Brainseq Transposons
modENCODE developmental timecourses
3D Structure of the Genome

Roger Alexander

Gerstein lab group meeting

Thurs 11 Aug 2011

Brainseq Transposons

modENCODE developmental timecourses

3D Structure of the Genome

Brainseq Dataset

Region	Acronym
Amygdala	AMY
Cerebellum	CBL (CBC)
Hippocampus	HIP
Striatum	STR
Thalamus	THM (MD)
Primary auditory (A1) cortex	A1C
Inferior temporal cortex	ITC
Posterior inferior parietal cortex	IPC
Primary somatosensory (S1) cortex	S1C
Posterior superior temporal cortex	STC
Primary visual (V1) cortex	V1C
Dorsolateral prefrontal cortex	DFC
Primary motor (M1) cortex	M1C
Medial prefrontal cortex	MFC
Orbital prefrontal cortex	OFC
Ventrolateral prefrontal cortex	VFC

We have 75nt single-end RNAseq reads in 16 brain regions from 6 people.

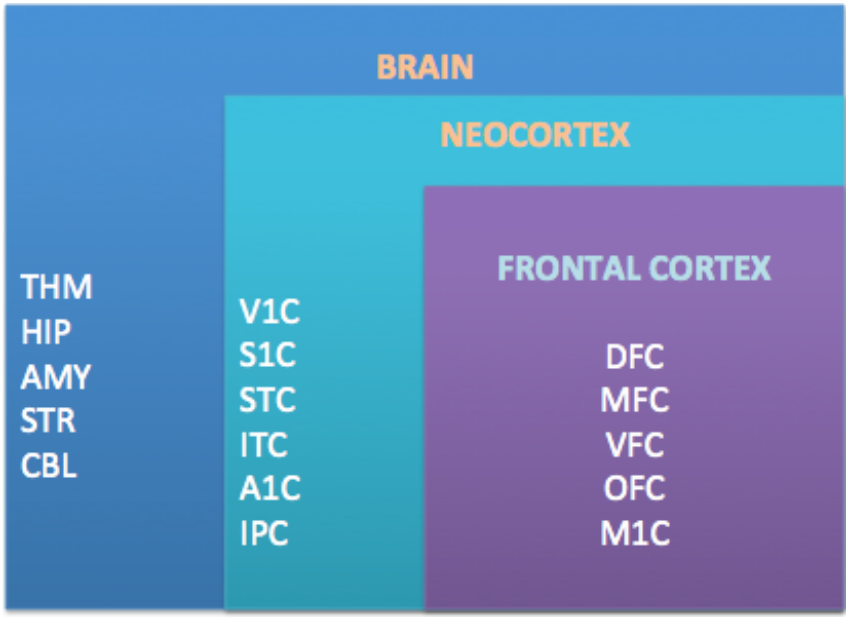


table courtesy of ASB

L1 retrotransposition in human neural progenitor cells

Neural progenitor cells (NPCs) give rise to 3 main lineages of the nervous system:
- neurons
- astrocytes
- oligodendrocytes

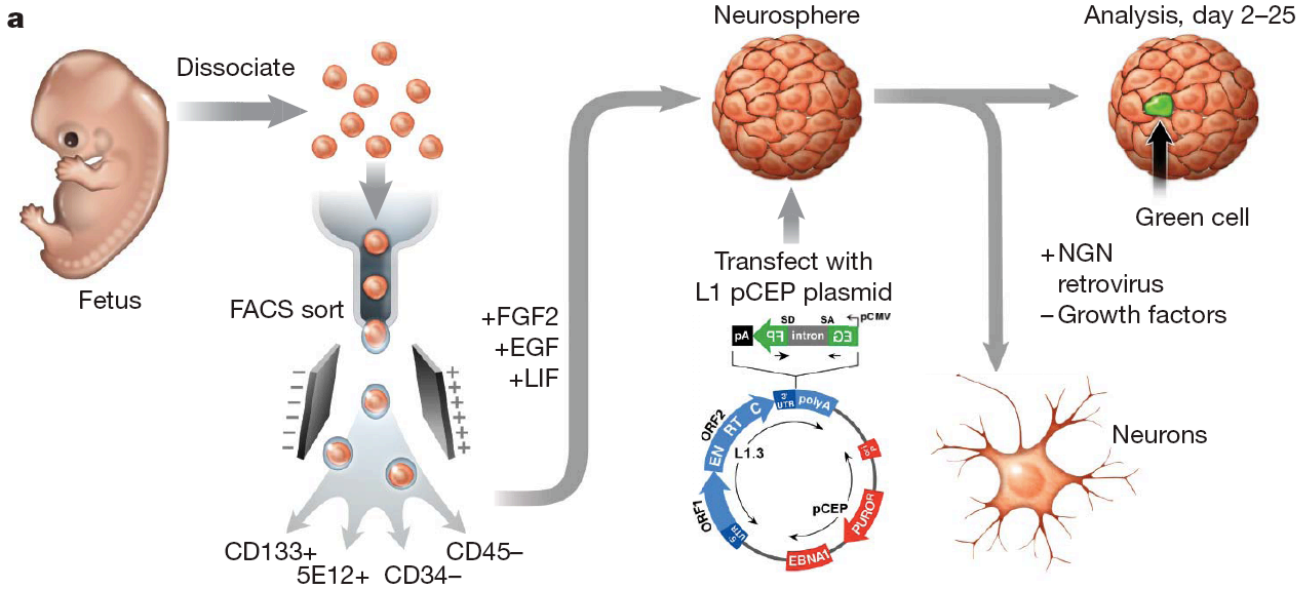
To determine whether human NPCs can support L1 retrotransposition, Gage transfected human fetal brain stem cells (hCNS-SCns) with expression construct containing retrotransposition-competent human L1 (RC-L1) driven from its native promoter (L1RP).

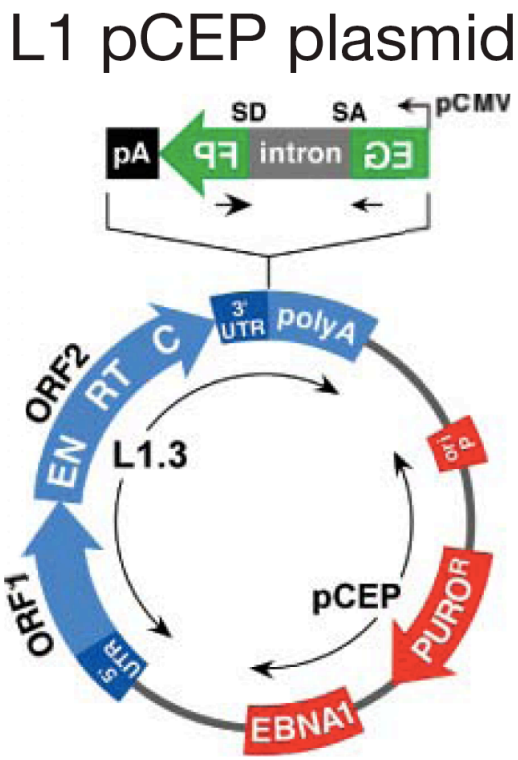
- LINE L1 retrotransposition often occurs at
  - L1 endonuclease consensus cleavage sites (59-TTTT/A and derivatives)
- flanking target site duplications
- 16 of 19 events were <100 kb from a gene, some expressed in neurons

L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127

work by Fred Gage lab, Salk Institute, San Diego

L1 retrotransposition in human neural progenitor cells

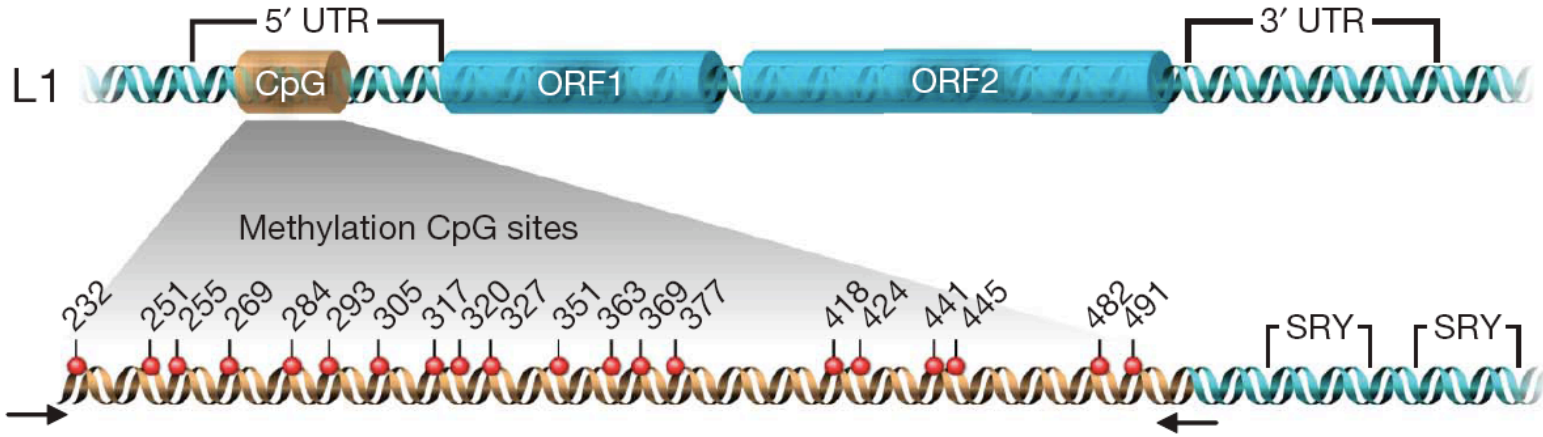


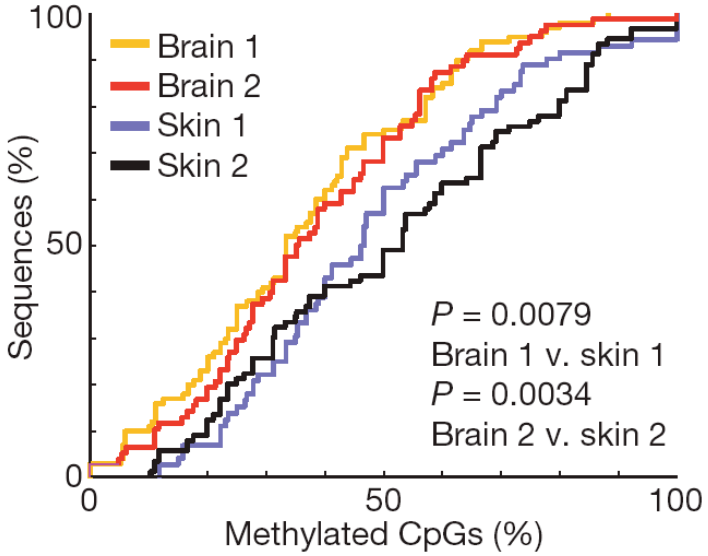


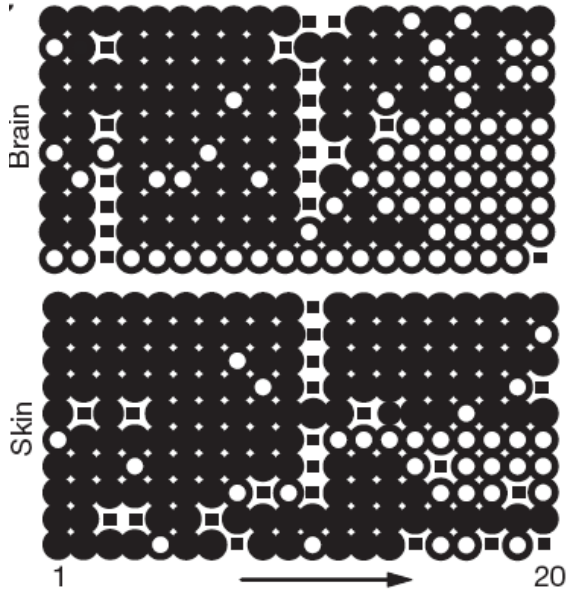
L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127

work by Fred Gage lab, Salk Institute, San Diego

L1 retrotransposition in human neural progenitor cells







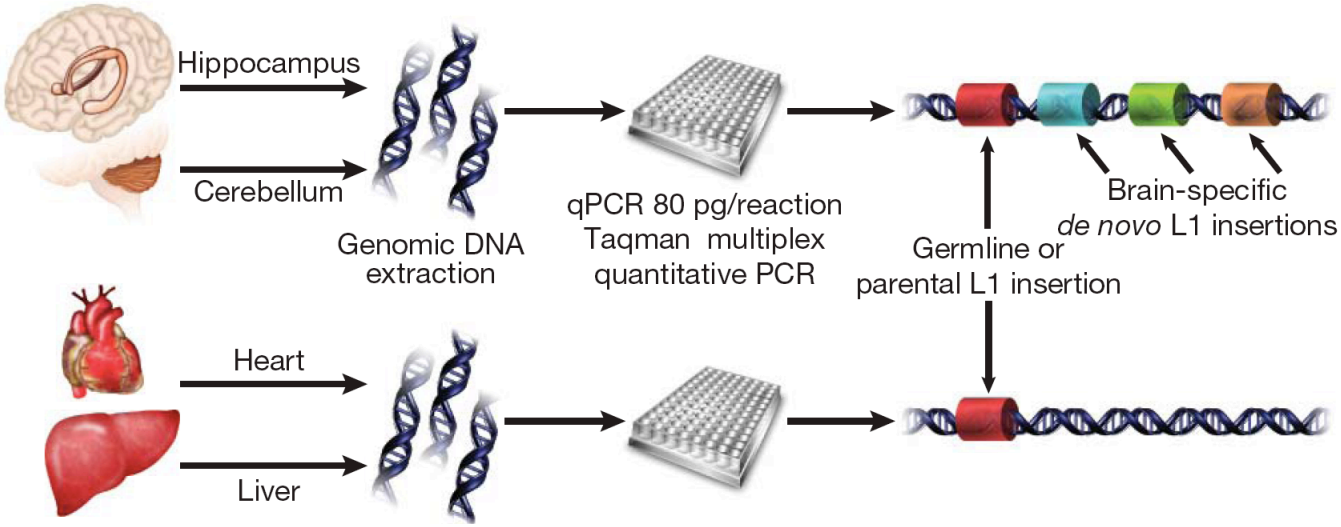
CpG islands in L1 5' UTR
are less methylated
in brain than
in skin tissue
in two fetal samples.
=> higher transcription

L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127

L1 retrotransposition in human neural progenitor cells

"We propose that less L1 promoter methylation in the developing brain may correlate with increased L1 transcription and perhaps L1 retrotransposition, and the differential interaction of SOX2 and MECP2 with L1 regulatory sequences may modulate L1 activity in different neuronal cell types.

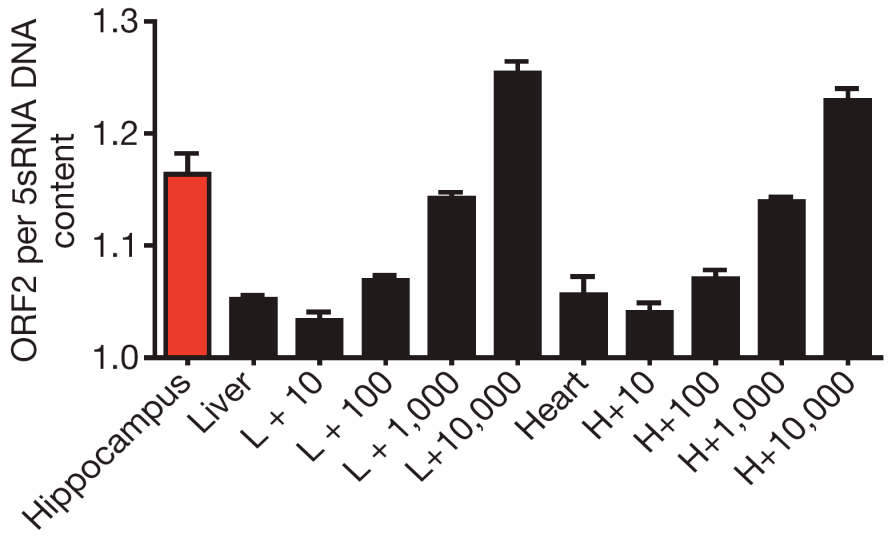
Although NPCs are useful to monitor L1 activity, they only allow monitoring of a single L1 expressed from a privileged context. By comparison, the average human genome contains ~80–100 active L1s, the expression of which may be affected by chromatin structure."



L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127

L1 retrotransposition in human neural progenitor cells

"hippocampus samples contained ~1,000 more L1 copies than the heart or liver genomic DNAs, suggesting a theoretical increase in ORF2 of approximately 80 copies per cell."



Ultimate proof that endogenous L1s are retrotransposing in the brain requires identification of new retrotransposition events in individual somatic cells."

L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127

Jumping of Active Transposons can lead to Somatic Genome Variation during Development

Hypothesis: Active retro-transposons have positioned themselves differently in the somatic genomes of different brain tissues in the same individual.

=> search for transposons present in one region and missing in others in the same individual

L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127
Do Jumping Genes Spawn Diversity?
Science (2011) 332: 300

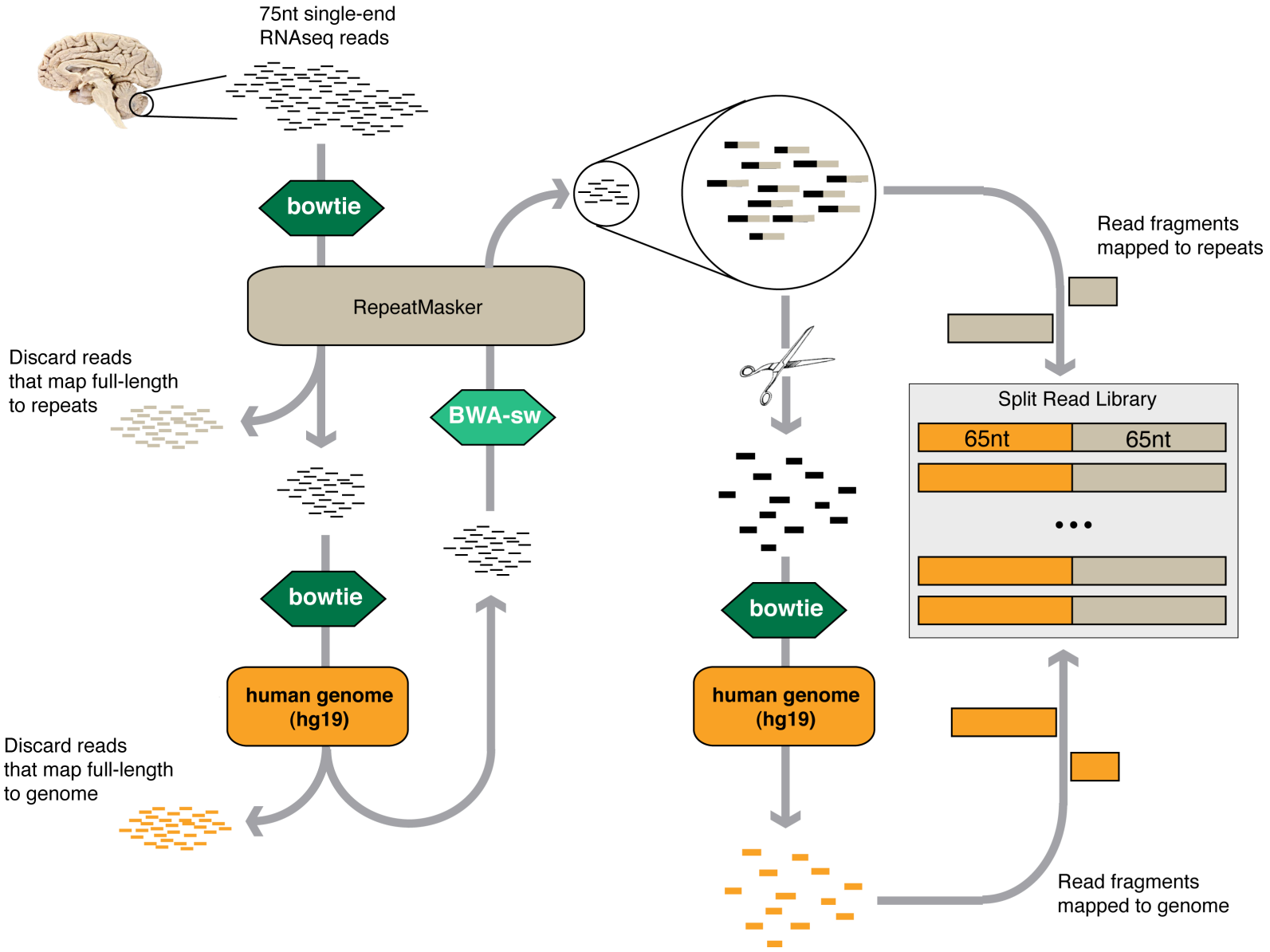
Jumping of Active Transposons can lead to Somatic Genome Variation during Development

Method:

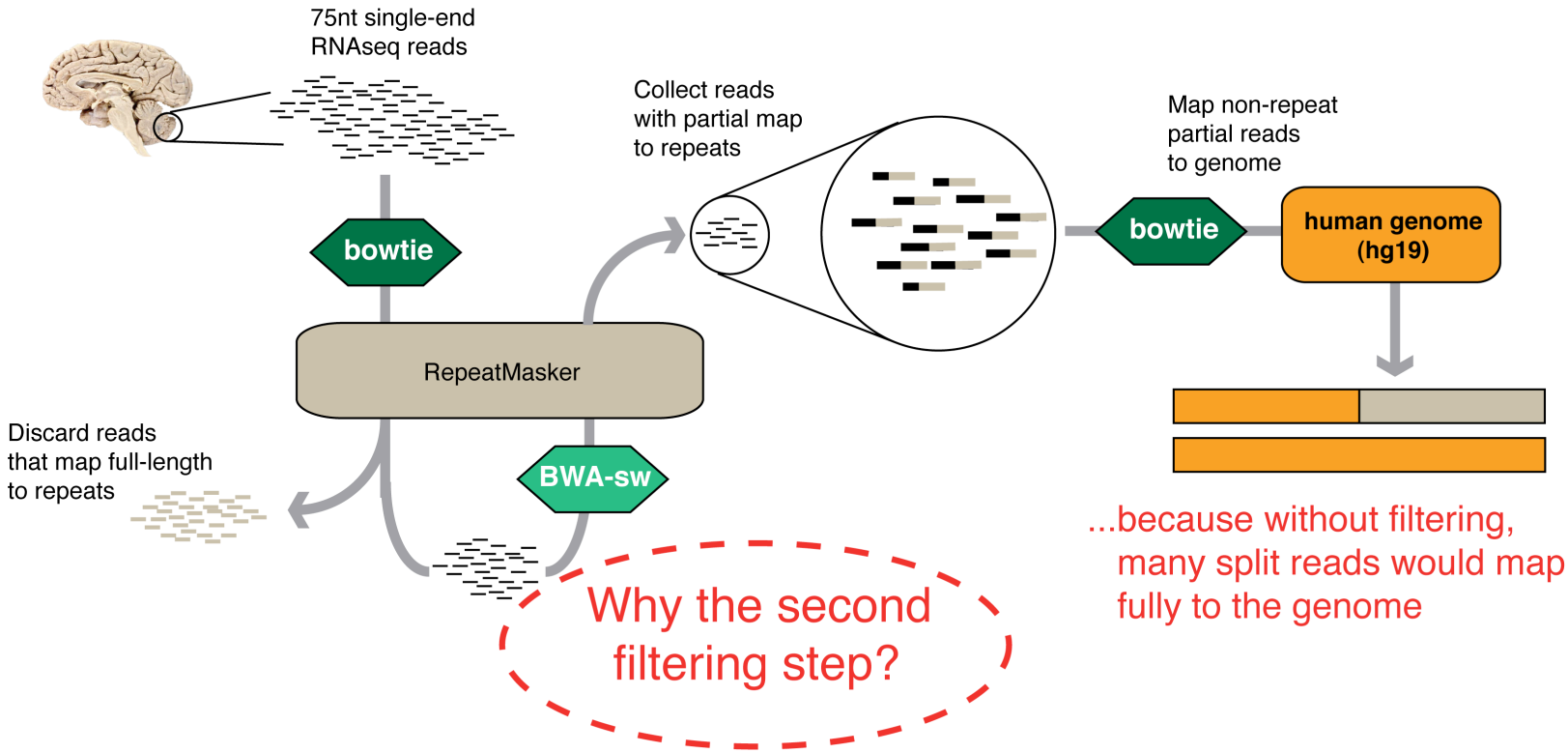
- Start with HSB123 sample RNAseq data for 16 brain regions
- Filter out completely repetitive RNAseq reads (bowtie on hg19 Repeat Masker)
- Filter out RNAseq reads that map completely to genome (bowtie on hg19)
- Scan remaining reads for partial overlap with hg19 Repeat Masker (BWA)
- Map non-repeat portion of split reads against hg19 (bowtie)
- Generate split read library and map all reads from each region against it (bowtie)
- Look for split reads with differential mapping between regions
- Analyze other brain samples – are the retrotransposon insertion sites random or common across individuals (i.e. regulated)?

L1 retrotransposition in human neural progenitor cells
Nature (2009) 460: 1127
Do Jumping Genes Spawn Diversity?
Science (2011) 332: 300

Method: Assemble a library of repeat / non-repeat split reads



Assemble a library of repeat / non-repeat split reads

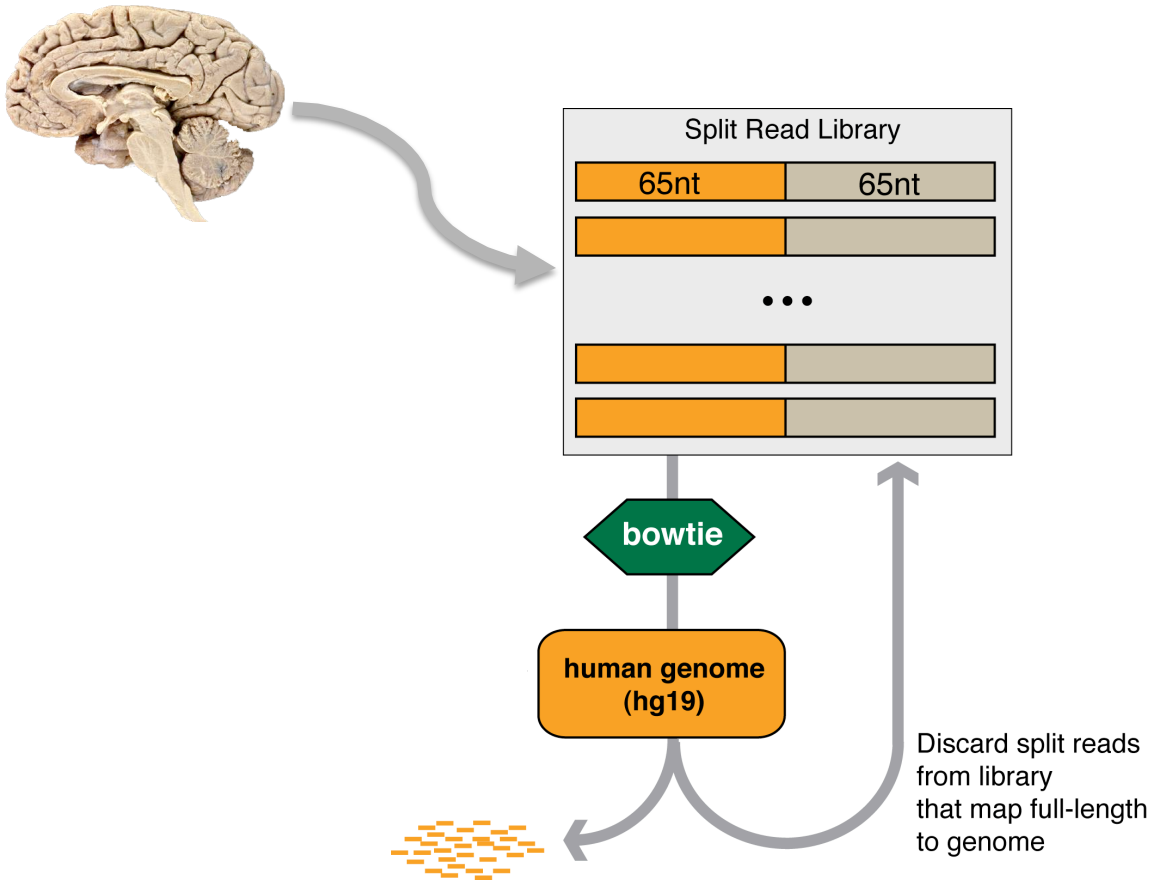


Example

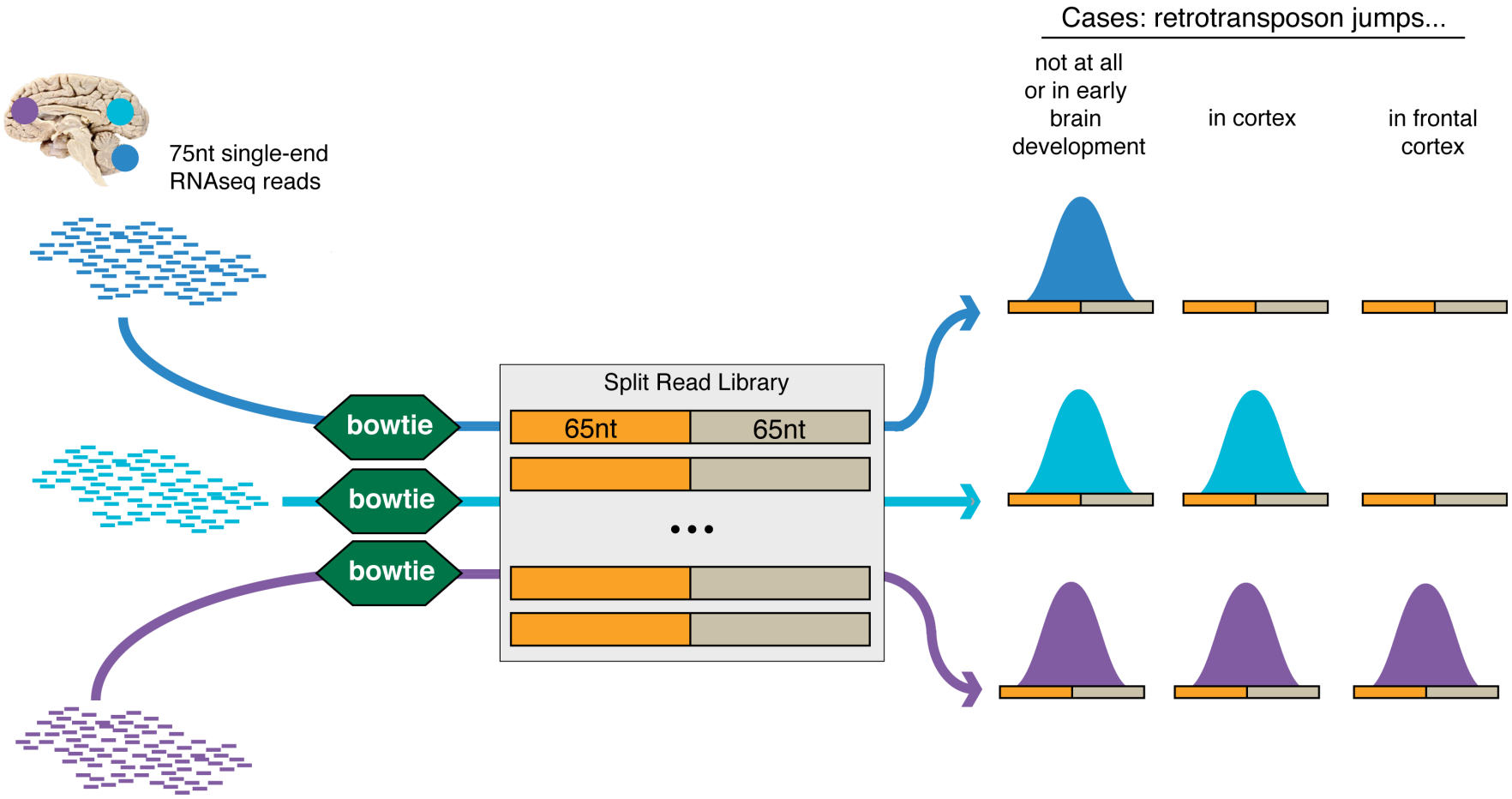
```
RNAseq read       AGAAATCAATGGCATTTGAACCACCAAAAAGAAATAAAGGGCTGAGTGTGGTGGCTCACGCCTGTAATCCCAGCA
match on chr18    AGAAATCAATGGCATTTGAACCACCAAAAAGAAATAAAGGGCTGAGTGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG...
repeat match                                            GGCTGAGTGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG...
(AluSq2_SINE_Alu_4 on chrX)
```

Assemble a library of repeat / non-repeat split reads

One final filtering step allows us to map our read sets
against just the split read library
and still expect ~unique mappings.

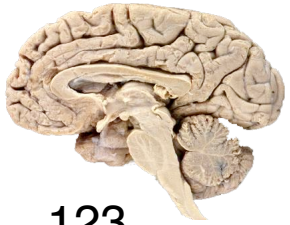


Map reads from different regions to split read library

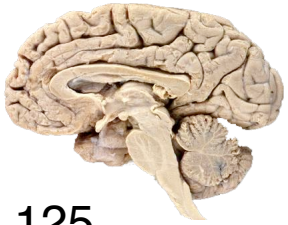


Compare location of retrotranspositon events across brains:

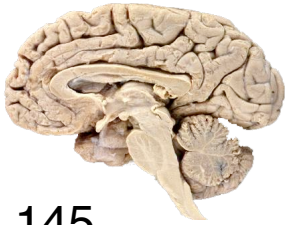
Are the locations shared, and thus somehow regulated, or random?



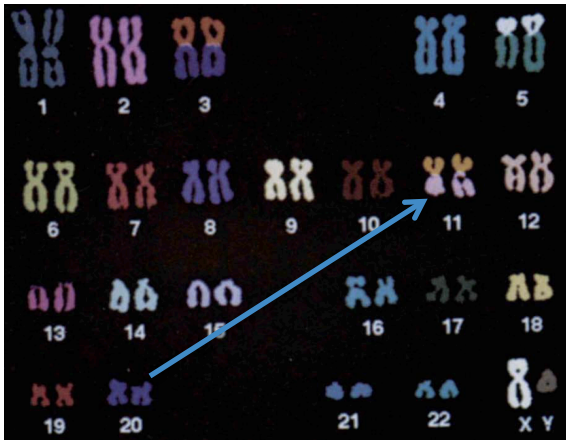
123

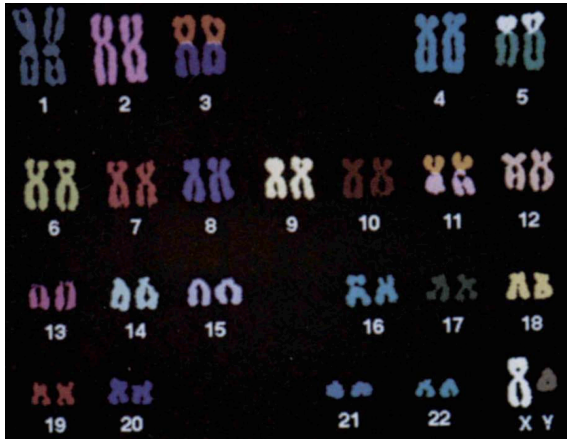


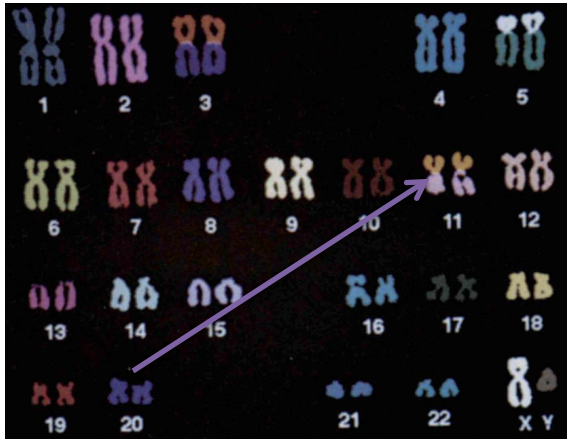
125



145





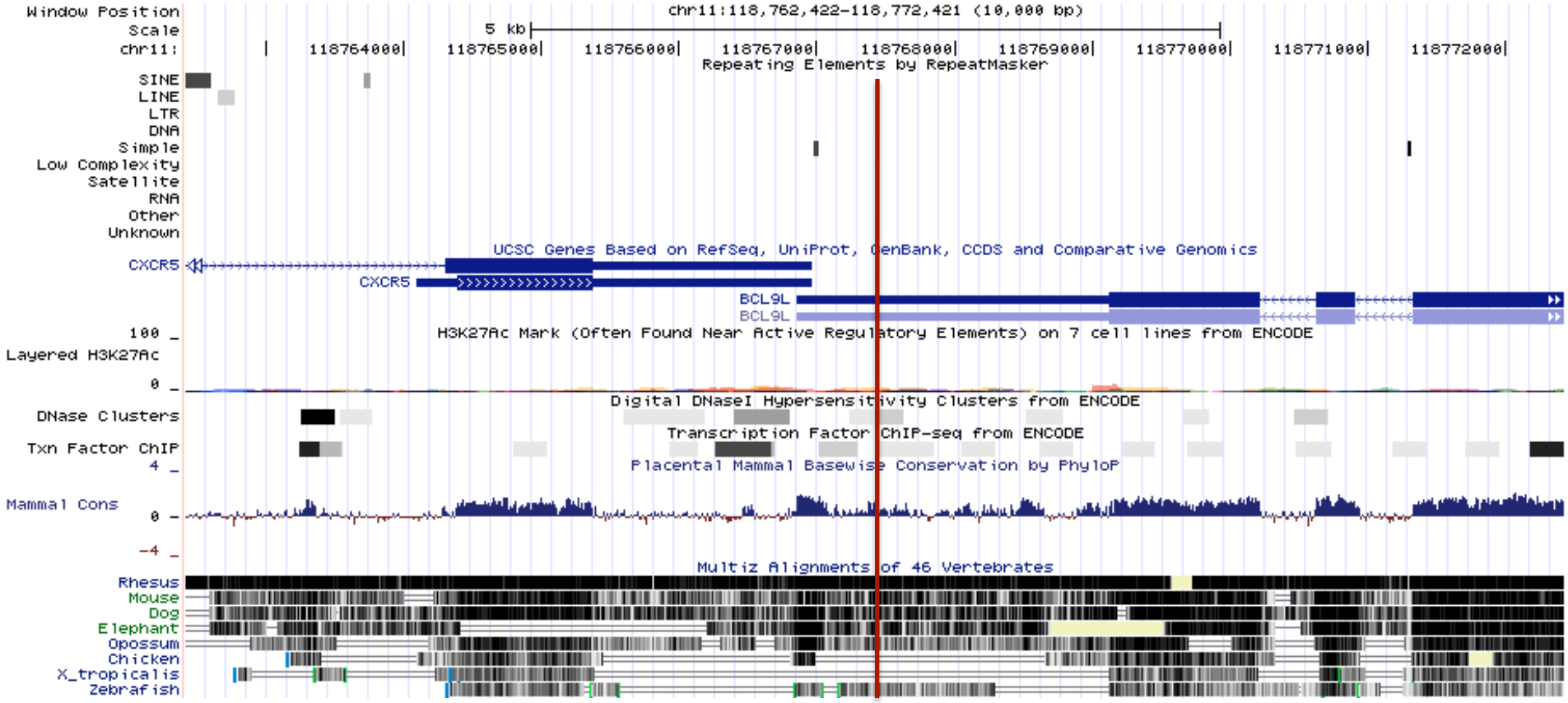


Nat. Genet. (1996) 12: 368

Example in HSB123 hippocampus

```
non-repeat: chr11:118767422-118767448      3'UTR of BCL9
repeat:     chr20:60762775-60761884        L1MEf_LINE_L1_2
```

```
read        AGCTGCCAGTGCACGACCAGGCCCTCCTCTCGTTTTCCTGATCAGGCTAGAAGTGTGTTAGTTAGCTTGTCTNNT
non-rep     AGCTGCCAGTGCAAGACCAGGCCCTC
rep 26S47M2S                          CTCTCGTTTTCCTGATCAGGCTAGAAGTGTGTTAGTTAGCTTGTCTC
```



function of BCL9: Transcriptional activator of beta-catenin; plays a role in tumorigenesis. Found in a complex with CDC73; CTNNB1 and PYGO1. Interacts with CTNNB1 and enhances its neoplastic transforming activity.

Example in HSB123 hippocampus

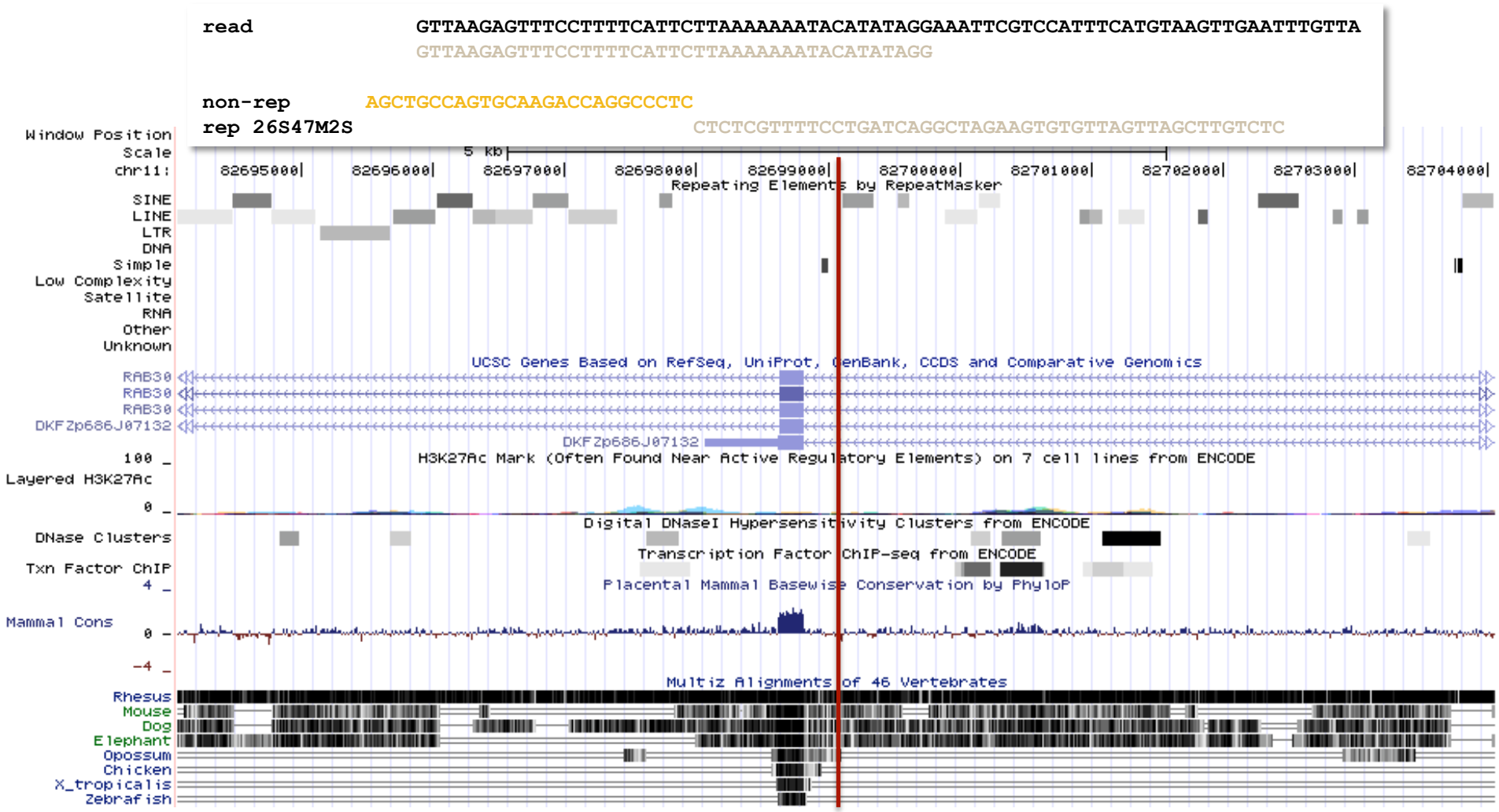
non-repeat: **chr11:118767422-118767448** **3′UTR of BCL9**
repeat: **chr20:60762775-60761884** **L1MEf_LINE_L1_2**

```
read          AGCTGCCAGTGCACGACCAGGCCCTCCTCTCGTTTTCCTGATCAGGCTAGAAGTGTGTTAGTTAGCTTGTCTNNT
non-rep       AGCTGCCAGTGCAAGACCAGGCCCTC
rep 26S47M2S                              CTCTCGTTTTCCTGATCAGGCTAGAAGTGTGTTAGTTAGCTTGTCTC
```

Sample	Region	Non-rep Strand (on chr11)	Rep Strand (on chr20)	chr11 pos	chr20 pos	Distance from chr11:118767422	Repeat Element Name
HSB145	OFC	+	-	82699062	60761884	**-36068360**	**L1MEf_LINE_L1_2**
…						**(36Mb)**	
HSB126	S1C	-	+	118485586	14588616	-281836	AluSq2_SINE_Alu_2
HSB145	MD	-	+	118485586	14588616	-281836	AluSq2_SINE_Alu_2
HSB123	IPC	-	-	118514600	33495966	-252822	L2a_LINE_L2_2
HSB123	VFC	-	+	118517002	21651144	-250420	AluYk4_SINE_Alu_2
HSB123	S1C	+	+	118527875	54189037	-239547	(A)n_Simple_repeat_Simple_repeat_2
HSB144	MFC	-	+	118622277	15026365	-145145	(T)n_Simple_repeat_Simple_repeat_2
HSB123	HIP	-	-	118623158	36207939	-144264	AluSx_SINE_Alu_2
HSB123	DFC	+	+	118624014	35270278	-143408	AluSx_SINE_Alu_2
HSB123	STR	-	+	118645431	56028200	-121991	AluJb_SINE_Alu_2
HSB144	AMY	-	+	118666636	42170568	-100786	AluSx_SINE_Alu_2
HSB123	**HIP**	**-**	**-**	**118767422**	**60761884**	**0**	**L1MEf_LINE_L1_2**
HSB144	STC	+	+	118829570	35670229	62148	AluY_SINE_Alu_2
HSB123	MFC	+	-	118871482	31358416	104060	AluSp_SINE_Alu_2
HSB144	S1C	-	-	118871580	49221469	104158	AluJo_SINE_Alu_2
HSB123	DFC	+	-	118871580	49261363	104158	AluSq_SINE_Alu_2
HSB123	S1C	+	-	118872468	57554507	105046	AluJr4_SINE_Alu_2
HSB123	A1C	-	+	118872609	43910012	105187	AluSz_SINE_Alu_2
HSB145	VFC	+	+	118872616	43910012	105194	AluSz_SINE_Alu_2
HSB123	MFC	+	+	118873358	10648740	105936	(A)n_Simple_repeat_Simple_repeat_2
HSB145	M1C	-	-	118874289	39867632	106867	AluY_SINE_Alu_2

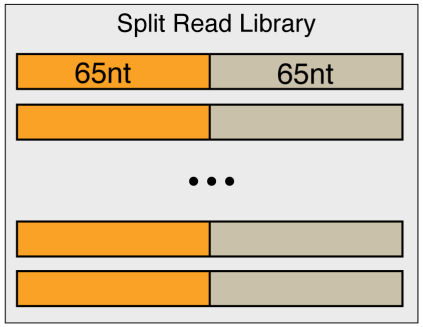
Same L1mEf_LINE L1 in HSB145 OFC

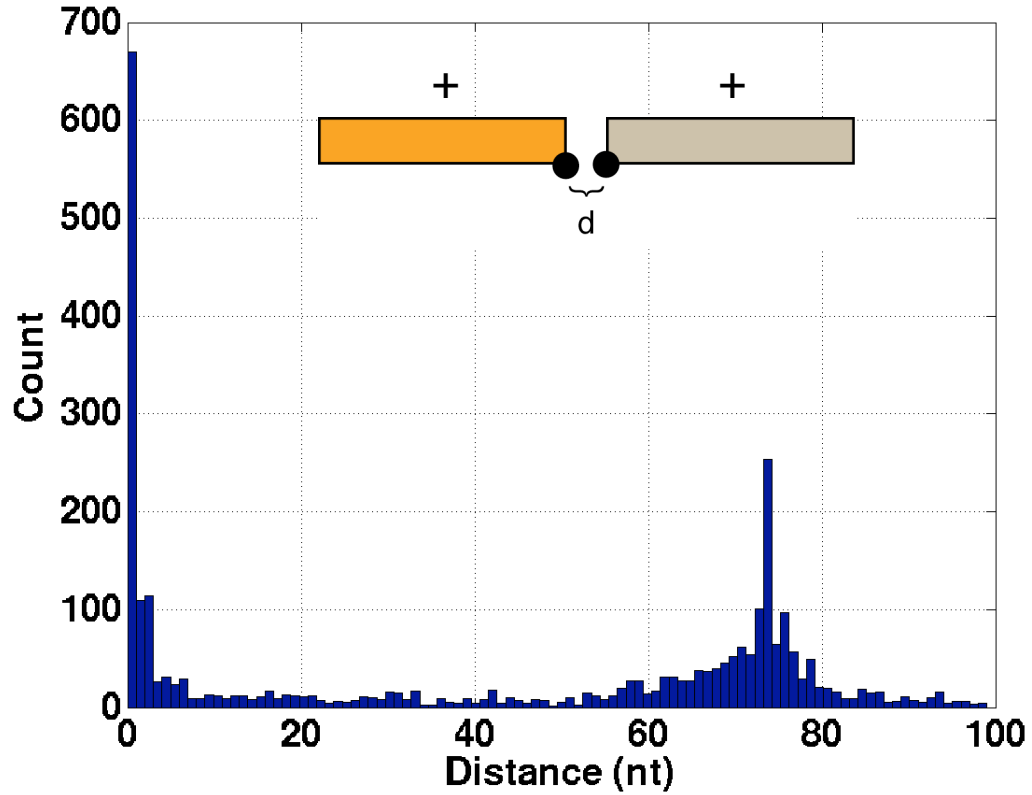
non-repeat: chr11:82694062 intron of RAB30
repeat: chr20:60762775-60761884 L1MEf_LINE_L1_2

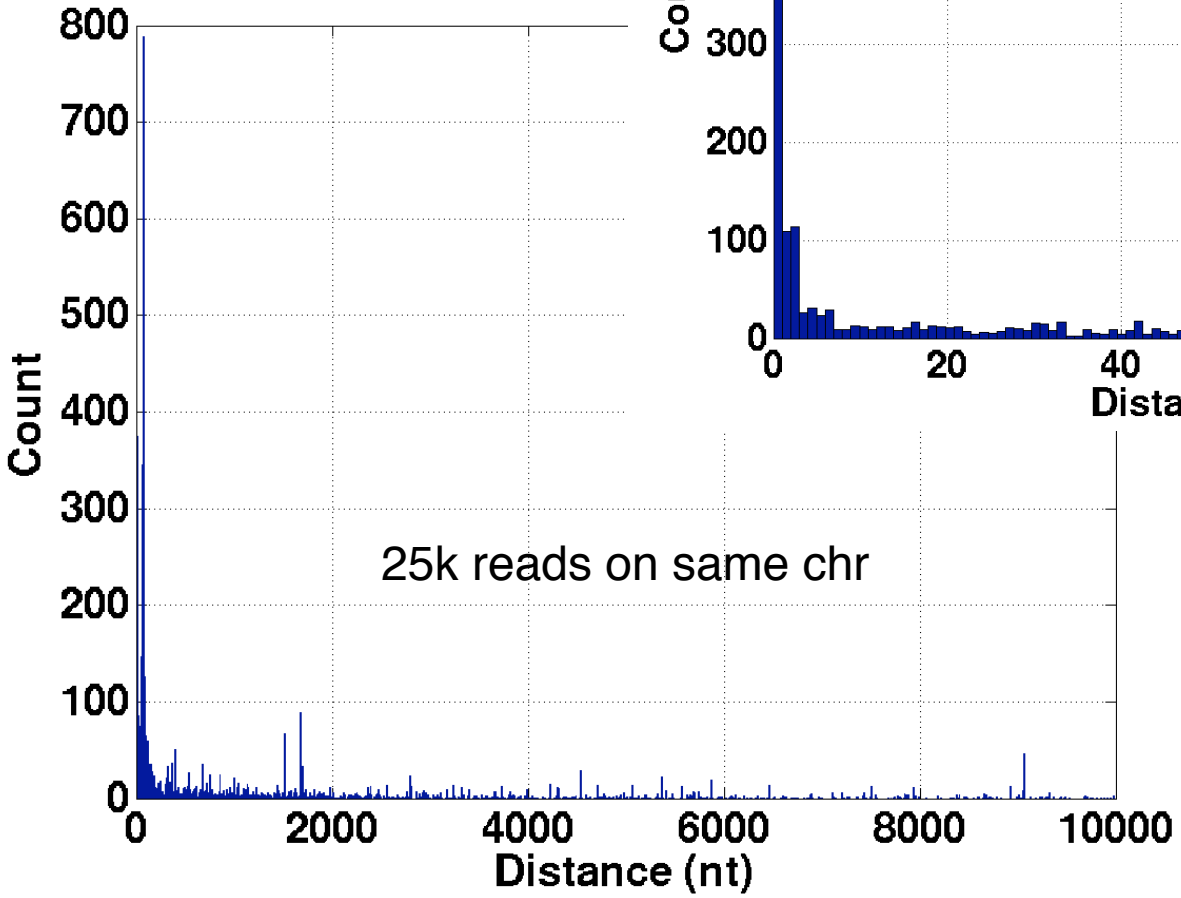


function of RAB30: isoform CRA_a, member RAS oncogene small GTPase family

I am currently stuck building the split read library..



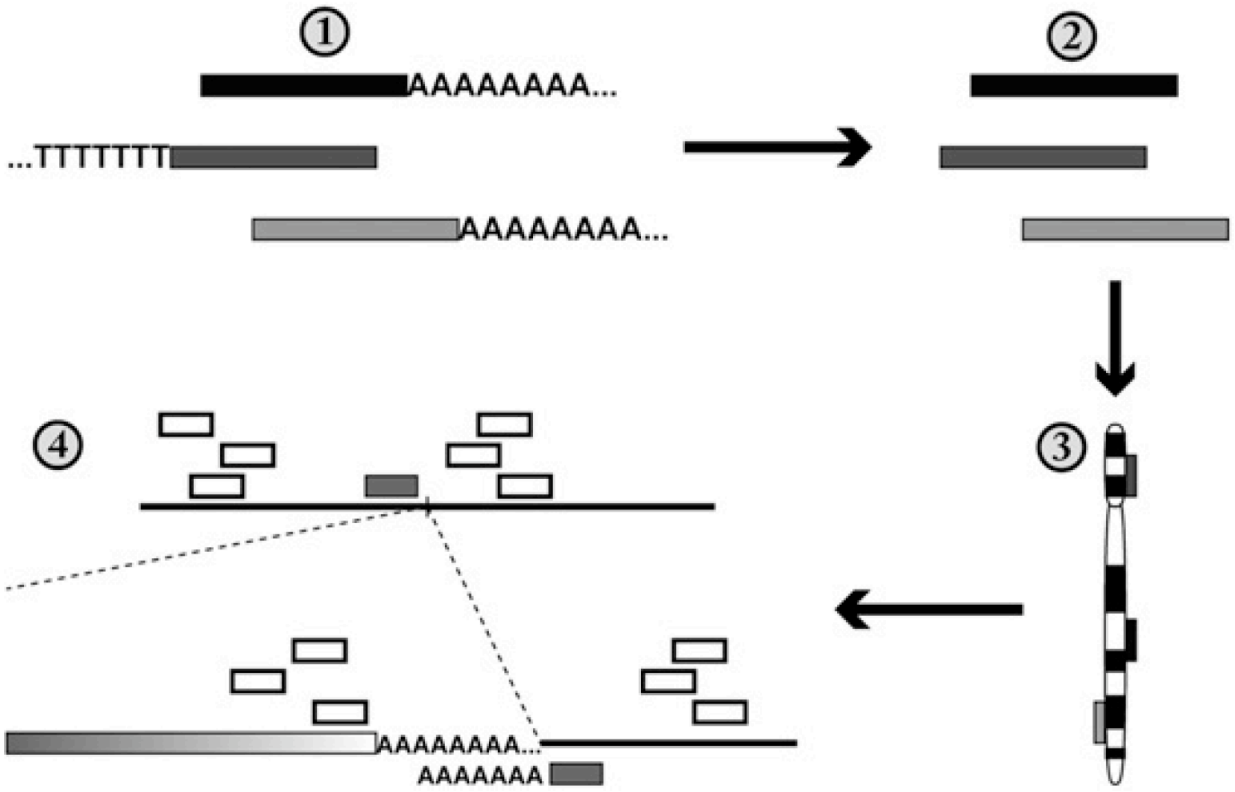




25k reads on same chr

(32k reads split on different chromosomes)

Suggestion from Alex:
Search for polyA ends of LINE elements



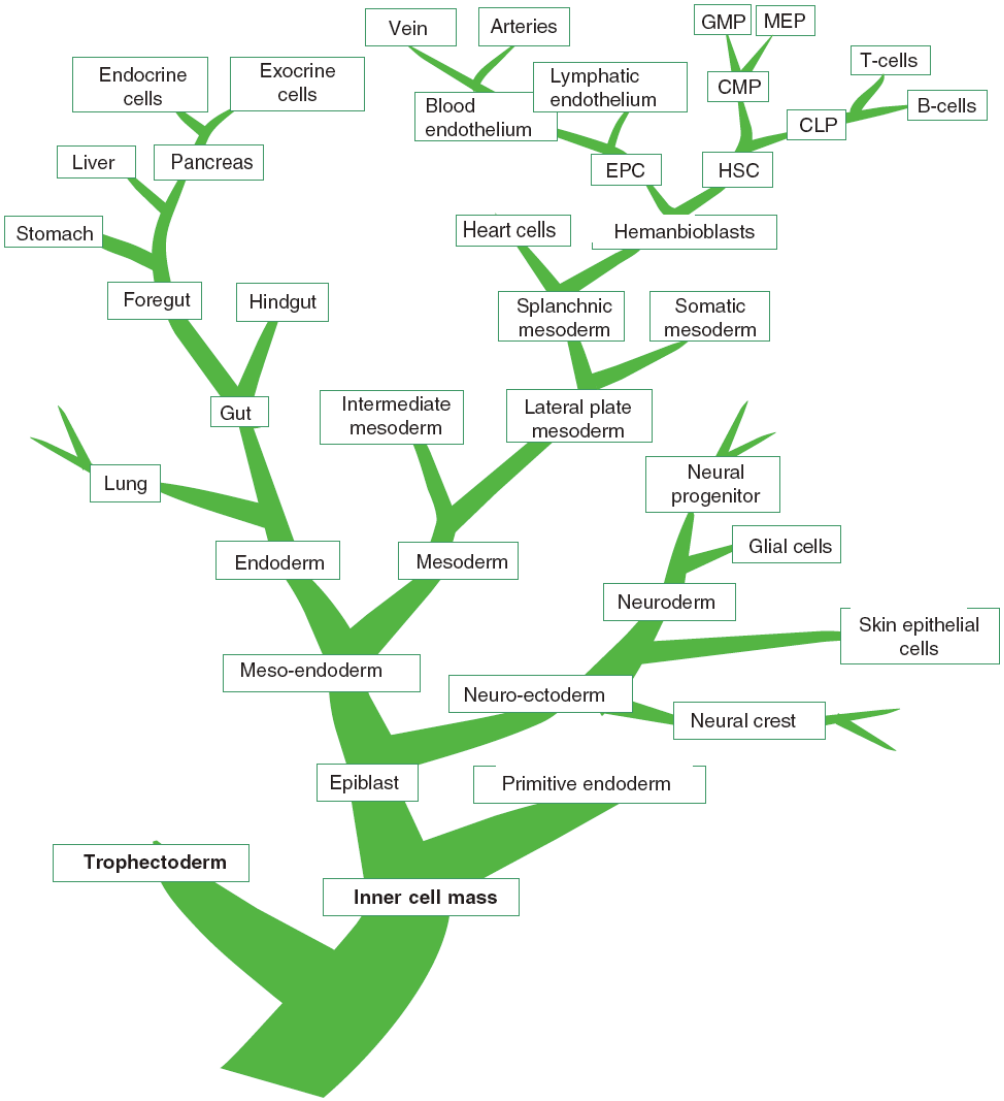
Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans
Genome Res. (2011) 21: 985

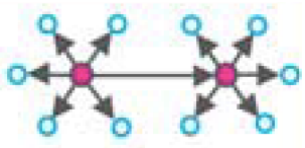
Brainseq Transposons

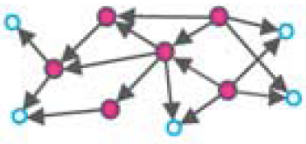
modENCODE developmental timecourses

3D Structure of the Genome

Ontogenetic Tree of Cell Types in Human Development

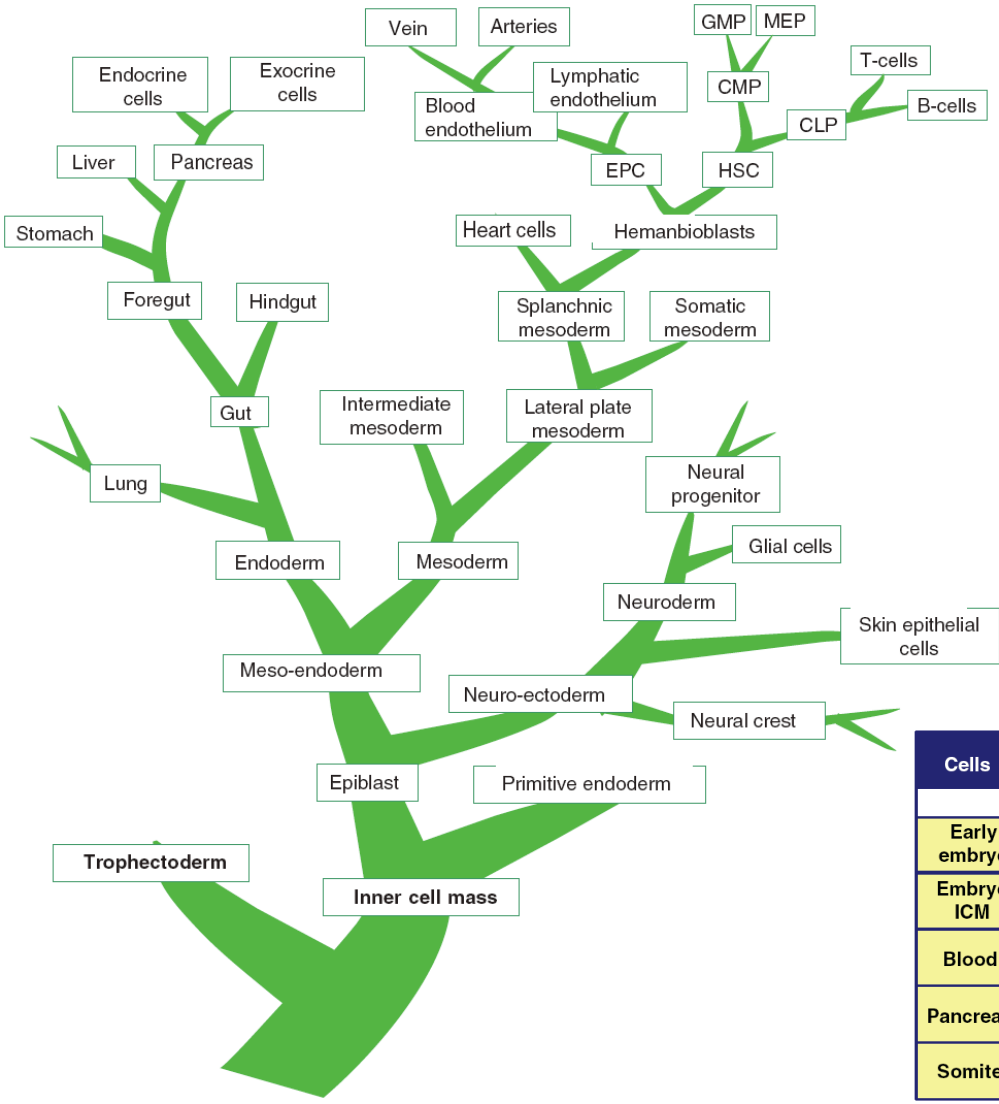


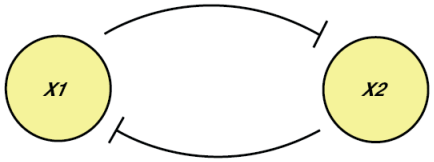


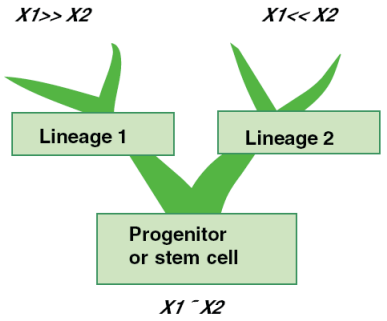


Understanding gene circuits at cell-fate branch points for rational cell reprogramming
Trends Genet. (2011) 27: 55

Ontogenetic Tree of Cell Types in Human Development



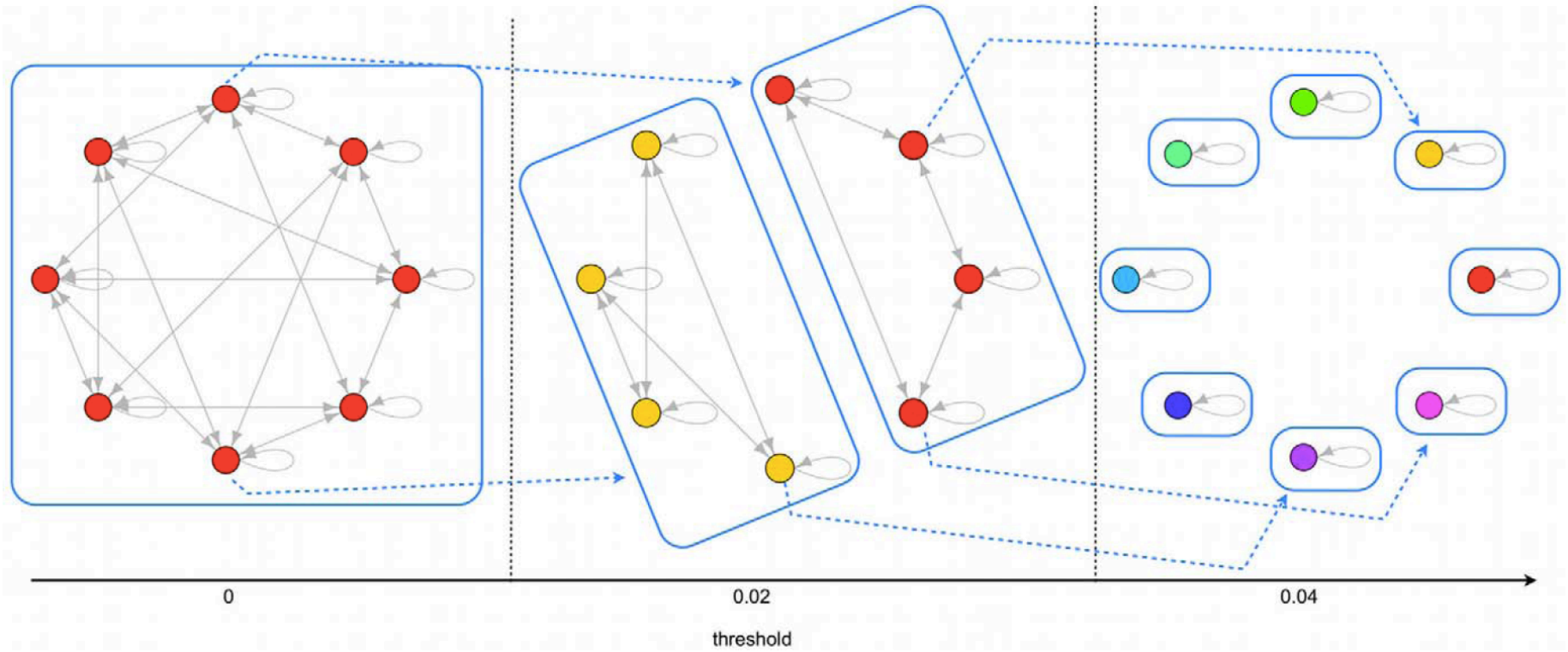




Cells	Transcription factors		Cell fates		
	X1	X2	X1> X2	X1~ X2	X1< X2
Early embryo	Cdx2	Oct4	Trophectoderm	Totipotent embryo	Inner cell mass
Embryo ICM	GATA6	Nanog	Primitive endoderm	Inner cell mass	Epiblast
Blood	GATA1	PU.1	Erythroid cells	Common myeloid progenitor	Myeloid cells
Pancreas	Ptf1a	Nkx6	Exocrine cells	Pancreatic progenitor	Endocrine cells
Somite	Pax3	Foxc2	Myogenic cells	Dermomyotome progenitor	Vascular cells

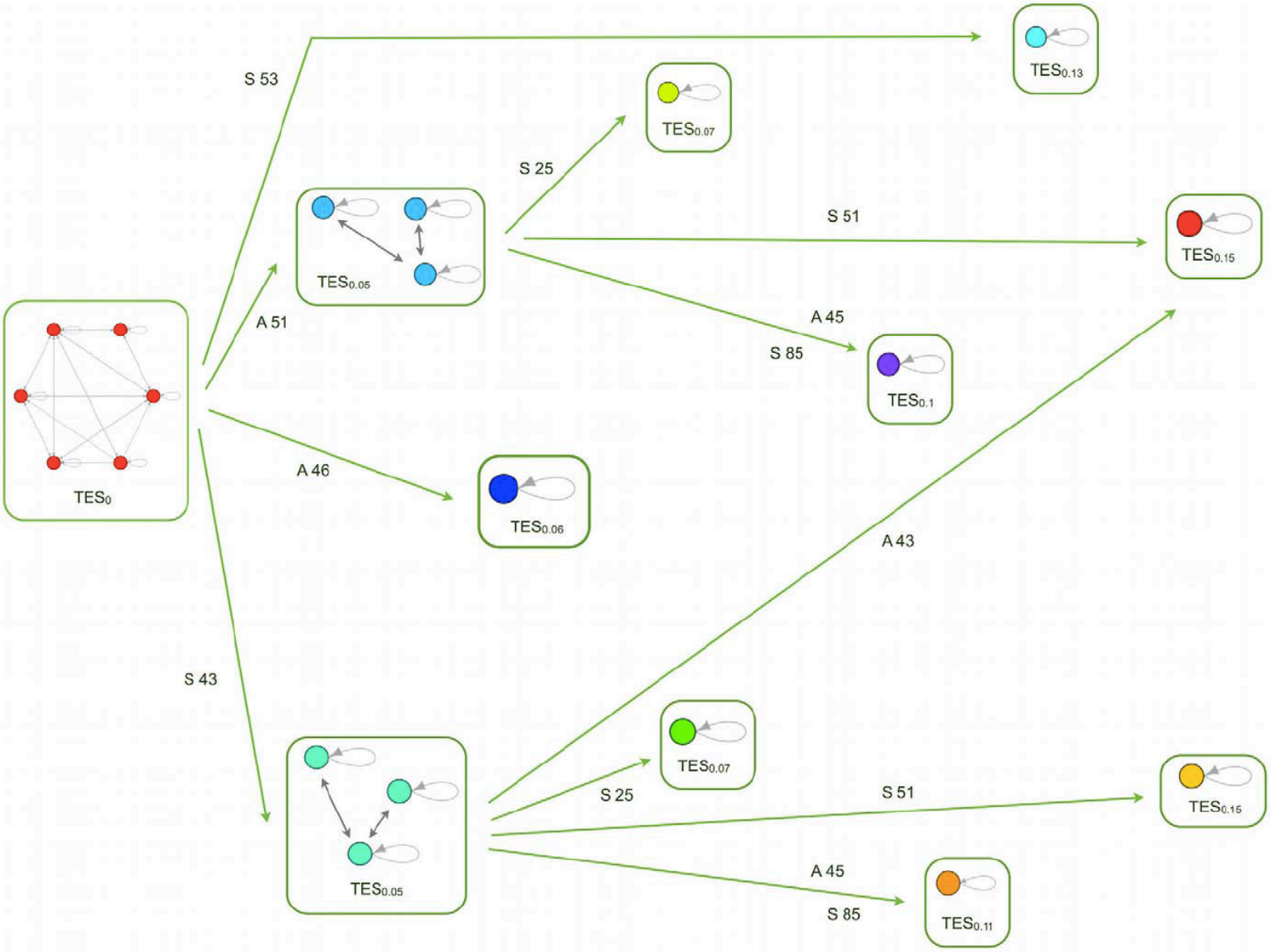
Understanding gene circuits at cell-fate branch points for rational cell reprogramming
Trends Genet. (2011) 27: 55

Dynamical Model of Cell Differentiation Networks



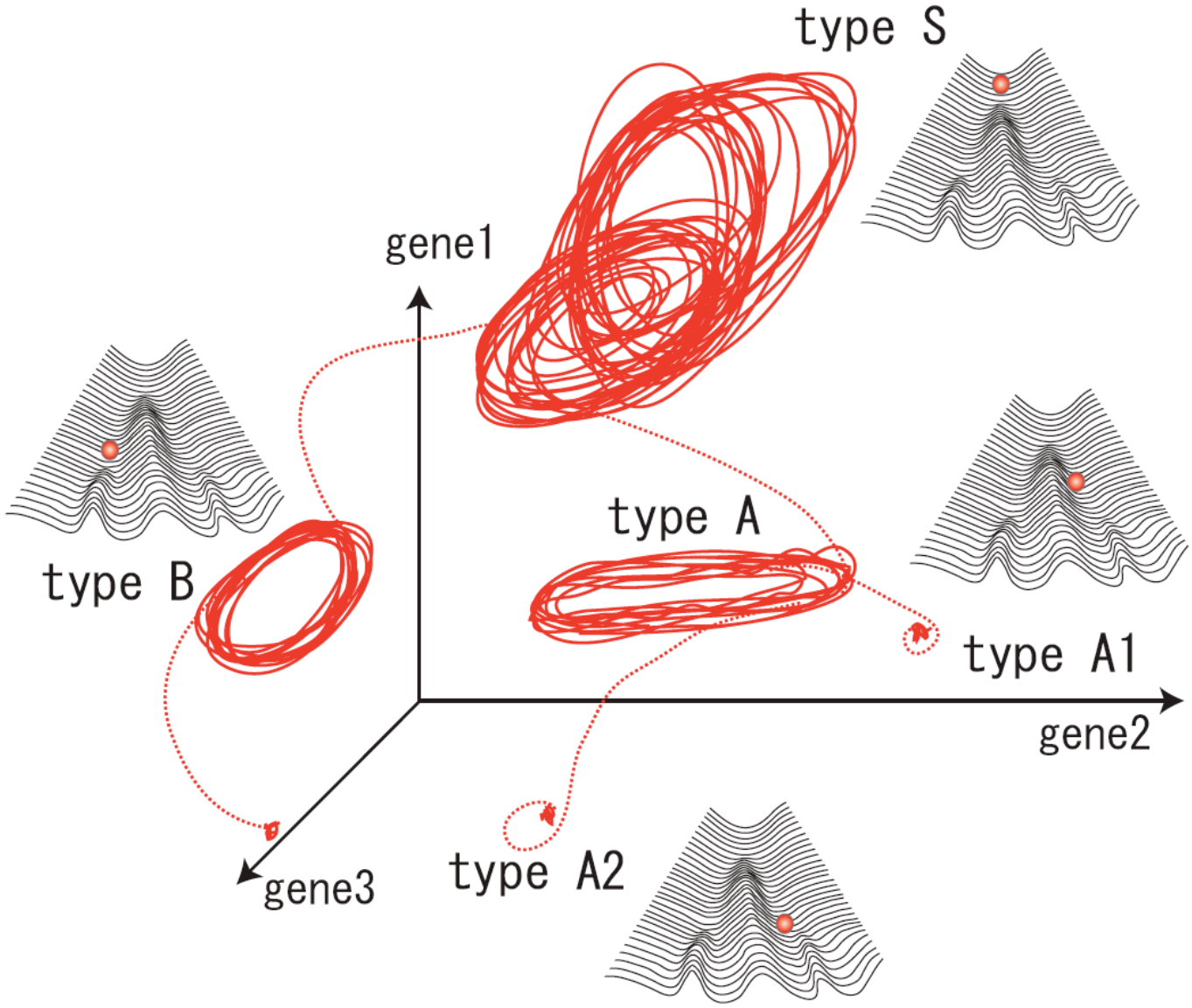
A Dynamical Model of Genetic Networks for Cell Differentiation
PLoS ONE (2011) 6: e17703

Dynamical Model of Cell Differentiation Networks



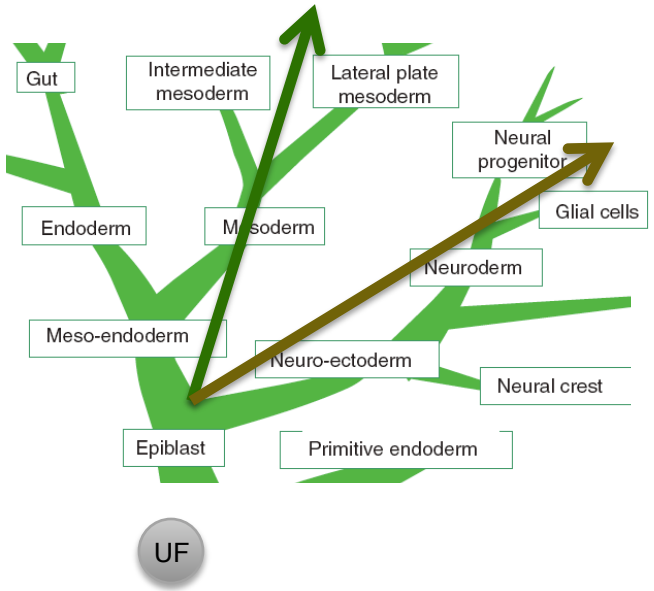
A Dynamical Model of Genetic Networks for Cell Differentiation
PLoS ONE (2011) 6: e17703

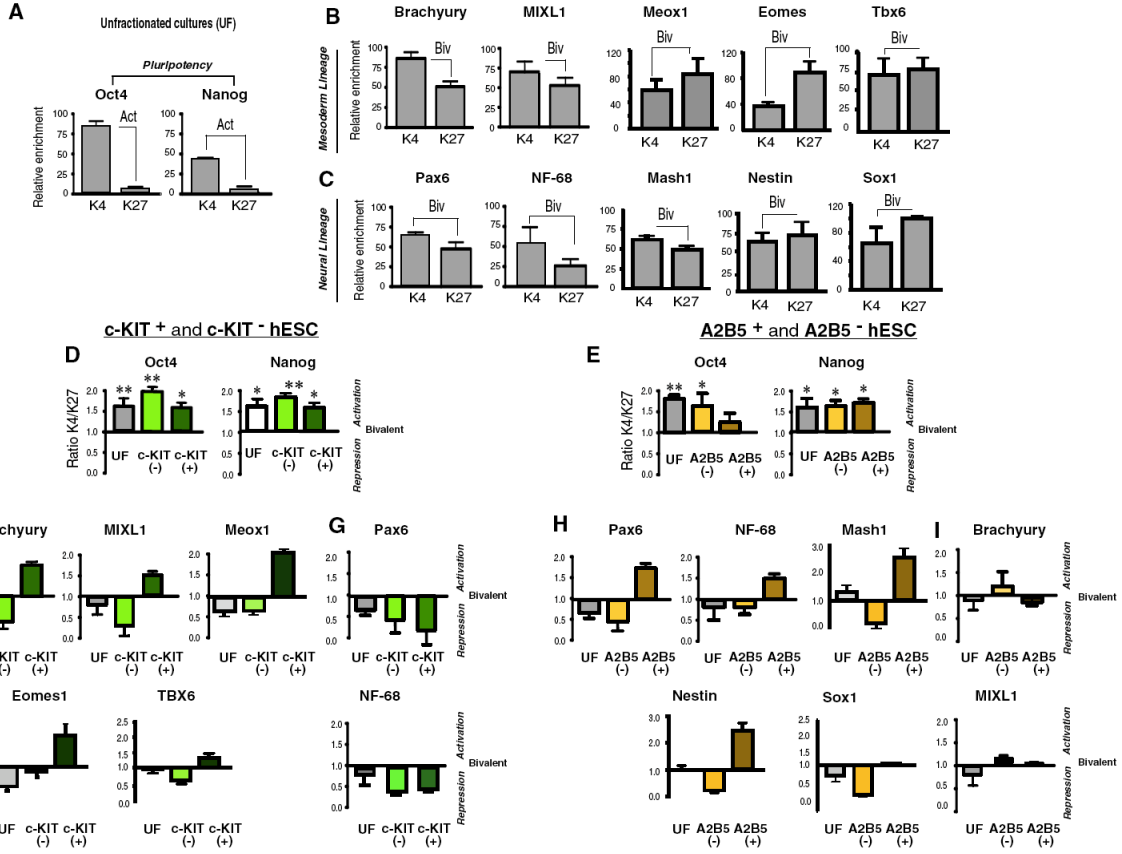
Chaotic Dynamics in Pluripotent Cell Differentiation



Chaotic expression dynamics implies pluripotency: when theory and experiment meet
Biol. Direct (2009) 4: 17

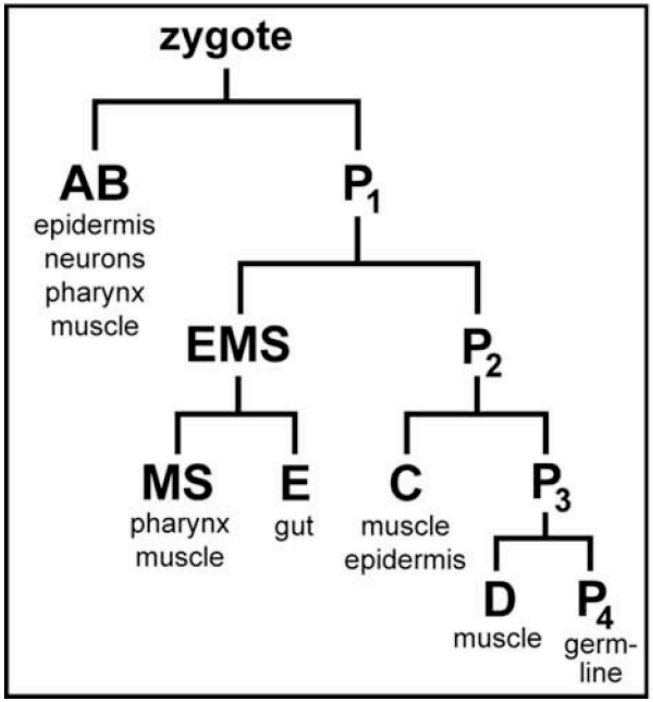
Ontogenetic Tree of Cell Types in Human Development

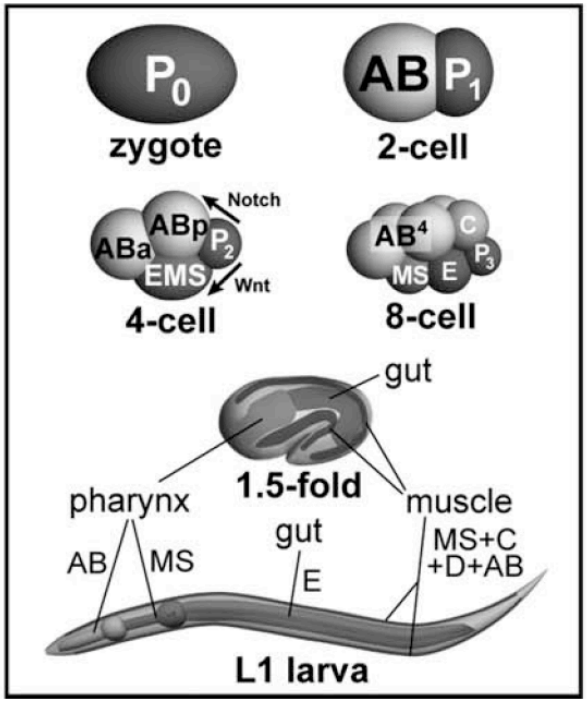


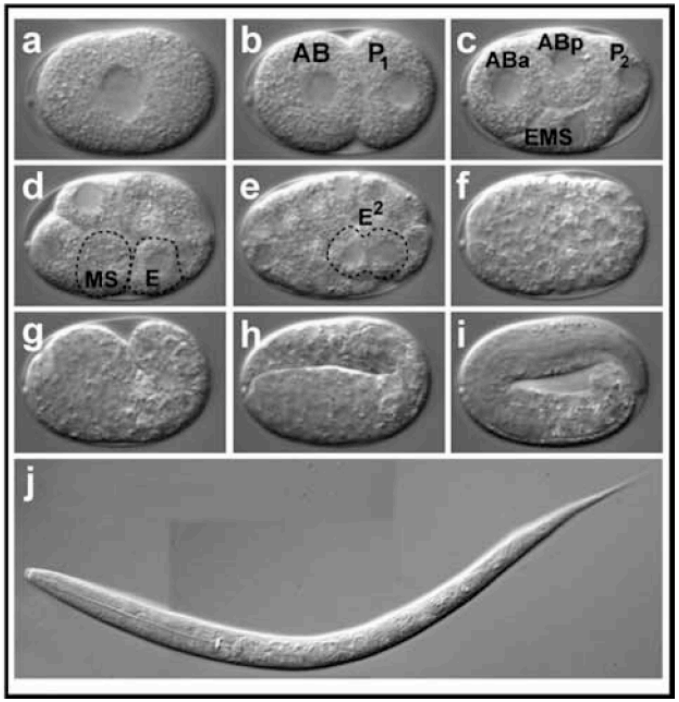


Cell Fate Potential of Human Pluripotent Stem Cells Is Encoded by Histone Modifications
Cell Stem Cell (2011) 9: 24

Regulation of Early Embryogenesis in *C. elegans*

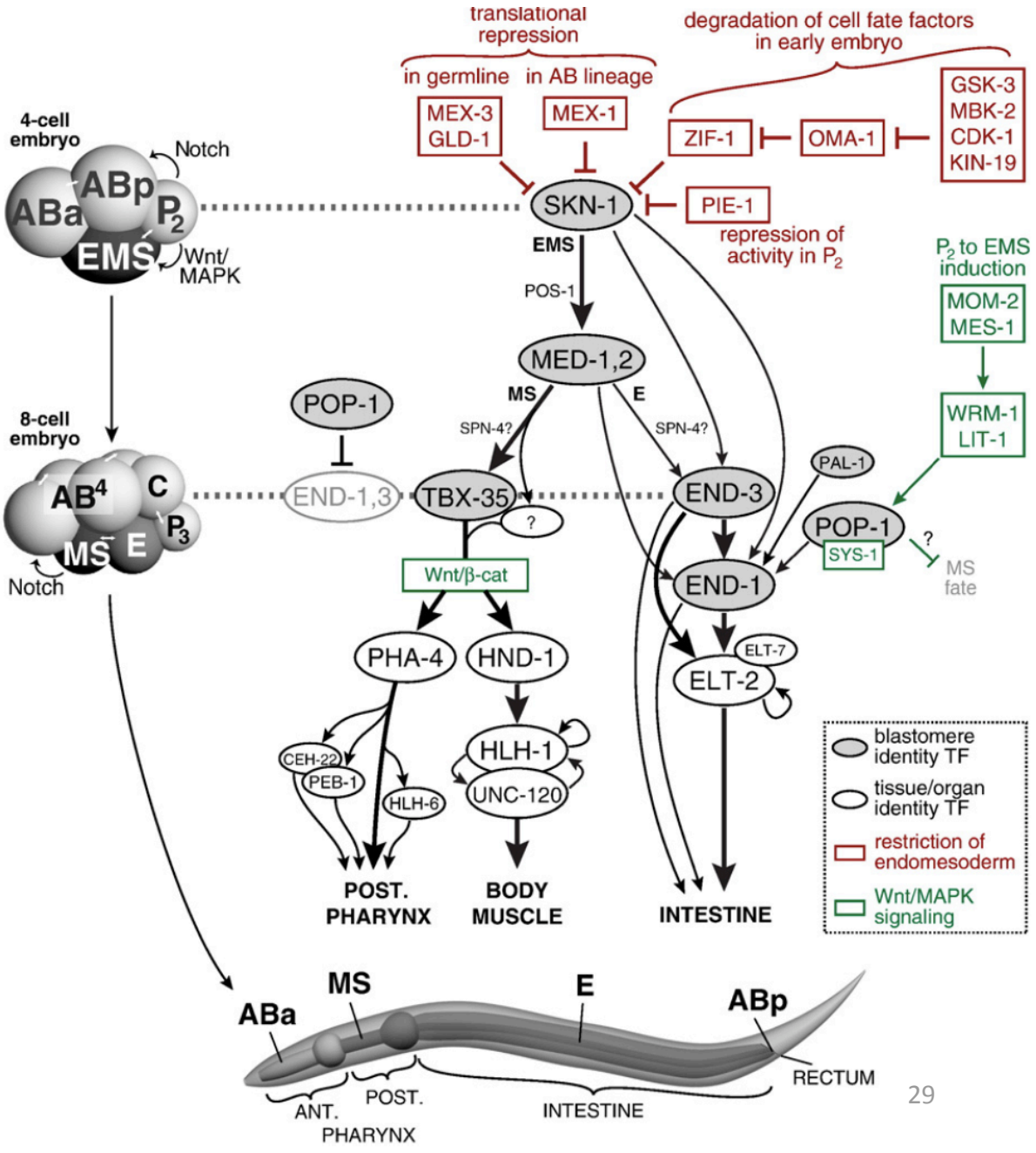






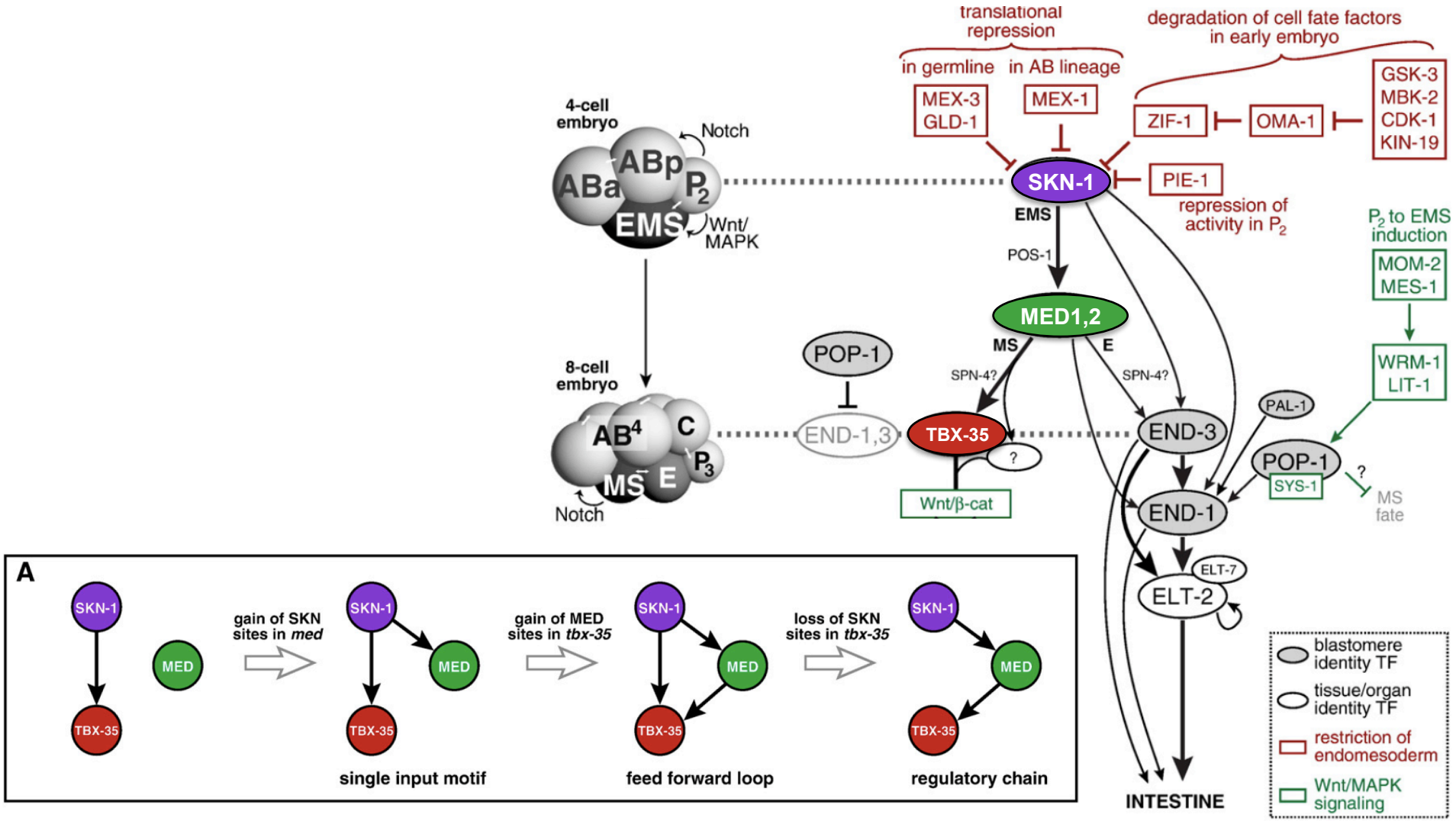
Cell Fate Specification in the C. elegans Embryo
Developmental Dynamics (2010) 239: 1315

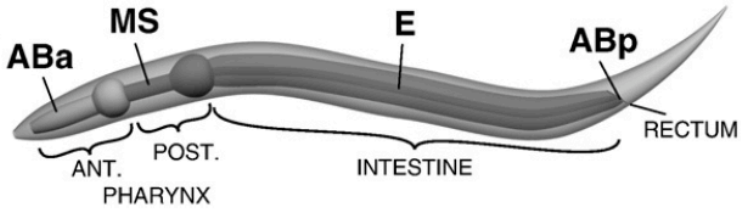
Regulation of Early Embryogenesis in *C. elegans*



Structure and evolution of the *C. elegans* embryonic endomesoderm network
BBA-Gene Regul Mech (2009) 1789: 250

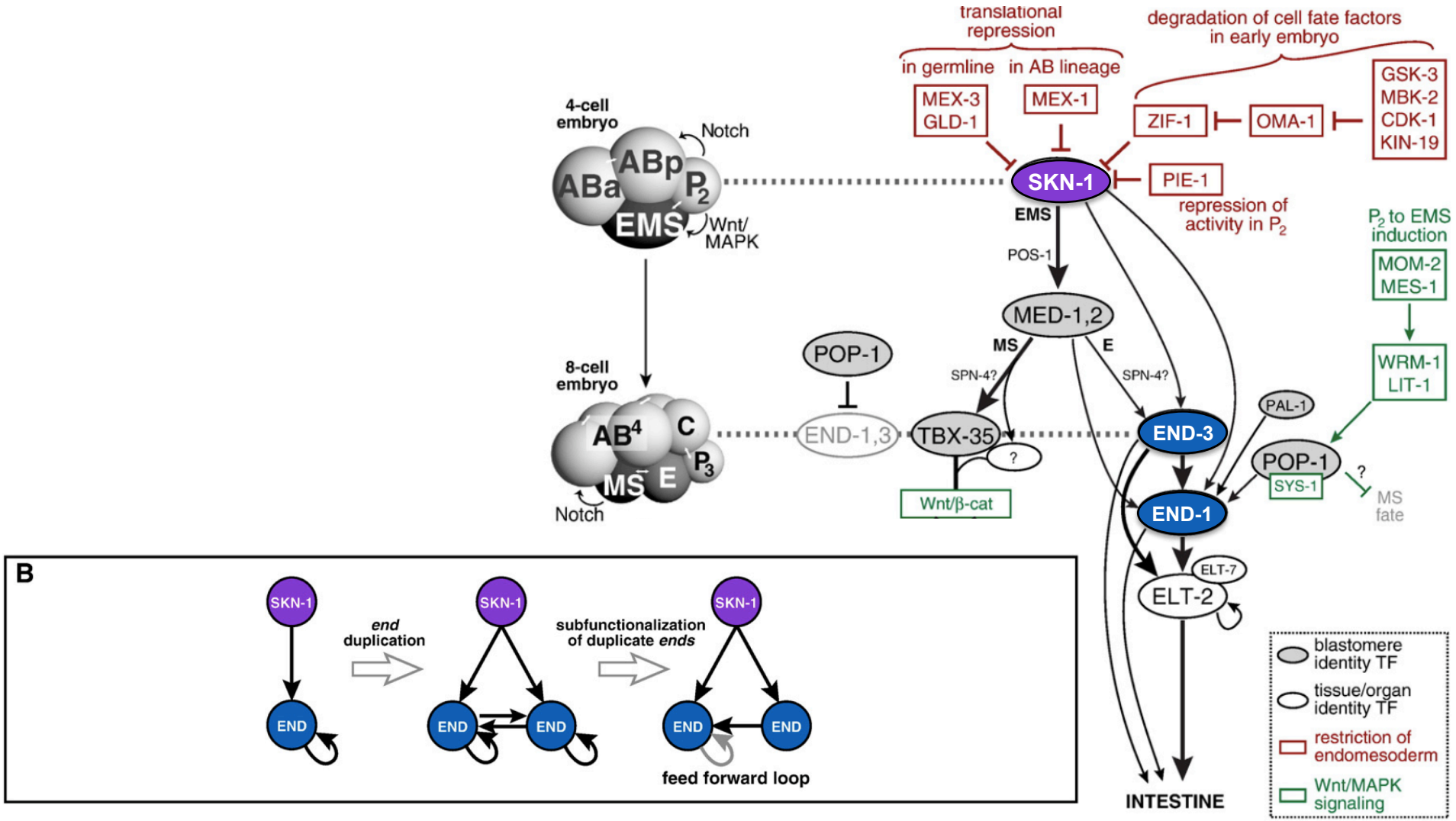
Regulation of Early Embryogenesis in *C. elegans*

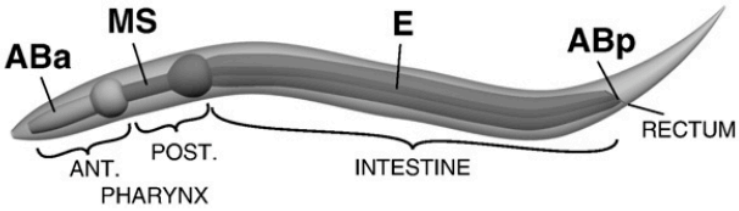




Structure and evolution of the *C. elegans* embryonic endomesoderm network
BBA-Gene Regul Mech (2009) 1789: 250

Regulation of Early Embryogenesis in *C. elegans*

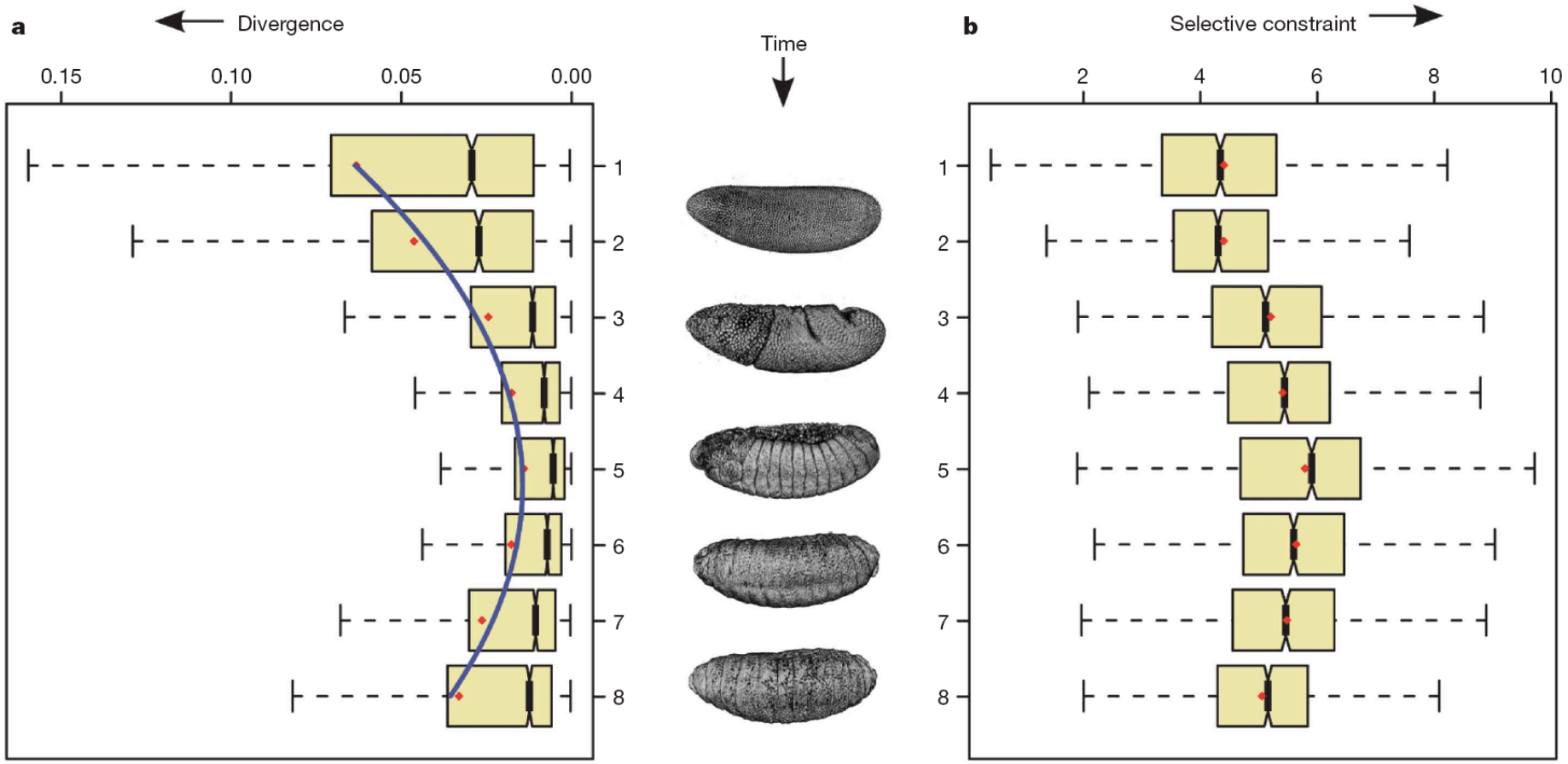




Structure and evolution of the *C. elegans* embryonic endomesoderm network
BBA-Gene Regul Mech (2009) 1789: 250

Jing's JC about Developmental Hourglass model

- phylotypic period



Nature (2010) 468: 811

Jing's JC about Developmental Hourglass model

- phylotypic period
- "To explain the resistance to evolutionary changes of this period, one hypothesis suggests that it is characterized by a high level of interactions…We propose that the phylotypic period may rather be the expression at the morphological level of strong conservation of molecular processes earlier in development."

Evol. Devel. (2010) 12: 144

Idea

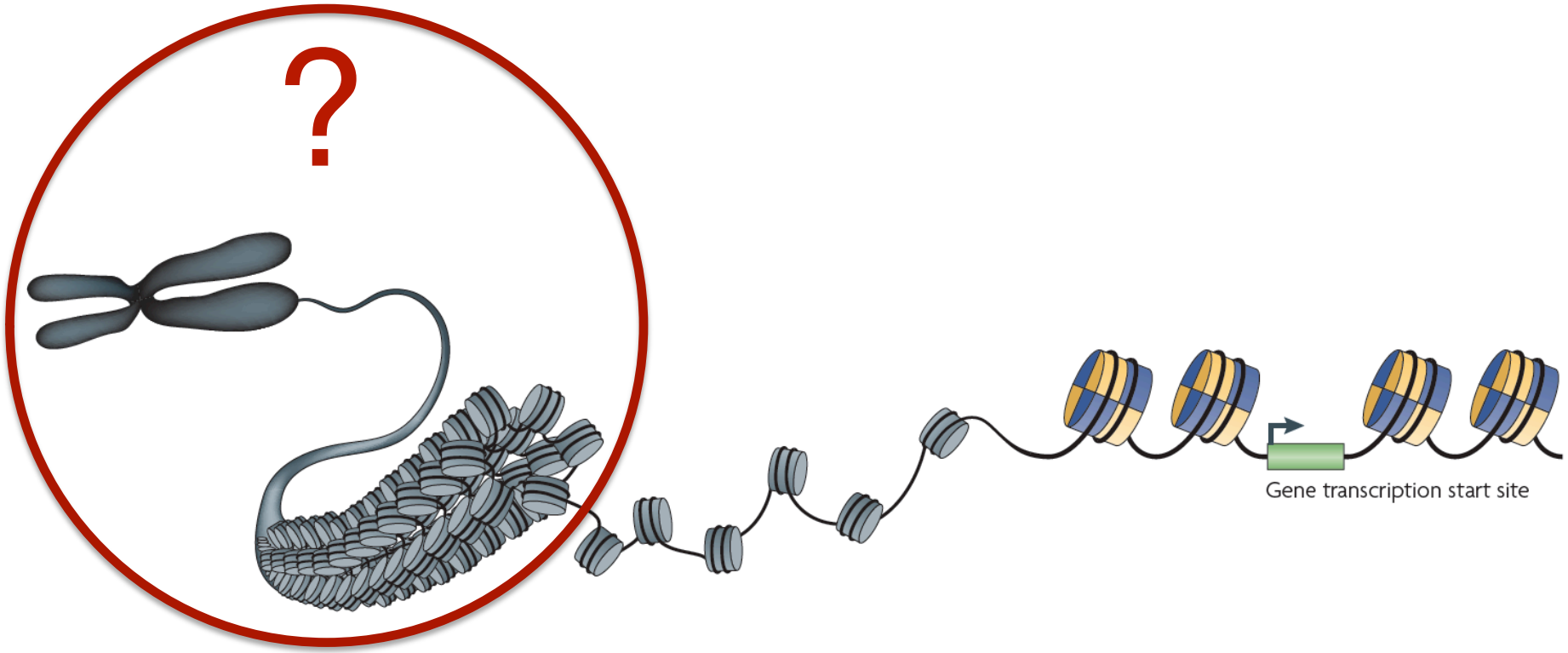
- the phylotypic period may…be the expression at the morphological level of strong conservation of molecular processes earlier in development
- Development proceeds from early, multipotent cognitive stages, to lineage specification, through to terminal differentiation cascades. Recall SANDY.
- If zebrafish authors are correct, then early developmental stages should be characterized by “cognitive” network motifs, e.g. FFLs; see Yarden 2007 Nat Genet article
- => look at PPI density in Dmel, Cele devel timecourses
- => look at network motif distribution in Dmel, Cele timecourses

Brainseq Transposons

modENCODE developmental timecourses

3D Structure of the Genome

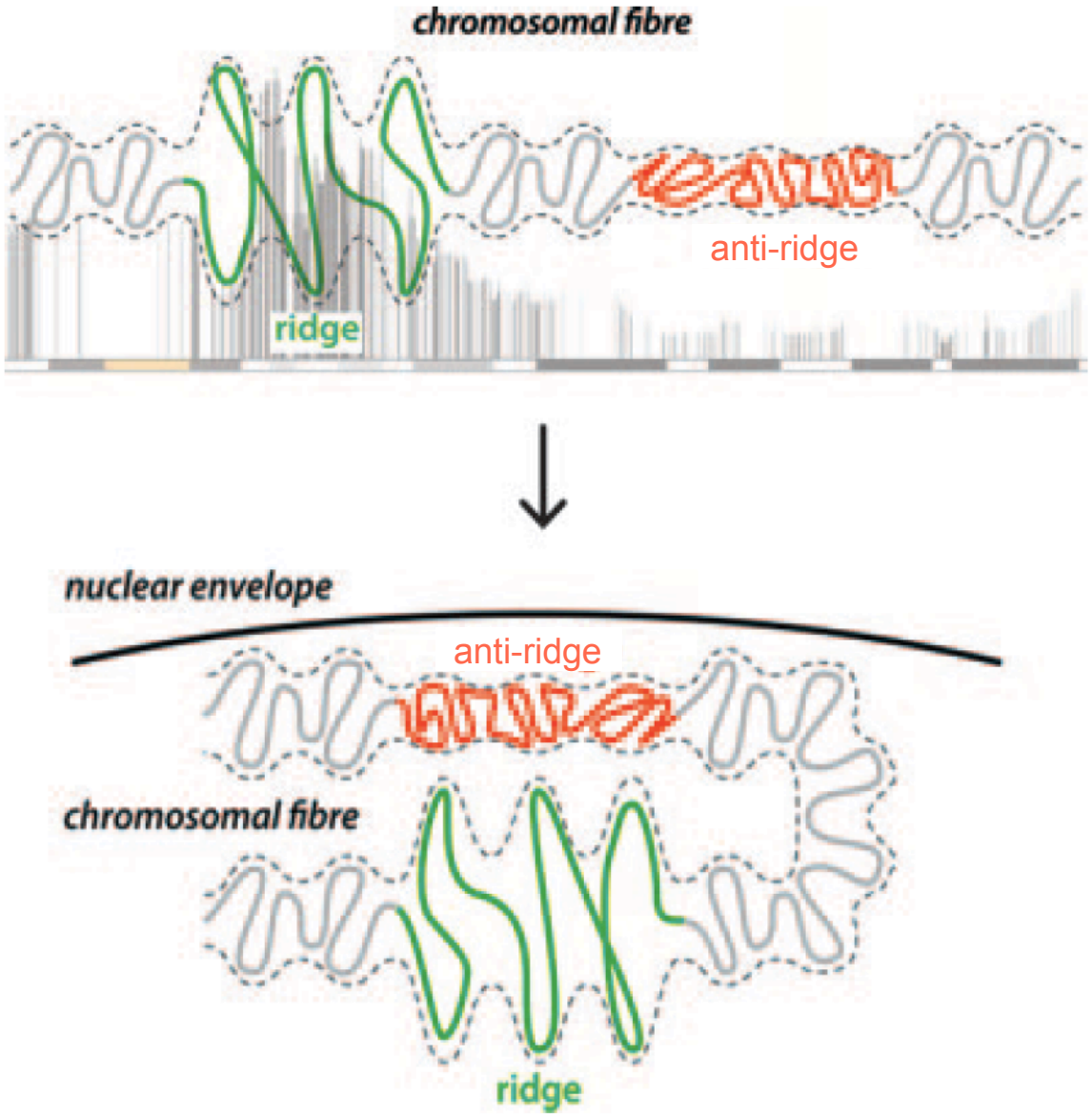
3D Structure of the Genome



At the DC ENCODE meeting, I had the idea that folding chromatin segmentations into a polymer model of chromatin dynamics might help predict higher-order chromatin structure.

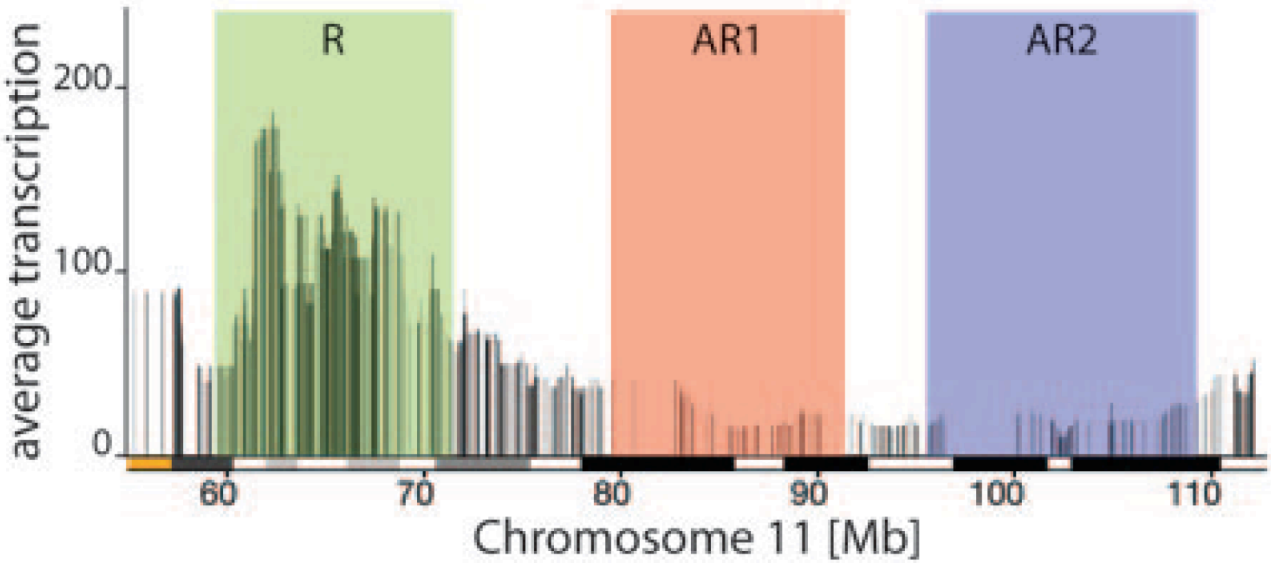
Nat. Rev. Genet. (2010) 11: 559

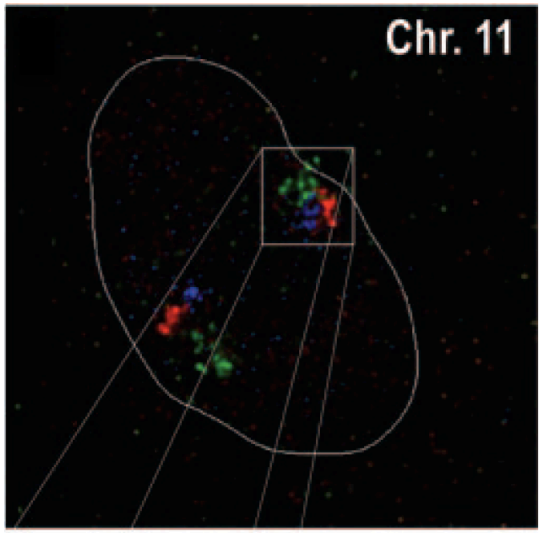
3D Structure of the Genome: Ridges and Anti-Ridges

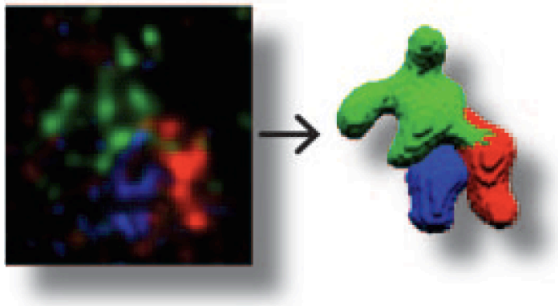


Molec. Cell. Biol. (2007) 27: 4475

3D Structure of the Genome: Ridges and Anti-Ridges

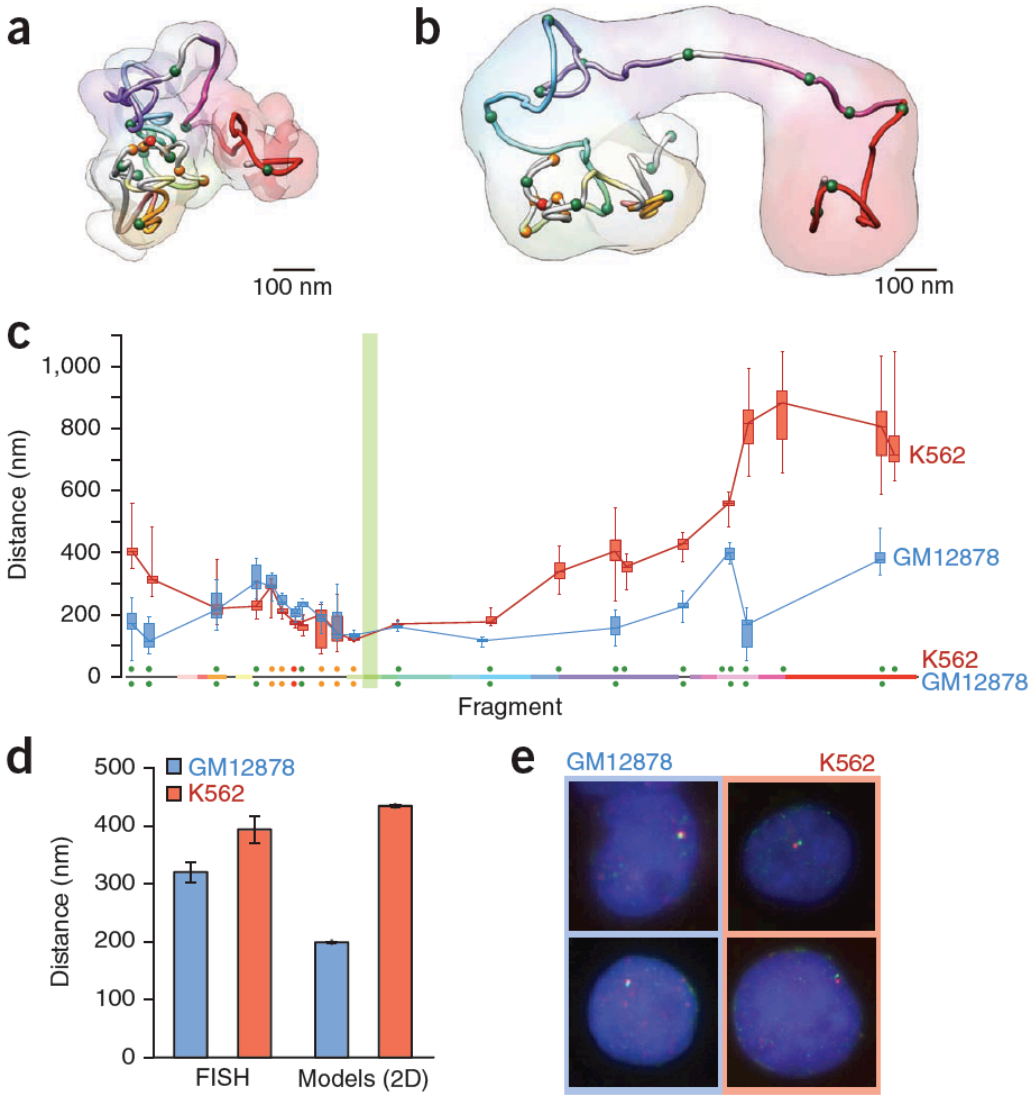






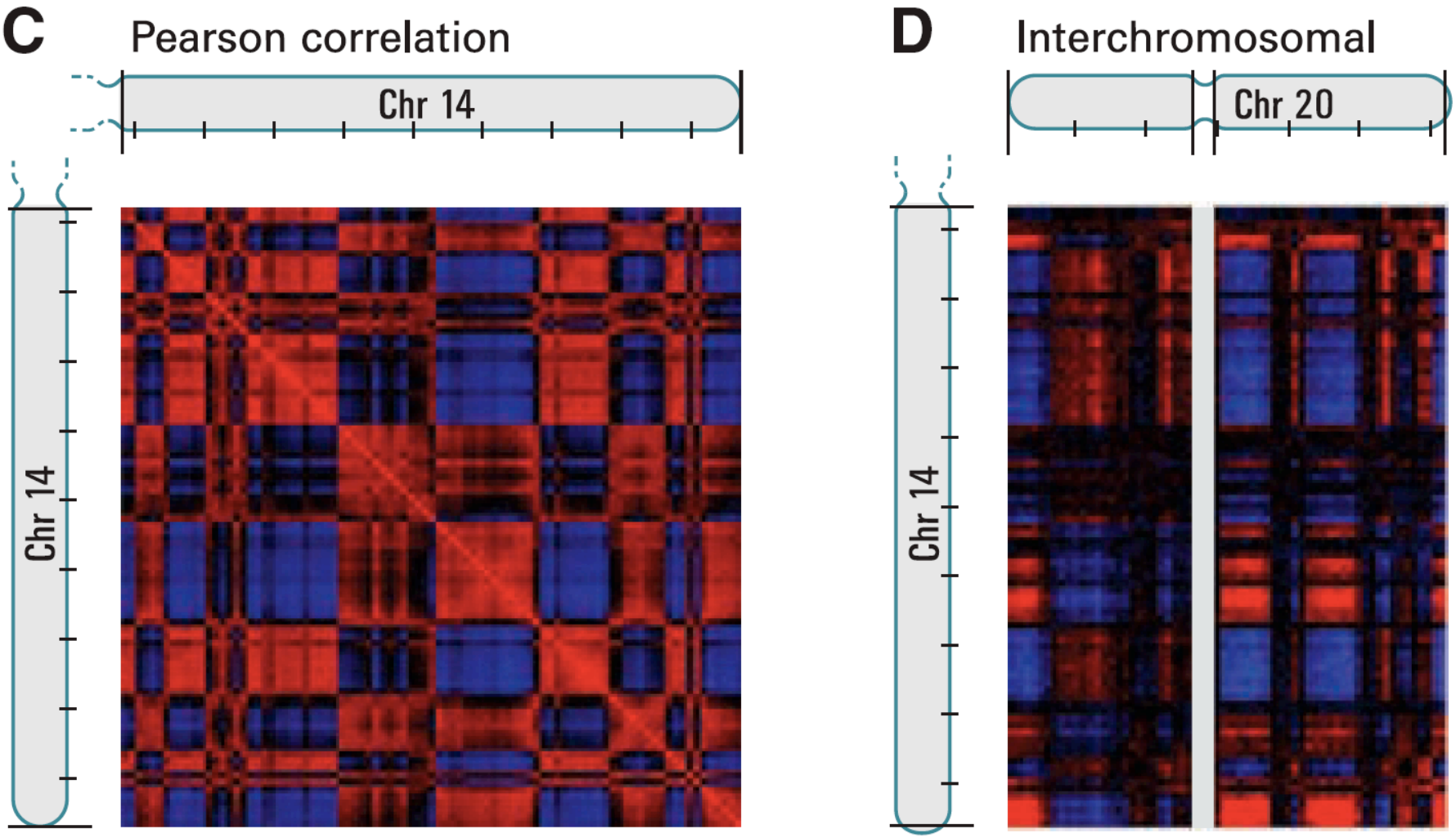
Molec. Cell. Biol. (2007) 27: 4475

3D Structure of the Genome



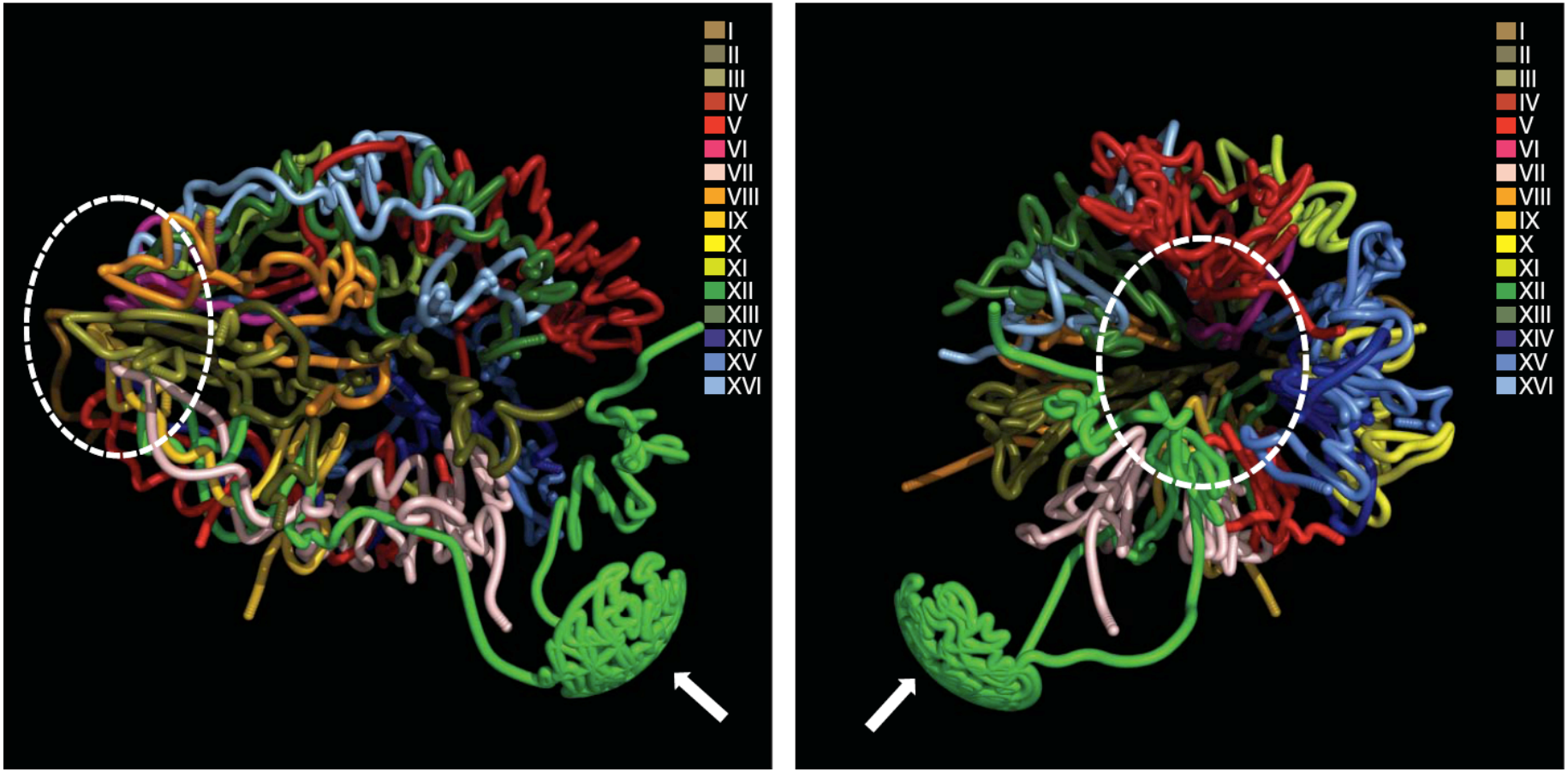
The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules
Nat. Struct. Molec. Biol. (2011) 18 107

3D Structure of the Genome

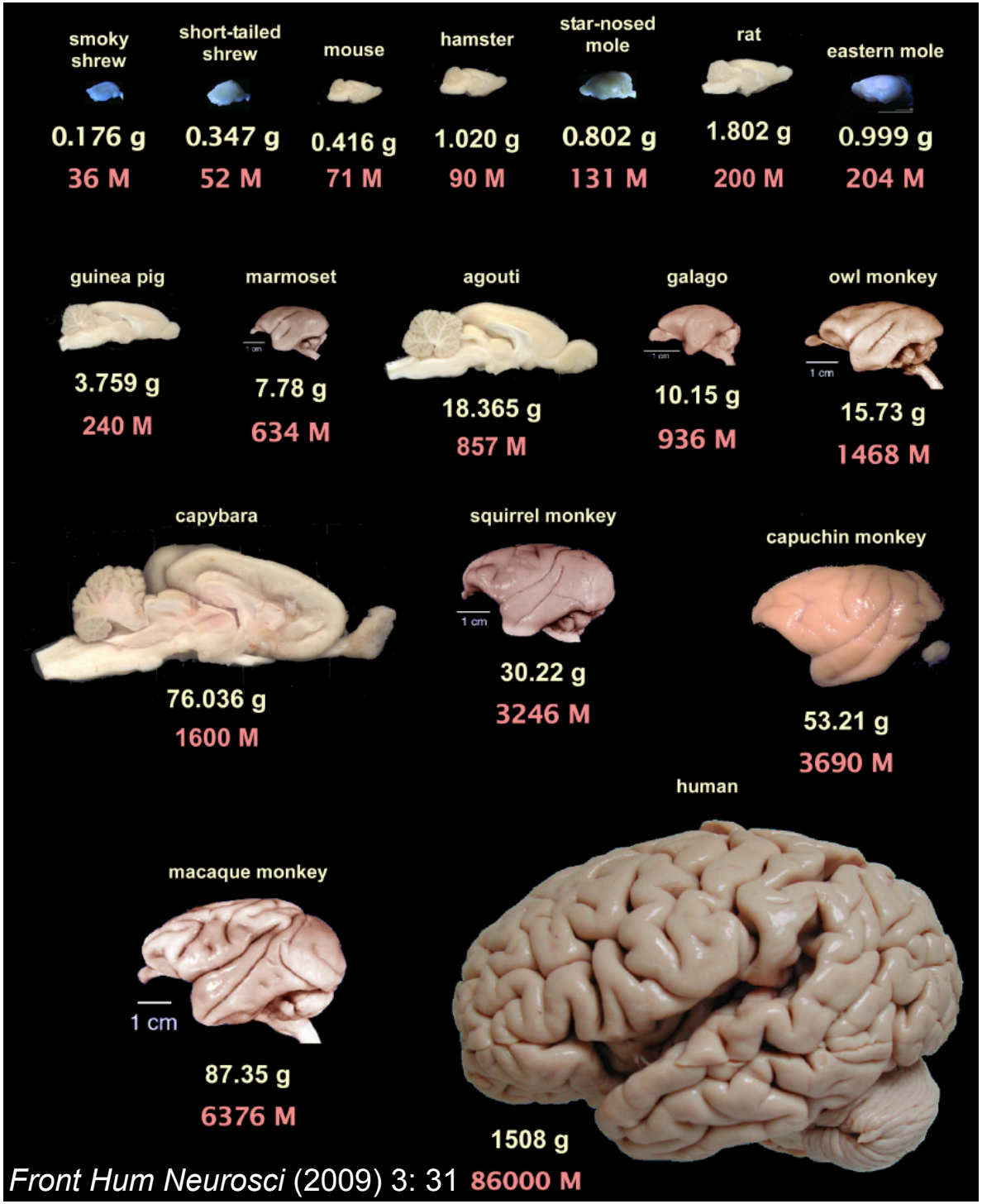


Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome
Science (2009) 326: 289

3D Structure of the Genome



A three-dimensional model of the yeast genome
Nature (2010) 465: 363



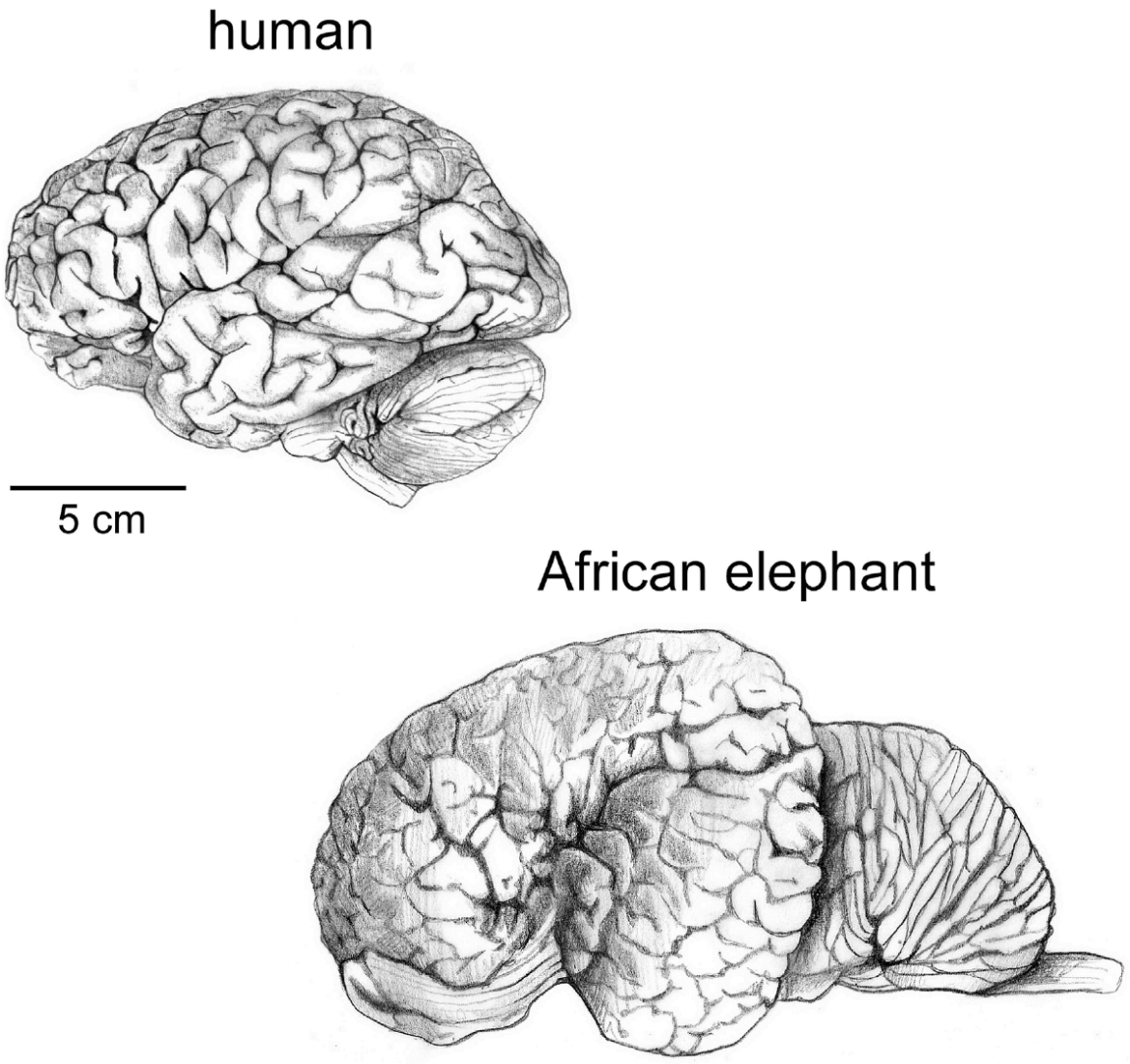
Front Hum Neurosci (2009) 3: 31

Brainspan

Acknowledgements

- Andrea
- Alex
- Jasmine
- Lukas
- Mark

- Sestan lab



Front Hum Neurosci (2009) 3: 31