SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Icicle representation and analysis in MiSTIC

A) MiSTIC Introductory web page. The main menu of MiSTIC is presented at the top in bold with 3 tabs for access to the "Datasets and Icicle" view, to "Single gene correlation analysis" (waterfall format) and to "Pair-wise correlation scatterplots". The list of datasets available on the server can be sorted or filtered in the Datasets and Icicle view to identify the appropriate dataset to visualize. Each row corresponds to a dataset available on the MiSTIC server. The first four columns (dataset, tissue, project and technology) are name tags for which the key is displayed as column header. Columns can be reordered, sorted and each one can be used for reordering datasets alphabetically (here, the project tags were used). The search box dynamically limits the dataset presented in this view by filtering on all tags. 'n' indicates the number of samples used for each dataset and the 'icicle' link leads to the icicle page (blue arrow). B) Basic icicle view visualizing the correlation peaks in the selected dataset. By clicking on individual grey grid circles, the corresponding threshold(s) can be indicated with bold lines for easier reference. Enrichment in a specific gene set can be mapped onto the icicle by entering the appropriate type, category, and geneset terms in the search boxes in the geneset enrichment tool (red arrow). C) A specific gene can be located in the cluster with which it correlates most using the "Locate" tool (red arrow). The corresponding peak or layer appears in red in the icicle representation. D) Selection of another dataset in the "Dataset comparison" tool (red arrow) superimposes a colour-coded conservation score onto each cluster; the scale ranges from light blue (for the least conserved clusters) to dark blue (for the most conserved ones). The Kappa coefficient associated with each cluster appears when holding the cursor on this cluster.

Figure S2: Gene correlation analysis in MiSTIC

A) Clusters of co-expressed genes correspond to sub-graphs of the minimum spanning tree and can be viewed in a graph form with edges representing the optimal pair-wise correlation between genes (nodes); a table of enriched gene sets is shown at the upper right side of the screen). P values (Fisher's exact test) and odds ratios are provided for each gene set, as well as its name, type category and ID. Upon selection of a gene set, pertinent genes in the cluster are highlighted in orange with a list at the top of the graph in orange letters (selection). Alternatively, individual genes can be iteratively selected by clicking on the corresponding tags in the graph, providing links to different databases (lower right side of the screen) for the last selection. Clicking on the Scatterplot tab generates scatterplots for each pair of selected genes (blue arrow) B) Pair-wise scatterplots are presented in a table format with gene names shown on the diagonal. Each scatter plot represents the relation between expression levels of the two corresponding genes in all patient samples of the current data set. Scales are logarithmic and use untransformed data, with all values below 0.0105 being replaced by 0.01. Samples can be highlighted by clicking on any dot or using the gating tool (which displays the proportion of samples in the selection). Specimen identifiers are shown at the bottom of the screen in plain text (HTML) to allow simple export of these selections using copy/paste. Samples can also be highlighted by selection of a clinical feature ("***Select a characteristic"*** menu). A sample selection can be saved by clicking on the + sign in the "***Highlight groups"*** tool box (blue arrow). Enrichment in clinical characteristics appears in the "***Sample term enrichment"*** tab (red arrow). C) The sample selection saved in the first group appears in orange by default. The color can be modified by clicking on the colored dot in the "***Highlight groups"*** tool box (red arrow), and a name can be attributed to the sample selection (Group 1 by default). Groups of samples can be removed from this selection by clicking on the – sign.

The selection can be removed using the trash icon. Finally, the samples in the selection can be highlighted in green by clicking on the arrow. New groups can be added to track simultaneously several groups of samples (see examples in **Fig 7** and Figs **S6C&8C**) D) in the "**Single gene correlations"** display (red arrow), both dataset and query gene are selected from the menu ("***Dataset"*** and "***Gene"*** tabs). The most correlated or anticorrelated genes (10 by default, but the number of labels to display can be increased in the "***More options"*** menu) are shown in a waterfall display. Enrichment in selected gene sets ("***Gene set tab"***) is visualised by blue ticks on the X axis of the graph.

Figure S3: Comparing gene expression clusters in cancer and/or normal tissues with MiSTIC.

A) Leucegene AML (dataset 1, **Table S1**) icicle highlighted for conservation with CD34+ cells (dataset 5, **Table S1**). B) TCGA breast cancer (dataset 7, **Table S1**) icicle highlighted for conservation with colon cancer (dataset 24, **Table S1**).

Figure S4: The AML HOXA/B cluster includes transcription factors PBX and Meis1

A) Minimum spanning tree representation of the *HOXA-MEIS-PBX* cluster from the NK-AML dataset (#4, **Table S1**). B-C) Waterfall representation of genes correlated to *MEIS1* in leukemia (dataset #4, **Table S1**) (B) and in lung cancer from TCGA (dataset #23, **Table S1**) (C).

Figure S5: Use of the Toggle label option for improved visualization of the correlation graph structure.

The cell cycle cluster in the TCGA breast cancer dataset is visualized with (A) or without (B) labels.

Figure S6: TP63 and FOXC1 correlation clusters enriched in the PAM50 gene set.

Genes present in the correlation peaks containing *TP63* and *FOXC1* are shown.

Figure S7: Use of the Locate gene set tool for visualization of the genes in a set.

Genes in the PAM50 (A) or Oncotype Dx (B) gene sets are visualized on the TCGA breast cancer icicle**,** revealing genes in these signatures that are found in different correlation peaks or in large clusters of less correlated genes.

Figure S8: Analysis of sample enrichment in the breast cancer dataset

A-B) Samples in the breast cancer dataset (#7) were ordered according to expression levels of *AURKA* and *CENPA*, which are highly correlated (r=0.83). Windows corresponding to low (A) or high (B) expression levels of these genes were manually selected. The sample term enrichment list indicates strong enrichment in PAM50 subtypes LumA and Basal. C) Tumors annotated as LumA, LumB, HER2+ and Basal were selected from the ***"Select a characteristic"*** list and highlighted in different colors in the "***Highlight groups"*** tool-box by adding selected samples with the + tool. Names were attributed to each group by clicking on the colored dot and entering the name in the label box.

Figure S9: Discrepancies between ER and HER2 levels defined by immuno-histochemistry and by mRNA levels.

A) Breast tumors were sorted according to expression levels of *ESR1* and *ERBB2*, defining four subpopulations, i.e. *ESR1*^{hi}*ERBB2*^{lo} (orange) *ESR1*^{hi}*ERBB2*^{hi} (blue) *ESR1*^{lo}*ERBB2*^{hi} (purple) *ESR1*^{lo}*ERBB2*^{lo} (dark grey). B-C) ER+ tumors were selected (B) and added to a new highlight group of tumors. *ESR1*^{hi} tumor group were then selected and removed from this group to generate an ER+*ESR1*^{lo} group of tumors, highlighted in red (C). D) Sorting the ER+ *ESR1*^{lo} tumors for expression levels of *GREB1* and *ESR1* indicates that these tumors express mostly low levels of *GREB1* in spite of being classified as ER+ tumors. E) Conversely, ER-*ESR1*^{hi} tumors were highlighted by selection of ER- tumors and removal of the *ESR1*^{lo} groups. F) Expression levels of *GREB1* are high in most ER-*ESR1*^{hi} tumors, suggesting active ER signaling. G) HER2+*ERBB2*^{lo} tumors were highlighted by selection of HER2+ tumors and removal of the *ERBB2*^{hi} group. H) HER2+*ERBB2*^{lo} tumors express low levels of *GRB7*.

Figure S10: Analysis of sample enrichment in the AML Leucegene dataset

A-B) Samples in the AML Leucegene dataset (#1) were ordered according to expression levels of *AURKA* and *CENPA*, which are highly correlated (r=0.87). Windows corresponding to low (A) or high (B) expression levels of these genes were manually selected. The sample term enrichment list indicates strong enrichment in categories Heparinized blood and Bone marrow. C) Tumors annotated as Blood or Bone marrow were selected from the ***"Select a characteristic"*** list and highlighted in different colors in the "***Highlight groups"*** tool-box by adding selected samples with the + tool. This was done in an iterative manner to pool the two Blood or the two Bone marrow categories. Names were added to the group by clicking on the colored dot and entering Blood or Bone marrow in the label box. Characteristic "Disease free survival < 3 yrs" was selected in the "***Select a characteristic"*** window, highlighting relevant samples in green.

Figure S11: Analysis of cell line transcriptome datasets.

A) Icicle comparison between the CCLE and GNE datasets B) Icicle comparison between the CCLE and breast cancer datasets C) enrichment in the gene set associated with epithelial differentiation in normal breast tissue in a peak of the CCLE dataset. D) GRHL2 and RAB25, two genes of the epithelial differentiation gene set, separate groups of cell lines enriched in hematopoietic-lymphoid origin vs carcinoma annotations in the CCLE dataset. E) Distribution of breast cancer, lung cancer and melanoma cell lines between the two clusters, suggesting mesenchymal characteristics for variable proportions of cell lines for each cancer type in the CCLE dataset.

Figure S12: Analyzing correlations between miRNAs and protein-coding genes.

A) Icicle generated from the TCGA breast cancer miRNA dataset (dataset 17 in **Table S1**). Comparison with the TCGA breast normal miRNA dataset (dataset 16 in **Table S1**) is shown. The number of features is lower and fewer peaks are formed compared to protein-coding gene datasets. Enrichment analysis can be performed as above to assess whether miRNA clusters are linked to individual chromosomal loci (whether due to amplification/deletion or to co-localization in miRNA clusters) or coregulated by common TFs. B) Icicle generated from the fusion of the TCGA breast normal datasets for miRNAs and protein-coding genes, resulting in the "BRCA miRNA coding normal" dataset (dataset 19 in **Table S1**). The miRNAs *MIR200A*, *MIR200B*, *MIR200C* and *MIR141* are found in the "Epithelial differentiation" peak (highlighted in blue). C) The waterfall tool indicates high anti-correlations of *MIR200A/B/C* with several genes (highlighted in blue) found in the "mesenchymal/adipocyte differentiation" peak (highlighted in red in B). These miRNAs are known to target *ZEB1/ZEB2* (1, 2), which

are part of this cluster. D) ZEB/AREB6 sites are enriched in the "Epithelial differentiation" gene cluster, consistent with a reciprocal transcriptional/post-transcriptional regulation between genes in these two clusters. Accordingly, reciprocal down-regulation between *GRHL2* and *ZEB1* has recently been described (3), and *MIR200A/B/C* and several "Epithelial differentiation" genes have been shown to be down-regulated by ZEB1/2 (1).

SUPPLEMENTARY REFERENCES

1. Bracken CP*, et al.* (2008) A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer research* 68(19):7846-7854.
2. Diaz-Lopez A, Moreno-Bueno G, & Cano A (2014) Role of microRNA in epithelial to mesenchymal transition and metastasis and clinical perspectives. *Cancer Manag Res* 6:205-216.
3. Cieply B, Farris J, Denvir J, Ford HL, & Frisch SM (2013) Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer research* 73(20):6299-6309.

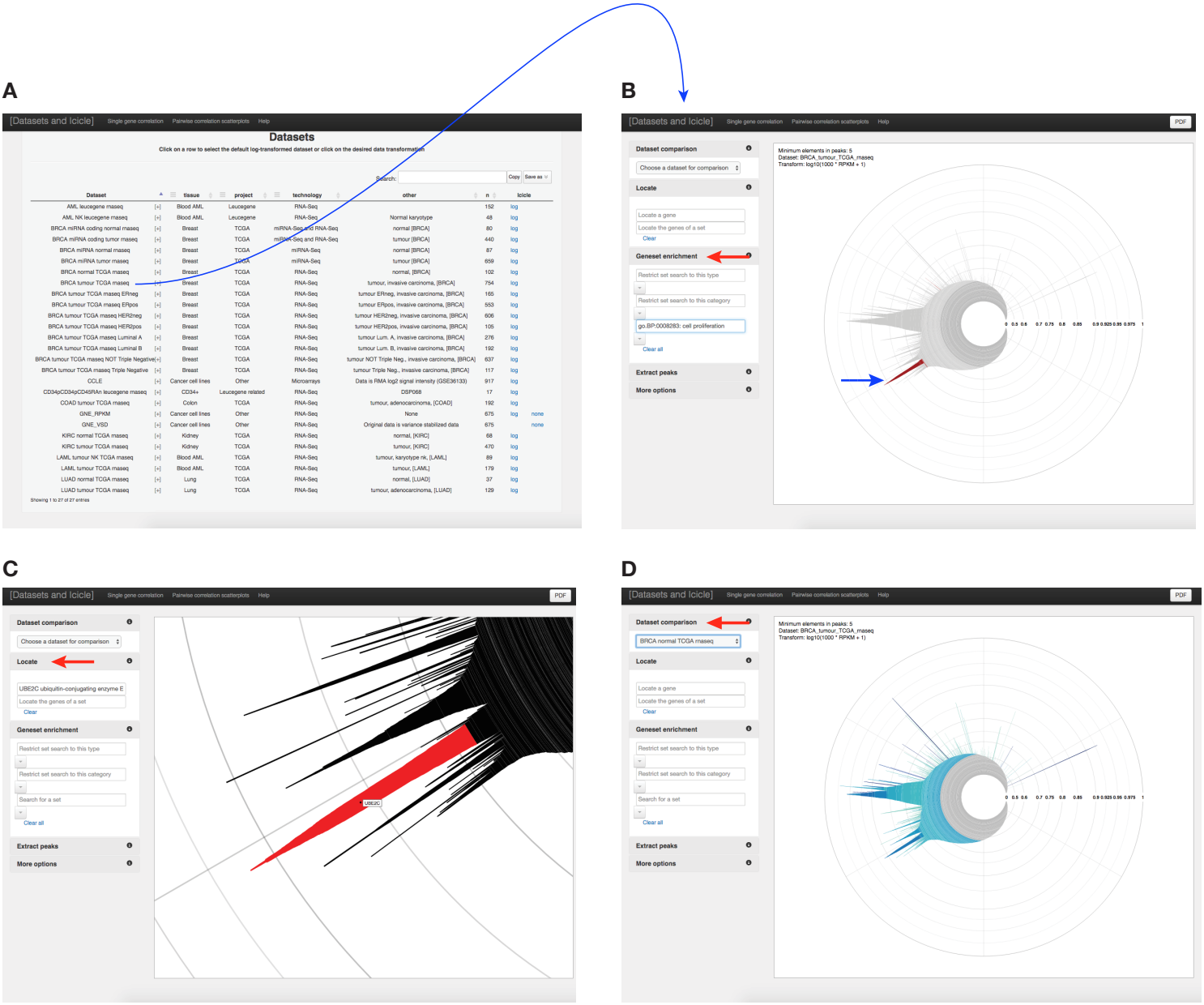

A
B
C
D
[Datasets and Icicle]
Single gene correlation
Pairwise correlation scatterplots
Help
PDF
Datasets
Click on a row to select the default log-transformed dataset or click on the desired data transformation
Search:
Copy
Save as
Dataset
tissue
project
technology
other
n
icicle
AML leucegene maseq
AML NK leucegene maseq
BRCA miRNA coding normal maseq
BRCA miRNA coding tumor maseq
BRCA miRNA normal maseq
BRCA miRNA tumor maseq
BRCA normal TCGA maseq
BRCA tumour TCGA maseq
BRCA tumour TCGA maseq ERneg
BRCA tumour TCGA maseq ERpos
BRCA tumour TCGA maseq HER2neg
BRCA tumour TCGA maseq HER2pos
BRCA tumour TCGA maseq Luminal A
BRCA tumour TCGA maseq Luminal B
BRCA tumour TCGA maseq NOT Triple Negative
BRCA tumour TCGA maseq Triple Negative
CCLE
CD34pCD34pCD45RAn leucegene maseq
COAD tumour TCGA maseq
GNE_RPKM
GNE_VSD
KIRC normal TCGA maseq
KIRC tumour TCGA maseq
LAML tumour NK TCGA maseq
LAML tumour TCGA maseq
LUAD normal TCGA maseq
LUAD tumour TCGA maseq
Showing 1 to 27 of 27 entries
Dataset comparison
Choose a dataset for comparison
Locate
Locate a gene
Locate the genes of a set
Clear
Geneset enrichment
Restrict set search to this type
Restrict set search to this category
go.BP:0008283: cell proliferation
Clear all
Extract peaks
More options
Minimum elements in peaks: 5
Dataset: BRCA_tumour_TCGA_maseq
Transform: log10(1000 * RPKM + 1)
BRCA normal TCGA maseq
UBE2C ubiquitin-conjugating enzyme E
Search for a set
UBE2C

Fig S1

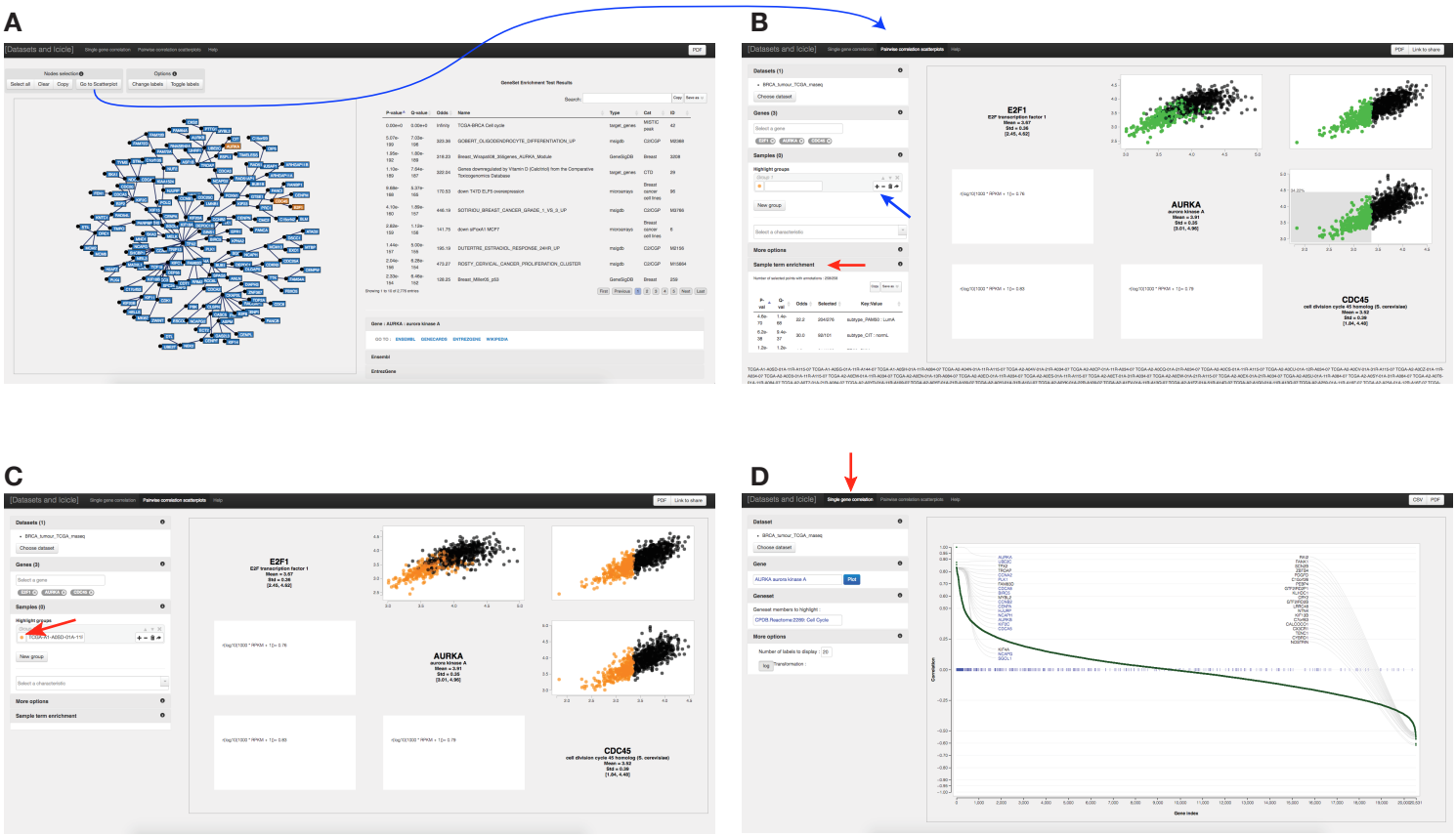


Fig S2

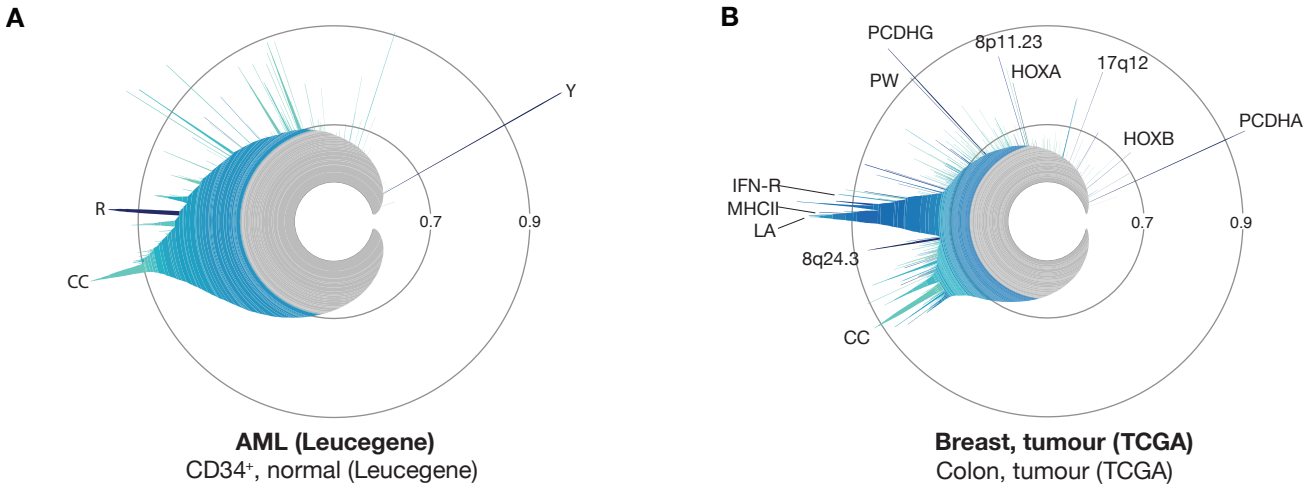

A
Y
R
CC
0.7
0.9
AML (Leucegene)
CD34+, normal (Leucegene)
B
PCDHG
8p11.23
17q12
PW
HOXA
PCDHA
HOXB
IFN-R
MHCII
LA
8q24.3
CC
0.7
0.9
Breast, tumour (TCGA)
Colon, tumour (TCGA)

Fig S3

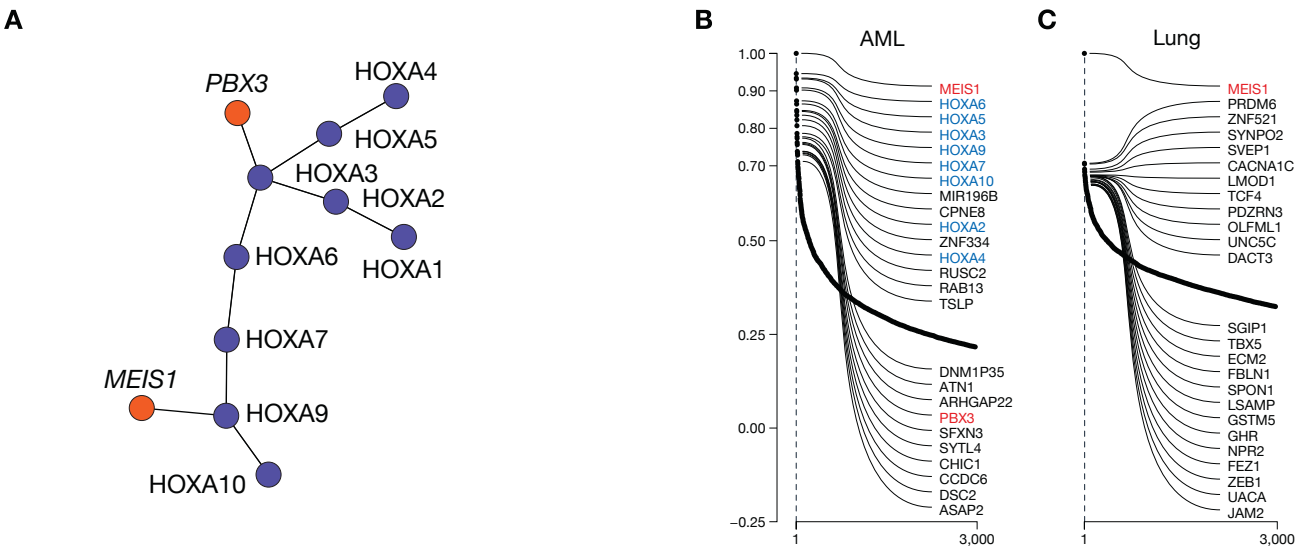

A
PBX3
HOXA4
HOXA5
HOXA3
HOXA2
HOXA1
HOXA6
HOXA7
MEIS1
HOXA9
HOXA10
B
AML
1.00
0.90
0.80
0.70
0.50
0.25
0.00
−0.25
1
3,000
MEIS1
HOXA6
HOXA5
HOXA3
HOXA9
HOXA7
HOXA10
MIR196B
CPNE8
HOXA2
ZNF334
HOXA4
RUSC2
RAB13
TSLP
DNM1P35
ATN1
ARHGAP22
PBX3
SFXN3
SYTL4
CHIC1
CCDC6
DSC2
ASAP2
C
Lung
1
3,000
MEIS1
PRDM6
ZNF521
SYNPO2
SVEP1
CACNA1C
LMOD1
TCF4
PDZRN3
OLFML1
UNC5C
DACT3
SGIP1
TBX5
ECM2
FBLN1
SPON1
LSAMP
GSTM5
GHR
NPR2
FEZ1
ZEB1
UACA
JAM2

Fig S4

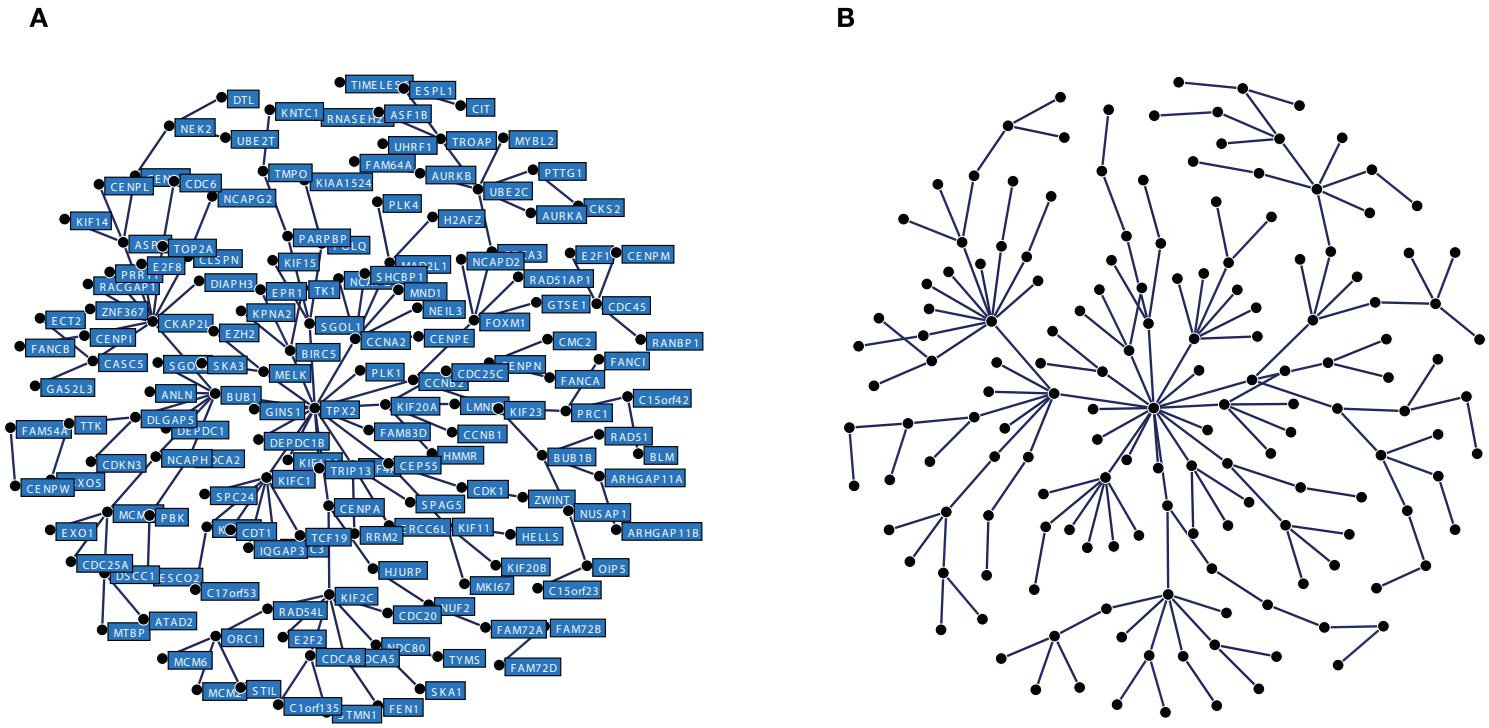
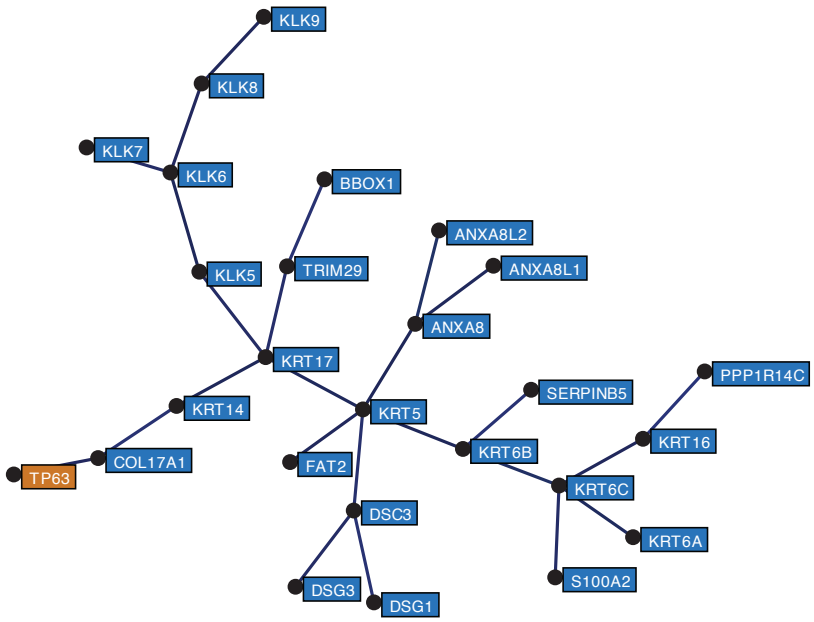

A
B
TIMELESS
ESPL1
CIT
DTL
KNTC1
RNASEH2A
ASF1B
NEK2
UBE2T
UHRF1
TROAP
MYBL2
FAM64A
TMPO
KIAA1524
AURKB
PTTG1
UBE2C
CENPL
CDC6
NCAPG2
PLK4
AURKA
CKS2
KIF14
H2AFZ
PARPBP
TOP2A
KIF15
NCAPD2
E2F1
CENPM
E2F8
DIAPH3
SHCBP1
RAD51AP1
RACGAP1
EPR1
TK1
MND1
GTSE1
CDC45
ZNF367
KPNA2
NEIL3
ECT2
CKAP2L
SGOL1
FOXM1
CENPI
EZH2
CCNA2
CENPE
CMC2
RANBP1
FANCB
BIRC5
CASC5
SKA3
MELK
PLK1
CDC25C
FANCI
GAS2L3
ANLN
BUB1
FANCA
C15orf42
GINS1
TPX2
KIF20A
KIF23
PRC1
FAM54A
TTK
DLGAP5
FAM83D
CCNB1
RAD51
DEPDC1B
CDKN3
HMMR
BUB1B
BLM
KIFC1
TRIP13
CEP55
ARHGAP11A
CENPW
SPC24
CDK1
ZWINT
CENPA
SPAG5
NUSAP1
EXO1
PBK
CDT1
TCF19
RRM2
KIF11
HELLS
ARHGAP11B
CDC25A
HJURP
KIF20B
DIP5
C17orf53
KIF2C
MKI67
C15orf23
ATAD2
RAD54L
CDC20
FAM72A
FAM72B
MTBP
ORC1
E2F2
MCM6
CDCA8
TYMS
FAM72D
STIL
SKA1
C1orf135
TMN1
FEN1

Fig S5

A


KLK9
KLK8
KLK7
KLK6
BBOX1
ANXA8L2
KLK5
TRIM29
ANXA8L1
ANXA8
KRT17
PPP1R14C
SERPINB5
KRT14
KRT5
KRT16
KRT6B
COL17A1
FAT2
TP63
KRT6C
DSC3
KRT6A
S100A2
DSG3
DSG1

B

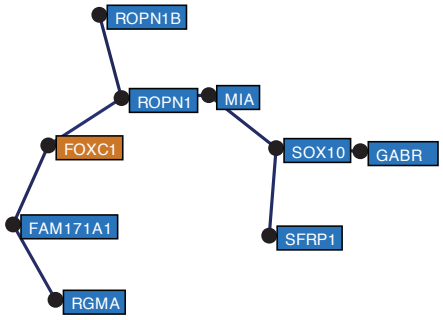
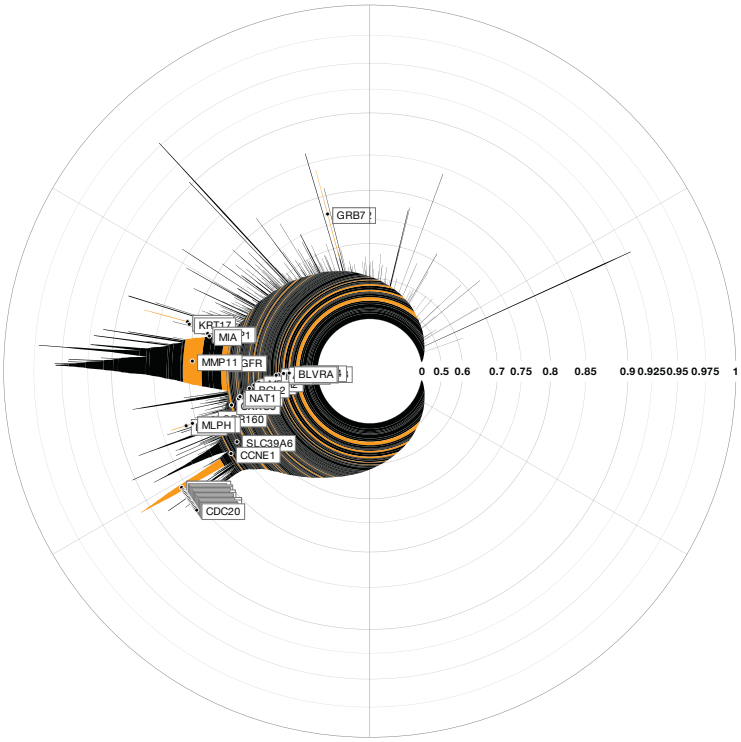

ROPN1B
ROPN1
MIA
FOXC1
SOX10
GABR
FAM171A1
SFRP1
RGMA

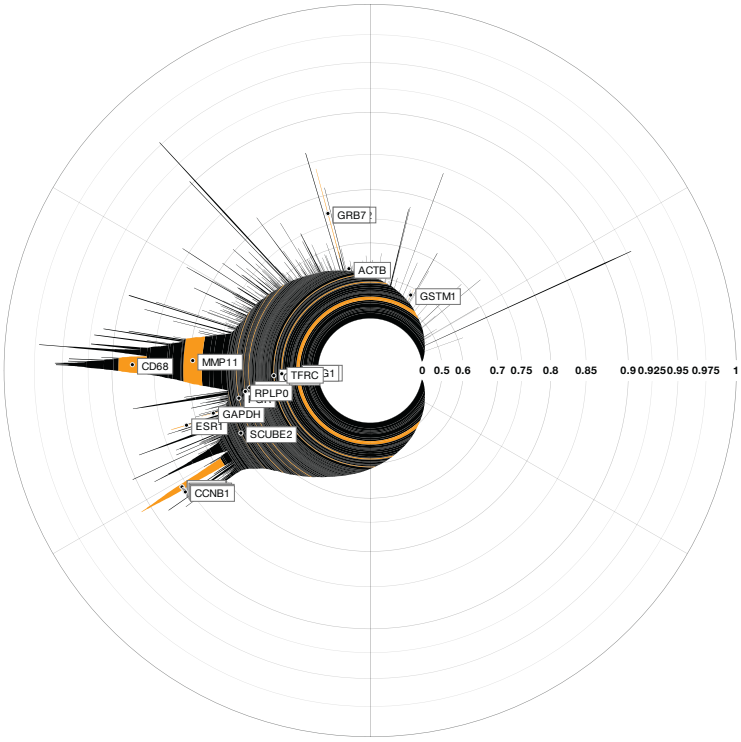
Fig S6

A



PAM50

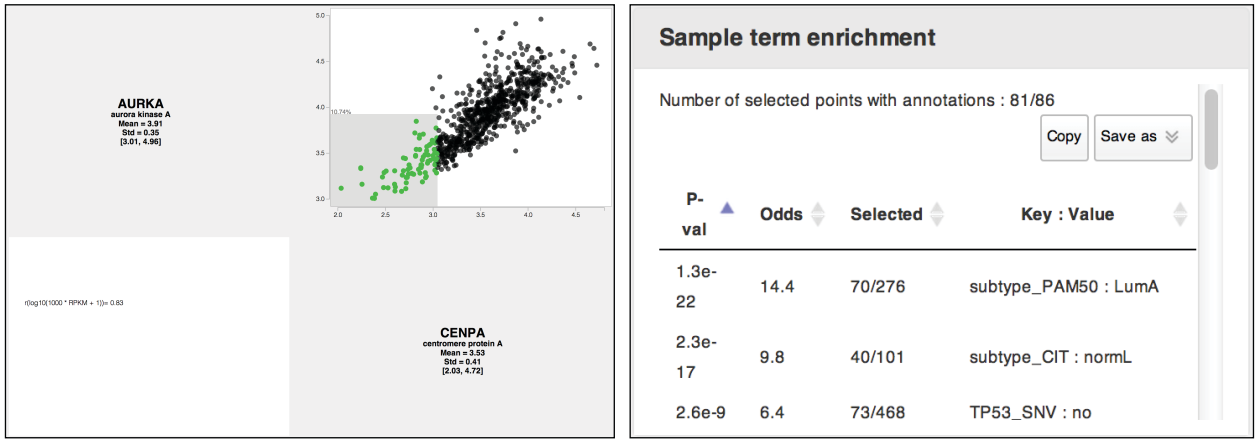
B



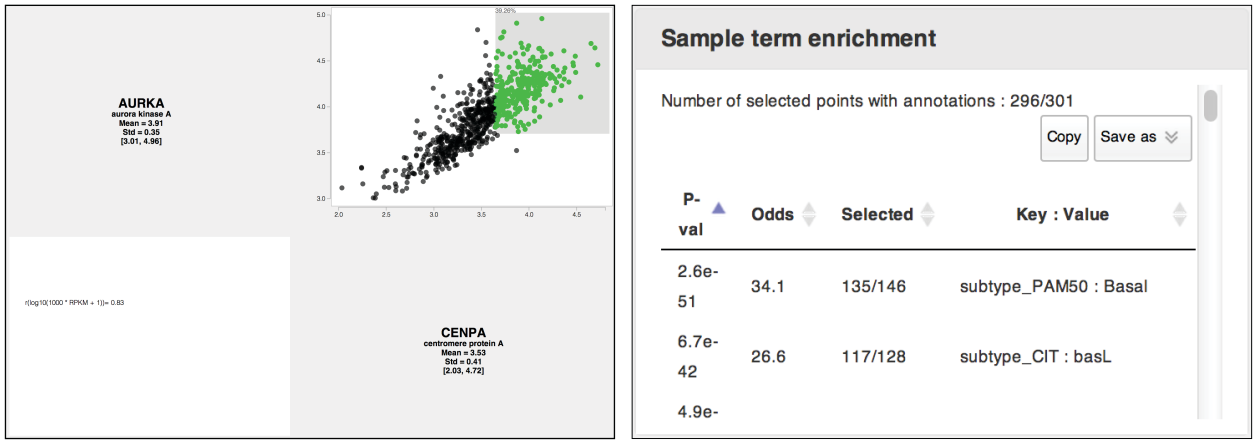
Oncotype Dx

Fig S7

A


AURKA
aurora kinase A
Mean = 3.91
Std = 0.35
[3.01, 4.96]
r(log10(1000 * RPKM + 1))= 0.83
CENPA
centromere protein A
Mean = 3.53
Std = 0.41
[2.03, 4.72]
Sample term enrichment
Number of selected points with annotations : 81/86
Copy
Save as
P-val
Odds
Selected
Key : Value
1.3e-22
14.4
70/276
subtype_PAM50 : LumA
2.3e-17
9.8
40/101
subtype_CIT : normL
2.6e-9
6.4
73/468
TP53_SNV : no

B


AURKA
aurora kinase A
Mean = 3.91
Std = 0.35
[3.01, 4.96]
r(log10(1000 * RPKM + 1))= 0.83
CENPA
centromere protein A
Mean = 3.53
Std = 0.41
[2.03, 4.72]
Sample term enrichment
Number of selected points with annotations : 296/301
Copy
Save as
P-val
Odds
Selected
Key : Value
2.6e-51
34.1
135/146
subtype_PAM50 : Basal
6.7e-42
26.6
117/128
subtype_CIT : basL
4.9e-

C


Datasets (1)
BRCA_tumour_TCGA_maseq
Choose dataset
Genes (2)
Select a gene
AURKA
CENPA
Samples (712)
Highlight groups
Luminal A
TCGA-A1-A0SD-01A-11R
Luminal B
TCGA-A1-A0SN-01A-11R
HER2+
TCGA-A1-A0SI-01A-11R-
Basal-like
TCGA-A1-A0SK-01A-12R
New group
subtype_PAM50 : Basal
More options
Sample term enrichment
AURKA
aurora kinase A
Mean = 3.91
Std = 0.35
[3.01, 4.96]
r(log10(1000 * RPKM + 1))= 0.83
CENPA
centromere protein A
Mean = 3.53
Std = 0.41
[2.03, 4.72]

Fig S8

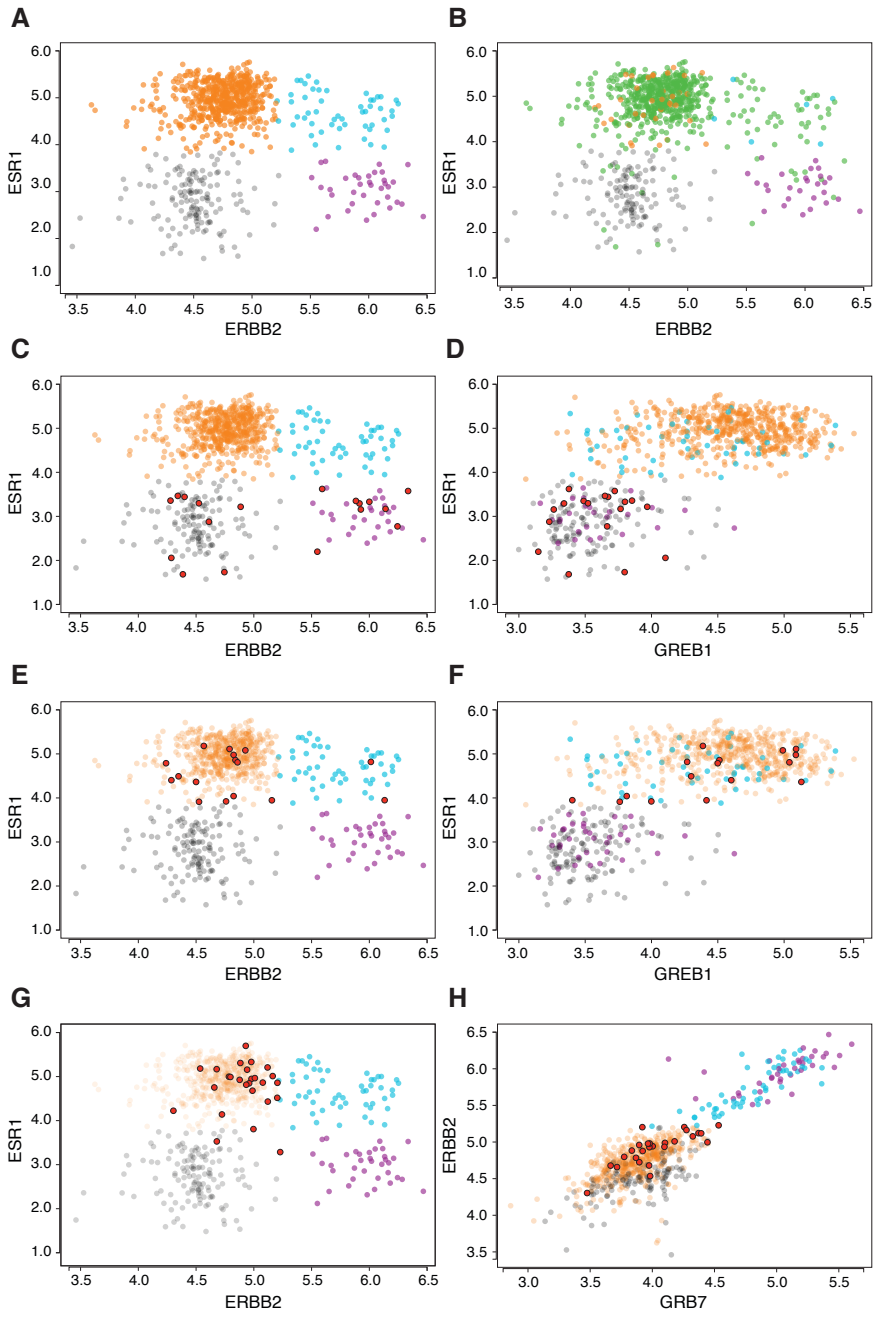
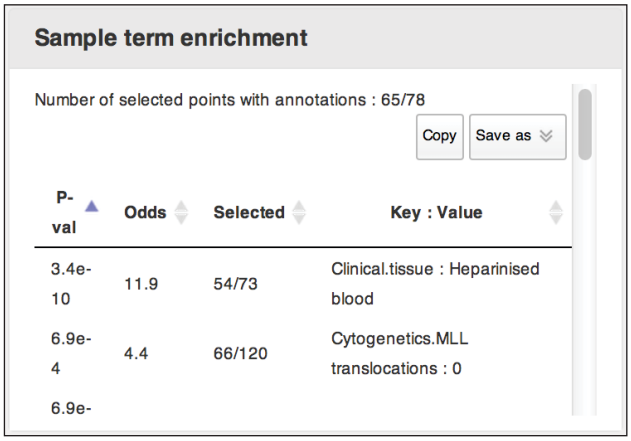


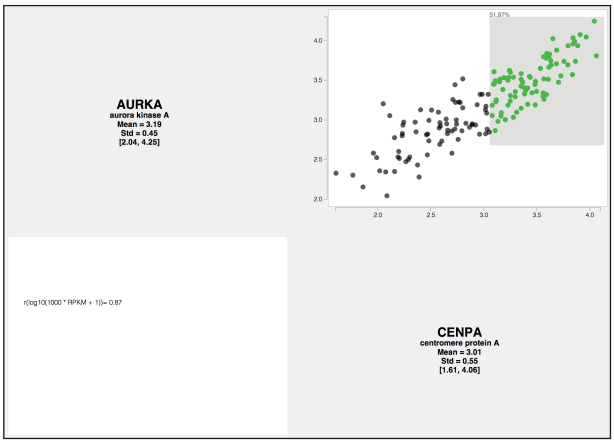
Fig S9

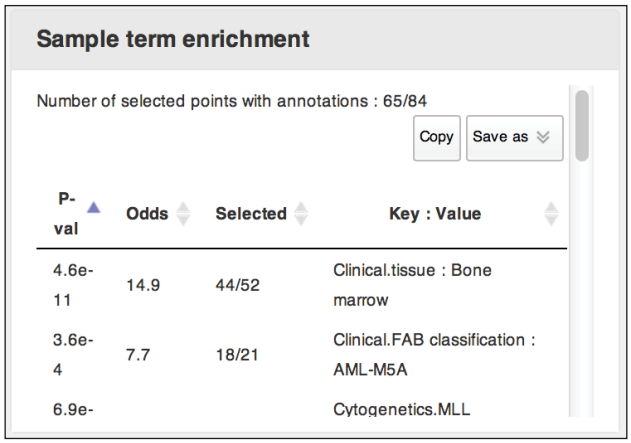
A


AURKA
aurora kinase A
Mean = 3.19
Std = 0.45
[2.04, 4.25]
CENPA
centromere protein A
Mean = 3.01
Std = 0.55
[1.61, 4.06]

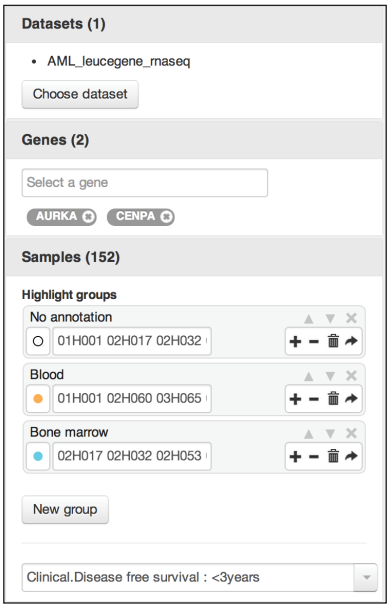

Sample term enrichment
Number of selected points with annotations : 65/78
Copy
Save as
P-val
Odds
Selected
Key : Value
3.4e-10
11.9
54/73
Clinical.tissue : Heparinised blood
6.9e-4
4.4
66/120
Cytogenetics.MLL translocations : 0
6.9e-

B


AURKA
aurora kinase A
Mean = 3.19
Std = 0.45
[2.04, 4.25]
CENPA
centromere protein A
Mean = 3.01
Std = 0.55
[1.61, 4.06]


Sample term enrichment
Number of selected points with annotations : 65/84
Copy
Save as
P-val
Odds
Selected
Key : Value
4.6e-11
14.9
44/52
Clinical.tissue : Bone marrow
3.6e-4
7.7
18/21
Clinical.FAB classification : AML-M5A
6.9e-
Cytogenetics.MLL

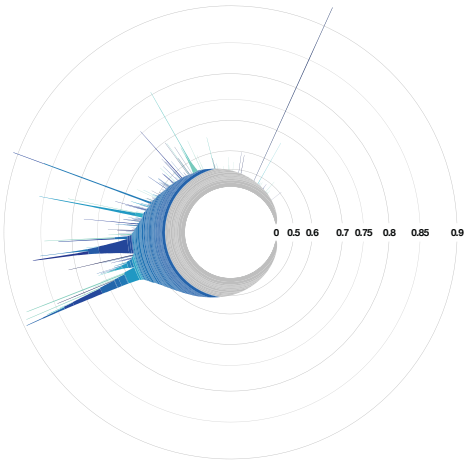
C


Datasets (1)
AML_leucegene_rnaseq
Choose dataset
Genes (2)
Select a gene
AURKA
CENPA
Samples (152)
Highlight groups
No annotation
01H001 02H017 02H032
Blood
01H001 02H060 03H065
Bone marrow
02H017 02H032 02H053
New group
Clinical.Disease free survival : <3years


AURKA
aurora kinase A
Mean = 3.19
Std = 0.45
[2.04, 4.25]
CENPA
centromere protein A
Mean = 3.01
Std = 0.55
[1.61, 4.06]

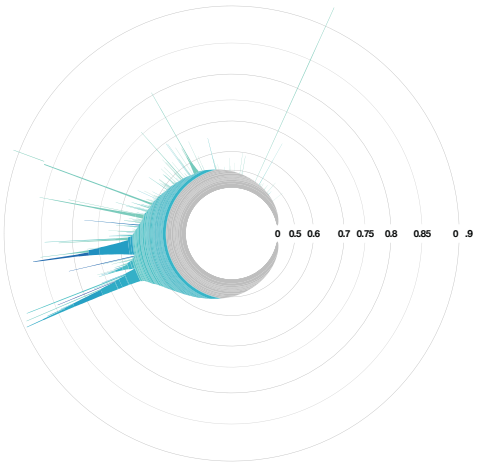
Fig S10

A



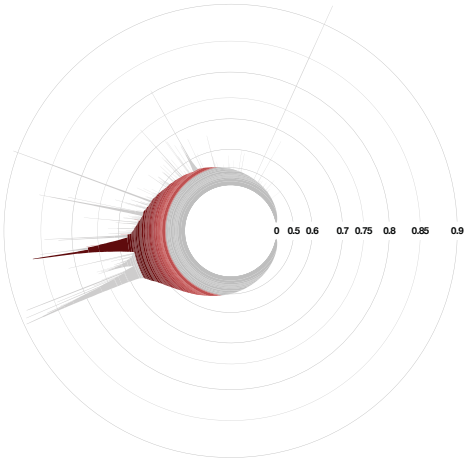
CCLE
GNE

B

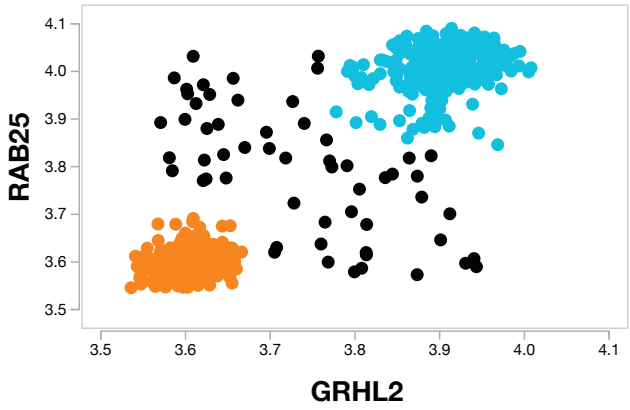


CCLE
Breast, tumour (TCGA)

C



D



E

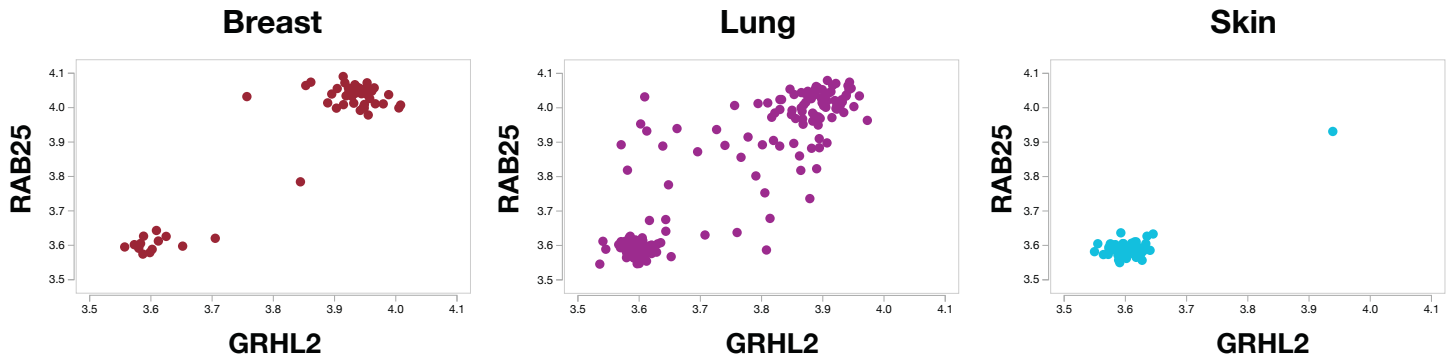


Fig S11

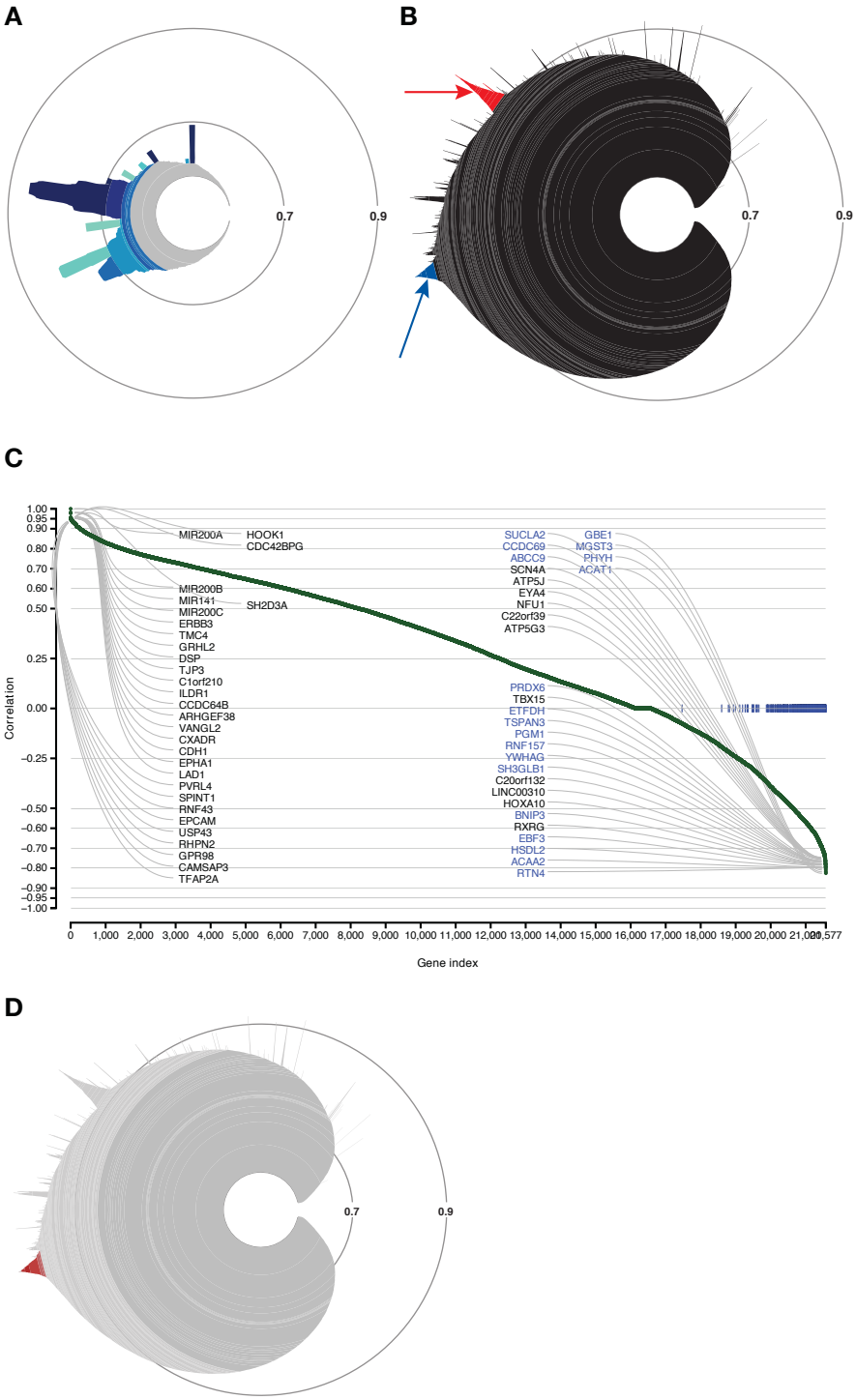

A
B
0.7
0.9
0.7
0.9
C
Correlation
Gene index
MIR200A
HOOK1
CDC42BPG
MIR200B
MIR141
MIR200C
ERBB3
TMC4
GRHL2
DSP
TJP3
C1orf210
ILDR1
CCDC64B
ARHGEF38
VANGL2
CXADR
CDH1
EPHA1
LAD1
PVRL4
SPINT1
RNF43
EPCAM
USP43
RHPN2
GPR98
CAMSAP3
TFAP2A
SH2D3A
SUCLA2
CCDC69
ABCC9
SCN4A
ATP5J
EYA4
NFU1
C22orf39
ATP5G3
GBE1
MGST3
PHYH
ACAT1
PRDX6
TBX15
ETFDH
TSPAN3
PGM1
RNF157
YWHAG
SH3GLB1
C20orf132
LINC00310
HOXA10
BNIP3
RXRG
EBF3
HSDL2
ACAA2
RTN4
D
0.7
0.9

Fig S12

Supp Table 1: Lists of datasets

Project and Dataset names	Dataset id	Tissue/Type of Cancer	Sub category	Source	Nb. Specimens	Technology	Description
AML Leucegene RNASeq	1	AML	Various karyotypes	BCLQ	152	RNASeq	Gene Expresion Omnibus (GSE49642, GSE52656, GSE62190, GSE67040, GSE67039
AML NK Leucegene RNASeq	2	AML	Normal karyotype	BCLQ	48	RNASeq	Gene Expresion Omnibus (GSE49642, GSE52656, GSE62190, GSE67040, GSE67039
LAML tumor TCGA RNASeq	3	AML	Various karyotypes	TCGA	179	RNASeq	PMID: 23634996
LAML tumor NK TCGA RNASeq	4	AML	Normal karyotype	TCGA	89	RNASeq	PMID: 23634996
CD34+/CD34+CD45RA- Leucegene RNASeq	5	Cord blood	CD34+	Héma-Qc	17	RNASeq	Gene Expresion Omnibus (GSE48846)
BRCA normal TCGA RNASeq	6	Breast	Normal	TCGA	102	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq	7	Breast	Cancer	TCGA	754	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq ER+	8	Breast	Cancer - ER+	TCGA	553	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq ER-	9	Breast	Cancer - ER-	TCGA	165	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq HER2+	10	Breast	Cancer - HER2+	TCGA	105	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq HER2-	11	Breast	Cancer - HER2-	TCGA	606	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq triple negative	12	Breast	Cancer - Triple -	TCGA	117	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq not triple negative	13	Breast	Cancer - Not triple -	TCGA	637	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq luminal A	14	Breast	Cancer - Luminal A	TCGA	276	RNASeq	PMID: 23000897
BRCA tumor TCGA RNASeq luminal B	15	Breast	Cancer - Luminal B	TCGA	192	RNASeq	PMID: 23000897
BRCA miRNA normal	16	Breast	Normal	TCGA	87	RNASeq	PMID: 23000897
BRCA miRNA tumor	17	Breast	Cancer	TCGA	659	RNASeq	PMID: 23000897
BRCA miRNA coding normal	18	Breast	Normal	TCGA	80	RNASeq	PMID: 23000897
BRCA miRNA coding tumor	19	Breast	Cancer	TCGA	440	RNASeq	PMID: 23000897
KIRC normal TCGA RNASeq	20	Kidney	Normal	TCGA	68	RNASeq	PMID: 23792563
KIRC tumor TCGA RNAseq	21	Kidney	Cancer	TCGA	470	RNASeq	PMID: 23792563
LUAD normal TCGA RNASeq	22	Lung	Normal	TCGA	37	RNASeq	PMID: 22980975
LUAD tumor TCGA RNASeq	23	Lung	Cancer	TCGA	129	RNASeq	PMID: 22980975
COAD tumor TCGA RNASeq	24	Colon	Cancer	TCGA	192	RNASeq	PMID: 22810696
CCLE	25	Cell lines	Cancer	CCLE	917	Affymetrix	PMID: 22460905
GNE_RPKM	26	Cell lines	Cancer	GDSC	675	RNASeq	PMID: 25485619
GNE_VSD	27	Cell lines	Cancer	GDSC	675	RNASeq	PMID: 25485619

Supp. Table 2. Patient annotations.

BCLQ ID	Exome sequenced (HiSeq)	Transcriptome sequenced (HiSeq)	Chemical screen	DX (WHO)	DX (short)	DX (FAB)	Age	Sex	Karyotype (seq)	Normal karyotype	Blasts (%)	Tissue	WBC x 10^9/ L	NPM1 (exon Mut2)	FLT-3-ITD	CD34 (%)	Total reads	Passing filters	Mapped reads	Basepairs	% Mapped reads	Mean Exon RPKM
02H053	1	1	0	AML without maturation	AML	M1	60	M	46,XY[20]	1	96%	Bone Marrow	99.0	Mut	Mut	0	2.54E+08	2.44E+08	1.66E+08	1.38E+10	68%	19.72
02H066	1	1	0	AML without maturation	AML	M1	53	F	46,XX[22]	1	95%	Bone Marrow	125.0	Mut	Mut	2	2.02E+08	1.90E+08	1.38E+08	1.10E+10	73%	18.76
03H041	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5	47	F	46,XX[22]	1	83%	Bone Marrow	102.0	WT	Mut	28	1.39E+08	1.34E+08	9.89E+07	8.31E+09	74%	18.52
03H116	1	1	0	AML without maturation	AML	M1	55	F	46,XX[21]	1	97%	Bone Marrow	120.0	Mut	Mut	1	2.10E+08	2.03E+08	1.62E+08	1.16E+10	80%	14.71
03H119	1	1	0	AML without maturation	AML	M1	68	M	46,XY[20]	1	92%	Bone Marrow	6.4	WT	WT	98	2.40E+08	2.27E+08	1.70E+08	1.35E+10	75%	17.60
04H024	1	1	0	AML without maturation	AML	M1	45	F	46,XX[21]	1	76%	Bone Marrow	231.0	Mut	WT	8	2.36E+08	2.19E+08	1.69E+08	1.39E+10	77%	17.76
04H112	1	1	0	AML without maturation	AML	M1	64	F	46,XX[21]	1	91%	Bone Marrow	361.2	Mut	Mut	36	3.14E+08	2.90E+08	2.11E+08	1.75E+10	73%	18.73
04H133	1	1	0	AML without maturation	AML	M1	62	F	46,XX[20]	1	91%	Bone Marrow	60.3	Mut	Mut	1	2.54E+08	2.39E+08	1.85E+08	1.48E+10	77%	18.75
05H050	1	1	0	Acute myelomonocytic leukaemia	AML	M4	61	M	46,XY[20]	1	94%	Bone Marrow	80.6	WT	Mut	10	2.44E+08	2.13E+08	1.62E+08	1.31E+10	76%	18.09
05H094	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	35	M	46,XY[23]	1	94%	Bone Marrow	68.4	Mut	WT	24	2.41E+07	2.25E+07	1.61E+07	1.19E+09	72%	15.49
05H149	1	1	0	AML without maturation	AML	M1	55	M	46,XY[20]	1	80%	Blood	30.5	WT	WT	98	1.35E+08	1.17E+08	8.95E+07	6.36E+09	76%	15.38
05H163	1	1	0	AML without maturation	AML	M1	20	M	46,XY[22]	1	86%	Blood	136.0	WT	WT	96	1.31E+08	1.29E+08	1.03E+08	7.46E+09	80%	15.77
05H181	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	65	F	46,XX[11]	1	80%	Blood	77.3	Mut	Mut	2	1.57E+08	1.51E+08	1.18E+08	9.39E+09	78%	16.78
06H028	1	1	0	AML without maturation	AML	M1	70	F	46,XX[20]	1	95%	Bone Marrow	258.8	Mut	WT	2	2.40E+08	2.32E+08	1.92E+08	1.27E+10	83%	13.40
06H045	1	1	1	AML with maturation	AML	M2	30	F	46,XX[22]	1	70%	Bone Marrow	16.6	WT	WT	70	9.58E+07	9.41E+07	6.89E+07	5.07E+09	73%	15.74
06H144	1	1	0	AML without maturation	AML	M1	71	F	46,XX[20]	1	90%	Blood	56.1	Mut	WT	2	2.75E+08	2.62E+08	2.09E+08	1.34E+10	80%	11.95
07H042	1	1	1	Acute myeloid leukaemia, NOS	AML	NC	23	M	46,XY[20]	1	83%	Blood	69.0	Mut	Mut	7	1.40E+08	1.37E+08	1.06E+08	7.88E+09	77%	16.67
07H062	1	1	0	AML without maturation	AML	M1	58	M	46,XY[20]	1	90%	Blood	105.0	Mut	Mut	28	1.53E+08	1.51E+08	1.22E+08	8.82E+09	81%	16.51
07H135	1	1	0	AML without maturation	AML	M1	67	M	46,XY[20]	1	97%	Blood	106.7	Mut	Mut	2	2.38E+08	2.24E+08	1.79E+08	1.32E+10	80%	15.91
08H048	1	1	1	AML without maturation	AML	M1	45	M	46,XY[21]	1	96%	Blood	127.7	WT	WT	41	2.20E+08	2.14E+08	1.59E+08	1.32E+10	74%	18.28
08H112	1	1	1	AML with myelodysplasia-related changes	AML	NC	52	M	46,XY[20]	1	85%	Blood	28.3	WT	WT	91	2.46E+08	2.31E+08	1.65E+08	1.23E+10	72%	16.98
09H031	1	1	1	AML without maturation	AML	M1	54	F	46,XX[20]	1	85%	Blood	48.9	Mut	Mut	10	2.39E+08	2.32E+08	1.65E+08	1.31E+10	71%	17.83
09H043	1	1	0	AML without maturation	AML	M1	53	M	46,XY[21]	1	80%	Bone Marrow	33.2	Mut	Mut	8	2.00E+08	1.91E+08	1.49E+08	1.16E+10	78%	17.48
09H083	1	1	0	AML without maturation	AML	M1	63	F	46,XX[20]	1	94%	Blood	65.3	Mut	Mut	5	2.73E+08	2.56E+08	2.10E+08	1.21E+10	82%	13.76
09H111	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	59	F	46,XX[21]	1	80%	Blood	46.8	Mut	WT	2	1.98E+08	1.90E+08	1.54E+08	1.22E+10	81%	16.14
09H113	1	1	1	AML without maturation	AML	M1	56	M	46,XY[22]	1	95%	Bone Marrow	68.7	WT	WT	96	2.02E+08	1.91E+08	1.54E+08	9.47E+09	81%	14.54
09H115	1	1	0	AML without maturation	AML	M1	46	M	46,XY[24]	1	90%	Blood	101.0	WT	Mut	3	1.78E+08	1.72E+08	1.40E+08	8.64E+09	82%	15.00
10H031	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	25	F	46,XX[27]	1	73%	Blood	184.6	WT	WT	47	2.94E+08	2.77E+08	2.28E+08	1.60E+10	82%	14.51
10H038	1	1	0	AML with minimal differentiation	AML	M0	67	F	46,XX[20]	1	91%	Blood	147.8	WT	WT	99	2.78E+08	2.61E+08	2.04E+08	1.28E+10	78%	13.98
10H052	1	1	0	Acute myeloid leukaemia, NOS	AML	NC	60	F	46,XX[20]	1	66%	Blood	6.6	Mut	Mut	2	2.46E+08	2.35E+08	1.56E+08	1.04E+10	67%	15.41
10H056	1	1	0	AML without maturation	AML	M1	57	F	46,XX[18]	1	83%	Bone Marrow	6.3	WT	WT	97	1.49E+08	1.45E+08	1.10E+08	8.34E+09	76%	20.04
10H072	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	69	M	46,XY[20]	1	77%	Bone Marrow	23.5	Mut	WT	1	2.00E+08	1.92E+08	1.61E+08	1.08E+10	84%	15.71
10H089	1	1	0	Acute myeloid leukaemia, NOS	AML	NC	49	F	46,XX[26]	1	80%	Blood	4.0	WT	WT	11	3.45E+08	3.32E+08	2.53E+08	1.62E+10	76%	14.41
10H092	1	1	0	AML without maturation	AML	M1	69	F	46,XX[21]	1	90%	Bone Marrow	72.9	Mut	Mut	0	1.32E+08	1.11E+08	8.65E+07	6.37E+09	78%	16.14
10H095	1	1	0	AML without maturation	AML	M1	65	F	46,XX[24]	1	91%	Blood	52.0	Mut	Mut	0	1.08E+08	1.00E+08	8.09E+07	5.47E+09	81%	13.64
10H101	1	1	0	AML with myelodysplasia-related changes	AML	M1	49	F	46,XX[22]	1	70%	Bone Marrow	24.5	Mut	Mut	1	1.87E+08	1.80E+08	1.42E+08	8.53E+09	79%	15.06
10H115	1	1	0	AML without maturation	AML	M1	43	M	46,XY[23]	1	88%	Bone Marrow	21.6	Mut	Mut	0	2.33E+08	2.22E+08	1.76E+08	1.06E+10	79%	14.41
10H166	1	1	0	Acute myelomonocytic leukaemia	AML	M4	63	M	46,XY[20]	1	89%	Blood	226.2	Mut	Mut	1	4.73E+07	4.51E+07	3.61E+07	2.62E+09	80%	15.57
11H006	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5A	26	F	46,XX[23]	1	94%	Bone Marrow	33.0	Mut	WT	1	1.97E+08	1.66E+08	1.36E+08	1.09E+10	82%	17.39
11H009	1	1	0	AML with maturation	AML	M2	53	M	46,XY[20]	1	70%	Blood	17.6	WT	Mut	6	1.26E+08	1.22E+08	9.75E+07	5.78E+09	80%	14.50
11H021	1	1	0	AML with maturation	AML	M2	21	F	46,XX[20]	1	70%	Blood	23.1	WT	Mut	97	9.90E+07	9.30E+07	7.20E+07	4.99E+09	77%	15.03
11H058	0	1	0	AML without maturation	AML	M1	58	M	46,XY[20]	1	90%	Bone Marrow	1.6	Mut	Mut	1	2.13E+08	2.07E+08	1.59E+08	1.23E+10	77%	17.48
11H072	1	1	0	AML with maturation	AML	M2	52	F	46,XX[20]	1	80%	Blood	32.2	Mut	Mut	0	1.54E+08	1.47E+08	1.16E+08	7.69E+09	79%	14.50
11H083	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5A	52	M	46,XY[20]	1	80%	Blood	77.1	Mut	Mut	0	1.48E+08	1.41E+08	1.09E+08	7.83E+09	78%	15.93
11H095	1	1	0	Therapy-related myeloid neoplasms	t-AML	M5A	50	M	46,XY[20]	1	87%	Bone Marrow	1.4	WT	WT	0	8.47E+07	8.03E+07	6.40E+07	4.99E+09	80%	17.71
11H126	1	1	0	Acute monoblastic and monocytic leukaemia	AML	M5B	74	M	46,XY[21]	1	68%	Bone Marrow	12.9	Mut	WT	3	1.16E+08	1.10E+08	9.08E+07	7.00E+09	83%	16.14
11H142	1	1	0	AML without maturation	AML	M1	52	F	46,XX[21]	1	96%	Blood	119.2	Mut	Mut	0	1.82E+08	1.75E+08	1.42E+08	8.63E+09	81%	14.50
11H151	1	1	1	AML without maturation	AML	M1	61	M	46,XY[21]	1	78%	Blood	53.3	Mut	Wt	5	2.40E+08	2.33E+08	1.77E+08	1.30E+10	76%	15.57
11H160	1	1	0	Acute myelomonocytic leukaemia	AML	M4	36	F	46,XX[22]	1	65%	Blood	84.8	WT	Mut	56	3.16E+08	3.02E+08	2.48E+08	1.93E+10	82%	16.44
12H030	1	1	1	AML with minimal differentiation	AML	M0	78	M	46,XY[20]	1	93%	Blood	144.8	WT	WT	94	2.36E+08	2.29E+08	1.76E+08	1.29E+10	77%	16.15

DX (FAB): NC, Not classifiable by FAB criteria

Supplementary Table 3. Genesets used for enrichment analysis in MiSTIC

geneset type	description	# genesets	website	version	main references (PMID)
chipseq	ChIPseq regions within 25kb, 10kb or 5kb of transcription start site of refseq genes and mirna	291	http://www.pazar.info/ http://genome.ucsc.edu/ENCODE/		18971253 22955616 23953112 23375374
Chromosome	chromosome and cytoband for refseq genes and mirna (UCSC cytoBand, refGene and wgRna annotation tracks)	841	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/		24270787 12700172 15608248 17991681 21527027 21720365 22895193 22960745 22980975 23000897
CNV	TCGA Copy Number Analysis	469	http://gdac.broadinstitute.org/runs/analyses__2014_01_15/		23622249 23634996
CPDB	Consensus-PathDB database integrating human molecular interaction networks (protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target)	4576	http://ConsensusPathDB.org http://cpdb.molgen.mpg.de/CPDB/	27 (12.08.2013)	23143270
GeneFamilies	Gene Families / Transcription Factor	1383	http://www.genenames.org/ http://tfclass.bioinf.med.uni-goettingen.de/tfclass		23245209 23180794
GeneSigDB	Gene Signature DataBase	3485	http://compbio.dfci.harvard.edu/genesigdb/	4	22110038
GO	Gene Ontology	12600	http://www.geneontology.org/		10802651
microarrays	Gene sets extracted from microarray experiments	85			
miRNA	miRecords miRNA-target interactions	1257	http://mirecords.biolead.org/	4	18996891
MSigDB	Molecular Signatures Database	8513	http://www.broadinstitute.org/gsea/msigdb/index.jsp	3.1	16199517
Target genes	Target genes sets	47	http://ctdbase.org/ http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=searchTFGeneForm		17202159 23093600 20887958
TFBS	Transcription factor binding site predictions within 10kb, 5kb or 2.5kb of transcription start site of refseq genes and mirna	2568			17986456 16381825

chipseq

Cat	Name	pmid
TF	AR LNCaP cells DHT	18971253
TF	ARNT LoVo cells	23953112
TF	ATF3 K562 cells	22955616
TF	BCL3 K562 cells	22955616
TF	BCLAF1 K562 cells	22955616
TF	CBX3 K562 cells	22955616
TF	CEBPB K562 cells	22955616
TF	CEBPB MCF7 cells	22955616
TF	ChIP-chip AR MCF7 cells	23375374
TF	ChIP-chip SPDEF MCF7 cells	23375374
TF	ChIP-chip XBP1 MCF7 cells	23375374
TF	CREB1 K562 cells	22955616
TF	CREB1 LoVo cells	23953112
TF	CREB3L4 LoVo cells	23953112
TF	CTCF K562 cells	22955616
TF	CTCF MCF7 cells	22955616
TF	CTCFL K562 cells	22955616
TF	E2f1 HeLa-S3 cells	22955616
TF	E2F4 Human mesenchymal stem cells	18971253
TF	E2F6 K562 cells	22955616
TF	EGR1 K562 cells	22955616
TF	EGR1 K562 cells PMA	18971253
TF	EGR1 MCF7 cells	22955616
TF	ELF1 K562 cells	22955616
TF	ELF1 MCF7 cells	22955616
TF	ESR1 LoVo cells	23953112
TF	ESR1 MCF7 cells E2	18971253
TF	ETS1 K562 cells	22955616
TF	ETS1 RWPE-1 cells	18971253
TF	FOS MCF7 cells E2	21179027
TF	FOSL1 K562 cells	22955616
TF	FOSL2 MCF7 cells	22955616
TF	FOXA1 MCF7 cells E2	18971253
TF	FOXM1 MCF7 cells	22955616
TF	GABPA HL60 cells	22955616
TF	GABPA K562 cells	22955616
TF	GABPA MCF7 cells	22955616
TF	GATA2 K562 cells	22955616
TF	GATA3 MCF7 cells	22955616
TF	GATA3 MCF7 cells E2	18971253
TF	GR A549 cells 100nM DEX	22955616
TF	GRHL2 primary bronchial epithelial cells	23690579
TF	HEY1 K562 cells	22955616
TF	HOXA1 LoVo cells	23953112
TF	HOXA10 LoVo cells	23953112
TF	JUND MCF7 cells	22955616
TF	KLF4 embryonic stem cells	18971253
TF	MAX K562 cells	22955616
TF	MAX MCF7 cells	22955616
TF	MCF7 cells RARA	20080953
TF	MEF2 K562 cells	22955616
TF	MYB LoVo cells	23953112
TF	Mybl2 HepG2 cells	22955616
TF	MYBL2 LoVo cells	23953112
TF	Nfkb GM12891 cells	22955616
TF	NFKB2 LoVo cells	23953112
TF	NKX2.2 LoVo cells	23953112
TF	NKX31 LNCaP cells	18971253
TF	NR2F2 K562 cells	22955616
TF	NR2F2 MCF7 cells	22955616
TF	NRSF HL60 cells	22955616
TF	NRSF K562 cells	22955616
TF	NRSF MCF7 cells	22955616
TF	PBX1 MCF7 cells	18971253
TF	PML K562 cells	22955616
TF	PML MCF7 cells	22955616
TF	PU.1 HL60 cells	22955616
TF	PU.1 K562 cells	22955616
TF	RAD21 K562 cells	22955616
TF	RAD21 MCF7 cells	22955616
TF	RARG LoVo cells	23953112
TF	Rxra HepG2 cells	22955616

TF	SIN3AK20 K562 cells	22955616
TF	SIN3AK20 MCF7 cells	22955616
TF	SIX5 K562 cells	22955616
TF	SOX2 LoVo cells	23953112
TF	SP1 K562 cells	22955616
TF	SP2 K562 cells	22955616
TF	SPDEF MCF7 cells	24043118
TF	SRF K562 cells	22955616
TF	SRF MCF7 cells	22955616
TF	STAT5A K562 cells	22955616
TF	TAF1 K562 cells	22955616
TF	TAF1 MCF7 cells	22955616
TF	TAF7 K562 cells	22955616
TF	TCF12 MCF7 cells	22955616
TF	TEAD4 K562 cells	22955616
TF	TEAD4 MCF7 cells	22955616
TF	TFAP2C MCF7 cells	18971253
TF	THAP1 K562 cells	22955616
TF	TP53 IMR90 cells	18971253
TF	TR4 HepG2 cells	22955616
TF	TRIM28 K562 cells	22955616
TF	USF1 K562 cells	22955616
TF	YY1 K562 cells	22955616
TF	ZBTB33 K562 cells	22955616
TF	ZBTB7A K562 cells	22955616

microarrays

Cat	Name	pmid
Breast cancer cell lines	down MCF7 12h vs 0h E2	17013392
Breast cancer cell lines	down MCF7 24h E2	17986456
Breast cancer cell lines	down MCF7 24h E2 CHX	17986456
Breast cancer cell lines	down MCF7 3h vs 0h E2	17013392
Breast cancer cell lines	down MCF7 6h vs 0h E2	17013392
Breast cancer cell lines	down MCF7 CHX RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	down MCF7 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	down MCF7 TSA 24h	Ismail H et al. in prep.
Breast cancer cell lines	down MCF7 TSA 3h	Ismail H et al. in prep.
Breast cancer cell lines	down siESR1 MCF7	22174013
Breast cancer cell lines	down siFoxA1 MCF7	21151129
Breast cancer cell lines	down siFoxA1 MCF7 E2	21151129
Breast cancer cell lines	down siGATA3 MCF7	23172872
Breast cancer cell lines	down siGATA3 MCF7 E2	23172872
Breast cancer cell lines	down SKBR3 CHX RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 FOXA1 overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 FOXO3A overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 GATA3 overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 Herceptin 24h	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 RA 24h	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 RA+Herceptin 24h	Rozendaal M et al. in prep.
Breast cancer cell lines	down SKBR3 SOX9	Rozendaal M et al. in prep.
Breast cancer cell lines	reg MCF7 E2 24h	17986456
Breast cancer cell lines	reg MCF7 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	reg SKBR3 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	up MCF7 12h vs 0h E2	17013392
Breast cancer cell lines	up MCF7 24h E2	17986456
Breast cancer cell lines	up MCF7 24h E2 CHX	17986456
Breast cancer cell lines	up MCF7 3h vs 0h E2	17013392
Breast cancer cell lines	up MCF7 6h vs 0h E2	17013392
Breast cancer cell lines	up MCF7 CHX RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	up MCF7 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	up MCF7 TSA 24h	Ismail H et al. in prep.
Breast cancer cell lines	up MCF7 TSA 3h	Ismail H et al. in prep.
Breast cancer cell lines	up siESR1 MCF7	22174013
Breast cancer cell lines	up siFoxA1 MCF7	21151129
Breast cancer cell lines	up siFoxA1 MCF7 E2	21151129
Breast cancer cell lines	up siGATA3 MCF7	23172872
Breast cancer cell lines	up siGATA3 MCF7 E2	23172872
Breast cancer cell lines	up SKBR3 CHX RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 ELF3 overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 FOXA1 overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 FOXO3A overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 GATA3 overexpression	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 Herceptin	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 RA 24h	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 RA 8h	Rozendaal M et al. in prep.
Breast cancer cell lines	up SKBR3 RA+Herceptin 24h	Rozendaal M et al. in prep.
Breast cancer cell lines	up SOX9 SKBR3 overexpression	Rozendaal M et al. in prep.
Breast Cancer Co-Expressior	module #1 ER: estrogen signaling	24516633
Breast Cancer Co-Expressior	module #10 ECM : extracellular matrix	24516633
Breast Cancer Co-Expressior	module #11 Prolif: cell proliferation	24516633
Breast Cancer Co-Expressior	module #2 Dev/Basal: development/differentiation	24516633
Breast Cancer Co-Expressior	module #3 Immune	24516633
Breast Cancer Co-Expressior	module #4 Immune	24516633
Breast Cancer Co-Expressior	module #5 Immune	24516633
Breast Cancer Co-Expressior	module #6 Histone	24516633
Breast Cancer Co-Expressior	module #7 ERBB2	24516633
Breast Cancer Co-Expressior	module #8 Mixed	24516633
Breast Cancer Co-Expressior	module #9 ECM/Dev/Immune	24516633
Signature	Breast Cancer ENDOPREDICT	21807638
Signature	Breast Cancer GGI	16478745
Signature	Breast Cancer Mammaprint	11823860
Signature	Breast Cancer molecular subtype [CIT]	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster a (basL)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster b (cycle/proliferation)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster c (lumA)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster d (AR)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster e1 (lumB)	21785460

Signature	Breast Cancer molecular subtype [CIT] - cluster e2 (ESR1)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster f (lumA/normL)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster g (mApo)	21785460
Signature	Breast Cancer molecular subtype [CIT] - cluster h (normL)	21785460
Signature	Breast Cancer molecular subtype [claudin-low vs others]	20813035
Signature	Breast Cancer molecular subtype [PAM50]	19204204
Signature	Breast Cancer molecular subtype [TNBC]	21633166
Signature	Breast Cancer molecular subtype [TNBC] - BASAL -LIKE 2 (cluster 3)	21633166
Signature	Breast Cancer molecular subtype [TNBC] - BASAL-LIKE1 (cluster 2)	21633166
Signature	Breast Cancer molecular subtype [TNBC] - IMMUNOMODULATORY (clus	21633166
Signature	Breast Cancer molecular subtype [TNBC] - LUMINAL AR (cluster 7)	21633166
Signature	Breast Cancer molecular subtype [TNBC] - MESENCHYMAL (cluster 5)	21633166
Signature	Breast Cancer molecular subtype [TNBC] - MESENCHYMAL STEM-LIKE (c	21633166
Signature	Breast Cancer molecular subtype [TNBC] - UNSTABLE (cluster 1)	21633166
Signature	Breast Cancer OncotypeDx	15591335

target genes

Cat	Name	pmid
CTD	Genes downregulated by Dihydrotestosterone (DHT) from the Compara	23093600
CTD	Genes downregulated by Doxorubicin from the Comparative Toxicogen	23093600
CTD	Genes downregulated by estrogen from the Comparative Toxicogenomi	23093600
CTD	Genes downregulated by retinoic acid from the Comparative Toxicogen	23093600
CTD	Genes downregulated by Tamoxifen from the Comparative Toxicogenor	23093600
CTD	Genes downregulated by TSA from the Comparative Toxicogenomics Da	23093600
CTD	Genes downregulated by Vitamin D (Calcitriol) from the Comparative Tc	23093600
CTD	Genes regulated by Dihydrotestosterone (DHT) from the Comparative T	23093600
CTD	Genes regulated by Doxorubicin from the Comparative Toxicogenomics	23093600
CTD	Genes regulated by estrogen from the Comparative Toxicogenomics Da	23093600
CTD	Genes regulated by retinoic acid from the Comparative Toxicogenomics	23093600
CTD	Genes regulated by Tamoxifen from the Comparative Toxicogenomics D	23093600
CTD	Genes regulated by TSA from the Comparative Toxicogenomics Databas	23093600
CTD	Genes regulated by Vitamin D (Calcitriol) from the Comparative Toxicog	23093600
CTD	Genes upregulated by Dihydrotestosterone (DHT) from the Comparativ	23093600
CTD	Genes upregulated by Doxorubicin from the Comparative Toxicogenom	23093600
CTD	Genes upregulated by estrogen from the Comparative Toxicogenomics I	23093600
CTD	Genes upregulated by retinoic acid from the Comparative Toxicogenom	23093600
CTD	Genes upregulated by Tamoxifen from the Comparative Toxicogenomic	23093600
CTD	Genes upregulated by TSA from the Comparative Toxicogenomics Datat	23093600
CTD	Genes upregulated by Vitamin D (Calcitriol) from the Comparative Toxic	23093600
E2F ChIP-chip	E2F regions 1.5kb TSS MCF7	17908821
Erg ChIPseq	Erg regions within 50kb of a gene in HPC-7 cells	20887958
ESR1	Estrogen targets	15001666 17986456
ESR1	Known ERE	17986456
Fli-1 ChIPseq	Fli-1 regions within 50kb of a gene in HPC-7 cells	20887958
Gata2 ChIPseq	Gata2 regions within 50kb of a gene in HPC-7 cells	20887958
Gfi1b ChIPseq	Gfi1b regions within 50kb of a gene in HPC-7 cells	20887958
Lmo2 ChIPseq	Lmo2 regions within 50kb of a gene in HPC-7 cells	20887958
Lyl1 ChIPseq	Lyl1 regions within 50kb of a gene in HPC-7 cells	20887958
Meis1 ChIPseq	Meis1 regions within 50kb of a gene in HPC-7 cells	20887958
MiSTIC peak	TCGA-BRCA Cell cycle	
MiSTIC peak	TCGA-BRCA EGR/FOS normal	
MiSTIC peak	TCGA-BRCA FOXC1 Triple negative	
MiSTIC peak	TCGA-BRCA GRHL2 normal	
MiSTIC peak	TCGA-BRCA ZEB1 normal	
Pu.1 ChIPseq	Pu.1 regions within 50kb of a gene in HPC-7 cells	20887958
RAR (RARa, RARb or RARg)	Known RARE	16081280
RAR (RARa, RARb or RARg)	Retinoic acid targets	12401878
Runx1 ChIPseq	Runx1 regions within 50kb of a gene in HPC-7 cells	20887958
Scl ChIPseq	Scl regions within 50kb of a gene in HPC-7 cells	20887958
Transcriptional Regulatory E	AP1 targets	17202159
Transcriptional Regulatory E	AP2 targets	17202159
Transcriptional Regulatory E	E2F targets	17202159
Transcriptional Regulatory E	EGR targets	17202159
bimodal genes	TCGA-BRCA tumor Bimodality Index >1.5	
bimodal genes	TCGA-BRCA tumor Bimodality Index >1.2	

Supplementary Table 4: Geneset enrichment in the epithelial, mesenchymal/adipocyte and

Selected enriched gene sets in the Epithelial differentiation peak (second cluster from the top)

P-value	Odds	Name	Type	Cat	ID
8.29e-12	4.90	GRHL2 primary bronchial **epithelial** cells [25kb TSS]	chipseq	TF	60
9.98e-18	8.23	GRHL2 primary bronchial **epithelial** cells [10kb TSS]	chipseq	TF	237
2.83e-14	7.85	GRHL2 primary bronchial **epithelial** cells [5kb TSS]	chipseq	TF	139
8.29e-12	4.90	GRHL2 primary bronchial **epithelial** cells [25kb TSS]	chipseq	TF	60
1.17e-7	30.73	JECHLINGER_**EPITHELIAL**_TO_MESENCHYMAL_TRANSITION_DN	msigdb	C2/CGP	M1417
1.01e-4	18.39	BOSCO_**EPITHELIAL**_DIFFERENTIATION_MODULE	msigdb	C2/CGP	M2533
1.41e-3	41.67	REACTOME_TGF_BETA_RECEPTOR_SIGNALING_IN_EMT_**EPITHELIAL**_1	msigdb	C2/CP:REACTOME	M627
1.59e-3	38.89	TGF-beta receptor signaling in EMT (**epithelial** to mesenchymal transitio	CPDB	Reactome	3870
7.00e-3	288.08	columnar/cuboidal **epithelial** cell development	go	BP	2066
7.00e-3	288.08	**epithelial**-mesenchymal signaling involved in prostate gland developmer	go	BP	60738
1.05e-2	144.03	intestinal **epithelial** cell development	go	BP	60576
1.39e-2	96.02	**epithelial** cell maturation involved in prostate gland development	go	BP	60743
1.39e-2	96.02	negative regulation of **epithelial** cell migration	go	BP	10633
1.39e-2	96.02	secretory columnal luminar **epithelial** cell differentiation involved in pros	go	BP	60528
1.69e-2	5.84	SARRIO_**EPITHELIAL**_MESENCHYMAL_TRANSITION_DN	msigdb	C2/CGP	M11513
1.74e-2	72.01	regulation of **epithelial** cell proliferation	go	BP	50678
2.77e-2	41.14	**epithelial** cell-cell adhesion	go	BP	90136
2.77e-2	41.14	lung **epithelial** cell differentiation	go	BP	60487
3.45e-2	32.00	**epithelial** tube branching involved in lung morphogenesis	go	BP	60441
3.45e-2	32.00	**EPITHELIAL**_CELL_DIFFERENTIATION	msigdb	C5/BP	M3
3.79e-2	28.80	negative regulation of **epithelial** to mesenchymal transition	go	BP	10719
8.59e-12	12.08	cell **junction**	go	CC	30054
8.06e-10	31.58	tight **junction**	go	CC	5923
9.95e-10	44.54	REACTOME_CELL_CELL_**JUNCTION**_ORGANIZATION	msigdb	C2/CP:REACTOME	M820
1.45e-9	41.96	Cell-cell **junction** organization	CPDB	Reactome	2328
1.07e-8	30.71	REACTOME_CELL_**JUNCTION**_ORGANIZATION	msigdb	C2/CP:REACTOME	M19248
1.80e-8	28.30	Cell **junction** organization	CPDB	Reactome	2505
4.53e-7	17.12	Tight **junction** - Homo sapiens (human)	CPDB	KEGG	980
4.53e-7	17.12	KEGG_TIGHT_**JUNCTION**	msigdb	C2/CP:KEGG	M11355
3.10e-6	47.91	REACTOME_TIGHT_**JUNCTION**_INTERACTIONS	msigdb	C2/CP:REACTOME	M17967
3.57e-6	46.06	Tight **junction** interactions	CPDB	Reactome	2327
5.97e-6	39.92	adherens **junction** organization	go	BP	34332
1.02e-5	19.66	CELL_**JUNCTION**	msigdb	C5/CC	M11246
4.78e-5	22.57	cell-cell **junction** organization	go	BP	45216
8.03e-5	19.60	INTERCELLULAR_**JUNCTION**	msigdb	C5/CC	M15597
1.14e-4	36.93	REACTOME_ADHERENS_**JUNCTIONS**_INTERACTIONS	msigdb	C2/CP:REACTOME	M11980
1.40e-4	16.83	cell **junction** assembly	go	BP	34329
1.42e-4	34.09	Adherens **junctions** interactions	CPDB	Reactome	2414
1.73e-4	31.65	TIGHT_**JUNCTION**	msigdb	C5/CC	M2166
2.29e-4	28.58	APICAL_**JUNCTION**_COMPLEX	msigdb	C5/CC	M4718
3.35e-4	97.27	apical **junction** complex	go	CC	43296
4.31e-4	22.71	adherens **junction**	go	CC	5912
2.92e-3	27.77	ADHERENS_**JUNCTION**	msigdb	C5/CC	M9849
3.73e-3	24.30	cell-cell adherens **junction**	go	CC	5913
9.97e-3	14.21	Stabilization and expansion of the E-cadherin adherens **junction**	CPDB	PID	2958
1.71e-2	10.59	Posttranslational regulation of adherens **junction** stability and dissassem	CPDB	PID	1966
2.08e-2	57.61	Tight **junction** interactions	CPDB	PID	1332
2.71e-2	8.20	Adherens **junction** - Homo sapiens (human)	CPDB	KEGG	956
2.85e-2	7.97	KEGG_ADHERENS_**JUNCTION**	msigdb	C2/CP:KEGG	M638
3.90e-2	6.68	cell-cell **junction**	go	CC	5911

Selected enriched gene sets in the Adipocyte differentiation peak (first cluster from the top)

9.10e-19	6.10	Breast Cancer molecular subtype [TNBC] - **MESENCHYMAL** STEM-LIKE	microarrays	Signature	77
7.63e-9	5.31	**Mesenchymal**StemCells_Yamamoto08_1252genes	GeneSigDB	**Mesenchymal Ster**	2308
2.58e-4	4.23	CHARAFE_BREAST_CANCER_LUMINAL_VS_**MESENCHYMAL**_DN	msigdb	C2/CGP	M9192
9.86e-3	14.50	StemCells_Secco09_29genes	GeneSigDB	**Mesenchymal Ster**	3478
2.44e-10	27.42	REACTOME_TRANSCRIPTIONAL_REGULATION_OF_WHITE_**ADIPOCYTE**_	msigdb	C2/CP:REACTOME	M1008
8.18e-9	23.06	URS_**ADIPOCYTE**_DIFFERENTIATION_UP	msigdb	C2/CGP	M15476
1.11e-2	179.97	Transport of Glycerol from **Adipocytes** to the Liver by Aquaporins	CPDB	Reactome	1880
2.50e-2	8.62	Transcriptional Regulation of White **Adipocyte** Differentiation	CPDB	Reactome	1406
3.42e-20	215.40	STEGER_**ADIPOGENESIS**_UP	msigdb	C2/CGP	M2270
4.67e-16	20.29	BURTON_**ADIPOGENESIS**_6	msigdb	C2/CGP	M1678
1.36e-13	49.70	GERHOLD_**ADIPOGENESIS**_UP	msigdb	C2/CGP	M1675
1.58e-11	6.27	WAKABAYASHI_**ADIPOGENESIS**_PPARG_RXRA_BOUND_8D	msigdb	C2/CGP	M2417
3.11e-10	102.28	LI_**ADIPOGENESIS**_BY_ACTIVATED_PPARG	msigdb	C2/CGP	M1645
3.34e-10	20.57	NAKAMURA_**ADIPOGENESIS**_LATE_UP	msigdb	C2/CGP	M2026
3.02e-9	16.09	**Adipogenesis**	CPDB	Wikipathways	818
5.40e-8	66.40	VERNOCHET_**ADIPOGENESIS**	msigdb	C2/CGP	M2432
1.24e-7	12.66	WAKABAYASHI_**ADIPOGENESIS**_PPARG_RXRA_BOUND_WITH_H4K20M	msigdb	C2/CGP	M2420
7.06e-4	7.64	BURTON_**ADIPOGENESIS**_5	msigdb	C2/CGP	M1564
1.14e-3	3.23	WAKABAYASHI_**ADIPOGENESIS**_PPARG_BOUND_8D	msigdb	C2/CGP	M2419
1.04e-2	4.95	WAKABAYASHI_**ADIPOGENESIS**_PPARG_RXRA_BOUND_36HR	msigdb	C2/CGP	M2415
1.21e-2	12.94	BURTON_**ADIPOGENESIS**_10	msigdb	C2/CGP	M1500
1.48e-2	6.14	BURTON_**ADIPOGENESIS**_9	msigdb	C2/CGP	M1666

Selected enriched gene sets in the Lymphocyte activation peak (cluster at the base of the peak)

1.74e-31	11.04	**Immune**_Kong10_5640genes_ImmPort_ComprehensiveListofImmune-R	GeneSigDB	Immune	111
2.69e-20	10.40	**Immune** System	CPDB	Reactome	1358
7.46e-20	42.71	**Immune**_Carmody02_151genes	GeneSigDB	Immune	2456
1.15e-19	13.60	Adaptive **Immune** System	CPDB	Reactome	1573
2.18e-18	18.35	**Immune**_Rioja04_438genes	GeneSigDB	Immune	415
3.17e-18	17.99	**IMMUNE**_SYSTEM_PROCESS	msigdb	C5/BP	M13664
3.00e-17	12.77	REACTOME_ADAPTIVE_**IMMUNE**_SYSTEM	msigdb	C2/CP:REACTOME	M1058
1.28e-16	9.21	REACTOME_**IMMUNE**_SYSTEM	msigdb	C2/CP:REACTOME	M1045
6.24e-14	40.02	regulation of **immune** response	go	BP	50776
3.36e-11	11.82	**immune** response	go	BP	6955
4.75e-11	25.51	**Immune**_Allen10_106genes	GeneSigDB	Immune	1950
1.07e-10	30.21	module #5 **Immune**	microarrays	Breast Cancer Co-E›	51
9.48e-10	12.97	**IMMUNE**_RESPONSE	msigdb	C5/BP	M19817
1.83e-9	7.42	Innate **Immune** System	CPDB	Reactome	4527
4.28e-9	25.13	module #4 **Immune**	microarrays	Breast Cancer Co-E›	50
1.84e-8	9.81	innate **immune** response	go	BP	45087
5.10e-8	17.86	**Immune**_Allen10_162genes	GeneSigDB	Immune	983
1.20e-7	15.86	**Immune**_Kong10_135genes_ImmPort_NaturalKillerCellCytotoxicity	GeneSigDB	Immune	475
2.08e-6	18.21	**IMMUNE**_SYSTEM_DEVELOPMENT	msigdb	C5/BP	M3457
2.54e-6	3.60	**Immune**_Kong10_2491genes_ImmPort_CD40LandAnti-IgMvsControl	GeneSigDB	Immune	2434
3.15e-6	12.55	**Immune**_Kong10_148genes_ImmPort_AntigenProcessingandPresentatic	GeneSigDB	Immune	950
3.81e-6	24.53	POSITIVE_REGULATION_OF_**IMMUNE**_SYSTEM_PROCESS	msigdb	C5/BP	M4046
1.48e-5	18.18	REGULATION_OF_**IMMUNE**_SYSTEM_PROCESS	msigdb	C5/BP	M5017
4.68e-5	6.80	REACTOME_INNATE_**IMMUNE**_SYSTEM	msigdb	C2/CP:REACTOME	M1036
5.23e-5	2.98	**Immune**_Allen10_2900genes	GeneSigDB	Immune	217
6.24e-5	47.45	ACTIVATION_OF_**IMMUNE**_RESPONSE	msigdb	C5/BP	M19789
6.34e-5	438.75	**immune** response-inhibiting cell surface receptor signaling pathway	go	BP	2767
1.75e-4	4.47	**Immune**_Kong10_855genes_ImmPort_CD40LvsControl	GeneSigDB	Immune	3129
2.36e-4	14.64	**Immune**_Kong10_79genes_Immune_Kong10_genes_ImmPort_Centroc	GeneSigDB	Immune	3428
2.96e-4	4.59	**Immune**_Kong10_544genes_ImmPort_Antimicrobials	GeneSigDB	Immune	817
3.22e-4	25.54	POSITIVE_REGULATION_OF_**IMMUNE**_RESPONSE	msigdb	C5/BP	M4618
4.74e-4	22.13	REGULATION_OF_**IMMUNE**_RESPONSE	msigdb	C5/BP	M12336
8.26e-4	2.91	**Immune**_Kong10_1790genes_ImmPort_Anti-IgMvsControl	GeneSigDB	Immune	1021
2.14e-3	33.73	adaptive **immune** response	go	BP	2250
4.63e-3	Infinity	positive regulation of gamma-delta T cell activation involved in **immune**	go	BP	2001193
5.48e-3	19.92	ADAPTIVE_**IMMUNE**_RESPONSE_GO_0002460	msigdb	C5/BP	M16548
5.93e-3	19.06	ADAPTIVE_**IMMUNE**_RESPONSE	msigdb	C5/BP	M13450
9.23e-3	217.23	regulation of humoral **immune** response	go	BP	2920
9.23e-3	217.23	leukocyte migration involved in **immune** response	go	BP	2522
9.23e-3	217.23	innate **immune** response-activating signal transduction	go	BP	2758
1.02e-2	14.13	**Immune**_Allen10_42genes	GeneSigDB	Immune	107
1.34e-2	12.17	**IMMUNE**_EFFECTOR_PROCESS	msigdb	C5/BP	M14818
1.38e-2	108.61	positive regulation of adaptive **immune** response	go	BP	2821
2.29e-2	54.30	negative regulation of cytokine secretion involved in **immune** response	go	BP	2740
2.29e-2	54.30	T cell proliferation involved in **immune** response	go	BP	2309
4.09e-2	27.14	positive regulation of **immune** response	go	BP	50778
4.97e-2	21.71	regulation of innate **immune** response	go	BP	45088
6.97e-20	27.46	LEE_DIFFERENTIATING_T_**LYMPHOCYTE**	msigdb	C2/CGP	M2200
1.31e-8	29.85	**LYMPHOCYTE**_ACTIVATION	msigdb	C5/BP	M2297
1.65e-8	28.78	HADDAD_T_**LYMPHOCYTE**_AND_NK_PROGENITOR_DN	msigdb	C2/CGP	M999
5.97e-6	40.65	**LYMPHOCYTE**_DIFFERENTIATION	msigdb	C5/BP	M18190
1.55e-5	83.06	BIOCARTA_B**LYMPHOCYTE**_PATHWAY	msigdb	C2/CP:BIOCARTA	M1473
1.06e-4	18.23	MORI_IMMATURE_B_**LYMPHOCYTE**_UP	msigdb	C2/CGP	M11262
5.65e-4	20.74	REGULATION_OF_**LYMPHOCYTE**_ACTIVATION	msigdb	C5/BP	M16183
6.15e-4	20.11	SIG_PIP3_SIGNALING_IN_B_**LYMPHOCYTES**	msigdb	C2/CP	M1315
2.47e-3	12.06	MORI_LARGE_PRE_BII_**LYMPHOCYTE**_DN	msigdb	C2/CGP	M4856
4.63e-3	Infinity	positive regulation of **lymphocyte** differentiation	go	BP	45621
4.63e-3	Infinity	regulation of **lymphocyte** activation	go	BP	51249
4.63e-3	Infinity	positive regulation of **lymphocyte** activation	go	BP	51251
5.48e-3	19.92	POSITIVE_REGULATION_OF_**LYMPHOCYTE**_ACTIVATION	msigdb	C5/BP	M18946
5.49e-3	8.95	MORI_PRE_BI_**LYMPHOCYTE**_DN	msigdb	C2/CGP	M6455
6.11e-3	8.60	MORI_PRE_BI_**LYMPHOCYTE**_UP	msigdb	C2/CGP	M1395
8.45e-3	7.61	MORI_MATURE_B_**LYMPHOCYTE**_UP	msigdb	C2/CGP	M17083
4.09e-2	27.14	**lymphocyte** chemotaxis	go	BP	48247
4.18e-2	6.43	HOFFMANN_SMALL_PRE_BII_TO_IMMATURE_B_**LYMPHOCYTE**_UP	msigdb	C2/CGP	M13760
4.39e-2	6.25	HOFFMANN_LARGE_TO_SMALL_PRE_BII_**LYMPHOCYTE**_DN	msigdb	C2/CGP	M19523
4.53e-2	24.13	VISALA_AGING_**LYMPHOCYTE**_UP	msigdb	C2/CGP	M2050

Supplementary Table 5: Genes in the CC cluster containing regions associated with E2F1, Mybl2 or FOXM1

E2F1 HeLa 5 kb (ChIPSeq #157)	Mybl2 HepG2 5 kb (ChIPSeq #147)	FOXM1 MCF7 cells 5 kb (ChIP-Seq #120)	# TFs associated with 5kb flanking regions
ANLN	ANLN	ANLN	3
ASF1B	ASF1B	ASF1B	3
	ASPM	ASPM	2
ATAD2	ATAD2	ATAD2	3
AURKA	AURKA	AURKA	3
	AURKB	AURKB	2
	BIRC5	BIRC5	2
BLM	BLM		2
	BUB1		1
BUB1B	BUB1B		2
	C15orf23	C15orf23	2
	C15orf42	C15orf42	2
C17orf53	C17orf53		2
C1orf135	C1orf135		2
	CASC5		1
	CCNA2	CCNA2	2
	CCNB1	CCNB1	2
	CCNB2	CCNB2	2
	CDC20	CDC20	2
CDC25A	CDC25A		2
	CDC25C	CDC25C	2
CDC45	CDC45		2
CDC6	CDC6		2
CDCA2	CDCA2	CDCA2	3
	CDCA3		1
	CDCA5		1
CDCA8	CDCA8	CDCA8	3
CDK1	CDK1	CDK1	3
	CDKN3	CDKN3	2
	CENPA	CENPA	2
	CENPE		1
CENPF	CENPF	CENPF	3
CENPI	CENPI		2
CENPL	CENPL	CENPL	3
CENPM	CENPM		2
	CENPN		1
	CENPW		1
CEP55	CEP55	CEP55	3
	CIT	CIT	2
	CKAP2L	CKAP2L	2
	CKS2	CKS2	2
CLSPN			1
	CMC2		1
	DEPDC1	DEPDC1	2
DEPDC1B	DEPDC1B	DEPDC1B	3
	DLGAP5	DLGAP5	2
DSCC1			1
DTL	DTL		2
E2F1	E2F1		2
E2F2			1
E2F8	E2F8		2
	ECT2	ECT2	2
ESCO2	ESCO2		2
	ESPL1	ESPL1	2
EXO1	EXO1		2
EZH2	EZH2	EZH2	3
	FAM54A	FAM54A	2
		FAM64A	1
	FAM83D	FAM83D	2
FANCA	FANCA		2
FANCB			1
FANCI		FANCI	2
FBXO5	FBXO5	FBXO5	3
FEN1	FEN1		2
FOXM1	FOXM1		2
	GAS2L3	GAS2L3	2
GINS1	GINS1		2
GTSE1	GTSE1	GTSE1	3
H2AFZ	H2AFZ		2
HELLS			1
HJURP	HJURP	HJURP	3
	HMMR	HMMR	2
	IQGAP3		1
	KIAA1524	KIAA1524	2
KIF11	KIF11	KIF11	3
	KIF14	KIF14	2
KIF15	KIF15	KIF15	3
KIF18A	KIF18A	KIF18A	3
KIF18B	KIF18B	KIF18B	3
	KIF20A	KIF20A	2
	KIF20B	KIF20B	2
KIF23	KIF23	KIF23	3
KIF2C	KIF2C	KIF2C	3
KIF4A	KIF4A	KIF4A	3
KIFC1	KIFC1	KIFC1	3
KNTC1	KNTC1		2
	KPNA2	KPNA2	2
LMNB1	LMNB1	LMNB1	3
MAD2L1	MAD2L1	MAD2L1	3
MCM2			1
MCM6			1
		MCM10	1
MELK	MELK		2
MKI67	MKI67	MKI67	3
	MTBP		1
MND1			1
MYBL2			1
NCAPD2	NCAPD2	NCAPD2	3
NCAPG	NCAPG		2
NCAPG2	NCAPG2	NCAPG2	3
NCAPH	NCAPH	NCAPH	3
NDC80	NDC80	NDC80	3
	NEIL3	NEIL3	2
	NEK2	NEK2	2
NUF2	NUF2	NUF2	3
NUSAP1	NUSAP1	NUSAP1	3
OIP5	OIP5	OIP5	3
	ORC1		1
	PARPBP	PARPBP	2
PBK	PBK	PBK	3
	PLK1	PLK1	2
PLK4	PLK4		2
	POLQ		1
PRC1	PRC1	PRC1	3
PRR11	PRR11	PRR11	3
	PTTG1	PTTG1	2
RACGAP1	RACGAP1	RACGAP1	3
RAD51			1
	RAD51AP1		1
RAD54L	RAD54L		2
RANBP1	RANBP1		2
RNASEH2A	RNASEH2A		2
RRM2	RRM2	RRM2	3
	SGOL1	SGOL1	2
	SGOL2	SGOL2	2
	SHCBP1		1
SKA3	SKA3	SKA3	3
SPAG5	SPAG5	SPAG5	3
	SPC24	SPC24	2
STIL	STIL	STIL	3
STMN1	STMN1		2
	TACC3	TACC3	2
TCF19	TCF19		2
TIMELESS			1
	TK1		1
TMPO	TMPO	TMPO	3
	TOP2A	TOP2A	2
	TPX2	TPX2	2
	TROAP	TROAP	2
	TTK	TTK	2
TYMS	TYMS		2
	UBE2C	UBE2C	2
UBE2T	UBE2T	UBE2T	3
UHRF1	UHRF1	UHRF1	3
ZNF367	ZNF367		2
ZWINT	ZWINT		2

Supplementary Table 6: Genes with chromatin regions associated with TFs in the FOXA1/ESR1 cluster

Gene	Subcluster	ChIPSeq ER 10kb	ChIPSeq FOXA1 10 kb	ChIPSeq GATA3 10 kb	ChIPSeq SPDEF 5kb	ChIPSeq AR 10 kb
ABCC8			ABCC8		ABCC8	
AFF3		AFF3			AFF3	AFF3
AGR2	FOXA1	AGR2	AGR2	AGR2		AGR2
AGR3	ESR1	AGR3	AGR3	AGR3		AGR3
ANXA9			ANXA9	ANXA9	ANXA9	
APBB2		ANXA9		APBB2	APBB2	APBB2
AR			AR	AR	AR	AR
ARSG		ARSG	ARSG		ARSG	ARSG
C17orf28		C17orf28	C17orf28		C17orf28	C17orf28
C1orf64		C1orf64				
C6orf211			C6orf211		C6orf211	
C9orf152		C9orf152	C9orf152	C9orf152	C9orf152	C9orf152
CA12	ESR1	CA12	CA12	CA12	CA12	
CCDC170	ESR1	CCDC170	CCDC170	CCDC170	CCDC170	
CMBL		CMBL	CMBL		CMBL	CMBL
CT62	ESR1	CT62	CT62		CT62	CT62
DACH1						
DEGS2		DEGS2		DEGS2	DEGS2	DEGS2
DNALI1						DNALI1
ESR1	ESR1	ESR1	ESR1	ESR1	ESR1	ESR1
FAM63A		FAM63A	FAM63A		FAM63A	FAM63A
FBP1		FBP1	FBP1	FBP1	FBP1	FBP1
FLJ45983	ESR1	FLJ45983	FLJ45983	FLJ45983	FLJ45983	
FOXA1	FOXA1	FOXA1	FOXA1	FOXA1	FOXA1	FOXA1
FSIP1		FSIP1	FSIP1	FSIP1	FSIP1	FSIP1
GATA3	ESR1	GATA3	GATA3	GATA3	GATA3	
GPR160		GPR160	GPR160	GPR160	GPR160	GPR160
GPR77	ESR1	GPR77		GPR77	GPR77	
KCNJ11			KCNJ11		KCNJ11	
LOC145837	FOXA1	LOC145837	LOC145837	LOC145837	LOC145837	LOC145837
MLPH	FOXA1	MLPH	MLPH	MLPH	MLPH	MLPH
PRR15	FOXA1	PRR15			PRR15	PRR15
RAB17		RAB17			RAB17	RAB17
RABEP1		RABEP1			RABEP1	
RMND1			RMND1		RMND1	
SIDT1		SIDT1	SIDT1	SIDT1		SIDT1
SLC16A6		SLC16A6	SLC16A6		SLC16A6	SLC16A6
SLC44A4		SLC44A4			SLC44A4	
SLC7A8		SLC7A8	SLC7A8	SLC7A8	SLC7A8	SLC7A8
SPDEF	FOXA1	SPDEF	SPDEF	SPDEF	SPDEF	SPDEF
TBC1D9	ESR1	TBC1D9			TBC1D9	
TFF1	FOXA1	TFF1	TFF1	TFF1	TFF1	TFF1
TFF3	FOXA1	TFF3	TFF3	TFF3	TFF3	TFF3
THSD4	ESR1					THSD4
XBP1	FOXA1	XBP1	XBP1	XBP1	XBP1	XBP1

Supplementary Table 7: Enrichement in basal genesets in the p63 cluster

P-value	Odds	Name	Type	Cat	ID
8.32E-28	182.32	module #2 Dev/Basal: development/differentiation	microarrays	Breast Cancer Co-Expression module	48
2.25E-22	77.16	SMID_BREAST_CANCER_BASAL_UP	msigdb	C2/CGP	M8124
1.41E-17	57.66	CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_DN	msigdb	C2/CGP	M14507
7.63E-16	224.15	Breast_Huper07_66genes_BasalCells	GeneSigDB	Breast	2535
1.05E-15	214.39	HUPER_BREAST_BASAL_VS_LUMINAL_UP	msigdb	C2/CGP	M13422
8.67E-13	87.01	CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_UP	msigdb	C2/CGP	M12795
1.56E-09	12.11	Breast_Farmer05_3198genes_basal_apocrine_luminal	GeneSigDB	Breast	1241
5.81E-09	101.57	Breast_Blick10_57genes_CD24_BasalB	GeneSigDB	Breast	1230
1.82E-06	20.29	FARMER_BREAST_CANCER_BASAL_VS_LULMINAL	msigdb	C2/CGP	M5652
3.54E-05	16.05	FARMER_BREAST_CANCER_APOCRINE_VS_BASAL	msigdb	C2/CGP	M2631
3.62E-05	6.11	Breast_Kreike07_3712genes_Basal_mets	GeneSigDB	Breast	3362
3.56E-04	84.14	Breast_Blick10_35genes_CD24_BasalB	GeneSigDB	Breast	1048
2.33E-03	889.36	squamous basal epithelial stem cell differentiation involved in prostate gland acinus development	go	BP	60529
3.21E-03	11.33	Breast_Farmer05_269genes__basal_apocrine_luminal	GeneSigDB	Breast	690

Supplementary Table 8: enrichment in basal genesets in the FOXC1 cluster.

P-value	Odds	Name	Type	Cat	ID
6.34E-13	323.74	module #2 Dev/Basal: development/differentiation	microarrays	Breast Cancer Co-Expression module	48
5.87E-10	118	SMID_BREAST_CANCER_BASAL_UP	msigdb	C2/CGP	M8124
2.46E-04	111.25	HUPER_BREAST_BASAL_VS_LUMINAL_DN	msigdb	C2/CGP	M5505
2.54E-04	33.52	FARMER_BREAST_CANCER_BASAL_VS_LULMINAL	msigdb	C2/CGP	M5652
3.59E-04	12.78	Breast_Farmer05_3198genes_basal_apocrine_luminal	GeneSigDB	Breast	1241
4.50E-03	24.74	Breast_Farmer05_269genes__basal_apocrine_luminal	GeneSigDB	Breast	690
7.36E-03	19.1	FARMER_BREAST_CANCER_APOCRINE_VS_BASAL	msigdb	C2/CGP	M2631
1.08E-02	6.65	Breast_Kreike07_3712genes_Basal_mets	GeneSigDB	Breast	3362
1.36E-02	13.75	CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_DN	msigdb	C2/CGP	M14507
2.00E-02	56.64	CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_DN	msigdb	C2/CGP	M12895
2.08E-02	54.41	Breast_Sorlie06_54genes_Luminal_v_Basal	GeneSigDB	Breast	3126