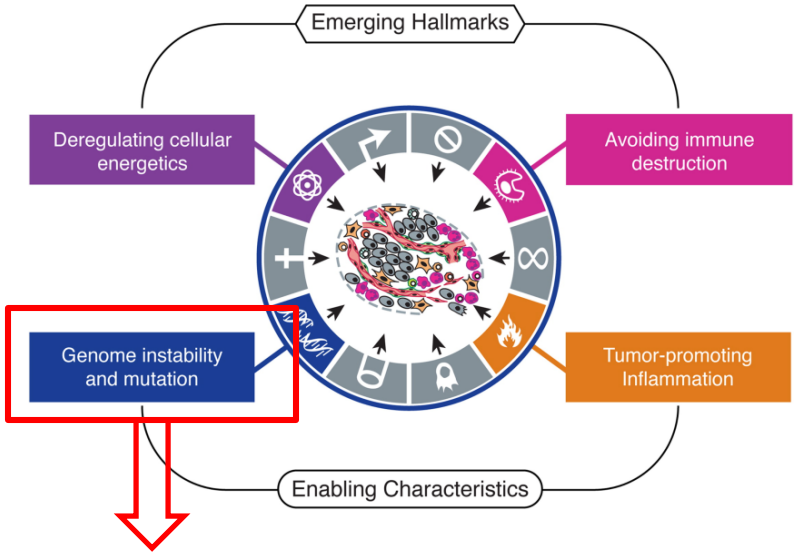
The impact of environmental and endogenous damage on human somatic mutation load

Natalie Saini
Mechanisms of Genome Dynamics Group
Genome Integrity and Structural Biology Laboratory

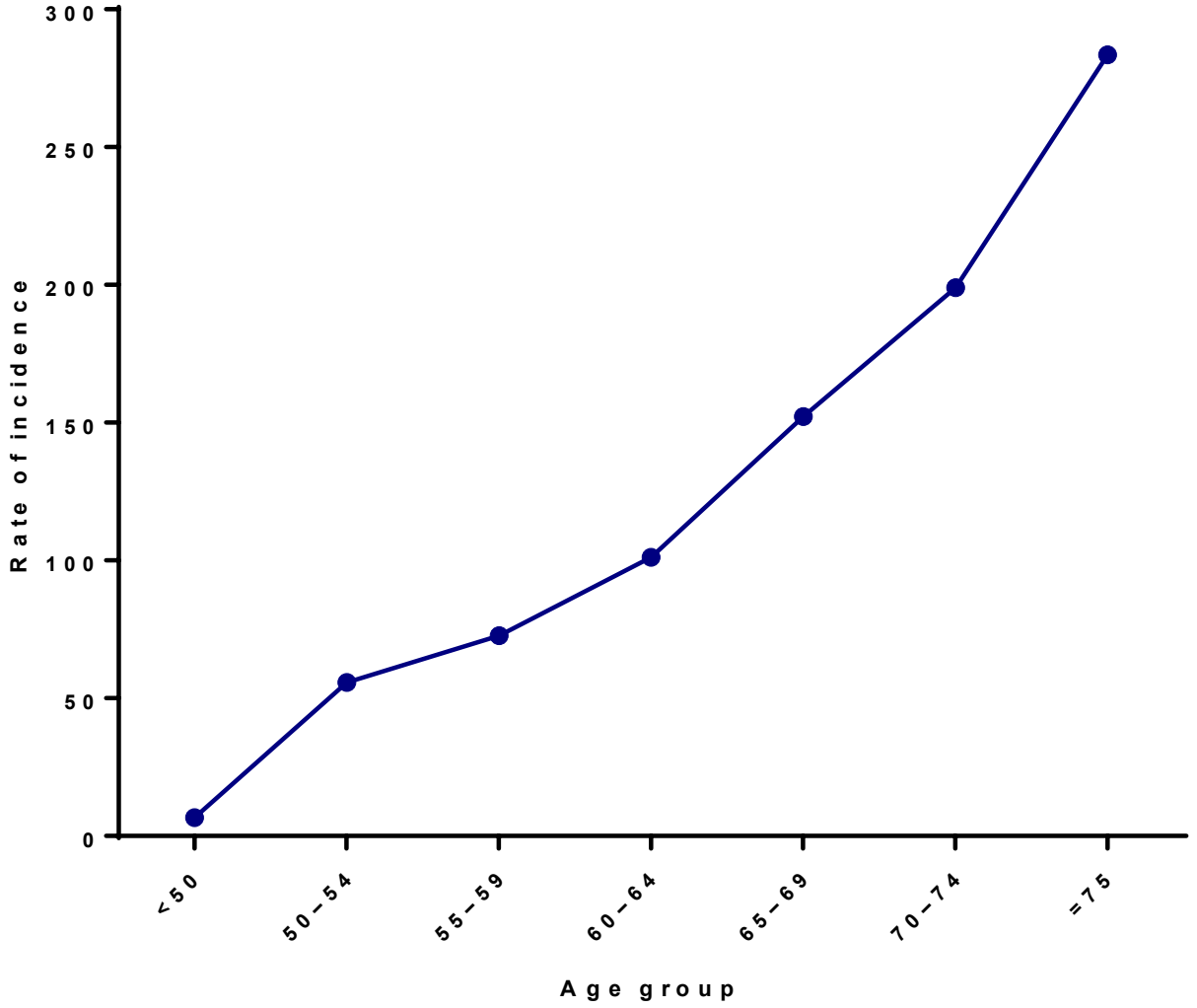
Somatic genome changes are associated with various diseases



Learning Disability	• Many different gene mutations and structural abnormalities
Autism	• Mosaic chromosomal structural abnormalities
Schizophrenia	• Aneuploidy of chromosome 1, 18 and X in brain
Alzheimer's disease	• Trisomy 21, or mutations in presinilin 1
Huntington’s disease	• Triplet repeat expansion
Friedreich’s ataxia	• Triplet repeat expansion
Ataxia-telangiectasia	• Chromosome 14-specific breaks, rearrangements
Primary biliary cirrhosis	• Mosaic monosomy of chromosome X
Autoimmune thyroid disease	• Mosaic monosomy of chromosome X

Cancer risk increases with age

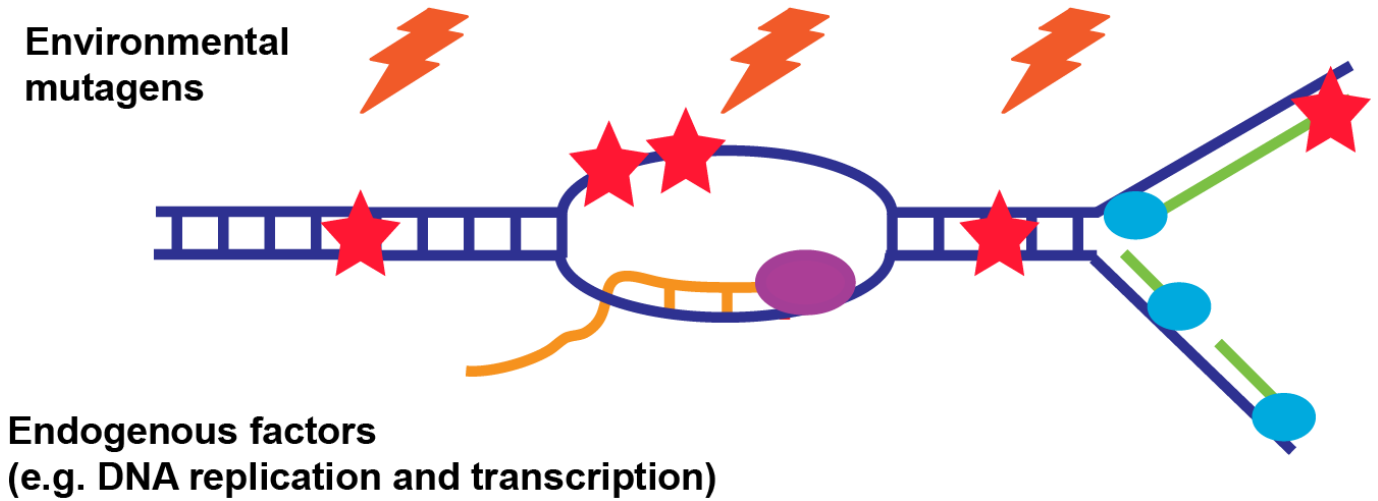
Colorectal cancer incidence rates per 100,000 population, by age group — USA, 2008



www.cdc.gov

Mutations accumulate during the lifetime of a person due to environmental and endogenous factors

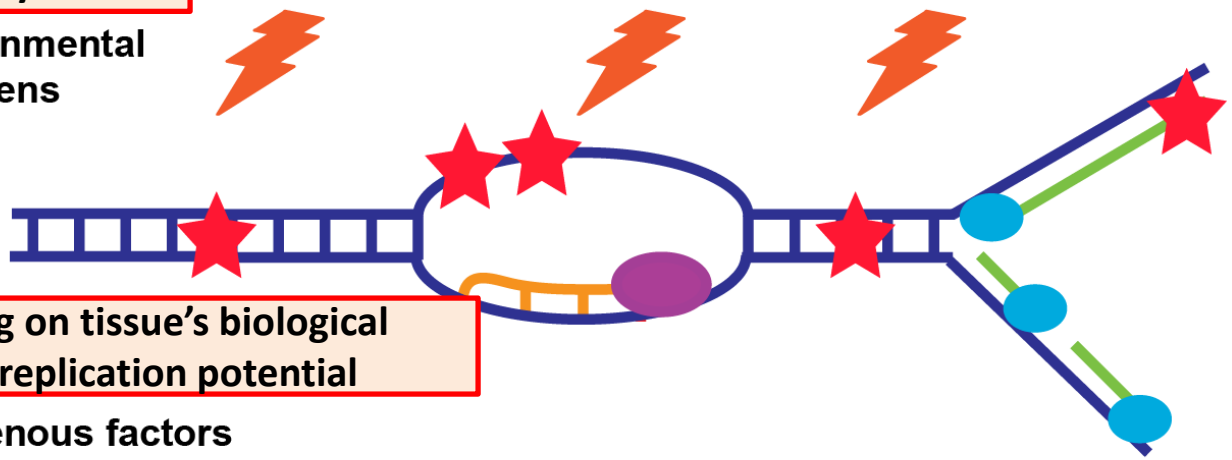
Mutation load



Mutations accumulate during the lifetime of a person due to environmental and endogenous factors

Exposure varies across the body

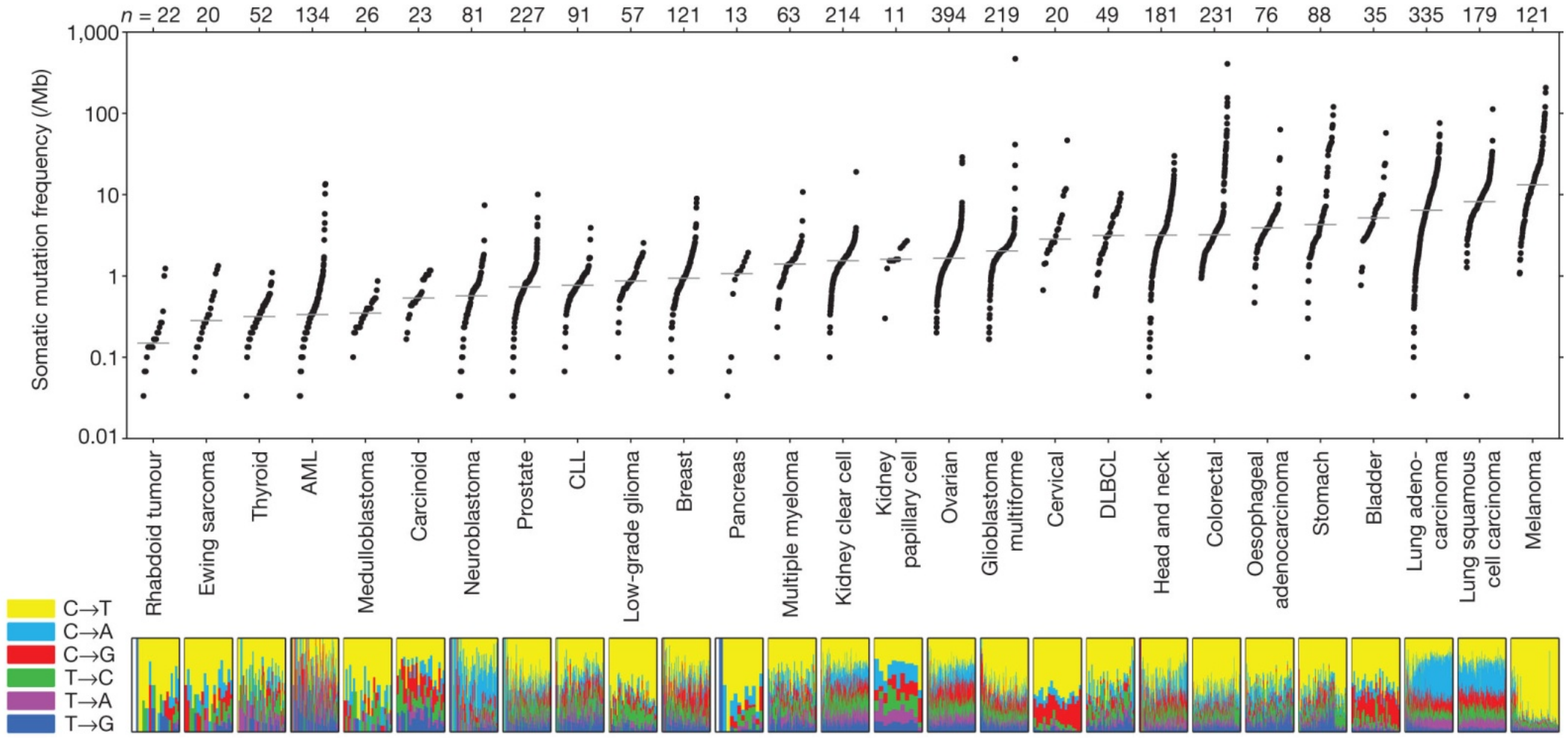
Environmental mutagens



Vary depending on tissue's biological features e.g. replication potential

Endogenous factors
(e.g. DNA replication and transcription)

Cancers genomes – wide variation in mutation load and spectra



Base substitutions in cancers range from: 10^2 to 10^6/genome

Lawrence *et al. Nature* 2013

Mutation load and signatures vary across the body in healthy individuals

BRAIN

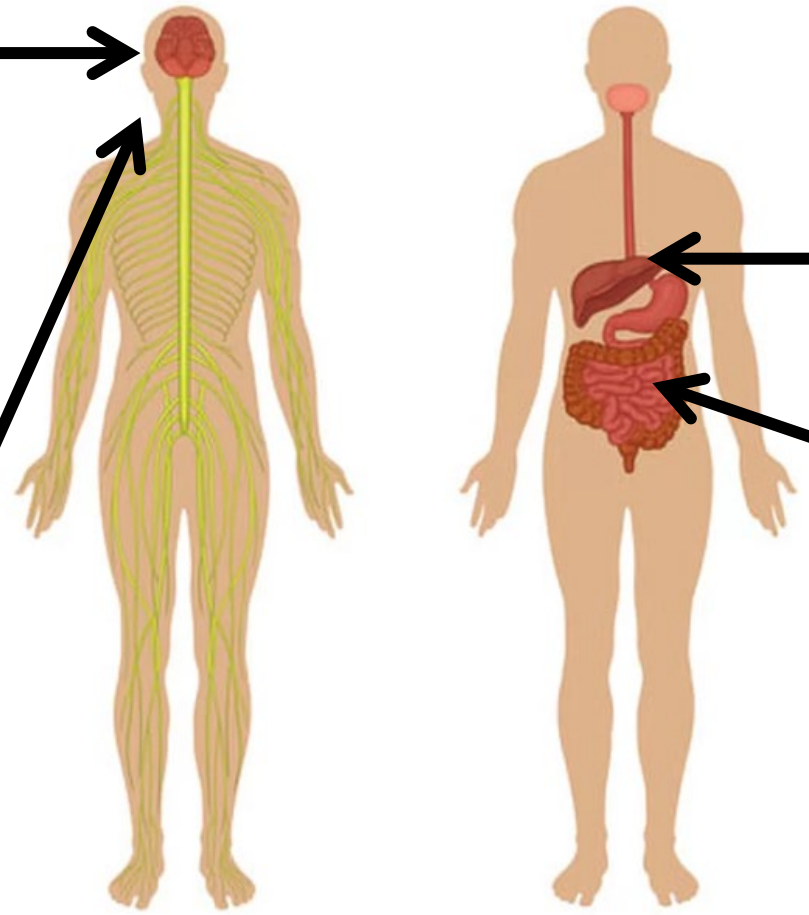
- ~1000 mutations per neuron.
- No increase with age.
- C→T changes at CpG motif

Lodato et al. Science. 2015

SKIN

- Deep sequencing of 74 genes - 3760 mutations found across the 234 samples from four individuals.
- C→T changes and CC→TT changes – UV signature

Martincorena *et al.*, Science. 2015



LIVER

- 1000-2000 mutations per cell.
- Linear increase with age.
- Unknown mutation signature.

SMALL INTESTINE and COLON

- 1000-3000 mutations per cell.
- Linear increase with age
- C→T changes at CpG motif

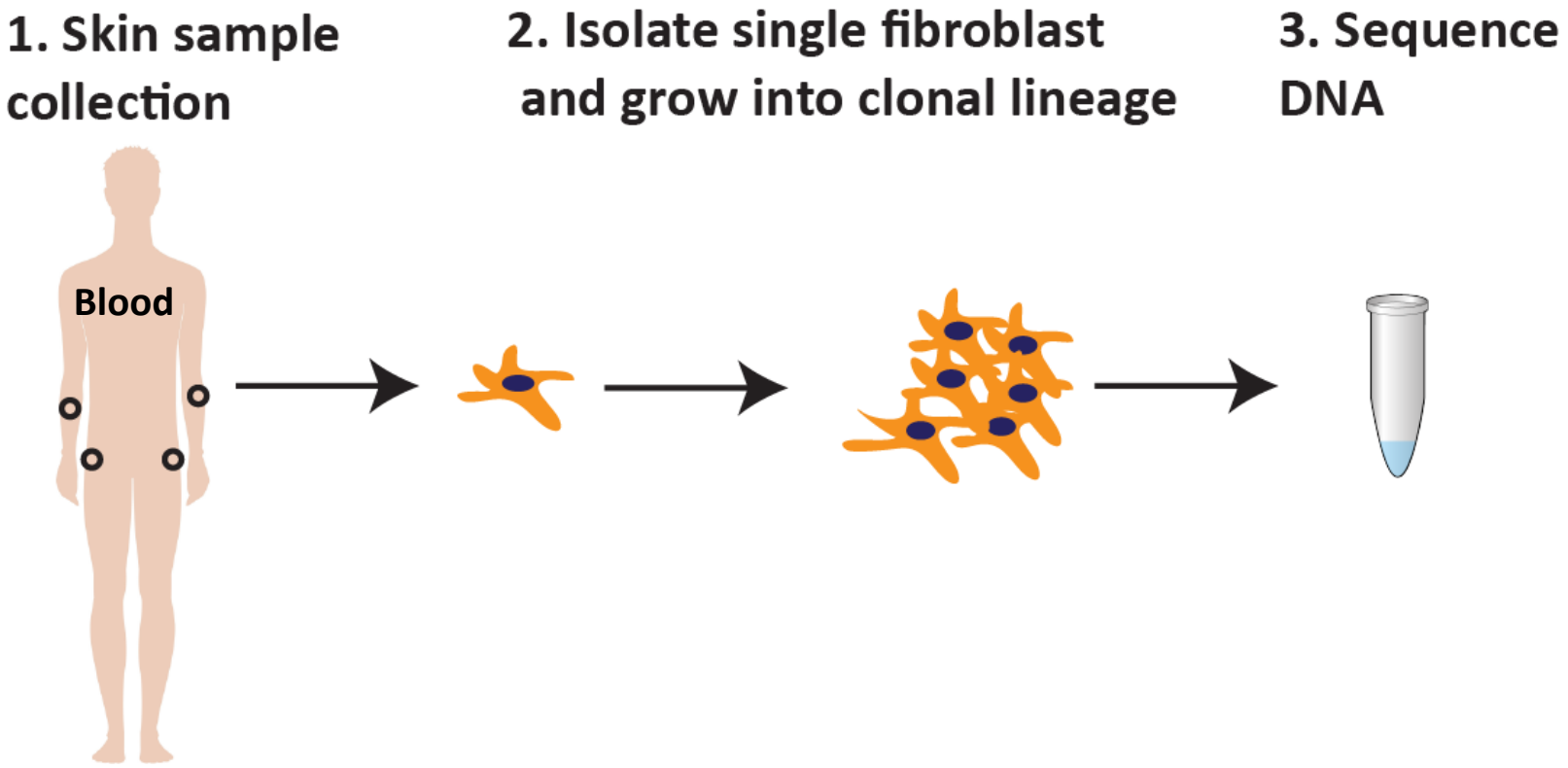
Blokzijl *et al.,* Nature. 2016

1. **What is the genome wide magnitude, spectrum and landscape of genomic changes accumulated in a healthy person?**

2. **Does exposure to an environmental DNA damaging agent affect mutation load?**

3. **What is the relative impact of environmental and endogenous factors to the mutation burden?**

Schematics of sample collection



Subjects: Caucasian males,

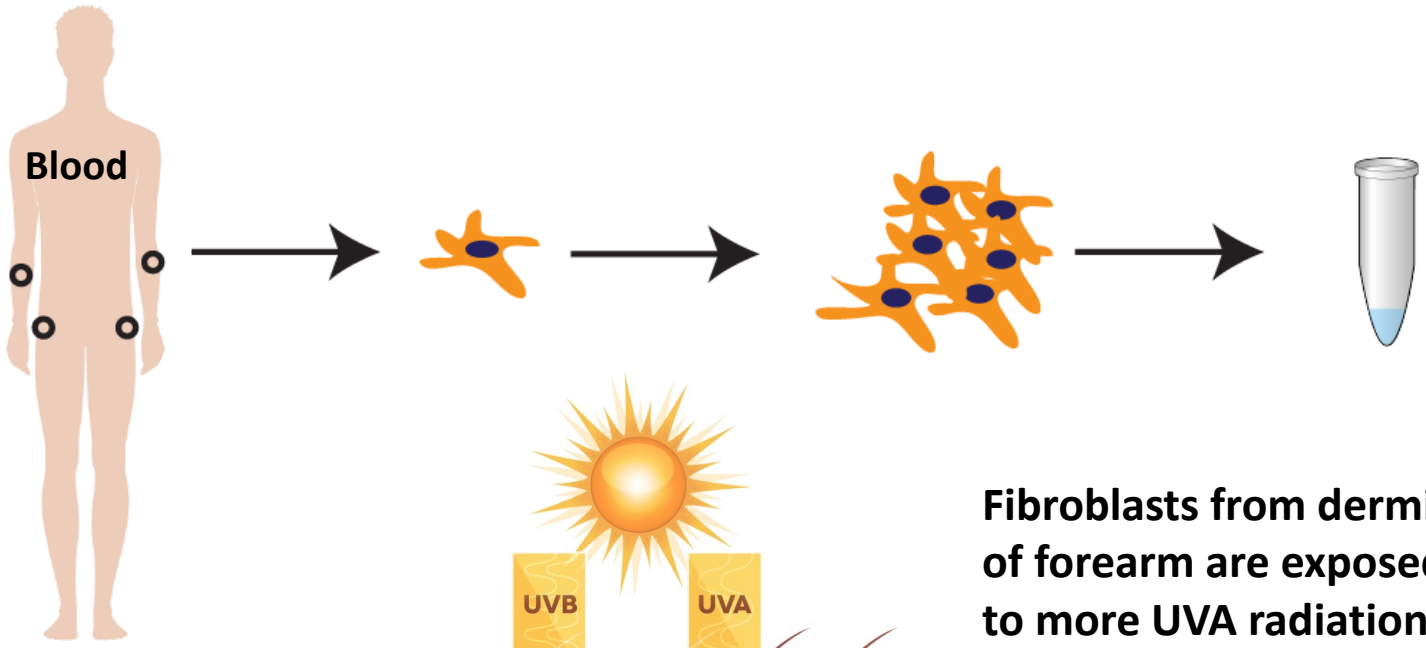
ages 58 and 62 yr

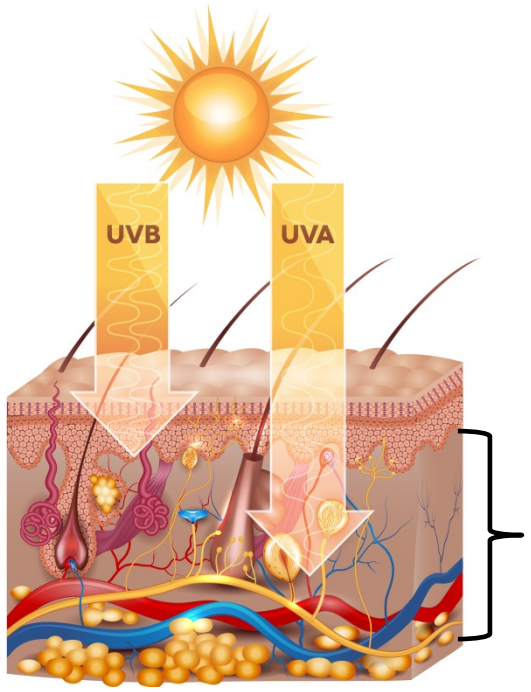
Schematics of sample collection

1. Skin sample collection

2. Isolate single fibroblast and grow into clonal lineage

3. Sequence DNA



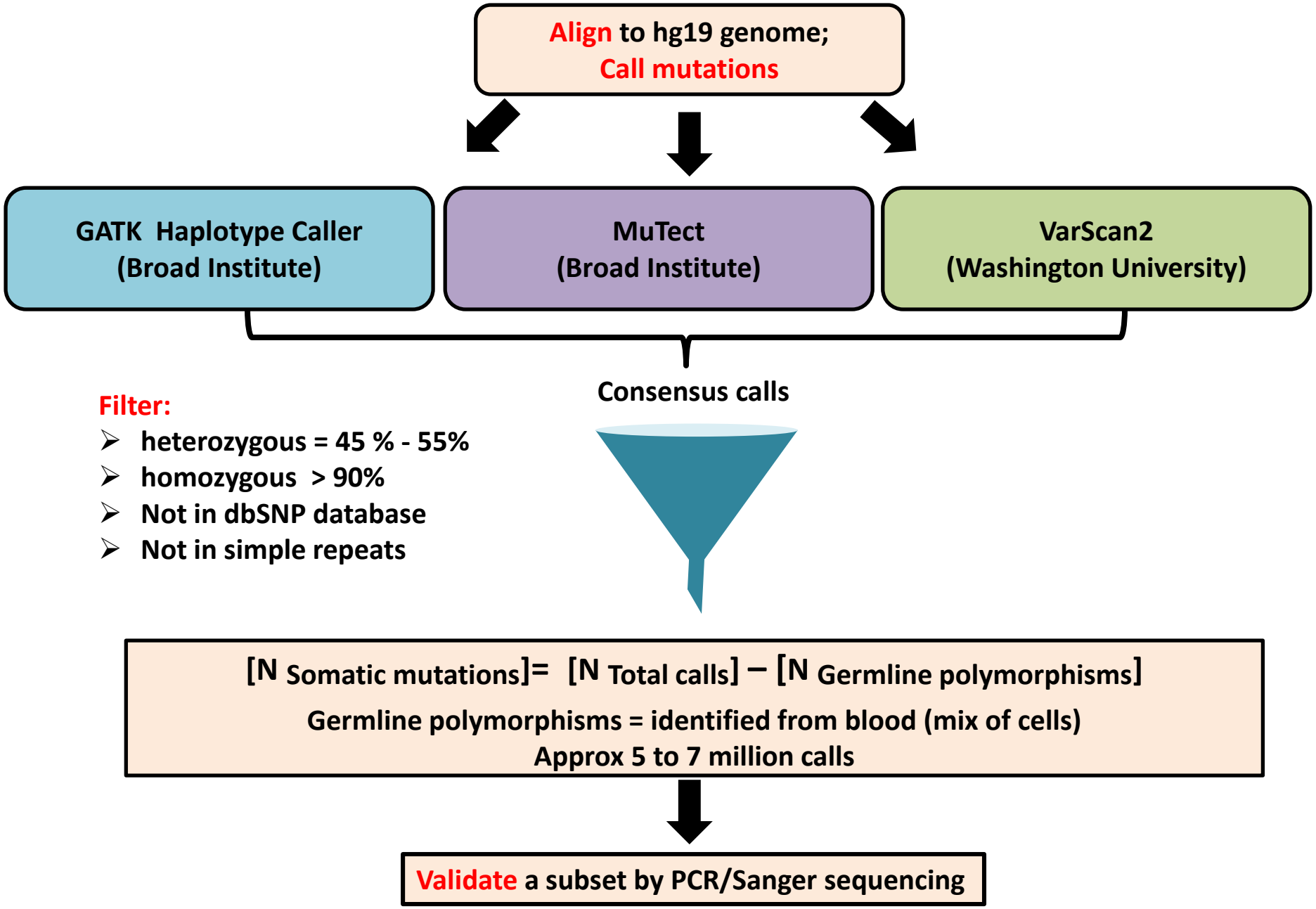


Fibroblasts from dermis of forearm are exposed to more UVA radiation

Subjects: Caucasian males, ages 58 and 62 yr

Dermis

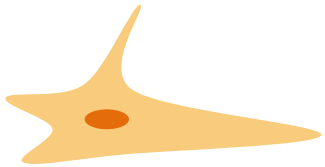
Accurate minimum estimate of somatic mutation load in a clone

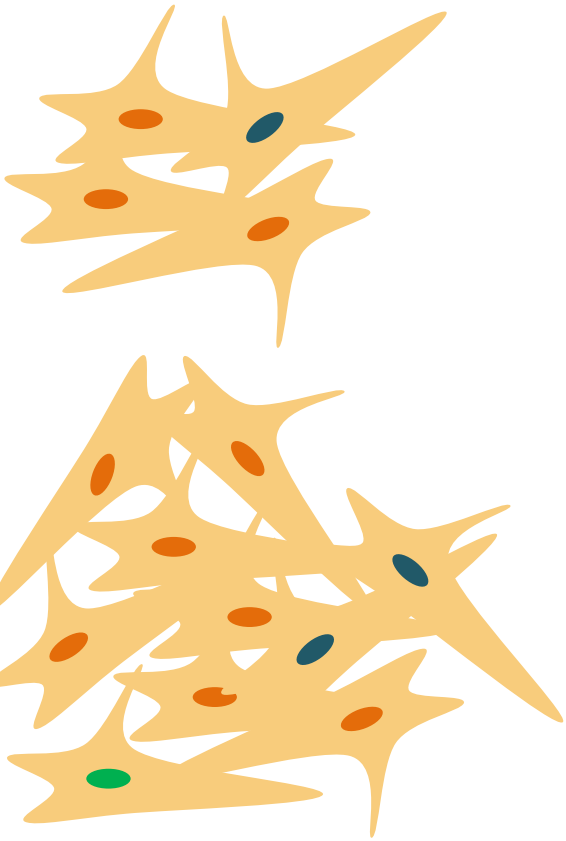


Filtering by allele frequency removes mutations added during growth of the cells in culture

Mutations in original cell

Allele frequency for:
Heterozygous allele = 50%
Homozygous allele = 100%





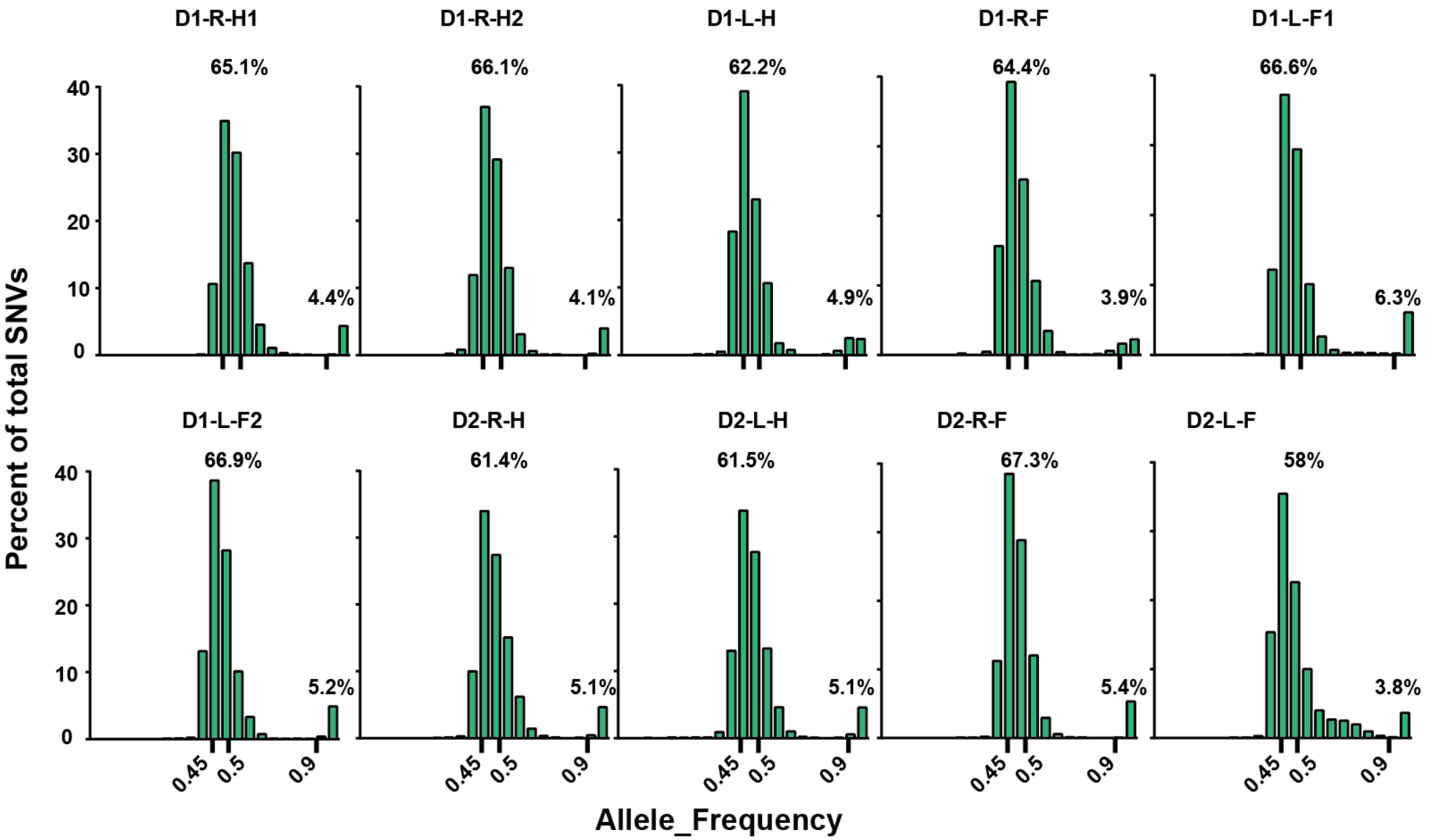
Mutations arising during culture

Allele frequency for:
Heterozygous allele < 50%
Homozygous allele < 100%

Allele frequency for:
Heterozygous allele = 45 - 50%
Homozygous allele = 90 - 100%

Allele frequency for:
Heterozygous allele < 50%
Homozygous allele < 100%

Majority of the consensus SNVs are clonal

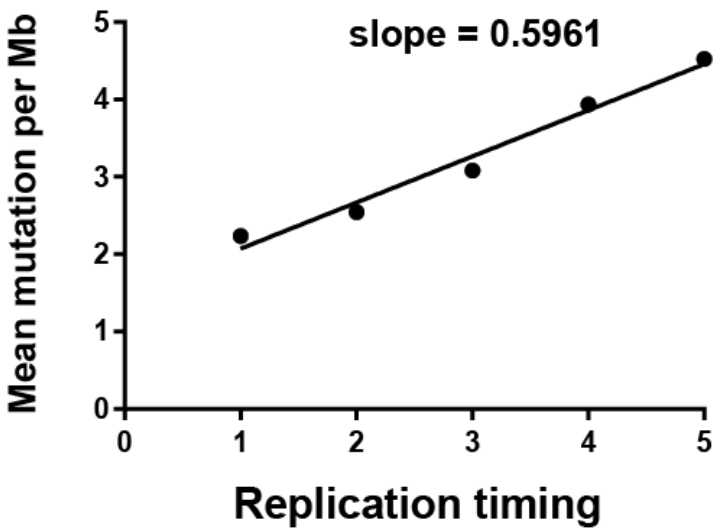


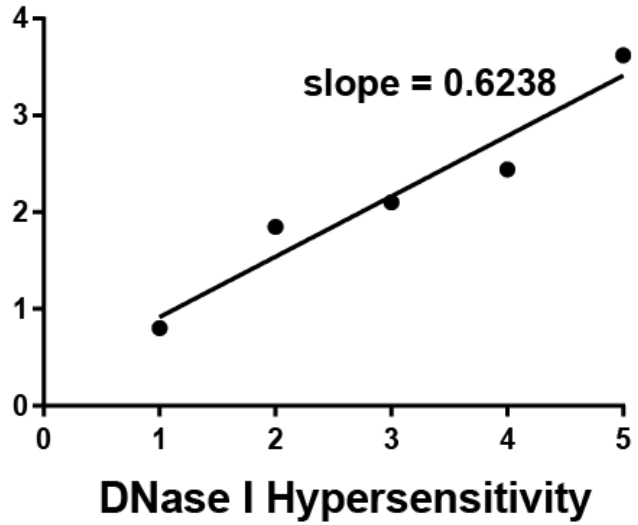
~1000 somatic base substitutions and at least 1 rearrangement are present in normal skin fibroblasts

Donor	Site	ID	Mutations	Rearrangements
Donor1	Right hip	D1-R-H1	1373	1
	Right hip	D1-R-H2	707	4
	Left hip	D1-L-H	581	3
	Right forearm	D1-R-F	1056	5
	Left forearm	D1-L-F1	5309	25
	Left forearm	D1-L-F2	3879	5
Donor2	Right hip	D2-R-H	1981	1
	Left hip	D2-L-H	4612	2
	Right forearm	D2-R-F	12743	5
	Left forearm	D2-L-F	8600	8

- Similar to median mutation load in cancers
- Mutations in forearms are more than mutations in hips

Mutation load depends on replication timing and chromatin status



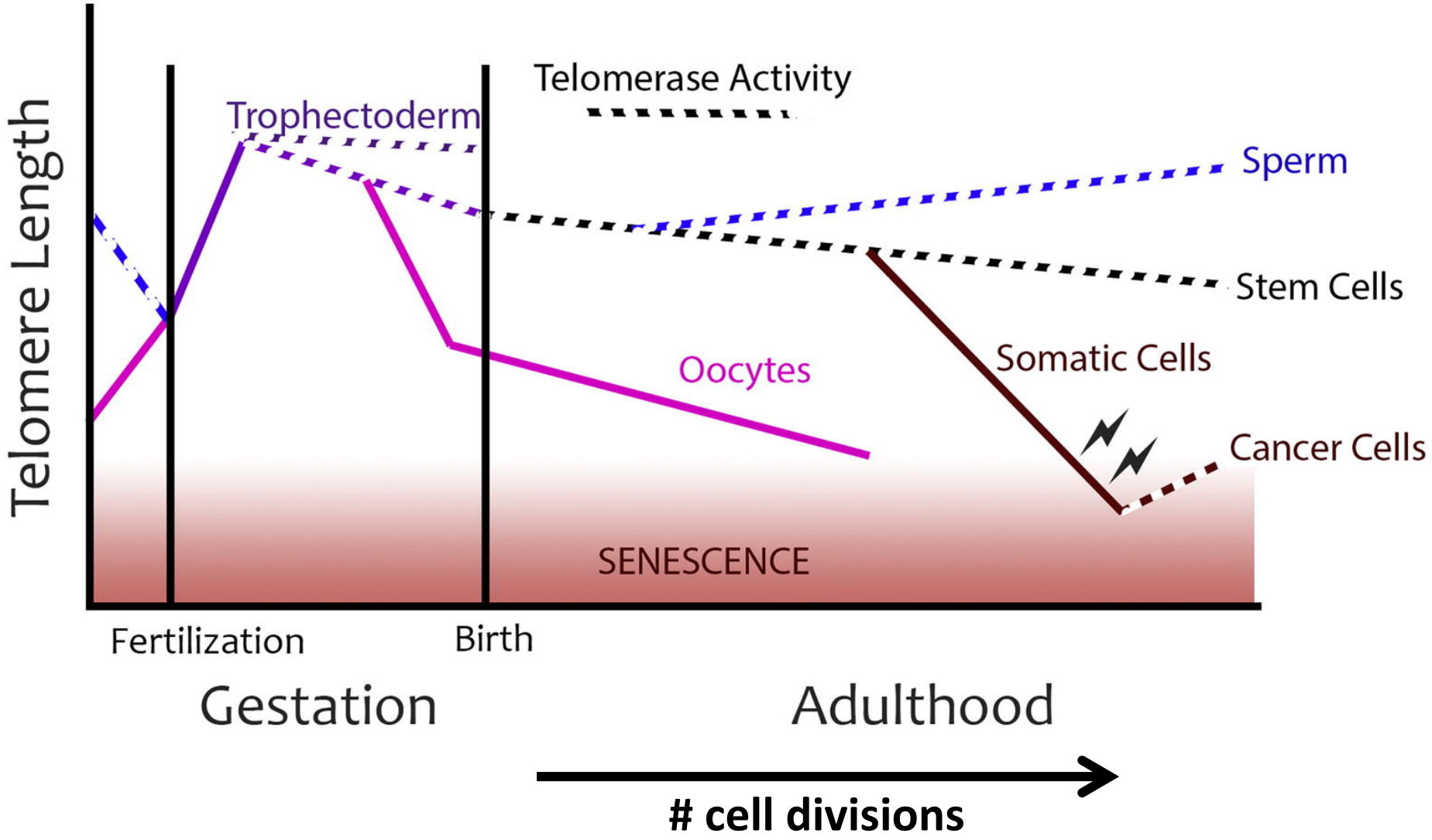


Early → Late

Open Chromatin → Closed Chromatin

Mechanisms leading to mutations in healthy fibroblasts are similar to cancers.

Telomere length as a proxy for number of cell divisions in human cells



Kalmbach *et al.* Fertility and Sterility, 2013

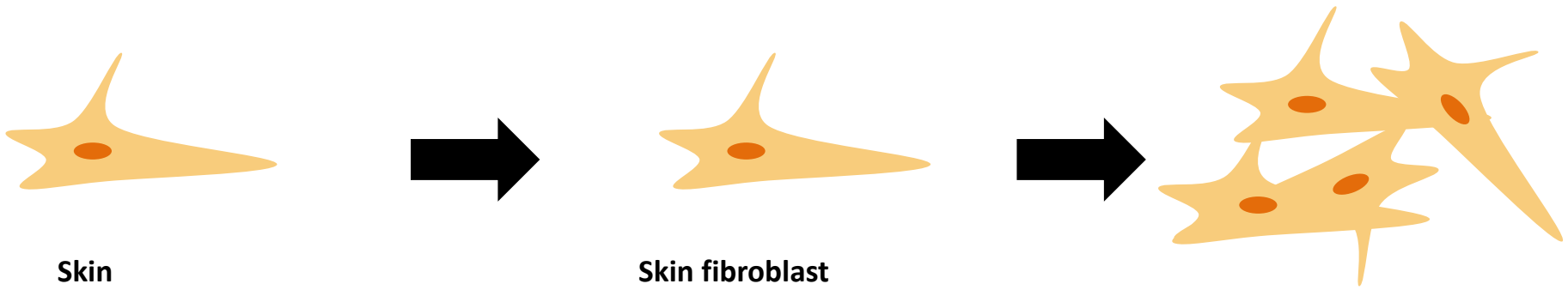
Estimation of telomere length in clones

Samples	Mutations	Estimated telomere length (kb)
D1-R-H1	1373	3.04
D1-R-H2	707	2.21
D1-L-H	581	2.25
D1-R-F	1056	2.25
D1-L-F1	5309	2.29
D1-L-F2	3879	2.03
D2-R-H	1981	2.26
D2-L-H	4612	2.40
D2-R-F	12743	2.12
D2-L-F	8600	2.07

Software used : Telseq

Telomere length as a proxy for number of cell divisions in human cells

- 50 – 100bp telomeres lost per cell division.
- At birth average telomere length is 10.5 -11kb.



Skin fibroblast at birth

Skin fibroblast isolated from Donor at age 58 – 62y

Clonal population sequenced

Telomere length as a proxy for number of cell divisions in human cells

- 50 – 100bp telomeres lost per cell division.
- At birth average telomere length is 10.5 -11kb.

Skin fibroblast at birth

Skin fibroblast isolated from Donor at age 58 – 62y

Clonal population sequenced

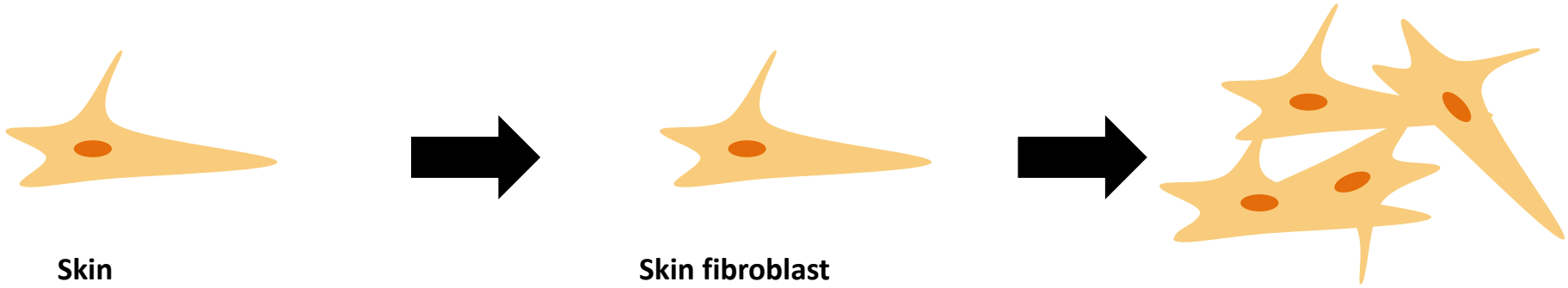
Telomere size ~4kb

20 cell divisions

Telomere size ~2kb

Telomere length as a proxy for number of cell divisions in human cells

- 50 – 100bp telomeres lost per cell division.
- At birth average telomere length is 10.5 -11kb.



Skin fibroblast at birth

Skin fibroblast isolated from Donor at age 58 – 62y

Clonal population sequenced

Telomere size ~11kb

65 – 70 cell divisions

Telomere size ~4kb

20 cell divisions

Telomere size ~2kb

Somatic mutation rates per cell division in clones

Samples	Mutations	Estimated telomere length (kb)	Number of cell divisions (from birth)	Mutation/genome/ cell division (from birth)
D1-R-H1	1373	3.04	60	23
D1-R-H2	707	2.21	68	10
D1-L-H	581	2.25	67	9
D1-R-F	1056	2.25	68	16
D1-L-F1	5309	2.29	67	79
D1-L-F2	3879	2.03	70	56
D2-R-H	1981	2.26	67	29
D2-L-H	4612	2.40	66	70
D2-R-F	12743	2.12	69	185
D2-L-F	8600	2.07	69	124

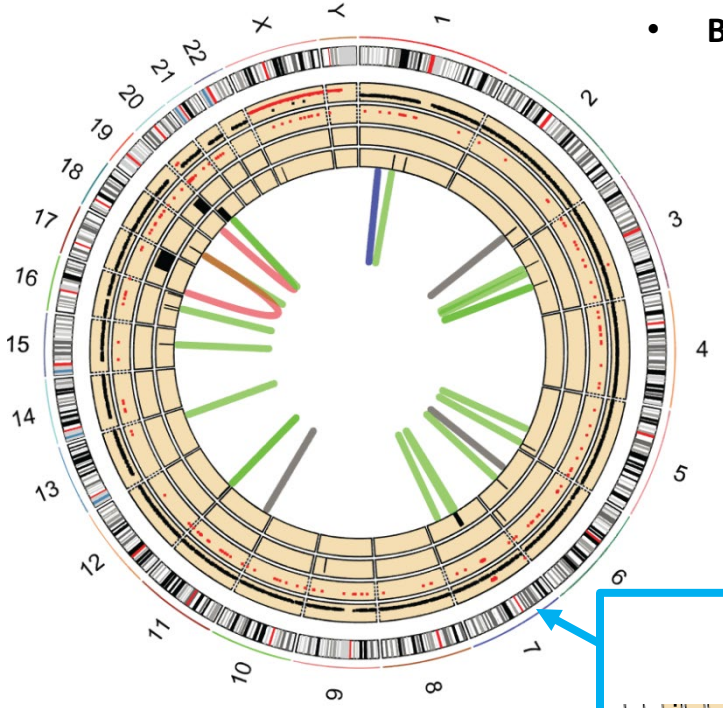
Estimates are maximal mutations/genome/cell division.

Number of cell divisions from zygote to newborn are not known.

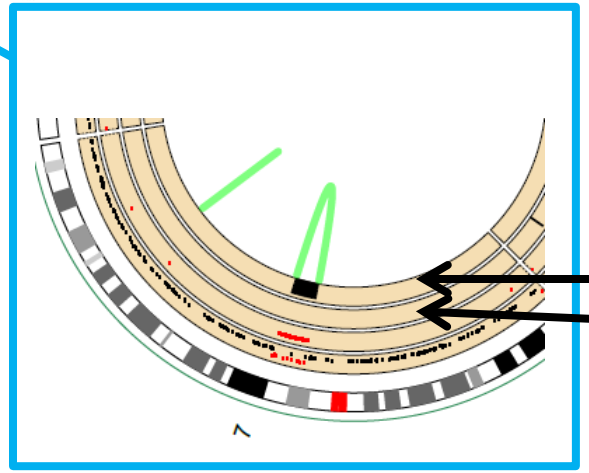
Software used : Telseq

Example of rearrangements seen in the clones

- Green = Deletion,
- Red = Translocation
- Blue = Inversion
- Black = Duplication



D1-L-F1



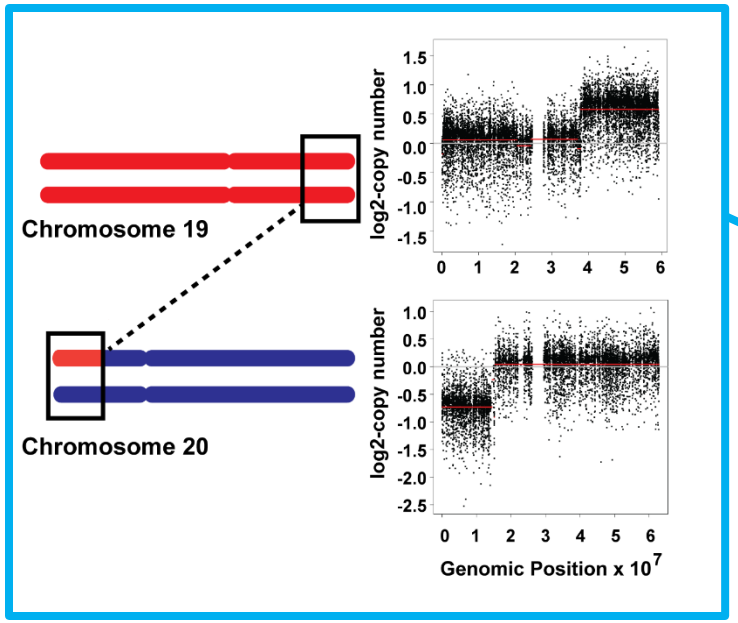
CNV – 1n

CNV – 3n

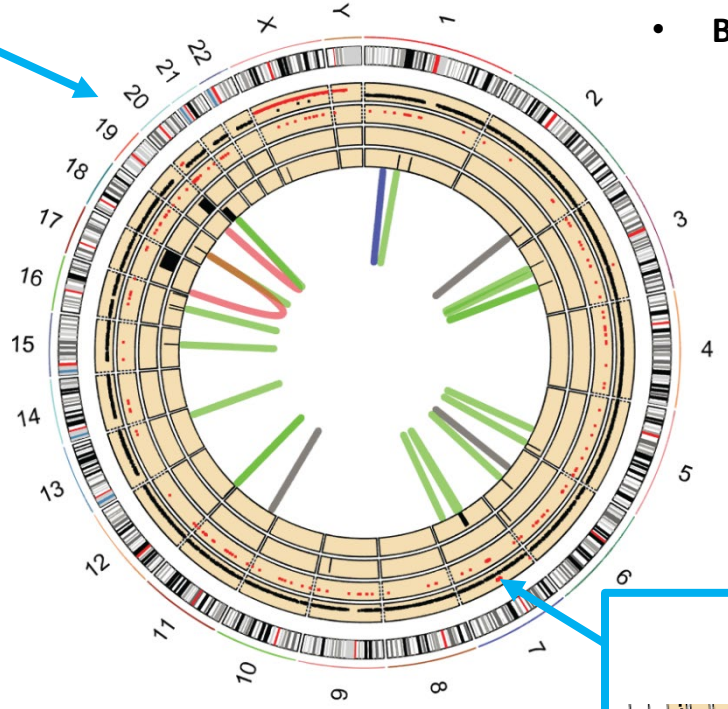
Red dots = Position of LOH/Homozygous mutations

Black dots = heterozygous mutations

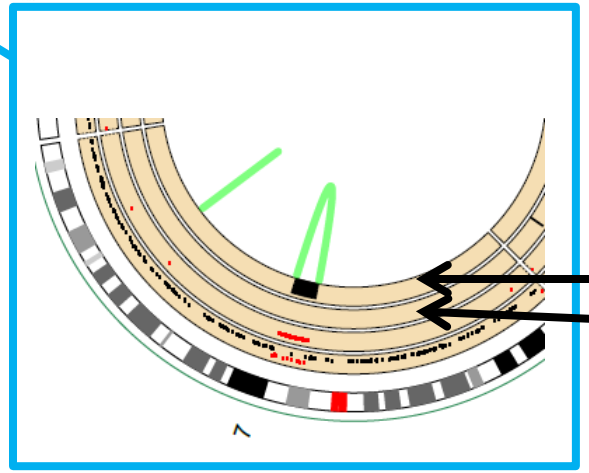
Example of rearrangements seen in the clones



- Green = Deletion,
- Red = Translocation
- Blue = Inversion
- Black = Duplication



D1-L-F1



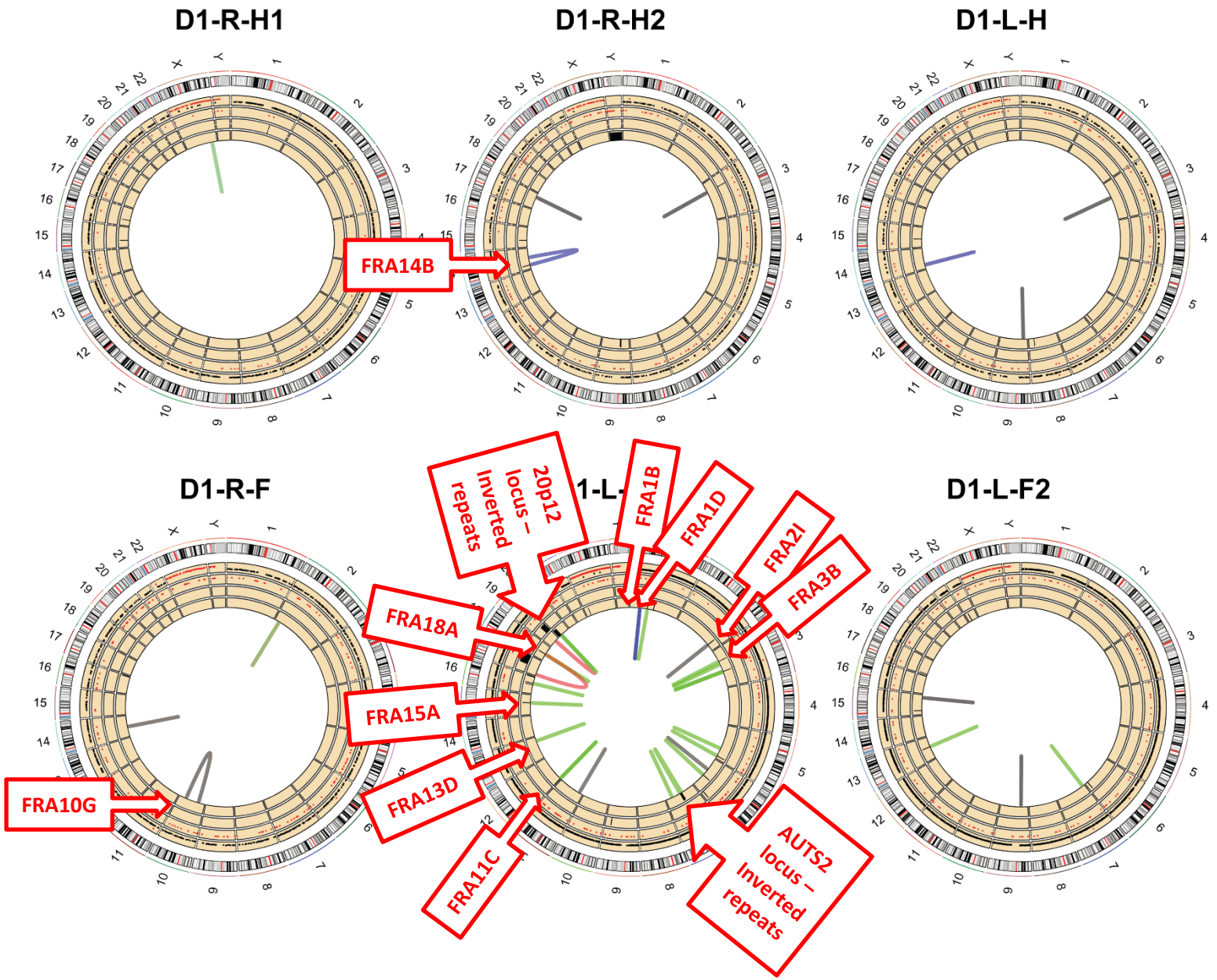
CNV – 1n

CNV – 3n

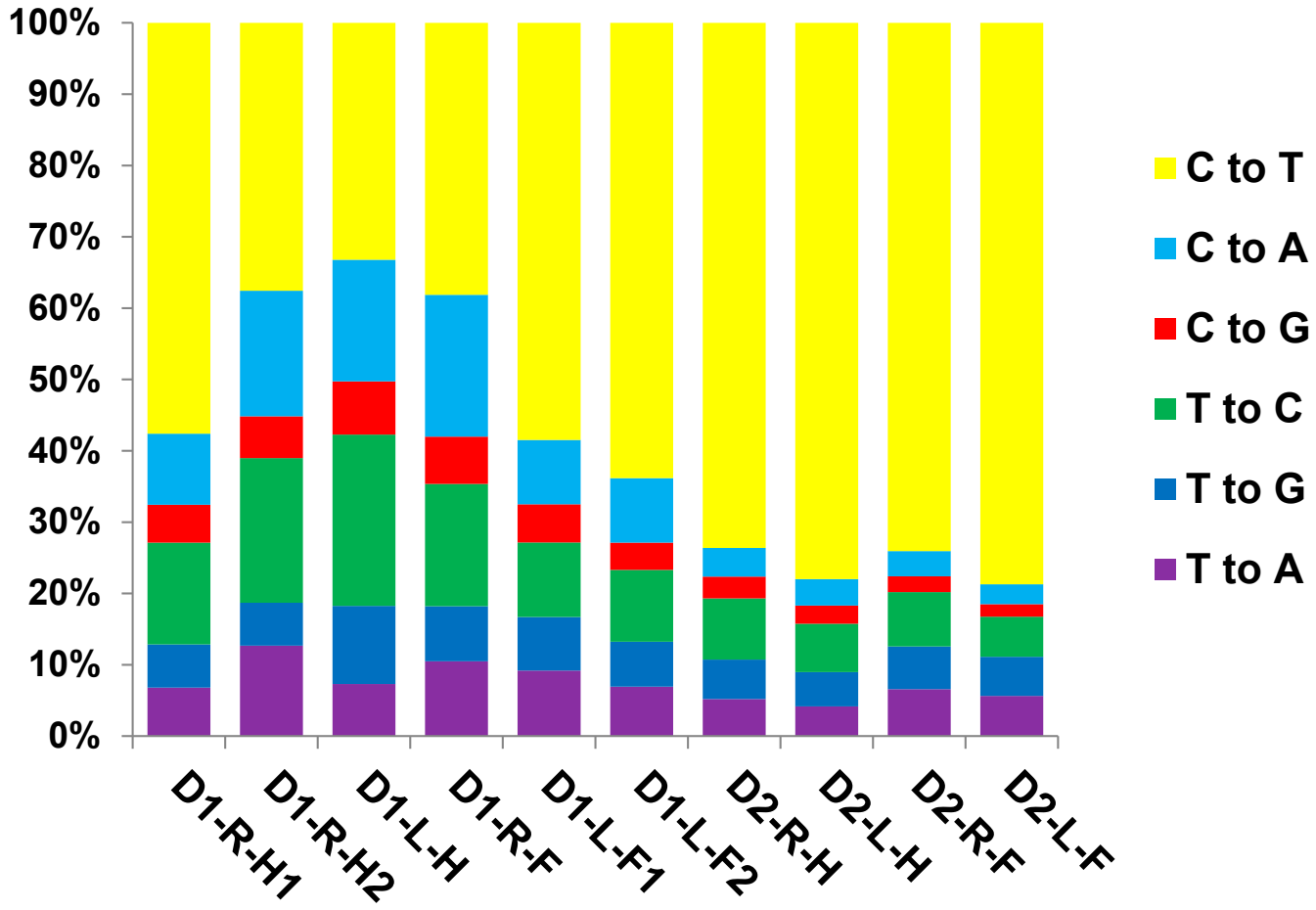
Red dots = Position of LOH/Homozygous mutations

Black dots = heterozygous mutations

Rearrangements were often in vicinity of common fragile sites



C→T changes are prevalent in all samples



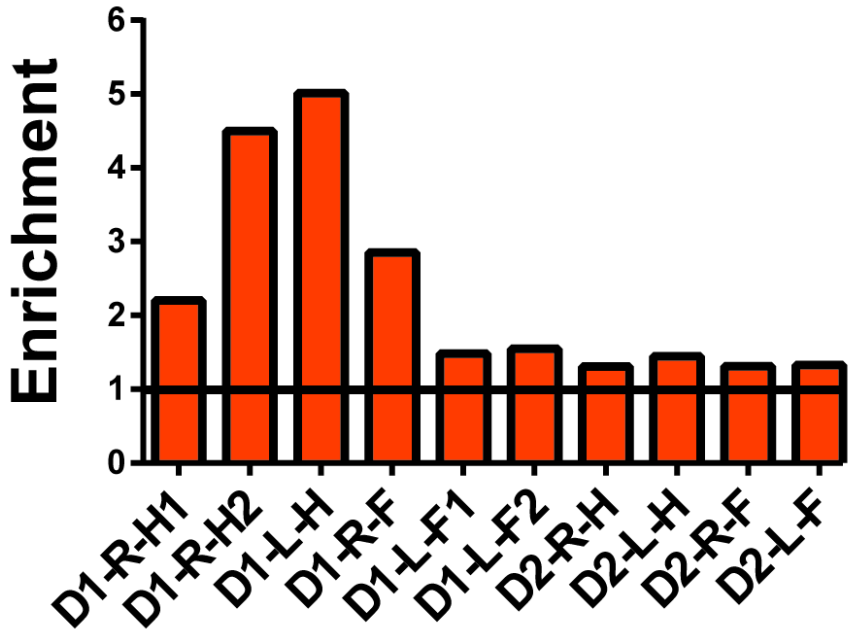
Are there more mutations in a motif than expected from random mutagenesis?

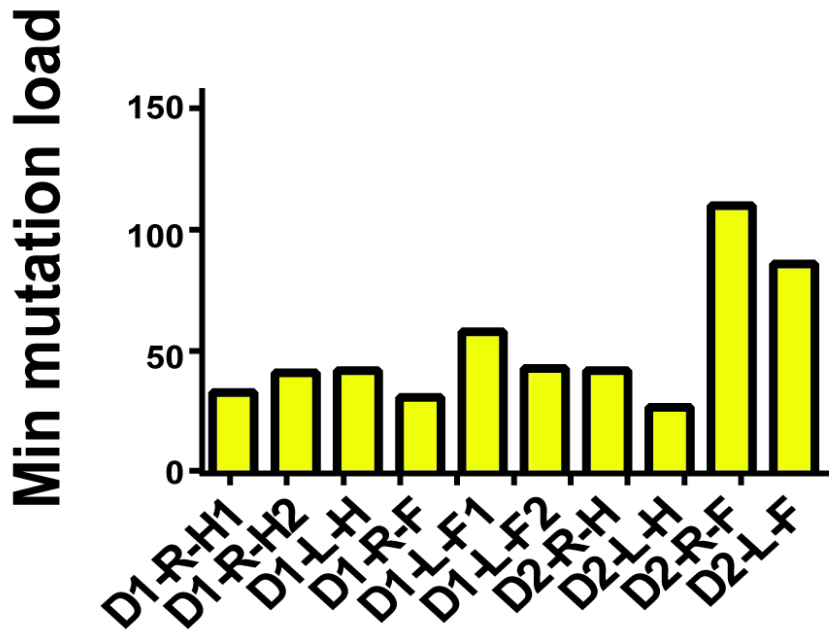
Calculating fold enrichment with mutations in a motif

tgttgcccaggctgggatttCggtgtgagccactgcactt
cagcacttggaacggtcggcCagctccaacggcccagggca
ccatttagctctgcttcctcCtgtcccaatacgttccatc
caggtacctgggaagagactCgtgctgtttcttacataccg
aaggcctatgccagtctaatCatgtgatctctagagttgca

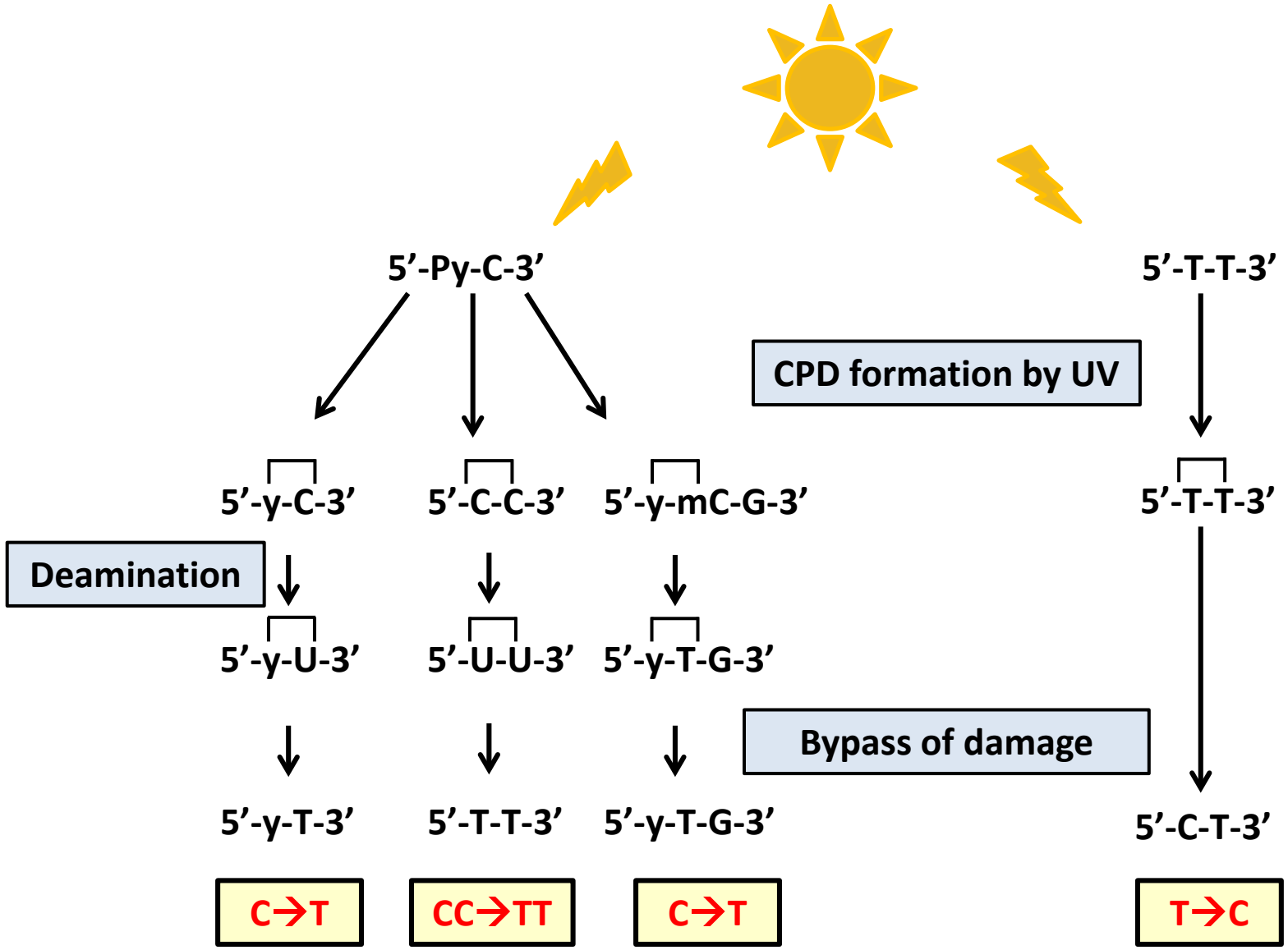
$$Enrichment = \frac{Mutations_{(yCn \rightarrow yTn)} \times Context_{(c)}}{Mutations_{(C \rightarrow T)} \times Context_{(ycn)}}$$

nCg→nTg mutations (clock-like changes associated with aging): Enriched in all samples

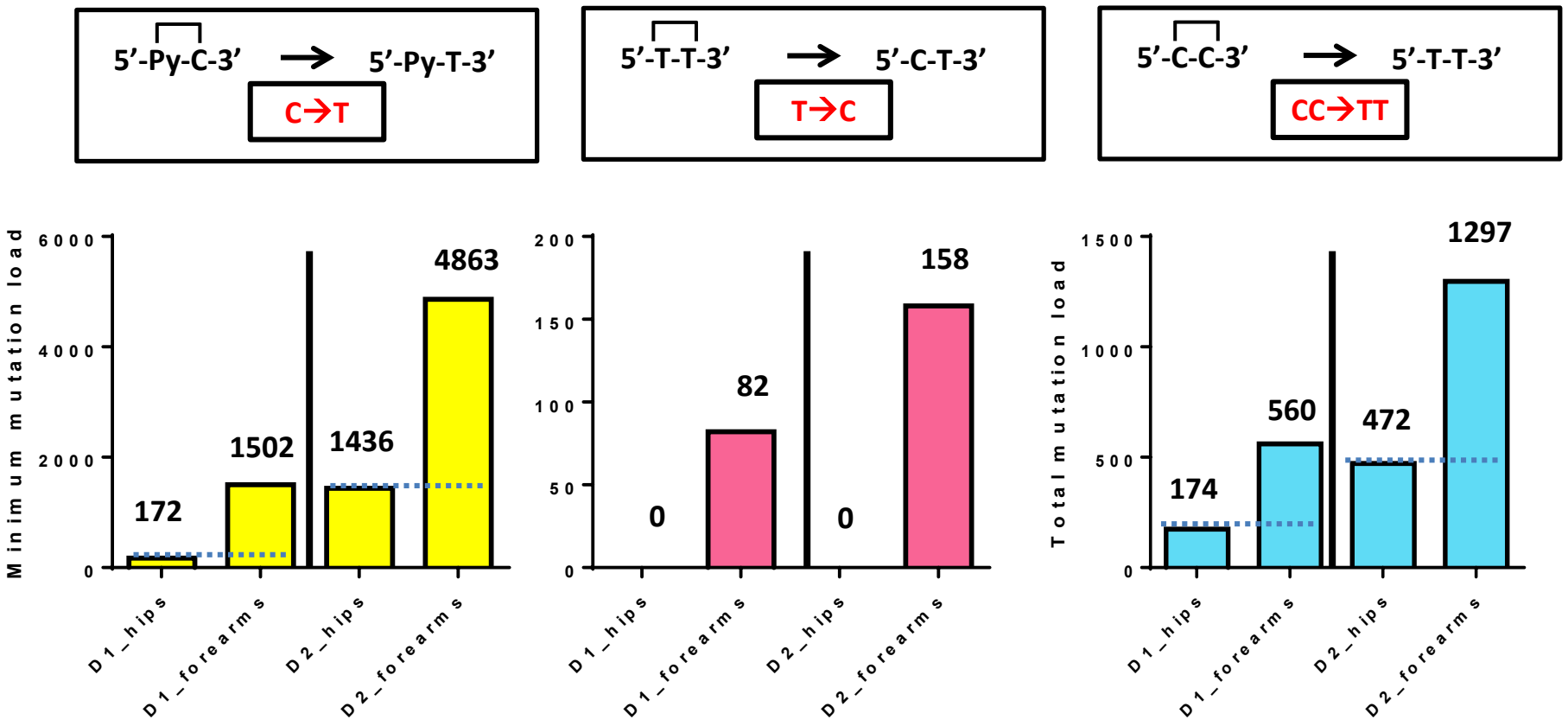




UV induces C→T mutations in a di-pyrimidine context and T→C mutations in a TT context



Forearm samples have increased UV-signature mutations



Model organisms, and biochemistry with Pol η = 3'T mutated in TT dimers

5' tTn 3' → 5'tCn 3'

in vivo

– *E. coli, M13mp7L1 phage*

Table 2. Nucleotide sequence data from SOS-induced cells of SMH10 transfected with vectors carrying either isomer 1 or isomer 2 (Dewar valence isomer) of the pyrimidine-pyrimidone(6-4) adduct located at the T-T target site in the sequence 5′-GCAAGTTGGAG-3′

	Sequences, no. (%)		
		Isomer 2	
	Isomer 1	Sample 1	Sample 2
T-T	16 (9)	17 (31)	49 (58)
A-T	0	5 (9)	4 (5)
C-T	2 (1)	3 (5)	7 (13)
G-T	2 (1)	1 (2)	1 (1)
T-A	0	3 (5)	4 (5)
T-C	158 (85)	23 (42)	11 (13)
T-G	1 (½)	1 (2)	2 (2)
−T*	1 (½)	0	0
Double†	5 (3)	2 (4)	7 (8)
Other‡	0	0	0
Total	185	55	85

LeClerc *et.al* , PNAS (1991)

– *yeast*

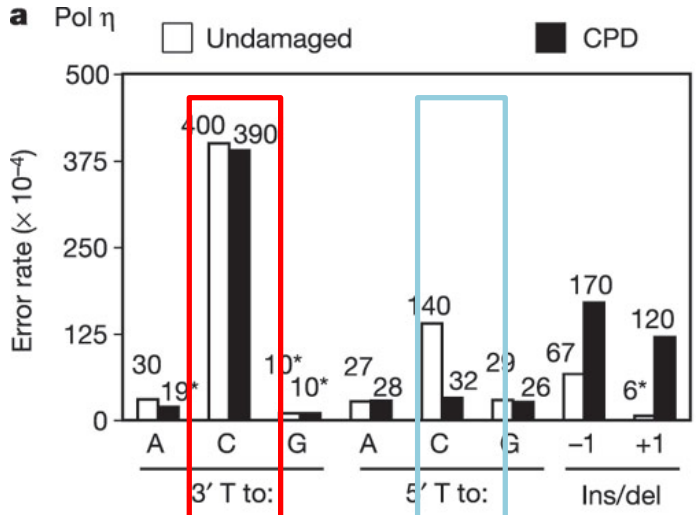
1262–1267 Nucleic Acids Research, 2002, Vol. 30, No. 5

UV-induced T→C transition at a TT photoproduct site is dependent on *Saccharomyces cerevisiae* polymerase η *in vivo*

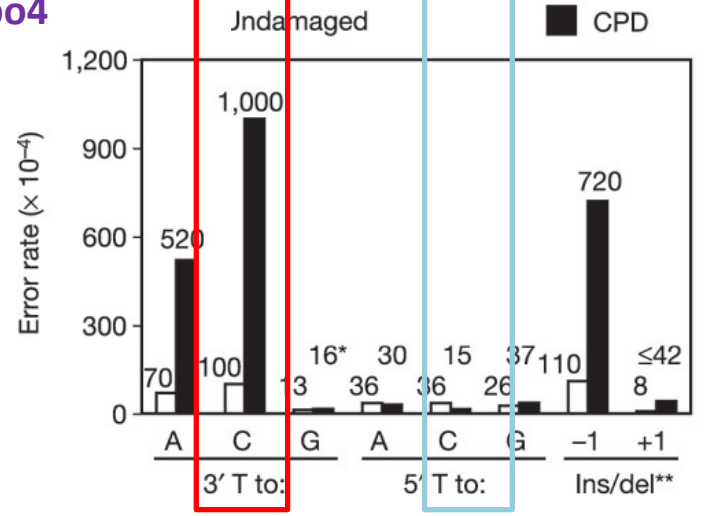
Hong Zhang and Wolfram Siede*

in vitro – CPD bypass

Human Pol η



S. Solfataricus Dpo4



McCulloch SD *et.al* , Nature (2004)

Model organisms, and biochemistry with Pol η = 3'T mutated in TT dimers

5' tTn 3' → 5'tCn 3'

in vivo

– *E. coli, M13mp7L1 phage*

Table 2. Nucleotide sequence data from SOS-induced cells of SMH10 transfected with vectors carrying either isomer 1 or isomer 2 (Dewar valence isomer) of the pyrimidine-pyrimidone(6-4) adduct located at the T-T target site in the sequence 5′-GCAAGTTGGAG-3′

	Sequences, no. (%)		
		Isomer 2	
Other+	0	0	0
Total	185	55	85

LeClerc *et.al* , PNAS (1991)

– *yeast*

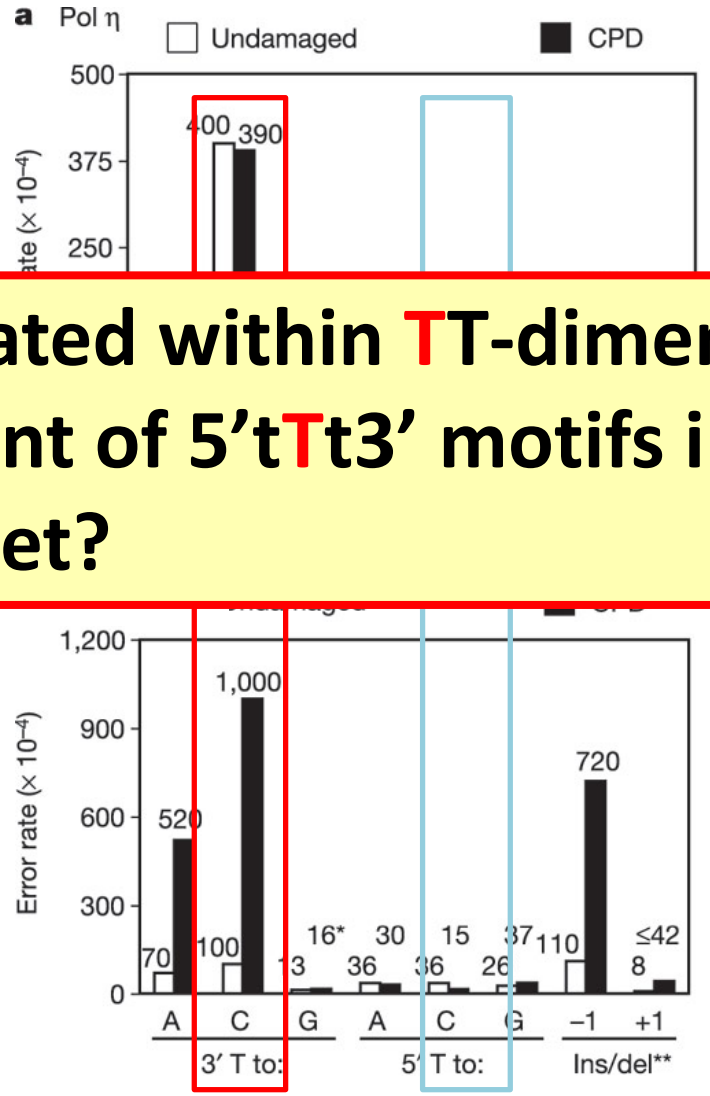
1262–1267 Nucleic Acids Research, 2002, Vol. 30, No. 5

UV-induced T→C transition at a TT photoproduct site is dependent on *Saccharomyces cerevisiae* polymerase η *in vivo*

Hong Zhang and Wolfram Siede*

in vitro – CPD bypass

Human Pol η

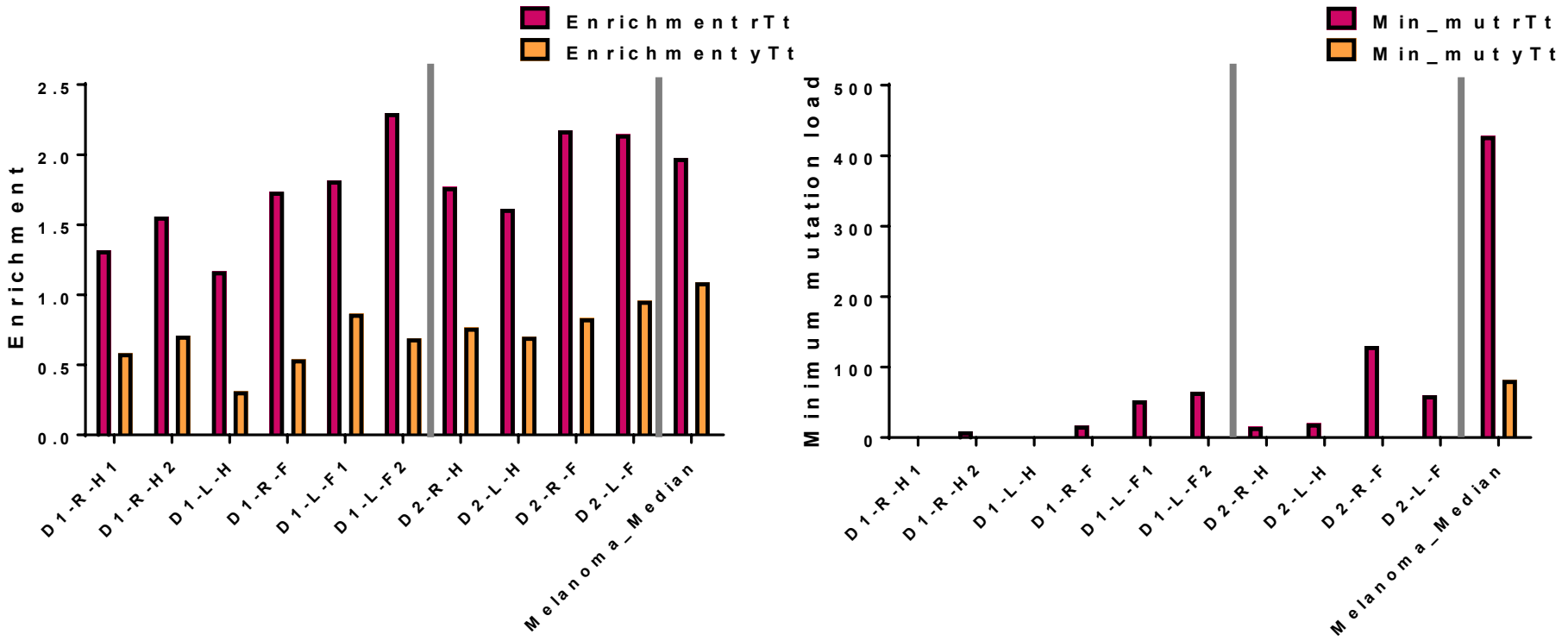


McCulloch SD *et.al* , Nature (2004)

Is the bias towards 5'T mutated within TT-dimers in humans due to enrichment of 5'tTt3' motifs in the dataset?

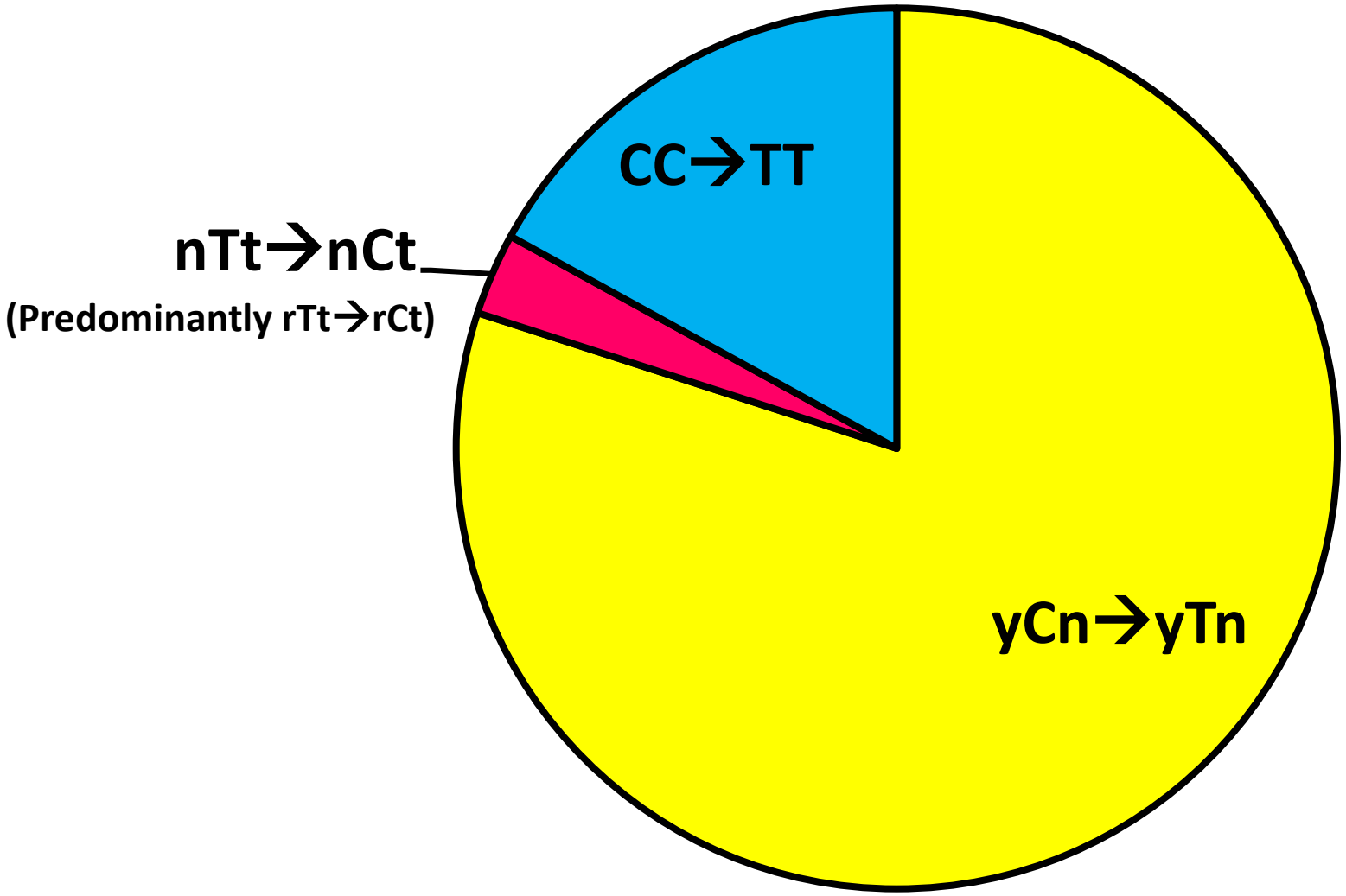
Analysis of +1 nt in nTt →nCt mutation signature reveals higher mutation loads by rTt →rCt

r = a or g, y = t or c

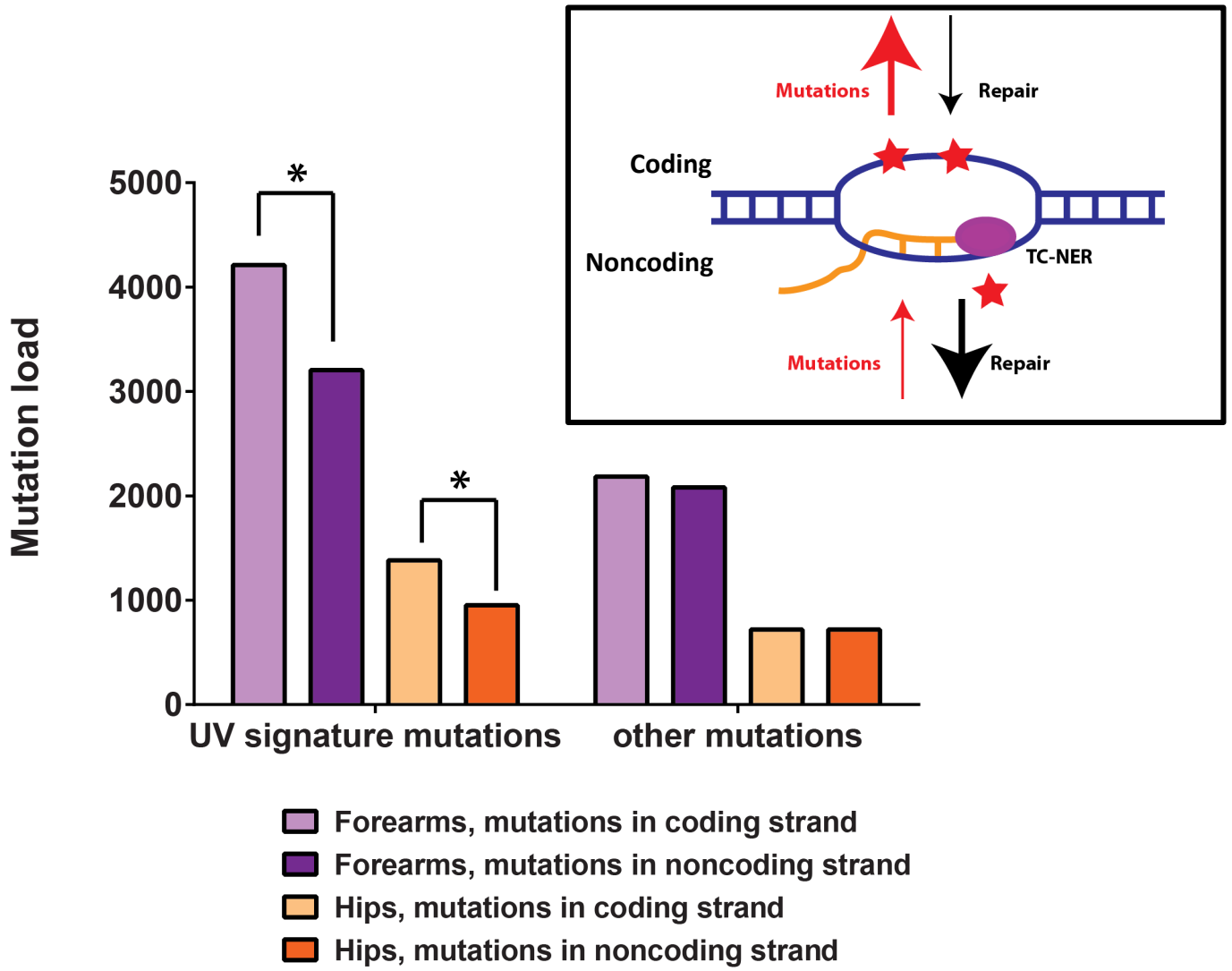


Melanoma data from: Fredriksson *et al.* Nature Genetics 2014; 46:1258-1263

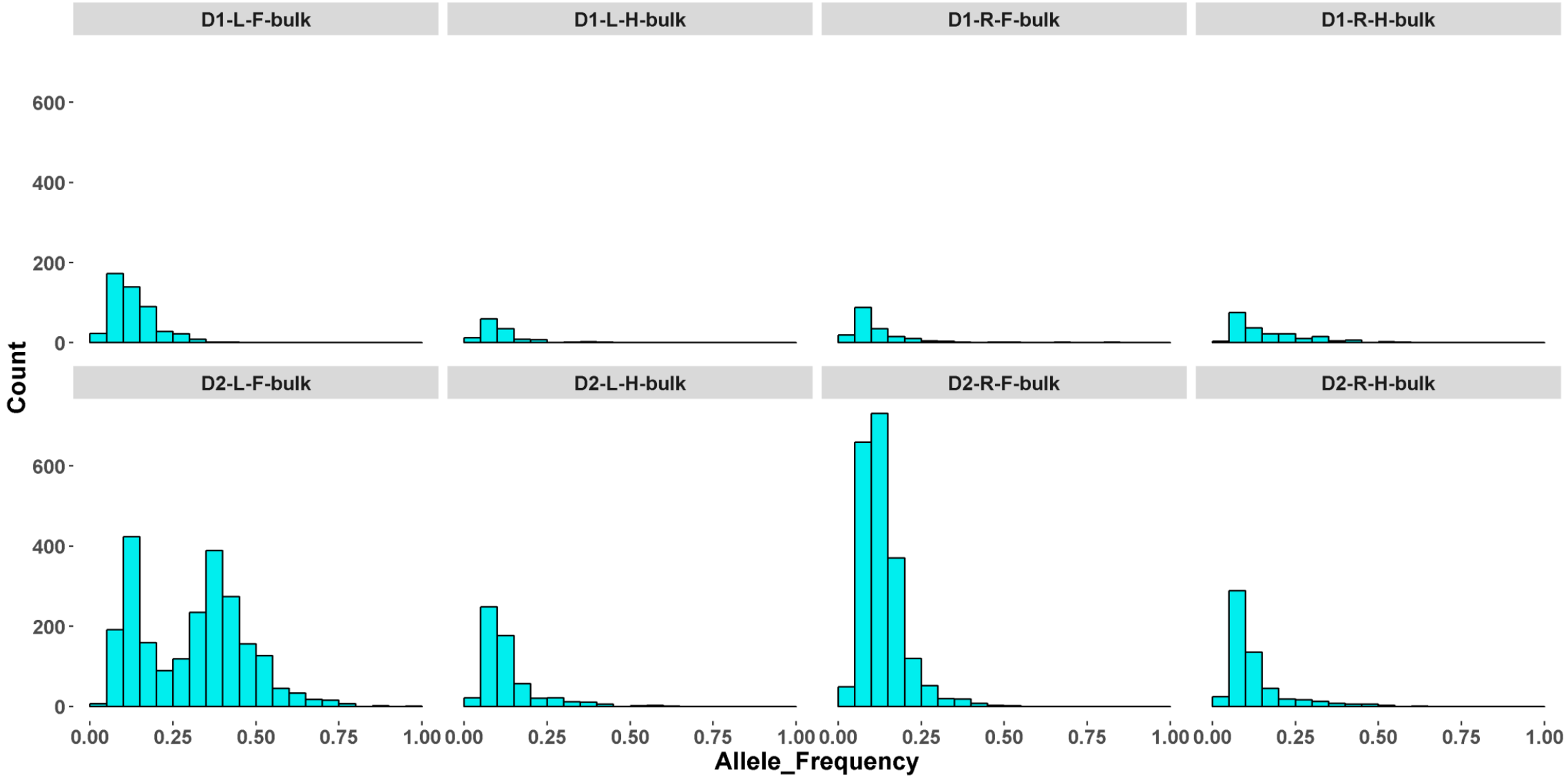
UV mutation signature in human skin fibroblasts



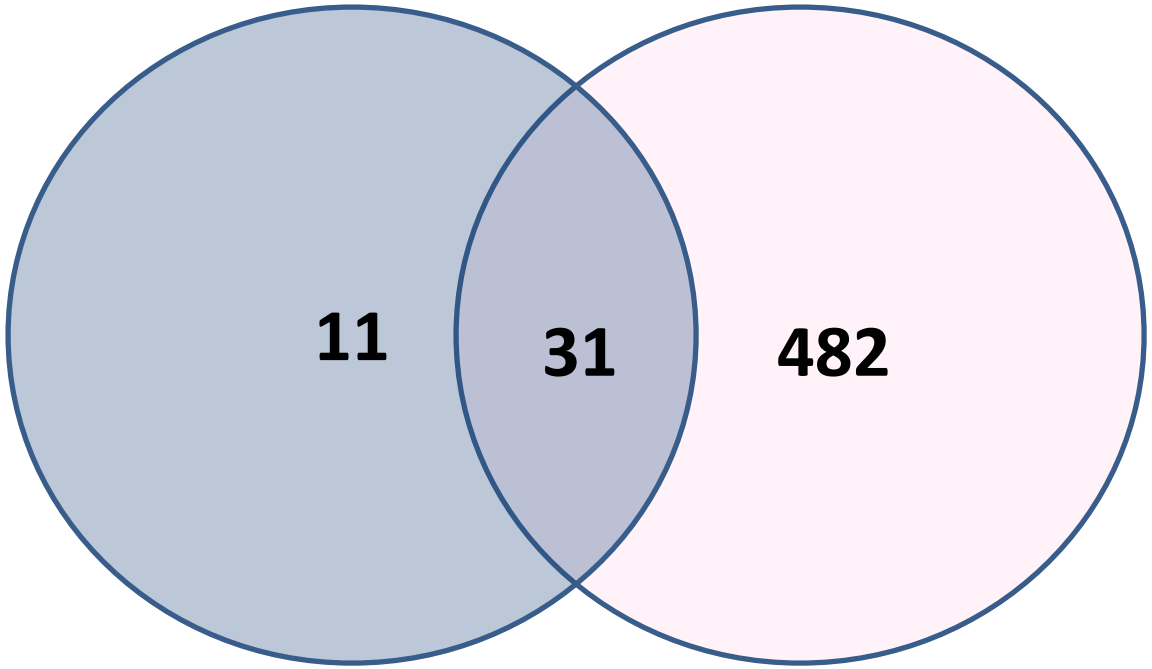
UV-signature mutations demonstrate a bias towards the coding strand



Exomes from bulk fibroblasts – Allele frequencies
Majority of alleles are at ~10-20%



Many of SNVs in clones were identical to those in the bulk samples and at ~10% frequency



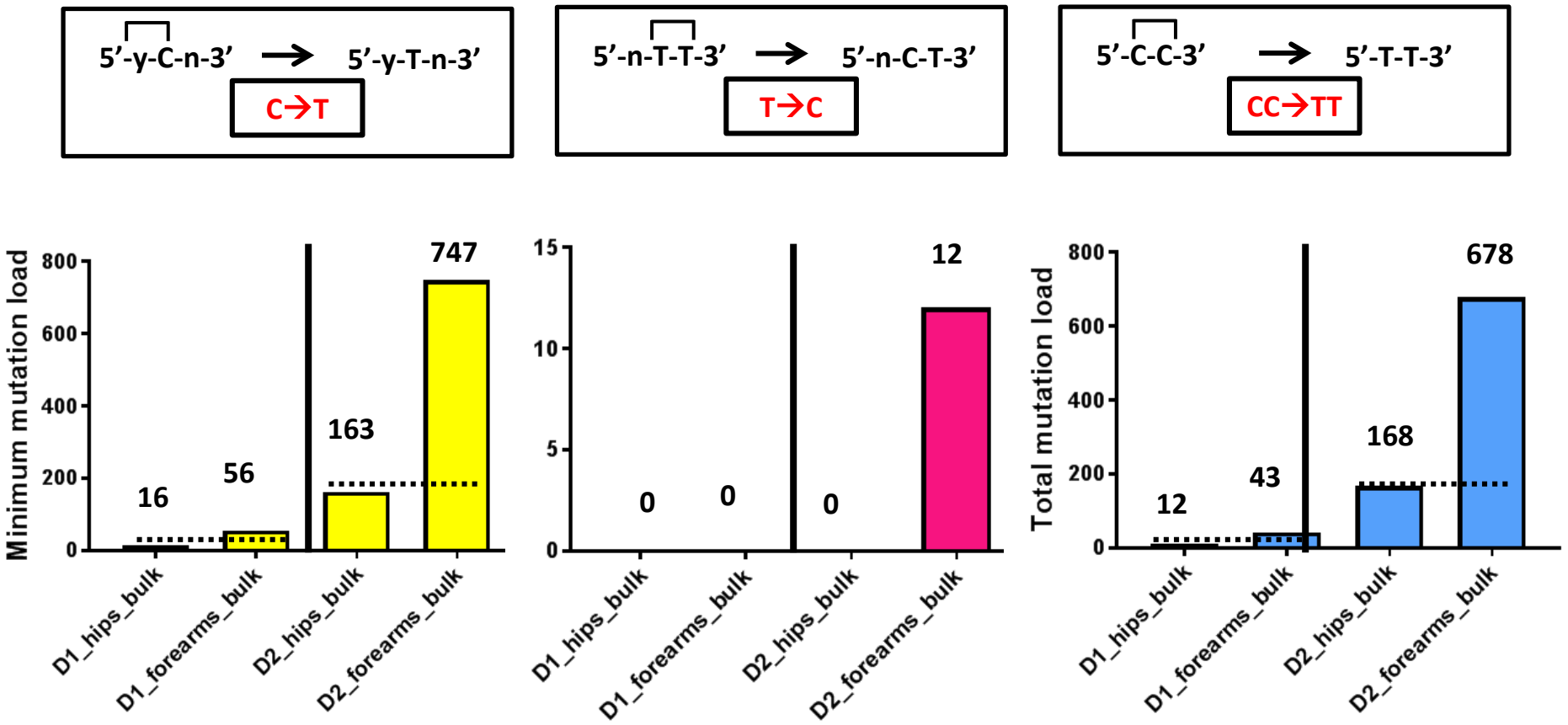
D1-L-F2
Mutations in exons only

D1-L-F2 - bulk

Skin biopsies were made up of at least 10 clonal lineages

Site	Mutations in Bulk	Mutations in clones	Expected in 10 clones
D1-R-H1	197	12	120
D1-L-H	125	10	100
D1-R-F	179	7	70
D1-L-F2	485	41	410
D2-R-H	560	23	230
D2-L-H	576	55	550
D2-R-F	2030	109	1090
D2-L-F	2286	83	830

Forearm samples show increased UV-signature mutations even in exome datasets



Conclusions

Magnitude

- ~1000 – 13000 somatic mutations
- 1-25 gross chromosomal rearrangements

Landscape

- Enriched in late replicating heterochromatin containing regions
- Rearrangements are associated with common fragile sites

Spectrum

- UV mutations – caused by sunlight exposure
- Universal presence of CpG→ TpG signature

Future

1. **Define “normal” and “pathological” levels of somatic genome instability in humans.**

2. **Assess genetic and environmental factors that impact mutagenesis in normal somatic cells as well as in cancers.**

3. **Develop strategies for using somatic mutation data for individuals as a “dosimeter” of lifetime genotoxic exposures.**

Acknowledgements

Mechanisms of Genome Dynamics Group
Dmitry Gordenin
Joan Sterling
Cynthia Sakofsky

Steven Roberts
(NIEHS → Washington State University, Pullman)
Kin Chan
(NIEHS → University of Ottawa)

Integrative Bioinformatics Support Group
David Fargo
Les Klimczak
Sara Grimm
Shuangshuang Dai

Molecular & Genetic Epidemiology Group
Jack Taylor

UNC - Chapel Hill
William Kaufmann
Jayne Boyer

High Throughput Sequencing Facility (UNC – Chapel Hill)
Piotr Mieczkowski
Ewa Malc

Harvard Medical School (Boston)
Peter Park
Eunjung (Alice) Lee
Isidro Cortes-Ciriano

NIEHS Clinical Research Unit
Shepherd Schurman

Chromatin and Gene Expression Group
Lois Annab

Anonymous donors

NIH National Institute of Environmental Health Sciences
Your Environment. Your Health.