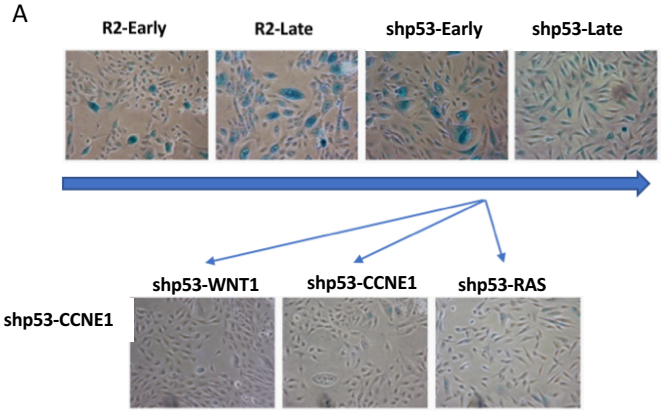
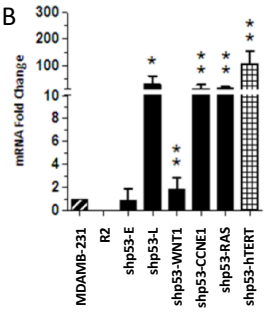
A

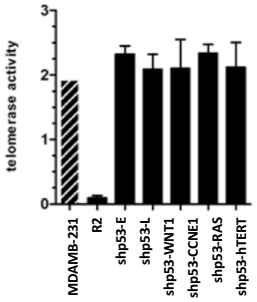


B



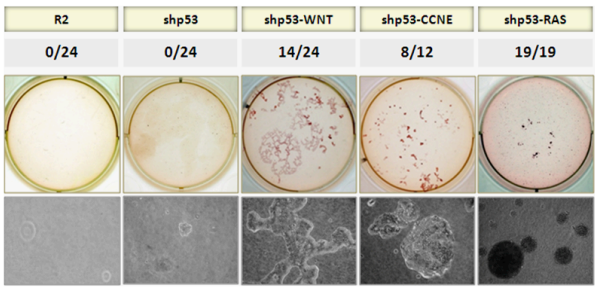
hTERT mRNA expression
In shp53 and shp53 + oncogene sublines

C



Telomerase activity
In shp53 and shp53 + oncogene sublines

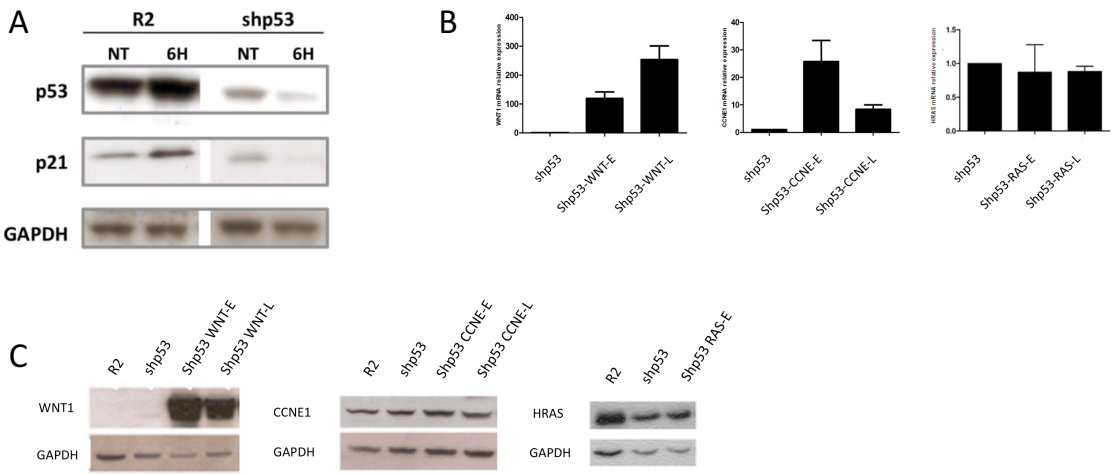
D



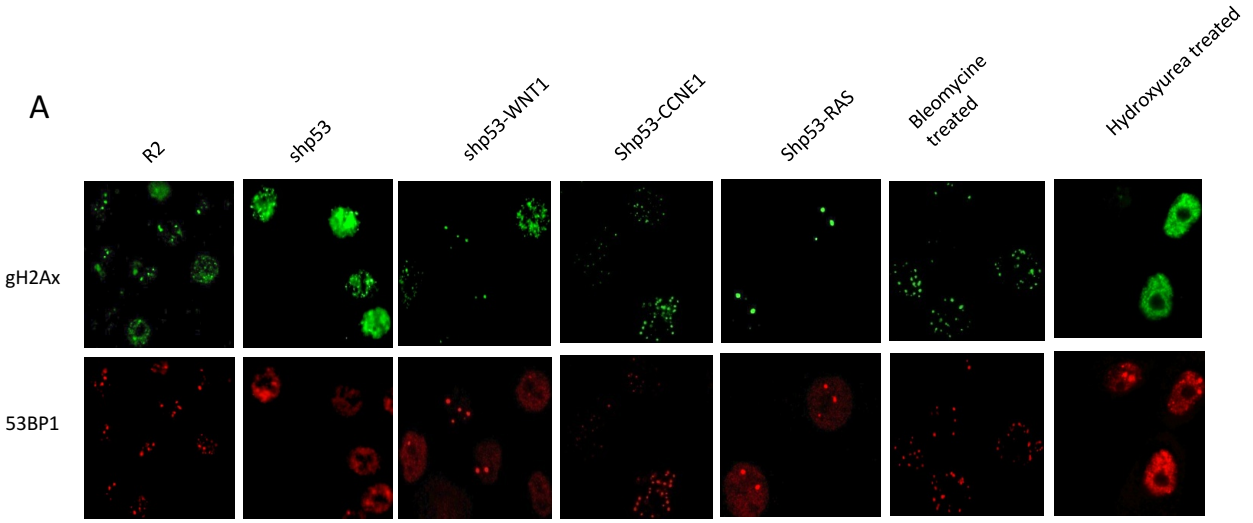
E

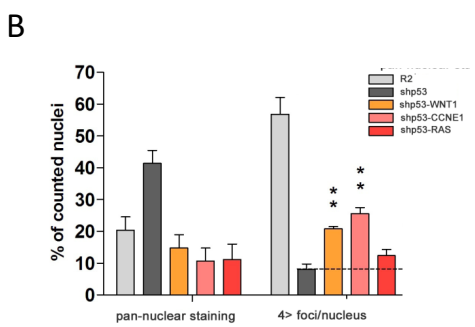
	Tumors in nude mice	Tumors in SCID mice	Tumors in SCID mice
Number of cells injected	1.5×10^6	1.5×10^6	5×10^6
R2.shp53-WNT1.SA	0/6	0/12	0/4
R2.shp53-CCNE1.SA	0/6	0/12	0/4
R2.shp53-RAS.SA	0/6	0/12	0/4

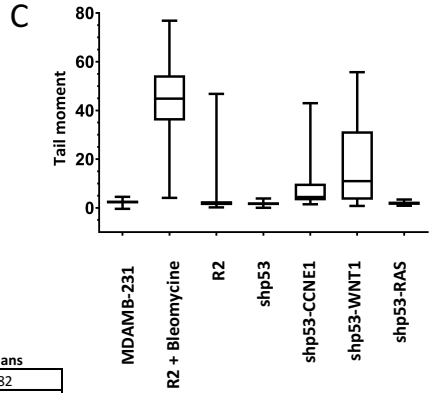
Supplementary Fig. 1 : **Gradual immortalization and transformation of the shp53 HMEC sublines. A,** b-galactosidase staining; **B, C**: HTERT mRNA expression and telomerase enzymatic activity, the shp53 HTERT cells are presented as a positive control of TERT expression; **D**: anchorage independent growth in soft agar, number of positive experiments out of number of attempts, pictures of the corresponding soft agar petri dish and blow up of foci formed by the respective sublines. **E**: shp53 sublines did not form tumor upon injection in immunocompromised mice. First line indicates the total number of cells injected into the interscapular fat pad. Each subline was injected in parallel in 6 mice and tumor growth monitored for 5 months. Statistical significance was determined by the Anova Dunn's Multiple Comparison test. Error bars correspond to s.e.m.



Supplementary Fig. 2: Attenuation of p53 after shRNA transduction and overexpression of oncogenes in oncogene transduced sublines. A: attenuation of p53 was ascertained by challenging primary R2 and R2-shp53 cells with Bleomycin for 6 hours, accumulation of p53 and induction of p21 were used as read outs for p53 functionality. **B**: QPCR verification of WNT1, CCNE1 and RAS mRNA expression in R2-shp53-WNT1, R2-shp53-CCNE1 and R2-shp53-RAS respectively. **C**: protein expression levels by western blotting. E: early; L: Late.



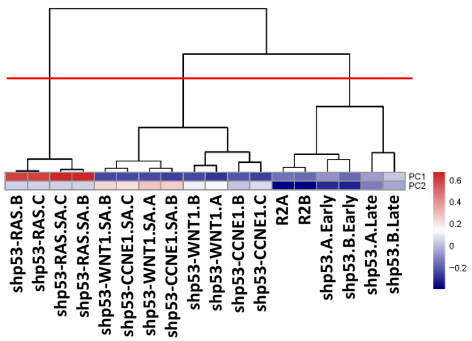




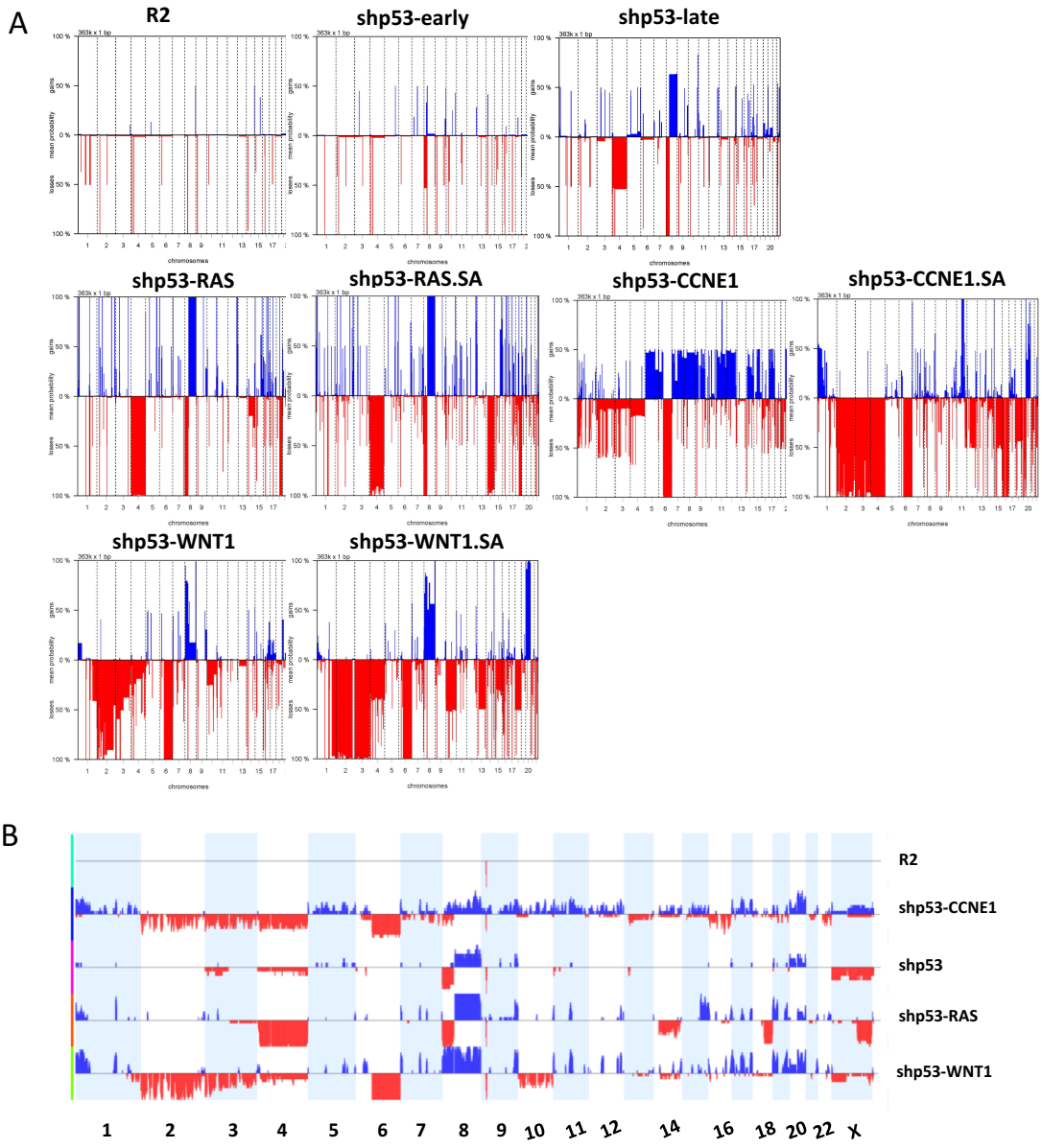
D

	Means	StD	Err. Std	Medians
R2	2.689	5.924	0.7647	1.582
R2 + bleo	44.57	13.98	1.563	44.85
ShP53	1.740	0.6688	0.07477	1.612
shp53.CCNE1	9.036	10.68	1.379	4.393
shp53.WNT1	17.40	15.63	1.748	10.94
shp53.RAS	1.947	0.5349	0.05981	1.911
Anova -Dunn's Multiple Comparisons test	**Mean rank diff.**	**Significant? P<0.001**	**Summary**	**Adjusted p value**
R2 vs shp53	10.89583	No	ns	>0.99
R2 vs shp53.CCNE1	-138.5333	Yes	***	<0.001
R2 vs shp53.WNT1	-161.3167	Yes	***	<0.001
R2 vs shp53.RAS	-16.20417	No	ns	>0.99
shp53 vs shp53.CCNE1	-149.4292	Yes	***	<0.001
shp53 vs shp53.WNT1	-172.2125	Yes	***	<0.001
shp53 vs shp53.RAS	-27.10000	No	ns	>0.99
shp53.CCNE1 vs shp53.WNT1	-22.78333	No	ns	>0.99
shp53.CCNE1 vs shp53.RAS	122.3292	Yes	***	<0.001
shp53.WNT1 vs shp53.RAS	145.1125	Yes	***	<0.001

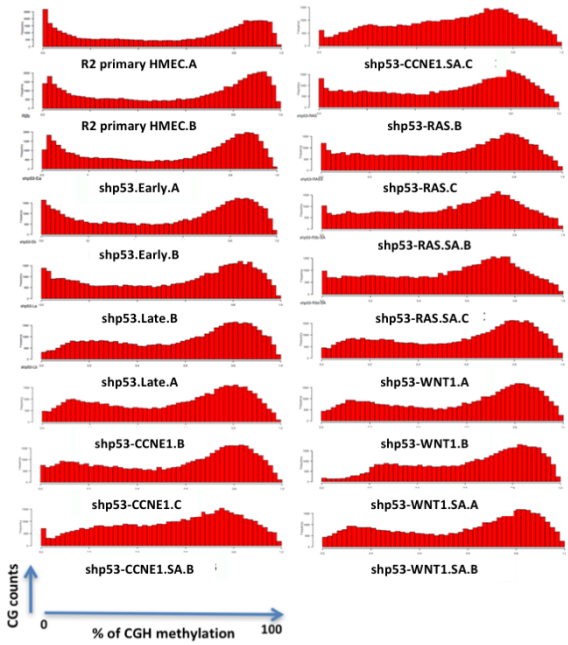
Supplementary Fig. 3: Spontaneous DNA damage in shp53 sublines. **A**: gammaH2Ax and 53BP1 staining patterns, note the pannuclear H2Ax staining in shp53 HMECs, whereas shp53-WNT1, shp53-CCNE1 and shp53-RAS show predominantly H2Ax foci. Hydroxyurea treatment was used as a control of H2Ax staining as a condition of severe replication stress. It is of note that primary R2 HMEC presented a sizeable number of foci positive cells, which can be attributed to telomere attrition in these cells. Bleomycine treatment was used as a control of double strand breaks. **B**: fraction of cells showing H2Ax pan-nuclear staining or more than 4 foci. **C**: Tail moment measurements box plot. **D**: Statistics of Neutral CometAssay tail moment measurement for 3 independent experiments in shP53 and shp53-oncogene sublines (top table) and Anova Dunn's multiple comparison test (bottom table). shP53-CCNE1 and shP53-WNT1 sublines present a significantly higher number of double strand breaks than shP53-RAS.



Supplementary Fig.4: HMEC models form 3 clusters based on the oncogene transduced.

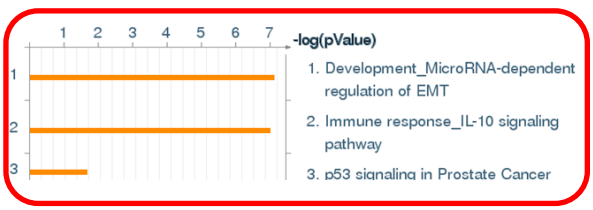


Supplementary Fig.5: CNA plots of HMEC models. A: CNA at different steps of transformation. CNAs are represented for each chromosome, red for losses, blue for gains. The hight of the bars indicates the probability of occurrence. B: **cumulated CNA plots of HMEC models**. CNAs are represented for each chromosome, red for losses, blue for gains. The hight of the bars indicates the amplitude of the copy number change.

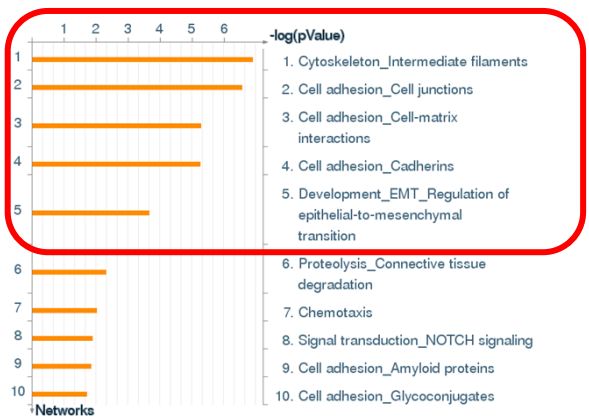


Supplementary Fig.6: density histograms of RRBS methylation scores at CpGs sites (at least 5 contiguous CG) genome wide in primary R2 HMECs, shp53.Early, shp53.Late, shp53-CCNE1, shp53-CCNE1.SA (Soft Agar), shp53-WNT1, shp53-WNT1.SA, shp53-RAS, shp53-RAS.SA

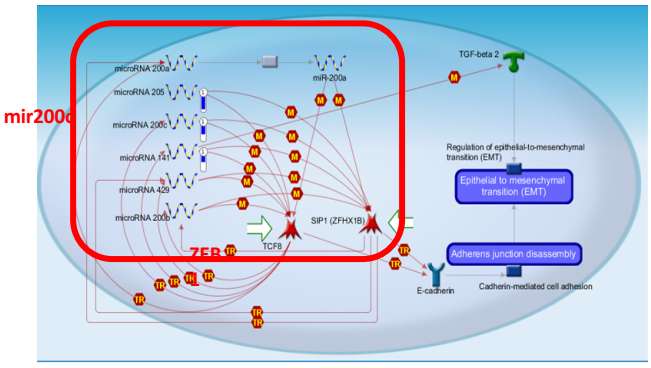
miR



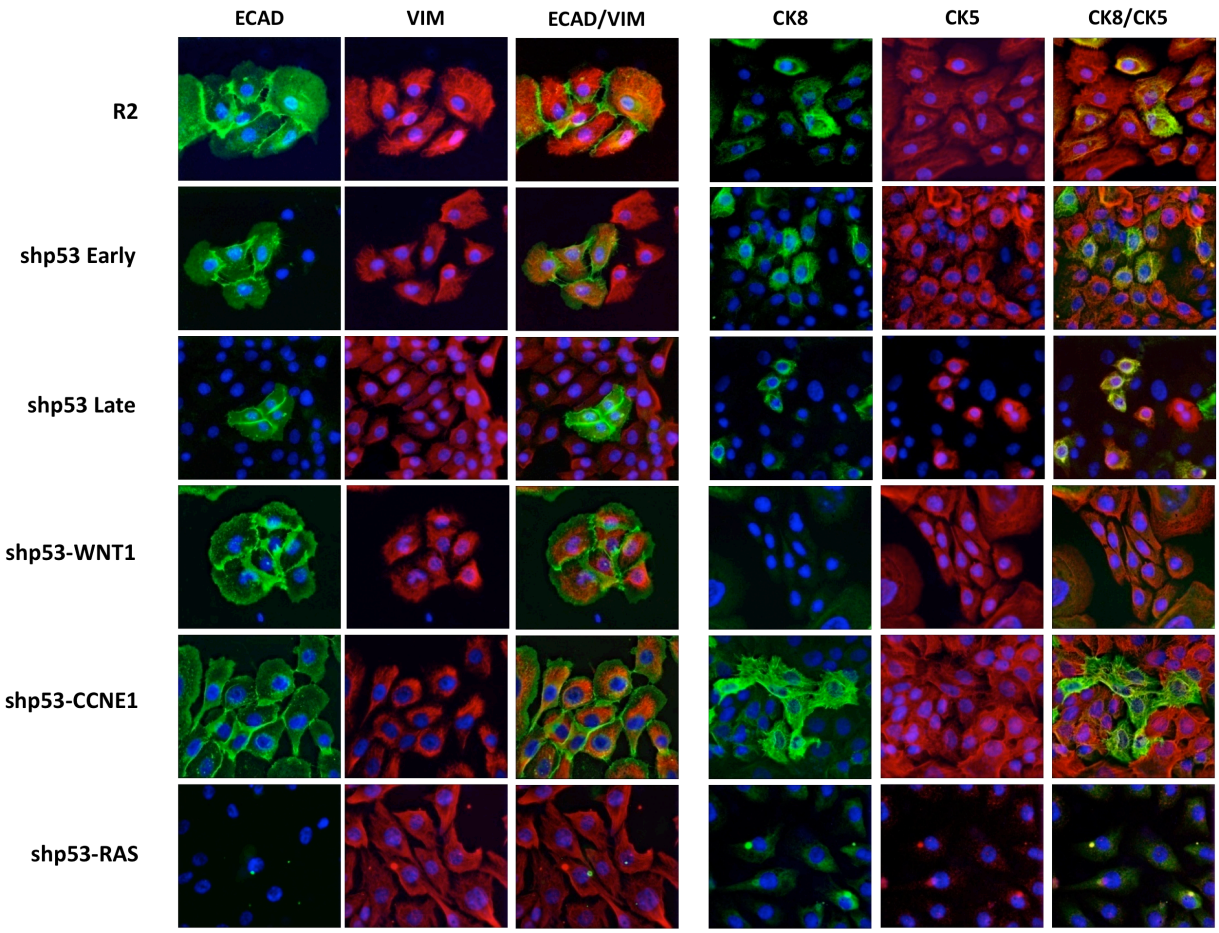
mRNA top N% = 185 genes



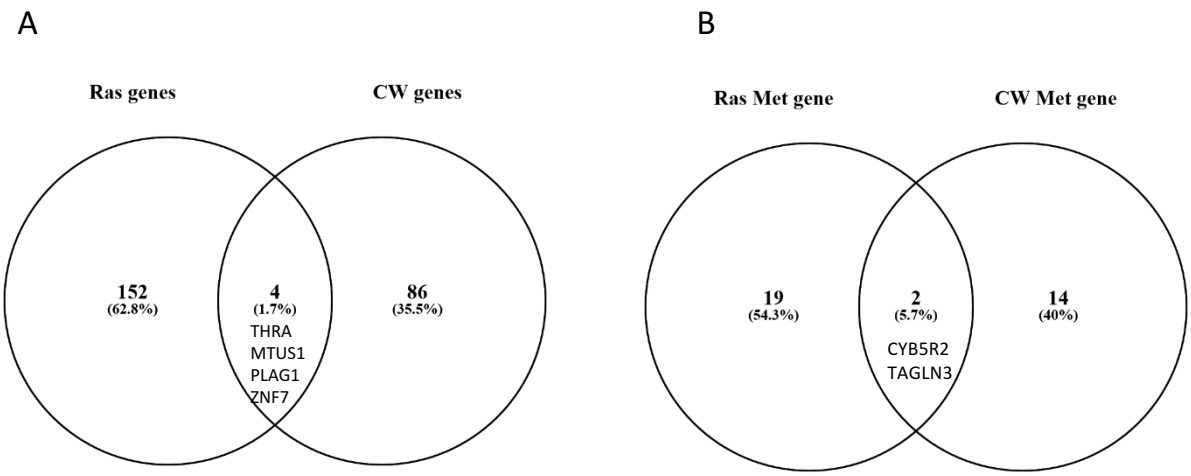
Regulation network



Supplementary Fig.7: pathways and regulation networks principally affected in the different HMEC sublines.

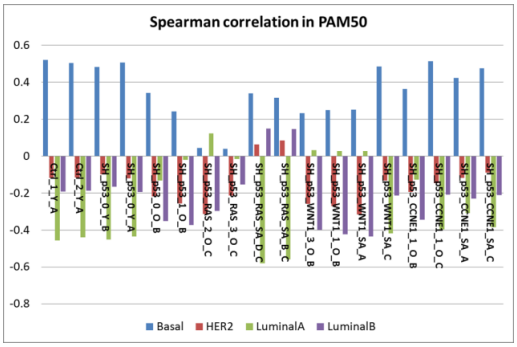


Supplementary Fig. 8: **phenotypic characteristics of shp53 HMEC sublines.** Cells were stained by immunofluorescence (IF) for the expression of ECAD (E-Cadherin, green) which is a marker of epithelial cells, VIM (Vimentin, red) marker of mesenchymal cells, CK8 (Cytokeratin 8) marker of luminal breast epithelial cells, CK5 (Cytokeratin 5) marker of basal breast epithelial cells. Differential expression patterns can be observed according to the genetic elements expressed and the stage of the culture. Normal HMEC and Early shp53 co-express ECAD and VIM and are mosaiec for CK5 and CK8 expression. In shp53 late HMEC tended to lose ECAD expression and become mesenchymal, but kept a mosaic CK5/CK8 pattern. In shp53-WNT1 and shp53-CCNE1 ECAD and VIM were co-expressed in all cells indicating the conservation of an epithelial phenotype. Interestingly, whereas shp53-WNT1 expressed only CK5 and no CK8, shp53-CCNE1 preserved the original mosaic phenotype of the shp53 HMECs. Expression of RAS-v12 produced drastic changes as illustrated by the concomitant loss of expression of ECAD and both CK5 and CK8.

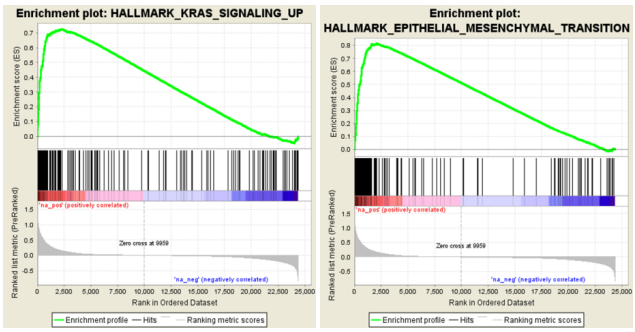


Supplementary Fig.9: Genes modified by CNA or differential methylation in R2$^{shp53\text{-}RAS}$ and R2$^{shp53\text{-}CCNE1}$/ R2$^{shp53\text{-}WNT1}$ show little overlap. MTUS1 is strongly underexpressed in RAS (–2,95;about 1/10 x) whereas it is overexpressed in CW (0,87, about 1,8 x). PLAG1 is strongly overexpressed in RAS (3,25; about 9,5 X) and moderatly in CW (0,84 about 1,8 x) similarly to THRA (RAS change =1,89 about 3,7x; CW change =0,89 about 1,85x) . ZNF7 expression change is equivalent in both clusters.

A



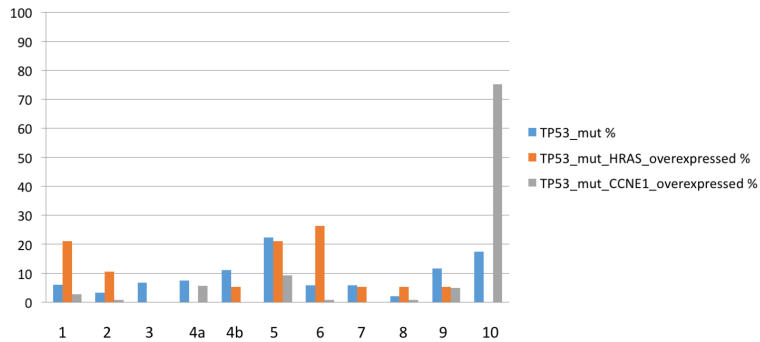
B



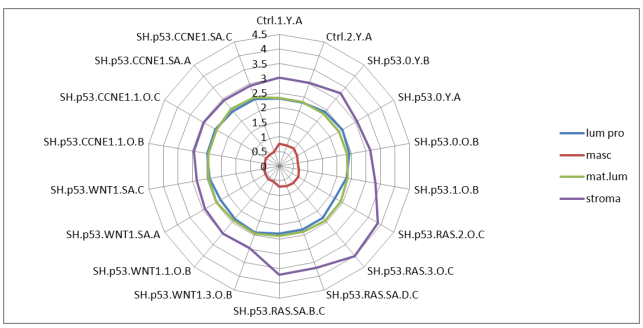
C

NAME	ES	NES	NOM p-val	FDR q-val	FWER p-val
HALLMARK_COMPLEMENT	0.7478657	1.8861527	0.0	8.095238E-4	0.001
HALLMARK_ALLOGRAFT_REJECTION	0.8131348	1.788368	0.0	0.0043991245	0.008
HALLMARK_INFLAMMATORY_RESPONSE	0.7507069	1.7823858	0.0	0.0029327495	0.008
HALLMARK_IL2_STAT5_SIGNALING	0.6613298	1.7780305	0.0	0.0021995623	0.008
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.75866187	1.7713786	0.0	0.0027169033	0.011
HALLMARK_KRAS_SIGNALING_UP	0.7131533	1.7637798	0.0	0.0025974195	0.012
HALLMARK_COAGULATION	0.71557105	1.7216653	0.0	0.0031839232	0.017
HALLMARK_APOPTOSIS	0.632473	1.7103698	0.0	0.004203047	0.022
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.82959616	1.7038239	0.0	0.0045748223	0.026
HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.68856674	1.6477119	0.0062	0.0100955255	0.054
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	0.7981844	1.592707	0.0062	0.021111127	0.105
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.8284832	1.5537144	0.022869023	0.030158127	0.167

D



E



Supplementary Fig.10: HMEC models present a high luminal progenitor and luminal mature score and Basal-like and claudin-low breast cancer resemble shp53-CCNE1/WNT1 and shp53-RAS HMEC models respectively. **A:** HMEC models were classified according to breast cancer molecular subtypes using the PAM50 classifier and correlate with the Basal-like subtype. **B-C**: principal pathways activated in shp53-RAS models highlighting the importance of the RAS and EMT pathways. **D:** Fraction of breast tumors showing a p53 mutation (blue bars), p53 mutation and RAS overexpression (orange bars, p53 mutation and CCNE1 overexpression (grey bar) in 10 Inclust molecular subgroups. **E**: radar plot of the mammary differentiation scores as defined by E. Lim et al (2009)

.

Supplementary Informations

Materials

Derived sublines

#	Model name	days in culture	doubling time	Work nomenclature	profiled in moGSA
1	R2.A	50-70	48-55 hrs	R2#1	1
2	R2.B	50-70	48-55 hrs	R2#2	1
3.E	shp53.A.Early	40-60 **	36-40 hrs	R2_shP53#0_Y	1
3.L	shp53.A.Late	250-300 **	32-36 hrs	R2_shP53#0_O	1
4.E	shp53.B.Early	40-60 **	36-40 hrs	R2_shP53#1_O	1
4.L	shp53.B.Late	250-300 **	32-36 hrs	R2_shP53#1_Y	1
5.E	shp53-CCNE1B.Early	120-150 **	30-32 hrs	R2_shP53-C#1_Y	0
5.L	shp53-CCNE1.B.Late	250-300 **	30-32 hrs	R2_shP53-C#1_O	1
6.E	shp53-CCNE1C.Early	120-150 **	30-32 hrs	R2_shP53-C#3_Y	0
6.L	shp53-CCNE1.C.Late	250-300 **	30-32 hrs	R2_shP53-C#3_O	1
7	shp53-CCNE1.SA.A	300-320 **	26-30 hrs	R2_shP53-C_SA#1A	1
8	shp53-CCNE1.SA.C	300-320 **	26-30 hrs	R2_shP53-C_SA#1C	1
9.E	shp53-RAS.B.Early	120-150 **	26-30 hrs	R2_shP53-R#2_Y	0
9.L	shp53-RAS.B.Late	250-300 **	24-28 hrs	R2_shP53-R#2_O	1
10.E	shp53-RAS.C.Early	120-150 **	26-30 hrs	R2_shP53-R#3_Y	0
10.L	shp53-RAS.C.Late	250-300 **	24-28 hrs	R2_shP53-R#3_O	1
11	shp53-RAS.SA.B	300-320 **	24-28 hrs	R2_shP53-R_SA#3B	1
12	shp53-RAS.SA.C	300-320 **	24-28 hrs	R2_shP53-R_SA#3D	1
13.E	shp53-WNT1.A.Early	120-150 **	28-32 hrs	R2_shP53-W#1_Y	0
13.L	shp53-WNT1.A.Late	250-300 **	28-32 hrs	R2_shP53-W#1_O	1
14.E	shp53-WNT1.C.Early	120-150 **	28-32 hrs	R2_shP53-W#3_Y	0
14.L	shp53-WNT1.C.Late	250-300 **	28-32 hrs	R2_shP53-W#3_O	1
15	shp53-WNT1.SA.A	300-320 **	28-32 hrs	R2_shP53-W_SA#1A	1
16	shp53-WNT1.SA.C	300-320 **	28-32 hrs	R2_shP53-W_SA#1C	1

In total 16 sublines are described in this work. Some sublines have been split in two to distinguish Early (xx.E) and Late (xx.L) passages, but they correspond to the same starting material. Eighteen (18) samples (one biological duplicate by condition) have been fully characterized by omics (CGH, miR, mRNA and RRBS).

Methods

Generation of cellular models and culture conditions

Human Mammary Epithelial Cells (HMECs) were isolated from mammary gland explants obtained from plastic surgical after informed consent from the patient. This work was approved by the Ethics committee of the University of Montpellier. Cell suspensions were produced by mechanical and enzymatic dissociation with 1% collagenase. Fibroblasts were eliminated by multiple centrifugations and epithelial organoids kept for a second round of dissociation using 0.25% trypsine. Resulting cell suspensions were cultured at 37°C in 5% CO2, in MEBM medium supplemented with antibiotics (Gibco/Thermo-Fisher, Illkirch-Graffenstaden, France), 1M HEPES, hydrocortisone, insulin, EGF, BPE

and Gentamycine (*MEGM single Quots*, Lonza, Levallois-Perret, France). Primary HMEC cultures were transduced with amphotropic retroviral supernatants produced by 293T cells transfected with one of the following retroviral constructs: pSUPER.retro.hygro-shp53, pBABE.neo-CCNE1, pLNC-WNT1, pBABE.puro-HRASV12. Transduction was followed by a 3 weeks period of antibiotic selection. Following local regulations retroviral transduction was done in an L3 facility and transduced cells kept in the L3 for 1 month until they were negative for reverse transcriptase.

Chromosome counts

Cells were arrested in metaphase with 5µM Nocodazole (Sigma-Aldrich, St Quentin Fallavier, France) for 4h, trypsinized and resuspendedn in 10 ml of a hypotonic solution (SVF 15%, KCL 0,01M) at 37°C for 10 min and fixed min and fixed in 3:1 methanol:glacial acetic acid. Metaphase spread are obtained by dropping a drop on glass slides.

DNA and RNA extraction

DNA and RNA were isolated using the QIAmp DNA Mini kit and Rneasy Mini Kit (Qiagen S.A. France, Courtaboeuf, France). Each DNA sample was quantified by nanospectrophotometry (NanoView, GE Healthcare, Orsay, France) and qualified by 0.8% agarose electrophoresis. Qualification of mRNA was performed using a Bioanalyser (Agilent, Santa Clara, CA, USA). For miR profiling total RNA were extracted uzing Trizol (Invitrogen/Thermo-Fisher, Illkirch-Graffenstaden, France) according to manufacturer's instructions.

RT-qPCR

Total RNA were reverse-transcribed using *RT-superscript III* (Invitrogen) and oligo-dT primers. QPCR was performed on the cDNAs using the *Power SYBR*-Green mix (Applied Biosytem,) on the *7300 real-time PCR system* (Applied Biosytem). Expression level was normalized on 28S rRNA. Primers used were as follows

	Sens	Anti-sens
Hygromycine B	CGGGGATTCCCAATACGAGG	CTACACAGCCATCGGTCCAG
WNT1	GACCTGCGCTTCCTCATG	GGTTGCCGTACAGGACGC
CCNE1	GGTATCAGTGGTGCGACATAGA	CGCTGCTCTGCTTCTTACC
HRAS	ATGACGGAATATAAGCTGGTGG	CTGTACTGGTGGATGTCCTCAA

PCR primers used to verify the integration of the retroviral construct

	Sens	Anti-sens
p21	CGAAGTCAGTTCCTTGTGGAG	CATGGGTTCTGACGGACAT
p53	AGGCCTTGGAACTCAAGGAT	CCCTTTTTGGACTTCAGGTG
WNT1	CGCTGGAACTGTCCCACT	AACGCCGTTTCTCGACAG
CCNE1	GGCCAAAATCGACAGGAC	GGGTCTGCACAGACTGCAT
HRAS	GGACGAATACGACCCCACTA	GCACGTCTCCCCATCAAT
HTERT	CACGCGAAAACCTTCCTC	ACCACTGTCTTCCGCAAGTT
28S	AGCCAAGCTCAGCGCAAC	CGATCCATCATCCGCAATG

QPCR primers

Protein extraction and Western Blotting

Cells were lysed in TE at pH8, 40mM NAPPi, 50mM NaF, 5mM MgCl2, 100µM Na3VO4, 1% Triton supplemented with protease inhibitor (Fermentas). Protein concentrations were determined with *BCA protein assay reagent* kit (Pierce). 25µg/sample of extract was loaded on denaturing polyacrylamide gel and transfered on PVDF membrane (Biotrace). The PVDF membrane was saturated in 5% dry-powdered milk containing 0.1% TBS-tween and incubated overnight with the primary antibody, rinsed in 0.1% TBS-tween and incubated 2h with the secondary antibody and revealed by chemioluminescence with *Western Lightning Plus-ECL* kit (PerkinElmer). The list of antibodies was as follows.

Antibodies used for Western blotting and Immunofluorescence

Anticorps	Dilution	Dealer
Mouse anti-p53 (DO7)	1/1000e	Dako
Rabbit anti-p21 (C-19)	1/500e	Santa Cruz
Rabbit anti-p16 (C-20)	1/500e	Santa Cruz
Rabbit anti-WNT1	1/500e	Abcam
Rabbit anti-CCNE1	1/500e	Abcam
Rabbit anti-HRAS (C-20)	1/500e	Santa Cruz
Mouse anti-GAPDH (6C5)	1/1000e	Santa Cruz
Anti-Mouse peroxydase	1/5000e	Santa Cruz
Anti-Rabbit peroxydase	1/5000e	Santa Cruz
Mouse anti-γH2AX	1/100e	Millipore
Mouse anti-E-cadhérine	1/100e	Abcam
Rabbit anti-53BP1	1/100e	Cell Signaling
Rabbit anti-Vimentine	1/100e	Abcam
Mouse anti-CK8	1/100e	Ozyme
Rabbit anti-CK5	1/100e	Ozyme
Alexa Fluor 488 anti-mouse	1/1000e	Invitrogen
Alexa Fluor 488 anti-mouse	1/1000e	Invitrogen

Array-CGH

Array-CGH was done using HG18 CGH 385K Whole Genome v2.0 array (Roche NimbleGen, Madison, WI, USA). DNA from a pool of 20 normal females was used as reference. For hybridization, 1 mg of genomic DNA and reference DNA were labeled using NimbleGen Dual-Color DNA Labeling Kit (Roche Diagnostics, Meylan, France). Labeling products were precipitated with isopropanol and resuspended in water. Test (Cy3) and reference (Cy5) samples were combined in 40 ml of NimbleGen Hybridization buffer. Hybridization was performed in a NimbleGen Hybridization system 4 for 48 h at 42 C with agitation mode B and washed using NimbleGen Wash Buffer kit according to manufacturer's instructions. Arrays were scanned at 5 mm resolution using the GenePix4000B scanner (Axon Instruments, Molecular Devices Corp., Sunnyvale, CA). Data were extracted from scanned images using NimbleScan 2.5 extraction software (Roche NimbleGen, Madison, WI, USA), which allows automated grid alignment, extraction, normalization, and export of data files. Normalized files were used as input for the Nexus 6.1 Software (Biodiscovery, El Segundo, CA, USA). Analysis settings for data segmentation and calling were the following: significant threshold for FASTST2 Segmentation algorithm: 1.0E-7, Max Continuous Probe Spacing: 1000, Min number of probes per segment: 10, high level gain: 0.485, gain: 0.17, loss: 0.2, homozygous copy loss: 0.485. Hierarchical clustering was done using Nexus 6.1 using average linkage setting. Interval files from each of the 18 samples were exported and converted to a BedGraph format and merged (using the Merge BedGraph files tool (v 0.1.1, http://galaxy.sb-roscoff.fr) in order to define common interval between samples. Intervals smaller than 2 MB were removed resulting in a set of 268 intervals with

CNA changes in at least one sample. This file was used as input of the MoCluster algorithm.

mRNA and miRNA expression profiling

Biotinylated cRNA were prepared according to the Affymetrix IVT Express protocol from 100 or 200 ng total RNA and hybridization was done as follows. CRNA were fragmented, 12 mg hybridized for 16 h at 45 C, washed and stained in the Affymetrix Fluidics Station 450 with Hybridization Wash & Stain kit. GeneChips were scanned using the Affymetrix GeneChip Scanner 3000 7G. Raw feature data were normalized using Robust Multi-array Average (RMA) method (R package affy). All subsequent analyses were performed on normalized datasets. To determine genes differentially expressed between sublines we used sam_multiclass command with one class for two biological duplicates (samr R package median). Lines containing identical GeneSymbol were collapsed with a max function. This file containing expression values for 18 samples and 2063 genes was used as input in MoCluster.

Preparation and processing of RRBS libraries

RRBS libraries were prepared as previously described (Auclair et al, 2014). Briefly, genomic DNA was digested for 5 h with MspI (Thermo Scientific) followed by end-repair, A- tailing (with Klenow fragment, Thermo Scientific) and ligation to paired-end methylated adapters (with T4 DNA ligase, Thermo Scientific) in Tango 1X buffer. We purified fragments in the range 150 to 400 bp by electrophoresis on a 3% (w/v) agarose 0.5X TBE gel with the MinElute gel extraction kit (Qiagen), and performed two rounds of bisulfite conversion with the EpiTect kit (Qiagen). RRBS libraries were generated with PfUTurbo Cx hotstart DNA polymerase (Agilent) and indexed PE Illumina primers using the following PCR conditions: 95°C for 2 minutes, 12 to 15 cycles (95°C for 30 s, 65°C for 30 s, 72°C for 45 s), 72°C for 7 minutes. The libraries were purified with AMPure magnetic beads (Beckman Coulter) and sequenced (2 × 75 bp) on an Illumina HiSeq2000 by Integragen SA (Evry, France) to generate between 20 and 30 million pairs of reads per sample. The processing of reads was performed as described (Auclair et al, 2014). We aligned reads to the human genome (hg19) with BSMAP and only retained the CpGs sequenced at least 8X.

To determine differentially methylated regions (DMRs) between sublines, we restricted the analysis to CGI comprising at least 5 contiguous CpGs and filtered for methylation differences of at least 0.2 between any of the samples. The frequency of the methylation levels was calculated and displayed as histogram in 40 classes to evaluate whole genomic variation of the DNA methylation (figure 3A). Then we used sam_multiclass function (samr R package median FDR= 0.0449). The result was a file of 892 DNA segments that was used as an input for MoCluster. Accordingly, we explored global DNA methylation variation genome-wide. In the expression correlation analysis we selected DMRs close to

TSS (+/- 1000 bp to TSS).

MoGSA

To integrate the omics data of different origins (CNA, mIR and Mrna and DNA Methylation), we used the MoGSA package (Meng, 2017a,b) to identify Joint Patterns Across Multiple Omics Data Sets. Mogsa have proved to be particularly efficient in term of computational time and compared favorably to icluster (Meng et al., 2016). We used consensus PCA from the MoGSA R package and displayed the results from the first and second principal component either for each omic dataset (CGH, DNA, mRNA or miR) or all together. The features with highest coefficient in the definition of the first axis of the PCA were selected and submitted to unsupervised Ward clustering and presented as heatmaps. Features varied according to the analysis and were as follows;
Coefficient higher or equal to 0.07 in Fig 2C, coefficient higher or equal to 0.05 in Fig 3C, coefficient higher or equal to 0.07 in Fig 4b, coefficient higher or equal to 0.06 in Fig 4D.

Gene expression dependent of gene copy numbers

To discover genes with gene copy number dependent expression, we calculated the Spearman correlation between expression levels and gene copy numbers (including gene located in aberration of less than 2 Mb). Genes with Spearman correlation higher than 0.47 ($p<0.05$) were kept (2182 genes). Then, two comparisons were done $R2^{shp53\text{-}RAS}$ vs. (R2+ $R2^{shp53y)}$ and $R2^{shp53\text{-}CCNE1}$/$R2^{shp53\text{-}WNT1}$ vs. (R2+ $R2^{shp53y).}$

Genes with copy number changes and concordant expression changes were considered as reflecting expression dependent of gene dosage. The following thresholds in log2 scale were used (copy number change <-0.15 or > 0.15; expression change <-0.5 or expression change >0.5; uncorrected t-test for expression data <0.05).

For the $R2^{shp53\text{-}CCNE1}$/$R2^{shp53\text{-}WNT1}$ comparison 156 genes (64 overexpressed and gained and 92 underexpressed and lost) were identified. Among them 34 were annotated (see below) with function related to cancer (including mapping in common regions of amplification).

For the $R2^{shp53\text{-}RAS}$ comparison 90 genes (36 overexpressed and gained and 54 underexpressed and lost) were detected.

Gene expression modulated by DNA methylation

To identify genes with DNA methylation dependent expression, we restricted our analysis to TSS +/- 1000 bp. Spearman correlation between expression level and DNA methylation was calculated. Using a threshold of -0.41 ($p<0.05$), we found 823 DNA regions corresponding to 85 genes. As in the expression/gene dosage analysis two comparisons were performed $R2^{shp53\text{-}RAS}$ vs. (R2+ $R2^{shp53y)}$ and $R2^{shp53\text{-}CCNE1}$/$R2^{shp53\text{-}WNT1}$ vs. (R2+ $R2^{shp53y).}$

We restricted our analysis to methylation variation (+ or -0.2) and expression variation of a least 2fold (log2=1). This revealed 15 DNA regions corresponding to 7 genes.

Annotation

To annotate the gene with expression submitted to gene dosage, we compiled several sources. For Oncogene column, we merged the Oncogene list from GSEA site and the oncogene list from Vogelstein (Vogelstein et al., 2013). For TSGs, we merged the TSG list from GSEA site and oncogene list from Vogelstein (Vogelstein et al., Science, 2013). For cancer genes, we compiled the driver lists from too large studies encompassing more than 2000 breast cancer cases (Stephens, 2012; Pereira, 2016). Finally, we merged the cancer gene, oncogene and TSG lists in a single Cancer Genes list. For amplicons we used the list of 30 amplicon (encompassing 1,747 genes) described by Nikolsky and coworkers (2008) and frequency of amplification from Ciriello and coauthors (2013).

Availability of genomic data

Raw array and RRBS data have been deposited on GEO and can be accessed at GSE114849 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114849 using the token cjarcymuxlmtvif. R code used for production of Figures 2 and 3 is available upon request stanislas.dumanoir@inserm.fr.

References

Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013 Oct;45(10):1127-33. doi: 10.1038/ng.2762. PubMed PMID: 24071851; PubMed Central PMCID: PMC4320046.

Meng C. (2017). *mogsa: Multiple omics data integrative clustering and gene set analysis*. R package version 1.11.0.

Chen Meng, Bernhard Kuster, Bjoern Peters, Aedin C Culhane, Amin Moghaddas Gholami. moGSA: integrative single sample gene-set analysis of multiple omics data. bioRxiv 046904; doi: https://doi.org/10.1101/046904

Meng C, Helm D, Frejno M, Kuster B. moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets. *J Proteome Res* 2016; **15**: 755–765.

Nikolsky Y nad al Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res. 2008 Nov 15;68(22):9532-40. doi:10.1158/0008-5472.CAN-08-3082. PubMed PMID: 19010930.

Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ, Tsui DW, Liu B, Dawson SJ, Abraham J, Northen H, Peden JF, Mukherjee A, Turashvili G, Green AR, McKinney S, Oloumi A, Shah S, Rosenfeld N, Murphy L, Bentley DR, Ellis IO, Purushotham A, Pinder SE, Børresen-Dale AL, Earl HM, Pharoah PD, Ross MT, Aparicio S, Caldas C. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016 May 10;7:11479

Stephens PJ, and al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012 May 16;486(7403):400-4.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58. doi: 10.1126/science.1235122.

uncorrected t-test for expression data <0.05).

Legends to Supplementary Figures and Tables

Supplementary Figure S1: Gradual immortalization and transformation of the shp53 HMEC sublines. A: b-galactosidase staining**. B-C:** HTERT mRNA expression and telomerase enzymatic activity, the shp53 HTERT cells are presented as a positive control of TERT expression**. D:** anchorage independent growth in soft agar, number of positive experiments out of number of attempts, pictures of the corresponding soft agar petri dish and blow up of foci formed by the respective sublines. **E:** shp53 sublines did not form tumor upon injection in immunocompromised mice. First line indicates the total number of cells injected into the interscapular fat pad. Statistical significance was determined by the Anova Dunn's Multiple Comparison test. Error bars correspond to s.e.m.

Supplementary Figure S2: Attenuation of p53 after shRNA transduction and overexpression of oncogenes in oncogene transduced sublines. A: attenuation of p53 was ascertained by challenging primary R2 and R2-shp53 cells with Bleomycin for 6 hours, accumulation of p53 and induction of p21 were used as read outs for p53 functionality. **B:** QPCR verification of WNT1, CCNE1 and RAS mRNA expression in R2-shp53-WNT1, R2-shp53-CCNE1 and R2-shp53-RAS respectively. **C:** protein expression levels by western blotting. E: early; L: Late.

Supplementary Figure S3: Spontaneous DNA damage in shp53 sublines. A: gammaH2Ax and 53BP1 staining patterns, note the pannuclear H2Ax staining in shp53 HMECs, whereas shp53-WNT1, shp53-CCNE1 and shp53-RAS show predominantly H2Ax foci. Hydroxyurea treatment was used as a control of H2Ax staining as a condition of severe replication stress. It is of note that primary R2 HMEC presented a sizeable number of foci positive cells, which can be attributed to telomere attrition in these cells. Bleomycine treatment was used as a control of double strand breaks. **B:** fraction of cells showing H2Ax pan-nuclear staining or more than 4 foci. **C:** Tail moment measurements box plot. **D:** Statistics of Neutral CometAssay tail moment measurement for 3 independent experiments in shP53 and shp53-oncogene sublines (top table) and Anova multiple comparison test (bottom table). shP53-CCNE1 and shP53-WNT1 sublines present a significantly higher number of double strand breaks than shP53-RAS.

Supplementary Figure S4: HMEC models form 3 clusters based on the oncogene transduced.

Supplementary Figure S5: CNA plots of HMEC models. A: CNA at different steps of transformation. CNAs are represented for each chromosome, red for losses, blue for gains. The hight of the bars indicates the probability of occurrence. **B:** cumulated CNA plots of HMEC models. CNAs are represented for each chromosome, red for losses, blue for gains. The hight of the bars indicates the amplitude of the copy number change.

Supplementary Figure S6: density histograms of RRBS methylation scores at CpGs sites (at least 5 contiguous CG) genome wide in primary R2 HMECs, shp53.Early, shp53.Late, shp53-CCNE1, shp53-CCNE1.SA (Soft Agar), shp53-WNT1, shp53-WNT1.SA, shp53-RAS, shp53-RAS.SA

Supplementary Figure S7: pathways and regulation networks principally affected in the different HMEC sublines.

Supplementary Figure S8: phenotypic characteristics of shp53 HMEC sublines. Cells were stained by immunofluorescence (IF) for the expression of ECAD (E-Cadherin, green) which is a marker of epithelial cells, VIM (Vimentin, red) marker of mesenchymal cells, CK8 (Cytokeratin 8) marker of luminal breast epithelial cells, CK5 (Cytokeratin 5) marker of basal breast epithelial cells. Differential expression patterns can be observed according to the genetic elements expressed and the stage of the culture. Normal HMEC and Early shp53 co-express ECAD and VIM and are mosaiec for CK5 and CK8 expression. In shp53 late HMEC tended to lose ECAD expression and become mesenchymal, but kept a mosaic CK5/CK8 pattern. In shp53-WNT1 and shp53-CCNE1 ECAD and VIM were co-expressed in all cells indicating the conservation of an epithelial phenotype. Interestingly, whereas shp53-WNT1 expressed only CK5 and no CK8, shp53-CCNE1 preserved the original mosaic phenotype of the shp53 HMECs. Expression of RAS-v12 produced drastic changes as illustrated by the concomitant loss of expression of ECAD and both CK5 and CK8.

Supplementary Figure S9: Genes modified by CNA or differential methylation in Shp53^{-RAS} and Shp53^{-CCNE1}/ Shp53^{-WNT1} show little overlap. Of note MTUS1 is strongly underexpressed in RAS (–2,95; about 1/10 x) whereas it is overexpressed in CW (0,87, about 1,8 x). PLAG1 is strongly overexpressed in RAS (3,25; about 9,5 X) and moderately in CW (0,84 about 1,8 x) similarly to THRA (RAS change =1,89 about 3,7x; CW change =0,89 about 1,85x). ZNF7 expression change is equivalent in both clusters.

Supplementary Figure S10: HMEC models present a high luminal progenitor and luminal mature score and Basal-like and claudin-low breast cancer resemble shp53-CCNE1/WNT1 and shp53-RAS HMEC models respectively. A: HMEC models were classified according to breast cancer molecular subtypes using the PAM50 classifier and correlate with the Basal-like subtype. **B-C:** principal pathways activated in shp53-RAS models highlighting the importance of the RAS and EMT pathways**. D:** Fraction of breast tumors showing a p53 mutation (blue bars), p53 mutation and RAS overexpression (orange bars, p53 mutation and CCNE1 overexpression (grey bar) in 10 Inclust molecular subgroups. **E:** radar plot of the mammary differentiation scores as defined by E. Lim et al (2009)

Supplementary Figure S11: RAS transformation is not restricted to phenotypic subset, whereas WNT1 preferentially transforms epithelial of shp53 HMEC. A: experimental scheme for the isolation

of epithelial and mesenchymal shp53 HMEC clones. **B:** FACS analysis by means of the epithelial CD24 and the mesenchymal CD44 markers of the different shp53 clones and derived shp53-RAS and shp53-WNT1 cells. **C:** RT-QPCR quantification of CDH1 and VIM mRNA expression levels. **D:** ZEB1 and ZEB2 mRNA expression levels. Expression levels were normalized on those measured in the primary R2 HMECS. Error bars correspond to s.e.m.

Supplementary Table S1A: genes with copy number dependent expression changes in the shp53-ras cluster vs. The shp53 and R2 cluster. Genes were selected after a Spearman correlation test on the complete dataset and 90 genes from this comparison corresponded to genes with log2 scale changes (copy number change <-0.15 or > 0.15; expression change <-0.5 or expression change >0.5; uncorrected t-test for expression data <0.05)

Supplementary Table S1B: genes with copy number dependent expression changes in the shp53-ccne/wnt cluster vs. The shp53 and R2 cluster. Genes were selected after a Spearman correlation test on the complete dataset and 156 genes from this comparison corresponded to genes with log2 scale changes (copy number change <-0.15 or > 0.15; expression change <-0.5 or expression change >0.5; uncorrected t-test for expression data <0.05).