NIEHS
National Institute of Environmental Health Sciences

Pan-Cancer Analysis of APOBEC Mutagenesis

Dmitry Gordenin

National Institute of Environmental Health Sciences (NIH, DHHS)

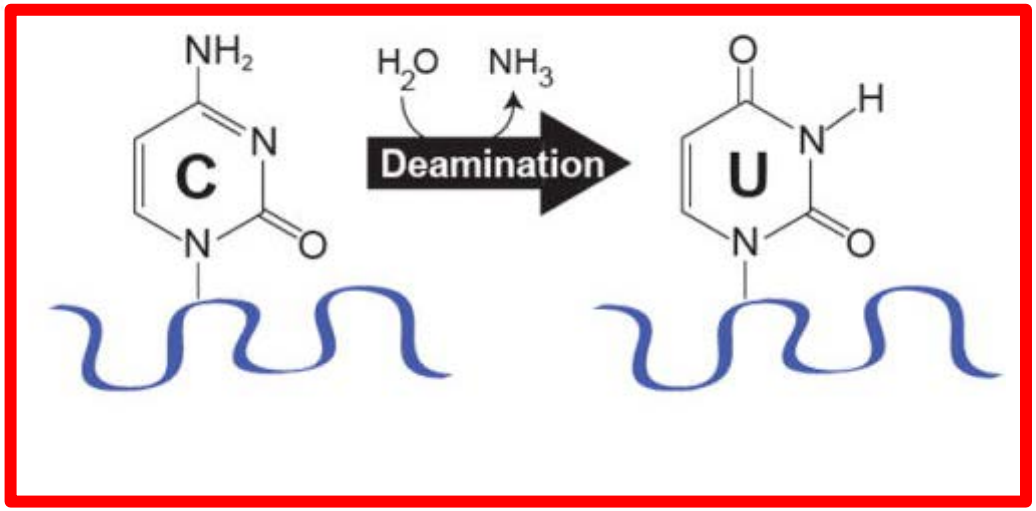
Laboratory of Molecular Genetics

See also poster 77, **Steve Roberts et al.**

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

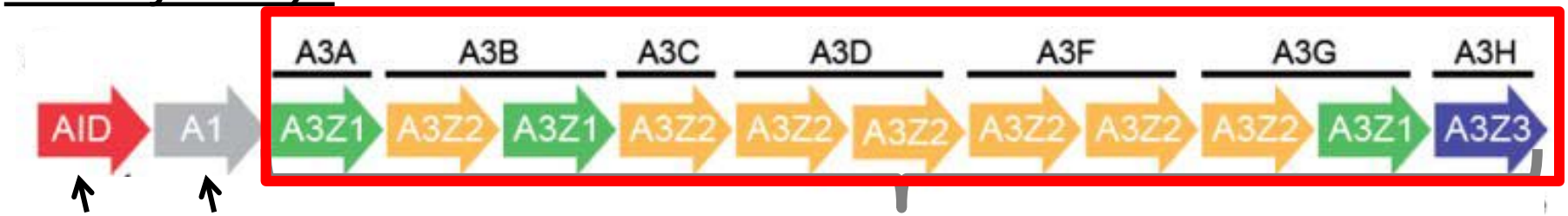
ssDNA-specific AID/APOBEC cytidine deaminases - an endogenous strong mutagen in human cancers

Biochemistry:



- *in vivo APOBECs* can cause C→T and C→G substitutions
- **APOBECs 1/3A/3B/3C/3DE/3F/3H** deaminate C in TCW motif

Gene family:



adaptive immunity

RNA editing

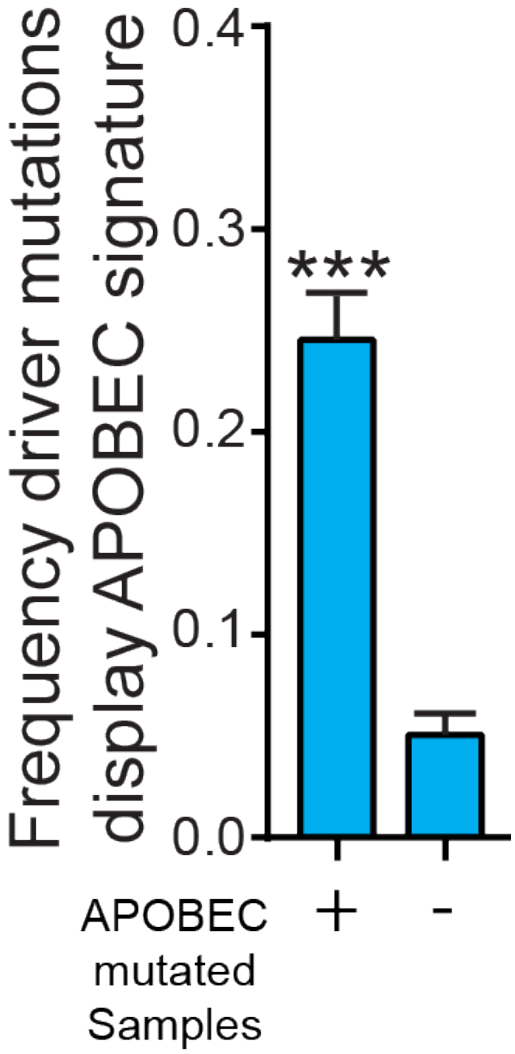
-- innate immunity: restriction of retroviruses/retrotransposons

-- hypermutation by accidental access to chromosomal DNA

APOBEC3B and/or APOBEC3A are the current prime suspects for mutagenesis in cancer

Graphics adapted from Refsland and Harris, 2013

APOBEC: a new type of carcinogenic mutagen



Roberts et al., Nature Genetics, 2013

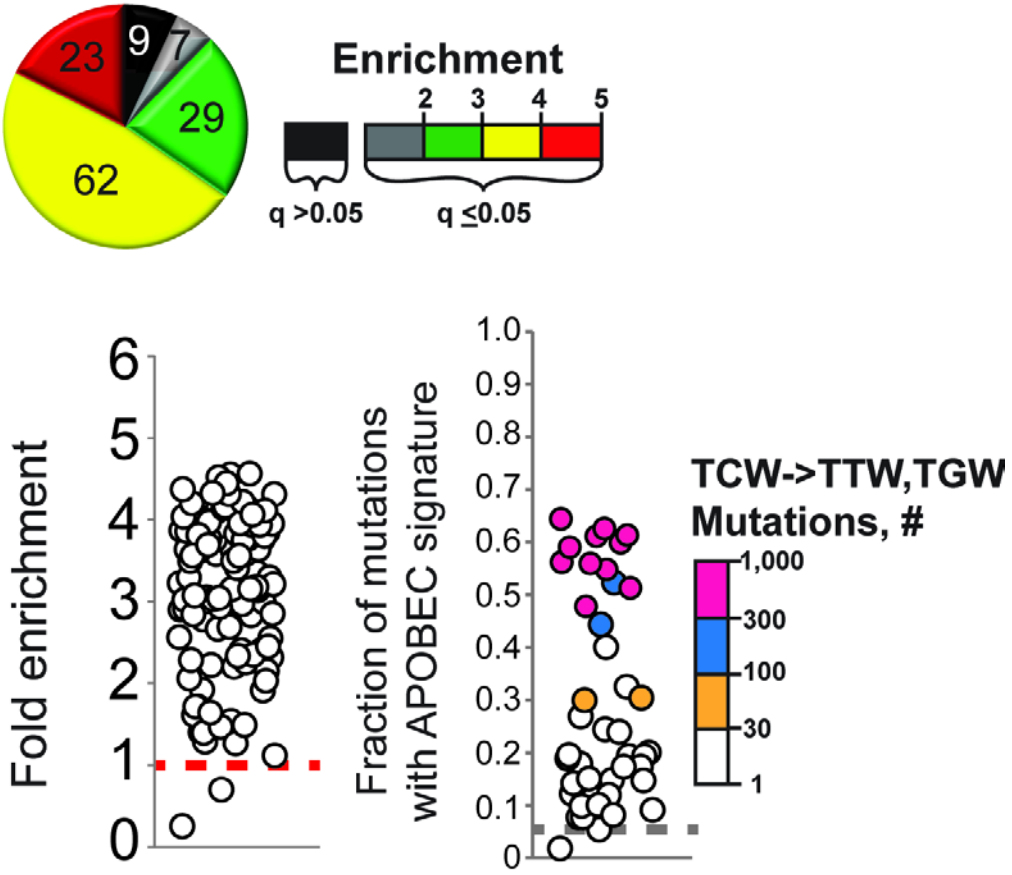
***Carcinogenic mutagens*:**

Exogenous:
UV, tobacco, aristolochic acid, occupational chemicals

Endogenous: **APOBEC**

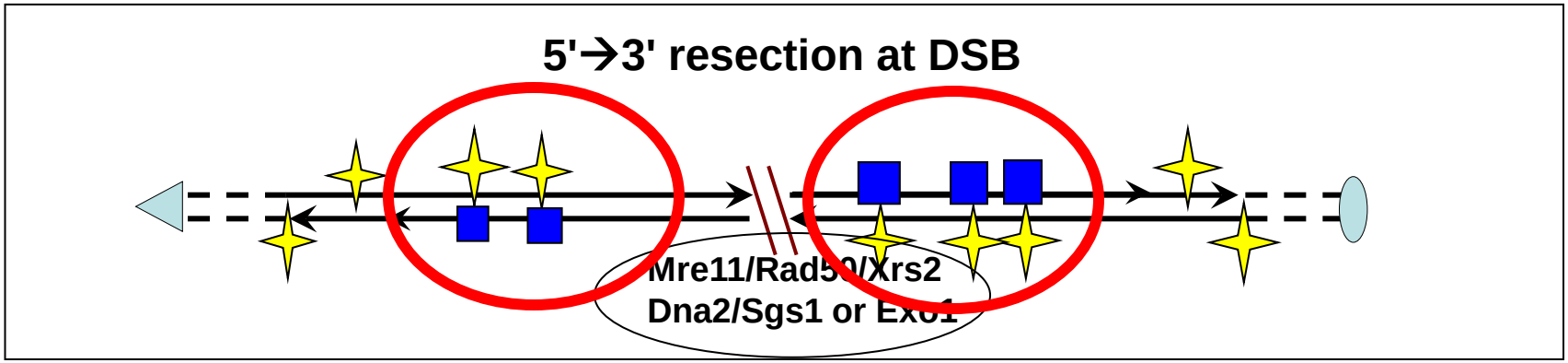
Roberts *et al*., 2012, 2013;
Nik-Zainal *et al*. 2012; Alexandrov *et al*. 2013;
Burns et al., 2013a,b

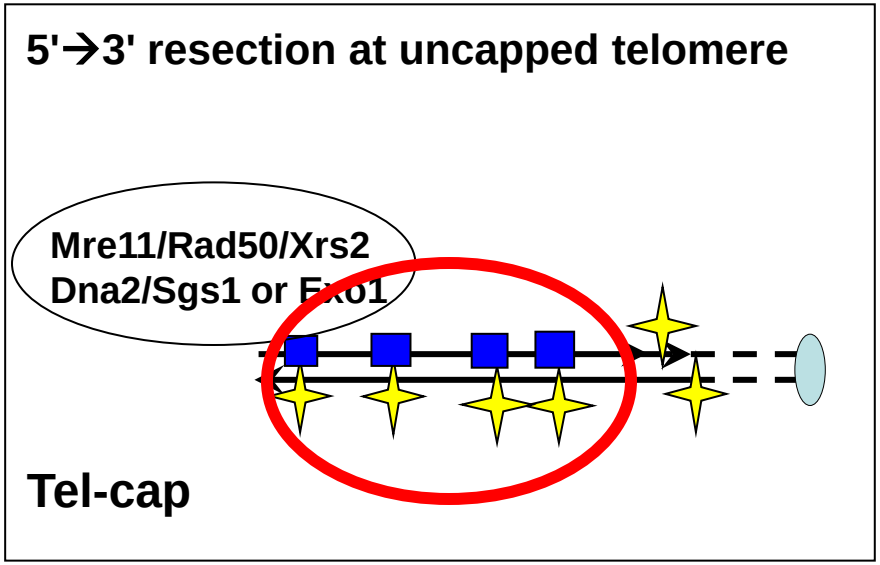
Example of very strong APOBEC mutagenesis –
- muscle invasive urothelial carcinomas of bladder (BLCA).

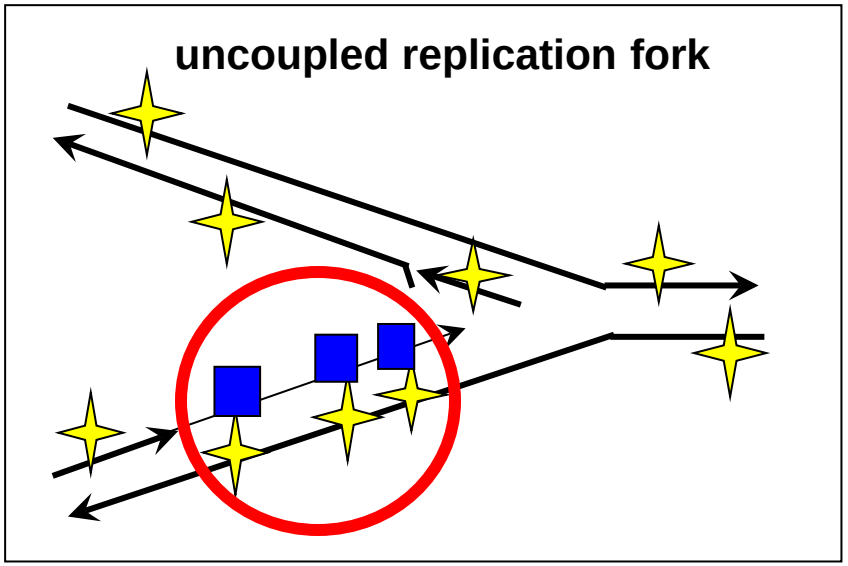


Gordenin et al., for TCGA BLCA Marker Paper, Nature, 2014

Findings in model yeast systems: Error-prone Translesion Synthesis in Damaged Long ssDNA can be a Source of ***Localized Hypermutation and Strand-Coordinated Mutation Clusters***





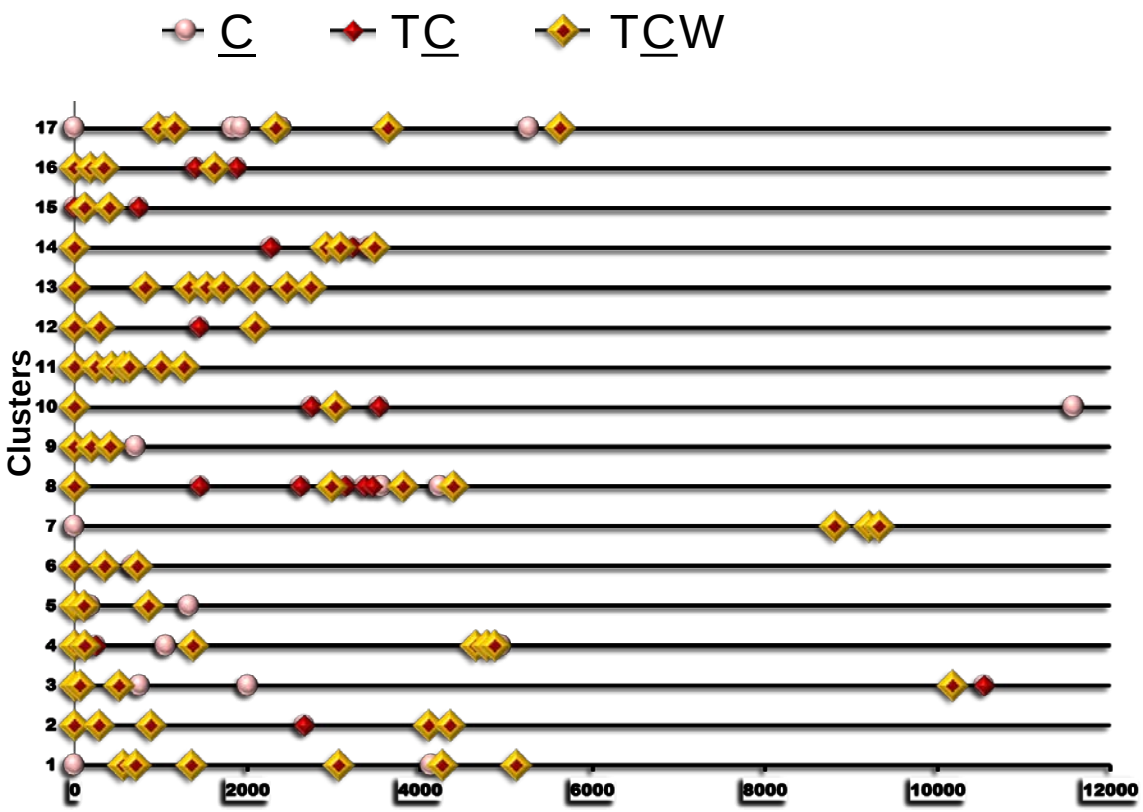


PLoS Genetics, 2008; 2012; *DNA Repair* 2010; 2013; *NAR* 2013; *Mol. Cell*, 2012

Mutation Motifs in C-coordinated clusters.

Roberts et al. *Mol.Cell,* 2012

(example - multiple myelomas)



C-coordinated mutations

84% are **TC/GA**

⬇

81% are **TCW/WGA**

87% are **C→T** or **C→G**

APOBEC Family Enzymes:

- Edit RNA & restrict retroelements
- **APOBEC 1/3A/3B/3C/3DE/3F/3H** – **TCW** – mutation motif
- strongly prefer **ssDNA** vs dsDNA
- activity in ssDNA gaps results in **C to G** and **C to T** substitutions

Similar events (termed "*kataegis*") also reported in breast cancer.
Nik-Zainal…and Stratton, *Cell*, 2012

Hypothesis-based statistics of APOBEC mutagenesis

Hypothesis driven Analysis

Enrichment (E): a Metric for APOBEC mutation pattern:

Roberts et al., *Mol. Cell* 2012

$$E = \frac{MUTATIONS_{TCW \rightarrow TTW \text{ or } TGW} \times CONTEXT_{C}}{MUTATIONS_{C \rightarrow T \text{ or } G} \times CONTEXT_{TCW}}$$

cagcagtccttgctgaccctCcaaggaacacttcagccaga
tacttgtggtagacgtcgatGatagagttcaaggctttctc
gttccaacatcatcttctctCagttcctctggtcggctgac
gacttgagcccacttgctttGagagaactgaagatattgga
agattatcttgtgttagctcAaccacccttactacttcctg

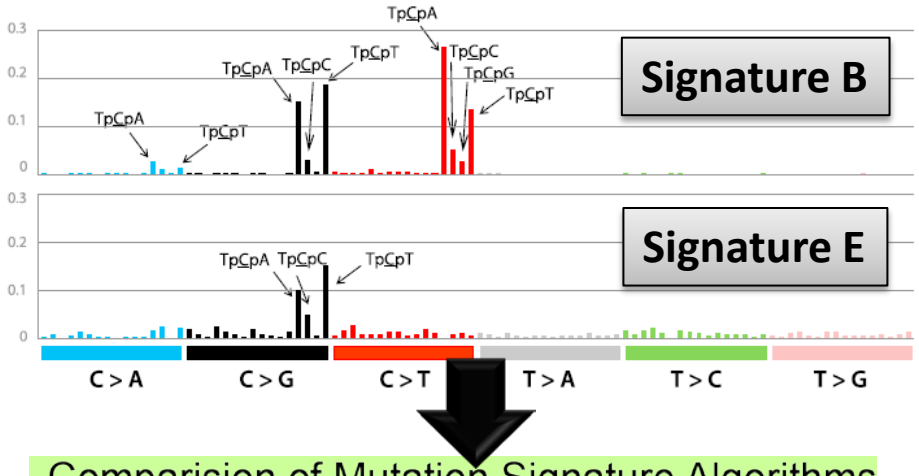
mutations

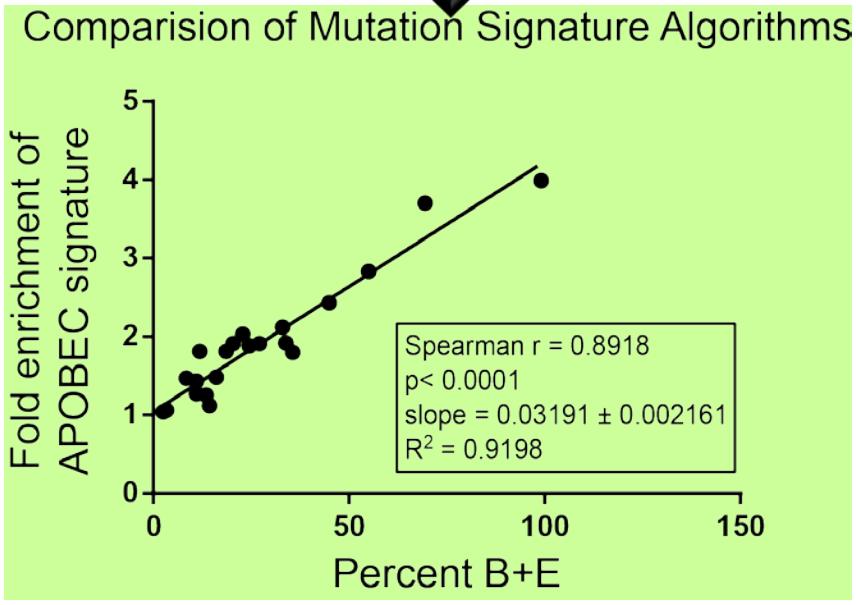
context

- Reflects short scanning mode of APOBEC enzymes
- Concentrates on "sequence-able" part of the genome
- **Produces sample-specific P-values**
  (works even for exome MAFs)

De Novo Pattern Recognition

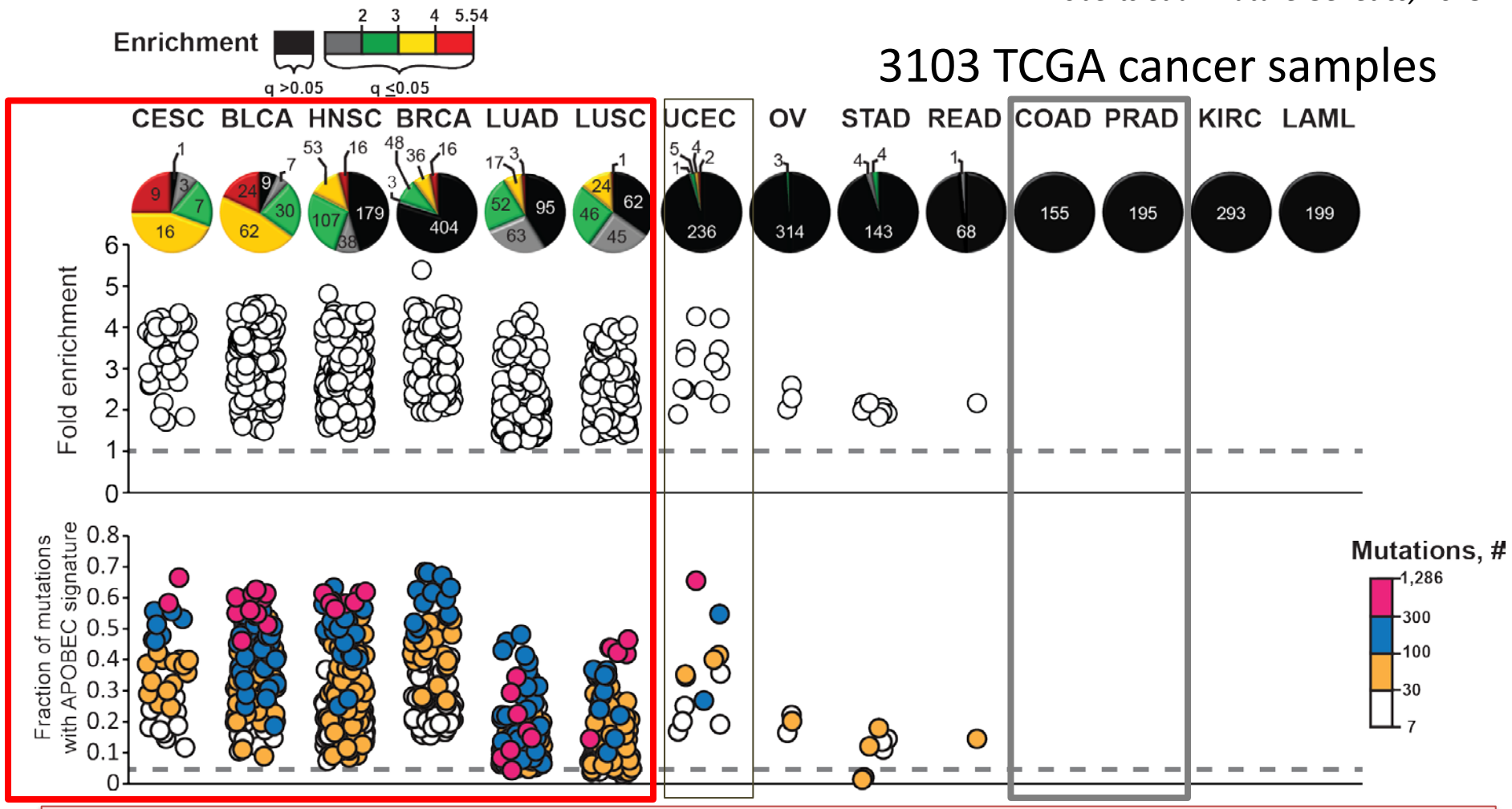
21 BRCA genomes: Nik-Zainal … Campbell, Stratton, *Cell* 2012





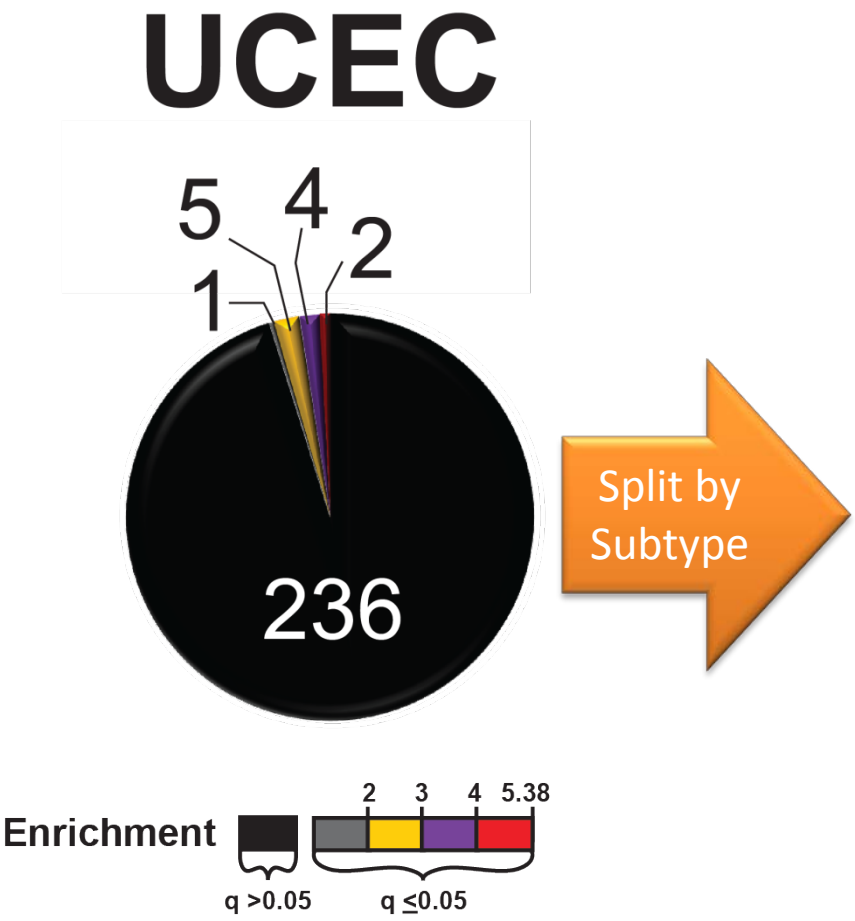
APOBEC mutation pattern is abundant in *cervical, bladder, head and neck, breast,* and *lung* cancer types

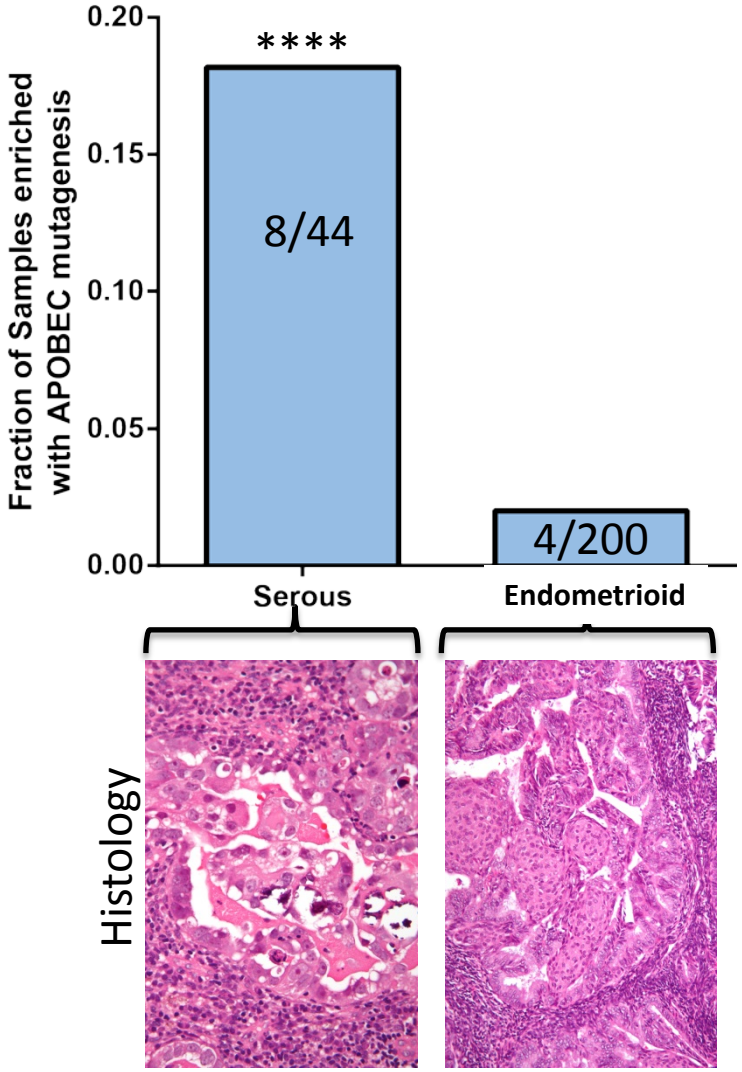
Roberts et al. Nature Genetics, 2013



APOBEC mutagenesis likely occurs in the background of all cancer types, but is more abundant in specific types.

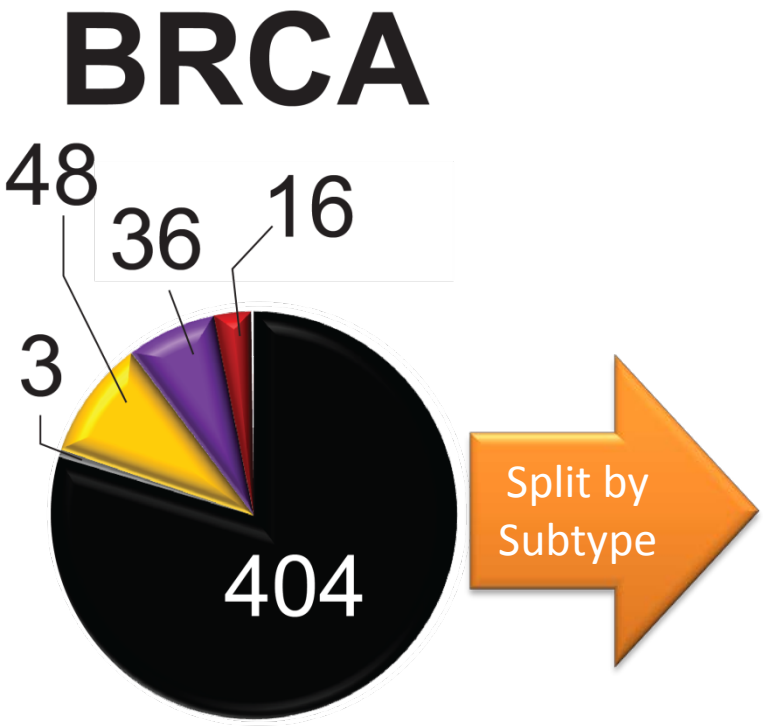
Highlighting rare APOBEC-mutated samples: example – Uterine Corpus Endometrioid Carcinoma (UCEC)



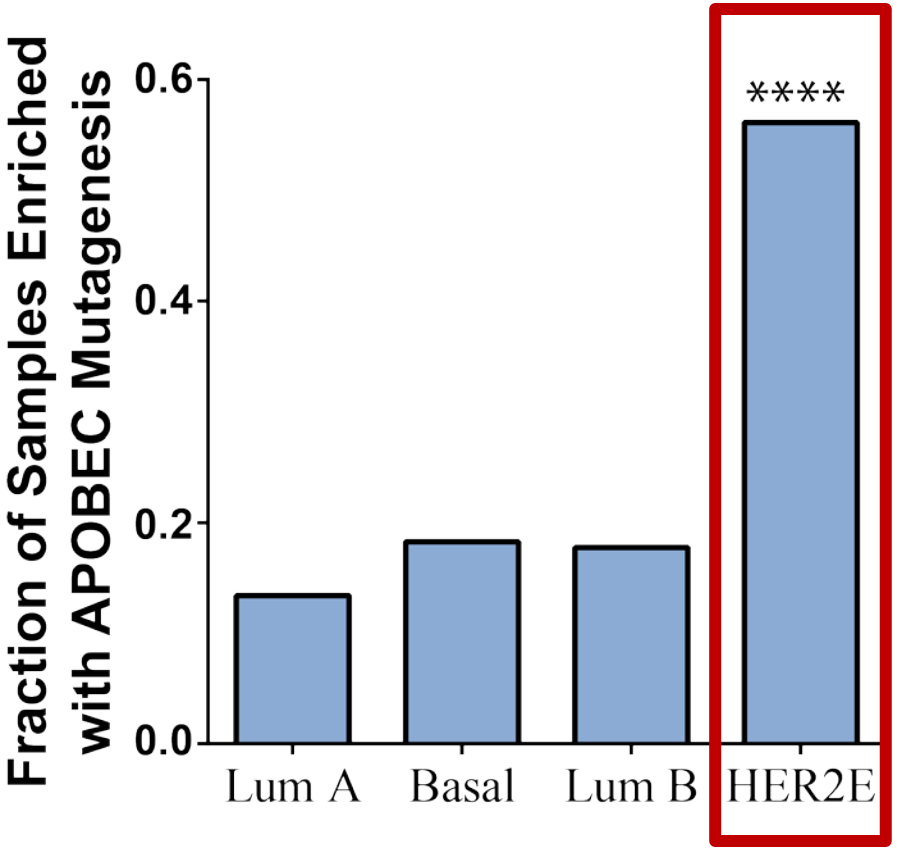


Highlighting APOBEC-hypermutated cancer subtypes: example - HER2-enriched subtype of breast cancer

Roberts *et al.*, 2013

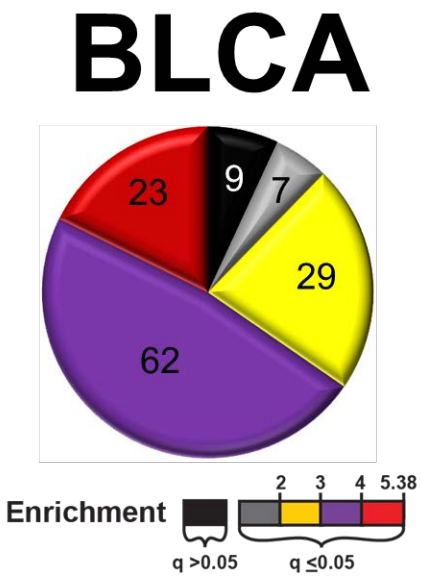


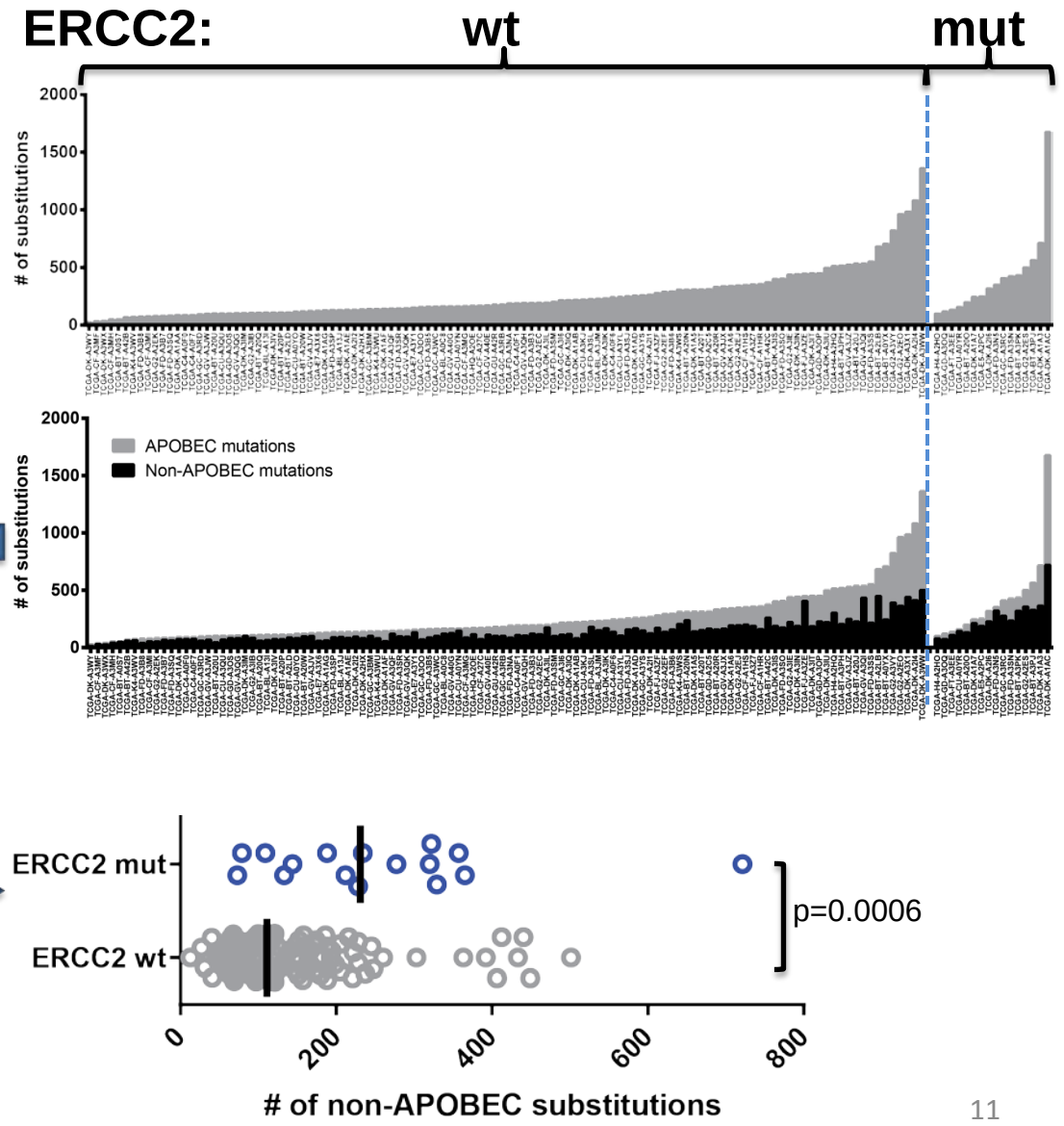
Split by Subtype



Deconvoluting mutagenic mechanisms within a cancer type: example - increased mutagenesis in ERCC2-mutated bladder cancers (BLCA)

- **BLCA SMGs include**: *TP53, MLL2, ARID1A, KDM6A, PIC3CA, EP300, CDKN1A, RB1, **ERCC2**…*
- ***ERCC2*** = Excision Repair
  - mutation = little effect on the total # of substitutions
- **Subtract APOBEC mutations to reveal mutagenesis in ERCC2 mutated samples**



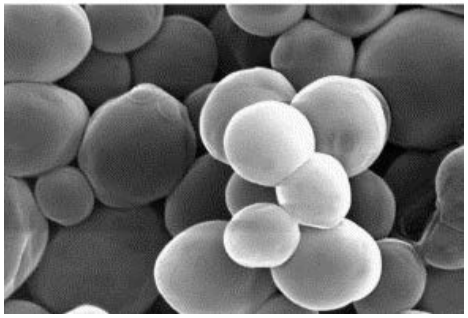


TCGA-related efforts

- Input into cancer specifics AWGs
- Integrate analysis of APOBEC mutagenesis in cancer exome MAFs into Firehose
- Analyse updated and new TCGA exome MAFs

Combining mechanistic and bioinformatics approaches to understand mutation processes operating in cancer

Yeast



Mechanistic knowledge used to build **stringent statistical hypothesis**

Tumors

Apply to bioinformatics exploration of large databases of clinical mutations (*e.g.* TCGA) **to understand mutagenic mechanisms** relevant for disease

Acknowledgements

y - Yeast model study

h – Analysis of mutations in human cancer genomes

- *NIEHS, CSG*:
  Steve Roberts-y,h
  Kin Chan - y
  Joan Sterling – y
  Cole Thompson – y
  Yong Yang – y
  Francesca Storici - y
  Michael Resnick – y

- *NIEHS, SRA*:
  Shawn Harris - y,h
  Deepak Mav – y,h
  Arpit Tandon – y,h
  Ruchir Shah – y,h

- ***NIEHS, Bioinformatics:***
  Les Klimczak – y,h
  Sara Grimm - h
  David Fargo - h

- *Broad/Harvard*:
  Jaegil Kim - h
  David Kwiatkowski - h
  Gregory Kryukov – h
  Adam Kiezun – h
  Mike Lawrence – h
  Petar Stojanov – h
  Scott Carter – h
  Gad Getz – h

- *UNC, Chapel Hill*:
  Ewa Malc - y
  Piotr Mieczkowski - y