Interplay between Genome and Epigenome of Prostate Cancer

Jing Li

Mentor: Yinghao Sun

In this study, we synthesized the first

Chinese Prostate Cancer Genome and Epigenome Atlas (CPGEA).

Data items

Figure 1A: Genomic and epigenomic abnormality across all tumor specimen (multi-panel bar graph).

Figure 1B: Mutation signature.

Figure 2: The Big Big Circle.

Figure 3: Non-coding SNV and recurrent ncSNV.

Figure 4: FOXA1

Figure 5: DNA methylation

Figure 6: Subtyping

Figure 7: Pathway

Table 1: Clinical and pathological characterization of specimen

Supplementary Figure 1: Diagram of the study design/scope

Supplementary Figure 2A: GISTIC results

Supplementary Figure 2B: Examples of amplification/deletion (METTL24 or a new one)

Supplementary Figure 3ABC: Example of fusion

Supplementary Figure 4: SMG waterfall plot

Supplementary Figure 5: FOXA1 validation

Supplementary Figure 6: Non-CG methylation differences

Supplementary Figure 7: DMR stats

Supplementary Figure 8: DMR enrichment

Supplementary Figure 9: epi-mutation and IDH/TET genes

Supplementary Table 1: Molecular and pathological estimates of specimen

Supplementary Table 2: Datasets generated in this study

Supplementary Table 3: Statistics underlying Figure 1

Supplementary Table 4: GISTIC result, arm level

Supplementary Table 5: GISTIC, focal (with gene names)

Supplementary Table 6: Fusion events

Supplementary Table 7: RT-PCR primers for fusion detection

Supplementary Table 8: Description of 13 cohorts

Supplementary Table 9: Statistics underlying Figure 2

Supplementary Table 10: Genomic distribution of non-coding SNV

Supplementary Table 11: Potential functional ncSNV

Supplementary Table 12: FOXA1 validation

First Goal：direct comparison between Chinese and Western

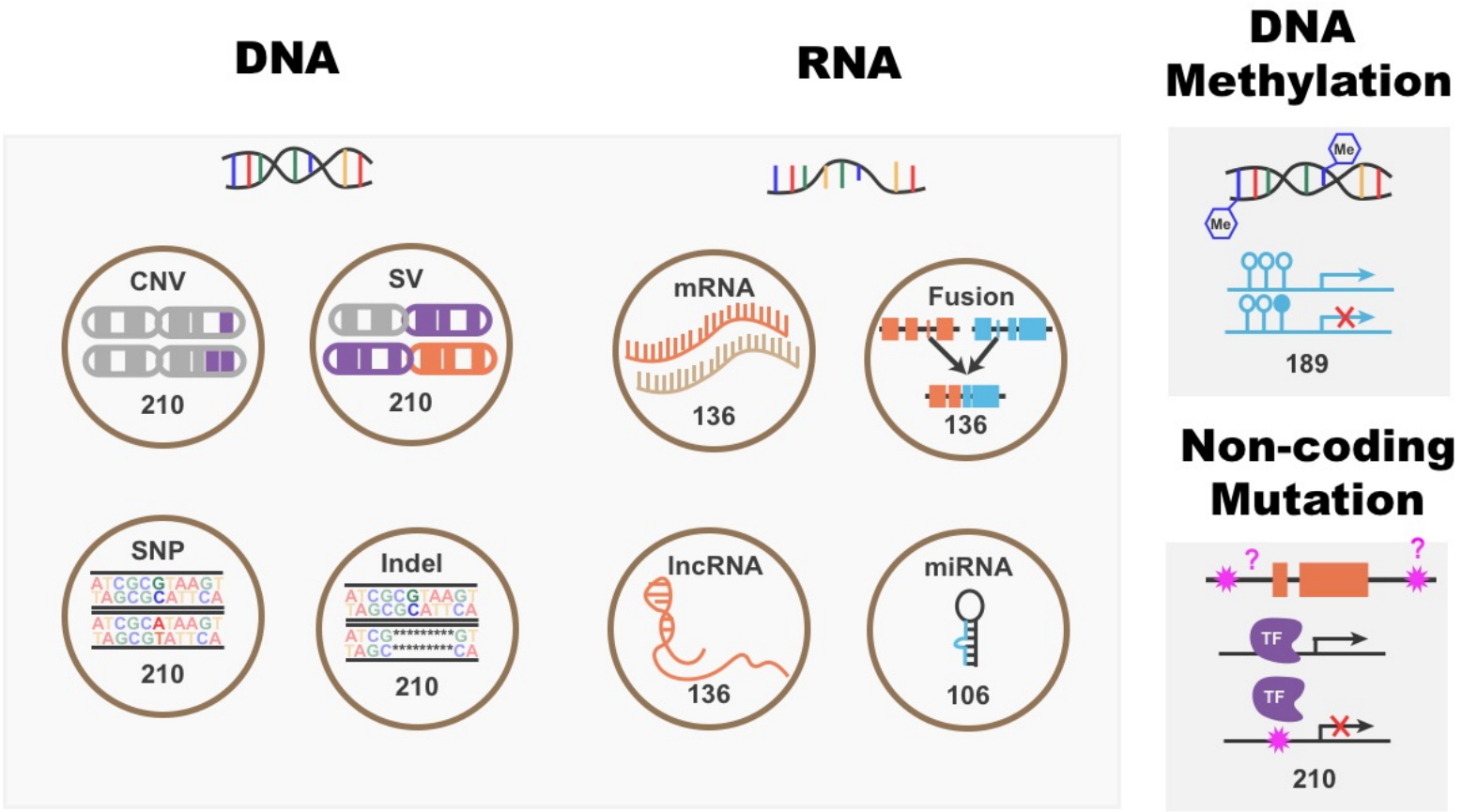
We integrated existing genomic datasets from 2,775 prostate tumors representing 13 western cohorts

Clinical samples and histopathologic data

Table 1. Clinical characteristics comparison

	CPGEA	TCGA[5]
Clinical Feature		
Age	69 (50-88)	61 (43-76)
Pre-operative PSA	18.72 (1.25-1161.92)	7.4 (1.6-87.0)
Gleason Score		
3+3	17	65
3+4	56	102
4+3	41	78
⩾8	95	88
Pathologic Stage		
pT2a/b	28	18
pT2c	78	111
pT3a	61	110
pT3b	33	82
pT4	10	6
PSA Recurrence		
Yes	61	33
No	128	248
Not available	21	47
Lymph Nodes Metastasis		
Positive	25	
Negative	154	
Not available	31	
Bone Metastasis		
Positive	15	
Negative	195	
Not available	/	
Margin Status		
Positive	83	69
Negative	125	193
Not available	2	71

We applied four omics technologies to characterize isolated biomolecules from the Chinese prostate cancer cohort: whole genome sequencing (WGS, 112.03× on average per sample), whole genome bisulfite sequencing (WGBS, 29.5× average coverage per CpG per sample), RNA sequencing (138.51M mapped paired-end reads on average) and miRNA sequencing (19.75M mapped reads on average). Overall, our first Chinese Prostate Cancer Genome and Epigenome Atlas was consisted of 420 WGS, 272 RNA-seq, 378 WGBS, and 212 miRNA-seq datasets (Supplementary Figure 1, Supplementary Table 2).



Landscape

CNV

- CNV arm level
- CNV gene level
- CNV comparison

Fusion

- Fusion
- Fusion comparison

SNV

- SMG
- SMG comparison

Cluster

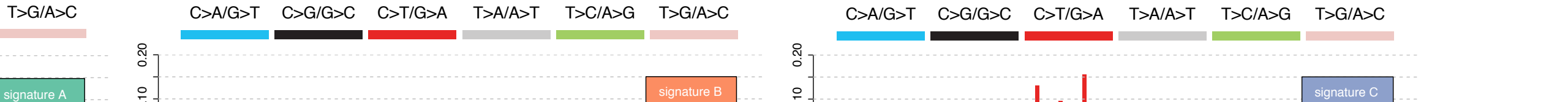
Non-coding

DNA methylation

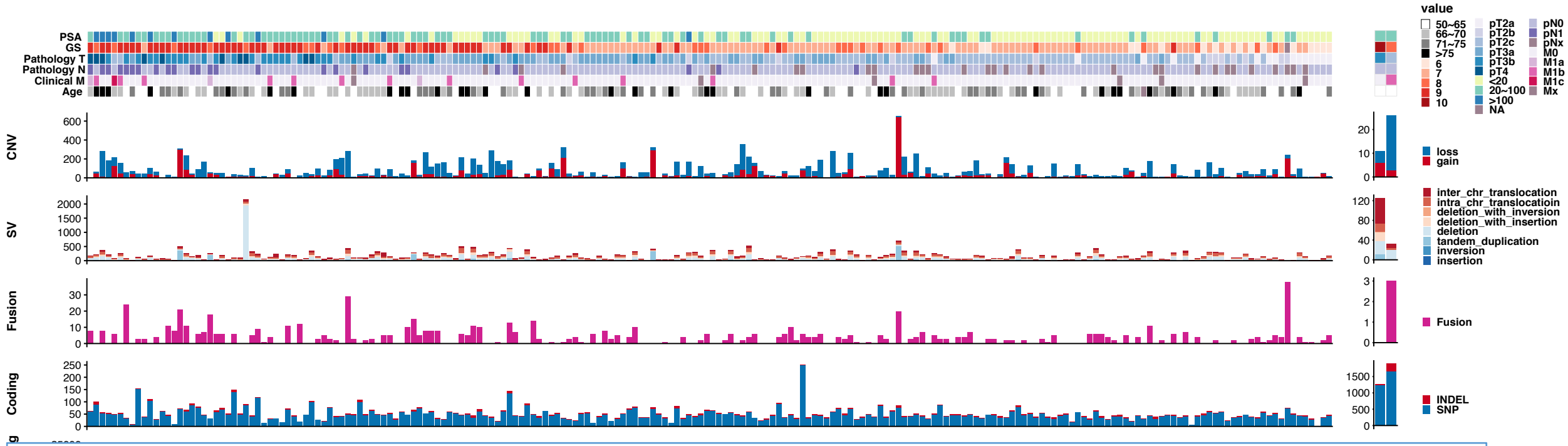
Somatic mutation landscape of prostate cancer of Chinese and Western men

Somatic mutation landscape of prostate cancer of Chinese and Western men

We observed two cases with extreme hypermutation phenotypes. Two out of three were identified the presence of DNA repair genes' mutation. Patient T13 had a germline mutation of MSH6 (NM_000179.2(MSH6):c.3991C>T (p.Arg1331Ter). While, patient T502 has a somatic indel of MLH1. To reduce the effect of mutation burden confounding, we excluded outlier samples for the further analysis.



Somatic mutation landscape of prostate cancer of Chinese and Western men

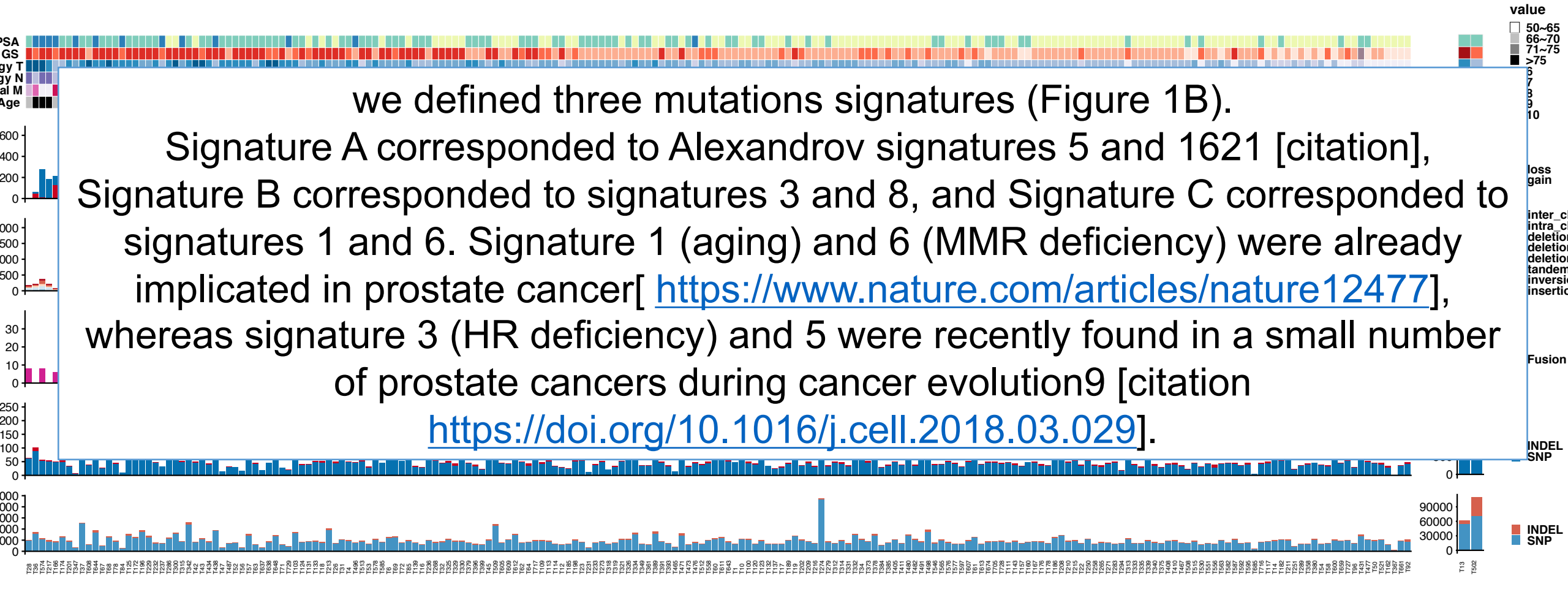


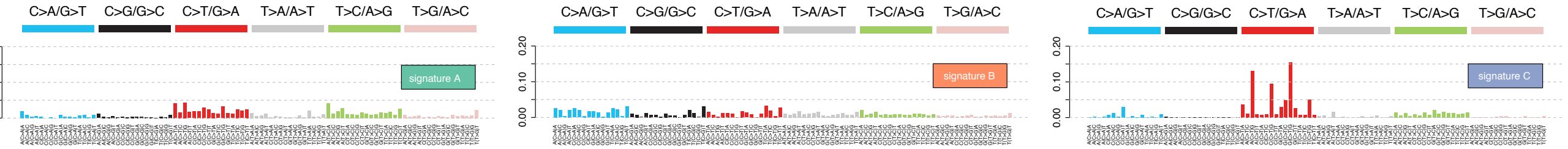
The average genome-wide substitution rate was 1.5197 per Mb (range 0.1728-7.4079 per Mb). The mutational burden across our cohort was 1.10750 per Mb (range 0.02778-4.58333 per Mb), which corresponded to **39.87 non-synonymous mutations per tumor genome (median:36; 25th:29 and 75th:46 percentiles respectively).**

This estimate was remarkably consistent with mutation burden estimated from exome sequencing from western cohort[19]

Somatic mutation landscape of prostate cancer of Chinese and Western men

we defined three mutations signatures (Figure 1B). Signature A corresponded to Alexandrov signatures 5 and 1621 [citation], Signature B corresponded to signatures 3 and 8, and Signature C corresponded to signatures 1 and 6. Signature 1 (aging) and 6 (MMR deficiency) were already implicated in prostate cancer[https://www.nature.com/articles/nature12477], whereas signature 3 (HR deficiency) and 5 were recently found in a small number of prostate cancers during cancer evolution9 [citation https://doi.org/10.1016/j.cell.2018.03.029].





Landscape

CNV
- CNV arm level
- CNV gene level
- CNV comparison

Fusion
- Fusion
- Fusion comparison

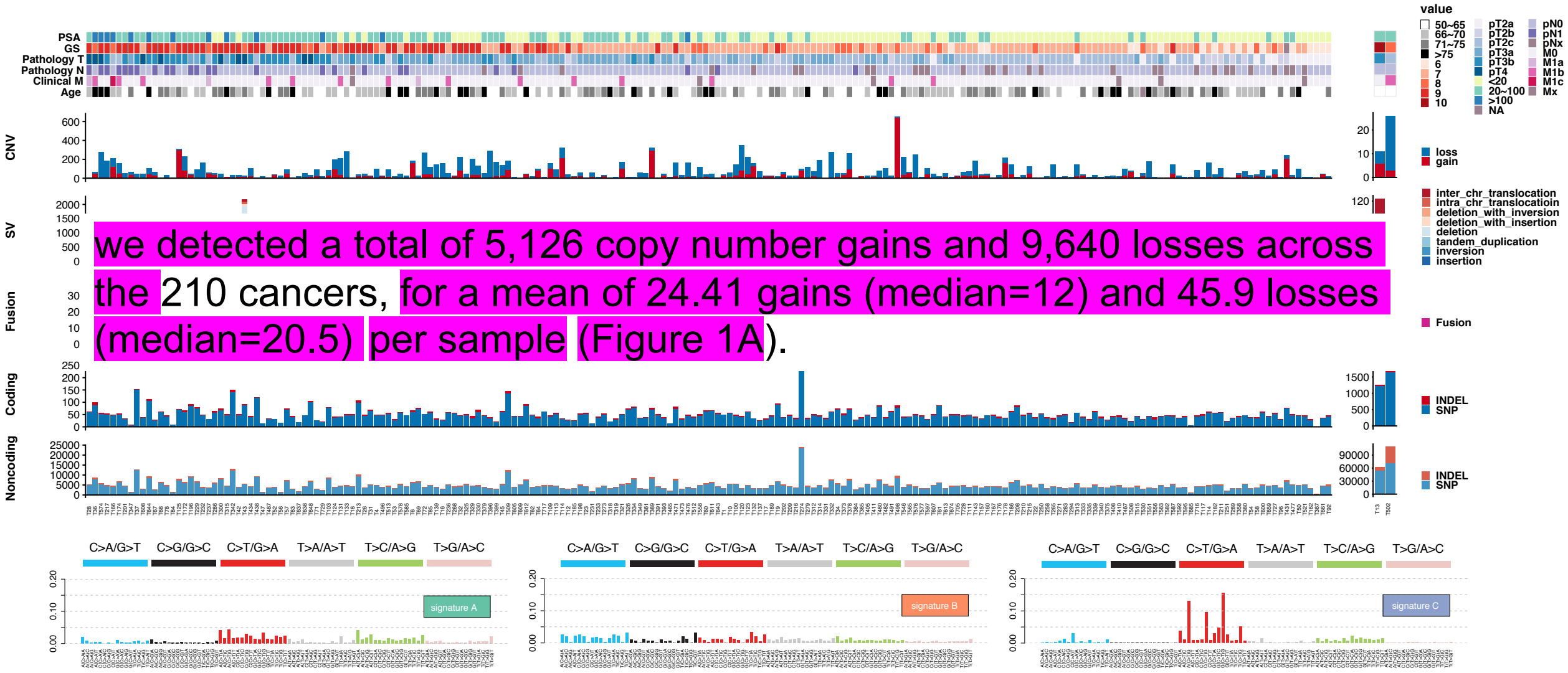
SNV
- SMG
- SMG comparison

Cluster

Non-coding

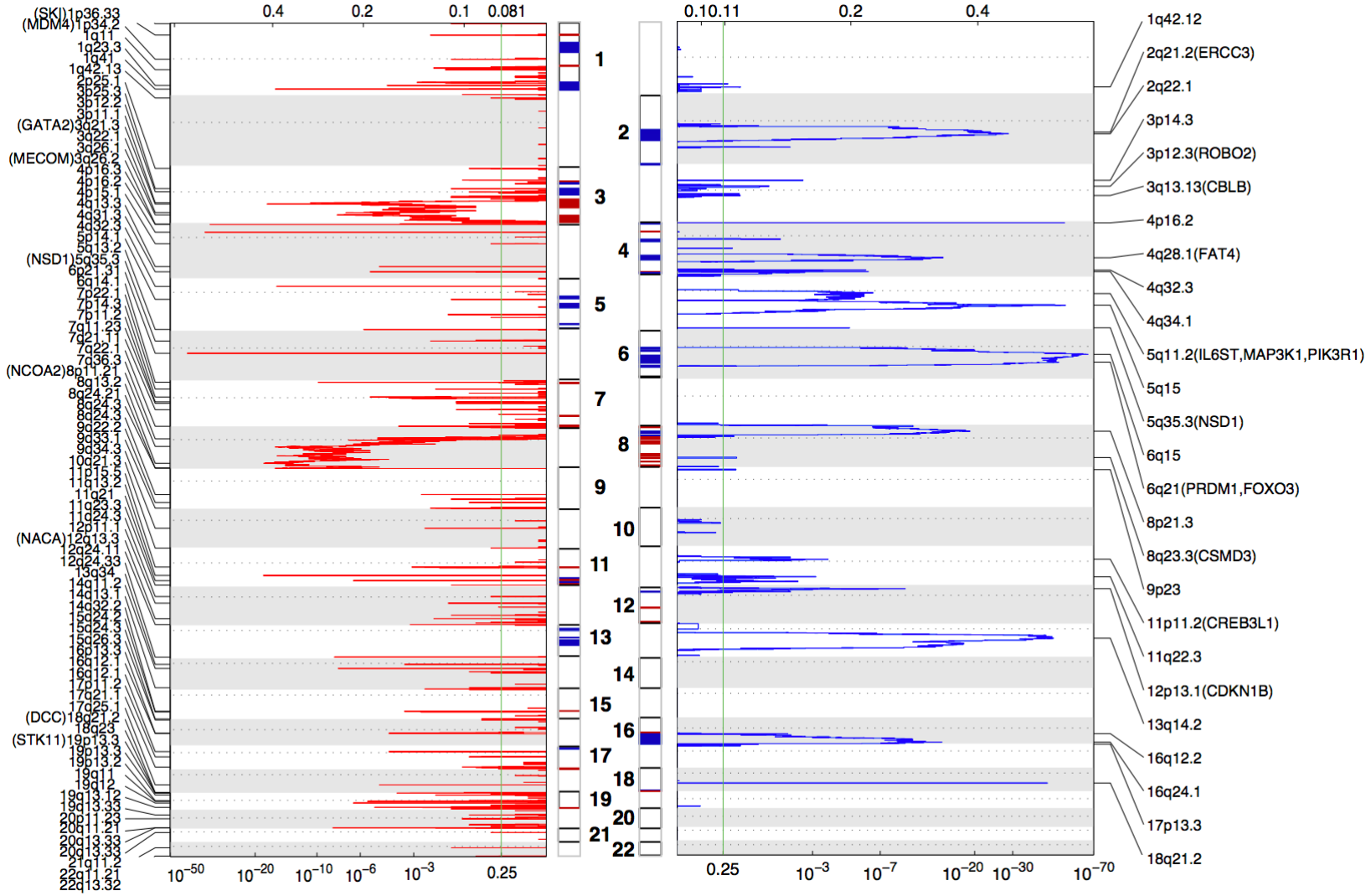
DNA methylation

Somatic mutation landscape of prostate cancer of Chinese and Western men

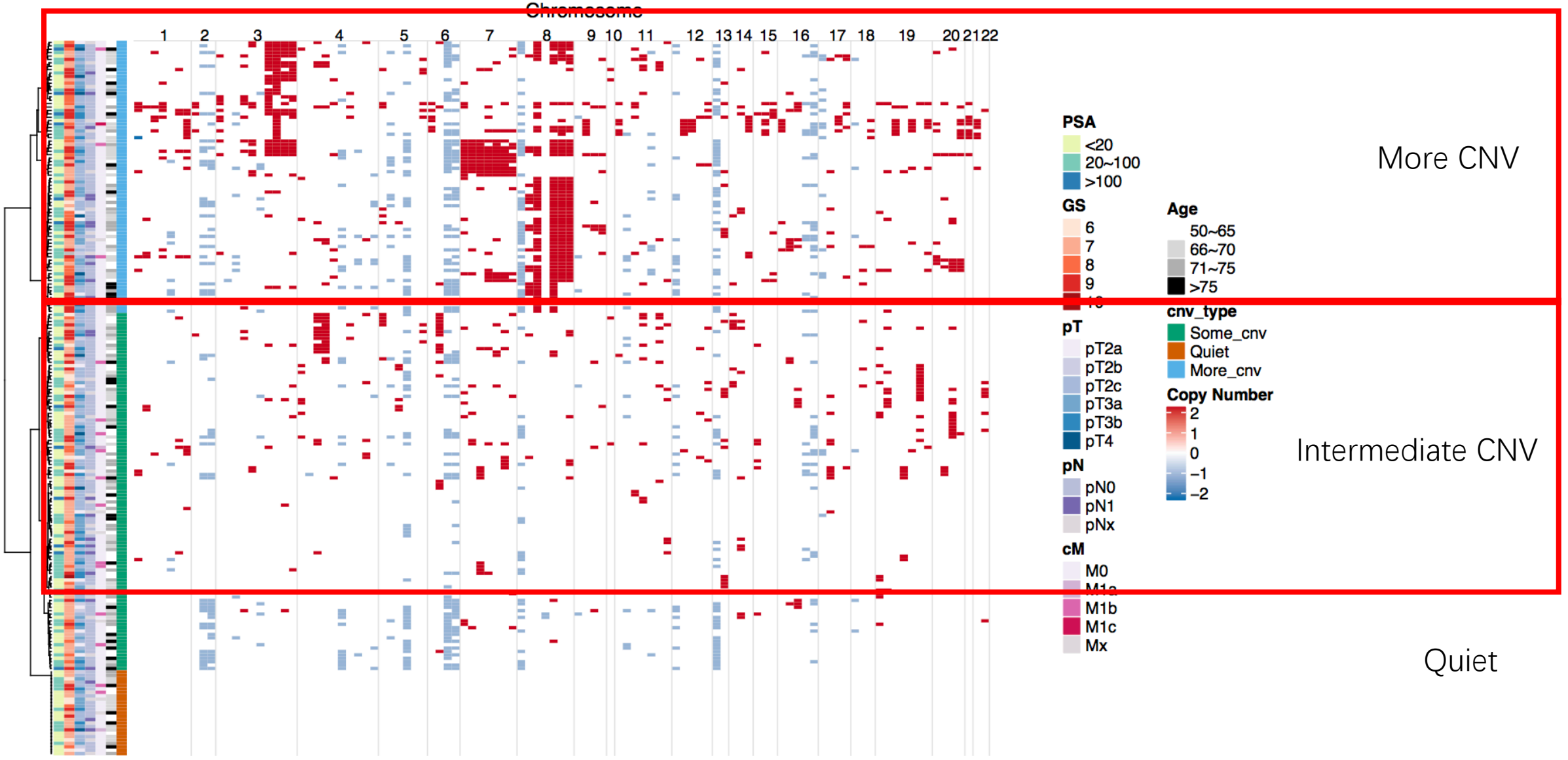


we detected a total of 5,126 copy number gains and 9,640 losses across the 210 cancers, for a mean of 24.41 gains (median=12) and 45.9 losses (median=20.5) per sample (Figure 1A).

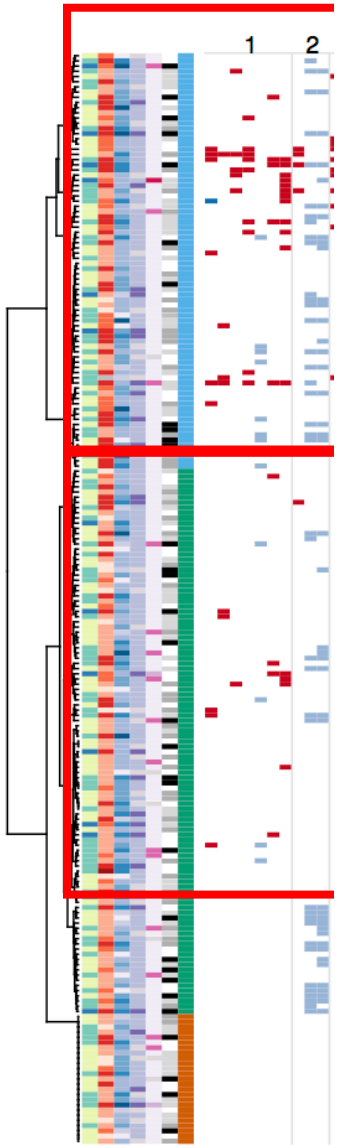
The most significantly recurrent (q<10−6) arm-level gains occurred on chromosome arms 8q (16%), 3q (8%),3p (7%), 7q(7%), 8p(7%) and 7p(6%). The most significant losses were of 8p (16%) and 16q(6%)

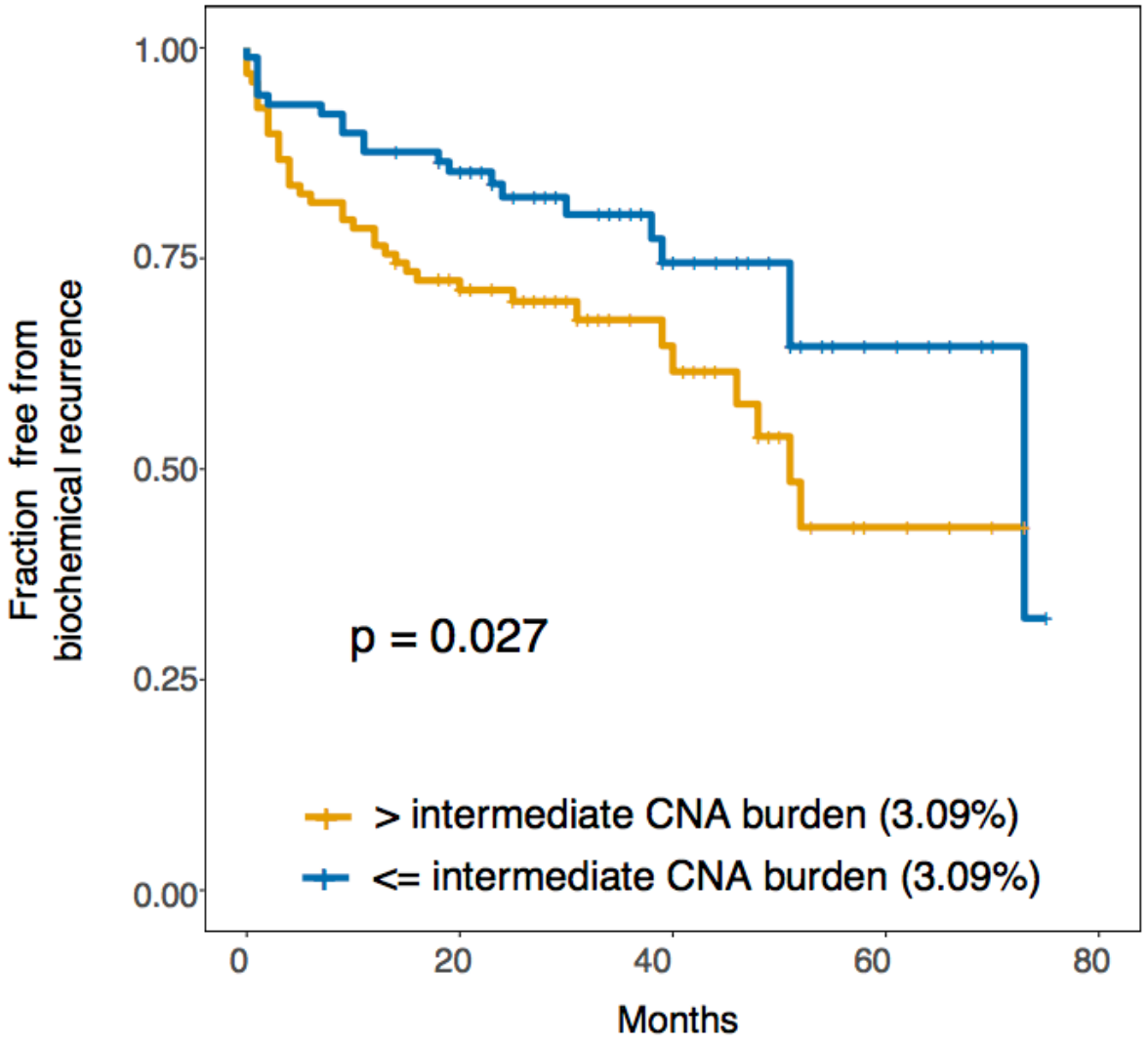


We did the clustering based on chromosomal arm level alterations as the previous TCGA paper[5]. Consistent with TCGA result, we could clearly see the trend of “more SCNA”, “Some SCNA” and “Quiet” (Supplementary Figure 2C).



We did the clustering based on chromosomal arm level alterations as the previous TCGA paper[5]. Consistent with TCGA result, we could clearly see the trend of “more SCNA”, “Some SCNA” and “Quiet” (Supplementary Figure 2C).



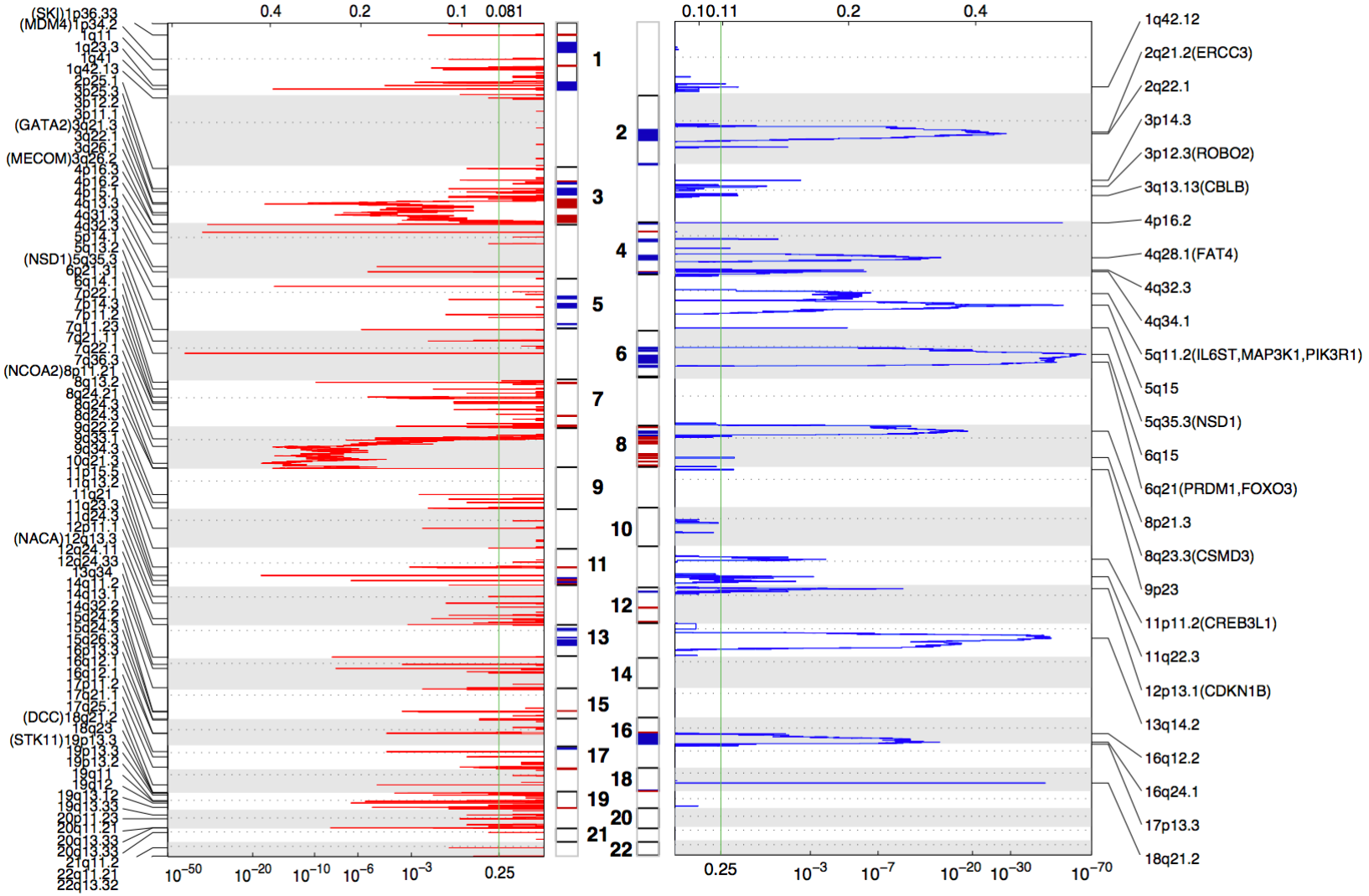


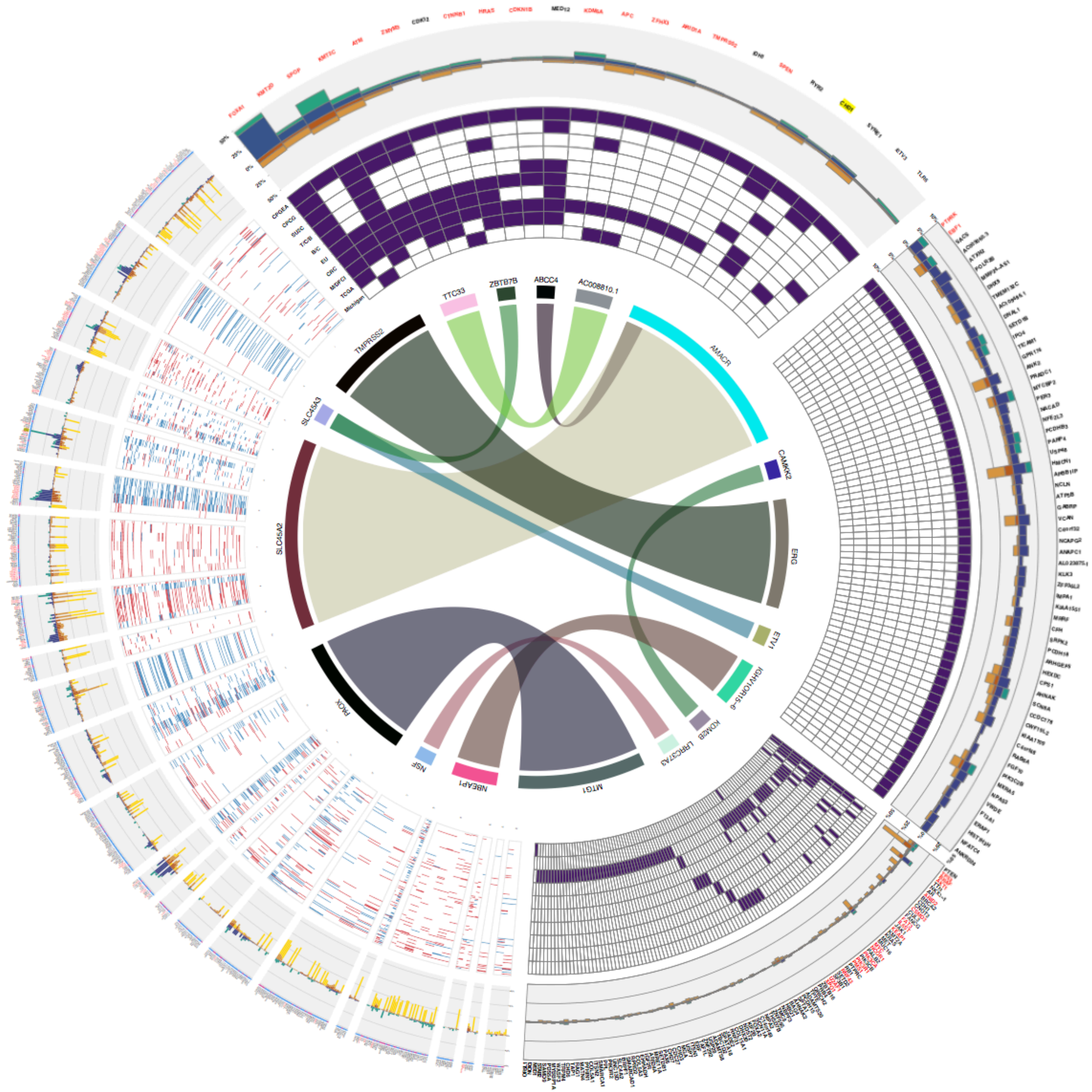
More CNV

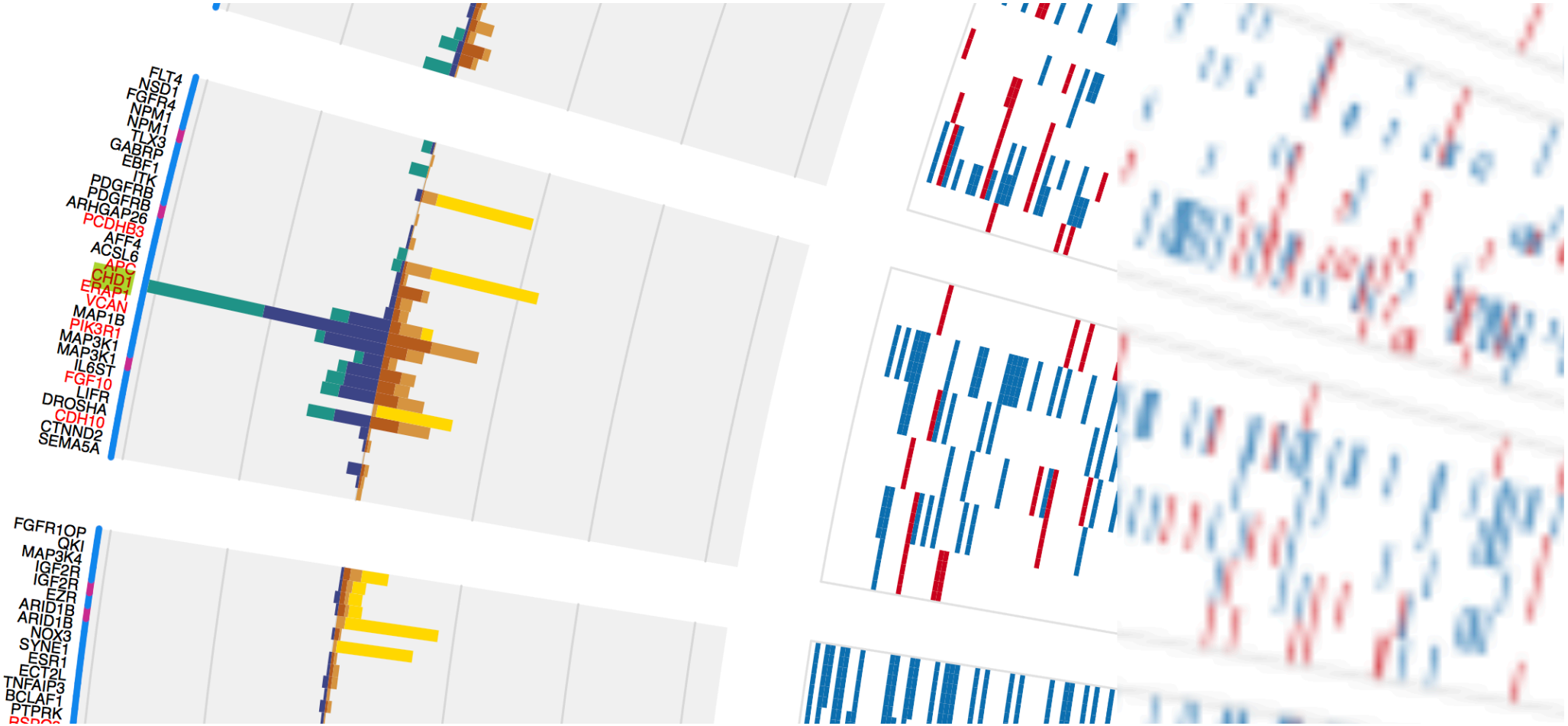
Intermediate CNV

Quiet

These results were consistent with previous studies25 [Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data, citation]. We identified 79 gains and 26 losses of significant focal copy-number alteration (Supplementary Figure 2A, Supplementary Table 5).




TMPRSS2
TTC33
ZBTB7B
ABCC4
AC008810.1
AMACR
CAMKK2
ERG
ETV1
KDM2B
LRRC37A3
MTG1
NBEAP1
NSF
PACX
SLC45A2
SLC45A3


FLT4
NSD1
FGFR4
NPM1
NPM1
TLX3
GABRP
EBF1
ITK
PDGFRB
PDGFRB
ARHGAP26
PCDHB3
AFF4
ACSL6
APC
CHD1
ERAP1
VCAN
MAP1B
PIK3R1
MAP3K1
MAP3K1
IL6ST
FGF10
LIFR
DROSHA
CDH10
CTNND2
SEMA5A
FGFR1OP
QKI
MAP3K4
IGF2R
IGF2R
EZR
ARID1B
ARID1B
NOX3
SYNE1
ESR1
ECT2L
TNFAIP3
BCLAF1
PTPRK

Landscape

CNV

- CNV arm level
- CNV gene level
- CNV comparison

Fusion

- Fusion
- Fusion comparison

SNV

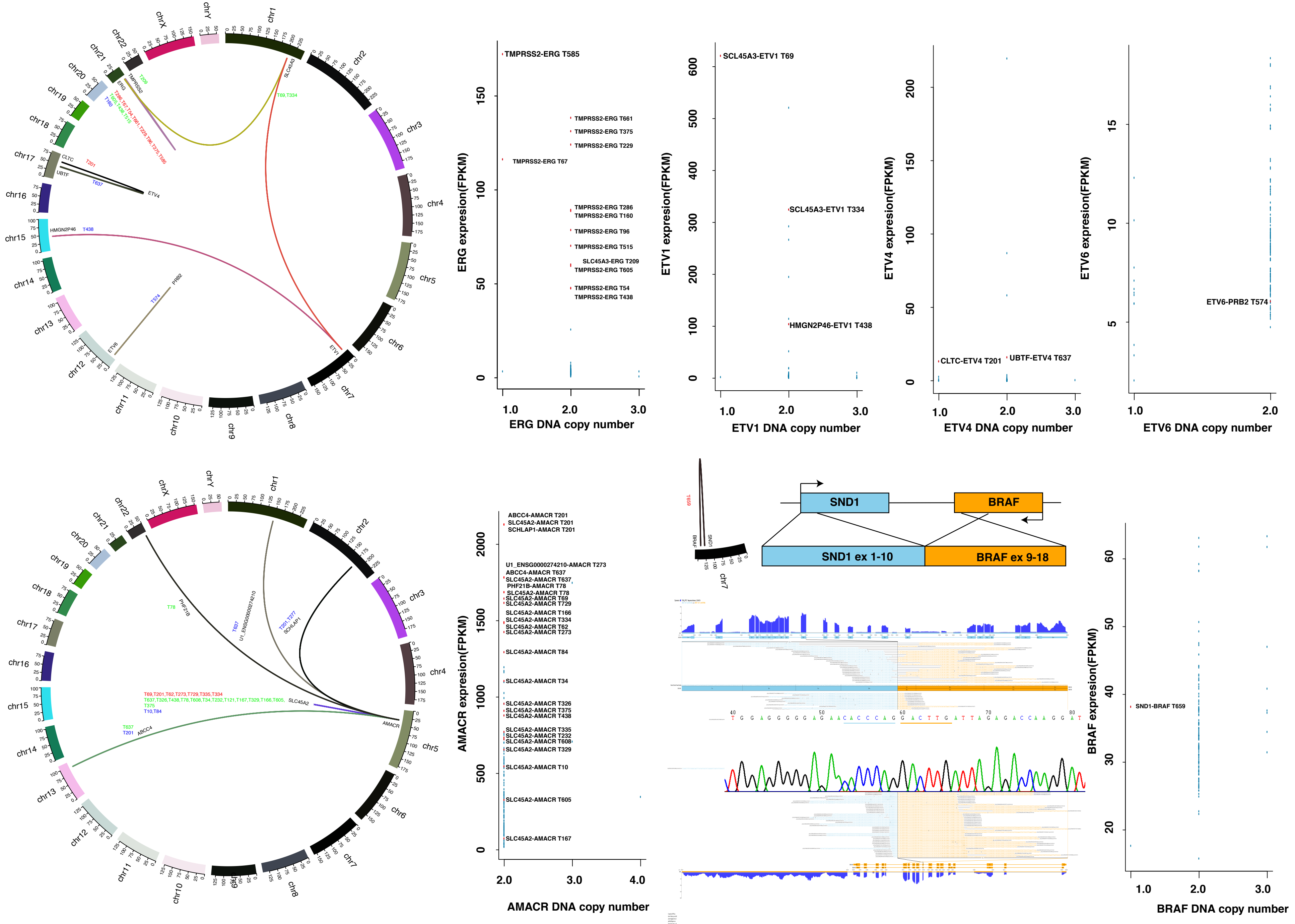
- SMG
- SMG comparison

Cluster

Non-coding

DNA methylation

We usedSOAPfuse26 [SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data, 10.1186/gb-2013-14-2-r12] and PRADA27 [PRADA: pipeline for RNA sequencing data analysis ,10.1093/bioinformatics/btu169] for fusion detection with a stringent filtering workflow (Methods). We were able to detect 182 gene-to-gene fusions (Supplementary Table 6), 46 of which were previously reported in different cancers, and 136 were novel to our study (Supplementary Table 6). 148 pairs were on the same chromosome, whereas 34 pairs were trans-chromosomal. Because fusion detection usually had high false positive rate, we experimentally validated 20 gene pairs using a total of 129 sets of RT-PCR probes (Methods, Supplementary Table 7). 91 sets were scored positive, validating 19 out of the 21 fusion candidates. In stark contrast to the Western prostate cancers, the most well-known hallmark TMPRSS2-ETS gene fusion was much lower in the Chinese cohort (12/136, 8.82%). The most prominent gene fusion event in our Chinese cohort was the AMACR rearrangement, occurring in 19.12% (26/136) of the cases (Supplementary Figure 3). SLC45A2-AMACR was reported in progressive prostate cancers (15.44%, 21/136)



TMPRSS2-ERG T585
TMPRSS2-ERG T661
TMPRSS2-ERG T375
TMPRSS2-ERG T229
TMPRSS2-ERG T67
TMPRSS2-ERG T286
TMPRSS2-ERG T160
TMPRSS2-ERG T96
TMPRSS2-ERG T515
SLC45A3-ERG T209
TMPRSS2-ERG T605
TMPRSS2-ERG T54
TMPRSS2-ERG T438
ERG expresion(FPKM)
ERG DNA copy number
SCL45A3-ETV1 T69
SCL45A3-ETV1 T334
HMGN2P46-ETV1 T438
ETV1 expresion(FPKM)
ETV1 DNA copy number
CLTC-ETV4 T201
UBTF-ETV4 T637
ETV4 expresion(FPKM)
ETV4 DNA copy number
ETV6-PRB2 T574
ETV6 expresion(FPKM)
ETV6 DNA copy number
ABCC4-AMACR T201
SLC45A2-AMACR T201
SCHLAP1-AMACR T201
U1_ENSG0000274210-AMACR T273
ABCC4-AMACR T637
SLC45A2-AMACR T637
PHF21B-AMACR T78
SLC45A2-AMACR T78
SLC45A2-AMACR T69
SLC45A2-AMACR T729
SLC45A2-AMACR T166
SLC45A2-AMACR T334
SLC45A2-AMACR T62
SLC45A2-AMACR T273
SLC45A2-AMACR T84
SLC45A2-AMACR T34
SLC45A2-AMACR T326
SLC45A2-AMACR T375
SLC45A2-AMACR T438
SLC45A2-AMACR T335
SLC45A2-AMACR T232
SLC45A2-AMACR T608
SLC45A2-AMACR T329
SLC45A2-AMACR T10
SLC45A2-AMACR T605
SLC45A2-AMACR T167
AMACR expresion(FPKM)
AMACR DNA copy number
SND1
BRAF
SND1 ex 1-10
BRAF ex 9-18
SND1-BRAF T659
BRAF expresion(FPKM)
BRAF DNA copy number

Landscape

CNV
- CNV arm level
- CNV gene level
- CNV comparison

Fusion
- Fusion
- Fusion comparison

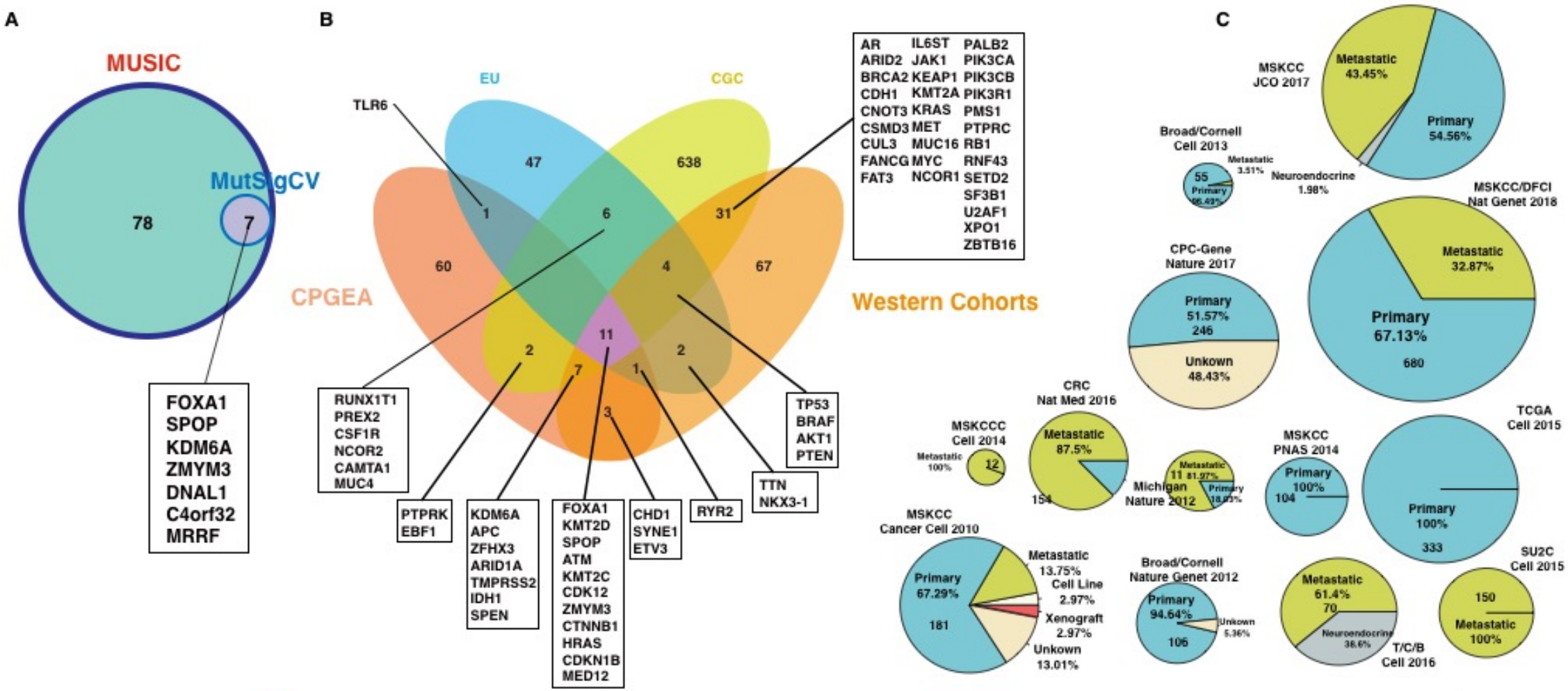
SNV
- SMG
- SMG comparison

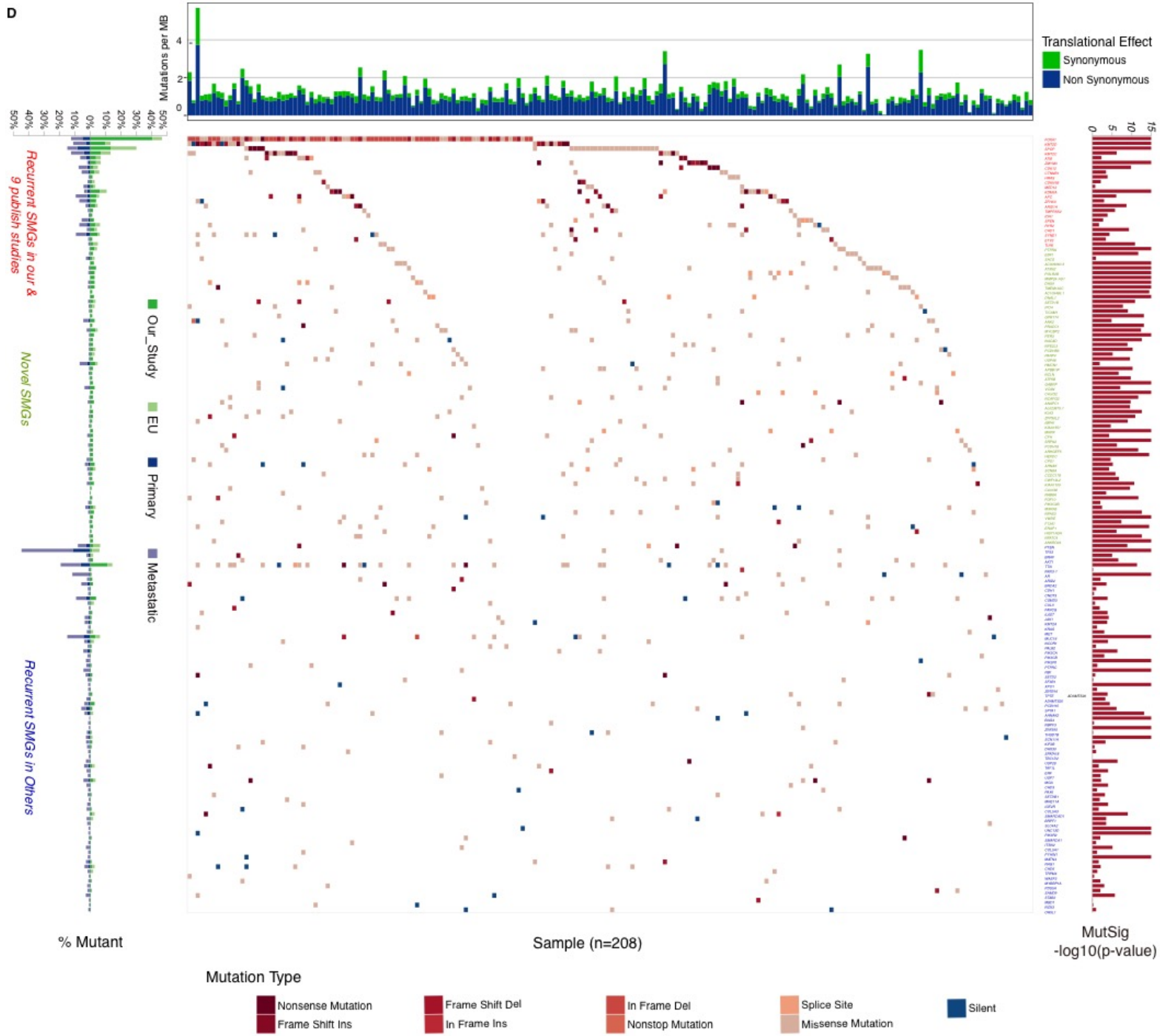
Cluster

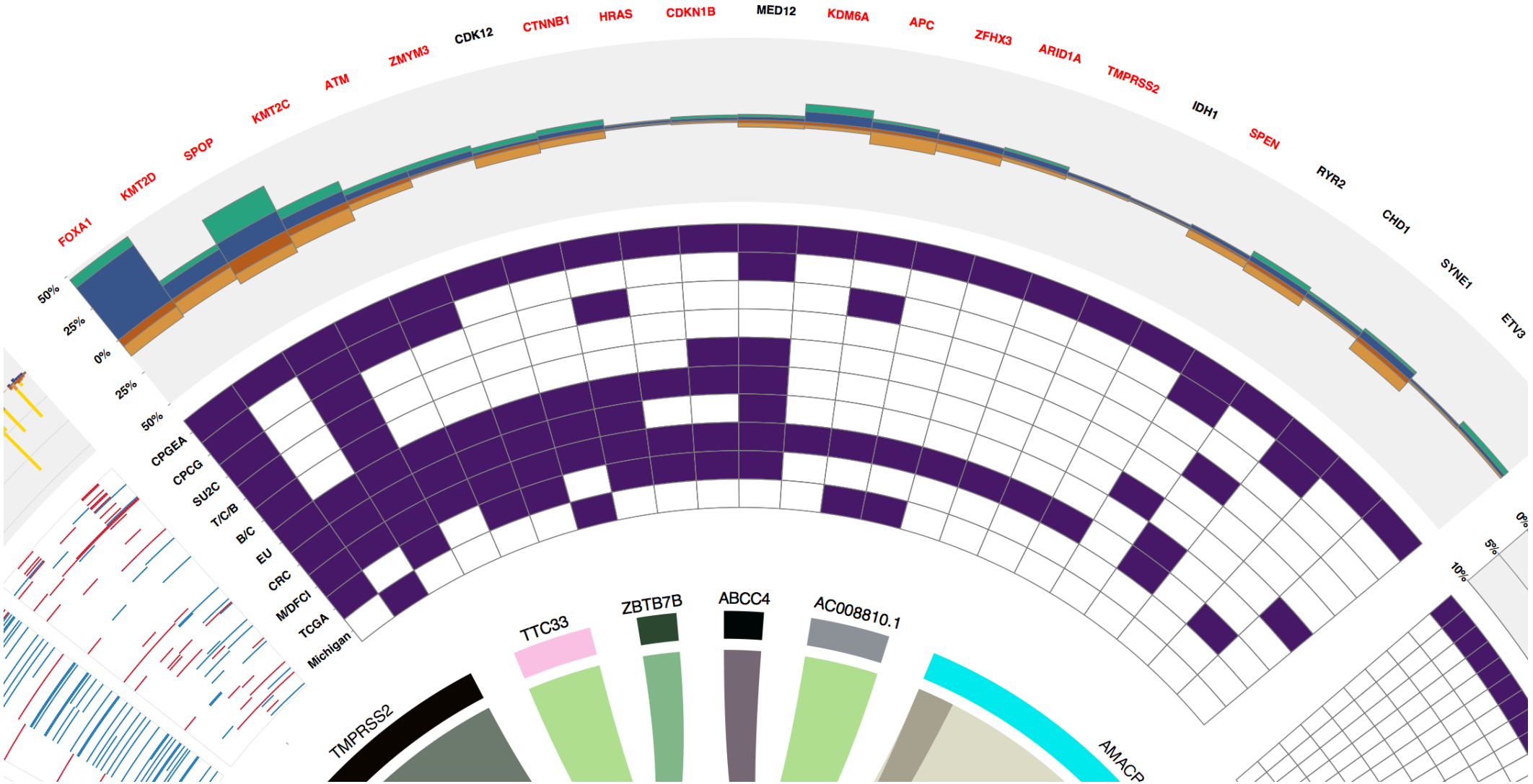
Non-coding

DNA methylation

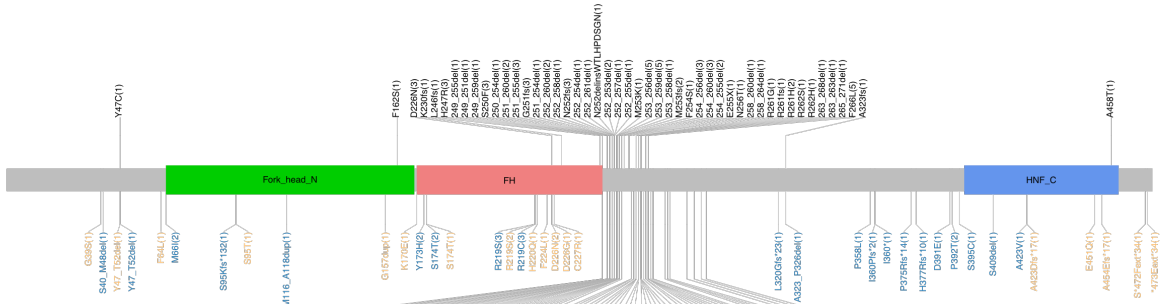
To detect significantly mutated genes (SMGs), we applied MuSiC[28] [MuSiC: identifying mutational significance in cancer genomes 10.1101/gr.134635.111] and MutSigCV[29] [Mutational heterogeneity in cancer and the search for new cancer-associated genes 10.1038/nature12213], and found 78 SMGs including well-known cancer genes such as SPOP, FOXA1, KDM6A, and ZMYM3 (Supplementary Figure 4A). The striking difference between TMRPSS2-EGR fusion events motivated us to comprehensively compare mutation signatures between our Chinese cohort and those established in previous studies. We collected a total of 13 cohorts, including a pilot cohort of Chinese samples published by our own group[18] [citation], and the remaining was mostly consisted of Caucasians[5-16] [citations] (Supplementary Table 8). We integrated 9 studies with large western cohorts to compile a list of 180 SMGs. We also used 699 consensus cancer genes from and Cancer Gene Census (CGC)[30][A census of human cancer genes 10.1038/nrc1299 citation] as an additional confidence filter. Altogether, we were able to define 11 core prostate cancer genes, 9 important Chinese prostate cancer genes, and 31 important Western prostate cancer genes (Supplementary Table 9). Because the SMGs we compiled from previous studies included metastatic tumors, we carefully curated their metadata and were able to separate primary and metastatic samples. Thus, we were able to calculate mutation frequency of these prostate cancer genes in the Chinese cohort and the Western cohorts. Altogether we determined that 135 genes were significantly differentially mutated in primary prostate cancers in Chinese cohort than in Western cohorts (Figure 2). We also reported differences in genes affected by focal copy number changes and gene fusion between Chinese and Western cohorts (Figure 2).


A
MUSIC
MutSigCV
78
7
FOXA1
SPOP
KDM6A
ZMYM3
DNAL1
C4orf32
MRRF
B
EU
CGC
CPGEA
Western Cohorts
TLR6
47
638
60
67
1
6
31
4
11
2
7
1
2
3
RUNX1T1
PREX2
CSF1R
NCOR2
CAMTA1
MUC4
PTPRK
EBF1
KDM6A
APC
ZFHX3
ARID1A
TMPRSS2
IDH1
SPEN
FOXA1
KMT2D
SPOP
ATM
KMT2C
CDK12
ZMYM3
CTNNB1
HRAS
CDKN1B
MED12
CHD1
SYNE1
ETV3
RYR2
TTN
NKX3-1
TP53
BRAF
AKT1
PTEN
AR
ARID2
BRCA2
CDH1
CNOT3
CSMD3
CUL3
FANCG
FAT3
IL6ST
JAK1
KEAP1
KMT2A
KRAS
MET
MUC16
MYC
NCOR1
PALB2
PIK3CA
PIK3CB
PIK3R1
PMS1
PTPRC
RB1
RNF43
SETD2
SF3B1
U2AF1
XPO1
ZBTB16
C
MSKCC
JCO 2017
Metastatic
43.45%
Primary
54.56%
Neuroendocrine
1.98%
Broad/Cornell
Cell 2013
55
Primary
96.49%
Metastatic
3.51%
MSKCC/DFCI
Nat Genet 2018
Metastatic
32.87%
Primary
67.13%
680
CPC-Gene
Nature 2017
Primary
51.57%
246
Unkown
48.43%
CRC
Nat Med 2016
Metastatic
87.5%
154
MSKCCC
Cell 2014
Metastatic
100%
12
Michigan
Nature 2012
Metastatic
11 81.97%
Primary
18.03%
MSKCC
PNAS 2014
Primary
100%
104
TCGA
Cell 2015
Primary
100%
333
MSKCC
Cancer Cell 2010
Primary
67.29%
181
Metastatic
13.75%
Cell Line
2.97%
Xenograft
2.97%
Unkown
13.01%
Broad/Cornell
Nature Genet 2012
Primary
94.64%
106
Unkown
5.36%
Metastatic
61.4%
70
Neuroendocrine
38.6%
T/C/B
Cell 2016
SU2C
Cell 2015
150
Metastatic
100%


D
Mutations per MB
Translational Effect
Synonymous
Non Synonymous
Recurrent SMGs in our & 9 publish studies
Novel SMGs
Recurrent SMGs in Others
Our_Study
EU
Primary
Metastatic
% Mutant
Sample (n=208)
MutSig
-log10(p-value)
Mutation Type
Nonsense Mutation
Frame Shift Ins
Frame Shift Del
In Frame Ins
In Frame Del
Nonstop Mutation
Splice Site
Missense Mutation
Silent

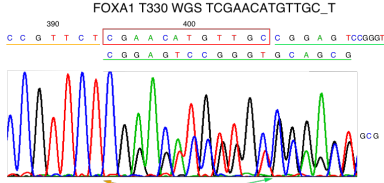

FOXA1
KMT2D
SPOP
KMT2C
ATM
ZMYM3
CDK12
CTNNB1
HRAS
CDKN1B
MED12
KDM6A
APC
ZFHX3
ARID1A
TMPRSS2
IDH1
SPEN
RYR2
CHD1
SYNE1
ETV3
50%
25%
0%
25%
50%
CPGEA
CPCG
SU2C
T/C/B
B/C
EU
CRC
M/DFCI
TCGA
Michigan
0%
5%
10%
TMPRSS2
TTC33
ZBTB7B
ABCC4
AC008810.1
AMACR

CPGEA

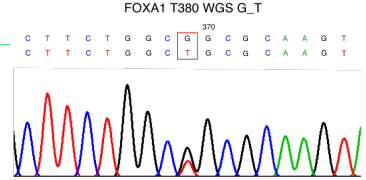


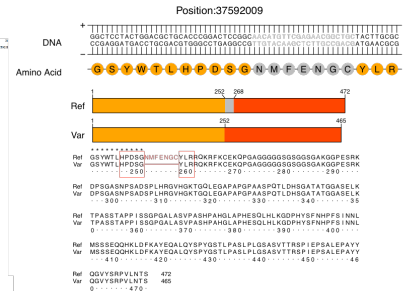
Western Cohort

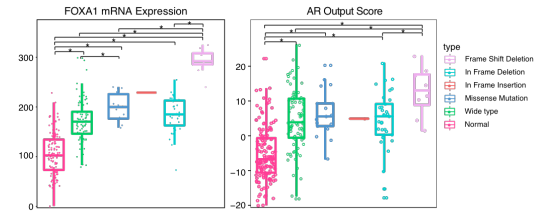
FOXA1 T330 WGS TCGAACATGTTGC_T

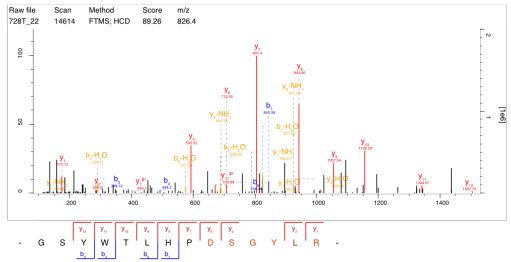


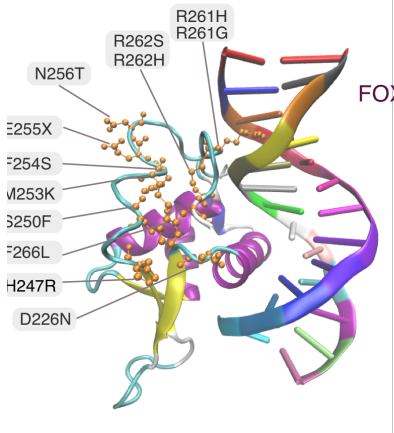
FOXA1 T380 WGS G_T





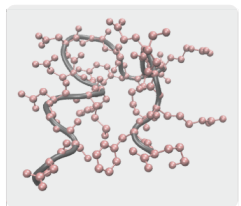


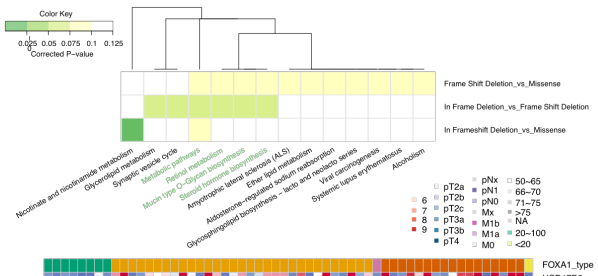


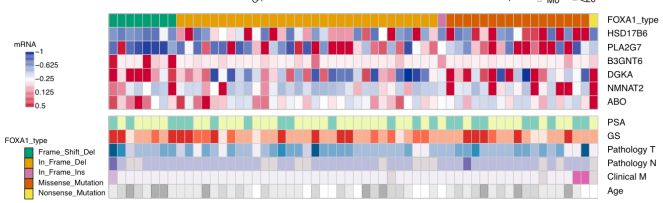


FOXA1

In Frame Deletion
Residue 249-267aa







Landscape

CNV
- CNV arm level
- CNV gene level
- CNV comparison

Fusion
- Fusion
- Fusion comparison

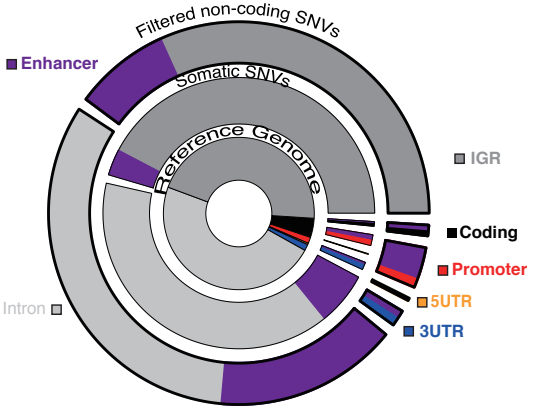
SNV
- SMG
- SMG comparison

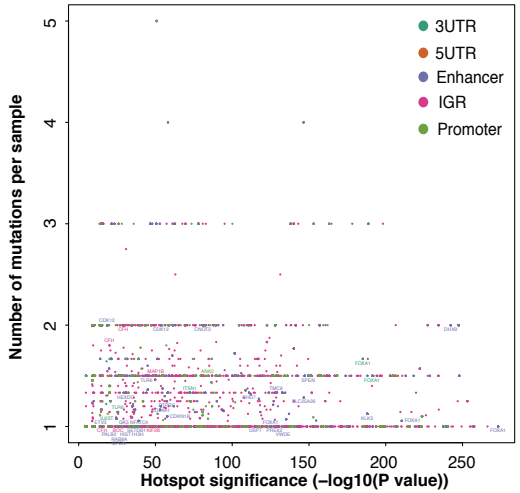
Cluster

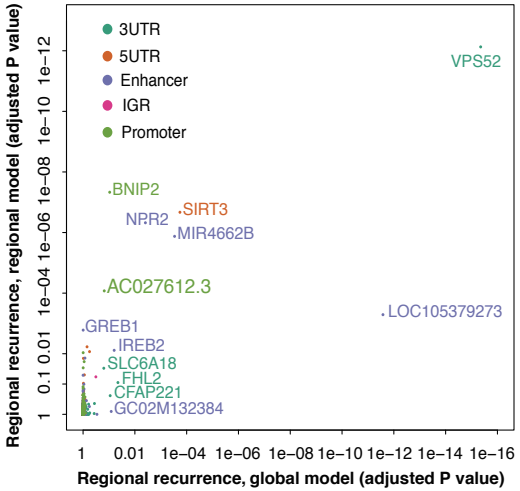
Non-coding

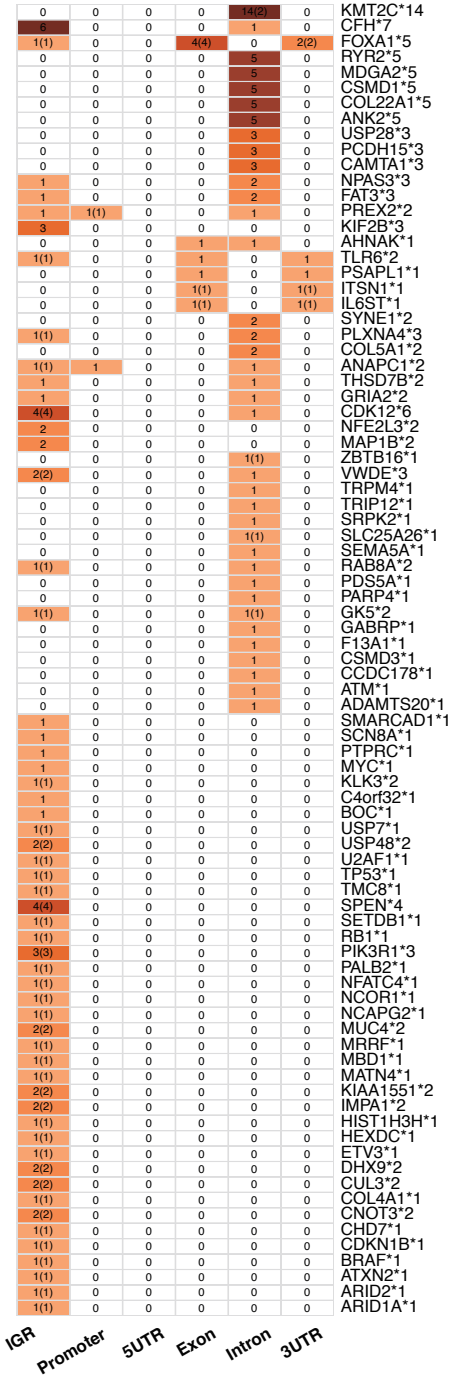
DNA methylation

Recurrently mutated non-coding regulatory elements in prostate cancer









Landscape

CNV
- CNV arm level
- CNV gene level
- CNV comparison

Fusion
- Fusion
- Fusion comparison

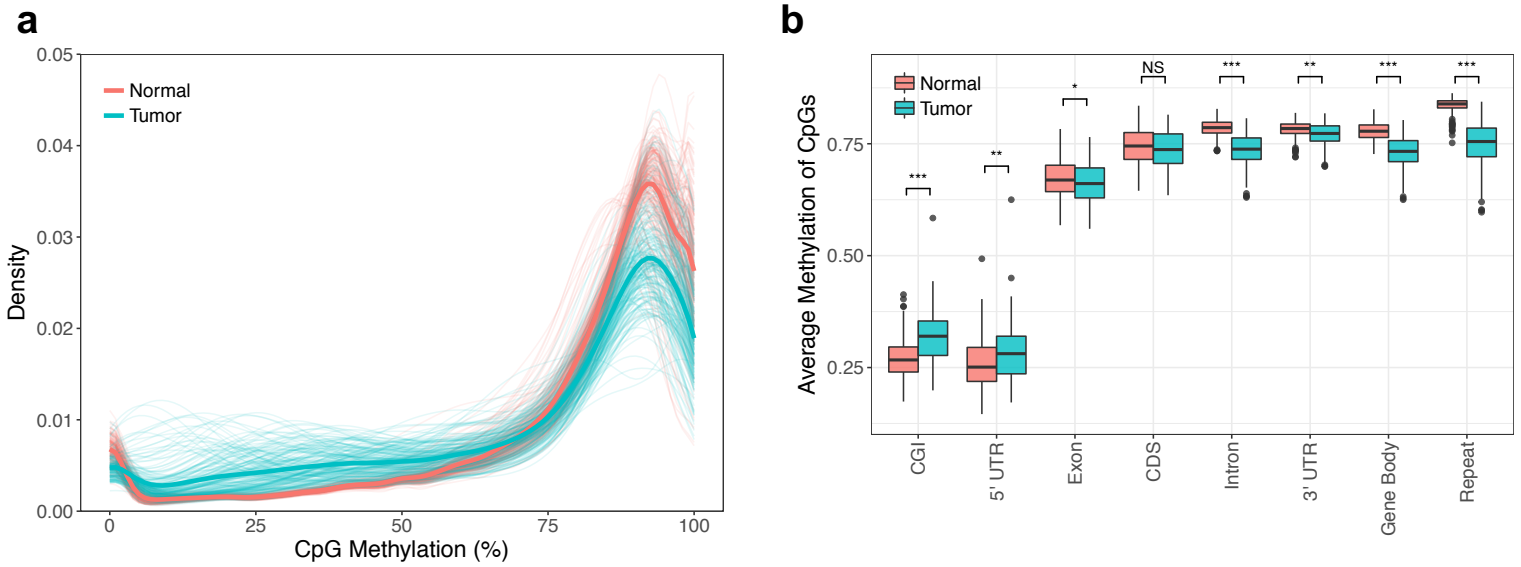
SNV
- SMG
- SMG comparison

Cluster

Non-coding

DNA methylation

Fig5



Compared to normal prostate tissues, prostate tumor DNA methylomes were significantly hypomethylated (mean methylation levels of normal: 78.3%; tumor: 68.8%; p-value = $1.3x10^{-32}$, Wilcoxon signed-rank test). Global CpG methylation level of normal prostates exhibited classical bimodal distribution, whereas CpG methylation of tumors was strongly shifted towards lower levels (Figure 5a, Supplementary Fig. 6a). Hypomethylation was significant across most genomic features including exons, introns, 3'UTRs, and all families of repetitive elements (Figure 5b, Supplementary Fig. 6b).

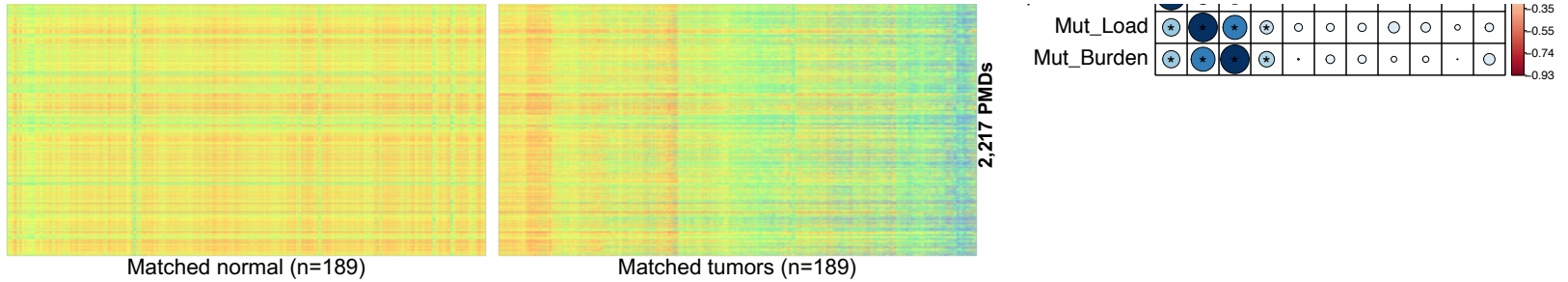
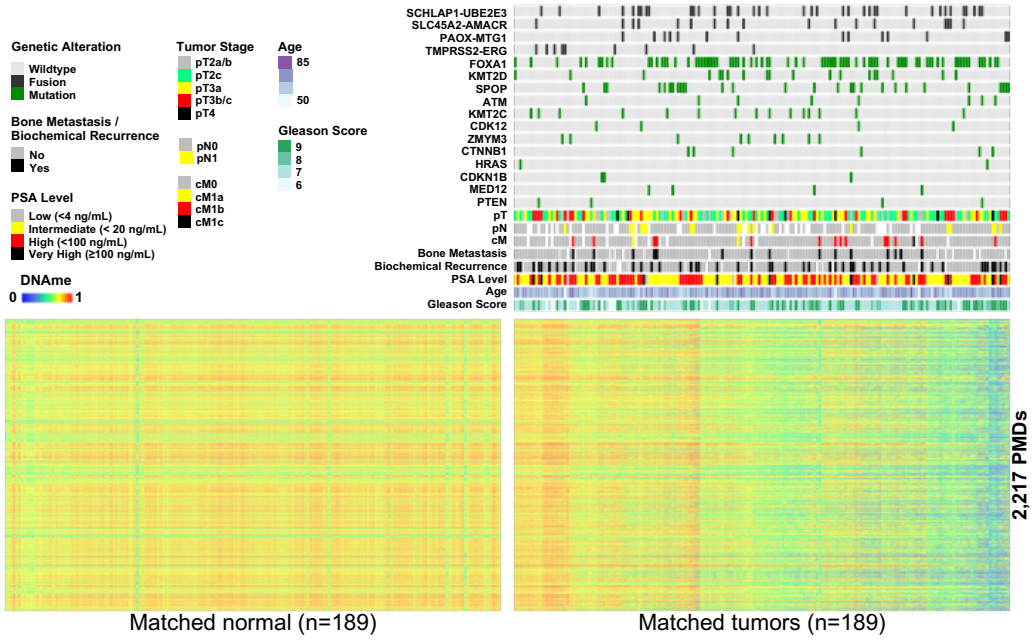


Fig5

a

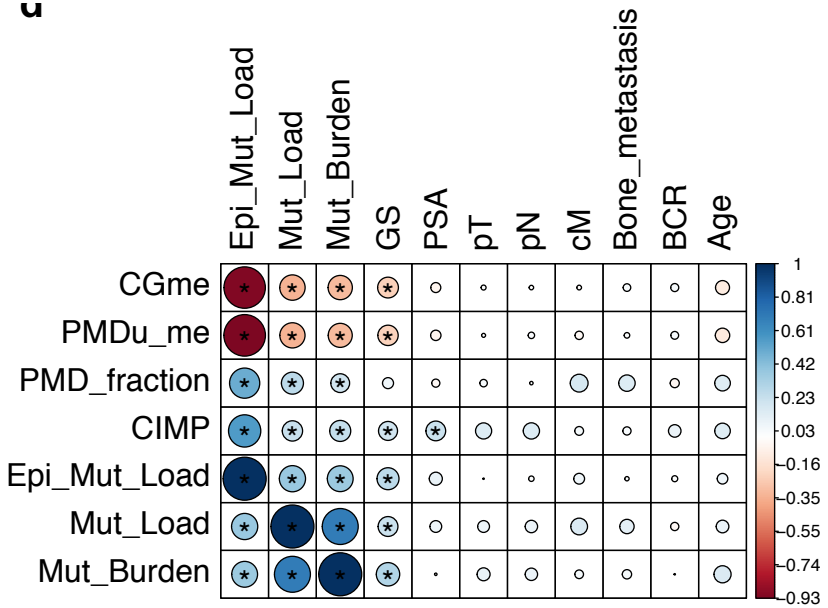
Partially methylated domains (PMD) have previously been characterized in cultured cells and are associated with gene repression and inactive chromatin marks [citations]. Using MethPipe [citations], we defined the pattern of PMD in our samples and found that PMDs were widespread in prostate cancers. These PMDs [description, size 940kb mean, methylation level, etc]. Strikingly, tumors with Gleason scores greater than 6 exhibited much lower DNA methylation levels in their PDMs than those with a Gleason score of 6 (Figure 5). [concluding remark]

https://doi.org/10.1038/ng.865 (Irizarry / Feinberg group, 2011)
https://doi.org/10.1101/gr.125872.111 (Ren group, 2012)
https://doi.org/10.1038/ng.969 (Peter Laird group, 2012)



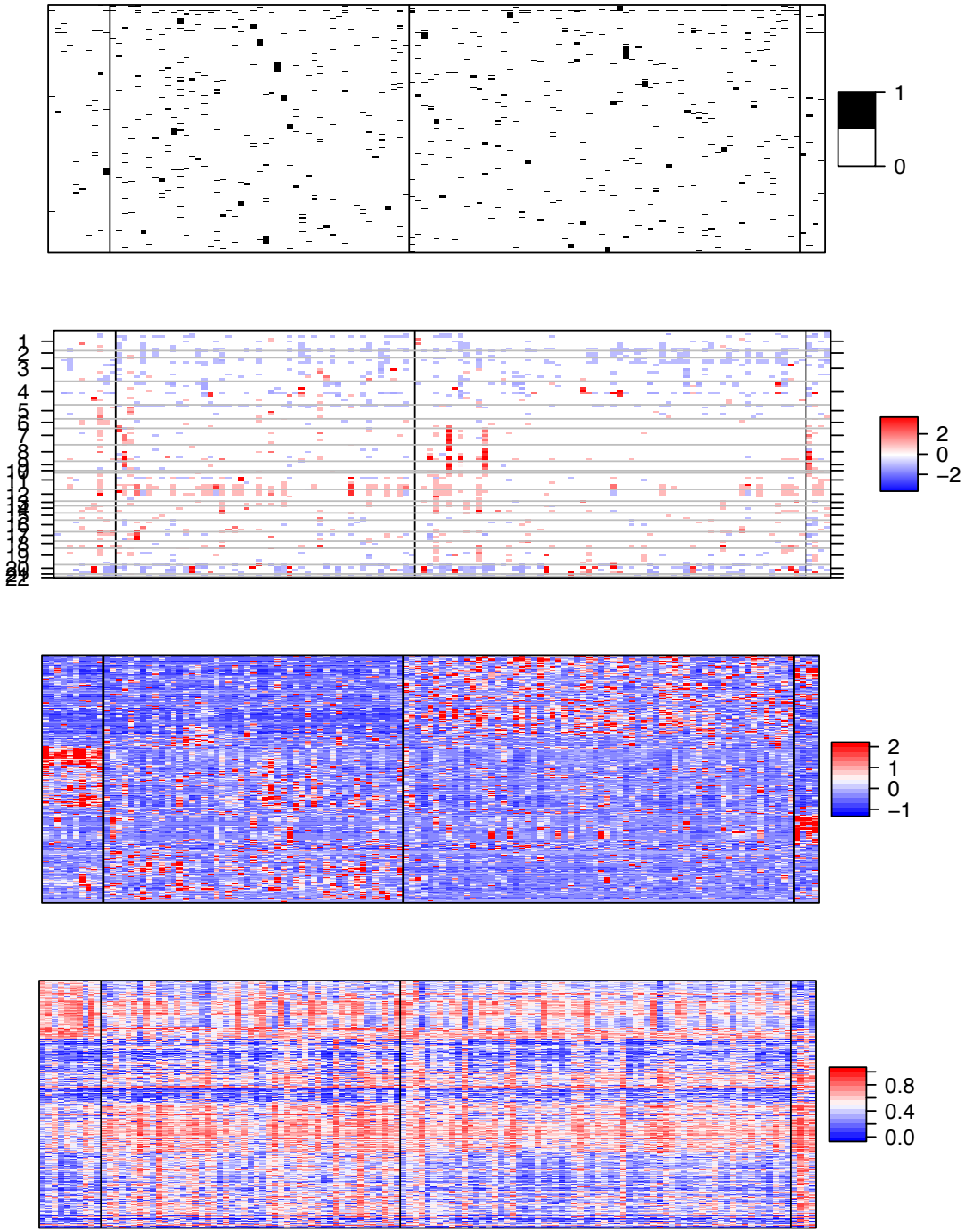
c

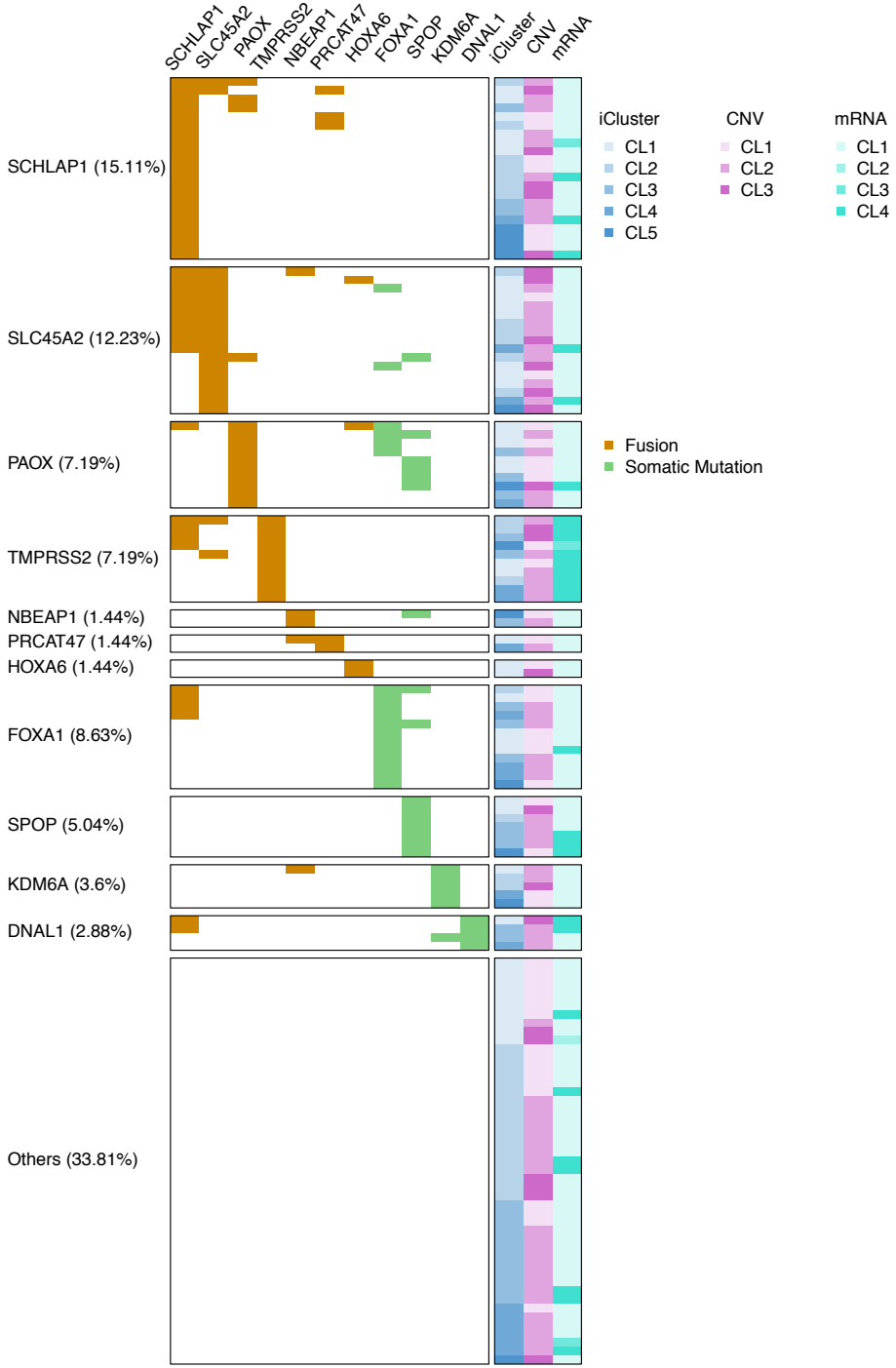
d



Ongoing…

- Icluster （CNV problem)
- pathway


1
0
2
0
−2
2
1
0
−1
0.8
0.4
0.0


SCHLAP1
SLC45A2
PAOX
TMPRSS2
NBEAP1
PRCAT47
HOXA6
FOXA1
SPOP
KDM6A
DNAL1
iCluster
CNV
mRNA
SCHLAP1 (15.11%)
SLC45A2 (12.23%)
PAOX (7.19%)
TMPRSS2 (7.19%)
NBEAP1 (1.44%)
PRCAT47 (1.44%)
HOXA6 (1.44%)
FOXA1 (8.63%)
SPOP (5.04%)
KDM6A (3.6%)
DNAL1 (2.88%)
Others (33.81%)
iCluster
CL1
CL2
CL3
CL4
CL5
CNV
CL1
CL2
CL3
mRNA
CL1
CL2
CL3
CL4
Fusion
Somatic Mutation

Pathway….