Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Gain of 20q11.21 in Human Pluripotent Stem Cells Impairs TGF-β-Dependent Neuroectodermal Commitment

C. Markouli,[1] E. Couvreu De Deckersberg,[1] M. Regin,[1] H.T. Nguyen,[2] F. Zambelli,[1,3] A. Keller,[1] D. Dziedzicka,[1] J. De Kock,[4] L. Tilleman,[5] F. Van Nieuwerburgh,[5] L. Franceschini,[6] K. Sermon,[1] M. Geens,[1] and C. Spits[1,*]

[1]Research Group Reproduction and Genetics, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Laarbeeklaan 103, 1090 Brussels, Belgium
[2]Center for Molecular Biology, Institute of Research and Development, Duy Tan University, K7/25 Quang Trung, Danang 550000, Vietnam
[3]Clínica EUGIN, Travessera de les Corts 322, 08029 Barcelona, Spain
[4]Department of In Vitro Toxicology & Dermato-Cosmetology, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Laarbeeklaan 103, 1090 Brussels, Belgium
[5]Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, 9000 Ghent, Belgium
[6]Laboratory of Molecular & Cellular Therapy, Department of Immunology - Physiology, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Laarbeeklaan 103, 1090 Brussels, Belgium
*Correspondence: claudia.spits@vub.be
https://doi.org/10.1016/j.stemcr.2019.05.005

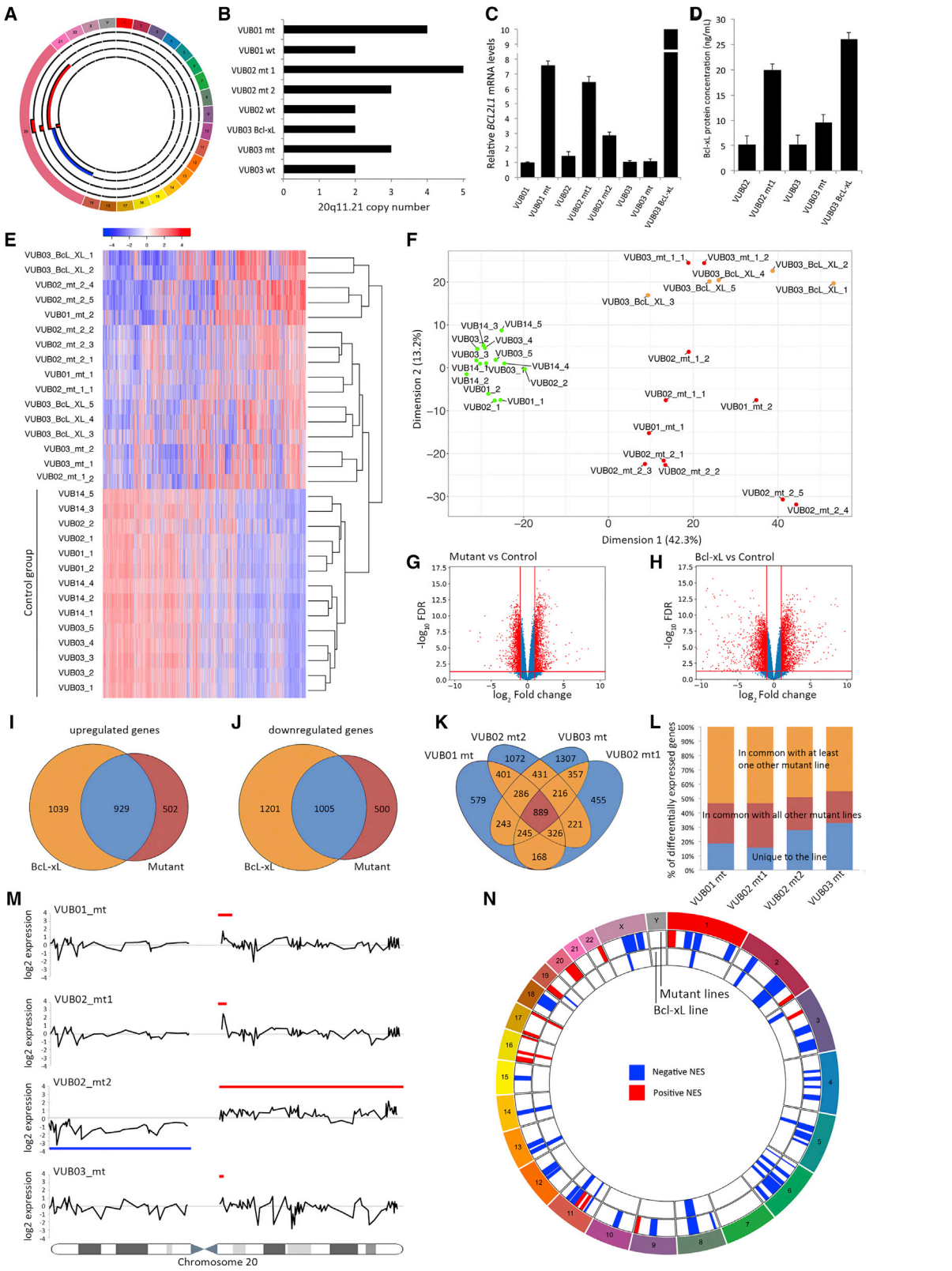
SUMMARY

Gain of 20q11.21 is one of the most common recurrent genomic aberrations in human pluripotent stem cells. Although it is known that overexpression of the antiapoptotic gene *Bcl-xL* confers a survival advantage to the abnormal cells, their differentiation capacity has not been fully investigated. RNA sequencing of mutant and control hESC lines, and a line transgenically overexpressing *Bcl-xL*, shows that overexpression of *Bcl-xL* is sufficient to cause most transcriptional changes induced by the gain of 20q11.21. Moreover, the differentially expressed genes in mutant and *Bcl-xL* overexpressing lines are enriched for genes involved in TGF-β- and SMAD-mediated signaling, and neuron differentiation. Finally, we show that this altered signaling has a dramatic negative effect on neuroectodermal differentiation, while the cells maintain their ability to differentiate to mesendoderm derivatives. These findings stress the importance of thorough genetic testing of the lines before their use in research or the clinic.

INTRODUCTION

Human embryonic stem cells (hESCs) can be cultured *in vitro* for extended periods of time without losing their ability to differentiate into all three germ layers. Their terminally differentiated derivatives have become a potent tool in disease modeling and may play important future roles in regenerative medicine, toxicology tests, and drug screening. As all these applications will require between millions and billions of cells, prolonged culture periods are almost inevitable (Avior et al., 2016; Serra et al., 2012). hESC cultures tend to acquire chromosomal abnormalities that vary in size from small structural variants to full chromosome gains and losses (Keller et al., 2018; Nguyen et al., 2013). Several aberrations are known to be recurrent, such as gains of chromosomes 1, 12, 17, and X, or parts thereof (Amps et al., 2011; Baker et al., 2007; Cowan et al., 2004; Draper et al., 2004; Herszfeld et al., 2006; Inzunza et al., 2004; Maitra et al., 2005; Mitalipova et al., 2005; Yang et al., 2010). Of special interest is the gain of 20q11.21 found in more than 20% of lines worldwide (Amps et al., 2011; Laurent et al., 2011; Lefort et al., 2008; Närvä et al., 2010; Spits et al., 2008; Wu et al., 2008). The smallest region of amplification includes the antiapoptotic gene *BCL2L1* (Amps et al., 2011). We, and others, have found that the mutant cells overexpress *Bcl-xL*, the predominantly expressed isoform of *BCL2L1*. This confers a survival advantage to these mutant cells that allows them to quickly take over the culture (Avery et al., 2013; Nguyen et al., 2014).

Although it is generally assumed that chromosome abnormalities influence the functional characteristics of hESCs, particularly on differentiation, only a handful of reports on this topic have been published (Ben-David et al., 2014; Fazeli et al., 2011; Herszfeld et al., 2006; Keller et al., 2018; Werbowetski-Ogilvie et al., 2009; Yang et al., 2008). Undifferentiated chromosomally abnormal hESCs have been found to acquire characteristics such as increased cloning efficiency, decreased population doubling times (Fazeli et al., 2011), increased capacity for cell proliferation and self-renewal, antiapoptotic properties (Baker et al., 2007; Yang et al., 2008), growth factor independence, and an increase in teratoma-initiating cells (Werbowetski-Ogilvie et al., 2009). These studies also demonstrated that hESCs carrying chromosomal abnormalities show altered gene expression patterns when compared with normal cell lines, while retaining their—possibly reduced—differentiation capacity (Fazeli et al., 2011; Werbowetski-Ogilvie et al., 2009). Karyotypically abnormal hESCs tend to produce immature teratomas when injected *in vivo*, containing a higher proportion of poorly differentiated or undifferentiated cells when compared with normal hESCs (Herszfeld et al., 2006; Werbowetski-Ogilvie et al., 2009; Yang et al., 2008). These



(legend on next page)

studies also show an upregulation of a number of oncogenes, concurrent with a downregulation of genes related to differentiation, which has been interpreted as a first step toward malignant transformation (Gopalakrishna-Pillai and Iverson, 2010; Werbowetski-Ogilvie et al., 2009; Yang et al., 2008). Furthermore, similarly to malignant cancer cells, mutant hESCs migrate faster in three-dimensional collagen gels than normal hESCs (Werbowetski-Ogilvie et al., 2009). Overall, these studies were limited in that they were carried out on a small number of lines carrying different abnormalities, or even one hESC line carrying a 20q11.21 duplication (Werbowetski-Ogilvie et al., 2009), while only one study consistently described one type of aberration using multiple carrier lines (Ben-David et al., 2014).

Our aim was to elucidate if and how the gain of 20q11.21 impacts the differentiation capacity of hESCs. This is of particular importance given that most of these gains fall below the size detection limit of conventional G-banding, which is still the most commonly used method for monitoring the karyotype of human pluripotent stem cells (hPSCs). We studied nine hESC lines, of which four carry a gain of 20q11.21 and one line transgenically overexpresses *Bcl-xL*. Two other genes present in the smallest common region of gain, *ID1* and *HM13*, are not further considered in this work because, as already described, ID1 protein level does not differ in the mutant cells as compared with normal lines and the function of HM13 is limited and relates to immune function, making it a poor candidate in relation to differentiation (Nguyen et al., 2014).

Our results show that hESC lines carrying a 20q11.21 amplification have a transcriptome that is significantly different from their chromosomally normal isogenic counterparts as reflected in multiple deregulated pathways, in particular for transforming growth factor β (TGF-β)- and SMAD-mediated signaling genes. We show that *Bcl-xL* overexpression is the main driver of these changes. Finally, we show that this results in a strong impairment of ectodermal differentiation in the mutant cells, while their capacity for differentiation into mesendodermal derivatives remains intact.

RESULTS

hESCs with a Gain of 20q11.21 Show an Altered Transcriptomic Profile Similar to the *Bcl-xL* Overexpressing Line

As a first step, we investigated the transcriptomic differences between lines with a gain of 20q11.21 and their genetically normal counterparts, and the role of *Bcl-xL* in these differences. Our study included nine hESC lines derived and kept in culture in our laboratory: four lines with gains of 20q11.21 (VUB01_mt, VUB02_mt1, VUB02_mt2, and VUB03_mt) of different sizes, their chromosomally balanced isogenic counterparts (VUB01, VUB02, and VUB03), an additional normal line (VUB14) and a line transgenically overexpressing *Bcl-xL* (VUB03_Bcl-xL) characterized by Nguyen et al. (2014) (Table S1). The karyotypes of all our lines were assessed by array comparative genomic hybridization (aCGH) before the experiments and the number of copies of 20q11.21 was established by real-time qPCR. Figure 1A shows a diagram of the chromosomal content of the mutant lines with a magnified chromosome 20. The size

Figure 1. HESCS with a Gain of 20q11.21 Show an Altered Transcriptomic Profile Similar to VUB03_BcL-xL

(A) Graphical representation of the chromosomal content of the mutant lines in this study, established by aCGH. Chromosome 20 is enlarged to show the breakpoints and region of gain. From outside to inside: VUB01_mt, VUB02_mt1, VUB02_mt2, and VUB03_mt. Regions of gain are represented in red, loss in blue.

(B) Copy number of the 20q11.21 region established by qPCR in all studied lines (n = 3).

(C and D) (C) *BCL2L1* mRNA (n = 3) and (D) Bcl-xL protein levels in mutant and control hESCs and in VUB03_BcL-xL as established by sandwich ELISA. Data are shown as means ± SEM (n = 3).

(E and F) (E) Unsupervised cluster analysis and (F) multidimensional scaling plot of dimension 1 versus dimension 2 of all coding genes with a count per million greater than one in at least two samples.

(G and H) Volcano plots of the differential gene expression analysis with a cutoff value of $|\log_2$ fold change$| > 1$ and FDR < 0.05 for control versus mutant (G) and control versus VUB03_BcL-xL (H).

(I and J) Venn diagrams comparing the deregulated genes in VUB03_BcL-xL with those deregulated in the mutant lines. (I) Shows the upregulated genes and (J) the downregulated genes.

(K) Venn diagram comparing the deregulated genes in all mutant lines.

(L) Shows the percentage of deregulated genes that are unique or common, or common between at least two lines.

(M) Smoothened gene expression data of all genes on chromosome 20, for the four lines carrying a gain of 20q11.21. The blue lines show the region of chromosomal gain, the red line the region of loss.

(N) Gene set enrichment analysis using the MSigDB database C1, for the deregulated genes in the mutant lines and in VUB03_BcL-xL. The plot shows the chromosomal positions with significant enrichment in deregulated genes. In blue are regions with a negative normalized enrichment score (NES) and in red, with a positive NES.

of the amplification ranged between 0.9 and 4 Mb with an identical proximal and a variable distal breakpoint. VUB02_mt2 carried an isochromosome 20 (loss of the p arm and duplication of the q arm). All the control lines had a fully balanced genetic content, and VUB03_BcL-xL also carried a gain of 1q32.1q41. The exact breakpoints of the chromosomal abnormalities, along with the passage range of each hESC line used in the study can be found in Table S1. The gain of 20q11.21 is often an amplification ranging from three to five copies rather than a simple duplication (Figure 1B). The *Bcl-xL* mRNA levels in the overexpressing line were nearly 70 times higher than in the control lines and were 10-fold that of VUB01_mt and VUB02_mt1 (Figure 1C), although the protein levels were similar to those of VUB02_mt1 (Figure 1D).

We carried out RNA sequencing on two to five replicates per line collected from different dishes. We only considered coding genes with a count per million greater than one in at least two samples. Unsupervised hierarchical cluster analysis using all expressed genes shows that the mutant hESC lines and the VUB03_Bcl-xL line display an expression motif different from the control lines, resulting in the clustering of the mutant and the *Bcl-xL* lines separate from the control lines (Figure 1E). The same pattern appears in the multidimensional scaling plot generated considering all expressed genes, showing the clustering of the control lines separate from the mutant lines and together with VUB03_BcL-xL (Figure 1F), demonstrating that overexpression of *Bcl-xL* is sufficient to result in a transcriptome close to that of a mutant line. Furthermore, the clustering pattern of VUB03_Bcl-xL suggests that the impact of the overexpression of *Bcl-xL* on the transcriptome of the cells is stronger than that of the gain of 1q32.1q41, which the line also carries.

Differential gene expression analysis with a cutoff value of $|\log_2$ fold change$| > 1$ and false discovery rate q value (FDR) < 0.05 shows 1,431 upregulated and 1,505 downregulated genes in mutant versus control lines. VUB03_BcL-xL shows 1,968 upregulated and 2,206 downregulated genes compared with the normal lines (Figures 1G and 1H). Approximately 65% of the genes that are differentially expressed by the mutant lines are in common with the deregulated genes in VUB03_BcL-xL (Figures 1I and 1J). Next, we plotted the overlap in deregulated genes of the mutant lines compared with the control lines. All mutant lines have a subset of deregulated genes that are unique to the line, representing between 15% and 32% of the differential gene expression. Conversely, 22% to 30% of genes were common to all lines, and between 44% and 53% were in common with at least one other mutant line (Figures 1K and 1L). This suggests that a core of 889 genes are deregulated due to the overexpression of genes located in the common region of 20q11.21 irrespective of the line-to-line variation. Furthermore, 763 of these 889 genes (85.8%) are also deregulated in the *Bcl-xL* overexpressing line, strongly suggesting that Bcl-xL plays a predominant role in the transcriptomic profile of the cells with a gain of 20q11.21.

Finally, we mapped the differential gene expression to chromosomal position. First, we calculated and plotted the smoothened $\log_2$ fold-change expression of all genes on chromosome 20 for each mutant line individually, and compared this with the control lines as a group (Figure 1M, raw data in Figure S1). In VUB01_mt, VUB02_mt1, and VUB02_mt2 the overexpression of genes located in the region of gain of chromosome 20 is clearly observed. VUB03_mt carries the smallest region of gain and no differences can be observed in the expression in that region. Second, we used the gene set enrichment analysis using the MSigDB database C1 for the deregulated genes in the mutant lines and in VUB03_BcL-xL. The mutant lines have a positive enrichment score for 20q, which is absent in VUB03_BcL-xL. However, there is a similar enrichment score across the genome of the mutant lines and VUB03_BcL-xL, showing 16 of the 18 enriched regions to be in common (Figure 1N). These results illustrate a number of salient points. Expectedly, the higher copy number has an effect on the gene expression of most—but not all—genes in the amplified region, which is why gene expression data can be used to karyotype cells (Ben-David et al., 2013). However, smaller gains fall under the detection level unless they induce very strong gene expression changes. More importantly, a variant copy number of one region can strongly act in *trans* on the expression of genes located on other chromosomes.

In summary, mutant lines show a different transcriptome from their normal counterparts, which can largely be attributed to the overexpression of *Bcl-xL*. The overexpression of other genes located within the mutation has a minimal, if any, effect.

The 20q11.21 Mutation Affects TGF-β- and SMAD-Mediated Signaling through *Bcl-xL* Overexpression

In a next step, we carried out pathway analysis using multiple bioinformatic tools. First, we carried out ingenuity pathway analysis (IPA) using all genes with a $|\log_2$ fold change$| > 1$ and FDR < 0.05, for both the mutant versus control groups, and the VUB03_BcL-xL line versus a control group. This analysis predicts the state of upstream regulators based on the expression of their downstream targets. IPA considers a pathway to be activated with a Z score >2, and inhibited with a score less than −2. When the p value is significant and the activation Z score is between −2 and 2, it is considered to be significantly affected, without being able to establish whether it is activated or inhibited. In our dataset, the activation score in

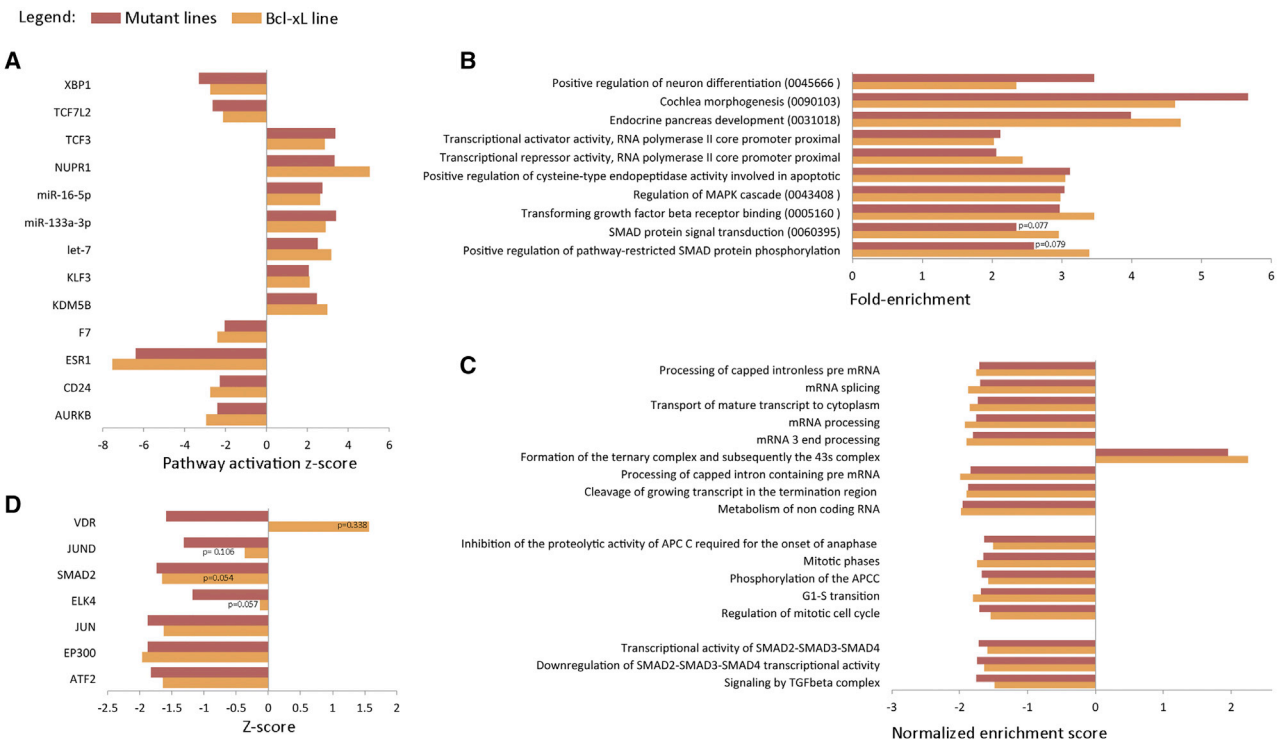


Figure 2. Pathway Analysis in hESCs with a Gain of 20q11.21 and in VUB03_BcL-xL

(A) Ingenuity Pathway Analysis of the upstream regulators of pathways with an activation score between −2 and 2 and a p value < 0.05 in the wild-type (WT) versus mutant (MT) lines and their behavior in VUB03_BcL-xL.

(B) DAVID Enrichment Analysis of the 1,000 top deregulated genes in the mutant group and VUB03_BcL-xL. In parenthesis are the gene-ontology term numbers.

(C and D) Gene set enrichment analysis using the MSigDB C2 library (C) and Enrichr-based prediction of protein-protein interactions for transcription factors (D). All results are statistically significant with a p value < 0.05 unless the value is given. All the analyses were performed using the differentially expressed genes between WT versus mutant group and WT versus VUB03_BcL-xL with a cutoff value of $|\log_2$ fold change$| > 1$ and FDR < 0.05.

VUB03_BcL-xL follows that of the mutant lines in all cases, and, in the case of pathways not reaching the *Z* score >2 or less than −2, the p value is still always significant (Figure 2A).

We then subjected the 1,000 genes with the highest log fold change in absolute value (563 up- and 437 downregulated for the 20q11.21 lines, 485 up- and 515 downregulated for VUB03_BcL-xL) to DAVID functional annotation enrichment analysis. The analysis retrieved a total of 75 annotations, of which we selected those gene ontology terms referring to cellular processes and that had a fold enrichment >2, and a p value < 0.05 in at least one of the groups (Figure 2B). Interestingly, the deregulated genes in both mutant lines and VUB03_BcL-xL are enriched for factors in the TGF-β and SMAD signaling, as well as in genes involved in neuron differentiation and cochlea morphogenesis, which are full or partial ectodermal derivatives, respectively. Next, we carried out gene set enrichment analysis using the C2 library (full list in Table S2). The results did not yield any significantly deregulated pathways when using the Kyoto Encyclopedia of Genes and Genomes, but revealed 17 significantly enriched processes as annotated in the Reactome library. These can broadly be categorized into related to mRNA processing, cell cycle, and TGF-β and SMAD signaling (Figure 2C).

Subsequently, we analyzed the transcription factor enrichment and protein-protein interactions in the differentially expressed genes in the mutant lines and in VUB03_BcL-xL with a $|\log_2$ fold change$| > 1$ and FDR < 0.05, using Enrichr (Chen et al., 2013; Kuleshov et al., 2016). Our datasets show a negative *Z* score for protein interaction with SMAD2 (Figures 2D and S2).

Overall, these results suggest that hESCs with a gain of 20q11.21 appear to have a perturbed TGF-β and SMAD signaling and that *Bcl-xL* overexpression is responsible for most of these changes. Given the pivotal role of TGF-β and SMAD signaling in gastrulation and early lineage commitment, we hypothesized that the gain of 20q11.21

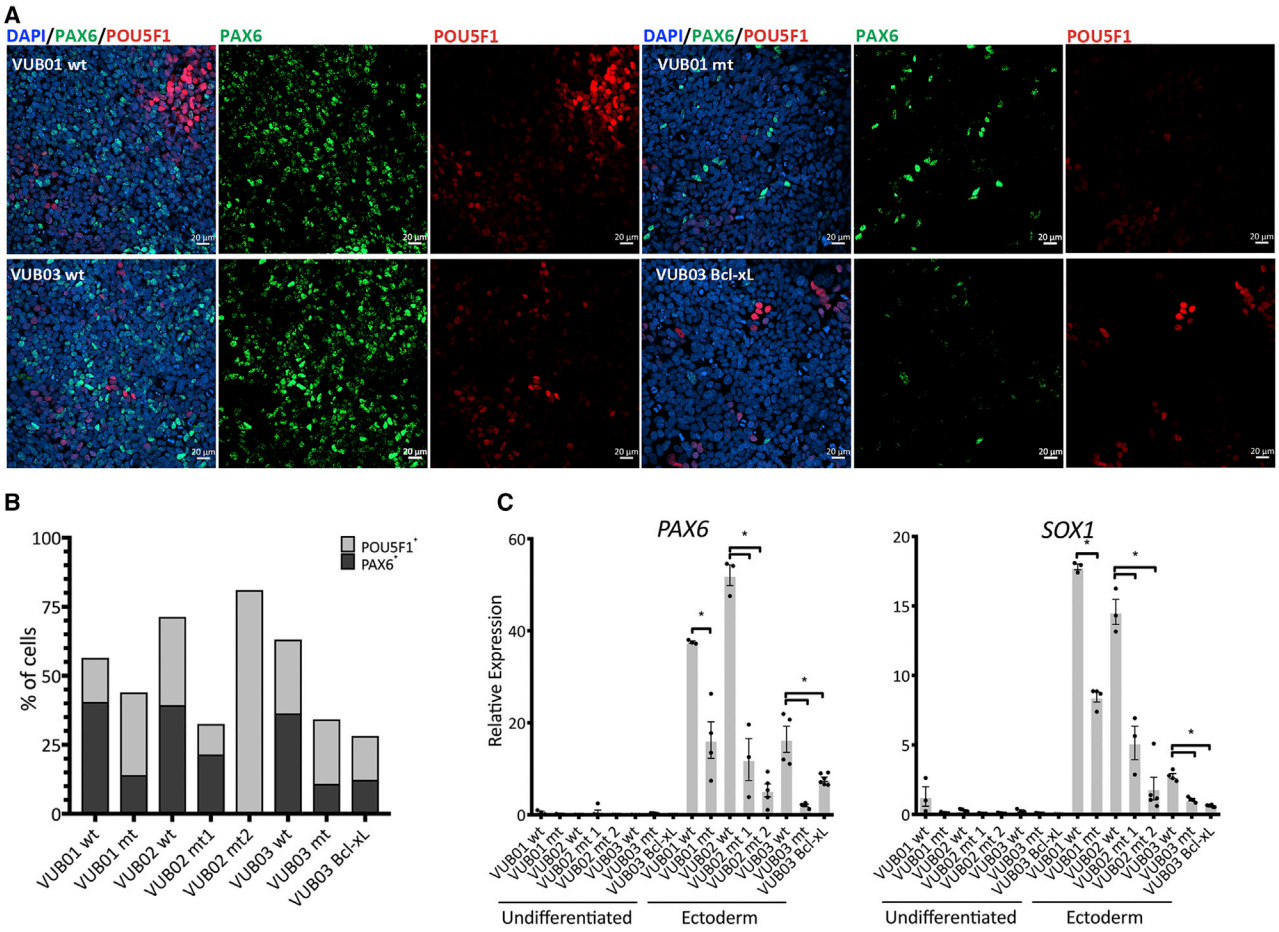


Figure 3. *Bcl-xL* Overexpression Due to a Gain of 20q11.21 Results in Impaired Neuroectoderm Differentiation

(A) Examples of immunostaining for PAX6 (green) and POU5F1 (red) in mutant and control lines and in VUB03_BcL-xL, after 4 days of neuroectoderm differentiation.

(B) Percentage of PAX6- and POU5F1-positive cells. The mutant lines and VUB03_BcL-xL yield less PAX6-positive cells than the control lines.

(C) Relative mRNA expression as measured by qPCR for ectoderm markers *PAX6* and *SOX1* (n = 3–6). Data are shown as means ± SEM, each dot represents an independent differentiation experiment and the horizontal bars with asterisks represent statistical significance between samples ($p < 0.05$, t test).

may have an impact on the differentiation capacity of the cells.

Gain of 20q11.21 Results in Impaired Neuroectoderm Differentiation

We first assessed the neuroectodermal differentiation efficiency of all our control and mutant lines and that of VUB03_BcL-xL. We performed a 4-day induction protocol recapitulating neuroectoderm commitment using Noggin and SB431542, both of which are TGF-β superfamily antagonists (Chambers et al., 2009; Chetty et al., 2013). We differentiated all our lines in three to six independent differentiation experiments to ensure the reproducibility and consistency of the experiments. At the fourth day of neuroectoderm induction, we analyzed the mRNA of markers for each lineage (endo-, neuroecto-, and mesoderm), and the stem cell markers *POU5F1* and *NANOG*. We stained for PAX6 as neuroectodermal marker and POU5F1 as pluripotency marker and quantified the number of positive cells (Figure 3A). The data show that all samples derived from mutant lines and VUB03_BcL-xL display statistically significantly lower levels of *PAX6* and *SOX1* mRNA (Figure 3C), and between 1.8 and 3.3 times less PAX6-positive cells (Figures 3B and S3). Next to the neuroectoderm and stem cell markers, we also tested for the expression of mesendoderm markers (Figure S3). The results show that, while the mutant cells do exit the pluripotent state, they do not differentiate into any early, specific germ line detectable with the

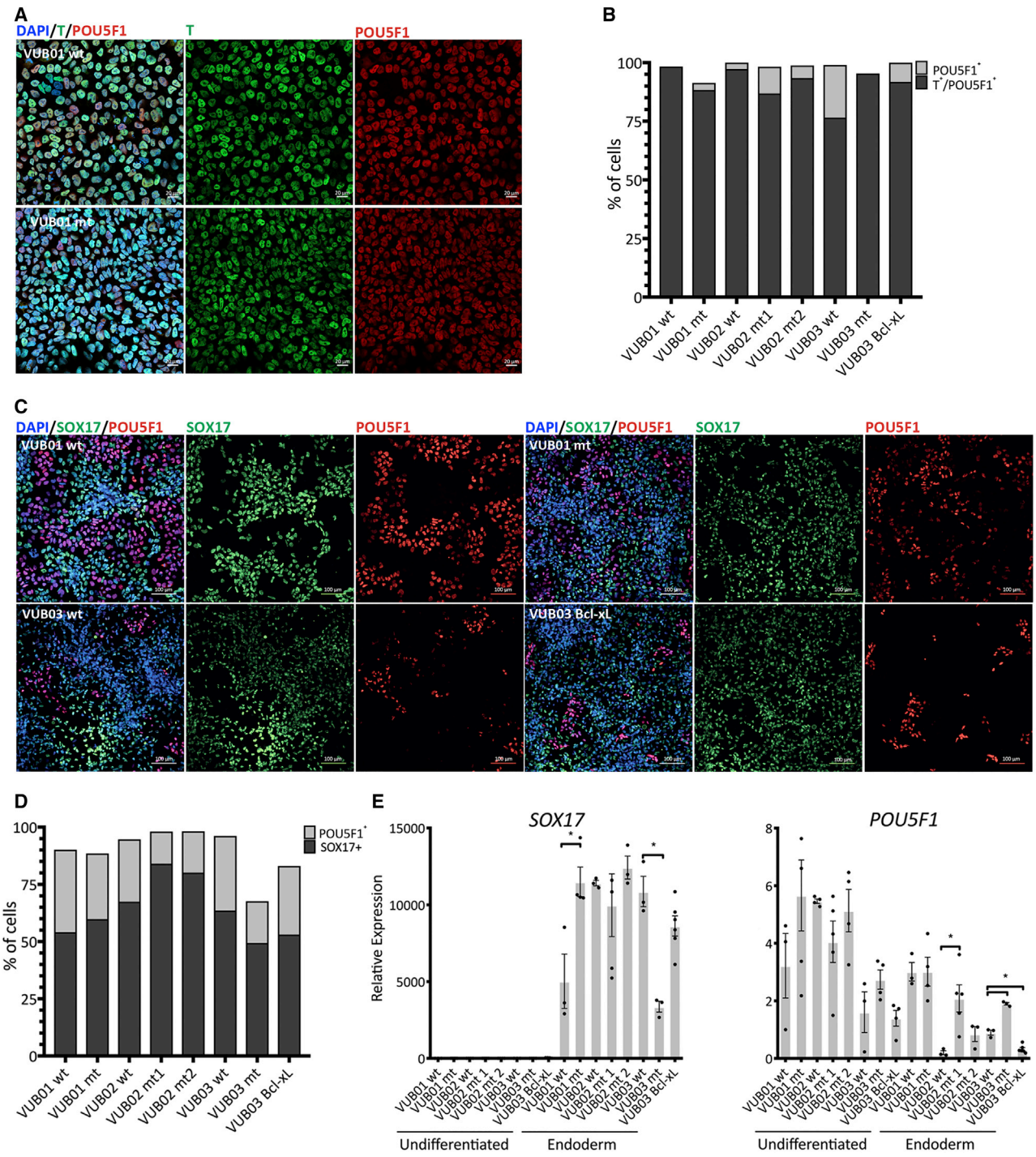


Figure 4. Control and Mutant Lines and VUB03_BcL-xL Equally Differentiate to Mesendoderm Derivatives

(A) Representative examples of immunostaining for T (green) and POU5F1 (red) in control (VUB01_wt) and mutant (VUB01_mt) lines after mesendoderm induction.

(B) Percentage of T/POU5F1 double-positive (mesendoderm cells) and POU5F1-only positive cells.

(C) Examples of immunostaining for SOX17 (green) and POU5F1 (red) in control (VUB01_wt and VUB03_wt) and mutant (VUB01_wt and VUB03_mt) lines, and VUB03_BcL-xL after definitive endoderm differentiation.

(legend continued on next page)

markers we used. The fact that VUB03_BcL-xL follows the same pattern as the mutant cells, showing practically no expression of neuroectodermal markers after the 4-day induction, supports the hypothesis that the overexpression of *Bcl-xL* alone is sufficient to cause this abnormal differentiation. Finally, we also differentiated control and mutant hESCs to neuroectoderm in a 12-day protocol, confirming the difference between the control and mutant hESCs after extended differentiation (Figure S3).

Mutant and Control Lines Equally Differentiate toward Mesendoderm Derivates

Finally, we assessed the impact of the mutation and *Bcl-xL* overexpression on mesendodermal lineage commitment using a 24-h protocol that drives the differentiation through the first events of the primitive streak formation by TGF-β and WNT activation. We measured the protein levels of Brachyury (T) in our lines and found that they all underwent mesendoderm induction with an equally good efficiency (Figures 4A and 4B). To investigate the ability of the lines to further commit to derivatives of this germ layer, we directed the cells toward definitive endoderm. We differentiated all eight lines (with three to six replicates per line) and evaluated the gene expression at the mRNA and protein levels. Figure 4E shows the mRNA expression results, which are in line with the immunostaining results for SOX17 and POUF51 (Figures 4C and 4D). All the lines increased their expression of endoderm markers, and no consistent expression of markers for other lineages was observed (Figure S5). *NANOG* and *POU5F1* were overall slightly but not completely downregulated, as expected given their important roles in early endoderm differentiation (Teo et al., 2011; Ying et al., 2015). Overall, in all cases, control and mutant lines as well as VUB03_BcL-xL readily and similarly differentiated to definitive endoderm.

Downregulation of *CHCHD2* Is a Transcriptomic Marker for the Gain of 20q11.21

The deregulation of the TGF-β and the SMAD signaling suggested by the DAVID analysis and the well-established relationship of this pathway to pluripotency and ectoderm differentiation prompted us to further investigate the differentially expressed genes to understand the driving mechanism for this deregulation. For this, we annotated the function of the top 100 differentially expressed genes (Table S3) in both mutant and *Bcl-xL* overexpressing cells, and searched the literature for their potential relationship to *Bcl-xL* and the TGF-β/SMAD signaling. The most interesting candidate in this list is *CHCHD2* because the protein binds Bcl-xL, resulting in a decreased oligomerization of apoptotic activator Bax (Liu et al., 2015). Furthermore, low levels of *CHCHD2* have been mechanistically linked to disrupted TGF-β/SMAD signaling and decreased neuroectodermal commitment in hPSCs (Zhu et al., 2016). Our RNA sequencing data showed that *CHCHD2* is strongly downregulated in the lines with a gain of 20q11.21 and in VUB03_BcL-xL. Hence, we decided to investigate if *CHCHD2* was effectively the mediator of the decreased neuroectoderm commitment of the mutant cells.

We validated the *CHCHD2* RNA sequencing findings by measuring the mRNA and protein levels in normal and mutant lines as well as VUB03_BcL-xL. All mutant lines and VUB03_BcL-xL showed statistically significantly lower levels of *CHCHD2* at the mRNA level, while the endogenous expression levels varied up to a 4-fold change among the control lines (Figure 5A). Immunocytochemistry of normal and mutant lines and VUB03_BcL-xL showed absence of CHCHD2 protein in the mutant and VUB03_BcL-xL cells (Figure 5B). To assess if this downregulation affected solely the pluripotent state, we differentiated the VUB03_wt, VUB03_mt, and VUB03_BcL-xL to neuroectoderm and definitive endoderm. Figure 5C shows the real-time qPCR results showing that *CHCHD2* remains stably downregulated after differentiation in both VUB03_mt and VUB03_BcL-xL.

To functionally address the potential link between *CHCHD2* downregulation and decreased neuroectoderm commitment, we used small interfering RNA (siRNA) to knockdown *CHCHD2* in control hESCs, and mRNA transfection to exogenously overexpress *CHCHD2* in mutant cells. Figure 6A shows the mRNA levels of *CHCHD2* in undifferentiated VUB01_wt cells 24, 48, and 72 h after transfection, and Figure 6B shows the immunostaining for the protein 72 h posttransfection. The results show downregulation both at the mRNA and protein levels. We induced neuroectoderm commitment in VUB01_wt treated with siRNA against *CHCHD2* as well as untreated cells and cells treated with a nontargeting siRNA. We measured *CHCHD2*, *PAX6*, and *SOX1* expression each day of a 4-day differentiation. This experiment was repeated four times; the results are shown in Figures 6C and 6D. We found that the knockdown of *CHCHD2* did not change the ability of the cells to differentiate to neuroectoderm in any of the replicates. We also exogenously overexpressed *CHCHD2* in VUB01_mt (Figure 6E) and measured the *PAX6* and *SOX1* expression on day 0, 2, and 4 of neuroectoderm induction (Figure 6F).

(D) Percentage of SOX17- and POU5F1-positive cells.

(E) Relative mRNA expression as measured by qPCR (n = 3 to 6). Data are shown as means ± SEM, each dot represents an independent differentiation experiment and the horizontal bars with asterisks represent statistical significance between samples ($p < 0.05$, t test).

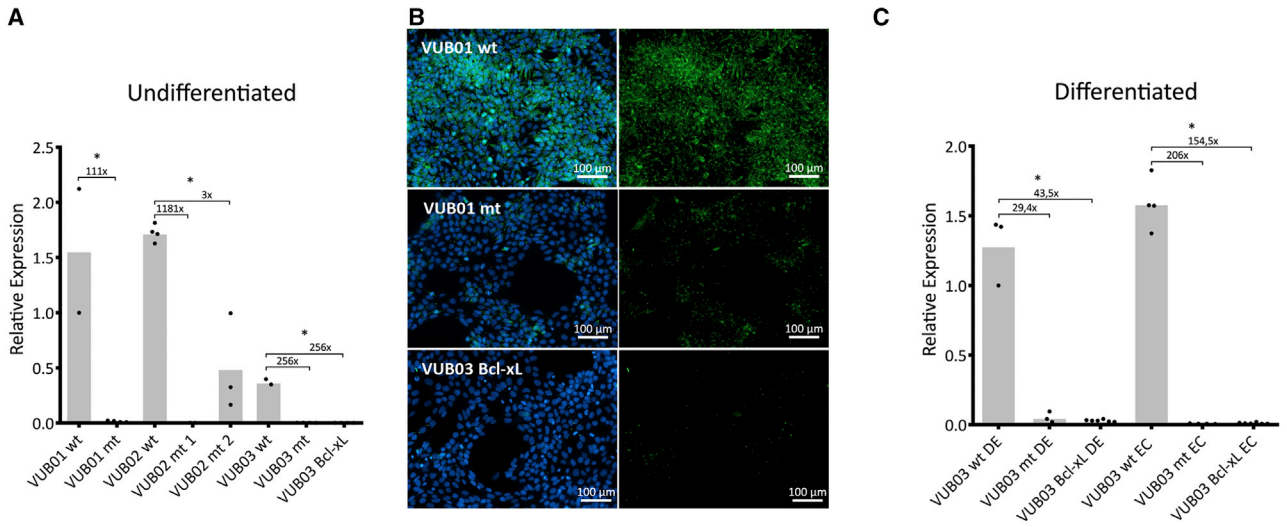


Figure 5. Lines with a Gain of 20q11.21 Stably Downregulate *CHCHD2* Due to *Bcl-xL* Overexpression

(A and B) (A) *CHCHD2* expression as measured by qPCR in undifferentiated cells. All mutant lines show a significantly decreased expression compared with their normal counterparts (n = 4–9). (B) CHCHD2 immunostaining (in green) in VUB01 showing the presence of the protein in the control subline and its absence in the mutant subline.

(C) *CHCHD2* expression as measured by qPCR in the three sublines of VUB03 (control, mutant, and VUB03_BcL-xL) in the differentiated state. DE, definite endoderm; EC, neuroectoderm (n = 7–10).

(A and C) Data are shown as means ± SEM, each dot represents an independent differentiation experiment and the horizontal bars with asterisks represent statistical significance between samples ($p < 0.05$, t test).

In line with the siRNA data, the results show that exogenous *CHCHD2* expression does not rescue the effect of the gain of 20q11.21.

In summary, while *CHCHD2* is consistently downregulated in mutant hESCs both before and after differentiation, our results strongly suggest that it is not the link between the *Bcl-xL* upregulation in the mutant cells and their decreased neuroectodermal commitment.

DISCUSSION

The aim of our study was to investigate the impact of one of the most commonly found chromosomal abnormalities in hPSCs, a gain of 20q11.21, on their differentiation capacity (Amps et al., 2011). The awareness of genome instability in hPSCs has increased in the past years, along with the realization that we lack answers on what the functional consequences of these mutations are on differentiating cells, especially when considering clinical applications (Andrews et al., 2017). Our work was prompted by the clear need for systematic studies of the impact of recurrent chromosomal abnormalities on the differentiation capacity and malignant potential of hPSCs.

To reach our aim, we included four mutant lines and their isogenic normal counterparts, and one *Bcl-xL* overexpressing line. Remarkably, transcriptomic analysis showed that all mutant lines have a significantly different transcriptome from their chromosomally normal isogenic counterparts, and which is very similar to that of cells transgenically overexpressing *Bcl-xL*. This is despite the fact that the mutant lines carried different copy numbers, leading to a difference in overexpression of genes within the region. This strongly suggests that overexpression of *Bcl-xL* alone, which is also the driver for the selective advantage of the cells in their undifferentiated state (Avery et al., 2013; Nguyen et al., 2014), is sufficient to explain the majority of transcriptomic changes. Interestingly, an important part of the differential expression was of genes related to the TGF-β- and SMAD-mediated signaling, which led us to investigate the behavior of the mutant hESCs during the lineage commitment to neuroectoderm and mesendoderm. Our differentiation experiments show that the lines with a gain of 20q11.21 and those overexpressing *Bcl-xL* have an impaired neuroectodermal lineage commitment, with no differences in their capacity for differentiation into mesendodermal derivatives.

Next, *CHCHD2* was identified as a prime candidate linking *Bcl-xL* to the transcriptomic changes, because it is consistently downregulated in mutant cells and the protein binds Bcl-xL, resulting in a decreased oligomerization of Bax (Liu et al., 2015). This downregulation may be through a self-regulatory feedback loop, as CHCHD2 is able to translocate to the nucleus to transactivate itself as well as other genes that contain oxygen-responsive

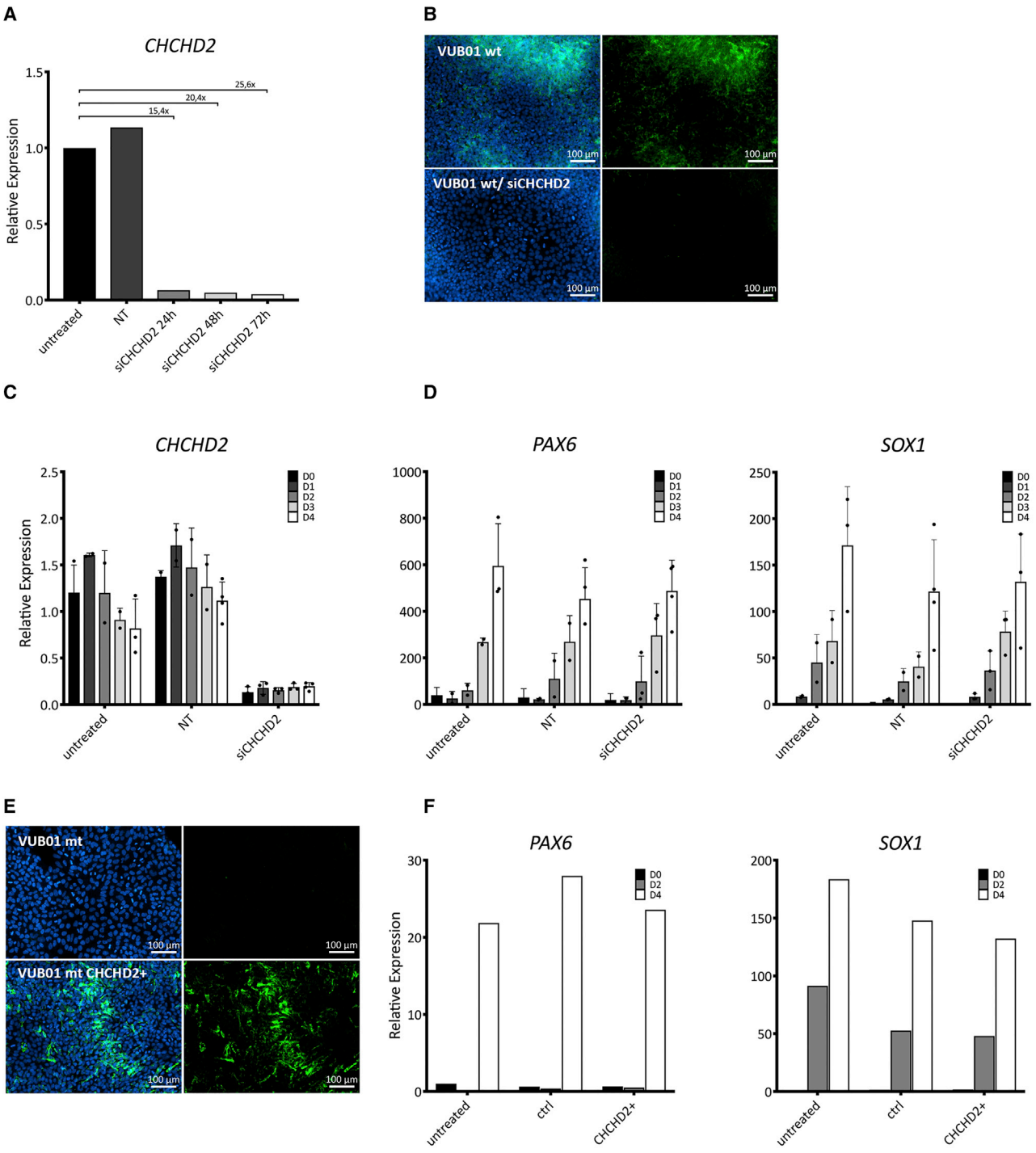


Figure 6. Knockdown and Overexpression of *CHCHD2* Do Not Affect Neuroectoderm Differentiation

(A) *CHCHD2* expression as measured by qPCR in VUB01_wt 24, 48, and 72 h after transfection with siRNA against *CHCHD2* (n = 1).

(B) Representative immunostaining of untreated and transfected cells 72 h posttransfection (the CHCHD2 protein can be seen in green).

(C and D) *CHCHD2* (C) and *PAX6* and *SOX1* (D) expression as measured by qPCR in VUB01 WT during a 4-day induction to neuroectoderm. NT, nontargeting siRNA (n = 1).

(E) Immunostaining of untreated cells and after 72 h of exogenous *CHCHD2* expression (in green).

(F) *PAX6* and *SOX1* expression as measured by qPCR in VUB01_mt during a 4-day induction to neuroectoderm in untreated, control (GFP) and *CHCHD2*-overexpressing cells (n = 1).

element sequences in their promoter region (Aras et al., 2013; Grossman et al., 2017). Another interesting function of CHCHD2 is its capacity to interact with the SMADs (2/3/4) and thus modulate the transcription of TGF-β-mediated target genes (Zhu et al., 2016). Zhu et al. also showed that endogenous low levels of CHCHD2 in hPSCs result in a decreased capacity for ectodermal differentiation. The findings in our transcriptome analysis showing deregulation of both TGF-β/SMAD signaling and genes involved in neural differentiation in hESCs with a 20q11.21 mutation and *Bcl-xL* overexpressing cells are consistent with the known functions of CHCHD2. Finally, the TGF-β/SMAD signaling is known to have a pivotal role during the first-lineage commitment (Sumi et al., 2008). We therefore carried out functional tests, assessing the impact on neuroectoderm differentiation of siRNA knockdown of *CHCHD2* in control cells and *CHCHD2* mRNA transfection in mutant cells. Unexpectedly, we found that *CHCHD2* does not mediate the effect on differentiation under our experimental conditions. We have no immediate explanation for these differences, although it is worth noting that Zhu et al. did not assess ectodermal differentiation in the same manner as we did, i.e., using a TGF-β inhibitor. In their work, the differentiation bias was assessed using spontaneous differentiation without modulating the TGF-β signaling, which may explain some of the differences we see. Another point of interest is that, in the last part of their work, they show that the use of the TGF-β inhibitor SB431542 during differentiation cancels the effect of downregulating *CHCHD2* using siRNA. This conclusion is based on qPCR results for *PAX6* between control and siRNA-treated cells. Although the authors found a statistically significant upregulation of *PAX6*, the relative fold change was less than 0.3 which may be not biologically relevant. This is in line with our results for siRNA *CHCHD2* knockdown. Taken together, this suggests that, although Zhu et al. clearly prove that CHCHD2 interacts with SMAD4, the relationship between CHCHD2 and neuroectoderm commitment is less direct than hypothesized. Nevertheless, this gene appears as an excellent transcriptomic marker for the gain of 20q11.21.

In summary, our work shows the significant impact genetic changes can have on hPSCs, and provides, to the best of our knowledge, one of the few studies pinpointing a specific effect of a recurrent aberration and identifies the mechanisms behind it. The results of our study are of particular importance when taking into account that the 20q11.21 amplification occurs in more than 20% of hPSC lines worldwide and that G-banding as the most widely used method for karyotyping fails to detect it. The mutation may bias experimental data in a research setting if unnoticed, although it is unlikely that it would be undetected in a clinical setting, because in clinical trials the genetic screening protocols are more stringent. Nevertheless, if the mutation is in the process of culture take-over, it could remain undetected and result in poor differentiation in the clinically relevant cell types.

EXPERIMENTAL PROCEDURES

hESC Lines and Culture

hESC were derived and characterized as described previously (Mateizel et al., 2006, 2009). All lines are registered with the EU hPSC registry (https://hpscreg.eu/). hESCs were cultured on dishes coated with 10 μg/mL laminin-521 (Biolamina) in NutriStem hESC XF medium (NS medium; Biological Industries) with 100 U/mL penicillin/streptomycin (Thermo Fisher Scientific), and passaged as single cells in a 1:10 to 1:100 ration using TrypLE Express (Thermo Fisher Scientific) when 70%–80% confluent. The cells were grown at 37°C in 5% CO_2. VUB03_BcL-xL was generated by lentiviral transduction, as described by Nguyen et al. (2014) and in the Supplemental Information.

A large bulk of frozen cells was generated for each line. After thawing, the lines were used at the lowest passage possible (Table S1), and were checked for the 20q11.21 amplification by copy-number assay 3 to 4 passages after thawing.

aCGH

Oligonucleotide aCGH was carried out based on the protocol provided by the manufacturer (Agilent Technologies). A total of 400 ng of DNA was labeled with Cy3, while the reference DNA (Promega) was labeled with Cy5. The samples are hybridized on the microarray slide (4 × 44K Human Genome CGH Microarray, Agilent Technologies). The slides were scanned using an Agilent dual laser DNA microarray scanner G2566AA. Only arrays with an SD $\leq$ 0.20, signal intensity > 50, background noise < 5, and a derivative log-ratio < 0.2 were taken into account. Cutoff values were set at three consecutive probes with an average $\log_2$ ratio over 0.3 for gains and of −0.45 for loss.

Gene Copy Number Using Real-Time qPCR

Copy-number quantification was performed on the ViiA 7 thermocycler (Thermo Fisher Scientific), ViiA 7 software v.1.2 (Thermo Fisher Scientific), and Applied Biosystems Copy Caller v.2.1. We used the copy-number assays for RNAseP (4403326) as a reference and POFUT1 (Hs02487189_cn) to assess the number of copies of the 20q11.21 region.

RNA Sequencing

RNA (150 ng) was used to perform an Illumina sequencing library preparation using the QuantSeq 3′ mRNA-Seq Library Prep Kits (Lexogen) according to manufacturer's protocol. During library preparation 17 PCR cycles were used. Libraries were quantified by qPCR, according to Illumina's protocol "Sequencing Library qPCR Quantification protocol guide," version February 2011. A high-sensitivity DNA chip (Agilent Technologies) was used to control the library's size distribution and quality. Sequencing was performed on a high-throughput Illumina NextSeq 500 flow cell generating 75 bp single reads.

Details on the bioinformatics processing can be found in the Supplemental Information.

hESC Differentiation

The protocol for neuroectoderm was adapted from Chetty et al. (2013) and Chambers et al. (2009). hESCs were passaged on laminin-521, as described above, 1–2 days before EC differentiation in a ratio of 50,000 cells per cm^2 so that they were 90% confluent on the starting day. The neuroectoderm differentiation medium was refreshed daily, and consisted of KnockOut DMEM (Thermo Fisher Scientific) with 10% KnockOut Serum Replacement (Thermo Fisher Scientific) and supplemented with 500 ng/mL Recombinant Human Noggin Protein (R&D Systems) and 10 μM SB431542 (Tocris). Endoderm differentiation was carried out using a protocol based on Sui et al. (2012). hESCs were passaged on laminin-521, as described above, 1–2 days before differentiation in a ratio of 30,000 cells per cm^2 so that they were 80%–90% confluent on the starting day. Definitive endoderm differentiation medium contains RPMI 1640 supplemented by GlutaMAX (Thermo Fisher Scientific), 0.5% B27 supplement (Thermo Fisher Scientific), 100 ng recombinant human/mouse/rat activin A (R&D Systems) and 3 μM CHIR99021 (Stemgent). One day later the medium was changed to differentiation medium without CHIR99021 and the cells were cultured for 2 more days. For mesendoderm induction, the same protocol as for endoderm was used, for 1 day.

Protein and mRNA Analysis

Real-time qPCR was carried out on a ViiA 7 thermocycler (Thermo Fisher Scientific) and using standard protocol as provided by the manufacturer. Details on the probes, assays and primers are listed in the Supplemental Information. Sandwich ELISA was conducted using the human total Bcl-xL ELISA kit for Bcl-xL (DuoSet IC, R&D Systems) following the manufacturer's protocol. All samples were analyzed in triplicate. Immunostaining was carried out on cells fixed and permeabilized with 4% paraformaldehyde and 100% methanol (Sigma-Aldrich), and blocked with fetal bovine serum (Thermo Fischer Scientific). The list with antibodies can be found in the Supplemental Information. Imaging was performed on a LSM800 confocal microscope (Carl Zeiss), and cell counts were done using the Zen 2 (blue edition) imaging software.

Statistics

Data are presented as means. Unpaired two-tailed t test (GraphPad Prism 5.0b) was used to determine significance between groups. For all statistical tests, the 0.05 level of confidence was accepted for statistical significance.

siRNA Knockdown

The siRNAs used in this study were purchased from Healthcare/Dharmacon: ON-TARGET*plus* Human *CHCHD2* siRNA SMARTpool (cat. no L-019120-100005) and ON-TARGET*plus* no-targeting siRNA as a negative control (cat. no. D-001810-01-05). The siRNA was dissolved in 1× siRNA buffer and a final concentration of 50 nM was transfected with lipofectamine RNAiMAX (13778150) in Opti-MEM (11058021) following the manufacturer's protocols (Thermo Fischer Scientific).

Plasmid Preparation and *In Vitro* Transcription of Capped mRNA

The synthetic gene encoding for *CHCHD2* was purchased from Integrated DNA Technologies and cloned as NcoI-XhoI fragment in the vector pEtheRNA-v2 as previously described (Van Lint et al., 2016). *In vitro* mRNA transcription and quality controls were performed as described in Bonehill et al. (2004). In brief, pEtheRNA-v2 plasmid was linearized with BfuAI (Fermentas) and *in vitro* mRNA transcription was performed with T7 polymerase according to the manufacturer instructions. mRNA integrity, concentration, and purity were tested on Agilent 2100 Bioanalyzer (Agilent Technologies).

ACCESSION NUMBERS

All data have been deposited in the GEO repository with accession number GEO: GSE116372.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2019.05.005.

AUTHOR CONTRIBUTIONS

C.M. carried out all of the experiments unless stated otherwise and co-wrote the manuscript. E.C.D.D. did the bioinformatics analysis. M.R. performed the microscopy and cell counting. H.T.N. provided the *Bcl-xL* overexpressing line. F.Z. assisted with the gene expression analysis. A.K. provided the isochromosome line. D.D. provided the control samples for RNA sequencing. J.D.K. assisted with the microscopy. L.T. and F.V.N. performed the RNA sequencing and assisted in the bio-informatic analysis. L.F. supervised the mRNA production. K.S. proofread the paper. M.G. proofread the paper and supervised the experimental work. C.S. co-wrote the paper, designed and supervised the experimental work.

ACKNOWLEDGMENTS

The authors acknowledge their colleagues from the human embryonic stem cell lab for the derivation and culture of the lines and professor Kris Thielemans from the LMCT laboratory of the VUB for providing the mRNA for overexpressing *CHCHD2*. This work was supported by the Fonds for Scientific Research in Flanders (Fonds Wetenschappelijk Onderzoek – Vlaanderen [FWO]) and the Methusalem Grant of the Research Council of the VUB. C.M. is a doctoral fellow supported by the Instituut voor Innovatie door Wetenschap en Technologie (IWT). A.K. and D.D. are doctoral fellows at the FWO.

Received: July 10, 2018
Revised: May 6, 2019
Accepted: May 6, 2019
Published: June 6, 2019

REFERENCES

Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., Beil, S., et al.

(2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat. Biotechnol. *29*, 1132–1144.

Andrews, P.W., Ben-David, U., Benvenisty, N., Coffey, P., Eggan, K., Knowles, B.B., Nagy, A., Pera, M., Reubinoff, B., Rugg-Gunn, P.J., et al. (2017). Assessing the safety of human pluripotent stem cells and their derivatives for clinical applications. Stem Cell Reports *9*, 1–4.

Aras, S., Pak, O., Sommer, N., Finley, R., Hüttemann, M., Weissmann, N., Grossman, L.I., and Grossman, L.I. (2013). Oxygen-dependent expression of cytochrome c oxidase subunit 4-2 gene expression is mediated by transcription factors RBPJ, CXXC5 and CHCHD2. Nucleic Acids Res. *41*, 2255–2266.

Avery, S., Hirst, A.J., Baker, D., Lim, C.Y., Alagaratnam, S., Skotheim, R.I., Lothe, R.A., Pera, M.F., Colman, A., Robson, P., et al. (2013). BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. Stem Cell Reports *1*, 1–8.

Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Mol. Cell Biol. *17*, 170–182.

Baker, D.E.C., Harrison, N.J., Maltby, E., Smith, K., Moore, H.D., Shaw, P.J., Heath, P.R., Holden, H., and Andrews, P.W. (2007). Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. Nat. Biotechnol. *25*, 207–215.

Ben-David, U., Mayshar, Y., and Benvenisty, N. (2013). Virtual karyotyping of pluripotent stem cells on the basis of their global gene expression profiles. Nat. Protoc. *8*, 989–997.

Ben-David, U., Arad, G., Weissbein, U., Mandefro, B., Maimon, A., Golan-Lev, T., Narwani, K., Clark, A.T., Andrews, P.W., Benvenisty, N., et al. (2014). Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. Nat. Commun. *5*, 4825.

Bonehill, A., Heirman, C., Tuyaerts, S., Michiels, A., Breckpot, K., Brasseur, F., Zhang, Y., Van Der Bruggen, P., and Thielemans, K. (2004). Messenger RNA-electroporated dendritic cells presenting MAGE-A3 simultaneously in HLA class I and class II molecules. J. Immunol. *172*, 6649–6657.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chetty, S., Pagliuca, F.W., Honore, C., Kweudjeu, A., Rezania, A., and Melton, D.A. (2013). A simple tool to improve pluripotent stem cell differentiation. Nat. Methods *10*, 553–556.

Cowan, C.A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J.P., Wang, S., Morton, C.C., McMahon, A.P., Powers, D., et al. (2004). Derivation of embryonic stem-cell lines from human blastocysts. N. Engl. J. Med. *350*, 1353–1356.

Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A., and Andrews, P.W. (2004). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat. Biotechnol. *22*, 53–54.

Fazeli, A., Liew, C.-G., Matin, M.M., Elliott, S., Jeanmeure, L.F.C., Wright, P.C., Moore, H., and Andrews, P.W. (2011). Altered patterns of differentiation in karyotypically abnormal human embryonic stem cells. Int. J. Dev. Biol. *55*, 175–180.

Gopalakrishna-Pillai, S., and Iverson, L.E. (2010). Astrocytes derived from trisomic human embryonic stem cells express markers of astrocytic cancer cells and premalignant stem-like progenitors. BMC Med. Genomics *3*, 12.

Grossman, L.I., Purandare, N., Arshad, R., Gladyck, S., Somayajulu, M., Hüttemann, M., and Aras, S. (2017). MNRR1, a biorganellar regulator of mitochondria. Oxid. Med. Cell. Longev. *2017*, 6739236.

Herszfeld, D., Wolvetang, E., Langton-Bunker, E., Chung, T.-L., Filipczyk, A.A., Houssami, S., Jamshidi, P., Koh, K., Laslett, A.L., Michalska, A., et al. (2006). CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. Nat. Biotechnol. *24*, 351–357.

Inzunza, J., Sahlén, S., Holmberg, K., Strömberg, A.M., Teerijoki, H., Blennow, E., Hovatta, O., and Malmgren, H. (2004). Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. Mol. Hum. Reprod. *10*, 461–466.

Keller, A., Dziedzicka, D., Zambelli, F., Markouli, C., Sermon, K., Spits, C., and Geens, M. (2018). Genetic and epigenetic factors which modulate differentiation propensity in human pluripotent stem cells. Hum. Reprod. Update *24*, 162–175.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J.V., Lee, S., Barrero, M.J., et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell *8*, 106–118.

Lefort, N., Feyeux, M., Bas, C., Féraud, O., Bennaceur-Griscelli, A., Tachdjian, G., Peschanski, M., and Perrier, A.L. (2008). Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. Nat. Biotechnol. *26*, 1364–1366.

Van Lint, S., Renmans, D., Broos, K., Goethals, L., Maenhout, S., Benteyn, D., Goyvaerts, C., Du Four, S., Van der Jeught, K., Bialkowski, L., et al. (2016). Intratumoral delivery of TriMix mRNA results in T-cell activation by cross-presenting dendritic cells. Cancer Immunol. Res. *4*, 146–156.

Liu, Y., Clegg, H.V., Leslie, P.L., Di, J., Tollini, L.A., He, Y., Kim, T.-H., Jin, A., Graves, L.M., Zheng, J., et al. (2015). CHCHD2 inhibits apoptosis by interacting with Bcl-x L to regulate Bax activation. Cell Death Differ. *22*, 1035–1046.

Maitra, A., Arking, D.E., Shivapurkar, N., Ikeda, M., Stastny, V., Kassauei, K., Sui, G., Cutler, D.J., Liu, Y., Brimble, S.N., et al. (2005). Genomic alterations in cultured human embryonic stem cells. Nat. Genet. *37*, 1099–1103.

Mateizel, I., De Temmerman, N., Ullmann, U., Cauffman, G., Sermon, K., Van de Velde, H., De Rycke, M., Degreef, E., Devroey, P., Liebaers, I., et al. (2006). Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. Hum. Reprod. *21*, 503–511.

Mateizel, I., Spits, C., Verloes, A., Mertzanidou, A., Liebaers, I., and Sermon, K. (2009). Characterization of CD30 expression in human embryonic stem cell lines cultured in serum-free media and passaged mechanically. Hum. Reprod. *24*, 2477–2489.

Mitalipova, M.M., Rao, R.R., Hoyer, D.M., Johnson, J.A., Meisner, L.F., Jones, K.L., Dalton, S., and Stice, S.L. (2005). Preserving the genetic integrity of human embryonic stem cells. Nat. Biotechnol. *23*, 19–20.

Närvä, E., Autio, R., Rahkonen, N., Kong, L., Harrison, N., Kitsberg, D., Borghese, L., Itskovitz-Eldor, J., Rasool, O., Dvorak, P., et al. (2010). High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. Nat. Biotechnol. *28*, 371–377.

Nguyen, H.T., Geens, M., and Spits, C. (2013). Genetic and epigenetic instability in human pluripotent stem cells. Hum. Reprod. Update *19*, 187–205.

Nguyen, H.T., Geens, M., Mertzanidou, A., Jacobs, K., Heirman, C., Breckpot, K., and Spits, C. (2014). Gain of 20q11.21 in human embryonic stem cells improves cell survival by increased expression of Bcl-xL. Mol. Hum. Reprod. *20*, 168–177.

Serra, M., Brito, C., Correia, C., and Alves, P.M. (2012). Process engineering of human pluripotent stem cells for clinical application. Trends Biotechnol. *30*, 350–359.

Spits, C., Mateizel, I., Geens, M., Mertzanidou, A., Staessen, C., Vandeskelde, Y., Van der Elst, J., Liebaers, I., and Sermon, K. (2008). Recurrent chromosomal abnormalities in human embryonic stem cells. Nat. Biotechnol. *26*, 1361–1363.

Sui, L., Mfopou, J.K., Geens, M., Sermon, K., and Bouwens, L. (2012). FGF signaling via MAPK is required early and improves Activin A-induced definitive endoderm formation from human embryonic stem cells. Biochem. Biophys. Res. Commun. *426*, 380–385.

Sumi, T., Tsuneyoshi, N., Nakatsuji, N., and Suemori, H. (2008). Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. Development *135*, 2969–2979.

Teo, A.K.K., Arnold, S.J., Trotter, M.W.B., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. *25*, 238–250.

Werbowetski-Ogilvie, T.E., Bossé, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., Wynder, T., Smith, M.-J., Dingwall, S., Carter, T., et al. (2009). Characterization of human embryonic stem cells with features of neoplastic progression. Nat. Biotechnol. *27*, 91–97.

Wu, H., Kim, K.J., Mehta, K., Paxia, S., Sundstrom, A., Anantharaman, T., Kuraishy, A.I., Doan, T., Ghosh, J., Pyle, A.D., et al. (2008). Copy number variant analysis of human embryonic stem cells. Stem Cells *26*, 1484–1489.

Yang, S., Lin, G., Tan, Y., Zhou, D., Deng, L., Cheng, D., Luo, S., Liu, T., Zhou, X., Sun, Z., et al. (2008). Tumor progression of culture-adapted human embryonic stem cells during long-term culture. genes. Chromosomes Cancer *47*, 665–679.

Yang, S., Lin, G., Tan, Y.-Q., Deng, L.-Y., Yuan, D., and Lu, G.-X. (2010). Differences between karyotypically normal and abnormal human embryonic stem cells. Cell Prolif. *43*, 195–206.

Ying, L., Mills, J.A., French, D.L., and Gadue, P. (2015). OCT4 coordinates with WNT signaling to pre-pattern chromatin at the SOX17 locus during human ES cell differentiation into definitive endoderm. Stem Cell Reports *5*, 490–498.

Zhu, L., Gomez-Duran, A., Saretzki, G., Jin, S., Tilgner, K., Melguizo-Sanchis, D., Anyfantis, G., Al-Aama, J., Vallier, L., Chinnery, P., et al. (2016). The mitochondrial protein CHCHD2 primes the differentiation potential of human induced pluripotent stem cells to neuroectodermal lineages. J. Cell Biol. *215*, 187–202.

Stem Cell Reports, Volume *13*

Supplemental Information

Gain of 20q11.21 in Human Pluripotent Stem Cells Impairs TGF-β-Dependent Neuroectodermal Commitment

C. Markouli, E. Couvreu De Deckersberg, M. Regin, H.T. Nguyen, F. Zambelli, A. Keller, D. Dziedzicka, J. De Kock, L. Tilleman, F. Van Nieuwerburgh, L. Franceschini, K. Sermon, M. Geens, and C. Spits

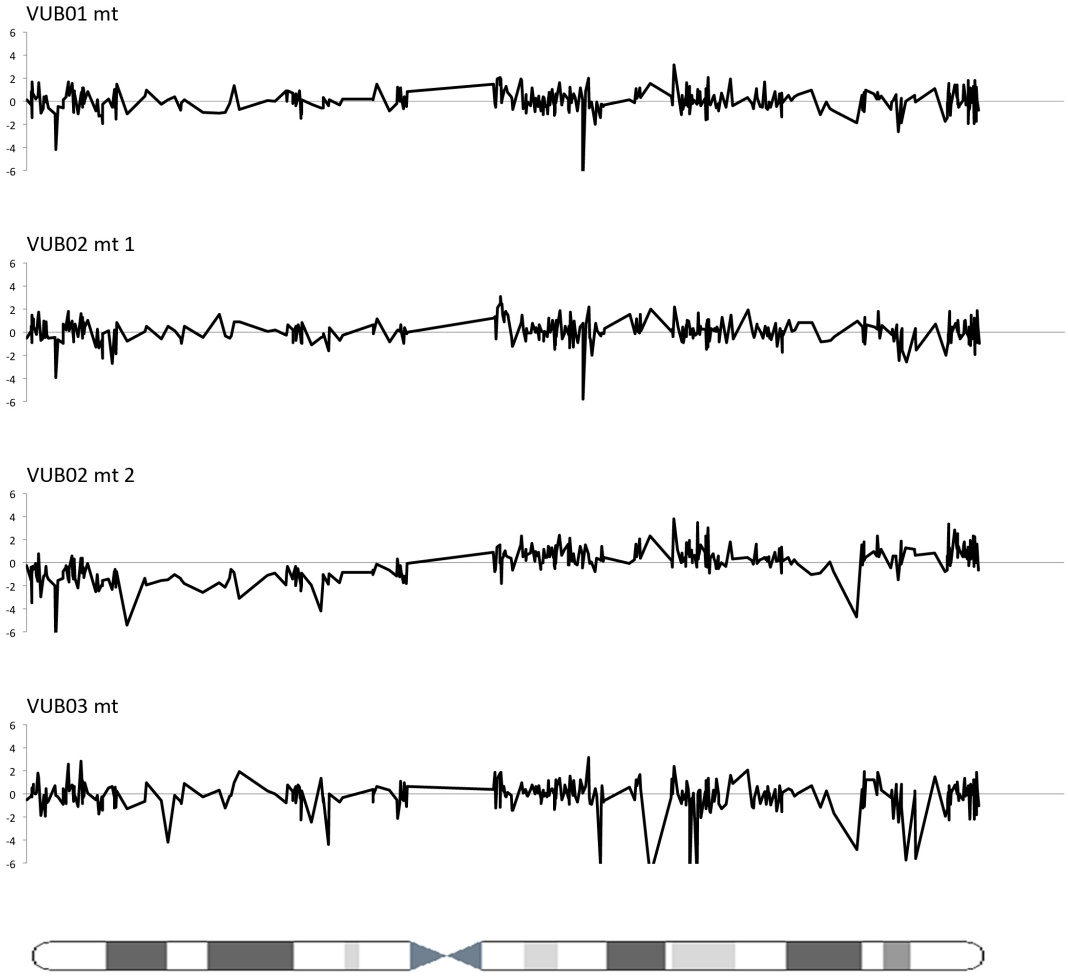


Figure S1 (relates to figure 1 in the main text). Raw gene expression data of all genes expressed on chromosome 20, for the four lines carrying a gain of 20q11.21. Each plot represents the differential gene expression for that line against the whole group of control lines. Data plotted as a Log_2 Fold Change.

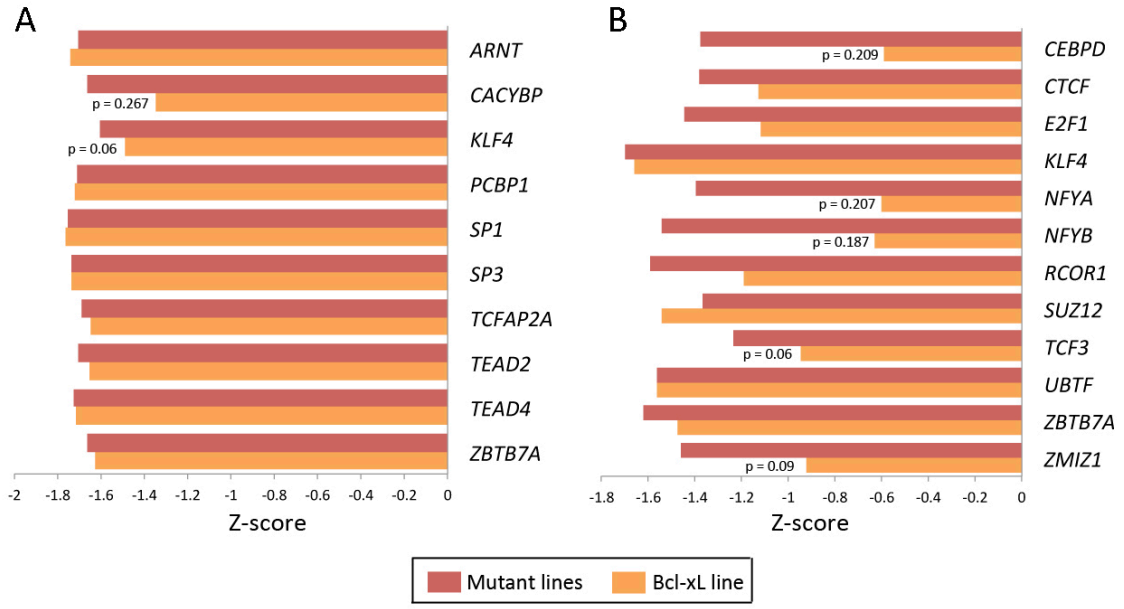


Figure S2 (relates to figure 2 in the main text). Supplementary Figure 2A shows an *in silico* prediction of the differentially activated transcription factors, based on the binding motifs identified at the promoter region of each gene that is differentially expressed between wt vs mt or wt vs VUB03_Bcl-xL (with a $|\log_2$ Fold Change$|>1$ and FDR<0.05). The prediction was done using the TRANSFAC and JASPAR databases in Enrichr (Chen et al., 2013; Kuleshov et al., 2016). The plot shows the Z-scores for the transcription factors. All results are statistically significant with a p-value<0.05 unless the value is specified. Supplementary figure 2B shows the results of the analysis using ENCODE and ChEA consensus, which searches for transcription binding models based on the ENCODE project and already published CHIP-seq data. The plot shows the Z-score for the transcription factors predicted to be differentially activated. All results are statistically significant with a p-value<0.05), unless otherwise indicated.

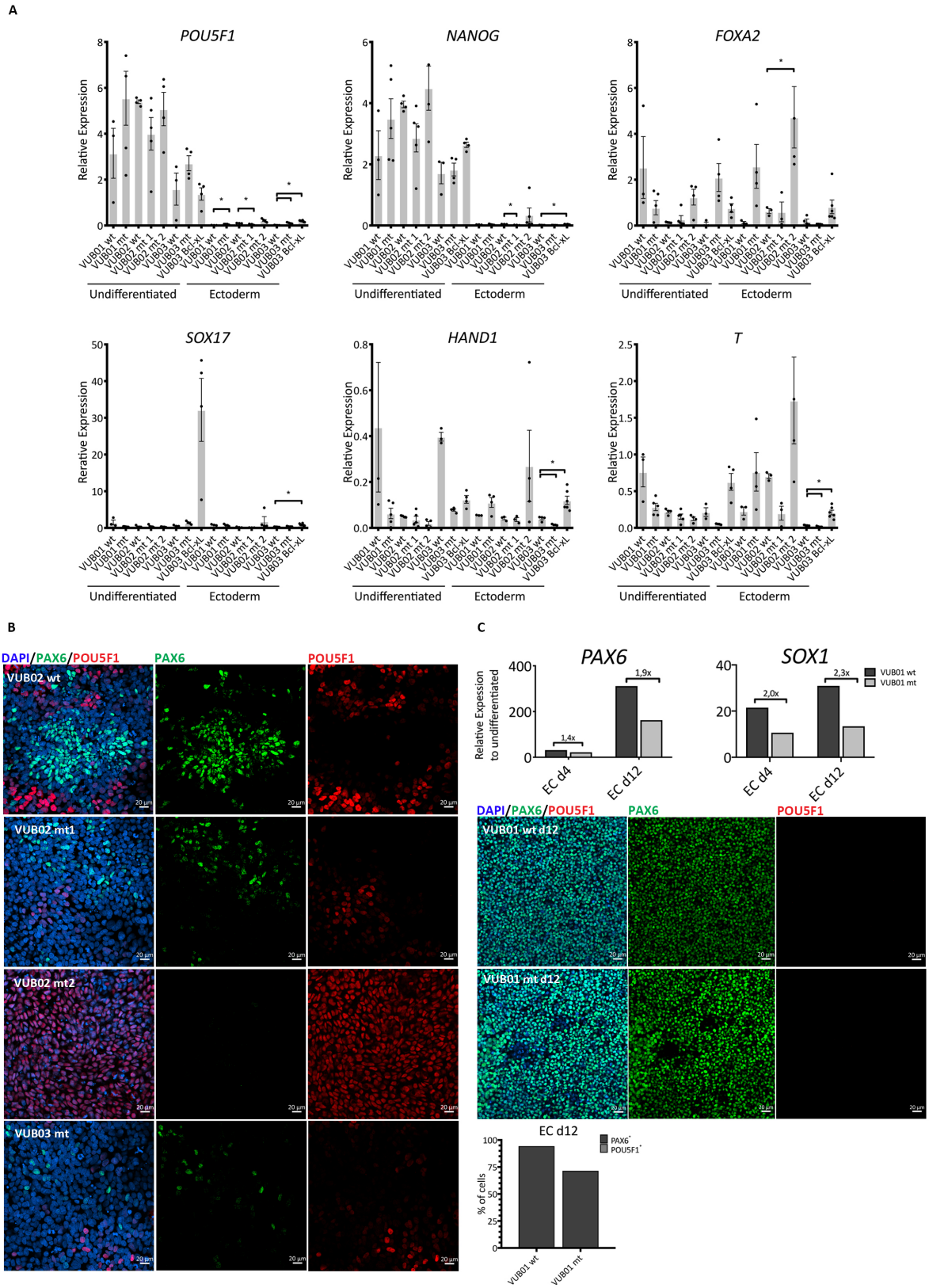


Figure S3 (relates to figure 3 in the main text). Figure S3A shows relative mRNA expression as measured by qPCR for *POU5F1, NANOG, FOXA2, SOX17, HAND1* and *T* after 4 days of neuroectoderm differentiation (n=3 to 5). Data are shown as mean ±SEM, each dot represents an independent differentiation experiment and the horizontal bars with asterisks represent statistical significance between samples (P<0.05, t-test).

Figure S3B shows examples of immunostaining for PAX6 (green) and POU5F1 (red) in the lines not shown in figure 3 of the main text (VUB02_wt, VUB02_mt1 and VUB02_mt2, VUB03_mt), after 4 days of neuroectoderm differentiation. Figure 3SC shows the results of a 12-day neuroectoderm differentiation. The top panel shows the relative mRNA expression as measured by qPCR for *PAX6* and *SOX1* in VUB01_wt and VUB01_mt after 12 days of differentiation, and as compared to a 4-days differentiation. Under, are examples of immunostaining for PAX6 (green) and POU5F1 (red) in VUB01_wt and VUB01_mt after 12 days of neuroectoderm differentiation. The lower panel shows a plot with the percentages of PAX6 and POU5F1-positive cells in VUB01_wt and VUB01_mt after 12 days of neuroectoderm differentiation (n=1).

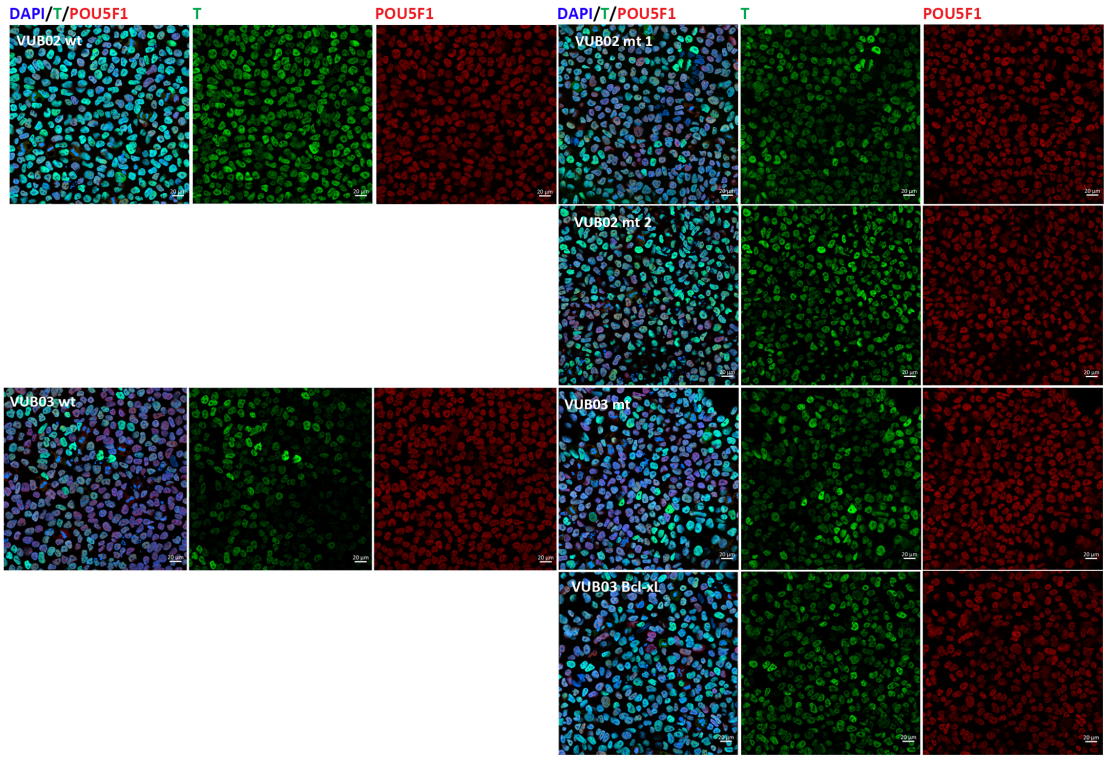


Figure S4, relates to figure 4 in main text. Immunostaining for T (green) and POU5F1 (red) after 24h mesendoderm induction for the lines not shown in figure 4A (VUB02_wt, VUB02_mt1 and VUB02_mt2, VUB03_wt versus VUB03_mt and VUB03_Bcl-xL.

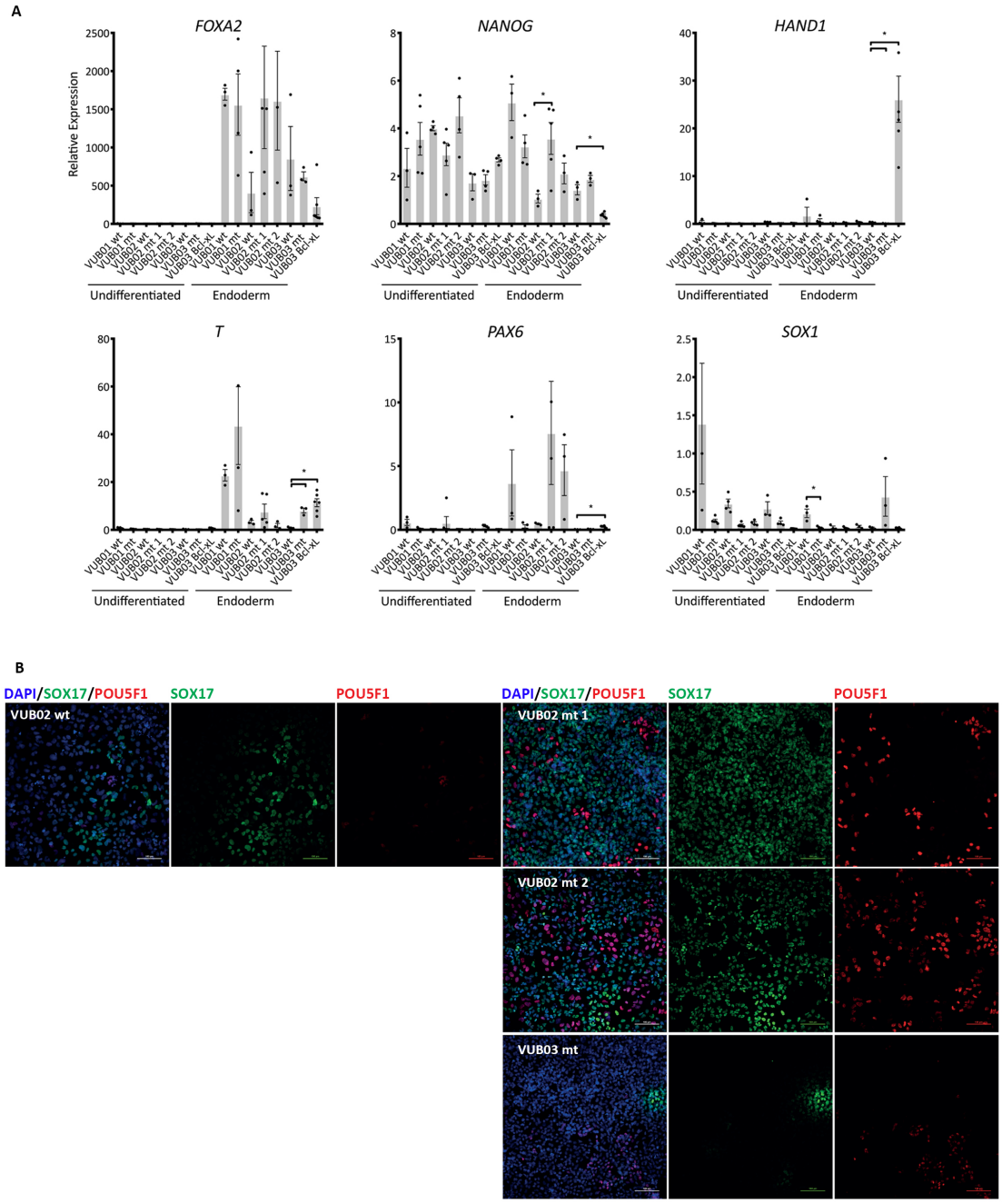


Figure S5 (relates to figure 4 in the main text). Figure S5A shows relative mRNA expression as measured by qPCR for *FOXA2, NANOG, HAND1, T, PAX6* and *SOX1* after definitive endoderm differentiation for all the lines (n=3 to 5). Data are shown as mean ±SEM, each dot represents an independent differentiation experiment and the horizontal bars with asterisks represent statistical significance between samples ($P<0.05$, t-test). Figure S5B shows immunostaining for SOX17 (green) and POU5F1 (red) after definitive endoderm differentiation in lines not shown in Figure 4 (VUB02_wt, VUB02_mt1, VUB02_mt2 and VUB03_mt).

Table S1. List of lines and sublines used in the study, passage range and details on the genetic content of the lines

Line	Passage range	Karyotype	Breakpoints (size)
VUB01	75-89	46, XY	
VUB01_mt	289-303	46, XY,dup(20)(q11.21)	20: 31300536 – 35306783 (3.9Mb)
VUB02	15	46, XY	
VUB02_mt1	377-383	46, XY,dup(20)(q11.21)	20: 31300536 – 32533536 (1.2Mb)
VUB02_mt2	19-25	46, XY,i(20)(p11.1)	20: 90359 – 26095364 (26Mb)
			20: 26108364 – 64262203 (36.3Mb)
VUB03	23-29	46, XX	
VUB03_mt	207-209	46, XX,dup(20)(q11.21)	20: 31300536 – 32212536 (0.9Mb)
VUB03_Bcl-xL	106-109	46, XX,dup(1)(q32.1q41)	1: 19705020 – 21919120 (22.1Mb)
VUB14	22-25	46, XX	

Table S2. Gene set enrichment analysis for the C2 library

NAME	SIZE	NES	FDR q-val
SHEN SMARCA2 TARGETS UP	413	-2.706	0.00
OUELLET OVARIAN CANCER INVASIVE VS LMP UP	113	-2.565	0.00
MILI PSEUDOPODIA HAPTOTAXIS UP	476	-2.508	0.00
SEIDEN ONCOGENESIS BY MET	82	-2.441	0.00
SENGUPTA NASOPHARYNGEAL CARCINOMA UP	254	-2.400	0.00
JOHNSTONE PARVB TARGETS 2 DN	309	-2.370	0.00
ZHANG BREAST CANCER PROGENITORS UP	394	-2.369	0.00
KAMMINGA EZH2 TARGETS	40	-2.333	0.00
DING LUNG CANCER EXPRESSION BY COPY NUMBER	96	-2.252	0.00
BIDUS METASTASIS UP	203	-2.230	0.00
DE YY1 TARGETS DN	88	-2.189	0.00
DAZARD RESPONSE TO UV SCC UP	106	-2.182	0.00
BORCZUK MALIGNANT MESOTHELIOMA UP	288	-2.130	0.00
ABRAMSON INTERACT WITH AIRE	42	-2.126	0.00
MALONEY RESPONSE TO 17AAG DN	77	-2.124	0.00

CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN	178	-2.114	0.00
RHEIN ALL GLUCOCORTICOID THERAPY DN	342	-2.107	0.00
GENTILE UV RESPONSE CLUSTER D4	54	-2.104	0.00
FISCHER G2 M CELL CYCLE	216	-2.101	0.00
SAKAI TUMOR INFILTRATING MONOCYTES DN	73	-2.100	0.00
SCHLOSSER MYC TARGETS REPRESSED BY SERUM	153	-2.099	0.00
MARTINEZ RESPONSE TO TRABECTEDIN DN	264	-2.098	0.00
IKEDA MIR30 TARGETS UP	111	-2.096	0.00
JAATINEN HEMATOPOIETIC STEM CELL UP	267	-2.095	0.00
DACOSTA UV RESPONSE VIA ERCC3 COMMON DN	463	-2.087	0.00
PYEON CANCER HEAD AND NECK VS CERVICAL UP	162	-2.068	0.00
ENK UV RESPONSE KERATINOCYTE DN	471	-2.041	0.00
GABRIELY MIR21 TARGETS	269	-2.038	0.00
RAMALHO STEMNESS UP	197	-2.036	0.00
WINNEPENNINCKX MELANOMA METASTASIS UP	149	-2.036	0.00
KENNY CTNNB1 TARGETS UP	41	-2.035	0.00
GARY CD5 TARGETS DN	394	-2.028	0.00
WAMUNYOKOLI OVARIAN CANCER LMP DN	176	-2.019	0.00
ONDER CDH1 TARGETS 1 DN	138	-2.017	0.00
IKEDA MIR1 TARGETS UP	52	-2.011	0.00
CHEOK RESPONSE TO HD MTX DN	23	-2.007	0.00
SPIRA SMOKERS LUNG CANCER UP	31	-2.006	0.00
HAHTOLA MYCOSIS FUNGOIDES CD4 DN	108	-2.005	0.00
LI WILMS TUMOR VS FETAL KIDNEY 2 UP	24	-2.000	0.00
NIKOLSKY BREAST CANCER 8Q12 Q22 AMPLICON	105	-1.986	0.00
DEBIASI APOPTOSIS BY REOVIRUS INFECTION UP	285	-1.984	0.00
YU MYC TARGETS UP	39	-1.980	0.00
WELCSH BRCA1 TARGETS UP	186	-1.978	0.00
FEVR CTNNB1 TARGETS DN	499	-1.978	0.00
GENTILE UV HIGH DOSE DN	290	-1.975	0.00
FLECHNER BIOPSY KIDNEY TRANSPLANT REJECTED VS OK DN	479	-1.973	0.00

BOYAULT LIVER CANCER SUBCLASS G3 UP	181	-1.968	0.00
REACTOME METABOLISM OF NON CODING RNA	47	-1.958	0.00
WILCOX RESPONSE TO PROGESTERONE UP	125	-1.951	0.00
DAZARD RESPONSE TO UV NHEK DN	296	-1.951	0.00
GREENBAUM E2A TARGETS UP	31	-1.947	0.00
PENG GLUCOSE DEPRIVATION DN	150	-1.945	0.00
CAIRO LIVER DEVELOPMENT UP	148	-1.942	0.00
FLORIO NEOCORTEX BASAL RADIAL GLIA DN	182	-1.940	0.00
XU HGF TARGETS INDUCED BY AKT1 48HR DN	24	-1.939	0.00
HORIUCHI WTAP TARGETS DN	284	-1.938	0.00
FERREIRA EWINGS SARCOMA UNSTABLE VS STABLE UP	144	-1.938	0.00
SENGUPTA NASOPHARYNGEAL CARCINOMA WITH LMP1 UP	329	-1.937	0.00
HU ANGIOGENESIS DN	35	-1.935	0.00
WHITFIELD CELL CYCLE S	144	-1.935	0.00
CHEN HOXA5 TARGETS 9HR UP	201	-1.934	0.00
LOPEZ MBD TARGETS IMPRINTED AND X LINKED	15	-1.932	0.00
GRAHAM NORMAL QUIESCENT VS NORMAL DIVIDING DN	80	-1.922	0.00
PUJANA XPRSS INT NETWORK	157	-1.922	0.00
BONOME OVARIAN CANCER POOR SURVIVAL UP	30	-1.919	0.00
VANTVEER BREAST CANCER METASTASIS DN	106	-1.918	0.00
KIM WT1 TARGETS DN	410	-1.914	0.00
TIEN INTESTINE PROBIOTICS 2HR DN	84	-1.912	0.01
KEGG CELL CYCLE	120	-1.903	0.01
ZHENG FOXP3 TARGETS IN T LYMPHOCYTE DN	29	-1.900	0.01
IKEDA MIR133 TARGETS UP	41	-1.896	0.01
KANG DOXORUBICIN RESISTANCE UP	50	-1.882	0.01
RHODES UNDIFFERENTIATED CANCER	65	-1.881	0.01
ROME INSULIN TARGETS IN MUSCLE UP	391	-1.880	0.01
SAKAI CHRONIC HEPATITIS VS LIVER CANCER UP	73	-1.879	0.01
ZHANG TLX TARGETS 36HR DN	181	-1.878	0.01
REACTOME CLEAVAGE OF GROWING TRANSCRIPT IN THE TERMINATION REGION	42	-1.875	0.01

CHIARADONNA NEOPLASTIC TRANSFORMATION KRAS UP	111	-1.870	0.01
LEE EARLY T LYMPHOCYTE UP	88	-1.869	0.01
RODWELL AGING KIDNEY NO BLOOD DN	138	-1.868	0.01
MORI IMMATURE B LYMPHOCYTE DN	87	-1.867	0.01
LEE LIVER CANCER SURVIVAL DN	163	-1.860	0.01
LY AGING OLD DN	54	-1.854	0.01
BURTON ADIPOGENESIS 12	31	-1.848	0.01
ALONSO METASTASIS UP	173	-1.846	0.01
REACTOME PROCESSING OF CAPPED INTRON CONTAINING PRE MRNA	132	-1.844	0.01
CHIARETTI T ALL RELAPSE PROGNOSIS	16	-1.844	0.01
WU APOPTOSIS BY CDKN1A VIA TP53	52	-1.843	0.01
VERHAAK GLIOBLASTOMA NEURAL	106	-1.842	0.01
MARKEY RB1 ACUTE LOF UP	218	-1.836	0.01
MMS MOUSE LYMPH HIGH 4HRS UP	32	-1.829	0.01
MOREAUX MULTIPLE MYELOMA BY TACI DN	153	-1.829	0.01
BROWNE HCMV INFECTION 10HR DN	48	-1.828	0.01
TOOKER GEMCITABINE RESISTANCE DN	118	-1.827	0.01
LEE RECENT THYMIC EMIGRANT	182	-1.823	0.01
RICKMAN METASTASIS UP	309	-1.822	0.01
MAYBURD RESPONSE TO L663536 DN	50	-1.822	0.01
OSMAN BLADDER CANCER UP	360	-1.821	0.01
WHITEFORD PEDIATRIC CANCER MARKERS	110	-1.821	0.01
EPPERT PROGENITOR	119	-1.818	0.01
TURASHVILI BREAST DUCTAL CARCINOMA VS LOBULAR NORMAL UP	69	-1.818	0.01
MISSIAGLIA REGULATED BY METHYLATION DN	113	-1.818	0.01
ZHANG TLX TARGETS DN	87	-1.815	0.01
PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA DN	122	-1.814	0.01
ZHENG FOXP3 TARGETS UP	22	-1.812	0.01
ROSTY CERVICAL CANCER PROLIFERATION CLUSTER	132	-1.811	0.01
BENPORATH ES 1	355	-1.808	0.02

ZHANG TLX TARGETS 60HR DN	257	-1.807	0.02
ZHENG FOXP3 TARGETS IN THYMUS UP	170	-1.805	0.02
FINETTI BREAST CANCERS KINOME GRAY	15	-1.804	0.02
REACTOME MRNA 3 END PROCESSING	33	-1.799	0.02
SEIDEN MET SIGNALING	19	-1.798	0.02
MORI PRE BI LYMPHOCYTE UP	72	-1.795	0.02
PECE MAMMARY STEM CELL DN	127	-1.795	0.02
BENPORATH PROLIFERATION	134	-1.795	0.02
BOYAULT LIVER CANCER SUBCLASS G23 UP	49	-1.794	0.02
THUM SYSTOLIC HEART FAILURE UP	313	-1.793	0.02
LE EGR2 TARGETS UP	103	-1.793	0.02
FLECHNER PBL KIDNEY TRANSPLANT OK VS DONOR DN	36	-1.791	0.02
BOYAULT LIVER CANCER SUBCLASS G123 UP	44	-1.791	0.02
BLALOCK ALZHEIMERS DISEASE INCIPIENT DN	151	-1.790	0.02
TARTE PLASMA CELL VS PLASMABLAST DN	288	-1.789	0.02
BIOCARTA PROTEASOME PATHWAY	28	-1.787	0.02
BECKER TAMOXIFEN RESISTANCE DN	45	-1.776	0.02
BIOCARTA CHEMICAL PATHWAY	22	-1.775	0.02
HEDENFALK BREAST CANCER BRCA1 VS BRCA2	153	-1.775	0.02
HADDAD T LYMPHOCYTE AND NK PROGENITOR UP	68	-1.774	0.02
CHIARADONNA NEOPLASTIC TRANSFORMATION CDC25 UP	104	-1.770	0.02
ALONSO METASTASIS EMT UP	30	-1.770	0.02
DEN INTERACT WITH LCA5	25	-1.770	0.02
CHANG CYCLING GENES	135	-1.769	0.02
CHESLER BRAIN HIGHEST EXPRESSION	36	-1.766	0.02
BAELDE DIABETIC NEPHROPATHY DN	400	-1.766	0.02
TURASHVILI BREAST DUCTAL CARCINOMA VS DUCTAL NORMAL UP	40	-1.764	0.02
VILLANUEVA LIVER CANCER KRT19 UP	158	-1.762	0.02
REICHERT MITOSIS LIN9 TARGETS	27	-1.761	0.02
REACTOME MRNA PROCESSING	149	-1.759	0.02
NIKOLSKY BREAST CANCER 6P24 P22 AMPLICON	19	-1.758	0.02

WANG SMARCE1 TARGETS DN	328	-1.758	0.02
BOYAULT LIVER CANCER SUBCLASS G12 UP	36	-1.757	0.02
REACTOME SIGNALING BY TGF BETA RECEPTOR COMPLEX	60	-1.757	0.02
JI RESPONSE TO FSH DN	56	-1.757	0.02
KYNG WERNER SYNDROM AND NORMAL AGING UP	81	-1.754	0.02
TCGA GLIOBLASTOMA COPY NUMBER DN	24	-1.753	0.02
ZHOU CELL CYCLE GENES IN IR RESPONSE 24HR	116	-1.750	0.02
FINETTI BREAST CANCER KINOME RED	16	-1.749	0.02
LU EZH2 TARGETS DN	338	-1.748	0.02
GRAHAM CML DIVIDING VS NORMAL QUIESCENT UP	161	-1.746	0.02
VERNELL RETINOBLASTOMA PATHWAY UP	68	-1.746	0.02
KIM GERMINAL CENTER T HELPER UP	55	-1.743	0.02
REACTOME DOWNREGULATION OF SMAD2 3 SMAD4 TRANSCRIPTIONAL ACTIVITY	19	-1.742	0.03
PUJANA BRCA CENTERED NETWORK	114	-1.741	0.03
ODONNELL TFRC TARGETS DN	114	-1.740	0.03
KRIEG KDM3A TARGETS NOT HYPOXIA	166	-1.739	0.03
APPIERTO RESPONSE TO FENRETINIDE DN	46	-1.735	0.03
REACTOME TRANSPORT OF MATURE TRANSCRIPT TO CYTOPLASM	51	-1.733	0.03
LINDGREN BLADDER CANCER CLUSTER 3 UP	293	-1.723	0.03
WANG CLIM2 TARGETS DN	173	-1.721	0.03
REACTOME TRANSCRIPTIONAL ACTIVITY OF SMAD2 SMAD3 SMAD4 HETEROTRIMER	36	-1.721	0.03
GENTILE UV RESPONSE CLUSTER D6	34	-1.718	0.03
LAIHO COLORECTAL CANCER SERRATED UP	102	-1.716	0.03
REACTOME REGULATION OF MITOTIC CELL CYCLE	75	-1.715	0.03
CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP	160	-1.711	0.03
MARTORIATI MDM4 TARGETS NEUROEPITHELIUM DN	121	-1.706	0.03
SHEDDEN LUNG CANCER POOR SURVIVAL A6	410	-1.706	0.03
REACTOME PROCESSING OF CAPPED INTRONLESS PRE MRNA	23	-1.705	0.03
KOBAYASHI EGFR SIGNALING 24HR DN	238	-1.705	0.03
SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP	141	-1.705	0.03

SARRIO EPITHELIAL MESENCHYMAL TRANSITION UP	168	-1.704	0.03
BROWNE HCMV INFECTION 18HR UP	157	-1.704	0.03
PEDERSEN METASTASIS BY ERBB2 ISOFORM 3	15	-1.704	0.03
YANG BREAST CANCER ESR1 LASER DN	47	-1.704	0.03
RODWELL AGING KIDNEY DN	125	-1.703	0.03
AIYAR COBRA1 TARGETS UP	33	-1.701	0.03
REACTOME MRNA SPLICING	104	-1.701	0.03
STEIN ESR1 TARGETS	78	-1.699	0.04
DITTMER PTHLH TARGETS UP	108	-1.699	0.04
ZHENG BOUND BY FOXP3	396	-1.698	0.04
GOLDRATH ANTIGEN RESPONSE	285	-1.697	0.04
YANAGIHARA ESX1 TARGETS	27	-1.695	0.04
PUIFFE INVASION INHIBITED BY ASCITES DN	136	-1.692	0.04
WANG TUMOR INVASIVENESS DN	196	-1.690	0.04
ISHIDA E2F TARGETS	49	-1.690	0.04
GENTILE UV RESPONSE CLUSTER D2	39	-1.690	0.04
GOLDRATH HOMEOSTATIC PROLIFERATION	159	-1.690	0.04
CHARAFE BREAST CANCER LUMINAL VS BASAL DN	381	-1.689	0.04
REACTOME G1 S TRANSITION	106	-1.689	0.04
FOURNIER ACINAR DEVELOPMENT LATE 2	264	-1.689	0.04
IGLESIAS E2F TARGETS UP	132	-1.686	0.04
TERAO AOX4 TARGETS SKIN DN	19	-1.686	0.04
PETRETTO CARDIAC HYPERTROPHY	33	-1.684	0.04
DEURIG T CELL PROLYMPHOCYTIC LEUKEMIA DN	244	-1.683	0.04
EGUCHI CELL CYCLE RB1 TARGETS	23	-1.682	0.04
MOHANKUMAR HOXA1 TARGETS UP	370	-1.678	0.04
WONG EMBRYONIC STEM CELL CORE	330	-1.678	0.04
RUIZ TNC TARGETS DN	132	-1.677	0.04
REACTOME PHOSPHORYLATION OF THE APC C	16	-1.675	0.04
AMUNDSON GAMMA RADIATION RESPONSE	38	-1.675	0.04
BROWNE HCMV INFECTION 6HR DN	143	-1.674	0.04

DACOSTA UV RESPONSE VIA ERCC3 XPCS DN	79	-1.671	0.04
MOLENAAR TARGETS OF CCND1 AND CDK4 DN	49	-1.668	0.04
PAL PRMT5 TARGETS UP	194	-1.667	0.04
MATZUK EMBRYONIC GERM CELL	15	-1.667	0.04
BONCI TARGETS OF MIR15A AND MIR16 1	81	-1.667	0.04
SCIBETTA KDM5B TARGETS DN	77	-1.666	0.04
GAZDA DIAMOND BLACKFAN ANEMIA PROGENITOR DN	61	-1.665	0.04
BIOCARTA CTCF PATHWAY	20	-1.664	0.04
PETROVA ENDOTHELIUM LYMPHATIC VS BLOOD UP	114	-1.664	0.04
HOSHIDA LIVER CANCER SUBCLASS S2	109	-1.663	0.04
SASAKI ADULT T CELL LEUKEMIA	159	-1.662	0.04
KAYO AGING MUSCLE DN	111	-1.661	0.04
EHLERS ANEUPLOIDY UP	36	-1.661	0.04
FARMER BREAST CANCER CLUSTER 2	32	-1.661	0.04
DAZARD UV RESPONSE CLUSTER G6	142	-1.659	0.04
KONG E2F3 TARGETS	87	-1.659	0.04
BIOCARTA CCR3 PATHWAY	18	-1.658	0.04
MATTIOLI MGUS VS PCL	84	-1.658	0.04
JIANG AGING CEREBRAL CORTEX UP	32	-1.658	0.04
BURTON ADIPOGENESIS 3	98	-1.653	0.05
REACTOME MITOTIC G1 G1 S PHASES	128	-1.649	0.05
PLASARI TGFB1 TARGETS 10HR DN	201	-1.649	0.05
SUNG METASTASIS STROMA DN	45	-1.649	0.05
FRASOR RESPONSE TO SERM OR FULVESTRANT UP	20	-1.648	0.05
KOKKINAKIS METHIONINE DEPRIVATION 96HR UP	104	-1.647	0.05
STEARMAN LUNG CANCER EARLY VS LATE UP	117	-1.645	0.05
XU HGF SIGNALING NOT VIA AKT1 48HR DN	20	-1.645	0.05
PUJANA BRCA2 PCC NETWORK	380	-1.644	0.05
PID NFKAPPAB CANONICAL PATHWAY	21	-1.644	0.05
REACTOME INHIBITION OF THE PROTEOLYTIC ACTIVITY OF APC C REQUIRED FOR THE ONSET OF ANAPHASE BY MITOTIC SPINDLE CHECKPOINT COMPONENTS	17	-1.641	0.05

BERENJENO TRANSFORMED BY RHOA UP	491	-1.641	0.05
NIKOLSKY BREAST CANCER 16P13 AMPLICON	87	2.136	0.01
LU EZH2 TARGETS UP	239	2.041	0.02
NIKOLSKY BREAST CANCER 7P22 AMPLICON	30	2.005	0.03
LI DCP2 BOUND MRNA	83	2.005	0.02
SCHLOSSER SERUM RESPONSE UP	114	1.997	0.02
SENGUPTA EBNA1 ANTICORRELATED	122	1.997	0.01
DIRMEIER LMP1 RESPONSE LATE DN	28	1.990	0.01
WALLACE JAK2 TARGETS UP	22	1.983	0.01
NIKOLSKY BREAST CANCER 16Q24 AMPLICON	43	1.958	0.02
REACTOME FORMATION OF THE TERNARY COMPLEX AND SUBSEQUENTLY THE 43S COMPLEX	48	1.950	0.02
DACOSTA UV RESPONSE VIA ERCC3 UP	292	1.947	0.02
RICKMAN METASTASIS DN	211	1.931	0.02
CERIBELLI GENES INACTIVE AND BOUND BY NFY	20	1.930	0.02
HAMAI APOPTOSIS VIA TRAIL DN	155	1.926	0.02
MARTENS TRETINOIN RESPONSE UP	440	1.896	0.03
CHNG MULTIPLE MYELOMA HYPERPLOID UP	50	1.872	0.03
MIKKELSEN MCV6 HCP WITH H3K27ME3	292	1.860	0.04
KESHELAVA MULTIPLE DRUG RESISTANCE	68	1.854	0.04
GINESTIER BREAST CANCER ZNF217 AMPLIFIED DN	293	1.847	0.04
SCHLOSSER MYC AND SERUM RESPONSE SYNERGY	32	1.845	0.04
MEDINA SMARCA4 TARGETS	33	1.843	0.04
HOLLEMAN ASPARAGINASE RESISTANCE B ALL UP	24	1.828	0.05
KORKOLA CORRELATED WITH POU5F1	32	1.826	0.04
NAKAMURA METASTASIS MODEL UP	35	1.817	0.05

Table S3. List of 100 top deregulated genes

Transcript name	LogFC	FDR	Gene_function
NPIPA8	7,46	1,4E-05	nuclear pore complex interacting protein family member A8 [Source:HGNC Symbol;Acc:HGNC:41983]
HIST1H4K	7,18	5,2E-04	histone cluster 1 H4 family member k [Source:HGNC Symbol;Acc:HGNC:4784]
CSAG1	6,80	3,5E-05	chondrosarcoma associated gene 1 [Source:HGNC Symbol;Acc:HGNC:24294]
INO80B-WBP1	6,72	9,7E-04	INO80B-WBP1 readthrough (NMD candidate) [Source:HGNC Symbol;Acc:HGNC:49199]
TTR	6,06	3,8E-08	transthyretin [Source:HGNC Symbol;Acc:HGNC:12405]
C15orf38-AP3S2	5,76	1,7E-06	C15orf38-AP3S2 readthrough [Source:HGNC Symbol;Acc:HGNC:38824]
CSAG2	5,70	1,1E-04	CSAG family member 2 [Source:HGNC Symbol;Acc:HGNC:16847]
CSAG3	5,70	1,1E-04	CSAG family member 3 [Source:HGNC Symbol;Acc:HGNC:26237]
MGAT2	5,54	2,7E-03	mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase [Source:HGNC Symbol;Acc:HGNC:7045]
MAGEA12	5,32	1,9E-04	MAGE family member A12 [Source:HGNC Symbol;Acc:HGNC:6799]
TBX1	5,27	3,2E-06	T-box 1 [Source:HGNC Symbol;Acc:HGNC:11592]
CXCL11	5,25	2,2E-04	C-X-C motif chemokine ligand 11 [Source:HGNC Symbol;Acc:HGNC:10638]
DAZ1	4,06	6,0E-04	deleted in azoospermia 1 [Source:HGNC Symbol;Acc:HGNC:2682]
TUBB8	3,97	2,6E-09	tubulin beta 8 class VIII [Source:HGNC Symbol;Acc:HGNC:20773]
NKX6-1	3,96	8,0E-04	NK6 homeobox 1 [Source:HGNC Symbol;Acc:HGNC:7839]
NOXO1	3,87	3,0E-05	NADPH oxidase organizer 1 [Source:HGNC Symbol;Acc:HGNC:19404]
RAC3	3,84	1,6E-12	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3) [Source:HGNC Symbol;Acc:HGNC:9803]
HS3ST6	3,76	2,8E-04	heparan sulfate-glucosamine 3-sulfotransferase 6 [Source:HGNC Symbol;Acc:HGNC:14178]
OC90	3,62	2,4E-06	otoconin 90 [Source:HGNC Symbol;Acc:HGNC:8100]
CHKB-CPT1B	3,61	1,5E-03	CHKB-CPT1B readthrough (NMD candidate) [Source:HGNC Symbol;Acc:HGNC:41998]
DEFB124	3,57	1,6E-04	defensin beta 124 [Source:HGNC Symbol;Acc:HGNC:18104]
NPHP3-ACAD11	3,55	1,9E-04	NPHP3-ACAD11 readthrough (NMD candidate) [Source:HGNC Symbol;Acc:HGNC:48351]
MYBL2	3,27	7,5E-12	MYB proto-oncogene like 2 [Source:HGNC Symbol;Acc:HGNC:7548]
NLRP5	3,27	2,7E-03	NLR family pyrin domain containing 5 [Source:HGNC Symbol;Acc:HGNC:21269]
MEF2B	3,26	2,2E-03	myocyte enhancer factor 2B [Source:HGNC Symbol;Acc:HGNC:6995]
EMX2	3,25	1,3E-02	empty spiracles homeobox 2 [Source:HGNC Symbol;Acc:HGNC:3341]
SGCA	3,20	1,9E-03	sarcoglycan alpha [Source:HGNC Symbol;Acc:HGNC:10805]
GLS2	3,17	4,5E-05	glutaminase 2 [Source:HGNC Symbol;Acc:HGNC:29570]
RBAK-RBAKDN	3,15	1,2E-02	RBAK-RBAKDN readthrough [Source:HGNC Symbol;Acc:HGNC:42971]

POLR2J2	3,08	1,7E-02	RNA polymerase II subunit J2 [Source:NCBI gene;Acc:246721]
TFAP2E	3,06	2,2E-03	transcription factor AP-2 epsilon [Source:HGNC Symbol;Acc:HGNC:30774]
POTEE	3,05	2,8E-03	POTE ankyrin domain family member E [Source:HGNC Symbol;Acc:HGNC:33895]
RNASEK-C17orf49	3,04	1,8E-13	RNASEK-C17orf49 readthrough [Source:HGNC Symbol;Acc:HGNC:44419]
TMEM121	3,04	7,8E-10	transmembrane protein 121 [Source:HGNC Symbol;Acc:HGNC:20511]
DPP7	3,01	4,9E-11	dipeptidyl peptidase 7 [Source:HGNC Symbol;Acc:HGNC:14892]
GDF5	2,99	2,2E-04	growth differentiation factor 5 [Source:HGNC Symbol;Acc:HGNC:4220]
PNMA6F	2,97	9,1E-05	paraneoplastic Ma antigen family member 6F [Source:HGNC Symbol;Acc:HGNC:53119]
ZIC1	2,96	2,4E-03	Zic family member 1 [Source:HGNC Symbol;Acc:HGNC:12872]
RNPEPL1	2,95	3,9E-12	arginyl aminopeptidase like 1 [Source:HGNC Symbol;Acc:HGNC:10079]
PHACTR3	2,95	1,9E-04	phosphatase and actin regulator 3 [Source:HGNC Symbol;Acc:HGNC:15833]
IGLON5	2,94	8,9E-14	IgLON family member 5 [Source:HGNC Symbol;Acc:HGNC:34550]
CDC34	2,93	6,4E-18	cell division cycle 34 [Source:HGNC Symbol;Acc:HGNC:1734]
EVA1B	2,93	3,2E-08	eva-1 homolog B [Source:HGNC Symbol;Acc:HGNC:25558]
RBM46	2,92	6,2E-05	RNA binding motif protein 46 [Source:HGNC Symbol;Acc:HGNC:28401]
SP140L	2,91	2,1E-04	SP140 nuclear body protein like [Source:HGNC Symbol;Acc:HGNC:25105]
NTAN1	2,90	1,0E-06	N-terminal asparagine amidase [Source:HGNC Symbol;Acc:HGNC:29909]
GOLGA6D	2,89	6,9E-03	golgin A6 family member D [Source:HGNC Symbol;Acc:HGNC:32204]
GRIN1	2,89	2,1E-04	glutamate ionotropic receptor NMDA type subunit 1 [Source:HGNC Symbol;Acc:HGNC:4584]
UTF1	2,88	1,6E-06	undifferentiated embryonic cell transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12634]
CLDN16	2,88	3,5E-03	claudin 16 [Source:HGNC Symbol;Acc:HGNC:2037]
HHEX	-2,53	4,5E-04	hematopoietically expressed homeobox [Source:HGNC Symbol;Acc:HGNC:4901]
ZNF280C	-2,55	2,0E-07	zinc finger protein 280C [Source:HGNC Symbol;Acc:HGNC:25955]
PLD6	-2,57	5,3E-03	phospholipase D family member 6 [Source:HGNC Symbol;Acc:HGNC:30447]
S100A6	-2,57	3,4E-04	S100 calcium binding protein A6 [Source:HGNC Symbol;Acc:HGNC:10496]
ISY1-RAB43	-2,59	3,0E-03	ISY1-RAB43 readthrough [Source:HGNC Symbol;Acc:HGNC:42969]
MAGEL2	-2,61	4,0E-05	MAGE family member L2 [Source:HGNC Symbol;Acc:HGNC:6814]
EPHA6	-2,63	5,5E-04	EPH receptor A6 [Source:HGNC Symbol;Acc:HGNC:19296]
CHCHD2	-2,69	2,5E-03	coiled-coil-helix-coiled-coil-helix domain containing 2 [Source:HGNC Symbol;Acc:HGNC:21645]
WFDC1	-2,69	1,7E-04	WAP four-disulfide core domain 1 [Source:HGNC Symbol;Acc:HGNC:15466]
RWDD2B	-2,71	2,7E-03	RWD domain containing 2B [Source:HGNC

			Symbol;Acc:HGNC:1302]
NPIPA7	-2,77	2,3E-02	nuclear pore complex interacting protein family member A7 [Source:HGNC Symbol;Acc:HGNC:41982]
FAM71D	-2,79	4,5E-05	family with sequence similarity 71 member D [Source:HGNC Symbol;Acc:HGNC:20101]
VANGL2	-2,80	2,0E-10	VANGL planar cell polarity protein 2 [Source:HGNC Symbol;Acc:HGNC:15511]
INPP4B	-2,80	5,3E-03	inositol polyphosphate-4-phosphatase type II B [Source:HGNC Symbol;Acc:HGNC:6075]
ANKS1B	-2,83	1,9E-08	ankyrin repeat and sterile alpha motif domain containing 1B [Source:HGNC Symbol;Acc:HGNC:24600]
NTS	-2,84	5,8E-07	neurotensin [Source:HGNC Symbol;Acc:HGNC:8038]
MAGEB3	-2,89	1,8E-03	MAGE family member B3 [Source:HGNC Symbol;Acc:HGNC:6810]
HIBADH	-2,90	2,8E-07	3-hydroxyisobutyrate dehydrogenase [Source:HGNC Symbol;Acc:HGNC:4907]
HIST1H1A	-2,90	5,5E-04	histone cluster 1 H1 family member a [Source:HGNC Symbol;Acc:HGNC:4715]
ANXA1	-2,92	2,6E-05	annexin A1 [Source:HGNC Symbol;Acc:HGNC:533]
NR1I2	-2,94	1,9E-04	nuclear receptor subfamily 1 group I member 2 [Source:HGNC Symbol;Acc:HGNC:7968]
BCL2L10	-2,97	4,7E-04	BCL2 like 10 [Source:HGNC Symbol;Acc:HGNC:993]
CTAGE8	-3,02	7,7E-05	CTAGE family member 8 [Source:HGNC Symbol;Acc:HGNC:37294]
PCDH15	-3,02	3,0E-03	protocadherin related 15 [Source:HGNC Symbol;Acc:HGNC:14674]
SERPINB4	-3,21	1,8E-04	serpin family B member 4 [Source:HGNC Symbol;Acc:HGNC:10570]
RIPPLY2	-3,33	1,7E-04	ripply transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:21390]
VEGFC	-3,39	2,8E-05	vascular endothelial growth factor C [Source:HGNC Symbol;Acc:HGNC:12682]
RNF128	-3,39	6,8E-07	ring finger protein 128, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:21153]
SRR	-3,43	5,8E-10	serine racemase [Source:HGNC Symbol;Acc:HGNC:14398]
CRTC2	-3,50	1,2E-07	CREB regulated transcription coactivator 2 [Source:HGNC Symbol;Acc:HGNC:27301]
SIK1	-3,52	1,3E-03	salt inducible kinase 1 [Source:HGNC Symbol;Acc:HGNC:11142]
MTRNR2L1	-3,56	7,0E-04	MT-RNR2-like 1 [Source:HGNC Symbol;Acc:HGNC:37155]
ESRRA	-3,70	2,4E-07	estrogen related receptor alpha [Source:HGNC Symbol;Acc:HGNC:3471]
FRMD1	-3,73	4,2E-04	FERM domain containing 1 [Source:HGNC Symbol;Acc:HGNC:21240]
ARHGAP36	-3,75	2,6E-05	Rho GTPase activating protein 36 [Source:HGNC Symbol;Acc:HGNC:26388]
NKX2-5	-3,80	9,7E-03	NK2 homeobox 5 [Source:HGNC Symbol;Acc:HGNC:2488]
PDCL2	-3,81	1,5E-04	phosducin like 2 [Source:HGNC Symbol;Acc:HGNC:29524]
CXCR5	-3,84	1,4E-04	C-X-C motif chemokine receptor 5 [Source:HGNC Symbol;Acc:HGNC:1060]
ZFP36	-3,84	6,1E-08	ZFP36 ring finger protein [Source:HGNC Symbol;Acc:HGNC:12862]
PRR16	-4,11	1,7E-07	proline rich 16 [Source:HGNC Symbol;Acc:HGNC:29654]
SERPINB3	-4,19	5,3E-04	serpin family B member 3 [Source:HGNC Symbol;Acc:HGNC:10569]
HSPB8	-4,22	6,6E-07	heat shock protein family B (small) member 8 [Source:HGNC

			Symbol;Acc:HGNC:30171]
RHOXF2	-5,01	5,4E-03	Rhox homeobox family member 2 [Source:HGNC Symbol;Acc:HGNC:30011]
LGALS7B	-5,05	5,7E-03	galectin 7B [Source:HGNC Symbol;Acc:HGNC:34447]
ZBED6	-5,24	1,5E-02	zinc finger BED-type containing 6 [Source:HGNC Symbol;Acc:HGNC:33273]
SYNC	-5,86	1,0E-10	syncoilin, intermediate filament protein [Source:HGNC Symbol;Acc:HGNC:28897]
HBD	-5,88	5,4E-03	hemoglobin subunit delta [Source:HGNC Symbol;Acc:HGNC:4829]
ZNF717	-6,54	4,3E-08	zinc finger protein 717 [Source:HGNC Symbol;Acc:HGNC:29448]
NAP1L5	-6,79	9,7E-07	nucleosome assembly protein 1 like 5 [Source:HGNC Symbol;Acc:HGNC:19968]
TSPYL5	-7,91	2,3E-08	TSPY like 5 [Source:HGNC Symbol;Acc:HGNC:29367]

Supplemental experimental procedures

Transgenic overexpression of *Bcl-xL* in hESC

The human *Bcl-xL* gene was cloned into the LV500A-1 lentiviral vector pCDH (System Biosciences), expressed by the activity of the elongation factor-1alpha (EF-1alpha) promoter. The expression of GFP and puromycin resistance gene was driven by the phosphoglycerate kinase I gene promoter (PGK). The packaging plasmid (pCMVDR8.9) and encoding plasmid (VSV.G/ pMD.G) were donated by D. Trono (University of Geneva). Lentiviral particles were generated by transient co-transfection of human embryonic kidney 293T cells. The viral particles were 300-fold concentrated by ultracentrifugation, suspended in PBS containing 10mg/ml protamine sulphate (LeoPharma) and stored at -80^{0}C. Virus titers were determined by measuring the reverse transcriptase activity. A genetically normal subline of VUB03 was used for double transduction, first with 10 mlof1 × 10^6 TU/ml viral particles, and 6–8 h later 30 mlof 1 × 10^6 TU/ml viral particles. GFP-positive cells were visualized using an IX-81 fluorescent microscope (Olympus) and selected by continuously adding 1 mg/ml of puromycin (Sigma-Aldrich) to the culture medium. After two passages during which the puromycin-containing medium was daily refreshed, 80% of the cells were GFP positive.

Fastq files

After RNA sequencing, on average 14.3 x 10^6 ± 6.4 x 10^6 reads were generated per sample. First, the reads were trimmed using cutadapt version 1.11 to remove the "QuantSEQ FWD" adaptor sequence. To assess the quality of the reads, the FastQC algorithm was used on our sequences (Love et al., 2014). A Quality Control tool for High Throughput Sequence Data, website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Mapping Sequences

To obtain the count table for the genes, the reads were mapped against the Genome Reference Consortium Human Build 38 patch release 10 (GRCh38.p10) combined with a general transfer format (GTF) file, both downloaded from the ensembl database(Zerbino et al., 2018). The software used for the mapping was STAR (version 2.5.3)(Dobin et al., 2013).

Count Tables

The RNA- Seq by Expectation Maximization (RSEM)(Li and Dewey, 2011) software (version 1.3.0) was used to produce the count table for each sample. RSEM algorithm was chosen because it is optimized for multi-mapped reads. On the 63967 ensembl's genes, only the 19847 coding genes were considered, and we removed the long non-coding RNAs, which were not useful for our analysis.

RNA-seq analysis

The RNA-seq analysis was performed using the R software (version 3.3.2) with the edgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014) libraries. Only genes with a count per million (cpm) greater than 1 in at least two samples were considered. The raw counts were normalized using the trimmed mean of M values (Robinson and Oshlack, 2010) (TMM) algorithm. For each comparison, a different general linear model was built. Statistical testing was done using the empirical Bayes quasi-likelihood F-test. The normalized counts were then transformed in a $\log_2$ fold-change ($\log_2$FC) table with their associated statistics, p-value and false discovery rate (FDR). In each comparison, genes with a $|\log_2 FC| > 1$ and an FDR < 0.05 were considered as significantly differentially expressed. A $|\log_2 FC| > 1$ means at least two times more or two times less transcript in the mutant group in comparison to the wild-type group.

The data was represented using two different methods; in both cases, all expressed genes were included. First, a heatmap and unsupervised hierarchical clustering was created using the heatmap.2 function from the gplot R library. Second, the data were represented using a multidimensional scaling (MDS) plot of distances between digital gene expression profiles. With the MDS method, the distance between the samples was calculated based on the $\log_2$FC. In the MDS, the Euclidean distances between samples were calculated and were then are regressed against the original distance matrix and the predicted ordination distances for each pair.

Ingenuity pathway analysis

Ingenuity Pathway Analysis (Krämer et al., 2014) (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) was used for the pathway analysis based on the differential gene expression between groups. The data were uploaded with their respective $\log_2$FC, FDR and p-value. IPA predicts the activation state of regulators by correlating literature reported effects with observed gene expression. In order to predict if a pathway is activated or inhibited, it computes the z-score. A z-score>2 means activated, z-score<-2 means inhibited and between means affected. Each z-score is associated with a p-value. To compute this z-score, for each prediction for each gene x_i, a score is associated to it.

Enrichr analysis

The enrichment signatures was done using Enrichr (Chen et al., 2013; Kuleshov et al., 2016). The deregulated genes were used as input ($|\log_2 FC|>1$, FDR<0.05). The outputs were selected from two libraries, 1) ENCODE and ChEA Consensus TFs from ChIP-X and 2) TRANSFAC and JASPAR PWMs. Only the enrichment with p-value<0.05 in both comparisons were taken.

David analysis

The pathway enrichment analysis was done using DAVID 6.8(Huang et al., 2009a, 2009b). The top 1'000 genes with the highest $\log_2$FC in absolute value were selected. In the different categories, only the GO term were taken for account. The cutoff value was at least one of the two comparisons had to have a p-value below 0.05 for each GO-term.

Gene set enrichment analysis

The Gene set enrichment analysis (GSEA) software was downloaded from (http://software.broadinstitute.org/gsea/). The ranking score for each score was computed for each coding gene CPM>1 in at least two samples. The parameters set for each analysis were: enrichment statistic as weighted, number of permutation was 1000, exclude sets larger than 500 and exclude sets smaller than 15. The libraries used from Molecular Signatures Database v6.2 (MSigDB) were: positional gene sets (C1) and curated gene sets (C2). The gene sets were statistically relevant if their FDR was below 0.05. The gene sets were considered as positively enriched if their normalized enriched score (NES) was above 1.4 and negatively enriched if their NES<-1.4 (Subramanian et al., 2005)

Probes, assays and primers used for qPCR.

Gene	Taqman Assay / Sequence
GUSB	Hs99999908_m1
PAX6	Hs00240871_m1
SOX1	Hs01057642_s1
SOX17	Hs00751752_s1
FOXA2	Hs00232764_m1
T	Hs00610080_m1
HAND1	Hs02330376_s1
CHCHD2	Hs00853326_g1
GAPDH	Forward 5'-ATG-GAA-ATC-CCA-TCA-CCA-TCT-T-3' Reverse 5'-CGC-CCC-ACT-TGA-TTT-TGG-3' Probe 6-FAM- CAG-GAG-CGA-GAT-CC-MGB
OCT3A	Forward 5'-GGA-CAC-CTG-GCT-TCG-GAT-TT-3' Reverse 5'-CAT-CAC-CTC-CAC-CAC-CTG-G-3' Probe 6-FAM- GCC-TTC-TCG-CCC-CC-MGB
NANOG	Forward 5'-TGC-AAA-TGT-CTT-CTG-CTG-AGA-TG-3' Reverse 5'-TCC-TGA-ATA-AGC-AGA-TCC-ATG-GA-3' Probe 6-FAM- CAG-AGA-CTG-TCT-CTC-CTC-MGB
UBC	Forward 5'-CGC-AGC-CGG-GAT-TTG-3' Reverse 5'-TCA-AGT-GAC-GAT-CAC-AGC-GA-3' Probe 6-FAM- TCG-CAG-TTC-TTG-TTT-GTG-MGB

Antibodies

Primary antibodies were incubated overnight: PAX6 (Mouse Monoclonal IgG, Abcam, Cat# ab78545), OCT3A (Rabbit Monoclonal IgG, Cell Signalling, Cat# C30A3), SOX17 (Goat Polyclonal IgG, R&D Systems, Cat# AF1924), CHCHD2 (Rabbit Polyclonal, Proteintech #19424-I-AP) T (Goat Polyclonal, R & D Systems #AF2085). Secondary antibodies were incubated for 2-3h: Goat anti-Mouse (H+L) Alexa Fluor 488 (Thermo Fisher Scientific, Cat# A11001), Donkey anti-Goat IgG (H+L) Alexa Fluor 488 (Thermo Fischer Scientific, Cat# A-11055), Goat anri-Rabbit (H+L) Alexa Fluor 488 (Thermo Fischer Scientific, Cat# A-11034), Donkey anti-Rabbit (H+L) Alexa Fluor 546 (Thermo Fisher Scientific, Cat# A10040), Donkey anti-Mouse IgG (H+L) Alexa Fluor 594 (Thermo Fischer Scientific, Cat# R37115). Nuclear staining was performed with Hoechst 33342 (Thermo Fischer Scientific).

References

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523–530.

Kuleshov, M. V, Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-7.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.