Discovering Structural Variant Signatures in Cancer

Greg Raskind

1 Background and Motivation

DDR Pathways are Critical For Maintaining Genomic Integrity

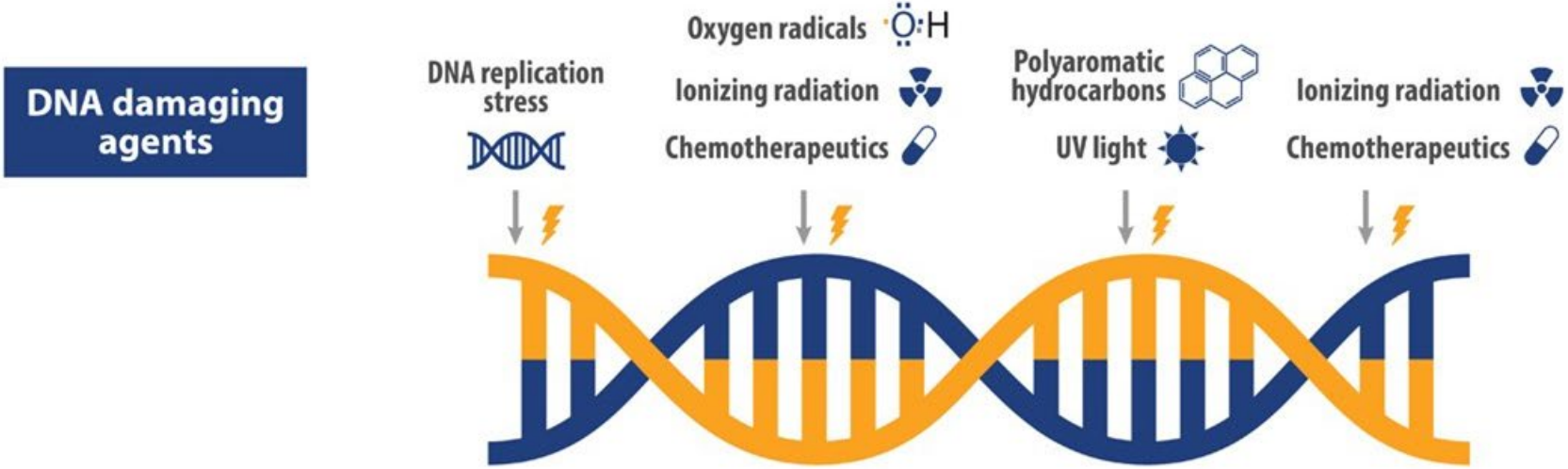


Image Credit: https://blog.crownbio.com/dna-damage-response

DDR Pathways are Critical For Maintaining Genomic Integrity

DNA damage types

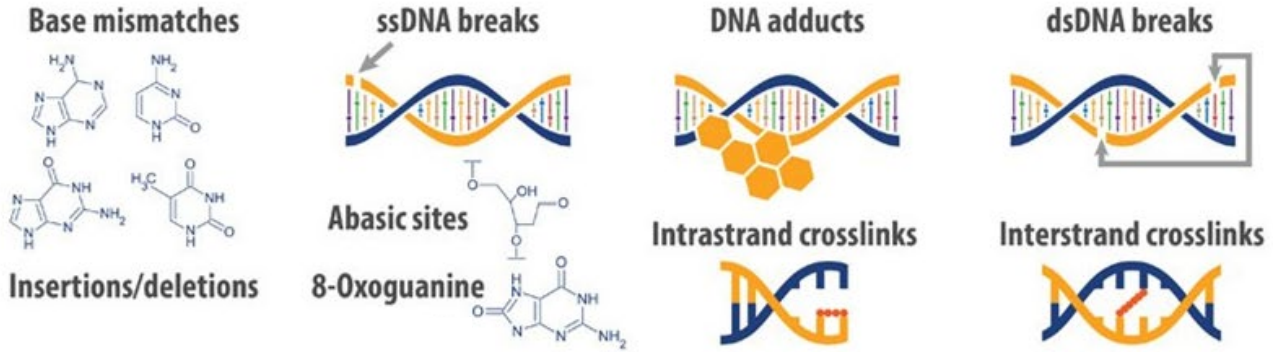


Image Credit: https://blog.crownbio.com/dna-damage-response

DDR Pathways are Critical For Maintaining Genomic Integrity

Image Credit: https://blog.crownbio.com/dna-damage-response

HR deficiency is a biomarker for PARP inhibitor sensitivity

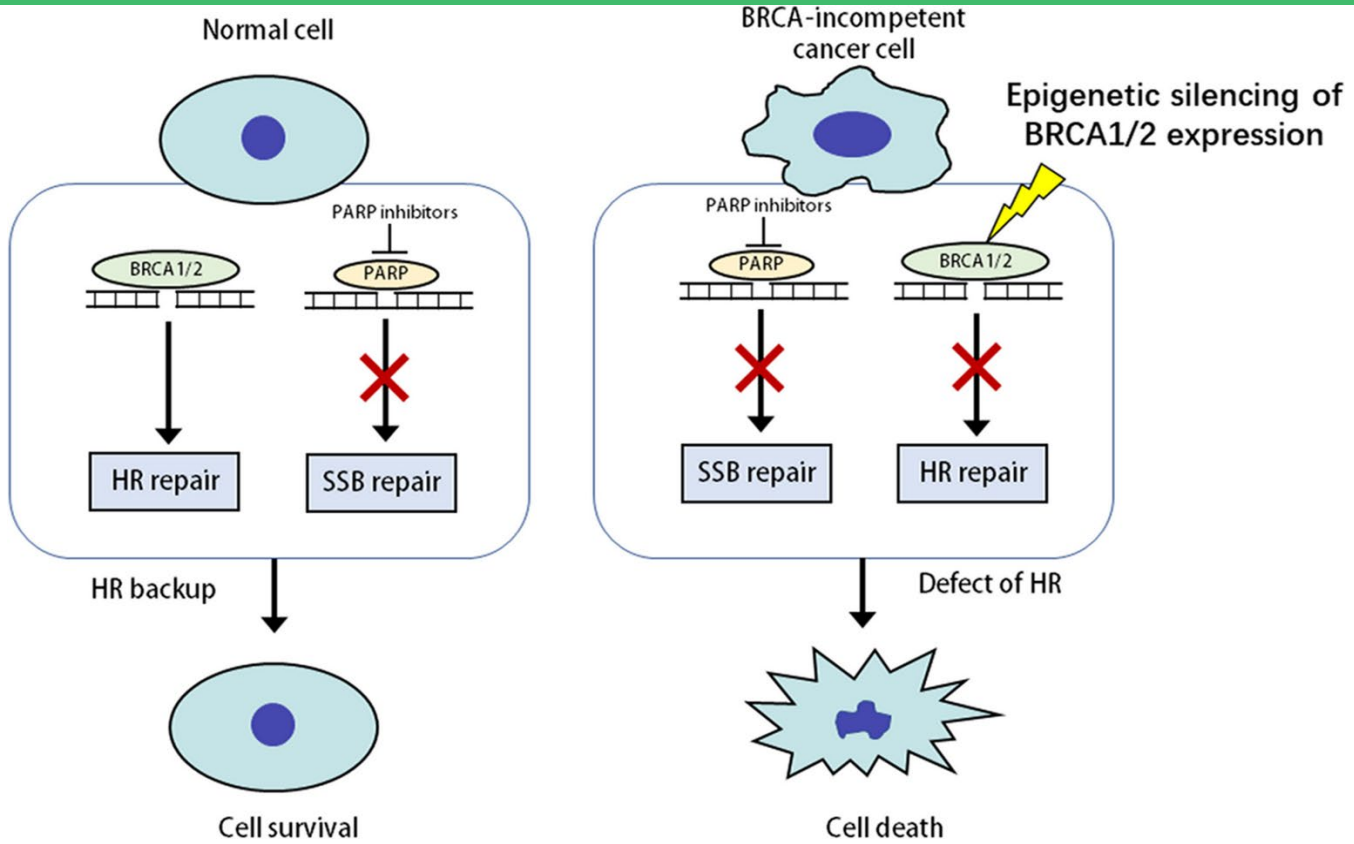


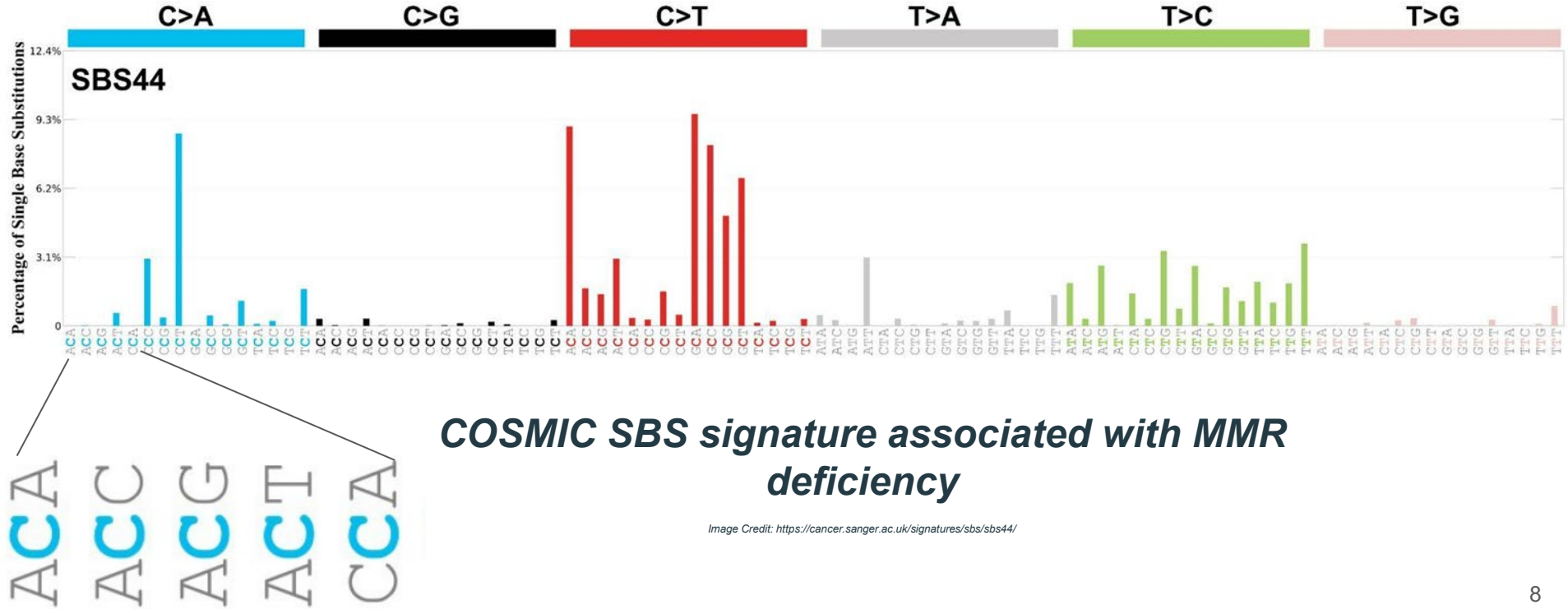
Image credit: Hu, Y. & Guo, M. *Cancer Sci.* (2020)

DDR defects in cancer can be exploited therapeutically

- MGMT methylation predicts improved response to radiation/chemotherapy in glioblastoma patients
- MMR deficiency is predictive of immunotherapy response in some cancers (e.g. colorectal)

Mutational signatures can serve as biomarkers of DDR state

Mutational signatures are patterns of mutations left behind by DNA damage/repair processes



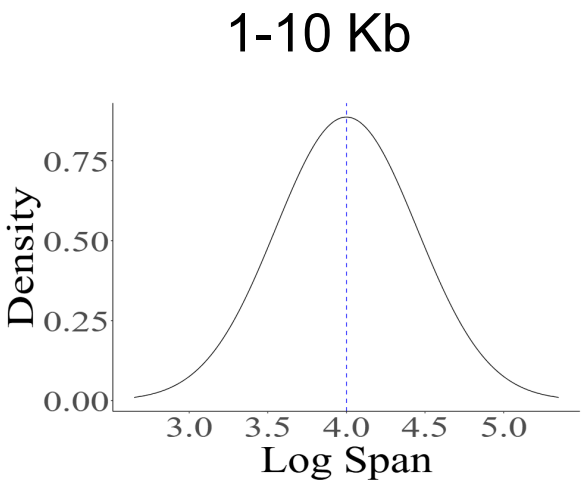
COSMIC SBS signature associated with MMR deficiency

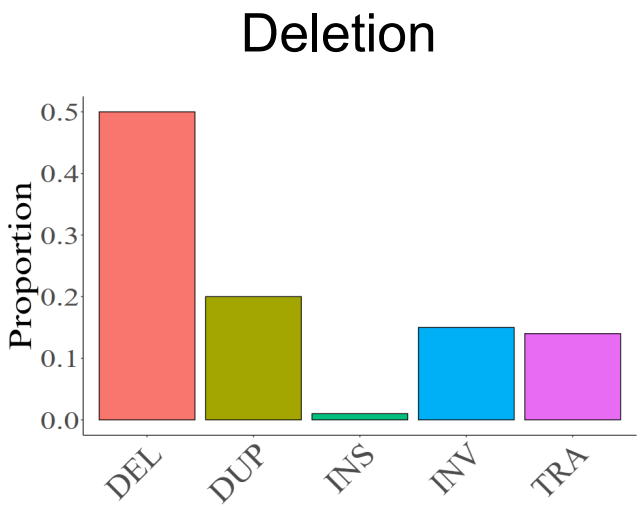
Image Credit: https://cancer.sanger.ac.uk/signatures/sbs/sbs44/

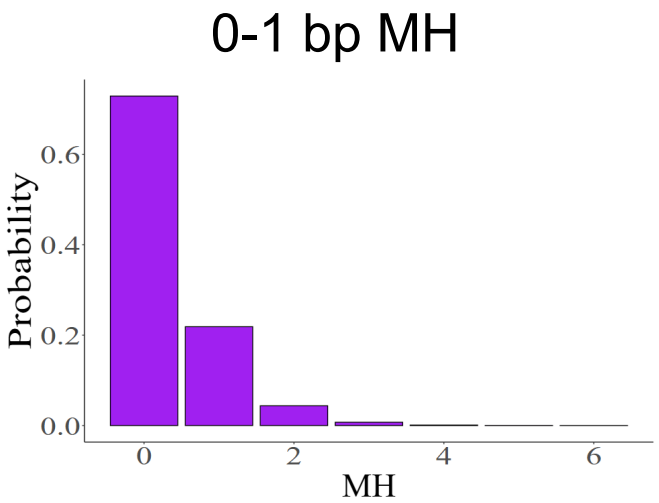
Structural variant signatures are understudied

- Current signature detection approaches largely focus on point mutations
- However, many DDR alterations are more likely to be reflected in structural variant (SV) signatures
  - Particularly those involved in double strand break repair

Our Approach: Represent features in a way that reflects underlying mechar



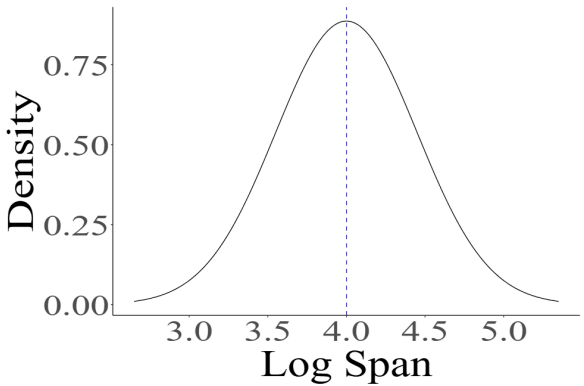




- Chosen distributions reflect our knowledge of the underlying biology
- Natural representation of quantitative/continuous variables
  - E.g. 1 Kb deletion is more similar to 10 Kb deletion than 10 Mb deletion

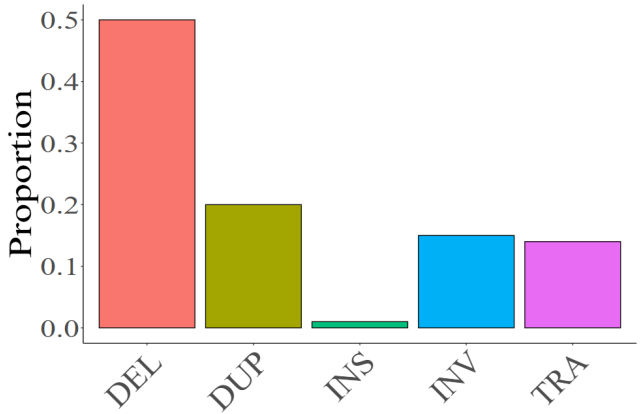
Our Approach: Fit model parameters that best explain observed data

1-10 Kb



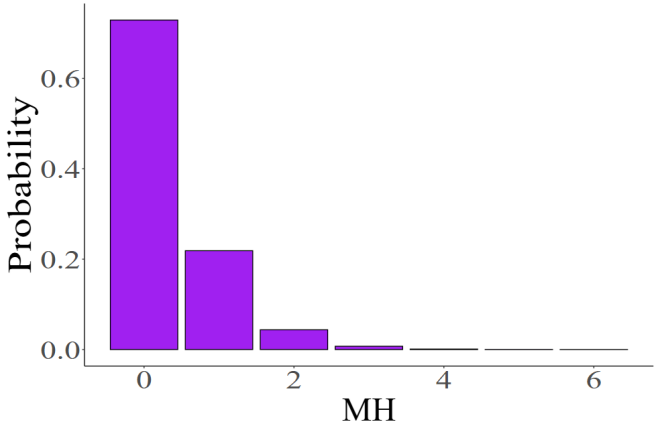
μ, σ

Deletion



c1, c2, c3, c4, c5

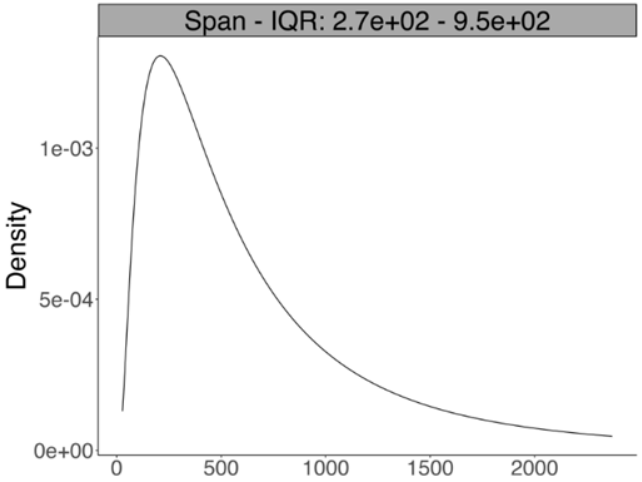
0-1 bp MH

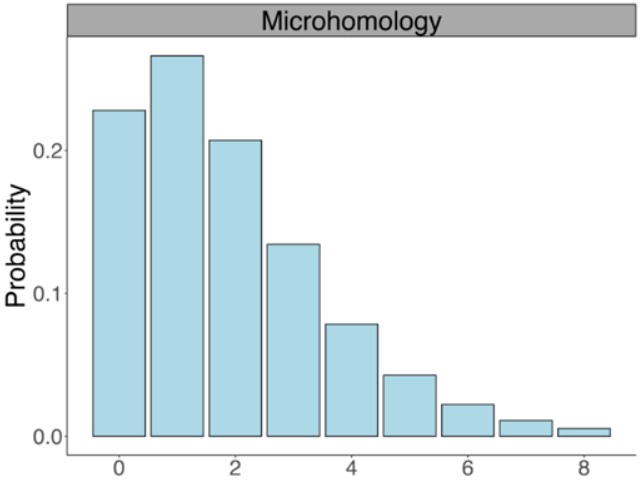


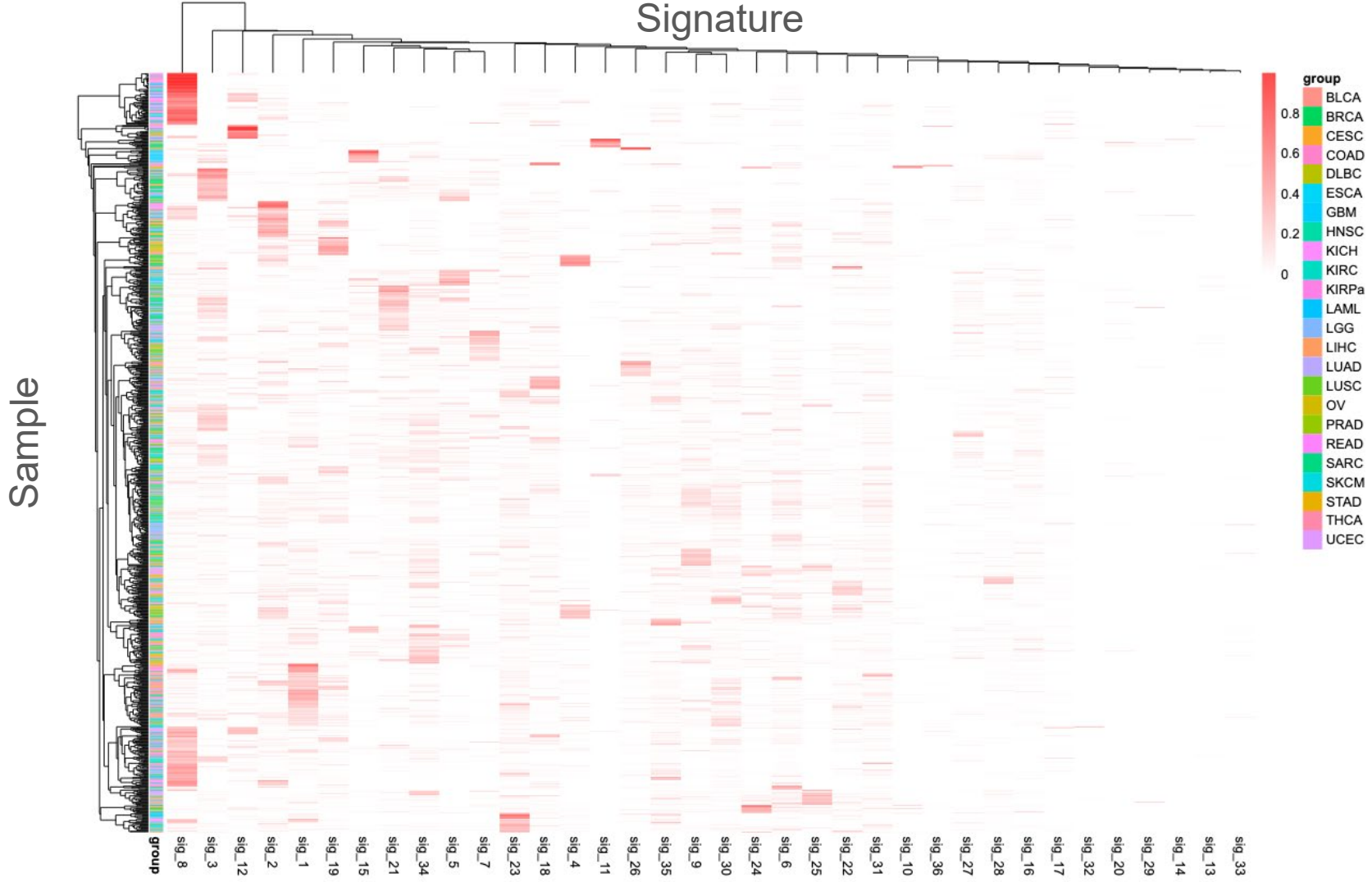
r, p

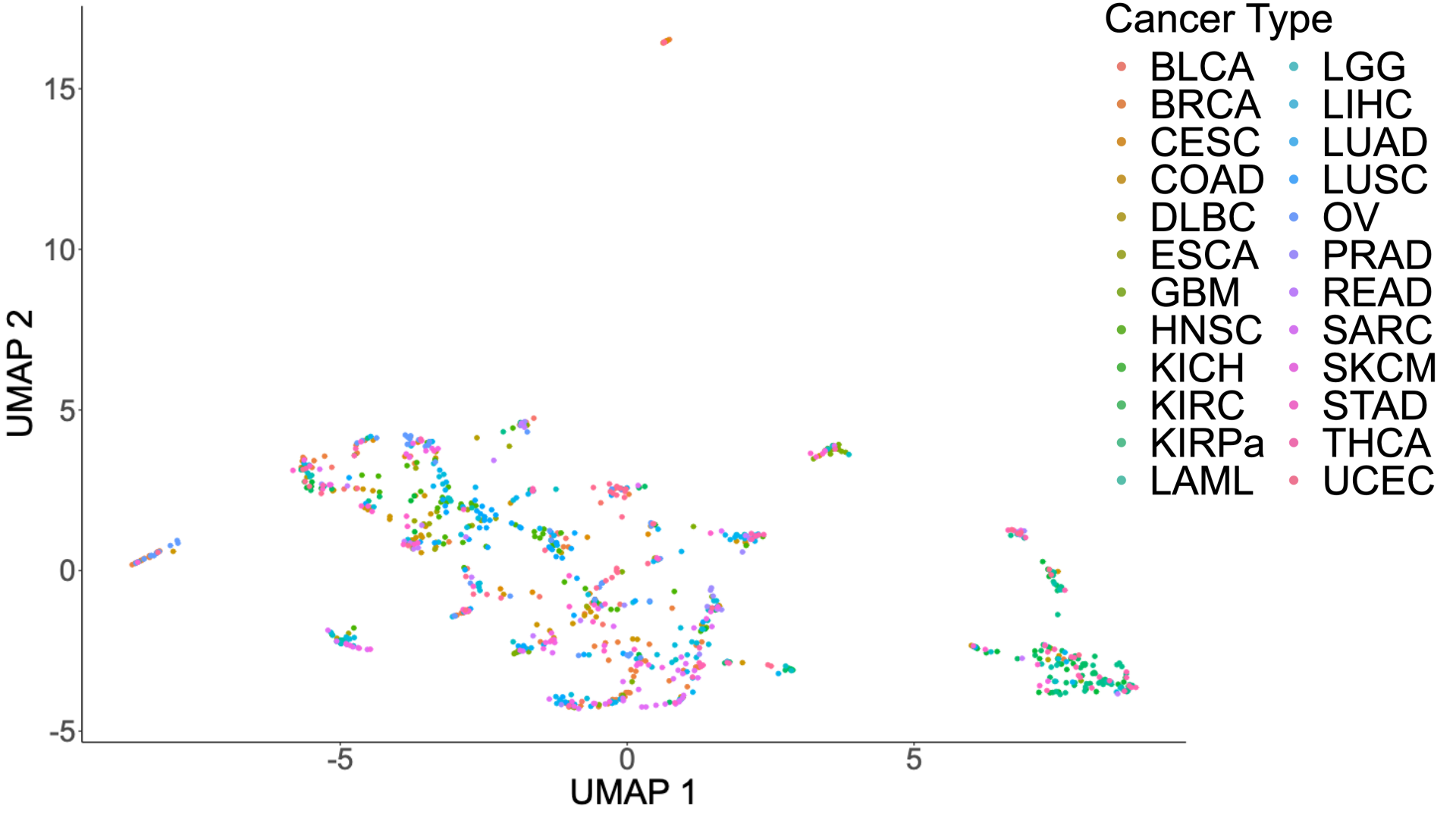
- Each feature only requires learning a few more parameters
- New features can be added without requiring a large increase in sample size to learn signatures

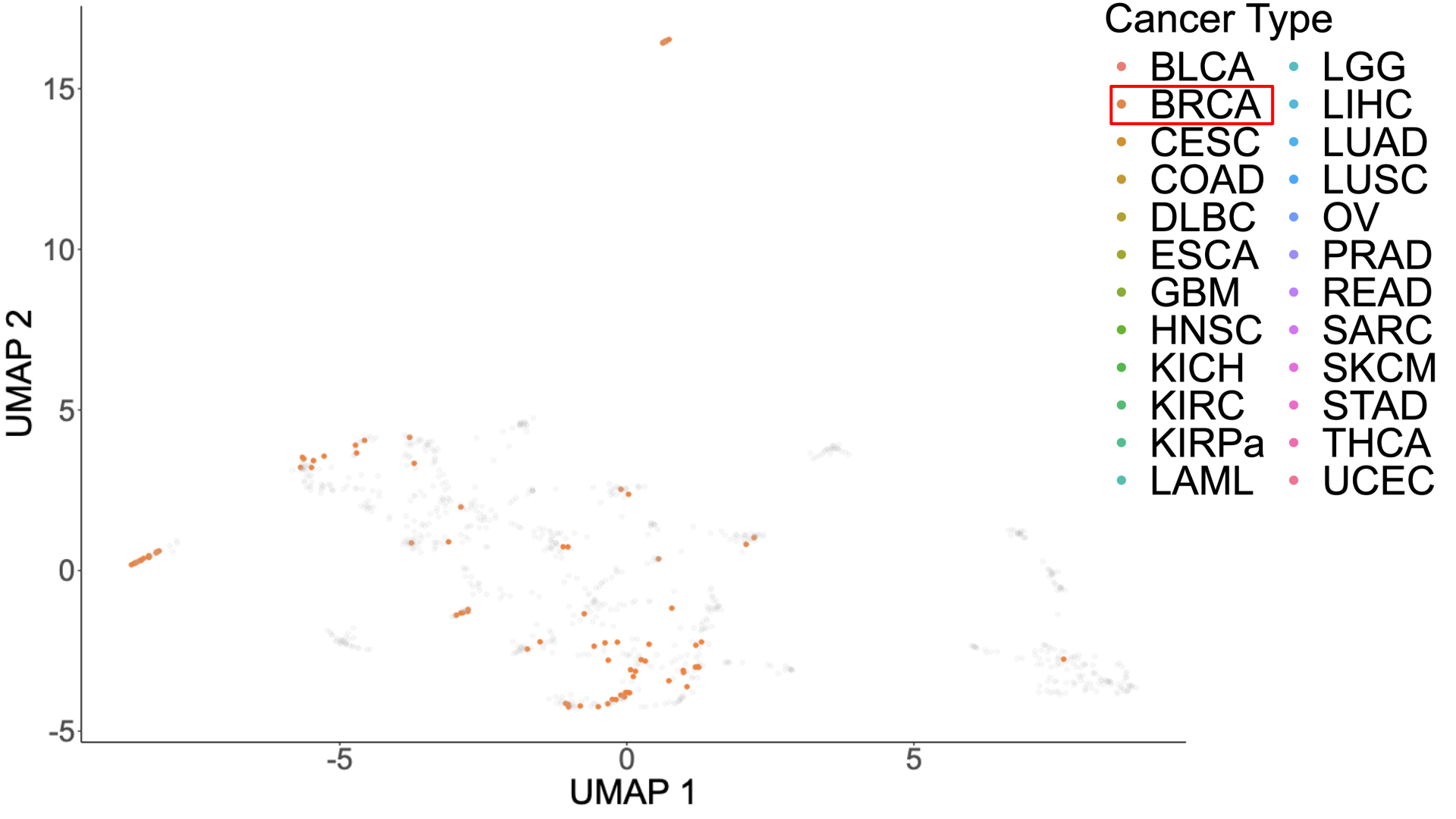
Analyzed 115,161 SV calls from 946 samples spanning 24 cancer types

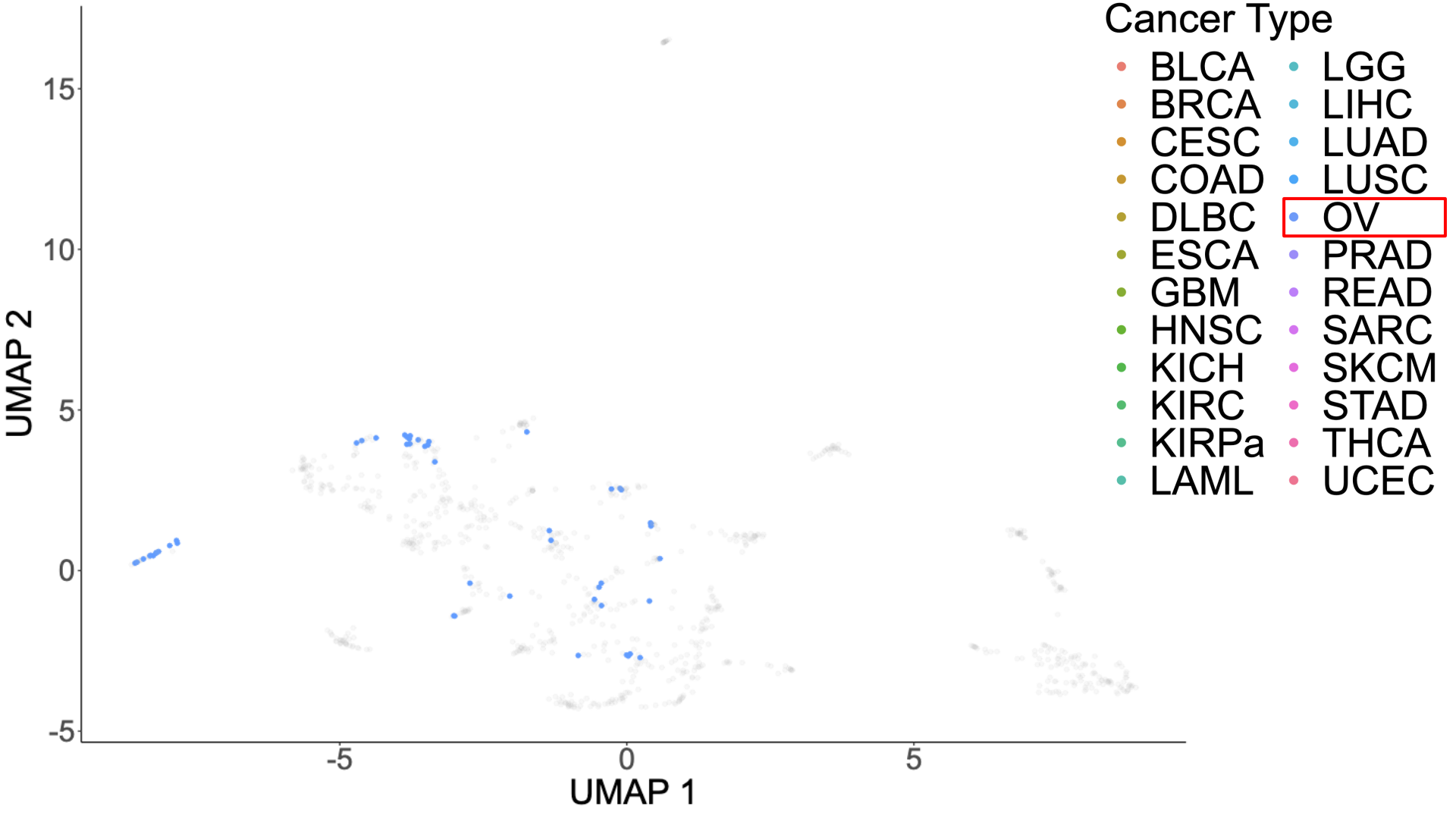


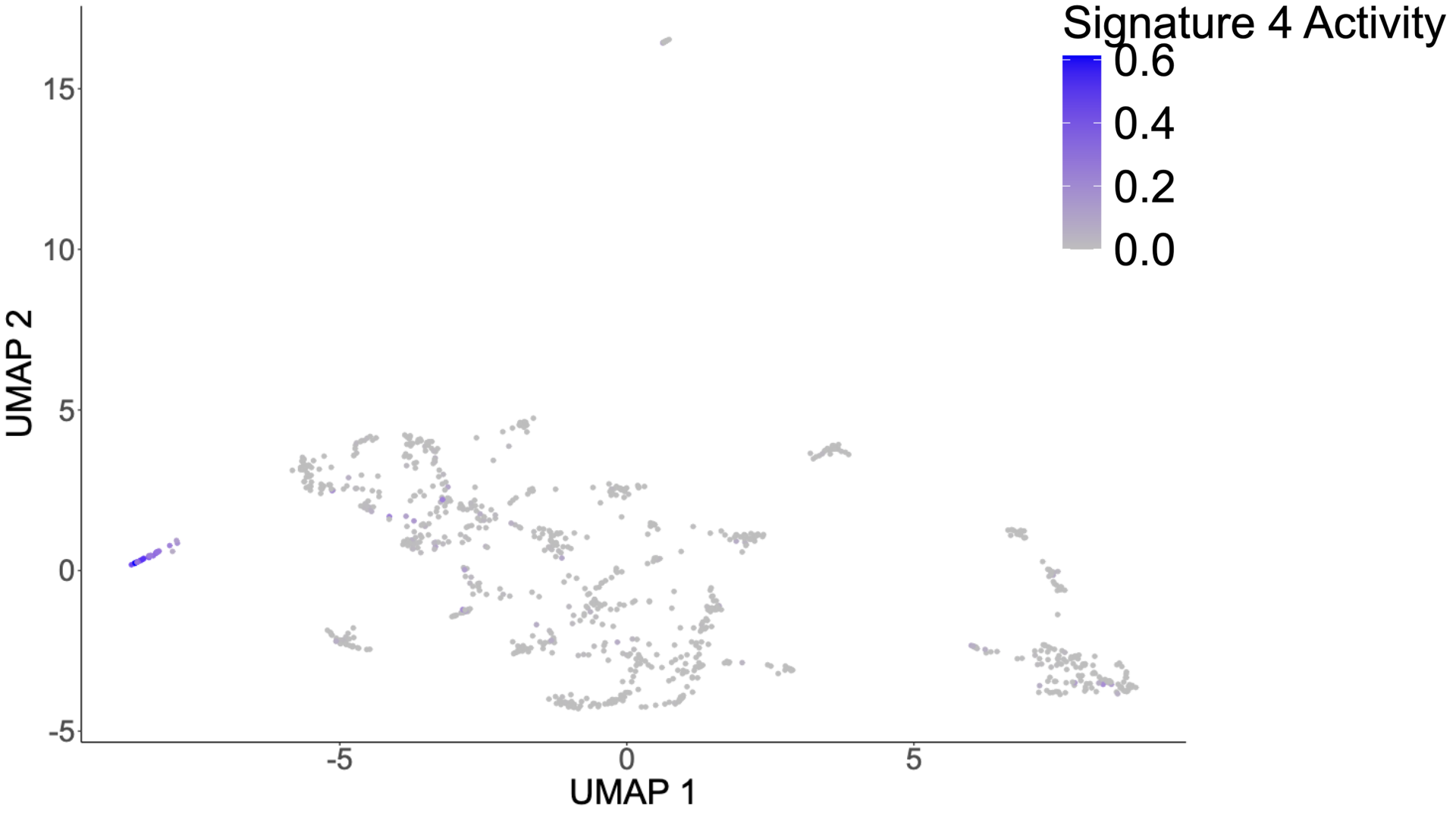


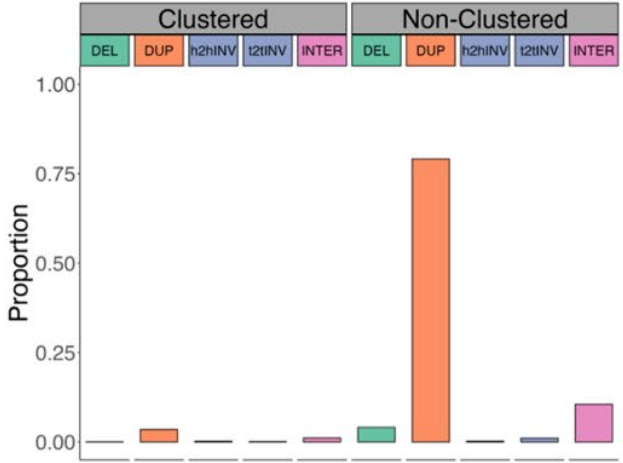

Signature
Sample
group
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
0.8
0.6
0.4
0.2
0
group
sig_8
sig_3
sig_12
sig_2
sig_1
sig_19
sig_15
sig_21
sig_34
sig_5
sig_7
sig_23
sig_18
sig_4
sig_11
sig_26
sig_35
sig_9
sig_30
sig_24
sig_6
sig_25
sig_22
sig_31
sig_10
sig_36
sig_27
sig_28
sig_16
sig_17
sig_32
sig_20
sig_29
sig_14
sig_13
sig_33

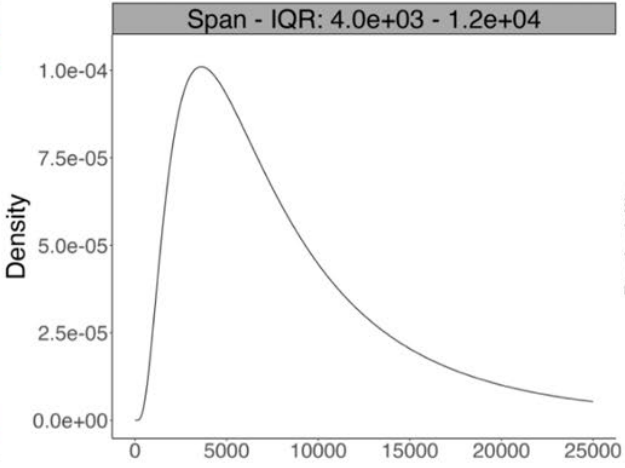

Cancer Type
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
UMAP 2
15
10
5
0
-5
UMAP 1
-5
0
5

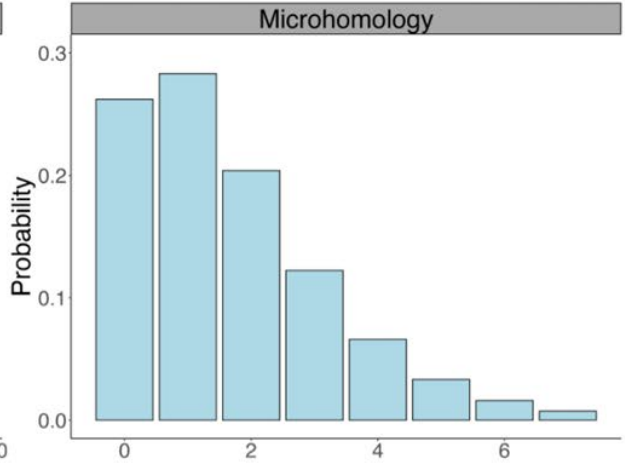

Cancer Type
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
UMAP 2
15
10
5
0
-5
-5
0
5
UMAP 1

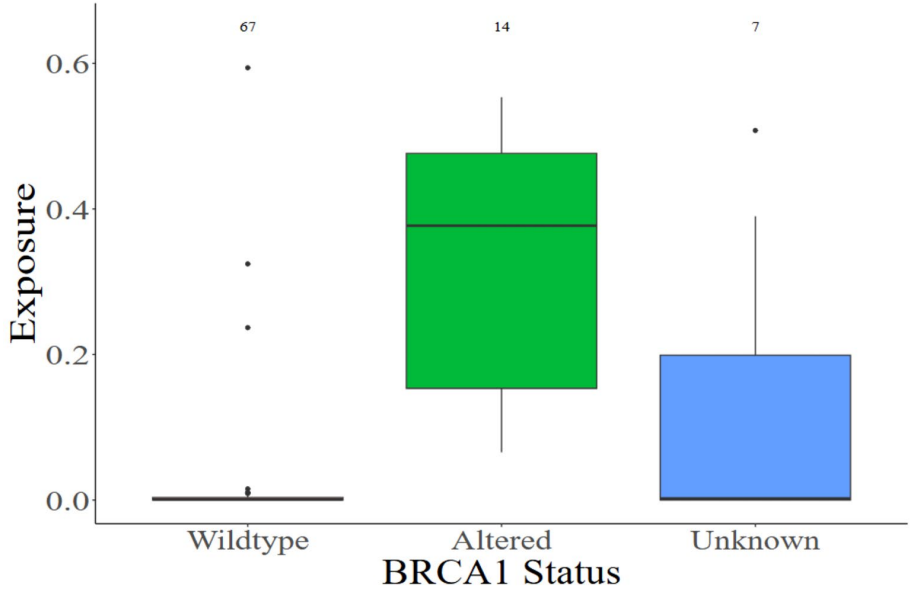

Cancer Type
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
UMAP 2
15
10
5
0
-5
-5
0
5
UMAP 1

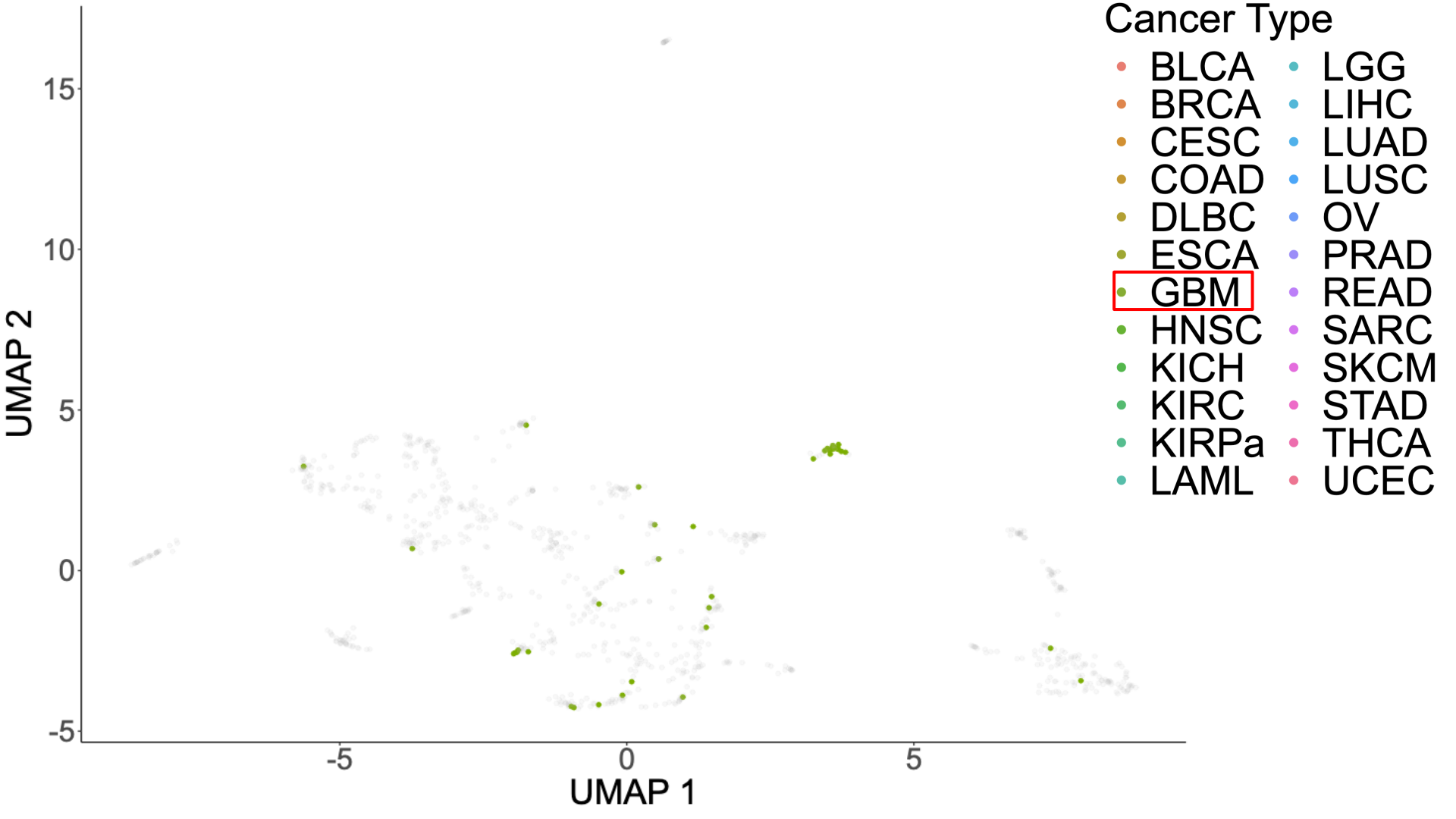

Signature 4 Activity
0.6
0.4
0.2
0.0
15
10
5
0
-5
UMAP 2
-5
0
5
UMAP 1

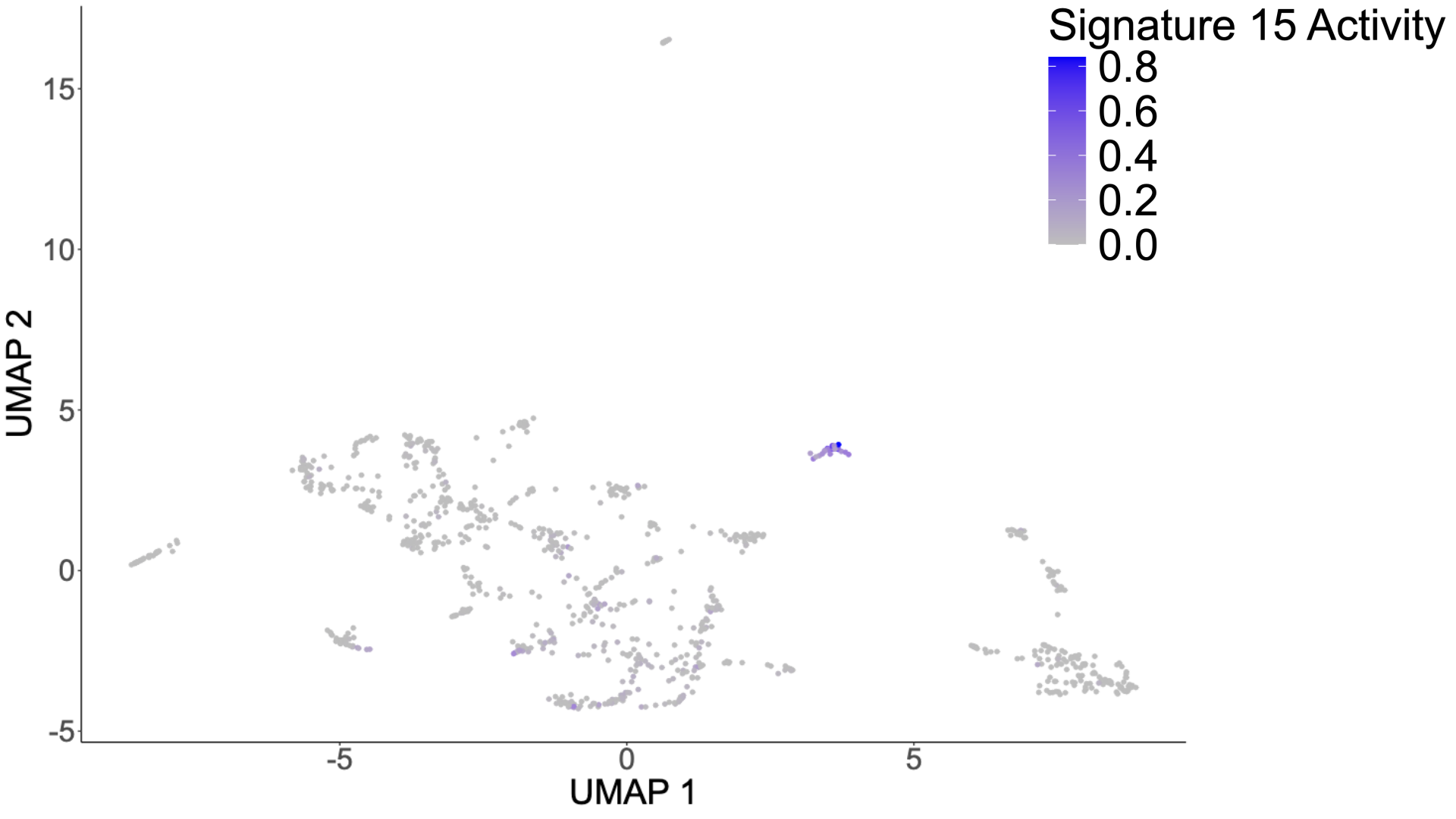

Clustered
Non-Clustered
DEL
DUP
h2hINV
t2tINV
INTER
Proportion
1.00
0.75
0.50
0.25
0.00

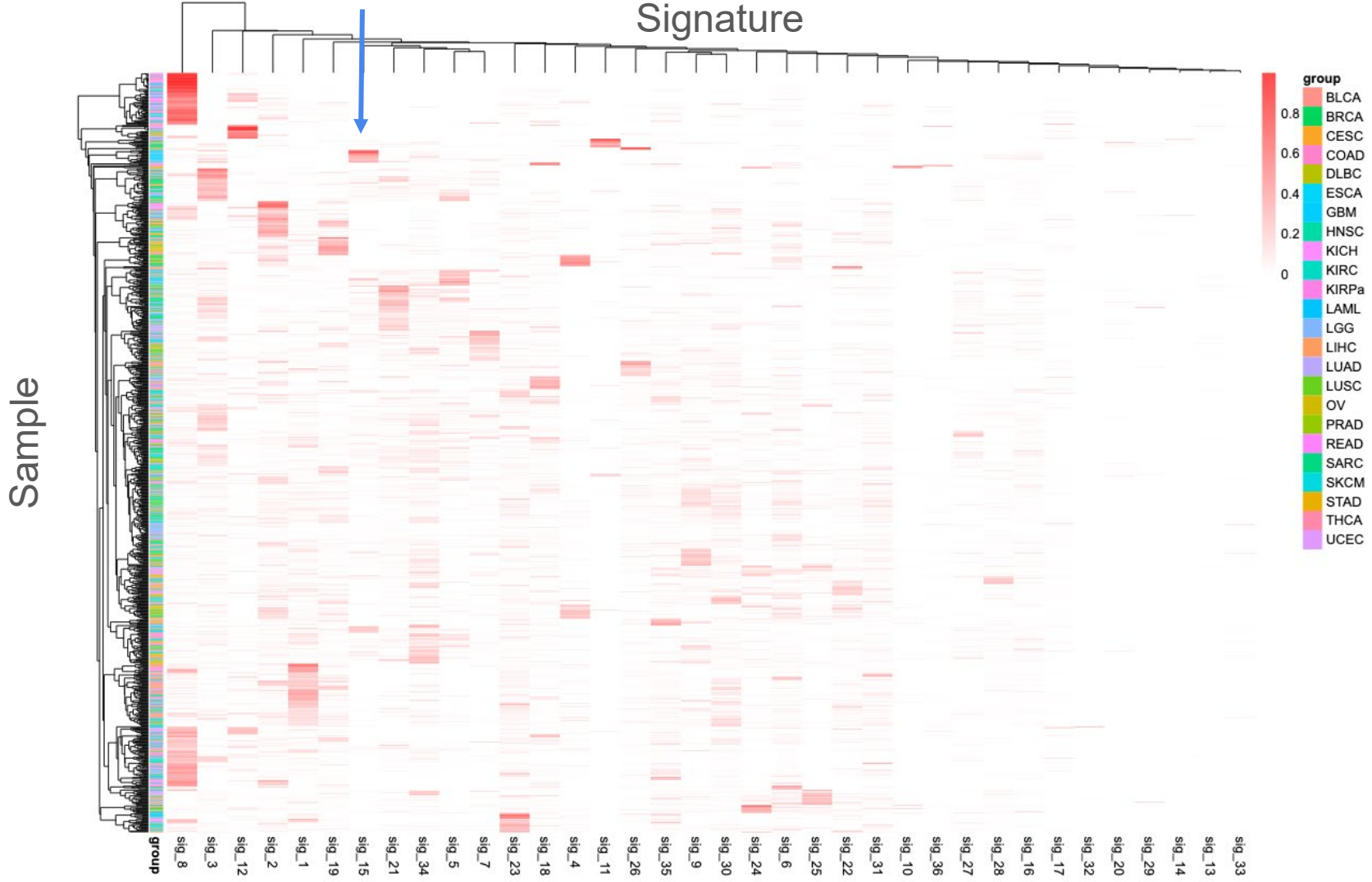

Span - IQR: 4.0e+03 - 1.2e+04
Density
1.0e-04
7.5e-05
5.0e-05
2.5e-05
0.0e+00
0
5000
10000
15000
20000
25000

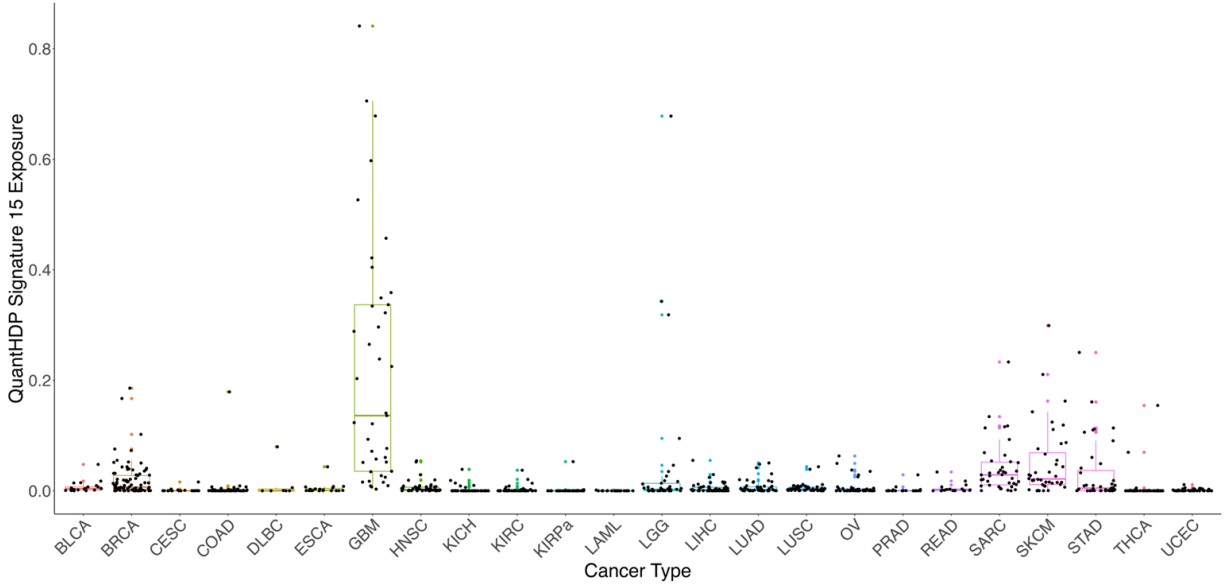

Microhomology
Probability
0.3
0.2
0.1
0.0
0
2
4
6

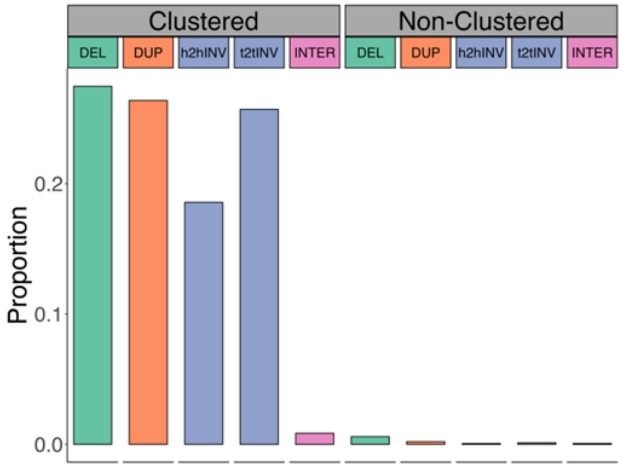

67
14
7
Exposure
0.6
0.4
0.2
0.0
Wildtype
Altered
Unknown
BRCA1 Status

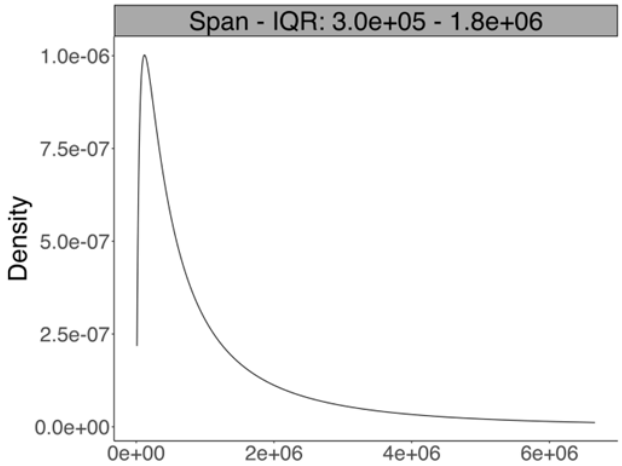

Cancer Type
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
UMAP 2
15
10
5
0
-5
UMAP 1
-5
0
5

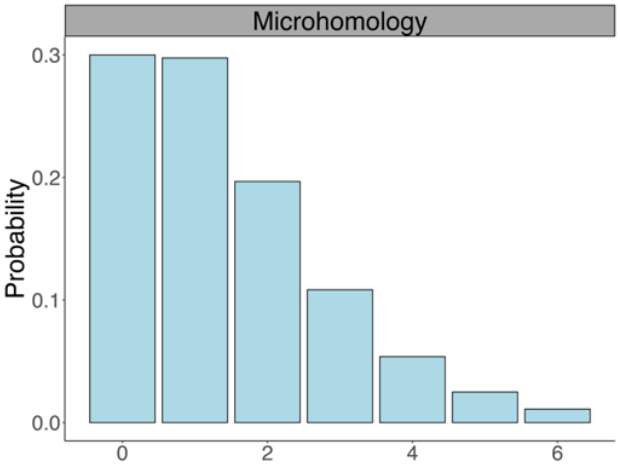

Signature 15 Activity
0.8
0.6
0.4
0.2
0.0
15
10
5
0
-5
UMAP 2
-5
0
5
UMAP 1


Signature
Sample
group
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
0.8
0.6
0.4
0.2
0
group
sig_8
sig_3
sig_12
sig_2
sig_1
sig_19
sig_15
sig_21
sig_34
sig_5
sig_7
sig_23
sig_18
sig_4
sig_11
sig_26
sig_35
sig_9
sig_30
sig_24
sig_6
sig_25
sig_22
sig_31
sig_10
sig_36
sig_27
sig_28
sig_16
sig_17
sig_32
sig_20
sig_29
sig_14
sig_13
sig_33


QuantHDP Signature 15 Exposure
0.0
0.2
0.4
0.6
0.8
BLCA
BRCA
CESC
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRPa
LAML
LGG
LIHC
LUAD
LUSC
OV
PRAD
READ
SARC
SKCM
STAD
THCA
UCEC
Cancer Type


Clustered
Non-Clustered
DEL
DUP
h2hINV
t2tINV
INTER
Proportion
0.0
0.1
0.2


Span - IQR: 3.0e+05 - 1.8e+06
Density
0.0e+00
2.5e-07
5.0e-07
7.5e-07
1.0e-06
0e+00
2e+06
4e+06
6e+06


Microhomology
Probability
0.0
0.1
0.2
0.3
0
2
4
6

Future Directions

- Validate signatures through analysis of independent datasets, including a pan-cancer cohort of metastatic tumors from the Hartwig Medical Foundation
- Establish signature etiologies
  - Correlate signature activities with CRISPR dependencies
  - Identify up/down-regulated gene expression modules in samples with high signature activity

Acknowledgements

- **Rameen Beroukhim**
- Peter Zheng
- Siyun Lee
- Anthony Zhao
- Simona Dalin
- Shahab Sarmashghi
- Sophie Webster
- Wolu Chukwu
- Narmen Azazmeh
- Everyone else in the Beroukhim Lab! :)
- Pratiti Bandopadhayay
- Keith Ligon
- Meraj Ramezani
- Ryan Capouellez
- Ethan Curtis
- Stephen Jasina
- Jonathon Jacobs
- Noor Sohail
- Daniel Kaiser

Dissertation Advisory Committee

- David Pellman
- Peter Park
- Gad Getz