Cell Reports, Volume *14*

Supplemental Information

YAP Induces Human Naive Pluripotency

Han Qin, Miroslav Hejna, Yanxia Liu, Michelle Percharde, Mark Wossidlo, Laure Blouin, Jens Durruthy-Durruthy, Priscilla Wong, Zhongxia Qi, Jingwei Yu, Lei S. Qi, Vittorio Sebastiano, Jun S. Song, and Miguel Ramalho-Santos

YAP Induces Human Naïve Pluripotency

Supplemental Information Inventory

Supplemental Data – Figure S1-4, including legends (related to Figure 1-4 respectively); Table S1; Extended Experimental Procedures (related to Experimental Procedures); Supplemental References.

Data S1 – Genes differentially expressed in human Yin-PSCs and Lin-PSCs compared with parental primed PSCs, related to Figure 3.

Data S2 – Functional category enrichment analysis of genes up-regulated in human Yin-PSCs and Lin-PSCs, related to Figure 3.

Data S3 – Functional category enrichment analysis of genes down-regulated in human Yin-PSCs and Lin-PSCs, related to Figure 3.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: YAP Over-Expression Promotes a Naïve State in Multiple Human PSC lines, Related to Figure 1

(A) YAP expression is significantly higher in trophectoderm cells than ICM cells in early human blastocysts, but this pattern shifts and YAP is up-regulated in epiblast cells of late blastocysts, as assessed by single-cell qRT-PCR. Data are averages of normalized single-cell data (n = 21 for early TE cells, n = 67 for cells of early ICM, n = 23 for late TE cells, n = 63 for cells of late ICM), and error bars represent standard deviation. (B) YAP protein is localized to the nucleus in both the trophectoderm and ICM of human blastocysts, as shown by immunofluorescence. Intensity profile graphs suggest that YAP is diffusely distributed in the nucleus of OCT4+/CDX2- ICM cells (top panels) and has a more heterogeneous nuclear pattern in OCT4-/CDX2+ trophectoderm cells (bottom panels). Scale bar, 20 μm. (C) OCT4 and NANOG expression in H9-YAP cells in N2B27+2iFL, as assessed by Western Blotting. Tubulin was used as loading control. (D-E) H1-YAP (B) and IMR-90 iPSC-YAP (C) in mTeSR+2iFL have a naïve-specific dome-like colony morphology, and show strong positive immunostaining for pluripotency markers SSEA3, SSEA4, TRA-1-60 and TRA-1-81. Black scale bar, 500 μm; White scale bar, 150 μm. (F) H1-YAP in mTeSR+2iFL have a normal male karyotype (46, XY). (G) IMR-90 iPSC-YAP in mTeSR+2iFL have a normal female karyotype (46, XX). (H-I) Pluripotency of H1-YAP (F) and IMR-90 iPSC-YAP (G) in mTeSR+2iFL was confirmed by their ability to form teratomas comprised of tissues derived from all three germ layers. (a) Gut-like epithelium (endoderm). (b) Cartilage (mesoderm). (c) Neural tissue (ectoderm). Scale bar, 200 μm.

Figure S2: Karyotyping of Lin-PSCs, Related to Figure 2

(A) Lentiviral YAP over-expression in human ESCs H9 and IMR-90 iPSCs is validated by Western Blotting. Tubulin was used as loading control. (B) A line of H9 cells in mTeSR+2iFL+LPA with a normal female karyotype (46, XX).

Figure S3: Functional Categories of Genes Differentially Expressed in Yin-PSCs and Lin-PSCs relative to Primed PSCs, Related to Figure 3

(A) Principal Component Analysis (PCA) using all genes confirms the distinct transcriptional profile of Yin-PSCs and Lin-PSCs vs primed PSCs, notably separated along PC1. (B-C) Gene ontology (GO) analysis of genes significantly up-regulated (B) or down-regulated (C) in Yin-PSCs and Lin-PSCs. -log10 P-value is plotted for significantly enriched GO terms. GO terms with a P Value of <0.05 are shown. (D) Gene Set Enrichment Analysis (GSEA) reveals that Yin-PSCs and Lin-PSCs have gene expression profiles concordant with naïve PSCs from other studies. Upper panel is enrichment plot for the gene set of Theunissen et al. down-regulated in naïve by log2 FC < -3.0 (Theunissen et al., 2014). Lower panel is enrichment plot for the gene set of Takahashi et al. down-regulated in naïve by log2 FC < -3.0 (Takashima et al., 2014). (E) IMR-90 iPSC-YAP cells in mTeSR+2iFL express markers specific for naïve pluripotency, as confirmed by qRT-PCR. Values were normalized to *GAPDH* and *UBB*, and then compared to H9 in primed medium (left panel), or H1 in primed medium (right panel). Data are averages of triplicate PCR reactions, and error bars represent standard deviation. (F) Heat map plot of standardized expression of genes with Blastocyst specificity [(expression in Blastocyst+1)/(average expression in 2Cell, 4Cell, 8Cell +1)] > 2, and naive specificity (average expression in naive cells-average expression in primed cells) > 1. Genes are sorted by Blastocyst specificity. Expression is standardized across developmental stages.

Figure S4: H3K9me3 Foci are Lost in Multiple Yin-PSCs, and GSK3 Inhibition Induces Morphological Differentiation of ESCs, Related to Figure 4

(A) Loss of H3K9me3 foci in H1-YAP and IMR-90 iPSC-YAP in mTeSR+2iFL, as assessed by Immunofluorescence. Primed cells were used as control. Scale bar, 20 μm. (B) CHIR99021 induces rapid differentiation of H9 ESCs. PD0325901 and CHIR99021 were added to DF12+bFGF primed medium combined (2i) or separately. Scale bar, 300 μm. (C) LPA suppresses CHIR99021-induced differentiation in H9 ESCs, as shown by qRT-PCR. Values were normalized to *GAPDH* and *UBB*, and then compared to H9 in mTeSR+1iFL (PD0325901, Forskolin and hLIF). Data are averages of triplicate

PCR reactions, and error bars represent standard deviation. (D) Loss of YAP expression in YAP-/- (ΔYAP) cells is validated by Western Blotting. Tubulin was used as loading control. (E) Sequencing results of the YAP locus in ΔYAP cells. Mutations are highlighted in red.

EXTENDED EXPERIMENTAL PROCEDURES

Generation of human iPSCs

Human primary male newborn foreskin (BJ) and female lung (IMR-90) fibroblasts were obtained from ATCC and cultured in DMEM with 10% FBS, 1x glutamine, 1x non-essential amino acids, 1x sodium pyruvate, 1x penicillin/streptomycin, and 0.06 mM β-mercaptoethanol (fibroblast medium). Cells were seeded at 100,000 cells per well on a 6-well plate the day before infection. Cells were infected with retroviruses (Harvard Gene Therapy Initiative) leading to the over-expression of OCT4, SOX2, KLF4 and c-MYC. Cells remained in the presence of virus for 48 hours, and fibroblast medium was added on the day after virus addition. 48 hours after infection, virus was removed and cells were cultured in DF12+bFGF primed medium.

Obtaining human embryos

Supernumerary human blastocysts from successful IVF cycles, subsequently donated for non-stem-cell research, were obtained with written informed consent from the Stanford University RENEW Biobank. De-identification was performed according to the Stanford University Institutional Review Board-approved protocol #10466 entitled 'The RENEW Biobank' and the molecular analysis of the embryos was in compliance with institutional regulations. Approximately 25% of the embryos were from couples that used donor gametes and the most common cause of infertility was unexplained at 35% of couples. No protected health information was associated with any of the embryos. The embryos were cultured at 37 °C with 6% CO2, 5% O2 and 89% N2 under standard human embryo culture conditions in accordance with current clinical IVF practice. Embryos used in this study were DPF 5–6.

Immunostaining and Western Blotting

For AP staining, cells were fixed directly in tissue culture plates using 4% paraformaldehyde, washed with PBS, and stained for AP using VECTOR Red Alkaline Phosphatase (AP) Substrate Kit (Vector Laboratories) according to the manufacturer's instructions. For Immunofluorescence, cells and human blastocysts were fixed in 4% paraformaldehyde and permeabilized with 0.1% Triton X-100. Cells were then stained with primary antibodies against OCT4 (sc-8628, Santa Cruz, for blastocysts), CDX2 (ab15258, Abcam), OCT4 (ab19857, Abcam), NANOG (AF1997, R&D), SSEA-3 (MAB4303, Millipore), SSEA-4 (MAB4304, Millipore), TRA-1-60 (FCMAB115F, Millipore), Tra1-81 (MAB4381, Millipore), YAP (14074, Cell Signaling) and H3K9me3 (39161, Active Motif). Respective secondary antibodies conjugated to Alexa Fluor 488, Alexa Fluor 568 and Alexa Fluor 647 (Invitrogen) were used at 1:500. For Western Blotting, primary antibodies against H3K9me3 (39161, Active Motif), YAP (14074, Cell Signal), Unphosphorylated (Active) β–Catenin (8814, Cell Signal) and alpha-Tubulin (ab18251, Abcam) were used. Respective secondary antibodies conjugated with HRP (Abcam) were used according to standard protocols.

Cytogenetic Analysis

Human PSCs were treated with 10 ng/mL of Colcemid (Invitrogen) overnight at 37 °C. Cells were harvested and G-banded according to a standard cytogenetic protocol (Rooney, 2001). Metaphase cells were analyzed under the microscope and karyotyped according to an International System for Human Cytogenetic Nomenclature (Shaffer et al., 2013) using CytoVision system (Applied Imaging). A minimum of 20 cells were analyzed per sample.

Teratoma induction

Human PSCs were grown to 70–80% confluency. One entire six-well plate-worth of cells was used to inject one immunocompromised SCID/Beige mouse subcutaneously into two sites near the hind flanks. Cells were pelleted and resuspended in 140 µl of DMEM/F12 and, immediately before injection, mixed with 60 µl of Matrigel (BD Biosciences) for a total volume of 200 µl. The cell/Matrigel

mix (100 µl) was injected into each site. Tumours developed after 6–9 weeks and were processed for histological analysis.

Quantitative Real-Time RT-PCR

For analyses of bulk cell populations RNA was isolated using the RNeasy Mini RNA Isolation kit (Qiagen) and reverse-transcribed using the High-Capacity cDNA Reverse Transcription kit (Applied BioSystems). The cDNA reaction was diluted 1:5 in TE (10mM Tris-Cl/1mM EDTA, pH 7.6) and used in Sybr Green real-time PCR reactions (BioRad or Applied BioSystems). PCR primers were designed to amplify 100-400 bp fragments spanning exons. Reactions were run in triplicates on a 7900HT machine (Applied BioSystems) according to the manufacturer's instructions. Only samples with single and matching end-point melting curve peaks were used for subsequent analysis. Cycle threshold values were imported into the REST software for fold-change calculations. Values were normalized to *GAPDH* and *UBB*, and then compared to control cells. Data are from triplicate PCR reactions, and error bars represent standard deviation. Primer sequences are listed in Table S1.

For single cell analyses of human blastocysts we used the C1 Single-Cell Auto Prep System (Fluidigm, Inc.) for cell capture, pre-amplification and subsequent qRT-PCR according to the manufacturer's instructions (protocol # PN 100-4904). Determination of limit of detection (LOD) value, quality assessment and normalization of single-cell expression values were performed as previously described (Durruthy-Durruthy et al., 2015).

Microarray Analysis

Total RNA was extracted using RNeasy Mini RNA Isolation kit with DNase1 treatment (Qiagen), and transcriptional profling was carried out using HumanHT-12 v4.0 Expression BeadChip arrays (Illumina) according to the manufacturer's protocols at the UCLA Neuroscience Genomics Core. Gene expression data were quantile-normalized and log2-transformed, then differentially expressed genes were identified in R using the Bioconductor (www.bioconductor.org) package linear models for microarray data (limma). Differentially expressed genes were identified using thresholds of $P < 0.05$

and log2 Fold Change (FC) >0.7 and are listed in Data S1. 3D PCA plots were generated using the R package "rgl". Microarray data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO). The accession number for the microarray data reported in this paper is GEO:GSE69200. Gene Ontology (GO) analysis was conducted using DAVID (Huang et al., 2009), and Gene Set Enrichment Analysis was performed using gene sets derived from published data (Takashima et al., 2014; Theunissen et al., 2014) as described (Subramanian et al., 2005).

Weighted Gene Co-Expression Network Analysis and Information Theoretic Distance

We constructed a weighted gene co-expression network as previously described (Zhang and Horvath, 2005) using 2000 most variable genes across the developmental stages, as measured by single-cell RNA-seq (Yan et al., 2013). The weight between two genes was computed as $a_{ij} = \max(0, cor(i,j))$, where $cor(i,j)$ denotes the Pearson correlation between genes i and j. In order to reduce the effect of uneven number of replicates among the different cell types, we used a boostrap sampling method: correlation between two genes was evaluated using 6 replicates, and when more than 6 samples are available, we bootstrapped sampled groups of 6 distinct samples and averaged the correlations. To further remove potential platform-dependent batch effects, we quantile normalized the correlation data to a reference Beta distribution Beta(4,4). Following (Zhang and Horvath, 2005), we raised a_{ij} to the power $\beta \in [4,6]$, chosen so that the resulting network is approximately scale invariant, and performed hierarchical clustering of genes using the topological overlap matrix. Cell types with fewer than five replicates (Oocyte, Zygote) were excluded from the analysis, as the Pearson correlation cannot be reliably computed for these small sample sizes. The resulting gene co-expression network was used to partition the genes into 5 modules. We computed the distance between two cell types' respective partitions of genes into correlation modules by using Variation of Information, an information-based distance that measures the amount of information lost and gained between two clustering results (Meilă, 2007). For each pair of cell types with partitions C_1 and C_2, respectively, we computed the joint probability $P(k_1, k_2) = |k_1 \cap k_2|/N$ that a randomly selected gene falls into the modules $k_1 \in C_1$ and

$k_2 \in C_2$, where N denotes the total number of genes in the networks. From the joint probability distribution $P(k_1, k_2)$, we computed the Variation of Information defined as $VI(C_1, C_2) = H(C_1|C_2) + H(C_2|C_1) = -\sum_{k_1 \in C_1, k_2 \in C_2} P(k_1, k_2)\, log\left(\frac{P(k_1,k_2)^2}{P(k_1)P(k_2)}\right)$, where $H(C_i|C_j)$ is the conditional entropy of partition C_i given C_j, and $P(k_1)$ and $P(k_2)$ are the marginal probability distributions derived from $P(k_1, k_2)$. Intuitively, Variation of Information measures how much information is lost and gained when changing between clustering results C_1 and C_2 and provides a notion of distance between two partitions.

CRISPR/Cas9-Mediated Generation of H9 YAP-/- Cells

Guide RNAs sgYAP-1 (GATCAGACAACAACATGGC) and sgYAP-2 (GACGTTCATCTGGGACAGCA) were cloned into a single Cas9/sgRNA-expressing vector pX330 (42230, Addgene) as described (Qi et al., 2013). sgGFP (GTCGACAGGTAATGGTTGTC) was used as control. Primed H9 ESCs cells were trypsinized to single cells and electroporated with px330-sgLUC, pX330-sgYAP-1 or pX330-sgYAP-2 separately using Human Stem Cells Nucleofector kit 2 (VPH-5022, Amaxa) with program B-016 according to the manufacturer's protocols. After electroporation, cells were plated onto MEFs in 6 cm dishes in primed culture medium in the presence of the ROCK inhibitor Y-27632. 10 days later, single colonies were picked and expanded. Immunofluorescence for YAP (14074, Cell Signal) was performed to screen clones with YAP expression loss prior to genomic DNA preparation using the Puregene Cell Kit (Qiagen). Genomic DNA was PCR amplified with High-Fidelity 2X PCR Master Mix (NEB) using two primers flanking the sgRNA target region (forward primer sequence: GCTTTGTGTAAGAACATGTTATTCAGGAGCC; reverse primer sequence: CTTCAGAACCAAATCTTAAAGAACCACTGGC). PCR products were cloned into T-vector (Promega), and sequenced. YAP-/- clones ΔYAP-1 and ΔYAP-2 are generated from sgYAP-1; ΔYAP-3 and ΔYAP-4 are generated from sgYAP-2.

SUPPLEMENTAL REFERENCES

Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E.J., Davila, J., Mall, M., Wong, W.H., Wysocka, J., et al. (2015). The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. Nat. Genet.

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

Meilă, M. (2007). Comparing clusterings—an information based distance. Journal of Multivariate Analysis *98*, 873–895.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Rooney, D.E. (2001). Human Cytogenetics: constitutional analysis, a practical approach (Oxford University Press).

Shaffer, L.G., McGowan-Jordan, J., and Schmid, M. (2013). ISCN 2013: An International System for Human Cytogenetic Nomenclature (2013).

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.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.

Table S1. Sequences for qRT-PCR Primers Used in This Study, Related to Experimental Procedures

Gene name	Forward primer sequence	Reverse primer sequence
GAPDH	CAATGACCCCTTCATTGACC	GACAAGCTTCCCGTTCTCAG
UBB	TTGTTGGGTGAGCTTGTTTG	GTCTTGCCGGTAAGGGTTTT
OCT4	TGTACTCCTCGGTCCCTTTC	TCCAGGTTTTCTTTCCCTAGC
NANOG	CAGTCTGGACACTGGCTGAA	CTCGCTGATTAGGCTCCAAC
GP130	GGAGTGAAGAAGCAAGTGGGA	AGGCAATGTCTTCCACACGA
TBX3	CGGACATACTTGTTCCCCGA	GCAGGGTGAGCTGTTTTCTTTT
SOCS3	ATTCGGGACCAGCCCCC	AAACTTGCTGTGGGTGACCA
TFCP2L1	CTCAGGTGCTGACTTGCTGA	ATGGCGTGGTACACAGACAG
STELLA	TCTCCACAAATGCTCACCGA	TCTTCTTTCATGCGTACGAACTCC
HERVH-Gag	ACGCTTTACAGCCCTAGACC	GTCGGGAGCAGATTGGGTAA
HERVH-Pol	CGCCCTTCTTCCCAATCCAA	GCCAAGGAGGGAGTAGAGGT
CDX2	CGGCAGCCAAGTGAAAACC	CTCCGGATGGTGATGTAGCG
T	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
SOX17	GGCGCAGCAGAATCCAGA	CCACGACTTGCCCAGCAT
SOX1	ATGCACCGCTACGACATGG	CTCATGTAGCCCTGCGAGTTG
SOX9	CCCCAACAGATCGCCTACAG	GAGTTCTGGTCGGTGTAGTC
MYC	AATGAAAAGGCCCCCAAGGTAGTTATCC	GTCGTTTCCGCAACAAGTCCTCTTC
AXIN2	CTGGCTCCAGAAGATCACAAAG	CATCCTCCCAGATCTCCTCAAA
EPHB3	GTCATCGCTATCGTCTGCCT	AATTCCCCAGCTCCGATCA
NKD1	ACTTTCGGCTGGAAGTGGC	GTGACCTTGCCGTTGTTGTC
LEF1	CCCGTGAAGAGCAGGCTAAA	TTCTTGGACCTGTACCTGATGC
YAP	GGCTAGACCCAAGGCTTGAC	GGCTGTTTCACTGGAGCACT

Figure S1

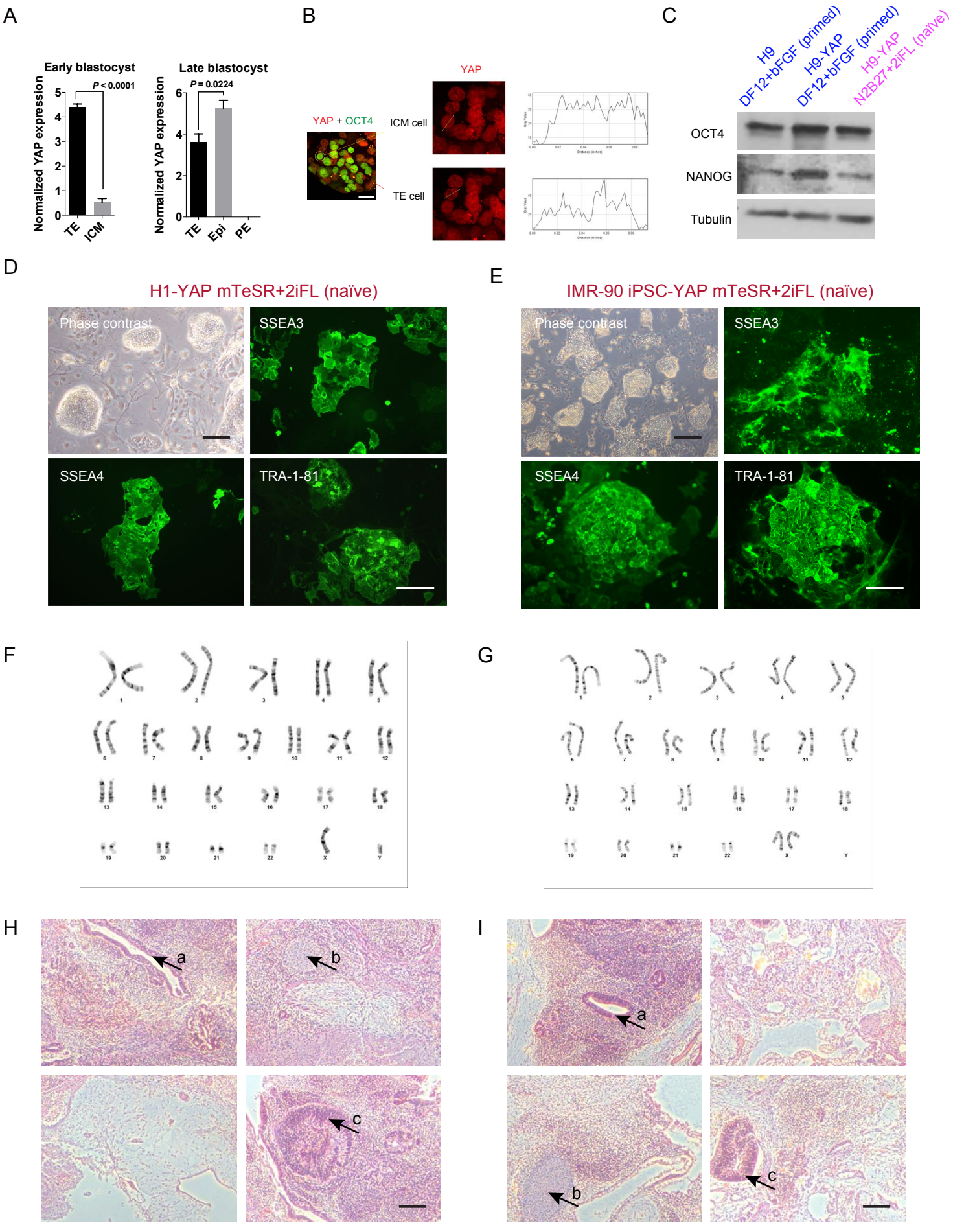
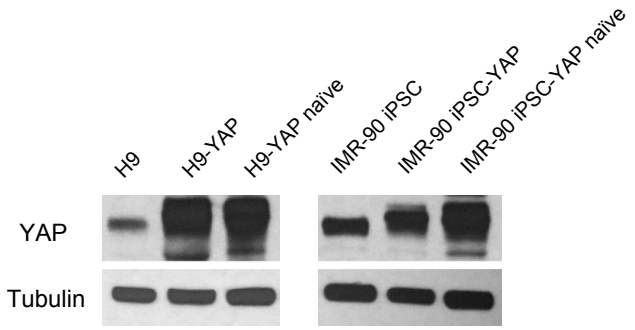

A
Early blastocyst
P < 0.0001
Normalized YAP expression
TE
ICM
Late blastocyst
P = 0.0224
Normalized YAP expression
TE
Epi
PE
B
YAP
YAP + OCT4
ICM cell
TE cell
C
H9
DF12+bFGF (primed)
H9-YAP
DF12+bFGF (primed)
H9-YAP
N2B27+2iFL (naïve)
OCT4
NANOG
Tubulin
D
H1-YAP mTeSR+2iFL (naïve)
Phase contrast
SSEA3
SSEA4
TRA-1-81
E
IMR-90 iPSC-YAP mTeSR+2iFL (naïve)
Phase contrast
SSEA3
SSEA4
TRA-1-81
F
G
H
a
b
c
I
a
b
c

Figure S2

A


H9
H9-YAP
H9-YAP naïve
IMR-90 iPSC
IMR-90 iPSC-YAP
IMR-90 iPSC-YAP naïve
YAP
Tubulin

B

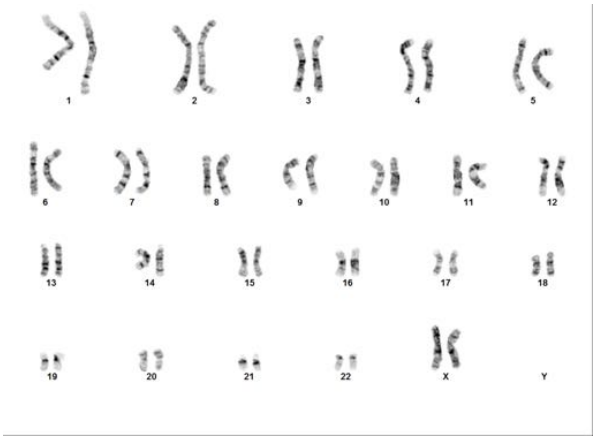
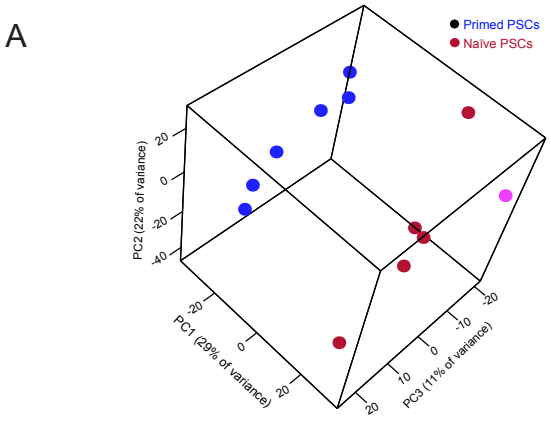
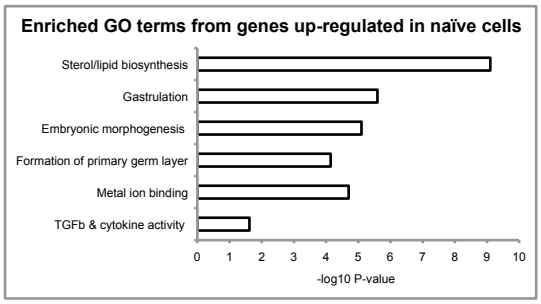


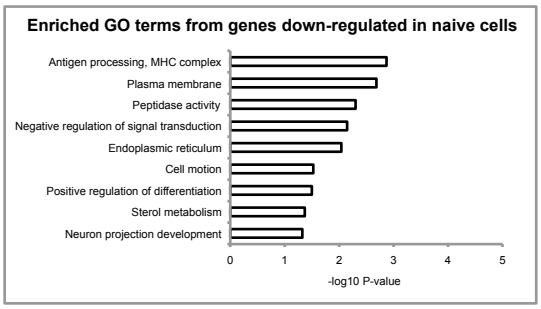
Figure S3



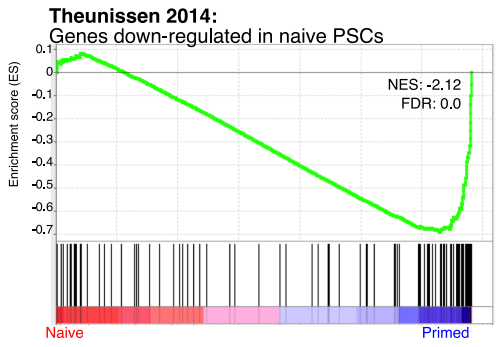
B

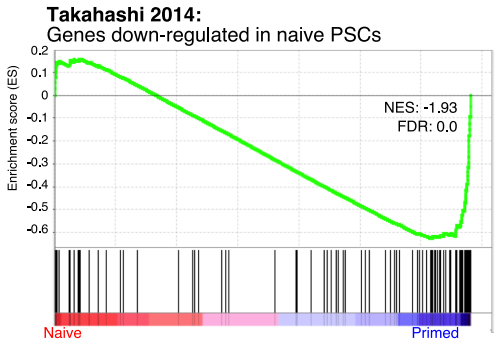


C

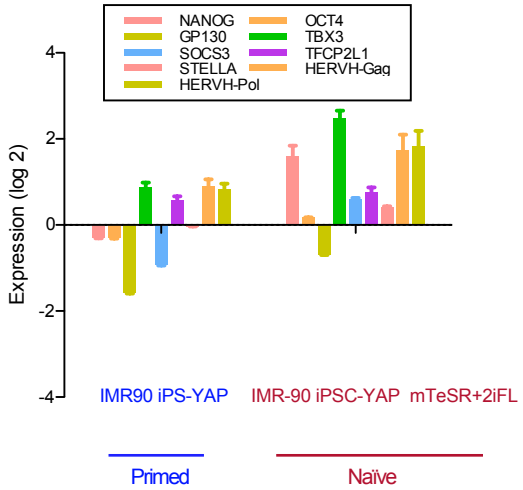


D





E



F

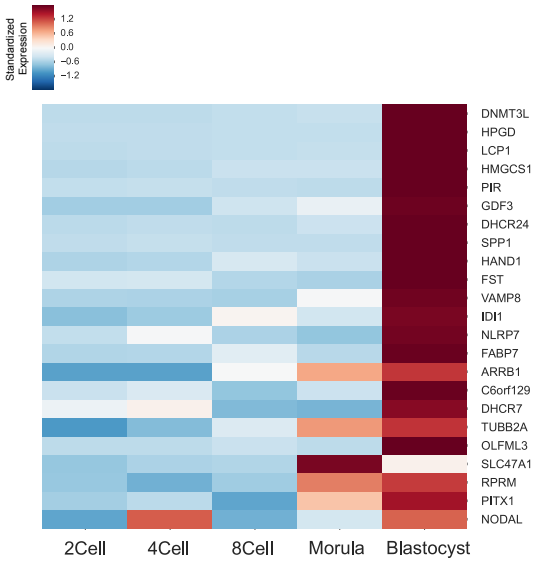
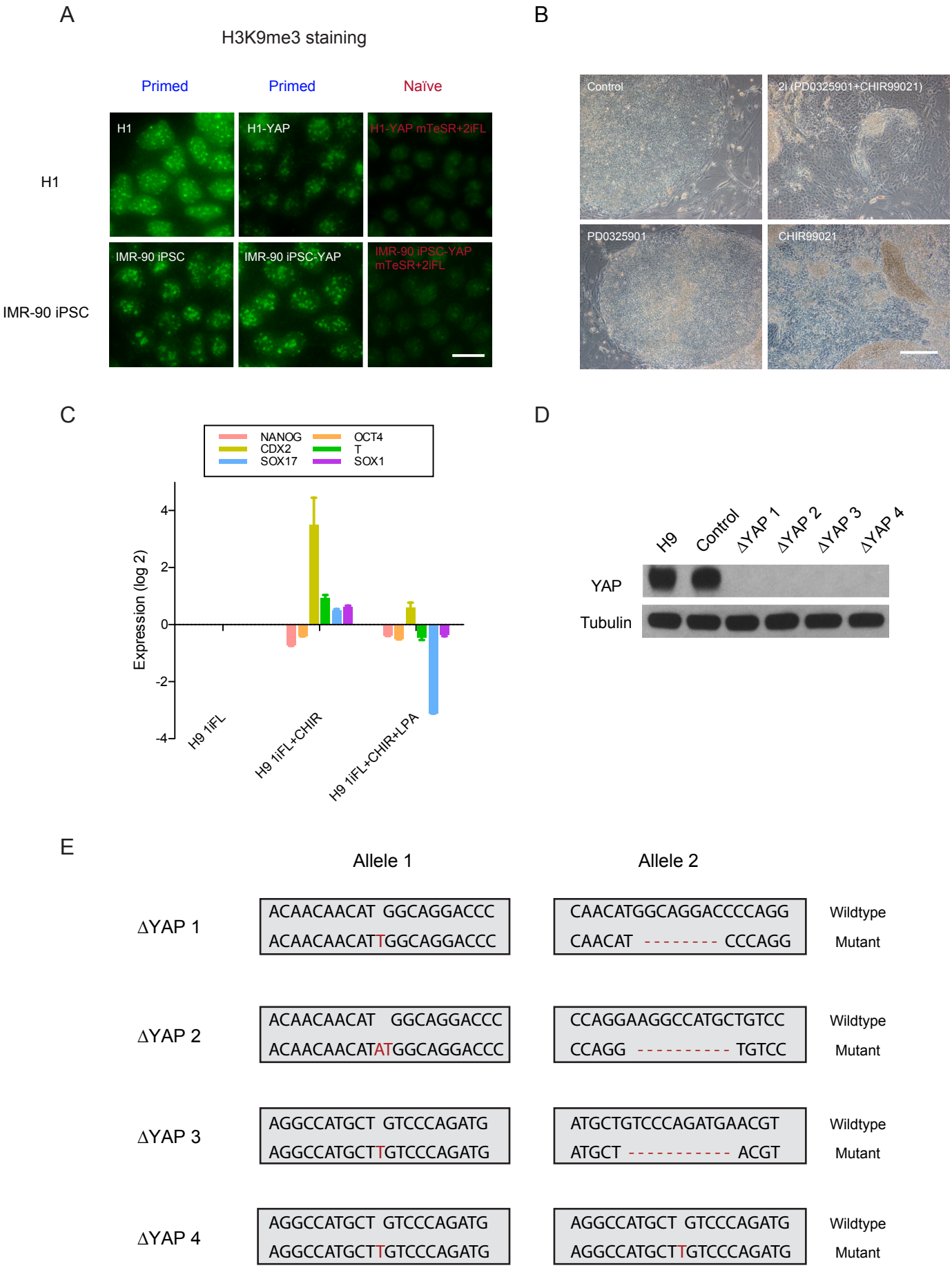


Figure S4


A
H3K9me3 staining
Primed
Primed
Naïve
H1
IMR-90 iPSC
H1
H1-YAP
H1-YAP mTeSR+2iFL
IMR-90 iPSC
IMR-90 iPSC-YAP
IMR-90 iPSC-YAP mTeSR+2iFL
B
Control
2i (PD0325901+CHIR99021)
PD0325901
CHIR99021
C
NANOG
CDX2
SOX17
OCT4
T
SOX1
Expression (log 2)
4
2
0
-2
-4
H9 1iFL
H9 1iFL+CHIR
H9 1iFL+CHIR+LPA
D
H9
Control
ΔYAP 1
ΔYAP 2
ΔYAP 3
ΔYAP 4
YAP
Tubulin
E
Allele 1
Allele 2
ΔYAP 1
ACAACAACAT GGCAGGACCC
ACAACAACATTGGCAGGACCC
CAACATGGCAGGACCCCAGG
CAACAT -------- CCCAGG
Wildtype
Mutant
ΔYAP 2
ACAACAACAT GGCAGGACCC
ACAACAACATATGGCAGGACCC
CCAGGAAGGCCATGCTGTCC
CCAGG ---------- TGTCC
Wildtype
Mutant
ΔYAP 3
AGGCCATGCT GTCCCAGATG
AGGCCATGCTTGTCCCAGATG
ATGCTGTCCCAGATGAACGT
ATGCT ---------- ACGT
Wildtype
Mutant
ΔYAP 4
AGGCCATGCT GTCCCAGATG
AGGCCATGCTTGTCCCAGATG
AGGCCATGCT GTCCCAGATG
AGGCCATGCTTGTCCCAGATG
Wildtype
Mutant