Virtual Karyotyping of Genetically Diverse Mouse Embryonic Stem Cells Reveals Chromosomal Sex Bias in Genomic Stability and Totipotent Subpopulations

A thesis submitted by

Alexander R. Stanton

in partial fulfillment of the requirements for the degree of

Ph.D.

in

Genetics

Tufts University

Graduate School of Biomedical Sciences

February 2023

Advisor: Martin F. Pera, Ph.D.

Abstract

In vitro culture of pluripotent stem cells (PSCs) is an invaluable tool for both research and clinical use - such as the ability to generate developmental or disease models and be employed in cell replacement therapy. However, these cell types are prone to developing chromosomal aberrations over time that can inhibit differentiation potential and provide a proliferative/survival advantage, posing a severe hurdle for research applications and dangers for clinical ones. Not all cell lines are as susceptible to developing these abnormalities, but little is known about the underpinnings of this variability. Here, bulk RNA-seq and mass spectrometry proteomics data sets from a panel of 170 mouse embryonic stem cell (mESC) lines derived from the Diversity Outbred (DO) mouse population were assayed for abnormal karyotype by multi-omic virtual karyotyping and interrogated for common genetic features of the aneuploid lines. Our methodology identified 62 lines harboring detectable aneuploid subpopulations consisting primarily of duplications for Chromosomes 8 and 11, and to a lesser degree Chromosome 1 and 6. We also identified 21 of the 88 originally XX cell lines in which a portion of the culture lost a Chromosome X (XO). No loci were found to be associated with duplication of specific chromosomes, however a quantitative trait loci (QTL) at Chromosome X was significantly associated with overall aneuploidy. Most strikingly, a significantly lower proportion of XX lines were aneuploid, compared to the XY or XO lines, and the proportion of aneuploid lines in these populations correlated to the expression of Chromosome X genes. These findings were corroborated by relatively lower expression of X-linked genes in the trisomic cells from a scRNA-seq data set of an additional 12 XY DO mESC lines and by a lower proportion of aneuploidies in human PSCs karyotyped by the WiCell Research Institute. Additionally, XX lines showed higher expression of genes upregulated in 2-cell-like cells (2CLCs) and several X-linked tumor

suppressor genes with associations to DNA damage repair. Follow up scRNA-seq of early and aged XX, XY, and XO isogenic cell lines validated a reduced accumulation of aneuploidies in XX lines where the remaining euploid cells were enriched for genes indicative of naïve pluripotency, suggesting a role for pluripotency state in maintaining genomic stability.

Dedication

To my family and friends who are the only reason I have been able to be here and make this accomplishment.

Acknowledgements

First, I would like to thank my thesis advisor Martin Pera for providing a wonderful and supportive space to develop as a growing scientist during doctoral training. Matin's mentorship meant having the creative freedom to explore scientific inquiries as they inspired me, while always being there to ask questions and provide insight and encouragement when needed. In a world that too easily builds imposter syndrome, this kind of training is one that helps maintain confident curiosity and a framework for positive growth. For this I am incredibly grateful.

Second, I would like to thank my thesis committee members past and present: Gregory Carter, Donna Slonim, Christopher Baker, and Gregory Cox. Every single one of them has been there through every challenge I faced during my doctoral studies to ensure I had the resources and guidance necessary to achieve my goals. Even when there seemed to be no light at the end of tunnel and it was hard to keep moving along, my committee found tough, but ultimately necessary, ways to make sure I reached the end.

Third, I would like to thank the Jackson Laboratory community. No matter how large or small of a role, there were numerous people who played a part in my personal and professional development. Thank you to Steven Munger, Selcan Aydin, Lisa Somes, Catherine Brunton, Tracy McGarr, Daniel Cortes, Nurten Akturk, Queen Igomu, Sammira Rais-Rohani – the lab members who housed me, fielded daily questions, and were consistent moral support during my time at The Jackson Laboratory. Thank you to Laura Reinholdt, Gary Churchill, Daniel Skelly, Anne Czechanski, Greg Keele, Duy Pham, Candice Byers, Lauren Kuffler, Catrina Spruce, and other mmbers already mentioned from the Seq-ers group that helped me in various ways to develop my thesis project into what it became. All your expertise, suggestions, and encouragement were invaluable

guidance. Thank you to all the graduate students who gone through everything a PhD has to offer and more alongside me – especially Micheal McLellan and Kate Foley who have been the most incredible cohort one could ask for. Thank you to everyone in scientific services, education departments, and administration who made sure everything kept running smoothly.

Fourth, I would like to thank everyone at other institutions who've been an integral part of my research. Thank you to Peter Andrews, Ivana Barbaric, and Dylan Stavish of Sheffeild University for providing council on human models, with which I'm less acquainted. Thank you to Tenneille Ludwig, Kim Leonhard, Seth Taapken, and Erik McIntire at The WiCell Institute for validating data we collected and discussing bigger connections from our findings. And thank you to Ted Choi of Predictive Biology, without whom the data that spurred my project would not exist.

Fifth, I would like to thank all of my friends and family who are the reasons I am where I am today. Thank you to my mother, Sherry Swenson, who raised me as a single parent with unwavering support and commitment and who has never ceased being supportive of all my life choices – even when they lead to furthest state away from home for the last seven years. Thank you to Robert and Maureen Minner, Stacey, Pat, Brett, Ryan, and Nick Poliquin. You have been the best tight knit family anyone could ask for and have always been there for me since I can remember. Thank you to the many other distant family members who have been there for me, no matter the geographic separation.

Lastly, thank you to my partner Heather Brown. Without you, none of this would be possible. You have been an invaluable font of reassurance, inspiration, and a helping hand when everything felt too insurmountable.

It is impossible to enumerate everyone who has made me or my project what they are today. Thank you to the countless others for everything you've done for me.

Table of Contents

Title Page i

Abstract ii

Dedication iv

Acknowledgements v

Table of Contents viii

List of Tables xi

List of Figures xii

List of Abbreviations xiv

Chapter 1: Introduction 1
1.1. Early embryonic development 1
1.1.1. Narrowing differentiation potential 1
1.1.2. Applications for ESCs and iPSCs 3
1.2. Genomic instability during prolonged culture 5
1.2.1. Aberrations and consequences 5
1.2.2. Methodologies for aneuploid screening 7
1.2.2.1. Direct karyotyping 8
1.2.2.2. Indirect karyotyping 9
1.3. Heterogeneity, natural variations, and effects of culture practices 11
1.3.1. Genetic background 11
1.3.2. Chromosomal sex 12
1.3.3. Variability and subpopulations 14
1.3.4. Culture conditions 17
1.4. Project rationale 20
1.5. Research projects 20
1.5.1. Annotation and characterization of CNV trends in DO mESC data sets 21
1.5.2. Analysis of factors associated with the development of aneuploidy in the DO mESC data sets 21
1.5.3. Characterizing the sexual dimorphism in aneuploidy development of mESCs 22

Chapter 2: Materials and Methods 23
2.1. DO mESC derivation and culture 23
2.2. DO mESC paired-end bulk RNA-seq 23
2.3. DO mESC mass spectrometry proteomics 23
2.4. DO mESC single-cell RNA-seq 24
2.5. Virtual karyotyping using bulk RNA-seq and proteomics 24
2.6. Virtual karyotyping using single-cell RNA-seq 24
2.7. Estimating aneuploid proportions 25
2.8. Mouse-Human Synteny 26
2.9. Differential Expression Analyses of DO mESCs 26
2.10. Allele-Specific Expression (ASE) Analyses 26
2.11. Genetic Mapping Analyses 28

2.12. 2CLC Gene Expression 28
2.13. Inbred mESC derivation and culture 28
2.14. Inbred mESC flow cytometry 29
2.15. Inbred mESC single-cell RNA-seq and analyses 29
2.15.1. Early and aged data set processing 30
2.15.2. Cell cycle phase annotations 31
2.15.3. Identification and characterization of pluripotent state, differentiation primed, MEF, and 2CLC clusters 31
2.15.4. Virtual karyotyping using single-cell RNA-seq 32
2.15.5. Differential expression in aged euploid XX populations 32
2.16. Author contributions 32

Chapter 3: Results 33
3.1. Annotation and characterization of CNV trends in DO mESC data sets 33
3.1.1. Characterization of DO mESC data sets 33
3.1.2. Virtual karyotyping with RNA-seq and proteomics 34
3.1.2.1. Evidence of chromosomal copy number variation 34
3.1.2.2. Clustering and annotation of autosome duplications using a multi-omic approach 38
3.1.2.3. Summary of aneuploid prevalence in bulk data sets 39
3.1.2.4. Annotation of XX lines becoming mosaic XX/XO 40
3.1.3. Virtual karyotyping with scRNA-seq 41
3.1.3.1. Single-cell RNA-seq virtual karyotyping methodology 41
3.1.3.2. Summary of aneuploid prevalence in scRNA-seq 42
3.1.4. Estimated aneuploid proportions of cell lines from bulk data sets 44
3.1.4.1. Comparing gene expression differences on duplicated chromosomes in single cells 44
3.1.4.2. Calculating aneuploid proportion using empirical reference 46
3.2. Analysis of factors associated with the development of aneuploidy in the DO mESC 47
3.2.1. Comparison with hPSC duplications 47
3.2.2. Haplotype analysis of allele-specific expression data 50
3.2.3. Genetic mapping of aneuploidy as a binary trait 53
3.2.4. Chromosomal sex bias in DO mESC aneuploidy 55
3.2.5. Chromosomal sex bias in hPSC aneuploidy 58
3.2.6. Two active X Chromosomes increases 2CLC gene expression 61
3.2.7. Two active X Chromosomes increases expression of tumor suppressor genes involved in DNA damage response 64
3.3. Characterizing the sexual dimorphism in aneuploidy development of mESCs 68
3.3.1. Experimental design for further chromosomal sex bias investigation 68
3.3.2. Long-term low dose retinoic acid treatment yields differentiation 69
3.3.3. Identification of pluripotent cell clusters and demultiplexing chromosomal sex by gene expression 74
3.3.4. Aneuploidies accumulate at a slower rate in XX C3H/HeSn-Paf/J PSCs 78
3.3.5. Aged XX mESCs are enriched for naïve pluripotency-associated genes 80
3.4. Author contributions 82

Chapter 4: Discussion 83
4.1. Proteomics provides additional modalities for virtual karyotyping approaches 83
4.2. Prevalence and timing of aneuploidies developing in culture 84

4.3. Identifying genetic determinants of genomic instability 85
4.3.1. No determinants of specific chromosomal duplications 85
4.3.2. A potentially reciprocal relationship between Chromosome X dosage and autosomal CNVs 86
4.4. Intersections of chromosomal sex with cell culture and heterogeneity 88
4.4.1. The naive culture conditions are not the same as the XX naïve state 88
4.4.2. Prospective improvements to naïve culture media 89
4.5. Conclusions and Future Directions 91

Chapter 5: Bibliography 92

List of Tables

Table 3.1 - Shapiro-Wilk Normality Test results for each chromosome and stage of gene expression 36
Table 3.2 - Aneuploidies in WiCell Institute karyotyped hPSCs 60

List of Figures

Figure 3.1 - DO mESC data sets. 33
Figure 3.2 - Accumulated chromosomal copy number variations can be observed through chromosome-wide gene expression changes. 34
Figure 3.3 - Recurrent CNVs disrupt the normal distribution of median chromosome expression in the DO mESC population. 35
Figure 3.4 - Workflow for virtual karyotyping using RNA-seq and proteomics. 38
Figure 3.5 - Distribution and annotation of DO mESC lines from chromosomes 1, 6, 8, and 11. 40
Figure 3.6 - Autosomal aneuploidies detected in DO mESC lines using bulk data sets are predominantly single chromosome aneuploidies and duplications of chromosomes 8 and 11. 41
Figure 3.7 - Distribution and annotation of originally XX DO mESC lines for chromosome X. 42
Figure 3.8 - Autosomal aneuploidies detected in DO mESC lines using scRNA-seq largely recapitulate the patterns using bulk data sets. 43
Figure 3.9 - Autosome duplications cause attenuated expression increases in the respective chromosomes, but not in other non-duplicated autosomes. 45
Figure 3.10 - Calculations of aneuploid content estimate low percentages of abnormal cells. 47
Figure 3.11 - Recurrent mESC duplications except chromosome 8 share a many of genes with hPSC duplications. 49
Figure 3.12 - Differentially expressed pathways in trisomy 8 single cells. 50
Figure 3.13 – No distinctly overrepresented founder alleles on specifically duplicated chromosomes. 52
Figure 3.14 - Chromosome X is implicated in the development of general aneuploidy. 54
Figure 3.15 - XX DO mESCs have a lower proportion of aneuploid lines compared to XY or XO lines. 55
Figure 3.16 - Individual Aneuploid XY cells have lower chromosome X expression than euploid cells. 56
Figure 3.17 - Directionality of X-linked gene expression differences is shared between individual chromosome trisomies. 59
Figure 3.18 - Fewer XX hiPSC lines, but not hESCs, are aneuploid in G-band karyotyped data from the WiCell Institute. 60
Figure 3.19 - Chromosome X gene Dusp9 is highly dose-sensitive and expressed lower in aneuploid lines. 62
Figure 3.20 - XX DO mESCs have increased 2CLC gene expression profiles compared to XY or XO. 63
Figure 3.21 - XX mESC lines have a higher proportion of 2CLCs compared to XY or XO mESC lines from the same genetic background. 64
Figure 3.22 - Chromosome X tumor suppressor gene Fhl1 is dose-sensitive and expressed lower in aneuploid lines. 66
Figure 3.23 - Aneuploid cells have similarly decreased expression of several chromosome X TSGs. 67
Figure 3.24 - Schematic for aging, treatment, and assays for XX, XO, and XY C3H/HeSn-Paf/J mESC lines. 68
Figure 3.25 - Retinoic acid (RA) induction of 2CL state in XY mESCs. 71
Figure 3.26 - 24-hour treatment with 100nM retinoic acid (RA) promotes lasting 2CLC increase. 72

Figure 3.27 - Sudden spontaneous differentiation occurs in XY mESCs following intermittent retinoic acid treatment. 73
Figure 3.28 – Identification of non-pluripotent scRNA-seq cell clusters. 75
Figure 3.29 - Identification of XX and XY pluripotent mESC cells using unsupervised clustering. 76
Figure 3.30 - Changing karyotype landscape between early and aged mESCs 78
Figure 3.31 - Less accumulation of autosomal aneuploidies occurs in XX mESCs over time in culture. 80
Figure 3.32 – Differentially expressed pluripotency and tumor suppressor genes. 81

List of Abbreviations

2C – 2-cell
2CL – 2C-like
2CLC – 2CL cells
2i – 2 inhibitors
4C – 4-cell
5iLA – 5 inhibitors, LIF, and activin
5iLAF – 5iLA and Fgf
8C – 8-cell
8CL – 8C-like
8CLC – 8CL cells
a2i – alternative 2i
Amer1 – APC membrane recruitment protein 1
ASE – Allele-specific expression
BCL-XL – B-cell lymphoma-extra large
Bcorl1 – BCL6 corepressor like 1
Bex2 – Brain expressed X-linked 2
BSA – Bovine serum albumin
Btk – Bruton tyrosine kinase
cDNA – Copy DNA
CDX2 – Caudal type homeobox 2
CGH – Comparative genome hybridization
CITE-seq – Cellular indexing of transcriptomes and epitopes by sequencing
CMA – Chromosomal microarray
CNV – Copy number variation
CONICSmat – Copy-number analysis in single-cell RNA-seq from matrix
Dax1 – Dosage sensitive sex reversal, adrenal hypoplasia critical region, on Chromosome X, gene 1
Ddx3x – DEAD-box helicase 3 X-linked
Dmd – Dystrophin
DMSO – Dimethyl sulfoxide
Dnmt3a/b – DNA methyltransferase
DO – Diversity Outbred
Dppa3 – Developmental pluripotency associated factor 3
Dusp6/9 – Dual specificity phosphatase 6/9
Eda2r – Ectodysplasin A2 receptor
EDTA – Ethylenediaminetetraacetic acid
Eras – Embryonic stem cell expressed Ras
Esrrb – Estrogen related receptor beta
ERV – Endogenous retrovirus
ESC – Embryonic stem cell
FBS – Fetal bovine serum
Fgf4/5 – Fibroblast growth factor 4/5
Fhl1 – Four and a half LIM domains 1
FISH – Fluorescence in situ hybridization
Flna – Filamin A
Foxo4 – Forkhead box protein O4
Foxp3 – Forkhead box protein P3
GBRS - Genotype reconstruction by RNA-sequencing

GENCODE – Generalized coding (graphics communications association)
GO – Gene ontology
Gpc3 – Glypican 3
Gsk3 – Glycogen synthase kinase
Gata3/6 – GATA-binding protein 3/6
hESC – Human ESC
hPSC – Human PSC
Hormad1 – HORMA domain containing 1
ICM – Inner cell mass
IgG – Immunoglobulin
iPSC – Induced PSC
JNK – c-Jun-N terminal kinase
Kdm6a – Lysine demethylase 6A
Klf2/4/5 – Krüppel-like factor 2/4/5
Lefty1 – Left-right determination factor 1
LIF – Leukemia inhibitory factor
LMO – Lipid modified oligonucleotide
LOD – Linkage of disequilibrium
LTR – Long terminal repeat
MAPK – Mitogen activated protein kinase
MEF – Murine embryonic fibroblasts
Mek1/2 – MAPK kinase
mERV – Mouse ERV
mESC – Mouse ESC
mPSC – Mouse PSC
MGI – Mouse Genome Informatics
MLL1 – Mixed lineage leukemia 1
MULTI-seq – scRNA-seq and snRNA-seq sample multiplexing using lipid-tagging indices
MZT – Maternal-to-zygotic transition
NHSM – Naïve human stem cell medium
Nono – Non-POU domain containing octamer binding
Nr0b1 – Nuclear receptor subfamily 0 group B member 1
Nr5a2 – Nuclear receptor subfamily 5 group A member 2
Oct3/4 – Octamer-binding transcription factor 3/4
OSKM – Oct4, Sox2, Klf4, Myc
Otx2 – Orthodenticle homeobox 2
PCR – Polymerase chain reaction
PFA – Paraformaldehyde
Phf6 – PHD finger protein 6
PKC – Protein kinase C
Pou3f1 – POU domain, class 3, transcription factor 1
Pou5f1 – POU domain, class 5, transcription factor 1 (alternatively Oct3/4)
Prdm1 – PR/SET domain 1
PrE – Primitive endoderm
PSC – Pluripotent stem cell
PXGL – PD0325901, XAV939, Gö6983, and LIF
QQ plot – quantile-quantile plot
RA – Retinoic acid/vitamin A
Rbmx – RNA binding motif protein X-linked
Rbbp7 – RB binding protein 7

RNP - Ribonuceoprotein
ROCK – Rho-associated, coiled-coil-containing protein kinase
Rpl10 – Ribosomal protein 10
Rps6ka6 – Ribosomal protein S6 kinase, 90kDa, polypeptide 6
scRNA-seq – Single-cell RNA-seq
SKY – Spectral karyotyping
SNP – Single nucleotide polymorphism
Sox2 – Sry-box transcription factor 2
SSEA4 – Stage specific embryonic antigen 4
Srpx – Sushi repeat containing protein X-linked
T – T box transcription factor T
t2iL+Gö – Titrated 2i, LIF, Gö6983
Tbx3 – T box transcription factor 3
Tceal7 – Transcription elongation factor A like 7
TE – Trophectoderm
Tet2 – Tet methylcytosine dioxygenase 2
Tfe3 – Transcription factor binding to IGHM enhancer 3
Thy1 – Thymocyte antigen 1
TNKS1/2 – Tankyrase 1/2
Trex2 – Three prime repair exonuclease 2
TSG – Tumor suppressor gene
Uhrf1 – Ubiquitin-like with PHD and ring finger domains 1
WGS – Whole genome sequencing
X – Chromosome X
Xa – X active
XCI – X Chromosome inactivation
XCR – X Chromosome reactivation
Xe – X eroded
Xi – X inactive
XIE – Xi erosion
Xist – Xi specific transcript
Xlr3a/c – X-linked lymphocyte regulated 3A/C
Yap1 – Yes associated protein 1
Zfp185 – Zinc finger protein 185
Zfx – Zinc finger protein X-linked
ZGA – Zygotic genome activation
Zscan4 – Zinc finger and SCAN domain containing 4

Chapter 1: Introduction

1.1. Early embryonic development

1.1.1. Narrowing differentiation potential

Cells of the developing embryo hold within them an unrestricted potential for cell fate shortly after fertilization which becomes increasingly restricted over time. Not long after the formation of the zygote, a two-stage maternal-to-zygotic transition (MZT) occurs wherein the zygote switches from maternal transcripts and proteins supplied by the egg to its own actively expressed genome. Soon a subset of these cells will become committed to an extraembryonic developmental trajectory and the remaining pool of embryonic cells will have those lineages foreclosed from them. These two populations of embryonic cells have become the subject of intense study and application due to their ability to become any cell type of the embryo or body.

The MZT process is accomplished via the zygotic genome activation (ZGA) minor and major waves, resulting in cells that are totipotent: capable of developing into trophectoderm (TE) and primitive endoderm (PrE) embryonic support membranes as well as endoderm, mesoderm, and ectoderm germ layers of the developing embryo. Low-level, enhancer/promoter-independent transcription using maternally supplied machinery begins to alleviate the existing repressive chromatin states during minor ZGA. During major ZGA the repressive genome methylation and histone marks are largely removed, ushering in more abundant transcription of thousands of genes from the zygotic genome (Burton and Torres-Padilla, 2014; Zhou and Dean, 2015; Lim et al, 2016). The developmental timing of these processes varies by species. Here, as in the rest of this chapter, we'll focus on mice and humans since the majority of stem cell research uses these two organisms. In mice, minor ZGA begins in the 1-cell state hours after fertilization, whereas in humans minor ZGA doesn't begin until the 2-cell (2C) or 4-

cell (4C) stage (Aoki et al, 1997; Bouniol et al, 1995; Matsumoto et al, 1994; Vassena et al, 2011; Xue et al, 2013). Consequently, the hypomethylation and active chromatin status that facilitates the broad differentiation potential of totipotency is established by the 2C stage in mice and around the 8-cell (8C) stage in humans (Guo et al, 2014; Smith et al, 2014; Suwińska 2012). Functionally these totipotent cells can contribute to both extraembryonic as well as embryonic lineages and can be identified experimentally by specific molecular hallmarks such as *Zscan4*-cluster and long terminal repeat (LTR)-flanked endogenous retroviral (ERV) gene expression in addition to unique global hypomethylation (Falco et al, 2007; Reik 2018; Svoboda et al, 2004).

By the formation of the trophoblast and preimplantation epiblast, some totipotent cells have differentiated into the extraembryonic TE and PrE that encase the inner cell mass (ICM) that is comprised of cells that are pluripotent: capable of developing into the three germ layers and the germline. These pluripotent stem cells (PSCs) maintain global hypomethylation and begin to express transcription factors unique to pluripotent cells such as *Nanog*, *Sox2*, and *Oct4* which inhibit *Gata6* expression to distinguish these cells from PrE cells (Chazaud and Yamanaka 2016; Leung and Zernika-Goetz 2015; Reik 2018). Heterogeneity exists in PSC transcriptional profiles mainly between what are considered naïve and primed pluripotency. Naïve PSCs are transcriptionally similar to the early preimplantation epiblasts, while primed PSCs have moved closer to lineage-primed pluripotency similar to postimplantation epiblast cells and might even have trouble forming germ line cells (Nichols and Smith 2009; Singh et al, 2016). The developmental window where the naïve PSCs exist before becoming multipotent (limited to a specific lineage) is short, but researchers have been able to directly capture and maintain these cells or recapitulate them through genetic manipulation.

There are two separate methods for acquiring PSC lines. Embryonic stem cells (ESCs) are PSCs derived from the ICM of the preimplantation epiblast. These cells are already in the naïve pluripotent state and are extracted, transferred to cell culture vessels coated with a layer of feeder cells, and cultured in media supplemented with LIF and/or small molecules that restrict spontaneous differentiation. Induced pluripotent stem cells (iPSCs) are PSCs derived from reprogramming somatic tissues – typically skin fibroblasts because they are easy to acquire. The expression of OSKM or Yamanaka factors (*Oct3/4*, *Sox2*, *Klf4*, and *c-Myc*) in these cells are induced through transfection to stimulate pathways that drive pluripotency gene expression networks (Takahashi and Yamanaka, 2006; Takahashi et al, 2007; Yu et al, 2007). The *in vitro* criteria used to determine the pluripotency status of these derived cell lines is a topic of debate, but generally centers around testing for their continued, long-term self-replication, capacity to differentiate into all embryonic tissues, and transcriptional similarity to PSCs of the preimplantation ICM (Singh et al, 2015a; Pera 2010; Theunissen et al, 2016; Yilmaz and Benvenisty 2019).

1.1.2. Applications for ESCs and iPSCs

Because of their high capacity for self-renewal and broad differentiation capacity, pluripotent stem cells are an important research tool for developmental studies, disease modeling, and generation of transgenic model organisms. PSCs have been used to study the progressive cellular changes during differentiation trajectories from pluripotency to defined terminal cell type (Ben-David et al, 2012). In addition to experiments observing normal development, ESCs or iPSCs generated either from individuals with genetic variants of interest or genetic variants induced through gene editing technologies are incredibly useful for studying how those variants perturb normal processes. For example, several years after iPSCs were first derived, iPSCs from

patients with familial dysautonomia were used to discover how the causal point mutation affected expression of the mutated gene and the downstream consequences on specific cell-type differentiation and function (Lee et al, 2009). These benefits can also be translated to organismal models by inducing desired genetic changes in ESC/iPSCs and transplanting them into the ICM of a developing host embryo where they will differentiate and contribute to a chimeric organism that is a mosaic of host and transplanted cells (Capecchi 2005). These chimeras can then be selectively bred to generate stable transgenic lines with the genetic variant of interest.

PSCs are also an incredibly useful technology in many clinical applications for regenerative medicine. Certain disorders are characterized by the absence, deficiency, or decreased functionality of particular cell types. One example that spans multiple of these classifications is the autoimmune depletion of pancreatic beta cells in type I diabetic patients and the decreased insulin secretion of pancreatic beta cells in type II diabetic patients. Using hESCs, researchers were able to differentiate pancreatic endoderm cells that were glucose-responsive, secreted insulin in physiologically relevant levels, and could prevent induced hyperglycemia when tested in mouse models (Kroon et al, 2008). Other contexts like neurodegenerative disorders can benefit from research into deriving functional tissues to substitute for the dysfunctional ones (reviewed in Ben-David et al, 2012, Mahla 2016; Singh et al, 2015b). Given the ability to reprogram patient fibroblasts, patients can even receive replacement cells derived from their own tissue which reduces the chances of immunological rejection of the replacement cells (Guha et al, 2013; Taylor et al, 2011). Patient-specific iPSCs are also optimal for pharmacological screening before prescribing any drug regiments.

1.2. Genomic instability during prolonged culture

1.2.1. Aberrations and consequences

Despite the incredible modeling and therapeutic potential of PSCs, they are plagued by a predisposition for accumulating aneuploidies – an imbalanced karyotype due to gains or losses in segments or entire chromosomes – over time during culture. The nature of these copy number variations (CNVs) differs based on the species of origin. In mouse PSCs, the most common aneuploidies are gains of Chromosomes 1, 6, 8, and 11, although gains or losses in several other autosomes have been observed with less frequency (Gatzelumendi and Nogues 2014; Sugawara et al, 2006; Guo et al, 2018; Codner et al, 2016; Rebuzzini et al, 2008; Kim et al, 2013). It is also common for XX mPSCs to lose an X Chromosome and somewhat less common for XY mPSCs to lose a Y Chromosome (Choi et al, 2017a; Ooi et al, 2010; Pasque et al, 2018; Zvetkova et al, 2010). Humans have a much more diverse array of aneuploidies and it is common to see gains in small regions, an arm of, or entire Chromosomes 1, 7, 12, 14, 17, 20, and X (Baker et al, 2007; Cowan et al, 2004; Draper et al, 2004; Maitra et al, 2005; Catalina et al, 2008; Lefort et al, 2009; Spits et al, 2008; Mitalipova et al, 2005; and unpublished work from our WiCell Institute collaborators). Because mouse chromosomes are telocentric – having their centromeres at one end of the chromosome – they are more often gained or lost in their entirety. These CNVs can occur through a variety of mitotic segregation errors, including the missegregation of chromosomes incurring double-stranded breaks (Babariya et al, 2017; Palmerola et al, 2022; Potapova and Gorbsky 2017). These abnormal karyotypes can take over cultures and induce undesirable or dangerous phenotypic consequences. It's not understood why, but while some human PSCs gain structural variants quite readily, others can be cultured at length with few

issues – suggesting some intrinsic variability in genomic instability (Amps et al, 2011; Taapken et al, 2011).

Aneuploidies that arise and entrench themselves in PSC cultures primarily provide a selective advantage and an impose a hurdle to differentiation. Studies of common aneuploidies in both mouse and human contexts have demonstrated a significantly increased proliferation rate in aneuploid cells relative to euploid cells of the same line (Ben-David et al, 2014; Zhang et al, 2016). Subsequent research has shown that these cells are not only out-proliferating the euploid cells, but some duplications may also reduce susceptibility to apoptosis through upregulating genes such as *Bcl-xl* on human Chromosome 20 (Avery et al, 2013; Zhang et al, 2019). These phenotypes not only ensure quick usurpation of a cell line, but pose a serious threat to patients who might receive abnormal cells as a result of their increased tumorigenicity (Ben-David et al, 2014; Na et al, 2014; Werbowetski-Oglivie et al, 2009; Zhang et al, 2016). Aneuploid cells can also display a lower potential for differentiation, posing an issue on using them for developmental research and preventing them from contributing to host embryos when developing transgenic models (Zhang et al, 2016). However, some studies have failed to replicate the differentiation block of aneuploidies and some claim that it can be overcome in mosaic euploid-aneuploid cultures (Gatzelumendi and Nogués 2014; Zhang et al, 2016).

These CNVs are not completely a derivative of *in vitro* culture as aneuploidies can be seen in the *in vivo* embryonic model as well. Aneuploidies in the developing embryo are the leading cause of miscarriage during pregnancy as the only viable systemic aneuploidies are Down Syndrome (trisomy 21), Edward Syndrome (trisomy 18), Patau Syndrome (trisomy 13), and Chromosome X copy number variations such as Turner Syndrome (XO) and Klinefelter Syndrome (XXY). Karyotype analysis of recovered

blastocysts from *in vivo* and *in vitro* fertilization showed that only 63-64% of embryos were euploid or low-grade aneuploid mosaics, the rest displayed extensive or complete aneuploidy (Munné et al, 2019). Some researchers have suggested that endogenous aneuploidy correction mechanisms may exist during development, citing the occasional reversal of abnormal karyotype during early preimplantation stages of development (Amano et al, 2015; Bazrgar et al, 2013). However, the frequent development of abnormal karyotypes remains a significant burden on developmental research, therapeutic applications, and successful embryonic development.

1.2.2. Methodologies for aneuploid screening

Because CNVs are so common after long term PSC culture and their effects can be so detrimental to research and health, there is an ongoing conversation regarding how to conduct routine testing among the stem cell community. Studies using PSCs that are not directly related to genomic instability do not always validate the karyotype of their cell lines, which can completely change the interpretation of results. While it is accepted that testing should occur, the specifics of how often and with what technologies is still being sorted out in a way that can improve PSC research without being a significant barrier to conducting research (reviewed in Rossi et al, 2022). Different techniques have different requirements, costs, and can assay different types of aneuploidies at different resolutions, if at all.

Karyotyping analyses can be broken up into two categories, direct and indirect, based on what the technique is measuring to determine CNVs. Here I'll be referring to techniques that visualize or measure DNA content to assess its quantity as direct karyotyping. Techniques that determine copy number without actually interfacing explicitly with the DNA itself will be referred to as indirect karyotyping. The former

category makes up the majority of the options and will likely be more accessible if validation is the only goal.

1.2.2.1. Direct karyotyping

There are a couple variations of cytogenetic techniques to identify CNVs visually. G-banded karyotyping is a technique that involves preparing metaphase spreads of gently trypsinized chromosomes from individual cells, staining them with a Giemsa or Leishman stain that preferentially incorporates into euchromatic regions, then chromosome number and structural arrangements are identified by specific banding patterns (McIntire et al, 2021). G-banding is a labor intensive and cannot identify changes with a resolution lower than 5-10 Mb (McIntire et al, 2020). Fluorescence in situ hybridization (FISH) – often called spectral karyotyping (SKY) or chromosome painting in this context – is a technique also performed on metaphase spreads, utilizing short oligos complimentary to a target genomic region tagged with a fluorescent marker. FISH/SKY assays paint each entire chromosome a unique color based on cohybridization of multiple probes of different colors that can be parsed by image analysis software (Padilla-Nash et al, 2006; Shakoori 2017). These chromosome painting techniques have a higher resolution limit around 1-2 Mb and make it easier to determine the chromosome origins of translocations (Imataka and Arisaka 2012).

Assays using chromosomal microarrays are much more high-throughput and high resolution but have their own limitations. The typical arrays used for this purpose are comparative genome hybridization (CGH) and single-nucleotide polymorphism (SNP) arrays. CMA technologies compare abundance of genomic regions to a euploid reference sample to determine over- or under-representation. This approach can achieve much higher resolution (around 0.5 Mb) but cannot identify structural rearrangements because it has no mechanism to test for position (McIntire et al, 2020).

In a similar vein, quantitative PCR can be used in a similar fashion if the screening is only intending to identify a few targets. Using a few PCR probes representative of often duplicated chromosomes or regions can be informative about key CNVs in a cheap and accessible manner (Codner et al, 2016; Fan and Quake 2007; Tan et al, 2019).

Lastly, whole genome sequencing (WGS) approaches have the least technical drawbacks of the direct karyotyping assays but are much more expensive. Sequencing allows for individual base pair resolution as one would get from CMA platforms. Additionally, WGS affords the ability to identify structural changes from individual reads that map to two previously unconnected regions of the genome. While WGS is the costliest approach, the specifics on cost, ability to detect mosaicism, and so on will depend on what technologies are applied and at what depth samples are being sequenced.

1.2.2.2. Indirect karyotyping

Indirect karyotyping is possible due to downstream consequences of gene dosage on the expression of gene products in the CNV. It's well-studied that copy number changes in genes results in increased transcript and protein. It's on this basis that posttranscriptional buffering of protein abundance was identified in karyotypically abnormal tumor samples (Gonçalves et al, 2017). To date, we do not know of any global mechanism in eukaryotes to buffer changes in gene dosage at the transcript level. However, there does appear to be a bimodal distribution of proteins consisting of one group whose increase in protein matches the copy number change and one whose protein abundance is lower than expected (Chick et al, 2016; Dephoure et al, 2014; Donnelly et al, 2014; Gonçalves et al, 2017; Torres et al, 2007). And generally, interspecies protein abundances correlate much better than interspecies transcript abundance or even intraspecies transcript and protein abundance (Maier et al, 2009;

Vogel and Marcotte 2012). Because the abundance of transcripts tracks so well with copy number of a gene, that means segmental or chromosomal aneuploidy should be reflected in proportionally increased/decreased RNA from the affected genes.

The method that has come from this understanding is called virtual karyotyping (sometimes expression virtual karyotyping) and infers aneuploidy from comparative transcriptome analysis. Virtual karyotyping can be conducted using gene expression microarray or RNA-seq data (Ben-David et al, 2013; Mayshar et al, 2010; Weissbein et al, 2016). While the sensitivity and resolution of virtual karyotyping are on the lower end compared to direct karyotyping techniques, virtual karyotyping can be performed cheaply and can use data that was acquired for purposes other than assessing CNVs (Ben-David et al, 2013). This latter point makes virtual karyotyping not only an efficient use of resources in research settings, but also in clinical settings where cells (i.e. from biopsies) are limited and invasive to come by. In these settings, it may be particularly necessary to limit the number of assays run on a given sample, so being able to acquire gene expression data and karyotype analysis from one technology is extremely helpful. Ideally, gene expression data from known euploid samples would be used as a reference by which to judge expression changes in test samples, but with a large enough population and well-understood genomic changes it is possible to infer euploid and aneuploid lines without reference. Single-cell RNA-seq (scRNA-seq) is particularly useful for this purpose because each cell is either euploid or aneuploid, so it contains its own euploid reference and variation from cell culture mosaics are not present (Griffiths et al, 2017). To our knowledge, only transcript data has been used for virtual karyotyping, although it should be possible with proteomics data – although proteome buffering might decrease sensitivity.

1.3. Heterogeneity, natural variations, and effects of culture practices

PSC cultures have several intrinsic or externally imposed sources of variation that are important to note. Most of these factors are not completely independent of each other and knowing how they intersect better informs our understanding of the complex dynamics within the diverse cellular community of PSC cultures. Where relevant, I will discuss how genetic background, chromosomal sex, different cellular subpopulations, and cell culture media/treatments can influence each other and how they have been shown to relate to aneuploidy.

1.3.1. Genetic background

Historic precedent has led to little diversity in the genetic background of mESCs used in the literature. Originally, the derivation of mESCs from ICM cells could only be achieved using limited mouse strains (Evans and Kaufman 1981; Kawase et al, 1994; Martin 1981). Some mouse strains were not amenable to the first methods using LIF and did not maintain a naïve pluripotent state until later protocols identified necessary inhibitors to facilitate this (Czechanski et al, 2014; Kawase et al, 1994). The strains where ESCs can be derived using only LIF supplementation are referred to as permissive strains and the gap before a method for nonpermissive strain derivation was established meant that many cell lines are only from permissive strains – typically 129S1/SvImJ or C57BL/6J (or their derivatives). For example, in a study of aneuploidy development and its consequences in mESCs, three separate cell lines were used, however they were from the highly related 129/Ola, 129/terSv, and 129/SVEV strains (Gatzelumendi and Nogués 2014). This is similar to studying development, aneuploidy, and so on in one or two human families. Human ESCs and iPSCs on the other hand are much more variable because each person they are collected from will have a completely unique genome.

1.3.2. Chromosomal sex

Most research using mESCs make use of XY cell lines, likely due to concerns over XX lines' higher likelihood of becoming XO than XY lines, but there are other differences between mESC lines of different chromosomal sexes that should be considered as well. Differences in Chromosome X expression mainly stem from dynamics of X Chromosome inactivation (XCI) and X Chromosome reactivation (XCR) during the early stages of embryonic development. XCI is a process by which supernumerary X Chromosomes are silenced through repressive epigenetic modifications initiated by the cis-acting coating of the inactivated chromosome by the long noncoding RNA *Xist* (reviewed in Augui et al, 2011). Before fertilization, the paternal Chromosome X is inactivated through an *Xist*-independent mechanism but is later inactivated through traditional XCI by the time of ZGA in mice (Okamoto et al, 2004; Turner et al, 2002). By the preimplantation epiblast stage, the ESCs of the ICM go through XCR, which erases most of the epigenetic imprinting on the reactivated paternal Chromosome X, and genes are expressed from both copies (Heard 2004; Mak et al, 2004; Okamoto et al, 2004). As ESCs are primed and begin the process of differentiating, XCI begins again and silences all but one Chromosome X randomly and that chromosome stays inactivated from then on (Augui et al, 2011). There is some debate in the field regarding whether Chromosome X dosage control in the preimplantation epiblast is the same in mouse and human (Saiba et al, 2018). Some papers have suggested that both copies of human Chromosome X are incompletely inactivated or 'dampened,' whereas others have argued for a traditional model of XCI as seen in the mouse (Moreira de Mello et al, 2017; Petropoulos et al, 2016). By either model, a similar level of dosage compensation is achieved as in mice during this stage of development.

When both X Chromosomes are active in the ICM, PSCs have several key differences in epigenetic signatures, expression of pluripotency genes, and timing of pluripotency exit. Studies have linked a global reduction of methylation relative to XY PSCs to dose-based differences in X-linked genes *Dusp9* and/or *Elk1* (Bruck et al, 2013; Choi et al, 2017). Expression differences in these X-linked genes lead to differences in *Uhrf* and *Dnmt3a/b* which facilitate global demethylation of the XX ESC genome (Bruck et al, 2013; Choi et al, 2017; Milagre et al, 2017; Ooi et al, 2010; Yagi et al, 2017; Zvetkova et al, 2005). The regions demethylated in XX relative to XY mESCs are enriched for MYC-associating domains (Panoupolos et al, 2017). In addition to being more demethylated, XX mESCs also express higher levels of key genes associated with pluripotency through increased inhibition of MAPK and GSK3 signaling pathways (Bruck et al, 2013; Schulz et al, 2014). The factors that control these pluripotency differences have not been pinned down, but strong candidates are *Dusp9*, *Eras*, *Nono*, *Tfe3*, *Nr0b1/Dax*, or *Zfx* – the former being the strongest candidate due to its ability to replicate most of these features through its sole overexpression (Schulz 2017). These is also a notable lag in XX ESCs exiting pluripotency that can be seen *in vitro* as well as in the timing of developmental milestones of XY embryos after preimplantation – however developmental timing differences can be somewhat attributable to dosage of Chromosome Y in addition to X (Burgoyne et al, 1995; Song et al, 2019).

Most hPSCs are not cultured in the same state of pluripotency, so their Chromosome X dosage differences are more subtle than those in mPSCs or the ICM of the mammalian preimplantation epiblast. Phenotypic and transcriptome comparisons have indicated that most common hPSC culture practices maintain cells that are more similar to a primed pluripotency state reminiscent of the late postimplantation epiblast when XCI has begun (Nakamura et al, 2016; Nichols and Smith 2009). As such, the

state of dosage compensation in these cells is highly variable and has been identified in various different combinations of active, inactive, and even partially active states (Anguera et al, 2012; Bar et al, 2019; Barakat et al, 2015; Hall et al, 2008; Hoffman et al, 2005; Kim et al, 2014; Lengner et al, 2010; Mekhoubad et al, 2012; Nazor et al, 2012; Pomp et al, 2011; Shen et al, 2008; Silva et al, 2008; Tchieu et al, 2010; Tomoda et al, 2012; Vallot et al, 2015; Xie et al, 2016). Unlike in the mouse context, it is uncommon for hPSCs to have two completely active X Chromosomes *in vitro*. Human PSCs also are subject to a process called Xi erosion (XIE), where a previously inactivated Chromosome X begins regaining expression of genes that are more distal from the X inactivation center. Between the two hPSC origins, it has been demonstrated that hESCs predominantly favor the XaXe condition and hiPSCs favor the XaXi condition, with XaXa being the least represented in either cell type (Bar et al, 2019). XIE is a progressive reactivation through demethylation of regions flanking the 'escapees' – genes that remain actively expressed even during inactivation – resulting in reactivating more of the chromosome over time (Bansal et al, 2021). The wide variety of XCI and erosion states, sometimes within the same culture, make the Chromosome X dosage effects difficult to compare between hPSCs and mPSCs as the dosages are muddied by often unreported states of activation.

1.3.3. Variability and subpopulations

Pluripotency factors are heterogeneously expressed in cultured PSCs. Several studies have found that while certain pluripotency-associated geneslike *Oct4* are expressed highly in almost all cells within a culture, other factors like *Essrb* are expressed in a wider range of intensity (Hayashi et al, 2008; Kumar et al, 2014; Messmer et al, 2019; Torres-Padilla and Chambers 2014; Toyooka et al, 2008). Even *Nanog*, one of the most commonly used markers for the naïve pluripotent state is not

expressed in a uniform manner (Chambers et al, 2007; Filipczyk et al, 2013; Hatano et al, 2005). As a result of these observations, some studies have suggested that ESC cultures are composed of cells fluctuating between naïve and primed pluripotency while being prevented from differentiation (Hayashi et al, 2008). The spectrum of pluripotency states that PSCs occupy can also be shifted depending on different strains, chromosomal sexes, and culture conditions. As mentioned in 1.3.1, nonpermissive strains naturally tend toward a more primed state of pluripotency. While this is not completely understood, examination of genetically diverse mice has discovered genetic variants that affect ground state pluripotency through changes to chromatin accessibility as well as the transcriptome and proteome – notably through differences in LIF receptor expression (Aydin et al, 2022; Skelly et al, 2020). As mentioned in 1.3.2, having two active copies of Chromosome X directs cells toward a more naïve pluripotent state, while a singular Chromosome X directs cells toward a more primed pluripotent state. Interestingly, the 2i/LIF culture condition can help equilibrate pluripotency differences due to Chromosome X dosage, suggesting that signaling from X-linked genes may be important for maintaining naïve pluripotency (Schulz et al, 2014; Schulz 2017).

Recently, a totipotent metastable state has been identified in mouse PSC cultures. Originally the ability of some ESC injections in host ICM to contribute to extraembryonic tissue, although inefficiently, was thought to be due to contaminating cells from those lineages (Beddington et al 1989). However, later a rare population of cells transcriptionally similar to the totipotent 2C stage of embryonic development was discovered in mESC cultures (Macfarlan et al, 2012). This population was termed the 2C-like (2CL) state due to their expression of totipotency genes such as *Zscan4* and endogenous retroviral elements, their downregulation of pluripotency genes such as *Nanog*, *Oct4*, and *Sox2*, and their ability to differentiate into an expanded range of

lineages (Carter et al, 2008; Falco et al, 2007; Macfarlan et al, 2012; Zhang et al, 2019). Generally, 2CL cells (2CLCs) are not more than a few percent of a culture, although it is a transient state that all PSCs will cycle through over the course of around nine passages (Carter et al, 2008; Macfarlan et al, 2012; Zalzman et al, 2010). Much of the consequences of 2CL cycling on mESCs is still unknown, however *Zscan4* functions in *Tert*-independent telomere elongation and it has been associated with genomic stability through *Zscan4c* knockdown in ESC cultures and correlations with culture conditions that are preferential to this subpopulation (Amano et al, 2013; Guo et al, 2018; Jiang et al, 2013; Zalzman et al, 2010).

While the existence of a totipotent subpopulation was thought to be a mouse exclusive feature, within the last year one has been discovered in hPSC cultures under specific conditions. As discussed, human PSCs do not exhibit truly naïve pluripotency during *in vitro* culture, and these primed pluripotent cultures had no discernable totipotent subpopulation. However, culture media designed to promote naïve pluripotency in hPSCs had similarly rare subpopulations of cells with expression of *Zscan4* and endogenous retroviral genes as well as decreased pluripotency-associated genes(Mazid et al, 2022; Moya-Jódar et al, 2022; Taubenschmid-Stowers et al, 2022; Yu et al, 2022). Due to differences in the developmental timeline between mouse and humans, these cells resemble the totipotent 8C stage of human embryonic development and are therefore termed 8C-like cells (8CLCs). Realization that certain hPSC cultures contain 8CLCs is new enough that only four publications have been published, mostly focusing on identification, characterization of cell state, and methods for induction (Mazid et al, 2022; Moya-Jódar et al, 2022; Taubenschmid-Stowers et al, 2022). Interestingly, this suggests that naïve pluripotency is more amenable to totipotent subpopulations than primed pluripotency is, which makes sense relative to the developmental timeline.

1.3.4. Culture conditions

Variety in mPSC culture conditions occurs mostly within the context of feeders and inhibitor cocktails which affect permissivity, heterogeneity, and genomic stability. Originally, PSC cultures were grown on a previously plated layer of mitotically-inactivated mouse embryonic fibroblasts (MEFs) which would secrete necessary signaling factors such as BMP and FSTL1 for maintenance of pluripotency – reminiscent of the embryonic context (Evans and Kaufman 1981; Guo et al, 2018). In a push to develop chemically defined culture conditions that would enhance replicability, MEF feeders were often removed because researchers could not reliably control what secreted factors they contribute to the media (Hayashi et al, 2008; Kumar et al, 2014; Torres-Padilla and Chambers et al, 2014; Xu et al, 2001). Cells grown in feeder-free cultures however are more likely to develop CNVs and have reduced 2CLC populations than feeder-based cultures with the same media (Guo et al, 2018). The first derived mESC lines were cultured in media containing serum and leukemia inhibitory factor (LIF) but without any supplemented inhibitors (Ser/LIF media), but eventually MEK1/2 and GSK3 inhibitors were added to derive mESCs from nonpermissive strains of mice (2i media) (Silva et al, 2008). Trying to improve upon the original 2i media, another group developed a less common use of a SRC inhibitor paired with the aforementioned GSK3 inhibitor (alternative 2i or a2i) (Shimizu et al, 2012). While both inhibitor cocktails have the desired effect of maintaining cells in pluripotency regardless of strain permissivity, reducing heterogeneity, and a2i media better replicates the transcriptional profile of ICM cells, there are detrimental side effects (Kolodziejczyk et al, 2015; Kumar et al, 2014; Singer et al, 2014). Cells cultured in these media are more prone to develop abnormal karyotypes and in homogenizing culture composition, 2i media essentially ablates the 2CLC population of mESCs (Garitaonandia et al, 2015; Guo et al, 2018).

There have been several media invented recently to culture human PSCs in a naïve pluripotent state rather than a primed pluripotent state. Some of the prominent culture media for naïve cultures are 5iLA or 5iLAF, t2iL+Gö, PXGL, and RSet which is a commercially available version of naïve human stem cell medium (NHSM) (Bredenkamp et al, 2019; Gafni et al, 2013; Takashima et al, 2014; Theunissen et al, 2014). All of these use the 2i/LIF media used to culture mouse PSCs in naïve pluripotency as a starting point and augment it in various ways directed by drug screening to adapt for human cells. NHSM was the earliest formulation and is composed of the 2i/LIF media plus inhibitors for P38/JNK, protein kinase C (PKC), and ROCK, as well as supplemented Activin and FGF (Gafni et al, 2013). 5iLA is the 2i/LIF media with inhibitors for BRAF, ROCK, and SRC supplemented with Activin, and also FGF in 5iLAF (Theuinissen et al, 2014). The t2iL+Gö media is the 2i/LIF media with a lower dose of the GSK3 inhibitor and a PKC inhibitor (Takashima et al, 2014). And lastly, PXGL media is the MEK inhibitor from 2i, LIF, the PKC inhibitor from t2iL+Gö media, a Rho-associated kinase inhibitor, and a tankyrase inhibitor/WNT pathway agonist (Bredenkamp et al, 2019). Naïve human PSC media have shown more heterogeneity in pluripotency-associated genes like SSEA4 and even some extraembryonic priming via expression of GATA6, GATA3, and CDX2 (Dong et al, 2020; Guo et al, 2016; Linneberg-Agerholm et al, 2019; Pastor et al, 2016). These media are relatively new and not used as often as the traditional primed media in part because they seem to also have a negative impact on genomic stability, similar to the 2i conditions in mouse cell lines (Pastor et al, 2016; Theunissen et al, 2014). It is worth pointing out that in both mouse and human, 2i or 2i-related media meant to artificially drive naïve pluripotency all result in worsened genomic stability (Ghimire et al, 2018; Guo et al, 2018; Pastor et al, 2016; Theunissen et al, 2014). This may be explained in part through these media's use of

MEK inhibitors – where prolonged MEK inhibition in mouse leads to worsened genomic stability and reduced MEK inhibition in human improves genomic stability (Di Stefano et al, 2018; Choi et al, 2017).

A number of methods have been developed for enriching the population of totipotent cells in culture through specific media, drug treatment, or metabolite supplementation. In the same year, two groups published papers describing derivation of what they called either extended or expanded pluripotent stem cells that were able to contribute to both embryonic and extraembryonic lineages (Yang et al, 2017a; Yang et al, 2017b). One group reported media containing LIF, (S)-(+)-dimethindene maleate, and minocycline hydrochloride in addition to the canonical GSK3 inhibitor present in 2i (Yang et al, 2007a). The other group reported media containing inhibitors for MEK1, JNK, P38, SRC, pyrrolopyrimidine, TNKS1/2, and YAP1 in addition to the canonical 2i inhibitors (Yang et al, 2007b). Culture media can also be supplemented with low doses of retinoic acid (RA)/vitamin A, small molecule inhibitors of MLL-1, acetate, L-lactate, D-ribose, spliceosome inhibitors (Iturbide et al, 2021; Rodriguez-Terrones et al, 2020; Sharova et al, 2016; Shen et al, 2021; Tagliaferri et al, 2016; Taglaferri et al, 2020; Wang et al, 2021; Zhang et al, 2019). One study has even induced 2CLC populations using nanograted topography of the adhesion surface to promote domed colony formation (Natale et al, 2020). None of these methods have demonstrated an ability to reprogram an entire culture of PSCs to totipotency in a stable, long-term fashion – most of them only converting a moderately sized population. Also, even some of the more common methods such as induction with RA-treatment have quite variable ranges of concentrations used, from as low as 0.16 – 0.53μM in Iturbide et al, 2021 to 5μM in Wang et al, 2021 (Iturbide et al, 2021; Tagliaferri et al, 2016; Tagliaferri et al, 2020; Wang et al, 2021). The range of techniques and dosages used to increase the 2CLC

population demonstrates that this is still a developing methodology, and the best practice has yet to be settled upon or may be dependent upon other contexts such as cell line, chromosomal sex, or culture practices.

1.4. Project rationale

Despite the vast academic and clinical usefulness of PSCs, they are still marred by the propensity to accumulate potentially restrictive and dangerous copy number variations. While there appears to be evidence of variable susceptibility to acquire aneuploidies, little research has been conducted to dissect the influences of genetic variation and chromosomal sex on this phenomenon – despite myriad sources of heterogeneity ranging from what culture media is used, to genetic variation, to chromosomal sex. It is also clear from the poor genomic stability of cells in chemically defined naïve pluripotency media that better understanding of aneuploidy variation is needed to improve culture conditions. The hypothesis of this project is that computational interrogation of existing multi-omic expression data from genetically diverse mESC panels can discover natural genetic variants or features that reduce accumulation of aneuploidies in mESCs.

1.5. Research projects

To achieve this goal, this project makes use of mESC expression data from the heterologous Diversity Outbred (DO) mouse population featured in Skelly et al, 2020 and Aydin et al, 2022. These mice are a multiparent population generated through crosses and successive outbreeding between eight founder mouse strains (Churchill et al, 2012; Svenson et al, 2012). Five of these parental strains are classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ) and three of them are more recently inbred wild-derived strains (CAST/EiJ, PWK/PhJ, WSB/EiJ). Due to outbreeding, these mice are all genetically unique and are heterozygous at most genomic locations, thus

better mimicking the variation one would see in humans, but comprised of previously known SNPs with balanced allele ratios that facilitate genomic mapping studies. First, we used bulk transcriptome and proteome data to indirectly karyotype a large population DO mESC lines, validating the uncovered patterns using single-cell transcriptomic data from a smaller DO mESC panel. Next, we leveraged the natural diversity in these panels to interrogate associations between genetic variants and chromosomal sex with aneuploidy status. Finally, differences in aneuploidy accumulation between chromosomal sexes are validated using single-cell transcriptomic analysis of early and aged isogenic mESCs and we investigated differentially expressed genes between key subpopulations.

1.5.1. Annotation and characterization of CNV trends in DO mESC data sets

Which CNVs exist in an mESC population can vary from study to study, with similar trends, and is dependent upon culture conditions and age of the culture. We used multi-omic expression data in the form of bulk RNA-seq and mass spectrometry proteomics to annotate DO mESC lines that have accrued significant populations of aneuploid cells in our mESC panel. The patterns of accumulated aneuploidy were validated using existing virtual karyotyping methods on a scRNA-seq data set from a small additional panel of DO mESCs cultured under the same conditions. We additionally identify a number of previously XX mESC lines to be reclassified as XO based on the same methodology and lack of *Xist* upregulation.

1.5.2. Analysis of factors associated with the development of aneuploidy in the DO mESC data sets

Within the DO mESC population, some samples manage to retain euploid chromosome counts over their time in culture while others develop substantial aneuploidy. To understand this disparity, we assayed for SNP over-representation on

duplicated chromosomes identified through an allele-specific expression (ASE) pipeline as well as genetic mapping to find associations between DO mESC genotypes and aneuploid annotations. We discovered no over-represented SNPs or genotypes associated with specific chromosome duplications, but identified a locus on Chromosome X associated with general aneuploidy. These observations led to the discovery of a chromosomal sex bias in aneuploidy based on Chromosome X dosage that was also verified in human iPSC data from a collaborator. Using both bulk and single-cell data we identified potential causal connections from a Chromosome X dosage bias in 2CLC population size and expression of X-linked tumor suppressor genes with functions in DNA damage response.

1.5.3. Characterizing the sexual dimorphism in aneuploidy development of mESCs

In this project, we expand upon the associations garnered from analyzing genetically diverse data sets and applied them to a controlled genetic context for validation and expanded analysis. We acquired three individually derived XX, XY, and XO mESC cell lines (nine in total) of the same passage number from a single genetic background (C3H/HeSn-Paf/J) and performed scRNA-seq at an early and aged time-point to evaluate how chromosomal sex modifies aneuploidy accumulation during prolonged culture. Here, we validate lesser aneuploid accumulation in XX mESC lines and show an enrichment for naïve pluripotency genes in the XX mESCs that remain euploid.

Chapter 2: Materials and Methods

2.1. DO mESC derivation and culture

The DO mESC lines used in this study were the same lines published in Skelly et al, 2020 and Aydin et al, 2022 and derived, cultured, and harvested as described in those methods sections. Briefly, DO mice were purchased from The Jackson Laboratory and bred at Predictive Biology Inc. The ICM from the resulting embryos were prepared as described in Czechanski et al, 2014 to derive mESCs. Initially in serum + 2i/LIF media during derivation, mESCs were transferred to a serum + 1i/LIF media (GSK3 inhibitor CHIR99021 only) and slowly tapered off of feeder-based culture until collection for bulk or scRNA-seq and mass spectrometry proteomics.

2.2. DO mESC paired-end bulk RNA-seq

RNA-seq data was collected, assayed, and processed as described in Skelly et al, 2020 and Aydin et al, 2022 with the following exceptions. Whereas Skelly et al, 2020 uses the single-end version of this data set, we use the paired-end data used in Aydin et al, 2022. Both previous papers used reads that had been rank Z transformed, but this project skips this step as this transformation removes the inherent structure necessary to identify copy number variations from non-normally distributed expression values. Before analyses, genes not expressed in all cell lines, gene models, and Riken genes were excluded.

2.3. DO mESC mass spectrometry proteomics

Mass spectrometry proteomics data was collected and assayed as described in Aydin et al, 2022 and processed as in Chick et al, 2016 with the following exceptions. Like the bulk RNA-seq data, these data have not been rank Z transformed in order to preserve the structure necessary to identify copy number variations from non-normally

distributed expression values. Before analyses, genes not expressed in all cell lines, gene models, and Riken genes were excluded.

2.4. DO mESC single-cell RNA-seq

DO mESC 10X Genomics scRNA-seq gene x cell matrix was provided by Ted Choi of predictive Biology. Data processing was performed mostly as described in the standard R/Seurat package tutorial. Only cells with less than 10% of total reads from mitochondrial genes, greater than 2,000 detected genes, and that had unique cell line identification were kept for further analysis. Cell cycle scoring was performed using mouse cell cycle genes from https://raw.githubusercontent.com/hbc/tinyatlas/master/cell_cycle/Mus_musculus.csv, but no significant effects were identified that would necessitate regression.

2.5. Virtual karyotyping using bulk RNA-seq and proteomics

A summary statistic as a metric of chromosome-wide gene expression was calculated by taking the median of expression for all autosomes and the sum of expression for Chromosome X. The distribution of these summary statistics was assessed using the Shapiro-Wilk normality test and chromosomes were considered non-normally distributed if $p \leq 0.01$. Chromosomes non-normally distributed for both RNA-seq and proteomics were tested further. Expression values for genes on these chromosomes were bootstrapped to determine separate RNA-seq and proteomics p-values, which were combined using the Z-transformed option for the ‘combine.test’ function from the R/survcomp package. A stringent threshold of combined $p \leq 0.0001$ was used to annotate high confidence aneuploid cell lines.

2.6. Virtual karyotyping using single-cell RNA-seq

Identification of cells harboring trisomies in the scRNA-seq mESC lines was performed using the R/CONICSmat package (Müller et al, 2018). Before virtual

karyotyping, cells were split by cell line, log normalized using a 1e6 scaling factor, and then normalized using negative binomial regression with the 'SCTransform' function all in the R/Seurat package. Each cell line was converted to log2(CPM/10+1) expression counts and only genes expressed in at least 5 cells were used for karyotype analysis. Cells were annotated as trisomic for a chromosome if the posterior probability of it belonging to the highest expression mixed Gaussian distribution was ≥ 0.9.

2.7. Estimating aneuploid proportions

The ratio of gene expression from cells aneuploid for a specific chromosome compared to entirely euploid cells was performed using the 'detectBreakpoints' function in the R/CONICSmat package. Gene expression ratios mapped across the chromosome were visually inspected for evidence of segmental duplication. If no segmental duplications were observed, an overall chromosomal expression aneuploid:euploid expression ratio was calculated by taking the median of all values for genes on that chromosome. Once the trisomy-induced change in expression ('Change' in the formula below) was calculated, the aneuploid content of bulk RNA-seq cell lines was calculated with the following formula:

$$Proportion_{aneup.} = \frac{(Line - Median_{eup.})/Median_{eup.}}{Change}$$

Where '*Proportion* $_{aneup.}$' is the proportion of cells in the sample that have the chromosome in question duplicated, '*Line*' is the summary statistic (as described in Chapter 2.5) for the cell line whose aneuploid proportion is being calculated, '*Median* $_{eup.}$' is the median of the summary statistic for lines annotated as euploid, and 'Change' is the proportional increase in expression expected if all cells in the sample were aneuploid for the chromosome in question, determined empirically.

2.8. Mouse-Human Synteny

Lists of genes with human synteny from mouse Chromosomes 1, 6, 8, and 11 were acquired from The Jackson Laboratory Synteny Browser (syntenybrowser.jax.org). Circos plots displaying syntenic gene connections between mouse and human genome coordinates were generated using the R/BioCircos package (Cui et al, 2016). Whole chromosome and chromosomal regions duplicated in human pluripotent stem cells were identified in Baker et al, 2007, Cowan et al, 2004, Draper et al, 2004, Maitra et al, 2005, Mikkola et al, 2006, Catalina et al, 2008, Lefort et al, 2009, Spits et al, 2008, Mitalipova et al, 2005, and unpublished work from our WiCell Institute collaborators.

2.9. Differential Expression Analyses of DO mESCs

Differential expression analysis of scRNA-seq data was performed using the R/Seurat package. The function 'FindMarkers' was used with the default Wilcoxon rank sum test parameter to find differentially expressed genes between each comparison. Biological process ontologies of differentially expressed genes with log2 fold-change > |0.3| between trisomy 8 and completely euploid cells were visualized using the 'gseGO' function in the R/clusterProfiler package (minGSSize = 10, maxGSSize = 800, pvalueCutoff = 0.1) showing 16 categories split between activated and suppressed.

2.10. Allele-Specific Expression (ASE) Analyses

Reconstruction of DO mESC phased genotypes and allele-specific expression analysis were conducted using the GBRS pipeline output that assigns RNA-seq reads to strain specific alleles. The imputed alleles were reassembled using the principles of parsimony – the first two genotypes are randomly assigned to chromosomes A or B, if one or both of the same genotypes are present at the next location then they are assigned to the same chromosome. This process is repeated until either (a) both alleles are different or (b) the genotype becomes homozygous. In situation (a), alleles are

randomly assigned to chromosome A or B again. In situation (b), allele assignments don't matter again until a location with heterozygosity is reached again. This continues until the end of the chromosome and repeated for all autosomes and Chromosome X of each DO mESC cell line. Once completed, the locations of ambiguity are identified (locations of heterozygosity after stretches of homozygosity and locations where both alleles changed) and the median ASE of regions between them are calculated. Regions with the highest ASE ratio are then assigned to the same chromosome (chromosome A) since it is likely that if one chromosome is duplicated and therefore relatively overexpressed it would continue throughout the chromosome, and if neither chromosome was duplicated, stochastic ASE differences shouldn't skew the bootstrapping assessment of duplication likelihood.

Once phased chromosome reconstruction was complete, ASE ratios were used to identify likelihood of a chromosome being duplicated (being skewed toward 2:1 expression). Median ASE ratios were calculated for chromosome A:chromosome B by calculating the chromosome B/total expression (0.0 means all expression is from chromosome A and 1.0 means all expression is from chromosome B, with the ratio equaling 0.33 if 100% of cells have chromosome A duplicated). ASE values were bootstrapped 2,000 times between chromosomes and p-values were assigned based on position in the distribution of possible median chromosomal ASE ratios. Chromosomes with $p < 0.01$ were considered specifically duplicated relative to the other chromosome in that sample. For each chromosome, the alleles for the duplicated copies were compared to the total pool of alleles in the population to test for overrepresentation using a multivariate hypergeometric test.

2.11. Genetic Mapping Analyses

Genetic mapping was performed using the R/qtl2 package (Broman et al, 2019). Aneuploid and euploid annotations for any chromosome, Chromosome 8, and Chromosome 11 were used as a binary trait, correcting the sex covariates for XX lines identified as XO. To identify significance thresholds for LOD scores, 1,000 permutations were run on each trait's genome scan independently. Genomic coordinates were identified as the region spanning a 1.5 LOD drop on either side of the identified significant peak. Genes within this region were acquired from the Mouse Genome Informatics (MGI) online database (informatics.jax.org).

2.12. 2CLC Gene Expression

Lists of transcripts differentially upregulated in 2CLCs were obtained from the supplemental materials of Macfarlan et al, 2012 and Amano et al, 2013. Both gene sets were filtered for genes with transcript expression in our bulk RNA-seq data set. For each gene set, expression z-scores were calculated on each gene for each cell line relative to the entire 170 cell line population. The resulting z-scores were summed within samples. Statistical significance was calculated by independent 2-group Welch's T-test comparing XX lines to XY and XO lines separately using the R/stats package.

2.13. Inbred mESC derivation and culture

C3H/HeSn-Paf/J mESC lines were obtained at the same passage number from Anne Czechanski in the lab of Laura Reinholdt where they were derived as described in Czechanski et al, 2014. All cell lines were cultured in the same media they were derived in, regardless of chromosomal sex. Culture dishes were treated with gelatin for at least 20 minutes at 37°C before plating an even layer of γ-irradiated MEF feeders the day before adding mESCs. Cells were allowed to grow for 2-3 days before passaging by

dissociation with trypsin-EDTA (Gibco 25300-054 or 25200-072) or accutase (Millipore SCR005) when performing flow cytometry or scRNA-seq cell preparation.

2.14. Inbred mESC flow cytometry

Cells were dissociated and turned into a single-cell solution using accutase (StemCell Technologies 07920). Dead cells were stained incubation with ghost dye 780 (Tonbo Biosciences 13-0865-T100) for 20 minutes. Following ghost dye incubation, cells were fixed in 4% PFA (Electron Microscopy Sciences 15710-S) for 15 minutes on ice. Washes following fixation were supplemented with 125 mM Glycine pH 3.0 to minimize autofluorescence. Cells were permeabilized using 1X Perm/Wash buffer (BD Biosciences 554723) for 15 minutes on ice before incubation with 1:600 MERVL-gag primary antibody (EpiGentek A-2801-050) and/or 1:300 OCT3/4 primary antibody (BD Biosciences 611202) for 1 hour at 4°C with shaking. Finally, cells were incubated with Alexa 488 donkey anti-rabbit IgG secondary antibody (Invitrogen A21206) or Alexa 647 donkey anti-mouse IgG (heavy and light chain) secondary antibody (A32787) for 20 minutes on ice. Stained cells were then assayed on the BD FAC Symphony A5.

2.15. Inbred mESC single-cell RNA-seq and analyses

For sequencing, cells were dissociated as they were for flow cytometry. Once lifted, cells were allowed to settle in a 10cm culture dish to remove MEF feeders due to the differential attachment rate of mESCs and feeders. The remaining cells were then passed through a 40 micron filter to remove clumping cells. Then, 500,000 cells from each cell line were labeled with lipid modified oligos (LMOs) as described in McGinnis et al, 2018. Briefly, cells were washed and resuspended in ice-cold PBS before incubating with 1:1 molar ratio of barcode oligo and LMO anchor (supplied by Chris Baker from the lab of Chris McGinnis) for 5 minutes at room temperature. Then, cells were incubated with LMO co-anchor (also from Chris Baker) on ice for 5 minutes. LMOs were blocked

with 1% BSA and 1% FBS in ice-cold PBS, washed in the same solution, and sent for 10X library prep and sequencing in The Jackson Laboratory single-cell biology core facility.

Cell viability was assessed on a Luna FX7 automated cell counter (Logos Biosystems), and up to 40,000 cells (4,000 cells per LMO) were loaded onto one lane of a 10X Chromium Controller. Single cell capture, barcoding and library preparation were performed using the 10X Chromium platform (Zheng et al, 2017) version 3.1 NEXTGEM chemistry and according to the manufacturer's protocol (#CG00315) with modifications for generating the LMO libraries. cDNA and libraries were checked for quality on Agilent 4200 Tapestation, quantified by KAPA qPCR, and libraries were pooled using a ratio of 95% gene expression library and 5% LMO library before sequencing; each gene expression-LMO library pair was sequenced at 91% of an Illumina NovaSeq 6000 S4 flow cell lane, targeting 20,000 barcoded cells with an average sequencing depth of 100,000 read pairs per cell.

Illumina base call files for all libraries were converted to FASTQs using bcl2fastq v2.20.0.422 (Illumina) and FASTQ files associated with the gene expression libraries were aligned to the GRCm38.93 reference assembly with GENCODE vM23 annotations (10x Genomics mm10 reference 2020-A) using the version 6.1.2 Cell Ranger count pipeline (10x Genomics). FASTQ files representing the LMO libraries were processed into cell x LMO digital count matrices using CITE-Seq-Count (version 1.4.5; https://zenodo.org/badge/latestdoi/99617772).

2.15.1. Early and aged data set processing

Data processing was performed mostly as described in the standard R/Seurat package tutorial. Only cells with less than 4% of total reads from mitochondrial genes, greater than 4,500 detected genes in the early data set and 4,000 detected genes in the

aged data set were kept for further analysis. LMO barcode demultiplexing was performed using the 'MULTIseqDemux' function in the R/Seurat package. These cell identifications were found not to coincide with unsupervised clustering or XX biased gene expression in downstream analyses and were regarded as uninformative due to the low LMO barcode read counts.

2.15.2. Cell cycle phase annotations

Cell cycle scoring was performed the same as in the DO mESC scRNA-seq data set – however, to ensure no subtle differences between chromosomal sexes due to cell cycling were observed, the difference between G2M and S phase scores were regressed out during scaling of the data.

2.15.3. Identification and characterization of pluripotent state, differentiation primed, MEF, and 2CLC clusters

Pluripotent clusters were identified using expression of several key genes associated with general, naïve, and primed pluripotency. Additionally, several genes differentially expressed in XX and XY mESCs were chosen to aid in identifying chromosomal sex. Genes were selected using information from Endoh and Niwa 2022, Genolet et al, 2021, Lentini et al, 2022, Schulz et al, 2014, Song et al, 2019, and Wang et al, 2020. MEFs were identified using expression of *Thy1* and by predominant G1 cell cycle phase. 2CLCs were identified using expression of the two Zinc finger and SCAN-domain containing 4 (*Zscan4*) genes expressed in preimplantation development: *Zscan4c* and *Zscan4d*. Clustering resolution was set just sensitive enough that the 2CLC cluster separated out, as it was the smallest cluster of interest.

2.15.4. Virtual karyotyping using single-cell RNA-seq

Cells with chromosomal CNVs were identified using the same R/CONICSmat pipeline used for the DO mESC scRNA-seq data except that cells were not processed with the 'SCTransform' function.

2.15.5. Differential expression in aged euploid XX populations

The XX cell population clusters from each data set were isolated using the 'SplitObject' function in R/Seurat and merged into a single object where original data set identity was regressed out during rescaling of the expression data for combined analysis. Differential expression analysis of scRNA-seq data was performed using the R/Seurat package. The function 'FindMarkers' was used with the default DESeq2 option to find differentially expressed genes between cells marked as XX and euploid in the aged data set compared to XX and euploid cells in the early data set.

2.16. Author contributions

Materials and methods were written and performed by Alexander Stanton with input from Sandy Daigle, Greg Perry, and Bill Flynn of The Jackson Laboratory as well as Dylan Stavish of Sheffield University and Ted Choi of Predictive Biology Inc. Specifically, Sandy Daigle and Greg Perry were responsible for preparation and quality control of the 10X scRNA-seq libraries generated from C3H/HeSn-Paf/J samples that I prepared and LMO-tagged; Bill Flynn was responsible for sequencing, scRNA-seq quality control, aligning sequencing reads, and provided the writeup of these materials and methods; Dylan Stavish provided G-banded karyotyping data collected from mutual collaborators at The WiCell Institute; and Ted Choi provided the DO mESC scRNA-seq data.

Chapter 3: Results

3.1. Annotation and characterization of CNV trends in DO mESC data sets

3.1.1. Characterization of DO mESC data sets

To better understand variability in genetic stability, we began by analyzing data sets from two panels of DO mESCs (Figure 3.1). The primary data sets used were paired-end bulk RNA-seq and multiplexed mass spectrometry proteomics data from 170 unique DO mESC lines with a balanced split of chromosomal sexes: 88 XX and 82 XY lines. We also obtained 10X single-cell RNA-seq data from 12 additional XY DO mESC lines, none of which were cell lines represented in the bulk expression data sets. All cells were cultured for multiple passages in serum + LIF media supplemented with the GSK3 inhibitor CHIR99021 and were weaned off of feeders after derivation (see chapter 2.1 for more details). After filtering out genes not expressed above a threshold value, genes not expressed in all cell lines, gene models, and Riken genes in both RNA-seq and mass spectrometry data, the bulk RNA-seq data consisted of 12,579 transcripts and the proteomics data consisted of 5,141 proteins. The single-cell RNA-seq data consisted of

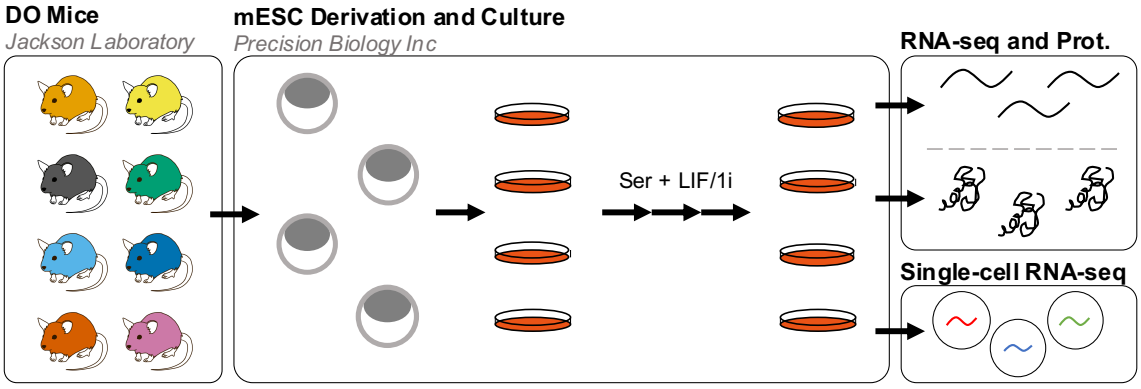


Figure 3.1 - DO mESC data sets.
DO mice from The Jackson Laboratory were purchased and maintained by Predictive Biology Inc, where they were bred to generate the embryos used to derive mESC lines. The derived mESCs were then cultured in Ser + LIF/1i (GSK3 inhibitor) while being weaned off of feeders. Then 170 mESC lines were harvested for bulk RNA-sequencing and proteomics, while an additional 12 mESC lines were collected for scRNA-seq.

18,954 transcripts and 5,992 total cells representing between 427 to 557 cells per cell line. While cell lines with gross aneuploidies were filtered out at the time of genotyping (before growing out cells in serum + LIF/GSK3 inhibitor media), none of these cell lines had corresponding karyotype data from the time of collection.

3.1.2. Virtual karyotyping with RNA-seq and proteomics

3.1.2.1. Evidence of chromosomal copy number variation

Because these data sets do not have direct karyotyping data associated with them, we employed virtual karyotyping as an indirect means of interrogating copy number variations in these cell lines. Our first step was to compare chromosome-wide gene expression levels between all cell lines in the bulk RNA-seq and proteomics data sets. To do this, the expression of each gene in each cell line was normalized to the

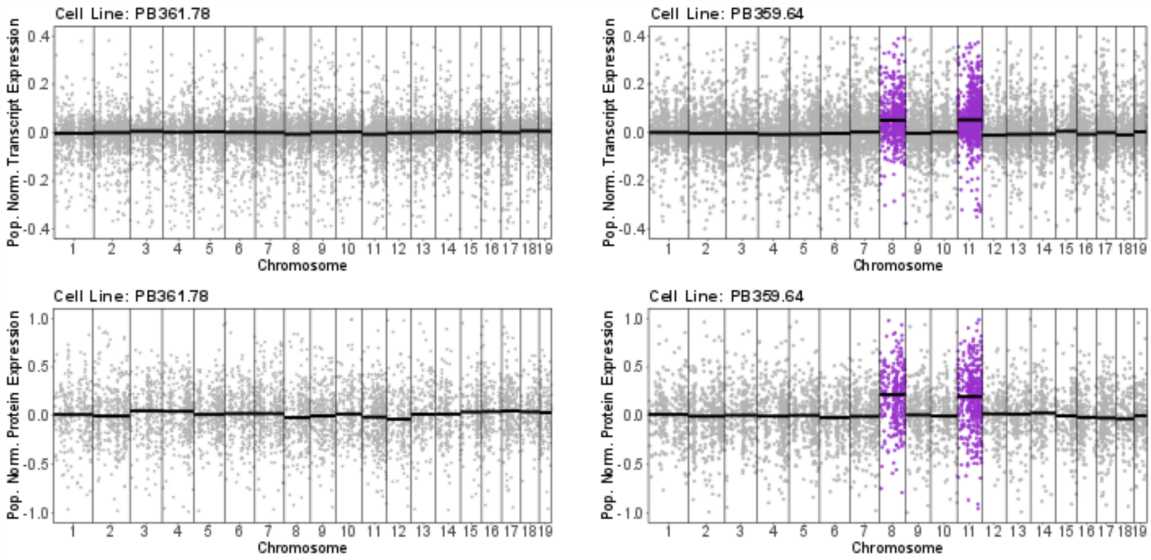


Figure 3.2 - Accumulated chromosomal copy number variations can be observed through chromosome-wide gene expression changes.
Gene expression values were population normalized to the median expression of that gene across all cell lines to align them along the x-axis. Points represent individual genes' population normalized expression; points are purple if they represent a chromosome later annotated as aneuploid. Black bars represent the median of expression of all genes on a chromosome. PB361.78 is an example of an entirely euploid DO mESC line, while PB359.64 is an example of a DO mESC line with detectable duplications of chromosomes 8 and 11.

median expression of that gene across all 170 cell lines. These population normalized gene expression values were plotted equidistantly in order of gene position along each chromosome and overlayed with the median of those values to better visualize trends among the variation (Figure 3.2). While most samples had chromosome median expressions that were very similar to that of the population (cell line PB361.78 in figure 3.2), certain cell lines displayed higher than expected median expression of specific chromosomes using both transcript and proteomic metrics (cell line PB359.64 in figure 3.2). This observation indicated to us that there were a number of lines in these data sets that had begun accumulating subpopulations of cells harboring chromosomal duplications.

We next wanted to determine which chromosomes were recurringly duplicated within the DO mESC lines. Due to the balanced genetic structure of the DO population, we assumed that if a panel of cell lines had normal karyotype, then their median expression for a given chromosome would fit a normal distribution. Under this assumption, if the distribution for any chromosome is positively or negatively skewed,

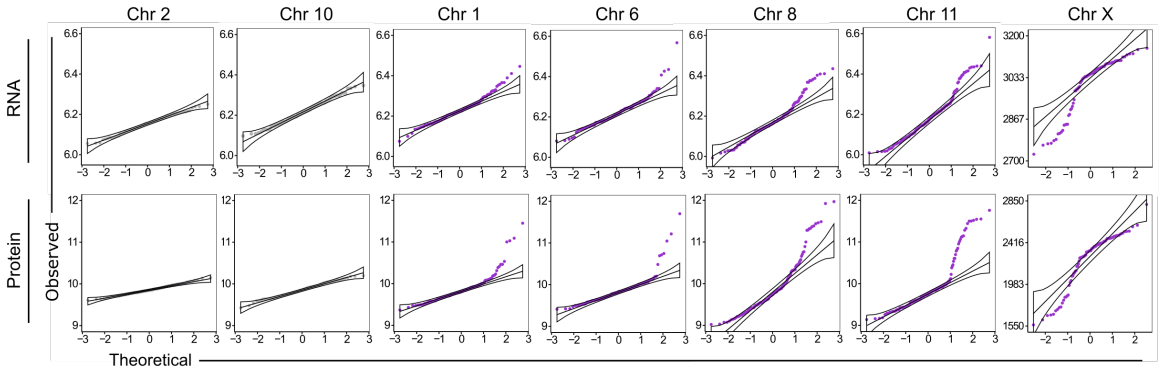


Figure 3.3 - Recurrent CNVs disrupt the normal distribution of median chromosome expression in the DO mESC population.
Shown are QQ plots for the median chromosome expression values (top row is RNA, bottom row is protein) for specific chromosomes of each cell line in our bulk data sets, plotting observed values against a theoretical normal distribution. All 170 DO mESC lines were tested for autosomes, whereas only the 88 originally XX lines were tested for chromosome X.

Table 3.1 - Shapiro-Wilk Normality Test results for each chromosome and stage of gene expression.
Each test is labeled 'N' for normally distributed ($p > 0.01$), 'NN' for non-normally distributed ($p \leq 0.01$) with no visible skew, '+S' for a non-normal positive skew, and '-S' for a non-normal negative skew. All 170 DO mESC lines were tested for autosomes, whereas only the 88 originally XX lines were tested for chromosome X. Underneath the skew are the Shapiro-Wilks test p-values.

	Chromosomes				
	1	**2**	**3**	**4**	**5**
RNA	+S (p = 9.7e-4)	N (p = 0.46)	N (p = 0.49)	N (p = 0.85)	N (p = 0.29)
Prot	+S (p = 1.3e-13)	N (p = 0.52)	N (p = 0.84)	N (p = 0.60)	N (p = 26)
	6	**7**	**8**	**9**	**10**
RNA	+S (p = 2.7e-8)	N (p = 0.39)	+S (p = 2.2e-5)	N (p = 0.71)	N (p = 0.61)
Prot	+S (p = 2.0e-15)	N (p = 0.09)	+S (p = 3.0e-10)	N (p = 0.72)	N (p = 0.46)
	11	**12**	**13**	**14**	**15**
RNA	+S (p = 3.3e-6)	N (p = 0.09)	N (p = 0.11)	N (p = 0.44)	N (p = 0.90)
Prot	+S (p = 7.6e-14)	NN (p = 2.8e-7)	N (p = 0.03)	N (p = 0.39)	N (p = 0.25)
	16	**17**	**18**	**19**	**X**
RNA	N (p = 0.16)	N (p = 0.33)	N (p = 0.17)	N (p = 0.07)	-S (p = 5.0e-8)
Prot	N (p = 0.16)	N (p = 0.29)	N (p = 0.50)	N (p = 0.69)	-S (p = 1.58e-6)

then it indicates the presence of recurrent duplications or losses of that chromosome, respectively. We find that chromosomes which are less likely to be duplicated or lost in mESCs based on previous reports, such as Chromosomes 2 and 10, tightly fit a normal distribution (Shapiro-Wilk normality test $p > 0.01$, Figure 3.3). Chromosomes 1, 6, 8, and 11 were the only autosomes with non-normal distributions in both median chromosome transcript and protein expression (Shapiro-Wilk normality test $p \leq 0.01$; Figure 3.3 and Table 3.1). All four showed a positively skewed distribution, indicating that multiple

samples had higher than expected expression from those chromosomes – indicating that all autosomal aneuploidies in these data sets are from duplications. Chromosome 12 was the only other autosome with non-normal distribution for proteomics (Shapiro-Wilk normality test $p \leq 0.01$), however there were no visibly notable deviations from the diagonal in a QQ plot and the transcriptomics data was normally distributed (Shapiro-Wilk normality test $p > 0.01$).

Because Chromosome X is often lost in originally XX mESC lines, we assessed the normality of Chromosome X expression as well. Because naïve pluripotent mESCs have reactivated their second Chromosome X we were able to treat it similarly to the analysis for autosomes. However, there are two differences. One, we limited the comparison to only cell lines originally genotyped as XX because the XY lines' single Chromosome X would confound the normality test. Two, we used the sum of Chromosome X transcript or protein expression instead of the median as it provides a more stable metric for Chromosome X. There was significant enough noise in the median Chromosome X expression metric to blur the distinction between XX and XY lines, even when accounting for the possibility of XO mosaic lines. The use of Chromosome X gene expression sums, as recommended by lab colleagues resolved this issue. It is not fully clear why this noise exists for Chromosome X genes; however, it may be linked to low levels of non-uniform dampened X-linked gene expression from leaky *Xist* expression or non-uniform upregulation of X-linked genes in XY lines (Lentini et al, 2022). The 88 originally XX cell lines also failed to fit a normal distribution, but this time with a negatively skewed distribution as we expect if some cell lines were losing Chromosome X (Figure 3.3). Loss of Chromosome X in a number of our originally XX cell lines meant that we would also need to annotate sex chromosome composition to update the metadata on chromosomal sex.

3.1.2.2. Clustering and annotation of autosome duplications using a multi-omic approach

Typically, virtual karyotyping relies solely on transcript expression data due to the ubiquity of RNA-seq analysis. We are in the somewhat unique position of having both transcript and protein expression data for this panel of cell lines. Therefore, we utilized a virtual karyotyping approach that integrates statistical analysis of both data sets such that corroboration from each context is required to identify a cell line as aneuploid (Figure 3.4). First, a bootstrapping approach was used to identify which observed chromosome summary statistics were significantly higher than expected. Second, the bootstrapped p-values were combined to obtain a single metric incorporating both layers of evidence. Third, an ultraconservative threshold ($p \leq 0.0001$) is used to distinguish

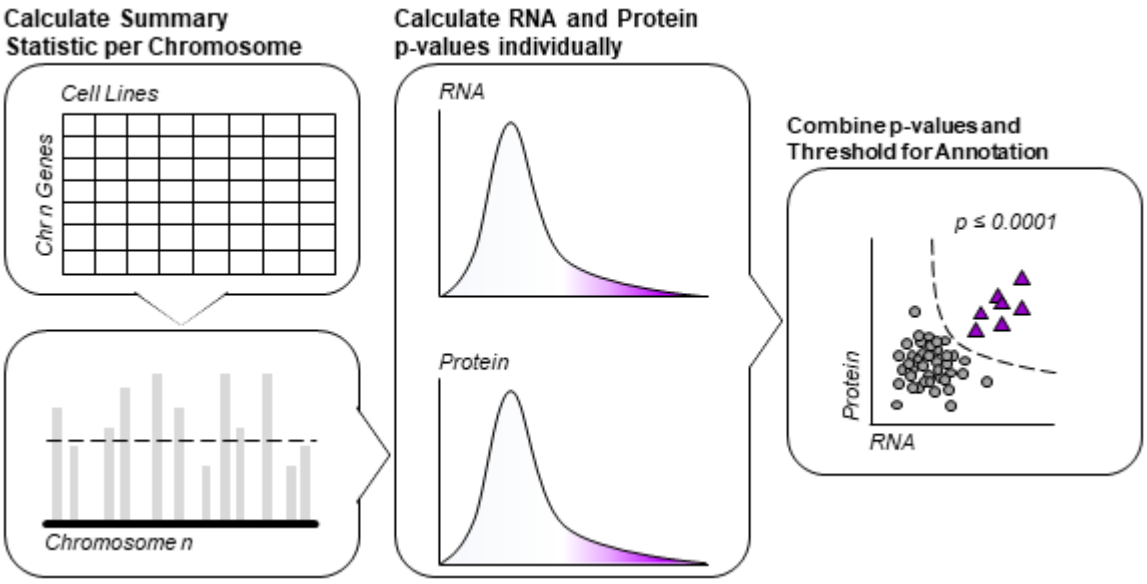


Figure 3.4 - Workflow for virtual karyotyping using RNA-seq and proteomics. Summary chromosomal expression statistics are calculated from each data set. Summary statistics are the median of all genes on an autosome or the sum of all genes on chromosome X. Gene expression values are shuffled between cell lines and summary statistics are recalculated 2,000 times to get bootstrapping probabilities for the observed values. Then p-values for RNA-seq and proteomics are combined using the z-transformed method and cell lines with observed summary statistics' $p \leq 0.0001$ are annotated as aneuploid.

samples with significantly higher transcript and protein expression for a given chromosome.

Using this methodology, we were able to better separate cell lines that have begun developing aneuploidies and annotate them accordingly. When plotting summary statistics for transcript and protein of the normally distributed Chromosomes 2 and 10, it can be seen that cell lines all cluster together tightly when there are no chromosomal copy number variations (Figure 3.5 A). In the cases of Chromosomes 1, 6, 8, and 11 we see that most cell lines are euploid and cluster together, but there is a distinct trail of cell lines separating from the euploid cluster with relatively increased transcript and protein expression (Figure 3.5 B). On a single axis some of the cell lines annotated as aneuploid blend into the periphery of the euploid cell line cluster, however with both transcript and protein axes the aneuploid cell lines are much more distinct. Conversely, for Chromosome 11 there are cell lines with high transcript or protein expression individually that is not seen in the other metric, therefore they are not considered aneuploid in this methodology because the strength of both evidences does not meet the threshold.

3.1.2.3. Summary of aneuploid prevalence in bulk data sets

Altogether we annotated 62 cell lines with detectable accumulations of cells harboring autosomal aneuploidies. Of these 62 cell lines, there were 10, 8, 35, and 24 cell lines with duplications for Chromosomes 1, 6, 8, and 11 respectively (Figure 3.5 B and Figure 3.6). The majority of cell lines were not annotated as aneuploid, displaying no statistically significant increase in autosomal gene expression (108/170, 63.5%; Figure 3.6). Most of the annotated aneuploidies were only aneuploid for a single chromosome. Chromosomes 8 and 11 were the most commonly duplicated, cumulatively 52 (83.9%) of the aneuploid cell lines had at least one of those two duplicated, including all of the cell lines with multiple duplications (Figure 3.6).

3.1.2.4. Annotation of XX lines becoming mosaic XX/XO

We applied the same methodology to identify cell lines with detectable subpopulations of XO cells. A cluster of 66 cell lines with high Chromosome X transcript and protein expression could be seen distinct from a trail of 22 cell lines separating from the XX cluster with relatively decreased Chromosome X transcript and protein expression (left panel of Figure 3.7 A). When the XY cell lines were overlayed with all the originally XX cell lines, they formed a separate cluster that the low Chromosome X expressing cell lines were moving toward and merging with (right panel of Figure 3.7 A). Additionally, the expression of *Xist* from the originally XX cell lines with low Chromosome X expression is not statistically higher than the cluster of XX cell lines, ruling out improper XCI in these cell lines accounting for decreased Chromosome X expression

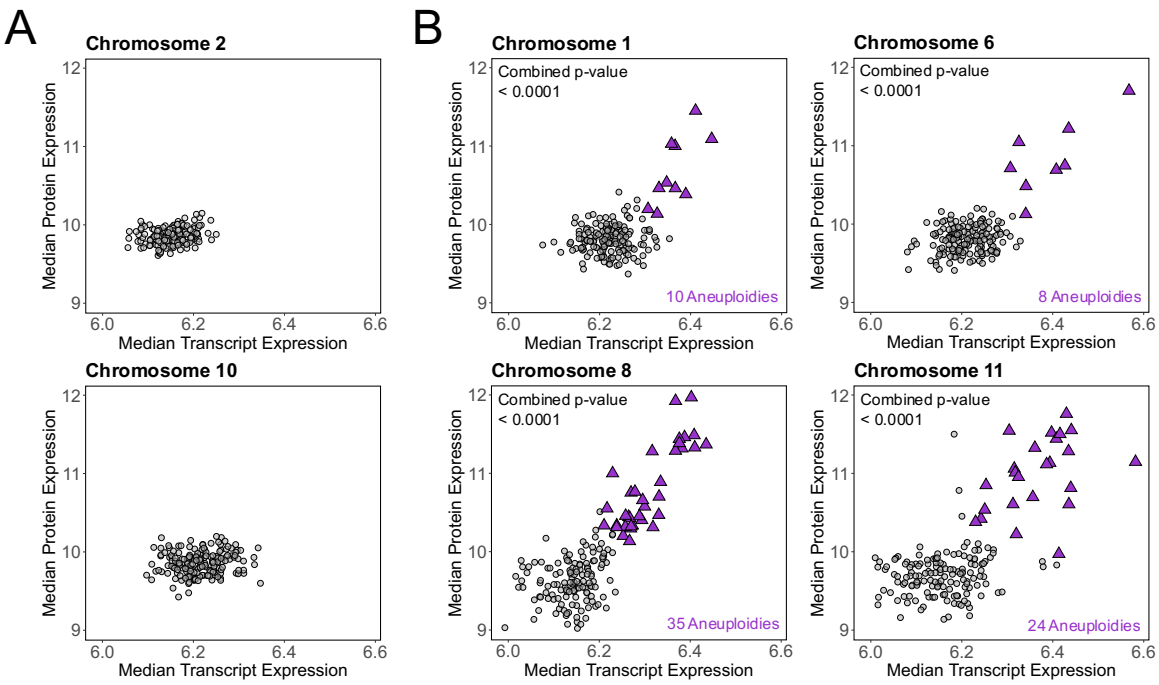


Figure 3.5 - Distribution and annotation of DO mESC lines from chromosomes 1, 6, 8, and 11.
Plots showing the clustering of cell lines by summary transcript and protein statistics for (A) normally distributed chromosomes 2 and 10, as well as (B) positively skewed chromosomes 1, 6, 8, and 11. Cell lines annotated as aneuploid are represented with purple triangles, cell lines annotated as euploid are represented with grey circles.

(Figure 3.7 B). Together, the lack of *Xist* upregulation and the Chromosome X expression becoming comparable to the XY cell lines were consistent with a mosaic of XX and XO cells in these cell lines (hereafter referred to as solely as XO). The updated chromosomal sex distribution was therefore 66 XX, 22 XO, and 82 XY cell lines.

3.1.3. Virtual karyotyping with scRNA-seq

3.1.3.1. Single-cell RNA-seq virtual karyotyping methodology

We also annotated the cells with abnormal karyotype in the DO scRNA-seq data set as well. For this analysis we used the R/CONICSmat (Copy-Number analysis In single-Cell RNA-sequencing from an expression matrix) package, designed to identify cells with CNVs using single-cell expression data without an explicit normal karyotype control data set (Müller et al, 2018). Analysis was conducted on the data split by cell line to avoid noise coming from variation in gene expression due to genetic background

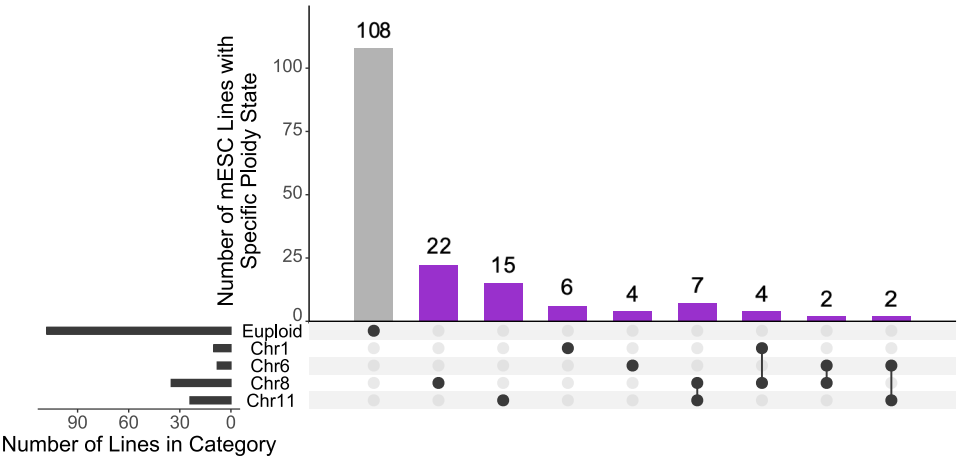


Figure 3.6 - Autosomal aneuploidies detected in DO mESC lines using bulk data sets are predominantly single chromosome aneuploidies and duplications of chromosomes 8 and 11.
Summary of the aneuploid categories that each cell line belongs to. Possible categories (euploid and aneuploid for chromosome 1, 6, 8, and/or 11) are listed along the side. Each vertical bar represents the number of cell lines in the category or combination of categories denoted by the filled points beneath the bar. Each horizontal bar represents the number of cell lines in each individual category, regardless of combinations with other categories.

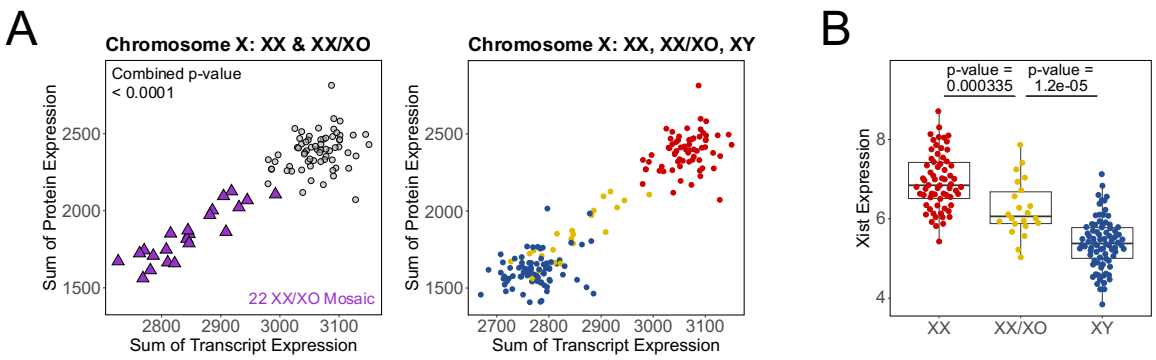


Figure 3.7 - Distribution and annotation of originally XX DO mESC lines for chromosome X.
(A) Plots showing the clustering of cell lines by summary transcript and protein statistics for chromosome X. On the left, only originally XX lines are shown. On the right, XY lines are included to show relative chromosome X expression to lines with a single active X. (B) Expression of the Xist lncRNA compared between high (XX) and low (XO) chromosome X expressing originally XX lines and the XY lines. Cell lines annotated as potentially XO are represented with purple triangles or yellow circles, cell lines annotated as XX are represented with grey or red circles, XY cell lines are represented with blue circles.

rather than CNVs and then log normalized. A clear separation of cells was obvious in the Chromosomes 1, 6, 8, and 11 – the chromosomes with recurrent duplications in the bulk data cell lines – but not in the typically euploid chromosomes. R/CONICSmat uses a mixed Gaussian model to assign posterior probabilities of each distribution to individual cells, which can be individually thresholded by the user. We used a threshold that a cell must have a probability of $p \geq 0.9$ of belonging to the higher expression distribution to be considered as a chromosome duplication.

3.1.3.2. Summary of aneuploid prevalence in scRNA-seq

After annotating individual aneuploid cells in the scRNA-seq data, we observed very similar distributions to the annotations from the bulk data virtual karyotyping (Figure 3.8). Again, the single category with the most cells were euploid cells, but in this data set that only represented around a third of all cells (2,109/5,992, 35.2%) and only slightly outnumber all cells containing a Chromosome 8 or 11 duplication (2,006 and 1,901 cells,

respectively). Single chromosome duplications, which were 72.5% of all cells with duplications (2,815/3,883), were more prevalent than multiple chromosome duplications. And chromosome 8 and 11 duplications were even more relatively common in this data set than in the bulk data sets, constituting 90.2% of all single chromosome duplications (2,538/2,815) and 98.1% of multiple chromosome duplications having at least 8 or 11 duplicated (1,048/1,068). We do not have passage number metadata for the DO data sets and the single-cell data were collected separately from the bulk data, so it is possible that the lower euploid cell proportion in the single-cell data is due to longer time in culture. Overall, the scRNA-seq data set corroborates the previous observation that Chromosomes 1, 6, 8, and 11 are the recurring duplications in DO mESCs and the dominant prevalence of Chromosome 8 and 11 duplications relative to the other two.

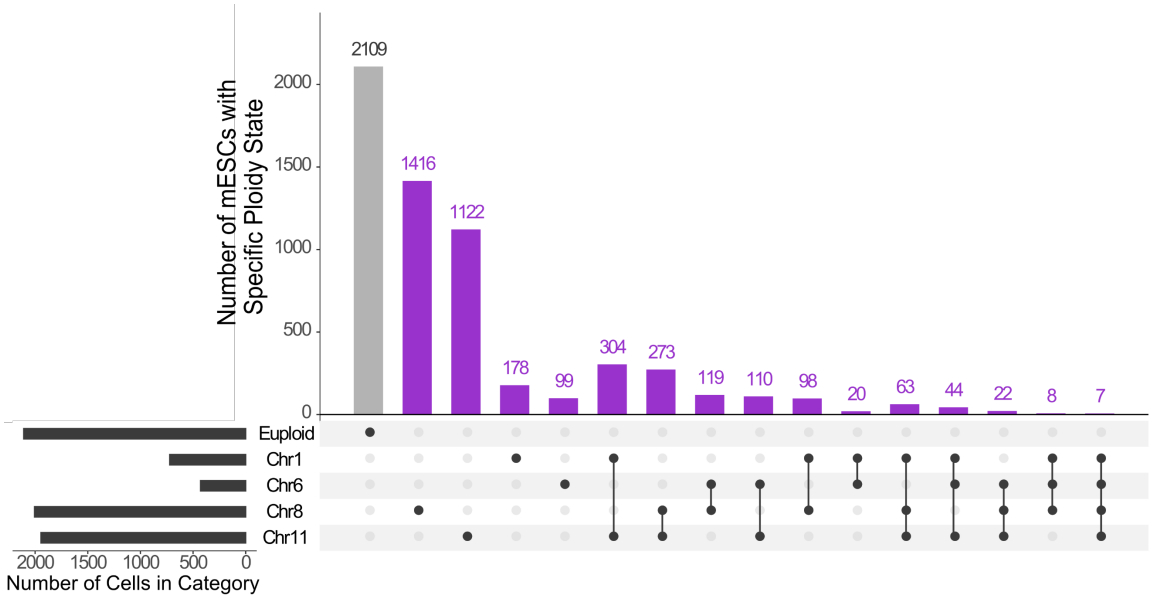


Figure 3.8 - Autosomal aneuploidies detected in DO mESC lines using scRNA-seq largely recapitulate the patterns using bulk data sets.
Summary of the aneuploid categories that each cell line belongs to. Possible categories (euploid and aneuploid for chromosome 1, 6, 8, and/or 11) are listed along the side. Each vertical bar represents the number of cell lines in the category or combination of categories denoted by the filled points beneath the bar. Each horizontal bar represents the number of cell lines in each individual category, regardless of combinations with other categories.

3.1.4. Estimated aneuploid proportions of cell lines from bulk data sets

3.1.4.1. Comparing gene expression differences on duplicated chromosomes in single cells

To estimate the proportions of trisomic cells in the culture mosaic of our aneuploid bulk data DO mESC lines, we compared their changes in duplicated gene expression to aneuploid DO single cells. While it is typically expected that a chromosomal duplication will lead to a 1.5x upregulation of genes from the duplicated chromosome, there is conflicting evidence in certain human trisomies. For example, several studies of human trisomy 18 and 21 have demonstrated lower than expected average changes in multiple affected cell types (Gonzales et al, 2018; Mao et al, 2005; Pelleri et al, 2018; Slonim et al, 2009). Therefore, we determined the effect size of these changes comparing aneuploid single cells to euploid single cells as described in Chapter 2.7.

Empirical quantification of expression changes due to trisomy were assessed in the scRNA-seq data using the R/CONICSmat package. By specifying the categories to compare, the ‘detectBreakPoints’ function calculates the expression changes to identify any potential segmental duplications/losses but can be repurposed to determine the median amount of change in expression across the chromosome. First, we compared the expression of non-duplicated chromosomes in cells harboring any duplication(s) to the entirely euploid cells to determine if we need to account for any global changes in gene expression due to aneuploidy. We found a 0.98x expression from non-duplicated chromosomes in aneuploid cells and therefore did not observe a notable global change in gene expression (Figure 3.9 A). When comparing cells with one specific trisomy to entirely euploid cells, we did not observe any observable evidence of segmental duplication (Figure 3.9 B). These cells had the expected increase in gene expression

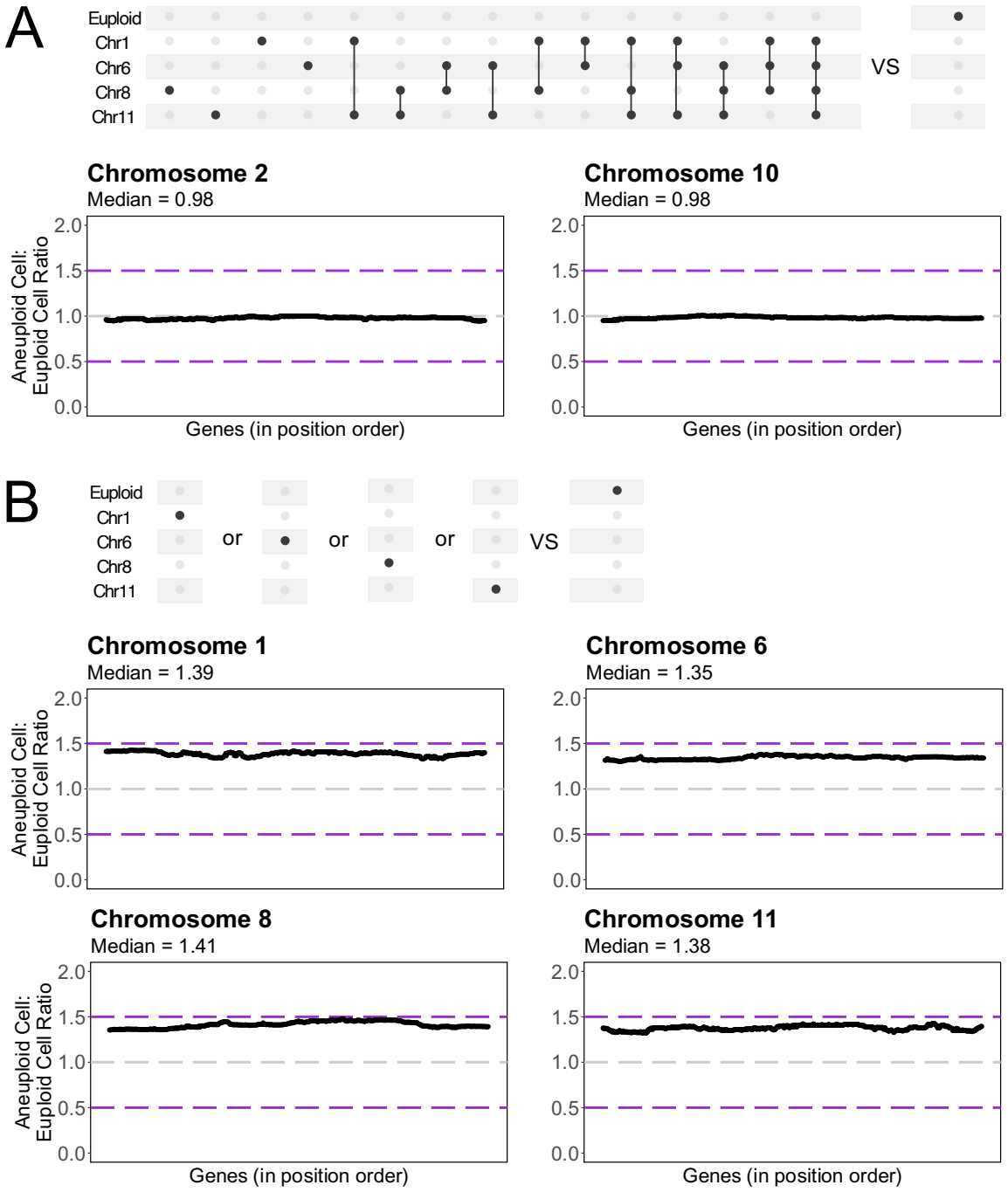


Figure 3.9 - Autosome duplications cause attenuated expression increases in the respective chromosomes, but not in other non-duplicated autosomes.
Top of each panel represents the comparisons being made in the bottom of the panel. Bottom of each panel represents the ratio of expression from (A) all aneuploid cells or (B) induvial trisomic cells compared to entirely euploid cells for the indicated chromosomes. Genes are plotted equidistantly in order of chromosomal location. Dotted purple lines represent the theoretical change in expression for duplications (1.5) or losses (0.5) of a chromosome. The grey dotted line represents no change in gene expression between categories (1.0).

from the respective duplicated chromosomes, all falling short of the theoretical 1.5x change in gene expression (1.39x, 1.35x, 1.41x, and 1.38x for Chromosomes 1,6, 8, and 11 respectively) indicating that there is some mild level of attenuation of steady state transcript levels. There is possible error in the R/CONICSmat pipeline since the euploid and aneuploid cell populations form a mixed Gaussian distribution, rather than wholly separate categories, and therefore some cells with intermediate chromosome expression may be miscategorized. However, all duplicated chromosomes fall around a 1.4x change (or 0.4x increase), so we used this as the difference in expression expected for an entirely aneuploid mESC culture when estimating aneuploid proportions using the methods described in Chapter 2.7).

3.1.4.2. Calculating aneuploid proportion using empirical reference

We then estimated the aneuploid proportions using median euploid expression, the empirical change in expression from single-cell data, and the transcript expression values of aneuploid annotated cell lines. Because we do not have single-cell proteomics data, we do not know the empirical change in transcript abundance due to whole chromosome aneuploidy – and the extent of attenuation is higher in at least a subset of genes at the translational or post-translational level (Torres et al, 2007; Dephoure et al, 2014; Chick et al, 2016; Gonçalves et al, 2017). Therefore, these estimates only consider the RNA measurements from these cell lines. We calculated that cell lines with Chromosome 1 aneuploidies had between 3.44 – 9.06% cells with Chromosome 1 duplications, cell lines with Chromosome 6 aneuploidies had between 3.9 – 14.39% cells with Chromosome 6 duplications, cell lines with Chromosome 8 aneuploidies had between 2.50 – 11.63% cells with Chromosome 8 duplications, and cell lines with Chromosome 11 aneuploidies had between 2.99 – 17.27% cells with Chromosome 11 duplications (Figure 3.10). Given the low percentages and how quickly aneuploid cells

can spread with their proliferative/survival advantages, these aneuploidies likely represent a very early stage of emergence within these lines. These are rough estimates due to the constraints of the data we have available and the variability in the euploid line

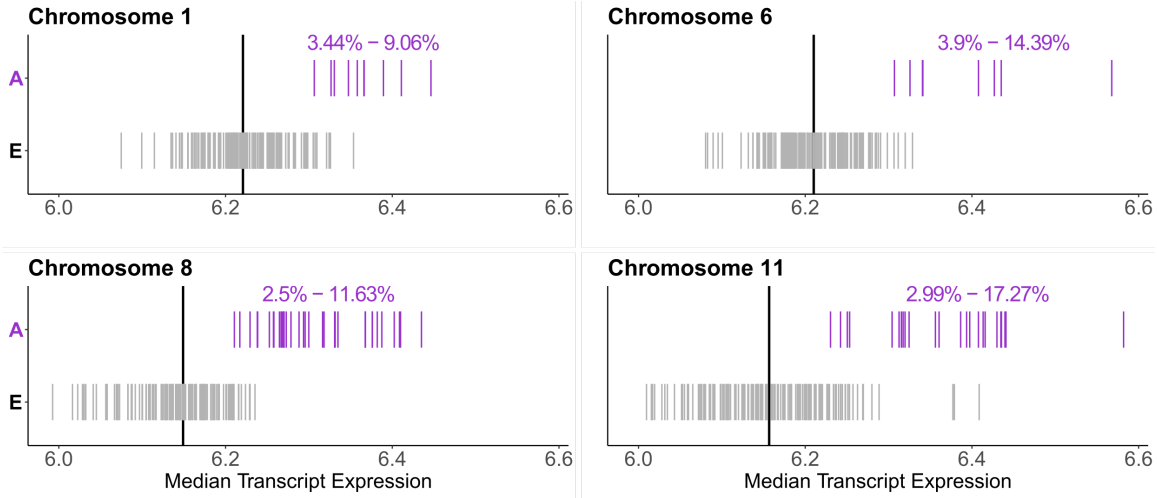


Figure 3.10 - Calculations of aneuploid content estimate low percentages of abnormal cells.
Each vertical line represents a single cell line's summary transcript expression for the specified chromosome – grey lines represent euploid annotated lines, purple lines represent aneuploid annotated lines ('E' is euploid and 'A' is aneuploid). Aneuploid proportion ranges were calculated as described in 2.7).

cluster due to genetic background. For this reason, we are using these only as an indication of how extensive the aneuploid subpopulation is within the aneuploid annotated lines we will be interrogating.

3.2. Analysis of factors associated with the development of aneuploidy in the DO mESC

3.2.1. Comparison with hPSC duplications

Before investigating how genetic diversity affects the development of aneuploidy, we first wanted to understand how the gene dosage changes due to duplications in mESCs compared to the gene dosage changes in the hPSC context. Sets of genes from mouse Chromosomes 1, 6, 8, and 11 with human synteny were acquired from The

Jackson Laboratory's Synteny Browser (syntenybrowser.jax.org, Kolishovski et al, 2019) and mapped top their corresponding location in the human genome to determine how extensively mouse and human PSC duplications shared genes (Figure 3.11 A). On Chromosomes 1, 6, and 11, between 46.11 – 79.63% of syntenic genes were located within chromosomes or segmental regions duplicated in hPSCs (Figure 3.11 B). Between 14.04 – 22.38% of these genes had synteny with segmental duplications. These findings suggest that duplications of mouse Chromosomes 1, 6, and 11 in mESCs are likely to alter the cells in similar ways as hPSC duplications by acting on largely the same set of genes. In contrast to these broad similarities, mouse Chromosome 8 has very few syntenic genes whose human correlate is duplicated. Only 4.40% of genes on mouse Chromosome 8 with human synteny (36/819) matched hPSC whole chromosome duplications – none of those were located in the smaller segmental duplications. Despite being the most commonly duplicated chromosome among both panels of DO mESC lines, Chromosome 8 duplications act on very few of the same genes. We hypothesized that although the mouse Chromosome 8 duplication was not targeting the same genes, it would influence similar pathways through alternative targets.

To understand what pathways were being targeted by Chromosome 8 duplication in DO mESCs, we differential expression analysis was performed comparing individual cells annotated with only trisomy 8 to those that were classified as entirely euploid. The scRNA-seq data was used for this analysis because it would provide a cleaner comparison between aneuploid and euploid conditions. The list of activated biological process GO terms includes mostly terms expected for aneuploid PSCs (Figure 3.12). Terms like DNA replication, ribosome biogenesis, RNA processing, nucleic acid metabolic process, and various terms regarding tRNA, rRNA, and ribonucleoprotein

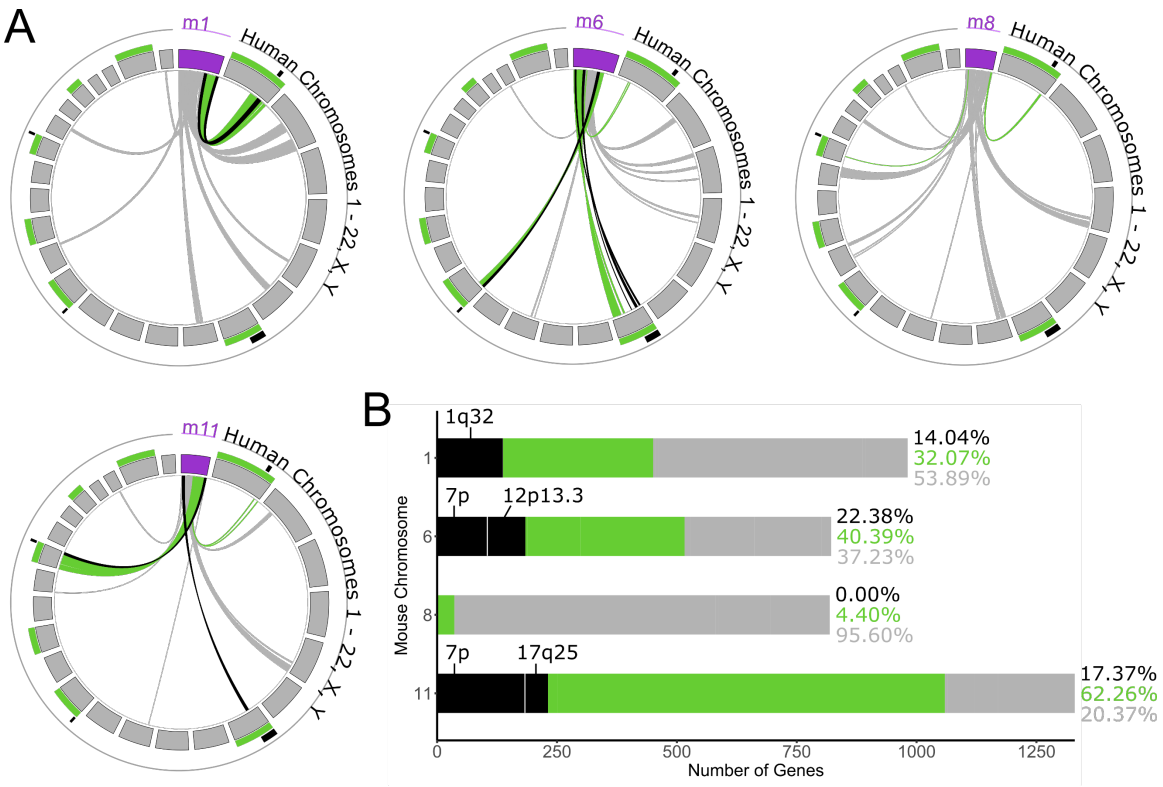


Figure 3.11 - Recurrent mESC duplications except chromosome 8 share a many of genes with hPSC duplications.
(A) Genes syntenic between duplications found in the DO mESC data sets with arcs connecting their locations on the mouse chromosome to their location in the human genome. Arcs are colored based on where they connect in the human genome. The track of inside bars are chromosomes – the purple bar represents the duplicated mouse chromosome being used as a reference and the grey bars represent human chromosomes. The middle track of green bars represents common whole chromosome duplications in hPSCs (chromosomes 1, 7, 12, 14, 17, 20, and X). The outside track of black bars represents common smaller segmental duplications in hPSCs (chromosomal regions 1q32, 7p, 12p13.3, and 17q25). (B) Quantification of the number of mouse-human syntenic genes by their relationship to hPSC duplications.

(RNP) processing make sense because these cells need to handle a large-scale increase in gene dosage. Similarly, DNA replication is in line with the known proliferative advantages described in aneuploid PSCs. The inclusion of cellular response to DNA damage is also a process identified by colleagues at Sheffield University as a key to the increased survival of aneuploid human PSCs. Curiously, all of the suppressed processes appear to be immune related with the exception of cell development – however it is not

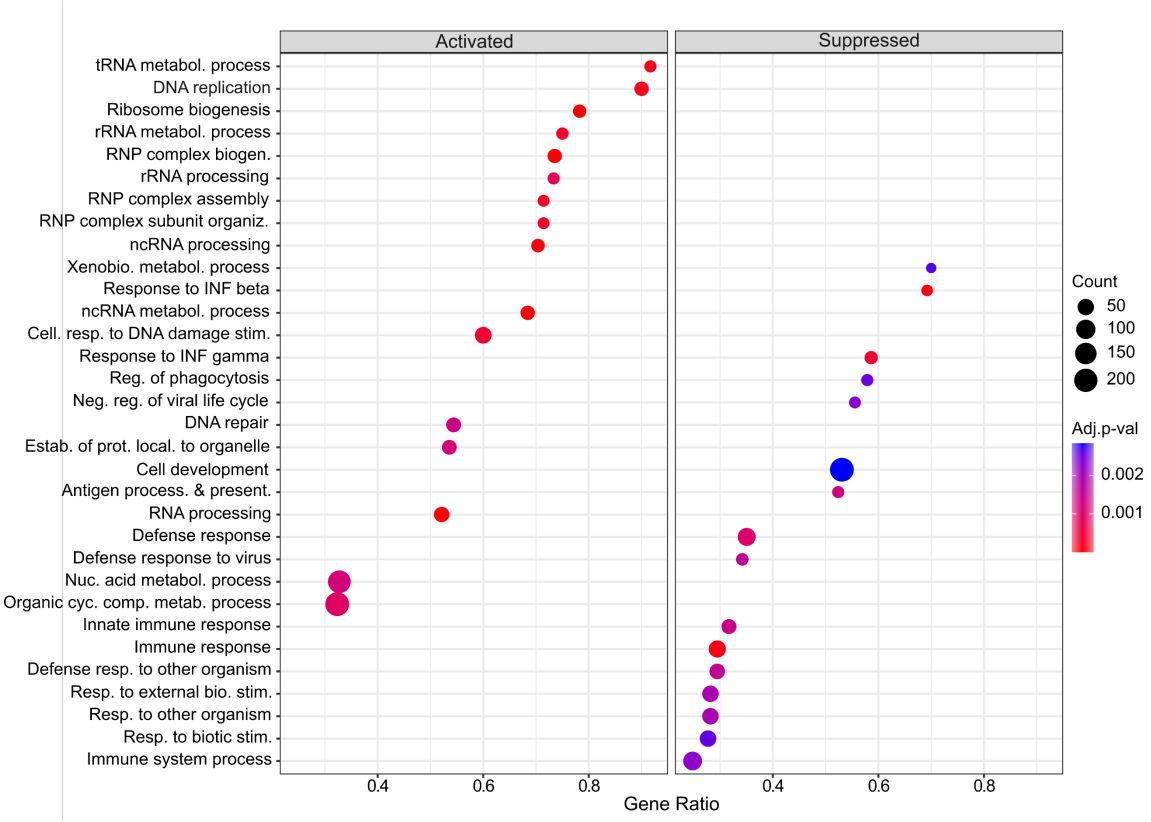


Figure 3.12 - Differentially expressed pathways in trisomy 8 single cells.
Top activated and suppressed pathways in the scRNA-seq data set when comparing cells with only trisomy 8 and completely euploid cells.

clear how these pathways relate to the known changes during aneuploidy in PSCs, although a better understanding of the relationship between aneuploidy and immune response in cancer is being developed (Kuang and Li 2022; Xian et al, 2021). These results suggest that the upregulated pathways are what link mESCs with Chromosome 8 aneuploidy to the other aneuploidies in mouse or human PSCs.

3.2.2. Haplotype analysis of allele-specific expression data

We began assessing genetic factors influencing aneuploidy in mESCs by investigating whether we can identify the specific duplicated chromosome in a sample and then identify cis-acting alleles that increase susceptibility to duplication on that chromosome. Genotyping the DO mESCs using the Genotype Reconstruction by RNA-

Sequencing (GBRS) pipeline generates allele-specific expression (ASE) data for each cell line that can be used to identify translational disparities between alleles (Choi et al, 2020). If we can approximate reconstructions of phased chromosomes using the genotype data, we can determine if either chromosome demonstrates skewed global expression relative to the other. In aggregating the chromosomes that appear to be specifically duplicated, we can then test for statistical over-representation of founder alleles in that subset. High representation of certain alleles would suggest either those chromosomes with those alleles are more often duplicated or more likely that cells which duplicate chromosomes containing those alleles have a selective advantage in cell culture.

Before testing for chromosome-biased ASE, we needed to reconstruct the phased chromosomes of each of the DO mESC lines. Under ideal conditions this would be done with long-read sequencing where it could be clear which alleles were represented on the same strand. Since we do not have that data, we created a methodology that uses the principles of parsimony – assuming the fewest number of cross over events – to approximate which alleles belong to which chromosome. Regions where neither allele were present on the chromosome up to that point do not have a most parsimonious resolution. In these cases, alleles were randomly assigned at that SNP until reconstruction was complete. After the first pass of reconstruction the median ASE ratio of each region between ambiguous site was calculated, then all regions of higher ASE were assigned to the same chromosome. This decision was made under the belief that if a chromosome were duplicated, this would make it easier to identify since the skewed ASE would not shift between chromosomes. If there was not a chromosome bias in

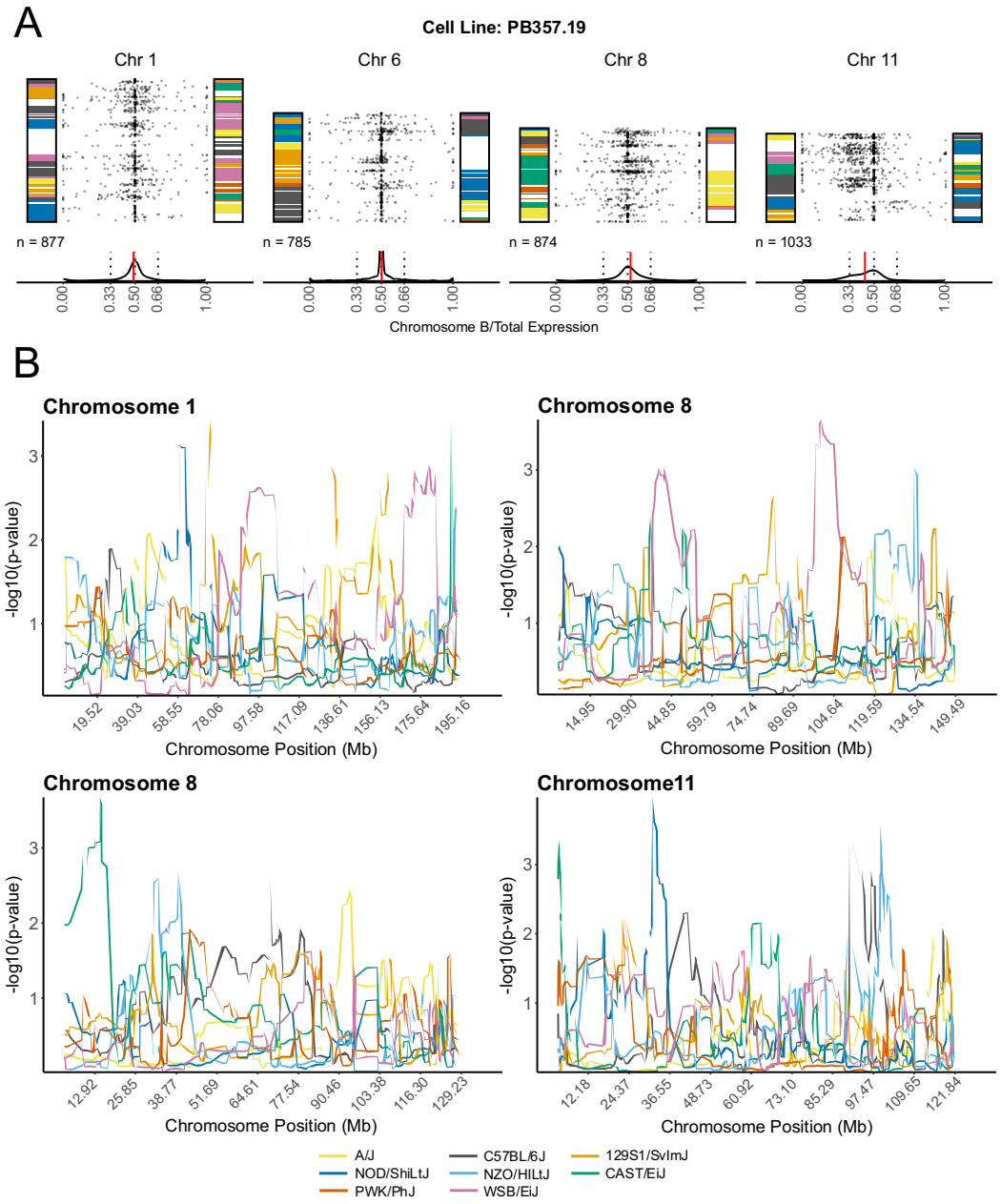


Figure 3.13 – No distinctly overrepresented founder alleles on specifically duplicated chromosomes.
(A) Depicted is an example DO mESC line with evidence of a specific duplication of Chromosome 11. Each black point represents a single gene, positioned at its genomic location along the y-axis and its allele-specific expression (ASE) ratio on the x-axis. Underneath the chromosomes are the number of genes on the given chromosome pair that are heterozygous in this example line. The density plots show the distribution of ASE ratios, with a red line indicating the median ASE ratio. (B) Results of the multivariate hypergeometric test for each chromosome with evidence of duplications in the ASE data for the DO mESC panel. Each line represents the statistical significance of that founder's allelic representation at the given locus.

ASE, then this methodology should not artificially represent that.

Following chromosome reconstruction, Chromosomes 1, 6, 8, and 11 were screened for chromosome-biased ASE and those chromosomes tested for allelic over-representation. ASE ratios were calculated at each location relative to a reference chromosome, where an ASE ratio of 0.5 meant balanced expression and a 0.0 or 1.0 meant sole expression from one chromosome (Figure 3.13 A). A bootstrapping approach was used to determine whether the mean ASE ratio of a sample's chromosome is significantly different from a random assignment of alleles to each chromosome. Using this method, we were able to identify 72 chromosomes that were specifically duplicated (p-value < 0.01), but there were also 7 cell lines that were excluded due to ASE data showing bias in all chromosomes. Without those cell lines, there were 4, 6, 13, and 24 duplicated haplotypes for Chromosomes 1, 6, 8, and 11 respectively. Using a multivariate hypergeometric test to determine statistically significant representation, we did not clearly identify any founder alleles more present on duplicated chromosomes than in the general population (Figure 3.13 B).

3.2.3. Genetic mapping of aneuploidy as a binary trait

Since the DO mice were designed as a genetic mapping population, we can perform a QTL scan for aneuploidy in the bulk data panel of 170 mESC lines as an additional line of evidence for cis-acting factors as well as discover any trans-acting ones. While no associations were identified for Chromosomes 1 and 6, these analyses were likely underpowered due to having few lines with these (10 and 8 lines respectively). We also performed a genetic mapping of aneuploidy as a binary trait for all aneuploidies, Chromosome 8 duplications, and Chromosome 11 duplications correcting the sex covariates to reflect the detected changes in chromosomal sex (Figure 3.7). No statistically significant loci were discovered for Chromosome 8 or 11 duplications,

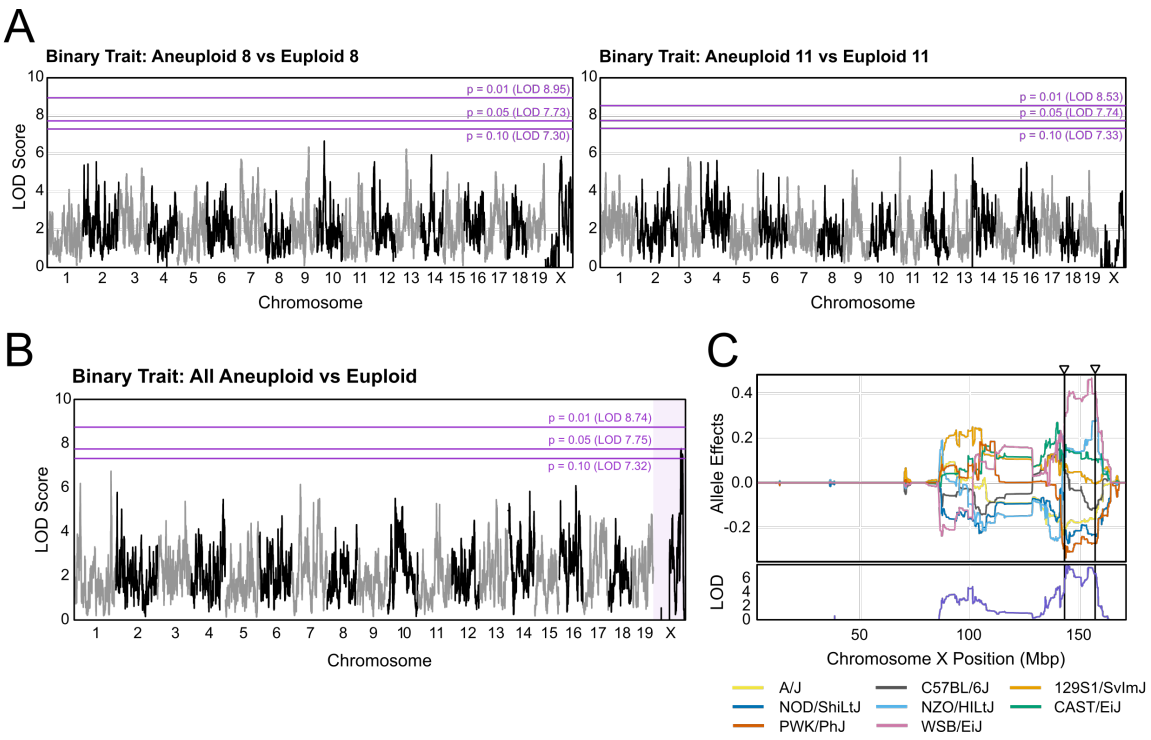


Figure 3.14 - Chromosome X is implicated in the development of general aneuploidy.
(A-B) Manhattan plots for SNP association with aneuploid chromosome 8 or 11 (A) or all aneuploidies (B) as a binary trait and correcting sex covariates for XX samples losing a chromosome X. Purple bars represent statistical significance thresholds for aneuploidy association determined by permutation testing. Light purple background behind a chromosome indicates presence of a statistically significant locus on that chromosome. (C) The upper panel is the founder allele associations with aneuploidy by position on chromosome X. Higher scores indicate association with aneuploidy, lower scores indicate association with euploidy. The lower panel displays a zoomed in version of the chromosome X LOD scores from the Manhattan plot. Vertical black bars and arrows bracket the identified chromosome X locus (Chr X: 143,530,457 – 157,497,091).

however a locus at position 144,485,716 on Chromosome X with LOD 7.77 ($p < 0.05$) was associated with generalized aneuploidy (Figure 3.14 A-B). The genomic range covered by a LOD drop of 1.5 is around a 13.97 Mbp region between 143,530,457 – 157,497,091 where the founder alleles from WSB/EiJ are associated with developing aneuploidy, but where the other alleles occupy a spread of effect sizes not clearly split into groups (Figure 3.14 C). These mapping analyses show that there are likely no genetic factors that drive duplication of any specific chromosomes we can test in this

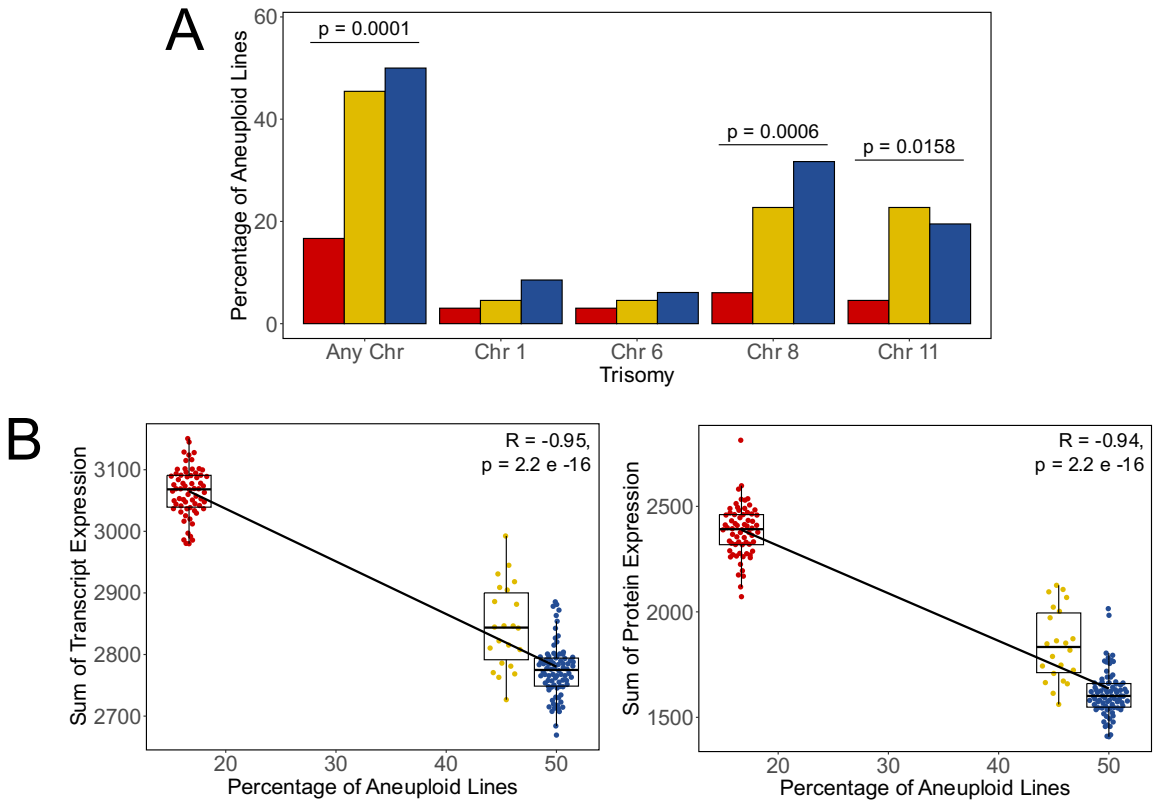


Figure 3.15 - XX DO mESCs have a lower proportion of aneuploid lines compared to XY or XO lines.
(A) Proportion of aneuploid to euploid cell lines compared between XX, XO, and XY chromosomal sexes. (B) Correlation between proportion of aneuploid to euploid cell lines and sum of chromosome X gene expression within each chromosomal sex. Red bars or points represent XX cell lines, yellow bars or points represent XO cell lines, and blue bars or points represent XY cell lines.

data set, whereas one or more genes on Chromosome X influence chromosomal duplications generally.

3.2.4. Chromosomal sex bias in DO mESC aneuploidy

Following from the identified aneuploidy locus on Chromosome X, we assessed whether there existed any disparity in aneuploidy between the chromosomal sex categories. The percentage of XX lines annotated as aneuploid were significantly lower than the percentage of aneuploid XY or even XO lines (16.67% compared to 50.00% or 45.45%, respectively – p = 0.0001; Figure 3.15 A). The difference in aneuploid

percentage was also observed between XX and XY or XO lines for Chromosome 8 and 11 duplications as well (p = 0.0006 and 0.0158, respectively). The number of cell lines with duplications of Chromosome 1 or 6 were too few to observe any difference between lines of different chromosomal sexes. With the exception of Chromosome 11, the XO lines had proportions of aneuploid lines that were near the XY proportions, but still between XX and XY. Since these lines spent some of their history as wholly XX and a portion of their mosaic might still be XX, the aneuploidy differences might be due to the

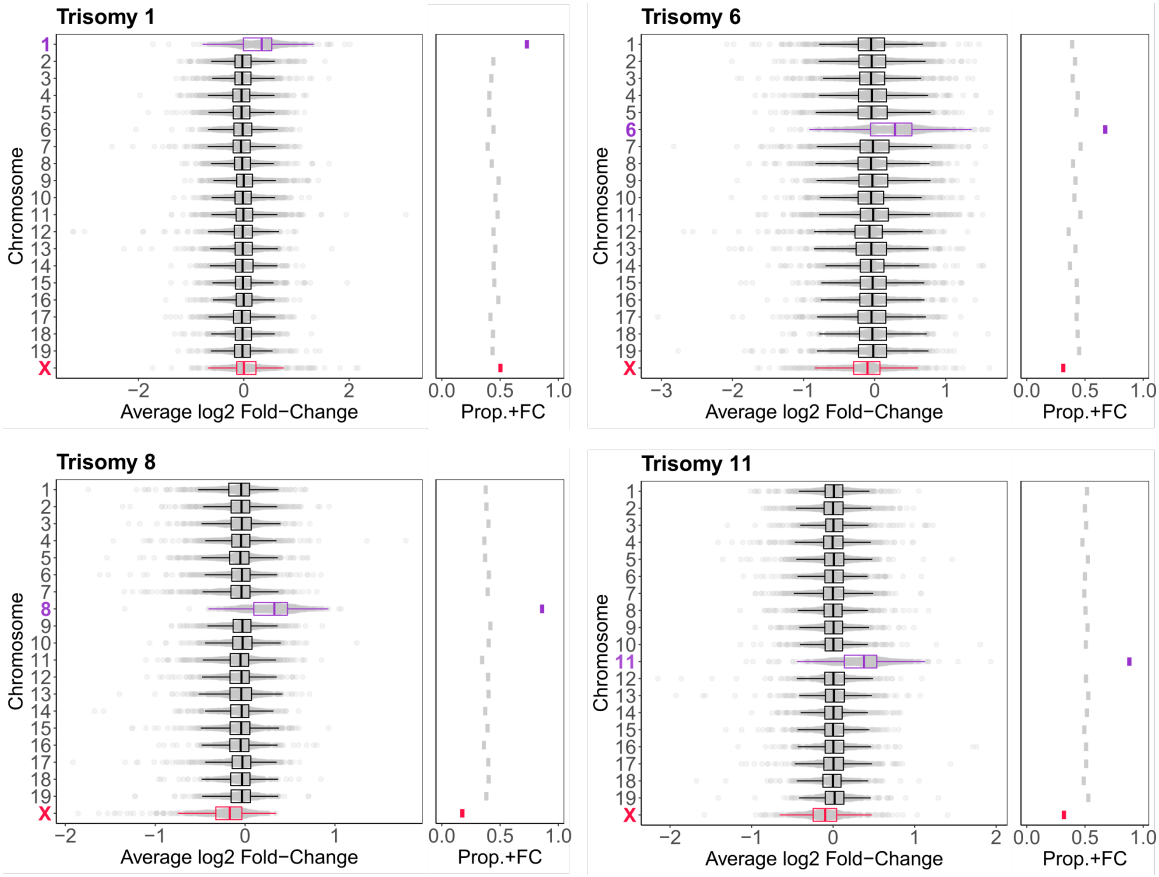


Figure 3.16 - Individual Aneuploid XY cells have lower chromosome X expression than euploid cells.
For each plot, differential expression analysis was performed on individual cells with only the specified trisomy compared to completely euploid cells. The left panels display log2 fold-change distributions for genes on each chromosome. The right panels display the proportion of genes on each chromosome with a positive log2 fold-change. Purple bars and points represent the duplicated autosome in each comparison. Blue bars and points represent chromosome X – since all cells in the scRNA-seq data set are XY.

loss of their second Chromosome X. There is a strong negative correlation between the expression of chromosome transcript or protein and the proportion of aneuploid lines between XX, XO, and XY chromosomal sexes ($R = -0.95$, $p < 0.01$ and $R = -0.94$, $p < 0.01$ for transcript and protein respectively; Figure 3.15 B). These results suggest that the dosage of Chromosome X, and consequently the extent of expression of X-linked genes, influences the development of aneuploidy in mESC lines.

While there is not a diversity of chromosomal sex in the single-cell data to run the same comparisons on, we surprisingly found additional evidence of an inverse relationship between X-linked gene expression and aneuploidy within those XY lines as well. Differential expression analysis was performed to compare individual cells that were aneuploid for only one specific autosomal duplication to cells that were entirely euploid. The results were then broken down by chromosome to understand how the gene expression changes are distributed throughout the genome (Figure 3.16). As expected, the genes from the duplicated chromosome show a distribution with a higher magnitude of log2 fold-change and a higher proportion have positive log2 fold-changes as well, compared to other autosomes. The other autosomes all have near zero log2 fold-change distributions and positive log2 fold-change proportions that are around 0.5. However, with the exception of trisomy 1, the changes in Chromosome X gene expression seems to be lower compared to the euploid cells. For trisomy 6, 8, and 11 the average log2 fold-change is lower than the non-duplicated autosomes and there is a noticeably lower proportion of genes with positive log2 fold-change. This indicates that the aneuploid cells in the scRNA-seq data set have lower expression of Chromosome X genes than the euploid cells. Because we only have one time-point and all the cell lines began with the same Chromosome X dosage, we cannot differentiate whether individual cells with stochastically lower Chromosome X expression were more likely to develop

autosomal duplications similar to what we observe in the bulk data or autosomal duplications induce downregulation of Chromosome X genes.

To understand whether the lower Chromosome X expression in aneuploid cells was due to random or shared gene expression differences, we compared X-linked gene expression changes between different trisomies. There was statistically significant correlation of gene expression changes in comparisons between trisomy 6, 8, and 11 ($p < 0.001$; Figure 3.17 A). There was also a much less significant negative correlation between trisomy 1 and trisomy 6, although this effect was small and may be due to these trisomies having the least and most influence on Chromosome X genes, respectively (Figure 3.16 and Figure 3.17 A). The number of upregulated Chromosome X genes shared between trisomies decreases as you increase the number of cases compared, while the number of downregulated Chromosome X genes increases considerably as you compare more trisomies – even including trisomy 1 (Figure 3.17 B). A probability distribution was generated by maintaining the number of genes with positive, negative, and zero log2 fold-change, but randomly assigning which of the 620 Chromosome X genes they correspond to. Sharing 136/620 Chromosome X genes with lesser expression was a statistically significant similarity ($p < 0.01$), suggesting that the change in Chromosome X gene expression in aneuploid cells is non-random.

3.2.5. Chromosomal sex bias in hPSC aneuploidy

While both mESC data sets display an inverse relationship between expression of Chromosome X genes and abnormal karyotype, we wanted to understand if this pattern could be observed in hPSCs as well. Because hPSCs are unlikely to have both X Chromosomes active, but can still express a subset of genes on an eroded

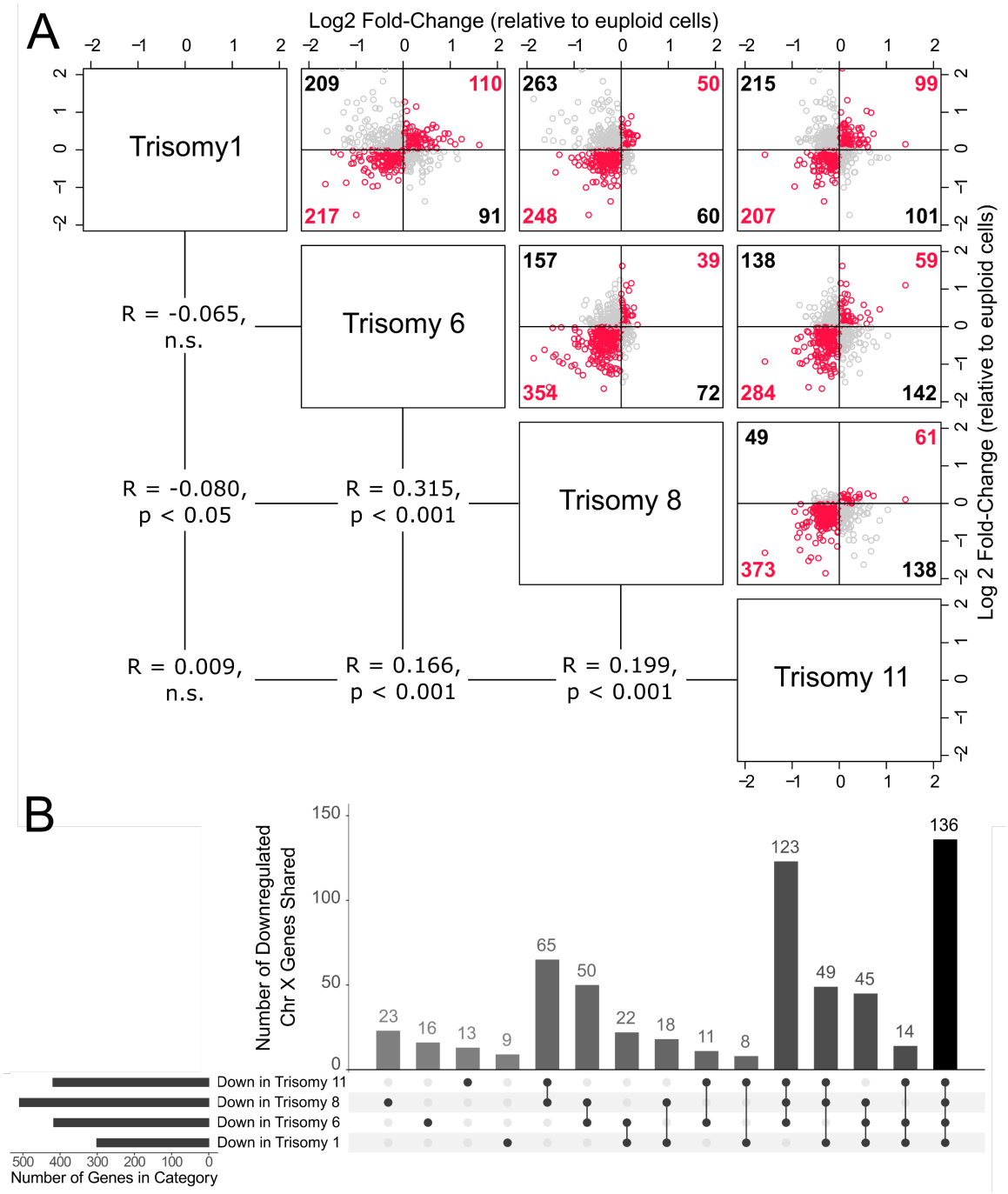


Figure 3.17 - Directionality of X-linked gene expression differences is shared between individual chromosome trisomies.
(A) Correlations between chromosome X gene expression changes in each trisomy (log2 fold-change). The numbers in the corners of each plot represent the number of genes in their respective quadrant. (B) The number of genes with negative log2 fold-change shared between different categories of trisomy.

Chromosome X or through XCI escape, there was uncertainty about whether Chromosome X dosage would have a significant effect in the human context. To address this question, we partnered with collaborators at The WiCell Institute who have performed G-banded karyotyping on thousands of user-submitted hESCs and hiPSCs. Cell lines were grouped based on chromosomal sex and cell type – which was confirmed

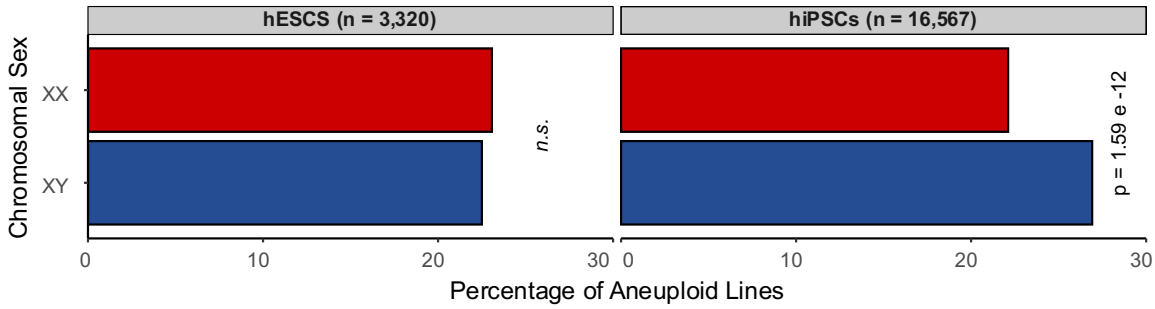


Figure 3.18 - Fewer XX hiPSC lines, but not hESCs, are aneuploid in G-band karyotyped data from the WiCell Institute.
Red bars represent XX hPSCs, blue bars represent XY hPSCs (confirmed during karyotyping). Data are split between hESCS and hiPSCS – primarily the latter – cultured in a variety of conditions.

Table 3.2 - Aneuploidies in WiCell Institute karyotyped hPSCs. G-banded karyotyping data from The WiCell Institute reported by chromosomal sex and whether cells were hESCs or hiPSCs.

	XX	XY
hESCs	**1547**	**1773**
Euploid	1190	1374
Aneuploid	357 (23.08%)	399 (22.50%)
hiPSCs	**7145**	**9422**
Euploid	5565	6886
Aneuploid	1580 (22.11%)	2536 (26.92%)

during karyotyping – then annotated as aneuploid if the line contained any chromosomal duplications or losses. The WiCell data set is primarily made up of iPSC samples, which are most likely in an XaXi and to a lesser extent XaXe status (Bar et al, 2019). However, the XX hiPSCs had statistically fewer aneuploid lines than the XY hiPSCs, but not in the case of the hESCs (22.11% compared to 26.92%, p = 1.59 e -12; Figure 3.18). Chromosomal sex has a much more subtle effect in hiPSCs, but this data set is an amalgam of lines maintained in a myriad of culture conditions which may obscure any effect different practices may have on how the chromosomal sex bias emerges.

3.2.6. Two active X Chromosomes increases 2CLC gene expression

Since the chromosomal sex bias in aneuploidy appears to be a consequence of Chromosome X dosage, we assessed the dosed sensitivity of X-linked genes within our RNA-seq and proteomics data sets. Since some genes can have gene dosage differences buffered at the protein level as explained in the introduction, the correlation between transcript and protein abundance can help discover genes with the least attenuation. We specifically keyed into six candidate genes – one of which was not present in the proteomics data set – from a review on chromosomal sex bias in pluripotency (Schulz 2017). The MAPK superfamily inhibitor dual specificity phosphatase 9 (*Dusp9*) had much higher transcript-protein correlation than the other four genes (Figure 3.19 A and B). As expected, based on gene dosage, the XY and XO lines cluster together by transcript and protein expression of *Dusp9* (Figure 3.19 B). Intriguingly, the expression of *Dusp9* is not only lower in XY and XO lines, but is also expressed lower in aneuploid lines compared to euploid lines of the same chromosomal sex (ANOVA chromosomal sex $p < 2 e -16$ for transcript and protein, karyotype $p = 0.109$ for transcript and $p = 0.0435$ for protein; Figure 3.19 C). *Dusp9* is a compelling clue regarding the

mechanism for the chromosomal sex bias in aneuploidy since it has lower abundance in both aneuploid lines and the lines more likely to become aneuploid.

The identification of Dusp9 lead us to consider whether differences in the 2-cell-like (2CL) state might mediate the differences in aneuploid development. As noted in the introduction, the feeder cell cultures that were most beneficial for promoting this state in mESCs resulted in a reduction of *Dusp9* expression. To address this question, we compared the expression of transcripts upregulated in *Zscan4*$^{+}$ 2CL cells (2CLCs)

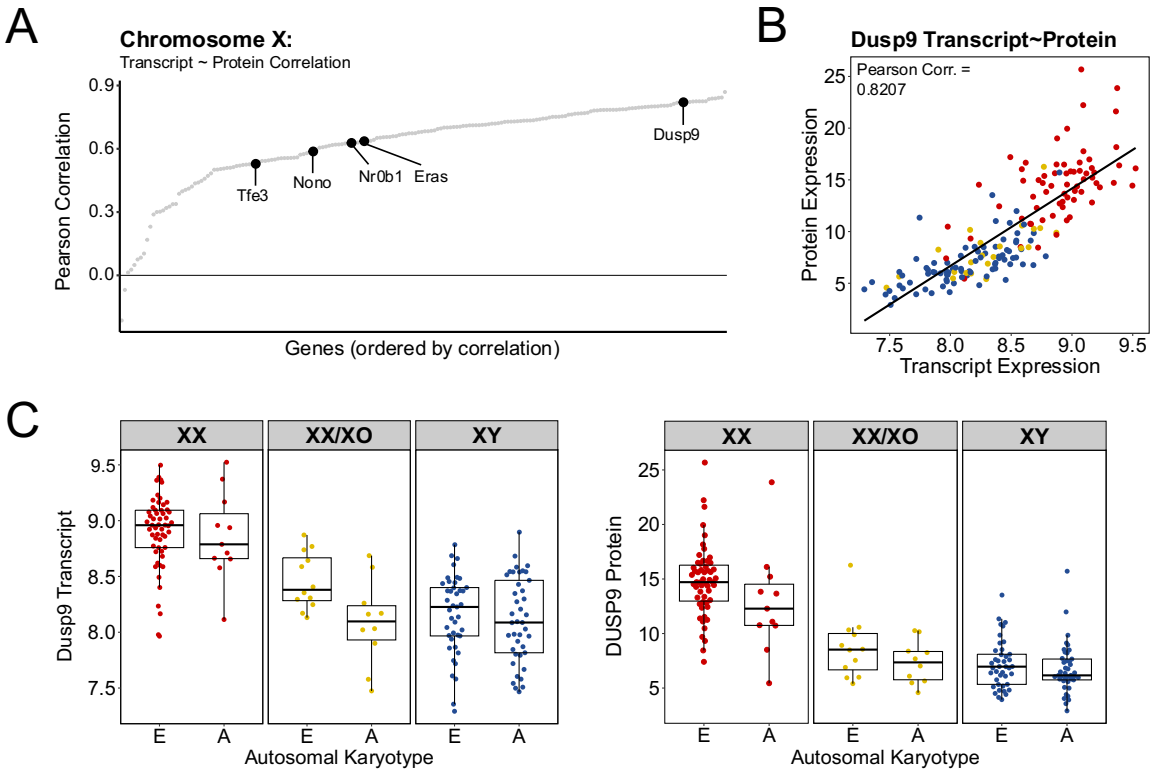


Figure 3.19 - Chromosome X gene Dusp9 is highly dose-sensitive and expressed lower in aneuploid lines.
(A) Pearson correlation of chromosome X gene transcript and protein expression to determine gene dosage sensitivity. Each point represents a gene with detected transcript and protein expression in the bulk data set. Black points represent genes identified in Schulz 2017 as candidates for sex bias in mESC pluripotency (Zfx protein was not detected). (B) Detailed view of Dusp9 transcript and protein correlation by cell line showing separation of expression by chromosomal sex. (C) Dusp9 transcript and protein expression split by chromosomal sex and autosomal karyotype ('E' is euploid and 'A' is aneuploid). Red points represent XX cell lines, yellow points represent XO cell lines, and blue points represent XY cell lines.

published in Macfarlan et al, 2012 and Amano et al, 2013 – of which, 65.52% and 45.45% were represented in our RNA-seq data set, respectively. The distribution of the sum of 2CLC z-scores is significantly higher for XX lines than in the XY or XO lines (Figure 3.20). The higher 2CL character of the bulk transcriptome suggests that these lines have either a higher percentage of 2CLCs or stronger expression of 2CL genes from their 2CLC population. Because these data are bulk measurements and the scRNA-seq data are all XY lines with too few cells to see significant differences in this rare population, we cannot directly test for population size differences in the DO context with these data sets.

To test whether mESC lines with dual Chromosome X dosage have larger 2CLC subpopulations, we directly quantified the number of 2CLCs in mESCs of different chromosomal sex from the same C3H/HeSn-Paf/J genetic background. We acquired nine independently derived mESC lines (three of each XX, XO, and XY), cultured them in the same in the same conditions (feeder plates in serum + LIF media with no inhibitors), and performed flow cytometry using antibodies against the gag protein of the

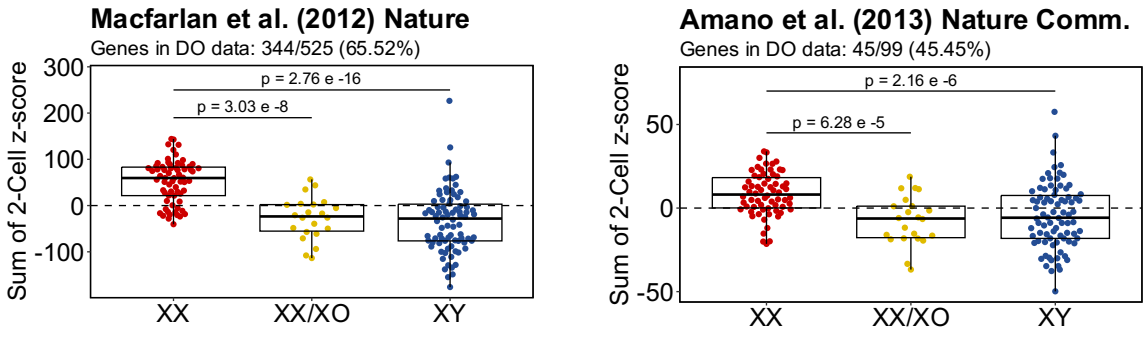


Figure 3.20 - XX DO mESCs have increased 2CLC gene expression profiles compared to XY or XO.
Expression of genes upregulated in Zscan4^{+} 2CLCs (gene sets from Macfarlan et al, 2012 and Amano et al, 2013) compared between XX, XO, and XY chromosomal sexes. Points represent the sum of z-scores for 2CLC genes found in our bulk RNA-seq data set for an individual cell line. Red points represent XX cell lines, yellow points represent XO cell lines, and blue points represent XY cell lines.

2CLC marker murine endogenous retrovirus L (MERVL). All six of the cell lines with a single Chromosome X had similar populations of 2CLCs, between 0.16 – 0.27% of each culture (bottom two rows of Figure 3.21 A and right side of Figure 3.21 B). The three XX lines had a large variance in 2CLC population – ranging from 0.39 – 0.94% – but was sizeable difference from the single Chromosome X lines (p = 0.0896; Figure 3.21 A and B). While these population differences are subtle, 2CLCs within mESC cultures are a rare population that are often not often representing more than a few percent of cells depending on genetic background or culture conditions.

3.2.7. Two active X Chromosomes increases expression of tumor suppressor genes involved in DNA damage response

Because no totipotent populations have been observed in hPSCs cultured in primed pluripotency conditions, we investigated whether expression of any X-linked

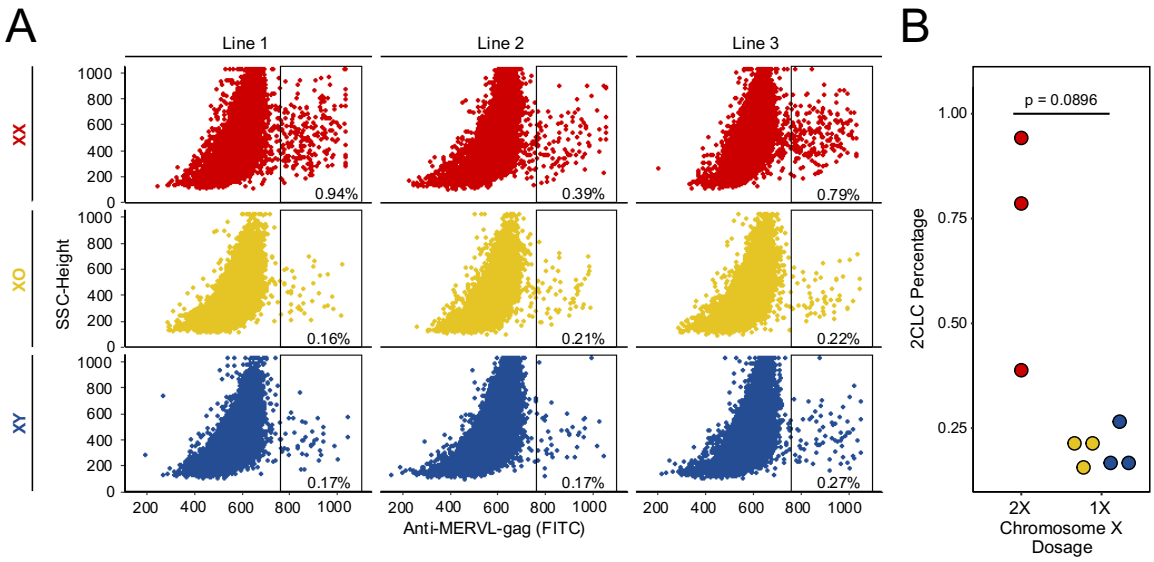


Figure 3.21 - XX mESC lines have a higher proportion of 2CLCs compared to XY or XO mESC lines from the same genetic background.
(A) Flow cytometry of C3H/HeSn-Paf/J mESC lines stained for the 2CLC marker MERVL gag protein. Black boxes are drawn around the gates used to identify 2CLCs in each cell line. (B) The percentage of cells in each line identified as 2CLCs grouped by number of X Chromosomes. Red points represent XX cell lines, yellow points represent XO cell lines, and blue points represent XY cell lines.

tumor suppressor genes (TSGs) could also play a role in the observed chromosomal sex bias. Here we have taken direction from the cancer literature since aneuploidy is a hallmark of cancer as well as long term PSC culture. As in our model, certain cancers show a chromosomal sex bias driven by Chromosome X dosage as exhibited by increased tumor risk in XO compared to XX individuals and lower risk in XXY compared to XY individuals (Ji et al 2016). One mechanism proposed for this bias is the differential expression of XCI escape of TSGs located on Chromosome X (Dunford et al, 2017). A list of 24 protein coding TSGs was acquired from The Tumor Suppressor Gene Database and we repeated the analysis used to identify dosage sensitive genes with decreased expression in aneuploid cell lines of matched chromosomal sex. Intriguingly, *Dusp9* is considered a tumor suppressor gene as well as a main contributor to chromosomal sex differences between mESCs. Two isoforms of *Fhl1* and *Kdm6a* had similar dosage sensitivity to *Dusp9* (Fig 3.22 A). Of these, only *Fhl1* was significantly correlated with both chromosomal sex and aneuploidy in the bulk data sets (ANOVA chromosomal sex $p < 2 e -16$ for transcript and $p = 1.1 e -14$ for protein, karyotype $p = 0.0013$ for transcript and $p = 0.0017$ for protein; Fig 3.22 B). *Gpc3*, *Dmd*, and *Rbbp7* showed intermediate dosage sensitivity, whereas the rest of the TSGs fell near the steep drop off of transcript-protein correlation.

We also assessed the expression of these X-linked TSGs in the DO scRNA-seq data set to determine if the decreased Chromosome X expression those XY aneuploid cells included these genes. When looking at the log2 fold-change and p-values for differential expression of the X-linked TSGs in all aneuploid cells compared to all entirely euploid cells, there was a clear grouping of lower expressed genes including *Eda2r*, *Fhl1*, *Flna*, *Rbmx*, *Dusp9*, *Kdm6a*, *Zfp185*, *Rbbp7*, *Dmd*, *Rps6ka6*, and *Rpl10* (Figure 3.23 A). The three genes noted previously (*Dusp9*, *Fhl1*, and *Kdm6a*) all fall within the

top six most differentially expressed. Because not all trisomies are represented equally in this scRNA-seq data set, we next compared the log2 fold-change in expression between the trisomies that had a noticeably lower expression of Chromosome X (Figure 3.23 B). From these comparisons, it appears that most of the TSGs in this group have variable differential expression between trisomies. For example, *Dusp9* is mostly low in the context of trisomy 11 and *Fhl1* is mostly low in the context of trisomy 8. Intriguingly, *Flna*, *Kdm6a*, *Rbmx*, and *Zfp185* are all connected to DNA damage response (Adamson

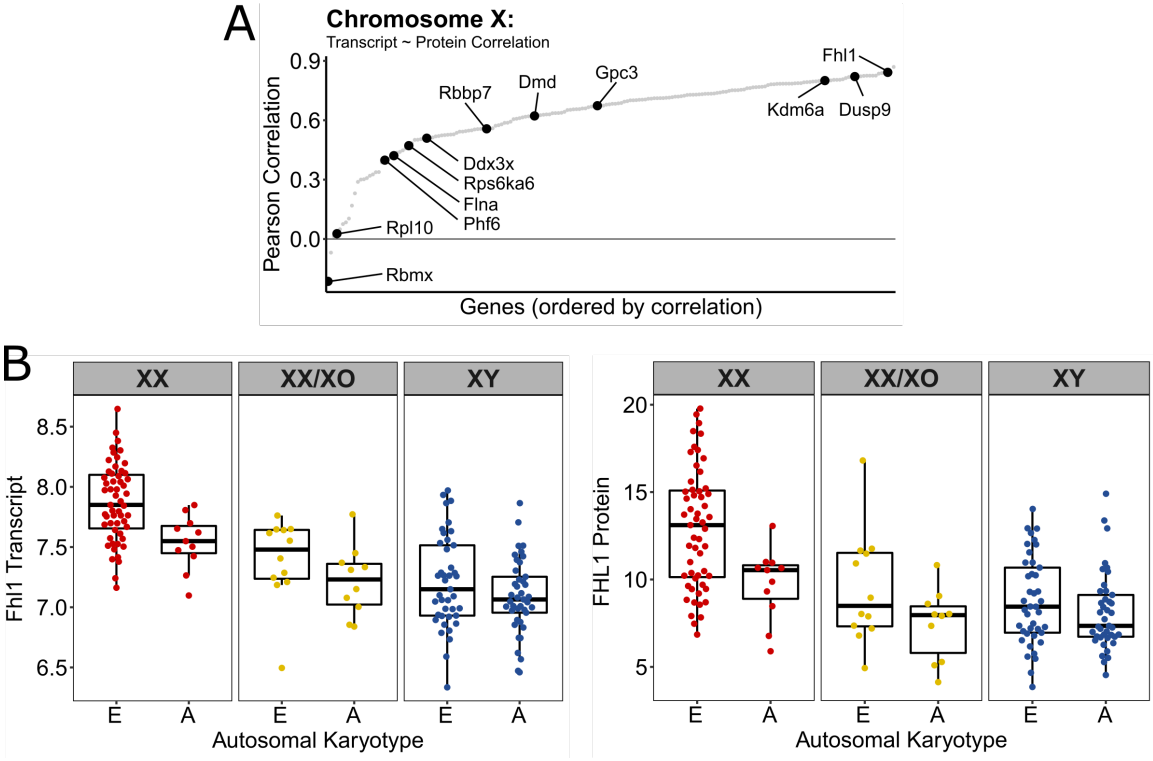


Figure 3.22 - Chromosome X tumor suppressor gene Fhl1 is dose-sensitive and expressed lower in aneuploid lines.
(A) Pearson correlation of chromosome X gene transcript and protein expression to determine gene dosage sensitivity. Each point represents a gene with detected transcript and protein expression in the bulk data set. Black points represent tumor suppressor genes identified by The Tumor Suppressor Gene Database. (B) Fhl1 transcript and protein (isoform ENSMUSP00000110417.2) expression split by chromosomal sex and autosomal karyotype ('E' is euploid and 'A' is aneuploid). Red points represent XX cell lines, yellow points represent XO cell lines, and blue points represent XY cell lines. Only the highest correlated Fhl1 isoform is represented graphically, although both have very similar correlation.

et al, 2012; Klaric et al, 2021; Koch et al, 2021; Matsunaga et al, 2012; Smirnov et al, 2018; Velkova et al, 2010; Zheng et al, 2020). One particularly interesting connection is the role of *Rbmx* in sister chromatid cohesion where its depletion leads to aberrant mitosis and improper chromosome segregation (Cho et al, 2018; Matsunaga et al,

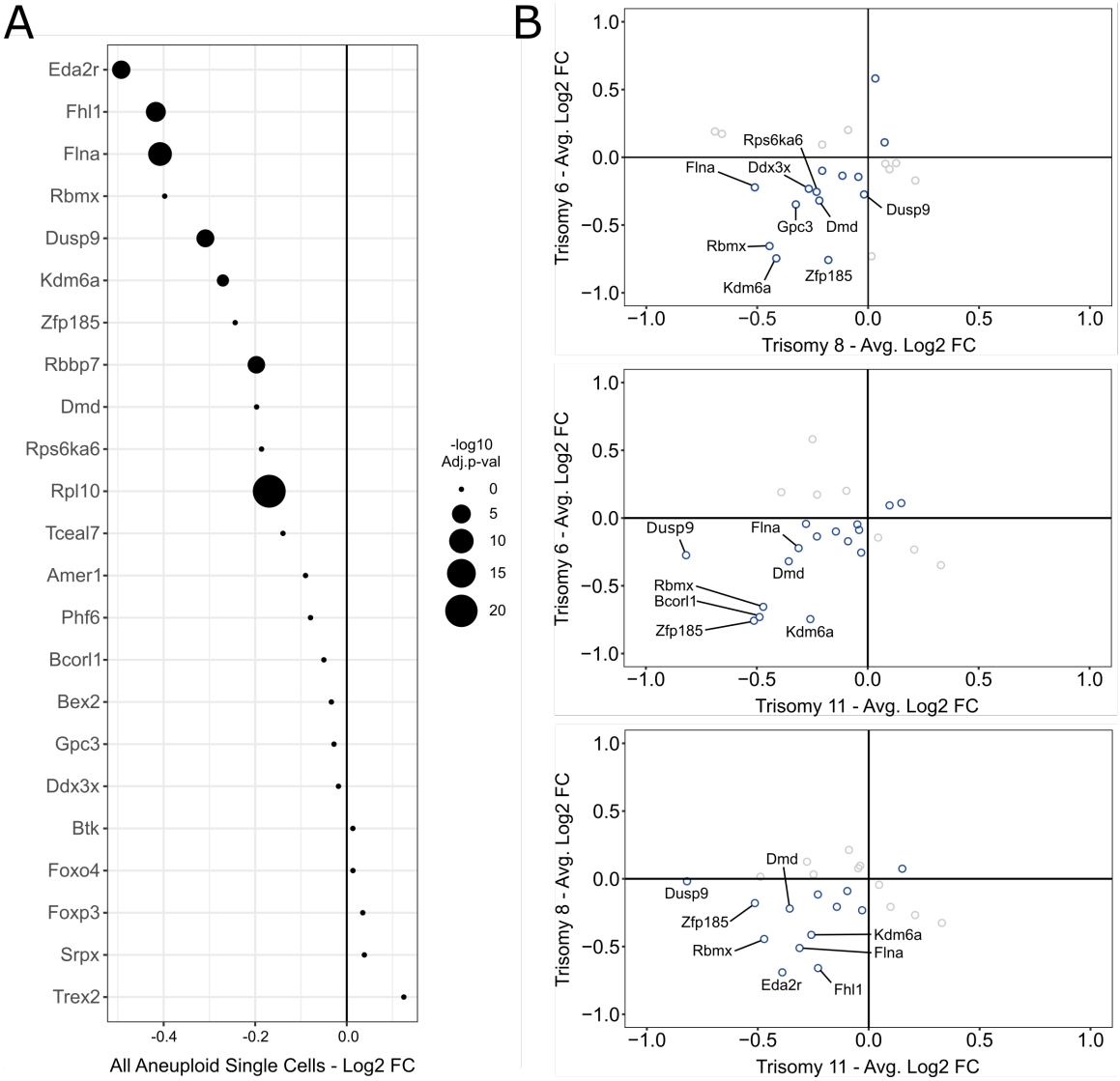


Figure 3.23 - Aneuploid cells have similarly decreased expression of several chromosome X TSGs.
(A) Log2 fold-change in expression of X-linked tumor suppressor genes when comparing cells with any combination of aneuploidy to entirely euploid cells. (B) Comparisons between X-linked TSG expression changes in the three aneuploidies with decreased chromosome X expression – chromosomes 6, 8, and 11.

2012). Additionally, deficiency of *Dmd* has also been demonstrated to lead to genomic instability in hPSCs through reactive oxygen species induced DNA damage (Jelinkova et al, 2019). While Dunford et al, 2017 has shown XCI escape of *Kdm6a* in human tumor data, escape can be cell type specific and we do not have data to compare whether these other TSGs of interest are expressed in hPSCs during XCI.

3.3. Characterizing the sexual dimorphism in aneuploidy development of mESCs

3.3.1. Experimental design for further chromosomal sex bias investigation

Given the results of Chapters 3.1 and 3.2, we first wanted to validate the findings that dual Chromosome X mESC lines accumulate less aneuploidy during prolonged culture than single Chromosome X mESC lines. To test this, we compared aneuploid cell

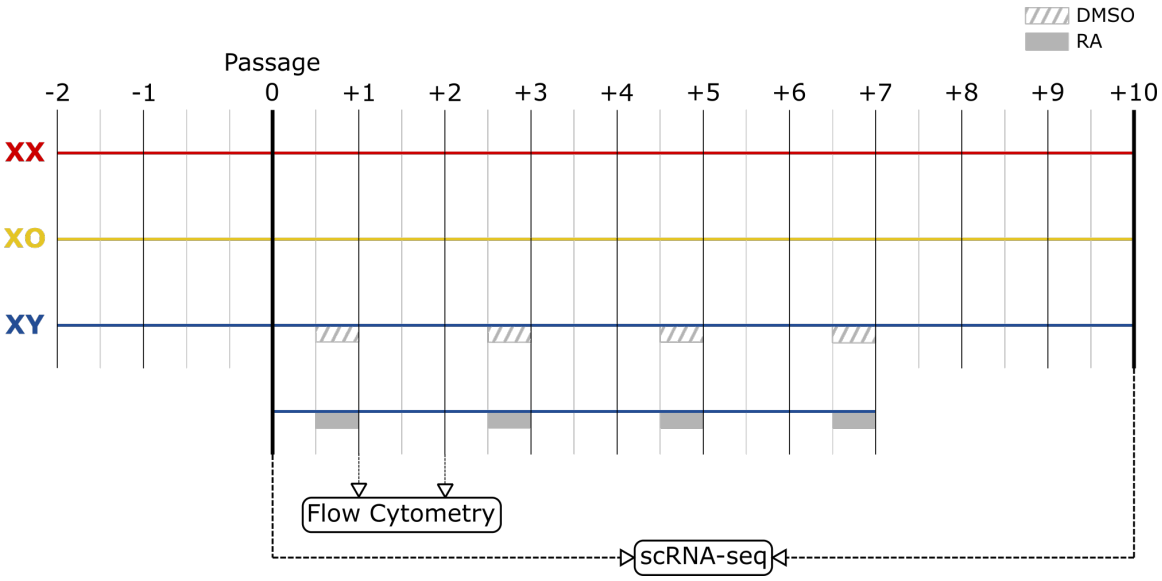


Figure 3.24 - Schematic for aging, treatment, and assays for XX, XO, and XY C3H/HeSn-Paf/J mESC lines.
The three cell lines of each chromosomal sex were acclimated from cryopreservation, then aged for ten passages between two scRNA-seq collection time-points. After the first collection, all XY cell lines were split into two cultures – treated with retinoic acid (RA) or an equal volume of DMSO for 24 hours before splitting every other passage. XY cells were analyzed by flow cytometry for the 2CLC marker MERVL-gag and the pluripotency-associated genes OCT3/4 immediately following and 48 hours after removal of the first RA treatment to assess 2CLC induction and return to pluripotency. Spaces between vertical lines represents 24-hour periods. The red line represents XX cell lines, the yellow line represents XO cell lines, and the blue lines represent XY cell lines.

proportions in virtually karyotyped scRNA-seq data collected from the nine inbred mESC lines used in Chapter 3.2.6 at a low passage number (early timepoint, P0) and ten passages later (aged timepoint, P+10) (Figure 3.24). Second, we wanted to test the hypothesis that elevating the proportion of 2CLCs in the single Chromosome X mESC lines to a similar abundance as the dual Chromosome X mESC lines is sufficient to reduce the accumulation of aneuploidies. To test this, we also cultured XY mESC lines treated with periodic low doses of retinoic acid (RA) to increase the number of 2CLCs. Ultimately the RA-treated cells were abandoned due to rapid, spontaneous differentiation – however, the details of this attempt will still be described to understand why we couldn't directly connect 2CLC proportions and aneuploidy accumulation, as well as a record of why this approach to artificially bolstering 2CLC proportions is not feasible as an approach to reduce aneuploidy. Third, we wanted to better understand which genes were enriched in cells that retained normal karyotype over time. To test this, we identified differentially expressed genes in aged euploid XX mESCs compared to early euploid mESCs.

3.3.2. Long-term low dose retinoic acid treatment yields differentiation

Because the cell lines used in this experiment are from a genetic background not represented in previous RA-treatment publications, we needed to experimentally determine a dosage and duration that would induce a higher, but physiologically normal proportion of 2CLCs in our XY mESC lines. Doses used in the literature increase the population significantly higher than our target since they are intending to enrich the 2CLC population for downstream applications. For this reason, we tested the low doses of 50nM and 100nM RA for a brief window of 24 hours. The proportion of 2CLCs in the untreated and vector control were similar to what was observed in the XY lines during the initial comparison of chromosomal sexes (Figure 3.21 and Figure 3.25). Short term

RA treatment at both doses seemed to induce a similar increase in 2CLCs (Figure 3.25). We opted to continue forward using the higher dose of 100nM RA to best avoid any human error bringing the 2CLC induction closer to normal XY range.

An additional result of RA treatment was a difference in pluripotency-associated gene expression from the low MERVL-gag cells. We also tested for OCT3/4 expression because its expression decreases in the 2CL state, unlike other pluripotency-associated genes. The point of highest density within the non-2CLCs becomes bimodal in RA-treated cultures, representing a population decreasing expression of OCT3/4, but not expressing higher MERVL-gag (Figure 3.25). Because there is an inverse relationship between these two genes in naïve pluripotent cells and 2CLCs, we believe this to represent cells in an intermediate state, cycling in or out of 2CLC identity.

Following the first RA treatment we performed OCT3/4 and MERVL-gag flow cytometry on all XY lines at the end of the first 24 hours with S/L+RA and 48 hours after return to standard S/L media. Our intermittent low-dose RA exposures were intended to artificially increase 2CL cycling, not to force mESCs into sustained 2CLC identity, therefore we wanted to ensure that cells returned to pluripotency during the recovery period. The RA-treated XY lines had the expected increase in 2CLC population accompanied by a downward shift in OCT3/4 from the naïve population as seen before, compared to a low 2CLC population and a single OCT3/4 high peak density in the DMSO-treated XY lines (Figure 3.26). Interestingly, although the increase in OCT3/4 & MERVL-gag low transitional state returns to what can be seen in the DMSO-treated lines, the 2CLC population remains inflated 48 hours after the removal of RA from the media (Figure 3.26). Another group reported some 2CLC induction remaining after three passages, although the spliceosome inhibitor Pladienolide B was used instead of RA, this demonstrates that our observation is not outside of the expectations (Shen et al,

2021). A sustained higher population of 2CLCs in these XY lines was not the intent of RA treatment, but they were within a normal physiological range as seen in the XX lines in Figure 3.21, and because the OCT3/4 low population is not also maintained we did not consider this cause to worry about inducing differentiation. Overall, the cells not in the 2CL state are returning to pluripotency despite the 2CLC population remaining enriched.

Rapid, spontaneous differentiation was observed in only the RA-treated XY lines in the latter half of the aging experiment. Following the sustained 2CLC induction noticed in Figure 3.26, we kept a close eye on the morphology of the XY lines, which maintained normal colony formation for the five additional passages equaling 10 days (first and third

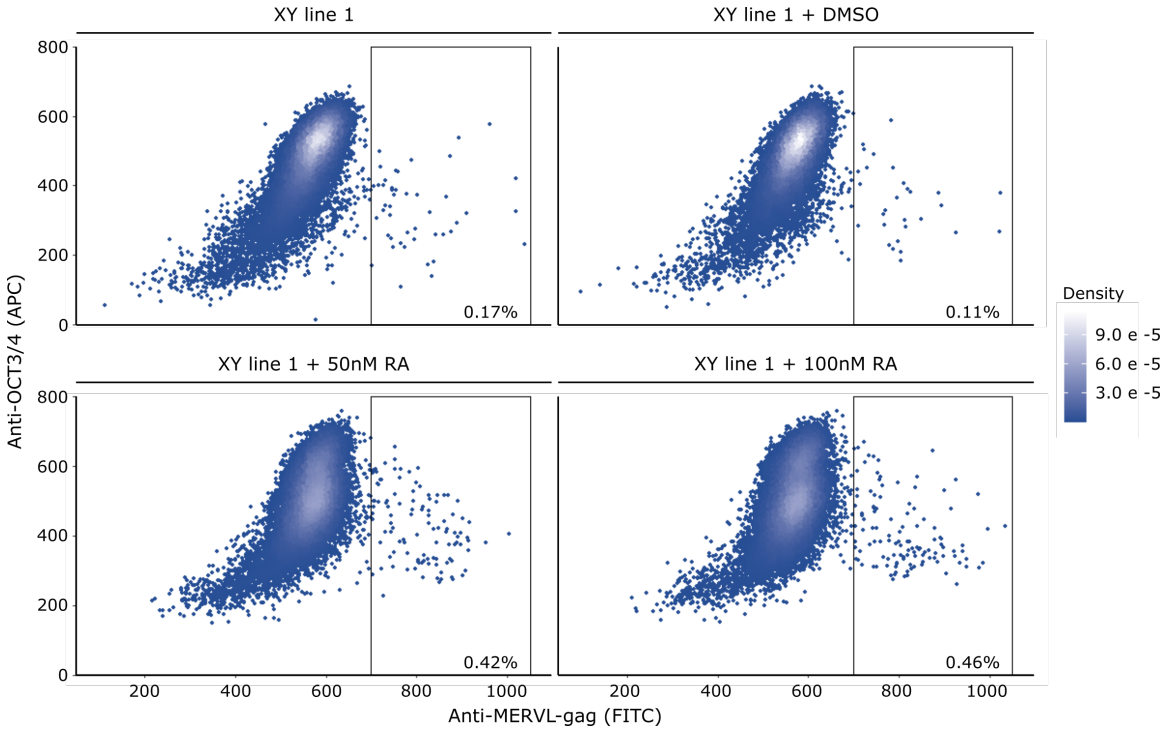


Figure 3.25 - Retinoic acid (RA) induction of 2CL state in XY mESCs.
Flow cytometry for the 2CLC marker MERVL-gag and the pluripotency-associated genesOCT3/4 in XY cell line 1 cultured in serum + LIF media or serum + LIF media supplemented with DMSO, 50nM RA, or 100nM RA for 24 hours (top-left to bottom-right). Black boxes are drawn around the gates used to identify 2CLCs in each cell line.

rows of Figure 3.27). During the fourth RA-treatment in P+6 no visible differentiation was observed. The next day after passaging into P+7 there were many extensive patches of cells with cobblestone-like morphology surrounding the mESC colonies in all the RA-treated XY lines, but not the DMSO-treated XY lines, occurring overnight (second and fourth rows of Figure 3.27). Because of the sudden onset of this event, we were unable to analyze the changes in the differentiating cells to determine their developmental trajectory. These cultures no longer represented an aging mESC culture as the experiment was designed, so we unfortunately had to continue forward without testing

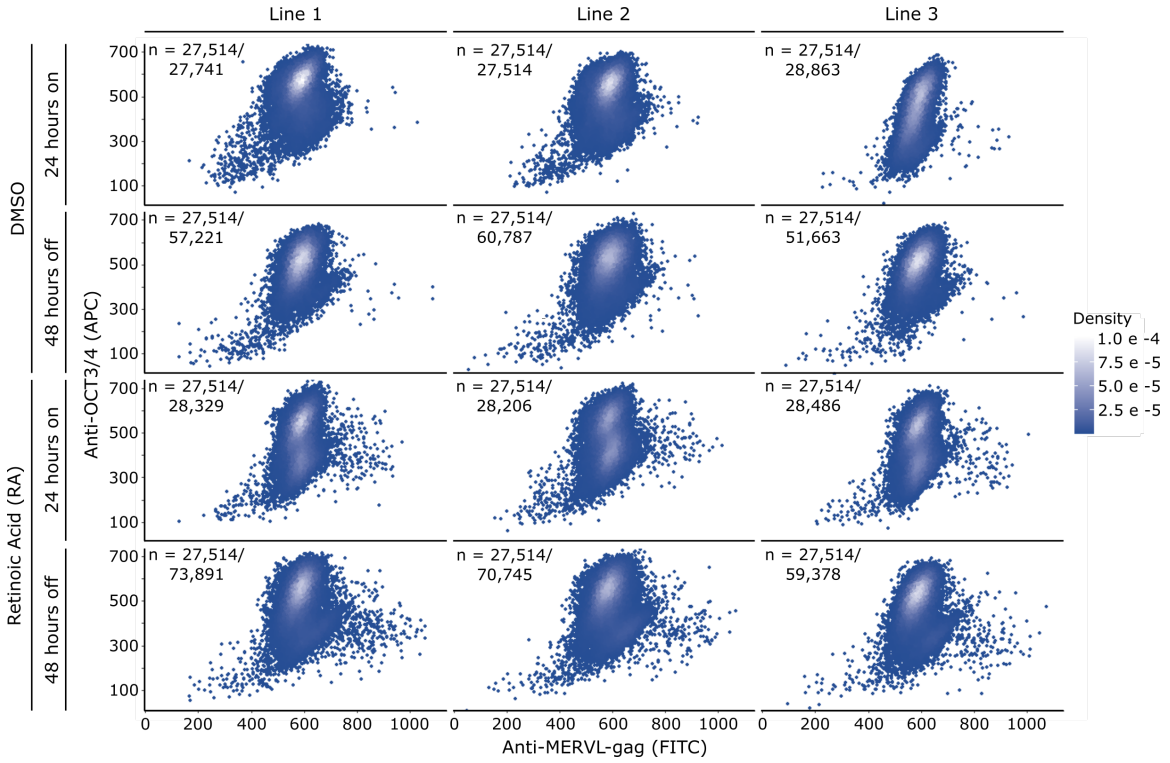


Figure 3.26 - 24-hour treatment with 100nM retinoic acid (RA) promotes lasting 2CLC increase.
Flow cytometry for the 2CLC marker MERVL-gag and the pluripotency-associated genesOCT3/4 in the three XY cell lines during the mESC aging workflow described in Figure 3.26. Upper two rows represent control DMSO-treated XY lines and lower two rows represent RA-treated XY lines. The first and third rows represent XY lines immediately following 24-hour treatment with DMSO or RA. The second and fourth rows represent XY lines 48 hours changing back to non-treatment media. Due to differences in number of cells counted, each plot was down-sampled for better visual comparison (plotted cell count and total sample count noted in the top-left corner of each plot).

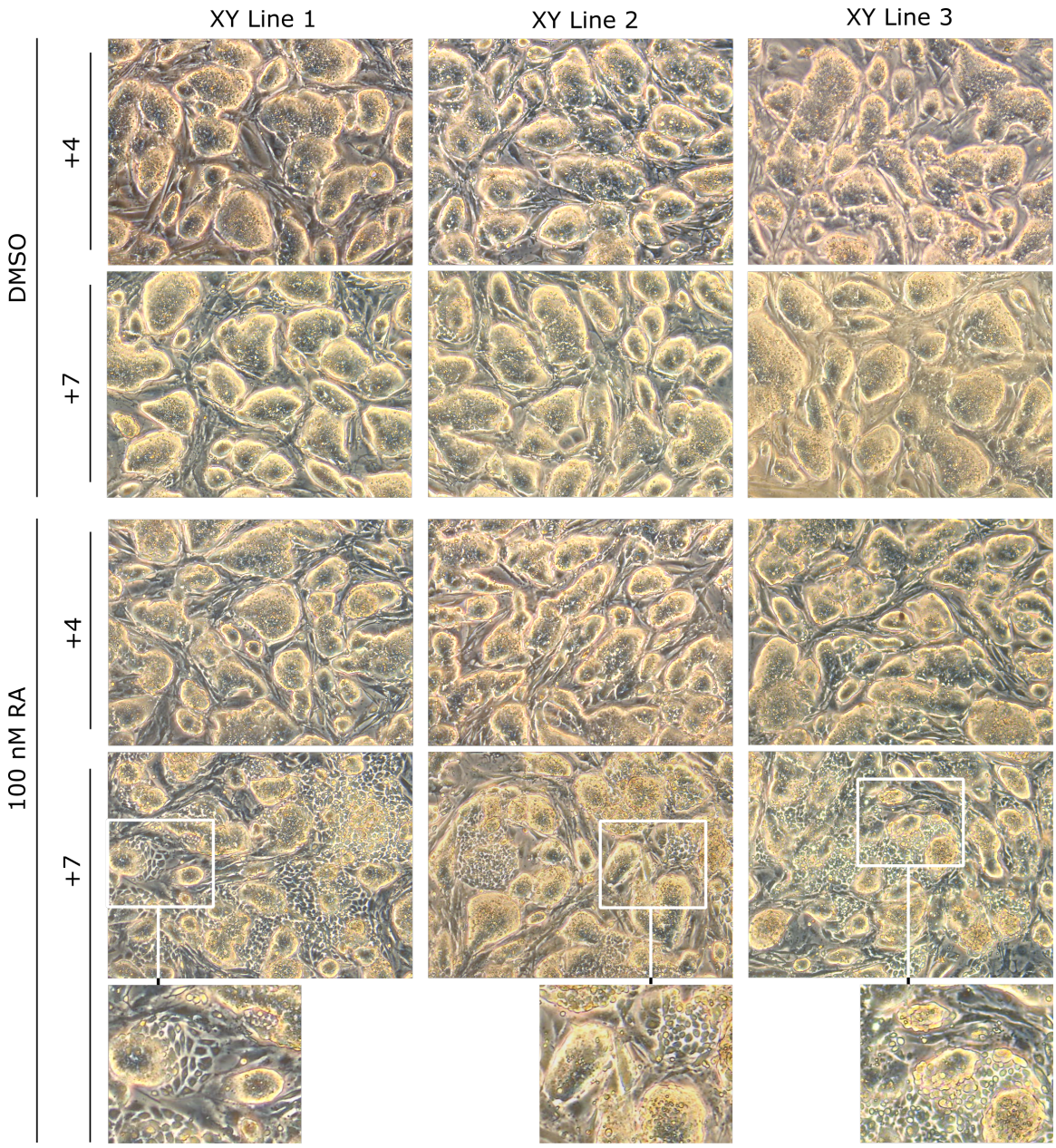


Figure 3.27 - Sudden spontaneous differentiation occurs in XY mESCs following intermittent retinoic acid treatment.
Images of the three independently derived XY C3H/HeSn-Paf/J mESC lines treated with intermittent DMSO or 100nM retinoic acid (RA) treatment. Smaller images in the bottom row are enlarged sections of the images of XY + RA treatment seven passages after first scRNA-seq collection, chosen to show contrast between the morphology of mESC colony, feeder cell layer, and the differentiated cell morphologies.

for the sufficiency of 2CLC induction to rescue aneuploidy accumulation.

3.3.3. Identification of pluripotent cell clusters and demultiplexing chromosomal sex by gene expression

After filtering both the early and aged scRNA-seq data sets of low-quality cells and genes expressed in too few cells to be informative, our early and aged timepoints contained 19,522 and 26,003 cells with expression data for 25,573 genes and 25,452 genes respectively. Unsupervised clustering identified nine clusters in the early mESC cultures and six clusters in the aged mESC cultures. Unfortunately, the detection of LMO barcodes was too low to result in sufficient demultiplexing, so the chromosomal sex of cells in these data sets were genetically demultiplexed by identifying non-ESC clusters, then classifying XX, XY, and XO clusters using known distinct XX- and XY-biased gene expression signatures.

We were able to identify several clusters that were not representative of pluripotent ESC populations. In both early and aged timepoints, there were contaminating MEFs and 2CLCs. Early cluster 6 and aged cluster 3 were identifiable as containing a small number of contaminating MEF cells based on them being mostly in G1 and their expression of *Thy1*, which distinguish fibroblasts and multipotent cells from ESC populations (Brambrink et al, 2008; Polo et al, 2012; Stadtfeld et al, 2008) (Figure 3.28 B-C). Early cluster 7 and aged cluster 5 were identifiable as cells transitioning to or already in the 2CL state due to unique expression of *Zscan4c* and *Zscan4d*, the most expressed of the *Zscan4* genes (Figure 3.28 C). In the early data set, there was also a population represented by early cluster 5 (which may match some cells included in aged cluster 2) that appear to be going through differentiation based on much higher expression of genes such as *T* than expected in the primed pluripotent state (Figure 3.28 C). Because these clusters are contaminating cells, not identifiable as any chromosomal

sex, and/or not pluripotent ESCs, they were not virtually karyotyped to compare aneuploid accumulation between chromosomal sexes.

In each timepoint, only three clusters appear distinguishable by chromosomal sex with corresponding Chromosome X ratios and expression of genes indicative of their respective places along the pluripotency spectrum. The clusters distinctly representing the XX, XY, and XO pluripotent mESCs are clusters 0-2 from each timepoint. In the early

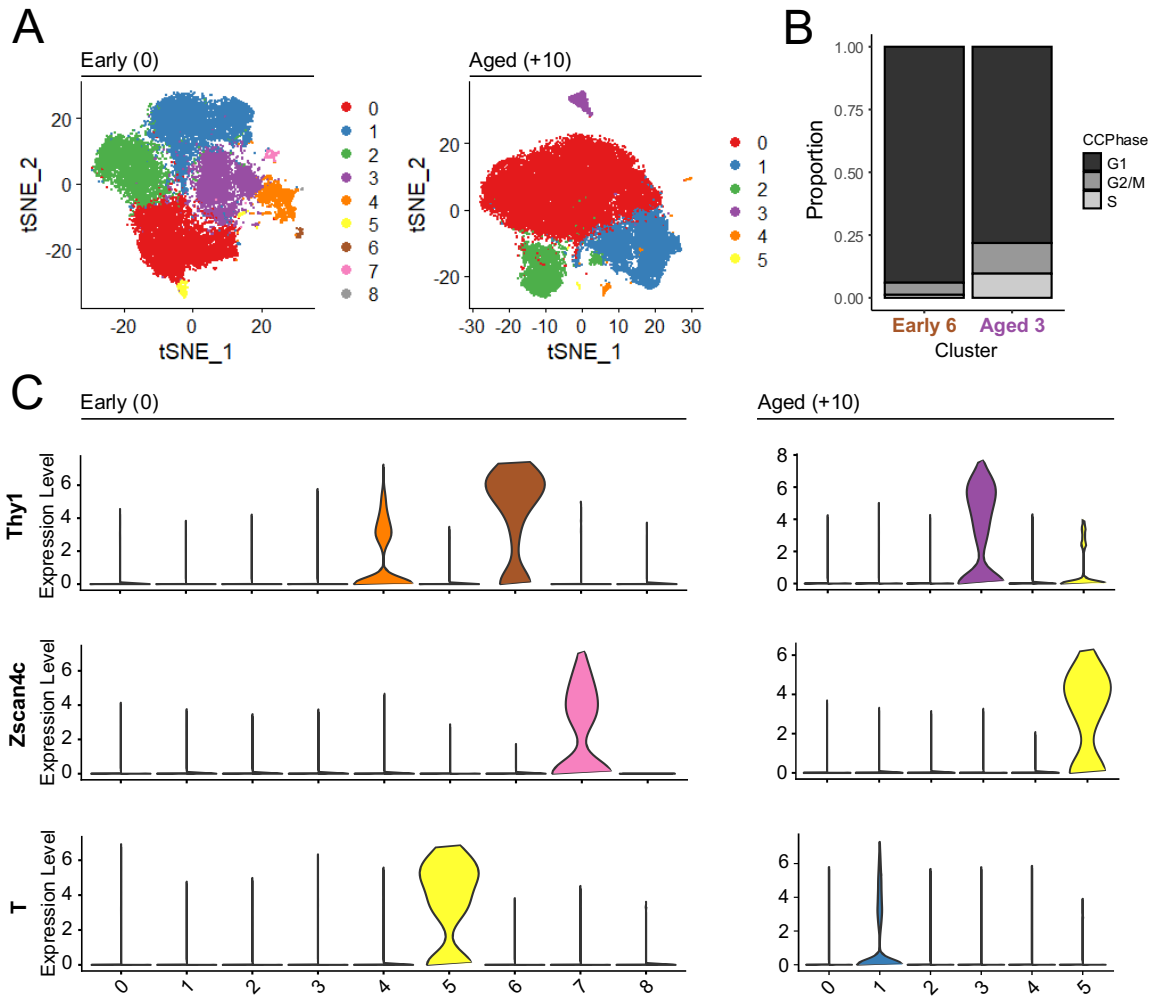


Figure 3.28 – Identification of non-pluripotent scRNA-seq cell clusters.
(A) Unsupervised clustering of early and aged mESC lines based on first 20 principal components. (B) Cell cycle phase proportions of MEF clusters in each timepoint. (C) Cluster specific gene expression for Thy1 (MEF-associated gene), Zscan4c (2CLC marker), and T (lineage specification gene).

data set, the median Chromosome X read percentage in clusters 0 and 2 are nearly half of that from cluster 1 (3.04% and 2.97% compared to 6.20%), and the same can be said

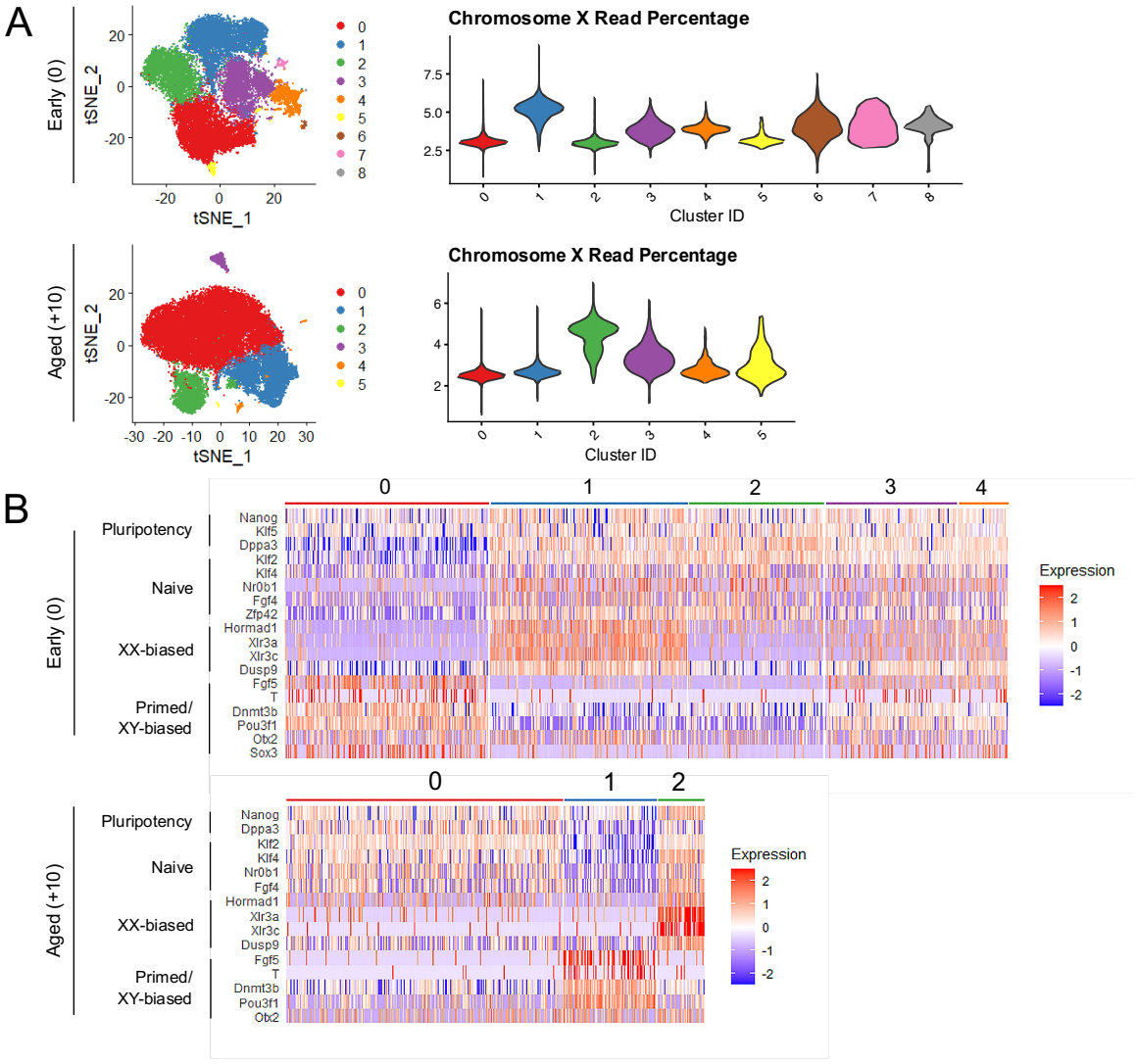


Figure 3.29 - Identification of XX and XY pluripotent mESC cells using unsupervised clustering.
Demultiplexing chromosomal sex from early and aged mESc cultures using XX and XY-biased features. (A) Unsupervised clustering of early and aged mESC lines based on first 20 principal components and cluster specific percentage of reads mapping back to chromosome X. (B) Heatmap of scaled expression of pluripotency-associated genes(Nanog, Klf5, Dppa3), naïve pluripotency-associated genes(Klf2, Klf4, Nr0b1, Fgf4, Zfp42), XX-biased genes (Hormad1, Xlr3a, Xlr3c, Dusp9), and primed pluripotency-associated genes/XY-biased genes (Fgf5, T, Dnmt3b, Pou3f1, Otx2). Data for Klf5 and Zfp42 were not top variable genes in aged mESCs, so were not included in the scaled data.

of clusters 0 and 1 relative to cluster 2 in the aged data set (2.50% and 2.66% compared to 4.58%), reflective of one versus two active X Chromosomes in these cell lines (Figure 3.29 A). The clusters with high Chromosome X read percentages also showed relatively higher expression of XX-biased genes and genes associated with naive pluripotency (Figure 3.29 B). Early cluster 0 and aged cluster 1 were distinguishable as XY by their lack of XX-biased genes and relatively higher expression of genes associated with primed pluripotency (Figure 3.29 B). This leaves early cluster 2 and aged cluster 0 as XO mESCs due to their lack of distinct expression of XX- or XY-biased genes while expressing intermediate levels of pluripotency associated genes - however it should be noted that we could not find documentation of any distinct XO-biased genes in the literature (Figure 3.29 B). The disproportionate size of the aged XO cluster likely indicates that it is composed partially of XX and XY cells which have become XO through loss of their second Chromosome X or Chromosome Y respectively.

Early clusters 3 and 4 were unable to be distinctly identified. Some of the larger clusters show unclear expression of naïve or primed pluripotency genes making it difficult to discern what their cell state is (Figure 3.29 C). Additionally, it is not clear from the percent of Chromosome X reads or comparable expression of both XX and XY biased gene expression whether cells in these clusters belong to XX, XY, or XO lines (Figure 3.29 A). It is possible that these clusters are made up of cells from different cell lines as they lose distinction with the onset of XCI and lineage specification. Because of the ambiguity of these clusters, only the clusters that were identifiable as a certain chromosomal sex were used to compare Chromosome X dosage effects on aneuploidy. In the early data set, this means we have 5,200 XX cells (26.6% of all early cells), 5,375 XY cells (27.5%), and 3,564 XO cells (18.3%), and then in the aged data set we have 2,832 XX cells (10.9%), 5,511 XY cells (21.2%), and 16,600 XO cells (63.8%).

3.3.4. Aneuploidies accumulate at a slower rate in XX C3H/HeSn-Paf/J PSCs

To test whether Chromosome X dosage protects against the development and expansion of aneuploidies in pluripotent cells, we performed virtual karyotyping on each of the identified early and aged pluripotent clusters (Figure 3.30). Each cluster in each data set was analyzed separately to reduce any transcriptional differences between lines from chromosomal sex or factors besides autosome duplication. Although 86.6% of identifiable early cells were annotated as entirely euploid, we detected a small number of cells with preexisting aneuploidies primarily in Chromosomes 8 and 11 as we would expect given the aneuploidy annotations in the DO data sets (Figure 3.6 and Figure 3.8). In the aged populations only 19.7% of cells were identified as euploid after an expansion of the aneuploid subpopulations.

There were two unexpected findings in the aged mESCs: a selection for Chromosome 1 duplication instead of Chromosome 11 and the detection of chromosome duplications not seen in the early mESCs or DO data sets. While

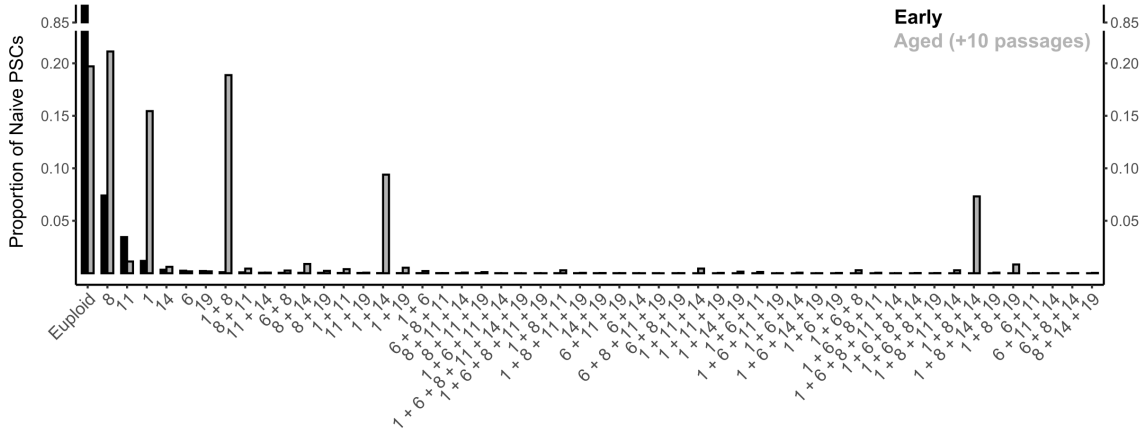


Figure 3.30 - Changing karyotype landscape between early and aged mESCs. Proportion of naïve pluripotent cells classified as euploid or contain combinations of CNVs irrespective of chromosomal sex. Black bars represent cells from the early and gray bars represent cells from the aged scRNA-seq collection.

Chromosome 8 continued to be the most dominant autosomal aneuploidy – with duplication of Chromosome 8 in 51.31% of aged cells – surprisingly Chromosome 1 duplications were the second largest aneuploid population. Additionally, most of the new autosomal duplications of Chromosomes 14 and 19 also contained duplication of Chromosome 1 and or 8. While Chromosome 14 was the third most abundant aneuploidy in the aged data set, 93.57% of those cells with trisomy 14 were trisomy 1+14 or trisomy 1+8+14. Likewise, the most substantial category with Chromosome 19 duplications were trisomy 1+8+19 which made up 37.76% of trisomy 19 cells. This aging study therefore shows that these duplications identified in small numbers during previously published studies may be more likely to accumulate at later timepoints as passenger aneuploidies in cells with previously existing chromosomal imbalances.

As we anticipated, the XX cell lines did not accumulate aneuploidies as extensively as the XY or XO cell lines during the course of ten passages. By the time of the first, early scRNA-seq collection there were already fewer XX cells with autosomal duplications (4.88% compared to 12.47% for XY and 27.16% in XO; Figure 3.31 A). By the time of the second, aged timepoint the aneuploid subpopulation in the XX cell lines increased by 30.46% while the XY and XO increased by 65.87% and 61.45% respectively (Figure 3.31 A). These cell lines had some low-level of aneuploidy from the time of our first collection and the aneuploidies present show evidence of stratified aneuploidy development that only widens over the course of their time in culture. While the Chromosome X dosage distinction is clearest when looking at overall aneuploid proportions, these patterns are not as clear when looking at individual aneuploidies in these cultures (Figure 3.31 B). The acquisition of specific chromosomal duplications is likely a stochastic process that has a selective pressure only after the duplication, in which duplications that are beneficial for survival or proliferation are retained and

expanded within the population. Under this model, there could be variability in the proportions of individual duplications due to random chance, but the overall incidence of duplications happening and their relative advantage would be reflected in the final aneuploid proportions and the change in aneuploid proportions over time.

3.3.5. Aged XX mESCs are enriched for naïve pluripotency-associated genes

To better understand the protective effect displayed by XX mESCs, we performed differential expression analysis to identify genes that are enriched or lost as the XX cells age. For this comparison, only cells annotated as entirely euploid were used to compare aged XX against early XX mESC populations. This DEA found 262 upregulated and 353

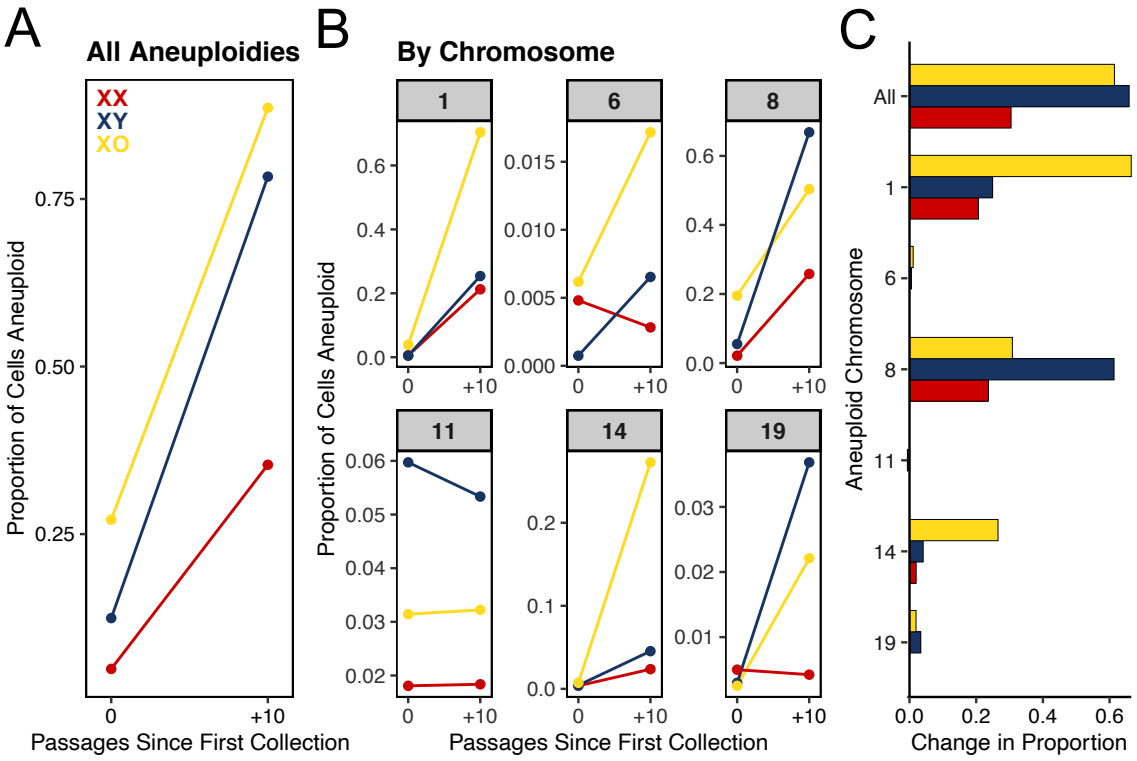


Figure 3.31 - Less accumulation of autosomal aneuploidies occurs in XX mESCs over time in culture.
(A-B) Overall and per chromosome aneuploid proportions at early and aged scRNA-seq collection for each chromosomal sex. (C) Aneuploidy proportion difference between early and aged timepoints. Red lines and bars represent XX cell lines, yellow lines and bars represent XO cell lines, and blue lines and bars represent XY cell lines.

downregulated genes (average log2 fold-change ≥ |0.5| and p ≤ 0.01; Figure 3.32 A). Within the upregulated genes were genes associated with the general and naïve pluripotent state such as *Klf4*, *Klf5*, *Tet2*, *Tbx3*, *Fgf4*, *Nr0b1*, *Esrrb*, *Nanog*, and *Nr5a2*. In the genes with lower aged expression are those associated with primed pluripotency such as *Prdm1*, *Lefty1*, *Pou3f1*, *Dusp6*, and *Fgf5* as well as the TSG *Rbbp7*. While *Xist* is also downregulated in aged XX mESCs, potentially suggesting a selective preference for cells maintaining both active chromosomes, the expression of Chromosome X over time does decrease and this may just be reflective of that change (Figure 3.32 B). Overall, this suggests that cells with more naïve-like character are the ones likely to maintain their genomic integrity over long periods of culture.

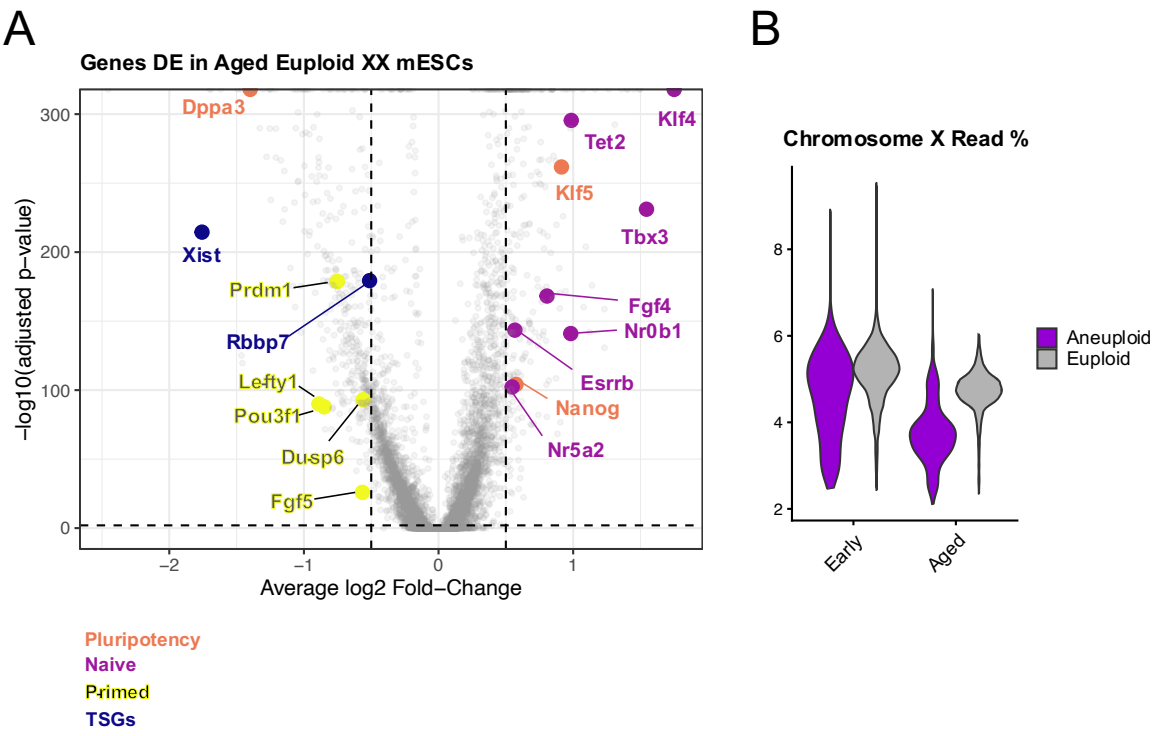


Figure 3.32 – Differentially expressed pluripotency and tumor suppressor genes. (A) Volcano plot for differential expression of genes in aged euploid XX mESCs compared to early euploid XX mESCs. (B) Chromosome X read percentage of XX mESCs split between different timepoints and karyotype.

3.4. Author contributions

All experiments and analyses were performed by Alexander Stanton with the following contributions. DO mESC bulk RNA-seq data and mass spectrometry proteomics data was provided by Steven Munger of The Jackson Laboratory. These data sets were collected by Precision Biology Inc, the proteomics were run and preprocessed by the lab of Steve Gygi of Harvard Medical School, and the RNA-seq was preprocessed by the lab of Gary Churchill at The Jackson Laboratory. DO scRNA-seq was collected, run, aligned, and provided by Precision Biology Inc. Human ESC and iPSC karyotyping was performed by The WiCell Institute and tabulated data was provided by Dylan Stavish and Ivana Barbaric of Sheffield University. Inbred mESC lines were derived and provided by Anne Czechanski and Laura Reinholdt of The Jackson Laboratory. Inbred scRNA-seq libraries were generated by Sandy Daigle, Greg Perry of The Jackson Laboratory and those libraries were sequenced, pre-filtered, and aligned by Bill Flynn of The Jackson Laboratory.

Chapter 4: Discussion

4.1. Proteomics provides additional modalities for virtual karyotyping approaches

As a facet of this research project, we have established an example of how to use proteomics data to support the karyotyping of cell populations. To the best of our knowledge, we are first to use proteomics data to karyotype cell lines. Other labs have used known copy number variations in entire chromosomes or chromosomal regions to better understand the resulting consequences on the proteome relative to the consequences on the transcriptome (Donnelly et al, 2014; Gonçalves et al, 2017; Stingele et al, 2021). While those studies have established that the proteins for most genes will reflect dosage differences similarly to transcripts, as far as we know we are the first to incorporate this evidence into a virtual karyotyping pipeline. One of the primary benefits of virtual karyotyping to evaluate the genomic integrity of a sample is that it utilizes data that has already been collected for other purposes (as is the case for the DO mESC lines described here) or can be used for other purposes later. As the stem cell research community urges reporting on karyotypes alongside published research, it is important to consider how to make these guidelines as accessible as possible. While collecting data on the proteome is ultimately not the most cost-effective option, it does present an option other than transcriptomics where the costs are not solely devoted to quality control. Additionally, there are occasions where clinical samples are analyzed by proteomics, such as for biomarker identification in cancer, where integration of karyotype could yield more nuanced understanding of what mutations the discovered biomarkers actually represent.

Regarding the usage of proteomics data in such pipelines, it should be noted that entirely proteomics-based assessments of CNVs will likely be less sensitive than ones directly assessing DNA or even RNA. Transcriptome-based virtual karyotyping is already

subject to transcriptional and post-transcriptional regulation, but the proteome is subject to translational and post-translational regulation as well and evidence has shown that most of the gene dosage attenuation in eukaryotes occurs there (Chick et al, 2016; Dephoure et al, 2014; Donnelly et al, 2014; Gonçalves et al, 2017; Stingele et al, 2012; Torres et al, 2007). While substantial buffering may be limited to a subset of proteins – around 30% by some reports – this may hamper capability to detect smaller segmental variations, particularly if those regions contain a high number of genes for affected categories such as complex-forming proteins (Chick et al, 2016; Gonçalves et al, 2017; Ryan et al, 2017). Conversely, the increase in X-linked gene production due to erosion of XCI in hPSCs over time results in a compensatory increase in autosomal proteins, but not transcripts (Brenes et al, 2021). If one were surveying for Chromosome X activity as well as copy number, this may skew data when observing X:autosome gene expression ratios as some researchers have chosen to do as a rough proxy for karyotype.

4.2. Prevalence and timing of aneuploidies developing in culture

Using both a custom multi-omic analysis pipeline and the preexisting scRNA-seq karyotyping R/CONICSmat package, we have provided a picture of the aneuploid landscape in mESCs with a diverse genetic context. The exact abnormalities reported in other research often centered around duplications of Chromosomes 8 and 11 as well as loss of Chromosome X and Y in XX and XY mPSCs respectively, with other abnormalities being reported less regularly (Gatzelumendi and Nogues 2014; Sugawara et al, 2006; Guo et al, 2018; Codner et al, 2016; Rebuzzini et al, 2008; Kim et al, 2013). We find that Chromosomes 8, 11, and X in our large panel of genetically diverse DO mESCs are the most frequent and the reason for the inconsistent reporting of other abnormalities is due to their lower prevalence (Figure 3.6, Figure 3.8, Figure 3.30, and Figure 3.31). Curiously, we also found potential evidence for a temporal aspect to

aneuploidy prevalence. Later we report a failure of Chromosome 11 aneuploidies to expand over time, whereas new aneuploidies arise (Figure 3.29). The new Chromosome 14 and 19 aneuploidies identified have been reported in low numbers before, and we propose that this may be due to their dependence upon other chromosomal aberrations. If these aberrations alone don't provide a selective advantage to cells, they may provide additional growth or survival advantages to cells already harboring aneuploidies.

4.3. Identifying genetic determinants of genomic instability

4.3.1. No determinants of specific chromosomal duplications

The primary purpose of this research was to address the variability in the incidence of PSC aneuploidies. One conceivable mechanism was that genetic variation on the duplicated chromosome caused disproportionate missegregation during mitosis, resulting in it being preferentially added to daughter cells. ASE-based methods could not identify any distinct variants on duplicated chromosomes and our genetic mapping of individual chromosome duplications corroborate this by not showing any loci on the duplicated chromosomes (Figure 3.14 and Figure 3.15). However, as we developed the ASE-based methodology, we were unable to find existing data sets to use as a positive control and did not have the additional time to generate one. So, while we were unable to identify any alleles/loci on the duplicated haplotypes, we would need a control that allows us to validate whether we can positively identify known overrepresented regions on duplicated chromosomes. Ideally, this positive control data set would have true haplotypes known and allele-specific RNA-seq data for samples containing a mutation which causes missegregation on at least two genetic backgrounds. Additionally, if any cis-acting elements were responsible for the duplications we see, then we might expect a progressive expansion of those chromosomes' copy number up to the point that it became a deleterious transcriptional and translational burden (Zhu et al, 2018). On top

of not finding any cis-acting genetic elements that would cause duplications of specific chromosomes, we also found no evidence of trans-acting elements that duplicate any of the commonly occurring chromosomes preferentially. Therefore, it appears that the events driving aneuploidy are chromosome agnostic in the genetic context of the eight DO founders, but instead is driven by selective advantages conferred once a chromosomal copy number variation has already occurred. The consistent reporting of increased proliferation, diminished susceptibility to apoptosis, and the similar genes/pathways augmented between species all point to this explanation as well (Ben-David et al, 2014; Werbowetski-Oglivie et al, 2009; Zhang et al, 2016; Figure 3.11 and Figure 3.12).

4.3.2. A potentially reciprocal relationship between Chromosome X dosage and autosomal CNVs

Although we did not find any genetic variation affecting duplication of individual chromosomes, our major finding is that dosage of Chromosome X modulates susceptibility to aneuploidy broadly. Our genetic mapping analysis identified a wide locus on Chromosome X that led to the discovery of a chromosomal sex bias in aneuploid cell line proportions – where XX lines had a lower fraction of aneuploidies compared to XY lines, with XX lines that became XO displaying proportions more similar to the XY lines (Figure 3.15). Later we were able to verify these findings in isogenic cell lines with different sex chromosome combinations and were even able to validate this in the human context with thousands of traditionally karyotyped hiPSCs from a variety of stem cell labs (Figure 3.18 and Figure 3.31). It is still unresolved why this prediction panned out in hiPSCs and not hESCs when the published research would suggest that hiPSC lines are likely downregulating Chromosome X genes to a larger extent than hESCs (Bar et al, 2019). We do not have expression data from these cell lines and the XCI status of

human cell lines are incredibly variable (Anguera et al, 2012; Bar et al, 2019; Barakat et al, 2015; Hall et al, 2008; Hoffman et al, 2005; Kim et al, 2014; Lengner et al, 2010; Mekhoubad et al, 2012; Nazor et al, 2012; Pomp et al, 2011; Shen et al, 2008; Silva et al, 2008; Tchieu et al, 2010; Tomoda et al, 2012; Vallot et al, 2015; Xie et al, 2016), so we cannot make that claim for certain and further examination of this dual X protective effect in hPSCs must be conducted to fully understand this nuance. To resolve this, ideally scRNA-seq data would be collected on XX human ESCs or iPSCs so that their virtual karyotypes could be matched to their Chromosome X activation status (Xa, Xi, Xe).

In addition to the aforementioned evidence showing improved genomic stability with two active X Chromosomes, we have also shown evidence that this relationship may have a reciprocal nature. Our analysis showed reduced X-linked gene expression in aneuploid cells that began with the same single dose of Chromosome X. There is an interpretation of this that would presume these XY cells express these genes in a spectrum of intensities and the lower end is more susceptible to develop copy number variations – akin to a more subtle variation of the XX and XY model. However, the literature around copy number variations in cancer would suggest that there may be more complicated feedback between sex chromosome and autosome dosage. To resolve this, XX and XY mESC lines with autosomal aneuploidies could be generated using microcell-mediated chromosome transfer, then their Chromosome X gene expression could be compared to the euploid parent cell lines.

Even without changes in chromosome copy number in hPSCs, we have referenced research that shows increases in the abundance of autosomal proteins in response to increased X-linked gene expression during XIE (Brenes et al, 2021). Investigation into the X:autosome gene expression ratios in publicly available data sets led to the

discovery that many tumor types have lower than expected Chromosome X expression (Zhang et al, 2018). And even in a non-mammalian model, experimentally induced CNVs in Drosophila stem cells resulted in decreased expression of X-linked genes (Clemente-Ruiz et al, 2016). These studies in conjunction with our data presented here make the case that changes in genome dosage between Chromosome X and autosomes is bidirectional – that low X-linked gene expression increases susceptibility to autosomal CNVs and autosomal CNVs in turn can downregulate X-linked genes.

4.4. Intersections of chromosomal sex with cell culture and heterogeneity

4.4.1. The naive culture conditions are not the same as the XX naïve state

One of the best described differences between XX and XY mPSCs is what state of pluripotency the cells occupy. While the 2i condition mimics the endogenous state of the XX lines, there are key differences between the artificial and natural naïve state. The GSK3 and MEK1 signaling pathways in XX mPSCs are intrinsically diminished compared to XY and XO mPSCs and the 2i inhibitors are able to achieve a similar transcriptome to the preimplantation epiblast that the XX cell lines achieve with LIF alone (reviewed in Schulz 2017). One intended consequence of the 2i condition is that cell cultures become more homogenous. The loss of the heterogeneity described in Chapter 1 also coincides with an unintended consequence: decreased genomic integrity (Kolodziejczyk et al, 2015). Here we have shown that part of the heterogeneity of XX mPSCs is an increased proportion of 2CLCs as well as increased naïve pluripotency-associated gene expression in cells maintaining karyotype over time, while other studies have shown that transcription factors associated with the naïve state, like *Dppa3*, are also associated with the 2CLC populations (Figure 3.20, Figure 3.21, and Figure 3.32; Zhang et al, 2022). One publication showed more 2CLCs in 2i compared to a2i and serum/LIF, but the opposite effect has been observed in multiple other studies since

(Ghimire et al, 2018; Guo et al, 2018). Since the 2CL state has been implicated in multiple studies, it is our assertion that the loss of 2CLCs in 2i and a2i naïve cells plays a role in their development of CNVs. While we do not have much information about totipotent cell states in naïve human media, the decreased genomic stability in these media leads us to believe that the same mechanisms may be at play here.

4.4.2. Prospective improvements to naïve culture media

Given the connections between the naïve pluripotent state, 2CLCs, and genomic stability, I think improvements on naïve culture media will hinge on improving dynamic cell signaling. Research on improving the 2i condition has largely focused on the inhibition of MEK signaling, with one group showing that prolonged suppression in mESCs results in loss of genomic stability and another showing in response that in the human context that naïve media with reduced MEK inhibition are better for genomic stability (Choi et al, 2017; Di Stefano et al, 2018). It seems that the typical strength of inhibition locks cells into a rigid homogenous naïve state that is less facilitative of allowing cells to transition through metastable states like the 2CLCs. Finding a mechanism by which the naïve pluripotency networks can be fostered while leaving room for natural dynamic regulation to exist might be better than inducing a tightly regulated population of cells.

Methods directly fostering a subpopulation of 2CLCs could also ensure their positive impact on genomic stability. We know *Zscan4* improves genome maintenance through telomere elongation (Guo et al, 2018). Other papers have suggested more lofty, unsubstantiated claims that it may enact supernumerary chromosome removal – showing this through overexpression of Zscan4 in trisomic cells *in vitro* or by pointing out that embryos lose aneuploid cells during a similar time of development as totipotency (Amano et al, 2015; Bazrgar et al, 2013). However, the latter correlations may not prove

to be related to *Zscan4* as recent findings suggest the loss of aneuploid cells in the developing embryo is due to differential cell death (Yang et al, 2021). Regardless, the expression of *Zscan4* has been shown to be associated with more stable genomes (Amano et al, 2013; Guo et al, 2018; Jiang et al, 2013; Zalzman et al, 2010). There have been many methods to enrich this population substantially for research purposes, but none for sustained smaller impacts. We have attempted to achieve this through periodic low dose RA treatment, but were met with sudden differentiation (Figure 3.24, Figure 3.25, Figure 3.26, and Figure 3.27). This negative outcome happened in serum/LIF media and it is possible that under 2i conditions the mESCs may have been prevented from exiting pluripotency. It is not clear how efficacious this treatment method would be at inducing 2CLCs under 2i conditions, however.

Given the evidence from cancer research about X-linked TSGs that escape XCI, there may be merit to the continued study of these genes and their relationship to aneuploidy. We found associations between Fhl1 and aneuploidy in the DO bulk RNA-seq and proteomics data, and variable differential expression of a panel of other factors in the DO scRNA-seq data, however only one X-linked TSG was differentially expressed in the aging scRNA-seq and it was downregulated (Figures 3.22, Figure 3.23, and Figure 3.32). These results are mixed as to the importance of any one individual factor in the development of aneuploidy in the context of natural heterogeneity and variation. However, given that aneuploidies arise from mitotic missegregation errors, factors such as *Rbmx* that ensure proper sister chromatid cohesion could be interesting targets and the focus of future research. Promoting genes like *Rbmx* could be a method by which proper checkpoints against missegregation are stimulated without trapping PSCs in a tightly defined homogenous state. However, future work would need to be done on the efficacy of targeted upregulation of these genes in the context of pluripotency.

4.5. Conclusions and Future Directions

In these investigations of PSC aneuploidy, I have provided evidence for a relationship between the gene dosage of Chromosome X and CNVs – not only showing a protective effect of dual Chromosome X dosage against autosomal aneuploidy, but also potentially a feedback response where Chromosome X is downregulated in response to chromosomal duplications. This observation led to discovering a connection between Chromosome X dosage and the 2CL state present in mouse PSC populations and may have implications regarding the newly discovered 8CLC population in human PSCs. In validating these findings, I also showed that the XX mESCs protected against aneuploidy also better preserve the naïve pluripotent state – a state which is most beneficial for the preservation of the 2CLC subpopulation. The implications of this research tie together chromosomal sex, pluripotent state heterogeneity, cell culture conditions, and genomic integrity in ways that will be informative about the best context for these cell types going forward with research and clinical applications. Hopefully work can continue to build on these observations to discover the best minimal factors that can be added to culture media to promote naïve pluripotency regardless of chromosomal sex of the cell line in a way that still supports the heterogenous cell populations that may be responsible for maintaining proper karyotype.

Chapter 5: Bibliography

Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W., & Elledge, S. J. (2012). A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nature Cell Biology*, *14*(3), 318–328. https://doi.org/10.1038/ncb2426

Amano, T., Hirata, T., Falco, G., Monti, M., Sharova, L. V., Amano, M., Sheer, S., Hoang, H. G., Piao, Y., Stagg, C. A., Yamamizu, K., Akiyama, T., & Ko, M. S. H. (2013). Zscan4 restores the developmental potency of embryonic stem cells. *Nature Communications*, *4*, 1966. https://doi.org/10.1038/ncomms2966

Amano, T., Jeffries, E., Amano, M., Ko, A. C., Yu, H., & Ko, M. S. H. (2015). Correction of Down syndrome and Edwards syndrome aneuploidies in human cell cultures. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *22*(5), 331–342. https://doi.org/10.1093/dnares/dsv016

Anguera, M. C., Sadreyev, R., Zhang, Z., Szanto, A., Payer, B., Sheridan, S. D., Kwok, S., Haggarty, S. J., Sur, M., Alvarez, J., Gimelbrant, A., Mitalipova, M., Kirby, J. E., & Lee, J. T. (2012). Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell*, *11*(1), 75–90. https://doi.org/10.1016/j.stem.2012.03.008

Aoki, F., Worrad, D. M., & Schultz, R. M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Developmental Biology*, *181*(2), 296–307. https://doi.org/10.1006/dbio.1996.8466

Augui, S., Nora, E. P., & Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews. Genetics*, *12*(6), 429–442. https://doi.org/10.1038/nrg2987

Avery, S., Hirst, A. J., Baker, D., Lim, C. Y., Alagaratnam, S., Skotheim, R. I., Lothe, R. A., Pera, M. F., Colman, A., Robson, P., Andrews, P. W., & Knowles, B. B. (2013). BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. *Stem cell reports*, *1*(5), 379–386. https://doi.org/10.1016/j.stemcr.2013.10.005

Aydin, S., Pham, D. T., Zhang, T., Keele, G. R., Skelly, D. A., Pankratz, M., Choi, T., Gygi, S. P., Reinholdt, L. G., Baker, C. L., Churchill, G. A., & Munger, S. C. (2022). *Genetic dissection of the pluripotent proteome through multi-omics data integration* [Preprint]. Genetics. https://doi.org/10.1101/2022.04.22.489216

Babariya, D., Fragouli, E., Alfarawati, S., Spath, K., & Wells, D. (2017). The incidence and origin of segmental aneuploidy in human oocytes and preimplantation embryos. *Human Reproduction (Oxford, England)*, *32*(12), 2549–2560. https://doi.org/10.1093/humrep/dex324

Baker, D. E. C., Harrison, N. J., Maltby, E., Smith, K., Moore, H. D., Shaw, P. J., Heath, P. R., Holden, H., & Andrews, P. W. (2007). Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nature Biotechnology*, *25*(2), 207–215. https://doi.org/10.1038/nbt1285

Bansal, P., Ahern, D. T., Kondaveeti, Y., Qiu, C. W., & Pinter, S. F. (2021). Contiguous erosion of the inactive X in human pluripotency concludes with global DNA hypomethylation. *Cell Reports*, *35*(10), 109215. https://doi.org/10.1016/j.celrep.2021.109215

Bar, S., Seaton, L. R., Weissbein, U., Eldar-Geva, T., & Benvenisty, N. (2019). Global Characterization of X Chromosome Inactivation in Human Pluripotent Stem Cells. *Cell Reports*, *27*(1), 20-29.e3. https://doi.org/10.1016/j.celrep.2019.03.019

Barakat, T. S., Ghazvini, M., de Hoon, B., Li, T., Eussen, B., Douben, H., van der Linden, R., van der Stap, N., Boter, M., Laven, J. S., Galjaard, R.-J., Grootegoed, J. A., de Klein, A., & Gribnau, J. (2015). Stable X chromosome reactivation in female human induced pluripotent stem cells. *Stem Cell Reports*, *4*(2), 199–208. https://doi.org/10.1016/j.stemcr.2014.12.012

Bazrgar, M., Gourabi, H., Valojerdi, M. R., Yazdi, P. E., & Baharvand, H. (2013). Self-correction of chromosomal abnormalities in human preimplantation embryos and embryonic stem cells. *Stem Cells and Development*, *22*(17), 2449–2456. https://doi.org/10.1089/scd.2013.0053

Beddington, R. S., & Robertson, E. J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development (Cambridge, England)*, *105*(4), 733–737. https://doi.org/10.1242/dev.105.4.733

Ben-David, U., Arad, G., Weissbein, U., Mandefro, B., Maimon, A., Golan-Lev, T., Narwani, K., Clark, A. T., Andrews, P. W., Benvenisty, N., & Carlos Biancotti, J. (2014). Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. *Nature Communications*, *5*, 4825. https://doi.org/10.1038/ncomms5825

Ben-David, U., Kopper, O., & Benvenisty, N. (2012). Expanding the boundaries of embryonic stem cells. *Cell Stem Cell*, *10*(6), 666–677. https://doi.org/10.1016/j.stem.2012.05.003

Bolton, H., Graham, S. J. L., Van der Aa, N., Kumar, P., Theunis, K., Fernandez Gallardo, E., Voet, T., & Zernicka-Goetz, M. (2016). Mouse model of chromosome mosaicism reveals lineage-specific depletion of aneuploid cells and normal developmental potential. *Nature Communications*, *7*, 11165. https://doi.org/10.1038/ncomms11165

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., & Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell stem cell*, *2*(2), 151–159. https://doi.org/10.1016/j.stem.2008.01.004

Bredenkamp, N., Yang, J., Clarke, J., Stirparo, G. G., von Meyenn, F., Dietmann, S., Baker, D., Drummond, R., Ren, Y., Li, D., Wu, C., Rostovskaya, M., Eminli-Meissner, S., Smith, A., & Guo, G. (2019). Wnt Inhibition Facilitates RNA-Mediated Reprogramming of Human Somatic Cells to Naive Pluripotency. *Stem Cell Reports*, *13*(6), 1083–1098. https://doi.org/10.1016/j.stemcr.2019.10.009

Brenes, A. J., Yoshikawa, H., Bensaddek, D., Mirauta, B., Seaton, D., Hukelmann, J. L., Jiang, H., Stegle, O., & Lamond, A. I. (2021). Erosion of human X chromosome inactivation causes major remodeling of the iPSC proteome. *Cell Reports*, *35*(4), 109032. https://doi.org/10.1016/j.celrep.2021.109032

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś., Yandell, B. S., & Churchill, G. A. (2019). R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics*, *211*(2), 495–502. https://doi.org/10.1534/genetics.118.301595

Bruck, T., Yanuka, O., & Benvenisty, N. (2013). Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-Linked ELK-1 gene. *Cell Reports*, *4*(2), 262–270. https://doi.org/10.1016/j.celrep.2013.06.026

Burgoyne, P. S., Thornhill, A. R., Boudrean, S. K., Darling, S. M., Bishop, C. E., & Evans, E. P. (1995). The genetic basis of XX-XY differences present before gonadal sex differentiation in the mouse. *Philosophical Transactions of the Royal*

Society of London. Series B, Biological Sciences, *350*(1333), 253-260 discussion 260-261. https://doi.org/10.1098/rstb.1995.0159
Burton, A., & Torres-Padilla, M.-E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nature Reviews. Molecular Cell Biology*, *15*(11), 723–734. https://doi.org/10.1038/nrm3885
Capecchi, M. R. (2005). Gene targeting in mice: Functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews. Genetics*, *6*(6), 507–512. https://doi.org/10.1038/nrg1619
Carter, M. G., Stagg, C. A., Falco, G., Yoshikawa, T., Bassey, U. C., Aiba, K., Sharova, L. V., Shaik, N., & Ko, M. S. H. (2008). An in situ hybridization-based screen for heterogeneously expressed genes in mouse ES cells. *Gene Expression Patterns: GEP*, *8*(3), 181–198. https://doi.org/10.1016/j.gep.2007.10.009
Catalina, P., Montes, R., Ligero, G., Sanchez, L., de la Cueva, T., Bueno, C., Leone, P. E., & Menendez, P. (2008). Human ESCs predisposition to karyotypic instability: Is a matter of culture adaptation or differential vulnerability among hESC lines due to inherent properties? *Molecular Cancer*, *7*, 76. https://doi.org/10.1186/1476-4598-7-76
Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., & Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature*, *450*(7173), 1230–1234. https://doi.org/10.1038/nature06403
Chazaud, C., & Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development (Cambridge, England)*, *143*(7), 1063–1074. https://doi.org/10.1242/dev.128314
Chick, J. M., Munger, S. C., Simecek, P., Huttlin, E. L., Choi, K., Gatti, D. M., Raghupathy, N., Svenson, K. L., Churchill, G. A., & Gygi, S. P. (2016). Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, *534*(7608), 500–505. https://doi.org/10.1038/nature18270
Choi, J., Clement, K., Huebner, A. J., Webster, J., Rose, C. M., Brumbaugh, J., Walsh, R. M., Lee, S., Savol, A., Etchegaray, J.-P., Gu, H., Boyle, P., Elling, U., Mostoslavsky, R., Sadreyev, R., Park, P. J., Gygi, S. P., Meissner, A., & Hochedlinger, K. (2017). DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. *Cell Stem Cell*, *20*(5), 706-719.e7. https://doi.org/10.1016/j.stem.2017.03.002
Choi, J., Huebner, A. J., Clement, K., Walsh, R. M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.-Y., Sharif, J., Rose, C. M., Mohammad, A., Odajima, J., Charron, J., Shioda, T., Gnirke, A., Gygi, S., Koseki, H., … Hochedlinger, K. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature*, *548*(7666), 219–223. https://doi.org/10.1038/nature23274
Choi, K., He, H., Gatti, D. M., Philip, V. M., Raghupathy, N., Munger, S. C., Chesler, E. J., & Churchill, G. A. (2020). *Genotype-free individual genome reconstruction of Multiparental Population Models by RNA sequencing data* (p. 2020.10.11.335323). bioRxiv. https://doi.org/10.1101/2020.10.11.335323
Churchill, G. A., Gatti, D. M., Munger, S. C., & Svenson, K. L. (2012). The Diversity Outbred mouse population. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, *23*(9–10), 713–718. https://doi.org/10.1007/s00335-012-9414-2
Clemente-Ruiz, M., Murillo-Maldonado, J. M., Benhra, N., Barrio, L., Pérez, L., Quiroga, G., Nebreda, A. R., & Milán, M. (2016). Gene Dosage Imbalance Contributes to

Chromosomal Instability-Induced Tumorigenesis. *Developmental Cell*, *36*(3), 290–302. https://doi.org/10.1016/j.devcel.2016.01.008
Codner, G. F., Lindner, L., Caulder, A., Wattenhofer-Donzé, M., Radage, A., Mertz, A., Eisenmann, B., Mianné, J., Evans, E. P., Beechey, C. V., Fray, M. D., Birling, M.-C., Hérault, Y., Pavlovic, G., & Teboul, L. (2016). Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction. *BMC Cell Biology*, *17*(1), 30. https://doi.org/10.1186/s12860-016-0108-6
Cowan, C. A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J. P., Wang, S., Morton, C. C., McMahon, A. P., Powers, D., & Melton, D. A. (2004). Derivation of embryonic stem-cell lines from human blastocysts. *The New England Journal of Medicine*, *350*(13), 1353–1356. https://doi.org/10.1056/NEJMsr040330
Cui, Y., Chen, X., Luo, H., Fan, Z., Luo, J., He, S., Yue, H., Zhang, P., & Chen, R. (2016). BioCircos.js: An interactive Circos JavaScript library for biological data visualization on web applications. *Bioinformatics (Oxford, England)*, *32*(11), 1740–1742. https://doi.org/10.1093/bioinformatics/btw041
Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L. R., Hadjantonakis, A.-K., & Reinholdt, L. G. (2014). Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. *Nature Protocols*, *9*(3), 559–574. https://doi.org/10.1038/nprot.2014.030
Dephoure, N., Hwang, S., O'Sullivan, C., Dodgson, S. E., Gygi, S. P., Amon, A., & Torres, E. M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *ELife*, *3*, e03023. https://doi.org/10.7554/eLife.03023
Di Stefano, B., Ueda, M., Sabri, S., Brumbaugh, J., Huebner, A. J., Sahakyan, A., Clement, K., Clowers, K. J., Erickson, A. R., Shioda, K., Gygi, S. P., Gu, H., Shioda, T., Meissner, A., Takashima, Y., Plath, K., & Hochedlinger, K. (2018). Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. *Nature Methods*, *15*(9), 732–740. https://doi.org/10.1038/s41592-018-0104-1
Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L. A., Khan, S. A., Park, K.-M., Yoon, E.-J., Xing, X., Kommagani, R., Wang, T., Solnica-Krezel, L., & Theunissen, T. W. (2020). Derivation of trophoblast stem cells from naïve human pluripotent stem cells. *ELife*, *9*, e52504. https://doi.org/10.7554/eLife.52504
Donnelly, N., & Storchová, Z. (2014). Dynamic karyotype, dynamic proteome: Buffering the effects of aneuploidy. *Biochimica Et Biophysica Acta*, *1843*(2), 473–481. https://doi.org/10.1016/j.bbamcr.2013.11.017
Draper, J. S., Moore, H. D., Ruban, L. N., Gokhale, P. J., & Andrews, P. W. (2004). Culture and characterization of human embryonic stem cells. *Stem Cells and Development*, *13*(4), 325–336. https://doi.org/10.1089/scd.2004.13.325
Dunford, A., Weinstock, D. M., Savova, V., Schumacher, S. E., Cleary, J. P., Yoda, A., Sullivan, T. J., Hess, J. M., Gimelbrant, A. A., Beroukhim, R., Lawrence, M. S., Getz, G., & Lane, A. A. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nature Genetics*, *49*(1), 10–16. https://doi.org/10.1038/ng.3726
Eckersley-Maslin, M. A., Alda-Catalinas, C., & Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nature Reviews. Molecular Cell Biology*, *19*(7), 436–450. https://doi.org/10.1038/s41580-018-0008-z
Endoh, M., & Niwa, H. (2022). Stepwise pluripotency transitions in mouse stem cells. *EMBO Reports*, e55010. https://doi.org/10.15252/embr.202255010

Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, *292*(5819), 154–156. https://doi.org/10.1038/292154a0

Falco, G., Lee, S.-L., Stanghellini, I., Bassey, U. C., Hamatani, T., & Ko, M. S. H. (2007). Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Developmental Biology*, *307*(2), 539–550. https://doi.org/10.1016/j.ydbio.2007.05.003

Fan, H. C., & Quake, S. R. (2007). Detection of aneuploidy with digital polymerase chain reaction. *Analytical Chemistry*, *79*(19), 7576–7579. https://doi.org/10.1021/ac0709394

Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P. S., Lickert, H., Anastassiadis, K., & Schroeder, T. (2013). Biallelic expression of nanog protein in mouse embryonic stem cells. *Cell Stem Cell*, *13*(1), 12–13. https://doi.org/10.1016/j.stem.2013.04.025

Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais, Y., Shipony, Z., Mukamel, Z., Krupalnik, V., Zerbib, M., Geula, S., Caspi, I., Schneir, D., Shwartz, T., … Hanna, J. H. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature*, *504*(7479), 282–286. https://doi.org/10.1038/nature12745

Garitaonandia, I., Amir, H., Boscolo, F. S., Wambua, G. K., Schultheisz, H. L., Sabatini, K., Morey, R., Waltz, S., Wang, Y.-C., Tran, H., Leonardo, T. R., Nazor, K., Slavin, I., Lynch, C., Li, Y., Coleman, R., Gallego Romero, I., Altun, G., Reynolds, D., … Laurent, L. C. (2015). Increased risk of genetic and epigenetic instability in human embryonic stem cells associated with specific culture conditions. *PloS One*, *10*(2), e0118307. https://doi.org/10.1371/journal.pone.0118307

Gaztelumendi, N., & Nogués, C. (2014). Chromosome instability in mouse embryonic stem cells. *Scientific Reports*, *4*, 5324. https://doi.org/10.1038/srep05324

Genolet, O., Monaco, A. A., Dunkel, I., Boettcher, M., & Schulz, E. G. (2021). Identification of X-chromosomal genes that drive sex differences in embryonic stem cells through a hierarchical CRISPR screening approach. *Genome Biology*, *22*(1), 110. https://doi.org/10.1186/s13059-021-02321-2

Ghimire, S., Van der Jeught, M., Neupane, J., Roost, M. S., Anckaert, J., Popovic, M., Van Nieuwerburgh, F., Mestdagh, P., Vandesompele, J., Deforce, D., Menten, B., Chuva de Sousa Lopes, S., De Sutter, P., & Heindryckx, B. (2018). Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Scientific Reports*, *8*(1), 5884. https://doi.org/10.1038/s41598-018-24051-5

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., & Beltrao, P. (2017). Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. *Cell Systems*, *5*(4), 386-398.e4. https://doi.org/10.1016/j.cels.2017.08.013

Gonzales, P. K., Roberts, C. M., Fonte, V., Jacobsen, C., Stein, G. H., & Link, C. D. (2018). Transcriptome analysis of genetically matched human induced pluripotent stem cells disomic or trisomic for chromosome 21. *PloS one*, *13*(3), e0194581. https://doi.org/10.1371/journal.pone.0194581

Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2017). Mosaic autosomal aneuploidies are detectable from single-cell RNAseq data. *BMC Genomics*, *18*(1), 904. https://doi.org/10.1186/s12864-017-4253-x

Guha, P., Morgan, J. W., Mostoslavsky, G., Rodrigues, N. P., & Boyd, A. S. (2017). Lack of Immune Response to Differentiated Cells Derived from Syngeneic Induced

Pluripotent Stem Cells. *Cell Stem Cell*, *21*(1), 144–148. https://doi.org/10.1016/j.stem.2017.03.012
Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., & Nichols, J. (2016). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports*, *6*(4), 437–446. https://doi.org/10.1016/j.stemcr.2016.02.005
Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., … Qiao, J. (2014). The DNA methylation landscape of human early embryos. *Nature*, *511*(7511), 606–610. https://doi.org/10.1038/nature13544
Guo, R., Ye, X., Yang, J., Zhou, Z., Tian, C., Wang, H., Wang, H., Fu, H., Liu, C., Zeng, M., Yang, J., & Liu, L. (2018). Feeders facilitate telomere maintenance and chromosomal stability of embryonic stem cells. *Nature Communications*, *9*(1), 2620. https://doi.org/10.1038/s41467-018-05038-2
Hall, L. L., Byron, M., Butler, J., Becker, K. A., Nelson, A., Amit, M., Itskovitz-Eldor, J., Stein, J., Stein, G., Ware, C., & Lawrence, J. B. (2008). X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. *Journal of Cellular Physiology*, *216*(2), 445–452. https://doi.org/10.1002/jcp.21411
Hatano, S.-Y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., Suemori, H., Nakatsuji, N., & Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mechanisms of Development*, *122*(1), 67–79. https://doi.org/10.1016/j.mod.2004.08.008
Hayashi, K., de Sousa Lopes, S. M. C., Tang, F., Lao, K., & Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*, *3*(4), 391–401. https://doi.org/10.1016/j.stem.2008.07.027
Heard, E. (2004). Recent advances in X-chromosome inactivation. *Current Opinion in Cell Biology*, *16*(3), 247–255. https://doi.org/10.1016/j.ceb.2004.03.005
Hoffman, L. M., Hall, L., Batten, J. L., Young, H., Pardasani, D., Baetge, E. E., Lawrence, J., & Carpenter, M. K. (2005). X-inactivation status varies in human embryonic stem cell lines. *Stem Cells (Dayton, Ohio)*, *23*(10), 1468–1478. https://doi.org/10.1634/stemcells.2004-0371
Imataka, G., & Arisaka, O. (2012). Chromosome analysis using spectral karyotyping (SKY). *Cell Biochemistry and Biophysics*, *62*(1), 13–17. https://doi.org/10.1007/s12013-011-9285-2
Iturbide, A., Ruiz Tejada Segura, M. L., Noll, C., Schorpp, K., Rothenaigner, I., Ruiz-Morales, E. R., Lubatti, G., Agami, A., Hadian, K., Scialdone, A., & Torres-Padilla, M.-E. (2021). Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nature Structural & Molecular Biology*, *28*(6), 521–532. https://doi.org/10.1038/s41594-021-00590-w
Jelinkova, S., Fojtik, P., Kohutova, A., Vilotic, A., Marková, L., Pesl, M., Jurakova, T., Kruta, M., Vrbsky, J., Gaillyova, R., Valášková, I., Frák, I., Lacampagne, A., Forte, G., Dvorak, P., Meli, A. C., & Rotrekl, V. (2019). Dystrophin Deficiency Leads to Genomic Instability in Human Pluripotent Stem Cells via NO Synthase-Induced Oxidative Stress. *Cells*, *8*(1), E53. https://doi.org/10.3390/cells8010053
Jiang, J., Lv, W., Ye, X., Wang, L., Zhang, M., Yang, H., Okuka, M., Zhou, C., Zhang, X., Liu, L., & Li, J. (2013). Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. *Cell Research*, *23*(1), 92–106. https://doi.org/10.1038/cr.2012.157

Kawase, E., Suemori, H., Takahashi, N., Okazaki, K., Hashimoto, K., & Nakatsuji, N. (1994). Strain difference in establishment of mouse embryonic stem (ES) cell lines. *The International Journal of Developmental Biology*, *38*(2), 385–390.

Kim, K.-Y., Hysolli, E., Tanaka, Y., Wang, B., Jung, Y.-W., Pan, X., Weissman, S. M., & Park, I.-H. (2014). X Chromosome of female cells shows dynamic changes in status during human somatic cell reprogramming. *Stem Cell Reports*, *2*(6), 896–909. https://doi.org/10.1016/j.stemcr.2014.04.003

Kim, Y. M., Lee, J.-Y., Xia, L., Mulvihill, J. J., & Li, S. (2013). Trisomy 8: A common finding in mouse embryonic stem (ES) cell lines. *Molecular Cytogenetics*, *6*(1), 3. https://doi.org/10.1186/1755-8166-6-3

Klaric, J. A., Wüst, S., & Panier, S. (2021). New Faces of old Friends: Emerging new Roles of RNA-Binding Proteins in the DNA Double-Strand Break Response. *Frontiers in Molecular Biosciences*, *8*, 668821. https://doi.org/10.3389/fmolb.2021.668821

Koch, J., Lang, A., Whongsiri, P., Schulz, W. A., Hoffmann, M. J., & Greife, A. (2021). KDM6A mutations promote acute cytoplasmic DNA release, DNA damage response and mitosis defects. *BMC Molecular and Cell Biology*, *22*(1), 54. https://doi.org/10.1186/s12860-021-00394-2

Kolishovski, G., Lamoureux, A., Hale, P., Richardson, J. E., Recla, J. M., Adesanya, O., Simons, A., Kunde-Ramamoorthy, G., & Bult, C. J. (2019). The JAX Synteny Browser for mouse-human comparative genomics. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, *30*(11–12), 353–361. https://doi.org/10.1007/s00335-019-09821-4

Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C. H., Ilicic, T., Henriksson, J., Natarajan, K. N., Tuck, A. C., Gao, X., Bühler, M., Liu, P., Marioni, J. C., & Teichmann, S. A. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*, *17*(4), 471–485. https://doi.org/10.1016/j.stem.2015.09.011

Kroon, E., Martinson, L. A., Kadoya, K., Bang, A. G., Kelly, O. G., Eliazer, S., Young, H., Richardson, M., Smart, N. G., Cunningham, J., Agulnick, A. D., D'Amour, K. A., Carpenter, M. K., & Baetge, E. E. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nature Biotechnology*, *26*(4), 443–452. https://doi.org/10.1038/nbt1393

Kuang, X., & Li, J. (2022). Chromosome instability and aneuploidy as context-dependent activators or inhibitors of antitumor immunity. *Frontiers in Immunology*, *13*. https://www.frontiersin.org/articles/10.3389/fimmu.2022.895961

Kumar, R. M., Cahan, P., Shalek, A. K., Satija, R., DaleyKeyser, Aj., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J. J., Ferrante, T. C., Regev, A., Daley, G. Q., & Collins, J. J. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, *516*(7529), 56–61. https://doi.org/10.1038/nature13920

Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A., Boutz, D. R., Huse, H. K., Nozue, K., Walia, H., Whiteley, M., Ronald, P. C., & Marcotte, E. M. (2010). Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics*, *10*(23), 4209–4212. https://doi.org/10.1002/pmic.201000327

Lee, G., Papapetrou, E. P., Kim, H., Chambers, S. M., Tomishima, M. J., Fasano, C. A., Ganat, Y. M., Menon, J., Shimizu, F., Viale, A., Tabar, V., Sadelain, M., & Studer, L. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature*, *461*(7262), 402–406. https://doi.org/10.1038/nature08320

Lefort, N., Feyeux, M., Bas, C., Féraud, O., Bennaceur-Griscelli, A., Tachdjian, G., Peschanski, M., & Perrier, A. L. (2008). Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nature Biotechnology*, *26*(12), 1364–1366. https://doi.org/10.1038/nbt.1509

Lengner, C. J., Gimelbrant, A. A., Erwin, J. A., Cheng, A. W., Guenther, M. G., Welstead, G. G., Alagappan, R., Frampton, G. M., Xu, P., Muffat, J., Santagata, S., Powers, D., Barrett, C. B., Young, R. A., Lee, J. T., Jaenisch, R., & Mitalipova, M. (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell*, *141*(5), 872–883. https://doi.org/10.1016/j.cell.2010.04.010

Lentini, A., Cheng, H., Noble, J. C., Papanicolaou, N., Coucoravas, C., Andrews, N., Deng, Q., Enge, M., & Reinius, B. (2022). Elastic dosage compensation by X-chromosome upregulation. *Nature Communications*, *13*(1), 1854. https://doi.org/10.1038/s41467-022-29414-1

Leung, C. Y., & Zernicka-Goetz, M. (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. *Current Opinion in Genetics & Development*, *34*, 71–76. https://doi.org/10.1016/j.gde.2015.08.002

Lim, C. Y., Knowles, B. B., Solter, D., & Messerschmidt, D. M. (2016). Epigenetic Control of Early Mouse Development. *Current Topics in Developmental Biology*, *120*, 311–360. https://doi.org/10.1016/bs.ctdb.2016.05.002

Linneberg-Agerholm, M., Wong, Y. F., Romero Herrera, J. A., Monteiro, R. S., Anderson, K. G. V., & Brickman, J. M. (2019). Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development (Cambridge, England)*, *146*(24), dev180620. https://doi.org/10.1242/dev.180620

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., & Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, *487*(7405), 57–63. https://doi.org/10.1038/nature11244

Mahla, R. S. (2016). Stem Cells Applications in Regenerative Medicine and Disease Therapeutics. *International Journal of Cell Biology*, *2016*, 6940283. https://doi.org/10.1155/2016/6940283

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, *583*(24), 3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036

Maitra, A., Arking, D. E., Shivapurkar, N., Ikeda, M., Stastny, V., Kassauei, K., Sui, G., Cutler, D. J., Liu, Y., Brimble, S. N., Noaksson, K., Hyllner, J., Schulz, T. C., Zeng, X., Freed, W. J., Crook, J., Abraham, S., Colman, A., Sartipy, P., … Chakravarti, A. (2005). Genomic alterations in cultured human embryonic stem cells. *Nature Genetics*, *37*(10), 1099–1103. https://doi.org/10.1038/ng1631

Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., & Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science (New York, N.Y.)*, *303*(5658), 666–669. https://doi.org/10.1126/science.1092674

Mao, R., Wang, X., Spitznagel, E. L., Jr, Frelin, L. P., Ting, J. C., Ding, H., Kim, J. W., Ruczinski, I., Downey, T. J., & Pevsner, J. (2005). Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. *Genome biology*, *6*(13), R107. https://doi.org/10.1186/gb-2005-6-13-r107

Matsumoto, K., Anzai, M., Nakagata, N., Takahashi, A., Takahashi, Y., & Miyata, K. (1994). Onset of paternal gene activation in early mouse embryos fertilized with

transgenic mouse sperm. *Molecular Reproduction and Development*, *39*(2), 136–140. https://doi.org/10.1002/mrd.1080390203

Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., Mizusawa, E., Yamakawa, M., Ashida, M., Matsunaga, T. M., Azuma, T., Uchiyama, S., & Fukui, K. (2012). RBMX: A regulator for maintenance and centromeric protection of sister chromatid cohesion. *Cell Reports*, *1*(4), 299–308. https://doi.org/10.1016/j.celrep.2012.02.005

Mazid, M. A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., Liu, H., Fu, L., Yang, Y., Ibañez, D. P., Lai, J., Wei, X., An, J., Guo, P., Yuan, Y., … Esteban, M. A. (2022). Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature*, *605*(7909), 315–324. https://doi.org/10.1038/s41586-022-04625-0

McGinnis, C. S., Patterson, D. M., Winkler, J., Conrad, D. N., Hein, M. Y., Srivastava, V., Hu, J. L., Murrow, L. M., Weissman, J. S., Werb, Z., Chow, E. D., & Gartner, Z. J. (2019). MULTI-seq: Sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nature Methods*, *16*(7), 619–626. https://doi.org/10.1038/s41592-019-0433-8

McIntire, E., Leonhard, K., Taapken, S., & Larson, A. L. (2021). G-Banded Karyotyping of Human Pluripotent Stem Cell Cultures. *Methods in Molecular Biology (Clifton, N.J.)*, *2239*, 251–268. https://doi.org/10.1007/978-1-0716-1084-8_16

McIntire, E., Taapken, S., Leonhard, K., & Larson, A. L. (2020). Genomic Stability Testing of Pluripotent Stem Cells. *Current Protocols in Stem Cell Biology*, *52*(1), e107. https://doi.org/10.1002/cpsc.107

Mekhoubad, S., Bock, C., de Boer, A. S., Kiskinis, E., Meissner, A., & Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*, *10*(5), 595–609. https://doi.org/10.1016/j.stem.2012.02.014

Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A. T. L., Marioni, J. C., & Reik, W. (2019). Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. *Cell Reports*, *26*(4), 815-824.e4. https://doi.org/10.1016/j.celrep.2018.12.099

Milagre, I., Stubbs, T. M., King, M. R., Spindel, J., Santos, F., Krueger, F., Bachman, M., Segonds-Pichon, A., Balasubramanian, S., Andrews, S. R., Dean, W., & Reik, W. (2017). Gender Differences in Global but Not Targeted Demethylation in iPSC Reprogramming. *Cell Reports*, *18*(5), 1079–1089. https://doi.org/10.1016/j.celrep.2017.01.008

Mitalipova, M. M., Rao, R. R., Hoyer, D. M., Johnson, J. A., Meisner, L. F., Jones, K. L., Dalton, S., & Stice, S. L. (2005). Preserving the genetic integrity of human embryonic stem cells. *Nature Biotechnology*, *23*(1), 19–20. https://doi.org/10.1038/nbt0105-19

Moreira de Mello, J. C., Fernandes, G. R., Vibranovski, M. D., & Pereira, L. V. (2017). Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing. *Scientific Reports*, *7*(1), 10794. https://doi.org/10.1038/s41598-017-11044-z

Moya-Jódar, M., Ullate-Agote, A., Barlabé, P., Rodríguez-Madoz, J. R., Abizanda, G., Barreda, C., Carvajal-Vergara, X., Vilas-Zornoza, A., Romero, J. P., Garate, L., Agirre, X., Coppiello, G., Prósper, F., & Aranguren, X. L. (2022). *Revealing cell populations catching the early stages of the human embryo development in naïve pluripotent stem cells* (p. 2022.03.06.482812). bioRxiv. https://doi.org/10.1101/2022.03.06.482812

Müller, S., Cho, A., Liu, S. J., Lim, D. A., & Diaz, A. (2018). CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics (Oxford, England)*, *34*(18), 3217–3219. https://doi.org/10.1093/bioinformatics/bty316

Munné, S., Nakajima, S. T., Najmabadi, S., Sauer, M. V., Angle, M. J., Rivas, J. L., Mendieta, L. V., Macaso, T. M., Sawarkar, S., Nadal, A., Choudhary, K., Nezhat, C., Carson, S. A., & Buster, J. E. (2020). First PGT-A using human in vivo blastocysts recovered by uterine lavage: Comparison with matched IVF embryo controls†. *Human Reproduction (Oxford, England)*, *35*(1), 70–80. https://doi.org/10.1093/humrep/dez242

Na, J., Baker, D., Zhang, J., Andrews, P. W., & Barbaric, I. (2014). Aneuploidy in pluripotent stem cells and implications for cancerous transformation. *Protein & Cell*, *5*(8), 569–579. https://doi.org/10.1007/s13238-014-0073-9

Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T., & Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, *537*(7618), 57–62. https://doi.org/10.1038/nature19096

Natale, C. F., Angrisano, T., Pistelli, L., Falco, G., Calabrò, V., Netti, P. A., & Ventre, M. (2020). Topographic Cues Impact on Embryonic Stem Cell Zscan4-Metastate. *Frontiers in Bioengineering and Biotechnology*, *8*, 178. https://doi.org/10.3389/fbioe.2020.00178

Nazor, K. L., Altun, G., Lynch, C., Tran, H., Harness, J. V., Slavin, I., Garitaonandia, I., Müller, F.-J., Wang, Y.-C., Boscolo, F. S., Fakunle, E., Dumevska, B., Lee, S., Park, H. S., Olee, T., D'Lima, D. D., Semechkin, R., Parast, M. M., Galat, V., … Laurent, L. C. (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell*, *10*(5), 620–634. https://doi.org/10.1016/j.stem.2012.02.013

Nichols, J., & Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell*, *4*(6), 487–492. https://doi.org/10.1016/j.stem.2009.05.015

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., & Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science (New York, N.Y.)*, *303*(5658), 644–649. https://doi.org/10.1126/science.1092727

Ooi, S. K., Wolf, D., Hartung, O., Agarwal, S., Daley, G. Q., Goff, S. P., & Bestor, T. H. (2010). Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics & Chromatin*, *3*(1), 17. https://doi.org/10.1186/1756-8935-3-17

Padilla-Nash, H. M., Barenboim-Stapleton, L., Difilippantonio, M. J., & Ried, T. (2006). Spectral karyotyping analysis of human and mouse chromosomes. *Nature Protocols*, *1*(6), 3129–3142. https://doi.org/10.1038/nprot.2006.358

Palmerola, K. L., Amrane, S., De Los Angeles, A., Xu, S., Wang, N., de Pinho, J., Zuccaro, M. V., Taglialatela, A., Massey, D. J., Turocy, J., Robles, A., Subbiah, A., Prosser, B., Lobo, R., Ciccia, A., Koren, A., Baslan, T., & Egli, D. (2022). Replication stress impairs chromosome segregation and preimplantation development in human embryos. *Cell*, S0092-8674(22)00780-2. https://doi.org/10.1016/j.cell.2022.06.028

Panopoulos, A. D., Smith, E. N., Arias, A. D., Shepard, P. J., Hishida, Y., Modesto, V., Diffenderfer, K. E., Conner, C., Biggs, W., Sandoval, E., D'Antonio-Chronowska, A., Berggren, W. T., Izpisua Belmonte, J. C., & Frazer, K. A. (2017). Aberrant DNA Methylation in Human iPSCs Associates with MYC-Binding Motifs in a Clone-Specific Manner Independent of Genetics. *Cell Stem Cell*, *20*(4), 505-517.e6. https://doi.org/10.1016/j.stem.2017.03.010

Pasque, V., Karnik, R., Chronis, C., Petrella, P., Langerman, J., Bonora, G., Song, J., Vanheer, L., Sadhu Dimashkie, A., Meissner, A., & Plath, K. (2018). X Chromosome Dosage Influences DNA Methylation Dynamics during Reprogramming to Mouse iPSCs. *Stem Cell Reports*, *10*(5), 1537–1550. https://doi.org/10.1016/j.stemcr.2018.03.019

Pastor, W. A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S. E., & Clark, A. T. (2016). Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell*, *18*(3), 323–329. https://doi.org/10.1016/j.stem.2016.01.019

Pelleri, M. C., Cattani, C., Vitale, L., Antonaros, F., Strippoli, P., Locatelli, C., Cocchi, G., Piovesan, A., & Caracausi, M. (2018). Integrated Quantitative Transcriptome Maps of Human Trisomy 21 Tissues and Cells. *Frontiers in genetics*, *9*, 125. https://doi.org/10.3389/fgene.2018.00125

Pera, M. F. (2010). Defining pluripotency. *Nature Methods*, *7*(11), 885–887. https://doi.org/10.1038/nmeth1110-885

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., & Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*, *165*(4), 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023

Polo, J. M., Anderssen, E., Walsh, R. M., Schwarz, B. A., Nefzger, C. M., Lim, S. M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., Bar-Nur, O., Cheloufi, S., Stadtfeld, M., Figueroa, M. E., Robinton, D., Natesan, S., Melnick, A., Zhu, J., Ramaswamy, S., & Hochedlinger, K. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell*, *151*(7), 1617–1632. https://doi.org/10.1016/j.cell.2012.11.039

Pomp, O., Dreesen, O., Leong, D. F. M., Meller-Pomp, O., Tan, T. T., Zhou, F., & Colman, A. (2011). Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. *Cell Stem Cell*, *9*(2), 156–165. https://doi.org/10.1016/j.stem.2011.06.004

Potapova, T., & Gorbsky, G. J. (2017). The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. *Biology*, *6*(1), E12. https://doi.org/10.3390/biology6010012

Rebuzzini, P., Neri, T., Mazzini, G., Zuccotti, M., Redi, C. A., & Garagna, S. (2008). Karyotype analysis of the euploid cell population of a mouse embryonic stem cell line revealed a high incidence of chromosome abnormalities that varied during culture. *Cytogenetic and Genome Research*, *121*(1), 18–24. https://doi.org/10.1159/000124377

Rodriguez-Terrones, D., Hartleben, G., Gaume, X., Eid, A., Guthmann, M., Iturbide, A., & Torres-Padilla, M.-E. (2020). A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Reports*, *21*(1), e48354. https://doi.org/10.15252/embr.201948354

Rossi, A., Lickfett, S., Martins, S., & Prigione, A. (2022). A call for consensus guidelines on monitoring the integrity of nuclear and mitochondrial genomes in human pluripotent stem cells. *Stem Cell Reports*, *17*(4), 707–710. https://doi.org/10.1016/j.stemcr.2022.01.019

Ryan, C. J., Kennedy, S., Bajrami, I., Matallanas, D., & Lord, C. J. (2017). A Compendium of Co-regulated Protein Complexes in Breast Cancer Reveals Collateral Loss Events. *Cell Systems*, *5*(4), 399-409.e5. https://doi.org/10.1016/j.cels.2017.09.011

Saiba, R., Arava, M., & Gayen, S. (2018). Dosage compensation in human pre-implantation embryos: X-chromosome inactivation or dampening? *EMBO Reports*, *19*(8), e46294. https://doi.org/10.15252/embr.201846294

Schulz, E. G. (2017). X-chromosome dosage as a modulator of pluripotency, signalling and differentiation? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1733), 20160366. https://doi.org/10.1098/rstb.2016.0366

Schulz, E. G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., & Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell*, *14*(2), 203–216. https://doi.org/10.1016/j.stem.2013.11.022

Shakoori, A. R. (2017). Fluorescence In Situ Hybridization (FISH) and Its Applications. *Chromosome Structure and Aberrations*, 343–367. https://doi.org/10.1007/978-81-322-3673-3_16

Sharova, L. V., Sharov, A. A., Piao, Y., Stagg, C. A., Amano, T., Qian, Y., Dudekula, D., Schlessinger, D., & Ko, M. S. H. (2016). Emergence of undifferentiated colonies from mouse embryonic stem cells undergoing differentiation by retinoic acid treatment. *In Vitro Cellular & Developmental Biology. Animal*, *52*(5), 616–624. https://doi.org/10.1007/s11626-016-0013-5

Shen, H., Yang, M., Li, S., Zhang, J., Peng, B., Wang, C., Chang, Z., Ong, J., & Du, P. (2021). Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell*, *184*(11), 2843-2859.e20. https://doi.org/10.1016/j.cell.2021.04.020

Shen, Y., Matsuno, Y., Fouse, S. D., Rao, N., Root, S., Xu, R., Pellegrini, M., Riggs, A. D., & Fan, G. (2008). X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4709–4714. https://doi.org/10.1073/pnas.0712018105

Shimizu, T., Ueda, J., Ho, J. C., Iwasaki, K., Poellinger, L., Harada, I., & Sawada, Y. (2012). Dual inhibition of Src and GSK3 maintains mouse embryonic stem cells, whose differentiation is mechanically regulated by Src signaling. *Stem Cells (Dayton, Ohio)*, *30*(7), 1394–1404. https://doi.org/10.1002/stem.1119

Silva, S. S., Rowntree, R. K., Mekhoubad, S., & Lee, J. T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4820–4825. https://doi.org/10.1073/pnas.0712136105

Singer, Z. S., Yong, J., Tischler, J., Hackett, J. A., Altinok, A., Surani, M. A., Cai, L., & Elowitz, M. B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular Cell*, *55*(2), 319–331. https://doi.org/10.1016/j.molcel.2014.06.029

Singh, V. K., Kalsan, M., Kumar, N., Saini, A., & Chandra, R. (2015). Induced pluripotent stem cells: Applications in regenerative medicine, disease modeling, and drug discovery. *Frontiers in Cell and Developmental Biology*, *3*, 2. https://doi.org/10.3389/fcell.2015.00002

Singh, V. K., Saini, A., Kalsan, M., Kumar, N., & Chandra, R. (2016). Describing the Stem Cell Potency: The Various Methods of Functional Assessment and In silico Diagnostics. *Frontiers in Cell and Developmental Biology*, *4*, 134. https://doi.org/10.3389/fcell.2016.00134

Skelly, D. A., Czechanski, A., Byers, C., Aydin, S., Spruce, C., Olivier, C., Choi, K., Gatti, D. M., Raghupathy, N., Keele, G. R., Stanton, A., Vincent, M., Dion, S., Greenstein, I., Pankratz, M., Porter, D. K., Martin, W., O'Connor, C., Qin, W., … Reinholdt, L. G. (2020). Mapping the Effects of Genetic Variation on Chromatin State and Gene Expression Reveals Loci That Control Ground State Pluripotency. *Cell Stem Cell*, *27*(3), 459-469.e8. https://doi.org/10.1016/j.stem.2020.07.005

Slonim, D. K., Koide, K., Johnson, K. L., Tantravahi, U., Cowan, J. M., Jarrah, Z., & Bianchi, D. W. (2009). Functional genomic analysis of amniotic fluid cell-free mRNA suggests that oxidative stress is significant in Down syndrome fetuses. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9425–9429. https://doi.org/10.1073/pnas.0903909106

Smirnov, A., Cappello, A., Lena, A. M., Anemona, L., Mauriello, A., Di Daniele, N., Annicchiarico-Petruzzelli, M., Melino, G., & Candi, E. (2018). ZNF185 is a p53 target gene following DNA damage. *Aging*, *10*(11), 3308–3326. https://doi.org/10.18632/aging.101639

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., & Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, *511*(7511), 611–615. https://doi.org/10.1038/nature13581

Song, J., Janiszewski, A., De Geest, N., Vanheer, L., Talon, I., El Bakkali, M., Oh, T., & Pasque, V. (2019). X-Chromosome Dosage Modulates Multiple Molecular and Cellular Properties of Mouse Pluripotent Stem Cells Independently of Global DNA Methylation Levels. *Stem Cell Reports*, *12*(2), 333–350. https://doi.org/10.1016/j.stemcr.2018.12.004

Spits, C., Mateizel, I., Geens, M., Mertzanidou, A., Staessen, C., Vandeskelde, Y., Van der Elst, J., Liebaers, I., & Sermon, K. (2008). Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature Biotechnology*, *26*(12), 1361–1363. https://doi.org/10.1038/nbt.1510

Stadtfeld, M., Maherali, N., Breault, D. T., & Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell stem cell*, *2*(3), 230–240. https://doi.org/10.1016/j.stem.2008.02.001

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., & Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Molecular Systems Biology*, *8*, 608. https://doi.org/10.1038/msb.2012.40

Sugawara, A., Goto, K., Sotomaru, Y., Sofuni, T., & Ito, T. (2006). Current status of chromosomal abnormalities in mouse embryonic stem cell lines used in Japan. *Comparative Medicine*, *56*(1), 31–34.

Suwińska, A. (2012). Preimplantation mouse embryo: Developmental fate and potency of blastomeres. *Results and Problems in Cell Differentiation*, *55*, 141–163. https://doi.org/10.1007/978-3-642-30406-4_8

Svenson, K. L., Gatti, D. M., Valdar, W., Welsh, C. E., Cheng, R., Chesler, E. J., Palmer, A. A., McMillan, L., & Churchill, G. A. (2012). High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics*, *190*(2), 437–447. https://doi.org/10.1534/genetics.111.132597

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J., & Schultz, R. M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Developmental Biology*, *269*(1), 276–285. https://doi.org/10.1016/j.ydbio.2004.01.028

Tagliaferri, D., De Angelis, M. T., Russo, N. A., Marotta, M., Ceccarelli, M., Del Vecchio, L., De Felice, M., & Falco, G. (2016). Retinoic Acid Specifically Enhances Embryonic Stem Cell Metastate Marked by Zscan4. *PloS One*, *11*(2), e0147683. https://doi.org/10.1371/journal.pone.0147683

Tagliaferri, D., Mazzone, P., Noviello, T. M. R., Addeo, M., Angrisano, T., Del Vecchio, L., Visconte, F., Ruggieri, V., Russi, S., Caivano, A., Cantone, I., De Felice, M., Ceccarelli, M., Cerulo, L., & Falco, G. (2019). Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of Dux and Duxbl1. *Frontiers in Cell and Developmental Biology*, *7*, 385. https://doi.org/10.3389/fcell.2019.00385

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., & Smith, A. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, *158*(6), 1254–1269. https://doi.org/10.1016/j.cell.2014.08.029

Tan, C., Chen, X., Wang, F., Wang, D., Cao, Z., Zhu, X., Lu, C., Yang, W., Gao, N., Gao, H., Guo, Y., & Zhu, L. (2019). A multiplex droplet digital PCR assay for non-invasive prenatal testing of fetal aneuploidies. *Analyst*, *144*(7), 2239–2247. https://doi.org/10.1039/C8AN02018C

Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., & Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell*, *29*(3), 449-459.e6. https://doi.org/10.1016/j.stem.2022.01.014

Taylor, C. J., Bolton, E. M., & Bradley, J. A. (2011). Immunological considerations for embryonic and induced pluripotent stem cell banking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1575), 2312–2322. https://doi.org/10.1098/rstb.2011.0030

Tchieu, J., Kuoy, E., Chin, M. H., Trinh, H., Patterson, M., Sherman, S. P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J. A., Clark, A. T., Pyle, A. D., Lowry, W. E., & Plath, K. (2010). Female human iPSCs retain an inactive X chromosome. *Cell Stem Cell*, *7*(3), 329–342. https://doi.org/10.1016/j.stem.2010.06.024

Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., Duc, J., Cohen, M. A., Wert, K. J., Castanon, R., Zhang, Z., Huang, Y., Nery, J. R., Drotar, J., Lungjangwa, T., … Jaenisch, R. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell*, *19*(4), 502–515. https://doi.org/10.1016/j.stem.2016.06.011

Theunissen, T. W., Powell, B. E., Wang, H., Mitalipova, M., Faddah, D. A., Reddy, J., Fan, Z. P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y., Rangarajan, S., D'Alessio, A., Zhang, J., Gao, Q., Dawlaty, M. M., Young, R. A., … Jaenisch, R. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*, *15*(4), 471–487. https://doi.org/10.1016/j.stem.2014.07.002

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N. A., Eilertson, K. E., Tsang, P., Baba, S., White, M. P., Sami, S., Srivastava, D., Conklin, B. R.,

Panning, B., & Yamanaka, S. (2012). Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. *Cell Stem Cell*, *11*(1), 91–99. https://doi.org/10.1016/j.stem.2012.05.019

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., & Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science (New York, N.Y.)*, *317*(5840), 916–924. https://doi.org/10.1126/science.1142210

Torres-Padilla, M.-E., & Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: A stochastic advantage. *Development (Cambridge, England)*, *141*(11), 2173–2181. https://doi.org/10.1242/dev.102624

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., & Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development (Cambridge, England)*, *135*(5), 909–918. https://doi.org/10.1242/dev.017400

Turner, J. M. A., Mahadevaiah, S. K., Elliott, D. J., Garchon, H.-J., Pehrson, J. R., Jaenisch, R., & Burgoyne, P. S. (2002). Meiotic sex chromosome inactivation in male mice with targeted disruptions of Xist. *Journal of Cell Science*, *115*(Pt 21), 4097–4105. https://doi.org/10.1242/jcs.00111

Vallot, C., Ouimette, J.-F., Makhlouf, M., Féraud, O., Pontis, J., Côme, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., & Rougeulle, C. (2015). Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell*, *16*(5), 533–546. https://doi.org/10.1016/j.stem.2015.03.016

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., & Izpisua Belmonte, J. C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development (Cambridge, England)*, *138*(17), 3699–3709. https://doi.org/10.1242/dev.064741

Velkova, A., Carvalho, M. A., Johnson, J. O., Tavtigian, S. V., & Monteiro, A. N. A. (2010). Identification of Filamin A as a BRCA1-interacting protein required for efficient DNA repair. *Cell Cycle (Georgetown, Tex.)*, *9*(7), 1421–1433. https://doi.org/10.4161/cc.9.7.11256

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227–232. https://doi.org/10.1038/nrg3185

Wang, W., Lu, G., Su, X., Tang, C., Li, H., Xiong, Z., Leung, C.-K., Wong, M.-S., Liu, H., Ma, J.-L., Cheung, H.-H., Kung, H.-F., Chen, Z.-J., & Chan, W.-Y. (2020). Pten-mediated Gsk3β modulates the naïve pluripotency maintenance in embryonic stem cells. *Cell Death & Disease*, *11*(2), 1–15. https://doi.org/10.1038/s41419-020-2271-0

Wang, Y., Na, Q., Li, X., Tee, W.-W., Wu, B., & Bao, S. (2021). Retinoic acid induces NELFA-mediated 2C-like state of mouse embryonic stem cells associates with epigenetic modifications and metabolic processes in chemically defined media. *Cell Proliferation*, *54*(6), e13049. https://doi.org/10.1111/cpr.13049

Weissbein, U., Schachter, M., Egli, D., & Benvenisty, N. (2016). Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq. *Nature Communications*, *7*, 12144. https://doi.org/10.1038/ncomms12144

Werbowetski-Ogilvie, T. E., Bossé, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., Wynder, T., Smith, M.-J., Dingwall, S., Carter, T., Williams, C., Harris, C., Dolling, J., Wynder, C., Boreham, D., & Bhatia, M. (2009). Characterization of

human embryonic stem cells with features of neoplastic progression. *Nature Biotechnology*, *27*(1), 91–97. https://doi.org/10.1038/nbt.1516
Werner, R. J., Schultz, B. M., Huhn, J. M., Jelinek, J., Madzo, J., & Engel, N. (2017). Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biology of Sex Differences*, *8*(1), 28. https://doi.org/10.1186/s13293-017-0150-x
Xian, S., Dosset, M., Almanza, G., Searles, S., Sahani, P., Waller, T. C., Jepsen, K., Carter, H., & Zanetti, M. (2021). The unfolded protein response links tumor aneuploidy to local immune dysregulation. *EMBO Reports*, *22*(12), e52509. https://doi.org/10.15252/embr.202152509
Xie, P., Ouyang, Q., Leng, L., Hu, L., Cheng, D., Tan, Y., Lu, G., & Lin, G. (2016). The dynamic changes of X chromosome inactivation during early culture of human embryonic stem cells. *Stem Cell Research*, *17*(1), 84–92. https://doi.org/10.1016/j.scr.2016.05.011
Xu, C., Inokuma, M. S., Denham, J., Golds, K., Kundu, P., Gold, J. D., & Carpenter, M. K. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology*, *19*(10), 971–974. https://doi.org/10.1038/nbt1001-971
Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J., Horvath, S., & Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, *500*(7464), 593–597. https://doi.org/10.1038/nature12364
Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., & Yamada, Y. (2017). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature*, *548*(7666), 224–227. https://doi.org/10.1038/nature23286
Yang, J., Ryan, D. J., Wang, W., Tsang, J. C.-H., Lan, G., Masaki, H., Gao, X., Antunes, L., Yu, Y., Zhu, Z., Wang, J., Kolodziejczyk, A. A., Campos, L. S., Wang, C., Yang, F., Zhong, Z., Fu, B., Eckersley-Maslin, M. A., Woods, M., … Liu, P. (2017). Establishment of mouse expanded potential stem cells. *Nature*, *550*(7676), 393–397. https://doi.org/10.1038/nature24052
Yang, M., Rito, T., Metzger, J., Naftaly, J., Soman, R., Hu, J., Albertini, D. F., Barad, D. H., Brivanlou, A. H., & Gleicher, N. (2021). Depletion of aneuploid cells in human embryos and gastruloids. *Nature Cell Biology*, *23*(4), 314–321. https://doi.org/10.1038/s41556-021-00660-7
Yang, Y., Liu, B., Xu, J., Wang, J., Wu, J., Shi, C., Xu, Y., Dong, J., Wang, C., Lai, W., Zhu, J., Xiong, L., Zhu, D., Li, X., Yang, W., Yamauchi, T., Sugawara, A., Li, Z., Sun, F., … Deng, H. (2017). Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell*, *169*(2), 243-257.e25. https://doi.org/10.1016/j.cell.2017.02.005
Yilmaz, A., & Benvenisty, N. (2019). Defining Human Pluripotency. *Cell Stem Cell*, *25*(1), 9–22. https://doi.org/10.1016/j.stem.2019.06.010
Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science (New York, N.Y.)*, *318*(5858), 1917–1920. https://doi.org/10.1126/science.1151526
Yu, X., Liang, S., Chen, M., Yu, H., Li, R., Qu, Y., Kong, X., Guo, R., Zheng, R., Izsvák, Z., Sun, C., Yang, M., & Wang, J. (2022). Recapitulating early human development with 8C-like cells. *Cell Reports*, *39*(12), 110994. https://doi.org/10.1016/j.celrep.2022.110994

Zalzman, M., Falco, G., Sharova, L. V., Nishiyama, A., Thomas, M., Lee, S.-L., Stagg, C. A., Hoang, H. G., Yang, H.-T., Indig, F. E., Wersto, R. P., & Ko, M. S. H. (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature*, *464*(7290), 858–863. https://doi.org/10.1038/nature08882

Zhang, C., Wen, H., Liu, S., Fu, E., Yu, L., Chen, S., Han, Q., Li, Z., & Liu, N. (2022). Maternal Factor Dppa3 Activates 2C-Like Genes and Depresses DNA Methylation in Mouse Embryonic Stem Cells. *Frontiers in Cell and Developmental Biology*, *10*, 882671. https://doi.org/10.3389/fcell.2022.882671

Zhang, H., Yang, X., Feng, X., Xu, H., Yang, Q., Zou, L., Yan, M., Liu, D., Su, X., & Jiao, B. (2018). Chromosome-wide gene dosage rebalance may benefit tumor progression. *Molecular Genetics and Genomics: MGG*, *293*(4), 895–906. https://doi.org/10.1007/s00438-018-1429-2

Zhang, J., Hirst, A. J., Duan, F., Qiu, H., Huang, R., Ji, Y., Bai, L., Zhang, F., Robinson, D., Jones, M., Li, L., Wang, P., Jiang, P., Andrews, P. W., Barbaric, I., & Na, J. (2019). Anti-apoptotic Mutations Desensitize Human Pluripotent Stem Cells to Mitotic Stress and Enable Aneuploid Cell Survival. *Stem Cell Reports*, *12*(3), 557–571. https://doi.org/10.1016/j.stemcr.2019.01.013

Zhang, M., Cheng, L., Jia, Y., Liu, G., Li, C., Song, S., Bradley, A., & Huang, Y. (2016). Aneuploid embryonic stem cells exhibit impaired differentiation and increased neoplastic potential. *The EMBO Journal*, *35*(21), 2285–2300. https://doi.org/10.15252/embj.201593103

Zhang, W., Chen, F., Chen, R., Xie, D., Yang, J., Zhao, X., Guo, R., Zhang, Y., Shen, Y., Göke, J., Liu, L., & Lu, X. (2019). Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Research*, *47*(16), 8485–8501. https://doi.org/10.1093/nar/gkz594

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049. https://doi.org/10.1038/ncomms14049

Zheng, T., Zhou, H., Li, X., Peng, D., Yang, Y., Zeng, Y., Liu, H., Ren, J., & Zhao, Y. (2020). RBMX is required for activation of ATR on repetitive DNAs to maintain genome stability. *Cell Death and Differentiation*, *27*(11), 3162–3176. https://doi.org/10.1038/s41418-020-0570-8

Zhou, L., & Dean, J. (2015). Reprogramming the genome to totipotency in mouse embryos. *Trends in Cell Biology*, *25*(2), 82–91. https://doi.org/10.1016/j.tcb.2014.09.006

Zhu, J., Tsai, H.-J., Gordon, M. R., & Li, R. (2018). Cellular Stress Associated with Aneuploidy. *Developmental Cell*, *44*(4), 420–431. https://doi.org/10.1016/j.devcel.2018.02.002

Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J. E., Crompton, L. A., John, R., Feil, R., & Brockdorff, N. (2005). Global hypomethylation of the genome in XX embryonic stem cells. *Nature Genetics*, *37*(11), 1274–1279. https://doi.org/10.1038/ng1663