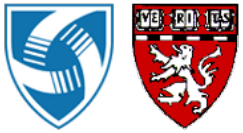
Using ENCODE to interpret mutational patterns in cancer

Shamil Sunyaev



Division of Genetics
Department of Medicine
Brigham and Women's Hospital / Harvard Medical School

Broad Institute of M.I.T. and Harvard

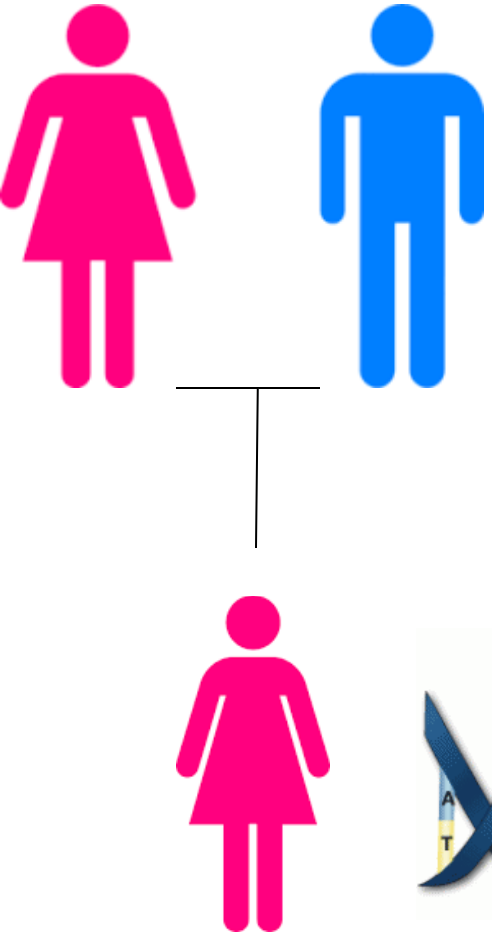
Disclosures

This work was funded by NIH

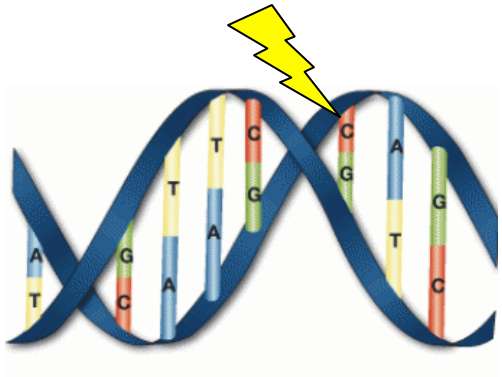
Nothing to disclose

Data on *de novo* mutations

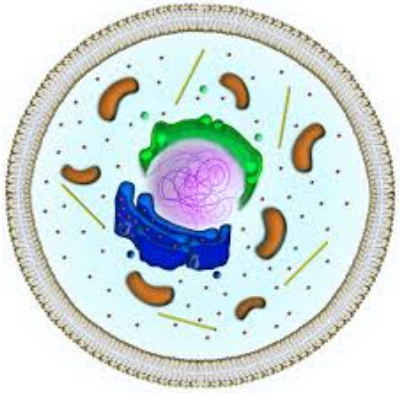
Germ-line mutations

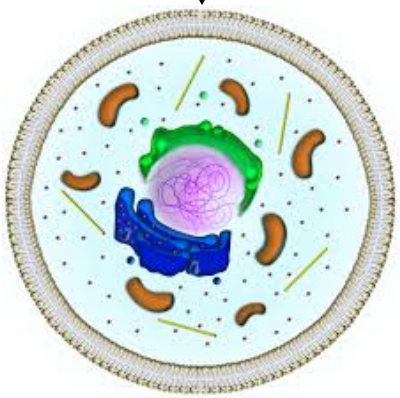


Change in DNA sequence

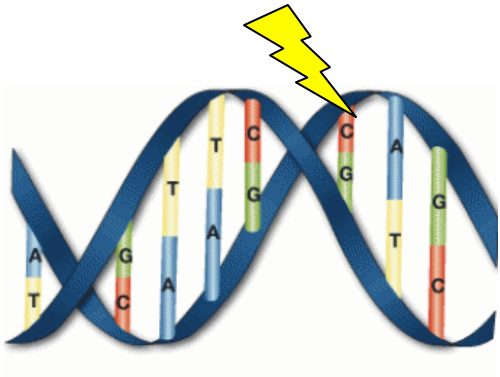


Somatic mutations





Change in DNA sequence



Why is this of any interest?

Statistical genetics of cancer → Methods

Evolution
- As a key parameter
- Of mutation rate

Biology
- Of DNA repair
- Of DNA replication

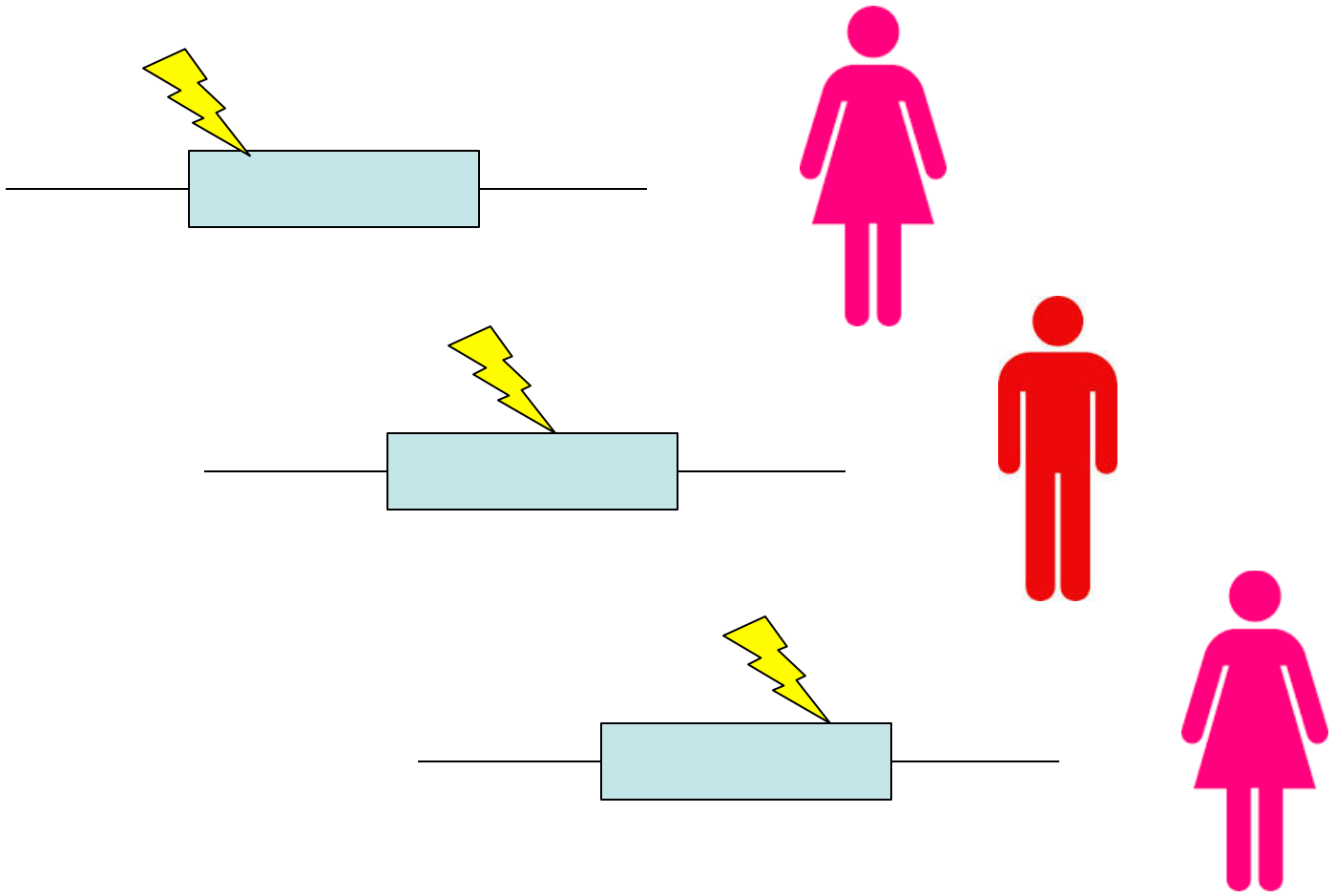
Genetic mapping and mapping by natural selection

Most of gene mapping methods (linkage, association) rely on recombination and are only applicable to sexual systems

Many methods to detect selection signals (selective sweeps, extended haplotypes) are similarly limited to sexual populations

What about cancer?

Identifying selected genes/functional units by recurrence



Everything is more difficult in search for non-coding drivers

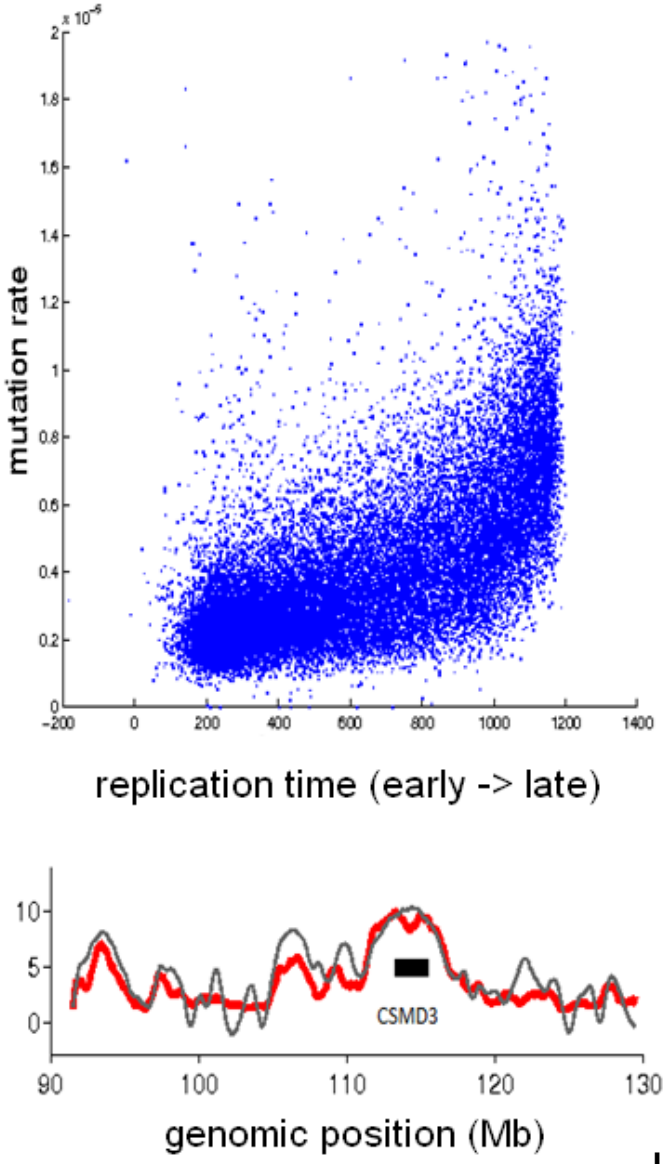
This signature of selection is completely (!!!) confounded by mutation rate variation

“Non-functional” regions may not serve as an ideal null model if mutation rate is correlated with “functionality”

Other samples may not serve as an ideal null model if mutation rate variation is sample-specific

All of this is exacerbated in the search for non-coding drivers!

Somatic cancer mutation density is associated with replication timing



Lawrence, *et al.*, *Nature* 2013

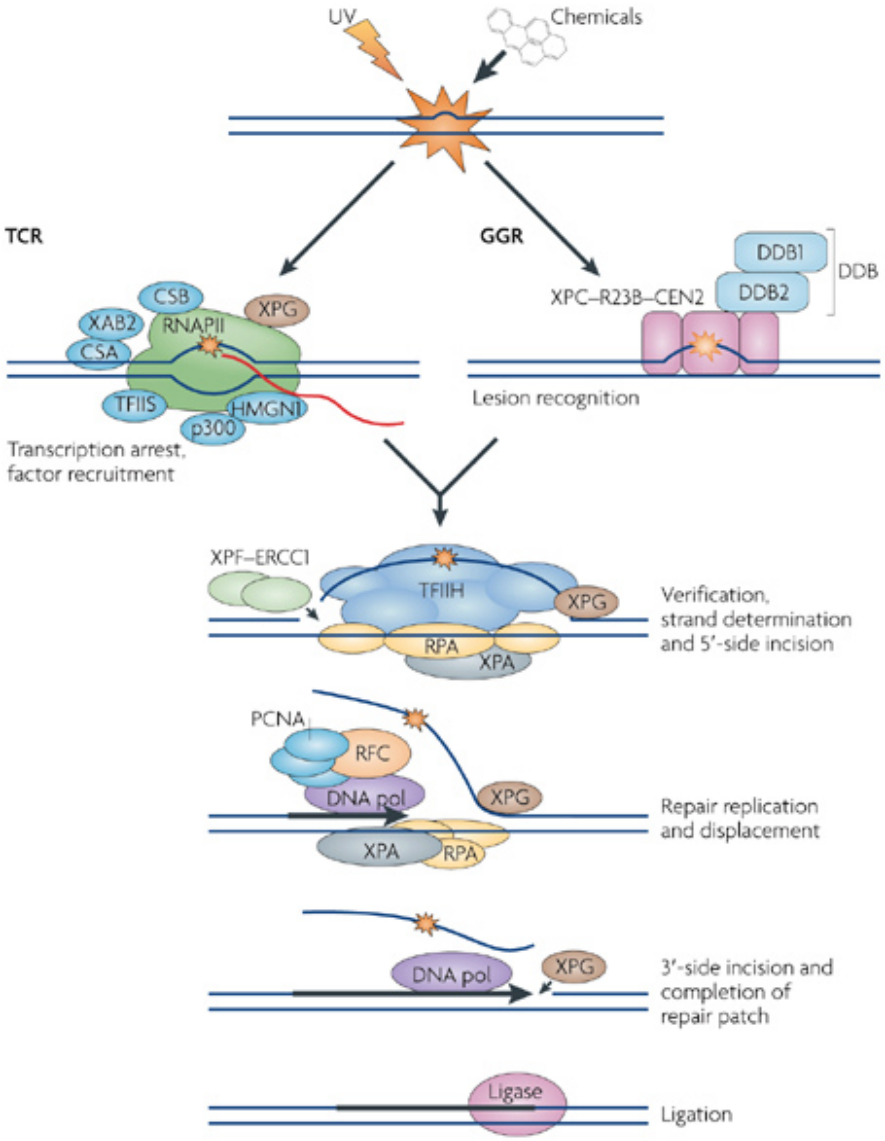
Somatic mutation rate depends on expression

Mutation rate is reduced in transcribed regions compared to intergenic regions

The reduction of mutation rate is proportional to expression level

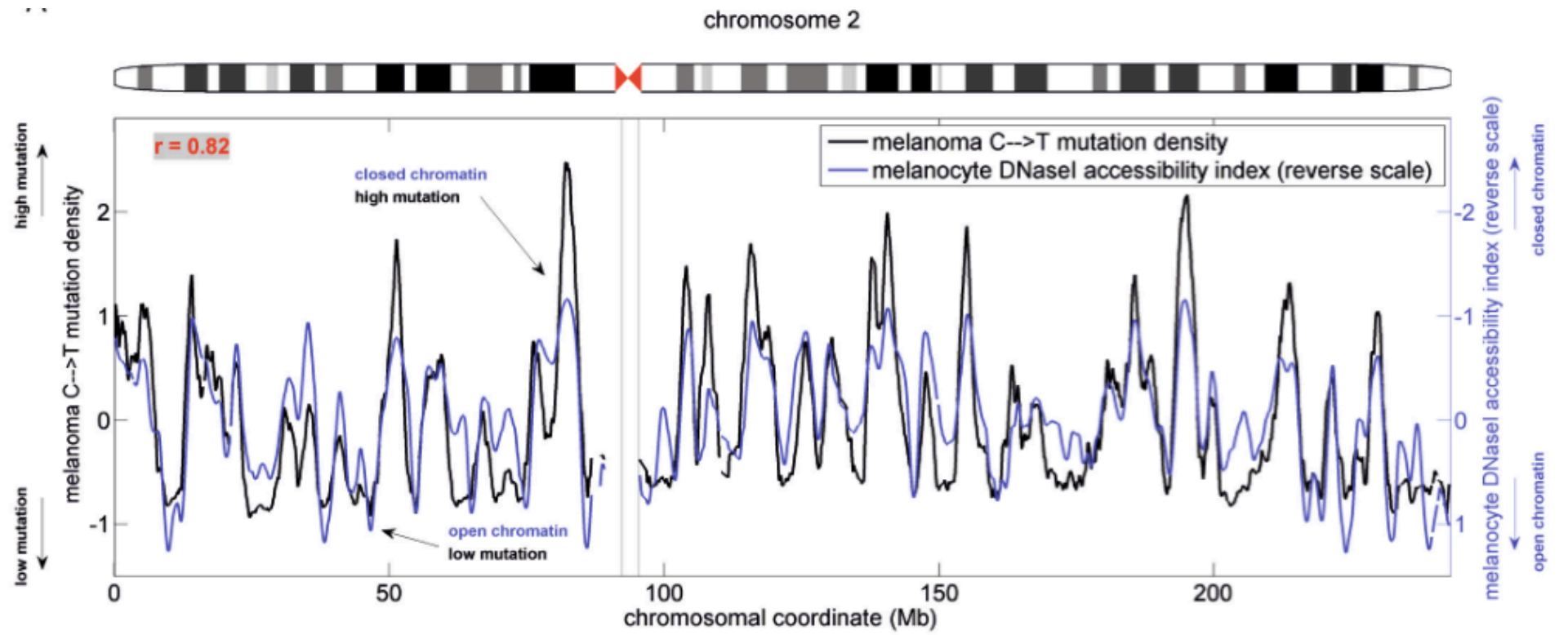
The effect is attributed to transcription coupled repair (TCR), which is supported by the strand bias

Hanawalt & Spivak, *Nat Rev Mol Cell Biol* 2008

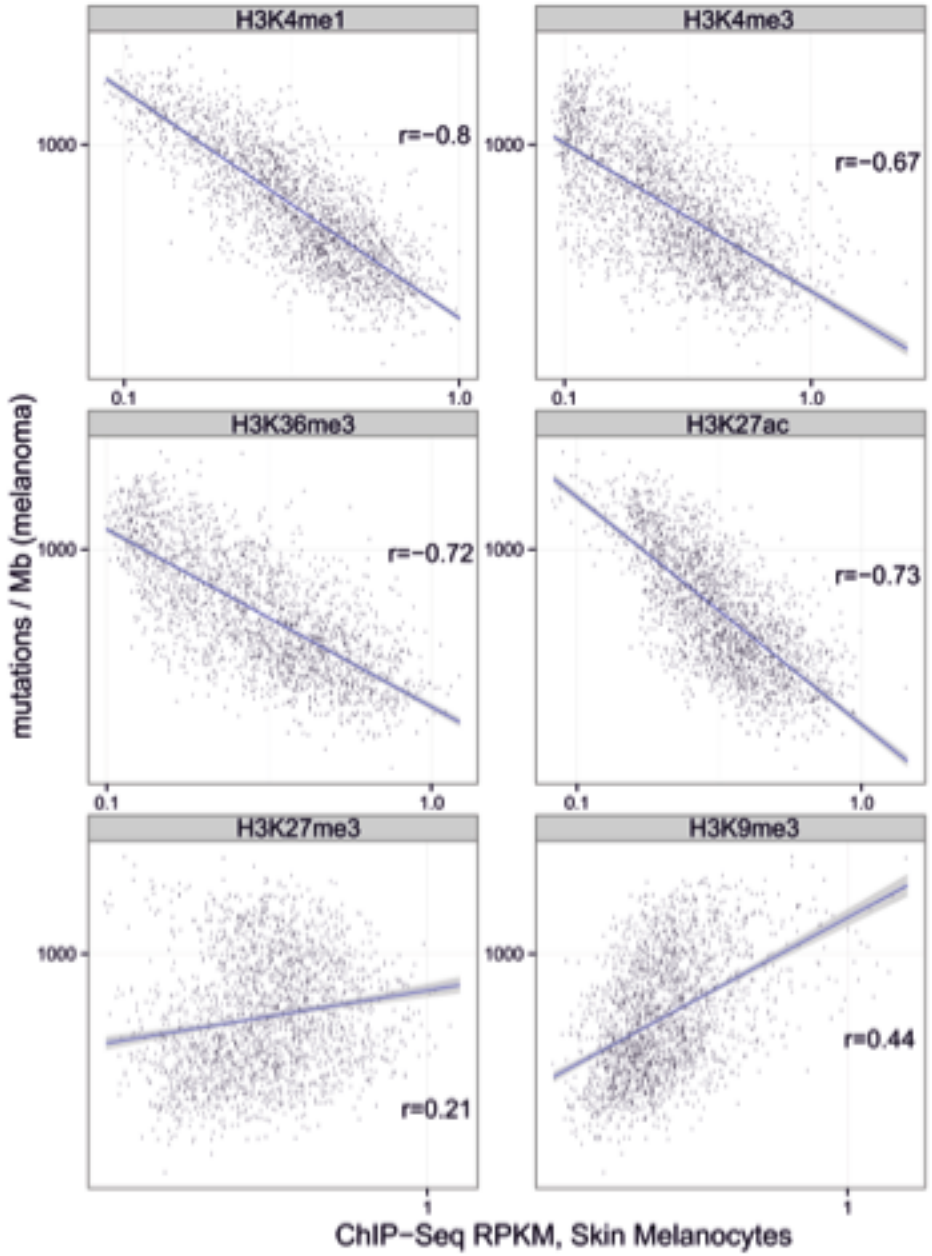


Nature Reviews | Molecular Cell Biology

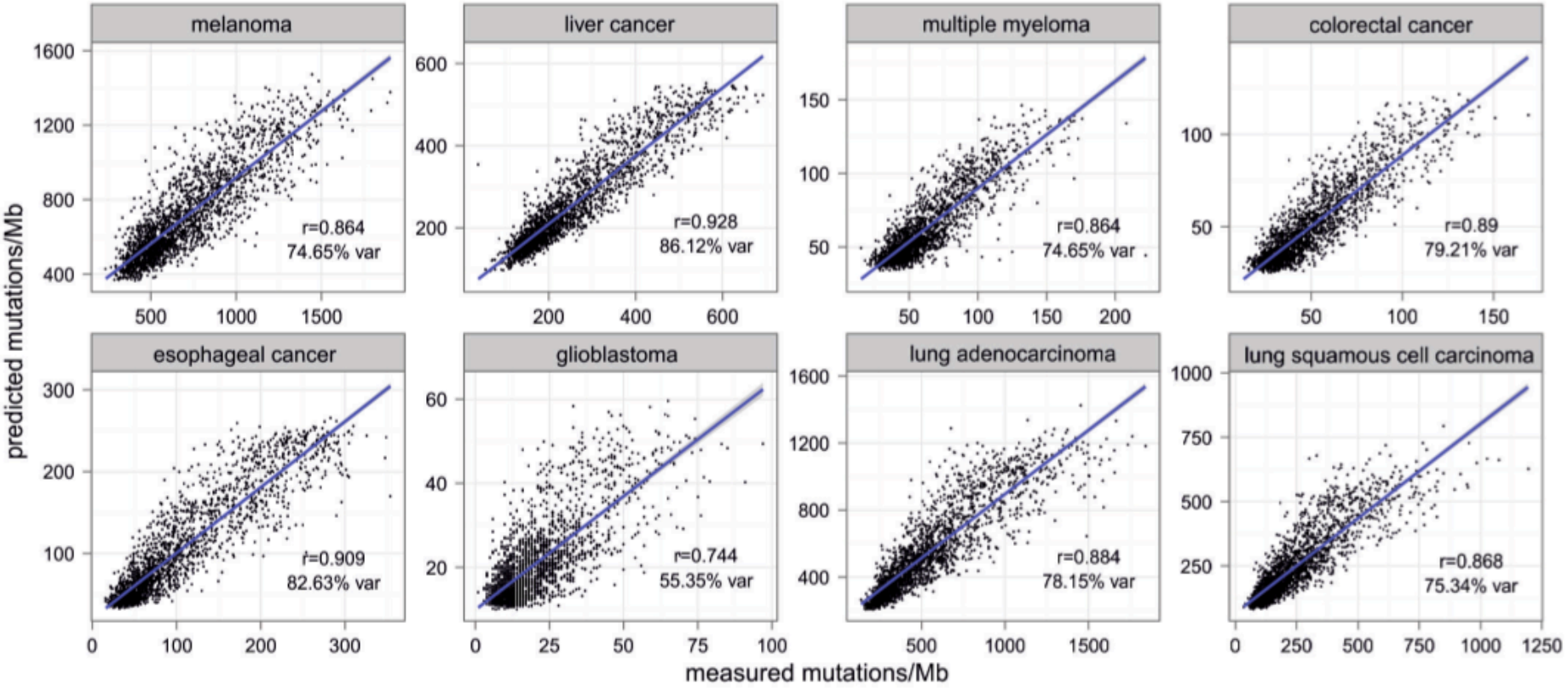
Predicting local mutation rate at 1Mb scale



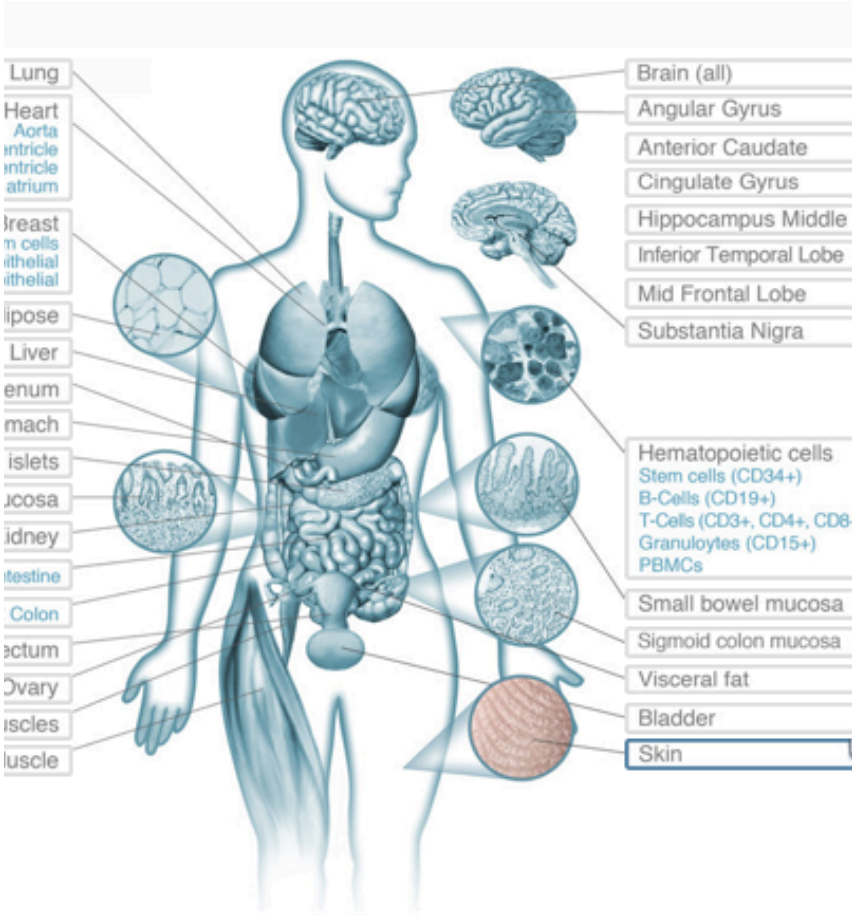
Polak, Karlic *et al.*, *Nature* 2015

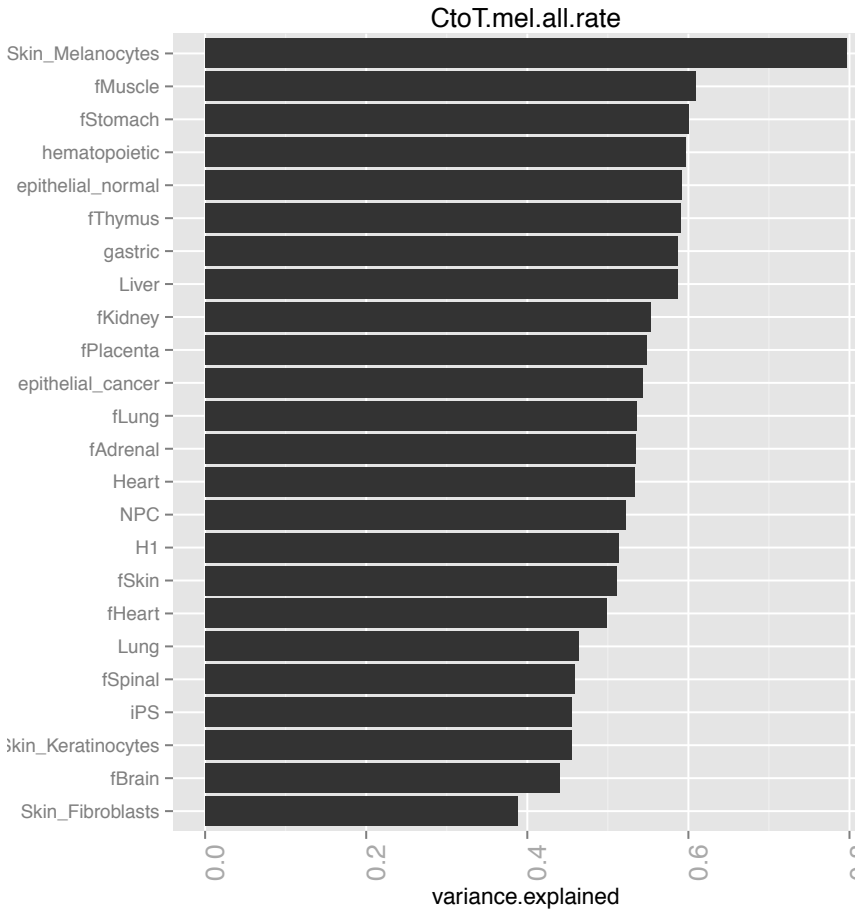

H3K4me1
r=−0.8
H3K4me3
r=−0.67
H3K36me3
r=−0.72
H3K27ac
r=−0.73
H3K27me3
r=0.21
H3K9me3
r=0.44
1000
0.1
1.0
1
mutations / Mb (melanoma)
ChIP−Seq RPKM, Skin Melanocytes

55-86% of regional variation is explained by 184 chromatin tracks from more than 80 tissues

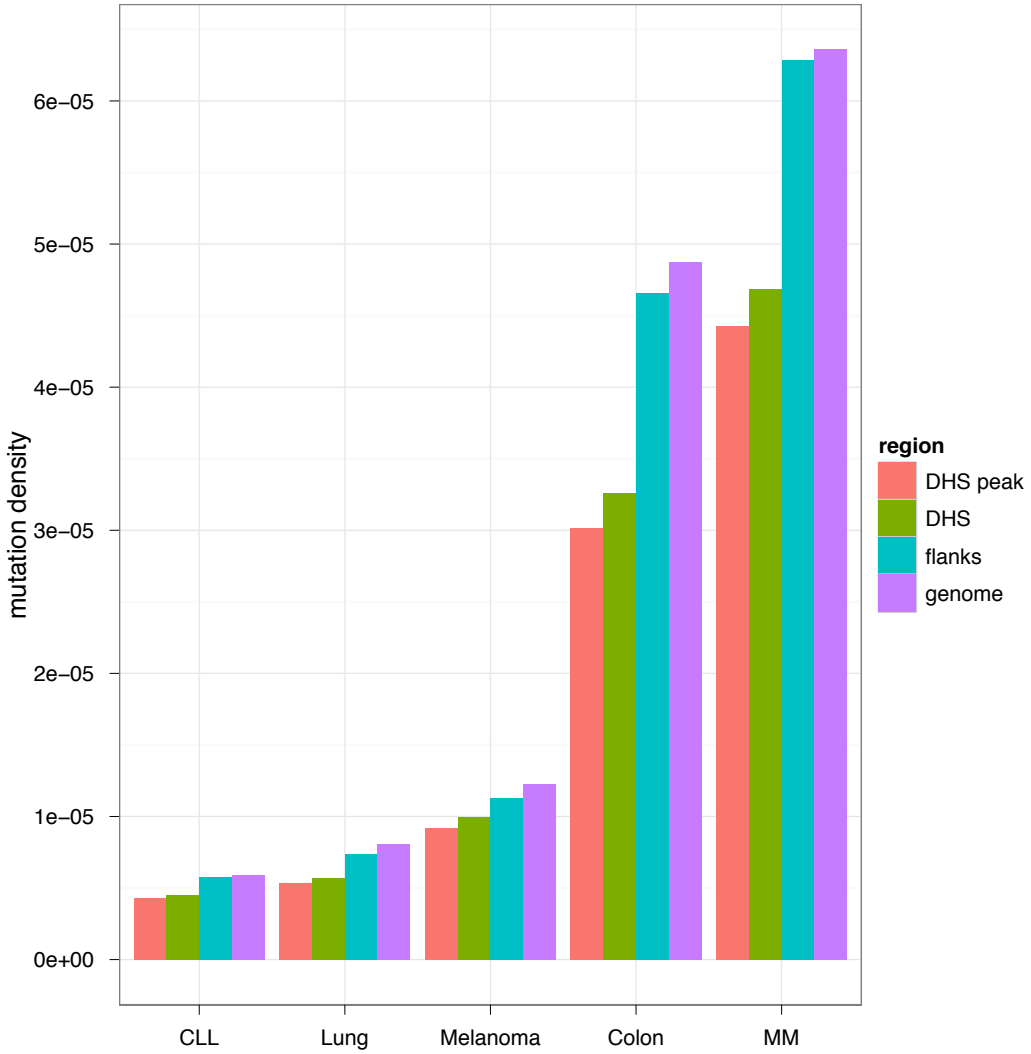


Cell type specificity

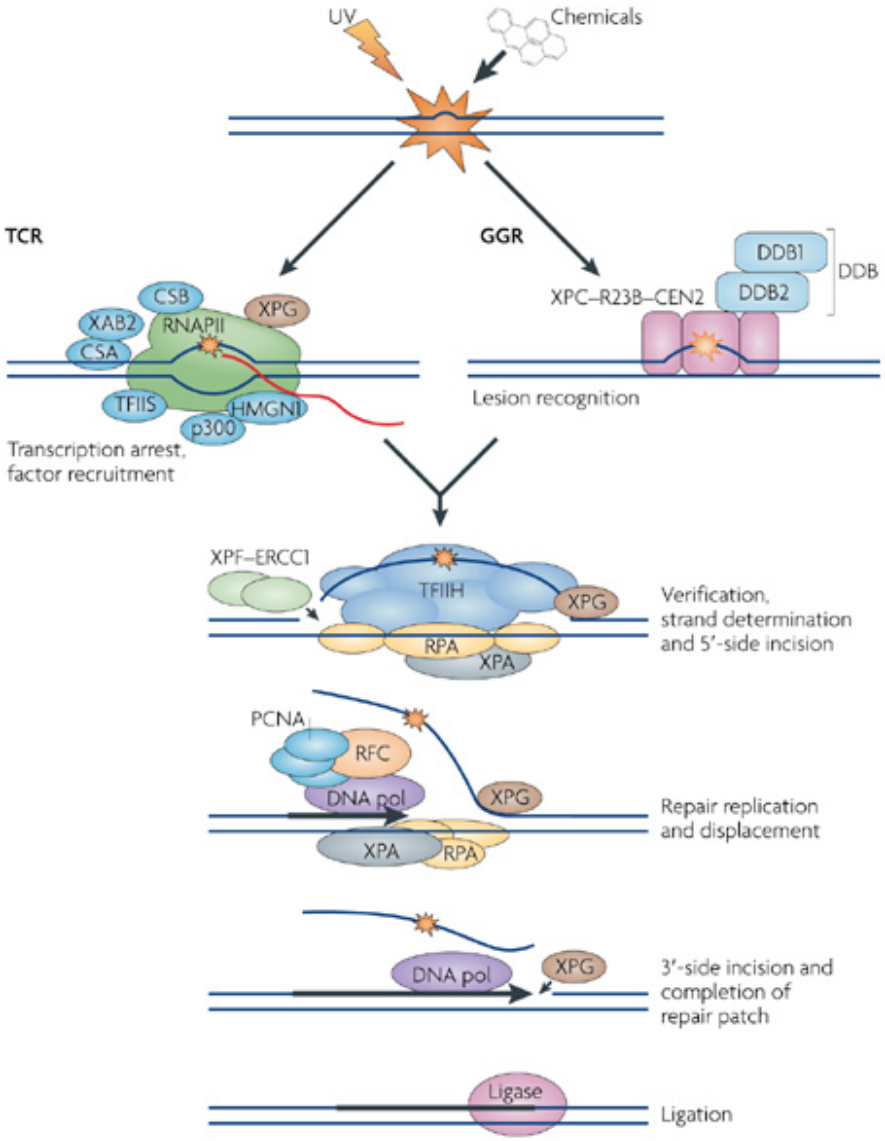




Mutation rate is reduced in regulatory regions marked by accessible chromatin

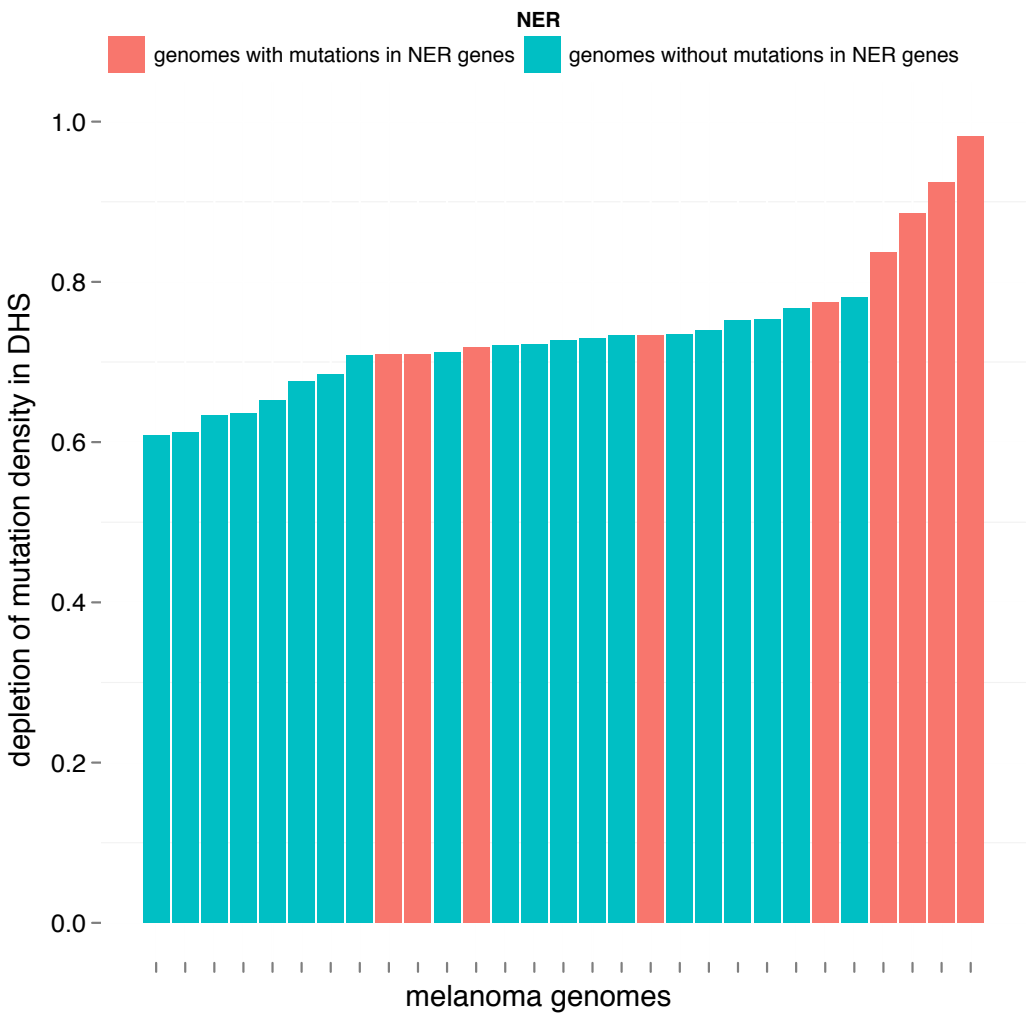


Polak *et al.*, *Nature Biotech. 2014*



Nature Reviews | Molecular Cell Biology

Implicating nucleotide excision repair (NER)



High mutation density in TFBS due to NER

LETTER

doi:10.1038/nature17661

Nucleotide excision repair is impaired by binding of transcription factors to DNA

Radhakrishnan Sabarinathan[1], Loris Mularoni[1], Jordi Deu-Pons[1], Abel Gonzalez-Perez[1] & Núria López-Bigas[1,2]

LETTER

doi:10.1038/nature17437

Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes

Dilmi Perera[1]*, Rebecca C. Poulos[1]*, Anushi Shah[1], Dominik Beck[1], John E. Pimanda[1,2] & Jason W. H. Wong[1]

Overall, mutation density is low in early replicating regions, active regulatory elements and highly expressed genes.

This is aggregate. How about the effects of individual mutagens?

How about individual samples?

Example: APOBEC mutagenesis

APOBECs are cytidine deaminases involved in cancer mutagenesis (primarily APOBEC3A)

APOBEC creates strand-coordinated mutation clusters. APOBEC acts on ssDNA.

APOBEC has a characteristic signature:
TCW→TTW or TCW→TGW

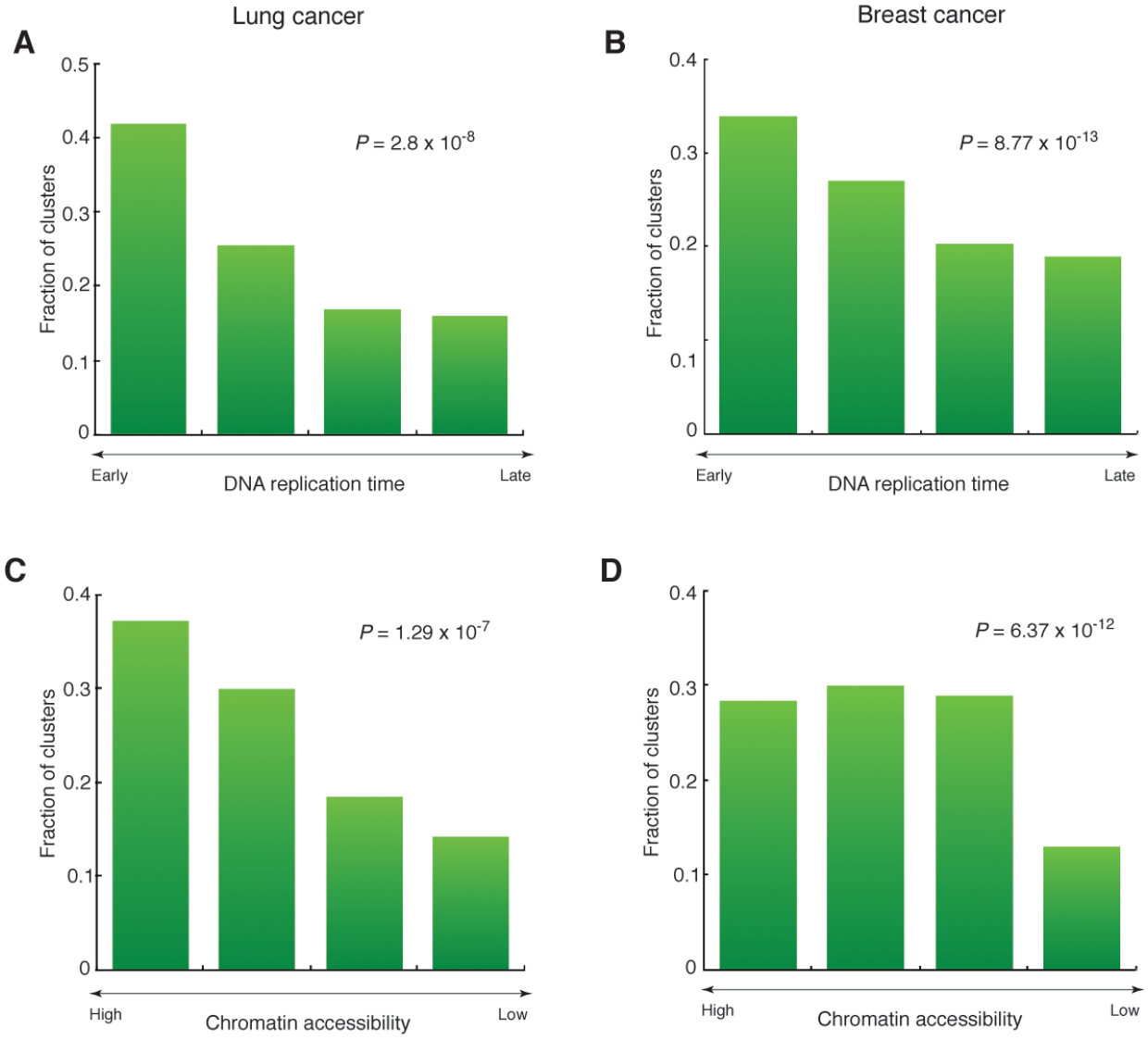
Focus on APOBEC

Enrichment of APOBEC signature and of cluster density varies by cancer types

It also varies by sample within cancer type

APOBEC activity results on sample-specific mutation properties

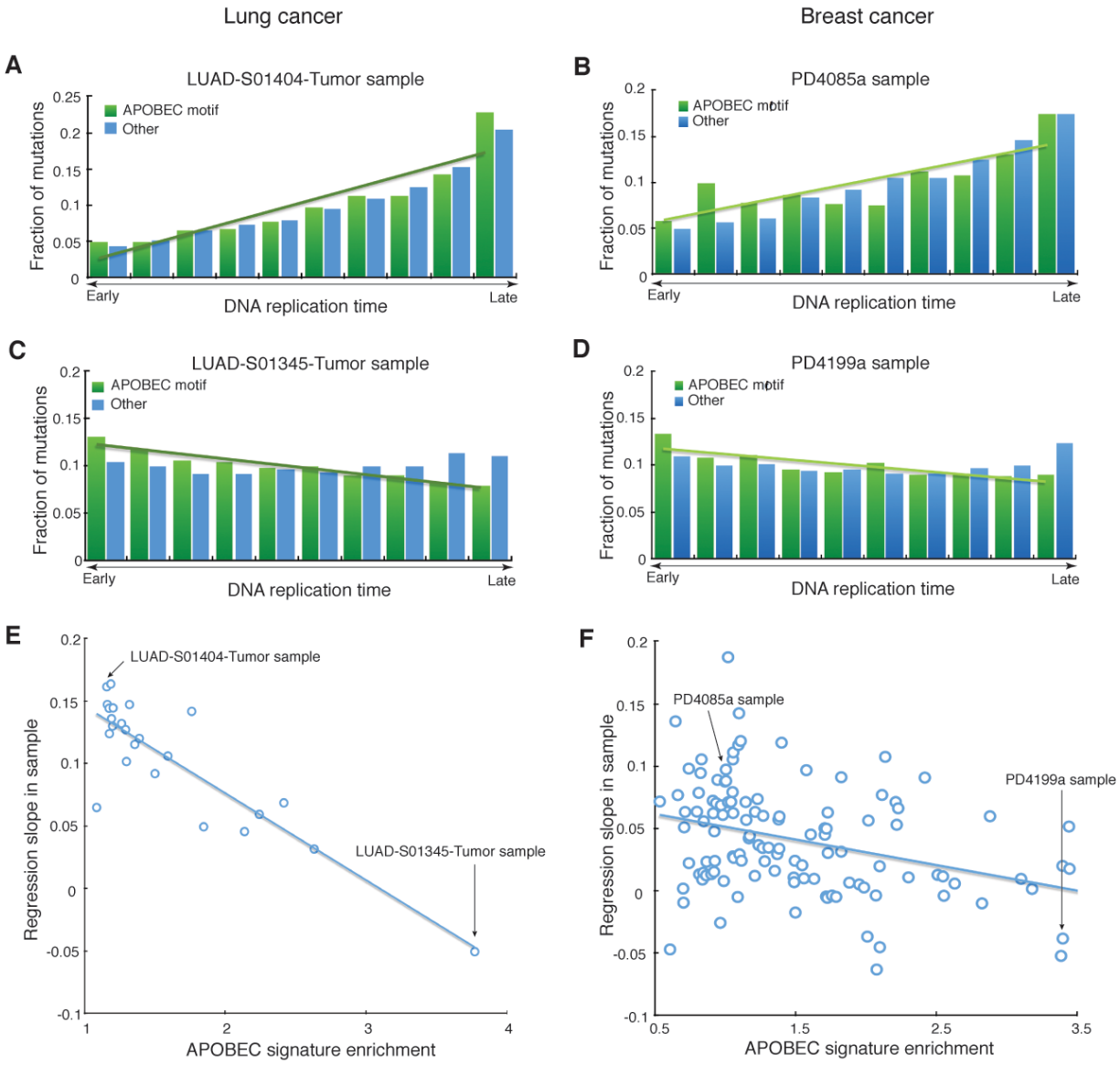
For mutations in clusters



119 breast and 24 lung cancer samples

Kazanov *et al.*, *Cell Reports* 2015

Dependency on the enrichment of the APOBEC signature

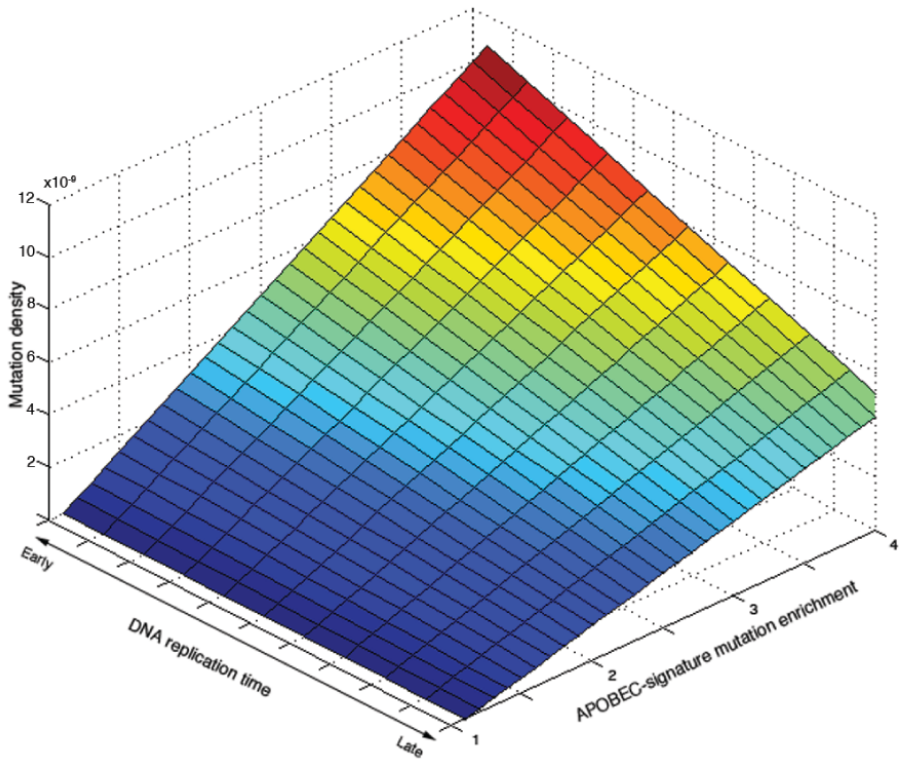


Modelling a mixture of mutation origins

We can model mutations within APOBEC signature as a mixture of APOBEC induced mutations and other mutations

$$M(x,s) \sim \begin{cases} \alpha(s)(\beta_{A0} + \beta_{A1} f(x)) + (1-\alpha(s))(\beta_{N0} + \beta_{N1} f(x)), & if\ x \in T\underline{C}W \\ \beta_{N0} + \beta_{N1} f(x), & if\ x \notin T\underline{C}W \end{cases}$$

Joint analysis of all samples



Conclusions

The effect of epigenomic features on cancer mutations may be mutagen-dependent

APOBEC mutations are unique in the genomic distribution

Mutation models have to be sample-specific

Search for non-coding drivers

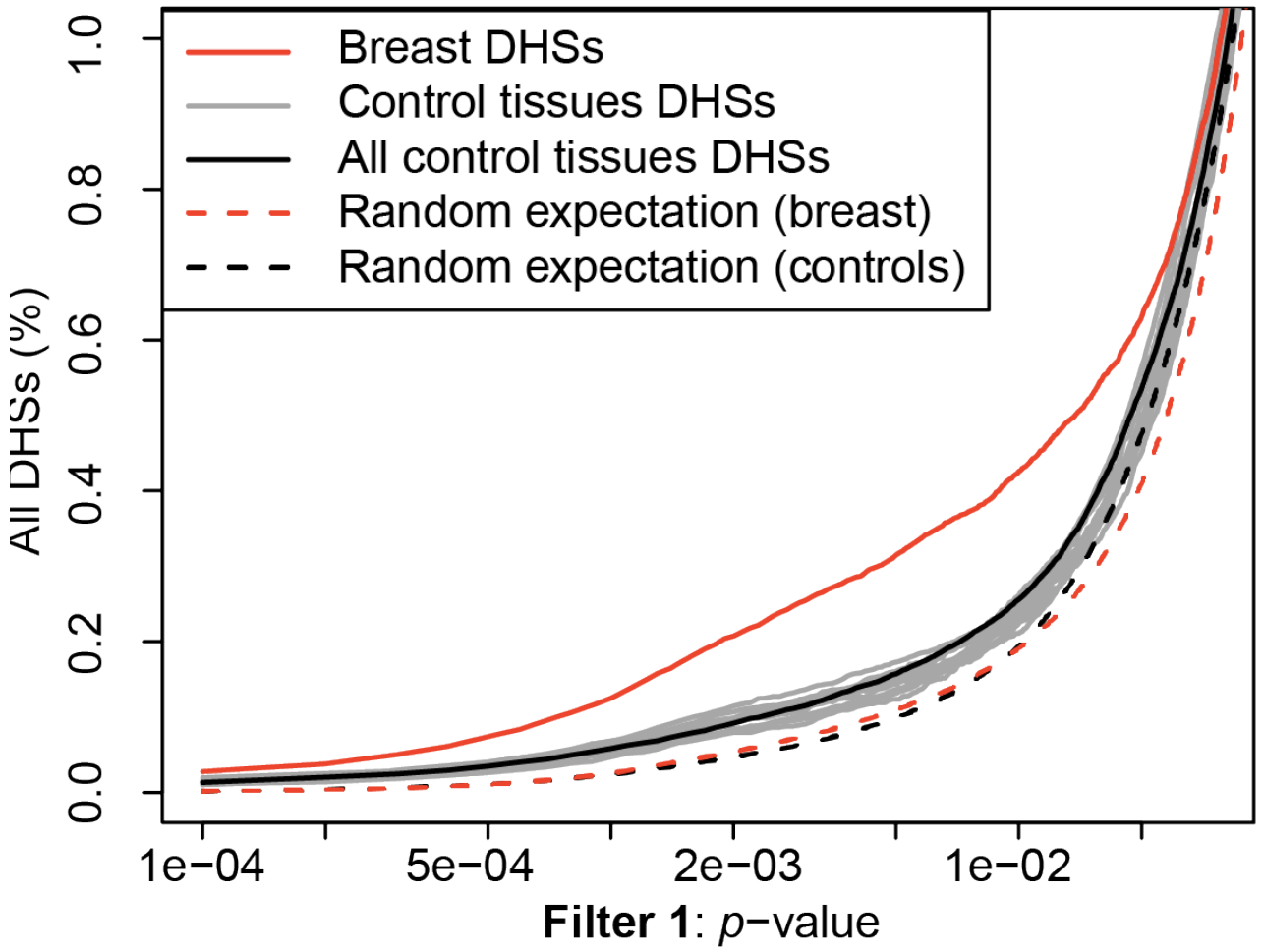
Cluster regulatory elements by all covariates

Assume Poisson statistics within clusters

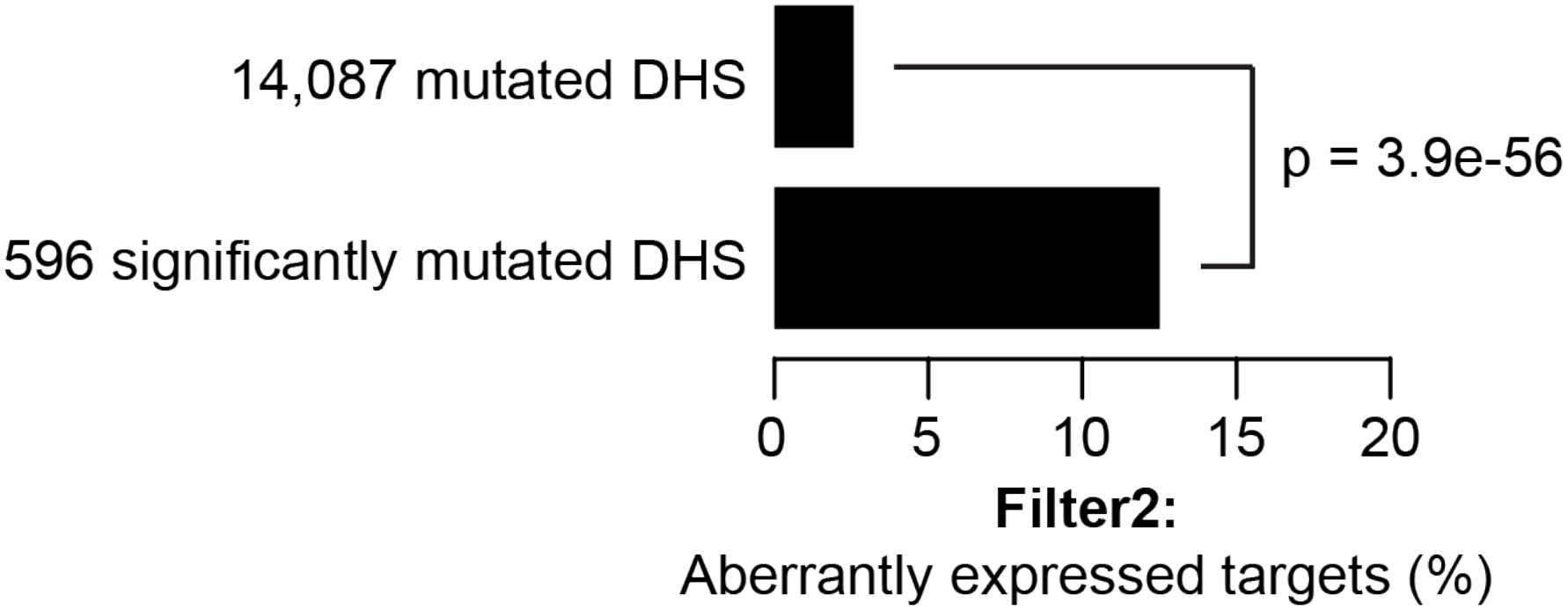
Use “wrong” tissue types as a control to derive FDR

D’Antonio, Weghorn *et al.,* (in review)

Search for non-coding drivers



Search for non-coding drivers



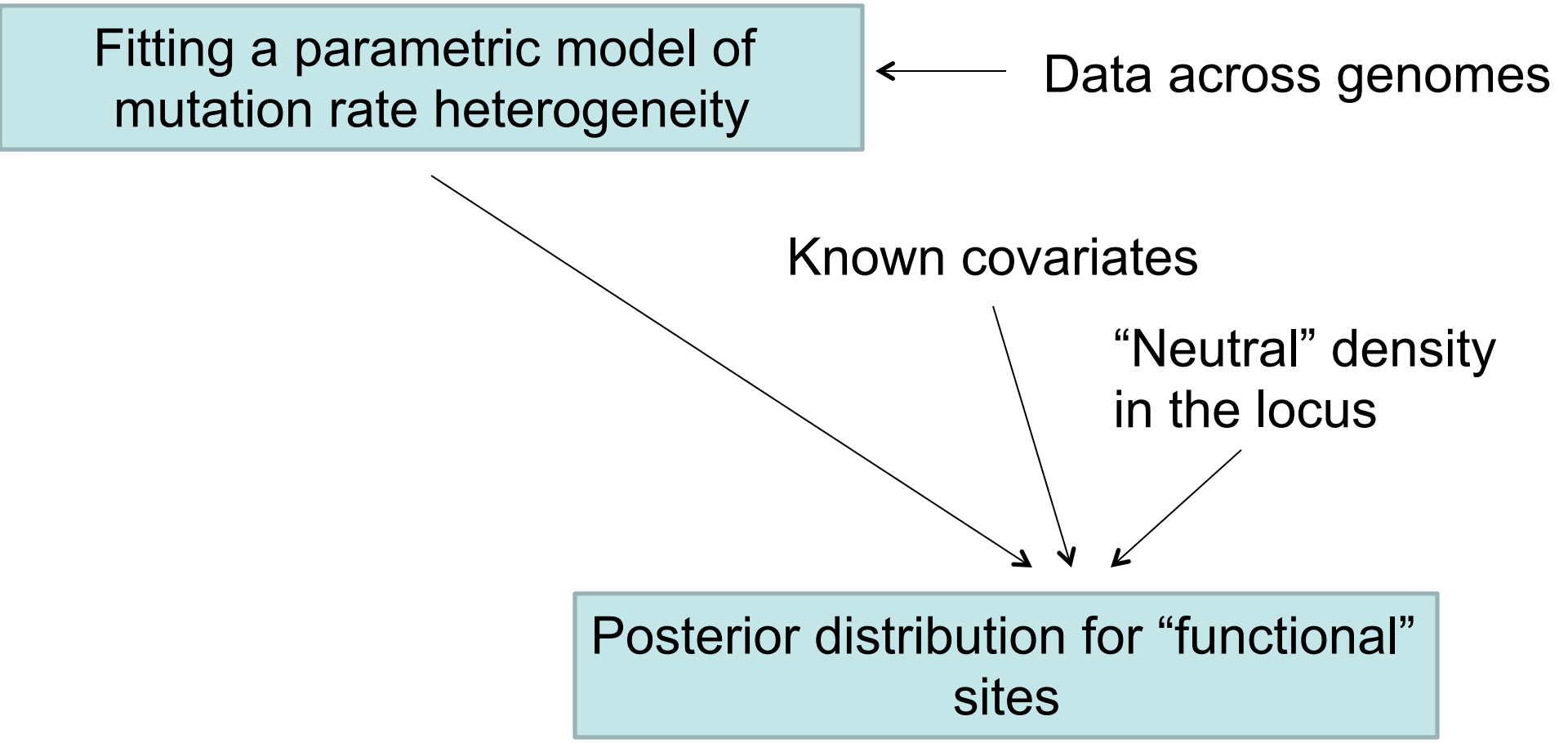
In search for a better statistical approach

Precisely estimating local mutation rate is very difficult

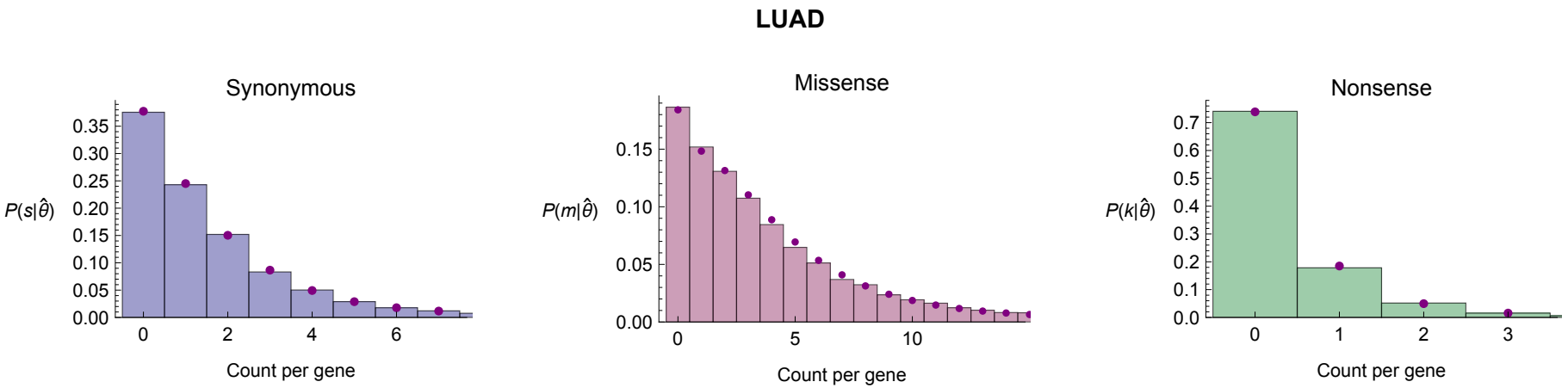
It is practical to model a set of samples rather than a cancer type or an individual sample

We opt for a hierarchical model

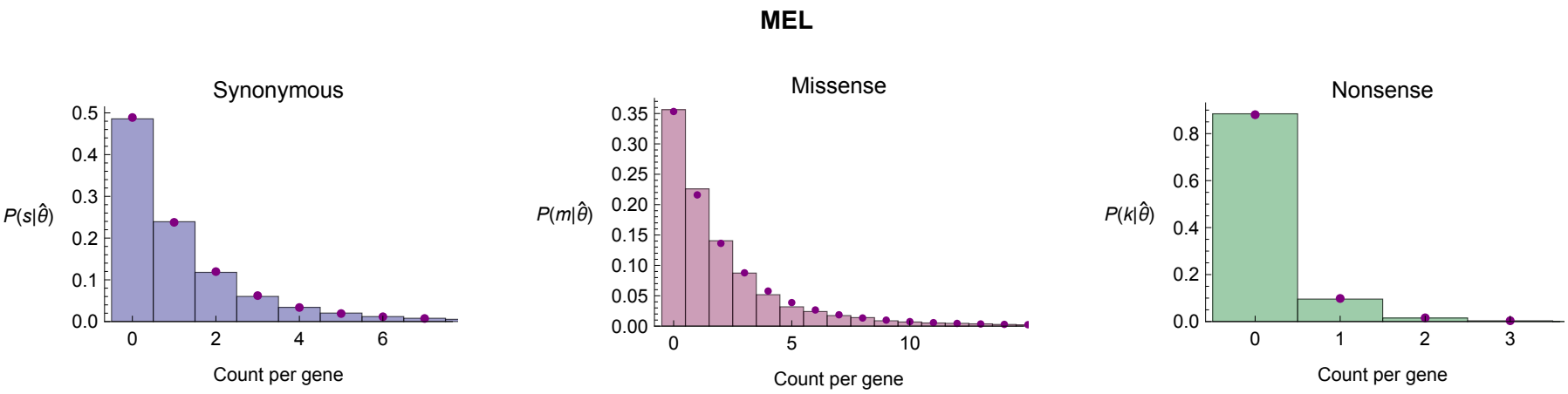
In search for a better statistical approach



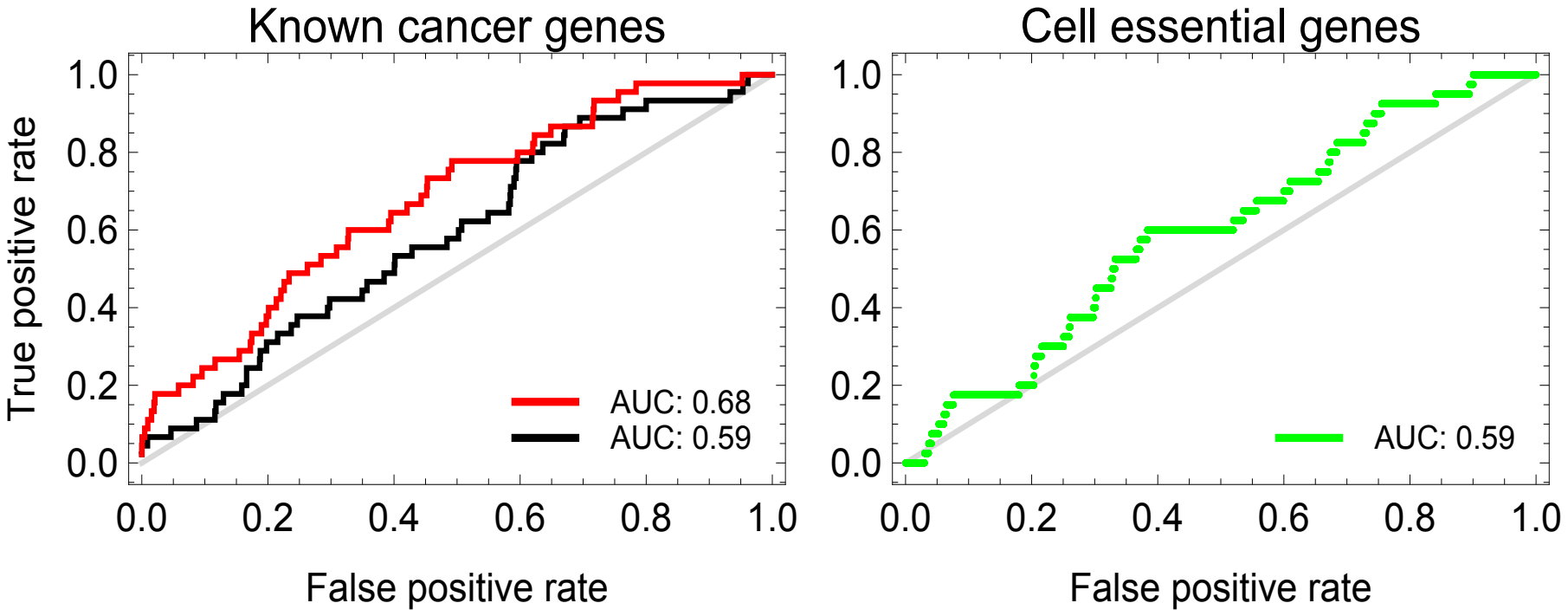
Modelling heterogeneity of mutation rates



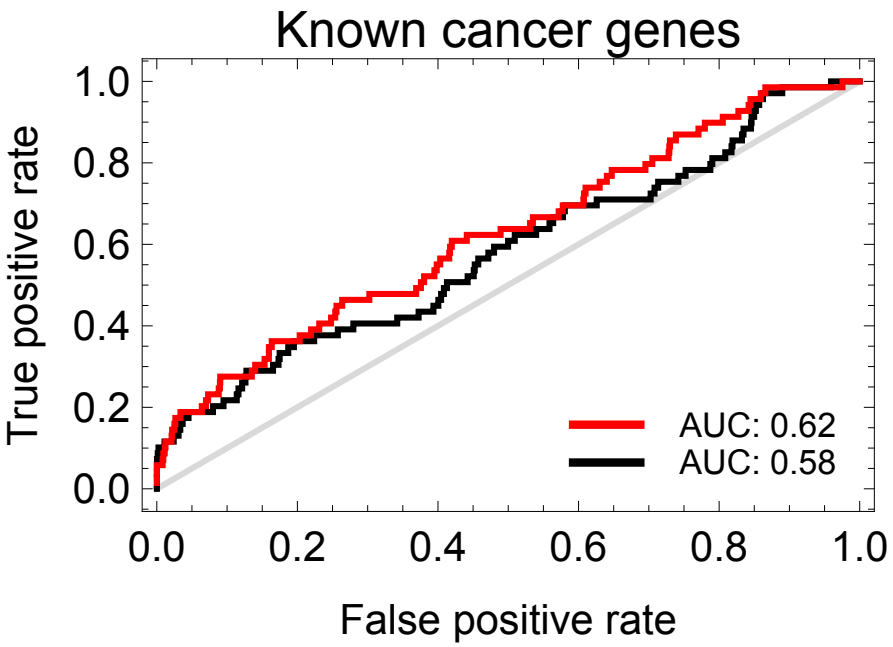
Modelling heterogeneity of mutation rates

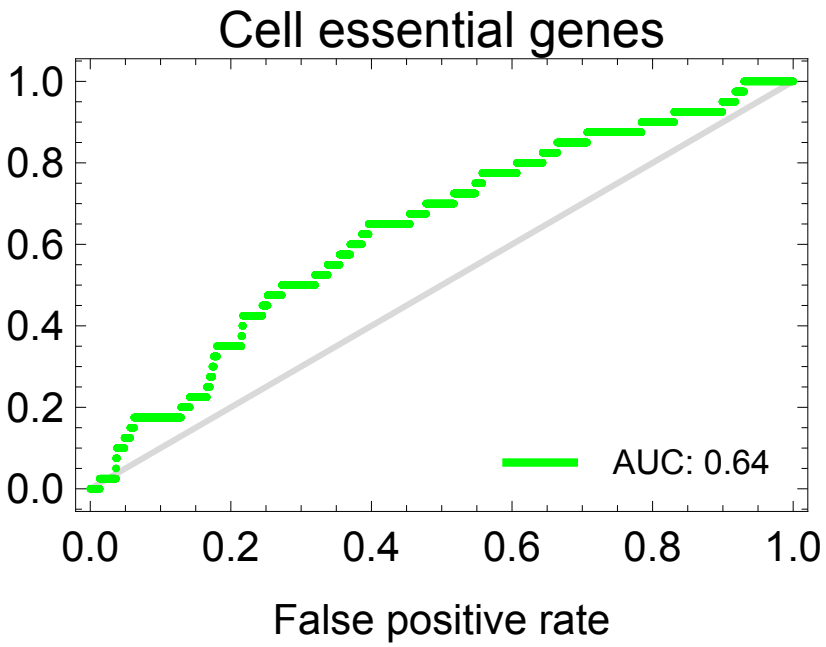


Finding genes under selection (LUSC)

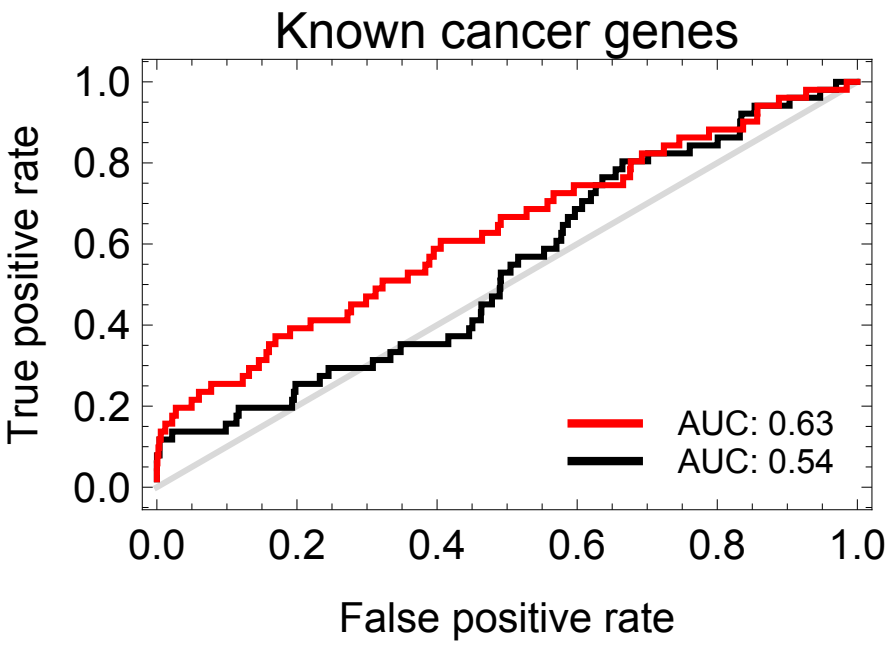


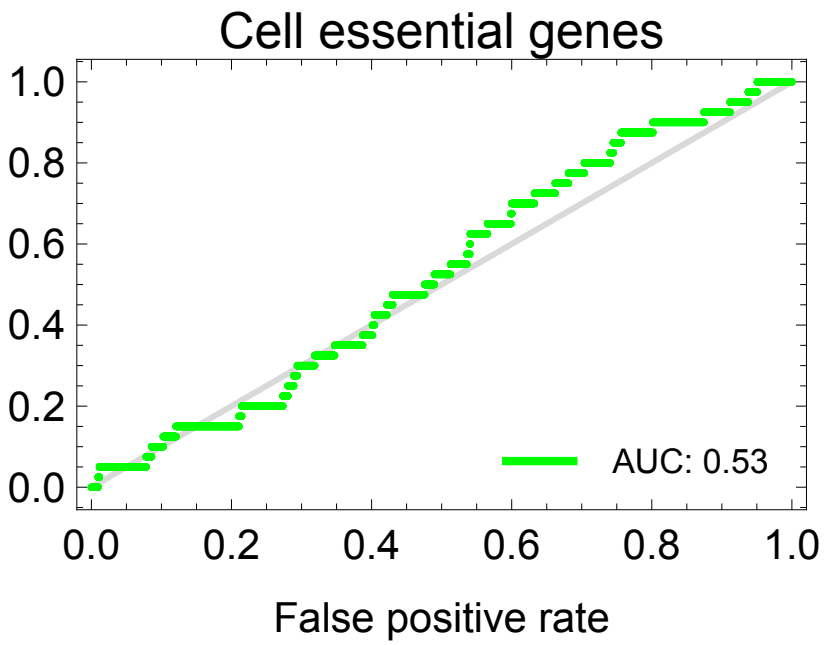
Finding genes under selection (LUAD)





Finding genes under selection (MEL)





Conclusions

Understanding mutation rate heterogeneity helps understand basic biology and develop statistical methods of cancer genomics

Mutation rate varies by cancer type and by sample

Epigenomic features are key

We need better statistical approaches

Acknowledgments

Acknowledgments

The lab: Daniel Jordan, Ivan Adzhubei, Paz Polak, Donate Weghorn, Mashaal Sohail, Dana Vuzman, Daniel Balick, Sung Chun, Jae-Hoon Sul, Chris Cassa, Sebastian Akle, David Radke

Collaborators:
John Stamatoyannopoulos, Bob Thurman,
Rosa Karlic, Amnon Koren,
Dmitri Gordenin, Marat Kazanov, Steven Roberts