Passenger Mutations in >2500 cancer genomes: Overall functional impact

Slides freely downloadable from **Lectures.GersteinLab.org** & “tweetable” (via **@markgerstein**)

See last slide for more info.

Canonical model of drivers & passengers in cancer

Drivers
directly confer a selective growth advantage to the tumor cell.

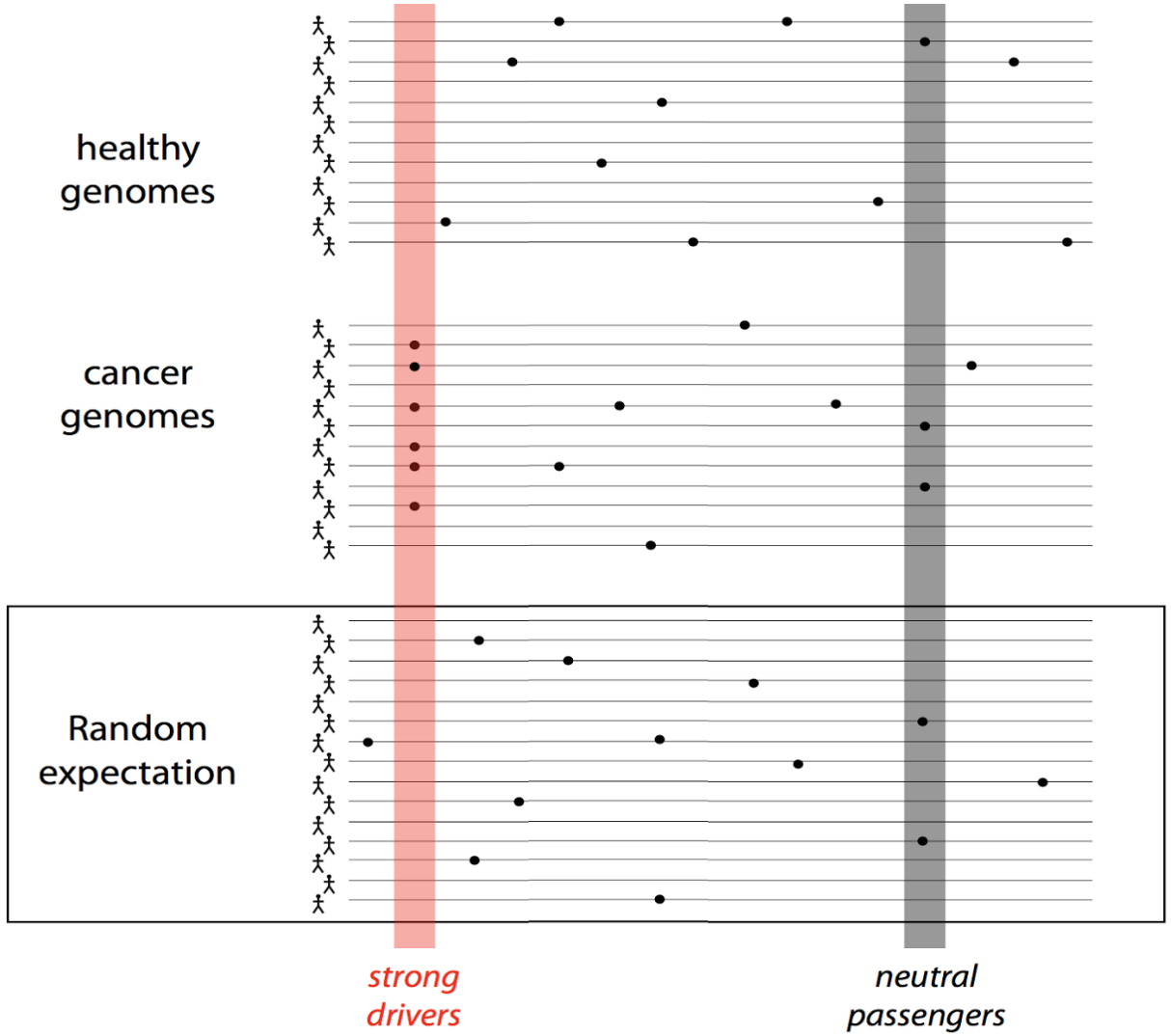
A typical tumor contains 2-8 drivers.

identified through signals of positive selection.

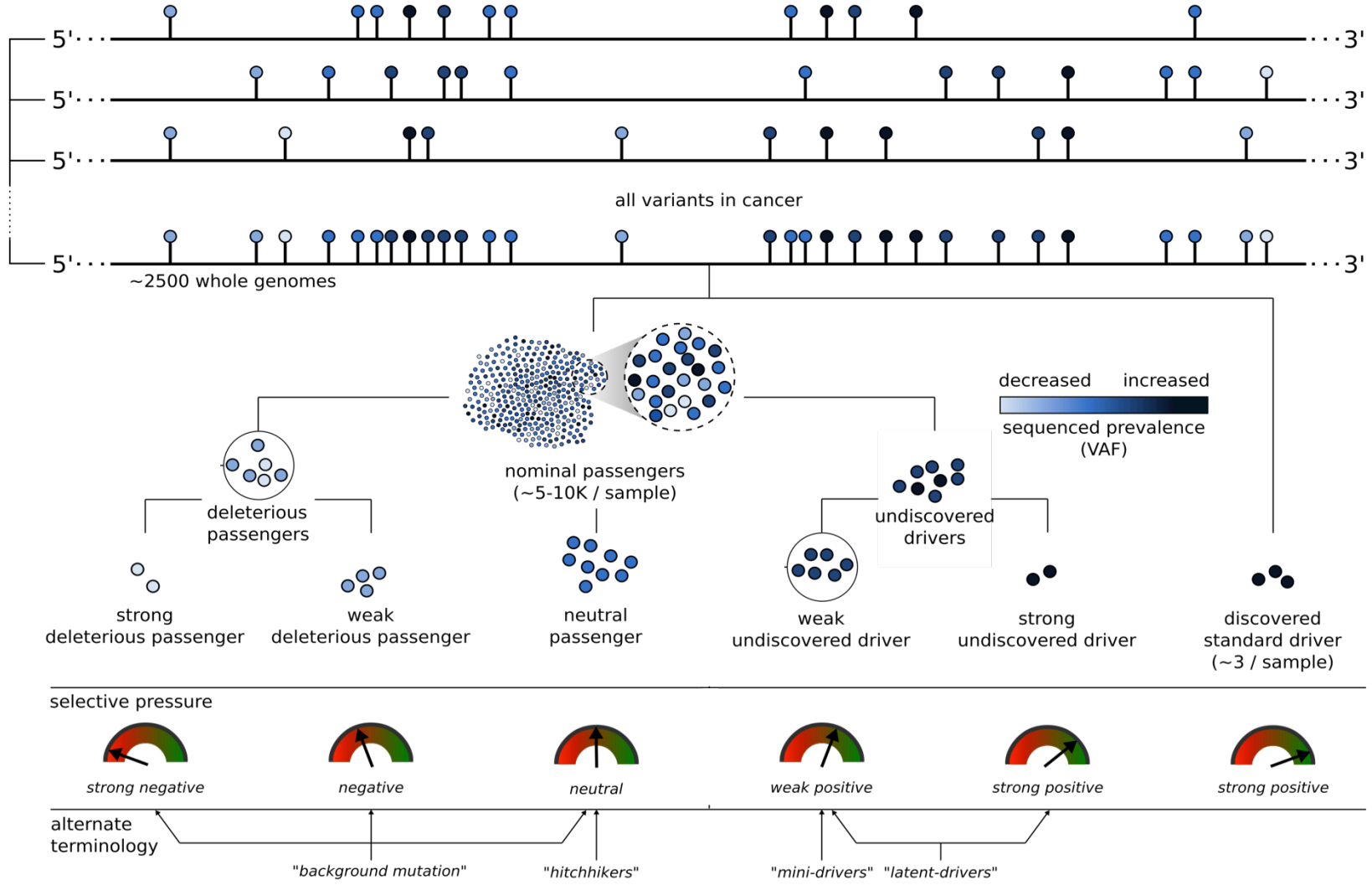
Existing cohorts of ~100s give enough power to identify

Passengers
Conceptually, a passenger mutation has no direct or indirect effect on tumor progression.

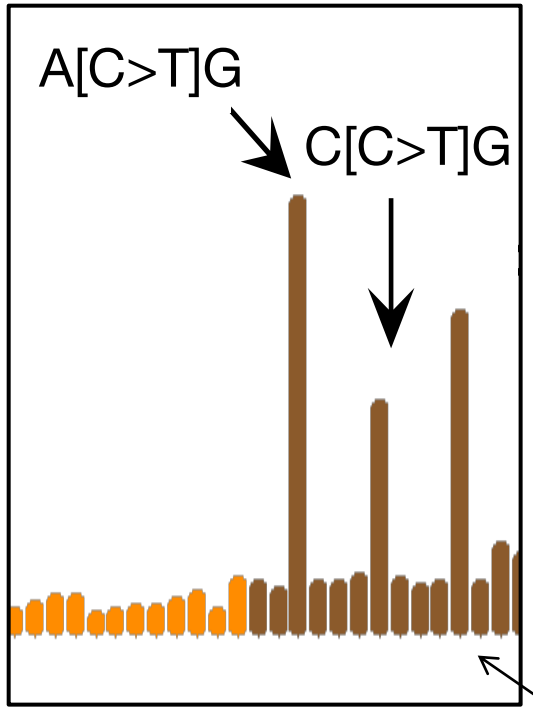
There are 1000s of passengers in a typical cancer genome.
[Vogelstein Science 2013. 339:1546]

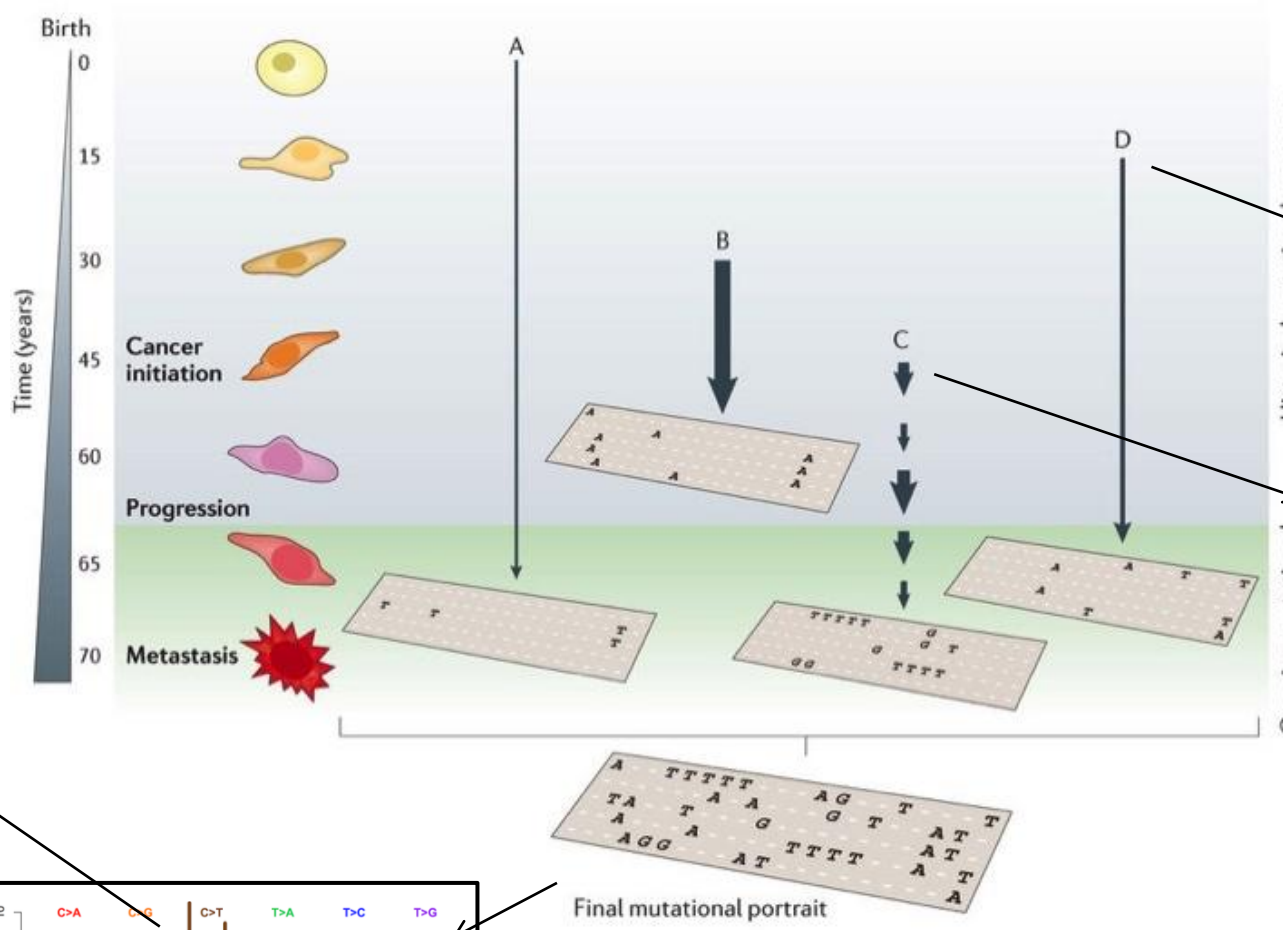


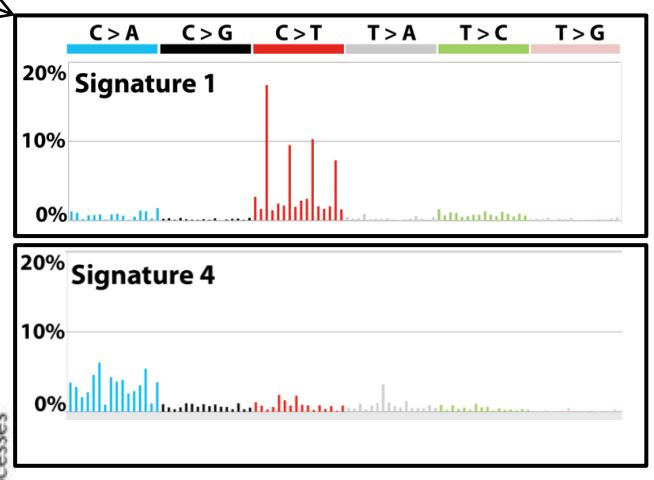
Conceptual extension of the canonical model of drivers & passengers



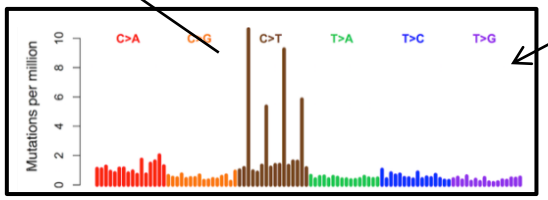
Mutational processes carry context-specific signatures







S: Mutation signature
inferred

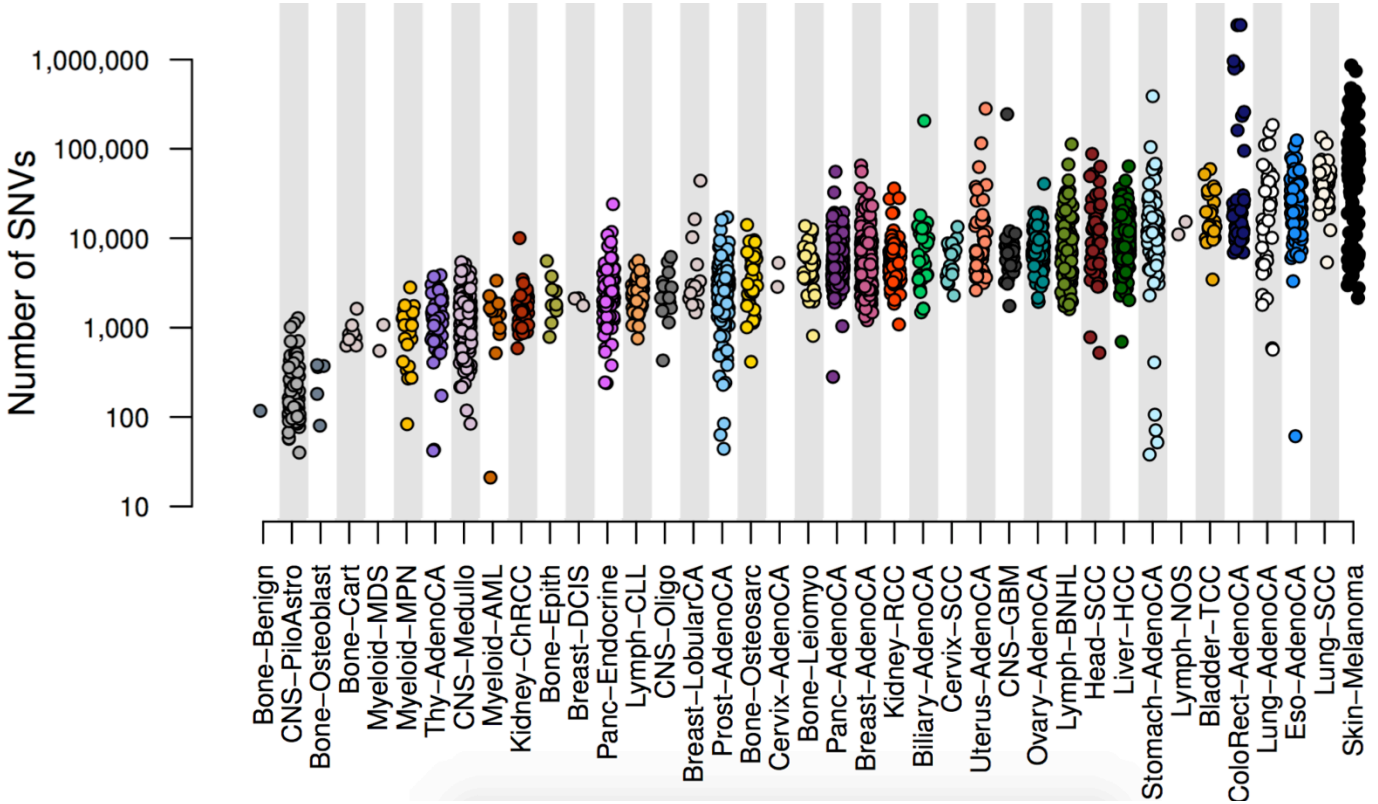


M: Mutation spectrum
observed

$$M = S \times W + \varepsilon$$

[T. Helleday, S. Eshtad & S. Nik-Zainal,
Nature Reviews Genetics ('14), L. Alexandrov et al., Nature ('13)]

PCAWG : most comprehensive resource for cancer whole genome analysis



Adapted from Campbell et. al., bioRxiv ('17)

Project Goals:

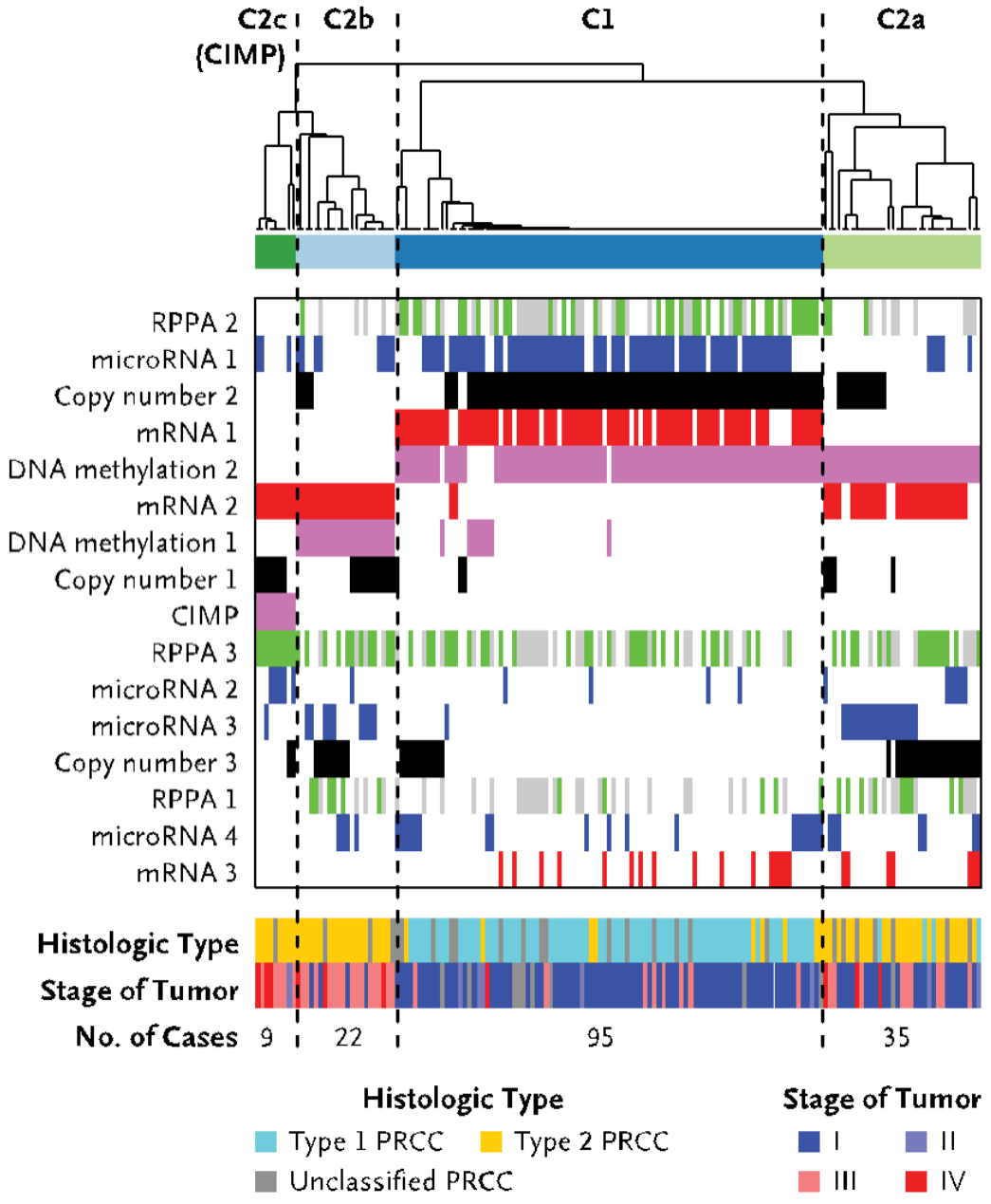
- To understand role of non-coding regions of cancer genomes in cancer progression.
- Union of TCGA-ICGC efforts

- Jointly analyzing ~2800 whole genome tumor/normal pairs
  - > 580 researchers
  - 16 thematic working groups
  - ~30M total somatic SNVs

PCAWG
PanCancer Analysis
OF WHOLE GENOMES

A case study: pRCC

- Kidney cancer lifetime risk of 1.6% & the papillary type (pRCC) counts for ~10% of all cases
- TCGA project sequenced 161 pRCC exomes & classified them into subtypes
- Also, 35 WGS of TN pairs



[Cancer Genome Atlas Research Network N Engl J Med. ('16)]

Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

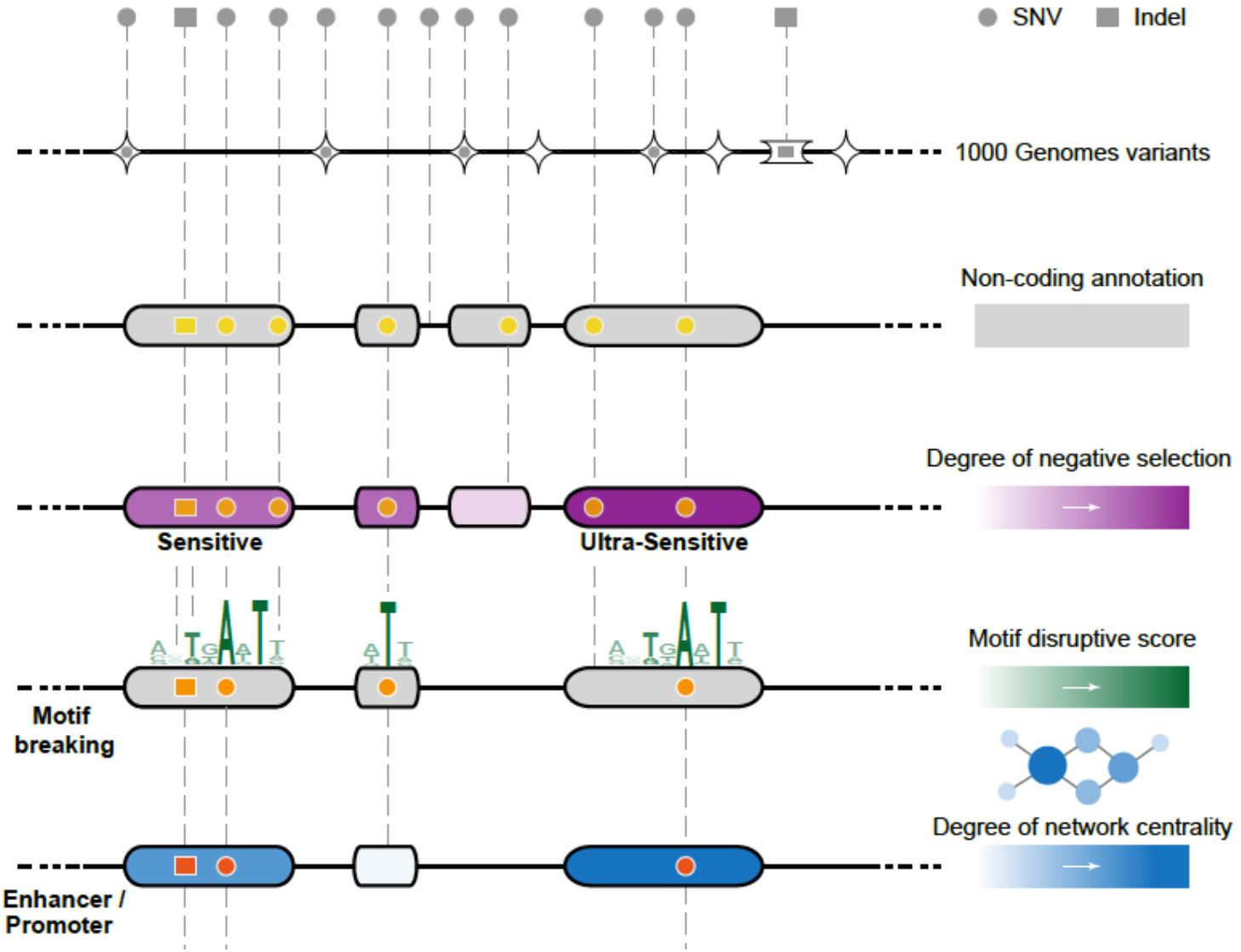
Funseq: a flexible framework to determine functional impact & use this to prioritize variants

Annotation (tf binding sites open chromatin, ncRNAs) & Chromatin Dynamics

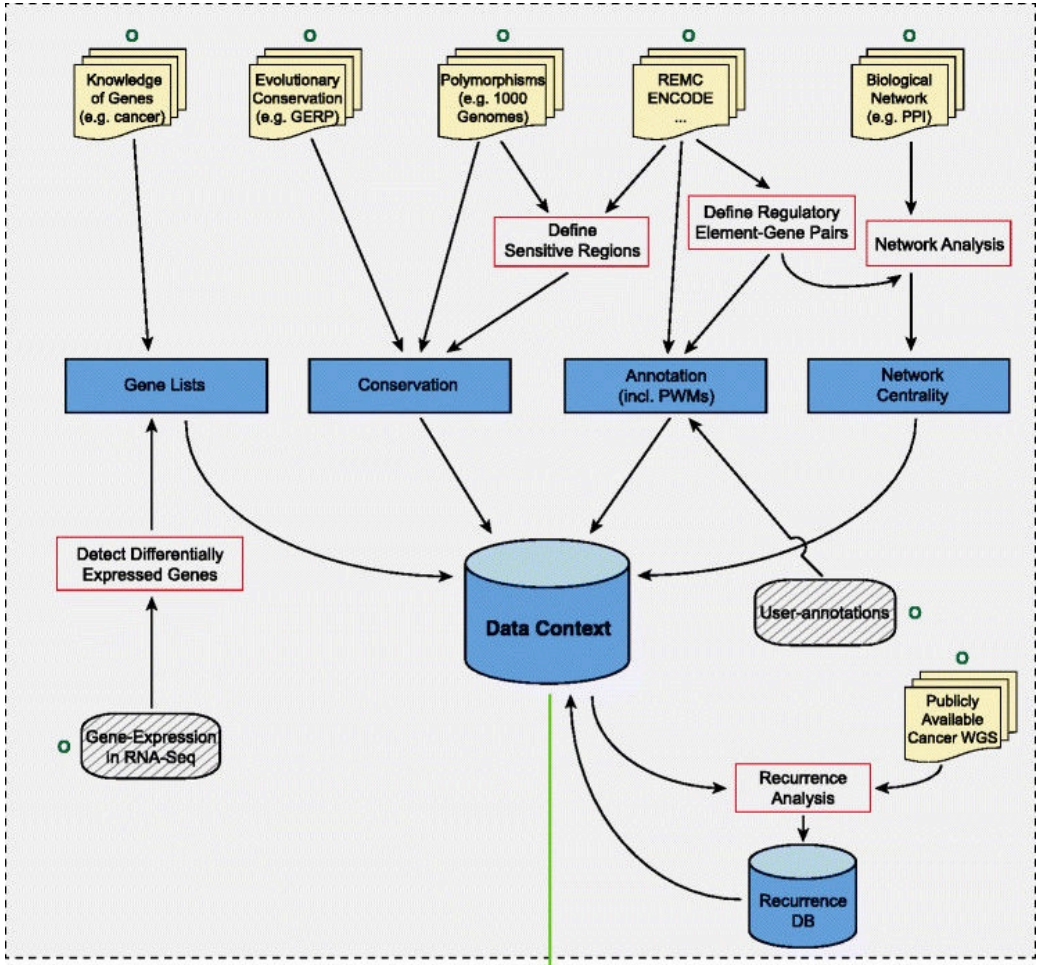
Conservation (GERP, allele freq.)

Mutational impact (motif breaking, Lof)

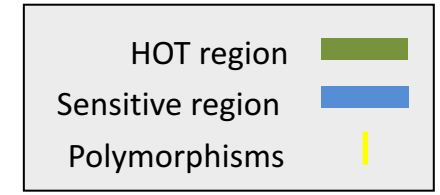
Network (centrality position)



[Fu et al., GenomeBiology ('14), , Khurana et al., Science ('13)]



FunSeq.gersteinlab.org

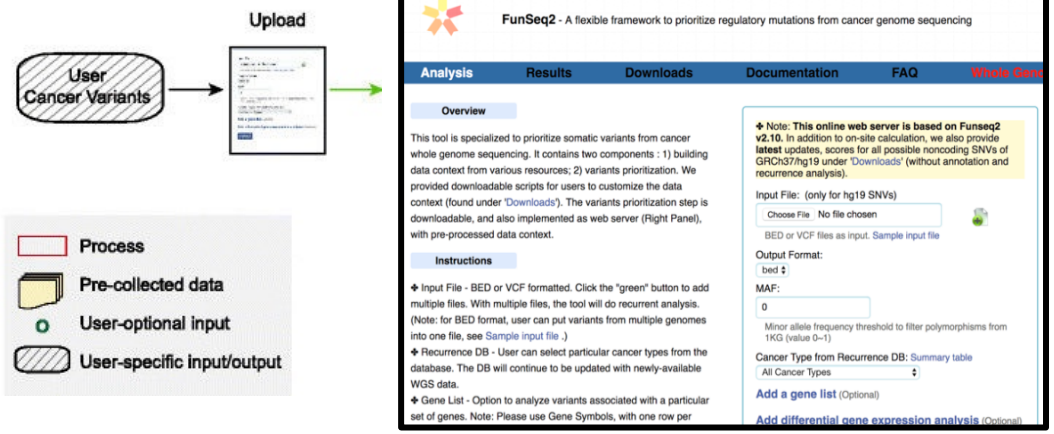


Genome

$$w_d = 1 + p_d log_2 p_d + (1 - p_d) log_2 (1 - p_d)$$

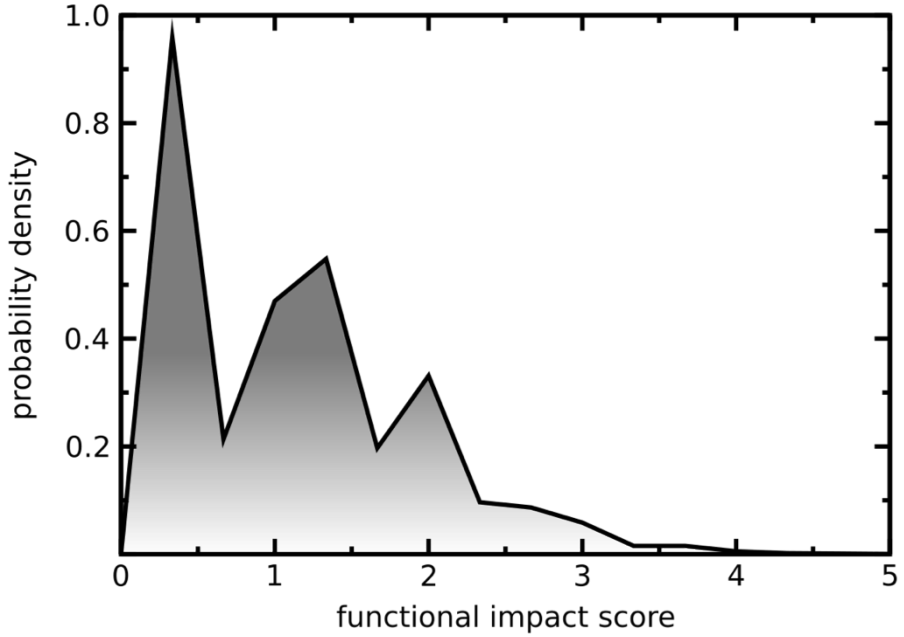
- Entropy based method for weighting consistently many genomic features

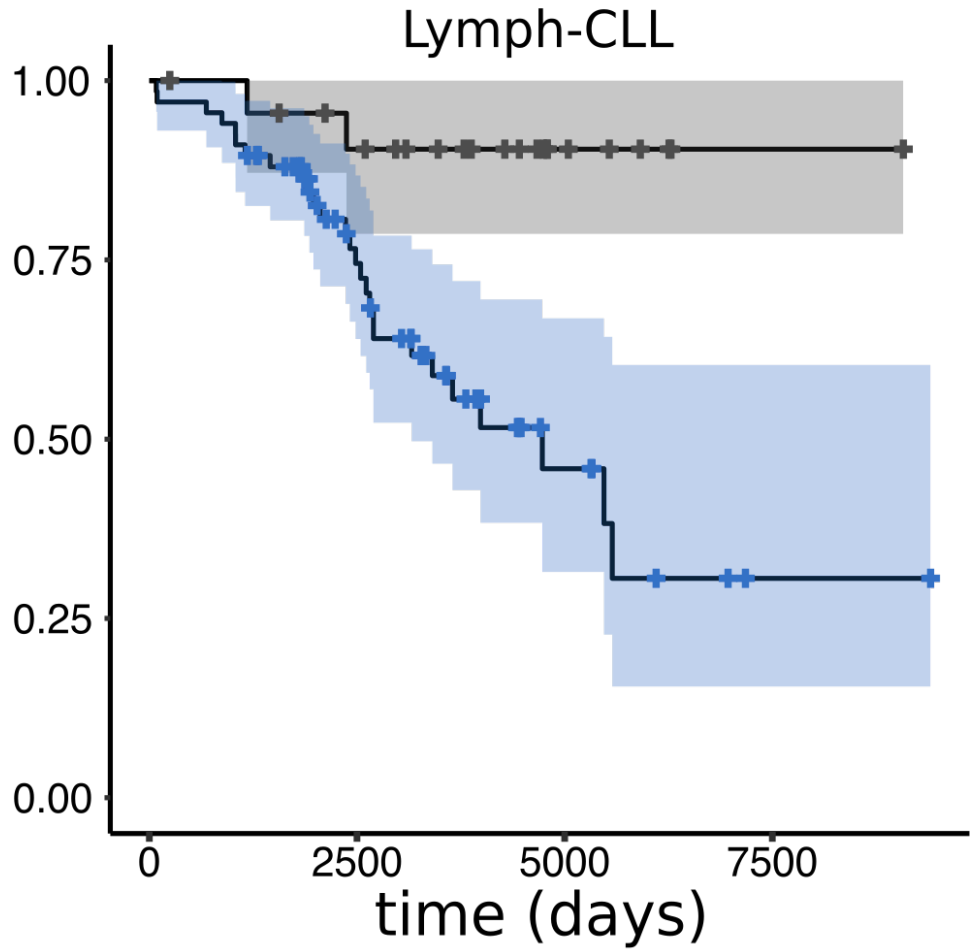
- Practical web server
- Submission of variants & pre-computed large data context from uniformly processing large-scale datasets



[Fu et al., GenomeBiology ('14)]

Overall functional impact distribution of PCAWG mutations

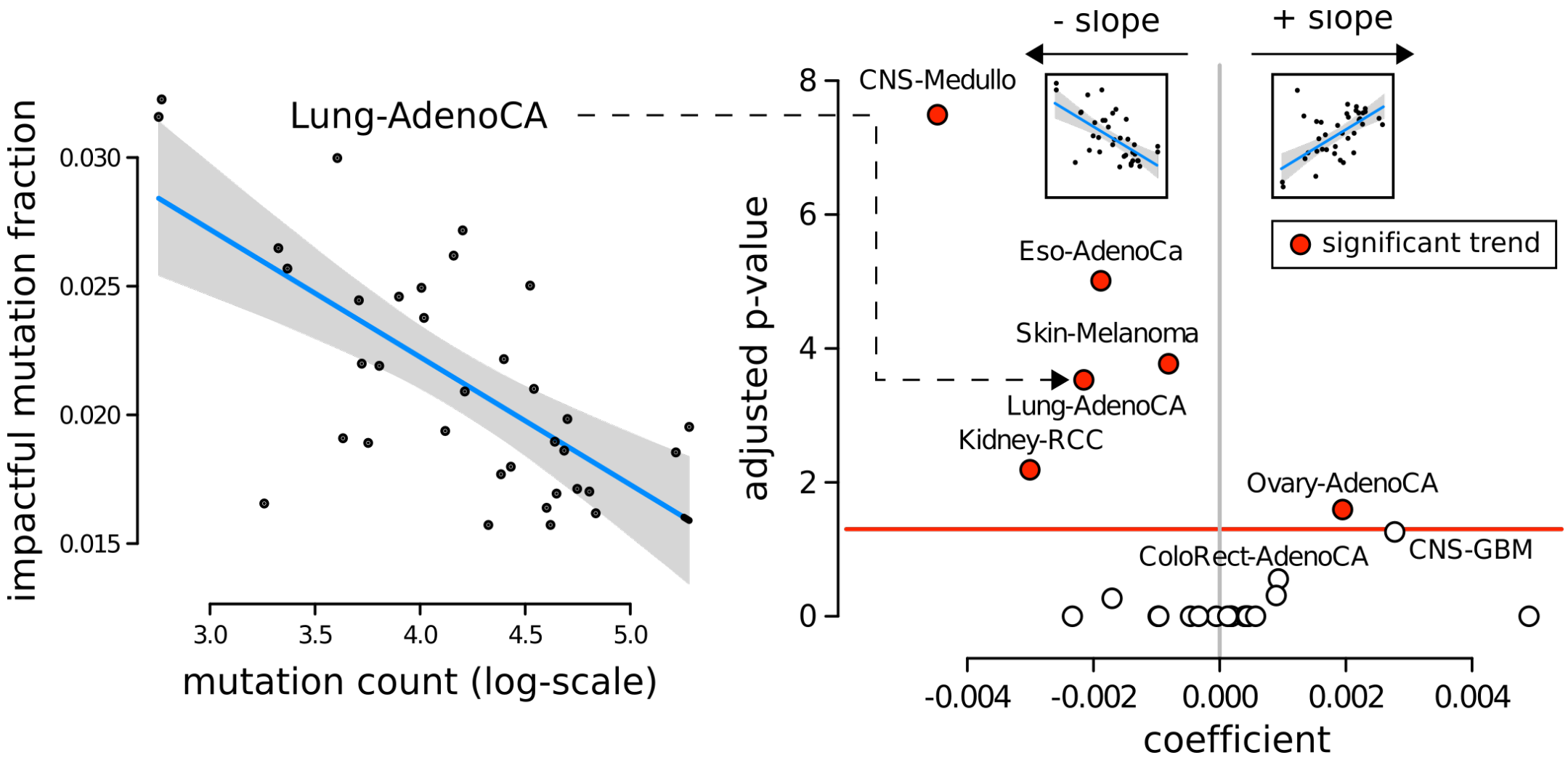




- Funseq molecular functional impact of ~30M variants in >2500 PCAWG samples

Division of PCAWG Lymph-CLL cohort based on average impact of non-driver variants (high v **low**) [A result of selection?]

In many PCAWG cohorts, the fraction of impactful "passengers" decreases with increase in total mutation burden (A result of selection?)

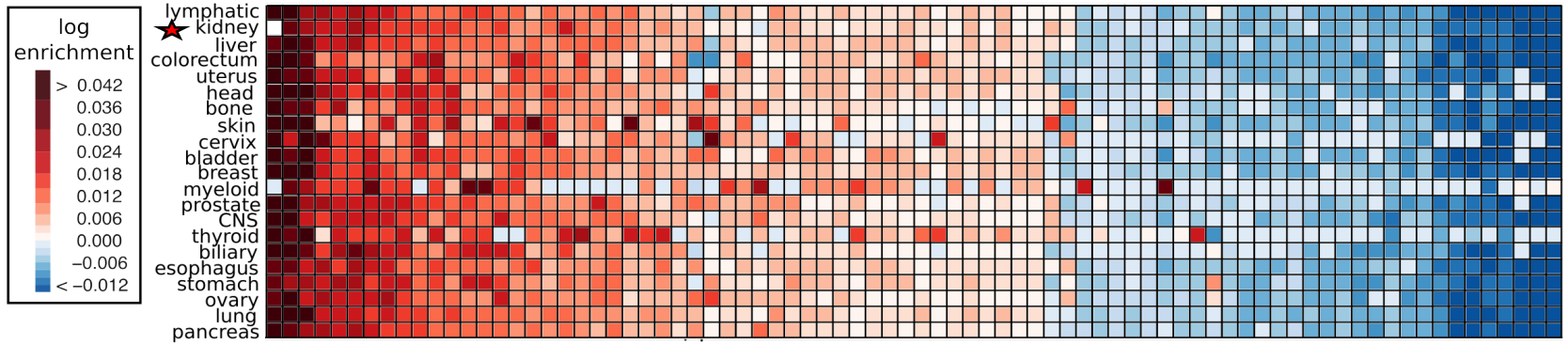


Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

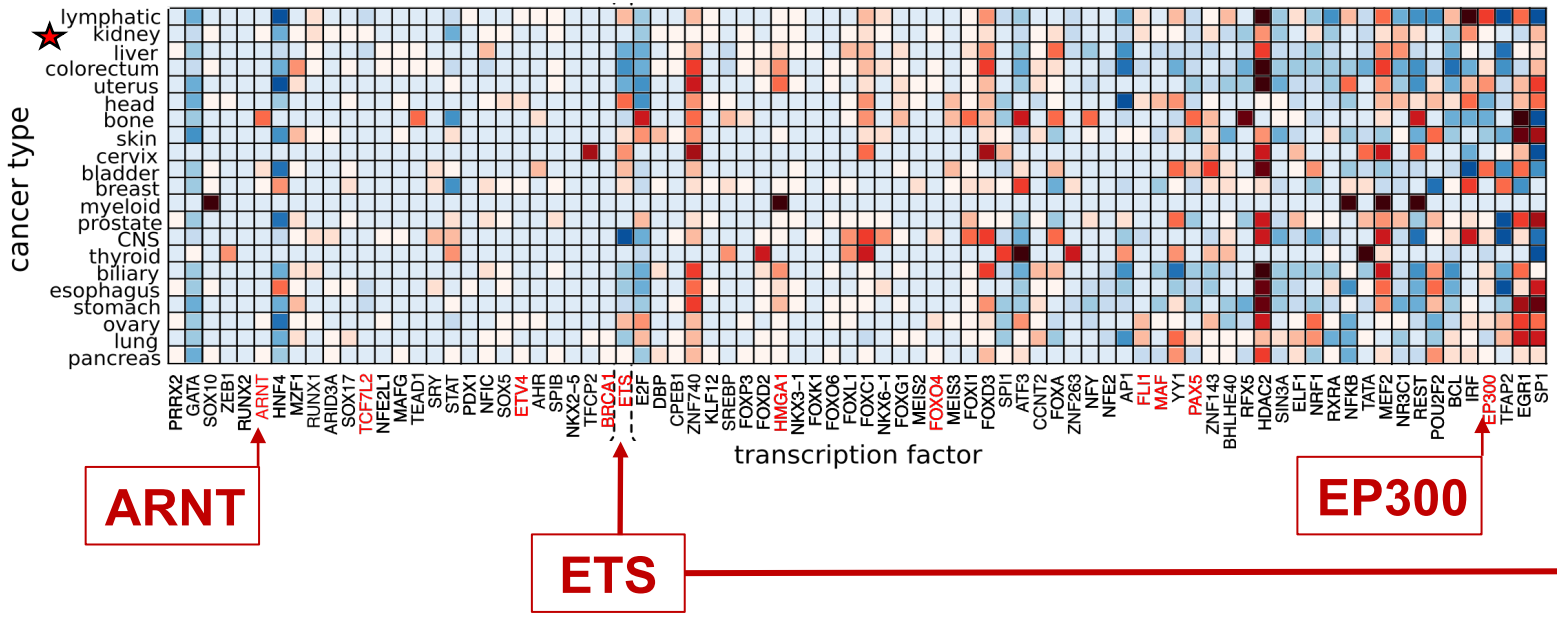
- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

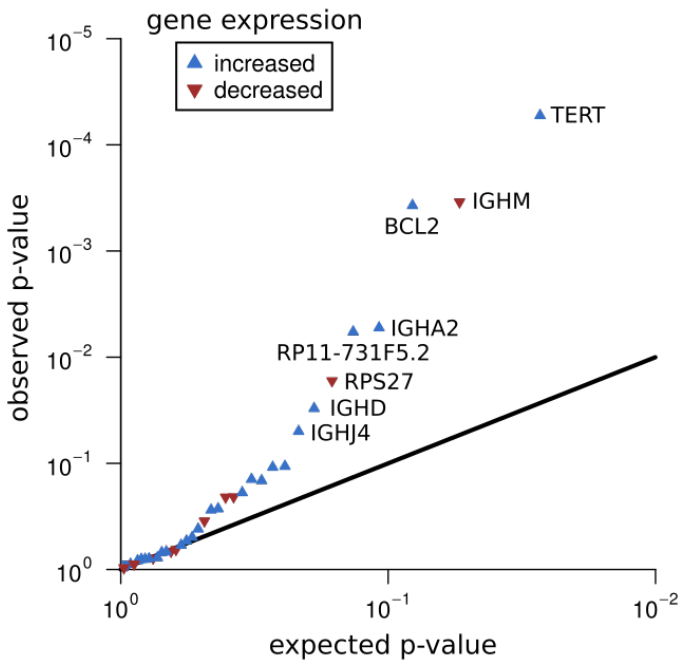
Differential Mutational burdening of TF-subnetworks due to SNVs breaking & creating binding sites

GAIN



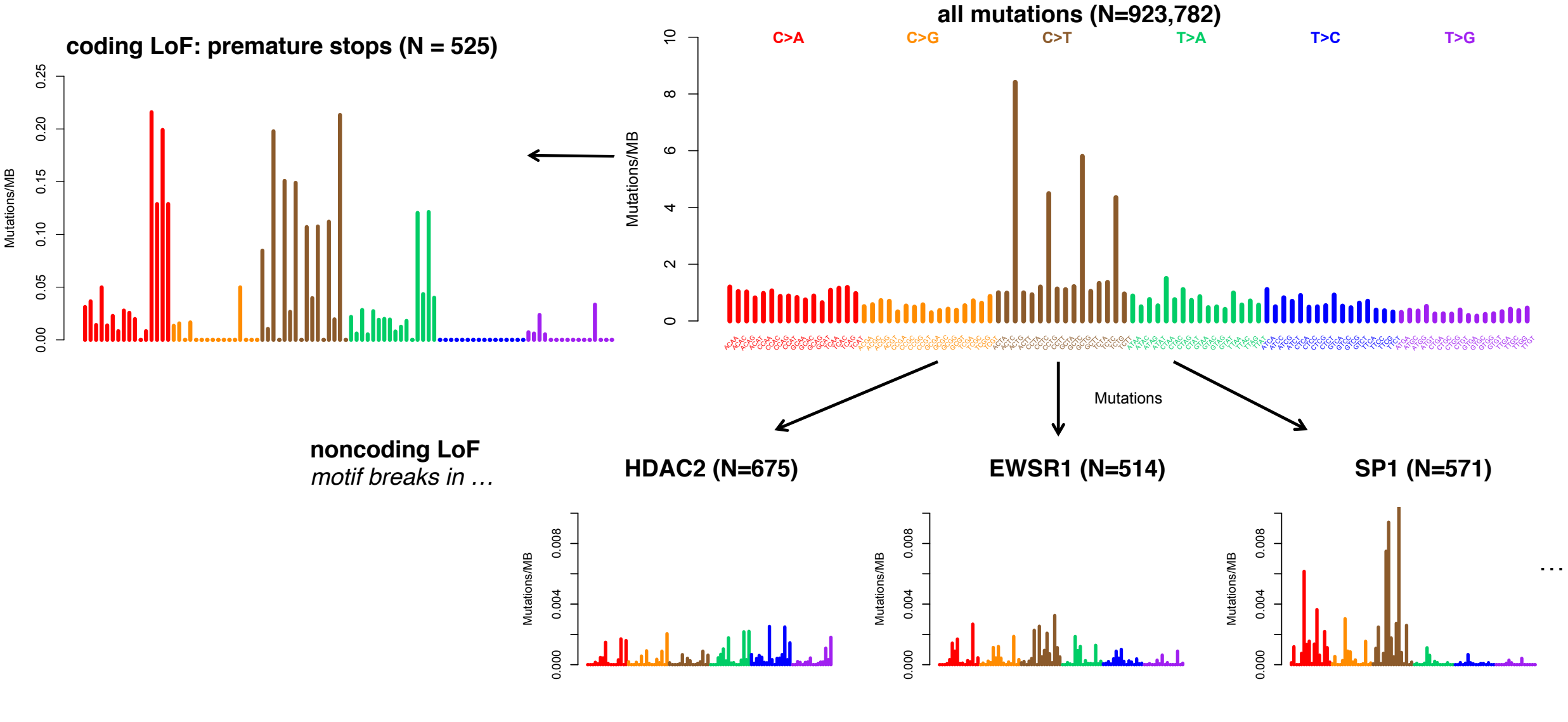
LOSS





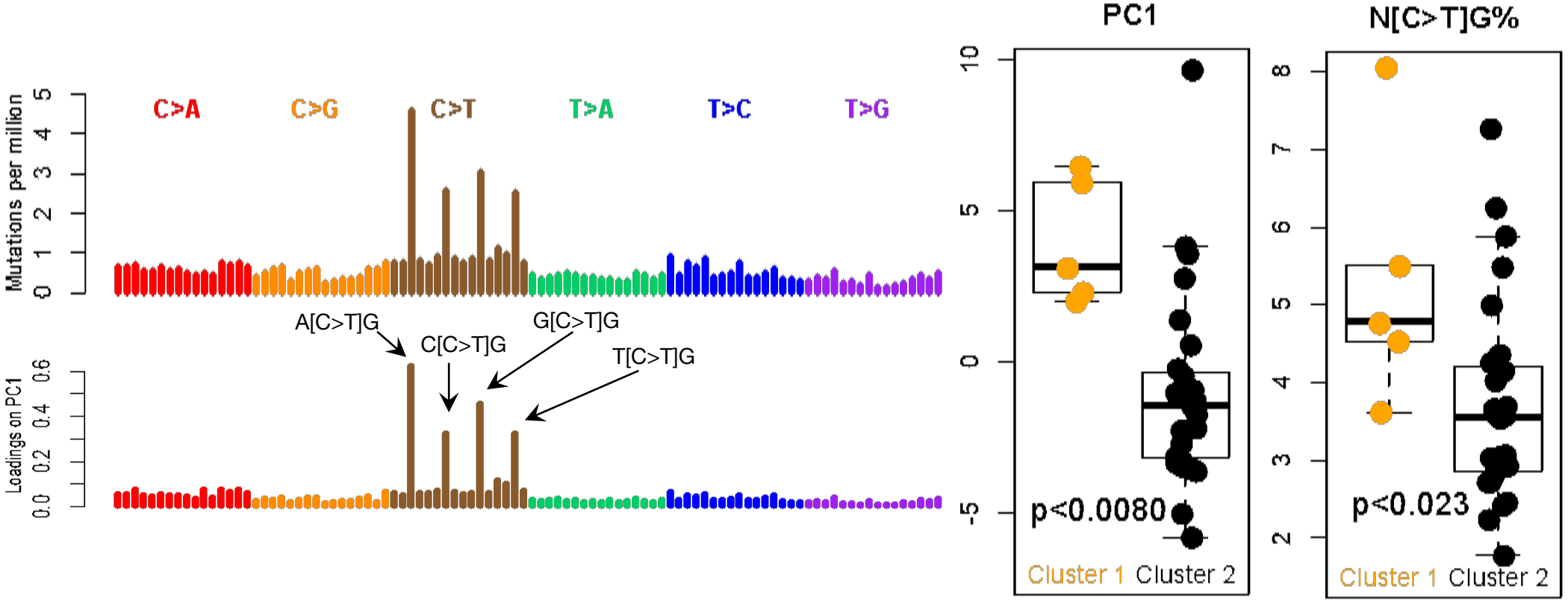
ETS regulated genes

Kidney cancer as an example: differential burdening correlates with mutational spectrum

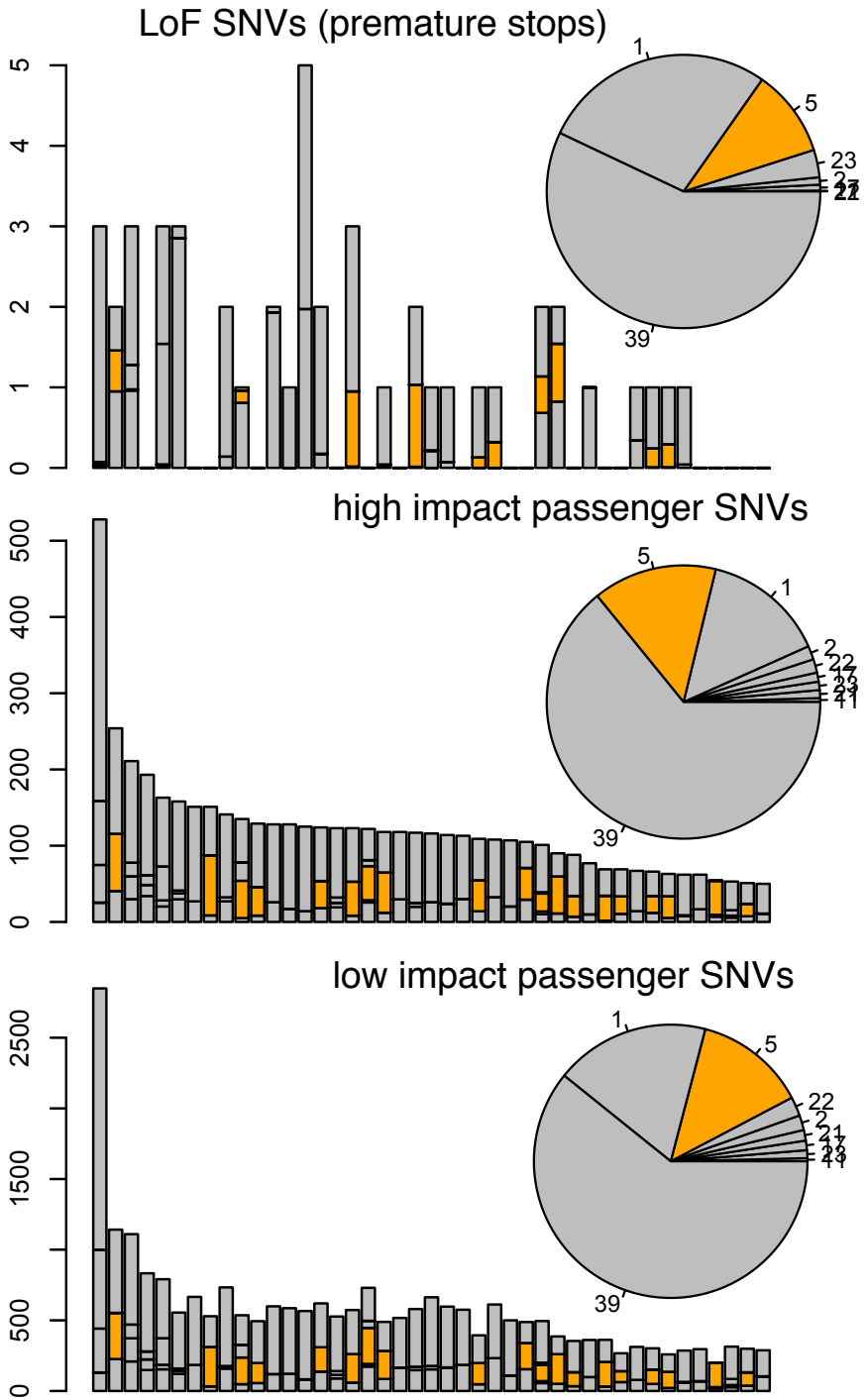


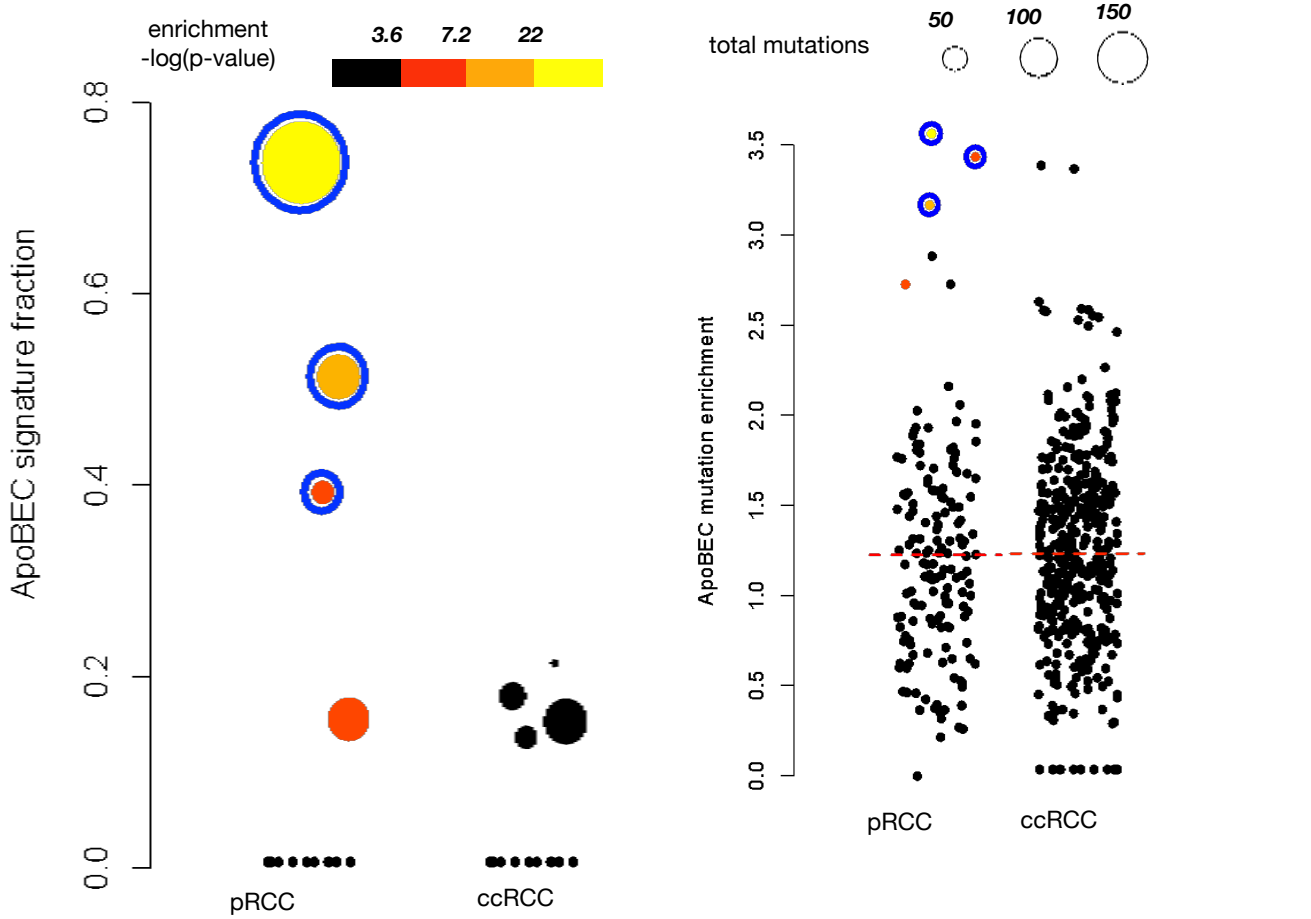
CpGs drive inter-patient variation in pRCC mutational spectra

- The loadings on PC1 are mostly [C>T]G
- Confirmed by higher C>T% in CpGs in the hypermethylated group (cluster1)



[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]

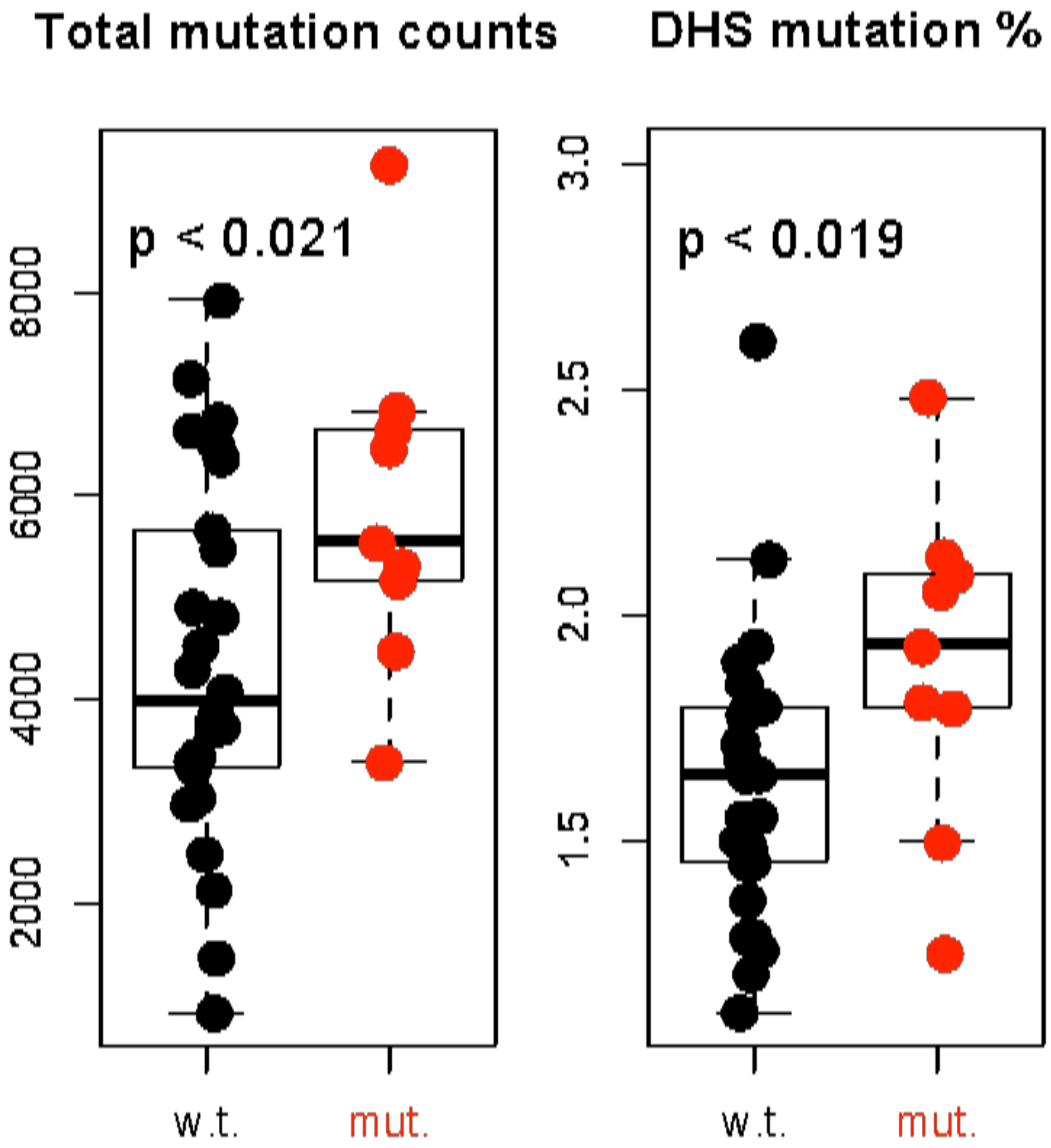




Signatures in pRCC

- Signatures burden the genome disproportionally
- We found 1 pRCC has ApoBEC signature, but nothing in a larger ccRCC cohort

[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]



Key mutation affects mutational landscape which, in turn, affects overall burden in pRCC

- Chromatin remodeling defect ("mut") leads to more mutations in open chromatin (raw number & fraction) in those pRCC cases w/ the mutation

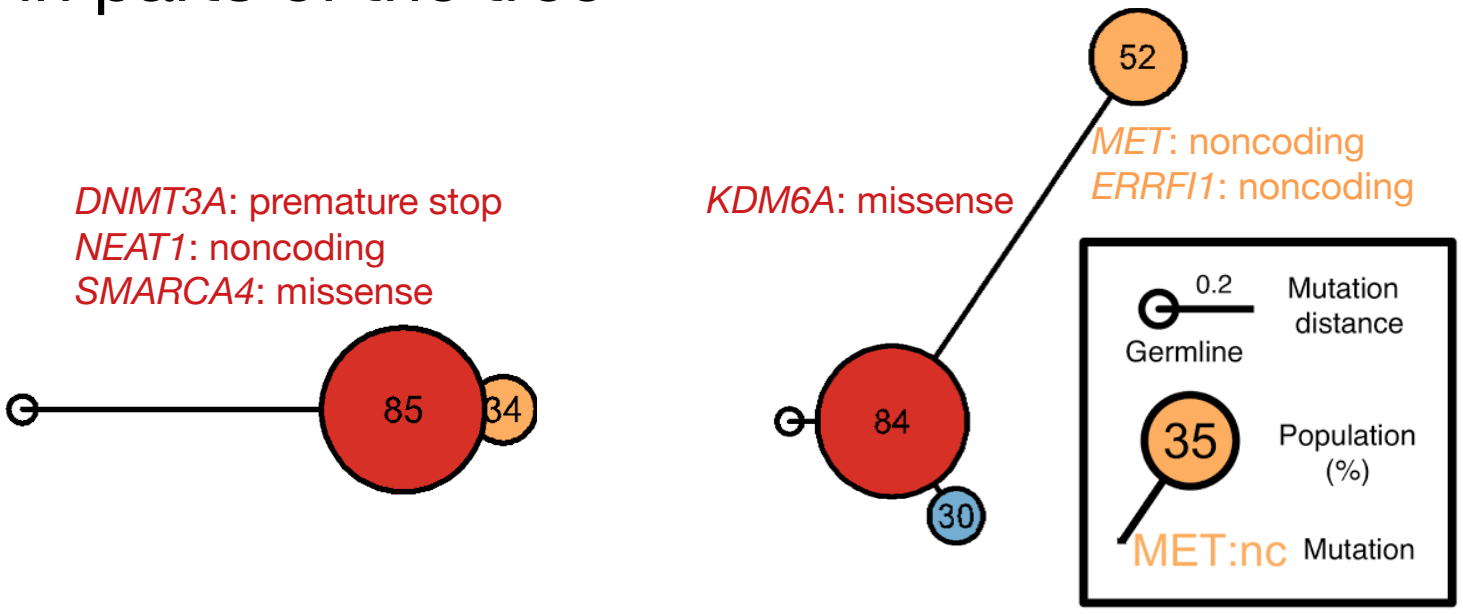
[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]

Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

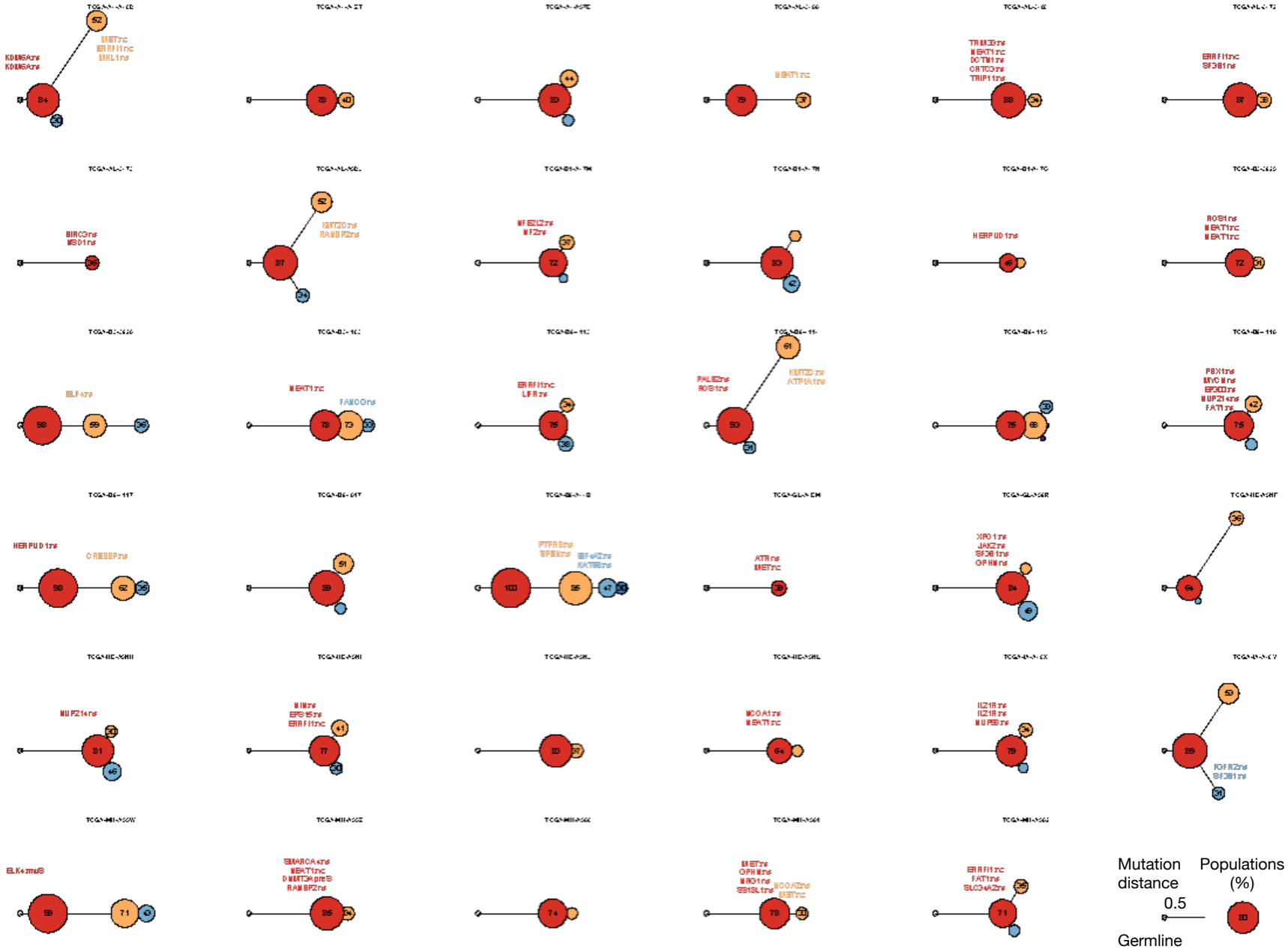
- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

Constructing evolutionary trees in pRCC

- Infer mutation order (eg early v late) & tree toplogy based on mutation abundance (PhyloWGS, Deshwar et al., 2015)
- Some key mutations occur in all the clones while others are just in parts of the tree



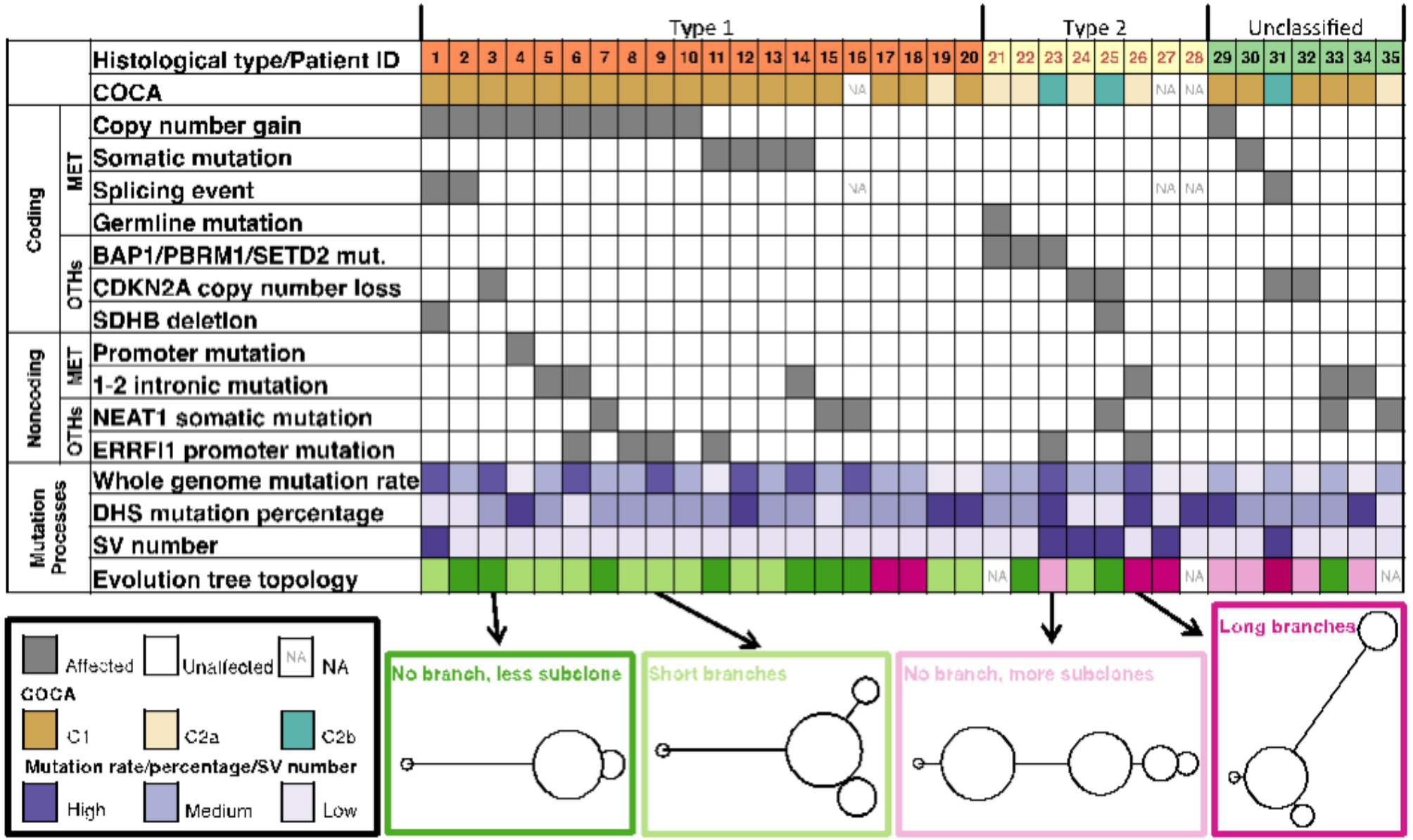
[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]



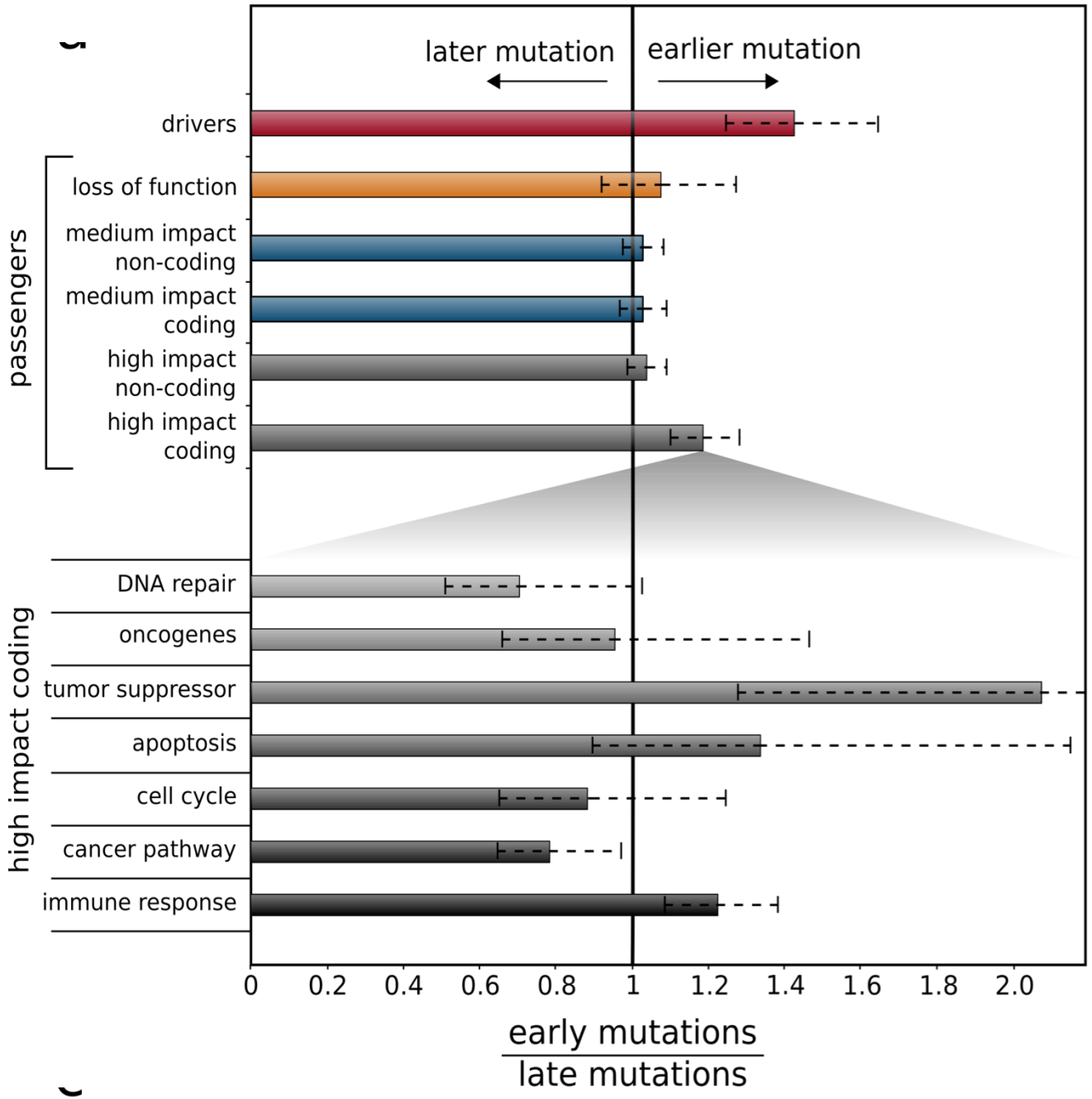
[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]

[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]

Tree topology correlates with molecular subtypes



[Li et al., PLOS Genetics ('17)]

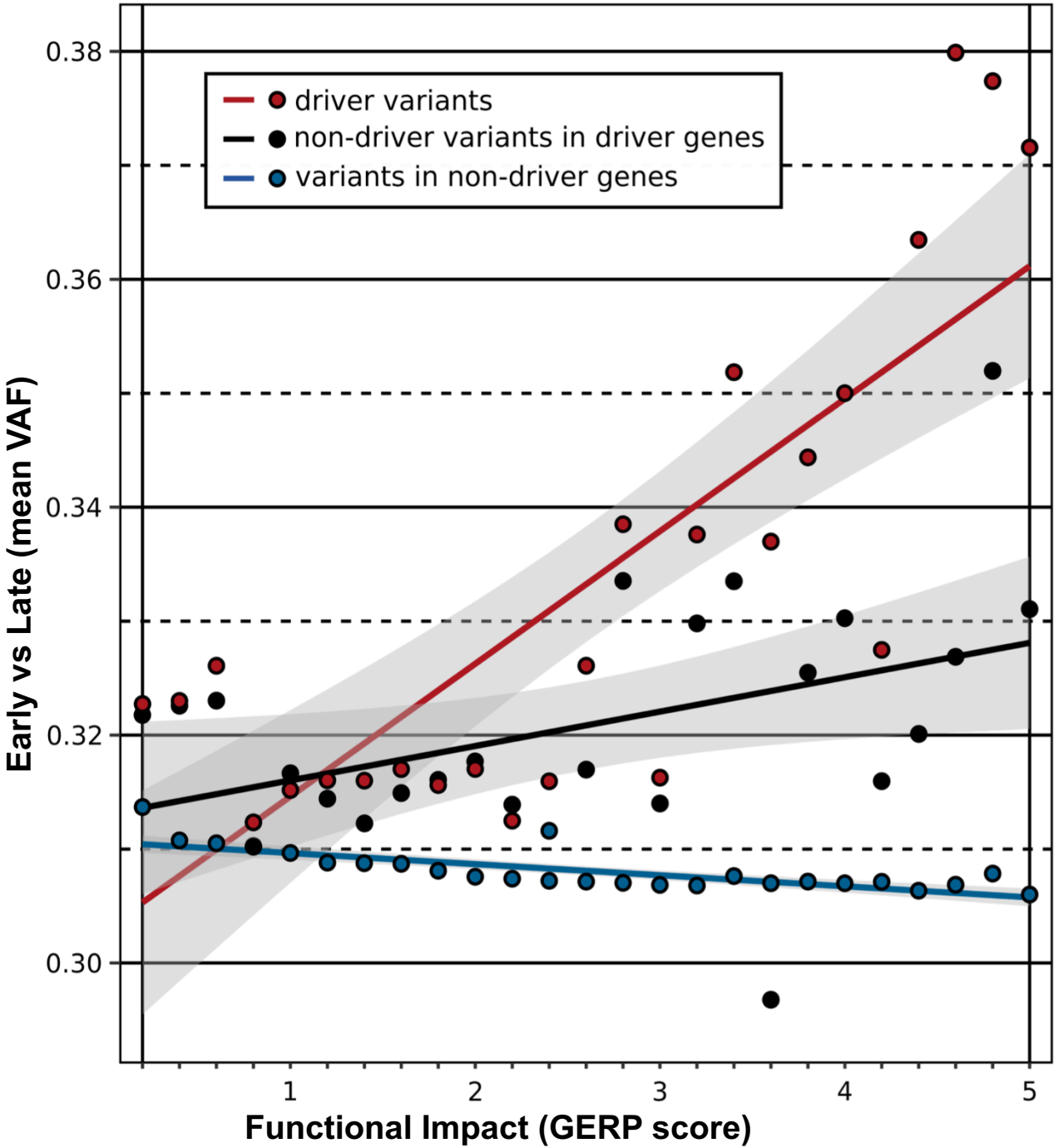


Sub-clonal architecture of mutations in PCAWG

As expected, drivers are enriched in earlier subclones. Overall, no such enrichment among passengers.

High impact passengers are slightly enriched among early subclones (weak drivers?)

Particularly, passengers in tumor suppressor (in contrast to oncogenes, which require specific mutations).



Continuous correlation of functional impact & VAF

Among mutations in driver genes:
higher impact mutation

Still true after removing all known driver variants from driver genes.
(Latent drivers?)

Outside driver genes:
higher impact mutation
(Deleterious passengers?)

Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

Passenger mutations in >2500 cancer genomes: Overall molecular functional impact

- **Introduction**
  - Background: driver-passenger model (w/ conceptual extension) + mutational spectra & signatures
  - Data source: PCAWG comprehensive WGS on >2.5K + focus on 35 pRCC WGS
- **Overall functional impact of variants**
  - FunSeq entropy-weights multiple features to evaluate the functional impact of SNVs
  - Investigating how the fraction of high-impact (non-strong-driver) SNVs scales & how it relates to survival
- **Differential burdening from various mutational processes**
  - Diff. burdening of TF sub-networks results from spectra & signatures differentially affecting binding motifs
  - High & low impact mutations assoc. w/ diff. signatures
  - Number of mutations in DHSes assoc. w/ specific chromatin mod. mutation
- **Functional impact & tumor evolution**
  - Mutational timing & tree topology classifies pRCC subtypes
  - Differences in functional impact betw. early & late passenger mutations (eg in TSGs & oncogenes)

Acknowledgements

PanCancer.info Functional impact

S **Kumar,** J **Warrell**, W Meyerson, P McGillivary, L Salichos, S Li, A Fundichely, E Khurana, C Chan, M Nielsen, C Herrman, A Harmanci, L Lochovsky,Y Zhang, X Li,

PCAWG Drivers & Functional Interpretation Group

(leaders: G Getz, J Pedersen, J Stuart, B Rapheal, N Lopez Bigas, D Wheeler), ICGC/TCGA PCAWG Network

FunSeq.gersteinlab.org

Y **Fu**, E **Khurana**, Z Liu, S Lou, J Bedford, XJ Mu, KY Yip

pRCC

S **Li**, B Shuch

Hiring Postdocs, See **JOBS**.gersteinlab.org

Info about this talk

General PERMISSIONS

- This Presentation is copyright Mark Gerstein, Yale University, 2016.
- Please read permissions statement at
  gersteinlab.org/misc/permissions.html .
- Feel free to use slides & images in the talk with PROPER acknowledgement (via citation to relevant papers or link to appropriate site).
  Paper references in the talk probably from Papers.GersteinLab.org.

PHOTOS & IMAGES

For thoughts on the source and permissions of many of the photos and clipped images in this presentation see streams.gerstein.info . In particular, many of the images have particular EXIF tags, such as kwpotppt , that can be easily queried from flickr, viz:
flickr.com/photos/mbgmbg/tags/kwpotppt