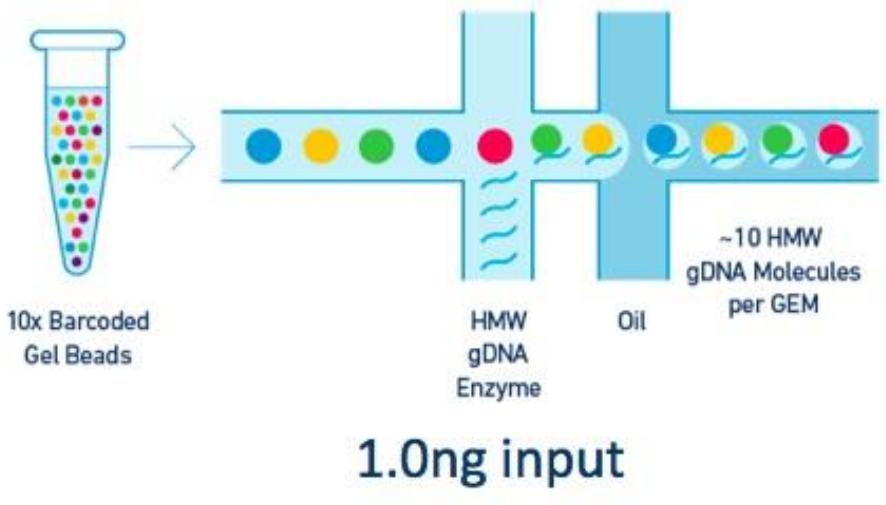
CANCER GENOMICS AND MOSAIC DISEASE

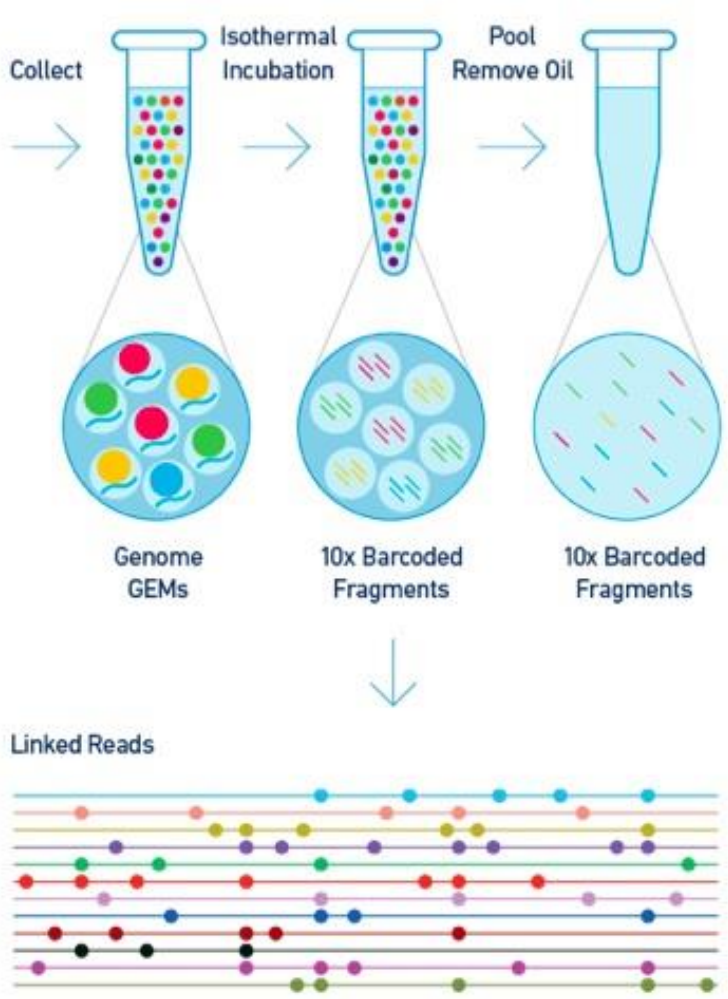
Computational Genomics: Applied Comparative Genomics
Charlotte Darby 4/24/18

0. Linked Read Sequencing

10X Genomics

Linked reads are short paired-end reads with additional barcodes attached in library preparation





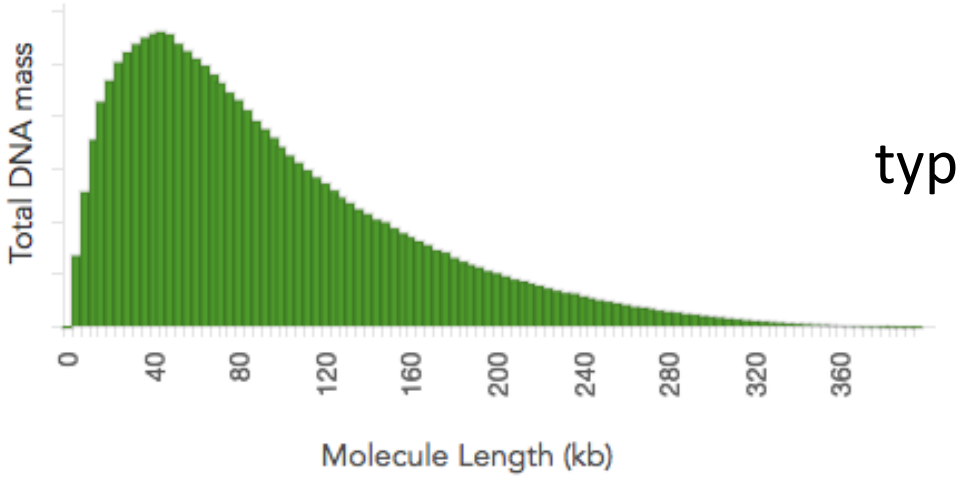
- High-quality and contiguous haplotype assemblies
- Diploid *de novo* genome assembly
- Structural variant breakpoints
- Single-cell applications

https://wheaton5.github.io/projects/tenx

genome

fragments

Molecule Length	μ 85,667 bp
DNA in Molecules >20kb	91.6%
DNA in Molecules >100kb	38.6%
Corrected Estimated of DNA Loaded	1.23 ng



typical length is 10's to 100's of kb
(mean 50-100kb)

http://loupe.10xgenomics.com/loupe/ NA12878 Germline Whole Genome

genome

fragments

barcodes

short-read sequencing

Ordinary short reads, 100’s of bp insert size

genome

fragments

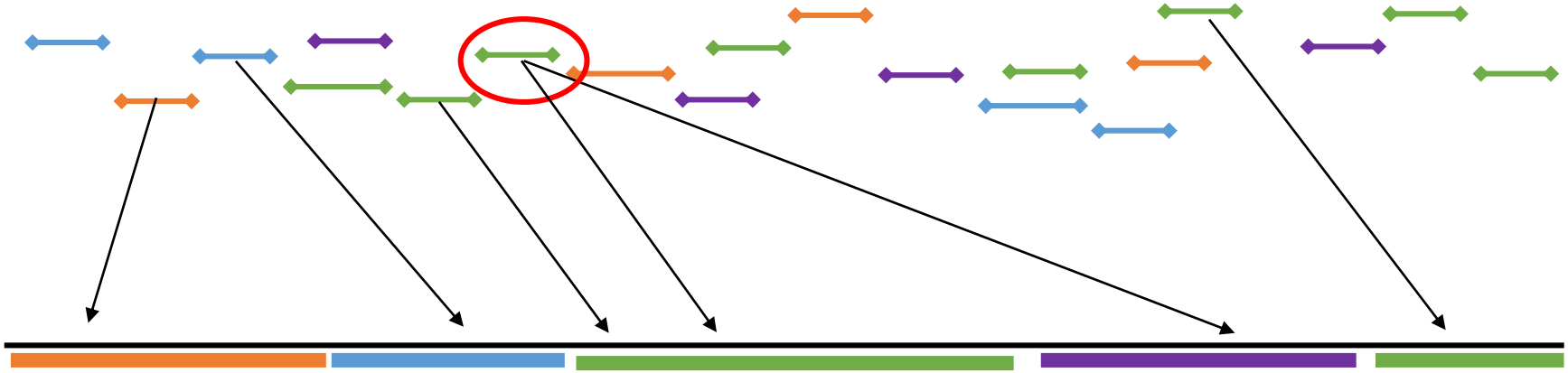
barcodes

10X Genomics Chromium

short-read sequencing

Illumina instrument

alignment

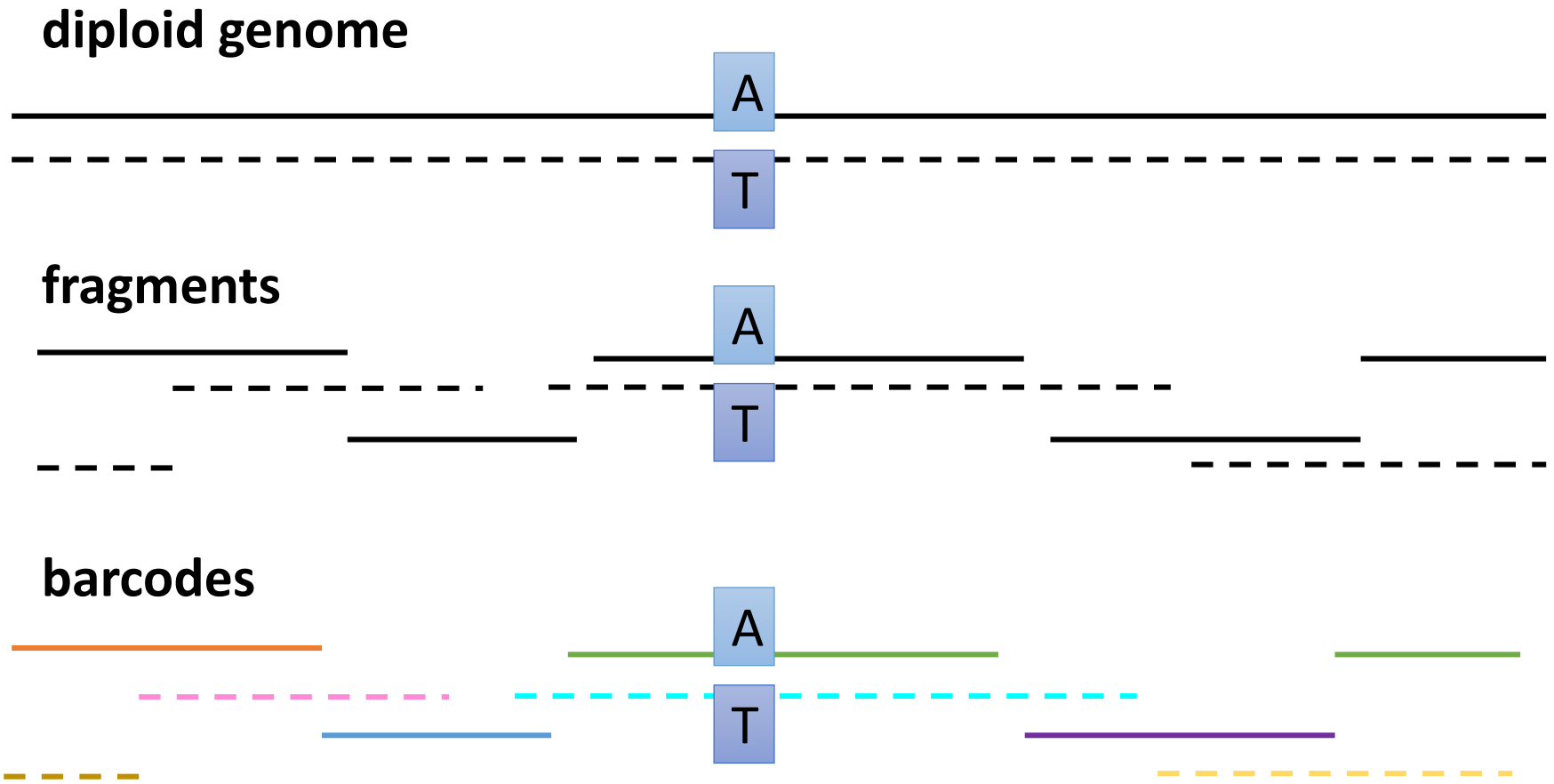


reference genome

Resolve multi-mapping reads
using barcodes

variant calling

haplotype assembly



- Each linked read has ~0.1X “coverage”
- Likely that least one constituent read overlaps a heterozygous site, at least in long fragments

Resources

Biotechnology, alignment (Lariat) and haplotype phasing (Longranger)
Haplotyping germline and cancer genomes with high-throughput linked-read sequencing.
Zheng et al., Nat Biotech (2016)

Diploid assembly (Supernova)
Direct determination of diploid genome sequences.
Weisenfeld et al., Genome Res (2017)

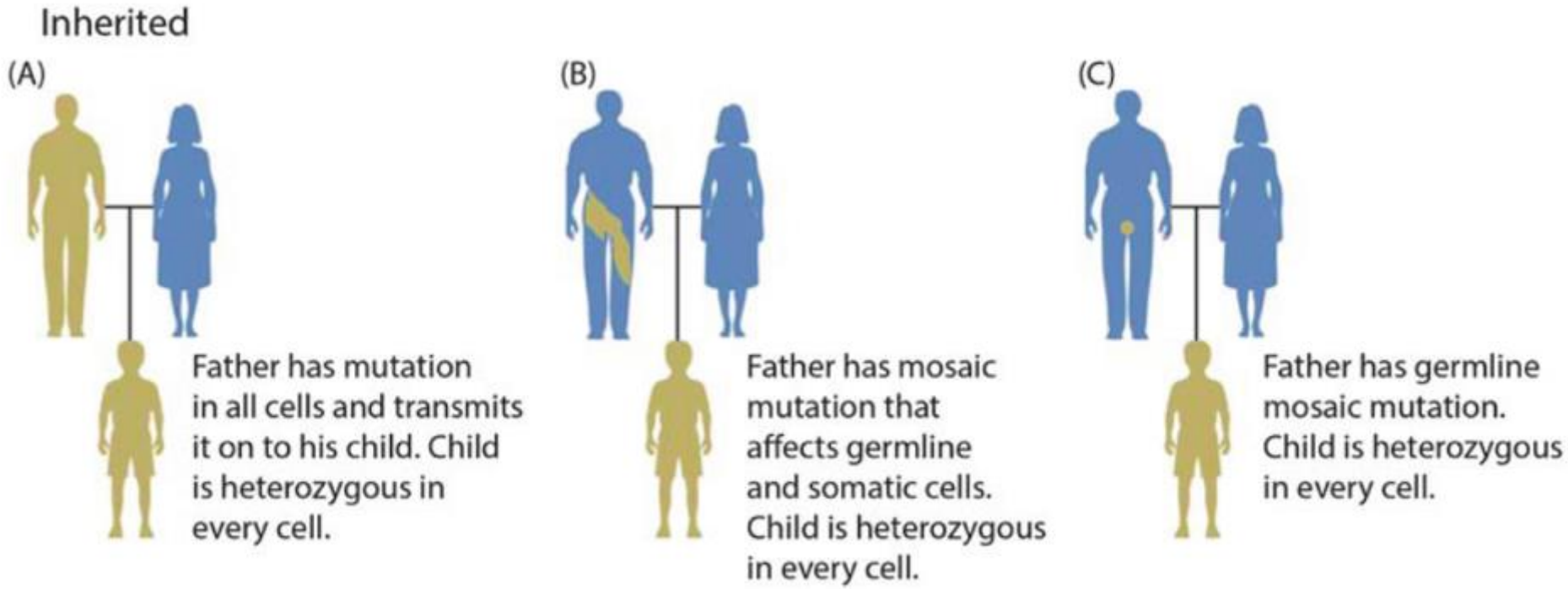
Simulation
LRSim: A Linked-Reads Simulator Generating Insights for Better Genome Partitioning
Luo et al., Comp Struct Biotech J (2018)

SV calling
Genome-wide reconstruction of complex structural variants using read clouds
Spies et al., Nat Methods (2017)

https://www.10xgenomics.com/resources/publications/

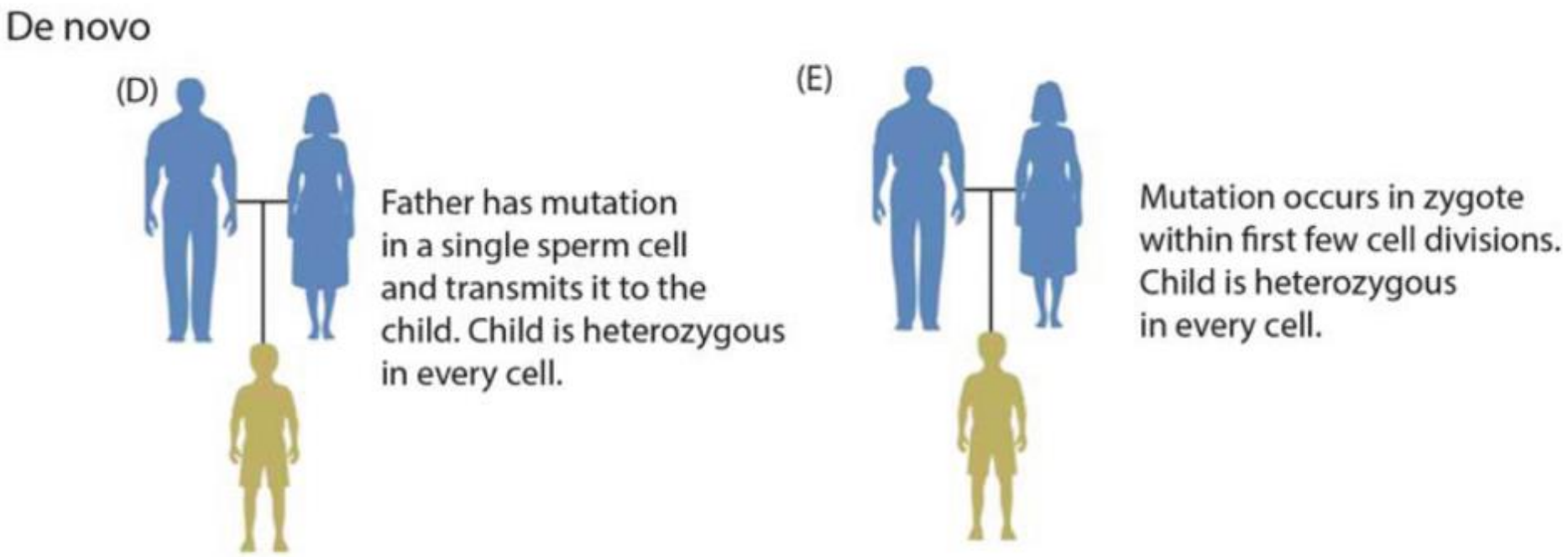
1. What is a somatic / mosaic mutation?

Inherited variants are present in at least some of the cells of a parent



Freed, Stevens and Pevsner, Genes (2014)

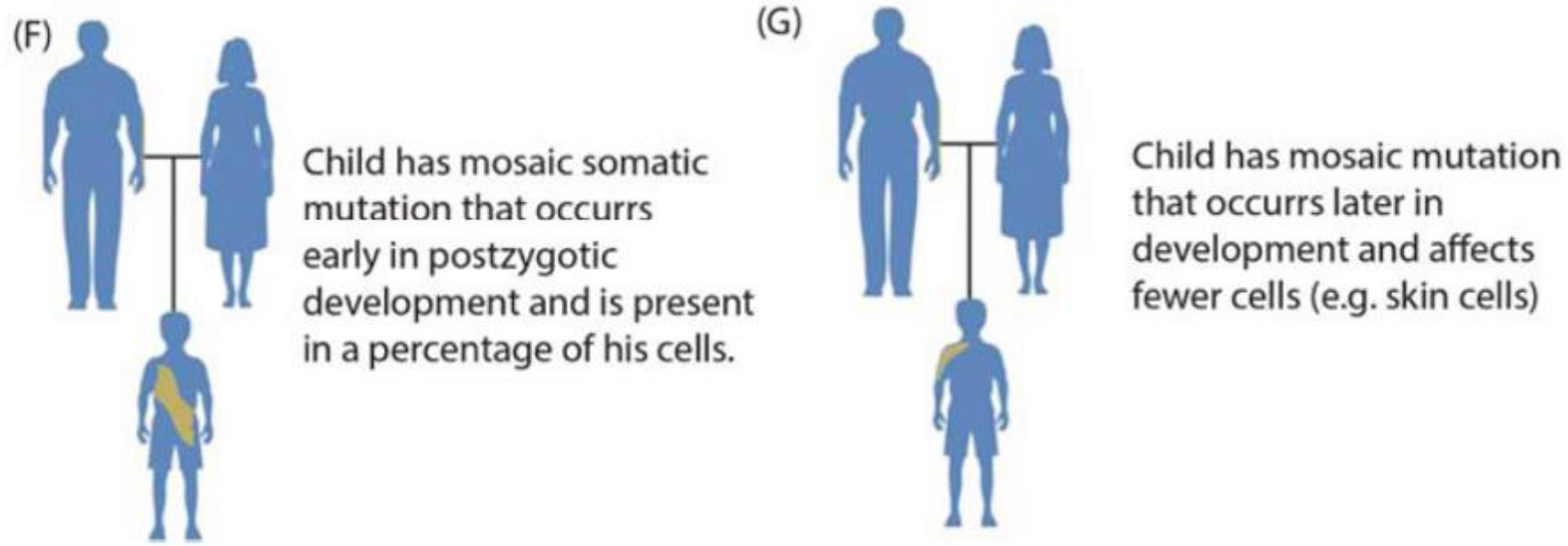
De novo mutations are present in all of the cells of the child



Freed, Stevens and Pevsner, Genes (2014)

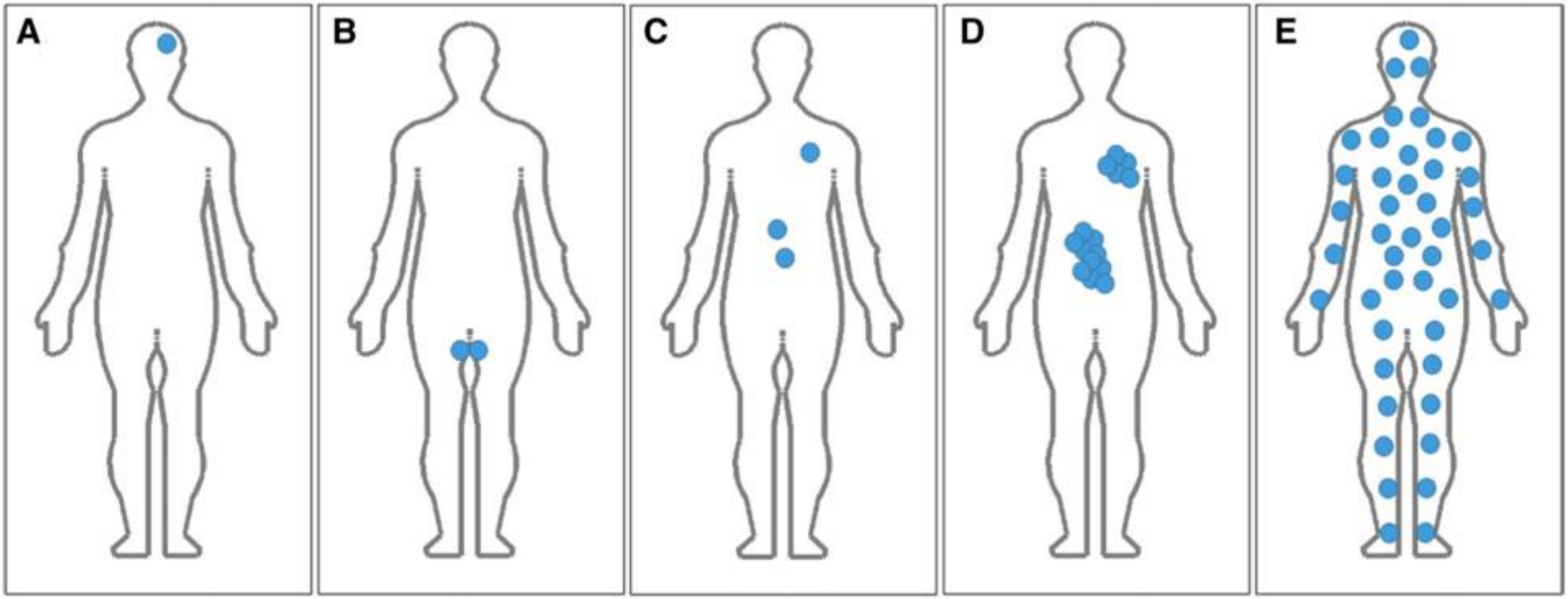
Somatic mutations are present in none of the cells of the parents and only some of the cells of the child

Somatic



Freed, Stevens and Pevsner, Genes (2014)

2. How do somatic mutations manifest?



somatic mosaicism

gonadal mosaicism

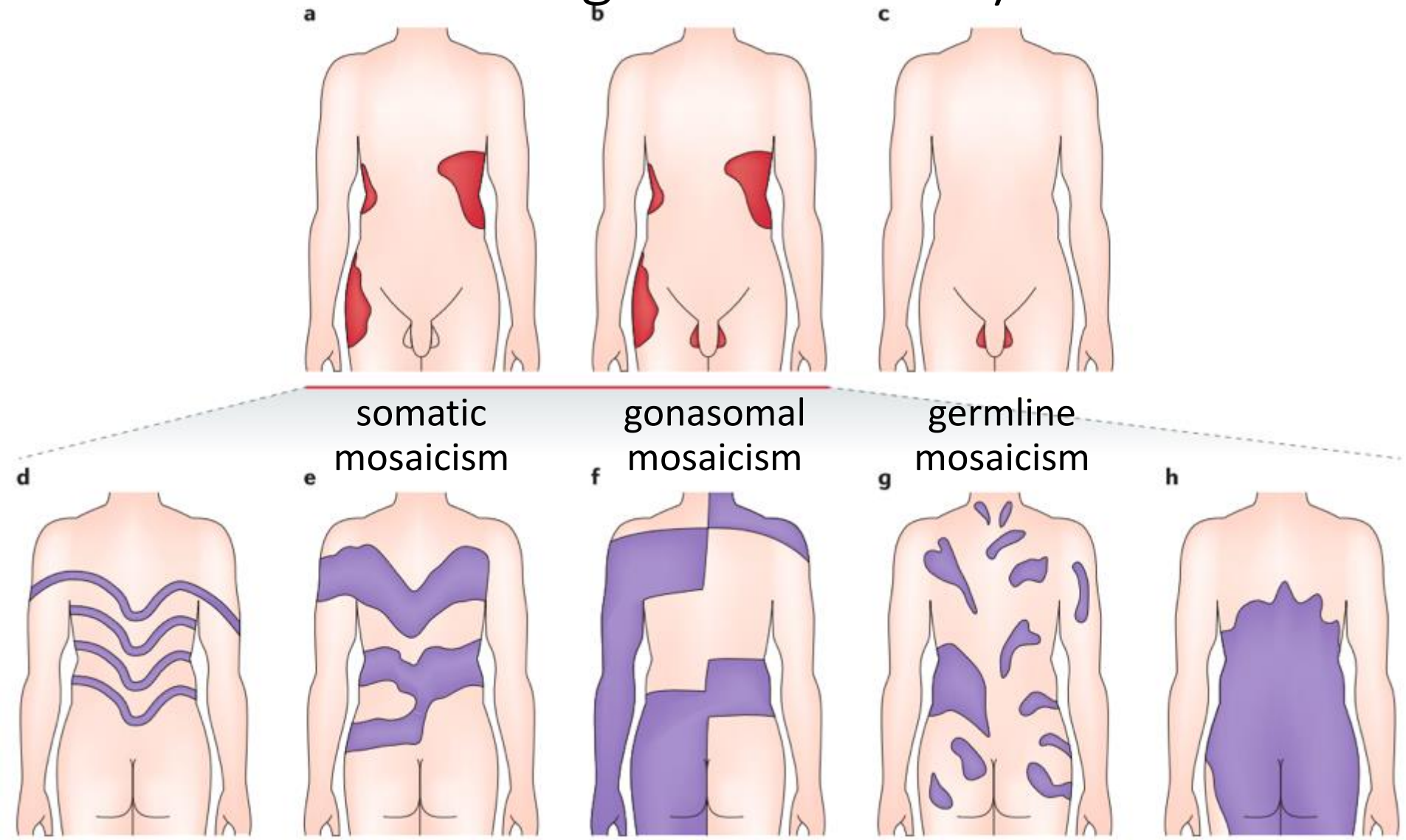
low-grade somatic mosaicism (endoderm)

medium-grade somatic mosaicism (endoderm)

constitutional somatic mosaicism

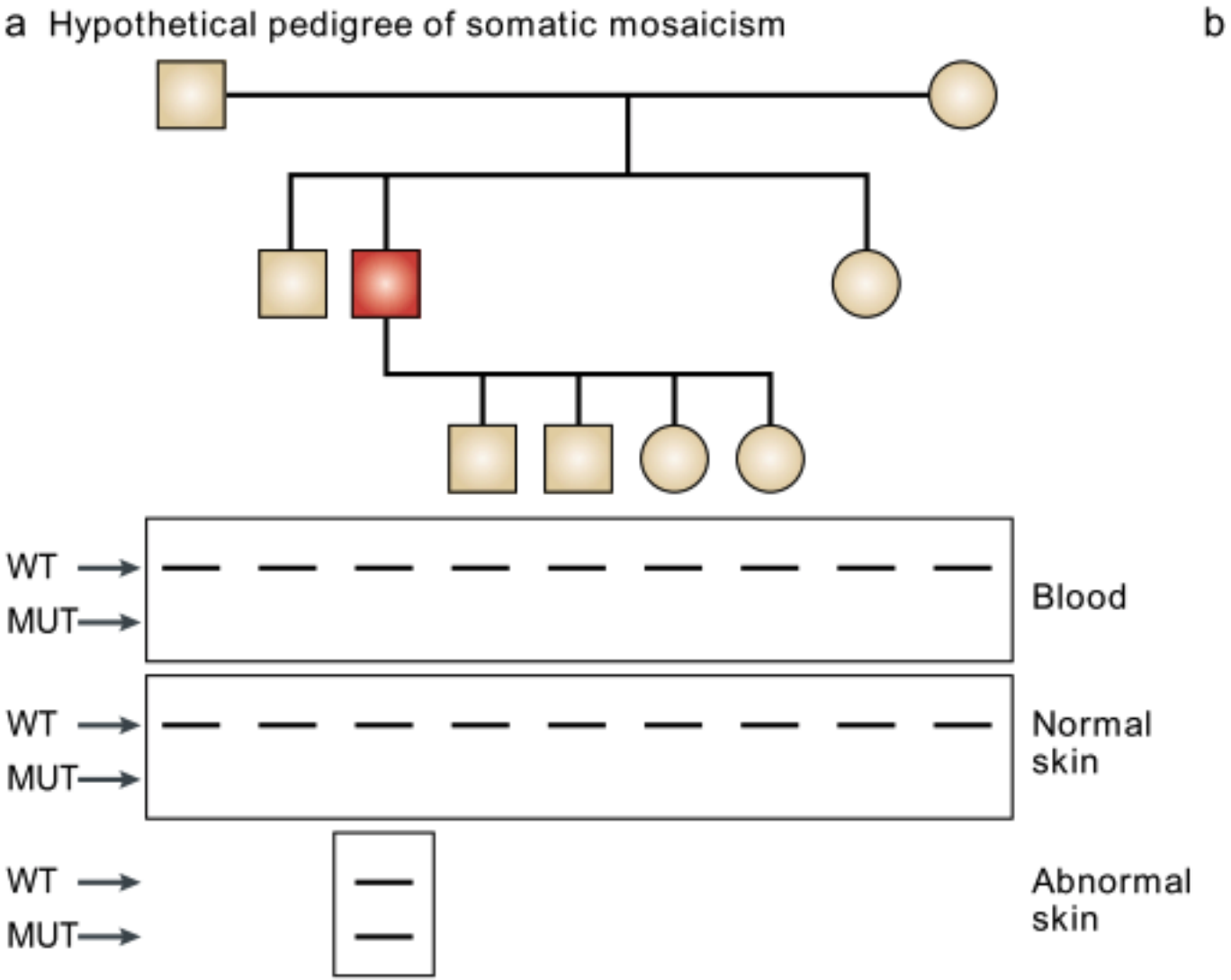
Gajecka, Mol Genet Genomics (2015)

Somatic mutations can have different distributions throughout the body



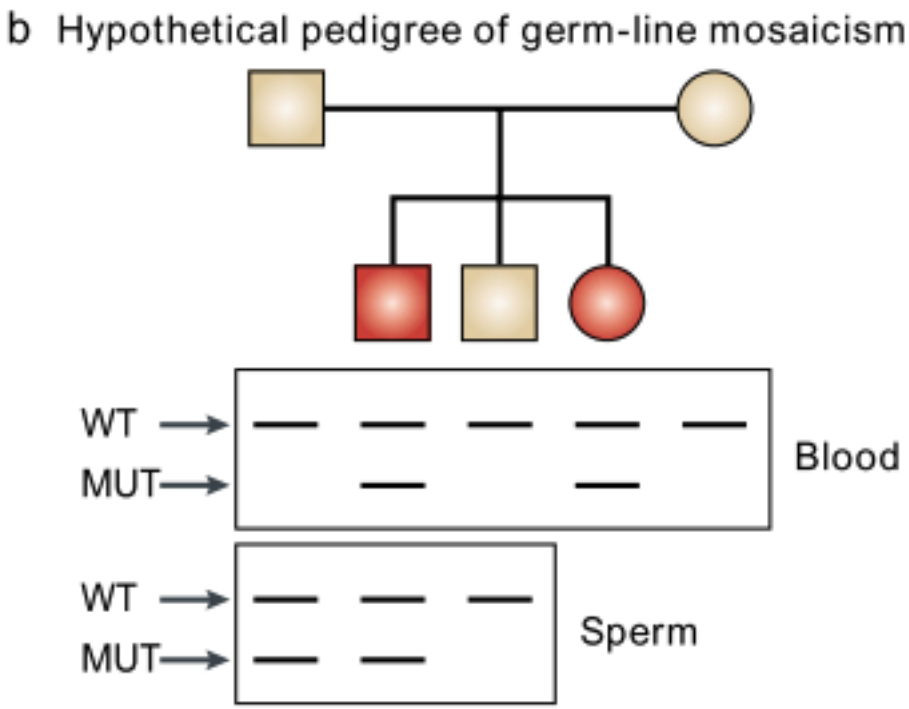
Biesecker and Spinner, Nat Rev Genet (2013)

An individual with somatic mosaicism in skin does not pass the trait to his offspring



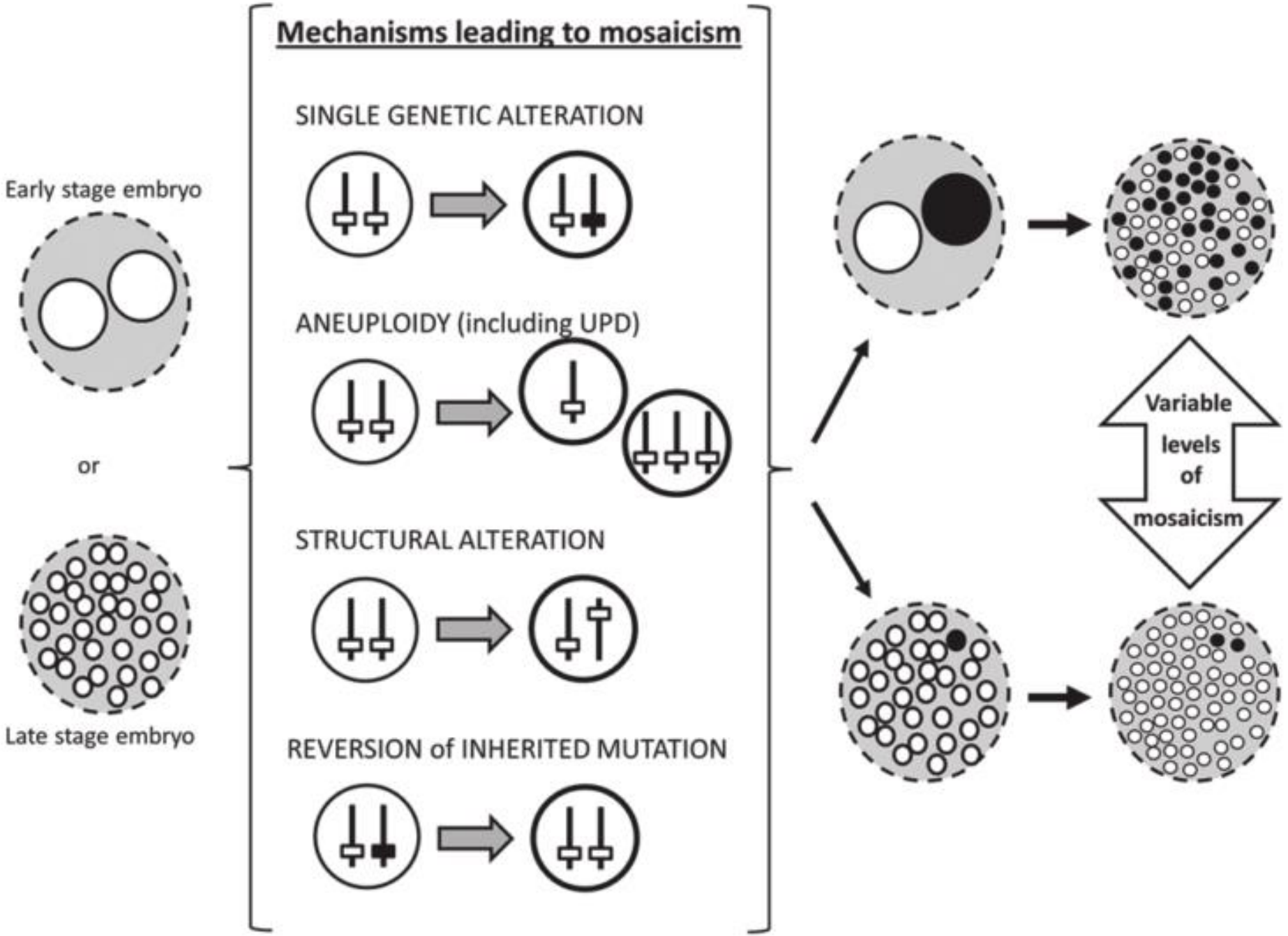
Biesecker and Spinner, Nat Rev Genet (2013)

A father with gonasomal mosaicism is affected and has some affected children



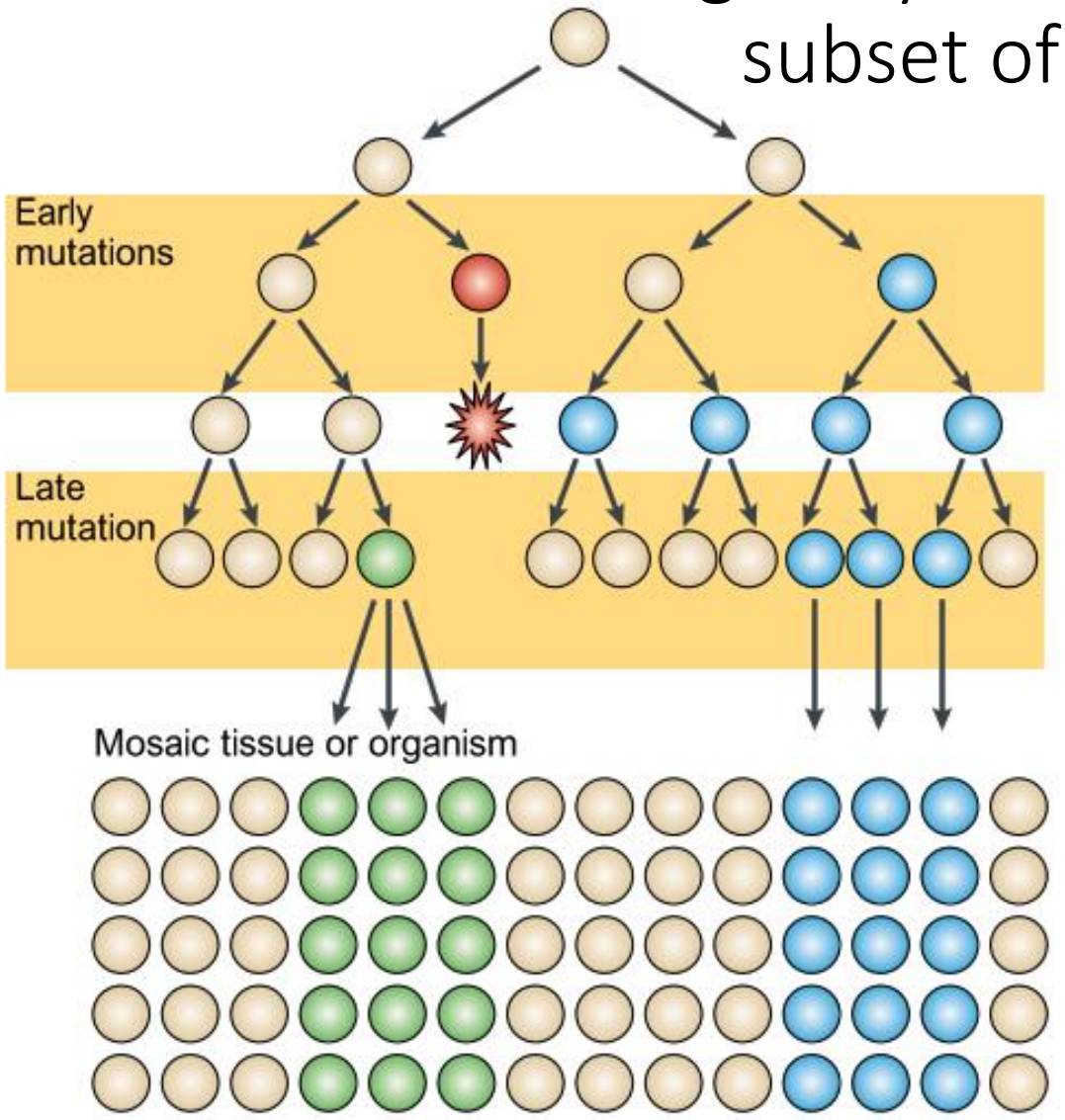
Biesecker and Spinner, Nat Rev Genet (2013)

3. What are possible mechanisms of somatic mutation?

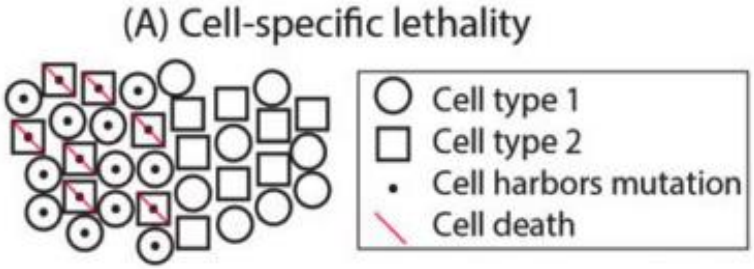


Cohen et al., Clin Genet (2015)

Mutations occurring early in development are in a subset of cells of the adult



"Model for how new mutations are selected in the generation of mosaic phenotypes. Mosaic populations arise if new mutations that occur early in development (blue circles) do not compromise growth or, alternatively, new mutations that occur relatively late in development (green circles) confer a proliferative advantage. Mutations that compromise cell growth early on most likely will not contribute to the mosaic phenotype (red circle)."



Youssoufian and Pyeritz, Nat Rev Genet (2002)

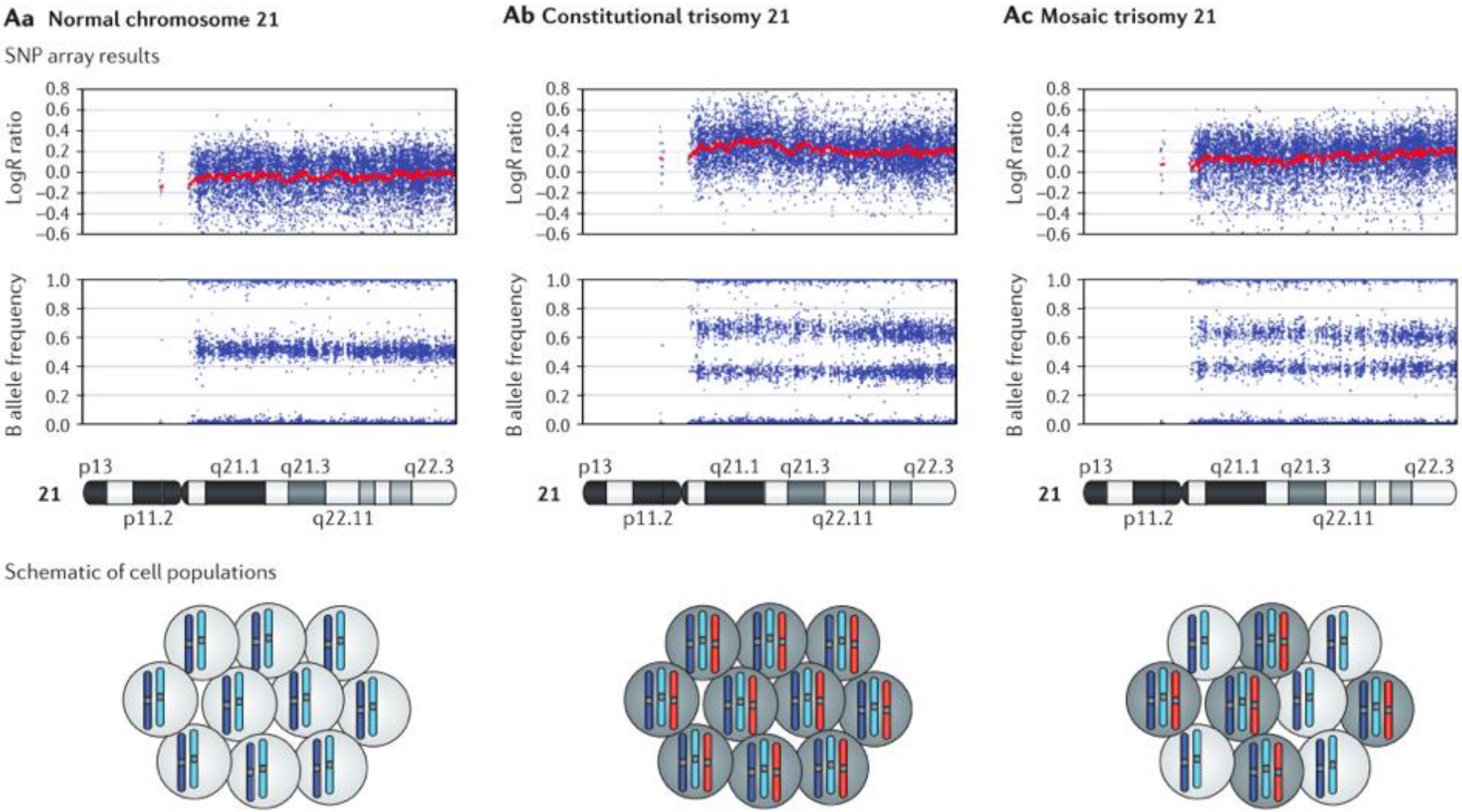
Freed, Stevens and Pevsner, Genes (2014)

4. Chromosome-scale mutations

Chromosome-scale mutations

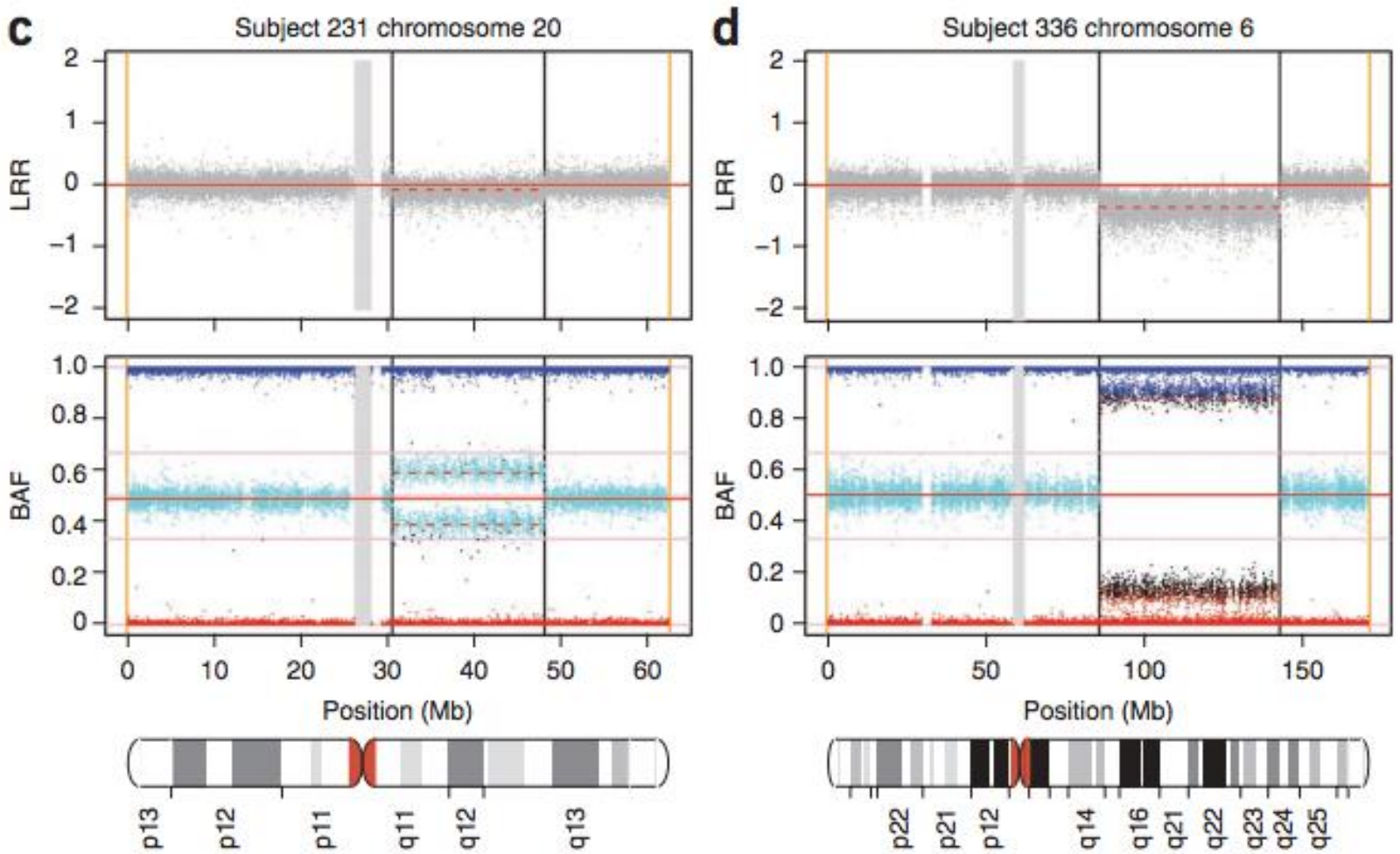
- Uniparental disomy
  - Both copies of one chromosome are from the same parent
- Aneuploidy
  - Not diploid
- Loss of heterozygosity
- Rearrangement

SNP array detection of aneuploidy



Biesecker and Spinner, Nat Rev Genet (2013)

SNP array detection of large-scale mosaicism



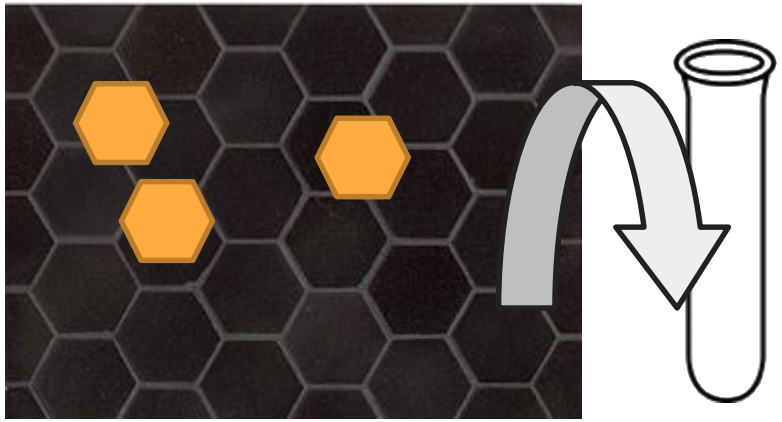
(c) A mosaic deletion at 20q is indicated by a narrow split in the intermediate BAF band along with a small decrease in LRR. A nonmosaic heterozygous deletion would have no intermediate BAF bands and a larger decrease in LRR. (d) A mosaic deletion at 6q is indicated by a wide split in the intermediate BAF band along with a large decrease in LRR. The mosaic in d has a greater proportion of cells containing the deletion than the one in c.

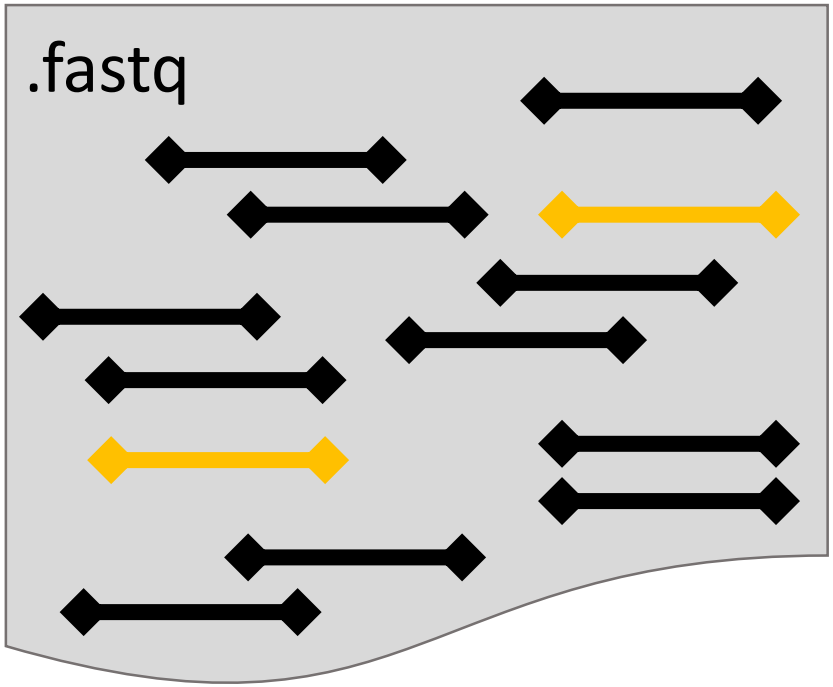
Laurie et al., Nat Genet (2012)

5. Smaller-scale mutations

Connections to my own research

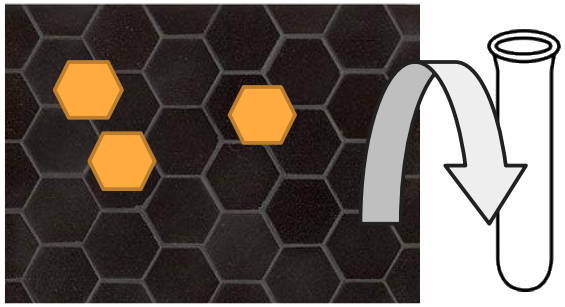
Somatic mutations have low allele frequency in the sequencing reads





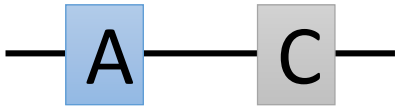
Allele(s) on the haplotype	Fraction of total reads
A — C	40%
T — C	40%
A — C	10%
T — G	10%

somatic variant



All cells you sequence have a heterozygous (A/T) variant

20% of cells you sequence have a somatic (C/T) variant where the somatic variant allele G always co-occurs with the heterozygous variant allele T

Allele(s) on the haplotype	Fraction of total reads
	40%
	40%
	10%
	10%

Allele fraction at somatic mutation site

G	C
10%	90%

Allele fraction partitioned by haplotype

	G	C
A	0%	50%
T	10%	40%

Haplotype-discordant reads

haplotype \ allele	0	1
0	H_{00}	H_{01}
1	H_{10}	H_{11}

$$\frac{\min \begin{cases} H_{00} + H_{11} \\ H_{01} + H_{10} \\ H_{00} + H_{10} \\ H_{01} + H_{11} \end{cases}}{H_{00} + H_{01} + H_{10} + H_{11}}$$

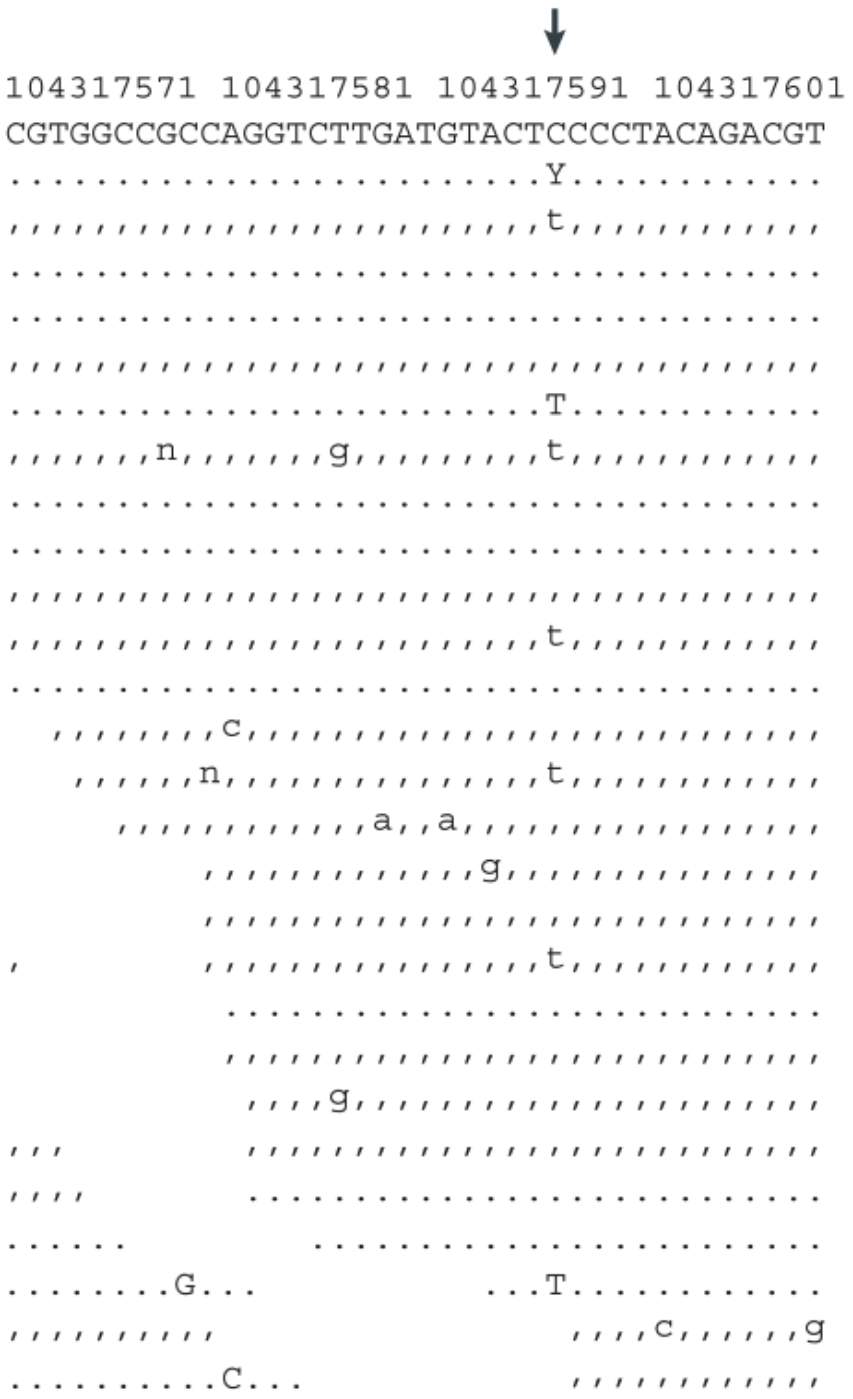
A high fraction of haplotype-discordant reads indicates that the alleles on each haplotype are not uniform

BUT … Uniformity is expected at heterozygous/homozygous sites

NGS-based detection of sites with low variant allele frequency (VAF)

- Bayesian statistical modeling
- Classification approaches
- Error filters

- VAF closer to 0% and 50% are harder to detect
- Higher sequencing depth gives more detection power
- Paired “normal” sample or trio information can also be used



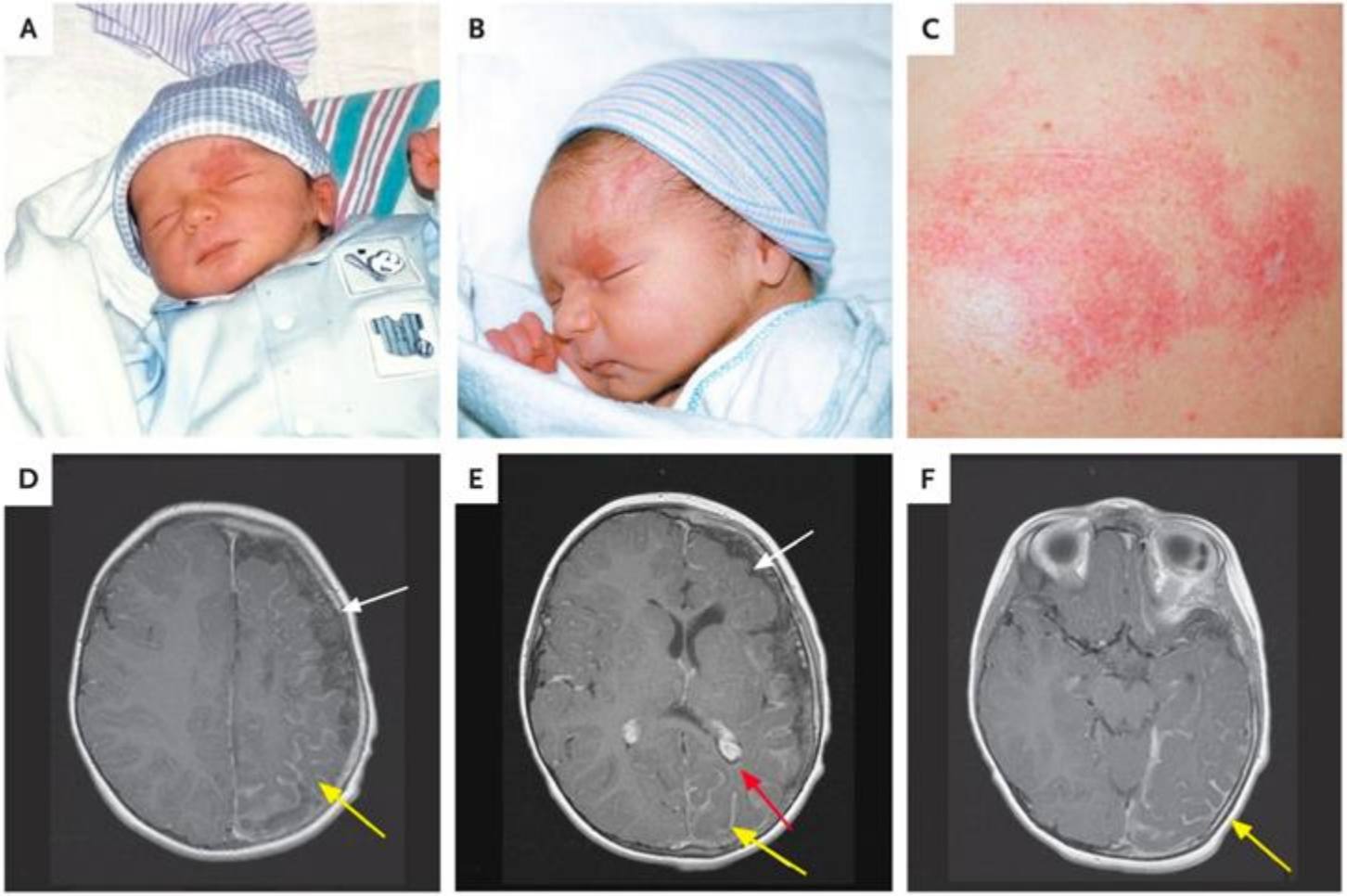
6. Mosaic disease

“Mendelian” diseases may also be caused by somatic mutation

- Disease has been studied and specific causal gene(s) are known
- Phenotype may be intermediate or patchy
- Specific gene can be studied closely for mosaicism

Note: A certain mutation (chromosomal or smaller-scale) might manifest only as a somatic mutation because if every cell had this mutation, the embryo would not survive
(e.g. McCune-Albright Syndrome)

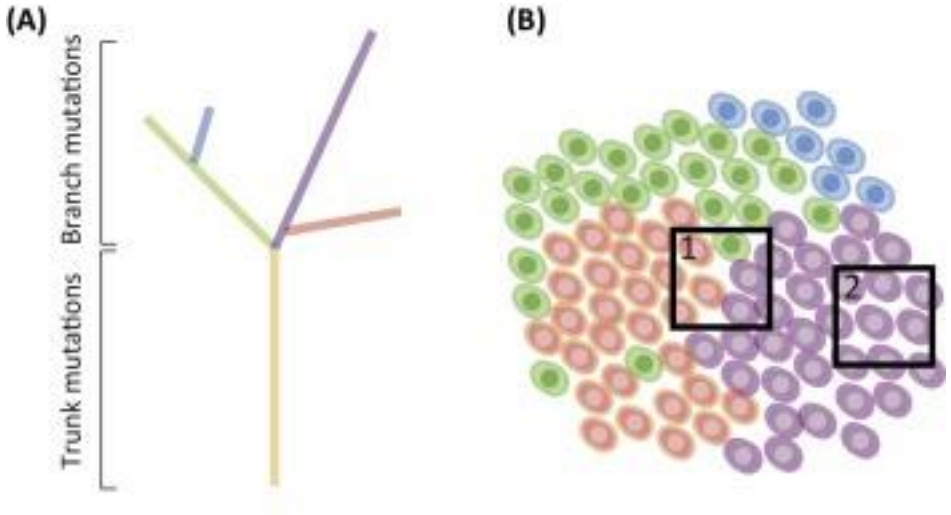
Sturge-Weber syndrome



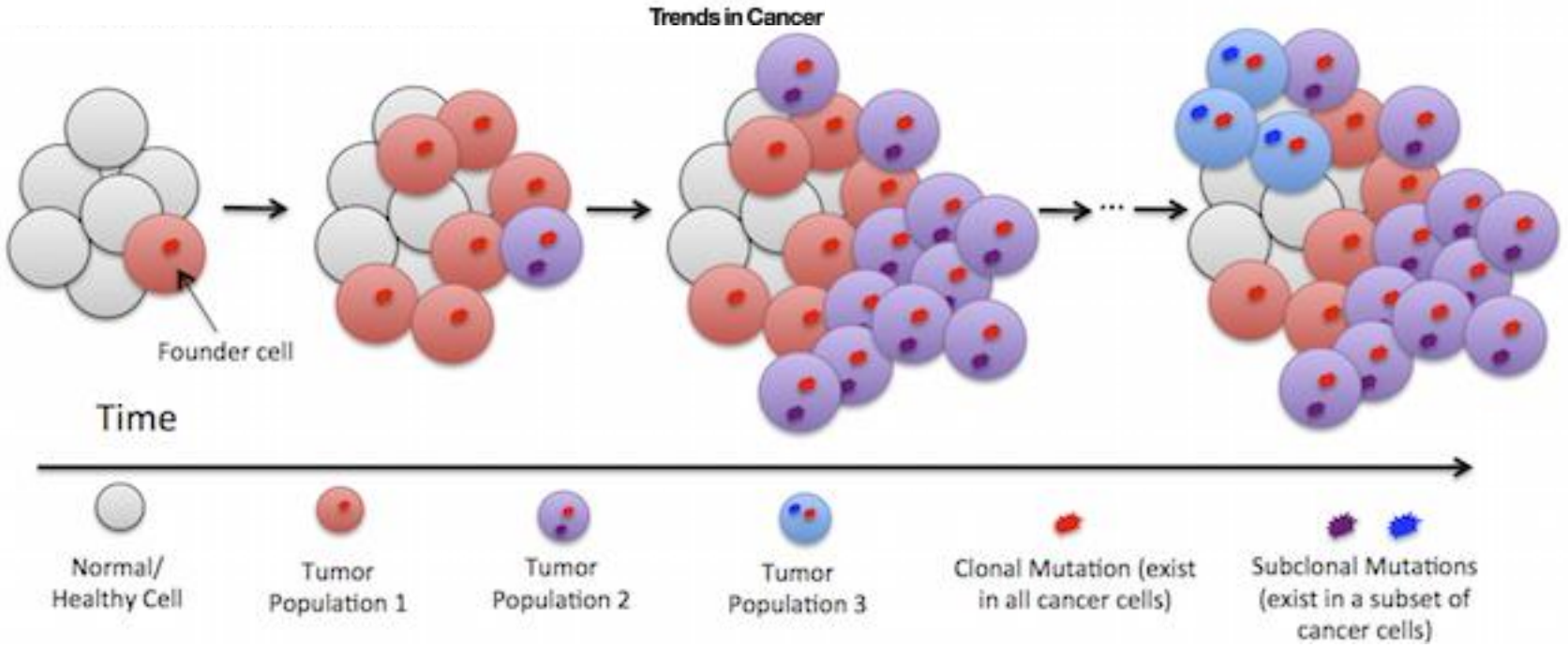
The Sturge–Weber syndrome and port-wine stains are caused by a somatic activating mutation in *GNAQ*.

Shirley … Pevsner et al., NEJM (2013)

Burrell and Swanton, Trends in Cancer (2016)



Tumor cells are different from normal cells and from each other



http://blog.creative-bioarray.com/introduction-on-the-heterogeneity-of-tumor-cell/

7. Are mosaic mutations normal?

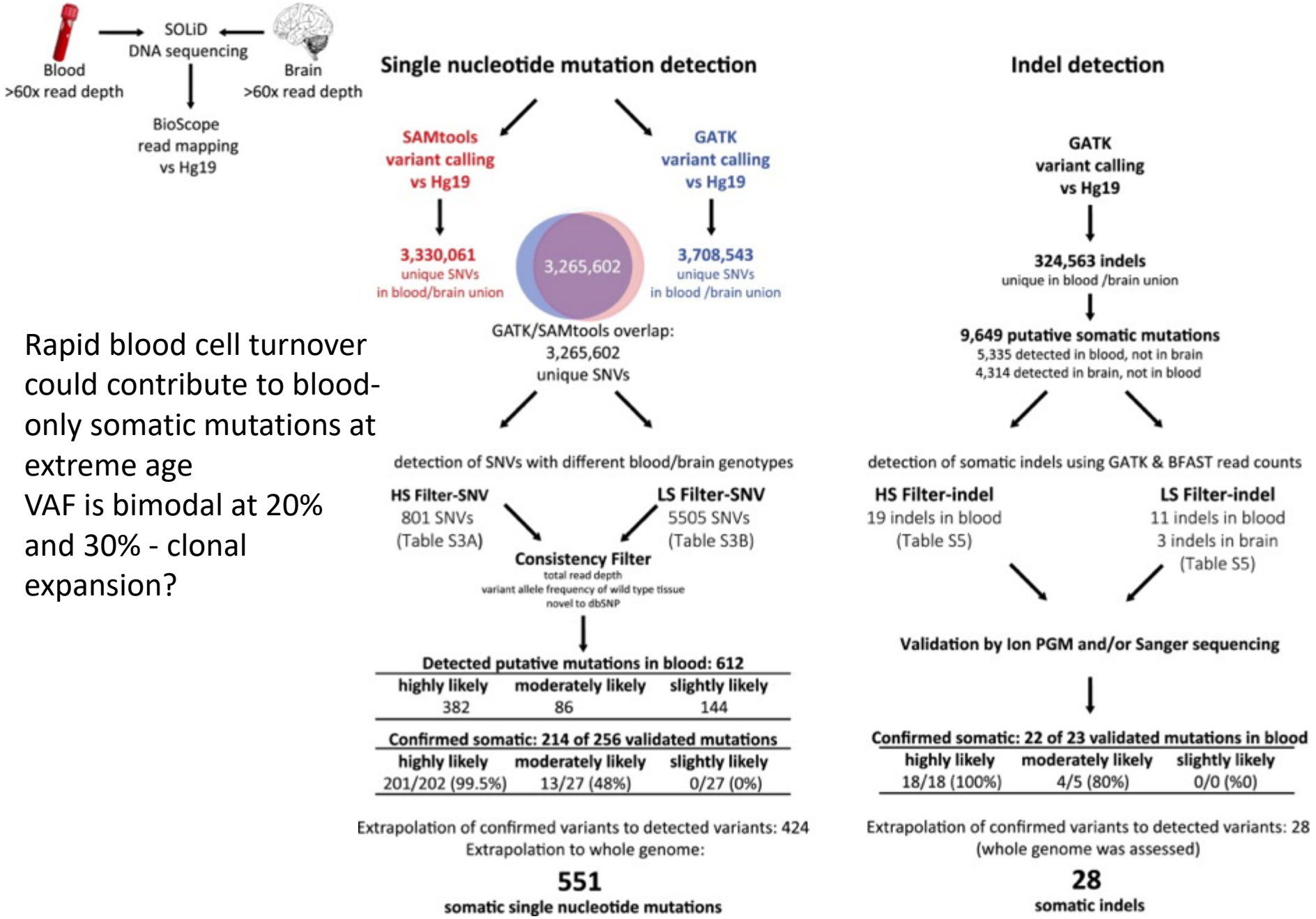
Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis

Henne Holstege,[1,10] Wayne Pfeiffer,[2] Daoud Sie,[3] Marc Hulsman,[4] Thomas J. Nicholas,[5] Clarence C. Lee,[6] Tristen Ross,[6] Jue Lin,[7] Mark A. Miller,[2] Bauke Ylstra,[3] Hanne Meijers-Heijboer,[1] Martijn H. Brugman,[8] Frank J.T. Staal,[8] Gert Holstege,[9] Marcel J.T. Reinders,[4] Timothy T. Harkins,[6] Samuel Levy,[5] and Erik A. Sistermans[1]

[1]Department of Clinical Genetics, VU University Medical Center, 1007 MB Amsterdam, The Netherlands; [2]San Diego Supercomputer Center, UCSD, La Jolla, California 92093, USA; [3]Department of Pathology, VU University Medical Center, 1007 MB Amsterdam, The Netherlands; [4]Delft Bioinformatics Laboratory, Delft University of Technology, 2628 CD Delft, The Netherlands; [5]Department of Molecular and Experimental Medicine, Scripps Translational Science Institute, San Diego, California 92037, USA; [6]Advanced Applications Group, Life Technologies, Beverly, Massachusetts 01915, USA; [7]Department of Biochemistry and Biophysics UCSF, San Francisco, California 94143, USA; [8]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; [9]Centre for Clinical Research, University of Queensland, Herston, QLD 4006, Australia

"The detected mutations appear to have been harmless passenger mutations: They were enriched in noncoding, AT-rich regions that are not evolutionarily conserved, and they were depleted for genomic elements where mutations might have favorable or adverse effects on cellular fitness, such as regions with actively transcribed genes."

Genome Res. (2014)



Rapid blood cell turnover could contribute to blood-only somatic mutations at extreme age

VAF is bimodal at 20% and 30% - clonal expansion?

Holstege et al., Genome Res. (2014)

REVIEWS | REVIEW

Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network

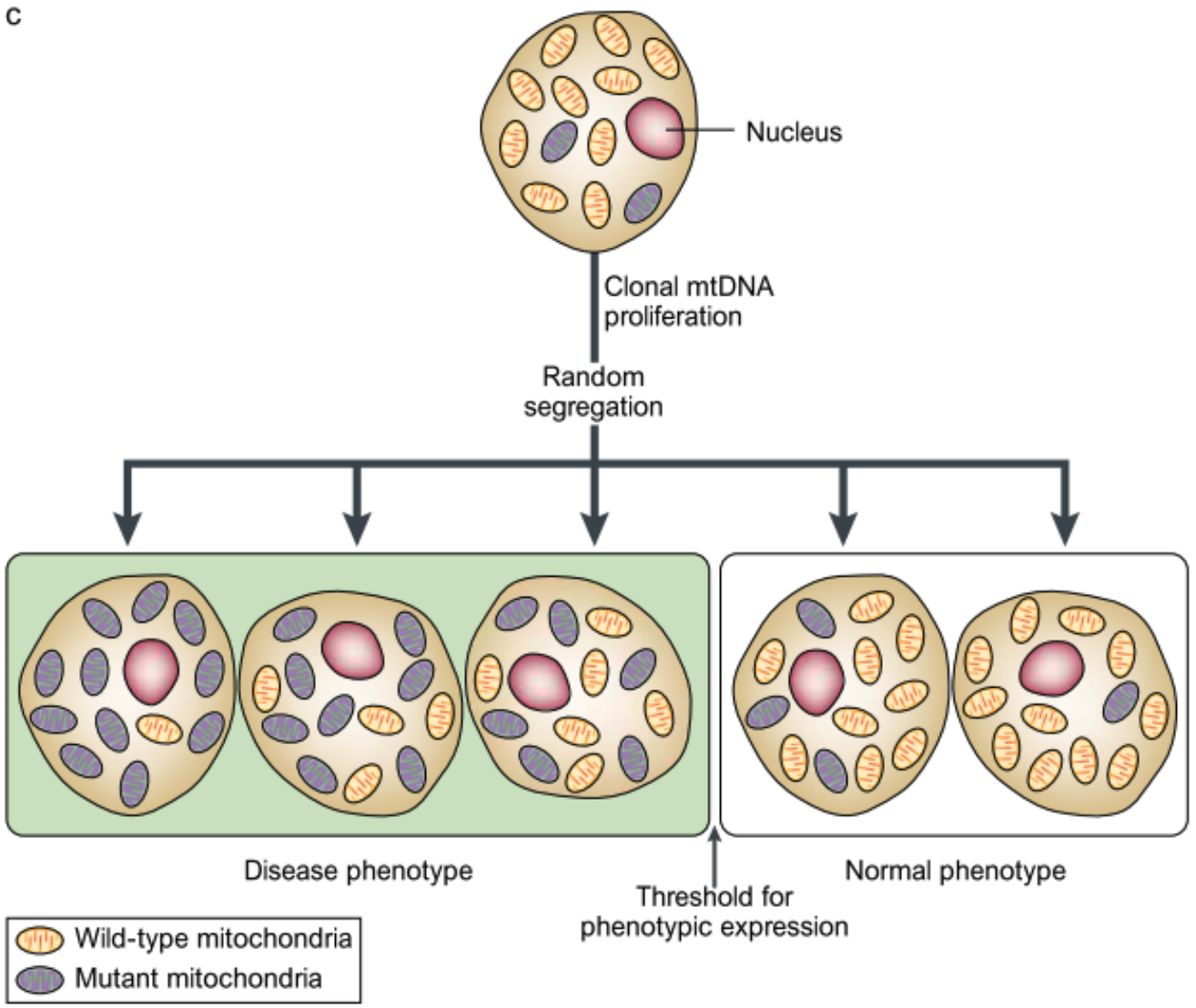
Michael J. McConnell[1,*,†], John V. Moran[2,3,*,†], Alexej Abyzov[4], Schahram Akbarian[5], Taejeong Bae[4], Isidro Cortes-Ciriano[6], Jennifer A. Erwin[7], Liana Fasching[8], Diane A. Flasch[2], Donald Freed[9,10], Javier Ganz[11,12], Andrew E. Jaffe[13], Kenneth Y. Kwan[2,14], Minseok Kwon[6], Michael A. Lodato[11,12], Ryan E. Mills[2,15], Apua C. M. Paquola[7], Rachel E. Rodin[11,12], Chaggai Rosenbluh[16], Nenad Sestan[17], Maxwell A. Sherman[6], Joo Heon Shin[13], Saera Song[18,19], Richard E. Straub[13], Jeremy Thorpe[9,10], Daniel R. Weinberger[13,20,21], Alexander E. Urban[22], Bo Zhou[22], Fred H. Gage[7], Thomas Lehner[23], Geetha Senthil[23], Christopher A. Walsh[11,12], Andrew Chess[16], Eric Courchesne[24], Joseph G. Gleeson[18,19], Jeffrey M. Kidd[2,15], Peter J. Park[6], Jonathan Pevsner[9,10], Flora M. Vaccarino[8,25], Brain Somatic Mosaicism Network[‡]

"The BSMN will examine large collections of postmortem brain tissue from neurotypical individuals and patients with neuropsychiatric disorders. By sequencing brain DNA and single neuronal genomes directly, rather than genomic DNA derived from peripheral blood or other somatic tissues, the BSMN will test the hypothesis that brain somatic variants contribute to neuropsychiatric disease."

Science (2017)

http://brainsomaticmosaicism.org/

Mitochondrial heteroplasmy



Youssoufian and Pyeritz, Nat Rev Genet (2002)