Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

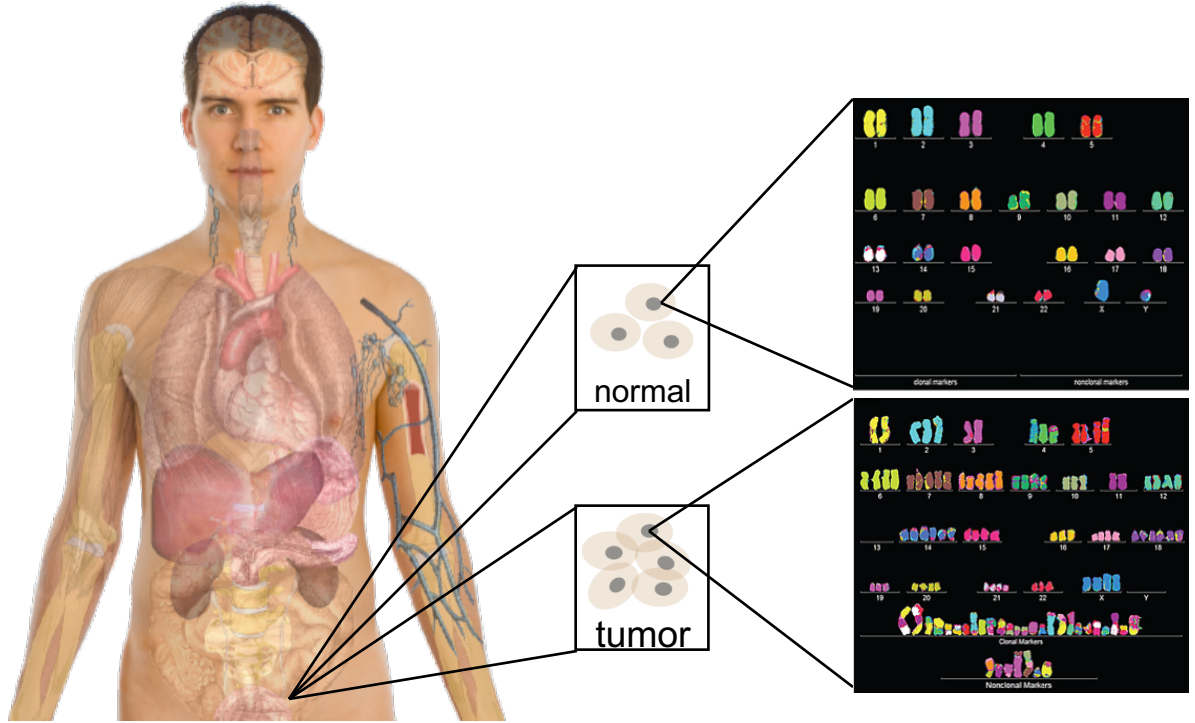
Mark Gerstein
Yale

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

See last slide for more info.

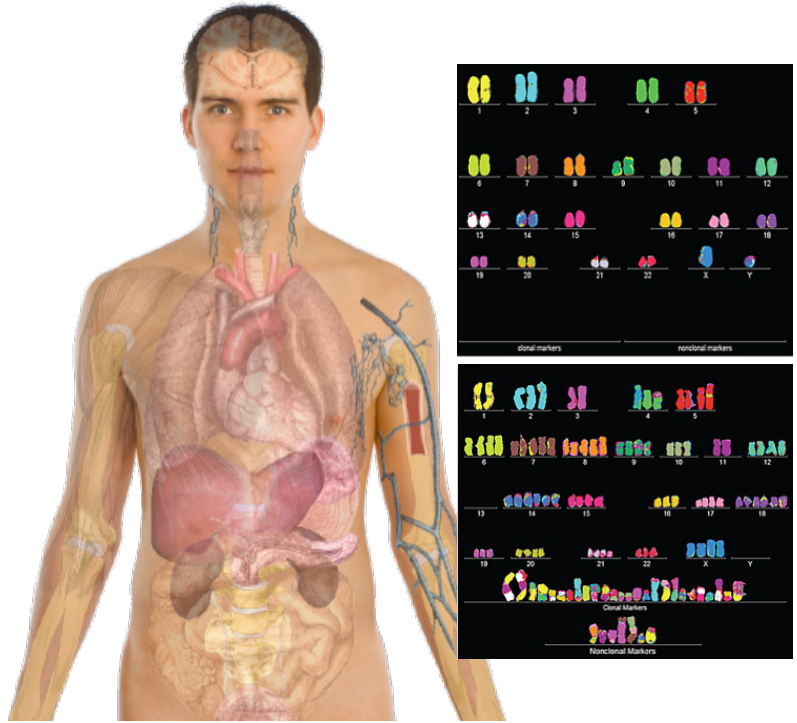
Personal Genomics as a Gateway into Biology

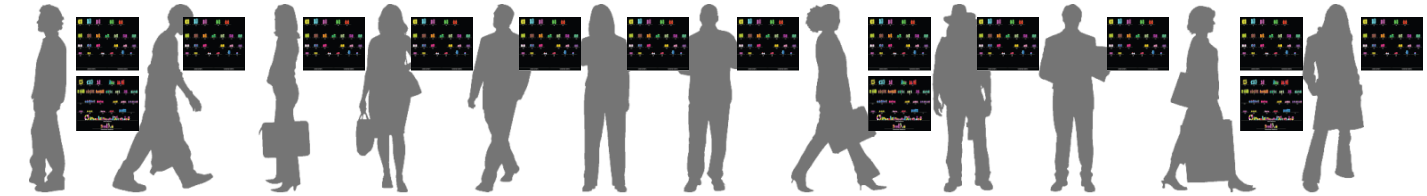
Personal genomes will soon become a commonplace part of medical research & eventually treatment (esp. for cancer). They will provide a primary connection for biological science to the general public.

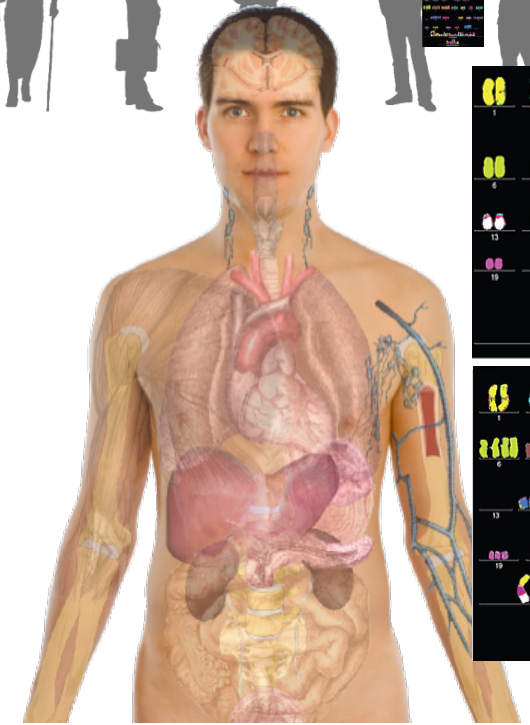


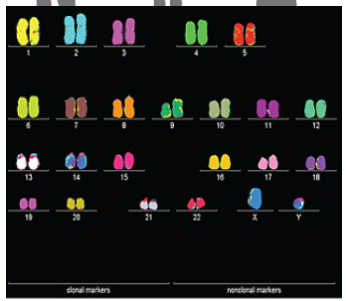
Personal Genomics as a Gateway into Biology

Personal genomes will soon become a commonplace part of medical research & eventually treatment (esp. for cancer). They will provide a primary connection for biological science to the general public.





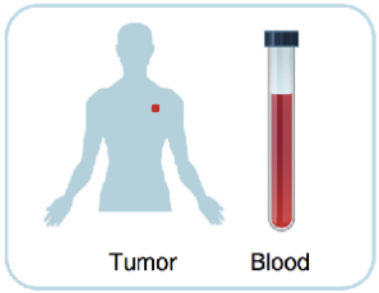



clonal markers
nonclonal markers

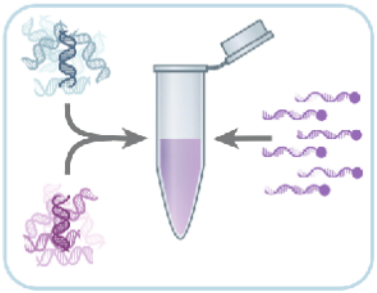

Clonal Markers
Nonclonal Markers

Key variants will increasingly play essential roles in precision medicine

1. General diagnosis

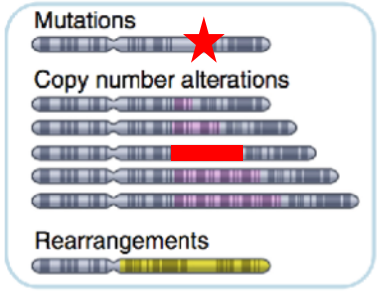


2. Sample extraction



3. Sample preparation

4. Sequencing



5. Analysis

6. Review

Database of variants

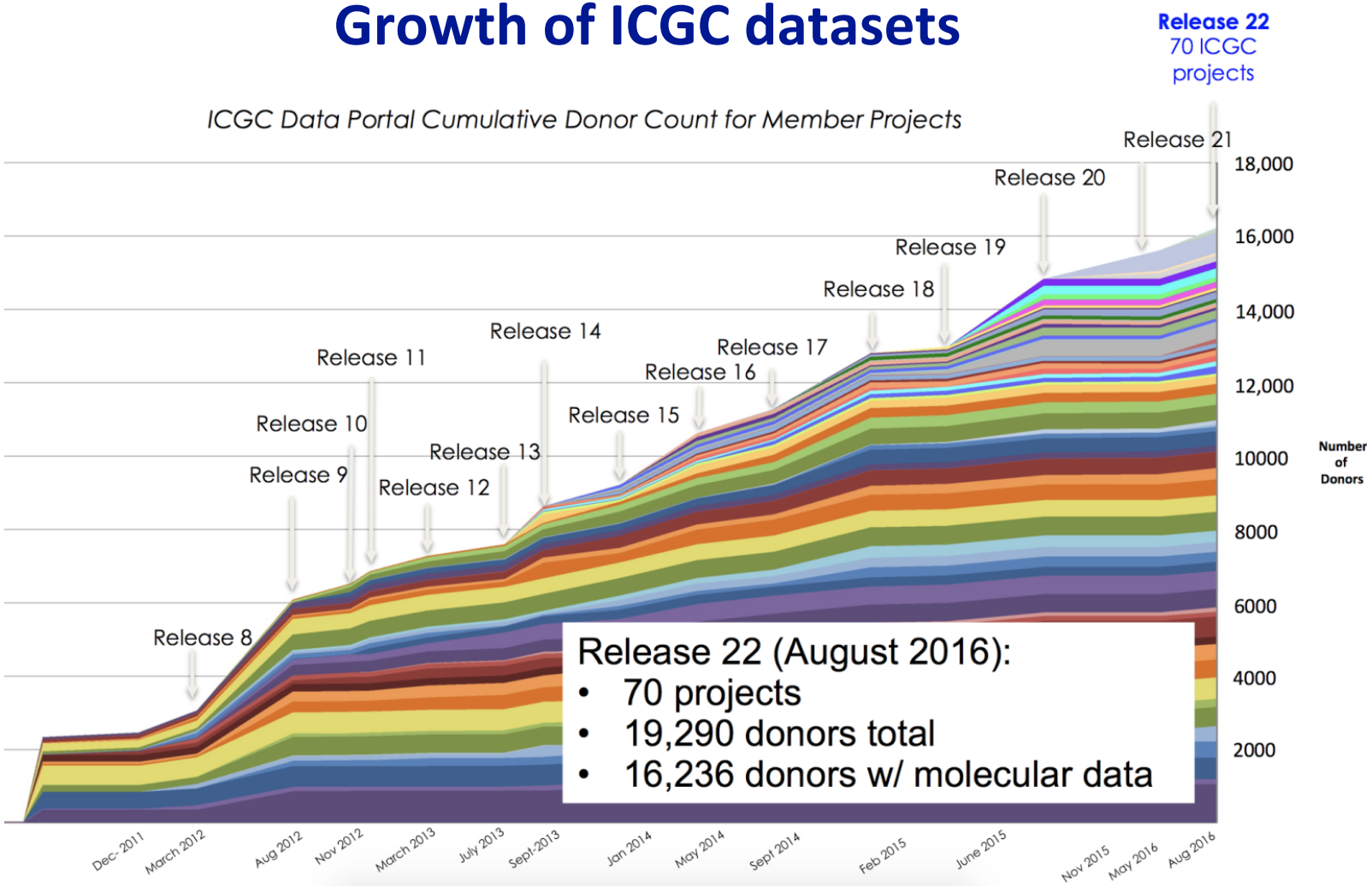
Clinical report

Trial matching

Refined diagnosis (ex: sub-cancer type)

Modified from A. Zehir et al, Nat. Med (2017)

Growth of ICGC datasets



Hardeep Nahal , 12th Scientific ICGC Workshop (Sept 2016)

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.

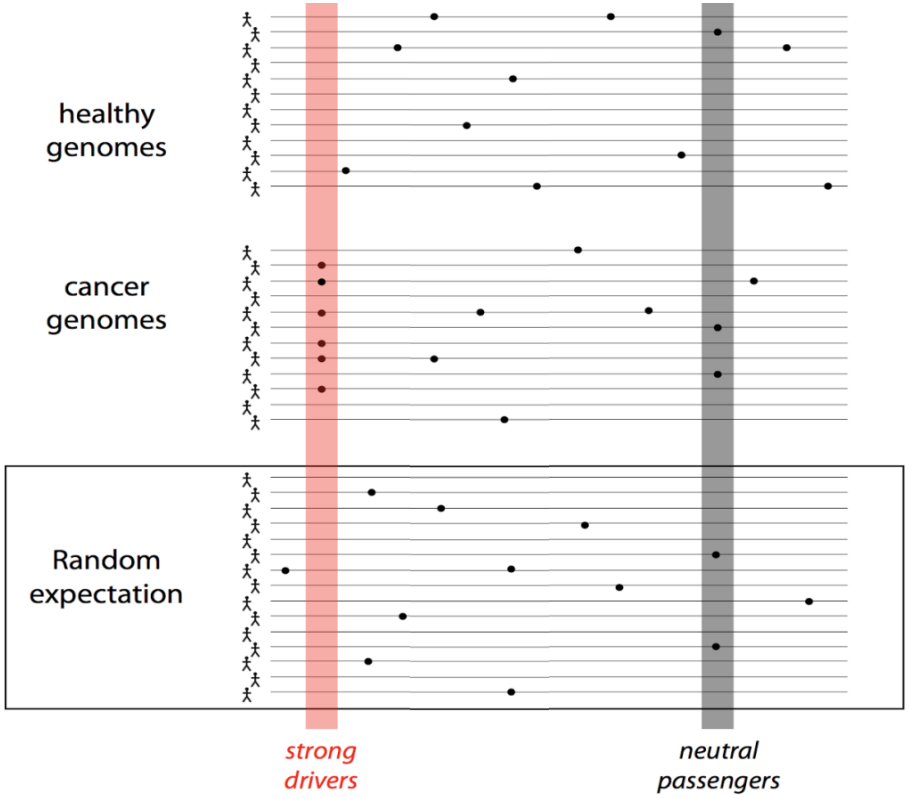
identified through signals of positive selection.

Existing cohorts of ~100s give enough power to identify

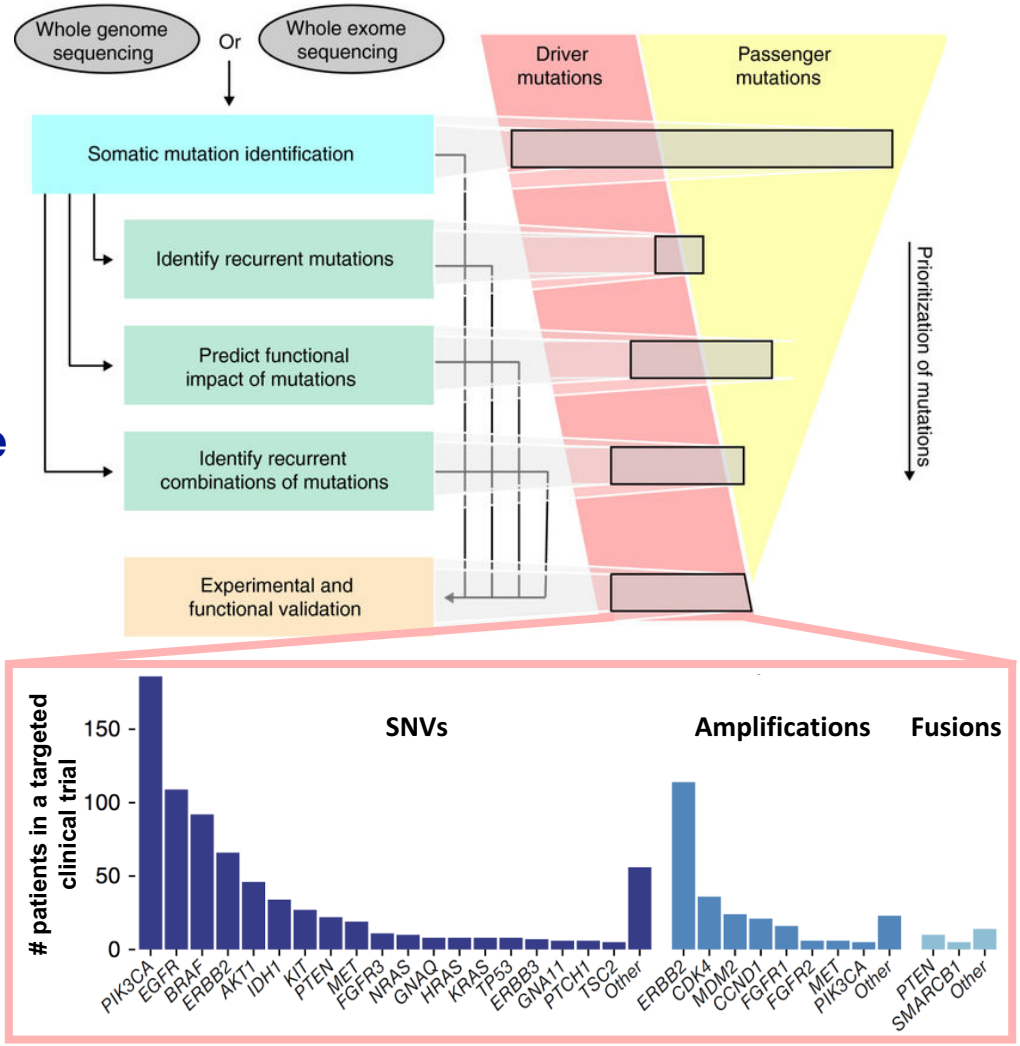
Passengers
Conceptually, a passenger mutation has no direct or indirect effect on tumor progression.

There are 1000s of passengers in a typical cancer genome.

[Vogelstein Science 2013. 339:1546]



Prioritizing key variants identifies drivers to better enable more precise diagnostics and targeted therapies



Identifying select driver variants from the large pool of candidate variants

Number of patients in matched clinical trials identified on the basis of actionable variants in different genes

Top: Raphael, et al., Genome Med. (2014)
Bottom: Modified from Zehir et al, Nat. Med (2017)

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Variant Annotation Tool (VAT)

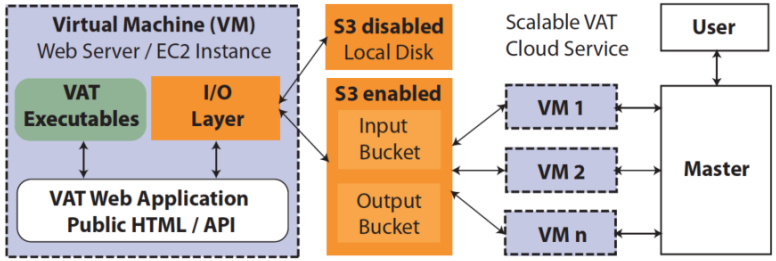
VCF Input

Output:

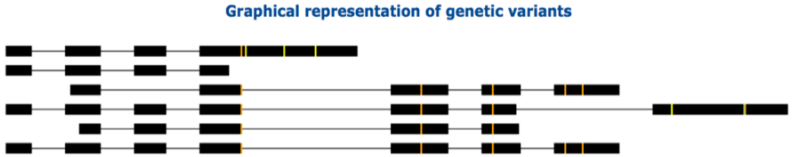
- Annotated VCFs
- Graphical representations of functional impact on transcripts

Access:

- Webserver
- AWS cloud instance
- Source freely available



CLOUD APPLICATION



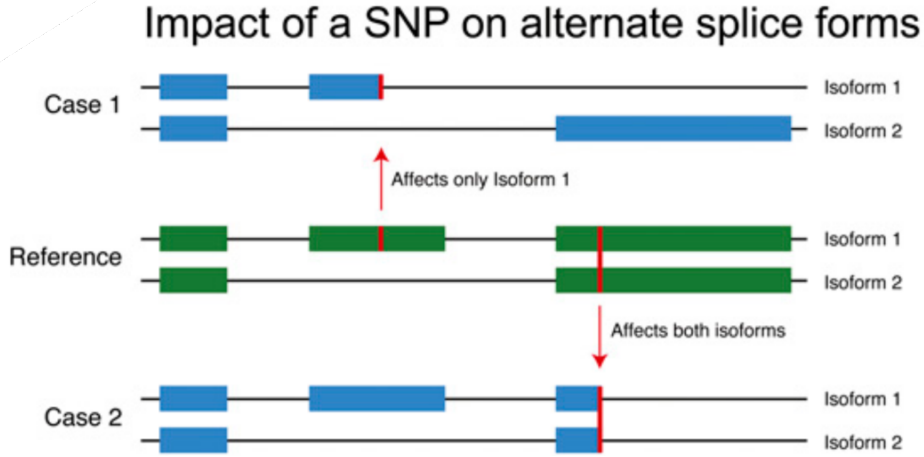
vat.gersteinlab.org

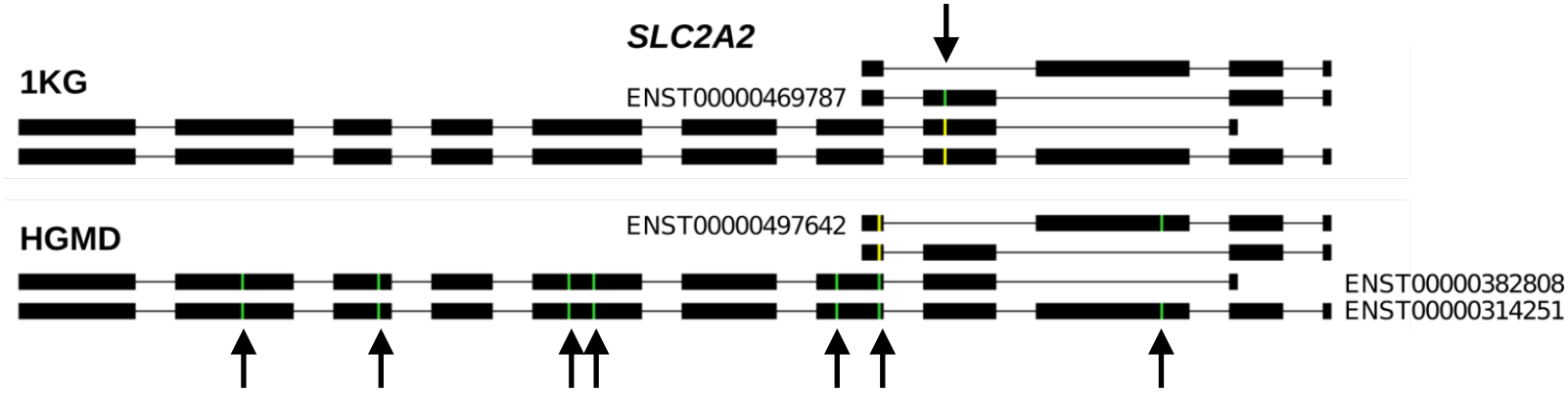
Habegger L., Balasubramanian S.*, et al. Bioinformatics, 2012*

Complexities in LOF annotation

Transcript isoforms,
distance to stop,
functional domains,
protein folding,
etc.

Balasubramanian S. et al., *Genes Dev.,* '11
Balasubramanian S.*, Fu Y.* et al., *NComms.*, '17





Annotation of Loss-of-Function Transcripts (ALoFT)

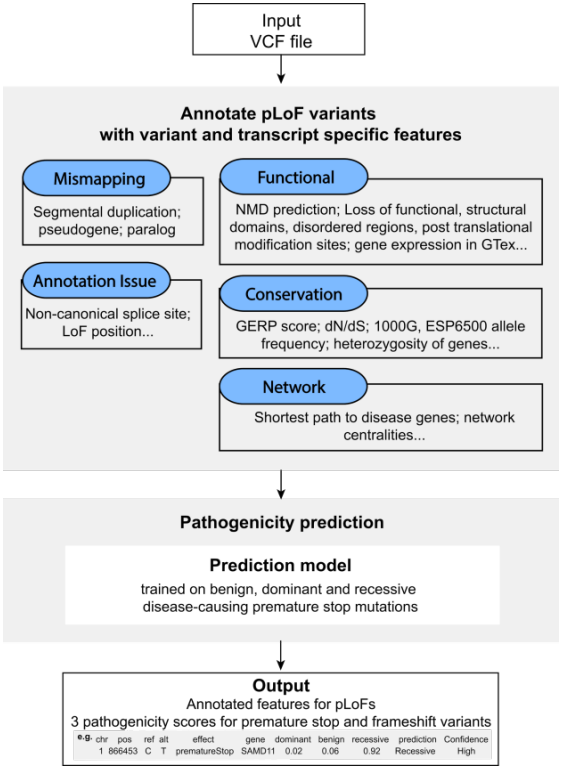
Runs on top of VAT

Output:

- Impact score: benign or deleterious.
- Confidence level.
- Annotated VCF.

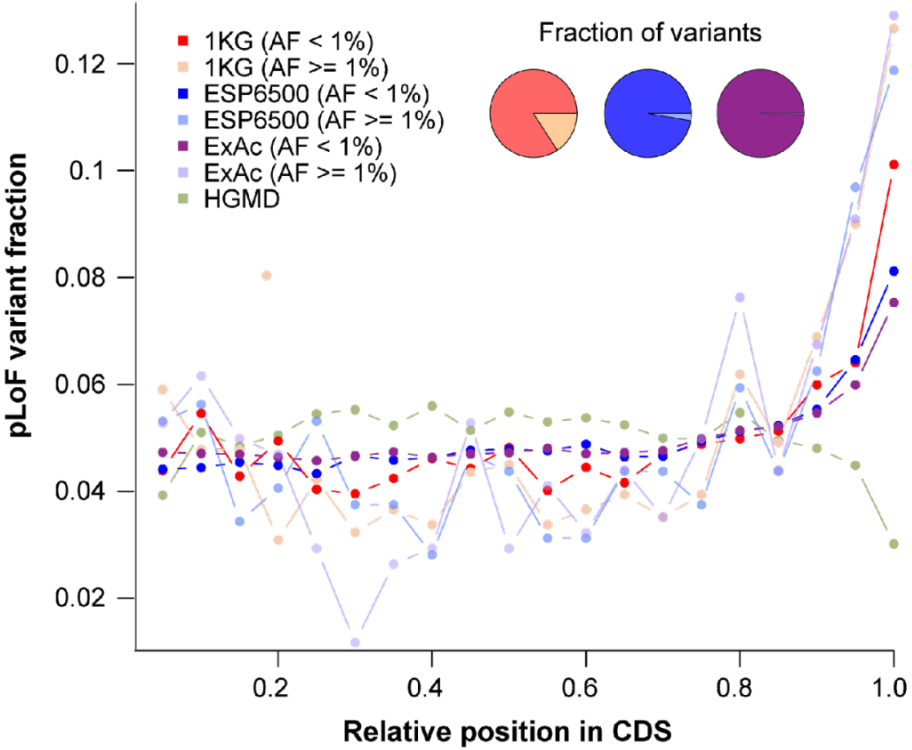
Access:

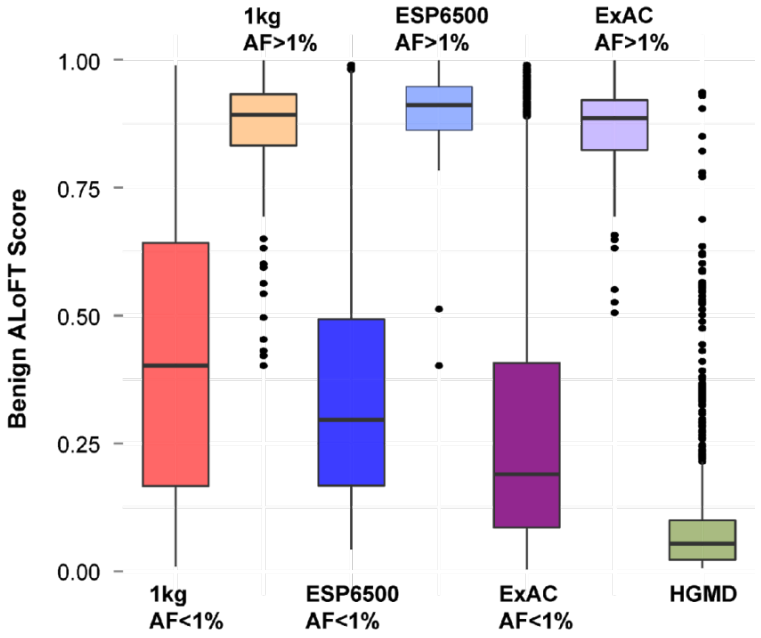
- Software package: **aloft.gersteinlab.org**
- GitHub: **github.com/gersteinlab/aloft**



Balasubramanian S.*, Fu Y.* et al., *NComms.*, '17

LoF distribution varies as expected by mutation set (from healthy people v from disease)





Balasubramanian S.*, Fu Y.* et al., *NComms.*, '17

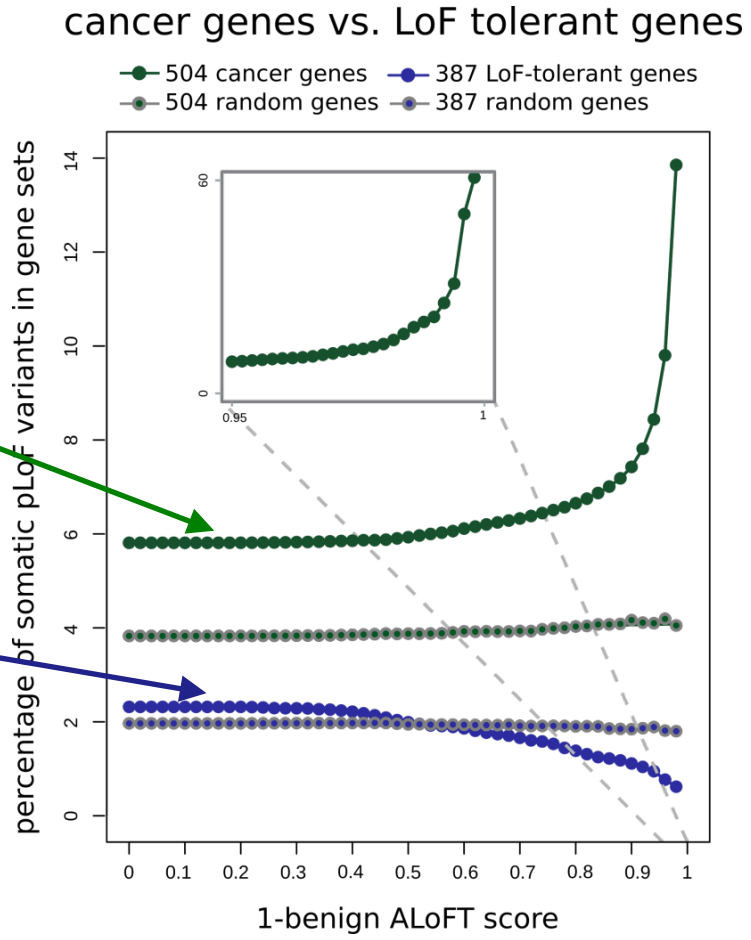
ALoFT identifies deleterious somatic LoF variants

Cancer genes:

- COSMIC consensus.
- *Enriched in deleterious LoFs.*

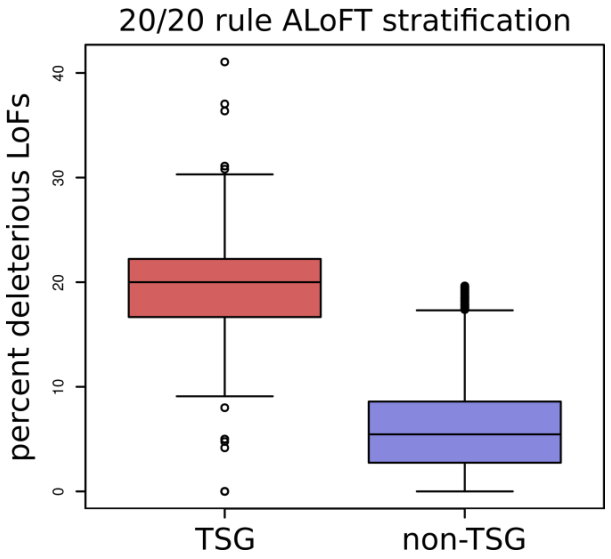
LoF tolerant genes:

- LoF in the 1KG cohort.
- *Depleted in deleterious LoFs.*



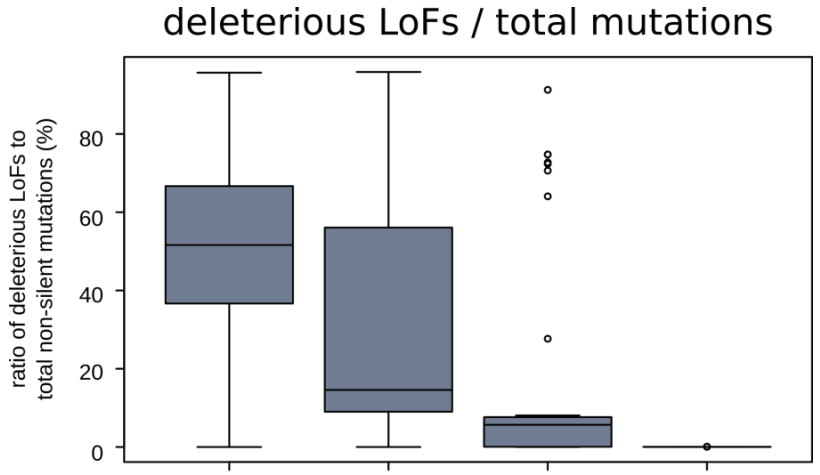
Balasubramanian S.*, Fu Y.* et al., *NComms.*, '17

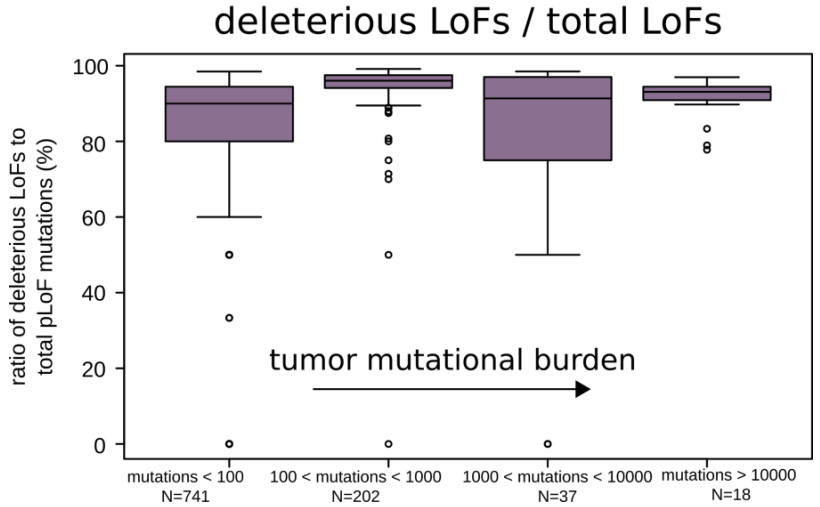
ALoFT refines cancer mutation characterization



Vogelstein et al. '13: if >20% of mutations in gene inactivating → tumor suppressor gene (TSG).

ALoFT further refines 20/20 rule predictions.

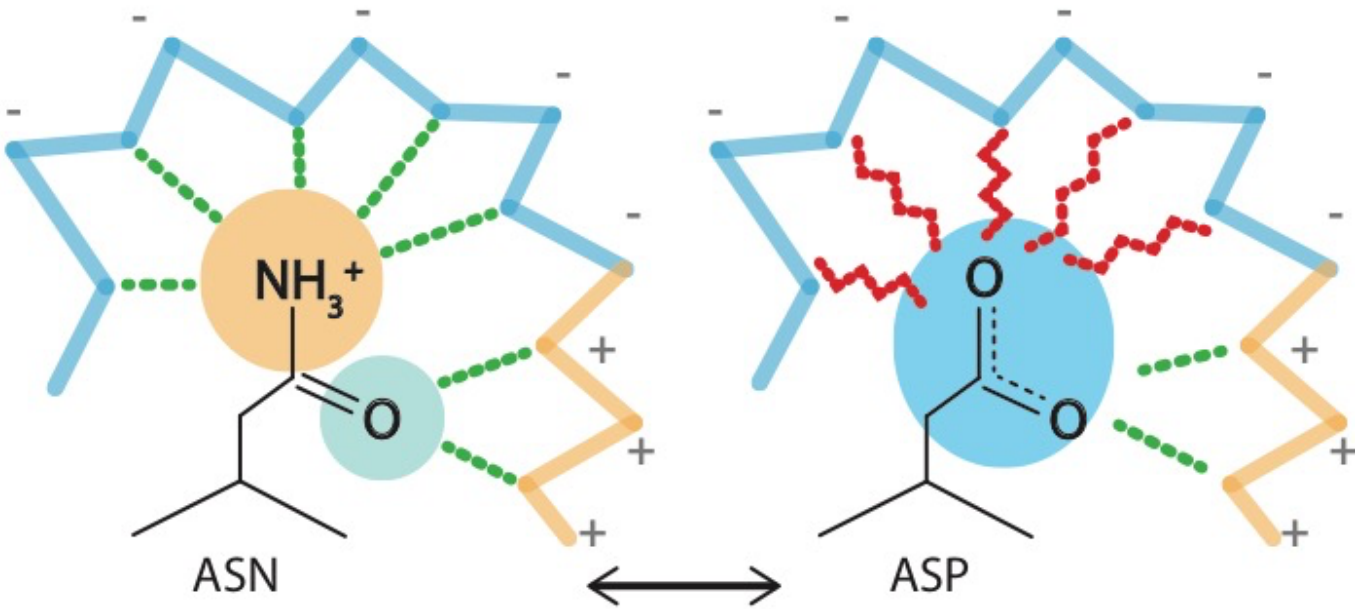


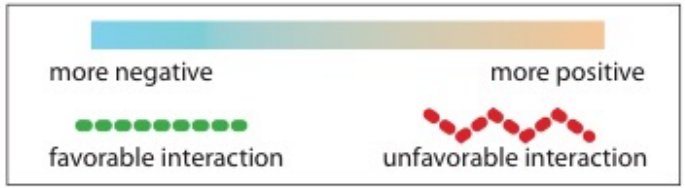


Balasubramanian S.*, Fu Y.* et al., *NComms.*, '17

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

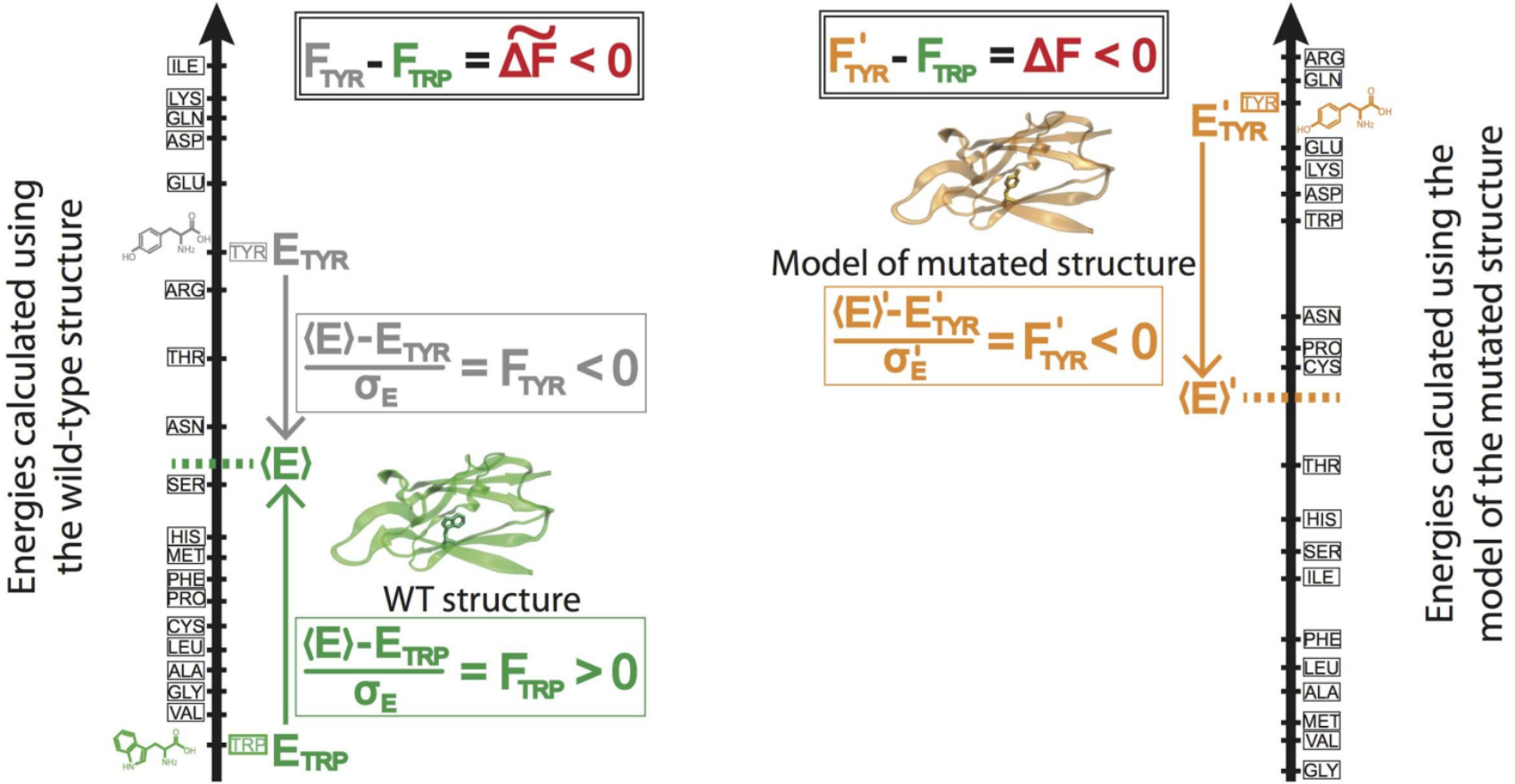




What is localized frustration ?

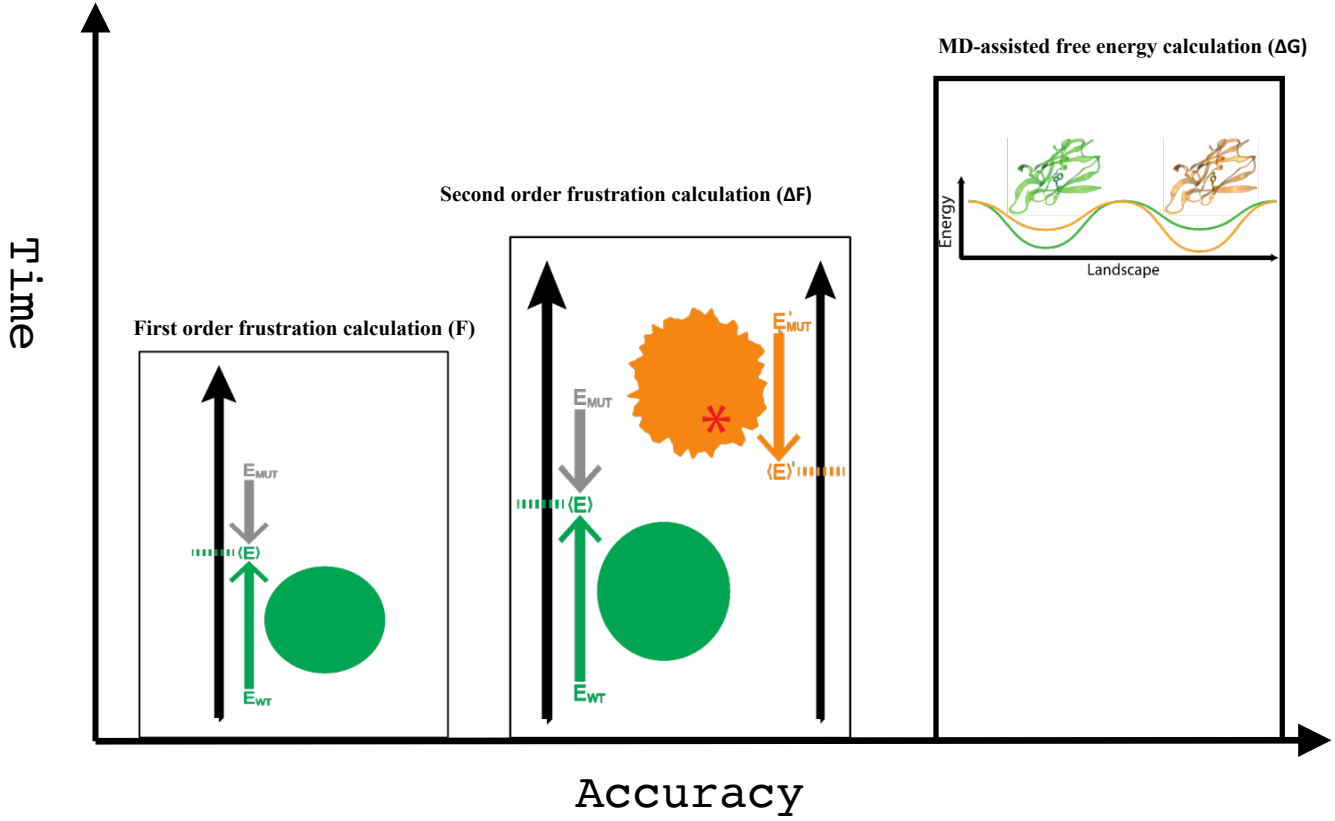
[Ferreiro et al., *PNAS* ('07)]

Workflow for evaluating localized frustration changes (ΔF)

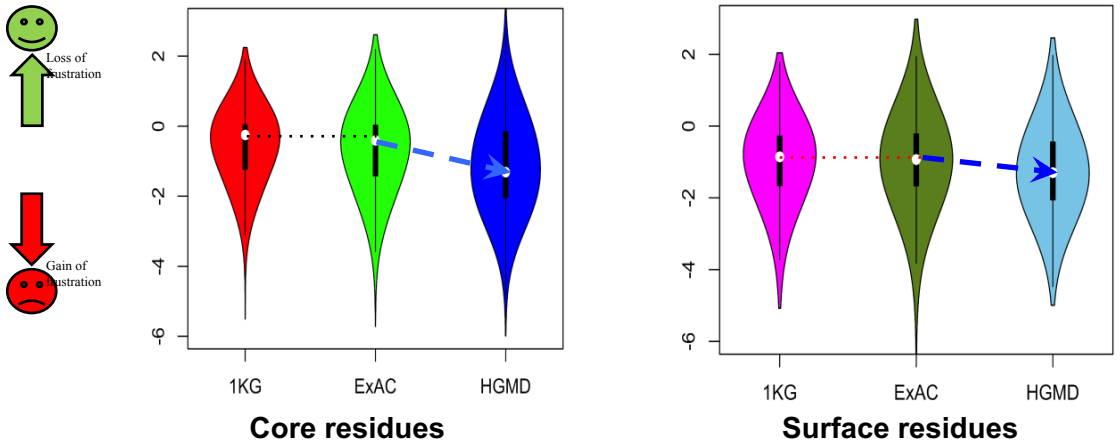


[Kumar et al. *NAR* (2016)]

Complexity of the second order frustration calculation



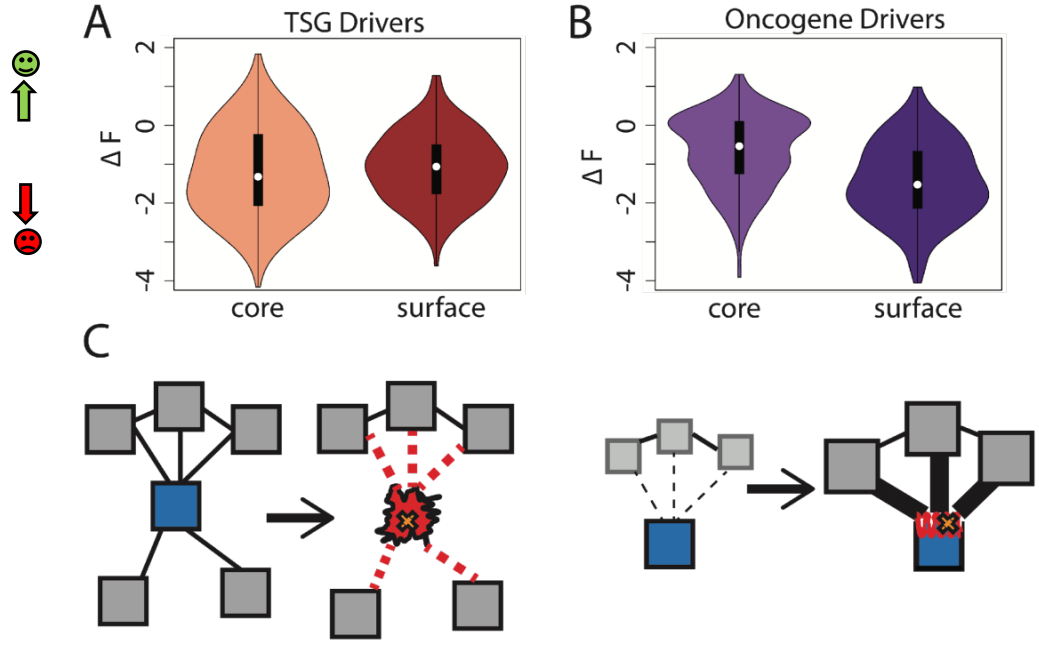
Comparing ΔF values across different SNV categories: disease v normal



Normal mutations (1000G) tend to unfavorably frustrate (less frustrated) surface more than core, but for disease mutations (HGMD) no trend & greater changes

[Kumar et al, *NAR* (2016)]

Comparison between ΔF distributions: TSGs v. oncogenes



SNVs in TSGs change frustration more in core than the surface, whereas those associated with oncogenes manifest the opposite pattern. This is consistent with differences in LOF v GOF mechanisms.

[Kumar et al, *NAR* (2016)]

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

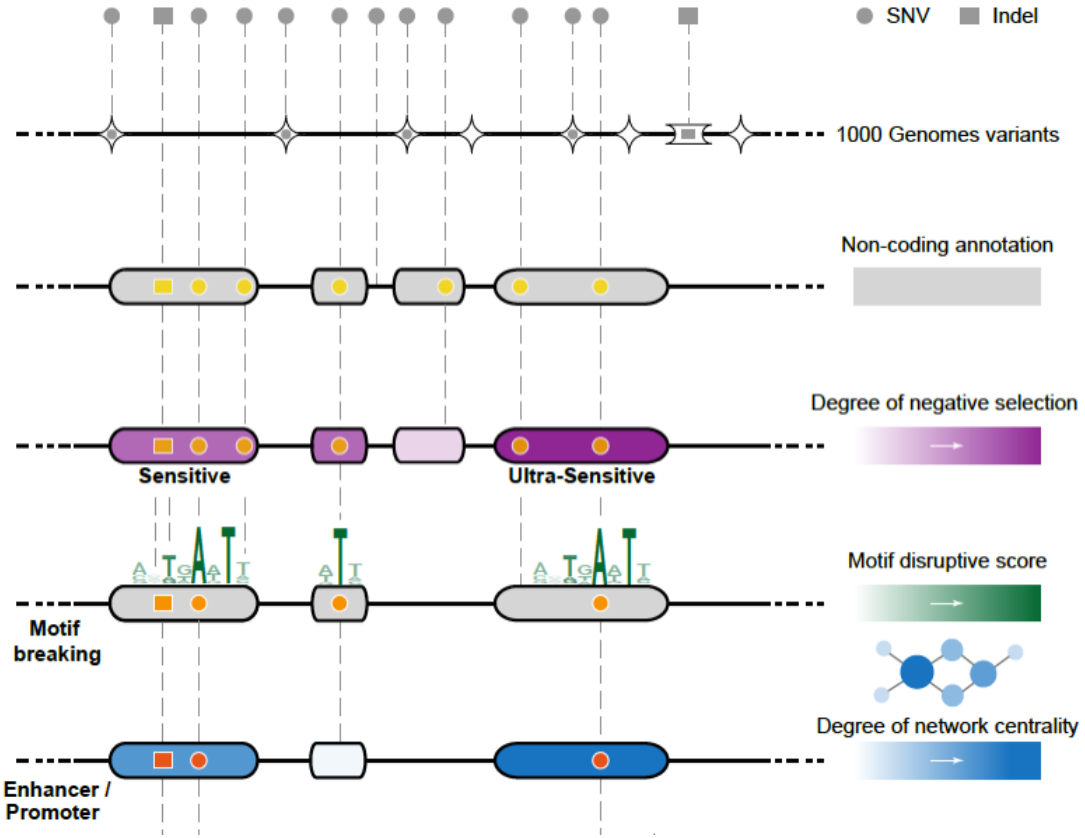
Funseq: a flexible framework to determine functional impact & use this to prioritize variants

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

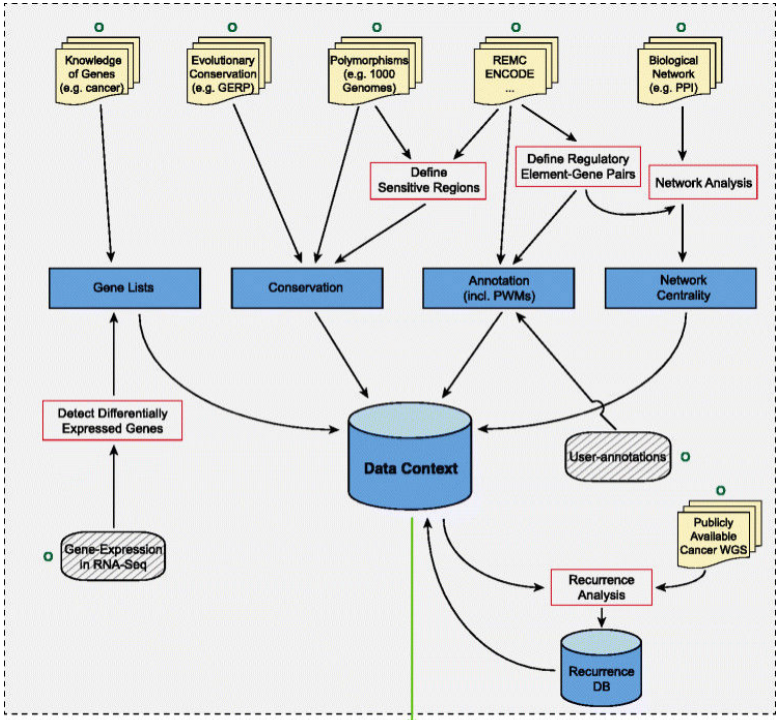
Conservation (GERP, allele freq.)

Mutational impact (motif breaking, Lof)

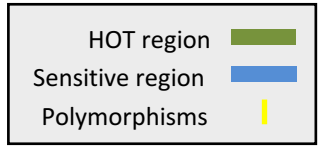
Network (centrality position)



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



FunSeq.gersteinlab.org

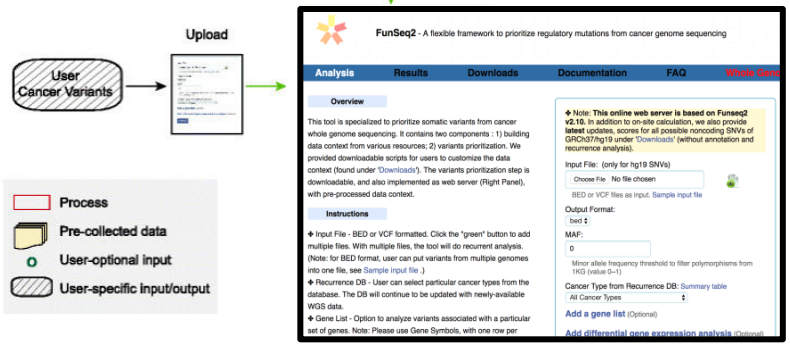


Genome

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

- 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

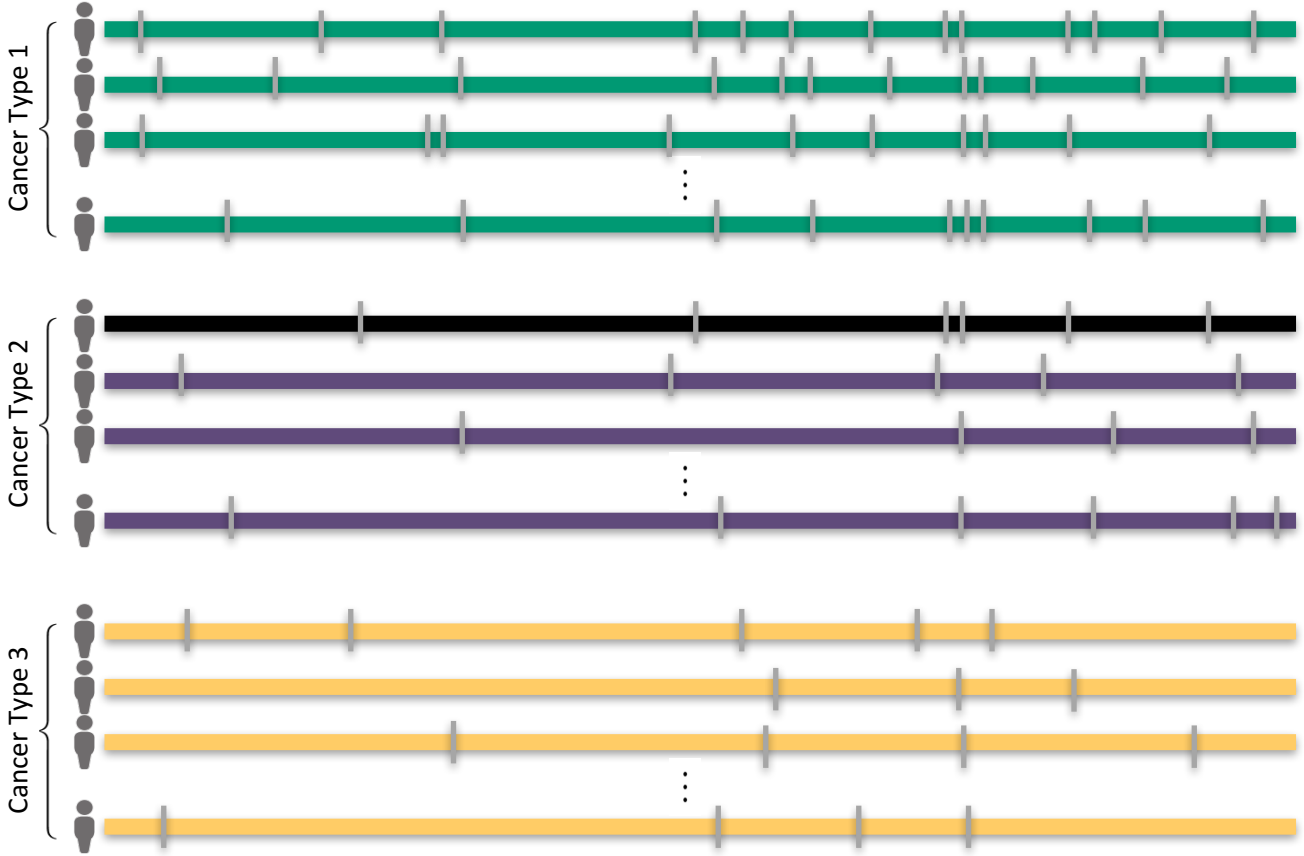


[Fu et al., GenomeBiology ('14)]

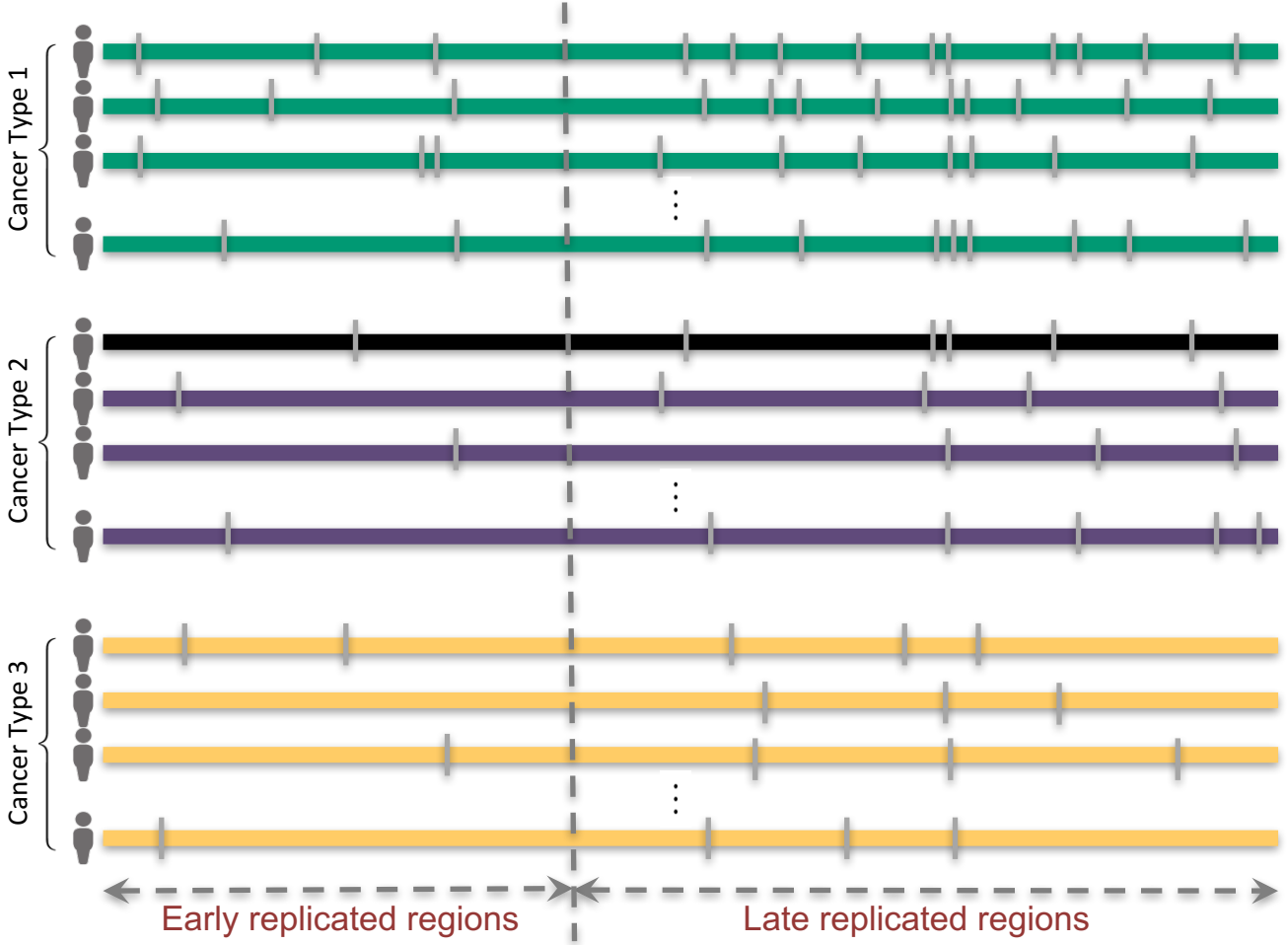
Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

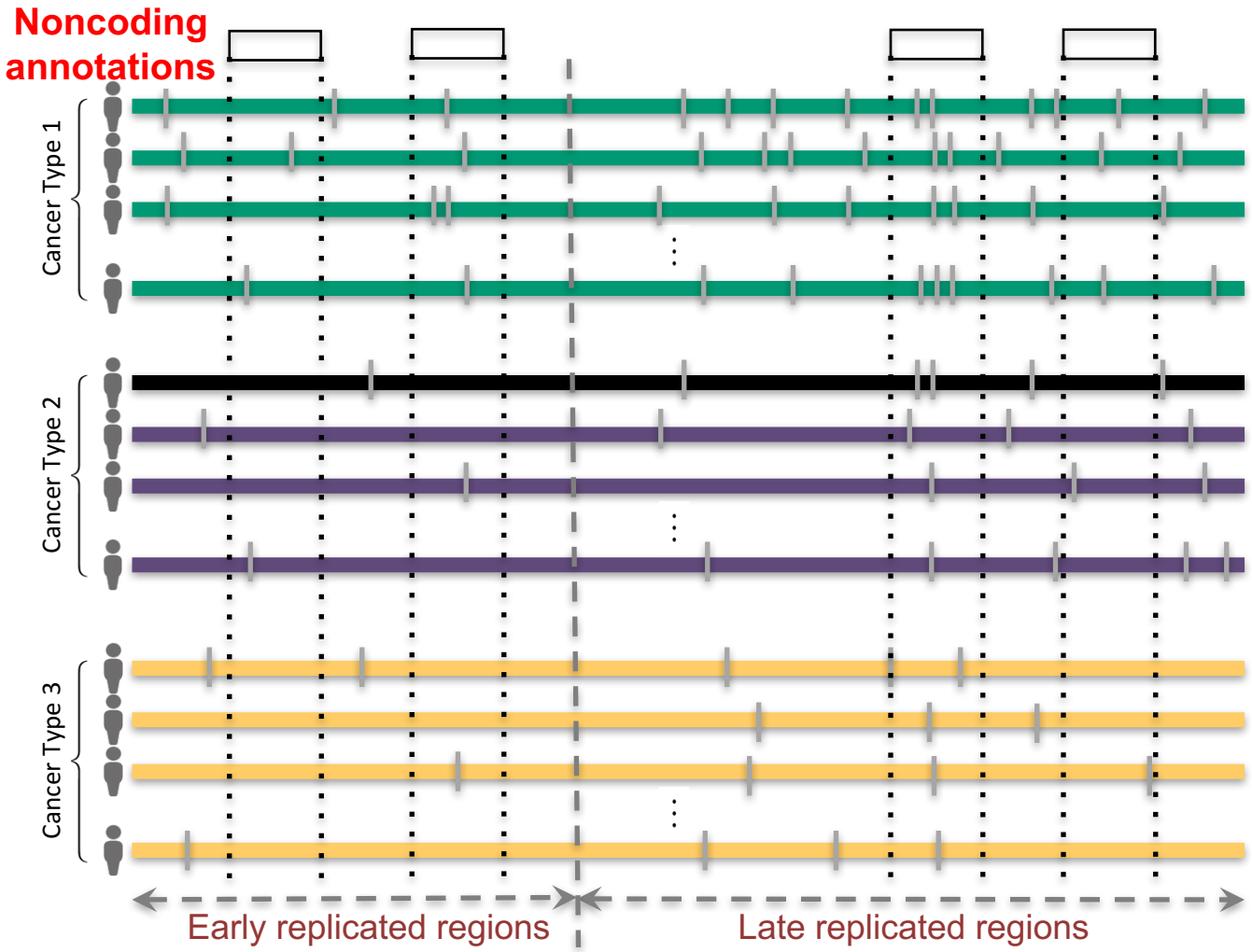
- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

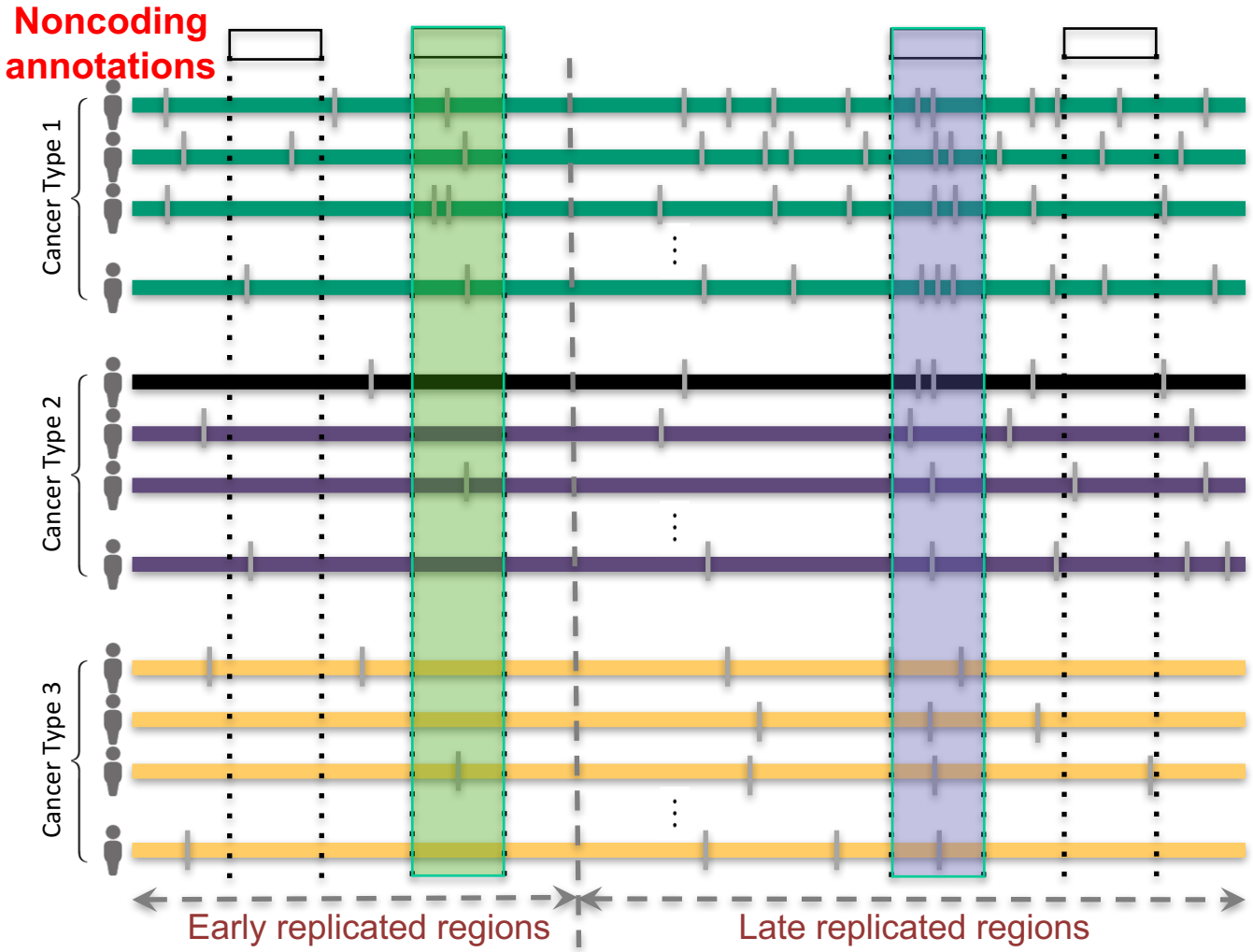
Mutation recurrence



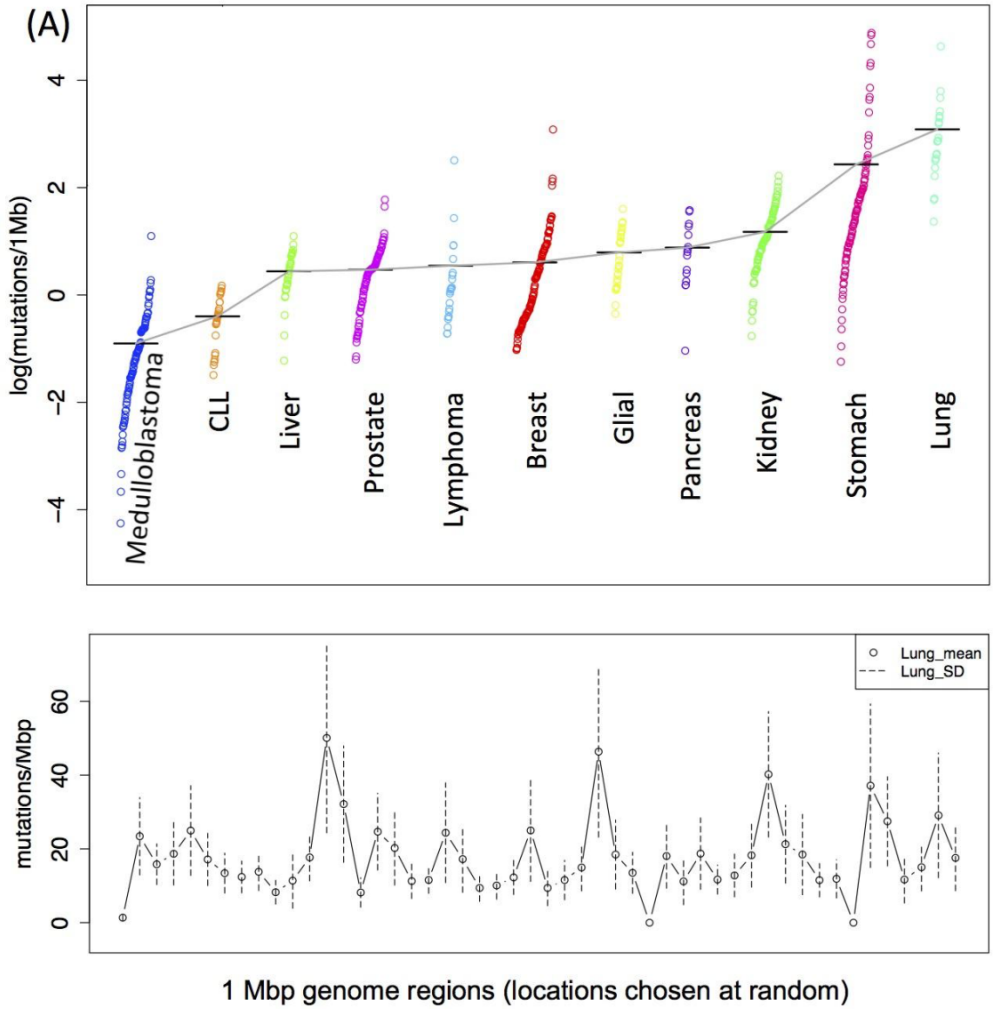
Mutation recurrence



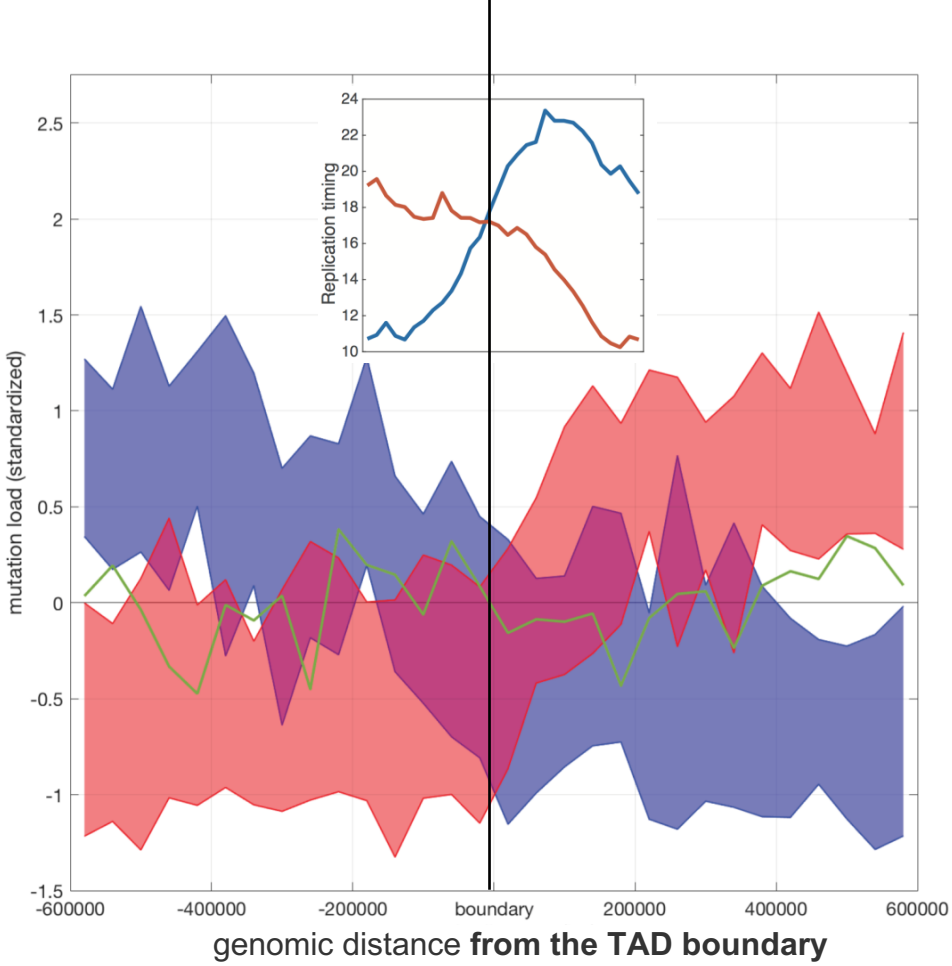

Noncoding annotations
Cancer Type 1
Cancer Type 2
Cancer Type 3
Early replicated regions
Late replicated regions


Noncoding annotations
Cancer Type 1
Cancer Type 2
Cancer Type 3
Early replicated regions
Late replicated regions

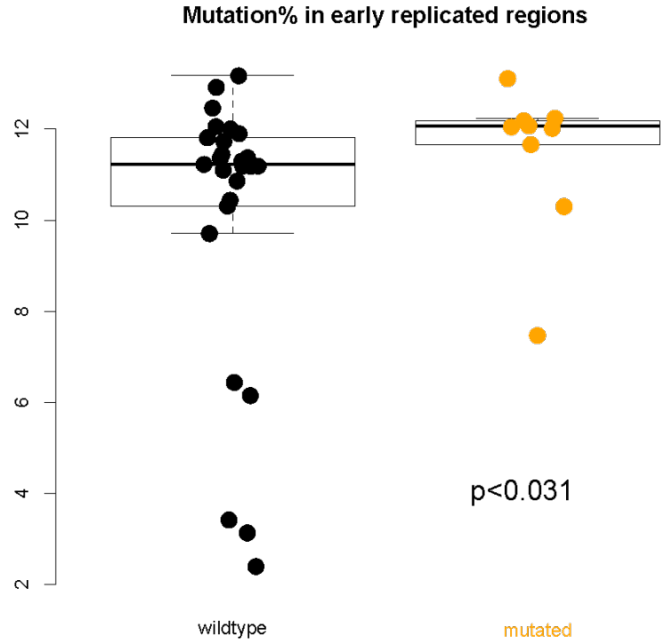
Cancer Somatic Mutational Heterogeneity, across cancer types, samples & regions



[Lochovsky et al. *NAR* ('15)]



[Yan et al., *PLOS Comp. Bio. ('17)*; S. Li et al., PLOS Genetics ('17)]]

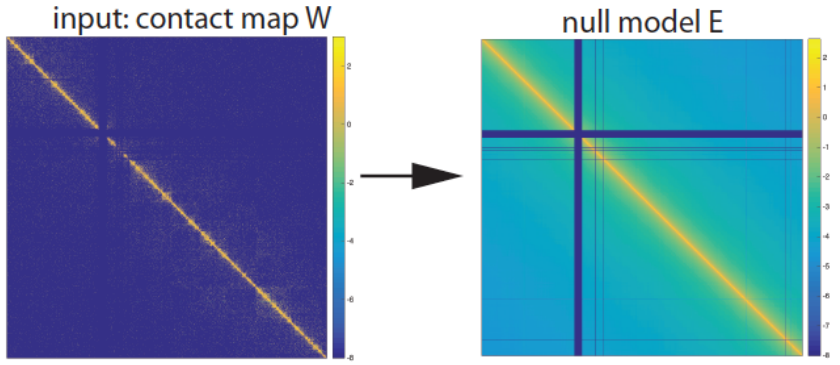


Chromatin remodeling failure leads to more mutations in early-replicating regions

Variation in somatic mutations is closely associated with chromatin structure (TADs) & replication timing

mrTADFinder:

Identifying TADs at multiple resolutions by maximizing modularity vs appropriate null



Choose a particular resolution γ
Optimize Q over all possible partitions

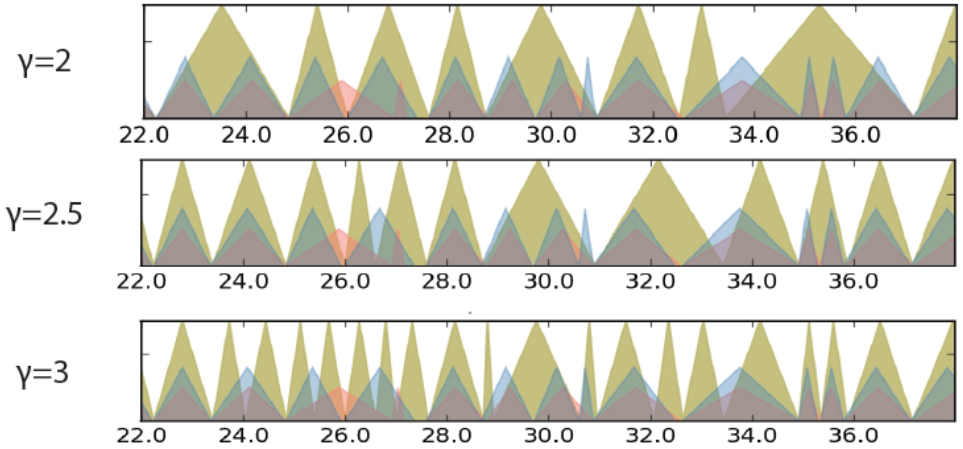
$$Q = \frac{1}{2N}\sum_{ij}(W_{ij} - \gamma E_{ij})\delta_{\sigma_i\sigma_j}$$

γ: resolution parameter

Multiple runs to define boundary scores for all pairs of adjacient bins

consensus boundaries based on the boundary scores

consensus TADs → output



[Yan et al., *PLOS Comp. Bio. ('17)*]

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Cancer Somatic Mutation Modeling

PARAMETRIC MODELS

Model 1: Constant Background Mutation Rate (Model from Previous Work)

$x_i : Binomial(n_i, p)$

Model 2a: Varying Mutation Rate with Single Covariate Correction

$x_i : Binomial(n_i, p_i)$

$p_i : Beta(\mu|R_i, \sigma|R_i)$

$\mu|R_i, \sigma|R_i$: constant within the same covariate rank

Model 2b: Varying Mutation Rate with Multiple Covariate Correction

$x_i : Binomial(n_i, p_i)$

$p_i : Beta(\mu|\boldsymbol{R_i}, \sigma|\boldsymbol{R_i})$

$\mu|\boldsymbol{R_i}, \sigma|\boldsymbol{R_i}$: constant within the same covariate rank

[Lochovsky et al. *NAR* ('15)]

- Suppose there are *k* genome elements. For element *i*, define:
  - n_i: total number of nucleotides
  - x_i: the number of mutations within the element
  - p: the mutation rate
  - R_i: the covariate rank of the element
- Non-parametric model is useful when covariate data is missing for the studied annotations
  - Also sidesteps issue of properly identifying and modeling every relevant covariate (possibly hundreds)

NON-PARAMETRIC MODELS

Assume constant background mutation rate in local regions.

Model 3a: Random Permutation of Input Annotations

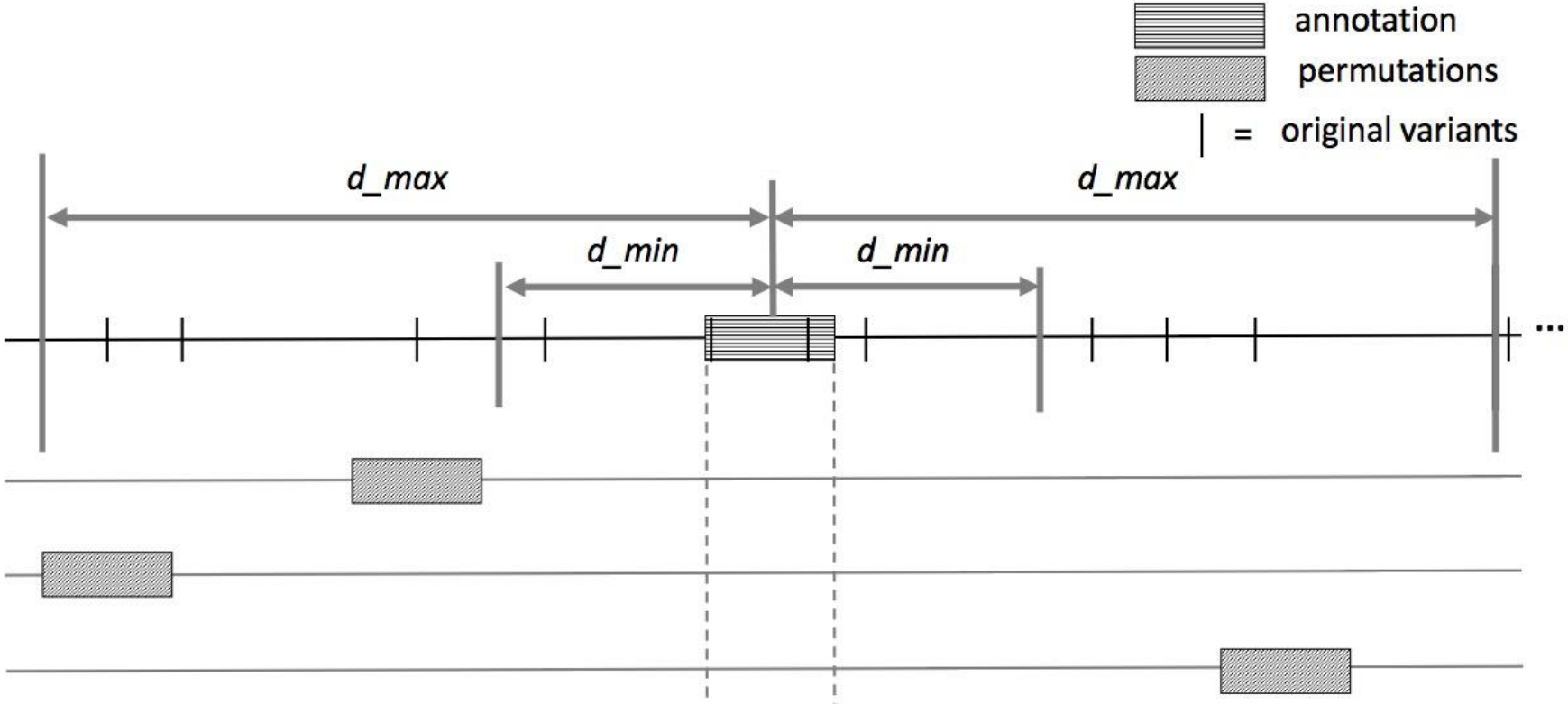
Shuffle annotations within local region to assess background mutation rate.

Model 3b: Random Permutation of Input Variants

Shuffle variants within local region to assess background mutation rate.

[Lochovsky et al. under review]

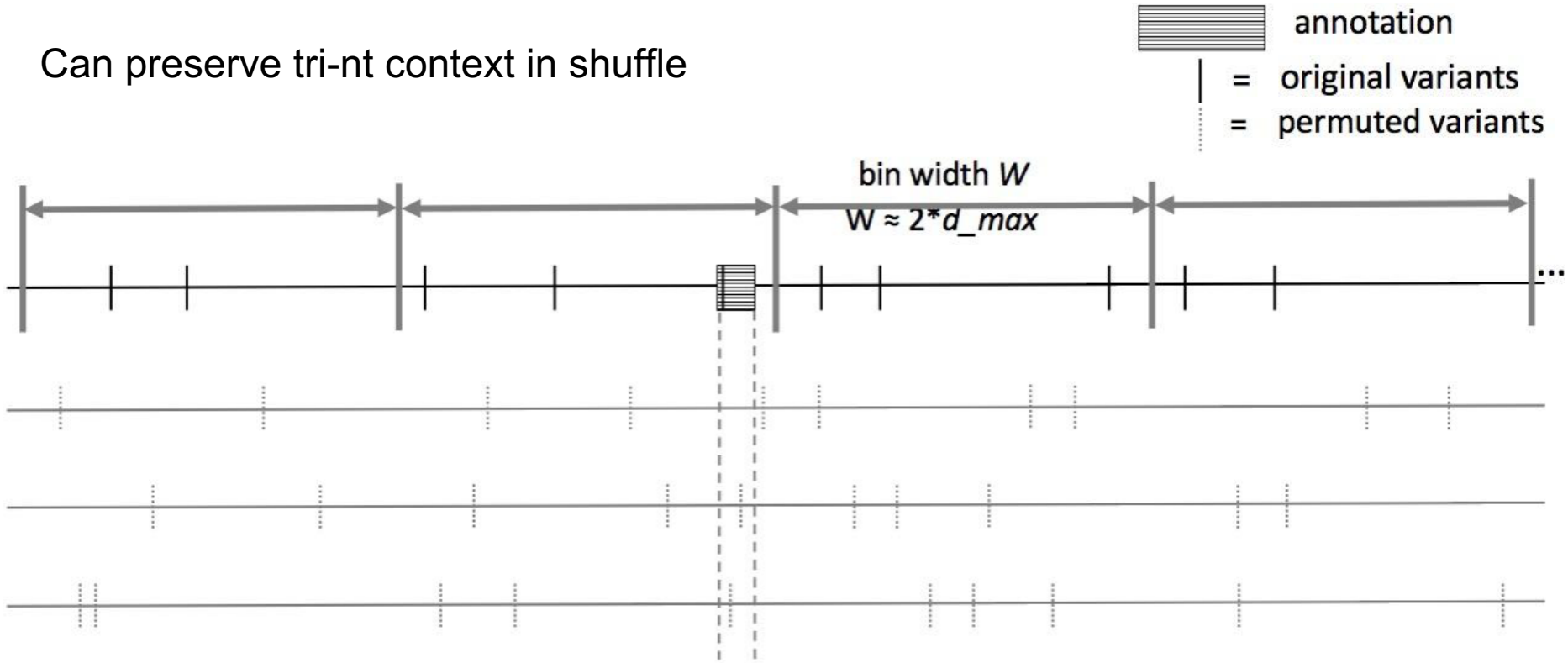
MOAT-a: Annotation-based permutation



[Lochovsky et al. under review]

MOAT-v: Variant-based Permutation

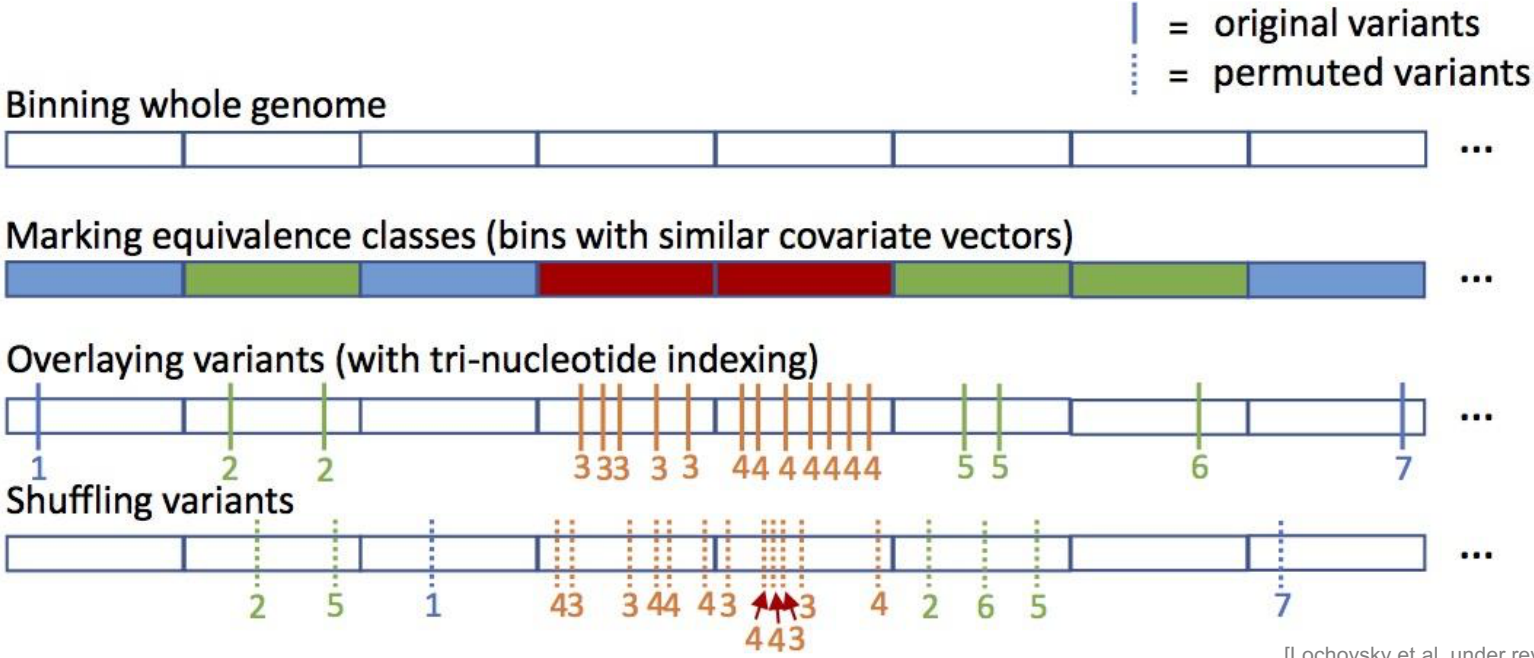
Can preserve tri-nt context in shuffle



[Lochovsky et al. under review]

MOAT-s: a variant on MOAT-v

- A somatic variant simulator
  - Given a set of input variants, shuffle to new locations, taking genome structure into account

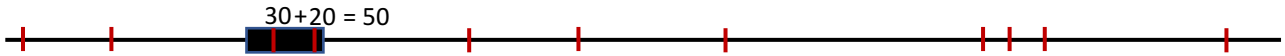


[Lochovsky et al. under review]

Funseq Integration with MOAT

- Run Funseq over the whole genome
  - Produce signal track that is the maximum score at each position

  - Calculate an annotation signal by summing the intersecting variants' scores



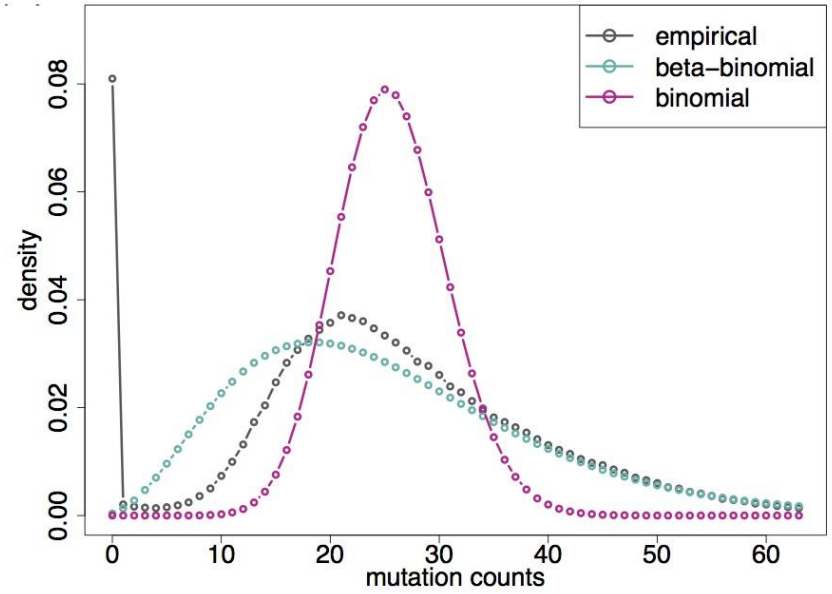
  - Use the same procedure on permuted data

  - These are background scores to determine if the observed score is significantly elevated

[Lochovsky et al. under review]

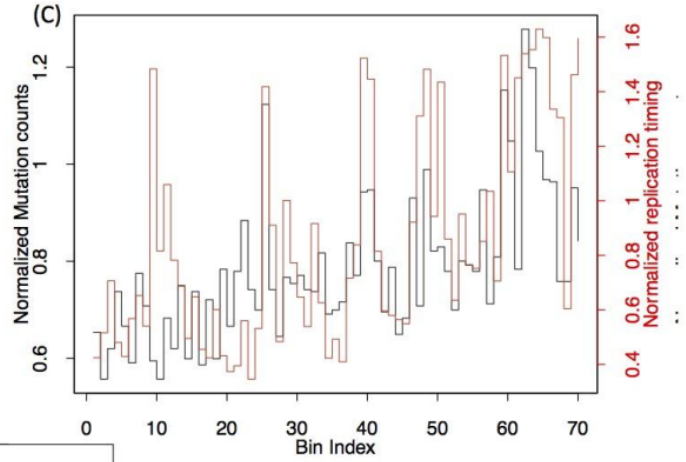
LARVA Model Comparison

- Comparison of mutation count frequency implied by the binomial model (model 1) and the beta-binomial model (model 2) relative to the empirical distribution
- The beta-binomial distribution is significantly better, especially for accurately modeling the over-dispersion of the empirical distribution

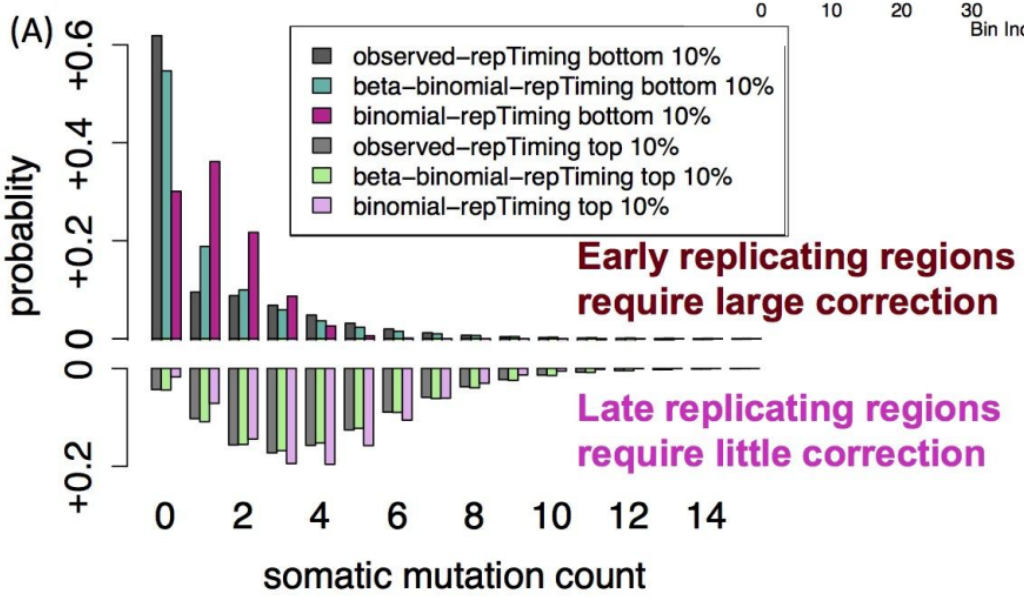


[Lochovsky et al. *NAR* ('15)]

Adding DNA replication timing correction further improves the beta-binomial model



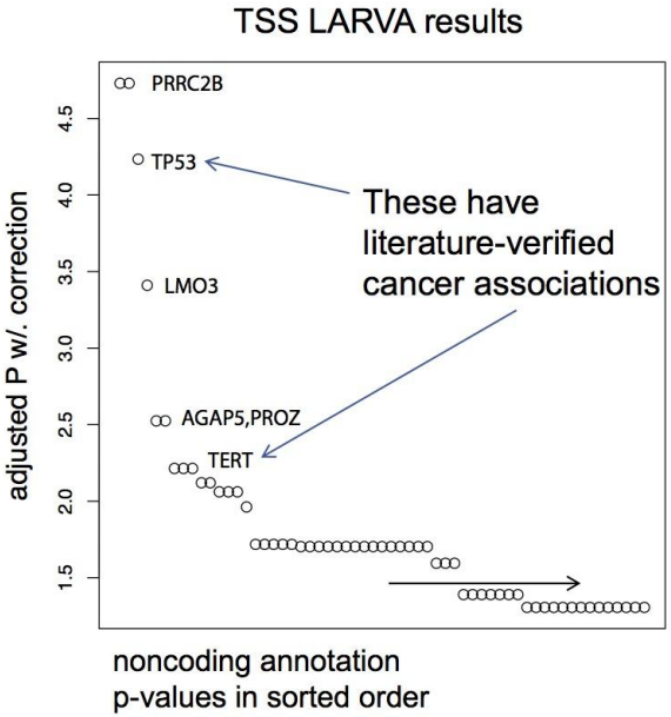
[Lochovsky et al. *NAR* ('15)]

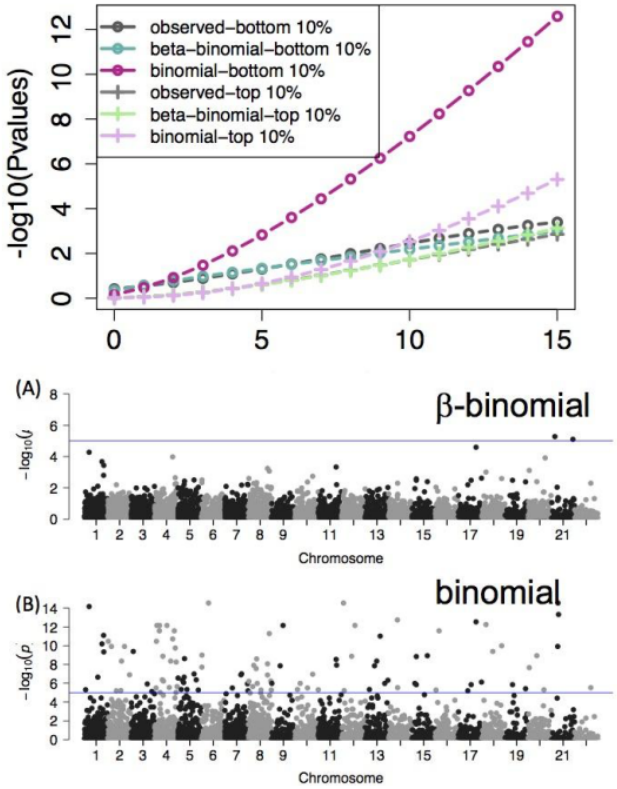


Early replicating regions require large correction

Late replicating regions require little correction

LARVA Results





[Lochovsky et al. *NAR* ('15)]

MOAT: recapitulates LARVA with GPU-driven runtime scalability

Gene Name	Documented role with cancer	Pubmed ID
SLC3A1	Cysteine transporter SLC3A1 promotes breast cancer tumorigenesis	28382174
ADRA2B	reduce cancer cell proliferation, invasion, and migration	25026350
SIL1	subtype-specific proteins in breast cancer	23386393
TCF24	NA	NA
AGAP5	significant mutation hotspots in cancer	25261935
TMPRSS13	Type II transmembrane serine proteases in cancer and viral infections	19581128
ERO1L	Overexpression of ERO1L is Associated with Poor Prognosis of Gastric Cancer	26987398

.
.
.

MOAT's high mutation burden elements recapitulate LARVA's results & published noncoding cancer-associated elements.

Computational efficiency of MOAT's NVIDIA™ CUDA™ version, with respect to the number of permutations, is dramatically enhanced compared to CPU version.

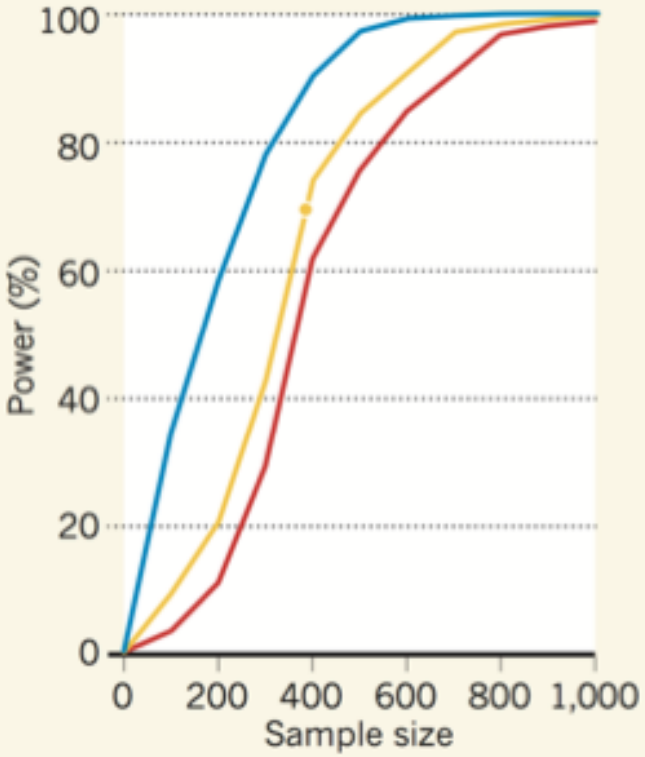
Number of permutations	Fold speedup of CUDA version
1k	14x
10k	100x
100k	256x

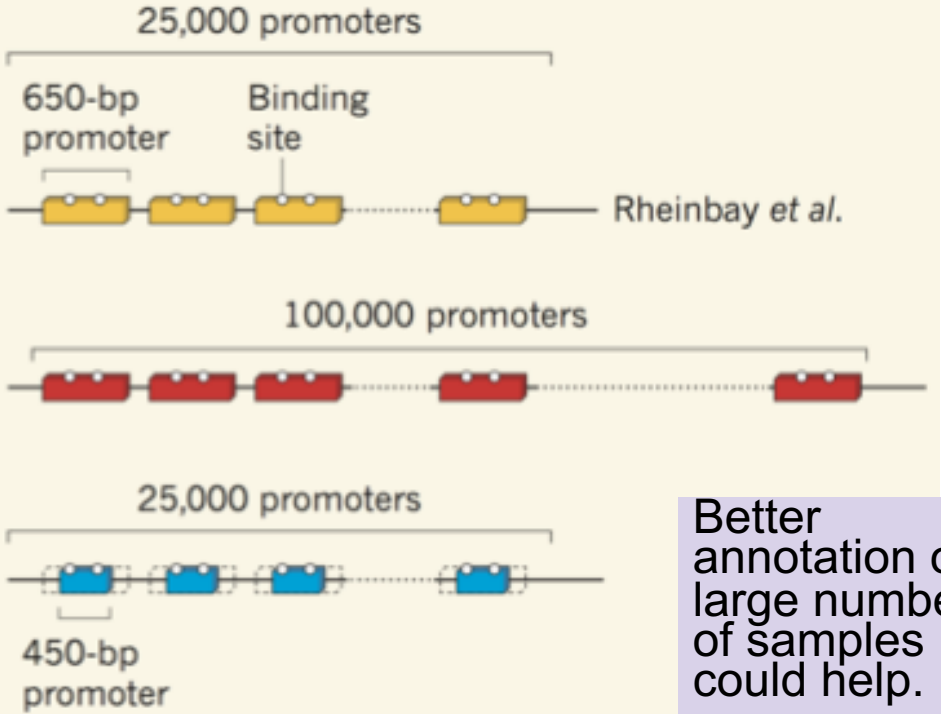
[Lochovsky et al. under review]

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Power, as an issue in driver discovery



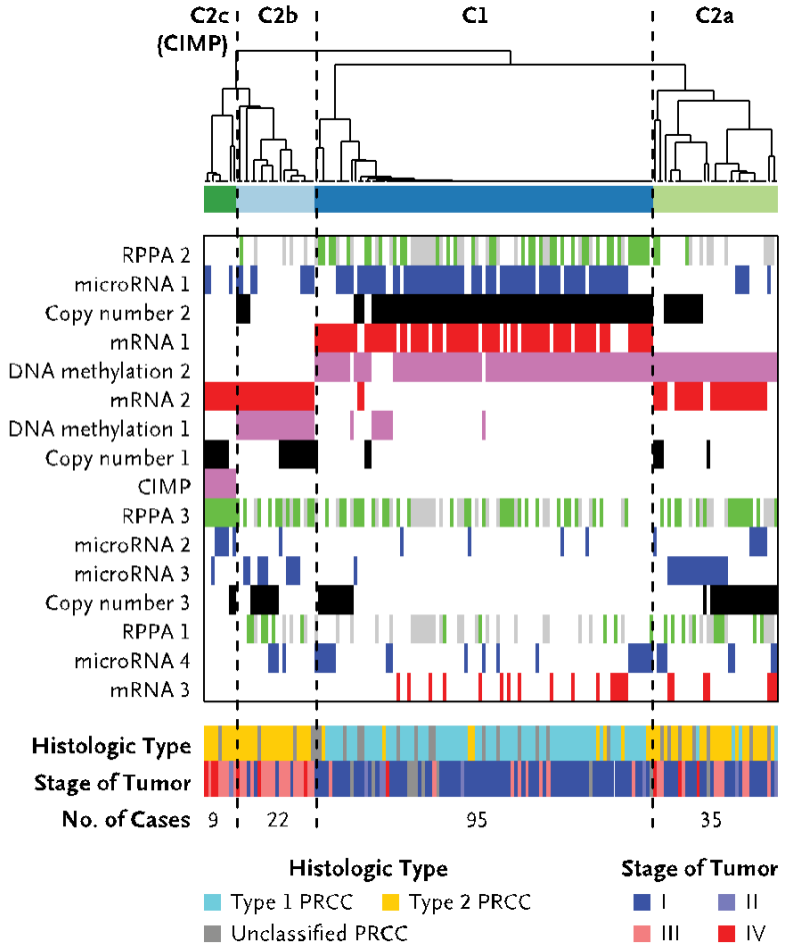


Better annotation or large number of samples could help.

[Kumar & Gerstein, *Nature* ('17)]

An (underpowered) case study: pRCC

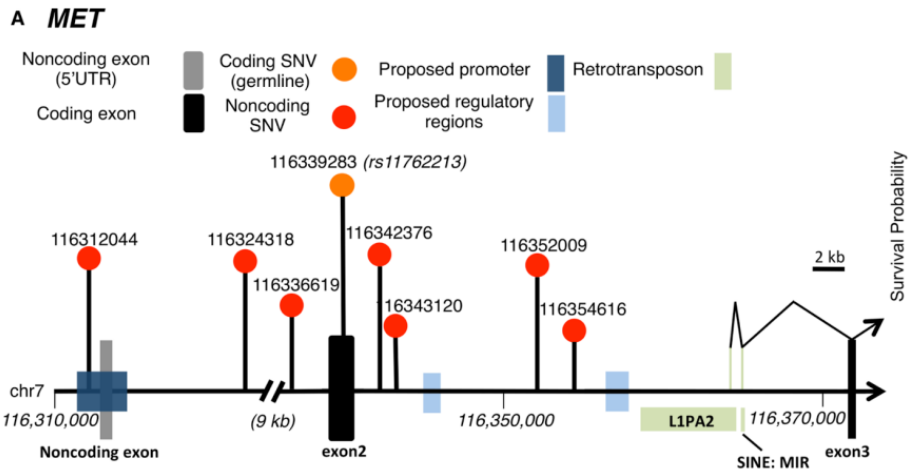
- Kidney cancer lifetime risk of 1.6% & the papillary type (pRCC) counts for ~10% of all cases
- TCGA project sequenced 161 pRCC exomes & classified them into subtypes
  - Yet, cannot pin down the cause for a significant portion of cases....
- 35 WGS of TN pairs, perhaps useful?

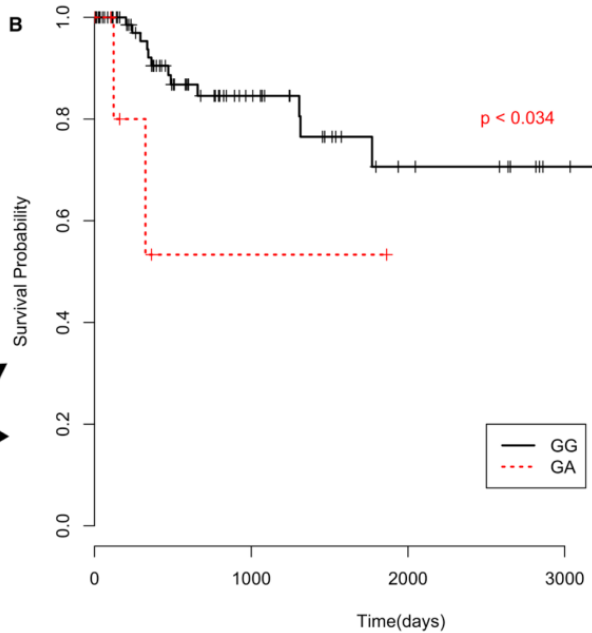


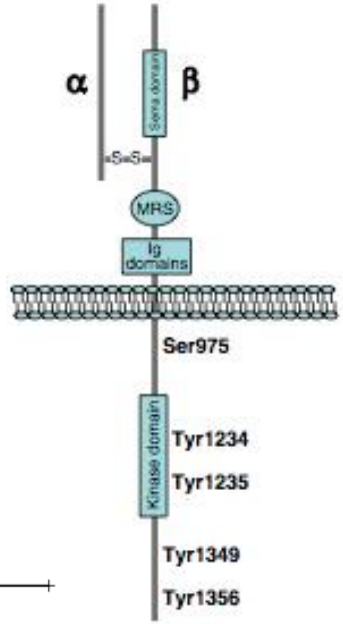
[Cancer Genome Atlas Research Network N Engl J Med. ('16)]

Tyr-kinase MET: Known Facts & New Results

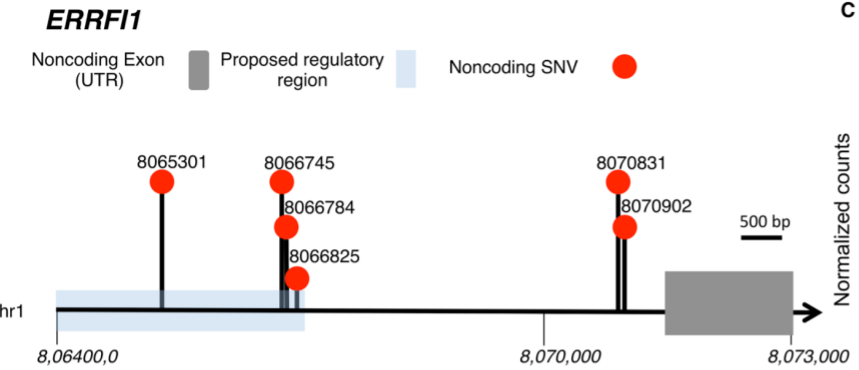
- MET is long known pRCC driver
- In MET, TCGA found somatic SNVs, duplications & an alt. splicing event as drivers (43/161).
- In addition, from 35 WGS we found
  - A noncoding hotspot associated with *MET*
  - Lack of SV and breakpoint disrupting *MET*
  - Germline SNP (rs11762213) predicts survival in type 2 patients

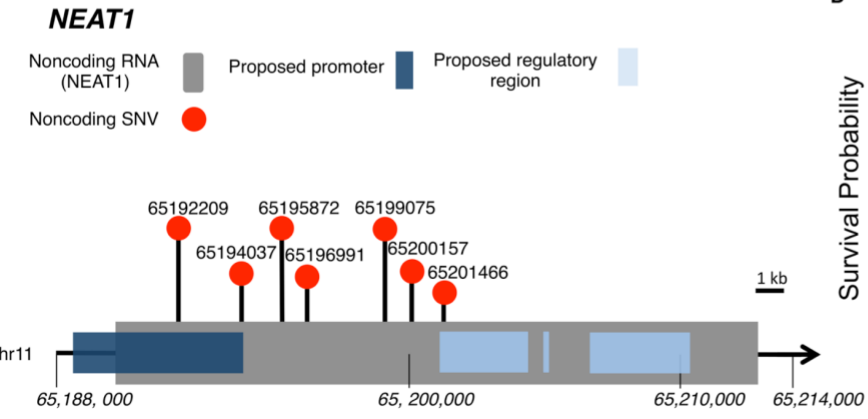


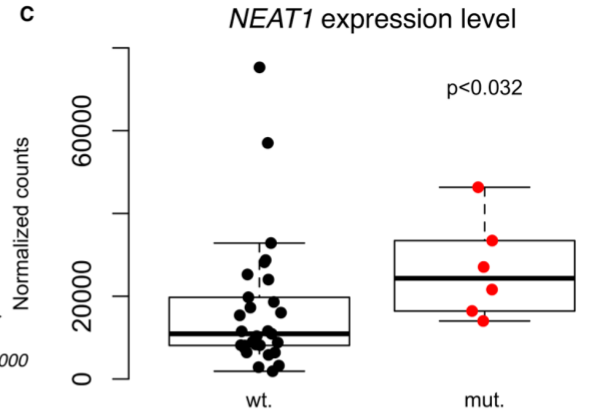


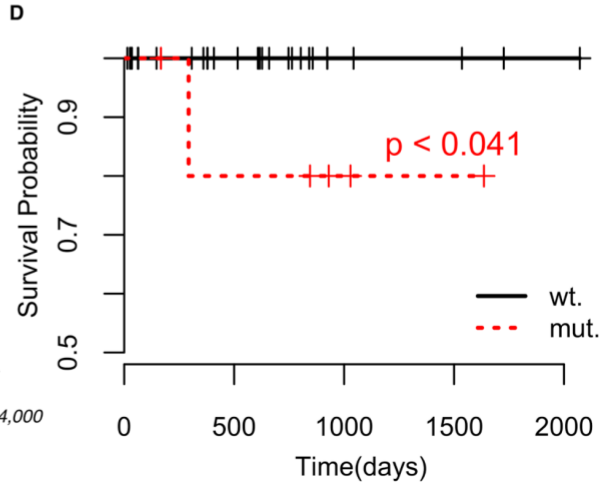


[A. Gentile, L. Trusolino and PM. Comoglio, Cancer and Metastasis Reviews ('08); S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]







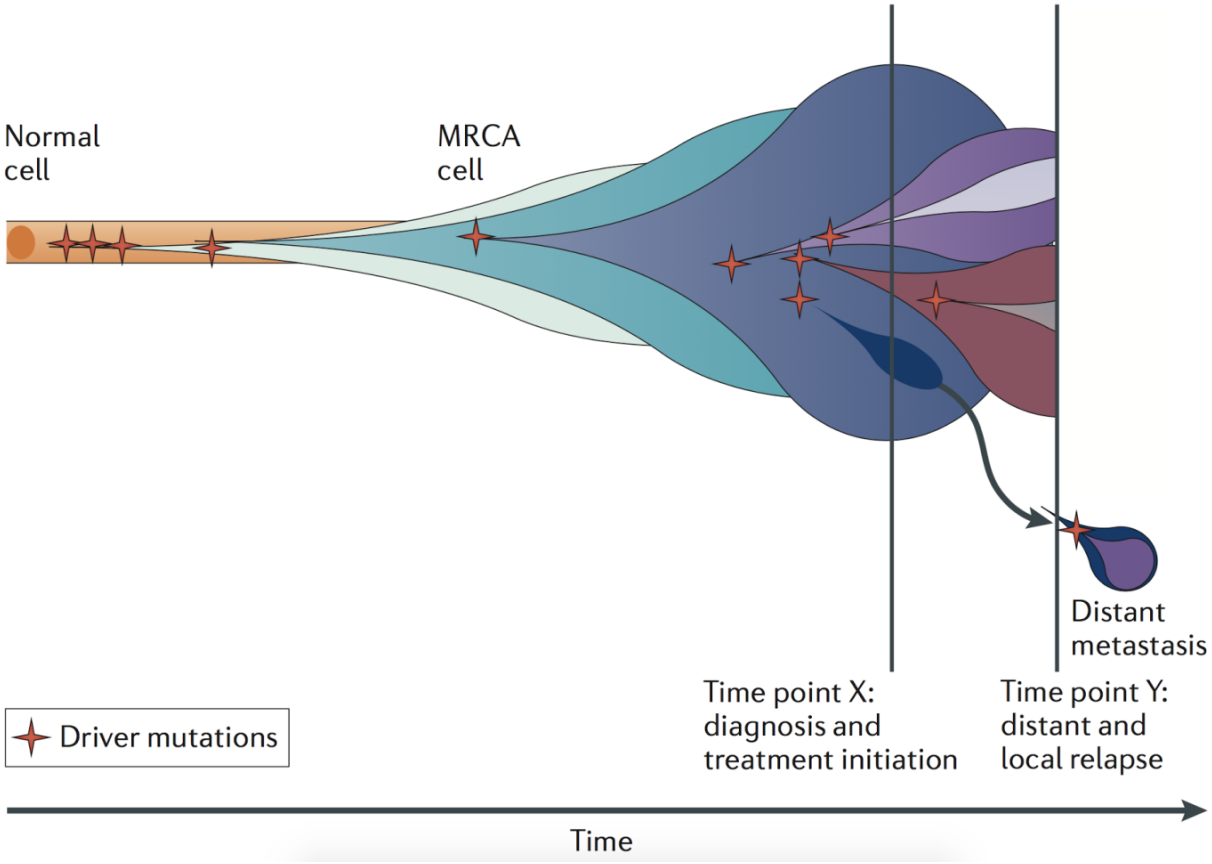


Beyond *MET*: 2 non-coding hotspots in NEAT & ERRFI1,

supported by expr. changes & survival analysis

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

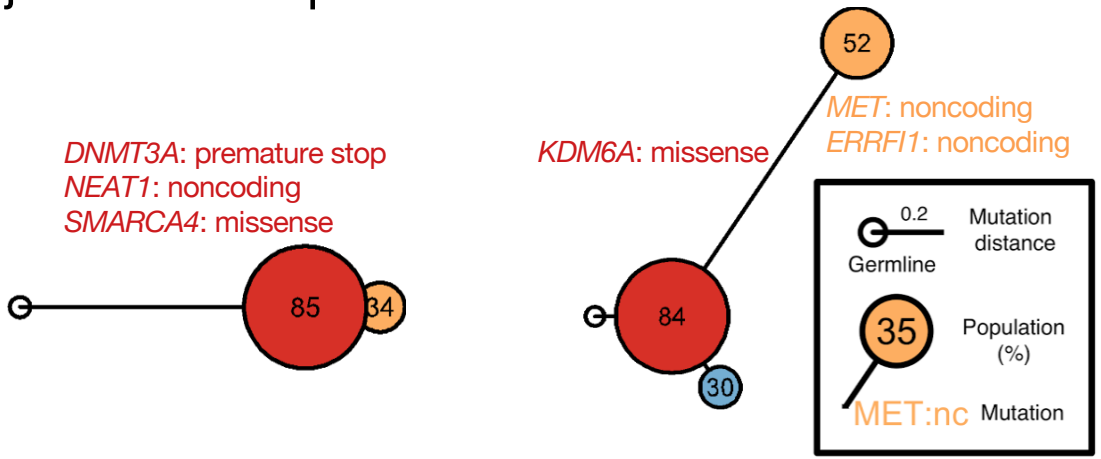
Tumor Evolution: Highlight the Ordering of Key Mutations



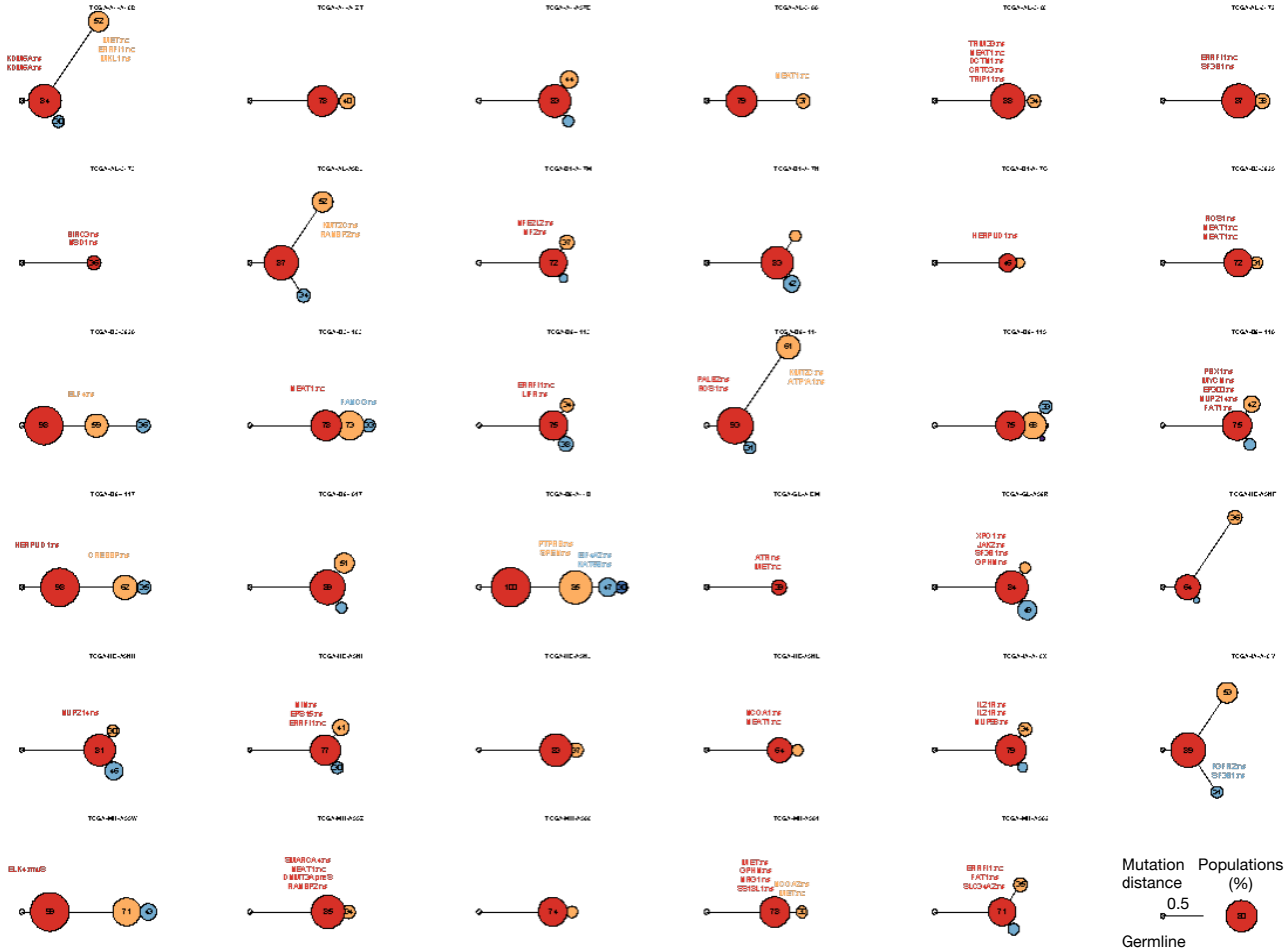
Yates et al, NRG (2012)

Construct evolutionary trees in pRCC

- Infer mutation order and tree structure based on mutation abundance (PhyloWGS, Deshwar et al., 2015)
- Some of the key mutations occur in all the clones while others are just in some parts of the tree



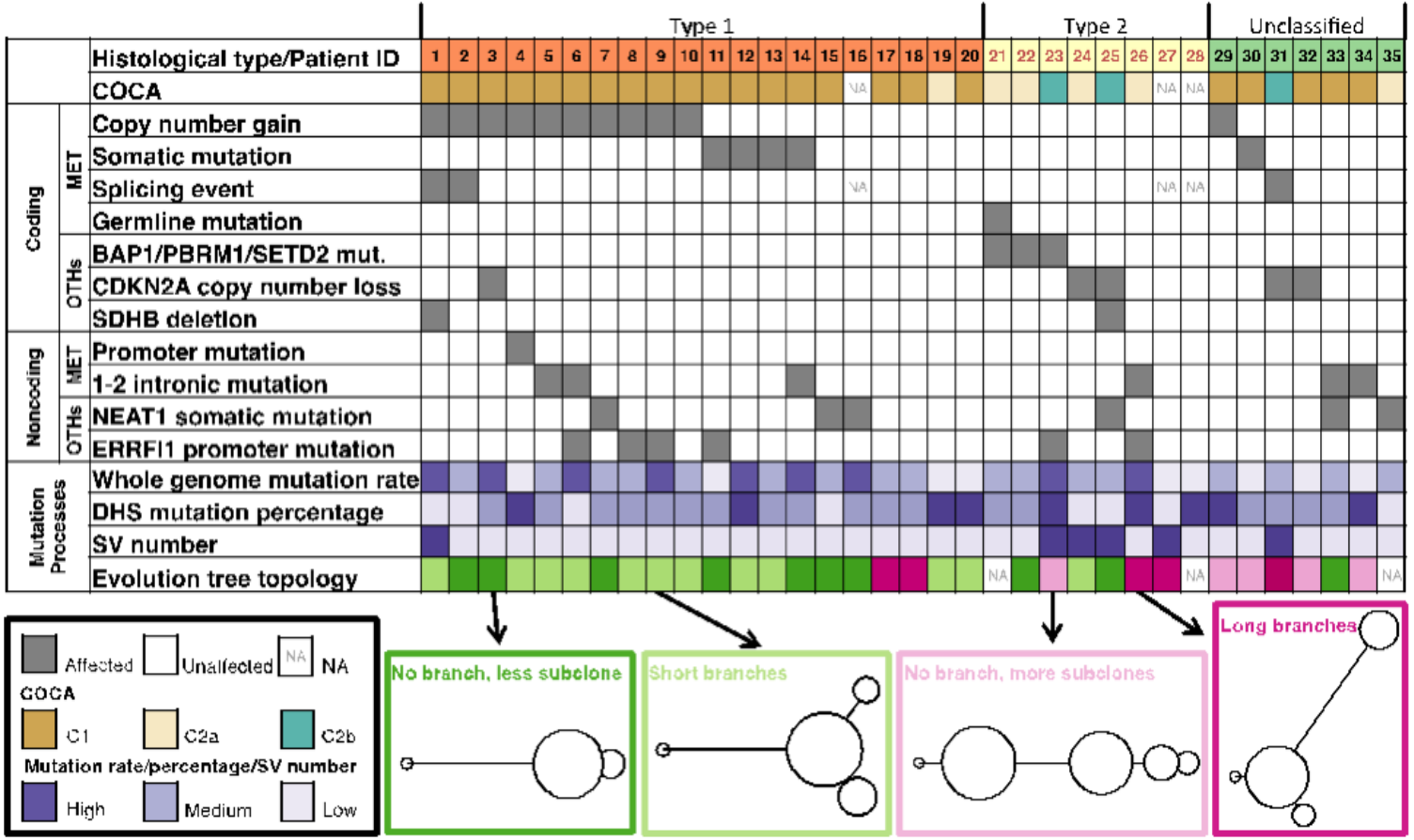
[S. Li, B. Shuch and M. Gerstein PLOS Genetics ('17)]



[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



Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Prioritizing somatic variants: Approaches to identifying key variants through functional impact & recurrence

- Introduction
  - Large growth in cancer genome data
  - Mining the data to prioritize variants for key drivers
- Functional impact #1: Coding
  - **ALoFT**: Annotation of Loss-of-Function Transcripts. LOF annotation as a complex problem.
  - Finding deleterious LoF SNVs
  - **Frustration** as a localized metric of SNV impact. Differential profiles for oncogenes vs. TSGs
- Functional impact #2: Non-coding
  - **FunSeq** integrates evidence, with an entropy based weighting scheme
- Recurrence #1: Statistics for driver identification
  - **Background mutation rate** significantly varies & is correlated with replication timing & TADs
  - Developed a variety of parametric & non-parametric methods taking this into account
  - **LARVA** uses parametric beta-binomial model, explicitly modeling covariates
  - **MOAT** does a variety of non-parm. shuffles (annotation, variants, &c). Useful when explicit covariates not available. Slower than but speeded up w/ GPUs
- Recurrence #2: (Low-power) application to **pRCC**
  - WGS finds additional facts on the canonical driver, MET. Other suggestive non-coding hotspots.
  - Tumor evolution analysis of the timing of key mutations helps with classification

Acknowledgments

github.com/gersteinlab/**Frustration**
S **Kumar**, D Clarke

github.com/gersteinlab/**MrTADfinder**
KK **Yan**, S Lou

VAT.gersteinlab.org
L **Habegger**, S Balasubramanian,
DZ Chen, E Khurana, A Sboner,
A Harmanci, J Rozowsky, D Clarke, M Snyder

ALoFT.gersteinlab.org
S **Balasubramanian**, Y **Fu**,
M Pawashe, P McGillivray, M Jin, J Liu,
KJ Karczewski, DG MacArthur

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

pRCC
S **Li**, B Shuch

LARVA.gersteinlab.org
L **Lochovsky**,
J **Zhang**,
Y Fu, E Khurana

MOAT.gersteinlab.org
L **Lochovsky**,
J **Zhang**

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 gersteinlab.org). Paper references in the talk were mostly 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