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

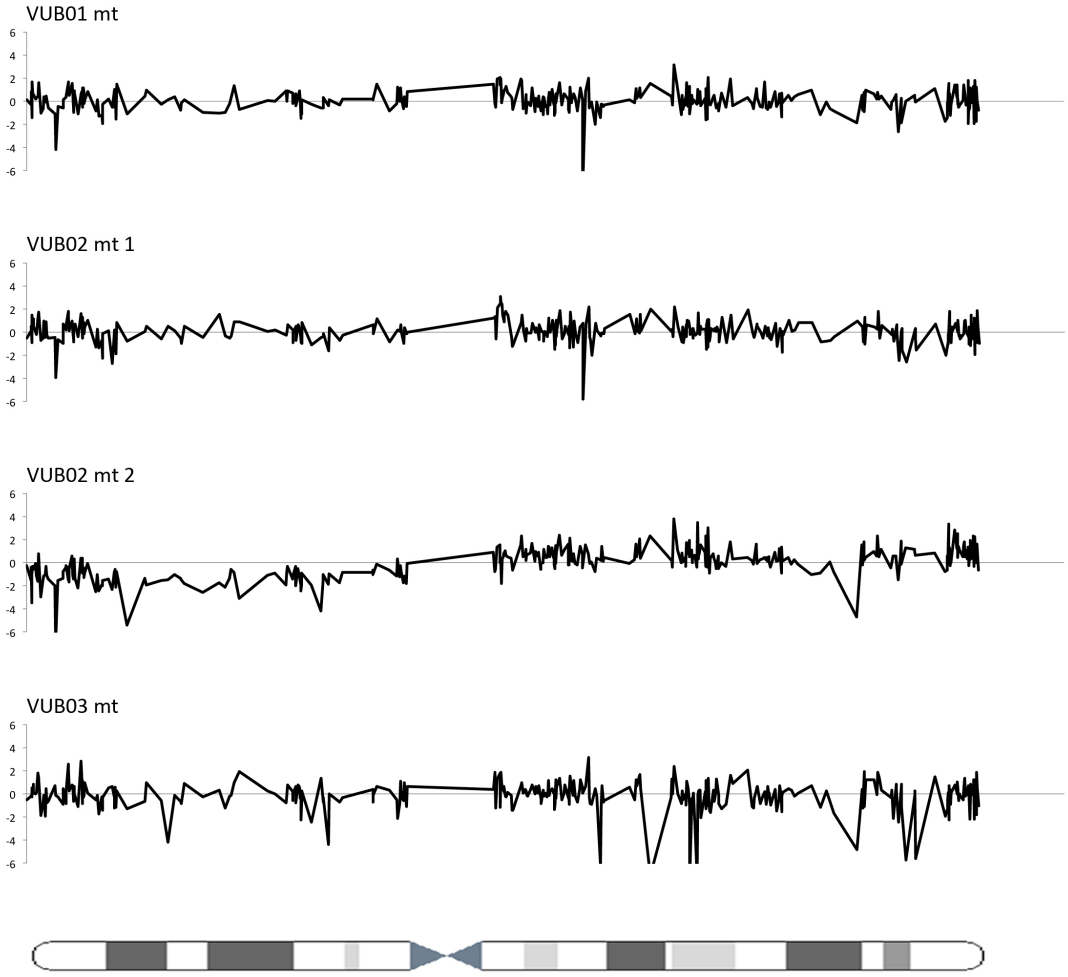

Figure S1 (relates to figure 1 in the main text). Raw gene expression data of all genes expressed on chromosome 20, for the four lines carrying a gain of 20q11.21. Each plot represents the differential gene expression for that line against the whole group of control lines. Data plotted as a Log_2 Fold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

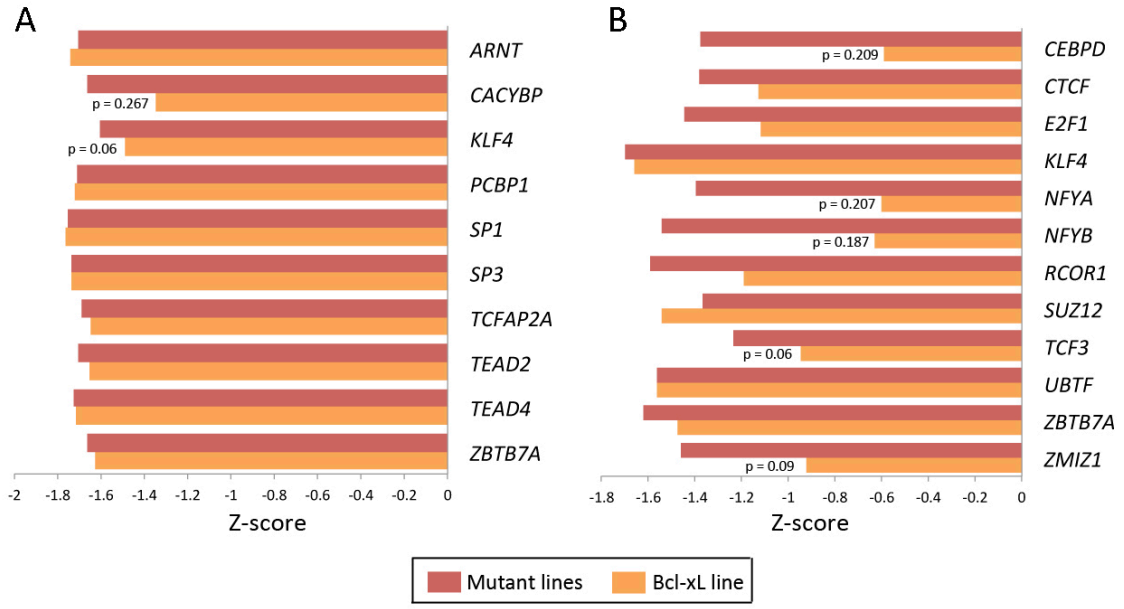

Figure S2 (relates to figure 2 in the main text). Supplementary Figure 2A shows an *in silico* prediction of the differentially activated transcription factors, based on the binding motifs identified at the promoter region of each gene that is differentially expressed between wt vs mt or wt vs VUB03_Bcl-xL (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.

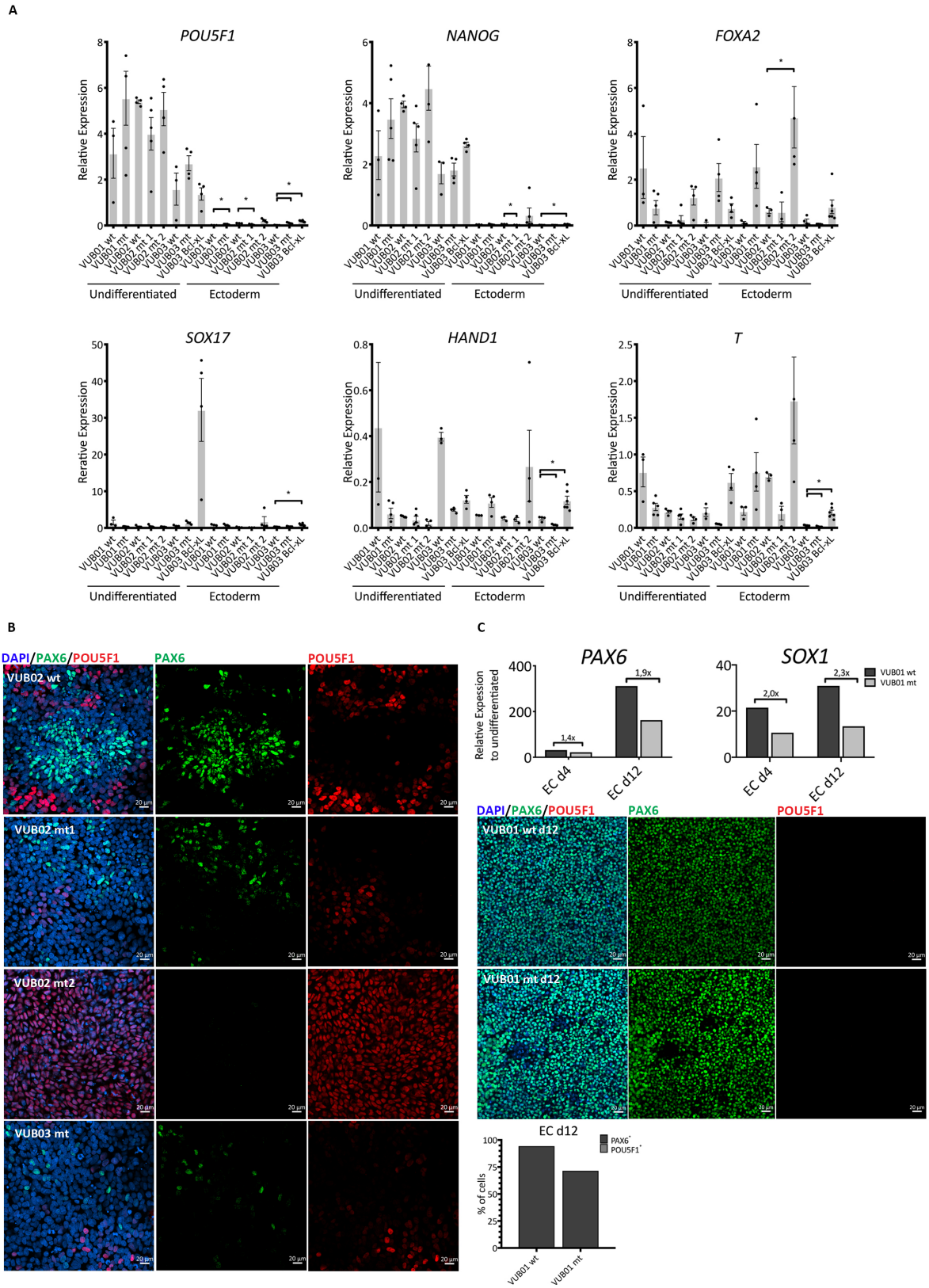

Figure S3 (relates to figure 3 in the main text). Figure S3A shows relative mRNA expression as measured by qPCR for *POU5F1, NANOG, FOXA2, SOX17, HAND1* and *T* after 4 days of neuroectoderm differentiation (n=3 to 5). Data are shown as mean ±SEM, each dot represents an independent differentiation experiment 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).

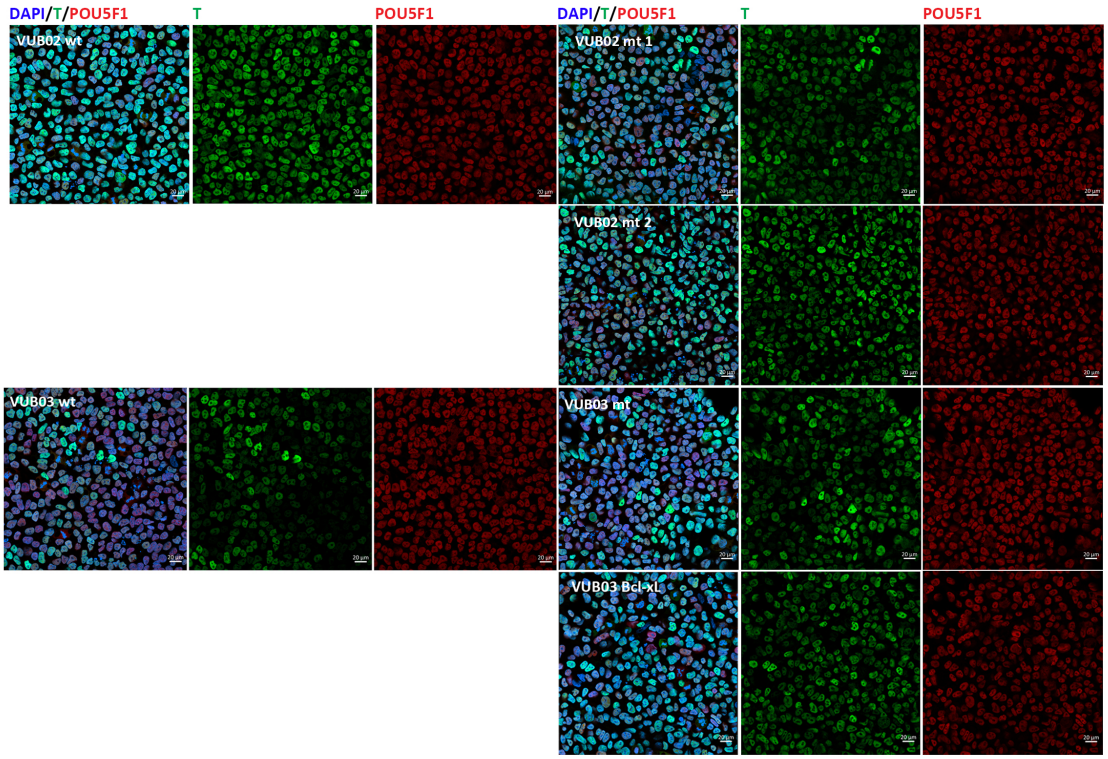

Figure S4, relates to figure 4 in main text. Immunostaining for T (green) and POU5F1 (red) after 24h mesendoderm induction for the lines not shown in figure 4A (VUB02_wt, VUB02_mt1 and VUB02_mt2, VUB03_wt versus VUB03_mt and VUB03_Bcl-xL.

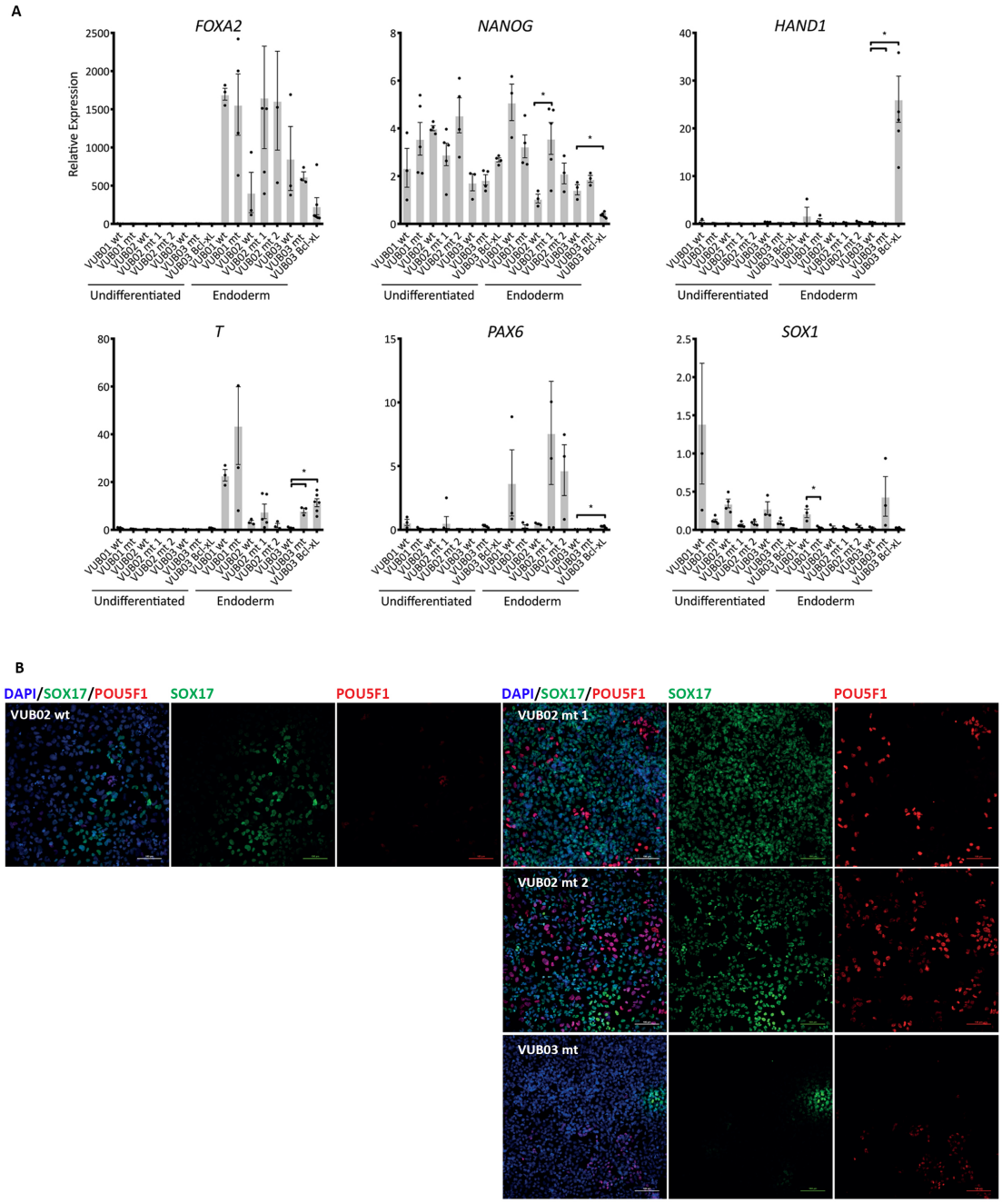

Figure S5 (relates to figure 4 in the main text). Figure S5A shows relative mRNA expression as measured by qPCR for *FOXA2, NANOG, HAND1, T, PAX6* and *SOX1* after definitive endoderm differentiation for all the lines (n=3 to 5). Data are shown as mean ±SEM, each dot represents an independent differentiation 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.