High-Resolution Comparative analysis of Great Ape Genomes

Susie (PAB)
Sumatran Orangutan

Susie (GGO)
Gorilla

Clint (PTR)
Chimpanzee

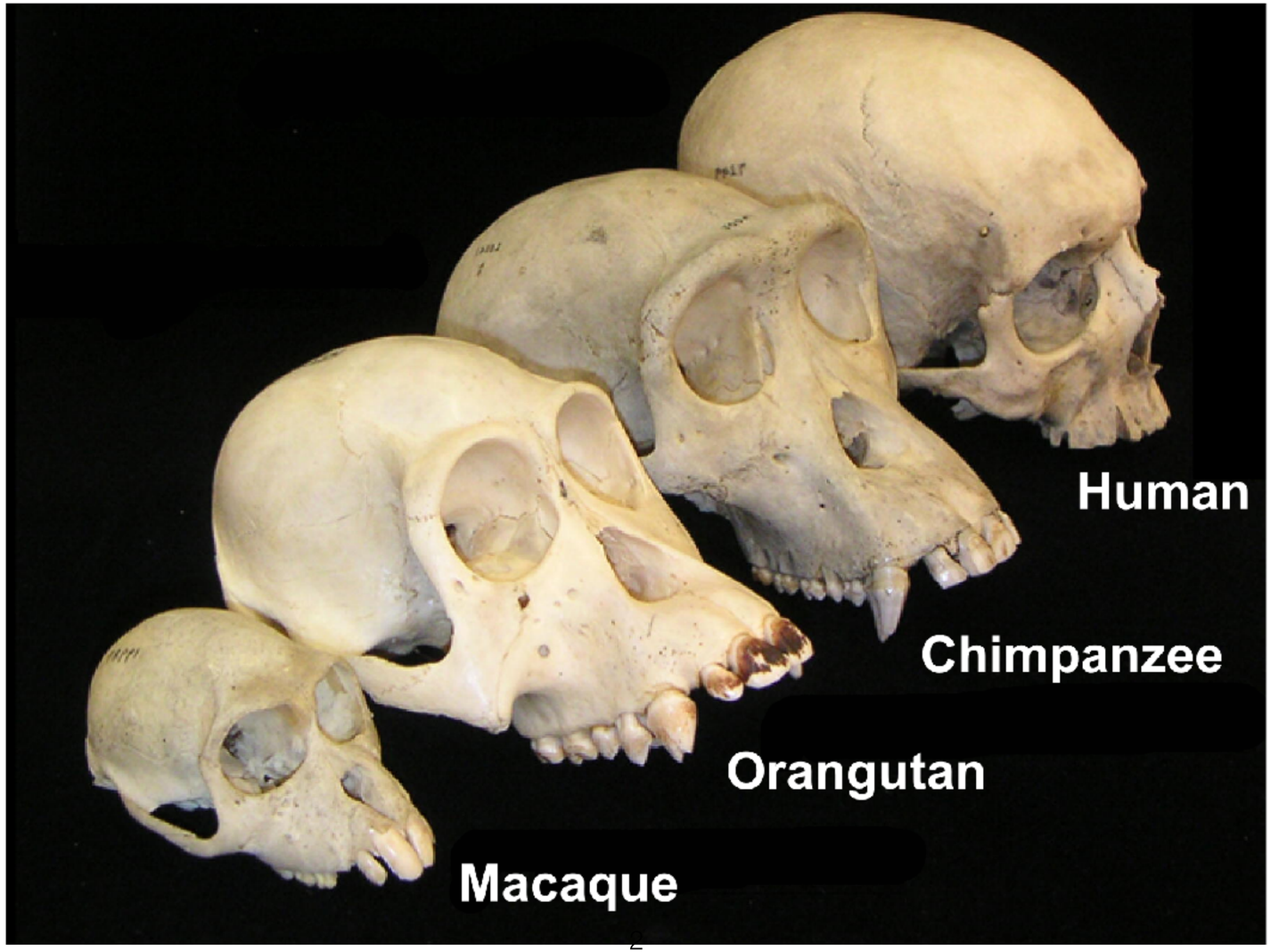
Dr. Eichler
“Super Human”

Zev Kronenberg

Conflict of interest : employee of Phase Genomics

@zevkronenberg

Identifying genetic differences that makes us human through comparative genomics



Structural variation is the largest source of genetic novelty

- Relative to the human reference genome:
  - Every human has ~3 Megabases of single nucleotide variants
  - Every human has ~20 Megabases of structural variation

Studying structural variation in short read assemblies is hard

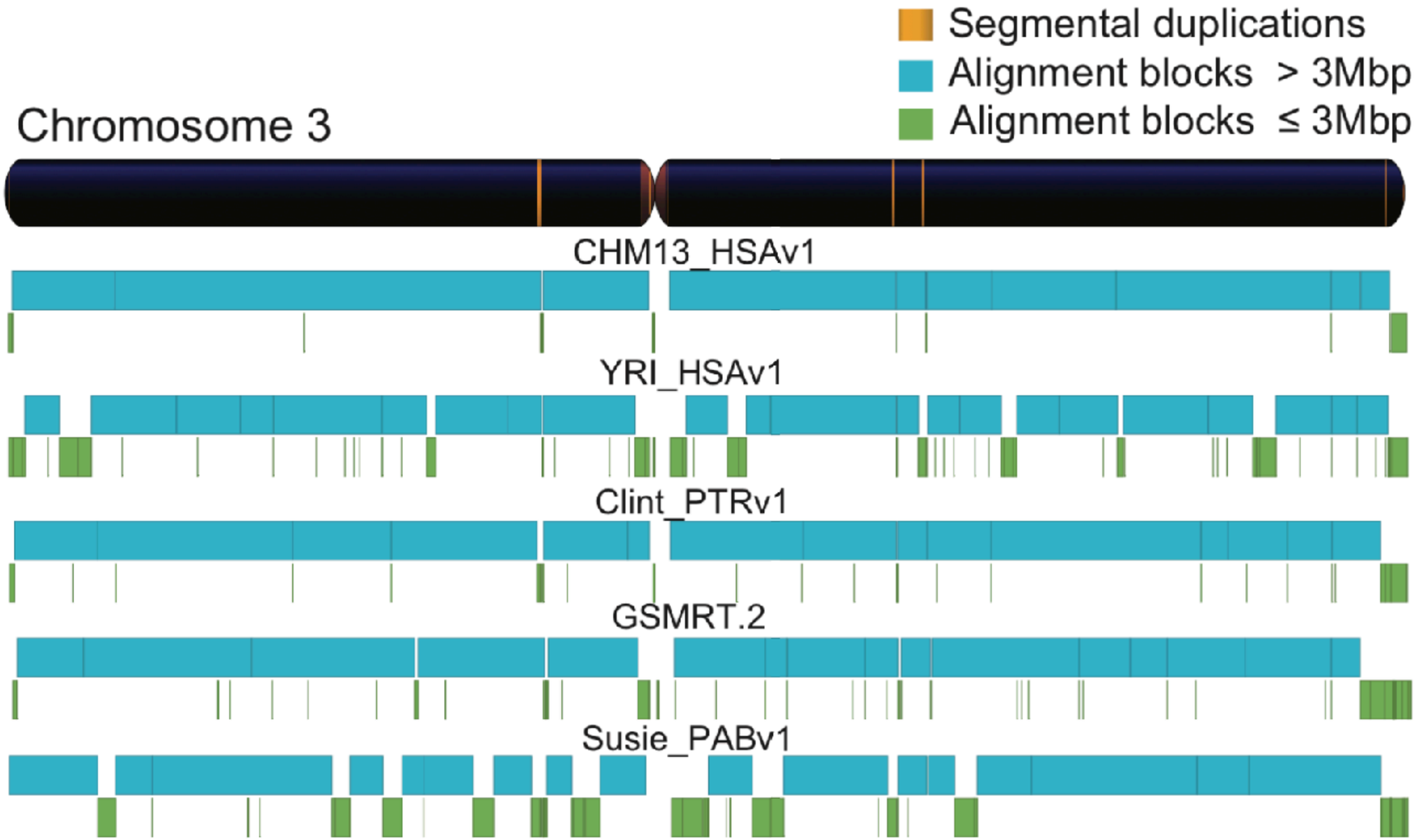
	Orangutan	Gorilla	Chimpanzee
Short read N50	21 kb	21 kb	401 kb

Single Molecule Real Time (SMRT) genome assembly is drastically improving SV comparative studies

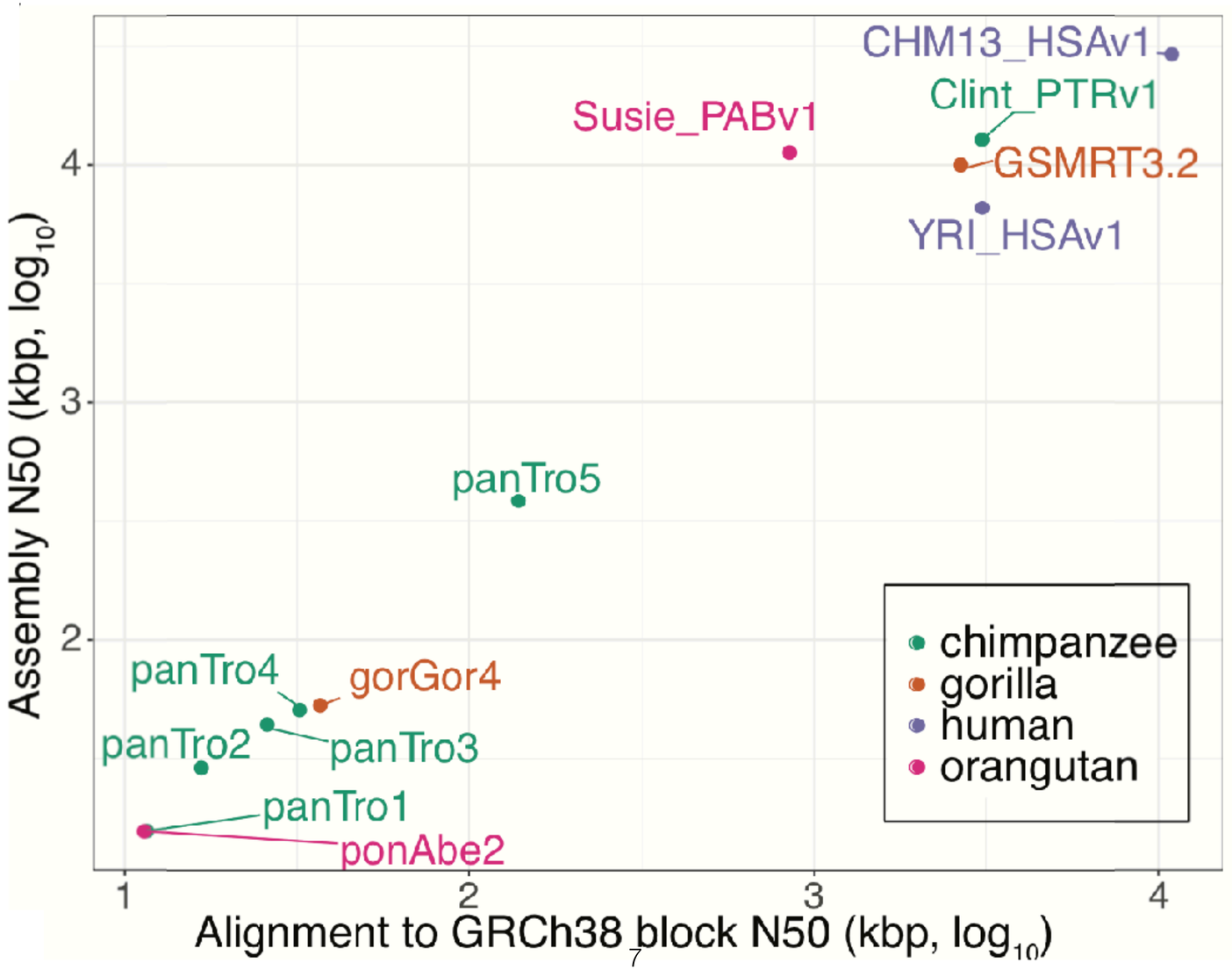
- Closed ~300k gaps in orangutan (96%)
- Closed 30,000 gaps in chimpanzee (50%)
- Removed > 27Mb of extemporaneous sequence in panTro5

	Orangutan	Gorilla	Chimpanzee
Short read N50	21 kb	21 kb	401 kb
PacBio (P6) depth of coverage	74X	74X	97X
SMRT N50 (Falcon asembly)	11.2	11.6 Mb	12 Mb
Improvement	530X	560X	32X

High contiguity of SMRT ape assemblies

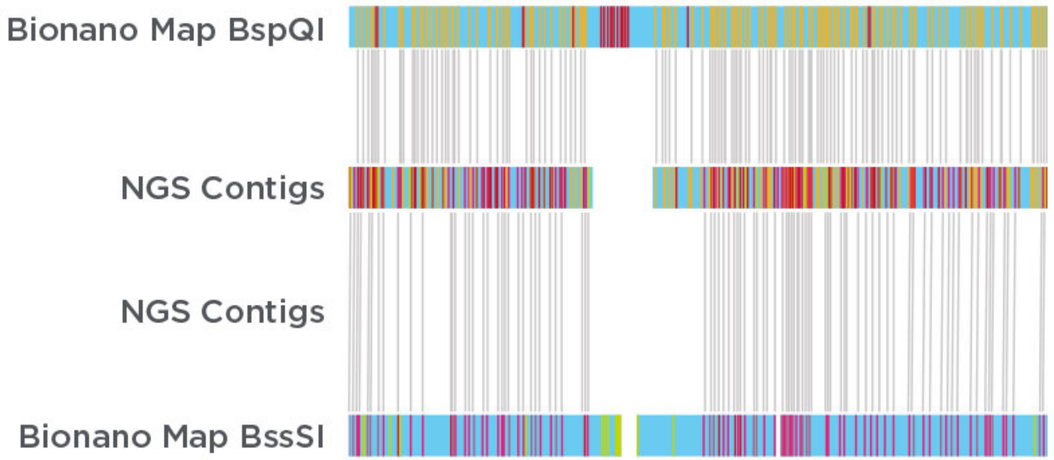


The SMRT great ape assemblies increase the size of syntenic blocks

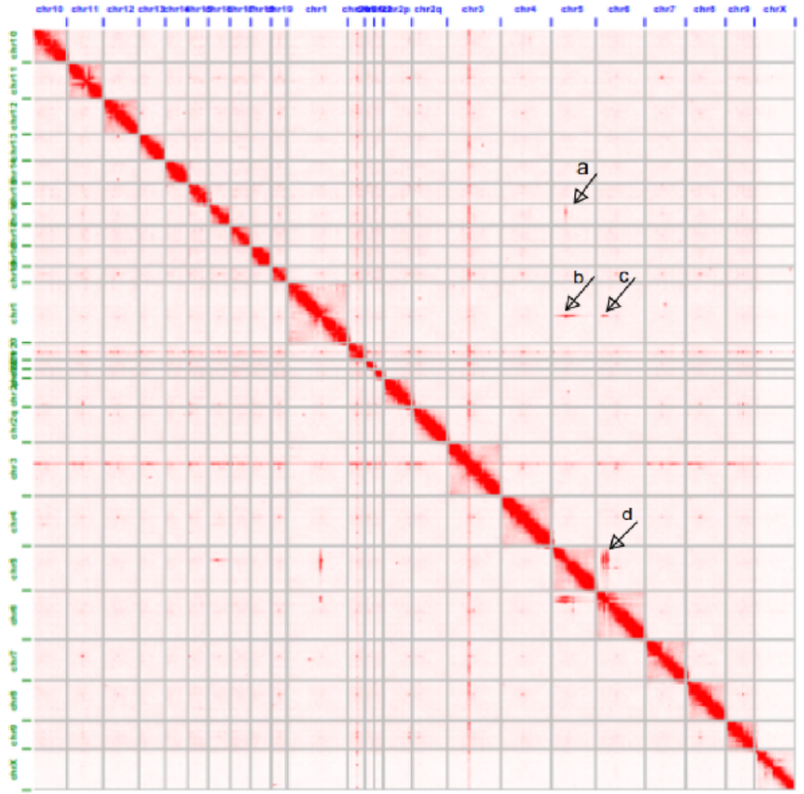


Removing the bias of the human reference genome through integrated scaffolding

Optical maps (Bionano)

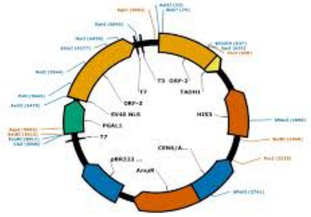


Hi-C (DNAse)

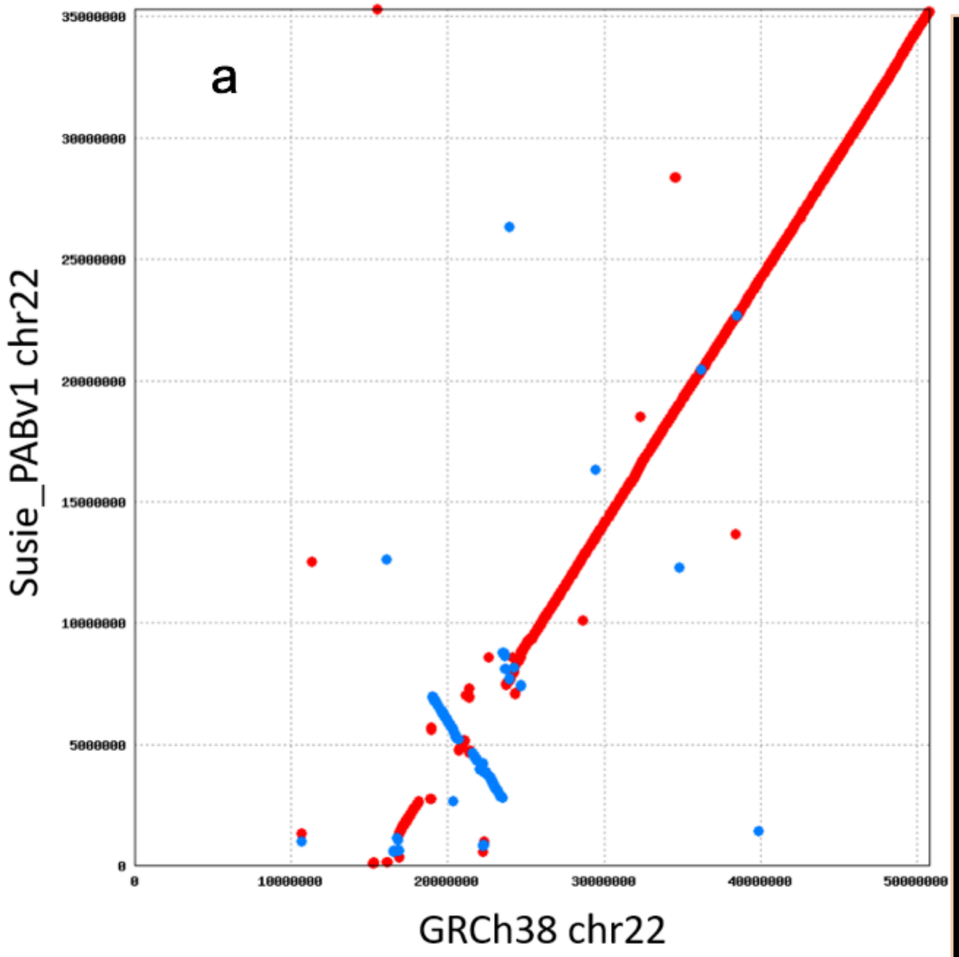


FISH'ed BACs

BAC end sequencing

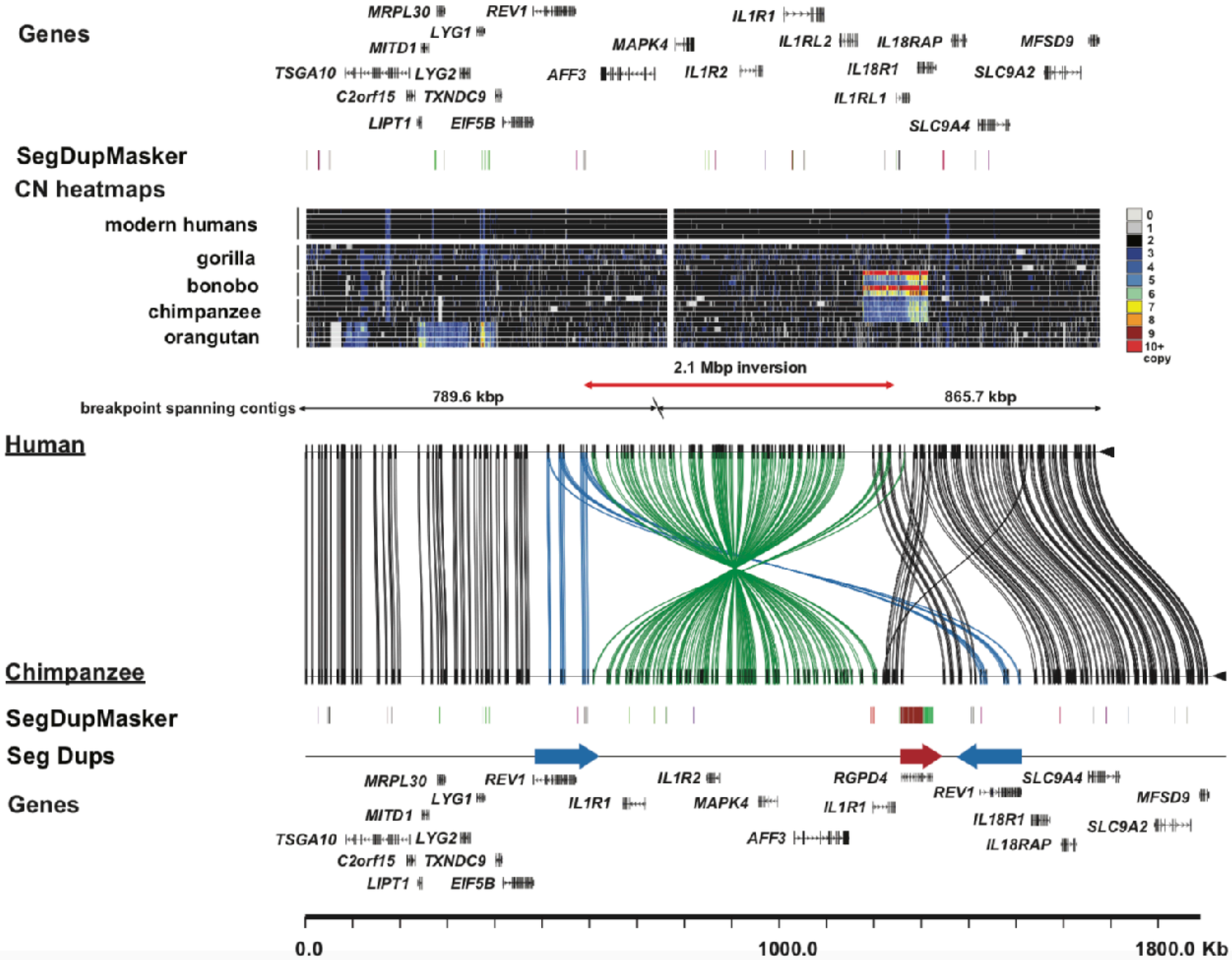


Chromosome-level scaffolds reveal Mb+ sized evolutionary events



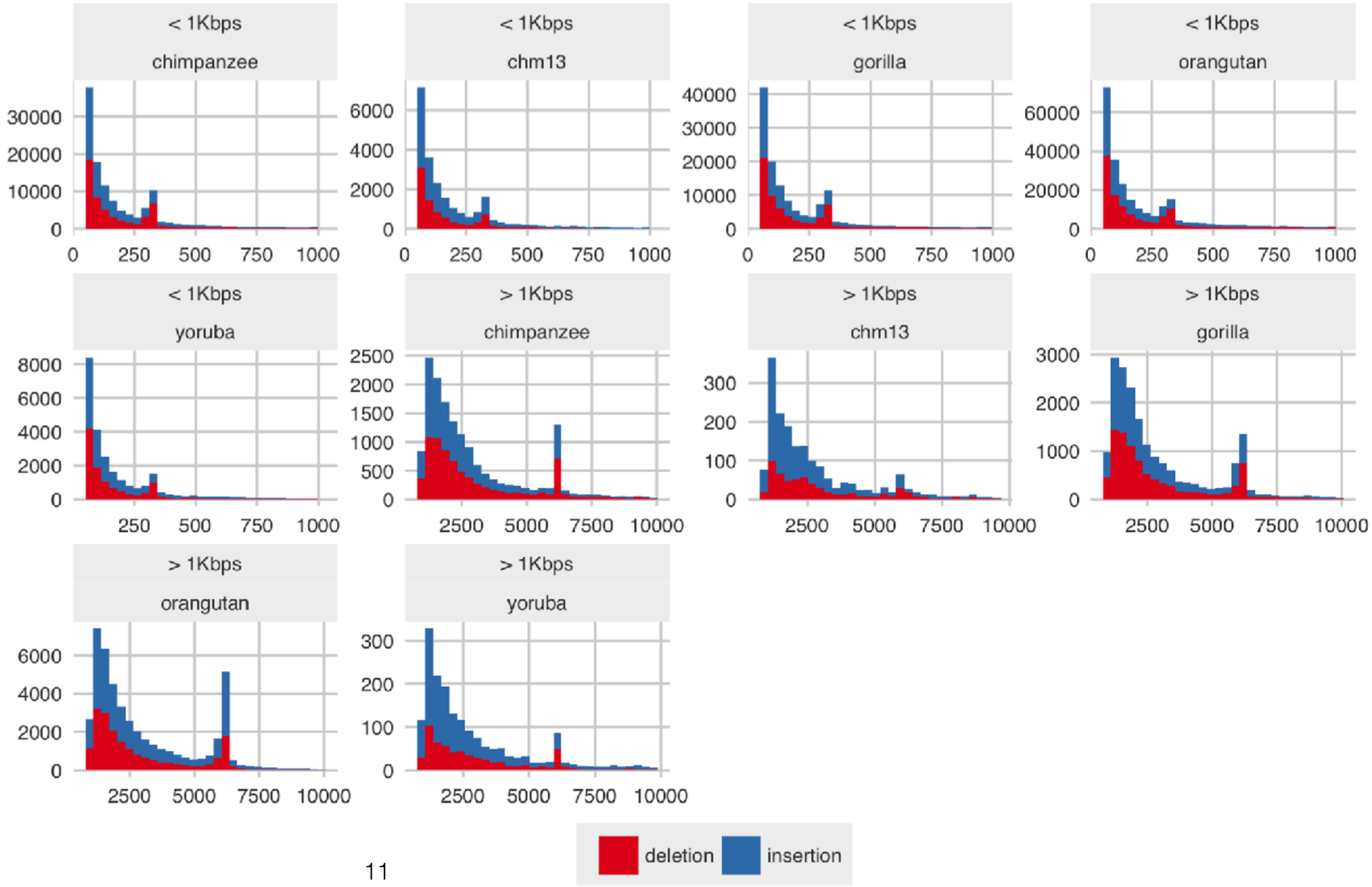
Yunis and Prakash, Science 1982

Optical maps revealed dozens of ~Mb sized evolutionary events



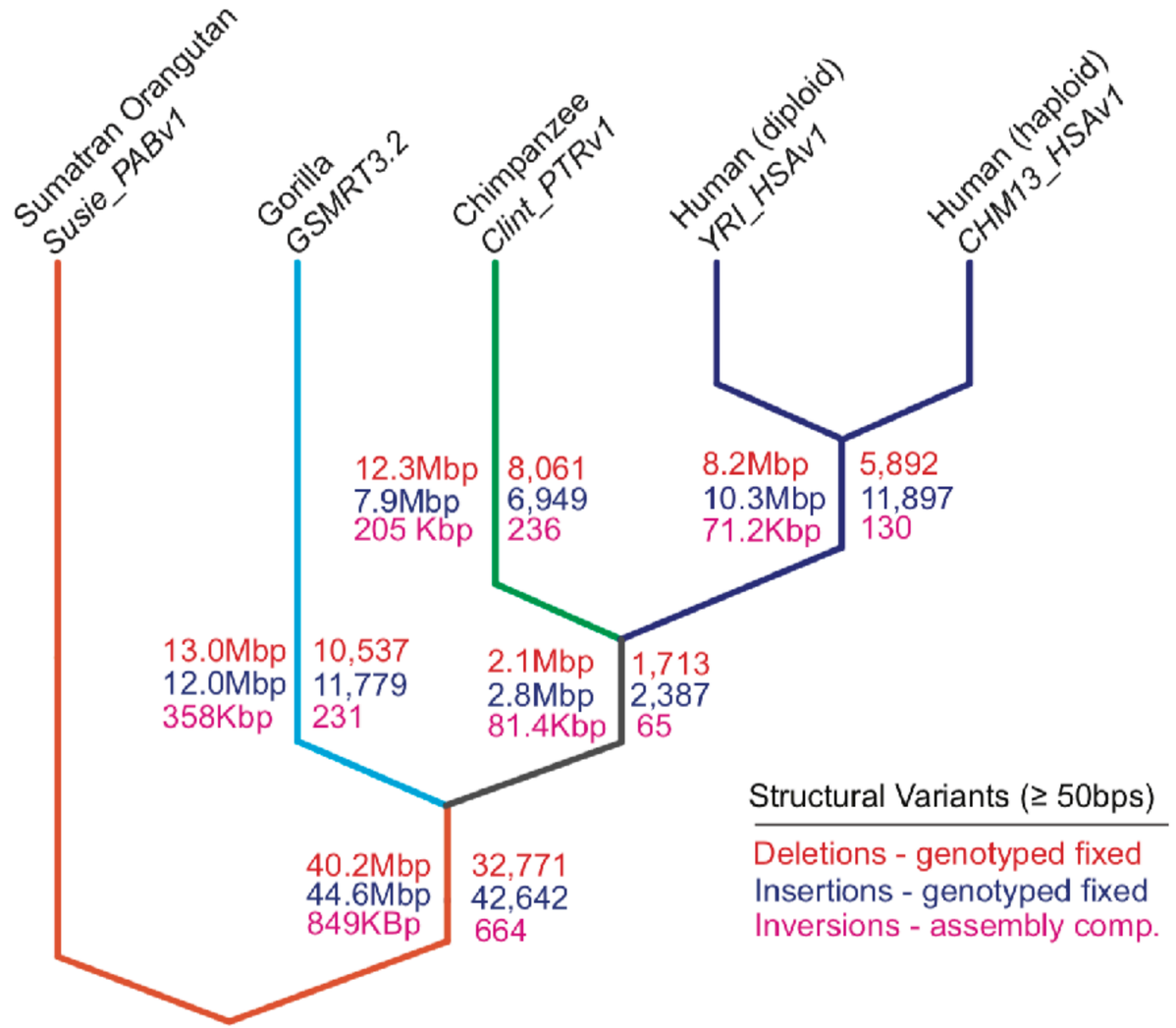
SMRTie-SV a new cross-species tool for discovering SVs

- > 600K SVs greater than 50bp were discovered
- > 90% validation rate
- Open source: https://github.com/zeeev/smartie-sv
- Snakemake pipeline that can run on most clusters
- Uses a modified version of BLASR (Chaisson)
- Largest SV discovered ~ 60K



Defined > 60Mb of lineage specific structural variation

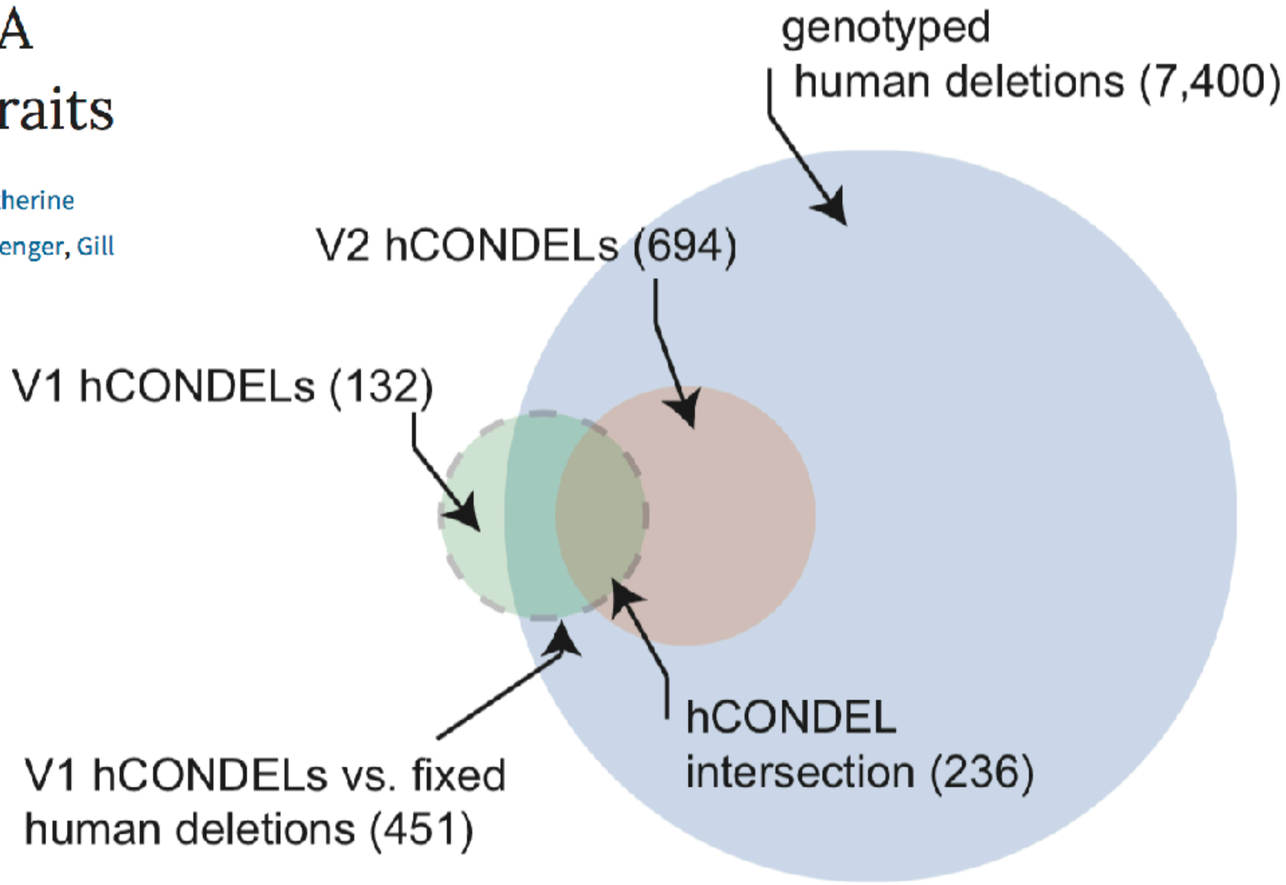
- SMARTie-SV
- Genotyping w. 100 primates
  - SVTyper
  - dCGH
  - Lineage Specific (VST or FST > 0.8)



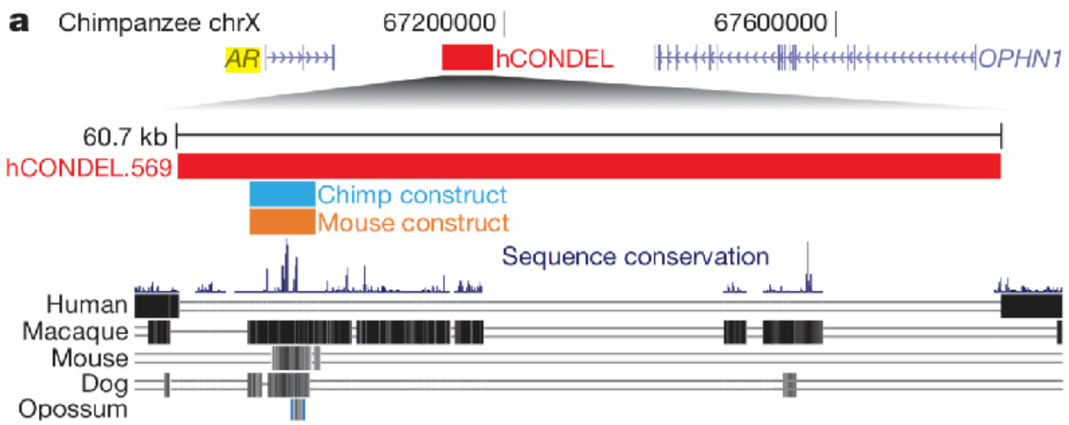
We doubled the number of human conserved deletions (hCONDELs)

Human-specific loss of regulatory DNA and the evolution of human-specific traits

Cory Y. McLean, Philip L. Reno, Alex A. Pollen, Abraham I. Bassan, Terence D. Capellini, Catherine Guenther, Vahan B. Indjeian, Xinhong Lim, Douglas B. Menke, Bruce T. Schaar, Aaron M. Wenger, Gill Bejerano ✉ & David M. Kingsley ✉

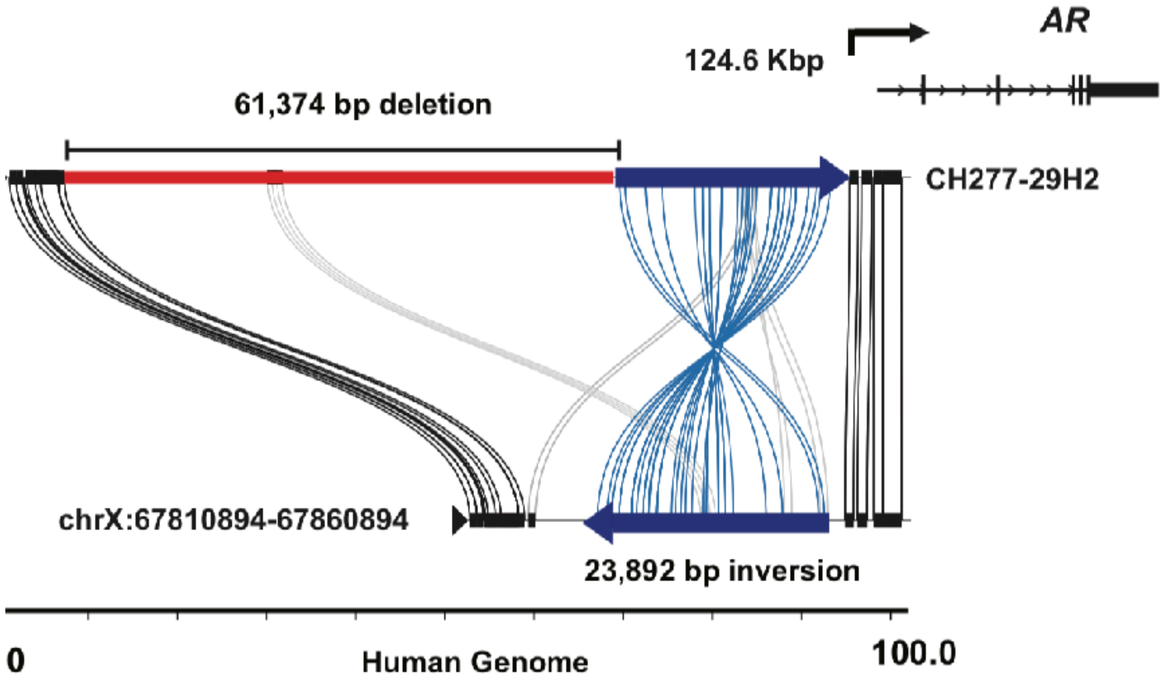


Allelic variation in gorilla near the androgen receptor hCONDEL

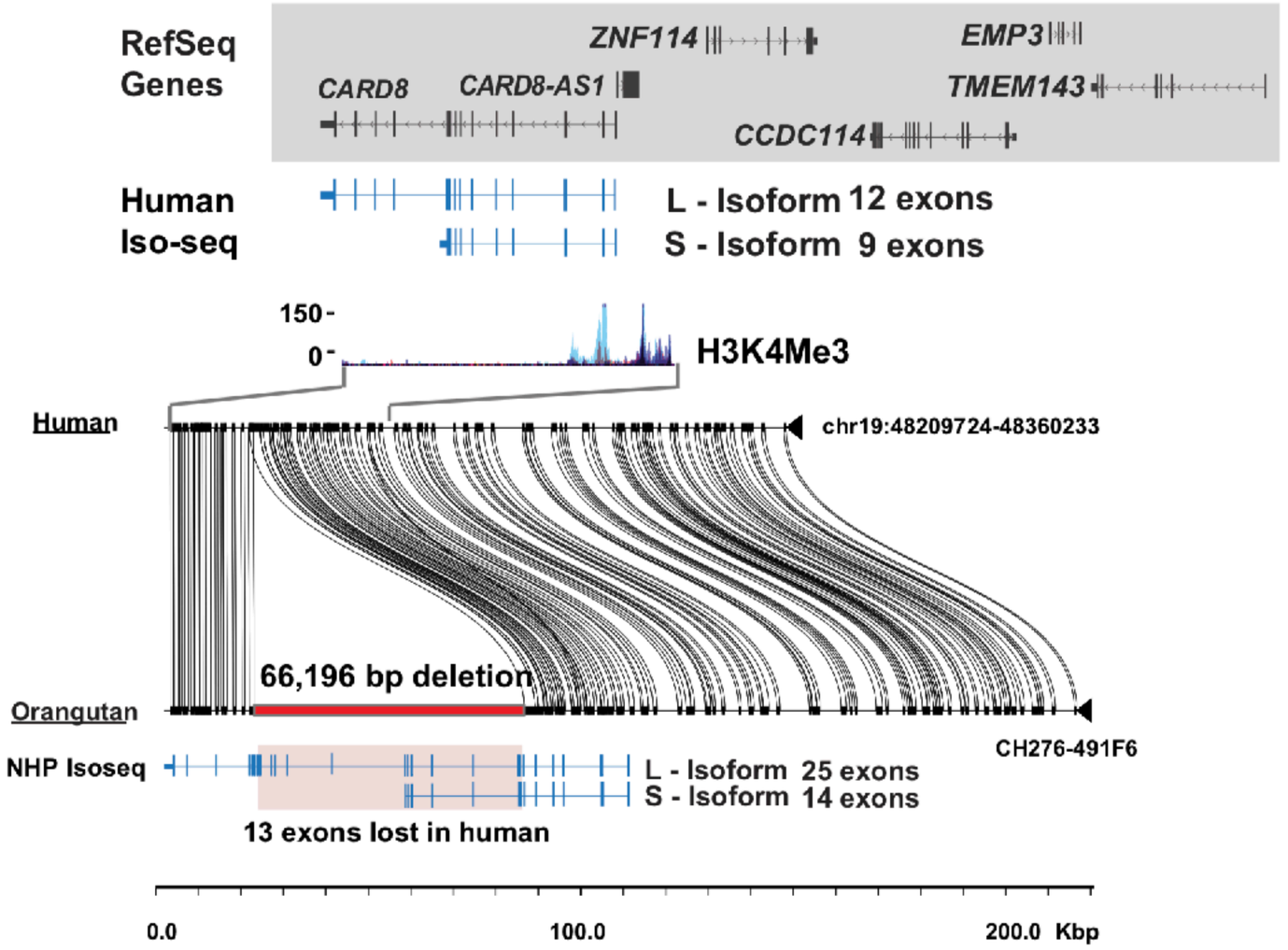


Human-specific loss of regulatory DNA and the evolution of human-specific traits

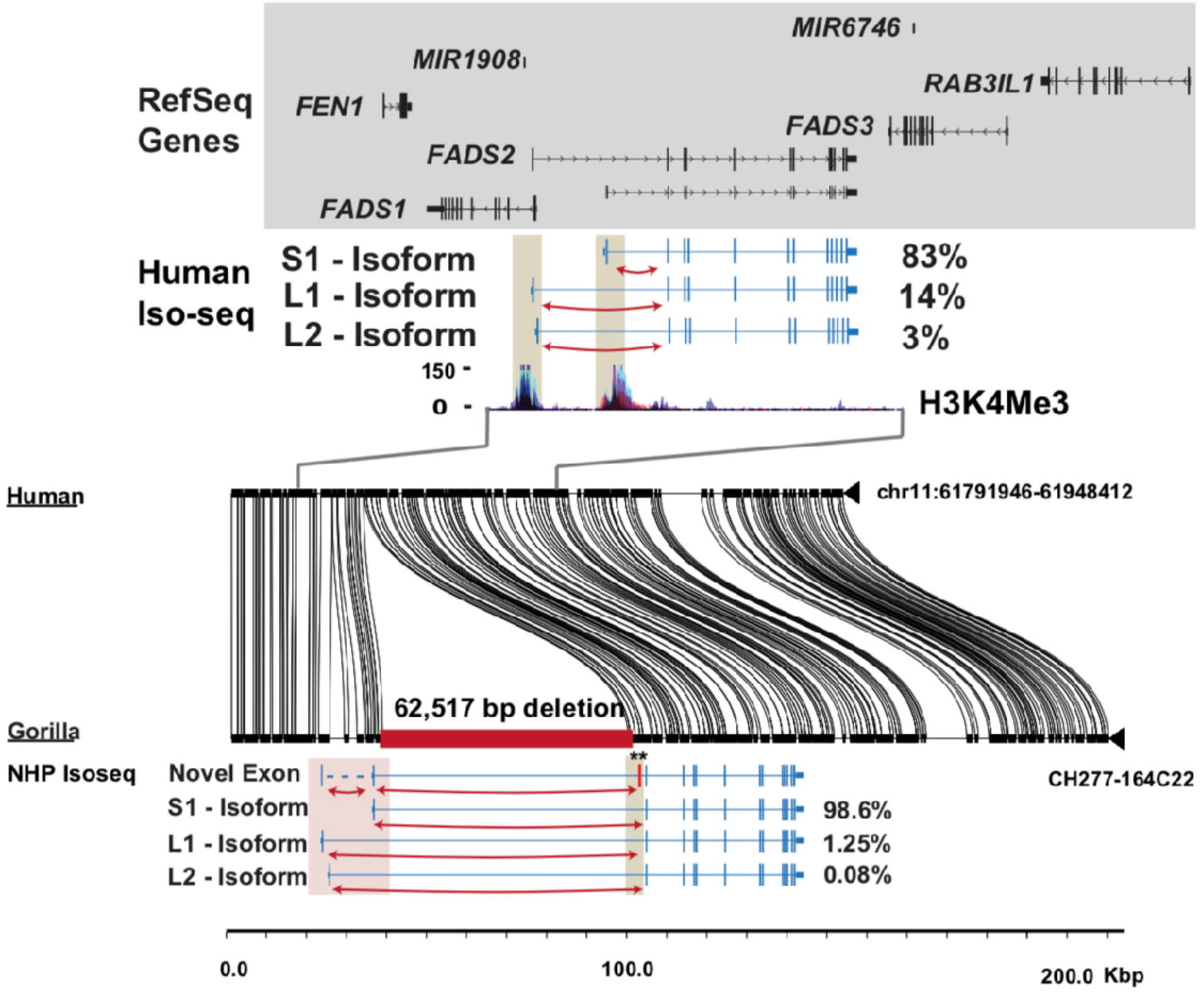
Cory Y. McLean, Philip L. Reno, Alex A. Pollen, Abraham I. Bassan, Terence D. Capellini, Catherine Guenther, Vahan B. Indjeian, Xinhong Lim, Douglas B. Menke, Bruce T. Schaar, Aaron M. Wenger, Gill Bejerano ✉ & David M. Kingsley ✉



A novel human specific deletion of *CARD8*



A simple and yet complex human specific deletion in the *FADS* gene cluster

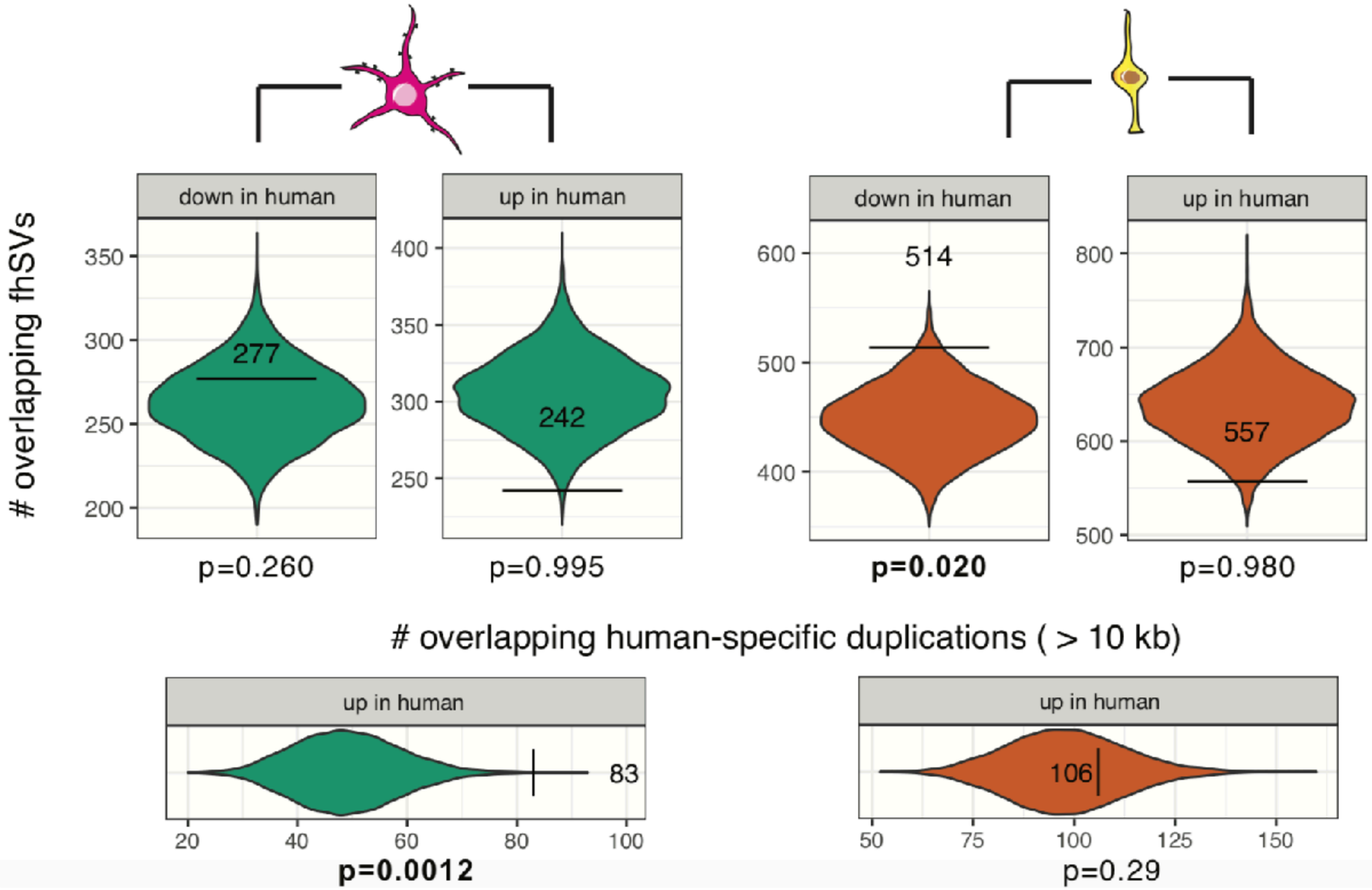


HUMAN GENETICS

Greenlandic Inuit show genetic signatures of diet and climate adaptation

Matteo Fumagalli,[1,2*] Ida Moltke,[3*] Niels Grarup,[4] Fernando Racimo,[2] Peter Bjerregaard,[5,6] Marit E. Jørgensen,[5,7] Thorfinn S. Korneliussen,[8] Pascale Gerbault,[1,9] Line Skotte,[3] Allan Linneberg,[10,11,12] Cramer Christensen,[13] Ivan Brandslund,[14,15] Torben Jørgensen,[10,16,17] Emilia Huerta-Sánchez,[18] Erik B. Schmidt,[17,19] Oluf Pedersen,[4] Torben Hansen,[4†] Anders Albrechtsen,[3†] Rasmus Nielsen[2,20†]

Human specific SVs were associated with down regulated genes in Radial Glia



In conclusion

- SMRT assemblies enable structural variation detection and comparative genomics
- SMRTie-SV is an accurate tools for identifying structural variation
- We identified an association between gene regulation and structural variation in radial glia and excitatory neurons
- There are several human specific variants that warrant further investigation

Acknowledgements

Salk Institute

Ahmet Denli
Rusty Gage

*N*ashU

Tina Graves-Lindsay
Wes Warren
Rick Wilson

UW

Max Dougherty
Brad Nelson
Shwetha Murali
David Gordon
Stuart Cantsilieris
Mark Chaisson
AnneMarie Welch
Kendra Hoekzema
John Huddleston
Melanie Scofield
Naheed Arang
Carl Baker
Vy Dang
Katy Munson
Chris Hill
Jay Shendure
Evan Eichler

UCSC

Ian Fiddes
Benedict Paten
Mark Diekhans
David Haussler

UCSF

Alex Pollen
Olivia Meyerson
Arnold Kriegstein

bionano GENOMICS

Alex Hastie
Andy Pang

PACBIO®

Elizabeth Tseng
Jason Chin

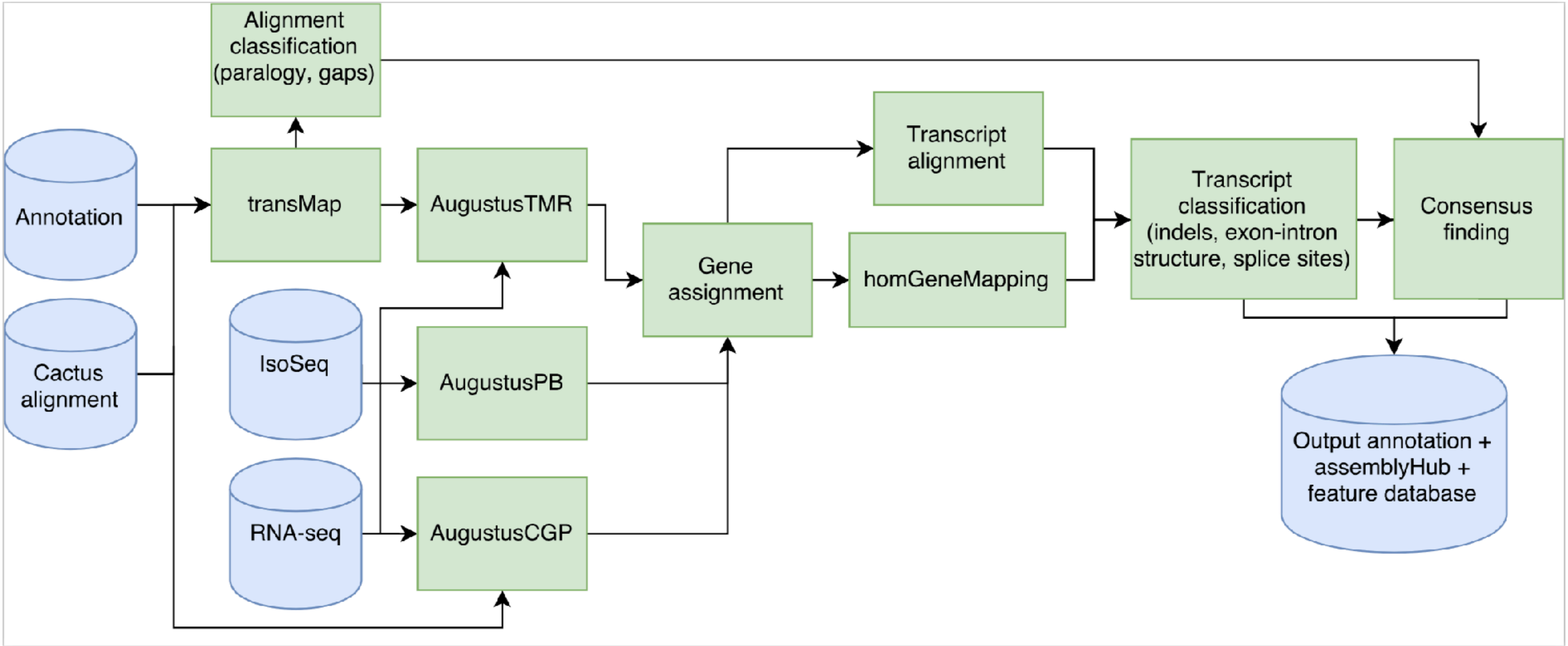
Funding

NIH National Institute of Allergy and Infectious Diseases
U24

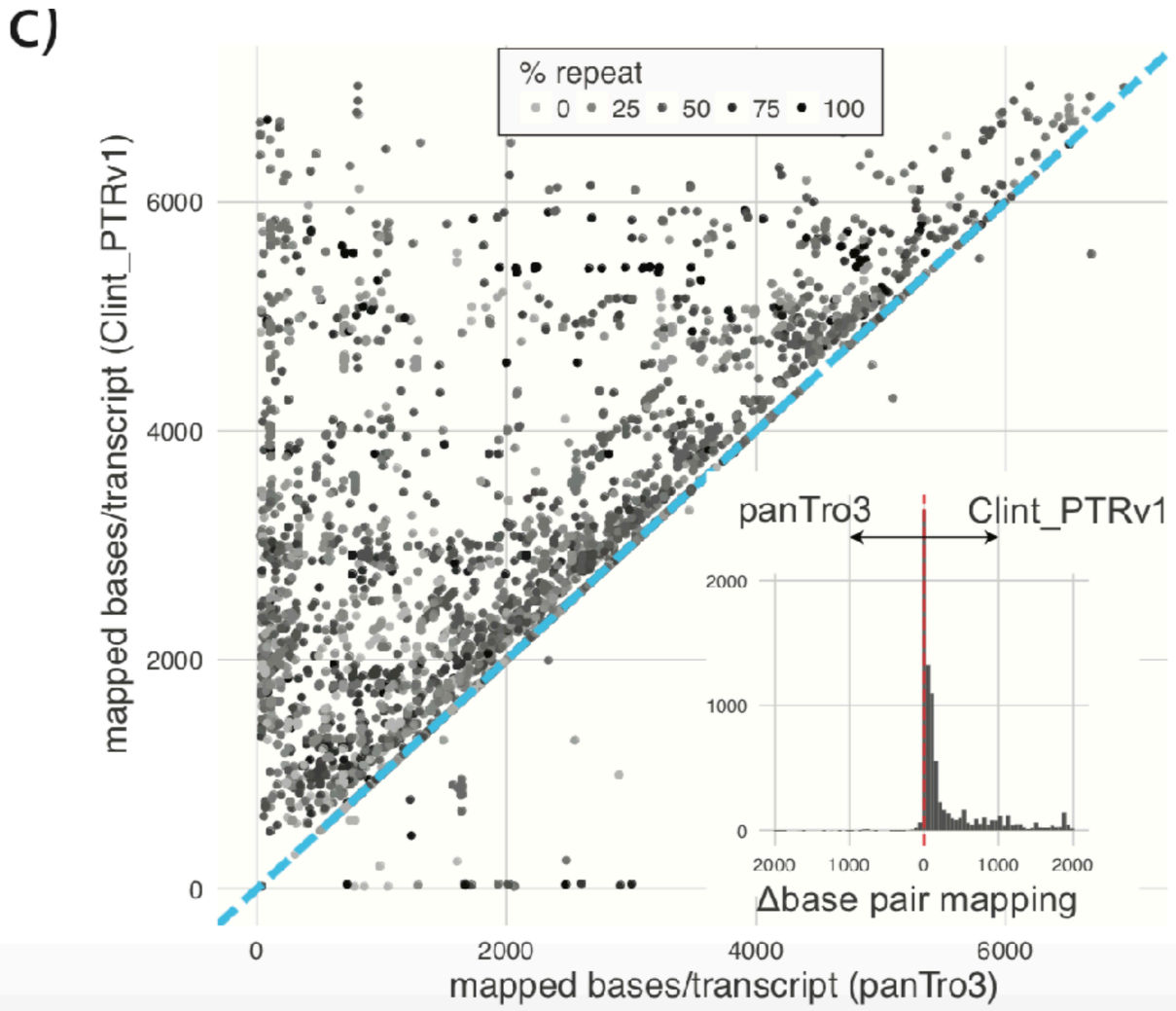
hhmi

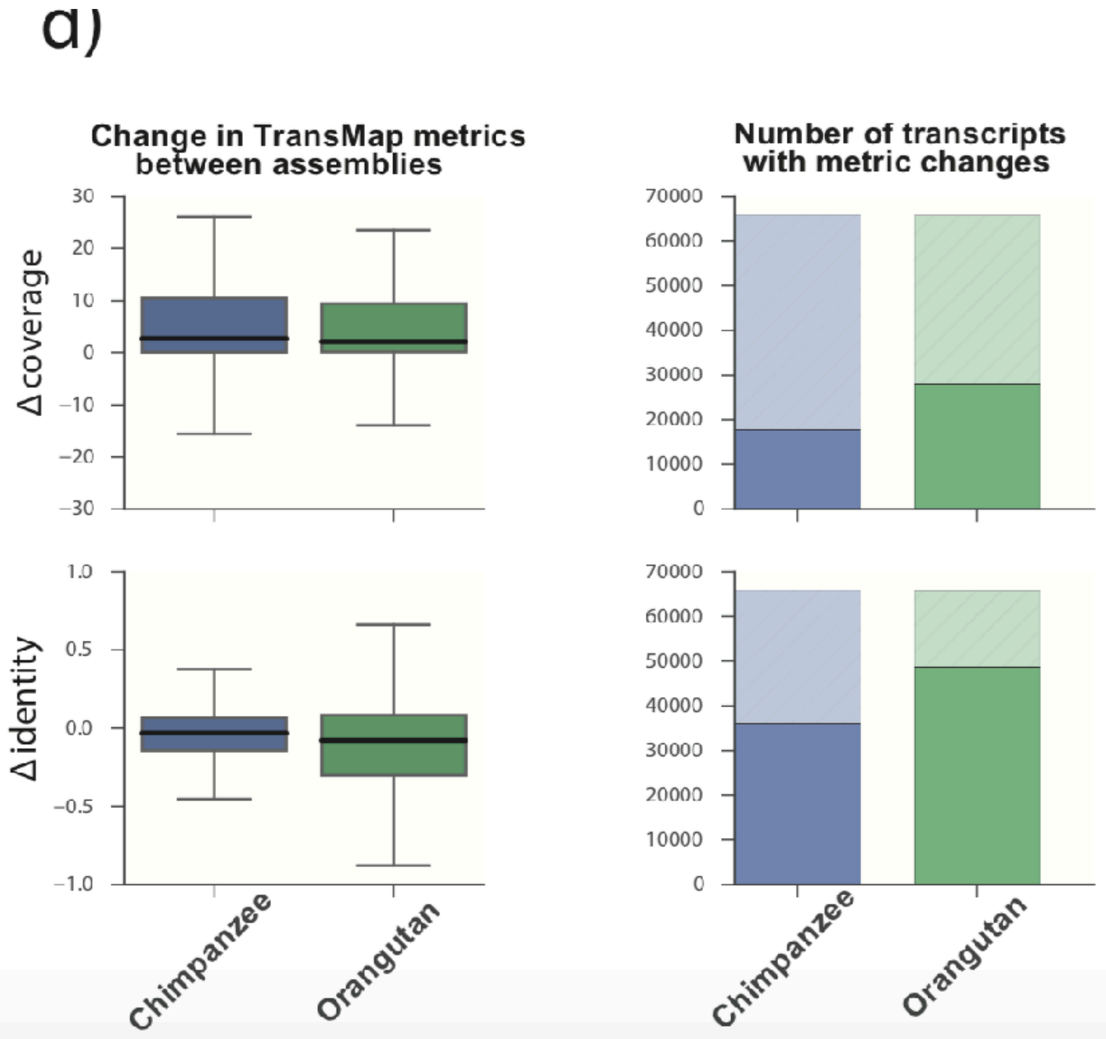
Questions?

Gene annotation using CAT and IsoSeq



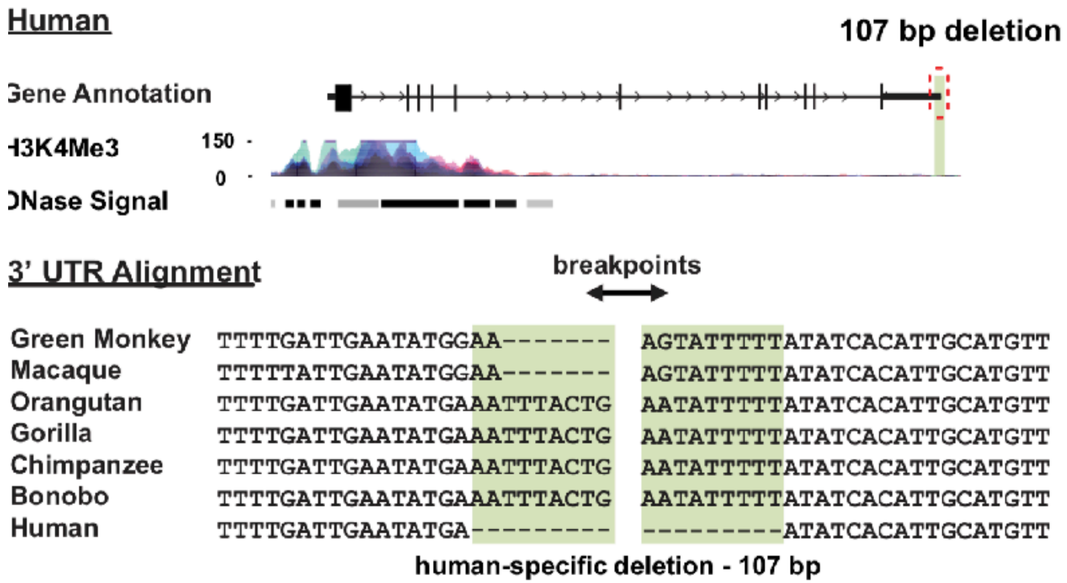
Improvements in transcript mapping and annotations



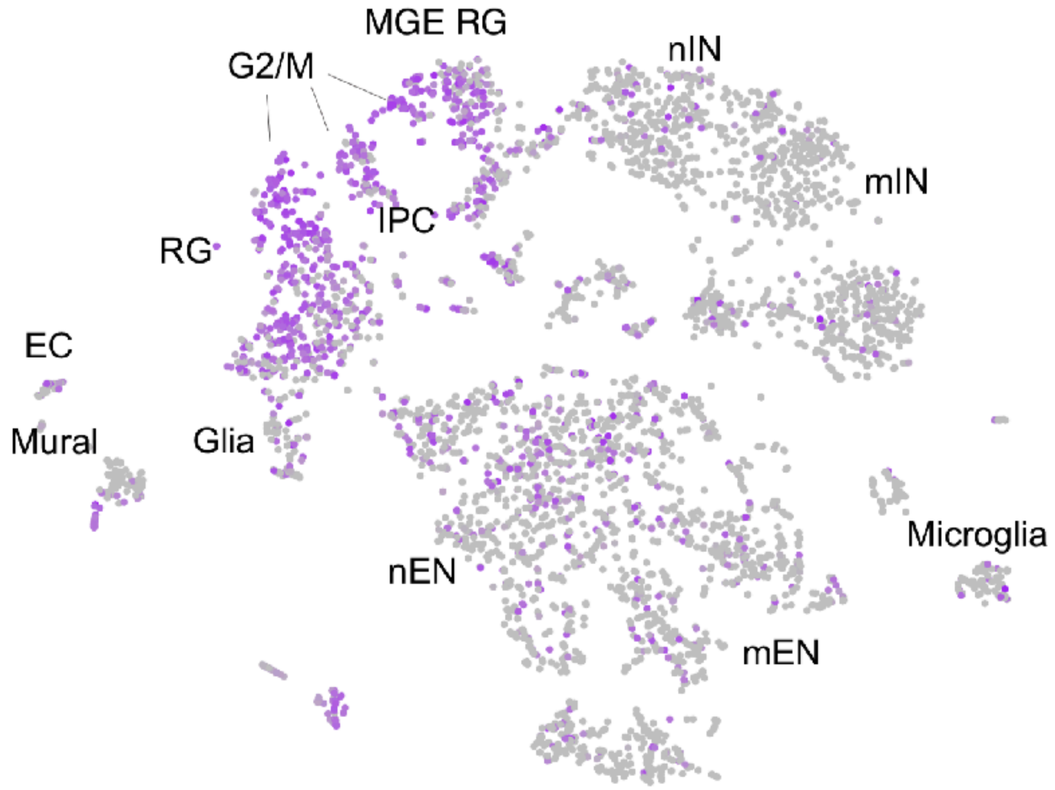


A significant association between human specific structural variants and neuronal gene expression

Example: *WEE1* is unregulated in cortical radial glia

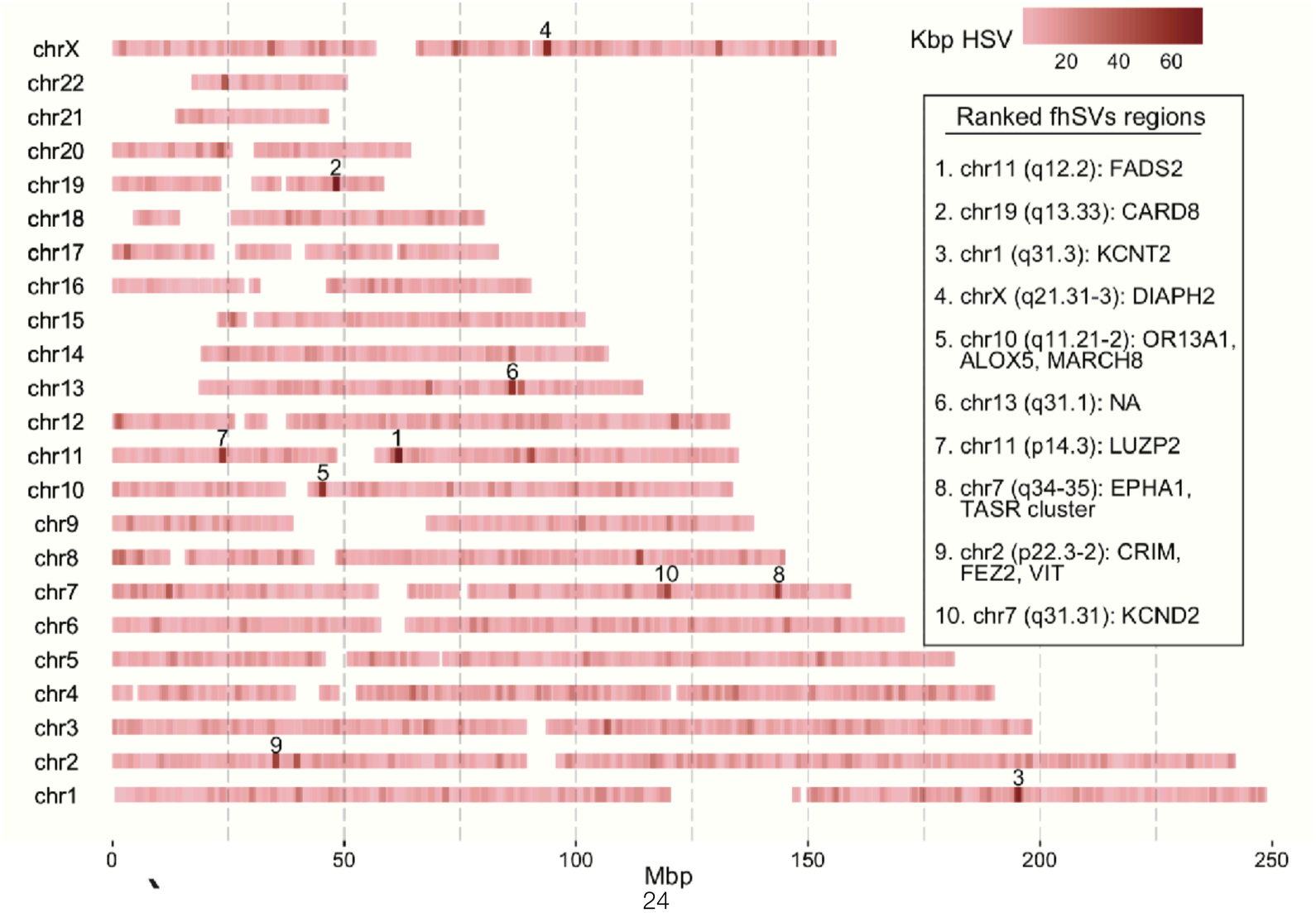


WEE1



Chimpanzee

A map of human specific structural variation reveals hotspots



PanTro5 false inversion