro* results, the ESCs from tertiary clone 9-35 (23 kb deletion, normal phenotype *in vitro*) contributed to the generation of chimeric 14.5 dpc fetuses and to viable newborn mice with an overall proportion of 75% and 30%, respectively (3/4 at E14.5 and 3/10 at birth and adulthood; **Table IV**). Sixty-seven pups, derived from crosses between two chimeric males (~75% and ~40% coat color chimerism) and C57BL/6 females, were genotyped for the transmission of the engineered allele. Germ-line transmission of these ESCs was documented by coat color analysis in 3 pups although the perpetuation of the deleted chromosome was not documented in any of the 67 pups. Of interest, embryos injected with ESCs from clone 9-18 (limited potential to differentiate *in vitro*, **Figure 2-2a**) showed a high mortality rate at 14.5 dpc (46% viable, **Table IV**) with undetectable ES-derived contribution (DNA analysis and coat color) evaluated in six 14.5 dpc fetuses and in 31 adults (**Table IV**, **Figure 2-2d-e** and data not shown). Thus, within the limit of these analyses, there is a good correlation between the EB formation competency *in vitro* and contribution to chimerism *in vivo* for our deleted ESC clones. These results also document that a subgroup of ESCs which have undergone our procedures remain competent for the creation of chimeras thus paving the way to use this strategy for *in vivo* studies.

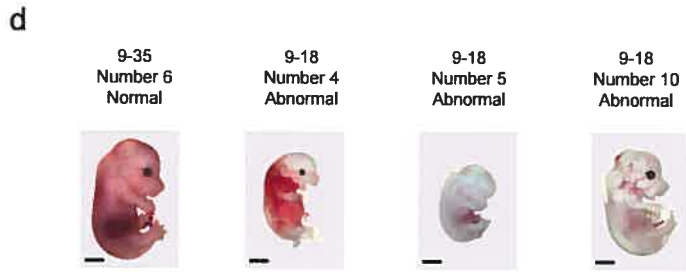
**Figure 2-2 *In vitro* and *in vivo* differentiation of ESC clones with deletions.**

(a) Day 7-8 embryoid body (EB) formation (mean  $\pm$  s.e.m.) of parental R1 ESCs, primary clone no.9 and selected tertiary clones. n=1-4 experiments as indicated. \*Low frequency of EB formation is observed with 10-100X higher seeding density. (b) Examples of day 7-8 EBs generated for selected clones. Scale bar: 250  $\mu$ m. (c) Fluorescence in situ hybridization (FISH) using BAC RP23-109P21 as a probe. Note the relative signal distribution in undifferentiated (left panels) and in differentiated (right panels) cells. Green signals were enhanced using Adobe Photoshop CS imaging tool to replicate visualization on LCD monitor. Scale bar: 10  $\mu$ m. Recombination and clonal analyses (*Kpn*I and *Bgl*III restriction enzymes, respectively) of DNA extracted from indicated ESCs (ES) and EBs. (d) Pictures of 14.5 dpc fetuses. Scale bar: 2 mm. (e) Southern blot analysis (*Bgl*III restriction digests; neomycin (neo) probe) or PCR studies of genomic DNA extracted from the indicated cells. These included ESC clones 9, 9-35, 9-18, R1 control and 14.5 dpc chimeric fetal livers (upper panel), heads (lower panel) or hematopoietic colonies derived from fetal liver cells (numbers shown between upper and lower panels). Note the absence of contribution for clone 9-18 to the chimeric fetuses. For “c” and “e”, unmodified blot are presented in **Supplementary Figure 2-7**.

*In vitro*



*In vivo*



**e**

	Primary 9				Tertiary 9-35				Tertiary 9-18						
	% in R1		chimeras		% in R1		chimeras		% in R1		chimeras				
ES cells and fetal livers	0	5	100	100	0	5	100	100	0	5	100	100			
Hematopoietic colonies	<i>neo</i> <sup>+</sup> (PCR)		2	0	2	2	0		0	2	1	0		0	0
	total analyzed		2	4	4	2	3		3	2	2	3		3	4
Heads	[Southern blot images]														

**Table IV • Chimera analysis**

Reimplantation	Fetus 14.5 dpc				Mice	
	No. embryos implanted	No. fetus observed/expected	Proportion normal (%)	Proportion of chimeric fetus <sup>a</sup>	No. neonates observed/expected	Proportion of chimeric mice <sup>b</sup>
no. 9	173	19 / 32	89	10 / 18	56 / 141	7 / 52
no. 9-35	36	7 / 8	100	3 / 4	11 / 28	3 / 10
no. 9-18	67	15 / 16	46	0 / 6	31 / 51	0 / 31

<sup>a</sup>According to Southern blot or PCR analysis of DNA or eye pigmentation analysis. <sup>b</sup>According to coat color analysis. No., number; dpc, days postcoitum; %, percentage.

## 2.4.5 Discussion and Conclusions

In this study, we have developed a system, entirely based on retroviruses, to engineer chromosomal deletions in the genome of mammalian cells. Since this technology relies on a pair of complementary retroviruses to deliver *loxP* sites, it bypasses the laborious step of homologous recombination, considerably accelerating the creation of large deletions that can easily be mapped through I-PCR. Ten different viruses were tested before an effective pair was identified. We have validated the functionality of this system through the analysis of 11 different families including several independent clones and provided evidence for an efficiency of deletions nearing 64% for clones that are selected based on sensitivity to puromycin and hygromycin. We also show that the average deletion in these clones is ~2.9 Mb in size thus suggesting that a complexity of  $10^3$  primary clones could cover a haploid genome in the mouse, providing that our anchor virus shows no preference for integration. A better estimate of the number of primary clones necessary to cover the mouse genome will be available when a larger collection of deletions is mapped around several anchor sites, allowing to take into consideration the preferential retroviral integration sites, the presence of haploinsufficient regions detrimental to ESCs, and finally the physical constraints such as chromatin organization that might affect the efficiency of Cre-*loxP* recombination.

The method described in our paper complements other functional genomics strategies applicable in mammalian cells<sup>9,14,20,35,146,147</sup>. Although certain limitations of the proposed method remain to be determined (integration of retrovirus in gene-poor regions in ESCs, epigenetic changes in long-term cultured ESCs, ability to produce homozygous deletions using high G418 concentrations<sup>117</sup>), our procedure should easily be amenable to high-

throughput screens. We suspect that our complementary viruses and deletion strategy will be particularly useful for functional screens that involve cells which show poor frequencies in homologous recombination (e.g., human ESCs) and to identify fragments of DNA involved in tumor progression.

## 2.5 Methods

**Retroviral constructs.** We generated the A1 retroviral construct (plasmid no.1647) by inserting both a *loxP*-ATG-less-*neo*<sup>R</sup> PCR cassette from pPNT<sup>148</sup> and a SV40 early mRNA polyadenylation signal (pA) fragment from pDsRed2-N1 (Clontech) into pRETRO-SUPER<sup>149</sup> linearized by *XhoI*-*EcoRI* (blunt) as indicated (**Supplementary Figure 2-3**). For S1 virus (plasmid no. 1643), a *Pgk-kozac-ATG-LoxP* fragment was placed in reverse orientation to a Hygro cassette in *HpaI*-linearized MSCV vector in which the *neo*<sup>R</sup> gene was removed.

**Inverse PCR, sequencing and mapping.** 0.5 ug of genomic DNA was linearized with 20 U of a restriction enzyme, in a total volume of 20 ul. Either *EcoRI* or *StuI* (Invitrogen) was used for the primary clones and *BstEII* (Invitrogen) or the double digest *BglII*-*BamHI* (Invitrogen) for the tertiary clones. After ethanol precipitation in presence of 0.5 ul of linear polyacrylamid carrier<sup>150</sup>, DNA was resuspended in 24 ul of HPCL grade water (J.T. Baker). 4ul of this linear DNA was put aside to be used as the PCR negative control and 20 ul was circularized using the T4 DNA ligase (Invitrogen, 4 U in a total volume of 45 ul, incubated overnight at 16°C). The ligated product was precipitated with ethanol (and carrier) and resuspended again in 24ul of HPCL grade water. The first PCR round was carried out using 4ul of ligated DNA, 1X PC2 reaction buffer (AbPeptides), 0.25 mM of each dNTPs (Invitrogen), 2 mM MgCl<sub>2</sub>, 20 pmol of forward and reverse primers (BioCorp), 5U of KlenTaq LA-16 DNA polymerase<sup>151</sup> (Mix 15:1 of KlenTaq1 from AbPeptides and Pfu from Stratagene), in a total volume of 50 ul. PCR was performed in a Perkin Elmer Termocycler using the following parameters: 2 min at 94°C for one cycle, 20 sec at 94°C\_30 sec at 63°C\_15 min at 68°C for 10 cycles, 20 sec at 94°C\_30 sec at 63°C\_15 min at 68°C with a 20 sec auto-extension for 20 cycles and finally, an extension of 30 min at 68°C. The PCR product was diluted 1:10 000 to 1:50 000 and used in a second PCR round with nested primers. The same settings were employed but with the annealing temperature at 65°C. For both primary and tertiary clones, the PCR primers for the first PCR round were (longNEO-F2) 5'-tggccgcttttctggattcatcgactgtgg-3' and (long-NEO-R) 5'-aagcggccggagaacctgctgcaatc-3'. The second round of PCR was done with primers (longPgk-R) 5'-ggcgcctaccgggtggatgtggaatgtgtg-3' and (long-NEO-R) for primary clones, and with (longPack-R) 5'-ggcggatggaggaagaggaggcggagg-3' and (longPgk-

R) for tertiary clones, respectively. PCR products were separated on 0.8% agarose gel and purified with the QIAEX II Gel Extraction Kit (Qiagen). Fragments were subcloned in pBluescript (Stratagene, T3 and T7 sequencing primers) or sequenced directly using one of the nested primers. Samples were processed using a dideoxy chain termination method and the 3730XL DNA Analyzer system (ABI), at the Genome Quebec Innovation Center (McGill University, Montreal, CA). Mapping and deletion analyses were done using the UCSC Genome Browser (<http://genome.ucsc.edu/>, NCBI mouse Build 33)<sup>152,153</sup>. Ensembl Genome Browser was used for schematic representations of deletions (<http://www.ensembl.org/index.html>, v.32-Jul 2005)<sup>143</sup>.

**Additional methods.** Description of pCX-*Cre* plasmid, cell culture, viral production and transfection, RNA and DNA analyses, aCGH, spectral karyotyping and FISH, chimeras production, equipment and settings is available in **Supplementary Methods**.

**Accession codes.** Gene Expression Omnibus (GEO): GSE6706.

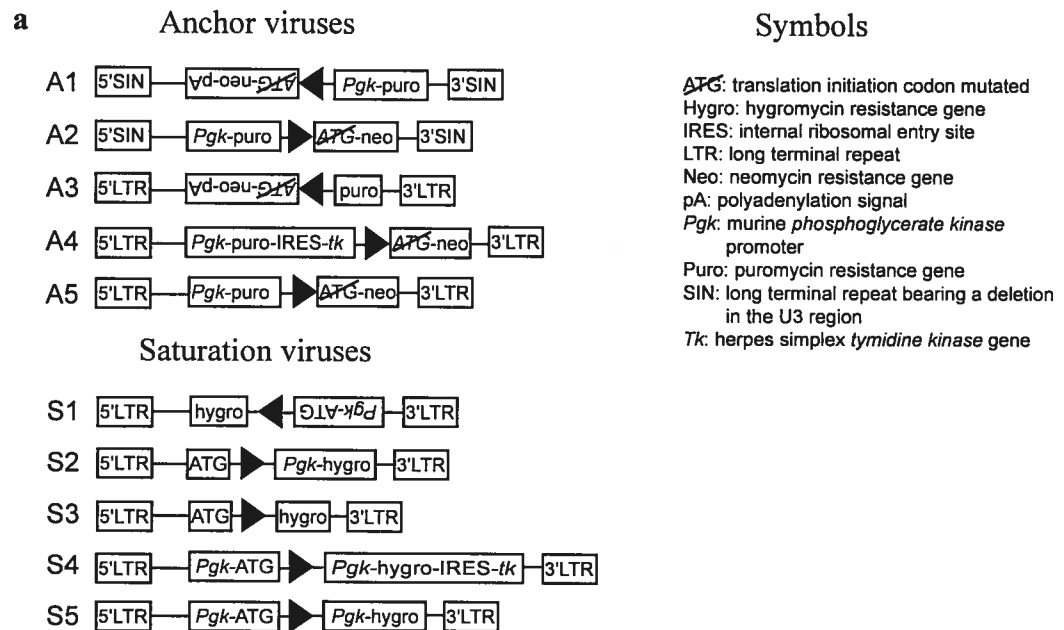
## 2.6 Acknowledgments

We thank Pierre Chartrand, Marc Therrien and colleagues for critically reading the manuscript; Véronique Paradis, Sébastien Harton and Éric Milot from the transgenic facility of IRIC; John Cowell, Michael R. Rossi and Devin McQuaid for the aCGH service (Roswell Park Cancer Institute, Buffalo); Jean-Philippe Laverdure from the bioinformatics service of IRIC; Christian Charbonneau from the imaging service of IRIC; Nadine Fradet, Mélanie Fréchette, Amélie Fredette, Edlie St-Hilaire, Tara MacRae, Claude Rondeau and Pauline Lussier for technical assistance; Andras Nagy for providing the R1 ESCs and the pCX-EYFP construct (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto); Allan Bradley for the pOG231 construct (Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton); Robert G. Hawley for the MSCV vectors (The George Washington University Medical Center); Maarten van Lohuizen for the pRETRO-SUPER construct (The Netherlands Cancer Institute) and Rudolf Jaenisch for the DR-4 mouse strain (Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology). This work was mostly supported by a grant from Génome Québec to Guy Sauvageau and in part by a grant from the Réseau de Recherche en Transgénèse du Québec to Josée Hébert. Mélanie Bilodeau is a recipient of a Canadian Institutes of Health Research (CIHR) studentship and Guy Sauvageau is a recipient of a Canada Research Chair in molecular genetics of stem cells and a scholar of the Leukemia Lymphoma Society of America.

## 2.7 Supplementary Figures

**Figure 2-3 Generation of retroviral vectors.**

Generation of retroviral vectors. (a) Anchor and saturation viruses carry a *loxP* site (▶) together with the selector genes as indicated. (b) Table showing the tests conducted to select the best viruses for our recombination system. Only virus A1, S1, and S2 were suitable for our system. See **Appendix II** for more details about characterization.



**b**

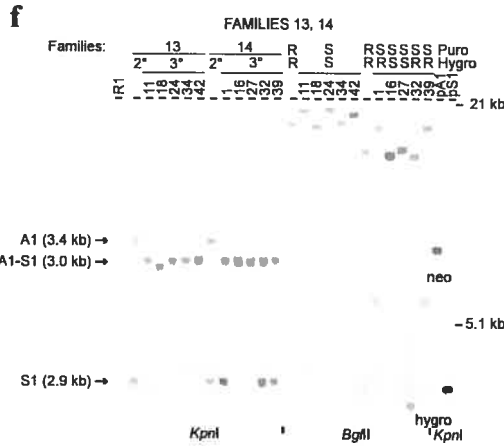
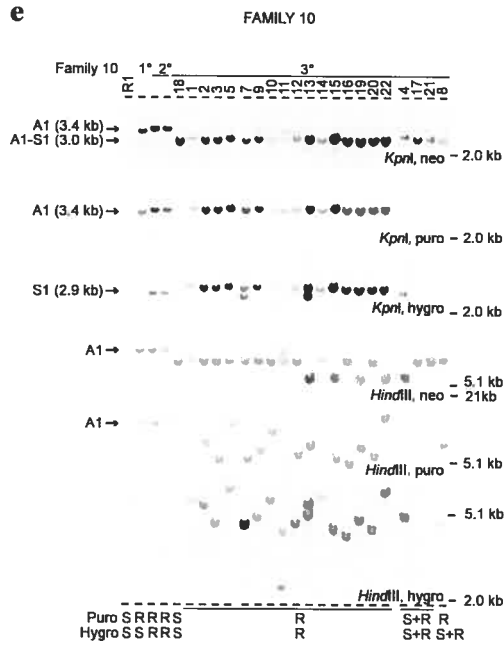
Table of the tests conducted in ES cells for the selection of retroviruses

Viruses	Packaging cell lines	Frequency of rearranged proviruses <sup>a</sup>	Selected yes/no	Undesired G418 <sup>R</sup> frequency <sup>c</sup>	Selected yes/no
<b>Anchors</b>					
A1	GP+E-86	0.1	yes→	<0.0001→	yes
A2	GP+E-86	0.2	yes→	0.01→	no
A3	GP+E-86	1.0	no		
A4	293GPG	0.7	no		
A5	GP+E-86	0.1	yes→	1.0→	no
<b>Saturation</b>					
S1	GP+E-86	<0.1	yes		
S2	GP+E-86	<0.1	yes		
S3	GP+E-86	n.d. <sup>b</sup>	no		
S4	293GPG	≥0.9	no		
S5	293GPG	≥0.9	no		

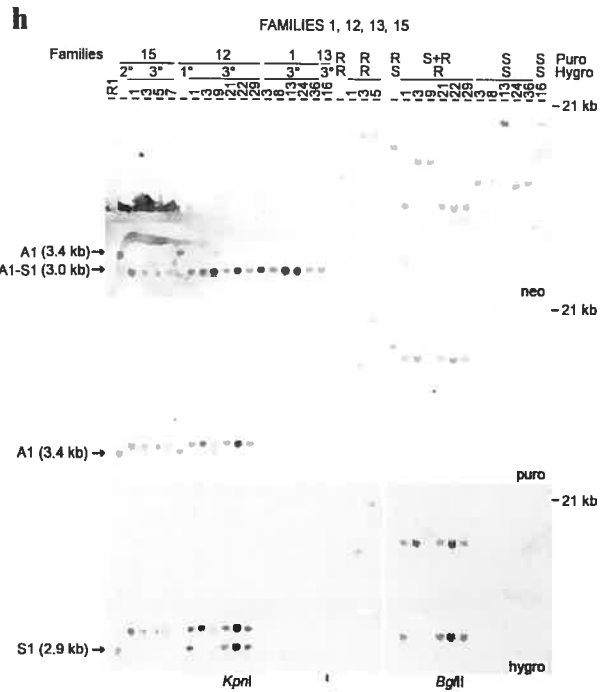
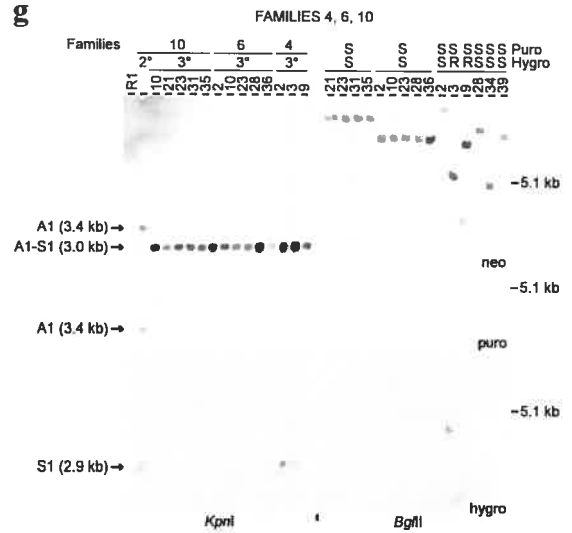
<sup>a</sup>Frequency of provirus rearrangement evaluated by Southern blot analysis of genomic DNA extracted from infected ES cell clones and/or from a polyclonal population. <sup>b</sup>Virus S3 did not infect ES cells properly. <sup>c</sup>Frequency evaluated by G418 (*geneticin*) selection of ES cell clones and/or of a polyclonal population infected by an anchor virus. N d: not determined





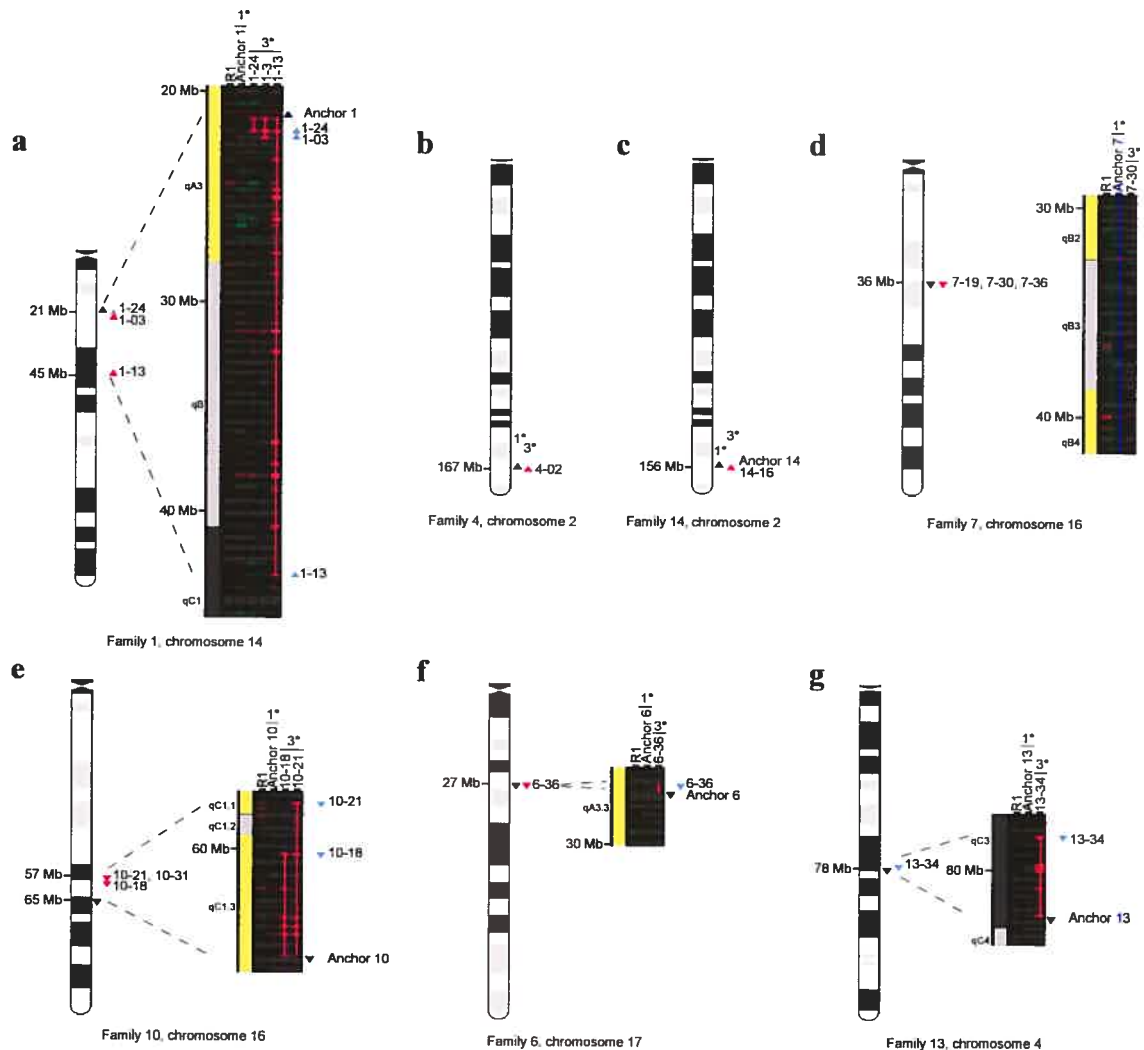


	Digests	Probes
Rearrangement analyses	KpnI	neo
	KpnI	puro
	KpnI	hygro
	HindIII	neo
Clonal analyses	HindIII	puro
	HindIII	hygro
	BglII	neo
	BglII	puro
	BglII	hygro



**Figure 2-5 Display showing confirmed chromosomal deletions in ESCs.**

Display showing the chromosomal deletions confirmed in ESCs. (a-g) CGHAnalyzer representations (right panel) linked to the chromosomal localization of confirmed deletions (Ensembl Karyoview, left panel, <http://www.ensembl.org/index.html>, v.32-Jul 2005) for the indicated families. Red lines represent nested deletions. ▲, anchor site; ▲, proviral integration site of a deletion confirmed by inverse PCR; ▲, proviral integration site of a deletion suggested by aCGH. Numbers above CGHAnalyzer representations correspond to the parental ESCs (R1), primary (1°) clones and tertiary (3°) clones.



## Figure 2-6 Evaluation of interchromosomal recombination events.

Evaluation of interchromosomal recombination events. (a) Table compiling rearrangements expected to have occurred in *trans*. (b) Clone 14-32 contains an unbalanced translocation t(2;16), as shown by aCGH and spectral karyotyping (SKY). (c) Schematic representation of transchromosomal rearrangements occurring when retroviruses are located on non-homologous chromosomes, in G<sub>1</sub> or G<sub>2</sub> cell cycle stage (see next page).

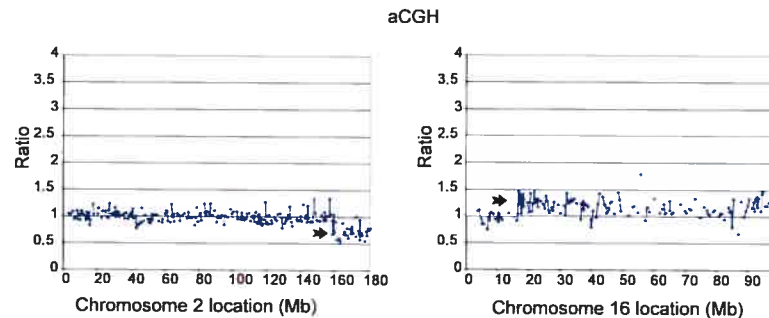
a

Confirmed or suspected recombination events in *trans*

Tertiary clone ID <sup>a</sup>	Sensitivity		Virus A1 integration			Virus S1 integration			Karyotype performed by	Certainty		
	<i>Puro</i>	<i>Hygro</i>	Chromosome	Start coordinate	End coordinate	Virus orientation (strand +/-)	Chromosome	Start coordinate			End coordinate	Virus orientation (strand +/-)
2-03	S	R	5	63,019,373	63,019,748	(+)	19	6,920,757	6,921,025	(-)	n.d.	C
4-03	S	R	2	167,486,315	167,486,681	(+)	n.d.	n.d.	n.d.	n.d.	SKY	S
4-09	S	R	2	167,486,315	167,486,681	(+)	n.d.	n.d.	n.d.	n.d.	n.d.	S
9-36	R	S	18	57,155,297	57,155,985	(+)	n.d.	n.d.	n.d.	n.d.	n.d.	S
9-40	R	S	18	57,155,297	57,155,985	(+)	n.d.	n.d.	n.d.	n.d.	n.d.	S
9-62	S	S	18	57,155,297	57,155,985	(+)	18	77,983,260	77,983,459	(-)	SKY and aCGH	S
14-01 and 14-39	S	R	2	156,503,302	156,503,387	(+)	n.d.	n.d.	n.d.	n.d.	aCGH <sup>b</sup>	S
14-27	S	S	2	156,503,302	156,503,387	(+)	16	16,998,336	16,998,873	(+)	n.d.	C
14-32	S	R	2	156,503,302	156,503,387	(+)	16	16,998,336	16,998,899	(+)	SKY and aCGH	C

<sup>a</sup>Clones showing rearrangement redundancy are on the same lane. <sup>b</sup>aCGH only for 14-39. ID, identification; *Puro*, puromycin; *Hygro*, hygromycin; S, sensitive; R, resistant; n.d., not determined.

b Example: tertiary 14-32



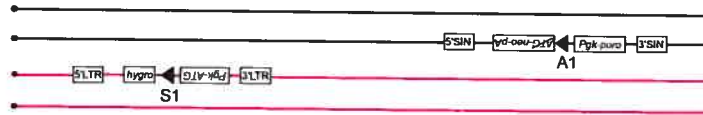
SKY



6/8 40,XY,t(2;16),del(14)

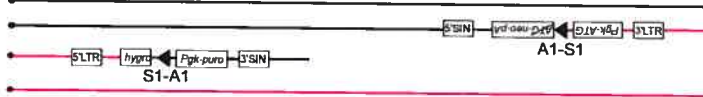
C

**Recombination in G<sub>1</sub>**

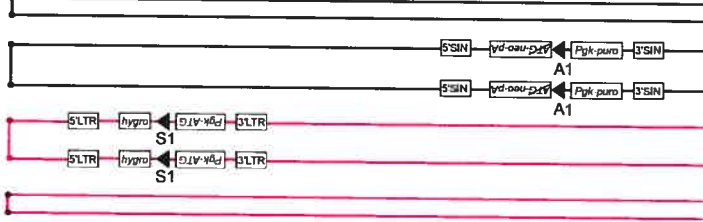


↓ Cre

Outcome: tertiary clones *neo<sup>R</sup> puro<sup>R</sup> hygro<sup>R</sup>*, balanced translocation

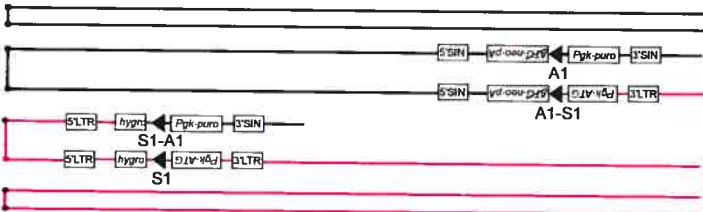


**Recombination in G<sub>2</sub>**



↓ Cre

One of the possible intermediate stage



↓

Outcome 1 (X-segregation): tertiary clones *neo<sup>R</sup> puro<sup>S</sup> hygro<sup>R</sup>*, unbalanced translocation (example : 14-32)

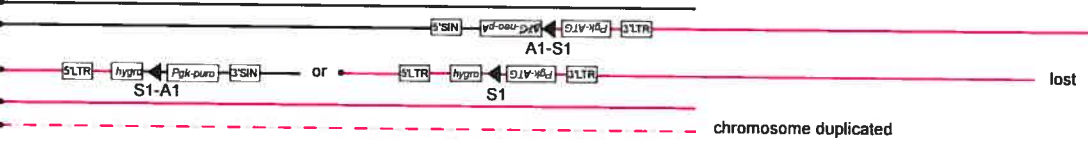


Outcome 2 (Z-segregation): tertiary clones *neo<sup>R</sup> puro<sup>R</sup> hygro<sup>R</sup>*, balanced translocation



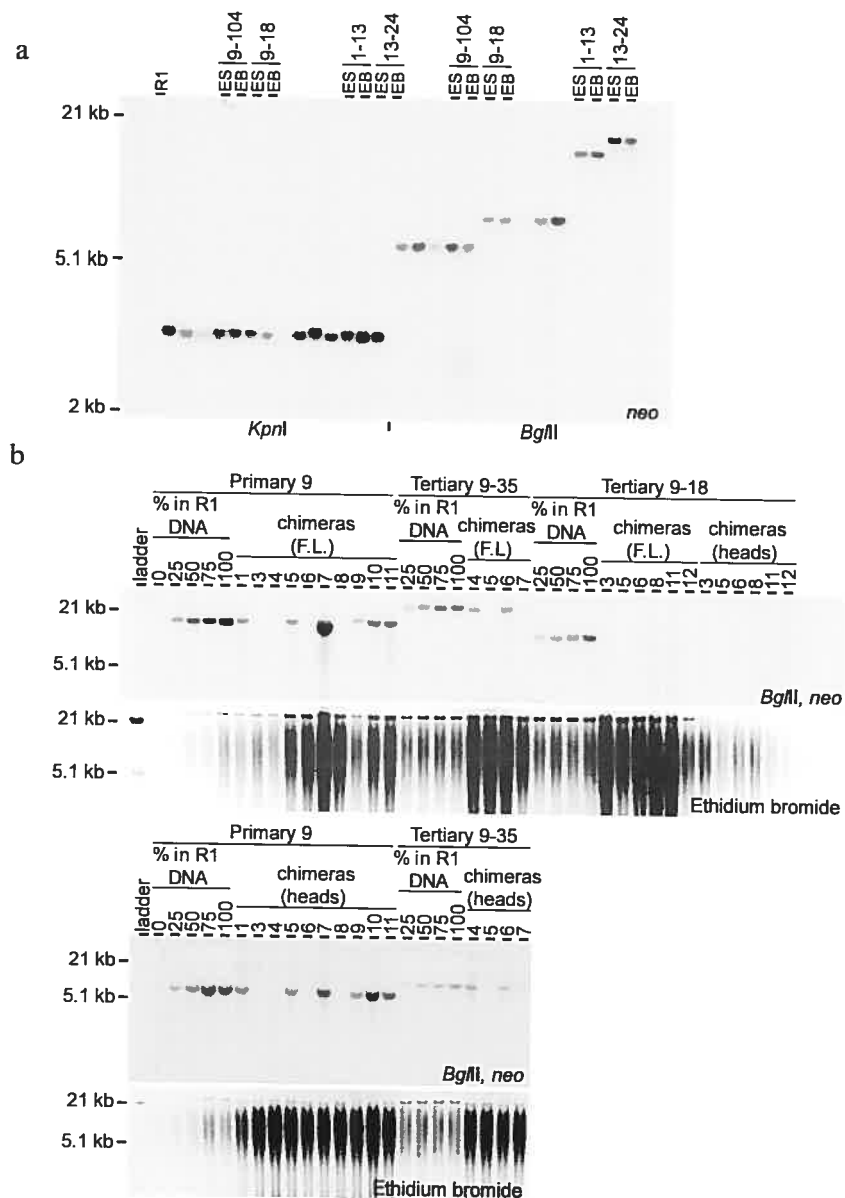
**Subsequent events to recombination in G<sub>1</sub> or G<sub>2</sub>**

We propose that infrequent tertiary clones *neo<sup>R</sup> puro<sup>S</sup> hygro<sup>S</sup>* can be obtained following trans-chromosomal rearrangements, by a mechanism involving the loss of a chromosome accompanied by the duplication of the homologous chromosome (example: 14-27).



### Figure 2-7 Full-length gels and blots.

Full-length gels and blots. (a) Full-length blot from **Figure 2-2c**. Recombination and clonal analyses (*KpnI* and *BglII* restriction digests, respectively) of DNA extracted from indicated ESCs and EBs. R1, parental ESC line (negative control). (b) Full-length gels and blots from **Figure 2-2e**. Southern blot analyses of DNA extracted from undifferentiated ESCs (clones 9, 9-35, 9-18 and R1 control), livers (F.L.) or heads of chimeric fetuses at 14.5 dpc. *BglII* restriction digests combined to a neomycin (neo) probe were used to visualize the contribution of the clones 9 and 9-35 to chimeric fetuses whereas no contribution was observed for 9-18, both in the livers and in the heads.



## 2.8 Supplementary Tables

**Table V • Summary of the Cre-mediated recombination around 11 randomly chosen loci**

Part I	Virus A1 integration				Karyo- type	Introduction of the virus S1				Puromycin sensitivity of recombined clones	
	Chro- mo- some	Start coordinate	End coordinate	Virus orien- tation (strand +/-)		aCGH	Infec- tion effi- ciency (%)	Estimated diversity <sup>f</sup>	Estimated no. of colonies after <i>cre</i> <sup>f</sup>	Frequency of G418 <sup>R</sup> colonies <sup>g</sup>	Num-ber of clones ana-lyzed
1	14	22164853	22165099	(+)	N	3	8.0E+03	2.2E+05	8.5E-04	42	0.48
2	5	63019373	63019748	(+)	N	4	1.6E+04	4.2E+05	1.4E-04	48	0.54
4	2	167486315	167486681	(+)	N	5	1.1E+04	3.0E+05	1.7E-04	31	0.42
6	17	27622141	27622184	(-)	N	2	7.0E+03	2.5E+05	6.1E-04	47	0.36
7	16	36011960	36012543	(-)	N	5	1.2E+04	2.9E+05	1.7E-04	31	0.23
9 A <sup>a</sup>	18	57155297	57155985	(+)	N <sup>b</sup>	5	1.8E+04	3.4E+05	5.9E-05	5	0.60
9 B						5	1.6E+04	2.7E+05	2.0E-04	10	0.50
9 C						4	1.5E+04	2.9E+05	1.8E-04	18	0.17
9 D						5	1.6E+04	2.9E+05	2.5E-04	45	0.22
10 A <sup>a</sup>	16	65165857	65166035	(-)	A <sup>c</sup>	6	2.6E+04	1.5E+05	3.3E-05	5	0.00
10 A						6	2.6E+04	4.2E+05	8.3E-05	34	0.18
10 B						6	2.6E+04	1.2E+05	1.5E-04	18	0.05
12	X	40497593	40497726	(+)	n.d.	3	1.1E+04	3.1E+05	1.8E-04	32	0.03
13	4	83558072	83558421	(-)	A <sup>d</sup>	10	3.8E+04	4.5E+05	2.7E-04	48	0.23
14	2	156503302	156503387	(+)	N <sup>e</sup>	3	1.3E+04	3.6E+05	2.4E-04	48	0.29
15	11	68627686	68627918	(-)	n.d.	7	3.8E+04	3.3E+05	2.1E-05	7	0.00
Sum										469	
Avera- ge <sup>a</sup>						5	1.8E+04	3.1E+05	2.5E-04		0.26
SD						2	1.0E+04	8.7E+04	2.2E-04		0.18

Part II	Analysis of <i>Puro</i> <sup>S</sup> clones (26% of clones)					Analysis of <i>Puro</i> <sup>R</sup> clones (74% of clones)		
		Hygro <sup>S</sup>		Hygro <sup>R</sup>			Hygro <sup>R</sup>	Hygro <sup>S</sup>
Clone id	Number of clones analyzed by S.B. (no. of ind. rearrangements)	Fre-quency	Type of rearrangement_ no. of clones confirmed and (ind. rearrangements)	Fre-quency	Type of rearrangements_ no. of clones confirmed and (ind. rearrangements)	Number of clones analyzed (no. of ind. rearrangements)	Fre-quency	Fre-quency
1	5 (2)	1.0	deletion_3 (2)	0		0	n.d.	n.d.
2	5 (3)	0.8	n.d.	0.2	translocation_1 (1)	0	n.d.	n.d.
4	6 (6)	0.7	deletion_1 (1)	0.3	uncertain_2 (2)	0	n.d.	n.d.
6	5 (1)	1.0	deletion_1 (1)	0		0	n.d.	n.d.
7	5 (1)	1.0	deletion_3 (1)	0		0	n.d.	n.d.
9 A <sup>a</sup>	6 (4)	1.0	deletion_4 (4)	0		2 <sup>h</sup>	0.5 <sup>h</sup>	0.5 <sup>h</sup>
9 B	11 (4)	1.0	deletion_2 (2) uncertain_1 (1)	0		8 <sup>h</sup>	0.88 <sup>h</sup>	0
9 C	6 (3)	1.0	deletion_3 (3)	0		18 <sup>h</sup>	1.0 <sup>h</sup>	0
9 D	11(2)	1.0	deletion_2 (2)	0		15 (8)	0.93	0.07
10 A <sup>a</sup>	0					5 (4)	1.0	0
10 A	4 (1)	1.0	deletion_2 (1)	0		0	n.d.	n.d.
10 B	1 (1)	1.0	deletion_1 (1)	0		13 (10)	1.0	0
12	0	n.d.	n.d.	n.d.	n.d.	6 (3)	1.0	0
13	6 (4-5)	1.0	deletion_1 (1)	0		0	n.d.	n.d.
14	5 (4)	0.4	deletion_1 (1) translocation_1 (1)	0.6	translocation_1 (1) uncertain_2 (1)	0	n.d.	n.d.
15	0	n.d.	n.d.	n.d.	n.d.	4 (2-3)	1.0	0
Sum	76					71		
Average <sup>a</sup>		0.9		0.1			0.97	0.01
SD		0.2		0.2			0.05	0.03

<sup>a</sup>pOG231 *cre* plasmid employed rather than pCX-*cre*; these populations were excluded in average analyses. <sup>b</sup>Also tested by SKY: 8 mitoses 40,XY and 4 mitoses 39,X,-Y out of 15 analyzed. <sup>c</sup>Trisomy chromosome (chr) 1. <sup>d</sup>Amplification and deletion on chr 4, position 52-76 Mb and an amplification of  $\geq 7$ Mb at the telomeric end of chromosome X. <sup>e</sup>Normal except the loss of a BAC on chr 2, position 164 Mb. <sup>f</sup>See the **Supplementary Methods** section for the calculation. <sup>h</sup>No. of G418<sup>R</sup> colonies obtained / no. of colonies after *cre* electroporation. <sup>h</sup>Determined by functional testing. Id, identification; A-D, labels of populations independently infected with the virus S1; S.B., Southern Blot analysis; no., number; ind., independent; G418<sup>R</sup>, *geneticin* resistant; Puro<sup>S</sup> or <sup>R</sup>, *puromycin* sensitive or resistant; Hygro<sup>S</sup> or <sup>R</sup>, *hygromycin* sensitive or resistant; %, percentage; N, normal; A, abnormal; n.d., not determined; SD, standard deviation; Mb, megabase pairs.



Table VI • *In vitro* differentiation of primary and tertiary clones carrying deletions.

Clone id <sup>a</sup>	Deletion size (kb)	Phenotypic analysis and (no. of experiments)
Primary 1	0	N (n=1)
1-03, 1-08, 1-24, 1-36	1 545	N (n=1-2 for each)
1-13	22 773	A (n=3)
Primary 2	0	N (n=1)
2-10, 2-22, 2-48	n.d.	N (n=1 for each)
2-41	n.d.	N (n=1)
Primary 4	0	N (n=2)
4-02	1 414	N (n=1)
4-28	n.d.	N (n=1)
Primary 6	0	N (n=1)
6-02, 6-10, 6-23, 6-28, 6-36	666	N <sup>b</sup> (n=1 for each)
Primary 7	0	N (n=1)
7-19, 7-27, 7-36, 7-30, 7-37	94	N (n=1 for each)
Primary 9	0	N (n=5)
9-31	6	N (n=1)
9-107	10	N (n=1)
9-17, 9-71	19	N (n=1)
9-68	19	N (n=1)
9-29	22	N (n=1)
9-35	23	N (n=1)
9-90	317	N (n=1)
9-104	4 182	A (n=3)
9-37	4 313	A (n=4)
9-18	5 049	A (n=5)
Primary 10	0	N <sup>c</sup> (n=2)
10-18	5 417	N (n=1)
10-21, 10-23, 10-31, 10-35	7 858	N <sup>d</sup> (n=1 for each)
Primary 13	0	N (n=3)
13-11	n.d.	N (n=2)
13-18	n.d.	N (n=1)
13-42	n.d.	N (n=3)
13-34	3 956	N (n=2)
13-24	n.d.	A (n=4)
Primary 14	0	N (n=1)
14-16	568	N (n=1)

<sup>a</sup>Clones showing a redundant rearrangement are grouped together. <sup>b</sup>Two clones were normal and three did not present a sufficient number of embryoid bodies (EBs) for phenotypic analysis. <sup>c</sup>EBs were smaller in size than normal. <sup>d</sup>EBs had outgrowths. Id, identification; kb, kilobase pairs; no., number; N, normal differentiation; A, abnormal differentiation.

## 2.9 Supplementary Methods

**Construction of pCX-Cre plasmid.** The coding sequence of Cre (*XhoI-MluI* blunted fragment from pBS185, Invitrogen) was subcloned in the *EcoRI* site (blunted) of pCX-EYFP<sup>154</sup> (a gift from A. Nagy, Samuel Lunenfeld Research Institute, Toronto, CA), replacing EYFP.

**Cell culture.** Male R1 ESCs<sup>61</sup> (provided by A. Nagy, Samuel Lunenfeld Research Institute, Toronto, CA) were grown on irradiated DR-4 mouse embryonic fibroblasts (MEFs made from DR-4 mouse strain<sup>155</sup>) or on gelatin coated dishes, in a media (DMEM high glucose with L-glutamine and pyruvate (Invitrogen), 15% fetal calf serum (Invitrogen),  $1.5 \times 10^{-4}$  M  $\alpha$ -monothioglycerol (Sigma) and  $1 \times 10^{-4}$  M non-essential amino acids (Invitrogen)) supplemented with 1000 U ml<sup>-1</sup> of ESGRO (Chemicon) or leukemia inhibitory factor (LIF, conditioned media from transfected COS cells). Serum replacement (Invitrogen) was used for extended growth period on gelatin. The number of ESCs replating was estimated by counting the colonies included in 0.09 cm<sup>2</sup> areas (n=3 minimum). ESCs were differentiated into embryoid bodies as described<sup>156</sup>, after one passage or more on gelatinized dishes. Briefly, single cell suspensions (between  $5 \times 10^3$  -  $5 \times 10^5$  cells) were plated on bacterial-grade 35 mm dishes without LIF (IMDM (Sigma), 15% fetal calf serum (HyClone), 5% serum-free and protein-free media for hybridoma culture (Invitrogen),  $2 \times 10^{-3}$  M L-glutamine (Invitrogen), 50 ug/ml ascorbic acid (Sigma) and  $3 \times 10^{-4}$  M  $\alpha$ -monothioglycerol).

**Viral producer cell lines and infection of target cells.** Retroviruses A1, A2, A3, A5, S1, S2 and S3 were generated with the GP+E-86 ecotropic packaging cell line<sup>128</sup> and maintained as described<sup>157</sup>, except for the linearized constructs (*DraI*) that were directly transfected into the producers using lipofectamine. Hygromycin (Roche, 200 ug ml<sup>-1</sup>) or puromycin (Sigma, 1.8 ug ml<sup>-1</sup>) selection was started on day 2 and maintained throughout the expansion in HXM media<sup>128</sup>. One passage prior to the infection, selection was stopped and cells were maintained in the presence of 10% newborn calf serum (NCS, Invitrogen) in DMEM (Invitrogen). Twenty-four hours before infection, producers' media were changed for complete ESC medium. Twenty-four hours infections were carried out using 4ug ml<sup>-1</sup> of polybren (Sigma). Fresh media was added the next day and selection started 48h after infection (1.5 ug ml<sup>-1</sup> puromycin and 150 ug ml<sup>-1</sup> hygromycin for ESCs). Numbers of ESC colonies surviving selection were estimated and compared to the one inferred for the replating, in order to calculate the percentage of infection. Viral titers were kept low to ensure low infection rate and to minimize chances of multiple integrations. For ESCs infected with S1,

the total number of colonies surviving hygromycin selection corresponded to the estimated integration diversity since the cells were not split. A4, S4, and S5 were VSV-G pseudotyped retroviruses produced by the amphotropic packaging cell line 293GPG<sup>129</sup>, as previously described<sup>157</sup>.

**Cre-induced recombination in ESCs.**  $10^7$  ESCs were transfected with a *cre* plasmid. Cells were resuspended in 800  $\mu$ l of ESC media and electroporated with 25  $\mu$ g of supercoiled pCX-*cre* or pOG231<sup>34</sup>, using 225 V and 950  $\mu$ F or 230 V and 500  $\mu$ F parameters, respectively. After a 20 minutes incubation at room temperature, cells were distributed on 3 dishes (100 mm) covered with irradiated MEFs. For every set of electroporation, at least one population similarly received the supercoiled pCX-*EYFP*<sup>154</sup> (225 V, 950  $\mu$ F), acting for both as an electroporation control and a negative control for G418 selection.  $2 \times 10^6$  cells of the population no.9A, no.10A, and no.10B ( $\geq 9 \times 10^4$  colonies for each) were also used as a negative control for selection. Forty-eight hours after electroporation, the replating efficiency was estimated and G418 selection started at 300  $\mu$ g ml<sup>-1</sup>. At that time, electroporation controls showed  $\geq 50\%$  of YFP positivity under the UV microscope. 7-9 days later, colonies were counted (to determine the frequency of G418<sup>R</sup> colonies), isolated in 96-well plates, and expanded in 24-well dishes before freezing. G418 selection was maintained throughout the expansion. Fractions of these clones were put aside on gelatin coated plates and tested functionally for the losses of puromycin and hygromycin, or used to extract DNA or RNA.

**DNA and RNA analysis.** Genomic DNA was isolated using DNAzol and total cellular RNA with Trizol, according to the manufacturer instructions (Invitrogen). Southern Blot and RNA analyses were performed as described previously<sup>158</sup>.

**aCGH.** aCGH was performed at the Microarray and Genomics Facility of the Roswell Park Cancer Institute (Buffalo, NY). Because the R1 ESCs have a XY genotype, the hybridizations were mainly carried with XX genomic DNA since the sex-mismatches provided internal controls. Occasionally, XY control genomic DNA was employed (samples: R1, 1-03, 9-18, 9-37, 9-90, 14-16). Deletions or amplifications were detected when adjacent BACs in the array presented a Log2 ratio  $\sim \leq -0.5$  or  $\geq 0.5$ , respectively. For viewing and comparison analyses of aCGH data from multiple clones, we used a modified version of the CGHAnalyzer<sup>142</sup>. This version was adapted to support mouse data and the use of local instances of publicly available database (kindly modified by Jean-Philippe Laverdure, BioneQ-Réseau québécois de bio-informatique, Université de Montréal, CA). aCGH data were submitted to the GEO database ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/))<sup>159</sup>.

**Spectral karyotyping and FISH.** Spectral karyotyping and FISH analyses were conducted by The Quebec Leukemia Cell Bank under the supervision of J.H. (<http://www.bclq.gouv.qc.ca/>, Maisonneuve-Rosemont Hospital, Montreal, CA).

**Chimeras' production.** Mouse chimeras were generated by the transgenic facility of IRIC. ES cells (129/Sv x 129/Sv-CP\_F1)<sup>61</sup> corresponding to primary clone no.9, tertiary clones 9-35 (no *in vitro* phenotype) and 9-18 (with *in vitro* phenotype), were injected into C57BL/6 blastocysts. ESCs from primary clone no.9 were also aggregated with CD1 morulas. Chimeric mice derived from primary clone no.9 (2 females with ~75% of coat color chimerism) and 9-35 (2 chimeric males with ~75% and ~40% of coat color chimerism) were bred with C57BL/6 mice to assess germ-line transmission of the engineered alleles (tested by PCR specific to the junction of virus A1 and the anchoring locus, or by PCR specific to the *neomycin* gene). Fifty-eight and sixty-seven pups were genotyped from each set of crosses, respectively.

**Equipment and settings.** We scanned blots and gels using a Duoscan T1200 AGFA scanner and the AGFA FotoLook 3.60.00 software (200 or 300 d.p.i.). We adjusted brightness and contrast and assembled the panels in Adobe Photoshop CS version 8.0 (Adobe Systems Incorporated). We reduced the images in Adobe Illustrator CS (Adobe Systems Incorporated).

For spectral karyotyping (SKY), we used a microscope (Axioplan 2 Imaging, Zeiss) equipped with a 63X / 1.40 immersion objective (Plan-Apochromat, Zeiss) and filters (Sky filter #1, DAPI and Cube filter, Zeiss). Images were acquired with a multi format CCD camera (C4880-80, Hamamatsu) and capture software (version 2.3, Spectral Imaging). Analyses were done with complementary software (Sky View version 1.61, Spectral Imaging).

For embryoid bodies (EBs) visualization (**Figure 2-2b**), we used an inverted microscope (Axiovert 25, Zeiss) equipped with a 5X / 0.12 (A-plan, Zeiss) and 2.5X / 0.075 (Plan-Neofluar, Zeiss) objectives, for R1 and primary clone no.9 samples, respectively. We set the camera to 4X magnification (G5, Canon). These pictures were 8-bit depth (RGB). For tertiary 9-35, we acquired the image with a microscope (DMIRB, Leica) equipped with a 10X / 0.30 objective (HC PL Fluotar, Leica), under darkfield illumination. The camera (Retiga EXi, QImaging) was equipped with a color module and link to an acquisition software (Northern Elite version 6, EMPIX Imaging Inc.). This image was 16-bit depth (RGB).

For FISH (**Figure 2-2c**), we employed a microscope (Axioskop 2 plus, Zeiss) with a 100X / 1.30 immersion objective (Iris Plan-Neofluar, Zeiss) and different wavelength filters (DAPI, FITC, TRITC, Cy3, Zeiss). Images were taken with a camera (CCD-CE-4912-5010, Applied Imaging), captured and analyzed with the same software (Cytovision version 3.6, Applied Imaging).

For fetus images (**Figure 2-2d**), we used a stereomicroscope (*MZFLIII*, Leica) coupled to a 0.63X lens (Planapo, Leica) with 1X magnification. The camera (Micropublisher 3.5, QImaging) was linked to the acquisition software (Northern Elite version 6, EMPIX Imaging Inc.). These images were 8-bit depth (RGB).

Calibration of each microscope was performed using a ruler or a micrometer microscope slide.

## ***Chapter 3* CREATION OF A LIBRARY OF ENGINEERED ESC CLONES SUITED FOR FUNCTIONAL ASSAYS**


Chapter 3 is an article in preparation (early phase), describing a team effort to generate a library of ESC clones containing deletions located broadly across the mouse genome. The strategy used to create these deletions was detailed in Chapter 2. Hopefully, this resource will be available for the scientific community in the coming months, following further characterization. These ESC clones were used to conduct preliminary screens (proliferation, differentiation) revealing interesting haploinsufficient regions. Mélanie Bilodeau participated to the project supervision, including the methodology elaboration, the team training, and technical preparation. In addition, Mélanie Bilodeau largely contributed to generating the results and their interpretation. Contribution of each author will be mentioned in the Author contributions' section.

**ARTICLE****DELES: a new library of nested chromosomal DELetions in mouse ES cells suited for functional screens**

**Mélanie Bilodeau**<sup>1</sup>, Valeria Azcoitia<sup>1</sup>, Simon Girard<sup>1</sup>, Nancy Ringuette<sup>1</sup>, Jana Kros<sup>1</sup>, Tara MacRae<sup>1</sup>, Nadine Mayotte<sup>1</sup>, Mélanie Fréchette<sup>1</sup>, Jalila Chagraoui<sup>1</sup>, Jean-Philippe Laverdure<sup>1</sup>, Jean Duchaine<sup>1</sup>, Pierre Chagnon<sup>1</sup> & Guy Sauvageau<sup>1,2,3</sup>.

<sup>1</sup>Laboratory of Molecular Genetics of Stem Cells, Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montréal, Québec, Canada, H2W 1R7. <sup>2</sup>Department of Medicine, Montréal, Québec, Canada, H3C 3J7. <sup>3</sup>Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, H1T 2M2.

Resource Article in preparation for *Cell*.

\*Correspondence: Guy Sauvageau, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada, H3C 3J7. Email : 

### 3.1 Author contributions

Experiments were designed by Mélanie Bilodeau, Nancy Ringuette, and Guy Sauvageau with the help of others. Valeria Azcoitia, Jana Krosi, Simon Girard, Mélanie Bilodeau, and Nancy Ringuette generated tertiary ESC clones. Tara MacRae, Mélanie Fréchette, and Nadine Mayotte supported ESC culture. Mélanie Bilodeau, Nadine Mayotte, Jalila Chagraoui, Tara MacRae, Valeria Azcoitia, Simon Girard, and Guy Sauvageau performed functional screens. Jean Duchaine and Pierre Chagnon were in charge of robotic cell culture and Q-PCR, respectively. Simon Girard determined retroviral integration sites by inverse-PCR. Tara MacRae extracted DNA and performed Southern blot analyses. Jean-Philippe Laverdure generated the database. The manuscript was written by Mélanie Bilodeau under Guy Sauvageau supervision. Mélanie Bilodeau, helped by Valeria Azcoitia and Jean-Philippe Laverdure, generated the tables and figures containing the team work, under Guy Sauvageau supervision.

### 3.2 Abstract

Using a previously described retroviral-based *Cre-loxP* system that efficiently creates haploid genomic deletions in mammalian cells, we now report the generation of a collection of ESC clones containing nested chromosomal deletions predicted to cover between 10-15% of the mouse genome. Except for the Y chromosome, proviral integrations were broadly distributed on all chromosomes indicating that our ability to functionally explore the ESC genome has not reached saturation. We screen this cell library containing over 1200 clones for phenotypes pertaining to cell survival, proliferation, and ability to differentiate into selected lineages such as hematopoiesis. We have identified the presence of several haploinsufficient regions for all the properties analyzed. Together, this collection of ESC clones with annotated phenotypes and chromosomal deletions (*ongoing*) is available in a centralized repository and compiled in a database accessible to the scientific community. Both these biological and bioinformatics resources will serve the scientific community for annotating coding and non-coding regions of the mouse genome.



### 3.3 Introduction

We have previously developed a retroviral-based Cre-*loxP* system that can be exploited to efficiently create haploid chromosomal deletions in mouse embryonic stem cells (ESCs)<sup>118</sup>. Most of the limited number of engineered ESC clones that were generated within this initial effort were competent in various differentiation assays and for the generation of chimeric mice<sup>118</sup>. We have also identified a significant subgroup of ESC clones whose differentiation was dominantly affected by the deleted chromosomal fragments thus indicating that several loci (or their combined deletion) are haploinsufficient. This observation is consistent with that of two genome-wide chemical screens performed with mutagenized mouse male gonads which showed an estimated frequency of dominant mutation at around 2%<sup>22,23</sup>. Based on recent estimation that approximately 600 mouse protein-coding genes (2.5%) could be imprinted<sup>160</sup> and that ESCs and their corresponding differentiated progeny demonstrate imprinting marks<sup>42,161,162</sup>, it is tempting to speculate that several haploid deletions generated in these cells are functional correspondent of hemizygous null alleles. Notably, Prader-Willi and Angelman syndromes are both associated with chromosomal deletions and imprinted gene dysregulation<sup>163,164</sup>.

Details about the potential of our approach and its applicability to large scale functional screens are provided in Chapter 1 (see section 1.3-1.4 and **Table I**) and Chapter 2. In brief and towards this goal, our approach shows the following advantages/characteristics: i) Anchoring of *loxP* sites is achieved by retroviral gene transfer rather than by gene targeting. Thus, the creation of a large library of ESC clones with chromosomal deletions can be achieved within 3-4 months including an initial *in vitro* functional screen; ii) the nested deletions cover on average 3-Mb and frequently include several contiguous genes thus offering the simultaneous interrogation of both protein-coding and non protein-coding elements (i.e., synthetic interactions); iii) the allele(s) are permanently deleted and not only silenced, and the mapping of deletions is precise.

This chapter documents our recent efforts to generate a library of nested chromosomal deletions in mouse ESCs (a project named DELES). Moreover, these ESC clones were used to perform a preliminary functional screen, highlighting several candidate regions for haploinsufficiency. DELES is an annotated resource of ESC clones, genetic material, and biological information, which will soon become available to the scientific community.

## 3.4 Results

### 3.4.1 Generation of a chromosomal deletion library in ESC clones

An overview of the methodology employed to generate deletion-containing ESC clones is provided in **Figure 3-1a,b**. The first *loxP* site was introduced in ESCs by retroviral gene transfer followed by puromycin selection (retrovirus A1, **Figure 3-1a,b**). Two hundred eighty-eight primary ESC clones (3 x 96-well plates) were thus generated and stored for further studies. A proportion of retroviral integration sites (n=102) was mapped using inverse-PCR (**Figure 3-2**). Retroviral integrations were localized on each chromosome, except chromosome Y (**Figure 3-2**). Aside from a limited sample size and pending statistical analyses, retroviral integration site distribution appeared fairly broad with the exception of a few specific chromosomal regions (e.g., see chromosomes 4 and 7) (**Figure 3-2**).

Bioinformatic analysis performed on predicted 3-Mb deletion per each primary clone, revealed the heterogeneity of different genomic features around these 102 anchor sites, such as the number of annotated CpG islands, genes, microRNAs, mRNAs, and highly conserved elements (**Table in Appendix III**). Some of these speculative regions are gene-poor (e.g.  $\leq 10$  annotated genes) or highly gene-rich (e.g.  $\geq 100$  annotated genes) (**Table in Appendix III**).

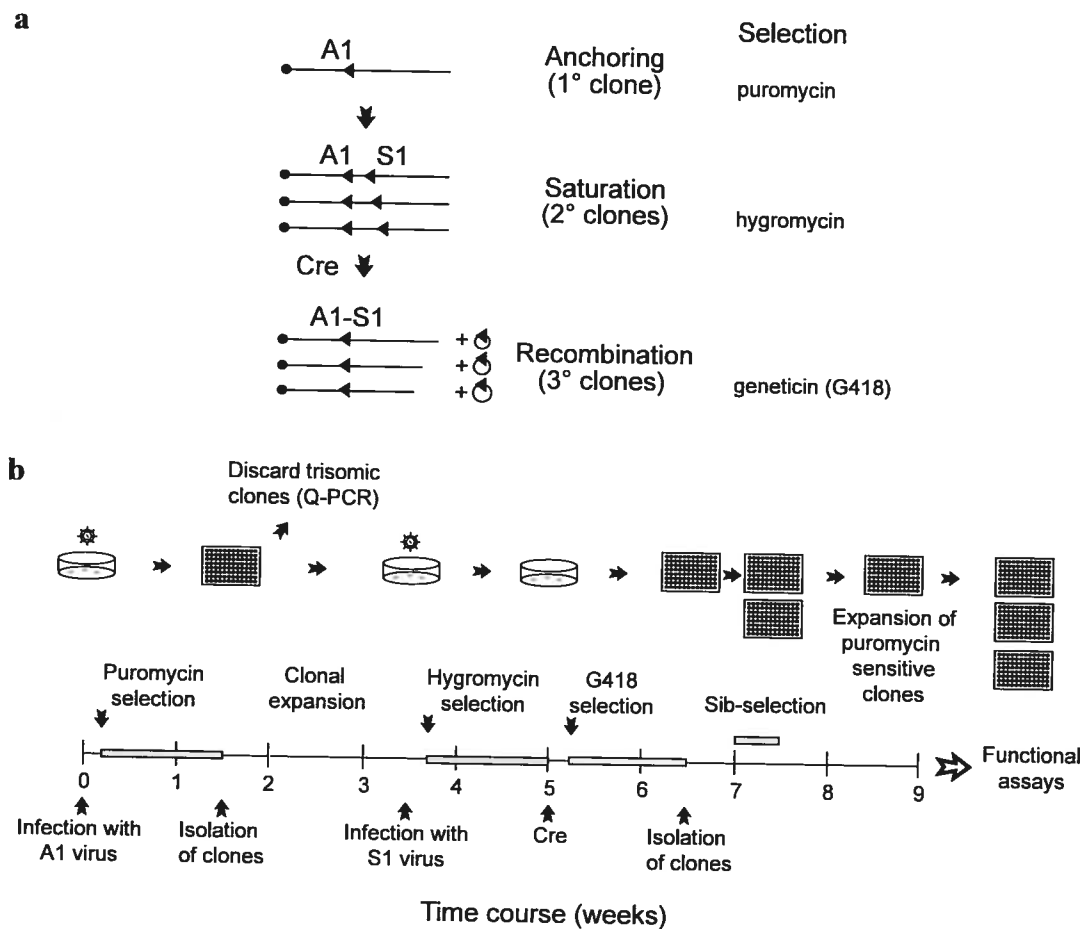
As expected from our previous work (Chapter 2), most of these primary clones contained a single A1 proviral integrant (data not shown). Prior to the second infection (retrovirus S1) (**Figure 3-1b**), we excluded primary ESC clones with chromosome 1, 8, 11, and 14 trisomies which were frequently observed in our culture conditions. For this, we used real-time quantitative PCR (Q-PCR) assays for several genes located on these chromosomes (see **Table IX** in Methods section). Normalization was done with Q-PCR assays for genes located on chromosome 3, since this chromosome is not frequently involved in mouse ESC trisomy<sup>115,118</sup> (see **Table IX** in Methods section). From these analyses, 19% of primary clones (52 out of 282 analyzed) were rejected (Chapter 1, **Table II**, primary clones D-E-F).

Five million cells from 187 independent primary clones were infected with the S1 saturation virus, then selected with hygromycin, generating the secondary populations used for Cre-electroporation (**Figure 3-1a,b and Table VII**). Following geneticin (G418) selection, 4929 tertiary clones related to 156 anchor sites were isolated (**Figure 3-1a,b and Table VII**). Thirty-one secondary populations did not form G418<sup>R</sup> tertiary clones. Aside from these

populations, an average of  $32 \pm 15$  (span: 1 to 44) geneticin-resistant (G418<sup>R</sup>) tertiary clones were generated per secondary populations (**Table VII**). According to the work described in Chapter 2, the loss of puromycin was highly predictive for chromosomal deletions in tertiary clones<sup>118</sup>. We thus conducted a puromycin sib-selection in our populations of tertiary clones to identify those which likely harbor chromosomal deletions at the expense of other rearrangements (e.g. inversions, duplications, etc.) (**Figure 3-1b**). With this approach, an average of  $11 \pm 10$  (span: 0 to 42 clones) puromycin-sensitive (puro<sup>S</sup>) tertiary clones were collected per family in 96-well plates (n=1670) (**Table VII**). The average proportion of G418<sup>R</sup> tertiary clones that lost puromycin expression was  $31 \pm 24\%$ , a number comparable to what we described previously<sup>118</sup> ( $26 \pm 18\%$ , Chapter 2, **Table V**). Of interest, we could not isolate puro<sup>S</sup> tertiary clones in 21 of the 156 families (**Table VII**). Further investigations will be required to validate these interesting observations, since the anchor virus might have integrated in the vicinity of ESC haplolethal loci, among other possibilities.

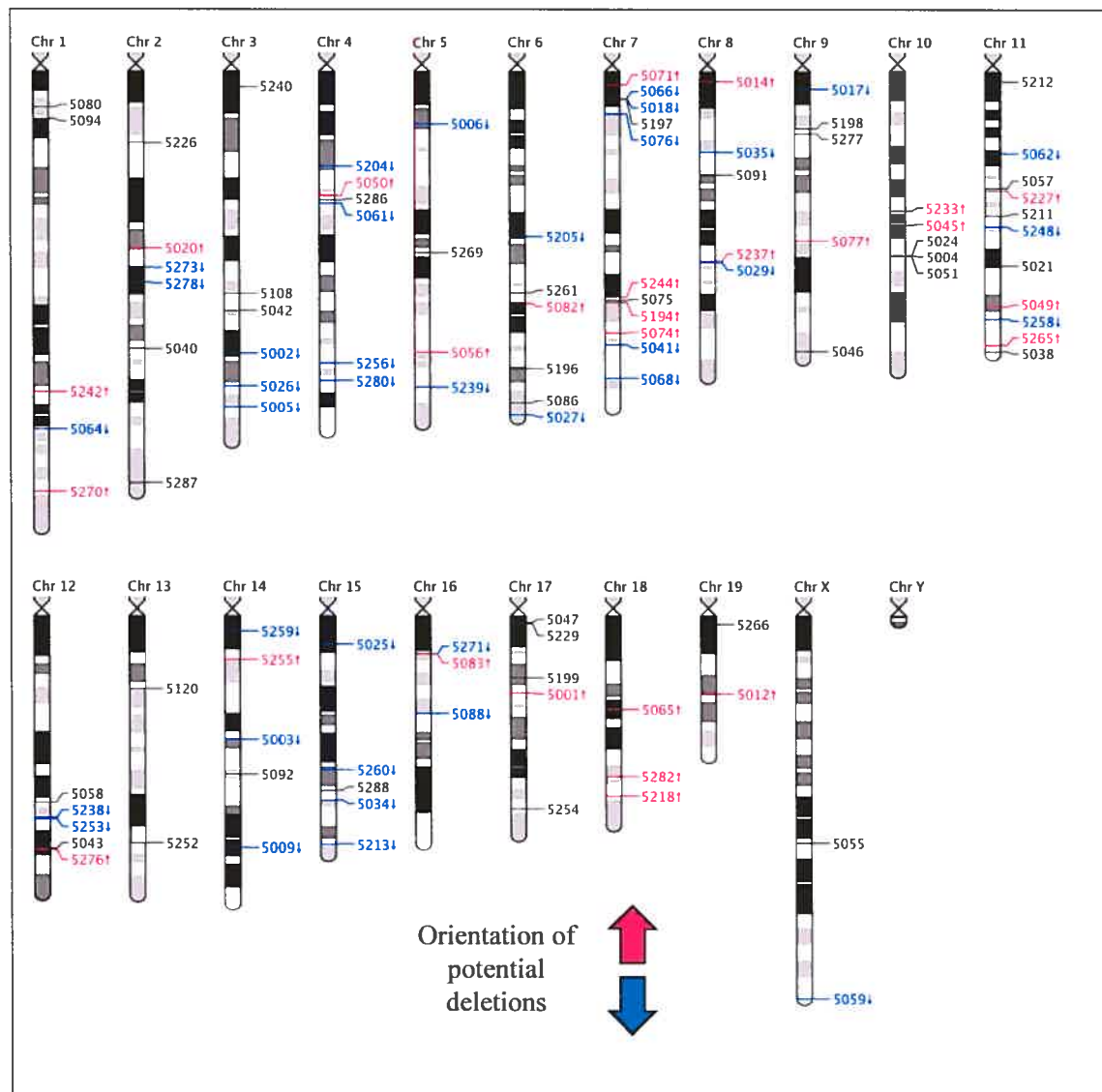
**Figure 3-1 Schematic representation of tertiary clone generation.**

(a) *LoxP* sites, successively introduced by compatible retroviruses (A1 anchor and S1 saturation viruses), are substrates for Cre-mediated recombination. Recombined clones are selected with geneticin (neomycin gene functional reconstitution) and the concomitant losses of puromycin and hygromycin genes pinpoint potential nested deletions. A family of clones contains the related parental primary (1°) clone, the secondary (2°) polyclonal population, and tertiary (3°) clones. (b) Overview of tertiary clone library generation. Note that DNA/RNA are extracted and cells frozen at various time-points (not shown), trisomic primary clones are discarded (chromosomes 1, 8, 11, and 14 trisomies suspected by Q-PCR), and a rapid puromycin selection is used to identify potential deletion-containing clones.



**Figure 3-2 Primary clone retroviral integration sites located by I-PCR.**

Note that 102 primary clone retroviral integration sites are currently mapped. The orientation of potential deletions is indicated (red and blue) for anchor sites related to chromosomal rearrangements found in G418<sup>R</sup> puro<sup>S</sup> tertiary clones that were used in functional assays (n=62, from a total of 104). Other mapped anchor sites (gray) are not related to clones for which preliminary functional assays were conducted (see **Table in Appendix III** for details).



**Table VII • Summary of G418<sup>R</sup> and G418<sup>R</sup> puro<sup>S</sup> tertiary clone generation. (Part 1 of 4)**

Family Id	Anchored chromosome <sup>a</sup>	Insertion position <sup>a</sup>	Potential deletion orientation <sup>b</sup>	Number of G418 <sup>R</sup> clones (up to 44)	Number of G418 <sup>R</sup> puro <sup>S</sup> clones	Proportion of puro <sup>S</sup> tertiary clones <sup>c</sup> (%)
5001	chr17	33507974	centromeric	44	18	41
5002	chr3	121802436	telomeric	44	7	16
5003	chr14	53526031	telomeric	44	20	45
5005	chr3	144547055	telomeric	16	2	13
5006	chr5	22919519	telomeric	44	4	9
5007				44	42	95
5008				44	10	23
5009	chr14	100119811	telomeric	44	16	36
5010				44	25	57
5011				44	2	5
5012	chr19	34101425	centromeric	11	2	18
5013				44	11	25
5014	chr8	4347514	centromeric	24	7	29
5016				44	19	43
5017	chr9	7720889	telomeric	12	6	50
5018	chr7	11840399	telomeric	36	12	33
5020	chr2	76438698	centromeric	36	2	6
5022				5	0	0
5023				44	10	23
5025	chr15	12662598	telomeric	33	13	39
5026	chr3	135937211	telomeric	44	21	48
5027	chr6	148168602	telomeric	44	8	18
5028				23	0	0
5029	chr8	82534844	telomeric	38	15	39
5030				44	23	52
5032				26	19	73
5034	chr15	79839426	telomeric	44	19	43
5035	chr8	35118318	telomeric	44	26	59
5036				4	0	0
5039				39	2	5
5041	chr7	118086289	telomeric	44	2	5
5044				44	3	7
5045	chr10	66400102	centromeric	44	28	64
5046	chr9	120614650	centromeric	44	0	0
5047	chr17	3113312	centromeric	5	0	0
5048				44	22	50
5049	chr11	101712127	centromeric	44	14	32
5050	chr4	53885955	centromeric	44	10	23

<sup>a</sup>Anchored chromosome and insertion position located by I-PCR. <sup>b</sup>Potential deletion orientation is assigned as centromeric or telomeric relatively to the A1 retroviral integration site. <sup>c</sup>The percentage of puro<sup>S</sup> clones is the ratio between the number of G418<sup>R</sup> puro<sup>S</sup> clones and the number of G418<sup>R</sup> tertiary clones. Id, identification.

---

**Part 2 of 4**


---

Family Id	Anchored chromosome <sup>a</sup>	Insertion position <sup>a</sup>	Potential deletion orientation <sup>b</sup>	Number of G418 <sup>R</sup> clones (up to 44)	Number of G418 <sup>R</sup> puro <sup>S</sup> clones	Proportion of puro <sup>S</sup> tertiary clones <sup>c</sup> (%)
5053				44	3	7
5056	chr5	121520891	centromeric	44	18	41
5059	chrX	165315983	telomeric	22	6	27
5060				13	0	0
5061	chr4	56984276	telomeric	44	30	68
5062	chr11	35263024	telomeric	44	12	27
5063				44	0	0
5064	chr1	153951498	telomeric	44	28	64
5065	chr18	40434159	centromeric	44	37	84
5066	chr7	11732942	telomeric	36	10	28
5067				44	9	20
5068	chr7	132432736	telomeric	44	19	43
5070				44	12	27
5071	chr7	5862669	centromeric	44	16	36
5072				44	10	23
5074	chr7	112839356	centromeric	44	17	39
5076	chr7	18344700	telomeric	44	18	41
5077	chr9	72920936	centromeric	44	13	30
5078				25	9	36
5079				44	5	11
5080	chr1	15775440	telomeric	8	0	0
5081				44	25	57
5082	chr6	100348725	centromeric	44	17	39
5083	chr16	16440516	centromeric	44	28	64
5084				44	15	34
5085				38	13	34
5086	chr6	143107733	centromeric	11	0	0
5087				44	31	70
5088	chr16	42675681	telomeric	44	19	43
5108	chr3	95991528	telomeric	11	3	27
5122				17	4	24
5123				7	3	43
5125				4	2	50
5126				4	0	0
5127				28	13	46
5128				44	3	7
5130				14	7	50
5133				42	9	21
5134				13	7	54
5138				28	5	18
5139				8	2	25
5140				44	13	30

<sup>a</sup>Achored chromosome and insertion position located by I-PCR. <sup>b</sup>Potential deletion orientation is assigned as centromeric or telomeric relatively to the A1 retroviral integration site. <sup>c</sup>The percentage of puro<sup>S</sup> clones is the ratio between the number of G418<sup>R</sup> puro<sup>S</sup> clones and the number of G418<sup>R</sup> tertiary clones. Id, identification.

---

<b>Part 3 of 4</b>						
Family Id	Anchored chromosome <sup>a</sup>	Insertion position <sup>a</sup>	Potential deletion orientation <sup>b</sup>	Number of G418 <sup>R</sup> clones (up to 44)	Number of G418 <sup>R</sup> puro <sup>S</sup> clones	Proportion of puro <sup>S</sup> tertiary clones <sup>c</sup> (%)
5142				29	1	3
5143				44	29	66
5144				22	19	86
5145				44	25	57
5146				44	35	80
5147				14	4	29
5151				7	0	0
5154				44	23	52
5157				44	27	61
5160				44	0	0
5161				44	2	5
5166				29	11	38
5168				44	16	36
5171				44	1	2
5172				18	6	33
5177				8	3	38
5178				36	9	25
5179				8	1	13
5183				3	0	0
5185				18	1	6
5187				3	2	67
5188				4	3	75
5189				8	0	0
5194	chr7	99820439	centromeric	44	1	2
5201				22	9	41
5202				44	7	16
5203				10	0	0
5204	chr4	41017224	telomeric	34	13	38
5205	chr6	71222644	telomeric	44	20	45
5206				15	1	7
5209				1	0	0
5213	chr15	98819220	telomeric	44	9	20
5214				44	32	73
5215				2	0	0
5216				5	1	20
5217				41	11	27
5218	chr18	77979316	centromeric	44	16	36
5219				22	0	0
5221				44	28	64
5224				28	6	21

<sup>a</sup>Anchored chromosome and insertion position located by I-PCR. <sup>b</sup>Potential deletion orientation is assigned as centromeric or telomeric relatively to the A1 retroviral integration site. <sup>c</sup>The percentage of puro<sup>S</sup> clones is the ratio between the number of G418<sup>R</sup> puro<sup>S</sup> clones and the number of G418<sup>R</sup> tertiary clones. Id, identification.



**Part 4 of 4**

Family Id	Anchored chromosome <sup>a</sup>	Insertion position <sup>a</sup>	Potential deletion orientation <sup>b</sup>	Number of G418 <sup>R</sup> clones (up to 44)	Number of G418 <sup>R</sup> puro <sup>S</sup> clones	Proportion of puro <sup>S</sup> tertiary clones <sup>c</sup> (%)
5225				44	31	70
5226	chr2	30983557	telomeric	5	0	0
5227	chr11	51463840	centromeric	44	8	18
5233	chr10	60208590	centromeric	26	3	12
5236				44	5	11
5237	chr8	81699072	centromeric	26	7	27
5238	chr12	87690206	telomeric	44	36	82
5239	chr5	136320687	telomeric	44	3	7
5241				44	35	80
5242	chr1	138279729	centromeric	27	7	26
5244	chr7	97304552	centromeric	44	7	16
5245				44	7	16
5246				17	7	41
5247				28	0	0
5248	chr11	66867224	telomeric	44	4	9
5249				44	19	43
5250				12	2	17
5253	chr12	88088474	telomeric	44	16	36
5255	chr14	19005222	centromeric	20	1	5
5256	chr4	125968987	telomeric	44	13	30
5257				30	11	37
5258	chr11	106585156	telomeric	14	4	29
5259	chr14	6848872	telomeric	28	10	36
5260	chr15	66901645	telomeric	22	3	14
5261	chr6	95681528	telomeric	22	0	0
5263				44	13	30
5265	chr11	118178073	centromeric	44	19	43
5270	chr1	181550084	centromeric	24	2	8
5271	chr16	16435791	telomeric	29	15	52
5272				7	3	43
5273	chr2	84628411	telomeric	3	2	67
5276	chr12	101234055	centromeric	19	6	32
5278	chr2	91161466	telomeric	44	36	82
5280	chr4	133587987	telomeric	44	23	52
5282	chr18	69906764	centromeric	18	2	11
5285				8	7	88
				G418 <sup>R</sup>	G418 <sup>R</sup> puro <sup>S</sup>	% puro <sup>S</sup>
			Sum	4929	1670	
			Average	32	11	31
			SD	15	10	24

<sup>a</sup>Achored chromosome and insertion position located by I-PCR. <sup>b</sup>Potential deletion orientation is assigned as centromeric or telomeric relatively to the A1 retroviral integration site. <sup>c</sup>The percentage of puro<sup>S</sup> clones is the ratio between the number of G418<sup>R</sup> puro<sup>S</sup> clones and the number of G418<sup>R</sup> tertiary clones. Id, identification.

### 3.4.2 Presentation of a functional screen performed with puro<sup>S</sup> tertiary clones

A significant portion of the generated puro<sup>S</sup> tertiary ESC clones (n=1307), containing chromosomal rearrangements anchored to 104 independent loci, was used to conduct preliminary functional screens. Analyses provided in Chapter 2 (11 independent families) revealed a 3-Mb average deletion size per puro<sup>S</sup> tertiary clone. We extrapolate that approximately 3-Mb x 104 families or  $\sim 3 \times 10^8$  bp might be deleted in the genome of our ESC collection, representing between 10-15% of the mouse genome. So far, 62 of these 104 anchor retroviral integration sites have been mapped by inverse-PCR (**Figure 3-2**).

Experimental approaches employed to perform functional screens are visually detailed in **Figure 3-3a,b**. Puro<sup>S</sup> tertiary ESC clones presenting similar proliferation rate were transferred in new 96-well plates, in order to get a more homogenous cell density for functional studies (**Figure 3-3a**). Five plate sets were generated (A, B, B\*, C, and D) (**Figure 3-3a**) based on the timing of harvest (A=earliest collection, D, latest). Following cell expansion, each of these normalized plates was used to seed cells for functional assays (**Figure 3-3b**).

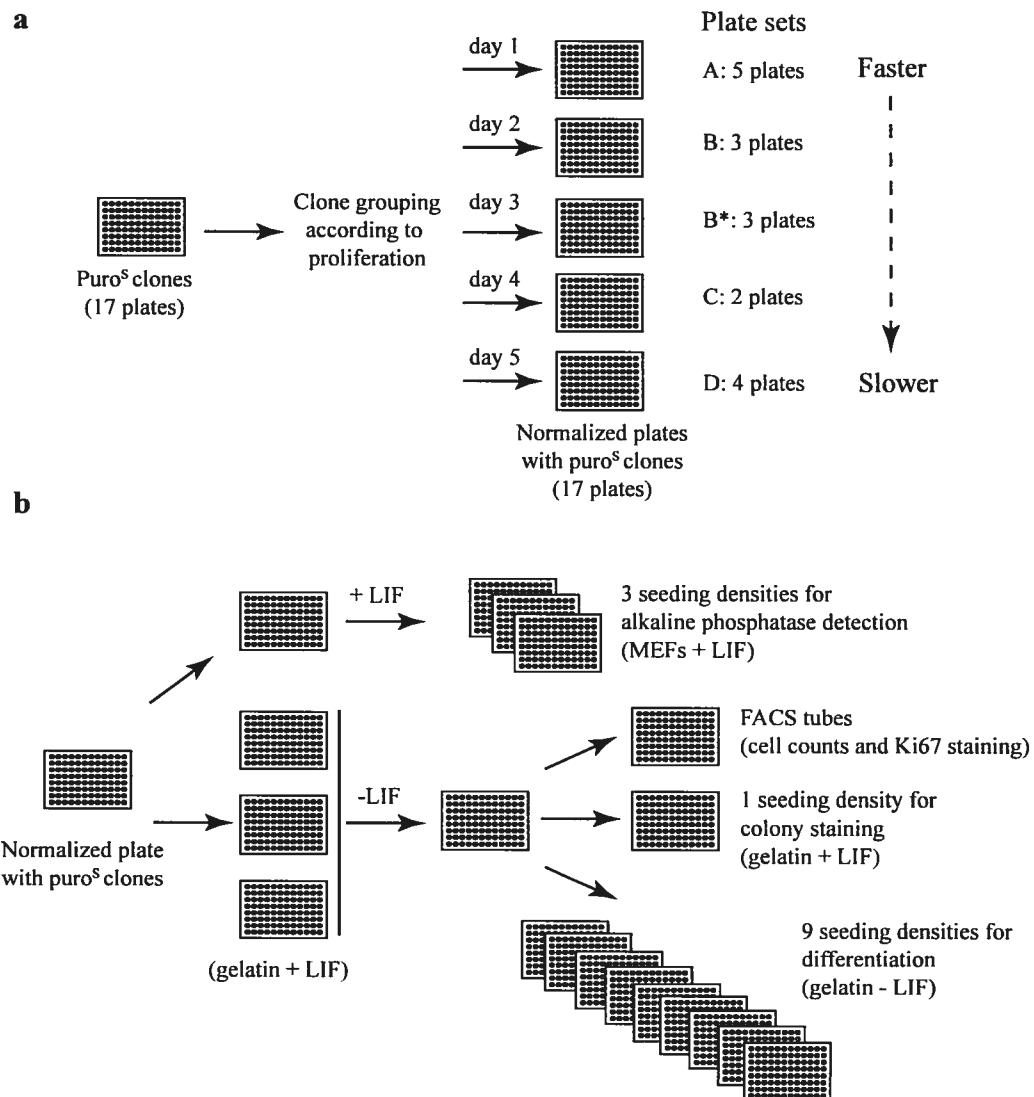
Detection of undifferentiated ESCs, maintained on a feeder layer in complete ESC media (+LIF), was achieved using alkaline phosphatase staining (**Figure 3-3b** and **Figure 3-4a**). Since ESCs differentiate when colonies fuse (over-plating), three seeding densities were prepared per tertiary clone (**Figure 3-3b**). Each puro<sup>S</sup> tertiary clone was scored in this assay and ranked from 1 (lowest percentage of undifferentiated cells) to 5 (highest percentage of undifferentiated cells) (**Figure 3-4a**).

Another portion of tertiary clone cells from normalized plates was counted by flow-cytometry, to determine the proportion of proliferating cells, and seeded for differentiation assays (**Figure 3-3b**). In order to insure the viability of cells plated for differentiation assays (culture media without LIF), a portion of puro<sup>S</sup> tertiary clone cells was seeded in ESC media (in presence of LIF) and colonies were stained the following day (**Figure 3-3b**). Cell counts correlated with estimated ESC density performed by colony staining (**Figure 3-4b**). The fraction of alive cycling ESC was determined by flow cytometry using Ki67 intranuclear staining (nuclear cell proliferation-associated antigen, expressed in G<sub>1</sub>-S-G<sub>2</sub>-M and not expressed in G<sub>0</sub> cells) (**Figure 3-4c**). For differentiation assays, puro<sup>S</sup> tertiary clone cells were seeded in 9 serial dilutions in order to perform phenotypic analyses on a range of

relevant densities (**Figure 3-3b**). Embryoid bodies (EBs) phenotypic analyses (see Chapter 2 for details) were performed at day 4 of differentiation (**Figure 3-3d**). Differentiated cell phenotypic analyses and hemoglobin histochemical staining were performed at day 8 of differentiation (**Figure 3-4e**).

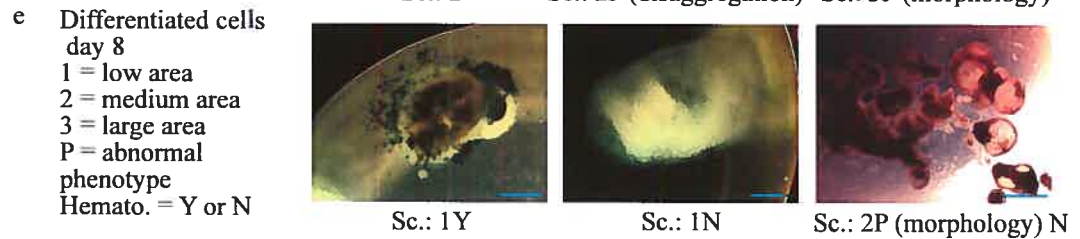
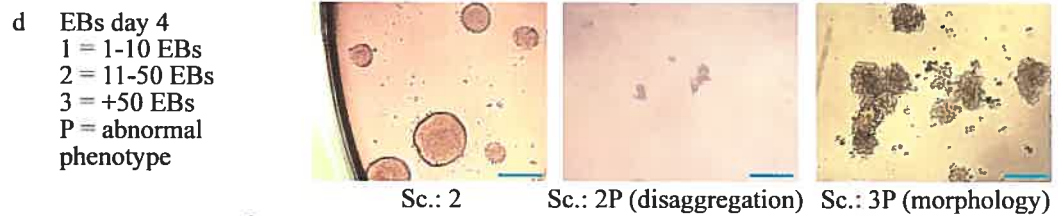
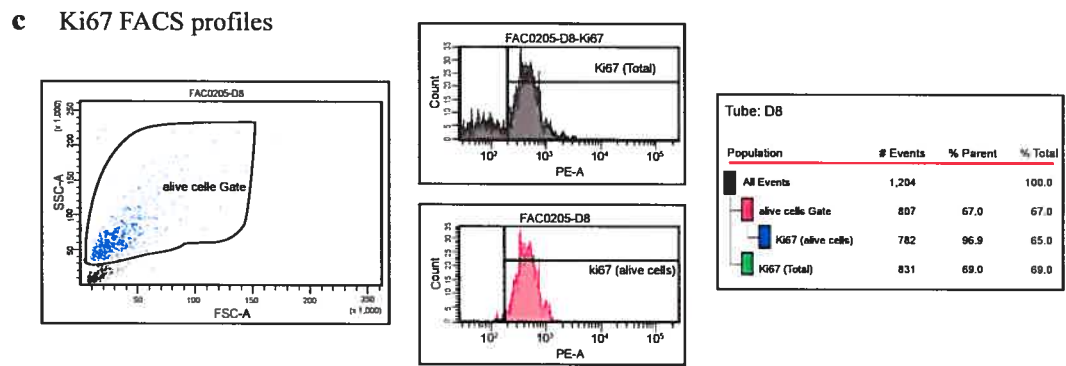
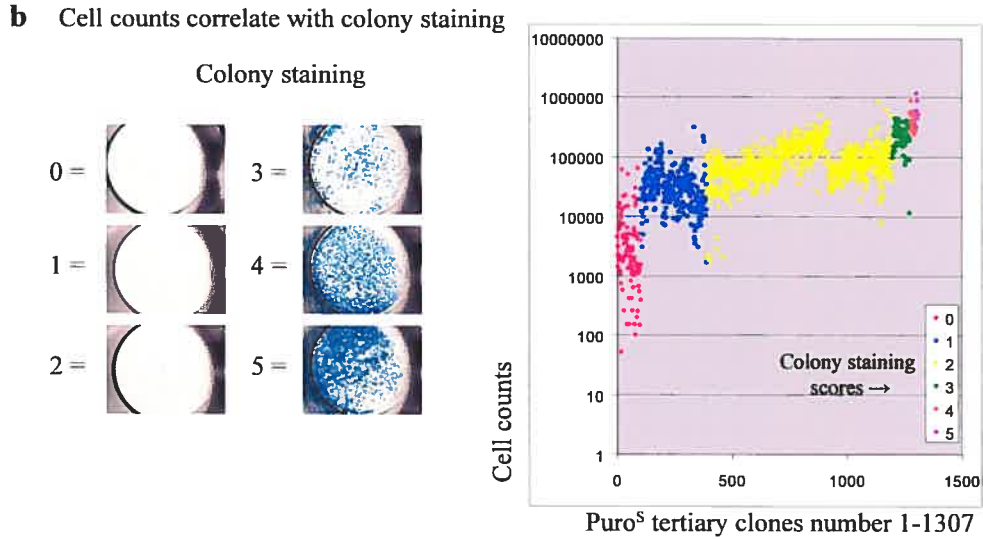
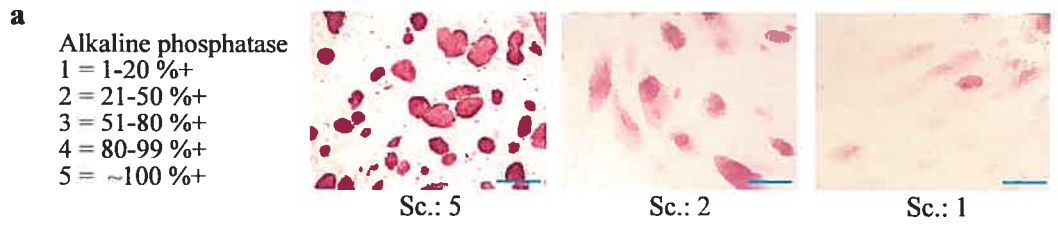
**Figure 3-3 Schematic representation of functional assay approaches.**

(a) Seventeen 96-well plates with puro<sup>S</sup> tertiary clones were thawed. Normalized 96-well plates were generated by pooling clones demonstrating similar proliferation rates (5 plate sets: A, B, B\*, C, and D). The last column (8 wells) of each normalized plate was kept for control cells included in the different assays. (b) Each normalized plate was expanded to get enough biological material for cell counts and functional assay seedings, as indicated. For clarity, frozen cells' plates and those for DNA extraction are not shown.



**Figure 3-4 Presentation of cell counts and performed functional screens.**

(a) Detection of undifferentiated cells by alkaline phosphatase staining (purple). %+, percentage of stained colonies in the 96-well. Scale bar: 250 microns. Sc., score. (b) Cell counts performed by flow cytometry correlate with estimated ESC density (colony staining with methylene blue, left side). (c) Representative FACS profiles of an ESC clone stained with Ki67 (proliferation-associated antigen) antibody. (d) Representative EBs scored for density (number of EBs) and phenotypic anomalies at day 4 of differentiation. Scale bar: 250 microns. (e) Representative differentiated cells scored for density (low, medium, and large area), phenotypic anomalies, and hematopoietic differentiation (hemato; hemoglobin detection by black benzidine staining) at day 8. Scale bar: 500 microns. Y, yes; N, no.



### 3.4.3 Global analyses of functional screens

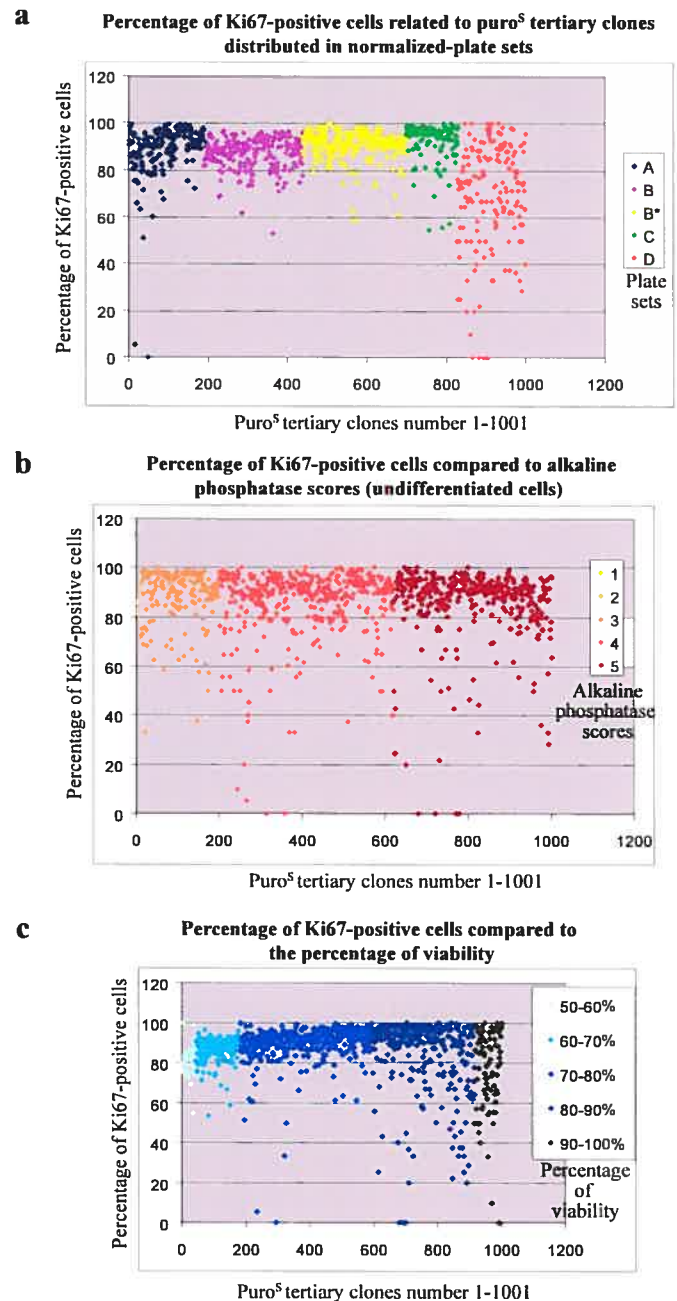
The majority of puro<sup>S</sup> tertiary clones demonstrated between 80-100% of Ki67-expressing cells (gated on live cells) (**Figure 3-5a**). However, some puro<sup>S</sup> tertiary clones presented <80% of Ki67-expressing cells, and accordingly were grouped in the normalized plates containing the slowly proliferating clones (plate set D) (**Figure 3-5a**). According to global comparison between Ki67 expression and alkaline phosphatase detection, some puro<sup>S</sup> clones presented <80% of cell positive for Ki67 expression, although they seem to contain a high proportion of undifferentiated cells (high alkaline phosphatase scores: 4-5) (**Figure 3-5b**). Inversely, several puro<sup>S</sup> clones that seemed more differentiated (low alkaline phosphatase scores: 1-2) presented a high proliferative index (>80% of Ki67-expressing cells) (**Figure 3-5b**). The percentage of Ki67-expressing cells was also compared to the percentage of viability (**Figure 3-5c**). Several puro<sup>S</sup> tertiary clones clustered in the following fashion: clones presenting 50-60%, or 60-80%, or >80% viable cells correlated with a proliferative index of ~80% or ~90%, or ~>95%, respectively (**Figure 3-5c**). Interestingly, the group presenting the highest viability (80-100%), also demonstrated a greater proportion of clones with <80% Ki67-positive cells. Together these interesting results raise the issue that a high ESC density, or an uneven distribution of ESC colonies in wells, could potentially negatively affect both the alkaline phosphatase detection and the proliferation since the cells could be differentiating and perhaps presenting a longer cell cycle (Chapter 1, **Figure 1-7**). Further cell-based proliferation analyses are needed to distinguish undifferentiated from differentiated cells (for example by using a marker such as Oct4, expressed by undifferentiated cells) on a single population (not two independent assays conducted in different conditions) to clarify this issue. Cell cycle profiling and specific apoptosis detection performed by flow cytometry will allow a better understanding of these preliminary observations. Finally, detailed analyses based on individual family of clones (e.g. tertiary clones related to the same anchor site) might highlight preponderant phenotypic observations in particular family (e.g. proliferation defects for example), hardly detectable by global analyses.

### Figure 3-5 Ki67 global analyses.

Plots presenting the correlation between the percentage of viable Ki67-expressing cells and the plate set distribution of analyzed puro<sup>S</sup> tertiary clones (a), or the alkaline phosphatase scores (b), or the percentage of viability among puro<sup>S</sup> tertiary clone cells (c).

#### 3.4.4 Specific analyses of functional screens

A database was created to compile information from each ESC clone generated in DELES project (e.g. primary, secondary, and tertiary clones), in addition to providing a user-friendly interface for interpretation of results merging from different sources (Figure 3-6a,b). A general phenotypic score and a degree of certainty were manually given to all puro<sup>S</sup> tertiary clones generated in this study (Figure 3-6b). The general phenotypic score given to puro<sup>S</sup> tertiary clones was one of the following: normal, survival (defect), proliferation (defect), differentiation (defect labeled as “no differentiation” although some differentiated cells were scored but presented phenotypic anomalies such as disaggregation), differentiation and proliferation (combination of both), hematopoietic differentiation defect, and finally other differentiation anomalies (such as morphological defect) (Figure 3-6b). Importantly, when the assignment was difficult (for examples: low amount of analyzed cells, contradictory results, technical problems observed, etc.), the mention “uncertain” was given to the score (Figure 3-6b).



**Figure 3-6 DELES database functionalities.**

(a) Database subsection “Plates” describes each ESC clone characteristics and details on the preserved biological material (cells, DNA, and RNA). (b) Database subsection “Scoring” aligns functional screen data for every *puro<sup>S</sup>* tertiary clone analyzed, grouped in their respective family. A phenotypic score was given to each tertiary clone and statistics are presented for *puro<sup>S</sup>* tertiary clones in each family.

**a** DelES : Engineered Deletions in Embryonic Stem Cells Log Out

**Plate Collections**

**Primary Cells**  
Qiagen DNA  
DNAzol DNA  
RNA

**Secondary Cells**

**Tertiary Cells**  
DNAzol DNA  
RNA

**Cherry Picked Cells**  
DNAzol DNA  
Cryovials

**Master Plates**  
Cells  
Cells No Mef  
DNAzol DNA

**Plate Search**

Plates Families Scoring Selection

◀ Back

Plate details for plate # PR10001-2 ◀ Previous Plate Next Plate ▶

	1	2	3	4	5	6	7	8	9	10
1	5001	5002	5003	5004	5005	5006	5007	5008	5009	5010
2	5011	5012	5013	5014	5015	5016	5017	5018	5019	5020
3	5021	5022	5023	5024	5025	5026	5027	5028	5029	5030
4	5031	5032	5033	5034	5035	5036	5037	5038	5039	5040
5	5041	5042	5043	5044	5045	5046	5047	5048	5049	5050
6	5051	5052	5053	5054	5055	5056	5057	5058	5059	5060
7	5061	5062	5063	5064	5065	5066	5067	5068	5069	5070
8	5071	5072	5073	5074	5075	5076	5077	5078	5079	5080
9	5081	5082	5083	5084	5085	5086	5087	5088	5089	5090
10	5091	5092	5093	5094	5095	5096	empty	empty	empty	empty

**Clone details for clone #5015**

**MEF Type :** DR4  
**Strain :** R1

**Trisomy Information :**

Chr 1	Chr 8	Chr 11	Chr 14
<input type="checkbox"/> no	<input checked="" type="checkbox"/> yes	<input type="checkbox"/> no	<input type="checkbox"/> no

**Neo southern blot Information:**  
Selected clone : single integration  
Band size : 7.43Kb

**Hygro southern blot :** not available  
**Hygro qPCR :** not available

**This clone is also present on the following plates:**

- PR10001-1
- PR10001-2
- DNA-E0001
- DNA0001
- RNA0001

IRIC, Institut de Recherche en Immunologie et en Cancérologie

DelES : Engineered Deletions in Embryonic Stem Cells Log Out

**Plate Collections**

**Primary Cells**  
Qiagen DNA  
DNAzol DNA  
RNA

**Secondary Cells**

**Tertiary Cells**  
DNAzol DNA  
RNA

**Cherry Picked Cells**  
DNAzol DNA  
Cryovials

**Master Plates**  
Cells  
Cells No Mef  
DNAzol DNA

**Plate Search**

Plates Families Scoring Selection

Show page : 1 of 3

PlateID	Format
DNA-B0125	96
DNA-B0126	96
DNA-B0127	96
DNA-B0128	96
DNA-B0129	96
DNA-B0130	96
DNA-B0131	96
DNA-B0132	96
DNA-B0133	96
DNA-B0134	96
DNA-B0136	96
DNA-B0137	96
DNA-B0138	96
DNA-B0139	96
DNA-B0140	96

IRIC, Institut de Recherche en Immunologie et en Cancérologie



DeES - Engineered Deletions in Embryonic Stem Cells

b

Plates Families Scoring Selection

Clones Type  
 Scored  
 Unscored  
 Survival  
 Proliferation  
 Differentiation  
 Diff. Hemato.  
 Other  
 Descriptor

Legend  
 All: All phenotypes  
 1-5: Opt. Density  
 Diff. Assay  
 A: Adherent  
 P: Proliferative  
 Y: Hemato.  
 C: Contamination  
 S: Colony #

Show family: 5025

Previous family Next Family

% of screened clones (1307)

Clone	Cell count	% K187	% Alive	Automated Colony Nb	Manual Colony Nb	Avg Size	% Surface
5025.02	286278.0	90.0	90.3	286	3	178.43	22.65
5025.08	102844.0	81.3	72.1	134	2	85.32	3.30
5025.10	46424.0	87.0	58.5	80	1	89.91	2.05
5025.11	160628.0	90.5	69.6	178	2	123.16	6.22
5025.18	67685.0	95.8	78.6	140	2	209.24	8.32
5025.23	129116.0	50.0	90.1	12	1	39.74	0.10
5025.27	38422.0	95.1	73.4	73	2	148.00	3.10
5025.28	5746.0	0.0	81.9	17	1	29.28	0.14
5025.30	22926.0	63.8	64.5	123	2	147.14	5.20
5025.33	173285.0	89.4	68.5	116	2	160.51	5.86

Clone	FACE				Methylene Blue				Alkaline Phosphatase				
	Cell count	% K187	% Alive	Automated Colony Nb	Manual Colony Nb	Avg Size	% Surface	Automated Colony Nb	Manual Colony Nb	[1]	[2]	[3]	[4]
5025.02	286278.0	90.0	90.3	286	3	178.43	22.65	3	0	3	0		
5025.08	102844.0	81.3	72.1	134	2	85.32	3.30	2	4	4	5		
5025.10	46424.0	87.0	58.5	80	1	89.91	2.05	1	5	5	5		
5025.11	160628.0	90.5	69.6	178	2	123.16	6.22	2	2	2	4		
5025.18	67685.0	95.8	78.6	140	2	209.24	8.32	2	5	5	5		
5025.23	129116.0	50.0	90.1	12	1	39.74	0.10	1	4	4	4		
5025.27	38422.0	95.1	73.4	73	2	148.00	3.10	2	4	4	5		
5025.28	5746.0	0.0	81.9	17	1	29.28	0.14	1	5	5	4		
5025.30	22926.0	63.8	64.5	123	2	147.14	5.20	2	4	4	4		
5025.33	173285.0	89.4	68.5	116	2	160.51	5.86	2	4	4	4		

Clone	Differentiation Assay												Score										
	Day 4			Day 8			Day 9			Day 9													
Name	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	Cell	Comment	
5025.02	2	3	2	2	N/A	A	N/A	N/A	N/A	3	3	A	A	A	N/A	A	A	A	N/A	N/A	N/A	Normal	
5025.08	A	A	A	1	N/A	N/A	A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Diff. Hemato.	
5025.10	A	A	A	1	N/A	N/A	A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Normal	
5025.11	A	A	A	2	N/A	N/A	A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Normal	
5025.18	3	A	A	A	A	N/A	A	N/A	N/A	N/A	A	N/A	1	1	1	1	1	1	1	1	1	Normal	
5025.23	0	0	0	0	N/A	0	N/A	N/A	N/A	0	0	0	0	0	0	0	0	0	0	0	0	Diff. Prost.	
5025.27	3	A	A	2	1	N/A	A	N/A	N/A	N/A	A	N/A	1	1	1	1	1	1	1	1	1	Normal	
5025.28	0	0	0	0	N/A	0	N/A	N/A	N/A	0	0	0	0	0	0	0	0	0	0	0	0	Diff. Prost.	upstream
5025.30	A	A	A	1	N/A	N/A	A	N/A	N/A	N/A	N/A	N/A	0	0	0	0	0	0	0	0	0	No Differentiation	
5025.33	A	A	A	1	N/A	N/A	A	N/A	N/A	N/A	N/A	N/A	1	1	1	1	1	1	1	1	1	Normal	

One of the database functionality is to combine the results of all functional assays with clonal analysis of provirus S1 integrations for each family of puro<sup>S</sup> tertiary clone analyzed. According to this preliminary classification, each family of puro<sup>S</sup> clones can be manually evaluated for the presence or absence of phenotypic anomalies. Four examples are provided below where two families demonstrate puro<sup>S</sup> clones associated with relatively normal phenotypes (**Figure 3-7** and **Figure 3-8**) and two families demonstrate a proportion of puro<sup>S</sup> clones associated with abnormal phenotypes (**Figure 3-9** and **Figure 3-10**). For each of these examples, ESC tertiary clones within a family that share the same integration of virus S1 are further analyzed in “subgroups” since they are believed to contain the same chromosomal rearrangement. Clones which cannot be categorized in a given subgroup are listed together for the moment (see clones missing Southern information in **Figure 3-8**). The validation of this clustering by aCGH or inverse-PCR is pending and thus, the rearrangements (deletion or other rearrangement) and the sizes of potentially deleted fragments are not known yet. In the four figures presented below, the scoring is as described in **Figure 3-4**. In addition, results presented in the column labeled “density” refer to cell counts performed by flow cytometry. Finally, note that these results remain preliminary since they have only been tested once and will merit confirmation using complementary experiments.

**Figure 3-7 Puro<sup>S</sup> tertiary clones from family no. 5077 present normal phenotypes.**

All the puro<sup>S</sup> tertiary clones in this family presented normal phenotypes. Thus, the chromosomal region that might be spanned by these unconfirmed deletions does not correlate with haploinsufficiency, according to the performed functional assays.

Plates		Families		Scoring		Admin	
Showing results for family : 5077							
<a href="#">Previous Family</a> <a href="#">Next Family</a>							
<b>Family 5077</b>							
<b>Sub-Family details :</b>	Southern Band Size :	2.6 Kb					
<b>Sub-Family A</b>							
<b>Sub-Family details :</b>	Southern Band Size :	10.13 Kb					
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5077.21	Normal		260860.0	90.2	2	5
	5077.22	Normal		49723.0	87.7	2	4
	5077.34	Normal		498955.0	98.0	4	3
<b>Sub-Family B</b>							
<b>Sub-Family details :</b>	Southern Band Size :	13.58 Kb					
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5077.44	Normal		36278.0	87.0	2	4
<b>Sub-Family C</b>							
<b>Sub-Family details :</b>	Southern Band Size :	9.08 Kb					
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5077.42	Normal		79000.0	84.0	2	3
<b>Sub-Family D</b>							
<b>Sub-Family details :</b>	Southern Band Size :	9.42 Kb					
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5077.39	Normal		393546.0	98.3	3	4
<b>Sub-Family E</b>							
<b>Sub-Family details :</b>	Southern Band Size :	9.77 Kb					
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5077.8	Normal		201375.0	93.3	2	4
	5077.16	Normal		41334.0	87.7	2	4
	5077.25	Normal		64223.0	89.5	2	3
	5077.37	Normal		55889.0	82.4	2	3

**Figure 3-8 Puro<sup>S</sup> tertiary clones from family no.5278 present normal phenotypes.**

Again most of the puro<sup>S</sup> clones in this family have a normal phenotype. The clustering according to Southern blot analyses suggest that the clone 5278.21 with the phenotypic score “normal uncertain” might be in fact “normal”. This example shows the additional information provided by subgrouping of clones.

Family 5278							
Sub-Family details :	Southern Band Size :	5.67 Kb					
<b>Sub-Family A</b>							
Sub-Family details :	Southern Band Size :	11.83 Kb					
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.25	Normal		99474.0	92.3	2	3
<b>Sub-Family B</b>							
Sub-Family details :	Southern Band Size :	11.99 Kb					
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.1	Normal		46925.0	80.9	2	4
<b>Sub-Family C</b>							
Sub-Family details :	Southern Band Size :	7.41 Kb					
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.2	Normal		69284.0	83.1	2	4
	5278.3	Normal		139068.0	90.3	2	3
	5278.4	Normal		127628.0	90.5	2	3
	5278.5	Normal		37422.0	89.3	2	3
	5278.10	Normal		42079.0	90.7	2	4
	5278.15	Normal		88628.0	87.2	2	3
	5278.16	Normal		53606.0	92.1	2	4
	5278.17	Normal		80085.0	86.9	2	3
	5278.19	Normal		79933.0	87.2	2	3
	5278.20	Normal		78051.0	89.1	2	3
	5278.21	Normal	uncertain	78509.0	88.5	2	3
	5278.22	Normal		48255.0	91.2	2	4
	5278.23	Normal		45204.0	93.5	2	4
	5278.24	Normal		64373.0	87.1	2	3
	5278.26	Normal		42711.0	89.8	2	4
	5278.27	Normal		40974.0	92.1	2	4
	5278.29	Normal		63153.0	81.8	2	3
	5278.31	Normal		72395.0	92.3	2	3
	5278.32	Normal		56848.0	87.6	2	3
	5278.35	Normal		84255.0	91.2	2	3
	5278.37	Normal		48790.0	91.2	2	4
	5278.38	Normal		49737.0	91.5	2	5
	5278.39	Normal		99553.0	92.4	2	3
	5278.40	Normal		34475.0	91.2	2	3
	5278.42	Normal		37628.0	94.8	2	4
	5278.43	Normal		78560.0	92.8	2	4
<b>Sub-Family D</b>							
Sub-Family details :	Southern Band Size :	8.56 Kb					
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.8	Normal		42356.0	89.7	2	3
	5278.12	Normal		65899.0	89.6	2	4
<b>Sub-Family E</b>							
Sub-Family details :	Southern Band Size :	9.53 Kb					
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.6	Normal		54204.0	92.5	2	3
	5278.13	Normal		27051.0	86.5	1	4
	5278.28	Normal		81967.0	87.4	2	3
<b>Clones Missing Southern Information</b>							
Member details :	CloneID	Phenotype	Comment	Density	Ki67	M.Blue	Alk.Po4
	5278.34	Normal		159563.0	92.9	3	4
	5278.36	Normal		67373.0	93.2	2	3
	5278.44	Normal		44645.0	95.9	2	4

**Figure 3-9 Certain puro<sup>S</sup> tertiary clones from family no.5032 present abnormal phenotypes.**

Clustering of puro<sup>S</sup> clones performed by Southern blot analyses highlight and strengthen the phenotypic anomalies observed in sub-family C (Diff. Prolif.= combined differentiation and proliferation defects). See the low Ki67 percentages (the differentiation data are not represented). More clones presenting these anomalies are not classified in sub-families yet (gray). At least one puro<sup>S</sup> clone (sub-family B) presented normal phenotype during the assays.

Plates		Families		Scoring		Admin	
Showing results for family : 5032				<a href="#">Previous Family</a> <a href="#">Next Family</a>			
<b>Family 5032</b>							
<b>Sub-Family details :</b>	Southern Band Size :		5.4 Kb				
<b>Sub-Family A</b>							
<b>Sub-Family details :</b>	Southern Band Size :		5.91 Kb				
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5032.23	Diff. Hemato.	uncertain	242746.0	88.4	3	4
<b>Sub-Family B</b>							
<b>Sub-Family details :</b>	Southern Band Size :		6.62 Kb				
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5032.9	Normal		34594.0	85.6	1	5
<b>Sub-Family C</b>							
<b>Sub-Family details :</b>	Southern Band Size :		7.79 Kb				
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5032.6	Diff. Prolif.		255.0	50.0	0	N/A
	5032.10	Diff. Prolif.		59797.0	45.5	2	4
	5032.11	Diff. Prolif.		25119.0	58.3	1	4
	5032.14	Diff. Prolif.		31526.0	45.5	1	4
	5032.15	Diff. Prolif.		4373.0	60.0	0	5
	5032.16	No Differentiation		110543.0	70.0	2	3
	5032.17	Diff. Prolif.		50187.0	54.5	1	5
<b>Clones Missing Southern Information</b>							
<b>Member details :</b>	<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
	5032.5	Diff. Prolif.		1170.0	15.0	0	N/A
	5032.20	Diff. Prolif.		7373.0	37.5	1	4
	5032.21	Diff. Prolif.		88984.0	40.0	1	4
	5032.24	Diff. Prolif.		10678.0	25.0	0	5
	5032.25	Diff. Prolif.		9560.0	20.0	0	5

**Figure 3-10 Certain puro<sup>S</sup> tertiary clones from family no.5276 present abnormal phenotypes.**

This family contains both puro<sup>S</sup> tertiary clones that showed normal phenotypes (sub-families A and D), or abnormal phenotypes (sub-families B, C, and E). The abnormal phenotypes are related to proliferation (percentage Ki67 below 80%) and differentiation defects. This family is an interesting candidate for further validation and characterization.

Plates		Families		Scoring		Admin		
Showing results for family : 5276				<a href="#">Previous Family</a> <a href="#">Next Family</a>				
<b>Family 5276</b>								
<b>Sub-Family details :</b>		Southern Band Size :		3.39 Kb				
<b>Sub-Family A</b>								
<b>Sub-Family details :</b>		Southern Band Size :		4.65 Kb				
<b>Member details :</b>		<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
		5276.12	Normal		239485.0	86.7	3	4
<b>Sub-Family B</b>								
<b>Sub-Family details :</b>		Southern Band Size :		4.94 Kb				
<b>Member details :</b>		<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
		5276.5	Proliferation		186550.0	57.1	3	3
<b>Sub-Family C</b>								
<b>Sub-Family details :</b>		Southern Band Size :		5.09 Kb				
<b>Member details :</b>		<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
		5276.1	Diff. Hemato.	uncertain	30388.0	86.6	2	5
		5276.2	Proliferation	uncertain	225450.0	73.9	3	3
<b>Sub-Family D</b>								
<b>Sub-Family details :</b>		Southern Band Size :		5.41 Kb				
<b>Member details :</b>		<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
		5276.8	Normal		443391.0	94.8	5	3
<b>Sub-Family E</b>								
<b>Sub-Family details :</b>		Southern Band Size :		5.93 Kb				
<b>Member details :</b>		<b>CloneID</b>	<b>Phenotype</b>	<b>Comment</b>	<b>Density</b>	<b>Ki67</b>	<b>M.Blue</b>	<b>Alk.Po4</b>
		5276.14	Diff. Prolif.		20200.0	28.6	1	5

### 3.5 Discussion and conclusions

In the short term, priority will be to confirm the chromosomal deletions and to map them in order to identify their genomic characteristics, particularly in families grouping puro<sup>S</sup> clones that demonstrated phenotypic anomalies in the functional assays. Then, interesting families could be validated and characterized, either based on tertiary clone phenotypic observations and/or the presence of interesting genomic features found in deleted segments (for examples: presence of specific genes or microRNAs, gene-poor region, etc.). As an example, the family no.5276 presented above contains puro<sup>S</sup> tertiary clones that showed proliferation/differentiation defects during the assays (**Figure 3-10**); using bioinformatic analysis, known genomic features found in the vicinity of rearrangement anchor site can be evaluated (**Table VIII**).

In conclusion, an annotated library of ESC clones containing potential deletions, anchored broadly across the mouse genome, was generated. Results from preliminary functional assays revealed that numerous haploinsufficient regions might have been pinpointed. Moreover, following the mapping of deletion end points, this library of ESC will be available to the scientific community and probably exploited in complementary functional assays, both *in vitro* and *in vivo*. Together this information load will contribute to the functional annotation of the mouse genome sequence.

**Table VIII • Genomic features found in a region spanning a virtual 3-Mb deletion anchored to the virus A1 retroviral integration site related to family no. 5276.**

Data extracted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32,152,153</sup>.

Family Id	Anchor Chr	Anchor site Start	3Mb region Stop	Number of CpG Islands	Number of RefSeq genes	Gene accession numbers	Gene names	Number of microRNAs	Number of mRNAs	Selected mouse phenotypes reported for this 3 Mb region	Number of highly conserved regions
5276	chr12	98234055	101234055	20	18	NM_008079, NM_008152, NM_029911, NM_178914, NM_011877, NM_001008506, NM_029334, NM_001081191, NM_029553, NM_198311, NM_183186, NM_030172, NM_028354, NM_146037, NM_008947, NM_009790, NM_153587, NM_145448	Galc, Gpr65, 1700024D23Rik, Spata7, Ptpn21, 2700069A02Rik, 2700069A02Rik, Emi5, Ttc8, Ttc8, Ches1, 2610021K21Rik, Tdp1, Konk13, Psmc1, Calm1, Rps6ka5, 9030617O03Rik	0	269	NM_008079_Galc:Immune, Cellular, Growth Size, Homeostasis, Life Span; NM_008152_Gpr65:Immune	292

Id, identification; Chr, chromosome.



### 3.6 Methods

Note that several methods were previously described in Chapter 2.

**A1 and S1 retroviruses, and pCX-Cre constructs; inverse-PCR, sequencing, and mapping;** were described previously<sup>118</sup>.

**Cell culture.** R1 ESCs<sup>61</sup> were maintained as described previously<sup>118</sup>. Penicillin (100 U ml<sup>-1</sup>)-Streptomycin (100 ug ml<sup>-1</sup>, Invitrogen) or occasionally Fongizole (100 U ml<sup>-1</sup> Penicillin, 100 ug ml<sup>-1</sup> Streptomycin, and 0.25 ug ml<sup>-1</sup> Amphotericin B, Sigma) were added to the culture media. For culture in 96-well plate, ESCs were either dissociated manually or with a Biomek FX robot (Beckman Coulter) placed in a sterile hood. ESCs in 96-well format were either frozen in 96-well polypropylene plate (Costar, Fisher Scientific) covered with a rubber mat (Fisher Scientific), or individually aliquoted in cryotubes labeled with a 2D bar code (CryoBank™, NUNC). ESC differentiation in attached embryoid bodies was performed in gelatinized 96-well plates (Sarsted), with a LIF-depleted media described previously<sup>118</sup>. The cellular equivalent of half a 96-well (ESCs grown on gelatin), was used to seed the first differentiation dilution. Then, 8 serial dilutions (1:4) were performed in order to obtain proper densities for each clone. Scoring was done manually using an inverted microscope.

**Viral producer cell lines and infection of target cells** were conducted as described previously<sup>118</sup>. Around 288 primary clones were generated, DNA/RNA were extracted, and Q-PCR assays were performed on genomic DNA to detect specific trisomies (see below). Primary clones with anomalies were rejected and others were expanded and frozen. Five million primary clone cells were infected with the virus S1 (supernatant containing S1 viruses diluted 1:12). Following hygromycin selection, these secondary populations were also frozen.

**Cre-induced recombination in ESCs.** Ten million cells from each secondary population were electroporated with 25 ug of supercoiled pCX-Cre and were maintained as described previously<sup>118</sup>. Up to 44 neomycin resistant tertiary clones were isolated per electroporated secondary population. Part of these clones was frozen (plates labeled TER0xxx) and RNA extracted. Puromycin selection was carried out to identify puromycin sensitive tertiary clones. Those were isolated, pooled in new 96-well plates (labeled CPC0xxx), expanded and frozen, and DNA was extracted. Normalized 96-well plates (labeled MPL0xxx) were generated with puromycin sensitive clones presenting similar proliferation (plate-sets A, B,

B\*, C, and D generated on succeeding days). These clones were expanded and frozen, used for functional assays, and DNA was extracted.

**Flow cytometry analyses.** The cellular equivalent of one 96-well was used for flow cytometry (ESCs grown on gelatin). Cell counts using TruCOUNT reference beads (BD Biosciences) were performed with half of these cells and Ki67 intranuclear staining (PE-conjugated mouse anti-human Ki67 monoclonal antibody; dilution 1:100, BD Biosciences) was done with the other half. The percentage of cell alive was gated according to forward and orthogonal light scatters.

**Colony staining.** The cellular equivalent to ~15% of a 96-well was seeded on gelatinized plate in ESC media (+LIF) and maintained for one day before staining. ESC colonies were directly stained in 96-well plates with 100ul of methylene blue solution (0.3% methylene blue in methanol, Sigma), at room temperature for 10 minutes. Then, plates were washed in water and dried. Scoring was done manually using an inverted microscope or by automated microscopy.

**Alkaline phosphatase detection.** For alkaline phosphatase detection, ESCs grown on gelatinized 96-well plate (+LIF) were seeded on MEFs in ESC media (+LIF)<sup>118</sup>. Three seeding densities were used: 2%, 4%, and 8% of cells from the donor plate (2%, 6%, and 18% only for normalized plate set A). Following three days of culture, alkaline phosphatase detection was performed according to the manufacturer's instructions (Chemicon). Scoring was done manually using an inverted microscope.

**Hemoglobin histochemical staining.** A 3% benzidine stock solution was made by diluting 4, 4'-Diaminobiphenyl (Sigma) in a 90% glacial acetic acid-10% water solution. Prior to use, the benzidine stock solution (1 part) was mixed with hydrogen peroxide concentrate (1 part, Sigma) and water (5 parts). EBs-day 8 hemoglobin staining was directly done in 96-well plates, by adding 15ul of benzidine solution per well to the 150ul EB differentiation media (dilution 1:10).

**DNA and RNA analyses.** Genomic DNA was isolated using DNAzol and total cellular RNA with Trizol, according to the manufacturer's instructions (Invitrogen). Southern blot analyses were performed as previously described<sup>118</sup>, using *EcoRI* or double *BglII/BamHI* restriction digests for primary or tertiary ESC clones, respectively.

**Chromosome 1, 8, 11, and 14 trisomy detections.** For Q-PCR analyses, genomic DNA was extracted with DNeasy 96 Blood & Tissue Kit (Qiagen). Gene copy number was determined using primer and probe sets from Universal ProbeLibrary (Exiqon TaqMan probes, Roche Diagnostics) (Table IX, assays in bold). 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 (IR160). A ratio of 1.5 (test versus chromosome 3 control) was indicative of a potential trisomy

**Table IX • Q-PCR assays employed to detect chromosome 1, 8, 11, and 14 trisomies.**

Chromosomal band	Gene	Accession number	Assay_Id	Primer_A	Primer_B	Universal ProbeLibrary probe number
1qB	<i>Actr1b</i>	NM_146107	<b>IR157</b>	atgcagccaagagtcagagc	tgaagagagtggggcaaac	21
1qD	<i>Hes6</i>	NM_019479	IR158	gggcataattctgcggta	tgggatggcaaccaaact	68
1qH3	<i>Cd244</i>	AK137505	IR159	catggctcaaagctcacaac	aggatgagccactgtaac	78
3qA1	<i>E2f5</i>	NM_007892	<b>IR160</b>	cctccagtgaccacattcagt	tgaactggagcctgtgtaa	55
3qE1	<i>Shox2</i>	NM_013665	IR161	gggaactaaaattcggctttgt	gccacactcctttgccaagt	60
3qG1	<i>Abca4</i>	NM_007378	IR162	acaccaggagtcacagtgga	ggtgagccagtgaaatttgg	25
8qA1.1	<i>Lamp1</i>	AK004637	<b>IR163</b>	ggcatctggctgggtaca	ggaaagtggcagctcacg	99
8qA4	<i>Msr1</i>	L04274	<b>IR164</b>	gtggtagtggagcccatga	ccagcgatcatcacagattg	31
8qD3	<i>Hsf4</i>	AB029349	IR165	agcaacgcctcctacttgg	caggcttttcagaggggatg	55
11qA1	<i>Ramp3</i>	NM_019511	<b>IR166</b>	ggctcggttccctagtttct	tcaggactagaaatgggtcagg	53
11qB1.3	<i>Hspa4</i>	D85904	IR167	gatggccaagggagacaacc	gccatcagaaggcacagc	66
11qD	<i>Hoxb4</i>	NM_010459	<b>IR168</b>	ctctcggaccgcctacaact	ggtagcgaatgtagtgaactcc	62
14qA1	<i>Pxk</i>	NM_145458	<b>IR169</b>	cattaaccacagataaagggttgc	aatgttggccctcctctac	102
14qC3	<i>Rnf17</i>	NM_001033043	IR170	tccegtttaccaccgtatc	ttcaactgcacggcaaac	82
14qE4	<i>Sox21</i>	NM_177753	IR171	tgaagatgcctctcaccaa	ctgaaaaacaggccaaaacag	52

Id, identification.

**Biological material tracking and database construction.** Frozen cells, DNA, RNA, and maintenance plates were identified with bar codes and a specific labeling. A database running on a MySQL server 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. A JSP web front-end running on a Tomcat 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.

### **3.7 Acknowledgments**

We thank Danièle Gagné from IRIC Flow Cytometry Core Facility; Anne-Sophie Guenier from IRIC High Throughput Screening Core Facility; Raphaëlle Lambert from IRIC Genomics Core Facility; Christian Charbonneau from the IRIC Bio-Imaging Core Facility; Edlie St-Hilaire for MEFs generation; Andras Nagy for providing the R1 ESCs and the pCX-EYFP construct (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto); Rudolf Jaenisch for the DR-4 mouse strain (Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology). This work was supported by Génome Québec and Canadian Institutes of Health Research (CIHR) grants to Guy Sauvageau. Mélanie Bilodeau is a recipient of a CIHR studentship and Guy Sauvageau is a recipient of a Canada Research Chair in molecular genetics of stem cells and a scholar of the Leukemia Lymphoma Society of America.

## ***Chapter 4* DISCUSSION AND PERSPECTIVES**

Since focused conclusions and discussions were provided through other parts of this thesis, Chapter 4 will consist in a more global discussion. Topics covered will be haploinsufficiency, pending optimizations with an emphasis on complementation, and finally potential applications of the system. Analyses, figures, and tables were generated by Mélanie Bilodeau under Guy Sauvageau supervision. Imprinted gene analysis was performed in collaboration with Jean-Philippe Laverdure.

## 4.1 Haploinsufficiency and imprinting

Several haploinsufficient elements or regions are probably associated with mutations in genes or in regulatory elements, altering expression levels. The exact proportion of haploinsufficient determinants is not known. According to two independent genome-wide ENU screens in mouse, the dominant mutation frequency was estimated around 2% according to the defects monitored: immunological anomalies, abnormal size or behavior, congenital malformations, etc<sup>22,23</sup>. In a way, this number could be an overestimation since each mouse could contain multiple ENU-induced mutations. However, the calculation was adjusted according to the frequency of mutations inherited dominantly, seemingly as a monogenic phenotype, following breeding of a proportion of founder F1s with wild-type mice<sup>22</sup>. Also, since this type of screen does not allow the recovery of haplolethal mutation, the frequency of haploinsufficiency could be underestimated.

Deletions engineered with the described methodology are physically haploid. In some instances, because of imprinting, some mutated alleles could be functionally null. Paternal or maternal allele expression of imprinted genes is regulated by differentially methylated domains (DMDs)<sup>161</sup>. Methylated imprints are initiated during gametogenesis, observable at the two-cell stage, and are thought to be resistant to global demethylation during embryonic cleavage (preimplantation period between fertilization and blastocyst stage)<sup>161,164</sup>. Already at the blastocyst stage, several imprinted genes show monoallelic expression<sup>164</sup>. ESCs and their corresponding EBs (or differentiated cells) demonstrate genomic imprinting according to methylation and expression analyses<sup>42,161,162</sup>. Several imprinted genes regulate fetal growth and development, sometimes in a tissue specific manner (for example: some genes are only imprinted in placenta)<sup>164</sup>. Bioinformatic analyses of all mouse protein-coding genes estimate that 2.5% (~600) of them could be imprinted<sup>160</sup>. Interestingly, the prediction model highlights correlation of imprinted genes with regions containing specific non protein-coding elements<sup>160</sup>. Notably, some human diseases such as Prader-Willi and Angelman syndromes are both associated with chromosomal deletions and imprinted gene dysregulation<sup>163,164</sup>.

From the work presented in Chapter 2, a bioinformatic analysis was performed to search for known and predicted imprinted genes<sup>160</sup> located inside the deleted fragments or 5 megabase pairs away from the largest deletion mapped in each family. The neighborhood of anchor sites without associated deletion was also screened (5 Mb according to the orientation of potential deletion). Currently, this limited sample size does not allow to correlate potential imprinted genes in regions related to abnormal phenotypes, nor a bias against recovering

deletions in known imprinted regions.

## 4.2 Pending optimizations

Some aspects of the presented project need further optimization, as discussed in this section. Among them, complementation is clearly a priority.

### 4.2.1 Complementation approaches

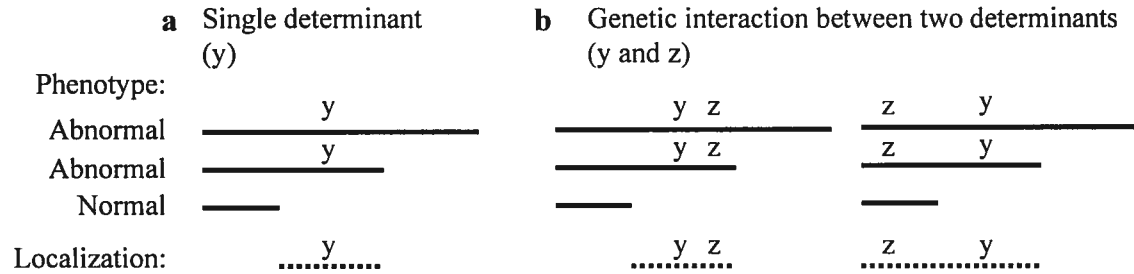
Although ESC clones seemingly possess euploid karyotypes according to SKY or aCGH analysis, they could also have mutations or epigenetic alterations undetectable with these assays. The formal proof that a deleted segment is causing an abnormal phenotype can only be obtained by a complementation study identifying the determinant(s) of interest.

#### 4.2.1.1 Identification of a minimal interval correlating with an abnormal phenotype.

According to an ideal scenario, a single determinant will cause a phenotypic anomaly (**Figure 4-1a**). In this case, nested deletions anchored to a specific site are very powerful to provide localization clues. The determinant will lie between the following endpoints: the largest deletion associated with normal phenotype and the smallest one associated with an abnormal phenotype (**Figure 4-1a**). However, oftentimes this scenario will likely be too simplistic because of a genetic synergy between two or several determinants located on the same chromosomal region. If two or more determinants genetically interact, they could be found anywhere between the anchor site and the end point of the smallest deletion associated with the abnormal phenotype (**Figure 4-1b**). The ratio between these two situations is unknown for the moment.

**Figure 4-1 Determination of a candidate region associated with an abnormal phenotype.**

Diagram representing deletion size and associated phenotype. Depending if a single (a) or a combination of interacting (b) determinants are responsible for the abnormal phenotype, the minimal localization need to be considered differently.



#### 4.2.1.2 Characterization of deleted segments

Since deletions are precisely localized with I-PCR and aCGH, several bioinformatic tools are available to characterize deleted segments. Public databases such as UCSC (<http://genome.ucsc.edu/>)<sup>32,152,153</sup> and Ensembl (<http://www.ensembl.org/index.html>)<sup>143</sup> Genome Browsers enclose information regarding genes, transcripts, microRNAs, CpG islands, BACs mapping, etc. SOURCE (<http://source.stanford.edu>)<sup>165</sup> is a database containing details such as aliases, Gene Ontology annotations, expression, and other information. Mouse tissue expression data are available at GNF SymAtlas (<http://symatlas.gnf.org/SymAtlas/>)<sup>166</sup>. Expression data related to ESC differentiation can be found in Stembase (<http://www.stembase.ca/>)<sup>167,168</sup>. Mouse Genome Informatics database (MGI, <http://www.informatics.jax.org/>)<sup>169</sup> presents descriptions of mutant alleles and phenotypes, and many other options. NCBI Mouse Genome Resources database (<http://www.ncbi.nlm.nih.gov/genome/guide/mouse/>)<sup>170</sup> allows comparative genome annotation (synteny), in addition to various features. Program such as Pathway Studio (<http://www.ariadnegenomics.com/products/pathway-studio/>)<sup>171</sup> can draw potential interactions between proteins according to literature survey. All these databases contain several more useful features and provide links to each other.

#### 4.2.1.3 Re-introduction of deleted DNA

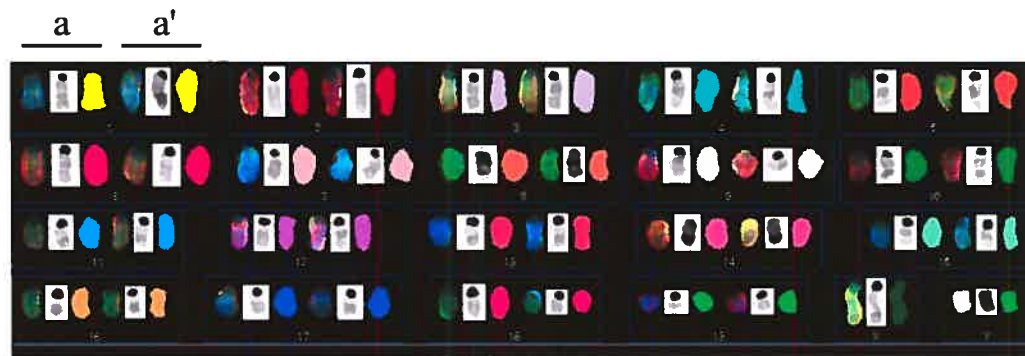
In Chapter 2, the differentiation assay was conducted with three seeding densities (2 log coverage). Every single clone (n=5) clearly presenting differentiation anomaly (dissagregation of EBs) demonstrated a few EBs differentiating normally when seeded



at high density. Southern blot analyses of these ESC clones and their corresponding rare EBs (**Supplementary Figure 2-7; 9-104, 9-18, 1-13, 13-24**) confirm that they were not contaminated by other clones. FISH analyses showed that rare EB cells from clone 9-104 reacquired at least a part of the deleted segment (**Figure 2-2c**), while keeping a normal karyotype for the mitotic cells detected by SKY (**Figure 4-2**). Interestingly, these cells could be further analyzed by aCGH and/or Q-PCR to characterize this reacquisition of genetic material correlating with the restored capacity to differentiate. These observations reveal two additional facts. First, differentiation defects are reversible, thus possibly amenable to complementation. Second, revertants will lead to complementation background and selected strategies need to be more efficient than the frequency of false-positive events. In other words, transduction of DNA must be very efficient and be achieved with vectors containing a selection marker gene.

**Figure 4-2 SKY analysis of rare EB cells derived from tertiary clone 9-104.**

ESC clone 9-104 contains an engineered deletion on chromosome 18. Rare EB mitotic cells (EBs day 8) present a haploid deletion on chromosome 18, but an otherwise normal karyotype (12/13 40, XY del (18q)). a and a' represent homologous chromosomes.



Currently, cDNA and BAC genomic DNA transduction with selectable vectors are the approaches envisioned to reintroduce deleted material. Both approaches have advantages and pitfalls, while complementing each other.

Transfection or retroviral transduction of cDNAs can be used to express known determinants. Annotated cDNAs libraries are commercially available, sometimes provided in expression vectors. Manipulation of these plasmids is simple. In addition, these libraries can be used globally to identify other determinants in the pathway(s) involved. However, potential

roles of non-expressed determinants are not addressed. There are additional limitations regarding *in vitro* differentiation for example. cDNA expression is frequently driven from a ubiquitous promoter, not necessarily reflecting the endogenous level, and without temporal or cell type specificity. To circumvent some of these limitations, transfection should be adjusted to cover a range of expression level and ideally driven from an inducible promoter. Different cDNA isoforms might also need to be tested.

BACs are advantageous to cover large regions (100-250 kb). In addition, they contain important regulatory elements including endogenous promoters directing the expression of alternative spliced transcripts. However, BAC modifications are not achievable with basic subcloning protocols and their transfection in cells is more challenging due to their large sizes. Nevertheless, different protocols exist to modify BAC for different purposes<sup>172</sup>. One important modification includes the addition of a selectable marker gene<sup>173</sup>. Moreover, a single or a few BACs can integrate in the genome following optimized transfection<sup>174</sup>, an important detail since phenotypic anomalies could be caused by abnormal gene dosage. BACs contain non-expressed sequences. However, their functionalities might not be observed following random integration. In fact, some ESC clone will probably be complemented only if the missing fragment is placed back in the original location. Fortunately, both BACs and tertiary clones are provided with a *loxP* site. The functional reconstitution of a *Pgk-ATG-loxP* found in tertiary clone with a *loxP-ATG-less-puromycin* gene present in a BAC vector could be Cre-mediated specifically in the original locus. A *loxP-ATG-less-puromycin* cassette compatible with the *Pgk-ATG-loxP* is available in the laboratory. However, this procedure will require a higher degree of optimization than a simple transfection.

In summary, sequences related to gene-poor regions will be reintroduced with BACs. Sequences related to gene rich regions can be reintroduced using either cDNAs or BACs. BACs are favorable because they are similar to the endogenous organization and regulation. However, the simplicity of cDNA approaches is valuable.

#### **4.2.1.4 Mapped regions correlating with differentiation anomalies**

Fully or partially mapped regions associated with differentiation anomalies were presented in Chapter 2. Initial studies revealed three genomic regions associated with ESC differentiation anomalies. Deletions on chromosome 4 pertaining to clones in family no.13 are partially mapped. One clone (13-24) presents an aberrant differentiation phenotype and is expected to contain a deletion larger than 3.9 megabase pairs (larger than the only deletion

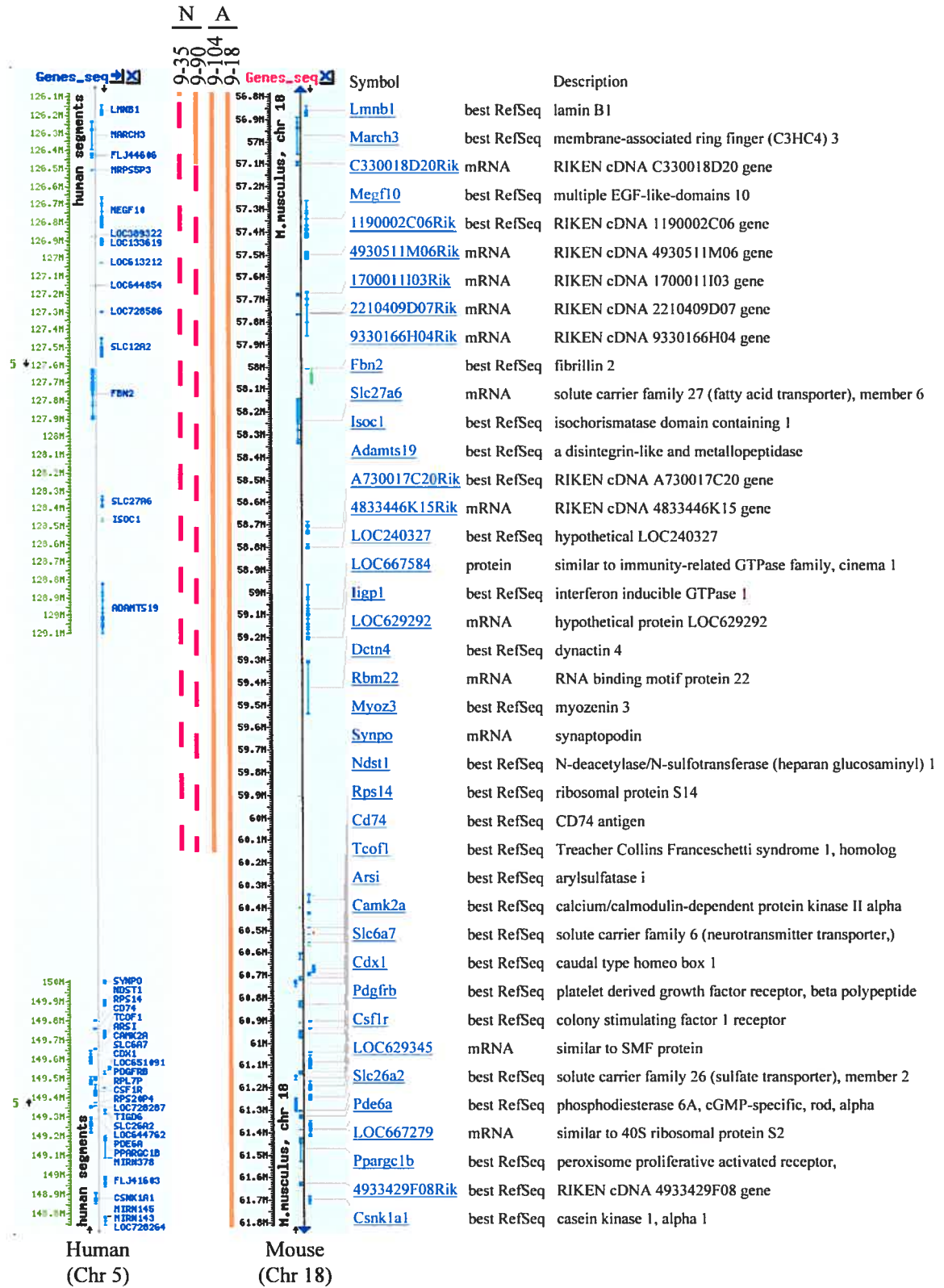
mapped, associated with relatively normal differentiation) (**Table III** and **Supplementary Table VI**). Deletions related to family no.1 were mapped on chromosome 14 (**Table III** and **Supplementary Table VI**). However, the clones with a smaller genomic deletion (1.5 megabase pairs, n=4 clones) differentiated normally while the one with the largest deletion (22.7 megabase pairs) did not (**Table III** and **Supplementary Table VI**). Consequently, the minimal interval is very broad. Finally, family no.9 related deletions, located on chromosome 18, were characterized to a greater extent. As presented previously, 3 independent clones with deletions (4.1-5.0 Mb) correlated with differentiation anomalies while the smallest deletions (6-317 kb, n=8 independent clones) did not (**Supplementary Table VI**). According to this, the confidence that an important determinant(s) resides inside the deleted interval rather than outside (bystander genetic or epigenetic alterations) is improved. Family no.9 will be used to illustrate a complementation approach.

#### 4.2.1.5 Characterization of deletions related to family no.9

In family no.9, if loss of a single determinant is causing the abnormal differentiation phenotype, the minimal interval in which this determinant is located is between deletion endpoints in clones 9-90 (largest deletion with a normal differentiation phenotype) and 9-104 (smallest deletion with an abnormal differentiation phenotype) (**Figure 4-3**). However, since clone 9-90 presents a chromosome 8 trisomy (**Table III**), a compensation for the deletion on chromosome 18 cannot be excluded. Therefore, a more conservative interval would be between the deletion endpoints in clones 9-35 and 9-104 (**Figure 4-3**). If two or more determinants are causing the phenotypic anomaly, they would be located anywhere between the anchor site and the endpoint of the smallest deletion associated with the abnormal phenotype. In this case, it would correspond to the size of the deletion found in clone 9-104 (**Figure 4-3**).

#### **Figure 4-3 Minimal intervals represented for family no.9.**

Representation adapted from NCBI Mouse Genome Resources (<http://www.ncbi.nlm.nih.gov/genome/guide/mouse/>)<sup>170</sup>. Mapping of deletions (orange lines) on mouse chromosome 18 aligned with the corresponding human syntenic chromosome 5 for tertiary clones 9-35, 9-90, 9-104, and 9-18. Pink dashed lines represent minimal intervals if a single determinant is involved (see text for details). However, if two or more determinants are involved, they can be found anywhere in the region covered by the clone 9-104 deletion. N, normal differentiation; A, abnormal differentiation; Chr, chromosome.



The abnormal phenotype monitored within this family occurs during differentiation. Clones with large deletions (9-104, 9-37, 9-18) initiate a visually normal differentiation up to day 3. Following day 3-4, EBs disaggregate rapidly. When re-introduced in blastocyst, the clone containing the largest deletion (9-18) failed to contribute to the tissues examined in chimeric fetuses (E14.5) and in adult mice (**Figure 2-2e, Table IV**). ESC clones 9-104 and 9-37 also presented an *in vitro* differentiation defect but were not chosen for chimera generation. An unstable karyotype was initially detected for clone 9-104 and the Y chromosome was lost in 9-37 (**Table III**). Since the deletion in clone 9-18 is larger than the two others, this additional deleted portion could also contribute both to the *in vivo* and the *in vitro* phenotypic anomalies. Nevertheless, both *in vitro* and *in vivo* data suggest an early differentiation defect.

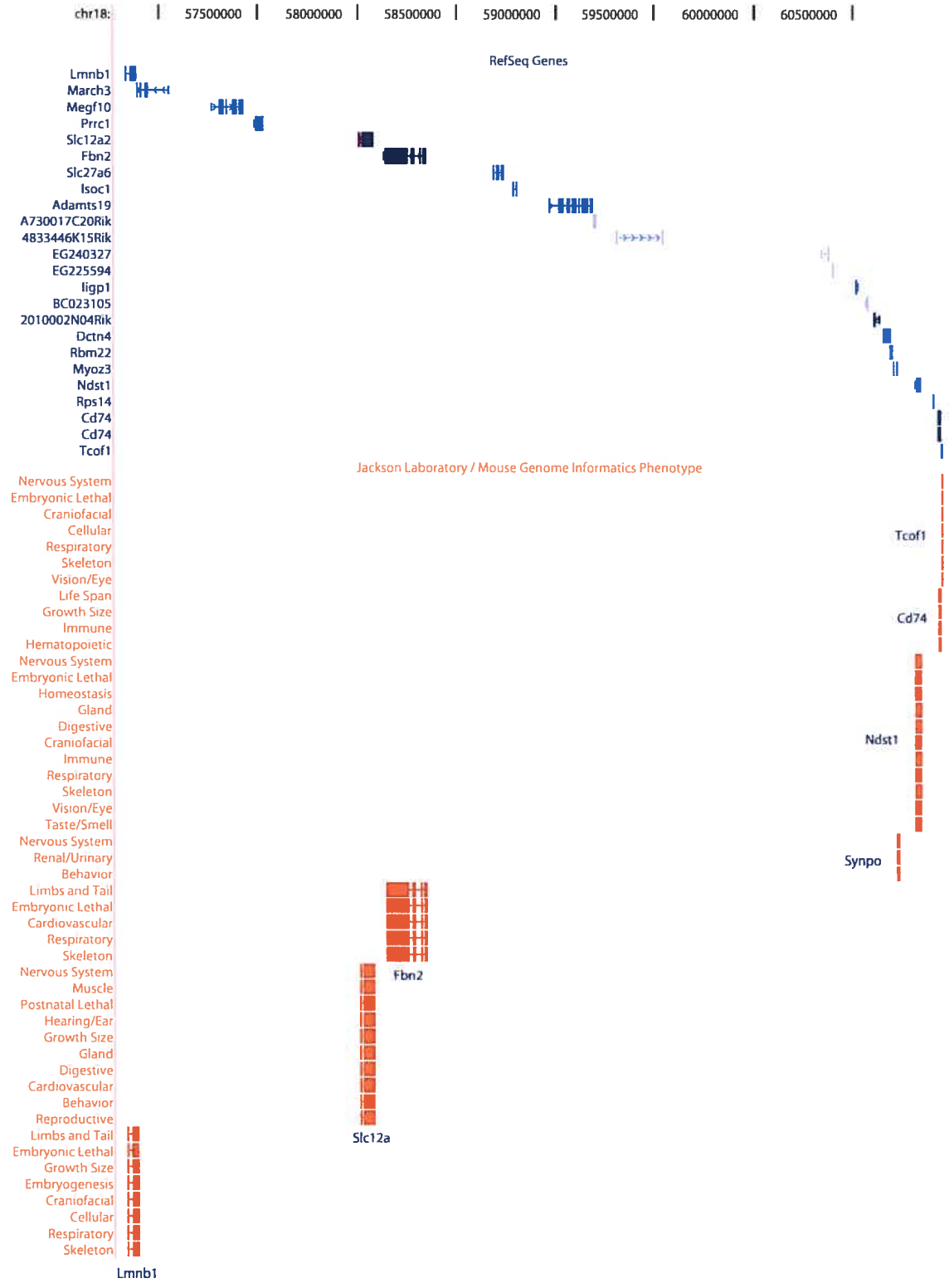
Assuming that the deleted determinant(s) causing the early differentiation defect is found in the minimal region corresponding to the deletion in clone 9-104 (4.1 megabase pairs), various analyses can be performed to characterize potential candidates.

Previously described mouse mutant alleles and their corresponding phenotypes found in this minimal region are shown in **Figure 4-4**. Interestingly, *Tcofl* is haploinsufficient; heterozygous embryos die of severe craniofacial defects (detected from E8)<sup>175</sup>. **Table X** presents known RefSeq genes found in the minimal interval and their associated functions. Twelve of them are expressed in blastocyst stage and early embryogenesis (E6.5-8.5) according to microarray analyses reported in GNF SymAtlas (<http://symatlas.gnf.org/SymAtlas/>)<sup>166</sup> (**Table X**). Interestingly, RIKEN cDNA 4833446K15 (recently annotated as a RefSeq gene, corresponding to mRNA AK019528) potentially encodes a mouse homologue of human CSS3 (chondroitin sulfate synthase 3). Gene Ontology annotations for CSS3 are N-acetylgalactosaminyl-proteoglycan 3-b-glucuronosyltransferase activity, glucuronyl-N-acetylgalactosaminylproteoglycan 4-beta-N-acetylgalactosaminyltransferase, and transmembrane localization in Golgi apparatus (<http://source.stanford.edu>)<sup>165</sup>. Importantly, this predicted protein seems involved in a process similar to the expressed *Ndst1*, also found in the deleted interval (**Table X**). Both are found in the Golgi apparatus membrane. These two determinants, *Ndst1* and the AK019528 predicted protein, could functionally interact to sustain a biological process. Synpo (Synaptopodin, AK034012) is expressed in the blastocyst stage and during early embryogenesis (E6.5-8.5) (<http://symatlas.gnf.org/SymAtlas/>)<sup>166</sup>, and is associated with mouse phenotypes (neuronal, behavior), but not annotated as a RefSeq gene (**Figure 4-4**). No microRNAs are detected in this region (<http://genome.ucsc.edu/>). Only *March3* is predicted to be imprinted<sup>160</sup>, but this gene is not expressed in early embryogenesis

according to GNF SymAtlas (**Table X**). Note that Stembase (<http://www.stembase.ca/>)<sup>167,168</sup> might provide pertinent gene expression profiling in differentiating EBs (registration to Stembase in process).

**Figure 4-4 Mouse mutant alleles and mapped phenotypes for family no. 9 minimal interval.**

Representation of RefSeq genes and corresponding mouse mutant alleles and phenotypes (Jackson Laboratory/Mouse Genome Informatics)<sup>169</sup> annotated in the UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32</sup>, corresponding to the minimal interval correlating with an abnormal differentiation phenotype for clones of family no.9 (region deleted in clone 9-104).



**Table X • Function and expression of RefSeq genes found in family no.9 minimal interval.**

**Part 1 of 3**

Name (symbol)	Gene Ontology Annotations	Functions
Lamin B1 (Lmnb1)	intermediate filament lamin filament nucleu s structural molecule activity	lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane, which is thought to provide a framework for the nuclear envelope and may also interact with chromatin.
Membrane-associated ring finger (C3HC4) 3 (March3)	integral to membrane protein ubiquitination ubiquitin ligase complex ubiquitin-protein ligase activity zinc ion binding	
Multiple EGF-like-domains 10 (Megf10)		plays important roles in both rhabdomere development and in photoreceptor cell survival. might function as a calcium- sequestering «sponge» to regulate the amount of free cytoplasmic calcium. it binds 0.3 mole of ca(2+) per mole of protein.
Proline-rich coiled-coil 1 (Prrc1)		
Solute carrier family 12, member 2 (Slc12a2)	amino acid transport amino acid-polyamine transporter activity basolateral plasma membrane carrier activity cation: chloride symporter activity chloride transport integral to membrane integral to plasma membrane ion transport membrane potassium ion transport sodium ion transport symporter activity transport transporter activity	electrically silent transporter system. mediates sodium and chloride reabsorption. plays a vital role in the regulation of ionic balance and cell volume.
Fibrillin 2 (Fbn2)	calcium ion binding embryonic limb morphogenesis extracellular matrix (sensu Metazoa) extracellular region extracellular space limb morphogenesis microfibril	fibrillins are structural components of 10-12 nm extracellular calcium-binding microfibrils, which occur either in association with elastin or in elastin-free bundles. fibrillin-2- containing microfibrils regulate the early process of elastic fiber assembly.

Name, symbol, Gene Ontology annotations and functions were determined using SOURCE database (<http://source.stanford.edu>)<sup>165</sup> for RefSeq genes mapped in the UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>153</sup>. Expression of these RefSeq genes in mouse balstocysts and embryos E6.5-8.5 was found in GNF SymAtlas (<http://symatlas.gnf.org/SymAtlas/>)<sup>166</sup>. Blue, expressed; no color, not expressed; green, not determined.



## Part 2 of 3

Name (symbol)	Gene Ontology Annotations	Functions
Solute carrier family 27 (fatty acid transporter), member 6 (Slc27a6)		acyl-coa synthetase probably involved in bile acid metabolism. proposed to activate c27 precursors of bile acids to their coa thioesters derivatives before side chain cleavage via peroxisomal beta-oxidation occurs. in vitro, activates 3-alpha,7- alpha,12-alpha-trihydroxy-5-beta-cholestanate (thca), the c27 precursor of cholic acid deriving from the de novo synthesis from cholesterol. does not utilize c24 bile acids as substrates. in vitro, also activates long- and branched-chain fatty acids and may have additional roles in fatty acid metabolism (by similarity). may be involved in translocation of long-chain fatty acids (lfca) across membranes.
Isochorismatase domain containing 1 (Isoc1)	catalytic activity metabolism	
A disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 19 (Adamts19)	extracellular matrix extracellular matrix (sensu Metazoa) hydrolase activity metallopeptidase activity metallopeptidase activity peptidase activity proteolysis and peptidolysis zinc ion binding	cleaves aggrecan, a cartilage proteoglycan, and may be involved in its turnover. has angiogenic inhibitor activity (by similarity). active metalloprotease, which may be associated with various inflammatory processes as well as development of cancer cachexia. may play a critical role in follicular rupture (by similarity).
RIKEN cDNA A730017C20 gene (A730017C20Rik)	integral to membrane	
RIKEN cDNA 4833446K15 gene (4833446K15Rik)	integral to membrane	
Predicted gene, EG240327 (EG240327)		
Similar to CDNA sequence BC023105 (LOC225594)		
Interferon inducible GTPase 1 (ligp1)	GDP binding GTP binding GTPase activity GTPase activity cellular component unknown cytokine and chemokine mediated signaling pathway protein self binding	
CDNA sequence BC023105 (BC023105)		
Name, symbol, Gene Ontology annotations and functions were determined using SOURCE database ( <a href="http://source.stanford.edu">http://source.stanford.edu</a> ) <sup>165</sup> for RefSeq genes mapped in the UCSC Genome Browser ( <a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a> ) <sup>153</sup> . Expression of these RefSeq genes in mouse balstocysts and embryos E6.5-8.5 was found in GNF SymAtlas ( <a href="http://symatlas.gnf.org/SymAtlas/">http://symatlas.gnf.org/SymAtlas/</a> ) <sup>166</sup> . Blue, expressed; no color, not expressed; green, not determined.		

## Part 3 of 3

Name (symbol)	Gene Ontology Annotations	Functions
RIKEN cDNA 2010002N04 gene (2010002N04Rik)	integral to membrane	
Dynactin 4 (Dctn4)	cytoplasmic dynein complex cytoskeleton protein binding	
RNA binding motif protein 22 (Rbm22)	RNA binding mRNA processing nuclear mRNA splicing, via spliceosome nucleic acid binding	involved in pre-mrna splicing. facilitates the cooperative formation of u2/u6 helix ii in association with stem ii in the spliceosome. binds to rna.
Myozenin 3 (Myoz3)	Z disc protein binding	
N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 1 (Ndst1)	Golgi apparatus [heparan sulfate]-glucosamine N-sulfotransferase activity cysteine-type endopeptidase activity integral to membrane organogenesis polysaccharide biosynthesis protein amino acid deacetylation protein amino acid sulfation protein amino acid sulfation proteolysis and peptidolysis respiratory gaseous exchange sulfotransferase activity transferase activity	essential bifunctional enzyme that catalyzes both the n- deacetylation and the n-sulfation of glucosamine (glcnac) of the glycosaminoglycan in heparan sulfate. modifies the glcnac-glca disaccharide repeating sugar backbone to make n-sulfated heparosan, a prerequisite substrate for later modifications in heparin biosynthesis. plays a role in determining the extent and pattern of sulfation of heparan sulfate. compared to other ndst enzymes, its presence is absolutely required. participates in biosynthesis of heparan sulfate that can ultimately serve as l- selectin ligands, thereby playing a role in inflammatory response.
Ribosomal protein S14 (Rps14)	RNA binding cytosolic ribosome (sensu Eukaryota) cytosolic small ribosomal subunit (sensu Eukaryota) intracellular protein biosynthesis ribonucleoprotein complex ribosome ribosome biogenesis structural constituent of ribosome	plays a critical role in mhc class ii antigen processing by stabilizing peptide-free class ii alpha/beta heterodimers in a complex soon after their synthesis and directing transport of the complex from the endoplasmic reticulum to compartments where peptide loading of class ii takes place.
Treacher Collins Franceschetti syndrome 1, homolog (Tcof1)	nucleolus nucleus transcription of nuclear rRNA large RNA polymerase I transcript transport	may be involved in nucleolar-cytoplasmic transport. may play a fundamental role in early embryonic development, particularly in development of the craniofacial complex (by similarity).

Name, symbol, Gene Ontology annotations and functions were determined using SOURCE database (<http://source.stanford.edu>)<sup>165</sup> for RefSeq genes mapped in the UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>153</sup>. Expression of these RefSeq genes in mouse blastocysts and embryos E6.5-8.5 was found in GNF SymAtlas (<http://symatlas.gnf.org/SymAtlas/>)<sup>166</sup>. Blue, expressed; no color, not expressed; green, not determined.

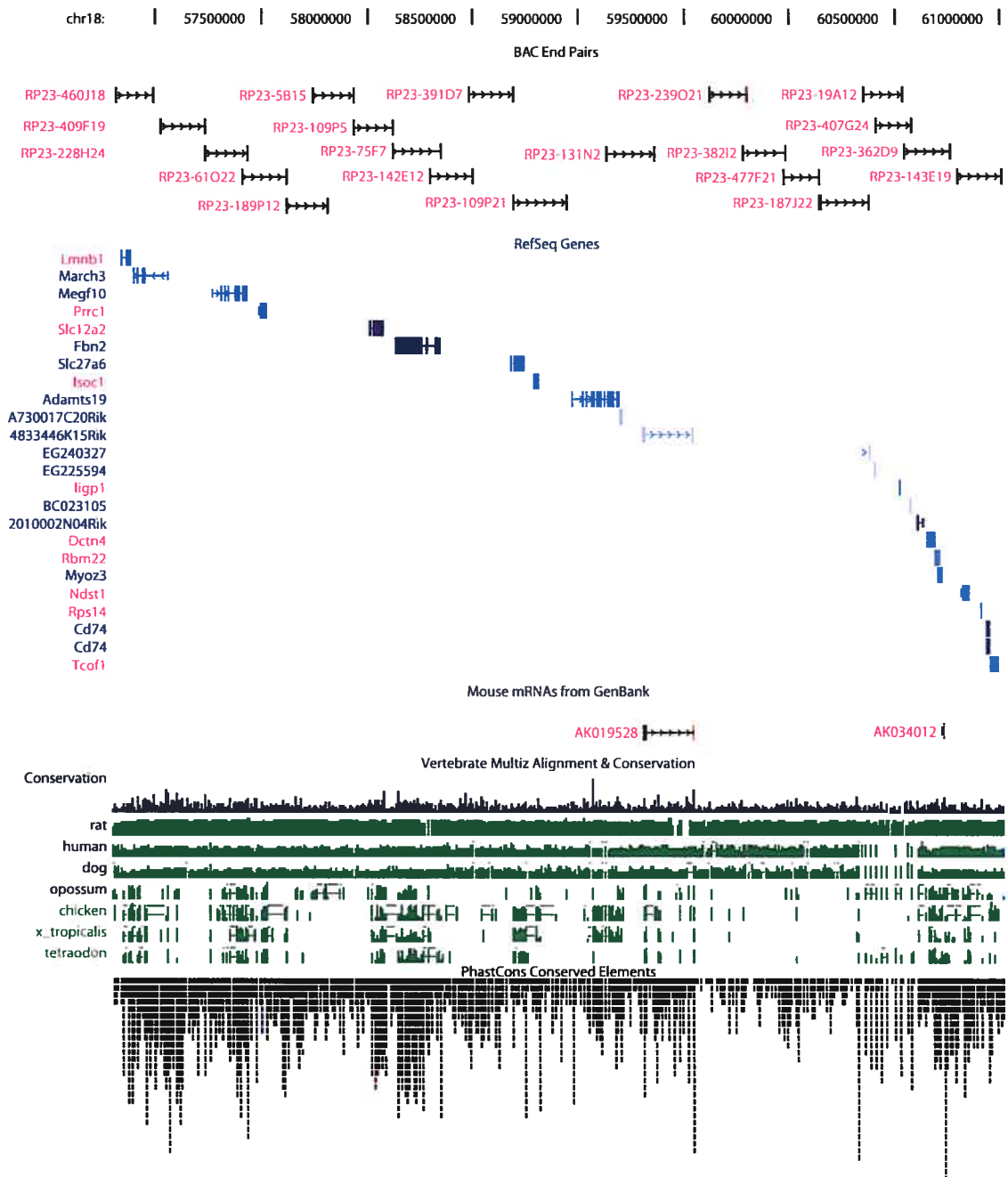
From these analyses, BACs and/or cDNAs corresponding to the expressed transcripts were purchased (**Figure 4-5**). Additional BACs covering intergenic regions were ordered (**Figure 4-5**). On a technical point of view, the first phase of complementation will be done *in vitro* with clones 9-104, 9-37, and 9-18. Clone 9-104 seems an obvious choice for complementation because it contains the smallest deletion associated with an abnormal phenotype. However, clone 9-37 and 9-18 karyotypes are more stable. Initially, expressed genes will be individually reintroduced either with BACs or cDNAs, and the combination *NDST1*/AK019528 will be tried. In the case where no single determinant (or the *NDST1*/AK019528 combination) can rescue the abnormal phenotype, combinations of expressed gene will be attempted. Subsequently, genes with undetermined expression or classified as non-expressing (according to the criteria used) will be reintroduced individually or in combination with expressed determinants. Or else, introduction of BACs in the original locus (*cis*-complementation) will be optimized to assess the role of non-transcribed regions.

Some regions are expected to be harder to complement than others, particularly if they are large and are containing numerous deleted determinants that could genetically interact or if they require *cis*-complementation, etc. Identification of minimal regions from the work presented in Chapter 3 will likely provide diverse complementation complexity levels.

Currently, plasmids containing BAC or cDNA are being engineered to carry a selection marker gene (puromycin, absent from clones containing deletions). Other complementation approaches were previously attempted for the family 9. Co-transfection of linearized BACs along with a selection marker gene (*Pgk*-puromycin) at a 3:1 ratio (fentomoles ratio) achieved BAC transfer in roughly half of the transfected ESC clonal populations (PCR detection of BAC plasmid). In some instances, isolated ESC subclones transfected with a BAC seemed rescued from the differentiation defect. However, FISH experiments revealed that their nucleus did not contain the selected BAC. Consequently, no convincing complementation during *in vitro* differentiation could be observed at a frequency higher than the frequency of natural revertants. Thus it was concluded that complementation approaches needed to be more efficient than the frequency of natural revertants and should be carried out by keeping cells in a polyclonal population following transfection and selection. Microcell-mediated chromosome transfer<sup>176</sup> was also considered to transfer an anchored parental chromosome in cells containing a related deletion (tertiary clone cells), but discarded since the revertants frequency would be higher than the chromosome transfer frequency.

**Figure 4-5 BACs and cDNAs selected to cover clone 9-104 deleted segment.**

cDNAs (red) corresponding to expressed determinants and BACs (RP23-, red) were purchased to cover the deleted region. Human cDNAs were purchased for *SLC12A2* and *NDST1*. High conservation score is observed for protein-coding elements and some segments outside of genes (possibly regulatory elements and other genomic features). *X\_tropicalis*; *Xenopus tropicalis*. Adapted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32</sup>.



## 4.2.2 Toward a recessive screen

Although some haploid deletions correlate with phenotypic anomalies during ESC differentiation (Chapter 2: 11% of our sample size), a large proportion does not demonstrate any effect. Exploitation of tertiary clones could be maximized by obtaining homozygous deletions. In the system presented in this thesis, the functional reconstitution of a neomycin gene (amplified from the pPNT template<sup>148</sup>, based in part on the neomycin gene derived from the bacterial transposon Tn5 found in pMCI Neo<sup>177</sup>) was chosen because it was previously employed to generate homozygous mutant ESCs from heterozygous targeted cells grown under high G418 concentration. Since this phenomenon involves extensive loss of heterozygosity<sup>117</sup>, this approach is attractive to potentially generate homozygous deletions *in vitro*. Likewise, the reconstituted *Pgk-ATG-loxP-neo* might be amenable to this strategy, but we have not attempted this experiment yet. Nevertheless, 6 candidate regions were selected from the work presented in Chapter 2 to attempt this loss of heterozygosity (Table XI). Small deletions will likely be less detrimental to ESCs, considering that the proportion of homozygous cell lethal genes is currently unknown. Therefore, three deletions spanning less than a megabase pairs were selected (ESC clones 9-35, 7-30, and 14-16, Table XI). Various chromosomal locations were chosen because the mechanism behind the loss of heterozygosity is not well known (Table XI). Finally, deletion-containing ESC clones with a normal karyotype were preferred (Table XI). Although issues such as imprinting might complicate characterization, proving that this approach can work in our system is very appealing.

**Table XI • Candidate haploid deletions that could be tested for loss of heterozygosity.**

Clone identification	Chromosome	Deletion size (kb)	Number of genes	Karyotype
9-35	18	23	0	N
7-30	16	94	3	N
14-16	2	568	9	N
4-02	2	1400	19	N.D.
1-03	14	1500	17	N
9-18	18	5000	32	N

Kb, kilobase pairs; N, normal; N.D., not determined.

### 4.2.3 Detection of ESC-derived progenies *in situ*

Ideally, ESCs used in the generation of chimeric mice should contain a marker that would allow the tracking of ESC-derived progenies. This marker should have the following characteristics: neutral for the cells, ubiquitously expressed, cell autonomous, and detectable *in situ* at the single cell level<sup>57</sup>. In combination with a fine tracking system, possibilities to generate chimeric or F0 mice (see Chapter 1) are complementary and allow the observation of various biological process. A few examples are mentioned below. For more, readers are invited to consult a recently published review (Tam, P.P. & Rossant, J., 2003)<sup>57</sup>. For example, chimeric mice could highlight the cell intrinsic properties of mutant ESCs giving rise to particular phenotypes in specific lineages and/or their exclusion from specific tissues<sup>57</sup>. Cell extrinsic properties could be suspected when abnormal phenotypes are observed in cells that are not ESC-derived<sup>57</sup>. Tetraploid complementation assay with normal embryo generates F0 mutant mice that sometimes could not be obtained following normal breeding, because of extraembryonic defects (trophoblast lineage or extraembryonic endoderm)<sup>57</sup>.

Chromosomal deletions were generated in non-labeled wild-type R1 ESCs<sup>61</sup> (for example, EGFP marker can be ubiquitously expressed in ESCs<sup>178</sup>). However following recombination, tertiary clones express neomycin which can be used as a marker. Accordingly, several anti-neomycin antibodies are commercially available. Neomycin detection by indirect immunofluorescence is possible on mouse paraffin sections<sup>179</sup>. For the purpose of *in vivo* experiments, it will be beneficial to test this detection *in vitro* and *in vivo*, to see if neomycin expression is detectable and ubiquitous, or if it is suppressed during ESC differentiation. However, even if it is not expressed, ESC-derived progenies contain this neomycin tag in their genomic DNA. Very sensitive *in situ* hybridization of low copy virus was achieved in paraffin sections using biotin-labeled cDNA probes, streptavidin-Nanogold, and silver acetate autometallography (localized and precise black staining)<sup>180</sup>. Finally, engineered ESCs could be injected in diploid ROSA26 mouse blastocysts (C57BL/6J background) expressing  $\beta$ -galactosidase ubiquitously<sup>181</sup>, in order to distinguish them (unlabeled) from host cells (labeled). ROSA26 mouse strain is also available in 129Sv background<sup>181</sup>, if R1 ESCs (129/Sv x 129/Sv-CP)<sup>61</sup> are used for teratocarcinoma formation.

### 4.3 Potential applications of the system

Delivering *loxP* sites with two complementary retroviruses could allow the genetic manipulation of cells that are not suited for gene targeting. Although Cre was transiently transfected by electroporation in ESCs, other delivery systems are available for sensitive cells. Retroviral gene transfer, using a Cre-encoded self-deleting retrovirus<sup>182,183</sup>, is an alternative, as well as the Cre protein fused to a membrane translocation sequence<sup>184</sup> (cell permeable). Several applications, different than those presented in this thesis, can be envisioned, either using ESCs or other cell lines. Basically, two components need to be assembled: a cell line permissive to retroviral infection and drug selection (Chapter 1, part III), and a screening methodology (**Table XII**). MSCV-based gene transfer is achievable for both mouse and human cells, provided that virions are encapsidated in appropriate packaging cell lines (tropism). Finally, although this methodology is suited for screening, it can simply be used to modify genomic regions in particular cells to create experimental models. Like it was mentioned in the previous section, if cells are reintroduced *in vivo*, they need to be distinguished from host cells.

**Table XII • Potential applications of retroviral-based Cre-loxP recombination.**

Cell lines	Screening methodologies or experimental models
Mouse ESCs	Teratomas-teratocarcinomas formation with engineered ESCs injected subcutaneously in syngenic mice (observe characteristics, aggressiveness, etc.). Look for genomic regions enhancing or suppressing tumor growth.
Mouse ESCs or somatic stem cells	Chimeric mice formation with engineered ESCs or somatic stem cells. Find genomic segments regulating differentiation and/or enhancing tumor formation (example: loss of tumor suppressor).
Cancer cell lines, metastatic	Re-introduction of engineered cancer cells in animal models (for example: breast cancer cell lines). Find genomic segments regulating aggressiveness, metastatic properties, etc.
Human and mouse ESCs	<i>In vitro</i> model (also <i>in vivo</i> in case of mouse ESCs) of human diseases associated with chromosomal anomalies <sup>163</sup> : deletions (example: Prader-Willi syndrome), translocations (some leukemias), or other aberrations. These models can also be used in screens (example: chemical compounds) or complementation.
Mouse ESCs	Observation of recombination frequencies (for example: intermolecular versus intramolecular) for different anchored regions (middle of chromosome, more telomeric or centromeric regions, etc.).

#### 4.4 Thesis conclusion

This thesis described the elaboration of a new functional genomic tool based on Cre-loxP recombination. Nested chromosomal deletions in mouse ESCs were obtained for various loci. This material will benefit the scientific community interested in stem cells, developmental biology, and tumorigenesis. Functional annotation of genes, non-protein coding transcripts, and non-expressed elements will be feasible with this methodology. Genetic interactions between contiguous elements acting cooperatively to sustain biological functions will be uncovered. The versatility of this system is a major advantage. Potential applications in other cell lines abound because the approach relies on retroviral gene transfer.



## REFERENCES (FOR CHAPTERS AND APPENDIXES)

1. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
2. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
3. Waterston, R.H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).
4. Lunter, G., Ponting, C.P. & Hein, J. Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput Biol* **2**, e5 (2006).
5. Lindblad-Toh, K. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-19 (2005).
6. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669-81 (2007).
7. Eisenstein, M. Uncovering hidden relationships. *Nat Methods* **2**, 806 (2005).
8. Ren, S.Y., Angrand, P.O. & Rijli, F.M. Targeted insertion results in a rhombomere 2-specific *Hoxa2* knockdown and ectopic activation of *Hoxa1* expression. *Dev Dyn* **225**, 305-15 (2002).
9. Adams, D.J. et al. Mutagenic insertion and chromosome engineering resource (MICER). *Nat Genet* **36**, 867-71 (2004).
10. Glaser, S., Anastassiadis, K. & Stewart, A.F. Current issues in mouse genome engineering. *Nat Genet* **37**, 1187-93 (2005).
11. Collins, F.S., Rossant, J. & Wurst, W. A mouse for all reasons. *Cell* **128**, 9-13 (2007).
12. Collins, F.S., Finnell, R.H., Rossant, J. & Wurst, W. A new partner for the international knockout mouse consortium. *Cell* **129**, 235 (2007).
13. Schnutgen, F. et al. Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* **102**, 7221-6 (2005).
14. Nord, A.S. et al. The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res* **34**, D642-8 (2006).
15. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8 (2006).
16. Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S. & Sabatini, D.M. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods* **3**, 715-9 (2006).
17. Bailey, S.N., Ali, S.M., Carpenter, A.E., Higgins, C.O. & Sabatini, D.M. Microarrays of lentiviruses for gene function screens in immortalized and primary cells. *Nat Methods* **3**, 117-22 (2006).
18. Kanellopoulou, C. et al. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**, 489-501 (2005).
19. Uren, A.G., Kool, J., Berns, A. & van Lohuizen, M. Retroviral insertional mutagenesis: past, present and future. *Oncogene* **24**, 7656-72 (2005).
20. Chen, Y. et al. Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat Genet* **24**, 314-7 (2000).

21. Munroe, R.J. et al. Mouse mutants from chemically mutagenized embryonic stem cells. *Nat Genet* **24**, 318-21 (2000).
22. Nolan, P.M. et al. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* **25**, 440-3 (2000).
23. Hrabe de Angelis, M.H. et al. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* **25**, 444-7 (2000).
24. Herron, B.J. et al. Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. *Nat Genet* **30**, 185-9 (2002).
25. Balling, R. ENU mutagenesis: analyzing gene function in mice. *Annu Rev Genomics Hum Genet* **2**, 463-92 (2001).
26. Yusa, K. et al. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**, 896-9 (2004).
27. Vivian, J.L., Chen, Y., Yee, D., Schneider, E. & Magnuson, T. An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **99**, 15542-7 (2002).
28. Chen, S. et al. Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A* **103**, 17266-71 (2006).
29. Aarts, M., Dekker, M., de Vries, S., van der Wal, A. & te Riele, H. Generation of a mouse mutant by oligonucleotide-mediated gene modification in ES cells. *Nucleic Acids Res* **34**, e147 (2006).
30. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55 (2003).
31. 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**, 6946-51 (2006).
32. Karolchik, D. et al. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**, 51-4 (2003).
33. Goodwin, N.C. et al. DelBank: a mouse ES-cell resource for generating deletions. *Nat. Genet.* **28**, 310-311 (2001).
34. Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720-724 (1995).
35. Brault, V., Pereira, P., Duchon, A. & Herault, Y. Modeling chromosomes in mouse to explore the function of genes, genomic disorders, and chromosomal organization. *PLoS Genet* **2**, e86 (2006).
36. Su, H., Wang, X. & Bradley, A. Nested chromosomal deletions induced with retroviral vectors in mice. *Nat. Genet.* **24**, 92-95 (2000).
37. LePage, D.F., Church, D.M., Millie, E., Hassold, T.J. & Conlon, R.A. Rapid generation of nested chromosomal deletions on mouse chromosome 2. *Proc. Natl. Acad. Sci. U.S.A* **97**, 10471-10476 (2000).
38. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6 (1981).
39. 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**, 7634-8 (1981).

40. Bradley, A., Evans, M., Kaufman, M.H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-6 (1984).
41. Smith, A.G. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435-62 (2001).
42. Zvetkova, I. et al. Global hypomethylation of the genome in XX embryonic stem cells. *Nat Genet* **37**, 1274-9 (2005).
43. Zwaka, T.P. & Thomson, J.A. A germ cell origin of embryonic stem cells? *Development* **132**, 227-33 (2005).
44. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55 (2005).
45. Ralston, A. & Rossant, J. Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet* **68**, 106-12 (2005).
46. Beddington, R.S. & Robertson, E.J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733-7 (1989).
47. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150-60 (2004).
48. Spivakov, M. & Fisher, A.G. Epigenetic signatures of stem-cell identity. *Nat Rev Genet* **8**, 263-71 (2007).
49. Boyer, L.A., Mathur, D. & Jaenisch, R. Molecular control of pluripotency. *Curr Opin Genet Dev* (2006).
50. 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**, 372-6 (2000).
51. Niwa, H. et al. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-29 (2005).
52. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).
53. Galan-Cardiad, J.M. et al. Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* **129**, 345-57 (2007).
54. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323**, 445-8 (1986).
55. Wang, Z. & Jaenisch, R. At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev Biol* **275**, 192-201 (2004).
56. Wood, S.A., Allen, N.D., Rossant, J., Auerbach, A. & Nagy, A. Non-injection methods for the production of embryonic stem cell-embryo chimaeras. *Nature* **365**, 87-9 (1993).
57. Tam, P.P. & Rossant, J. Mouse embryonic chimeras: tools for studying mammalian development. *Development* **130**, 6155-63 (2003).
58. Eggan, K. et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A* **98**, 6209-14 (2001).

59. Eakin, G.S., Hadjantonakis, A.K., Papaioannou, V.E. & Behringer, R.R. Developmental potential and behavior of tetraploid cells in the mouse embryo. *Dev Biol* **288**, 150-9 (2005).
60. Poueymirou, W.T. et al. F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat Biotechnol* **25**, 91-9 (2007).
61. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A* **90**, 8424-8428 (1993).
62. Egan, K. et al. Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. *Nat Biotechnol* **20**, 455-9 (2002).
63. Nakano, T., Kodama, H. & Honjo, T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098-101 (1994).
64. Maye, P., Becker, S., Kasameyer, E., Byrd, N. & Gabel, L. Indian hedgehog signaling in extraembryonic endoderm and ectoderm differentiation in ES embryoid bodies. *Mech Dev* **94**, 117-32 (2000).
65. Terada, N. et al. Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416**, 542-5 (2002).
66. Gadue, P., Huber, T.L., Nostro, M.C., Kattman, S. & Keller, G.M. Germ layer induction from embryonic stem cells. *Exp Hematol* **33**, 955-64 (2005).
67. Li, H., Roblin, G., Liu, H. & Heller, S. Generation of hair cells by stepwise differentiation of embryonic stem cells. *Proc Natl Acad Sci U S A* **100**, 13495-500 (2003).
68. Yamane, T., Hayashi, S., Mizoguchi, M., Yamazaki, H. & Kunisada, T. Derivation of melanocytes from embryonic stem cells in culture. *Dev Dyn* **216**, 450-8 (1999).
69. Lengerke, C. & Daley, G.Q. Patterning definitive hematopoietic stem cells from embryonic stem cells. *Exp Hematol* **33**, 971-9 (2005).
70. Takahashi, K., Mitsui, K. & Yamanaka, S. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**, 541-5 (2003).
71. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P.P. Pten is essential for embryonic development and tumour suppression. *Nat Genet* **19**, 348-55 (1998).
72. Yang, J.T., Rayburn, H. & Hynes, R.O. Embryonic mesodermal defects in alpha 5 integrin-deficient mice. *Development* **119**, 1093-105 (1993).
73. Taverna, D. & Hynes, R.O. Reduced blood vessel formation and tumor growth in alpha5-integrin-negative teratocarcinomas and embryoid bodies. *Cancer Res* **61**, 5255-61 (2001).
74. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**, 380-5 (2007).
75. Kaji, K. et al. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* **8**, 285-92 (2006).
76. Malumbres, M. & Barbacid, M. To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer* **1**, 222-31 (2001).

77. Savatier, P., Huang, S., Szekely, L., Wiman, K.G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9**, 809-18 (1994).
78. Jirmanova, L., Afanassieff, M., Gobert-Gosse, S., Markossian, S. & Savatier, P. Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells. *Oncogene* **21**, 5515-28 (2002).
79. Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B. & Samarut, J. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309-22 (1996).
80. Fujii-Yamamoto, H., Kim, J.M., Arai, K. & Masai, H. Cell cycle and developmental regulations of replication factors in mouse embryonic stem cells. *J Biol Chem* **280**, 12976-87 (2005).
81. Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432-8 (2002).
82. Faast, R. et al. Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). *Oncogene* **23**, 491-502 (2004).
83. Stead, E. et al. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320-33 (2002).
84. Malumbres, M. et al. Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell* **118**, 493-504 (2004).
85. Berthet, C. et al. Combined loss of Cdk2 and Cdk4 results in embryonic lethality and Rb hypophosphorylation. *Dev Cell* **10**, 563-73 (2006).
86. Kozar, K. et al. Mouse development and cell proliferation in the absence of D-cyclins. *Cell* **118**, 477-91 (2004).
87. Geng, Y. et al. Cyclin E ablation in the mouse. *Cell* **114**, 431-43 (2003).
88. Sage, J. et al. Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev* **14**, 3037-50 (2000).
89. Aladjem, M.I. et al. ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Curr Biol* **8**, 145-55 (1998).
90. Hong, Y. & Stambrook, P.J. Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation. *Proc Natl Acad Sci U S A* **101**, 14443-8 (2004).
91. Schmidt-Kastner, P.K., Jardine, K., Cormier, M. & McBurney, M.W. Absence of p53-dependent cell cycle regulation in pluripotent mouse cell lines. *Oncogene* **16**, 3003-11 (1998).
92. Lin, T. et al. p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7**, 165-71 (2005).
93. Beekman, C. et al. Evolutionarily conserved role of nucleostemin: controlling proliferation of stem/progenitor cells during early vertebrate development. *Mol Cell Biol* **26**, 9291-301 (2006).

94. Dannenberg, J.H., van Rossum, A., Schuijff, L. & te Riele, H. Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev* **14**, 3051-64 (2000).
95. Ding, H. et al. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell* **117**, 873-86 (2004).
96. Bartek, J., Lukas, C. & Lukas, J. Checking on DNA damage in S phase. *Nat Rev Mol Cell Biol* **5**, 792-804 (2004).
97. Prost, S., Bellamy, C.O., Clarke, A.R., Wyllie, A.H. & Harrison, D.J. p53-independent DNA repair and cell cycle arrest in embryonic stem cells. *FEBS Lett* **425**, 499-504 (1998).
98. Luo, G. et al. Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. *Proc Natl Acad Sci U S A* **96**, 7376-81 (1999).
99. Takai, H. et al. Aberrant cell cycle checkpoint function and early embryonic death in Chk1(-/-) mice. *Genes Dev* **14**, 1439-47 (2000).
100. Damelin, M., Sun, Y.E., Sodja, V.B. & Bestor, T.H. Decatenation checkpoint deficiency in stem and progenitor cells. *Cancer Cell* **8**, 479-84 (2005).
101. Thornton, B.R. & Toczyski, D.P. Precise destruction: an emerging picture of the APC. *Genes Dev* **20**, 3069-78 (2006).
102. Wirth, K.G. et al. Loss of the anaphase-promoting complex in quiescent cells causes unscheduled hepatocyte proliferation. *Genes Dev* **18**, 88-98 (2004).
103. Uhlmann, F. & Hopfner, K.P. Chromosome biology: the crux of the ring. *Curr Biol* **16**, R102-5 (2006).
104. Murray, A.W. Recycling the cell cycle: cyclins revisited. *Cell* **116**, 221-34 (2004).
105. Weaver, B.A. & Cleveland, D.W. Does aneuploidy cause cancer? *Curr Opin Cell Biol* **18**, 658-67 (2006).
106. Dobles, M., Liberal, V., Scott, M.L., Benezra, R. & Sorger, P.K. Chromosome missegregation and apoptosis in mice lacking the mitotic checkpoint protein Mad2. *Cell* **101**, 635-45 (2000).
107. Le Cam, L., Lacroix, M., Ciemerych, M.A., Sardet, C. & Sicinski, P. The E4F protein is required for mitotic progression during embryonic cell cycles. *Mol Cell Biol* **24**, 6467-75 (2004).
108. Schrott, G. et al. Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. *Mol Cell Biol* **21**, 2933-43 (2001).
109. Coats, S., Flanagan, W.M., Nourse, J. & Roberts, J.M. Requirement of p27Kip1 for restriction point control of the fibroblast cell cycle. *Science* **272**, 877-80 (1996).
110. Flores, I., Benetti, R. & Blasco, M.A. Telomerase regulation and stem cell behaviour. *Curr Opin Cell Biol* **18**, 254-60 (2006).
111. Armstrong, L., Lako, M., Lincoln, J., Cairns, P.M. & Hole, N. mTert expression correlates with telomerase activity during the differentiation of murine embryonic stem cells. *Mech Dev* **97**, 109-16 (2000).
112. Niida, H. et al. Severe growth defect in mouse cells lacking the telomerase RNA component. *Nat Genet* **19**, 203-6 (1998).

113. Niida, H. et al. Telomere maintenance in telomerase-deficient mouse embryonic stem cells: characterization of an amplified telomeric DNA. *Mol Cell Biol* **20**, 4115-27 (2000).
114. Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A. & Stambrook, P.J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci U S A* **99**, 3586-90 (2002).
115. Liu, X. et al. Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Dev Dyn* **209**, 85-91 (1997).
116. Mortensen, R.M., Conner, D.A., Chao, S., Geisterfer-Lowrance, A.A. & Seidman, J.G. Production of homozygous mutant ES cells with a single targeting construct. *Mol Cell Biol* **12**, 2391-5 (1992).
117. 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).
118. Bilodeau, M., Girard, S., Hebert, J. & Sauvageau, G. A retroviral strategy that efficiently creates chromosomal deletions in mammalian cells. *Nat Methods* **4**, 263-8 (2007).
119. Coffin, J.M., Hughes, S.H. & Varmus, H.E. *Retroviruses*, 843 (Cold Spring Harbor Laboratory Press, 1997).
120. Goff, S.P. Host factors exploited by retroviruses. *Nat Rev Microbiol* **5**, 253-63 (2007).
121. Suzuki, Y. & Craigie, R. The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol* **5**, 187-96 (2007).
122. Pages, J.C. & Bru, T. Toolbox for retrovectorologists. *J Gene Med* **6 Suppl 1**, S67-82 (2004).
123. Grez, M., Akgun, E., Hilberg, F. & Ostertag, W. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proc Natl Acad Sci U S A* **87**, 9202-6 (1990).
124. Cherry, S.R., Biniszkiwicz, D., van Parijs, L., Baltimore, D. & Jaenisch, R. Retroviral expression in embryonic stem cells and hematopoietic stem cells. *Mol Cell Biol* **20**, 7419-26 (2000).
125. Pannell, D. et al. Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *Embo J* **19**, 5884-94 (2000).
126. Grez, M., Zornig, M., Nowock, J. & Ziegler, M. A single point mutation activates the Moloney murine leukemia virus long terminal repeat in embryonal stem cells. *J Virol* **65**, 4691-8 (1991).
127. Hawley, R.G., Fong, A.Z., Burns, B.F. & Hawley, T.S. Transplantable myeloproliferative disease induced in mice by an interleukin 6 retrovirus. *J Exp Med* **176**, 1149-63 (1992).
128. Markowitz, D., Goff, S. & Bank, A. A safe packaging line for gene transfer: separating viral genes on two different plasmids. *J. Virol.* **62**, 1120-1124 (1988).
129. Ory, D.S., Neugeboren, B.A. & Mulligan, R.C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proc. Natl. Acad. Sci. U.S.A* **93**, 11400-11406 (1996).

130. Harris, R.S. & Liddament, M.T. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4**, 868-77 (2004).
131. Holmes, R.K., Malim, M.H. & Bishop, K.N. APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* (2007).
132. Nisole, S. & Saib, A. Early steps of retrovirus replicative cycle. *Retrovirology* **1**, 9 (2004).
133. Sullivan, C.S. & Ganem, D. MicroRNAs and viral infection. *Mol Cell* **20**, 3-7 (2005).
134. Swain, A. & Coffin, J.M. Mechanism of transduction by retroviruses. *Science* **255**, 841-5 (1992).
135. Zhang, J. & Temin, H.M. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **259**, 234-8 (1993).
136. Schwartz, J.R., Duesberg, S. & Duesberg, P.H. DNA recombination is sufficient for retroviral transduction. *Proc Natl Acad Sci USA* **92**, 2460-4 (1995).
137. Hajjar, A.M. & Linial, M.L. A model system for nonhomologous recombination between retroviral and cellular RNA. *J. Virol.* **67**, 3845-3853 (1993).
138. Bushman, F. et al. Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* **3**, 848-58 (2005).
139. Lewinski, M.K. et al. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* **2**, e60 (2006).
140. Fehse, B., Kuhlcke, K., Langer, A., Ostertag, W. & Lothar, H. Rapid and efficient cloning of proviral flanking fragments by kanamycin resistance gene complementation. *Nucleic Acids Res* **27**, 706-7 (1999).
141. Hui, E.K., Wang, P.C. & Lo, S.J. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell Mol Life Sci* **54**, 1403-11 (1998).
142. Margolin, A.A. et al. CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics.* **21**, 3308-3311 (2005).
143. Hubbard, T. et al. Ensembl 2005. *Nucleic Acids Res* **33**, D447-53 (2005).
144. Yu, Y. & Bradley, A. Engineering chromosomal rearrangements in mice. *Nat. Rev. Genet.* **2**, 780-790 (2001).
145. Liu, P., Jenkins, N.A. & Copeland, N.G. Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells. *Nat. Genet.* **30**, 66-72 (2002).
146. van der, W.L. & Bradley, A. Mouse Chromosome Engineering for Modeling Human Disease. *Annual Review of Genomics and Human Genetics* **7**, 247-276 (2006).
147. De-Zolt, S. et al. High-throughput trapping of secretory pathway genes in mouse embryonic stem cells. *Nucleic Acids Res* **34**, e25 (2006).
148. Tybulewicz, V.L., Crawford, C.E., Jackson, P.K., Bronson, R.T. & Mulligan, R.C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**, 1153-63 (1991).
149. Brummelkamp, T.R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243-247 (2002).
150. Gaillard, C. & Strauss, F. Ethanol precipitation of DNA with linear polyacrylamide as carrier. *Nucleic Acids Res* **18**, 378 (1990).



151. Barnes, W.M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A* **91**, 2216-20 (1994).
152. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
153. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493-6 (2004).
154. Hadjantonakis, A.K. & Nagy, A. FACS for the isolation of individual cells from transgenic mice harboring a fluorescent protein reporter. *Genesis*. **27**, 95-98 (2000).
155. Tucker, K.L., Wang, Y., Dausman, J. & Jaenisch, R. A transgenic mouse strain expressing four drug-selectable marker genes. *Nucleic Acids Res*. **25**, 3745-3746 (1997).
156. Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M.V. Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol Cell Biol* **13**, 473-86 (1993).
157. Sauvageau, G. et al. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. *Genes Dev* **9**, 1753-65 (1995).
158. Kroon, E. et al. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *Embo J* **17**, 3714-25 (1998).
159. Barrett, T. et al. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* **35**, D760-5 (2007).
160. Luedi, P.P., Hartemink, A.J. & Jirtle, R.L. Genome-wide prediction of imprinted murine genes. *Genome Res* **15**, 875-84 (2005).
161. Biniszkiwicz, D. et al. Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol Cell Biol* **22**, 2124-35 (2002).
162. Szabo, P. & Mann, J.R. Expression and methylation of imprinted genes during in vitro differentiation of mouse parthenogenetic and androgenetic embryonic stem cell lines. *Development* **120**, 1651-60 (1994).
163. van der Weyden, L. & Bradley, A. Mouse chromosome engineering for modeling human disease. *Annu Rev Genomics Hum Genet* **7**, 247-76 (2006).
164. Yang, X. et al. Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat Genet* **39**, 295-302 (2007).
165. Diehn, M. et al. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* **31**, 219-23 (2003).
166. Su, A.I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-7 (2004).
167. Perez-Iratxeta, C. et al. Study of stem cell function using microarray experiments. *FEBS Lett* **579**, 1795-801 (2005).
168. Sene, K.H. et al. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8**, 85 (2007).
169. Eppig, J.T. et al. The Mouse Genome Database (MGD): from genes to mice--a community resource for mouse biology. *Nucleic Acids Res* **33**, D471-5 (2005).
170. Wheeler, D.L. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **35**, D5-12 (2007).

171. Nikitin, A., Egorov, S., Daraselia, N. & Mazo, I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* **19**, 2155-7 (2003).
172. Copeland, N.G., Jenkins, N.A. & Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**, 769-79 (2001).
173. Wang, Z., Engler, P., Longacre, A. & Storb, U. An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res* **11**, 137-42 (2001).
174. Szeto, I.Y., Barton, S.C., Keverne, E.B. & Surani, A.M. Analysis of imprinted murine *Peg3* locus in transgenic mice. *Mamm Genome* **15**, 284-95 (2004).
175. Dixon, J., Brakebusch, C., Fassler, R. & Dixon, M.J. Increased levels of apoptosis in the pre-fusion neural folds underlie the craniofacial disorder, Treacher Collins syndrome. *Hum Mol Genet* **9**, 1473-80 (2000).
176. Meaburn, K.J., Parris, C.N. & Bridger, J.M. The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer. *Chromosoma* **114**, 263-74 (2005).
177. Thomas, K.R. & Capecchi, M.R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-12 (1987).
178. Hadjantonakis, A.K., Gertsenstein, M., Ikawa, M., Okabe, M. & Nagy, A. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech Dev* **76**, 79-90 (1998).
179. Hu, H., Tomasiwicz, H., Magnuson, T. & Rutishauser, U. The role of polysialic acid in migration of olfactory bulb interneuron precursors in the subventricular zone. *Neuron* **16**, 735-43 (1996).
180. Zehbe, I. et al. Sensitive in situ hybridization with catalyzed reporter deposition, streptavidin-Nanogold, and silver acetate autometallography: detection of single-copy human papillomavirus. *Am J Pathol* **150**, 1553-61 (1997).
181. Zambrowicz, B.P. et al. Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci USA* **94**, 3789-94 (1997).
182. Silver, D.P. & Livingston, D.M. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol. Cell* **8**, 233-243 (2001).
183. Pfeifer, A., Brandon, E.P., Kootstra, N., Gage, F.H. & Verma, I.M. Delivery of the Cre recombinase by a self-deleting lentiviral vector: efficient gene targeting in vivo. *Proc Natl Acad Sci USA* **98**, 11450-5 (2001).
184. Jo, D. et al. Epigenetic regulation of gene structure and function with a cell-permeable Cre recombinase. *Nat Biotechnol* **19**, 929-33 (2001).
185. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76 (2006).
186. Boiani, M. & Scholer, H.R. Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol* **6**, 872-84 (2005).
187. Hochedlinger, K. & Jaenisch, R. Nuclear reprogramming and pluripotency. *Nature* **441**, 1061-7 (2006).
188. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997-1001 (2006).

189. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035-8 (2002).
190. Rodolfa, K.T. & Eggan, K. A Transcriptional Logic for Nuclear Reprogramming. *Cell* **126**, 652-655 (2006).
191. Moffat, J. et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283-98 (2006).
192. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).
193. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
194. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321-5 (2004).
195. Zhang, J. & Sapp, C.M. Recombination between two identical sequences within the same retroviral RNA molecule. *J Virol* **73**, 5912-7 (1999).
196. Delviks, K.A. & Pathak, V.K. Development of murine leukemia virus-based self-activating vectors that efficiently delete the selectable drug resistance gene during reverse transcription. *J Virol* **73**, 8837-42 (1999).
197. Varela-Echavarria, A., Prorock, C.M., Ron, Y. & Dougherty, J.P. High rate of genetic rearrangement during replication of a Moloney murine leukemia virus-based vector. *J Virol* **67**, 6357-64 (1993).
198. Parthasarathi, S., Varela-Echavarria, A., Ron, Y., Preston, B.D. & Dougherty, J.P. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. *J Virol* **69**, 7991-8000 (1995).
199. Puech, A. et al. Normal cardiovascular development in mice deficient for 16 genes in 550 kb of the velocardiöfacial/DiGeorge syndrome region. *Proc.Natl.Acad.Sci. U.S.A* **97**, 10090-10095 (2000).
200. Mitalipova, M.M. et al. Preserving the genetic integrity of human embryonic stem cells. *Nat Biotechnol* **23**, 19-20 (2005).

# APPENDIXES

## APPENDIX I: Article

### Uncovering stemness

Mélanie Bilodeau<sup>1</sup> and Guy Sauvageau<sup>1,2</sup>

<sup>1</sup>Laboratory of Molecular Genetics of Stem Cells, Institut de Recherche en Immunologie et Cancérologie (IRIC), C.P. 6128 succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada; <sup>2</sup>Department of Medicine and Division of Hematology and Leukemia Cell Bank of Quebec, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada

Article published in *Nature Cell Biology* 8, 1048-9 (2006)

Reprinted by permission from Macmillan Publishers Ltd: NATURE CELL BIOLOGY [*Nature Cell Biology* 8, 1048-9 (2006)], copyright 2006.

Corresponding Author:

Guy Sauvageau M.D., Ph.D.

Institut de Recherche en Immunologie et Cancérologie

C.P. 6128, succursale Centre-Ville, Montréal, Québec

Canada H3C 3J7

Telephone: (514) 343-7134

Facsimile: (514) 343-7379

E-mail: [REDACTED]

### Author contributions

Mélanie Bilodeau wrote the manuscript and prepared the figure under Guy Sauvageau guidance.

## Abstract

Understanding how self-renewal and pluripotency - two key characteristics of stem cells - are controlled may allow generation of stem cell lines from somatic tissues, avoiding the ethically contentious need to derive them from embryos. A step in this direction was recently taken by two teams, who exploited recombinant retroviruses in gain and loss of function experiments to characterize candidate transcription factors with the potential to regulate “stemness”.

## News & Views

For years, many predicted that understanding the properties of murine (mESC) and human (hESC) embryonic stem cells could lead to the design of cell replacement therapies<sup>44</sup>. Using a list of candidate factors, Takahashi and Yamanaka<sup>185</sup> show that a specific combination of only four factors allows the generation of pluripotent stem cells from mouse adult fibroblast cultures, avoiding the controversial need of deriving them from an embryo. On the other hand, Ivanova *et al.*<sup>15</sup> chose a RNA interference strategy to identify known and novel transcriptional regulators from two distinct pathways that control mESC self-renewal<sup>15</sup>.

Both cell extrinsic and intrinsic factors regulate mESC self-renewal (maintenance of cell characteristics through division) and pluripotency (ability to form all lineages from all tissues). In terms of cell extrinsic factors, mESC maintenance *in vitro* requires the presence of the leukemia inhibitory factor (LIF) and bone morphogenic protein (BMP) that signal through Stat3 and Smad proteins, respectively<sup>186</sup>. Evidence indicates that Wnt signaling is also involved<sup>186</sup>. The cell intrinsic machinery includes the transcription factors *Oct4*, *Nanog* and *Sox2*<sup>49</sup>. The pluripotency state could also be regulated by epigenetic mechanisms involving members of the Polycomb group proteins<sup>49</sup>.

Up to now, only two relatively inefficient methods have allowed somatic cells reprogramming into a pluripotent state: nuclear transfer and cell fusion. Interestingly, reprogramming efficiency is influenced by the developmental/epigenetic stage of the donor cell<sup>187</sup> and might be enhanced by ectopic expression of *Nanog*<sup>188</sup>. Importantly, the exact blend of factors necessary to reprogram a cell remained unknown until recently<sup>185</sup>. Indeed, Takahashi and Yamanaka selected 24 candidate genes either with documented roles in ES self-renewal and pluripotency, or specifically expressed in these cells. These factors were expressed alone or in combination using retroviral gene transfer in mouse embryonic fibroblasts (MEFs) or

adult tail-tip fibroblasts (TTFs). Interestingly, cells used in these experiments were derived from a transgenic mouse bearing a  $\beta$ geo gene integrated into the *Fbx15* locus, which is mostly active in mESCs. The newly generated pluripotent clones engineered to express the subgroup of transcription factors were then selected with G418 for  $\beta$ geo expression. By testing increasingly narrow combinations of candidates, the authors identified a minimal set of genes necessary to obtain G418<sup>R</sup> pluripotent stem cells derived from MEFs (iPS-MEF) and TTFs (iPS-TTF) cultures. Only four factors turned out to be essential: Oct4, Sox2, c-Myc and Klf4<sup>185</sup>. iPS clones isolated resembled mESCs both in terms of morphological criteria and transcriptional signatures, as evaluated by RT-PCR and microarray analyses, as well as by their capacity to differentiate into the three germ layers both *in vitro* during embryonic body (EB) differentiation and *in vivo* by teratoma formation. Interestingly, the selected stem cell lines (iPSs) strictly needed LIF and feeders to remain undifferentiated, suggesting that the four transduced factors were not sufficient to confer cell autonomous properties. Although Nanog was not required for their generation, it might overcome the need for LIF and BMP (provided by the serum) and accentuate the ES-like properties of the reported iPSs. When introduced in mouse blastocyst, iPS-TTF could contribute to derivatives of the three germ layers up to day E13.5; however, no chimeric mice were born<sup>185</sup>.

It seems that somehow the formation of iPS is constrained. Using the data provided, we estimated the efficiency of reprogramming to be between ~1 per 2,500 to ~1 per 30,000 infected embryonic and adult fibroblasts, respectively. Because of the great scientific impact of these findings and the low frequency of the observed phenomenon, two critical issues arise: Which cell type(s) was reprogrammed? Was the reported combination of factors sufficient for reprogramming? Although the authors reasonably argued that the low frequency of iPS derivation is due to selective pressure for the rare cells that express appropriate levels of the factors, it will be necessary to experimentally demonstrate this in subsequent studies. As recently reported by Jeanisch and colleagues<sup>189</sup>, definitive proof of mature cell nuclear reprogramming will emerge when similar studies are performed with differentiated cells that are genetically marked.

The issue of whether these factors are sufficient for reprogramming remains open, especially considering the important range in frequency of iPS generation and the inability of these cells to behave as normal mESCs. The interesting possibility that insertional mutagenesis has contributed must be considered because ~20 integrations are reported per clone. The identification of common insertion sites in the various clones could thus reveal novel reprogramming factors. Since the diversity of the retroviral particles used in this study

was apparently generated by transfection of plasmid pools, the possibility that unexpected recombinant viruses have been generated remains. Other caveats not related to retroviral infection have already been discussed by Rodolfa and Eggan and will not be repeated here<sup>190</sup>. Nevertheless, even if some points need to be clarified concerning the iPS and the genes involved, the reported somatic cell reprogramming with just a handful of factors is of major interest and represents a breakthrough in the stem cell field.

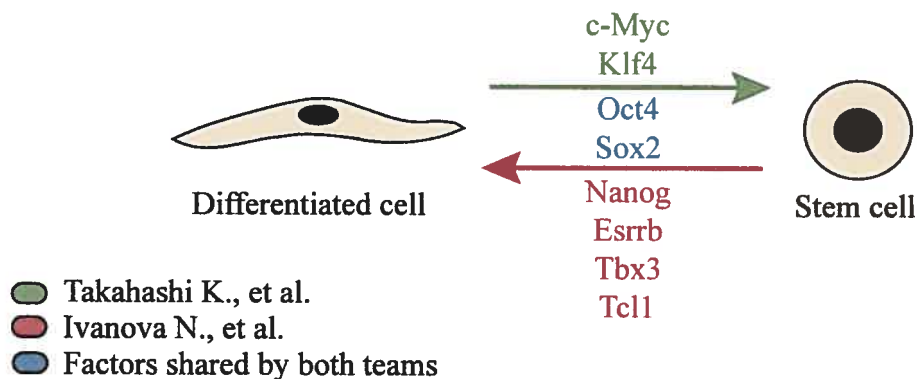
Ivanova *et al.*<sup>15</sup> used a complementary approach to identify mESC self-renewal genes. They report a list of around 65 candidate transcription factors from a selection of 901 genes down-regulated during retinoic acid-induced mESC differentiation and generated shRNA-based lentiviral vectors for these genes. The efficiency of mRNA knock-down of this approach compares favorably to most other methods reported to date<sup>191</sup>. In order to evaluate the effect on self-renewal capacity, they designed a competitive *in vitro* assay between transduced and untransduced cells which resulted in a final list of 10 candidate genes. Two of these genes were eliminated because they induce cell loss. The others were tested with an elegant rescue experiment using a lentiviral vector containing both the shRNA and its corresponding cDNA expressed under the control of an inducible promoter. Only one candidate failed in the rescue experiment. The functions of all the other candidates were characterized by knocking down their expression in ESCs or forcing their expression during differentiation. Analyses were then performed by RT-PCR for expression markers corresponding to the three germ layers and the trophoectoderm, or by the contribution of these modified ESCs to chimeric mice. The effect of knockdown experiments was also evaluated by microarray analysis. These results showed for the first time that two separate pathways seem to regulate self-renewal, one including *Nanog*, *Oct4* and *Sox2* and the other *Esrrb*, *Tbx3*, *Tcl1* and *Dppa4*. A hundred sixty candidates for positive regulators of differentiation were selected from the microarray data and overexpressed in ESCs; eighteen induced differentiation. The direct transcriptional link between these genes and the seven candidate self-renewal factors will need to be demonstrated. Nonetheless, these results represent an enormous endeavor in understanding the transcriptional complexity underlying mESC self-renewal.

Together, these studies confirm the involvement of *Oct4* and *Sox2* in the maintenance of ESC identity and further underscore their ability to induce nuclear reprogramming (**Figure 0-1**). Based on this, it should be of interest to test the reprogramming potential of the other factors identified by Ivanova *et al.*<sup>15</sup> and also the target genes shared by *Oct4* and *Sox2*<sup>192</sup>. Moreover, these results will become even more significant when epistatic and synthetic interactions are uncovered, as might be achieved through high-throughput ESC experiments.

Thus, it can now be concluded that a small set of transcription factors are key components regulating ESC fate decisions. Understanding their post-translational modification in self-renewing cells will become necessary. Non-coding microRNAs are also expected to be implicated in the regulatory network<sup>193</sup>. In addition, accumulating evidence indicate that up to ~2% of the genome involves non-coding regions kept under active evolutionary selection, with more than 5000 sequences of over 100 bp that are absolutely conserved between human and mouse<sup>194</sup>. Interestingly, some of these conserved elements, referred as “bivalent domains”, may silence developmental genes in ESCs but also keep them poised for activation<sup>52</sup>. It can be envisioned that chromosome engineering technology will become part of the growing list of complementary approaches to decipher the function of these conserved regions that likely control “stemness”<sup>35</sup>.

### Figure 0-1 Candidate transcription factors that determine stem cell identity

Oct4 and Sox2 (blue) contribute to reprogramming fibroblasts into a pluripotent state (green)<sup>185</sup>. Interestingly, these genes are also required for maintenance of mESC cell self-renewal (red)<sup>15</sup>.





## APPENDIX II: TOWARD THE DESIGN OF A SUCESSFUL RETROVIRAL SYSTEM


Appendix II describes the construction and validation of complementary retroviruses, carriers of *loxP* sites, designed specifically to create Cre-induced deletions in mammalian cells. This chapter is an article in preparation (Rapid communication) relating technical problems encountered in generating these viruses. These observations highlight the complexity of retroviral biology and the high degree of vigilance that these commonly used entities deserve. **Supplementary Figure 0-5** presents conclusive retrovirus testing and successful Cre-induced recombination in NIH 3T3 cells. These results paved the way to the work reported in the Chapter 2. Figures presented in this chapter do not overlap with those found in Chapter 2. However, a testing summary was presented in a supplementary figure of Chapter 2 manuscript (**Figure 2-3**). Melanie Bilodeau performed 100% the experiments, generated all the figures, and analyses under the supervision of Guy Sauvageau.

# ARTICLE: Shunning pitfalls in retroviral vectors design for functional genomics

Mélanie Bilodeau<sup>1</sup> and Guy Sauvageau<sup>1,2,3\*</sup>

<sup>1</sup>Laboratory of Molecular Genetics of Stem Cells, Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montréal, Québec, Canada, H2W 1R7. <sup>2</sup>Department of Medicine, Montréal, Québec, Canada, H3C 3J7. <sup>3</sup>Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montréal, Québec, Canada, H1T 2M2.

Rapid communication in preparation for *Virology*

\*Correspondence: Guy Sauvageau, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, Canada, H3C 3J7. Email : 

## Author contributions

Mélanie Bilodeau performed the experiments, prepared all the figures, and wrote the manuscript under Guy Sauvageau's guidance.

## Abstract

The design of specific retroviral vectors and packaging cells to generate recombinant virions which are suitable for functional genomic studies remains difficult to achieve even with today's technology. Rearranged proviruses were observed in embryonic stem cells, related to different retroviral constructs. Independently, a retroviral-like particle conferred a drug resistance gene to target cells, although this gene was absent from the designated viral construct. Finally, an inducible drug resistance cassette was leaky when placed downstream of promoter(s), resulting in undesired resistant cells. Together, these observations raise concerns about low incidence events following retroviral gene transfer and gene therapy mediated by retroviral vectors.

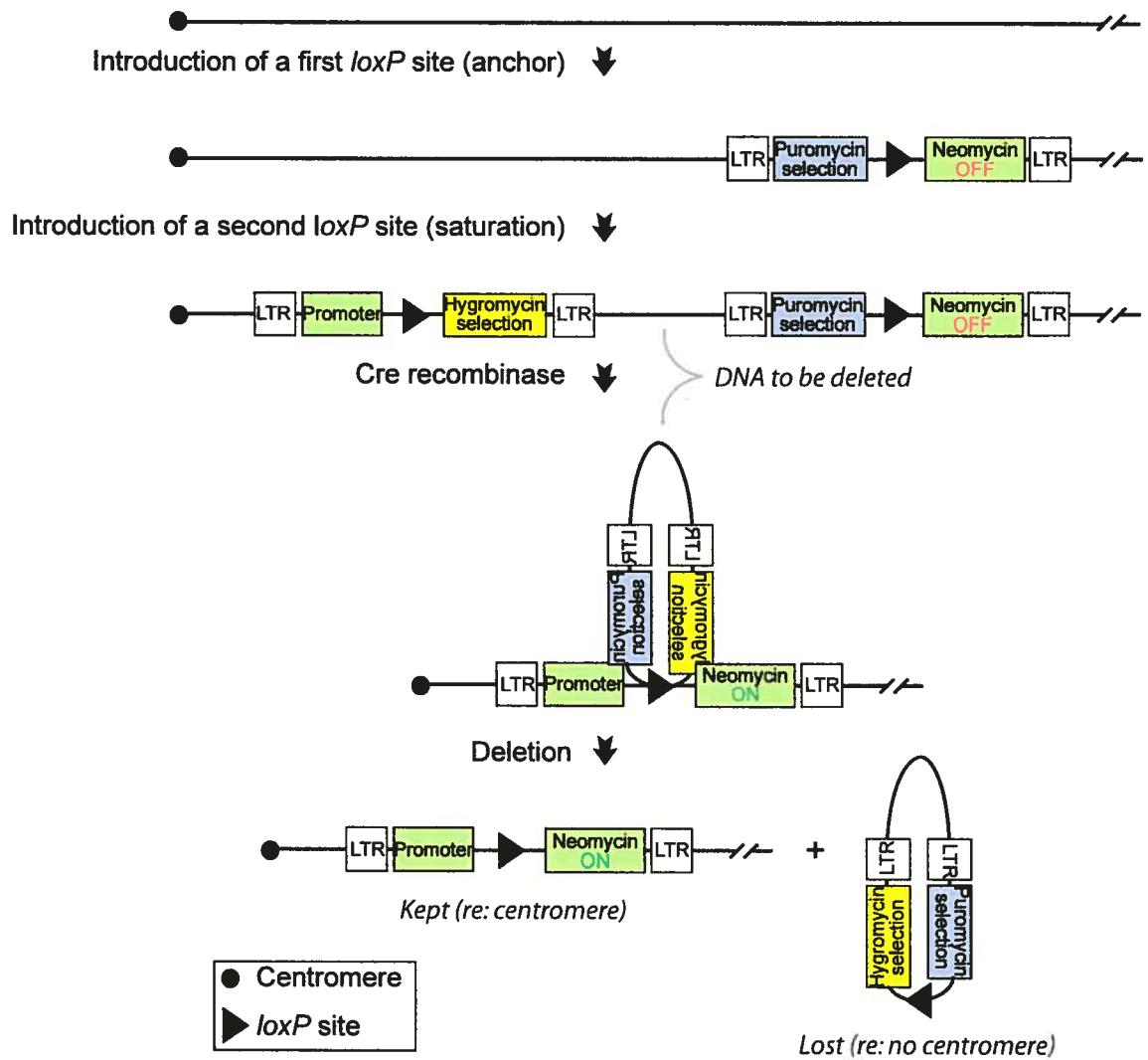
## Introduction

Compatible retroviral vectors were designed and were optimized to create deletions in embryonic stem cells (ESCs) using Cre-*loxP* recombination system<sup>118</sup>. Construct representations and a brief testing summary were published in the supplementary information of a manuscript<sup>118</sup>. These *loxP*-containing retroviral vectors were classified in two complementary sets: anchor and saturation viruses<sup>118</sup>. Properties common to anchor viruses were the presence of a *loxP* site, a puromycin gene to select for stable integration, and a promoter-less neomycin gene devoid of its first translation initiation codon (ATG) (**Figure 0-2**). Saturation viruses shared the following characteristics: a *loxP* site, a hygromycin gene to select integration, and a promoter coupled to an ATG to complement the inactive neomycin gene found in anchor viruses (**Figure 0-2**).

Challenging problems were encountered for several retroviral vectors and a specific packaging cell line. These difficulties included rearrangements of proviruses, transmission of a presumed retroviral-like particle, and unexpected gene expression. The entire analysis of these problems is presented to review and to illustrate some potential drawbacks of retroviral-based system design.

**Figure 0-2 Retroviral constructs were designed to mediate Cre-loxP recombination.**

Both the anchor and the saturation viruses deliver a *loxP* site in the genome of mammalian cells. The former is selected with puromycin and the latter with hygromycin. A promoter-ATG cassette is delivered by the saturation virus while an inactive neomycin gene (OFF) is introduced by the anchor virus, and the bipartite neomycin selection gene is reconstituted (ON) following Cre-induced recombination between *loxP* sites.



## Results

### Rearranged proviruses

Constructs leading to rearranged proviruses can be classified in to one predominant or to several rearrangements. For the first class, two constructs (S4 and S5) showed a similar rearrangement in almost all proviruses, according to Southern blot sensitivity (**Figure 0-3a**). Both S4 and S5 virions were Vesicular Stomatitis Virus G (VSV-G) pseudotyped in 293GPG packaging cells<sup>129</sup>. Southern blot analyses using *NheI* enzymatic digestion and hygromycin detection revealed that S4 provirus size in genomic DNA extracted from ESC polyclonal populations (lane 2 and 3, in absence or presence of the virus *Cre*, respectively) was smaller than expected (plasmid DNA, lane 4) (**Figure 0-3a**, top panel). Using the same enzymatic digestion and hybridization conditions, S5 proviruses in genomic DNA of GP+E86 or ESC target cells (lanes 7 and 8, respectively) were also of reduced size in comparison to plasmid DNA (pS5, lane 15, **Figure 0-3a**). However, GP+E-86 cells directly transfected with the virus S5 plasmid and selected with hygromycin presented a dominant band of expected size in addition to a smaller fragment as observed in target cell genomic DNA (**Figure 0-3a**, lanes 6-8). Suspecting that rearrangements (smaller bands) were caused by the two *Pgk-1* direct repeat sequences introduced in S4 and S5 constructs, the potential deletion was further defined. Southern blot analyses with *HindIII* enzymatic digestion combined to hygromycin detection were expected to show a 1.9 kb band as observed for control plasmid S5 (**Figure 0-3a**, lane 16) or a polyclonality smear if one of the *HindIII* site was loss. In fact, one of the *HindIII* site was loss in S5 proviruses integrated in ESCs genomic DNA (**Figure 0-3a**, lane 9). The *XhoI* restriction site was preserved, according to a 1.9 kb band detected with *XhoI-HindIII* double restriction digest (lane 10), a product similar to either *HindIII-HindIII* or *XhoI-HindIII* double restriction digests of plasmid S5 (lanes 16-17) (**Figure 0-3a**). This deletion probably occurred during reverse transcription, prior to integration in target cells, as suggested by others<sup>195,196</sup>. The minor rearrangement observed in GP+E-86 packaging cells transfected with S5 plasmid reveals either that the deletion might also occur directly in genomic DNA, or a minor proportion of the packaging cells can be infected with viruses that they produced. The first hypothesis is more likely since packaging cells are presumed to be resistant to the virions they themselves produce because viral envelop proteins block their surface receptors<sup>119</sup>.

For the second category (multiple rearrangements), both the saturation S3 and the anchor A3 viruses relied on the viral 5'LTR promoter to direct hygromycin or puromycin

expression, although other elements were found in between (**Figure 0-3b,c**). Polyclonal GP+E86 populations, stably transduced with corresponding plasmid DNA, were recovered following respective selection. Genomic DNA analysis by Southern blot presented fragments of appropriate sizes and polyclonality smears, as expected (**Figure 0-3b**, lane 1,3-4 and **c**, lane 2-3, 20-21 and 38). However, ESCs infections were ineffective and viral titers were low in both cases (data not shown). Nevertheless, some totally altered A3 proviruses permitted the recovery of puromycin resistant (puro<sup>R</sup>) ESC clones, although at low frequency (**Figure 0-3c**, lanes 5-19 and 23-37). These rearranged proviruses could have originated from multiple causes: alteration during plasmid DNA random integration, rearrangement with packaging cell endogenous elements, or rearrangement during reverse transcription or at the genomic level in target cells, etc. The large anchor virus A4 also demonstrated a high rearrangement frequency (**Figure 0-3d**). Clonal analysis by Southern blot with genomic DNA extracted from 92 ESC clones infected at low multiplicity of infection, presented frequent fragments smaller than the 4.1 kb minimal expected size (**Figure 0-3d**, *Bgl*III restriction digests and puromycin detection, top panels). Detection with a neomycin probe revealed suspected rearrangements since several ESC clones were lacking a signal (**Figure 0-3d**, second panels). From these limited observations, occurrence of rearranged proviruses in these ESC clones was estimated around 70%<sup>118</sup>. A couple of problems might explain the poor stability of this construct. First, it contains *herpes simplex thymidine kinase (tk)* gene that can cause rearrangements in proviruses<sup>197,198</sup>. This cause was probably minor in this experiment because the same *tk* gene found in the S4 construct (**Figure 0-3a**) did not perpetrate this type of rearrangement. Most probably, the cause was a small repetitive sequence (~190 bp) introduced between the puromycin coding sequence and the IRES (including *Nhe*I site) that corresponded in part to sequences in the vicinity and in the 3'LTR (but not the polyadenylation signal). Nonetheless, many altered puro<sup>R</sup> proviruses could be recovered in target cells. Further Southern Blot analyses revealed that although some genomic DNA extracted from ESC clones presented an intact *Nhe*I fragment detected with a puromycin probe (clones 33-34,38,40 corresponding to lanes 5,6,10,12, panel 3), none of them contained the neomycin gene detected either with *Bgl*III (panel 2), *Nhe*I (panel 4), or *Eco*RI (panel 6) restriction digests (**Figure 0-3d**). In fact, further analysis of genomic DNA from 1 of 10 clones demonstrated an intact *Eco*RI fragment revealed with puromycin detection (**Figure 0-3d**, clones 31-40, panel 5). However, none of these genomic DNA showed an intact *Nhe*I fragment detected with a neomycin probe (panel 4).

Rearrangement in proviruses can also be caused by splice donors and acceptors found in plasmid constructs<sup>119</sup>. We did not address this possibility experimentally. Together,

these results illustrate that inappropriate sequence organization in retroviral constructs can be associated with provirus rearrangements. The generation of replication-incompetent retrovirus and target cell infection create a range of observable proviral variants under selective conditions.

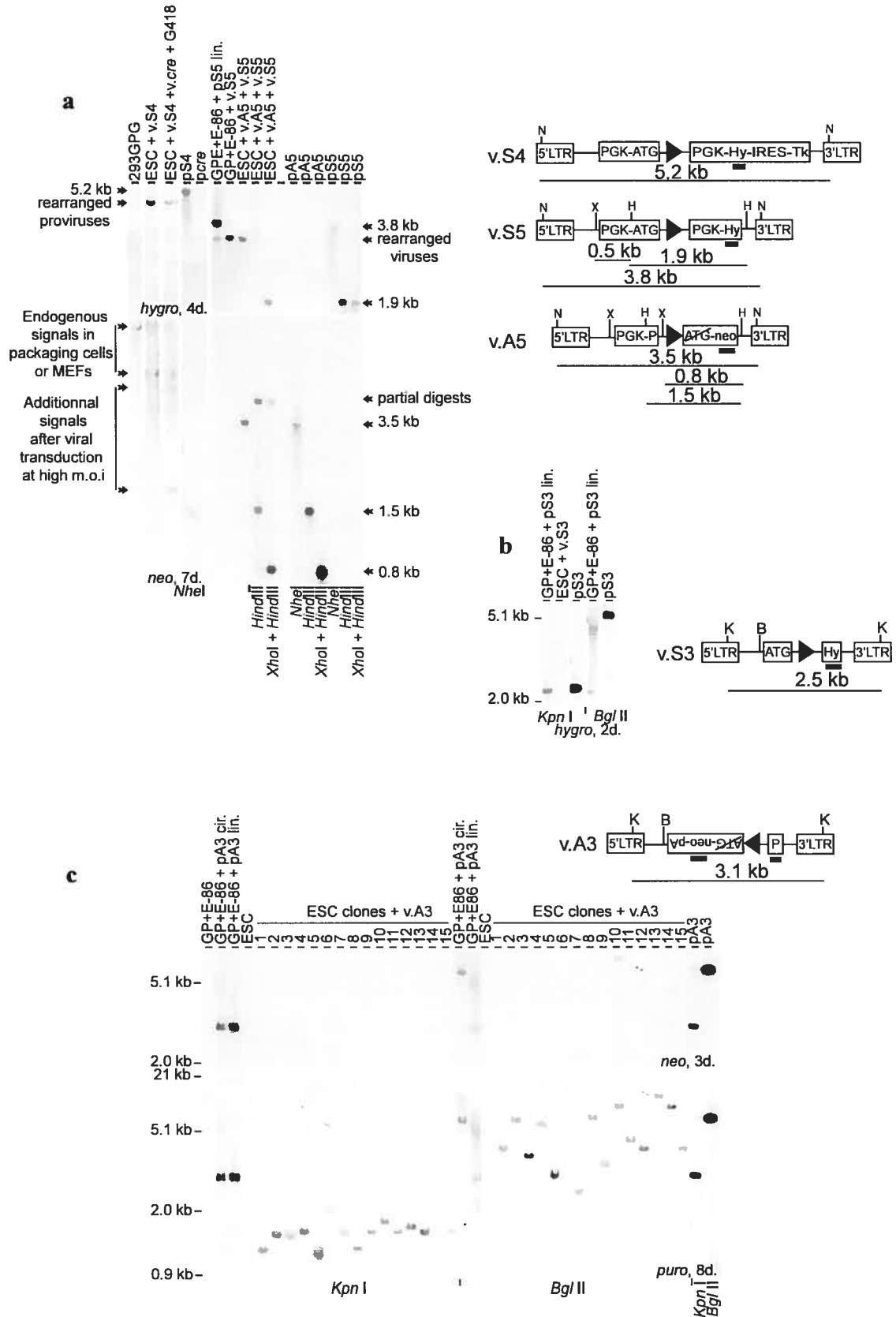
### **Transmission of undesired retroviral-like particle**

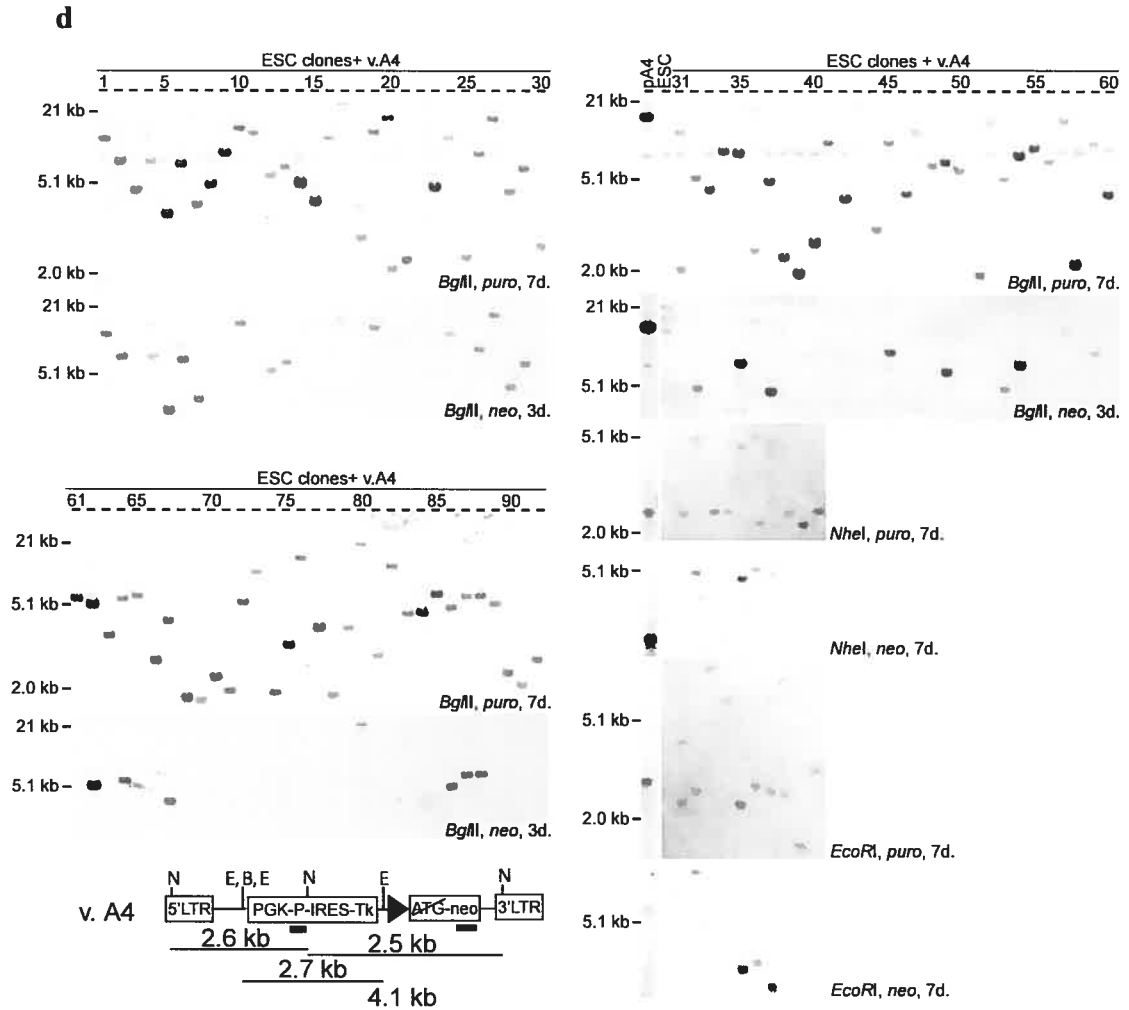
As a selection control for an experiment, ESCs successively infected with S4 and *Cre* retroviruses were treated with G418 (geneticin) (**Figure 0-3a**). Total sensitivity was expected since neither of these retroviral constructs contained the neomycin selection gene (**Figure 0-3a**, bottom panel, lanes 4-5). Surprisingly, ~0.4% of the ESC colonies were G418 resistant (G418<sup>R</sup>) when infected both with S4 and *Cre* viruses, but totally sensitive when infected only with the S4 retrovirus. Southern blot analyses with *NheI* restriction digests and neomycin hybridization, performed with genomic DNA extracted from this G418<sup>R</sup> ESC population, revealed two additional neomycin bands that were not observed in 293GPG or residual feeder cell (MEFs) genomic DNA (lanes 1-3) (**Figure 0-3a**, bottom panel). Presumably one or two dominant retroviral-like product(s) corresponded to the two specific neomycin fragments observed in the genomic DNA of G418<sup>R</sup> polyclonal ESC population (**Figure 0-3a**, bottom panel), and were also found in the genomic DNA of two other independent ESC populations infected with the same virus *Cre*<sup>182</sup> (data not shown). Most likely this retroviral-like particle(s) was present at low frequency in the virus *Cre* packaging cells, but only revealed following G418 selection of target cells. This phenomenon was not an isolated event. Neomycin resistance (frequency: 100-fold lower) was observed with an unrelated *Cre* viral construct tested on the same ESC populations (data not shown). These VSV-G pseudotyped virions were produced by the amphotropic packaging cell line 293GPG<sup>129</sup>. Aberrant G418<sup>R</sup> viruses might have been formed by recombination in 293GPG cells containing the neomycin gene co-integrated with the packaging functions<sup>129,135,137</sup>. The retroviral-like particle(s) observed by Southern blot analysis might corresponded to replication-competent retrovirus (helper), to satellite virus (replication-incompetent, encoding *env*), or to satellite RNA (replication-incompetent, not encoding *env*). In any case, neomycin resistance was transferred, resulting in a totally unacceptable background.

**Figure 0-3 Rearranged proviruses and transmission of a retroviral-like particle.**

(a-d) Southern blot analyses of genomic DNA extracted from 293GPG or GP+E-86 packaging cells transfected or not with circular (cir.) or linear (lin.) plasmid DNA (p), and ESC clones/polyclonal populations or GP+E-86 infected with the indicated viruses. Plasmids are larger (~3kb more) than the depicted proviruses (v.S4, v.S5, v.A5, v.S3, v.A3 and v.A4) because of their backbones. As a general rule, provirus rearrangement analysis was done with a single or a combination of enzymes to give a fragment of expected size (for example, *NheI* or *KpnI* that cut in both MSCV LTRs) following detection with an internal probe as indicated (hygro, neo, puro). For clonal analysis, usually the selected enzyme cut once in the provirus and elsewhere in genomic DNA, or cleaves outside of the provirus, and the probe is internal to the provirus.







### Symbols

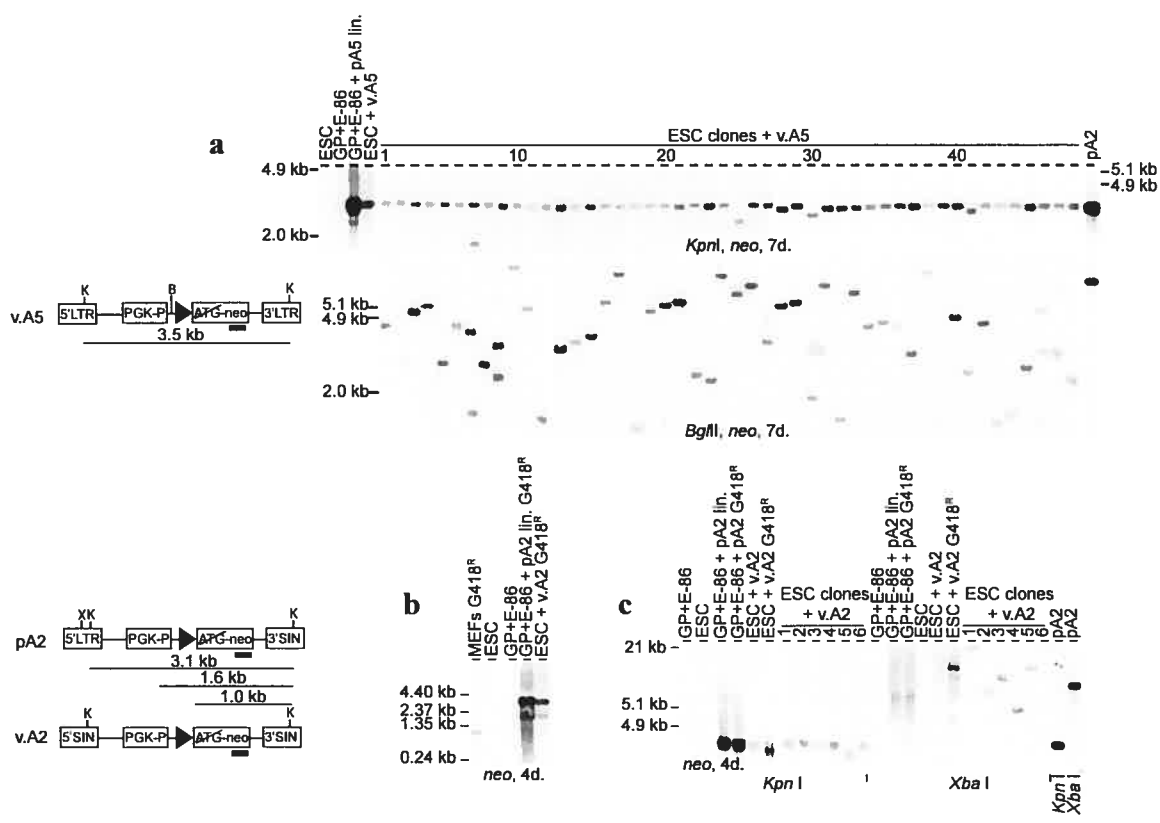
- |  |                               |
|--|-------------------------------|
| <b>ATG:</b> translation initiation codon mutated           | <b>v.:</b> virus              |
| <b>d.:</b> days of exposure                                | <b>N:</b> <i>NheI</i>         |
| <b>Hygro or Hy:</b> hygromycin resistance gene             | <b>X:</b> <i>XhoI</i>         |
| <b>IRES:</b> internal ribosomal entry site                 | <b>H:</b> <i>HindIII</i>      |
| <b>LTR:</b> long terminal repeat                           | <b>K:</b> <i>KpnI</i>         |
| <b>Neo:</b> neomycin resistance gene                       | <b>B:</b> <i>BglII</i>        |
| <b>p:</b> plasmid  | <b>E:</b> <i>EcoRI</i>        |
| <b>pA:</b> polyadenylation signal                          | <b>-:</b> Southern blot probe |
| <b>Pgk:</b> murine <i>phosphoglycerate kinase</i> promoter |                               |
| <b>Puro or P:</b> puromycin resistance gene                |                               |
| <b>Tk:</b> herpes simplex <i>thymidine kinase</i> gene     |                               |

## Undesired neomycin expression

A5 anchor construct was particularly promising since proviruses in ESC clones demonstrated low rearrangement frequency according to Southern blot analysis of genomic DNA (**Figure 0-4a**). Most of the time, an intact 3.5 kb *KpnI* fragment was detected with a neomycin probe (**Figure 0-4a**, top panel). However, cells containing A5 provirus were G418<sup>R</sup> (data not shown), which was incompatible with the proposed system (neomycin expression only upon Cre-induced recombination). To minimize the neomycin gene leakiness, a construct based on a self-inactivating (SIN) MSCV backbone<sup>149</sup> was generated to reduce or abolish expression from the viral 5'LTR (**Figure 0-4b,c**). This virus significantly reduced the frequency of undesired neomycin resistance (~100 fold reduction<sup>118</sup>). However, some G418<sup>R</sup> ESCs persisted in a population of infected cells, as assessed functionally and by northern blot analysis of total RNA (**Figure 0-4b**, lane 5). Neomycin expression seemed directed both from the 5'LTR and from the internal *Pgk-1* promoters in producers (as expected because of the intact 5'LTR in the plasmid) and in target cells (not expected because following reverse transcription, the 5'LTR is mutated) (**Figure 0-4b**, lanes 4-5). Recombination event(s) could explain the unexpected expression presumably directed from the 5'SIN LTR. Neomycin resistance correlated with a subtle genomic rearrangement (lane 6), but most isolated ESC clones (lanes 7-10 and 12) or unselected polyclonal ESC population (lane 5) presented an intact provirus according to Southern Blot analysis of genomic DNA (**Figure 0-4c**). Together these results showed that ATG-less-neomycin gene expression was not abolished when the cassette was placed downstream of promoters. In addition, discrete and infrequent rearrangements, observed under selective conditions, restored expression from the mutated 5'SIN LTR, an observation noticed before<sup>119</sup>.

### Figure 0-4 Retroviruses associated with unexpected neomycin resistance.

(a) Rearrangement (*KpnI* enzymatic digestion) and clonal analyses (*BglII* enzymatic digestion) of genomic DNA extracted from packaging cells (GP+E-86) transfected with A5 linear (lin.) plasmid (pA5) or from ESC clones infected with A5 virus (v.A5). Southern blot analyses were performed with a neomycin probe, most of the time revealing an intact and unique provirus in ESC clones. (b) Northern blot analysis revealing neomycin transcripts using RNA extracted from of G418<sup>R</sup> MEFs (hybridization control, lane 1), v.A2 producer cells (lane 4), and ESCs infected with v.A2 (lane 5). (c) Southern blot analysis performed with genomic DNA extracted from v.A2 viral producers (lanes 3-4 and 14-15) or ESC infected with v.A2 (lanes 5-6 and 17-18), before and after G418 selection. Six ESC clones infected with v.A2 are also shown. Virus integrity is revealed with *KpnI* restriction digests (3.1 kb bands), and clonal analysis with *XbaI* restriction digests. Note that G418<sup>R</sup> polyclonal ESC population presents slightly rearranged proviruses (lanes 6 and 18). Cartoons represent either proviruses or plasmid (without the backbone).



#### Symbols

ATG: translation initiation codon mutated  
 d.: days of exposure  
 ESC: embryonic stem cells  
 GP+E-86: GP+E-86 packaging cell line  
 G418<sup>R</sup>: G418 resistant  
 LTR: long terminal repeat  
 MEFs: mouse embryonic fibroblasts  
 Neo: neomycin resistance gene

p: plasmid  
*Pgk*: murine phosphoglycerate kinase promoter  
 Puro or P: puromycin resistance gene  
 SIN: long terminal repeat bearing a deletion in the U3 region  
 v.: virus  
 X: *XbaI*  
 K: *KpnI*  
 B: *BglII*  
 ▬: probe

## Discussion and Conclusions

Together these results underlined some difficulties faced when designing retroviral vectors in association with packaging cell lines. Three suitable retroviruses were generated for the proposed system (A1 and S1 described previously<sup>118</sup>, and S2) by resolving these problems (**Supplementary Figure 0-5**). First, GP+E-86 ecotropic packaging cell line (free of neomycin, puromycin, and hygromycin genes)<sup>128</sup> was used to avoid undesired resistance transmitted by retroviral-like particles. To minimize proviral rearrangement, repetitive sequences and *herpes simplex thymidine kinase* gene were avoided, and simple viral constructs were generated. In order to abrogate undesired neomycin expression, a self-inactivating MSCV backbone was used and the neomycin cassette was inserted in reverse orientation compared to the 5'LTR promoter and to the *Pgk*-puromycin cassette.

Creating proper retroviral constructs that can be used in functional genomic studies (i.e., low rearrangement frequency, robust inducible drug resistance cassette, etc.) is a controllable process. Packaging cells and the retroviral life cycle are not as manageable. Transmissible but altered retroviral-like entities can be produced by multiple ways: recombination or read-through transcripts or spliced transcripts initiated in packaging cells, recombination during reverse transcription involving foreign viral or cellular RNA co-encapsidated with the designated viral genomic RNA, and other possibilities<sup>119,134-137</sup>. Several of these events probably happen at low frequency. Unfortunately, some certainly stay invisible because of their lack of detectable/selectable characteristics. Results reported in this manuscript suggest careful interpretations of low incidence observations following retroviral gene transfer and raise concerns about gene therapy mediated by retroviral vectors.

## Methods

### Retroviral constructs.

**Virus S5.** *Pgk-1* promoter fragment from pMSCVpuro was ligated with an adapter *kozac-ATG* (oligonucleotides 5'-agcttaccatgg -3' and 5'-aattccatggta-3') and was placed in pBluescript (Stratagene) upstream of a *loxP* sequence (plasmid no.1535). (*LoxP* from plasmid *Pgk-loxP-Pgk-neo*<sup>199</sup> offered by Bruno Saint-Jore, INSERM U321, Paris, France). *Pgk-kozac-ATG-loxP* fragment was subcloned into the *HpaI* site of pMSCVhyg (plasmid no.1537 corresponding to virus S5).

**Virus S4.** An IRES fragment from pMSCV-IRES-YFP (Clontech) was ligated to the *herpes simplex tymidine kinase gene (tk)* (PCR fragment produced with oligonucleotides 5'-gagaattctcagttagcctccccatctc-3' and 5'-ggaagatctaccatggcttctgacctc-3') in pBluescript (Stratagene) (plasmid no.1539). IRES-*tk* fragment was subcloned in the *SalI* site of plasmid no.1537 (plasmid no.1538 corresponding to virus S4).

**Virus A4.** A *loxP-ATG-less-neo* fragment was PCR amplified from pPNT<sup>148</sup> (primers: 5'-gagaattcataacttcgtatagcatacattatacgaagttattaggatcgccattgaa-3' and 5'-gagaggatcctcagaagaactogtc-3') and inserted downstream of IRES-*tk* in plasmid no.1539 (plasmid no.1540) or in pBluescript (Stratagene) (plasmid no.1541). A *Pgk-puro* fragment (*EcoRV* restriction digest) was released from pMSCVpuro and subcloned upstream IRES-*tk-LoxP-ATG-less-neo* in plasmid no.1540 (plasmid no.1542). *Pgk-puro-IRES-tk-LoxP-ATG-less-neo* fragment was then subcloned into pMSCVneo linearized with *BgIII-BamHI* (plasmid no.1543 corresponding to virus A4).

**Virus A5.** *LoxP-ATG-less-neo* fragment from plasmid no.1541 was subcloned in a modified pMSCVneo depleted of the neomycin gene (plasmid no.1598). A *Pgk-puro* fragment (*XhoI-ClaI*) was extracted from pMSCVpuro and introduced upstream of *LoxP-ATG-less-neo* in plasmid no.1598 (plasmid no.1599 corresponding to virus A5).

**Virus S2.** One of the two *Pgk* (the one located upstream of the *loxP*) was removed from plasmid no.1537 (virus S5) creating plasmid no.1638 corresponding to the virus S2.

**Virus S3.** Hygromycin gene was PCR-amplified (primers: gctctagaatgaaaaagcctgaactc and tatctagactattcctttgcctcg) from pMSCVhyg. The hygromycin PCR product was subcloned downstream of *Pgk-kozac-ATG-LoxP* in plasmid no.1535 (plasmid no.1639). The *Pgk* of plasmid no.1639 was removed (plasmid no.1640). The *kozac-ATG-loxP-hygro* fragment from plasmid no.1640 was ligated in a modified pMSCVneo depleted of the neomycin gene (plasmid no.1641 corresponding to virus S3).

**Virus A2.** *Pgk-puro* fragment (*XhoI-ClaI*) obtained from pMSCVpuro was subcloned upstream of *LoxP-ATG-less-neo* in plasmid no.1541 (plasmid no.1600). The *Pgk-puro-LoxP-ATG-less-neo* fragment from plasmid no.1600 was subcloned into pRETRO-SUPER<sup>149</sup> linearized with *BgIII-ClaI* (plasmid no.1644 corresponding to virus A2).

**Virus A3.** The SV40 early mRNA polyadenylation signal (pA) from pDsRed2-N1 (Clontech), included in a 1 kb *HpaI-StuI* fragment, was subcloned downstream of *LoxP-ATG-less-neo* in plasmid no.1541 (plasmid no.1645). *LoxP-ATG-less-neo-pA* fragment from plasmid no.1645 was subcloned in reverse orientation in pMSCVpuro linearized with *EcoRI-HindIII* (plasmid no.1648 corresponding to virus A3).

**Construction of viruses S1 and A1; ESC culture; viral producer cell lines, transduction of target cells & DNA and RNA analyses.** Performed as described previously (Chapter 2 article)<sup>118</sup>. Viruses A4, S4, and S5 were produced by 293GPG packaging cells while A1, A2, A3, A5, S1, S2, and S3 were generated in GP+E-86<sup>118</sup>.

## Acknowledgments

We thank Andras Nagy for providing R1 ESCs and the pCX-EYFP construct (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto); Robert G. Hawley for the MSCV vectors (The George Washington University Medical Center); Maarten van Lohuizen for the pRETRO-SUPER construct (The Netherlands Cancer Institute), and Bruno Saint-Jore (INSERM U321, Paris) for *Pgk-LoxP-Pgk-neo* construct. Mélanie Bilodeau is a recipient of a Canadian Institutes of Health Research (CIHR) studentship and Guy Sauvageau is a recipient of a Canada Research Chair in molecular genetics of stem cells and a scholar of the Leukemia Lymphoma Society of America.

## Supplementary Figure

### Figure 0-5 Successful retroviruses and efficient Cre-loxP recombination.

(a) Cre-mediated recombination between appropriate viruses displayed at the molecular level. S1-A1 or S2-A1, recombination products between indicated proviruses (v.). (b) Southern blot analysis performed with genomic DNA extracted either from GP+E-86 packaging cell lines (G) or ESCs, and probed with a hygromycin probe. *KpnI* restriction digests, assessing proviral integrity, show 2.9 kb and 3.3 kb bands corresponding to S1 and S2 viruses, respectively. *BglII* restriction digests revealed the polyclonality of integration sites in packaging cell lines. (c) Southern blot analysis using a neomycin probe, to view the A1 virus integrity (*KpnI* restriction digests; 3.4 kb band) and polyclonality of integration sites in the packaging cell line genomic DNA (*XbaI* restriction digest). (d) Southern blot analysis using a neomycin probe, performed with 14 ESC clones. *KpnI* and *EcoRI* restriction digests evaluated integrity and single integration of A1 proviruses, respectively. (e) Southern blot analysis of genomic DNA showing vectors' complementary when NIH 3T3 cells are successively infected with indicated viruses, and exposed to transient Cre expression. Upper panel shows 3.4 kb or 3.0 kb signals with *KpnI* restriction digests, corresponding to unarranged A1 provirus or recombination between v.A1 and v.S1 or v.S2, respectively. Bottom panel (*NcoI* restriction digests): 0.6 kb signals represent rearrangement between v. S1 and v.A1 (lane 6) or v.S2 and v.A1 (lane 9), and polyclonality smears (not detectable) correspond to unrecombined A1 viruses (lane 2-5 and 7-8). Although both virus combinations gave the expected recombination products, qualitatively, the v.A1-v.S1 pair was slightly more efficient than the v.A1-v.S2 pair for giving rise to G418<sup>R</sup> NIH 3T3 cells. Consequently, this combination (v.A1-v.S1) was further used for engineering ESCs.





## Supplementary Methods

**Cell culture.** R1 ESC culture was described previously (Chapter 2 article)<sup>118</sup>. NIH 3T3 cells were maintained under standard conditions.

**Viral producer cell lines and transduction of target cells.** Viral producer cell lines and ESC viral transduction were described previously (Chapter 2 article)<sup>118</sup>. For NIH 3T3 cells viral transduction, producers media were changed 24h before infection (10% NCS in DMEM). 2 dishes (100 mm) of NIH 3T3 (10% confluence) were infected during 24h with v.A1 using 1:20 and 1:200 dilutions of viral supernatant in presence of 6  $\mu\text{g ml}^{-1}$  polybren (Sigma). Fresh media was added the next day and puromycin selection (Sigma, 1.2  $\mu\text{g ml}^{-1}$ ) started 48h after infection. Five dishes (100 mm) of each populations (50% confluence) were infected during 24h with v.S1 or v.S2 in presence of 6  $\mu\text{g ml}^{-1}$  polybren (Sigma) (dilution of viral supernatant; 1:2). The next day, cells were split, half were frozen, and the rest plated in 4 dishes (100 mm).

**Cre-loxP recombination in NIH 3T3 cells.** 48 h after the beginning of the second infection, 3 dishes (100 mm, >90% confluence) were transfected with circular pCX-EYFP<sup>154</sup> (1 dish) or pCX-cre<sup>118</sup> (2 dishes) using Lipofectamine (Invitrogen). The next day, cells were split and G418 selection (Invitrogen, 1mg  $\text{ml}^{-1}$ ) was started after 48h. As expected, every cells transfected with pCX-EYFP were sensitive to G418.

## APPENDIX III: Table

**Table presenting the genomic features of speculative 3 Mb-deletions anchored to virus A1 retroviral integration sites determined by I-PCR (part 1 of 4)**

Primary clone Id	Anchor Chr	3Mb deletion Start	3Mb deletion Stop	Number of CpG islands	Number of Refseq genes	Number of microRNAs	Number of mRNAs	Number of highly conserved regions
5001	chr17	30507974	33507974	43	62	0	545	196
5002	chr3	121802436	124802436	16	17	1	179	239
5003	chr14	53526031	56526031	56	91	2	735	438
5004	chr10	79449174	82449174	151	102	0	1735	351
5005	chr3	144547055	147547055	19	31	0	238	289
5006	chr5	22919519	25919519	41	33	1	445	368
5009	chr14	100119811	103119811	9	10	0	269	249
5012	chr19	31101425	34101425	9	14	0	223	116
5014	chr8	1347514	4347514	27	35	0	352	137
5017	chr9	7720889	10720889	10	9	0	118	165
5018	chr7	11840399	14840399	14	25	0	166	128
5020	chr2	73438698	76438698	32	29	1	244	559
5021	chr11	80905934	83905934	26	51	0	420	279
5024	chr10	76447973	79447973	67	81	0	878	184
5025	chr15	12662598	15662598	4	3	0	55	190
5026	chr3	135937211	138937211	10	21	0	255	187
5027	chr6	148168602	151168602	10	11	0	159	94
5029	chr8	82534844	85534844	9	11	0	115	255
5034	chr15	79839426	82839426	57	63	1	868	360
5035	chr8	35118318	38118318	19	18	0	224	158
5038	chr11	120670373	123670373	16	26	0	250	77
5040	chr2	120048741	123048741	40	52	0	537	484
5041	chr7	118086289	121086289	23	44	0	427	258
5042	chr3	103259159	106259159	25	39	0	376	348
5043	chr12	100588873	103588873	21	24	0	350	297

In blue, A1 retroviral integration sites related to families with G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In red, A1 retroviral integration sites related to families with G418<sup>R</sup> tertiary clones, but no G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In black, A1 retroviral integration sites that are not related to families with G418<sup>R</sup> tertiary clones. Id, identification; Chr, chromosome; Mb, megabase pairs. Data extracted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32,152,153</sup>.

---

**Part 2 of 4**


---

Primary clone Id	Anchor Chr	3Mb deletion Start	3Mb deletion Stop	Number of CpG islands	Number of Refseq genes	Number of microRNAs	Number of mRNAs	Number of highly conserved regions
5045	chr10	63400102	66400102	0	2	0	44	87
5046	chr9	117614650	120614650	36	39	1	391	270
5047	chr17	113312	3113312	2	1	0	31	20
5049	chr11	98712127	101712127	68	116	0	1078	578
5050	chr4	50885955	53885955	8	14	0	99	188
5051	chr10	79792992	82792992	117	82	0	1561	300
5055	chrX	95506072	98506072	24	34	1	311	343
5056	chr5	118520891	121520891	29	30	0	313	265
5057	chr11	50215746	53215746	33	45	0	434	279
5058	chr12	78144677	81144677	18	17	0	215	295
5059	chrX	165315983	168315983	6	2	0	31	6
5061	chr4	56984276	59984276	22	29	1	284	271
5062	chr11	35263024	38263024	10	6	2	83	606
5064	chr1	153951498	156951498	18	24	0	279	230
5065	chr18	37434159	40434159	46	61	0	441	234
5066	chr7	11732942	14732942	18	30	0	211	133
5068	chr7	132432736	135432736	16	23	0	268	214
5071	chr7	2862669	5862669	45	82	0	543	157
5074	chr7	109839356	112839356	18	20	0	289	349
5075	chr7	95913682	98913682	25	26	1	274	342
5076	chr7	18344700	21344700	47	41	0	374	125
5077	chr9	69920936	72920936	26	29	0	377	432
5080	chr1	15775440	18775440	13	18	0	136	216
5082	chr6	97348725	100348725	9	8	0	134	537
5083	chr16	13440516	16440516	23	31	1	298	324

In blue, A1 retroviral integration sites related to families with G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In red, A1 retroviral integration sites related to families with G418<sup>R</sup> tertiary clones, but no G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In black, A1 retroviral integration sites that are not related to families with G418<sup>R</sup> tertiary clones. Id, identification; Chr, chromosome; Mb, megabase pairs. Data extracted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32,152,153</sup>.

---

---

**Part 3 of 4**


---

Primary clone Id	Anchor Chr	3Mb deletion Start	3Mb deletion Stop	Number of CpG islands	Number of Refseq genes	Number of microRNAs	Number of mRNAs	Number of highly conserved regions
5086	chr6	140107733	143107733	11	29	0	181	246
5088	chr16	42675681	45675681	16	31	0	310	398
5091	chr8	41807504	44807504	9	21	0	186	178
5092	chr14	65726128	68726128	23	30	0	296	320
5094	chr1	20804732	23804732	12	18	2	128	171
5108	chr3	95991528	98991528	53	66	0	541	316
5120	chr13	31689643	34689643	13	25	0	183	153
5194	chr7	96820439	99820439	36	41	1	418	316
5196	chr6	125421248	128421248	23	29	0	314	203
5197	chr7	11956736	14956736	3	18	0	63	120
5198	chr9	24525625	27525625	11	12	0	162	182
5199	chr17	26764953	29764953	73	58	0	733	384
5204	chr4	41017224	44017224	55	79	1	855	425
5205	chr6	71222644	74222644	22	38	0	472	261
5211	chr11	59329278	62329278	52	60	0	633	390
5212	chr11	1396196	4396196	27	36	0	462	205
5213	chr15	98819220	101819220	64	83	0	749	427
5218	chr18	74979316	77979316	28	19	0	258	306
5226	chr2	30983557	33983557	65	55	1	591	492
5227	chr11	48463840	51463840	41	74	2	563	273
5229	chr17	512972	3512972	4	3	0	71	61
5233	chr10	57208590	60208590	26	33	0	512	286
5237	chr8	78699072	81699072	10	9	0	123	278
5238	chr12	87690206	90690206	16	20	0	263	339
5239	chr5	136320687	139320687	53	81	5	654	294

In blue, A1 retroviral integration sites related to families with G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In red, A1 retroviral integration sites related to families with G418<sup>R</sup> tertiary clones, but no G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In black, A1 retroviral integration sites that are not related to families with G418<sup>R</sup> tertiary clones. Id, identification; Chr, chromosome; Mb, megabase pairs. Data extracted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32,152,153</sup>.

---

---

**Part 4 of 4**


---

Primary clone Id	Anchor Chr	3Mb deletion Start	3Mb deletion Stop	Number of CpG islands	Number of Refseq genes	Number of microRNAs	Number of mRNAs	Number of highly conserved regions
5240	chr3	3906705	6906705	2	2	0	32	374
5242	chr1	135279729	138279729	44	51	0	522	321
5244	chr7	94304552	97304552	7	7	1	80	209
5248	chr11	66867224	69867224	68	95	1	857	699
5252	chr13	98168963	101168963	24	21	0	261	258
5253	chr12	88088474	91088474	9	13	0	188	328
5254	chr17	83783958	86783958	30	16	0	171	383
5255	chr14	16005222	19005222	13	12	0	143	140
5256	chr4	125968987	128968987	45	41	0	443	374
5258	chr11	106585156	109585156	27	29	0	365	310
5259	chr14	6848872	9848872	5	10	0	97	213
5260	chr15	66901645	69901645	4	3	2	43	85
5261	chr6	95681528	98681528	7	11	0	124	346
5265	chr11	115178073	118178073	83	76	0	967	424
5266	chr19	4163981	7163981	122	128	2	1486	523
5269	chr5	75329547	78329547	20	23	0	227	229
5270	chr1	178550084	181550084	18	17	1	239	287
5271	chr16	16435791	19435791	59	70	3	987	357
5273	chr2	84628411	87628411	10	135	0	145	82
5276	chr12	98234055	101234055	20	18	0	269	292
5277	chr9	23748642	26748642	10	13	0	128	137
5278	chr2	91161466	94161466	41	38	2	437	398
5280	chr4	133587987	136587987	60	54	1	637	330
5282	chr18	66906764	69906764	21	23	0	283	396
5286	chr4	55492994	58492994	17	21	1	229	221
5287	chr2	174697742	177697742	0	5	0	274	17
5288	chr15	72717127	75717127	32	42	2	272	157

In blue, A1 retroviral integration sites related to families with G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In red, A1 retroviral integration sites related to families with G418<sup>R</sup> tertiary clones, but no G418<sup>R</sup> puro<sup>S</sup> tertiary clones. In black, A1 retroviral integration sites that are not related to families with G418<sup>R</sup> tertiary clones. Id, identification; Chr, chromosome; Mb, megabase pairs. Data extracted from UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>32,152,153</sup>.

---

