Supplemental Figures for Shah et al: The clonal and mutational evolution spectrum of primary triple negative breast cancers

Figure 1: Distribution of sequence coverage of targeted exons for 54 tumour and normal exome capture libraries. Median of median coverage over targeted exons over all libraries was 60.5x.

{ sfig:exome˙coverage˙boxp

Figure 2: Distribution of the number of mutations per sample as it relates to cellularity according to pathologist review

{ sfig:mutations˙cellularity }

Figure 3: Somatic copy number aberration (CNA) landscape of triple negative breast cancers. **(a)** The somatic CNA landscape of Basal (top) and Other (bottom) triple negative cancers showing significant segmental CNA instability across the genome. The frequency alteration of each gene by Ensembl54 annotation is shown where deletions (blue) are reflected as negative frequencies and gains (red) are shown as positive frequencies. **(b)** Comparison of genomic instability between Basal and non-Basal cancers for each of the alteration states. Comparisons are shown as boxplots of the distribution of proportion of the genome altered for each CNA state over the population of 80 tumours that were classifiable by PAM50. All states except HLAMPs showed a statistically higher proportion of alteration in Basals than Others. **(c)**Proportion of homozygous somatic SNV mutations overlapping segmental CNAs, segregated by state. **(d)** RNASeq derived gene fusions supported by segmental CNA breakpoints in Basal (left) and Other (right) cancers depicted as circos diagrams. Chromosomes are arrayed around the circle and arcs between points on the circle represent gene fusions predicted by defuse [1].

{ sfig:CNA˙landscape }

Figure 4: The somatic CNA landscape of Basal (top) and Other (bottom) triple negative cancers from an external cohort of 122 TNBC tumours. The frequency alteration of each gene by Ensembl54 annotation is shown where deletions (blue) are reflected as negative frequencies and gains (red) are shown as positive frequencies. Patterns of alteration were consistent with the cohort of 104 TNBCs featured in this study.

{ sfig:cna˙freq˙METABRIC }

Figure 5: Predicted intragenic homozygous deletions in PARK2 on chr6q26 inferred from Affymetrix SNP 6.0 arrays. Predictions for 6 cases are shown where vertical grey bars denote the boundaries of PARK2. Orange box on the ideogram (bottom right) indicates the genomic coordinates of the view. Each datapoint corresponds to the log ratio of representing total copy number. Bright green points indicate homozygous deletions inferred by segmentation by HMM-Dosage (Supplemental methods). Dark green points are predicted hemizygous regions and blue points are predicted neutral regions. {sfig:park2˙homd}

Figure 6: Examples of four (multi-page PDF) somatic mutation-associated alternative splicing events indicated on the distribution of canonical and alternative exon junction usage. Specific cases that have a donor/acceptor dinucleotide splice site mutation are indicated as red dots. Distributions are shown for splice site mutation events in TP53, PIK3R1, LAMA2 and LAMB1. In all examples, cases harbouring a mutation in the splice sites show outlying alternative junction usage. {sfig:splice˙site˙examples}

Figure 7: List of deleterious variations occurring within RB predicted for the triple negative breast cancer set. The coordinates of the variations and of functional TFBS predicted with the reference sequence (without variation) are provided in the hg18 assembly. The most proximal gene is based on Ensembl annotation (release 63). In the two columns with sequences, the red nucleotides indicate a variation position. The last column highlights the variant position within the TFBS logo (red rectangle). The data are ranked by the ratio of the relative scores. The lower this value, the greater the predicted alteration to the recognition of the sequence by the TF. The rows with bold text reflect the most damaging impact predicted (ratio of relative score lower than 0.9 - see Supplemental methods). {sfig:tfbs˙mutation}

Figure 8: Clustering of extreme CNA event (HOMD, HLAMP) into gene modules and functional pathways by Reactome analysis. Nodes in the network are genes affected by a high level amplification or homozygous deletion. Size of the node is proportional to the frequency of abberation in the population of 104 tumours. Edges between nodes represent protein interactions as described in the Reactome FI database [2] {sfig:CNA˙modules˙pathwa

Figure 9: Ab-initio clustering of the mutational profiles of cases and pathways. Pathway-based features were calculated based on the frequency of mutation in each pathway for a given case. Conservatively, we included 48 pathways showing enrichment from the entire network for further analysis (Table S14) containing 40 somatically mutated genes. For each tumour, we counted the number of times a gene in a given pathway was found to be mutated, producing a tumour-pathway matrix. Hierarchical clustering reveals a population structure where distinct groupings of cases by their pathway-based mutational profiles are evident. Annotations of the number of mutations in each case (scaled in blue), the PAM50 subtype and TP53, Rb1 and PTEN mutation status are shown across the top of the heatmap. Rows in the matrix correspond to pathways shown in Figure 3b. The relationship between genes and pathways in this analysis is such that a small number of highly connected gene nodes (drivers) will implicate many pathways, however the strength of pathway association is also modified by the presence of additional mutated genes, shown as a grayscale for each tumour-pathway data point. {sfig:networkclustering}

Figure 10: Probabilistic graphical model of a Dirichlet process for clonal frequency estimation. Left the random variables in the model are shown as nodes in a graph where shade nodes are observed quantities and unshaded nodes are unobserved quantities we wish to estimate. The model inputs the total number of reference and variant reads at the mutation of interest, the observed copy number at the genomic position of each mutation i and then infers the genotype of the mutation (G_i), the clonal frequency (ϕ_i) and the number of variant reads belonging to each clone (A^i_{var}, D^i_{var}). The conditional probability distributions for the model are shown (right). {sfig:pyclone}

Figure 11: Clonal frequency analysis for all tumours with 10 more mutations in the TNBC cohort (multipage PDF). This figure shows on alternating pages, the clustering of the co-occurrence matrix for mutations to infer the assignment to clonal clusters and on the subsequent page the full inferred posterior distributions of the clonal frequency estimates as output from 2500 runs of the MCMC sampling of the Dirichlet process model. The mutations in the distribution plots are first ordered by the average clonal frequencies over each group of mutations. The mode of the clonal frequency estimate for each mutation is shown with a grey dot. {sfig:omnibus`evolution}

Figure 12: Distribution of clonal frequencies across all mutations (adjusted for predicted normal contamination). Specific clonal frequencies are shown for TP53 mutations and PIK3CA mutations and reveals that the distribution of TP53 mutation is incompatible with the notion that it is always the founder mutation (and therefore present in all cells). Instead the distribution is skewed in favour of TP53 residing in a dominant proportion of cells, but exhibits a tail suggesting that in some tumours it occurs later in the evolution of the cancer. {sfig:tp53clonality}

References

[1] McPherson, A. *et al.* deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput Biol* **7** (2011).

[2] Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology* **11**, R53 (2010).

Figure S1

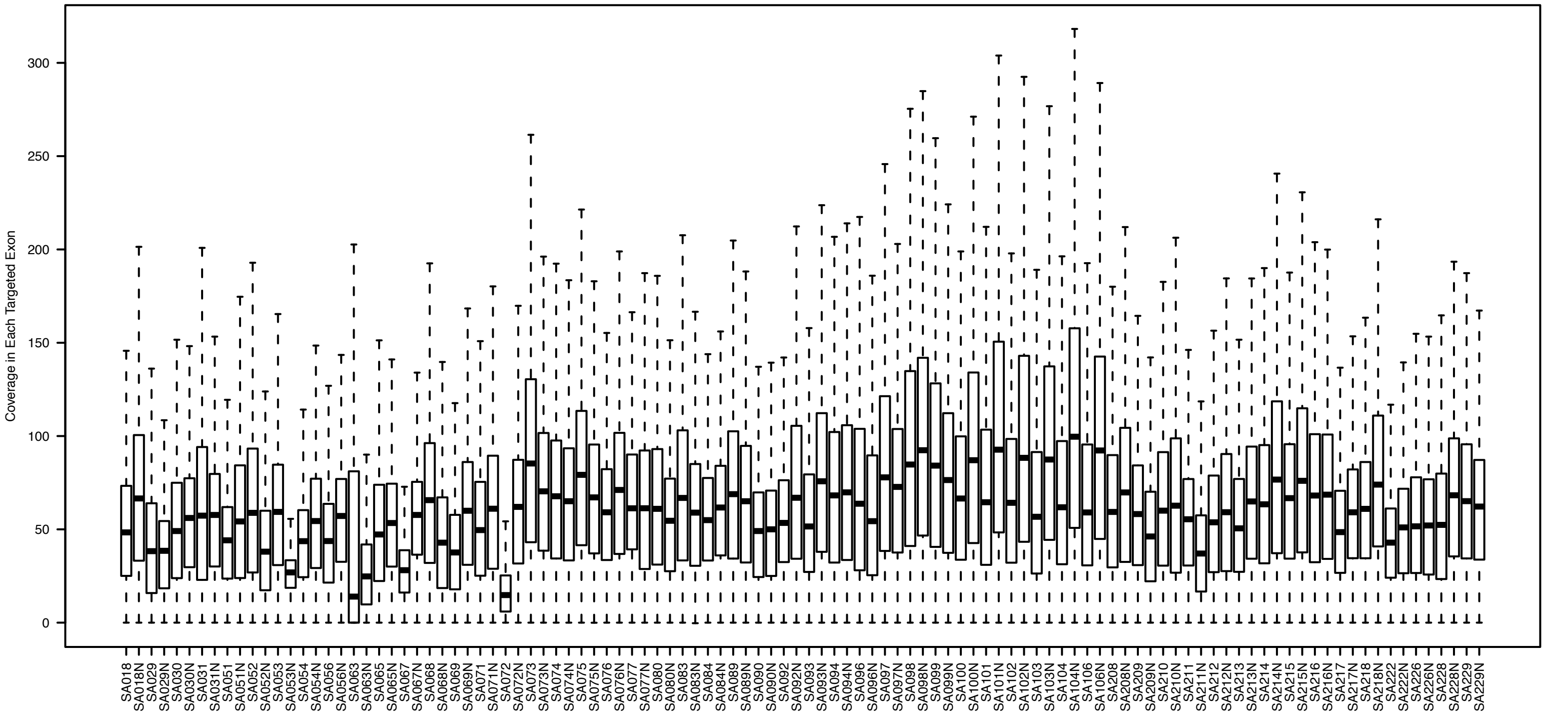

Coverage in Each Targeted Exon
0
50
100
150
200
250
300
SA018
SA018N
SA029
SA029N
SA030
SA030N
SA031
SA031N
SA051
SA051N
SA052
SA052N
SA053
SA053N
SA054
SA054N
SA056
SA056N
SA063
SA063N
SA065
SA065N
SA067
SA067N
SA068
SA068N
SA069
SA069N
SA071
SA071N
SA072
SA072N
SA073
SA073N
SA074
SA074N
SA075
SA075N
SA076
SA076N
SA077
SA077N
SA080
SA080N
SA083
SA083N
SA084
SA084N
SA089
SA089N
SA090
SA090N
SA092
SA092N
SA093
SA093N
SA094
SA094N
SA096
SA096N
SA097
SA097N
SA098
SA098N
SA099
SA099N
SA100
SA100N
SA101
SA101N
SA102
SA102N
SA103
SA103N
SA104
SA104N
SA106
SA106N
SA208
SA208N
SA209
SA209N
SA210
SA210N
SA211
SA211N
SA212
SA212N
SA213
SA213N
SA214
SA214N
SA215
SA215N
SA216
SA216N
SA217
SA217N
SA218
SA218N
SA222
SA222N
SA226
SA226N
SA228
SA228N
SA229
SA229N

Figure S2

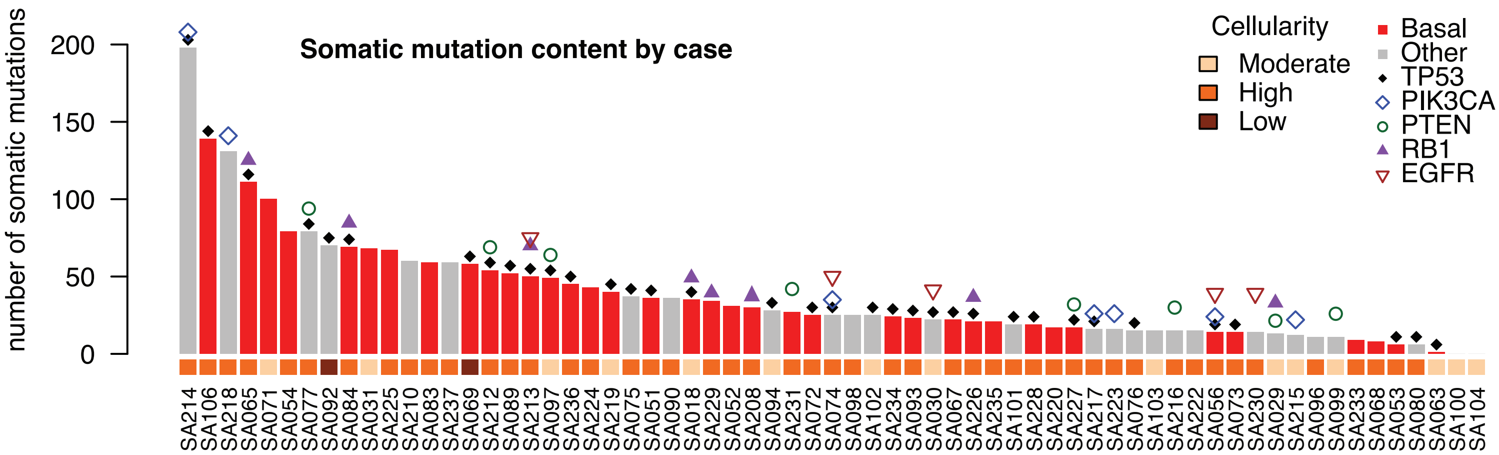
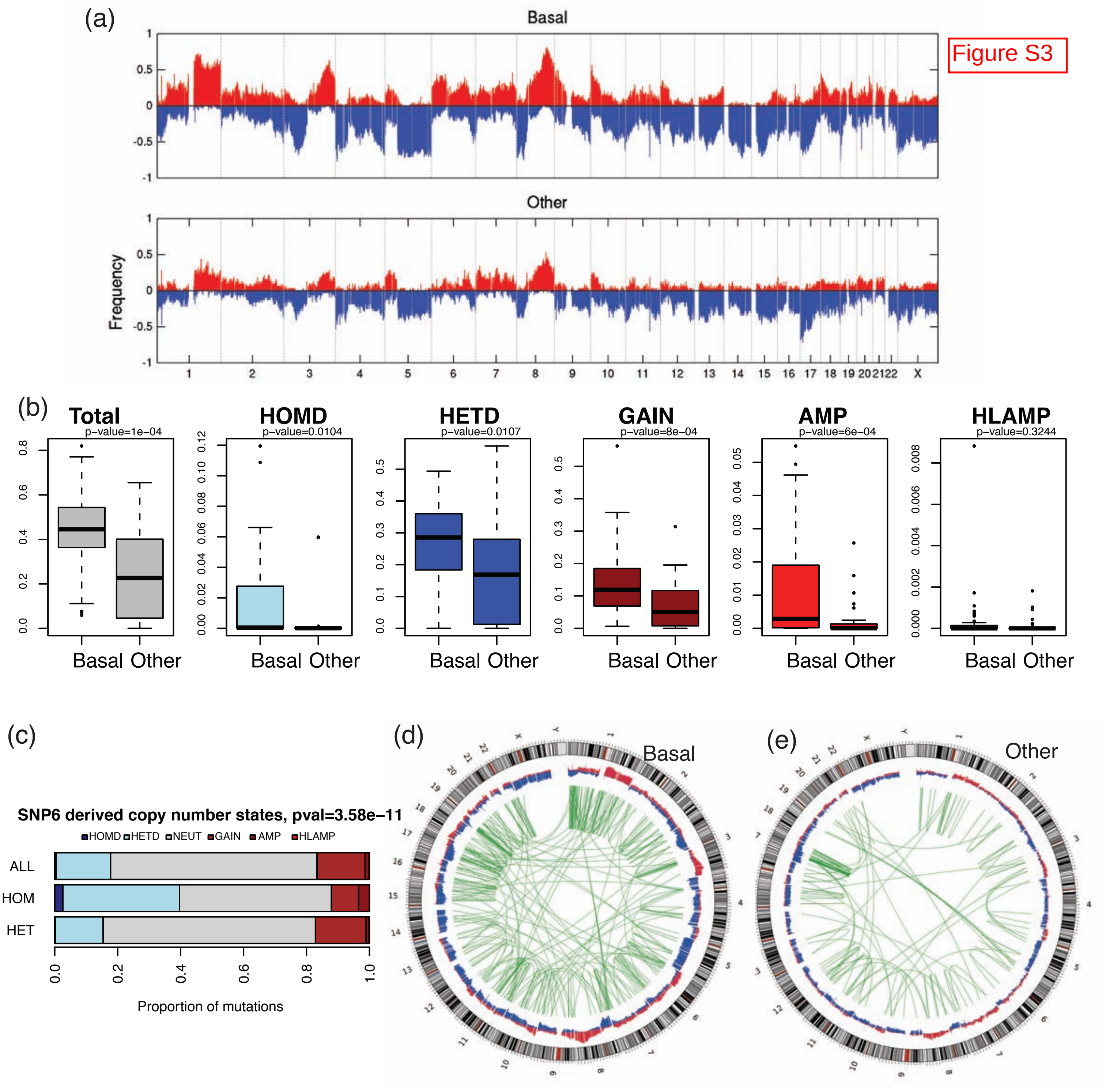

Somatic mutation content by case
number of somatic mutations
0
50
100
150
200
Cellularity
Moderate
High
Low
Basal
Other
TP53
PIK3CA
PTEN
RB1
EGFR
SA214
SA106
SA218
SA065
SA071
SA054
SA077
SA092
SA084
SA031
SA225
SA210
SA083
SA237
SA069
SA212
SA089
SA213
SA097
SA236
SA224
SA219
SA075
SA051
SA090
SA018
SA229
SA052
SA208
SA094
SA231
SA072
SA074
SA098
SA102
SA234
SA093
SA030
SA067
SA226
SA235
SA101
SA228
SA220
SA227
SA217
SA223
SA076
SA103
SA216
SA222
SA056
SA073
SA230
SA029
SA215
SA096
SA099
SA233
SA068
SA053
SA080
SA063
SA100
SA104

Figure S3



(a)
Basal
Other
Frequency
1
0.5
0
-0.5
-1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
(b)
Total
p–value=1e–04
HOMD
p–value=0.0104
HETD
p–value=0.0107
GAIN
p–value=8e–04
AMP
p–value=6e–04
HLAMP
p–value=0.3244
Basal Other
Basal Other
Basal Other
Basal Other
Basal Other
Basal Other
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.1
0.2
0.3
0.4
0.5
0.00
0.01
0.02
0.03
0.04
0.05
0.000
0.002
0.004
0.006
0.008
(c)
SNP6 derived copy number states, pval=3.58e–11
■HOMD ■HETD ■NEUT ■GAIN ■AMP ■HLAMP
ALL
HOM
HET
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of mutations
(d)
Basal
(e)
Other
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

Figure S4

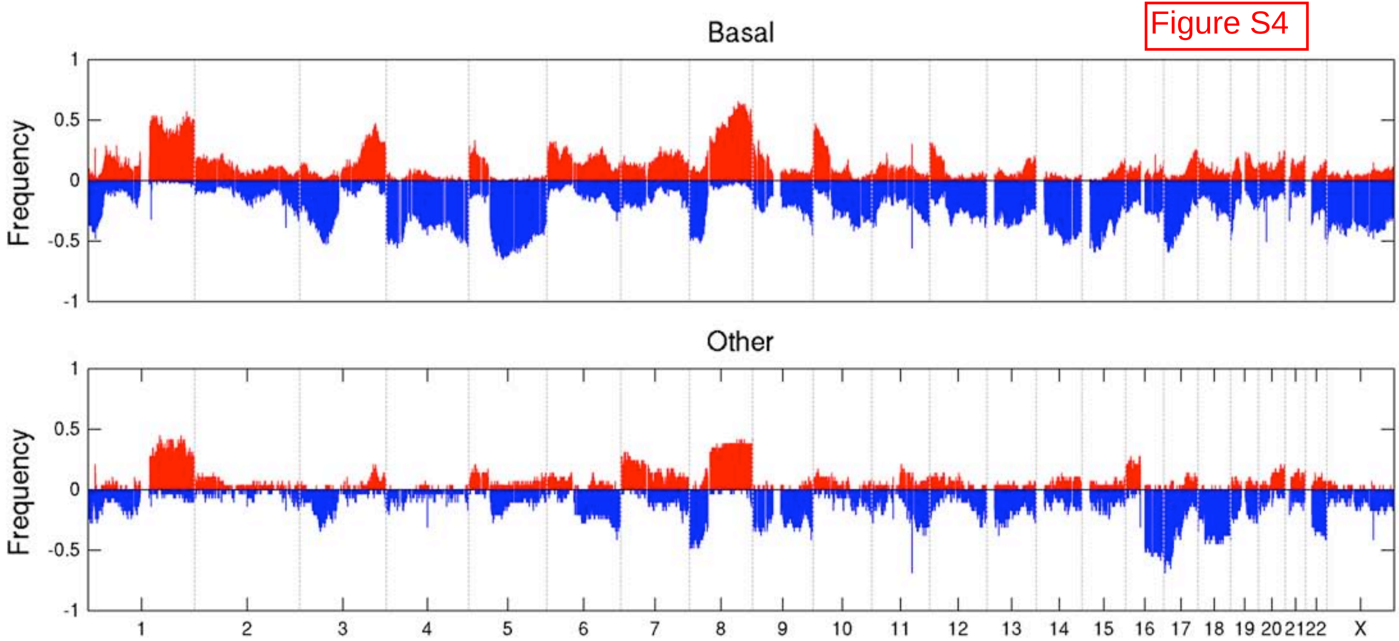

Basal
Frequency
1
0.5
0
-0.5
-1
Other
Frequency
1
0.5
0
-0.5
-1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X

Figure S5

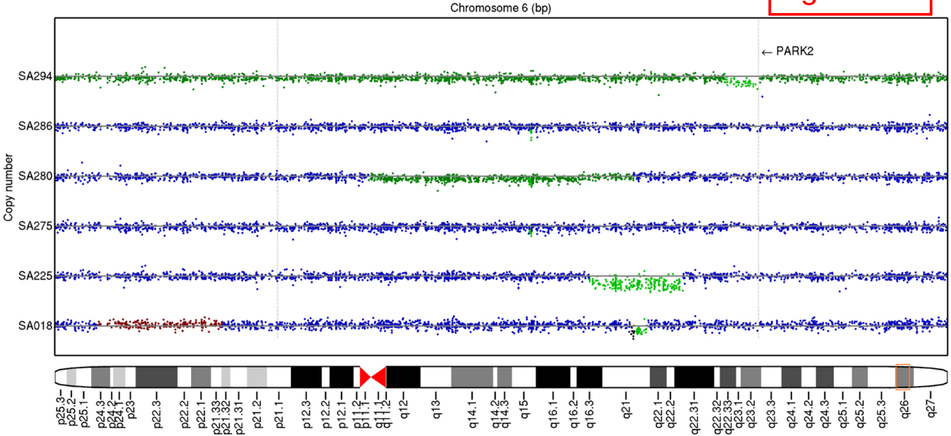
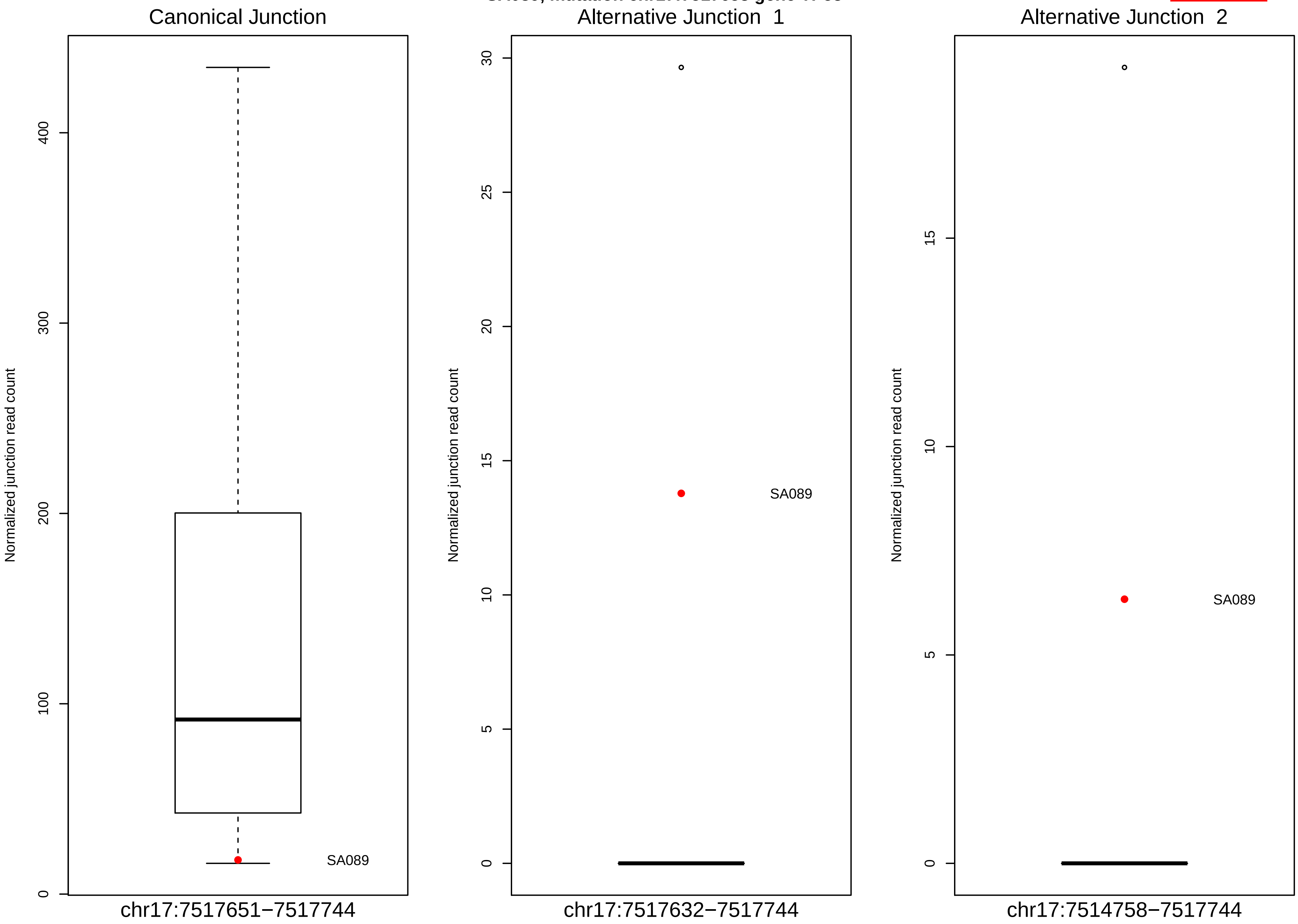
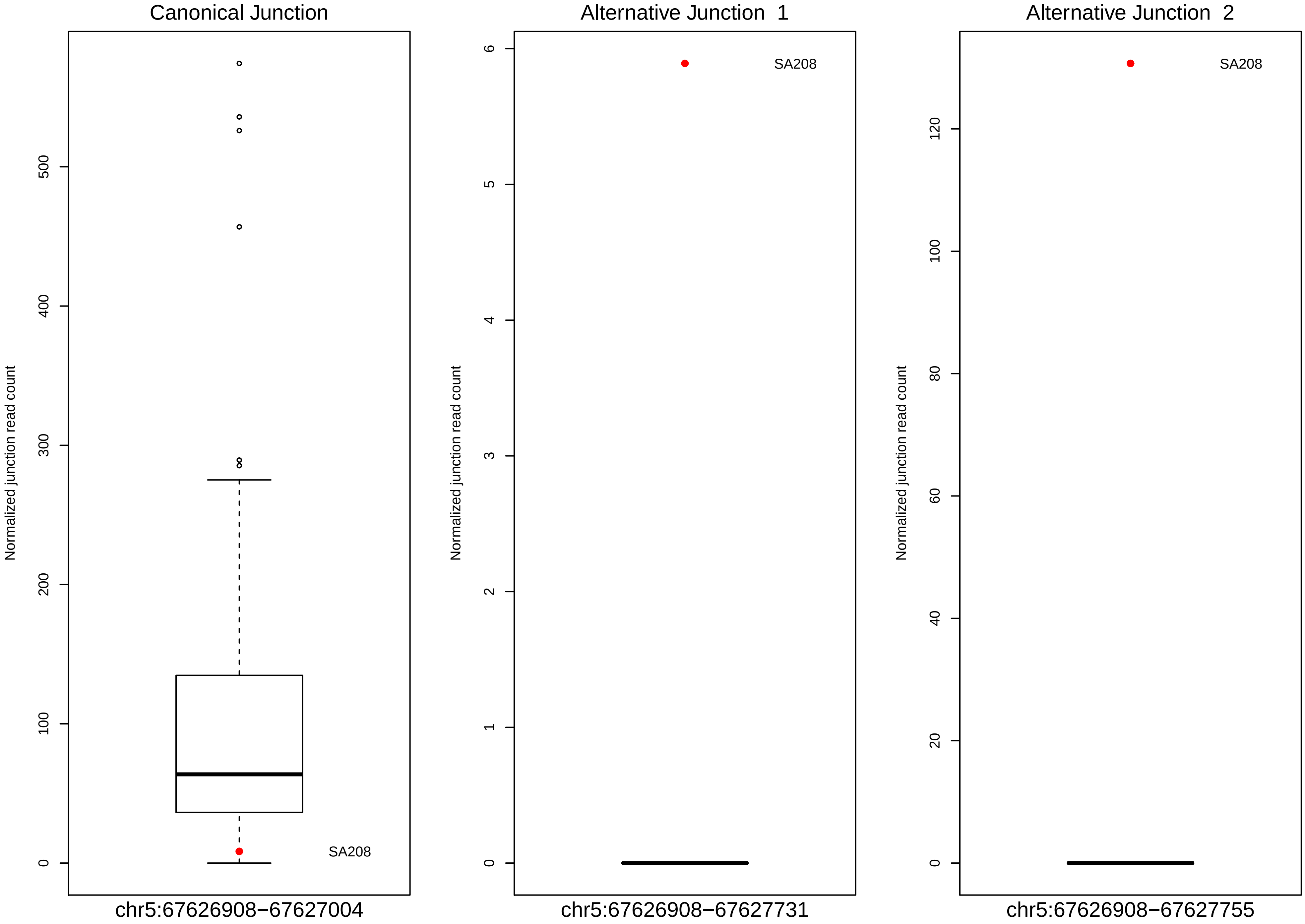

Chromosome 6 (bp)
← PARK2
Copy number
SA294
SA286
SA280
SA275
SA225
SA018
p25.3
p25.2
p25.1
p24.3
p24.2
p24.1
p23
p22.3
p22.2
p22.1
p21.33
p21.32
p21.31
p21.2
p21.1
p12.3
p12.2
p12.1
p11.2
p11.1
q11.1
q11.2
q12
q13
q14.1
q14.2
q14.3
q15
q16.1
q16.2
q16.3
q21
q22.1
q22.2
q22.31
q22.32
q22.33
q23.1
q23.2
q23.3
q24.1
q24.2
q24.3
q25.1
q25.2
q25.3
q26
q27

Figure 6
4 pages

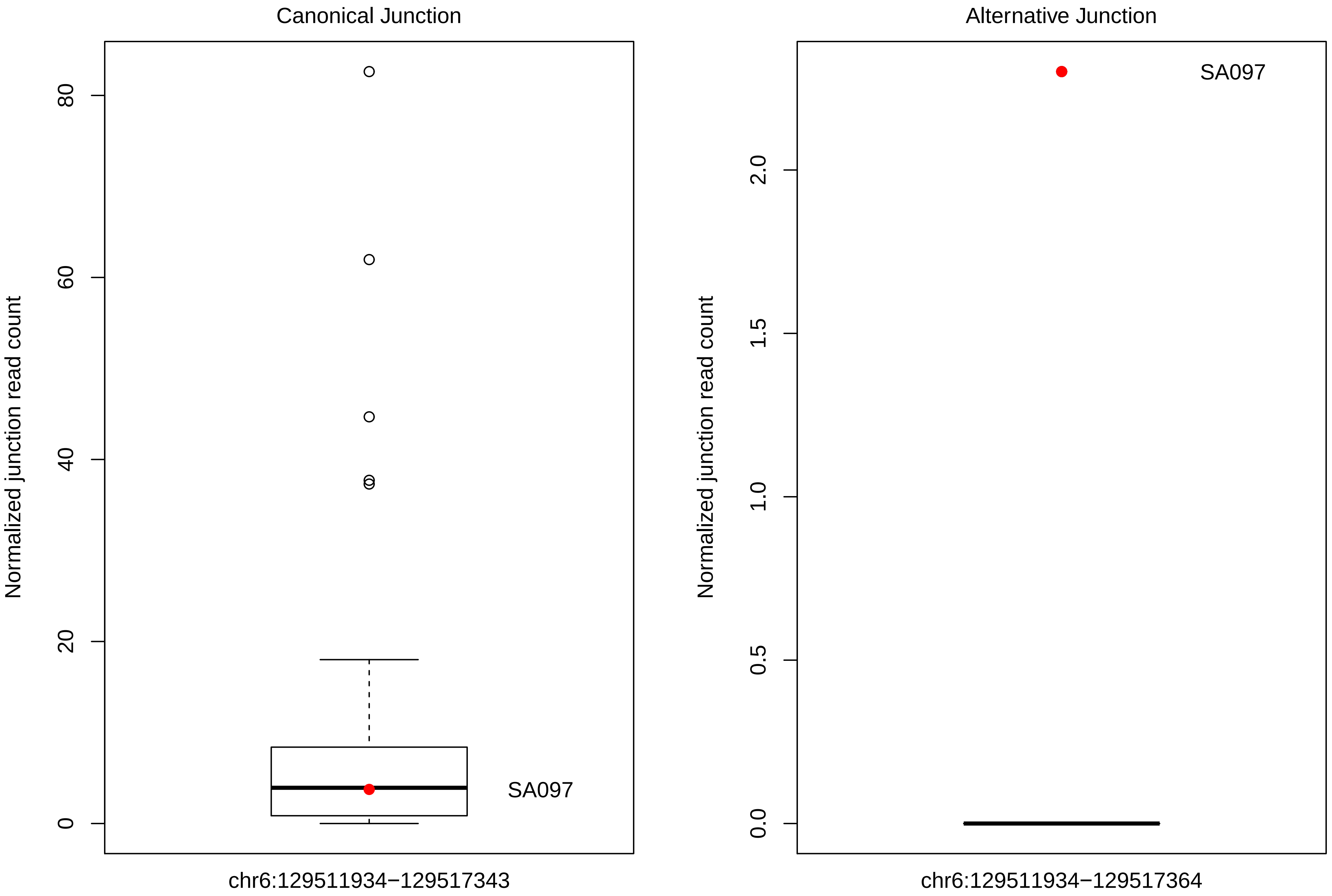
SA089, mutation chr17:7517653 gene TP53



SA208, mutation chr5:67627003 gene PIK3R1



SA097, mutation chr6:129517341 gene LAMA2



SA210, mutation chr7:107356887 gene LAMB1

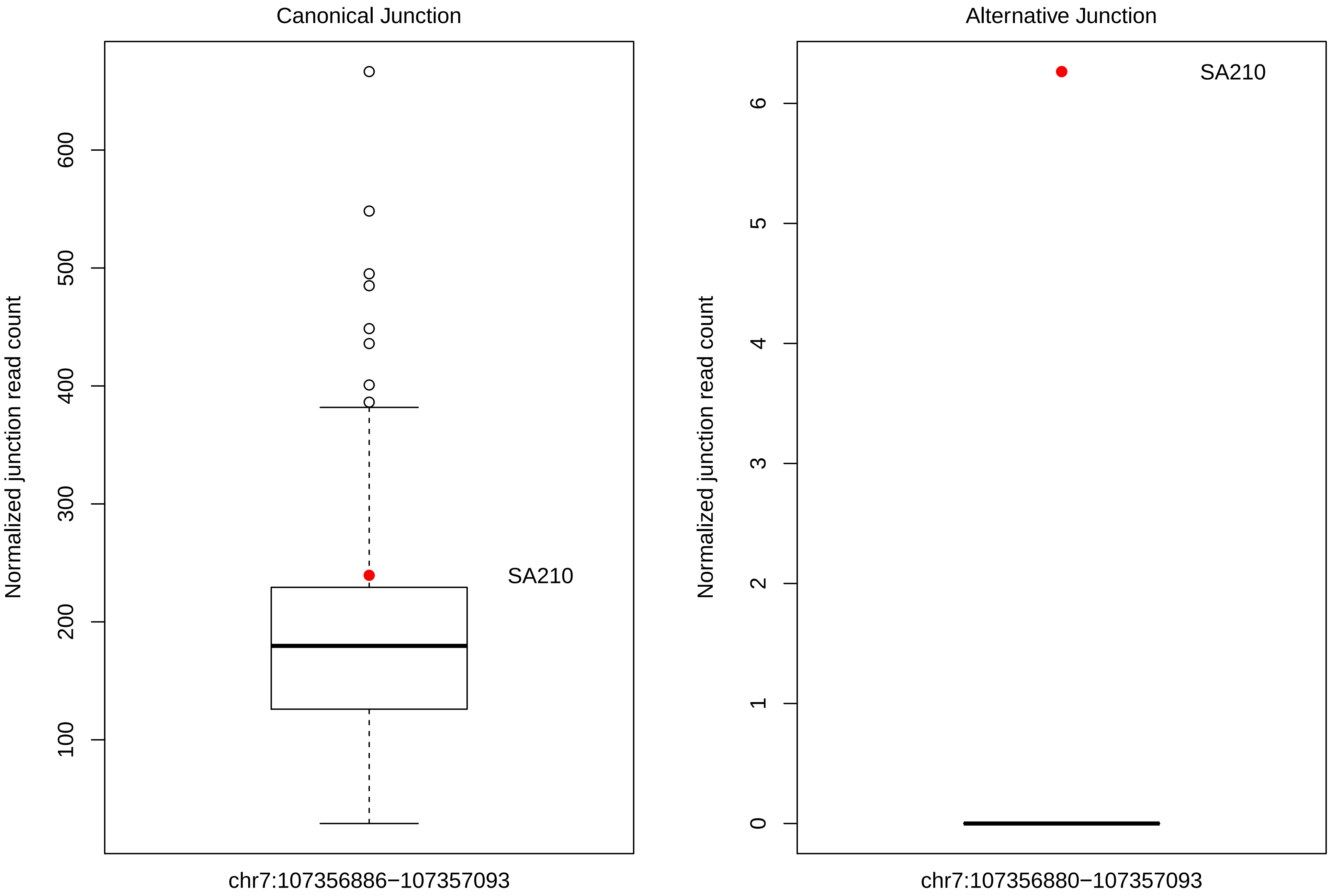


Figure S7

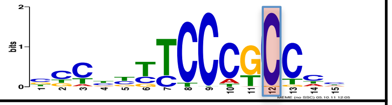
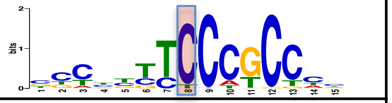
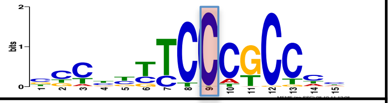
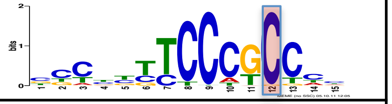
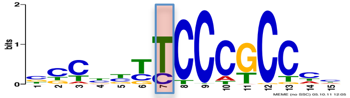
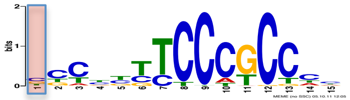
ID	Hg18 chromosome	Hg18 coordinate	TFBS coordinates	Closest gene Ensembl ID and gene name	Reference sequence	Reference relative score	Alternative sequence	Alternative relative score matrix	Ratio of relative score	Location of the variation in the TFBS
SA073	**1**	**215871678**	**215871667 215871681**	**ENSG00000162814 SPATA17**	**CTTTTTTTCTTCGAC**	**80.7**	**CTTTTTTTCTTTGAC**	**70.9**	**0.88**	**C>T** 
SA236	**1**	**24401404**	**24401397 24401411**	**ENSG00000230023 RP11-10N16.2**	**TTCTTGTCCAGGCCC**	**83.0**	**TTCTTGTGCAGGCCC**	**73.2**	**0.88**	**C>G** 
SA065	**21**	**34656187**	**34656178 34656192**	**ENSG00000159197 KCNE2**	**CCCTGTTCCCACCAG**	**88.2**	**CCCTGTTCCGACCAG**	**78.6**	**0.89**	**C>G** 
SA065	**19**	**48793075**	**48793064 48793078**	**ENSG00000226763 SRRM5**	**TGCACGCTCCTCTCC**	**82.1**	**TGCACGCTCCTGTCC**	**73.9**	**0.90**	**C>G** 
SA065	19	15390963	15390957 15390971	ENSG00000246896 AC020911.1	TGAGAGCCCCGCCTT	82.9	TGAGAGGCCCGCCTT	75.2	0.91	C>G 
SA237	5	97673262	97673262 97673276	ENSG00000176183 CH17-12M21.1	GTCTTTTCCTTTCTC	81.1	ATCTTTTCCTTTCTC	79.1	0.98	G>A 

Figure S8

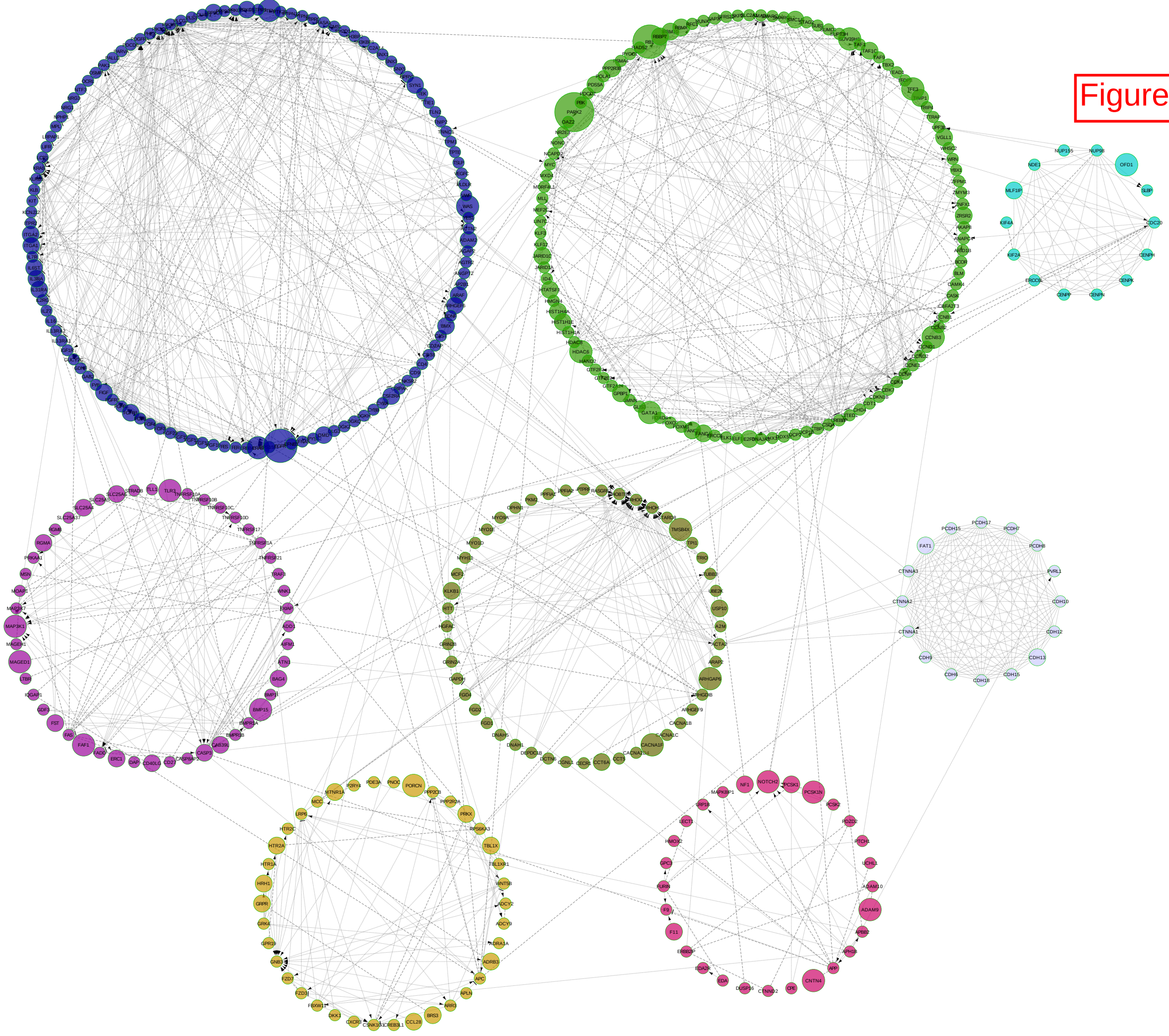

PLS3
PLCG2
PIK3R1
PIK3CB
PHEX
PDGFRB
PDCD6
PARVA
PALLD
PAK1
OSMR
OCRL
NTF3
NRG3
NRG1
NPHP1
MPL
LRPAP1
LIFR
LCP2
KRAS
KLRB1
KLB
KIT
KCNJ12
ITPR2
ITGA2
ITGA1
IL7R
IL6ST
IL3RA
IL31RA
IL2RG
IL27
IL16
IL13RA2
IL13RA1
IGF1R
GUCY2C
GDNF
GAB2
FYB
FIGF
FGFR3
FGFR2
FGFR1
FGF6
FGF4
FGF3
FGF23
FGF10
FES
FER
ERBB2
EGFR
EFNB1
EFNA5
DPYSL2
DMD
DLG3
DGKZ
DGKQ
DGKH
CYBB
CYBA
CSF2RA
CRIPAK
CNKSR2
CD9
CD4
CD38
CD2AP
CAST
BMX
BDNF
ARHGEF6
ARAF
AP2B1
ANGPT2
AGTR2
AGAP2
ADAM2
ACTN2
YES1
WAS
VWF
VLDLR
VEGFC
TSLP
TPTE
TPM1
TNNC1
TNIP2
TLN2
TIE1
TEK
SYN1
SPRY2
SNX5
SNX3
SNX1
SLC2A14
SH3KBP1
SH3BP2
SH2D1A
SCMS1
RASA1
PTPRC
PTPN6
PTPN12
PTK2B
PTEN
PRKG1
RUNX2
SAP30
SFRS15
SKP2
SLC2A4
SMARCC1
SMC1A
STAG2
SUB1
SUMO1
SUPT3H
SUV39H1
TAF1
TAF1C
TAF9
TBX2
TEAD1
TFDP3
TFE3
TIMP1
TRIP4
TTRAP
UPF3B
VGLL1
WHSC2
WRN
YBX1
ZFPM1
ZMYM3
ZNFX1
ZRSR2
AKAP8
ANAPC4
ARID1B
BCOR
BLM
CAMK4
CASK
CBFA2T3
CCNB1
CCNB2
CCNB3
CCND1
CCND2
CCNE1
CCNH
CDK4
CDK7
CDKN1B
CDT1
CHD4
CITED1
CREBBP
CSDA
CTBP1
DCP1B
DCP2
DDX5
DHX15
DNAJA3
E2F3
ELF1
ELK1
ERCC8
FANCA
FANCB
FOXM1
FOXO3
FOXO4
GATA1
GLIS3
GMNN
GPBP1
GTF2A2
GTF2E2
GTF2F2
HAND2
HDAC6
HDAC8
HIST1H1A
HIST1H1E
HIST1H4A
HMGN4
HTATSF1
ID4
JARID1A
JARID1C
KLF12
KLF3
LIN7C
MEF2C
MLL
MORF4L1
MXD4
MYC
NCAPD2
NONO
NR2E3
OAZ2
PARK2
PBK
PDCD7
PDS5A
POLA1
PPP2R3B
PSMA4
PYGO1
RAD52
RB1
RBBP7
RBM10
RBMX
RFC1
NUP155
NUP98
OFD1
NDE1
MLF1IP
SLBP
KIF4A
CDC20
KIF2A
CENPH
ERCC6L
CENPK
CENPP
CENPN
STRADB
TLL1
TLR3
TNFRSF10A
TNFRSF10B
TNFRSF10C
TNFRSF10D
TNFRSF17
TNFRSF1A
TNFRSF21
TRAF3
WNK1
XIAP
ADD1
AIFM1
ATN1
BAG4
BMP1
BMP15
BMPR1A
BMPR1B
CAB39L
CASP3
CASP8AP2
CD27
CD40LG
DAP
ERC1
FADD
FAF1
FAS
FST
GDF3
IQGAP1
LTBR
MAGED1
MAGEH1
MAP3K1
MAP3K7
MOAP1
MSN
PRKAA1
RGMA
RGMB
SLC25A37
SLC25A4
SLC25A5
SLC25A6
PPFIA1
PPFIA2
PTPRF
RASGRF1
RHOBTB2
RHOG
RHOH
STARD8
TMSB4X
TPI1
TRIO
TUBB3
UBE2K
USP10
A2M
ACTA2
ARAP2
ARHGAP6
ARHGDIB
ARHGEF9
CACNA1B
CACNA1C
CACNA1F
CACNA2D4
CCT5
CCT6A
CECR1
CGNL1
DCTN6
DEPDC1B
DNAH1
DNAH5
FGD1
FGD2
FGD4
GAPDH
GRIN2A
GRIN2B
HGFAC
HTT
KLKB1
MCF2
MYH11
MYO1D
MYO1E
MYO9A
OPHN1
PKM2
PCDH15
PCDH17
PCDH7
FAT1
PCDH8
CTNNA3
PVRL1
CTNNA2
CDH10
CTNNA1
CDH12
CDH9
CDH13
CDH6
CDH18
CDH15
P2RY4
PDE3A
PNOC
PORCN
MTNR1A
MCC
PPP2CB
PPP2R2A
LRP6
PRKX
HTR2C
RPS6KA3
HTR2A
TBL1X
HTR1A
TBL1XR1
HRH1
WNT5B
GRPR
ADCY2
GRK4
ADCY9
GPR19
ADRA1A
GNB3
ADRB3
FZD7
APC
FZD3
APLN
FBXW11
ARR3
DKK1
BRS3
CXCR3
CSNK1G1
CREB3L1
CCL28
NF1
NOTCH2
PCSK1
PCSK1N
MAPK8IP1
LRP1B
PCSK2
LECT1
PDZD2
HMOX2
PTCH1
GPC3
UCHL1
FURIN
ADAM10
F9
ADAM9
F11
APBB2
ERBB2IP
APH1B
EDA2R
APP
EDA
CNTN4
DUSP16
CTNND2
CPE

Figure S9

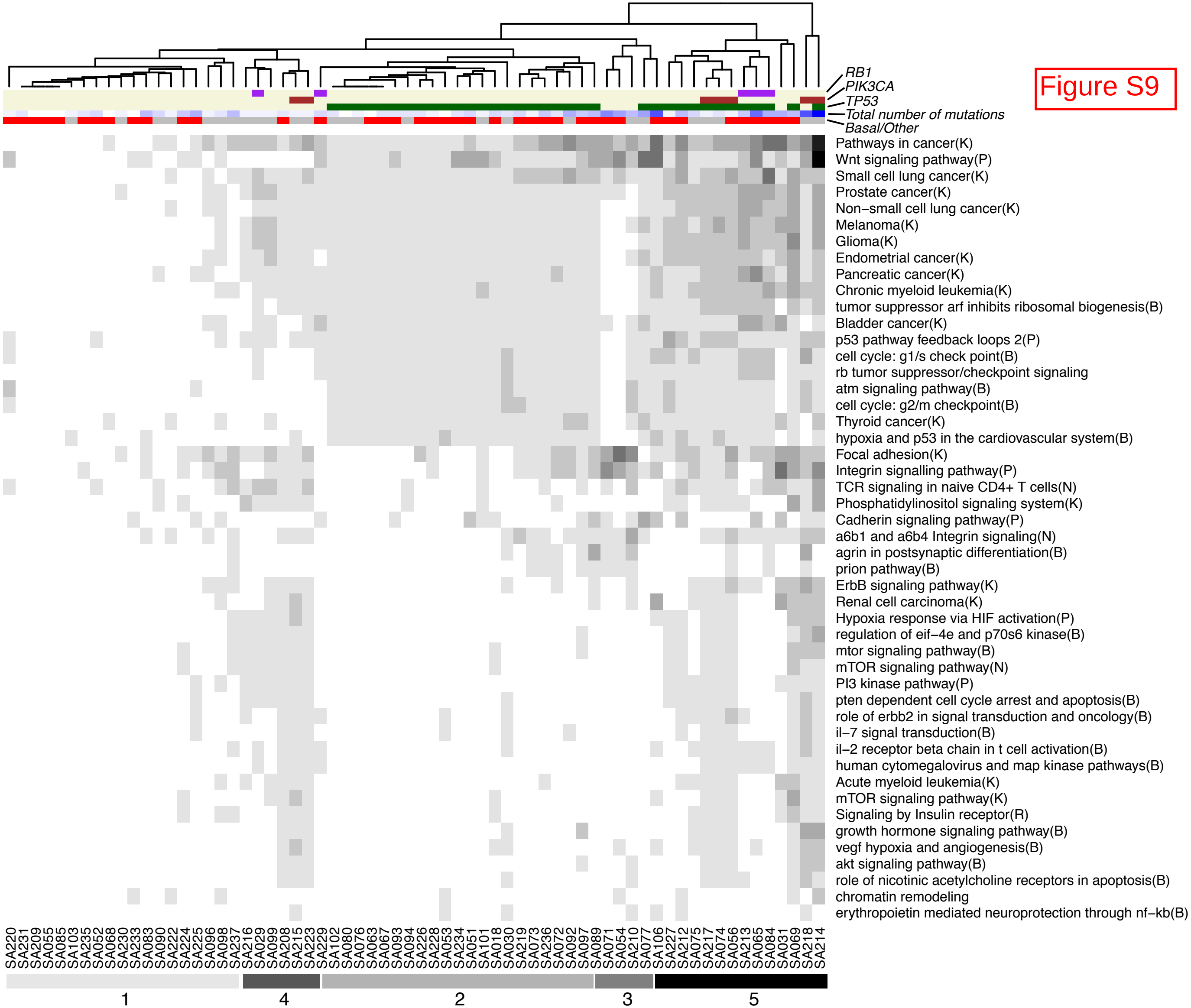
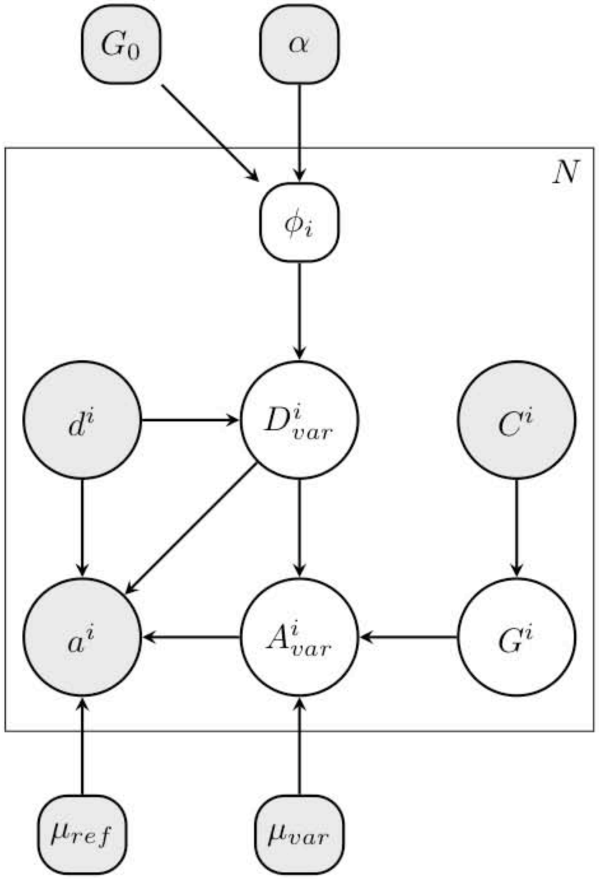
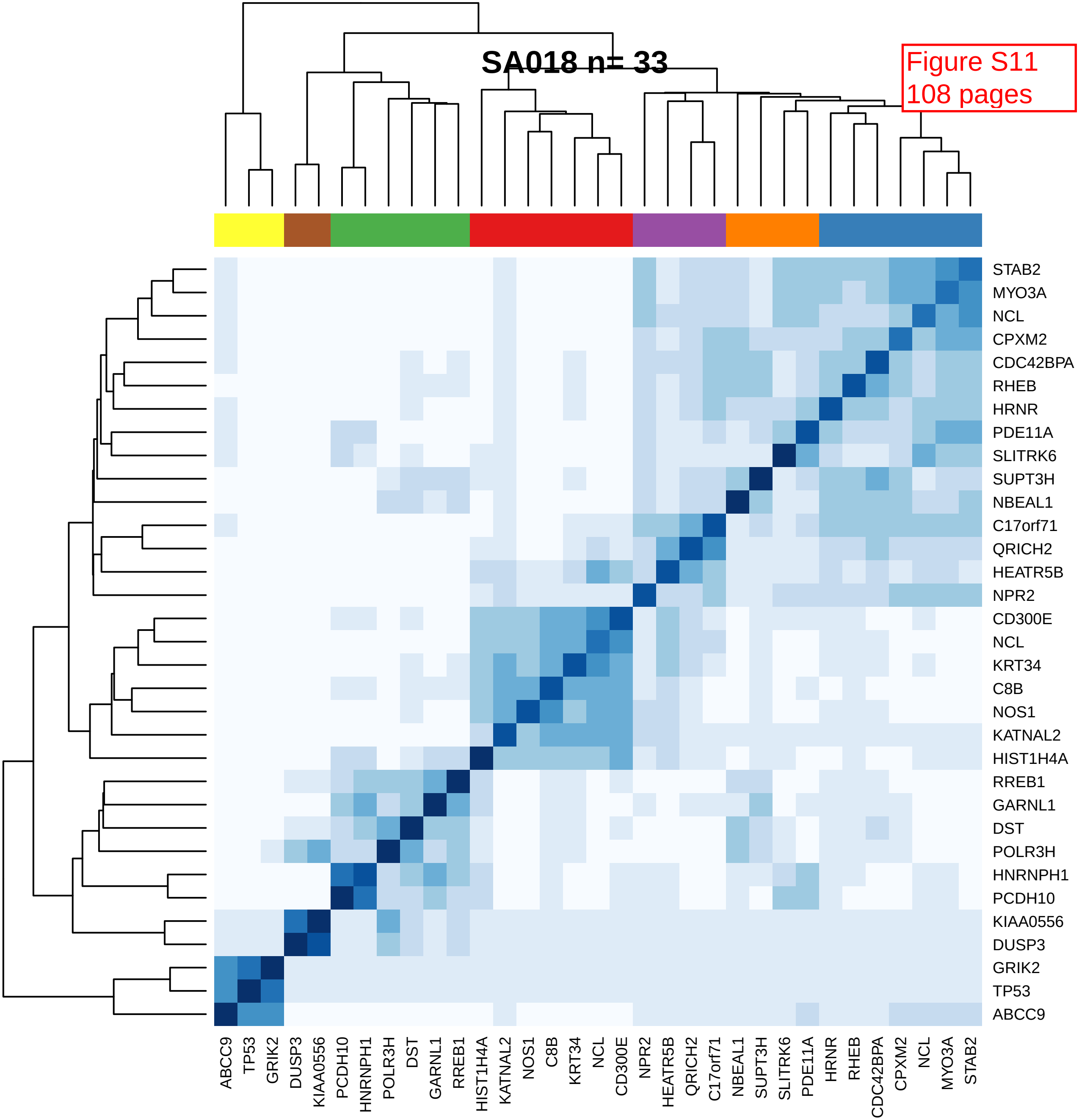

RB1
PIK3CA
TP53
Total number of mutations
Basal/Other
Pathways in cancer(K)
Wnt signaling pathway(P)
Small cell lung cancer(K)
Prostate cancer(K)
Non−small cell lung cancer(K)
Melanoma(K)
Glioma(K)
Endometrial cancer(K)
Pancreatic cancer(K)
Chronic myeloid leukemia(K)
tumor suppressor arf inhibits ribosomal biogenesis(B)
Bladder cancer(K)
p53 pathway feedback loops 2(P)
cell cycle: g1/s check point(B)
rb tumor suppressor/checkpoint signaling
atm signaling pathway(B)
cell cycle: g2/m checkpoint(B)
Thyroid cancer(K)
hypoxia and p53 in the cardiovascular system(B)
Focal adhesion(K)
Integrin signalling pathway(P)
TCR signaling in naive CD4+ T cells(N)
Phosphatidylinositol signaling system(K)
Cadherin signaling pathway(P)
a6b1 and a6b4 Integrin signaling(N)
agrin in postsynaptic differentiation(B)
prion pathway(B)
ErbB signaling pathway(K)
Renal cell carcinoma(K)
Hypoxia response via HIF activation(P)
regulation of eif−4e and p70s6 kinase(B)
mtor signaling pathway(B)
mTOR signaling pathway(N)
PI3 kinase pathway(P)
pten dependent cell cycle arrest and apoptosis(B)
role of erbb2 in signal transduction and oncology(B)
il−7 signal transduction(B)
il−2 receptor beta chain in t cell activation(B)
human cytomegalovirus and map kinase pathways(B)
Acute myeloid leukemia(K)
mTOR signaling pathway(K)
Signaling by Insulin receptor(R)
growth hormone signaling pathway(B)
vegf hypoxia and angiogenesis(B)
akt signaling pathway(B)
role of nicotinic acetylcholine receptors in apoptosis(B)
chromatin remodeling
erythropoietin mediated neuroprotection through nf−kb(B)
SA220
SA231
SA209
SA055
SA085
SA103
SA235
SA052
SA068
SA230
SA233
SA083
SA090
SA222
SA224
SA225
SA096
SA098
SA237
SA216
SA029
SA099
SA208
SA215
SA223
SA229
SA102
SA080
SA076
SA063
SA067
SA093
SA094
SA226
SA228
SA053
SA234
SA051
SA101
SA018
SA030
SA219
SA073
SA236
SA072
SA092
SA097
SA089
SA071
SA054
SA210
SA077
SA106
SA227
SA212
SA075
SA217
SA074
SA056
SA213
SA065
SA084
SA031
SA069
SA218
SA214
1
4
2
3
5

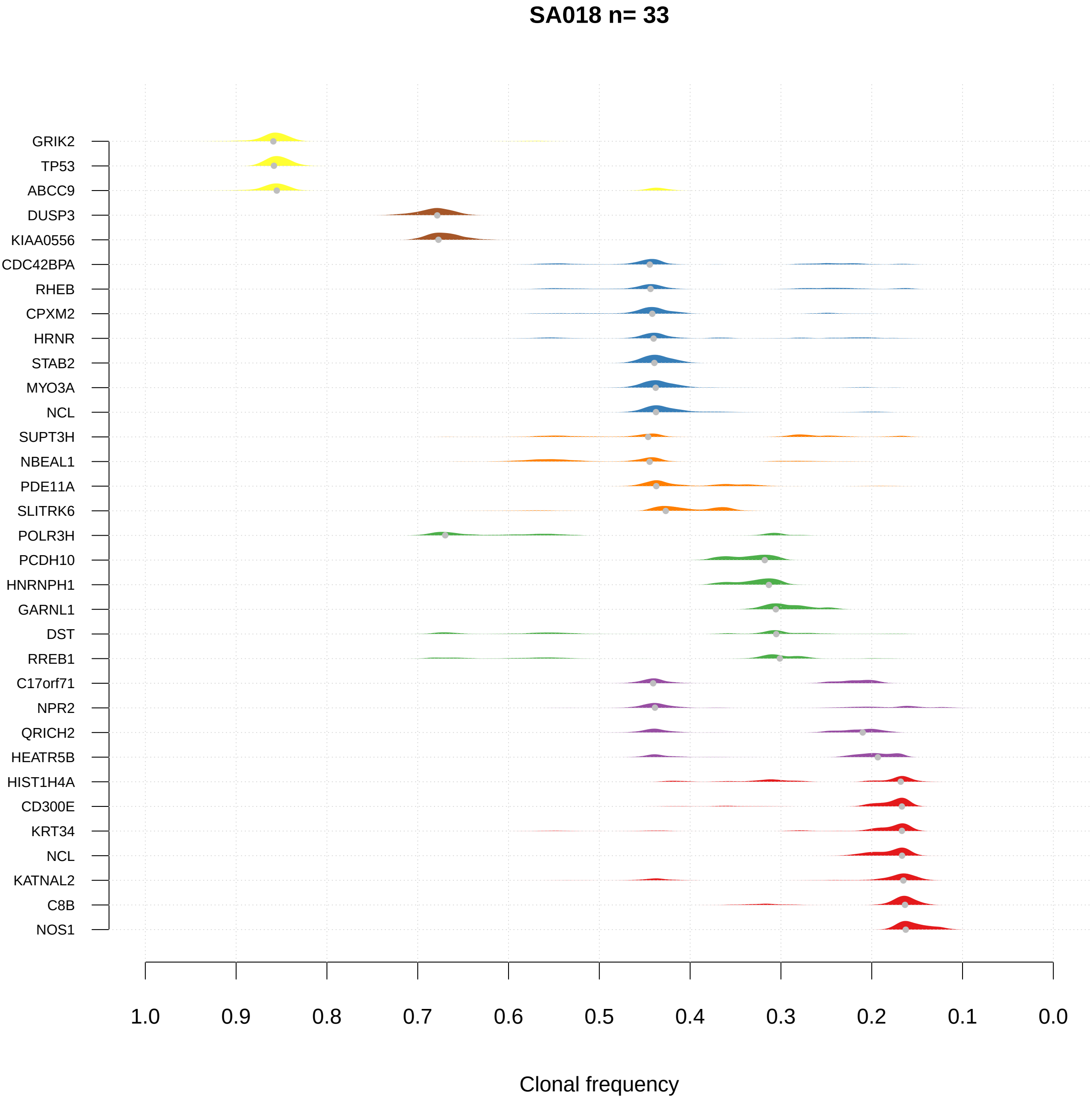
Figure S10

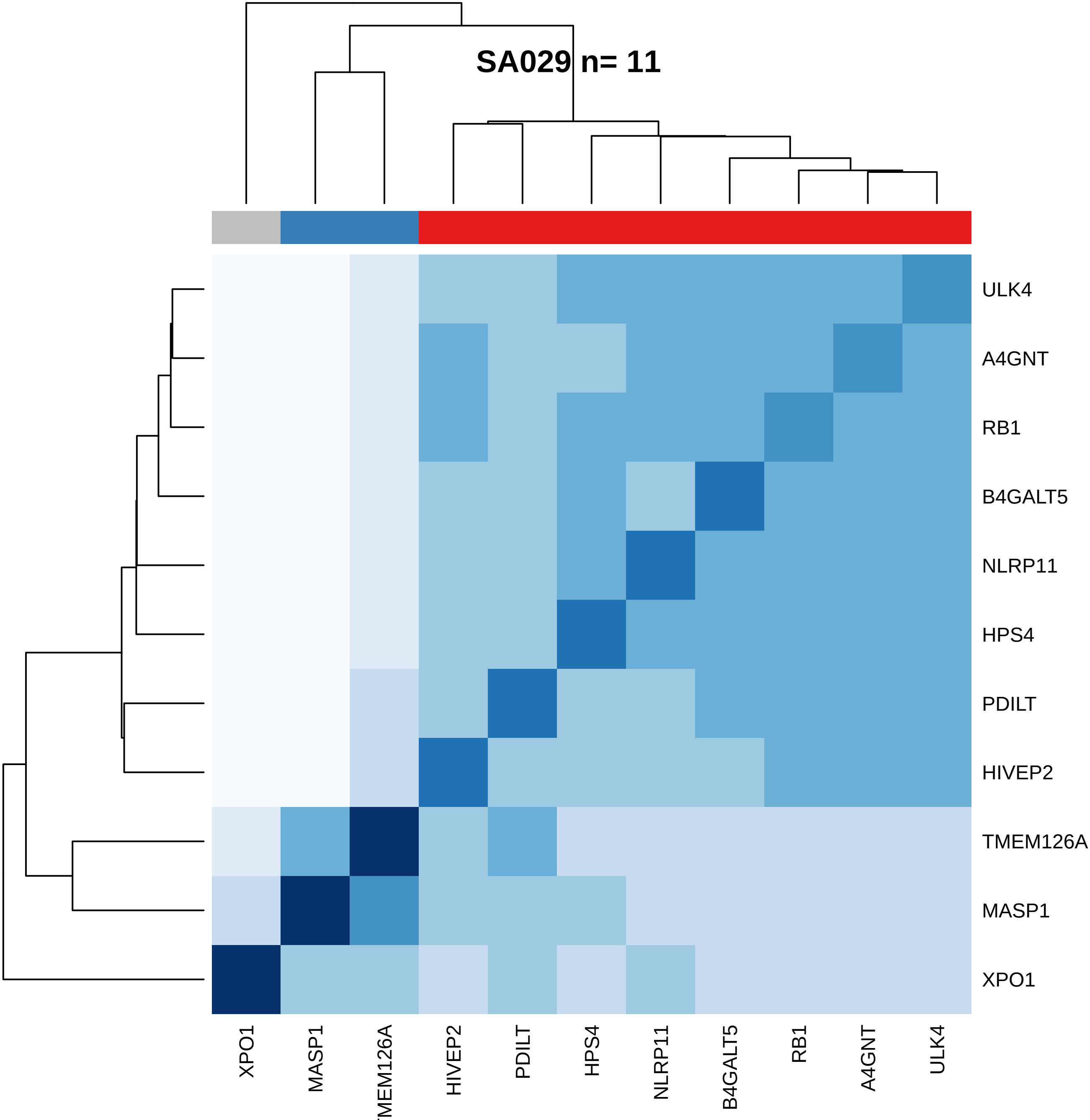


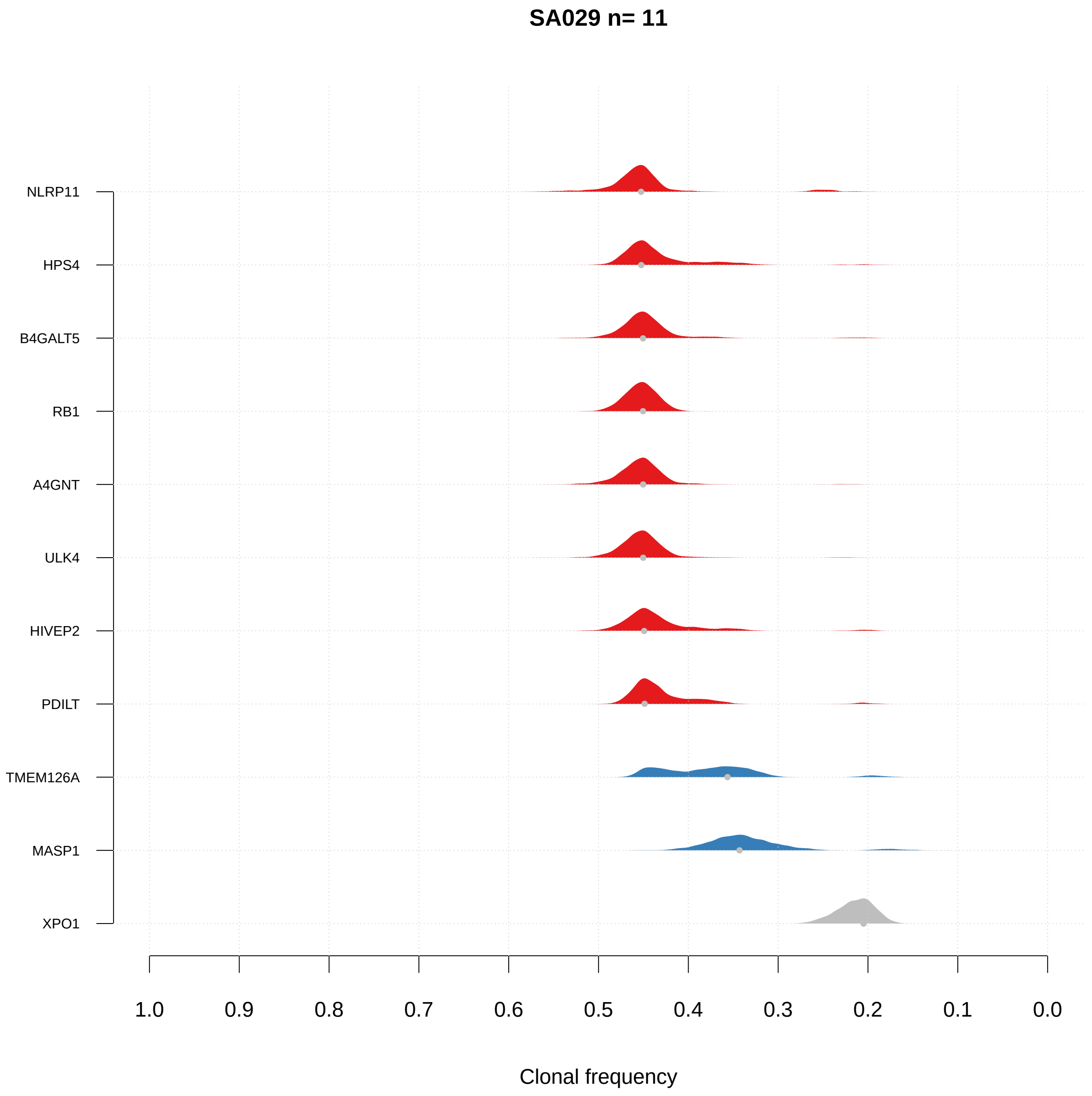
$$
\begin{aligned}
\phi^i &\sim DP(\alpha, G_0) \\
&= \text{Clonal frequency} \\
D^i_{var} | \phi^i, d^i &\sim Binomial(d^i_{var} | \phi^i, d^i) \\
&= \text{Number of reads sampled from clones} \\
A^i_{var} | G^i, \mu_{var}, D^i_{var} &\sim Binomial(a^i_{var} | \mu_{var:cg}, d^i_{var}) \\
&= \text{Number of matches to ref from clones} \\
G^i | C^i &\sim Discrete\,Uniform(g | 1, c) \\
&= \text{Genotype of clone} \\
a^i | A^i_{var}, d^i, D^i_{var}, \mu_{ref} &\sim Binomial(a^i - a^i_{var} | \mu_{ref}, d^i - d^i_{var}) \\
&= \text{Number of matches to ref}
\end{aligned}
$$

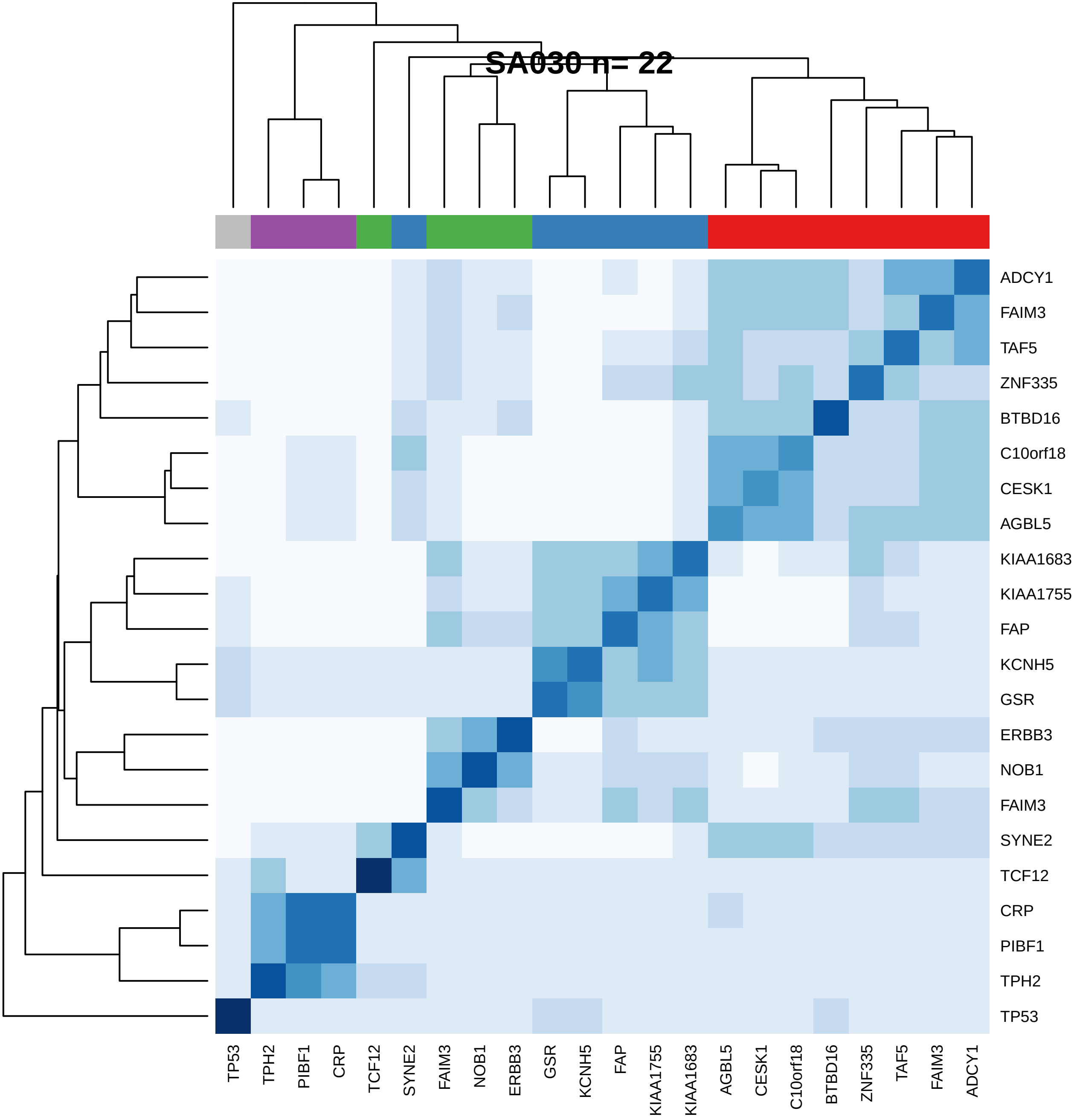
Figure S11
108 pages

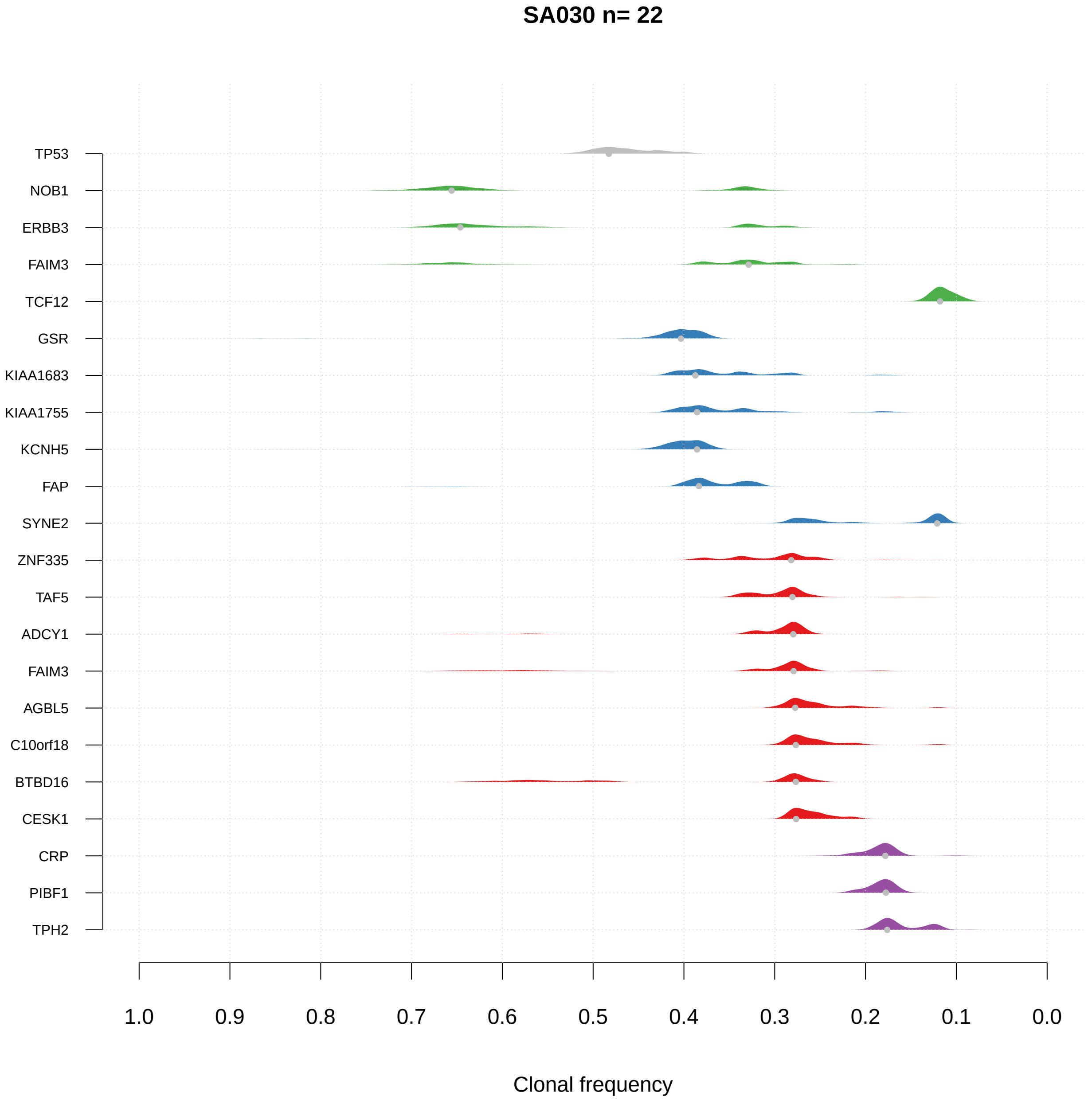


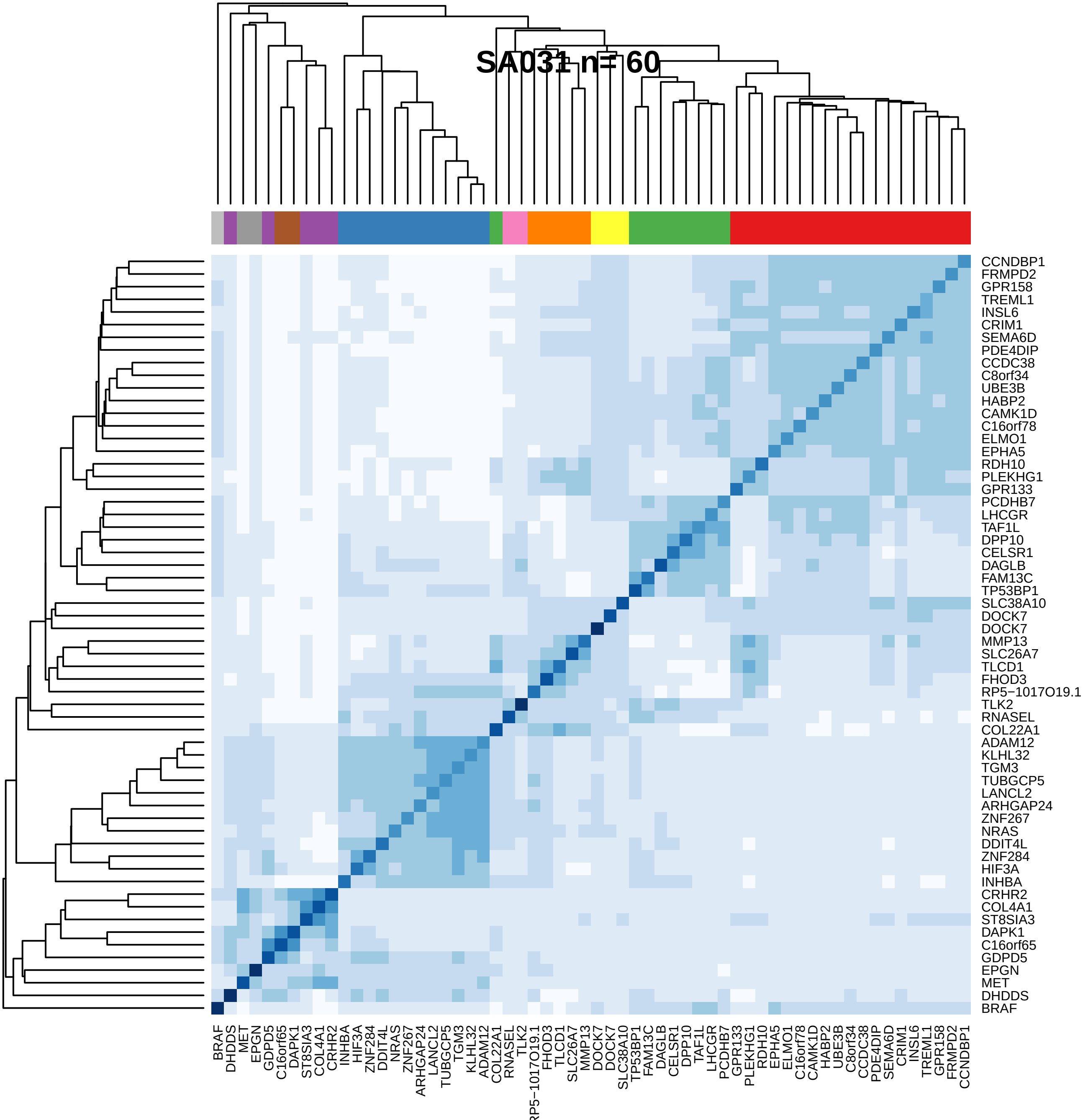

SA018 n= 33
GRIK2
TP53
ABCC9
DUSP3
KIAA0556
CDC42BPA
RHEB
CPXM2
HRNR
STAB2
MYO3A
NCL
SUPT3H
NBEAL1
PDE11A
SLITRK6
POLR3H
PCDH10
HNRNPH1
GARNL1
DST
RREB1
C17orf71
NPR2
QRICH2
HEATR5B
HIST1H4A
CD300E
KRT34
NCL
KATNAL2
C8B
NOS1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

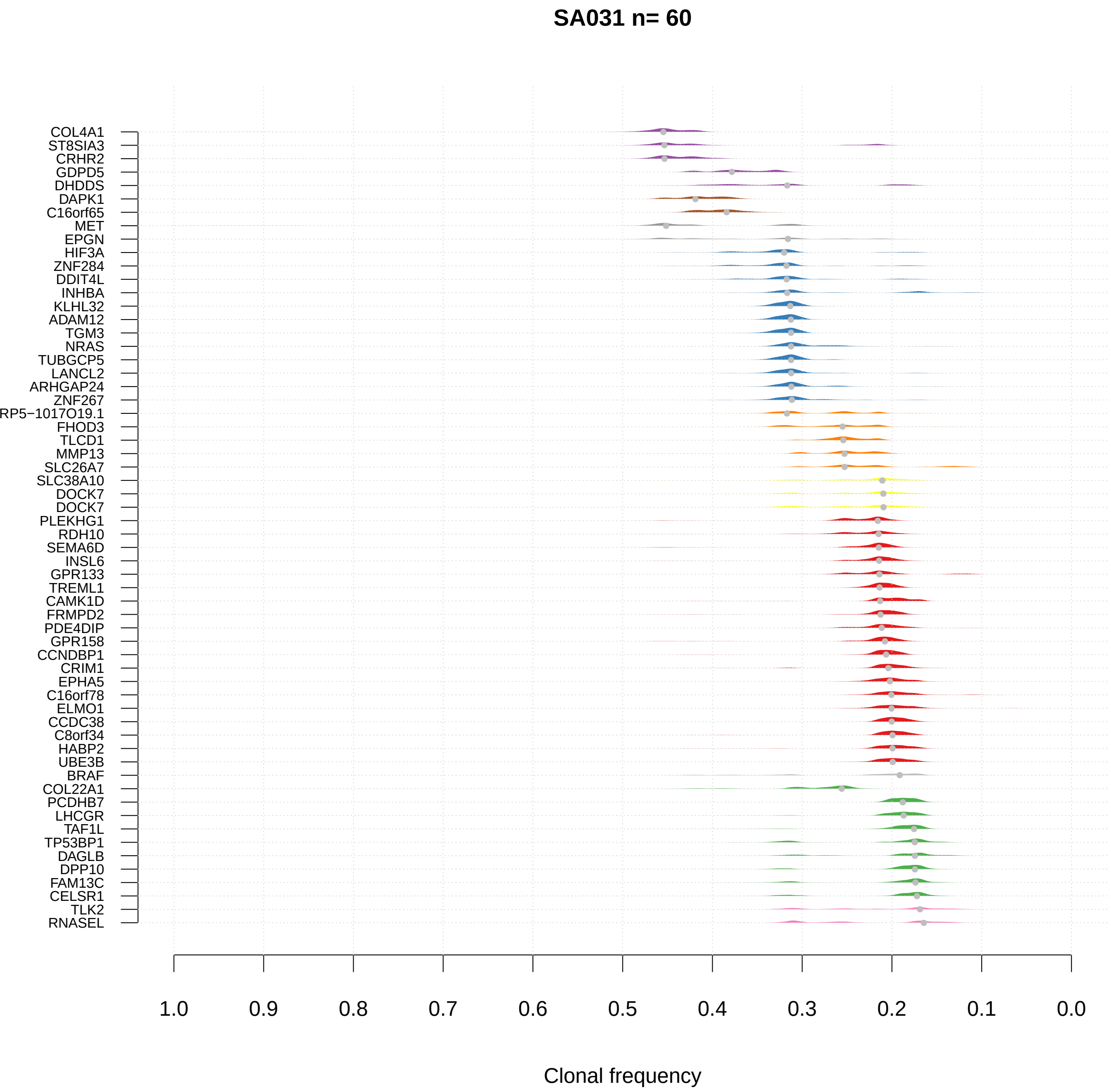

SA029 n= 11
ULK4
A4GNT
RB1
B4GALT5
NLRP11
HPS4
PDILT
HIVEP2
TMEM126A
MASP1
XPO1
XPO1
MASP1
TMEM126A
HIVEP2
PDILT
HPS4
NLRP11
B4GALT5
RB1
A4GNT
ULK4

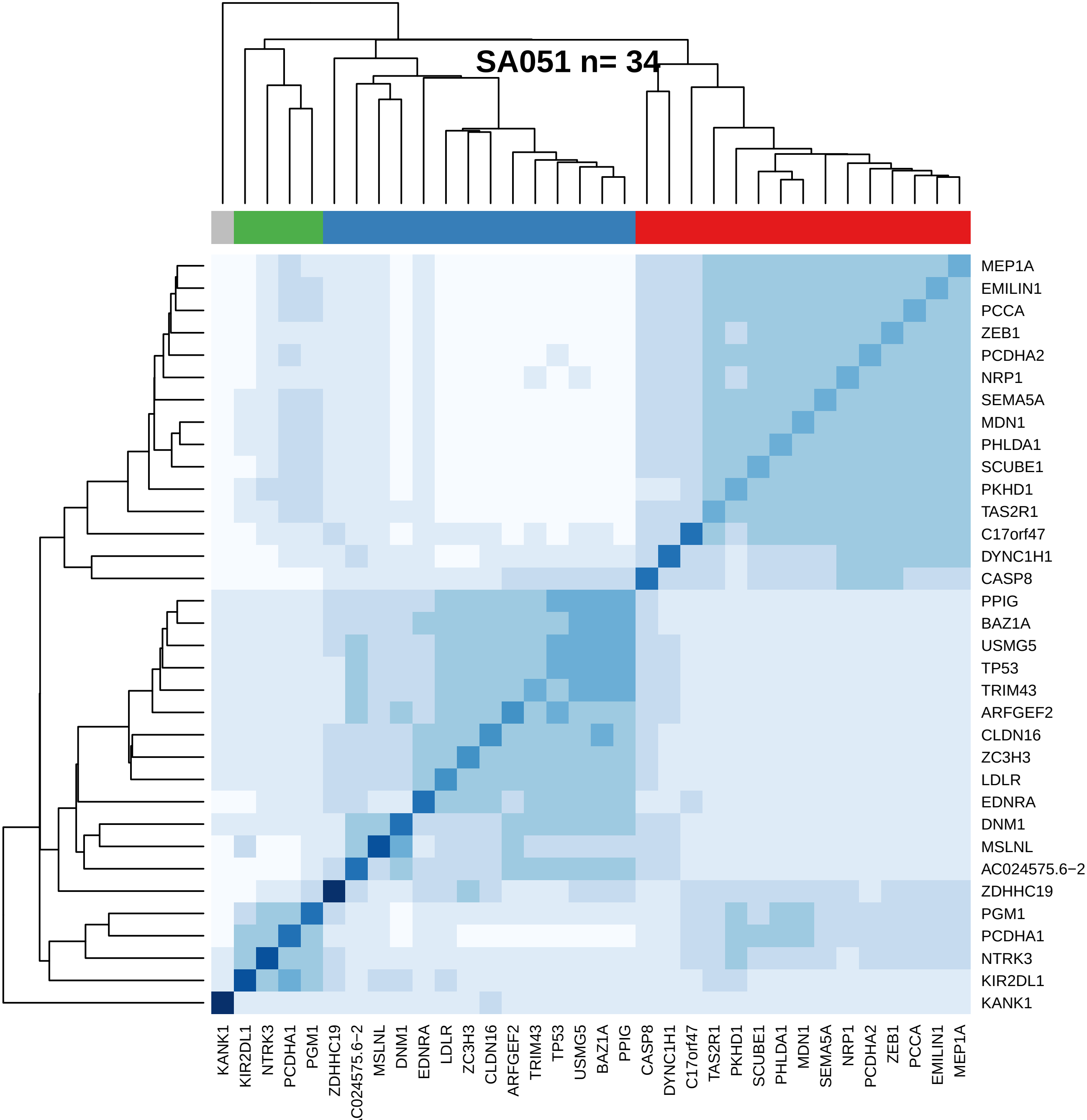

SA029 n= 11
NLRP11
HPS4
B4GALT5
RB1
A4GNT
ULK4
HIVEP2
PDILT
TMEM126A
MASP1
XPO1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

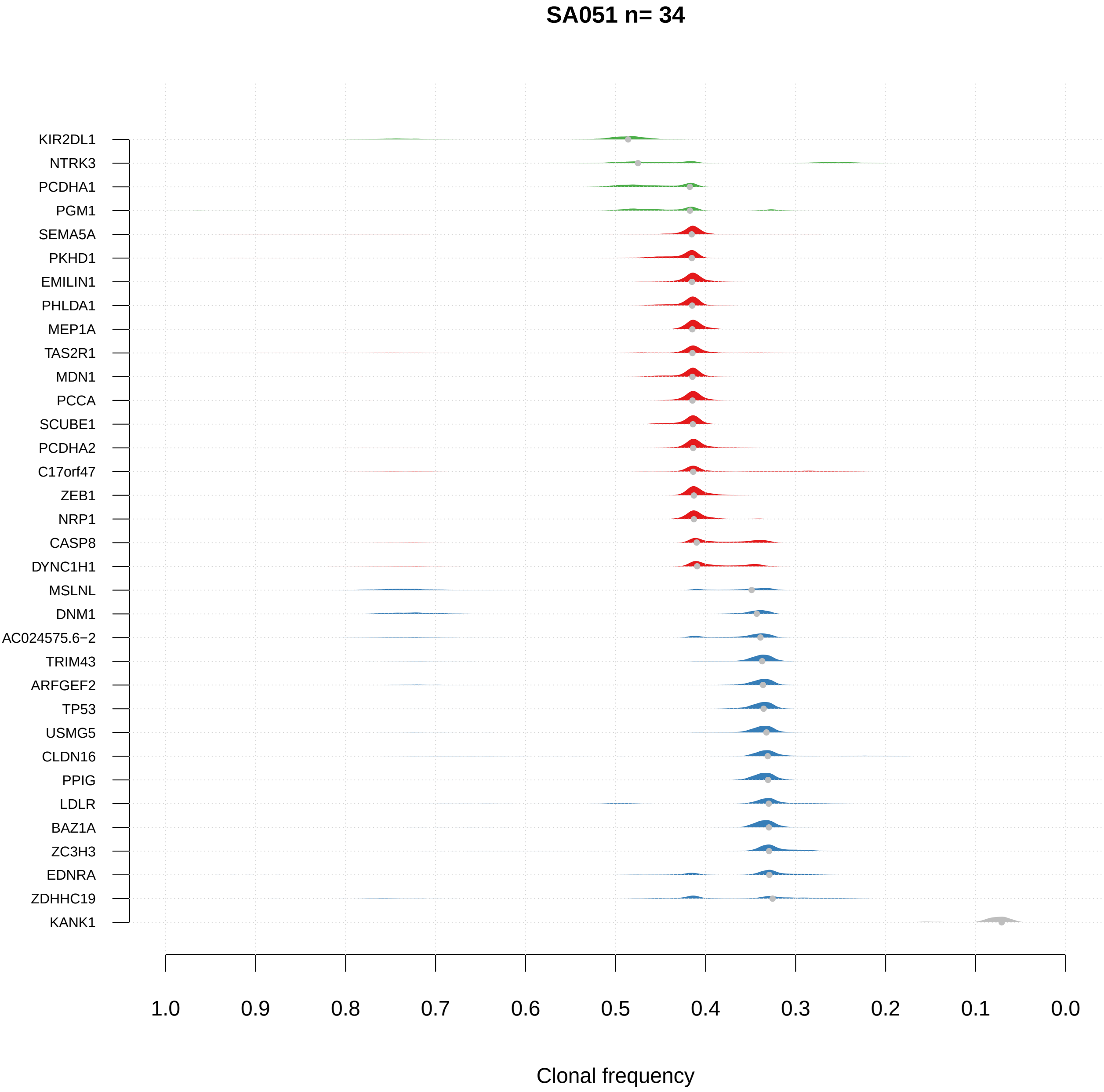

SA030 n= 22
ADCY1
FAIM3
TAF5
ZNF335
BTBD16
C10orf18
CESK1
AGBL5
KIAA1683
KIAA1755
FAP
KCNH5
GSR
ERBB3
NOB1
FAIM3
SYNE2
TCF12
CRP
PIBF1
TPH2
TP53
TP53
TPH2
PIBF1
CRP
TCF12
SYNE2
FAIM3
NOB1
ERBB3
GSR
KCNH5
FAP
KIAA1755
KIAA1683
AGBL5
CESK1
C10orf18
BTBD16
ZNF335
TAF5
FAIM3
ADCY1

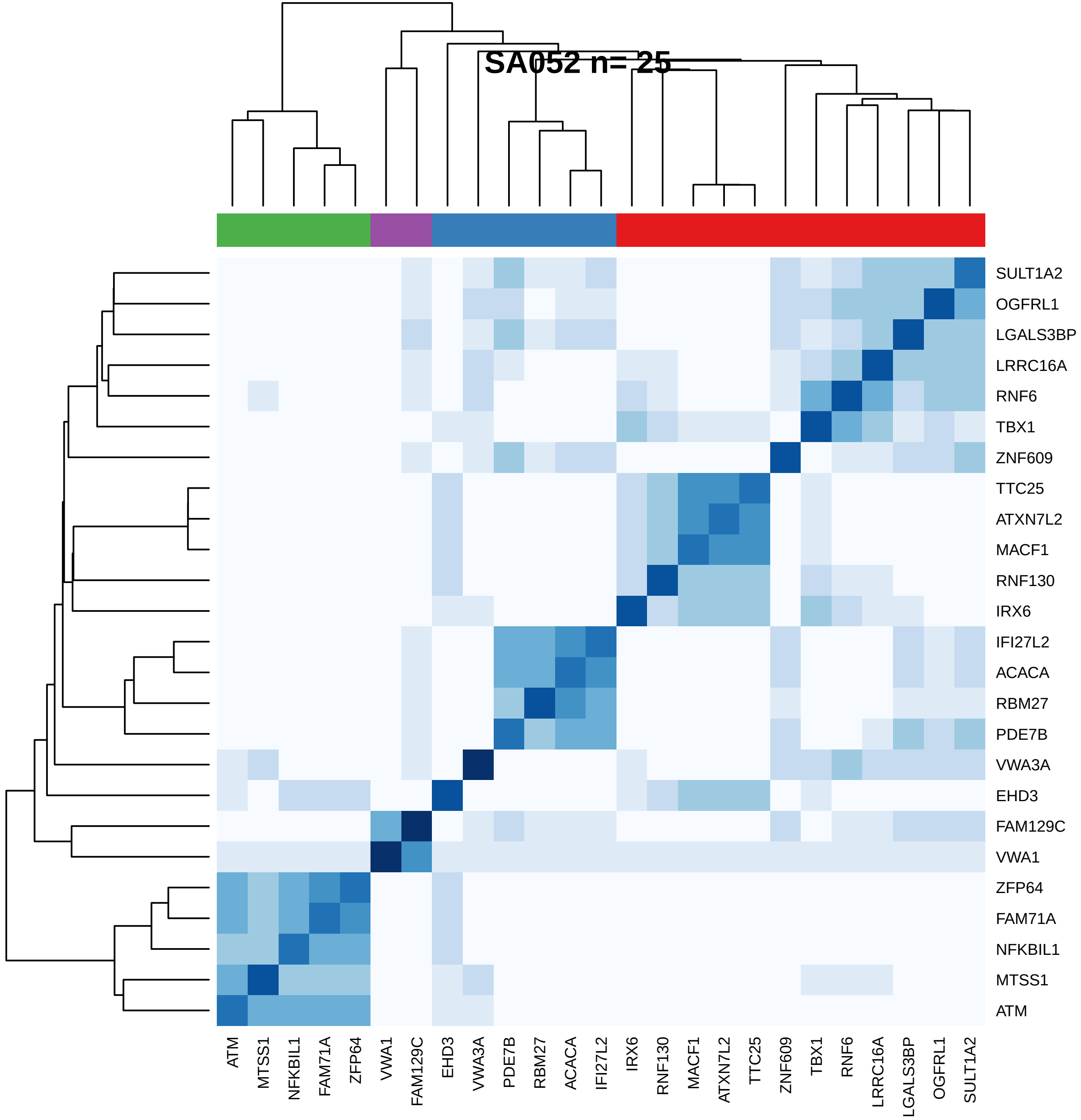

SA030 n= 22
TP53
NOB1
ERBB3
FAIM3
TCF12
GSR
KIAA1683
KIAA1755
KCNH5
FAP
SYNE2
ZNF335
TAF5
ADCY1
FAIM3
AGBL5
C10orf18
BTBD16
CESK1
CRP
PIBF1
TPH2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

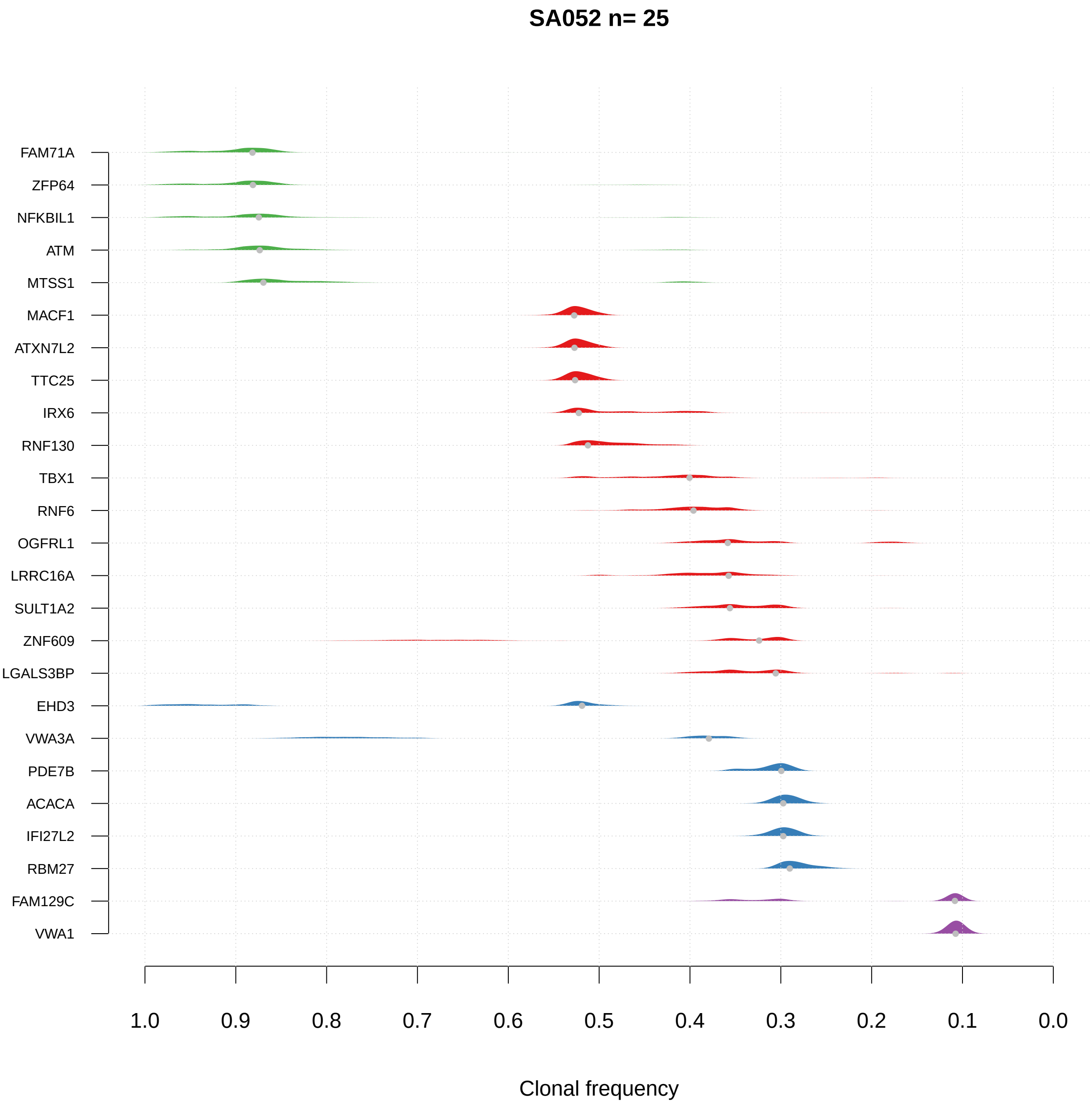

SA031 n= 60
CCNDBP1
FRMPD2
GPR158
TREML1
INSL6
CRIM1
SEMA6D
PDE4DIP
CCDC38
C8orf34
UBE3B
HABP2
CAMK1D
C16orf78
ELMO1
EPHA5
RDH10
PLEKHG1
GPR133
PCDHB7
LHCGR
TAF1L
DPP10
CELSR1
DAGLB
FAM13C
TP53BP1
SLC38A10
DOCK7
DOCK7
MMP13
SLC26A7
TLCD1
FHOD3
RP5−1017O19.1
TLK2
RNASEL
COL22A1
ADAM12
KLHL32
TGM3
TUBGCP5
LANCL2
ARHGAP24
ZNF267
NRAS
DDIT4L
ZNF284
HIF3A
INHBA
CRHR2
COL4A1
ST8SIA3
DAPK1
C16orf65
GDPD5
EPGN
MET
DHDDS
BRAF
BRAF
DHDDS
MET
EPGN
GDPD5
C16orf65
DAPK1
ST8SIA3
COL4A1
CRHR2
INHBA
HIF3A
ZNF284
DDIT4L
NRAS
ZNF267
ARHGAP24
LANCL2
TUBGCP5
TGM3
KLHL32
ADAM12
COL22A1
RNASEL
TLK2
RP5−1017O19.1
FHOD3
TLCD1
SLC26A7
MMP13
DOCK7
DOCK7
SLC38A10
TP53BP1
FAM13C
DAGLB
CELSR1
DPP10
TAF1L
LHCGR
PCDHB7
GPR133
PLEKHG1
RDH10
EPHA5
ELMO1
C16orf78
CAMK1D
HABP2
UBE3B
C8orf34
CCDC38
PDE4DIP
SEMA6D
CRIM1
INSL6
TREML1
GPR158
FRMPD2
CCNDBP1

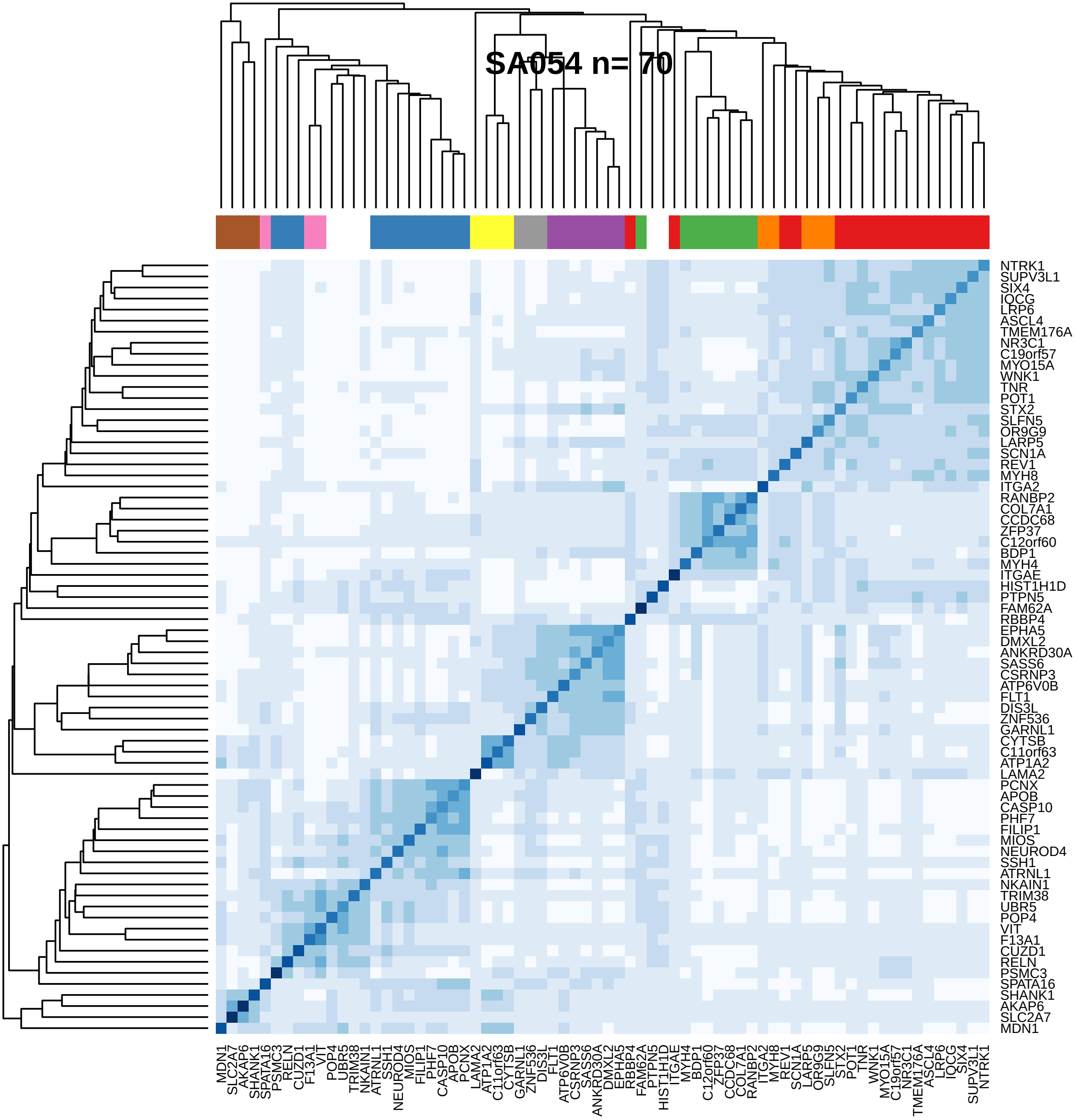

SA031 n= 60
COL4A1
ST8SIA3
CRHR2
GDPD5
DHDDS
DAPK1
C16orf65
MET
EPGN
HIF3A
ZNF284
DDIT4L
INHBA
KLHL32
ADAM12
TGM3
NRAS
TUBGCP5
LANCL2
ARHGAP24
ZNF267
RP5−1017O19.1
FHOD3
TLCD1
MMP13
SLC26A7
SLC38A10
DOCK7
DOCK7
PLEKHG1
RDH10
SEMA6D
INSL6
GPR133
TREML1
CAMK1D
FRMPD2
PDE4DIP
GPR158
CCNDBP1
CRIM1
EPHA5
C16orf78
ELMO1
CCDC38
C8orf34
HABP2
UBE3B
BRAF
COL22A1
PCDHB7
LHCGR
TAF1L
TP53BP1
DAGLB
DPP10
FAM13C
CELSR1
TLK2
RNASEL
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

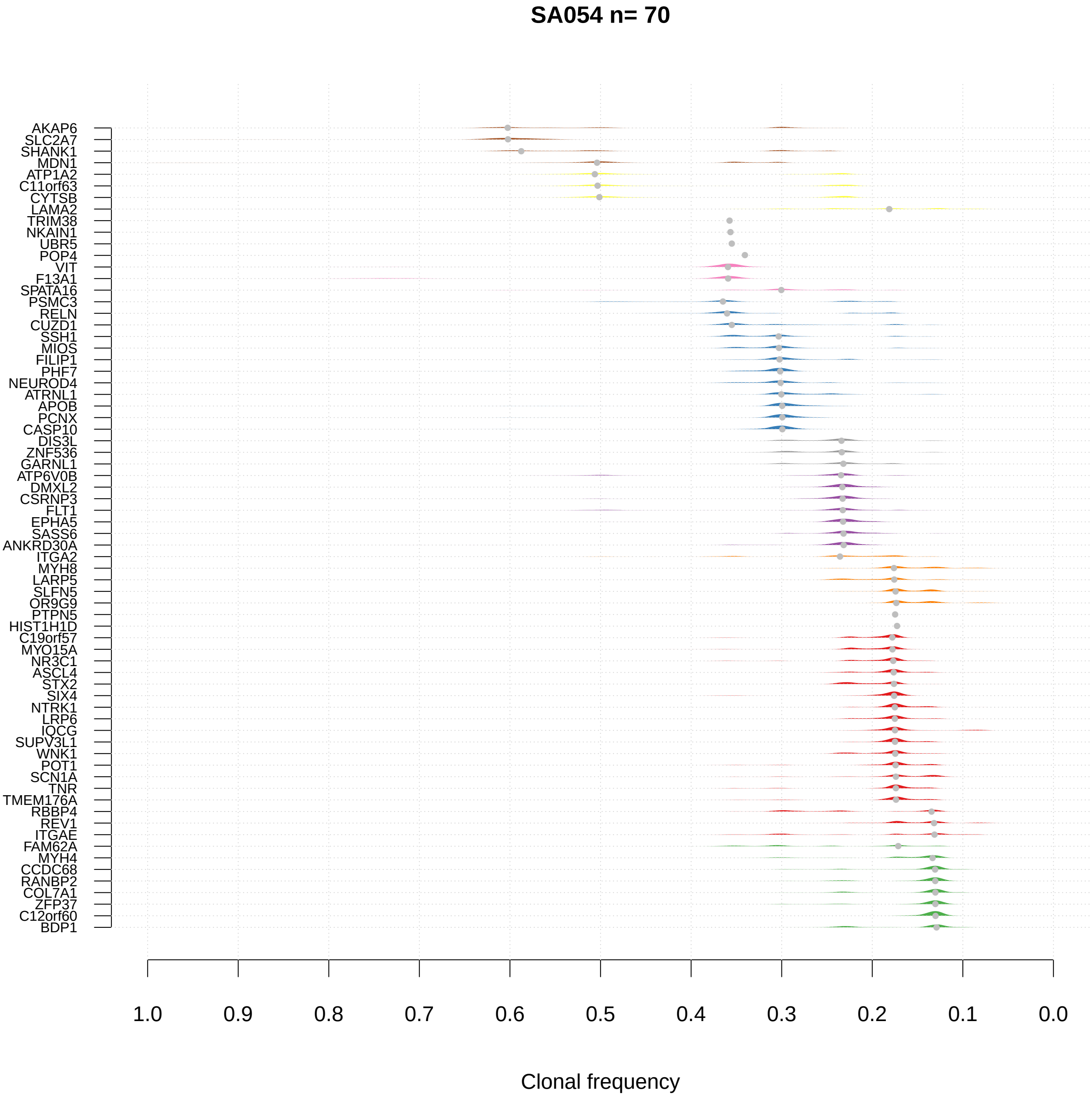

SA051 n= 34
MEP1A
EMILIN1
PCCA
ZEB1
PCDHA2
NRP1
SEMA5A
MDN1
PHLDA1
SCUBE1
PKHD1
TAS2R1
C17orf47
DYNC1H1
CASP8
PPIG
BAZ1A
USMG5
TP53
TRIM43
ARFGEF2
CLDN16
ZC3H3
LDLR
EDNRA
DNM1
MSLNL
AC024575.6−2
ZDHHC19
PGM1
PCDHA1
NTRK3
KIR2DL1
KANK1
KANK1
KIR2DL1
NTRK3
PCDHA1
PGM1
ZDHHC19
AC024575.6−2
MSLNL
DNM1
EDNRA
LDLR
ZC3H3
CLDN16
ARFGEF2
TRIM43
TP53
USMG5
BAZ1A
PPIG
CASP8
DYNC1H1
C17orf47
TAS2R1
PKHD1
SCUBE1
PHLDA1
MDN1
SEMA5A
NRP1
PCDHA2
ZEB1
PCCA
EMILIN1
MEP1A

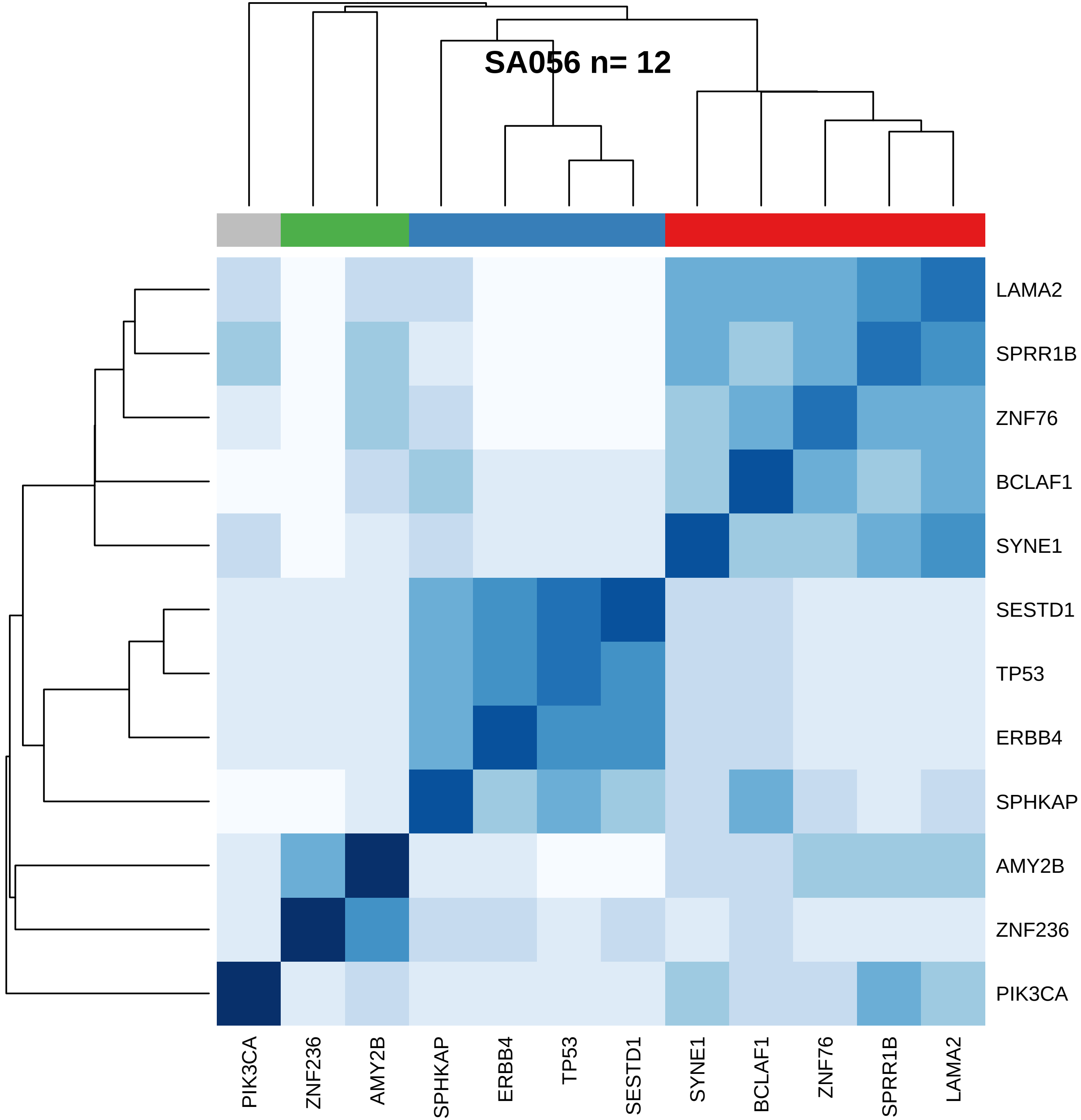

SA051 n= 34
KIR2DL1
NTRK3
PCDHA1
PGM1
SEMA5A
PKHD1
EMILIN1
PHLDA1
MEP1A
TAS2R1
MDN1
PCCA
SCUBE1
PCDHA2
C17orf47
ZEB1
NRP1
CASP8
DYNC1H1
MSLNL
DNM1
AC024575.6−2
TRIM43
ARFGEF2
TP53
USMG5
CLDN16
PPIG
LDLR
BAZ1A
ZC3H3
EDNRA
ZDHHC19
KANK1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

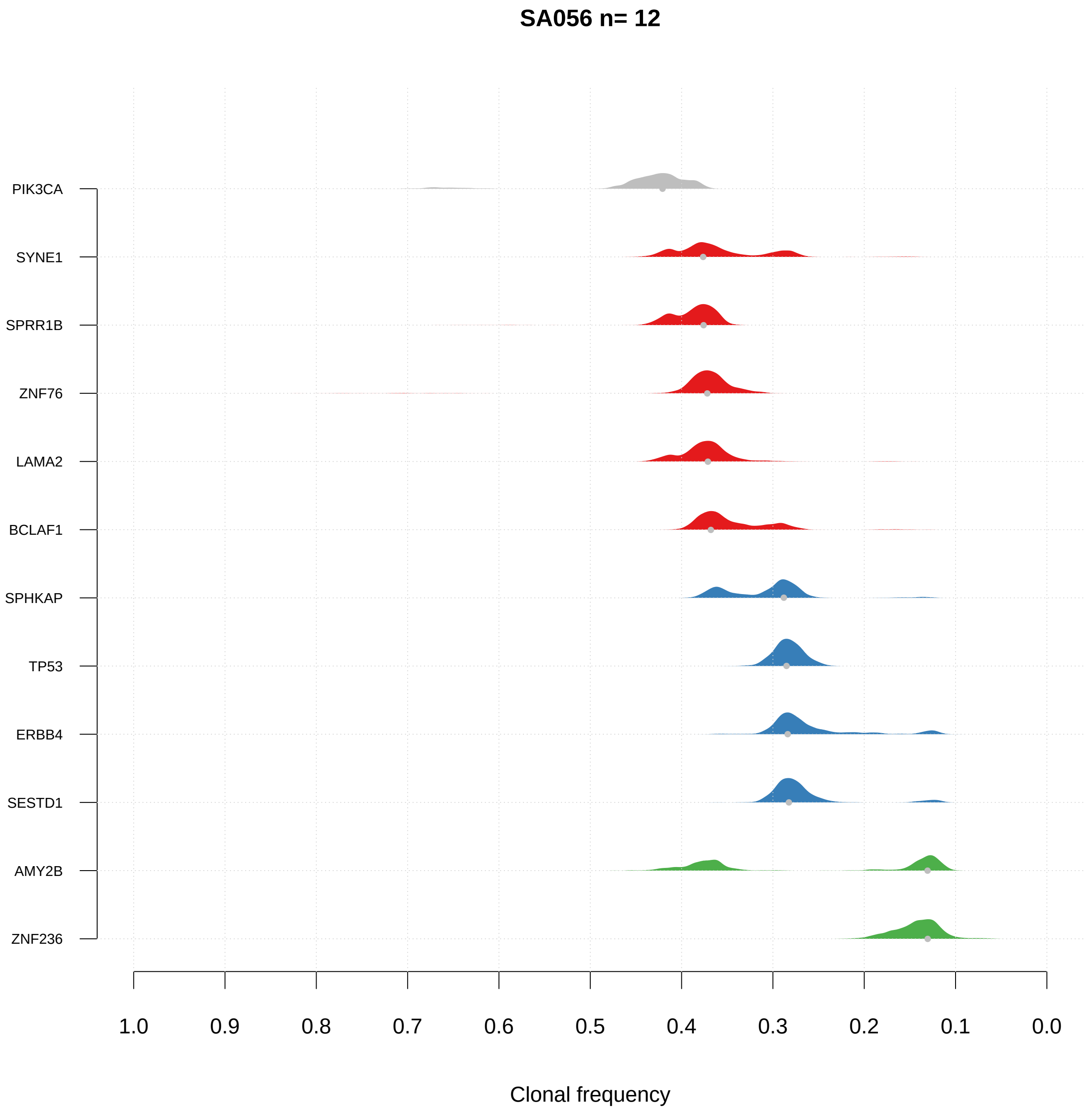

SA052 n= 25
SULT1A2
OGFRL1
LGALS3BP
LRRC16A
RNF6
TBX1
ZNF609
TTC25
ATXN7L2
MACF1
RNF130
IRX6
IFI27L2
ACACA
RBM27
PDE7B
VWA3A
EHD3
FAM129C
VWA1
ZFP64
FAM71A
NFKBIL1
MTSS1
ATM
ATM
MTSS1
NFKBIL1
FAM71A
ZFP64
VWA1
FAM129C
EHD3
VWA3A
PDE7B
RBM27
ACACA
IFI27L2
IRX6
RNF130
MACF1
ATXN7L2
TTC25
ZNF609
TBX1
RNF6
LRRC16A
LGALS3BP
OGFRL1
SULT1A2

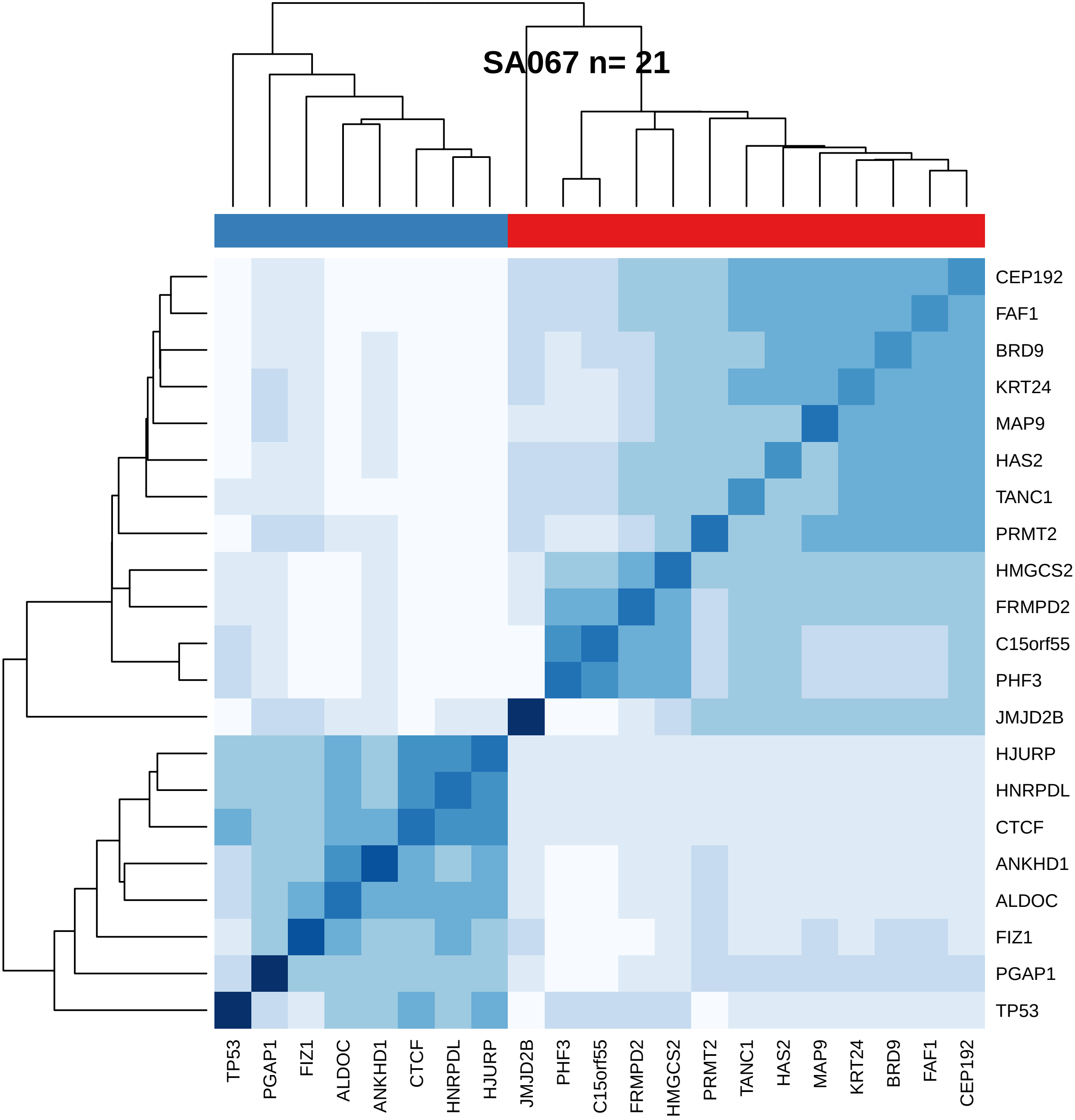

SA052 n= 25
FAM71A
ZFP64
NFKBIL1
ATM
MTSS1
MACF1
ATXN7L2
TTC25
IRX6
RNF130
TBX1
RNF6
OGFRL1
LRRC16A
SULT1A2
ZNF609
LGALS3BP
EHD3
VWA3A
PDE7B
ACACA
IFI27L2
RBM27
FAM129C
VWA1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

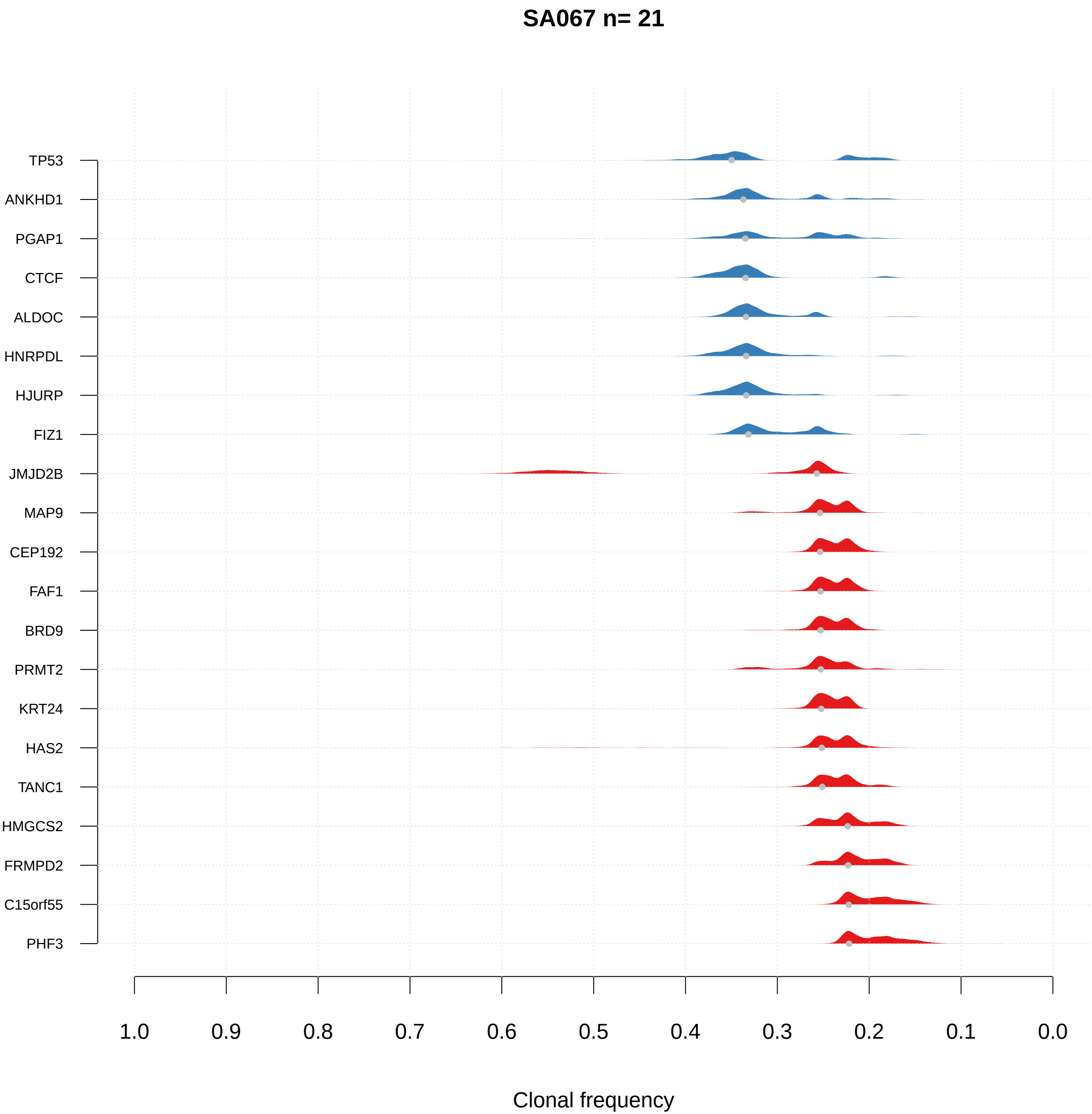

SA054 n= 70
NTRK1
SUPV3L1
SIX4
IOCG
LRP6
ASCL4
TMEM176A
NR3C1
C19orf57
MYO15A
WNK1
TNR
POT1
STX2
SLFN5
OR9G9
LARP5
SCN1A
REV1
MYH8
ITGA2
RANBP2
COL7A1
CCDC68
ZFP37
C12orf60
BDP1
MYH4
ITGAE
HIST1H1D
PTPN5
FAM62A
RBBP4
EPHA5
DMXL2
ANKRD30A
SASS6
CSRNP3
ATP6V0B
FLT1
DIS3L
ZNF536
GARNL1
CYTSB
C11orf63
ATP1A2
LAMA2
PCNX
APOB
CASP10
PHF7
FILIP1
MIOS
NEUROD4
SSH1
ATRNL1
NKAIN1
TRIM38
UBR5
POP4
VIT
F13A1
CUZD1
RELN
PSMC3
SPATA16
SHANK1
AKAP6
SLC2A7
MDN1
MDN1
SLC2A7
AKAP6
SHANK1
SPATA16
PSMC3
RELN
CUZD1
F13A1
VIT
POP4
UBR5
TRIM38
NKAIN1
ATRNL1
SSH1
NEUROD4
MIOS
FILIP1
PHF7
CASP10
APOB
PCNX
LAMA2
ATP1A2
C11orf63
CYTSB
GARNL1
ZNF536
DIS3L
FLT1
ATP6V0B
CSRNP3
SASS6
ANKRD30A
DMXL2
EPHA5
RBBP4
FAM62A
PTPN5
HIST1H1D
ITGAE
MYH4
BDP1
C12orf60
ZFP37
CCDC68
COL7A1
RANBP2
ITGA2
MYH8
REV1
SCN1A
LARP5
OR9G9
SLFN5
STX2
POT1
TNR
WNK1
MYO15A
C19orf57
NR3C1
TMEM176A
ASCL4
LRP6
IOCG
SIX4
SUPV3L1
NTRK1

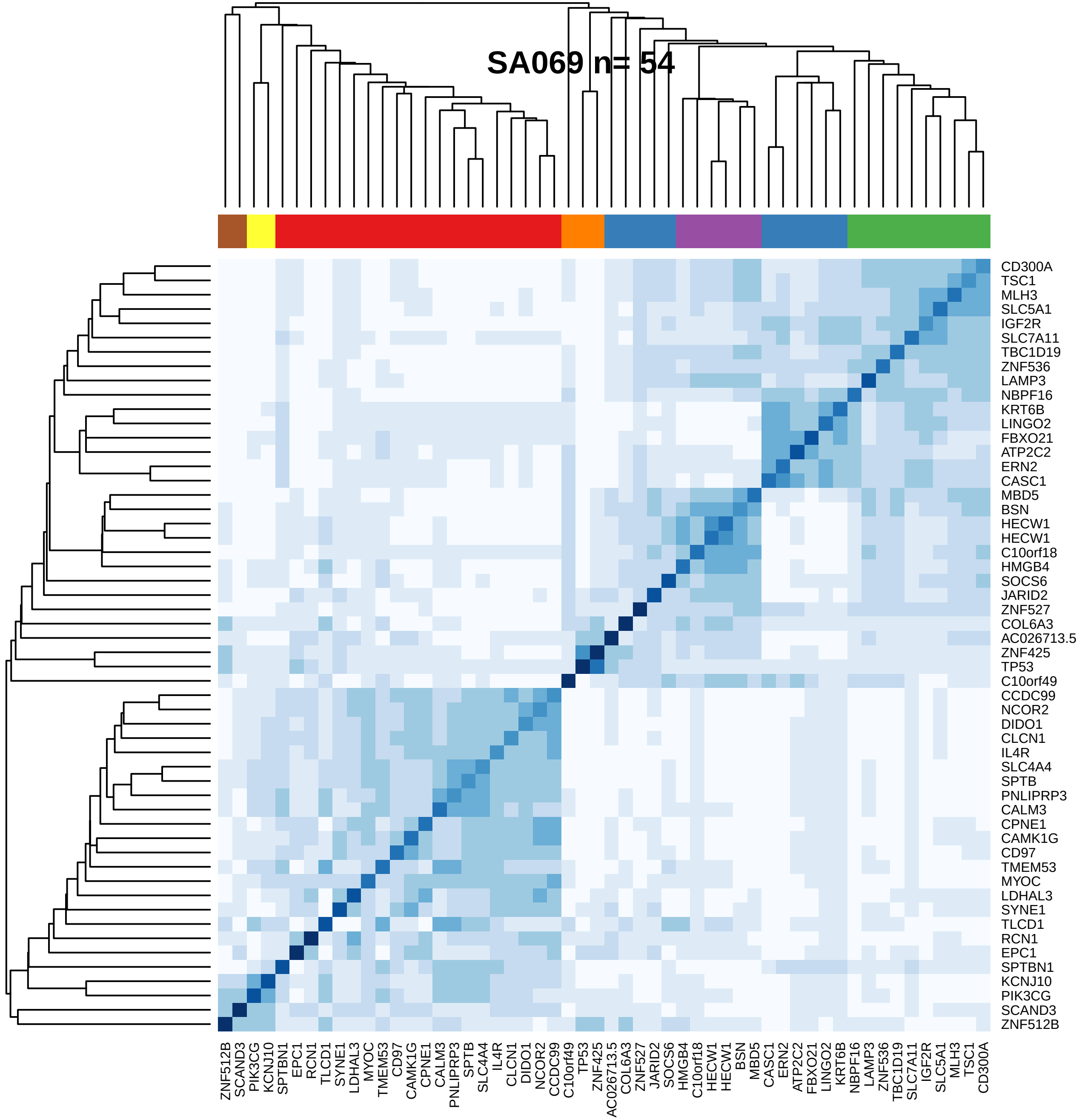

SA054 n= 70
AKAP6
SLC2A7
SHANK1
MDN1
ATP1A2
C11orf63
CYTSB
LAMA2
TRIM38
NKAIN1
UBR5
POP4
VIT
F13A1
SPATA16
PSMC3
RELN
CUZD1
SSH1
MIOS
FILIP1
PHF7
NEUROD4
ATRNL1
APOB
PCNX
CASP10
DIS3L
ZNF536
GARNL1
ATP6V0B
DMXL2
CSRNP3
FLT1
EPHA5
SASS6
ANKRD30A
ITGA2
MYH8
LARP5
SLFN5
OR9G9
PTPN5
HIST1H1D
C19orf57
MYO15A
NR3C1
ASCL4
STX2
SIX4
NTRK1
LRP6
IQCG
SUPV3L1
WNK1
POT1
SCN1A
TNR
TMEM176A
RBBP4
REV1
ITGAE
FAM62A
MYH4
CCDC68
RANBP2
COL7A1
ZFP37
C12orf60
BDP1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA056 n= 12
LAMA2
SPRR1B
ZNF76
BCLAF1
SYNE1
SESTD1
TP53
ERBB4
SPHKAP
AMY2B
ZNF236
PIK3CA
PIK3CA
ZNF236
AMY2B
SPHKAP
ERBB4
TP53
SESTD1
SYNE1
BCLAF1
ZNF76
SPRR1B
LAMA2


SA056 n= 12
PIK3CA
SYNE1
SPRR1B
ZNF76
LAMA2
BCLAF1
SPHKAP
TP53
ERBB4
SESTD1
AMY2B
ZNF236
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA067 n= 21
CEP192
FAF1
BRD9
KRT24
MAP9
HAS2
TANC1
PRMT2
HMGCS2
FRMPD2
C15orf55
PHF3
JMJD2B
HJURP
HNRPDL
CTCF
ANKHD1
ALDOC
FIZ1
PGAP1
TP53
TP53
PGAP1
FIZ1
ALDOC
ANKHD1
CTCF
HNRPDL
HJURP
JMJD2B
PHF3
C15orf55
FRMPD2
HMGCS2
PRMT2
TANC1
HAS2
MAP9
KRT24
BRD9
FAF1
CEP192

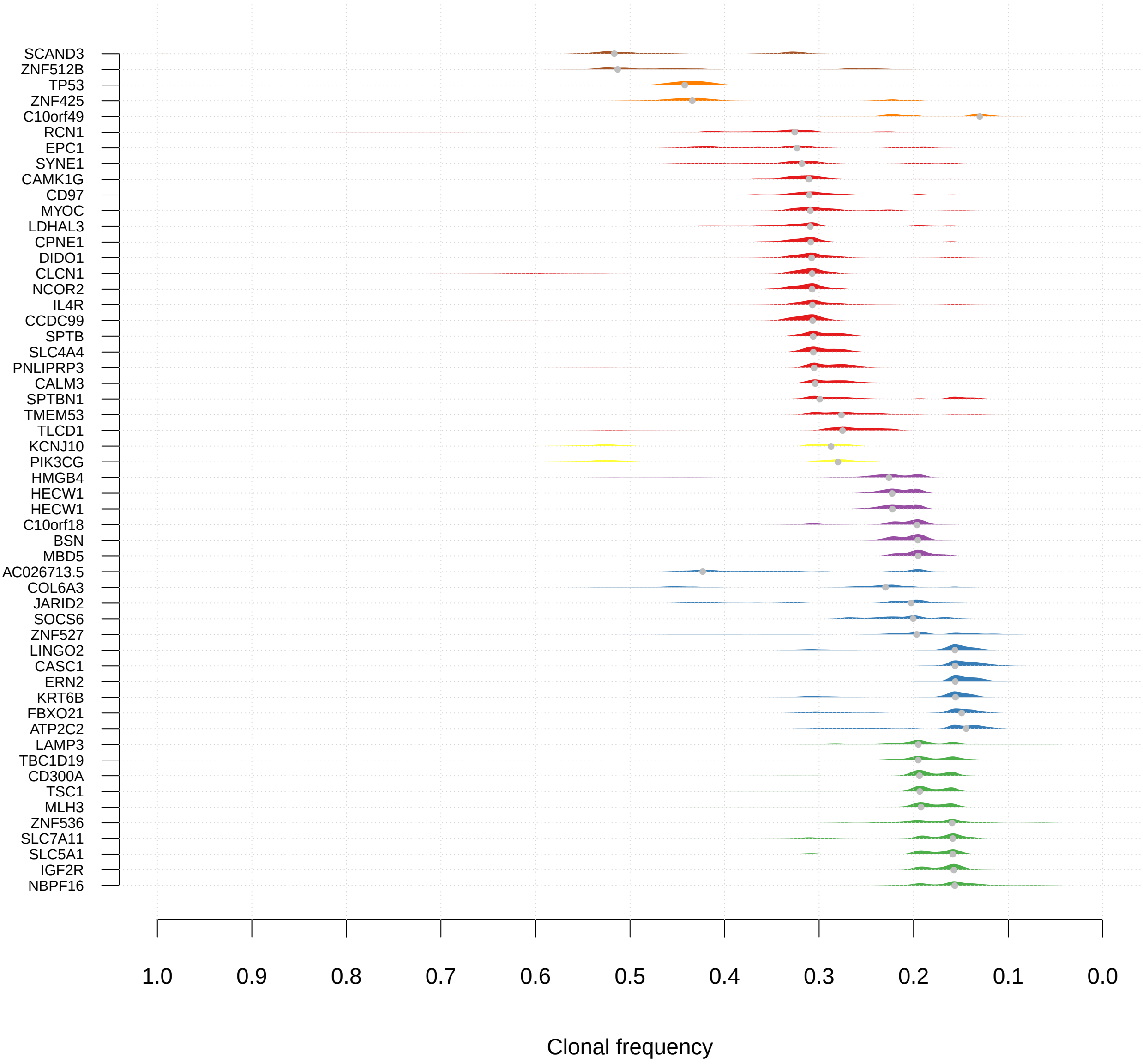

SA067 n= 21
TP53
ANKHD1
PGAP1
CTCF
ALDOC
HNRPDL
HJURP
FIZ1
JMJD2B
MAP9
CEP192
FAF1
BRD9
PRMT2
KRT24
HAS2
TANC1
HMGCS2
FRMPD2
C15orf55
PHF3
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

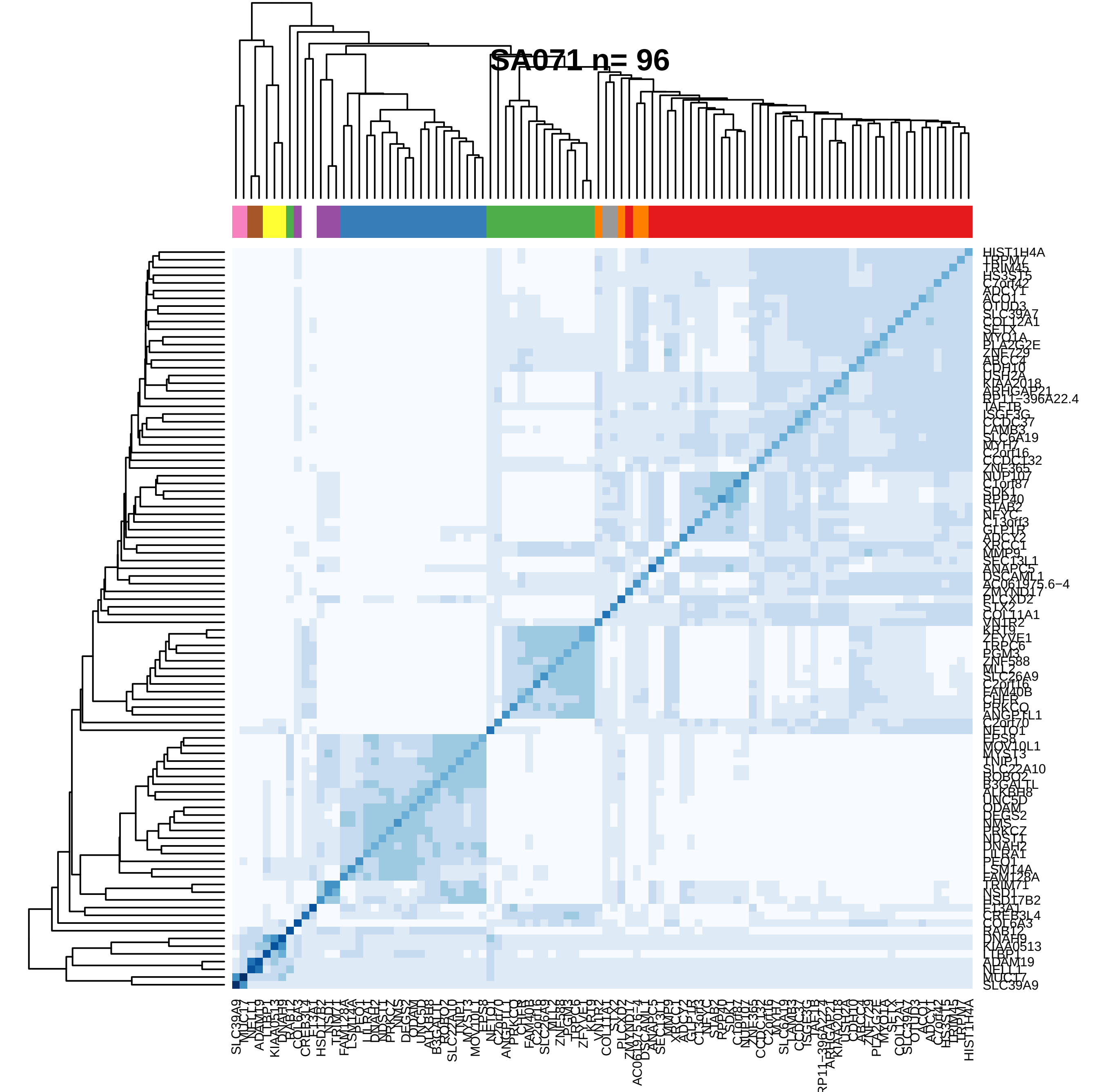


SA069 n= 54

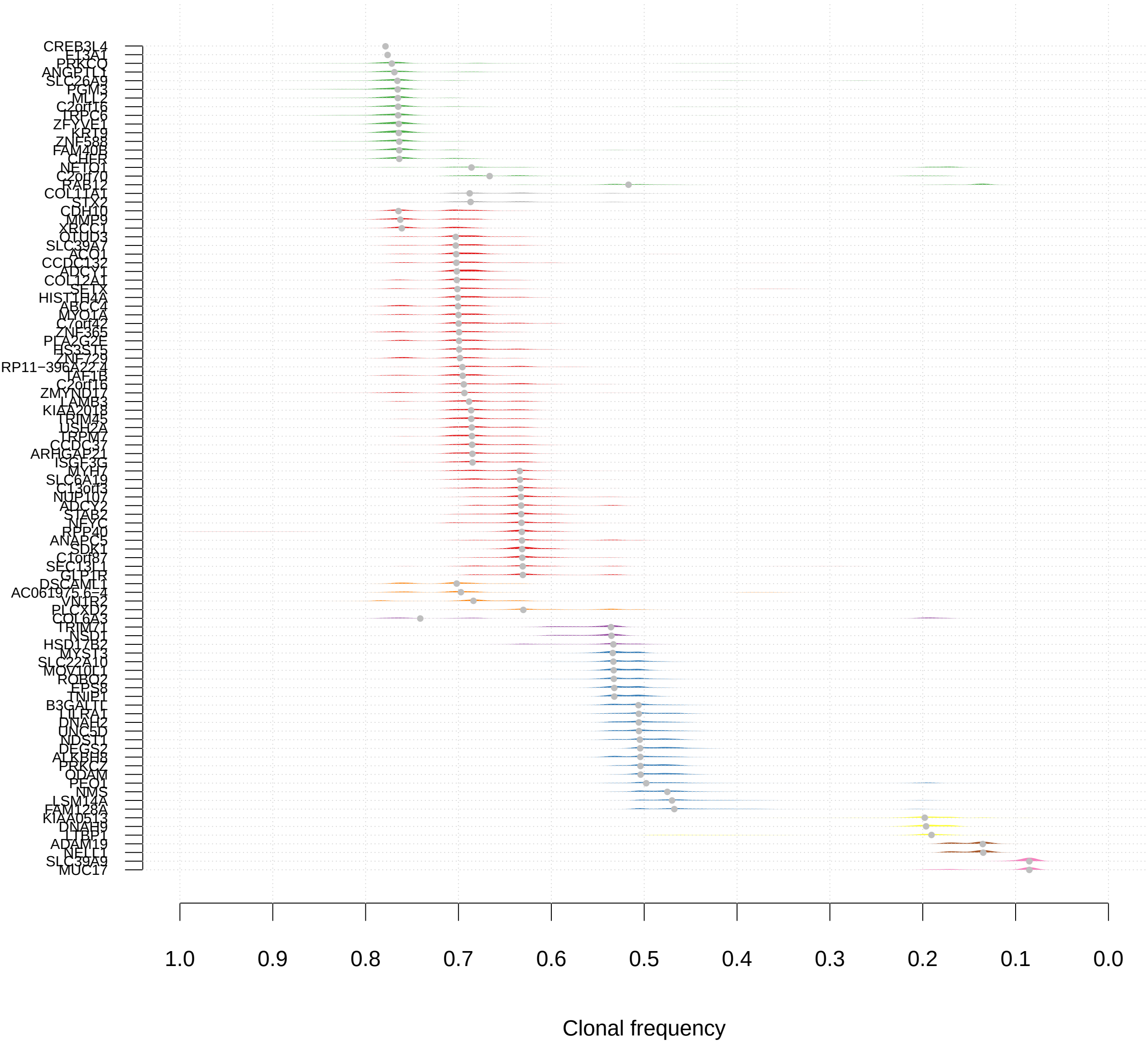
CD300A
TSC1
MLH3
SLC5A1
IGF2R
SLC7A11
TBC1D19
ZNF536
LAMP3
NBPF16
KRT6B
LINGO2
FBXO21
ATP2C2
ERN2
CASC1
MBD5
BSN
HECW1
HECW1
C10orf18
HMGB4
SOCS6
JARID2
ZNF527
COL6A3
AC026713.5
ZNF425
TP53
C10orf49
CCDC99
NCOR2
DIDO1
CLCN1
IL4R
SLC4A4
SPTB
PNLIPRP3
CALM3
CPNE1
CAMK1G
CD97
TMEM53
MYOC
LDHAL3
SYNE1
TLCD1
RCN1
EPC1
SPTBN1
KCNJ10
PIK3CG
SCAND3
ZNF512B
ZNF512B
SCAND3
PIK3CG
KCNJ10
SPTBN1
EPC1
RCN1
TLCD1
SYNE1
LDHAL3
MYOC
TMEM53
CD97
CAMK1G
CPNE1
CALM3
PNLIPRP3
SPTB
SLC4A4
IL4R
CLCN1
DIDO1
NCOR2
CCDC99
C10orf49
TP53
ZNF425
AC026713.5
COL6A3
ZNF527
JARID2
SOCS6
HMGB4
C10orf18
HECW1
HECW1
BSN
MBD5
CASC1
ERN2
ATP2C2
FBXO21
LINGO2
KRT6B
NBPF16
LAMP3
ZNF536
TBC1D19
SLC7A11
IGF2R
SLC5A1
MLH3
TSC1
CD300A

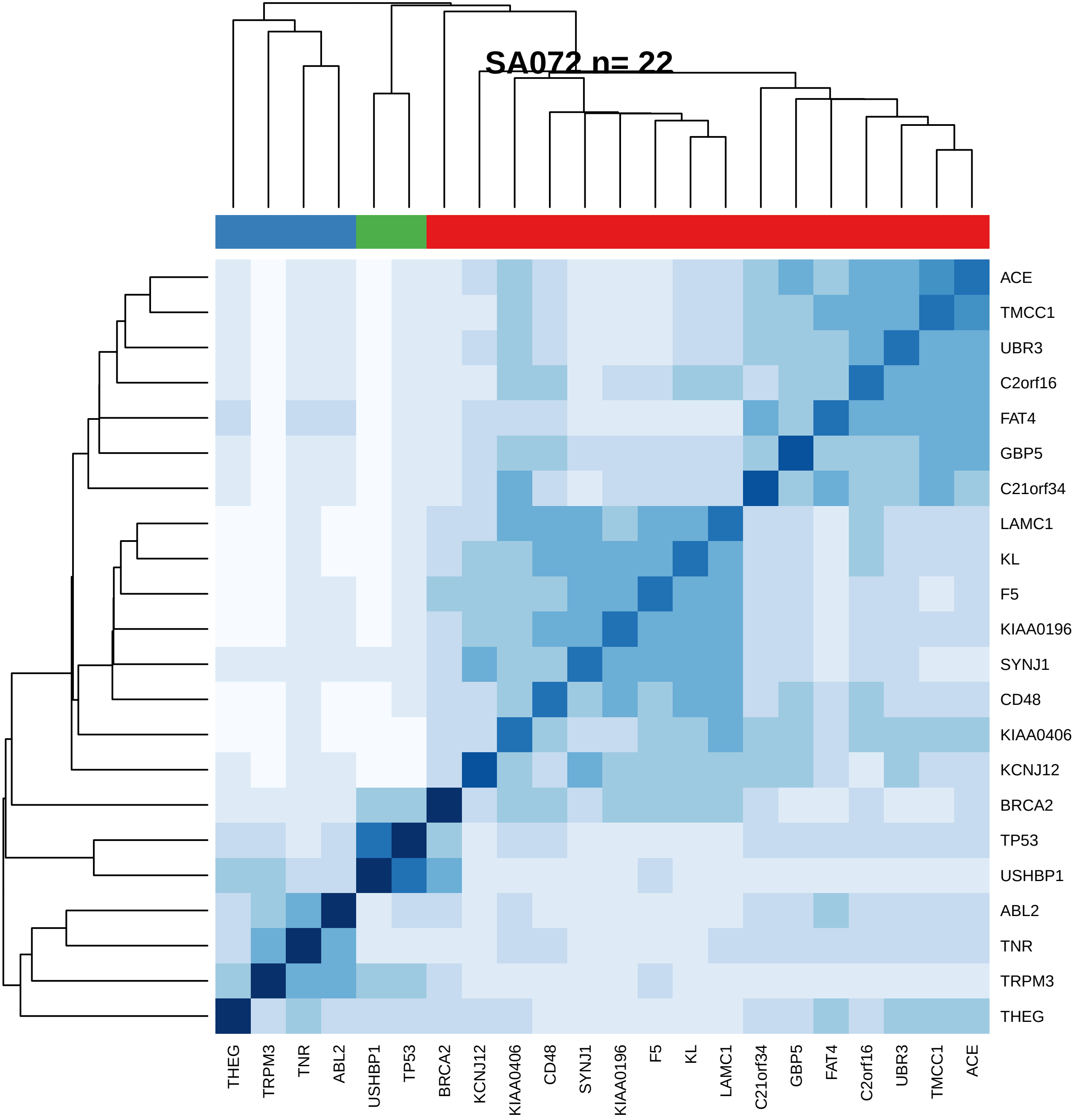
SA069 n= 54

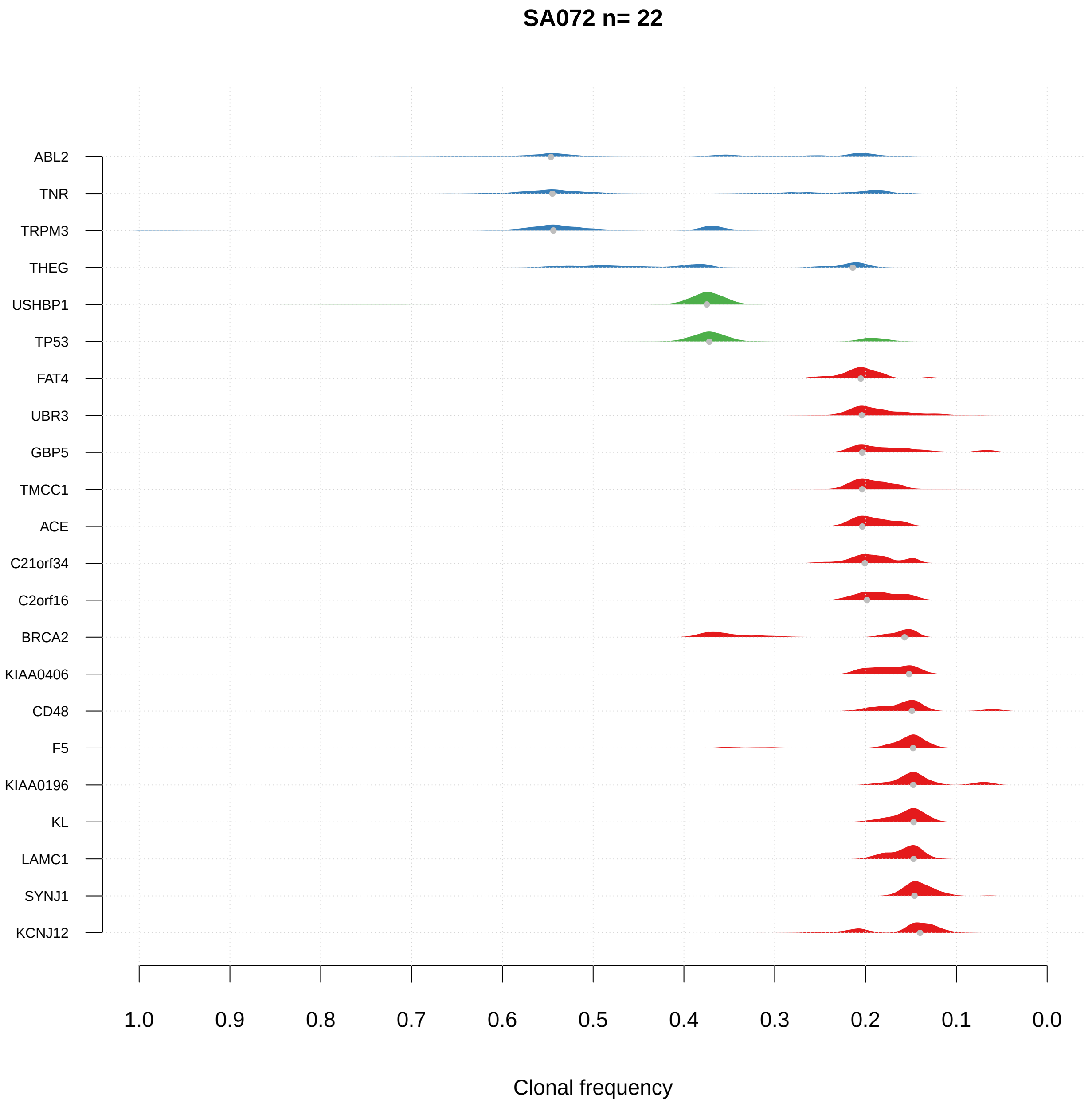

SCAND3
ZNF512B
TP53
ZNF425
C10orf49
RCN1
EPC1
SYNE1
CAMK1G
CD97
MYOC
LDHAL3
CPNE1
DIDO1
CLCN1
NCOR2
IL4R
CCDC99
SPTB
SLC4A4
PNLIPRP3
CALM3
SPTBN1
TMEM53
TLCD1
KCNJ10
PIK3CG
HMGB4
HECW1
HECW1
C10orf18
BSN
MBD5
AC026713.5
COL6A3
JARID2
SOCS6
ZNF527
LINGO2
CASC1
ERN2
KRT6B
FBXO21
ATP2C2
LAMP3
TBC1D19
CD300A
TSC1
MLH3
ZNF536
SLC7A11
SLC5A1
IGF2R
NBPF16
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

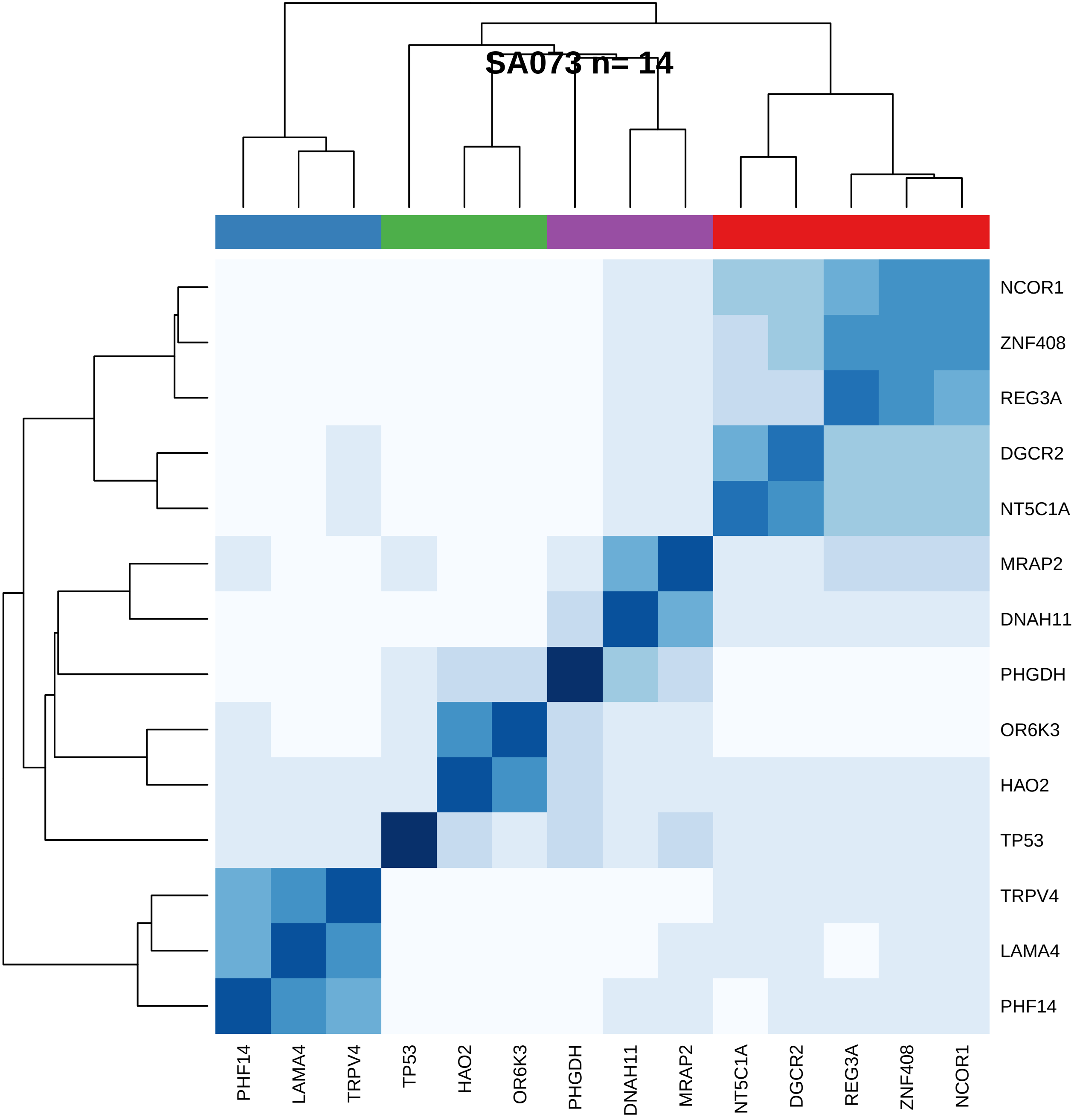

SA071 n= 96
HIST1H4A
TRPM7
TRIM45
HS3ST5
C7orf42
ADCY1
ACO1
OTUD3
SLC39A7
COL12A1
SETX
MYO1A
PLA2G2E
ZNF729
ABCC4
CDH10
USH2A
KIAA2018
ARHGAP21
RP11−396A22.4
TAF1B
ISGF3G
CCDC37
LAMB3
SLC6A19
MYH7
C2orf16
CCDC132
ZNF365
NUP107
C1orf87
SDK1
RPP40
STAB2
NFYC
C13orf3
GLP1R
ADCY2
XRCC1
MMP9
SEC13L1
ANAPC5
DSCAML1
AC061975.6−4
ZMYND17
PLCXD2
STX2
COL11A1
VN1R2
KRT9
ZFYVE1
TRPC6
PGM3
ZNF588
MLL2
SLC26A9
C2orf16
FAM40B
CHFR
PRKCQ
ANGPTL1
C2orf70
NETO1
EPS8
MOV10L1
MYST3
TNIP1
SLC22A10
ROBO2
B3GALTL
ALKBH8
UNC5D
ODAM
DEGS2
NMS
PRKCZ
NDST1
DNAH2
LILRA1
PEO1
LSM14A
FAM128A
TRIM71
NSD1
HSD17B2
F13A1
CREB3L4
COL6A3
RAB12
DNAH9
KIAA0513
LTBP1
ADAM19
NELL1
MUC17
SLC39A9
SLC39A9
MUC17
NELL1
ADAM19
LTBP1
KIAA0513
DNAH9
RAB12
COL6A3
CREB3L4
F13A1
HSD17B2
NSD1
TRIM71
FAM128A
LSM14A
PEO1
LILRA1
DNAH2
NDST1
PRKCZ
NMS
DEGS2
ODAM
UNC5D
ALKBH8
B3GALTL
ROBO2
SLC22A10
TNIP1
MYST3
MOV10L1
EPS8
NETO1
C2orf70
ANGPTL1
PRKCQ
CHFR
FAM40B
C2orf16
SLC26A9
MLL2
ZNF588
PGM3
TRPC6
ZFYVE1
KRT9
VN1R2
COL11A1
STX2
PLCXD2
ZMYND17
AC061975.6−4
DSCAML1
ANAPC5
SEC13L1
MMP9
XRCC1
ADCY2
GLP1R
C13orf3
NFYC
STAB2
RPP40
SDK1
C1orf87
NUP107
ZNF365
CCDC132
C2orf16
MYH7
SLC6A19
LAMB3
CCDC37
ISGF3G
TAF1B
RP11−396A22.4
ARHGAP21
KIAA2018
USH2A
CDH10
ABCC4
ZNF729
PLA2G2E
MYO1A
SETX
COL12A1
SLC39A7
OTUD3
ACO1
ADCY1
C7orf42
HS3ST5
TRIM45
TRPM7
HIST1H4A

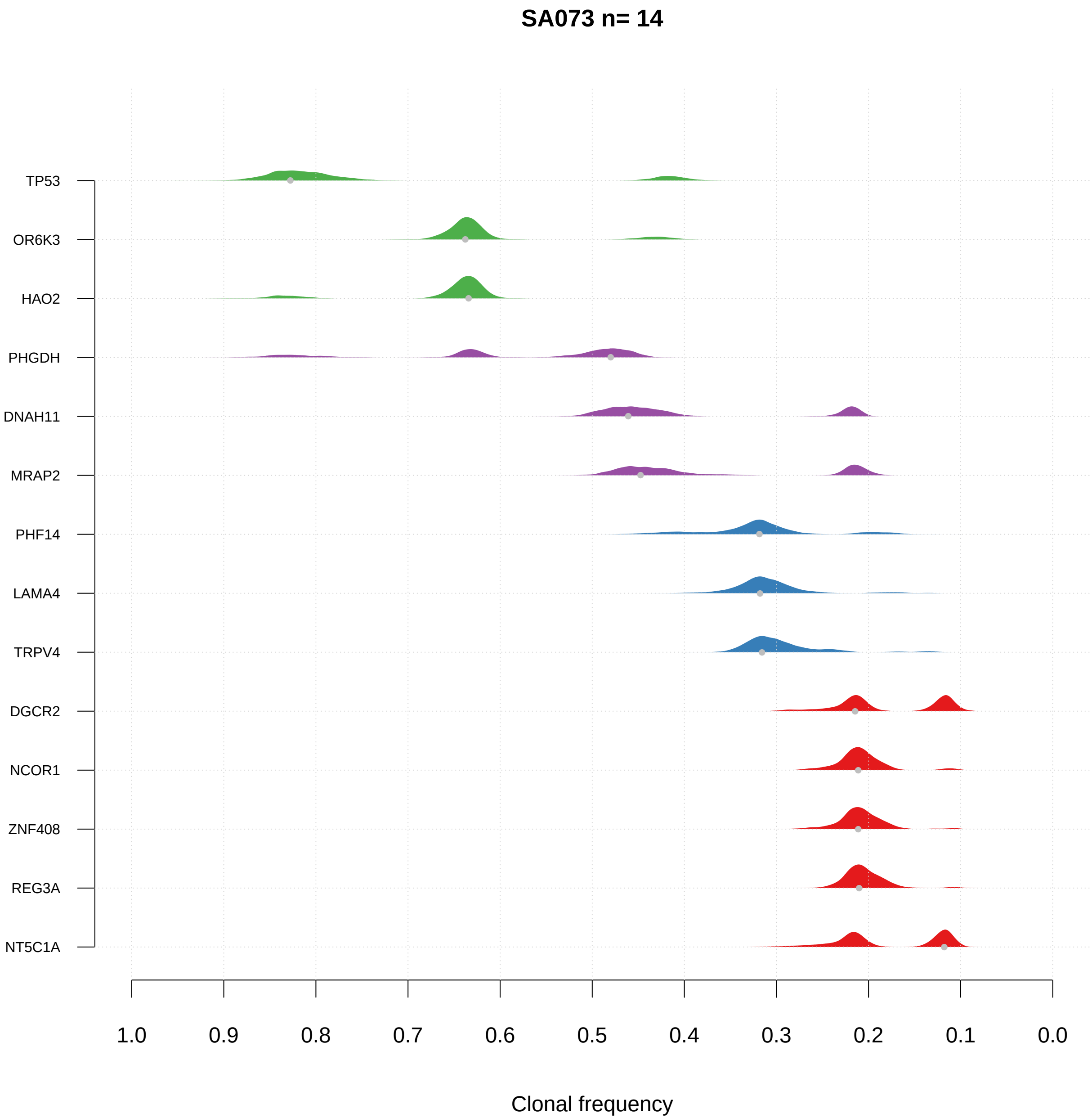
SA071 n= 96

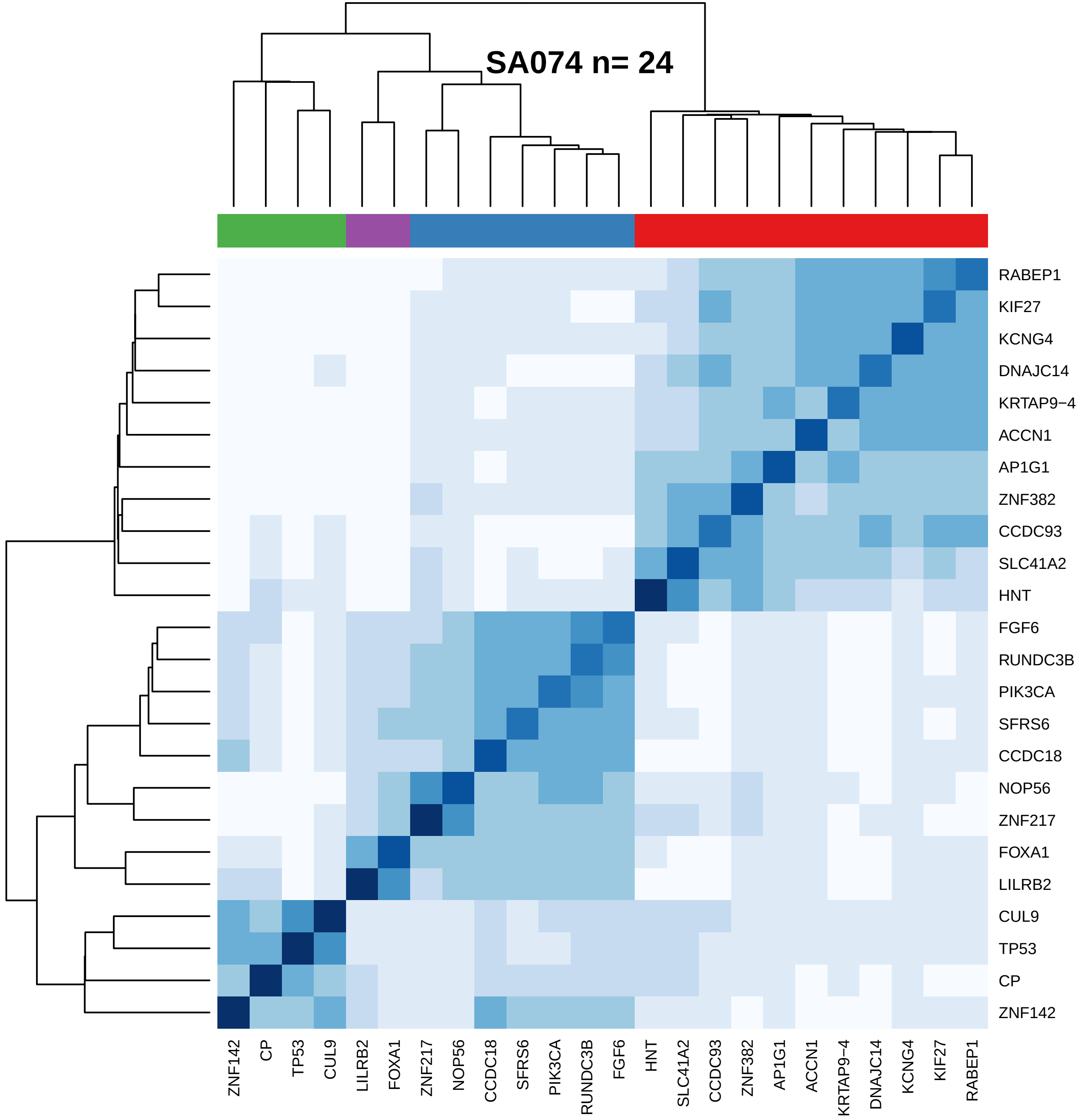

CREB3L4
F13A1
PRKCQ
ANGPTL1
SLC26A9
PGM3
MLL2
C2orf16
TRPC6
ZFYVE1
KRT9
ZNF588
FAM40B
CHFR
NETO1
C2orf70
RAB12
COL11A1
STX2
CDH10
MMP9
XRCC1
OTUD3
SLC39A7
ACO1
CCDC132
ADCY1
COL12A1
SETX
HIST1H4A
ABCC4
MYO1A
C7orf42
ZNF365
PLA2G2E
HS3ST5
ZNF729
RP11−396A22.4
TAF1B
C2orf16
ZMYND17
LAMB3
KIAA2018
TRIM45
USH2A
TRPM7
CCDC37
ARHGAP21
ISGF3G
MYH7
SLC6A19
C13orf3
NUP107
ADCY2
STAB2
NFYC
RPP40
ANAPC5
SDK1
C1orf87
SEC13L1
GLP1R
DSCAML1
AC061975.6−4
VN1R2
PLCXD2
COL6A3
TRIM71
NSD1
HSD17B2
MYST3
SLC22A10
MOV10L1
ROBO2
EPS8
TNIP1
B3GALTL
LILRA1
DNAH2
UNC5D
NDST1
DEGS2
ALKBH8
PRKCZ
ODAM
PEO1
NMS
LSM14A
FAM128A
KIAA0513
DNAH9
LTBP1
ADAM19
NELL1
SLC39A9
MUC17
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

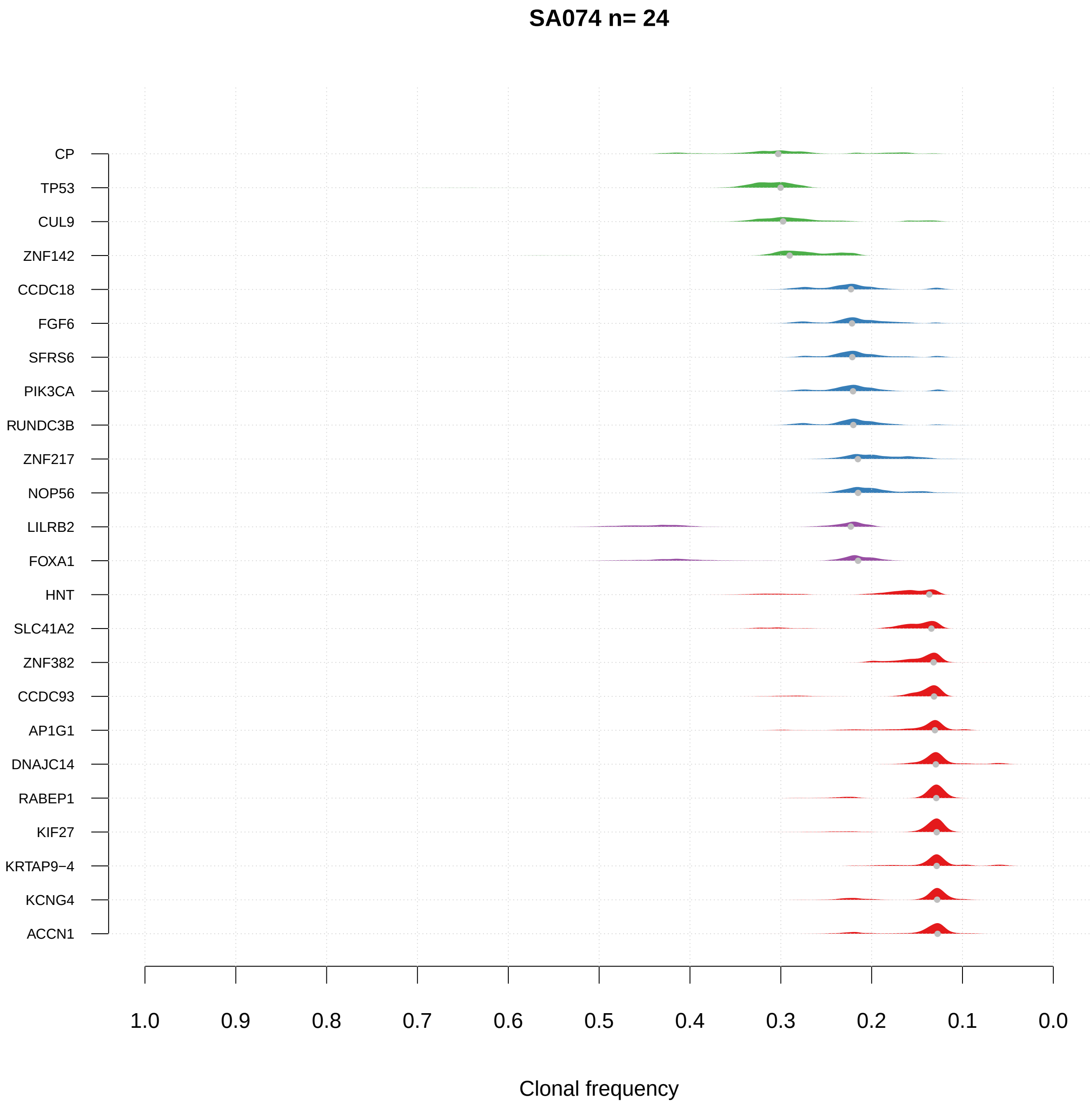

SA072 n= 22
ACE
TMCC1
UBR3
C2orf16
FAT4
GBP5
C21orf34
LAMC1
KL
F5
KIAA0196
SYNJ1
CD48
KIAA0406
KCNJ12
BRCA2
TP53
USHBP1
ABL2
TNR
TRPM3
THEG
THEG
TRPM3
TNR
ABL2
USHBP1
TP53
BRCA2
KCNJ12
KIAA0406
CD48
SYNJ1
KIAA0196
F5
KL
LAMC1
C21orf34
GBP5
FAT4
C2orf16
UBR3
TMCC1
ACE

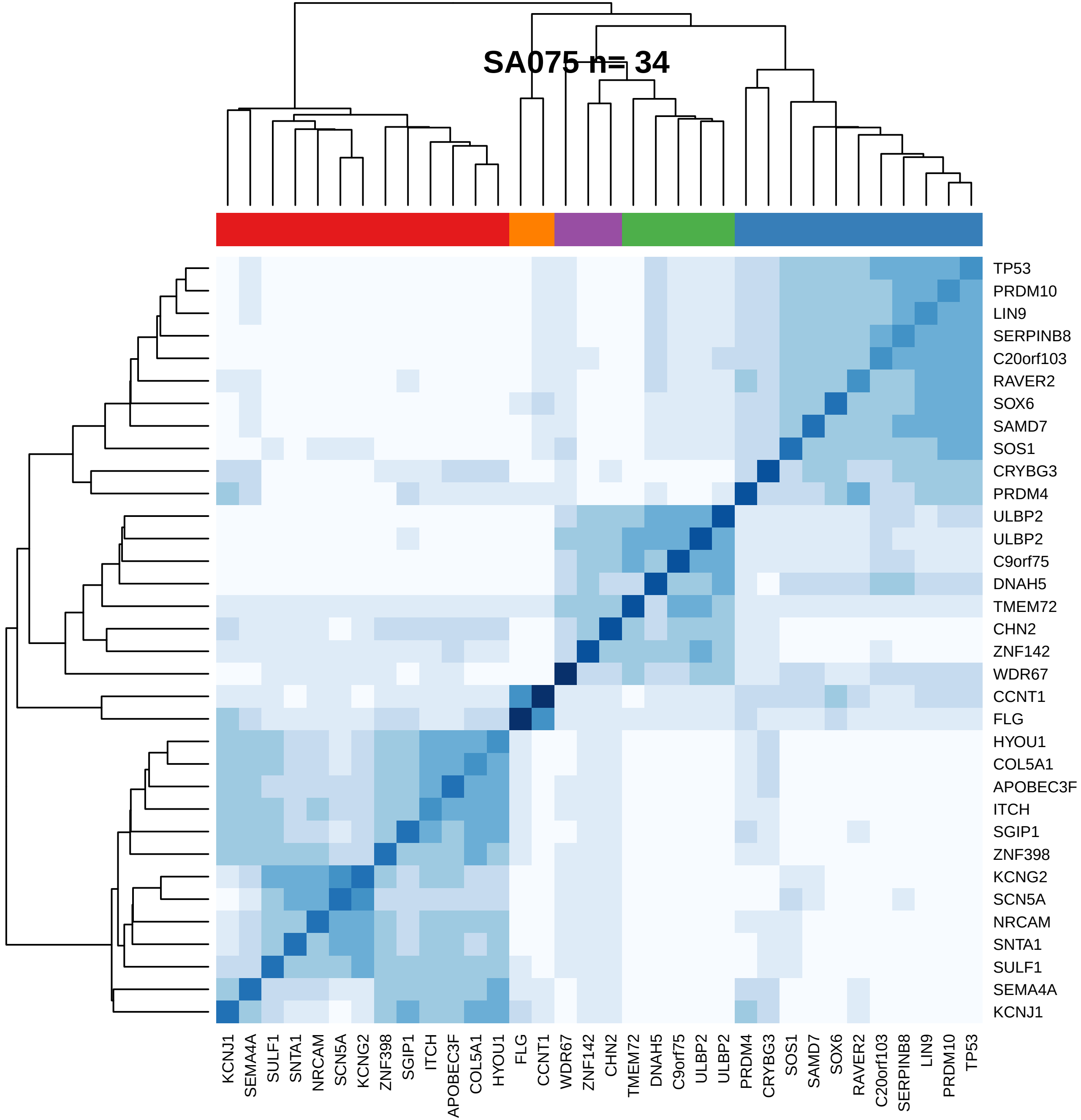

SA072 n= 22
ABL2
TNR
TRPM3
THEG
USHBP1
TP53
FAT4
UBR3
GBP5
TMCC1
ACE
C21orf34
C2orf16
BRCA2
KIAA0406
CD48
F5
KIAA0196
KL
LAMC1
SYNJ1
KCNJ12
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

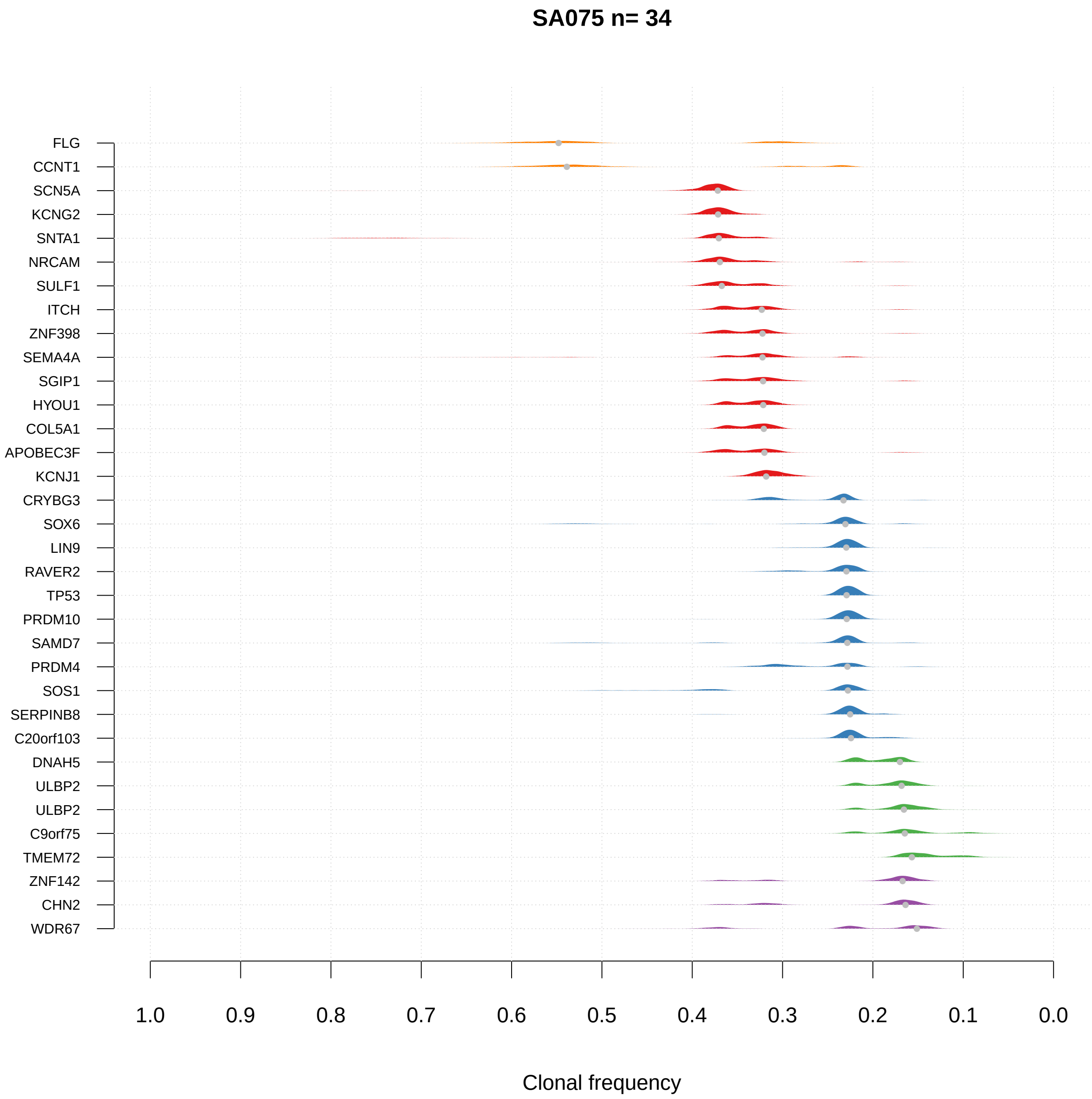

SA073 n= 14
NCOR1
ZNF408
REG3A
DGCR2
NT5C1A
MRAP2
DNAH11
PHGDH
OR6K3
HAO2
TP53
TRPV4
LAMA4
PHF14
PHF14
LAMA4
TRPV4
TP53
HAO2
OR6K3
PHGDH
DNAH11
MRAP2
NT5C1A
DGCR2
REG3A
ZNF408
NCOR1

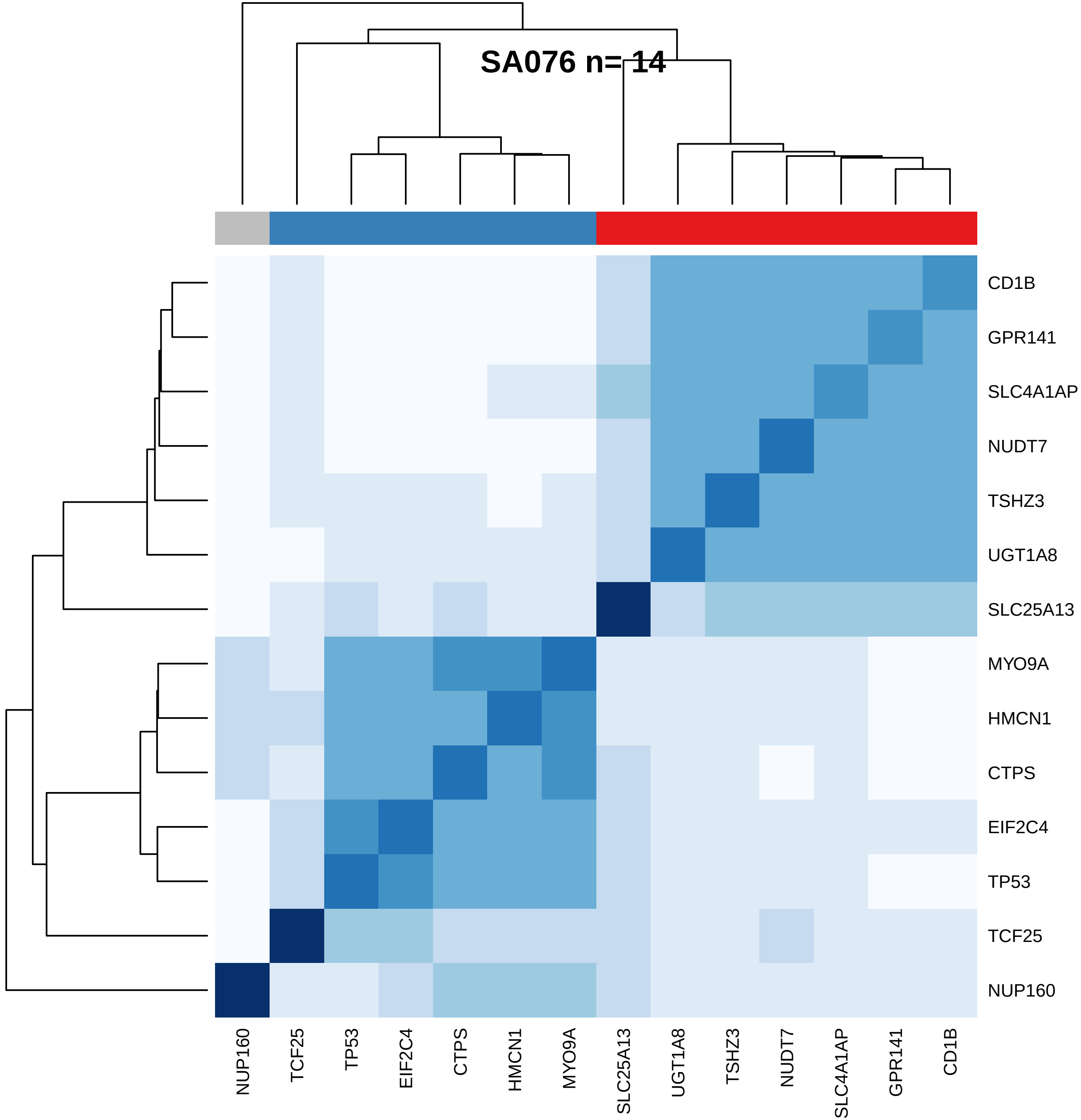

SA073 n= 14
TP53
OR6K3
HAO2
PHGDH
DNAH11
MRAP2
PHF14
LAMA4
TRPV4
DGCR2
NCOR1
ZNF408
REG3A
NT5C1A
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

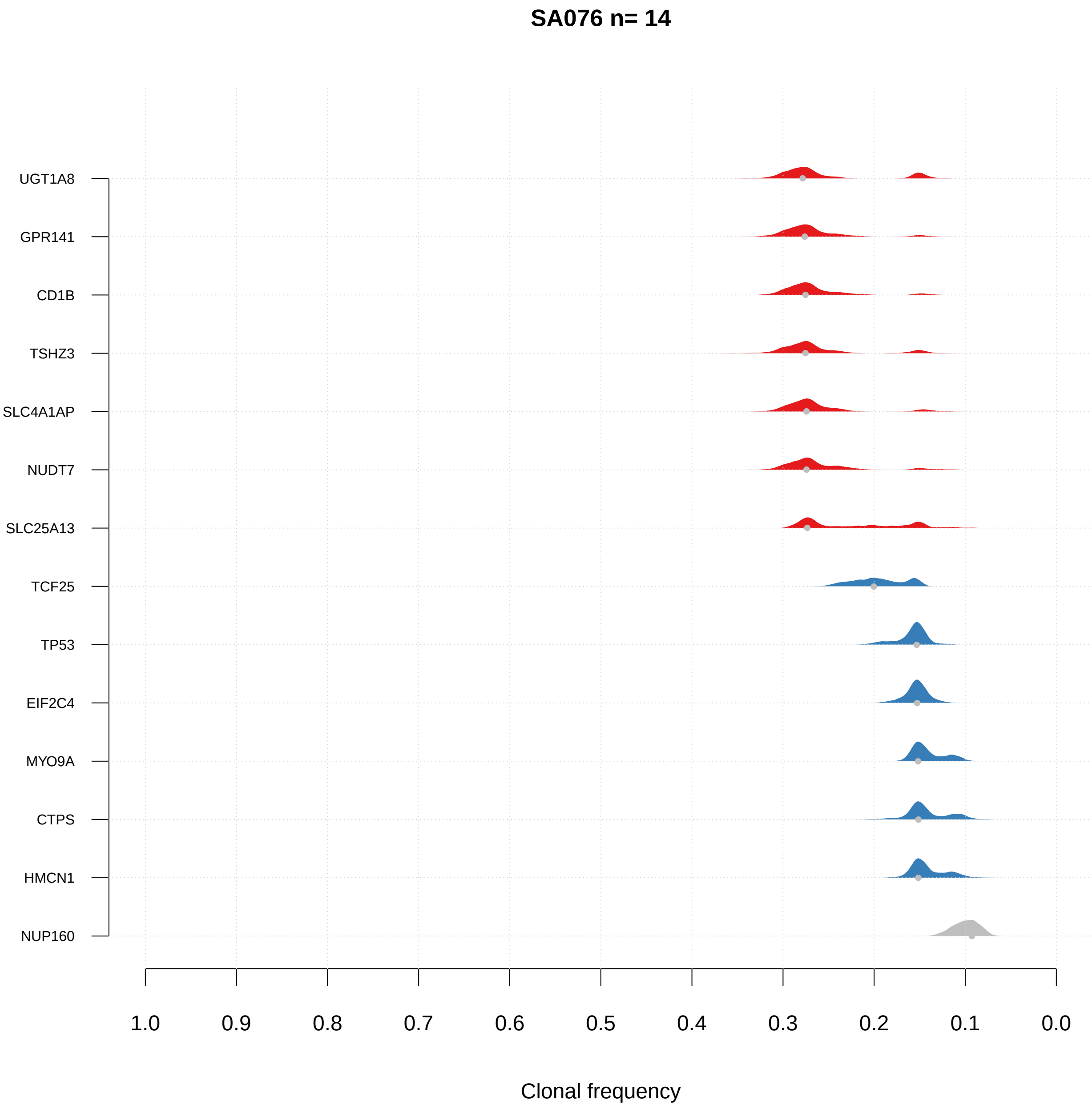

SA074 n= 24
RABEP1
KIF27
KCNG4
DNAJC14
KRTAP9−4
ACCN1
AP1G1
ZNF382
CCDC93
SLC41A2
HNT
FGF6
RUNDC3B
PIK3CA
SFRS6
CCDC18
NOP56
ZNF217
FOXA1
LILRB2
CUL9
TP53
CP
ZNF142
ZNF142
CP
TP53
CUL9
LILRB2
FOXA1
ZNF217
NOP56
CCDC18
SFRS6
PIK3CA
RUNDC3B
FGF6
HNT
SLC41A2
CCDC93
ZNF382
AP1G1
ACCN1
KRTAP9−4
DNAJC14
KCNG4
KIF27
RABEP1

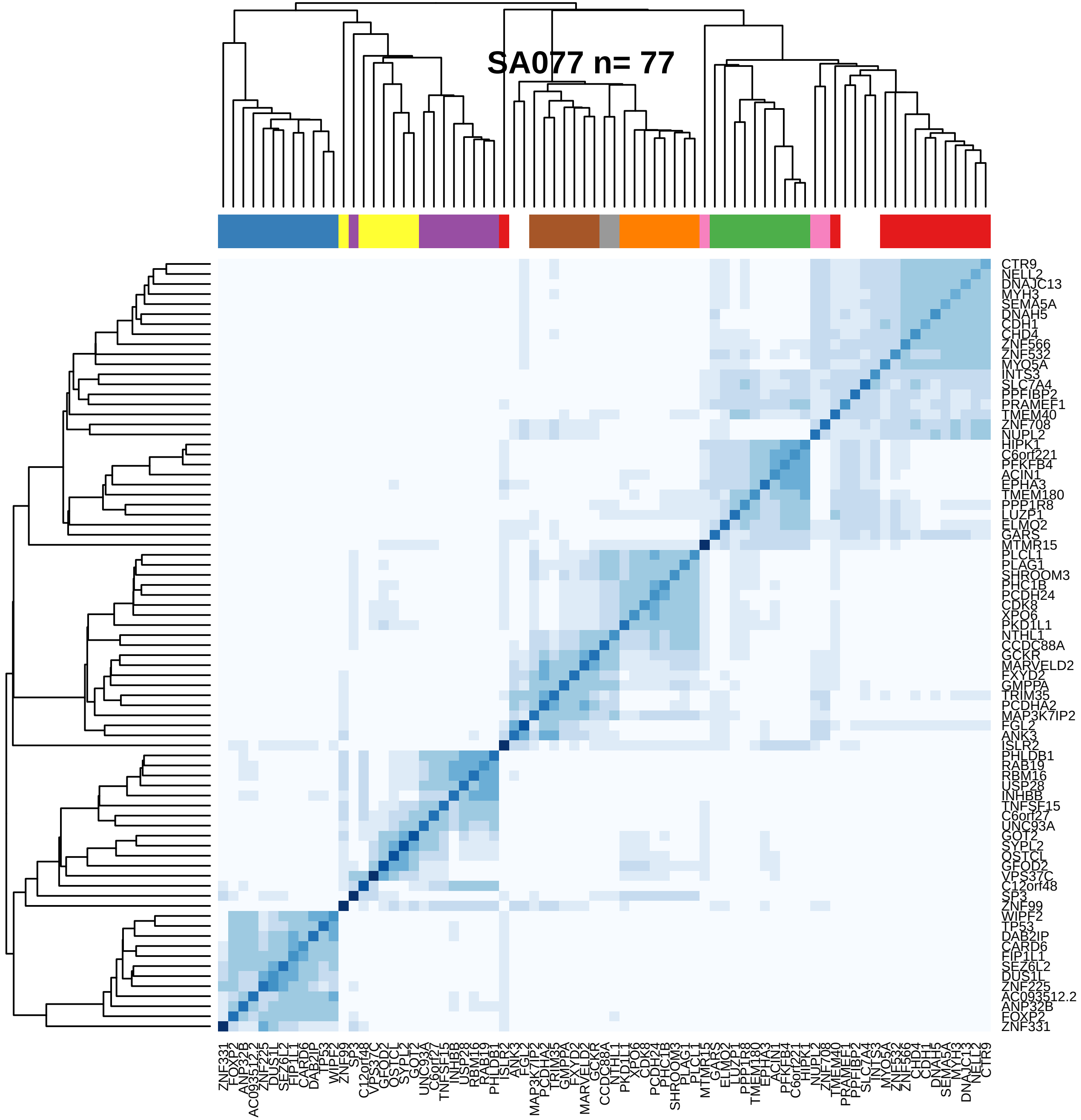

SA074 n= 24
CP
TP53
CUL9
ZNF142
CCDC18
FGF6
SFRS6
PIK3CA
RUNDC3B
ZNF217
NOP56
LILRB2
FOXA1
HNT
SLC41A2
ZNF382
CCDC93
AP1G1
DNAJC14
RABEP1
KIF27
KRTAP9−4
KCNG4
ACCN1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

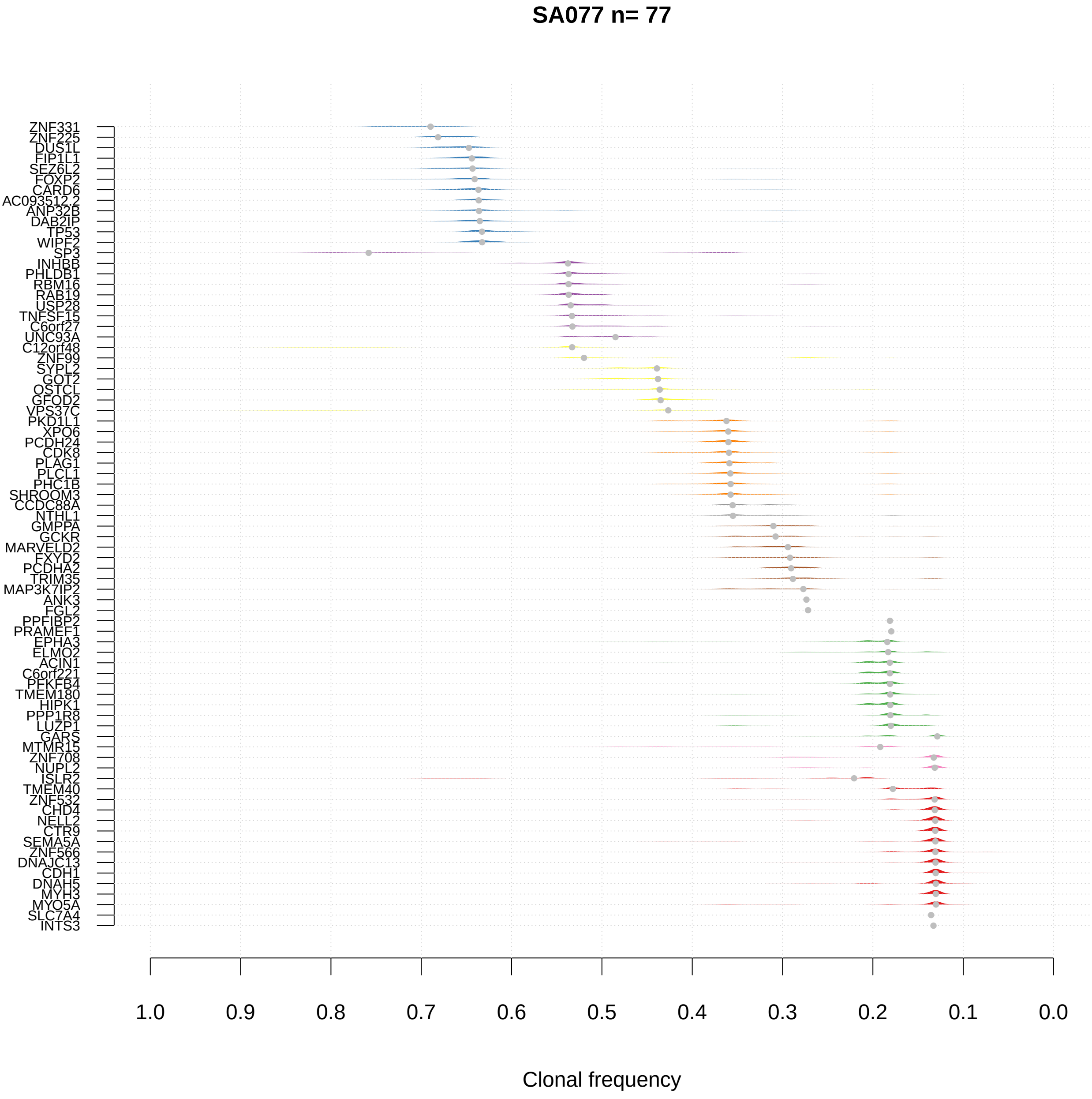

SA075 n= 34
TP53
PRDM10
LIN9
SERPINB8
C20orf103
RAVER2
SOX6
SAMD7
SOS1
CRYBG3
PRDM4
ULBP2
ULBP2
C9orf75
DNAH5
TMEM72
CHN2
ZNF142
WDR67
CCNT1
FLG
HYOU1
COL5A1
APOBEC3F
ITCH
SGIP1
ZNF398
KCNG2
SCN5A
NRCAM
SNTA1
SULF1
SEMA4A
KCNJ1
KCNJ1
SEMA4A
SULF1
SNTA1
NRCAM
SCN5A
KCNG2
ZNF398
SGIP1
ITCH
APOBEC3F
COL5A1
HYOU1
FLG
CCNT1
WDR67
ZNF142
CHN2
TMEM72
DNAH5
C9orf75
ULBP2
ULBP2
PRDM4
CRYBG3
SOS1
SAMD7
SOX6
RAVER2
C20orf103
SERPINB8
LIN9
PRDM10
TP53


SA075 n= 34
FLG
CCNT1
SCN5A
KCNG2
SNTA1
NRCAM
SULF1
ITCH
ZNF398
SEMA4A
SGIP1
HYOU1
COL5A1
APOBEC3F
KCNJ1
CRYBG3
SOX6
LIN9
RAVER2
TP53
PRDM10
SAMD7
PRDM4
SOS1
SERPINB8
C20orf103
DNAH5
ULBP2
ULBP2
C9orf75
TMEM72
ZNF142
CHN2
WDR67
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA076 n= 14
CD1B
GPR141
SLC4A1AP
NUDT7
TSHZ3
UGT1A8
SLC25A13
MYO9A
HMCN1
CTPS
EIF2C4
TP53
TCF25
NUP160
NUP160
TCF25
TP53
EIF2C4
CTPS
HMCN1
MYO9A
SLC25A13
UGT1A8
TSHZ3
NUDT7
SLC4A1AP
GPR141
CD1B

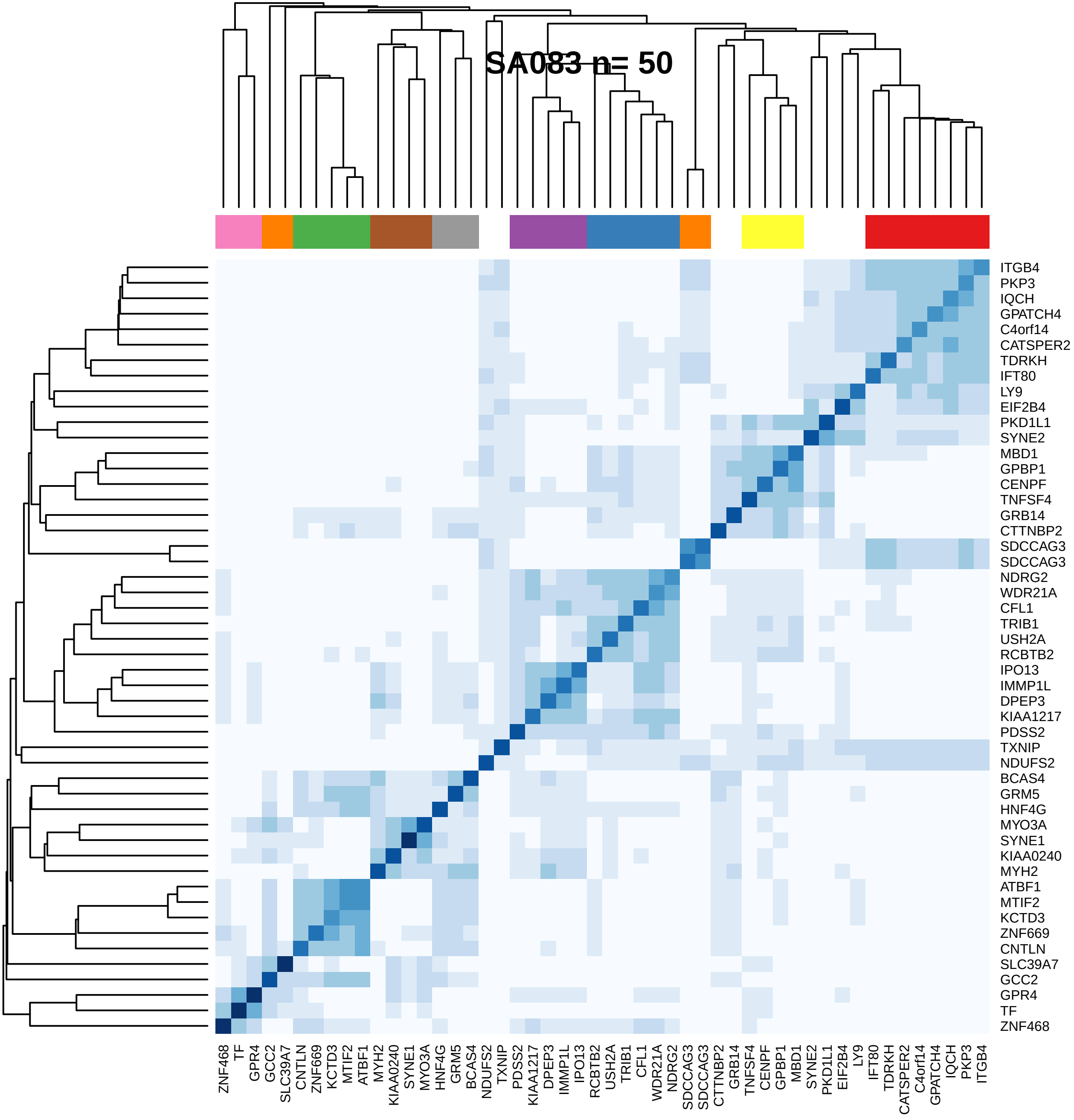

SA076 n= 14
UGT1A8
GPR141
CD1B
TSHZ3
SLC4A1AP
NUDT7
SLC25A13
TCF25
TP53
EIF2C4
MYO9A
CTPS
HMCN1
NUP160
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

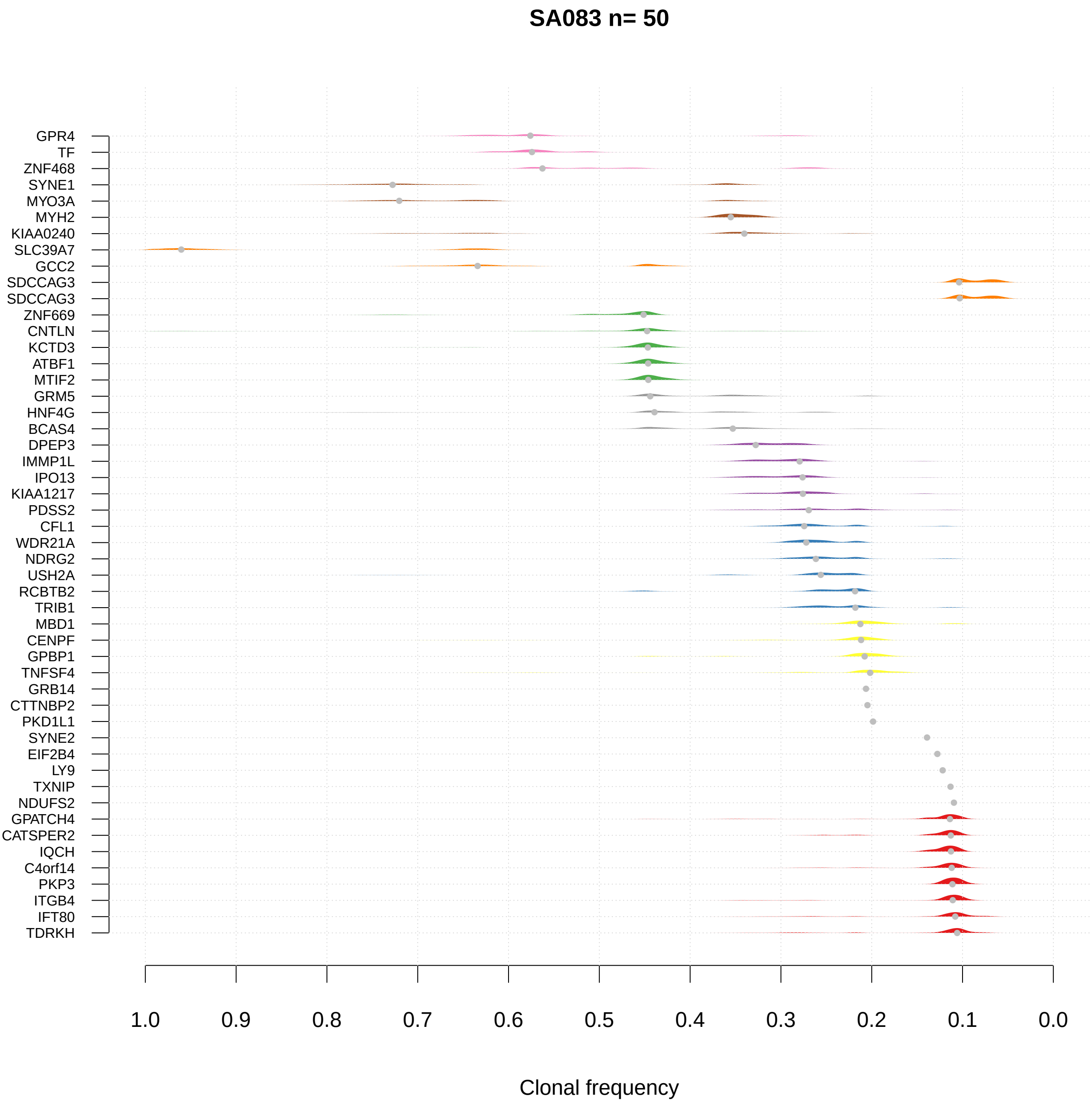

SA077 n= 77
CTR9
NELL2
DNAJC13
MYH3
SEMA5A
DNAH5
CDH1
CHD4
ZNF566
ZNF532
MYO5A
INTS3
SLC7A4
PPFIBP2
PRAMEF1
TMEM40
ZNF708
NUPL2
HIPK1
C6orf221
PFKFB4
ACIN1
EPHA3
TMEM180
PPP1R8
LUZP1
ELMO2
GARS
MTMR15
PLCL1
PLAG1
SHROOM3
PHC1B
PCDH24
CDK8
XPO6
PKD1L1
NTHL1
CCDC88A
GCKR
MARVELD2
FXYD2
GMPPA
TRIM35
PCDHA2
MAP3K7IP2
FGL2
ANK3
ISLR2
PHLDB1
RAB19
RBM16
USP28
INHBB
TNFSF15
C6orf27
UNC93A
GOT2
SYPL2
OSTCL
GFOD2
VPS37C
C12orf48
SP3
ZNF99
WIPF2
TP53
DAB2IP
CARD6
FIP1L1
SEZ6L2
DUS1L
ZNF225
AC093512.2
ANP32B
FOXP2
ZNF331
ZNF331
FOXP2
ANP32B
AC093512.2
ZNF225
DUS1L
SEZ6L2
FIP1L1
CARD6
DAB2IP
TP53
WIPF2
ZNF99
SP3
C12orf48
VPS37C
GFOD2
OSTCL
SYPL2
GOT2
UNC93A
C6orf27
TNFSF15
INHBB
USP28
RBM16
RAB19
PHLDB1
ISLR2
ANK3
FGL2
MAP3K7IP2
PCDHA2
TRIM35
GMPPA
FXYD2
MARVELD2
GCKR
CCDC88A
NTHL1
PKD1L1
XPO6
CDK8
PCDH24
PHC1B
SHROOM3
PLAG1
PLCL1
MTMR15
GARS
ELMO2
LUZP1
PPP1R8
TMEM180
EPHA3
ACIN1
PFKFB4
C6orf221
HIPK1
NUPL2
ZNF708
TMEM40
PRAMEF1
PPFIBP2
SLC7A4
INTS3
MYO5A
ZNF532
ZNF566
CHD4
CDH1
DNAH5
SEMA5A
MYH3
DNAJC13
NELL2
CTR9

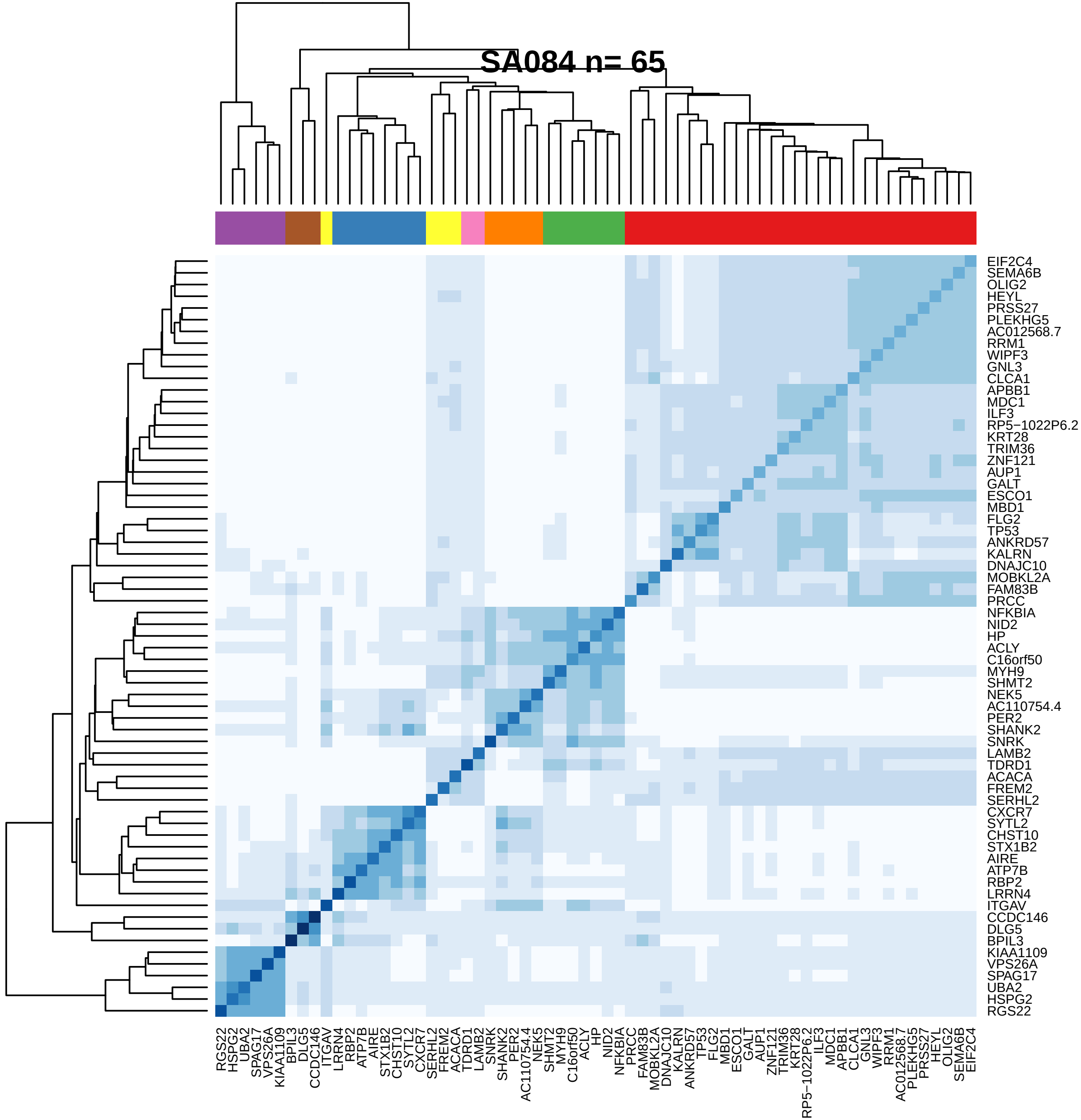


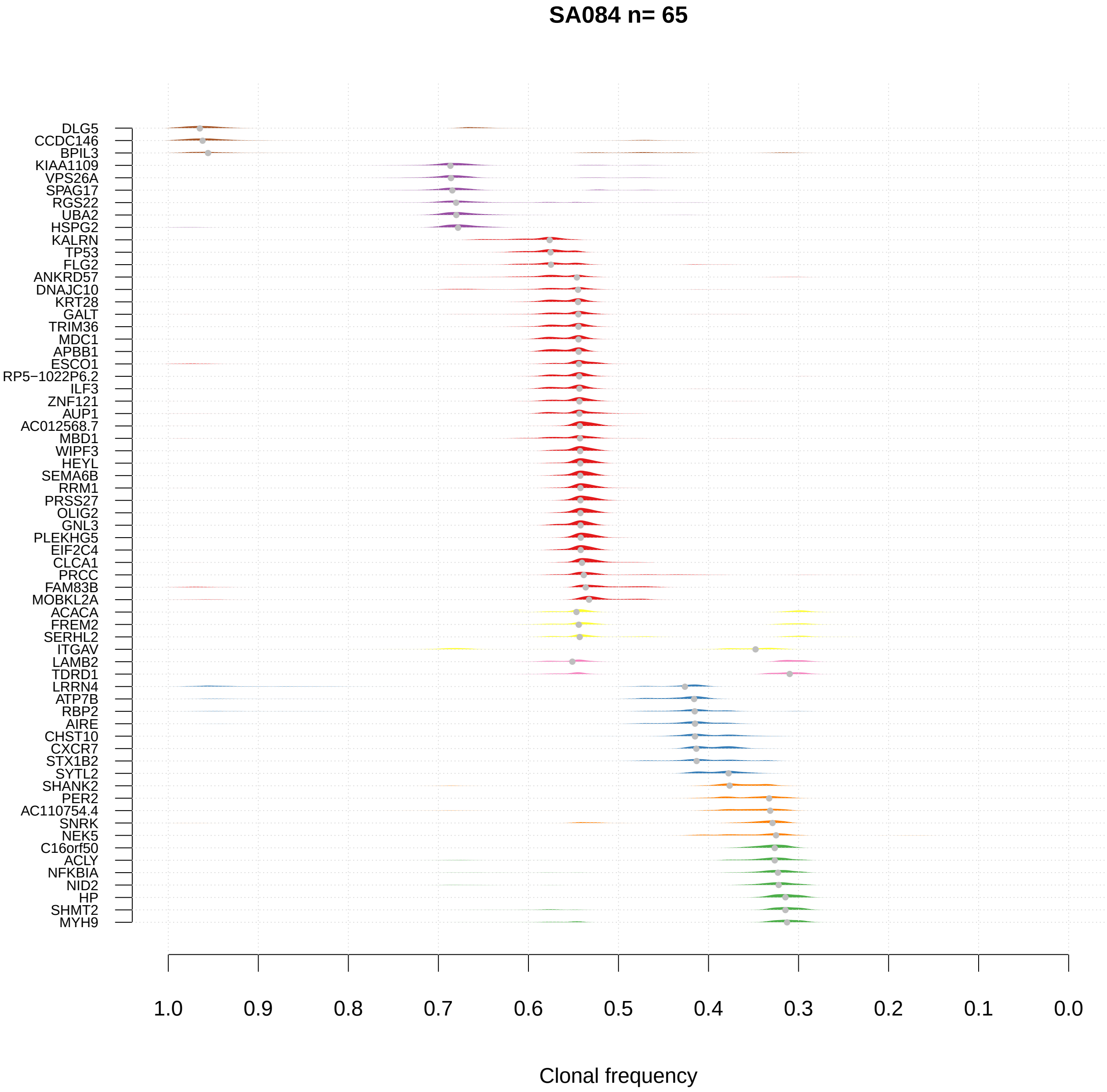
SA077 n= 77

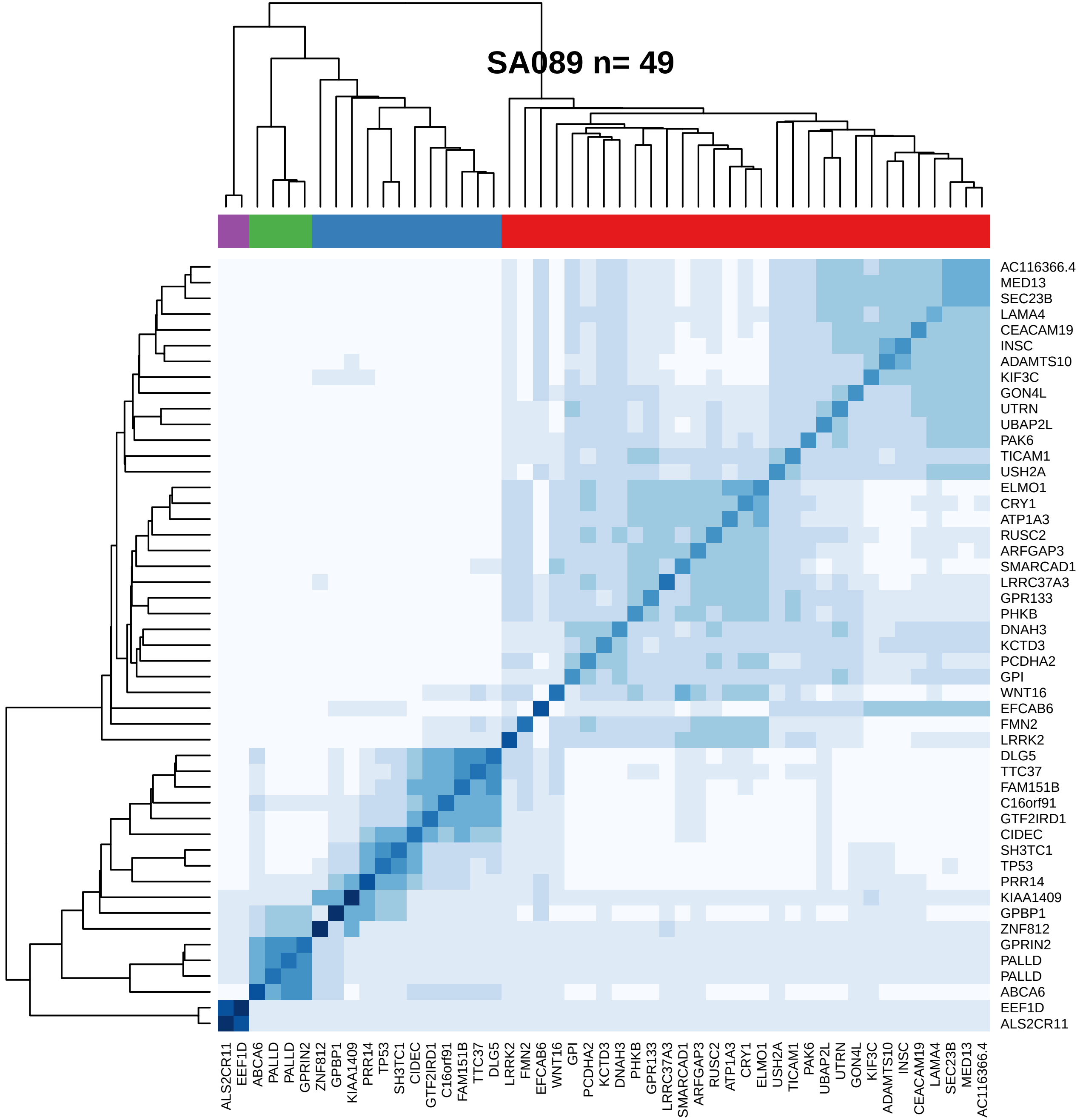
ZNF331
ZNF225
DUS1L
FIP1L1
SEZ6L2
FOXP2
CARD6
AC093512.2
ANP32B
DAB2IP
TP53
WIPF2
SP3
INHBB
PHLDB1
RBM16
RAB19
USP28
TNFSF15
C6orf27
UNC93A
C12orf48
ZNF99
SYPL2
GOT2
OSTCL
GFOD2
VPS37C
PKD1L1
XPO6
PCDH24
CDK8
PLAG1
PLCL1
PHC1B
SHROOM3
CCDC88A
NTHL1
GMPPA
GCKR
MARVELD2
FXYD2
PCDHA2
TRIM35
MAP3K7IP2
ANK3
FGL2
PPFIBP2
PRAMEF1
EPHA3
ELMO2
ACIN1
C6orf221
PFKFB4
TMEM180
HIPK1
PPP1R8
LUZP1
GARS
MTMR15
ZNF708
NUPL2
ISLR2
TMEM40
ZNF532
CHD4
NELL2
CTR9
SEMA5A
ZNF566
DNAJC13
CDH1
DNAH5
MYH3
MYO5A
SLC7A4
INTS3
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA083 n= 50
ITGB4
PKP3
IQCH
GPATCH4
C4orf14
CATSPER2
TDRKH
IFT80
LY9
EIF2B4
PKD1L1
SYNE2
MBD1
GPBP1
CENPF
TNFSF4
GRB14
CTTNBP2
SDCCAG3
SDCCAG3
NDRG2
WDR21A
CFL1
TRIB1
USH2A
RCBTB2
IPO13
IMMP1L
DPEP3
KIAA1217
PDSS2
TXNIP
NDUFS2
BCAS4
GRM5
HNF4G
MYO3A
SYNE1
KIAA0240
MYH2
ATBF1
MTIF2
KCTD3
ZNF669
CNTLN
SLC39A7
GCC2
GPR4
TF
ZNF468


SA083 n= 50
GPR4
TF
ZNF468
SYNE1
MYO3A
MYH2
KIAA0240
SLC39A7
GCC2
SDCCAG3
SDCCAG3
ZNF669
CNTLN
KCTD3
ATBF1
MTIF2
GRM5
HNF4G
BCAS4
DPEP3
IMMP1L
IPO13
KIAA1217
PDSS2
CFL1
WDR21A
NDRG2
USH2A
RCBTB2
TRIB1
MBD1
CENPF
GPBP1
TNFSF4
GRB14
CTTNBP2
PKD1L1
SYNE2
EIF2B4
LY9
TXNIP
NDUFS2
GPATCH4
CATSPER2
IQCH
C4orf14
PKP3
ITGB4
IFT80
TDRKH
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

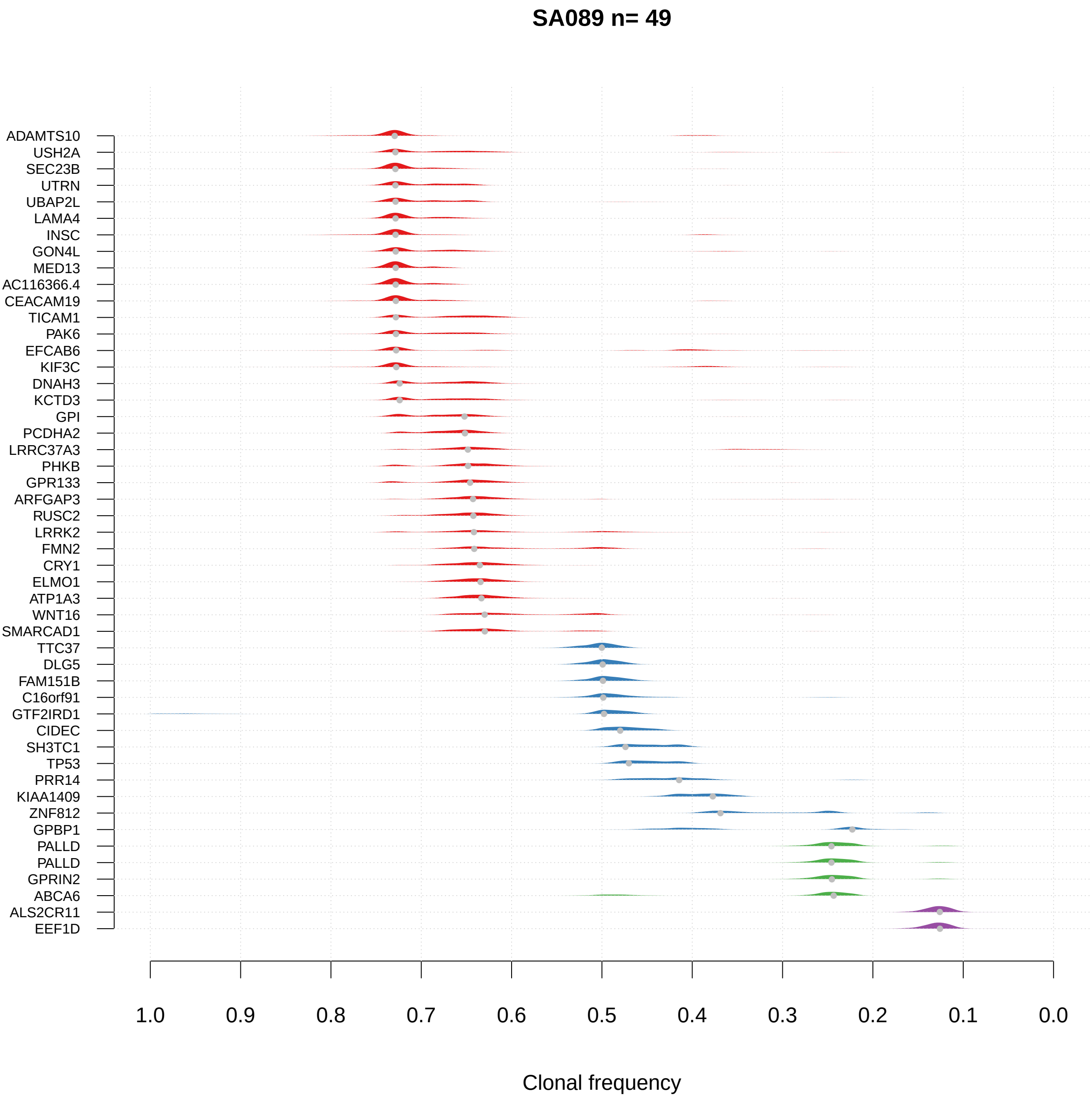

SA084 n= 65
EIF2C4
SEMA6B
OLIG2
HEYL
PRSS27
PLEKHG5
AC012568.7
RRM1
WIPF3
GNL3
CLCA1
APBB1
MDC1
ILF3
RP5−1022P6.2
KRT28
TRIM36
ZNF121
AUP1
GALT
ESCO1
MBD1
FLG2
TP53
ANKRD57
KALRN
DNAJC10
MOBKL2A
FAM83B
PRCC
NFKBIA
NID2
HP
ACLY
C16orf50
MYH9
SHMT2
NEK5
AC110754.4
PER2
SHANK2
SNRK
LAMB2
TDRD1
ACACA
FREM2
SERHL2
CXCR7
SYTL2
CHST10
STX1B2
AIRE
ATP7B
RBP2
LRRN4
ITGAV
CCDC146
DLG5
BPIL3
KIAA1109
VPS26A
SPAG17
UBA2
HSPG2
RGS22
RGS22
HSPG2
UBA2
SPAG17
VPS26A
KIAA1109
BPIL3
DLG5
CCDC146
ITGAV
LRRN4
RBP2
ATP7B
AIRE
STX1B2
CHST10
SYTL2
CXCR7
SERHL2
FREM2
ACACA
TDRD1
LAMB2
SNRK
SHANK2
PER2
AC110754.4
NEK5
SHMT2
MYH9
C16orf50
ACLY
HP
NID2
NFKBIA
PRCC
FAM83B
MOBKL2A
DNAJC10
KALRN
ANKRD57
TP53
FLG2
MBD1
ESCO1
GALT
AUP1
ZNF121
TRIM36
KRT28
RP5−1022P6.2
ILF3
MDC1
APBB1
CLCA1
GNL3
WIPF3
RRM1
AC012568.7
PLEKHG5
PRSS27
HEYL
OLIG2
SEMA6B
EIF2C4

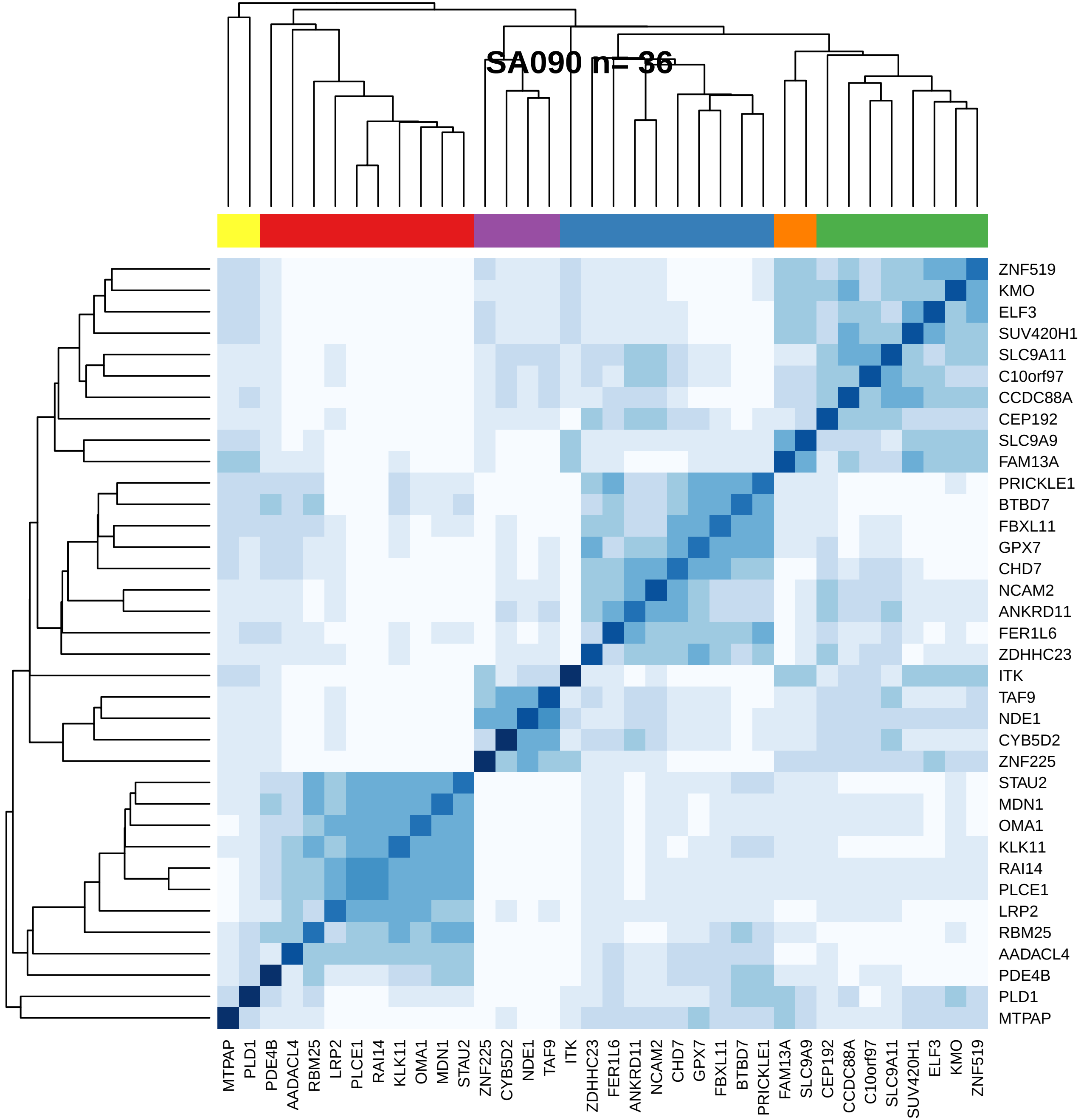

SA084 n= 65
DLG5
CCDC146
BPIL3
KIAA1109
VPS26A
SPAG17
RGS22
UBA2
HSPG2
KALRN
TP53
FLG2
ANKRD57
DNAJC10
KRT28
GALT
TRIM36
MDC1
APBB1
ESCO1
RP5−1022P6.2
ILF3
ZNF121
AUP1
AC012568.7
MBD1
WIPF3
HEYL
SEMA6B
RRM1
PRSS27
OLIG2
GNL3
PLEKHG5
EIF2C4
CLCA1
PRCC
FAM83B
MOBKL2A
ACACA
FREM2
SERHL2
ITGAV
LAMB2
TDRD1
LRRN4
ATP7B
RBP2
AIRE
CHST10
CXCR7
STX1B2
SYTL2
SHANK2
PER2
AC110754.4
SNRK
NEK5
C16orf50
ACLY
NFKBIA
NID2
HP
SHMT2
MYH9
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

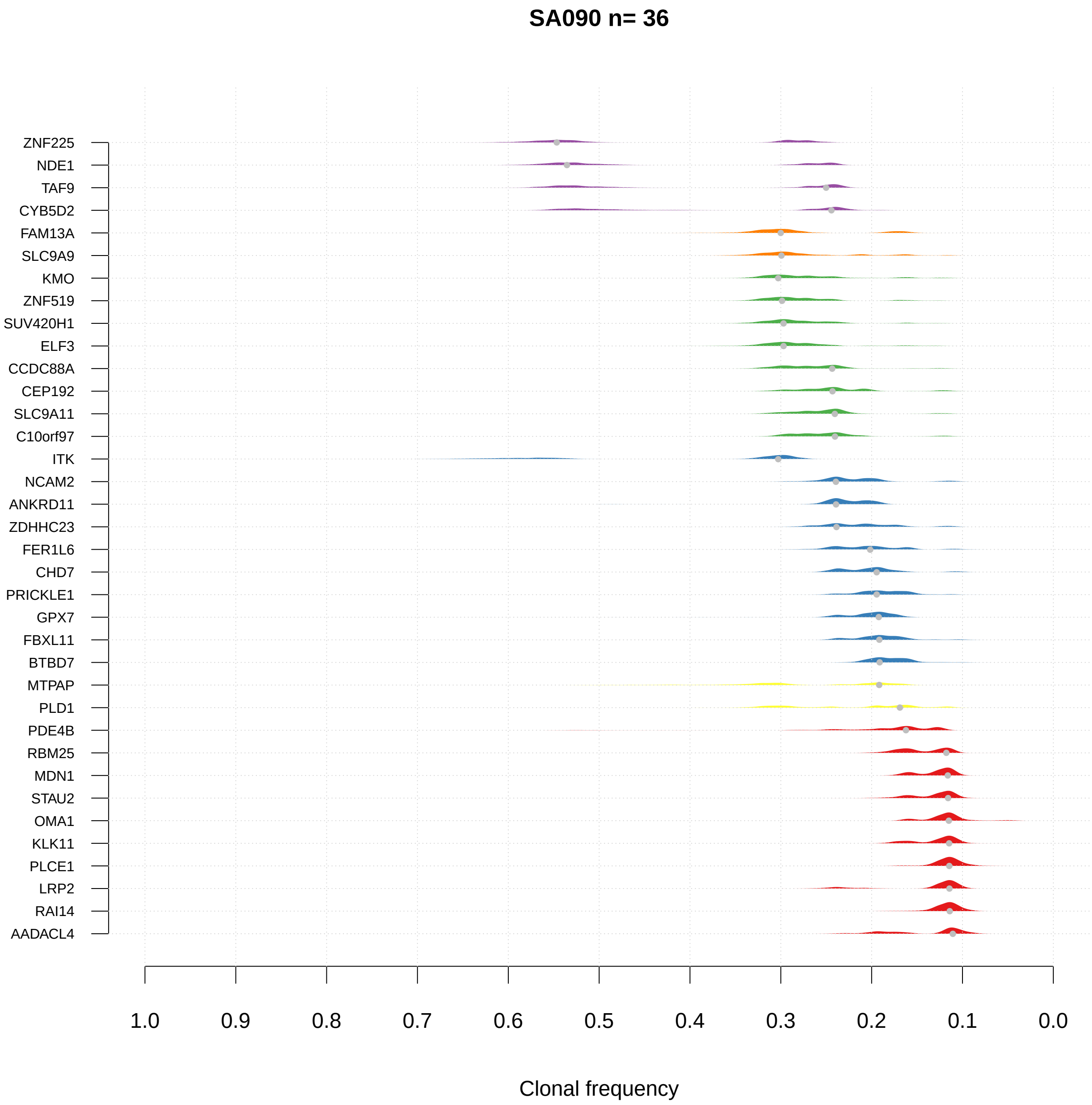


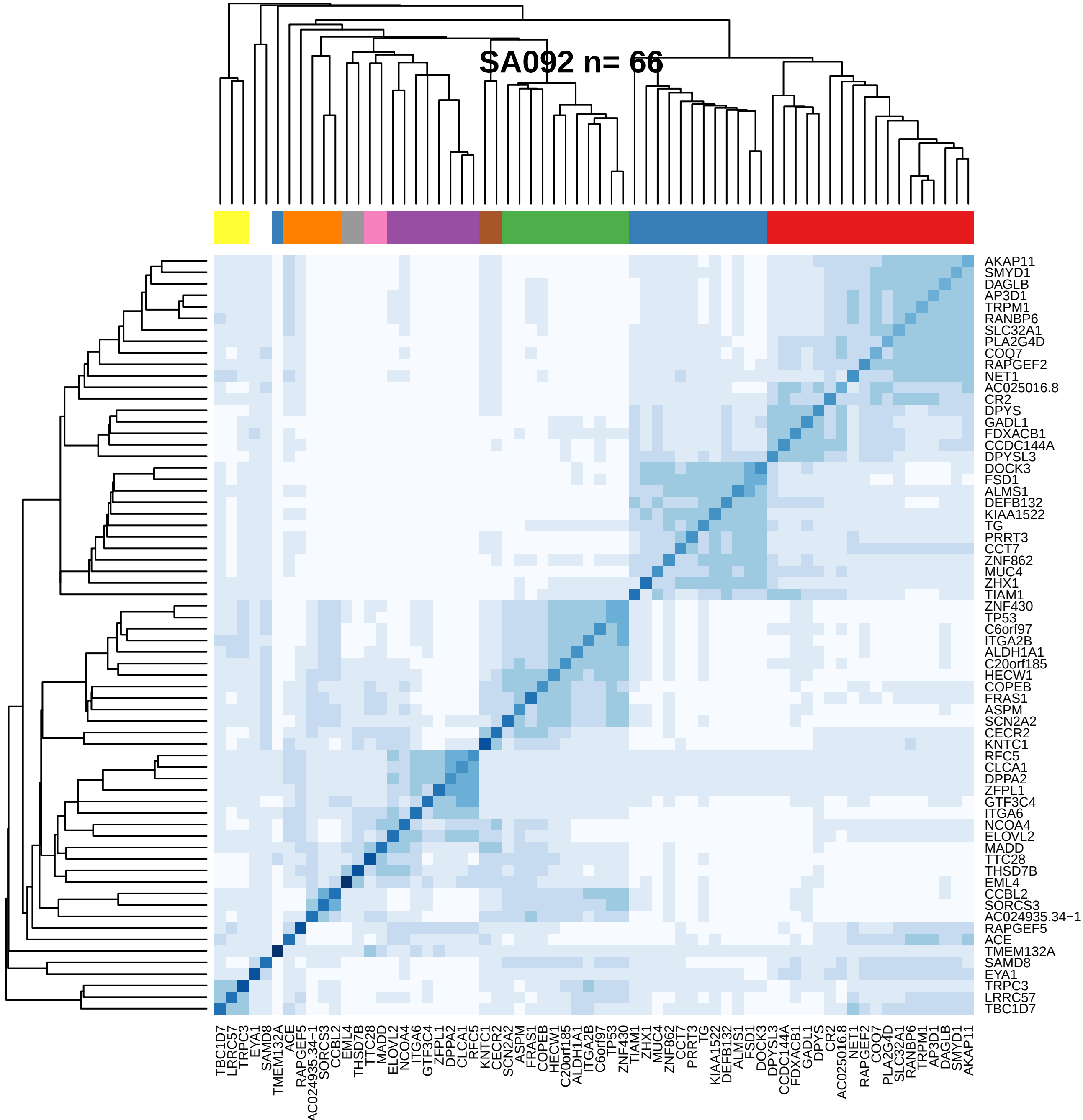
SA089 n= 49

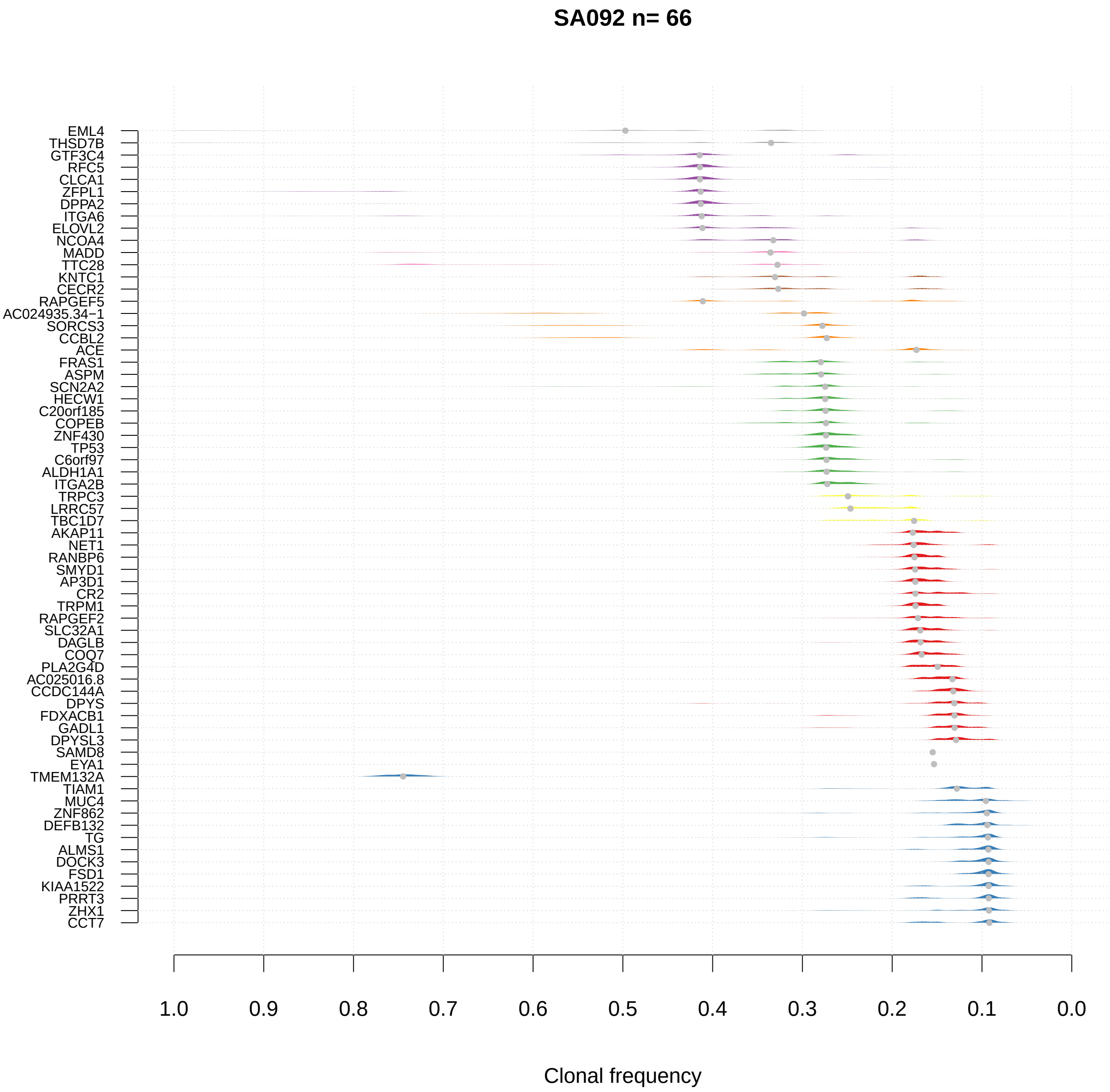
AC116366.4
MED13
SEC23B
LAMA4
CEACAM19
INSC
ADAMTS10
KIF3C
GON4L
UTRN
UBAP2L
PAK6
TICAM1
USH2A
ELMO1
CRY1
ATP1A3
RUSC2
ARFGAP3
SMARCAD1
LRRC37A3
GPR133
PHKB
DNAH3
KCTD3
PCDHA2
GPI
WNT16
EFCAB6
FMN2
LRRK2
DLG5
TTC37
FAM151B
C16orf91
GTF2IRD1
CIDEC
SH3TC1
TP53
PRR14
KIAA1409
GPBP1
ZNF812
GPRIN2
PALLD
PALLD
ABCA6
EEF1D
ALS2CR11
ALS2CR11
EEF1D
ABCA6
PALLD
PALLD
GPRIN2
ZNF812
GPBP1
KIAA1409
PRR14
TP53
SH3TC1
CIDEC
GTF2IRD1
C16orf91
FAM151B
TTC37
DLG5
LRRK2
FMN2
EFCAB6
WNT16
GPI
PCDHA2
KCTD3
DNAH3
PHKB
GPR133
LRRC37A3
SMARCAD1
ARFGAP3
RUSC2
ATP1A3
CRY1
ELMO1
USH2A
TICAM1
PAK6
UBAP2L
UTRN
GON4L
KIF3C
ADAMTS10
INSC
CEACAM19
LAMA4
SEC23B
MED13
AC116366.4

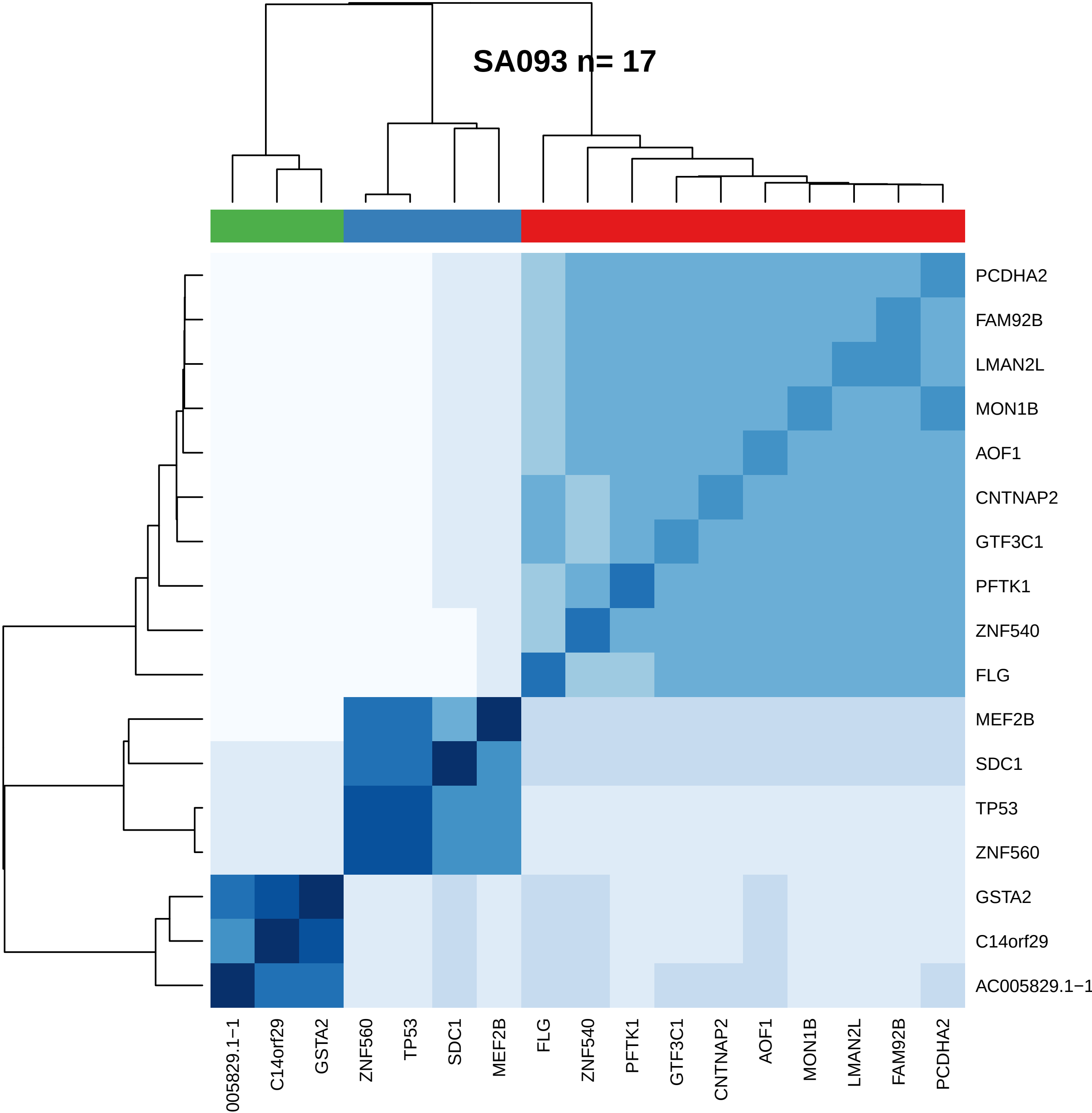

SA089 n= 49
ADAMTS10
USH2A
SEC23B
UTRN
UBAP2L
LAMA4
INSC
GON4L
MED13
AC116366.4
CEACAM19
TICAM1
PAK6
EFCAB6
KIF3C
DNAH3
KCTD3
GPI
PCDHA2
LRRC37A3
PHKB
GPR133
ARFGAP3
RUSC2
LRRK2
FMN2
CRY1
ELMO1
ATP1A3
WNT16
SMARCAD1
TTC37
DLG5
FAM151B
C16orf91
GTF2IRD1
CIDEC
SH3TC1
TP53
PRR14
KIAA1409
ZNF812
GPBP1
PALLD
PALLD
GPRIN2
ABCA6
ALS2CR11
EEF1D
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

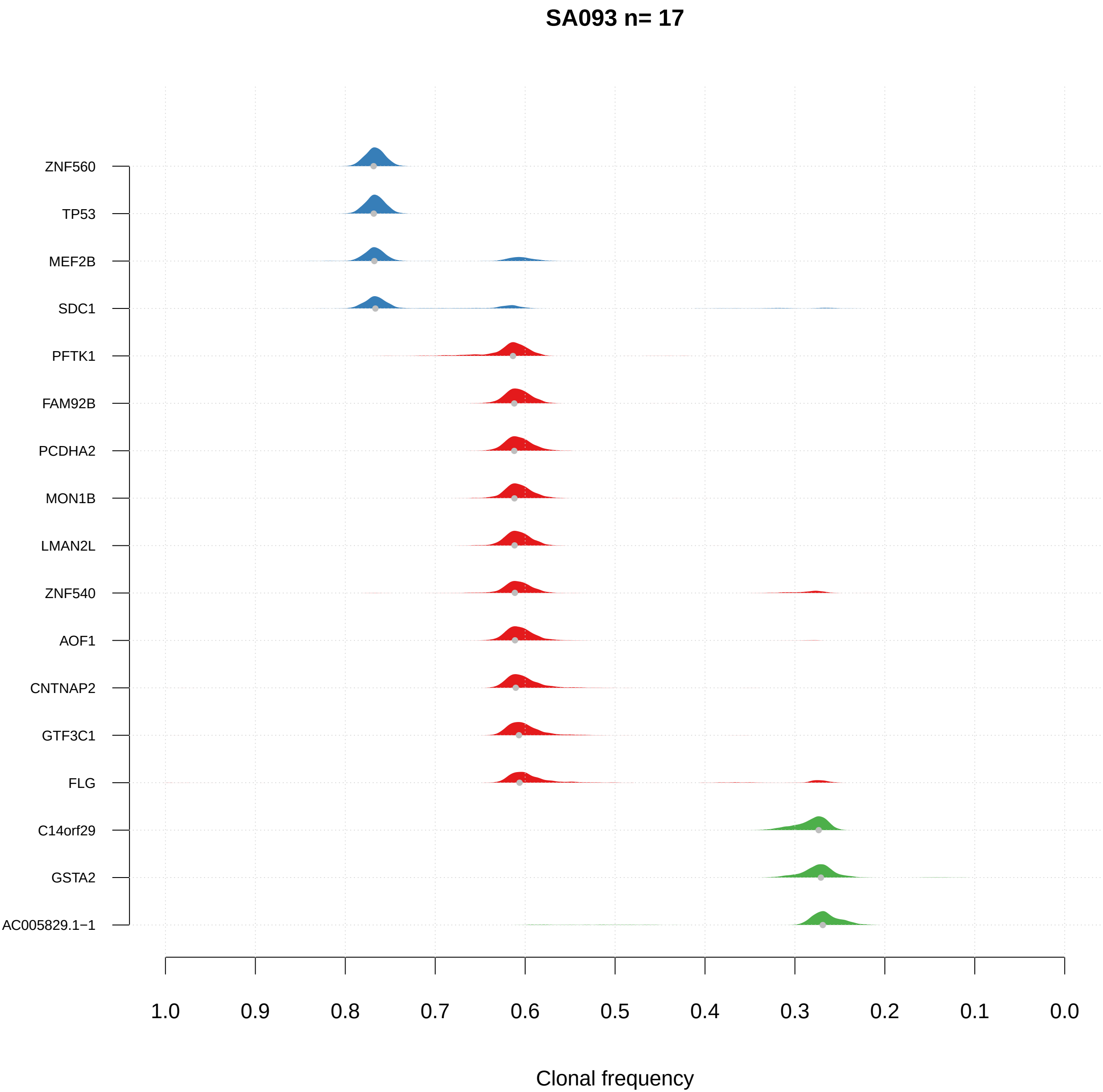

SA090 n= 36
ZNF519
KMO
ELF3
SUV420H1
SLC9A11
C10orf97
CCDC88A
CEP192
SLC9A9
FAM13A
PRICKLE1
BTBD7
FBXL11
GPX7
CHD7
NCAM2
ANKRD11
FER1L6
ZDHHC23
ITK
TAF9
NDE1
CYB5D2
ZNF225
STAU2
MDN1
OMA1
KLK11
RAI14
PLCE1
LRP2
RBM25
AADACL4
PDE4B
PLD1
MTPAP
MTPAP
PLD1
PDE4B
AADACL4
RBM25
LRP2
PLCE1
RAI14
KLK11
OMA1
MDN1
STAU2
ZNF225
CYB5D2
NDE1
TAF9
ITK
ZDHHC23
FER1L6
ANKRD11
NCAM2
CHD7
GPX7
FBXL11
BTBD7
PRICKLE1
FAM13A
SLC9A9
CEP192
CCDC88A
C10orf97
SLC9A11
SUV420H1
ELF3
KMO
ZNF519

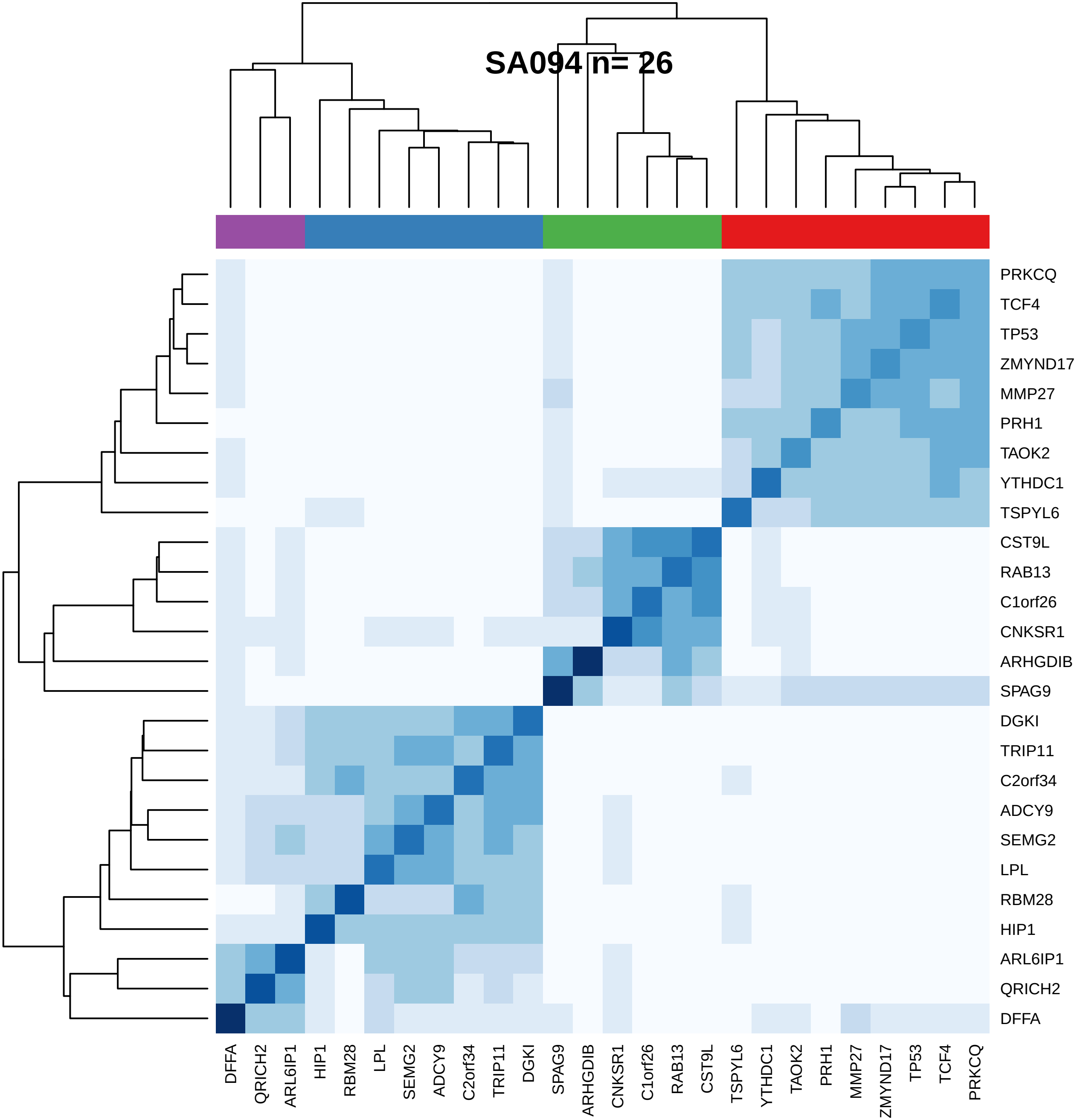

SA090 n= 36
ZNF225
NDE1
TAF9
CYB5D2
FAM13A
SLC9A9
KMO
ZNF519
SUV420H1
ELF3
CCDC88A
CEP192
SLC9A11
C10orf97
ITK
NCAM2
ANKRD11
ZDHHC23
FER1L6
CHD7
PRICKLE1
GPX7
FBXL11
BTBD7
MTPAP
PLD1
PDE4B
RBM25
MDN1
STAU2
OMA1
KLK11
PLCE1
LRP2
RAI14
AADACL4
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

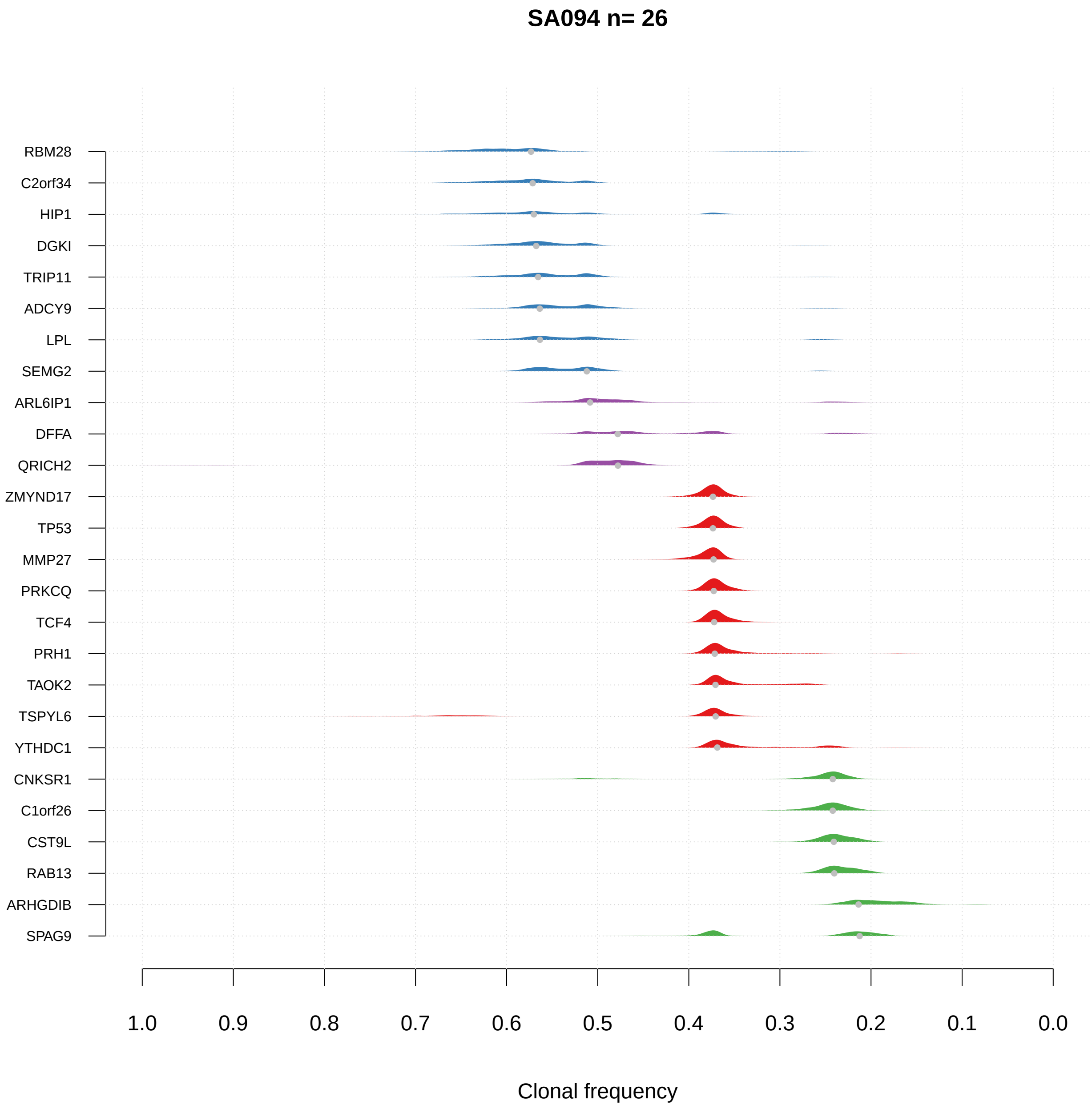

SA092 n= 66
AKAP11
SMYD1
DAGLB
AP3D1
TRPM1
RANBP6
SLC32A1
PLA2G4D
COQ7
RAPGEF2
NET1
AC025016.8
CR2
DPYS
GADL1
FDXACB1
CCDC144A
DPYSL3
DOCK3
FSD1
ALMS1
DEFB132
KIAA1522
TG
PRRT3
CCT7
ZNF862
MUC4
ZHX1
TIAM1
ZNF430
TP53
C6orf97
ITGA2B
ALDH1A1
C20orf185
HECW1
COPEB
FRAS1
ASPM
SCN2A2
CECR2
KNTC1
RFC5
CLCA1
DPPA2
ZFPL1
GTF3C4
ITGA6
NCOA4
ELOVL2
MADD
TTC28
THSD7B
EML4
CCBL2
SORCS3
AC024935.34−1
RAPGEF5
ACE
TMEM132A
SAMD8
EYA1
TRPC3
LRRC57
TBC1D7
TBC1D7
LRRC57
TRPC3
EYA1
SAMD8
TMEM132A
ACE
RAPGEF5
AC024935.34−1
SORCS3
CCBL2
EML4
THSD7B
TTC28
MADD
ELOVL2
NCOA4
ITGA6
GTF3C4
ZFPL1
DPPA2
CLCA1
RFC5
KNTC1
CECR2
SCN2A2
ASPM
FRAS1
COPEB
HECW1
C20orf185
ALDH1A1
ITGA2B
C6orf97
TP53
ZNF430
TIAM1
ZHX1
MUC4
ZNF862
CCT7
PRRT3
TG
KIAA1522
DEFB132
ALMS1
FSD1
DOCK3
DPYSL3
CCDC144A
FDXACB1
GADL1
DPYS
CR2
AC025016.8
NET1
RAPGEF2
COQ7
PLA2G4D
SLC32A1
RANBP6
TRPM1
AP3D1
DAGLB
SMYD1
AKAP11

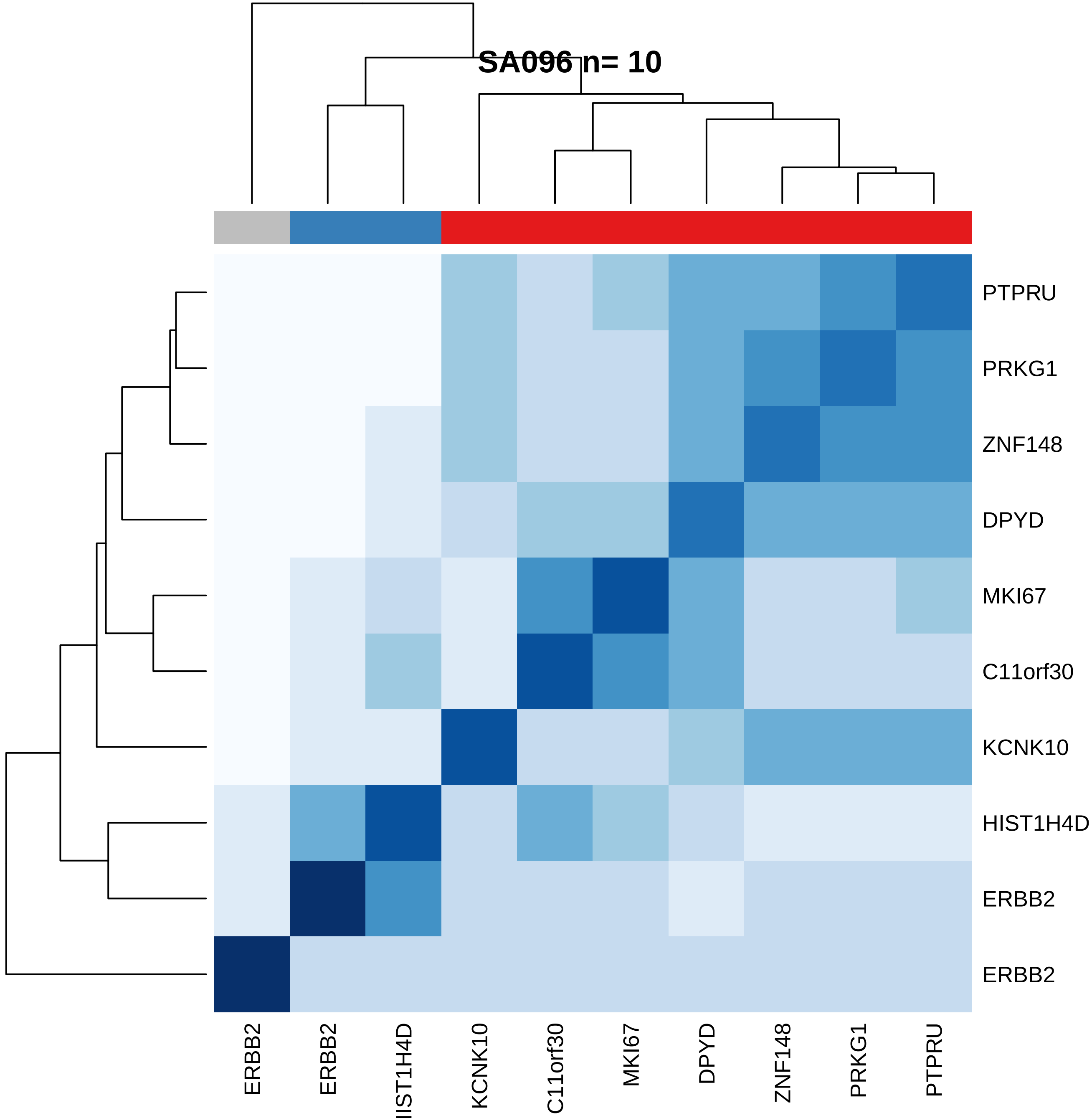

SA092 n= 66
EML4
THSD7B
GTF3C4
RFC5
CLCA1
ZFPL1
DPPA2
ITGA6
ELOVL2
NCOA4
MADD
TTC28
KNTC1
CECR2
RAPGEF5
AC024935.34−1
SORCS3
CCBL2
ACE
FRAS1
ASPM
SCN2A2
HECW1
C20orf185
COPEB
ZNF430
TP53
C6orf97
ALDH1A1
ITGA2B
TRPC3
LRRC57
TBC1D7
AKAP11
NET1
RANBP6
SMYD1
AP3D1
CR2
TRPM1
RAPGEF2
SLC32A1
DAGLB
COQ7
PLA2G4D
AC025016.8
CCDC144A
DPYS
FDXACB1
GADL1
DPYSL3
SAMD8
EYA1
TMEM132A
TIAM1
MUC4
ZNF862
DEFB132
TG
ALMS1
DOCK3
FSD1
KIAA1522
PRRT3
ZHX1
CCT7
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA093 n= 17
PCDHA2
FAM92B
LMAN2L
MON1B
AOF1
CNTNAP2
GTF3C1
PFTK1
ZNF540
FLG
MEF2B
SDC1
TP53
ZNF560
GSTA2
C14orf29
AC005829.1−1
005829.1−1
C14orf29
GSTA2
ZNF560
TP53
SDC1
MEF2B
FLG
ZNF540
PFTK1
GTF3C1
CNTNAP2
AOF1
MON1B
LMAN2L
FAM92B
PCDHA2


SA093 n= 17
ZNF560
TP53
MEF2B
SDC1
PFTK1
FAM92B
PCDHA2
MON1B
LMAN2L
ZNF540
AOF1
CNTNAP2
GTF3C1
FLG
C14orf29
GSTA2
AC005829.1−1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

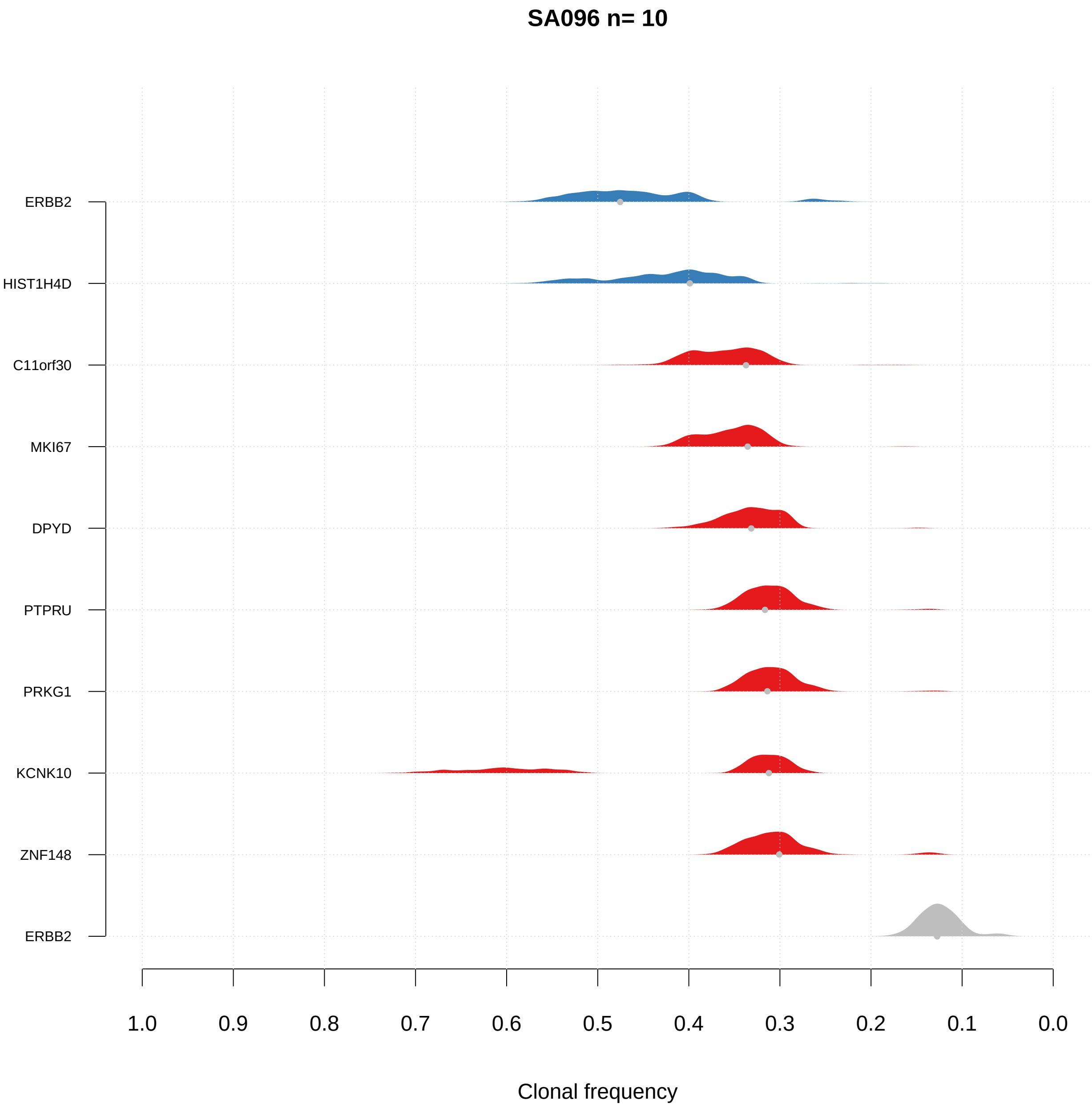

SA094 n= 26
PRKCQ
TCF4
TP53
ZMYND17
MMP27
PRH1
TAOK2
YTHDC1
TSPYL6
CST9L
RAB13
C1orf26
CNKSR1
ARHGDIB
SPAG9
DGKI
TRIP11
C2orf34
ADCY9
SEMG2
LPL
RBM28
HIP1
ARL6IP1
QRICH2
DFFA
DFFA
QRICH2
ARL6IP1
HIP1
RBM28
LPL
SEMG2
ADCY9
C2orf34
TRIP11
DGKI
SPAG9
ARHGDIB
CNKSR1
C1orf26
RAB13
CST9L
TSPYL6
YTHDC1
TAOK2
PRH1
MMP27
ZMYND17
TP53
TCF4
PRKCQ

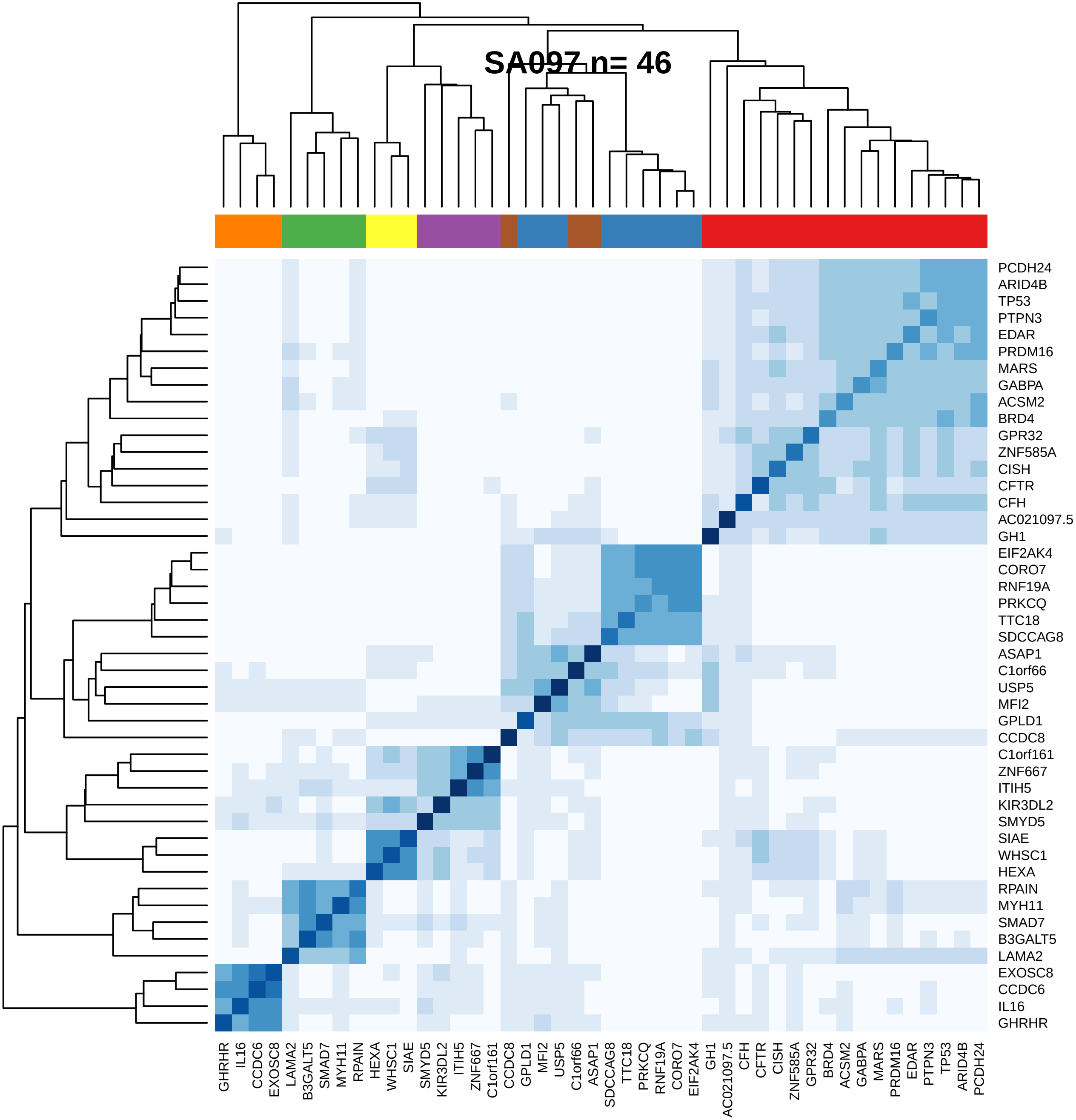

SA094 n= 26
RBM28
C2orf34
HIP1
DGKI
TRIP11
ADCY9
LPL
SEMG2
ARL6IP1
DFFA
QRICH2
ZMYND17
TP53
MMP27
PRKCQ
TCF4
PRH1
TAOK2
TSPYL6
YTHDC1
CNKSR1
C1orf26
CST9L
RAB13
ARHGDIB
SPAG9
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

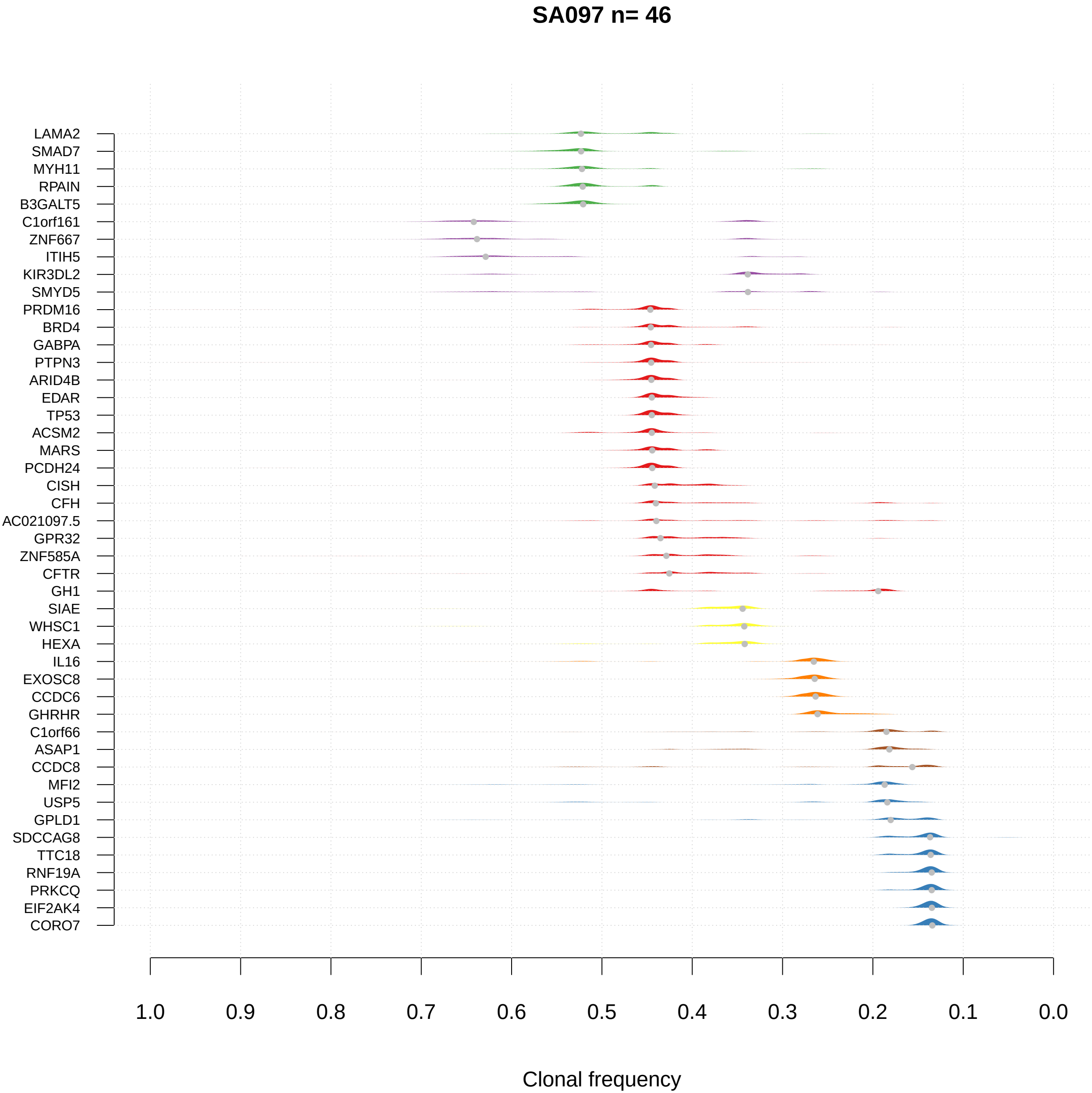


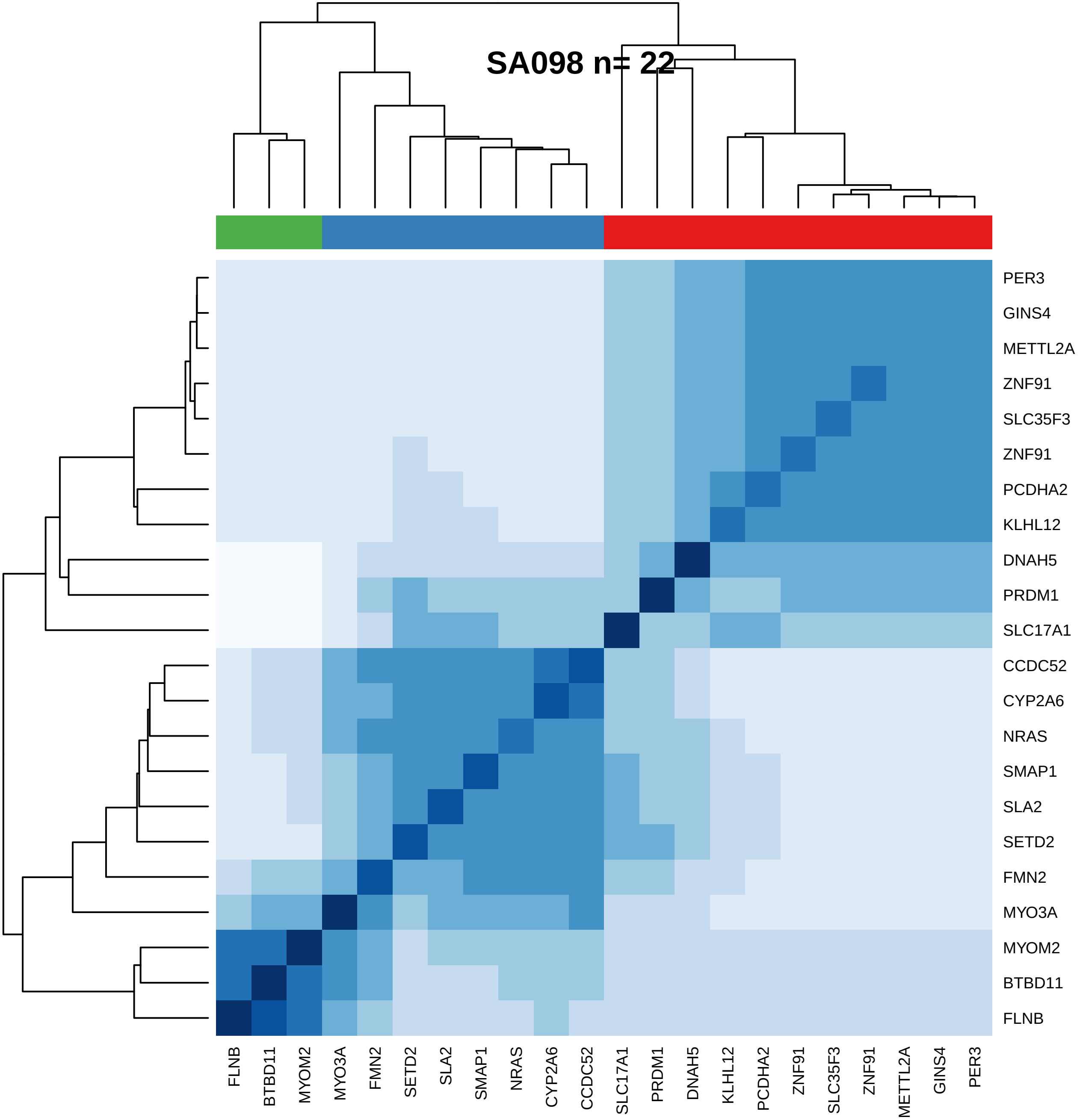
SA096 n= 10

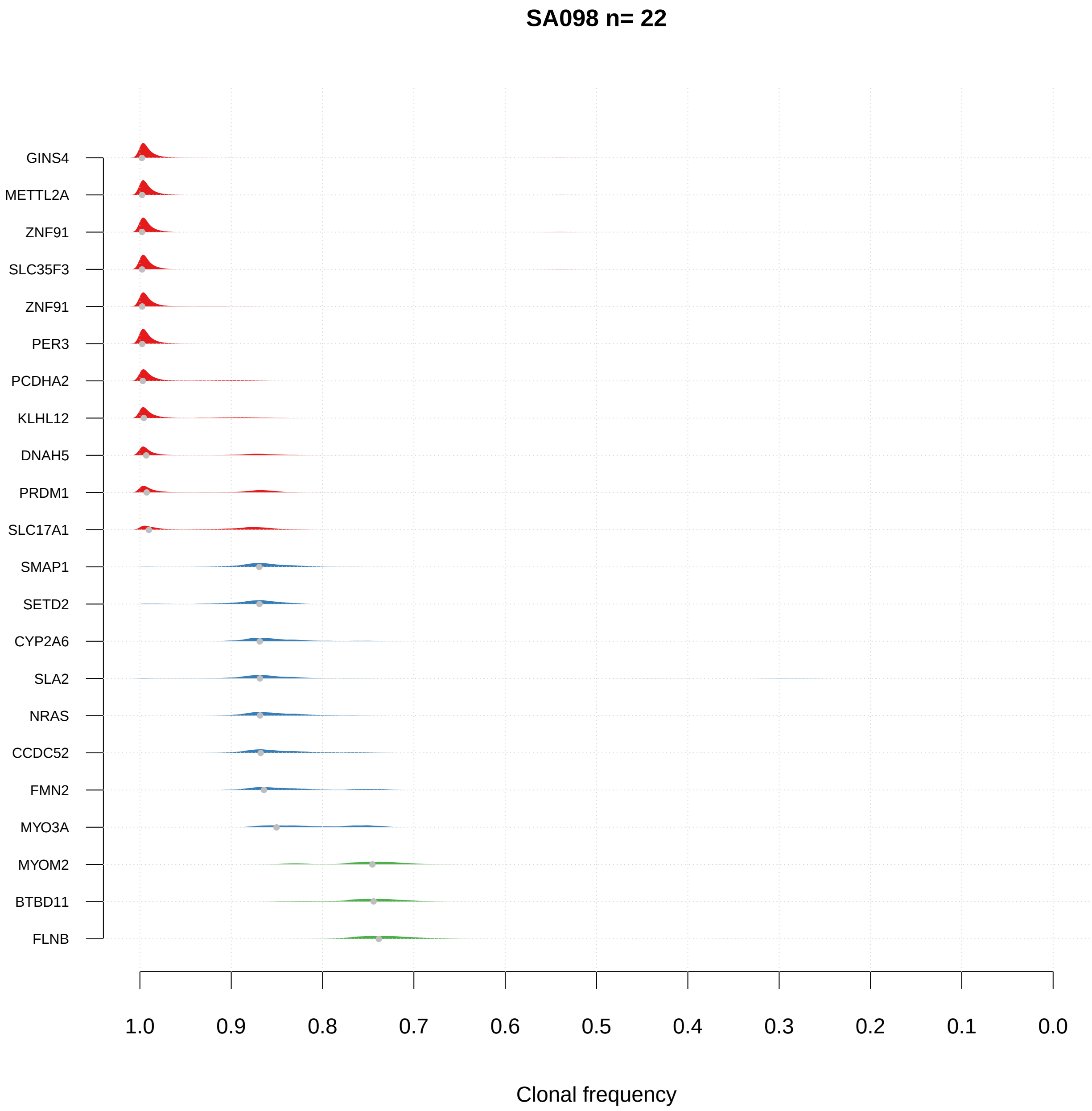
PTPRU
PRKG1
ZNF148
DPYD
MKI67
C11orf30
KCNK10
HIST1H4D
ERBB2
ERBB2
ERBB2
ERBB2
HIST1H4D
KCNK10
C11orf30
MKI67
DPYD
ZNF148
PRKG1
PTPRU

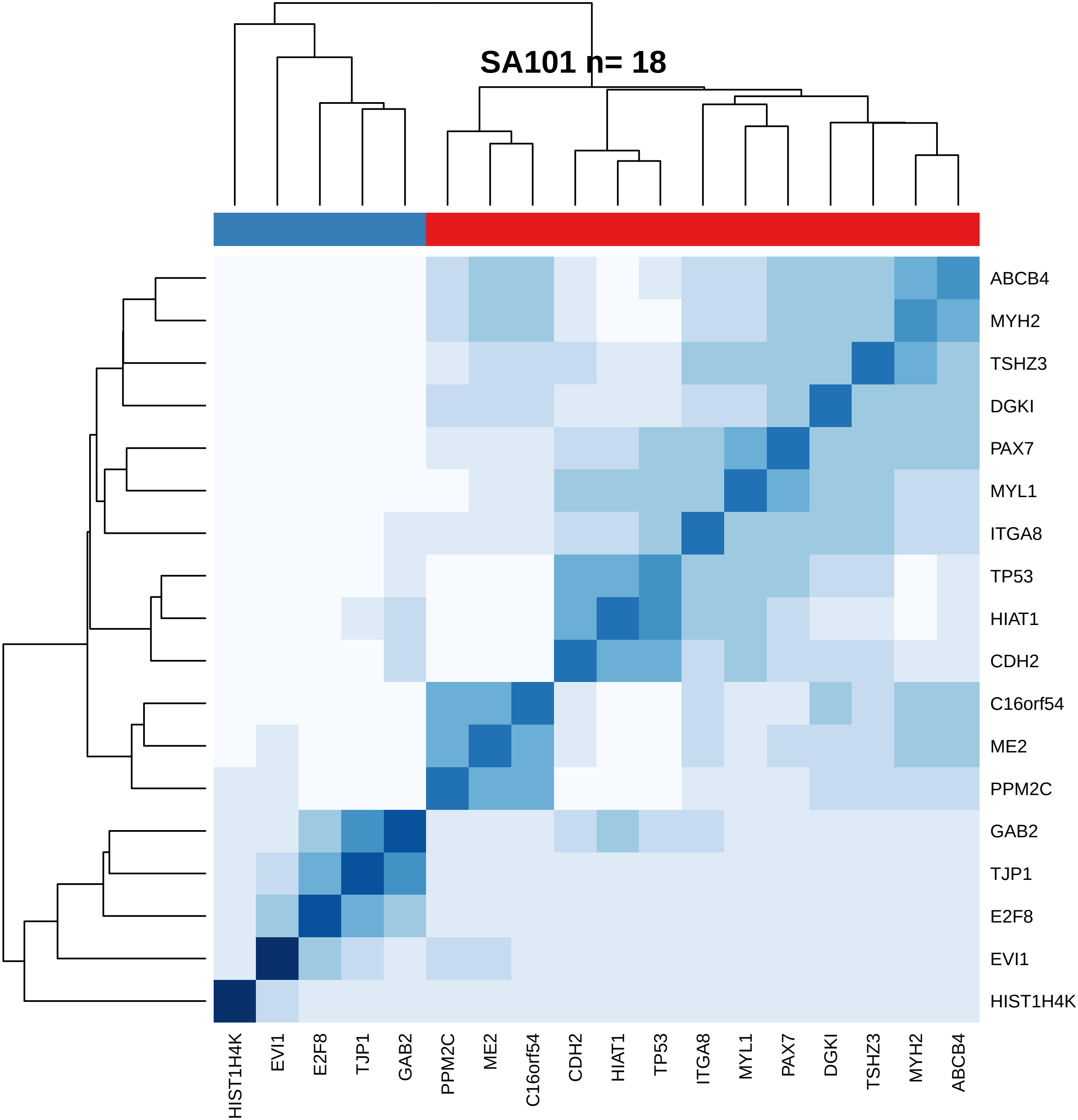

SA096 n= 10
ERBB2
HIST1H4D
C11orf30
MKI67
DPYD
PTPRU
PRKG1
KCNK10
ZNF148
ERBB2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

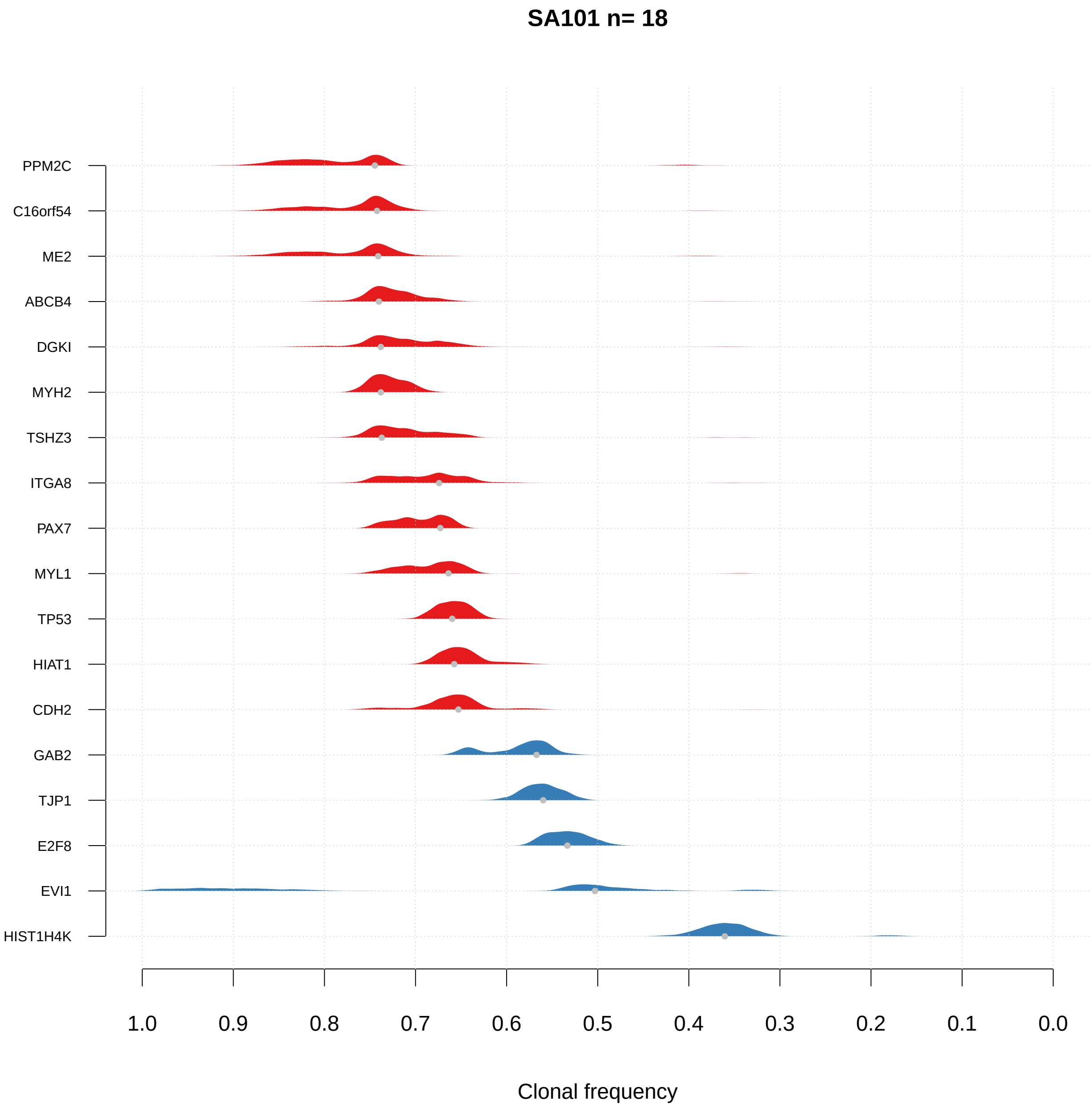

SA097 n= 46
PCDH24
ARID4B
TP53
PTPN3
EDAR
PRDM16
MARS
GABPA
ACSM2
BRD4
GPR32
ZNF585A
CISH
CFTR
CFH
AC021097.5
GH1
EIF2AK4
CORO7
RNF19A
PRKCQ
TTC18
SDCCAG8
ASAP1
C1orf66
USP5
MFI2
GPLD1
CCDC8
C1orf161
ZNF667
ITIH5
KIR3DL2
SMYD5
SIAE
WHSC1
HEXA
RPAIN
MYH11
SMAD7
B3GALT5
LAMA2
EXOSC8
CCDC6
IL16
GHRHR
GHRHR
IL16
CCDC6
EXOSC8
LAMA2
B3GALT5
SMAD7
MYH11
RPAIN
HEXA
WHSC1
SIAE
SMYD5
KIR3DL2
ITIH5
ZNF667
C1orf161
CCDC8
GPLD1
MFI2
USP5
C1orf66
ASAP1
SDCCAG8
TTC18
PRKCQ
RNF19A
CORO7
EIF2AK4
GH1
AC021097.5
CFH
CFTR
CISH
ZNF585A
GPR32
BRD4
ACSM2
GABPA
MARS
PRDM16
EDAR
PTPN3
TP53
ARID4B
PCDH24

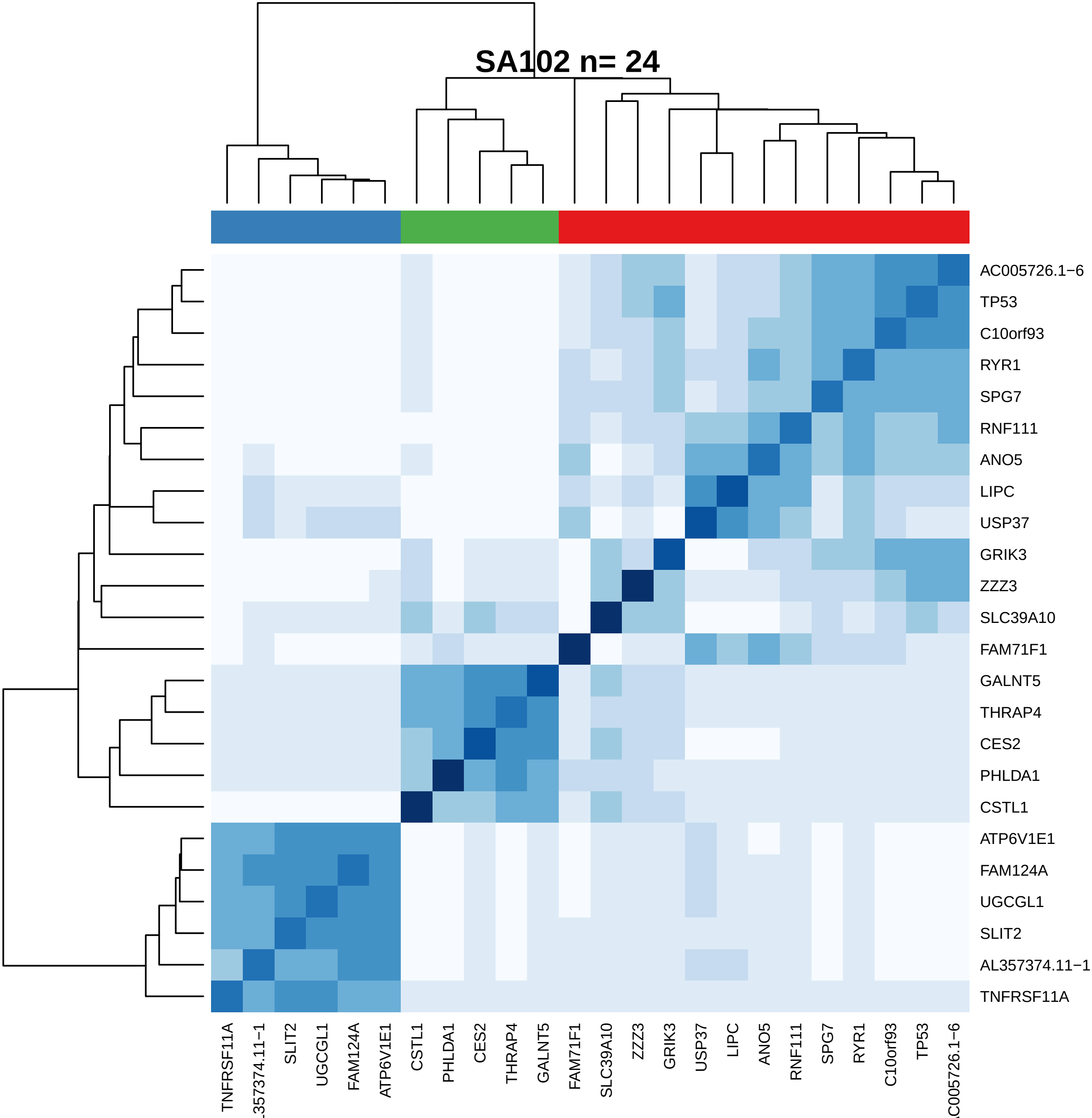

SA097 n= 46
LAMA2
SMAD7
MYH11
RPAIN
B3GALT5
C1orf161
ZNF667
ITIH5
KIR3DL2
SMYD5
PRDM16
BRD4
GABPA
PTPN3
ARID4B
EDAR
TP53
ACSM2
MARS
PCDH24
CISH
CFH
AC021097.5
GPR32
ZNF585A
CFTR
GH1
SIAE
WHSC1
HEXA
IL16
EXOSC8
CCDC6
GHRHR
C1orf66
ASAP1
CCDC8
MFI2
USP5
GPLD1
SDCCAG8
TTC18
RNF19A
PRKCQ
EIF2AK4
CORO7
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

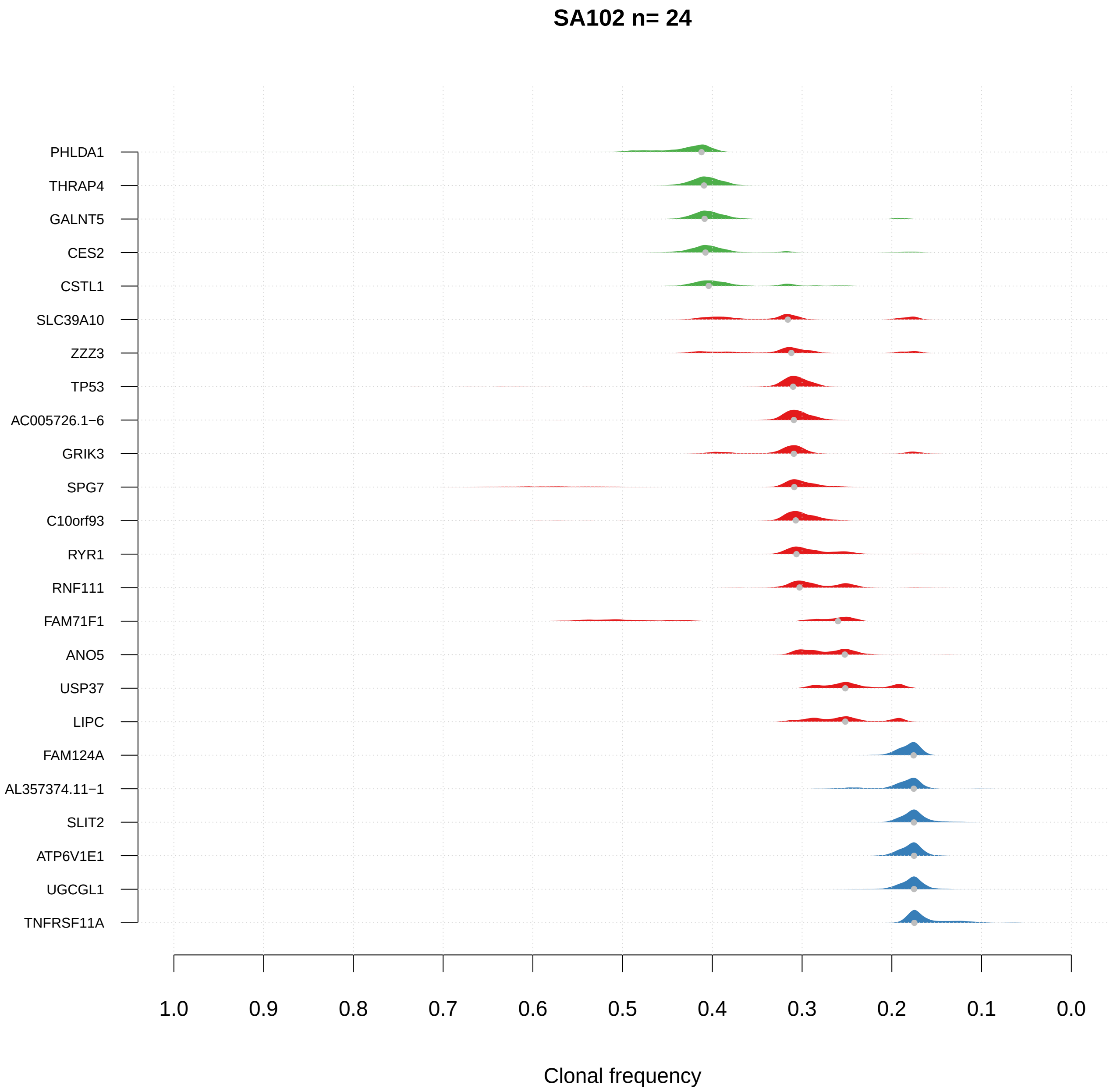

SA098 n= 22
PER3
GINS4
METTL2A
ZNF91
SLC35F3
ZNF91
PCDHA2
KLHL12
DNAH5
PRDM1
SLC17A1
CCDC52
CYP2A6
NRAS
SMAP1
SLA2
SETD2
FMN2
MYO3A
MYOM2
BTBD11
FLNB
FLNB
BTBD11
MYOM2
MYO3A
FMN2
SETD2
SLA2
SMAP1
NRAS
CYP2A6
CCDC52
SLC17A1
PRDM1
DNAH5
KLHL12
PCDHA2
ZNF91
SLC35F3
ZNF91
METTL2A
GINS4
PER3

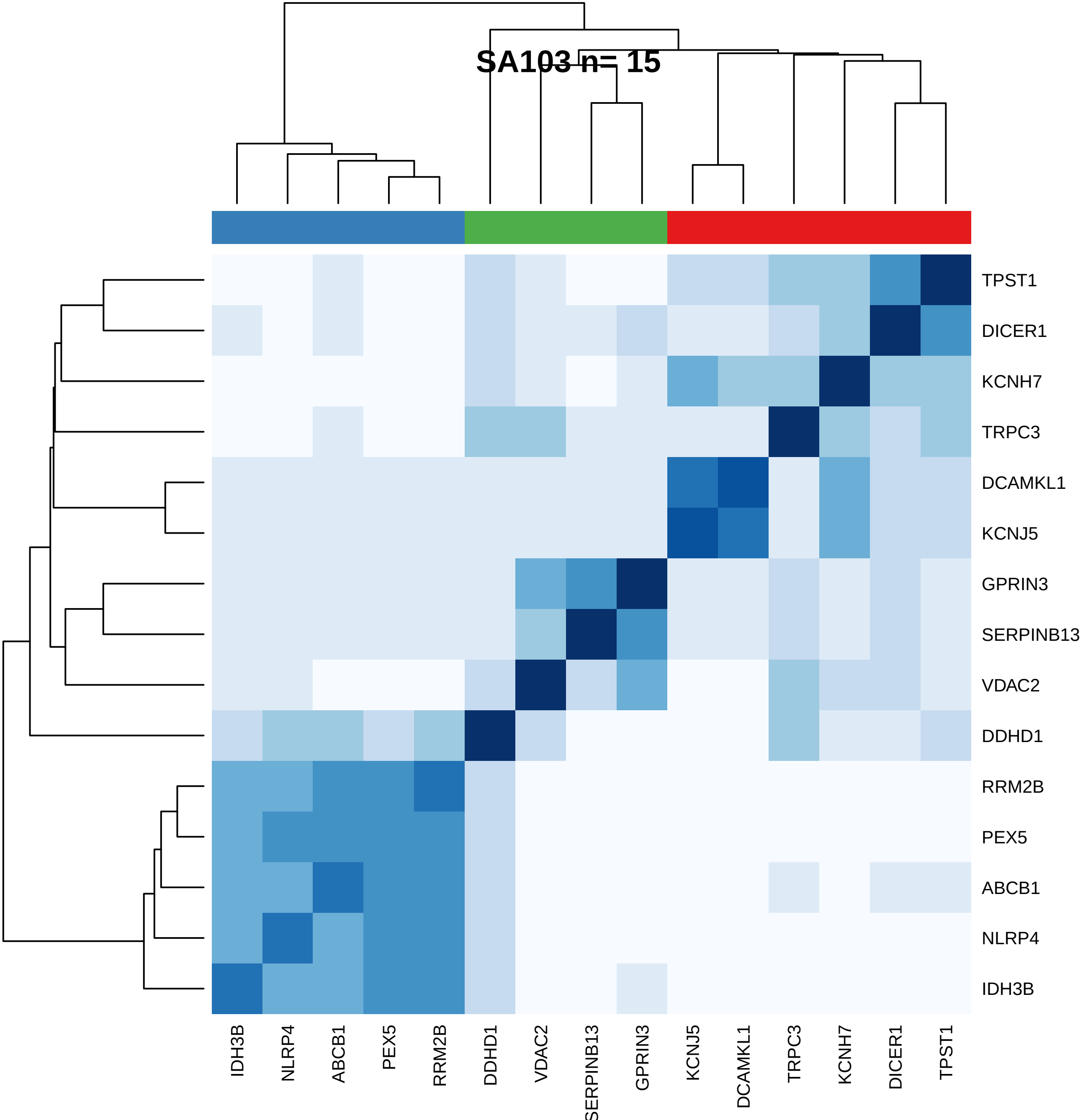

SA098 n= 22
GINS4
METTL2A
ZNF91
SLC35F3
ZNF91
PER3
PCDHA2
KLHL12
DNAH5
PRDM1
SLC17A1
SMAP1
SETD2
CYP2A6
SLA2
NRAS
CCDC52
FMN2
MYO3A
MYOM2
BTBD11
FLNB
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

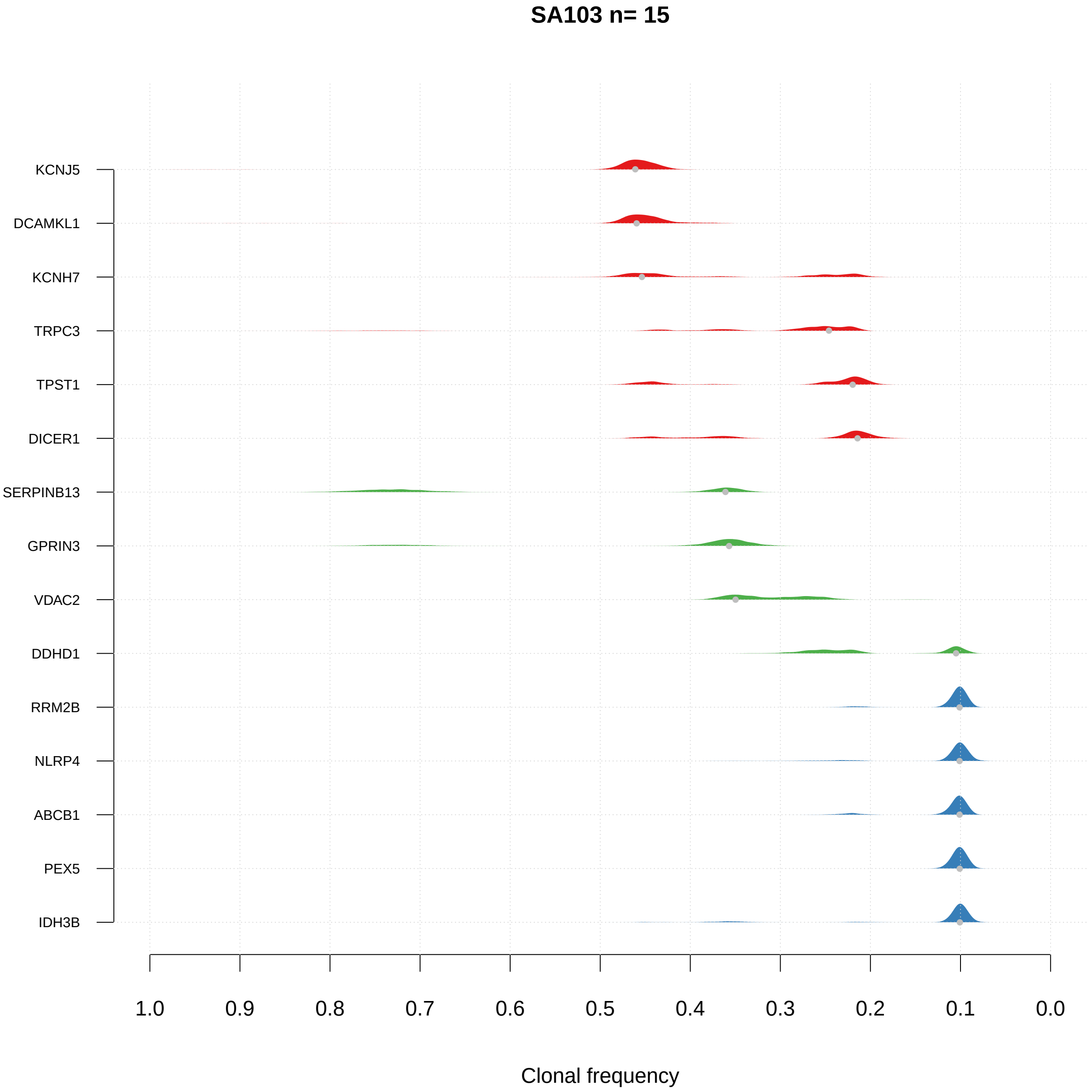

SA101 n= 18
ABCB4
MYH2
TSHZ3
DGKI
PAX7
MYL1
ITGA8
TP53
HIAT1
CDH2
C16orf54
ME2
PPM2C
GAB2
TJP1
E2F8
EVI1
HIST1H4K
HIST1H4K
EVI1
E2F8
TJP1
GAB2
PPM2C
ME2
C16orf54
CDH2
HIAT1
TP53
ITGA8
MYL1
PAX7
DGKI
TSHZ3
MYH2
ABCB4

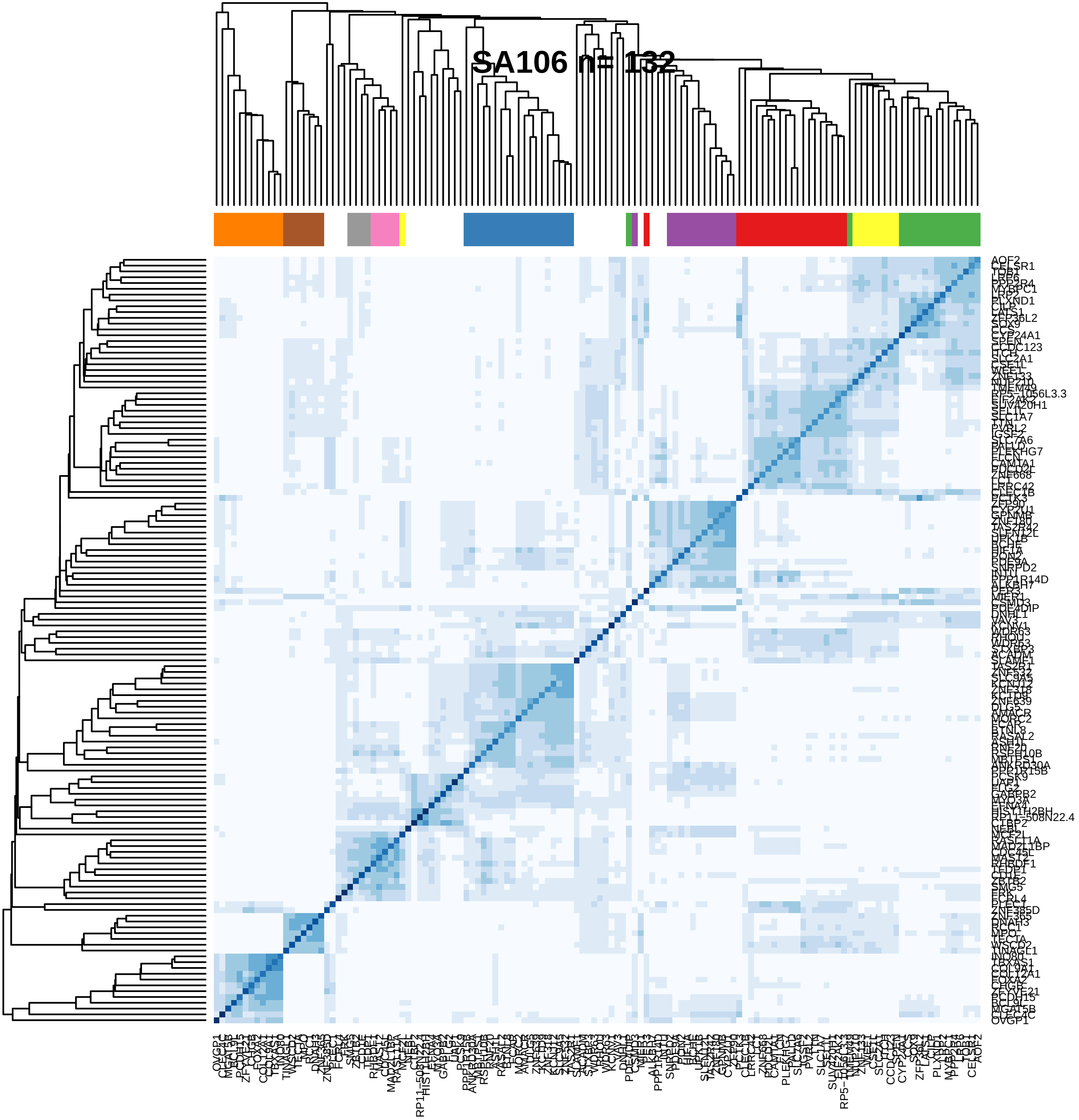

SA101 n= 18
PPM2C
C16orf54
ME2
ABCB4
DGKI
MYH2
TSHZ3
ITGA8
PAX7
MYL1
TP53
HIAT1
CDH2
GAB2
TJP1
E2F8
EVI1
HIST1H4K
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

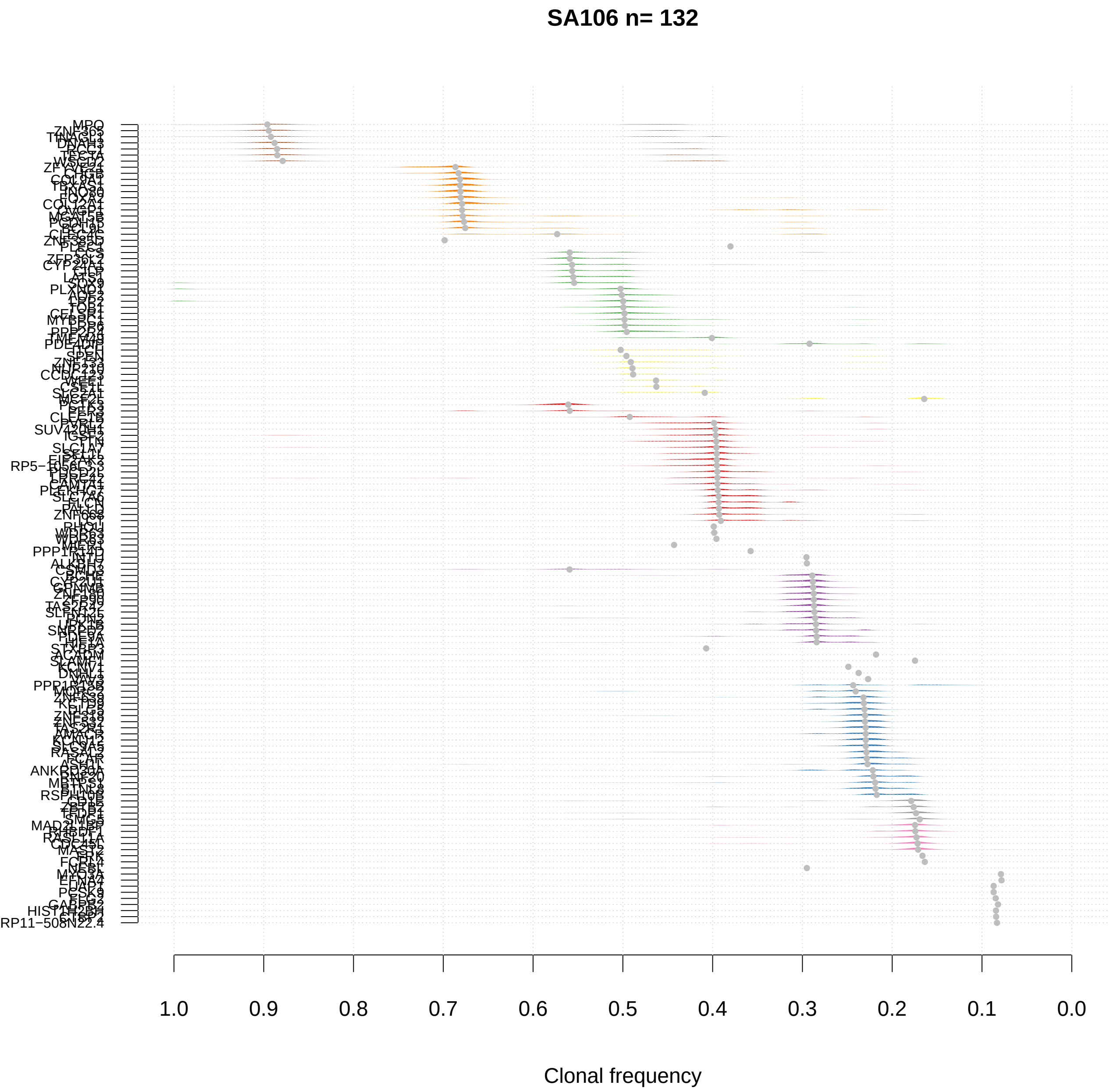

SA102 n= 24
AC005726.1−6
TP53
C10orf93
RYR1
SPG7
RNF111
ANO5
LIPC
USP37
GRIK3
ZZZ3
SLC39A10
FAM71F1
GALNT5
THRAP4
CES2
PHLDA1
CSTL1
ATP6V1E1
FAM124A
UGCGL1
SLIT2
AL357374.11−1
TNFRSF11A
TNFRSF11A
357374.11−1
SLIT2
UGCGL1
FAM124A
ATP6V1E1
CSTL1
PHLDA1
CES2
THRAP4
GALNT5
FAM71F1
SLC39A10
ZZZ3
GRIK3
USP37
LIPC
ANO5
RNF111
SPG7
RYR1
C10orf93
TP53
C005726.1−6

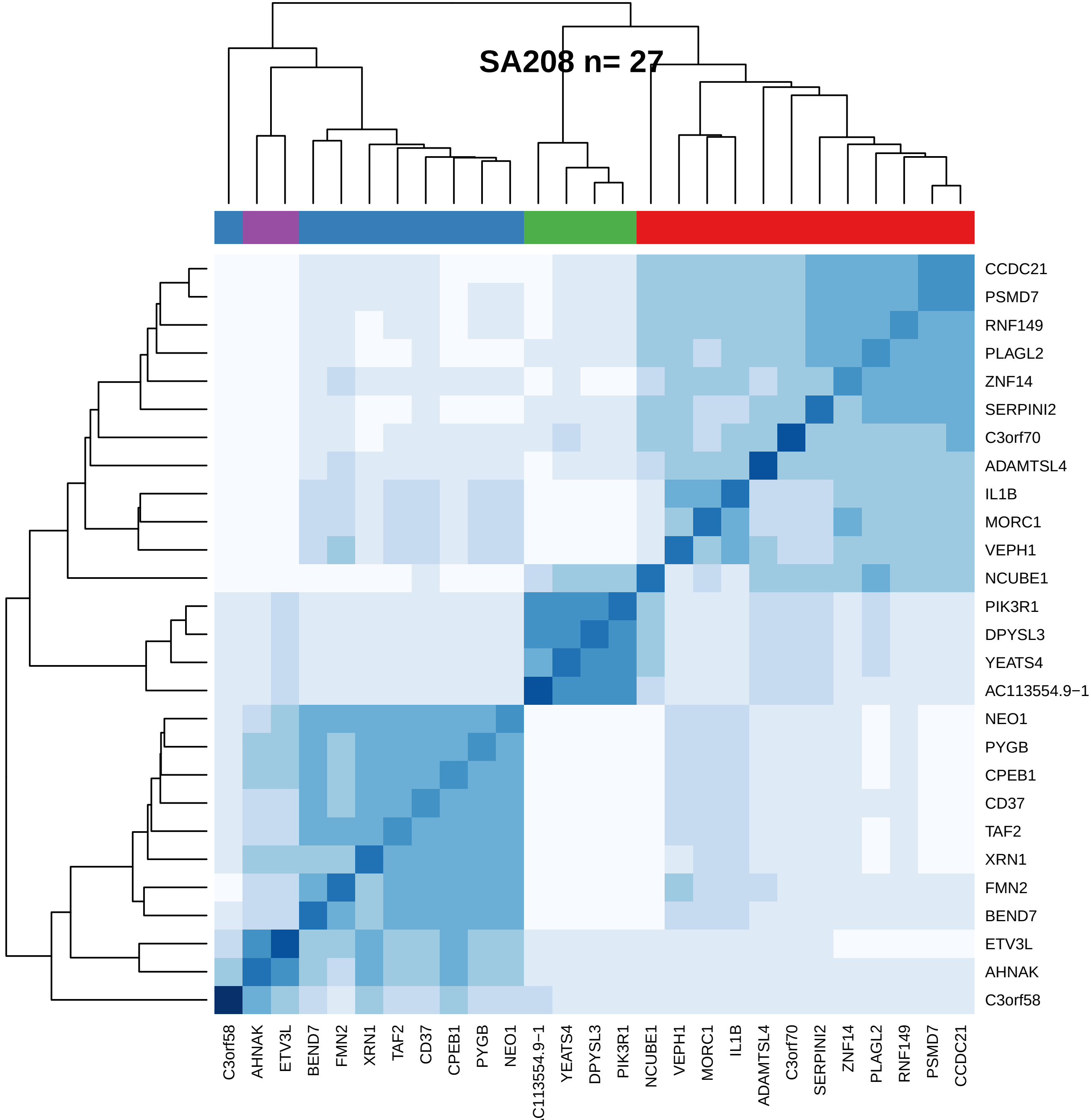

SA102 n= 24
PHLDA1
THRAP4
GALNT5
CES2
CSTL1
SLC39A10
ZZZ3
TP53
AC005726.1−6
GRIK3
SPG7
C10orf93
RYR1
RNF111
FAM71F1
ANO5
USP37
LIPC
FAM124A
AL357374.11−1
SLIT2
ATP6V1E1
UGCGL1
TNFRSF11A
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

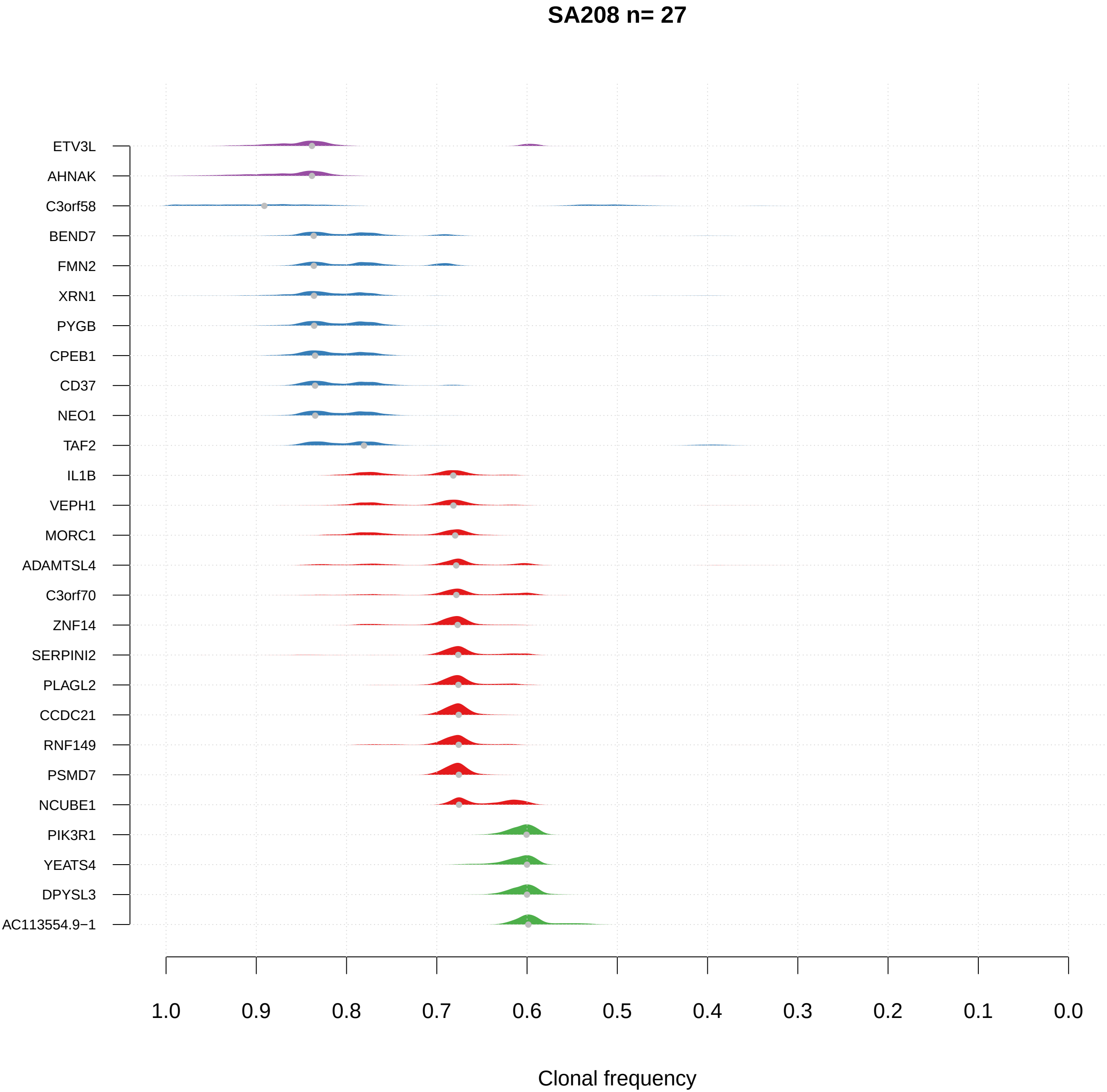

SA103 n= 15
TPST1
DICER1
KCNH7
TRPC3
DCAMKL1
KCNJ5
GPRIN3
SERPINB13
VDAC2
DDHD1
RRM2B
PEX5
ABCB1
NLRP4
IDH3B
IDH3B
NLRP4
ABCB1
PEX5
RRM2B
DDHD1
VDAC2
SERPINB13
GPRIN3
KCNJ5
DCAMKL1
TRPC3
KCNH7
DICER1
TPST1

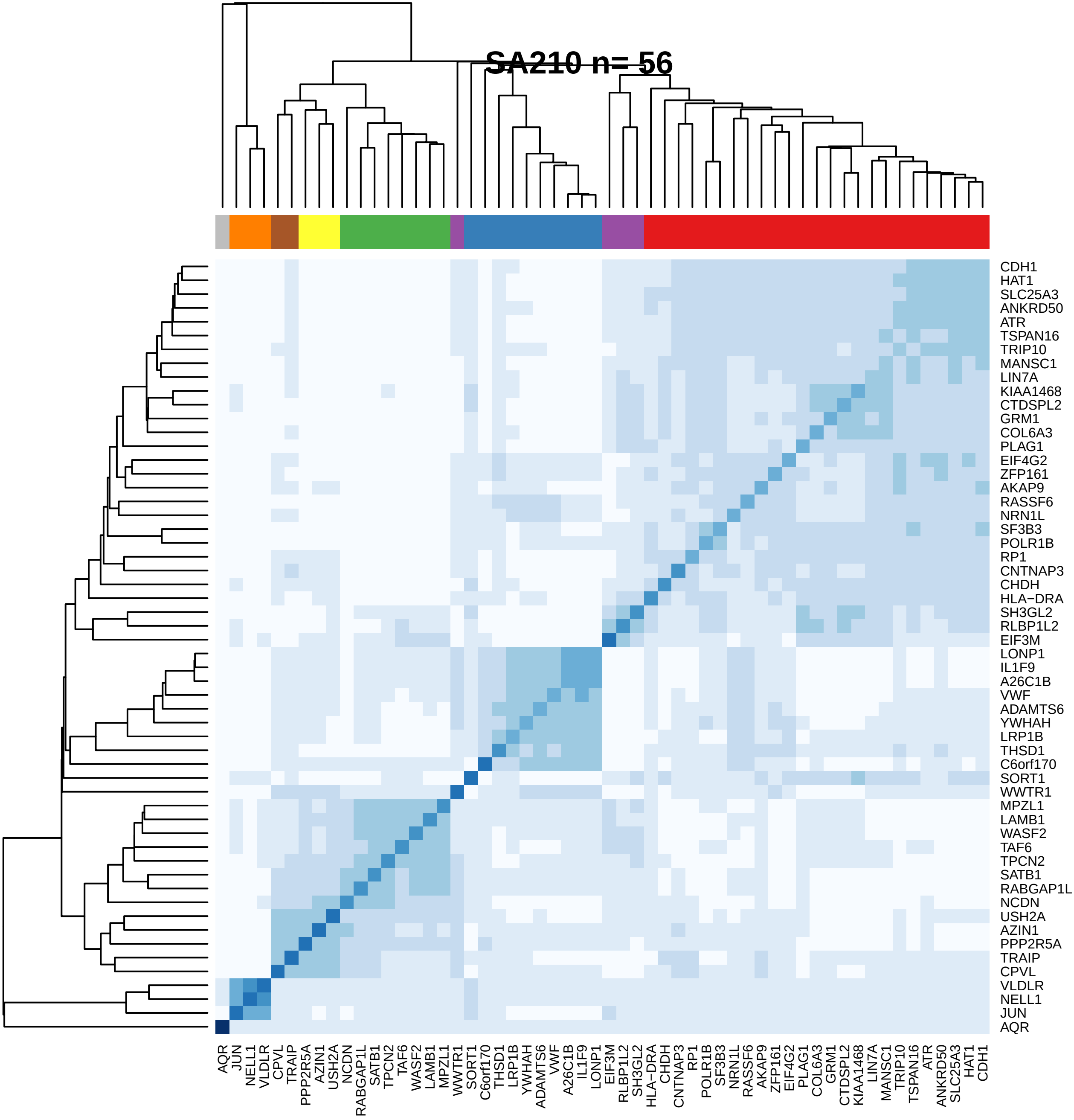

SA103 n= 15
KCNJ5
DCAMKL1
KCNH7
TRPC3
TPST1
DICER1
SERPINB13
GPRIN3
VDAC2
DDHD1
RRM2B
NLRP4
ABCB1
PEX5
IDH3B
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

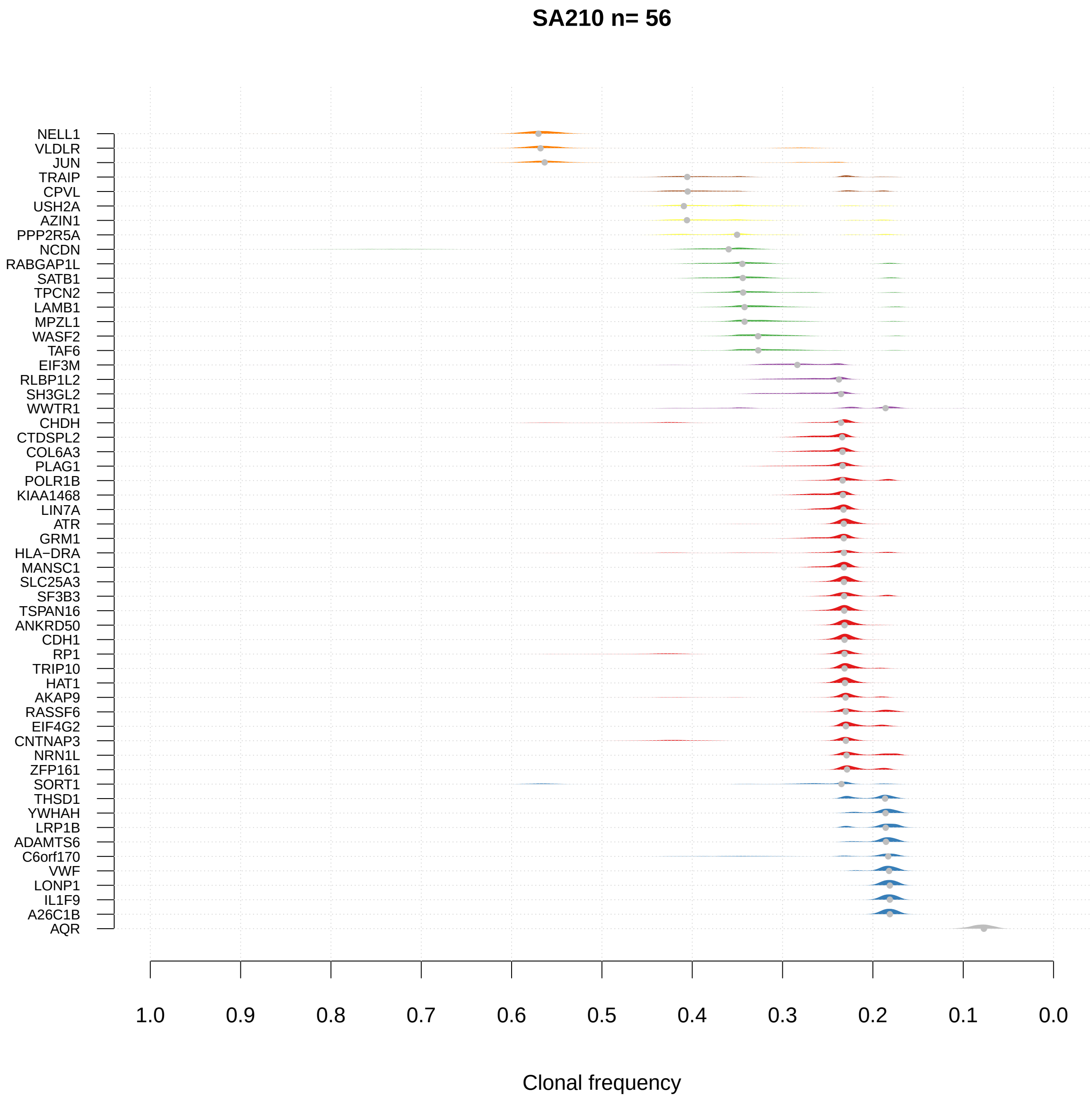

SA106 n= 132
AOF2
CELSR1
TOB1
LRP6
PPP2R4
MYBPC1
LRP2
PLXND1
CILP
LATS1
ZFP36L2
SOX9
CCS
CYP24A1
SPEN
CCDC123
ITCH
SLC2A1
CSE1L
WEE1
ZNF133
NUP210
TMEM49
RP5−1056L3.3
EIF2AK2
SUV420H1
SELL
SLC1A7
TTN
PVRL2
IGSF2
SLC7A6
PALLD
PLEKHG7
FLCN
CAMTA1
PDCD2L
ZNF668
LCT
LRRC42
CLEC1B
PCTK3
ZFP90
CYP2U1
GPNMB
ZNF180
TAS2R42
SLFN12L
UPK1B
BCHE
HIF1A
PON2
PDE9A
SNRPD2
INTU
PPP1R14D
ALKBH7
PER3
MIER1
CSMD3
PDE4DIP
DNHL1
VAV3
KCNV1
WDR63
RHOU
WDR63
STXBP3
ACADM
SLAMF1
TAS2R1
ZNF532
SLC9A5
KCNJ12
ZNF318
KCTD9
ZNF639
DLG5
AMACR
MORC2
FCAR
BTNL8
RASAL2
ASH1L
RNF20
RSPH10B
MBTPS1
ANKRD30A
PPP1R15B
PCSK9
UAP1
FLG2
GABPB2
MYO3A
EFNA4
HIST1H2BH
RP11−508N22.4
CTBP2
NEBL
MCF2L
RASL11A
MAD2L1BP
CDC45L
MAST2
RHBDF1
TEDP1
CD1E
ZBTB2
SMG5
ERK
FCRL4
PLEC1
ZNF385D
ZNF365
DNAH3
RCC1
MPO
TECTA
WSCD2
TINAGL1
INO80
TBXAS1
COL9A1
COL12A1
FOXA2
CHGB
ZFYVE21
PCDH15
BCL9L
MGAT5B
CLEC4C
OVGP1

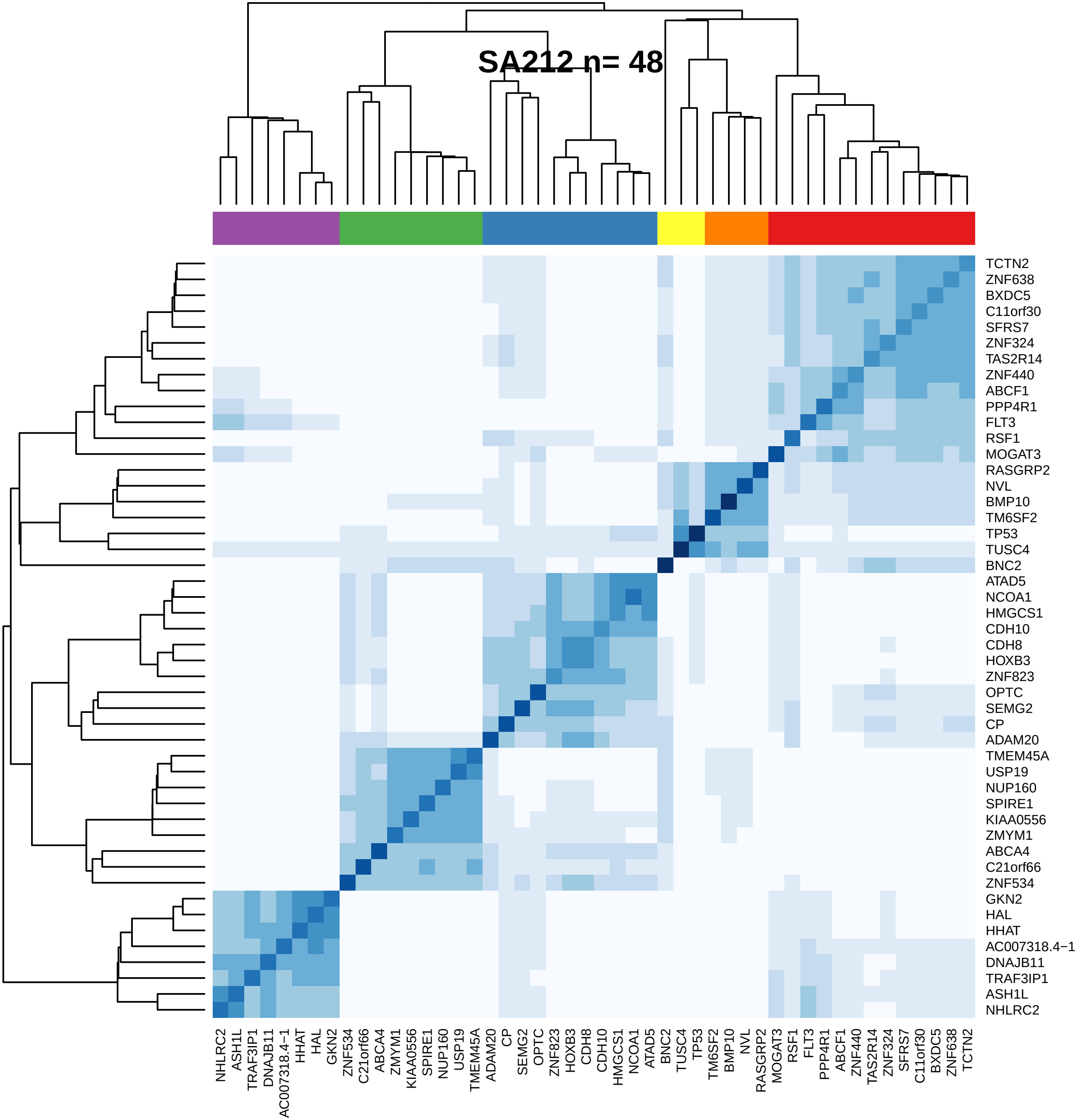

SA106 n= 132
MPO
ZNF365
TRNAU1
DNAH3
ZFYVE26
CHGB
COL6A1
TSKU
COL12A1
ZNF382
PLXND1
ANO2
MYBPC1
TMEM40
PDE4DIP
SPEN
ZNF157
CCDC13
MCF2L
CLEC18B
SUV420H1
IGSF11
SLC1A7
RP5−1057J7.6
PLEKHG7
ZNF606
PPP1R14D
CMIP
ZNF180
SLFN12L
SNRPB2
STX1B
ACADM
PPP1R12B
KCNC2
RASA3
ANKRD30A
RSPH10B
MAD2L1BP
MASP2
FCRL4
MYO3A
EDAR
PCSK9
HIST1H2BH
RP11−508N22.4
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

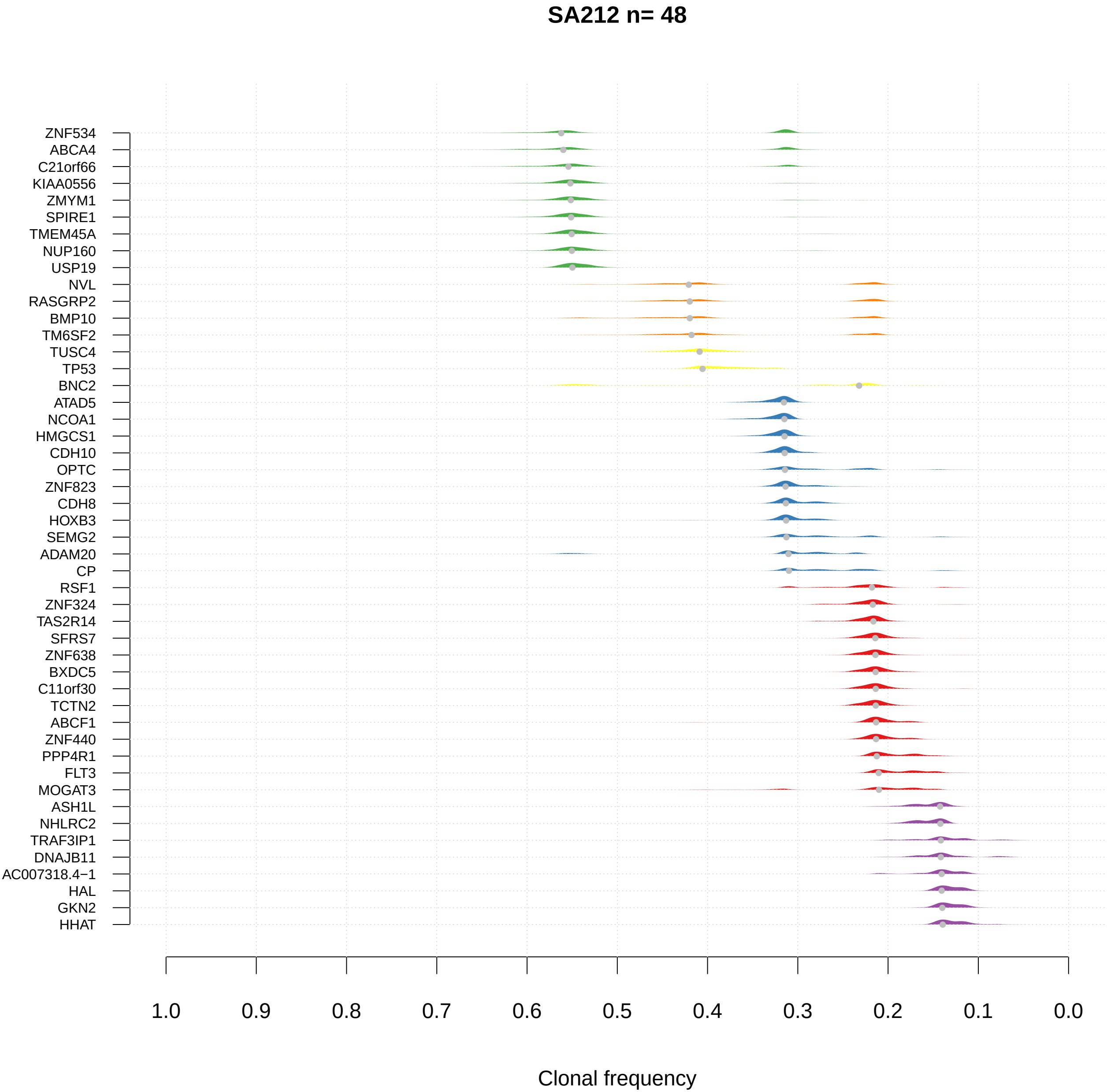

SA208 n= 27
CCDC21
PSMD7
RNF149
PLAGL2
ZNF14
SERPINI2
C3orf70
ADAMTSL4
IL1B
MORC1
VEPH1
NCUBE1
PIK3R1
DPYSL3
YEATS4
AC113554.9−1
NEO1
PYGB
CPEB1
CD37
TAF2
XRN1
FMN2
BEND7
ETV3L
AHNAK
C3orf58
C3orf58
AHNAK
ETV3L
BEND7
FMN2
XRN1
TAF2
CD37
CPEB1
PYGB
NEO1
AC113554.9−1
YEATS4
DPYSL3
PIK3R1
NCUBE1
VEPH1
MORC1
IL1B
ADAMTSL4
C3orf70
SERPINI2
ZNF14
PLAGL2
RNF149
PSMD7
CCDC21

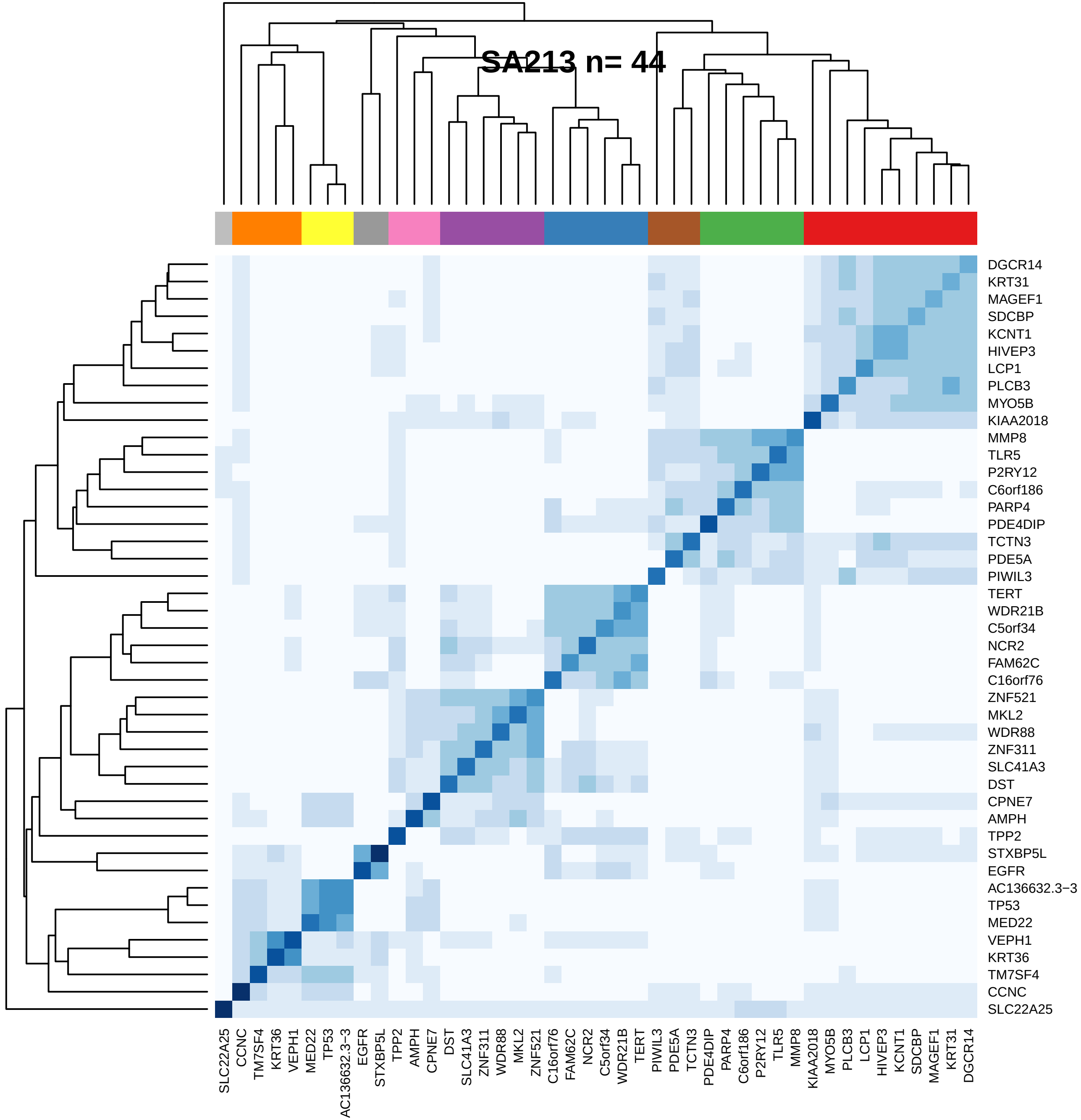

SA208 n= 27
ETV3L
AHNAK
C3orf58
BEND7
FMN2
XRN1
PYGB
CPEB1
CD37
NEO1
TAF2
IL1B
VEPH1
MORC1
ADAMTSL4
C3orf70
ZNF14
SERPINI2
PLAGL2
CCDC21
RNF149
PSMD7
NCUBE1
PIK3R1
YEATS4
DPYSL3
AC113554.9−1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

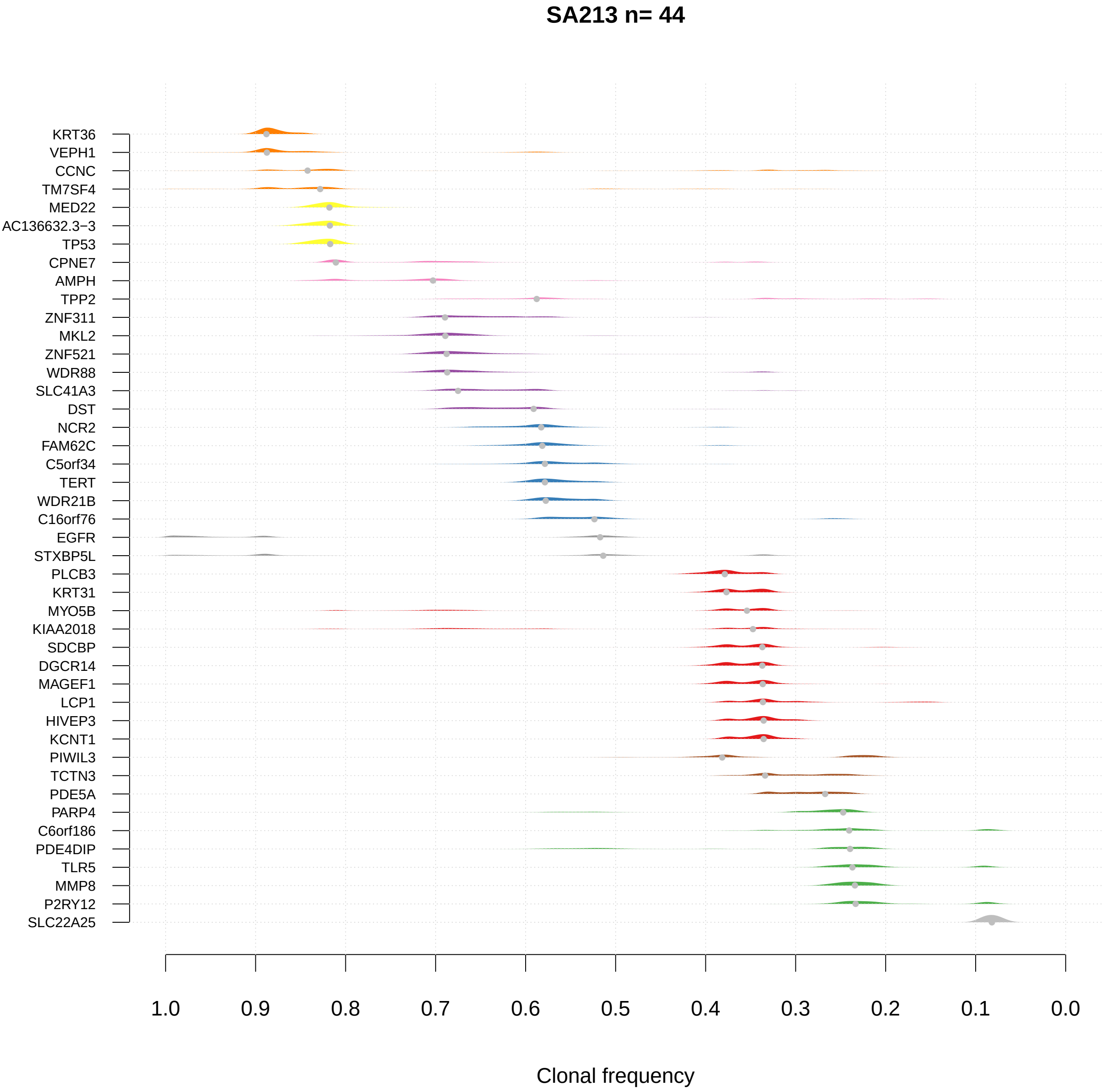

SA210 n= 56
CDH1
HAT1
SLC25A3
ANKRD50
ATR
TSPAN16
TRIP10
MANSC1
LIN7A
KIAA1468
CTDSPL2
GRM1
COL6A3
PLAG1
EIF4G2
ZFP161
AKAP9
RASSF6
NRN1L
SF3B3
POLR1B
RP1
CNTNAP3
CHDH
HLA−DRA
SH3GL2
RLBP1L2
EIF3M
LONP1
IL1F9
A26C1B
VWF
ADAMTS6
YWHAH
LRP1B
THSD1
C6orf170
SORT1
WWTR1
MPZL1
LAMB1
WASF2
TAF6
TPCN2
SATB1
RABGAP1L
NCDN
USH2A
AZIN1
PPP2R5A
TRAIP
CPVL
VLDLR
NELL1
JUN
AQR
AQR
JUN
NELL1
VLDLR
CPVL
TRAIP
PPP2R5A
AZIN1
USH2A
NCDN
RABGAP1L
SATB1
TPCN2
TAF6
WASF2
LAMB1
MPZL1
WWTR1
SORT1
C6orf170
THSD1
LRP1B
YWHAH
ADAMTS6
VWF
A26C1B
IL1F9
LONP1
EIF3M
RLBP1L2
SH3GL2
HLA−DRA
CHDH
CNTNAP3
RP1
POLR1B
SF3B3
NRN1L
RASSF6
AKAP9
ZFP161
EIF4G2
PLAG1
COL6A3
GRM1
CTDSPL2
KIAA1468
LIN7A
MANSC1
TRIP10
TSPAN16
ATR
ANKRD50
SLC25A3
HAT1
CDH1

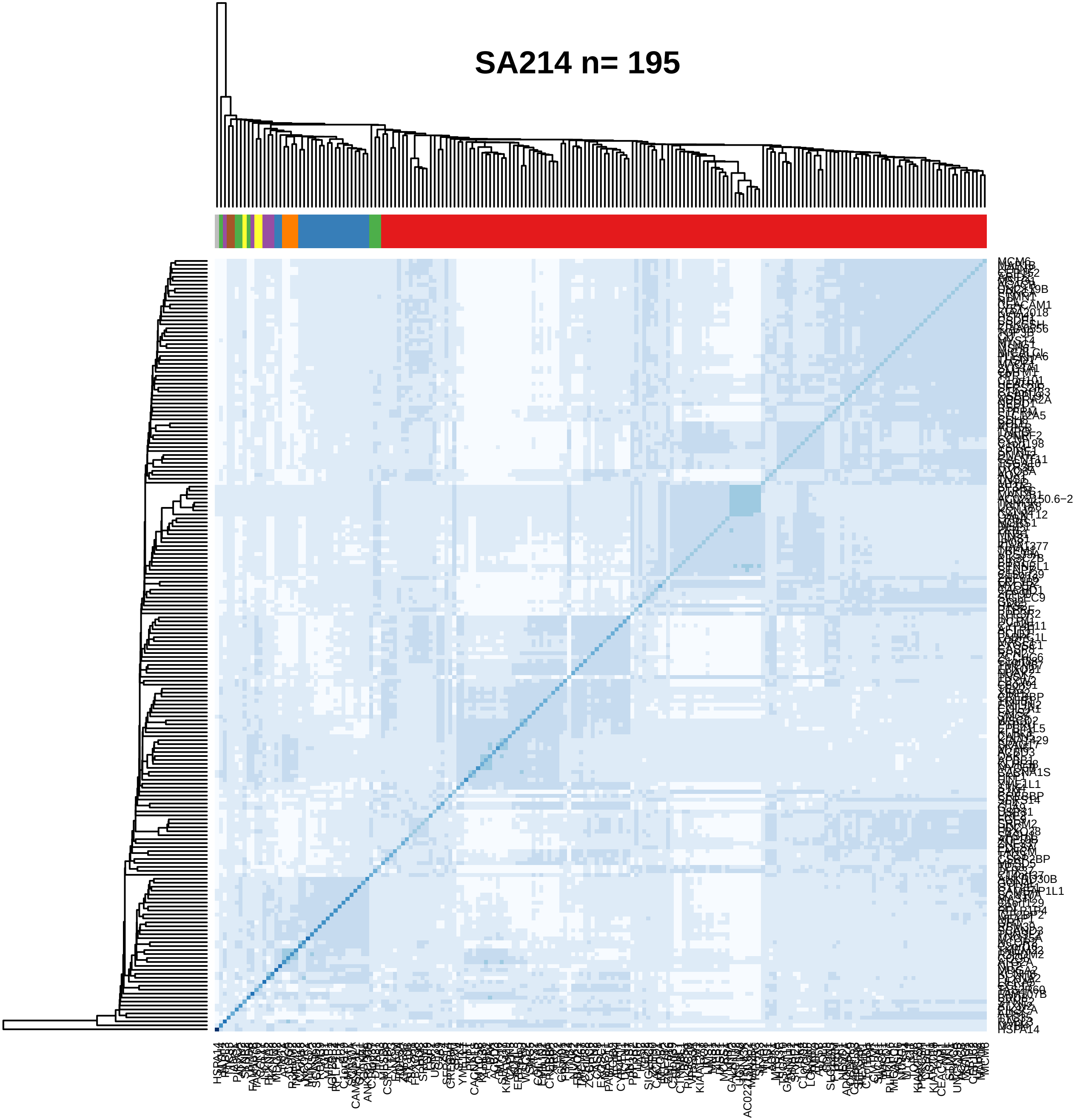

SA210 n= 56
NELL1
VLDLR
JUN
TRAIP
CPVL
USH2A
AZIN1
PPP2R5A
NCDN
RABGAP1L
SATB1
TPCN2
LAMB1
MPZL1
WASF2
TAF6
EIF3M
RLBP1L2
SH3GL2
WWTR1
CHDH
CTDSPL2
COL6A3
PLAG1
POLR1B
KIAA1468
LIN7A
ATR
GRM1
HLA−DRA
MANSC1
SLC25A3
SF3B3
TSPAN16
ANKRD50
CDH1
RP1
TRIP10
HAT1
AKAP9
RASSF6
EIF4G2
CNTNAP3
NRN1L
ZFP161
SORT1
THSD1
YWHAH
LRP1B
ADAMTS6
C6orf170
VWF
LONP1
IL1F9
A26C1B
AQR
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

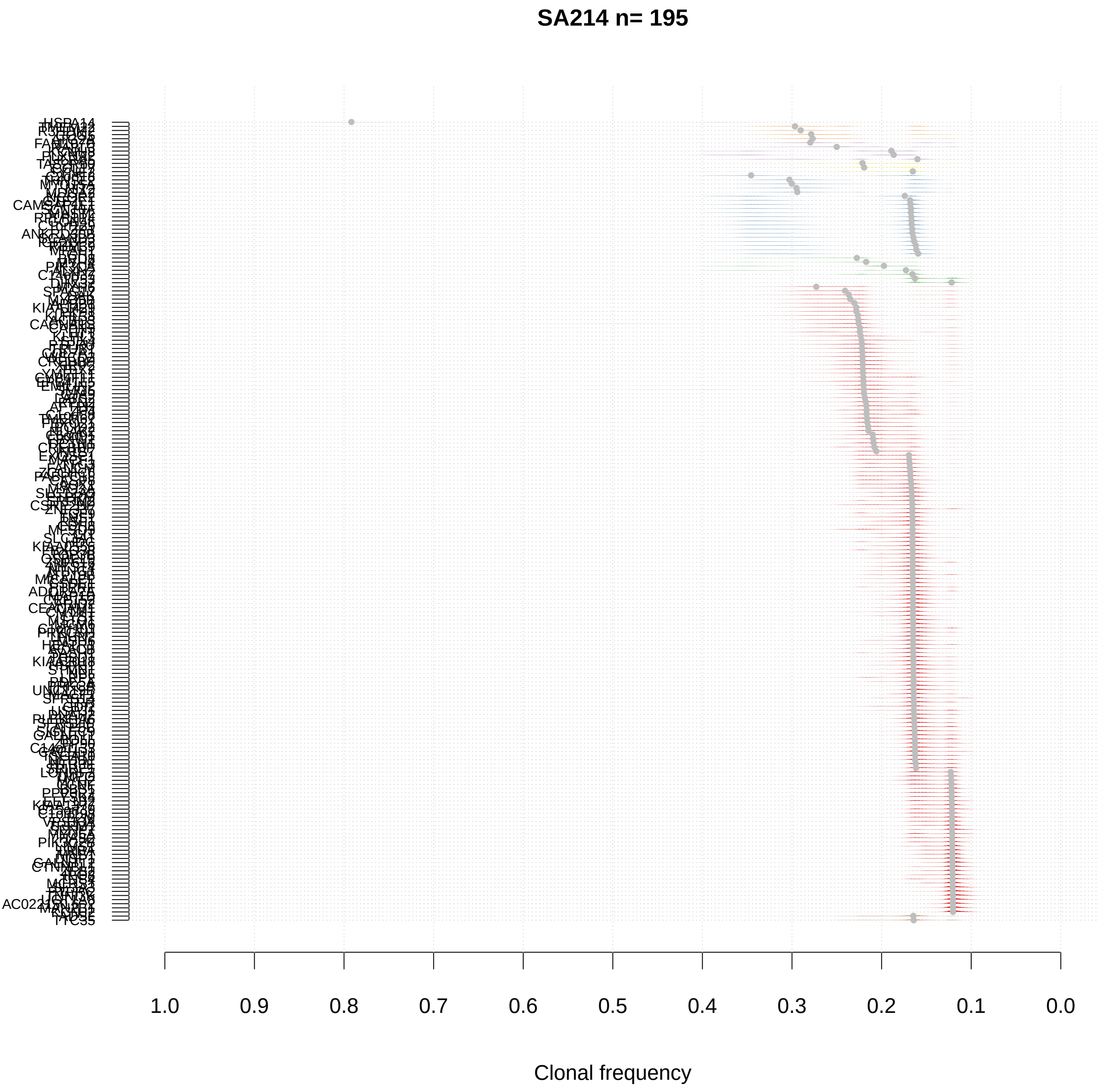

SA212 n= 48
TCTN2
ZNF638
BXDC5
C11orf30
SFRS7
ZNF324
TAS2R14
ZNF440
ABCF1
PPP4R1
FLT3
RSF1
MOGAT3
RASGRP2
NVL
BMP10
TM6SF2
TP53
TUSC4
BNC2
ATAD5
NCOA1
HMGCS1
CDH10
CDH8
HOXB3
ZNF823
OPTC
SEMG2
CP
ADAM20
TMEM45A
USP19
NUP160
SPIRE1
KIAA0556
ZMYM1
ABCA4
C21orf66
ZNF534
GKN2
HAL
HHAT
AC007318.4−1
DNAJB11
TRAF3IP1
ASH1L
NHLRC2
NHLRC2
ASH1L
TRAF3IP1
DNAJB11
AC007318.4−1
HHAT
HAL
GKN2
ZNF534
C21orf66
ABCA4
ZMYM1
KIAA0556
SPIRE1
NUP160
USP19
TMEM45A
ADAM20
CP
SEMG2
OPTC
ZNF823
HOXB3
CDH8
CDH10
HMGCS1
NCOA1
ATAD5
BNC2
TUSC4
TP53
TM6SF2
BMP10
NVL
RASGRP2
MOGAT3
RSF1
FLT3
PPP4R1
ABCF1
ZNF440
TAS2R14
ZNF324
SFRS7
C11orf30
BXDC5
ZNF638
TCTN2

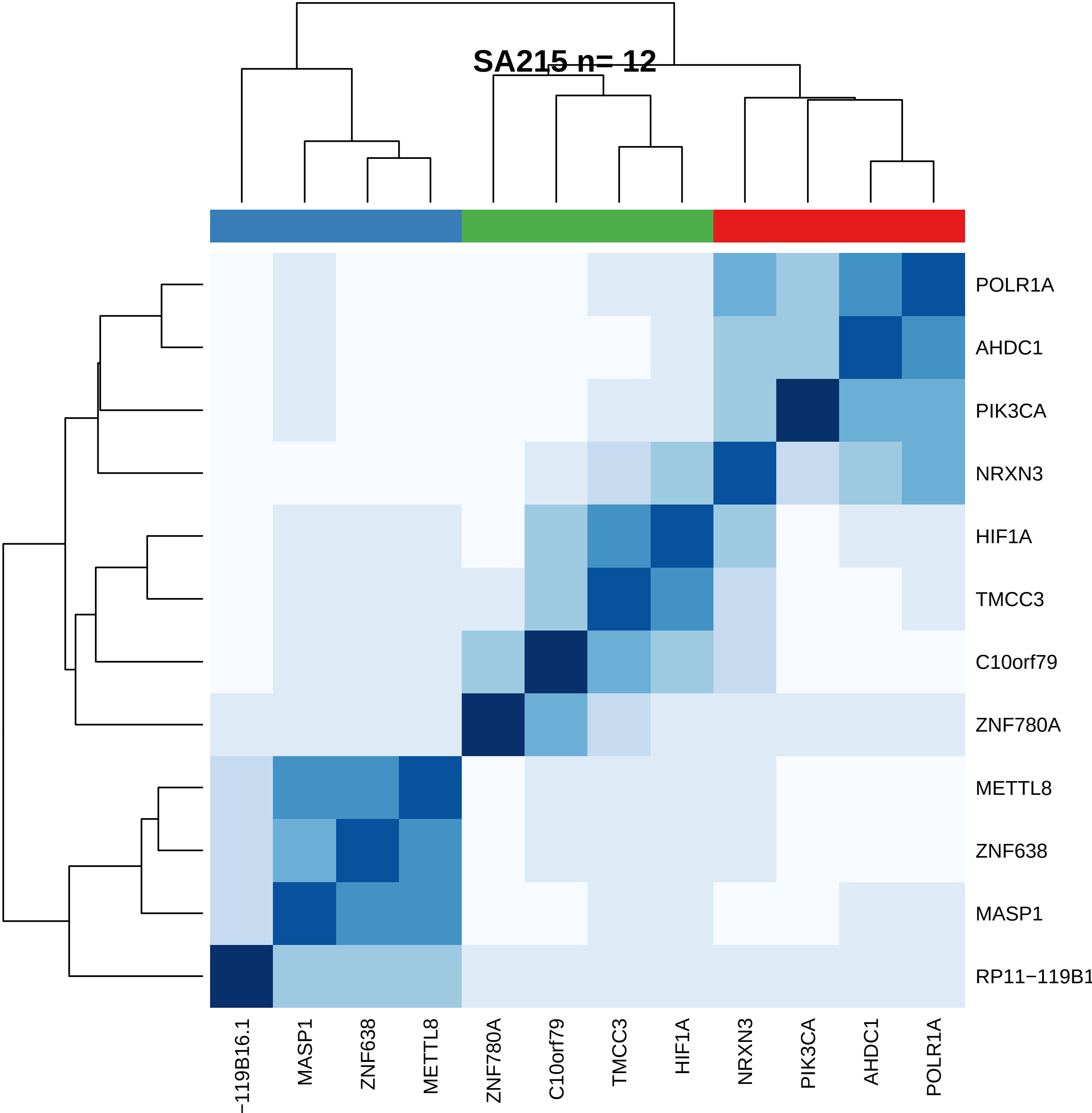

SA212 n= 48
ZNF534
ABCA4
C21orf66
KIAA0556
ZMYM1
SPIRE1
TMEM45A
NUP160
USP19
NVL
RASGRP2
BMP10
TM6SF2
TUSC4
TP53
BNC2
ATAD5
NCOA1
HMGCS1
CDH10
OPTC
ZNF823
CDH8
HOXB3
SEMG2
ADAM20
CP
RSF1
ZNF324
TAS2R14
SFRS7
ZNF638
BXDC5
C11orf30
TCTN2
ABCF1
ZNF440
PPP4R1
FLT3
MOGAT3
ASH1L
NHLRC2
TRAF3IP1
DNAJB11
AC007318.4−1
HAL
GKN2
HHAT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

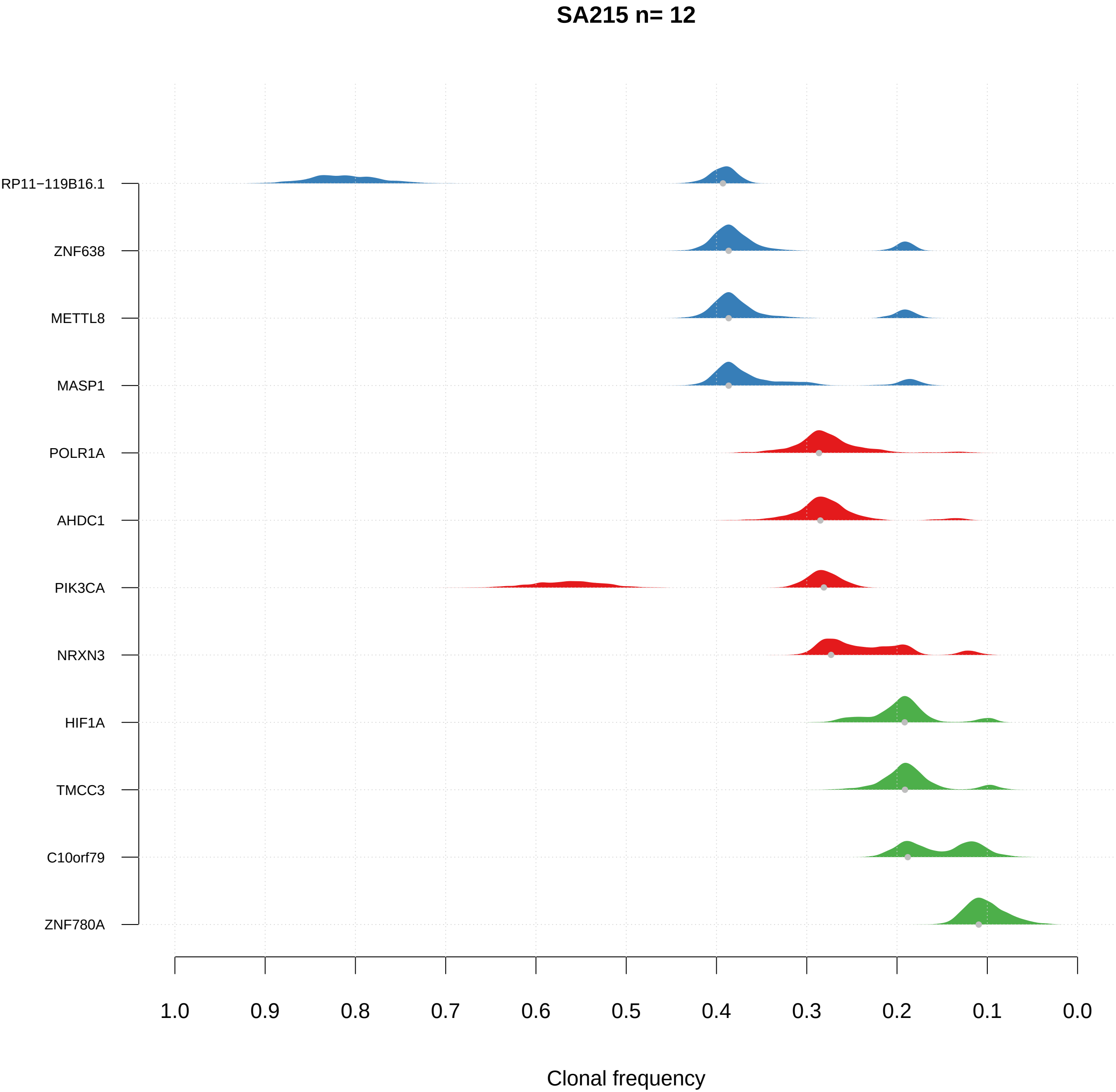

SA213 n= 44
DGCR14
KRT31
MAGEF1
SDCBP
KCNT1
HIVEP3
LCP1
PLCB3
MYO5B
KIAA2018
MMP8
TLR5
P2RY12
C6orf186
PARP4
PDE4DIP
TCTN3
PDE5A
PIWIL3
TERT
WDR21B
C5orf34
NCR2
FAM62C
C16orf76
ZNF521
MKL2
WDR88
ZNF311
SLC41A3
DST
CPNE7
AMPH
TPP2
STXBP5L
EGFR
AC136632.3−3
TP53
MED22
VEPH1
KRT36
TM7SF4
CCNC
SLC22A25
SLC22A25
CCNC
TM7SF4
KRT36
VEPH1
MED22
TP53
AC136632.3−3
EGFR
STXBP5L
TPP2
AMPH
CPNE7
DST
SLC41A3
ZNF311
WDR88
MKL2
ZNF521
C16orf76
FAM62C
NCR2
C5orf34
WDR21B
TERT
PIWIL3
PDE5A
TCTN3
PDE4DIP
PARP4
C6orf186
P2RY12
TLR5
MMP8
KIAA2018
MYO5B
PLCB3
LCP1
HIVEP3
KCNT1
SDCBP
MAGEF1
KRT31
DGCR14

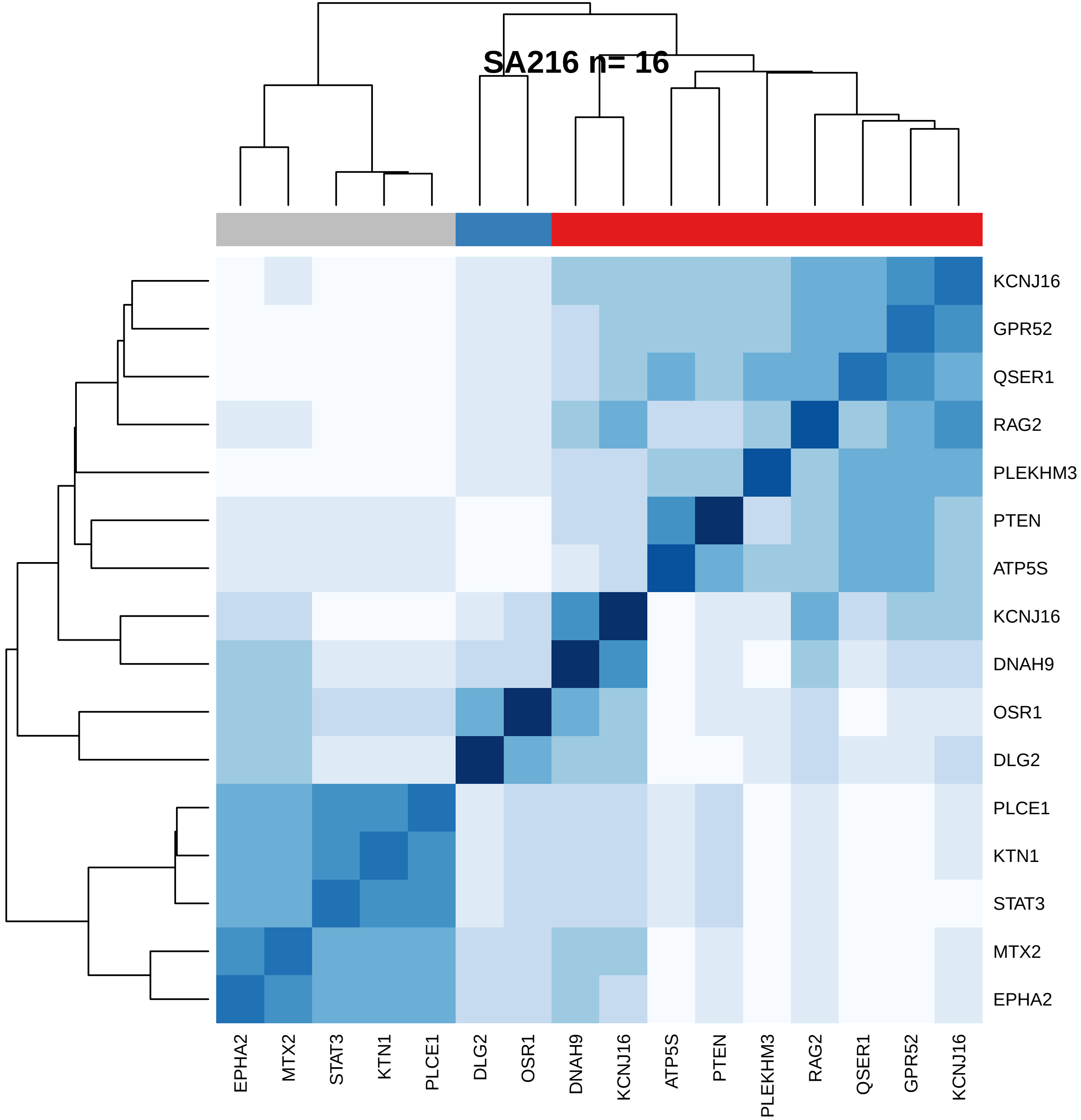

SA213 n= 44
KRT36
VEPH1
CCNC
TM7SF4
MED22
AC136632.3−3
TP53
CPNE7
AMPH
TPP2
ZNF311
MKL2
ZNF521
WDR88
SLC41A3
DST
NCR2
FAM62C
C5orf34
TERT
WDR21B
C16orf76
EGFR
STXBP5L
PLCB3
KRT31
MYO5B
KIAA2018
SDCBP
DGCR14
MAGEF1
LCP1
HIVEP3
KCNT1
PIWIL3
TCTN3
PDE5A
PARP4
C6orf186
PDE4DIP
TLR5
MMP8
P2RY12
SLC22A25
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

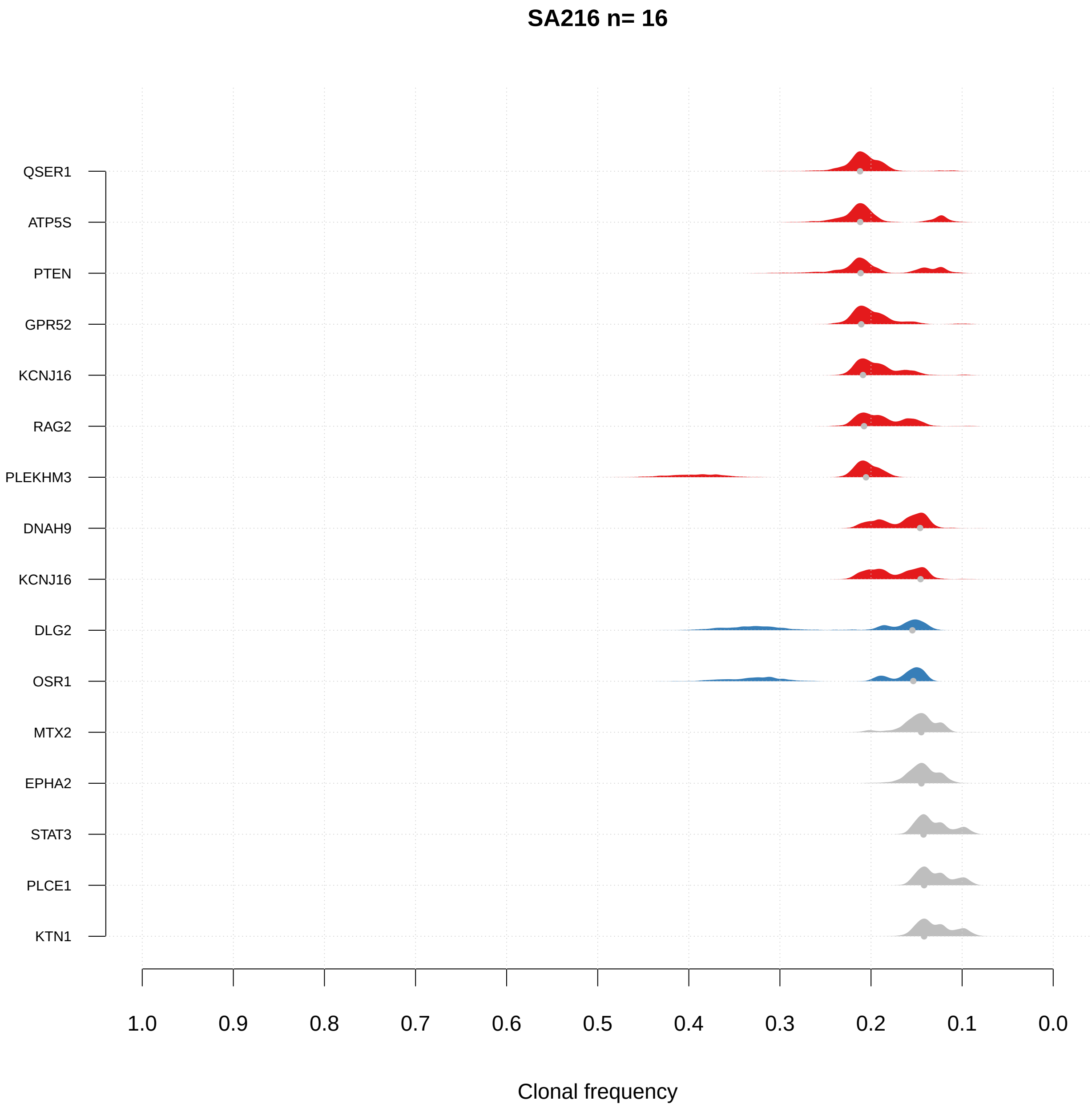

SA214 n= 195
MCM6
HSPA14
KIAA0556
VPS13D
CACNA1S
AC022150.6−2
ERCC6
FAM107B
HSPA14
MCM6

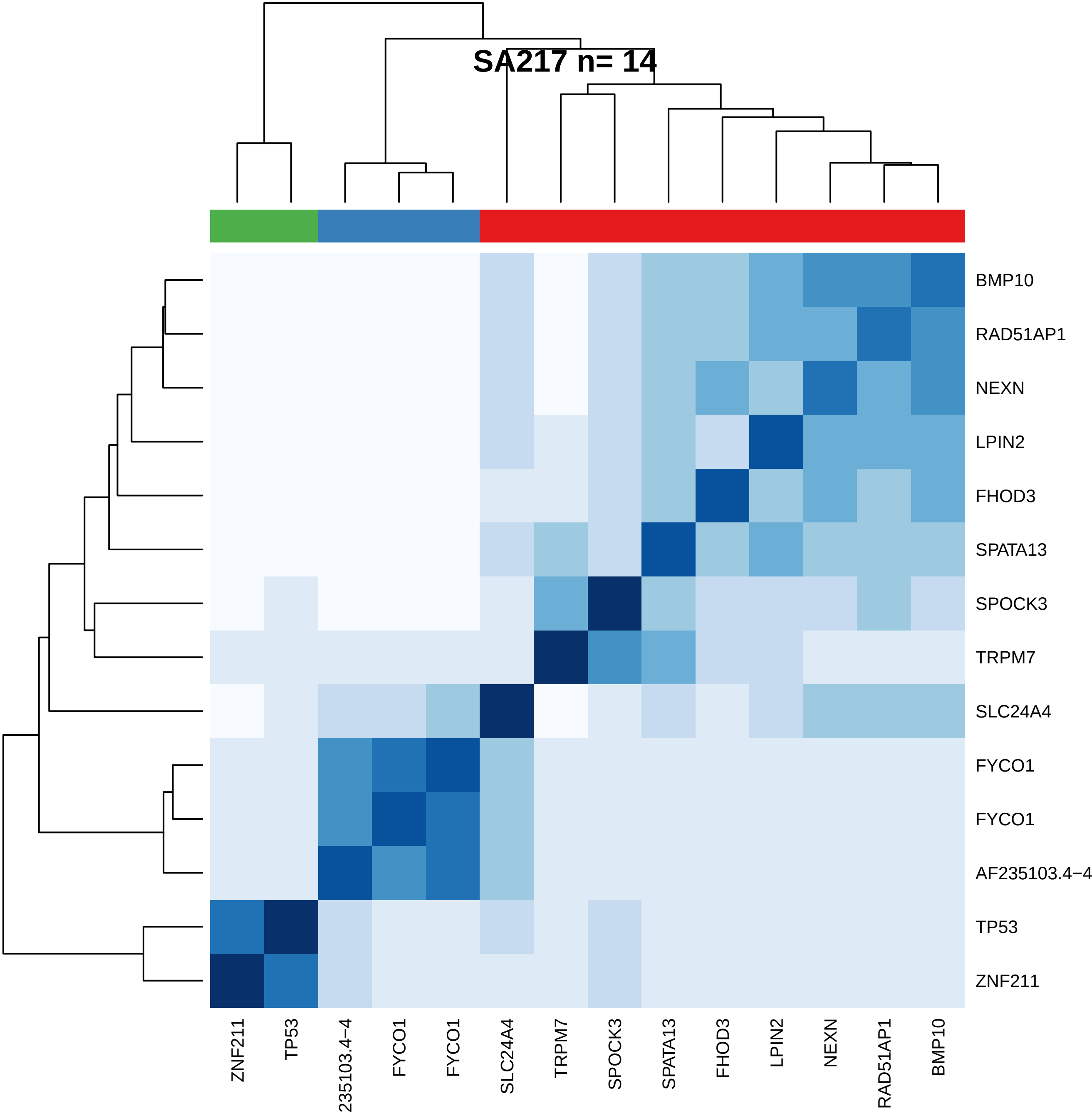

SA214 n= 195
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency
HSPA14
ITC35

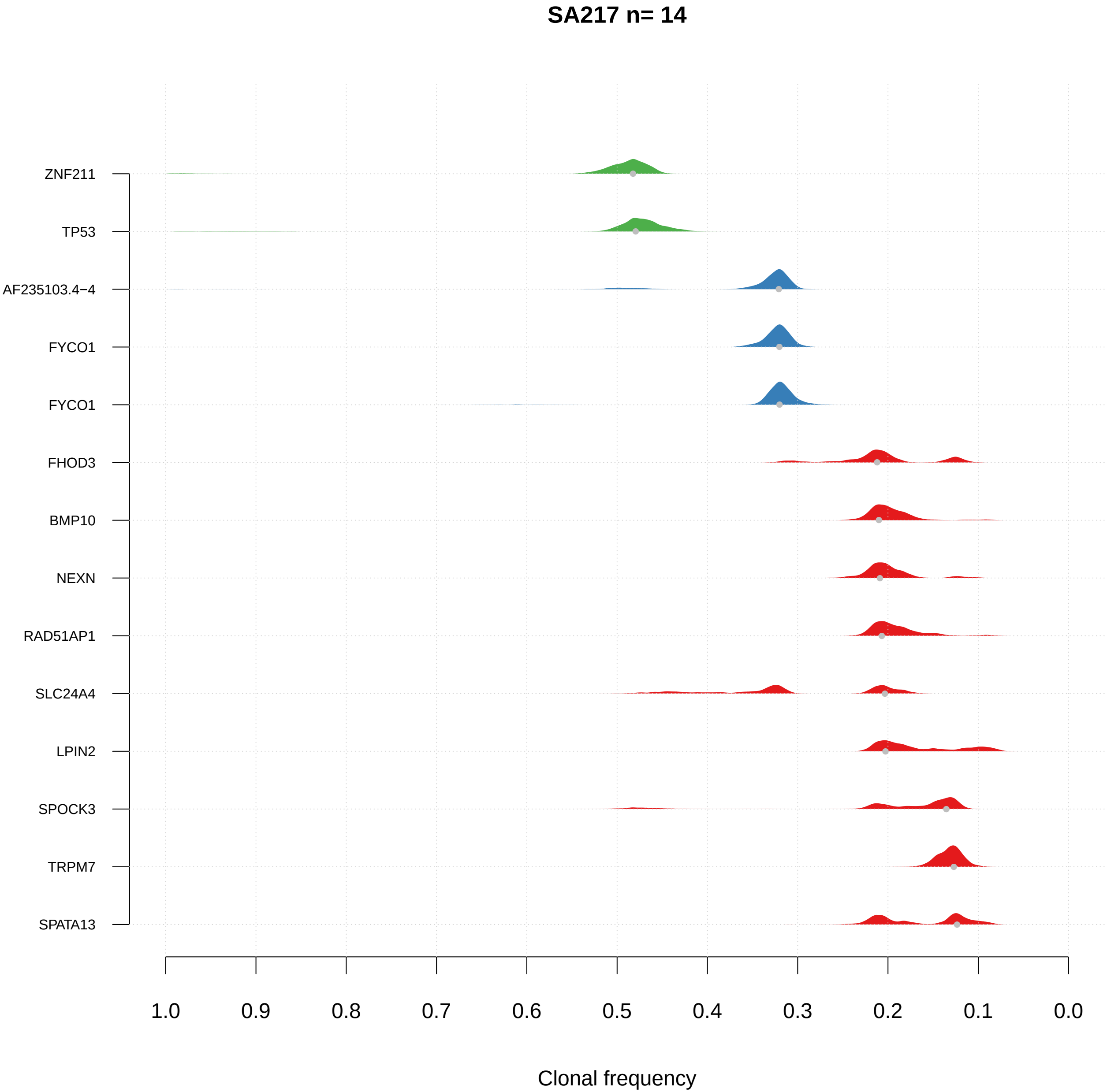

SA215 n= 12
POLR1A
AHDC1
PIK3CA
NRXN3
HIF1A
TMCC3
C10orf79
ZNF780A
METTL8
ZNF638
MASP1
RP11−119B1
−119B16.1
MASP1
ZNF638
METTL8
ZNF780A
C10orf79
TMCC3
HIF1A
NRXN3
PIK3CA
AHDC1
POLR1A

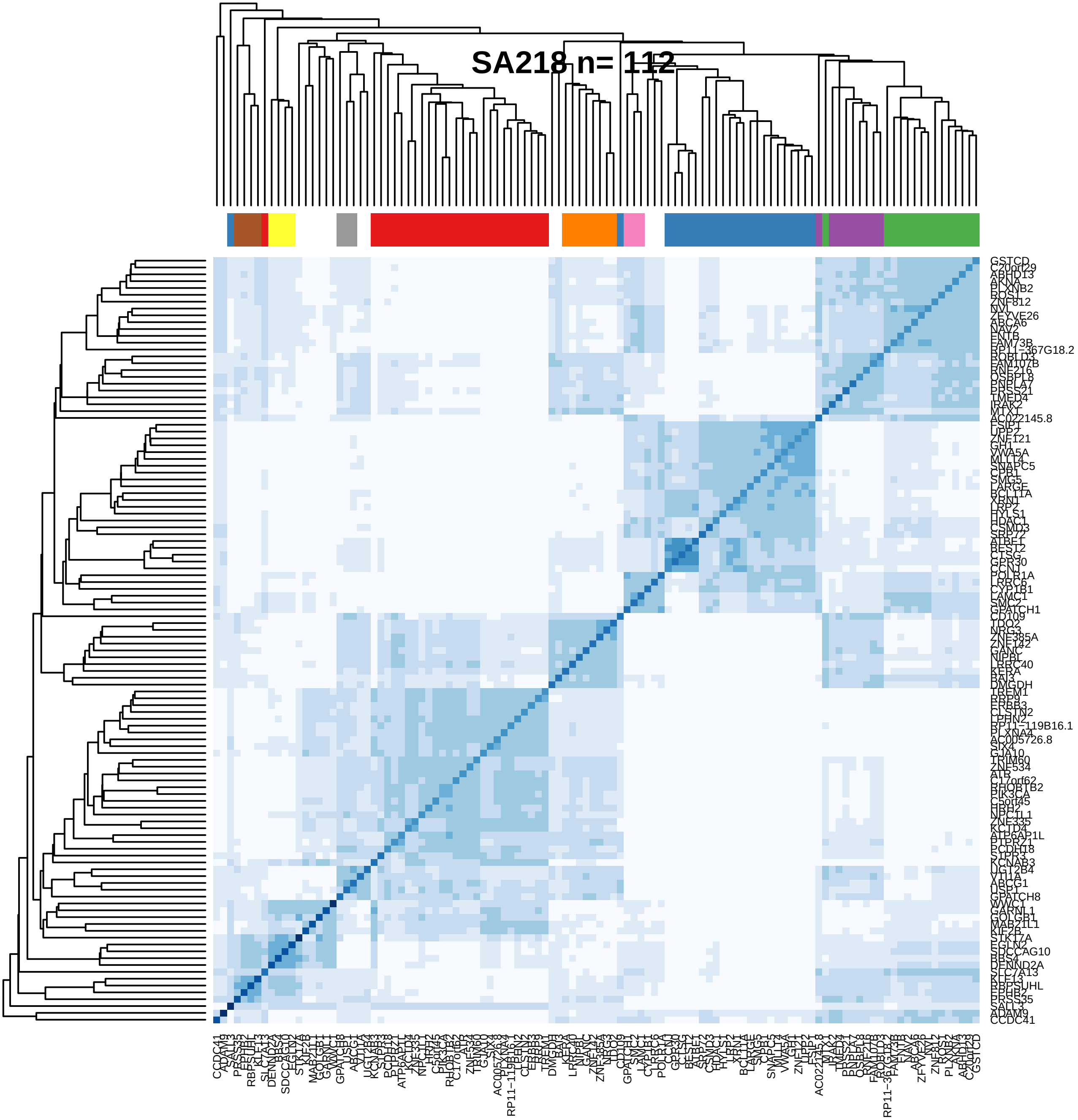

SA215 n= 12
RP11−119B16.1
ZNF638
METTL8
MASP1
POLR1A
AHDC1
PIK3CA
NRXN3
HIF1A
TMCC3
C10orf79
ZNF780A
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA216 n= 16
KCNJ16
GPR52
QSER1
RAG2
PLEKHM3
PTEN
ATP5S
KCNJ16
DNAH9
OSR1
DLG2
PLCE1
KTN1
STAT3
MTX2
EPHA2
EPHA2
MTX2
STAT3
KTN1
PLCE1
DLG2
OSR1
DNAH9
KCNJ16
ATP5S
PTEN
PLEKHM3
RAG2
QSER1
GPR52
KCNJ16

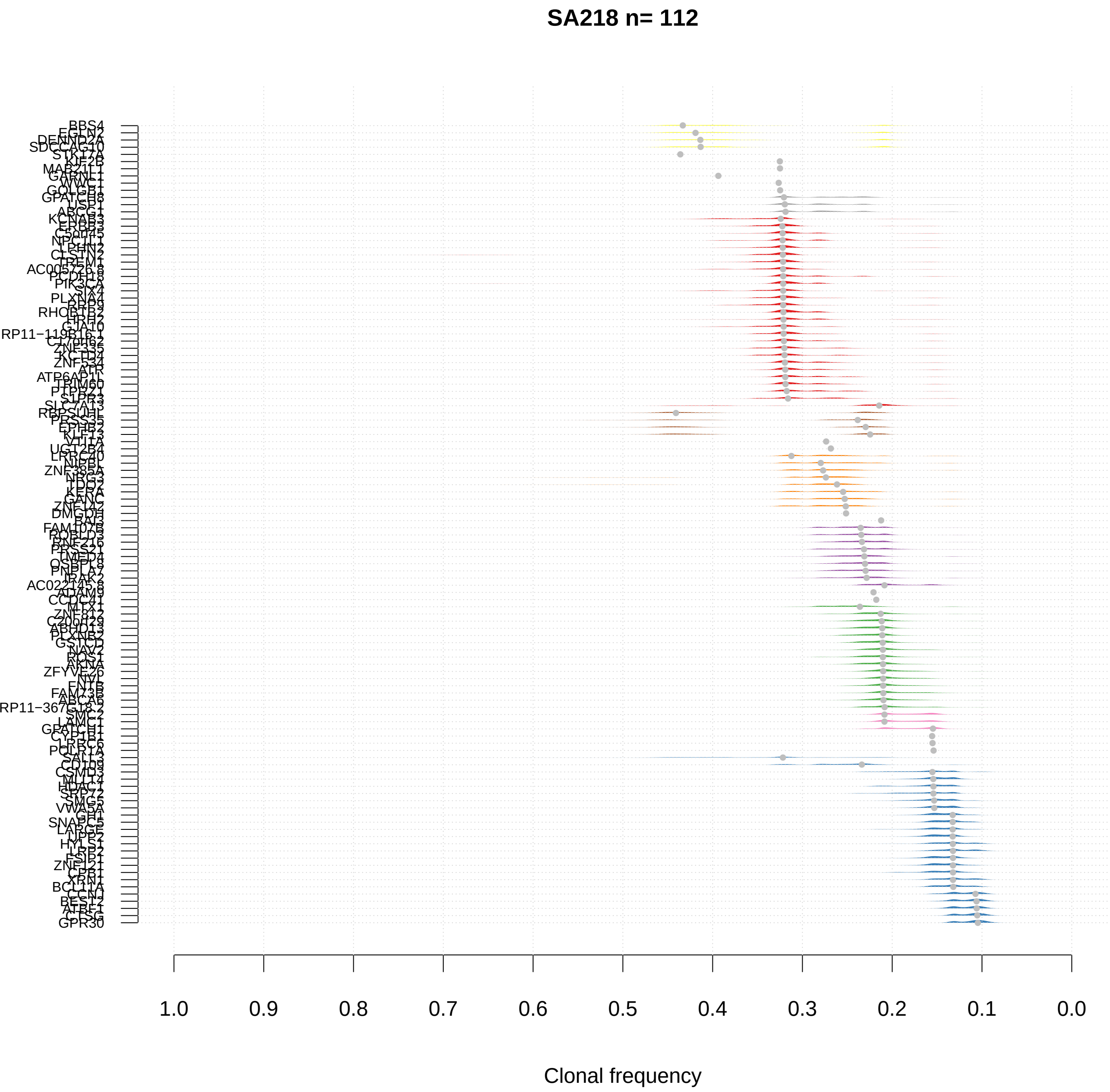

SA216 n= 16
QSER1
ATP5S
PTEN
GPR52
KCNJ16
RAG2
PLEKHM3
DNAH9
KCNJ16
DLG2
OSR1
MTX2
EPHA2
STAT3
PLCE1
KTN1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

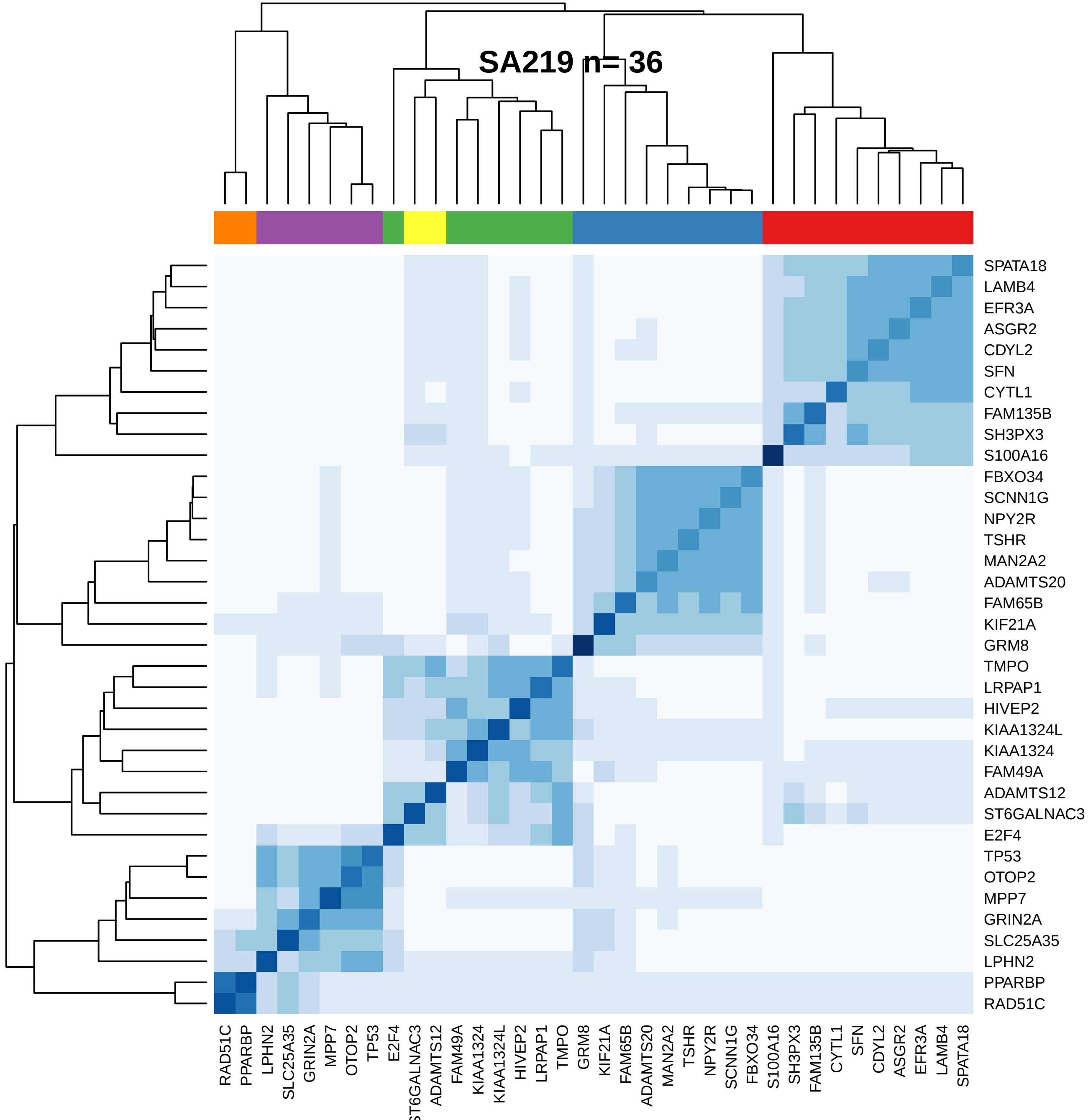

SA217 n= 14
BMP10
RAD51AP1
NEXN
LPIN2
FHOD3
SPATA13
SPOCK3
TRPM7
SLC24A4
FYCO1
FYCO1
AF235103.4−4
TP53
ZNF211
ZNF211
TP53
235103.4−4
FYCO1
FYCO1
SLC24A4
TRPM7
SPOCK3
SPATA13
FHOD3
LPIN2
NEXN
RAD51AP1
BMP10

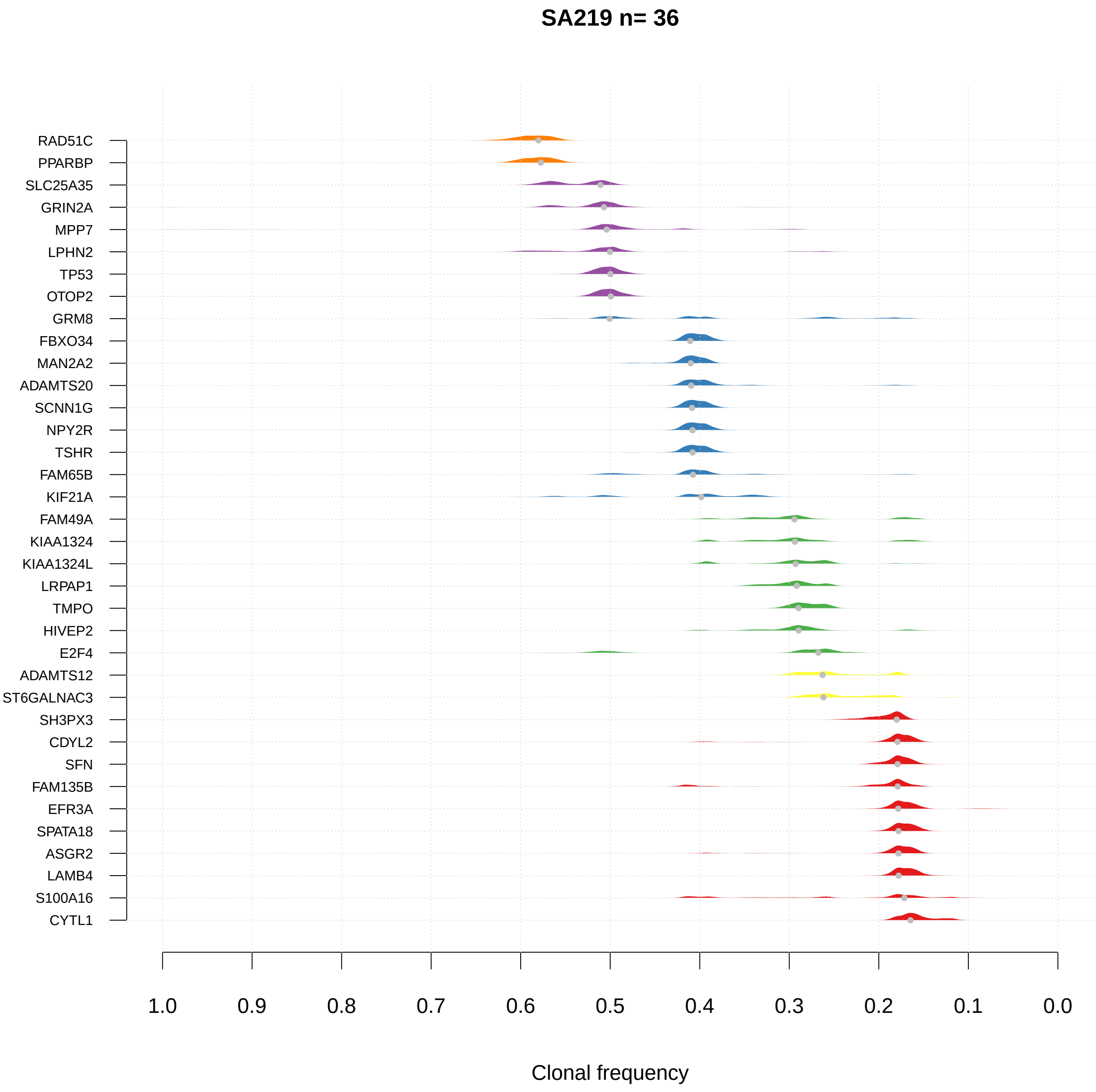

SA217 n= 14
ZNF211
TP53
AF235103.4−4
FYCO1
FYCO1
FHOD3
BMP10
NEXN
RAD51AP1
SLC24A4
LPIN2
SPOCK3
TRPM7
SPATA13
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

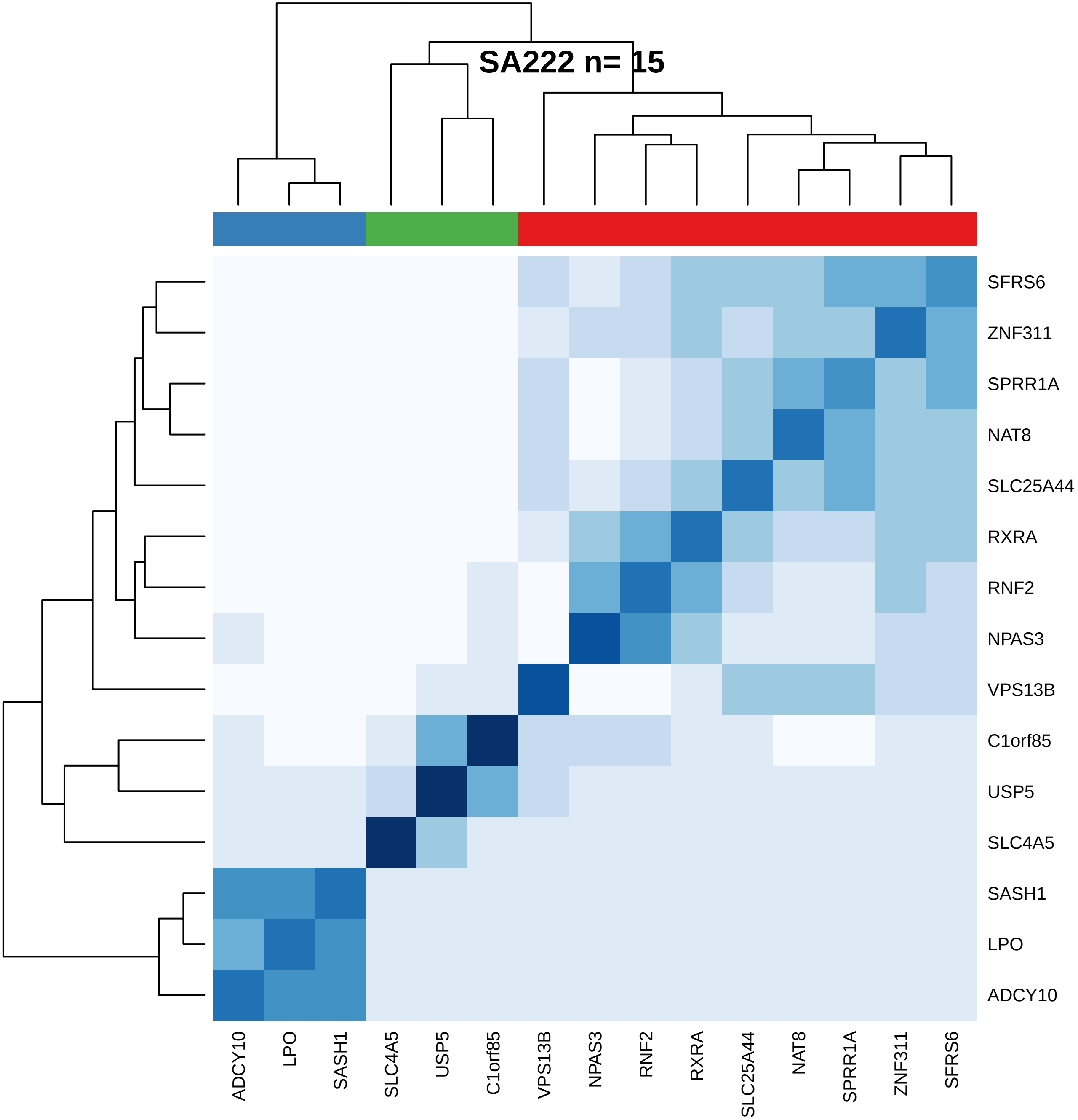


SA218 n= 112
GSTCD
C20orf29
ABHD13
AKNA
PLXNB2
ROS1
ZNF812
NVL
ZFYVE26
ABCA6
NAV2
ENTB
FAM73B
RP11−367G18.2
ROBLD3
FAM107B
RNF216
OSBPL8
PNPLA7
PRSS21
TMED4
IRAK2
MTX1
AC022145.8
FSIP1
UPP2
ZNF121
GH1
VWA5A
MLLT4
SNAPC5
CPB1
SMG5
LARGE
BCL11A
XRN1
LRP2
HYLS1
HDAC1
CSMD3
SRP72
ATBF1
BEST2
CTSG
GPR30
CCNJ
POLR1A
LRRC6
CYP1B1
LAMC1
SMC2
GPATCH1
CD109
TDO2
NRG3
ZNF385A
ZNF142
GANC
NIPBL
LRRC40
KERA
BAI3
DMGDH
TREM1
RRP9
ERBB3
CLSTN2
LPHN2
RP11−119B16.1
PLXNA4
AC005726.8
SIX4
GJA10
TRIM60
ZNF534
ATR
C17orf62
RHOBTB2
PIK3CA
C5orf45
HRH2
NPC1L1
ZNF335
KCTD4
ATP6AP1L
PTPRZ1
PCDH18
S1PR3
KCNAB3
UGT2B4
XTI1A
ABCG1
USP1
GPATCH8
WWC1
GARNL1
GOLGB1
MAB21L1
KIF2B
STK17A
EGLN2
SDCCAG10
BBS4
DENND2A
SLC7A13
KLF13
RBPSUHL
EPHB2
PRSS35
SALL3
ADAM9
CCDC41

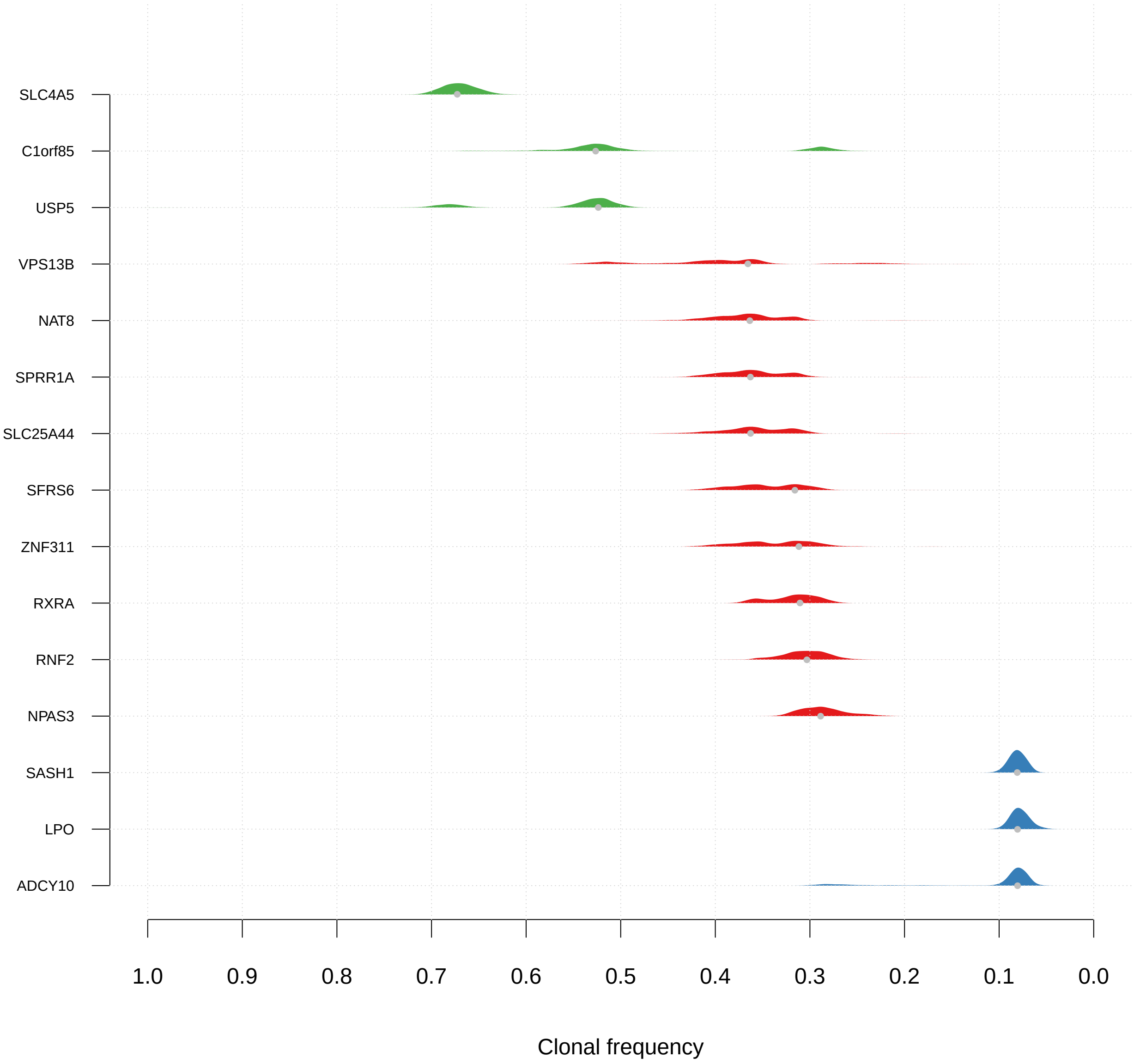

SA218 n= 112
BBS4
EGLN2
DENND2A
SDCCAG10
STK17A
KIF2B
MAB21L1
GARNL1
WWC1
GOLGB1
GPATCH8
USP1
ABCG1
KCNAB3
ERBB3
C5orf45
NPC1L1
LPHN2
CLSTN2
TREM1
AC005726.8
PCDH18
PIK3CA
SIX4
PLXNA4
RRP9
RHOBTB2
HRH2
GJA10
RP11−119B16.1
C17orf62
ZNF335
KCTD4
ZNF534
ATR
ATP6AP1L
TRIM60
PTPRZ1
S1PR3
SLC7A13
RBPSUHL
PRSS35
EPHB2
KLF13
VTI1A
UGT2B4
LRRC40
NIPBL
ZNF385A
NRG3
TDO2
KERA
GANC
ZNF142
DMGDH
BAI3
FAM107B
ROBLD3
RNF216
PRSS21
TMED4
OSBPL9
PNPLA7
IRAK2
AC022145.8
ADAM9
CCDC41
MTX1
ZNF812
C20orf29
ABHD13
PLXNB2
GSTCD
NAV2
ROS1
AKNA
ZFYVE26
NVL
FNTB
FAM73B
ABCA6
RP11−367G18.2
SMC2
LAMC1
GPATCH1
CYP1B1
LRRC6
POLR1A
SALL3
CD109
CSMD3
MLLT4
HDAC1
SRP72
SMG5
VWA5A
GH1
SNAPC5
LARGE
UPP2
HYLS1
LRP2
FSIP1
ZNF121
CPB1
XRN1
BCL11A
CCNJ
BEST2
ATBF1
CTSG
GPR30
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

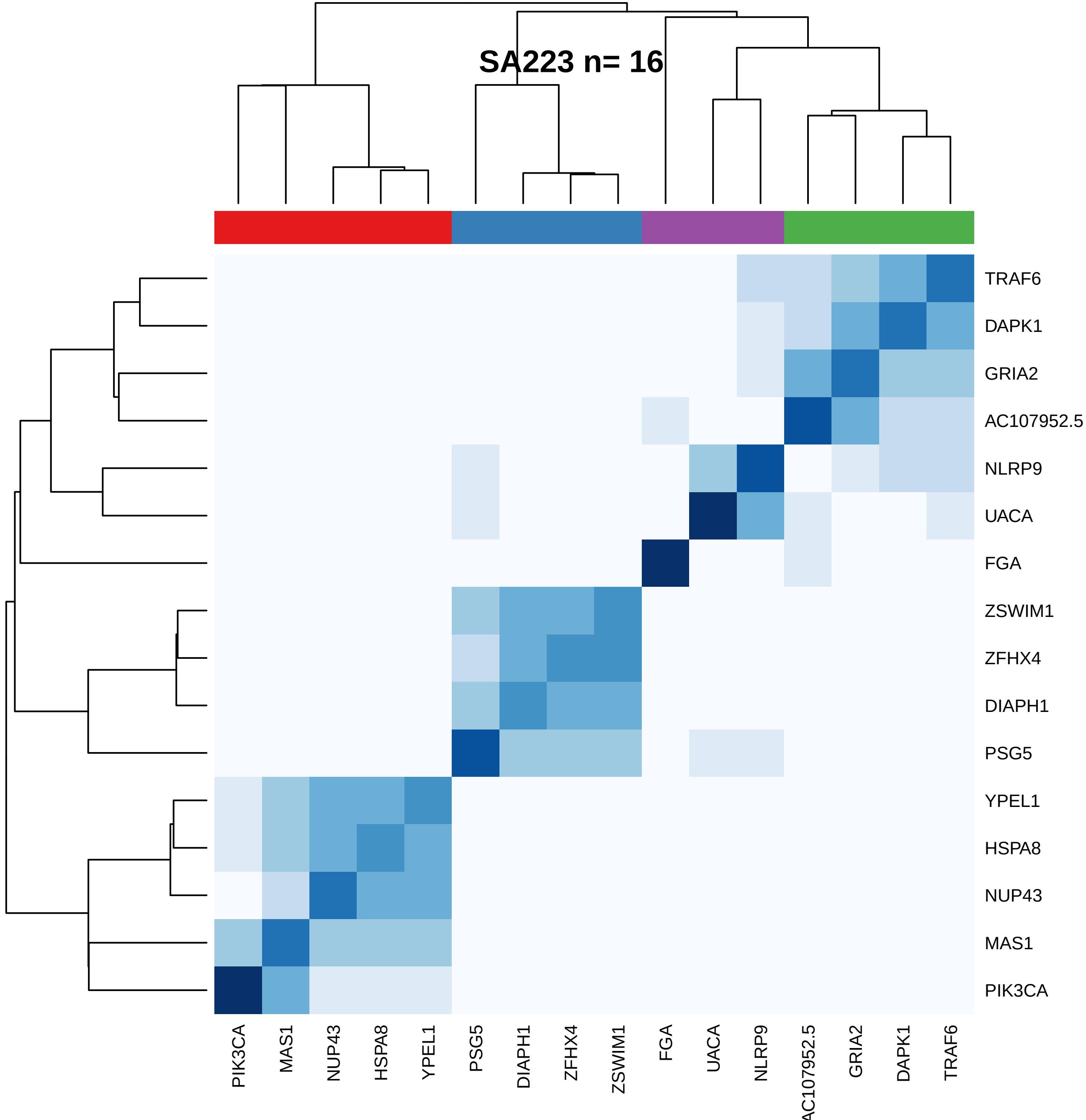

SA219 n= 36
SPATA18
LAMB4
EFR3A
ASGR2
CDYL2
SFN
CYTL1
FAM135B
SH3PX3
S100A16
FBXO34
SCNN1G
NPY2R
TSHR
MAN2A2
ADAMTS20
FAM65B
KIF21A
GRM8
TMPO
LRPAP1
HIVEP2
KIAA1324L
KIAA1324
FAM49A
ADAMTS12
ST6GALNAC3
E2F4
TP53
OTOP2
MPP7
GRIN2A
SLC25A35
LPHN2
PPARBP
RAD51C
RAD51C
PPARBP
LPHN2
SLC25A35
GRIN2A
MPP7
OTOP2
TP53
E2F4
ST6GALNAC3
ADAMTS12
FAM49A
KIAA1324
KIAA1324L
HIVEP2
LRPAP1
TMPO
GRM8
KIF21A
FAM65B
ADAMTS20
MAN2A2
TSHR
NPY2R
SCNN1G
FBXO34
S100A16
SH3PX3
FAM135B
CYTL1
SFN
CDYL2
ASGR2
EFR3A
LAMB4
SPATA18

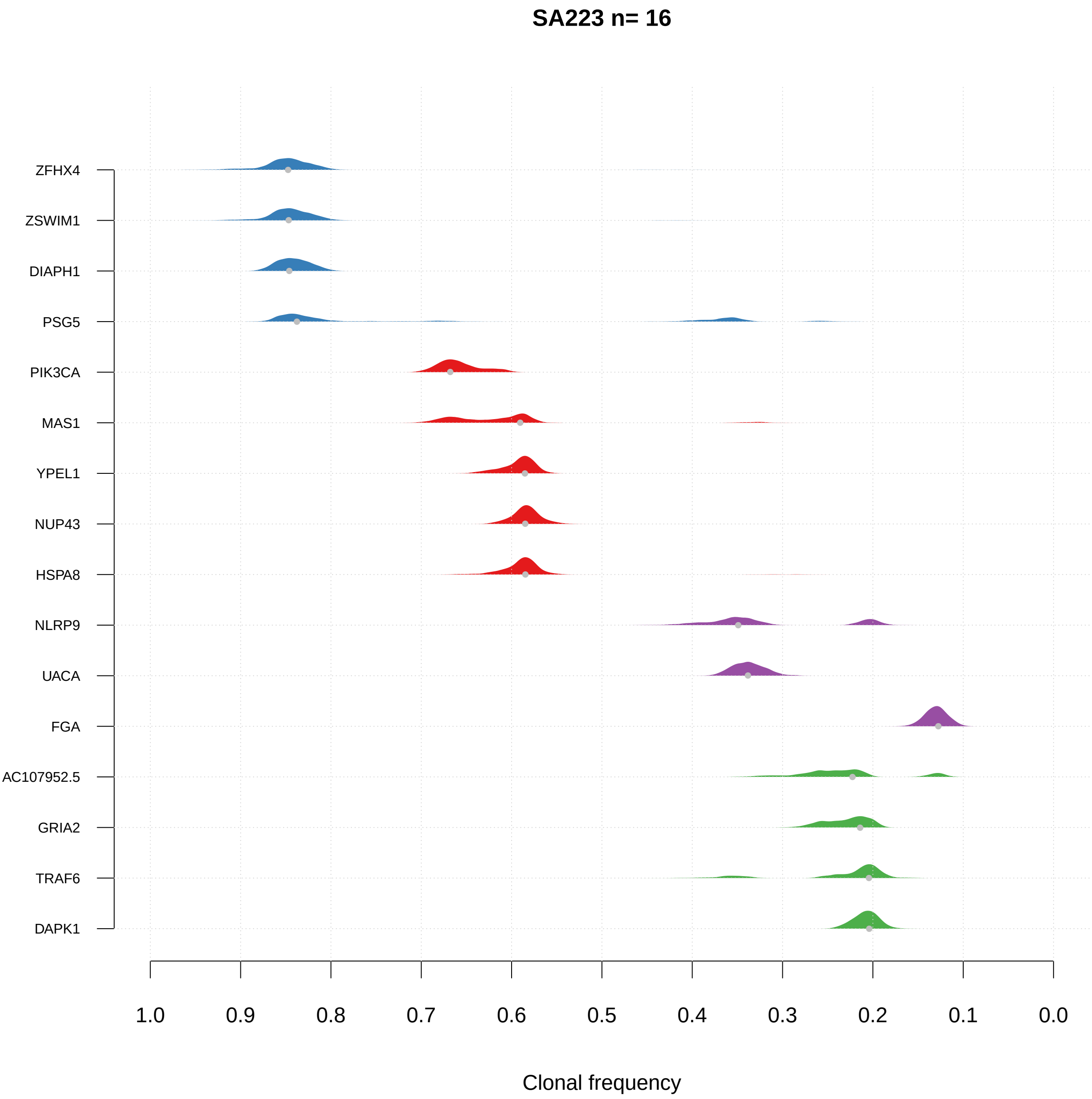

SA219 n= 36
RAD51C
PPARBP
SLC25A35
GRIN2A
MPP7
LPHN2
TP53
OTOP2
GRM8
FBXO34
MAN2A2
ADAMTS20
SCNN1G
NPY2R
TSHR
FAM65B
KIF21A
FAM49A
KIAA1324
KIAA1324L
LRPAP1
TMPO
HIVEP2
E2F4
ADAMTS12
ST6GALNAC3
SH3PX3
CDYL2
SFN
FAM135B
EFR3A
SPATA18
ASGR2
LAMB4
S100A16
CYTL1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

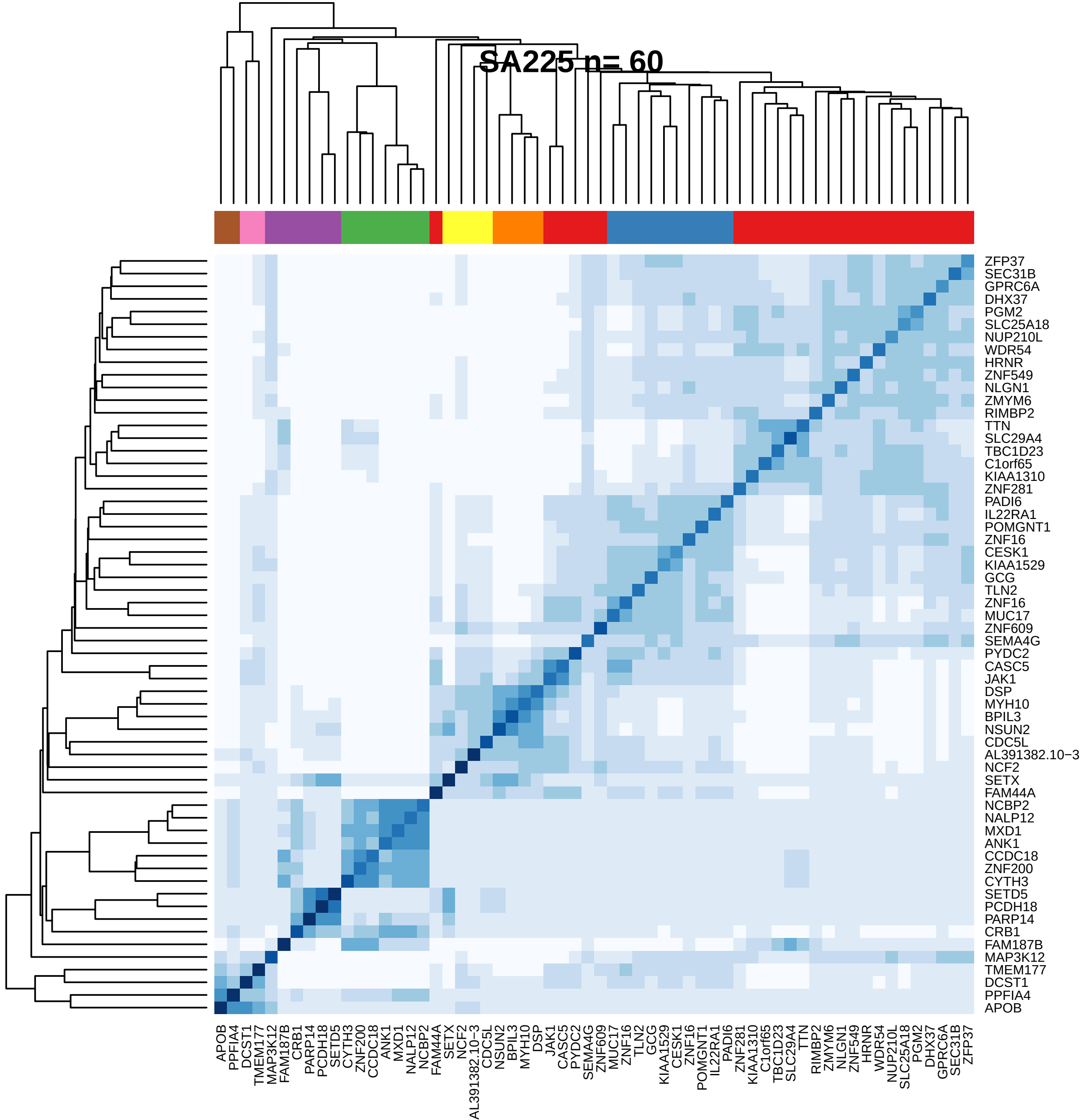

SA222 n= 15
SFRS6
ZNF311
SPRR1A
NAT8
SLC25A44
RXRA
RNF2
NPAS3
VPS13B
C1orf85
USP5
SLC4A5
SASH1
LPO
ADCY10
ADCY10
LPO
SASH1
SLC4A5
USP5
C1orf85
VPS13B
NPAS3
RNF2
RXRA
SLC25A44
NAT8
SPRR1A
ZNF311
SFRS6

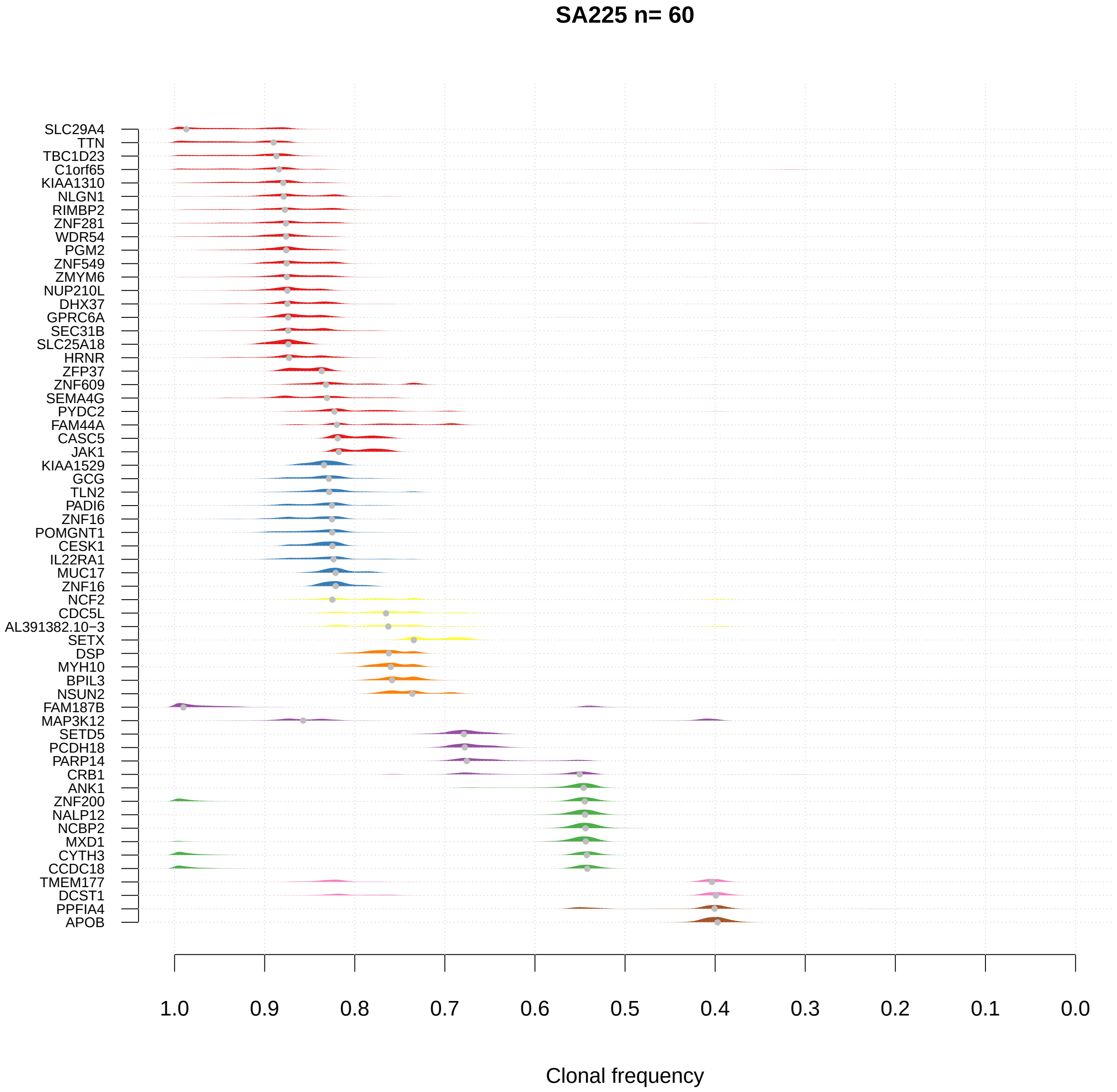
SA222 n= 15

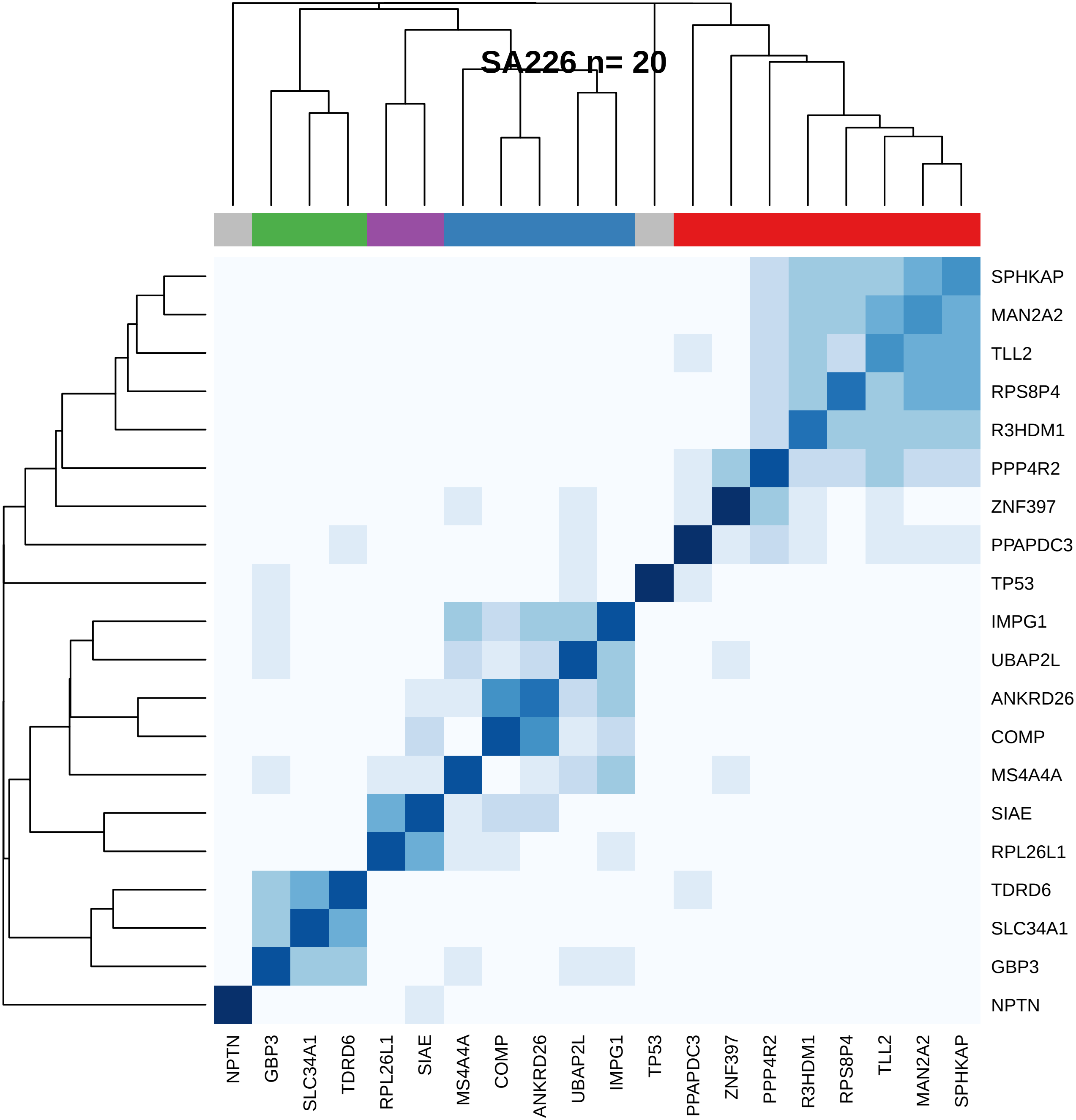

SLC4A5
C1orf85
USP5
VPS13B
NAT8
SPRR1A
SLC25A44
SFRS6
ZNF311
RXRA
RNF2
NPAS3
SASH1
LPO
ADCY10
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

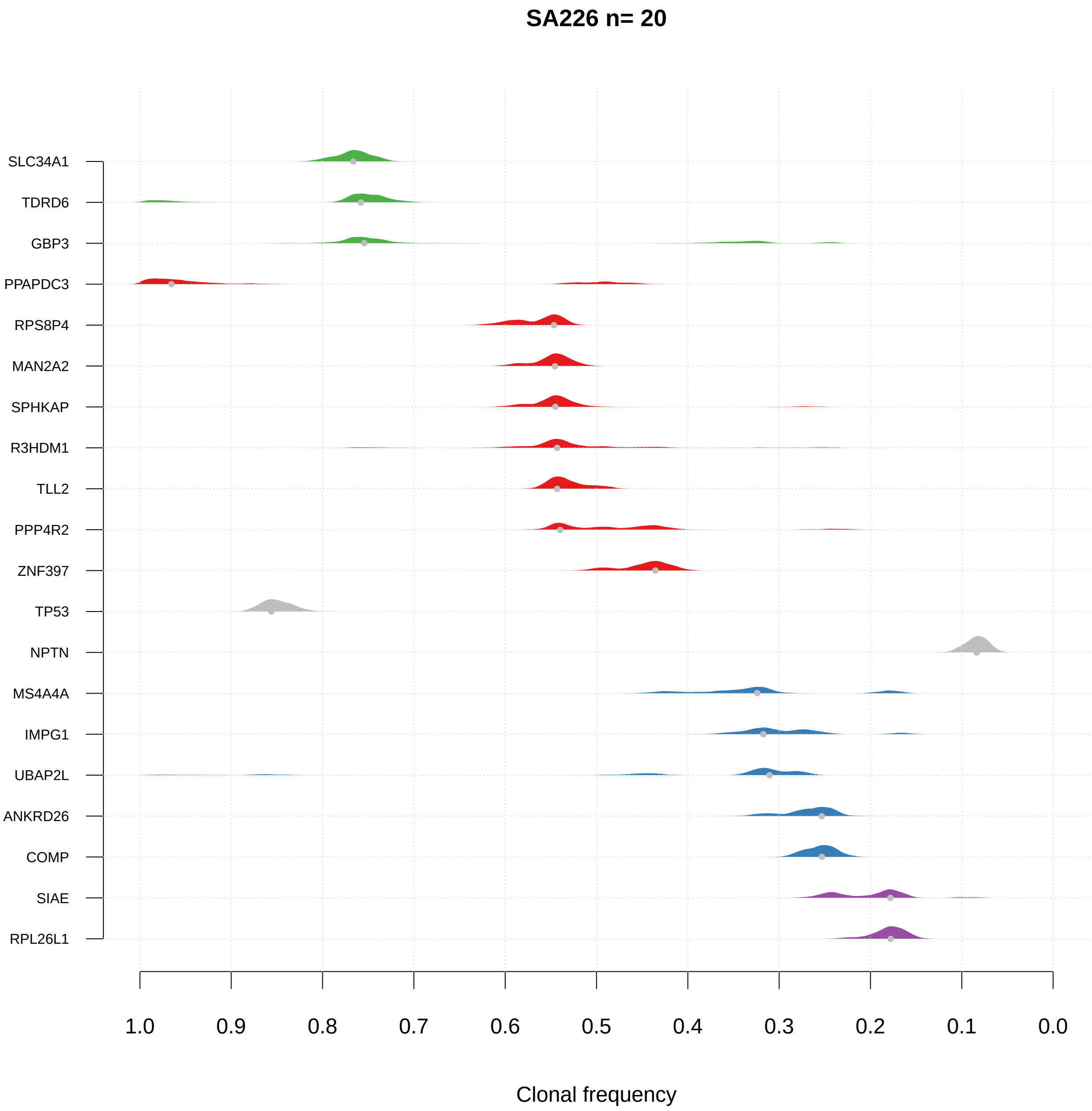

SA223 n= 16
TRAF6
DAPK1
GRIA2
AC107952.5
NLRP9
UACA
FGA
ZSWIM1
ZFHX4
DIAPH1
PSG5
YPEL1
HSPA8
NUP43
MAS1
PIK3CA
PIK3CA
MAS1
NUP43
HSPA8
YPEL1
PSG5
DIAPH1
ZFHX4
ZSWIM1
FGA
UACA
NLRP9
AC107952.5
GRIA2
DAPK1
TRAF6

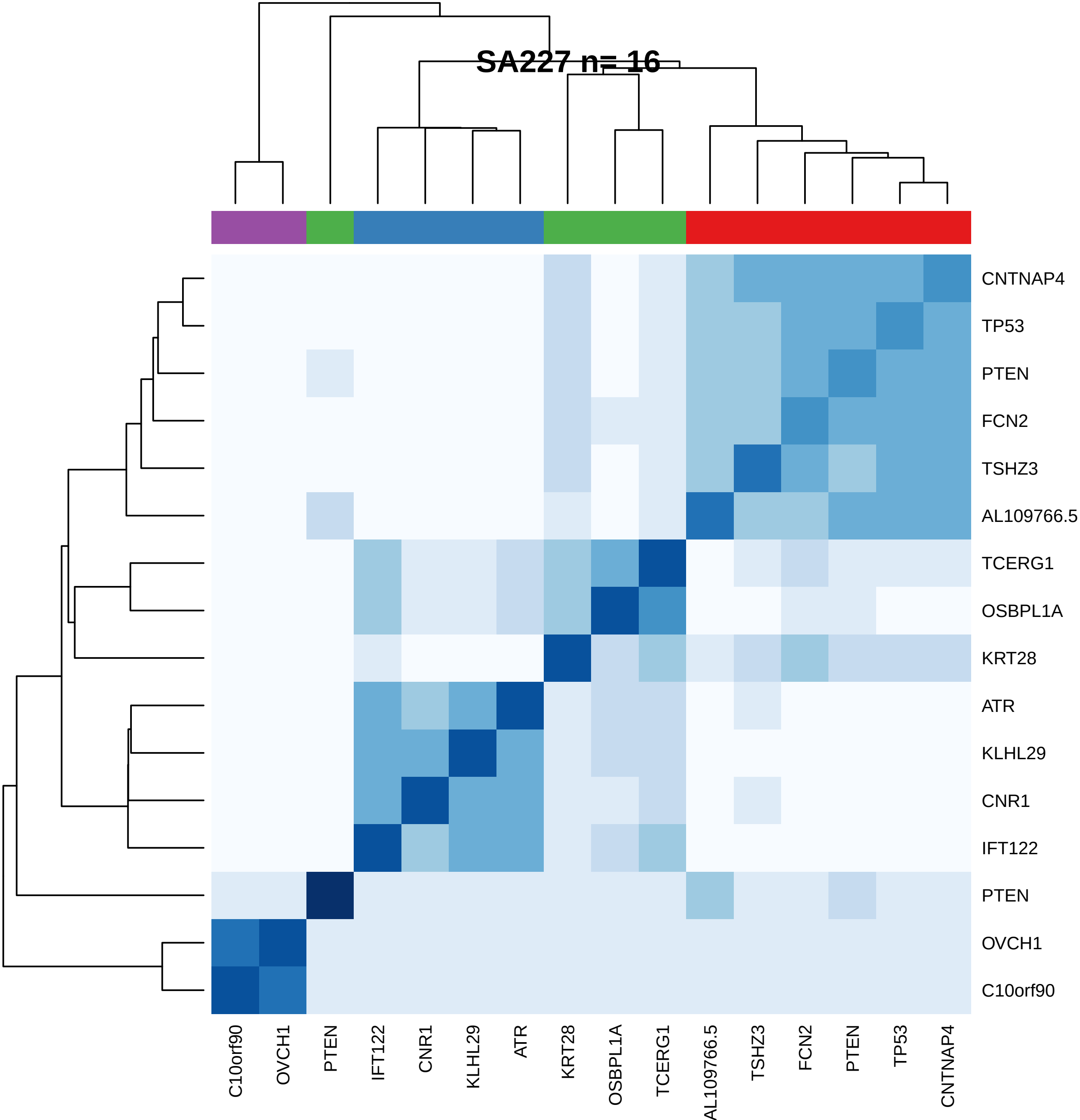

SA223 n= 16
ZFHX4
ZSWIM1
DIAPH1
PSG5
PIK3CA
MAS1
YPEL1
NUP43
HSPA8
NLRP9
UACA
FGA
AC107952.5
GRIA2
TRAF6
DAPK1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

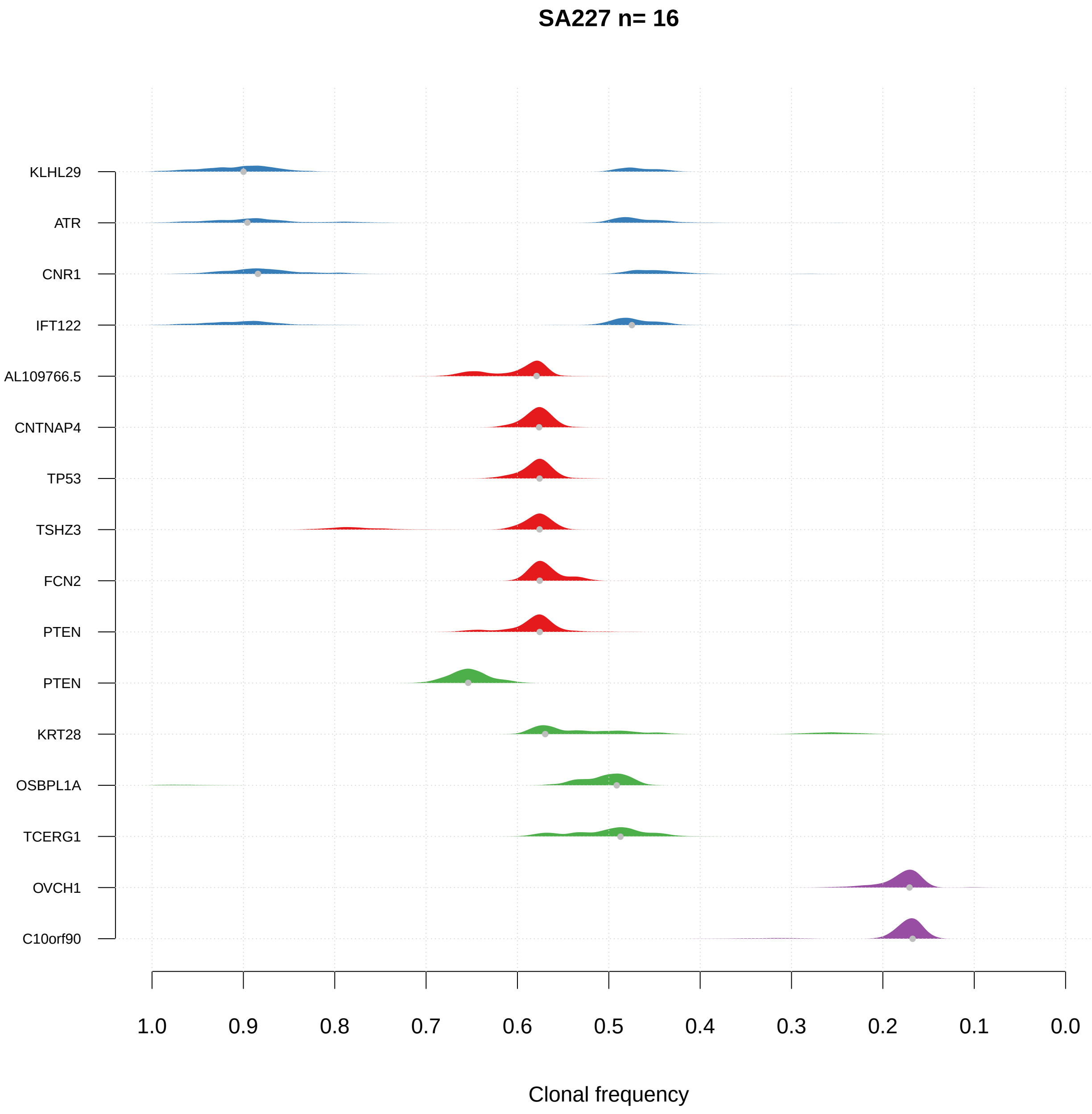

SA225 n= 60
ZFP37
SEC31B
GPRC6A
DHX37
PGM2
SLC25A18
NUP210L
WDR54
HRNR
ZNF549
NLGN1
ZMYM6
RIMBP2
TTN
SLC29A4
TBC1D23
C1orf65
KIAA1310
ZNF281
PADI6
IL22RA1
POMGNT1
ZNF16
CESK1
KIAA1529
GCG
TLN2
ZNF16
MUC17
ZNF609
SEMA4G
PYDC2
CASC5
JAK1
DSP
MYH10
BPIL3
NSUN2
CDC5L
AL391382.10−3
NCF2
SETX
FAM44A
NCBP2
NALP12
MXD1
ANK1
CCDC18
ZNF200
CYTH3
SETD5
PCDH18
PARP14
CRB1
FAM187B
MAP3K12
TMEM177
DCST1
PPFIA4
APOB
APOB
PPFIA4
DCST1
TMEM177
MAP3K12
FAM187B
CRB1
PARP14
PCDH18
SETD5
CYTH3
ZNF200
CCDC18
ANK1
MXD1
NALP12
NCBP2
FAM44A
SETX
NCF2
AL391382.10−3
CDC5L
NSUN2
BPIL3
MYH10
DSP
JAK1
CASC5
PYDC2
SEMA4G
ZNF609
MUC17
ZNF16
TLN2
GCG
KIAA1529
CESK1
ZNF16
POMGNT1
IL22RA1
PADI6
ZNF281
KIAA1310
C1orf65
TBC1D23
SLC29A4
TTN
RIMBP2
ZMYM6
NLGN1
ZNF549
HRNR
WDR54
NUP210L
SLC25A18
PGM2
DHX37
GPRC6A
SEC31B
ZFP37

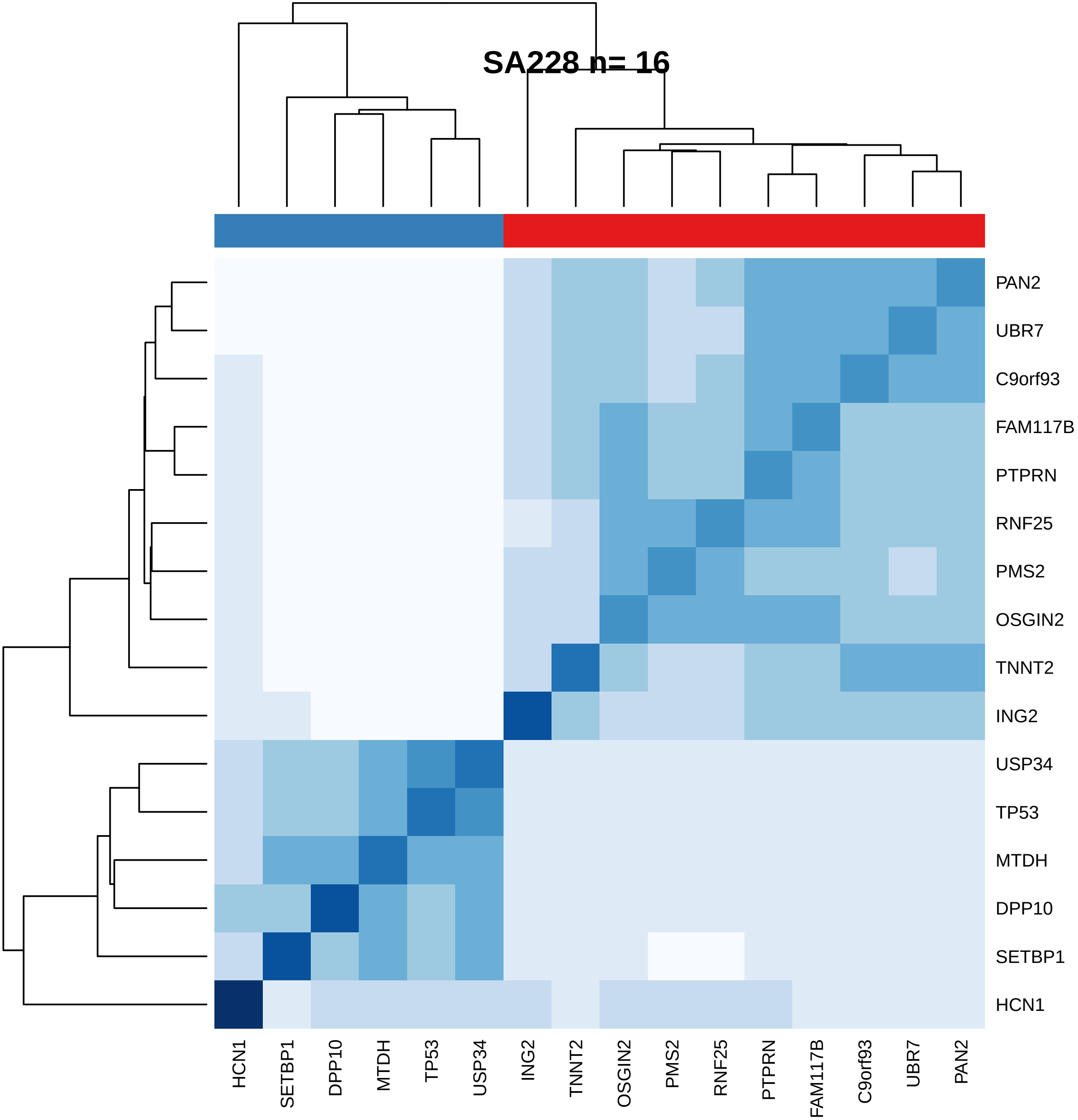

SA225 n= 60
SLC29A4
TTN
TBC1D23
C1orf65
KIAA1310
NLGN1
RIMBP2
ZNF281
WDR54
PGM2
ZNF549
ZMYM6
NUP210L
DHX37
GPRC6A
SEC31B
SLC25A18
HRNR
ZFP37
ZNF609
SEMA4G
PYDC2
FAM44A
CASC5
JAK1
KIAA1529
GCG
TLN2
PADI6
ZNF16
POMGNT1
CESK1
IL22RA1
MUC17
ZNF16
NCF2
CDC5L
AL391382.10−3
SETX
DSP
MYH10
BPIL3
NSUN2
FAM187B
MAP3K12
SETD5
PCDH18
PARP14
CRB1
ANK1
ZNF200
NALP12
NCBP2
MXD1
CYTH3
CCDC18
TMEM177
DCST1
PPFIA4
APOB
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

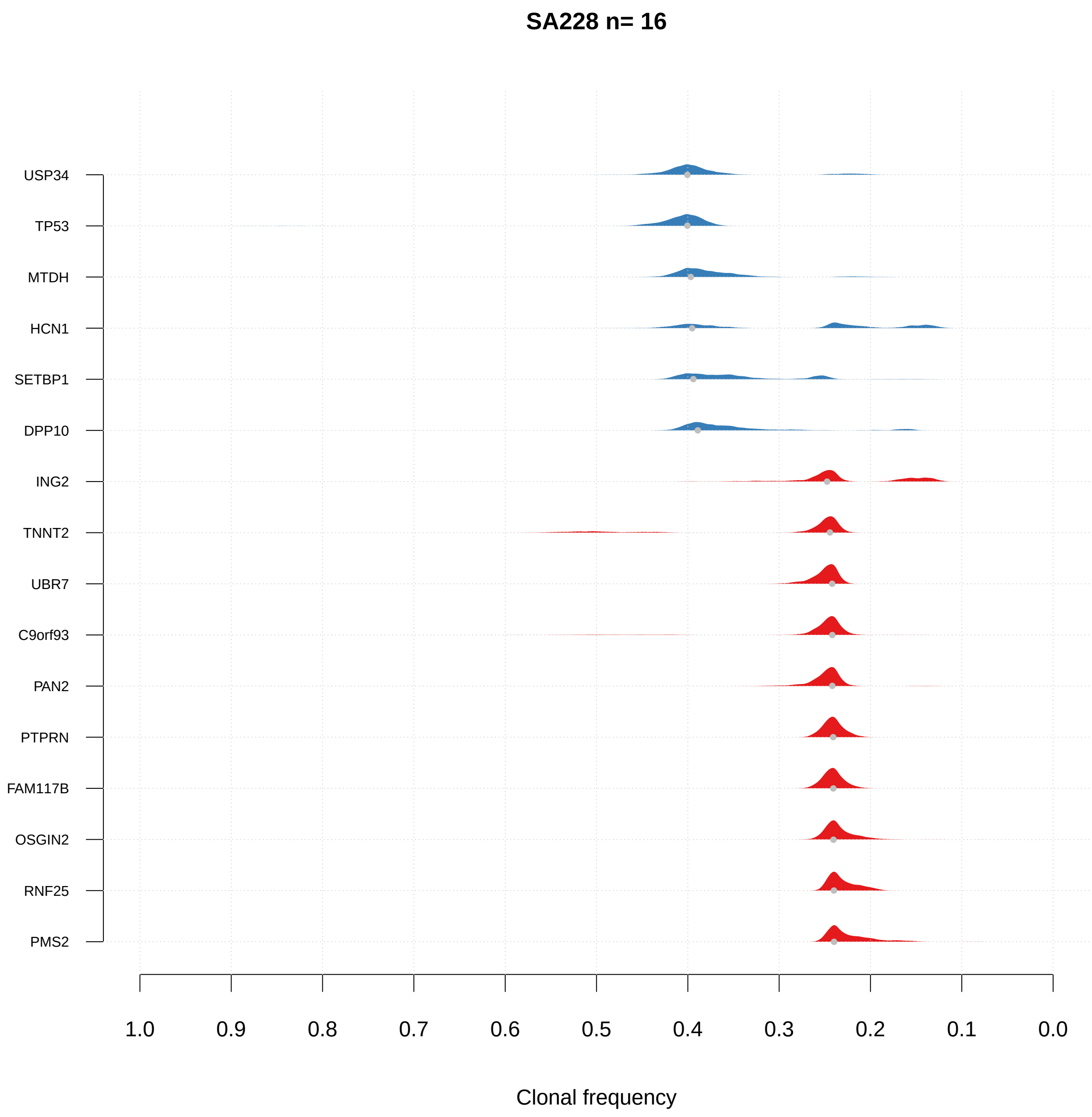

SA226 n= 20
SPHKAP
MAN2A2
TLL2
RPS8P4
R3HDM1
PPP4R2
ZNF397
PPAPDC3
TP53
IMPG1
UBAP2L
ANKRD26
COMP
MS4A4A
SIAE
RPL26L1
TDRD6
SLC34A1
GBP3
NPTN
NPTN
GBP3
SLC34A1
TDRD6
RPL26L1
SIAE
MS4A4A
COMP
ANKRD26
UBAP2L
IMPG1
TP53
PPAPDC3
ZNF397
PPP4R2
R3HDM1
RPS8P4
TLL2
MAN2A2
SPHKAP

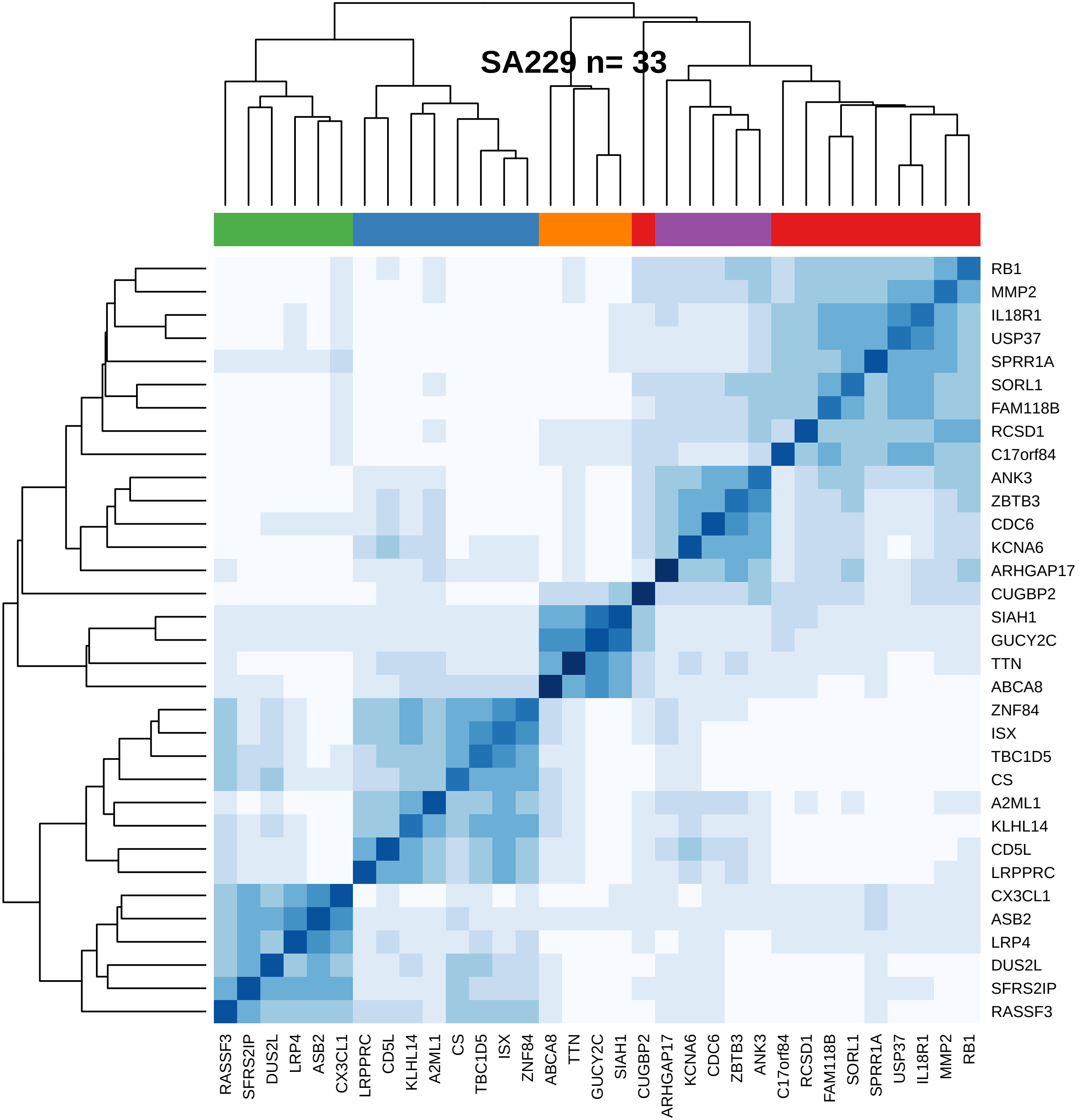

SA226 n= 20
SLC34A1
TDRD6
GBP3
PPAPDC3
RPS8P4
MAN2A2
SPHKAP
R3HDM1
TLL2
PPP4R2
ZNF397
TP53
NPTN
MS4A4A
IMPG1
UBAP2L
ANKRD26
COMP
SIAE
RPL26L1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA227 n= 16
CNTNAP4
TP53
PTEN
FCN2
TSHZ3
AL109766.5
TCERG1
OSBPL1A
KRT28
ATR
KLHL29
CNR1
IFT122
PTEN
OVCH1
C10orf90
C10orf90
OVCH1
PTEN
IFT122
CNR1
KLHL29
ATR
KRT28
OSBPL1A
TCERG1
AL109766.5
TSHZ3
FCN2
PTEN
TP53
CNTNAP4

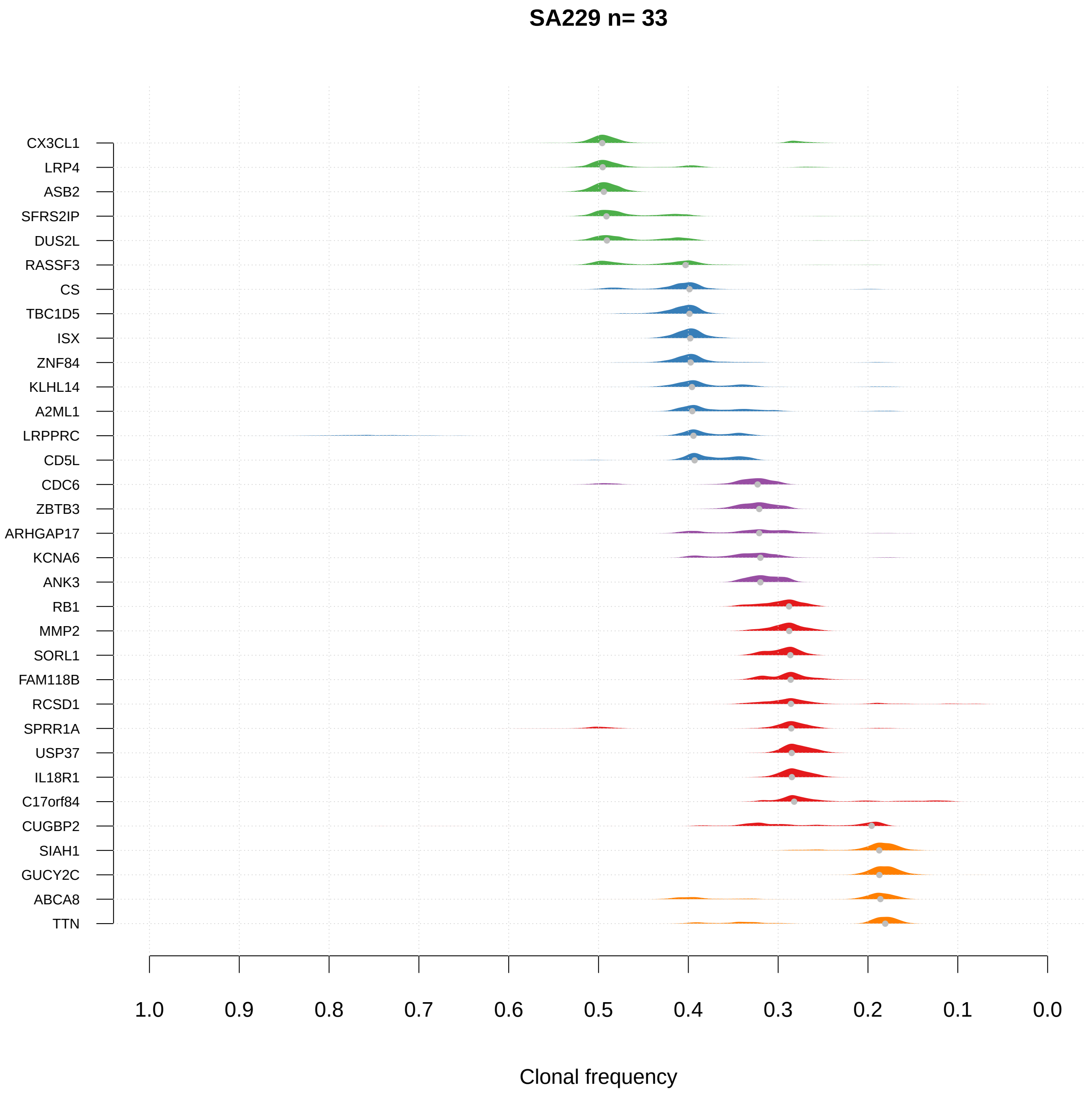

SA227 n= 16
KLHL29
ATR
CNR1
IFT122
AL109766.5
CNTNAP4
TP53
TSHZ3
FCN2
PTEN
PTEN
KRT28
OSBPL1A
TCERG1
OVCH1
C10orf90
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

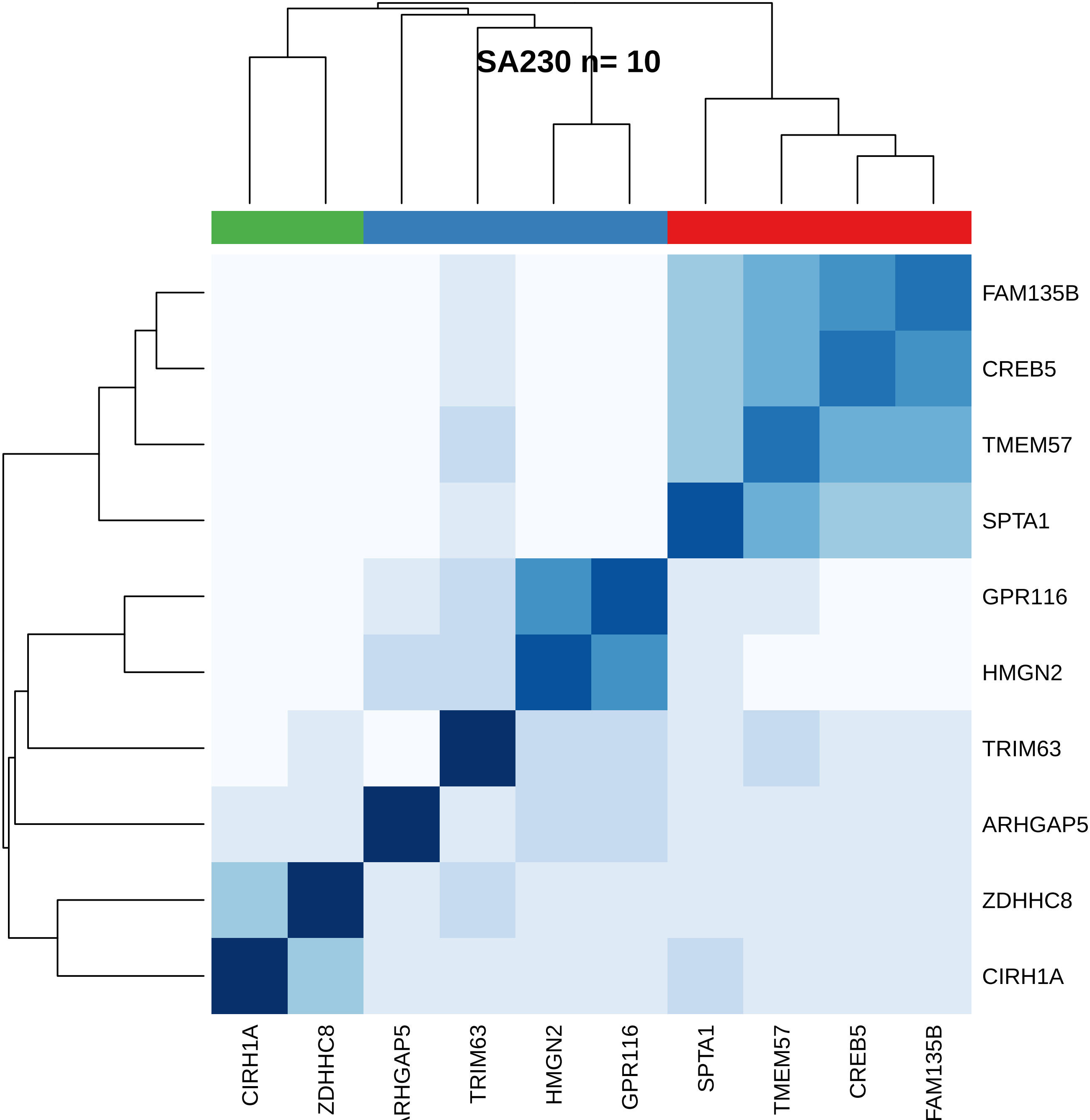

SA228 n= 16
PAN2
UBR7
C9orf93
FAM117B
PTPRN
RNF25
PMS2
OSGIN2
TNNT2
ING2
USP34
TP53
MTDH
DPP10
SETBP1
HCN1
HCN1
SETBP1
DPP10
MTDH
TP53
USP34
ING2
TNNT2
OSGIN2
PMS2
RNF25
PTPRN
FAM117B
C9orf93
UBR7
PAN2


SA228 n= 16
USP34
TP53
MTDH
HCN1
SETBP1
DPP10
ING2
TNNT2
UBR7
C9orf93
PAN2
PTPRN
FAM117B
OSGIN2
RNF25
PMS2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

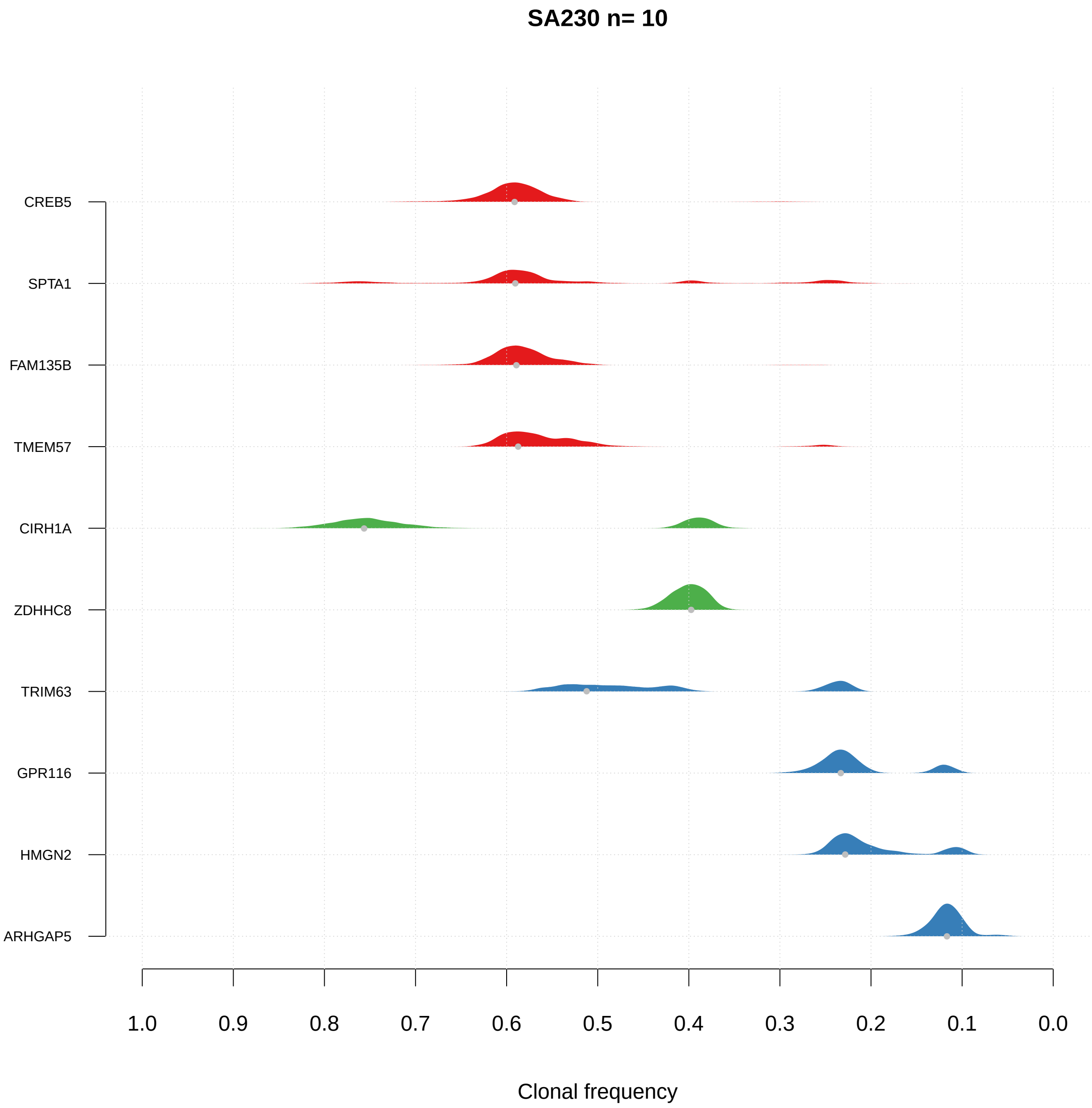


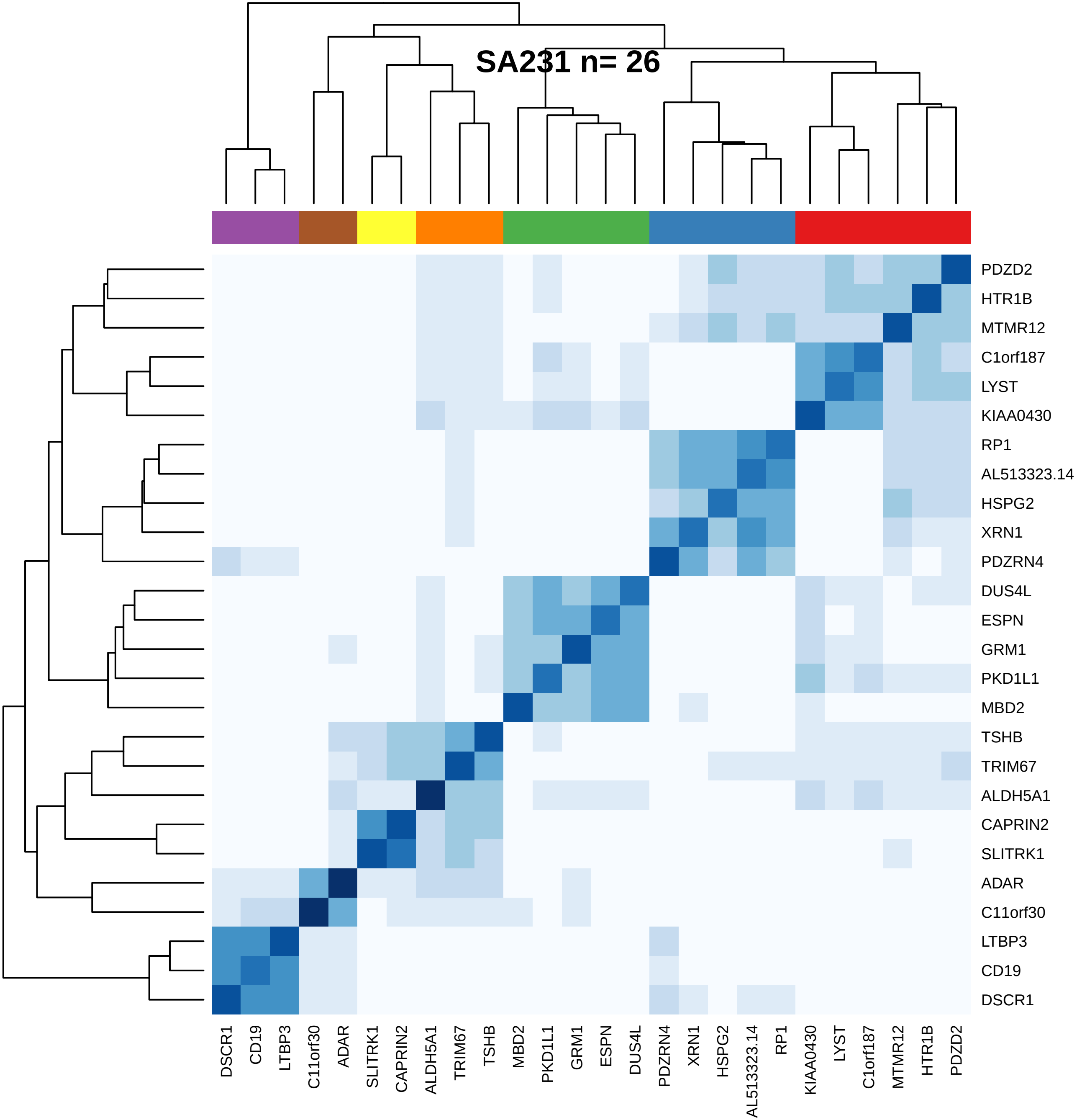
SA229 n= 33
RB1
MMP2
IL18R1
USP37
SPRR1A
SORL1
FAM118B
RCSD1
C17orf84
ANK3
ZBTB3
CDC6
KCNA6
ARHGAP17
CUGBP2
SIAH1
GUCY2C
TTN
ABCA8
ZNF84
ISX
TBC1D5
CS
A2ML1
KLHL14
CD5L
LRPPRC
CX3CL1
ASB2
LRP4
DUS2L
SFRS2IP
RASSF3
RASSF3
SFRS2IP
DUS2L
LRP4
ASB2
CX3CL1
LRPPRC
CD5L
KLHL14
A2ML1
CS
TBC1D5
ISX
ZNF84
ABCA8
TTN
GUCY2C
SIAH1
CUGBP2
ARHGAP17
KCNA6
CDC6
ZBTB3
ANK3
C17orf84
RCSD1
FAM118B
SORL1
SPRR1A
USP37
IL18R1
MMP2
RB1

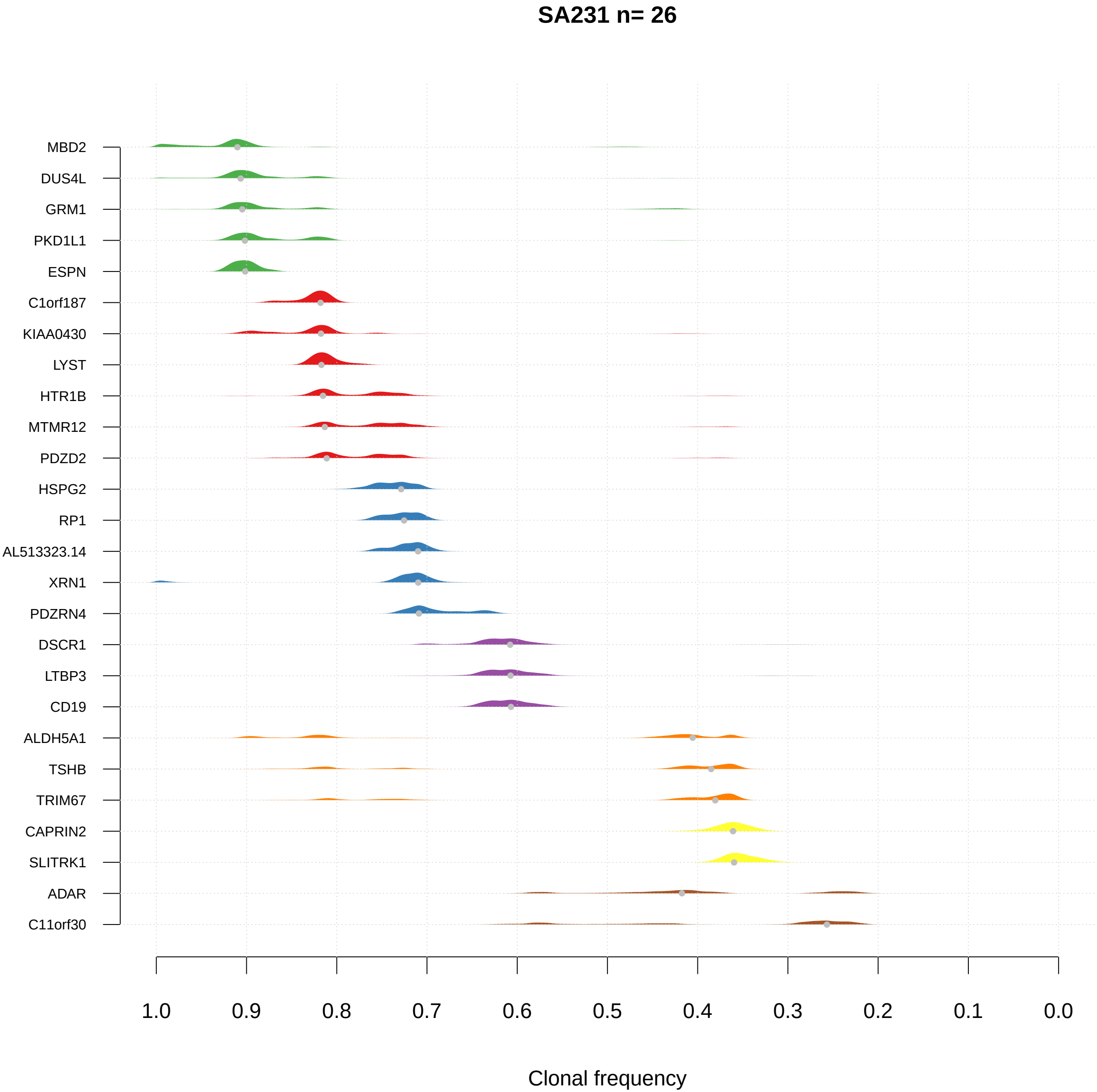

SA229 n= 33
CX3CL1
LRP4
ASB2
SFRS2IP
DUS2L
RASSF3
CS
TBC1D5
ISX
ZNF84
KLHL14
A2ML1
LRPPRC
CD5L
CDC6
ZBTB3
ARHGAP17
KCNA6
ANK3
RB1
MMP2
SORL1
FAM118B
RCSD1
SPRR1A
USP37
IL18R1
C17orf84
CUGBP2
SIAH1
GUCY2C
ABCA8
TTN
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

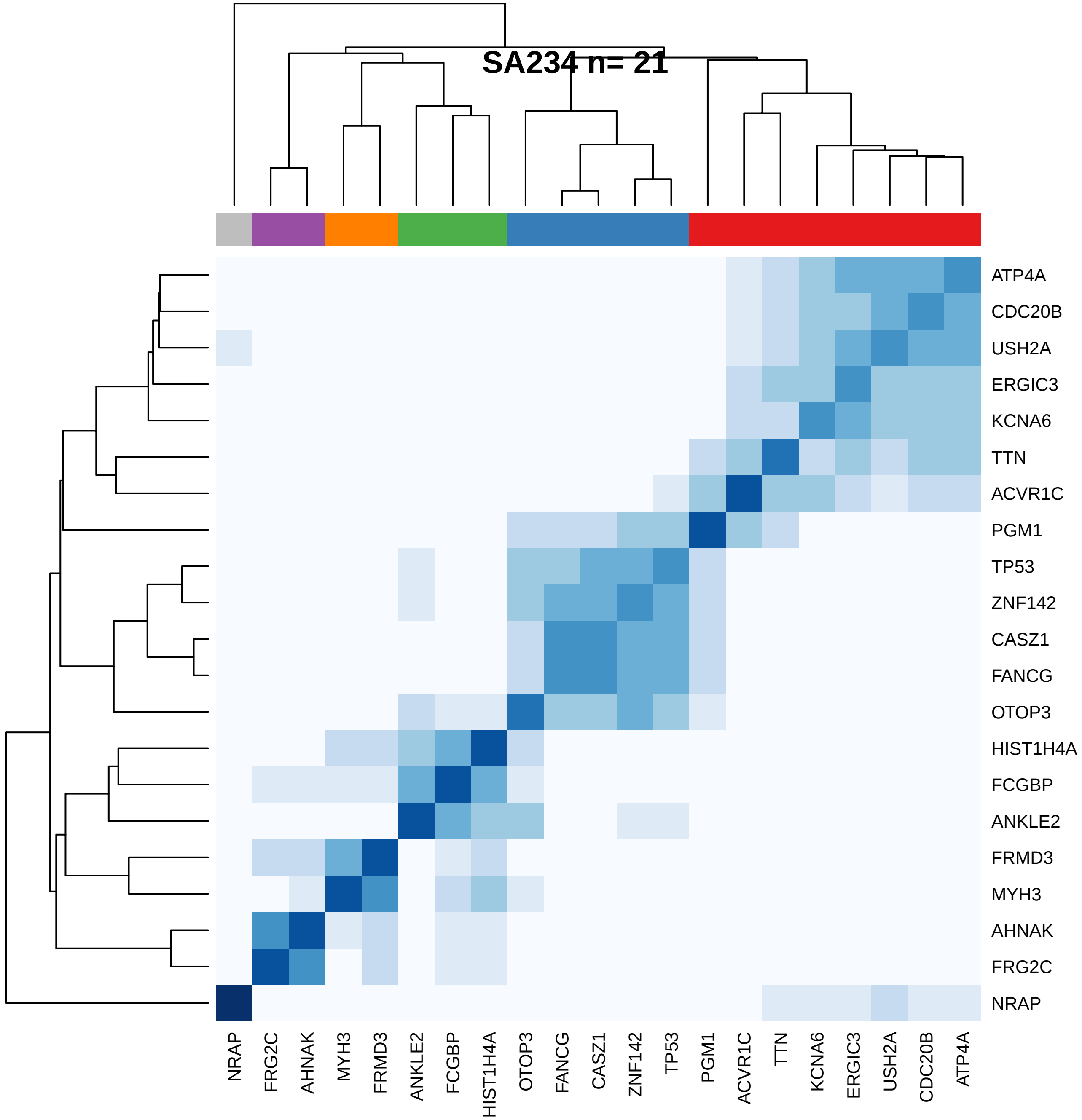


SA230 n= 10
FAM135B
CREB5
TMEM57
SPTA1
GPR116
HMGN2
TRIM63
ARHGAP5
ZDHHC8
CIRH1A
CIRH1A
ZDHHC8
RHGAP5
TRIM63
HMGN2
GPR116
SPTA1
TMEM57
CREB5
FAM135B

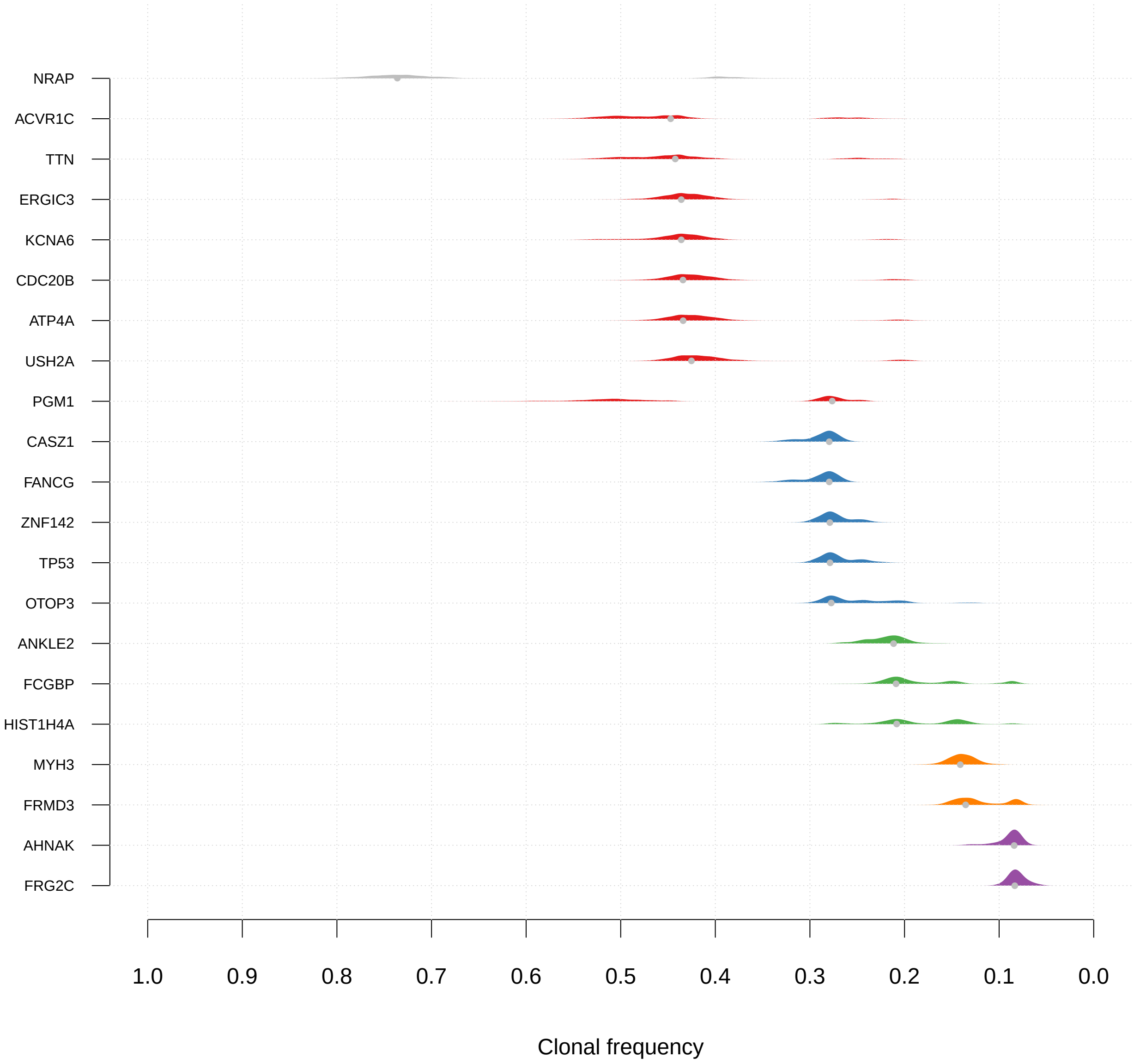

SA230 n= 10
CREB5
SPTA1
FAM135B
TMEM57
CIRH1A
ZDHHC8
TRIM63
GPR116
HMGN2
ARHGAP5
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

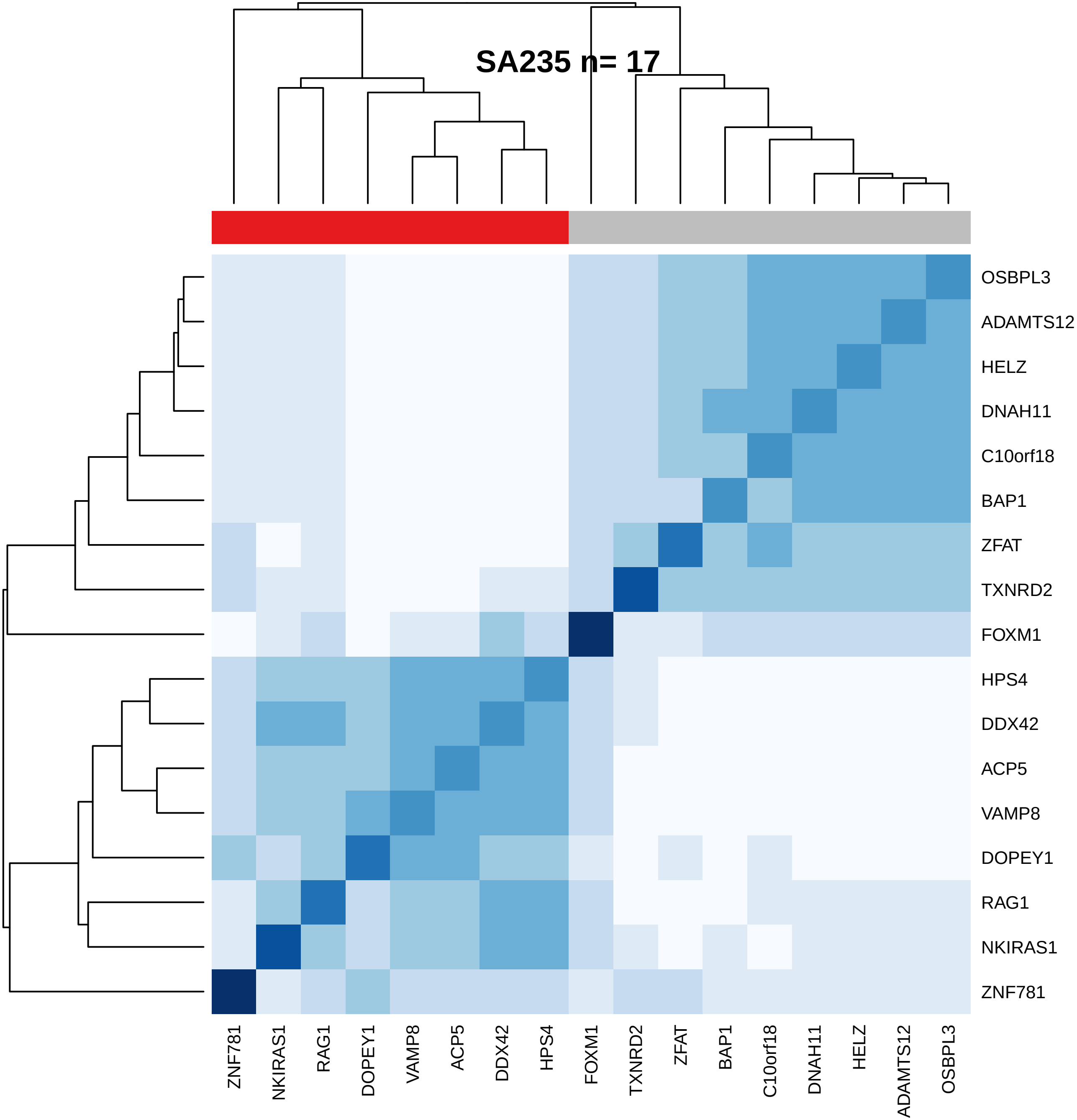

SA231 n= 26
PDZD2
HTR1B
MTMR12
C1orf187
LYST
KIAA0430
RP1
AL513323.14
HSPG2
XRN1
PDZRN4
DUS4L
ESPN
GRM1
PKD1L1
MBD2
TSHB
TRIM67
ALDH5A1
CAPRIN2
SLITRK1
ADAR
C11orf30
LTBP3
CD19
DSCR1
DSCR1
CD19
LTBP3
C11orf30
ADAR
SLITRK1
CAPRIN2
ALDH5A1
TRIM67
TSHB
MBD2
PKD1L1
GRM1
ESPN
DUS4L
PDZRN4
XRN1
HSPG2
AL513323.14
RP1
KIAA0430
LYST
C1orf187
MTMR12
HTR1B
PDZD2


SA231 n= 26
MBD2
DUS4L
GRM1
PKD1L1
ESPN
C1orf187
KIAA0430
LYST
HTR1B
MTMR12
PDZD2
HSPG2
RP1
AL513323.14
XRN1
PDZRN4
DSCR1
LTBP3
CD19
ALDH5A1
TSHB
TRIM67
CAPRIN2
SLITRK1
ADAR
C11orf30
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

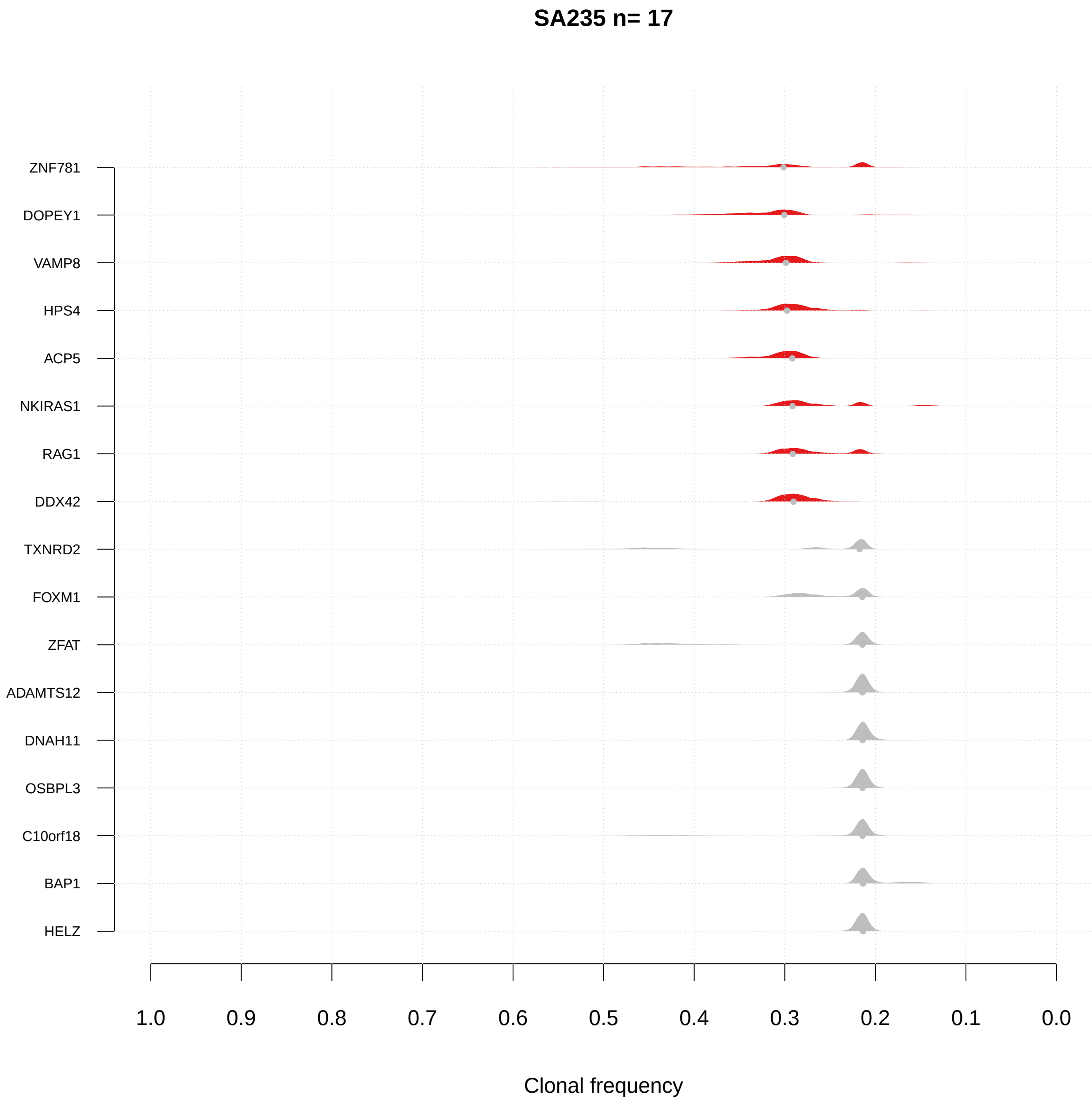

SA234 n= 21
ATP4A
CDC20B
USH2A
ERGIC3
KCNA6
TTN
ACVR1C
PGM1
TP53
ZNF142
CASZ1
FANCG
OTOP3
HIST1H4A
FCGBP
ANKLE2
FRMD3
MYH3
AHNAK
FRG2C
NRAP
NRAP
FRG2C
AHNAK
MYH3
FRMD3
ANKLE2
FCGBP
HIST1H4A
OTOP3
FANCG
CASZ1
ZNF142
TP53
PGM1
ACVR1C
TTN
KCNA6
ERGIC3
USH2A
CDC20B
ATP4A

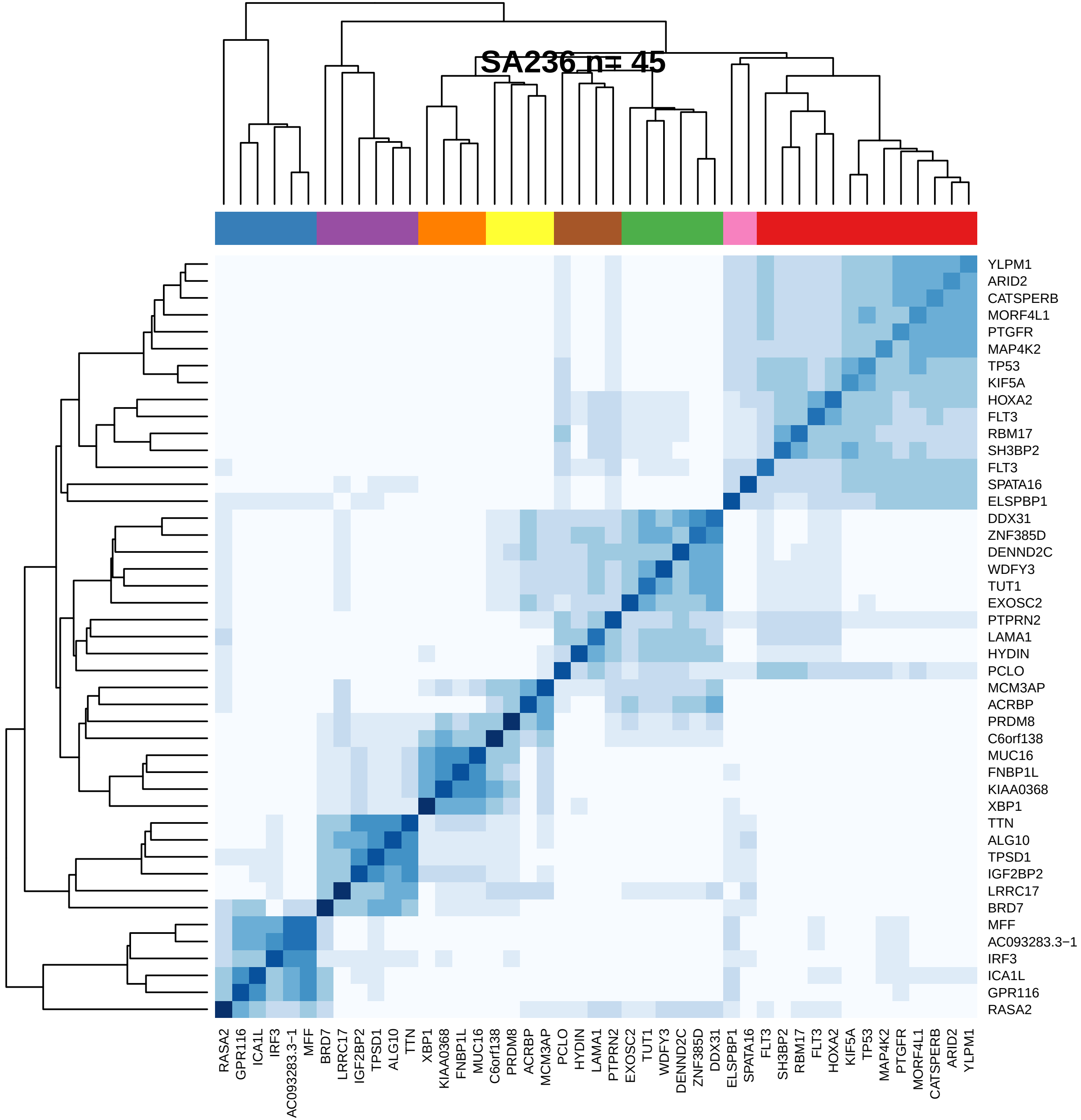
SA234 n= 21

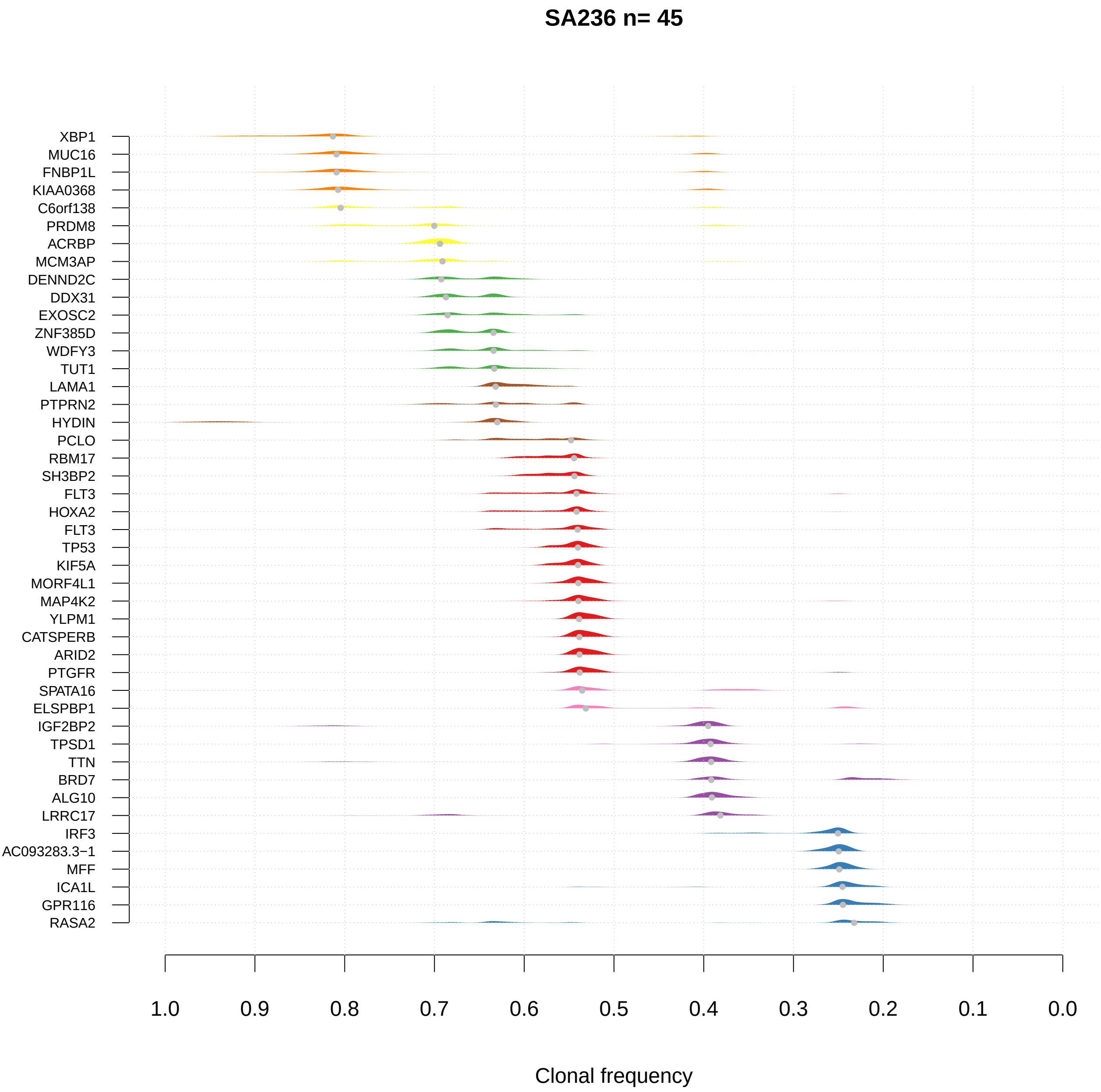

NRAP
ACVR1C
TTN
ERGIC3
KCNA6
CDC20B
ATP4A
USH2A
PGM1
CASZ1
FANCG
ZNF142
TP53
OTOP3
ANKLE2
FCGBP
HIST1H4A
MYH3
FRMD3
AHNAK
FRG2C
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

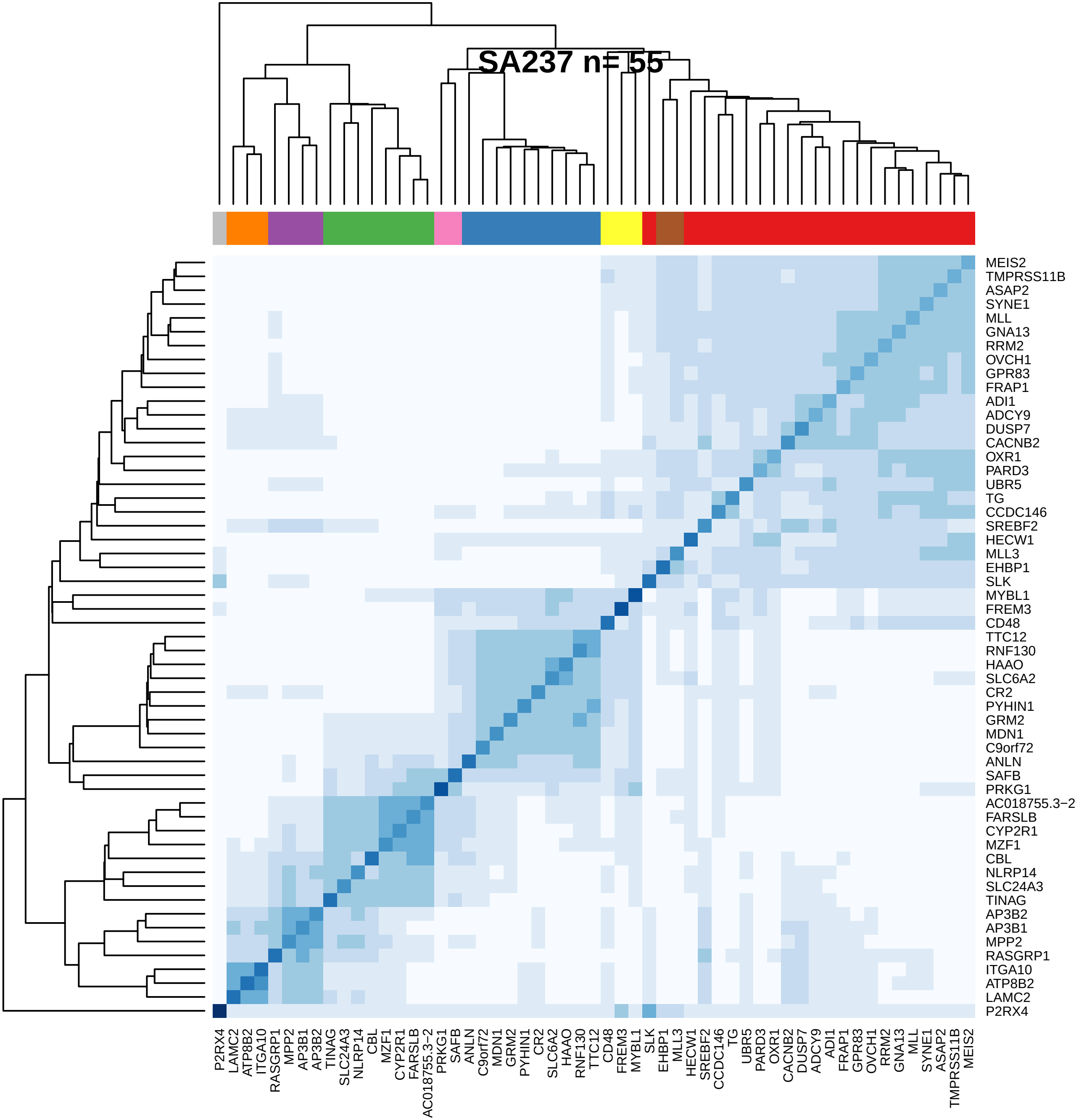


SA235 n= 17
OSBPL3
ADAMTS12
HELZ
DNAH11
C10orf18
BAP1
ZFAT
TXNRD2
FOXM1
HPS4
DDX42
ACP5
VAMP8
DOPEY1
RAG1
NKIRAS1
ZNF781
ZNF781
NKIRAS1
RAG1
DOPEY1
VAMP8
ACP5
DDX42
HPS4
FOXM1
TXNRD2
ZFAT
BAP1
C10orf18
DNAH11
HELZ
ADAMTS12
OSBPL3


SA235 n= 17
ZNF781
DOPEY1
VAMP8
HPS4
ACP5
NKIRAS1
RAG1
DDX42
TXNRD2
FOXM1
ZFAT
ADAMTS12
DNAH11
OSBPL3
C10orf18
BAP1
HELZ
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency


SA236 n= 45
YLPM1
ARID2
CATSPERB
MORF4L1
PTGFR
MAP4K2
TP53
KIF5A
HOXA2
FLT3
RBM17
SH3BP2
FLT3
SPATA16
ELSPBP1
DDX31
ZNF385D
DENND2C
WDFY3
TUT1
EXOSC2
PTPRN2
LAMA1
HYDIN
PCLO
MCM3AP
ACRBP
PRDM8
C6orf138
MUC16
FNBP1L
KIAA0368
XBP1
TTN
ALG10
TPSD1
IGF2BP2
LRRC17
BRD7
MFF
AC093283.3−1
IRF3
ICA1L
GPR116
RASA2
RASA2
GPR116
ICA1L
IRF3
AC093283.3−1
MFF
BRD7
LRRC17
IGF2BP2
TPSD1
ALG10
TTN
XBP1
KIAA0368
FNBP1L
MUC16
C6orf138
PRDM8
ACRBP
MCM3AP
PCLO
HYDIN
LAMA1
PTPRN2
EXOSC2
TUT1
WDFY3
DENND2C
ZNF385D
DDX31
ELSPBP1
SPATA16
FLT3
SH3BP2
RBM17
FLT3
HOXA2
KIF5A
TP53
MAP4K2
PTGFR
MORF4L1
CATSPERB
ARID2
YLPM1


SA236 n= 45
XBP1
MUC16
FNBP1L
KIAA0368
C6orf138
PRDM8
ACRBP
MCM3AP
DENND2C
DDX31
EXOSC2
ZNF385D
WDFY3
TUT1
LAMA1
PTPRN2
HYDIN
PCLO
RBM17
SH3BP2
FLT3
HOXA2
FLT3
TP53
KIF5A
MORF4L1
MAP4K2
YLPM1
CATSPERB
ARID2
PTGFR
SPATA16
ELSPBP1
IGF2BP2
TPSD1
TTN
BRD7
ALG10
LRRC17
IRF3
AC093283.3−1
MFF
ICA1L
GPR116
RASA2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency



SA237 n= 55

MEIS2
TMPRSS11B
ASAP2
SYNE1
MLL
GNA13
RRM2
OVCH1
GPR83
FRAP1
ADI1
ADCY9
DUSP7
CACNB2
OXR1
PARD3
UBR5
TG
CCDC146
SREBF2
HECW1
MLL3
EHBP1
SLK
MYBL1
FREM3
CD48
TTC12
RNF130
HAAO
SLC6A2
CR2
PYHIN1
GRM2
MDN1
C9orf72
ANLN
SAFB
PRKG1
AC018755.3−2
FARSLB
CYP2R1
MZF1
CBL
NLRP14
SLC24A3
TINAG
AP3B2
AP3B1
MPP2
RASGRP1
ITGA10
ATP8B2
LAMC2
P2RX4
P2RX4
LAMC2
ATP8B2
ITGA10
RASGRP1
MPP2
AP3B1
AP3B2
TINAG
SLC24A3
NLRP14
CBL
MZF1
CYP2R1
FARSLB
AC018755.3−2
PRKG1
SAFB
ANLN
C9orf72
MDN1
GRM2
PYHIN1
CR2
SLC6A2
HAAO
RNF130
TTC12
CD48
FREM3
MYBL1
SLK
EHBP1
MLL3
HECW1
SREBF2
CCDC146
TG
UBR5
PARD3
OXR1
CACNB2
DUSP7
ADCY9
ADI1
FRAP1
GPR83
OVCH1
RRM2
GNA13
MLL
SYNE1
ASAP2
TMPRSS11B
MEIS2

SA237 n= 55

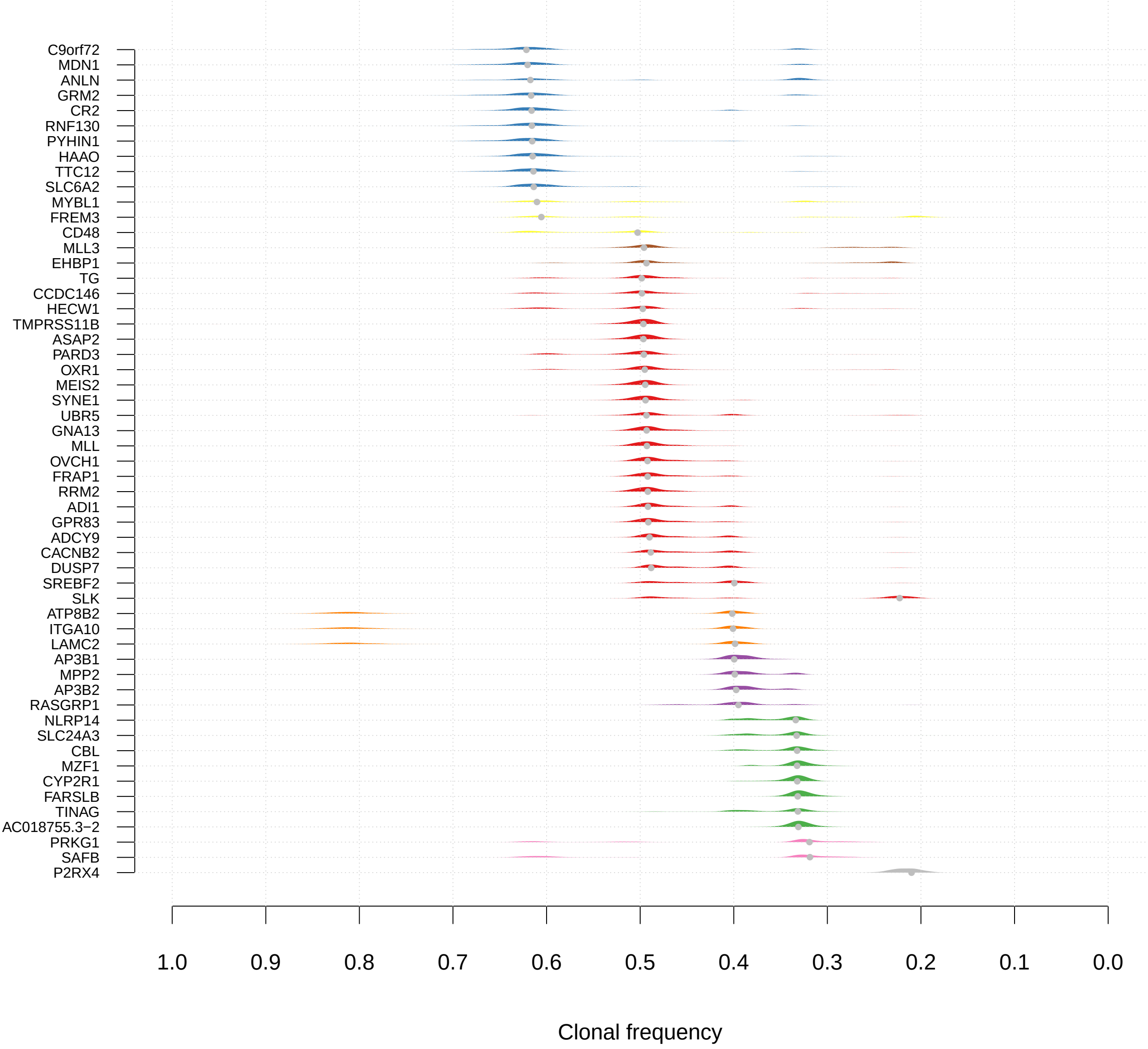
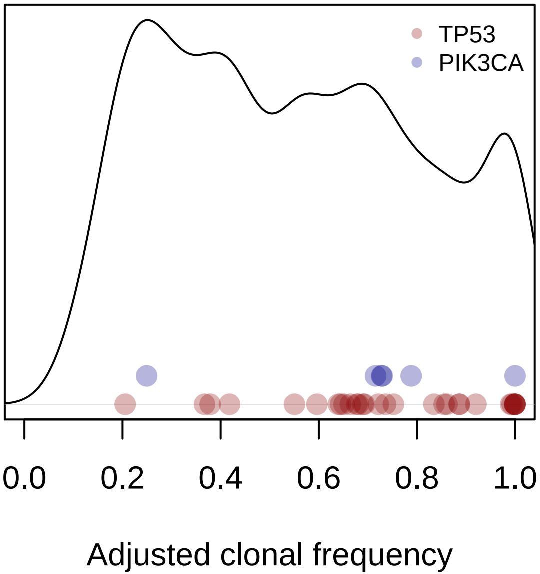

C9orf72
MDN1
ANLN
GRM2
CR2
RNF130
PYHIN1
HAAO
TTC12
SLC6A2
MYBL1
FREM3
CD48
MLL3
EHBP1
TG
CCDC146
HECW1
TMPRSS11B
ASAP2
PARD3
OXR1
MEIS2
SYNE1
UBR5
GNA13
MLL
OVCH1
FRAP1
RRM2
ADI1
GPR83
ADCY9
CACNB2
DUSP7
SREBF2
SLK
ATP8B2
ITGA10
LAMC2
AP3B1
MPP2
AP3B2
RASGRP1
NLRP14
SLC24A3
CBL
MZF1
CYP2R1
FARSLB
TINAG
AC018755.3−2
PRKG1
SAFB
P2RX4
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Clonal frequency

Figure S12



Supplemental Methods for Shah et al: The clonal and mutational evolution spectrum of primary triple negative breast cancers

1 Biospecimen collection and ethical consent

We assembled a collection of 149 clinically annotated primary fresh frozen triple negative breast cancer specimens (n = 149) each with a source of matched non malignant tissue for extraction of germline representative DNA. Tumour specimens were obtained from three tumour banks (BCCA Vancouver, breast tumour tissue repository; Alberta CBCF Breast Tumour Bank Edmonton, Cambridge UK, Addenbrooke's Hospital breast tumour bank) each with local REB/IRB approval for genomic studies of nucleic acids from breast cancer patients. This project was conducted under local BCCA REB/IRB projects H06-00289, H08-1230, H06-3199. The source of germline DNA was from peripheral blood lymphocytes in all but 4 cases. In these 4 cases histologically normal adjacent breast tissue was used. Initial case selection was based on clinical immunohistochemistry to define primary triple negative breast cancers obtained from surgical specimens, prior to the initiation of any chemotherapy or radiotherapy. Tumours typed as ER-, HER2- and PR- were initially selected for further review and re-validation of the IHC.

2 Histopathological review

Tissue sections were subject to expert histopathological review (GT) to assess the presence of invasive tumour, pre-malignant or benign changes, lymphocytic infiltration, necrosis and tumour cellularity. Tumour cellularity was scored visually in a semiquantitative fashion on sections taken from the cryosectioning runs used to isolate nucleic acids from each tumour. Cellularity values were binned such that 'low cellularity' corresponds to samples with $<$40% malignant cells, 'moderate cellularity' corresponds to 40% - 70% malignant cells, and samples with $>$70% malignant cells were considered to have 'high cellularity'. All but one sample classified as low cellularity were excluded from further analysis. The ER-, PR- HER2- immunophenotype was reconfirmed on sections or TMA cores from the cases included for analysis and additionally CK5/6 and EGFR were assessed by IHC. Subsequently SNP6.0 copy number analysis was also used to confirm the absence of HER2 amplification in each case.

3 Sequence data generation

Paired end RNASeq libraries were generated as described in Wiegand et al [1]. Exome capture was achieved through solution hybrid selection with the Human All Exon kit SureSelect Target Enrichment System (Agilent) version 1 for Illumina Genome Analyzer paired-end sequencing [2] and libraries were prepared as described in [3]. SOLiD whole genome shotgun libraries for 15 tumour/normal pairs were generated as previously described [4]. Details of per case coverage for each sequencing method are presented in Table S2. Targeted sequencing interrogating the exons and splice sites of 29 genes in 159 additional breast cancers (82 ER+ and 77 ER-ve, tumour and matched normal) was undertaken using exon-specific PCR and cDNA probe-based capture. The amplified/captured products were sequenced on the Illumina HiSeq 2000 platform.

4 Exome sequence data analysis and mutation validation

Sequence alignment and mutation calling Sequence reads were aligned to the human reference sequence (NCBI build 36, hg18) and a compacted reference composed of the targeted exons. Alignments were performed using Maq v0.7.1 [5] using default parameters. Uniquely aligned reads were kept for downstream analysis. Somatic SNVs were called using JointSNVMix (`http://compbio.bccrc.ca/software/jointsnvmix/`) [6] combined with mutationSeq [7], from the resulting tumour and normal pair of BAM files for each case. We used a marginal posterior probability threshold of 0.2 to nominate somatic SNVs, aiming to err on the side of sensitivity and minimize false negatives. All non-synonymous mutations as profiled by MutationAssessor [8] (`http://mutationassessor.org`) were carried forward for validation by deep amplicon sequencing. Indels were predicted using samtools v0.1.7 pileup command from BAM files. Indels present in at least 10% of reads covering the position of interest, predicted to affect a coding sequence, and for which there were 0 reads containing the same indel in the corresponding normal DNA library were carried forward for validation as candidate somatic indels.

4.1 Validation of mutations

{ sec:revalidation }

Revalidation sequencing Somatic mutation predictions were subjected to rigorous validation by targeted deep amplicon sequencing in the tumour and normal DNA templates. Genomic DNA was prepared as previously described [9] . Automated primer design was performed using Primer3 (Rozen and Skaletsky, 1998) and custom scripting. Primer pairs were designed to place the variant position within 75bps of either end of the amplicon and to be between 50-300bp in length. Primer pairs were independently validated by in silco PCR followed by BLAT against the human genome to ensure that the correct target was generated and that the resulting amplicon was unique within the genome. DNA primers were synthesized in 96-well plates at a 25nmol scale with standard desalting (IDT Coralville, IA USA). Polymerase cycling reactions were set up in 96-well plates and comprised of 0.5 μM forward primer, 0.5 μM reverse primer, 1-2 ng of gDNA template, 5X Phusion HF Buffer, 0.2 μM dNTPs, 3% DMSO, and 0.4 units of Phusion DNA polymerase (NEB, Ipswich, MA, USA). Reaction plates were cycled on a MJR Peltier Thermocycler (model PTC-225) with cycling conditions of a denaturation step at 98 °C for 30 sec, followed by 35 cycles of [98 °C for 10 sec, 69 °C for 15 sec, 72 °C for 15 sec] and a final extension step at 72 °C for 10 min. PCR reactions were visualized on 3% agarose (NuSieve) gels for 2hrs at 170V to assess PCR success. Successful reactions were manually pooled (4ul per well) by template and subjected to Illumina library construction using a modified paired-end protocol (Illumina, Hayward, USA). This involved A-tailing of the amplicons and ligation to Illumina PE adapters. Adapter-ligated products were purified on Qiaquick spin columns (Qiagen, Valencia, CA, USA) and PCR-amplified using Phusion DNA polymerase (NEB, Ipswich, MA, USA) in 10 cycles using PE primer 1.0 (Illumina) and a custom multiplexing PCR Primer [5-CAAGCAGAAGACGGCATACGAGAT NNNNNNCGGTCTCGGCATTCCTGC TGAACCGCTCTTCCGATCT-3] where *NNNNNN* was replaced with unique fault tolerant hexamer barcodes for each template. PCR products of the desired size range were purified away from adapter ligation artifacts using 8% PAGE gels, pooled and DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay (Agilent, Santa Clara CA, USA) and Nanodrop 7500 spectrophotometer (Nanodrop, Wilmington, DE, USA) and subsequently diluted to 10nM. The final concentration was confirmed using a Quant-iT dsDNA HS assay kit and Qubit fluorometer (Invitrogen, Carlsbad, CA, USA). For sequencing, clusters were generated on the Illumina cluster station using v4 cluster reagents and paired-end 75bp reads generated using v4 sequencing reagents on the Illumina GA$_{ii}$ platform following the manufacturer's instructions. Between the paired 75bp reads a third 7 base pair read was performed using the following custom sequencing primer [5- GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG] to sequence the hexamer barcode. Image analysis, base-calling and error calibration was performed using v1.60 of Illumina's Genome analysis pipeline.

Revalidation data analysis PCR primers were designed to revalidate candidate somatic mutations (SNVs and indels) on whole genome amplified material for both tumour and normal DNA. Resulting amplicons were indexed, pooled and sequenced deeply using 75bp paired end libraries as described above. Reads were separated out by 6bp index using a Hamming distance function with a maximum mismatch tolerance of 1, and aligned (Maq v0.7.1) to a reference database containing only the targeted loci for each case. (Note, we compared results to BWA [10] and there were no observable differences in the validated list of mutations). For each targeted sequence, we inferred the presence/absence of the targeted variants using a Binomial exact test as previously described [9]. In addition, we imposed the following criteria in order to classify a 'validated' somatic mutation:

- both tumour and normal data had a minimum of 50 reads covering the targeted position
- the Binomial exact test result (Benjamini Hochberg adjusted p-value) for the tumour < 0.01
- the Binomial exact test result (Benjamini Hochberg adjusted p-value) for the normal $>= 0.01$
- the proportion of reads indicating the variant in the tumour $\geq 5\%$

For indels, we required:

- a minimum of 10% of all reads aligned to the position containing the indel
- at least 3 reads containing the indel
- a minimum local realignment score of 300

One validation library was single end. For this library, we aligned the reads using GSNAP [11] - a gapped aligner capable of detecting indels in single end reads.

5 RNASeq data analysis and validation

Sequence alignment and SNV/indel calling for RNASeq SNVs for RNASeq were analysed as previously described [9] using SNVMix [12]. Paired end libraries were aligned using a modified version of BWA to a reference consisting of the human genome reference (NCBI build 36, hg18) and a database of known exon-exon junctions obtained from different annotation databases (Ensembl [13], RefSeq [14], AceView [15]). Sequences representing exon-exon junctions were designed to require at least a 4 base pair overlap for split-reads. Considering a read length of 50 base pairs, 46 base pairs on either side of the exon-exon junction were concatenated to represent each exon-exon junction. These alignments were used as the information source for differential expression analysis as well as measurement of exon abundance and intron usage. Additionally, de novo splice sites were predicted using HMMSplicer [16]; these predictions served to determine the effects of splice-site mutations in the transcriptome.

Alignment of RNA-Seq short reads was done using a modified version of BWA (base version 0.5.5 [17], patch unpublished). In order to apply the Burrows-Wheeler transform [18], upon which BWA is built, the reference sequence has to be concatenated into one string. When including exon-exon junction sequences to the alignment reference, these are appended to this string in order for the aligner to consider them as part of the target space. This procedure can cause mapping problems when working with paired end tags. In the last stages of alignment, BWA takes advantage of paired information to increase alignment specificity. When reads map to multiple places, whether with zero or more mismatches, pairing information can be used to determine the best alignment position for the pair. In order to optimize this process, BWA measures the distance between the reads of each pair at each mapping position using the concatenated reference as a base and comparing this distance to the estimated fragment size. Alignment positions in which the distance between the reads (too short or too long) or their alignment orientation do not match are discouraged. Given that the exon-exon junction sequences are appended to the genome reference, any pair where one or both reads map to a junction are broken due to the concatenated distance between them. Our modifications make BWA aware of the true position of each exon-exon junction in the reference sequence, thus eliminating some of the mapping issues described above.

Non synonymous SNVs and indels predicted from RNASeq screened against dbSNP and 1000genomes data that were recurrent by gene in at least 3 cases were submitted for validation using resequencing methods described as above for exome capture libraries.

Gene expression analysis Gene expression of Ensembl54 gene models was estimated by the number of fragments that covered them. A fragment was assigned to a gene when at least one read in the read-pair overlapped an exon in the gene, reads with BWA-assigned mapping quality lower than 10 were discarded. In order to make expression values comparable across libraries, we applied quantile normalization [19] of fragment counts in log2 space. Akin to microarray analysis, quantile normalization has been proposed as a viable method for RNA-Seq data that effectively protects against biased estimates due to over-representation of highly expressed genes [20].

PAM50 classification Normalized expression values (as described above) were used classify the triple negative breast cancer samples into two groups (*Basal* and *Other*) using the 50-gene classifier described by Parker et al. [21]. Of the 80 triple negative tumours 48 were classified as *basal*: and 32 as *other*; the latter group included any sample labeled by PAM50 as LuminalA, LuminalB, Her2 or Normal-like. Unsupervised hierarchical clustering (method: Ward, distance: spearman correlation) was performed to assess the reproducibility of PAM50 groupings. At $k = 3$ clusters there is a strong group of 16 samples labeled *Other*, a group formed by 6 libraries obtained from normal tissue and a group of 64 samples with mixed labels; at $k = 4$ this last group is split into a strong *Basal* cluster ($n = 28$) and a mixed one ($n = 36$). Unsupervised hierarchical clustering thus independently supported the primary classification by PAM50 into a relatively homogenous ”basal” expression group and a more heterogenous ”other” TNBC group.

Gene fusion analysis and PCR validation Gene fusions from paired end RNASeq data were predicted using *deFuse* v0.3.6 [22]. The deFuse predictions were filtered by thresholding on the deFuse probability at a value of 0.5, corresponding to a false positive rate of 0.93 and a false negative rate of 0.15. We identified, from the filtered deFuse predictions, those fusions that were corroborated by a genomic breakpoint in the Affymetrix SNP 6.0 copy number analysis. A fusion was considered corroborated by a genomic breakpoint if that breakpoint was predicted to occur within the boundaries of one of the genes involved in the gene fusion. From the list of genomic breakpoint corroborated fusions, we selected 64 events for validation.

Where available, we sought to identify evidence of genomic rearrangement to corroborate fusion predictions. We performed a targeted search for supporting Whole Genome Shotgun Sequencing (WGSS) read pairs in SOLiD data for 15 tumour genomes. For each fusion, a paired end alignment was considered supporting if it aligned less than 200,000 nucleotides downstream of the predicted fusion boundary in each gene. Supporting paired end alignments for each fusion were clustered, and the largest cluster for each was identified. Sequences for each breakpoint were assembled from paired end alignments as described previously [23]. In total, we identified 29 genomic breakpoints supporting gene fusions. In addition, 5 tumour genomes were sequenced using paired end whole genome shotgun sequencing at low coverage ($\sim$8x) using Illumina GA$_{ii}$. For these genomes, we used comrad v0.0.1 [23] (`http://compbio.cs.sfu.ca`) to identify fusions with supporting genomic breakpoints.

The expression of the fusion transcripts were validated by RT-PCR followed by Sanger sequencing. In brief, reverse transcription was performed using SuperScript III (Invitrogen Canada Inc., Burlington, ON) and 200 ng of RNA from the source tumour. PCR was carried out using Finnzymes PhusionTM High Fidelity DNA polymerase (New England BioLabs Inc., Pickering, Ontario) as well as forward T7 tagged and M13 reverse tagged PCR primers specifically designed to each predicted rearrangement. Products were visualized on 1.5% agarose gel prior to submission for Sanger sequencing at Macrogen (Rockville, MD, USA).

6 Network and pathway analysis using Reactome and EnrichmentMap

{ sec:cytoscape }

For network and pathway enrichment analysis for mutations, differential expression and copy number analysis, we used the Reactome FI [24] Cytoscape [25] plugin in Cytoscape v2.8.1. Significantly enriched pathway gene sets were exported and analysed using EnrichmentMap [26] to determine relationships between pathways. For this analysis, gene sets in GMT format were provided by the Reactome FI authors (Wu, personal communication). Patient-pathway clustering was performed by projecting a patient-gene matrix containing occurrences of mutations in a given gene in a given patient to a patient-pathway matrix containing occurrences of mutations in a given gene belonging to a pathway (or geneset) in a given patient. This patient-pathway matrix was then clustered using hierarchical agglomerative clustering with Euclidean distance and Ward linkage. We acknowledge the redundancy induced by the many-to-many relationship structure of genes in pathway gene sets, however to preserve biological interpretability and to maintain identifiable gene sets in Reactome, genesets were kept unaltered.

7 Affymetrix SNP6 analyses

7.1 Normalization of intensities

The Affymetrix SNP6.0 arrays for both tumour and normal samples were all independently normalized using the single array method CRMAv2 [27]. We applied the default settings using the following tags: ACC,ra,-XY,BPN,-XY,AVG,A+B,FLN,-XY. For each sample, allelic-crosstalk calibration, probe sequence effects normalization, probe-level summarization, and PCR fragment length normalization was performed. The following annotation files were used (Feb 14, 2008): Chip definition file (GenomeWideSNP_6.Full.cdf), Unit fragment-length (GenomeWideSNP_6,Full,na24,HB20080214.ufl), Unit genome position (GenomeWideSNP_6,Full,na24,HB

7.2 Normalization using a reference

Log ratios were computed for both tumours and normals by normalizing each array independently against a reference. Due to the incomplete pairings of tumours and matched normals, we needed to generate a pooled reference from all the available normals.. However, because we are also attempting to identify CNVs in addition to CNAs in the tumours, using a simple pooled reference (median intensity across all normals for each probe) can “subtract” inherently frequent ($> 50\%$) copy number events in the normals from the tumours. For example, using a reference which contains a region that is deleted in more than half the normal samples will result in an artificially amplified signal when computing the ratio for a true neutral signal in the tumour sample. Therefore, we generated a “masked” reference from the normal dataset that was free of inherent copy number polymorphisms, while accounting for systematic biases.

We started with CRMAv2 normalized intensities of the matched normal samples and computed log ratios using a randomly permuted reference. Next, probe-level discrete copy number calls were made using a 6-state hidden Markov model (described below). For each sample, we identify probes that are predicted as gains or losses. Next, we retrieved the original normalized intensities that we started with and masked out the probes we identified in the previous step. Finally, we computed the median (pooling) intensity value for each probe across all samples using only the remaining values after masking. The result is a single vector of median values, which makes up the masked reference. In this approach, the masked reference improves analysis of germline CNVs in the tumours by serving as a true base-line vector of intensities for log ratio calculations and allowing all true alteration signals in the data to be detected.

7.3 Segmentation and copy number analysis of 28 Affymetrix SNP6 normal samples using a 6-state HMM

We applied the 6-state HMM to 28 normal samples to detect germline CNVs by returning segment and probe-level copy number. The 6-state HMM is a modified version of CNA-HMMer [28] that that has been adapted to analyzing high-density genotyping arrays such as the Affymetrix SNP6.0 platform and whole genome shotgun data. Another significant extension is the inclusion of additional copy number states, which offers a more intuitive interpretation of DNA dosage in cancer. The new discrete copy number state space included neutral (NEUT), homozygous (HOMD) and hemizygous (HETD) deletions, gain (GAIN), amplification, (AMP) and high-level amplification (HLAMP).

$$K_{CNA} = \{HOMD, HETD, NEUT, GAIN, AMP, HLAMP\} \tag{1}$$

{eq:Kcna}

7.4 Segmentation and copy number analysis of 104 Affymetrix SNP6 tumour samples using HMM-Dosage

We performed copy number analysis on the 104 tumour samples using HMM-Dosage (Ha et al, manuscript in preparation). The alrogithm is designed to detect and distinguish the complete set of somatic and germline copy number events in cancer genomes interrogated by SNP array data. This model extends the 6-state HMM (above) by using 5 additional CNV states to represent the analogous copy number status

$$K_{CNV} = \{CNVHOMD, CNVHETD, CNVGAIN, CNVAMP, CNVHLAMP\}. \tag{2}$$

{eq:Kcnv}

The model performs segmentation on log ratios of intensity data and assigns a discrete copy number status $k_t \in \{K_{CNA}, K_{CNV}\}$ to the each latent variable, Z_t, for all probes $t \in \{1 \ldots T\}$. Log ratios, $Y = \{y_1, \ldots, y_T\}$, are observed values modeled using Student's-t densities in the emission component of the HMM, conditional on $Z_t = k_t$. Spatial information is captured using a non-stationary transition matrix, A_t, that encodes position-specific probabilities of CNVs and CNAs at each probe t. In order to distinguish between somatic copy number alterations (CNAs) and germline copy number variants (CNVs), HMM-Dosage probabilistically incorporates CNV information as a prior to the transition matrix.

The CNV prior was computed as probe-level CNV frequencies by combining (weighted averaging) 2 datasets: 482 normal samples from METABRIC [29] and an external dataset of 450 HapMap normal samples whose CNVs were predicted by [30]. Because the HapMap dataset contains males, we excluded chrX frequencies and simply used the chrX frequencies from the 482 normals.

The parameters of the Student's-t distributions are unobserved and estimated using the expectation maximization (EM) algorithm for each sample, independently. Initial parameters for the K_{CNA} states of the Student's-t mixture were empirically determined using 45 SNP6 breast cancer cell line samples in the COSMIC [31] dataset. First, independently for each cell-line sample, we fit the log ratios to a 6-state Gaussian mixture model using EM. Subsequently, the converged Gaussian mixture parameters are used as initial parameters for fitting a 6-state Student's-t mixtures in the 6-state HMM (again, using EM). Finally, we averaged the converged Student's-t parameters for each state across the 45 cell lines. These averaged parameters became the initial K_{CNA} parameters to Student's-t emission of HMM-Dosage. Initial parameters for K_{CNV} were set by hand to theoretical values $log2(\,[0.5, 1, 3, 5, 7]\,/2)$.

The optimal state sequence, $Z_{1:T}$, which represents the copy number prediction for each probe, is computed using the Viterbi algorithm.

In general, HMM-Dosage was tuned to be conservative in calling CNVs as to avoid misclassifying the more important CNA calls. Therefore, the results were post-processed by comparing against the CNV map (derived from 482 normals and HapMap270) such that CNA segments that had at least 25% reciprocal genomic overlap with a segment of the same copy number state in the CNV map were converted to germline state.

7.5 Genomic instability

For each of the 80 genomes analyzed by Affymetrix SNP6 arrays that have RNA-seq data, genomic instability (GI) was computed as the proportion of genome altered by basepair. For this analysis, we calculated the total length of all CNA events (HOMD, HETD, GAIN, AMP, HLAMP) by using the segmentation boundaries predicted by HMM-Dosage. This value is then divided by the total length of all predicted segments, including NEUT. This calculation is also done for each CNA type to reveal the relative contribution of each alteration event type to GI.

We also compared GI between basal and non-basal PAM50 subtypes. The basal subtype shows higher GI for all CNA types.

7.6 Gene alterations

In order to identify the genes that are altered by copy number changes, we searched for overlap of segments with gene regions. The gene annotations and coordinates are given by Ensembl 54 (hg18). The genes selected for analysis are those annotated as protein-coding, totaling to 20878 genes.

We generate two types of patient-by-gene copy number matrices to capture two perspectives of gene alterations. The first type is the 'call' matrix, $C \in \mathbb{Z}^{P \times G}$, which is populated with values representing discrete copy number state calls. The second is the log ratio matrix, which contains segment median log ratios. For this latter matrix, genes overlapping multiple segments are summarized using two different approaches to produce $L_a \in \mathbb{R}^{P \times G}$ and $L_s \in \mathbb{R}^{P \times G}$. For each patient $p \in P$ and each gene $g \in G$, we identify segment s that overlaps g and assign $C(p, g)$ with the copy number state of s and $L_a(p, g)$ and $L_s(p, g)$ with the median log ratio of the probes in s.

If gene g overlaps or is broken by a set of segments, $S = s_1, \ldots, s_k$, where $k \geq 2$, we assign $C(p, g)$ with the copy number state of the segment having max $severity$ based on the binary relation (Equation 5 and 6). For $L_a(p, g)$, the weighted sum of the median logRs (Equation 3) is used while for $L_s(p, g)$, the logR corresponding to the most severe segment is used (Equation 4).

$$\begin{aligned}
L_a(p,g) &= \Big(\sum_{s \in S} length(s) * medianLogR(s)\Big) / \sum_{s \in S} length(s) && (3)\\
L_s(p,g) &= logRvalueOf\left(\underset{s \in S}{\operatorname{argmax}}\{severity(CNstateOf(s))\}\right) && (4)\\
C(p,g) &= CNstateOf\left(\underset{s \in S}{\operatorname{argmax}}\{severity(CNstateOf(s))\}\right) && (5)\\
severity &= \{(\text{‘}NEUT\text{’}, 0), \\
&\quad (\text{‘}HOMD\text{’}, 8), (\text{‘}HETD\text{’}, 6), \\
&\quad (\text{‘}GAIN\text{’}, 5), (\text{‘}AMP\text{’}, 6), (\text{‘}HLAMP\text{’}, 8), \\
&\quad (\text{‘}CNVHOMD\text{’}, 4), (\text{‘}CNVHETD\text{’}, 3), \\
&\quad (CNVGAIN\text{’}, 2), (\text{‘}CNVAMP\text{’}, 3), (\text{‘}CNVHLAMP\text{’}, 4)\} && (6)
\end{aligned}$$

7.7 Subtype-specific analysis of copy number

A $\mathcal{X}^2$ (Chi-Square) test of independence was used to determine subtype association with copy number. For each gene g, copy number of the patients were divided into 3 levels: deletion (HOMD & HETD), gain (GAIN, AMP, HLAMP) and neutral. For the omnibus test, patients were divided into 2 levels (subtypes), Basal and non-Basal. No significant genes were found after using Bonferroni adjusted p-value cutoff of 0.1.

7.8 Outlier analysis

To identify genes whose expression was driven in cis by extreme copy number events, we fit Gaussian distributions to the 80 expression values of a gene across the patients in the cohort (for which RNASeq data was available), using the maximum likelihood estimates of the mean and variance. For genes with multiple probes, we summarized the expression level using the median. This resulted in 20878 Gaussian distributions, each corresponding to the expression profile of a gene.

We then sought to capture gene-patient (g, p) events where large amplitude copy number events (such as putative homozygous deletions and high level amplifications) drives expression into the extreme tails of the expression profile. We selected events in 5% left-tail or 5% right-tail of the expression distributions. Events in the right tail whose copy number state was either amplification or high-level amplification (HLAMP) by HMM-Dosage and those events in the left tail whose copy number state was homozygous deletion (HOMD) were reported.

8 WGSS analysis for copy number

The following analysis is implemented into an R package called HMMcopy available from `http://compbio.bccrc.ca/software/hmmcopy/`.

8.1 Segmentation and copy number analysis of 15 WGSS tumour samples using 6-state HMM

The WGSS-derived copy number results used as input in APOLLOH (see below) were generated using a modified version of a paired tumour-normal strategy described previously in [32] and [9]. The genome was divided into fixed windows of 1kb, and read depth is extract for each window in the tumour and normal. In this study, we applied two additional preprocessing steps prior to segmentation in order to achieve more accurate copy number estimates. First, we applied a filter to remove repetitive regions that are highly mappable. Second, we corrected GC content bias to remove wave-like patterns in the tumour and normal, separately, using a loess curve fit between GC content and read depth. A log ratio is computed for each window by computing proportion between the GC corrected tumour value and GC corrected normal value.

Divide the genome into fixed genomic windows

First, the genome was divided into fixed genomic windows of 1kb, reducing the analysis to a set of approximately 2.5 to 3 million loci, $\mathbf{R}$.

Extract read depth for normal and tumour

Next, separately for the normal and tumour genomes of each patient, we extracted the total read depth for each window in $\mathbf{R}$ using BAMtools [33], resulting in a vector of read count data from the normal, $\mathbf{N_R} = (n_1, \ldots, n_{|R|})$ and tumour, $\mathbf{T_R} = (t_1, \ldots, t_{|R|})$, respectively.

Removing windows that are highly mappable

Using ”ENCODE Duke Uniqueness of 35bp sequences” track from UCSC (`http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg18&g=wgEncodeMapability`), we filtered windows that were within repetitive regions, resulting in being highly mappable by aligners. Windows that had mappability score of ≥ 0.9 were excluded. This removed extreme amplified positions which would have otherwise posed as confounding outliers in downstream segmentation analysis.

GC content correction of normal and tumour read counts

We performed GC content bias correction to the tumour and normal of each patient separately. We applied a global loess fit between GC content and read depth for windows in $\mathbf{R}$. Due to computational restrictions of fitting 3 million data points, we further excluded outlier windows based on read depths being in the upper and lower 1% quantile, and randomly sampled 20,000 of the remaining windows for generating the loess curve. Finally, the read depth of all windows, $\mathbf{N_R}$ and $\mathbf{T_R}$, are corrected by scaling the observed value by the loess fitted value (Equation 7).

$$corrected\ read\ depth = \frac{observed\ read\ depth}{loess\ fitted\ value} \tag{7}$$

{ eq:loess˙correct }

Normalizing copy number in tumours

The GC corrected normal $\mathbf{N_R}$ and tumour $\mathbf{T_R}$ counts are normalized independently to generate $\mathbf{\bar{N}_R} = (\bar{n}_1, \ldots, \bar{n}_{|R|})$ and $\mathbf{\bar{T}_R} = (\bar{t}_1, \ldots, \bar{t}_{|R|})$, respectively, where $\bar{n}_i = \frac{n_i}{\sum_j n_j}$ and $\bar{t}_i = \frac{t_i}{\sum_j t_j}$, $i \in \{1, \ldots, |R|\}$. To obtain the final tumour copy number observed at each loci $r \in R$, we applied another normalization step by taking the log2 ratio between tumour and normal copy number, $\mathbf{T_N(i)} = log_2\left(\frac{\bar{t}_i}{\bar{n}_i}\right)$, $i \in \{1, \ldots, |R|\}$.

Segmentation and copy number prediction via HMM

The 6-state version of HMM-Dosage was used to segment the input data $\mathbf{T_N}$. Initialization and hyperparameters for the prior means of the Student's-t distribution used for this analysis were $log2(([1, 1.4, 2, 2.7, 3, 4.5])/2)$. Because the normalized the tumour counts with the match normal read count naturally excludes germline CNVs, using the 11-state HMM-Dosage to distinguish CNA from CNV was not required.

8.2 Segment-based correlation of WGSS and SNP6 predicted copy number

Copy number results from WGSS data was used to help validate HOMD and HLAMP predictions in the Affymetrix SNP6 data. In the 15 cases, for each HOMD and HLAMP predicted segment, s, boundaries s_{start} and s_{end} and segment median a_s are extracted from the SNP6 results. We compute the median log ratio copy number w_s across the set of regions $\boldsymbol{i} \in R$ that overlap these boundaries s_{start} and s_{end}.

$$w_s = median(T_N(r)), \ \{r | r \in R, r_{start} \geq s_{start}, r_{end} \leq s_{end}\}$$

Spearman correlation is computed on $w_{1:S}$ and $a_{1:S}$ where S is the total number HOMD/HLAMP segments in the SNP6 results.

9 Analysis of LOH

We determined loss of heterozygosity (LOH) in the 15 WGSS and 50 exome samples using a probabilistic model packaged in the software called APOLLOH (Ha, et al manuscript in preparation). Preprocessing of the data is performed in a paired tumour-normal approach by interrogating informative heterozygous loci identified in the normal genomes using UnifiedGenotyper from GATK [34]. Next, the tumour genome allelic counts in the same loci are input into APOLLOH which performs segmentation of the allelic ratios into regions of LOH, allele-specific copy number amplification (ASCNA) and heterozygosity (HET). Copy number was also input into APOLLOH to enhance predictions within regions that deviate from copy neutral. For the WGSS samples, we used copy number results predicted by the previously described depth-based WGSS approach. For the exome samples, we used copy number results from the SNP6 genotyping arrays. Software for APOLLOH is available from the authors' website at : http://compbio.bccrc.ca/software/apolloh/ .

10 Human regulatory region analysis

Analysis of 323 high-confidence somatic mutations in non-coding space showed that 72 were found to fall within experimentally determined human regulatory regions as described in [35]. For each variation contained within a ChIP-seq-defined region, the matrix for the corresponding TF was used to predict putative TFBSs overlapping the variation (reviewed in [36]). Two TFBS scores were generated, one corresponding to the reference genome sequence and the other corresponding to the observed cancer variation. A relative matrix score threshold of 80% was applied to define putative TFBSs. As the variations were selected to overlap regions of known transcription factor binding, we generally expect the reference allele to score higher for functionally significant changes (i.e. to go from a functional TFBS to a non-functional one). We called regulatory variations exceeding the 80% threshold as putatively impactful to transcription factor binding. In the TN breast cancer set, 2 such somatic mutations were predicted to affect MITF binding and 6 somatic mutations were predicted to affect RB binding.

11 DriverNet: Integrated analysis of copy number, mutations and gene expression data

To nominate mutations predicted to have significant impact on expression networks, we used the 681 Reactome FI gene sets containing 7521 genes to construct an 'influence graph' where genesets were used to form 'cliques' in the graph. We then constructed a bipartite graph such that each gene that was in the mutation list from sequencing or HOMD, HLAMP list from CNA analysis was placed in one partition of the graph and the outlying expression genes (as described above) was placed in the other partition. Edges between nodes in the two partitions were drawn according to the following criteria:

- a mutation and an outlying expression event co-occurred in the same tumour
- an edge between the two genes was present in the influence graph

With the influence graph constructed, we then sought to rank the genes in the 'mutation' partition such that nodes explaining the highest number of outlying expression expression events (by greedy algorithm) were nominated as putative driver mutations. The greedy algorithm is an efficient approximation to the combinatorial optimization problem induced by the bipartite graph construction that we wish to solve in order to predict the fewest number of mutations that disrupt the maximum number of expression signatures.

The statistical significance of the candidate genes are assessed using a randomization framework. Given the original dataset S, the algorithm is run on randomly generated datasets N times $(S_1, \ldots, S_N)$ and then the results on real data are assessed to see if they are significantly different from the randomized datasets. This in an indirect way of perturbing the bipartite graph corresponding to the original problem. To generate the random datasets, we keep the contents of patient-mutation, M, and patient-outlier, G', matrices the same but replace the gene symbols with a randomly selected set of genes from the Ensmbl 54 protein-coding gene list. Using the same influence graph, the algorithm is run on the new patient-mutation, $M_1 ... M_M$, and patient-outlier, $G'_1 ... G'_M$, matrices.

Suppose D is the result of driver mutation discovery algorithm on the input U. D contains a ranked list of driver genes with their corresponding node coverage in the bipartite graph, $\mathcal{B}$. The statistical significance of a gene $g \in D$ with a corresponding node coverage, COV_g, is the fraction of times that we observe driver genes in our random data runs D_i, with node coverage more than COV_g:

$$\text{pvalue}(g) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \delta[COV_{gj} > COV_g]}{N * \sum_{i=1}^{N} M_i} \tag{8}$$

where M_i is the number of drivers identified in run n of the algorithm, N is the total number of times that the algorithm on random datasets were run.

Software for this analysis called *DriverNet* (Haffari et al., manuscript in preparation) is available from the authors' website: `http://compbio.bccrc.ca/software/drivernet` . Pathway analysis of the top 50 genes output by the greedy algorithm was performed using Reactome as described above.

12 Clonal frequency analysis

{ sec:clonal }

We used a hierachical Bayesian model which accounts for uncertainty about genotype to estimate the frequency of cells with a given mutation. We imagine each mutation divides the population of cells into two classes; those without the mutation we call the reference population and those with we refer to as the variant population. We treat the reference population as though it has fixed genotype of AA. We model the genotype of the variant population at a given locus using one of 27 states allowing for copy numbers from 1-6 and differing zygosity status. We use information about copy number and LOH status derived from APOLLOH exome analysis to inform or prior beliefs on the genotype of the variant population. To allow multiple mutations to have the same frequency, we use a Dirichlet process prior on the class frequencies. Sites which are clustered by this model have cellular frequencies which are sufficiently similar that they are better explained as originating from a single class of cells. Inference for this model is performed using Markov chain Monte Carlo sampler. A full description of the hierarchical Bayesian model shown as a probabilistic graphical model is shown in Figure S10. Input to the model is a set of mutations and their allele abundance (reference, variant). In addition, the copy number and LOH status at the position of each mutation is given. The output of the model is a mutation co-occurrence matrix containing the proportion of MCMC samples in which each pair of mutations were grouped together, and the clonal frequency estimation for each of the MCMC samples, thus allowing inference of a posterior distribution of clonal frequency estimates.

Software implementing this approach called *PyClone* (Roth et al., manuscript in preparation) is available upon request from the authors. Cases with 10 more more mutations were included in the clonal frequency analysis. This led to estimates for 2109 mutations distributed across 54 cases. We clustered the mutation co-occurrence matrix for each case using hierarchical clustering and deterministically assessed the groupings of mutations into 'clonal clusters' using the R library dynamicTreeCut [37]. Clonal frequency distributions for each mutation were then plotted as in Figure 4 according to these groupings.

Each pathway in Reactome was also assessed for skewing of clonal frequency by Wilcoxon test. Clonal frequency estimates for each case were first adjusted by normalizing by the maximum clonal frequency value to remove the effect of normal cell contamination. The adjusted clonal frequencies were then used to assess whether the distributions of the mutations falling into a given pathway were different than the background distribution of all clonal frequency estimates using the R implementation of the Wilcoxon test `wilcox.test`. Resultant p-values were then adjusted using `p.adjust` with `method="BH"`.

References

[1] Wiegand, K. C. *et al.* Arid1a mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* **363**, 1532–1543 (2010).

[2] Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182–189 (2009).

[3] Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-hodgkin lymphoma. *Nature* **476**, 298–303 (2011).

[4] McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research* **19**, 1527–1541 (2009).

[5] Li, H., Ruan, J. & Durbin, R. Mapping short dna sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851–1858 (2008).

[6] Roth, A. *et al.* Jointsnvmix : A probabilistic model for accurate detection of somatic mutations in normal/tumour paired next generation sequencing data. *Bioinformatics* (2012).

[7] Ding, J. *et al.* Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* **28**, 167–175 (2012).

[8] Reva, B., Antipin, Y. & Sander, C. Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics. *Nucleic Acids Research* (2011, in press).

[9] Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–13 (2009).

[10] Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

[11] Wu, T. D. & Nacu, S. Fast and snp-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).

[12] Goya, R. *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* **26**, 730–736 (2010).

[13] Flicek, P. *et al.* Ensembl's 10th year. *Nucl. Acids Res.* **38**, D557–562 (2010).

[14] Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucl. Acids Res.* **35**, D61–65 (2007).

[15] Thierry-Mieg, D. & Thierry-Mieg, J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biology* **7**, S12 (2006).

[16] Dimon, M. T., Sorber, K. & DeRisi, J. L. HMMSplicer: a tool for efficient and sensitive discovery of known and novel splice junctions in RNA-Seq data. *PLoS ONE* **5**, e13875 (2010).

[17] Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

[18] Burrows, M. & Wheeler, D. J. A block-sorting lossless data compression algorithm (1994).

[19] Bolstad, B., Irizarry, R., Astrand, M. & Speed, T. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185 –193 (2003).

[20] Bullard, J., Purdom, E., Hansen, K. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).

[21] Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* **27**, 1160 –1167 (2009).

[22] McPherson, A. *et al.* deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput Biol* **7** (2011).

[23] McPherson, A. *et al.* Comrad: detection of expressed rearrangements by integrated analysis of rna-seq and low coverage genome sequence data. *Bioinformatics* **27**, 1481–1488 (2011).

[24] Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology* **11**, R53 (2010).

[25] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).

[26] Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).

[27] Bengtsson, H., Wirapati, P. & Speed, T. P. A single-array preprocessing method for estimating full-resolution raw copy numbers from all affymetrix genotyping arrays including genomewidesnp 5 & 6. *Bioinformatics* **25**, 2149–56 (2009).

[28] Shah, S. P. *et al.* Integrating copy number polymorphisms into array cgh analysis using a robust hmm. *Bioinformatics* **22**, e431–9 (2006).

[29] Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* (2012, accepted for publication).

[30] Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–12 (2010).

[31] Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).

[32] Chiang, D. Y. *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* **6**, 99–103 (2009).

[33] Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P. & Marth, G. T. Bamtools: a c++ api and toolkit for analyzing and managing bam files. *Bioinformatics* **27**, 1691–1692 (2011).

[34] McKenna, A. *et al.* The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res* **20**, 1297–1303 (2010).

[35] Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).

[36] Worsley-Hunt, R., Bernard, V. & Wasserman, W. W. Identification of cis-regulatory sequence variations in individual genome sequences. *Genome Med* **3**, 65–65 (2011).

[37] Langfelder, P., Zhang, B. & Horvath, S. ”defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics* **24**, 719–720 (2007).

Supplemental Tables for Shah et al: The clonal and mutational evolution spectrum of primary triple negative breast cancers

This file contains the legends describing Tables S1-17. All tables are available as an archived zip file in the Supplemental Data from the Nature website.

Table 1: Summary of data generation for 54 tumour and normal DNA exome libraries and 80 RNASeq transcriptome libraries. RNAseq and exome libraries were 50bp paired end Illumina sequence data. For RNASeq data, the number of aligned reads in exons, introns and intergenic regions of the genome are listed with three mapping quality categories. Mean, median and standard deviation of targeted exon coverage is listed for each exome library. Total acquired data for 15 SOLiD whole genome shotgun tumour/normal pairs are also listed.

{stab:data˙summary˙all}

Table 2: Summary of clinical characteristics of the patients. A full data dictionary describing all clinical variables is presented (Sheet 2)

{stab:clin˙data˙table}

Table 3: Omnibus table of validated somatic SNVs. All mutations were revalidated with deep amplicon resequencing. There were 2414 somatic point mutations that revalidated. Mutations were annotated with MutationAssessor (http://mutationassessor.org/, [1]). Column descriptions are as follows: **id**: unique patient/genome coordinate (NCBI build36, hg18); **Patient.ID**: case id; **Chr: chromosome**; **Coord**: chromosomal position; **Mutation**: nucleotide substitution; **RG.var.type**: missense, nonsense or silent mutation; **Func..Impact**: mutationassessor Functional impact class; **FI.score**: mutationassessor Functional impact score; **AA.variant**: amino acid substitution; **Gene**: gene name (HGNC); **Location**: chromosomal band; **Uniprot**: uniprot id; **Refseq**: refseq id; **mutations.in.COSMIC.position**: previously reported mutations at the same genomic coordinate in the COSMIC database; **types.of.cancer.in.COSMIC.position**: description of previously reported COSMIC mutations; **gene.s.known.role.in.cancer**: cancer gene census; **regions.position**: affected protein/domain localisation; **Ref.count.Tumour.**: resquencing reads from the tumour DNA matching the reference; **Alt.count.Tumour.**: resquencing reads from the tumour DNA matching the variant; **Total.count..Tumour.**: total resequencing reads from the tumour DNA; **Alt.frequency..Tumour.**: proportion of resequencing reads that are variant; **Binomial.test.p.value..Tumour.**: Binomial exact test p-value; **Binomial.test.corrected.p.value..Tumour.**: Binomial exact test corrected p-value; **Binomial.test.status..Tumour.**: Binomial exact test corrected p-value < 0.01?; **Ref.count..Normal.**: resquencing reads from the matched normal DNA matching the reference; **Alt.count..Normal.Total.count..Normal.**: resquencing reads from the normal DNA matching the variant; **Alt.frequency..Normal.**: proportion of resequencing reads matching the variant; **Binomial.test.p.value..Normal.**: Binomial exact test p-value; **Binomial.test.corrected.p.value.Normal.**: Binomial exact test corrected p-value; **Binomial.test.status.Normal.**: Binomial exact test corrected p-value < 0.01?; **SVNMix2.Call.RNASeq**: most probable genotype from SNVMix2 [2] (0 - not expressed, 1 - wildtype expressed, 2 - heterozygously expressed, 3 - homozygously expressed) ; **Ref.RNASeq.**: number of reference reads at position from the matching RNASeq data ; **Nonref.RNASeq.**: number of variant reads at position from the matching RNASeq data; **Ens54Gene:** Ensembl v54 gene name ; **Strand**: gene strand ; **CNCall**: copy number state ; **CNLogR**: SNP6.0 median log ratio for the overlapping segment at this position ; **PAM50**: PAM50 class (Basal/Other) for this case ; **Splice.site**: position is a dinucleotide donor/acceptor splice site (TRUE/FALSE); **HMMDosage_Call**: copy number state; **HMMDosage_LogR**: raw continuous copy number value; **OncoSNP_Call**: predicted OncoSNP [3] state ; **OncoSNP_CN**: predicted OncoSNP copy number; **apolloh.loh.ratio**: raw LOH value from APOLLOH analysis (see supplemental methods); **apolloh.call**: LOH call from APOLLOH from exomes; **APOLLOH_WGSS_Call**: LOH call from APOLLOH from genomes;**APOLLOH_WGSS_CN**: copy number used for APOLLOH analysis; **HMMcopy_WGSS_Call**: copy number inferred from genome data; **HMMcopy_WGSS_CN**: normalized logR input for genome-based copy number analysis; **mut_freq_GBM**: frequency of mutation in gene in TCGA GBM data; **mut_freq_HGS**: frequency of mutation in gene in TCGA HGS data; **clonality**: estimation of clonal frequency.

{stab:validated˙somatic˙SN

Table 4: Validated SNVs at dinucleotide donor/accptor splice sites (based on Ensembl v54 annotations) inducing alternative splicing or intron retention in the RNASeq data: legends as per Table S3.

{stab:validated˙somatic˙spli

Table 5: Validated somatic insertions and deletions. Scores are as predicted by `samtools pileup` [4].

{stab:validated˙somatic˙ind

Table 6: Table of CNA segments for 104 Affymetrix SNP6.0 arrays, predicted as described in Supplemental methods. Columns are CaseID; Chromosome; segment start coordinate (NCBI build 36); segment end coordinate; number of probes in segment; median logratio of segment; HMMDosage state (somatic states are HOMD: homozygous deletion, HETD: hemizygous deletion: GAIN: single copy gain, AMP: amplification, HLAMP: high-level amplification (see Supplemental methods); HGNC gene names of genes contained in the segment. The first six columns can be extracted and viewed in the Integrated Genomics Viewer Software (http://www.broadinstitute.org/igv/).

{stab:CNA˙frequency˙table

Table 7: Summary of 517 CNA breakpoint-corroborated gene fusions predicted by deFuse [5]. In addition to the columns described at http://compbio.bccrc.ca, 2 columns (gene_1_broken, gene_2_broken) identify which genes are interrupted by a CNV breakpoint, and 2 columns (transition_type_1, transition_type_2) identify the type of the transition.

{stab:defuse˙predictions}

Table 8: Summary of genomic breakpoints from a targeted search for supporting paired end reads in 15 SOLiD paired end tumour genomes. Shown are the case identifier, genes involved in the fusion, breakpoint sequence, and number of supporting paired end reads. {stab:solid˙genome˙breaks

Table 9: Integrated genome/transcriptome analysis for expressed gene fusions with corroborated genome rearrangements from 5 tumours (SA029, SA030, SA031, SA051, SA054) for which both RNASeq and low-pass whole genome shotgun data generated with Illumina paired end sequencing was obtained. Analysis was performed using Comrad (http://fusioncomrad.sf.net). A description of the table format can be found at http://fusioncomrad.sf.net {stab:comrad˙results}

Table 10: A subset of 35/63 selected gene fusion candidates from Table S7 that were validated with PCR. {stab:validated˙fusions}

Table 11: 72 non-coding mutations falling in regulatory regions showing 23 mutations in Rb transcription factor binding sites and relative scores of predicted Rb binding in wildtype vs mutant tfbs sequences {stab:regregions}

Table 12: Mutation frequency by gene for somatic SNV, indel and splice site mutations {stab:snv˙indel˙gene˙freque

Table 13: Mutation significance analysis showing genes predicted to be under selection through analysis of background synonymous mutations and non-synonymous mutations. All genes were analysed using methodology outlined in Youn and Simon [6]. {stab:mutsig}

Table 14: Enriched pathways from the network computed from recurrent somatically mutated genes in the Reactome functional interaction map. Enriched pathways with FDR <= 0.001 are included.

{stab:mutations˙pathways}

Table 15: Analysis of the top somatically aberrated genes (by node degree) connected (by Reactome gene sets) to genes that exhibited outlying expression from their population level distributions as computed by driverNet. Columns: Rank: by driverNet algorithm (Supplemental methods); Gene: somatically aberrated gene; gband: chromosomal band containing gene; SNV/Indel: number of cases harbouring an SNV or indel in the gene; HLAMP: number of cases harbouring a predicted high level amplification; HOMD: number of cases harbouring a predicted homozygous deletion; events: number of gene expression outliers (see Supplemental methods) coincident with a genomic aberration and where the outlying gene is connected to the aberrated gene; p-value: statistical significance based on a randomly generated background distribution (Supplemental methods).

{stab:drivernetall}

Table 16: Comparison of significantly enriched pathways (FDR < 0.005) by Reactome analysis from the top 50 genes predicted by driverNet (Supplemental methods) for TNBC and ovarian high grade serous (HGS) data from the TCGA [7]. The pathway, the initial p-value, the adjusted p-value the driverNet genes and the overlap status are shown, where status=1 indicates TNBC specific pathway, status=2 indicates HGS and TNBC enriched pathways and status=3 indicates HGS specific pathways.

{stab:tnbc˙ohgs}

Table 17: Analysis of the distributions of clonal frequency estimates of genes in Reactome pathways. For each set of genes belonging to a pathway, a Wilcoxon test was performed to assess whether the distribution of the clonal frequencies of genes in a pathway was different than the overall background distribution. For each pathway, the name of the Reactome pathway, the geneset of the pathway, the mutated genes that fall into the pathway, the number of such mutation events, the median clonal frequency estimate, the Wilcoxon p-value, and adjusted p-value (Benjamini-Hochberg) are shown.

{stab:pathway˙clonal˙freque

References

[1] Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39** (2011).

[2] Goya, R. *et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* **26**, 730–736 (2010).

[3] Yau, C. *et al.* A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biol* **11** (2010).

[4] Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

[5] McPherson, A. *et al.* deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. *PLoS Comput Biol* **7** (2011).

[6] Youn, A. & Simon, R. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics* **27**, 175–181 (2011).

[7] Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).