Reconstruction of Human Genome Evolution Predicts an Extensively Changed Neurodevelopmental Gene
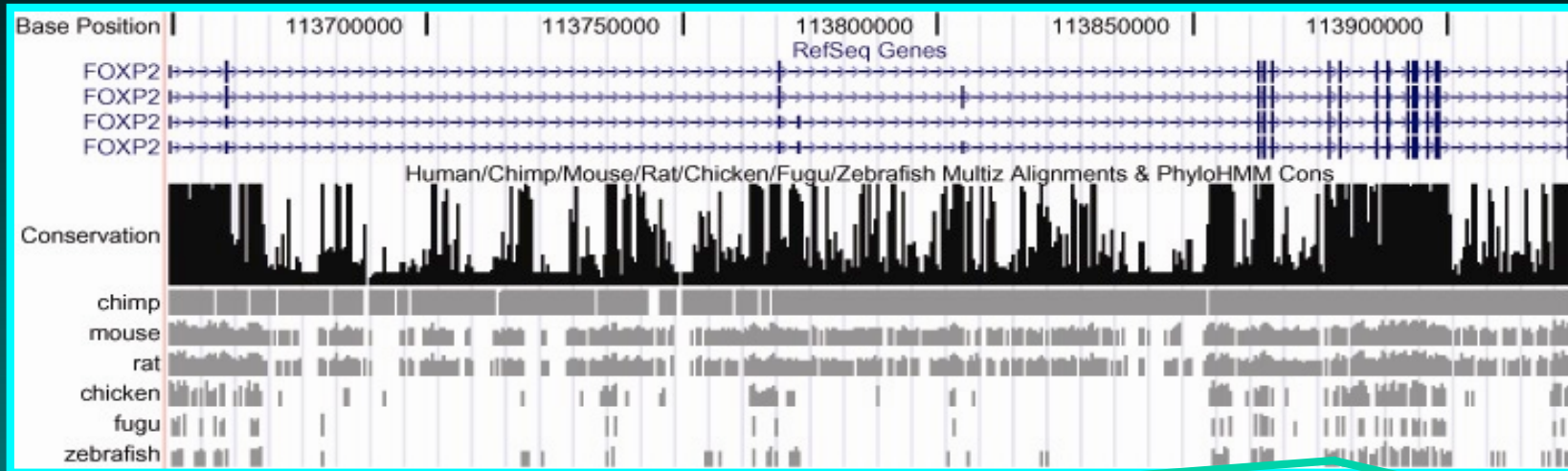
David Haussler

Howard Hughes Medical Institute

Center for Biomolecular Science and Engineering
University of California, Santa Cruz

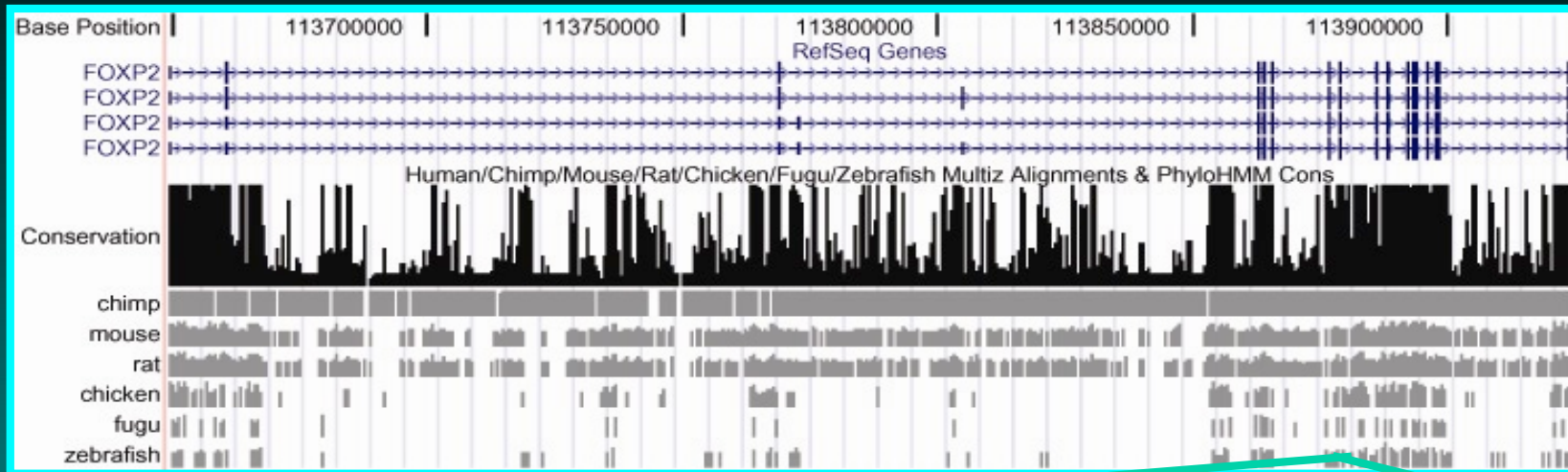# Neutral drift: a genetic change that does not affect the organism

- Mutations occur all the time in protein-coding regions; some do not change the protein, so do not affect the fitness of the organism

- Changing the third DNA base in this codon does not change the amino acid it encodes, alanine (A)

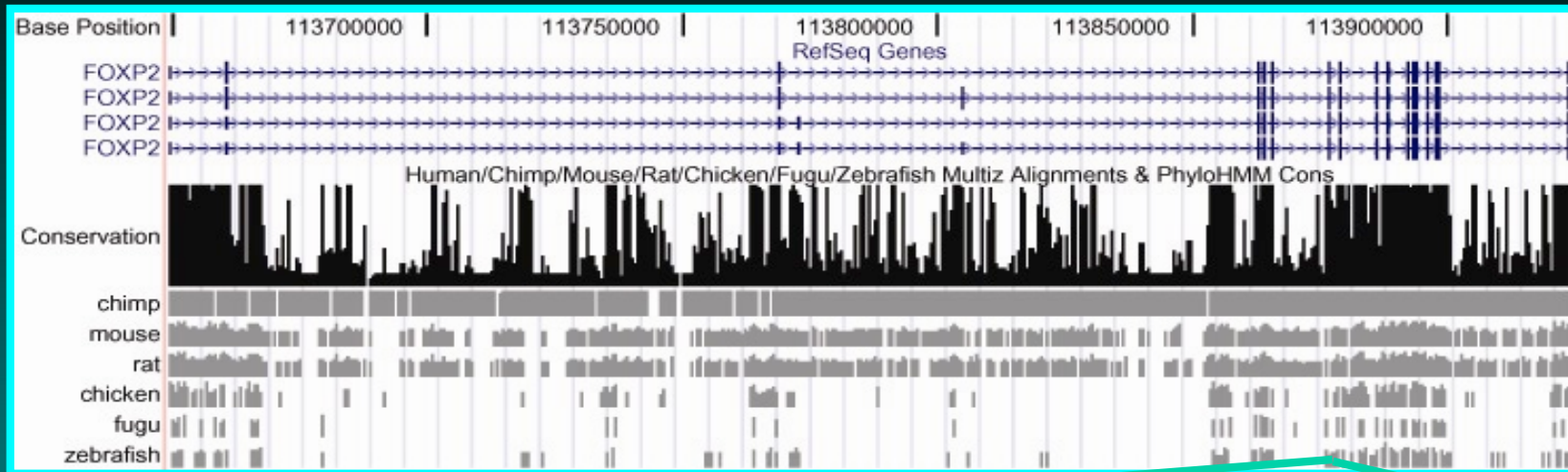Browser: Kent et al; conservation track: Siepel and Rosenbloom

Negative selection: rejecting a change that decreases fitness

- Some mutations would change the protein and thereby reduce fitness
- Such changes are rejected by natural selection, and the DNA is conserved

Browser: Kent et al; conservation track: Siepel and Rosenbloom

Positive selection: a genetic change that increases fitness

- Some mutations have a positive effect: This change from C to A in the gene FOXP2 changed the amino acid from threonine (T) to asparagine (N), which may have improved fitness

- Possible role in the evolution of speech

Browser: Kent et al; conservation track: Siepel and Rosenbloom; FOXP2 results: Enard et al, *Nature*, 2002

About 5% of the human genome is conserved enough with mouse to indicate negative selection to preserve function

Scaled density for human-mouse conservation levels in 50bp windows of orthologous transposon DNA

Density for human-mouse conservation levels in all 50bp windows

blue - red = 5% of genome

K. Roskin, R. Weber, F. Chiaromonte
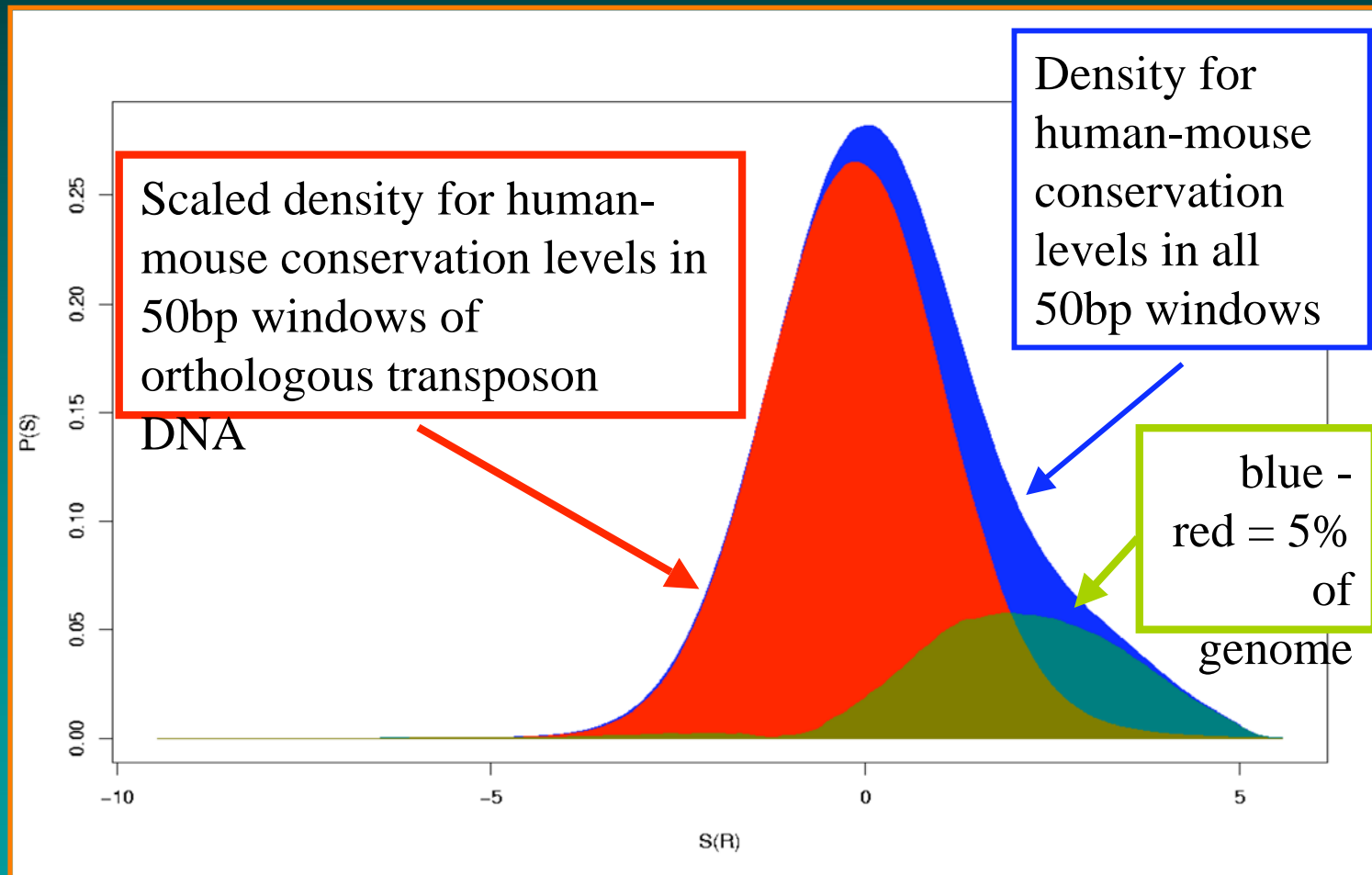
We found evidence for recent positive selection in a novel gene expressed during brain development

Human Accelerated Region 1

Katie Pollard and Sofie Salama

Timing, population genetics

- All 18 changes are fixed in the human population; no evidence of recent selective sweep

- Changes probably occurred between 1 and 7 million years ago

Sol Katzman, Andy Kern, Katie Pollard

Corresponding mouse transcripts

Sofie Salama

This gene produces a structural RNA sequence, not a protein

New interactions in the human version of this gene
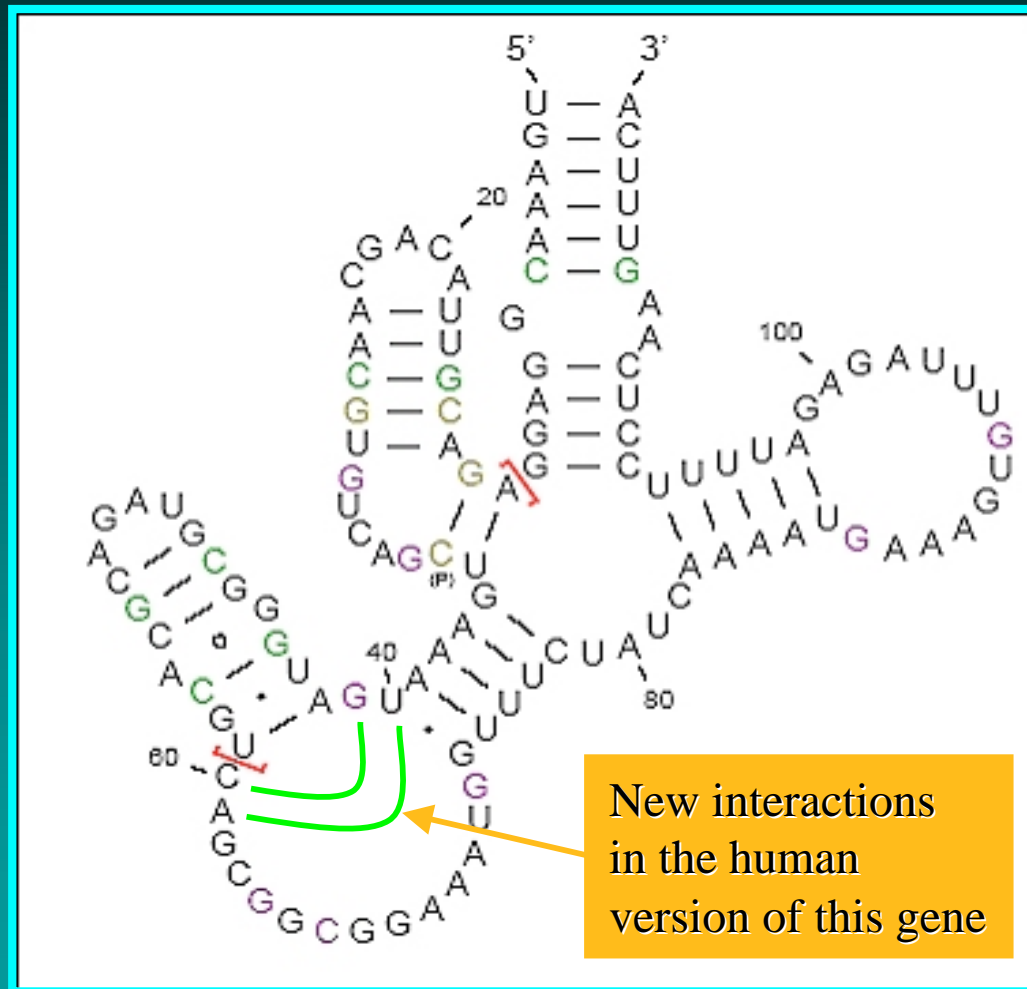
Computational prediction of structure conserved throughout amniotes

Jakob Pedersen

New human RNA structure verified by dimethyl sulfate probing and primer extension
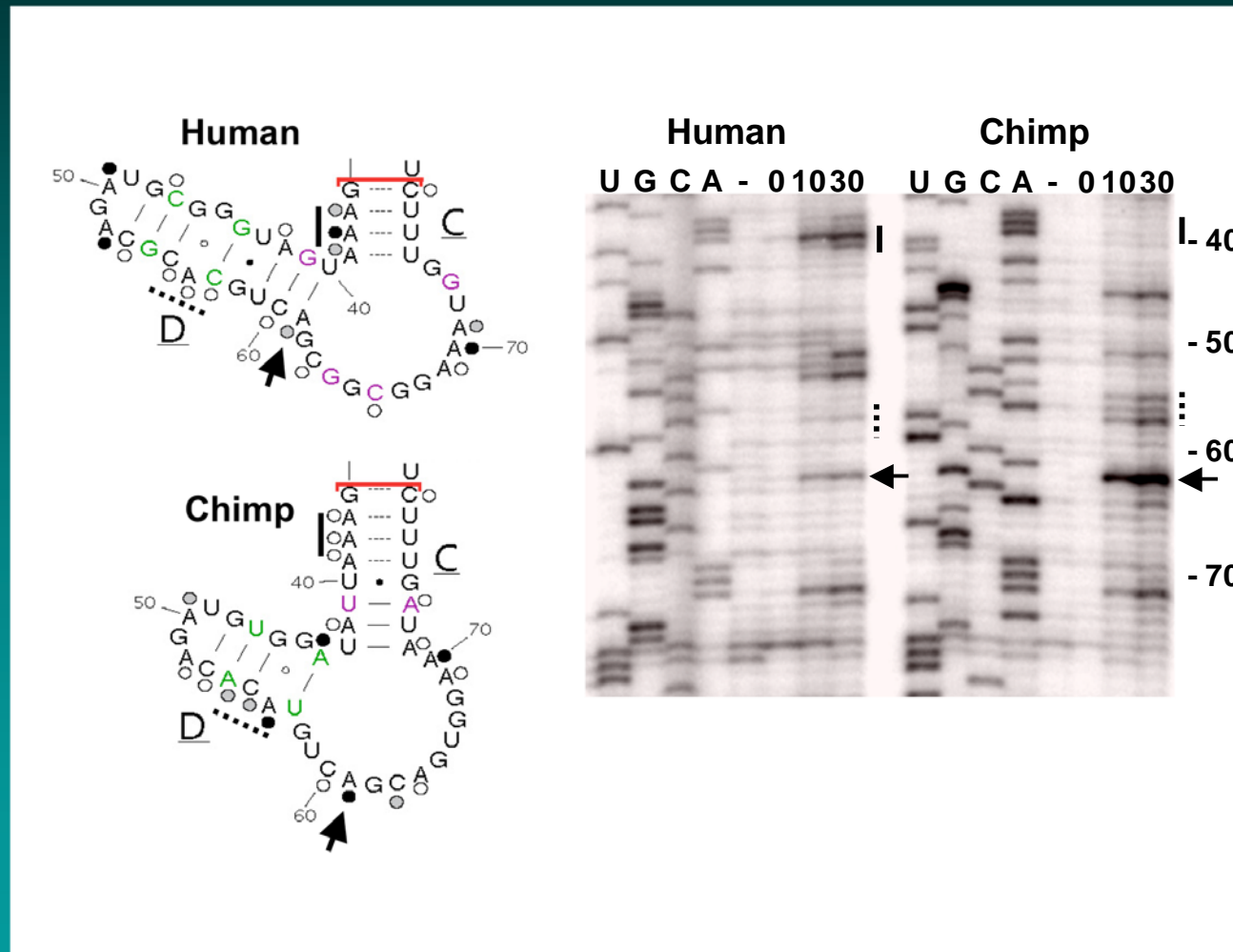
Haller Igel, Manny Ares, Jakob Pedersen, Sofie Salama

In human, HAR1 is expressed in adult brain and testes; HAR1F >> HAR1R in fetal brain

| Sample	HAR1F	HAR1R
cerebral cortex total RNA	1.0 (0.77, 1.29)	0.095 (0.054, 0.17)
frontal lobe total RNA	1.2 (0.70, 2.0)	0.12 (0.067, 0.22)
temporal lobe total RNA	0.72 (0.55, 0.95)	0.049 (0.029, 0.083)
parietal lobe total RNA	0.77 (0.61, 0.98)	0.083 (0.052, 0.13)
occipital pole total RNA	1.10 (0.92, 1.31)	0.12 (0.065, 0.21)
insula total RNA	0.91 (0.62, 1.32)	0.078 (0.045, 0.13)
hippocampus total RNA	0.65 (0.44, 0.96)	0.051 (0.031, 0.087)
pons total RNA	0.51 (0.35, 0.76)	0.12 (0.094, 0.14)
medulla oblongata total RNA	0.39 (0.27, 0.56)	0.043 (0.025, 0.074)
fetal brain total RNA	0.14 (0.11, 0.18)	0.003 (0.002, 0.005)
brain total RNA	0.96 (0.71, 1.3)	0.024 (0.015, 0.039)
testes total RNA	0.12 (0.10, 0.15)	0.12 (0.047, 0.31)
thalamus poly A RNA	4.9 (3.2, 7.5)	0.51 (0.27, 0.98)
hypothalamus poly A RNA	4.8 (3.5, 6.6)	0.54 (0.30, 0.96)

RNA levels of HAR1 relative to the level in cerebral cortex
A value of 1 is equivalent to 24,000 copies in 100 ng of template cDNA

Courtney Onodera, Bryan King

The dominant structure in the human brain is the cerebral cortex
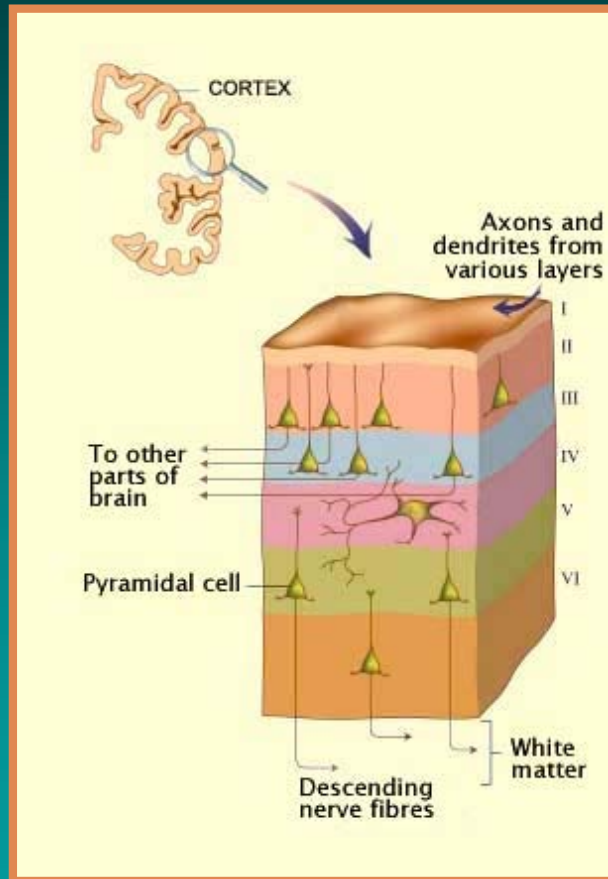
During development, the cerebral cortex is built "inside-out" by neurons migrating radially from the subventricular zone to the pial surface. This process is guided by the neurodevelopmental gene Reelin

Image: www.thebrain.mcgill.ca

The HAR1F gene is specifically expressed in the developing cerebral cortex
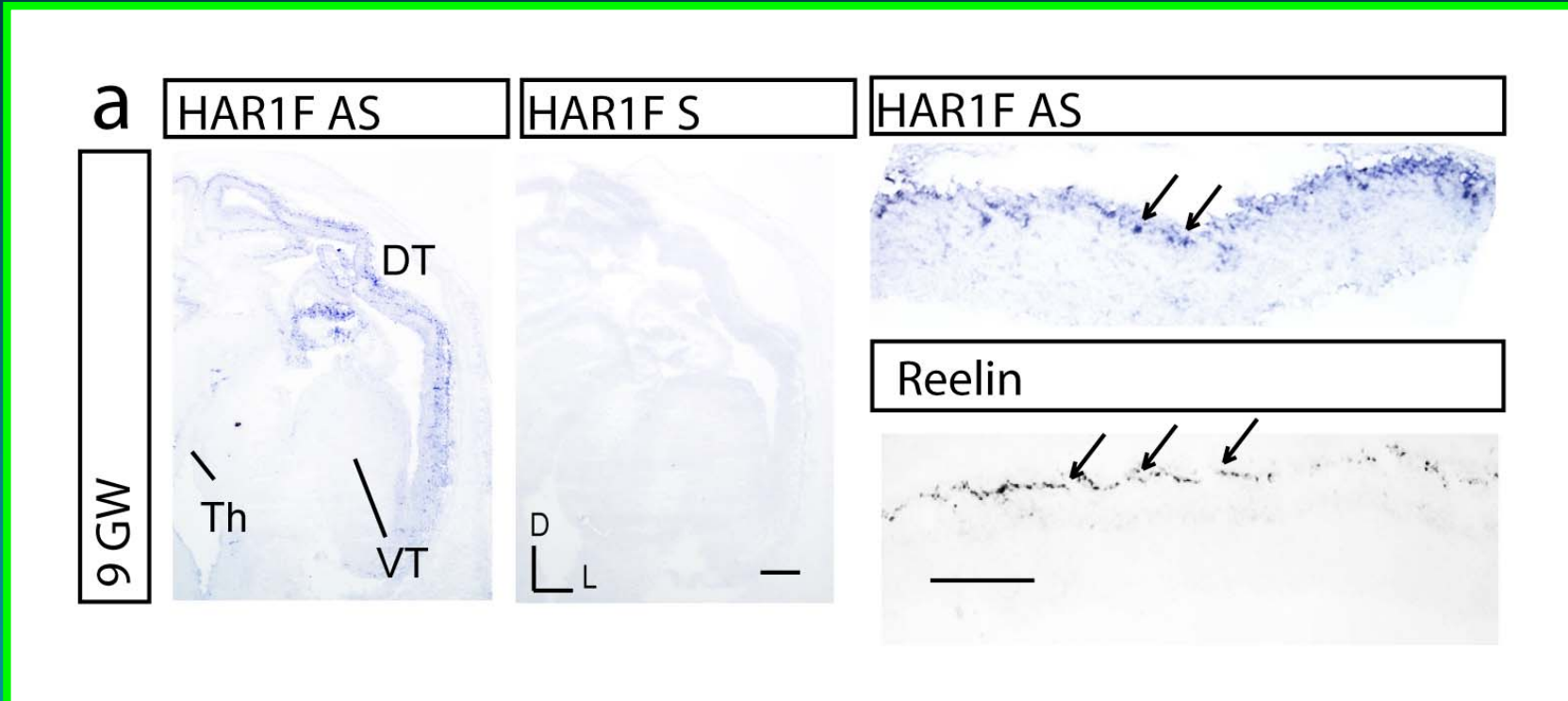
The gene HAR1F is expressed in Cajal-Retzius pyramidal neurons in the dorsal telencephalon, along with Reelin

Nelle Lambert, Marie-Alexandra Lambot, Sandra Coppens, Pierre Vanderhaeghen

Could it have something to do with the evolution of the human brain?

There are 48 other human accelerated regions

What do they do?

A bigger challenge:
Can we reconstruct the evolutionary history of every base of human DNA?

- The history of most bases can be reconstructed with at least 98% accuracy back about 100 million years, to the origins of placental mammals in the Cretaceous period

- In genes and ultraconserved noncoding regions, accurate reconstruction may be possible back >500 million years, to the Cambrian explosion of animal life

Webb Miller, Jian Ma, Brian Raney, Gill Bejerano, Mathieu Blanchette, Adam Siepel, Bernard Suh, Krishna Roskin, Broad Genome Center, Baylor Genome Center, WashU
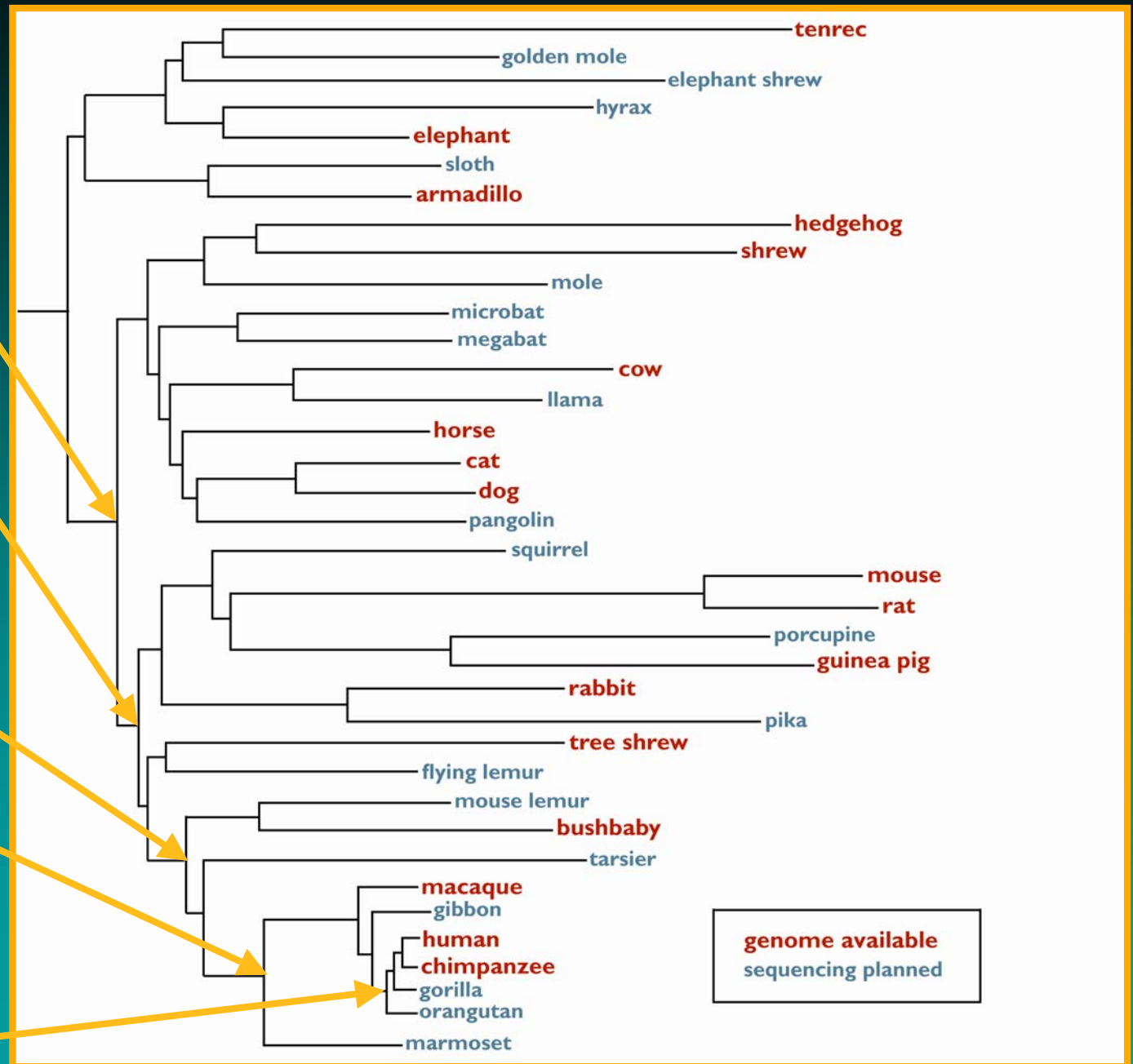
Sequences becoming available

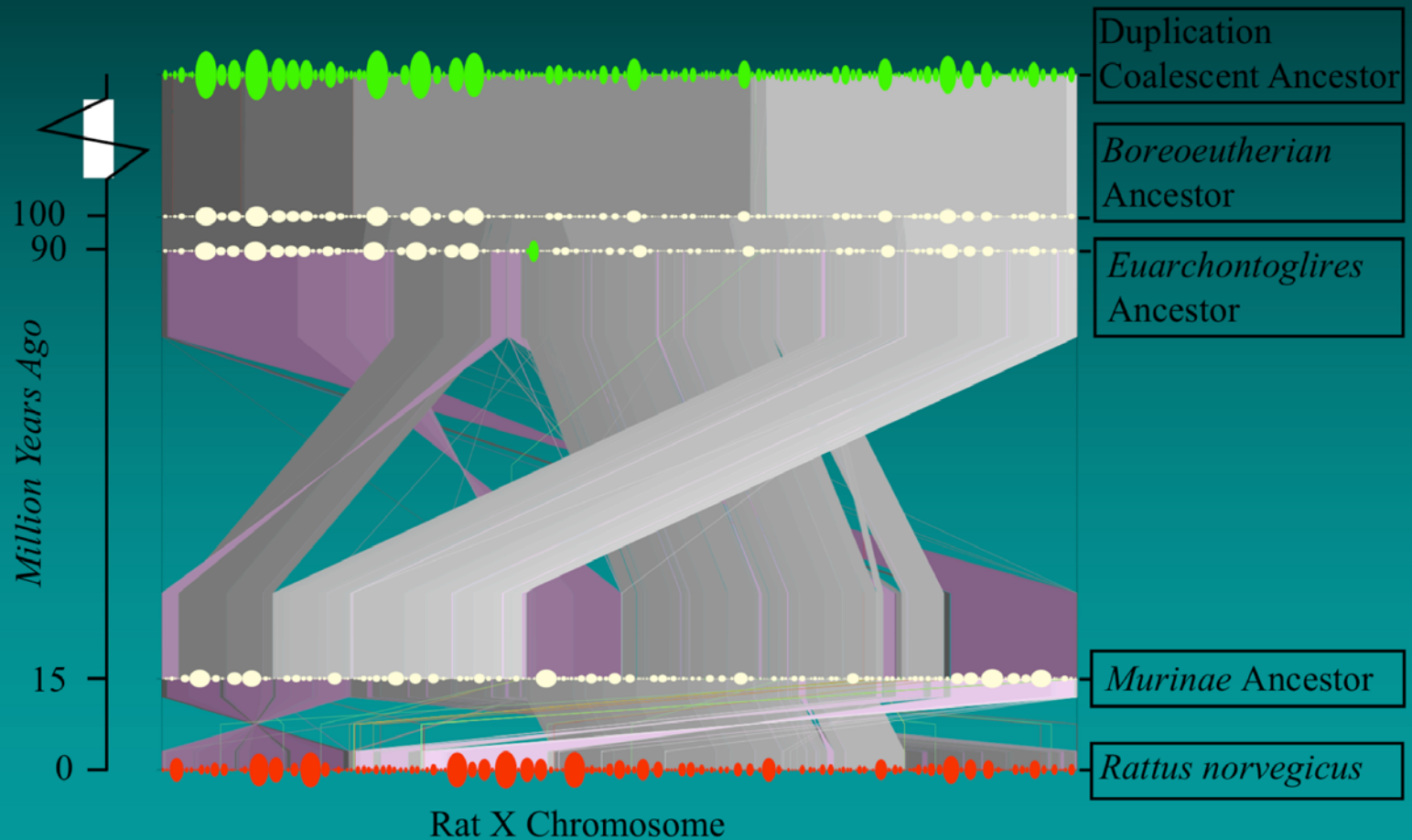**Boreoeutherian ancestor**

Euarchontoglire ancestor

Primate ancestor

Anthropoid ancestor

Hominoid (ape) ancestor

tenrec
golden mole
elephant shrew
hyrax
elephant
sloth
armadillo
hedgehog
shrew
mole
microbat
megabat
cow
llama
horse
cat
dog
pangolin
squirrel
mouse
rat
porcupine
guinea pig
rabbit
pika
tree shrew
flying lemur
mouse lemur
bushbaby
tarsier
macaque
gibbon
human
chimpanzee
gorilla
orangutan
marmoset

genome available
sequencing planned

Michelle Clamp

Example: evolutionary history of a mammalian chromosome

Duplication Coalescent Ancestor

Boreoeutherian Ancestor

Euarchontoglires Ancestor

Murinae Ancestor

Rattus norvegicus

Million Years Ago

100
90

15

0

Rat X Chromosome

History of rat chromosome X   Jian Ma, Bernard Suh, Brian Raney

Reconstructed Boreoeutherian ancestral genome
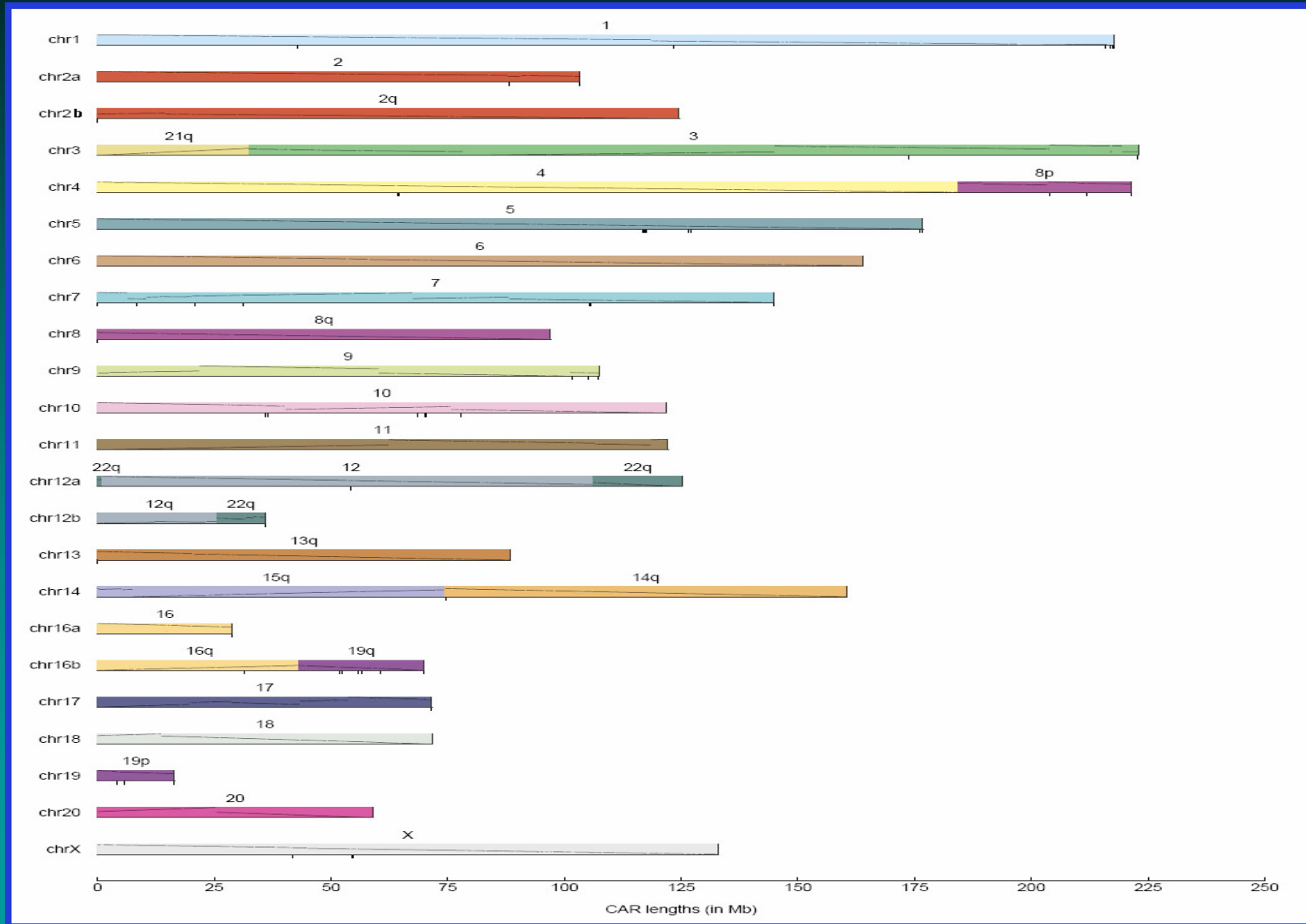
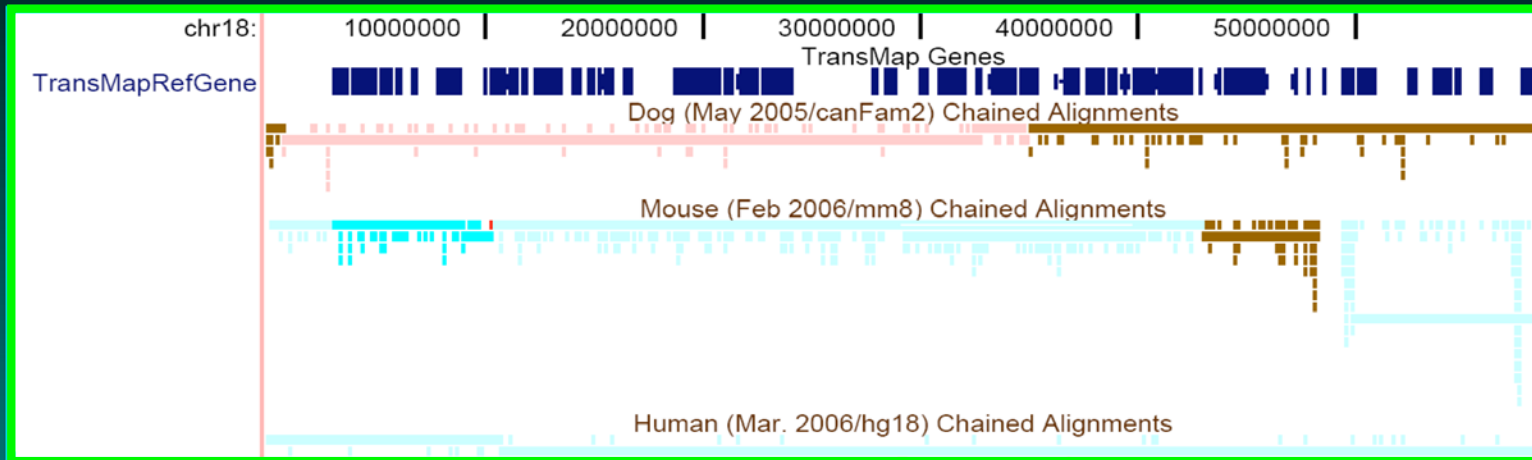# Size and complexity of the ancestral genome

- Human genome has 2.8 billon bases

- The reconstructed ancestral genome has 2.1 billion bases

- 91% of the 23,000 known human genes map to the ancestral genome sequence

Jian Ma, Rico Burhans, Brian Raney, Webb Miller, Mark Diekhans

Boreoeutherian ancestral genome browser

Alignment shows molecular history of human lineage

Boreoeutherian ancestral genome browser

G to A transition

Boreoeutherian ancestral genome browser

A mouse gene maps to this location, but no human gene does

Boreoeutherian ancestral genome browser

Tryptophan (W) changed to a stop codon before the human-chimp ancestor. This may have killed the gene. Proteins of this type (acyltransferase 3) appear in all branches of life; this was the last in the hominid genome.

Grand challenge of human molecular evolution

Reconstruct the evolutionary history of each base in the human genome

- Recognize functional elements from patterns of negative selection

- Find the origins of evolutionary innovations specific to the human lineage

The UCSC Team

Katie Pollard and Gill Bejerano

The wetlab

Sofie Salama

Extended Credits

Thanks to Jim Kent, Robert Baertsch, Galt Barber, Gill Bejerano, Hiram Clawson, Mark Diekhans, Jorge Garcia, Rachel Harte, Angie Hinrichs, Fan Hsu, Donna Karolchik, Sol Katzman, Andy Kern, Bryan King, Robert Kuhn, Victoria Lin, Andre Love, Craig Lowe, Yontao Lu, Jian Ma, Chester Manuel, Courtney Onodera, Jakob Pedersen, Andy Pohl, Brian Raney, Brooke Rhead, Kate Rosenbloom, Krishna Roskin, Sofie Salama, Zack Sanborn, Kayla Smith, Mario Stanke, Bernard Suh, Paul Tatarsky, Archana Thakkapallayil, Daryl Thomas, Heather Trumbower, Jason Underwood, Ting Wang, Erich Weiler, Chen-Hsiang Yeang, Jing Zhu, and Ann Zweig, in my group at UCSC

And to Nadav Ahituv, Manny Ares, Mathieu Blanchette, Rico Burhans, Michele Clamp, Richard Gibbs, Eric Green, Haller Igel, John Karro, Eric Lander, Kerstin Lindblad-Toh, Webb Miller, Jim Mullikin, Katie Pollard, Tom Pringle, Eddy Rubin, Armen Shamamian, Adam Siepel, Chuck Sugnet, Pierre Vanderhaeghen, and many other outside collaborators