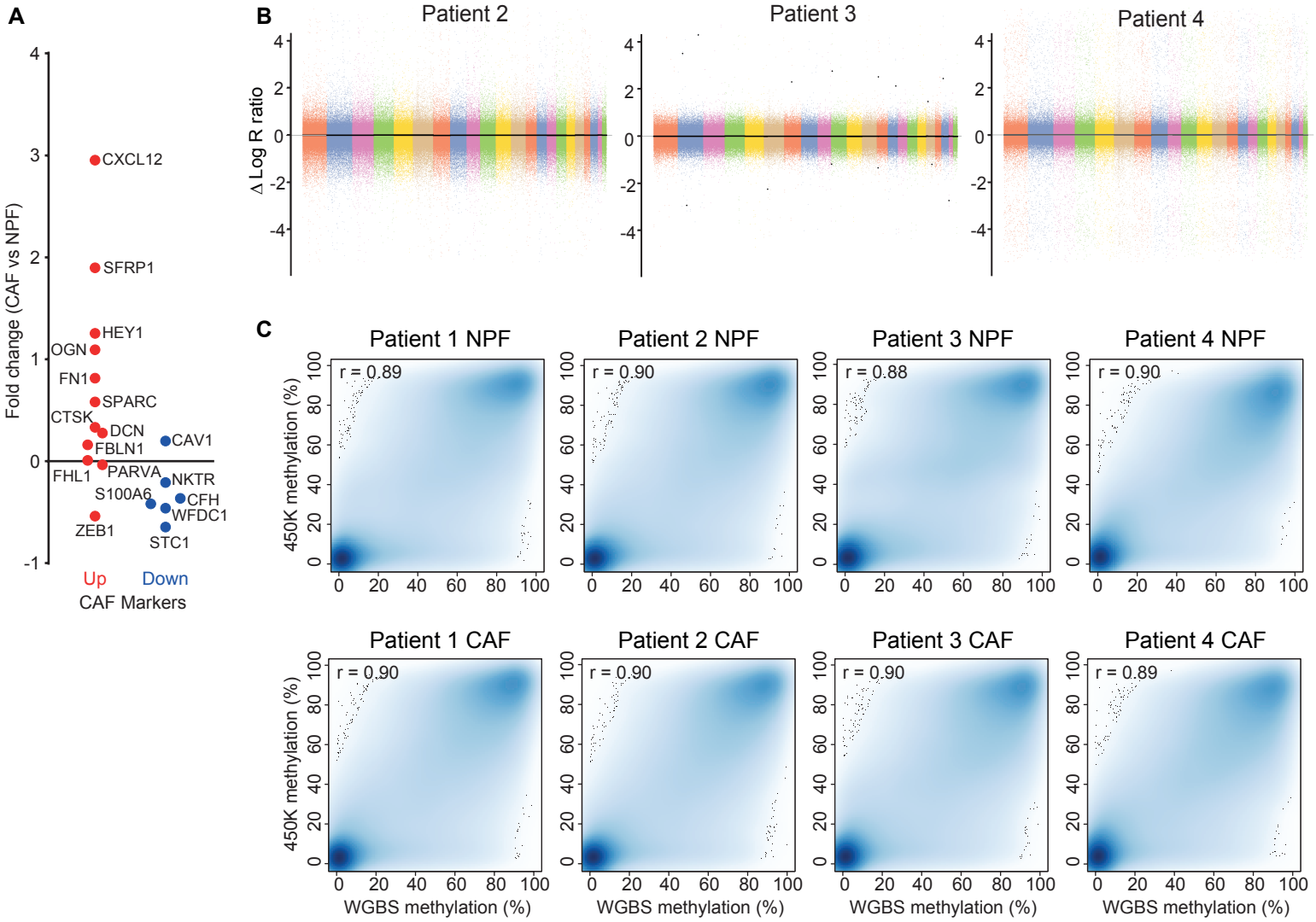
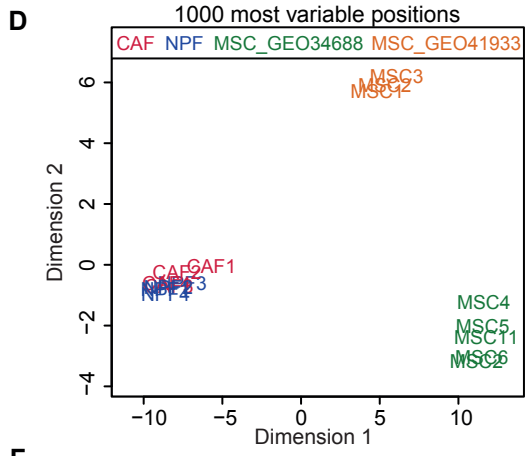
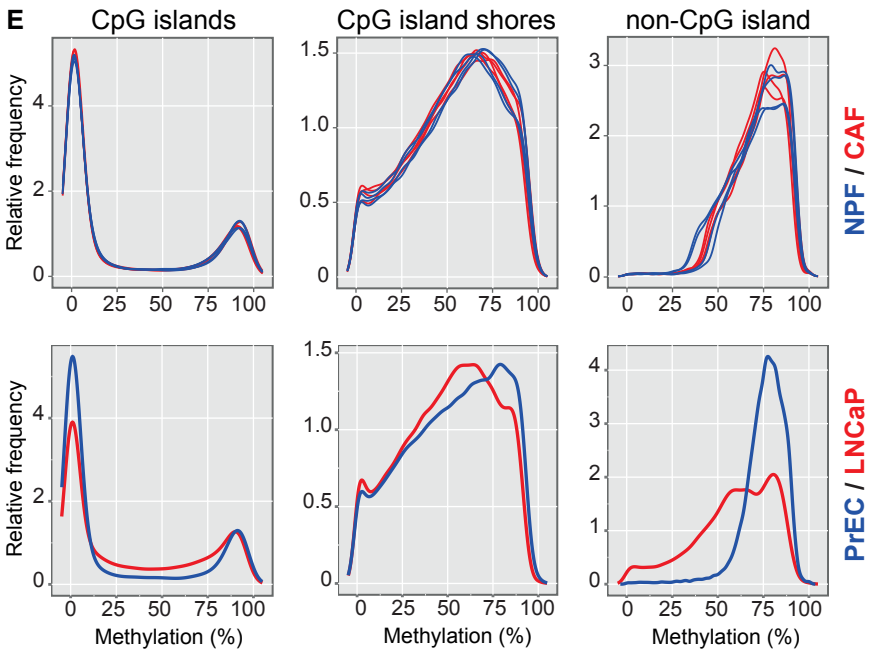
Supplemental figures and tables for
Pidsley et al., Enduring Epigenetic Landmarks Define the Cancer Microenvironment

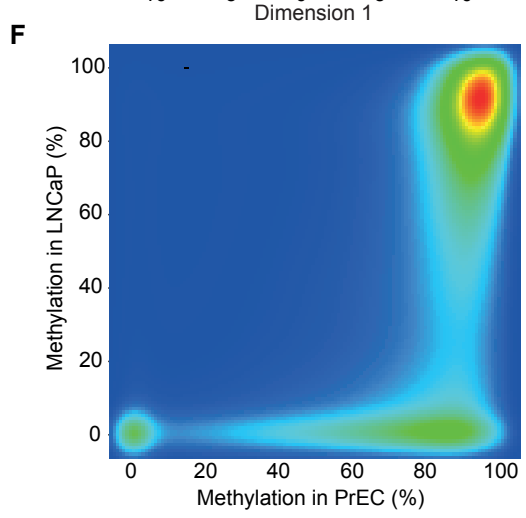
Table of Contents (Figures listed in order of citation in the text)

(Large Supplemental Tables 1-11 are available via separate links to .xlsx files at the Genome Research website).

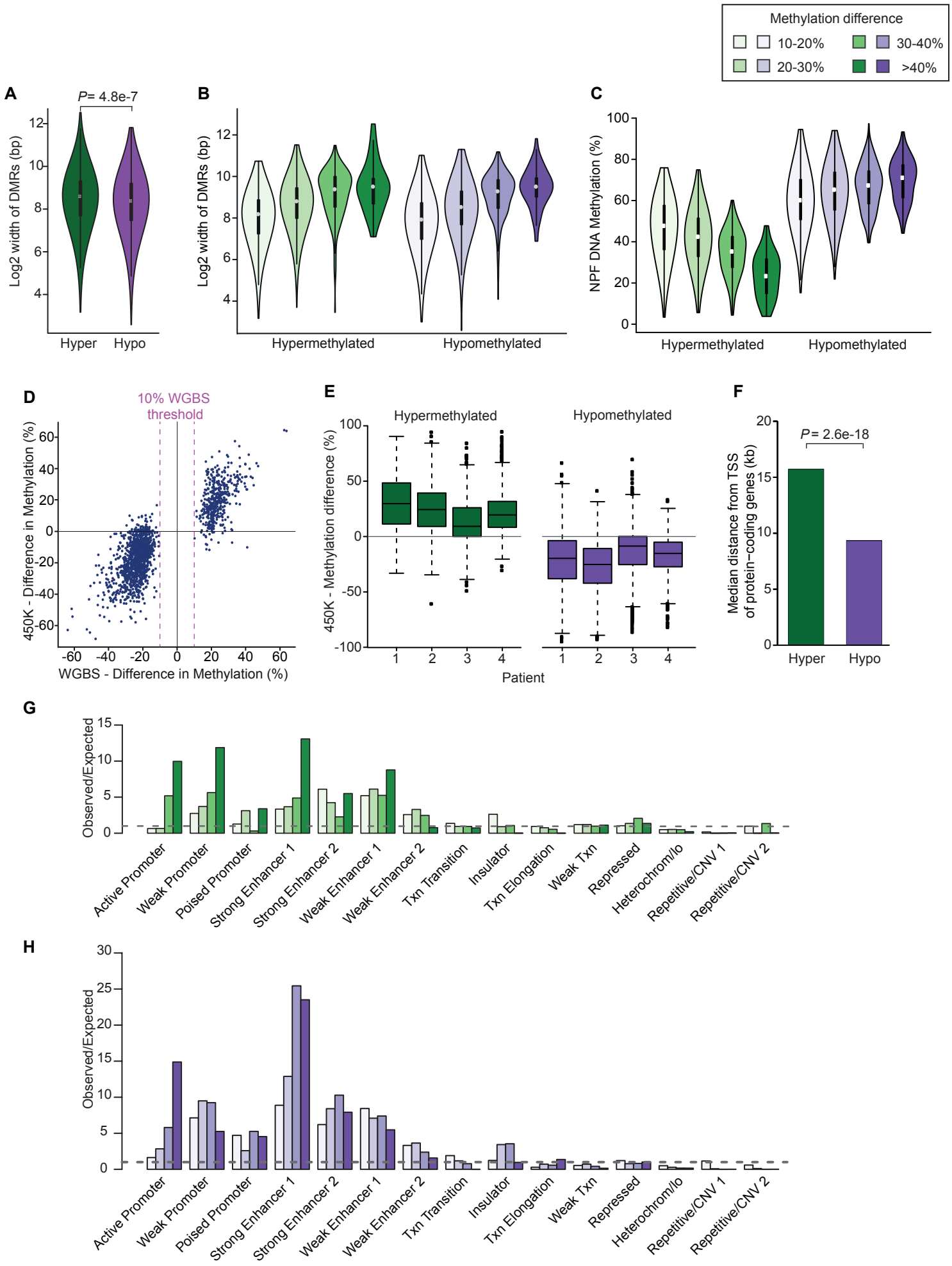

A
Fold change (CAF vs NPF)
CXCL12
SFRP1
HEY1
OGN
FN1
SPARC
CTSK
DCN
FBLN1
FHL1
PARVA
CAV1
NKTR
S100A6
CFH
WFDC1
ZEB1
STC1
Up
Down
CAF Markers
B
Patient 2
Patient 3
Patient 4
Δ Log R ratio
C
Patient 1 NPF
Patient 2 NPF
Patient 3 NPF
Patient 4 NPF
Patient 1 CAF
Patient 2 CAF
Patient 3 CAF
Patient 4 CAF
r = 0.89
r = 0.90
r = 0.88
r = 0.90
r = 0.90
r = 0.90
r = 0.90
r = 0.89
450K methylation (%)
WGBS methylation (%)


D
1000 most variable positions
CAF
NPF
MSC_GEO34688
MSC_GEO41933
MSC3
MSC1
CAF1
MSC4
MSC5
MSC11
MSC6
MSC2
Dimension 1
Dimension 2

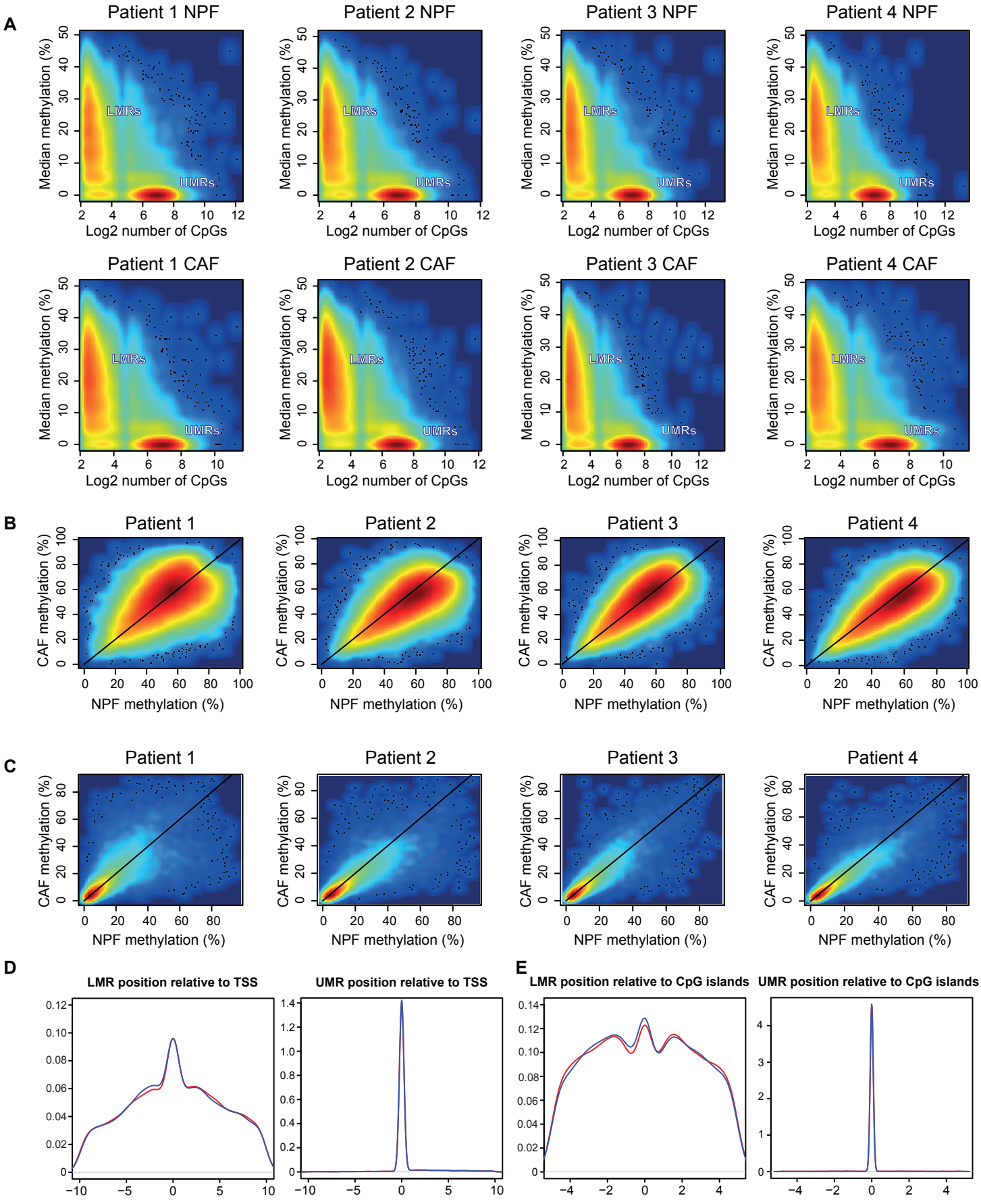

E
CpG islands
CpG island shores
non-CpG island
Relative frequency
NPF / CAF
PrEC / LNCaP
Methylation (%)


F
Methylation in LNCaP (%)
Methylation in PrEC (%)

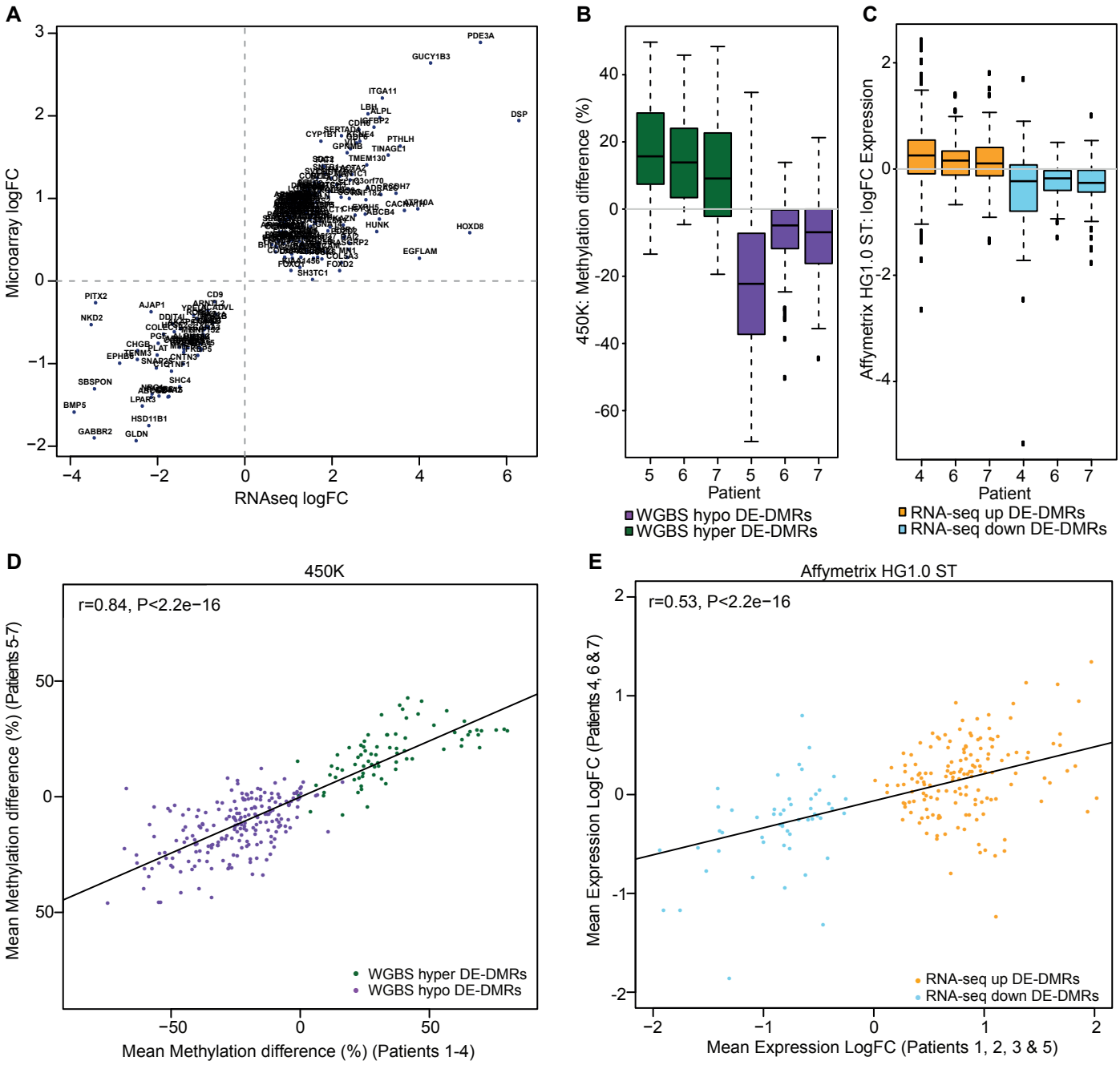
Supplemental Fig. S1. CAFs and NPFs lack genetic copy number changes and have methylation profiles characteristic of non-malignant human cells. (A) RNA-seq data for CAF markers shows similar expression differences to previous studies. (B) HumanOmni2.5-8 BeadChip genotyping data showing no evidence of large-scale copy number alterations within patient-matched prostate fibroblast samples. The difference in log R ratio between CAFs and NPFs (Patients 2-4) is plotted for each probe on the microarray, with genomic position along the x-axis and color changes indicating separate chromosomes. Circular binary segmentation of the smoothed log R ratio differences is represented by the black line, with y=0 indicating no copy number alteration. (C) Scatter plots of DNA methylation between cross-platform replicates of WGBS and 450K arrays at CpG sites interrogated by both platforms. (D) MDS plot of CAFs and NPFs compared to human bone marrow mesenchymal stromal cells from two published datasets. Plot is based on 1000 most variable positions from 450K data. (E) Density plots showing distribution of DNA methylation at CpG islands, shores and non-CpG island regions in CAFs versus NPFs and tumour epithelial (LNCaP) versus normal prostate epithelial (PrEC) cells. WGBS methylation values are smoothed separately for each sample across 100 bp windows. (F) Scatter plot of DNA methylation in LNCaP versus PrEC cells showing global hypomethylation in LNCaP cells.


Methylation difference
10-20%
20-30%
30-40%
>40%
A
P= 4.8e-7
Log2 width of DMRs (bp)
Hyper
Hypo
B
Log2 width of DMRs (bp)
Hypermethylated
Hypomethylated
C
NPF DNA Methylation (%)
Hypermethylated
Hypomethylated
D
10% WGBS threshold
450K - Difference in Methylation (%)
WGBS - Difference in Methylation (%)
E
Hypermethylated
Hypomethylated
450K - Methylation difference (%)
Patient
F
P= 2.6e-18
Median distance from TSS of protein−coding genes (kb)
Hyper
Hypo
G
Observed/Expected
Active Promoter
Weak Promoter
Poised Promoter
Strong Enhancer 1
Strong Enhancer 2
Weak Enhancer 1
Weak Enhancer 2
Txn Transition
Insulator
Txn Elongation
Weak Txn
Repressed
Heterochrom/lo
Repetitive/CNV 1
Repetitive/CNV 2
H
Observed/Expected
Active Promoter
Weak Promoter
Poised Promoter
Strong Enhancer 1
Strong Enhancer 2
Weak Enhancer 1
Weak Enhancer 2
Txn Transition
Insulator
Txn Elongation
Weak Txn
Repressed
Heterochrom/lo
Repetitive/CNV 1
Repetitive/CNV 2

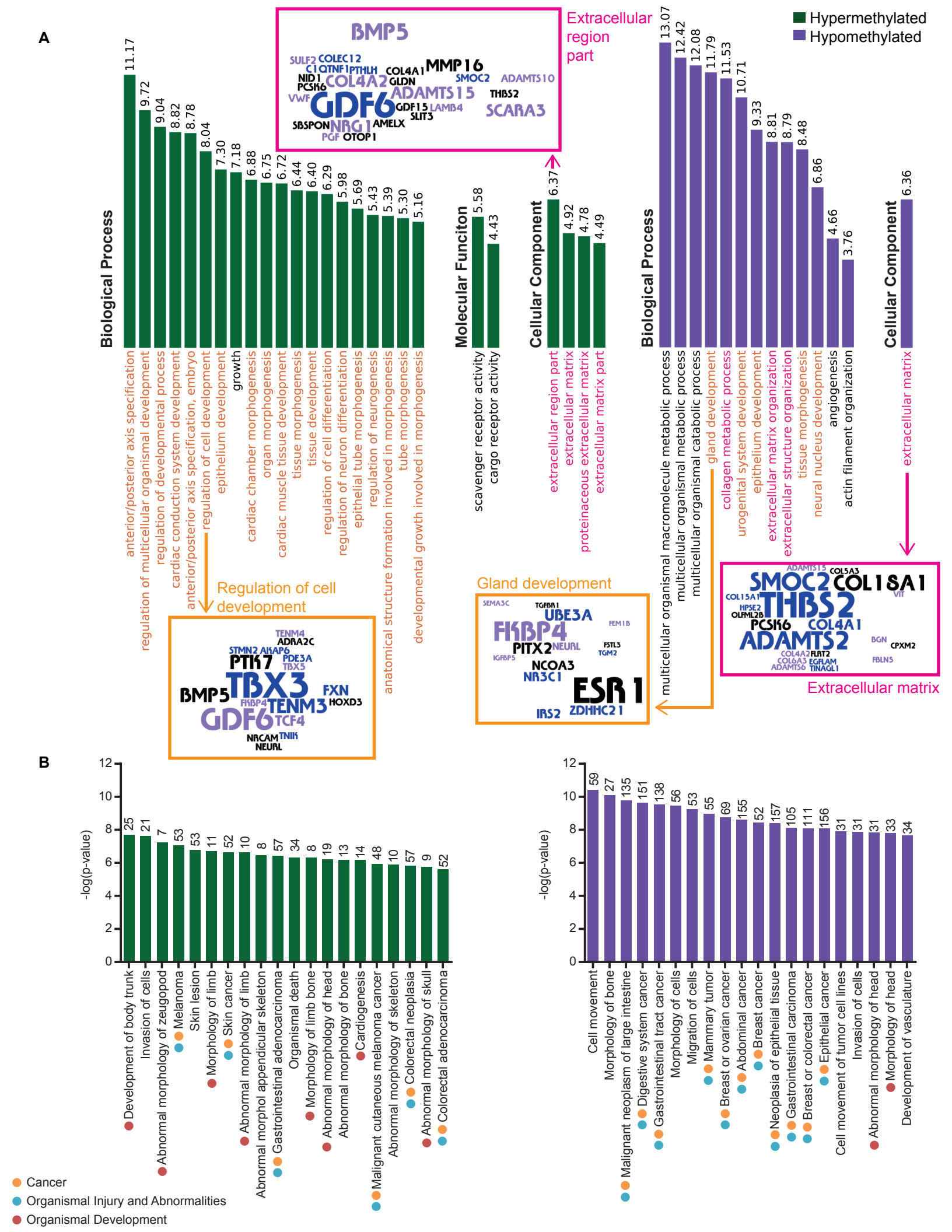
Supplemental Fig. S2. Characteristics of CAF-NPF DMRs identified using WGBS. (A) Violin plots of the log2 width of all DMRs. Hypermethylated (hyper, green) DMRs are significantly wider than hypomethylated (hypo, purple) DMRs (P=4.8 e-07, Mann-Whitney U test). (B-C) Violin plots of (B) the log2 width of DMRs and (C) NPF methylation level, grouped according to the percentage change in DNA methylation in CAFs compared to NPFs. (D) Cross-platform validation of WGBS and 450K methylation data at DMRs (Pearson's r=0.86, P<2.2e-16), for the 27% of DMRs overlapped by probes on the 450K array. Each dot represents the mean difference in DNA methylation (n=4 pairs) averaged across each DMR. (E) Boxplots showing the percentage DNA methylation difference between each patient-matched pair of CAFs and NPFs at 450K probes overlapping with WGBS hypermethylated and hypomethylated DMRs. (F) The median distance of DMRs to the transcription start site (TSS) of protein-coding genes (P=2.6 e-18, Mann- Whitney U test). (G-H) Bar graphs showing the observed/expected frequency of (G) hypermethylated and (H) hypomethylated CAF-NPF DMRs, subset according to the mean difference in methylation, at regulatory loci defined by ChromHMM data from normal fibroblasts. Note that ENCODE defines two types of strong enhancers, weak enhancers, and repetitive/CNV regions.



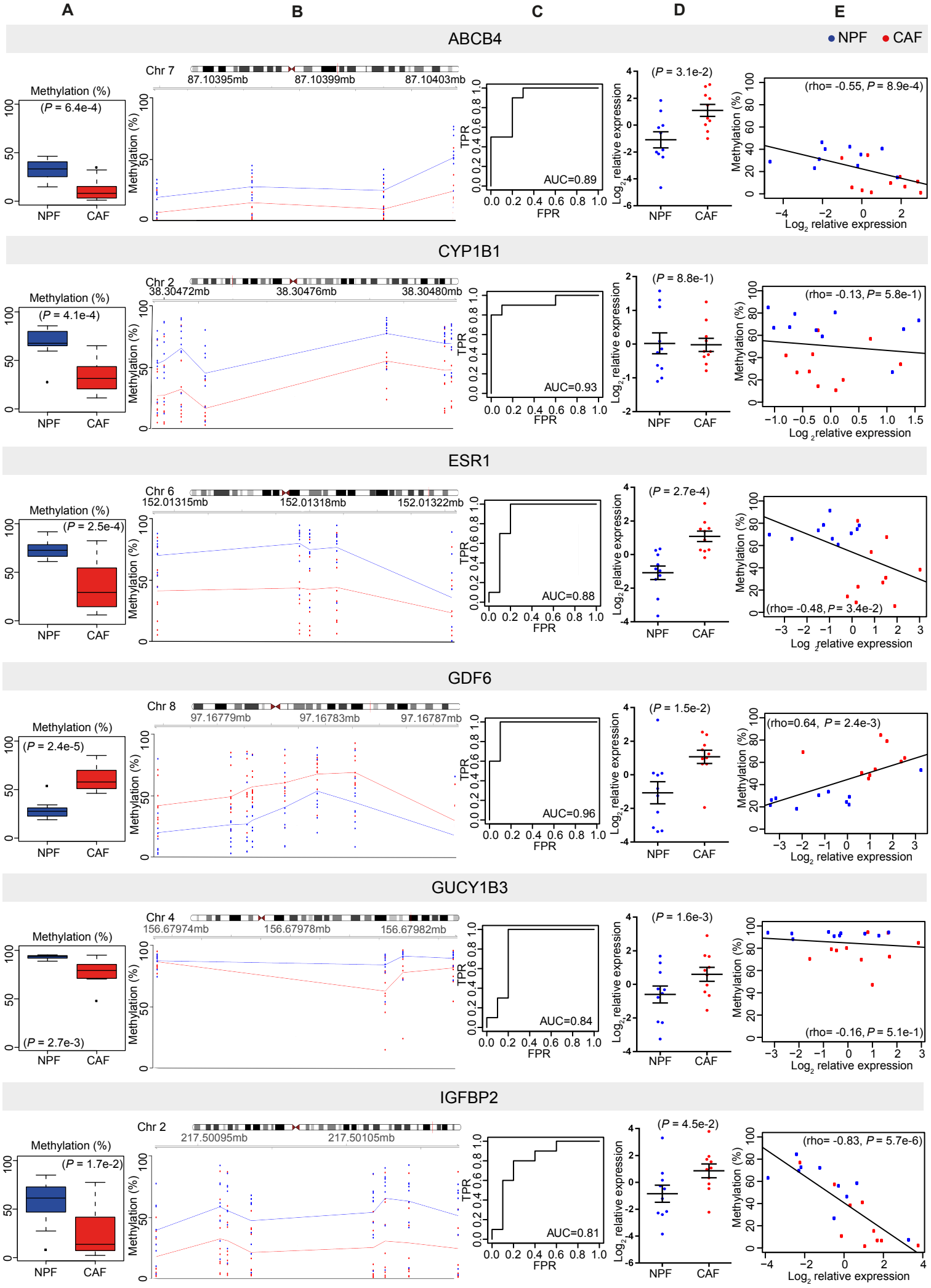
Supplemental Fig. S3. Identification of regions of LMRs and UMRs in CAFs and NPFs with typical genomic distributions. (A) Plots showing the log2 number of CpGs versus median methylation levels for all hypomethylated regions identified using the methylseekR computational tool. In all samples hypomethylated regions separate into two groups: CpG-poor, lowly methylated regions (LMRs) and CpG-rich, unmethylated regions (UMRs). Plots comparing (B) LMRs and (C) UMRs in each patient-matched pair of CAFs and NPFs. Plots showing the location of LMRs and UMRs in CAFs (red) and NPFs (blue) relative to (D) the transcription start site (TSS) of protein coding genes and (E) CpG islands.

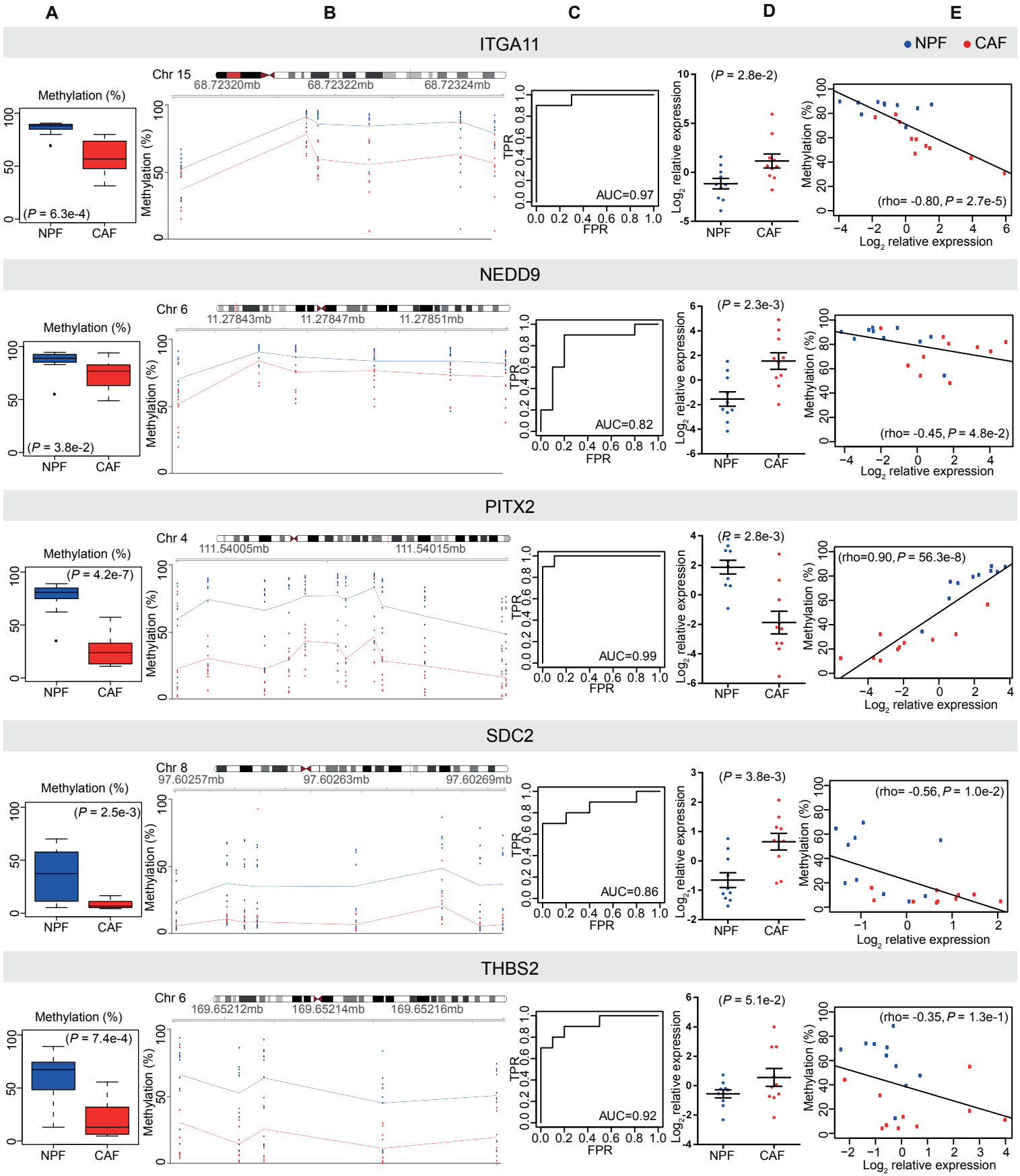


Supplemental Fig. S4. Cross-platform and biological validation of DE-DMRs. (A) The plot shows a strong correlation between gene expression differences measured with RNAseq and Affymetrix Human Gene 1.0ST Arrays (Pearson's r=0.88, P<2.2e-16). Data points represent average log fold changes (logFC) in the expression of each gene between matched CAFs and NPFs (n=4 pairs). (B) Illumina 450K BeadChip data from patients 5, 6 and 7, showing DNA methylation differences in CAFs versus NPFs at the DE-DMRs previously identified with WGBS in patients 1-4. (C) Affymetrix Human gene 1.0ST data from patients 4, 6 & 7, showing expression differences in CAFs versus NPFs at the DE-DMRs previously identified with RNA-seq in patients 1-3 & 5. (D) The relationship between the mean CAF-NPF methylation difference for patients 1-4 versus patients 5-7, measured on the 450K BeadChip. (E) The relationship between mean CAF-NPF expression Log fold change (LogFC) for patients 1-3 & 5 versus patients 4, 6 & 7, measured on the Affymetrix Human gene 1.0ST microarray.

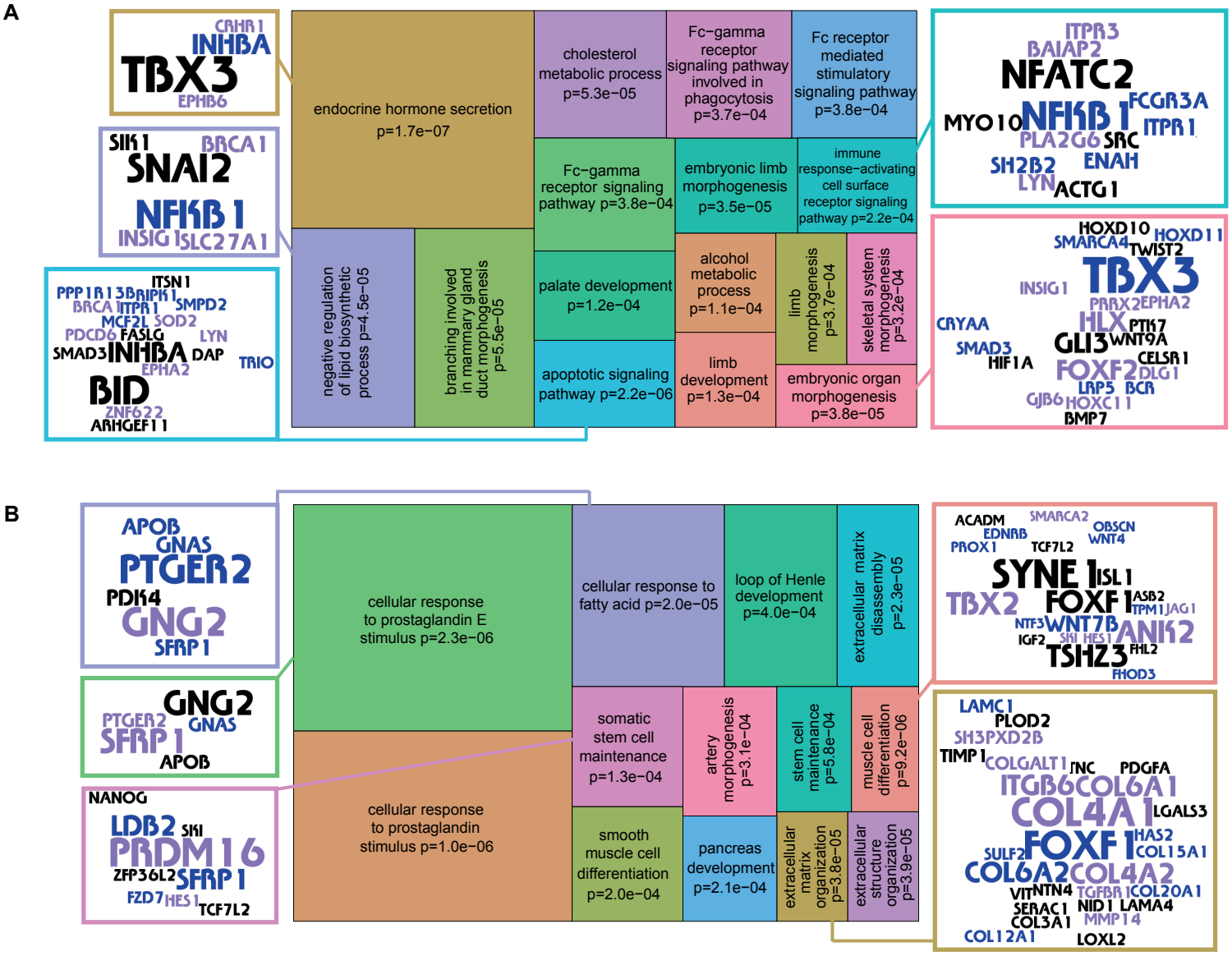

A
Hypermethylated
Hypomethylated
Biological Process
11.17
9.72
9.04
8.82
8.78
8.04
7.30
7.18
6.88
6.75
6.72
6.44
6.40
6.29
5.98
5.69
5.43
5.39
5.30
5.16
anterior/posterior axis specification
regulation of multicellular organismal development
regulation of developmental process
cardiac conduction system development
anterior/posterior axis specification, embryo
regulation of cell development
epithelium development
growth
cardiac chamber morphogenesis
organ morphogenesis
cardiac muscle tissue development
tissue morphogenesis
tissue development
regulation of cell differentiation
regulation of neuron differentiation
epithelial tube morphogenesis
regulation of neurogenesis
anatomical structure formation involved in morphogenesis
tube morphogenesis
developmental growth involved in morphogenesis
Molecular Funciton
5.58
4.43
scavenger receptor activity
cargo receptor activity
Cellular Component
6.37
4.92
4.78
4.49
extracellular region part
extracellular matrix
proteinaceous extracellular matrix
extracellular matrix part
Extracellular region part
BMP5
SULF2 COLEC12
C1QTNF1 PTHLH
COL4A1
MMP16
NID1
COL4A2
GLDN
SMOC2
ADAMTS10
PCSK6
VWF
GDF6
ADAMTS15
THBS2
GDF15
LAMB4
SCARA3
SLIT3
SBSPON
NRG1
AMELX
PGF
OTOP1
Regulation of cell development
TENM4
ADRA2C
STMN2
AKAP6
PTK7
PDE3A
TBX5
BMP5
TBX3
FXN
FKBP4
TENM3
HOXD3
GDF6
TCF4
NRCAM
TNIK
NEURL
Gland development
SEMA3C
TGFBR1
UBE3A
FKBP4
FEM1B
PITX2
NEURL
FSTL3
TGM2
IGFBP5
NCOA3
NR3C1
ESR1
IRS2
ZDHHC21
Biological Process
13.07
12.42
12.08
11.79
11.53
10.71
9.33
8.81
8.79
8.48
6.86
4.66
3.76
multicellular organismal macromolecule metabolic process
multicellular organismal metabolic process
multicellular organismal catabolic process
gland development
collagen metabolic process
urogenital system development
epithelium development
extracellular matrix organization
extracellular structure organization
tissue morphogenesis
neural nucleus development
angiogenesis
actin filament organization
Cellular Component
6.36
extracellular matrix
Extracellular matrix
ADAMTS15
COL5A3
SMOC2
COL1A1
COL15A1
VIT
THBS2
HPSE2
OLFML2B
COL4A1
PCSK6
ADAMTS2
BGN
CPXM2
COL4A2
FLRT2
COL6A3
EGFLAM
ADAMTS6
TINAGL1
FBLN5
B
-log(p-value)
0
2
4
6
8
10
12
25
21
7
53
53
11
52
10
8
57
34
8
19
13
14
48
10
57
9
52
Development of body trunk
Invasion of cells
Abnormal morphology of zeugopod
Melanoma
Skin lesion
Morphology of limb
Skin cancer
Abnormal morphology of limb
Abnormal morphol appendicular skeleton
Gastrointestinal adenocarcinoma
Organismal death
Morphology of limb bone
Abnormal morphology of head
Abnormal morphology of bone
Cardiogenesis
Malignant cutaneous melanoma cancer
Abnormal morphology of skeleton
Colorectal neoplasia
Abnormal morphology of skull
Colorectal adenocarcinoma
-log(p-value)
0
2
4
6
8
10
12
59
27
135
151
138
56
53
55
69
155
52
157
105
111
156
31
31
31
33
34
Cell movement
Morphology of bone
Malignant neoplasm of large intestine
Digestive system cancer
Gastrointestinal tract cancer
Morphology of cells
Migration of cells
Mammary tumor
Breast or ovarian cancer
Abdominal cancer
Breast cancer
Neoplasia of epithelial tissue
Gastrointestinal carcinoma
Breast or colorectal cancer
Epithelial cancer
Cell movement of tumor cell lines
Invasion of cells
Abnormal morphology of head
Morphology of head
Development of vasculature
Cancer
Organismal Injury and Abnormalities
Organismal Development

Supplemental Fig. S5. Gene ontology and pathways analysis of DE-DMRs in CAFs compared to NPFs. (A) Genomic regions enrichment (GREAT) analysis of DE-DMRs. Bar charts display -log10 P values from the binomial test for selected categories. Categories associated with developmental processes or transcription factor binding are highlighted in orange, and those associated with extracellular cellular matrix are in pink. Word clouds show the genes comprising each of the categories, with word size proportional to the number of proximal DMRs. (B) Ingenuity Pathway Analysis of DE-DMRs. Bar graphs showing the top 20 annotations by P value for diseases and biological functions that are enriched among DE-DMRs. The height of each bar represents the -log10 P value for each annotation. The number of genes associated with each annotation is also shown. Coloured circles denote annotations that belong to the top 3 categories among the dataset.


A
B
C
D
E
NPF
CAF
ABCB4
Methylation (%)
(P = 6.4e-4)
Chr 7
87.10395mb
87.10399mb
87.10403mb
AUC=0.89
TPR
FPR
(P = 3.1e-2)
Log2 relative expression
(rho= -0.55, P = 8.9e-4)
CYP1B1
(P = 4.1e-4)
Chr 2
38.30472mb
38.30476mb
38.30480mb
AUC=0.93
(P = 8.8e-1)
(rho= -0.13, P= 5.8e-1)
ESR1
(P = 2.5e-4)
Chr 6
152.01315mb
152.01318mb
152.01322mb
AUC=0.88
(P = 2.7e-4)
(rho= -0.48, P= 3.4e-2)
GDF6
(P = 2.4e-5)
Chr 8
97.16779mb
97.16783mb
97.16787mb
AUC=0.96
(P = 1.5e-2)
(rho=0.64, P = 2.4e-3)
GUCY1B3
(P = 2.7e-3)
Chr 4
156.67974mb
156.67978mb
156.67982mb
AUC=0.84
(P = 1.6e-3)
(rho= -0.16, P = 5.1e-1)
IGFBP2
(P = 1.7e-2)
Chr 2
217.50095mb
217.50105mb
AUC=0.81
(P = 4.5e-2)
(rho= -0.83, P = 5.7e-6)

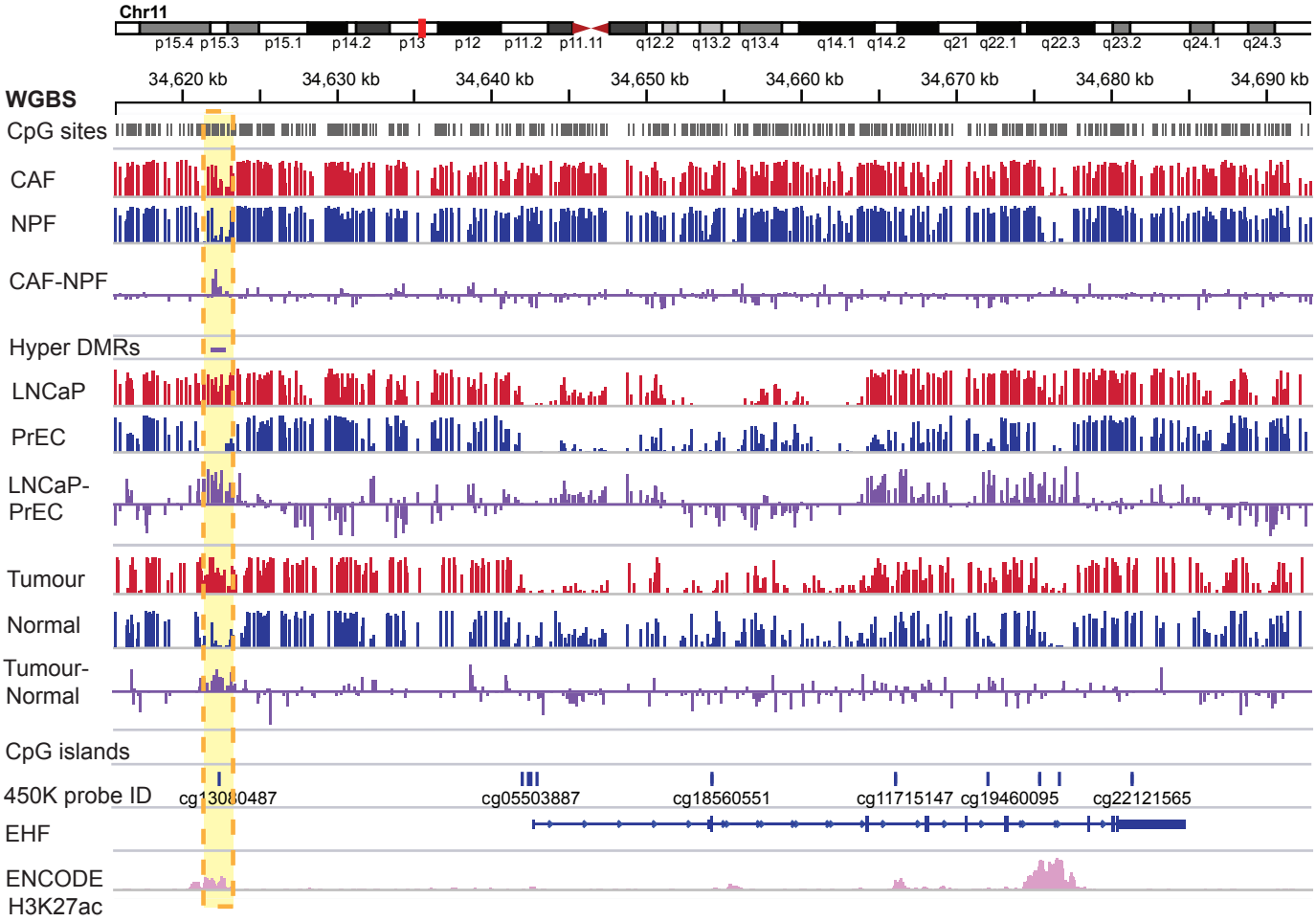


Supplemental Fig. S6. Validation of DE-DMRs in an independent cohort of patients. Candidate DMRs were examined with targeted bisulphite sequencing (A, B, C, E) and qPCR (D, E) in matched CAFs (red) and NPFs (blue) from patients 8-17 (n=10). (A) Boxplots show the mean percentage of DNA methylation across all CpG sites within each DE-DMR (paired one-sided t-test). (B) Line graphs of the DNA methylation of individual CpG sites within each DE-DMR (trendlines denote group mean methylation at each CpG). (C) The sensitivity and specificity of DNA methylation levels at each amplicon to distinguish between NPFs and CAFs. Receiver Operator Characteristic curves show the true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity) for each DMR, summarised by the Area Under the Curve. (D) qPCR data showing average log2 fold changes (±SEM) relative to the mean expression of each gene in all fibroblasts (paired one-sided t-tests). (E) Scatter plots comparing the mean percentage of DNA methylation (across all CpGs) to relative gene expression for each DE-DMR (Spearman's test). Each point represents an individual sample.

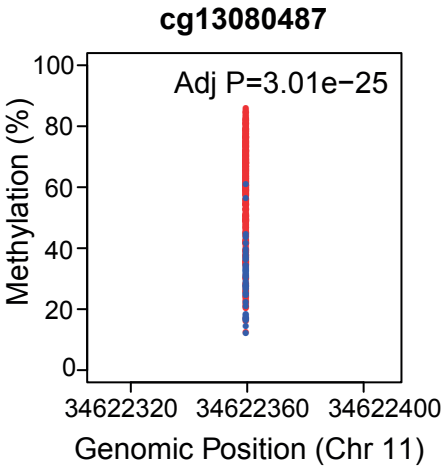

A
endocrine hormone secretion p=1.7e−07
cholesterol metabolic process p=5.3e−05
Fc−gamma receptor signaling pathway involved in phagocytosis p=3.7e−04
Fc receptor mediated stimulatory signaling pathway p=3.8e−04
Fc−gamma receptor signaling pathway p=3.8e−04
embryonic limb morphogenesis p=3.5e−05
immune response−activating cell surface receptor signaling pathway p=2.2e−04
negative regulation of lipid biosynthetic process p=4.5e−05
branching involved in mammary gland duct morphogenesis p=5.5e−05
palate development p=1.2e−04
alcohol metabolic process p=1.1e−04
limb morphogenesis p=3.7e−04
skeletal system morphogenesis p=3.2e−04
apoptotic signaling pathway p=2.2e−06
limb development p=1.3e−04
embryonic organ morphogenesis p=3.8e−05
TBX3
EPHB6
CRHR1
INHBA
SNAI2
SIK1
BRCA1
NFKB1
INSIG1
SLC27A1
ITSN1
PPP1R13B
RIPK1
SMPD2
BRCA1
ITPR1
MCF2L
SOD2
PDCD6
FASLG
LYN
SMAD3
INHBA
DAP
TRIO
EPHA2
BID
ZNF622
ARHGEF11
ITPR3
BAIAP2
NFATC2
MYO10
NFKB1
FCGR3A
ITPR1
PLA2G6
SRC
SH2B2
ENAH
LYN
ACTG1
HOXD10
HOXD11
SMARCA4
TWIST2
TBX3
INSIG1
PRRX2
EPHA2
CRYAA
HLX
PTK7
SMAD3
GLI3
WNT9A
HIF1A
FOXF2
CELSR1
DLG1
LRP5
BCR
GJB6
HOXC11
BMP7
B
cellular response to prostaglandin E stimulus p=2.3e−06
cellular response to fatty acid p=2.0e−05
loop of Henle development p=4.0e−04
extracellular matrix disassembly p=2.3e−05
somatic stem cell maintenance p=1.3e−04
artery morphogenesis p=3.1e−04
stem cell maintenance p=5.8e−04
muscle cell differentiation p=9.2e−06
cellular response to prostaglandin stimulus p=1.0e−06
smooth muscle cell differentiation p=2.0e−04
pancreas development p=2.1e−04
extracellular matrix organization p=3.8e−05
extracellular structure organization p=3.9e−05
APOB
GNAS
PTGER2
PDK4
GNG2
SFRP1
GNG2
PTGER2
GNAS
SFRP1
APOB
NANOG
LDB2
SIX1
PRDM16
ZFP36L2
SFRP1
FZD7
HES1
TCF7L2
ACADM
SMARCA2
OBSCN
EDNRB
WNT4
PROX1
TCF7L2
SYNE1
ISL1
TBX2
FOXF1
ASB2
TPM1
JAG1
NTF3
WNT7B
ANK2
IGF2
SIX1
HES1
TSHZ3
FHL2
FHOD3
LAMC1
PLOD2
SH3PXD2B
TIMP1
COLGALT1
TNC
PDGFA
ITGB6
COL6A1
COL4A1
LGALS3
FOXF1
HAS2
SULF2
COL15A1
COL6A2
COL4A2
VIT
NTN4
TGFBR1
COL20A1
SERAC1
NID1
LAMA4
COL3A1
MMP14
COL12A1
LOXL2
C
Hypermethylated
Hypomethylated
-log(p-value)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
78
58
103
41
47
42
54
61
226
215
209
64
52
212
76
190
66
100
210
75
Differentiation of cells
Cell proliferation of tumor cell lines
Proliferation of cells
Development of head
Apoptosis of tumor cell lines
Development of body axis
Cell death of tumor cell lines
Transcription of RNA
Cancer
Tumorigenesis of tissue
Abdominal neoplasm
Transcription
Transcription of DNA
Neoplasia of epithelial tissue
Necrosis
Adenocarcinoma
Expression of RNA
Neoplasia of cells
Epithelial cancer
Apoptosis
155
140
147
200
200
243
246
343
181
204
181
344
193
130
136
128
314
280
321
81
Uterine tumor
Endometrial cancer
Uterine cancer
Tumorigenesis of genital organ
Genital tumor
Abdominal adenocarcinoma
Abdominal carcinoma
Abdominal cancer
Tumorigenesis of reproductive tract
Pelvic cancer
Female genital neoplasm
Abdominal neoplasm
Genital tract cancer
Endometrial carcinoma
Uterine carcinoma
Adenocarcinoma in endometrium
Adenocarcinoma
Gastrointestinal tract cancer
Digestive system cancer
Gastroesophageal adenocarcinoma
Cancer
Organismal Injury and Abnormalities
Reproductive System Disease

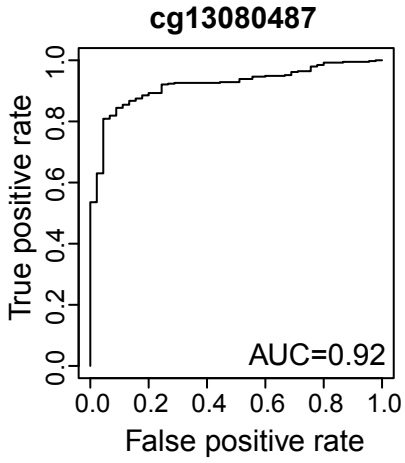
Supplemental Fig. S7. Gene ontology and pathways analysis of tsDMRs in CAFs compared to NPFs. Genomic regions enrichment (GREAT) analysis of (A) hypermethylated and (B) hypomethylated tsDMRs showing enriched biological processes with raw P-values from the binomial test. Word clouds show the genes comprising categories, with word size proportional to the number of proximal DMRs. (C) Ingenuity Pathway Analysis of tsDMRs. Bar graphs showing the top 20 annotations by P value for diseases and biological functions that are enriched among tsDMRs. The height of each bar represents the -log10 P value for each annotation. The number of genes associated with each annotation is also shown. Coloured circles denote annotations that belong to the top 3 categories among the dataset.

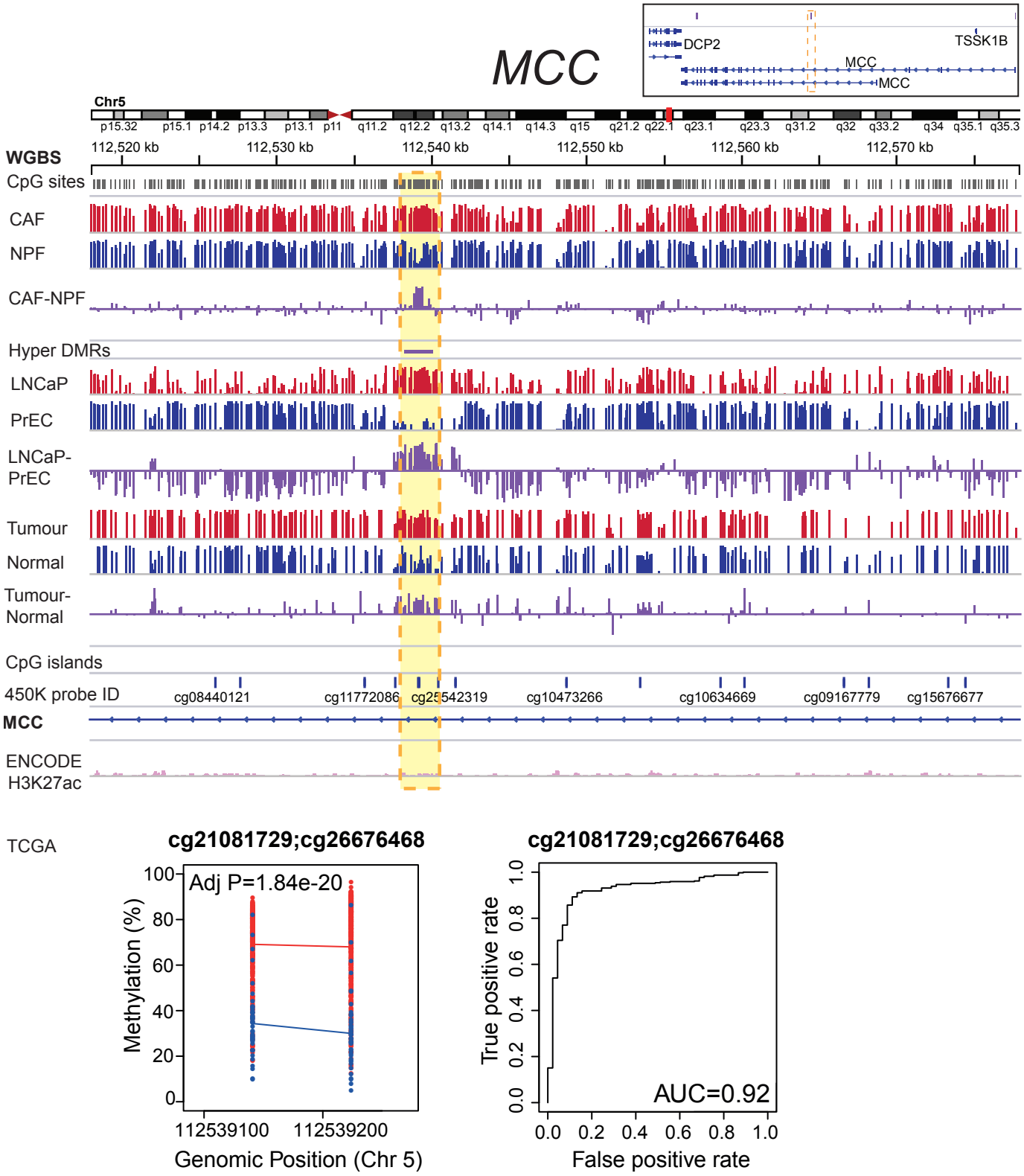
EHF

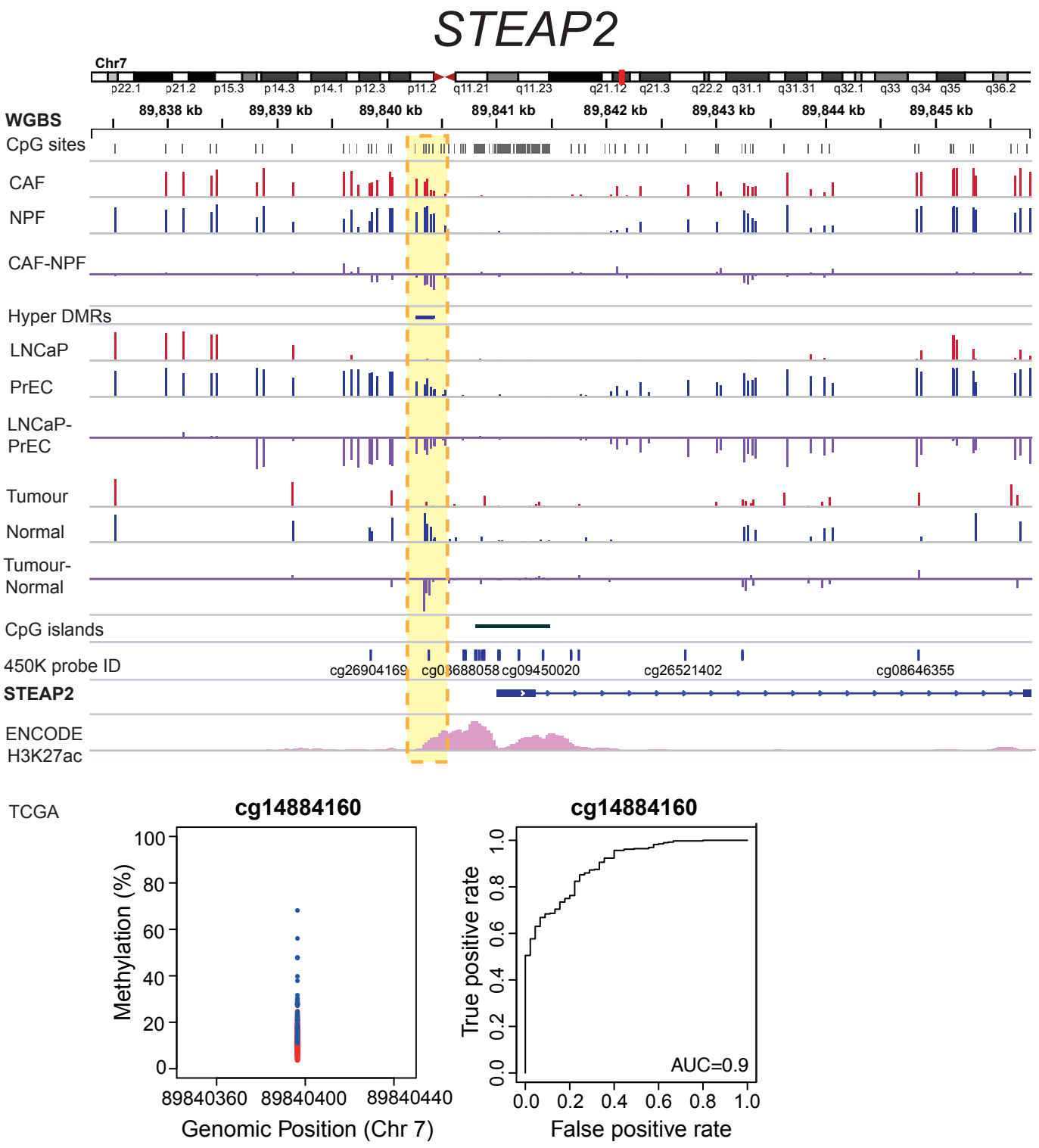

Chr11
p15.4 p15.3 p15.1 p14.2 p13 p12 p11.2 p11.11 q12.2 q13.2 q13.4 q14.1 q14.2 q21 q22.1 q22.3 q23.2 q24.1 q24.3
34,620 kb
34,630 kb
34,640 kb
34,650 kb
34,660 kb
34,670 kb
34,680 kb
34,690 kb
WGBS
CpG sites
CAF
NPF
CAF-NPF
Hyper DMRs
LNCaP
PrEC
LNCaP-PrEC
Tumour
Normal
Tumour-Normal
CpG islands
450K probe ID
cg13080487
cg05503887
cg18560551
cg11715147
cg19460095
cg22121565
EHF
ENCODE H3K27ac

TCGA


cg13080487
Adj P=3.01e−25
Methylation (%)
100
80
60
40
20
0
34622320
34622360
34622400
Genomic Position (Chr 11)


cg13080487
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.92
False positive rate


MCC
DCP2
TSSK1B
MCC
MCC
Chr5
p15.32 p15.1 p14.2 p13.3 p13.1 p11 q11.2 q12.2 q13.2 q14.1 q14.3 q15 q21.2 q22.1 q23.1 q23.3 q31.2 q32 q33.2 q34 q35.1 q35.3
112,520 kb
112,530 kb
112,540 kb
112,550 kb
112,560 kb
112,570 kb
WGBS
CpG sites
CAF
NPF
CAF-NPF
Hyper DMRs
LNCaP
PrEC
LNCaP-PrEC
Tumour
Normal
Tumour-Normal
CpG islands
450K probe ID
cg08440121
cg11772086
cg25542319
cg10473266
cg10634669
cg09167779
cg15676677
MCC
ENCODE H3K27ac
TCGA
cg21081729;cg26676468
Adj P=1.84e-20
Methylation (%)
100
80
60
40
20
0
112539100
112539200
Genomic Position (Chr 5)
cg21081729;cg26676468
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.92
False positive rate



Supplemental Fig. S8. Tumour-specific DMRs in CAFs and cancer cells. Summary of the DNA methylation profiles of DMRs in the EHF, MCC and STEAP2 gene. WBGS data for CAFs versus NPFs (n=4), LNCaP versus PrEC cells, and tumour versus matched normal clinical samples (n=4 patients) are shown. Yellow boxes highlight the tsDMRs. Corresponding DNA methylation levels are shown for TCGA samples (tumour = red, normal = blue) based on 450K probes within the tsDMRs. The ROC curves show the ability of the 450K data to discriminate between matched cancer versus normal TCGA tissues.