Induced pluripotent stem cell models of Kleefstra syndrome

Karla Manzano Vargas

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

September 2018

Department of Human Genetics
McGill University
Montreal, Quebec, Canada

Dedication

I dedicate this thesis to my mother, Alfonsina Leticia Vargas Ramirez.
Thank you, mom, for teaching me how to be a strong woman.

Dedico esta tesis a mi madre Alfonsina Leticia Vargas Ramírez.
Gracias mamá por enseñarme a ser una mujer fuerte.

Abstract

Kleefstra Syndrome (KS) is a neurodevelopmental disorder mainly defined by intellectual disability, characteristic facial features and hypotonia. KS is caused by a microscopic deletion of the chromosomal region 9q34.3 or by mutations in the euchromatic histone methyltransferase 1 (*EHMT1*). A detailed phenotypic description of KS has been achieved through the study of animal models, allowing us to uncover various molecular mechanisms of the disease. Despite great progress in the study of KS and its underlying etiology, treatments remain restricted to targeting specific symptoms (i.e., seizures and speech) rather than the underlying disease pathology. Recently, human induced pluripotent stem cells (iPSCs) have been shown to be good models for drug discovery as well as the study of disease pathology since a patient's somatic cells can be reprogrammed for investigation. Here we characterized iPSC lines from two KS syndrome patients and the patients' sex-matched siblings as controls. For these lines, quality control experiments were performed to validate the purity and integrity of our iPSCs. We also used CRISPR/Cas9 to generate a heterozygous knockout (KO) of *EHMT1* in one of our healthy sex-matched control siblings. Particularly, the first two exons of the *EHMT1* SET-domain were deleted in this healthy iPSC line to generate an artificial disease model. Overall, our patient and KO iPSC lines meet all the characteristics and requirements for pluripotency and lack any observable gross chromosomal abnormalities. The iPSC models we present in this thesis may be employed for drug discovery initiatives and clinical translational research for KS. Ultimately, the potential of our iPSC lines to be genetically engineered and differentiated into an array of cell lineages opens the door for new approaches to study and treat KS.

Résumé

Le syndrome de Kleefstra (SK) est un trouble neuro-développemental principalement caractérisé par un déficit intellectuel, une dysmorphie faciale caractéristique et une hypotonie infantile. Le SK est causé soit par une délétion microscopique de la région chromosomique 9q34.3 ou par une mutation ponctuelle du gène "*euchromatic histone methyltransferase 1*" (*EHMT1*). Une description phénotypique détaillée du SK dans des modèles animaux nous a permis de découvrir différents mécanismes moléculaires de la maladie. Malgré de grands progrès dans l'étude du SK et de son étiologie, les traitements se limitent à soigner des symptômes spécifiques (par exemple: les convulsions et les problèmes du langage) plutôt que la pathologie sous-jacente. Puisque les cellules somatiques d'un patient peuvent être reprogrammées pour fin d'investigation, les cellules souches pluripotentes induites (CSPi) ont récemment été désignées comme étant de bons modèles dans le domaine de la découverte de médicaments et l'étude des pathologies. Ici, nous avons généré des lignées CSPi provenant de deux patients atteints du SK et d'autres provenant de frères et sœurs des patients, jumelées selon le sexe, utilisées comme contrôles. Plusieurs expériences de contrôle de qualité ont été réalisées pour valider la pureté et l'intégrité de nos CSPi. En outre, en utilisant la lignée cellulaire d'un individu contrôle, nous avons utilisé *CRISPR/Cas9* pour générer un *knock-out* (KO) hétérozygote du gène *EHMT1*. De façon plus précise, pour générer un modèle de maladie artificielle, les deux premiers exons du domaine SET du gène *EHMT1* ont été supprimés dans cette lignée. Dans l'ensemble, nos lignées CSPi des patients et du KO répondent à toutes les caractéristiques et exigences de la pluripotence et ne présentent aucune anomalie chromosomique observable. Les modèles CSPi que nous présentons dans ce mémoire peuvent être utilisés pour des initiatives de découverte de médicaments et de recherche translationnelle clinique pour le SK. Finalement, le potentiel de nos lignées CSPi à être modifiées et différenciées génétiquement dans une gamme de lignées cellulaires ouvre la porte à de nouvelles approches pour étudier et traiter le SK.

Table of Contents

List of abbreviations

9qSTDS: 9q SubTelomeric Deletion Syndrome

ASD: Autism Spectrum Disorder

CDH1: Cadherin 1

CDK: Cyclin-Dependent Kinase

cDNA: complementary DNA

CNVs: Copy-Number Variants

COL1A: Collagen Type I Alpha 1 Chain

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DMEM: Dulbecco Modified Eagle's Medium

DNA: Deoxyriboucleic Acid

DNMT2: DNA MethylTransferase 2

EBs: Embryoid Bodies

EHMT1: Euchromatic Histone Lysine Methyltransferase 1

ESCs: Embryonic Stem Cells

FBS: Fetal Bovine Serum

FISH: Fluorescent In Situ Hybridization

H3K9: Histone 3 Lysine 9

HDR: Homology-Directed Repair

hTERT: Telomerase Reverse Transcriptase

IBD: Identity-By-Descent

ICC: Immunocytochemistry

ICM: Inner Cell Mass

ID: Intellectual Disability

IGF2: Insulin Like Growth Factor 2

iPSCs: Induced Pluripotent Stem Cells

KLF4: Krueppel-like factor 4

KS: Kleefstra Syndrome

KSR: KnockOut™ Serum Replacement

MLPA: Multiplex Ligation-Dependent Probe Amplification

NANOG: Nanog homeobox

NHEJ: Nonhomologous End Joining

OCT4: Octamer-binding Transcription Factor 4

OMIM: Online Mendelian Inheritance in Man

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

qPCR: quantitative Polymerase Chain Reaction

RNA: Ribonucleic Acid

SET-domain: Su(var)3-9, Enhancer-of-zeste and Trithorax domain

SNP: Single Nucleotide Polymorphisms

SSCs: Somatic Stem Cells

SSEA-4: Stage-specific Embryonic Antigen-4

STR: Short Tandem Repeat

TALENs: Transcription Activator-Like Effector Nucleases

ZFP42: Zinc Finger Protein 42

List of Figures

Chapter 3: Results

Chapter 5: Conclusions and Future Directions

List of Tables

Chapter 2: Materials and Methods

Acknowledgments

I would like to thank my supervisor, Dr. Carl Ernst for his guidance and continuous support through my Master´s studies. I would also like to express my deep thanks to my supervisory committee members Dr. Loydie Majewska and Dr. Yannis Trakadis for their insights on my project, support, and guidance.

I would like to thank the families, parents, and children, that participated in this study. Thank you to the GeneSpark foundation for kindly proving monetary support for carrying out this project. I am also grateful for the financial support I received from the Mexican National Council of Science and Technology, as well as from Mitacs.

I would like to acknowledge Dr. Huashan Peng for their support and training, as well as his collaboration in the study. I would also like to thank Malvin Jefri for his help and support in the laboratory. I would also like to express my deep thanks to my friend Dr. Heika Silveira, thank you for all the support you gave me in the difficult moments. I am also thankful for the daily technical and social support from my fellow lab members Zahia Aouabed, Jean-François Théroux, Nuwan Hettige, Jessie Poquérusse, Scott Bell and Hanrong Wu.

Estoy muy agradecida con mi familia por todo el amor que me ha dado a lo largo de mi vida, en especial, gracias a mi mama y mis hermanas, sin ustedes no estaría donde estoy hoy. Finalement, merci beaucoup à mon mari pour l’amour qu'il me donne et merci à ma belle-famille pour tout son soutien.

Preface and contribution of authors

The research presented in this traditional thesis was made possible through my collaboration with Dr. Huashan Peng who helped me to create the iPSC lines and created the *EHMT1* SET-domain heterozygous KO. Dr. Huashan Peng generated the immunocytochemistry profiles presented in Figure 6. Zahia Aouabed did the cell line authentication analysis. She also generated Figure 12. Malvin Jeffri ran the gel presented in Figure 13. Zahia Aouabed, Nuwan Hettige, Kellie MacDonald, and Jessie Poquérusse provided editorial support. The candidate conducted all the other experiments.

Contextualizing this thesis in the context of a larger *EHMT1* mutation study in Dr. Ernst's laboratory

This MSc project is part of a larger project funded by GeneSpark, an association initially founded by parents of children affected by Kleefstra Syndrome (KS). The main research objective of this project is to discover treatments for KS. The hypothesis of this MSc project was to test whether KS could be modeled in iPSCs. To our knowledge, only one other group in the world has successfully reported the establishment of a KS iPSC line (Varga, Nemes et al. 2016). We established several iPSC lines from KS patients and their sex-matched siblings. The iPSC lines we created have a diversity of *EHMT1* mutations. These lines are going to be used in neuron studies with a translational research approach. Modeling KS in iPSC-derived neurons represents a very important tool to understand the role of *EHMT1* gene mutations in the lead to the disease. We have started the reprogramming process of the patient iPSC lines to Neural Progenitor Cells (NPCs). These lines will soon be ready for gene expression analysis.

At a broad level, our *EHMT1* project is based on the hypothesis that a reduced *EHMT1* gene dosage alters gene expression patterns in the brain, which is supported by strong preliminary data and recent publications (Balemans, Ansar et al. 2014) (Nagy, Kobolak et al. 2017). The overall impact on understanding basic mechanisms of why mutations in the *EHMT1* gene lead to the disease is likely to be very high because it may lead to the discovery of novel molecular drug targets. This MSc project also led to the development and implementation of a pipeline for fast and low-cost quality control of iPSC lines, a very important tool that ensures the quality of all the iPSCs produced in our laboratory. Also, during this MSc project, we worked on a strategy to monitor *EHMT1* expression. We published this strategy in the International Journal of Neuropsychopharmacology under the title "Strategies to Advance Drug Discovery in Rare Monogenic Intellectual Disability Syndromes" (Hettige, Manzano-Vargas et al. 2018). Briefly, we introduce a 2A system at the end of the *EHMT1* gene, coupled with GFP. In this way, we use GFP levels as a read-out of *EHMT1* expression. We have been collaborating closely with researchers at the Montreal General Hospital to achieve this objective. This system is going to be applied to a high throughput screening approach looking for drugs that may increase *EHMT1* levels and bring it to normal expression levels. More details of the system can be found in the Future Directions section. This system is another example of the important contribution of this MSc project to the study of KS and highlights the importance of establishing high-quality KS iPSC lines.

Chapter I: Literature review and introduction

A. Kleefstra syndrome

Rare monogenic Intellectual Disability syndromes

Intellectual disability (ID) is characterized by deficits in intellectual and adaptive functioning arising before the age of eighteen (Chiurazzi and Pirozzi 2016). Patients affected with ID may exhibit other disorders such as Autism Spectrum Disorder (ASD), attention deficit disorder and behavioral problems (Chiurazzi and Pirozzi 2016). The estimated prevalence of ID is 14.3 out of 1000 individuals (Chiurazzi and Pirozzi 2016), affecting approximately 1% of the global population (Maulik, Mascarenhas et al. 2011).

ID could be classified as syndromic and non-syndromic, depending on whether or not other organs are affected and on the presence of typical facial characteristics (Chiurazzi and Pirozzi 2016). Environment and genetics may co-operate or act alone to cause ID, creating a great heterogeneity of clinical features. Genetically induced ID could be attributed to single or multiple genes, but monogenic ID syndromes have been widely studied in the past decades - examples include Fragile X Syndrome and Rett syndrome.

As our ability to understand and manipulate the human genome has rapidly increased, in the last years many disorders of previously unknown etiology have been genetically characterized (Carvill and Mefford 2015). For example, Kabuki syndrome, an autosomal dominant congenital syndrome, was a disorder of unknown etiology until 2010, when Sarah B Ng et al identified *KMT2D* as the gene causative of the disease and its characteristic phenotype (Ng, Bigham et al. 2010).

Kleefstra syndrome

Kleefstra Syndrome (KS) (OMIM 610253) is a rare monogenic neurodevelopmental disorder characterized by ID, autistic-like features, hypotonia and characteristic facial features (Willemsen, Vulto-van Silfhout et al. 2012). The prevalence of this disease is unknown. To date, 114 cases have been reported. Different research groups have described KS clinical features since the 90's, (Schimmenti, Berry et al. 1994) (Font-Montgomery, Weaver et al. 2004) (Harada, Visser et al. 2004). Even though characteristic clinical features of the previously known 9q34.3 terminal deletion syndrome (9qSTDS) have been broadly described, the molecular etiology of KS was unknown until 2005, when Kleefstra et al. identified the euchromatin histone methyltransferase 1 (*EHMT1*) gene as the gene responsible for the disease (Kleefstra, Smidt et al. 2005). Molecular

gene identification was achieved through genetic analysis of a patient with a *de novo* balanced translocation affecting the chromosomal region 9q34.3 (Kleefstra, Smidt et al. 2005). The translocation also disrupted the X chromosome gene *ZNF81*; however, the later gene had not been associated with the syndromic features of KS in any other studied patient. This evidence joined to other studies looking at overlapping deletions in several patients, led to the identification of *EHMT1* as the causative gene of KS (Kleefstra, Smidt et al. 2005). Particularly, KS is caused by a microscopic deletion in the chromosomal region 9q34.3. Point mutations in the *EHMT1* gene leading to aberrant protein function could also cause KS (Willemsen, Vulto-van Silfhout et al. 2012).

Clinical characteristics

KS is an autosomal dominant disorder characterized by distinctive facial features, including synophrys, microcephaly, midface hypoplasia, hypertelorism, prognathism, brachycephaly and protruding tongue (Willemsen, Vulto-van Silfhout et al. 2012). Other symptoms include heart and urogenital defects, febrile seizures, genital defects in males, renal defects and epilepsy (Kleefstra, van Zelst-Stams et al. 2009) (Willemsen, Vulto-van Silfhout et al. 2012). Patients may also show a regressive phenotype during or after adolescence (Willemsen, Vulto-van Silfhout et al. 2012).

Most of the reported KS patients have a microdeletion in the subtelomeric region 9q34.3 (Willemsen, Vulto-van Silfhout et al. 2012). In 2012, Willemsen et al. reported a group of twenty-nine KS patients, made up principally of patients with a 9q34.3 microdeletion and some patients with an *EHMT1* intragenic mutation. Deletions sizes varied widely among patients, from 270 kb to 3.85 Mb (Willemsen, Vulto-van Silfhout et al. 2012). The authors searched for genotype-phenotype correlations. While most patients with a 9q34.3 deletion suffered from abnormal speech development, some patients with *EHMT1* mutations were able to communicate with adequate sentences (Willemsen, Vulto-van Silfhout et al. 2012). Interestingly, while all KS patient with a point mutation in the *EHMT1* gene were able to walk, patients with a 9q34.3 microdeletion showed gait difficulties (Willemsen, Vulto-van Silfhout et al. 2012). Generally speaking, patients with deletion sizes larger than 1 Mb showed more severe clinical features compared to patients with smaller deletions or *EHMT1* point mutations (Willemsen, Vulto-van Silfhout et al. 2012) and, patients with *EHMT1* point mutations show similar clinical phenotypes compared to patients with deletions smaller than 1 Mb (Willemsen et al., 2012, Kleefstra, et al 2010). However, only partial

conclusions about phenotype-genotype correlations could be made, as the number of cases reported to date is too low.

Another important aspect of the KS clinical characteristics is the so-called "regressive phenotype". Verhoeven et al. described this phenomenon in two clinical reports addressing characteristic phenotypic information of adults affected with KS (Verhoeven, Kleefstra et al. 2010) (Verhoeven, Egger et al. 2011). The authors described behavioral and neuropsychiatric characteristics of a 53-year-old female and a 59-year-old male affected with KS. The described patients´ core symptoms include irregular circadian cycles and apathy. Patients exhibited sleeplessness at night and somnolence during the daytime (Verhoeven, Kleefstra et al. 2010). In comparison with controls, patients showed reduced purposeful behavior, independence, interest, and curiosity (Verhoeven, Kleefstra et al. 2010). The extent of behavioral and motor problems seemed to increase over time and deterioration became more apparent after adolescence. Neurological examination in one of the patients, including EEG-recording and CT-scanning, revealed no abnormalities. The other patient showed spike-and-wave discharges in the EEG recording, principally located on the left temporal.

Similar cases were reported by Verhoeven et al. in 2011; the authors described the behavioral characteristics of three female patients aged 19, 33, and 43 years (Verhoeven, Egger et al. 2011). During adolescence, individuals gradually lost language, body flexibility, motivation and general performance (Verhoeven, Egger et al. 2011). While all the symptoms above may be considered as an advanced stage in the pathophysiology of KS (Verhoeven, Egger et al. 2011), lack of longitudinal data made it difficult to determine to what extent other adult KS patients could be affected with similar symptoms (Verhoeven, Egger et al. 2011). Also, current information is scarce and insufficient to determine the life expectancy of individuals affected by KS (Kleefstra et al, GeneReviews). While some patients may live until their fifth decade of life, as seen in the previously described patients, death during childhood could occur as the result of heart or respiratory complications (Stewart and Kleefstra 2007).

Mode of inheritance

Most cases of KS are caused by *de novo* mutations in the *EHMT1* gene (Kleefstra, et al 2010). Even though familial cases are rare, there have been reports on parental mosaicism (Willemsen, Beunders et al. 2011). In 2011, Willemsen et al. described for the first time the familial mosaic

transmission of a 9q34.3 deletion. The authors performed fluorescent in situ hybridization (FISH) in one family, as well as multiplex ligation-dependent probe amplification (MLPA) analyses in a second family. Genetic information was obtained from fibroblasts, lymphocytes, and buccal swabs. Mosaic pattern detection showed that both mothers were mosaic carriers of the 9q34.3 deletion (Willemsen, Beunders et al. 2011). Mothers of children affected by KS showed mild learning disabilities and some characteristic KS facial features (Willemsen, Beunders et al. 2011). A recent study showed that 9q34.3 mosaic deletion carriers might be more vulnerable to different psychopathologies, such as Major Depressive Disorder, mood disorders, and ASD (de Boer, Vermeulen et al. 2018). Three adults with different levels of mosaicism were studied using different psychiatric and psychological tests. In addition to the symptoms above, patients manifested limited eye contact, minimal use of emotional gestures, impaired social behavior and limited mental flexibility (de Boer, Vermeulen et al. 2018).

In 2013, Rump et al. described a patient affected by KS that carried a novel mutation causing *EHMT1* aberrant splicing (Rump, Hildebrand et al. 2013). While the proband showed a well-defined heterozygous state, his mother showed different levels of mosaicism in blood cells and oral mucosa (Rump, Hildebrand et al. 2013). Mosaicism at the proband level has also been reported. Hervé et al. reported a patient with an unbalanced translocation between region 5q35 and region 9q34. Interestingly, even though saliva and lymphocytes showed a low to moderate level of mosaicism, the patient manifested typical KS symptoms, developmental delay, behavioral problems, and heart defects.

The *EHMT1* gene and its role in epigenetics

The *EHMT1* gene (NM_024757.4), also known as *GLP*, has an open reading frame of twenty-seven exons (Willemsen, Vulto-van Silfhout et al. 2012). *EHMT1* has a C3HC4 type zinc finger (RING finger) (C3HC4) domain, as well as seven ankyrin repeat domains. Closer to the C-terminal protein region is located the Pre-SET domain and the SET-domain (Willemsen, Vulto-van Silfhout et al. 2012). The ankyrin repeats are composed of thirty-three amino acids and are commonly found in tandem arrays which interact with each other to mediate protein-protein interactions and therefore facilitate molecular recognition (Al-Khodor, Price et al. 2010). The Su(var)3–9, Enhancer-of-zeste, trithorax (SET) domain is responsible for histone methylation. The domain received its name after its identification on three Drosophila melanogaster proteins, for instance,

Suppressor of variegation 3-9 (Su(var)3-9), the Polycomb-group chromatin regulator Enhancer of zeste (E(z)) and the trithorax-group chromatin regulator trithorax (Trx) (Dillon, Zhang et al. 2005).

Histone post-translation modification can influence how chromatin is organized in the nucleus, and thus how associated chromosomal genes are expressed (Dillon, Zhang et al. 2005). Histone methylation serves as a marker for the recruitment of protein complexes, which are responsible for chromatin organization. While histone acetylation is usually associated with open chromatin, and therefore active genes, histone methylation could be associated with actin or repressed chromatin depending on the maker number and position (Dillon, Zhang et al. 2005). To add a methyl group on a histone or other protein, the SET domain transfers a methyl group residue from S-adenosyl-L-methionine to the lysine amino acid of the destined protein, and S-adenosyl-L-homocysteine is produced as a byproduct (Dillon, Zhang et al. 2005). Histone lysine residues can be mono-, di- or tri-methylated. Depending on the lysine residue modified, methylation could lead to repressive heterochromatin, active genes and transcriptional activation, or euchromatic gene silencing (Dillon, Zhang et al. 2005). For example, lysine four histone three (K4H3) methylation is a marker for gene transcription.

The EHMT1 protein mono- and di-methylates the lysine nine of histone three (H3K9me1 and H3K9me2) (Database, GeneCards). Protein EHMT1 histone methylation markers are associated with transcriptional silencing of euchromatin. They play a crucial role in chromatin remodeling and the epigenetic landscape (Willemsen, Vulto-van Silfhout et al. 2012). The EHMT1 protein heterodimerizes with its paralog, protein EHMT2. In mice, the *EHMT2* gene is also known as the *G9a* gene. The EHMT1- EHMT2 heterodimeric protein complex writes the H3K9me2 mark, a repressive mark commonly found at the promoter of silenced genes. GLP protein-deficient mice exhibit similar phenotypes to the G9A-deficient mice. Similarities include embryonic lethality and reduction of H3K9me1 and H3K9me2 levels. (Tachibana, Ueda et al. 2005). Interestingly, even though GLP protein deficiency is lethal in mice, it is possible to detect early-stage embryos at E9.5. These embryos exhibited phenotypic anomalies when compared to same stage wild-type embryos. For instance, GLP-deficient embryos showed severe growth retardation and high reduction in H3K9me1 and H3K9me2 marks. A similar phenotype is found in G9A-deficient

early-stage embryos, pointing to a common role of the G9A and GLP proteins in embryonic development (Tachibana, Ueda et al. 2005).

The GLP/G9A complex is critical during brain development and cell differentiation (Benevento, van de Molengraft et al. 2015). Methylation on histone 3 has a reported role in memory formation and neural plasticity (Benevento, van de Molengraft et al. 2015). The H3K9me2 mark has a role in long-term memory, as well as fear memory formation (Benevento, van de Molengraft et al. 2015). In 2012, Gupta-Agarwal et al. reported increases of the H3K9me2 levels in the murine hippocampus and entorhinal cortex after a contextual fear conditioning test. Similar increases were seen in the lateral amygdala in response to auditory fear conditioning, suggesting a critical role of the G9a/GLP synaptic plasticity and memory formation (Gupta-Agarwal, Franklin et al. 2012).

Another important function of the GLP and G9A proteins on mice is their ability to bind to the same histone marks that they write. This affinity is mediated by the ankyrin repeat protein domains (Collins, Northrop et al. 2008). This histone mark reading function is critical for epigenetic regulation (Kramer 2016). It has been shown that the ankyrin mediated interaction to premethylated H3K9 nucleosomes stimulates the recruitment and action of GLP and G9A (Liu, Zhang et al. 2015). Mice containing mutations that reduce the ability of GLP to bind to H3K9me1 and H3K9me2 exhibited congenital problems and died soon after birth (Liu, Zhang et al. 2015). Interestingly, there is a report of a patient affected with KS whose *EHMT1* point mutation is found at the ankyrin repeat domains. Amino acid substitution on these repeats led to aberrant protein folding, and thus altered protein affinity and function (Blackburn, Tischer et al. 2017), causing the characteristic KS phenotype.

The GLP protein plays a crucial role in the developing mouse brain. The *Glp* gene is expressed in all the central nervous system during mouse embryogenesis, and its expression is restricted to particular adult brain structures, these are the piriform cortex, the ventricular wall, the olfactory bulb and the hippocampus (Kleefstra, Smidt et al. 2005). *Glp* gene expression levels decrease after birth, as was observed in young wild-type mice where GLP protein levels decrease by 10-fold within the first post-natal month (Balemans, Kasri et al. 2013). The strong decrease in protein levels after birth suggests that GLP plays a role in the developing brain. A similar mechanism is observed in humans after birth (Balemans, Kasri et al. 2013).

Animal models of KS

Drosophila model

While mammals have two *EHMT* paralogs (*EHMT1* and *EHMT2*), flies only have a single ortholog (Kramer, Kochinke et al. 2011). Similar to mammals, in flies, the Ehmt protein is an essential player in the epigenetic regulation of cognition and memory (Kramer, Kochinke et al. 2011). Kramer et al. deleted the *Ehmt* gene in flies and looked for changes in epigenetic markers (Kramer, Kochinke et al. 2011). Loss of the *Ehmt* gene has an impact on the dendrite development of sensory neurons, resulting in a reduced number of dendrite ends. Interestingly, this phenotype is also observed in GLP-deficient mice, indicating that the effect is conserved across species (Balemans, Kasri et al. 2013). Also, *Ehmt1*-deficient flies showed apparent difficulties crawling compared with wild-type flies, as well as deficits in different forms on learning, specifically in habituation and courtship memory (Kramer, Kochinke et al. 2011). This phenotype resembles the ID phenotype observed in patients with KS. Similarly to what has been observed in mammals, *Ehmt1*-deficient flies have an H3K9me2 dysregulation at discrete euchromatic loci (Kramer, Kochinke et al. 2011) (Tachibana, Ueda et al. 2005) (Balemans, Ansar et al. 2014).

Mouse model

The *EHMT1* human gene is a highly conserved gene across mammals. For instance, *Glp* the murine ortholog of *EHMT1*, has 97% similarity to the former one in the amino acid sequence (Balemans, Huibers et al. 2010). The *Glp* similarity to the human *EHMT1* gene provides an excellent framework for the characterization of a murine model of KS. In 2010, Balemans et al. reported a heterozygous *Glp* knockout mouse model of KS. Mice were evaluated in different behavioral and motor tests to assess resemblance to KS human phenotype. $Glp^{+/-}$ mice exhibited reduced exploration and increased anxiety under different behavioral tests; for instance, environment exploration results were lower compared to controls (Balemans, Huibers et al. 2010). Furthermore, juvenile $Glp^{+/-}$ mice showed decreased social play and diminished preference for establishing novel social interactions. On the other hand, $Glp^{+/-}$ mice did not show motor difficulties in comparison with wild-type animals (Balemans, Huibers et al. 2010). Therefore, while this model resembles the autistic-like features of KS, it seems that the model failed to resemble the characteristic hypotonia and motor difficulties seen in KS patients.

To further characterize the *Glp*$^{+/-}$ mouse model, Balemans et al. performed various memory and learning tests. In fear conditioning experiments using the startle response test with foot shock stimulus, *Glp*$^{+/-}$ mice showed more time freezing, in comparison with wild-type mice (Balemans, Kasri et al. 2013). This response was kept longer during the extinction phase, demonstrating increased sensitivity to external stimuli and learning impairments (Balemans, Kasri et al. 2013). Object and spatial recognition abilities were also impaired in *Glp*$^{+/-}$ mice compared to wild-type mice (Balemans, Kasri et al. 2013). Morphological examination of neurons of *Glp*$^{+/-}$ mice showed abnormalities in the dendritic arborization. Specifically, the CA1 pyramidal neurons showed reductions in the number of endings, total surface, and branch points (Balemans, Kasri et al. 2013). Neuron abnormalities may be associated with deficits in presynaptic release probability, as seen through voltage clamp experiments, which may explain to some extent the learning disabilities seen in the KS mouse model (Balemans, Kasri et al. 2013).

Balemans et al. continued to characterize the phenotype observed in the KS mouse model, with a focus on the core KS features, including ID, hypotonia and characteristic facial features observed in patients (Balemans, Ansar et al. 2014). *Glp*$^{+/-}$ mice showed postnatal development delay when compared to controls; for instance, the *Glp*$^{+/-}$ mice exhibited delayed ear and eye-opening, as well as delayed upper incisor eruption (Balemans, Ansar et al. 2014). The *Glp*$^{+/-}$ mice also exhibited unstable waking and difficulties in holding a bar for as long as the wild-type mice, pointing to hypotonia. Concerning the characteristic facial features observed in KS patients, *Glp*$^{+/-}$ mice exhibited a similar reduction in skull length, nasal, frontal and parietal bones, as well as more prominent frontal bones and a higher distance between the left and right anterolateral corners of the frontal bone. These cranial characteristics resemble the brachycephaly and hypertelorism observed in KS patients (Balemans, Ansar et al. 2014). In mice, *Glp* regulates the expression of different genes associated with bone tissue formation. Interestingly, a substantial reduction of H3K9me2 is observed around the transcription start sites of those genes, suggesting that H3K9me2 dysregulation caused by GLP protein deficiency has a direct effect on gene expression and bone formation (Balemans, Ansar et al. 2014). These molecular mechanisms offered an excellent model to explain the phenotype observed in KS patients.

Disadvantages of using animal models in the study of KS

A detailed phenotypic description of KS has been achieved using animal models allowing us to uncover various molecular mechanisms of the disease. Animal models of KS have undoubtedly facilitated the study of the disease and advanced our understanding of its pathophysiology. Mouse and fly models present many advantages to modeling a disease, such as high reproductive rate, large litter sizes, and accelerated life spans. However, they also present many limitations in uncovering some of the disease mechanisms. Murine cellular physiology, even though close to human physiology, still presents many limitations to study particular human cellular characteristics and human cell-specific phenotypes (Chadman, Yang et al. 2009).

Furthermore, animal models are incapable of resembling patients' genetic background, which is critical to studying the relationship between patient-specific phenotype and the genetic basis of the disease (Deng 2017). This is an issue in the study of KS because the disorder could originate from a variety of genetic mutations. Genotypic variation of the disease generates a wide range of phenotypes (Willemsen, Vulto-van Silfhout et al. 2012), which would be complicated to model in an animal model. Even though previous KS characterization and studies in animal models have allowed for the advancement of our understanding of the disease, we still lack a detailed molecular and cellular understanding of its molecular processes, making it necessary to create new ways to study the disease.

B. Human-induced Pluripotent Stem Cells

Recent fast-paced progress in the stem cell field has positioned stem cells as a key tool for understanding disease etiology and its molecular processes, as well as for therapeutic applications (Singh, Kalsan et al. 2015). Stem cells provide an excellent model to better understand complex molecular mechanisms seen in many diseases, as it is the case for KS. Stem cells are undifferentiated cells that have self-renewal ability and can give rise to multiple lineages. Stem cells are classified into two main groups: somatic stem cells (SSCs) and embryonic stem cells (ESCs). While SSCs are found in many adult tissues, ESCs are mainly found in the Inner Cell Mass (ICM) of fertilized embryos. SSCs are only able to differentiate into a restricted cellular lineage, whereas ESCs can differentiate into the three primary germ layers (Kanherkar, Bhatia-Dey et al. 2014).

History and development of cellular reprogramming

Throughout history, research has expanded our knowledge of the embryonic stem cell state and the cellular differentiation process. The discovery and isolation of ESCs motivated the search for artificial pluripotency induction (Kanherkar, Bhatia-Dey et al. 2014). The creation of induced pluripotent cells was the result of numerous scientific discoveries. Three main findings led to the creation of induced Pluripotent Stem Cells (iPSCs): 1) reprogramming by nuclear transfer, 2) the establishment of culture conditions that enable the maintenance and study of pluripotent lines and 3) the discovery of transcription factors as key regulators of the cell fate.

Reprogramming by nuclear transfer

In 1955, Briggs and King transferred nuclei isolated from late-stage embryos into enucleated oocytes (Briggs and King 1952). Years later, Gurdon et al. transferred nuclei from keratinized skin cells of adult Xenopus into enucleated frog eggs. They were able to obtain swimming tadpoles, showing that cell differentiation does not involve any permanent genetic change (Gurdon, Laskey et al. 1975). Similar results were observed in mammals with the creation of Dolly; Ian Wilmut et al. reported the birth of the first mammal created by somatic cloning of mammary epithelial cells (Wilmut, Schnieke et al. 1997). These findings have proven that cellular differentiation does not involve any irreversible changes and that oocytes contain factors that can reprogram adult somatic cells.

Pluripotent cell lines culture

In 1981, Evans et al. established a line of pluripotent cells from mouse embryos (Evans and Kaufman 1981). Similarly, Thomson et al. were able to establish embryonic stem cell lines derived from human blastocysts (Thomson, Itskovitz-Eldor et al. 1998). Since then, different conditions have been established to inhibit the differentiation of pluripotent embryonic stem cells and therefore, maintain pluripotency in culture conditions. For instance, Austin et al. identified the leukemia inhibitory factor as a regulator of long-term maintenance of pluripotency (Smith, Heath et al. 1988). Likewise, optimal culture conditions for the maintenance of human ESCs were established.

The discovery of lineage-specific transcription factors

This stream was based on the discovery of lineage-specific transcription factors which help to establish and maintain specific cellular identities. These so-called master transcription factors drive the expression of lineage-specific genes and suppress the expression of other genes (Stadtfeld

and Hochedlinger 2010). Different research groups found that specific lineage-specific transcription factors can change cell fate when ectopically expressed in other cell lines. Davis and Weintraub were pioneers in this field; they described the reprogramming of murine fibroblasts to myoblasts. The reprogramming was achieved through the transfection of the skeletal muscle gene *MyoD* (Davis, Weintraub et al. 1987). More recently, Xie and colleagues reprogrammed B cells into macrophages by enforced expression of the genes *Ebpα* and *Ebpβ* and described some of the involved molecular mechanisms to achieve reprogramming (Xie, Ye et al. 2004).

All these findings led Yamanaka and colleagues to hypothesize that a particular combination of transcription factors in ESCs allows for reprogramming of adult somatic cells into a pluripotent state (Yamanaka 2012). Takahashi and Yamanaka designed a genetic approach to assess the effects of various transcription factors in the induction of pluripotency. In this genetic test, the pluripotent state was monitored through the expression of the *Fbx15* gene, which is an ESC-specific gene. The authors inserted a neomycin resistance gene in the *Fbx15* gene locus and tested different combinations of candidate genes in mouse embryonic fibroblasts. Drug-resistant colonies were not observed when the murine genes *Oct3/4* and *Klf4* were removed from the mix. While removing the *Sox2* gene resulted in a decrease of pluripotent colonies, the absence of the gene *c-Myc* did not affect colony number. These observations led to the identification of the genes *Oct3/4*, *Klf4*, *Sox2* and *c-Myc* as key players in pluripotent induction (Takahashi and Yamanaka 2006). IPSC colonies generated with this set of core transcription factors expressed pluripotent characteristic markers, such as the proteins SSEA-1 and NANOG, and can generate teratomas in immunocompromised mice. Since then, the genes *Oct3/4*, *Klf4*, *Sox2*, and *c-Myc* have been known as the Yamanaka factors (Stadtfeld and Hochedlinger 2010). A year later, the same group reported the creation of iPSC lines from adult human fibroblasts (Takahashi, Tanabe et al. 2007). Interestingly, on the same day, another group reported the creation of human iPSCs using a different combination of human genes *OCT3/4, SOX2, NANOG* and *LIN28* (Yu, Vodyanik et al. 2007).

Initial methods to induce pluripotency relied on retroviral transduction and the use of modifications into the host genome to monitor pluripotency markers' expression (Takahashi and Yamanaka 2006). Moreover, iPSCs colonies were produced at a very low efficiency; for instance, in the initial Yamanaka's report, only 0.02% of the transfected cells resulted in iPS cells at 14 to 21 days post-transduction (Takahashi and Yamanaka 2006). Since the initial discovery phase in the iPSCs field,

many researchers groups have developed different methods to increase the efficiency of reprogramming, including new reprogramming factor expression vectors and new ways to deliver them into the cell without modifying the genome (Malik and Rao 2013). Since then, iPSCs have been generated from many species, including humans, rats, mice, and monkeys, using the four Yamanaka factors (Stadtfeld and Hochedlinger 2010), showing that induced pluripotency mechanisms are conserved across evolution. Furthermore, iPSCs have not only been derived from fibroblasts adult cells but also from many somatic cell lineages, such as lymphocytes, neurons, and keratinocytes (Stadtfeld and Hochedlinger 2010).

Reprogramming Methods

Integrating virus

The use of lentivirus is one of the most efficient methods to achieve pluripotency. These viruses integrate into the host genome and deliver the pluripotency factor cassettes and help to achieve stable expression of the reprogramming factors. Once pluripotency is achieved, viral sequences are silenced, and cells stop relying on the viral reprogramming factors and begin endogenously expressing the pluripotency genes (Wang and Na 2011).

Even though the use of lentivirus vectors has been shown to be very efficient, the method still has many disadvantages. For instance, there have been reports of reprogrammed cells whose viral genes were not properly silenced; after long periods of cell culture, these cells may exhibit viral gene expression (Wang and Na 2011). Reactivation of the gene *c-Myc* has been associated with increased chances to develop tumors in mice derived from integrating viruses-derived iPSCs (Okita, Ichisaka et al. 2007). Thus, reprogramming using lentivirus vectors is not ideal for clinical applications and makes it necessary to develop transgene-free reprogramming approaches (Wang and Na 2011).

Non-integrating virus

Non-integrating viruses offer an alternative to reprogramming methods that generate mutations in the genome. Different research groups have shown that inducing iPSCs with non-integrating viruses, i.e., adenovirus and Sendai virus, can lead to stable, karyotypically normal iPSCs lines (Zhou and Freed 2009) (Chen, Fukuda et al. 2013). Also, iPSC induction using non-integrating viruses has been achieved in several cell lines, like fibroblasts, peripheral blood cells, and heart cells (Miyamoto, Akiyama et al. 2018).

Adenovirus

While adenoviruses offer an integration-free reprogramming method, they also have important disadvantages, for instance, adenovirus-mediated reprogramming has low reprogramming efficiency. Reported reprogramming efficiency has been observed between 0.001–0.0001% in mouse cells, and only 0.0002% efficiency in human cells (Wang and Na 2011). Moreover, adenoviruses stop expressing in the host after cell division, forcing researchers to repeat adenovirus treatment on subsequent days (Zhou and Freed 2009).

Sendai virus

In contrast to the DNA Adenoviruses, Sendai viruses are RNA viruses and thus do not enter the host nucleus. After transfection, Sendai viruses usually dilute out of the cells before ten cellular passages. Reprogramming efficiencies have been reported as reaching up to 1% in fibroblasts (Wang and Na 2011). An enormous advantage of using Sendai virus for inducing pluripotency is that it only requires a modest initial number of cells, around $1x10^4$, allowing researchers to achieve reprogramming in less time, in comparison with other methods (Beers, Linask et al. 2015). However, Sendai viruses are expensive and potentially cytotoxic, limiting their use in the laboratory (Beers, Linask et al. 2015).

Non-viral reprogramming methods

mRNAs

Reprogramming factors can be delivered into the cells without using viruses. Synthetic mRNAs encoding the reprogramming factors could be introduced into somatic cells and induce pluripotency. Once inside the cytoplasm, synthetic mRNAs are translated in the ribosomes. No genetic change is produced after or during their use. However, synthetic mRNA is highly unstable, and thus requires repeated introduction for several days (Kim, Choi et al. 2011). Another major drawback of synthetic RNA is its cytotoxicity. Exogenous single-stranded RNA could trigger an innate immune response. Warren et al. circumvented this problem using interferon inhibitor as media supplementation and introducing RNA modifications at the nucleotide level (Warren, Manos et al. 2010).

microRNAs

Lin et al. reported pluripotency induction of skin cancer cells using microRNAs. A retroviral delivery approach was used to deliver the miR302 cluster into cells. Overexpression of the miR302

led to iPSCs colonies that met all characteristics of pluripotency (Lin, Chang et al. 2008). Later on, Miyoshi et al. used mature double-stranded microRNAs to induce pluripotency in mouse and human cells. A combination of miR200c, miR302, and miR369 was successfully employed to achieve reprogramming (Miyoshi, Ishii et al. 2011). One main disadvantage of using microRNAs to induce pluripotency is the lack of scientific knowledge about the miRNA networks and its mechanisms. It is difficult to predict whether the employed microRNAs may have unexpected adverse off-target effects (Wang and Na 2011).

PiggyBac transposon

PiggyBac is a transposon which can integrate into the genome, specifically in TTAA sites. Once the transposon is in the genome, the gene cassette that it carries can be read and translated. One major advantage of this system is that transposons can carry long cassettes, allowing for the introduction of the four Yamanaka factors at once. When transposon activity is no longer useful, they can be excised from the genome by expressing a transpose. As a result, the created iPSCs are transgene-free. However, transposon reprogramming efficiencies are very low, which have limited their use in research (Malik and Rao 2013).

Proteins

Recombinant proteins have been created in bacterial and human cells. Zhou et al. used *E. coli* to synthesize the Yamanaka proteins, purified and isolated them. Purified proteins were introduced into mouse cells, and with the help of small molecules, Zhou et al. were able to create iPSCs that met all the characteristics of pluripotency (Malik and Rao 2013). With regards to human cells, Kim et al. overexpressed the Yamanaka proteins in HEK293 cell lines and isolated recombinant proteins that were later used for inducing pluripotency in human newborn fibroblasts (Kim, Kim et al. 2009). Even though large amounts of recombinant proteins can be generated in the laboratory, this method has low efficacy results and is labor intensive, as several protein transductions are needed to achieve pluripotency.

Episomal plasmids

Episomal vectors have auto-replication abilities, allowing for the prolonged expression of the reprogramming factors in host cells. Using oriP/EBNA1 (Epstein-Barr nuclear antigen-1)-based episomal vectors, Yu et al. created iPSC lines free of vectors and transgene sequences, with an efficiency of around 0.0006%. Reprogramming was achieved using the canonical human

Yamanaka factors, *OCT3/4, SOX2, C-MYC,* and *KLF4,* in addition to *NANOG* and *LIN28*, all of which were cloned in oriP/EBNA1 vectors (Yu, Hu et al. 2009). Even though reprogramming efficiencies are low, when cells are cultured with defined E8 media, efficiency increases, depending on the somatic lineage transfected, from 0.006 to 1% (Malik and Rao 2013). Drozd et al. created iPSCs from urinary epithelial cells using the oriP/EBNA-1-based episomal vectors (Drozd, Walczak et al. 2015). Overall episomal vectors offer a good model to induce pluripotency in a wide variety of cell lineages. Moreover, when combined with small molecules this method can have better yields (Malik and Rao 2013).

Mechanisms underlying iPSC formation

Transcriptional and epigenetic changes are produced during the differentiation process. A large dynamic change in the transcriptional network, including not only RNAs but also microRNAs, is produced as a result of de-differentiation from somatic cells to iPSCs (Gonzalez and Huangfu 2016). Similarly, the epigenetic landscape plays a crucial role in the reprogramming process; DNA methylation and histone-specific methylation interact together with different transcription factors to achieve proper gene expression during this process (Kim, Choi et al. 2011). Within this large dynamic network, the Yamanaka factors and other pluripotency-related transcription factors play crucial roles in cellular identity, modeling changes in the transcriptional and epigenetic network.

For instance, the proteins OCT3/4, NANOG, SOX2, and KLF4 help activate pluripotency genes and repress somatic-related genes. These transcription factors help recruit the necessary epigenetic machinery to silence somatic genes and thus repress differentiation to somatic lineages. In this context, the Polycomb chromatin remodeling complex plays a crucial role in the epigenetic remodeling of somatic related genes. Polycomb catalyzes tri-methylation of lysine twenty-seven on histone three (H3K27me3), a histone mark that promotes gene expression silencing, and thus leads to gene repression of lineage-associated genes (Boyer, Plath et al. 2006).

Gene expression analysis in murine induced pluripotent stem cells showed that the reprogramming process is a multi-step process, including three main phases: initiation, maturation, and stabilization (Samavarchi-Tehrani, Golipour et al. 2010). The initiation phase is characterized by mesenchymal-to-epithelial transition. Whereas mesenchymal murine genes, *Slug, Zeb1/2*, and *Snai1/2*, are repressed, epithelial genes are upregulated, such as *Crb3, Cdh1*, and *Cldn3*. Early pluripotency genes, such as gene *Fbxo15*, are also upregulated at the initiation phase (Gonzalez

and Huangfu 2016). In mice, the initiation phase is also characterized by BMP signaling induction of several microRNAs, including the miR200 family and miR205 (Samavarchi-Tehrani, Golipour et al. 2010).

During the initiation phase of the reprogramming process, the C-MYC protein plays a role in regulating cell growth and proliferation, which is achieved through precise coordination of the Cyclin-dependent kinase (CDK) inhibitors and the cell cycle checkpoint factors (Meyer and Penn 2008). On the other hand, the KLF4 protein is present during the initiation of reprogramming, as well as in late stages. In the early stage, KLF4 helps to repress gene expression of somatic-related genes, while in late stages, it promotes gene activation of pluripotent-related genes, such as *Oct3/4* and *Klf5* (Gonzalez and Huangfu 2016). It has been proposed that pluripotency factors may participate in positive feedback loops to increase the expression of their genes and other pluripotency-related genes; these feed-forward loops are established when most core pluripotency factors are expressed at the end of the stabilization stage.

Mature iPSCs express pluripotency factors in a manner independent of transgene factor expression. The maturation stage is characterized by elongation of telomeres, reactivation of the silenced X chromosome in female cells and significant changes in DNA methylation, including remodeling of 5-methylcytosine marks (Stadtfeld and Hochedlinger 2010). IPSCs acquire specific epigenetic changes related to pluripotency which are different from somatic states and partially reprogrammed cells.

Characteristic epigenetic states in pluripotency include open and accessible chromatin and scattering heterochromatin foci (Gonzalez and Huangfu 2016). In humans, pluripotency-related gene promoters, such as *OCT3/4* and *NANOG* promoters are hypo or unmethylated, while somatic-related gene promoters are highly methylated (Gonzalez and Huangfu 2016). Several genes are essential for maintenance of pluripotency, such as the human genes *NANOG*, *DNMT3b*, *EST2*, and *ZFP42*, which play crucial roles in this process (Shi, Wang et al. 2006). The mitochondria play an important role in the establishment and homeostasis of pluripotency. The acidic zinc finger gene *ZFP42* plays a role in mitochondrial regulation in pluripotency. The ZFP42 protein promotes mitochondrial fission through phosphorylation of DRP1, which ultimately leads to a metabolic change towards glycolysis, a metabolic state characteristic of stem cells (Son, Choi et al. 2013).

Pluripotent stem cells have high gene expression levels of the human Telomerase Reverse Transcriptase (*HTERT*) gene, also known as *EST2*. *HTERT* is part of the telomerase enzymatic complex and helps to maintain telomere length. When cells start to differentiate, the *HTERT* gene is downregulated through a process of progressive promoter methylation. At the end of the differentiation process, the *HTERT* promoter is fully methylated. Inversely, pluripotency induction is accompanied by upregulation of *HTERT* expression (Ramlee, Wang et al. 2016). Limited telomerase activity in differentiated cells results in telomere length shortening, which limits proliferative capacity.

Undifferentiated stem cells have a characteristic glycosylation profile that distinguishes them from other lineages. While in the pluripotent state N-glycans are downregulated, new structures emerge during the differentiation process. One of the most characteristic glycosylation marks in pluripotent cells is fucosylation (Draper, Pigott et al. 2002). The unique glycosylation profile of the pluripotent cells makes it an excellent target to identify and isolate fully reprogrammed iPSCs. In this sense, the surface antigen markers offer great advantages to monitor the reprogramming progress (Abujarour, Valamehr et al. 2013). Antigens used for this purpose include those that are associated with carbohydrate epitopes, those linked to glycolipids, such SSEA4, and some which are associated with glycoproteins, as is the case of TRA-1-60 (Natunen, Satomaa et al. 2011).

Stage-specific embryonic antigen-4 (SSEA-4) is an epitope on a related glycosphingolipid, GL-7. This glycosphingolipid is a carbohydrate chain linked to ceramide found in the cell membrane. SSEA-4 specifically recognizes the terminal sialic acid on GL-7 (Brimble, Sherrer et al. 2007). SSEA-4 is widely expressed in embryonic stem cell and iPSCs, given that this antigen is downregulated during differentiation; it represents a good marker of pluripotency (Kannagi, Cochran et al. 1983). Similarly, the keratan sulfate TRA1-60 is a cell surface antigen that recognizes a carbohydrate epitope; TRA1-60 is routinely used to assess pluripotency (Natunen, Satomaa et al. 2011).

Assessment of iPSC differentiation potential

Rapid advances in iPSC development called for standard and efficient methods to characterize them. Research efforts shifted from new technologies to create iPSCs to the development of high-throughput and cost-effective tools for iPSC standardization (Fergus, Quintanilla et al. 2016). Common quality control practices include iPSC colony morphology assessment and monitoring

of the expression of markers of pluripotency and differentiation potential. In the past years, the ability of iPSCs to form teratomas established itself as a standard technique used to assess differentiation potential. In this technique, iPSCs are injected in mice to confirm the presence of the three germ layers in the tumor. This technique is complex and labor-intensive, as it requires extensive histological and immunohistochemical analysis of developed tumors (Gertow, Przyborski et al. 2007). Furthermore, teratoma formation is a lengthy process that usually requires 1 to 2 months. Technical difficulties associated with previous methods to standardize iPSC lines led to the development of new techniques for this purpose.

New animal-free high-throughput molecular techniques were developed by different research groups (Muller, Brandl et al. 2008). One of the main efforts in this area was the development of the PluriTest™. It is a molecular technique based on RNA sequencing analysis of self-renewal and lineage markers (Muller, Brandl et al. 2008). Another good technique for this purpose is the TaqMan® hPSC Scorecard Panel. This technique is also based on gene expression, but in contrast to the PluriTest, the Scorecard Panel is a RT-PCR-based assay. Pluripotency and differentiation potential is assessed through comparisons to the reference transcriptome information of several stem cell and iPSC lines. The Scorecard is a good tool to rapidly assess pluripotency and differentiation potential to the three germ layers (Bock, Kiskinis et al. 2011). However, it can be difficult to afford by many small laboratories. Besides, the Scorecard is not optimized for early germ layer differentiation and requires customized downstream computational analysis. Software to analyze the TaqMan RT-qPCR results is cloud-based - users do not have access to the code, making it difficult to customize the plate analysis for specific-needs (Tsankov, Akopian et al. 2015).

The creators of the TaqMan hPSC Scorecard Assay published a more affordable method to assess differentiation potential of iPSCs (Tsankov, Akopian et al. 2015). They described a qPCR assay based on the expression measurements of 96 genes. The authors published a list of selected genes to evaluate each germ layer. Also, they published a computational approach to analyze gene expression signatures that can be customized to various plates and combinations of genes (Tsankov, Akopian et al. 2015). Tsankov et al.'s list includes markers such as the Aldehyde Dehydrogenase 1 Family Member A1 (*ALDH1A*), which is a common marker of stem cell populations. The ALDH1A protein activity is associated with differentiation properties that define

stem cells (Ozbek, Calik et al. 2012), specifically in the ectodermal lineage (D'Antonio, Woodruff et al. 2017). Similarly, Tsankov et al.'s list included the protein Collagen Type I Alpha 1 Chain (COL1A) and the Aniridia Type II protein, also known as PAX6, both proteins were reported as excellent ectodermal markers.

Regarding mesodermal markers, among other proteins, the authors reported the human protein Insulin Like Growth Factor 2 (IGF2) as an excellent marker of mesoderm formation (Morali, Jouneau et al. 2000). During teratoma formation, expression of mesodermal markers is correlated with an increase of endogenous IGF2 levels (Morali, Jouneau et al. 2000). Likewise, upregulated gene expression of the Bone Morphogenetic Protein 4 (BMP4) has been observed during mesoderm formation (Zhang, Li et al. 2008). Induction of the *BMP4* gene can initiate mesoderm formation in human embryonic stem cells (Zhang, Li et al. 2008). Similarly, the Platelet Derived Growth Factor Receptor Alpha gene, also known as *PDGFRA*, is expressed in mesoderm during mouse embryogenesis (Sakurai, Era et al. 2006). The PDGFRA protein is used to isolate mesodermal progenitor cells (Zhang, Afink et al. 1998).

For the endodermal lineage, the authors included protein markers such as Cadherin 1 (CDH1), SOX7 and GATA6. *CDH1* gene expression has been observed in the endoderm lineage in various cell lines (D'Antonio, Woodruff et al. 2017). Similarly, the *SOX7* gene is expressed in the embryonic endoderm. Several reports show it is possible to establish endoderm progenitors by *SOX7* expression in human embryonic stem cells (Seguin, Draper et al. 2008). Likewise, *GATA6* gene expression is essential for endodermal formation, supported by a dynamic *GATA6* expression in early and mature endodermal formation (Cai, Capo-Chichi et al. 2008).

IPSC genomic integrity

Different studies have analyzed the presence of chromosomal abnormalities in iPSCs. In general, fibroblast-derived iPSCs do not show gross chromosomal abnormities upon G-banded karyotyping (Martins-Taylor and Xu 2012). However, some abnormalities have been reported, especially in iPSC lines that have been cultured for long periods of time (Martins-Taylor and Xu 2012). Among the "common" chromosomal abnormalities observed in iPSCs, there is trisomy of chromosomes twelve and eight (Martins-Taylor, Nisler et al. 2011). Besides, there have been reports of a recurrent chromosomal abnormality in chromosome twenty. It consists of amplification of 2.5 to 4.6 Mb at the chromosomal region 20q11.21, which has been related to oncogenic issues (Lefort,

Feyeux et al. 2008). The presence of these mutations on iPSC lines could generate problems during the reprogramming process and affect the study's results (D'Antonio, Woodruff et al. 2017). Therefore G-banded karyotype is a common practice in the iPSC field (D'Antonio, Woodruff et al. 2017). G-band karyotype is a cytogenetic technique that allows for staining-facilitated visualization of condensed chromosomes and their subsequent karyotyping. Chromosomal rearrangements of five Mb or more can be detected with this technique (D'Antonio, Woodruff et al. 2017). The G-band karyotype results serve to distinguish iPSCs best suited for research and therapeutic applications.

Therapeutic potential of iPSCs

Human iPSCs (hiPSC) have an enormous potential to be used in basic science and translational research. In the field of basic science, the use of hiPSCs is helping us to expand our knowledge of the molecular mechanisms involved in lineage differentiation (Yamanaka 2012). Regarding translational research, hiPSCs have been used to model patient-specific mutations in multiple genetic disorders. Also, hiPSCs have been employed in sophisticated drug discovery strategies and cytotoxicity screenings, examples of such strategies are described below (Yamanaka 2012).

Disease Modeling

Traditional human disease research has been performed in various models, including but not limited to animal models, immortalized cells and patient-derived cell lines. Even though these approaches have helped advance our understanding of disease pathology, they also present many disadvantages in the study of the disease molecular mechanisms. For instance, immortalized cell lines usually exhibit several chromosomal and genetic abnormalities, which could hinder data interpretation and biological meaning. Besides, immortalized lines do not provide insights on patient-specific genetic background (Ebert, Liang et al. 2012). Likewise, using patient samples as a platform for drug discovery is difficult, as these cells have limited expansion potential. Furthermore, obtaining cell lines from tissues that are difficult to access, such as neurons or cardiomyocytes, may require surgery, and in some cases, cell lines can only be obtained after a patient's death.

HiPSCs circumvent these disadvantages and provide a good model for studying cellular and molecular mechanisms of a disease (Varga, Nemes et al. 2016). Because patient-derived hiPSCs

have the same genetic background as the patient, it is easier to control for the individual's genetic makeup, which can affect response cytotoxic and treatment response. Moreover, iPSCs have the ability to self-renew indefinitely alongside the potential to give rise to many cellular lineages.

In the past years, researchers were able to model monogenic and complex diseases in iPSCs and have been able to differentiate them to many lineages, including different kinds of neurons, glial cells, cardiomyocytes, hepatocytes, hematopoietic cells, insulin- and glucagon-producing cells, muscle cells, among others (Ebert, Liang et al. 2012). In 2008 Dimos et al. reported, for the first time, neurons derived from iPSCs. These neurons were crucial in the study of amyotrophic lateral sclerosis (ALS) (Dimos, Rodolfa et al. 2008). Since then, different iPSC-derived neuron lines have been created to study the pathophysiology of many neurological diseases, like Alzheimer's disease (Sullivan and Young-Pearse 2017), Parkinson's disease (Nguyen, Byers et al. 2011) and schizophrenia (Windrem, Osipovitch et al. 2017). In the case of Parkinson's disease, Nguyen et al. created a line of patient iPSC-derived dopaminergic neurons. These neurons had a common mutation in the Leucine-Rich Repeat Kinase-2 (*LRRK2*) gene. Interestingly, the neurons showed increased susceptibility to oxidative stress and increased levels of α-synuclein protein. These traits resemble the disease pathology observed in humans, thus making them good targets for the study of potential treatments (Nguyen, Byers et al. 2011).

Platform for drug discovery

Given that hiPSCs have the potential to be successfully differentiated into many lineages and simulate disease phenotypes, they provide a good platform for testing drugs in small and high throughput settings. One great advantage of using iPSCs is that cell lines could be generated from patients with different mutations and disease severities (Ellis and Bhatia 2011). The use of hiPSCs could help us to study and control intra-species variation (Ellis and Bhatia 2011). Besides, the use of hiPSCs overcomes ethical concerns regarding the use of human embryos (Yamanaka 2012).

Several research groups have generated iPSC models of various monogenic and complex diseases. These iPSC lines have been used for drug screening assays. The field of ID disorders has not been an exception. In 2012 Marchetto et al. published an iPSC model of Rett syndrome and its use for drug testing. Rett syndrome hiPSC-derived glutamatergic neurons exhibited clear biomarkers that were manipulated to assay to rescue the disease phenotype. The authors observed a reduction in synaptic number, as well as a reduction of the frequency and amplitude of synaptic currents

compared to controls. Treatment with IGF1 helped to increase the synaptic number. Similarly, treatment with gentamicin increased full-length MECP2 protein production. Even though the particular doses and timing required to treat humans safely are still unknown, this study is a good example of how iPSCs could be used for drug discovery (Marchetto, Carromeu et al. 2010).

High-throughput drug screening assays could be done using iPSC-derived lines; Kaufmann et al. reported the use of a Fragile X syndrome iPSC line to assess 50,000 compounds in a large drug screening assay (Kaufmann, Schuffenhauer et al. 2015). They assessed modulation of expression of the *FMR1* gene, which in the disease state is silenced. To achieve this, the authors employed various chemical compounds, including epigenetic regulators that were hypothesized to restore *FMR1* expression. They measured *FMR1* expression in several cellular locations, including the nucleus, ring region, cytoplasm and the whole cell area. The authors found a histone deacetylase inhibitor that increased *FMR1* gene expression levels (Kaufmann, Schuffenhauer et al. 2015).

Similarly, a High-Throughput drug screening assay was reported in *SHANK3* haploinsufficiency syndrome, a syndromic form of autism spectrum disorder (Darville, Poulet et al. 2016). A total of 202 compounds were tested, including FDA-approved drugs, epigenetic regulators, and modulators of neurogenesis. The two main readouts of the study were SHANK3 protein levels and SHANK3 protein location in glutamatergic synapses. Valproic acid and lithium were identified as good modulators of SHANK3 activity. Lithium was administrated to a *SHANK3*-deficient patient for one year. After this period, the patient showed an improvement in his autistic core symptoms as evaluated using the social responsiveness scale and the autism diagnosis observational scale (Darville, Poulet et al. 2016). However, the patient also developed an attention deficit hyperactive disorder. In-depth investigation of the mechanisms of action of lithium is needed before lithium could be administrated to a large cohort. Ultimately, a randomized, double-blind clinical trial against a placebo would help to validate the effects of lithium in *SHANK3*-deficient patients.

C. Genetic engineering

Manipulation of the genome offers great advantages in the study of several natural phenomena, ranging from cell molecular mechanisms to the identification of disease treatments. Since the beginning of research on molecular mechanisms, scientists have investigated, modified and manipulated the genome. To achieve this, researchers have developed several methods for precise

editing and regulation of the genome. To create a nucleotide-specific modification in the genome, it is necessary to have two main components: an element that binds to the desired genome location in a specific manner and an element that creates the modification (Martinez-Lage, Torres-Ruiz et al. 2017). One of the first approaches to achieve this was the use of meganucleases. Nucleases normally recognize long specific nucleotide sequences, making them a good tool for genome editing. Nucleases are re-engineered to target desired genome locations. Another approach to editing the genome includes zinc-finger nucleases and transcription activator-like effector nucleases (TALENs). These tools relied on the use of transcription factors that are fused with a nuclease domain; this way, the transcription factor recognizes targeted sequences through its DNA-binding domain, and the nuclease domain creates a double-stranded break close to the targeted site (Wang, La Russa et al. 2016). The main disadvantage of these methods is that they required the engineering of protein sequences to achieve specific DNA-protein interactions. Protein engineering is a difficult process, as it requires complex and labor intensive molecular techniques, making it difficult to implement these methods in high-throughput studies.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

In contrast to previous methods, the CRISPR system uses a nuclease that is guided by RNA, the Cas9 nuclease, making it easier to manipulate it. Reprogramming of Cas9 to target new sites in the genome can be achieved by modifying the guide RNA; this task is faster and less complex than protein engineering.

The CRISPR system is naturally found in bacteria and archaea; it provides an immunity defense against virus and plasmids. Invading organisms may introduce foreign DNA into the host, and these fragments are known as protospacers. Foreign DNA is degraded by the CRISPR RNAs, which binds to foreign DNA through Watson–Crick interactions. The CRISPR system works in three main steps, the first step, called Foreign DNA acquisition, is characterized by introduction of foreign DNA into the CRISPR locus. In the second stage, CRISPR RNA processing, Cas9 nucleases are expressed and the CRISPR locus is expressed into pre-crRNA, which is further processed into mature crRNAs. Subsequently, in the third stage, RNA-guided targets viral elements and Cas9 mediates the cleavage of invading DNA (Jinek, Chylinski et al. 2012). The CRISPR system function depends on the presence of PAM sequences (protospacer adjacent motif) in the invading genome. These sequences are usually found adjacent to the crRNAs recognition

sequences. PAM sequences are not present in the host genome, thus avoiding self-cleavage and protecting the host from viral infection.

A dual-RNA structure guides Cas9 to introduce double-stranded breaks in a target site. This dual-RNA structure is formed of two RNAs: the crRNA and the tracrRNA. Once the crRNA recognizes foreign DNA, the tracrRNA hybridizes with the crRNA, and Cas9 binds to this site (Wang, La Russa et al. 2016). Cas9 target recognition also requires a GG-dinucleotide-containing PAM sequence adjacent to the recognition site. To achieve DNA cleavage, Cas9 uses two endonuclease domains, the HNH, and RuvC-like domains.

In 2012, Jinek et al. showed that Cas9 could be programmed to create double-stranded DNA breaks guided by a single RNA. The authors engineered RNA to fuse the crRNA–tracrRNA into a single guide RNA (Jinek, Chylinski et al. 2012). This re-engineered CRISPR system is easier to use and implement as it only requires customizing a single guide RNA to pair the DNA of interest; any location in the genome that contains a PAM sequence can be targeted. Once Cas9 has created a double-stranded break in the desired genome location, the DNA is repaired. This can be achieved by two cellular mechanisms, nonhomologous end joining (NHEJ) or homology-directed repair (HDR). Both mechanisms can be exploited in different ways to achieve desired outcomes. NHEJ may cause random mutations in the site; these can be advantageous when looking to create frameshifts or disrupt gene function. On the other hand, HDR can be used to replace or introduce a donor DNA template at the targeted locus, allowing several applications, such as mutation repair, introduce a reporter, create a specific mutation or introduce a specific deletion (Wang, La Russa et al. 2016).

The simplicity of the CRISPR/Cas9 system has made it nowadays widely used for several applications, including disease modeling, gene therapy, and gene function discovery. The CRISPR/Cas9 system can edit multiple target sites across the genome making it an ideal platform for high-throughput applications. Furthermore, the CRISPR/Cas9 applications are not restricted to knock-in or knock-out approaches, as the system can also be used to modify the epigenome, regulate gene expression, RNA editing, and live-cell imaging (Martinez-Lage, Torres-Ruiz et al. 2017).

Applications of CRISPR/Cas9 for disease modeling

Since the advent of the CRISPR/Cas9 system, the research community has been able to generate cellular models of multiple human diseases. Besides, the CRISPR/Cas9 technology has been widely used in animal models, allowing generating specific mutations in different organisms to study disease physiology. Finally, CRISPR has also been used to restore gene functionally in human cells, and ultimately, there have been attempts to use the system in gene therapy in vivo (Zhang, Wang et al. 2017).

In the study of ID, the CRISPR/Cas9 system has been very useful in uncovering disease mechanisms and looking for new ways to treat those diseases. For instance, the CRISPR/Cas9 system was used to simulate patient mutations in iPSCs. Heterozygous knockouts of the autism-related *CHD8* gene were done in cerebral organoids derived from iPSCs, uncovering transcriptional networks related to autism (Wang, Mokhtari et al. 2017). Comparison of dysregulated genes in autism spectrum disorders with other neurological diseases such as bipolar disorder and schizophrenia showed that there are common dysregulated genes, providing new approaches to develop treatments for genetically heterogeneous complex traits (Wang, Mokhtari et al. 2017). Regarding the use of CRISPR/Cas9 in therapeutic approaches for ID, attempts have been made to restore the silenced *FMR1* gene in Fragile X syndrome. This was achieved through the deletion of the expanded CGG-repeat of the fragile X chromosome in human iPSCs (Zhang, Wang et al. 2017).

Research rationale

Even though previous research on KS has advanced our understanding of the disease, we still lack a detailed molecular and cellular understanding of it. Recent fast-paced progress in the iPSC field has positioned them as key tools for understanding disease etiology and its molecular processes. IPSC models of KS would allow us to study the disease pathology as it develops in neuron cells while retaining the patient genetic background. Recently, our laboratory developed a rapid pipeline to induce pluripotency and simultaneously create gene-specific modification using CRISPR/Cas9 (Bell, Peng et al. 2016).

The hypothesis of this project is to test whether KS could be modeled in iPSCs. We plan to establish several iPSC lines from KS patients and their sex-matched siblings. These lines will serve to study KS molecular mechanisms and to make comparisons between individuals. Besides, we propose to create an *EHMT1* genetically modified iPSC line from a patient´s sex-matched sibling, which will add scientific power to our comparisons and reduce intra-individual variation.

This project is part of a larger project on EHMT1 haploinsufficiency and, as mentioned early in the section "*Contextualizing this thesis in the context of a larger EHMT1 mutation study in Dr. Ernst's laboratory*", the iPSC models presented in this thesis are going to be used to study gene expression patterns in iPSC-derived neurons. Ultimately, these lines are going to be used in a translational research approach to search for new treatments for KS.

Objectives

1. To create iPSC lines from a heterozygous *EHMT1*-deletion patient and his healthy brother.
2. To create iPSC lines from a heterozygous *EHMT1* point mutation patient and her healthy sister.
3. To simultaneously induce pluripotency and delete the first two exons of the *EHMT1* SET-domain in a healthy fibroblast line using the CRISPR/Cas9 system.
4. To perform quality control experiments in all the lines to ensure they are *bona fide* iPSCs.
   4.1 Assess the expression of pluripotency markers at the RNA and protein level.
   4.2 Ensure iPSCs are mycoplasma-free.
   4.3 Assess iPSC differentiation potential into each of the three germ layers.

4.4 Ensure iPSC lines do not carry gross chromosomal abnormalities that may have resulted from reprogramming or genetic editing.

4.5 Verify cell line integrity, i.e., ensure cell lines were not misidentified or cross-contaminated during the reprogramming process.

Chapter 2: Materials and Methods

Ethics statement

Our study was approved by the Douglas Mental Health University Institute Research Ethics Board. Families were asked to give their consent for study participation. The consent form code is IUSMD-15-14, the form title is "*The INVESTICATE project: Identification of New Variation, Establishment of Stem cells, and TIssue Collection to Advance Treatment Efforts*".

Collection of fibroblast cell lines

We received fibroblasts from a male affected with KS, patient ID 58204-del0.466Mb, and his healthy brother, *EHMT1* mutation-free, individual ID 58203-control. Cells were collected through skin biopsy at the SickKids Hospital, Toronto and sent to the Douglas Hospital. Fibroblast lines were expanded upon receipt.

Fibroblasts from a female patient affected with KS were also collected, patient ID 58071-p.R948W. We also received fibroblasts from her healthy sister, *EHMT1* mutation-free, individual ID 58070-control. Fibroblasts were collected through skin biopsy in the UT Southwestern Medical Center, Texas and sent to the Douglas Hospital.

Culturing Skin Fibroblasts

Fibroblasts were plated on 0.05% gelatin (Sigma-Aldrich) coated tissue culture dishes (Corning). Cells were maintained in Dulbecco modified Eagle's medium (DMEM) (Thermofisher) supplemented with 10% Fetal Bovine Serum (FBS) (Thermofisher). An antibiotic solution was added to the media to prevent contamination; the solution was composed of 100 U/mL Penicillin and 100 µg/mL Streptomycin (Mediatech). Once cells reached 80% confluency, they were passaged to fresh gelatin-coated dishes using 0.05% Trypsin-EDTA (Thermofisher).

DNA and RNA extraction

All DNA extractions were performed using QIAamp DNA Mini kit from Qiagen. DNA quantities were measured on a NanoDrop spectrophotometer (Thermofisher). RNA extractions were performed using the Direct-zol RNA MiniPrep Kit from Zymo Research. A DNase I treatment step was included in the process to eliminate any DNA contamination. RNA quantities were determined on a NanoDrop spectrophotometer.

cDNA synthesis

For each reaction, an initial RNA quantity of 10 ul RNA (100 ng/ul, 1 ug) was included. We included 13.5 ul of ultra-pure water, 0.5 ul of random hexamer primers (3 mg/ml) and 2 ul 10 mM dNTPs. The samples were heated at 65°C for 5 min in a PCR machine. Following, we placed the hot tubes on ice for at least 3 min. Next, we added 8 ul of 5x First Strand Buffer, 4 ul of 0.1M DTT, 2 ul of M-MLV Reverse Transcriptase (400 U/sample). Finally, cDNA was synthesized in a PCR machine, the reaction was incubated at 37°C for 1 h 15 min and at 70 °C for 15 min.

Relative quantitative real-time polymerase chain reaction

To measure *EHMT1* DNA signal, we designed primers on the first intron of the gene; primer optimal amplification temperature was determined to be 62 ºC. Relative qPCR was performed using the following primers (forward: 5'-TTCTCCCCAGACAATGCTGC-3') and (reverse: 5'-TGAGAATGAGCACGGGGAAG-3').

To measure *EHMT1* RNA signal, a PrimeTime® qPCR Primer assay was ordered from IDT; the assay identifier is Hs.PT.58.20041494. For both DNA and RNA signal quantifications, we used quantitative real-time PCR reactions. The experiments were performed in 384-well plates using a Quant Studio 6 Flex Real-time PCR machine (Life Technologies). As reference cDNA, we created a pool of cDNA from all the samples included in each experiment. Each well in the qPCR plate included 5 mL of the 2x gene expression master mix (2X Power SYBR Green PCR Master Mix, Applied Biosystems), 1 mL of 20X primer mix solution, including forward and reverse primers, 2 mL of RNAase free water and 2 mL of cDNA. To measure *EHMT1* RNA signal, the *GAPDH* gene was used as a reference gene for normalization.

Sequencing

We confirmed patient 58071-p.R948W point mutation through Sanger sequencing. Primers were designed to amplify the genomic region surrounding the patient's mutation. The primers used were (forward: 5'-TGAACATCCACGGAGACTCG-3') and (reverse: 5'-CACAAGACTGTCAAGCGGTC-3'). PCR products were purified using Agencourt AMPure XP beads; 100 µL of AMPure XP were added to a 50 µL PCR product plus 50 µL of water. DNA and AMPure beads mix was placed on a magnetic racket, allowing the separation of beads and DNA fragments from primer dimers, enzymes and nucleotides. Beads were washed twice with 70% ethanol. DNA was eluted and transferred to a new Eppendorf tube. Purified PCR products were

sent for Sanger Sequencing analysis to the McGill University and Genome Québec Innovation Centre (Montreal, QC) on a 3730xl DNA Analyzer (Applied Biosciences).

Mutation analysis

We used PolyPhen-2 (Polymorphism Phenotyping v2) (Adzhubei, Jordan et al. 2013) and SIFT (Sorting Intolerant from Tolerant) (Ng and Henikoff 2003) to predict the functional effect of the patient 58071-p.R948W point mutation on the biological function of the EHMT1 protein. The mutant EHMT1 protein structure was created using DynaMut (Rodrigues, Pires et al. 2018). Molecular graphics were performed using the PyMOL software (PyMOL Molecular Graphics System).

Conversion of fibroblasts to iPSCs

Fibroblasts reprogramming to iPSC was achieved using episomal vectors containing the genes *OCT3/4, SOX2, MYC3/4, KLF4, SHRNA P53* (ALSTEM) and a puromycin resistant gene. Episomal vectors were introduced into the cells using electroporation, in a Neon Transfection System (Invitrogen, Burlington). Approximately 500,000 cells were reprogramed using 5 µg of episomal vectors per reaction. Electroporation was achieved following the manufacturer's instructions, using the following parameters: 11,650 V, 10 ms, 3 pulses. Each reaction was plated in a well of a tissue culture 6-well plate coated with Matrigel (Corning) in 10% FBS DMEM. Twenty-four hours after electroporation, cell media was replaced by 10% FBS DMEM media supplemented with 2 µg/mL puromycin (Sigma-Aldrich). This step allowed for effectively selecting cells that were transfected with the episomal vectors. No transfected cells died after puromycin treatment. Puromycin selection was exercised for 48 hours. Two days after the electroporation, media was changed to fresh TesR-E7 media (Stem Cell Technologies, Vancouver). Daily media changes were done during the cells induction process, using TesR-E7 media. Approximately ten days after cell selection with puromycin, cell colonies started to form in the dish. Single iPSC colonies were observed and could be seen forming from a single fibroblast cell. After approximately three weeks of differentiation, cells formed robust colonies and distinct cell populations of approximately 500–1,000 µm in diameter. Once cell colonies reached this stage, they were detached using ReLeSR media (Stem Cell Technologies, Vancouver) and replated in fresh coated Matrigel plates with mTesR1 media supplemented with the ROCK inhibitor y-27632 (Sigma-Aldrich) at a final concentration of 10 µM. Pure iPSC colonies were either maintained in

culture in mTesR1 media or cryopreserved in fetal bovine serum with 10% DMSO (Dimethyl sulfoxide, Sigma-Aldrich) (Bell, Peng et al. 2017).

Assessment of pluripotency using absolute RT-PCR

Primers to assess the endogenous expression of the Yamanaka factors (*OCT3/4, SOX2, KLF4, C-MYC*) and *NANOG* have been previously described (Chan, Ratanasirintrawoot et al. 2009). We used those previously described sequences for our RT-PCR experiments. For the genes *DNMT3B, EST2*, and *ZFP42*, primers were designed and ordered from IDT. All the primer sequences are described in Table 1. Quantitative real-time PCR reactions were performed in 384-well plates using a Quant Studio 6 Flex Real-time PCR machine (Life Technologies). We used a reference pool of cDNA to generate a standard curve. Serials dilutions provided amounts ranging between 0.00064 to 10 ng.

Gene	Forward sequence	Reverse sequence
Nanog	TGAACCTCAGCTACAAACAG	TGGTGGTAGGAAGAGTAAAG
OCT 3/4	CCTCACTTCACTGCACTGTA	CAGGTTTTCTTTCCCTAGCT
SOX2	CCCAGCAGACTTCACATGT	CCTCCCATTTCCCTCGTTTT
c-Myc	TGCCTCAAATTGGACTTTGG	GATTGAAATTCTGTGTAACTGC
KLF4	GATGAACTGACCAGGCACTA	GTGGGTCATATCCACTGTCT
DNMT3B	CCATCGACCTCACAGACGAC	CATCTCCACTGTCTGCCTCC
EST2	CACGCACACCAGGCACT	CCACTACCGCGAGGTGC
ZFP42	ACTGGAGAGAAGCCGTTTCG	TGCGTTAGGATGTGGGCTTT

Table 1. Primers to assess the pluripotency state of our iPSC lines. Primers to measure the endogenous expression of *OCT3/4, SOX2, KLF4, C-MYC* and *NANOG* were taken from (Chan, Ratanasirintrawoot et al. 2009).

Assessment of pluripotency using immunofluorescence

Cells were cultured on Matrigel coated coverslips for five days with mTeSR1 medium. Subsequently, cells were washed three times with phosphate-buffered saline (PBS) and fixed with 3% paraformaldehyde (Sigma-Aldrich) on slides. Following fifteen minutes of incubation at room

temperature, samples were washed with PBS for fifteen minutes three times. To permeabilize the cells, we incubated them in PBS + 0.1% Triton X-100 at room temperature for 10 minutes. Permeabilization buffer was removed and replaced with a blocking solution, consisting of PBS plus 5% serum, and cells were incubated at room temperature for 60 minutes. Next, we removed the blocking solution and added the corresponding primary antibody in appropriate dilution (Table 2). Cells were incubated overnight. The next day, cells were washed with PBS, 0.5% PBS-BSA containing an appropriate dilution of secondary antibody was added to the samples and incubated at room temperature for 60 minutes. Subsequently, the samples were washed with PBS and stained with DAPI for 5 minutes at room temperature. Cells were visualized on an FV1200 Laser Scanning Microscope (Olympus). Images were analyzed using ImageJ.

Antigen	Host species	Isotype	Working dilution	Manufacturer	Catalog number
SSEA4	Mouse	IgG	1 in 100	Abcam	ab109884
OCT4	Rabbit	IgG	1 in 100	Abcam	ab109884
TRA-1-60	Mouse	IgM	1 in 100	Abcam	ab109884
NANOG	Rabbit	IgG	1 in 50	Abcam	ab109884

Table 2. Description of antibodies used for assessment of pluripotency.

Assessment of mycoplasma contamination

We decided to use a PCR-based commercial assay to assess mycoplasma contamination in our iPSC lines. We employed the EZ-PCR™ Mycoplasma Test Kit (Biological Industries) because of its ability to detect a great variety of mycoplasma species, including *M. fermentans*, *M. hyorhinis*, *M. arginini*, *M. orale, M. salivarium, M. hominis, M. pulmonis, M. arthritidis, M. bovis, M. pneumoniae, M. pirum, M. capricolum*, and Spiroplasma. Pure iPSC colonies were maintained in culture in mTesR1 media. We transfered 1 mL of mTesR1 media supernatant into a 2 mL centrifuge tube. Cell culture media supernatant was processed according to the manufacturer's instructions. PCR amplification products were run on a 1.5% agarose gel.

Assessment of differentiation potential

We assessed the differentiation potential of our iPSC lines using three strategies. The first method involved spontaneous differentiation of the iPSC into Embryoid Bodies (EBs) and subsequent analysis using the TaqMan hPSC Scorecard Assay. The second method was a trilineage

differentiation strategy and analysis with the TaqMan hPSC Scorecard Assay. Finally, we used a home-developed protocol involving three gene markers per germ layer.

Spontaneous differentiation of iPSCs into Embryoid Bodies

IPSC lines were cultured in Matrigel-coated dishes with mTesR1 media. When 90% confluency was reached, the mTesR1 media was changed for Embryoid Bodies (EBs) media. EBs media was created according to the TaqMan® hPSC Scorecard™ Panel protocol from Applied Biosystems, Life Technologies. EBs media was created using 79 mL of DMEM/F-12 (1X) with GlutaMAX™-I, 20 mL of KnockOut™ Serum Replacement (KSR), 1 mL of MEM Non-essential Amino Acids Solution (10 mM) and 100 µL β-Mercaptoethanol. Cells were gently detached from the plates and transferred into a 15 mL canonical tube. Subsequently, cells were transferred to non-tissue culture-treated dishes to prevent cell attachment and allow cell aggregation and EBs formation. Cells were incubated in a 37°C, 5% CO2 incubator for seven days, with daily EB media changes. On day seven, cells were transferred into a 15 mL canonical tube and allowed time to sediment down; the supernatant was removed, and cells pellets were processed for RNA extraction.

TaqMan hPSC Scorecard Assay

Seven-day EBs RNA was converted to cDNA. The later was further processed to be used on the qRT-PCR TaqMan® hPSC Scorecard™ Panel according to the manufacturer's instructions. We used the 384 well-plate TaqMan® hPSC Scorecard™ Panel and the 2X TaqMan® Gene Expression Master Mix to run the qRT-PCR. Gene expression results were analyzed in the web-based hPSC Scorecard™ Analysis Software.

Directed trilineage differentiation

We employed the STEMdiff™ Trilineage Differentiation Kit (Stem Cell Technologies, Vancouver) to achieve directed differentiation of our iPSCs lines to the three germ layers. To process the iPSC samples for trilineage differentiation, we followed the *Stem Cell Technologies*' protocol instructions. STEMdiff™ Trilineage Ectoderm Medium, STEMdiff™ Trilineage Mesoderm Medium, and STEMdiff™ Trilineage Endoderm Medium were supplemented with the ROCK inhibitor y-27632 (Sigma-Aldrich) at a final concentration of 10 µM. IPSC lines were cultured in Matrigel-coated dishes with mTesR1 media, when cells reached 70% confluency, they were visually examined to identify regions of differentiated cells, and those regions were removed from the plates. Cells were washed with PSC and detached from the wells using the Gentle Cell

Dissociation Reagent (Stem Cell Technologies). Cells aggregates were broken and transferred into a centrifuge tube with 1 mL of DMEM/F-12. Cells were centrifuged at 300xg for 5 minutes. Subsequently, cells were divided into three batches and resuspended in the appropriate STEMdiff™ Trilineage medium. Daily media changes were performed for seven days, after which, cell aggregates were processed for RNA extraction to assess differentiation into the three germ lineages.

Assessment of germ layer-specific markers

We decided to use Tsankov et al.'s list of markers for each germ layer (Tsankov, Akopian et al. 2015). Instead of using the 96 markers, we limited our analysis to nine markers, three markers per layer. We chose three good representative genes per layer to assess their expression on a PCR and electrophoresis gel experiment. This enables the simultaneous interrogation of various iPSC lines at less cost than the Scorecard Taqman assay. We chose the ectodermal markers *ALDH1A, COL1A* and *PAX6*, mesodermal markers *IGF2, BMP4* and *PDGFRA,* and endodermal markers *CDH1, SOX7* and *GATA6*. We used previously described primers for our PCR experiments (D'Antonio, Woodruff et al. 2017), except for the genes *PDGFRA* and *GATA6*, whose primers were designed and ordered from IDT. All PCR experiments to assess the expression of germ layer-specific markers were performed using the following conditions: 95°C for 4 minutes, 40 cycles of 95°C for 30 seconds, annealing temperature of 60°C for 30 seconds, 72°C for 30 seconds, followed by 72°C for 2 minutes. 5 µL of the reaction were run on a 1.5% agarose gel to confirm amplification.

Gene	Lineage	Forward sequence	Reverse sequence
ALDH1A	Ectoderm	CTGCTGGCGACAATGGAGT	CGCAATGTTTTGATGCAGCCT
COL1A	Ectoderm	GAGGGCCAAGACGAAGACATC	CAGATCACGTCATCGCACAAC
PAX6	Ectoderm	AACGATAACATACCAAGCGTGT	GGTCTGCCCGTTCAACATC
IGF2	Mesoderm	AGACGTACTGTGCTACCCC	TGCTTCCAGGTGTCATATTGG
BMP4	Mesoderm	GCACTGGTCTTGAGTATCCTG	TGCTGAGGTTAAAGAGGAAACG
PDGFRA	Mesoderm	AGGGATAGCTTCCTGAGCCA	TGTGCTTTCATCAGCAGGGT
CDH1	Endoderm	CGAGAGCTACACGTTCACGG	GGGTGTCGAGGGAAAAATAGG
SOX7	Endoderm	TCGACGCCCTGGATCAACT	CTGGGAGACCGGAACATGC
GATA6	Endoderm	CTCAGTTCCTACGCTTCGCA	GTCGAGGTCAGTGAACAGCA

Table 3. Primers to assess the differentiation potential

G-banded karyotyping

Karyotype analysis was performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. Trained technicians at the SickKids cytogenetics facility generated a cytogenomic research report for each iPSC line.

Cell line authentication

Cell line authentication was achieved using commercial microarrays to genotype common SNPs in the global population. Genomic DNA was diluted in nuclease-free water at a final concentration of 100 ng/µL, in a minimum volume of 20 µL. DNA was sent to the McGill University and Genome Québec Innovation Centre (Montreal, QC) to be analyzed on an Illumina Infinium® Global Screening Array. Microarray analysis was performed by bioinformaticians at Genome Quebec and results were uploaded in Nanuq platform. Individual relatedness was assessed using PLINK. Briefly, PLINK was used to find pairs of individuals who are genetically like each other, i.e., more than we would expect by chance in a random sample. Genome-wide identity-by-descent (IBD) scores were calculated per pairs of individuals.

CRISPR/Cas9 Gene Editing

We generated a Paprika RFP reporter (DNA 2.0) with a double nickase CRIPSR/Cas9 gene editing system and two gRNAs (DNA2.0) targeting the first two exons of the EHMT1 SET-domain. The section of the *EHMT1* gene-targeted and the gRNA sequences are shown in Table 4. The targeted sequence represents 902 base pairs of exonic sequence. We used 1 µg of the construct reaction per transfection reaction. Transfection was performed using the aforementioned parameters for iPSC reprogramming. We simultaneously transfected CRISPR/Cas9 constructs and iPSC reprogramming vectors, as previously described in (Bell, Peng et al. 2017). After transfection, cells were plated on Matrigel-coated plates and cultured with 10% FBS DMEM for 24 hours. Subsequently, cells were sorted using Fluorescence Activated Cell Sorting looking for RFP positive cells. RFP positive cells were replated with 10% FBS DMEM media supplemented with 2 µg/mL puromycin. This step allowed for effectively selecting cells that were transfected with the iPSC reprogramming episomal vectors. Puromycin selection was exercised for 48 hours. At two days post electroporation, media was changed to fresh TesR-E7 media. Colony formation and purification were performed as aforementioned for conversion of fibroblasts to iPSCs. In order to assess which iPSC colonies were successfully genetic engineered, we designed primers around the *EHMT1* SET-domain targeted sequence, primers are shown in Table 4, amplicon length is 958

bps. DNA was extracted and amplified; primers' optimal annealing temperature was 64 °C. Following, 5 µL of the PCR reaction were run on a 1.5% agarose gel to confirm amplification. Amplicons of genetically engineered iPSC colonies were sent for Sanger sequencing to Genome Quebec.

A) Section of the *EHMT1* SET-domain targeted	
CCTCATTTCAGTGACGTGGCACCAGCTGGCTTTTGCTGACCATTGTGGCAACAGG CTGAGGCTGTGGCCTGGGTTCACCACTACTCTCTATTTTTCAGGGCAAGGCTGCA GCTCTACCGGACGCGGGACATGGGCTGGGGCGTGCGGTCCCTGCAGGACATCCC ACCAGGCACCTTTGTCTGCGAGTGAGTGAGTCCCTGGGTCACCCCAAGCCTGGTG TCATTTCTGGGACGGAGGCCCATCTGTGTCTGTACTTCAGGAAGCCCCTCTGGGA GCAGGCACATCCCTGGCGTACAGCAGCGTGGGGTGGGGGCCACAGAGACCCTGG CCCCGAGAACCAAGCTCGTGCTGTCCTCAGTCCTCTTGCTGCTGCCCGTAACCAG CCCAGGAGTGCATTTAAGAAGGCGTGGTCCAGTTAGGAAGGCGTGGTCCAGTTA GGAAGGCGTGGTCCAGCATGGGCAGTCATCTCTAGGGGAGGACCATTCTGGGTT CTCAGGTACCAGACCGCAGACGCAGAGCTTTCGGTATCGTTATCATCATCCTCCG GACCCTCAGCTGAAACCCCAGTTCAACCCTGGGCACACCTCTCTAAACATGTTCC CTGCATGTTCTTCTGGTGTGCCCTGCTCTTTCCCTGTGGCTGCGGAGTCTGGGCTG TGCTTGTCTGTGGGCAGTGCTCGCTGCCTTCCAGGGCCTCACCTGCACCGCACCC TCTGCAGGTATGTTGGGGAGCTGATTTCAGACTCAGAAGCCGACGTTCGAGAGG AAGATTCTTACCTCTTTGATCTCGACAATAAGGTAATGTGTTTTGTGGGGTTGGG GCCACGCAGAACTTGTGAACTGTAAAACCTGAATGTGTTTGTCCCAGTAGGGCT GGGATTCAGAAGAGAGCTCTTACTGTT	
B) CRIPSR/Cas9 gRNA	
First exon of the SET-domain	
Forward sequence	Reverse sequence
CCCTGAAAAATAGAGAGTAG Reverse complement CTACTCTCTATTTTTCAGGG	GCTGCAGCTCTACCGGACGC
Second exon of the SET-domain	
Forward sequence	Reverse sequence
TACAGTTCACAAGTTCTGCG Reverse complement CGCAGAACTTGTGAACTGTA	CTGAATGTGTTTGTCCCAGT
C) Sanger sequencing primers	
Forward sequence	Reverse sequence
CATCGAACGCTTCGGATACA	CTATAGCAAGCCCACTCTTGTC

Table 4. Gene editing design. A) The two exons of the *EHMT1* SET-domain were highlighted in red letters. Sequences surrounding the two highlighted exons are *EHMT1* introns. B) CRIPSR/Cas9 gRNAs employed for gene editing with their sequences. The gRNA that targets *EHMT1* SET-domain first exon was highlighted in yellow on the *EHMT1* SET-domain targeted sequence; the gRNA targeting the second exon was highlighted in green. C) Primers used for amplification (amplicon length 958 bps), colony selection and Sanger sequencing.

Chapter 3: Results

We recruited four individuals for our study, two KS patients and their sex-matched siblings. The two KS patients showed different mutations in the *EHMT1* gene. One had a whole *EHMT1* deletion and the other had an *EHMT1* point mutation. Details about the patients are described below.

Description of patients

Patient 58204-del0.466Mb mutation information

Six-year-old male with hypotonia, global developmental delay and characteristic facial features of KS. Based on the cytogenetics report provided by the Clinical Cytogenetics Laboratory of SickKids hospital, patient 58204-del0.466Mb has a 0.466 Mb heterozygous deletion, in the chromosomal region 9q34.3. The clinic reported the following nucleotide positions of the deletion: 140,188,313-140,653,849 (genome build NCBI 37/hg 19 (2009)). Estimated patient 58204-del0.466Mb karyotype information is arrXp21.1(34,033,342-34,575,575)x2,9q34.3(140,188,313-140,653,849)x1, however, the exact breakpoints are unknown. There seem to be thirteen RefSeq genes in the deleted area; only two genes are OMIM Morbid Map genes. RefSeq genes involved in the deletion are *NRARP, EXD3, NOXAI, ENTPDS, MIR7114, PNPLAT, MRPL4l, DPH7, ZMYND19, ARRDC1* and *C9ORF37*. OMIM Morbid Map genes are *NSMF* and *EHMT1* (UCSC Genome Browser, hg19). Also, patient 58204-del0.466Mb has another copy number change, a 0.542 Mb duplication in chromosome region Xp21.1 involving the gene FAM47A. Based on the current literature, the nature of this structural change is unknown.

Individual 58203-control

Brother of patient 58204-del0.466Mb, young male phenotypically normal. No genetic abnormalities in the *EHMT1* gene.

Patient 58071-p.R948W mutation information

Five-year-old female with congenital heart disease, global developmental delay, oropharyngeal dysphagia, and hypotonia. The genetic testing company GeneDx performed Whole Exome Sequencing analysis to find out the mutation causing the KS phenotype in this patient. Additionally, the company performed an in silico structural analysis of the EHMT1 mutated protein. Based on the company's genetic testing report, individual 58071-p.R948W is heterozygous for the *de novo* p.R948W mutation in the *EHMT1* gene [variant p.Arg948Trp (CGG>TGG), c.2842 C>T in exon 19 (NM_024757.4)]. Furthermore, GeneDx reported that the

p.R948W mutation is damaging to the EHMT1 protein structure. According to the company´s genetic report, the amino acid change in polarity and charge has an impact on protein structure. This alteration causes protein dysfunction, resulting in only one copy of the protein working properly. GeneDx genetic report also stated that the p.R948W mutation was not observed with a significant frequency in approximately 6,500 individuals of European and African American ancestry in the NHLB1 Exome Sequencing Project.

Individual 58070-control
Sister of patient 58071-p.R948W, young female phenotypically normal. No genetic abnormalities in the *EHMT1* gene.

From each fibroblast line collected from each of the four individuals described above, we created two high-quality iPSC lines. In total, we characterized eight iPSC lines, two for each individual. We performed various quality control experiments to demonstrate that our lines are pluripotent, functional, free of any gross abnormalities and have the potential to differentiate into any of the three germ layers.

Confirmation of the presence of the patients' mutations
The first step in our quality control experiments was to demonstrate that the patient cells we received from the hospitals carry the expected mutations. To confirm the presence of the *EHMT1* deletion of patient 58204-del0.466Mb, we performed two relative RT-PCR experiments, one to examine the *EHMT1* DNA signal and one to examine the *EHMT1* RNA signal. For both experiments, we compared the patient's fibroblast *EHMT1* signal versus the *EHMT1* signal in the fibroblasts of the patient's brother (individual 58203-control). Results from the two qPCR experiments are presented in **Figure 1**.

As expected, there is a clear reduction in the patient's DNA and RNA signals compared to the control. The two experiments confirmed the presence of the *EHMT1* deletion in the patient's fibroblast we received from the hospital. While RNA and DNA signals are statistically reduced in patient cells versus control, we do not observe an exact 50% reduction in the signal; this is further discussed in the *Discussion* section.

To confirm the presence of patient 58071-p.R948W point mutation in the cell lines we received from the hospital, we designed primers around the mutation and analyzed the DNA region by

Sanger Sequencing (**Figure 2A)**. As expected, we observed a Thymine instead of the wild-type Cytosine normally observed at this genomic position. This was further confirmed by analyzing the sequenced reverse strand, where we observed an Adenosine instead of a Guanine. The patient's mutation was not observed in the patient's sister's DNA (individual 58070-control), neither in the forward nor the reverse strand.

The mutation was predicted to be "probably damaging" by the tool PolyPhen2, and analysis with SIFT showed that the substitution might affect protein function. We used PyMOL and DynaMut to visualize the mutant and wild-type EHMT1 protein structure. The substitution of a positively charged residue, Arginine, for the aromatic residue, Tryptophan, seems to disrupt amino acid bonds in this area (**Figure 2B**).

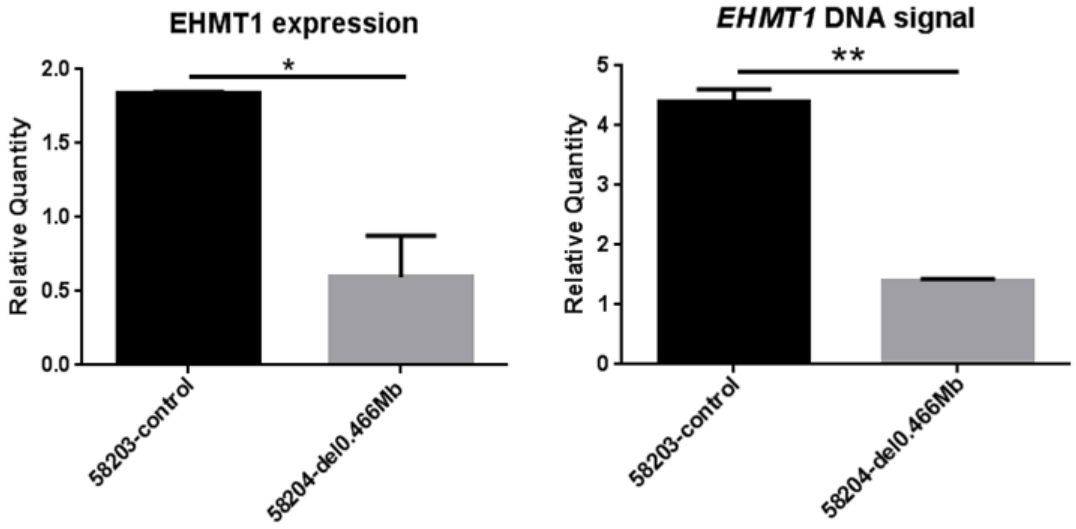
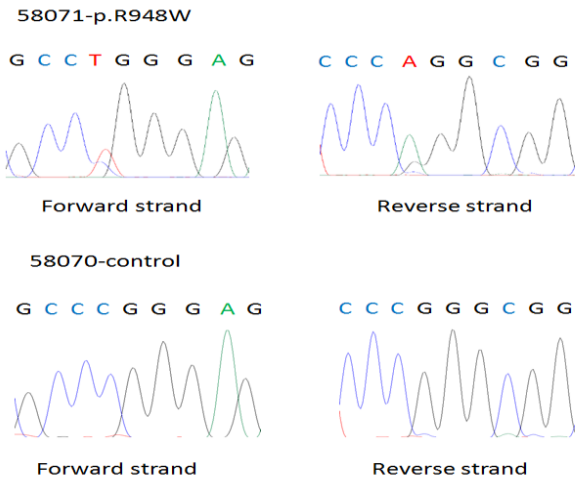


Figure 1. Characterization of patient 58204-del0.466Mb deletion. Quantification of the *EHMT1* signal at the DNA and RNA level. Comparison between patient 58204-del0.466Mb and control. Data are from two independent experiments. Error bars indicate standard error of the mean (SEM). Statistics was performed by unpaired t-test. *$p \leq 0.05$, **$p \leq 0.01$.

Analysis using the DynaMut server (Rodrigues, Pires et al. 2018) showed a variation in the vibrational entropy energy between the wild-type and the mutant structure of -4.457 kcal.mol^{-1}.K^{-1}, suggesting a decrease in molecule flexibility. DynaMut analysis also showed a variation in the Cutoff Scanning Matrix (mCSM) of -0.630 kcal/mol, suggesting a destabilizing effect on the protein (Pires, Ascher et al. 2014).

In summary, our analysis revealed that the conformation of the mutant protein was likely to be altered, resulting in protein dysfunction. These analyses supported the genetic testing company, GeneDx, report describing the p.R948W mutation as the disease-causing mutation.

A)



B)

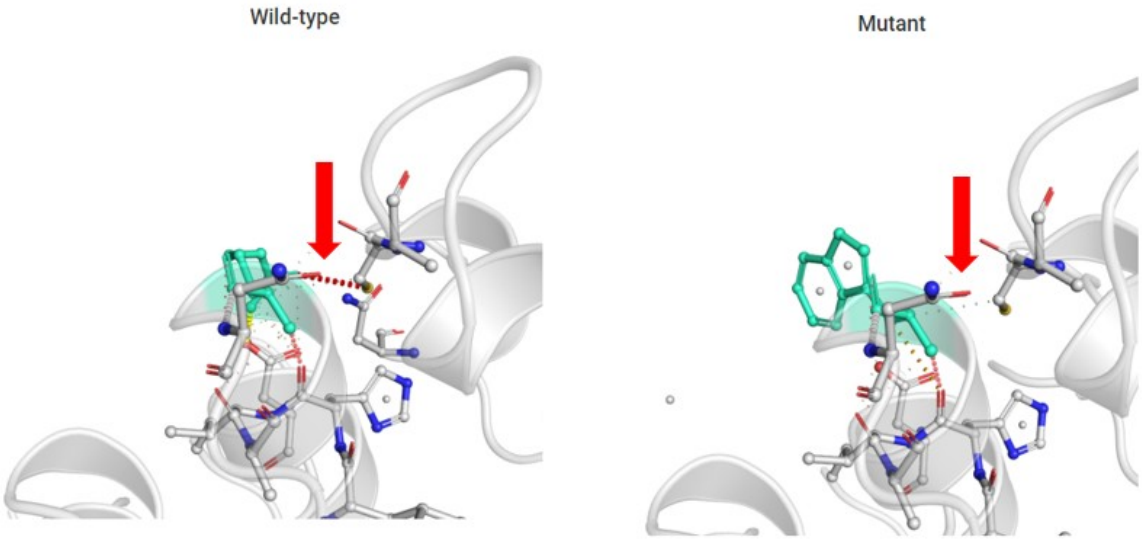


Figure 2. Description of patient 58071-p.R948W *EHMT1* point mutation. A) Chromatograms confirmed the presence of a heterozygous point mutation at *EHMT1*, producing a C to T change in patient DNA. The mutation was not observed in control DNA. **B)** Structure analyses of the p.R948 mutation in the EHMT1 protein. Substitution of an Arginine for the bulkier side-chain of Tryptophan disrupts amino acid interactions and seems to affect protein stability. The variation in the vibrational entropy energy between the wild-type and the mutant structure was determined to be -4.457 kcal.mol-1.K-1, suggesting a decrease in molecule flexibility.

Fibroblasts differentiation to iPSCs

IPSC lines were created using a previously described method (see *Materials and Methods*). Approximately ten days after cell selection with puromycin, cell colonies started to form in the dish. Single iPSC colonies were observed and could be seen forming from a single fibroblast cell. After approximately three weeks of differentiation, cells formed robust colonies and distinct cell populations of approximately 500–1,000 μm in diameter, as could be seen in **Figure 3**.

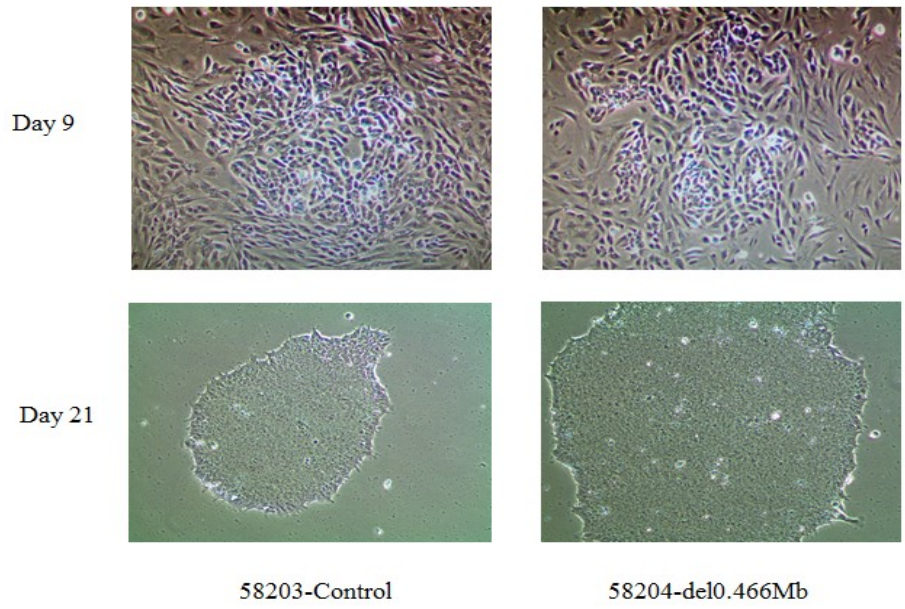


Figure 3. Conversion of patient fibroblasts to iPSCs

EHMT1 expression across cell lines

Once we obtained the iPSC lines, we wondered whether *EHMT1* would be expressed in these lines. Previous reports showed that *EHMT1* expression in stem cell lineages is crucial for correct embryonic development (Tachibana, Ueda et al. 2005). As mentioned in the introduction, *Glp* deficiency in mice leads to embryonic lethality and a severe reduction of the histone mark H3K9 (Tachibana, Ueda et al. 2005). To verify the expression of the *EHMT1* gene in our patient iPSCs, we performed a relative RT-PCR experiment using cDNA of fibroblasts and iPSCs of individuals 58203-control and 58204-del0.466Mb. As expected, the *EHMT1* gene was expressed in fibroblasts and iPSC lines (**Figure 4**).

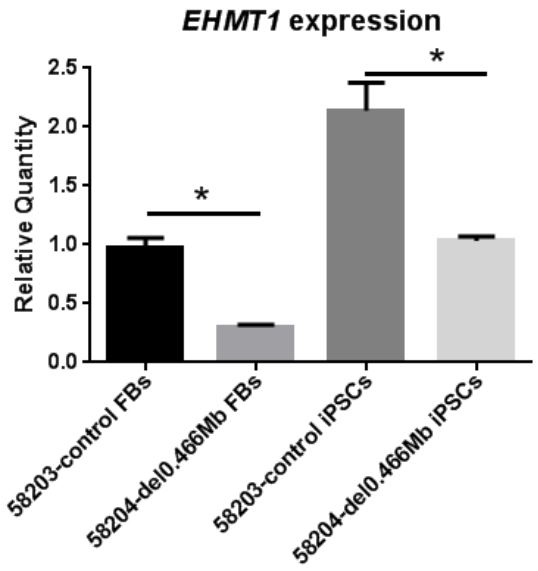


Figure 4. *EHMT1* expression on fibroblasts and iPSCs. EHMT1 haploinsufficiency is observed across cell lineages. Error bars indicate standard error of the mean (SEM). Statistics was performed by unpaired t-test. *$p \leq 0.05$.

Another important result from this experiment is that we observed the patient *EHMT1* haploinsufficiency in fibroblasts and iPSCs. Patient 58204-del0.466Mb cells showed statistically significant reduced levels of *EHMT1* expression compared to the control. We showed that *EHMT1* was expressed in our iPSC lines and that the patient cells that were initially used for reprogramming were not mislabeled during the process. This last objective was further studied in a cell line authentication experiment (section *Cell line authentication*).

Assessment of pluripotency

Endogenous expression of pluripotency markers

Gene expression analysis by quantitative RT-PCR was performed to assess the pluripotency potential of each iPSC line. We evaluated the endogenous expression of important genes associated with pluripotency, including the four Yamanaka factors, in addition to *NANOG, DNMT2, EST2,* and *ZFP42*. As discussed in the introduction, these proteins are excellent markers of pluripotency.

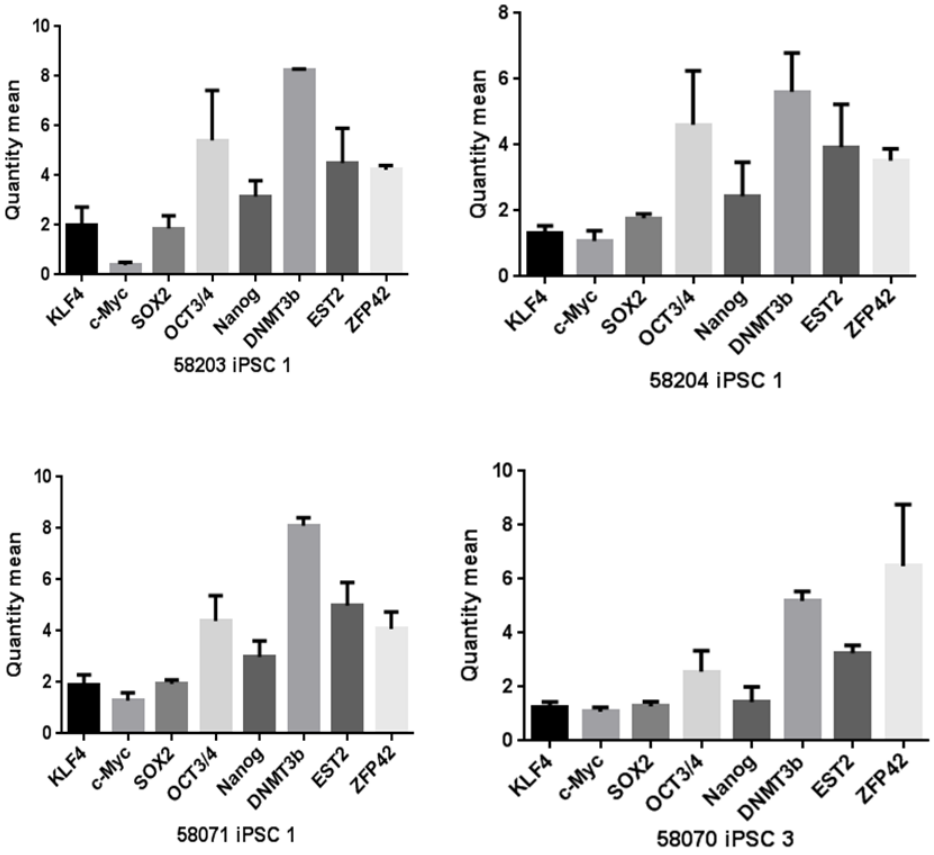


Figure 5. Endogenous expression of pluripotency markers in our iPSC lines. Quantification of the expression of various pluripotency gene markers. Error bars indicate standard error of the mean (SEM).

All the pluripotency markers are expressed in our iPSC lines. **Figure 5** shows the quantity mean expression of the human genes *OCT3/4, KLF4, SOX2, C-MYC*, *NANOG, DNMT3b, EST2* and *ZFP42* in four of our iPSC lines (one line per individual). The expression analysis results of the other four iPSC colonies can be found in the *Appendices* section (**Appendix I**).

Immunocytochemistry of pluripotency markers

Next, we assessed the immunocytochemical profile of four pluripotency-associated markers. We decided to evaluate the presence of the human proteins SSEA4, OCT3/4, NANOG, and TRA1-60, as it has been shown that these four markers are a good battery of markers to assess pluripotency and selectively isolate high-quality iPSCs using flow cytometry (Abujarour, Valamehr et al. 2013).

We assessed gene expression in different iPSC colonies using a combination of SSEA4 with OCT3/4, and TRA-1-60 with NANOG. In **Figure 6** we show the immunocytochemical profile of two iPSC lines, one of patient 58204-del0.466Mb and one of patient 58071-p.R948W. Immunocytochemical profiles of the other two characterized iPSC lines of these patients and the iPSC lines of their siblings can be found in the *Appendices* section (**Appendix II**).

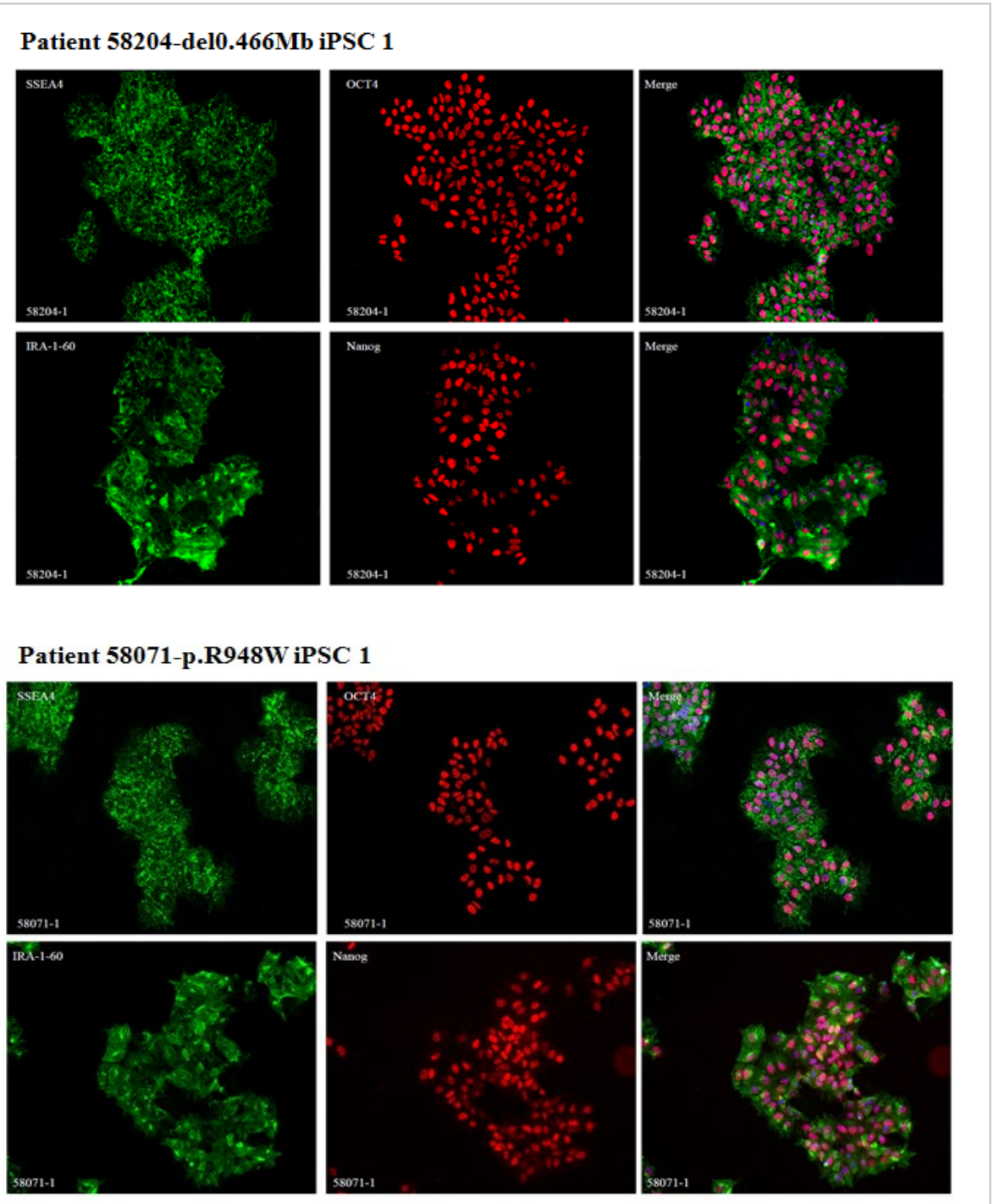


Figure 6. Immunocytochemistry of pluripotency markers. Immunocytochemical expression of four key pluripotent stem cell markers, SSEA4, OCT4, Nanog and TRA1-60. IPSC lines correspond to patients 58204-del0.466Mb and 58071-p.R948W. The immunocytochemistry profile of the other characterized iPSC lines can be found in *Appendix II*.

Assessment of mycoplasma contamination

We assessed mycoplasma contamination in our iPSC lines media. We included a positive control on the gel to check PCR efficiency. As it can be observed in **Figure 7**, none of the analyzed iPSC lines have mycoplasma contamination. The fact that all our iPSC lines are mycoplasma-free highlights the quality of our reprogramming process.

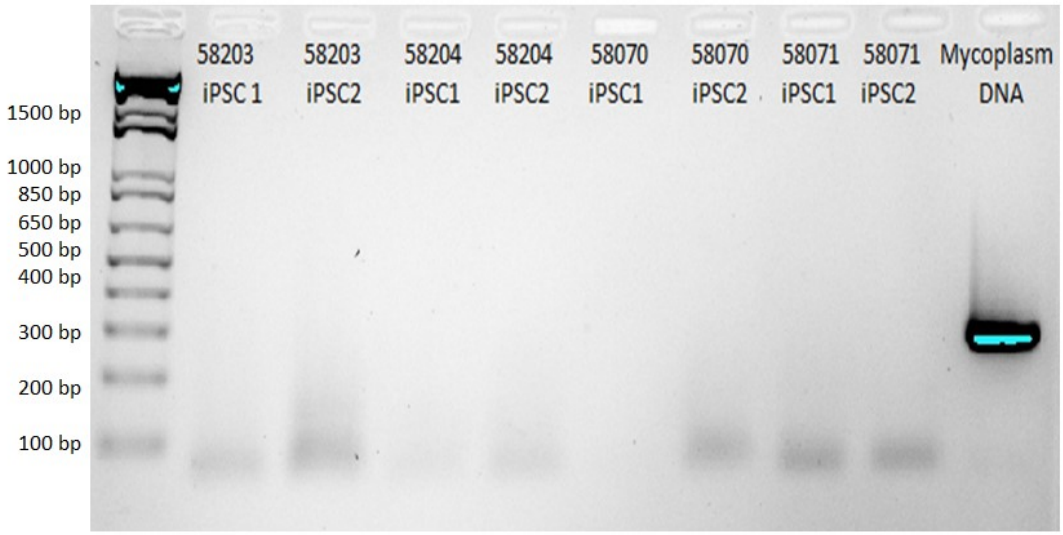


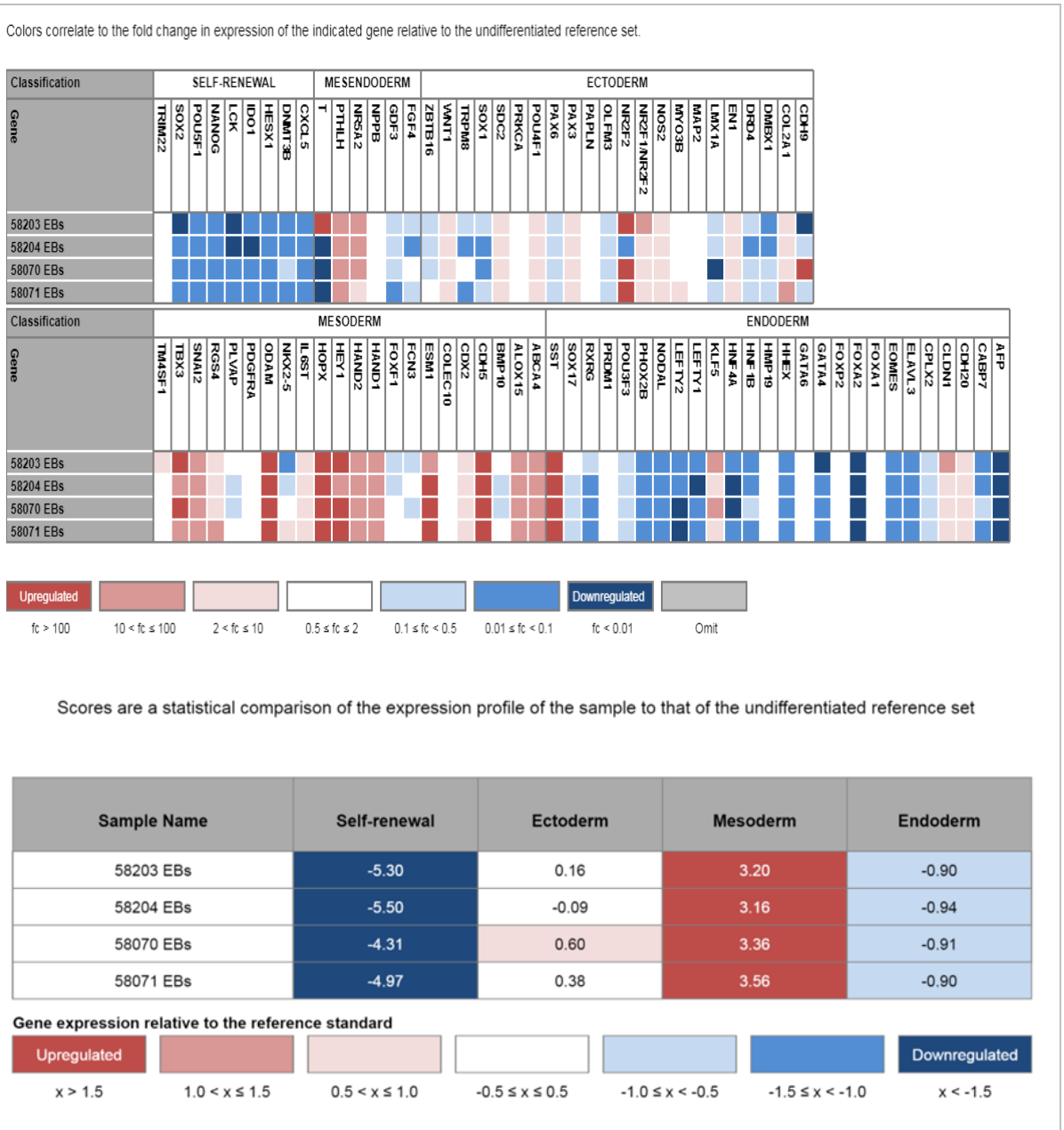
Figure 7. Assessment of mycoplasma contamination. None of our iPSC lines showed mycoplasma contamination. The size of the positive template PCR product is 270 bp.

Assessment of differentiation potential

The TaqMan hPSC Scorecard Assay in Embryoid Bodies

We allowed the spontaneous differentiation of the iPSC lines into EBs for seven days (see *Materials and Methods*). For one iPSC line of each individual, we assessed the in-vitro differentiation potential toward the three germ layers. Gene expression analysis of seven-day EBs showed a tendency of our iPSC lines to differentiate into the mesoderm (**Figure 8**). Self-renewal and endodermal markers were downregulated in the four iPSC lines analyzed. EBs of line 58070 showed a slight upregulation of genes associated with the ectoderm.

Given that all the four EB lines showed a tendency to differentiate to the mesoderm, we thought this strategy would be inappropriate to show the differentiation potential of our iPSC lines, so we decided to switch to another strategy. We chose to do a proof of concept experiment, in which we directly differentiated one iPSC line to each of the three germ layers and analyzed the gene expression patterns.



Scores are a statistical comparison of the expression profile of the sample to that of the undifferentiated reference set

Sample Name	Self-renewal	Ectoderm	Mesoderm	Endoderm
58203 EBs	-5.30	0.16	3.20	-0.90
58204 EBs	-5.50	-0.09	3.16	-0.94
58070 EBs	-4.31	0.60	3.36	-0.91
58071 EBs	-4.97	0.38	3.56	-0.90

Figure 8. TaqMan hPSC Scorecard Assay in Embryoid Bodies. One iPSC line of each individual was placed on media to allow spontaneous differentiation to EBs for seven days. The four iPSC lines analyzed showed a tendency to differentiate into mesoderm.

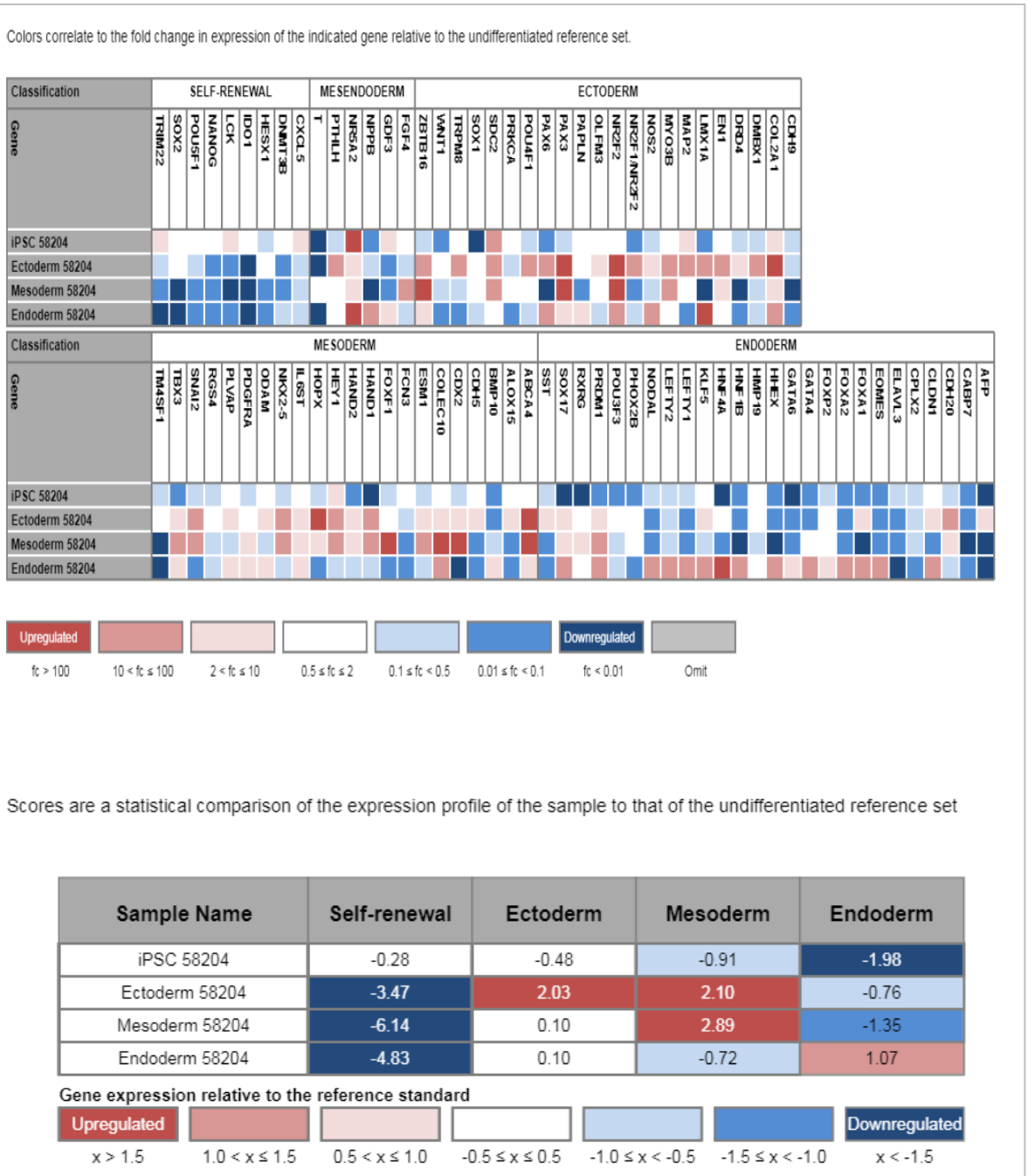
The TaqMan hPSC Scorecard Assay a trilineage differentiated line

We chose 58204-del0.466Mb iPSC line to be differentiated into the three germ layers because it is a patient line and thought it would be more interesting to prove its differentiation potential.

Ectoderm results showed upregulated expression of ectodermal markers. Interestingly, markers of the mesoderm were more upregulated than the ectodermal markers (mesoderm=2.10 and ectoderm= 2.03). Endodermal and self-renewal markers were downregulated.

Mesoderm results showed upregulated expression of genes associated with the mesoderm. Endodermal and self-renewal markers were downregulated. We did not observe a statistically significant upregulation or downregulation of the ectodermal markers compared to the reference.

Endoderm results showed upregulated expression of markers of the endoderm and downregulated expression of self-renewal and mesodermal markers. We did not observe a statistically significant upregulation or downregulation of the ectodermal markers compared to the reference.



Scores are a statistical comparison of the expression profile of the sample to that of the undifferentiated reference set

Sample Name	Self-renewal	Ectoderm	Mesoderm	Endoderm
iPSC 58204	-0.28	-0.48	-0.91	-1.98
Ectoderm 58204	-3.47	2.03	2.10	-0.76
Mesoderm 58204	-6.14	0.10	2.89	-1.35
Endoderm 58204	-4.83	0.10	-0.72	1.07

Gene expression relative to the reference standard

Upregulated						Downregulated
x > 1.5	1.0 < x ≤ 1.5	0.5 < x ≤ 1.0	-0.5 ≤ x ≤ 0.5	-1.0 ≤ x < -0.5	-1.5 ≤ x < -1.0	x < -1.5

Figure 9. Trilineage differentiation results of line 58204-del0.466Mb. The line was differentiated to the ectodermal, mesodermal, and endodermal lineages and analyzed by RT-PCR using the TaqMan hPSC Scorecard Assay. Results showed the line's capacity to form all three lineages.

Assessment of differentiation potential using a nine-gene PCR

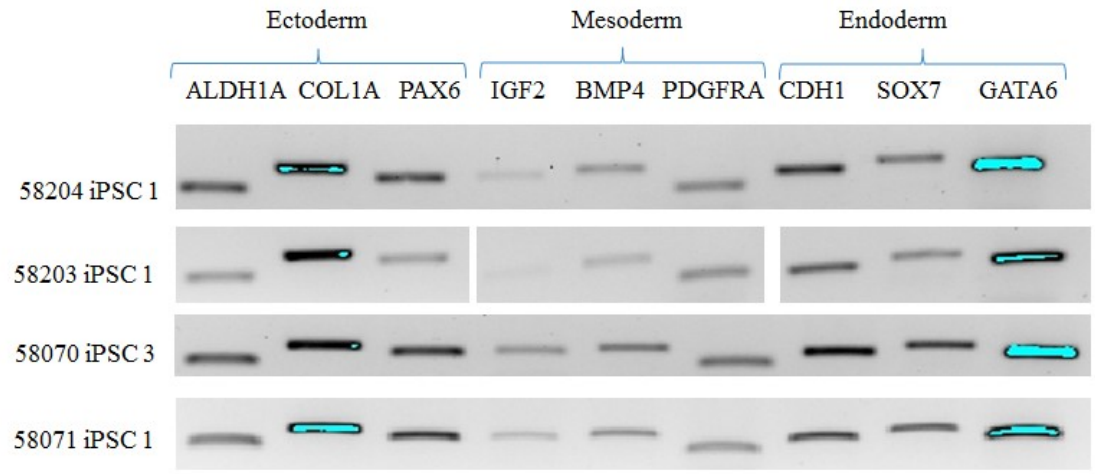


Figure 10. Assessment of differentiation potential using a nine-gene PCR. *In vitro* differentiation potential was examined by PCR using three markers per germ layer. We examined one iPSC line per individual; each line was differentiated to the three germ layers. Markers of each layer were expressed in their corresponding lineage.

We differentiated one iPSC line of each individual to the three germ layers; in total, we analyzed twelve samples. We chose the ectodermal markers *ALDH1A, COL1A* and *PAX6*, the mesodermal markers *IGF2, BMP4* and *PDGFRA,* and the endodermal markers *CDH1, SOX7* and *GATA6*. As mentioned in the *Introduction*, these markers have been reported in the literature as good markers of the germ layers (D'Antonio, Woodruff et al. 2017). All selected markers were expressed in their respective lineage (**Figure 10**).

G-banded karyotyping of iPSCs

G-banded karyotyping was performed in all our iPSC lines to rule out the presence of gross chromosomal abnormalities that could have originated during the iPSC reprogramming process. A normal, diploid, male human karyotype was reported for individual 58204-del0.466Mb and

58203-control. A normal, diploid, female human karyotype was also reported for individual 58071-p.R948W and 58070-control.

For all iPSC lines, a minimum of twenty metaphase cells was examined after G-banding. Of these metaphases, fifteen were scored, all of which or most of which (fourteen) displayed forty-six chromosomes with apparently normal karyotype after brief scoring for gross abnormalities. Also, five metaphases were karyotyped, all of which showed forty-six chromosomes with a normal diploid karyotype. The G-band karyotyping report of the iPSC line 1 of patient 58204-del0.466Mb is presented in **Figure 11**. The karyotype information of the other iPSC line characterized from this patient and the iPSC lines of individual 58203-control, 58070-control and patient 58071-p.R948W can be found in the *Appendices* section (**Appendix III**).

Interestingly, the G-band karyotyping results of iPSC line 2 individual 58203-control showed trisomy of chromosome twelve, which was detected in a single cell out of thirty cells examined. For this line, cytogenetic preparation was performed from a single submitted dish.

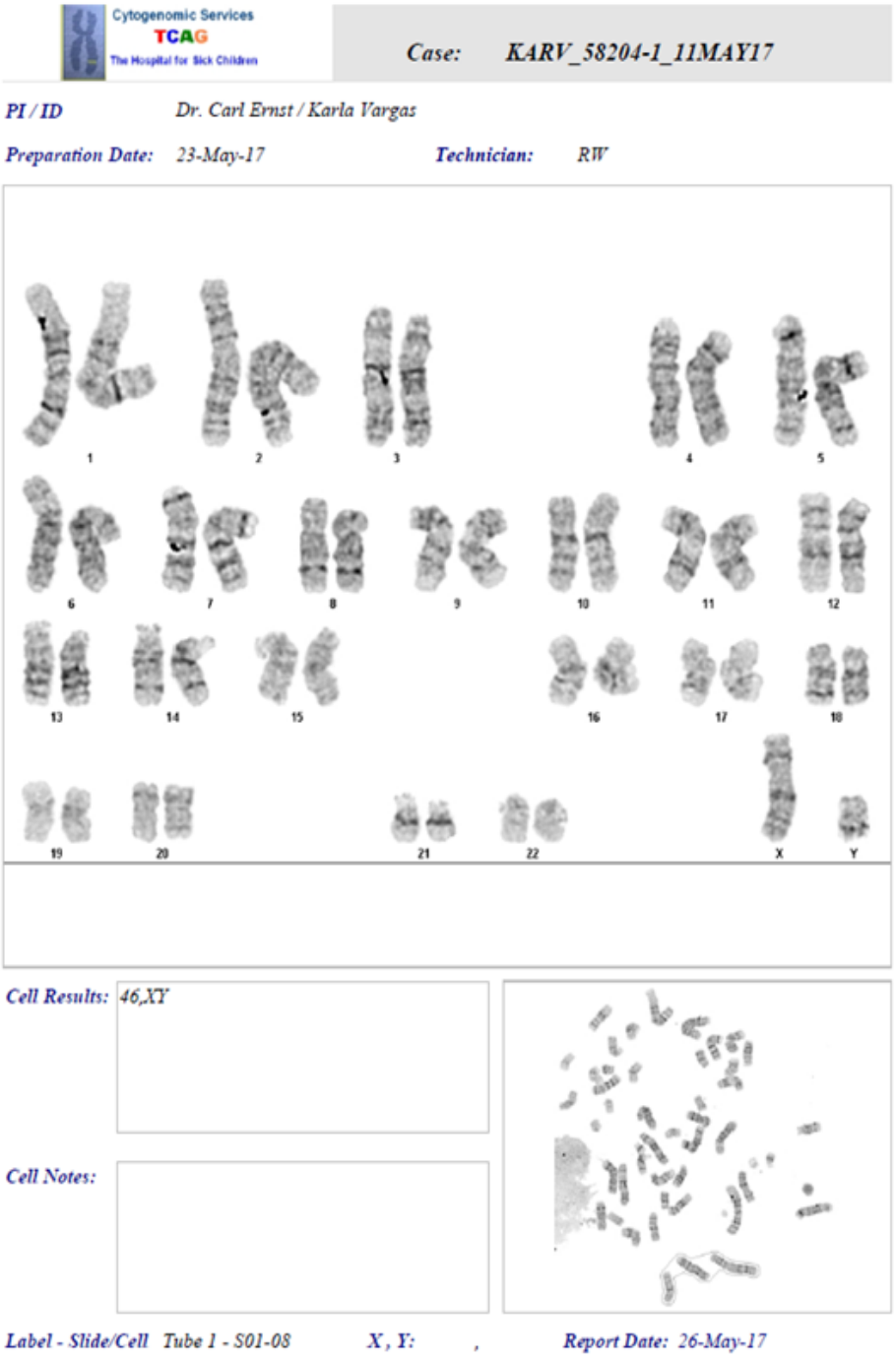


Figure 11. G-banded karyotyping of iPSC 1 patient 58204-del0.466Mb. Karyogram showing normal, diploid, male human karyotype (46, XY). Banding resolution: 400. The G-banded karyotyping report of the other characterized iPSC lines can be found in *Appendix III*.

Cell line authentication

Inadvertent cell line contamination is a problem in clinical research because it leads to false and non-reproducible results. To avoid cell line contamination, we authenticated our iPSC lines using the Infinium® Global Screening array, which genotypes common SNPs in the global population. We genotyped DNA from samples at different stages of the reprogramming process, i.e., fibroblasts, iPSCs, and CRISPR-edited lines. We performed this genotyping strategy for two purposes: to ensure that the initial cells used for reprogramming were not mislabeled during the process and to assess individual relatedness.

We used PLINK to evaluate identity and relatedness. The resulting PCA representation is shown in **Figure 12**. All the lines derived from an individual formed a cluster. For instance, all lines derived from patient 58071-p.R948W (FBs, iPSCs, and CRISPR-edited lines) mapped together. Same is true for the other three individuals, which indicated that we did not mislabel any sample during the reprogramming and editing process.

We next assessed individual relatedness to verify that the fibroblast lines we received from the hospital corresponded to samples from related individuals, i.e., 58204-del0.466Mb was related to individual 58203-control, and 58071-p.R948W was related to individual 58070-control. The analysis confirmed a relationship between the two pairs of individuals (**Table 5**).

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2
5820	58203	5820	58204	OT	0	0.1161	0.2061	0.6778
5820	58203	5807	58070	UN	NA	0.4277	0	0.5723
5820	58203	5807	58071	UN	NA	0.428	0	0.572
5820	58204	5807	58070	UN	NA	0.4297	0	0.5703
5820	58204	5807	58071	UN	NA	0.4279	0.0019	0.5702
5807	58070	5807	58071	OT	0	0.1494	0.1843	0.6663

Table 5. Relatedness assessment. Identity-by-descent (IBD) values calculated by PLINK. The analysis confirmed a relationship between the two pairs of individuals. The RT column shows the relationship type. OT stands for Other Related and UN stands for Unrelated. FID1= Family ID for first individual, IID1= Individual ID for first individual, FID2= Family ID for second individual, IID2= Individual ID for second individual, EZ= Expected IBD, Z0= P(IBD=0), Z1= P(IBD=1), Z2= P(IBD=2).

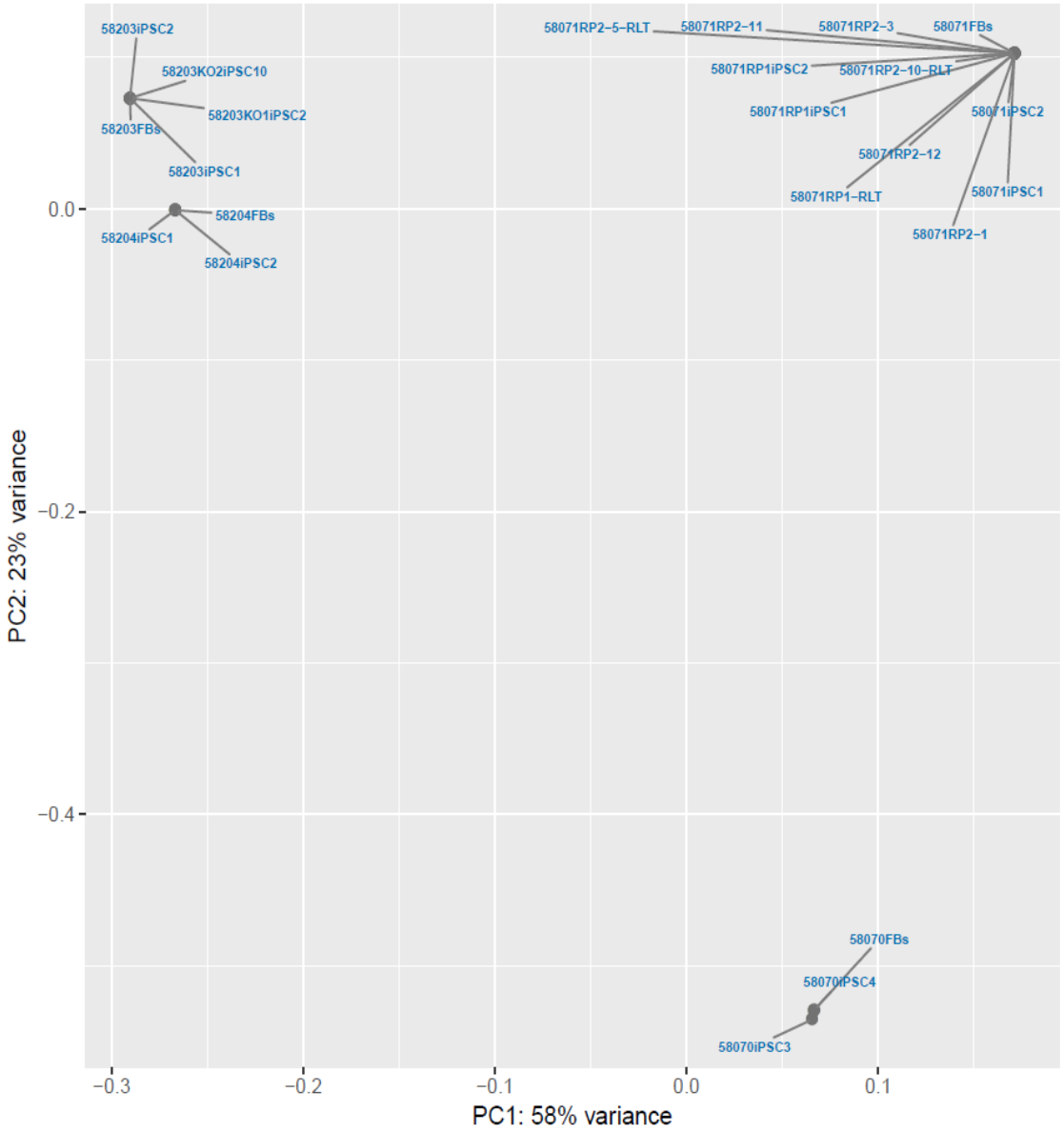


Figure 12. Cell line authentication using genotypic information. Samples derived from the same individual clustered together indicating proper sample labelling and identification. FBs = fibroblasts, KO = knock-out, RP = repair.

Generation of another model of KS: the SET-domain knock-out

We wanted to explore the effects of a SET-domain knock-out on our iPSC lines. To do this, we designed a strategy of simultaneous CRISPR engineering and iPSC reprogramming, following a protocol previously published in our laboratory (Bell, Peng et al. 2017). CRISPR engineering and iPSC reprogramming were carried out in a healthy individual fibroblast line; we chose individual 58203-control for this purpose (see *Materials and Methods*).

The *EHMT1* SET-domain is found across four exons in the *EHMT1* gene, exons twenty-four, twenty-five, twenty-six and twenty-seven. The first and the last two exons of the SET-domain are separated by a large intron of approximately 16,203 base pairs. The presence of this large intron in the middle made it difficult to design a CRISPR strategy to delete the four exons of the SET-domain in one experiment. Trying to induce a DNA deletion of this magnitude would have induced damage in the genome and probably led to gross chromosomal abnormalities. This is the reason why we decided to target the first two exons of the SET-domain. We generated a Paprika RFP reporter with a double nickase CRISPR/Cas9 gene editing system and two gRNAs targeting the first two exons of the *EHMT1* SET-domain.

After gene editing and induction of pluripotency, we obtained several iPSC colonies that could either harbor or not harbor the desired deletion. To assess which iPSC colonies were successfully genetically engineered, we amplified the genomic region surrounding the first two exons of the SET-domain (**Figure 13A**). The size product of a successful deletion was 169 base pairs. From the gel, we observed different modifications in the genome across the analyzed iPSC colonies. For instance, *EHMT1* KO v2-4 had a heterozygous deletion, one allele had a small deletion, and the other was a wild-type allele. Similarly, there was a small deletion in colony *EHMT1* KO v2-2. It was only colony *EHMT1* KO v2-10 that had a deletion of the expected size. We decided to continue further analyzing this colony, and as such, cut out this band from the gel and sent it for Sanger Sequencing at Genome Quebec.

We assigned this iPSC colony the identifier *SET-domain KO 2 colony 10*. This colony had a heterozygous deletion in the first two exons of the *EHMT1* SET-domain, as it was confirmed with the Sanger Sequencing results. We aligned the sequenced DNA against the human genome using

BLAT in the UCSC Genome Browser, assembly GRCh37/hg19, and observed a successful deletion of the first two exons of the *EHMT1* SET-domain. The Sanger Sequencing results and the BLAT alignment can be found in ***Appendix IV***.

Furthermore, the Sanger Sequencing results showed a frameshift mutation in the *SET-domain KO 2 colony 10* edited sequence. We, therefore, wondered whether this frameshift could affect the normal RNA production in the cells. To test this hypothesis, we extracted RNA from this colony and performed an RT-PCR experiment (**Figure 14C**). *SET-domain KO 2 colony 10* showed half levels of *EHMT1* expression compared with a control (58203-control). The *EHMT1* levels seen in the line were like those observed in patient 58204-del0.466Mb.

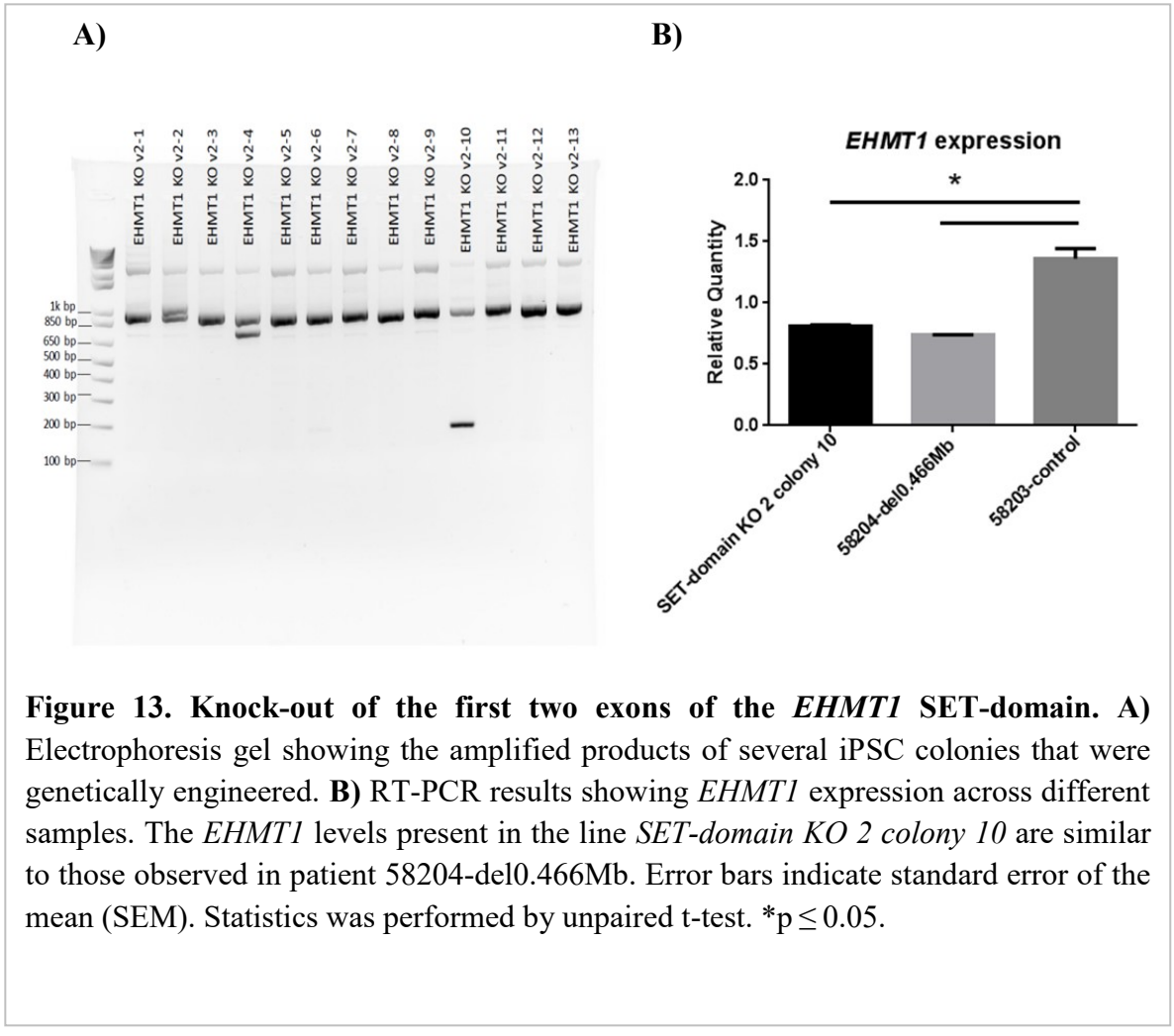


Figure 13. Knock-out of the first two exons of the *EHMT1* SET-domain. A) Electrophoresis gel showing the amplified products of several iPSC colonies that were genetically engineered. **B)** RT-PCR results showing *EHMT1* expression across different samples. The *EHMT1* levels present in the line *SET-domain KO 2 colony 10* are similar to those observed in patient 58204-del0.466Mb. Error bars indicate standard error of the mean (SEM). Statistics was performed by unpaired t-test. *$p \leq 0.05$.

We also wanted to make sure that this genetically engineered line was a good quality iPSC line, i.e., that it had all the characteristics of a pluripotent line and that it did not carry any gross genetic abnormalities. We performed all the quality control experiments described in the previous section on this cell line. To assess pluripotency, we checked for the endogenous expression of several pluripotency markers and performed immunocytochemistry experiments. We verified that the *SET-domain KO 2 colony 10* line was a mycoplasma-free line. Also, we assessed the differentiation potential of this line using our nine-gene PCR assay. Finally, we confirmed that the line did not have any gross chromosomal abnormalities using G-banded karyotyping and that the line was properly identified and labeled using the cell line authentication method previously described. All the quality control results of the *SET-domain KO 2 colony 10* line can be found in in the ***Appendix V***.

Chapter 4: Discussion

IPSC quality control strategies

The iPSC quality control strategy presented in this thesis represents a comprehensive approach to ensure the reliability and consistency of these lines. Our workflow for inducing pluripotency, creating a desired mutation with CRISPR/Cas9, and ensuring quality control, represents a good tool for the rapid generation of *bona fide* iPSC lines.

As previously mentioned, the first step of our quality control strategy was to make sure that the cells we received from the hospitals carry the expected mutations. We analyzed the presence of a patient *EHMT1* deletion and observed a reduction in the DNA and RNA signal. While the signals are statistically reduced in patient cells versus control, we did not observe an exact 50% reduction. We have performed numerous experiments of this sort and rarely observe a 50% reduction in haploinsufficiency syndromes, with ranges from ~30% to 70% loss. Similar results are reported in the literature (Talkowski, Rosenfeld et al. 2012) (Skraban, Wells et al. 2017). For instance, Talkowski et al. reported a disruption of gene *TCF4* in monozygotic twins with multiple developmental abnormalities. Two independent experiments on different *TCF4* exons showed reduced signal levels in patients; the observed reductions were on a range of 45% to 60%. It has been reported that RT-PCR efficiency could be affected by several conditions (Svec, Tichopad et al. 2015), for instance, primer design, experiment settings, and contaminants may cause variations in the RT-PCR results (Svec, Tichopad et al. 2015). These may be the reasons we do not observe an exact 50% reduction in the RNA and DNA signals.

Another important aspect of the quality of iPSC lines is their capacity to differentiate into all the three germ layers. To assess the differentiation potential of our lines, we initially used a spontaneous EB differentiation method. Interestingly, after this procedure, most of the lines showed a predominant differentiation potential towards the mesoderm. While this phenomenon was not observed when we directly differentiated one iPSC line toward each of the three germ layers, we still observed that ectoderm-differentiated cells showed higher levels of expression of mesodermal marker.

Several factors could explain our iPSC lines tendency to differentiate towards the mesoderm. The main factor that could be influencing this pattern could be the media we used to allow spontaneous

EB formation. It has been reported that different media compositions influence stem cell differentiation potential (Purpura, Morin et al. 2008). For instance, the use of fetal bovine serum and defined growth factor supplements affects germ layer differentiation (Nostro, Cheng et al. 2008). Bettiol et al. reported increased levels of cardiac differentiation of human embryonic stem cells after supplementation with fetal bovine serum (Bettiol, Sartiani et al. 2007). The fact that our iPSC lines tended to differentiate into mesoderm could be explained by the fact that we cryopreserved the iPSC lines in fetal bovine serum with 10% DMSO, and traces of the serum could have affected the differentiation process.

Another, and perhaps more likely explanation for this phenomenon is that the tendency to differentiate into each of the three germ layers may be dependent on the number of days allotted for spontaneous differentiation. In our case, we differentiated the cells for seven days. Longer periods of spontaneous differentiation may have led to other cell lineages. For instance, Bock et al reported that sixteen-day EB lines showed different differentiation propensities that are cell-line specific. While one of these lines showed the greatest propensity for endoderm differentiation, another line exhibited a greater tendency towards the neural lineage (Bock, Kiskinis et al. 2011). To decipher the reason why we observed preferential differentiation towards the mesoderm, different supplement conditions, and time length experiments could be performed with several lines. However, for quality control, we decided to directly differentiate our iPSC lines into each of the three germ layers and analyze specific markers of each layer. These results showed expression of germ layer-specific markers in all our lines. In summary, the three strategies we presented provided extensive evidence that our iPSC lines had the potential to be differentiated to all three germ layers.

Our quality control analysis indicates that all iPSC lines created in our laboratory are trustworthy and can be further used in several downstream analyses, including differentiation to other cellular lineages and use in drug discovery strategies. Moreover, cell line authentication showed that all our lines were properly labeled and handled, indicating a rigorous methodology for our reprogramming protocols and iPSC storage.

All the samples tested are free of gross chromosomal abnormalities indicating that our episomal vector transfection protocol did not generate gross mutations, as has been observed in other transfection methods involving integrating viruses (Schlaeger, Daheron et al. 2015). There have

been reports of reprogrammed cells whose viral genes were not properly silenced or reactive after long periods of cell culture (Wang and Na 2011). This problem was circumvented in our protocol by avoiding virus usage in pluripotency induction. Our episomal vector strategy did not involve genome modifications, and therefore, the chances of inducing mutations in the genome were low.

Genetic abnormalities are one of the main worries in the iPSC quality control because they are associated with dysregulated growth rates and tumorigenesis and affect the result analysis (Assou, Bouckenheimer et al. 2018). The induction of pluripotency is a relatively long process, ranging from three to five weeks, during which cell colonies are isolated and passaged several times (Lin and Xiao 2017). This process makes iPSC lines more susceptible to acquiring mutations and chromosomal abnormalities ranging from single-nucleotide mutations to aneuploidy (Assou, Bouckenheimer et al. 2018). Several reports have identified common aneuploidy events in iPSC lines, including a gain of chromosome eight, twelve, seventeen, and twenty (Taapken, Nisler et al. 2011). While all our lines are free of gross chromosomal mutations, we cannot affirm with full certainty that they do not carry single-nucleotide mutations or Copy Number Variations (CNVs) that could have originated during the reprogramming process. To exclude the presence of such mutations in our iPSC lines, it would be necessary to perform another kind of experiments, such as genomic hybridization (CGH), SNP genotypification and high throughput sequencing. A cost-benefit analysis should be considered when performing routine iPSC quality control. Though whole-genome sequencing would be a good strategy to discard mutations and chromosomal abnormalities, it is more expensive than G-banded karyotyping (D'Antonio, Woodruff et al. 2017). Detection of CNVs and single-nucleotide mutations using whole-genome sequencing would require high coverage, and thus higher costs, which would not be feasible in routine quality control strategies (Lin and Xiao 2017). We believe our G-banded strategy is a good balance between cost and performance.

IPSCs: a good model to study KS

One great advantage of our study is that we can create iPSC models that are patient-specific. Our patient and control lines could be further employed on different strategies for drug discovery and therapeutics. The iPSC lines of the patients' siblings are good controls for further studies because the genetic variability and heterogeneity observed between siblings are lower than other controls in the population (Germain and Testa 2017).

Interestingly, an iPSC model of KS was recently published by Varga et al. The line was derived from the blood of an 11-year-old KS patient (Varga, Nemes et al. 2016). Similar to our patient 58071-p.R948W, this individual carried a heterozygous *de novo* mutation, but in contrast with our patient, the mutation on the reported patient causes a premature termination codon. The patient reported by this group was another proof of the enormous array of phenotypic variation found among KS patients. While the patient exhibited common characteristics of the syndrome, such as autism, developmental delay, hypotonia, facial dimorphisms, behavioral and psychiatric disorders, the patient did not exhibit ID or epilepsy, making this atypical case very interesting. This is the reason why the authors decided to study the individual further; they differentiated the iPSC line to cortical neurons and studied the pathogenic neuron-specific mechanisms in these cells. Patient-derived neurons showed atypical neurites and arborization, altered cholinergic function and abnormal gene expression profiles (Nagy, Kobolak et al. 2017). Important genes in neuronal function and maturation were dysregulated in patient-derived neurons, including the genes *CUX2* and *SHANK1*, which are involved in the neural specification and synaptic function, respectively. In addition, the authors observed a reduction in the neurite length and the number of neuron roots (Nagy, Kobolak et al. 2017). Interestingly, similar neuronal characteristics were observed in the mouse model of KS, including reduced dendritic arborization and an abnormal number of mature spines (Balemans, Kasri et al. 2013).

The patient iPSC-derived neurons published by Nagy et al. represent a good model of the disease that could be used as an *in vitro* system to study KS and as a model to research new treatments for this syndrome (Nagy, Kobolak et al. 2017). Similarly, our iPSC lines are being differentiated into neurons. We have established a pipeline for neuron differentiation and quality control. We have published several reports showing the quality of our patient-derived neurons (Bell, Peng et al. 2017). As previously mentioned at the beginning of the thesis, our patient iPSC-derived neurons are going to be used in gene expression profile studies.

The relevance of the SET-domain knock-out in the study of KS

The SET-domain is one of the most important domains in EHMT1 protein because it is the catalytic site for lysine methylation. Interestingly, while mice with a full deletion of the *EHMT1* SET-

domain (resulting in catalytically inactive EHMT1 (GLP) protein) are viable, mice with the same mutation in *EHMT2* (G9a) have not been observed (Kramer 2016).

As previously mentioned in the introduction, The GLP/G9a complex is critical during brain development and cell differentiation (Benevento, van de Molengraft et al. 2015). GLP/G9a mediated H3K9 di-methylation has been implicated in long-term memory, as well as fear memory consolidation and formation through fine tuning of neuronal properties in different brain regions (Gupta-Agarwal, Franklin et al. 2012). The SET-domain knock-out we created in this study could help us to further understand the role of *EHMT1* in brain development and disease. Our line could be employed in ChIP-seq studies to analyze the differential patterns of H3K9 methylation across the genome; such analysis could uncover important epigenetic and gene expression profiles in KS.

Also, our SET-domain KO line represents a good isogenic model that could be employed as a control in drug discovery strategies. We could compare the effects of a given drug in the non-genetically engineered 58203-control iPSC line with those of the SET-domain KO line. In this scenario, the 58203-control non-engineered line would serve as a negative control; it would not produce any response to treatment and would serve as a baseline or background, defining compounds with no activity. On the other hand, the SET-domain KO line would serve as a positive control which joined to the 58204-del0.466Mb iPSC line would increase the validity of the assay.

Taken together, our SET-domain KO line is a unique cell line in the study of the KS pathology. Because the line was derived from a patient's phenotypically normal sibling, it represents a good control of the individual genetic background.

Chapter 5: Conclusions and Future Directions

Gene editing technologies such as CRISPR/Cas9, combined with the induction of pluripotency, could be used to generate iPSC models that recapitulate the exact mutation of a patient. Besides, this approach could also be used to correct the patient mutation, both resulting in isogenic control lines. Reduction in intra-species variation is important in research and drug discovery (Germain and Testa 2017), isogenic knock-in or knock-out cell lines could be used for this purpose.

Recently, we proposed a high throughput screening strategy to advance drug discovery in rare monogenic ID syndromes (Hettige, Manzano-Vargas et al. 2018). Our pipeline focuses on gene

dosage syndromes, i.e., syndromes caused by gene haploinsufficiency, using as an example *KMT2D*, the causative gene of Kabuki syndrome (Cheon and Ko 2015). We plan to genetically engineer patient 58204-del0.466Mb iPSC line and control iPSC line 58203-control to introduce a 2A system at the end of the *EHMT1* gene, coupled with GFP (Hettige, Manzano-Vargas et al. 2018). Self-cleaving is a property of 2A peptide-containing proteins, in this way, various proteins could be produced from a single messenger RNA, resulting in a one to one ratio (de Felipe, Luke et al. 2010). We are going to take advantage of this system to introduce GFP as a reporter of *EHMT1* expression. Given that both proteins are going to be expressed in similar quantities, levels of GFP fluorescence would be a read-out of *EHMT1* expression (**Figure 14**) (Hettige, Manzano-Vargas et al. 2018).

Given that the patient cells have only one functional *EHMT1* allele, and the other allele is deleted, GFP expression should be around 50% when compared with controls. On the premise that reduced *EHMT1* dosage leads to KS and that increasing the expression of the wild-type allele on patient cells could help to improve their clinical phenotype, with this read-out, we could test many small molecules in a high throughput screening approach looking for drugs that increase *EHMT1* levels and bring it to normal expression levels. Once we find small molecules that upregulate *EHMT1* expression, we could test them on a pre-clinical phase using a mouse model of KS.

This approach represents a new strategy to advance drug discovery in monogenic ID syndromes. The iPSC models developed during this project will serve as the basis for this, and other approaches. The potential of our iPSC lines to be genetically engineered and differentiated into an array of cell lineages opens the door for new approaches to study and treat KS.

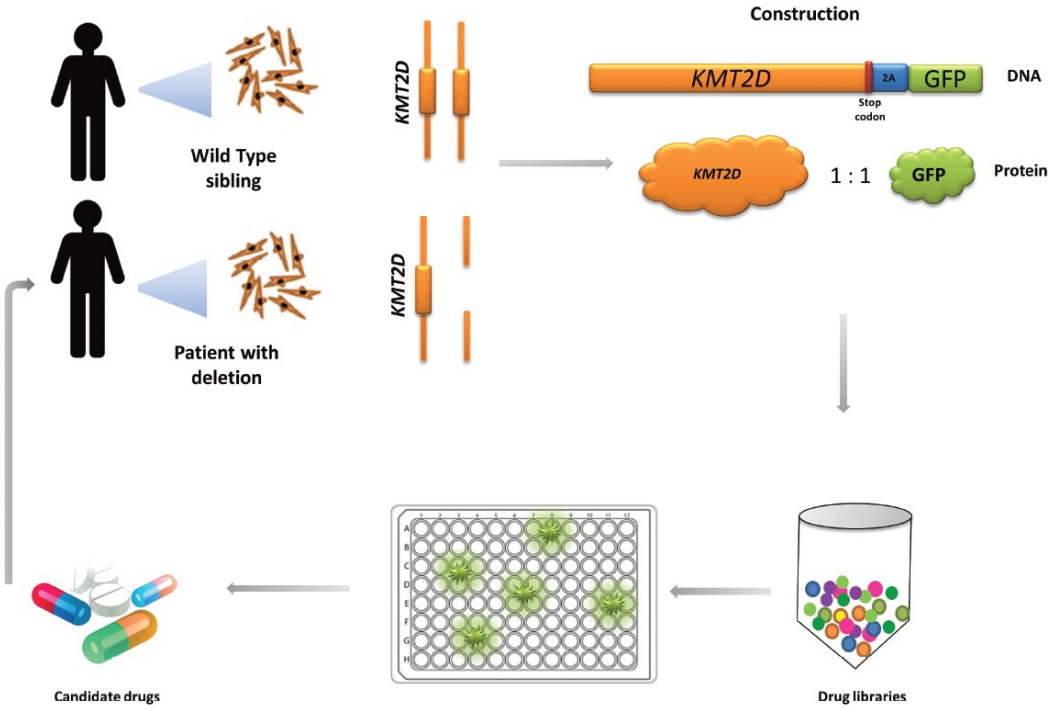


Figure 14. An iPSC and gene editing drug discovery strategy. The drug discovery pipeline presented here is focused on monogenic ID syndromes caused by haploinsufficiency, where one functional allele is still present. Figure taken from (Hettige, Manzano-Vargas et al. 2018).

Chapter 6: References

Abujarour, R., B. Valamehr, M. Robinson, B. Rezner, F. Vranceanu and P. Flynn (2013). "Optimized surface markers for the prospective isolation of high-quality hiPSCs using flow cytometry selection." Sci Rep **3**: 1179.

Adzhubei, I., D. M. Jordan and S. R. Sunyaev (2013). "Predicting functional effect of human missense mutations using PolyPhen-2." Curr Protoc Hum Genet **Chapter 7**: Unit7 20.

Al-Khodor, S., C. T. Price, A. Kalia and Y. Abu Kwaik (2010). "Functional diversity of ankyrin repeats in microbial proteins." Trends Microbiol **18**(3): 132-139.

Assou, S., J. Bouckenheimer and J. De Vos (2018). "Concise Review: Assessing the Genome Integrity of Human Induced Pluripotent Stem Cells: What Quality Control Metrics?" Stem Cells.

Balemans, M. C., M. Ansar, A. R. Oudakker, A. P. van Caam, B. Bakker, E. L. Vitters, P. M. van der Kraan, D. R. de Bruijn, S. M. Janssen, A. J. Kuipers, M. M. Huibers, E. M. Maliepaard, X. F. Walboomers, M. Benevento, N. Nadif Kasri, T. Kleefstra, H. Zhou, C. E. Van der Zee and H. van Bokhoven (2014). "Reduced Euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in Kleefstra syndrome mice." Dev Biol **386**(2): 395-407.

Balemans, M. C., M. M. Huibers, N. W. Eikelenboom, A. J. Kuipers, R. C. van Summeren, M. M. Pijpers, M. Tachibana, Y. Shinkai, H. van Bokhoven and C. E. Van der Zee (2010). "Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice." Behav Brain Res **208**(1): 47-55.

Balemans, M. C., N. N. Kasri, M. V. Kopanitsa, N. O. Afinowi, G. Ramakers, T. A. Peters, A. J. Beynon, S. M. Janssen, R. C. van Summeren, J. M. Eeftens, N. Eikelenboom, M. Benevento, M. Tachibana, Y. Shinkai, T. Kleefstra, H. van Bokhoven and C. E. Van der Zee (2013). "Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome." Hum Mol Genet **22**(5): 852-866.

Beers, J., K. L. Linask, J. A. Chen, L. I. Siniscalchi, Y. Lin, W. Zheng, M. Rao and G. Chen (2015). "A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture." Sci Rep **5**: 11319.

Bell, S., H. Peng, L. Crapper, I. Kolobova, G. Maussion, C. Vasuta, V. Yerko, T. P. Wong and C. Ernst (2016). "A Rapid Pipeline to Model Rare Neurodevelopmental Disorders with Simultaneous CRISPR/Cas9 Gene Editing." Stem Cells Transl Med.

Bell, S., H. Peng, L. Crapper, I. Kolobova, G. Maussion, C. Vasuta, V. Yerko, T. P. Wong and C. Ernst (2017). "A Rapid Pipeline to Model Rare Neurodevelopmental Disorders with Simultaneous CRISPR/Cas9 Gene Editing." Stem Cells Transl Med **6**(3): 886-896.

Benevento, M., M. van de Molengraft, R. van Westen, H. van Bokhoven and N. N. Kasri (2015). "The role of chromatin repressive marks in cognition and disease: A focus on the repressive complex GLP/G9a." Neurobiol Learn Mem **124**: 88-96.

Bettiol, E., L. Sartiani, L. Chicha, K. H. Krause, E. Cerbai and M. E. Jaconi (2007). "Fetal bovine serum enables cardiac differentiation of human embryonic stem cells." Differentiation **75**(8): 669-681.

Blackburn, P. R., A. Tischer, M. T. Zimmermann, J. L. Kemppainen, S. Sastry, A. E. Knight Johnson, M. A. Cousin, N. J. Boczek, G. Oliver, V. K. Misra, R. H. Gavrilova, G. Lomberk, M. Auton, R. Urrutia and E. W. Klee (2017). "A Novel Kleefstra Syndrome-associated Variant That Affects the Conserved TPLX Motif within the Ankyrin Repeat of EHMT1 Leads to Abnormal Protein Folding." J Biol Chem **292**(9): 3866-3876.

Bock, C., E. Kiskinis, G. Verstappen, H. Gu, G. Boulting, Z. D. Smith, M. Ziller, G. F. Croft, M. W. Amoroso, D. H. Oakley, A. Gnirke, K. Eggan and A. Meissner (2011). "Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines." Cell **144**(3): 439-452.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, G. W. Bell, A. P. Otte, M. Vidal, D. K. Gifford, R. A. Young and R. Jaenisch (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-353.
Briggs, R. and T. J. King (1952). "Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs." Proc Natl Acad Sci U S A **38**(5): 455-463.
Brimble, S. N., E. S. Sherrer, E. W. Uhl, E. Wang, S. Kelly, A. H. Merrill, Jr., A. J. Robins and T. C. Schulz (2007). "The cell surface glycosphingolipids SSEA-3 and SSEA-4 are not essential for human ESC pluripotency." Stem Cells **25**(1): 54-62.
Cai, K. Q., C. D. Capo-Chichi, M. E. Rula, D. H. Yang and X. X. Xu (2008). "Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis." Dev Dyn **237**(10): 2820-2829.
Carvill, G. L. and H. C. Mefford (2015). "Next-Generation Sequencing in Intellectual Disability." J Pediatr Genet **4**(3): 128-135.
Chadman, K. K., M. Yang and J. N. Crawley (2009). "Criteria for validating mouse models of psychiatric diseases." Am J Med Genet B Neuropsychiatr Genet **150B**(1): 1-11.
Chan, E. M., S. Ratanasirintrawoot, I. H. Park, P. D. Manos, Y. H. Loh, H. Huo, J. D. Miller, O. Hartung, J. Rho, T. A. Ince, G. Q. Daley and T. M. Schlaeger (2009). "Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells." Nat Biotechnol **27**(11): 1033-1037.
Chen, I. P., K. Fukuda, N. Fusaki, A. Iida, M. Hasegawa, A. Lichtler and E. J. Reichenberger (2013). "Induced pluripotent stem cell reprogramming by integration-free Sendai virus vectors from peripheral blood of patients with craniometaphyseal dysplasia." Cell Reprogram **15**(6): 503-513.
Cheon, C. K. and J. M. Ko (2015). "Kabuki syndrome: clinical and molecular characteristics." Korean J Pediatr **58**(9): 317-324.
Chiurazzi, P. and F. Pirozzi (2016). "Advances in understanding - genetic basis of intellectual disability." F1000Res **5**.
Collins, R. E., J. P. Northrop, J. R. Horton, D. Y. Lee, X. Zhang, M. R. Stallcup and X. Cheng (2008). "The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules." Nat Struct Mol Biol **15**(3): 245-250.
D'Antonio, M., G. Woodruff, J. L. Nathanson, A. D'Antonio-Chronowska, A. Arias, H. Matsui, R. Williams, C. Herrera, S. M. Reyna, G. W. Yeo, L. S. B. Goldstein, A. D. Panopoulos and K. A. Frazer (2017). "High-Throughput and Cost-Effective Characterization of Induced Pluripotent Stem Cells." Stem Cell Reports **8**(4): 1101-1111.
Darville, H., A. Poulet, F. Rodet-Amsellem, L. Chatrousse, J. Pernelle, C. Boissart, D. Heron, C. Nava, A. Perrier, M. Jarrige, F. Coge, M. J. Millan, T. Bourgeron, M. Peschanski, R. Delorme and A. Benchoua (2016). "Human Pluripotent Stem Cell-derived Cortical Neurons for High Throughput Medication Screening in Autism: A Proof of Concept Study in SHANK3 Haploinsufficiency Syndrome." EBioMedicine **9**: 293-305.
Davis, R. L., H. Weintraub and A. B. Lassar (1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts." Cell **51**(6): 987-1000.
de Boer, A., K. Vermeulen, J. I. M. Egger, J. G. E. Janzing, N. de Leeuw, H. E. Veenstra-Knol, N. S. den Hollander, H. van Bokhoven, W. Staal and T. Kleefstra (2018). "EHMT1 mosaicism in apparently unaffected parents is associated with autism spectrum disorder and neurocognitive dysfunction." Mol Autism **9**: 5.
de Felipe, P., G. A. Luke, J. D. Brown and M. D. Ryan (2010). "Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences." Biotechnol J **5**(2): 213-223.
Deng, B. (2017). "Mouse models and induced pluripotent stem cells in researching psychiatric disorders." Stem Cell Investig **4**: 62.
Dimos, J. T., K. T. Rodolfa, K. K. Niakan, L. M. Weisenthal, H. Mitsumoto, W. Chung, G. F. Croft, G. Saphier, R. Leibel, R. Goland, H. Wichterle, C. E. Henderson and K. Eggan (2008). "Induced pluripotent stem cells

generated from patients with ALS can be differentiated into motor neurons." Science **321**(5893): 1218-1221.
Draper, J. S., C. Pigott, J. A. Thomson and P. W. Andrews (2002). "Surface antigens of human embryonic stem cells: changes upon differentiation in culture." J Anat **200**(Pt 3): 249-258.
Drozd, A. M., M. P. Walczak, S. Piaskowski, E. Stoczynska-Fidelus, P. Rieske and D. P. Grzela (2015). "Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system." Stem Cell Res Ther **6**: 122.
Ebert, A. D., P. Liang and J. C. Wu (2012). "Induced pluripotent stem cells as a disease modeling and drug screening platform." J Cardiovasc Pharmacol **60**(4): 408-416.
Ellis, J. and M. Bhatia (2011). "iPSC technology: platform for drug discovery. Point." Clin Pharmacol Ther **89**(5): 639-641.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.
Fergus, J., R. Quintanilla and U. Lakshmipathy (2016). "Characterizing Pluripotent Stem Cells Using the TaqMan(R) hPSC Scorecard(TM) Panel." Methods Mol Biol **1307**: 25-37.
Font-Montgomery, E., D. D. Weaver, L. Walsh, C. Christensen and V. C. Thurston (2004). "Clinical and cytogenetic manifestations of subtelomeric aberrations: Report of six cases." Birth Defects Res A Clin Mol Teratol **70**(6): 408-415.
Germain, P. L. and G. Testa (2017). "Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling." Stem Cell Reports **8**(6): 1784-1796.
Gertow, K., S. Przyborski, J. F. Loring, J. M. Auerbach, O. Epifano, T. Otonkoski, I. Damjanov and L. Ahrlund-Richter (2007). "Isolation of human embryonic stem cell-derived teratomas for the assessment of pluripotency." Curr Protoc Stem Cell Biol **Chapter 1**: Unit1B 4.
Gonzalez, F. and D. Huangfu (2016). "Mechanisms underlying the formation of induced pluripotent stem cells." Wiley Interdiscip Rev Dev Biol **5**(1): 39-65.
Gupta-Agarwal, S., A. V. Franklin, T. Deramus, M. Wheelock, R. L. Davis, L. L. McMahon and F. D. Lubin (2012). "G9a/GLP histone lysine dimethyltransferase complex activity in the hippocampus and the entorhinal cortex is required for gene activation and silencing during memory consolidation." J Neurosci **32**(16): 5440-5453.
Gurdon, J. B., R. A. Laskey and O. R. Reeves (1975). "The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs." J Embryol Exp Morphol **34**(1): 93-112.
Harada, N., R. Visser, A. Dawson, M. Fukamachi, M. Iwakoshi, N. Okamoto, T. Kishino, N. Niikawa and N. Matsumoto (2004). "A 1-Mb critical region in six patients with 9q34.3 terminal deletion syndrome." J Hum Genet **49**(8): 440-444.
Hettige, N. C., K. Manzano-Vargas, M. Jefri and C. Ernst (2018). "Strategies to Advance Drug Discovery in Rare Monogenic Intellectual Disability Syndromes." Int J Neuropsychopharmacol **21**(3): 201-206.
Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.
Kanherkar, R. R., N. Bhatia-Dey, E. Makarev and A. B. Csoka (2014). "Cellular reprogramming for understanding and treating human disease." Front Cell Dev Biol **2**: 67.
Kannagi, R., N. A. Cochran, F. Ishigami, S. Hakomori, P. W. Andrews, B. B. Knowles and D. Solter (1983). "Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells." EMBO J **2**(12): 2355-2361.
Kaufmann, M., A. Schuffenhauer, I. Fruh, J. Klein, A. Thiemeyer, P. Rigo, B. Gomez-Mancilla, V. Heidinger-Millot, T. Bouwmeester, U. Schopfer, M. Mueller, B. D. Fodor and A. Cobos-Correa (2015). "High-Throughput Screening Using iPSC-Derived Neuronal Progenitors to Identify Compounds Counteracting Epigenetic Gene Silencing in Fragile X Syndrome." J Biomol Screen **20**(9): 1101-1111.

Kim, D., C. H. Kim, J. I. Moon, Y. G. Chung, M. Y. Chang, B. S. Han, S. Ko, E. Yang, K. Y. Cha, R. Lanza and K. S. Kim (2009). "Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins." Cell Stem Cell **4**(6): 472-476.
Kim, J. S., H. W. Choi, S. Choi and J. T. Do (2011). "Reprogrammed pluripotent stem cells from somatic cells." Int J Stem Cells **4**(1): 1-8.
Kleefstra, T., M. Smidt, M. J. Banning, A. R. Oudakker, H. Van Esch, A. P. de Brouwer, W. Nillesen, E. A. Sistermans, B. C. Hamel, D. de Bruijn, J. P. Fryns, H. G. Yntema, H. G. Brunner, B. B. de Vries and H. van Bokhoven (2005). "Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome." J Med Genet **42**(4): 299-306.
Kleefstra, T., W. A. van Zelst-Stams, W. M. Nillesen, V. Cormier-Daire, G. Houge, N. Foulds, M. van Dooren, M. H. Willemsen, R. Pfundt, A. Turner, M. Wilson, J. McGaughran, A. Rauch, M. Zenker, M. P. Adam, M. Innes, C. Davies, A. G. Lopez, R. Casalone, A. Weber, L. A. Brueton, A. D. Navarro, M. P. Bralo, H. Venselaar, S. P. Stegmann, H. G. Yntema, H. van Bokhoven and H. G. Brunner (2009). "Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype." J Med Genet **46**(9): 598-606.
Kramer, J. M. (2016). "Regulation of cell differentiation and function by the euchromatin histone methyltransferases G9a and GLP." Biochem Cell Biol **94**(1): 26-32.
Kramer, J. M., K. Kochinke, M. A. Oortveld, H. Marks, D. Kramer, E. K. de Jong, Z. Asztalos, J. T. Westwood, H. G. Stunnenberg, M. B. Sokolowski, K. Keleman, H. Zhou, H. van Bokhoven and A. Schenck (2011). "Epigenetic regulation of learning and memory by Drosophila EHMT/G9a." PLoS Biol **9**(1): e1000569.
Lefort, N., M. Feyeux, C. Bas, O. Feraud, A. Bennaceur-Griscelli, G. Tachdjian, M. Peschanski and A. L. Perrier (2008). "Human embryonic stem cells reveal recurrent genomic instability at 20q11.21." Nat Biotechnol **26**(12): 1364-1366.
Lin, K. and A. Z. Xiao (2017). "Quality control towards the application of induced pluripotent stem cells." Curr Opin Genet Dev **46**: 164-169.
Lin, S. L., D. C. Chang, S. Chang-Lin, C. H. Lin, D. T. Wu, D. T. Chen and S. Y. Ying (2008). "Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state." RNA **14**(10): 2115-2124.
Liu, N., Z. Zhang, H. Wu, Y. Jiang, L. Meng, J. Xiong, Z. Zhao, X. Zhou, J. Li, H. Li, Y. Zheng, S. Chen, T. Cai, S. Gao and B. Zhu (2015). "Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability." Genes Dev **29**(4): 379-393.
Malik, N. and M. S. Rao (2013). "A review of the methods for human iPSC derivation." Methods Mol Biol **997**: 23-33.
Marchetto, M. C., C. Carromeu, A. Acab, D. Yu, G. W. Yeo, Y. Mu, G. Chen, F. H. Gage and A. R. Muotri (2010). "A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells." Cell **143**(4): 527-539.
Martinez-Lage, M., R. Torres-Ruiz and S. Rodriguez-Perales (2017). "CRISPR/Cas9 Technology: Applications and Human Disease Modeling." Prog Mol Biol Transl Sci **152**: 23-48.
Martins-Taylor, K., B. S. Nisler, S. M. Taapken, T. Compton, L. Crandall, K. D. Montgomery, M. Lalande and R. H. Xu (2011). "Recurrent copy number variations in human induced pluripotent stem cells." Nat Biotechnol **29**(6): 488-491.
Martins-Taylor, K. and R. H. Xu (2012). "Concise review: Genomic stability of human induced pluripotent stem cells." Stem Cells **30**(1): 22-27.
Maulik, P. K., M. N. Mascarenhas, C. D. Mathers, T. Dua and S. Saxena (2011). "Prevalence of intellectual disability: a meta-analysis of population-based studies." Res Dev Disabil **32**(2): 419-436.
Meyer, N. and L. Z. Penn (2008). "Reflecting on 25 years with MYC." Nat Rev Cancer **8**(12): 976-990.
Miyamoto, K., M. Akiyama, F. Tamura, M. Isomi, H. Yamakawa, T. Sadahiro, N. Muraoka, H. Kojima, S. Haginiwa, S. Kurotsu, H. Tani, L. Wang, L. Qian, M. Inoue, Y. Ide, J. Kurokawa, T. Yamamoto, T. Seki, R.

Aeba, H. Yamagishi, K. Fukuda and M. Ieda (2018). "Direct In Vivo Reprogramming with Sendai Virus Vectors Improves Cardiac Function after Myocardial Infarction." Cell Stem Cell **22**(1): 91-103 e105.
Miyoshi, N., H. Ishii, H. Nagano, N. Haraguchi, D. L. Dewi, Y. Kano, S. Nishikawa, M. Tanemura, K. Mimori, F. Tanaka, T. Saito, J. Nishimura, I. Takemasa, T. Mizushima, M. Ikeda, H. Yamamoto, M. Sekimoto, Y. Doki and M. Mori (2011). "Reprogramming of mouse and human cells to pluripotency using mature microRNAs." Cell Stem Cell **8**(6): 633-638.
Morali, O. G., A. Jouneau, K. J. McLaughlin, J. P. Thiery and L. Larue (2000). "IGF-II promotes mesoderm formation." Dev Biol **227**(1): 133-145.
Muller, F. J., B. Brandl and J. F. Loring (2008). Assessment of human pluripotent stem cells with PluriTest. StemBook. Cambridge (MA).
Nagy, J., J. Kobolak, S. Berzsenyi, Z. Abraham, H. X. Avci, I. Bock, Z. Bekes, B. Hodoscsek, A. Chandrasekaran, A. Teglasi, P. Dezso, B. Kovanyi, E. T. Voros, L. Fodor, T. Szel, K. Nemeth, A. Balazs, A. Dinnyes, B. Lendvai, G. Levay and V. Roman (2017). "Altered neurite morphology and cholinergic function of induced pluripotent stem cell-derived neurons from a patient with Kleefstra syndrome and autism." Transl Psychiatry **7**(7): e1179.
Natunen, S., T. Satomaa, V. Pitkanen, H. Salo, M. Mikkola, J. Natunen, T. Otonkoski and L. Valmu (2011). "The binding specificity of the marker antibodies Tra-1-60 and Tra-1-81 reveals a novel pluripotency-associated type 1 lactosamine epitope." Glycobiology **21**(9): 1125-1130.
Ng, P. C. and S. Henikoff (2003). "SIFT: Predicting amino acid changes that affect protein function." Nucleic Acids Res **31**(13): 3812-3814.
Ng, S. B., A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad and J. Shendure (2010). "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." Nat Genet **42**(9): 790-793.
Nguyen, H. N., B. Byers, B. Cord, A. Shcheglovitov, J. Byrne, P. Gujar, K. Kee, B. Schule, R. E. Dolmetsch, W. Langston, T. D. Palmer and R. R. Pera (2011). "LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress." Cell Stem Cell **8**(3): 267-280.
Nostro, M. C., X. Cheng, G. M. Keller and P. Gadue (2008). "Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood." Cell Stem Cell **2**(1): 60-71.
Okita, K., T. Ichisaka and S. Yamanaka (2007). "Generation of germline-competent induced pluripotent stem cells." Nature **448**(7151): 313-317.
Ozbek, E., G. Calik, A. Otunctemur, T. Aliskan, S. Cakir, M. Dursun and A. Somay (2012). "Stem cell markers aldehyde dehydrogenase type 1 and nestin expressions in renal cell cancer." Arch Ital Urol Androl **84**(1): 7-11.
Pires, D. E., D. B. Ascher and T. L. Blundell (2014). "mCSM: predicting the effects of mutations in proteins using graph-based signatures." Bioinformatics **30**(3): 335-342.
Purpura, K. A., J. Morin and P. W. Zandstra (2008). "Analysis of the temporal and concentration-dependent effects of BMP-4, VEGF, and TPO on development of embryonic stem cell-derived mesoderm and blood progenitors in a defined, serum-free media." Exp Hematol **36**(9): 1186-1198.
Ramlee, M. K., J. Wang, W. X. Toh and S. Li (2016). "Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene." Genes (Basel) **7**(8).
Rodrigues, C. H., D. E. Pires and D. B. Ascher (2018). "DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability." Nucleic Acids Res **46**(W1): W350-W355.
Rump, A., L. Hildebrand, A. Tzschach, R. Ullmann, E. Schrock and D. Mitter (2013). "A mosaic maternal splice donor mutation in the EHMT1 gene leads to aberrant transcripts and to Kleefstra syndrome in the offspring." Eur J Hum Genet **21**(8): 887-890.

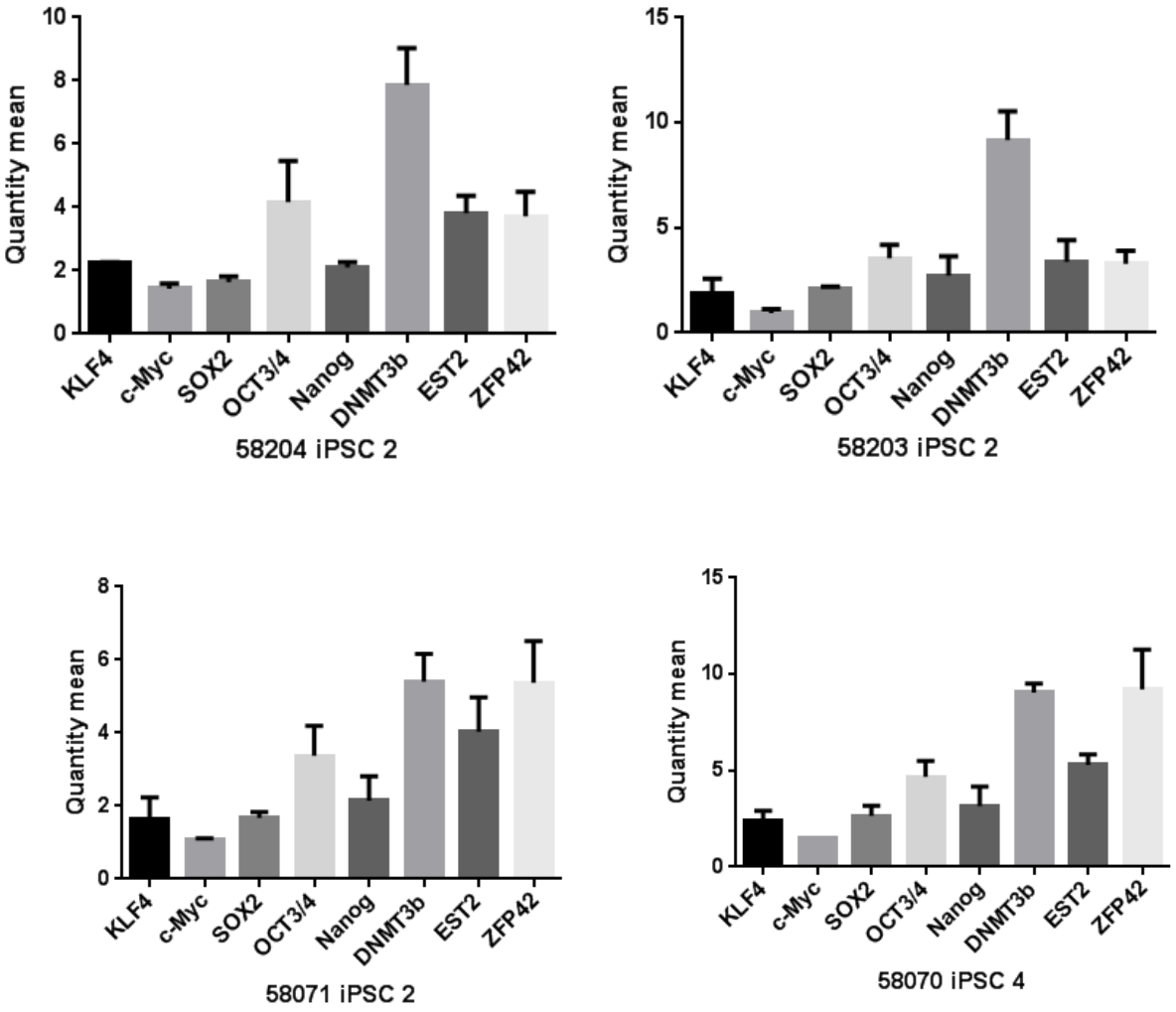
Sakurai, H., T. Era, L. M. Jakt, M. Okada, S. Nakai, S. Nishikawa and S. Nishikawa (2006). "In vitro modeling of paraxial and lateral mesoderm differentiation reveals early reversibility." Stem Cells **24**(3): 575-586.
Samavarchi-Tehrani, P., A. Golipour, L. David, H. K. Sung, T. A. Beyer, A. Datti, K. Woltjen, A. Nagy and J. L. Wrana (2010). "Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming." Cell Stem Cell **7**(1): 64-77.
Schimmenti, L. A., S. A. Berry, M. Tuchman and B. Hirsch (1994). "Infant with multiple congenital anomalies and deletion (9)(q34.3)." Am J Med Genet **51**(2): 140-142.
Schlaeger, T. M., L. Daheron, T. R. Brickler, S. Entwisle, K. Chan, A. Cianci, A. DeVine, A. Ettenger, K. Fitzgerald, M. Godfrey, D. Gupta, J. McPherson, P. Malwadkar, M. Gupta, B. Bell, A. Doi, N. Jung, X. Li, M. S. Lynes, E. Brookes, A. B. Cherry, D. Demirbas, A. M. Tsankov, L. I. Zon, L. L. Rubin, A. P. Feinberg, A. Meissner, C. A. Cowan and G. Q. Daley (2015). "A comparison of non-integrating reprogramming methods." Nat Biotechnol **33**(1): 58-63.
Seguin, C. A., J. S. Draper, A. Nagy and J. Rossant (2008). "Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells." Cell Stem Cell **3**(2): 182-195.
Shi, W., H. Wang, G. Pan, Y. Geng, Y. Guo and D. Pei (2006). "Regulation of the pluripotency marker Rex-1 by Nanog and Sox2." J Biol Chem **281**(33): 23319-23325.
Singh, V. K., M. Kalsan, N. Kumar, A. Saini and R. Chandra (2015). "Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery." Front Cell Dev Biol **3**: 2.
Skraban, C. M., C. F. Wells, P. Markose, M. T. Cho, A. I. Nesbitt, P. Y. B. Au, A. Begtrup, J. A. Bernat, L. M. Bird, K. Cao, A. P. M. de Brouwer, E. H. Denenberg, G. Douglas, K. M. Gibson, K. Grand, A. Goldenberg, A. M. Innes, J. Juusola, M. Kempers, E. Kinning, D. M. Markie, M. M. Owens, K. Payne, R. Person, R. Pfundt, A. Stocco, C. L. S. Turner, N. E. Verbeek, L. E. Walsh, T. C. Warner, P. G. Wheeler, D. Wieczorek, A. B. Wilkens, E. Zonneveld-Huijssoon, S. Deciphering Developmental Disorders, T. Kleefstra, S. P. Robertson, A. Santani, K. L. I. van Gassen and M. A. Deardorff (2017). "WDR26 Haploinsufficiency Causes a Recognizable Syndrome of Intellectual Disability, Seizures, Abnormal Gait, and Distinctive Facial Features." Am J Hum Genet **101**(1): 139-148.
Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl and D. Rogers (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.
Son, M. Y., H. Choi, Y. M. Han and Y. S. Cho (2013). "Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency." Stem Cells **31**(11): 2374-2387.
Stadtfeld, M. and K. Hochedlinger (2010). "Induced pluripotency: history, mechanisms, and applications." Genes Dev **24**(20): 2239-2263.
Stewart, D. R. and T. Kleefstra (2007). "The chromosome 9q subtelomere deletion syndrome." Am J Med Genet C Semin Med Genet **145C**(4): 383-392.
Sullivan, S. E. and T. L. Young-Pearse (2017). "Induced pluripotent stem cells as a discovery tool for Alzheimers disease." Brain Res **1656**: 98-106.
Svec, D., A. Tichopad, V. Novosadova, M. W. Pfaffl and M. Kubista (2015). "How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments." Biomol Detect Quantif **3**: 9-16.
Taapken, S. M., B. S. Nisler, M. A. Newton, T. L. Sampsell-Barron, K. A. Leonhard, E. M. McIntire and K. D. Montgomery (2011). "Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells." Nat Biotechnol **29**(4): 313-314.
Tachibana, M., J. Ueda, M. Fukuda, N. Takeda, T. Ohta, H. Iwanari, T. Sakihama, T. Kodama, T. Hamakubo and Y. Shinkai (2005). "Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9." Genes Dev **19**(7): 815-826.
Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda and S. Yamanaka (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
Talkowski, M. E., J. A. Rosenfeld, I. Blumenthal, V. Pillalamarri, C. Chiang, A. Heilbut, C. Ernst, C. Hanscom, E. Rossin, A. M. Lindgren, S. Pereira, D. Ruderfer, A. Kirby, S. Ripke, D. J. Harris, J. H. Lee, K. Ha, H. G. Kim, B. D. Solomon, A. L. Gropman, D. Lucente, K. Sims, T. K. Ohsumi, M. L. Borowsky, S. Loranger, B. Quade, K. Lage, J. Miles, B. L. Wu, Y. Shen, B. Neale, L. G. Shaffer, M. J. Daly, C. C. Morton and J. F. Gusella (2012). "Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries." Cell **149**(3): 525-537.
Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.
Tsankov, A. M., V. Akopian, R. Pop, S. Chetty, C. A. Gifford, L. Daheron, N. M. Tsankova and A. Meissner (2015). "A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells." Nat Biotechnol **33**(11): 1182-1192.
Varga, E., C. Nemes, Z. Tancos, I. Bock, S. Berzsenyi, G. Levay, V. Roman, J. Kobolak and A. Dinnyes (2016). "Establishment of EHMT1 mutant induced pluripotent stem cell (iPSC) line from a 11-year-old Kleefstra syndrome (KS) patient with autism and normal intellectual performance." Stem Cell Res **17**(3): 531-533.
Verhoeven, W. M., J. I. Egger, K. Vermeulen, B. P. van de Warrenburg and T. Kleefstra (2011). "Kleefstra syndrome in three adult patients: further delineation of the behavioral and neurological phenotype shows aspects of a neurodegenerative course." Am J Med Genet A **155A**(10): 2409-2415.
Verhoeven, W. M. A., T. Kleefstra and J. I. M. Egger (2010). "Behavioral phenotype in the 9q subtelomeric deletion syndrome: a report about two adult patients." Am J Med Genet B Neuropsychiatr Genet **153B**(2): 536-541.
Wang, H., M. La Russa and L. S. Qi (2016). "CRISPR/Cas9 in Genome Editing and Beyond." Annu Rev Biochem **85**: 227-264.
Wang, P., R. Mokhtari, E. Pedrosa, M. Kirschenbaum, C. Bayrak, D. Zheng and H. M. Lachman (2017). "CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in cerebral organoids derived from iPS cells." Mol Autism **8**: 11.
Wang, P. and J. Na (2011). "Mechanism and methods to induce pluripotency." Protein Cell **2**(10): 792-799.
Warren, L., P. D. Manos, T. Ahfeldt, Y. H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger and D. J. Rossi (2010). "Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA." Cell Stem Cell **7**(5): 618-630.
Willemsen, M. H., G. Beunders, M. Callaghan, N. de Leeuw, W. M. Nillesen, H. G. Yntema, J. M. van Hagen, A. W. Nieuwint, N. Morrison, S. T. Keijzers-Vloet, A. Hoischen, H. G. Brunner, J. Tolmie and T. Kleefstra (2011). "Familial Kleefstra syndrome due to maternal somatic mosaicism for interstitial 9q34.3 microdeletions." Clin Genet **80**(1): 31-38.
Willemsen, M. H., A. T. Vulto-van Silfhout, W. M. Nillesen, W. M. Wissink-Lindhout, H. van Bokhoven, N. Philip, E. M. Berry-Kravis, U. Kini, C. M. van Ravenswaaij-Arts, B. Delle Chiaie, A. M. Innes, G. Houge, T. Kosonen, K. Cremer, M. Fannemel, A. Stray-Pedersen, W. Reardon, J. Ignatius, K. Lachlan, C. Mircher, P. T. Helderman van den Enden, M. Mastebroek, P. E. Cohn-Hokke, H. G. Yntema, S. Drunat and T. Kleefstra (2012). "Update on Kleefstra Syndrome." Mol Syndromol **2**(3-5): 202-212.
Wilmut, I., A. E. Schnieke, J. McWhir, A. J. Kind and K. H. Campbell (1997). "Viable offspring derived from fetal and adult mammalian cells." Nature **385**(6619): 810-813.
Windrem, M. S., M. Osipovitch, Z. Liu, J. Bates, D. Chandler-Militello, L. Zou, J. Munir, S. Schanz, K. McCoy, R. H. Miller, S. Wang, M. Nedergaard, R. L. Findling, P. J. Tesar and S. A. Goldman (2017). "Human iPSC Glial Mouse Chimeras Reveal Glial Contributions to Schizophrenia." Cell Stem Cell **21**(2): 195-208 e196.
Xie, H., M. Ye, R. Feng and T. Graf (2004). "Stepwise reprogramming of B cells into macrophages." Cell **117**(5): 663-676.

Yamanaka, S. (2012). "Induced pluripotent stem cells: past, present, and future." Cell Stem Cell **10**(6): 678-684.
Yu, J., K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, Slukvin, II and J. A. Thomson (2009). "Human induced pluripotent stem cells free of vector and transgene sequences." Science **324**(5928): 797-801.
Yu, J., M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II and J. A. Thomson (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.
Zhang, P., J. Li, Z. Tan, C. Wang, T. Liu, L. Chen, J. Yong, W. Jiang, X. Sun, L. Du, M. Ding and H. Deng (2008). "Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells." Blood **111**(4): 1933-1941.
Zhang, X., L. Wang, M. Liu and D. Li (2017). "CRISPR/Cas9 system: a powerful technology for in vivo and ex vivo gene therapy." Sci China Life Sci **60**(5): 468-475.
Zhang, X. Q., G. B. Afink, K. Svensson, J. J. Jacobs, T. Gunther, K. Forsberg-Nilsson, E. J. van Zoelen, B. Westermark and M. Nister (1998). "Specific expression in mouse mesoderm- and neural crest-derived tissues of a human PDGFRA promoter/lacZ transgene." Mech Dev **70**(1-2): 167-180.
Zhou, W. and C. R. Freed (2009). "Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells." Stem Cells **27**(11): 2667-2674.

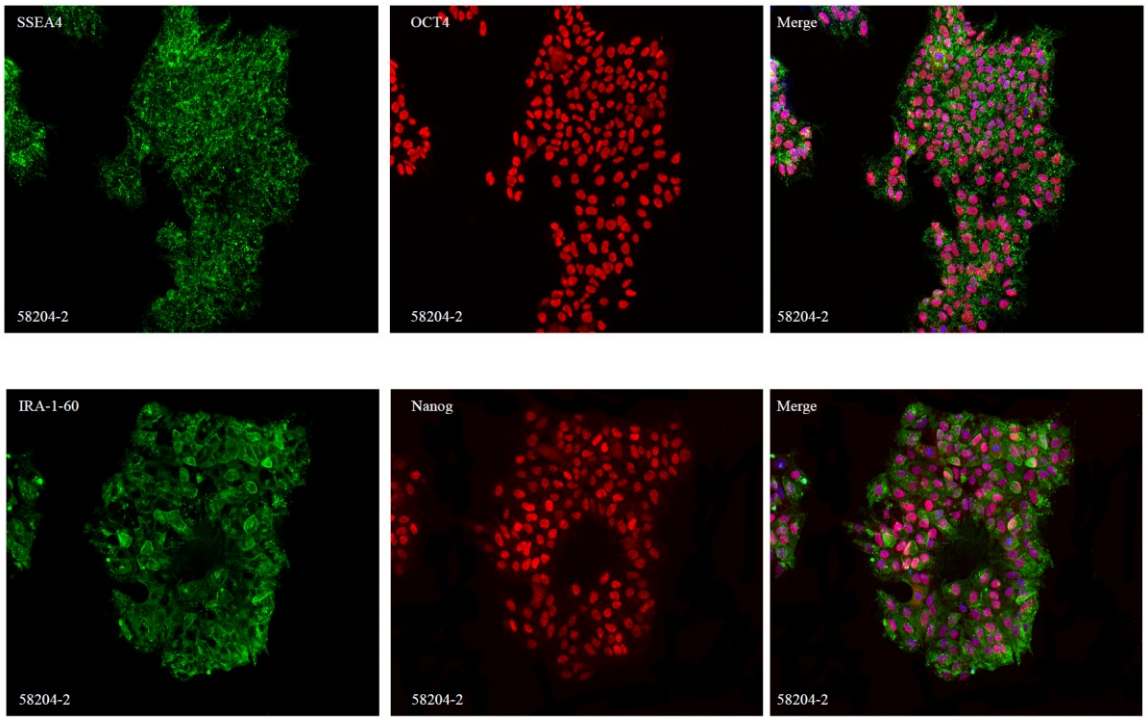
Appendices

Appendix I: Endogenous expression of pluripotency markers in the iPSC lines

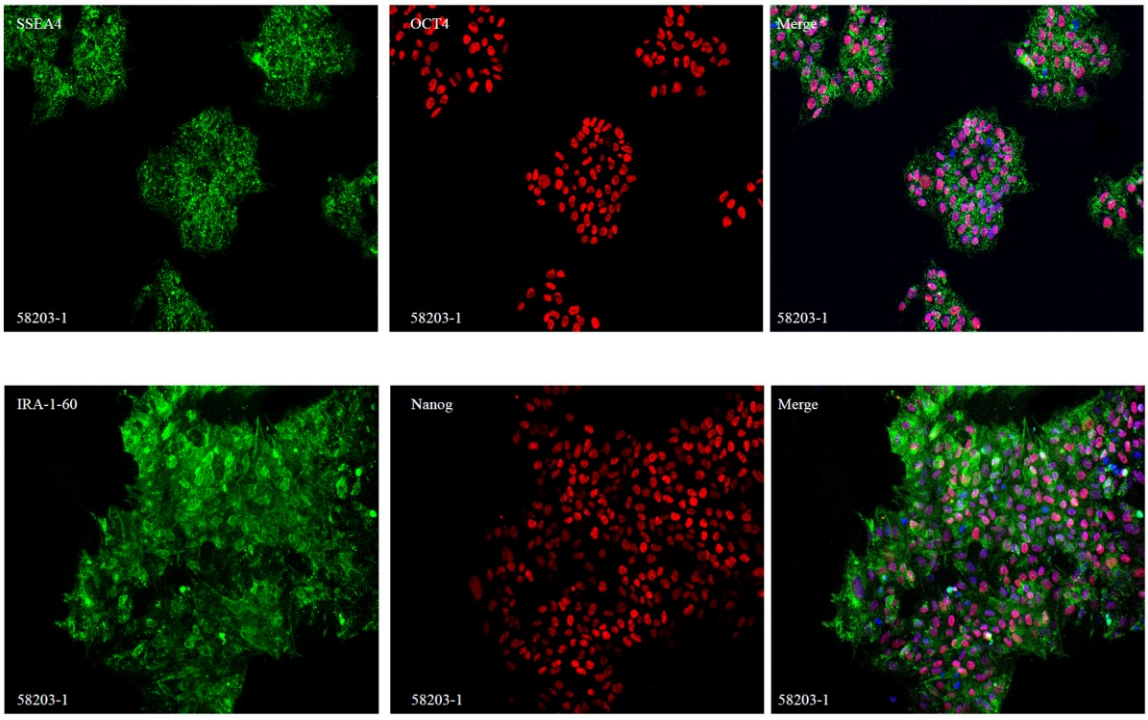


Endogenous expression of pluripotency markers in our iPSC lines. Quantification of the expression of various pluripotency gene markers. Error bars indicate standard error of the mean (SEM).

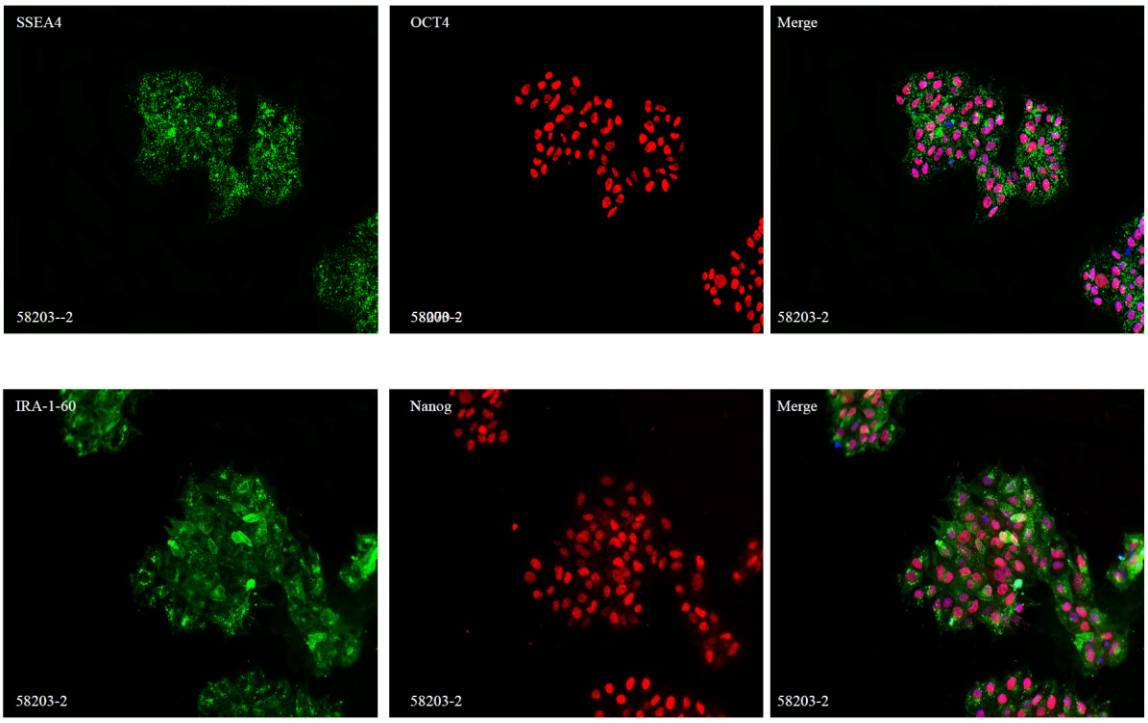
Appendix II: Immunocytochemistry of pluripotency markers
Patient 58204 iPSC 2



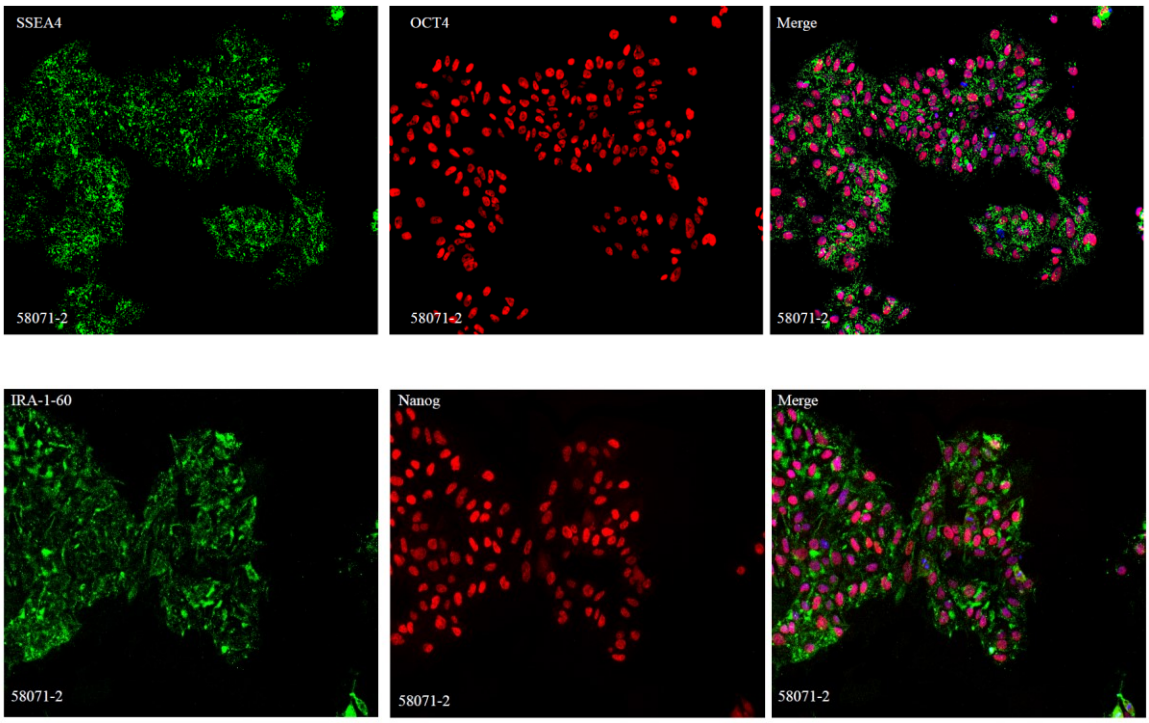
Individual 58203 iPSC 1



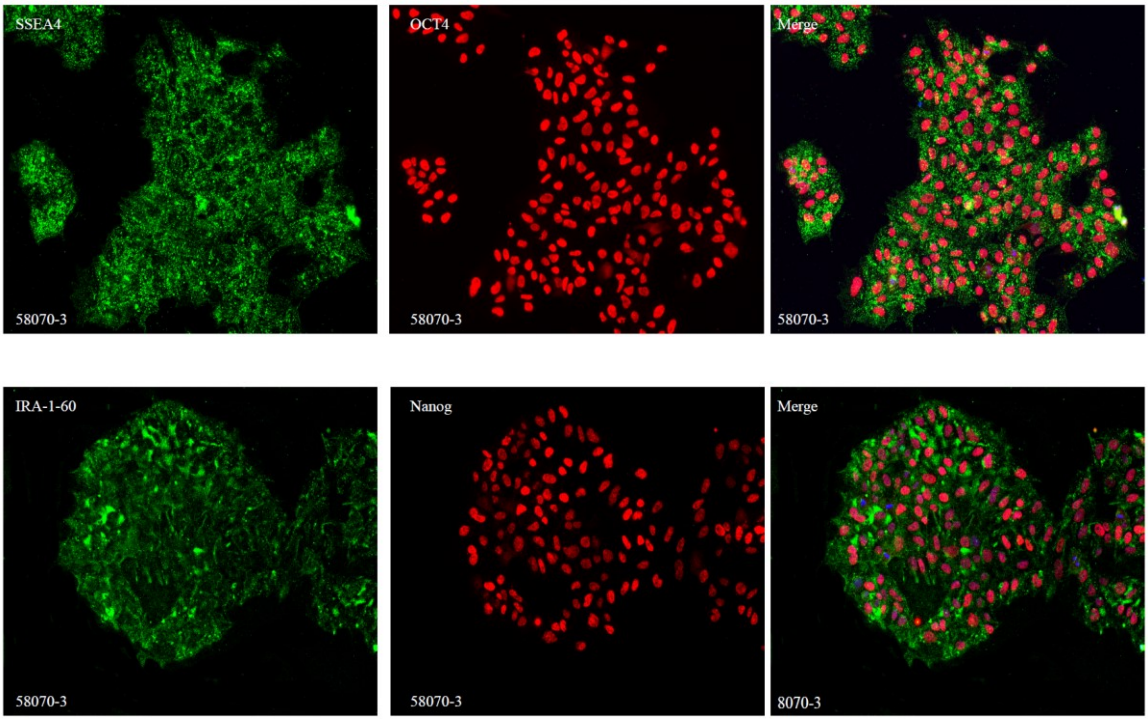
Individual 58203 iPSC 2



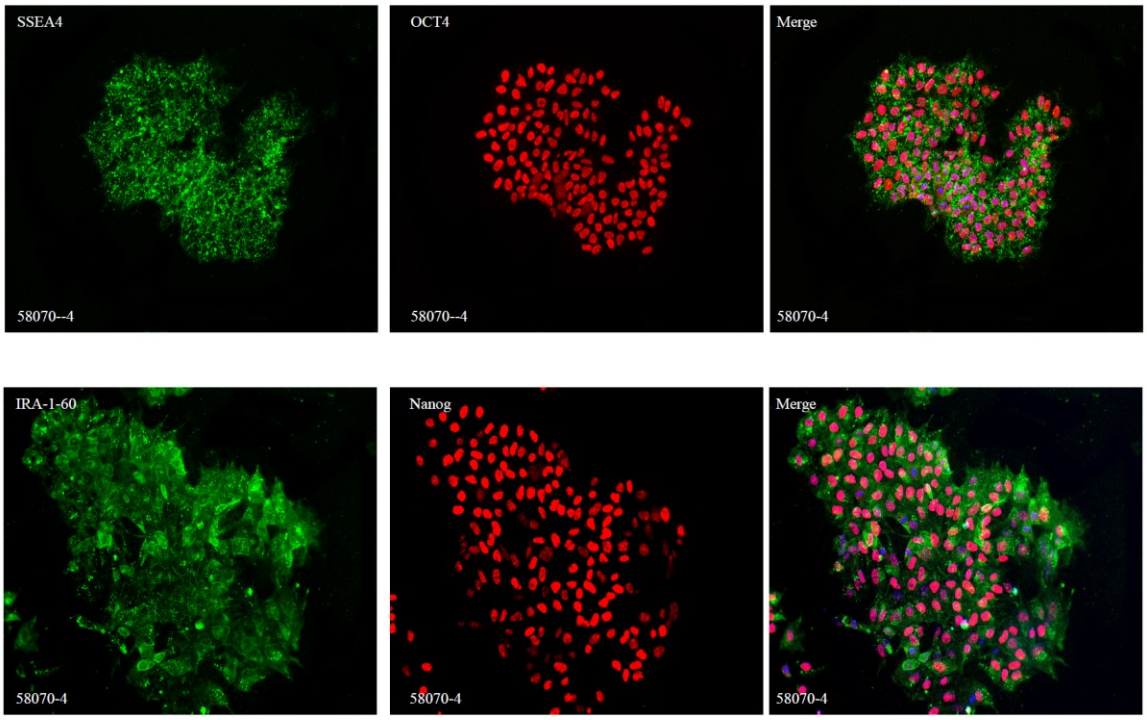
Patient 58071 iPSC 2



Individual 58070 iPSC 3



Individual 58070 iPSC 4



Appendix III: G-banded karyotyping of iPSCs

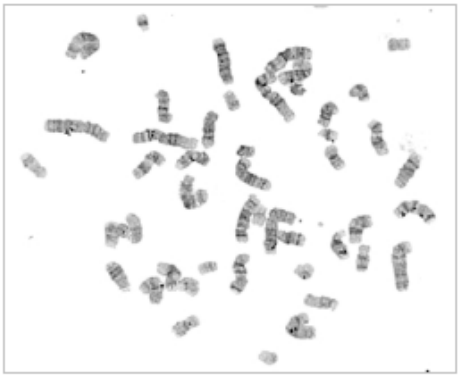
Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_58204-2_11MAY17

PI / ID *Dr. Carl Ernst / Karla Vargas*

Preparation Date: *24-May-17* ***Technician:*** *RW*

Cell Results: *46,XY*



Cell Notes:

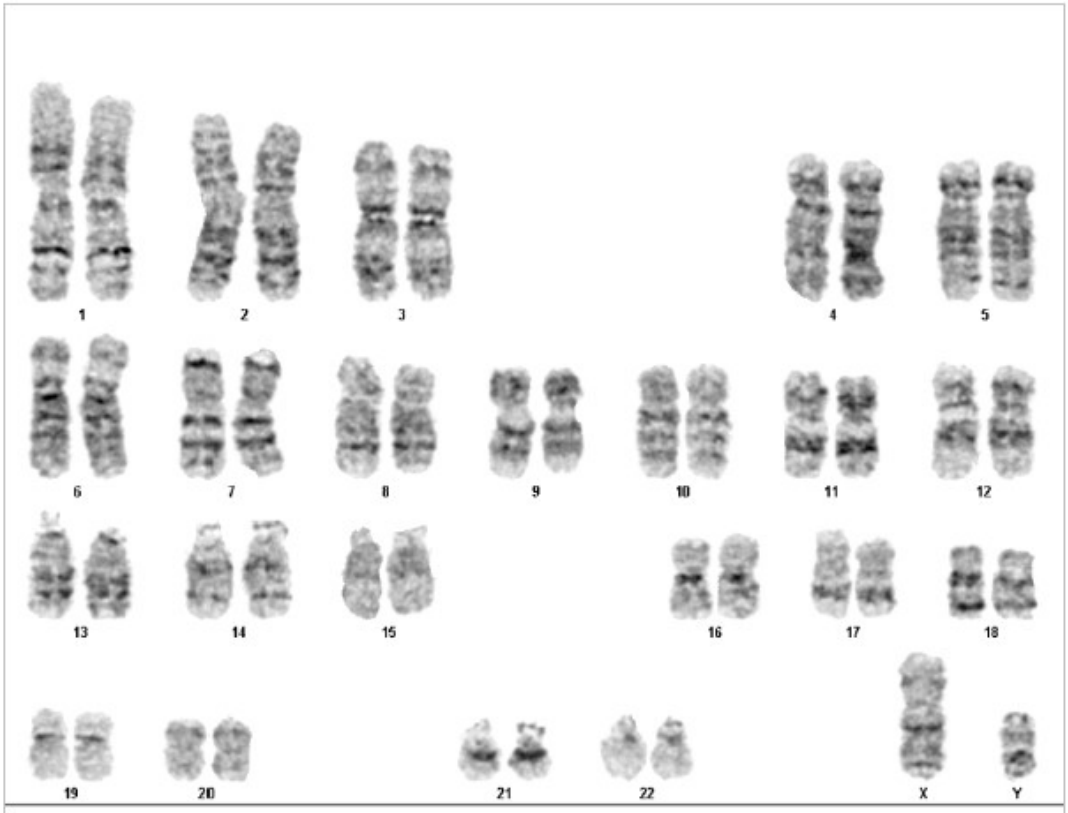
Label - Slide/Cell *Tube 1 - S02-07* ***X, Y:*** , ***Report Date:*** *26-May-17*

Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_58203-1_05APR17

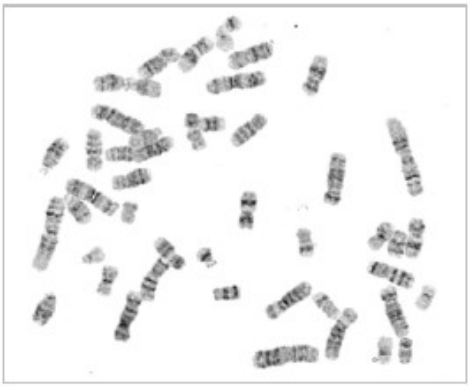
PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 13-Apr-17 **Technician:** RW



Cell Results: 46,XY

Cell Notes:



Label - Slide/Cell S01-09 **X , Y:** , **Report Date:** 20-Apr-17

Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_58203-2_05APR17

PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 12-Apr-17 **Technician:** RW

Cell Results: 46,XY

Cell Notes:

Label - Slide/Cell S01-01 **X , Y:** , **Report Date:** 21-Apr-17

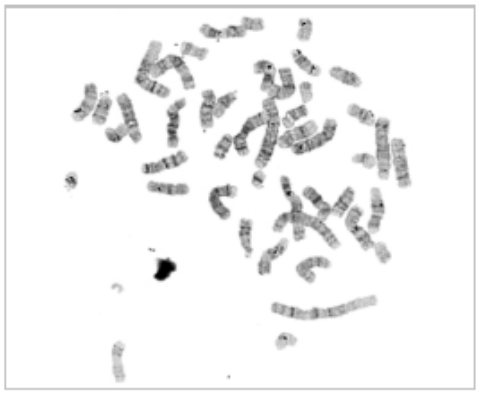
Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_5871-1_27APR17

PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 08-May-17 **Technician:** RW

Cell Results: 46,XX



Cell Notes:

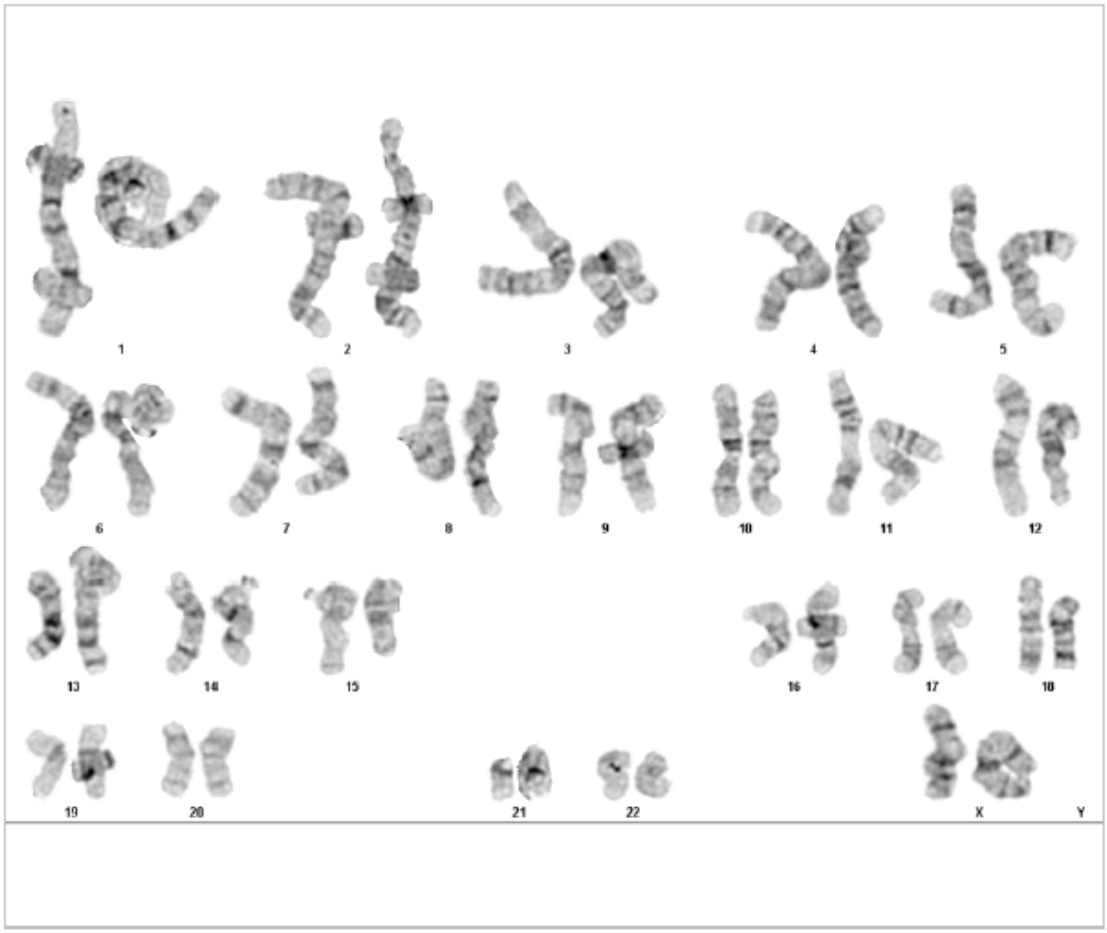
Label - Slide/Cell Tube 1 - S02-01 **X , Y:** , **Report Date:** 15-May-17

Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_5871-2_27APR17

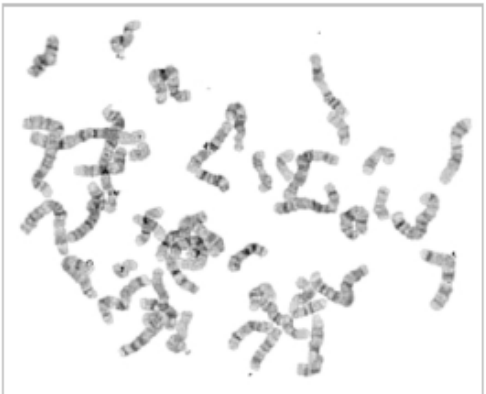
PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 09-May-17 **Technician:** RW



Cell Results: 46,XX

Cell Notes:



Label - Slide/Cell Tube 1 - S01-04 **X , Y:** , **Report Date:** 16-May-17

Cytogenomic Services
TCAG
The Hospital for Sick Children

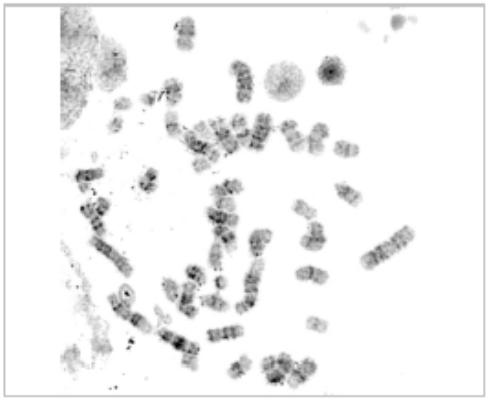
Case: KARV_58070_3 P14_13SEP17

PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 18-Sep-17 **Technician:** RW

Cell Results: 46,XX

Cell Notes:



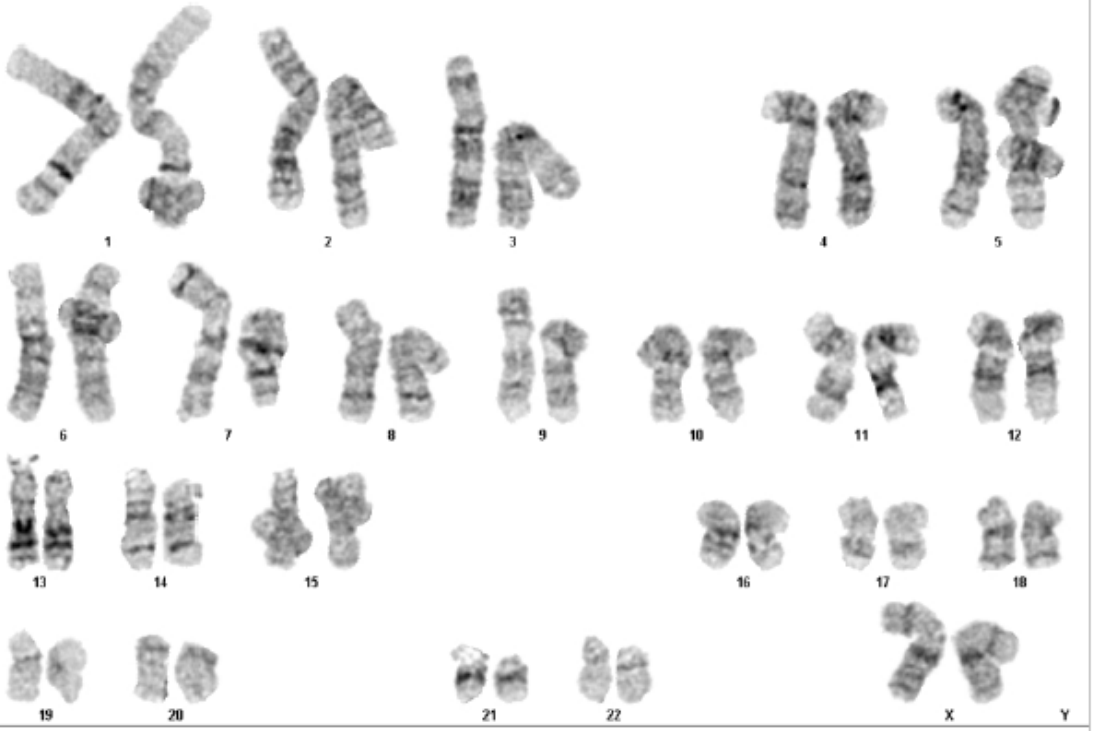
Label - Slide/Cell Tube 1 - S02-03 **X , Y:** , **Report Date:** 19-Sep-17

Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: KARV_5870-4_19APR17

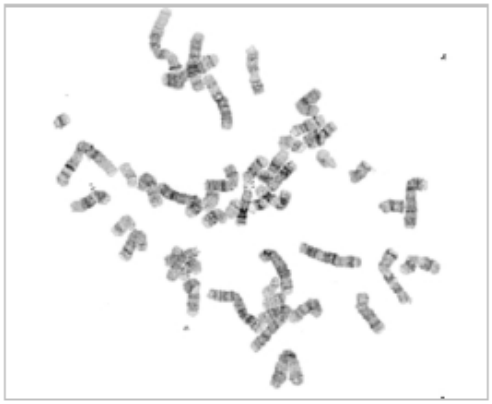
PI / ID Dr. Carl Ernst / Karla Vargas

Preparation Date: 03-May-17 **Technician:** RW



Cell Results: 46,XX

Cell Notes:



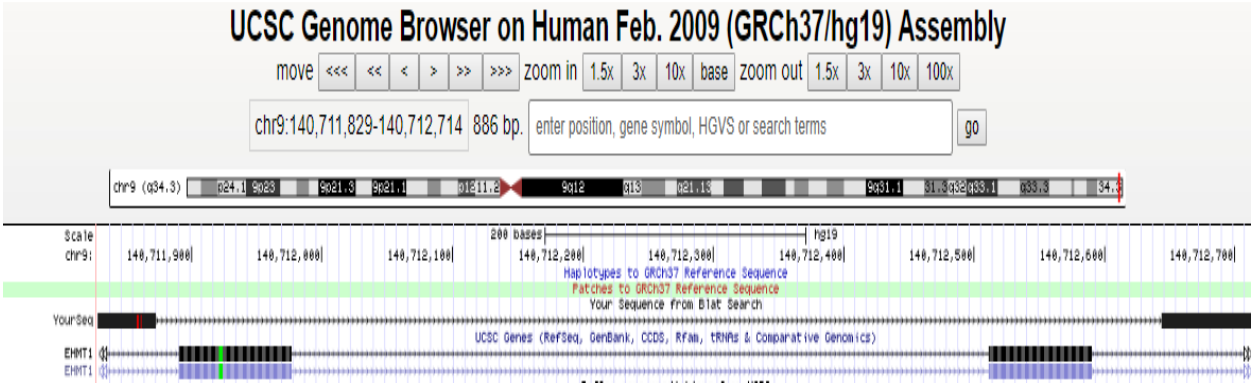
Label - Slide/Cell Tube 1 - S01-07 **X , Y:** , **Report Date:** 05-May-17

Appendix IV: SET-domain KO 2 colony 10 Sanger sequencing results

A)

>ID|10764092 Colony10_P1714184_093.ab1
NNNcNggctttNNtgaccattgtggcaacaggctgaggctgtgNNcNgggttcaccacaatgtgtttgtcccagtagggctgggattcagaagagagctcttactgttgacaagagtgggcttgctatNNN
>ID|10764095 Colony10_P1714184_094.ab1
NNNtcagccNNNtgggacaacacattgtggtgaacccaggccacagcctcagcctgttgccacaatggtcagcaaaagccagctggtgccacgtcactgaaatgaggaagaaaactggttctgtatccgaagcgttcgatgacNNNN

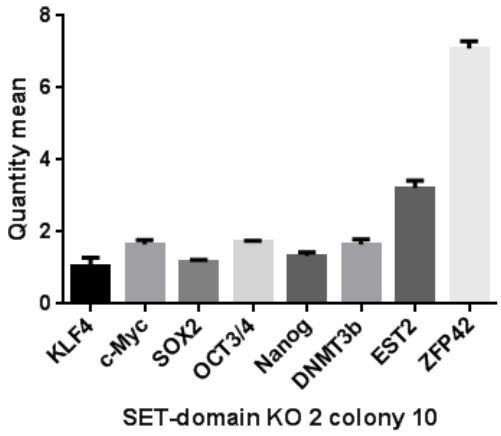
B)



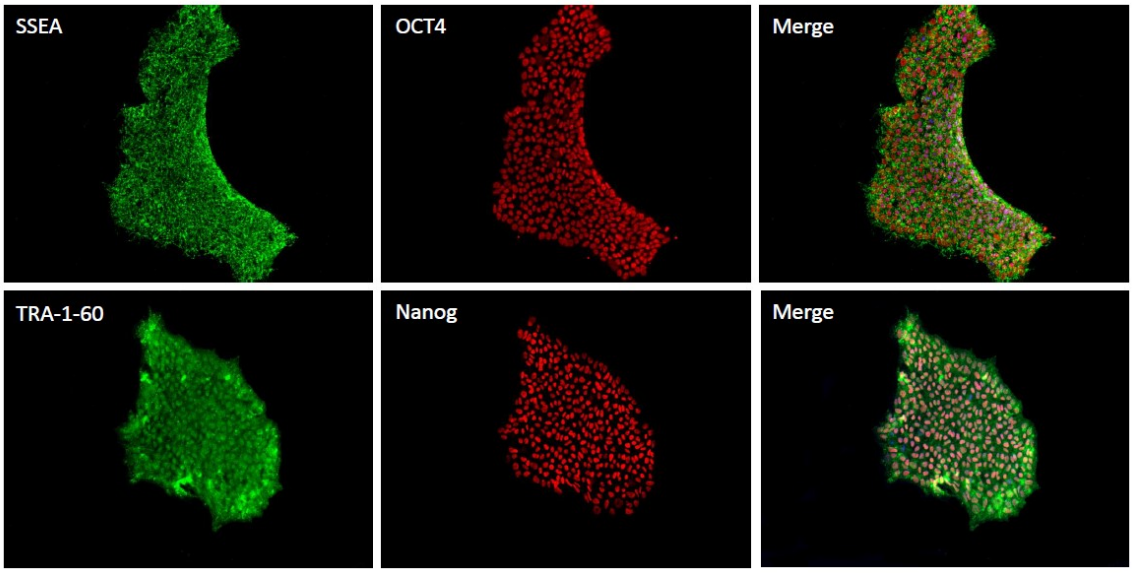
Blat alignment in the UCSC Genome Browser, assembly GRCh37/hg19. SET-domain KO 2 colony 10 sequenced DNA is represented in a black rectangle in the first lane of the browser, below EHMT1 RefSeq sequences. There is a big chunk of DNA missing between the two extremes of our sequence indicating that this part of the EHMT1 gene was successfully deleted.

Appendix V: SET-domain KO 2 colony 10 quality control

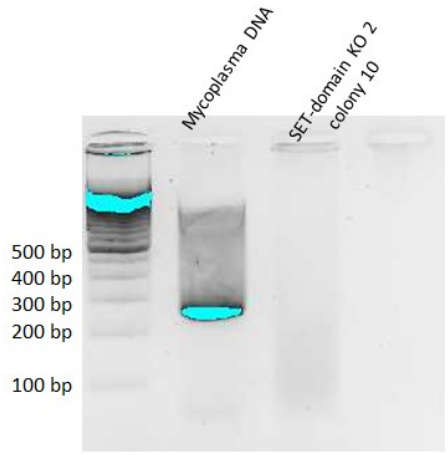
Endogenous expression of pluripotency markers



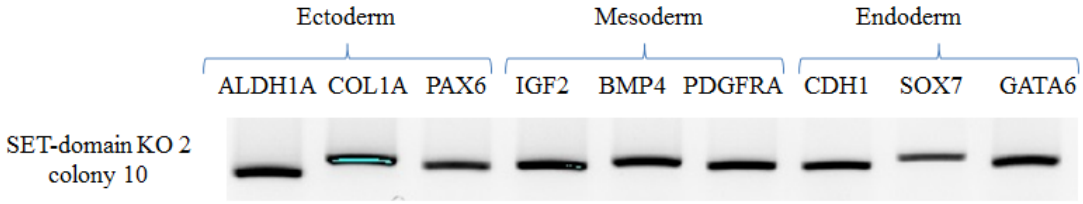
Immunocytochemistry of pluripotency markers



Assessment of mycoplasma contamination



Assessment of differentiation potential using a nine-gene PCR



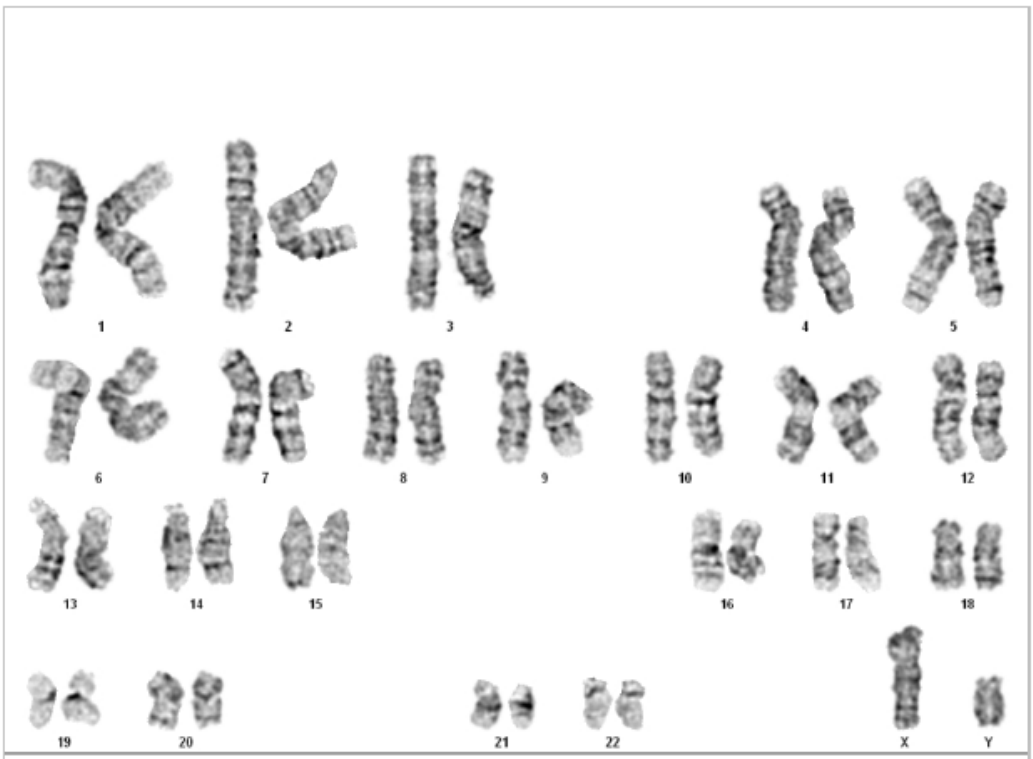
G-banded karyotyping

Cytogenomic Services
TCAG
The Hospital for Sick Children

Case: *KARV_SET-domain KO 2 colony 10_1*

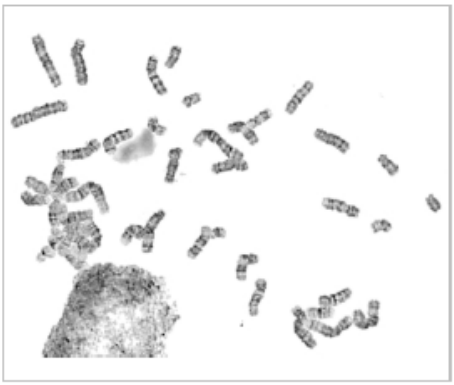
PI / ID *Dr. Carl Ernst / Karla Vargas*

Preparation Date: *25-May-18* *Technician:* *RW*



Cell Results: *46,XY*

Cell Notes:



Label - Slide/Cell *Tube 2 - S07-04* *X, Y:* , *Report Date:* *29-May-18*