
Discovering Copy Number Variation Across Multiple Cancer Types

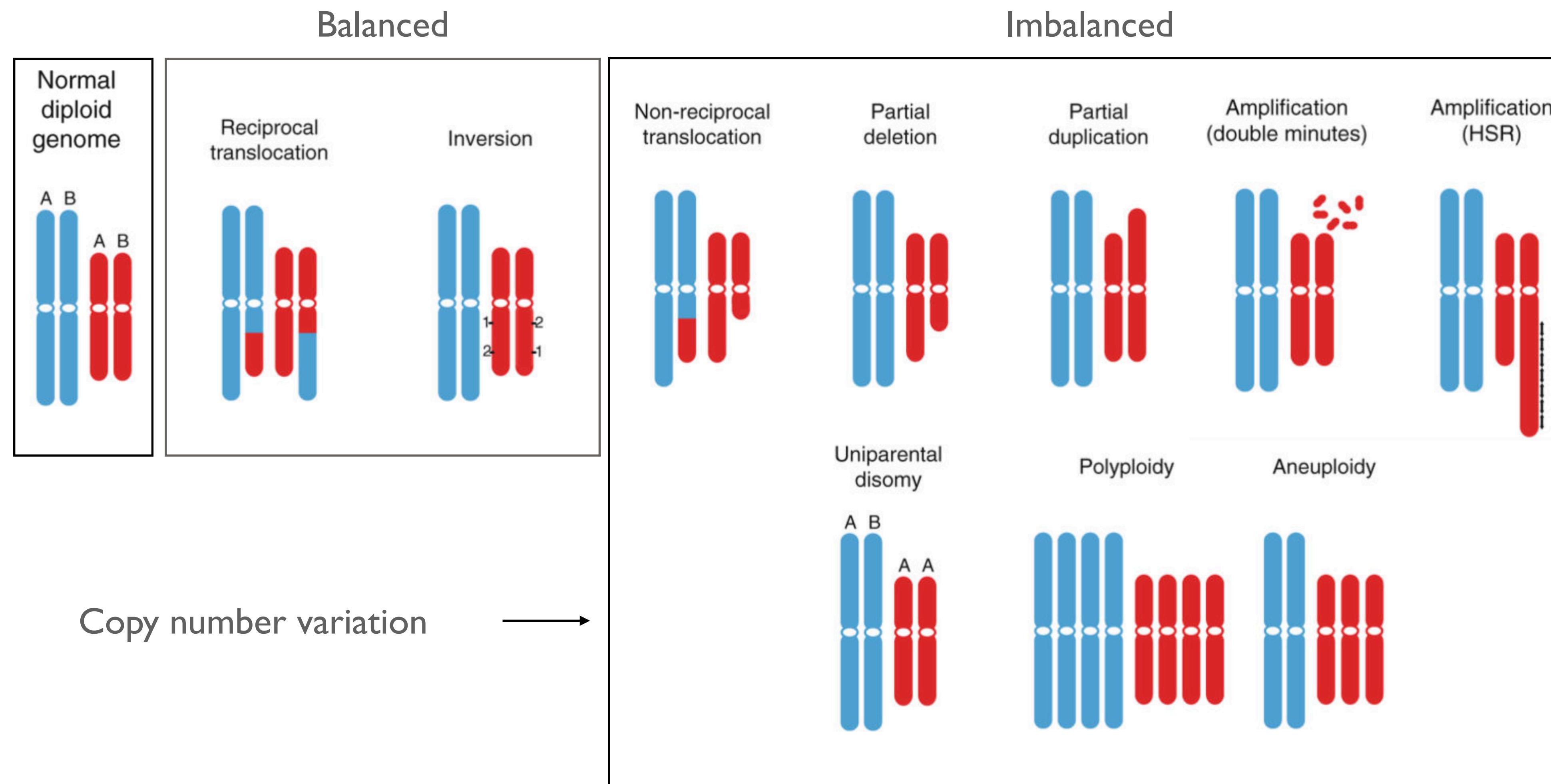
Qingyao Huang
Baudis group UZH
2021-3-25 Zurich Seminars in Bioinformatics

Presentation Outline

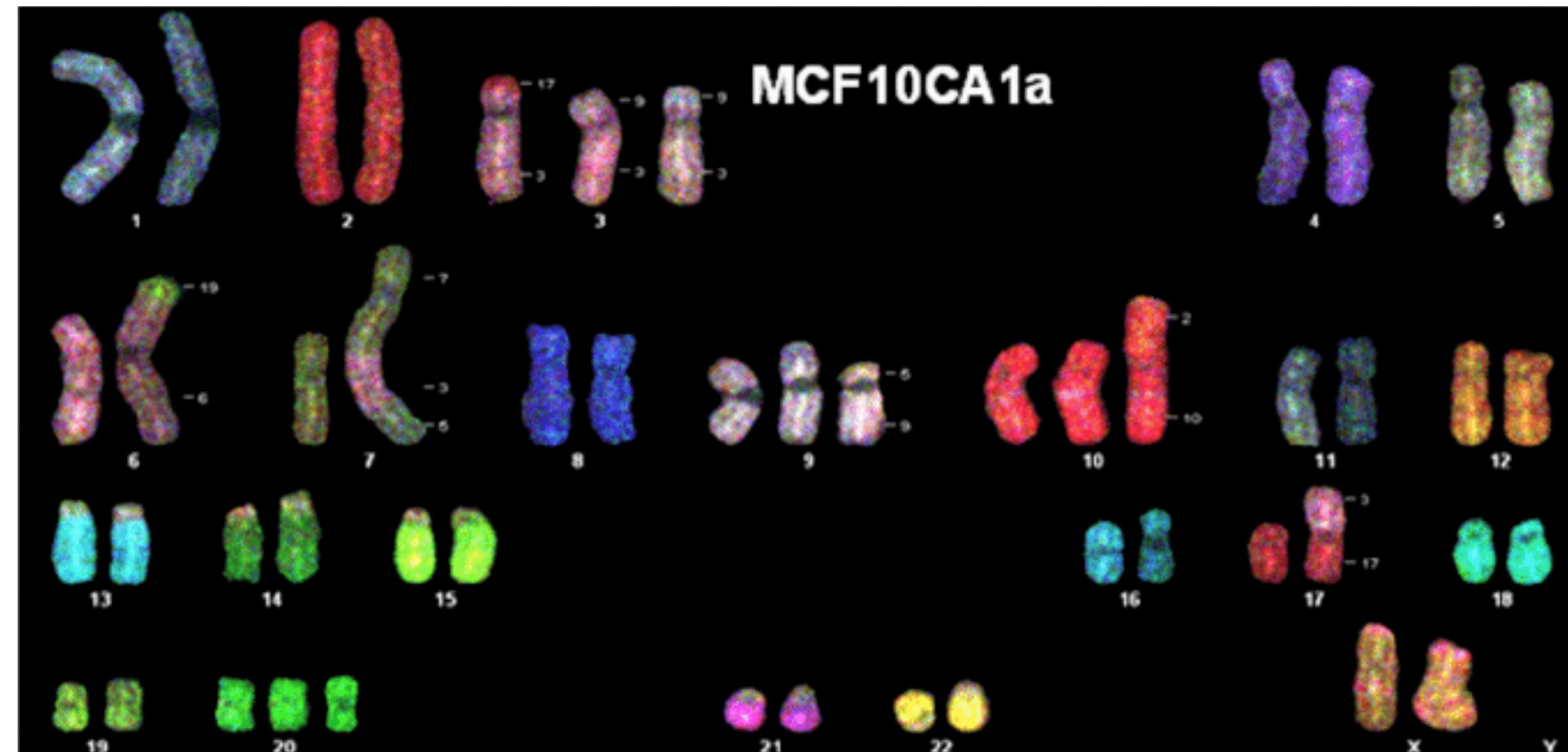
- What is copy number variation (CNV) in cancer?
 - Why is it difficult to interpret the data?
 - What properties are robust and not random?
 - How to capture these properties?
 - Evidences of the CNV-exerted mechanism.
-

What Is Copy Number Variation (CNV)?

Structural Variation in Cancer



Karyotype of Genome With Structural Variation



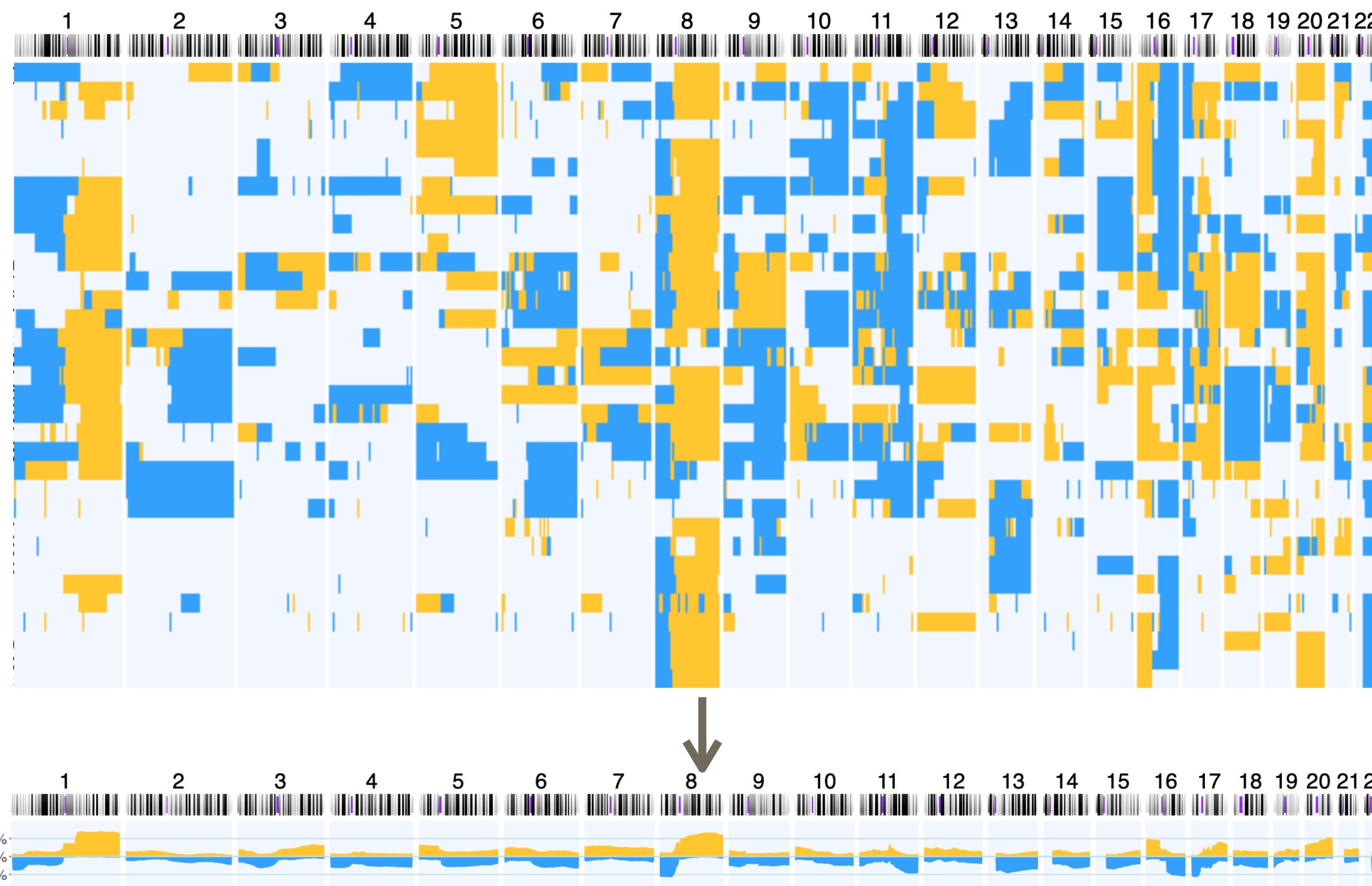
Marella et al., 2009 Cancer Research



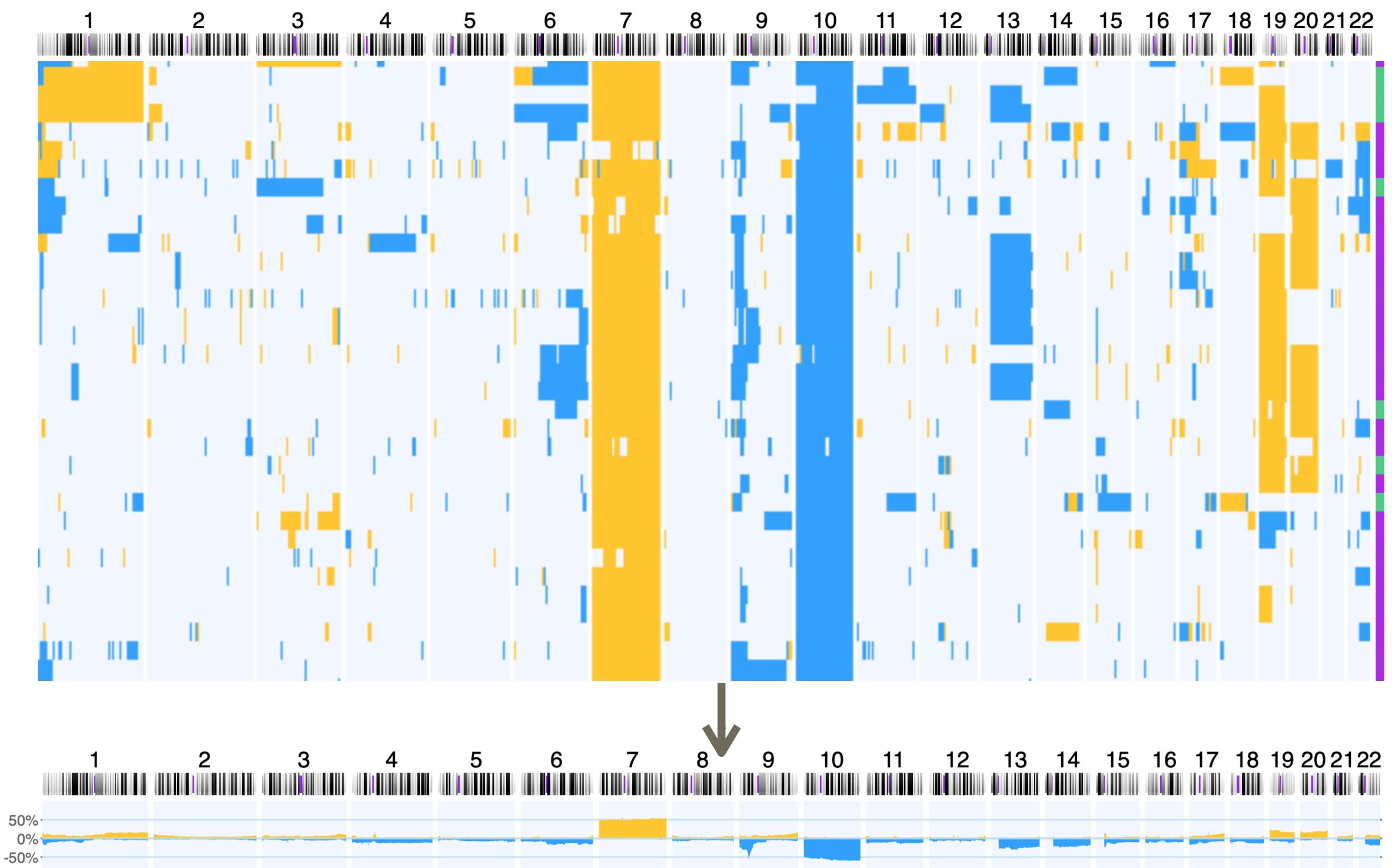
Bessette et al., 2015 PLoS one

Single Samples' Profiles Can Be Random

Ductal Breast Carcinoma



Glioblastoma

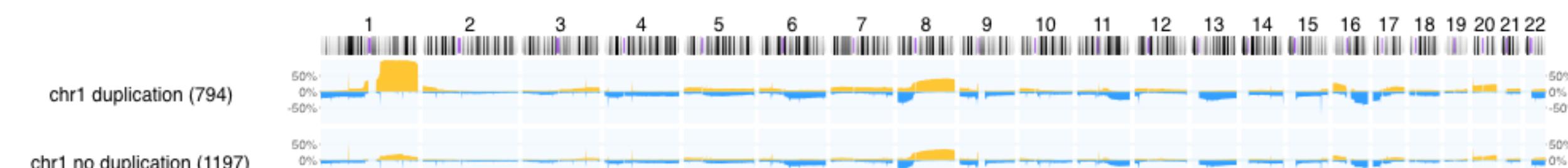


Thousands of genes involved (passengers)
Descriptive report with arbitrary cutoff

Collective CNV Pattern Is Not Random!

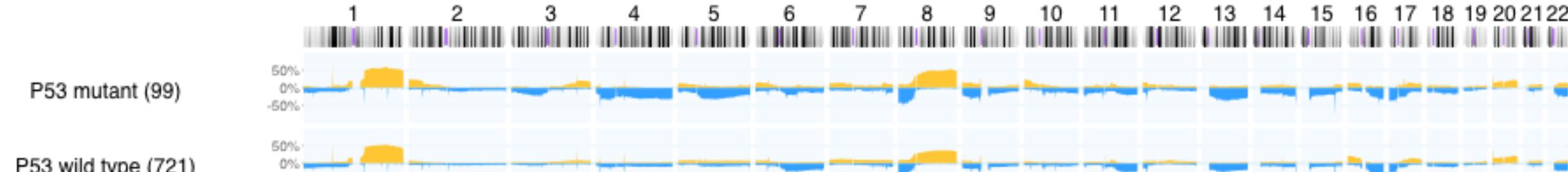
I. Robust

Independent evolution of CNV landscape



2. Regulated

P53 mutant slightly higher CNV



3. Functionally relevant

IntClust BRCA subtypes

A new genome-driven integrated classification of breast cancer and its implications

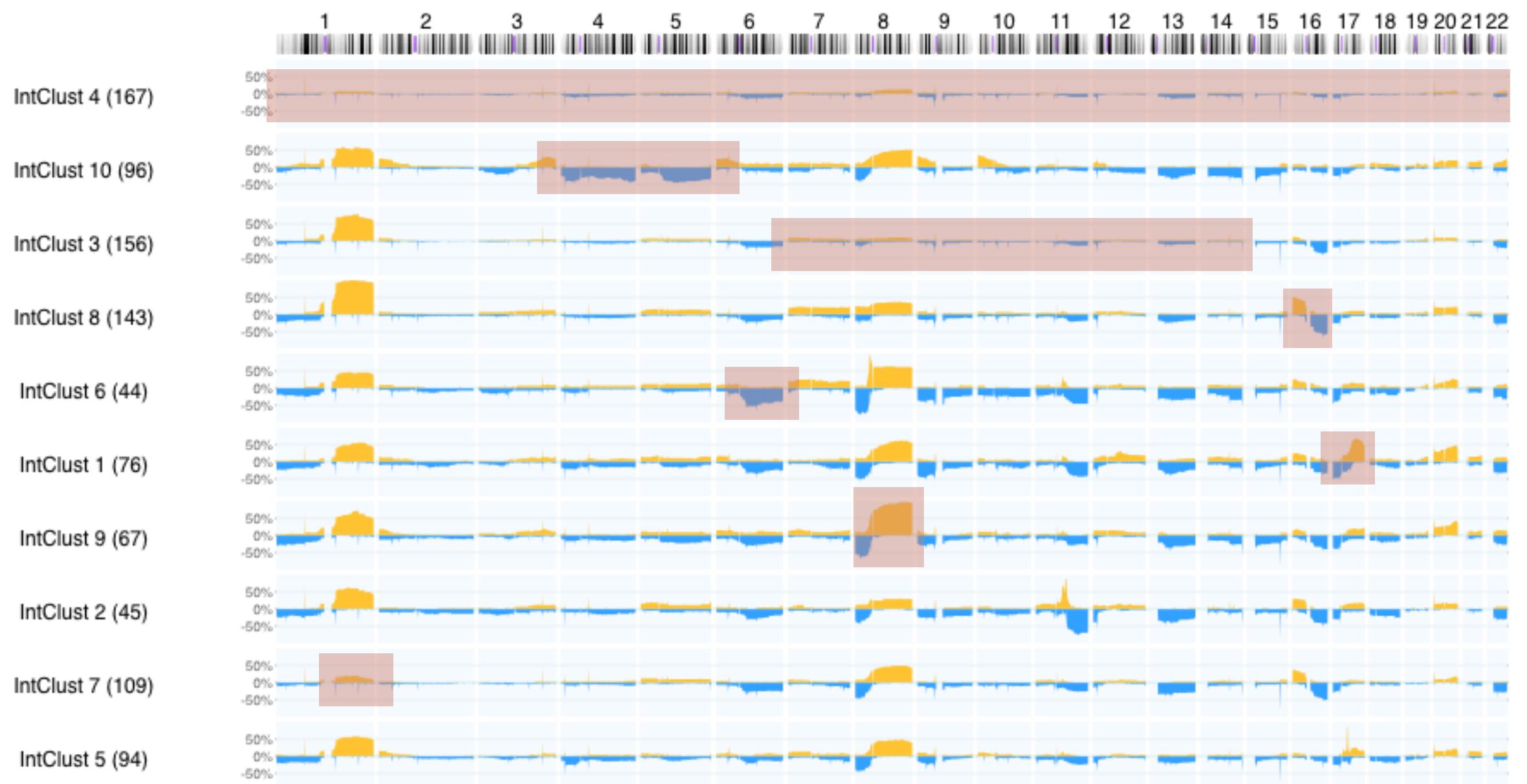
SJ Dawson, OM Rueda, S Aparicio, C Caldas - The EMBO journal, 2013 - embopress.org

Breast cancer is a group of heterogeneous diseases that show substantial variation in their molecular and clinical characteristics. This heterogeneity poses significant challenges not only in breast cancer management, but also in studying the biology of the disease. Recently, rapid progress has been made in understanding the genomic diversity of breast cancer. These advances led to the characterisation of a new genome-driven integrated

classification of breast cancer, which substantially refines the existing classification systems ...

☆ 99 Cited by 278 Related articles All 8 versions ☰

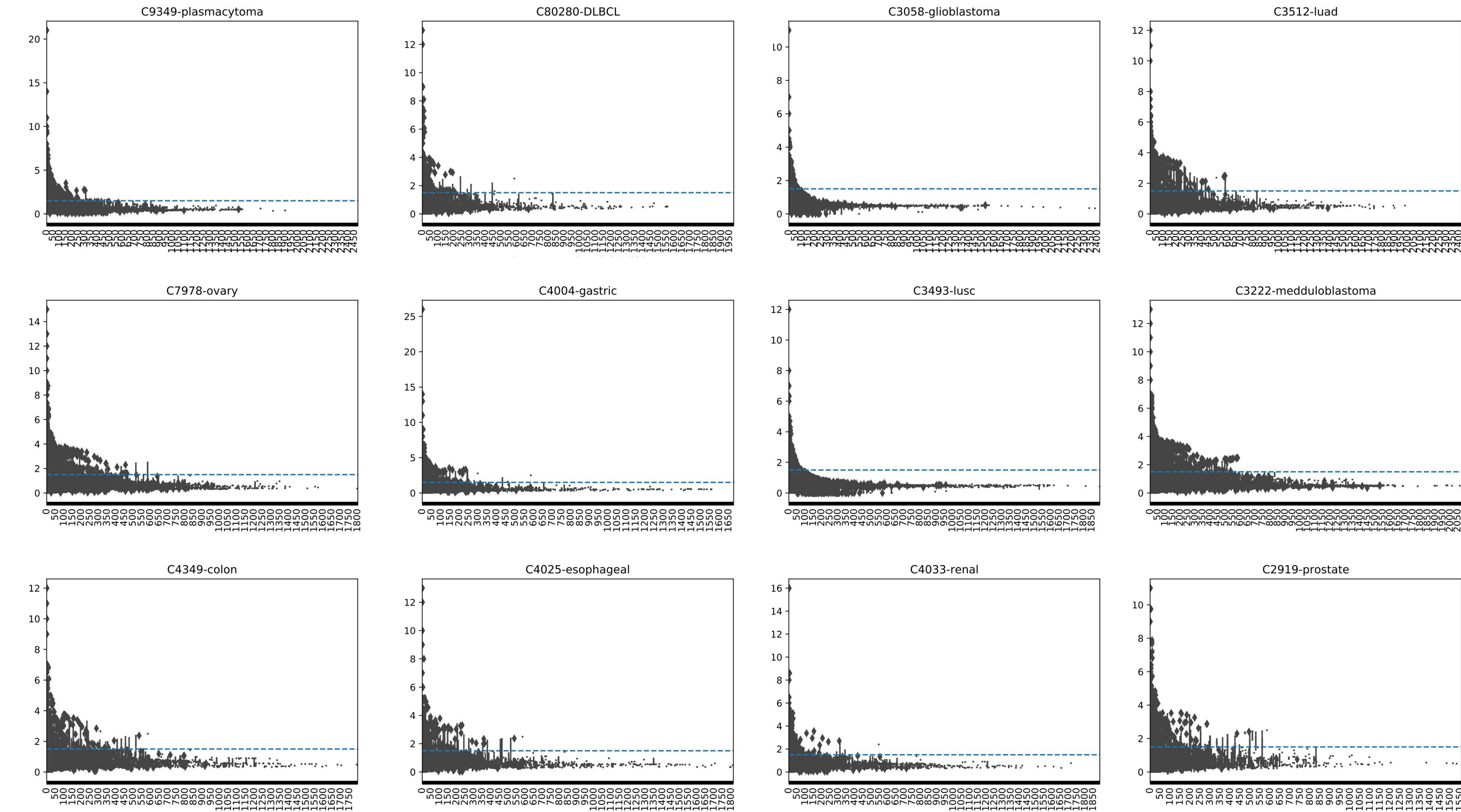
Distinctive CNV patterns among IntClust groups



Observations From Exploratory Analysis

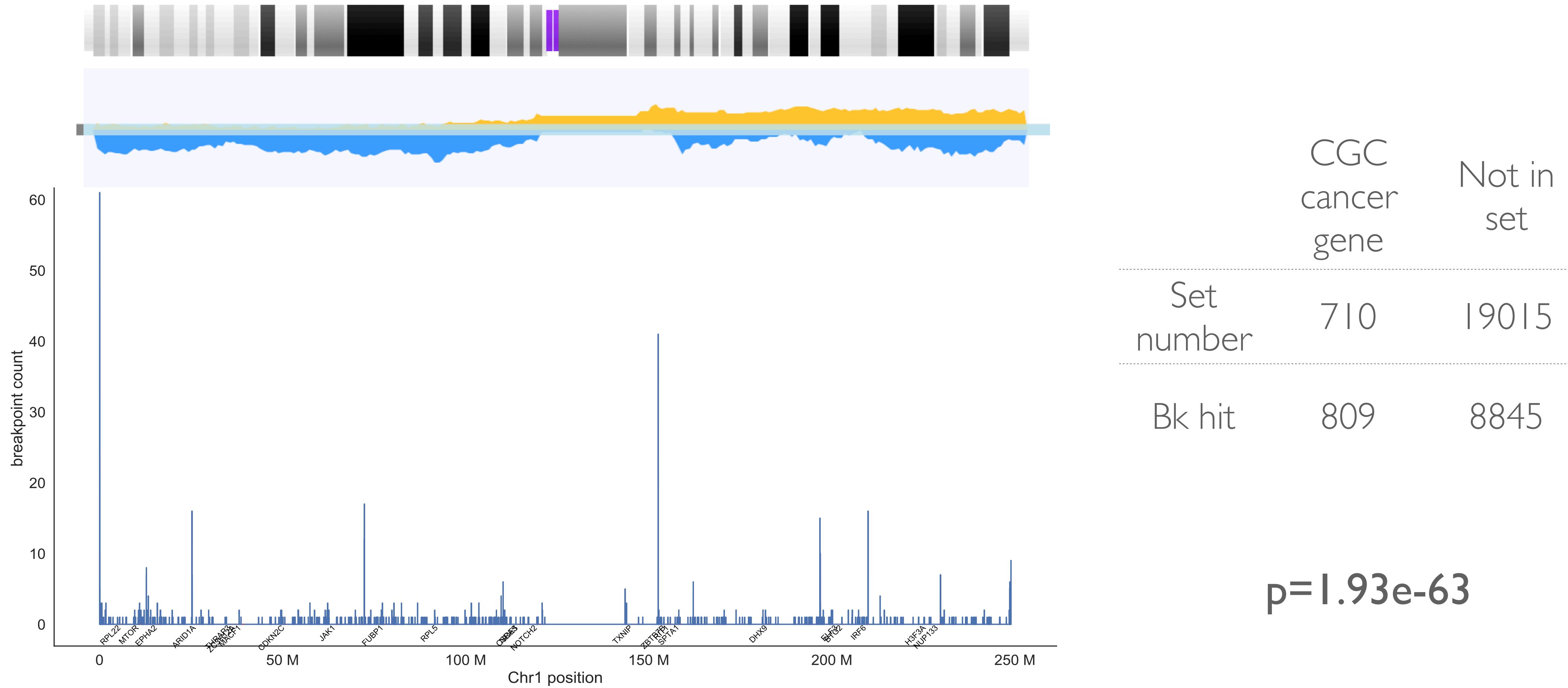
Gene Deletions Tend To Be Focal

Truncated/deleted gene per 100kb



CNV segment length (100kb bin)

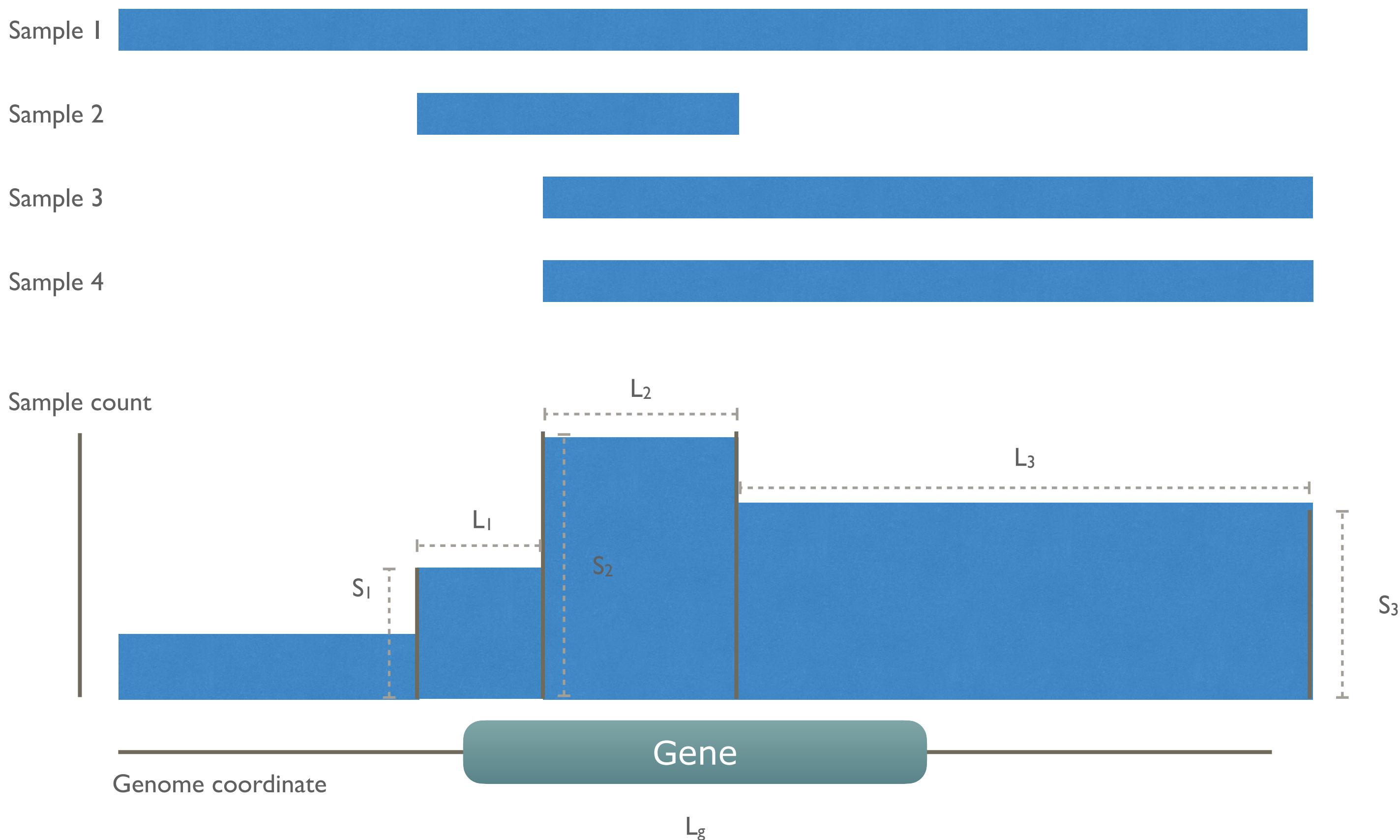
Copy Deletion Breaks Recur at Driver Genes



Metric Definition

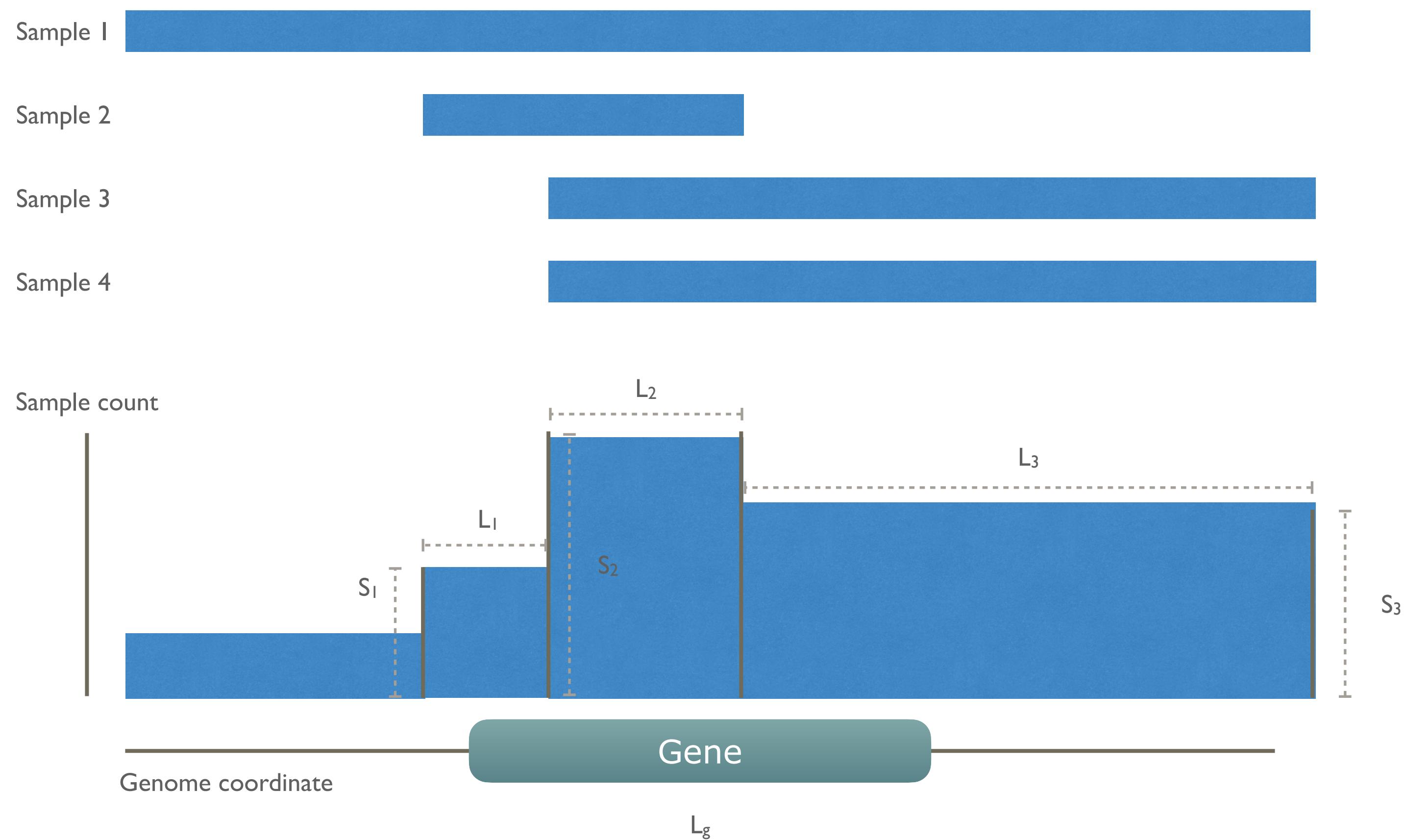
A Gene-Based Metric To Capture These Properties

- Collective landscape is robust
- Deletions tend to be focal
- Recurring breaks at driver genes
- Gene length as confounder

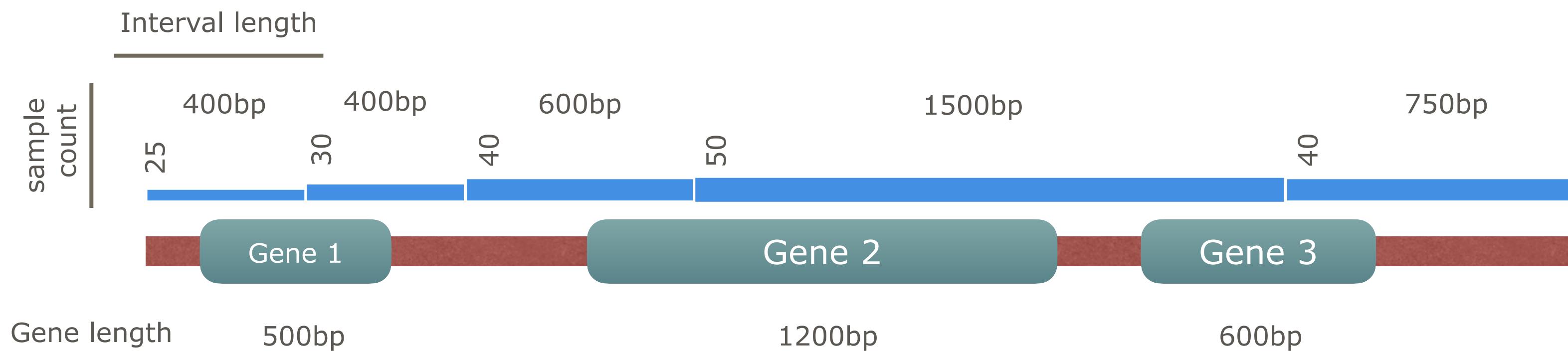


A Gene-Based Metric To Capture These Properties

$$Score_g = \sum_{i=1}^N \left(\frac{S_i}{L_i + L_g} \right)$$

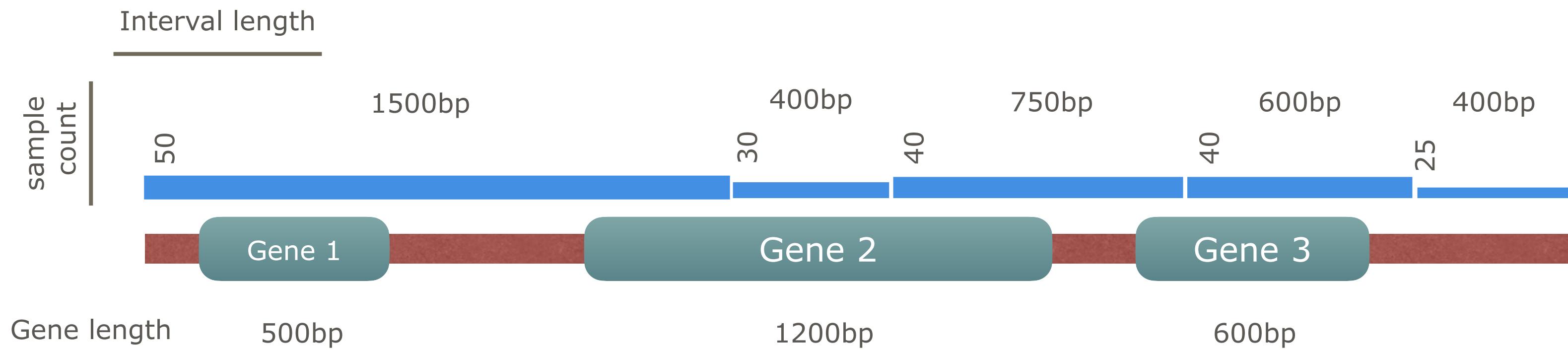


Random Shuffling To Keep the Same Genomic Background



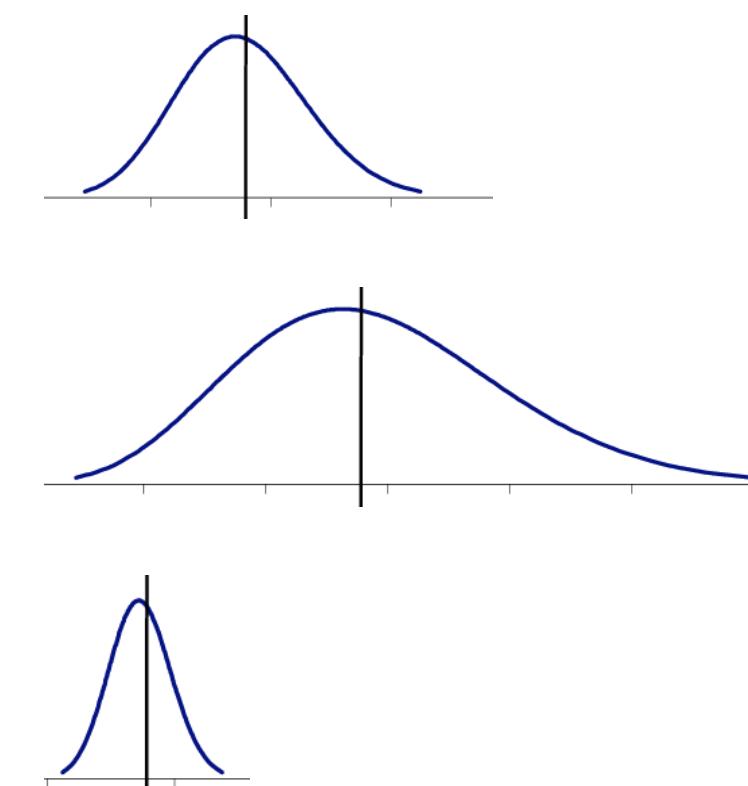
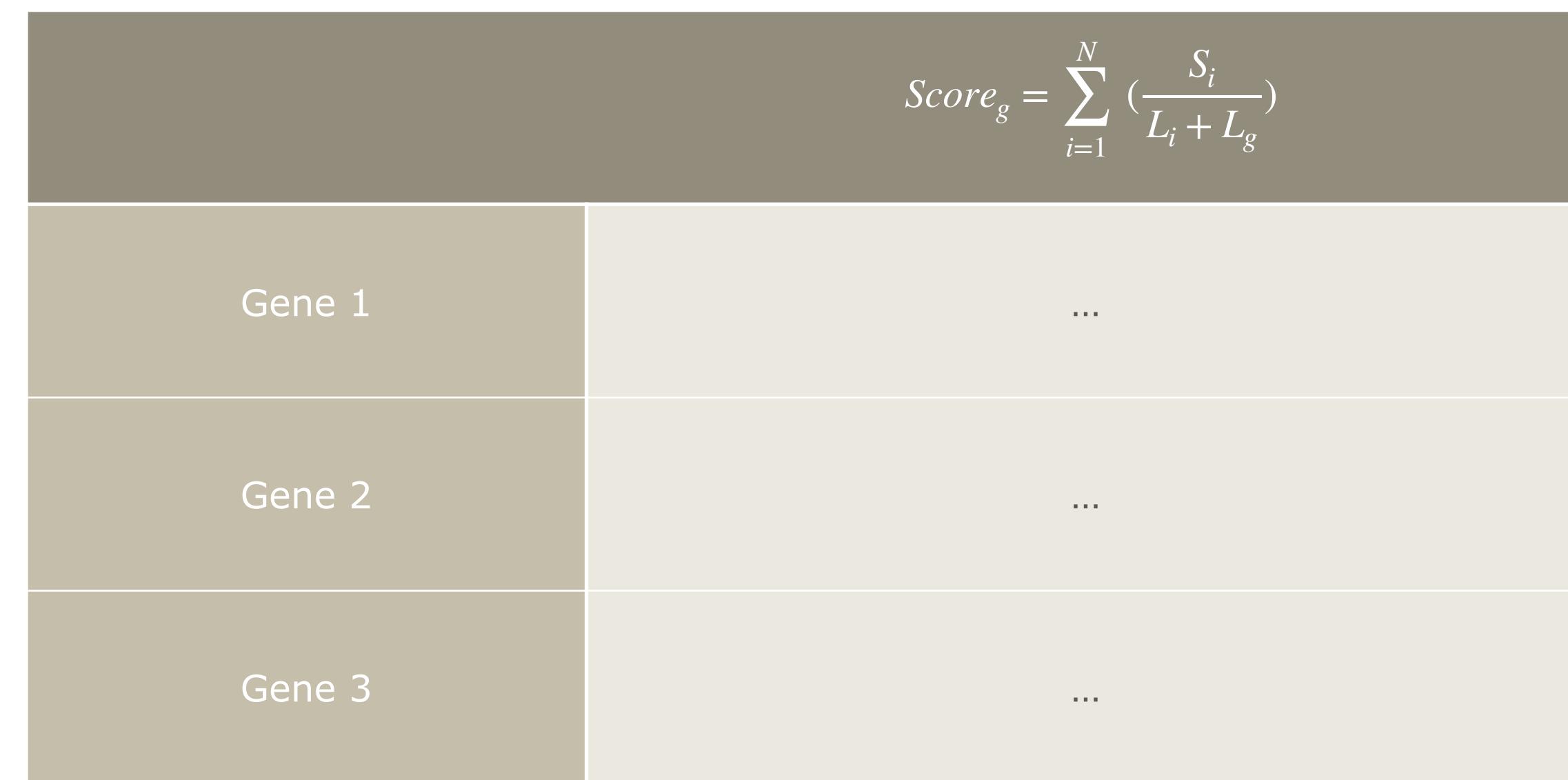
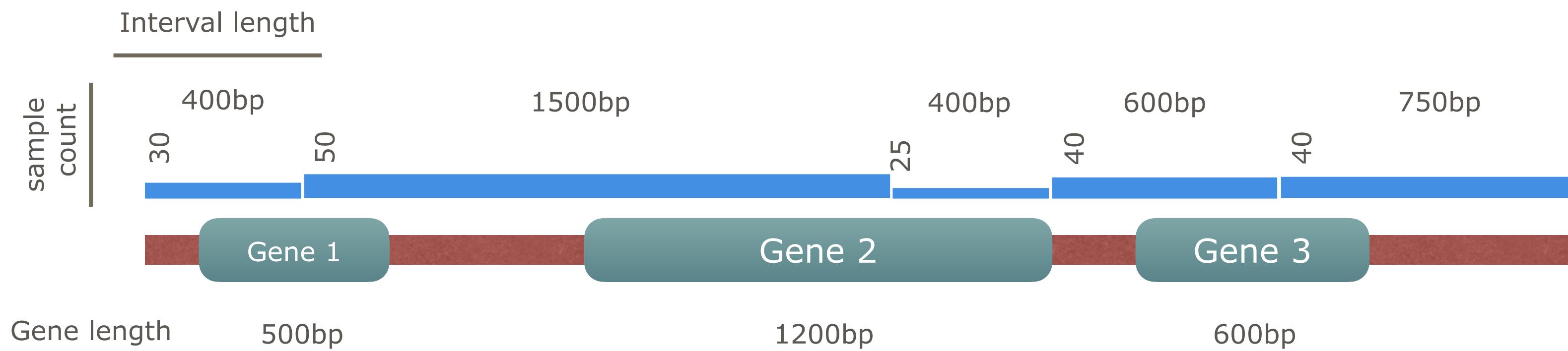
		$Score_g = \sum_{i=1}^N \left(\frac{S_i}{L_i + L_g} \right)$
Gene 1		$25/(400+500) + 30/(400+500) = \mathbf{0.061}$
Gene 2		$40/(600+1200) + 50/(1500+1200) = \mathbf{0.041}$
Gene 3		$50/(1500+600) + 40/(750+600) = \mathbf{0.053}$

Random Shuffling To Keep the Same Genomic Background

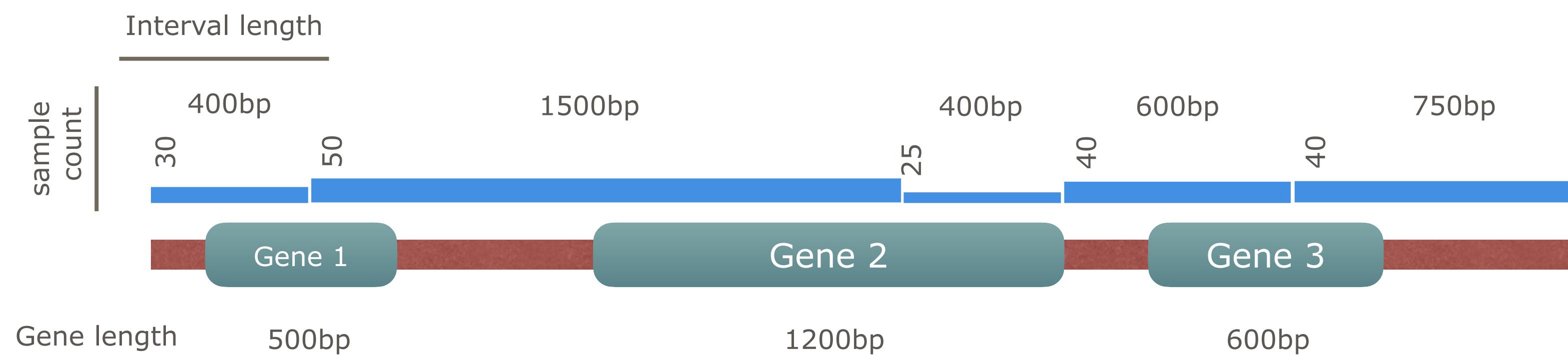


$Score_g = \sum_{i=1}^N \left(\frac{S_i}{L_i + L_g} \right)$	
Gene 1	...
Gene 2	...
Gene 3	...

Random Shuffling To Keep the Same Genomic Background



Random Shuffling To Keep the Same Genomic Background

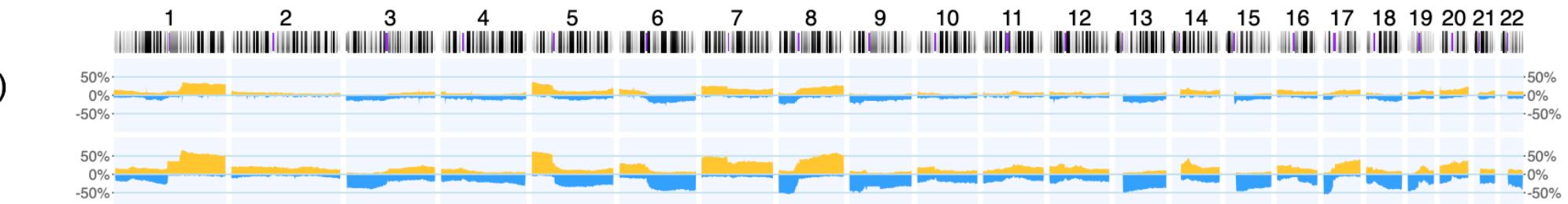


Keep gene neighbourhood context

Robust per CNV extent among samples

Robust per single sample segmentation noise

arrayMap (607)
TCGA (325)



Evidence of Pattern Discovery

- Comparison with GISTIC 2.0
 - Benchmark with multiple cancer types
 - Significant genes overlap with known drivers
 - Functional pathway enrichment
 - Clustering
-

GISTIC2.0

- Most widely-used CNV analysis tool
- Marker: focal regions
- P value for each marker

RESEARCH ARTICLE



Original tool
2007

Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma

Rameen Beroukhim, Gad Getz, Leia Nghiempuh, Jordi Barretina, Teli Hsueh, David Linhart, Igor Vivanco, Jeffrey C. Lee, Julie H. Huang, Sethu Alexander, Jinyan Du, Tweeny Kau, Roman K. Thomas, Kinjal Shah, Horacio Soto, Sven Perner, John Prensner, Ralph M. Debiasi, Francesca Demichelis, Charlie Hatton, Mark A. Rubin, Levi A. Garraway, Stan F. Nelson, Linda Liau, Paul S. Mischel, Tim F. Cloughesy, Matthew Meyerson, Todd A. Golub, Eric S. Lander, Ingo K. Mellinghoff, and William R. Sellers

PNAS December 11, 2007 104 (50) 20007-20012; <https://doi.org/10.1073/pnas.0710052104>

Method | [Open Access](#) | Published: 28 April 2011

GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers

Updated tool
2011

[Craig H Mermel](#), [Steven E Schumacher](#), [Barbara Hill](#), [Matthew L Meyerson](#), [Rameen Beroukhim](#)✉ & [Gad Getz](#)✉
Genome Biology 12, Article number: R41 (2011) | [Cite this article](#)

Article | Published: 18 February 2010

The landscape of somatic copy-number alteration across human cancers

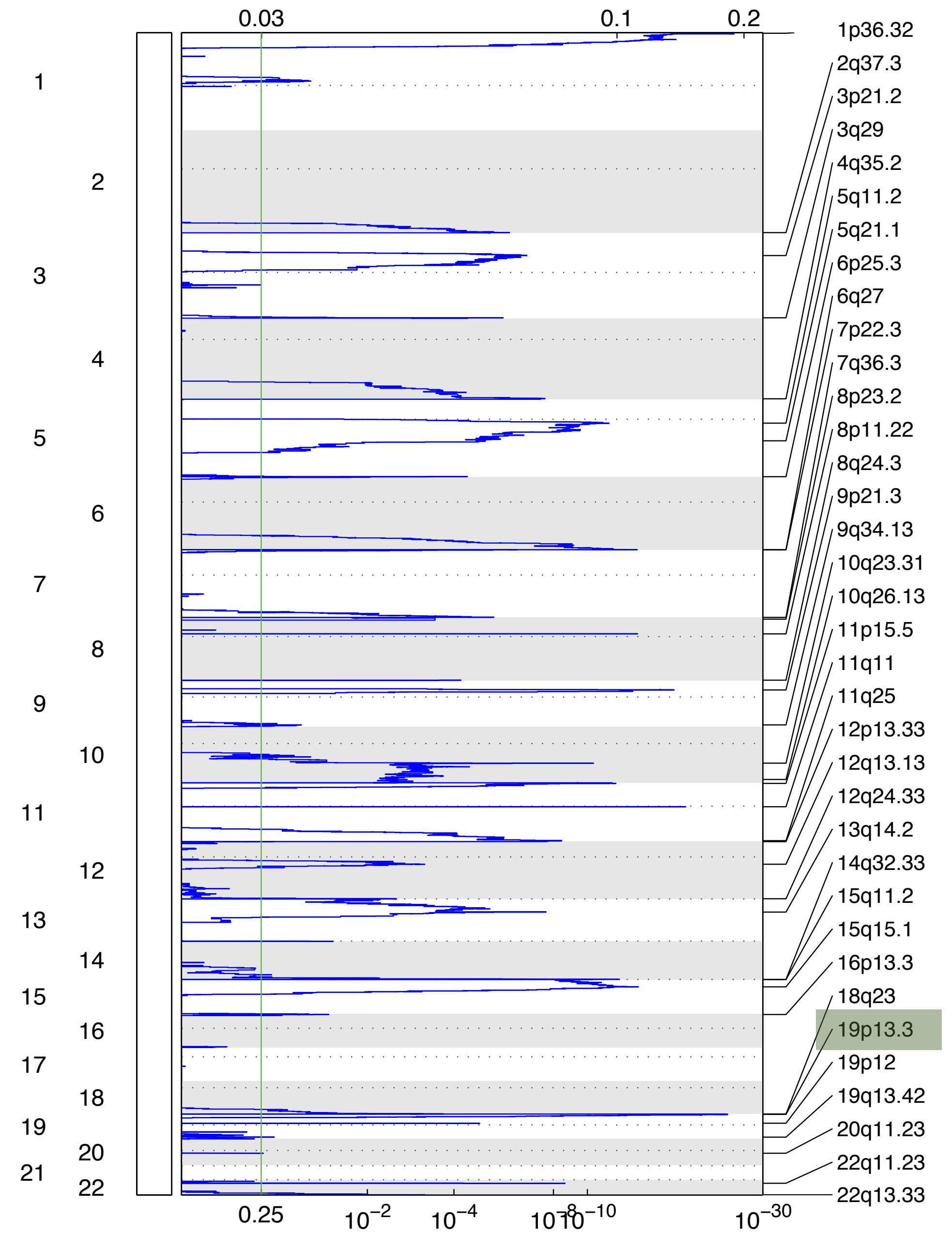
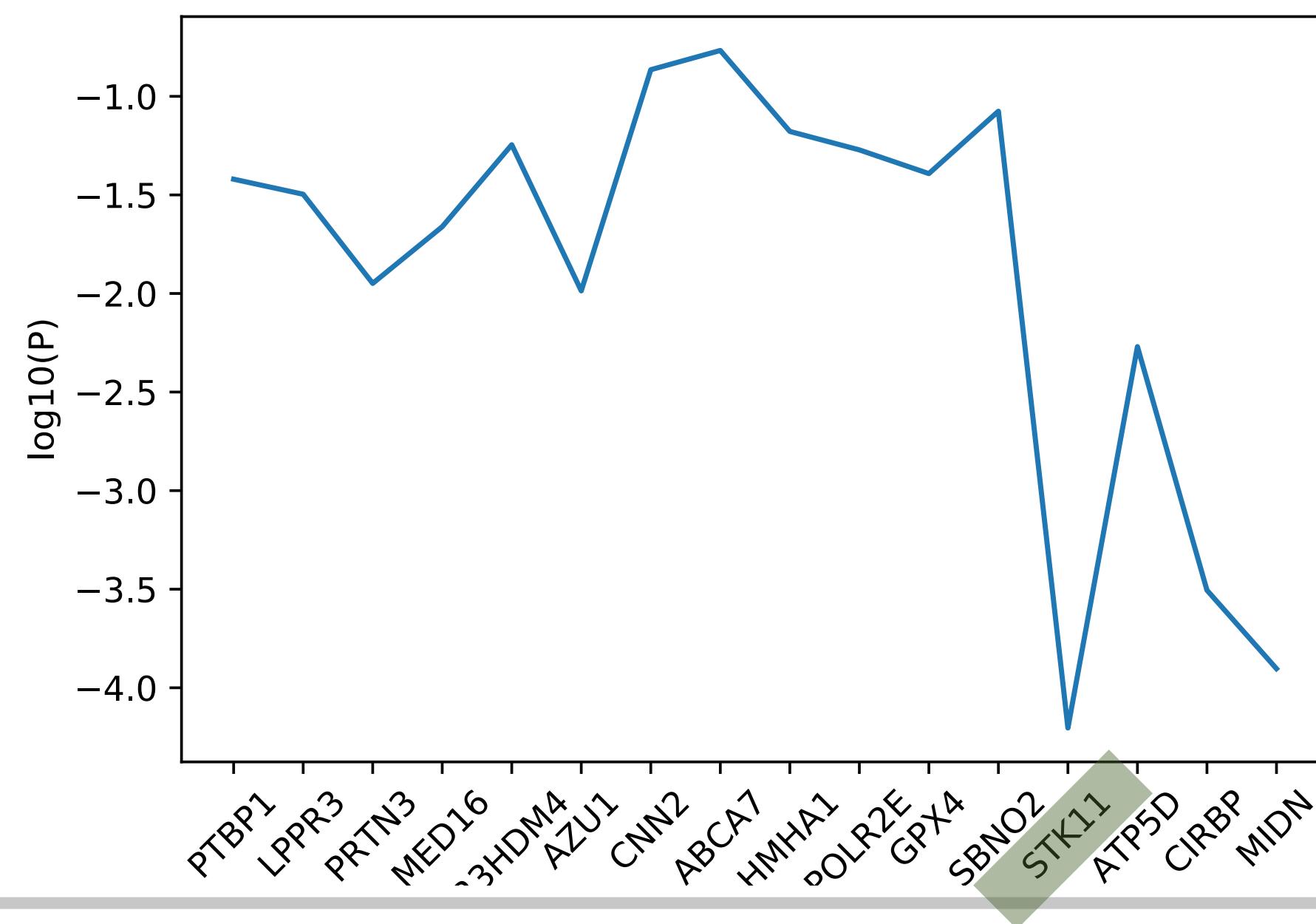
Functional analysis
2010

Rameen Beroukhim, Craig H. Mermel, [...] Matthew Meyerson✉

Nature 463, 899–905(2010) | [Cite this article](#)

Comparison to GISTIC2.0

- GISTIC peaks are retained
- Under same peak, known cancer genes are scored as more significant



Evidence of Pattern Discovery

- Comparison with GISTIC 2.0
- Benchmark with multiple cancer types
 - Significant genes overlap with known drivers
 - Functional pathway enrichment
 - Clustering

Multiple Cancer Types From Different Sources

- cBioPortal (WES), TCGA (SNP6 array), arrayMap (multiple array platforms)
 - C3058-glioblastoma, C3985I-bladder, C4349-colon,
 - C4017-ductal breast, C7978-ovary,
 - C3512-lung adeno, C3493-lung squamous,
 - C4033-renal, C2919-prostate.

Glioblastoma Multiforme

arrayMap (326)

cBioPortal (547)

TCGA (607)

arrayMap (326)

cBioPortal (547)

TCGA (607)

Colon adenocarcinoma

arrayMap (402)

TCGA (283)

cBioPortal (833)

arrayMap (402)

TCGA (283)

cBioPortal (833)

Clear cell renal cell carcinoma

TCGA (520)

arrayMap (242)

TCGA (520)

arrayMap (242)

Lung squamous cell carcinoma

arrayMap (358)

TCGA (467)

arrayMap (358)

TCGA (467)

Bladder urothelial carcinoma

TCGA (343)

cBioPortal (425)

TCGA (343)

cBioPortal (425)

Ovarian serous cystadenocarcinoma

TCGA (567)

arrayMap (324)

TCGA (567)

arrayMap (324)

Lung adenocarcinoma

arrayMap (607)

TCGA (325)

arrayMap (607)

TCGA (325)

Ductal breast carcinoma

TCGA (778)

arrayMap (816)

TCGA (778)

arrayMap (816)

Prostate adenocarcinoma

TCGA (290)

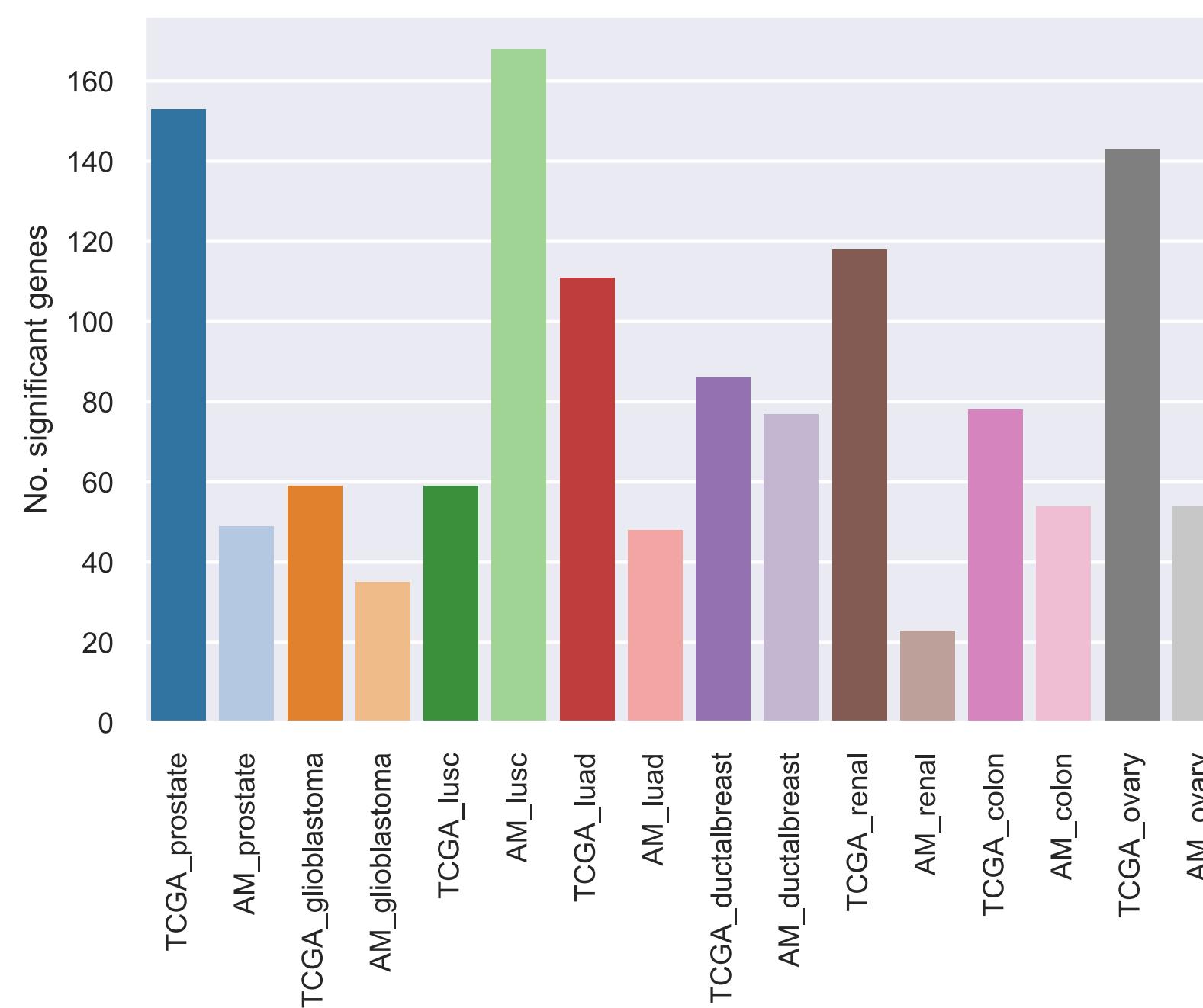
arrayMap (281)

TCGA (290)

arrayMap (281)

© 2021 progenetix.org

20 - 170 Genes Are Significant in Each Analysis



Bailey
299

Dietlein
461

CGC
724

[HTML] Comprehensive characterization of cancer driver genes and mutations

MH Bailey, C Tokheim, E Porta-Pardo, S Sengupta... - Cell, 2018 - Elsevier

Identifying molecular cancer drivers is critical for precision oncology. Multiple advanced algorithms to identify drivers now exist, but systematic attempts to combine and optimize them on large datasets are few. We report a PanCancer and PanSoftware analysis spanning 9,423 tumor exomes (comprising all 33 of The Cancer Genome Atlas projects) and using 26 computational tools to catalog driver genes and mutations. We identify 299 driver genes with implications regarding their anatomical sites and cancer/cell types. Sequence-and structure ...

☆ 99 Cited by 949 Related articles All 55 versions

[HTML] Identification of cancer driver genes based on nucleotide context

F Dietlein, D Weghorn, A Taylor-Weiner, A Richters... - Nature ..., 2020 - nature.com

Cancer genomes contain large numbers of somatic mutations but few of these mutations drive tumor development. Current approaches either identify driver genes on the basis of mutational recurrence or approximate the functional consequences of nonsynonymous mutations by using bioinformatic scores. Passenger mutations are enriched in characteristic nucleotide contexts, whereas driver mutations occur in functional positions, which are not necessarily surrounded by a particular nucleotide context. We observed that mutations in ...

☆ 99 Cited by 36 Related articles All 7 versions

[HTML] The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers

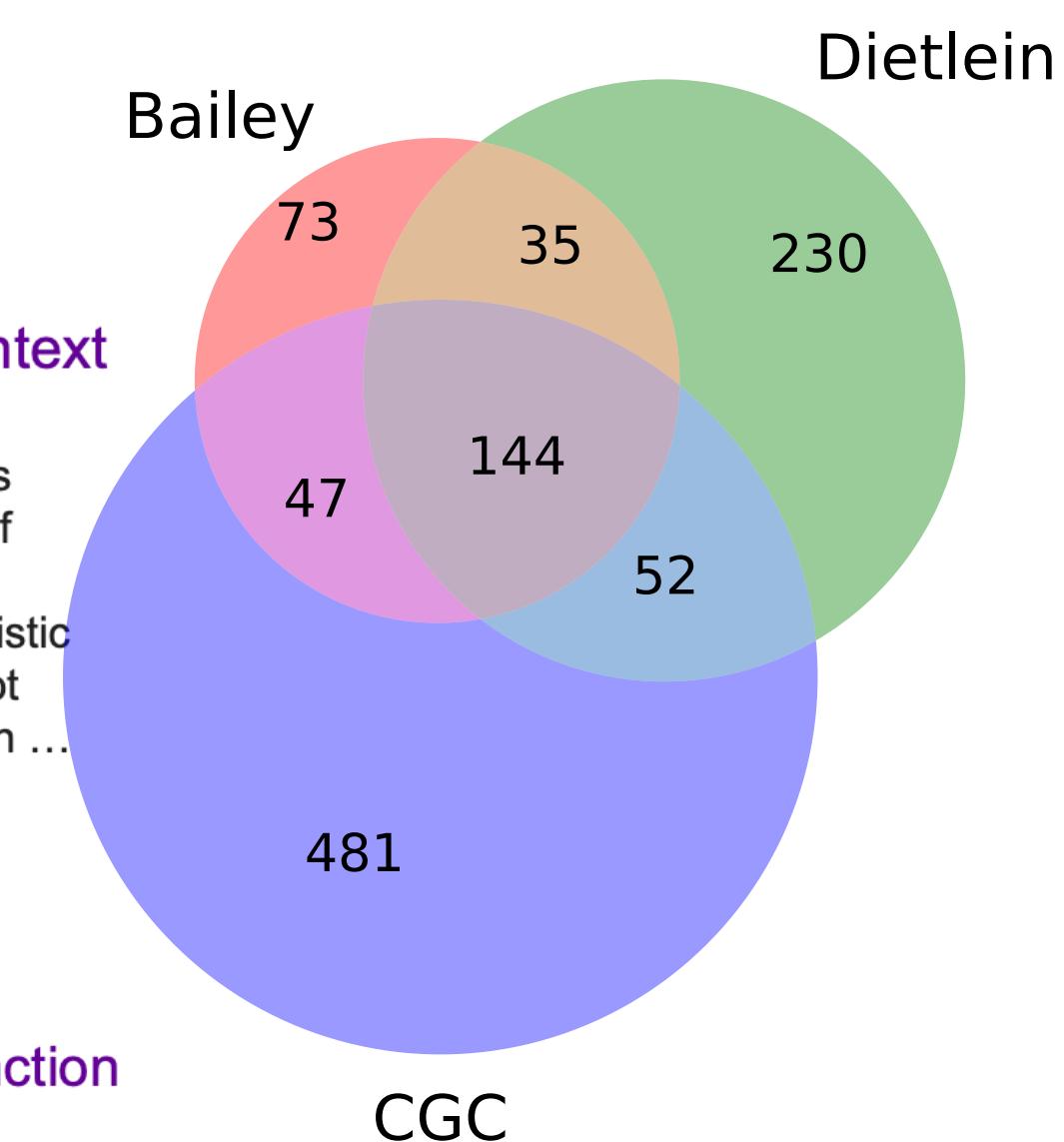
Z Sondka, S Bamford, CG Cole, SA Ward... - ... Reviews Cancer, 2018 - nature.com

The COSMIC **Cancer Gene Census**: describing **genetic** dysfunction across all human **cancers**.

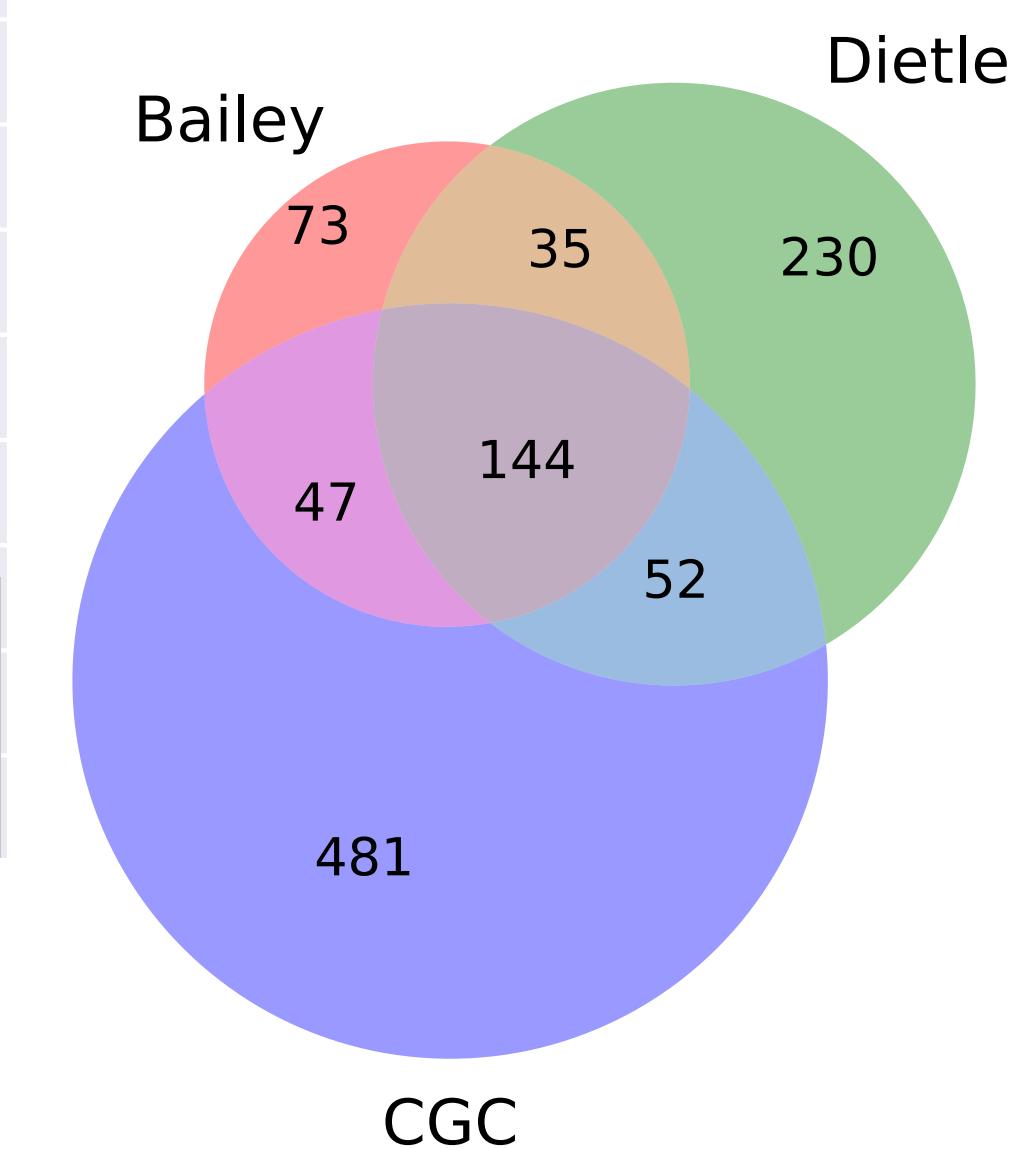
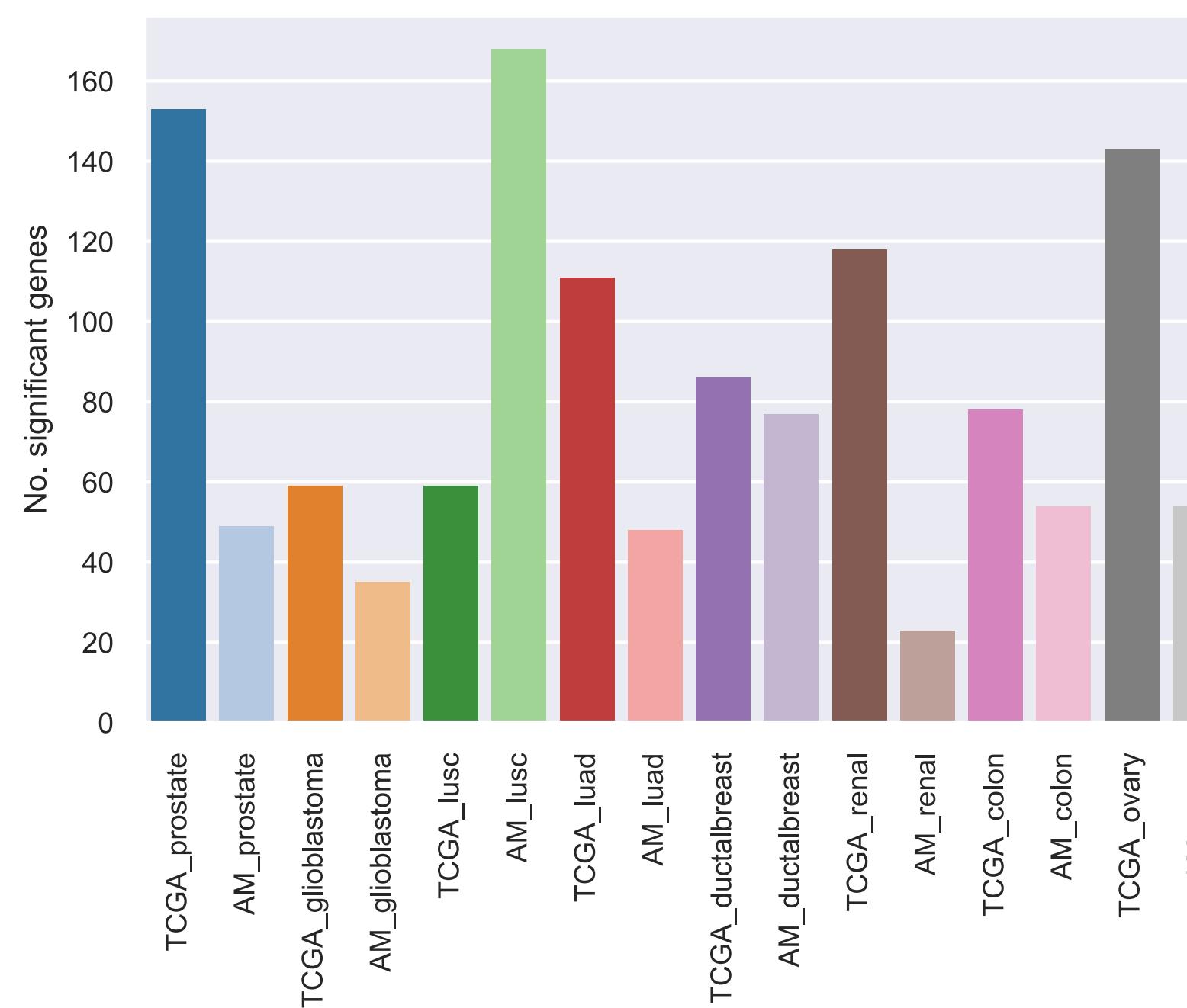
Download PDF. Review Article; Published: 06 October 2018. The COSMIC **Cancer Gene**

Census: describing **genetic** dysfunction across all human **cancers** ...

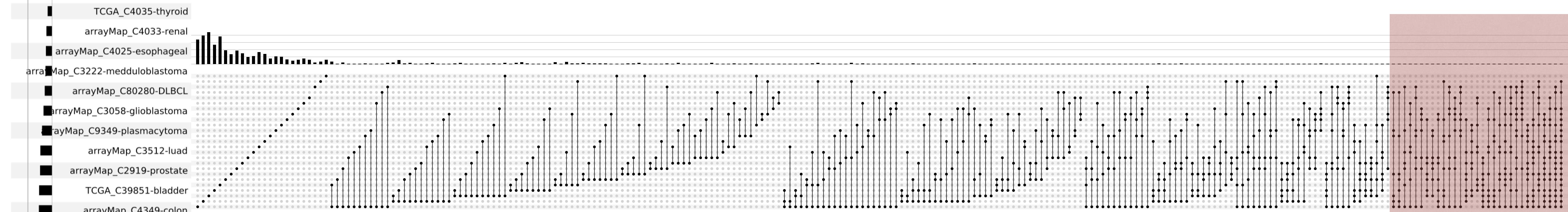
☆ 99 Cited by 353 Related articles All 5 versions



20 - 170 Genes Are Significant in Each Analysis



source	Cancer Type	Significant	in Bailey	Fisher p	in Dietlein	Fisher p	In CGC	Fisher p
TCGA	C2919-prostate	153	8	3.54E-03	15	7.38E-06	17	1.02E-04
arrayMap	C2919-prostate	49	5	9.82E-04	5	6.19E-03	8	4.42E-04
TCGA	C3058-glioblastoma	59	4	1.42E-02	4	5.5E-02	6	2.44E-02
arrayMap	C3058-glioblastoma	35	2	1.02E-01	2	2.02E-01	3	1.41E-01
TCGA	C3099-liver	127	8	1.03E-03	9	4.12E-03	12	3.58E-03
arrayMap	C3222-medulloblastoma	27	2	6.41E-02	2	1.33E-01	2	2.64E-01
TCGA	C3493-lusc	59	7	4.19E-05	6	3.08E-03	12	2.02E-06
arrayMap	C3493-lusc	168	10	3.32E-04	12	7.69E-04	16	6.06E-04
TCGA	C3512-luad	111	4	9.93E-02	6	5.4E-02	14	9.03E-05
arrayMap	C3512-luad	48	5	8.55E-04	3	1.04E-01	7	1.87E-03
TCGA	C39851-bladder	53	7	2.06E-05	6	1.78E-03	9	1.64E-04
arrayMap	C4004-gastric	86	13	1.06E-09	7	4.58E-03	18	2.98E-09
TCGA	C4017-ductalbreast	86	8	7.14E-05	6	1.85E-02	10	1.63E-03
arrayMap	C4017-ductalbreast	77	11	3.33E-08	9	8.99E-05	12	2.71E-05
arrayMap	C4025-esophageal	25	3	6.91E-03	0	1E+00	4	1.38E-02
TCGA	C4033-renal	118	5	4.17E-02	8	9.06E-03	13	6.73E-04
arrayMap	C4033-renal	23	0	1E+00	1	4.4E-01	1	6E-01
TCGA	C4035-thyroid	18	0	1E+00	0	1E+00	0	1E+00
TCGA	C4349-colon	78	6	1.6E-03	8	6.73E-04	11	1.97E-04
arrayMap	C4349-colon	54	4	9.6E-03	1	1E+00	3	4.53E-01
TCGA	C6287-endometrium	57	6	2.89E-04	4	4.98E-02	7	5.81E-03
TCGA	C7978-ovary	143	17	2.17E-10	16	5.44E-07	21	1.5E-07
arrayMap	C7978-ovary	54	7	2.08E-05	5	9.27E-03	8	8.6E-04
arrayMap	C80280-DLBCL	30	2	7.79E-02	4	5.35E-03	2	3.08E-01
arrayMap	C9349-plasmacytoma	41	5	4.5E-04	5	2.98E-03	3	2.01E-01



At least 7

Cancer gene	Count
RB1	16
PTEN	15
CDKN2A	12
PTPRD	11
SMAD2	9
NRG1	9
JAK2	8
FHIT	8
DLC1	8
SMAD4	7
MAP2K4	7
RET	6
LRP1B	6
BRCA2	5
EPHA7	5
MLLT3	5
KANSL1	5
CTNNB1	5
APC	5
FGFR1	5
NCOR1	5
FLCN	5

Cancer gene	Count
ROBO2	4
EPHA3	4
CDKN1B	4
GPS2	4
HLA-A	4
BAP1	4
CHEK2	4
ID3	4
RAD51B	4
PRDM1	4
MTOR	4
RASA1	4
EP300	4
ARHGEF10	4
IRF2	4
PCDH9	3
EYS	3
FLT3	3
PTCH1	3
PRSS50	3
FOXP1	3
TEC	3
DHX30	3
PDGFRA	3

Cancer gene	Count
PGR	3
WT1	3
TMPRSS2	3
KIT	3
ABL1	3
NF1	3
GOPC	3
DICER1	3
ROS1	3
MAP2K1	3
B2M	3
NPM1	3
ATM	3
CXCR4	3
TNFAIP3	3
TP53	3
MALT1	3
EIF4A2	3
PBRM1	3
FBXW7	3
CSF3R	3
PIK3R1	3
MAP3K1	3
CAMTA1	3
RABEP1	3

CCSER1	JAK2	OR52N1
CDH13	GSTM1	PARK2
CDKN2A	GSTM2	PDE4D
CDKN2B	LCE3B	PTEN
CSMD1	LPAR6	PTPRD
DLC1	MAP2K4	RBI
ELF2	MCPH1	RPII-145E5.5
ERICH1	MSRI	SGCZ
FBXO16	MTAP	SMAD4
FHIT	MTUS1	UGT2B15
FOCAD	NRGI	WWOX
		ZNF705G

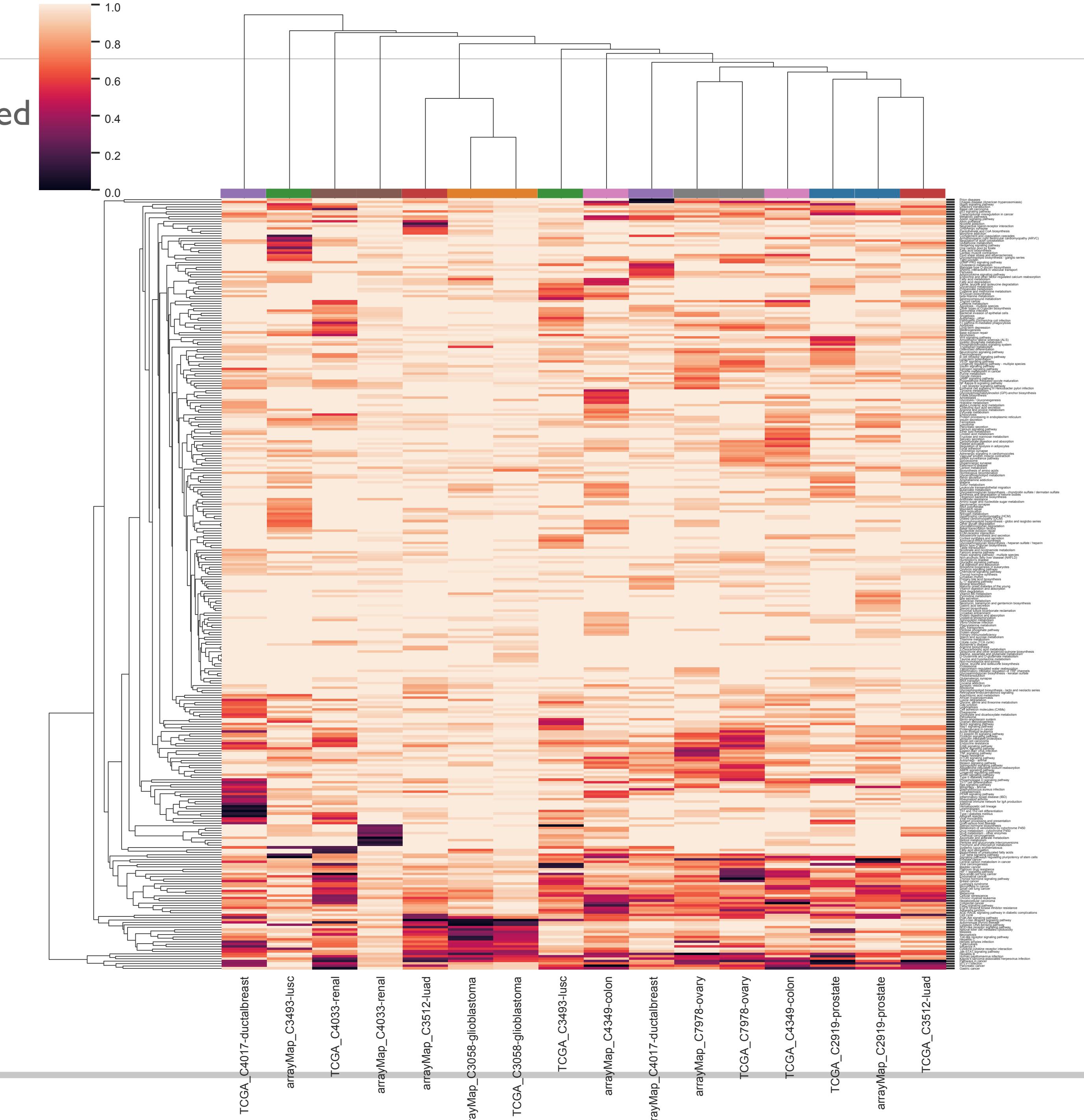
Mini Review

Common fragile sites, extremely large genes, neural development and cancer

David I Smith ^a Yu Zhu ^a, Sarah McAvoy ^a, Robert Kuhn ^b

KEGG Pathway Enrichment

Log P
standardised



Fisher exact test p value

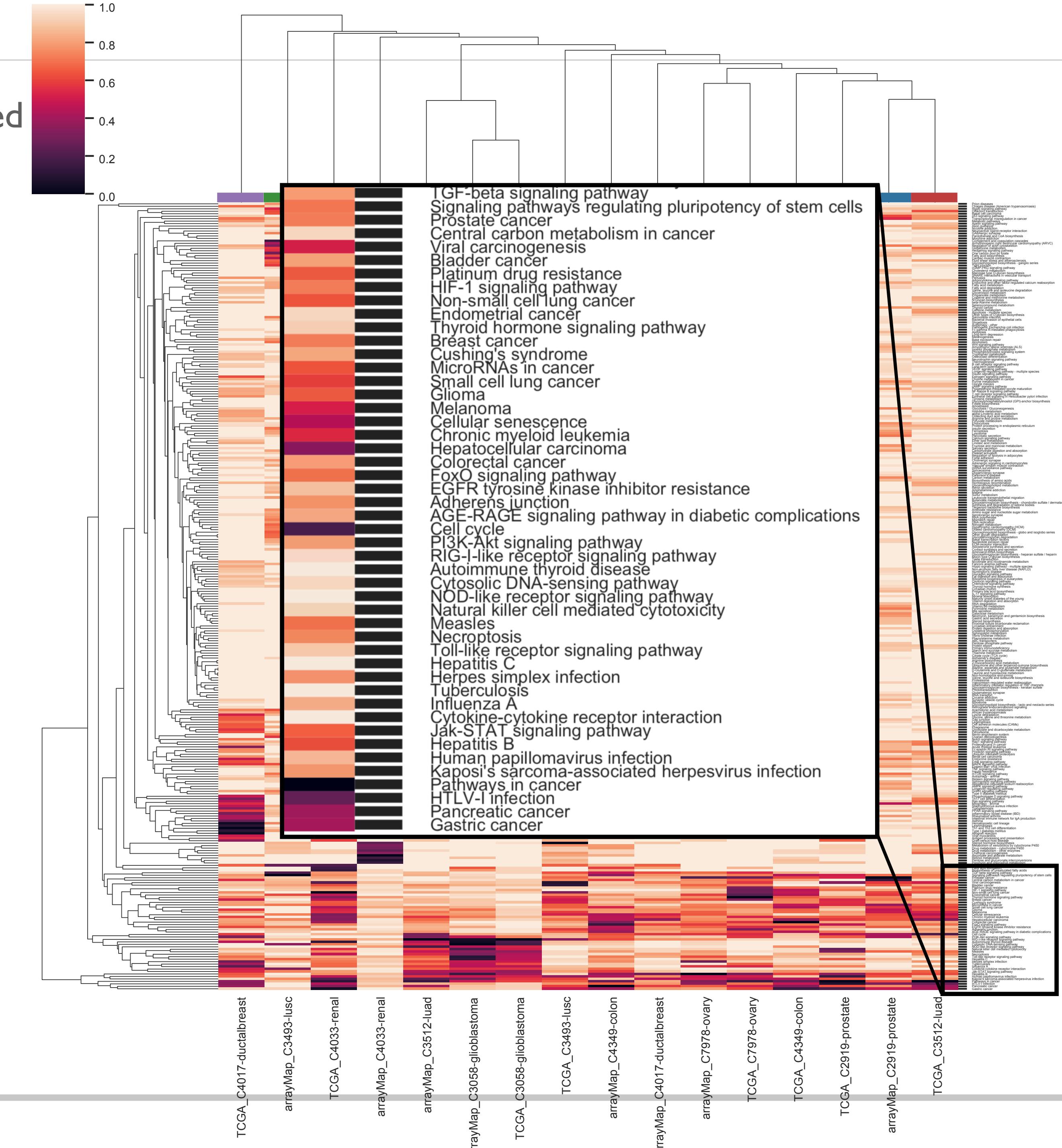
	In pathway	Not in pathway
Significant gene	20	40

	Whole genome
Significant gene	20
Whole genome	200
	19800

p-value < 2.2e-16

Cancer Pathway Enrichment

Log P
standardised

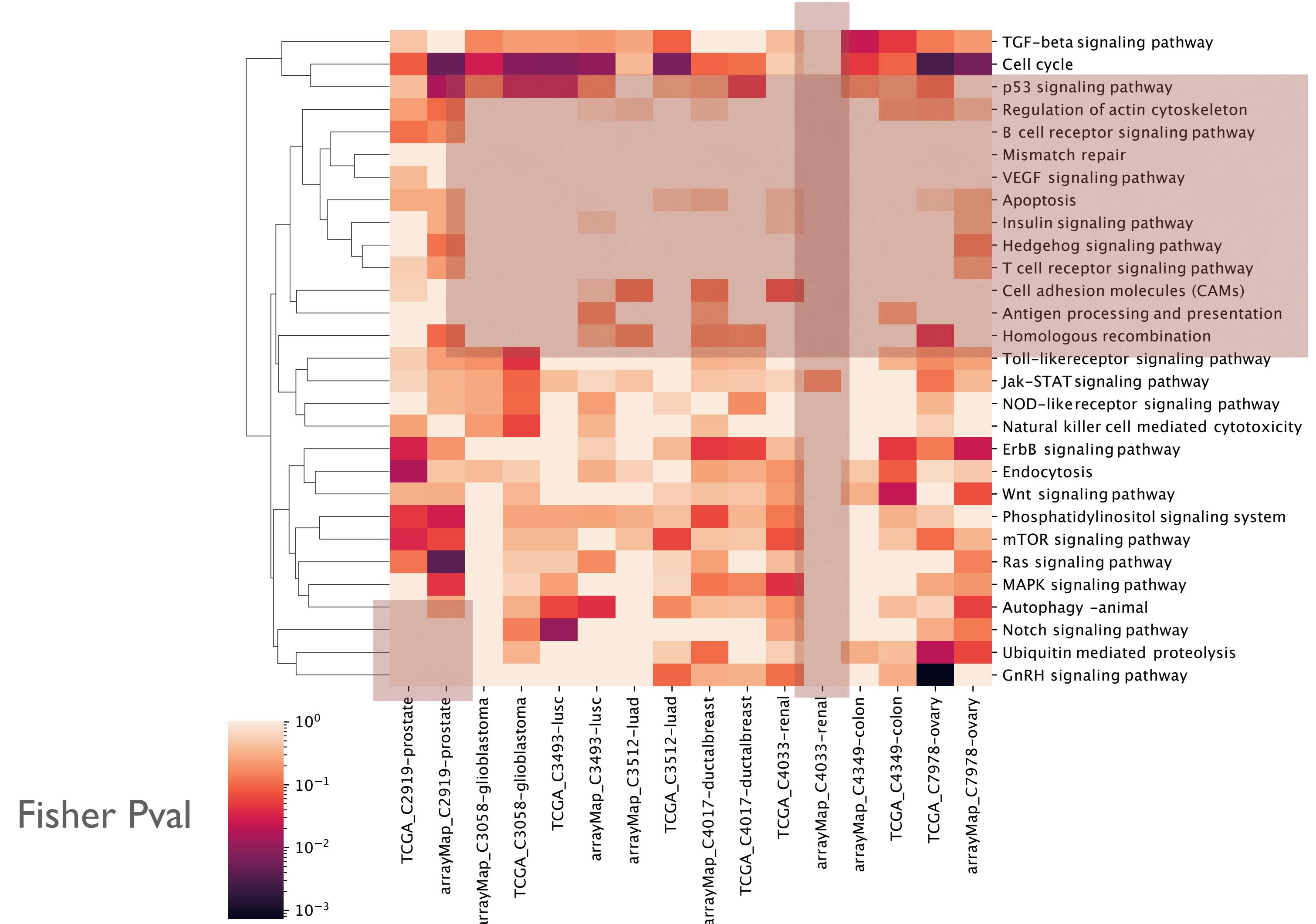


Fisher exact test p value

	In pathway	Not in pathway
Significant gene	20	40
Whole genome	200	19800

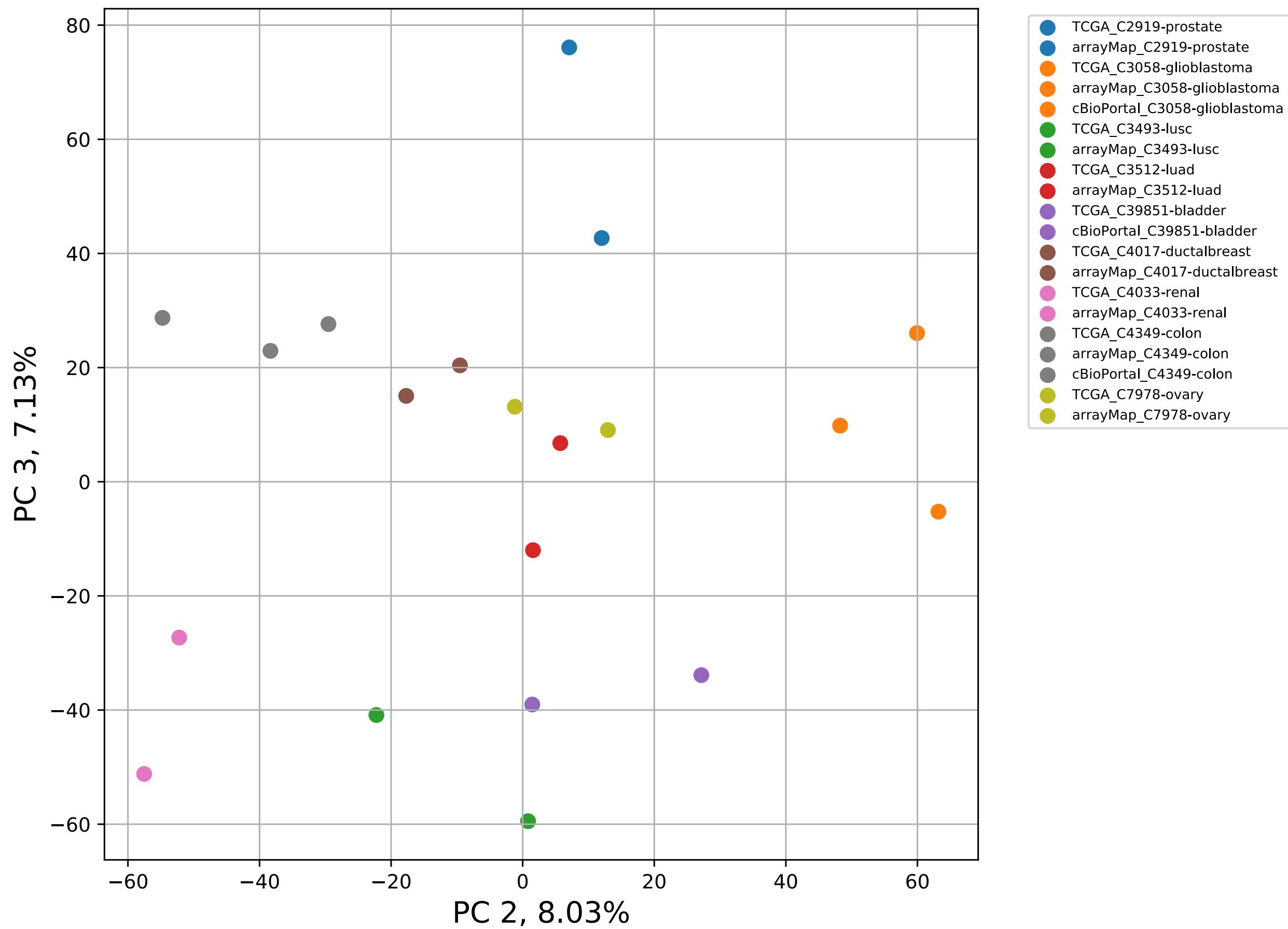
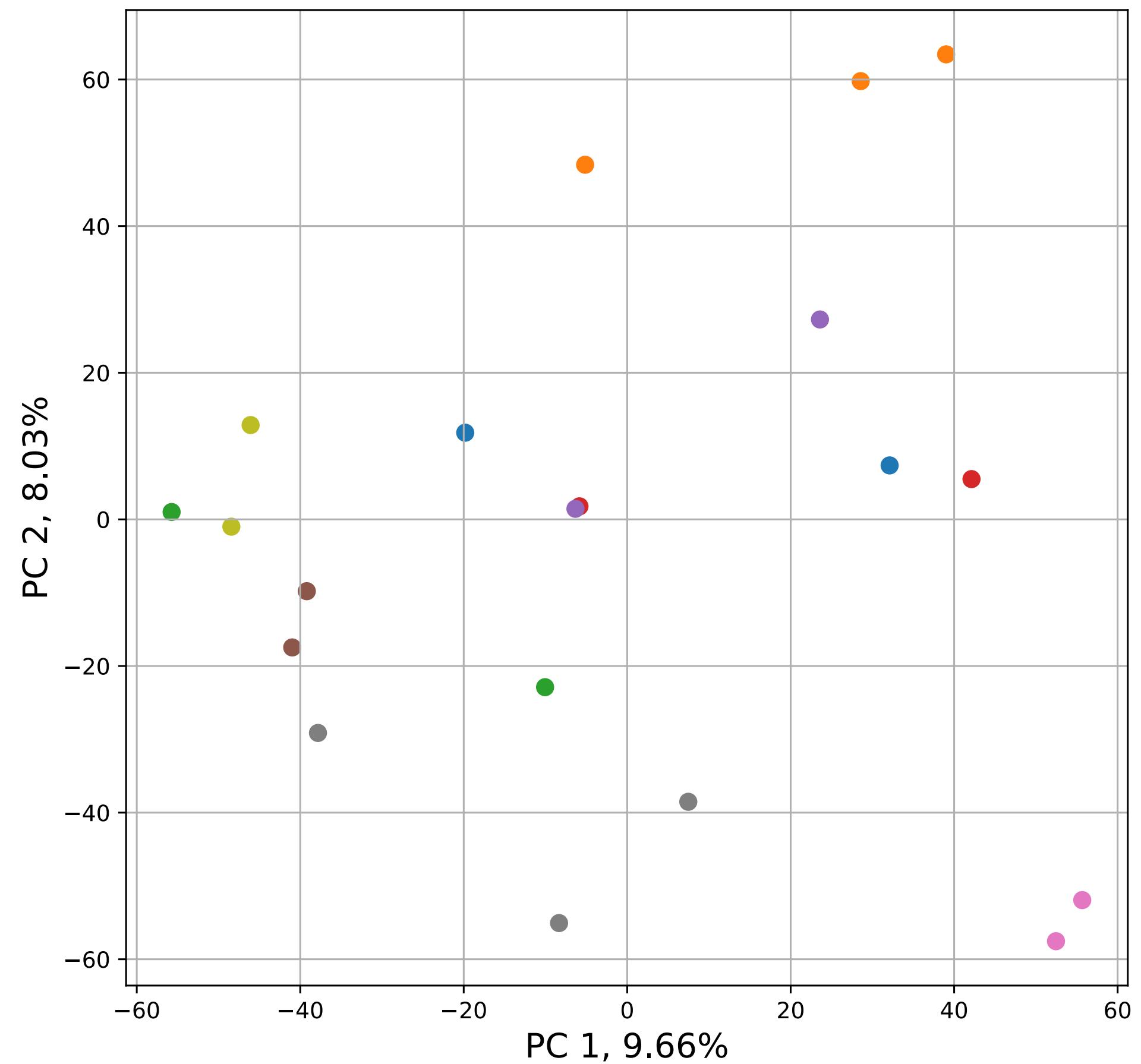
p-value < 2.2e-16

Pathways Related to Cancer Mechanisms



Vogelstein et, al. 2013

Clustering by Genome-Wide Significance



TCGA_C2919-prostate
arrayMap_C2919-prostate
TCGA_C3058-glioblastoma
arrayMap_C3058-glioblastoma
cBioPortal_C3058-glioblastoma
TCGA_C3493-lusc
arrayMap_C3493-lusc
TCGA_C3512-luad
arrayMap_C3512-luad
TCGA_C39851-bladder
arrayMap_C39851-bladder
TCGA_C4017-ductalbreast
arrayMap_C4017-ductalbreast
TCGA_C4033-renal
arrayMap_C4033-renal
TCGA_C4349-colon
arrayMap_C4349-colon
cBioPortal_C4349-colon
TCGA_C7978-ovary
arrayMap_C7978-ovary

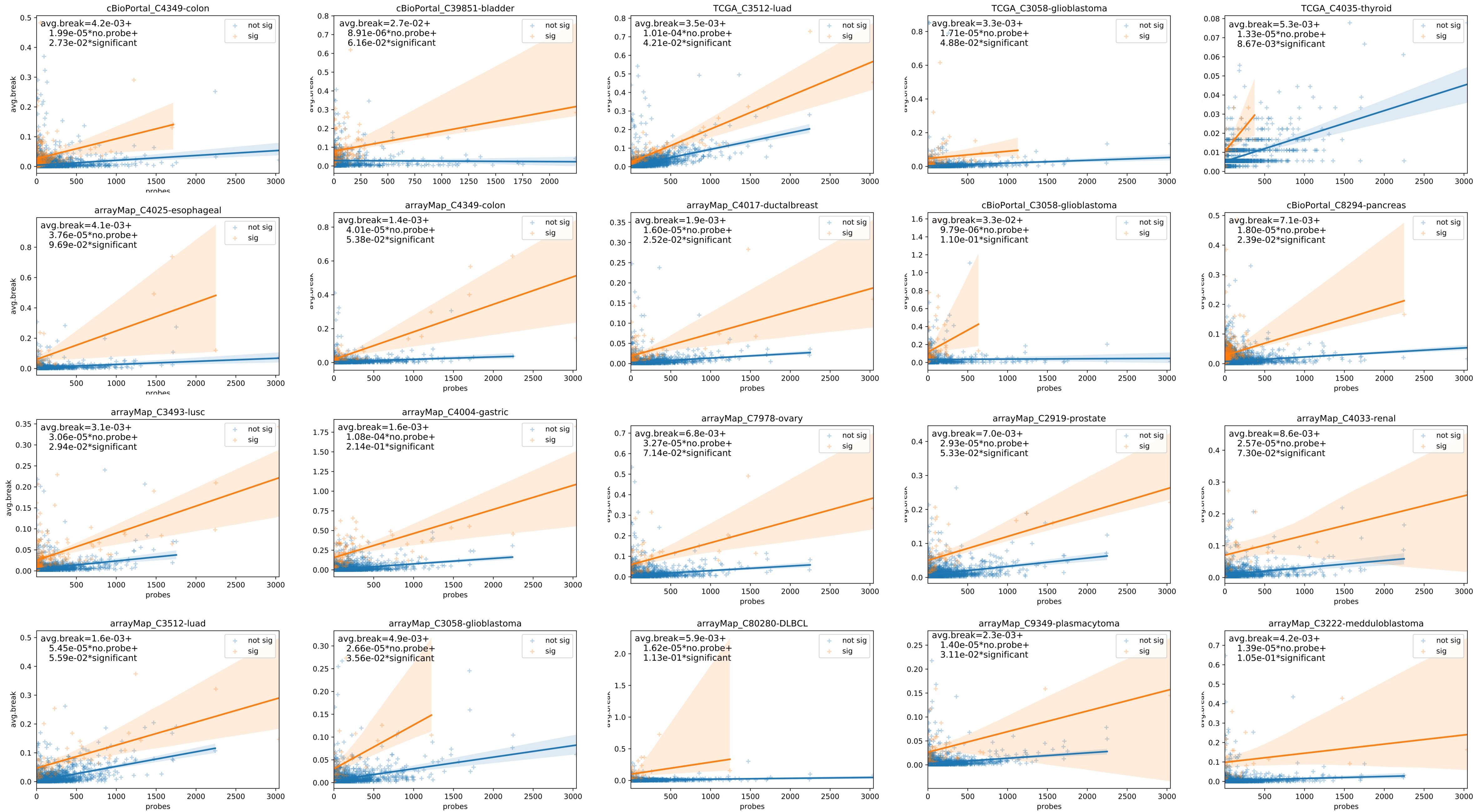
Conclusion

- Translate genome-wide CNV landscape to gene-based significance landscape
 - Re-discover known cancer-promoting genes and pathways
 - Distinguish between cancer types
-

Thank You for Your Attention!

Acknowledgement

Prof. Michael Baudis and Baudis group



331Mb gene rich
1523Mb gene poor

