

Integrating Diverse Transcriptomic Alterations to Identify Cancer-Relevant Genes

Natalie Davidson

Joint work with: PanCancer Analysis of Whole Genomes and Transcriptomes Working Group (PCAWG-3), PCAWG Consortium, Kjong-Van Lehmann^{1,2}, André Kahles^{1,2}, Alvis Brazma³, Angela

N. Brooks⁴, Claudia Calabrese³, Nuno A. Fonseca³, Jonathan Goke⁵, Roland F Schwarz^{3,6}, Gunnar Rätsch^{1,2}, Zemin Zhang^{7,8}

¹ETH Zürich, Computer Science Dept, Universitätstrasse 6, 8092 Zürich, Switzerland;

²Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York 10065, USA;

³European Molecular Biology Laboratory - European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK;

⁴University of California, Santa Cruz, CA 95060; ⁴Baylor

College of Medicine, Houston, TX, USA; ⁵Duke-NUS Graduate

Medical School, 8 College Road, Singapore 169857, Singapore; ⁶Berlin

Institute for Medical Systems Biology, Max Delbrück Center for Molecular

Medicine, Berlin, Germany; ⁷Peking-Tsinghua Center for Life Sciences, Peking

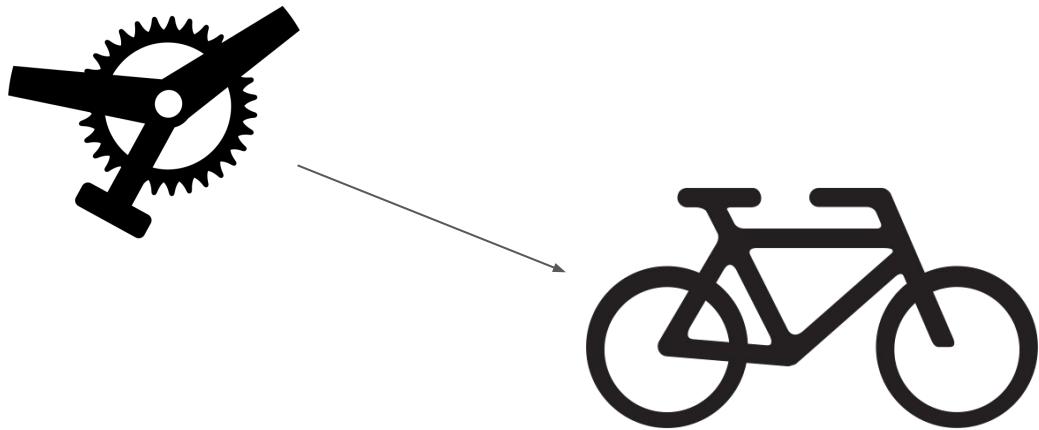
University, Beijing, 100871, China; ⁸Genome Institute of Singapore, 60 Biopolis Street,

Genome #02-01, Singapore 138672, Singapore;

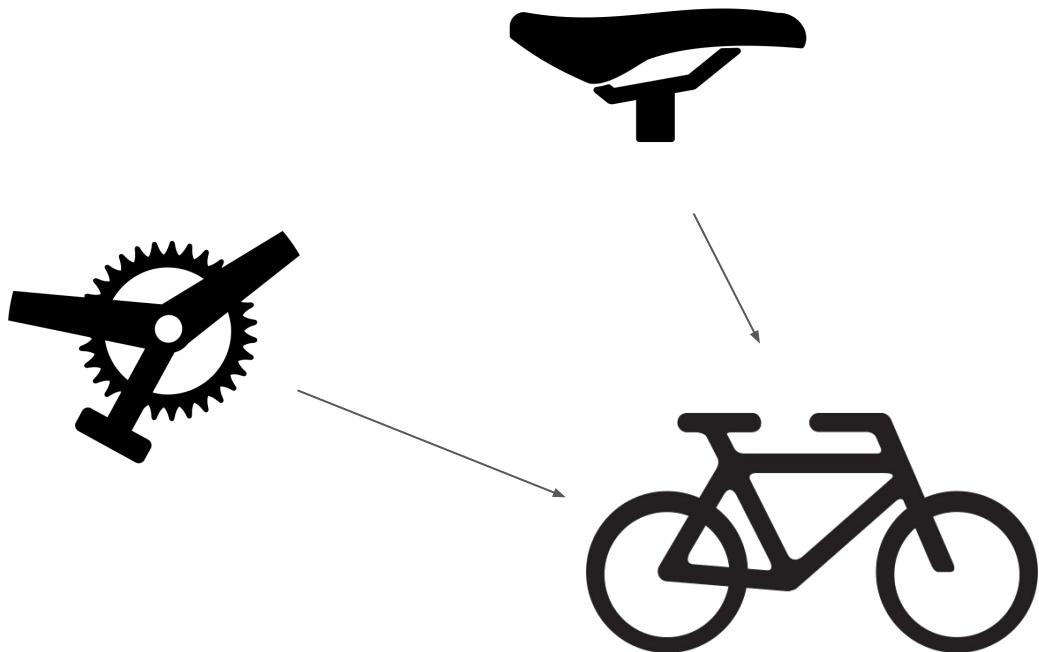
Many ways to break a bike



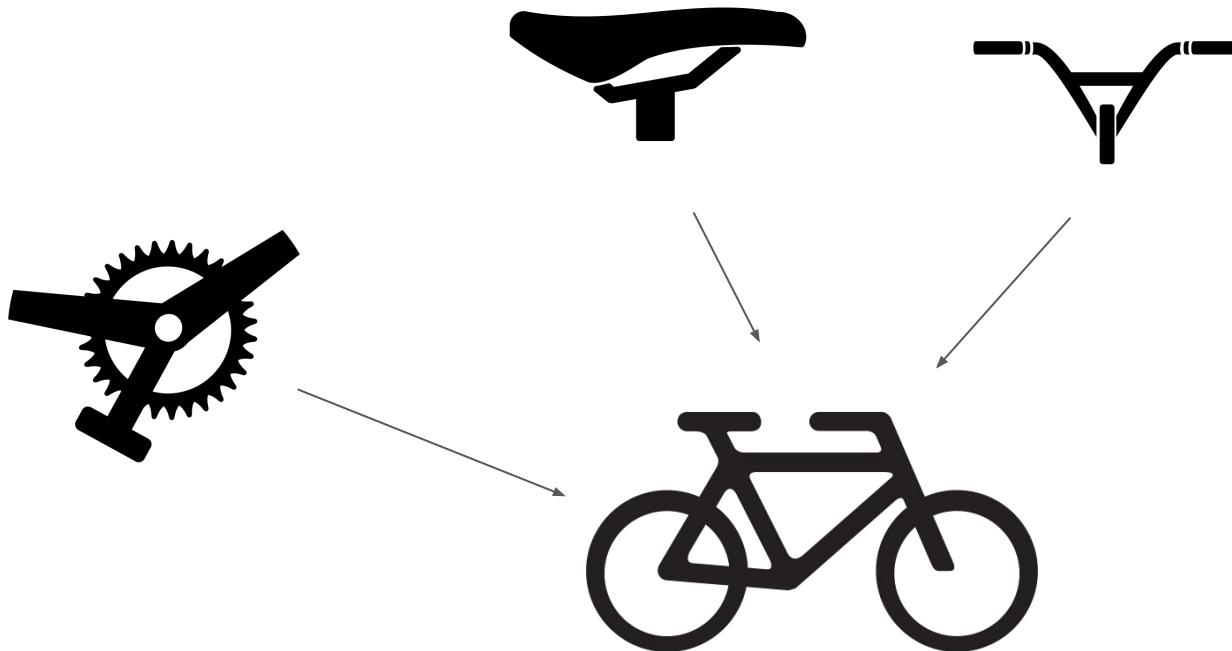
Many ways to break a bike



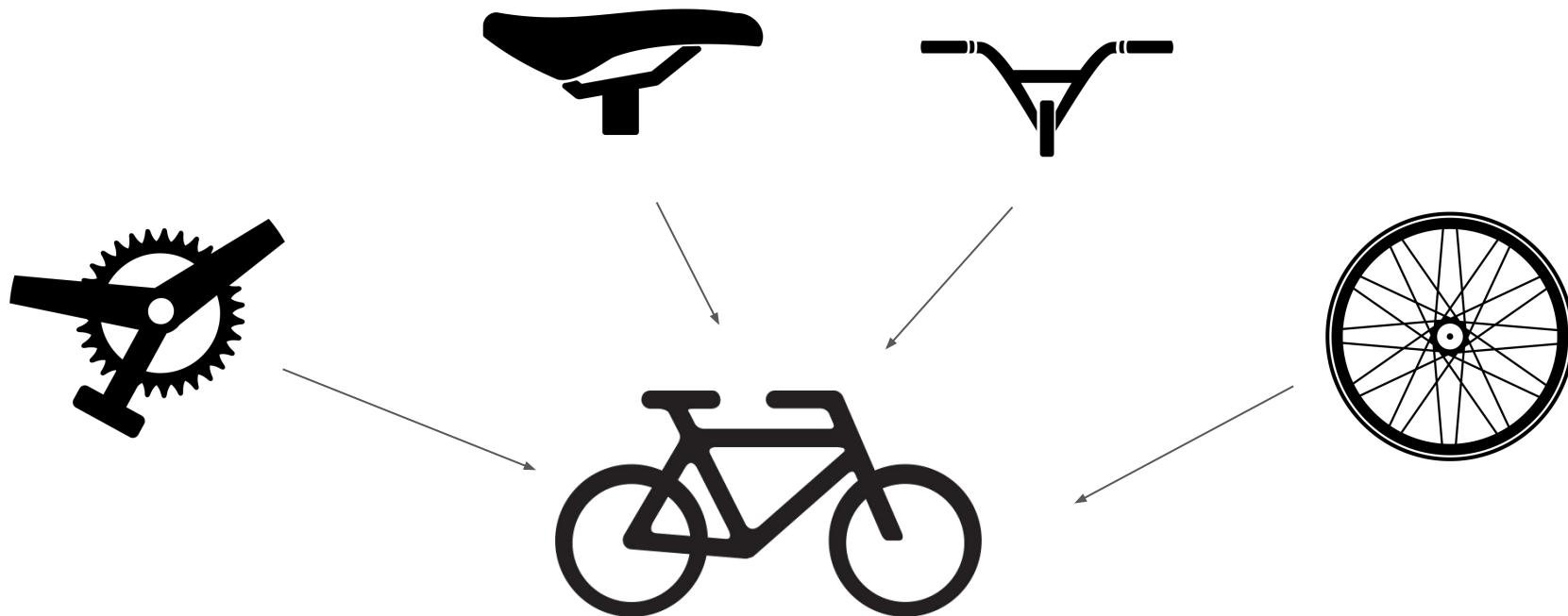
Many ways to break a bike



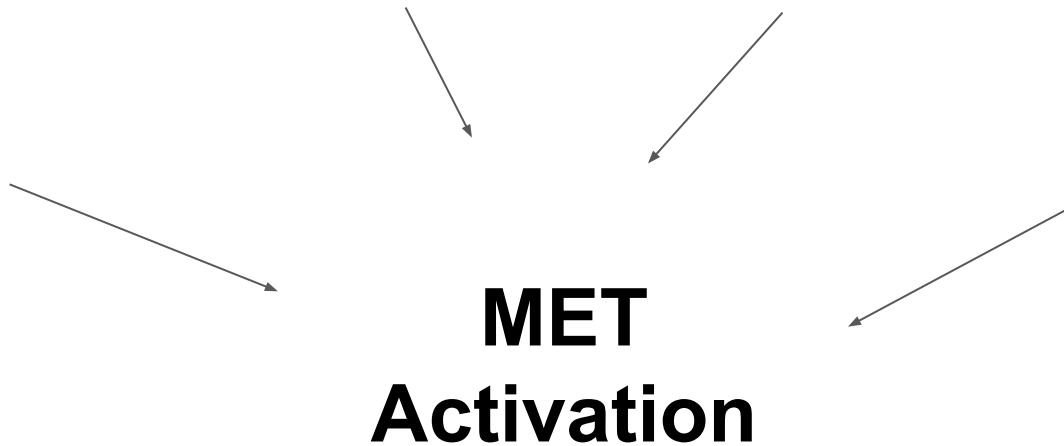
Many ways to break a bike



Many ways to break a bike



Many ways to induce oncogenic event



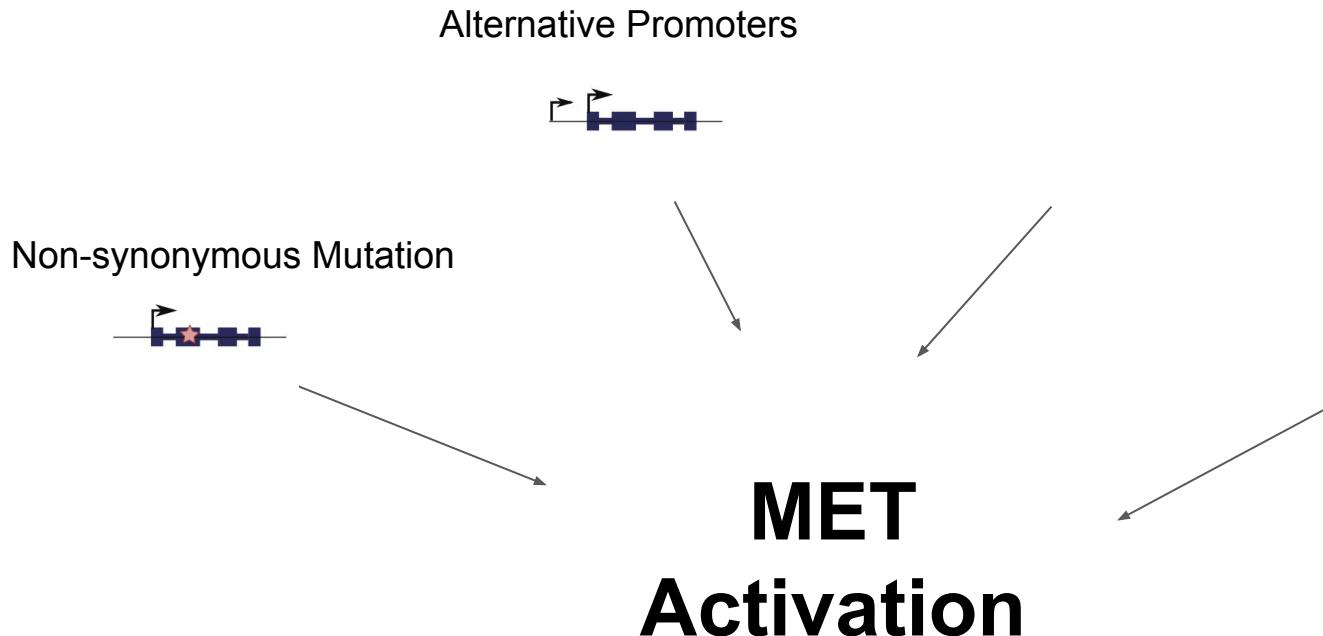
Many ways to induce oncogenic event

Non-synonymous Mutation



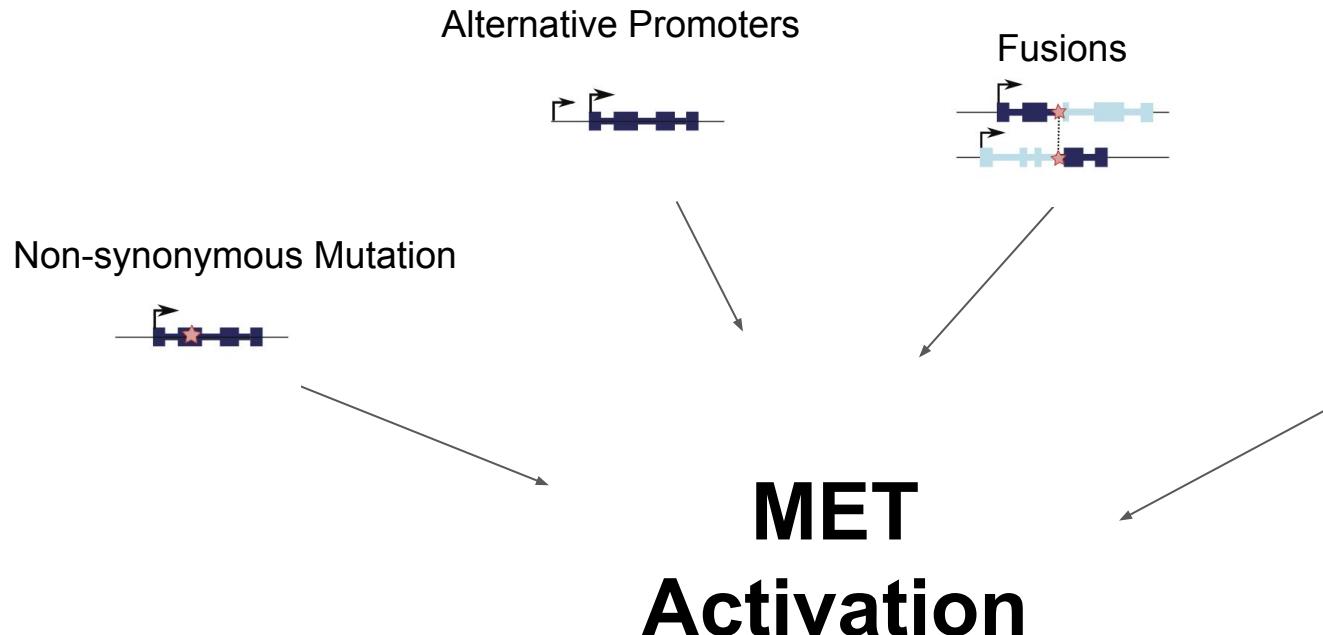
**MET
Activation**

Many ways to induce oncogenic event

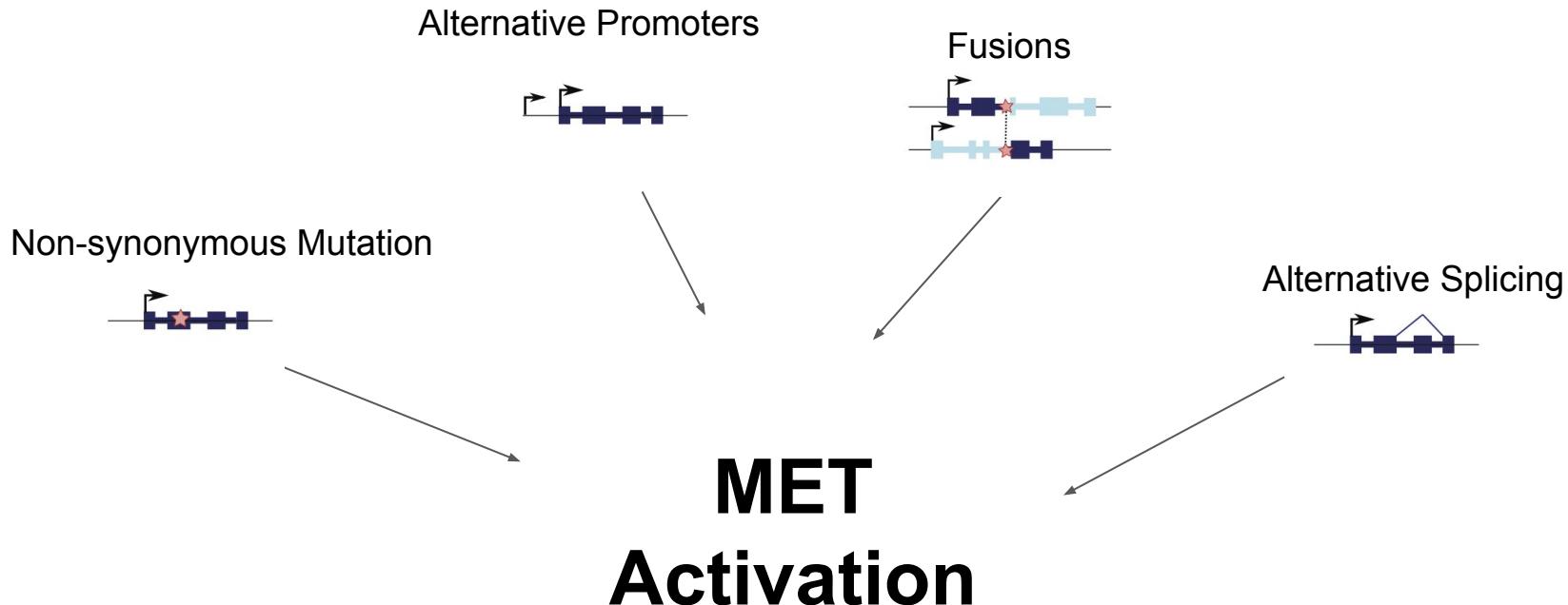


Stransky et al. Nat Comm, 2014
Muratani et al. Nat Comm, 2014
TCGA, Nature, 2014
TCGA, NEJM, 2016

Many ways to induce oncogenic event



Many ways to induce oncogenic event



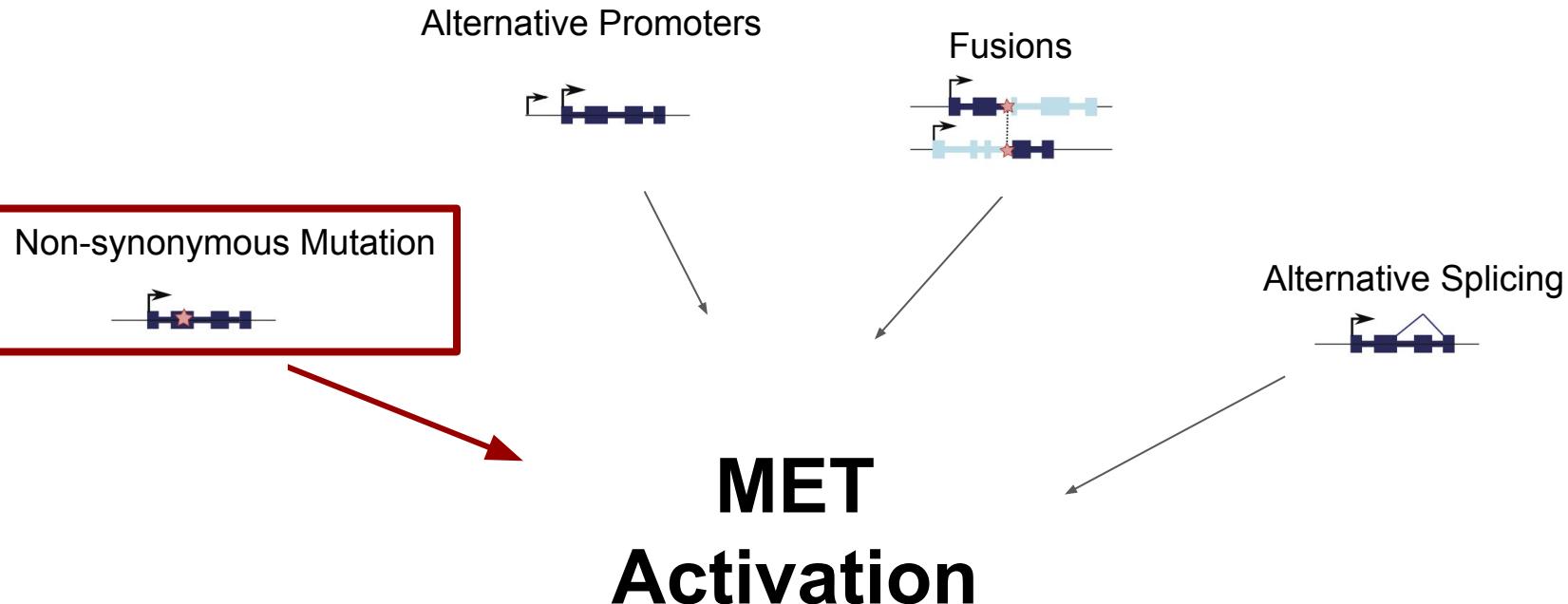
Stransky et al. Nat Comm, 2014

Muratani et al. Nat Comm, 2014

TCGA, Nature, 2014

TCGA, NEJM, 2016

Many ways to induce oncogenic event



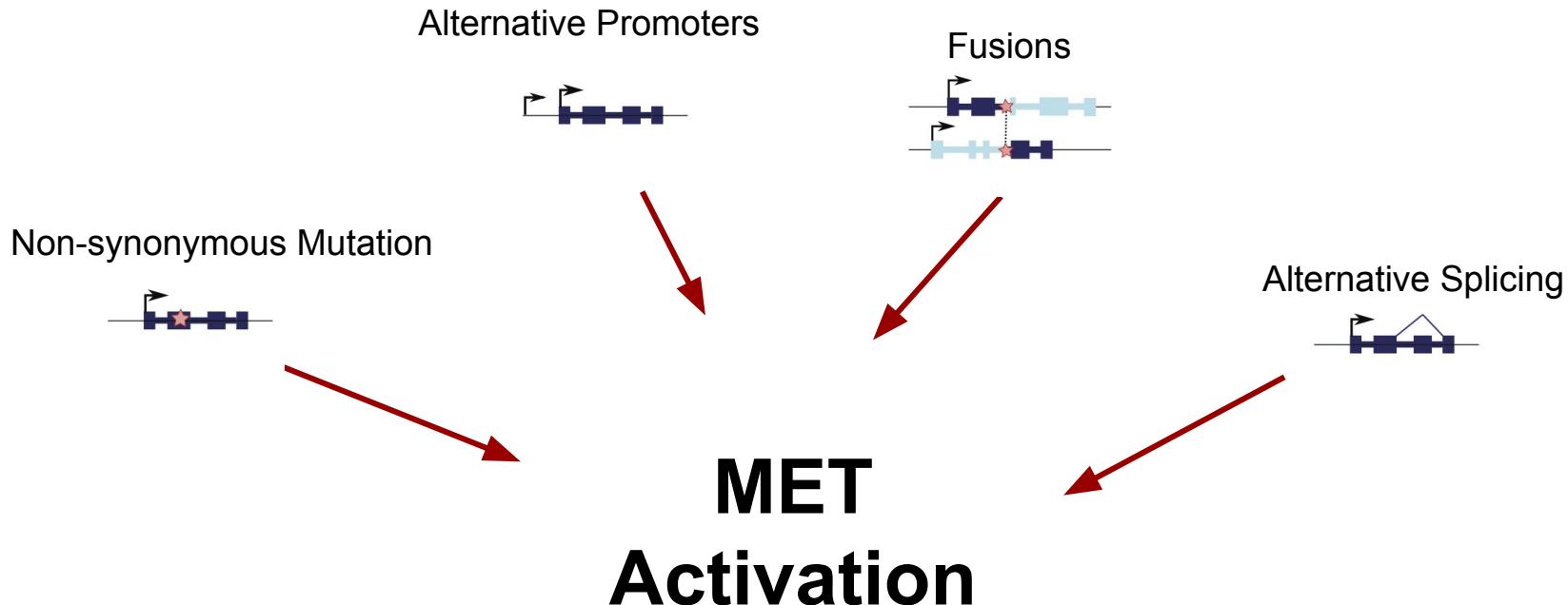
Stransky et al. Nat Comm, 2014

Muratani et al. Nat Comm, 2014

TCGA, Nature, 2014

TCGA, NEJM, 2016

Many ways to induce oncogenic event



ICGC/PCAWG-3 overview



International
Cancer Genome
Consortium



Grey = Collaboration

Overview of PCAWG-3 Samples

PCAWG RNA-Seq FASTQ

TopHat2 (v2.0.12)
processed at
EMBL-EBI

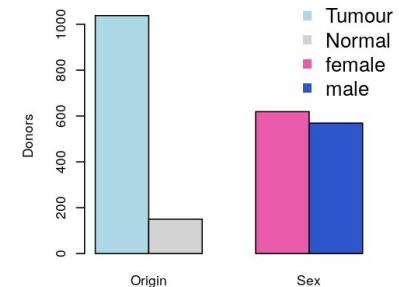
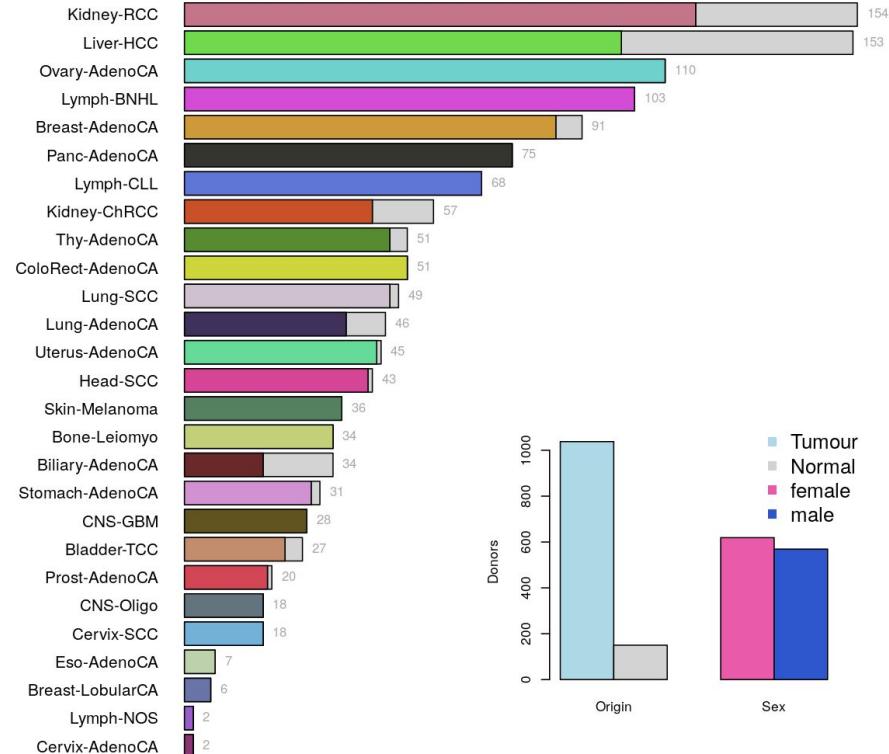
STAR 2 pass (v2.4.0i)
processed at
MSKCC

HTSeq

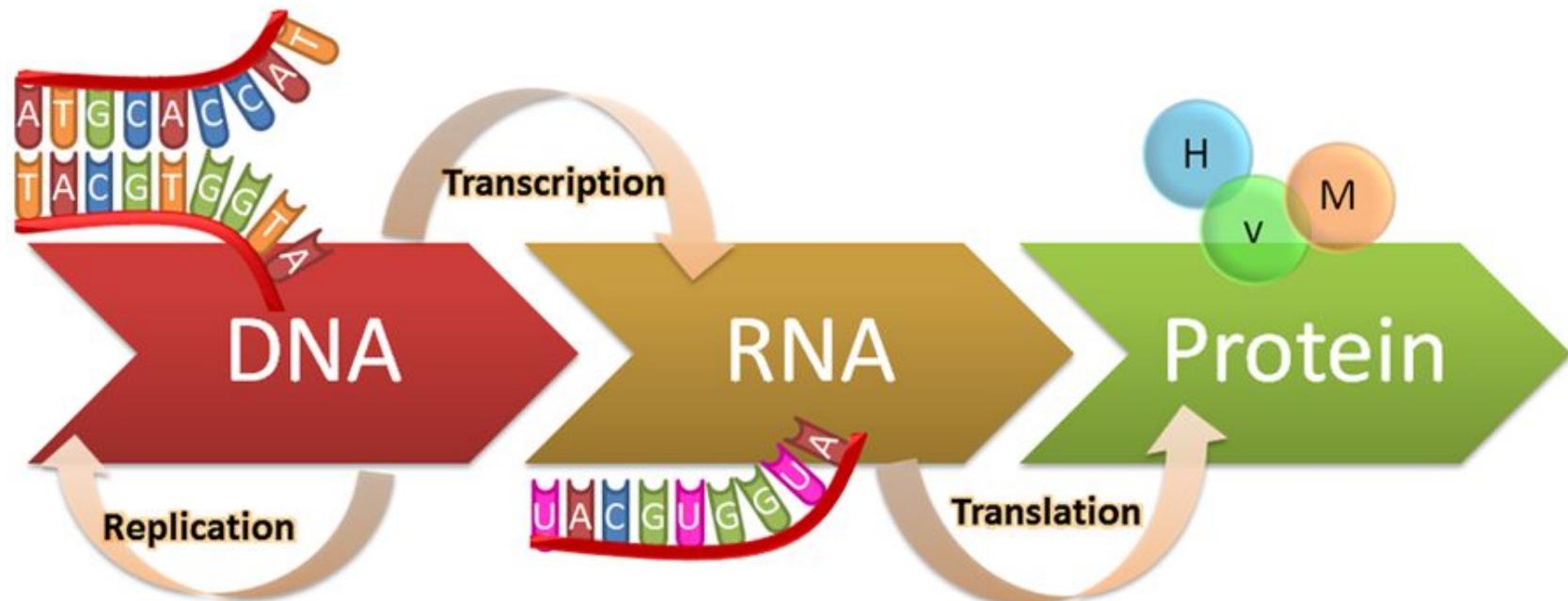
HTSeq

Combined
normalized
FPKM-UQ

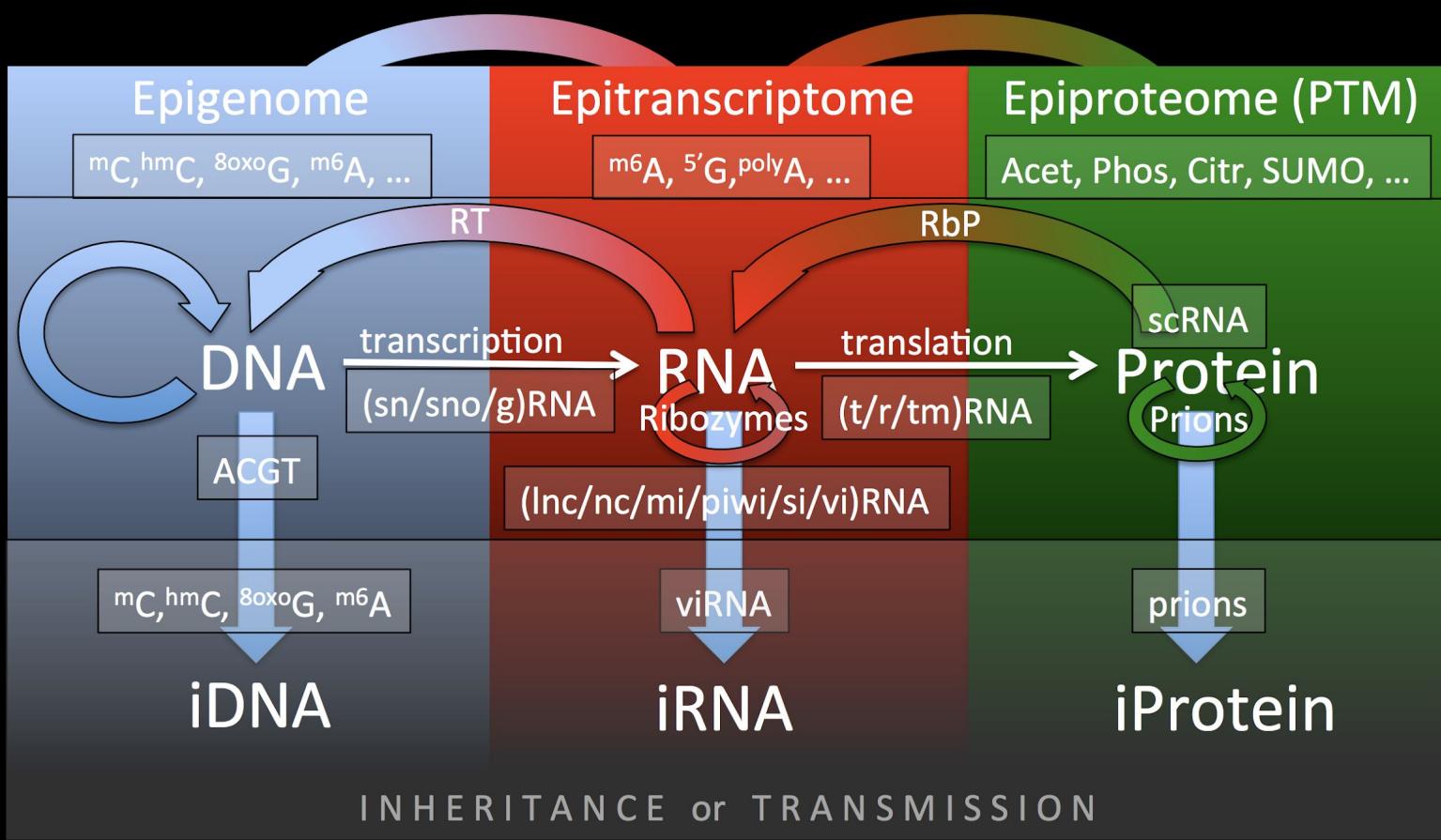
PCAWG RNA-seq data
Number of samples (tumour and normal) per histological type



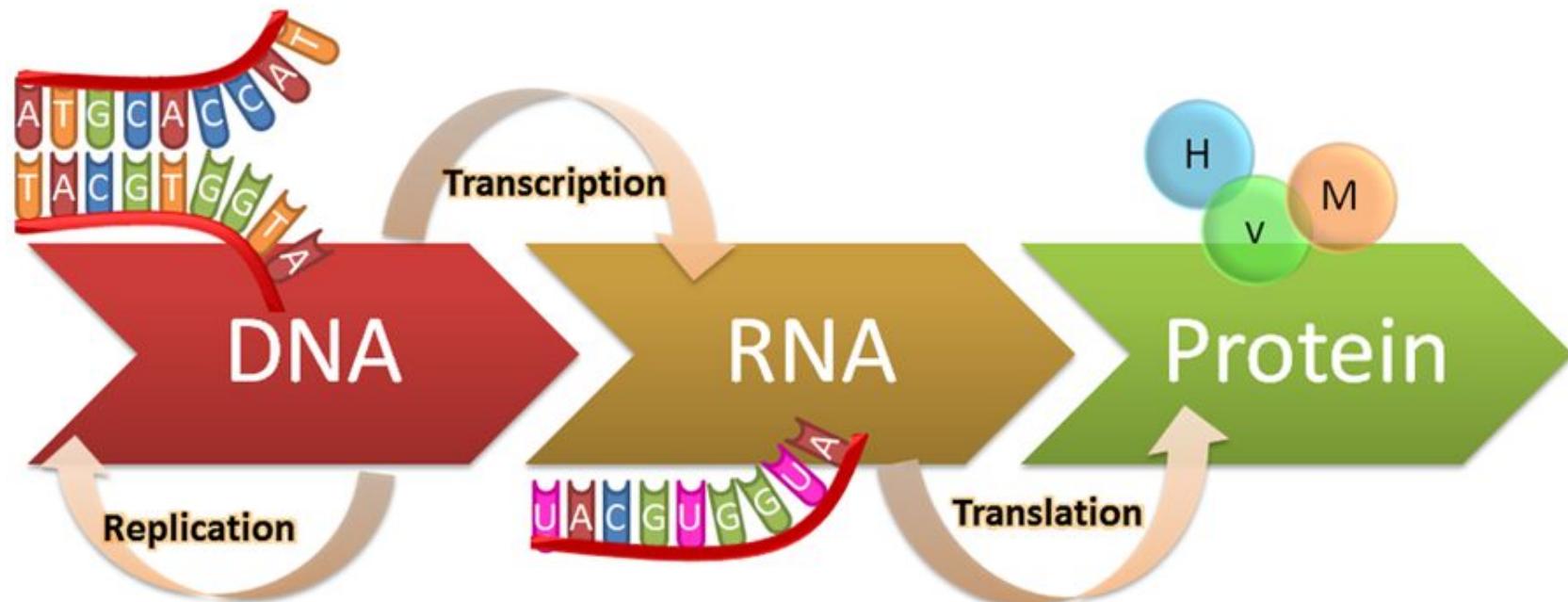
The Central Dogma



The Central Dogma



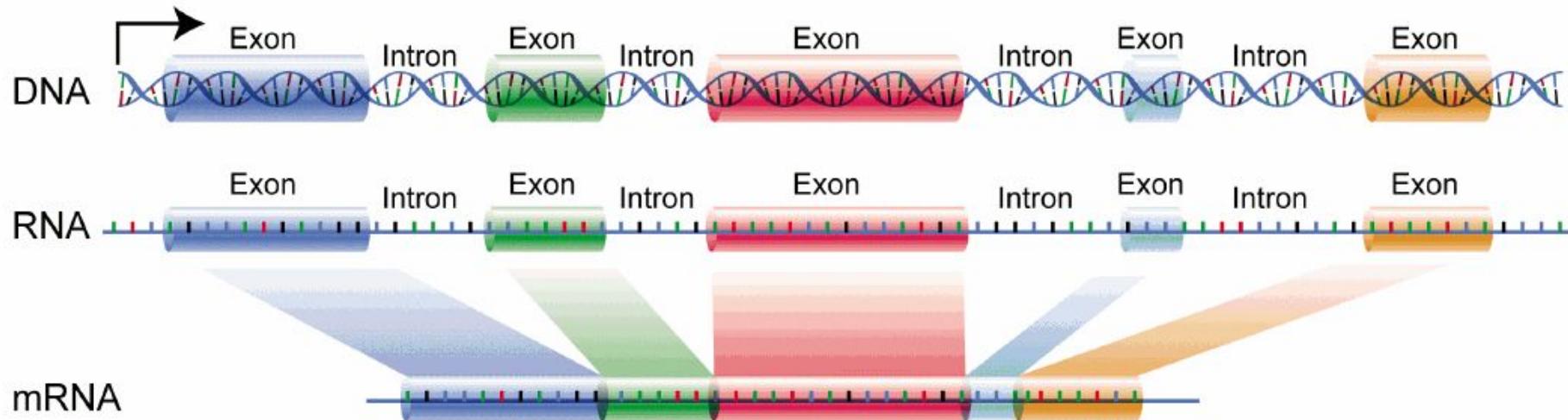
The Central Dogma



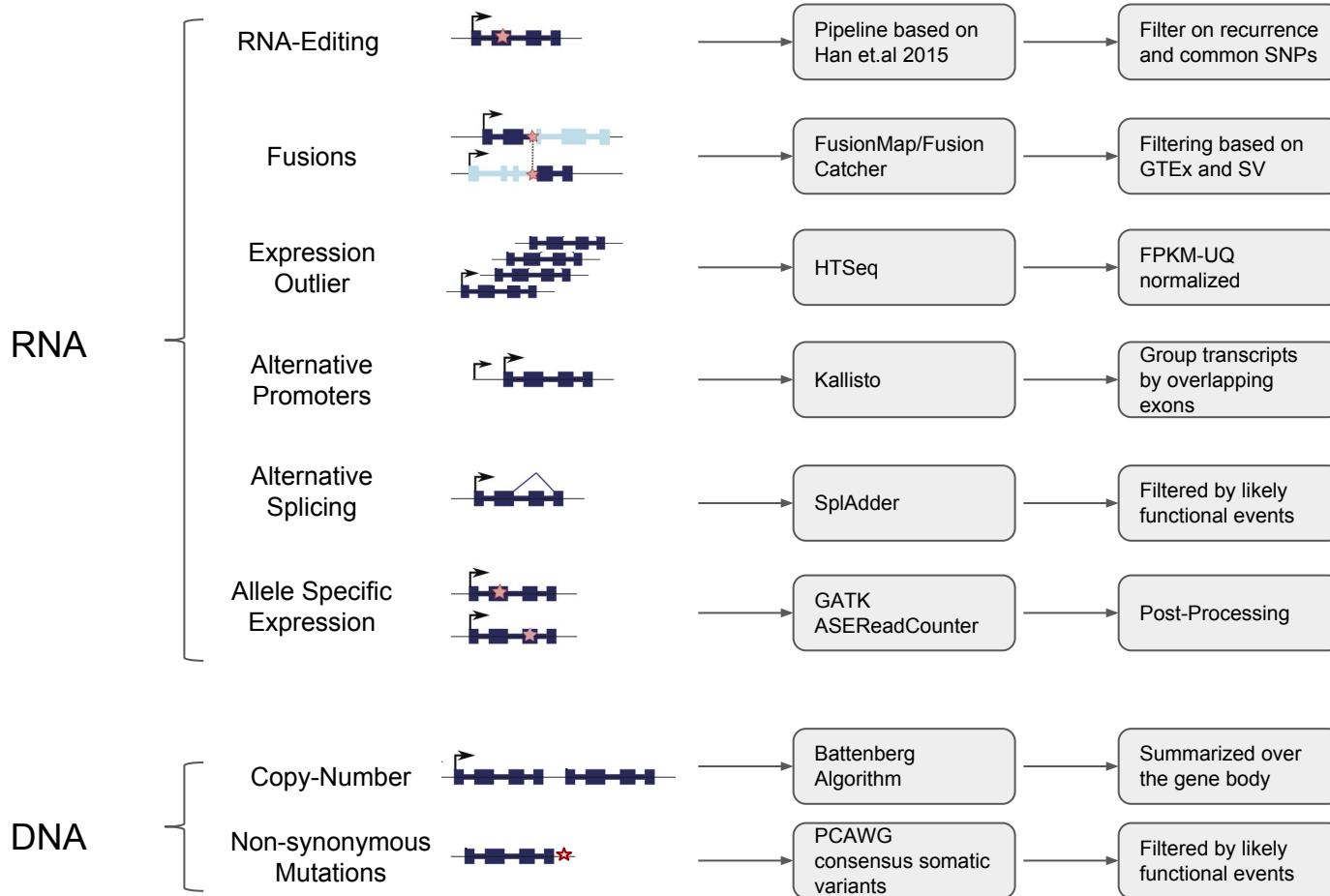
<http://genius.com/Biology-genius-the-central-dogma-annotated>

https://commons.wikimedia.org/wiki/File:DNA_exons_introns.gif

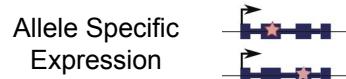
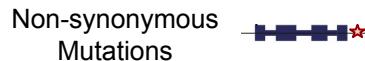
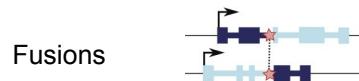
The Central Dogma



Overview of PCAWG Data used



Summarize alterations for a gene

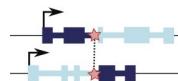


Summarize alterations for a gene

RNA-Editing



Fusions



Non-synonymous
Mutations



1, If an event occurs in the gene
0, else

Copy-Number



Expression
Outlier



Alternative
Promoters



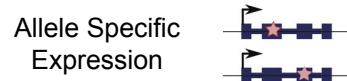
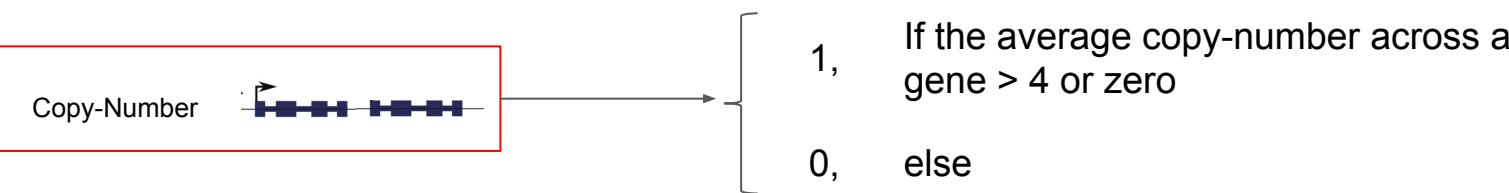
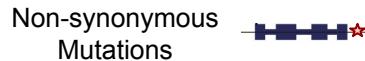
Alternative
Splicing



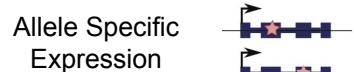
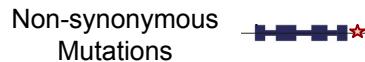
Allele Specific
Expression



Summarize alterations for a gene



Summarize alterations for a gene

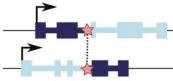
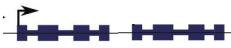


1,

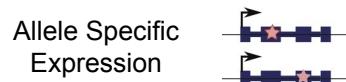
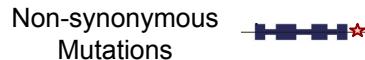
0,
else

If it is in the bottom or top percentile
within each cancer type.

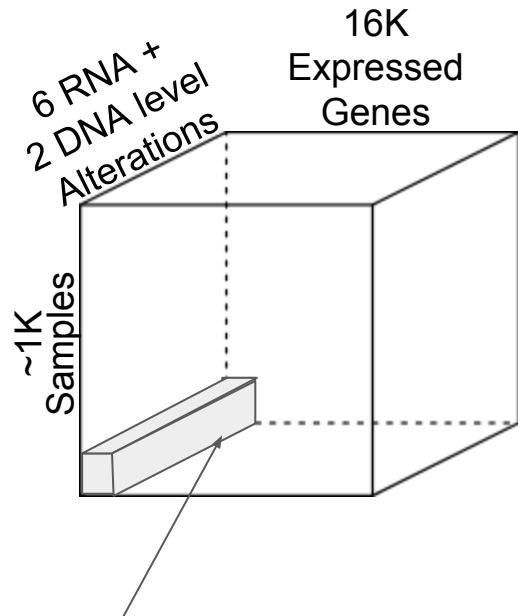
Summarize alterations for a gene

RNA-Editing		
Fusions		
Non-synonymous Mutations		
Copy-Number		
Expression Outlier		
Alternative Promoters		
Alternative Splicing		
Allele Specific Expression		

Summarize alterations for a gene

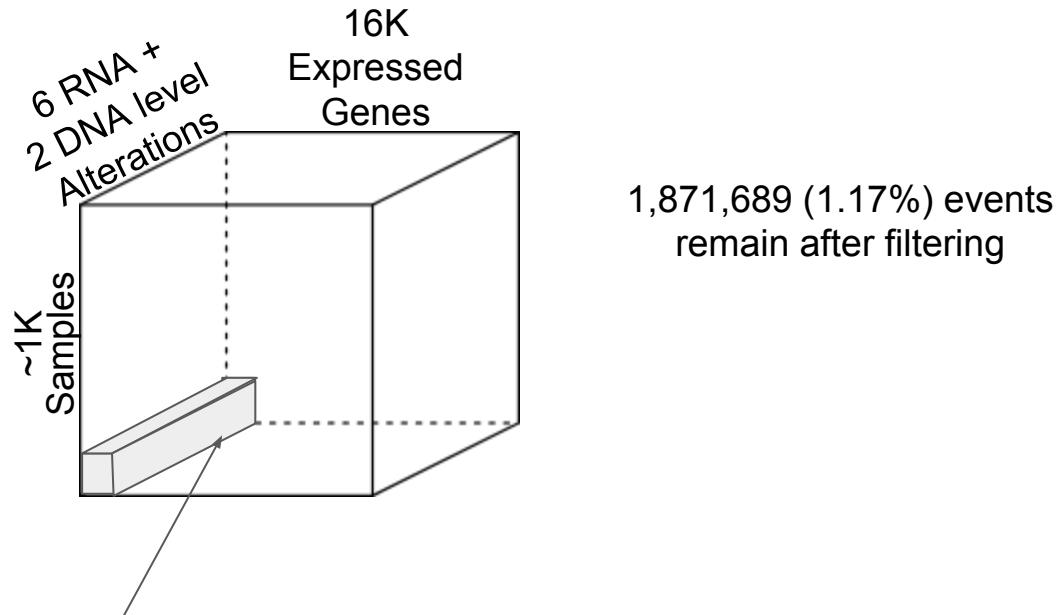


Summarize alterations across genes and samples



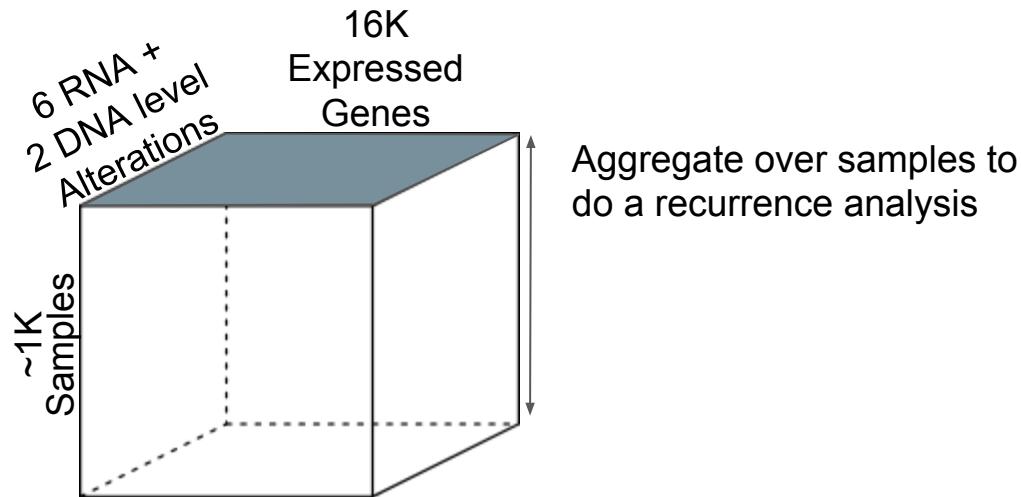
Binary value for each triplet (sample, gene, alteration).

Summarize alterations across genes and samples

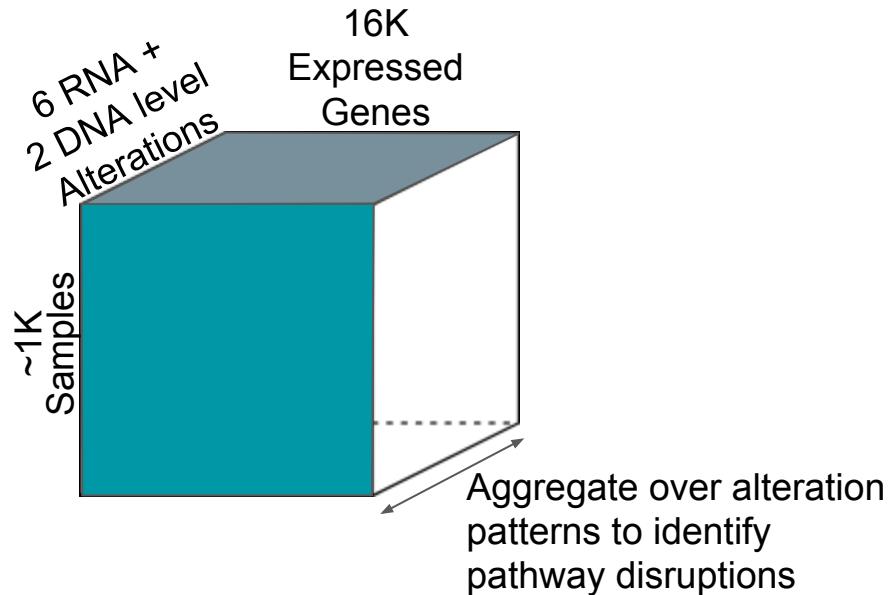


Binary value for each triplet (sample, gene, alteration).

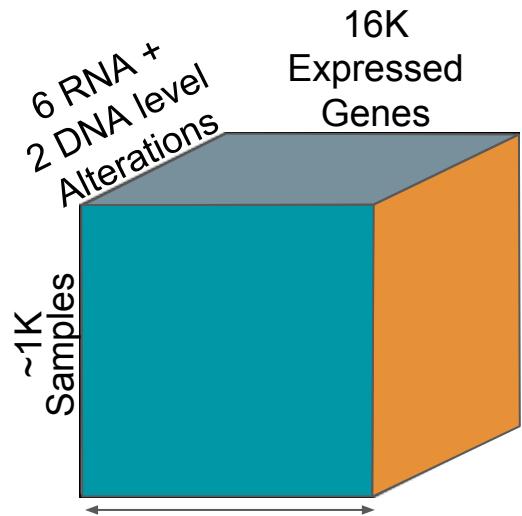
Summarize alterations across genes and samples



Summarize alterations across genes and samples

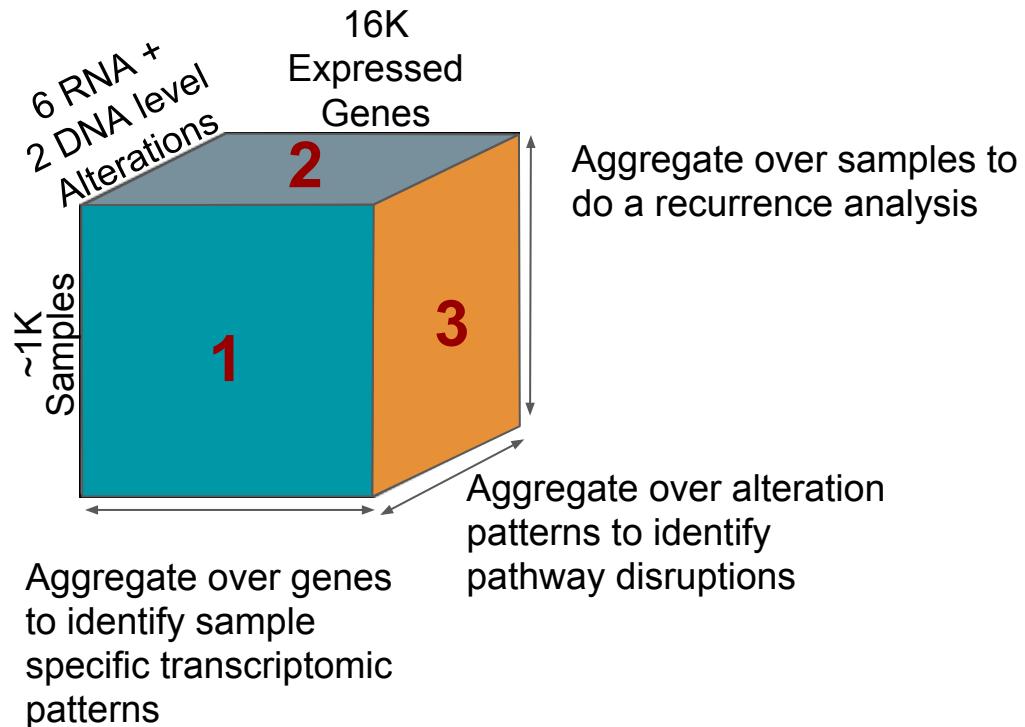


Combining the alteration types - gene-centric view



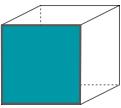
Aggregate over genes
to identify sample
specific transcriptomic
patterns

Summarize alterations across genes and samples

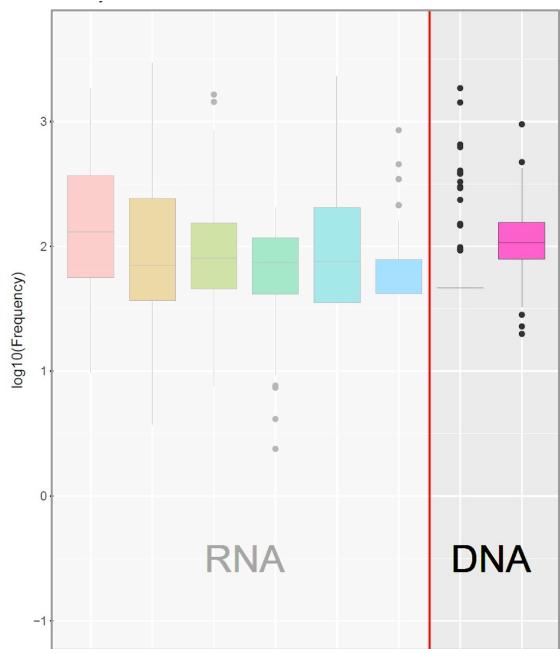


Compare and contrast samples

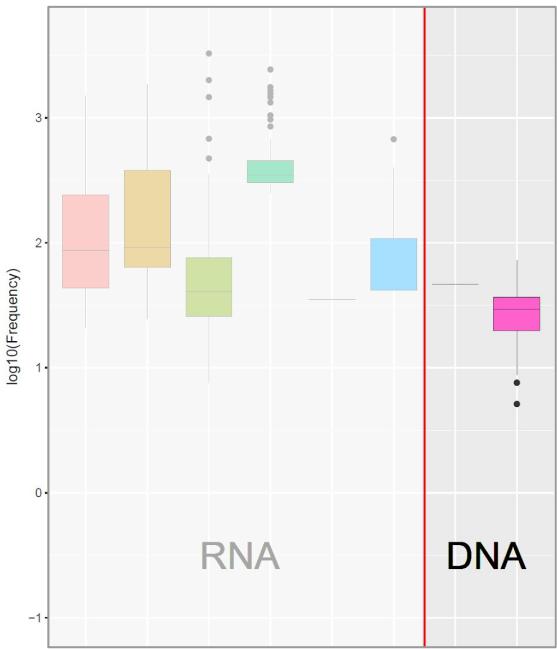
→Cancer Type



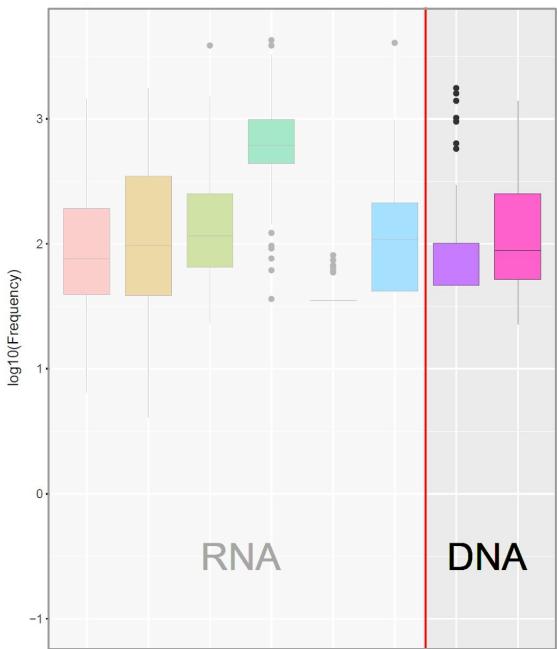
Kidney-RCC



Lymph-CLL

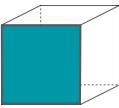


Lymph-BNHL

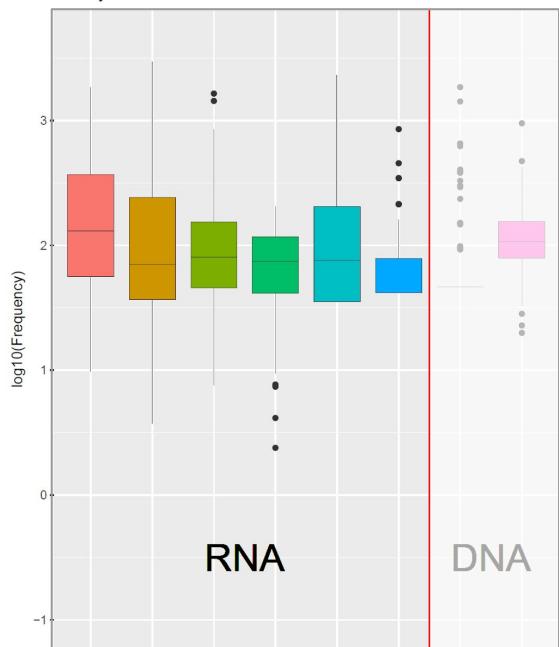


Compare and contrast samples

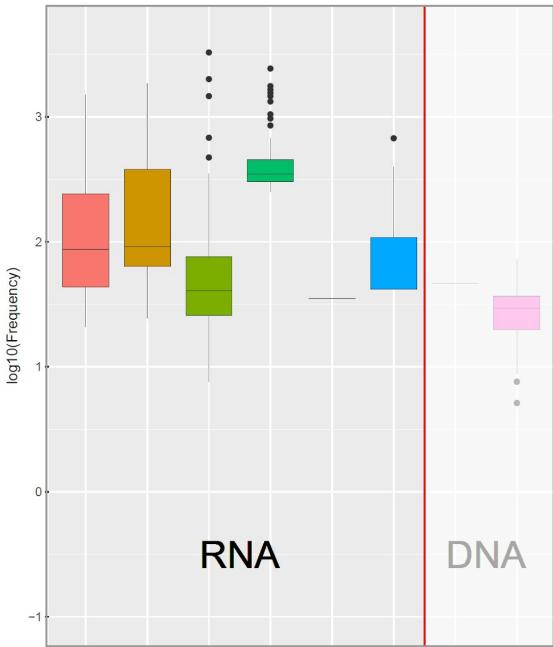
→Cancer Type



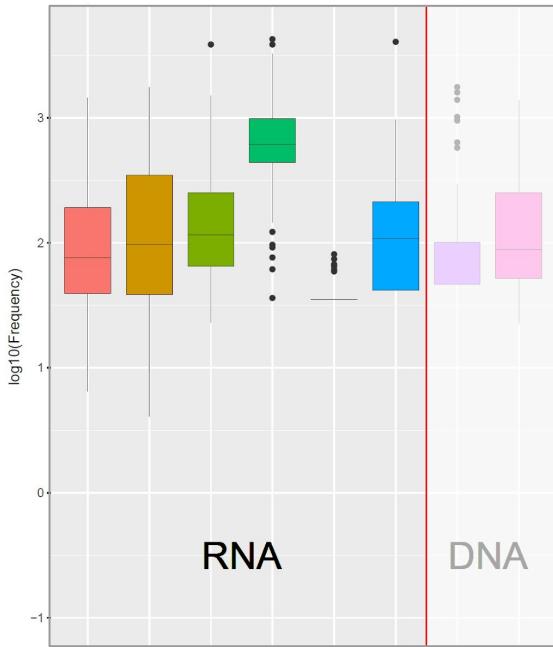
Kidney-RCC



Lymph-CLL

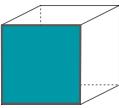


Lymph-BNHL

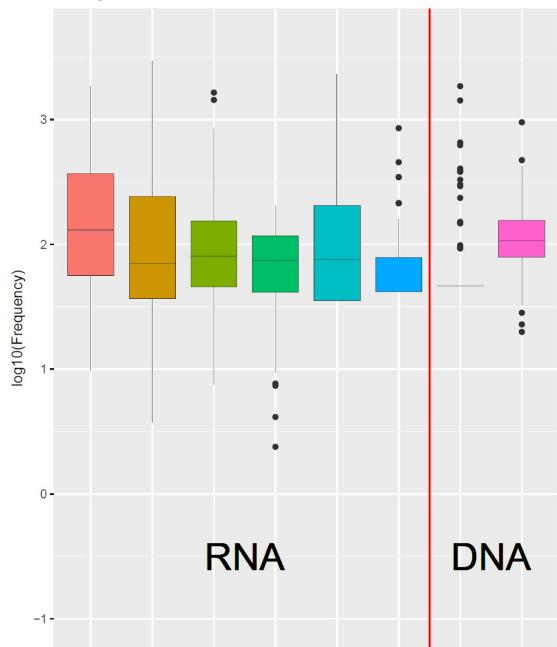


Compare and contrast samples

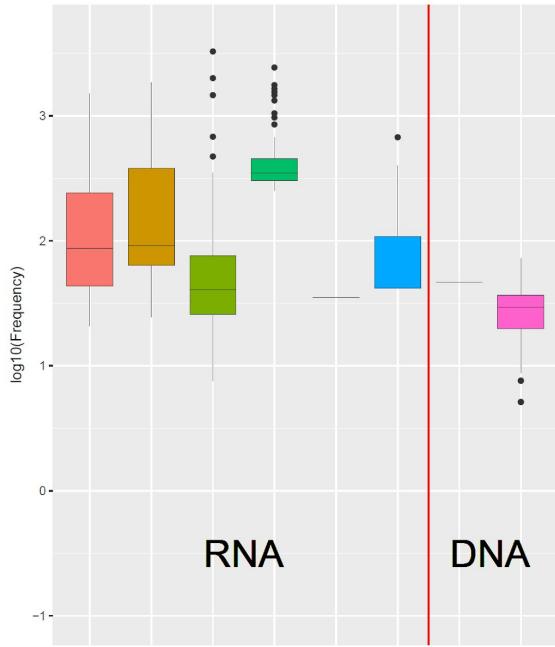
→Cancer Type



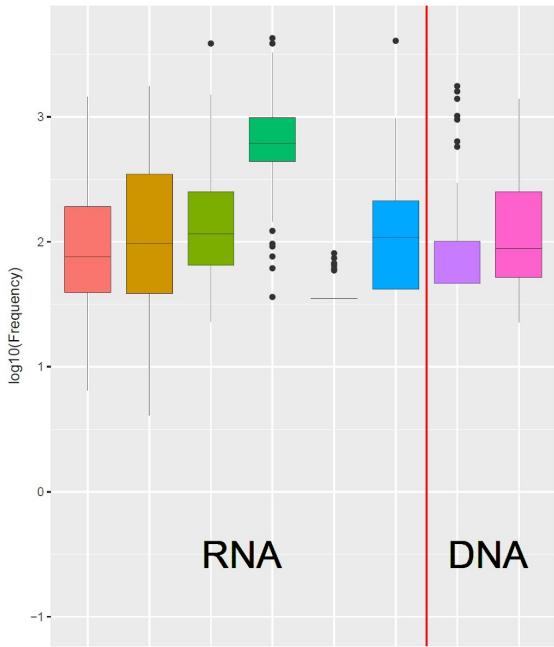
Kidney-RCC

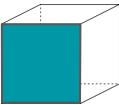


Lymph-CLL



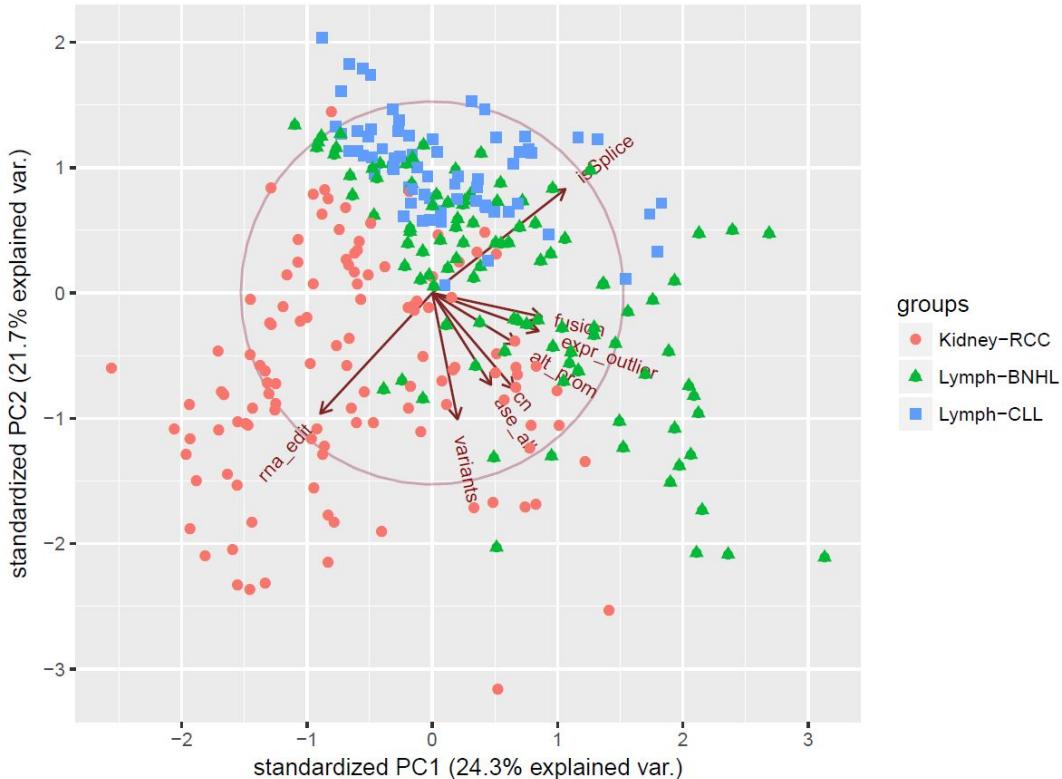
Lymph-BNHL





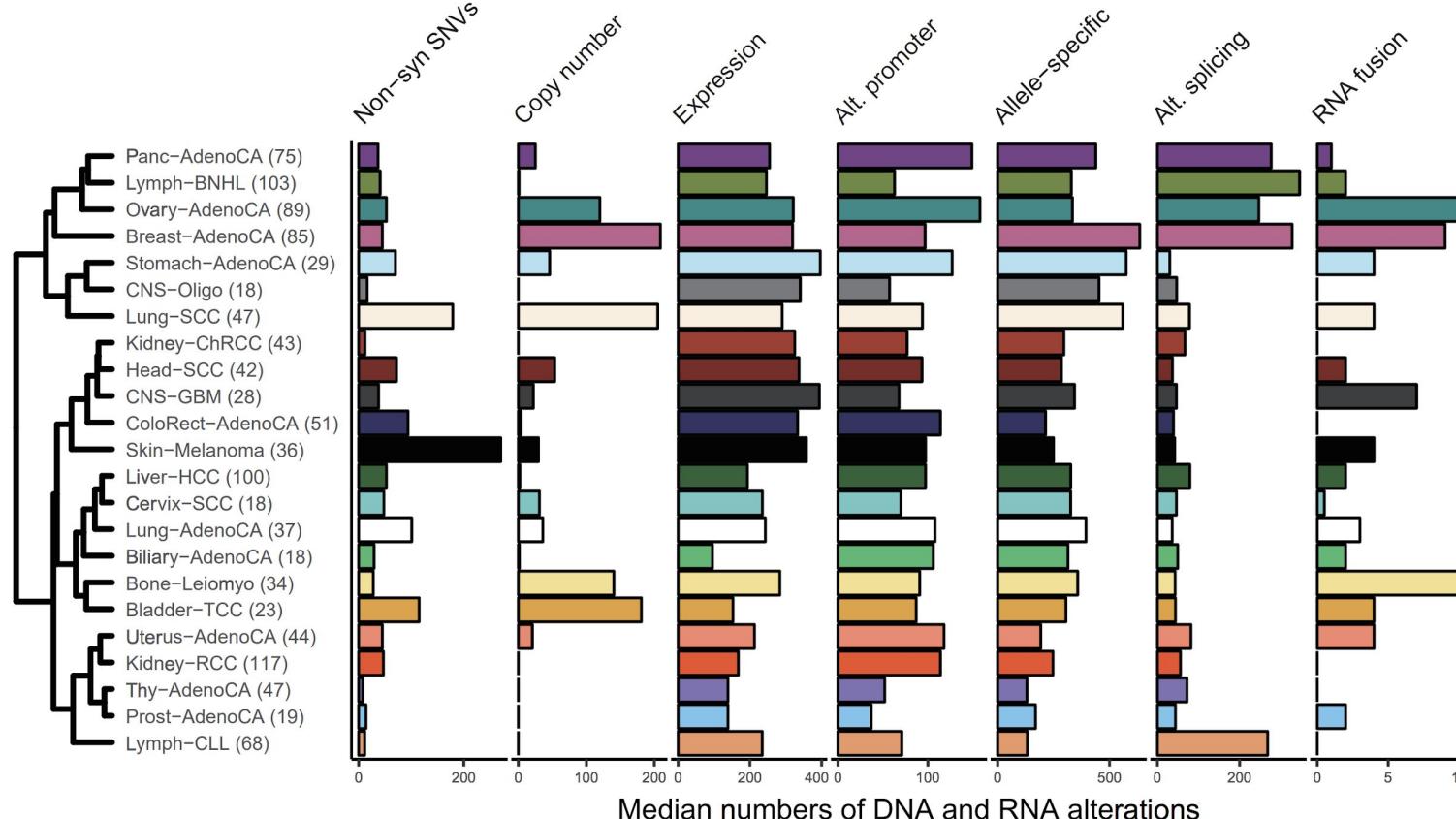
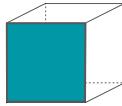
Compare and contrast samples

→Cancer Type

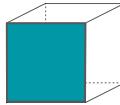


Compare and contrast samples

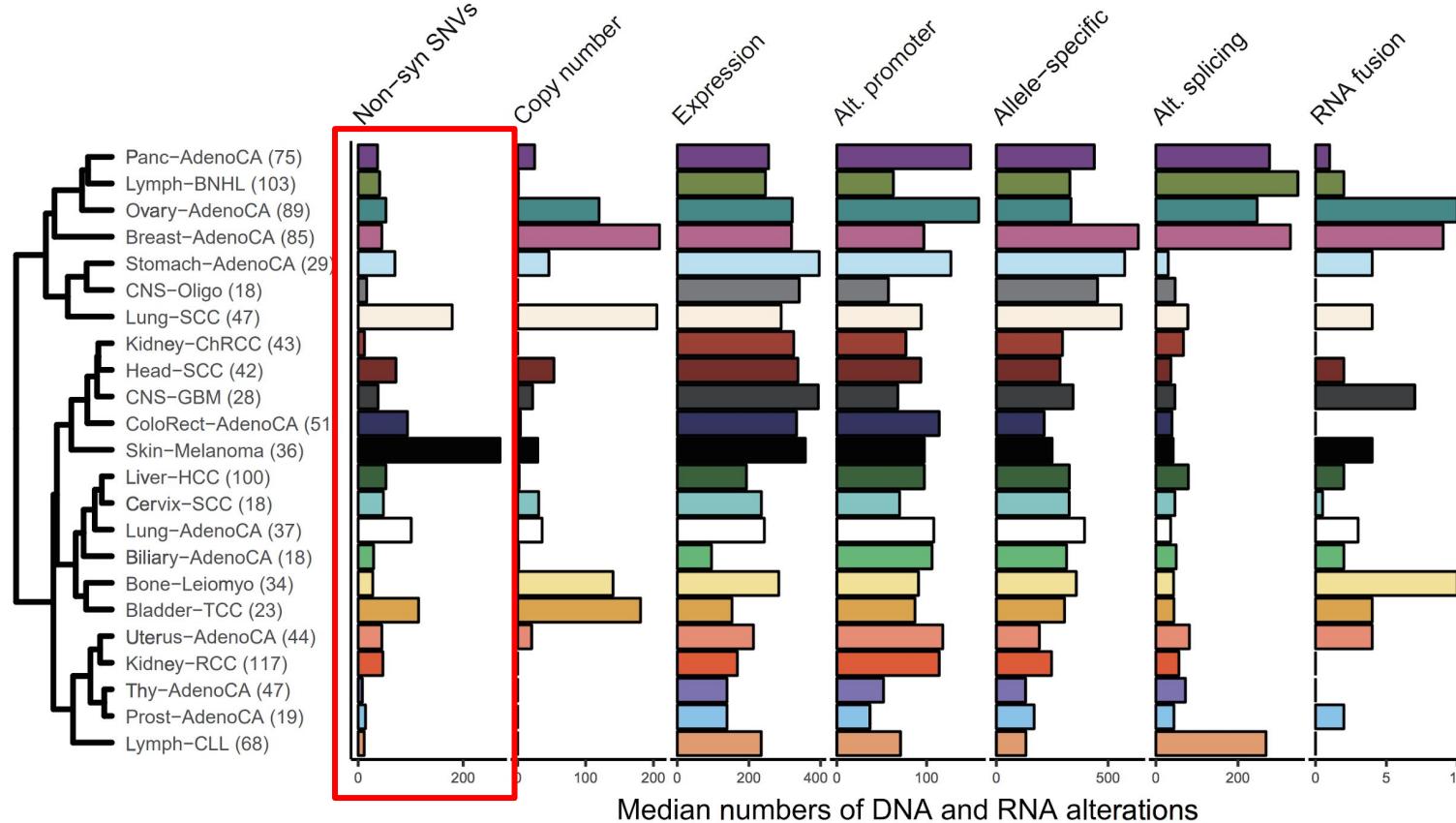
→Cancer Type

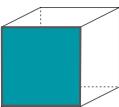


Compare and contrast samples



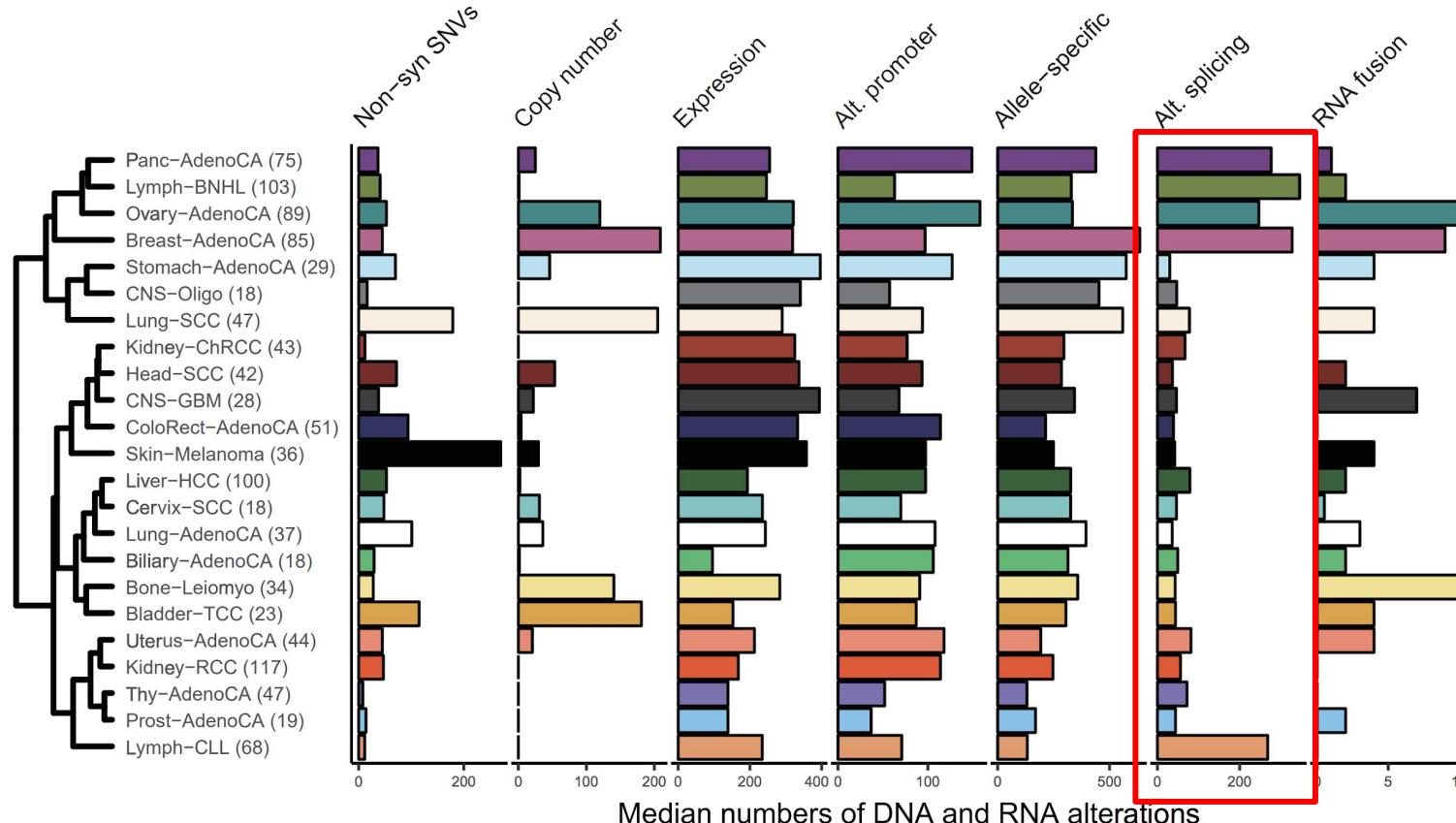
→Cancer Type



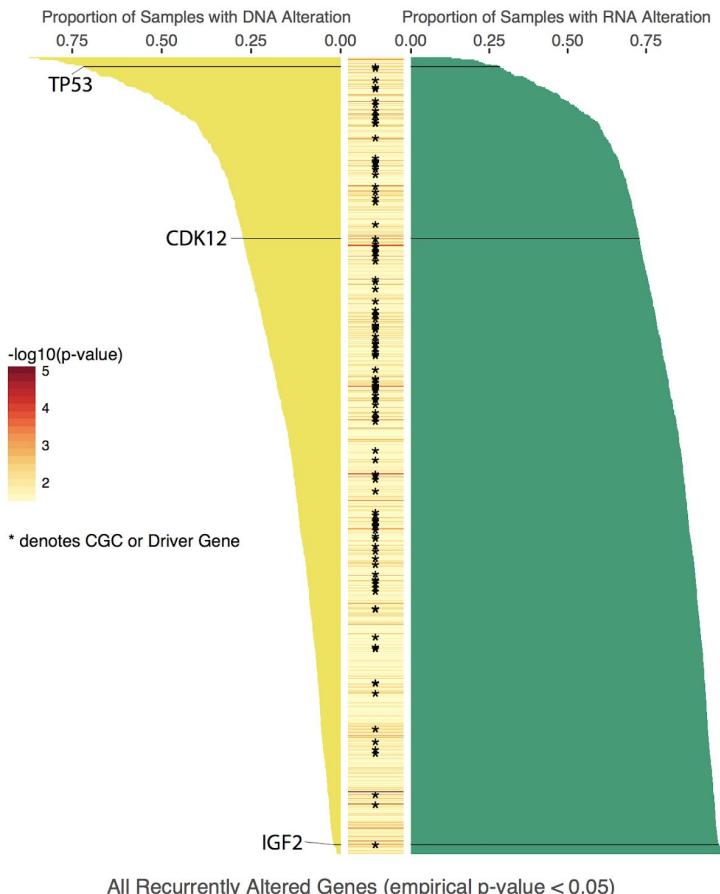
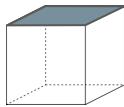


Compare and contrast samples

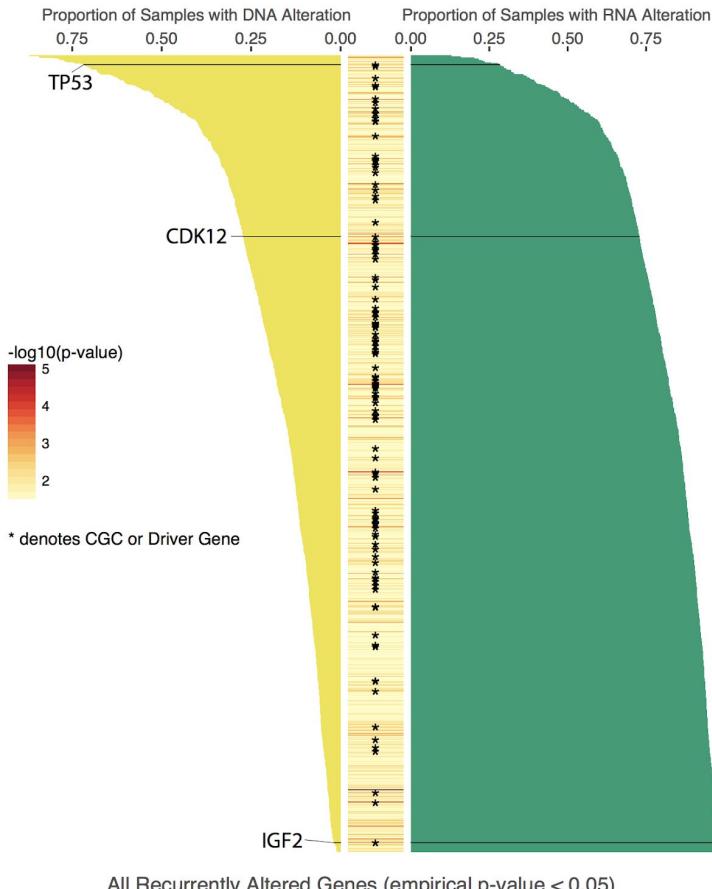
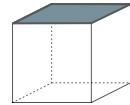
→Cancer Type



Identify known and new recurrently altered genes



Identify known and new recurrently altered genes

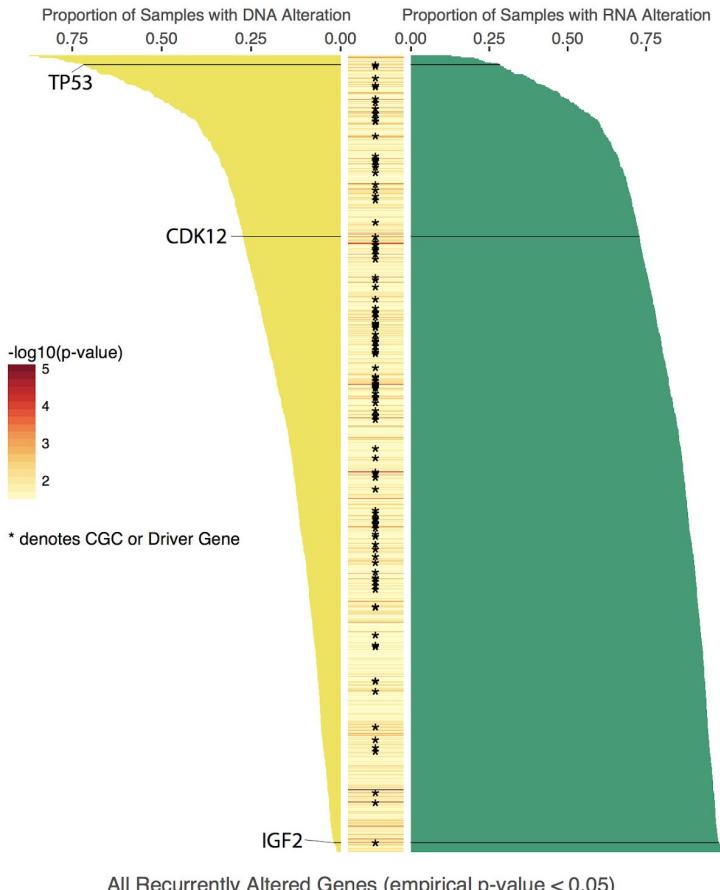
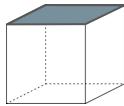


1,012 genes depicted have a significant recurrence score.

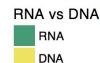
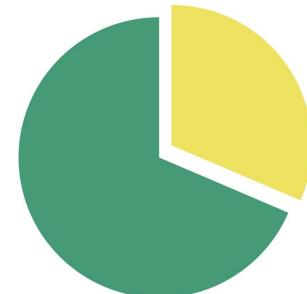
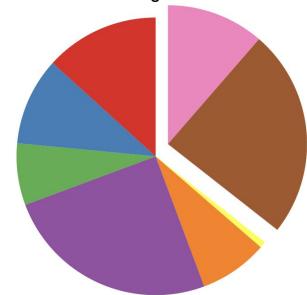
This score:

- Guards against frequent alterations dominating
- Prioritizes genes that are heterogeneously altered

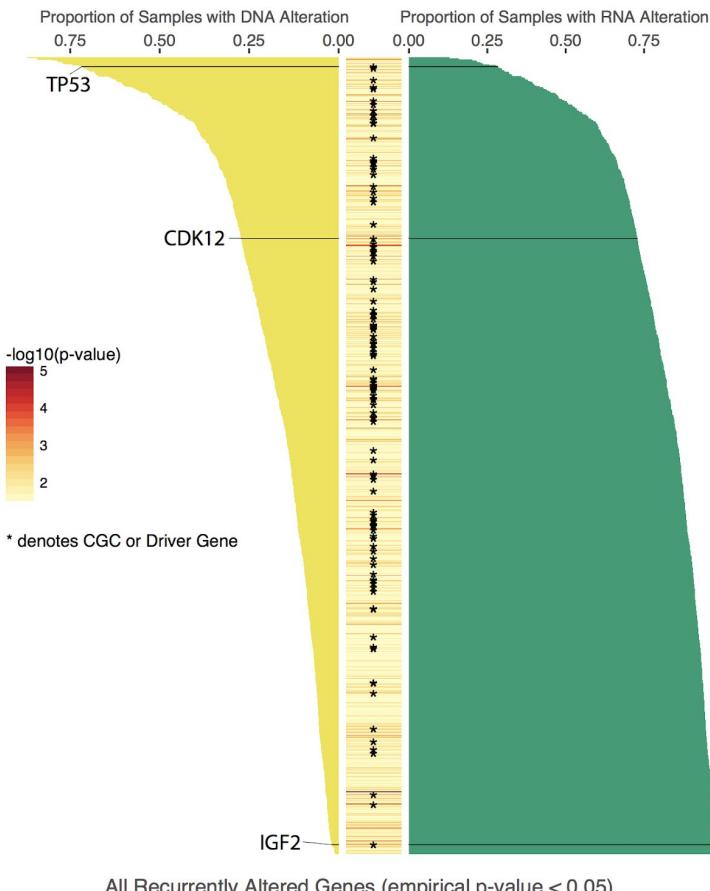
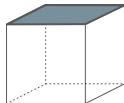
Identify known and new recurrently altered genes



Breakdown of Top 2 Alterations
Across Significant Genes



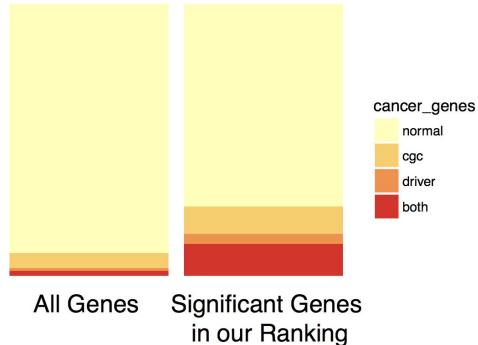
Identify known and new recurrently altered genes



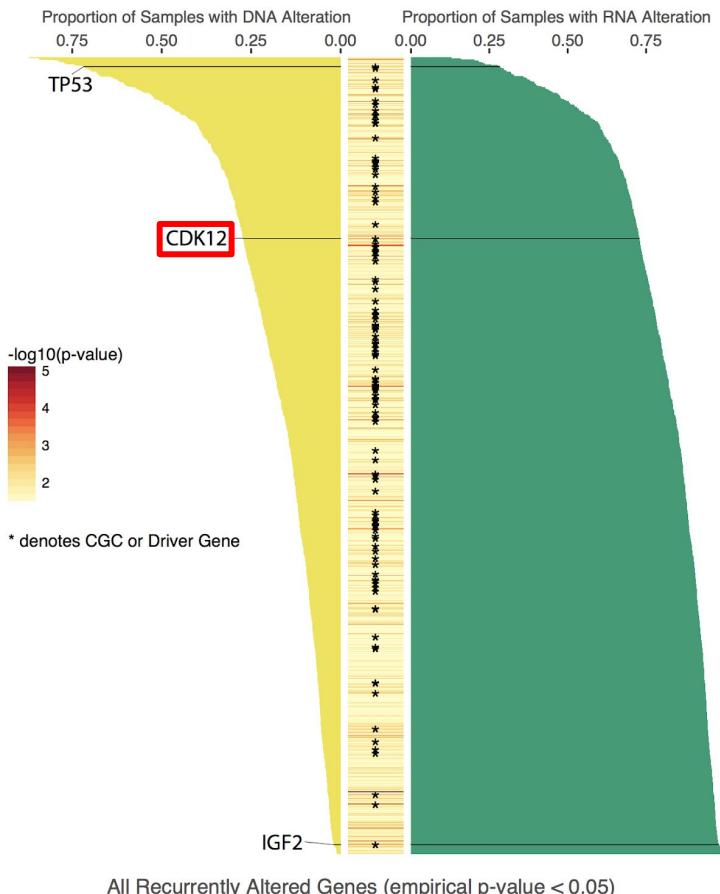
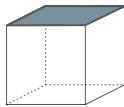
Significant Enrichment
for CGC and PCAWG-8
identified driver genes

- Enrichment: 2.82
- P-value: $5 \cdot 10^{-26}$

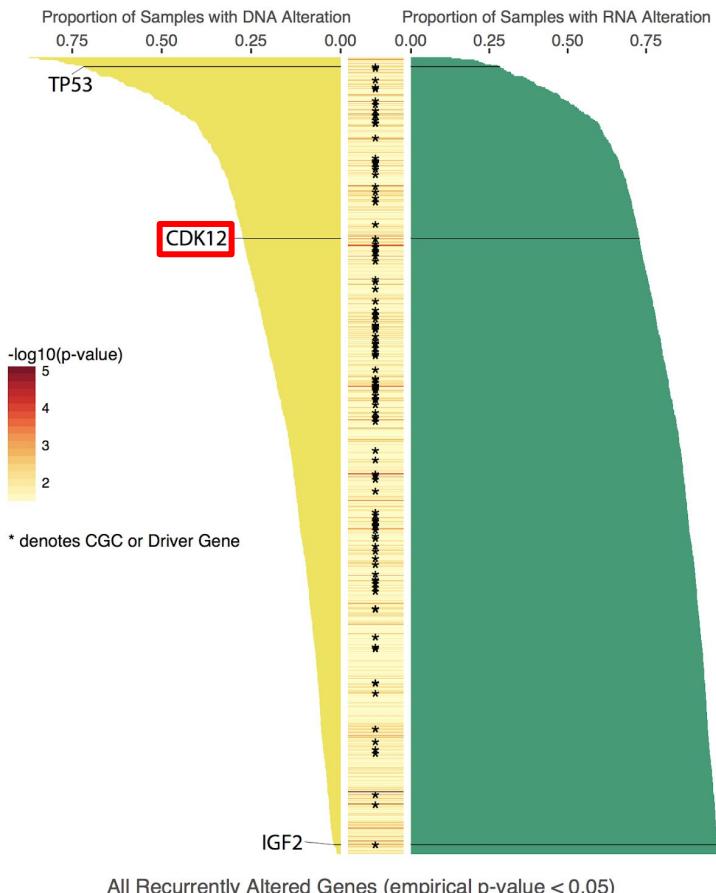
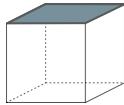
Proportion of Cancer Genes



Identify known and new recurrently altered genes



Identify known and new recurrently altered genes



Only 23 patients have a non-synonymous SNV

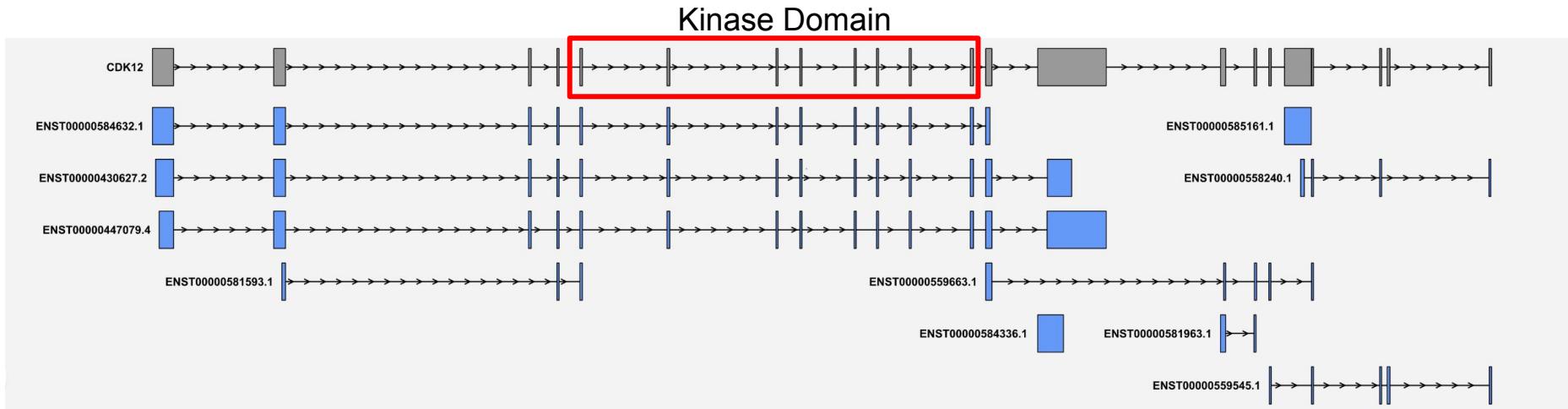
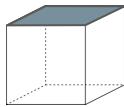
Using SNV+CN alone

- rank 1340

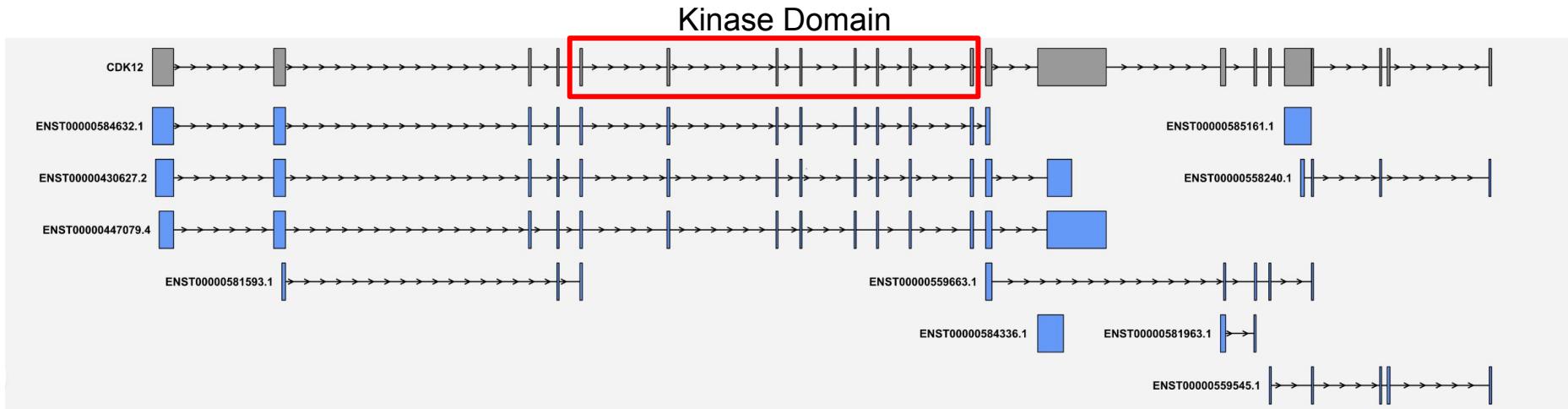
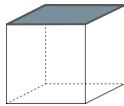
Using RNA+DNA

- rank 40

Identify known and new recurrently altered genes

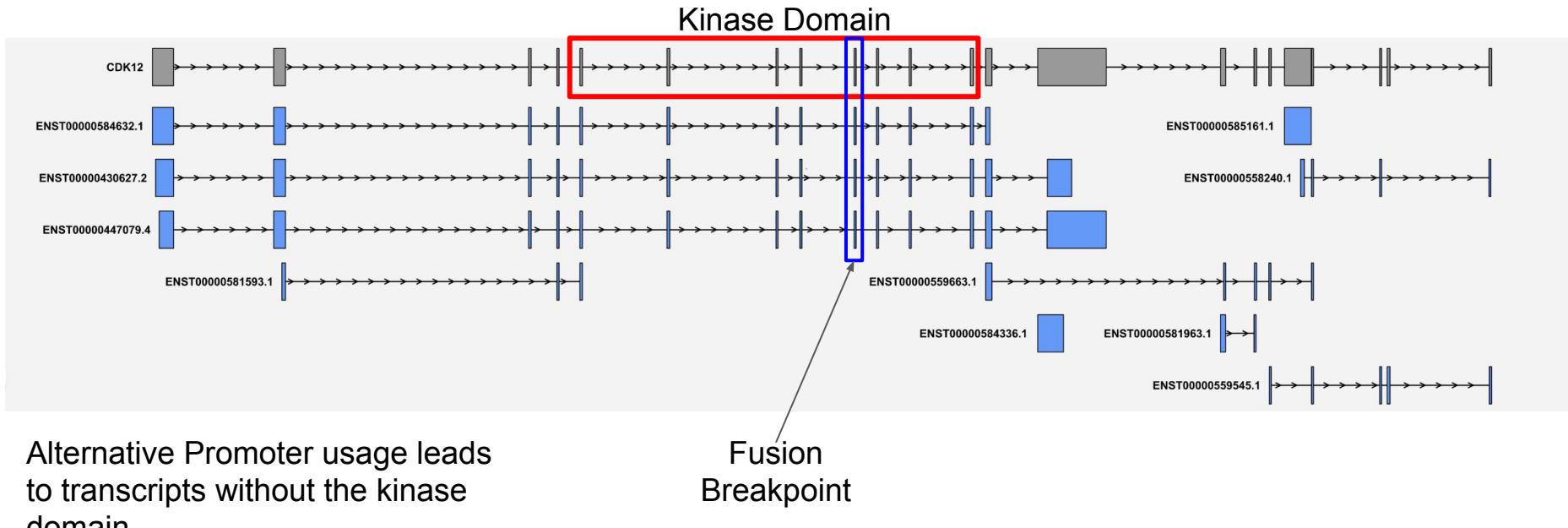
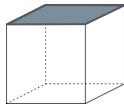


Identify known and new recurrently altered genes

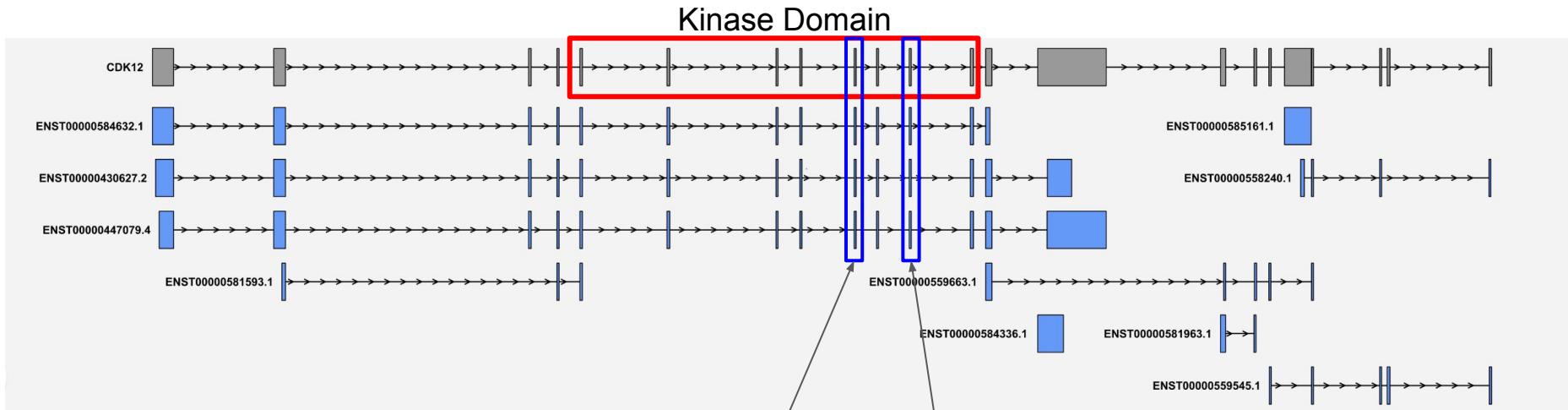
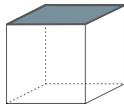


Alternative Promoter usage leads
to transcripts without the kinase
domain.

Identify known and new recurrently altered genes



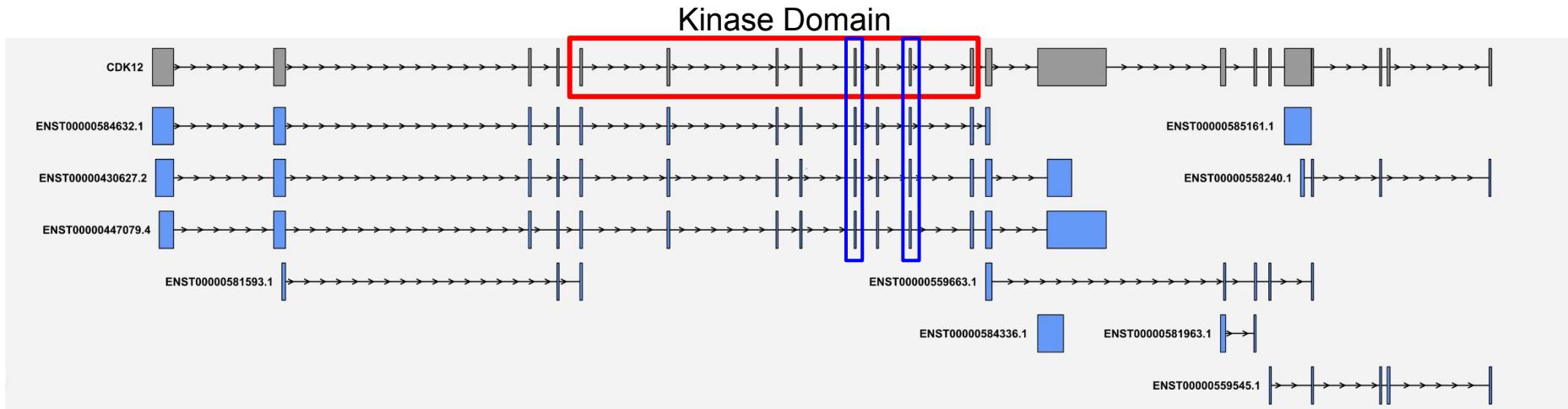
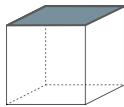
Identify known and new recurrently altered genes



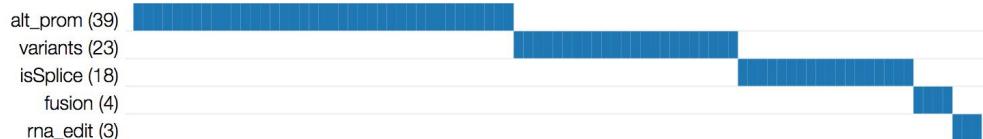
Alternative Promoter usage leads
to transcripts without the kinase
domain

Fusion
Breakpoint
Exon
Skip

Identify known and new recurrently altered genes

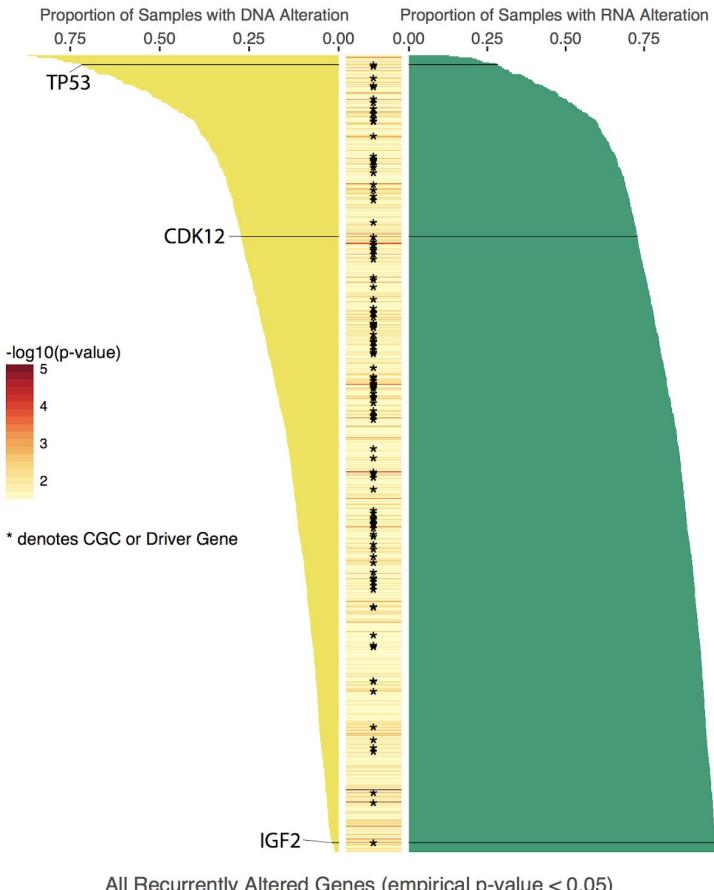
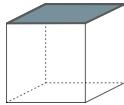


Alteration Matrix for CDK12 (p-value=0.07 WExT)



Coverage: 7.32% (87/1188)

Identify known and new recurrently altered genes

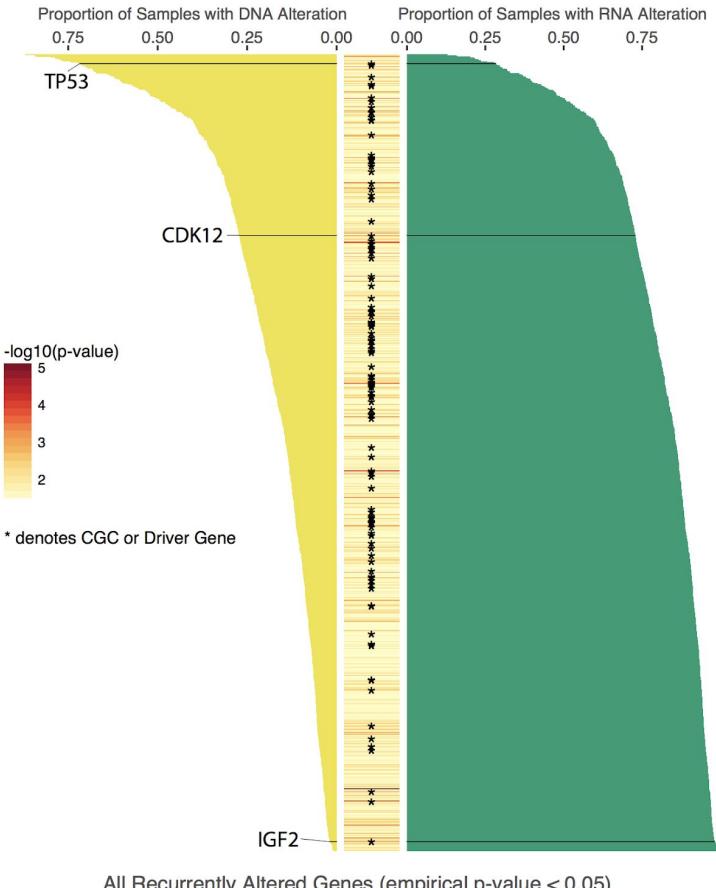
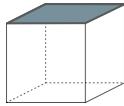


893 genes are neither cancer census nor PCAWG-8 driver genes.

Some of these genes have been implicated in previous cancer studies:

- KLF13
- IRF5
- ZFAT
- BCAS3
- TLK2
- COL6A3

Identify known and new recurrently altered genes



Cancer Cell Res. 2010 Jun;12(4):192-8. doi: 10.1159/000308303. Epub 2010 Jun 2.

Overexpression of KLF13 and FGFR3 in oral cancer cells.

Henzen BJ¹, Gorlin SM²

✉ Author information

Abstract

KLF13 and FGFR3 have important cellular functions and each is believed to play a role in cancer. KLF13 is a transcription factor required for the expression of several oncogenes. FGFR3 is a fibroblast growth factor receptor that initiates a signaling cascade leading to the activation of numerous cellular pathways. Here we show that KLF13 and FGFR3 are overexpressed in oral cancer cells. We also show that artificially reducing cellular levels of KLF13 and FGFR3 decreases cell proliferation and increases sensitivity to ionizing radiation. These data suggest that KLF13 and FGFR3 contribute to malignancy in oral cancer cells and may be useful biomarkers for early detection and possible targets for therapy.

2010 S. Karger AG, Basel.

PMID: 20539070 PMCID: PMC2889258 DOI: 10.1159/000308303

Am J Physiol Endocrinol Metab. 2008 Feb;294(2):E385-91. Epub 2007 Dec 4.

Regulation of Kruppel-like factor 4, 9, and 13 genes and the steroidogenic genes LDLR, StAR, and CYP11A in ovarian granulosa cells.

Natesanpalai S¹, Keriville J, Leung PC, Vedhanayagam JC

✉ Author information

Abstract

Kruppel-like factors (KLFs) are important Sp1-like eukaryotic transcription proteins. The LDLR, StAR, and CYP11A genes exhibit GC-rich Sp1-like sites, which have the potential to bind KLFs in multiprotein complexes. We now report that KLF4, KLF9, and KLF13 transcripts are expressed in and regulate ovarian cells. KLF4 and 13, but not KLF9, mRNA expression was induced and then repressed over time ($P < 0.001$). Combined LH and IGF-I stimulation increased KLF4 mRNA at 2 h ($P < 0.01$), whereas LH decreased KLF13 mRNA at 6 h ($P < 0.05$), and IGF-I reduced KLF13 at 24 h ($P < 0.01$) compared with untreated control. KLF9 was not regulated by either hormone. Transient transfection of KLF4, KLF9, and KLF13 suppressed LDLR/luc, StAR/luc, and CYP11A/luc by 80–90% ($P < 0.001$). Histone-deacetylase (HDAC) inhibitors stimulated LDLR/luc five- to sixfold and StAR/luc and CYP11A/luc activity twofold ($P < 0.001$) and partially reversed suppression by all three KLFs ($P < 0.001$). Deletion of the zinc finger domain of KLF13 abrogated repression of LDLR/luc. Lentiviral overexpression of the KLF13 gene suppressed LDLR mRNA ($P = 0.001$) and CYP11A mRNA ($P = 0.003$) but increased StAR mRNA ($P = 0.007$). Collectively, these data suggest that KLFs may recruit inhibitory complexes containing HDAC coresspressors, thereby repressing LDLR and CYP11A transcription. Conversely, KLF13 may recruit unknown coactivators or stabilize StAR mRNA, thereby explaining enhancement of in situ StAR gene expression. These data introduce new potent gonadal transregulators of genes encoding proteins that mediate steroid uptake and steroid biosynthesis.

PMID: 18056793 PMCID: PMC2747322 DOI: 10.1152/ajpendo.00480.2007

Oncogene. 2016 Oct 28;35(42):5565-5575. doi: 10.1038/onc.2016.97. Epub 2016 Apr 4.

KLF13 regulates the differentiation-dependent human papillomavirus life cycle in keratinocytes through STAT5 and IL-8.

Zhang W¹, Hondo S¹, Maniar KP², Cheng S¹, Jie C³, Bademkoker AW⁴, Krensky AM^{1,5}, Clayberger C¹

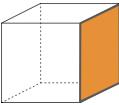
✉ Author information

Abstract

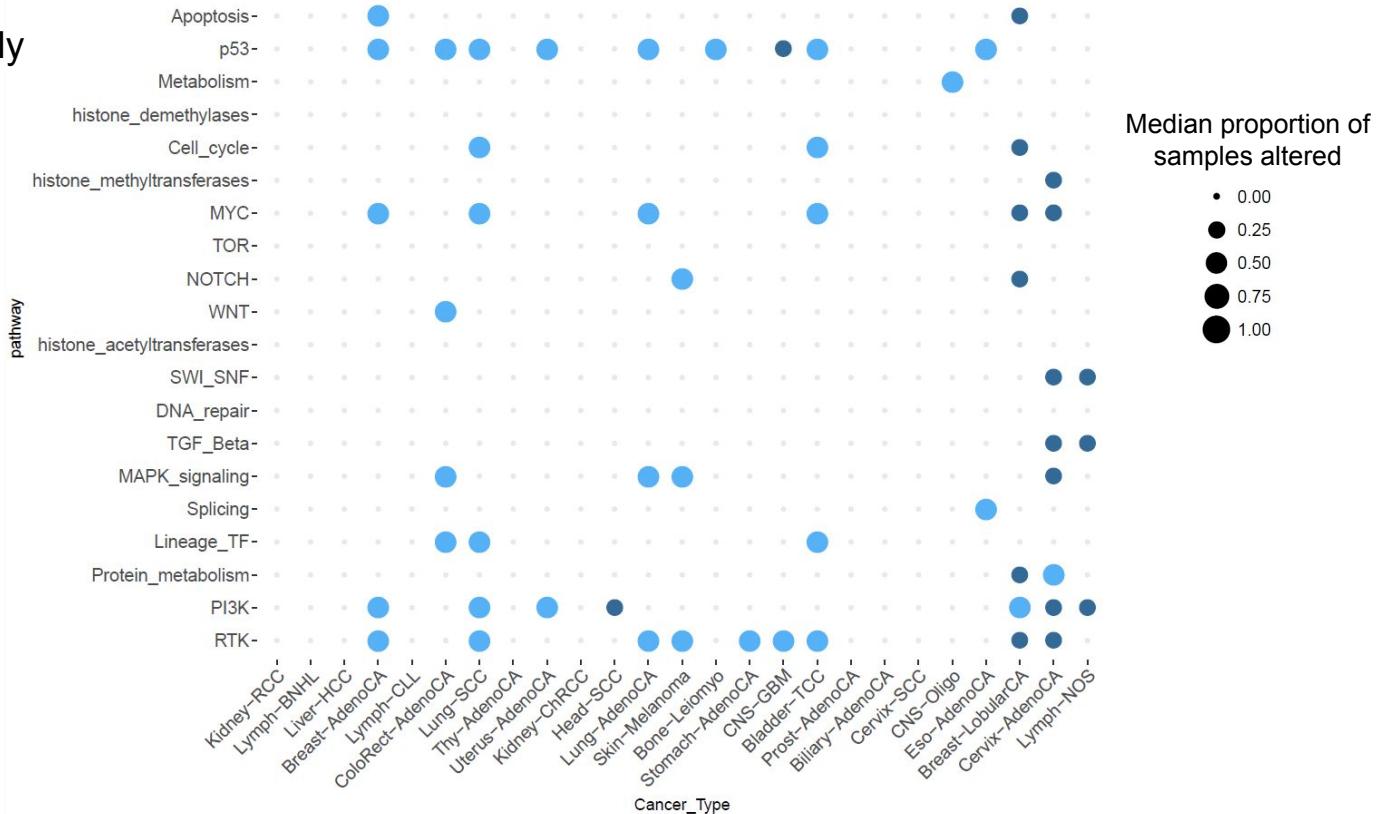
High-risk strains of human papillomavirus (HPV) are the causative agents of cervical and anogenital cancers and are associated with 5% of all human cancers. Although prophylactic vaccines targeting a subset of HPV types are available, they are ineffective in HPV-infected individuals. Elucidation of the mechanisms controlling HPV replication may allow development of novel anti-HPV therapeutics. Infectious HPV viruses are produced during terminal differentiation of host cells. The process of viral maturation requires synergistic interactions between viral and cellular proteins that leads to amplification of the viral genome and expression of late viral genes. Here we show that the transcription factor Kruppel-like factor 13 (KLF13) has a critical role in the HPV life cycle. KLF13 is overexpressed in HPV-positive keratinocytes and cervical cancer cell lines. Expression of KLF13 in normal cervical epithelium is low, but increases significantly in cervical intraepithelial neoplasia and invasive squamous cervical cancer. After HPV infection, the E7 protein suppresses ubiquitin ligase FBW7 expression leading to an increase in KLF13 expression. Reduction of KLF13 with short hairpin RNA in differentiating HPV-positive cells resulted in diminished levels of viral gene expression and genome amplification. Knockdown of KLF13 also reduced the level of the transcription factor signal transducer and activator of transcription 5, which led to the downregulation of the ataxia-telangiectasia mutated DNA damage pathway and the chemokine interleukin-8 (IL-8). In addition, neutralization of IL-8 diminished viral genome amplification in differentiating HPV-positive cells. Thus, KLF13 is critical for the activation of the HPV productive life cycle and is likely involved in initiation and progression of cervical cancer.

PMID: 27641562 DOI: 10.1038/onc.2016.97

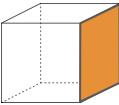
Identify alterations in cancer pathways



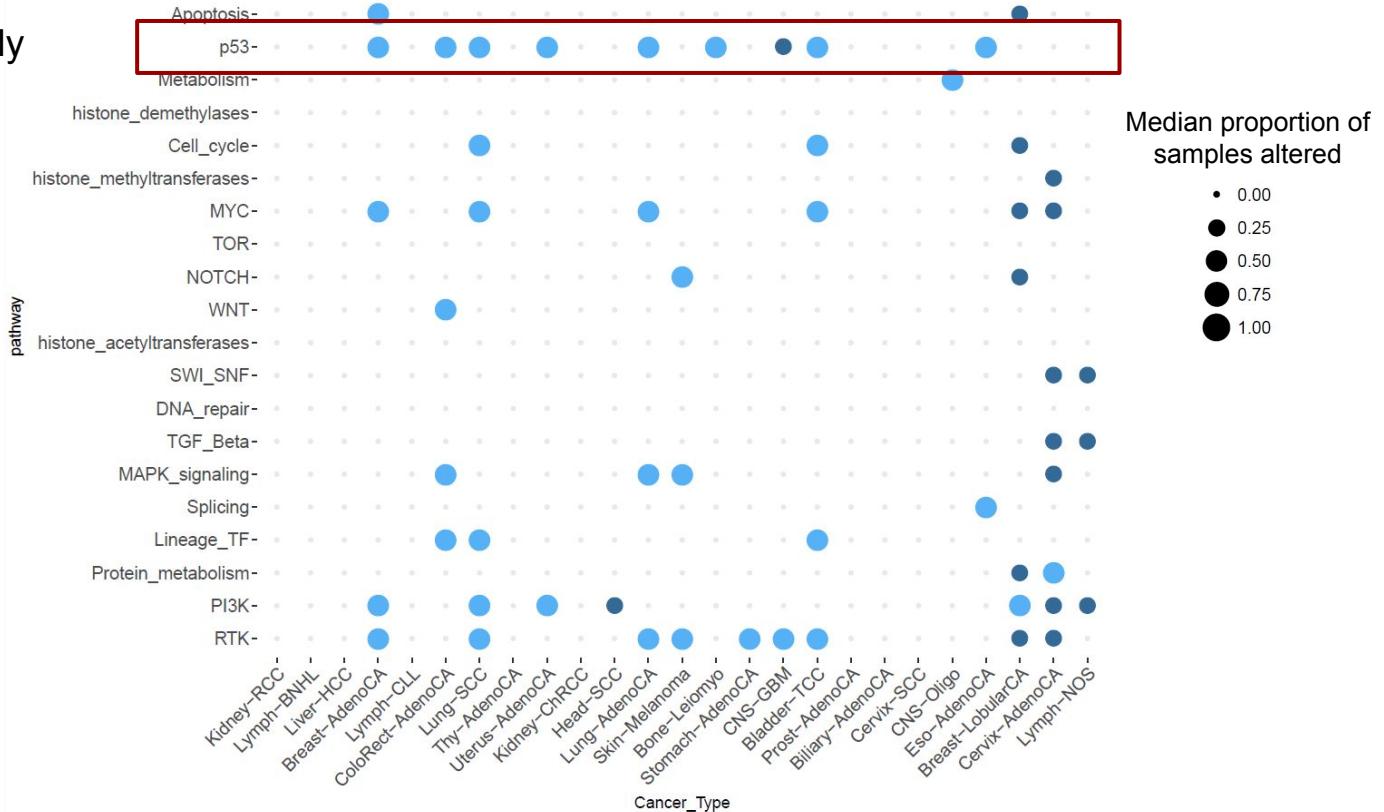
DNA alterations only



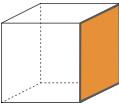
Identify alterations in cancer pathways



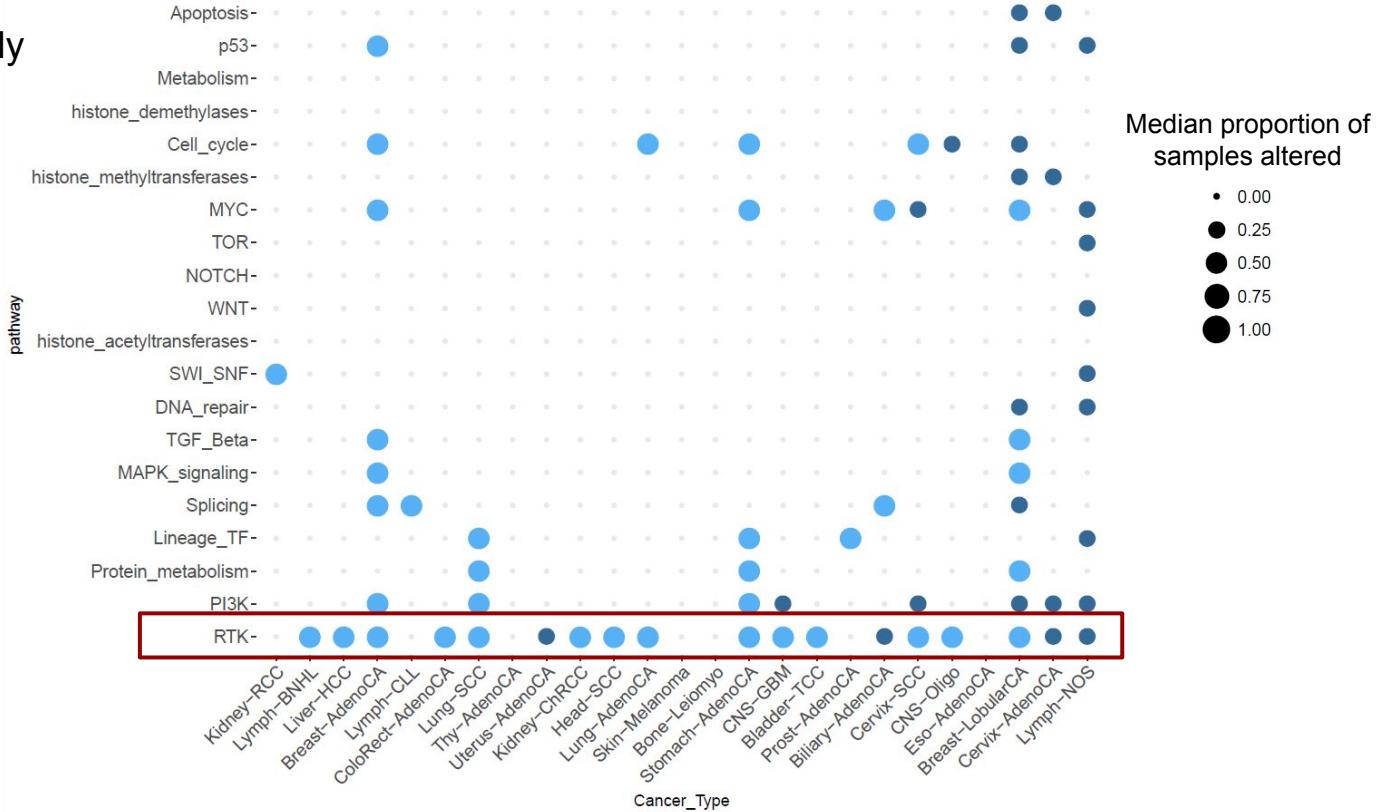
DNA alterations only



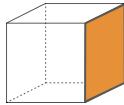
Identify alterations in cancer pathways



RNA alterations only



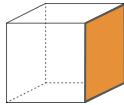
Identify alterations in cancer pathways



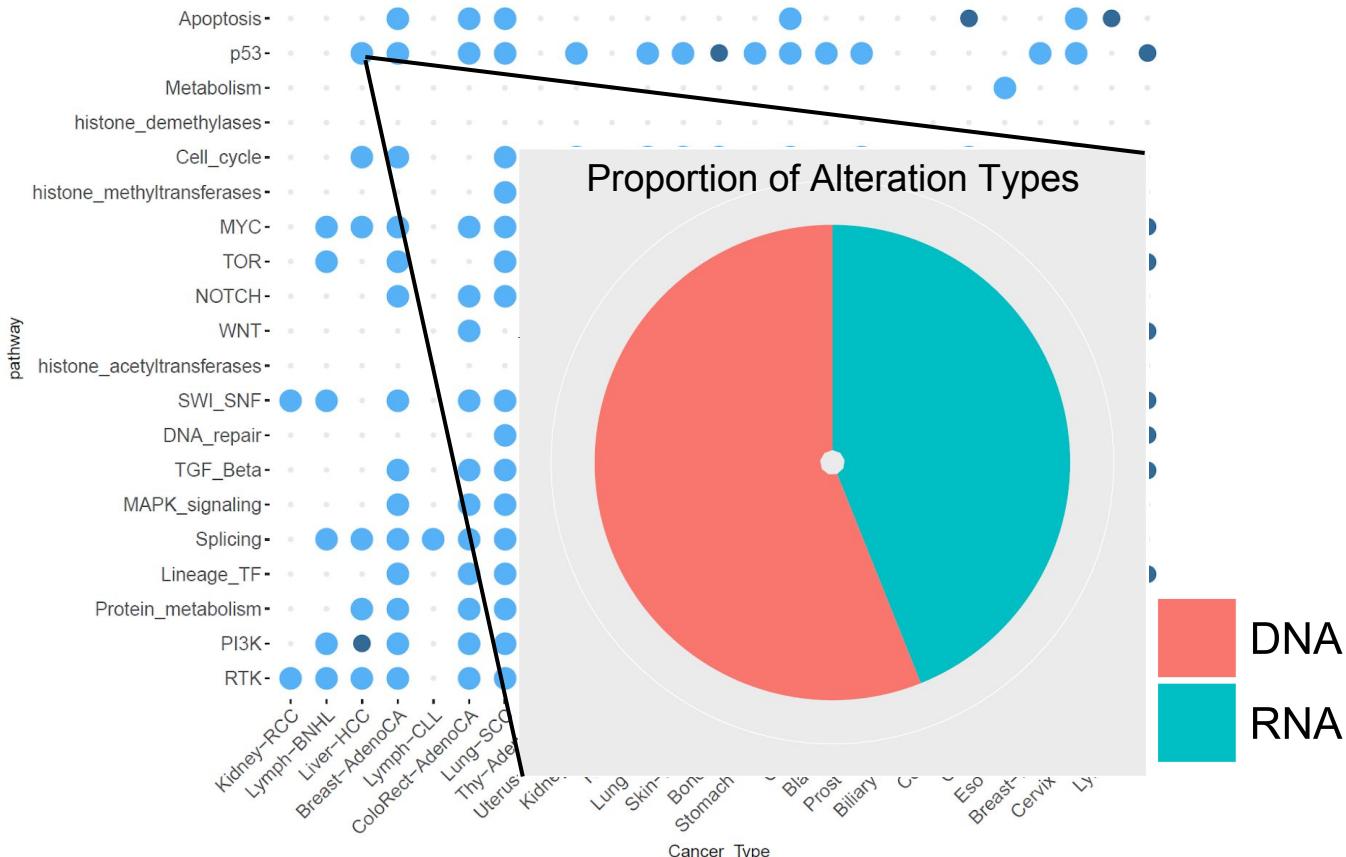
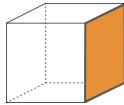
DNA + RNA



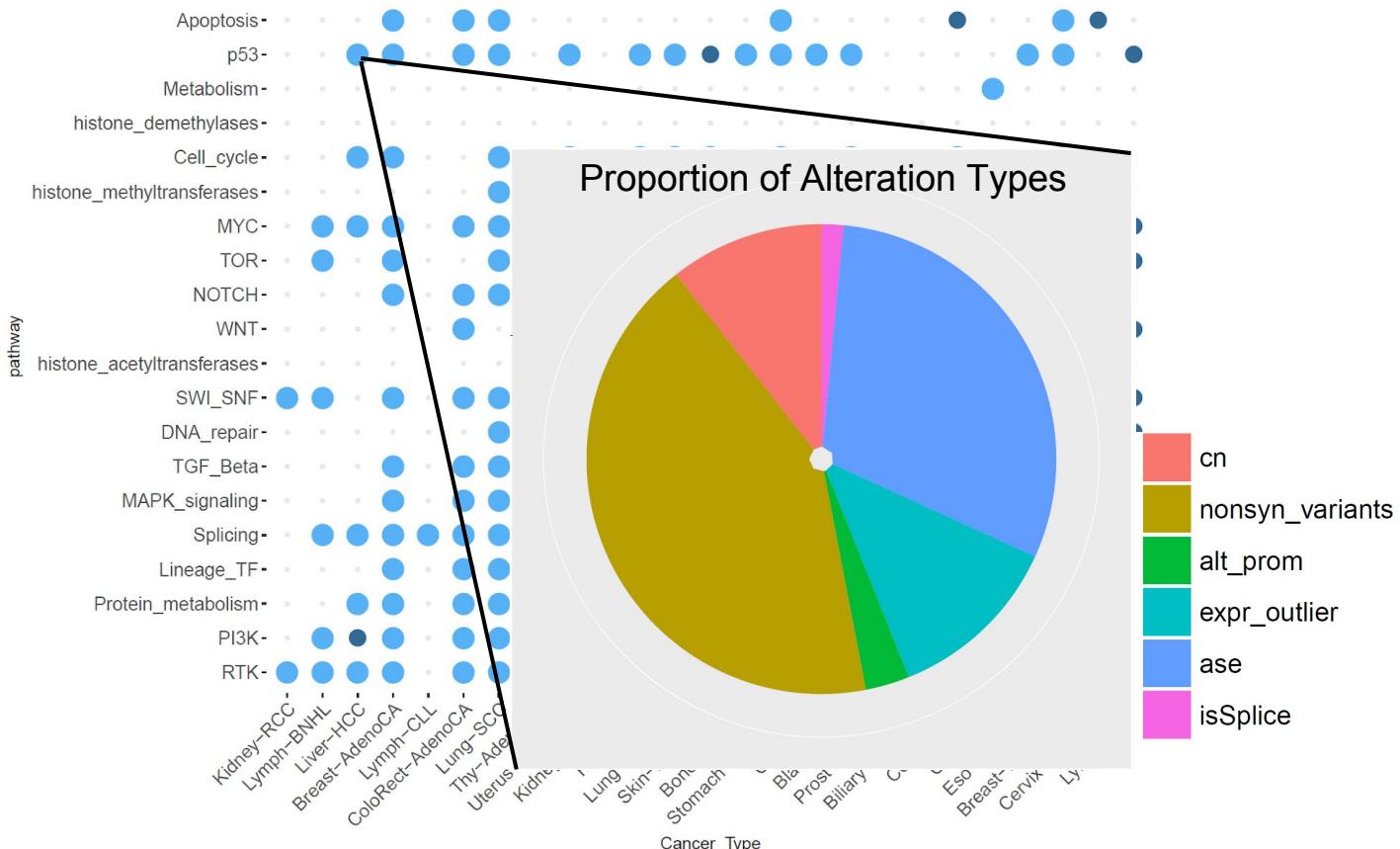
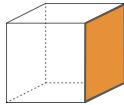
Identify alterations in cancer pathways



Identify alterations in cancer pathways



Identify alterations in cancer pathways



Conclusion

- Performed comprehensive analysis of transcriptomic alterations across 27 cancer types.

Conclusion

- Performed comprehensive analysis of transcriptomic alterations across 27 cancer types.
- Increased the number of samples found to contain an alteration in cancer relevant pathways.

Conclusion

- Performed comprehensive analysis of transcriptomic alterations across 27 cancer types.
- Increased the number of samples found to contain an alteration in cancer relevant pathways.
- Identified cancer relevant genes through a recurrence analysis using both DNA- and RNA-level aberrations.

Acknowledgements

Biomedical Informatics Group ETHZ

- **Kjong Van-Lehmann**
- **André Kahles**
- Stefan Stark
- **Gunnar Rätsch**

PanCancer Analysis of Whole Genomes Consortium
Transcriptomes Working Group (PCAWG-3)

- **Alvis Brazma**
- **Angela N. Brooks**
- Claudia Calabrese
- Nuno A. Fonseca
- Jonathan Goke
- Roland F. Schwarz
- Zemin Zhang

Tri-Institutional Training Program in Computational Biology and Medicine

ETH Zürich Informatik



**Genomic basis for RNA alterations revealed by whole-genome analyses of
27 cancer types**

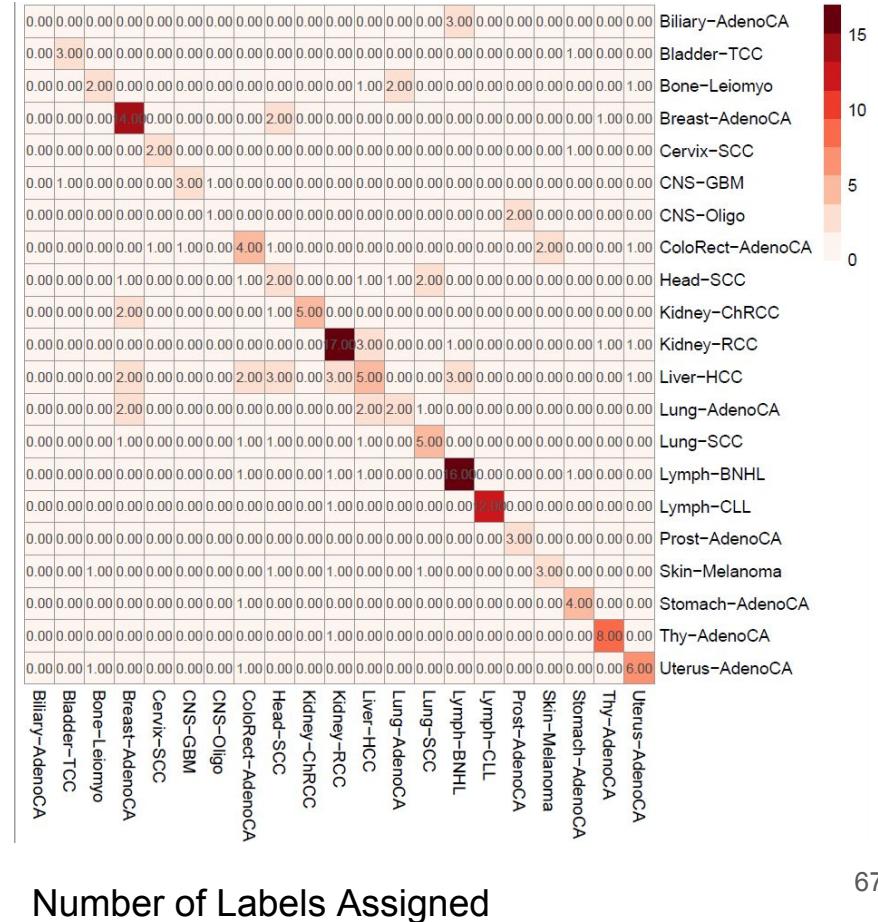
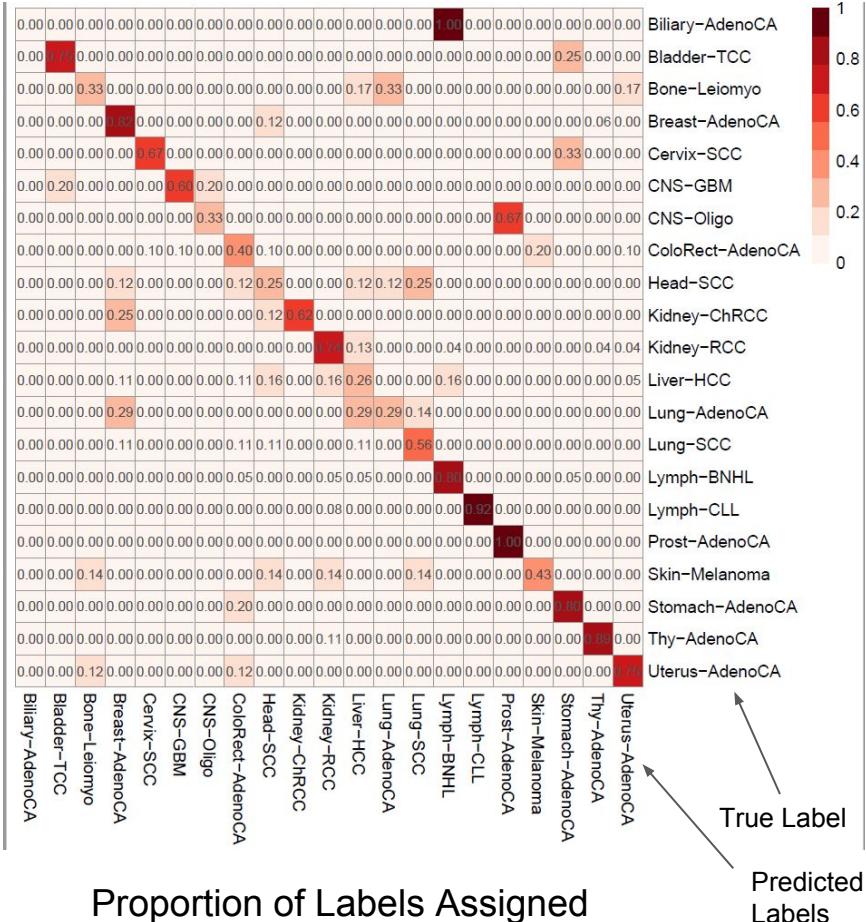


International
Cancer Genome
Consortium

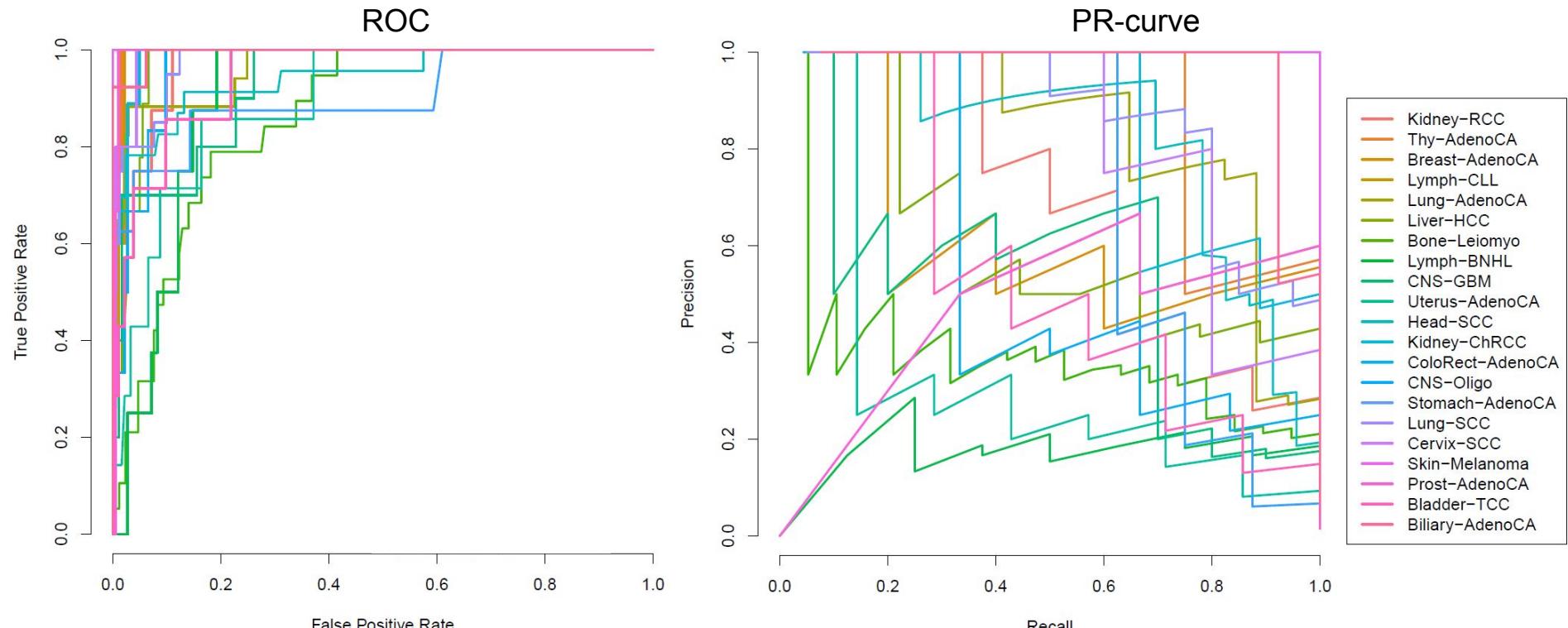


Questions?

Predict Histotype

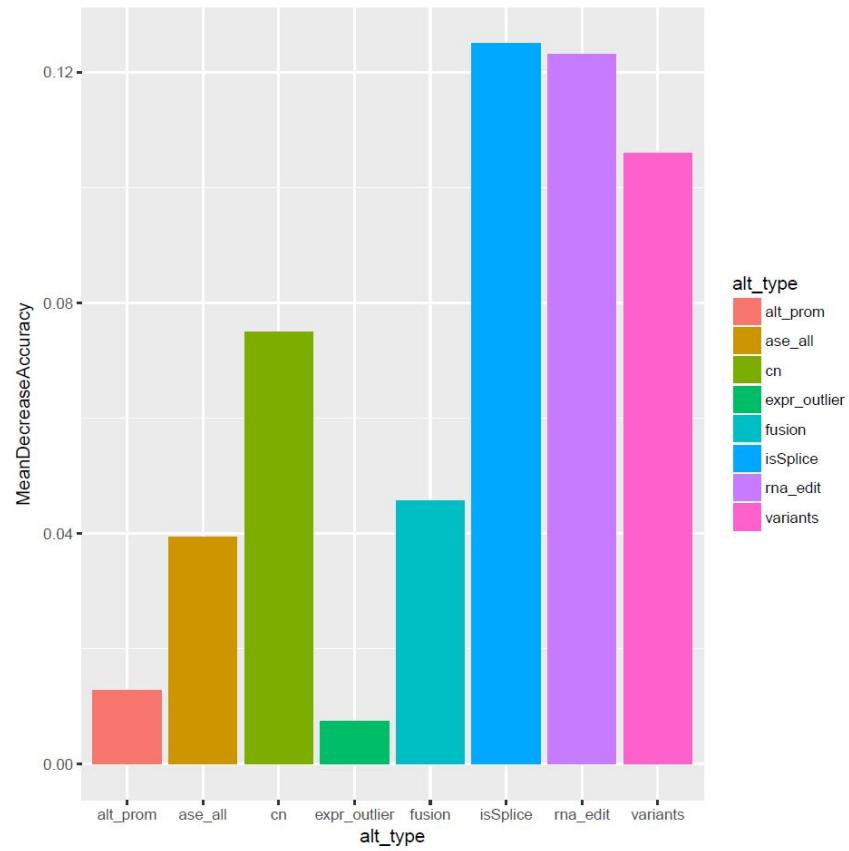
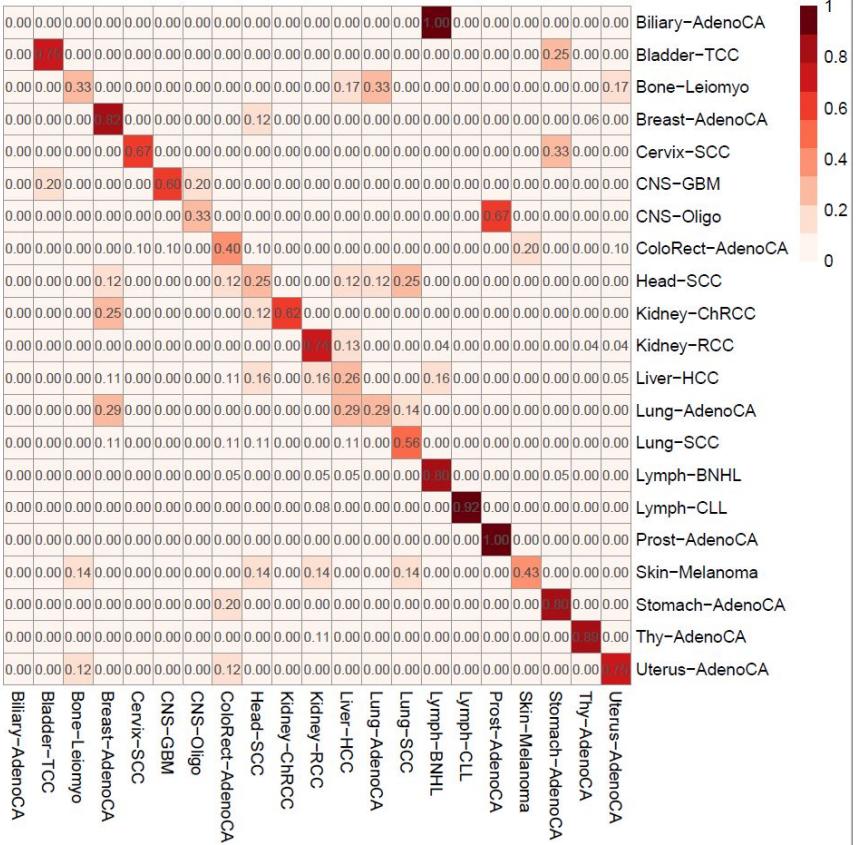


ROC and PR-curves for predicting Histotype



Here we see that even though the ROC-AUC seems to be high, the PR-AUC is not perfect. This is caused by a large class-size imbalance. 68

Predict Histotype

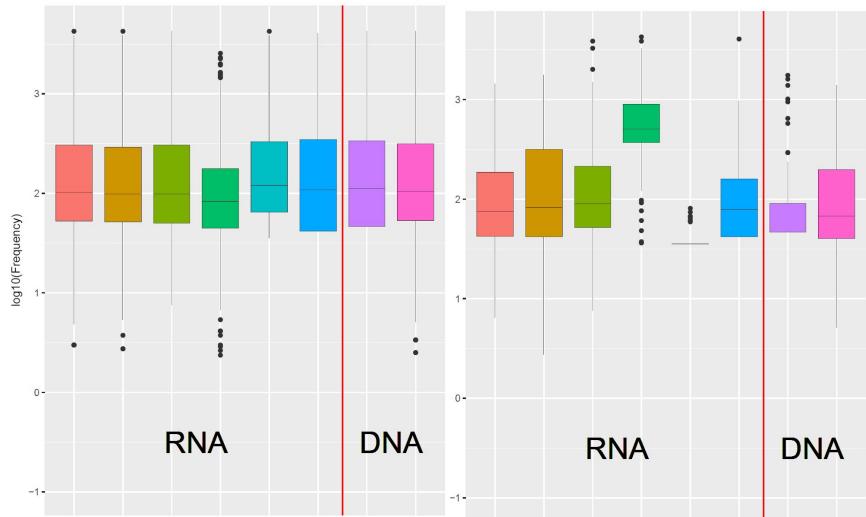


We see that we are able to correctly predict some histotypes using the transcriptomic alterations. Furthermore, we see that DNA-level aberrations aren't the only important features, but that splicing is very important in prediction.

Compare and contrast samples

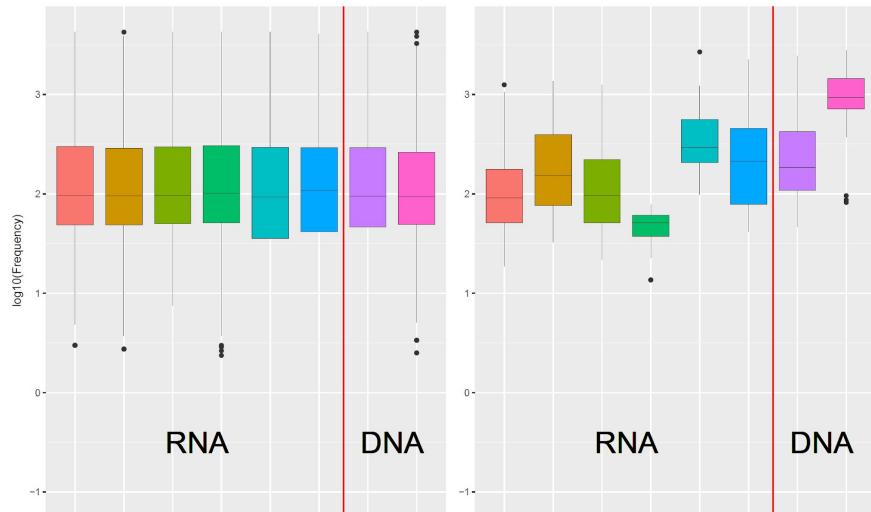
→ Mutational Signature

- alt_pattern
- alt_prom
- expr_outlier
- ase_all
- isSplice
- rna_edit
- fusion
- cn
- variants



Non-signature 9

signature 9

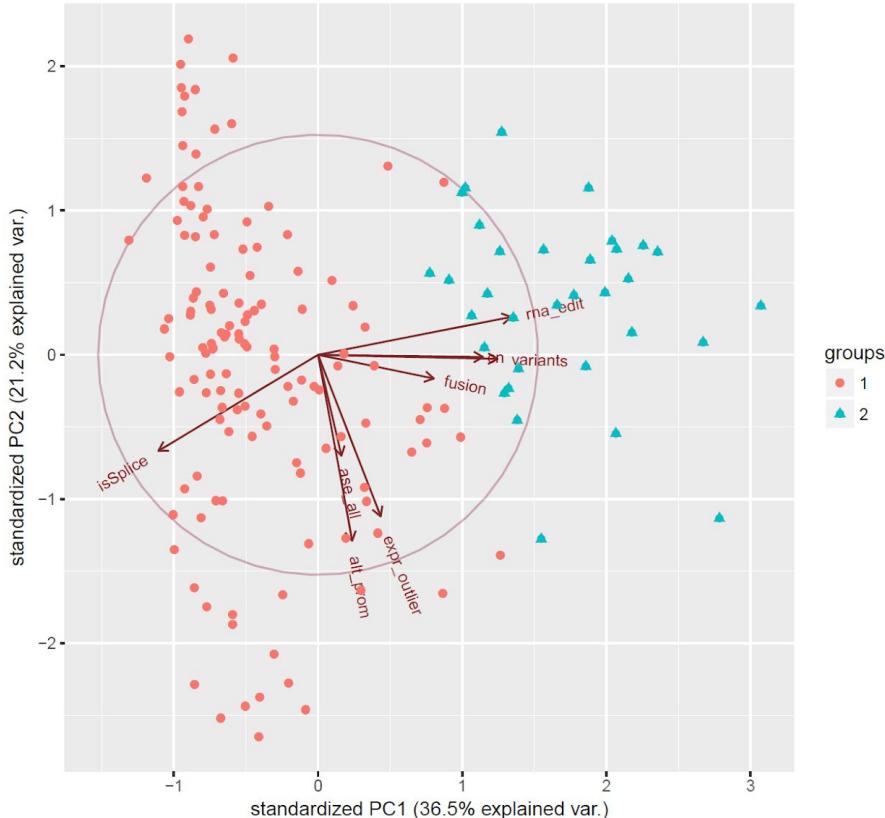


Non-signature 7

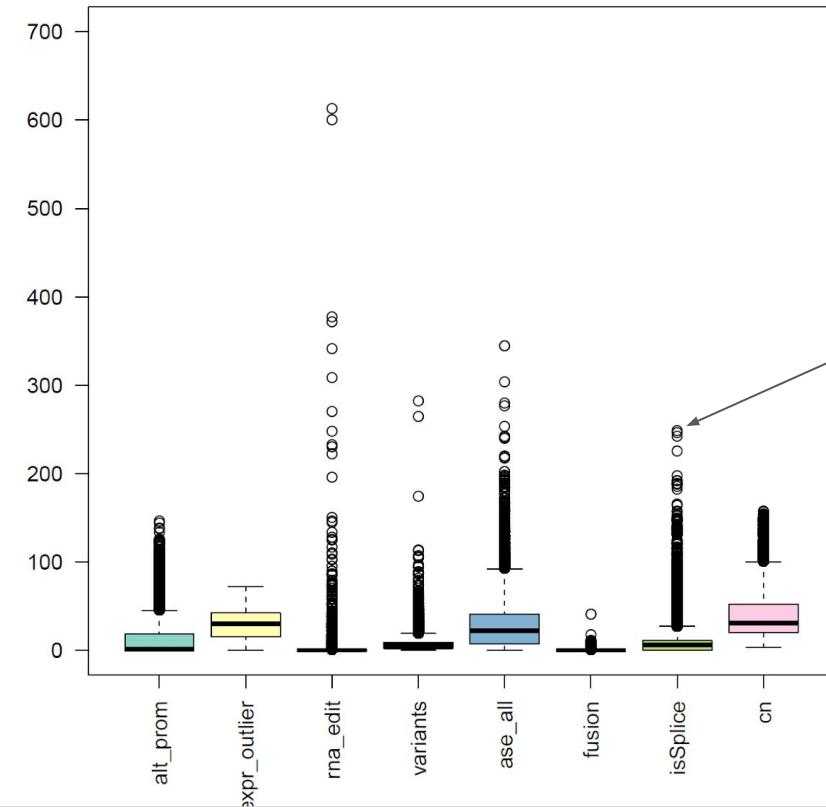
signature 7

Compare and contrast samples

→ Mutational Signature

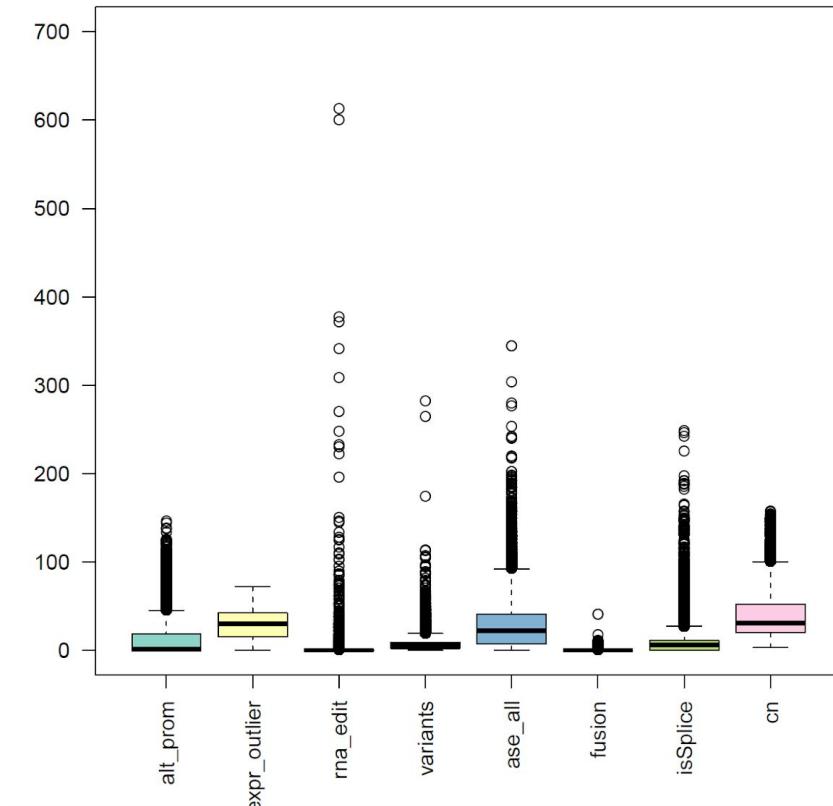


Overview of Alteration Frequencies



This point is the gene ZNF777 and it has a splicing alteration in 249 samples.

Overview of Alteration Frequencies



Aggregating across genes leads to frequent alterations dominating.

Scoring Method

Gene ID	alt1	alt2	alt3	alt4	sum
IGF2	0	1	95	1	97
PTEN	7	0	70	0	77
TP53	0	2	100	0	102
IGLL5	9	13	37	0	59

Scoring Method

Gene ID	alt1	alt2	alt3	alt4
IGF2	0	1	95	1
PTEN	7	0	70	0
TP53	0	2	100	0
IGLL5	9	13	37	0

Ranks →

Gene ID	alt1	alt2	alt3	alt4
IGF2	4	3	2	1
PTEN	2	4	3	4
TP53	4	2	1	4
IGLL5	1	1	4	4

Scoring Method

Gene ID	alt1	alt2	alt3	alt4
IGF2	0	1	95	1
PTEN	7	0	70	0
TP53	0	2	100	0
IGLL5	9	13	37	0

Ranks →

Gene ID	alt1	alt2	alt3	alt4
IGF2	4	3	2	1
PTEN	2	4	3	4
TP53	4	2	1	4
IGLL5	1	1	4	4

1, 2, 3, 4
2, 3, 4, 4
1, 2, 4, 4
1, 1, 4, 4

Scoring Method

Gene ID	alt1	alt2	alt3	alt4
IGF2	0	1	95	1
PTEN	7	0	70	0
TP53	0	2	100	0
IGLL5	9	13	37	0

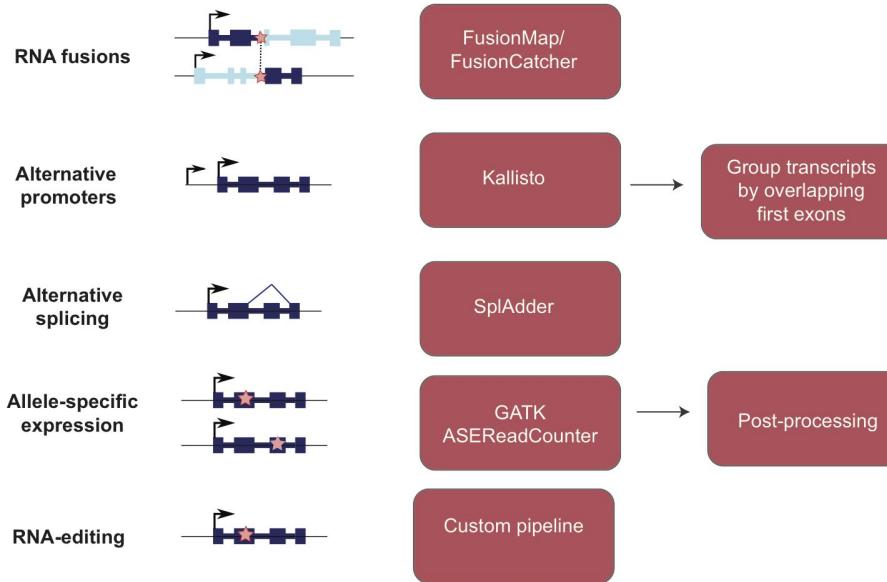
Ranks →

Gene ID	alt1	alt2	alt3	alt4
IGF2	4	3	2	1
PTEN	2	4	3	4
TP53	4	2	1	4
IGLL5	1	1	4	4

1, **2**, 3, 4
2, **3**, 4, 4
1, **2**, 4, 4
1, **1**, 4, 4

We take the second smallest rank to ensure heterogeneous alterations are prioritized.

Overview of PCAWG-3 Data

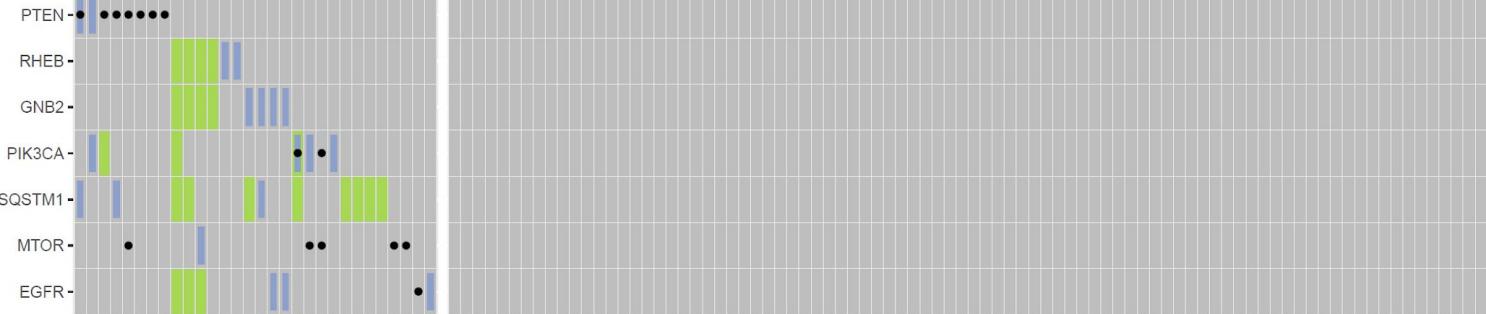


RNA alterations reveal additional driver gene and pathway alterations

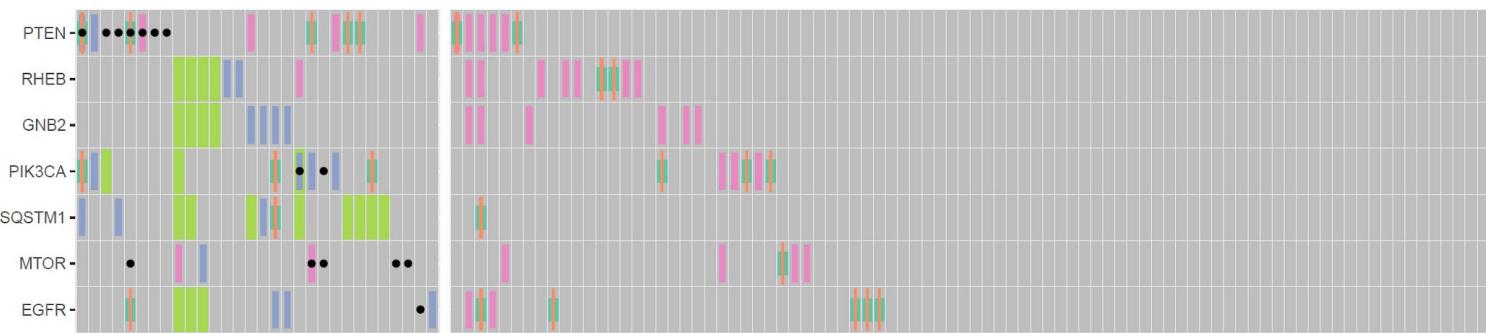
- Alterations
- Expression Outlier
 - Alt. Promoter
 - CN
 - fusions
 - Missense Variants
 - Allele Specific Exp.
 - Splice Outlier

30 Samples

27 Samples



Published Alterations
Only
(CN, variants,
expression outlier)



With All Alterations