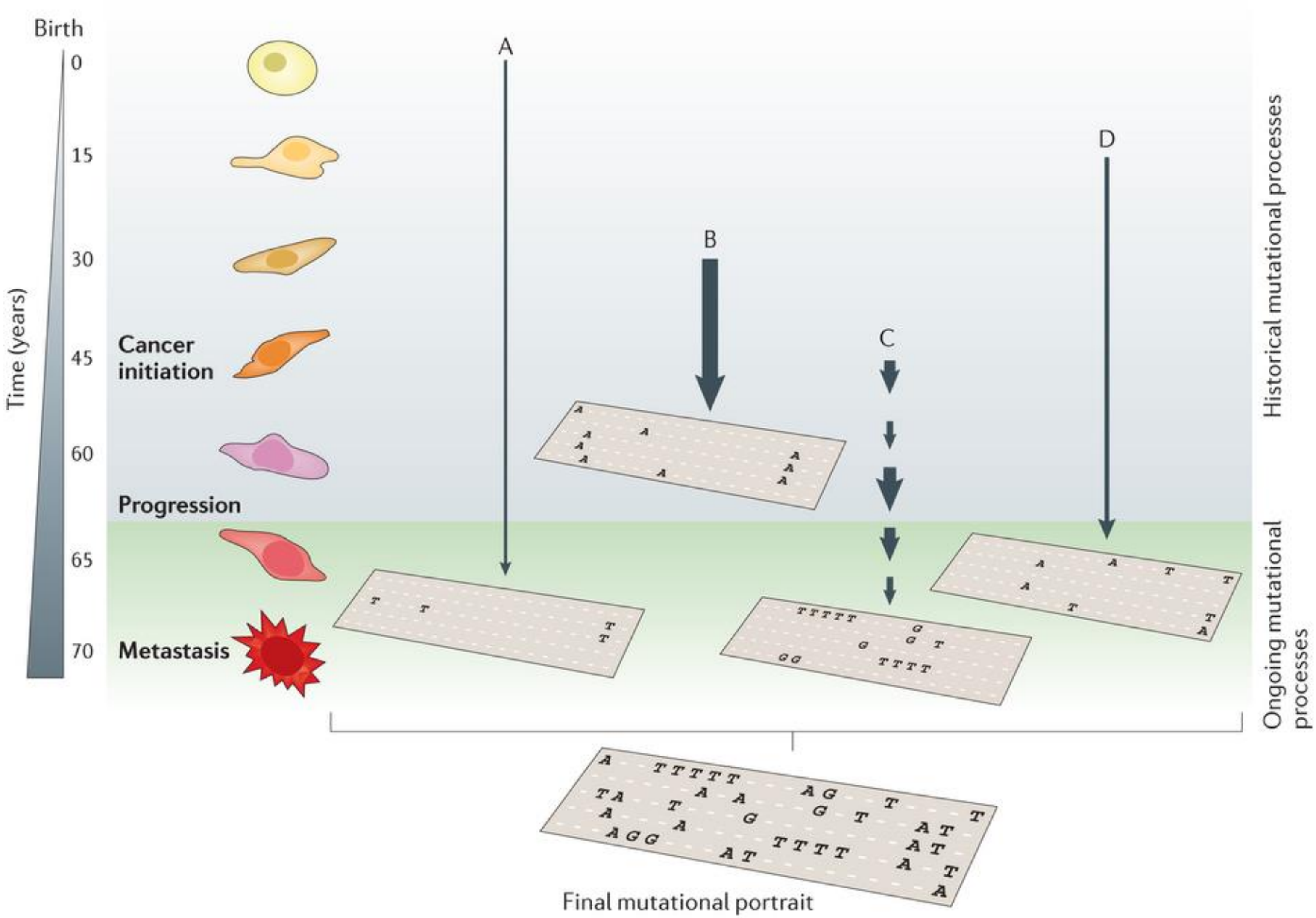
Multisample anlaysis

MUTATIONAL SIGNATURES – MODELLING MUTATION PROCESSES

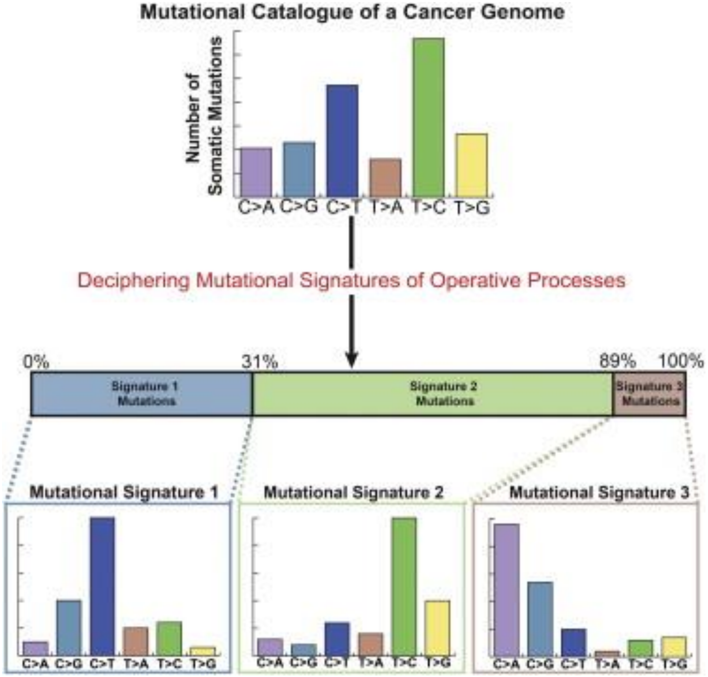
Signatures of mutational processes operative in cancer



Helleday et al., 2014

Nature Reviews | Genetics

Modelling signatures of mutational processes operative in cancer genomes using non-negative matrix factorization (NMF)



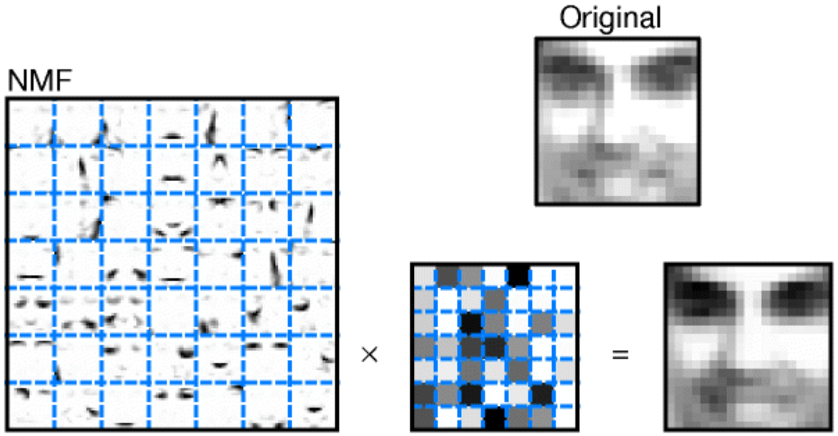
$$m_g^i \approx \sum_{n=1}^{N} p_n^i e_g^n.$$

$$\begin{bmatrix} m_1^1 & m_2^1 & \cdots & m_{G-1}^1 & m_G^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_1^K & m_2^K & \cdots & m_{G-1}^K & m_G^K \end{bmatrix} \approx \begin{bmatrix} p_1^1 & p_2^1 & \cdots & p_{N-1}^1 & p_N^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_1^K & p_2^K & \cdots & p_{N-1}^K & p_N^K \end{bmatrix}$$

$$\times \begin{bmatrix} e_1^1 & e_2^1 & \cdots & e_{G-1}^1 & e_G^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ e_1^N & e_2^N & \cdots & e_{G-1}^N & e_G^N \end{bmatrix}$$

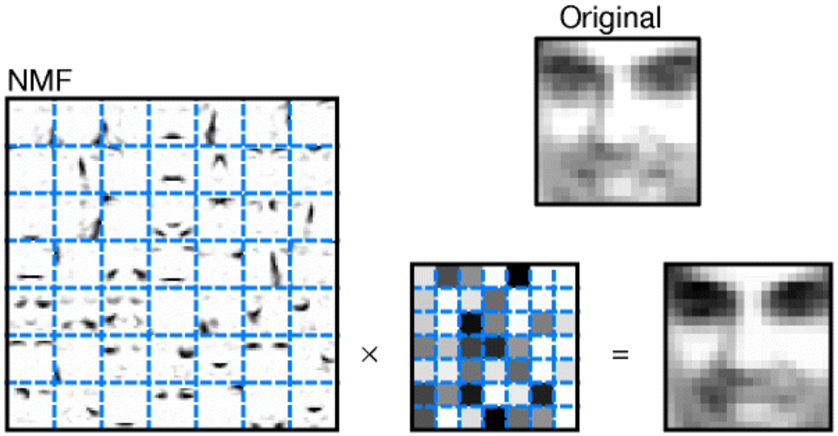
Alexandrov et al., Cell Reports 2013

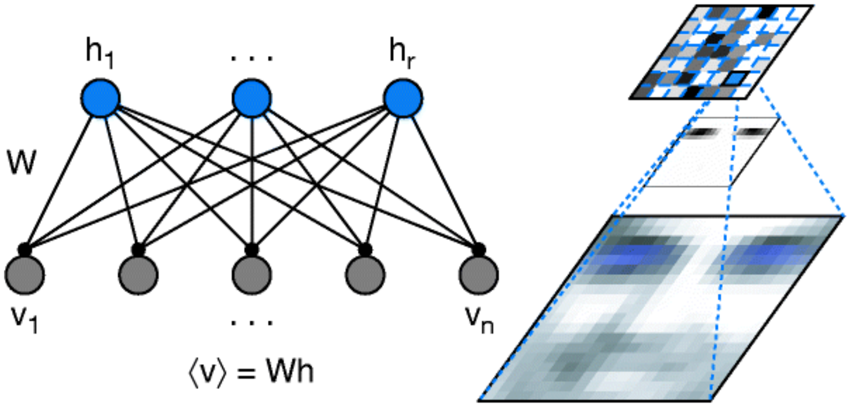
Non-negative matrix factorization



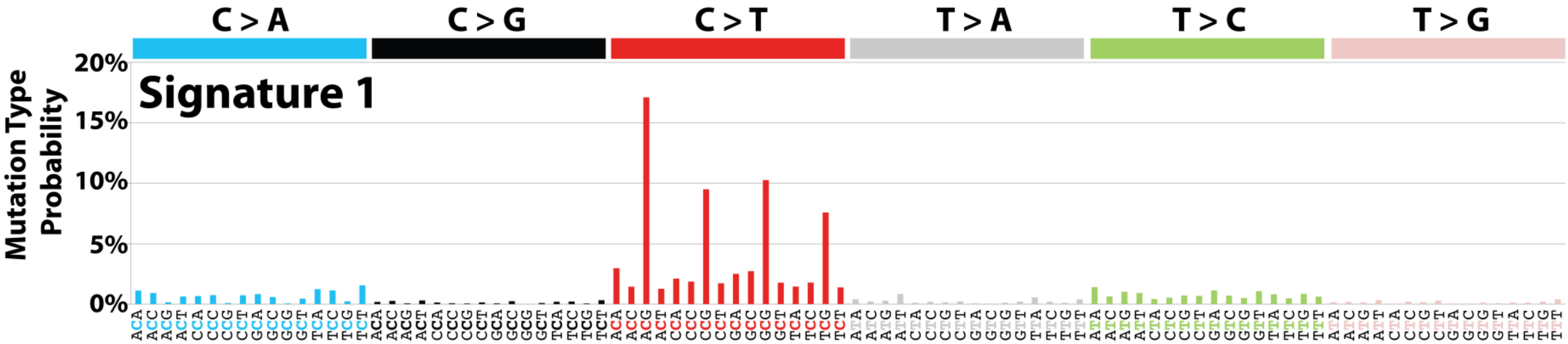
Lee and Seung, Nature 1999

Non-negative matrix factorization

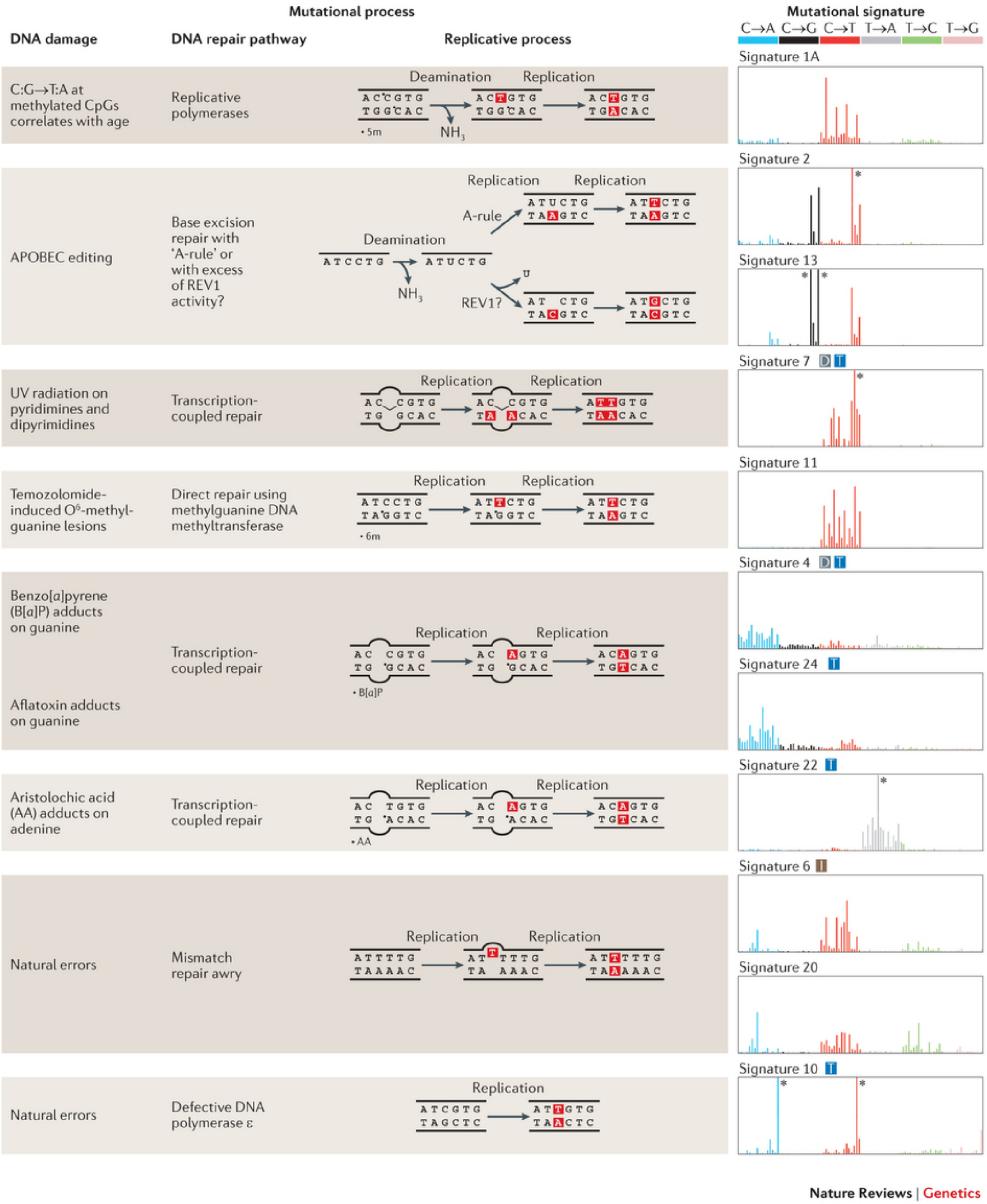




Lee and Seung, Nature 1999


C > A
C > G
C > T
T > A
T > C
T > G
Signature 1
Mutation Type Probability
20%
15%
10%
5%
0%
ACA ACC ACG ACT CCA CCC CCG CCT GCA GCC GCG GCT TCA TCC TCG TCT
ACA ACC ACG ACT CCA CCC CCG CCT GCA GCC GCG GCT TCA TCC TCG TCT
ACA ACC ACG ACT CCA CCC CCG CCT GCA GCC GCG GCT TCA TCC TCG TCT
ATA ATC ATG ATT CTA CTC CTG CTT GTA GTC GTG GTT TTA TTC TTG TTT
ATA ATC ATG ATT CTA CTC CTG CTT GTA GTC GTG GTT TTA TTC TTG TTT
ATA ATC ATG ATT CTA CTC CTG CTT GTA GTC GTG GTT TTA TTC TTG TTT

Summary of known mutational signatures



Helleday et al., 2014

Multisample analysis

MUTSIG(CV) – FINDING SIGNIFICANTLY MUTATED GENES

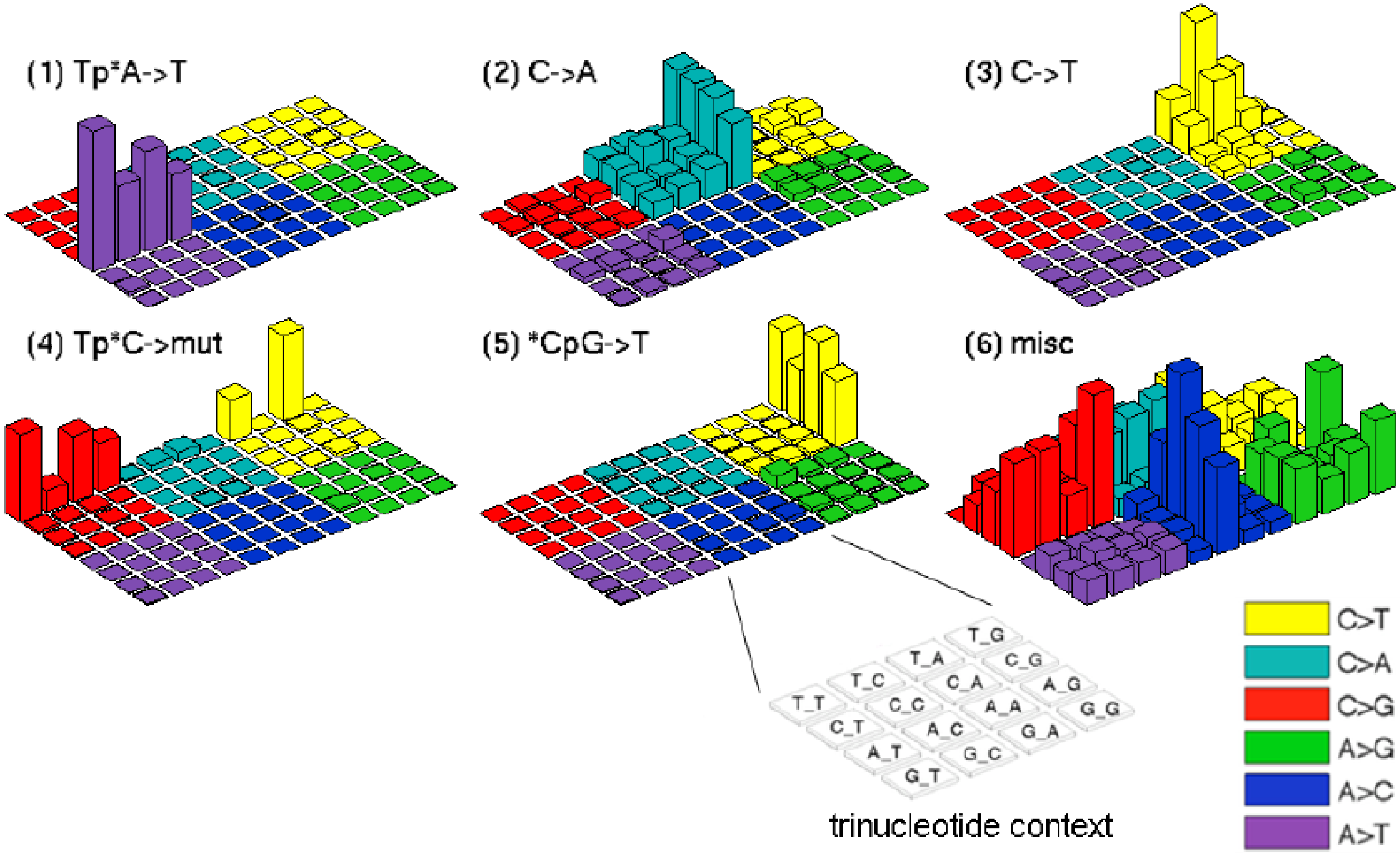
Identifying significantly mutated genes (drivers)

Key question: Is this gene mutated more often than expected by chance?

Key solution: More samples!

Problems? Yes, many. Can you think of them?

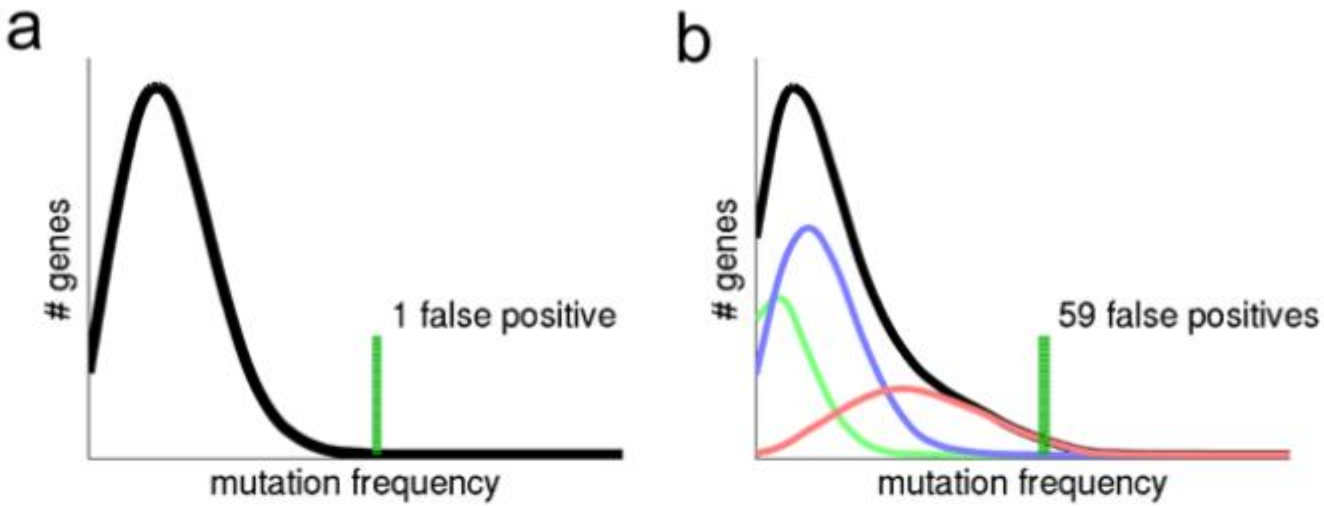
Mutsig


(1) Tp*A->T
(2) C->A
(3) C->T
(4) Tp*C->mut
(5) *CpG->T
(6) misc
T_G
T_A
C_G
T_C
C_A
A_G
T_T
C_C
A_A
G_G
C_T
A_C
G_A
A_T
G_C
G_T
trinucleotide context
C>T
C>A
C>G
A>G
A>C
A>T

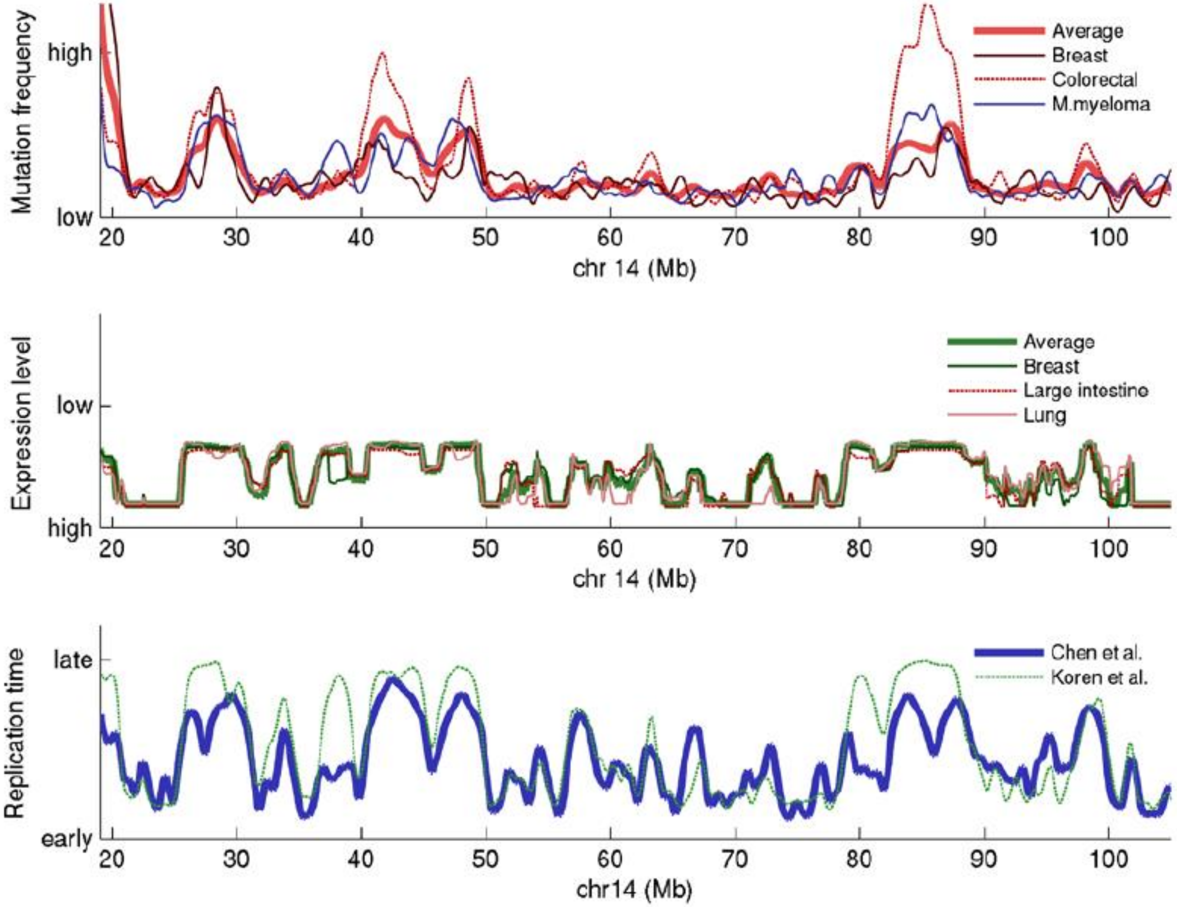
Mutsig - findings

- 450 novel significantly mutated genes!
  Wow!
- Many with suspicious biological function?
  Hmmmmm?
- One quarter were olfactory receptors?!
  Uh oh?!

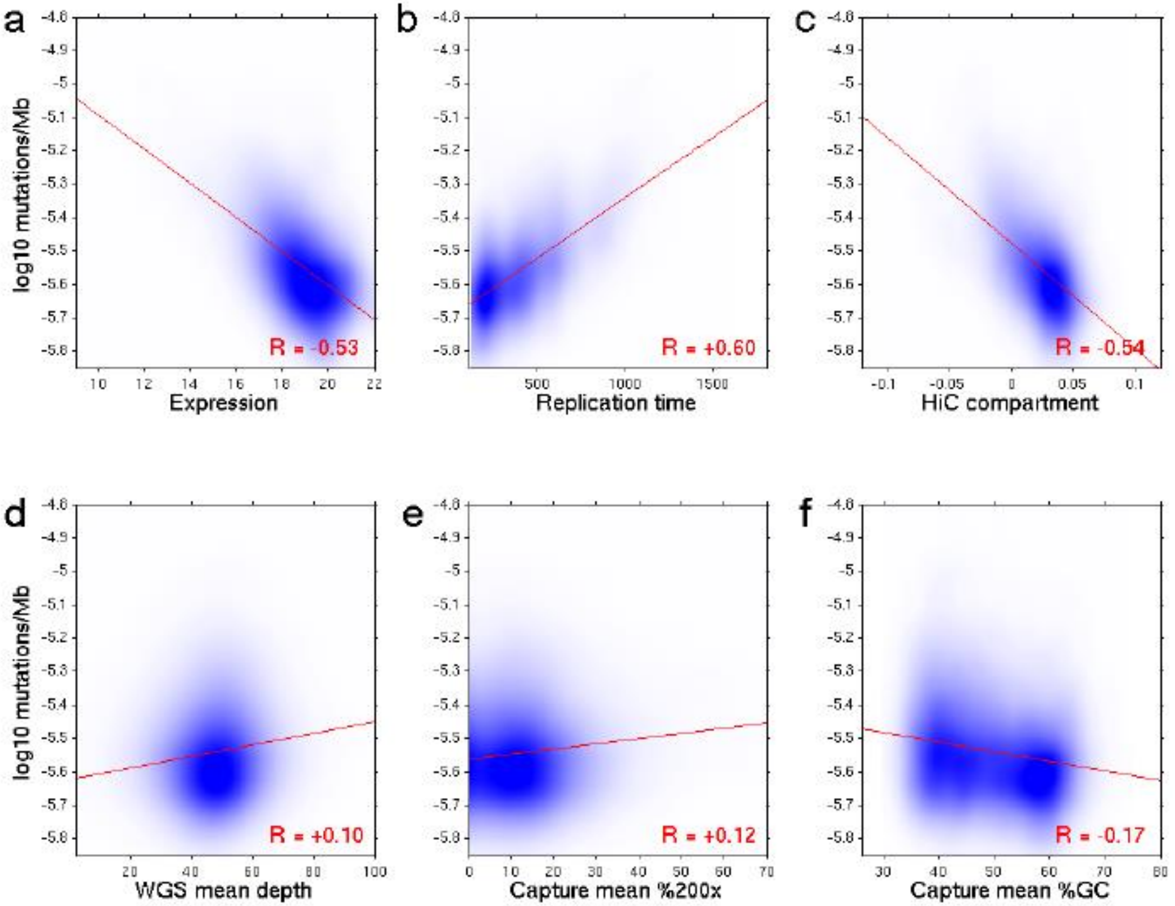
Mutation rate heterogeneity



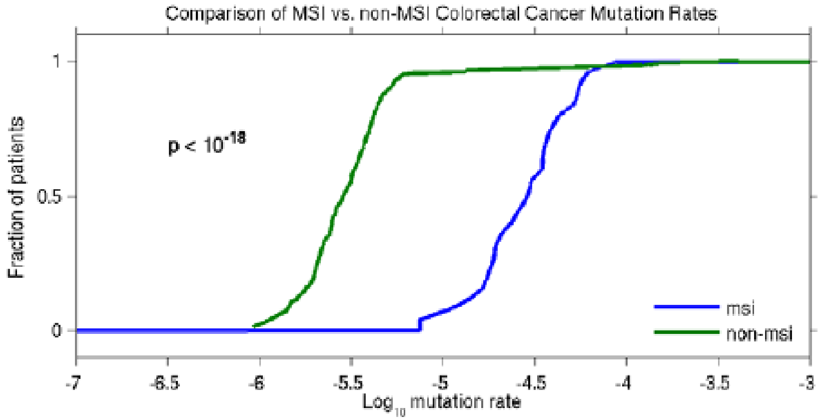
MutsigCV

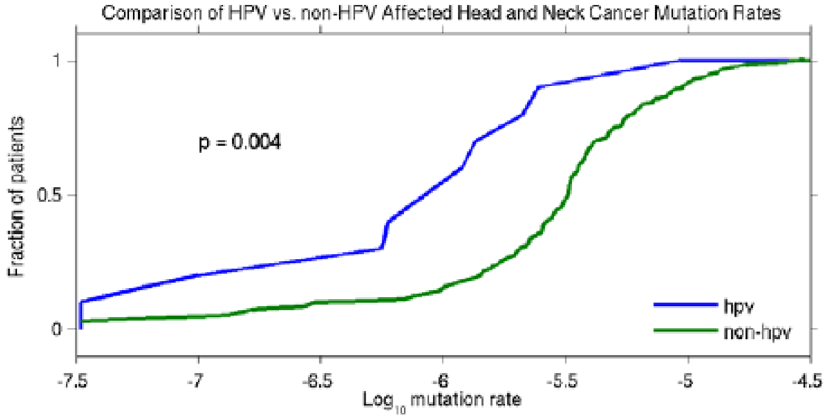

Mutation frequency
high
low
Average
Breast
Colorectal
M.myeloma
20
30
40
50
60
70
80
90
100
chr 14 (Mb)
Expression level
low
high
Average
Breast
Large intestine
Lung
chr 14 (Mb)
Replication time
late
early
Chen et al.
Koren et al.
chr14 (Mb)

MutsigCV


a
b
c
d
e
f
log10 mutations/Mb
-4.8
-4.9
-5
-5.1
-5.2
-5.3
-5.4
-5.5
-5.6
-5.7
-5.8
R = -0.53
R = +0.60
R = -0.54
R = +0.10
R = +0.12
R = -0.17
10 12 14 16 18 20 22
Expression
500 1000 1500
Replication time
-0.1 -0.05 0 0.05 0.1
HiC compartment
20 40 60 80 100
WGS mean depth
0 10 20 30 40 50 60 70
Capture mean %200x
30 40 50 60 70 80
Capture mean %GC

MutsigCV


Comparison of MSI vs. non-MSI Colorectal Cancer Mutation Rates


Comparison of HPV vs. non-HPV Affected Head and Neck Cancer Mutation Rates

MutsigCV

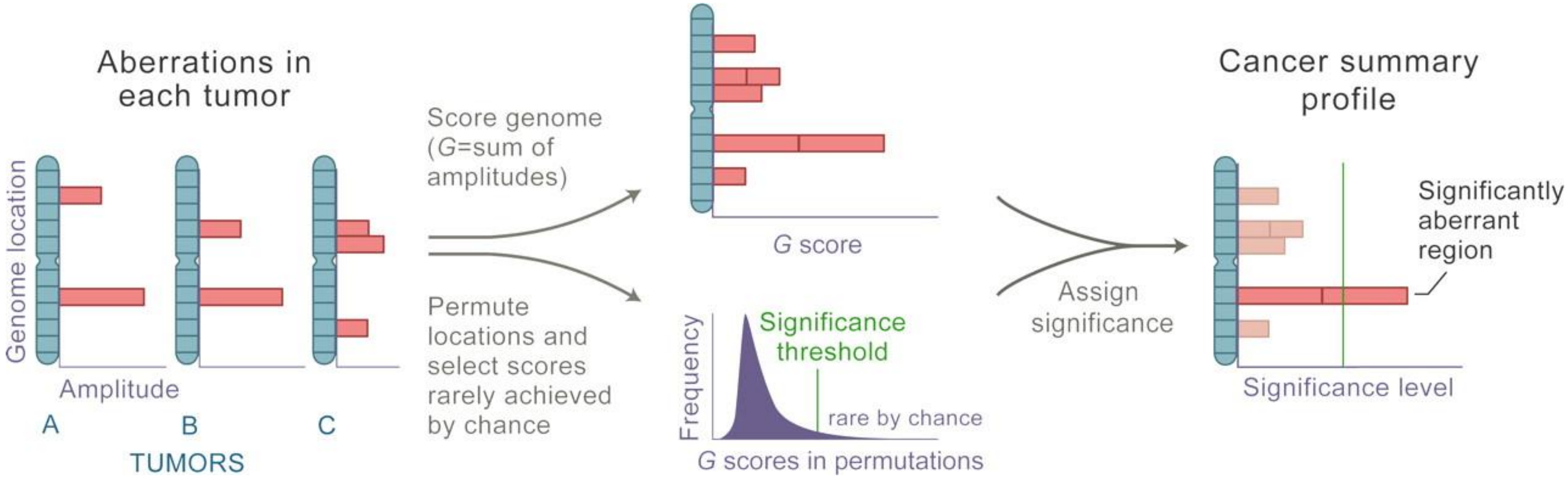
- http://www.broadinstitute.org/cancer/software/genepattern/modules/docs/MutSigCV

Multisample anlaysis

GISTIC – IDENTIFYING DRIVERS IN A SEA OF COPY-NUMBER CHANGE

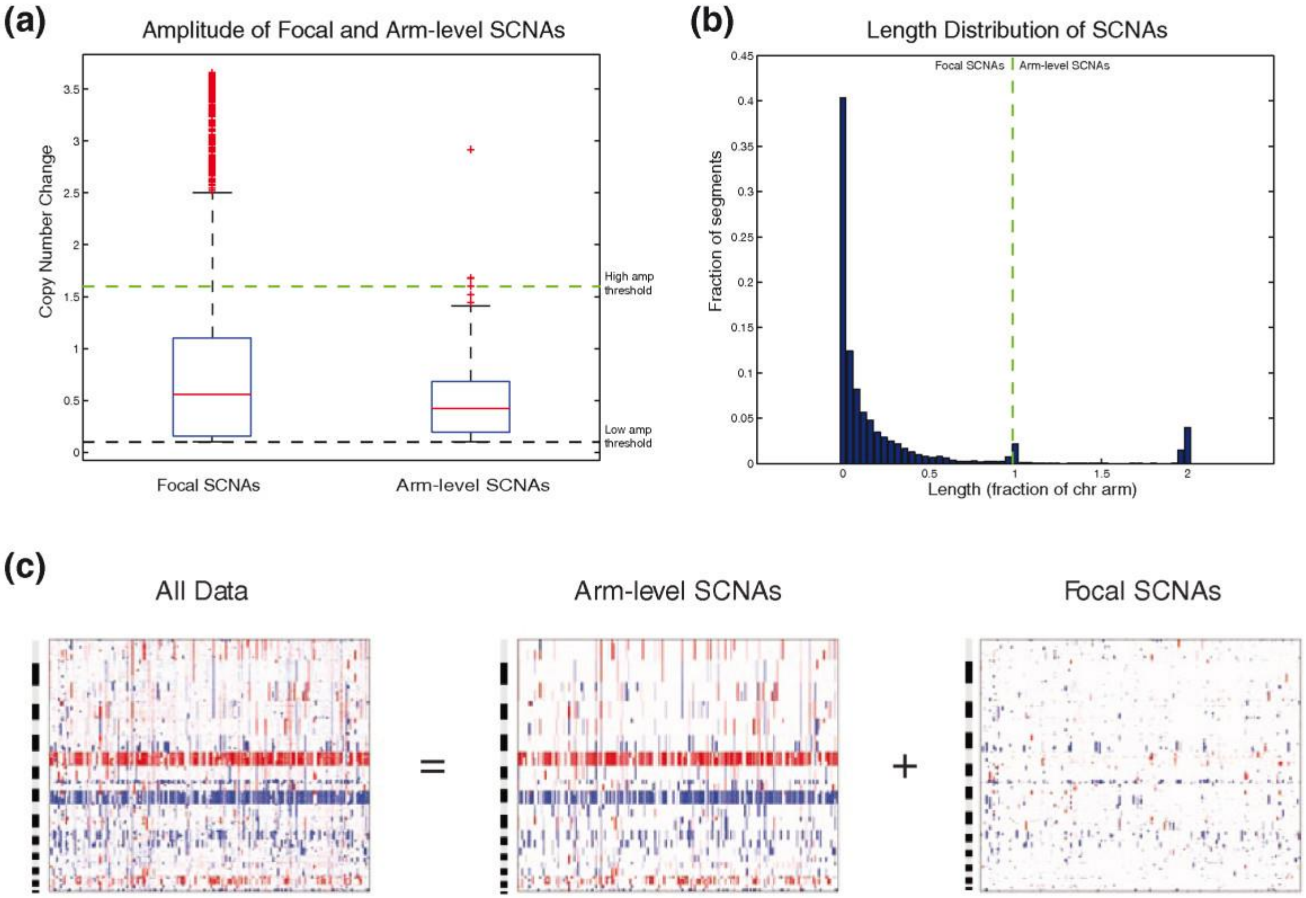
GISTIC (Genomic Identification of Significant Targets in Cancer)

GISTIC (Genomic Identification of Significant Targets in Cancer)



Rameen Beroukhim et al. PNAS 2007;104:20007-20012

GISTIC 2.0



(a)
Amplitude of Focal and Arm-level SCNAs
Copy Number Change
3.5
3
2.5
2
1.5
1
0.5
0
High amp threshold
Low amp threshold
Focal SCNAs
Arm-level SCNAs
(b)
Length Distribution of SCNAs
Fraction of segments
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Focal SCNAs
Arm-level SCNAs
0
0.5
1
1.5
2
Length (fraction of chr arm)
(c)
All Data
=
Arm-level SCNAs
+
Focal SCNAs

GISTIC

https://www.broadinstitute.org/software/cprg/?q=node/31