Human *de novo* mutation rates from a four-generation pedigree reference

孙砚青

浙江大学生命演化研究中心

Introduction

浙江大学 生命演化研究中心

Article

Human de novo mutation rates from a four-generation pedigree reference

https://doi.org/10.1038/s41586-025-08922-2
Received: 5 August 2024
Accepted: 20 March 2025
Published online: 23 April 2025
Open access
Check for updates

David Porubsky[1], Harriet Dashnow[2,3,22], Thomas A. Sasani[2,22], Glennis A. Logsdon[1,20,22], Pille Hallast[4,22], Michelle D. Noyes[1,22], Zev N. Kronenberg[5,22], Tom Mokveld[5,22], Nidhi Koundinya[1], Cillian Nolan[5], Cody J. Steely[2,6], Andrea Guarracino[7], Egor Dolzhenko[5], William T. Harvey[1], William J. Rowell[5], Kirill Grigorev[8,9], Thomas J. Nicholas[2], Michael E. Goldberg[2], Keisuke K. Oshima[10], Jiadong Lin[1], Peter Ebert[11,12], W. Scott Watkins[2], Tiffany Y. Leung[13], Vincent C. T. Hanlon[13], Sean McGee[1], Brent S. Pedersen[2], Hannah C. Happ[2], Hyeonsoo Jeong[1,21], Katherine M. Munson[1], Kendra Hoekzema[1], Daniel D. Chan[13], Yanni Wang[13], Jordan Knuth[1], Gage H. Garcia[1], Cairbre Fanslow[5], Christine Lambert[5], Charles Lee[4], Joshua D. Smith[1], Shawn Levy[14], Christopher E. Mason[15,16,17], Erik Garrison[7], Peter M. Lansdorp[13,18], Deborah W. Neklason[2], Lynn B. Jorde[2], Aaron R. Quinlan[2], Michael A. Eberle[5] & Evan E. Eichler[1,19] ✉

First & Co-first Authors:
David Porubsky, Harriet Dashnow, Thomas A. Sasani, Glennis A. Logsdon, et al. (7 authors)

Corresponding Author:
Evan E. Eichler

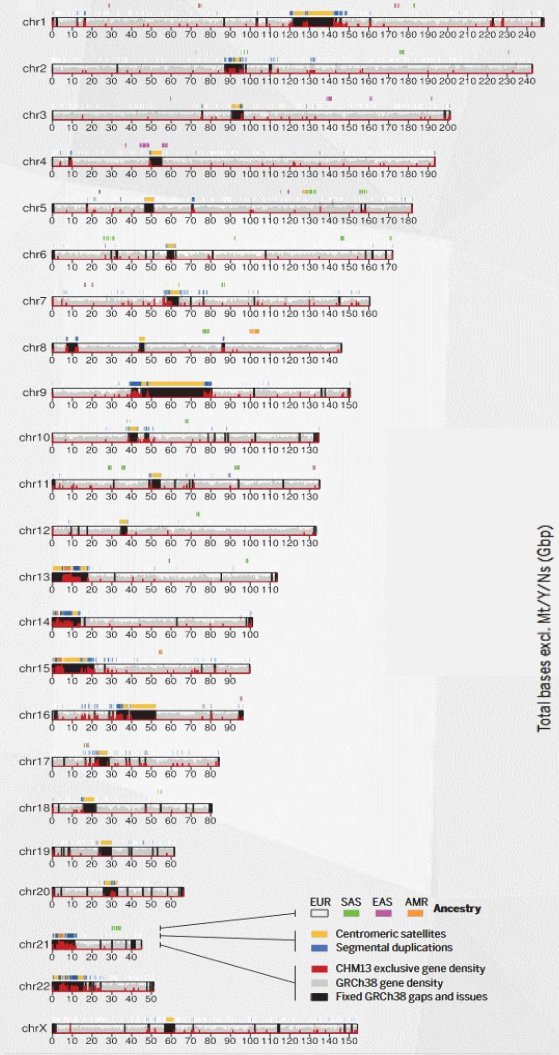
Lead Affiliations:
Department of Genome Sciences, University of Washington School of Medicine

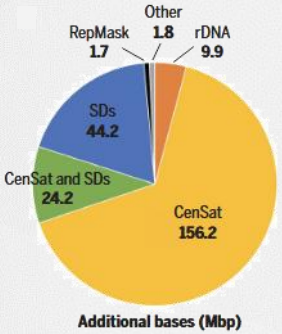
Outline

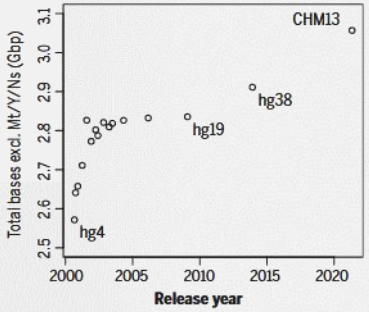
- Background
- Sequencing and assembly
- Main results
  - *De novo* SNVs and small indels
  - Hypermutable Regions
    - Tandem Repeats (TRs)
    - Centromeres
    - Y chromosome
  - Sequence resolved recombination map
- Summary & Discussion

Background

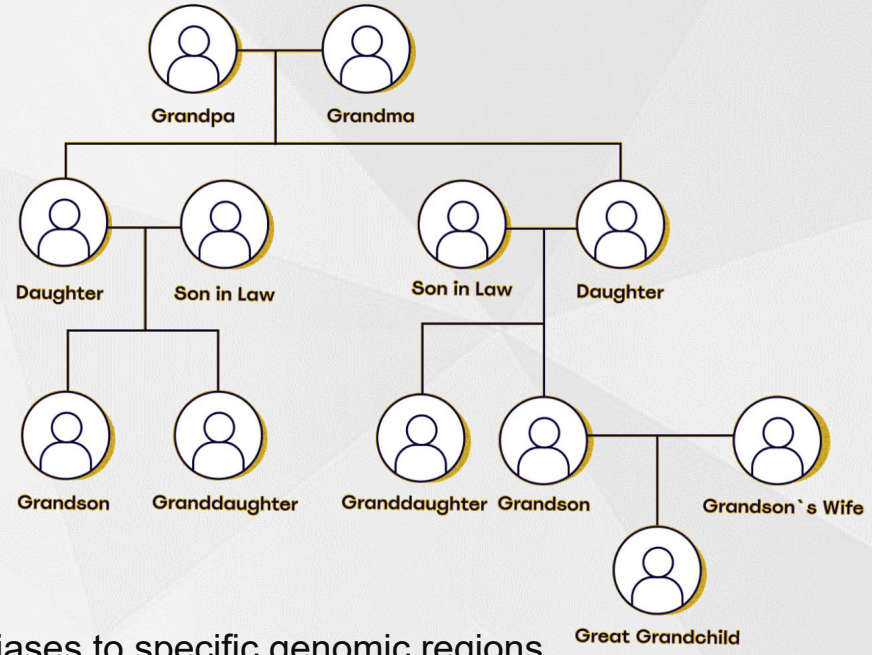
- T2T-CHM13 filled 8% of repeat-rich regions.







- Study human ***de novo* mutation (DNM)** and **recombination** by fully assembled and phased assemblies



- ✖ biases to specific genomic regions
- ✖ biases to classes of genetic variation
- ✖ reference genome effects

Background

- Four-generation, 28-member family CEPH 1463 has been intensively studied over the past three decades.

1990,CEPH COLLABORATIVE

PROGRAM DESCRIPTION

Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome

JEAN DAUSSET,* HOWARD CANN,* DANIEL COHEN,* MARK LATHROP,* JEAN-MARC LALOUEL,† AND RAY WHITE†,[1]

Centre d'Etude du Polymorphisme Humain (CEPH), 27 rue Juliette Dodu, 75010 Paris, France; and †Howard Hughes Medical Institute and Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, Utah 84132

Elsewhere in this issue of *Genomics* is the first consortium map, that of chromosome 10, from the CEPH collaboration to map the human genome genetically. The map is truly a collaborative achievement, in that the underlying genotypes represent the efforts of laboratories collaborating with each other and with CEPH to produce a primary genetic map of the genome, consisting of polymorphic markers placed at approximately 20-cM intervals along each of the human autosomes and the X chromosome. Such a map provides a tool for the systematic localization of genes that determine inherited diseases and of other genes of interest. Genetic lo-

genetic map with DNA polymorphisms (Botstein *et al.*, 1980). A key premise of the CEPH collaboration is that the human genetic map will be efficiently achieved by collaborative research on DNA from the *same* sample of families. To this end, CEPH provides to collaborating investigators high-quality cellular DNA produced from cultured lymphoblastoid cell lines (LCL) derived from each member of a reference panel of large nuclear families/pedigrees and a database contributed to and shared by these investigators. Collaborating investigators determine genotypes with their probes and the DNA from the CEPH panel to test the families for segregation of these genetic markers. They then contribute the genotypes to CEPH for preparation of a database which is returned to them for linkage analysis and map construction. As of October 1, 1989, 63 research laboratories in the United States (36), Canada (2), Europe (20), South Africa (2), Japan (2), and Australia (1) collaborate with CEPH in this manner.

CEPH Reference Family Panel

Families with large sibships, living parents, and grandparents are especially informative for linkage mapping (White *et al.*, 1985). From 100 families available from various sources, selected not for disease but for large sibship size, an initial group of 40 families was defined for the CEPH reference panel by the original group of collaborating investigators. Table 1 shows the geographic origins of these families and the contributors of the LCLs to CEPH. These are Caucasian families. The mean sibship size for these 40 families, based on those individuals for whom there are LCLs, is 8.3; no family has less than 6 offspring, and 23 families have 8 or more offspring. LCLs are available for all 4 grandparents in each of 29 families of the reference panel.

Resource

A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree

Michael A. Eberle,[1] Epameinondas Fritzilas,[2] Peter Krusche,[2] Morten Källberg,[2]
Benjamin L.
Sean J. Hum
Gil McVean,

[1]Illumina Incorpor
Nr Saffron Walder
Kingdom; [4]Big Da
United Kingdom

De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families

Jonathan R. Belyeu,[1] Harrison Brand,[2,3,4] Harold Wang,[2,3,4] Xuefang Zhao,[2,3,4] Brent S. Pedersen,[1] Julie Feusier,[5] Meenal Gupta,[1] Thomas J. Nicholas,[1] Joseph Brown,[1] Lisa Baird,[1] Bernie Devlin,[6] Stephan J. Sanders,[7,8] Lynn B. Jorde,[1,10] Michael E. Talkowski,[2,3,4,*] and Aaron R. Quinlan[1,9,10,*]

Summary

Each human genome includes *de novo* mutations that arose during gametogenesis. While these germline mutations represent a fundamental source of ne
mutations in the hu
tural variants (dnSV
CEPH-Utah families
tions detected by ali
and we observe a si
nearly 73% of *de n*
by mutational mech
significant correlati
trum of mutational
different rates and s

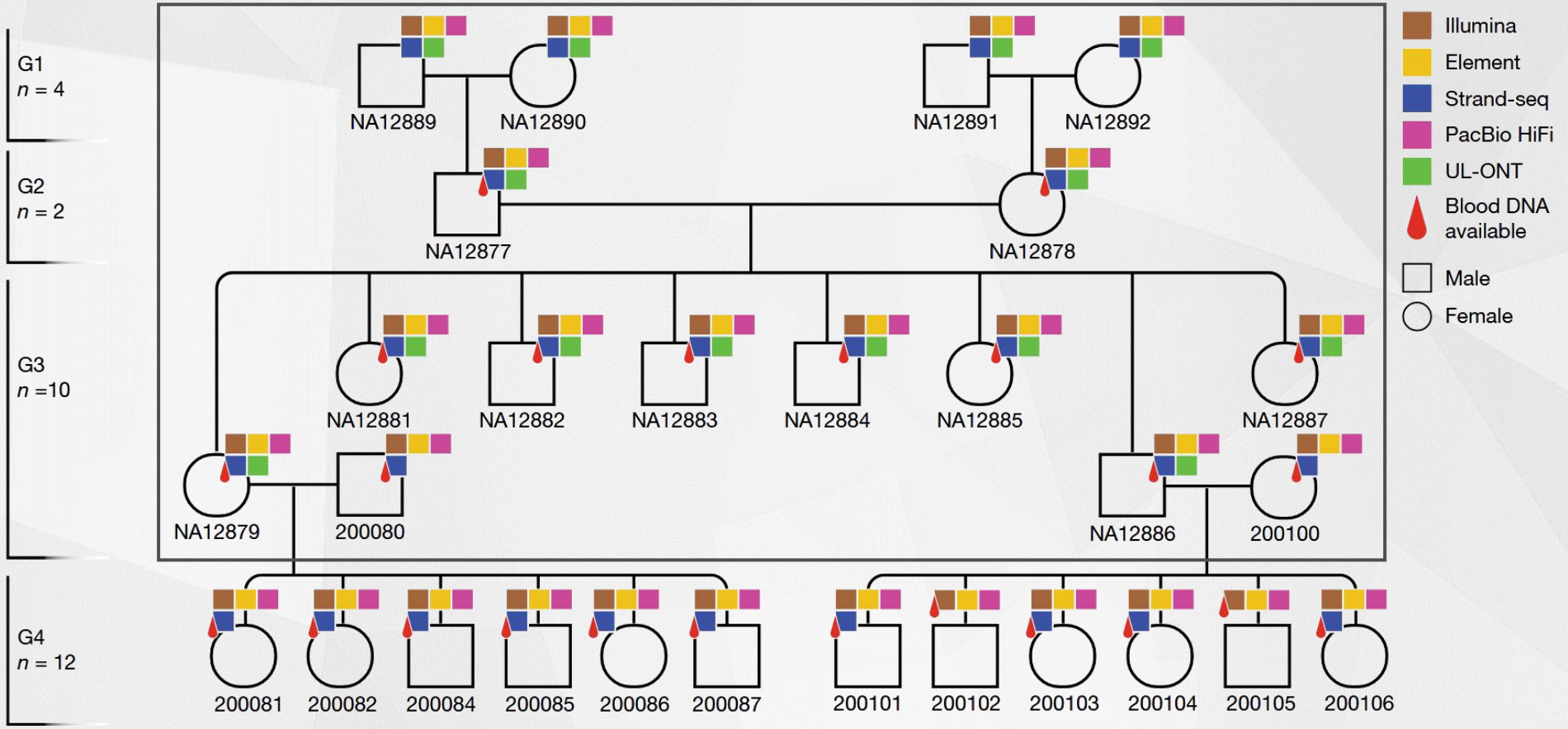
Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation

Thomas A Sasani[1]*, Brent S Pedersen[1], Ziyue Gao[2], Lisa Baird[1], Molly Przeworski[3,4], Lynn B Jorde[1,5]*, Aaron R Quinlan[1,5,6]*

[1]Department of Human Genetics, University of Utah, Salt Lake City, United States; [2]Howard Hughes Medical Institute and Department of Genetics, Stanford University, Stanford, United States; [3]Department of Biological Sciences, Columbia University, New York City, United States; [4]Department of Systems Biology, Columbia University, New York City, United States; [5]USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, United States; [6]Department of Biomedical Informatics, University of Utah, Salt Lake City, United States

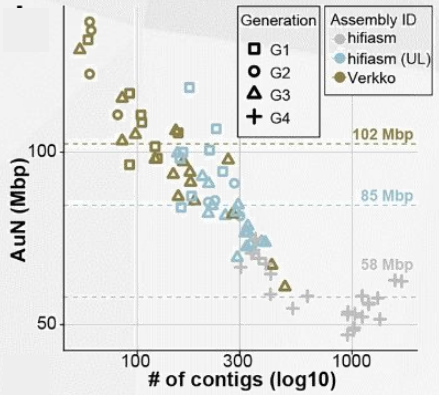
Genome sequence and assembly

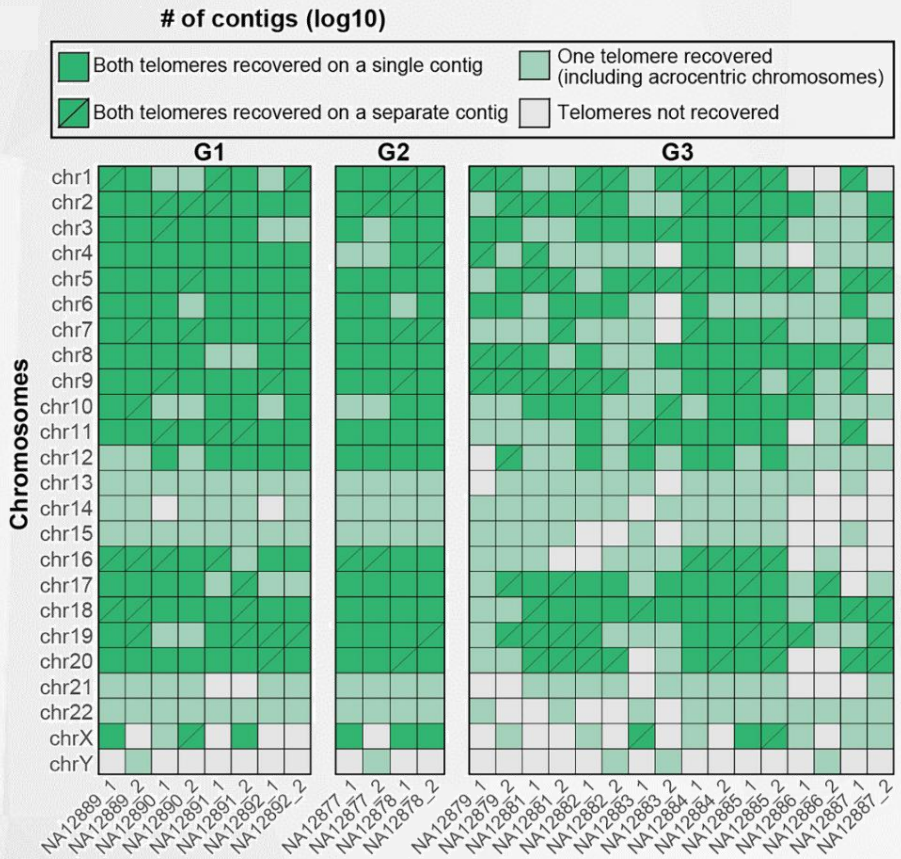
- Deep WGS data from multiple orthogonal sequencing platforms were generated.
- Primarily on the first three generations (G1–G3) and used the fourth generation (G4) to validate *de novo* germline variants.

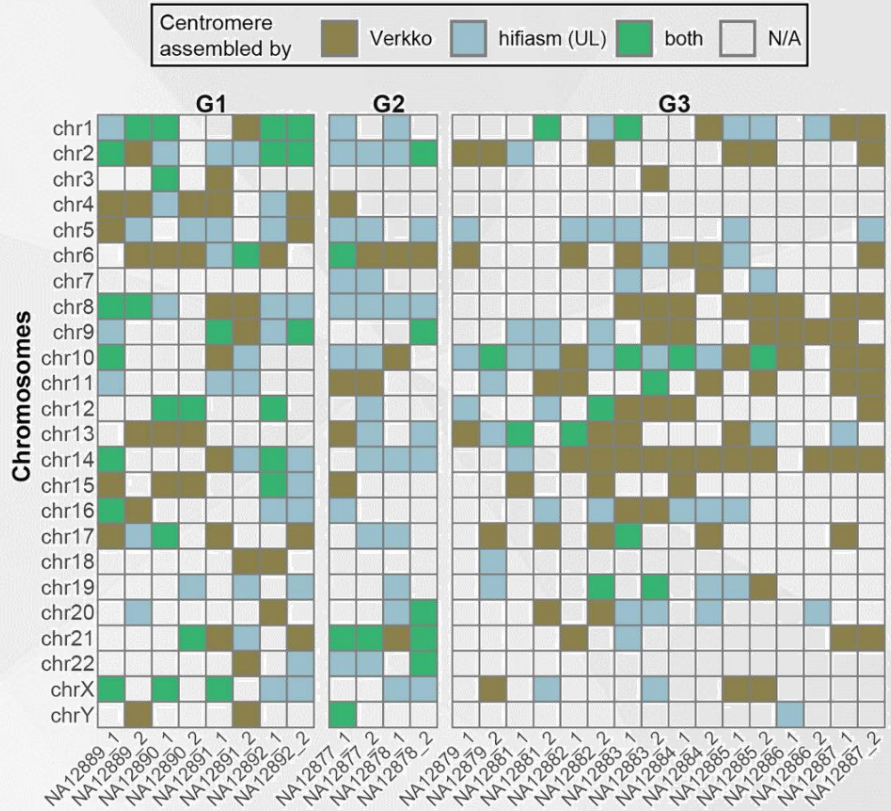


Genome sequence and assembly

- 63.3% (319 out of 504) of chromosomes across G1–G3 are near-T2T.
- 42.3% (213 out of 504) of non-acrocentric chromosomes are spanned in a single contig.
- 288 centromeres (44.7%, 288 out of 644) across G1–G3 were assembled.



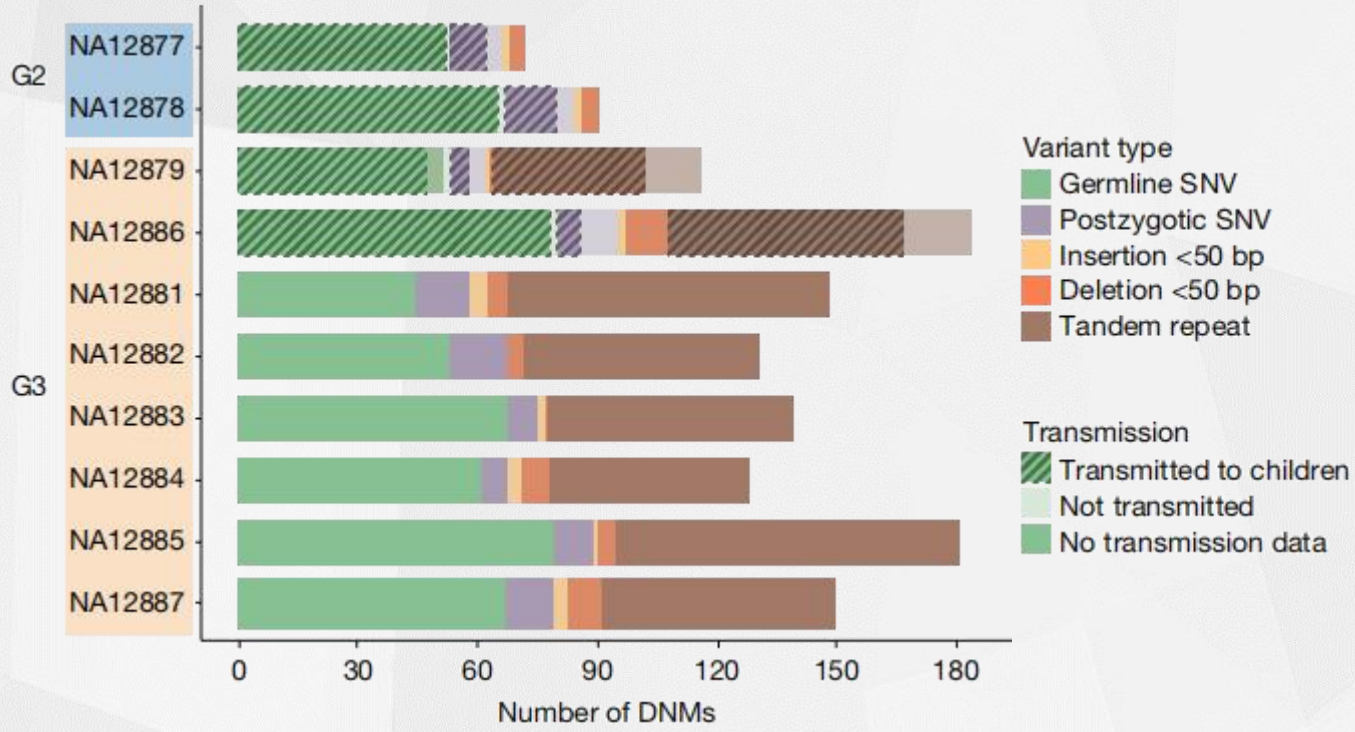




De novo SNVs and small indels

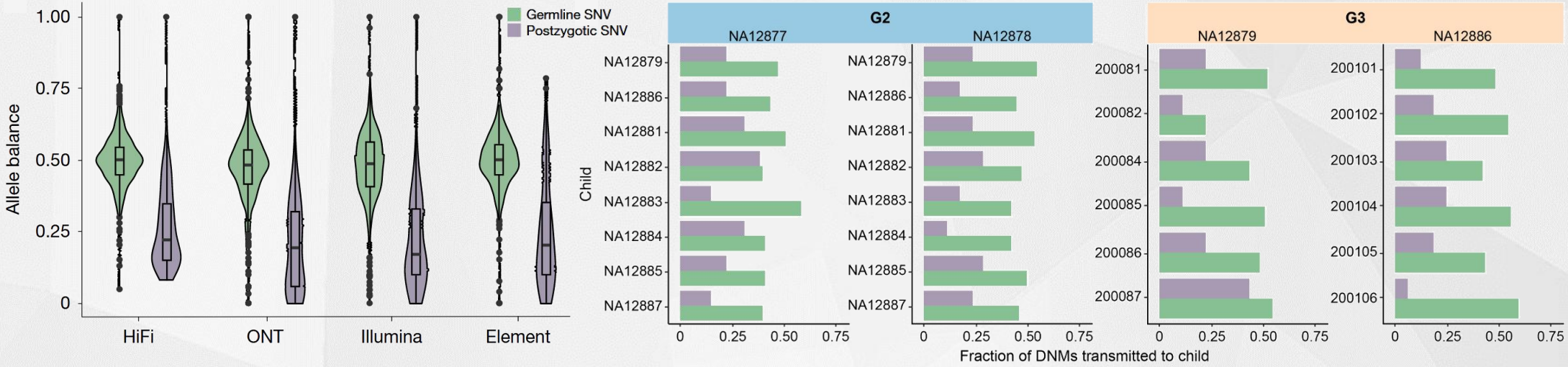
浙江大学 生命演化研究中心

- The *de novo* callset included 755 SNVs and 73 indels across the autosomes, and 27 SNVs and 1 indel on the X chromosome.



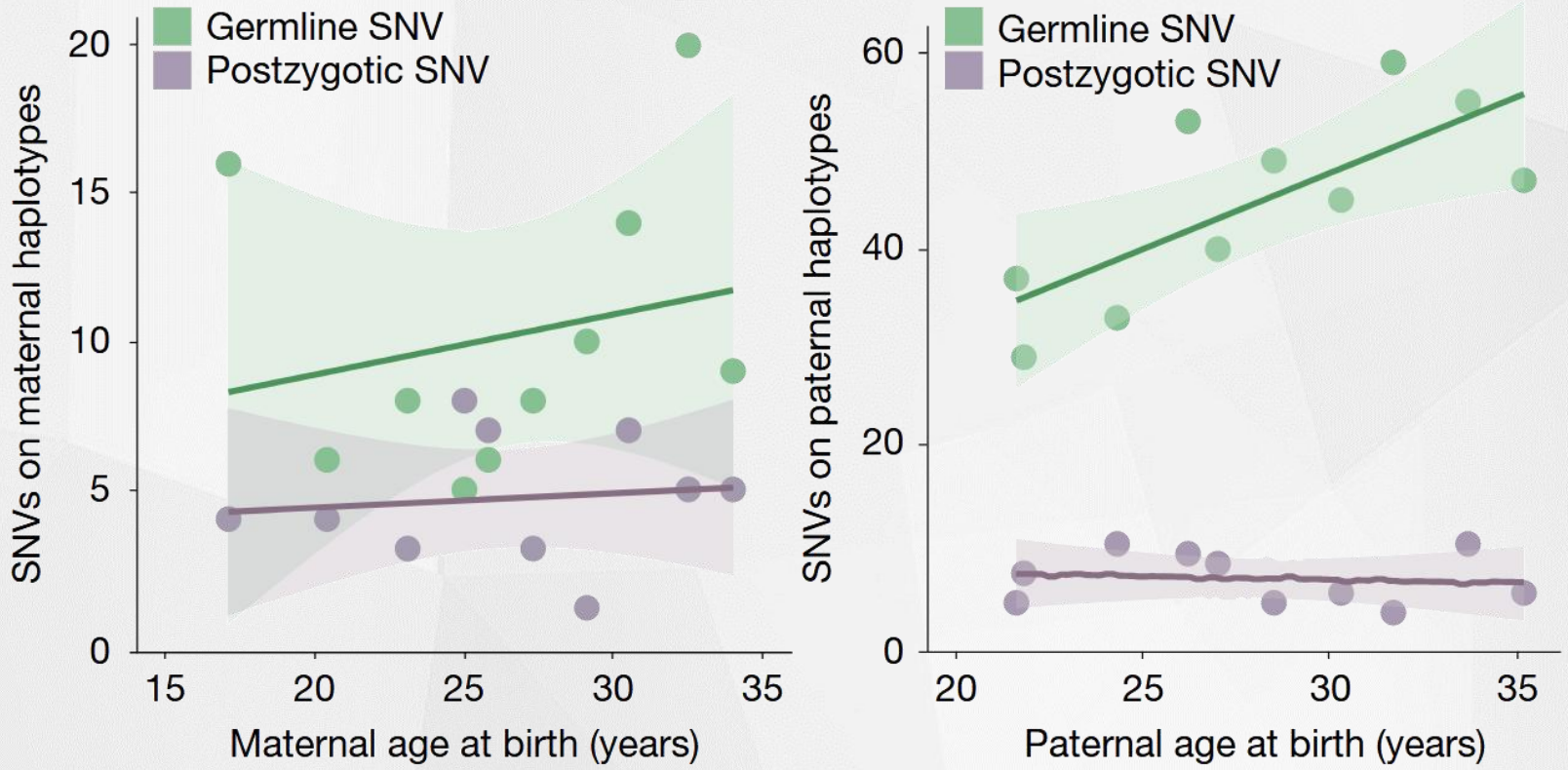
De novo SNVs and small indels

- **DNM originate from a parental gamete or the early embryo (postzygotic mutation, PZM)?**
  - using flanking SNVs to construct haplotypes, phase variants
- A final callset of 119 PZMs, accounting for 16% of total autosomal SNVs
- Of the 62 PZMs in these four samples, 64.5% (n = 40) are transmitted to the next generation, compared with 97.2% of germline SNVs and 100% of indels.



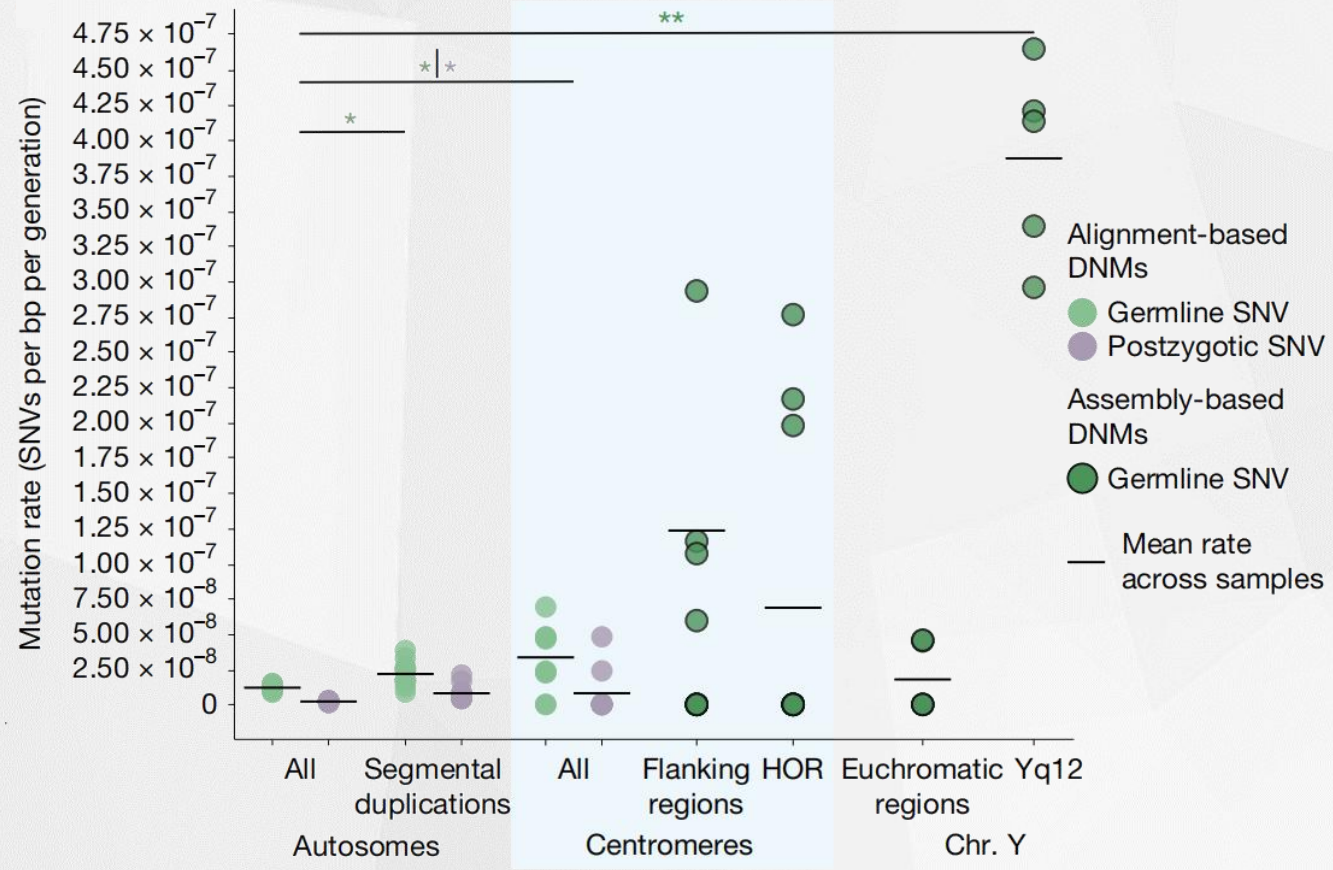
De novo SNVs and small indels

- In total, 81.4% of germline small DNMs originate on paternal haplotypes (**4.38:1 paternal:maternal ratio**)
- A significant parental age effect of 1.55 germline DNMs per additional year of paternal age.
- PZMs show no significant difference with respect to parental origin (1.38:1 paternal:maternal) and no parental age effects



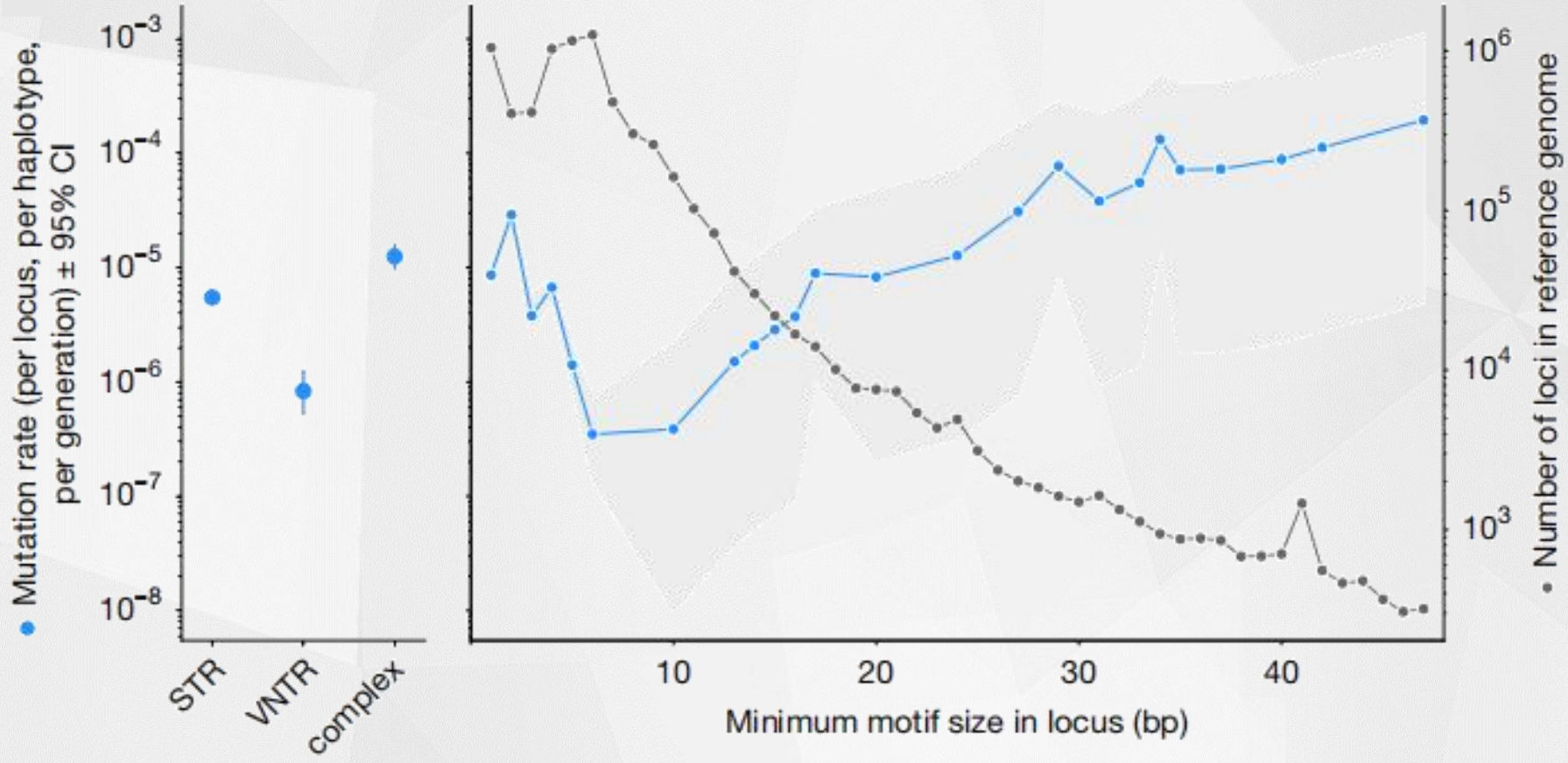
De novo SNVs and small indels

- Parental germline contributes 1.17×10^{-8} SNVs per bp per generation.
- *De novo* SNVs are significantly enriched in repetitive sequences, as much as 2.8-fold in centromeres and 1.9-fold in SDs.
- A lower PZM rate of 2.04×10^{-9} SNVs per bp per generation across the autosomes, yet enriched 3.9 fold in SD regions.



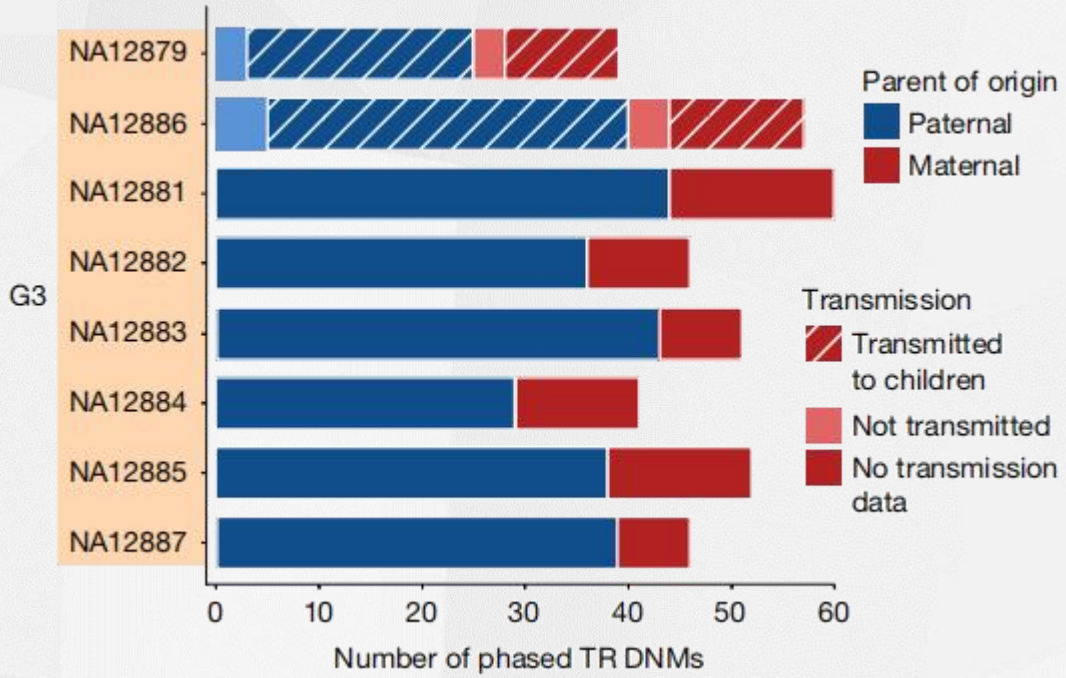
De novo TRs

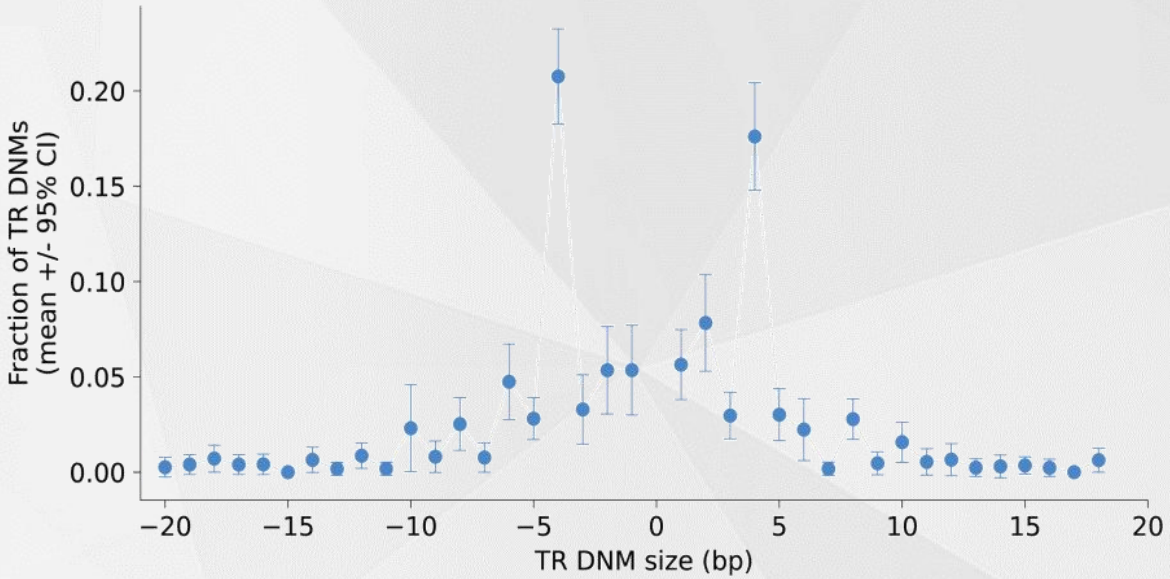
- An average of 65.3 TR DNMs (including STRs, VNTRs and complex loci) per sample and estimated a TR DNM rate of 4.74×10^{-6} per locus per haplotype per generation, with substantial variation across repeat motif sizes



De novo TRs

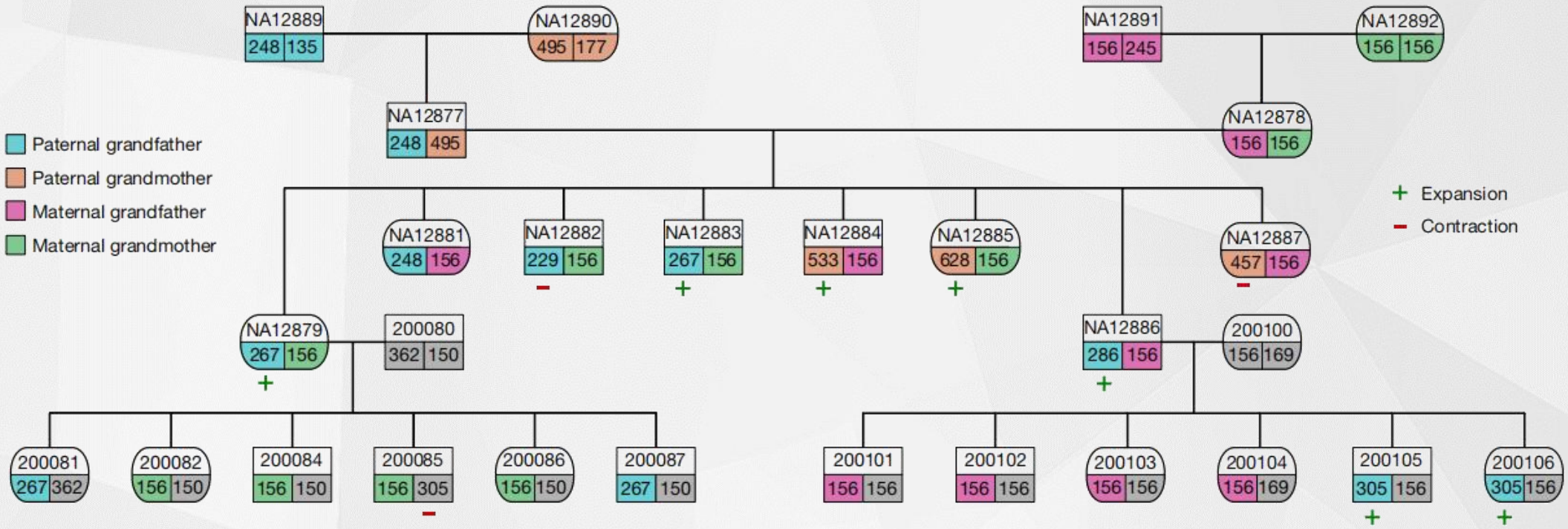
- Overall, 75.0% of phased *de novo* TR alleles were paternal in origin.
- No significant bias towards expansions or contractions was observed.





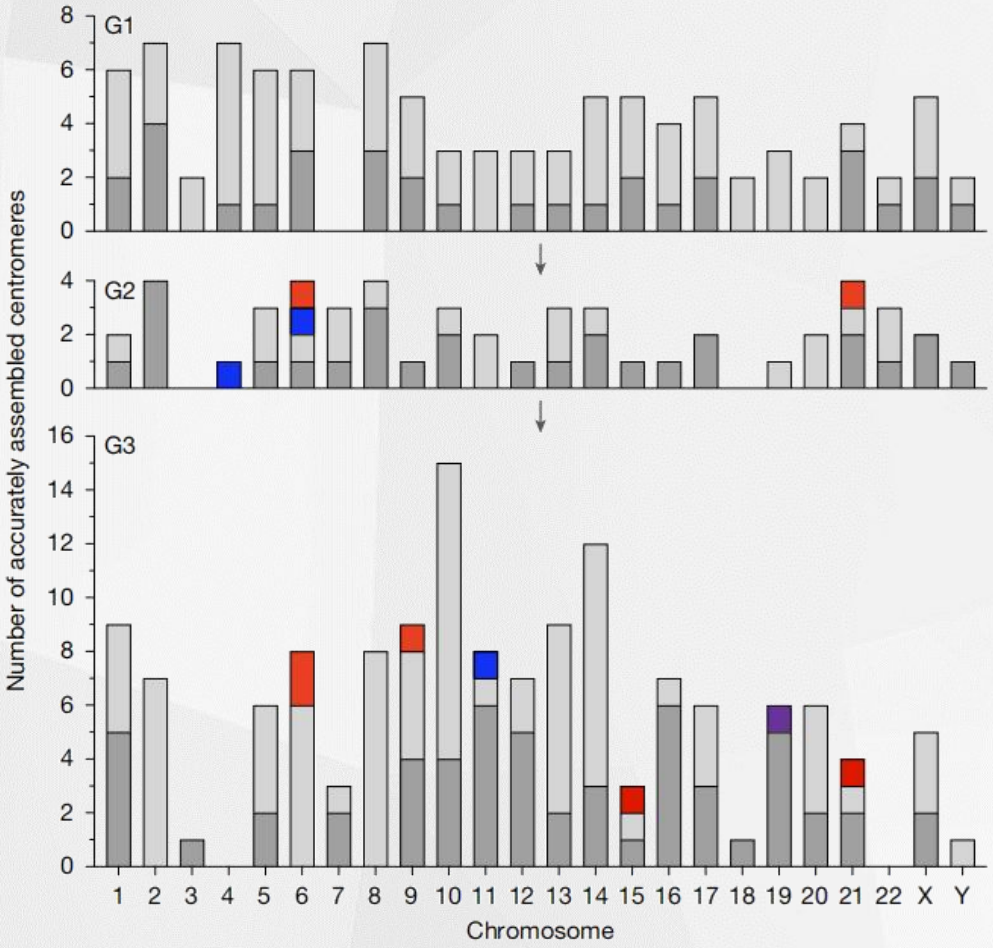
De novo TRs

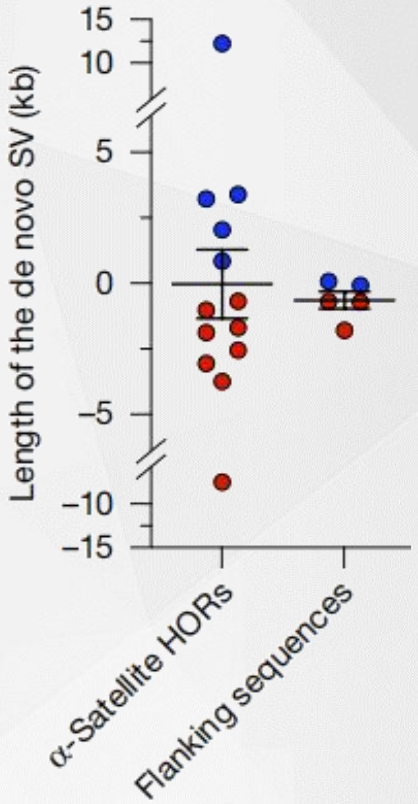
- a high-confidence set of 32 loci that were recurrently mutated among members of the pedigree.
- A recurrent VNTR locus at chromosome 8: 2376919–2377075(T2T-CHM13)



Centromere transmission and *de novo* SVs

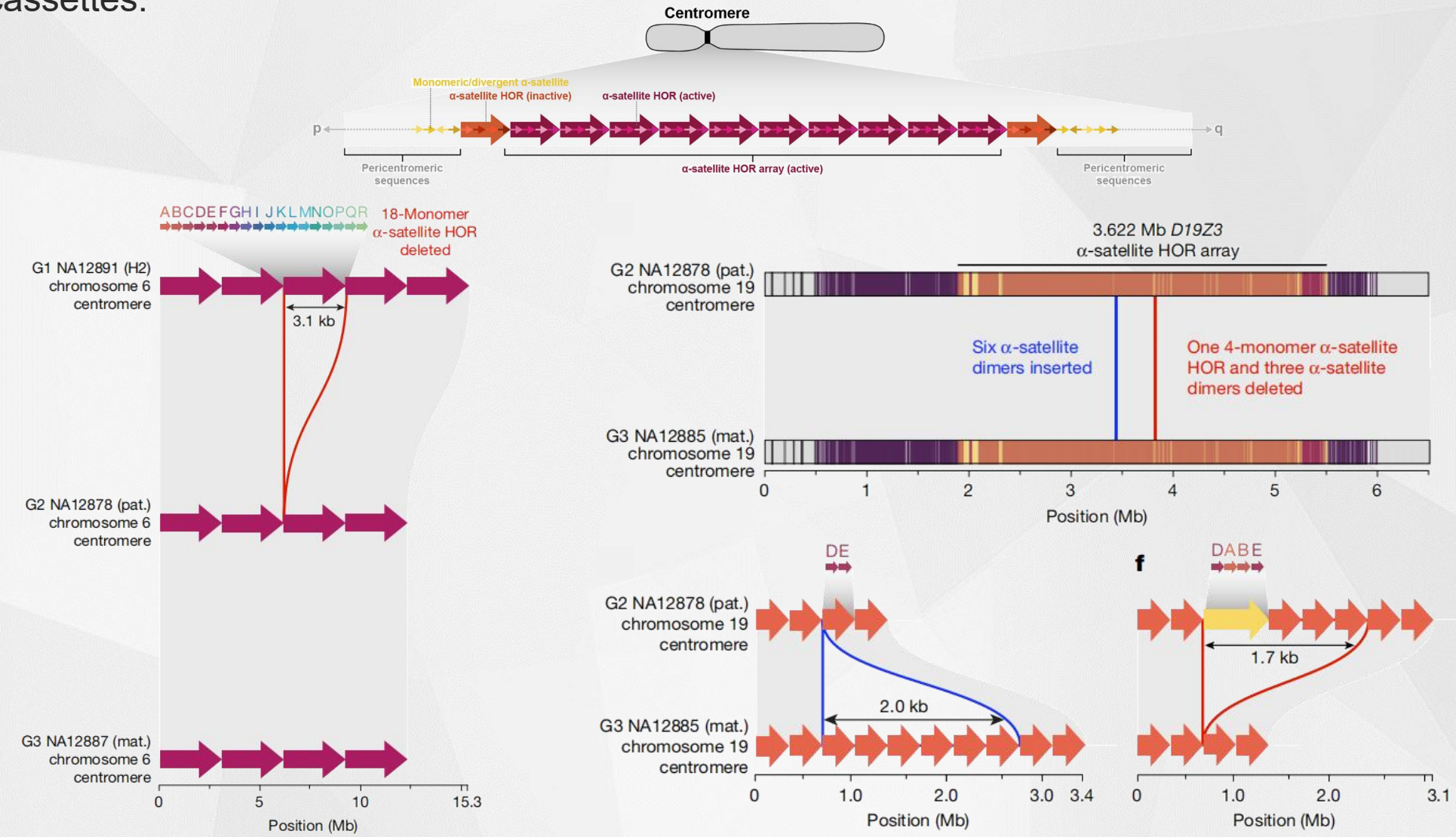
- Among the 288 completely assembled centromeres, 150 intergenerational transmissions were assessed.
- 18 (12%) *de novo* SVs with roughly equivalent numbers of insertions and deletions





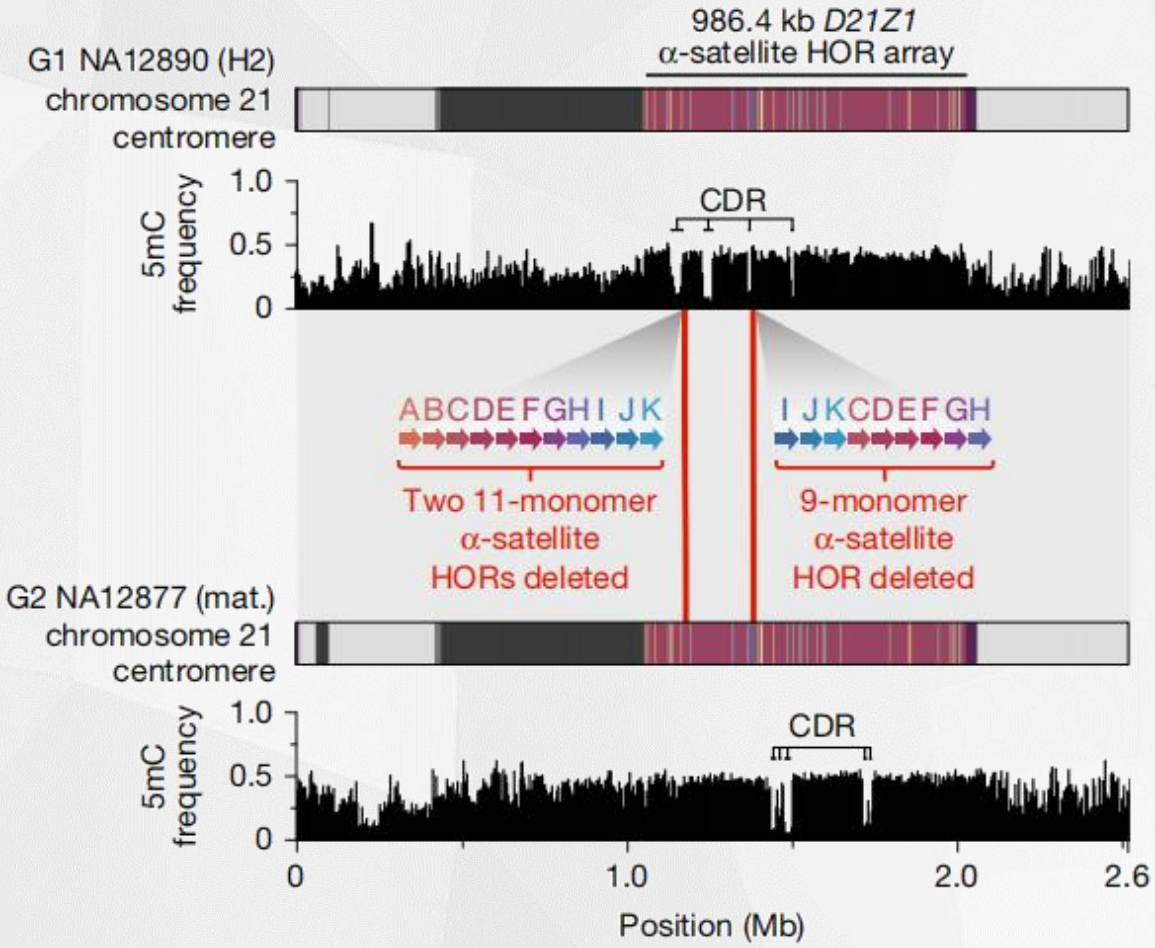
Centromere transmission and de novo SVs

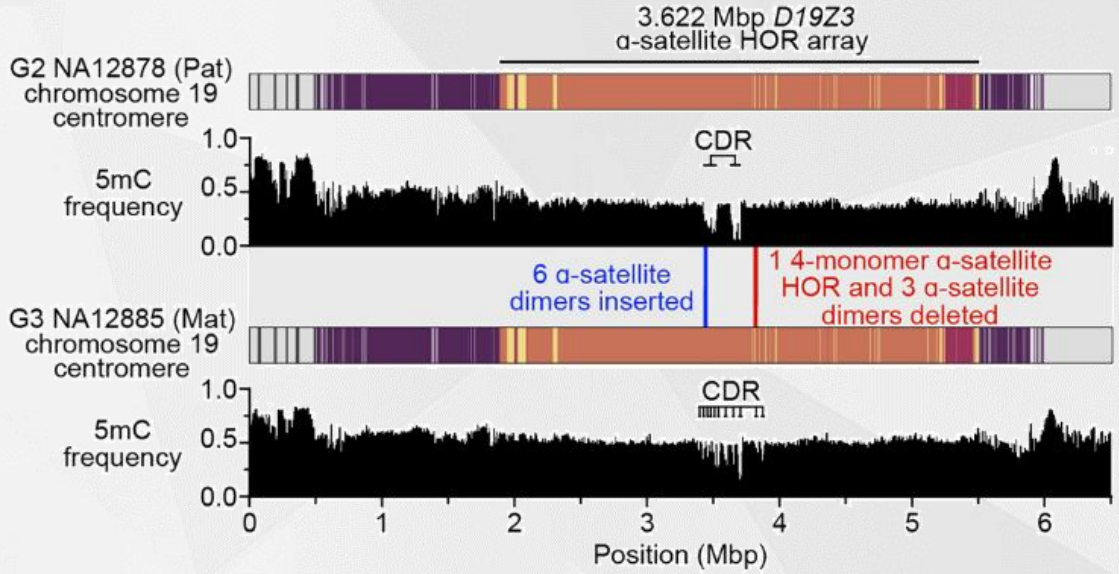
- All α-satellite HOR *de novo* SV events involve integer changes in the basic α-satellite HOR cassettes.



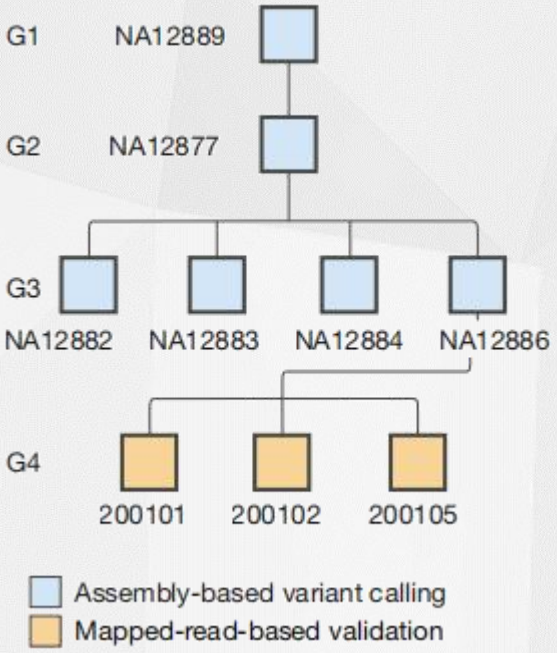
Centromere transmission and de novo SVs

- 18 SV events for their potential effect on the hypomethylation pocket associated with the centromere dip region (CDR)—a marker of the site of kinetochore attachment (动粒附着点）.

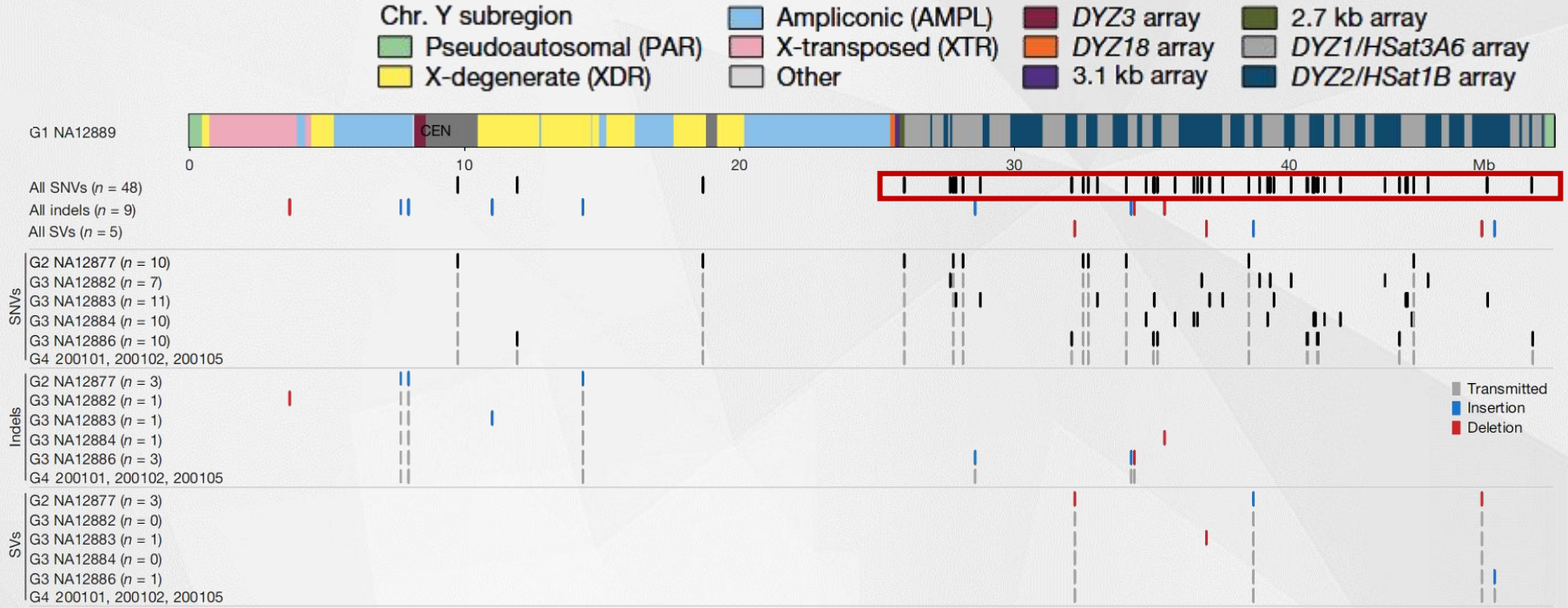




Y chromosome mutations

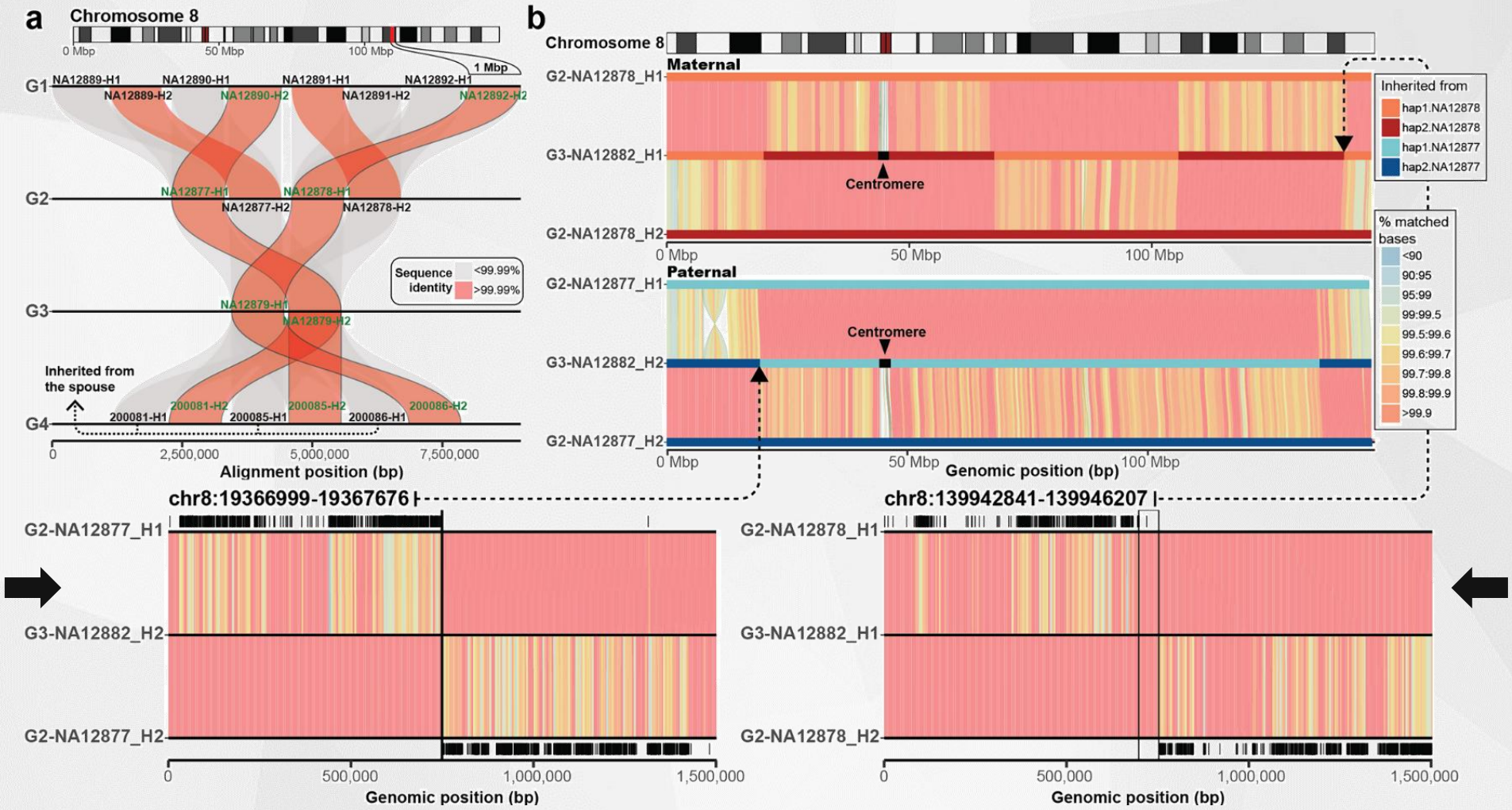


- 48 *de novo* SNVs in the male-specific Y-chromosomal region (MSY) across the 5 G2–G3 male individuals, ranging from 7 to 11 SNVs per Y transmission (mean, 9.6; median, 10)
- In total, a *de novo* SNV rate of 1.99×10^{-7} mutations per bp per generation, an order of magnitude higher than that previously reported.



Sequence-resolved recombination map

- 539 meiotic breakpoints in G3 with 99.8% supported by more than one approach were identified with a refined median size of about 2.5 kb.

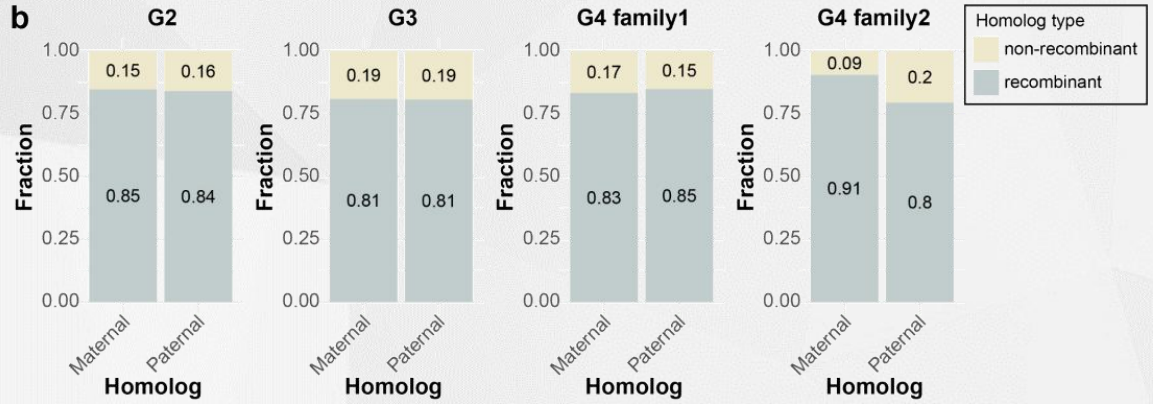


Sharp transition

An extended region of homology

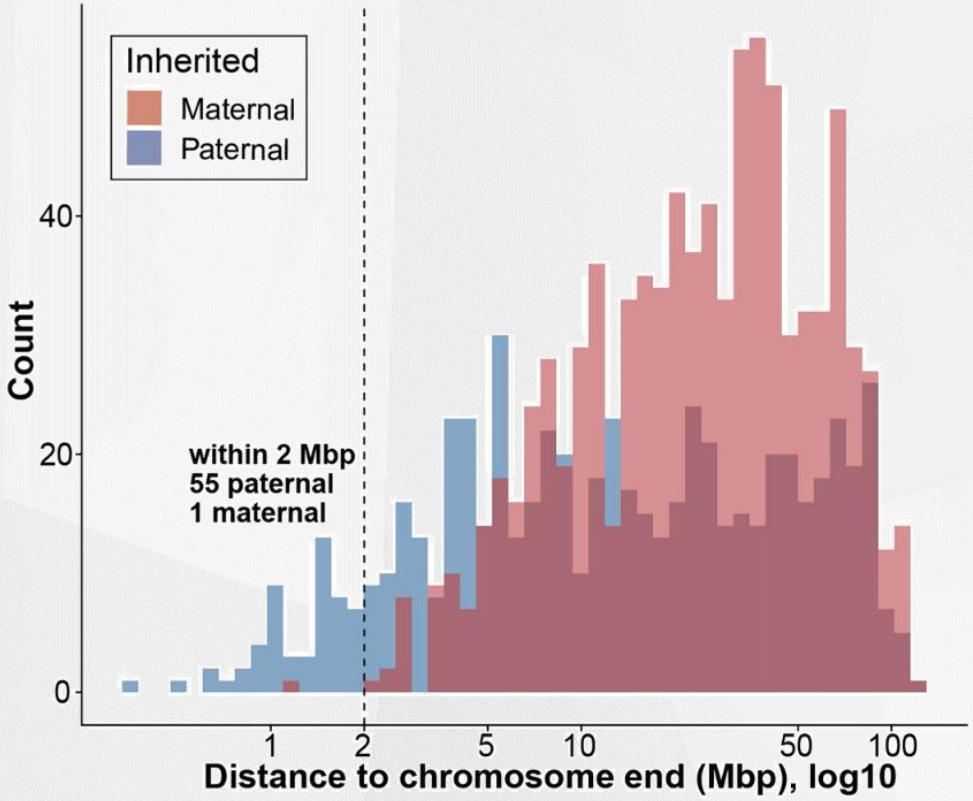
Sequence-resolved recombination map

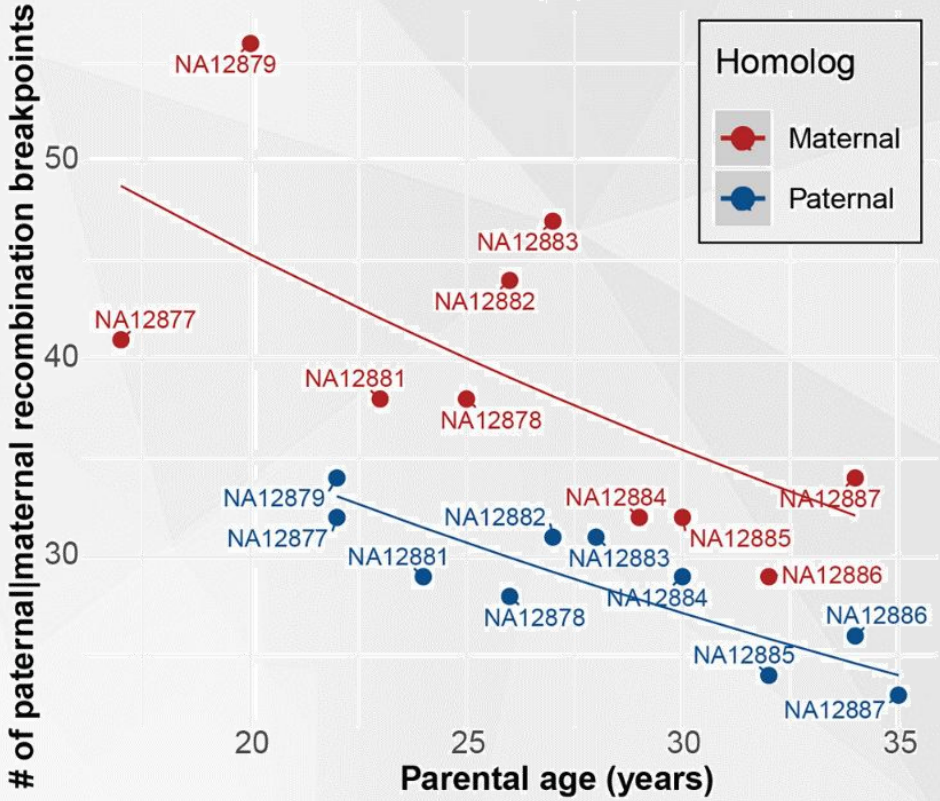
- Overall, 15–20% of paternal and maternal homologues are transmitted without a detectable meiotic breakpoint.



Sequence-resolved recombination map

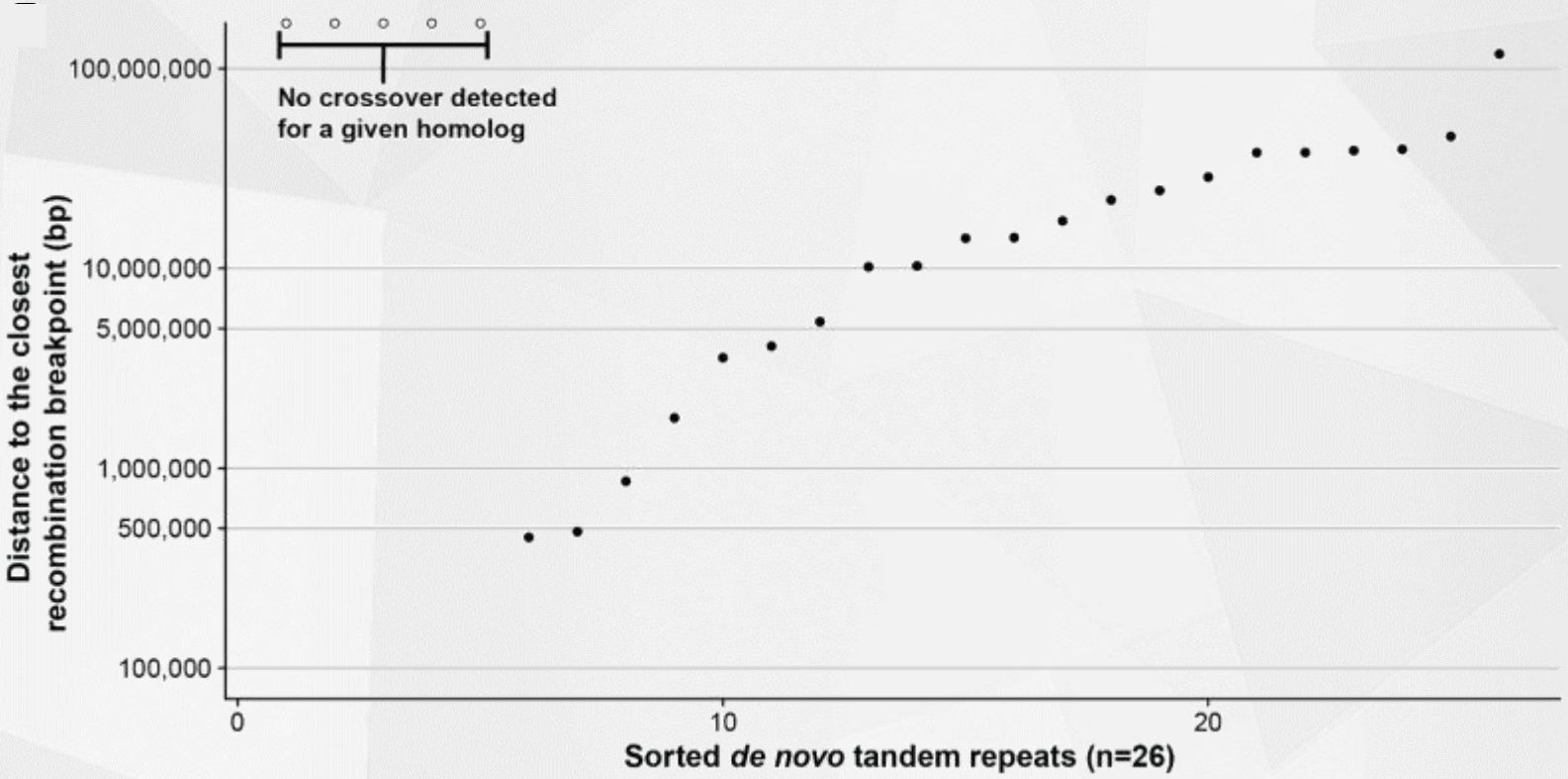
- Overall, 15–20% of paternal and maternal homologues are transmitted without a detectable meiotic breakpoint.
- Paternal recombination is significantly biased towards the ends of human chromosomes.
- In G2–G3, a decrease in crossover events with advancing parental age for both male and female germlines was observed.





De novo SVs

- None of the 27 euchromatic *de novo* SVs coincide with recombination crossovers.



Summary and Discussion

- This study estimate a range of 98–206 DNMs per transmission (average of 152 per generation)
- A strong paternal de novo bias (70–80%) and an increase with advancing paternal age, not only for SNVs but also for indels and SVs, including TR
- 16% of *de novo* SNVs as postzygotic in origin, accounting for 12% of all SNVs transmitted to the next generation , an increase over previous estimates

limitations:

- **Sequencing limitation:** homopolymers still remain challenging even with the use of Element data as longer alleles and motifs embedded in larger repeats are still not reliably assayed with short reads.
- **Super-complex genomic region limitation:** unable to characterize DNMs in the acrocentric regions due to the repetitive nature of the regions and rampant ectopic recombination
- **Sample limitations:** only one multigenerational family with single genetic background and used G4 only for validation purposes of transmitted variants.

浙江大学
生命演化研究中心

Thanks for your listening!