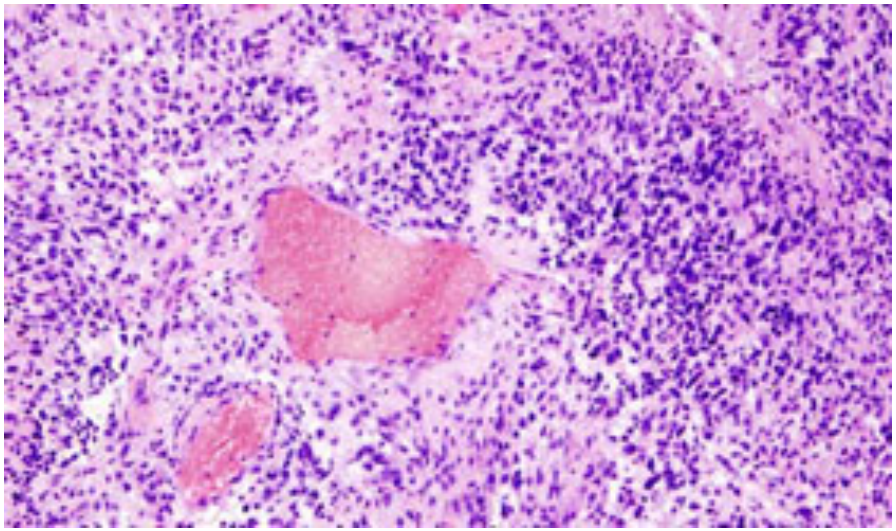
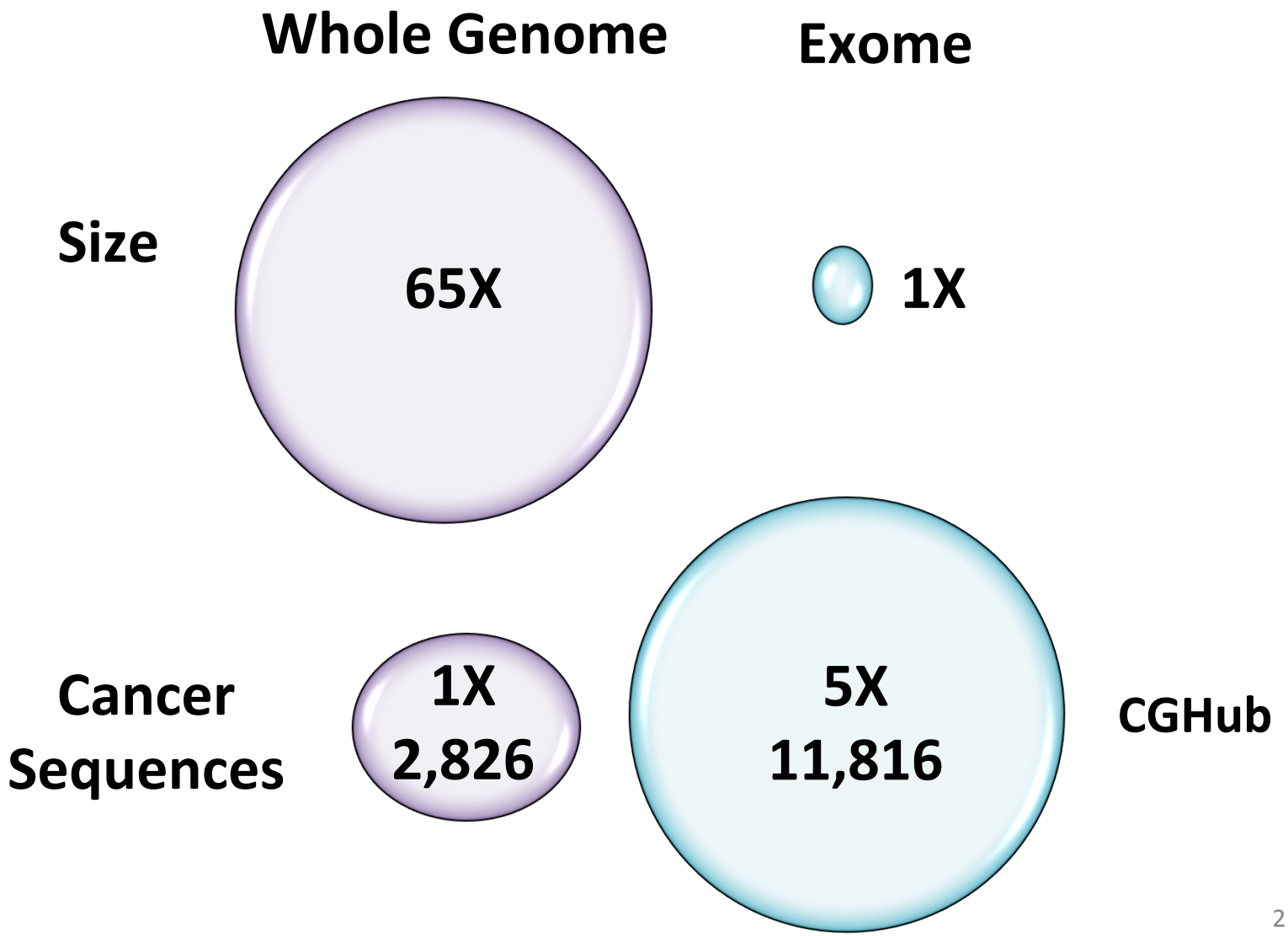
Recurrent Somatic Mutations in Regulatory and Other Regions of Human Cancer Genomes

Michael Snyder
Stanford University
June 9, 2016



Most Effort on Cancer is Performed on Exomes

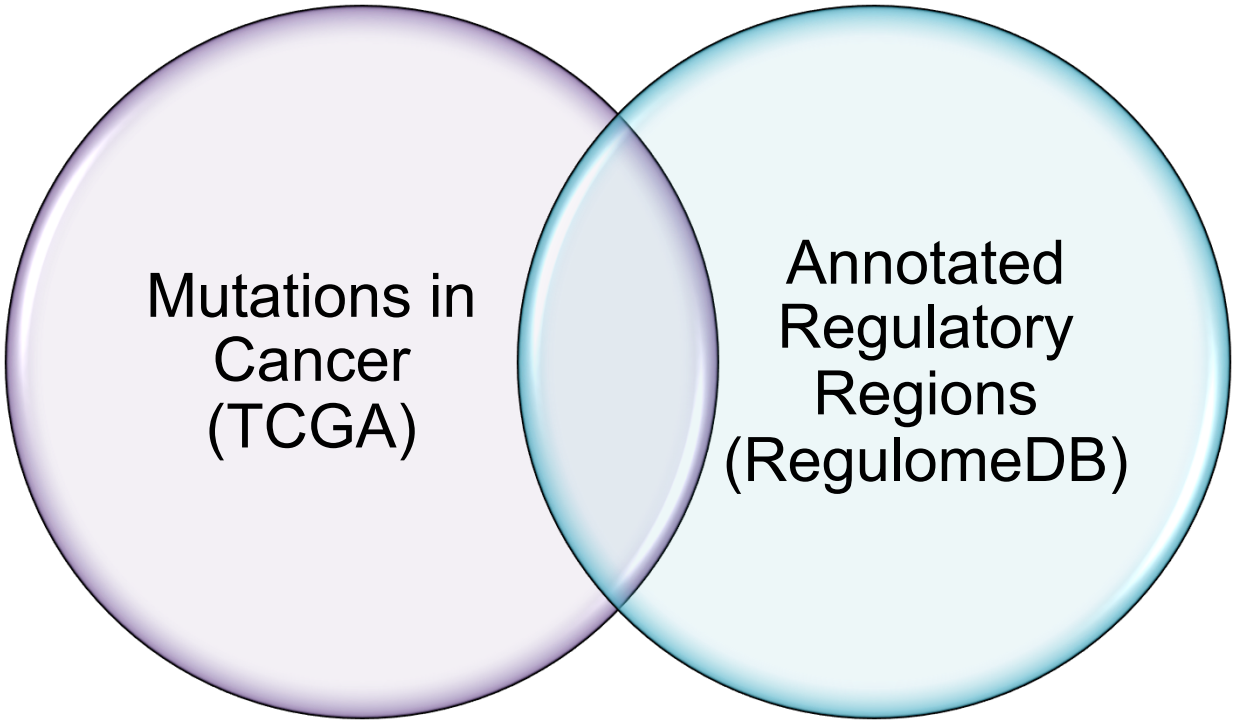


Two Approaches

1) Look for mutation enrichment using ENCODE data and fixed windows

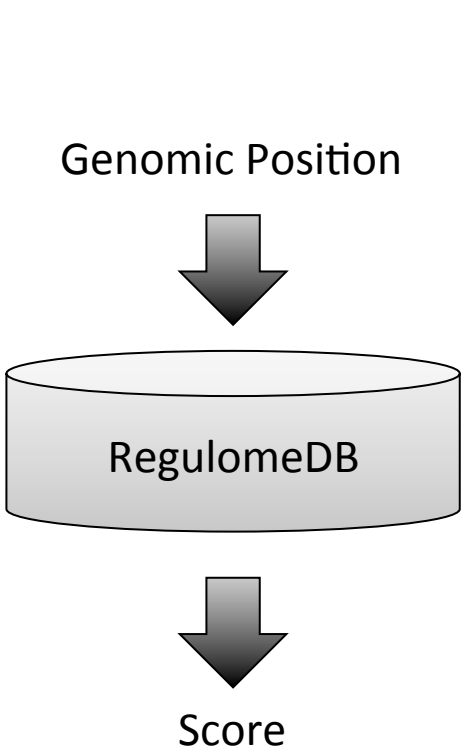
2) Look for increased density of mutations in variable sized windows

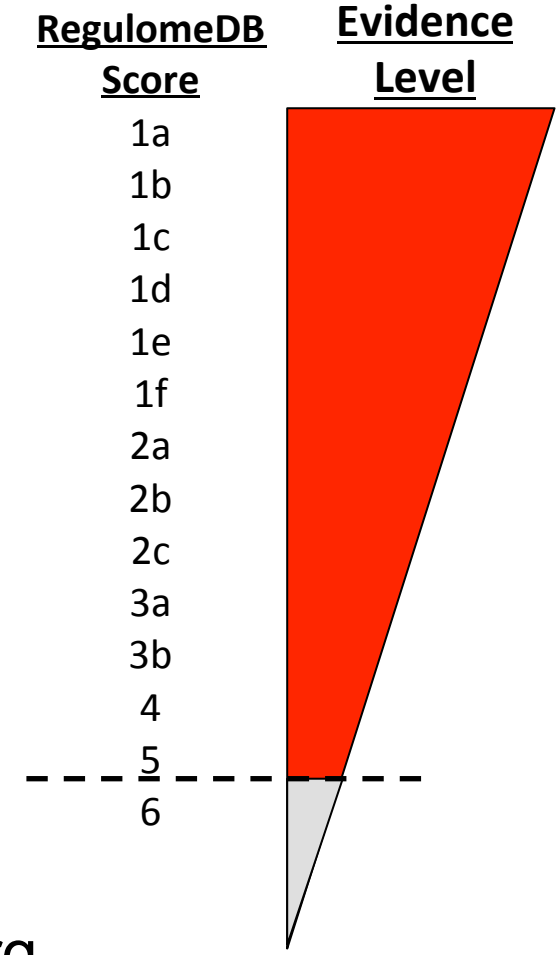
Are there functionally important mutations in regulatory regions in cancer?



Collin Melton

RegulomeDB Scores Are Used to Annotate Mutations with Regulatory Information



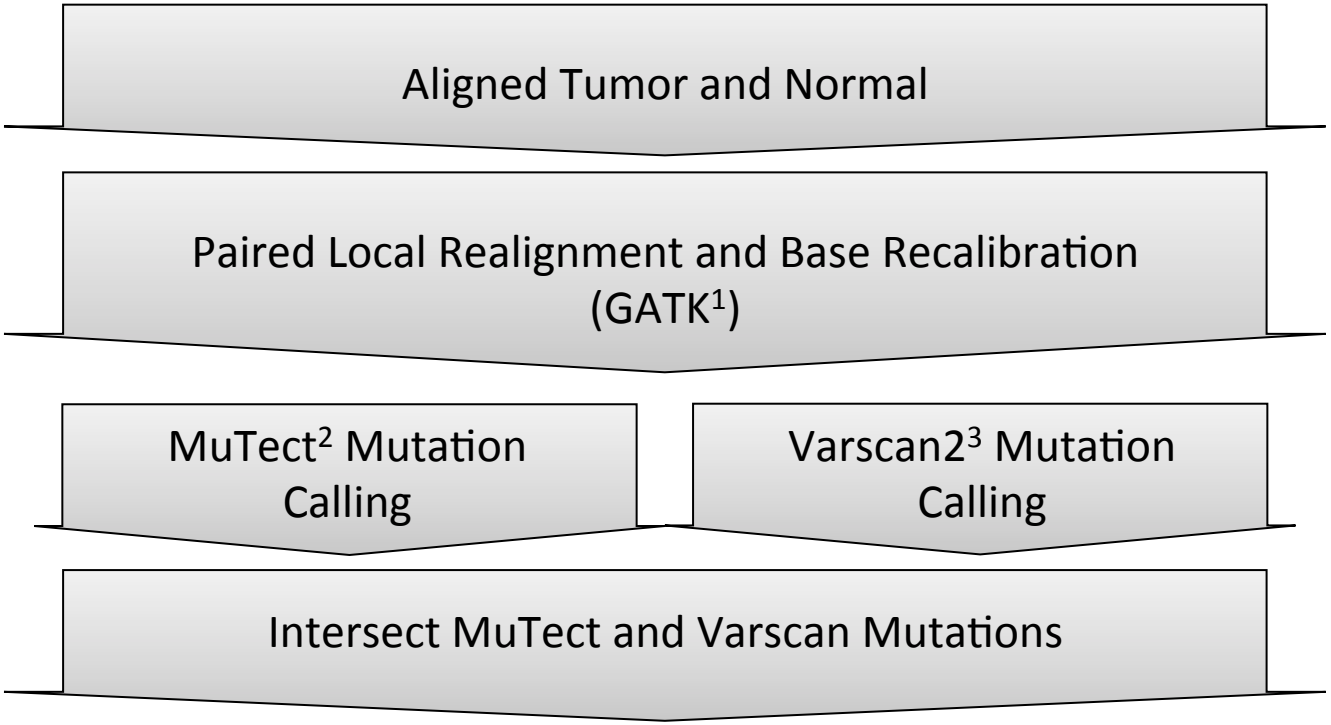


At least experimental evidence from ChIP Sequencing or DNAse Hypersensitivity Assays

http://www.regulomedb.org

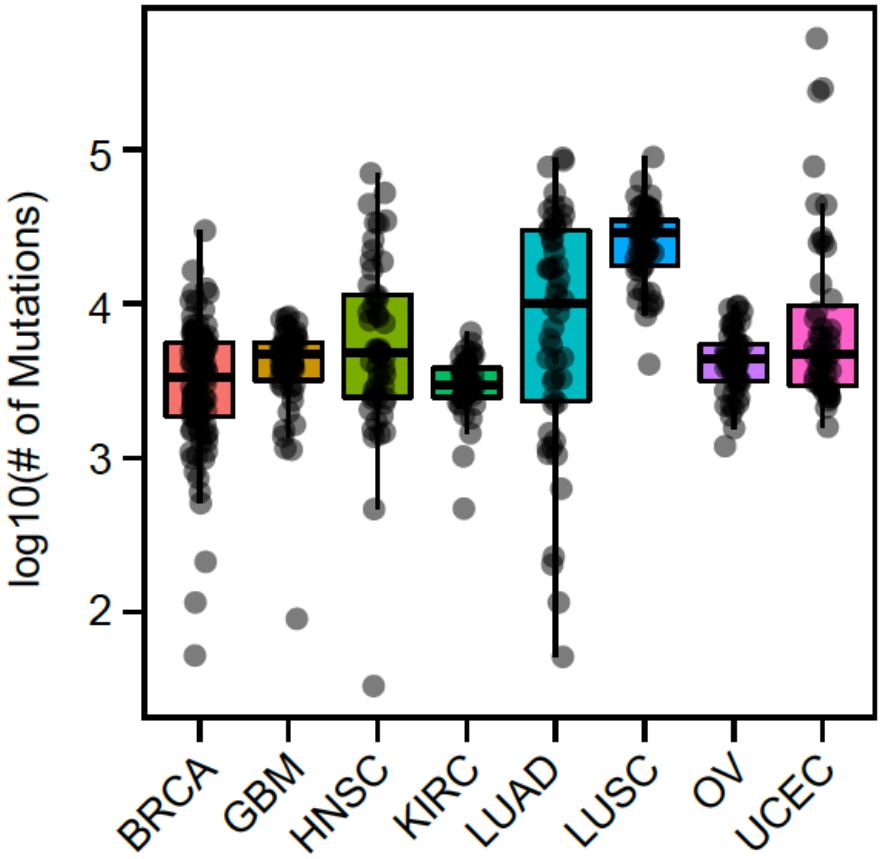
Boyle et al. Genome Res. 2012.

Workflow to Identify Cancer Mutations from WGS Data: 436 Genomes



1. McKenna et al. Genome Res. 2010.
2. Cibulskis K et al. Nat Biotechnol. 2013.
3. Koboldt, D. et al. Genome Res. 2012.

Identified Mutations in 436 Individuals from 8 Cancer Types



BRCA: Breast Cancer
GBM: Glioblastoma Multiforme
HNSC: Head and Neck Squamous Cell Carcinoma
KIRC: Kidney Clear Cell Renal Carcinoma
LUAD: Lung Adenomcarcinoma
LUSC: Lung Squamous Cell Carcinoma
OV: Ovarian Cancer
UCEC: Uterine Corpus Endometrial Carcinoma

Are Mutations Enriched in Regulatory Regions?

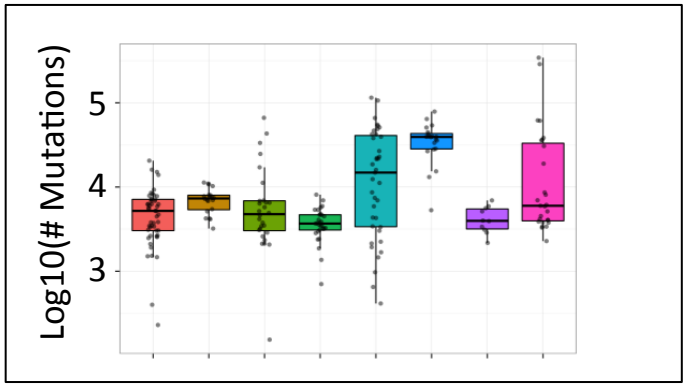
Fraction of Mutations in Regulatory Regions

vs

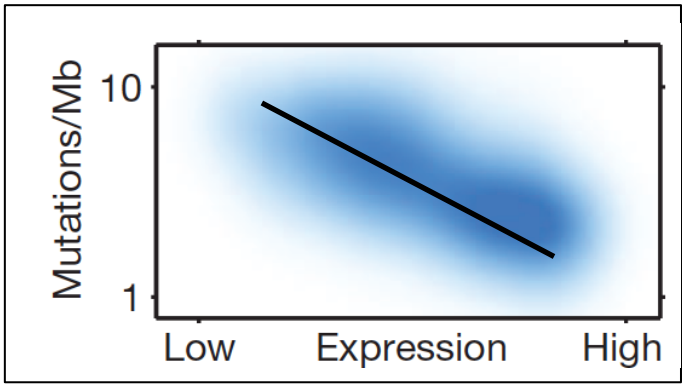
Fraction of Simulated Mutations in Regulatory Regions

Need to Simulate Mutations with Matched Covariates because Mutation Probability is not Uniform Across Samples and Genomic Sites

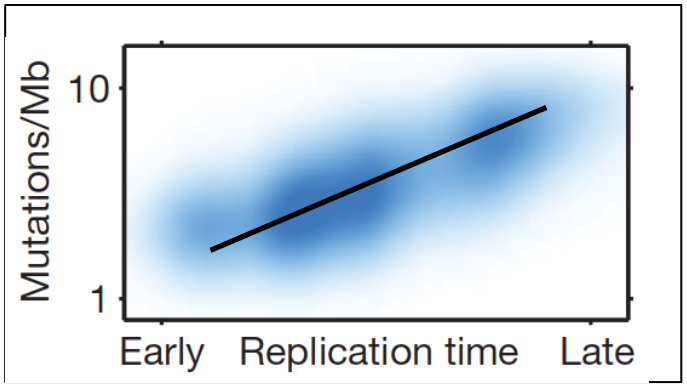
Variation Across Samples



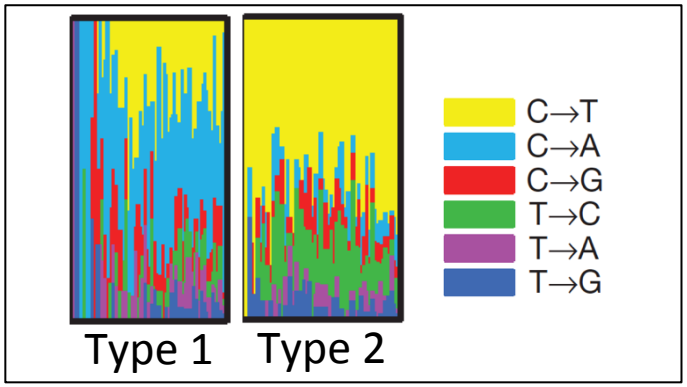
Variation With Expression



Variation with Replication Timing

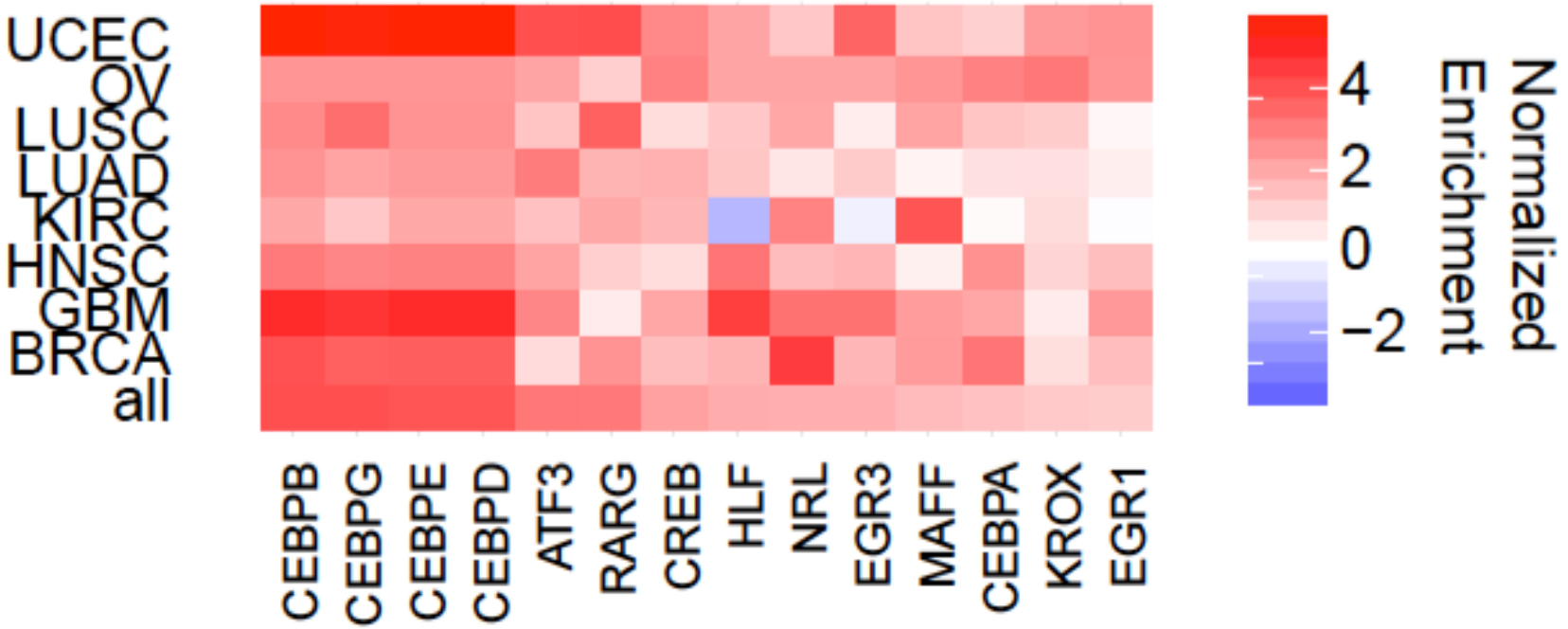


Variation With Base Type

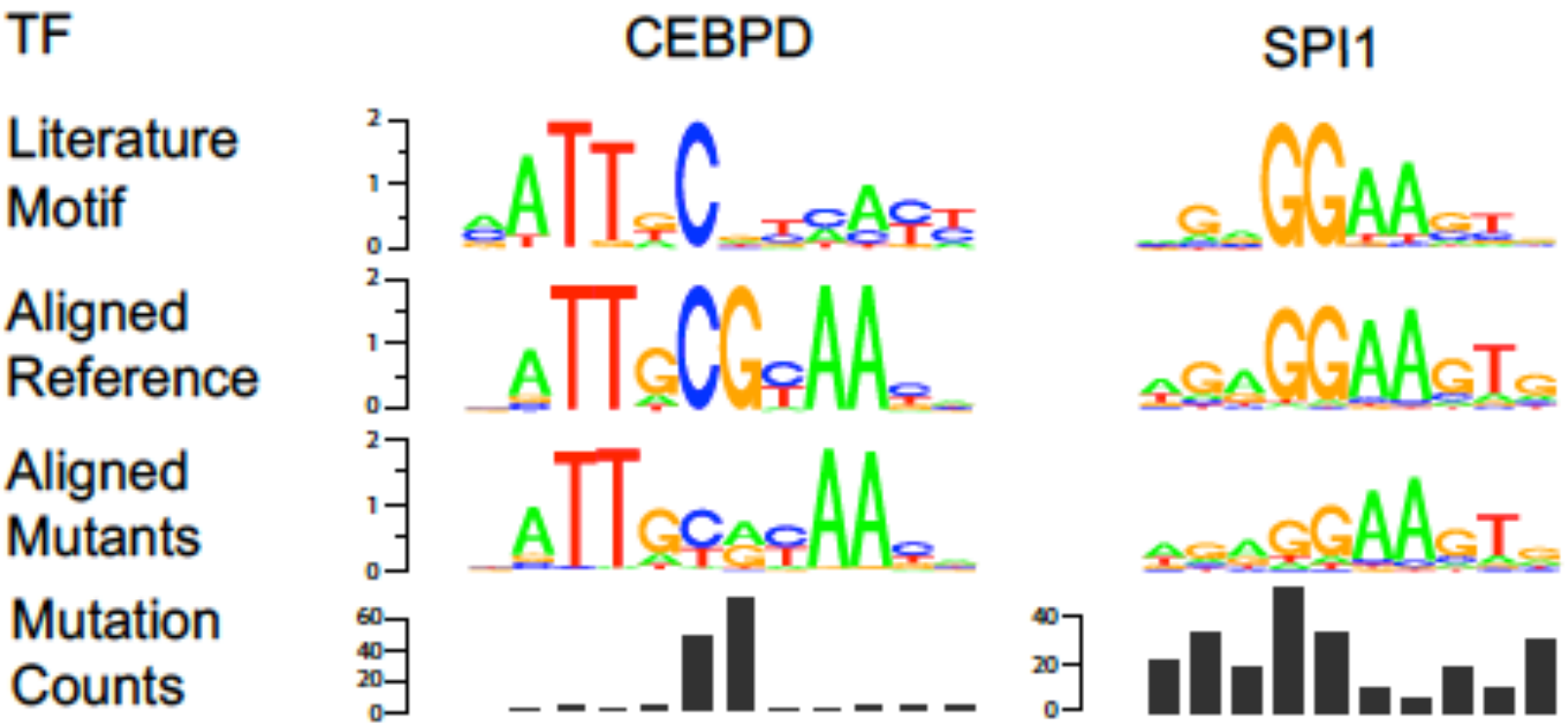


Lawrence *et al*. Nature 2013

General Characterization: Sites Bound by Specific Transcription Factors are Enriched for Mutations



Occasionally Specific Residues within a Motif are Selectively Mutated



Mutations in TF Binding Sites Usually Reduce the PWM Match Score

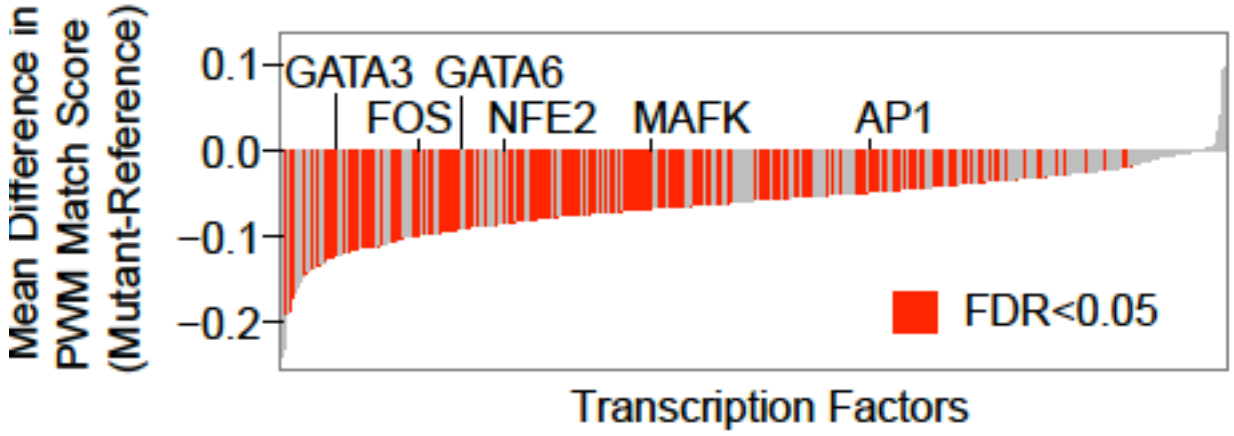
TF CEBPD

Literature Motif

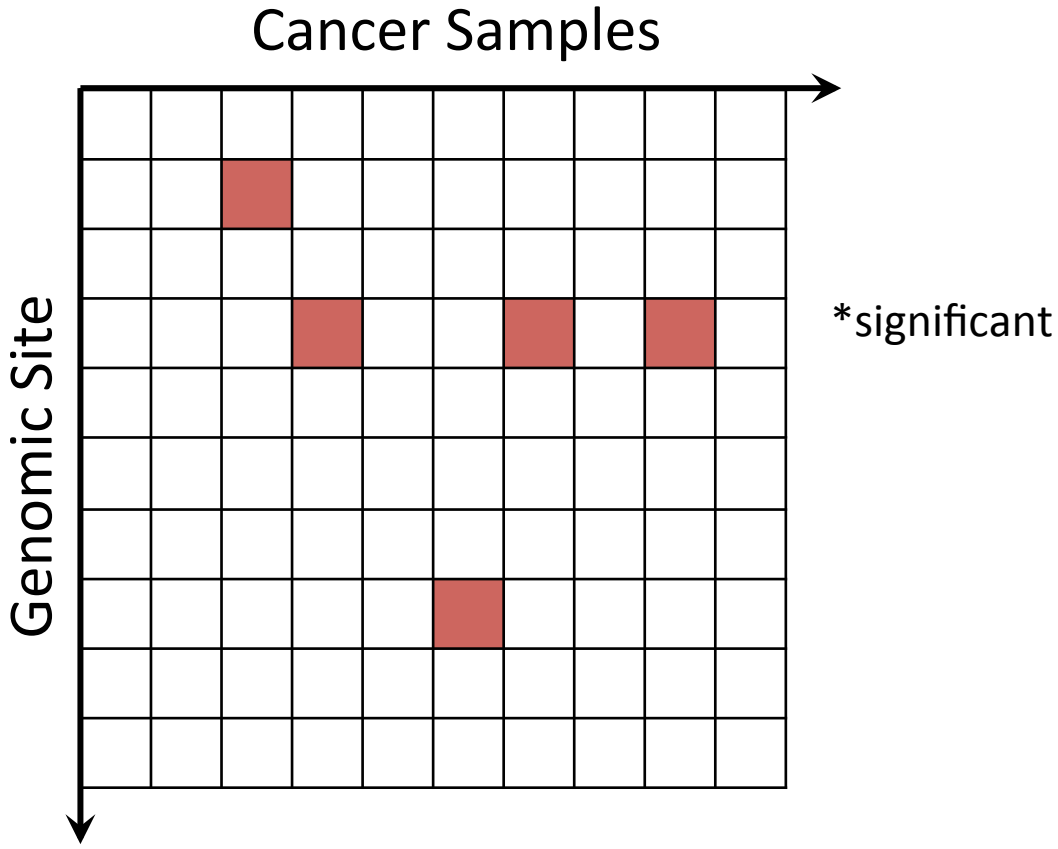
Ref: AATTGCGTCACT ⟶ Ref. Score

Mut: AATTGTGTCACT ⟶ Mut. Score

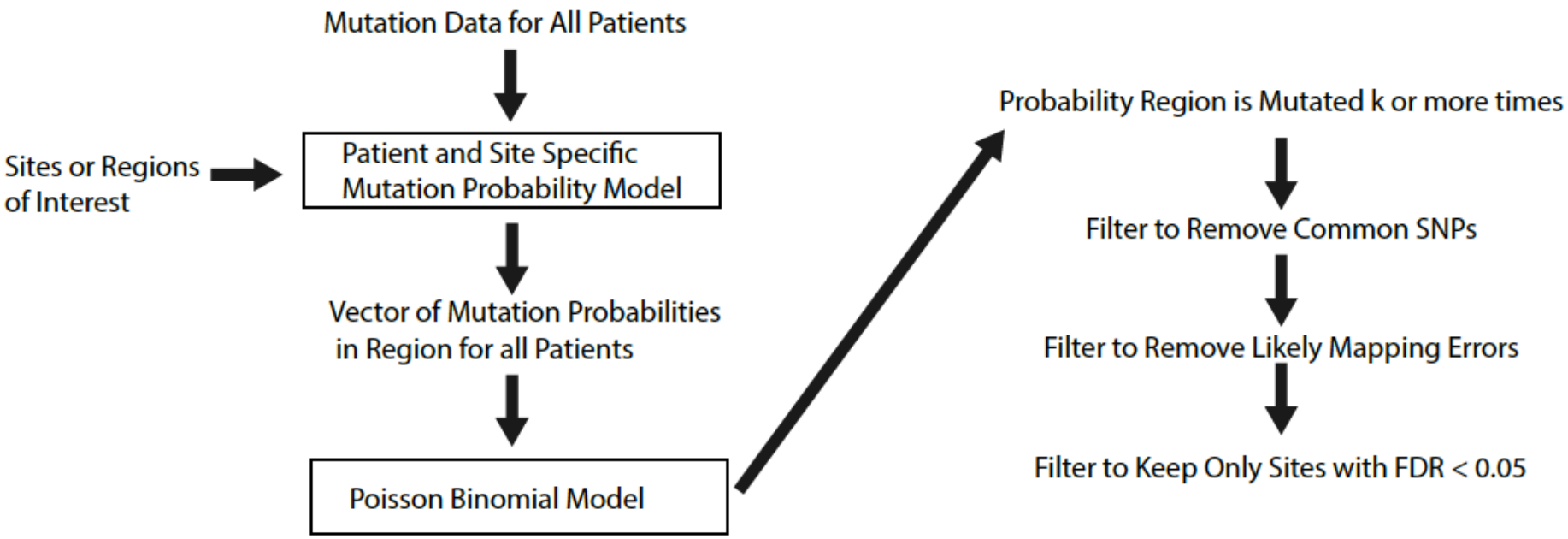
Random Mut: AATTGCGTAACT ⟶ Random Mut. Score



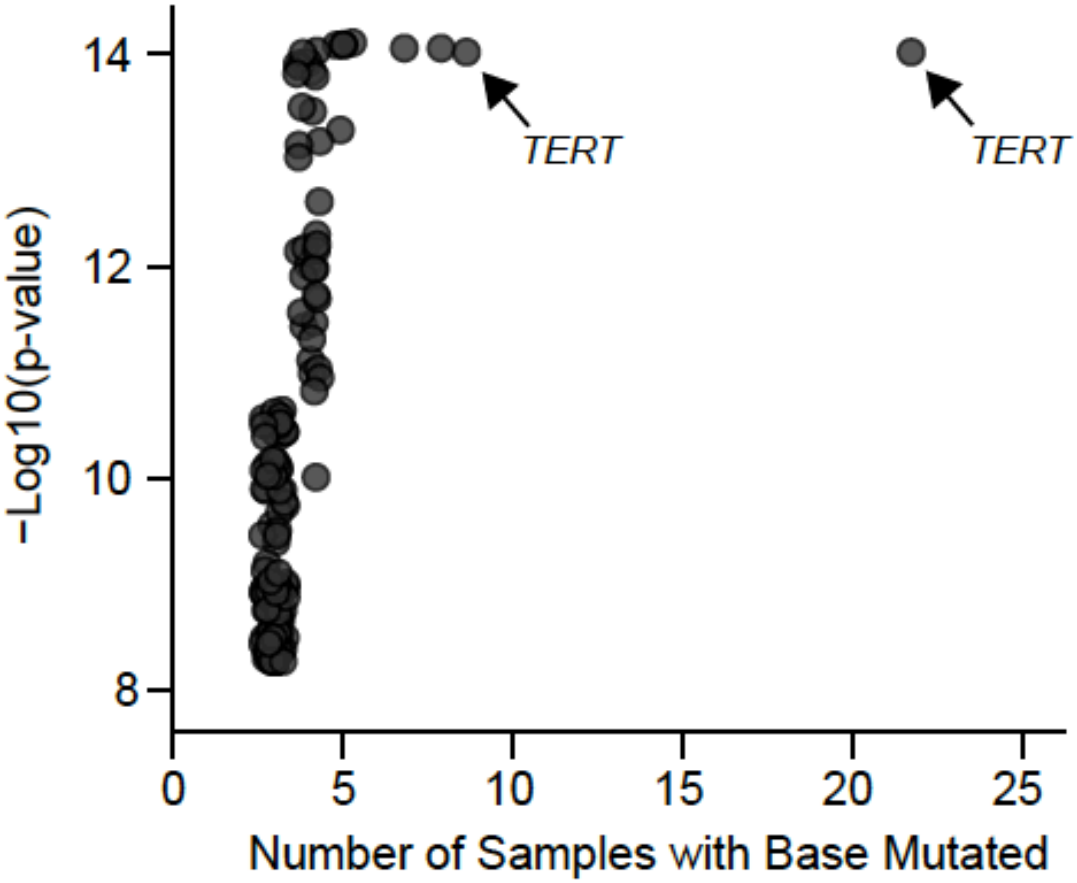
Are Mutations Recurrently Found at Specific Genomic Loci?



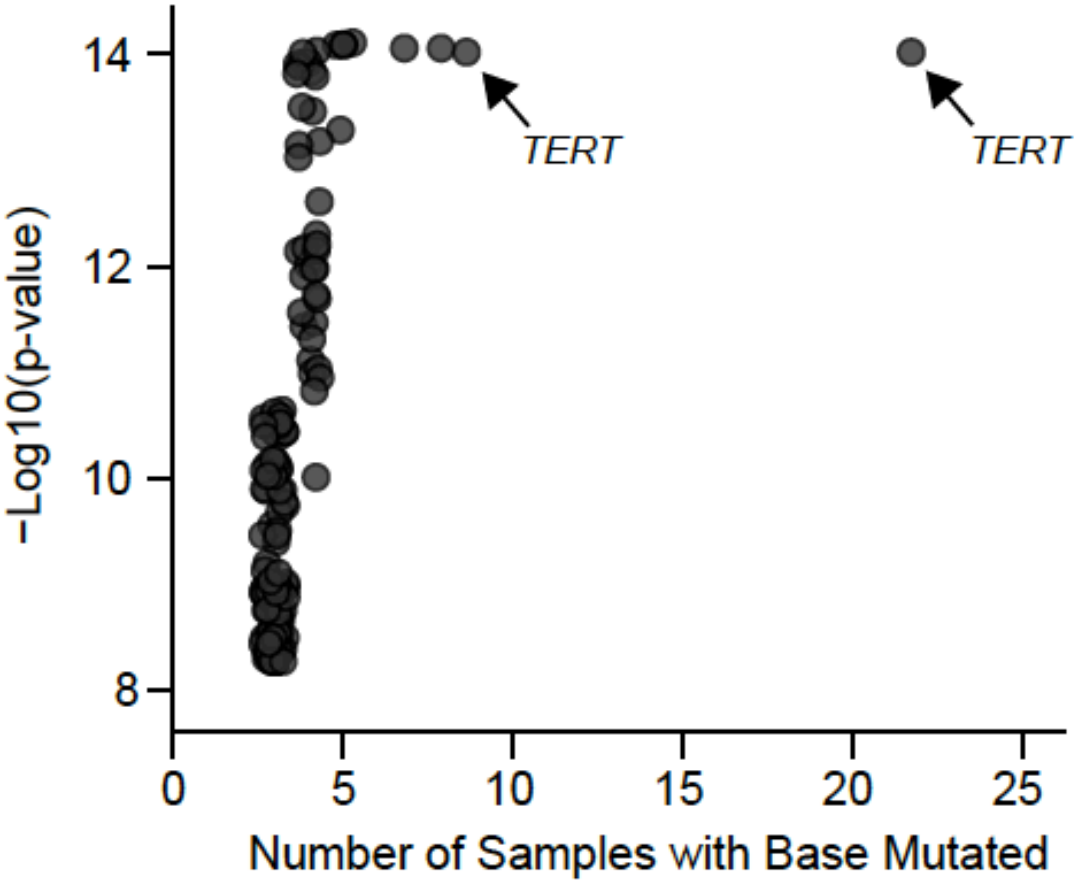
Approach to Identify Recurrently Mutated Sites



Significant Repeatedly Mutated Loci (~123 Regulatory)

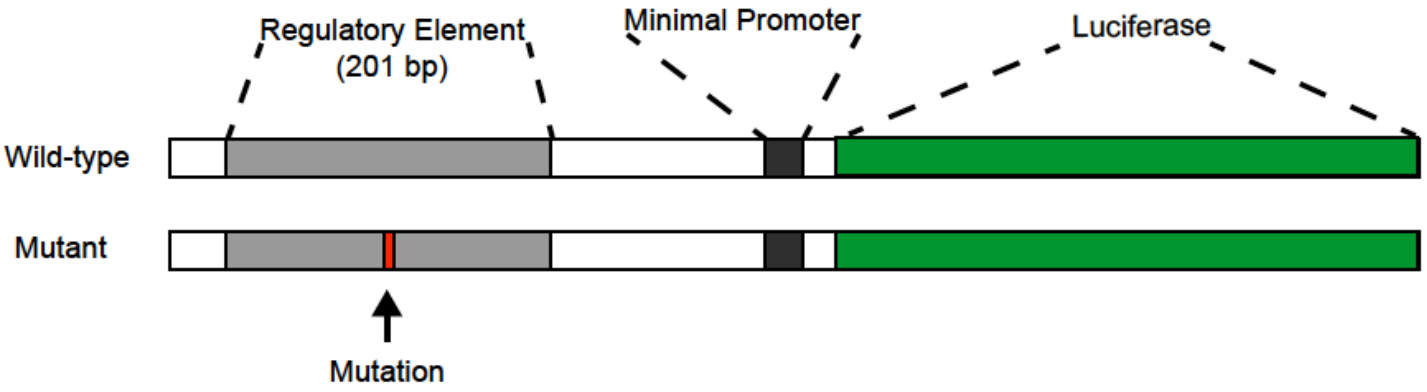


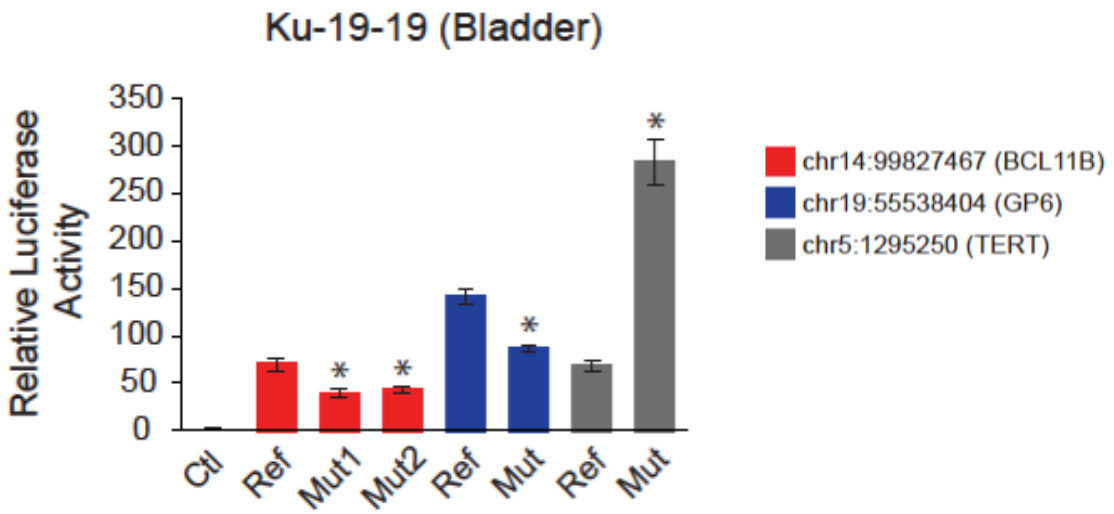
Significant Repeatedly Mutated Loci (~200)



Cancer Associated Genes in Vicinity of Regulatory Mutations
GNAS, INPP4B, MAP2K2, BCL11B, NEDD4L, ANKRD11, TRPM2, P2RY8

Mutations Alter Enhancer Function in Validation Assays





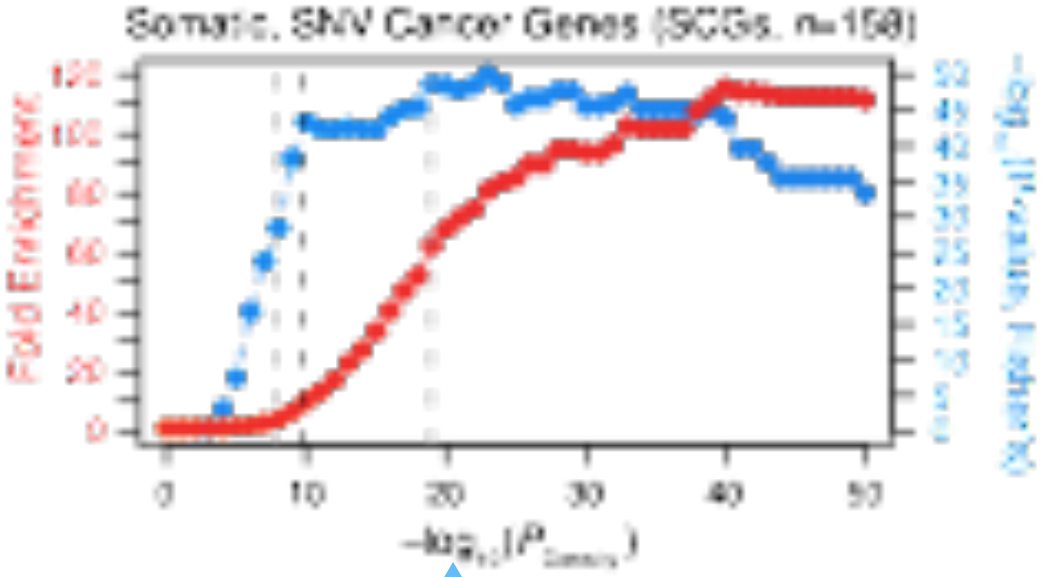
Two Approaches

1) Look for mutation enrichment in fixed windows

2) Look for increased density of mutations in variable sized windows

1. DBSCAN
2. ~4500 exomes
3. Corrections for sequence, replication etc

Density scores strongly enrich for known cancer genes

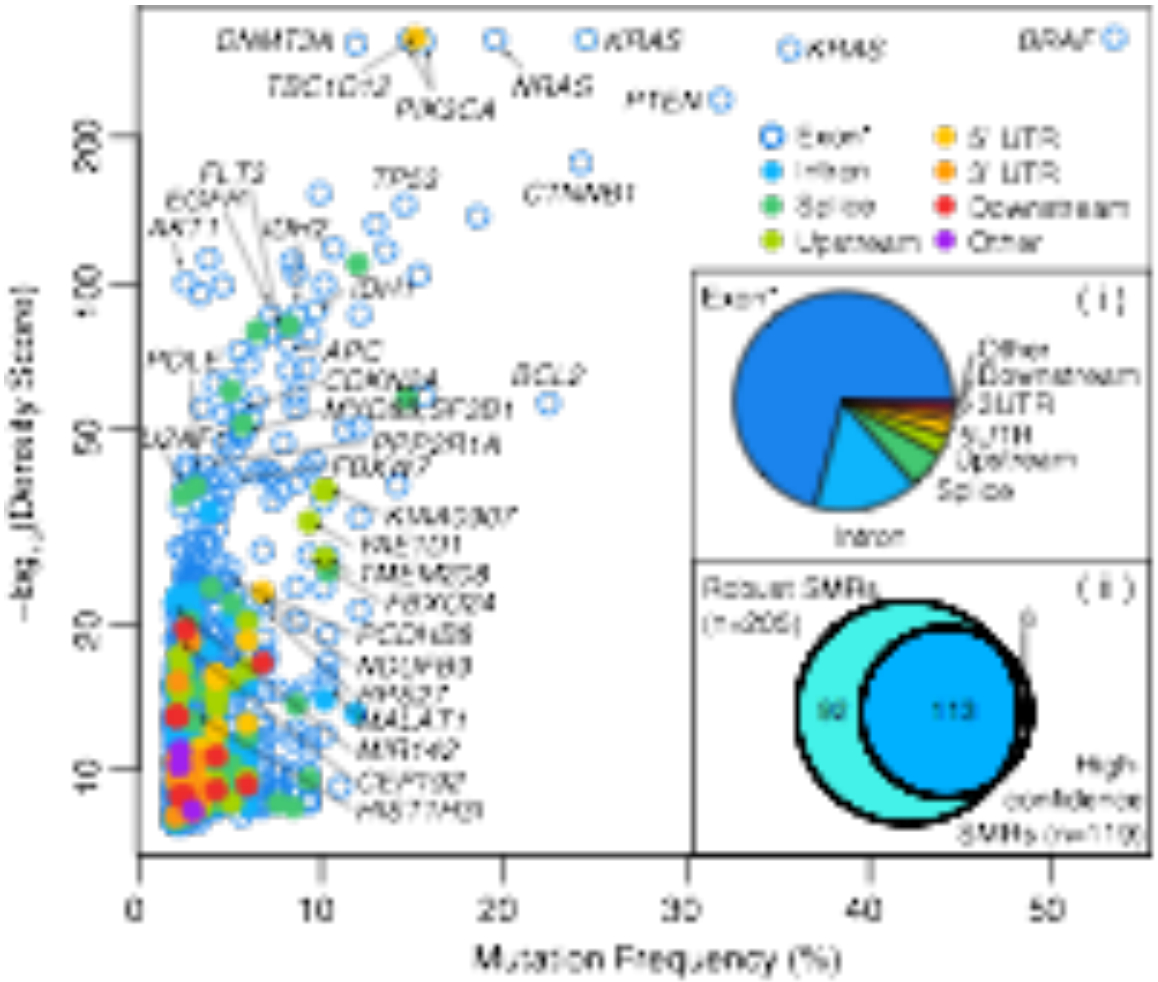


Modeling | Clustering | Scoring | Simulation | FDR (5%) | Analysis

- Computed density scores permit strong selective enrichments for somatically-altered *Cancer Gene Census* (CGC) genes:
  - For **somatic cancer genes (SCGs)** affected by point mutations (*n*=158) enrichments climb **up to 120x**.
  - Even at extreme density scores: **10% of genes are novel cancer drivers**.

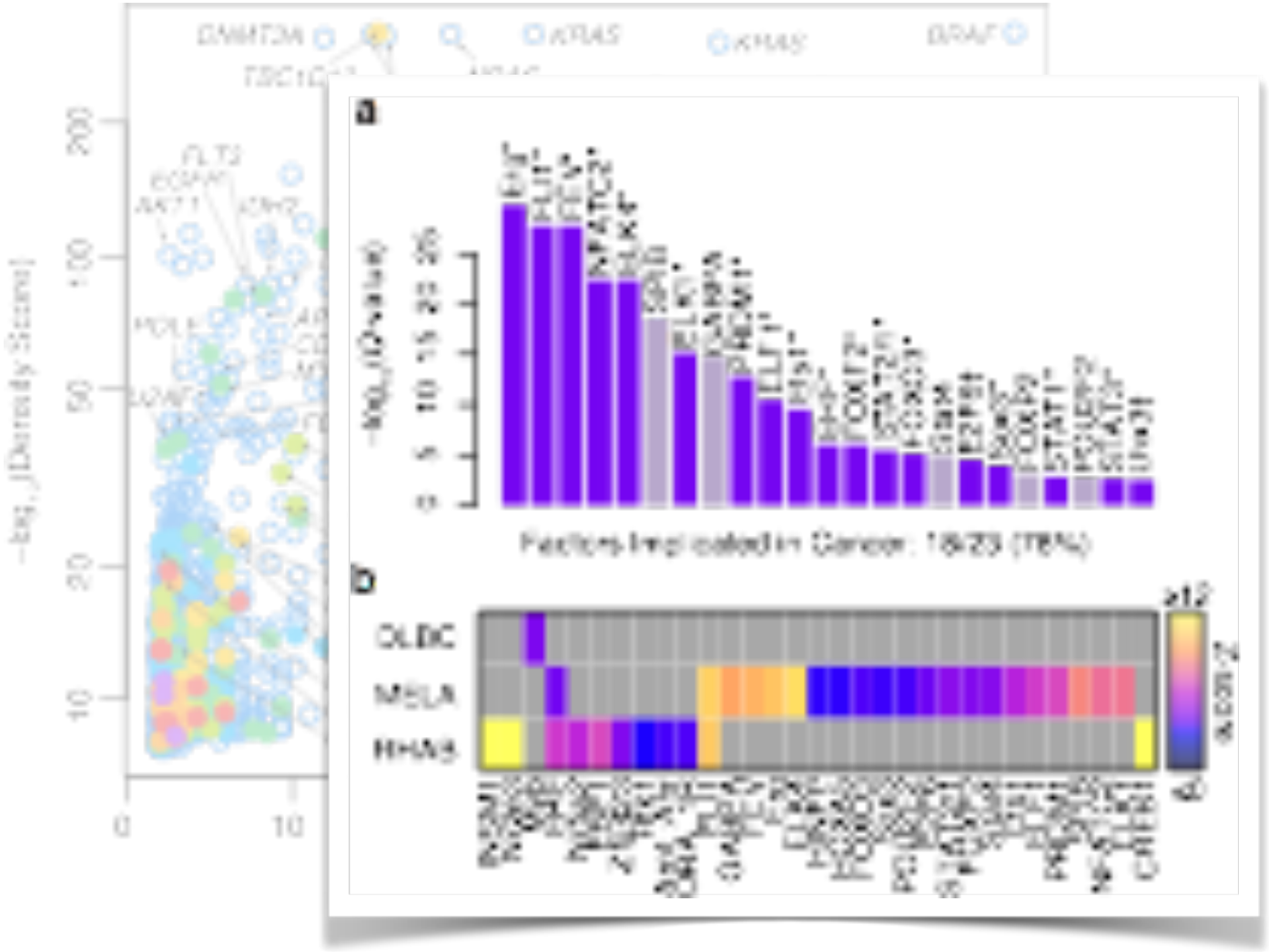
claraya@stanford.edu @claraya

Significantly mutated regions (SMRs) highlight the varied mechanisms of oncogenic misregulation



- SMRs classified into 185 high-, 496 medium-, and 191 low-confidence sets with corresponding (63x, 6.5x, 5.0x) enrichments for known cancer genes.

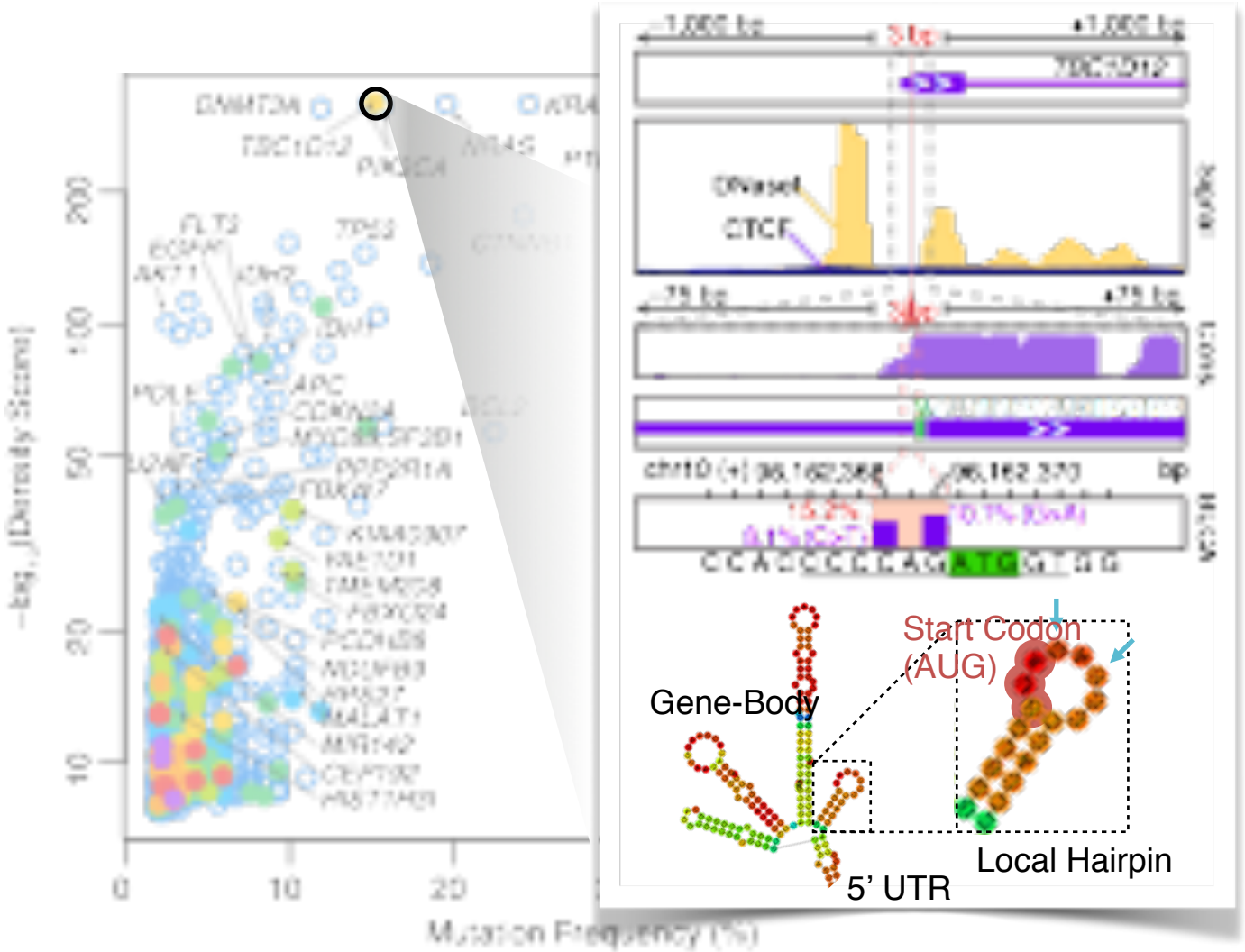
Oncogenic mutations often alter regulatory binding sites



- Sequence enrichment in small (≤25 bp) SMRs reveals enrichments for cancer-associated TFs (*Pscan*)

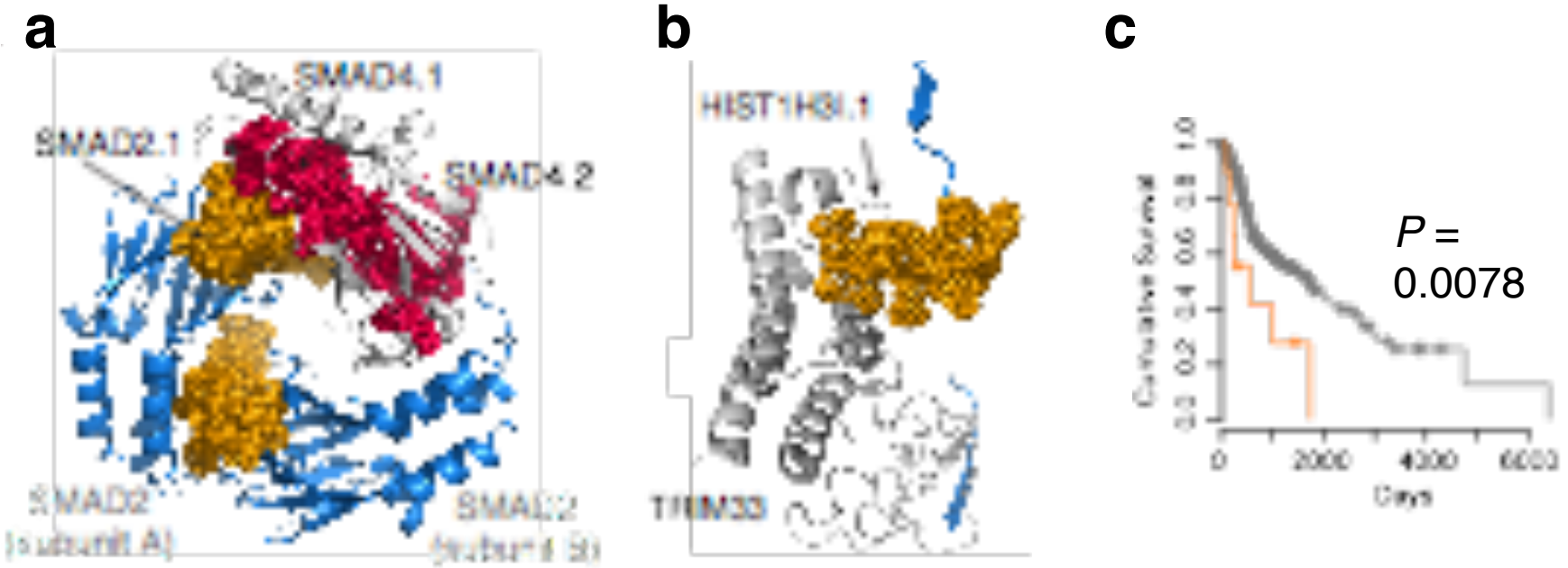
claraya@stanford.edu/claraya

Oncogenic mutations often alter non-coding/regulatory sites



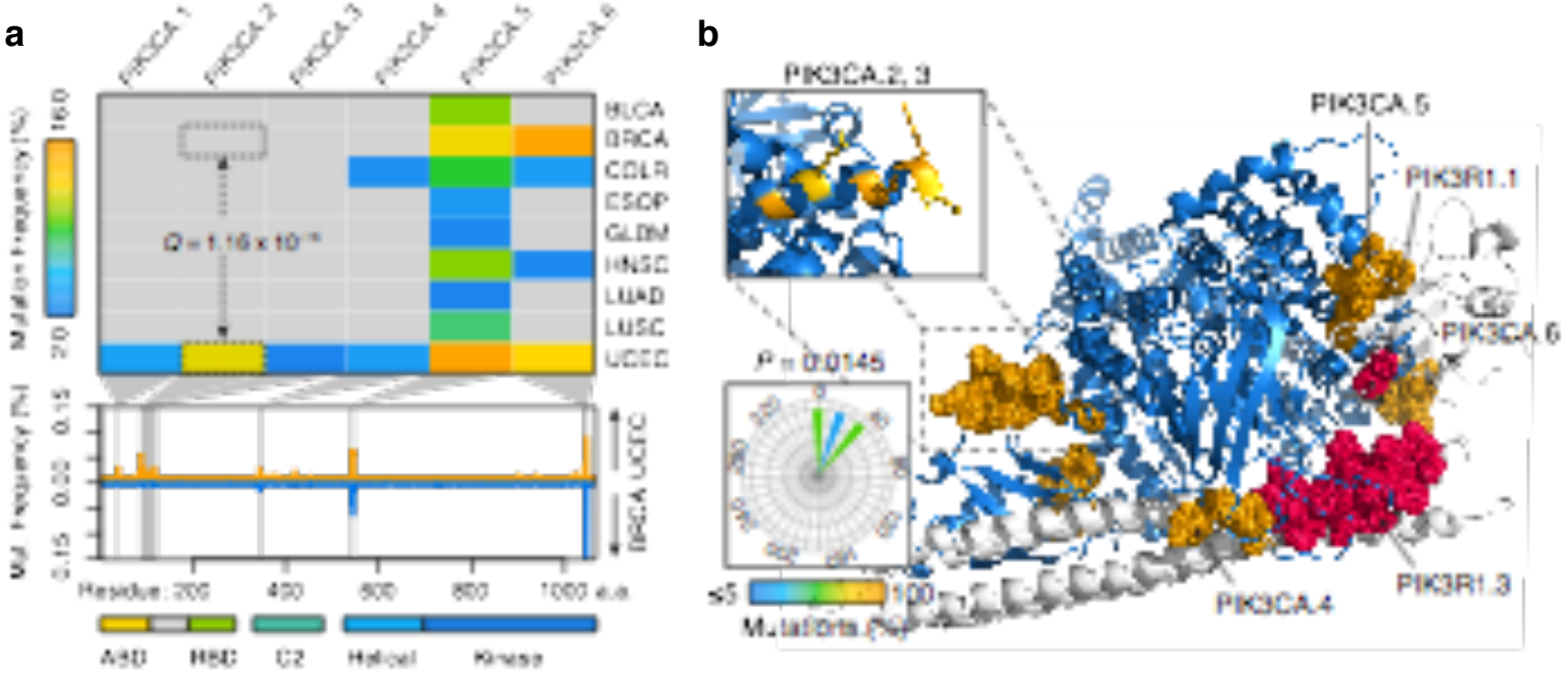
* 18% of bladder cancers harbor transitions in the 5' UTR of *TBC1D12*. (Mutated in WGS of 7 cancer types, including 2/20 bladder cancers, 2/40 lung adenomas, and 3/172 breast cancers.)

Structural mapping reveals recurrently targeted interfaces



- SMRs localized to ***n*=17** interfaces of protein-protein or DNA-protein interactions (**15/17** are known cancer-driver interfaces).
- SMRs pinpoint a novel interface of oncogenic alteration in a **histone H3/TRIM33** interface (with putative survival implications).

SMRs in *PIK3CA* are differentially mutated across cancers



- **6 SMRs** in the catalytic subunit (α) of the phosphoinositide 3-kinase oncogene, **PIK3CA** (p110α), including particularly recurrent alterations in the helical (**PIK3CA.5**) and kinase (**PIK3CA.6**) domains[1].
- SMRs (**PIK3CA.2, PIK3CA.3**) between the ABD and linker domains are affected in up **14% of endometrial carcinomas** (~7% inside the α-helix).
- **PIK3CA.2** (G106V, K111E) and **PIK3CA.3** (G118D) mutations have been shown to be kinase-activating.

claraya@stanford.edu @claraya

(1) Huang CH, et al. (2007). *Science* 318, 1744–1748

Conclusions

1) Can identify mutations in regulatory regions that are enriched for mutations

2) Often near genes involved in cancer

3) Mutations often disrupt binding motifs of specific TFs.

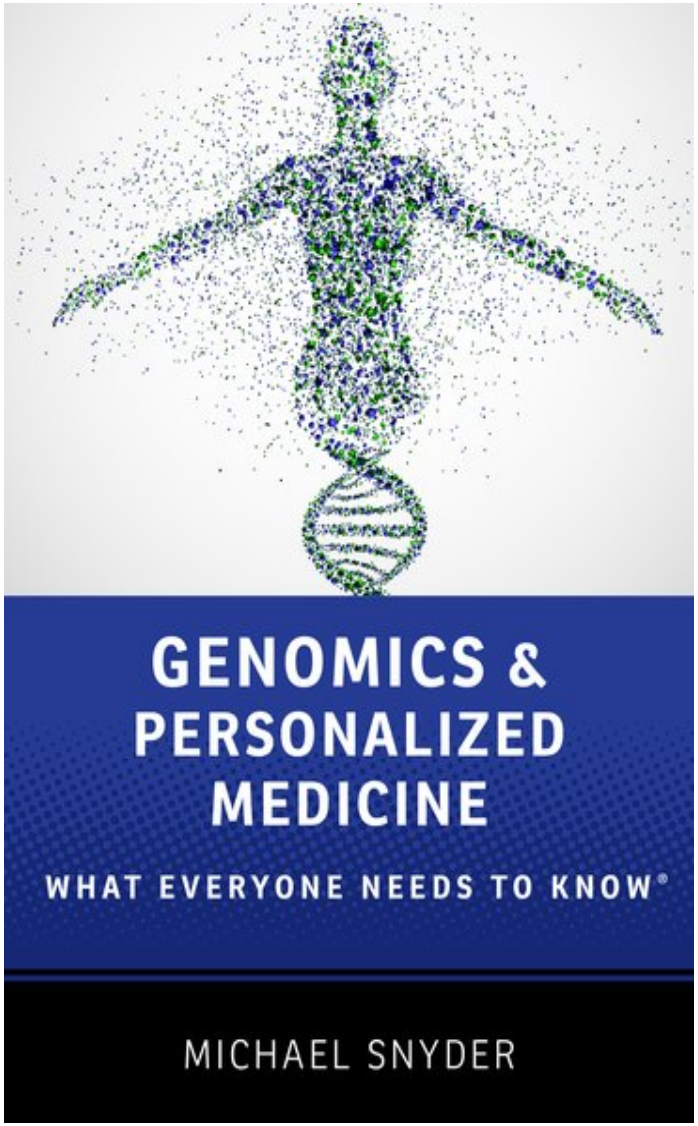
4) Using density analysis can find subsets of coding regions that are preferentially mutated in cancer; these can be cancer specific.

5) Will want to incorporate these into genome and clinical analyses

Acknowledgments

Collin Melton
Jason Reuter (reporter assays)
Damek Spacek
Carlos Araya
Can Cenik
Alan Boyle (RegulomeDB)
Will Greenleaf

TCGA and Volunteer Patients for Data



Genomics and Personalized Medicine

What Everyone Needs to Know®

Michael Snyder

Just released!

Available from Amazon.com