KSBi-BIML 2026

Bioinformatics & Machine Learning(BIML) Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

Somatic mutations in non-cancer diseases

김준호 _ 성균관대학교

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2026 워크샵을 목적으로
제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우
발생하는 **모든 법적 책임은 행위자 본인에게 있음**을 알립니다.

KSBi-BIML 2026

Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics &Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의가 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

강의개요

Somatic mutations in non-cancer diseases

체세포 변이 (somatic mutations)는 수정란의 발생 및 분화 이후 생성된 후천적인 돌연변이로, 부모에게서 물려받은 유전 변이와는 다르게 체내의 일부 세포만이 해당 돌연변이를 갖게 된다. 이러한 체세포 변이에 대한 연구는 주로 암 분야에서 진행되어 왔으며, 그 결과 오늘날 암 발생의 주요 요인이 체세포 변이의 축적임이 밝혀졌으며 실제 암을 일으키는 주요 체세포 변이의 규명 및 발굴 또한 활발히 진행되고 있다.

최근 다양한 연구를 통해 체세포 변이가 암 뿐만 아니라 다른 유형의 질병에도 기여할 수 있음이 점차 밝혀지고 있다. 본 강의에서는 암 이외의 질병에서 체세포 변이가 어떻게 기여하는지 관련된 연구와 분석 기법을 소개하고자 한다. 특히 발생 및 분화과정 초기에 생성되어 일정 세포군이 공유하게 되는 체세포 변이와, 분화 이후 개별 세포에 특이적으로 축적되는 체세포 변이를 나누어 각 변이의 특징과 이를 검출할 수 있는 시퀀싱 기술 및 분석 기법을 소개하고자 한다.

강의는 다음의 내용을 포함한다:

- 체세포 변이의 정의와 유형 및 분류
- 체세포 변이 검출을 위한 시퀀싱 기술 및 분석 기법 소개
- 암 이외의 질병에서 체세포 변이의 역할과 응용 사례
- 체세포 변이 검출 기술의 한계와 발전 방향

* 강의 난이도: 초급

* 강의: 김준호 교수 (성균관대학교 생명과학과)

Curriculum Vitae

Speaker Name: Junho Kim, Ph.D.

▶ **Personal Info**

Name	Junho Kim
Title	Associate Professor
Affiliation	Sungkyunkwan University

▶ **Contact Information**

Address	Department of Biological Sciences, Sungkyunkwan University, Suwon, South Korea
Email	junho.kim@skku.edu

Research Interest

Computational genomics, Bioinformatics, NGS analysis and method development

Educational Experience

2008	B.S. in Bio and Brain Engineering, KAIST, Daejeon, Korea
2009	M.S. in Bio and Brain Engineering, KAIST, Daejeon, Korea
2014	Ph.D. in Bio and Brain Engineering, KAIST, Daejeon, Korea

Professional Experience

2014-2018	Postdoctoral Researcher, Yonsei Biomedical Science Institute, Yonsei University College of Medicine, South Korea
2017-2021	Postdoctoral Research Fellow, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA
2021-2024	Assistant Professor, Department of Biological Sciences, Sungkyunkwan University
2024-	Associate Professor, Department of Biological Sciences, Sungkyunkwan University

Selected Publications (3 maximum)

1. **Junho Kim**, August Yue Huang, Shelby L. Johnson, Jenny Lai, Laura Isacco, Ailsa M. Jeffries, Michael B. Miller, Michael A. Lodato, Christopher A. Walsh* and Eunjung Alice Lee*, Prevalence and mechanisms of somatic deletions in single human neurons during normal aging and in DNA repair disorders, Nature Communications, 2022, 13.1: 5918.
2. **Junho Kim**, Boxun Zhao, August Yue Huang, Michael B. Miller, Michael A. Lodato, Christopher A. Walsh* and Eunjung Alice Lee*, APP gene copy number changes reflect exogenous contamination, Nature, 2020, 584.7821: E20-E28.
3. **Junho Kim**, Dachan Kim, Jae Seok Lim, Ju Heon Maeng, Hyeonju Son, Hoon-Chul Kang, Hojung Nam, Jeong Ho Lee* and Sangwoo Kim*, The use of technical replication for detection of low-level somatic mutations in next-generation sequencing, Nature Communications, 2019, 10.1: 1047.

KSBI | 한국생명정보학회

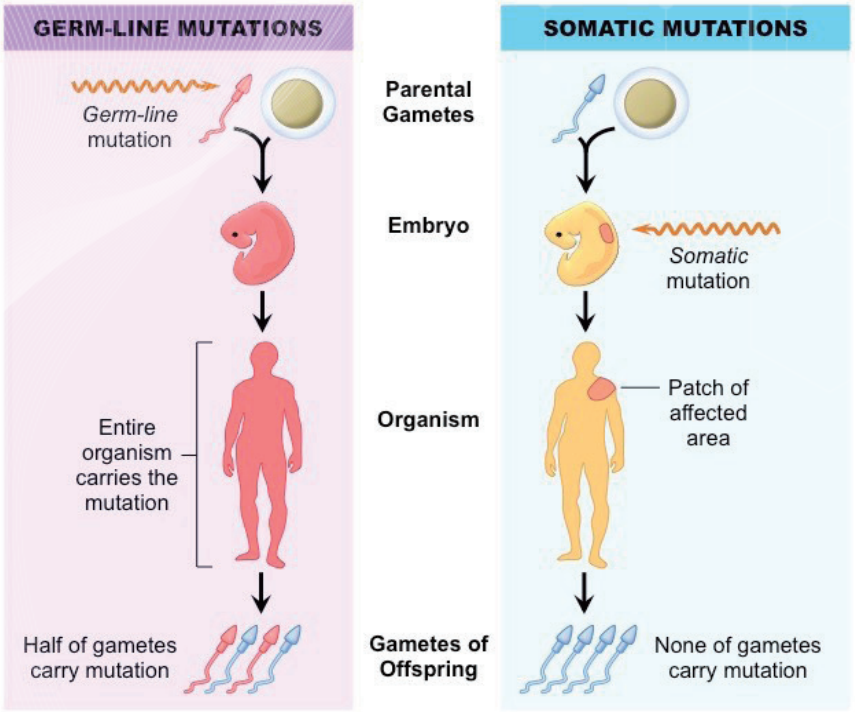
KSBi-BIML 2025

Somatic mutations in non-cancer diseases

Junho Kim, Ph.D.

Department of Biological Sciences
Sungkyunkwan University

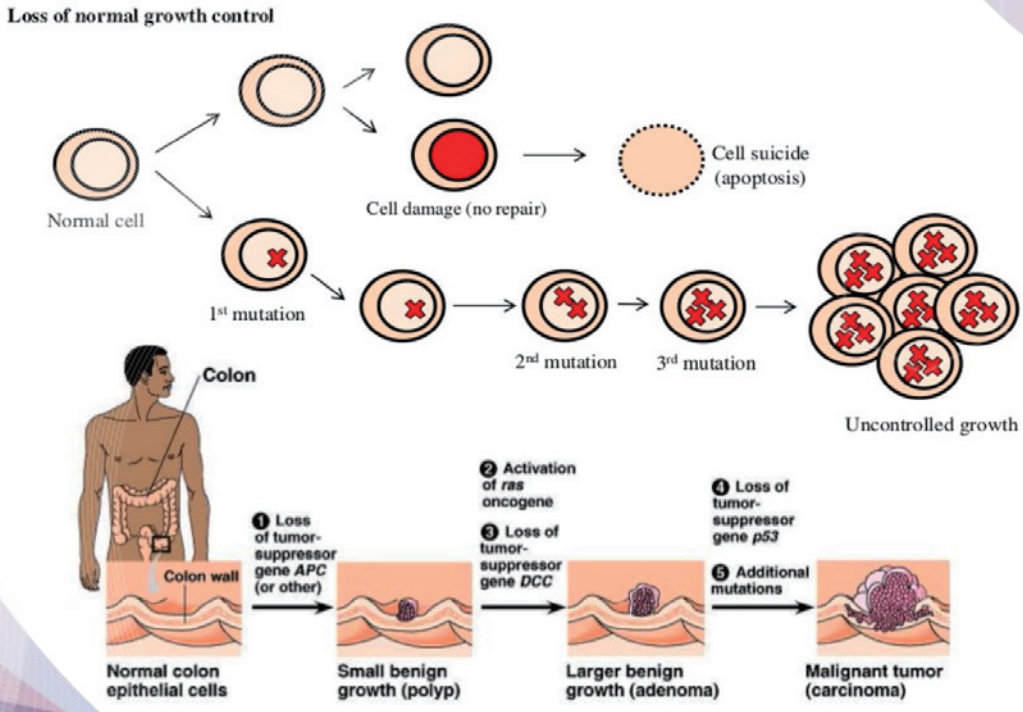
Germline vs. Somatic mutations



Cornell B. Somatic vs Germline mutations, 2016

2

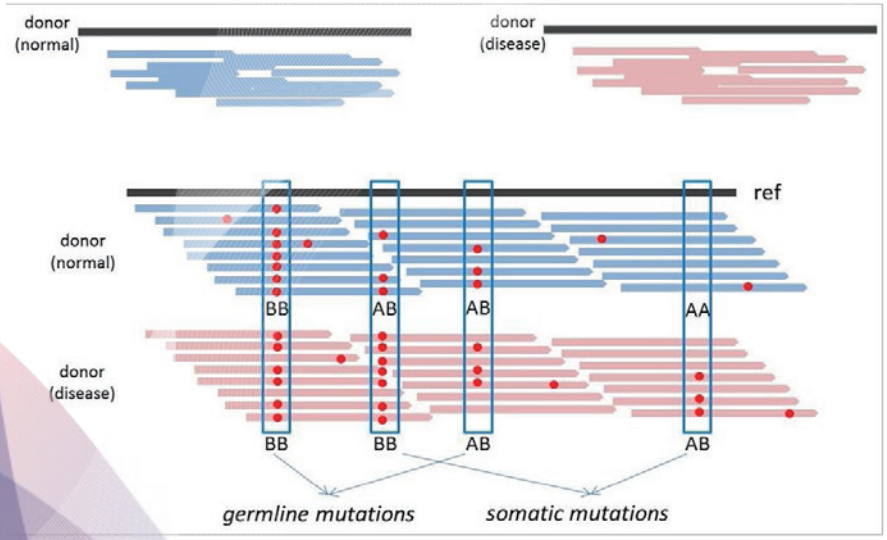
Cancer arises from somatic mutations in cells

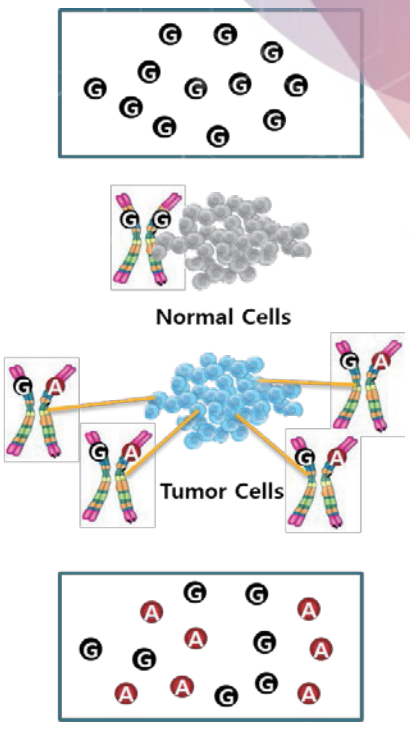


Yulug I. *Molecular basis of cancer,* 2006

3

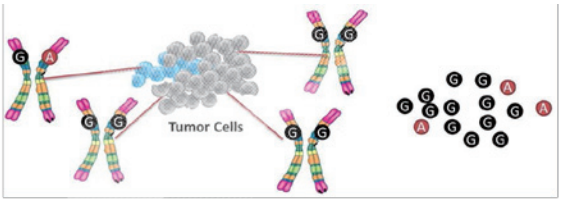
Somatic mutation detection from NGS data

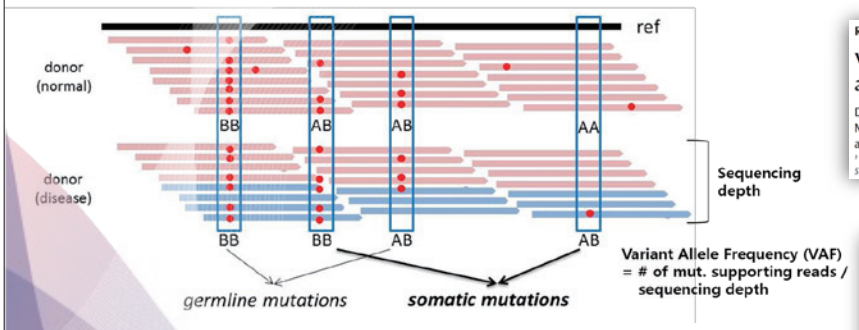


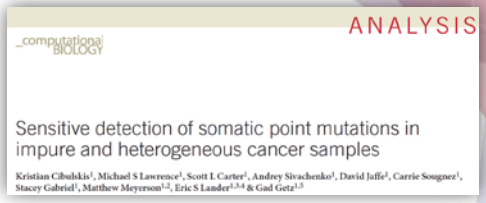


4

Low-level somatic mutation detection






ANALYSIS

computational BIOLOGY

Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples

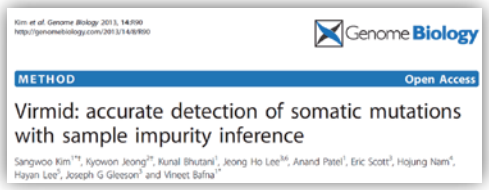
Kristian Cibulskis[1], Michael S Lawrence[1], Scott L Carter[1], Andrey Sivachenko[1], David Jaffe[1], Carrie Sougnez[1], Stacey Gabriel[1], Matthew Meyerson[1,2], Eric S Lander[1,3,4] & Gad Getz[1,5]


Resource

VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing

Daniel C. Koboldt,[1] Qunyuan Zhang,[1] David E. Larson,[1] Dong Shen,[1] Michael D. McLellan,[1] Ling Lin,[1] Christopher A. Miller,[1] Elaine R. Mardis,[1,2,3] Li Ding,[1,2,4] and Richard K. Wilson[1,2,3,4]

[1]The Genome Institute, Washington University, St. Louis, Missouri 63108, USA; [2]Department of Genetics, Washington University, St. Louis, Missouri 63110, USA; [3]Siteman Cancer Center, Washington University, St. Louis, Missouri 63110, USA


Kim et al. Genome Biology 2013, 14:R90
http://genomebiology.com/2013/14/8/R90

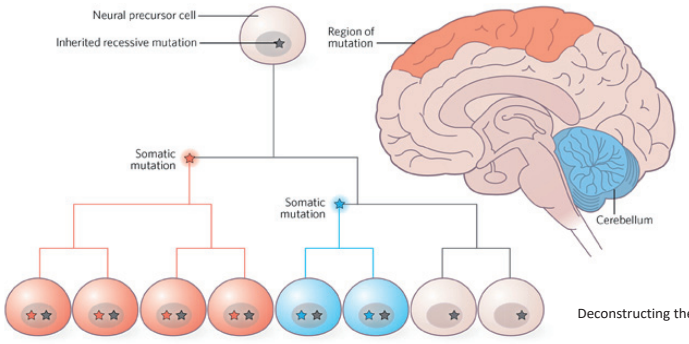
Genome Biology

METHOD Open Access

Virmid: accurate detection of somatic mutations with sample impurity inference

Sangwoo Kim[1*†], Kyowon Jeong[2*†], Kunal Bhutani[1], Jeong Ho Lee[3,6], Anand Patel[1], Eric Scott[3], Hojung Nam[4], Hayan Lee[5], Joseph G Gleeson[3] and Vineet Bafna[1*]

Mosaic brain hypothesis



Deconstructing the Mosaic Brain, Tom Curran, The Scientist, 2011

- In most tissues, cell turnover and regeneration improve the deleterious effects of somatic mutations
- However, most neurons cannot be replaced after neural development

Somatic mutations in the brain may have permanent effect as germline mutations

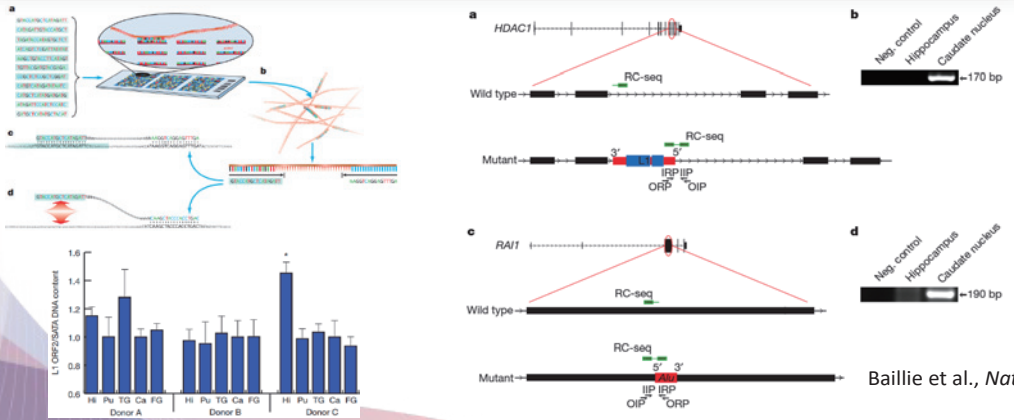
Somatic mutations in human brain

LETTER

doi:10.1038/nature10531

Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie[1]*, Mark W. Barnett[1]*, Kyle R. Upton[1]*, Daniel J. Gerhardt[2], Todd A. Richmond[2], Fioravante De Sapio[1], Paul Brennan[3], Patrizia Rizzu[4], Sarah Smith[1], Mark Fell[1], Richard T. Talbot[1], Stefano Gustincich[5], Thomas C. Freeman[1], John S. Mattick[6], David A. Hume[1], Peter Heutink[4], Piero Carninci[7], Jeffrey A. Jeddeloh[2] & Geoffrey J. Faulkner[1]



Baillie et al., *Nature,* 2011

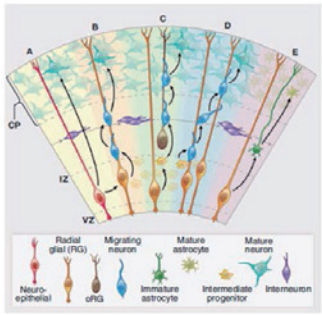
Somatic mutations cause neurodevelopmental diseases

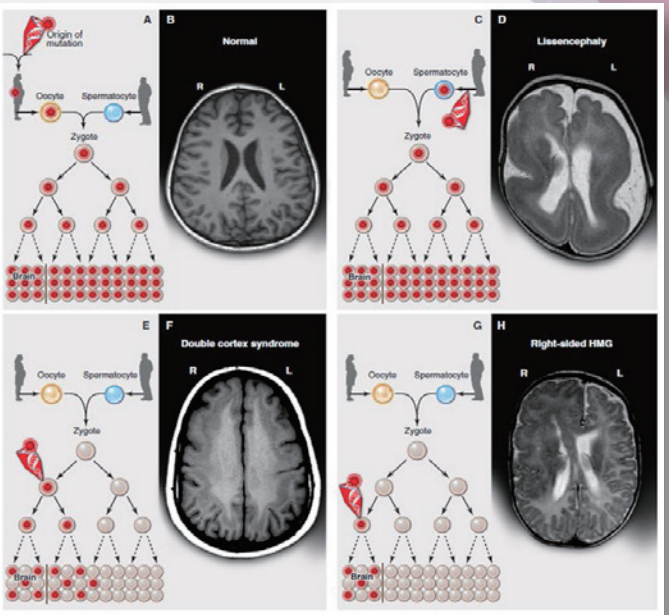
Somatic Mutation, Genomic Variation, and Neurological Disease

Annapurna Poduri, Gilad D. Evrony, Xuyu Cai, Christopher A. Walsh*

Background: Genetic mutations that cause human disease are conventionally considered to be inherited from one's parents and present in all somatic (body) cells. We do know, however, that most mutations that cause cancer arise somatically, and we are becoming increasingly aware of mutations that cause other diseases and that arise de novo, meaning they are undetectable in the parents. Some such de novo mutations arise in the gamete of a parent, but some arise after fertilization during embryonic development, generating somatic mutations. Somatic mutations occur in several neurodevelopmental diseases associated with epilepsy, autism spectrum disorders, and intellectual disability, although their broader relevance for neurological disease is unknown.

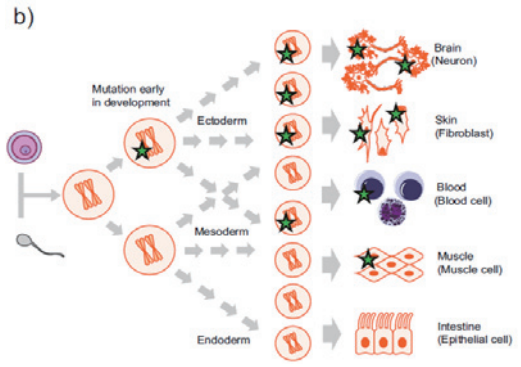
Advances: A key recent advance has been the increasing identification of somatic mutations in affected tissues. For example, somatic mutations in several genes (*PIK3CA*, *AKT3*, and *mTOR*) cause enlargement of just one hemisphere of the brain, a malformation called hemimegalencephaly that is highly associated with epilepsy. These mutations may or may not be found in the blood, the convenient source tissue for DNA analysis, thus presenting a challenge to disease gene identification. Remarkably, patients can show dysfunction of essentially an entire half of their cerebral cortex when only 8 to 35% of the brain cells carry the mutation, suggesting that a minority of cells with a somatic mutation can disrupt the function of widespread cortical

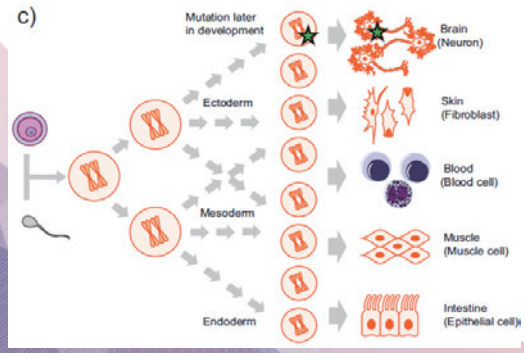


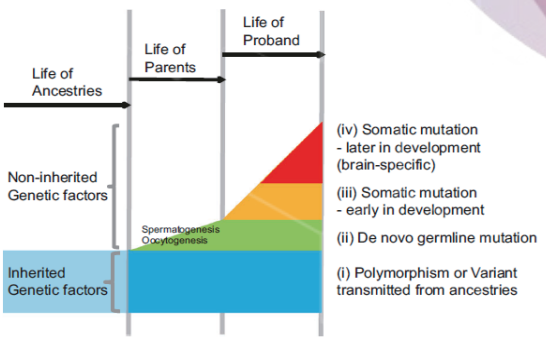


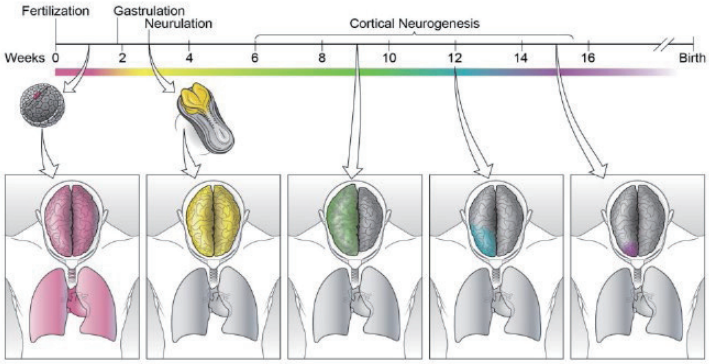
Poduri et al., *Science,* 2013

Somatic mutations cause neurodevelopmental diseases





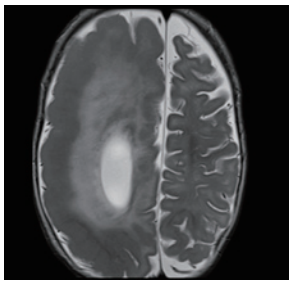




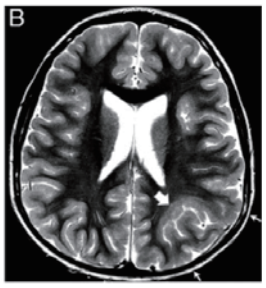
Nishioka et al., *Mol. Psychiatry,* 2018

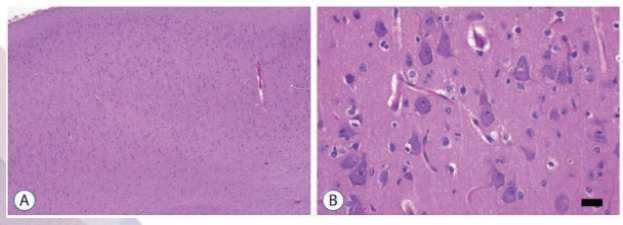
Hemimegalencephaly (HMG) and focal cortical dysplasia (FCD)

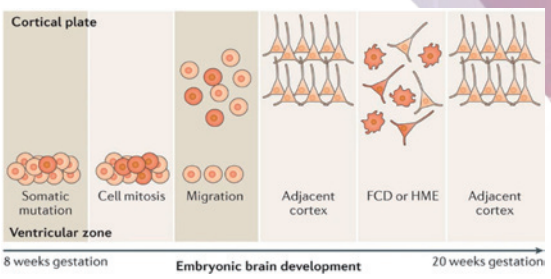
HMG

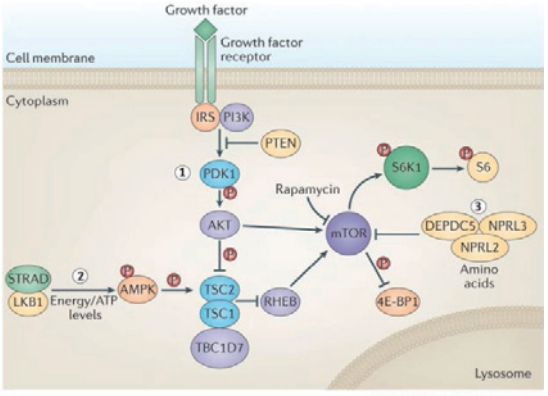


FCD

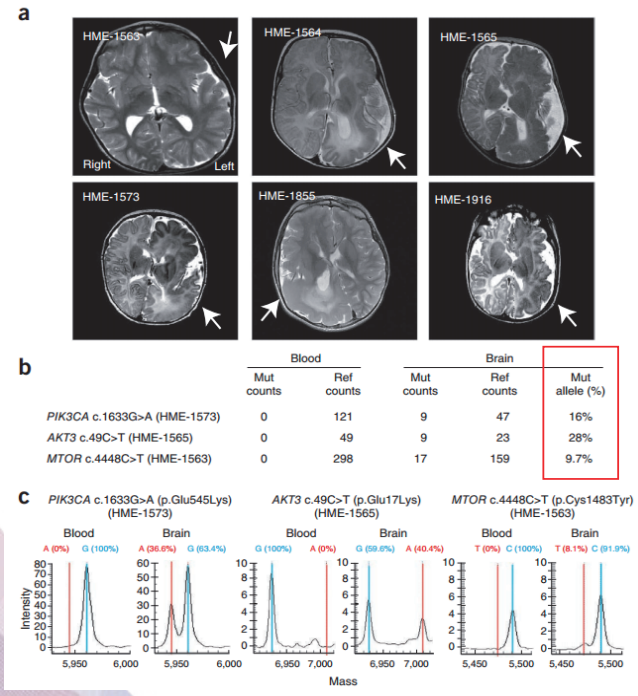




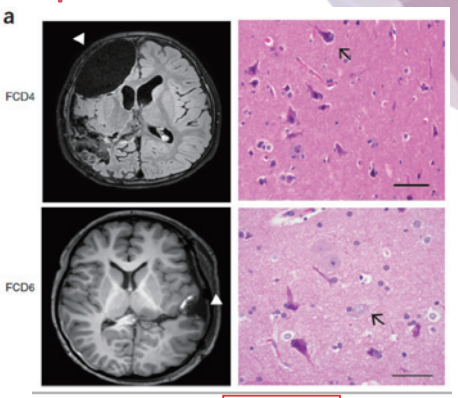




Low-level mutations in neurodevelopmental diseases



Lee J et al, *Nat. Genet.*, 2012

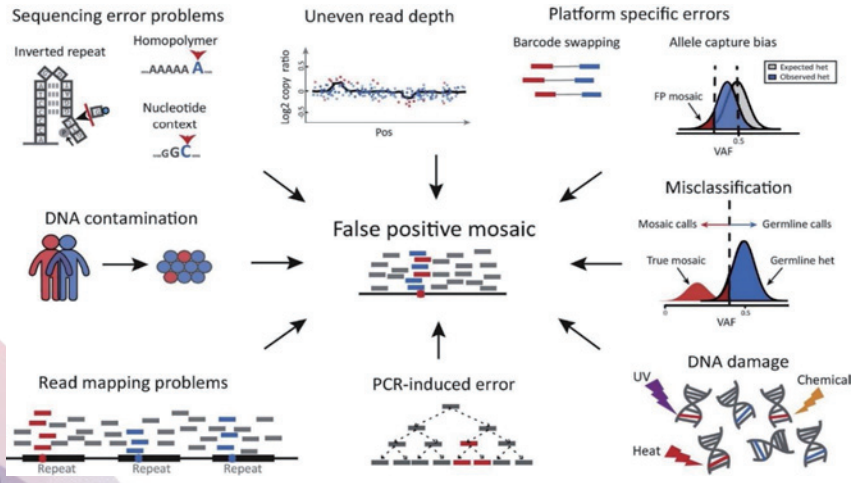


Nucleotide changes	Protein changes	Hybrid capture % mutated allele	PCR amplicon sequencing % mutated allele
c.7280T>C	p.Leu2427Pro	4.11–9.63%	6.57–12.63%
c.7280T>C	p.Leu2427Pro	4.57–6.90%	3.48–7.28%
c.6577C>T	p.Arg2193Cys	2.99%	1.26%
c.1871G>A	p.Arg624His	1.80%	4.41%
c.5126G>A	p.Arg1709His	1.63%	1.52%
c.6644C>T	p.Ser2215Phe	2.41%	2.11%
c.7280T>A	p.Leu2427Gln	3.05%	2.86~5.11%

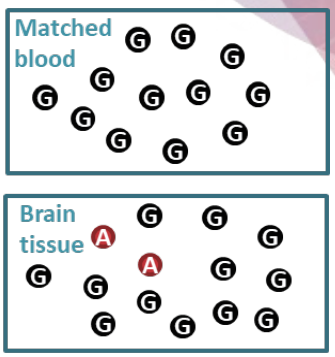
Lim J et al, *Nat. Med.*, 2015

11

Overall challenge: amount of signal may be comparable to noise



Dou et al., *Trends in Genetics*, 2018



No mutation.
Two 'A's are from errors.

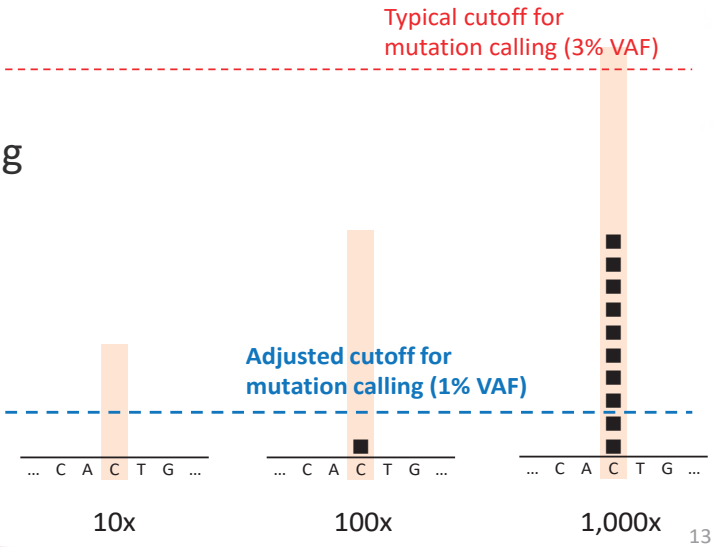
vs.

The sample has a low-level somatic mutation.

12

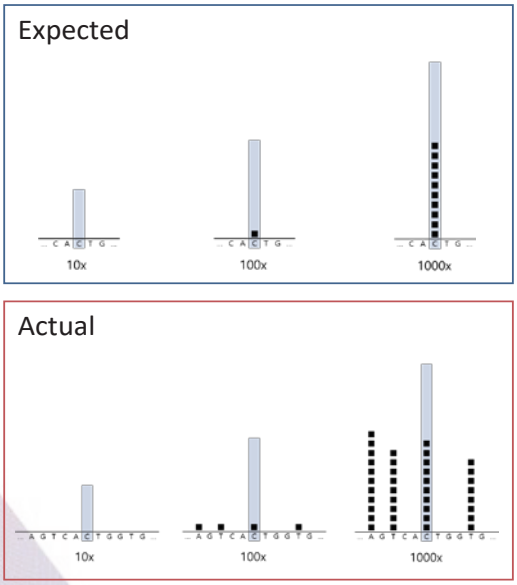
Detecting low-VAF allele - Is high depth sufficient?

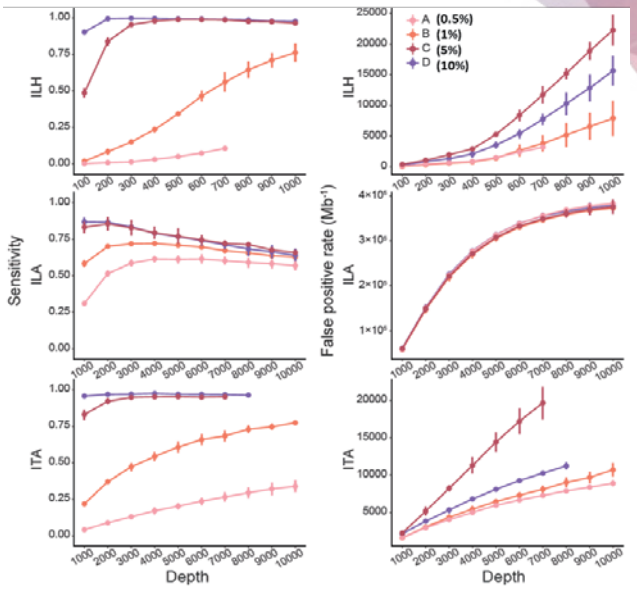
- Ultra-high sequencing depth (>10,000x) may resolve some part of the problem
  - For low-level mutation detection with 1% VAF
    - 1 read in 100x
    - 10 reads in 1,000x
    - 100 reads in 10,000x
- Conventional mutation calling with adjusted threshold
  - Problem solved?



13

Actual problem to detect low-level mutations

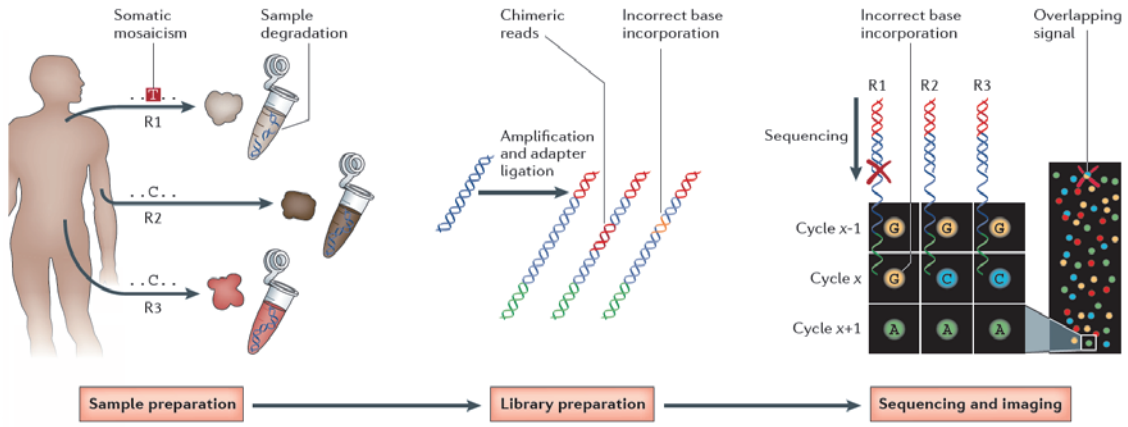




Kim et al., *Nat. Comm.*, 2019

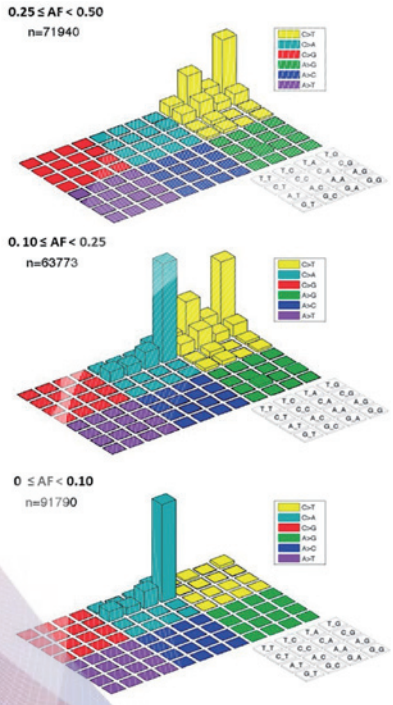
14

Innate technical errors under sequencing experiments

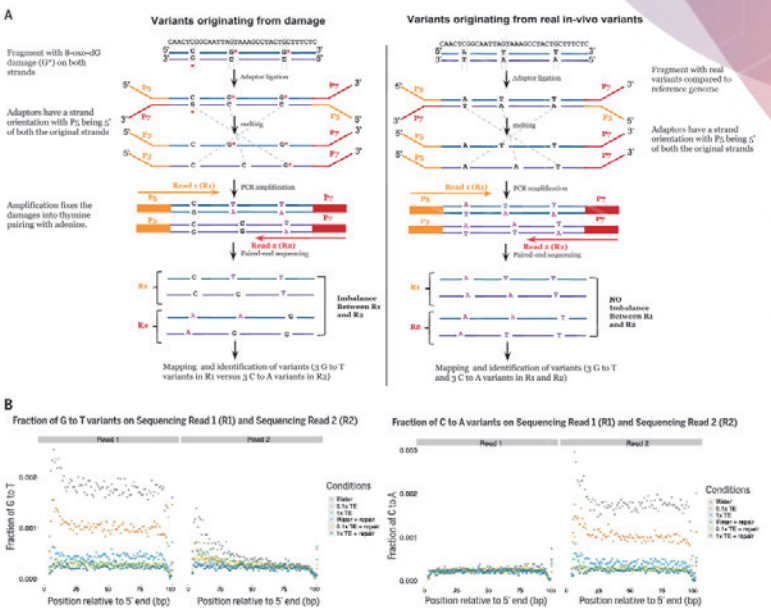


Robasky et al., *Nat. Rev. Genet.*, 2014

Oxidative DNA damage during sample preparation

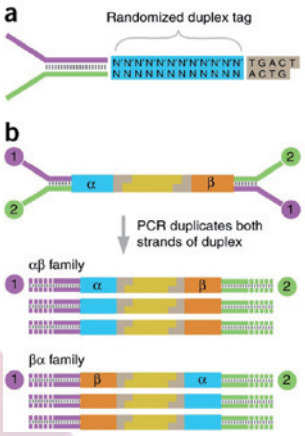


Costello et al., *Nucleic Acids Research*, 2012

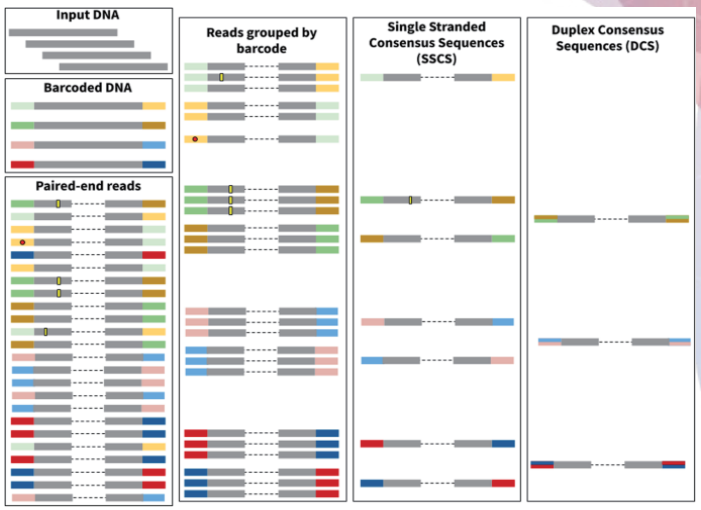


Chen et al., *Science*, 2017

Barcoded sequencing

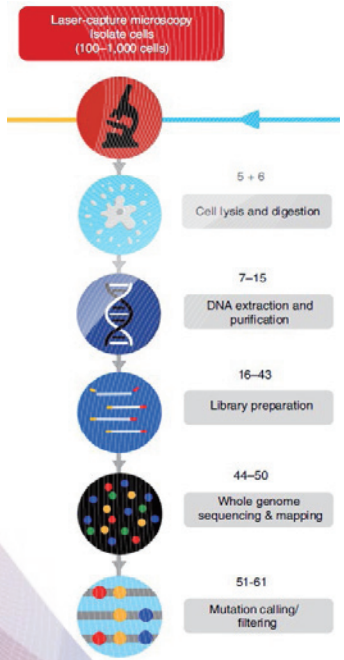


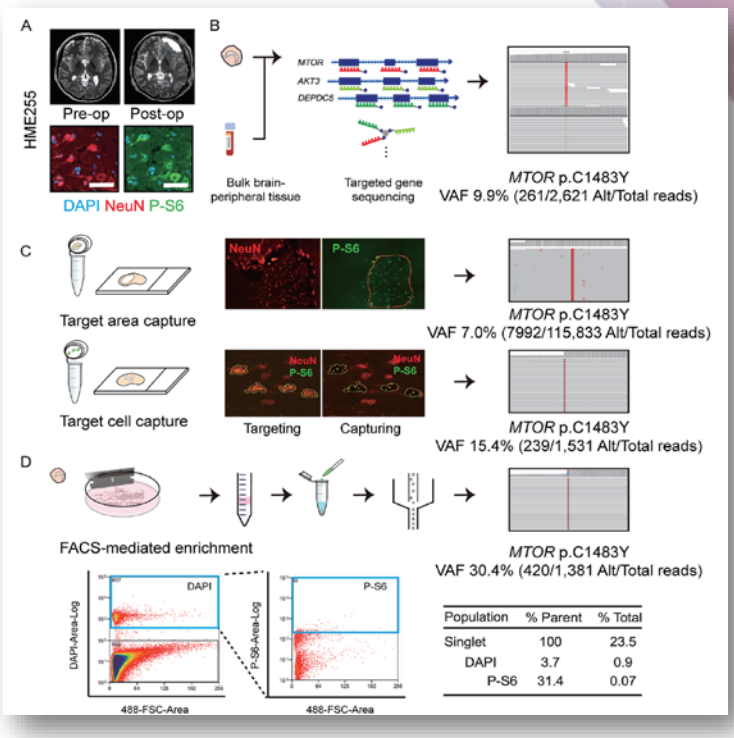
Kennedy et al., *Nat. Protoc.*, 2014



Stoler et al., *BMC Bioinform.*, 2020

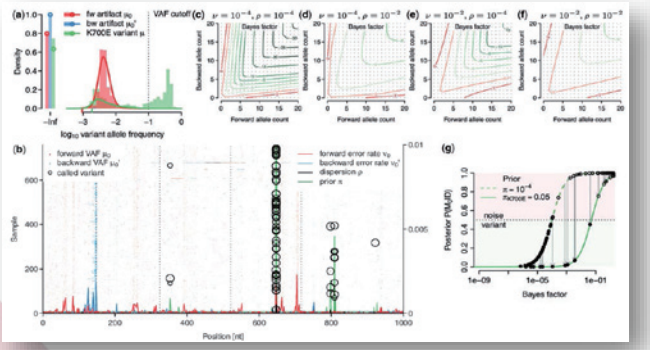
Enrichment of mutation-harboring cells



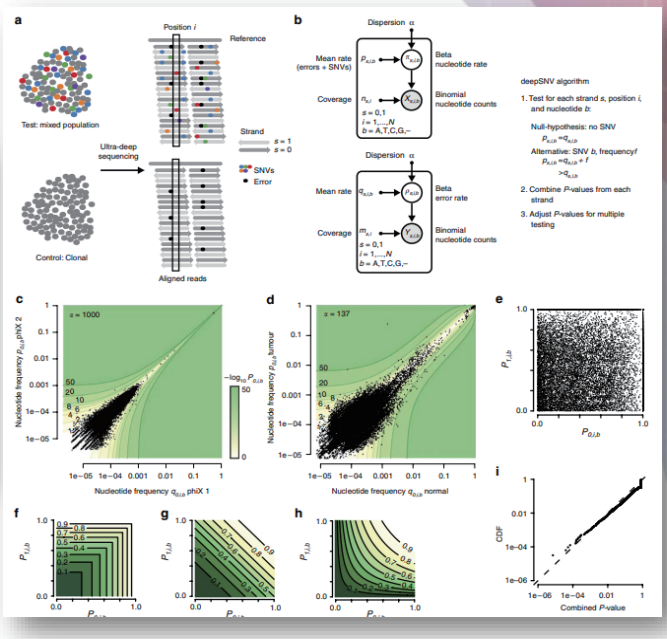


Kim et al., *Ann. Neurol.*, 2023

Base-specific error model



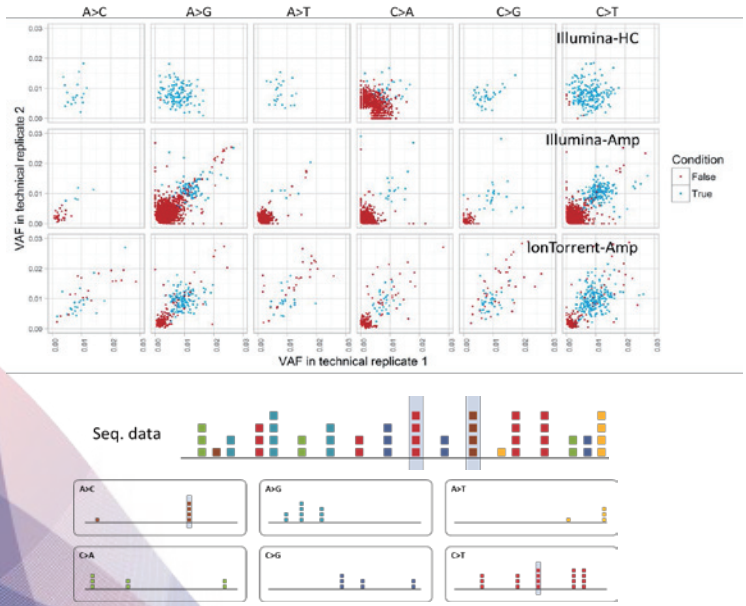
Martincorena et al., *Bioinformatics*, 2014



Gerstung et al., *Nat. Comm.*, 2012

19

Substitution-type-specific error model



For *i*-th position with b alleles,

$$x_i = b_i / n_i$$

$$P(x_i) = P(x_i \cap Mut) + P(x_i \cap TE)$$

$$= \boxed{P(x_i \mid Mut) \cdot P(Mut)} + \boxed{P(x_i \mid TE) \cdot P(TE)}$$

Binomial probability based on base qualities

Estimated by substitution-type-specific VAF profiles

$$S_i = \log\left(\frac{P(M_v)\prod_{j=1}^{k} P(x_j^i \mid M_v)}{P(M_0)\prod_{j=1}^{k} P(x_j^i \mid M_0)}\right)$$

$$= \log\left(\frac{\phi_v \prod_{j=1}^{k} N(x_{TEj}^{\ i}; \hat{\mu}_{TE}^{\ i}, \hat{\sigma}_{TE}^{\ i\,2})}{\phi_{TE} \prod_{j=1}^{k} Exp(x_{TEj}^{\ i}; \lambda_{TE})}\right)$$

Reference allele

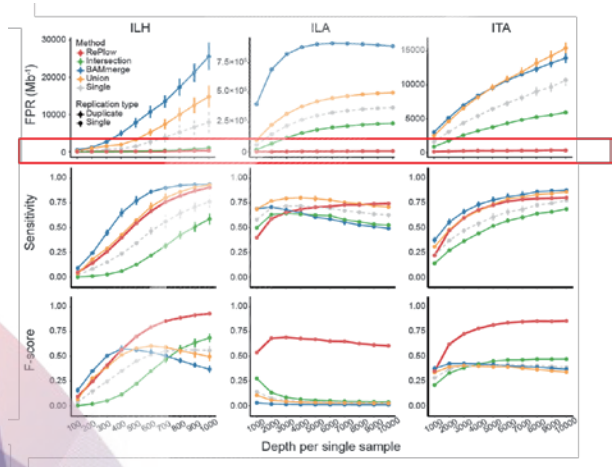
G

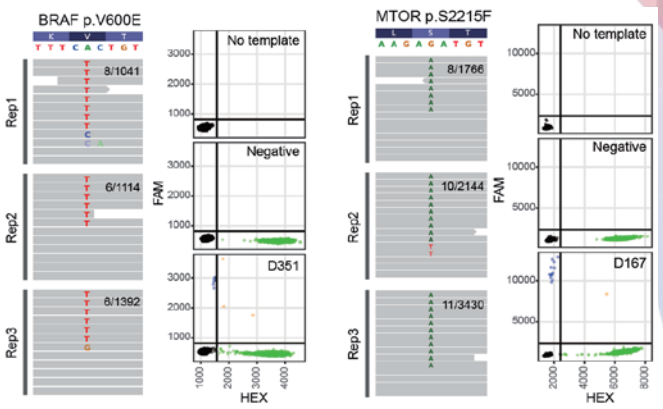
A A A A G G G G G G G G G G G G

Kim et al., *Nat. Comm.*, 2019

20

Substitution-type-specific error model





Kim et al., *Nat. Comm.*, 2019

Ultra low-level somatic SNV calling algorithms

LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets

Andreas Wilm[1], Pauline Poh Kim Aw[1], Denis Bertrand[1], Grace Hui Ting Yeo[1], Swee Hoe Ong[1], Chang Hua Wong[1], Chiea Chuen Khor[1], Rosemary Petric[2], Martin Lloyd Hibberd[1] and Niranjan Nagarajan[1,*]

[1]Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore and [2]Hoffmann-La Roche, Bldg 85/521340 Kingsland Street, Nutley, NJ 07110, USA

BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 15 2013, pages 1908–1909 doi:10.1093/bioinformatics/btt305

Sequence analysis Advance Access publication May 27, 2013

Mutascope: sensitive detection of somatic mutations from deep amplicon sequencing

Shawn E. Yost[1], Hakan Alakus[2,3], Hiroko Matsui[2,3], Richard B. Schwab[4,5], Kristen Jepsen[2,3], Kelly A. Frazer[2,3,4,6,7] and Olivier Harismendy[2,3,4,6,*]

[1]Bioinformatics Graduate Program, [2]Department of Pediatrics, [3]Rady Children's Hospital, [4]Moores UCSD Cancer Center, [5]Department of Medicine, [6]Clinical and Translational Research Institute and [7]Institute for Genomic Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, USA

Associate Editor: Inanc Birol

BIOINFORMATICS ORIGINAL PAPER Vol. 30 no. 9 2014, pages 1198–1204 doi:10.1093/bioinformatics/btt750

Genome analysis Advance Access publication January 16, 2014

Subclonal variant calling with multiple samples and prior knowledge

Moritz Gerstung[1], Elli Papaemmanuil[1] and Peter J. Campbell[1,2,3,*]

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK, [2]Department of Haematology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK and [3]Department of Haematology, University of Cambridge, Cambridge CB22XY, UK

Associate Editor: Michael Brudno

nature biotechnology ANALYSIS https://doi.org/10.1038/s41587-021-00861-3

A unified haplotype-based method for accurate and comprehensive variant calling

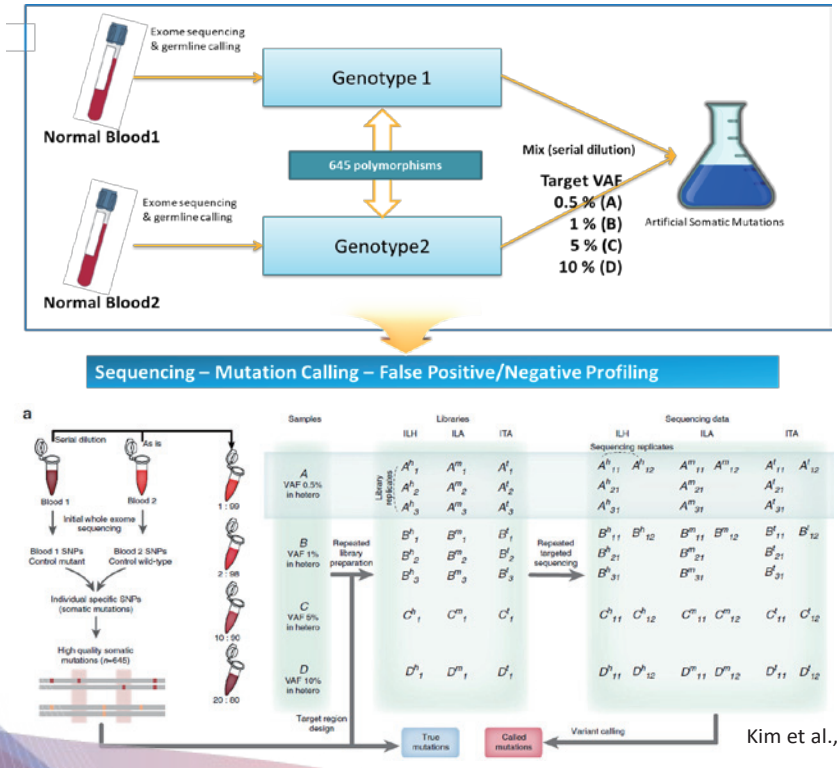
Daniel P. Cooke[1], David C. Wedge[2] and Gerton Lunter[1,3]

ARTICLE https://doi.org/10.1038/s41467-019-09026-y OPEN

The use of technical replication for detection of low-level somatic mutations in next-generation sequencing

Junho Kim[1], Dachan Kim[1], Jae Seok Lim[2], Ju Heon Maeng[1], Hyeonju Son[1], Hoon-Chul Kang[3], Hojung Nam[4], Jeong Ho Lee[2] & Sangwoo Kim[1]

Spike-in experiment for benchmarking



Kim et al., *Nat. Comm.*, 2019

Genome in a Bottle (GIAB)

DESCRIPTION

Consortium goals:

The Genome in a Bottle Consortium is a public-private-academic consortium hosted by NIST to develop the technical infrastructure (reference standards, reference methods, and reference data) to enable translation of whole human genome sequencing to clinical practice and innovations in technologies. The priority of GIAB is comprehensive characterization of several human genomes for use in benchmarking, including analytical validation and technology development, optimization, and demonstration.

Reference samples:

GIAB has currently characterized a pilot genome (NA12878/HG001) from the HapMap project, and two son/father/mother trios of Ashkenazi Jewish and Han Chinese ancestry from the Personal Genome Project (selected because, unlike the pilot genome, they are consented for commercial redistribution). These samples are available from NIST and Coriell, and their IDs from NIST, Coriell, and PGP are in this table (see FAQ for differences between NIST and Coriell samples).

Benchmark (or "High-confidence") variant calls and regions:

We developed an integration pipeline to utilize sequencing data generated by multiple technologies to generate variant calls and regions for use in benchmarking and validating variant calling pipelines. Currently, benchmark VCF and BED files for small variants are available for GRCh37 and GRCh38 under each genome at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/

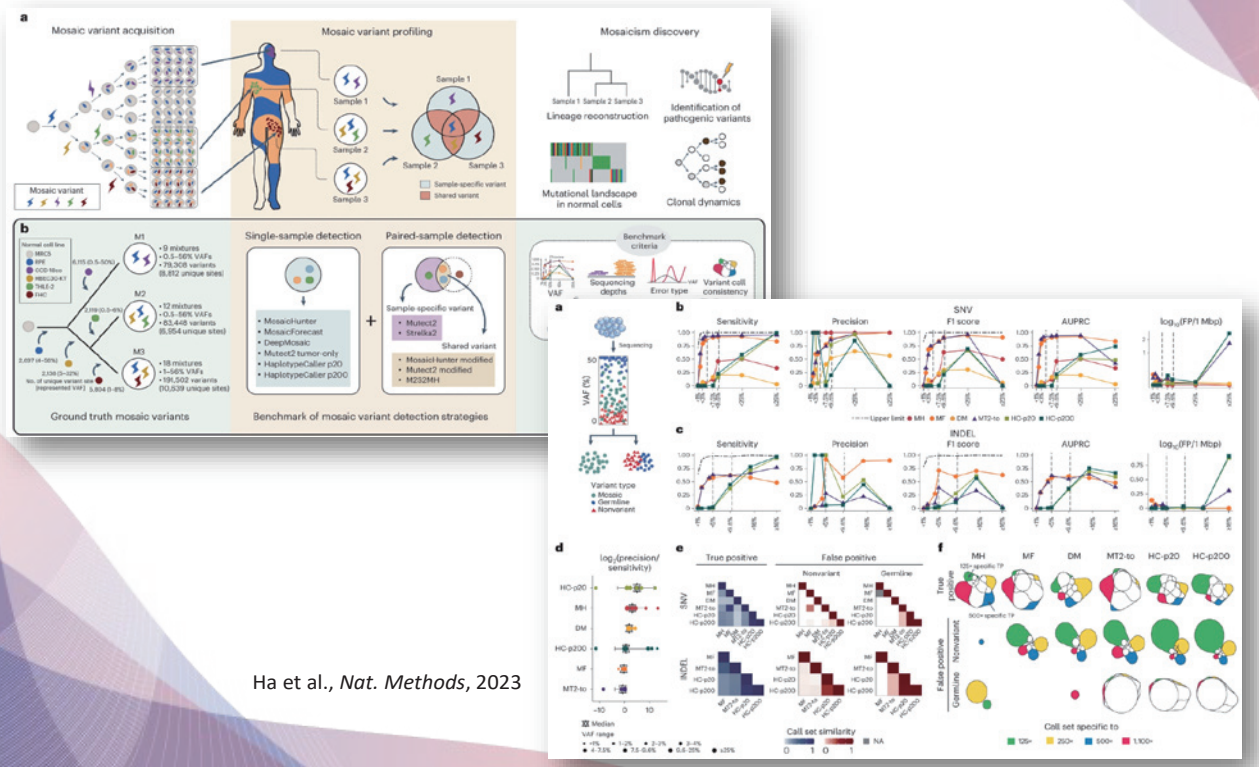
GIAB's versions of GRCh37 and GRCh38 reference fasta files, including a new GRCh38 reference in collaboration with the GRC that masks false duplications in GRCh38, are at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/references/.

AshkenazimTrio

Son:HG002 https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG002_NA24385_son/
Father:HG003 https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG003_NA24149_father/
Mother:HG004 https://ftp.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/HG004_NA24143_mother/

Sequencing Platform	Sequence	Alignment
Illumina WGS 2x150bp 300X per individual	All HG002 HG003 HG004	novoalign: All HG002 HG003 HG004
Illumina 6KB Matepair	All HG002 HG003 HG004	bwamem:hg19 All HG002 HG003 HG004
Illumina WGS 2X250bp	All HG002 HG003 HG004	isaac:hg19 All HG002 HG003 HG004 novoalign: All HG002 HG003 HG004
Moleculo	All HG002 HG003 HG004	
Illumina Whole Exome	-	bwamem:hg19 All HG002 HG003 HG004
SOLiD 60x for son	All HG002	LifeScope:hg19 All HG002
CompleteGenomics	-	CGAtools:hg19 All HG002 HG003 HG004
Ion Proton 1000x Exome	-	TMAP:hg19 All HG002 HG003 HG004
10X Genomics	-	bwamem:hg19 All HG002 HG003 HG004
10X Genomics ChromiumGenome	All HG002	LongRanger2.0:hg19 All HG002 HG003 HG004
BioNano	All:bnx HG002:bnx HG003:bnx HG004:bnx	All:cmap HG002 HG003 HG004
PacBio 70x/30x/30x	All HG002 HG003 HG004 All:hdf5 HG002 HG003 HG004	NGMLR:hg19 All HG002 HG003 HG004 minimap2: All HG002 HG003 HG004
PacBio CCS 10kb	All HG002	pbmm2:hg19 All HG002
PacBio CCS 11kb	All HG002	pbmm2:hg19 All HG002
PacBio CCS 15kb	All HG002	pbmm2:hg19 All HG002
PacBio CCS 15kb_20kb chemistry2	All HG002	pbmm2: All HG002 HG003 HG004

Comprehensive benchmarking of mosaic variant calling strategies



Ha et al., *Nat. Methods*, 2023

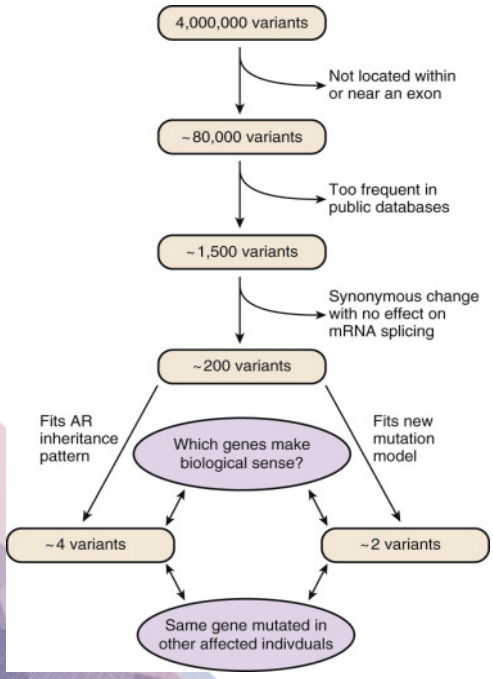
Key questions for somatic variant discovery

Objective II: Are they functional?
- Are they damaging? pathogenic?
- Are they related to phenotypes?

Bunch of candidate variants → **Many variants** → **A few variants**

Objective I: Do they really exist?
- Any mistakes in sequencing and variant calling?
- Any exogenous source that may mimic the variant?

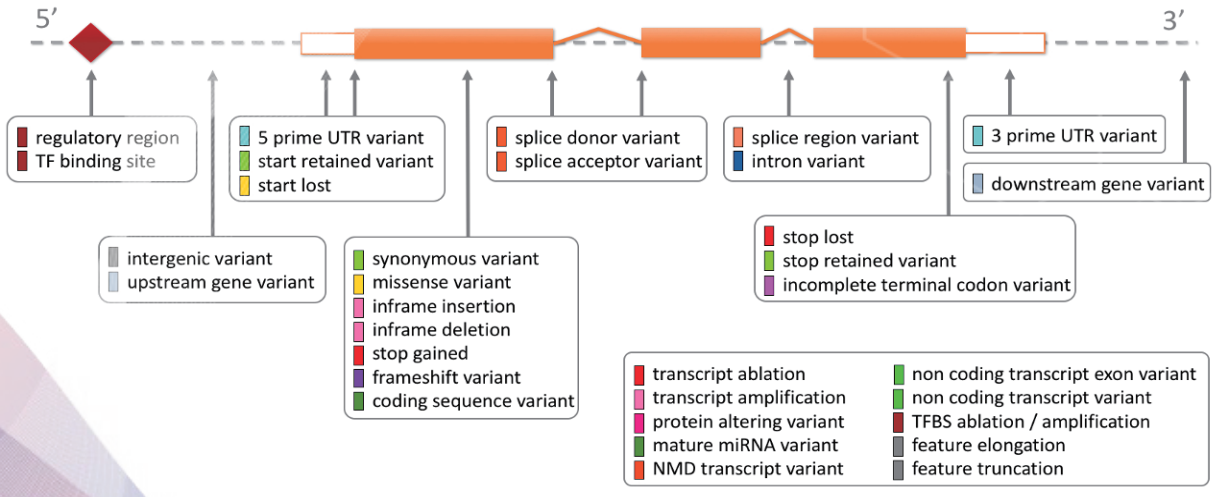
It's not easy to find disease-causing variants



- We usually have too many variants
  - Disease-causing mutations should be rare enough or recurrent across patients
- Factors considered in filtering variant candidates
  - Variant location with respect to protein-coding genes
  - Deleterious nature of the variant
  - Population frequency of a given variant

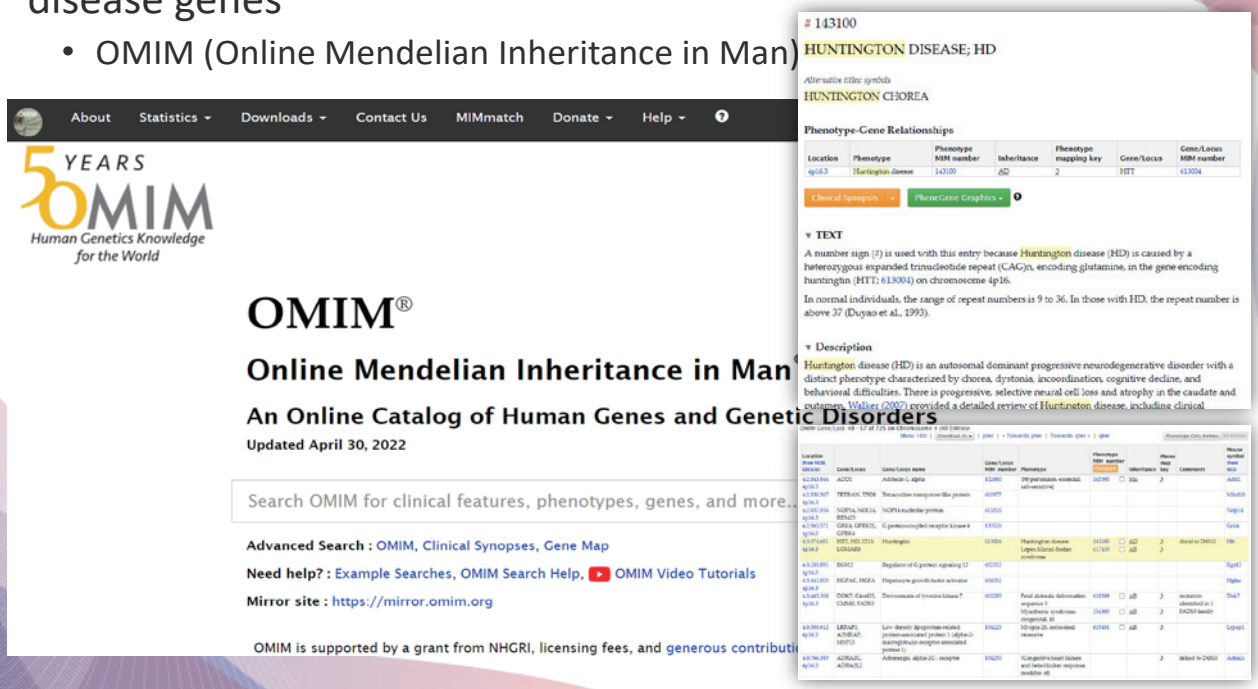
Variant annotation

- The process of assigning functional information to DNA variants
  - Evaluates which functional consequences will be made for variants

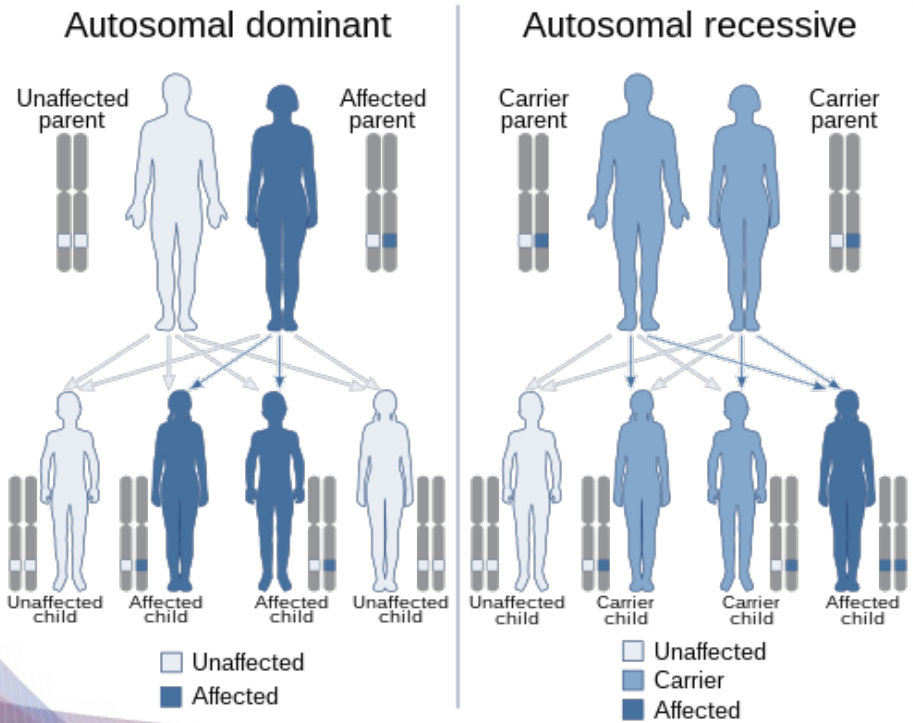


Variant filtering and prioritization

- Utilizing reported information of genetic disorders and related disease genes
  - OMIM (Online Mendelian Inheritance in Man)



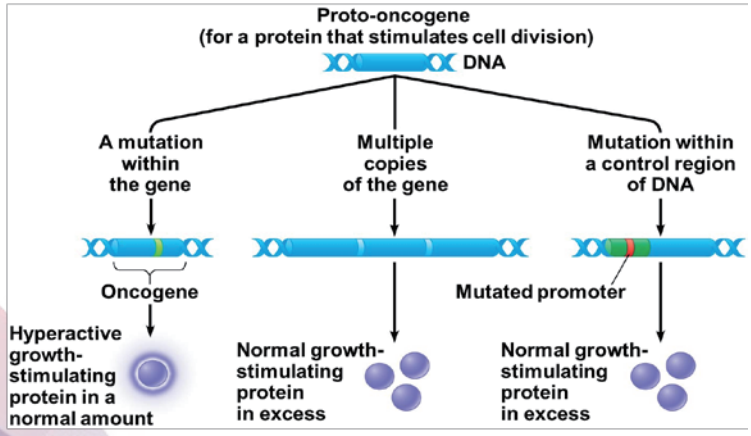
Mode of inheritance

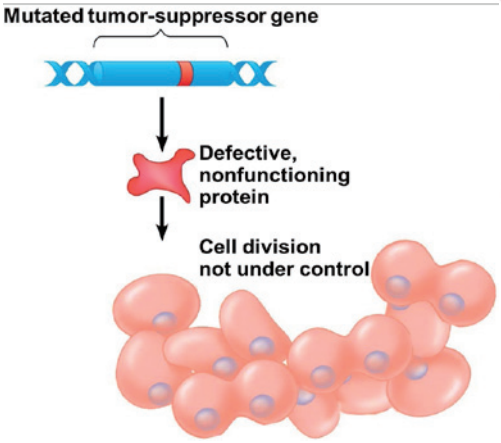


30

Functional effect of a variant

- Gain-of-function (GoF) vs. Loss-of-function (LoF)

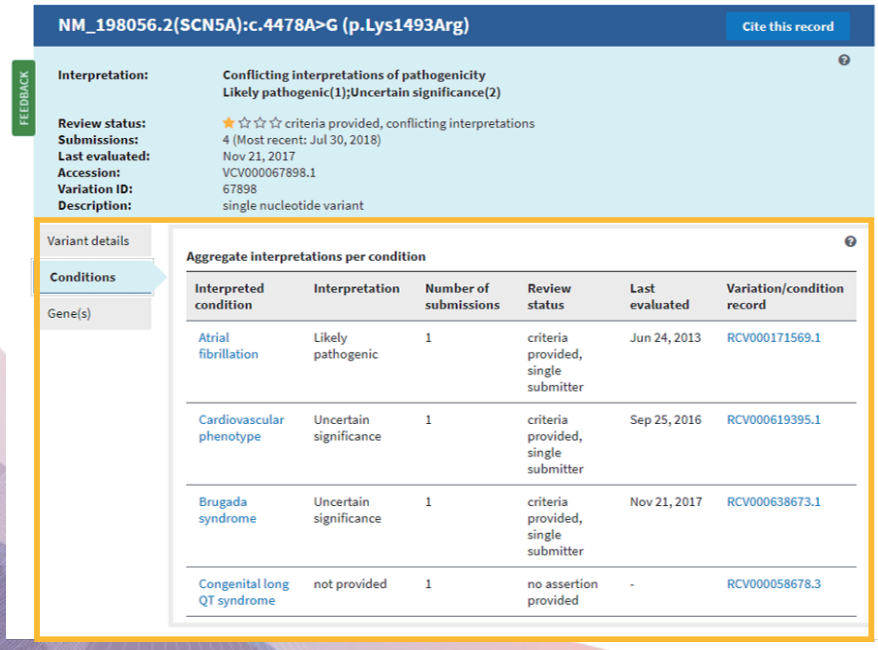




31

Variant filtering and prioritization

- Utilizing reported information about genomic variants and its relationship to human health



NM_198056.2(SCN5A):c.4478A>G (p.Lys1493Arg) — Cite this record

FEEDBACK

Interpretation:	Conflicting interpretations of pathogenicity Likely pathogenic(1);Uncertain significance(2)
Review status:	★☆☆☆ criteria provided, conflicting interpretations
Submissions:	4 (Most recent: Jul 30, 2018)
Last evaluated:	Nov 21, 2017
Accession:	VCV000067898.1
Variation ID:	67898
Description:	single nucleotide variant

Variant details | Conditions | Gene(s)

Aggregate interpretations per condition

Interpreted condition	Interpretation	Number of submissions	Review status	Last evaluated	Variation/condition record
Atrial fibrillation	Likely pathogenic	1	criteria provided, single submitter	Jun 24, 2013	RCV000171569.1
Cardiovascular phenotype	Uncertain significance	1	criteria provided, single submitter	Sep 25, 2016	RCV000619395.1
Brugada syndrome	Uncertain significance	1	criteria provided, single submitter	Nov 21, 2017	RCV000638673.1
Congenital long QT syndrome	not provided	1	no assertion provided	-	RCV000058678.3

- COSMIC
- HGMD
- ClinVar
- OMIM

...

COSMIC Catalogue Of Somatic Mutations In Cancer

OMIM®

ClinVar

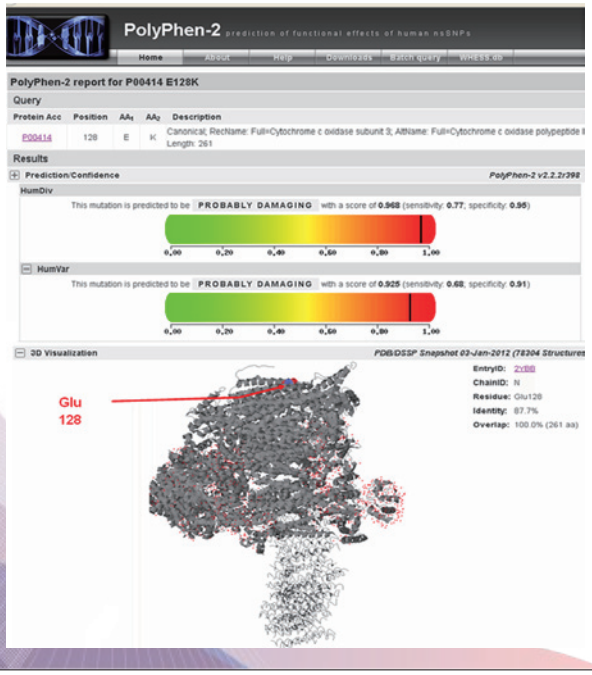
HGMD®

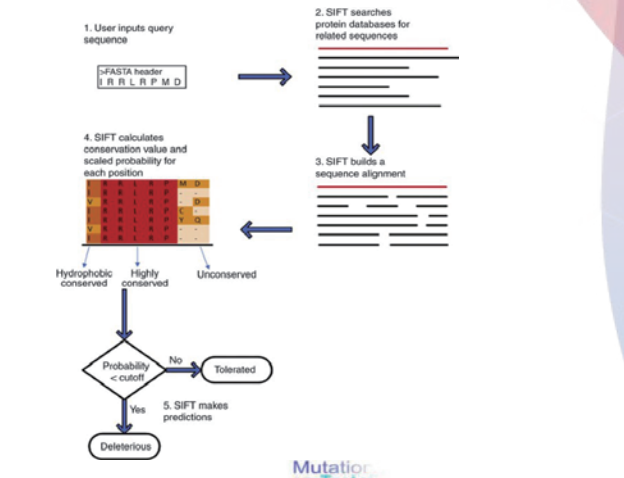
Potential Disease association

32

Variant filtering and prioritization

- Prediction of possible impact of an amino acid substitution on the structure and function of a human protein





Sorting Intolerant From Tolerant | MutationTaster | Human Splicing Finder | PolyPhen-2 | SPIDEX: SPLICING INDEX

Prediction of variant effect (amino acid changes or splicing)

Variant filtering and prioritization

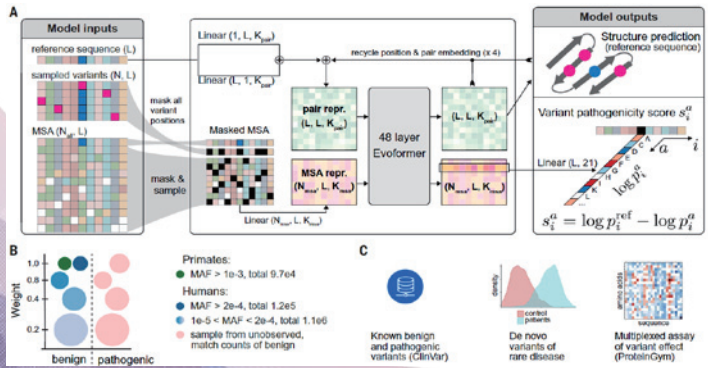
- AlphaMissense

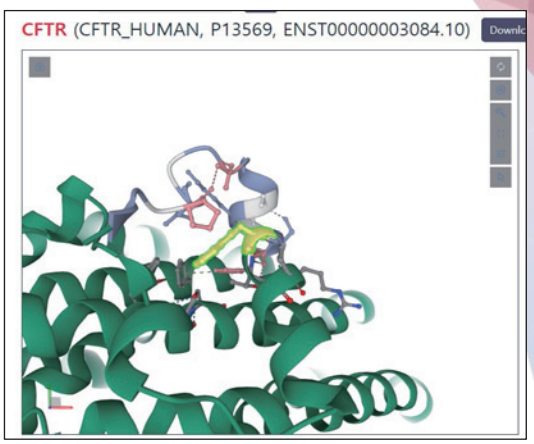
RESEARCH ARTICLE

MACHINE LEARNING

Accurate proteome-wide missense variant effect prediction with AlphaMissense

Jun Cheng*, Guido Novati, Joshua Pan†, Clare Bycroft†, Akvilė Žemgulytė†, Taylor Applebaum†, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli*, Žiga Avsec*

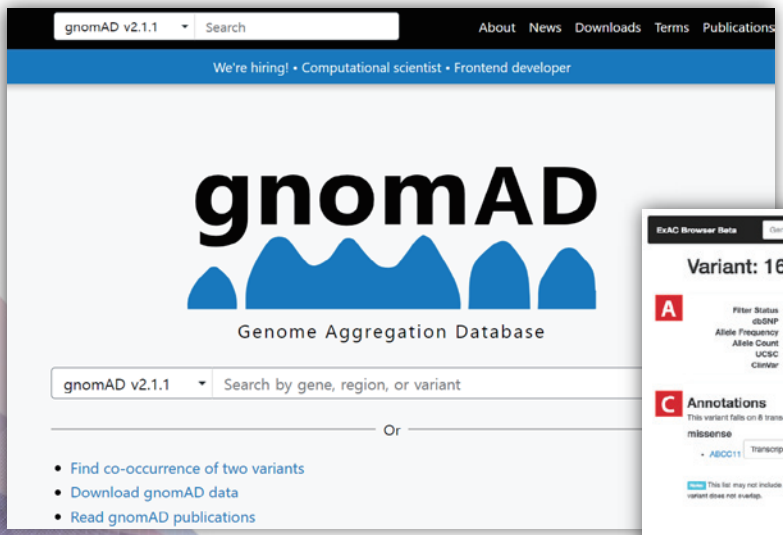


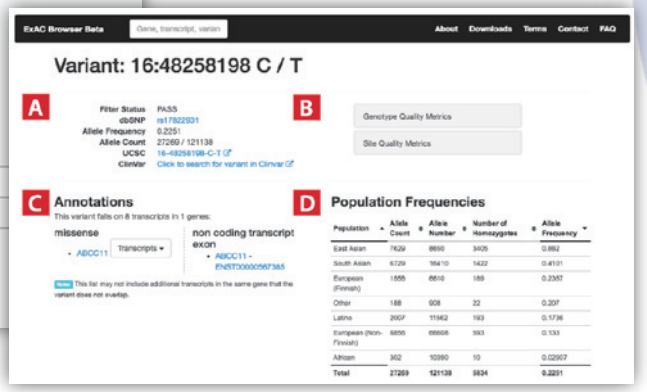


Cheng et al., *Science*, 2023

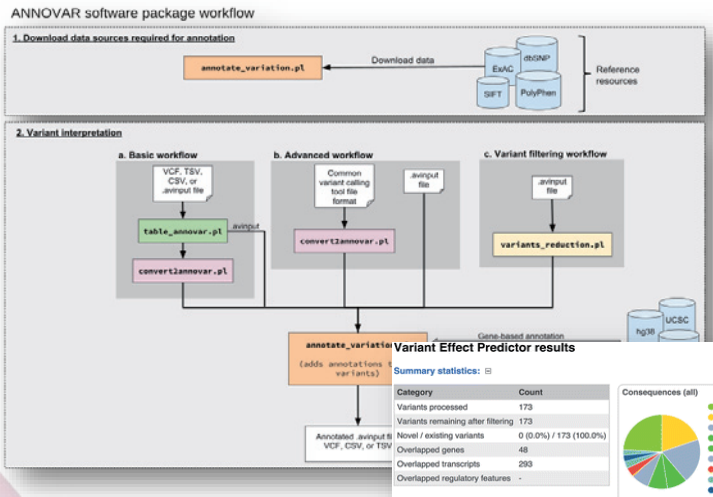
Population allele frequency in normal individuals

- The Genome Aggregation Database (gnomAD)
  - The aggregation and analysis of 125,748 WES and 15,708 WGS data for normal individuals of diverse ancestries





Variant annotation tools

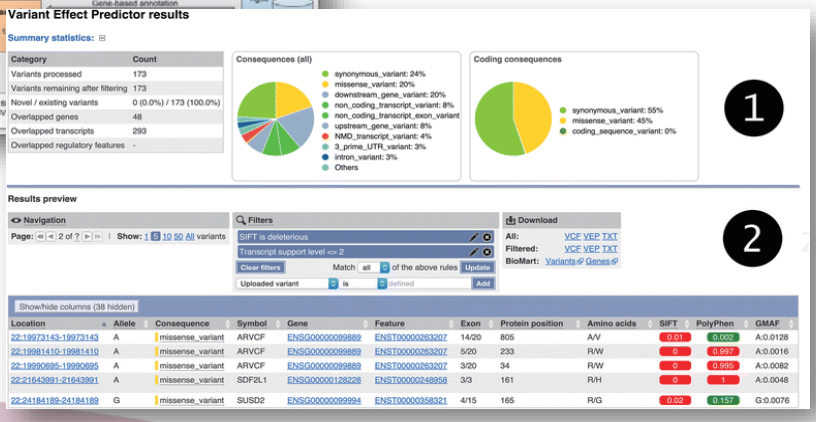


ANNOVAR

Wang et al., *Nucleic Acids Research,* 2010

VEP

McLaren et al., *Genome Biology,* 2016



ACMG guideline

© American College of Medical Genetics and Genomics **ACMG STANDARDS AND GUIDELINES** | Genetics inMedicine

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Richards et al., *Genetics in Medicine,* 2015

Table 3 Criteria for classifying pathogenic variants

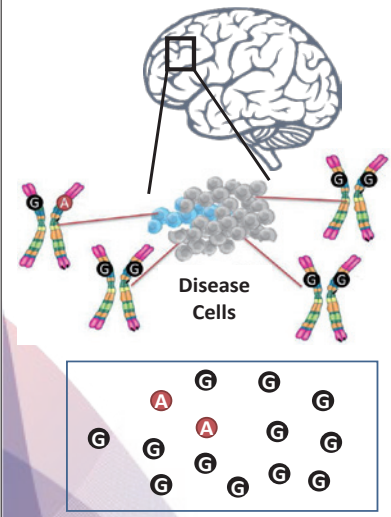
Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP, MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in

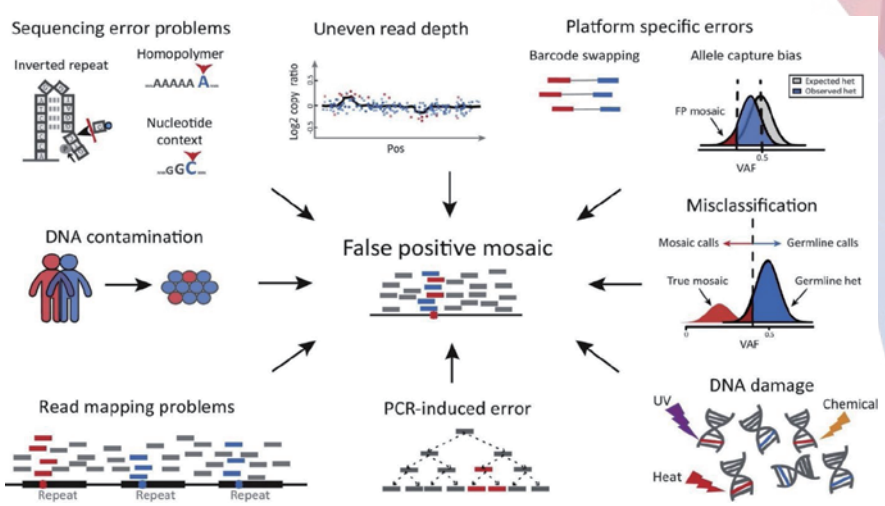
Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)

The detection of somatic mutation is difficult

- We always have false calls – **False Positives & False Negatives**





Dou Y et al, *Trends in Genetics,* 2018

The detection of somatic mutation is difficult

LETTER

doi:10.1038/nature10531

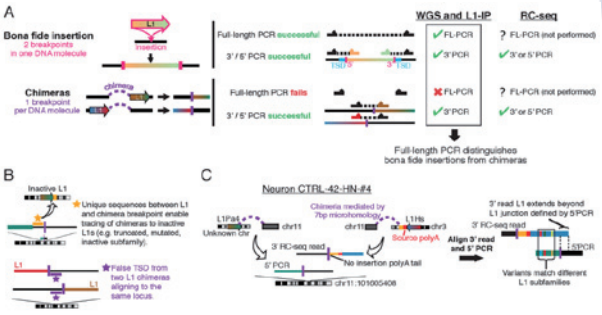
Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie[1]*, Mark W. Barnett[1]*, Kyle R. Upton[1]*, Daniel J. Gerhardt[2], Todd A. Richmond[2], Fioravante De Sapio[1], Paul Brennan[3], Patrizia Rizzu[4], Sarah Smith[1], Mark Fell[1], Richard T. Talbot[1], Stefano Gustincich[5], Thomas C. Freeman[1], John S. Mattick[6], David A. Hume[1], Peter Heutink[4], Piero Carninci[7], Jeffrey A. Jeddeloh[2] & Geoffrey J. Faulkner[1]

Resolving rates of mutation in the brain using single-neuron genomics

Gilad D Evrony[1,2,3,4,5†], Eunjung Lee[6,7†], Peter J Park[6,7*], Christopher A Walsh[1,2,3,4,5*]

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease, Boston Children's Hospital, Boston, United States; [2]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, United States; [3]Department of Neurology, Harvard Medical School, Boston, United States; [4]Department of Pediatrics, Harvard Medical School, Boston, United States; [5]Broad Institute of MIT and Harvard, Cambridge, United States; [6]Department of Biomedical Informatics, Harvard Medical School, Boston, United States; [7]Division of Genetics, Brigham and Women's Hospital, Boston, United States



Evrony GD et al, *eLife,* 2016

Exogenous errors – sample contamination

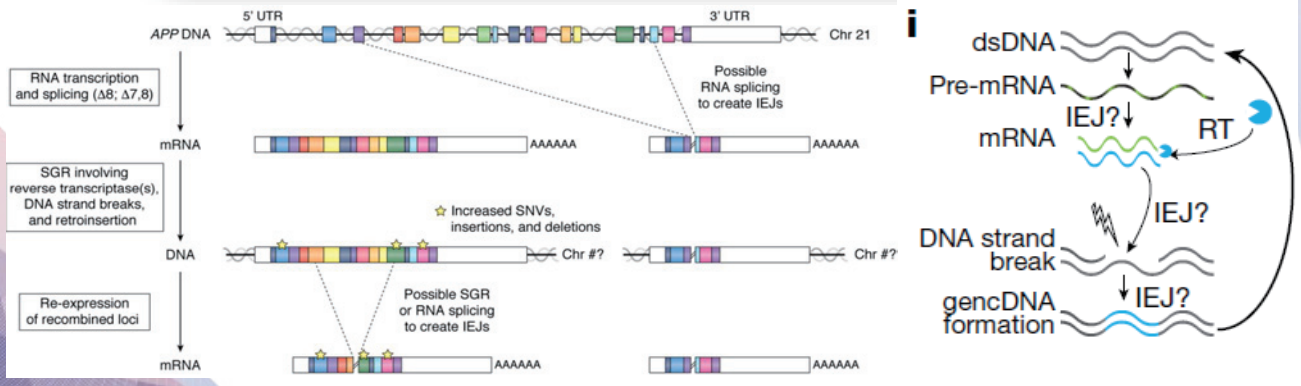
ARTICLE

Corrected: Publisher Correction

https://doi.org/10.1038/s41586-018-0718-6

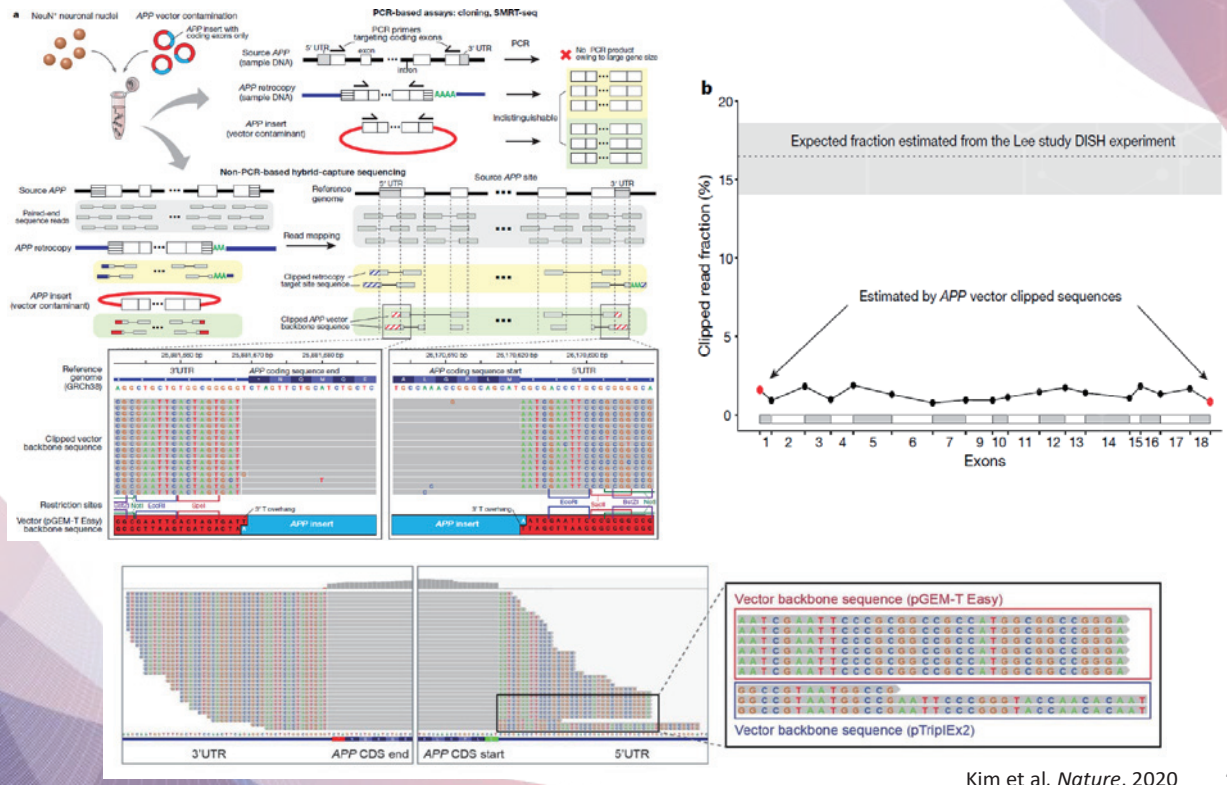
Somatic *APP* gene recombination in Alzheimer's disease and normal neurons

Ming-Hsiang Lee[1], Benjamin Siddoway[1,3], Gwendolyn E. Kaeser[1,2,3], Igor Segota[1,3], Richard Rivera[1], William J. Romanow[1], Christine S. Liu[1,2], Chris Park[1,2], Grace Kennedy[1], Tao Long[1] & Jerold Chun[1]*



Lee et al, *Nature*, 2018

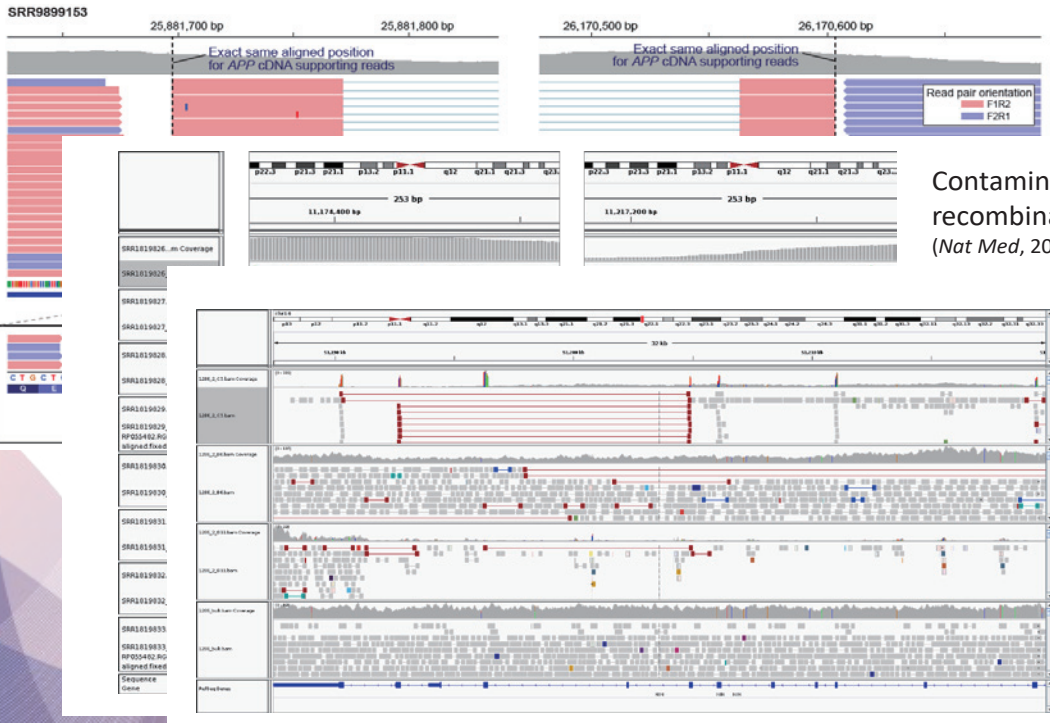
Exogenous errors – sample contamination



Kim et al, *Nature*, 2020

41

Exogenous errors – sample contamination



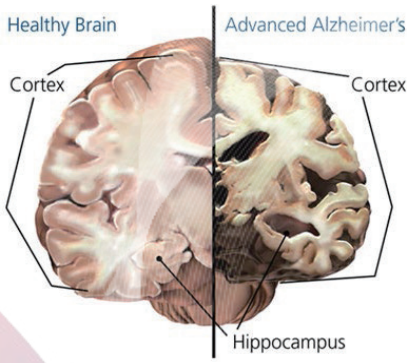
Contaminated by PCR amplicons (*Nature*, 2020)

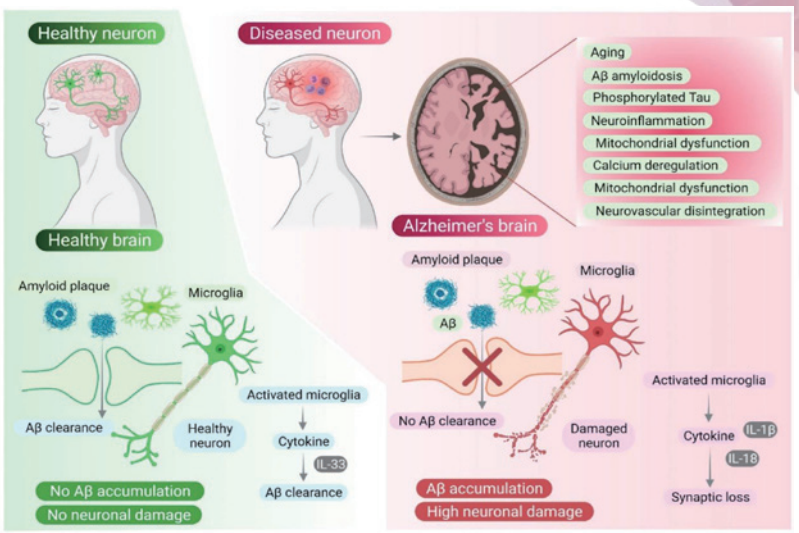
Contaminated by recombinant vector (*Nat Med*, 2015)

Contaminated by mouse mRNA(cDNA) (*Science*, 2018)

42

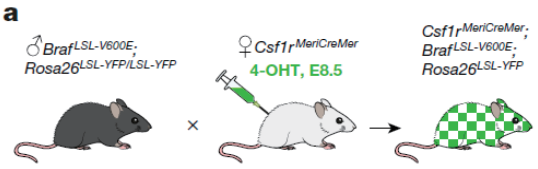
Can somatic mutations affect neurodegenerative diseases?

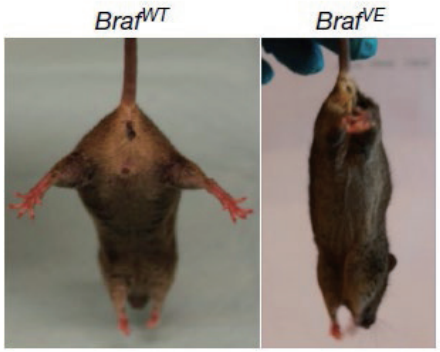


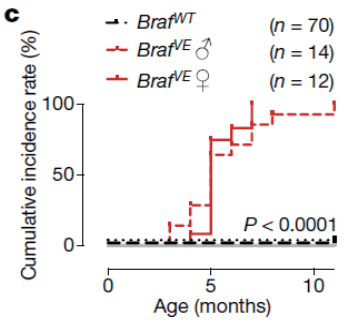


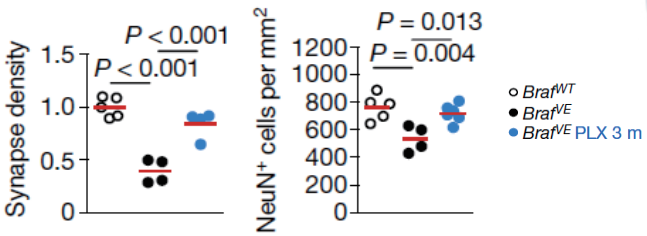
Prasanna et al., *Life*, 2021

Somatic mutations may cause neurodegenerative diseases





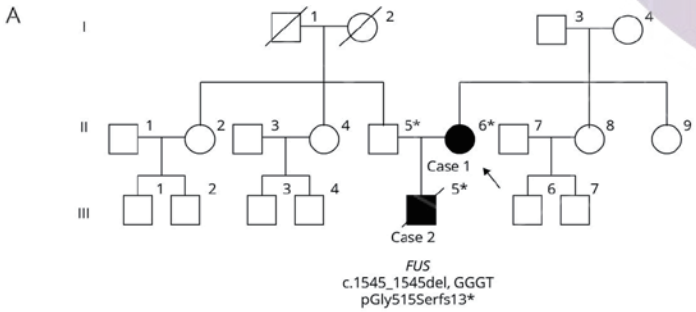


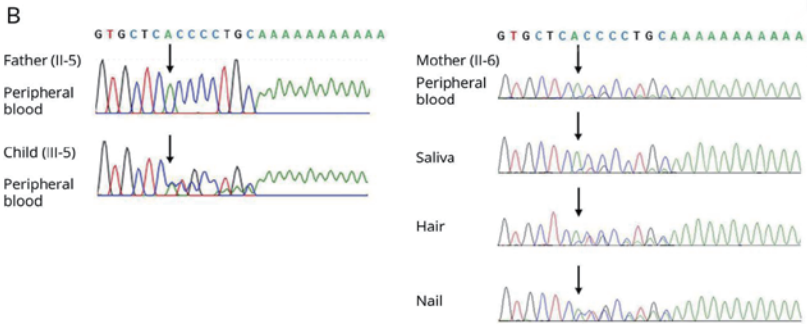


Mass et al, *Nature,* 2017

Possible somatic mosaicism in familial ALS

- The mother had adult-onset ALS
- The FUS (p.Gly515Serfs13*) somatic mutation had ~1% VAFs in blood and saliva samples and higher VAFs in hair and nail samples
- The son inherited the mutation and developed early-onset ALS

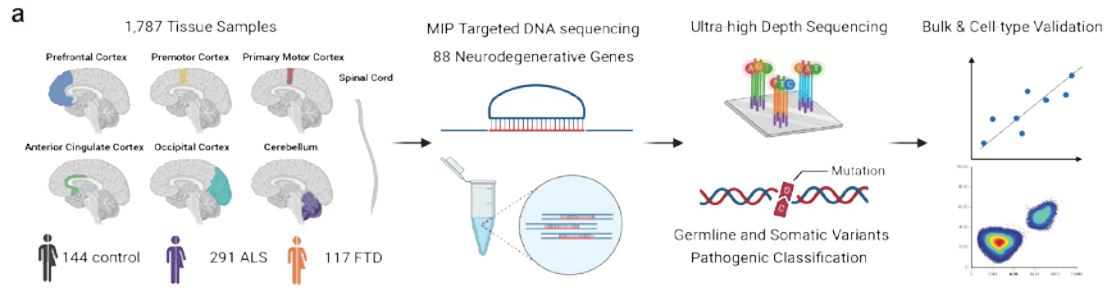


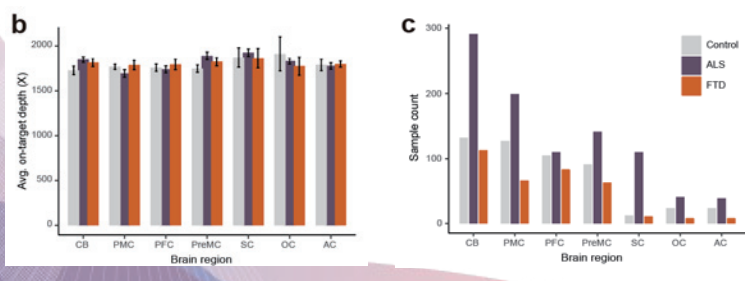


Hisahara et al., *Neurol. Genet.*, 2021

High-depth targeted sequencing of ALS and FTD brains

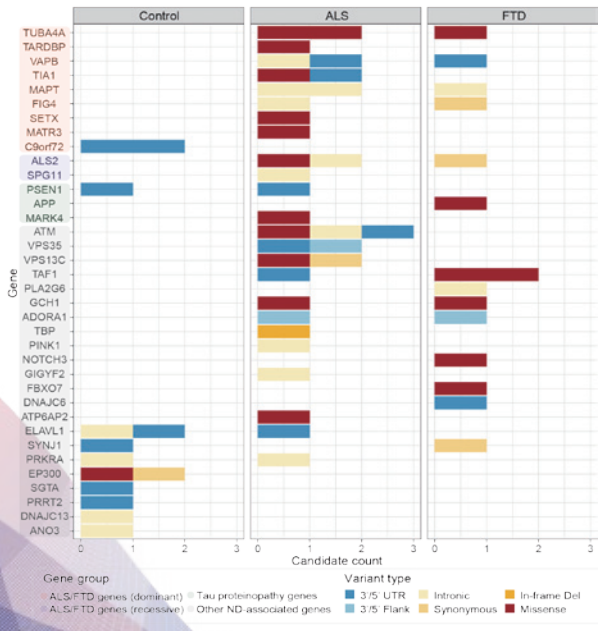
- 1,787 tissue samples from seven different regions
  - Samples from 144 control, 291 ALS, and 113 FTD individuals
  - Avg. ~1,800X coverage for 1,787 samples

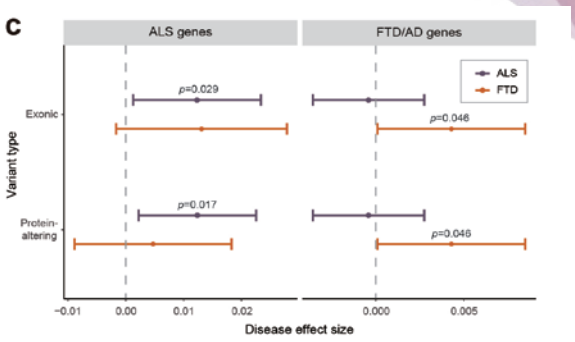


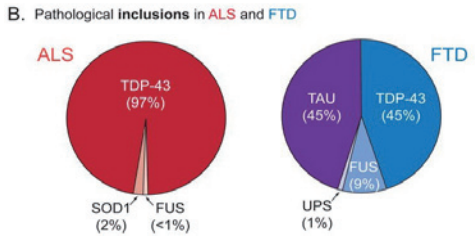


Zhou Z, Kim J et al., unpublished

Damaging somatic mutations are enriched in the related genes of ALS and FTD



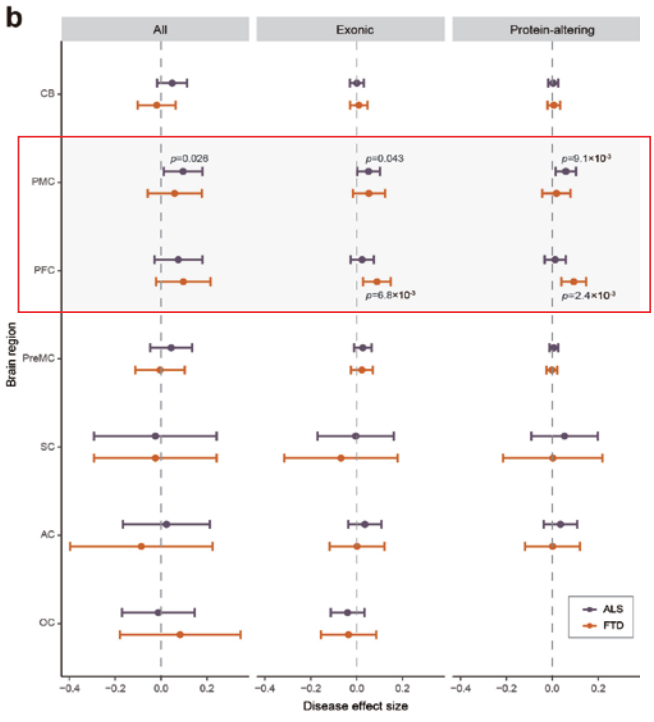


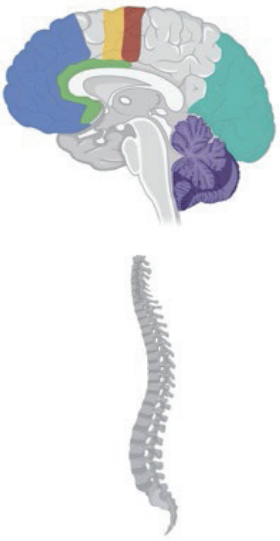


Zhou Z, Kim J et al., unpublished

47

Damaging somatic mutations are enriched in the affected brain regions of ALS and FTD

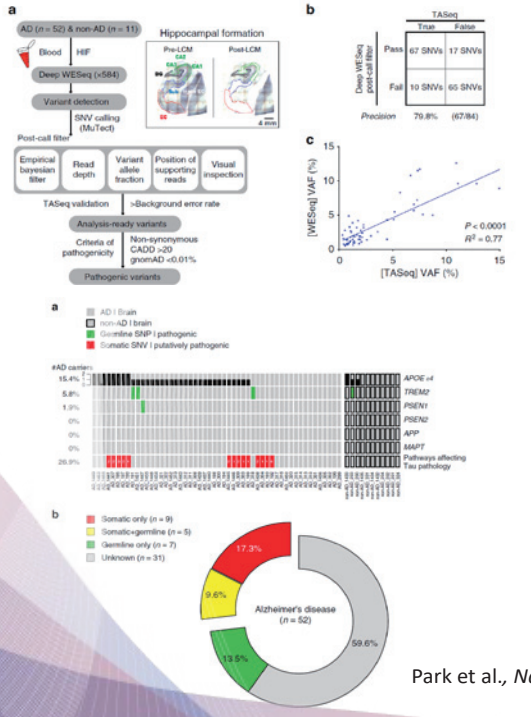




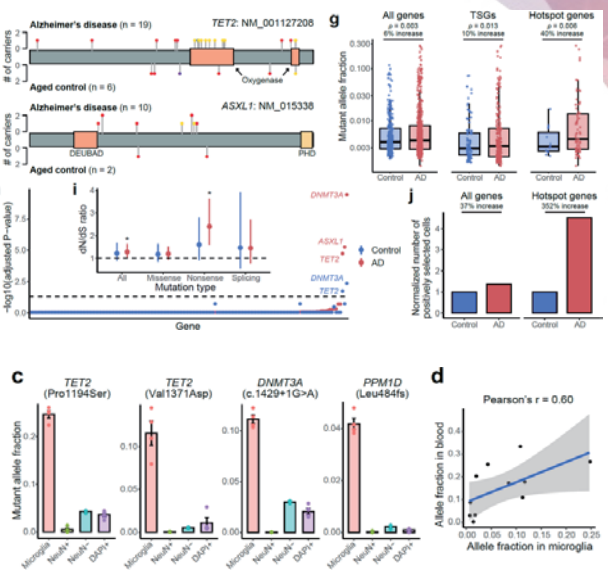
Zhou Z, Kim J et al., unpublished

48

Somatic mutations in Alzheimer's disease

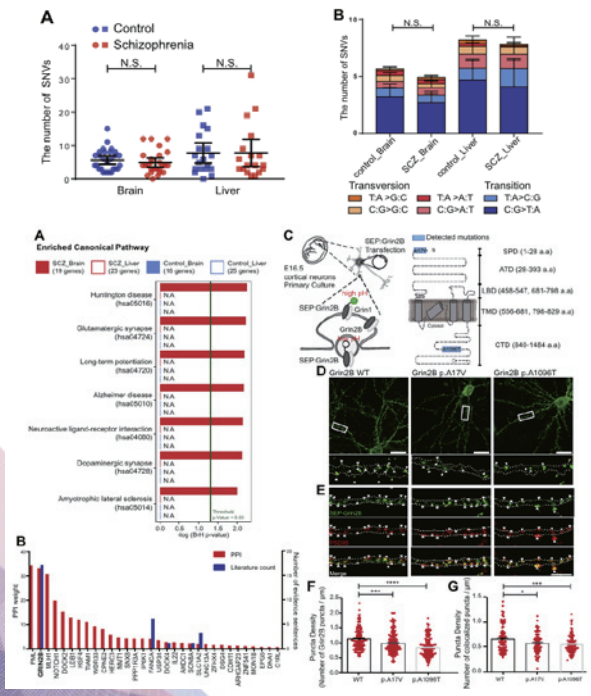


Park et al., *Nat. Comm.*, 2019

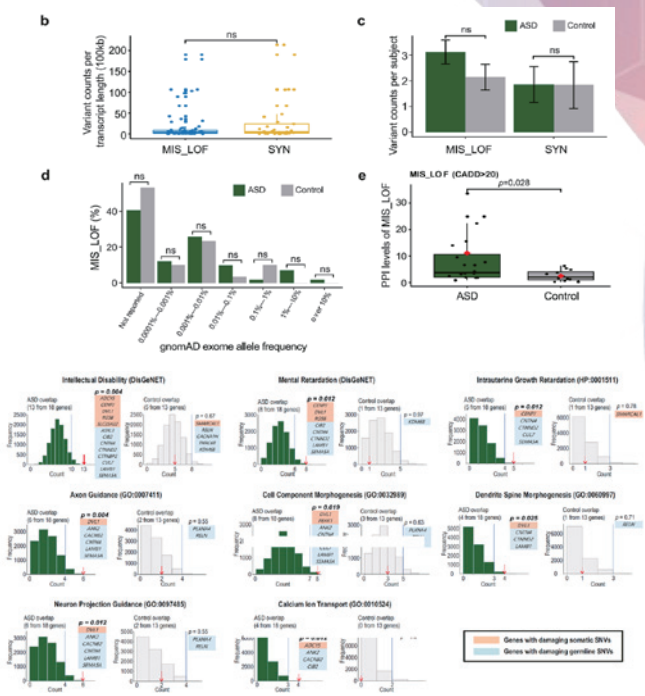


Huang et al., *bioRxiv*, 2024

Somatic mutations in neuropsychiatric diseases

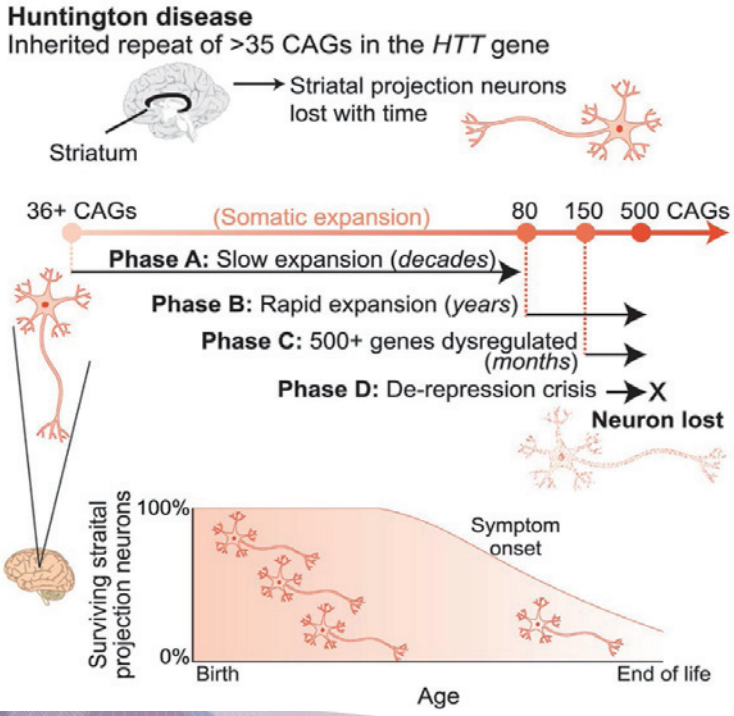


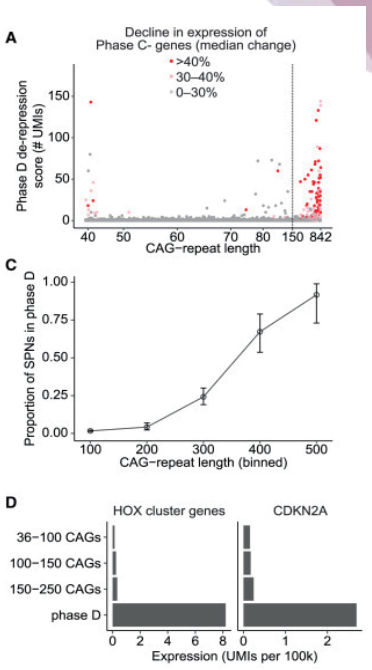
Park et al., *Biol. Psychiatry*, 2021



Kim et al., *Exp. Mol. Med.*, 2024

Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease

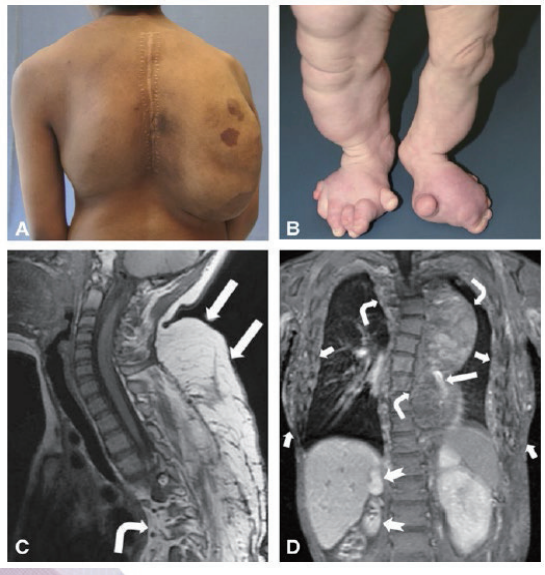


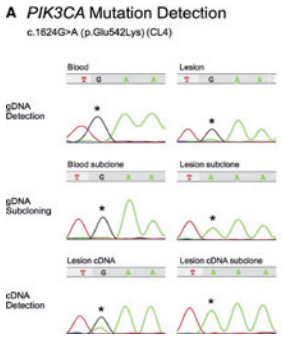


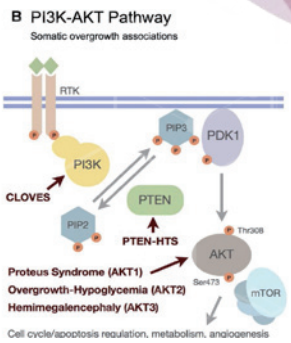
Handsaker et al., *Cell*, 2025

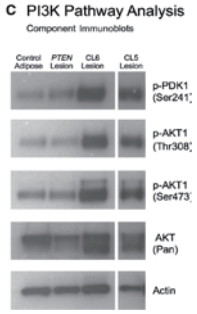
51

Somatic mutations in developmental diseases





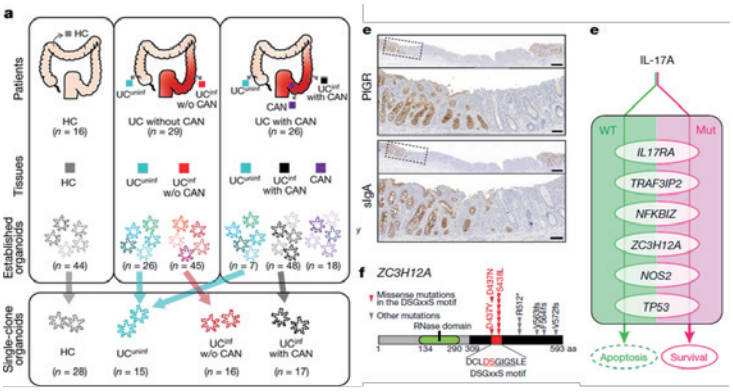




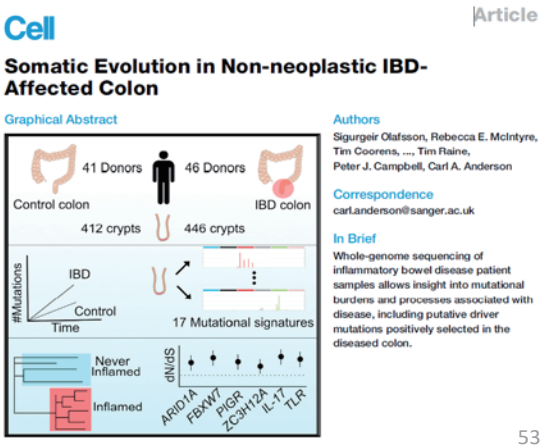
Kurek et al, *AJHG*, 2012

52

Somatic mutations in inflammatory bowel disease (IBD)



Nanki et al, *Nature,* 2020



Olafsson et al, *Cell,* 2020

Somatic mutations in normal human tissue

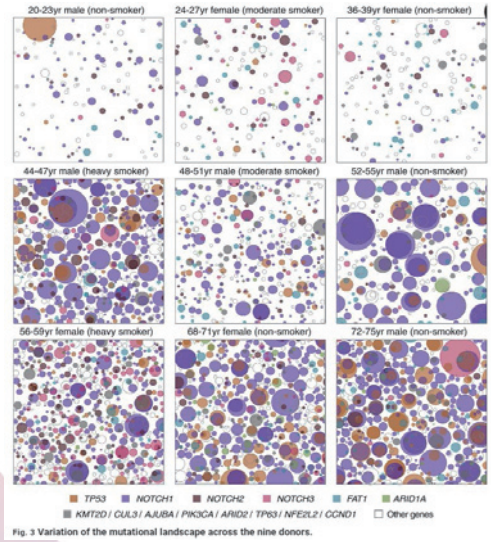
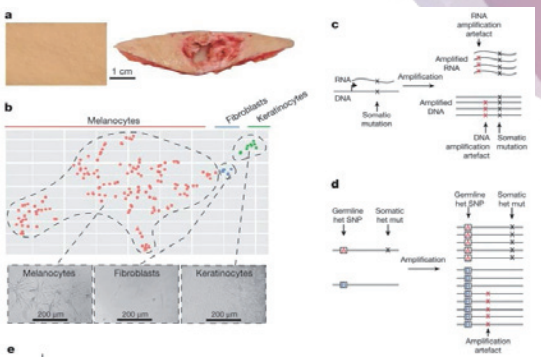
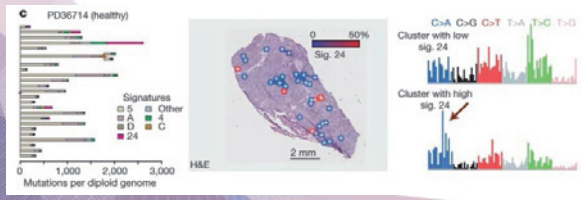


Fig. 3 Variation of the mutational landscape across the nine donors.

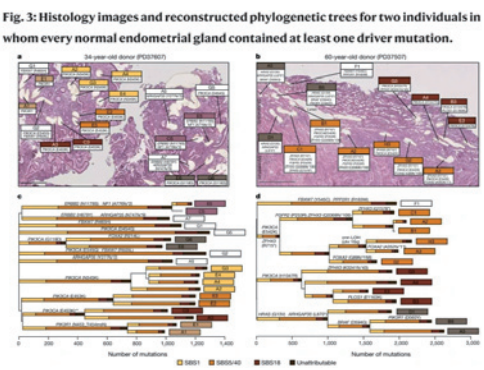
Human esophagus – Martincorena et al, *Science*, 2018



Melanocytes from human skin - Shain et al, *Nature,* 2020

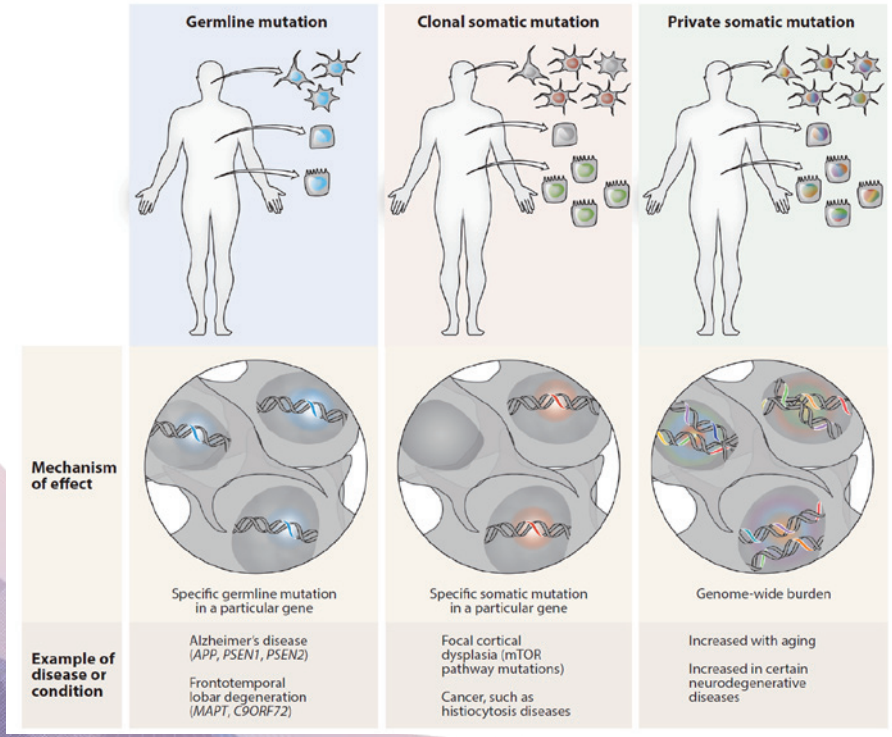


Healthy human liver - Brunner et al, *Nature,* 2019



Healthy human endometrial gland - Moore et al, *Nature,* 2020

Clonal vs. Private somatic mutations

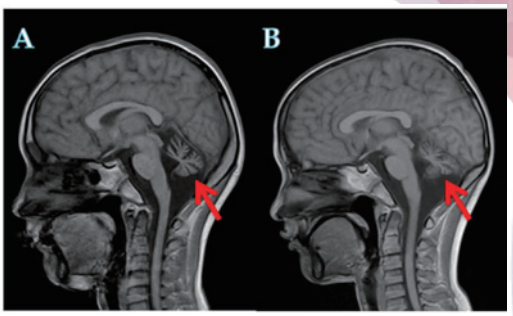


Miller et al,.
Annu. Rev. Genom. Hum. Genet., 2021

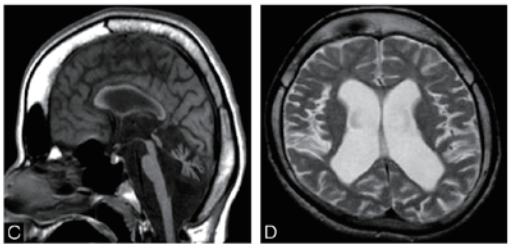
55

Defective DNA repair and neurodegeneration

- Many congenital diseases with premature aging and neurodegeneration involve defective DNA repair
  - Defective DNA double-strand break (DSB) repair in Ataxia Telangiectasia (AT)
  - Defective nucleotide excision repair (NER) in Cockayne Syndrome (CS) and Xeroderma Pigmentosum (XP)



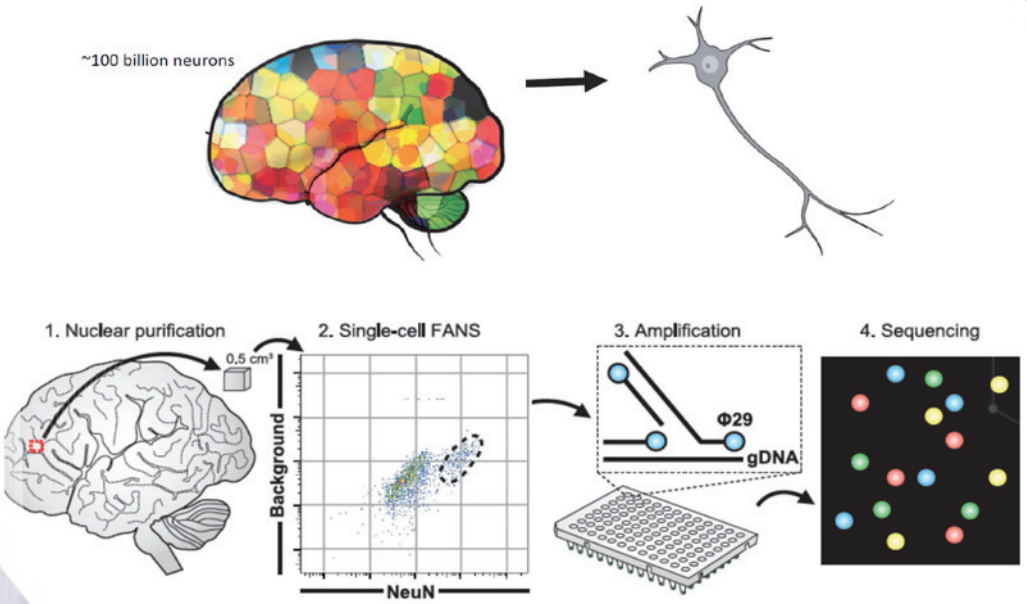
Ataxia Telangiectasia (AT)



Cockayne Syndrome (CS)

56

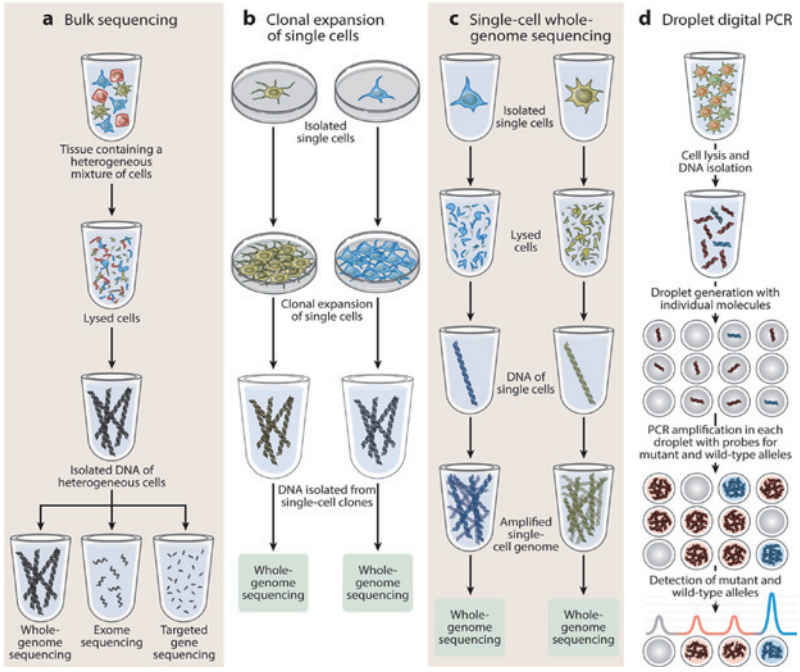
Single-cell whole-genome sequencing (scWGS)



Lodato et al., *Science*, 2015

57

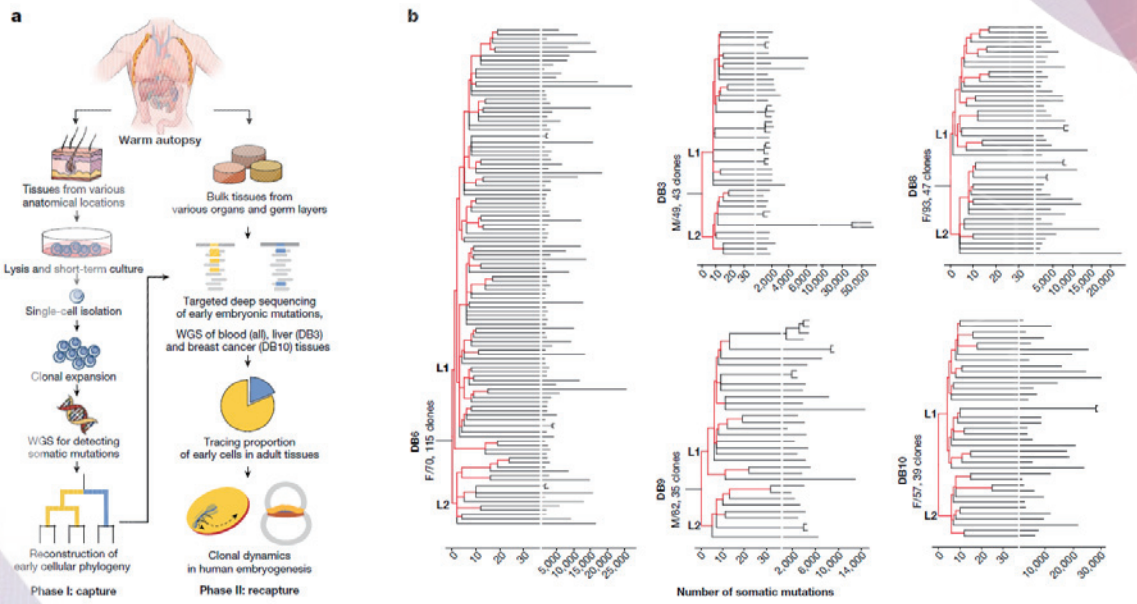
Methods for detecting somatic mutations in a single cell



Miller et al,.
Annu. Rev. Genom. Hum. Genet., 2021

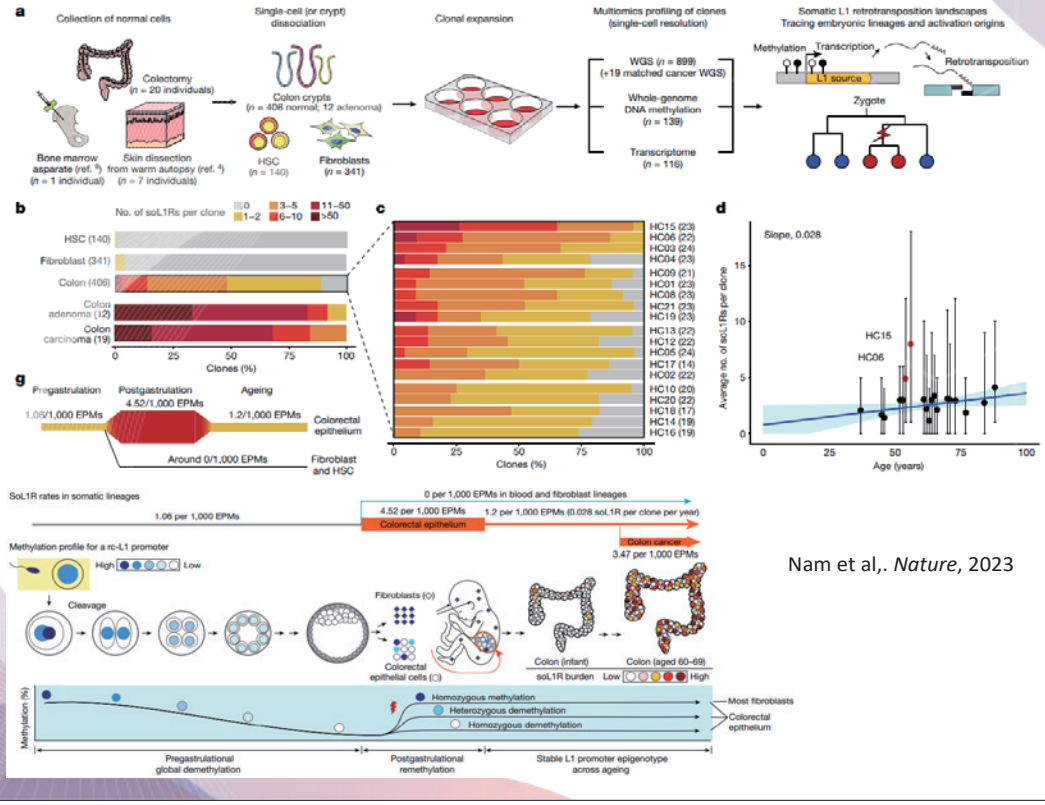
58

Tracing early cellular phylogenies using somatic mutations



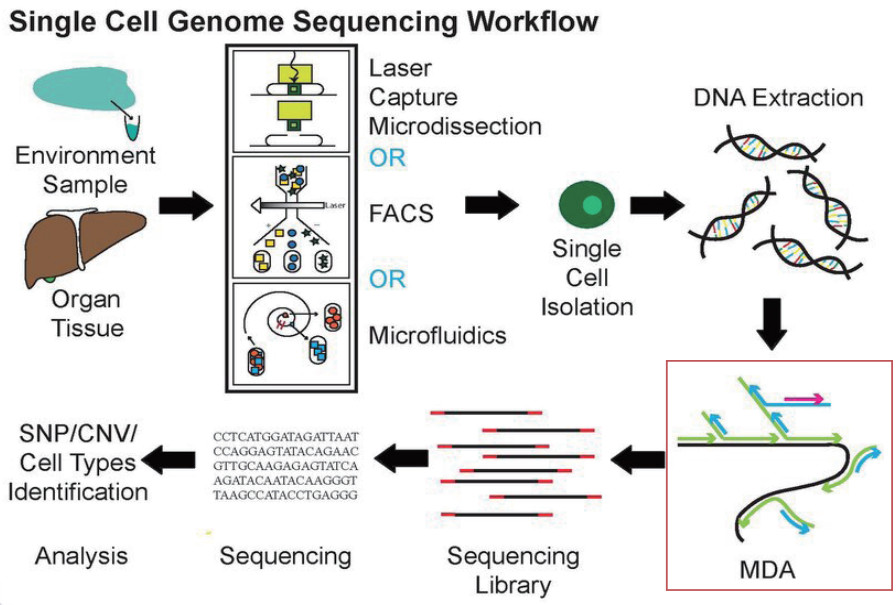
Park et al,. *Nature*, 2021

Somatic retrotransposition in normal colorectal epithelium



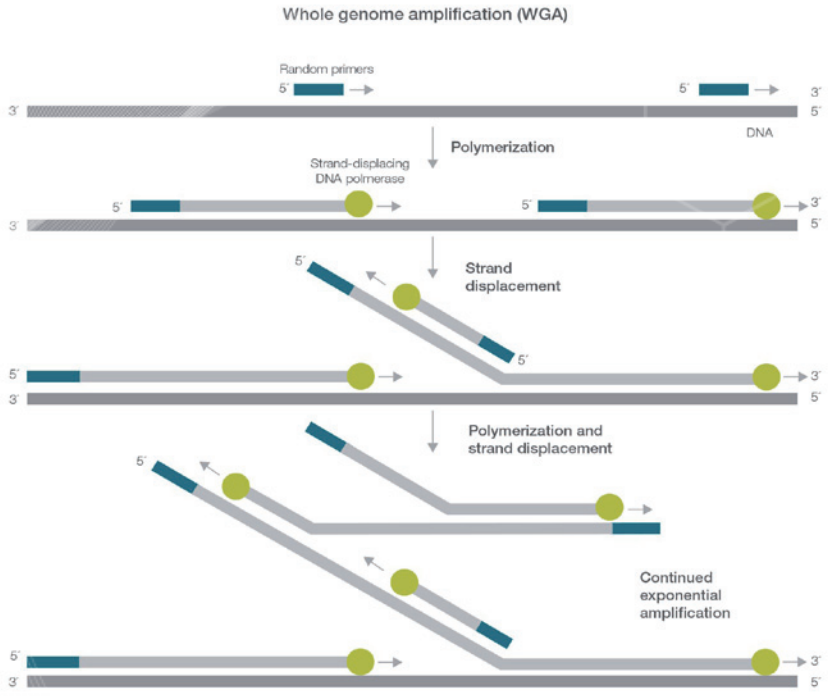
Nam et al,. *Nature*, 2023

Whole-genome amplification



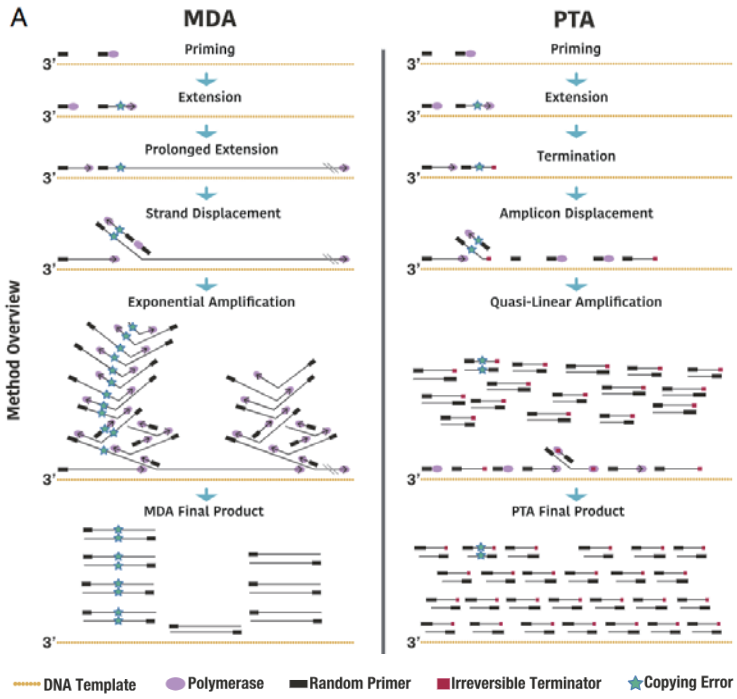
61

Multiple displacement amplification (MDA)



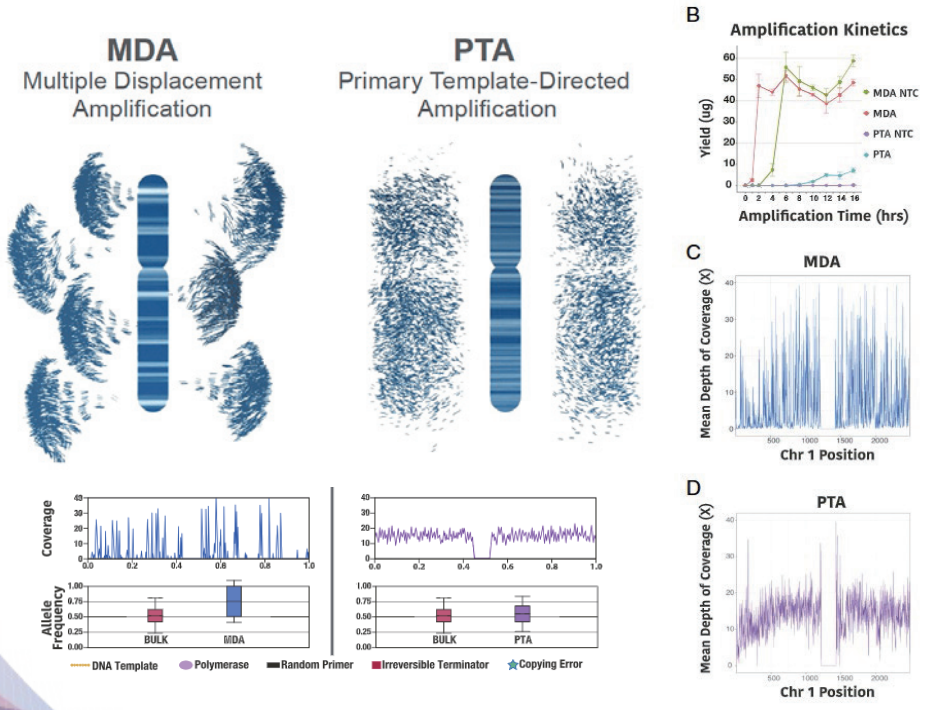
62

Primary template-directed amplification (PTA)



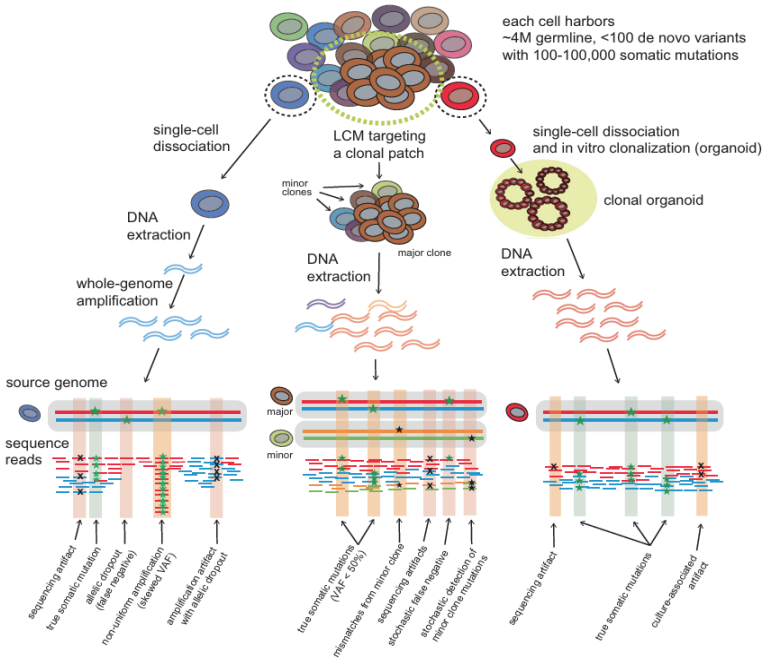
Gonzalez-Pena et al,. *PNAS*, 2021

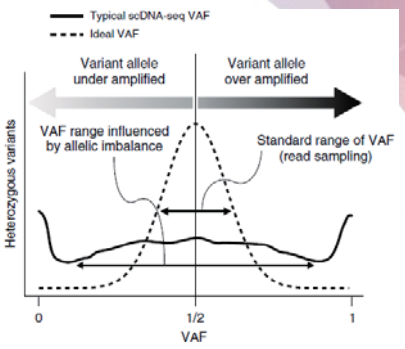
Primary template-directed amplification (PTA)

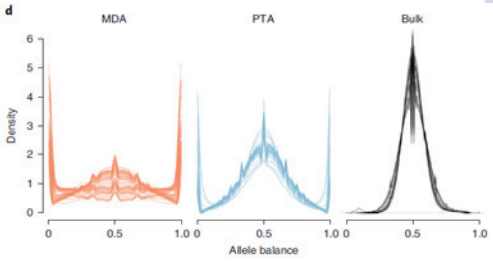


Gonzalez-Pena et al,. *PNAS*, 2021

Huge errors from whole-genome amplification





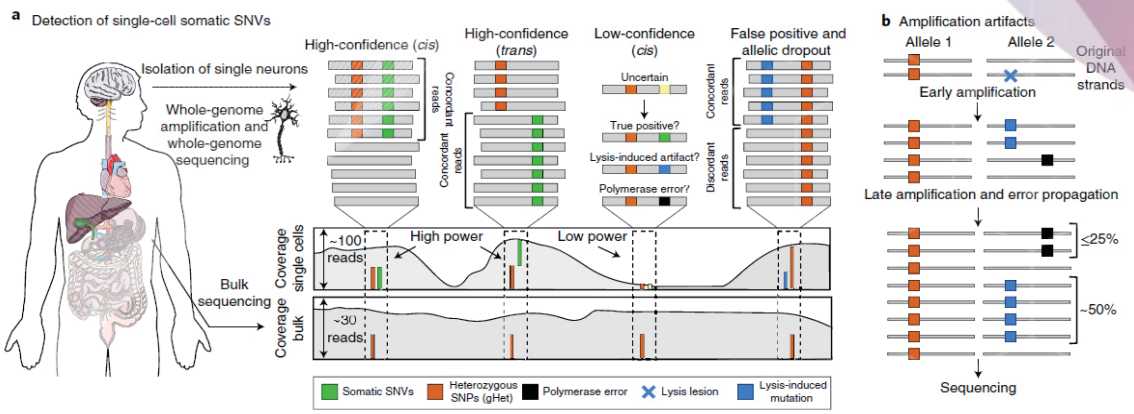


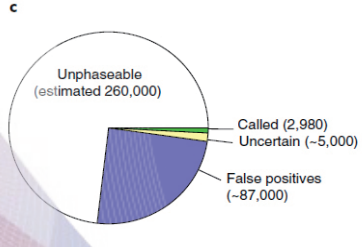
Youk et al., *Exp. Mol. Med.*, 2021

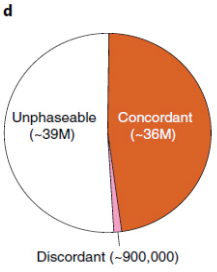
Luquette et al., *Nat. Genet.*, 2022

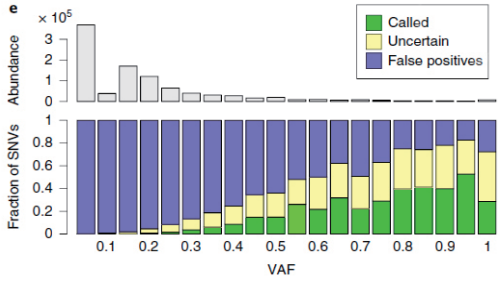
65

SNV detection with linkage-read analysis





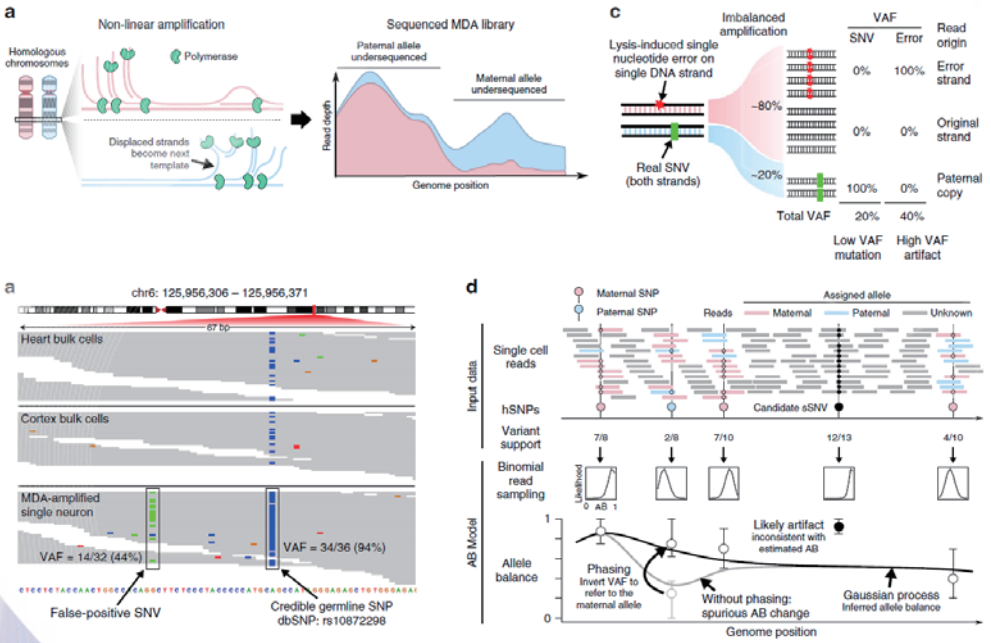




Bohrson et al., *Nat. Genet.*, 2019

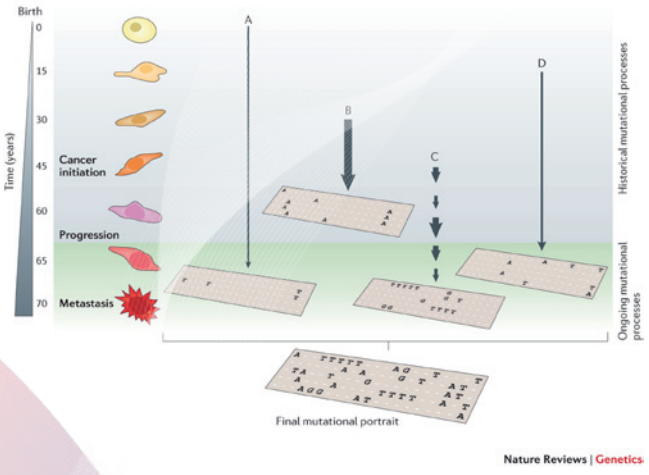
66

SNV detection with allelic imbalance modeling

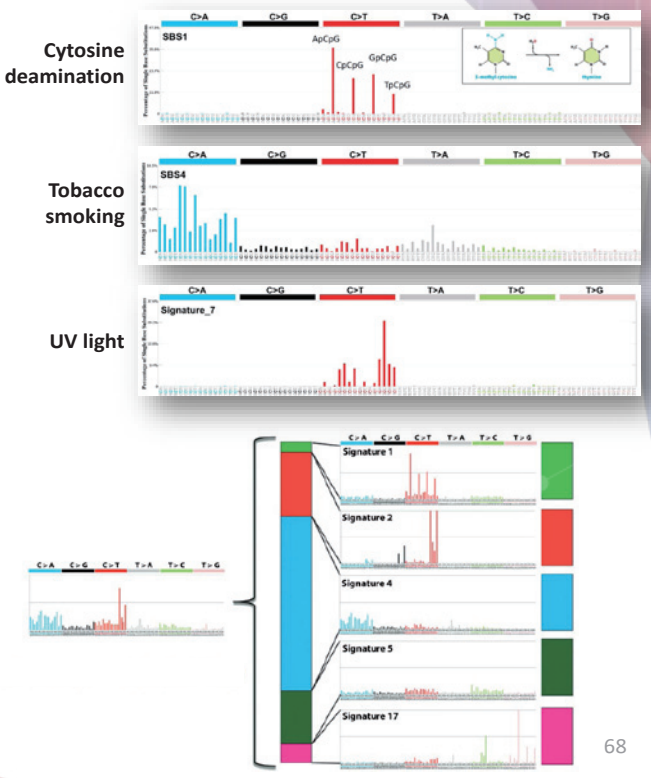


Luquette et al., *Nat. Comm.*, 2019

Mutation signature analysis

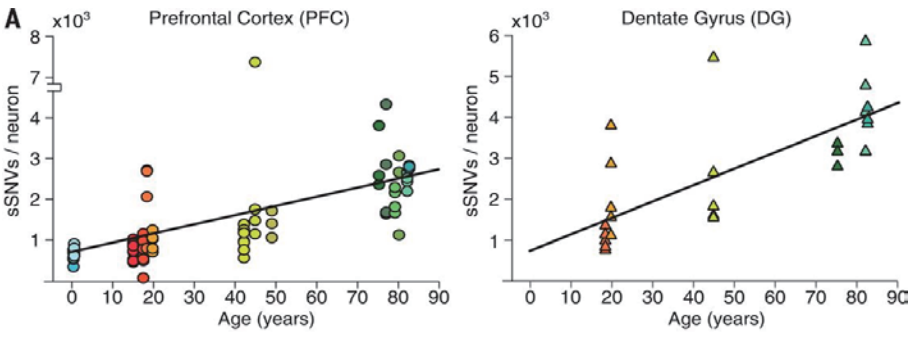


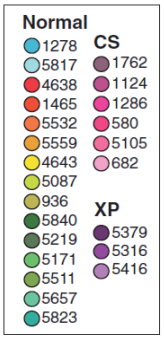
Helleday et al., *Nat. Rev. Genet.*, 2014

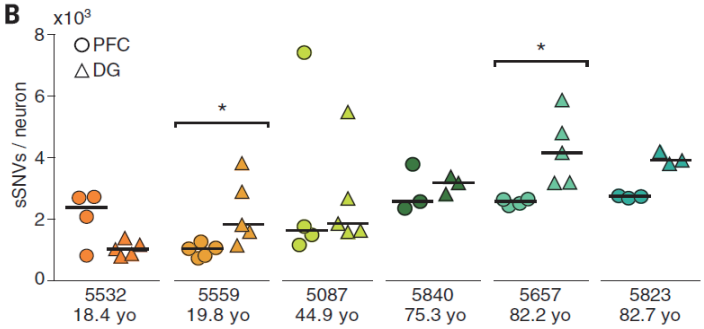


Mutation signatures of human cancer

Somatic SNVs increase with age in single neurons

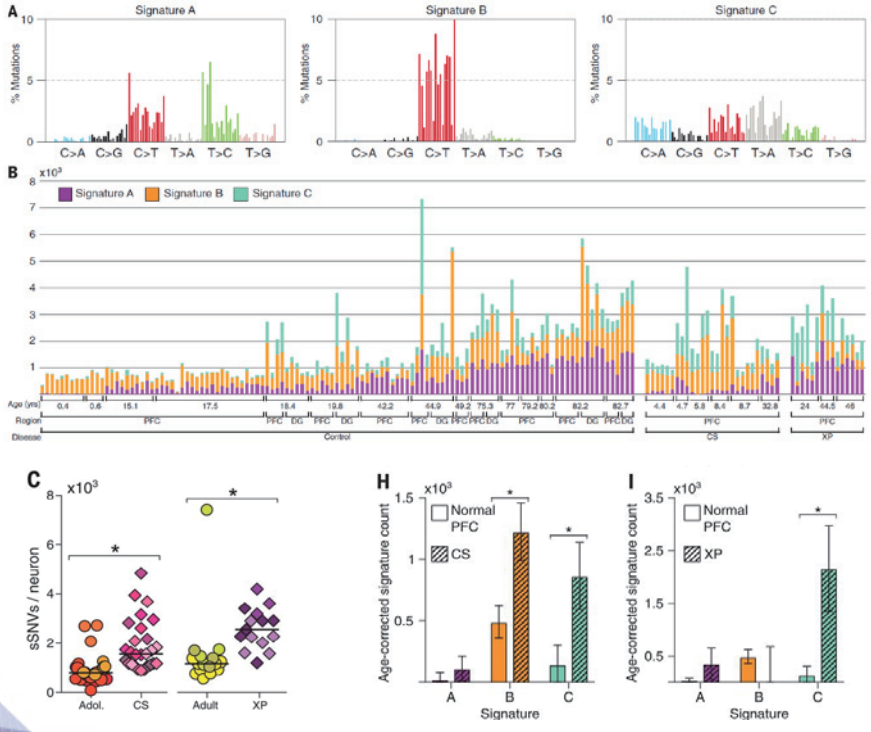






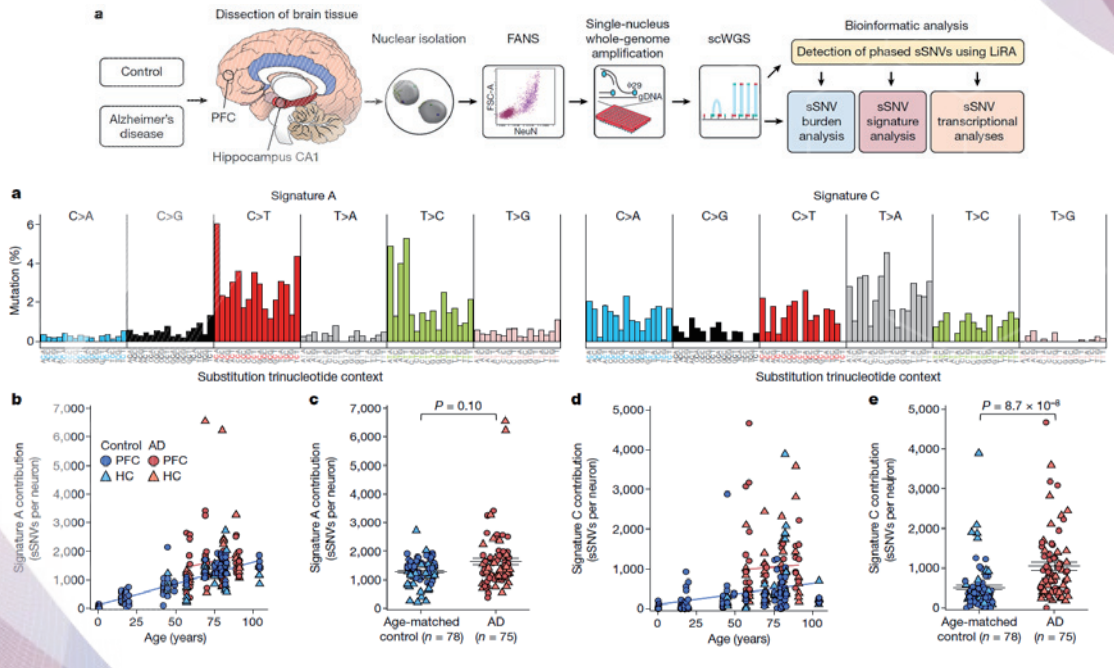
Lodato et al., *Science*, 2018

Somatic SNVs are elevated in CS and XP



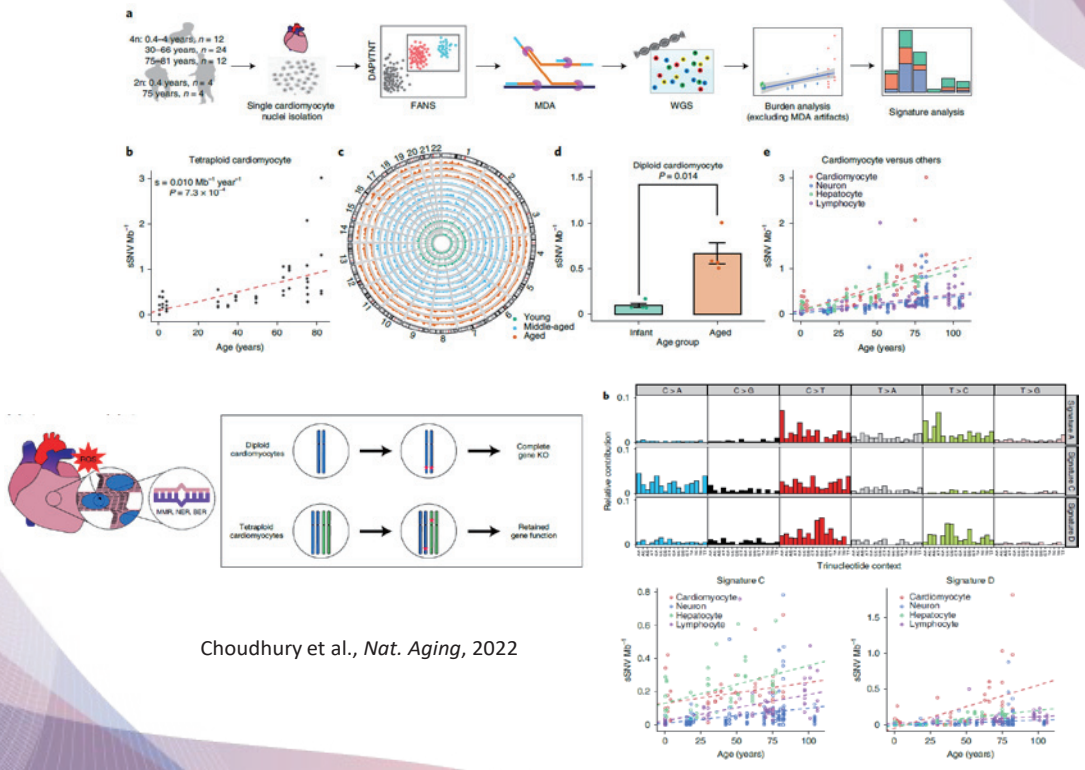
Lodato et al., *Science*, 2018

Somatic SNVs in Alzheimer's disease



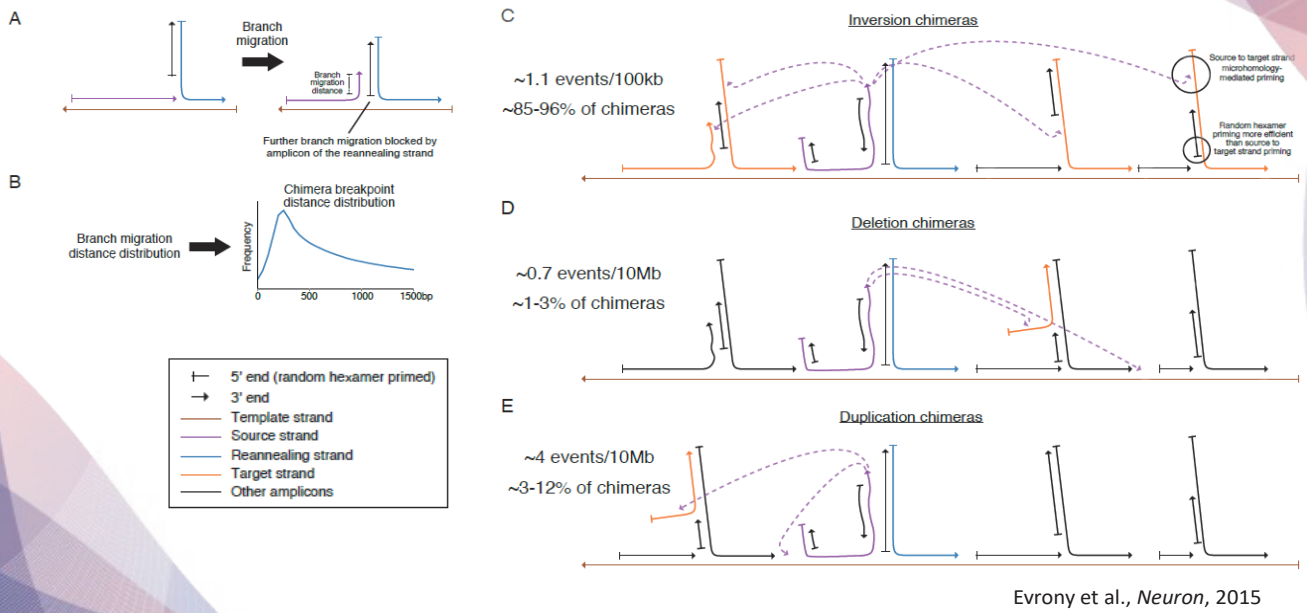
Miller et al., *Nature*, 2022

Somatic SNVs in cardiomyocytes



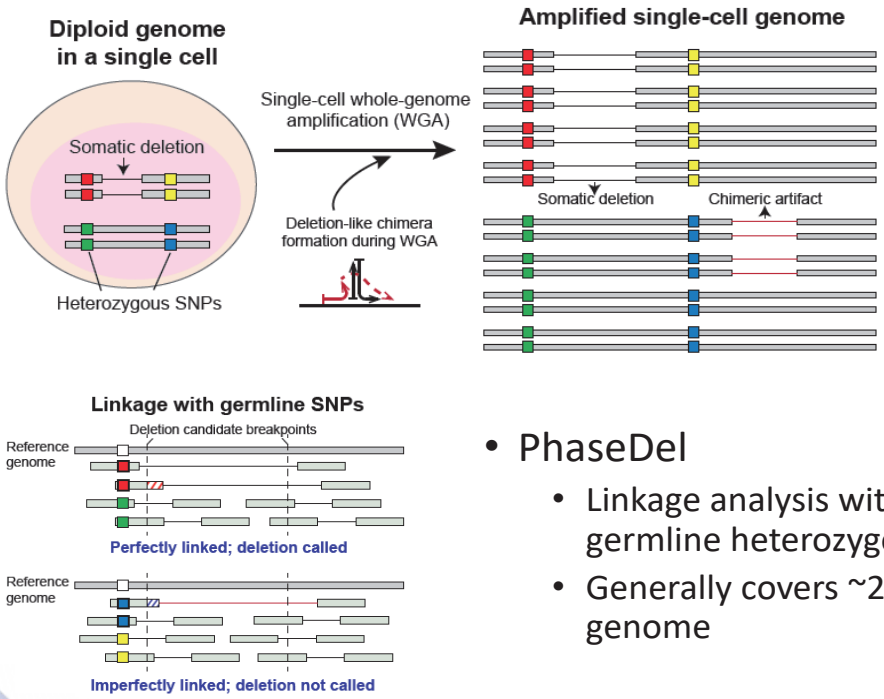
Choudhury et al., *Nat. Aging*, 2022

Chimeric artifacts hinder structural variation (SV) detection



Evrony et al., *Neuron*, 2015

Identification of somatic SVs using linkage information



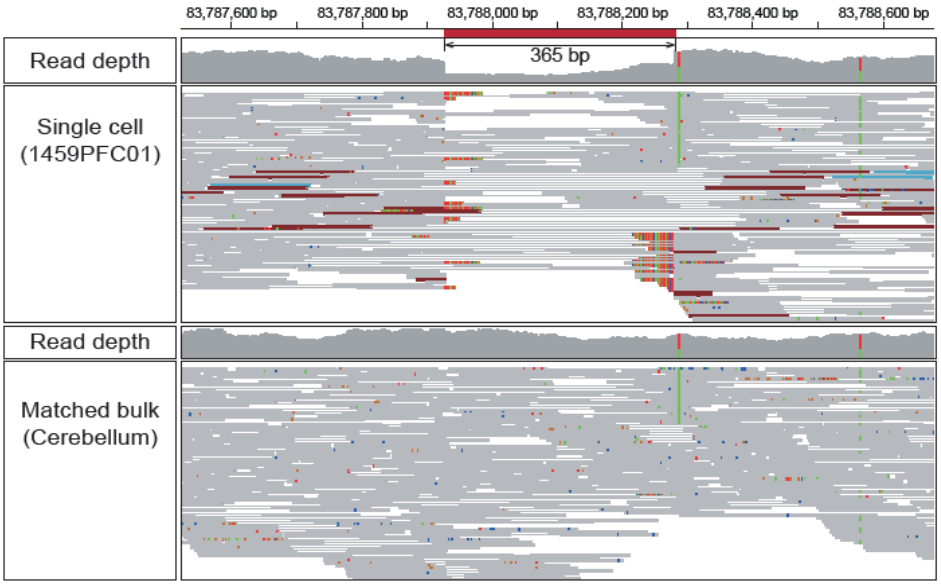
- PhaseDel
  - Linkage analysis with nearby germline heterozygous SNPs
  - Generally covers ~25% of the genome

Kim et al., *Nat. Comm.*, 2022

75

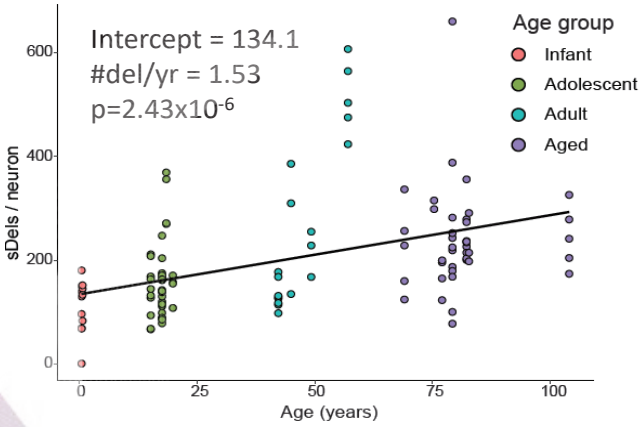
Somatic deletion in scWGS data

"*Nano*" deletion

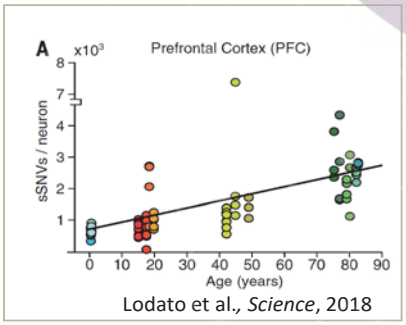


76

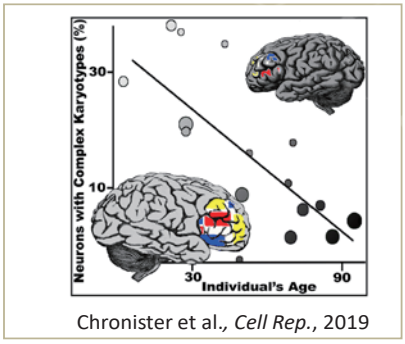
Somatic deletions accumulate with age



Kim et al., *Nat. Comm.*, 2022



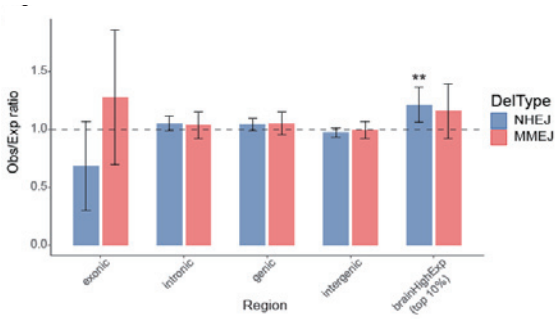
Lodato et al., *Science*, 2018

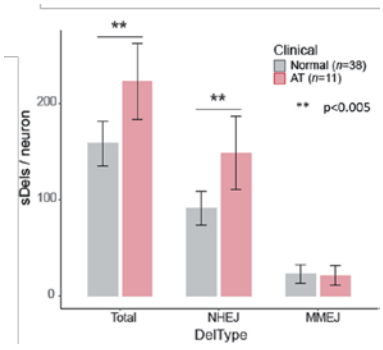


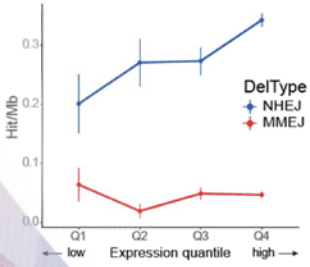
Chronister et al., *Cell Rep.*, 2019

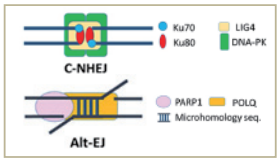
77

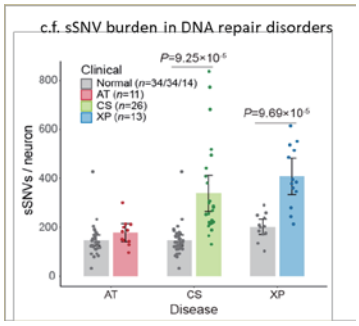
Somatic deletions increases with gene expression and in DNA repair disorders







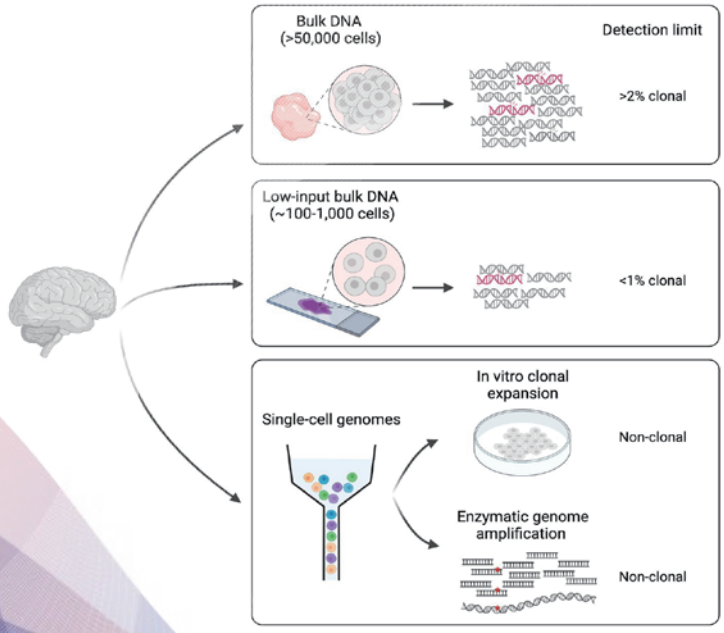




Kim et al., *Nat. Comm.*, 2022

78

The human brain through the lens of somatic mosaicism



- Two different approaches to study somatic mutations
  - Analysis of low-level clonal somatic mutations in bulk tissue
  - Somatic mutation analysis in single cell

Bizzotto et al. , *Front. Neurosci.*, 2023

79

Thank you

80