



HAL
open science

Distinct oncogenes drive different genome and epigenome alterations in human mammary epithelial cells

Claire Fonti, Anne Saumet, Amanda Abi-Khalil, Béatrice Orsetti, Elouan Cleroux, Ambre A. Bender, Michael Dumas, Emeline Schmitt, Jacques Colinge, William Jacot, et al.

► **To cite this version:**

Claire Fonti, Anne Saumet, Amanda Abi-Khalil, Béatrice Orsetti, Elouan Cleroux, et al.. Distinct oncogenes drive different genome and epigenome alterations in human mammary epithelial cells. *International Journal of Cancer*, 2019, 145 (5), pp.1299-1311. 10.1002/ijc.32413 . inserm-02177754

HAL Id: inserm-02177754

<https://inserm.hal.science/inserm-02177754>

Submitted on 9 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distinct oncogenes drive different genome and epigenome alterations in human mammary epithelial cells

Claire Fonti^{#1}, Anne Saumet^{#1}, Amanda Abi-Khalil¹, Béatrice Orsetti^{1,2}, Elouan Cleroux³, Ambre Bender³, Michael Dumas³, Emeline Schmitt¹, Jacques Colinge¹, William Jacot², Michael Weber³, Claude Sardet¹, Stanislas du Manoir¹, Charles Theillet^{1,2*}

Claire Fonti and Anne Saumet contributed equally to this work

* Correspondence to Charles Theillet, IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université de Montpellier, Montpellier, F-34298, France. charles.theillet@inserm.fr

Author affiliations

1. IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université de Montpellier, Montpellier, F-34298, France
2. ICM, Institut Régional du Cancer de Montpellier, Montpellier, F-34298, France
3. CNRS, University of Strasbourg, UMR 7242 Biotechnology and Cell Signaling, 300 Bd Sébastien Brant, CS BP 10413, 67412 Illkirch, France

Conflict of interest: The authors declare they have no potential conflicts of interest

Keywords: breast cancer, genome, modifications, oncogene, cell type

Abbreviations: HMEC; human mammary epithelial cells, CGH; comparative Genomic Hybridization, CNA; copy number aberrations, mRNA; messenger RNA, miR: micro-RNA, RRBS: Reduced representation bisulfite sequencing

Abstract

Molecular subtypes of breast cancer are defined on the basis of gene expression and genomic/epigenetic pattern differences. Different subtypes are thought to originate from distinct cell lineages, but the early activation of an oncogene could also play a role. It is difficult to discriminate the respective inputs of oncogene activation or cell type of origin. In this work, we wished to determine whether activation of distinct oncogenic pathways in human mammary epithelial cells (HMEC) could lead to different patterns of genetic and epigenetic changes. To this aim, we transduced shp53 immortalized HMECs in parallel with the CCNE1, WNT1 and RASv12 oncogenes which activate distinct oncogenic pathways and characterized them at sequential stages of transformation for changes in their genetic and epigenetic profiles. We show that initial activation of CCNE1, WNT1 and RASv12, in shp53 HMECs results in different and reproducible changes in mRNA and miRNA expression, copy number alterations (CNA) and DNA methylation profiles. Noticeably, HMECs transformed by RAS bore very specific profiles of CNAs and DNA methylation, clearly distinct from those shown by CCNE1 and WNT1 transformed HMECs. Genes impacted by CNAs and CpG methylation in the RAS and the CCNE1/WNT1 clusters showed clear differences, illustrating the activation of distinct pathways. Our data show that early activation of distinct oncogenic pathways leads to active adaptive events resulting in specific sets of CNAs and DNA methylation changes. We, thus, propose that activation of different oncogenes could have a role in reshaping the genetic landscape of breast cancer subtypes.

Novelty and impact: Different breast cancer molecular subtypes are assumed to originate from distinct cell lineages. The authors propose that founding oncogenic mutations could also have an impact on the genetic and epigenetic landscape of the tumor. Data presented herein, based on engineered primary HMEC models, show that transformation induced by distinct oncogenes resulted in different and reproducible patterns of genetic and DNA methylation changes, indicating that early activation of distinct oncogenic insults will not only impinge on the phenotypic characteristics of the resulting tumors, but also have a strong impact on their genomic and epigenetic profiles.

Introduction

Genetic instability lies at the core of neoplastic development with up to 85% of human cancers showing loss of chromosome integrity at varying levels such as aberrant copy numbers and aneuploidy.

Chromosomal instability has been observed in early stages of cancer¹ and rearrangement intensity correlated with disease aggressiveness². In addition to structural defects, cancer genomes undergo important epigenetic changes occurring at the chromatin and DNA levels³. At the DNA level, cancer associated epigenetic modifications involve genome wide cytosine methylation changes corresponding to demethylation of repetitive DNA sequences and hypermethylation of CpG enriched sequences^{4,5}.

Genetic instability results in stochastically occurring aberrations, of which a fraction will be selected according to the survival or growth advantage they confer. Hence, in tumors profiles of somatically acquired genetic and epigenetic changes and associated RNA expression modifications reflect the combined interactions of genetic instability and selective pressure. As a consequence, recurrent profiles of genomic and epigenetic aberrations should be structured around anomalies that confer maximum advantage in a given tissue and environment. Noticeably, cancers of distinct anatomical origins exhibit quite different profiles of genomic and epigenetic anomalies⁶. Recent work, based on human open reading frame library screening in 3 different cell types to identify proliferation drivers, elegantly showed the existence of a tissue specific selection for gains of function and loss of tumor suppressors⁷.

In breast cancer, molecular subtypes were defined on the basis of RNA expression, as well as of genomic anomalies and DNA methylation differences⁸⁻¹¹. Although definitive proof is still missing, it is generally proposed that the genetic and epigenetic differences in different breast tumor subtypes are dictated by distinct cell types of origin^{8,9}. Founding events due to early activation of distinct oncogenic pathways in a single cell type could also have an impact on genomic and epigenetic changes and induce the selection of anomalies functionally coherent with the activated pathway¹²⁻¹⁵. This is supported by studies showing that expression of RASv12 and of BRAFv600 resulted in the transcriptional repression and hypermethylation of distinct gene sets, involving different cascades of repressors and DNA methylases^{16,17}. Yet, it has not been experimentally demonstrated that distinct oncogenic events could lead to specific genomic rearrangements.

In this work, we sought to determine the impact of the early activation of distinct oncogenic pathways on genomic and epigenetic changes in immortalized human mammary epithelial cells (HMECs). To this aim, we adapted a previously devised experimental scheme¹⁸ and overexpressed by retroviral transduction three oncogenes WNT1, CCNE1 and RASv12, known to activate different oncogenic pathways, in shp53 immortalized human HMECs. Epigenetic and genetic changes were then monitored at different steps of cancer progression. The sequence of genetic and epigenetic alterations accompanying the transition between the normal and transformed states show that activation of these distinct oncogenes leads to the emergence of distinct and specific profiles of changes. These results thus support a model in which genetic and epigenetic changes in cancer cells reflect adaptive responses to the oncogenic driver.

Material and Methods

HMEC models

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. After elimination of fibroblasts, HMECs were cultured at 37°C in 5% CO₂, in conditioned MEBM medium supplemented with antibiotics (*MEGM single Quots*, Lonza, Levallois-Perret, France). Primary HMECs were transduced with amphotropic retroviral supernatants corresponding to: pSUPER.retro.hygro-shp53, pBABE.neo-CCNE1, pLNC-WNT1, pBABE.puro-HRASV12 followed by 3 weeks of antibiotic selection. This project has been reviewed and approved by the Ethics committee of the University of Montpellier.

β-galactosidase senescence test

β-galactosidase activity was assayed by histochemistry using the *Senescence Cells Histochemical Staining* kit (Sigma-Aldrich, St Quentin Fallavier, France) following manufacturer's instructions. β-galactosidase positive cells were quantified under the microscope in duplicate on 400 cells minimum.

Telomeric restriction fragment (TRF) analysis and telomerase activity test and chromosome counts

TRF were purified using the *TeloTAGGG Telomere Length Assay* (Sigma-Aldrich, St Quentin Fallavier, France) following manufacturer's instructions. TRF sizes determined by 0.8% agarose gel electrophoresis and Southern blotting. The Nylon membrane was hybridized telomeric DNA probe labeled with digoxigenine and revealed with anti-DIG antibodies coupled alkaline phosphatase and chemiluminescence. Medium TRF size was calculated using $TRF = \frac{\sum(OD_i)}{\sum(OD_i/L_i)}$ (OD_i = optical density at i , L_i = size at i) Telomerase activity was measured on 2×10^5 cells with the *TeloTAGGG Telomerase PCR ELISA* kit (Sigma-Aldrich, St Quentin Fallavier, France) following manufacturer's instructions.

Anchorage independence growth (AIG)

AIG was determined in 6 well plates containing 2 layers of low melting point agarose in MEBM (Lonza) at 0.75% and 0.45% on top. 15000 cells/well were seeded and plates incubated for 5 weeks at 37°C in 5% CO₂. Colonies were visualized using 0.01% crystal violet and counted. Prior staining colonies were isolated and put in culture to generate the Soft Agar (SA) clones.

Tumorigenicity

1.5 or 5×10^6 cells resuspended in a 1:1 Matrigel/PBS solution were injected subcutaneously in Swiss nude or SCID beige mice (Harlan/Envigo, Garnat France). Each subline tested was injected in 6 animals in parallel and tumor growth monitored for 5 months before animals were sacrificed. In vivo experiments were

systematically reviewed and approved by an internal animal ethics committee and the University of Montpellier animal ethics committee.

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).

Array-CGH and mRNA expression profiling

Array-CGH was done using HG18 CGH 385K Whole Genome v2.0 array (Roche NimbleGen, Madison, WI, USA). Methods and analysis are described in detail in the Supplementary Methods.

miR expression profiling

Biotinylated cRNA were prepared according to the Affymetrix IVT Express protocol. Detailed description of the methods is provided in the Supplementary Methods.

DNA methylation profiling

RRBS libraries were prepared as previously described¹⁹. Detailed description of the experimental method and DMR selection approach are available in the Supplementary Methods.

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. More detailed description is available in the Supplementary Methods.

Immunofluorescence

Cells were seeded onto glass slides and grown to reach 50-60% confluence. Prior immunofluorescence cells were fixed with either 2% paraformaldéhyde or with ice cold methanol and permeabilized with 1% PBS-Triton and rinsed with 2% serum-PBS before incubation with the primary antibody, being rinsed in 2% serum-PBS and stained with DAPI and incubated with the secondary antibody. γ H2AX/53BP1 foci were counted in triplicate on 400 nuclei. Antibodies are listed in the S Appendix.

High throughput data analysis

Detailed methods are presented in the Supporting Information 1. Raw array and RRBS data can be accessed at GSE114849 (see Supporting Information Supplementary Methods).

Results

Establishment of stepwise transformation cell models

Primary HMEC were isolated from fresh mammary tissue obtained from donors undergoing reductive plastic surgery (Supplementary Methods). We obtained 3 stable cell cultures that were propagated for at least 15 passages before cells started showing signs of senescence (reduced proliferation, large vacuoles and positive β -galactosidase staining) (Supporting Information Fig.1A). Of the 3 primary HMEC lines, we selected the R2 line to establish our models of stepwise cell transformation. Cells were genetically modified by sequential retroviral transductions with defined genetic elements (Fig.1A). In the first step (immortalization), we transduced a vector expressing an shRNA targeting the *TP53* gene (designated shp53 hereafter), which plays a key role in the senescence barrier. The shp53 efficiently knocked-down p53 protein expression and signaling (Supporting Information Fig.2) and cells grew rapidly. We derived three stable shp53 HMEC sublines that were used as models hereafter. The shp53 sublines showed reduced β -galactosidase staining, re-expression of the endogenous *hTERT* and stabilization of telomere length, indicating that the senescence program had been overcome spontaneously (Fig.1B-C, Supporting Information Fig.1B-C). This was in concordance with reports of p53 acting as a transcriptional repressor of *hTERT*^{20; 21} and of the spontaneous reactivation of telomerase in p53-nul HMECs²². In the second step, two of these shp53 sublines were independently transduced with vectors expressing either *CCNE1*, *WNT1* or *HRAS*^{v12}, 3 oncogenes belonging to distinct signaling pathways. Several independent sublines were derived for each oncogenic situation (shp53-WNT1, shp53-CCNE1, shp53-RAS). Each subline was then seeded in soft agar to determine anchorage independent growth as a standard read out of *in vitro* transformation (Supporting Information Fig.1D). Cell clones were isolated from soft agar foci and expanded (shp53-WNT1.SA, shp53-CCNE1.SA, shp53-RAS.SA). In total, we established 16 HMEC sublines (see Supplementary methods) corresponding to different steps of cancer transformation; immortalized (shp53), pre-transformed (shp53-oncogene) and transformed (shp53-oncogene.SA). However, it is of note that none of the shp53-oncogene.SA formed tumors in immunocompromised mice even after injection of up to 5×10^6 cells (Supporting Information Fig.1E).

Genetic instability in the immortalized and transformed cell lines

We first analyzed the different HMEC sublines by array-CGH. Using the fraction of the genome involved in copy number alterations (CNAs) as a metric, we noted that the level of genetic instability gradually increased from 2 to 6% in the immortalized shp53 to 12 to 13% in the shp53-WNT1, shp53-CCNE1 and shp53-RAS and reached up to 22 and 27% in shp53-WNT1.SA and shp53-CCNE1.SA, respectively (Fig.1D). Interestingly, patterns of CNA progression were coherent with ploidy changes. Metaphase spreads of shp53 and shp53-RAS sublines showed 92 chromosomes indicating tetraploidy, while, shp53-CCNE1 and shp53-WNT1 sublines became aneuploid showing 78 to 110 chromosomes (Fig.1E). To determine whether increased CNA levels were associated with increased genetic stress, we performed immunofluorescence staining of γ H2Ax and 53BP1 in R2, shp53 and shp53-oncogene HMECs (Supporting Information Fig.3A-B).

Two types of γ H2Ax staining patterns were observed; nuclear foci, considered as markers of DNA breaks, and pan-nuclear staining, which has been proposed to reveal widespread replication stress in the absence of double strand breaks²³. Immortalized shp53 HMECs predominantly showed pan-nuclear γ H2AX staining and low levels of γ H2Ax and 53BP1 foci, suggesting an elevation of genetic stress but low levels of DNA breaks in these cells. This contrasted with shp53-CCNE1 and shp53-WNT1, which essentially displayed γ H2Ax and 53BP1 nuclear foci (Supporting Information Fig.3A-B), consistent with a significant increase of DNA breaks confirmed by comet-assay (Supporting Information Fig.3C-D). Of note, shp53-RAS showed distinctly lower levels of γ H2Ax and 53BP1 nuclear foci and low levels of DNA breaks by comet-assay relative to shp53-CCNE1 and shp53-WNT1.

Immortalized and transformed HMECs form 3 distinct clusters defined by the activated oncogene

Next, we determined whether the different immortalized and transformed HMEC sublines presented distinct profiles of genetic and epigenetic anomalies. Copy Number Alterations (CNA), DNA Methylation, miRNA (miR) and mRNA expression levels were determined at different steps of cell immortalization and transformation. To integrate these data corresponding to diverse molecular features and different technological platforms, we used moCluster (moGSA package), allowing to produce integrative clustering on multiple omics data. This approach is based on multivariate latent variable decomposition to discover correlated global variance structure across datasets²⁴.

Two independent biological replicates of each immortalization or transformation step, (R2, shp53, shp53-WNT1, shp53-CCNE1, shp53-RAS and shp53-WNT1.SA, shp53-CCNE1.SA, shp53-RAS.SA HMECs) were included in this analysis. These analyses pointed at similarities and differences between distinct sublines and stages of transformation. These were determined by means of Principal Component Analysis (PCA). Changes defining the most variant PCA vector were subsequently analyzed by Ward clustering. The combined analysis of the four datasets (CNA, DNA methylation, mRNA and miRNA expression) defined 3 clusters (Fig.2A). The first one, at the trunk, organized around R2 and shp53 early and late passages. The second corresponded to the four shp53-RAS sublines. The third cluster included shp53-WNT1 and shp53-CCNE1 sublines (Fig.2A). The number of clusters was confirmed by Nbclust (Supporting Information Fig.4 and Supplementary methods). Altogether, these analyses revealed that the HMEC sublines transduced with different genetic elements showed clear differences in their genetic and epigenetic patterns. We next analyzed CNAs, DNA methylation, mRNA and miRNA datasets individually to verify whether pattern differences applied similarly to all datasets.

Overexpression of CCNE1, WNT1 or HRAS^{v12} oncogenes in immortalized HMECs result in distinct profiles of copy number alterations

Genomic regions involved in CNAs were identified with the Nexus 7.5 Software (Biodiscovery, CA., USA) and only CNAs covering at least 2 Mb were used in the moGSA analysis. PCA classification of HMEC sublines singled out three clusters. Cluster 1 positioned at the trunk encompassed R2 and shp53 replicates, cluster 2 formed by the shp53-RAS and shp53-RAS.SA sublines and a more dispersed cluster 3 comprising shp53-

WNT1, shp53-CCNE1, shp53-WNT1.SA and shp53-CCNE1.SA (Fig.2B). The distance along the PC1 vector in the principal component analysis, separating cluster 2 from cluster 3, illustrated the strong differences at the CNA level between the RAS and the WNT1 or CCNE1 transformed sublines (Fig.2B).

A clustering analysis of the CNAs defining the PC1 vector identified co-occurring loss of chromosome 4, loss at 8p and gain at 8q as the most significant anomalies characterizing the shp53-RAS HMECs (Fig.2C). Other events were gains at 12q and losses at 14 and 18q (Supporting Information Fig.5A-B). In clear contrast, shp53-WNT1 and shp53-CCNE1 HMECs were characterized by losses at chromosomes 2, 3 and 6, as well as focal gains at 11q13 and 20q13 (Fig.2C and Supporting Information Fig.5A-B). Interestingly, whereas the pre-transformed shp53-RAS showed little difference with their transformed shp53-RAS.SA counterpart, significant deviation was detected between shp53-CCNE1 and shp53-CCNE1.SA, as well as between shp53-WNT1 and shp53-WNT1.SA, with additional gains at chromosome 11q13-q14 and 20q11-q13, respectively (Fig.2C). These data indicated increased genetic instability in shp53-CCNE1.SA and shp53-WNT1.SA, whereas this was not the case when shp53-RAS were compared with shp53-RAS.SA (Fig.1D and Fig.2C).

Overexpression of CCNE1, WNT1 or HRAS^{v12} oncogenes in immortalized HMECs result in distinct profiles of DNA methylation

DNA methylation profiles of the HMEC sublines were analyzed by reduced representation bisulfite sequencing (RRBS). Variations in methylation levels were quantified genome-wide at sequences containing at least 5 contiguous CpG pairs. A global increase and change in distribution of the methylation pattern was detected between primary R2 HMECs and all the sublines derived there from (Supporting Information Fig.6). Variations were particularly pronounced in transformed cells after soft agar cloning (shp53-oncogene.SA). Using a methylation difference of 0.2 as inclusion criterion, we next performed a SAMR analysis of these data to identify differentially methylated regions (DMRs) that varied most significantly between sublines. Identified DMRs were analyzed in moGSA and used to classify cell lines by Principal Component Analysis (PCA). The resulting PCA classification was remarkably similar to that generated with CNAs, as shp53-RAS and shp53-RAS.SA defined a clearly distinct cluster of the shp53-CCNE1, shp53-WNT1, shp53-CCNE1.SA and shp53-WNT1.SA cluster (Fig.3A). Additional clustering analysis of the DMRs centered on CpGs methylation patterns, confirmed this trend, shp53-RAS HMECs showing an almost binary difference when compared to shp53-CCNE1 and shp53-WNT1 (Fig.3B). Noticeably, primary R2 and immortalized shp53 HMECs (Early and Late passages) clustered at the trunk (cluster 1 in Fig.3A). Altogether, both CNA and DNA methylation profiles indicate that an oncogenic form of *RAS* induces specific genomic and epigenetic alterations, that markedly differ from those induced by WNT1 or CCNE1, whose profiles co-clustered in both classifications.

Overexpression of CCNE1, WNT1, or HaRAS^{v12} oncogenes in Shp53 HMECs result in distinct transcriptional programs and phenotypical cell fates.

As observed with CNA and DNA methylation profiles, cluster analyses of both miR and mRNA expression profiles clearly distinguished the RAS and CCNE1/WNT1 transformed cells in two separate clusters (Fig.3C

and Fig.3E). Many genes presented an inverted pattern of expression in the RAS compared with the CCNE1/WNT1 sublines in the Ward clustering analysis of miR and mRNA expression changes (Fig.3D and Fig.3F). GeneGo Metacore (Thomson Reuter) network analyses of the miR and mRNA gene lists that discriminated cluster 2 and cluster 3 identified several differentially activated functional gene networks. The miR data pointed at strong differences in EMT-associated pathways, illustrated by the level of miR200c, a key regulator of the EMT transcription factor ZEB1²⁵, and the mesenchymal cell-specific miR143/miR145, that differed markedly between RAS and CCNE1/WNT1 cells²⁶ (Supporting Information Fig.7). Morphological features (cuboidal vs. fusiform) and immunostaining of sublines with epithelial E-Cadherin (ECAD) and mesenchymal Vimentin (VIM) markers, confirmed that primary R2, immortalized shp53, and transformed shp53-CCNE1 and shp53-WNT1 cells all maintained an epithelial phenotype (cuboid, ECAD^{high}, VIM^{low}), while shp53-RAS cells were fusiform, ECAD-negative and VIM^{high} indicating a clear mesenchymal conversion (Supporting Information Fig.8).

Chronology of the genetic and epigenetic changes occurring during transformation of HMECs.

Monitoring the genetic and epigenetic changes occurring at each step of the cell transformation process allowed us to reconstruct their chronology. This could be inferred from the PC2 vectors in the PCA classifications. Starting from normal R2 primary cells, moving to early and late immortalized cells and progressively to soft agar subclones, we noted that the different sublines bore different positions in these classifications according to the dataset considered. Indeed, the CNA classification indicated that R2 coclustered with early and late passage shp53 immortalized cells (Fig.2B), while that based on miR indicated that primary R2 were clearly distinct from all the other sublines (Fig.3C). In contrast, the mRNA and DMR based classifications showed that primary R2 co-clustered with early passage shp53 cells, whereas late passage shp53 were positioned closer to shp53-CCNE1 and shp53-WNT1, indicating a shift between early and late passages shp53 HMECs (Fig.3A and 3E).

These results suggest a temporal hierarchy of the genetic and epigenetic events that primed transformation in our models. Upon p53 inactivation in HMECs, miR expression appeared to be the first modified, whilst the modification of mRNA expression profiles and massive modifications of the methylation landscape occurred later. Upon transduction with RAS, WNT1 or CCNE1, additional DNA methylation changes occurred concomitantly with modifications in mRNA expression. Copy number changes emerged last and underwent a clear selection process during soft agar cloning (Fig.2B and Supporting Information Fig.5A).

Impact of copy number changes and differential DNA methylation on gene expression.

Next, we identified the genes in the RAS or the CCNE1/WNT1 clusters whose expression change correlated with CNAs and showed differential expression in comparison with R2 and shp53 HMECs. Ninety (90) genes (36 overexpressed with copy number gains and 54 underexpressed and copy loss) fitted these criteria in shp53-RAS HMECs and 156 genes (64 overexpressed and gained, 92 underexpressed and lost) in shp53-CCNE1/shp53-WNT1 (Supporting Information Table1). Noticeably, only 4 genes (1.7%) were found in both

the shp53-RAS and in the shp53-CCNE1/shp53-WNT1 lists (Supporting Information Fig.9A). Such a small overlap could not have happened by chance (hypergeometric test, $p=0.75$).

We selected annotated cancer genes (oncogenes and tumor suppressors) or genes belonging to the RAS pathway (Fig.4A and Supporting Information Table 1). In the shp53-RAS cluster, we identified 5 known oncogenes that were gained and over-expressed (*ERBB2*, *HEY1*, *PLAG1*, *ZBTB10*, *CRTC3*), and 6 genes lost and under-expressed. Of these 6 genes, 4 were part of the RAS pathway (*TMEM154*, *LRAT*, *AKAP6*, *MTUS1*) and 2 were tumor suppressors (*FBXW7*, *SERPINB5*). *FBXW7* is a negative regulator of oncogenic transcription factors such as MYC, FOS or NOTCH²⁷, whereas *SERPINB5* (*MASPIN*) is described as a metastasis suppressor in breast and other cancers²⁸.

In the shp53-CCNE1/shp53-WNT1 cluster, we identified 13 genes, 5 gained and over-expressed, 8 lost and under-expressed (Fig.4A), all of which classified as cancer genes. Of the 5 gained and over-expressed genes, *NUMA1*, *PMS2* and *CUX1* have well documented roles in mitotic spindle assembly or DNA repair and *STAT3* is a key signaling node for a number of growth factors and cytokines frequently involved in cancer²⁹. *PLAG1* was selected both in the RAS and CCNE1/WNT1 upregulated genes, but its expression level was 9 times higher in shp53-RAS compared to shp53-CCNE1/shp53-WNT1. Of the 8 lost and under-expressed genes, we noted *LATS1*, negative regulator of the Hippo pathway, *CTNNB1* (β -catenin), *FYN* homolog of *SRC* and *TGFBR2*.

Using a similar approach, we also searched for genes whose expression was impacted by CpG methylation. We restricted our analysis to DMRs located close to transcription start sites (TSS) of annotated genes showing at least a 20% change in methylation, associated to a 2fold variation in mRNA expression level (Fig.4B). Again, genes modified by DNA methylation in the shp53-RAS cluster showed little overlap with those in shp53-CCNE1/shp53-WNT1 HMECs (Hypergeometric test, $p = 0.66$, Supporting Information Fig.9B). Of the 17 genes hypermethylated and underexpressed in the shp53-RAS cluster, 7 were key in cell-adhesion and the epithelial phenotype (*KRT5*, *ITGB4*, *DMKN*, *PKP3*, *ACP*, *PROM2*, *KDF1*). Interestingly, 4 genes were hypomethylated and overexpressed, of which 2 (*ITGB1*, *FRMD4A*) corresponded to genes involved in cell invasion and metastasis^{30,31}. In the shp53-CCNE1/shp53-WNT1 HMECs, 16 genes all hypermethylated and downregulated were identified (Fig.4B), including 4 genes related to cell invasion or EMT (*ANGPTL4*, *ITGB3*, *EHD3*, *FBN2*). Interestingly, *ITGB3* and *EHD3*, interact physically in an activating regulatory loop.

Altogether, the RAS and the CCNE1/WNT1 transformed HMECs showed clear differences in genes whose expression was impacted by either CNAs or CpG methylation. The principal features in RAS HMECs were the repression of genes associated with the epithelial phenotype and cell adhesion and conversely the activation of genes favoring cell invasion. In clear contrast, CCNE1/WNT1 HMECs showed a downregulation of EMT or invasion associated genes, combined with the activation of DNA repair and cell division genes. These results indicate the activation of distinct pathways in the respective sublines, with an opposite trend concerning cell phenotype, pro-mesenchymal in shp53-RAS and pro-epithelial in shp53-CCNE1/shp53-WNT1.

RAS and CCNE1/WNT1 HMECs resemble claudin-low and Basal-like breast cancer, respectively

We wanted to determine how much our HMEC models paralleled with human breast cancer. Our 16 HMEC models were classified as Basal-like breast cancer according to the CIT and PAM50³² classifiers (Supporting Information Fig.10A). Noticeably, shp53-RAS were classified as Claudin-low and shp53-CCNE1 and shp53-WNT1 as epithelial basal-like (Fig.5A). Next, we sought to compare the characteristics of our HMEC models with that of the Claudin-low and the Basal-like subgroups in the human primary breast cancer METABRIC dataset. Restricting our analysis to *TP53* mutated and Basal-like of the invasive ductal carcinoma type, we selected 225 tumors, which we stratified in Claudin-low/Basal-like (63) and Basal-like/Non-Claudin-low (162) subsets. We determined that the KRAS pathway was among the top activated pathways in Claudin-low/Basal-like tumors (Supporting Information Fig.10B-C), and that Basal-like/Non-Claudin-low tumors presented a significantly higher incidence of CNAs and increased levels of *CCNE1* expression (Fig.5B-C-D). These data suggested that the differences in genomic profiles induced by distinct oncogenes observed in our HMEC models mimicked situations occurring in spontaneous human breast tumors: fewer genomic changes and elevated RAS in mesenchymal like claudin-low basal tumors, in comparison to non-claudin-low basal tumors which show more rearrangements and frequent *CCNE1* overexpression.

Discussion

Breast cancer can be broken down in at least 5 molecular subtypes on the basis of differences in mRNA expression and genetic anomaly profiles^{9,10}. It is generally assumed that the origin of these molecular subtypes lies in cell lineage differences, where the original tumorigenic insult took place^{8,9}. It is, however, of note that basal-like tumors from BRCA1 mutated patients were shown to originate from luminal progenitor cells^{33,34}, which undergo a differentiation shift from luminal to basal cells, giving rise to basal-like rather than luminal tumors³⁵. Similarly, expression of an activated *PikcaH1047R* allele in committed unipotent luminal cells in mouse mammary glands induced cell fate reprogramming and emergence of basal like tumors³⁶. Furthermore, mammary tumors from genetically modified mice were shown to bear different CNA profiles according to the driver mutation that initiated the tumor³⁷. Altogether, these data suggest that founding oncogenic mutations can also impinge on the genetic profile of breast cancers. Using primary HMECs, that classified as luminal progenitors³³ (Supporting Information Fig.10E), we show that cell transformation by way of distinct oncogenes resulted in different patterns of aberrations at both the CNA and DNA methylation levels, as well as distinct phenotypes. Most remarkably, HMECs transformed by RASv12 (shp53-RAS) presented clearly distinct patterns of genetic and epigenetic modifications compared to their shp53-CCNE1 or shp53-WNT1 counterparts. The latter two exhibited globally similar CNA and DNA methylation profiles, albeit some focal differences could be found.

In our model system, inactivation of the *TP53* gene was the initial step towards transformation. Given the role of *TP53* in genome integrity, onset of genetic instability was expected in shp53 HMECs, but no gross genomic anomalies were observed, even after more than one year in culture. However, shp53 HMECs rapidly became tetraploid, in line with the critical role of p53 in ploidy control³⁸. Tetraploidy is considered as a prelude to large scale chromosomal aberrations in cancer³⁹⁻⁴¹. Whole genome sequencing of cancer showed that whole genome doubling (WGD) was an early event in over 40% of breast cancers and occurred significantly more frequently in *TP53* mutated compared with *TP53* wild type tumors⁴². Thus, in our HMEC models p53 inactivation did not result in structural rearrangements, but favored genetic plasticity and laid the ground for genetic anomalies that occurred upon oncogene expression. Noticeably, shp53-CCNE1 and shp53-WNT1 became aneuploid and showed a significant increase in rearrangement levels after soft agar cloning. In remarkable contrast, shp53-RAS remained strictly tetraploid and did not acquire further anomalies after soft agar. These results, together with reduced numbers of γ H2Ax and 53BP1 foci in shp53-RAS, compared with shp53-CCNE1 or shp53-WNT1, indicated lower levels of genetic instability in RAS relative to CCNE1 or WNT1 HMECs. These observations are in line with recent work showing that primary mammary epithelial cells or immortalized HME expressing high levels of *ZEB1* kept stable genomes upon transduction of RASv12⁴³. In this model system, *ZEB1* was shown to control a ROS scavenging program that protected cells overexpressing RASv12 from DNA damage. The relative genetic stability of shp53-RAS HMECs could be linked to the strong EMT, associated with the activation of *ZEB1*, that characterized these

cells (Fig.5A). In contrast, shp53-CCNE1 and shp53-WNT1 HMECs showed an epithelial phenotype and low levels of *ZEB1* expression.

In this work, we inferred a chronology of the genetic changes that occurred at different steps of HMEC transformation. The first modified were miR and mRNA expression, followed by DNA methylation, while CNAs occurred last. This suggested that expression changes acted as drivers, modified the phenotype and impacted on the epigenetic and genetic landscape that finally locked the changes. In shp53-RAS HMECs DNA methylation largely favored EMT, by repressing *KRT5*, *ITGB4*, *NUAK1*, *ACP5* and *PROM2* and activating the pro-invasive *ITGBL1* and *FRMD4A* genes. Thus, epigenetic changes in shp53-RAS are in coherence with the deep shift towards a mesenchymal phenotype undergone by these cells, as well as with the proposition that DNA methylation changes are hallmarks of advanced stages of EMT⁴⁴. Interestingly, CNAs in shp53-RAS impacted two actors of the NOTCH pathway, with *HEY1* being gained and overexpressed and *FBXW7* being lost and downregulated. *FBXW7* ubiquitinates NOTCH and represses its activity²⁷. Interestingly, *FBXW7* ubiquitination activity is itself repressed by RAS activation⁴⁵. This suggests that the copy number loss of *FBXW7* could be an adaptive response to RAS activation, resulting in the activation of the NOTCH pathway. In shp53-CCNE1/shp53-WNT1 DNA methylation resulted in the repression of the proinvasive genes *ANGPTL4*, *ITGB3*, *EHD3* and *FBN2* and thus, appeared to consolidate the epithelial phenotype. This trend was reinforced by CNAs, as shown by the loss and downregulation of *TGFBR2* and *CTNNB1*, two known promoters of EMT. Further notable consequence of CNAs in shp53-CCNE1/shp53-WNT1 cells were the gain of the mitotic spindle assembly genes *NUMA1*, *CUX1* and loss of *AKAP12* negative regulator of Polo Kinase, whose depletion has been linked to aneuploidy⁴⁶ suggesting these anomalies could correspond to adaptive responses linked to increasing chromosomal instability.

Shp53-RAS were classified Claudin-low, whereas shp53-CCNE1 and shp53-WNT1 classified as Basal-like with epithelial dominance and we noted an analogy between shp53-RAS and basal-like/claudin-low breast cancer, whereas shp53-CCNE1 and shp53-WNT1 resembled non-claudin-low/basal-like tumors. Indeed, basal-like/claudin-low breast cancers show low CNA levels and a frequent activation of the RAS pathway, whereas non-claudin-low/basal-like tumors present high levels of CNAs and elevated *CCNE1* expression. It is clear that HMEC models do not sum up the complete spectrum of human breast cancers, however our data show that they mimic, at least partially, Basal-like and Claudin-low breast cancer. In conclusion, our data show that early activation of distinct oncogenic insults in a given cell type will not only impinge on the phenotypic characteristics of the resulting tumors, but also impact on their genomic and epigenetic landscapes. Indeed, the genes whose expression was modified either by DNA methylation or CNAs in our models were consistent with the dominant pathways activated and reflected the phenotypes of the respective models, mesenchymal in shp53-RAS, epithelial in shp53-CCNE1/shp53-WNT1 (Fig.6).

Grants and support: Agence Nationale pour la Recherche ANR projet blanc ESCATMO. Anne Saumet was supported by ESCATMO, MESR doctoral fellowship to Claire Fonti. SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553. European Research Council ERC Consolidator grant n°615371 to Michael Weber.

Acknowledgments: Authors wish to thank Anne Morel and Laurent Le Cam for providing retroviral constructs used in this project and Alexandre Djiane for critical reading. The expert input of the personnel of the microarray Center of the Genomics Research Unit Luxembourg Institute of Health is gratefully acknowledged.

Authors contributions: CF, AS, AAK: cell and molecular biology, retroviral transductions, biochemistry. BO: array-CGH and data analysis. ES, JC: biomathematical analyses of gene expression data. EC, AB, MD, MW: RBBS libraries and DNA methylation data production. WJ, CS: helped design and edited the manuscript. SdM: designed and performed bioinformatics and biomathematical analyses, edited the manuscript. CT: Study design and coordination, fund raising, manuscript preparation and writing.

Availability of data and materials

Raw array and RRBS data can be accessed at GSE114849 (see S Appendix).

References

1. Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, Rodriguez EG, Kuo W-L, Ljung BM, Chew K, Myambo K, Miranda M, Krig S, et al. In situ analyses of genome instability in breast cancer. *2004*;36:984–8.
2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *2009*;458:719–24.
3. Jones S, Wang T-L, Shih I-M, Mao T-L, Nakayama K, Roden R, Glas R, Slamon D, Diaz LA, Vogelstein B, Kinzler KW, Velculescu VE, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* 2010;330:228–31.
4. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer* 2004;4:143–53.
5. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128:683–92.
6. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *2013*;45:1127–33.
7. Sack LM, Davoli T, Li MZ, Li Y, Xu Q, Naxerova K, Wooten EC, Bernardi RJ, Martin TD, Chen T, Leng Y, Liang AC, et al. Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* 2018;;1–40.
8. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Molecular Oncology* 2011;5:5–23.
9. Guedj M, Marisa L, de Reynies A, Orsetti B, Schiappa R, Bibeau F, Macgrogan G, Lerebours F, Finetti P, Longy M, Bertheau P, Bertrand F, et al. A refined molecular taxonomy of breast cancer. *Oncogene* 2011;
10. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346–52.
11. Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jönsson G, Olsson H, Borg A, Ringnér M. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res* 2010;12:R36.
12. van Beers EH, van Welsem T, Wessels LFA, Li Y, Oldenburg RA, Devilee P, Cornelisse CJ, Verhoef S, Hogervorst FBL, van't Veer LJ, Nederlof PM. Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations. *Cancer Res* 2005;65:822–7.
13. Smolen GA, Muir B, Mohapatra G, Barmettler A, Kim WJ, Rivera MN, Haserlat SM, Okimoto RA, Kwak E, Dahiya S, Garber JE, Bell DW, et al. Frequent met oncogene amplification in a Brca1/Trp53 mouse model of mammary tumorigenesis. *Cancer Res* 2006;66:3452–5.
14. Gazin C, Wajapeyee N, Gobeil S, Virbasius C-M, Green MR. An elaborate pathway required for Ras-mediated epigenetic silencing. *2007*;449:1073–7.
15. Holstege H, van Beers E, Velds A, Liu X, Joosse SA, Klarenbeek S, Schut E, Kerkhoven R, Klijn CN, Wessels LFA, Nederlof PM, Jonkers J. Cross-species comparison of aCGH data from mouse and human BRCA1- and BRCA2-mutated breast cancers. *BMC Cancer* 2010;10:455.
16. Wajapeyee N, Malonia SK, Palakurthy RK, Green MR. Oncogenic RAS directs silencing of

tumor suppressor genes through ordered recruitment of transcriptional repressors. *Genes & Development* 2013;27:2221–6.

17. Fang M, Ou J, Hutchinson L, Green MR. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Molecular Cell* 2014;55:904–15.
18. Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, Popescu NC, Hahn WC, Weinberg RA. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. 2001;15:50–65.
19. Auclair G, Guibert S, Bender A, Weber M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* 2014;15:490.
20. Xu D, Wang Q, Gruber A, Björkholm M, Chen Z, Zaid A, Selivanova G, Peterson C, Wiman KG, Pisa P. Downregulation of telomerase reverse transcriptase mRNA expression by wild type p53 in human tumor cells. *Oncogene* 2000;19:5123–33.
21. Kanaya T, Kyo S, Hamada K, Takakura M, Kitagawa Y, Harada H, Inoue M. Adenoviral expression of p53 represses telomerase activity through down-regulation of human telomerase reverse transcriptase transcription. *Clin Cancer Res* 2000;6:1239–47.
22. Stampfer MR, Garbe J, Nijjar T, Wigington D, Swisshelm K, Yaswen P. Loss of p53 function accelerates acquisition of telomerase activity in indefinite lifespan human mammary epithelial cell lines. *Oncogene* 2003;22:5238–51.
23. Toledo LI, Murga M, Zur R, Soria R, Rodriguez A, Martinez S, Oyarzabal J, Pastor J, Bischoff JR, Fernandez-Capetillo O. A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations. *Nat Struct Mol Biol* 2011;18:721–7.
24. Meng C, Helm D, Frejno M, Kuster B. moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets. *J Proteome Res* 2016;15:755–65.
25. Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer? *EMBO Rep* 2010;11:670–7.
26. Kent OA, McCall MN, Cornish TC, Halushka MK. Lessons from miR-143/145: the importance of cell-type localization of miRNAs. *Nucleic Acids Res* 2014;42:7528–38.
27. Welcker M, Clurman BE. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat Rev Cancer* 2008;8:83–93.
28. Bailey CM, Khalkhali-Ellis Z, Seftor EA, Hendrix MJC. Biological functions of maspin. *J Cell Physiol* 2006;209:617–24.
29. Huynh J, Etemadi N, Hollande F, Ernst M, Buchert M. The JAK/STAT3 axis: A comprehensive drug target for solid malignancies. *Semin Cancer Biol* 2017;45:13–22.
30. Li XQ, Du X, Li DM, Kong PZ, Sun Y, Liu PF, Wang QS, Feng YM. ITGBL1 Is a Runx2 Transcriptional Target and Promotes Breast Cancer Bone Metastasis by Activating the TGF Signaling Pathway. *Cancer Res* 2015;75:3302–13.
31. Goldie SJ, Mulder KW, Tan DWM, Lyons SK, Sims AH, Watt FM. FRMD4A Upregulation in Human Squamous Cell Carcinoma Promotes Tumor Growth and Metastasis and Is

Associated with Poor Prognosis. *Cancer Res* 2012;72:3424–36.

32. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, Speed T, Spellman PT, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 2006;10:515–27.
33. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat M-L, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* 2009;15:907–13.
34. Molyneux G, Geyer FC, Magnay F-A, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A, Reis-Filho JS, Smalley MJ. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 2010;7:403–17.
35. Proia TA, Keller PJ, Gupta PB, Klebba I, Jones AD, Sedic M, Gilmore H, Tung N, Naber SP, Schnitt S, Lander ES, Kuperwasser C. Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* 2011;8:149–63.
36. Van Keymeulen A, Lee MY, Ousset M, Brohée S, Rorive S, Girardi RR, Wuidart A, Bouvencourt G, Dubois C, Salmon I, Sotiriou C, Phillips WA, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. 2015;525:119–23.
37. Ben-David U, Ha G, Khadka P, Jin X, Wong B, Franke L, Golub TR. The landscape of chromosomal aberrations in breast cancer mouse models reveals driver-specific routes to tumorigenesis. *Nat Comms* 2016;7:12160.
38. Aylon Y, Oren M. p53: guardian of ploidy. *Molecular Oncology* 2011;5:315–23.
39. Kuznetsova AY, Seget K, Moeller GK, de Pagter MS, de Roos JADM, Dürrbaum M, Kuffer C, Müller S, Zaman GJR, Kloosterman WP, Storchová Z. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* 2015;14:2810–20.
40. Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* 2005;437:1043–7.
41. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng C-Z, Wala J, Mermel CH, Sougnez C, Gabriel SB, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;45:1134–40.
42. Bielski CM, Zehir A, Penson AV, Donoghue MTA, Chatila W, Armenia J, Chang MT, Schram AM, Jonsson P, Bandlamudi C, Razavi P, Iyer G, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* 2018;:1–11.
43. Morel A-P, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, Devouassoux-Shisheboran M, Combaret V, Finetti P, Chassot C, Pinatel C, Fauvet F, et al. A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med* 2017;23:568–78.
44. Tam WL, Weinberg RA. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med* 2013;19:1438–49.
45. Minella AC, Welcker M, Clurman BE. Ras activity regulates cyclin E degradation by the Fbw7 pathway. *Proc Natl Acad Sci USA* 2005;102:9649–54.

46. Canton DA, Keene CD, Swinney K, Langeberg LK, Nguyen V, Pelletier L, Pawson T, Wordeman L, Stella N, Scott JD. Gravin Is a Transitory Effector of Polo-like Kinase 1 during Cell Division. *Molecular Cell* 2012;48:547–59.

Figure Legend:

Fig. 1: Stepwise transformation HMEC models. **A:** Experimental scheme of the subline production. **B:** percentage of B-galactosidase positive cells in the different HMEC variants. **C:** Mean length of telomeres estimated by Southern blotting show a stabilization of telomere length in shp53.Late and transformed HMECs. **D:** Fraction of the genome involved in Copy Number Alterations. Anova Multiple Comparisons test showed significant increase in genomic fraction involved in CNA between R2 primary HMEC and oncogene transduced sublines (R2 vs. shp53-Ras $p=0.0149$; R2 vs. shp53-CCNE1 and shp53-WNT1 $p=0.0193$). Comparison of pretransformed and transformed (after Soft Agar) sublines showed that genomic fraction involved in CNA were significantly increased in shp53-CCNE1.SA and shp53-WNT1.SA cells (shp53-CCNE1/shp53-WNT1 vs. shp53-CCNE1.SA/shp53-WNT1.SA $p= 0.0021$) but not in shp53-RAS.SA cells (shp53-Ras vs. shp53-Ras.SA $p= 0.9835$). **E:** Chromosome numbers were estimated on chromosome spreads. At least 50 karyotypes were scored in each subline.

Fig. 2: HMEC expressing different genetic elements form distinct clusters and show different profiles of copy number alterations. **A:** Joint PCA analysis including CNA, DNA methylation, miR and mRNA data. **B:** PCA analysis of the CNA data (HG18 CGH 385K Whole Genome v2.0 array) **C:** Ward clustering of the CNAs defining PC1 in the PCA analysis, color code of the heatmap is blue for loss, red for gain.

Fig. 3: CpG methylation, miR and mRNA expression profiles show coordinated variation according to the oncogene expressed. **A:** PCA analysis of the most significantly varying DMRs. **B:** Ward clustering of the DMRs (without constraint concerning their location) defining PC1 in the PCA analysis, color code of the heatmap is red for hypermethylation, blue for hypomethylation. **C:** PCA analysis of the most significantly varying miR. **D:** Ward clustering of the miR defining PC1 in the PCA analysis, color code of the heatmap is red for overexpression, blue for underexpression. **E:** PCA analysis of the most significantly varying mRNA. **F:** Ward clustering of the miR defining PC1 in the PCA analysis, color code of the heatmap is red for overexpression, blue for underexpression.

Fig. 4: Most significant gene expression differences due to copy number alterations (A) and differential methylation (B) in shp53-RAS and shp53-CCNE1/shp53-WNT1 HMECs. Expression changes (Exp ratio column) have been calculated relative to expression levels in the R2 and shp53 cluster and are indicated in log2 scale. Genes are listed for each cluster (RAS for shp53-RAS, C/W for shp53-CCNE1/shp53-WNT1). **A:** Genes modified as a consequence of CNA have been selected from a broader list on the basis of their assignment either as cancer or tumor suppressor genes (assembled under the cancer gene category) or as members of the RAS pathway. Highlighted in red, genes overexpressed in regions of gain (Gain/OE), in blue, genes underexpressed in regions of loss (Loss/UE). **B:** Genes modified as a consequence of differential methylation at the transcription start sites (TSS). In red, genes hypomethylated with increased expression (Hypo/OE), in blue, genes hypermethylated with reduced expression (Hyper/UE).

Fig. 5: The different HMEC models resemble basal breast cancers and show varying levels of mesenchymal traits. **A.** Clustering analysis using a 28 gene signature representative of claudin low tumors.

shp53-RAS models classified as strict claudin-low (mesenchymal), whereas shp53-CCNE1 and shp53-WNT1 were strictly epithelial. Of note, R2 (Ctrl here) and R2shp53 presented intermediate profiles combining expression of both epithelial and mesenchymal genes. **B-C:** CCNE1/WNT1 and RAS HMECs resemble respectively Basal-like and claudin-low breast cancers which show similar differences in CNA numbers. **D:** Differences in CCNE1 mRNA expression levels. This analysis was done on 259 Basal-like tumors (PAM 50 classification) selected from the Metabric dataset which were split in Basal-Like/Non claudin (176) and Claudin-low (83).

Fig. 6: Genetic, epigenetic and phenotypic modifications resulting from the transduction of distinct oncogenes in HMEC. The changes occurring are depicted at different steps starting from primary HMEC and ending at transformed cells isolated from soft agar colonies. For readability purposes, only principal events and genes are presented. Arrows indicate upregulation or downregulation of expression of genes modified by CNAs or differential methylation at their TSS, as well as up or downregulation of pathways as part of the phenotypic consequences of the activation of the respective oncogenes.

Legends to Supporting Information Figures and Tables

Supporting Information 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.

Supporting Information 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:** Quantification of the corresponding bands and ratio with the reference protein (GAPDH) **C:** QPCR verification of WNT1, CCNE1 and RAS mRNA expression in R2-shp53-WNT1, R2-shp53-CCNE1 and R2-shp53-RAS respectively. **D:** protein expression levels by western blotting. E: early; L: Late. **E:** Quantification of the blots

Supporting Information 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:** Growth curves of the HMEC models shown in this experiment **D:** Tail moment measurements box plot. **E:** 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.

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

Supporting Information 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 height 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 height of the bars indicates the amplitude of the copy number change.

Supporting Information 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

Supplementary Fig.7: principal pathways and regulation networks differentially expressed in the RAS compared with the WNT/CCNE cluster. **A-** differential pathways at the level of miR expression; the 81 miR contributing the most to the definition of first axis of the PCA in Fig 3c were used, the log value indicates the significance level. We selected the top 3 pathways. **B-** differential pathways at the level of mRNA expression; the 185 genes contributing the most to the definition of the first axis of the PCA in Fig 3e were used, the log value indicates the significance level. The top 10 are presented. **C-** modified regulation network as determined by Metacore showing the miR dependent regulation of EMT map. Blue thermometers shows decrease of miR205, miR200c, miR141 in RAS vs CW clusters. Enrichment analysis was done using Genego (version 6.29.68613)

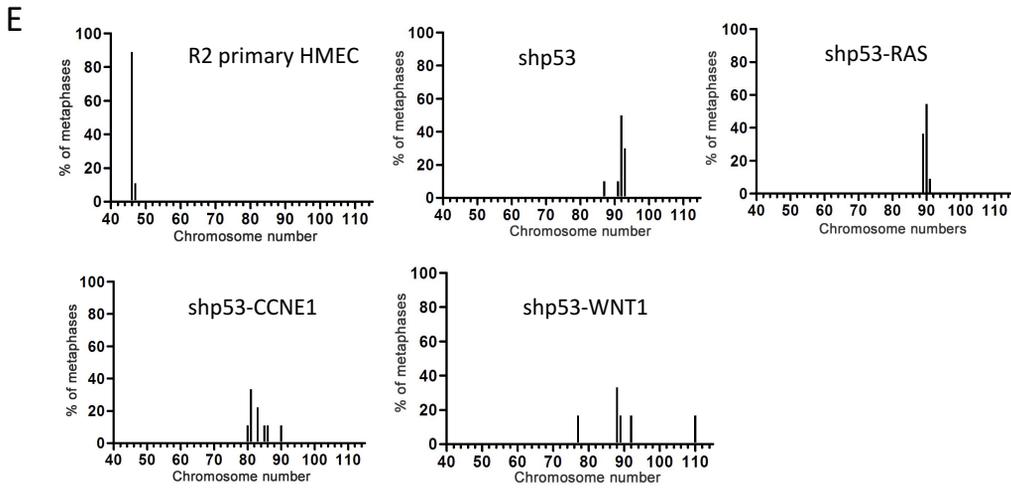
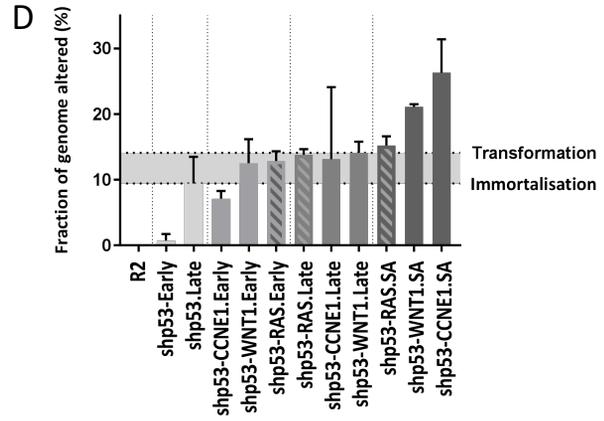
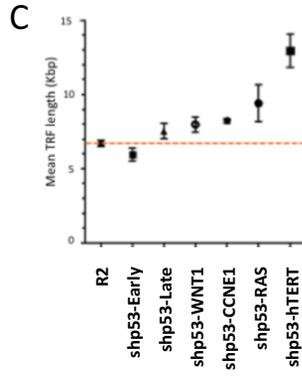
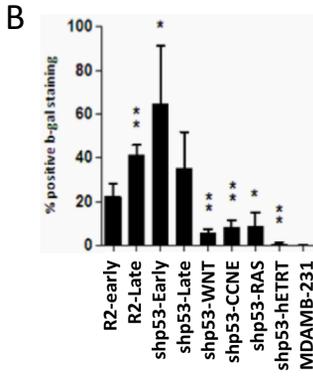
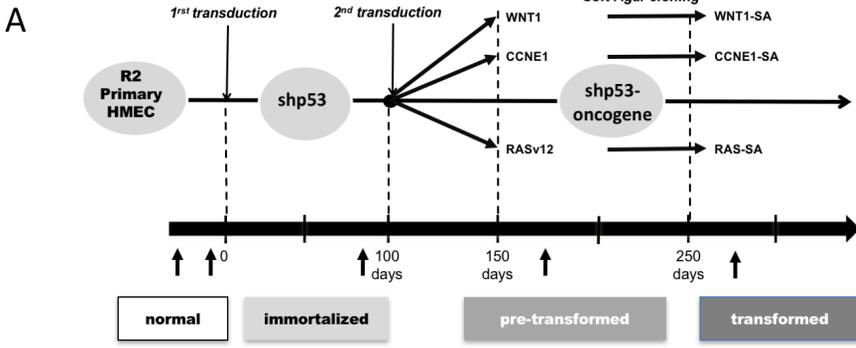
Supporting Information Fig.8: phenotypic characteristics of shp53 HMEC sublines. Cells were stained by immunofluorescence (IF) for the expression of ECAD (E-Cadherin, green) which is a marker of epithelial cells, VIM (Vimentin, red) marker of mesenchymal cells, CK8 (Cytokeratin 8) marker of luminal breast 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 mosaic 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.

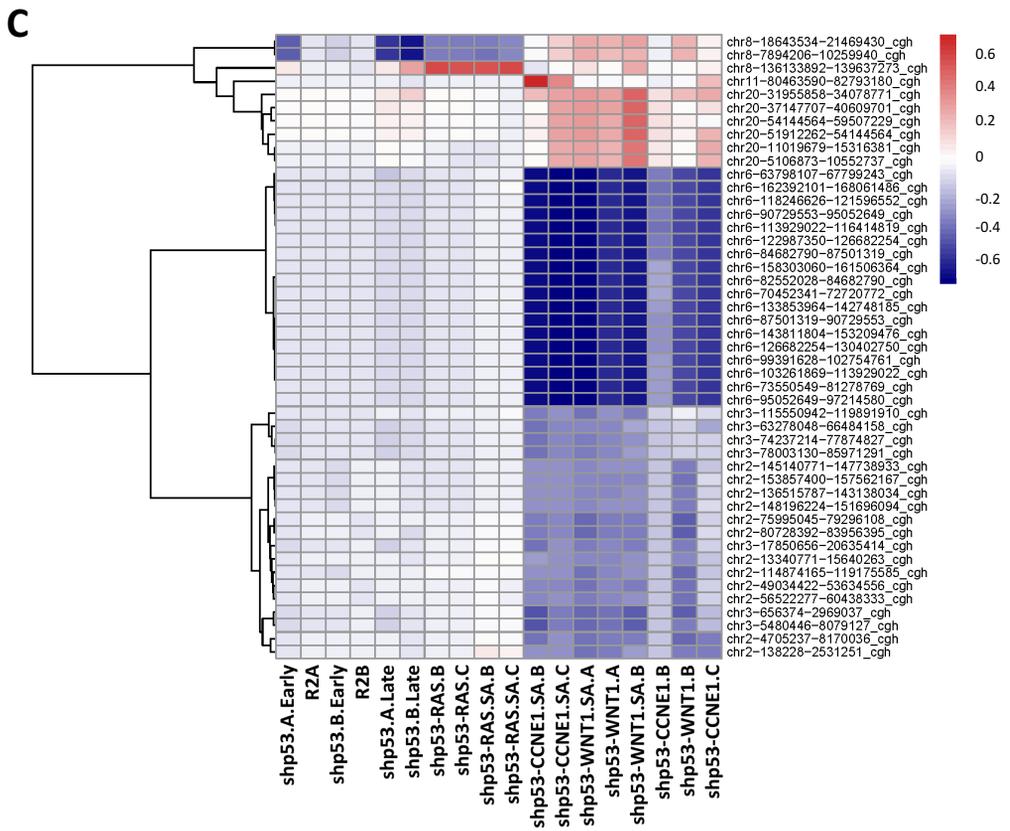
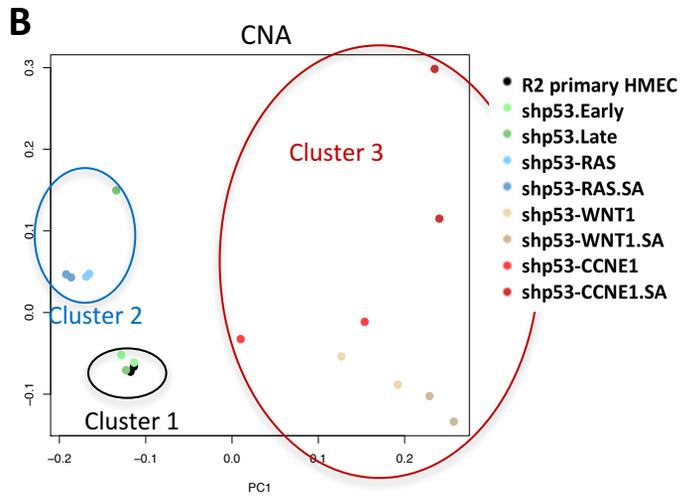
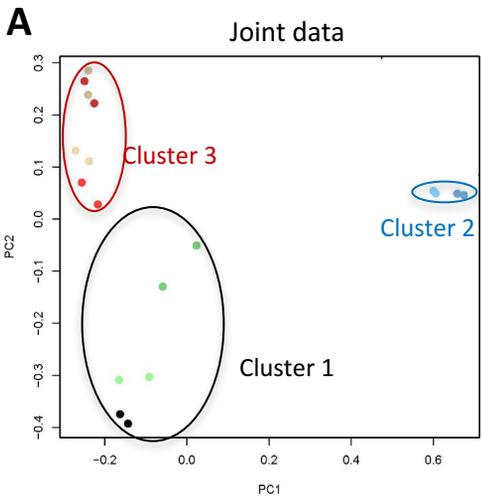
Supporting Information Fig.9: Genes modified by CNA or differential methylation in shp53-RAS and shp53-CCNE1/ shp53-WNT1 show little overlap. Of note MTUS1 is strongly underexpressed in RAS (-2,95; about 1/10 x) whereas it is overexpressed in CW (0,87, about 1,8 x). PLAG1 is strongly overexpressed in 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.

Supporting Information 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)

Supporting Information Table 1A: 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)

Supporting Information Table 1B: 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).





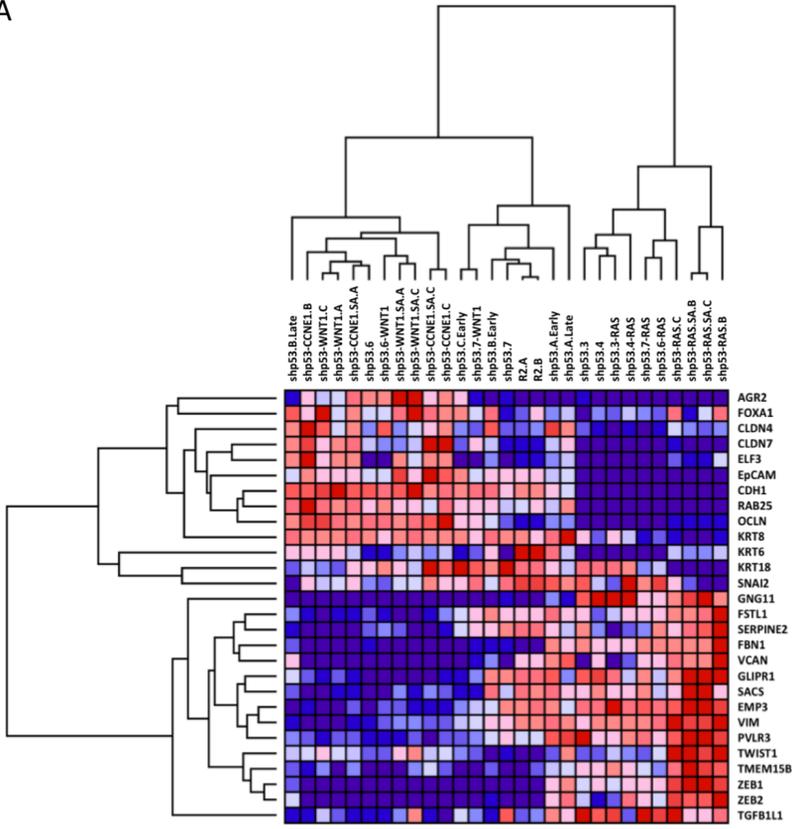
A

Cluster	Gene Name	Chrom band	CNA/Exp	Exp ratio	Cancer Gene Ras pathway	Complete name
RAS	ERBB2	17q12.1	Gain/OE	1.05		erb-b2 receptor tyrosine kinase 2
RAS	PLAG1	8q12.1	Gain/OE	3.25		PLAG1 zinc finger
RAS	HEY1	8q21.13	Gain/OE	2.8		hes related family bHLH transcription factor
RAS	ZBTB10	8q21.13	Gain/OE	0.9		zinc finger and BTB domain containing 10
RAS	CRTC3	15q26.1	Gain/OE	1.24		CREB regulated transcription coactivator 3
RAS	SERPINB5	18q21.33	Loss/UE	-1.3		serpin family B member 5 (Maspin)
RAS	TMEM154	4q31.3	Loss/UE	-2.7		transmembrane protein 154
RAS	LRAT	4q32.1	Loss/UE	-1.8		lecithin retinol acyltransferase
RAS	FBXW7	4q31.3	Loss/UE	-1.4		F-box and WD repeat domain containing 7
RAS	AKAP6	14q12	Loss/UE	-0.9		A-kinase anchoring protein 6
RAS	MTUS1	8p22	Loss/UE	-2.9		microtubule associated scaffold protein 1
C/W	NUMA1	11q13.4	Gain/OE	1.46		nuclear mitotic apparatus protein 1
C/W	PMS2	7p22.1	Gain/OE	0.96		PMS1 homolog 2, mismatch repair system component
C/W	PLAG1	8q12.1	Gain/OE	0.66		PLAG1 zinc finger
C/W	STAT3	17q21.2	Gain/OE	0.55		signal transducer and activator of transcription 3
C/W	CUX1	7q22.1	Gain/OE	0.61		cut like homeobox 1
C/W	TGFBR2	3p24.1	Loss/UE	-0.94		transforming growth factor beta receptor 2
C/W	CTNNB1	3p22.1	Loss/UE	-0.97		catenin beta 1
C/W	FYN	6q21	Loss/UE	-0.89		FYN proto-oncogene, Src family tyrosine kinase
C/W	PTPRK	6q22.33	Loss/UE	-0.62		protein tyrosine phosphatase, receptor type K
C/W	PLAGL1	6q24.2	Loss/UE	-2.83		PLAG1 like zinc finger 1
C/W	AKAP12	6q25.1	Loss/UE	-1.6		A-kinase anchoring protein 12
C/W	LATS1	6q25.1	Loss/UE	-0.53		large tumor suppressor kinase 1
C/W	MAP3K7	6q15	Loss/UE	-0.94		mitogen-activated protein kinase kinase kinase 7

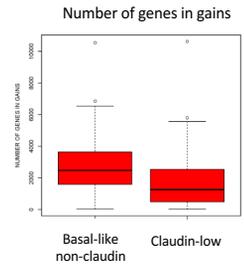
B

Cluster	Gene Name	Chrom Band	mCpG/Exp	Exp Ratio	Complete name
RAS	KRT5	12q13.13	Hyper/UE	-6.09	keratin 5
RAS	ITGB4	17q25.1	Hyper/UE	-3.8	integrin subunit beta 4
RAS	OCIAD2	4p11	Hyper/UE	-2.38	OCIA domain containing 2
RAS	SLFN13	17q12	Hyper/UE	-2.36	schlafen family member 13
RAS	DMKN	19q13.12	Hyper/UE	-2.08	dermokine
RAS	NUAK1	12q23.3	Hyper/UE	-1.95	NUAK family kinase 1
RAS	CYB5R2	11p15.4	Hyper/UE	-1.84	cytochrome b5 reductase 2
RAS	SLC7A5	16q24.2	Hyper/UE	-1.82	solute carrier family 7 member 5
RAS	PTPN6	12p13.31	Hyper/UE	-1.76	protein tyrosine phosphatase, non-receptor type 6
RAS	MYEF2	15q21.1	Hyper/UE	-1.72	myelin expression factor 2
RAS	TAGLN3	3q13.2	Hyper/UE	-1.52	transgelin 3
RAS	PKP3	11p15.5	Hyper/UE	-1.47	plakophilin 3
RAS	ACP5	19p13.2	Hyper/UE	-1.47	acid phosphatase 5, tartrate resistant
RAS	HK2	2p12	Hyper/UE	-1.37	hexokinase 2
RAS	PROM2	2q11.1	Hyper/UE	-1.35	prominin 2
RAS	KDF1	1p36.11	Hyper/UE	-1.11	keratinocyte differentiation factor 1
RAS	GRAMD2	15q23	Hyper/UE	-1.09	GRAM domain containing 2A
RAS	TBX18	6q14.3	Hypo/OE	4.96	T-box 18
RAS	ANPEP	15q26.1	Hypo/OE	2.55	alanyl aminopeptidase, membrane
RAS	ITGBL1	13q33.1	Hypo/OE	2.24	integrin subunit beta like 1
RAS	FRMD4A	10p13	Hypo/OE	1.03	FERM domain containing 4A
C/W	ANGPTL4	19p13.2	Hyper/UE	-3.41	angiopoietin like 4
C/W	FBN2	5q23.3	Hyper/UE	-3.37	fibrillin 2
C/W	PLAGL1	6q24.2	Hyper/UE	-2.88	PLAG1 like zinc finger 1
C/W	NRN1	6p25.1	Hyper/UE	-2.44	neurtin 1
C/W	BCAT1	12p12.1	Hyper/UE	-2.43	branched chain amino acid transaminase 1
C/W	UCHL1	4p13	Hyper/UE	-2.00	ubiquitin C-terminal hydrolase L1
C/W	EHD3	2p23.1	Hyper/UE	-1.95	EH domain containing 3
C/W	CYB5R2	11p15.4	Hyper/UE	-1.68	cytochrome b5 reductase 2
C/W	ITGB3	17q21.32	Hyper/UE	-1.58	integrin subunit beta 3
C/W	TAGLN3	3q13.2	Hyper/UE	-1.47	transgelin 3
C/W	KCNC1	11p15.1	Hyper/UE	-1.38	potassium voltage-gated channel subfamily C 1
C/W	RFTN1	3p24.3	Hyper/UE	-1.29	raftlin, lipid raft linker 1
C/W	DYSF	2p13.2	Hyper/UE	-1.29	dysferlin
C/W	MYADM	19q13.42	Hyper/UE	-1.19	myeloid associated differentiation marker
C/W	FGF5	4q21.21	Hyper/UE	-1.04	fibroblast growth factor 5
C/W	HBS1L	6q23.3	Hyper/UE	-1.03	HBS1 like translational GTPase

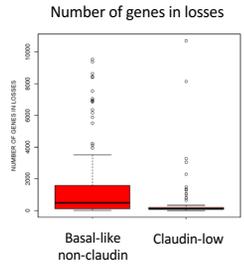
A



B



C



D

