Stem Cell Reports, Volume *14*

Supplemental Information

Recurrent Genetic Abnormalities in Human Pluripotent Stem Cells: Definition and Routine Detection in Culture Supernatant by Targeted Droplet Digital PCR

Said Assou, Nicolas Girault, Mathilde Plinet, Julien Bouckenheimer, Caroline Sansac, Marion Combe, Joffrey Mianné, Chloé Bourguignon, Mathieu Fieldes, Engi Ahmed, Thérèse Commes, Anthony Boureux, Jean-Marc Lemaître, and John De Vos

Figure S1

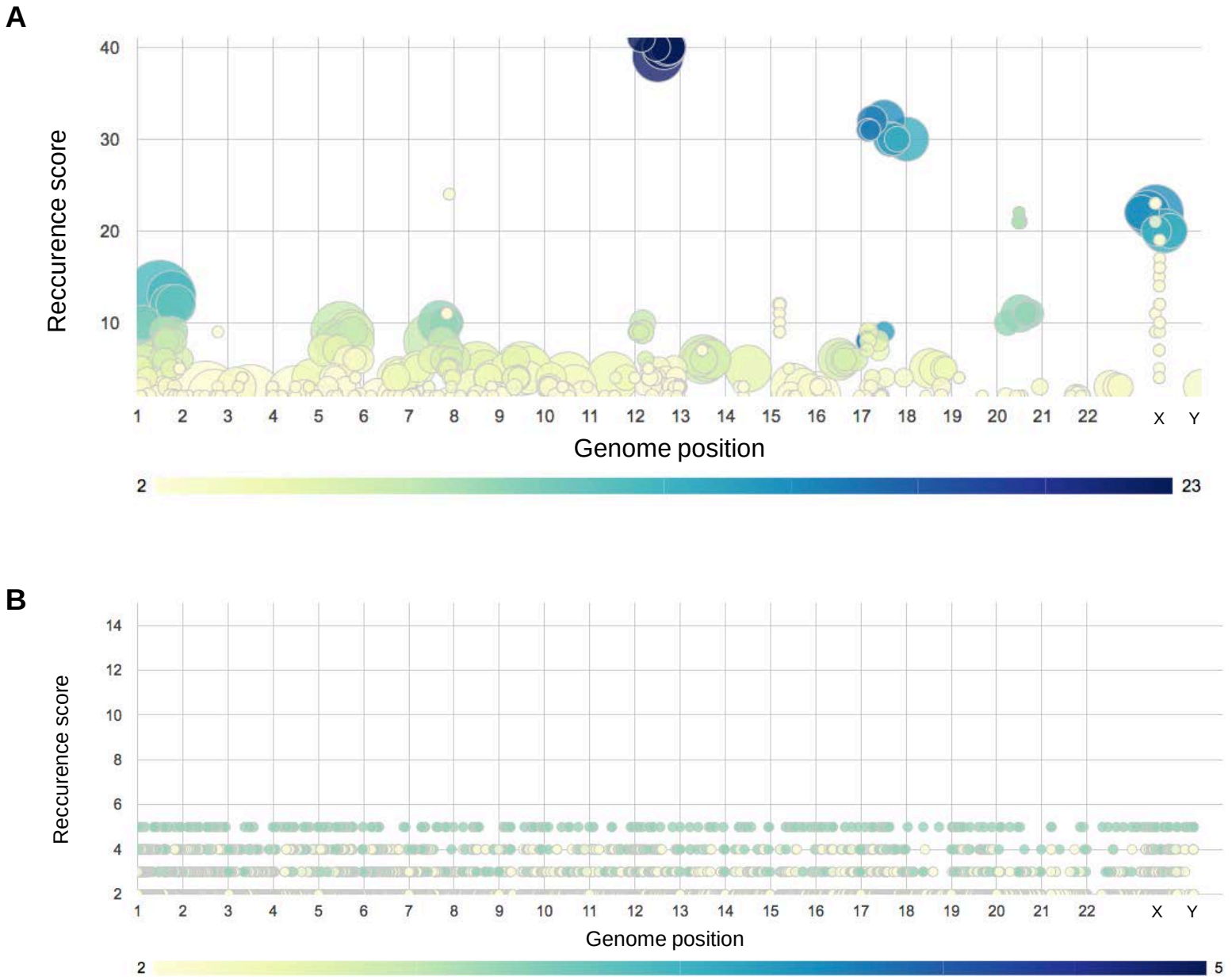

A
Reccurence score
40
30
20
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Genome position
2
23
B
Reccurence score
14
12
10
8
6
4
2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Genome position
2
5

Figure S2

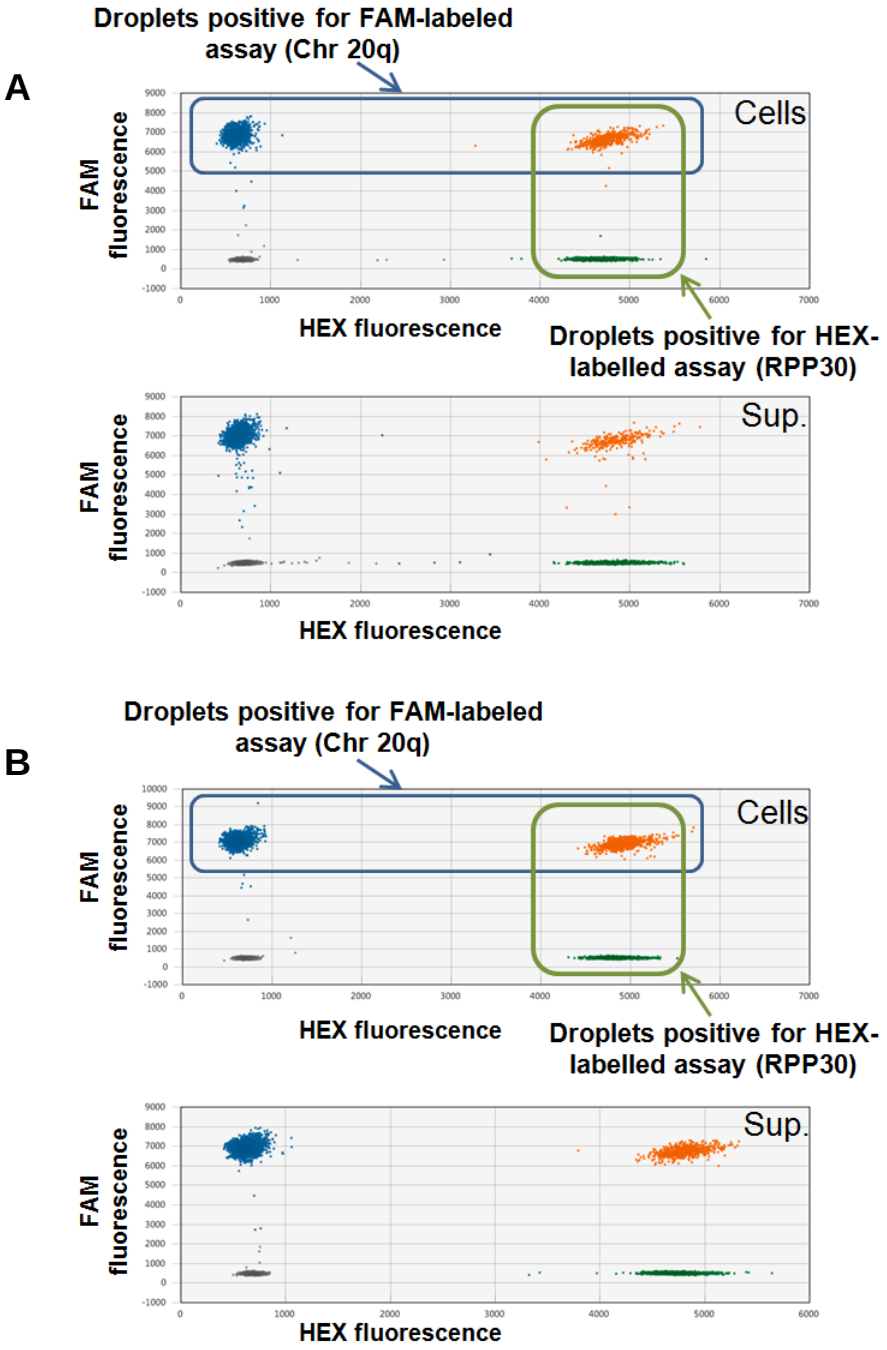

A
Droplets positive for FAM-labeled assay (Chr 20q)
Cells
FAM fluorescence
HEX fluorescence
Droplets positive for HEX-labelled assay (RPP30)
Sup.
FAM fluorescence
HEX fluorescence
B
Droplets positive for FAM-labeled assay (Chr 20q)
Cells
FAM fluorescence
HEX fluorescence
Droplets positive for HEX-labelled assay (RPP30)
Sup.
FAM fluorescence
HEX fluorescence

Figure S3

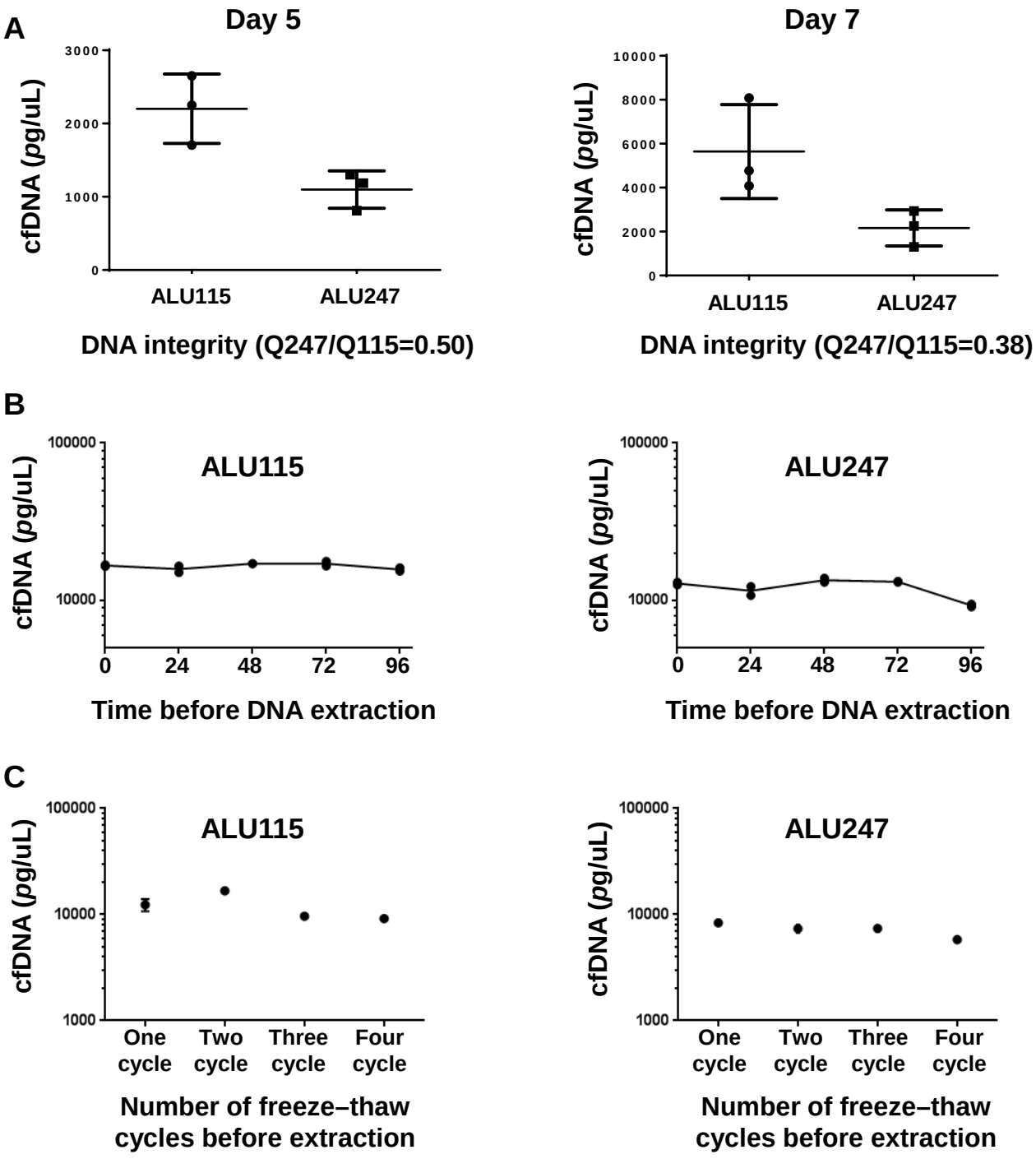
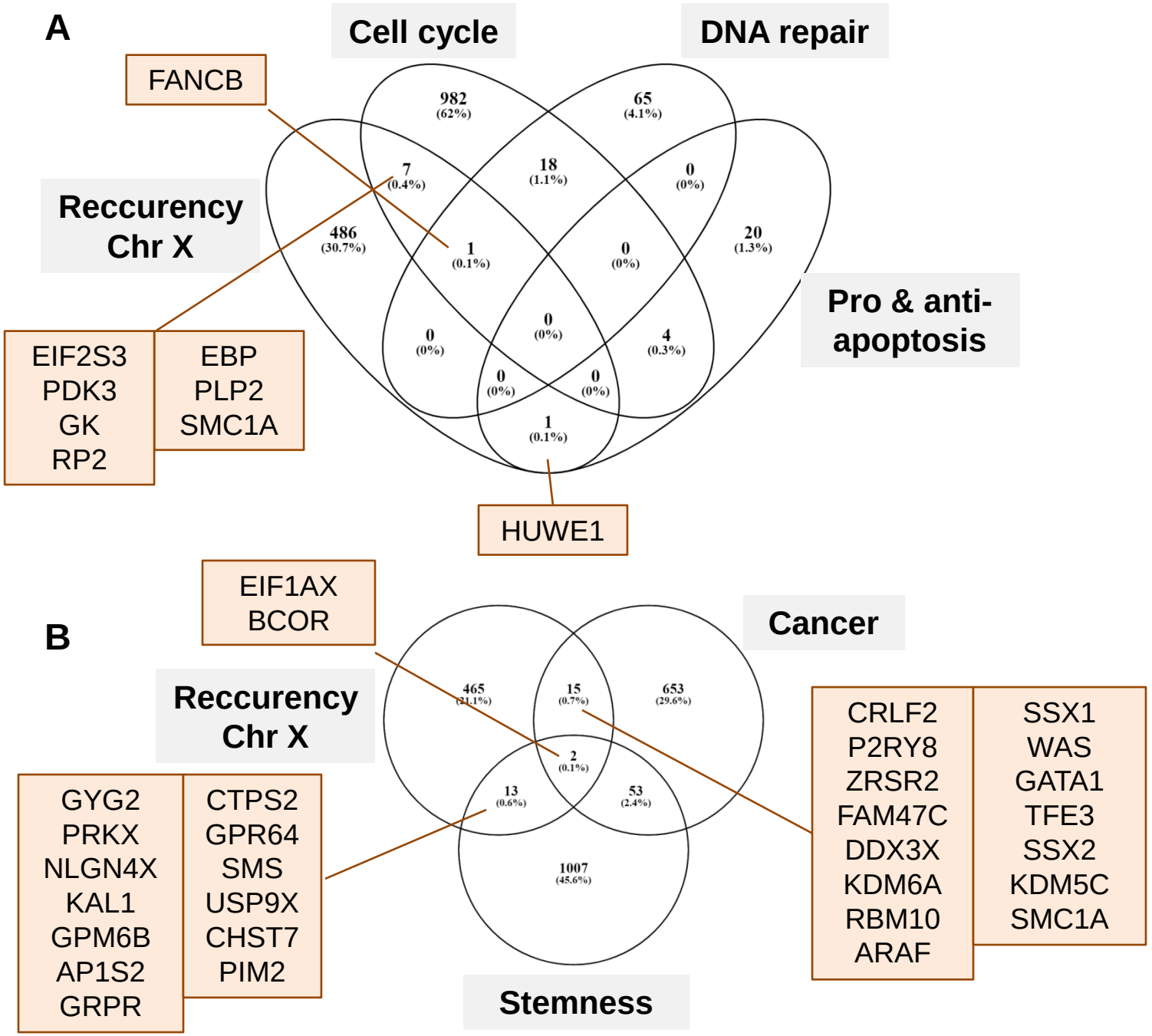

A
Day 5
cfDNA (pg/uL)
3000
2000
1000
0
ALU115
ALU247
DNA integrity (Q247/Q115=0.50)
Day 7
cfDNA (pg/uL)
10000
8000
6000
4000
2000
0
ALU115
ALU247
DNA integrity (Q247/Q115=0.38)
B
ALU115
cfDNA (pg/uL)
100000
10000
0
24
48
72
96
Time before DNA extraction
ALU247
cfDNA (pg/uL)
100000
10000
0
24
48
72
96
Time before DNA extraction
C
ALU115
cfDNA (pg/uL)
100000
10000
1000
One cycle
Two cycle
Three cycle
Four cycle
Number of freeze–thaw cycles before extraction
ALU247
cfDNA (pg/uL)
100000
10000
1000
One cycle
Two cycle
Three cycle
Four cycle
Number of freeze–thaw cycles before extraction

Figure S4


A
Cell cycle
DNA repair
FANCB
Reccurency Chr X
982 (62%)
65 (4.1%)
7 (0.4%)
18 (1.1%)
0 (0%)
486 (30.7%)
1 (0.1%)
0 (0%)
20 (1.3%)
0 (0%)
0 (0%)
4 (0.3%)
0 (0%)
0 (0%)
1 (0.1%)
Pro & anti-apoptosis
EIF2S3
PDK3
GK
RP2
EBP
PLP2
SMC1A
HUWE1
B
EIF1AX
BCOR
Cancer
Reccurency Chr X
465 (21.1%)
15 (0.7%)
653 (29.6%)
2 (0.1%)
13 (0.6%)
53 (2.4%)
1007 (45.6%)
CRLF2
P2RY8
ZRSR2
FAM47C
DDX3X
KDM6A
RBM10
ARAF
SSX1
WAS
GATA1
TFE3
SSX2
KDM5C
SMC1A
GYG2
PRKX
NLGN4X
KAL1
GPM6B
AP1S2
GRPR
CTPS2
GPR64
SMS
USP9X
CHST7
PIM2
Stemness

Supplemental figure legends

Figure S1

A. Bubble plot showing recurrent DNA abnormalities >10bp in hPSC lines. Each bubble represents a region where several DNA abnormalities from several publications are found. The bubble size is proportional to the region length and the color indicates the number of publications that reported abnormalities in that region. The horizontal axis corresponds to the genome position and the vertical axis corresponds to the recurrence score (see Material and Methods).

B. Same as in A, but for abnormalities and variants ≤10bp.

Figure S2

A. 2D plot showing a typical result obtained from quantifying chromosome 20q CNV in a euploid hPSC line (UHOMi001-A) using DNA from cells and supernatant (Sup.), as indicated. Blue: droplets positive for the target CNV (Chr20q); green: droplets positive for the reference gene (*RPP30*); grey: negative droplets (containing no target or reference genes); orange: droplets positive for both the reference and target genes.

B. FACS-like plot showing a typical result obtained from quantifying chromosome 20q CNV in an aneuploid hPSC line (RSP4) using DNA from cells and supernatant (Sup.), as indicated. Same color code as in A.

Figure S3

A. Concentration of DNA in supernatant samples and effects of various pre-analytical conditions. Quantification of supernatant-DNA by quantitative PCR with two sets of ALU primers (115 and 247 bp) that amplify DNA fragments of different lengths in supernatant samples collected from three hPSC lines at day 5 and day 7. The ALU 115 and ALU 247 values are significantly different ($p<0.05$). The Q247/Q115 ratio indicates the DNA integrity value (Q115 corresponds to the DNA concentration obtained using the ALU 115 primers and Q247 to the concentration obtained with the ALU 247 primers). The mean

Q247/Q115 ratio in supernatant samples collected at day 5 and day 7 was 0.50 and 0.38 respectively, suggesting that the DNA released in the supernatant originates mostly from apoptotic rather than necrotic cells.

B. Comparison of supernatant-DNA amount when supernatant was stored at room temperature for different times (24h, 48h, 72h and 96h) before extraction. DNA concentrations (pg/uL) were determined using the ALU115 and ALU247 primers. DNA concentration was not affected by keeping supernatant at room temperature.

C. Comparison of supernatant-DNA amount according to the number of freeze–thaw cycles before extraction. DNA concentration slightly decreased after the supernatants underwent 3-4 freeze–thaw cycles before extraction but the DNA quality was not affected.

Figure S4

A. Intersection of genes in the recurrent regions with cell cycle genes (signature obtained by comparison of samples with high proliferation index, such as rapidly dividing early CD71+erythroid progenitors and CD105+ endothelial cells, with somatic samples, (Assou et al., 2009), DNA repair genes (Wood et al., 2005), pro-and anti-apoptotic genes (BCL2 family members).

B. Intersection with pluripotency-associated genes (previously published data set with a consensus PSC stemness gene list, (Assou et al., 2007) and cancer genes (Cancer Gene Census, http://www.sanger.ac.uk/genetics/CGP/Census/). Venn diagrams show the number of genes in each comparison and the genes shared.

Supplemental Tables

Table S2: The most frequently used methods for genetic stability assessment with their main advantages and disadvantages.

	Cells		**DNA**		
	G-banding Karyotype	**FISH**	**Microarray**	**Whole genome/ exome sequencing**	**PCR/ddPCR**
Resolution	Poor	Excellent			
Quantity	Several cells in metaphase	Several cells	>1000 ng of DNA		>1ng of DNA
Sensitivity	≥10%	1%	10 - 20%	10 - 20%	10 - 20%
Price (dollars)	400-600	200-300	500	1500	100-200
Results	Require specialist staff		Need for a bio-informatician for analysis		Easy (software)
Timing	1-2 weeks				1 day
Advantages	- Gold standard for the detection of aneuploidy, polyploidy, and other large chromosomal imbalances	- High sensitivity and reproducibility	- Provide information on DNA regions with gains or losses	- Very high and scalable throughput, sensitivity and accuracy - Assess the whole genome at single-base resolution	- High precision for the CNV and SNV detection at a reasonable cost
Disadvantages	- High number of metaphases are needed - Cannot detect sub-karyotypic variants	- Does not allowthe comprehensive screening of chromosomal aberrations	- Cannot detect balanced rearrangements, such as inversions.	- Demanding computational power - Huge data analysis workload - Complex result interpretation -Expensive	- Balanced rearrangements not detected - Does not allow the comprehensive screening of chromosomal aberrations

Table S3: List Bio-Rad ddPCR assay ID

Chromosome	**Unique Assay ID (Bio-Rad)**	**Gene symbol**	**Locus**
20	#dHsaCP2506319	ID1	20q11
12	#dHsaCP1000374	NCAPD2	12p13
X	#dHsaCP2506654	STS	Xp22

17	# dHsaCP1000054	RPS6KB1	17q23
1	#dHsaCP1000482	SOAT1	1q25
5	#dHsaCNS501868922	PITX1	5q31
Reference	# dHsaCP2500350	RPP30	10q23

Supplemental Experimental Procedures

Recurrence scores

A first analysis was carried out using a recurrence score (RS) for data split in two datasets (>10 bp and ≤10 bp) from which polymorphic data (sequences present in dbSNP or DGV) were removed. RS of the first analysis was computed by comparing each abnormality to all the others and by identifying abnormalities with a reciprocal overlap of at least 0.2. Regions with a reciprocal overlap higher than 80% were merged. For each overlap, RS was computed as follows: RS = a * s, where (a) is the number of abnormalities that contributed to define this overlap (identical abnormalities from the same cell line in the same study were counted only once) and (s) the number of different studies from which these overlapping abnormalities came from.

To define recurrent genetic abnormalitiesn for each abnormality >10 bp we computed an overlap 'Ov' as follow: Ov = End Ov – Start Ov, where End Ov is the minimal value of the ends of both abnormalities that are compared, and Start Ov is the maximal of their starts. Equilibrated translocations were excluded.

Bedtools (Quinlan and Hall, 2010) was used to identify common abnormal regions. The common regions that cover the highest number of genetic abnormalities were calculated in a spreadsheet.

Cell reprogramming and cell passaging

The human hESC lines HD129 and HD291 were derived in our laboratory (Bai et al., 2015). The hiPSC lines UHOMi001-A (Ahmed et al., 2018), RSP4, iCOPD2A1, iCOPD9A2, and HY03 were reprogrammed using the Sendai virus and the CytoTune®-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific), and

display all the PSC features: grow as typical PSCs, are positive for pluripotency markers (*OCT4*, *NANOG*, *SOX2*, *TRA1-60*, *TRA1-81*, *SSEA3*, *SSEA4*) and for phosphatase alkaline activity, and can differentiate into cells of all three germ layers. Mechanical passaging was carried out under an inverted microscope in a hood using scalpels. For single-cell enzymatic passaging and clump passaging, colonies were pre-incubated with the Rho-associated protein kinase (ROCK) inhibitor Y-27632 for 1h, and then dissociated with TrypLE™ Select (Invitrogen) or EDTA (Versene Solution, Thermo Fisher Scientific) at 37°C for 10min.

Culture medium collection and nucleic acid extraction

Before passaging, supernatant (1.5 mL) was collected into a safe-lock tube (DNase-free) from cell cultures that were at least 70% confluent. DNA was extracted from 200 µL of supernatant using the QIAmp DNA Mini Blood Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Briefly, 20 µL proteinase K and 200 µL Buffer AL were added to each supernatant sample. After pulse vortexing for 15s, the lysis mixture was incubated in an Eppendorf tube (1.5 mL) at 56°C for 10min. The highly denaturing conditions and elevated temperatures favored the complete release of DNA from any bound proteins. After adding 200 µL of cold ethanol (100%) to the lysates, samples were transferred in QIAamp Mini columns and centrifuged at 6000g for 1min followed by two wash steps (in Buffer AW1 and Buffer AW2) to eliminate contaminants. Then, supernatant-DNA was eluted in 60 µL Buffer AE and stored at -20°C.

Quantification of supernatant-DNA by ALU-qPCR and QuBit

Supernatant-DNA was analyzed by qPCR (LC480, Roche)using the ALU 115 and ALU 247 primers, as previously described in (Umetani et al., 2006). The sequences of the ALU115 primers were: forward, 5′-CCTGAGGTCAGGAGTTCGAG-3′; reverse, 5′-CCCGAGTAGCTGGGATTACA-3′. The ALU247 primers were: forward, 5′-GTGGCTCACGCCTGTAATC-3′; reverse, 5′-CAGGCTGGAGTGCAGTGG-3′. One µL of each eluted supernatant-DNA sample was added to a reaction mixture containing 2X LightCycler480 SYBR Green I master mix (Roche Applied Science, Germany) and 0.25 µM of forward and reverse primers (ALU-115 and ALU-247) as described in (Umetani et al., 2006) in a total volume of 10 µL. Reactions were carried out in 96-well white plates using an EpMotion 5070 Liquid Handling Workstation (Eppendorf). All reactions were performed in triplicate. A negative control (RNAse/DNAse-free water) was included in each run. The supernatant-DNA concentration was determined using a standard curve obtained by successive dilutions of a commercial human genomic DNA sample. DNA integrity was calculated as the ratio of the qPCR results with the two primer sets (ALU115 and ALU247). The ALU115 set amplifies smaller fragments that result from apoptosis and the ALU247 set amplifies only larger fragments that result from necrosis. Supernatant-DNA concentration was quantified using the QuBit dsDNA HS Assay Kit and a Qubit 2.0 fluorometer following the manufacturer's instructions (Life Technologies).

Flow cytometric detection of apoptosis and necrosis using the Annexin-V and 7-AAD assay

The PE Annexin-V Apoptosis Detection Kit I (BD Pharmingen, Ref: 559763) was used to quantify the percentage of apoptotic and necrotic cells in supernatant samples. Briefly, samples were incubated with PE Annexin-V in buffer containing 7-

Amino-Actinomycin (7-AAD) according to the kit protocol (http://www.bdbiosciences.com/ds/pm/tds/559763.pdf), and analyzed by flow cytometry at the Montpellier Resources Imaging (MRI) facility (https://www.mri.cnrs.fr/en). HiPS cells are used as controls for FACS gating.

Digital droplet PCR (ddPCR)

The ddPCR workflow was performed according the Bio-Rad instructions (Bio-Rad QX200 system). Briefly, reactions were set up using one primer pair that targets the region of interest (for instance: CNV-chr20) and a second primer pair that targets the reference gene (*RPP30*). The two primer sets were labeled with different fluorophores (FAM and HEX). DNA (amount) from each sample was added to the TaqMan PCR reaction mixture (final volume of 20 µL) that included 2XSupermix No dUTP (Bio-Rad, Ref: 1863023) and the primer sets. Each reaction mixture was loaded in a disposable plastic cartridge (Bio-Rad) with 70 µL of droplet generation oil (Bio-Rad) and placed in the droplet generator (Bio-Rad). The cartridge was removed from the droplet generator, and the droplets collected in the droplet well were then manually transferred with a multichannel pipette to a 96-well PCR plate. The PCR amplification conditions were: 94°C for 10min, 40 cycles of 94°C for 30s, and 60°C for 1min, followed by 98°C for 10min and ending at 4°C. After amplification, the plate was loaded into the QX200 Droplet Reader (Bio-Rad). Copy number was assessed using the Quantasoft software. For testing the ddPCR sensitivity in detecting a CNV-12q gain, the UHOMi001-A diploid and the HD291 aneuploid line were used. Cells were grown on Geltrex matrix in E8 Medium prior to the experiment and then dissociated using trypsin and counted. After mixing the two cell lines to obtain increasing concentrations (from 0% to 100%) of abnormal cells within the diploid

population, each mixed sample was processed for genomic DNA extraction using the QIAmp DNA Mini Blood Kit (Qiagen, Hilden, Germany) and for ddPCR analysis. Reference for the designed BioRad ddPCR probes can be found in Table S3.

To establish the sensitivity of the droplet digital PCR method, we used sampleswith increasing percentages (from 0 to 100%) of hPSCs harboring a trisomy 12 within a sample of euploid hPSCs. A significant difference in the CNV copy number compared with control (0%) was observed in samples with at least 10% of abnormal cells (*p*-value <0.05, Student's t test). The panels represent three biological replicates. The error bars (generated by the QuantaSoft software) for each well represent the 95% confidence intervals using Poisson statistics with the total number of droplets. All values of copy number were corrected by adding -0.083 because more than 50 experiments on samples with a normal count of chromosome 12 have shown a bias with this probe with a median value of 2.083.

Generation of the FOXJ1_mCherry and CCDC40_KO iPSC lines using CRISPR/Cas9

A stock of HY03 (75k) M53Cl2SC6 cells was made and after thawing their euploidy was confirmed using the iCS-digital test for detection of CNV anomalies at mechanical passage M53, clumps passage Cl2, single cell passage SC11 for FOXJ1_mCherry tagging, and at mechanical passage M53, clumps passage Cl2, single cell passage SC17 for CCDC40_KO cells respectively. The day before transfection, 25000 HY03 (75k) M53Cl2SC7 cells per cm^2 were plated in a 6-well plate coated with Geltrex matrix and with E8 supplemented with Y-27632 (10µM). The day of transfection, medium was refreshed using the Lipofectamine Stem transfection reagent (Invitrogen) at least 2 hours before transfection, following the

manufacturer's instructions. Briefly, 2 µg of pSpCas9(BB)-2A-GFP (PX458) (gift from Feng Zhang, Addgene plasmid #48138) containing the FOXJ1 targeting sgRNA sequence 5'-GGGCCTTCTTGTAAGAGGCC-3' or the CCDC40 targeting sgRNA sequence 5'-CTCCTCGTTGGCGGCTGCGC-3' with 1 µg of MBX plasmid (gift from Linzhao Cheng, Addgene plasmid #64122) and 1 µg of homemade donor plasmid pUC19_FOXJ1_mCherry_cNEO for FOXJ1_mCherry tagging were mixed with 4 µL of Lipofectamine and left at room temperature for 10min to form complexes. The Lipofectamine-DNA complexes were added on top of the cells, distributed by gently swirling the plate, and incubated at 37˚C, 5% CO_2. The following day, the medium was changed with fresh E8 medium supplemented with G418 (200 µg/ml) for the FOXJ1_mCherry cells, and then changed daily for 6 days. Colonies were manually picked and transferred into a 96-well plate for amplification. At confluence, clones were passaged to a 24-well plate, and then to a 6-well plate. DNA was collected to screen clones by bridge-PCR, transgene copy counting, and Sanger sequencing for FOXJ1_mCherry tagging or by high resolution melt analysis (HRMA) followed by Sanger sequencing for CCDC40_KO. Finally, the presence of genomic abnormalities was checked using the iCS-digital test.

Supplemental references

Ahmed, E., Sansac, C., Fieldes, M., Bergougnoux, A., Bourguignon, C., Mianné, J., Arnould, C., Vachier, I., Assou, S., Bourdin, A., et al. (2018). Generation of the induced pluripotent stem cell line UHOMi001-A from a patient with mutations in CCDC40 gene causing Primary Ciliary Dyskinesia (PCD). Stem Cell Res. 33, 15–19.

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.

Martins-Taylor, K., Nisler, B.S., Taapken, S.M., Compton, T., Crandall, L., Montgomery, K.D., Lalande, M., and Xu, R.-H. (2011). Recurrent copy number variations in human induced pluripotent stem cells. Nat. Biotechnol. 29, 488–491.

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.-C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7, 521–531.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. 26, 841–842.

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.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. 26, 841–842.

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.

Taapken, S.M., Nisler, B.S., Newton, M.A., Sampsell-Barron, T.L., Leonhard, K.A., McIntire, E.M., and Montgomery, K.D. (2011). Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. Nat. Biotechnol. 29, 313–314.

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.

Wood, R.D., Mitchell, M., and Lindahl, T. (2005). Human DNA repair genes, 2005. Mutat. Res. 577, 275–283.