How modENCODE informs us about human chromatin

Jason Lieb
Carolina Center for Genome Sciences
University of North Carolina at Chapel Hill

What chromatin data did we generate?

A total of 291 Worm datasets,
mostly over 3 developmental timepoints:
Early embryo, 3rd Larval stage, and adult

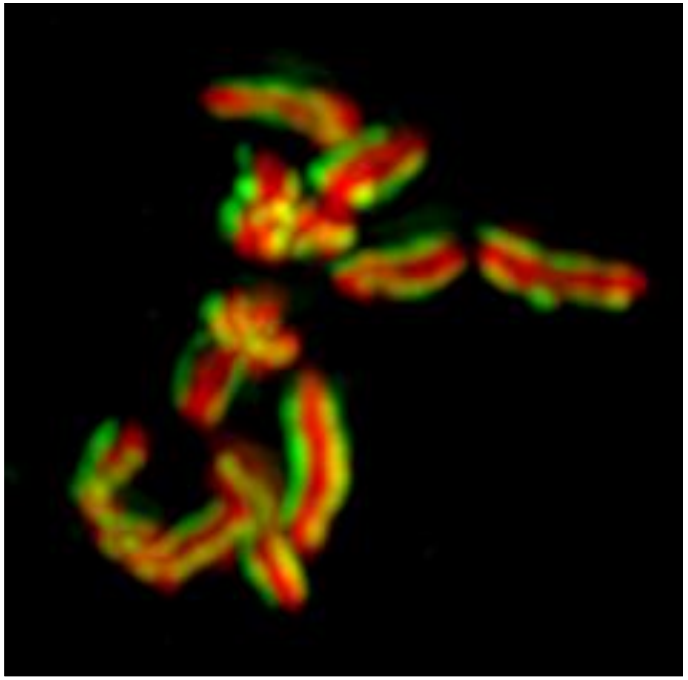
133 **profiles of 30 different histone marks**
147 **profiles of 72 different non-histone chromosomal proteins**
451 **polyclonal antibodies, 288 validated by at least one assay**

A total of 601 Fly datasets
mostly over 3 cell lines
plus early embryo, 3rd Larval stage, and adult
288 **profiles of 28 different histone marks**
313 **profiles of ~50 non-histone chromosomal proteins**

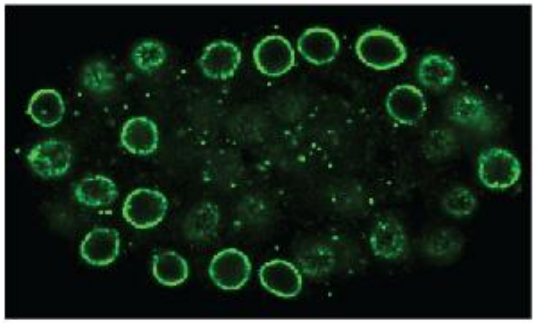
See talks by Gerstein (networks) & Kellis (states)

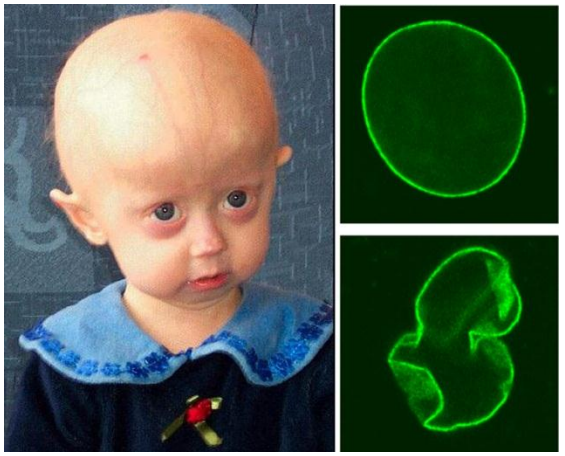
2 stories: Unique insights gained by modENCODE

(1) Centromere specification



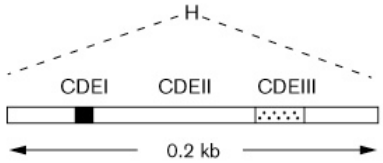
(2) Chromosome-membrane interactions



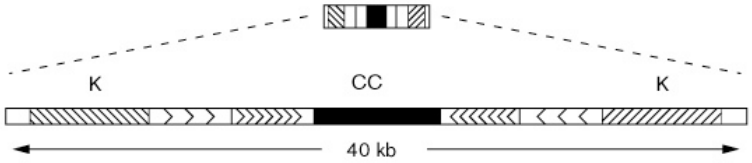


PLoS Biology Vol. 3/11/2005, e395

Centromeric DNA is not conserved

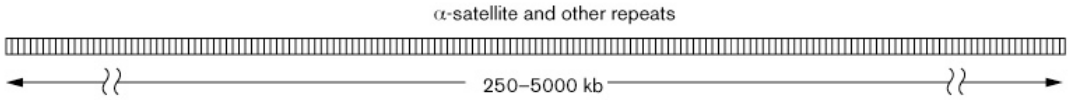


Budding Yeast: 125 bp



Fission Yeast: 40,000-100,000 bp

Fruit Fly: ~400,000 bp

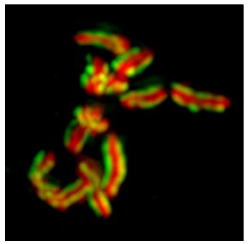


Humans: up to 5 million bp

α-satellite: NEITHER NECESSARY NOR SUFFICIENT

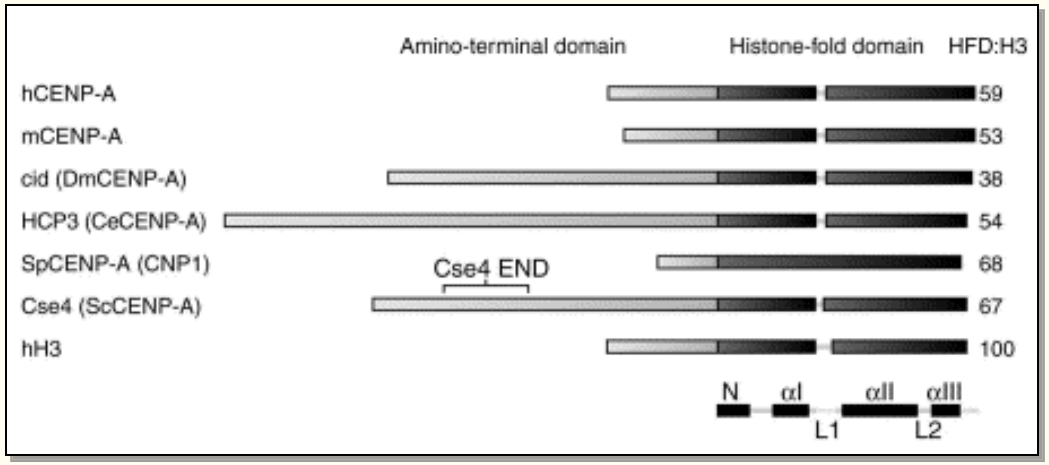
HOLOCENTRICS

some
Nematodes, Insects, and Plants



***C. elegans*: ~15 Mb Chromosomes**

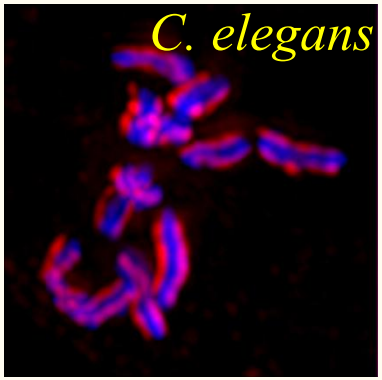
Kinetochores assemble on chromatin containing the histone H3 variant CENP-A (CenH3)

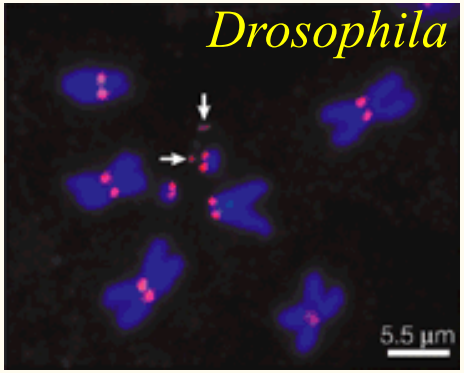


CENP-A and its chaperone HJURP over-expressed in many human cancers.

In flies, CENP-A over-expression results in ectopic sites, ectopic kinetochores, missegregation and aneuploidy.

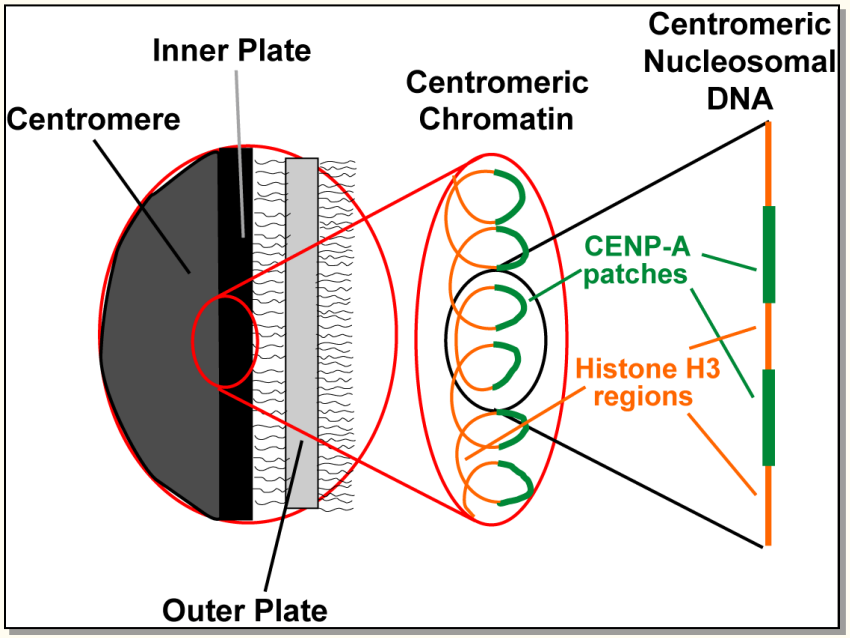
DNA/CENP-A

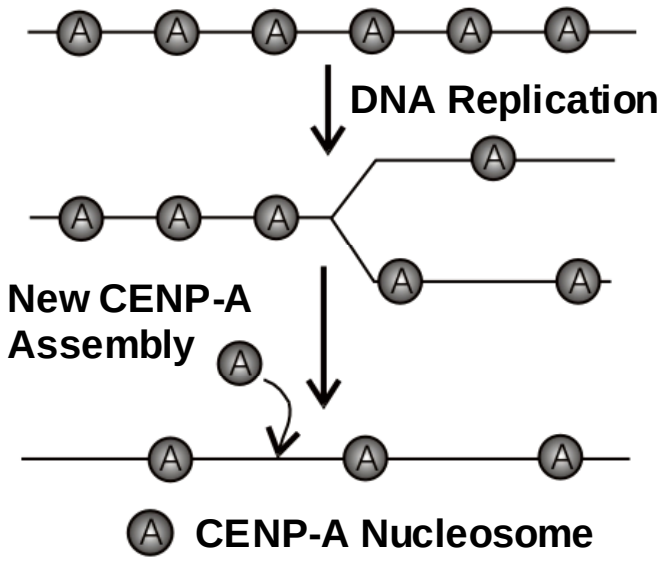




- **substitutes for H3 in (altered) nucleosomes**
- **located at the base of the kinetochore**

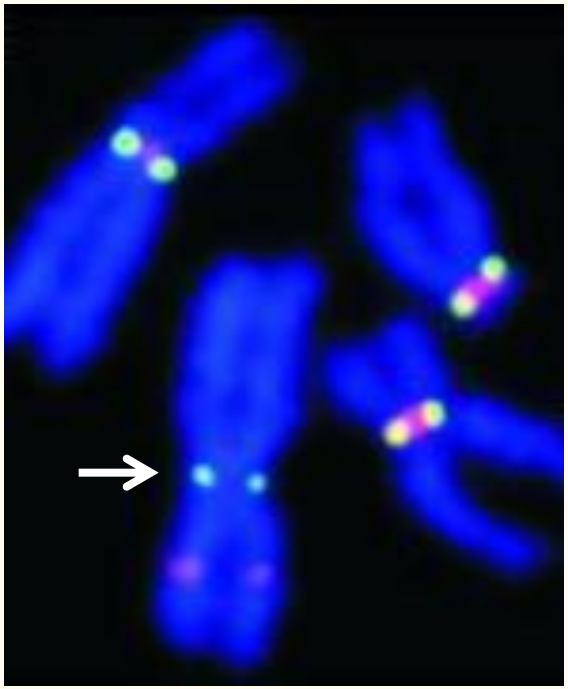
Inheritance of CENP-A nucleosome domains is postulated to propagate centromere identity ("OLD" → "NEW")

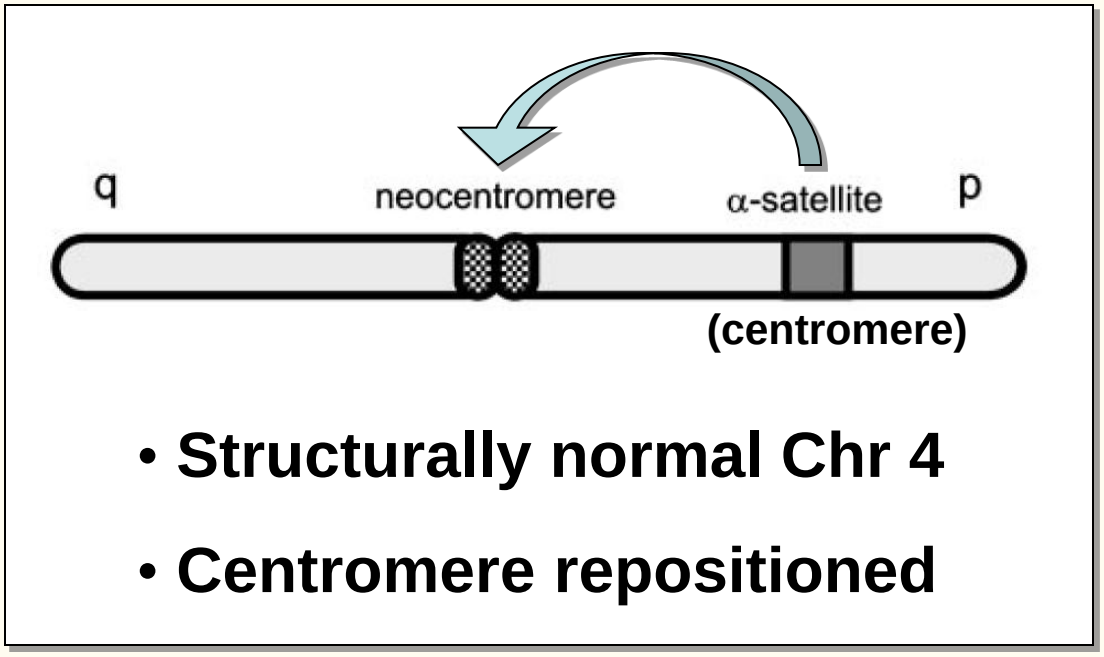




- Uncoupled from DNA replication
- Requires specialized machinery (e.g. KNL-2, HJURP/Scm3)
- Occurs in late anaphase-early G1 in mammals, *Drosophila*

Human Neocentromeres: Evidence for rare *de novo* CENP-A chromatin domain establishment





- Structurally normal Chr 4
- Centromere repositioned

CENP-A

Alpha-satellite (CEN DNA)

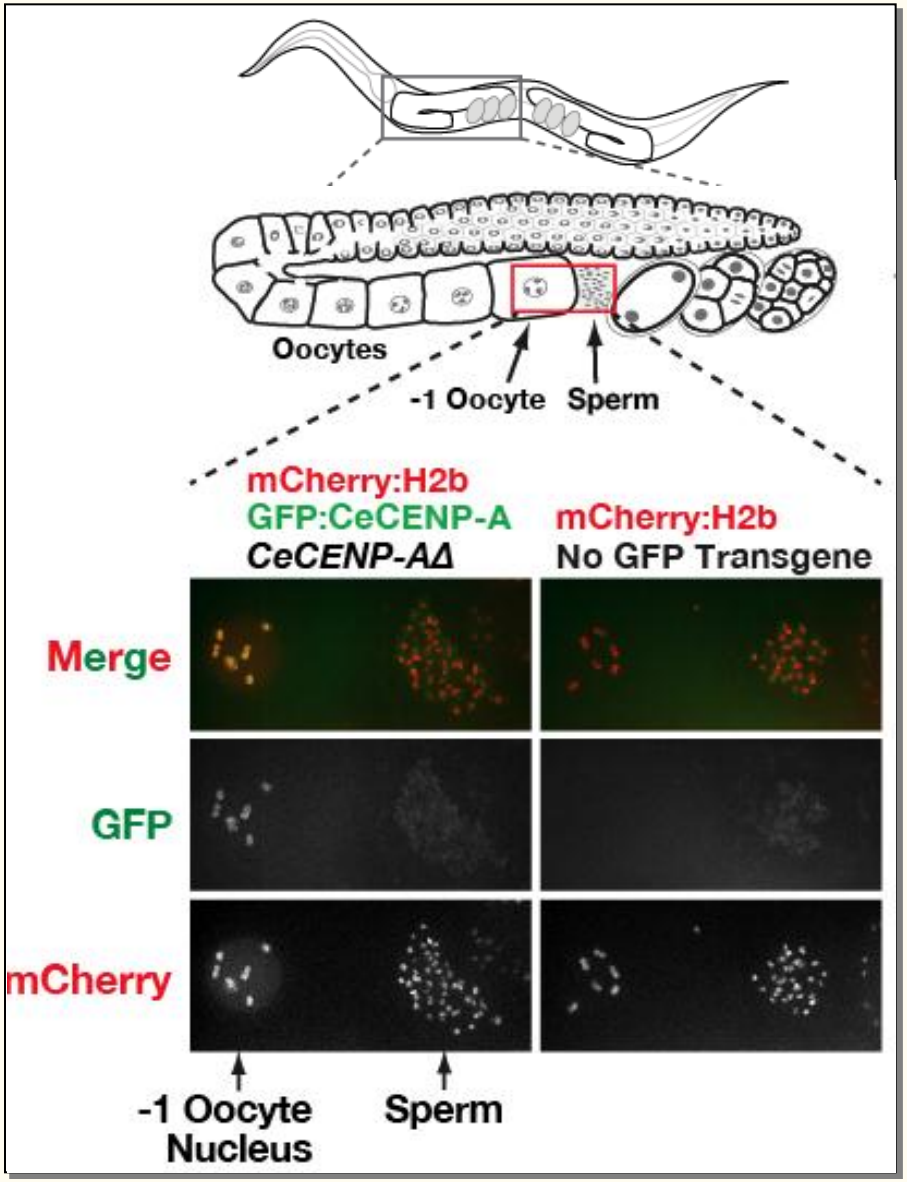
Andy Choo & colleagues

Amor et al. *PNAS* (2004)

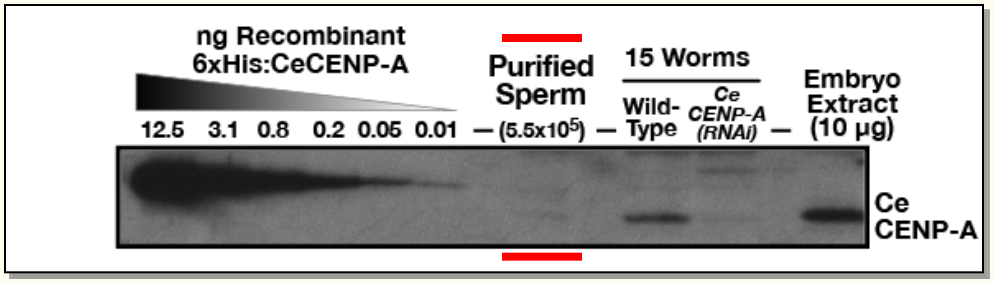
- Neocentromerization observed in diverse species
- Centromere repositioning common among related species

CENP-A is not detectable in *C. elegans* sperm

Microscopy

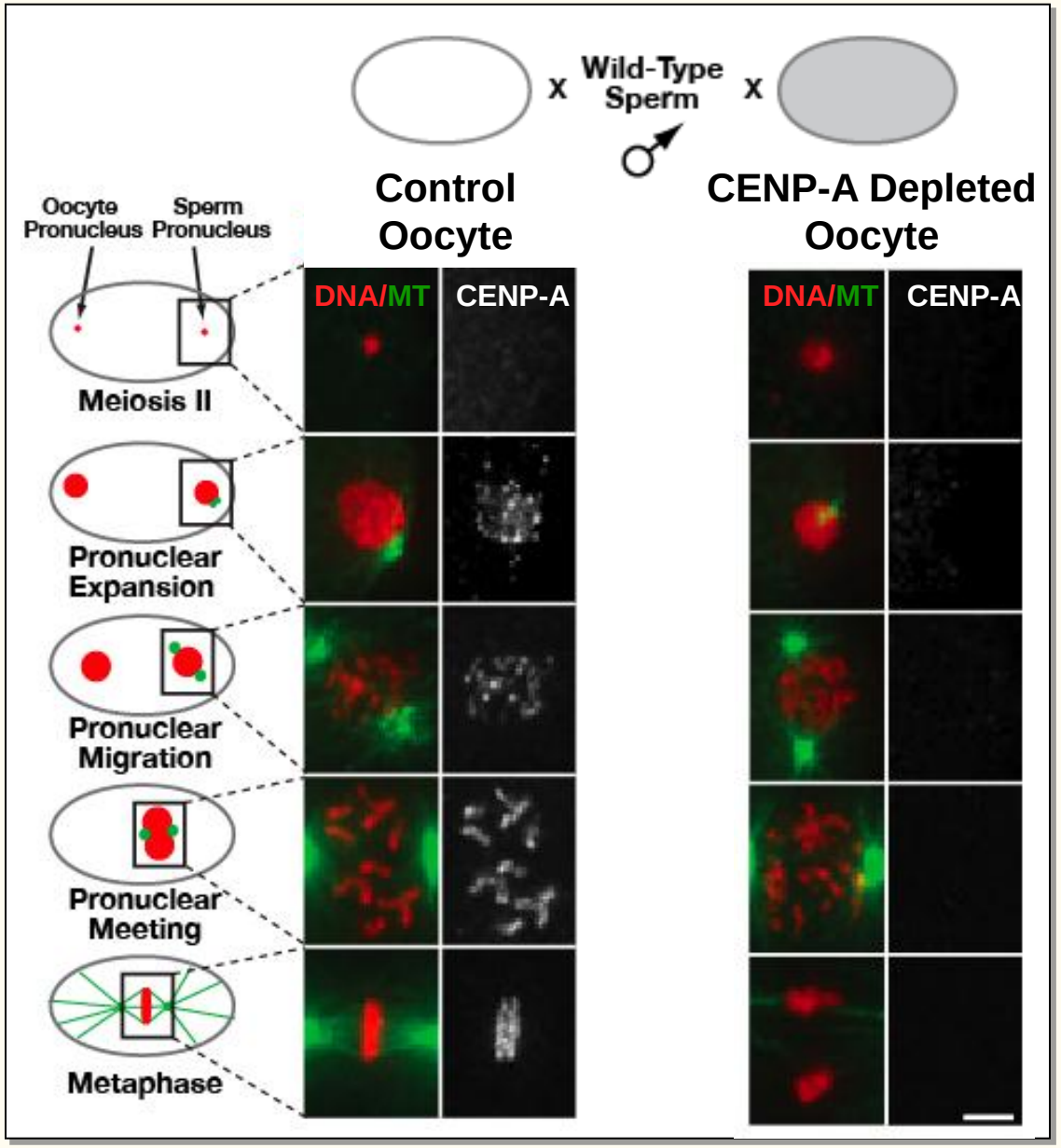


Quantitative immunoblotting



- **<300 CENP-A molecules / sperm**
- **Embryos have ~130,000 CENP-A molecules per nucleus (purified nuclei)**

Sperm chromatin recruits CENP-A from oocyte cytoplasm

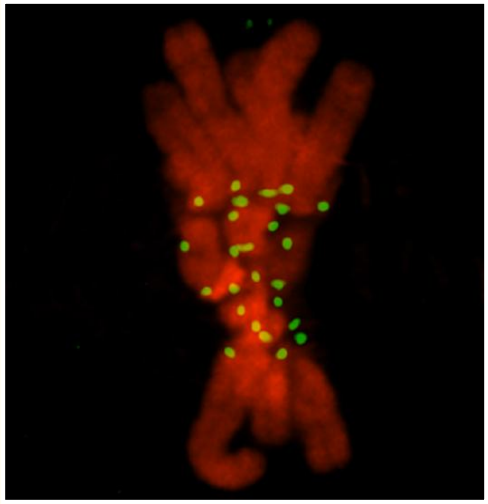


- **CENP-A Absent from Mature Sperm & Recruited From Oocyte Cytoplasm at Fertilization**
- **CENP-A Removal & Reloading in Meiotic Prophase**

Inconsistent with "OLD" →"NEW" Model

If not OLD CENP-A nucleosomes, what epigenetic features restrict targeting of CENP-A?

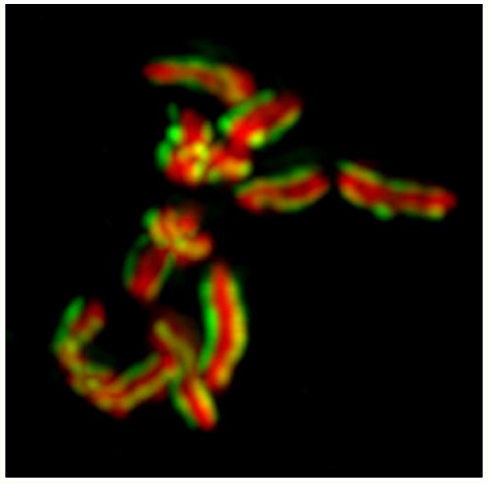
Monocentric



Repetitive Architecture of human centromeres has prevented mapping of CENP-A location

---

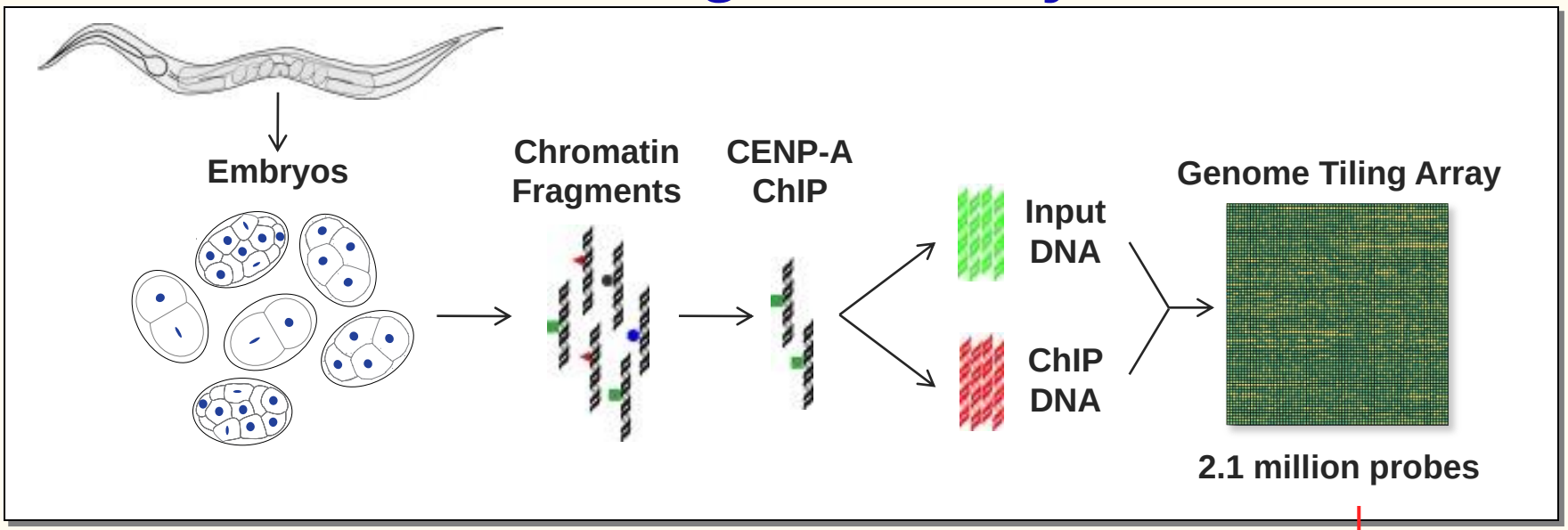
Holocentric

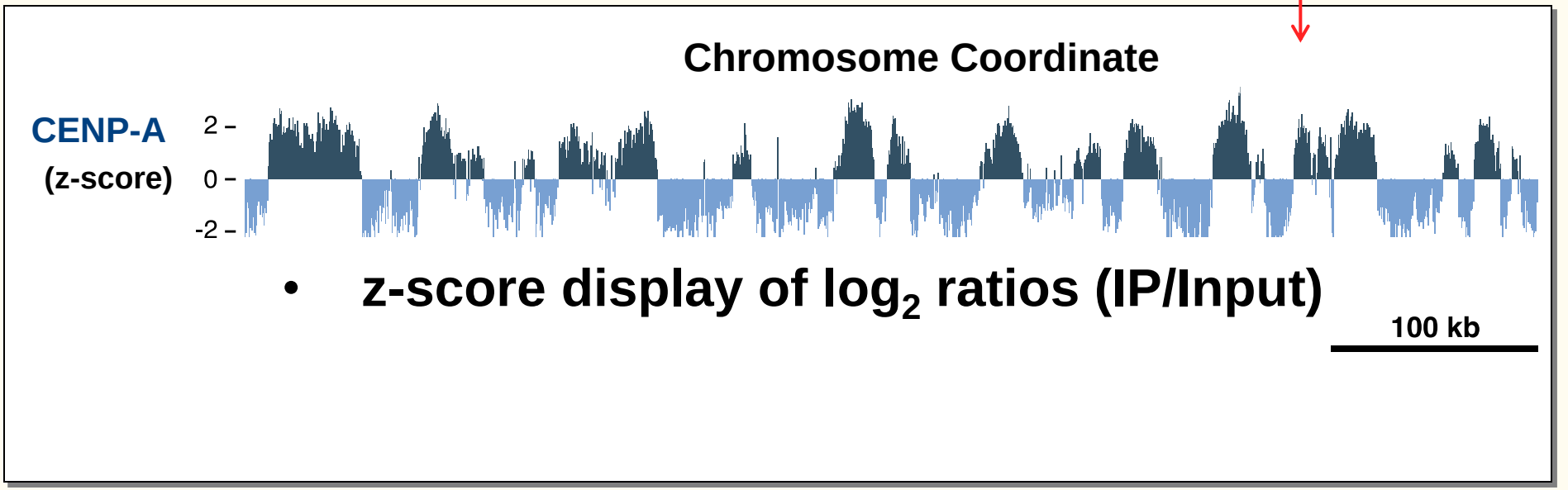


CENP-A Chromatin IP

Genomic Tiling Microarrays (Nimblegen)

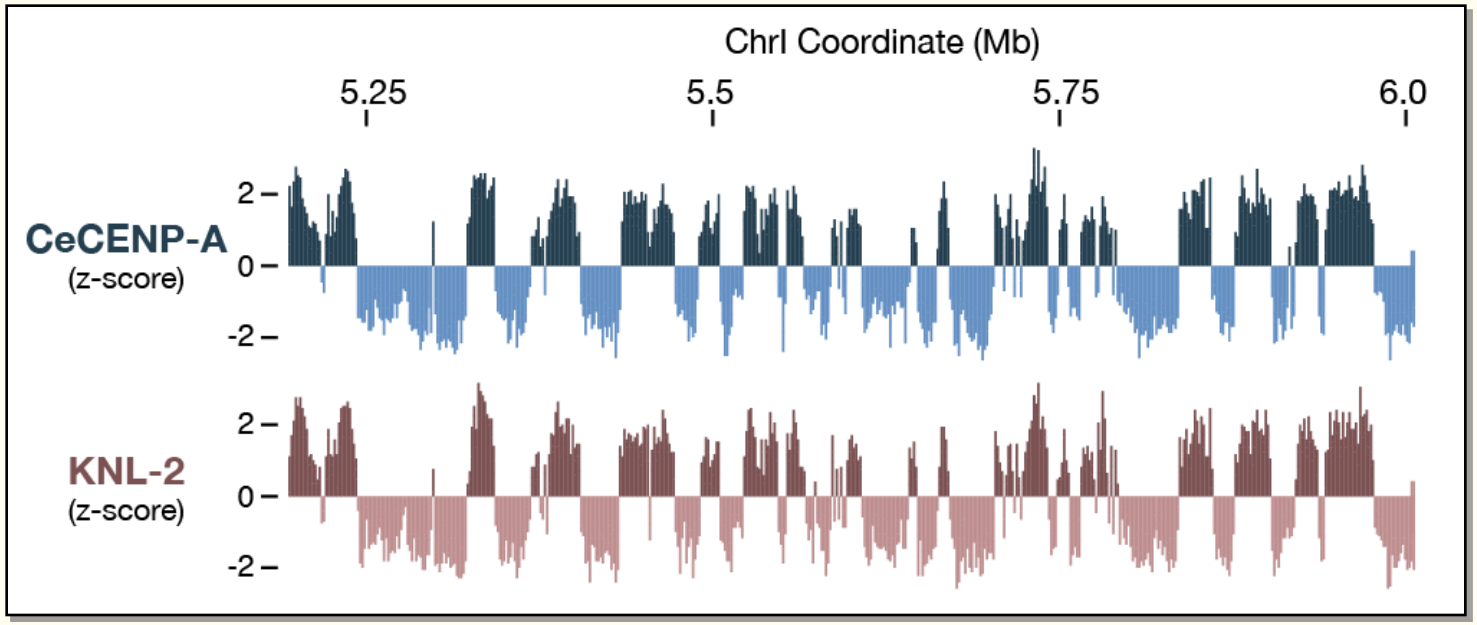
Mapping CENP-A (& KNL-2) location genome-wide in *C. elegans* embryos

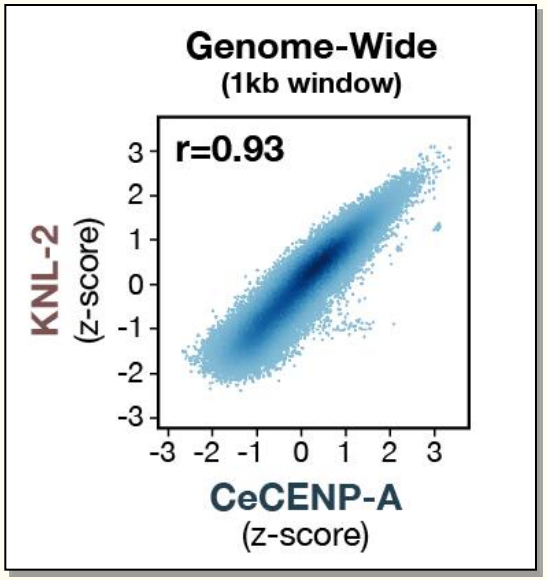




- z-score display of $\log_2$ ratios (IP/Input)

CENP-A & KNL-2 exhibit identical genomic distribution

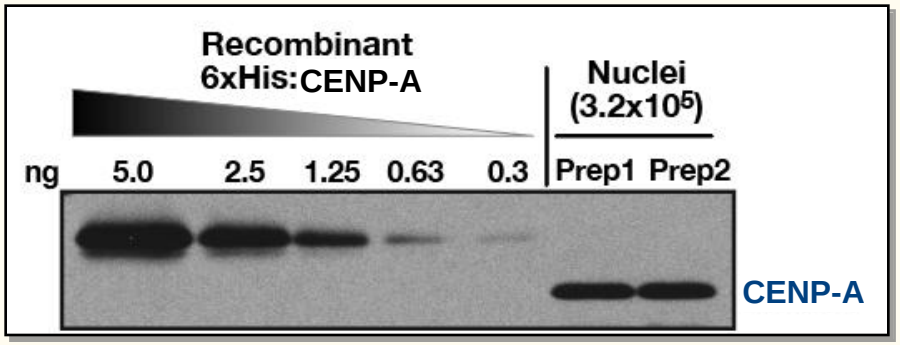




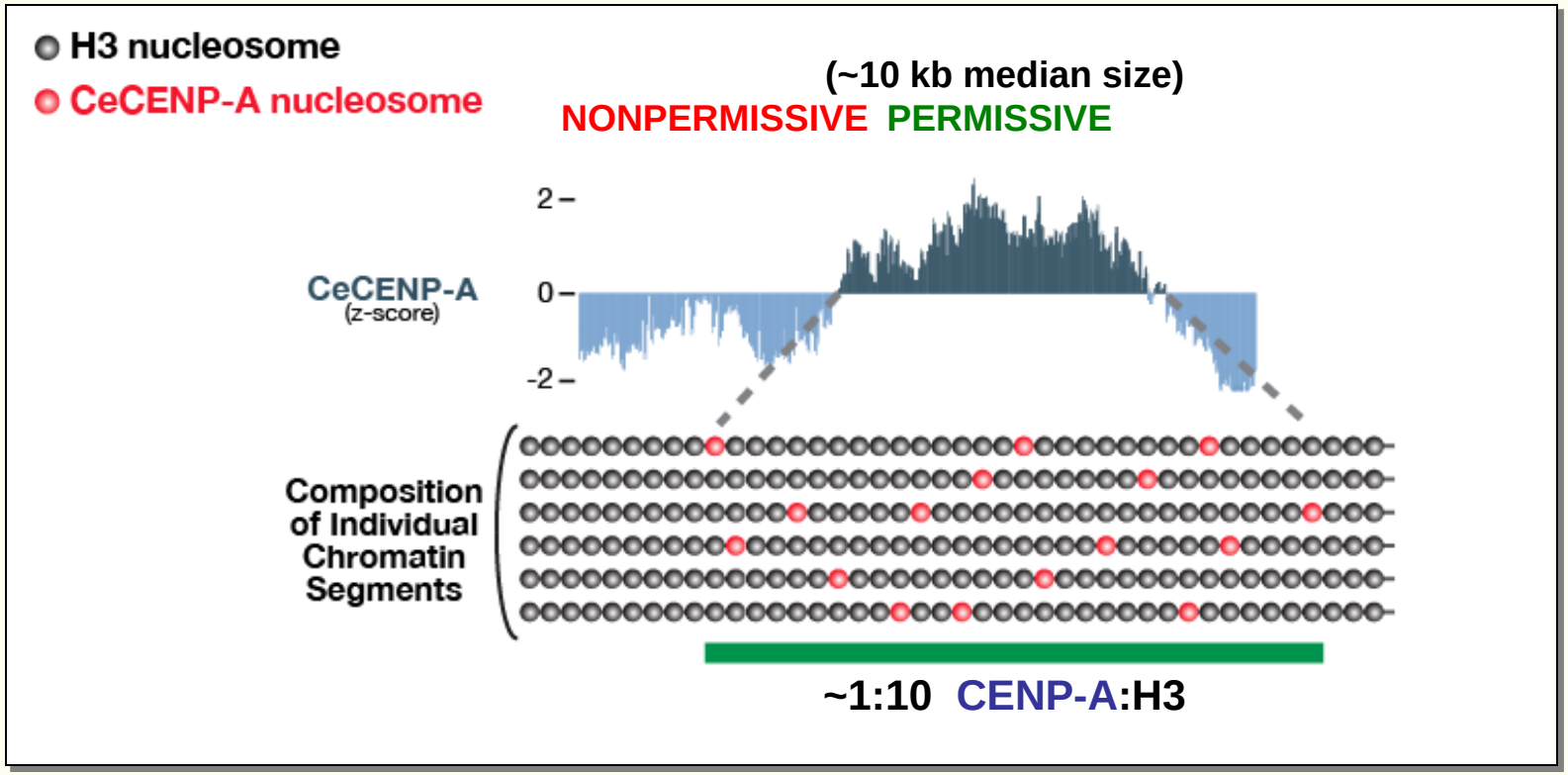
- **Broad Distribution (47% of Genome)**

Counting CENP-A molecules per nucleus to interpret structure of CENP-A enriched domains

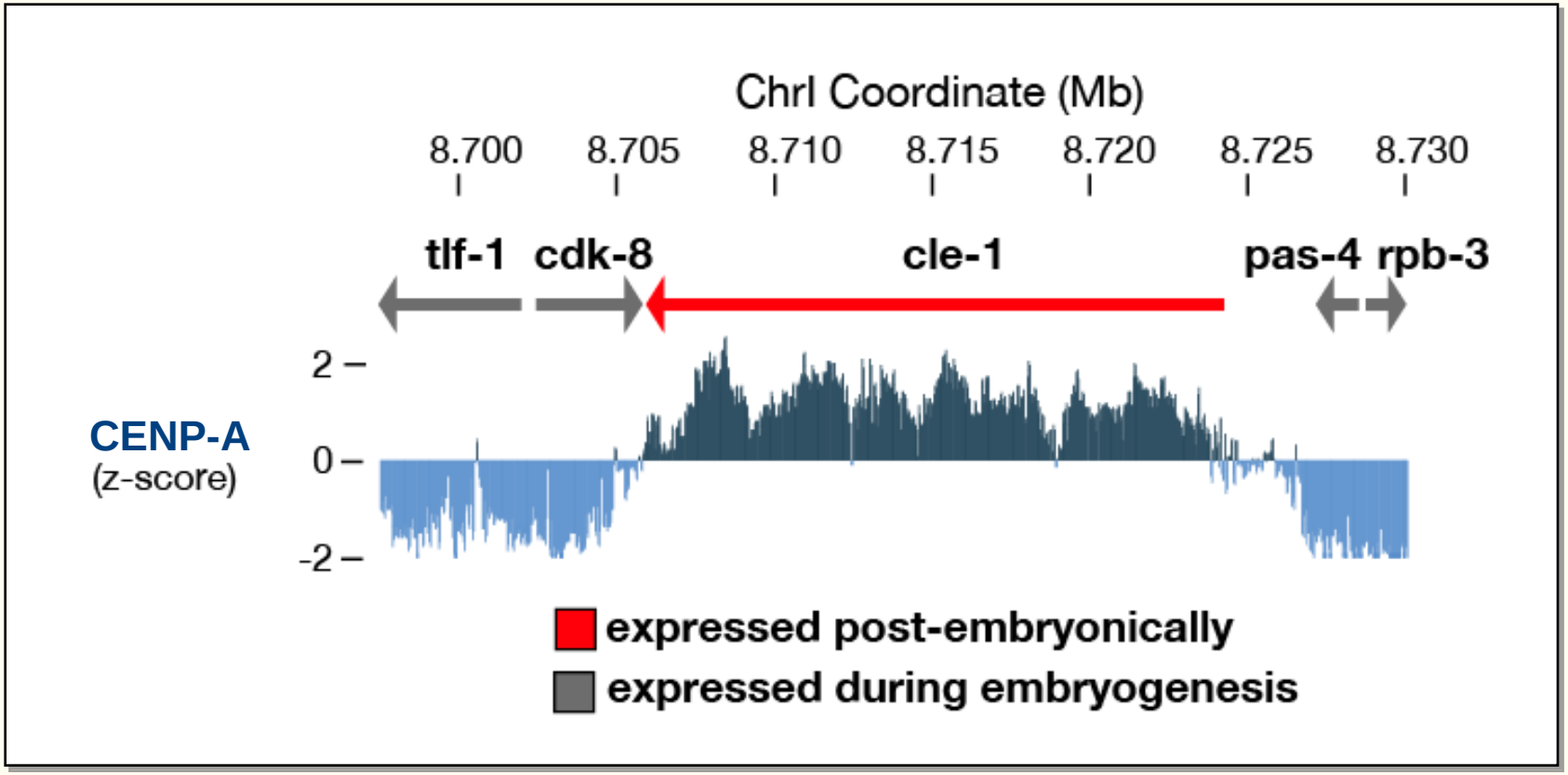
Maximally, 4% of the nucleosomes CAN be CENP-A nucleosomes



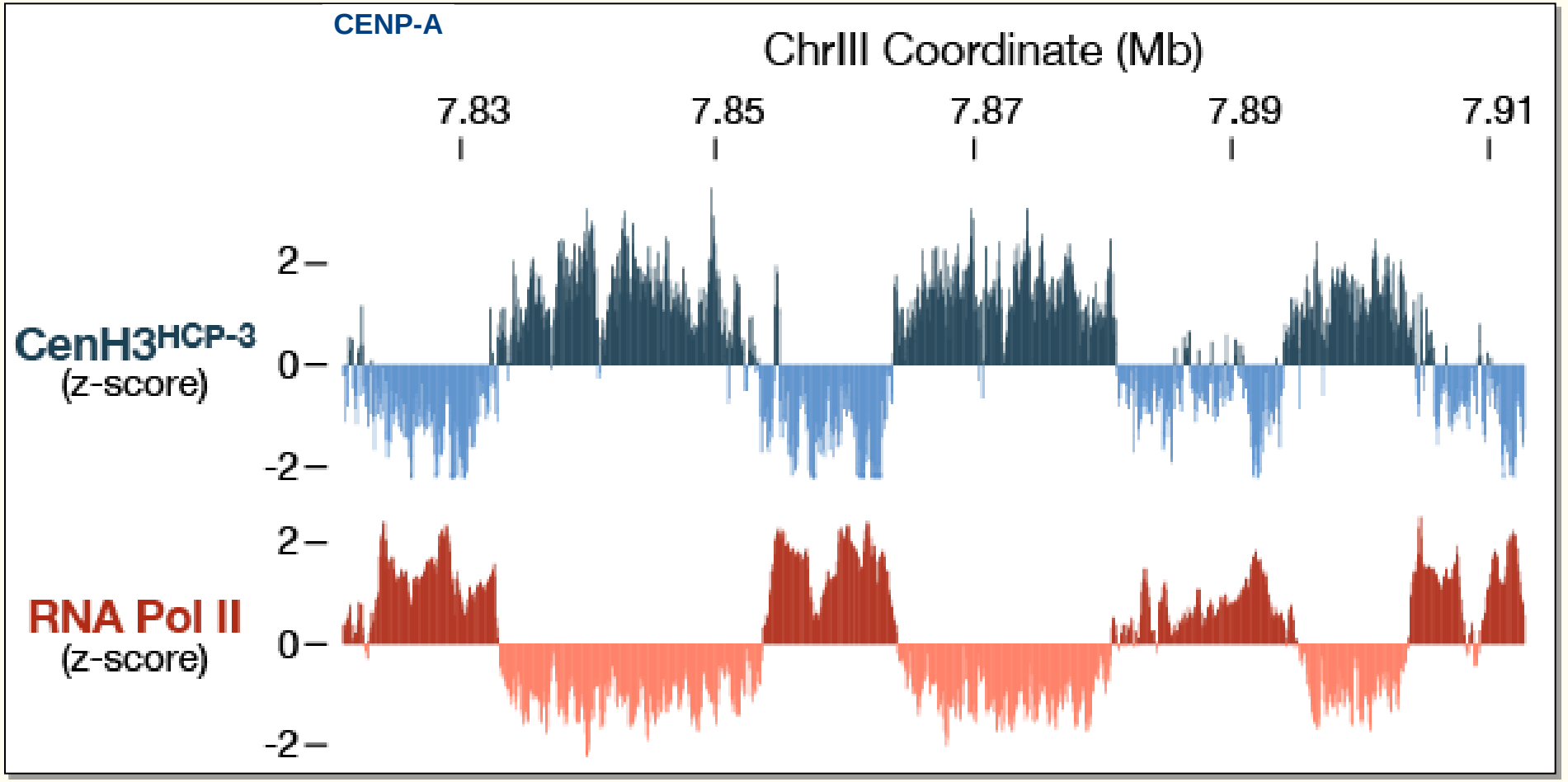
1.3 x 10^5 CENP-A molecules/nucleus



In embryos, CENP-A domains appear to be delimited by gene expression



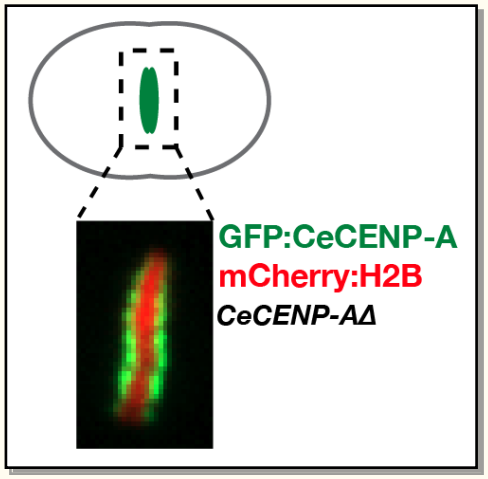
CENP-A & RNA Polymerase II Occupancy are Inversely Correlated



Is CENP-A simply excluded from regions of active transcription?

- **No significant zygotic transcription until 30-cell stage**

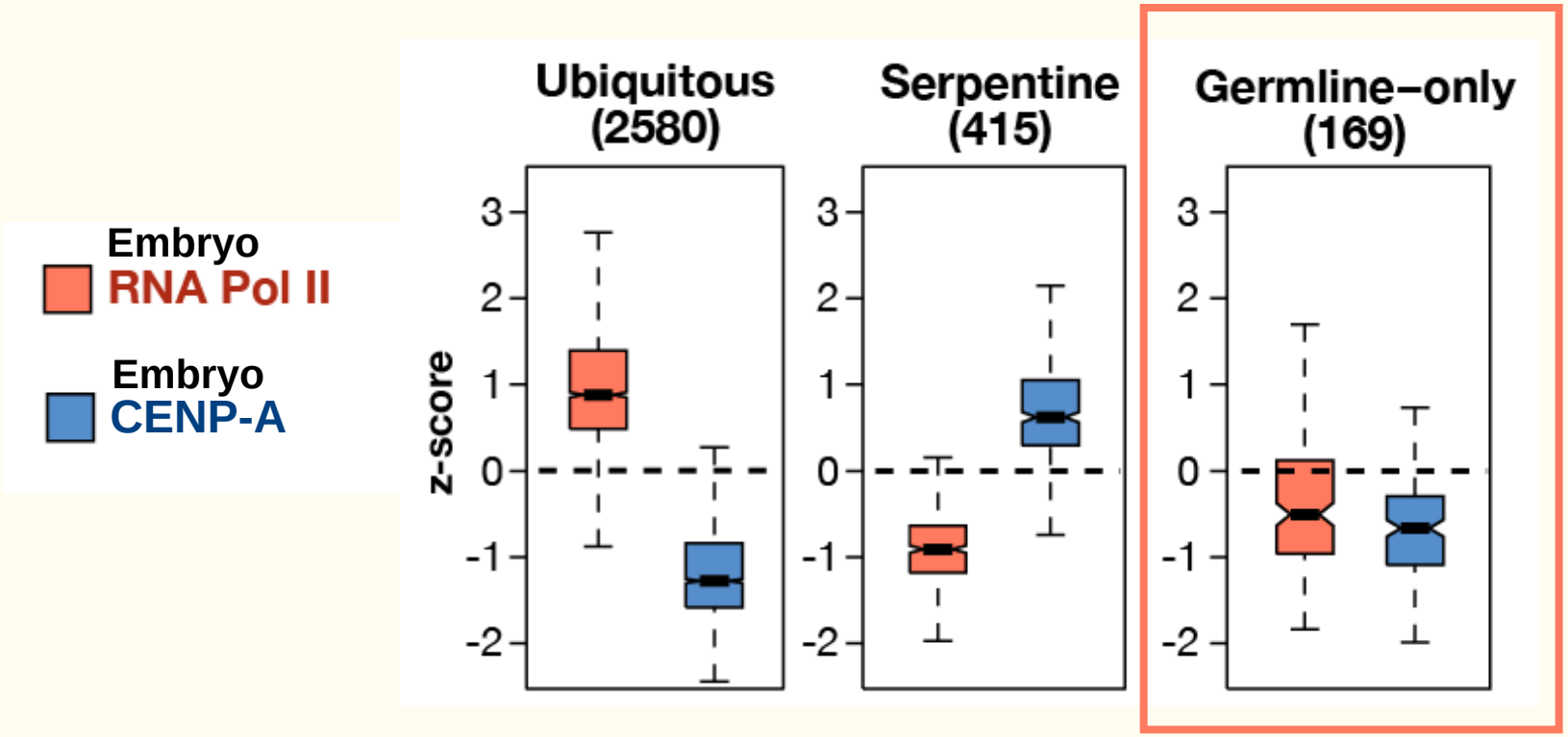
1-Cell Embryo



- **Early divisions are normal in pol II-inhibited embryos**

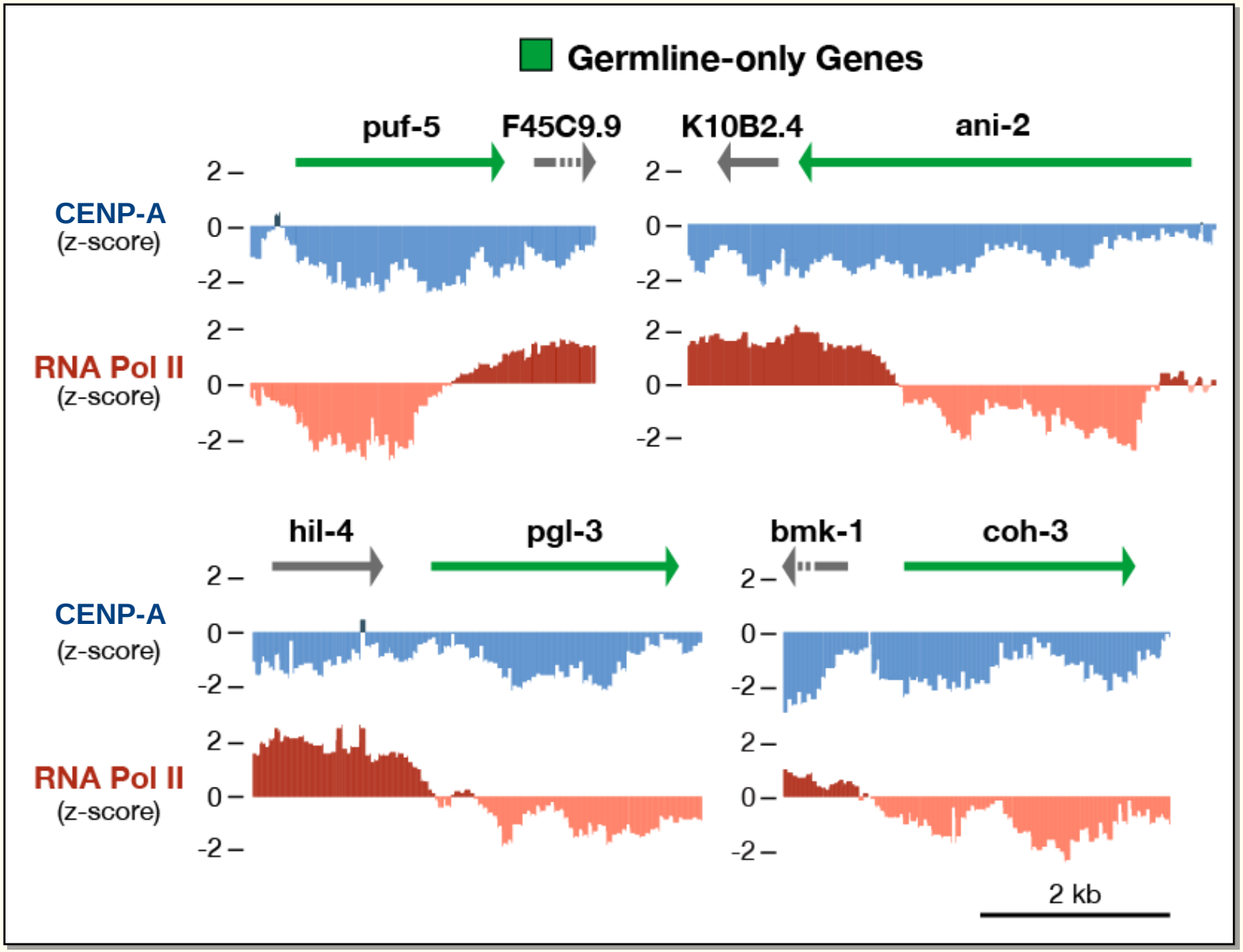
- **No change in CENP-A pattern during development (from 8-cell stage to >250 cell stage)**

Gene set analysis: A clue to the origin of the underlying pattern?

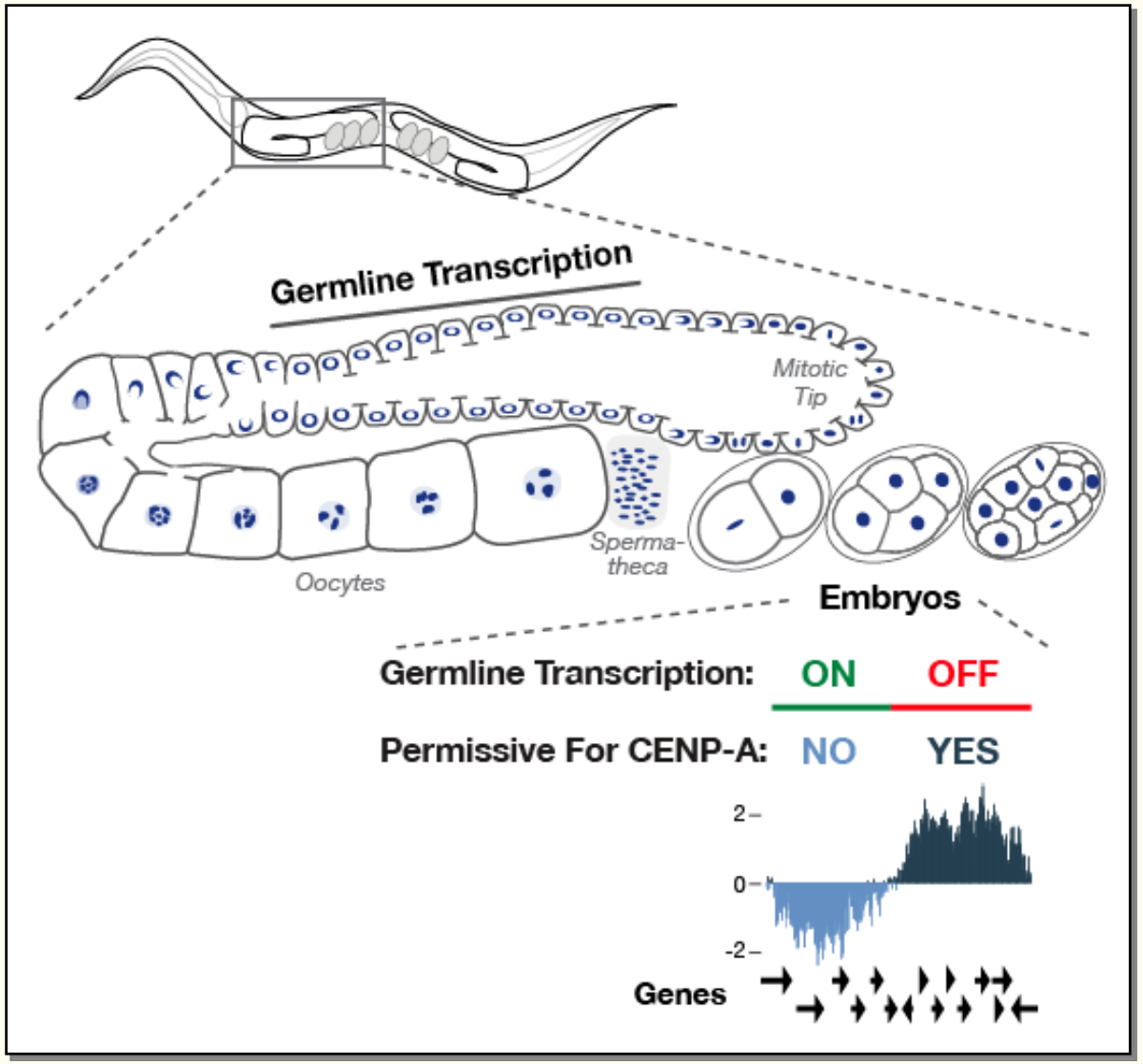


Germline-only = expressed in germline but not in embryos or other tissues

Example “DOUBLE NEGATIVE” germline-only genes

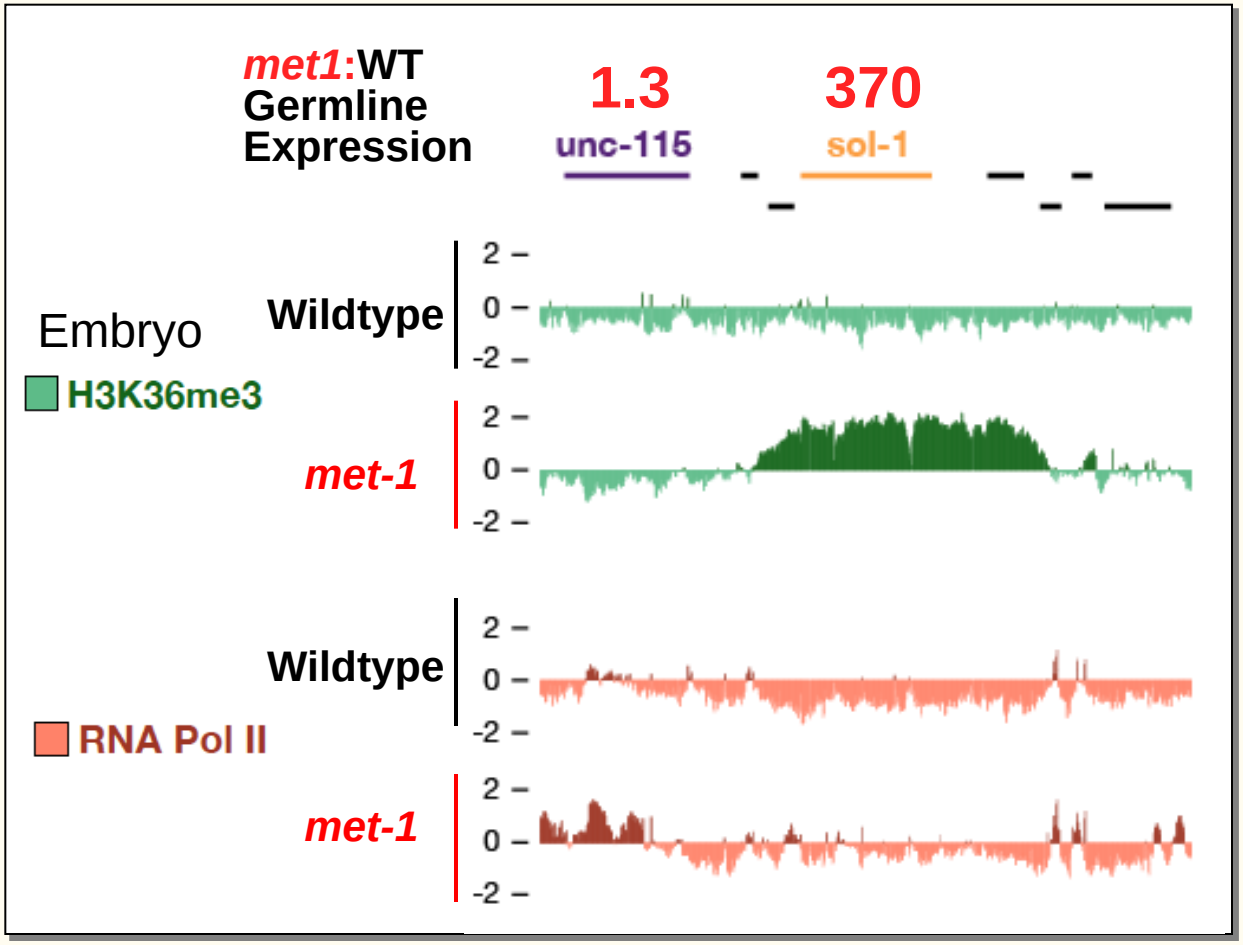


Model: Germline transcription defines regions of CENP-A exclusion throughout embryogenesis



Testing the Model: Does ectopic expression in the germline convert a CENP-A^{+} region into CENP-A^{-} in embryos?

- **Mutant with Altered Germline Expression**



(met-1)

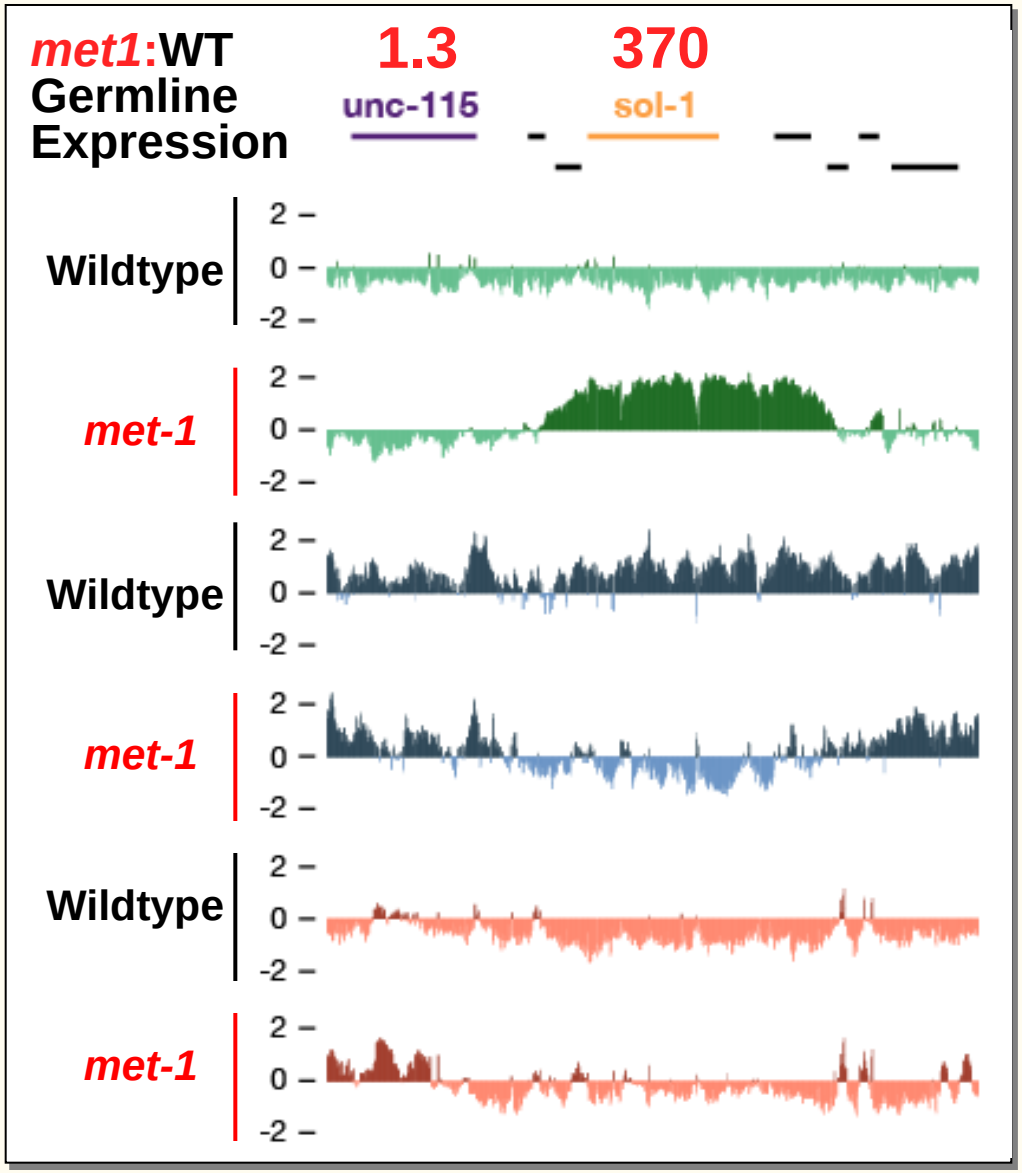
non-essential H3K36 methyltranferase

- **Germline Expression**
  Measured in Dissected Gonad Tissue

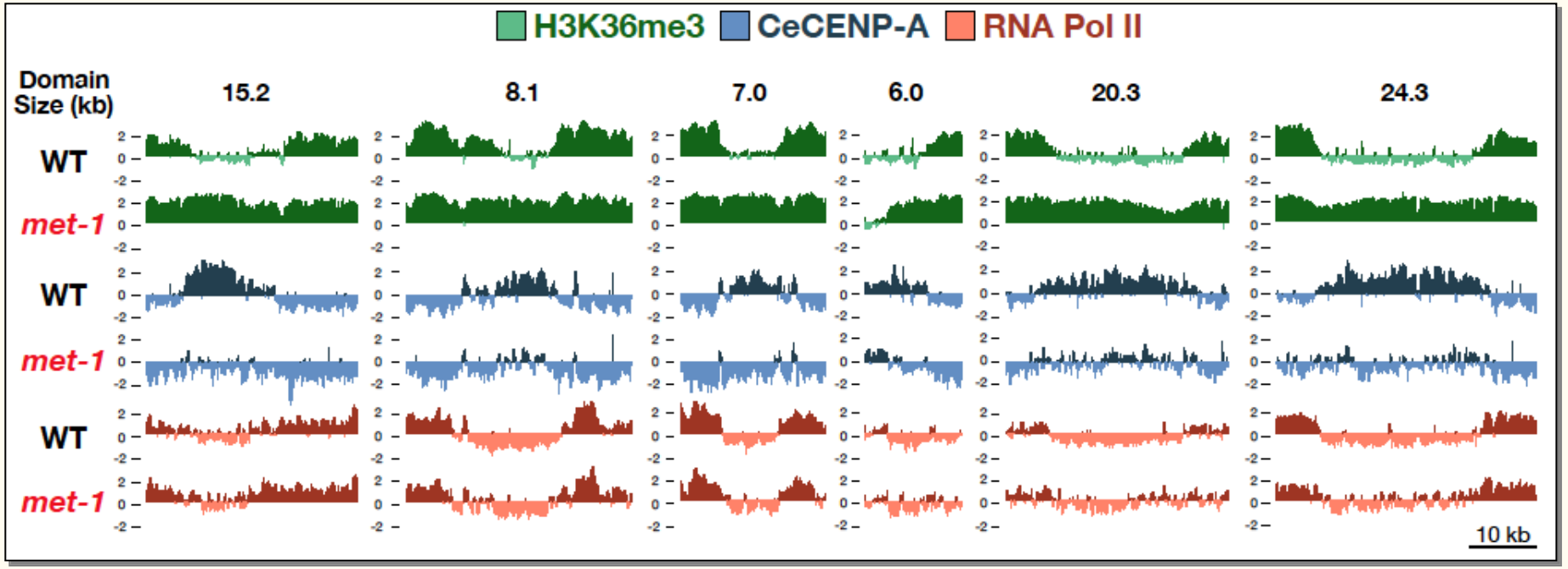
- **H3K36me3**
  Marks Germline-Expressed Genes Throughout Embryogenesis
  (Strome & Kelly Labs)

- **RNA Pol II: Marks Active Transcription in Embryos**

Ectopic germline expression leads to CENP-A loss without active transcription in embryos



Ectopic germline expression leads to CENP-A loss without active transcription in embryos



WT vs. *met-1*:

132 domains > 5kb size with H3K36me3 increase > 1 z-score in *met-1* vs WT

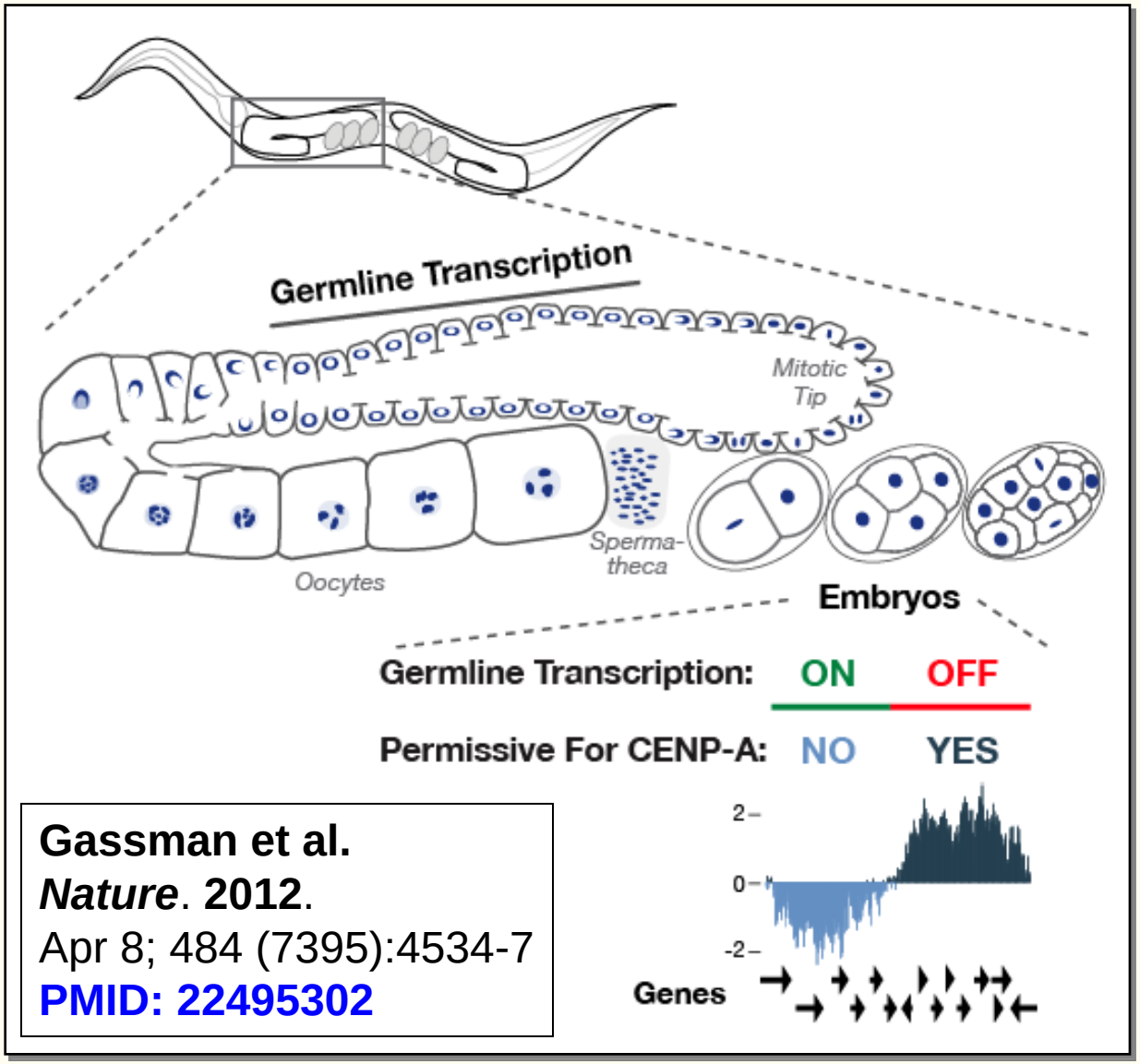
↓

75 domains with RNA Pol II z-score < 0.1 in *met-1*

↓

Average ΔCENP-A$^{met\text{-}1\ -\ WT}$ z-score = - 0.8

Model: Germline transcription defines regions of CENP-A exclusion throughout embryogenesis



Gassman et al.
***Nature*. 2012**.
Apr 8; 484 (7395):4534-7
PMID: 22495302

Mechanism Underlying "Memory"?

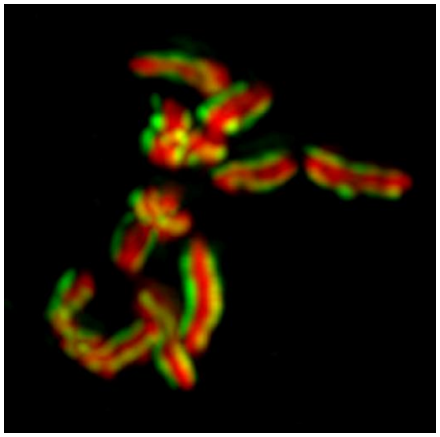
i) MES-4 (Major H3K36 Methyltransferase)
Strome & Kelly Labs

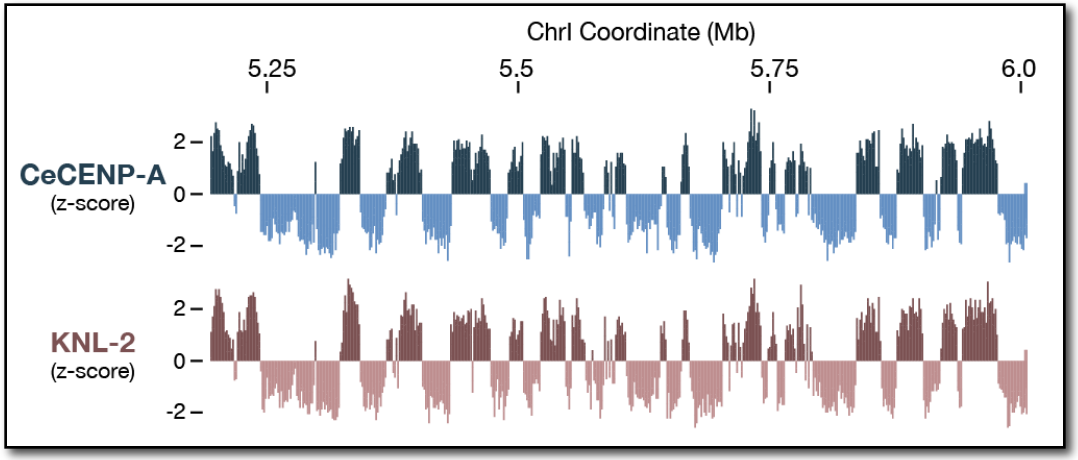
PLoS Genetics. 2010 Sep 2;6(9). e1001091.
PMID: 20824077

ii) CSR-1 Argonaute & 22G small RNAs
Mello Lab

Implications

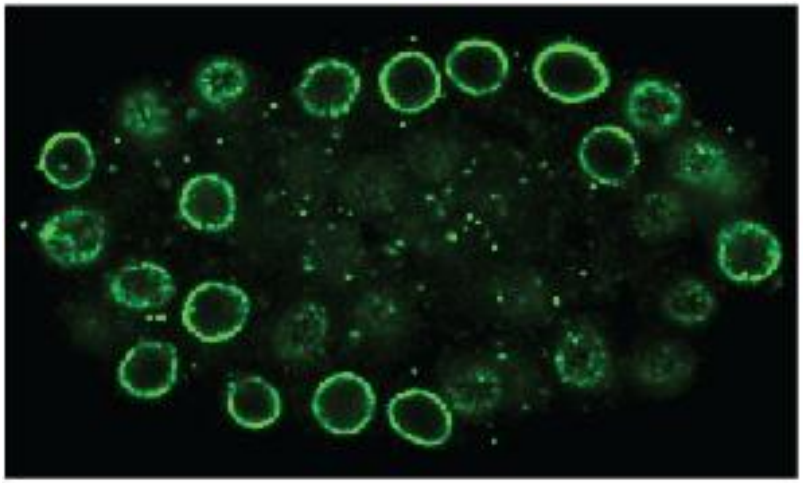
- **CENP-A nucleosomes can be "guided" by cues that are not pre-existing CENP-A nucleosomes.**
- **Trans-generational epigenetic memory of gene expression regulates histone variant incorporation.**
- **Germline expression may influence sites chosen for centromere repositioning during evolution.**



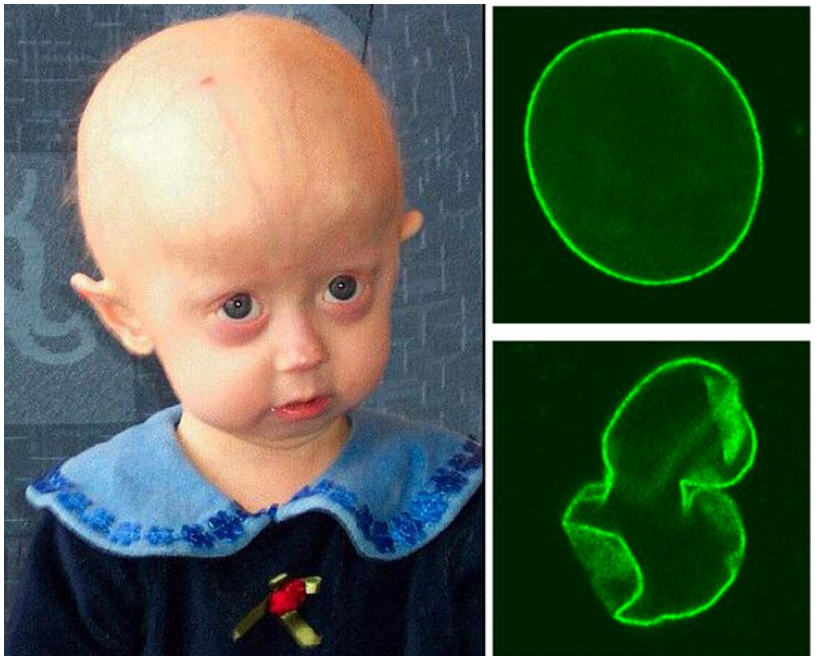


(2) Chromosome-membrane interactions

LEM-2 at the *C. elegans* nuclear envelope



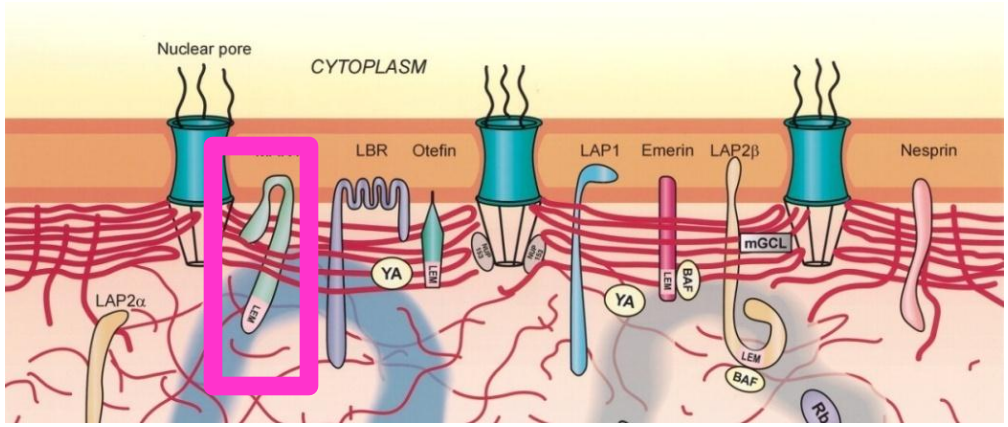
Hutchinson-Gilford Progeria, with aberrant nuclear morphology



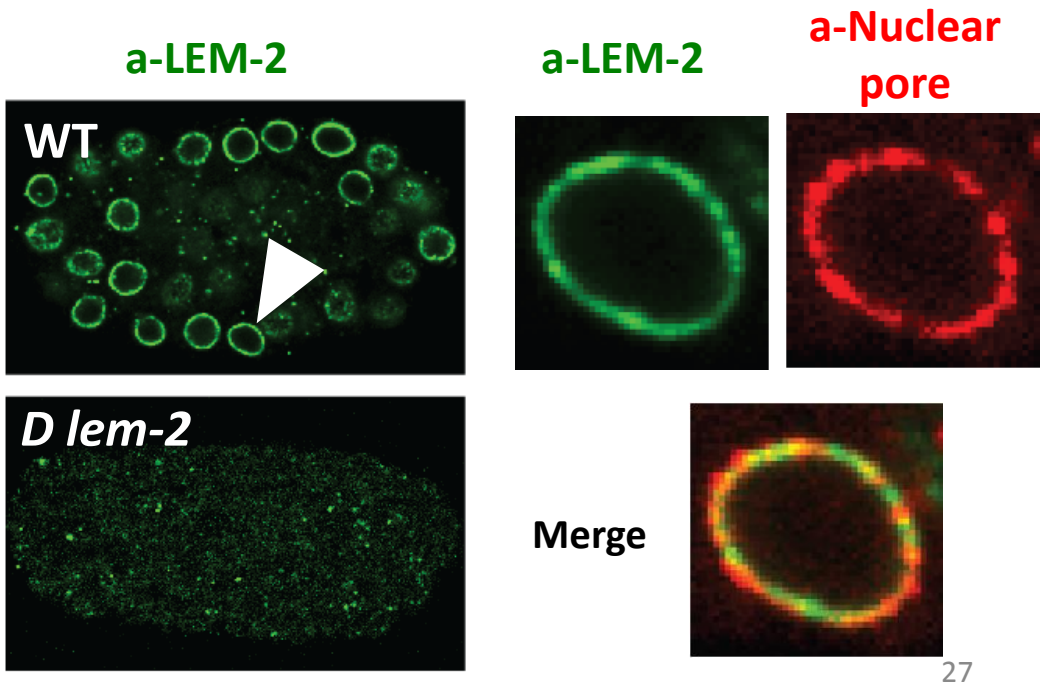
PLoS Biology Vol. 3/11/2005, e395

LEM-2, a nuclear membrane protein

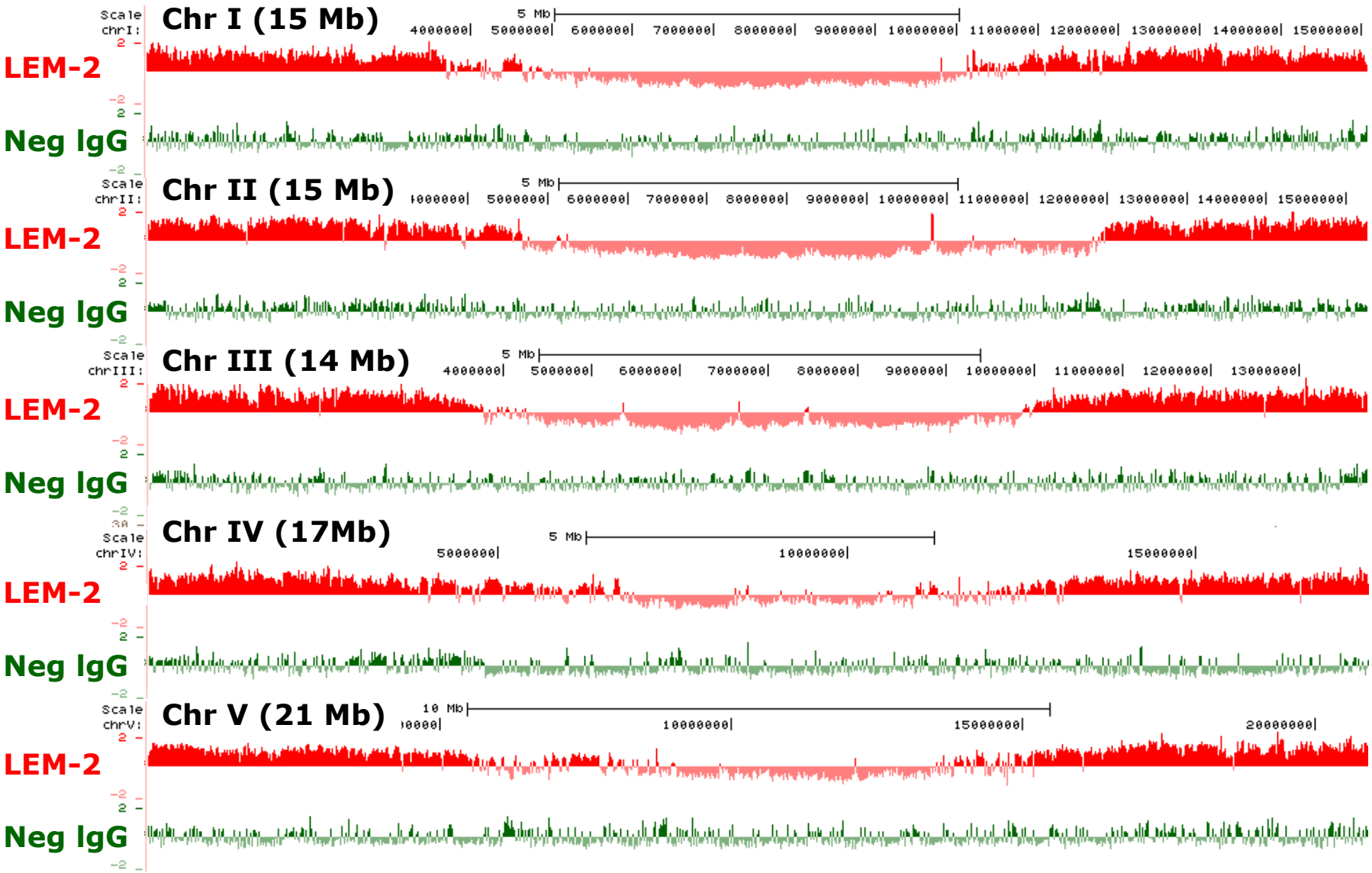
- Localized at the nuclear envelope; dependant on:
  - transmembrane domains
  - lamin A
- Conserved among yeast, *C. elegans* & mammals
- Ubiquitously expressed
- Not essential, but co-RNAi with emerin causes embryonic lethality in *C. elegans*
- Interacts w/DNA-binding protein BAF-1
- Associates w/ yeast chromatin



Goldman et al. *Genes Dev*. (2002)

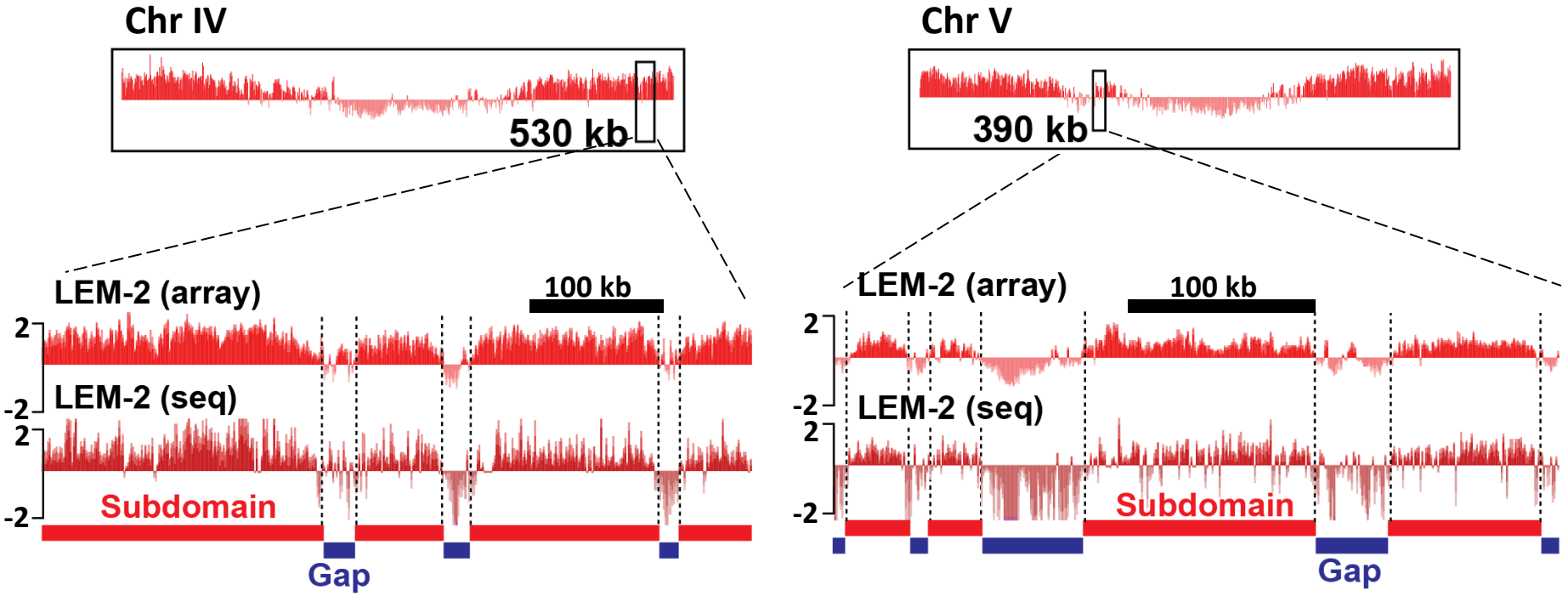


LEM-2 associates with chromosome arms



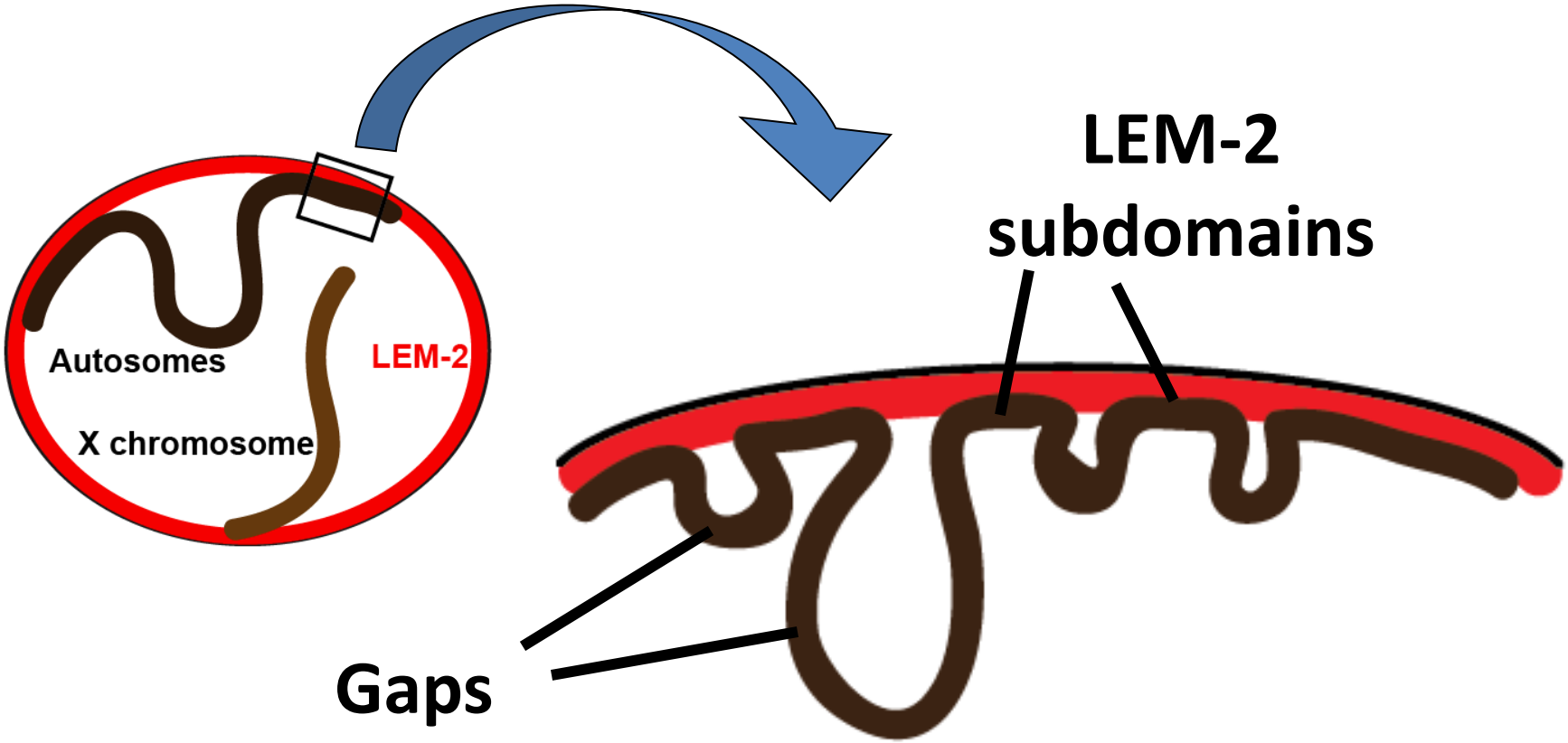
Ikegami *et al.*, Genome Biology (2010)

Large LEM-2 domains consist of small subdomains

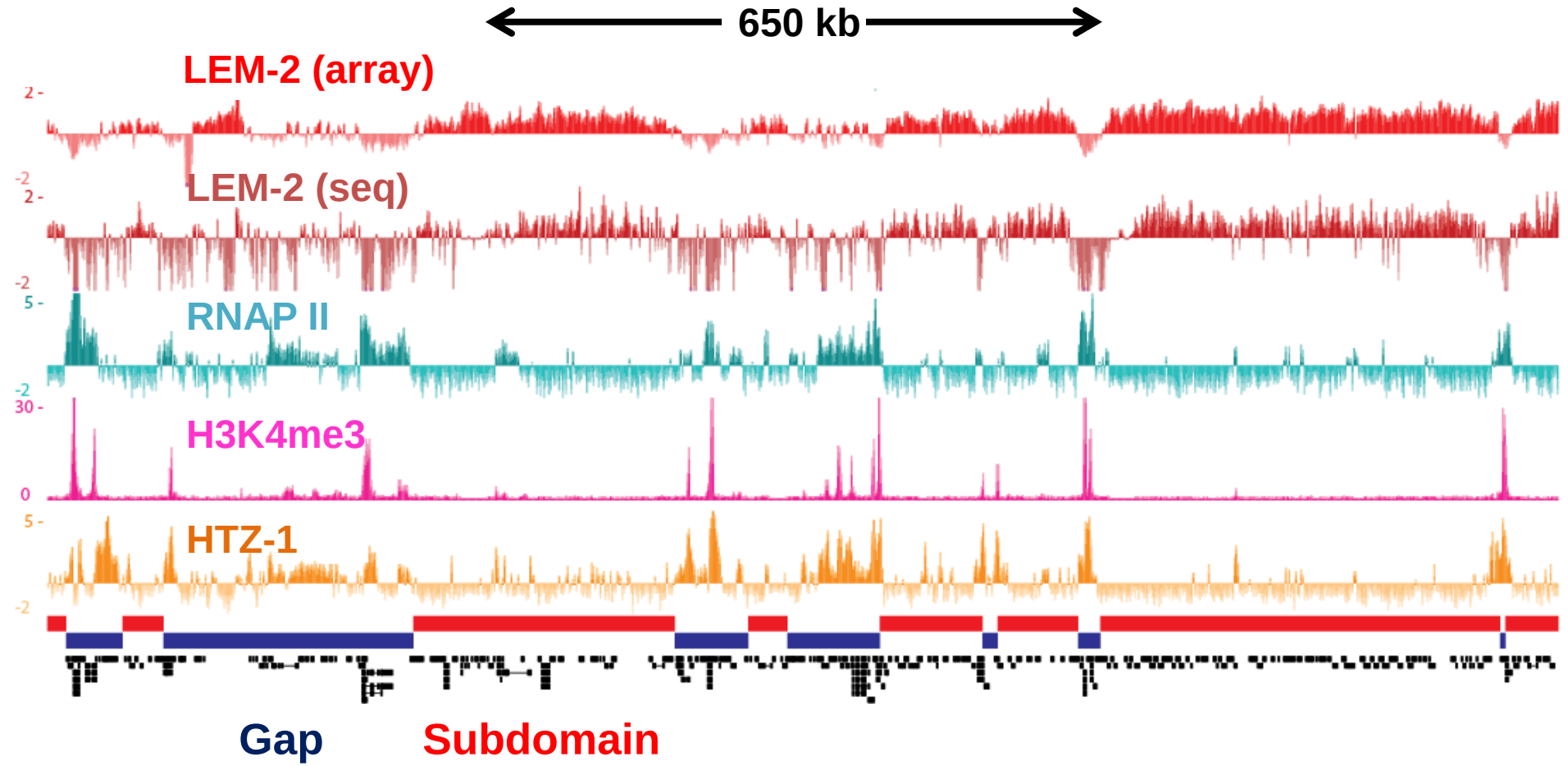


LEM-2 domain: ~4 Mb
LEM-2 subdomain: ~60 kb
Gap: ~12 kb

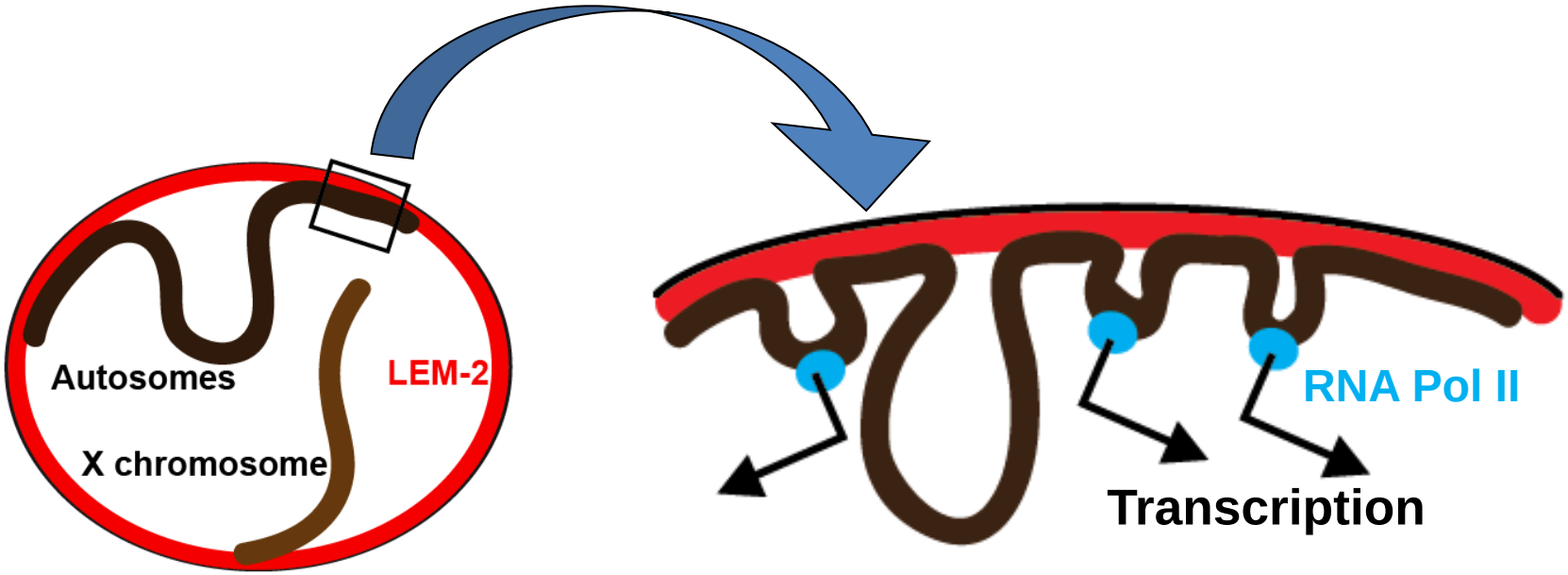
Domain-subdomain organization of nuclear membrane association



High RNA polymerase II, H3K4me3 and HTZ-1 (H2A.Z) levels at gaps between LEM-2 subdomains



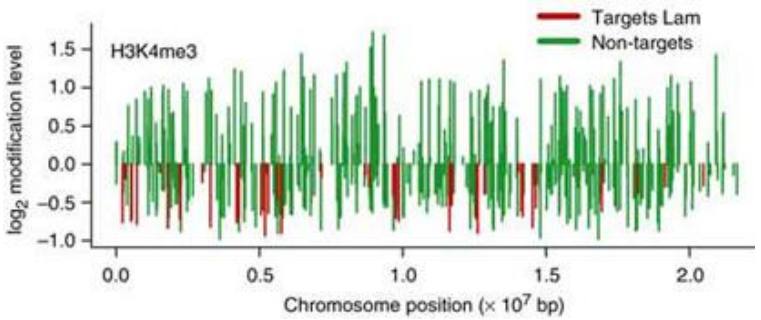
Small loops emerging from the nuclear membranes are transcriptionally active



- Small loops may be “wells” to concentrate factors, for higher residency times and component recycling
- Development of protocols for nuclear membrane genomics

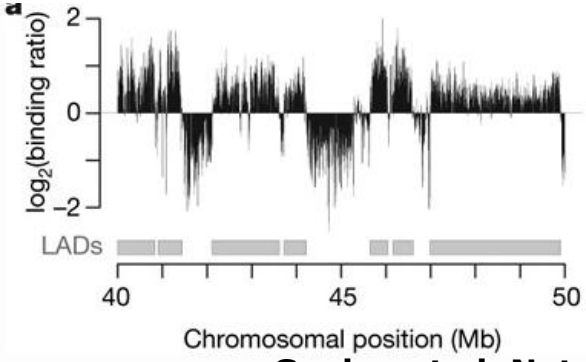
Chromosome interaction with the nuclear envelope

***Drosophila*:**
Repressive environment in Lamin B associated regions



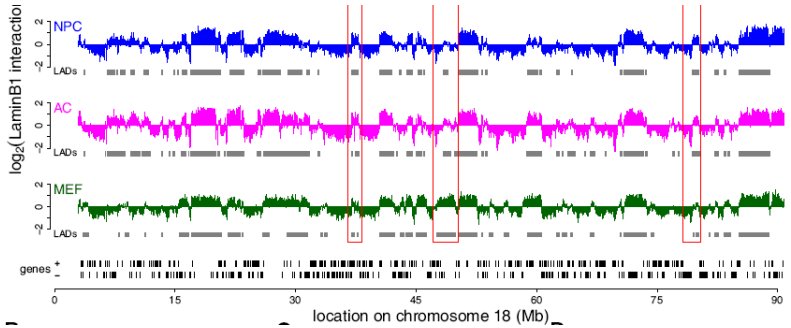
Pickersgill et al. Nat Genet 2006

Human:
Domain-wide (100 kb-10 Mb) Lamin B1 association



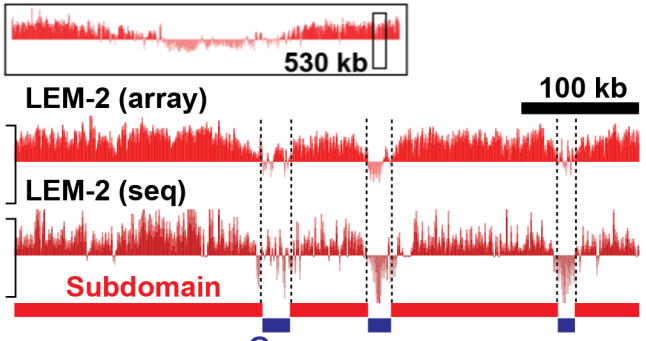
Guelen et al. Nature 2008

Mouse:
Cell-type specific Lamin B1 association



Peric-Hupkes et al., Mol Cell 2010

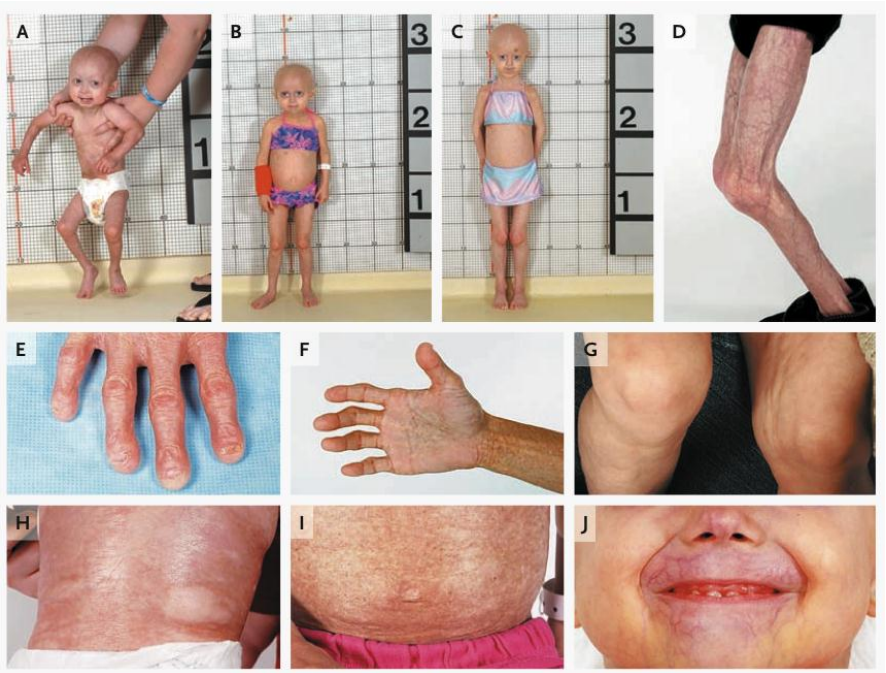
C. elegans
Domain-subdomain organization;
Active small loops; Local determinates



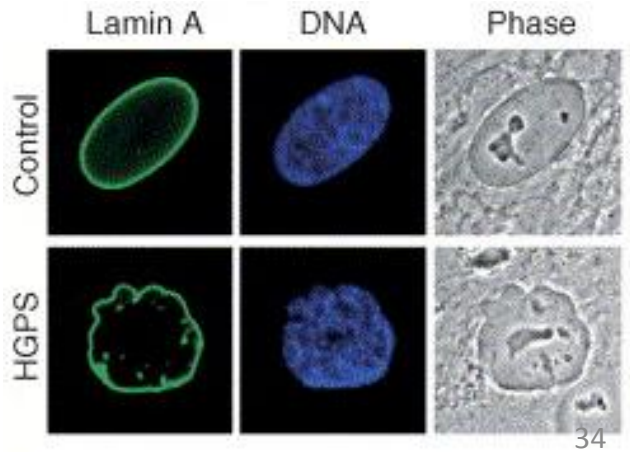
Ikegami et al. 2010

Hutchinson-Gilford Progeria Syndrome (HGPS)

- Phenotypes
  - Segmental premature aging
  - **Average lifespan of 13 years**
  - Affected tissues:
    - Bone (decreased bone mineral)
    - Fat (decreased body fat)
    - Cardiovascular system (elevated blood pressure)
    - Skin (dimpling, mottling (spots))
  - Unaffected tissues:
    - Immune system
    - Gastrointestinal system
  - **Major cause of death**
    - **Coronary atherosclerosis (heart attack)**
    - **Cerebrovascular arteries (stroke)**
- Rare disease
  - 1 in 4 million
  - 64 living patients identified world-wide (April, 2010)
- Cellular phenotypes
  - Gene expression changes
  - Aberrant nuclear membrane

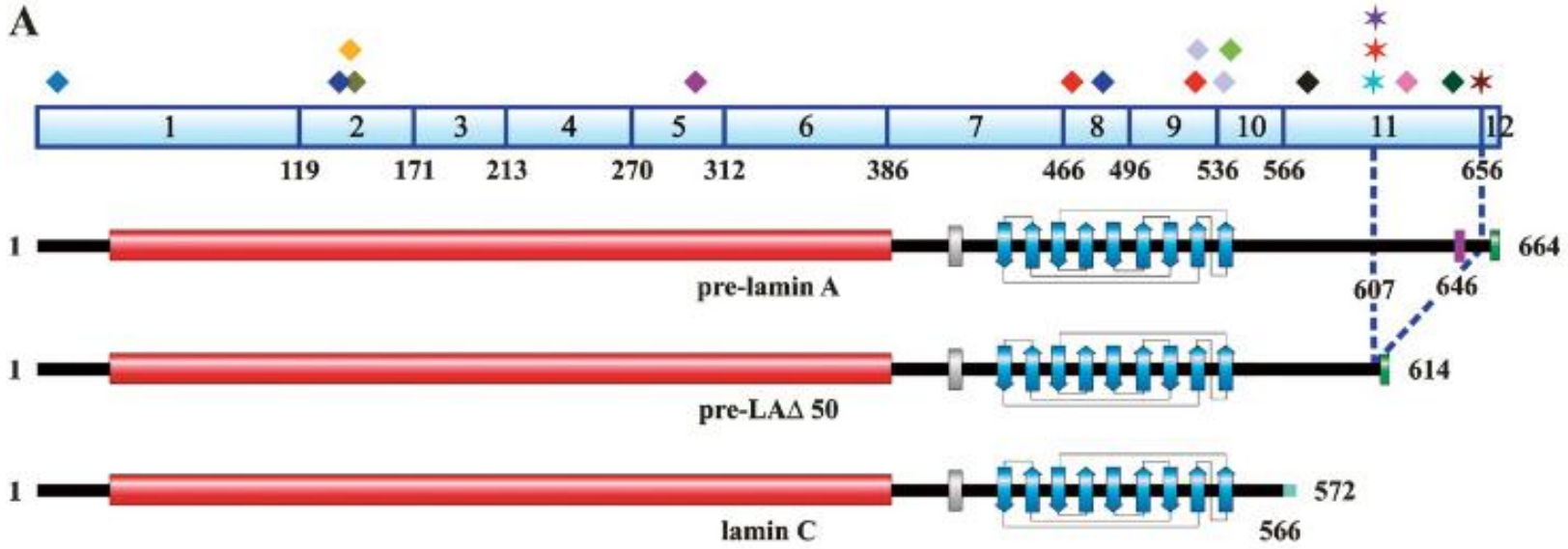


Merideth et al., N Engl J Med 2008



Mattout et al., Curr Opin Cell Biol 2006

HGPS is caused by mutations of the lamin A/C gene



Dechat et al., Gene Dev 2008

- Lamin A/C is one of two major lamin proteins
  - A-type: lamin A and C; B-type: lamin B1 and B2
- Mostly dominant *de novo* point mutations; C->T mutation (G608G) in exon11
  - Activation of a cryptic splicing site
  - Deletion of 50 amino acids => a mutant pre-lamin A, called “Progerin”
  - Progerin retains farnseylation

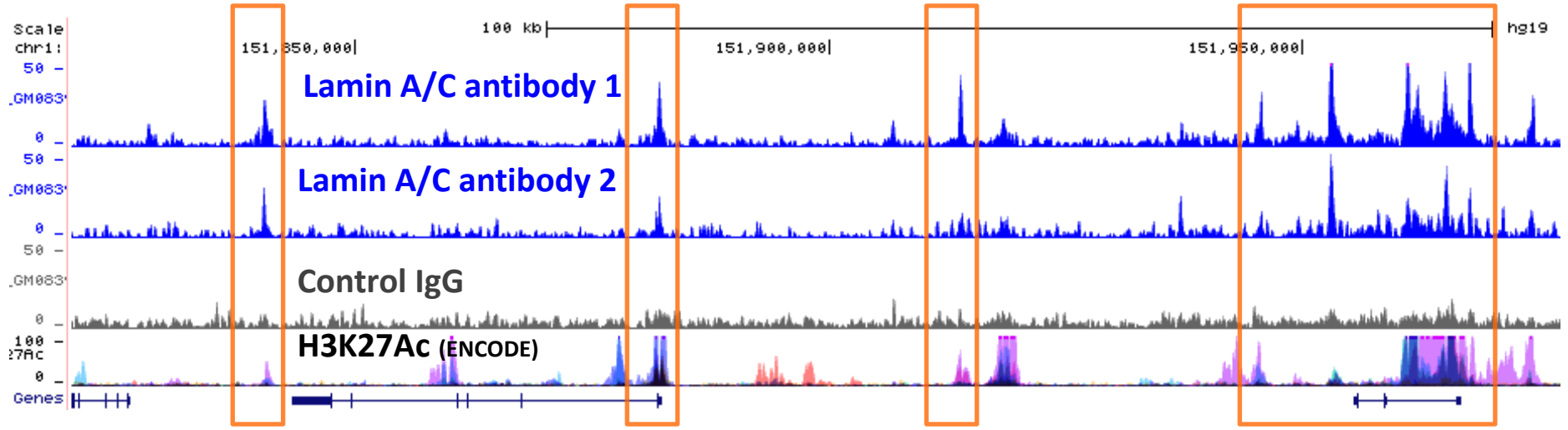
Many other LMNA mutations linked to diseases

Table 1
Diseases caused by mutations in genes encoding lamins and associated proteins

Disease	Mutation	Major disease phenotypes
Striated muscle diseases		
Autosomal dominant EDMD	*LMNA*	Muscle weakness and wasting in scapulohumeral-peroneal distribution; early joint contractures; dilated cardiomyopathy
Autosomal recessive EDMD	*LMNA*	Muscle weakness and wasting in scapulo-humeral peroneal distribution; early joint contractures; dilated cardiomyopathy
Cardiomyopathy dilated 1A	*LMNA*	Cardiomyopathy with minimal to no skeletal muscle involvement
Limb-girdle muscular dystrophy type 1B	*LMNA*	Muscle weakness and wasting in limb-girdle distribution; dilated cardiomyopathy
Congenital-type muscular dystrophy	*LMNA*	Severe relatively diffuse myopathy presenting in first year of life; later cardiomyopathy
"Heart-hand" syndrome (with limb defects)	*LMNA*	Brachydactyly with mild hand and more severe foot involvement; cardiomyopathy
X-linked EDMD	*EMD*	Muscle weakness and wasting in scapulo-humeral peroneal distribution; early joint contractures; and dilated cardiomyopathy
Partial lipodystrophy syndromes		
FPLD2	*LMNA*	Loss of subcutaneous fat from the extremities at puberty, followed by increased fat accumulation in the face and neck; insulin resistance; diabetes mellitus; hyptertriglyceridemia; hepatic steatosis
Lipoatrophy with diabetes, hepatic steatosis, hypertrophic cardiomyopathy, and leukomelanodermic papules	*LMNA*	Generalized fat loss; insulin-resistant diabetes, hypertriglyceridemia, hepatic steatosis, hypertrophic cardiomyopathy; disseminated whitish papules
Mandibuloacral dysplasia (also has features of progeria)	*LMNA*	Hypoplastic mandible with dental crowding, acroosteolysis, stiff joints, atrophy of the skin over hands and feet, hypoplastic clavicles; "Andy Gump" appearance; persistently wide cranial sutures and multiple wormian bones; alopecia and short stature; and partial lipodystrophy
Acquired partial lipodystrophy (Barraquer-Simons syndrome)	*LMNB2*	Progressive, sporadic lipodystrophy with phenotype similar to FPLD2 (above)

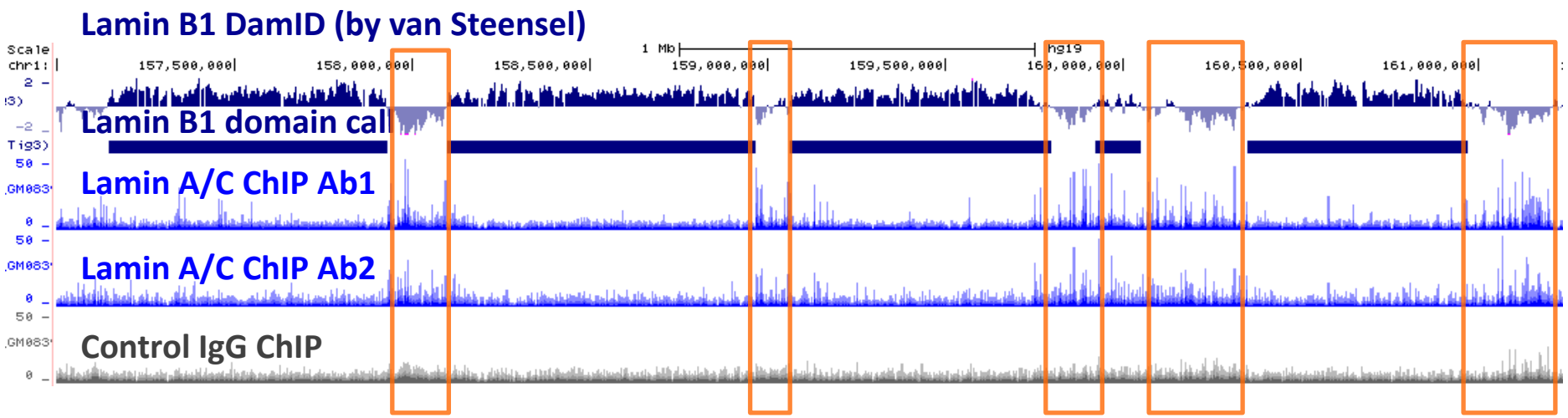
Worman et al., J. Clin. Invest 2009

Lamin A/C associates with active regulatory sites



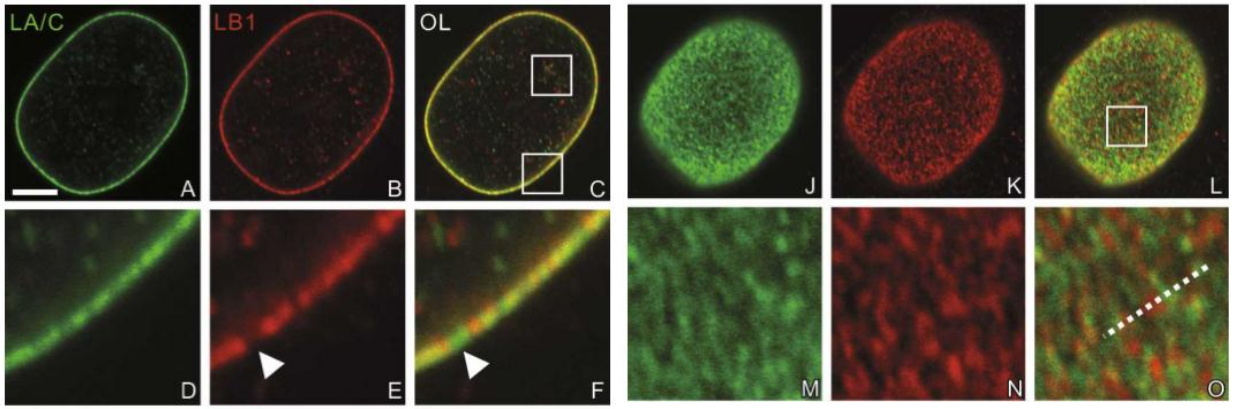
ChIP in primary human fibroblast cells

Lamin A/C distribution is inversely correlated with lamin B

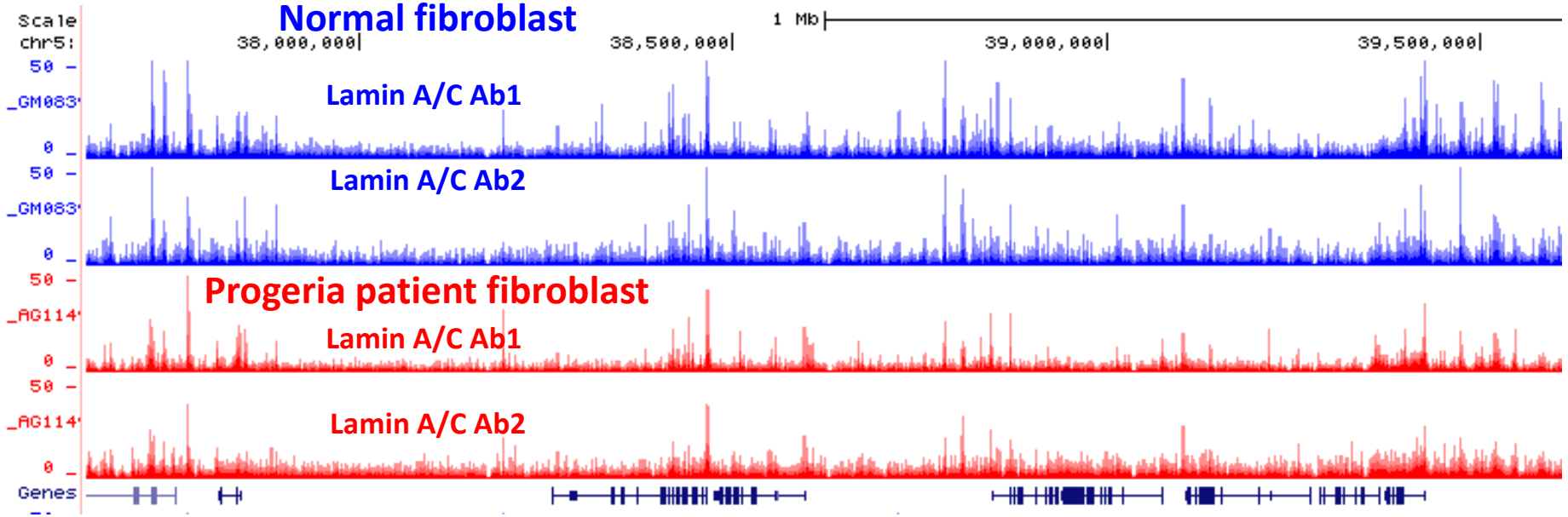


Both ChIP and DamID (Guelen et al., Nature 2008) are performed in primary human fibroblast cells

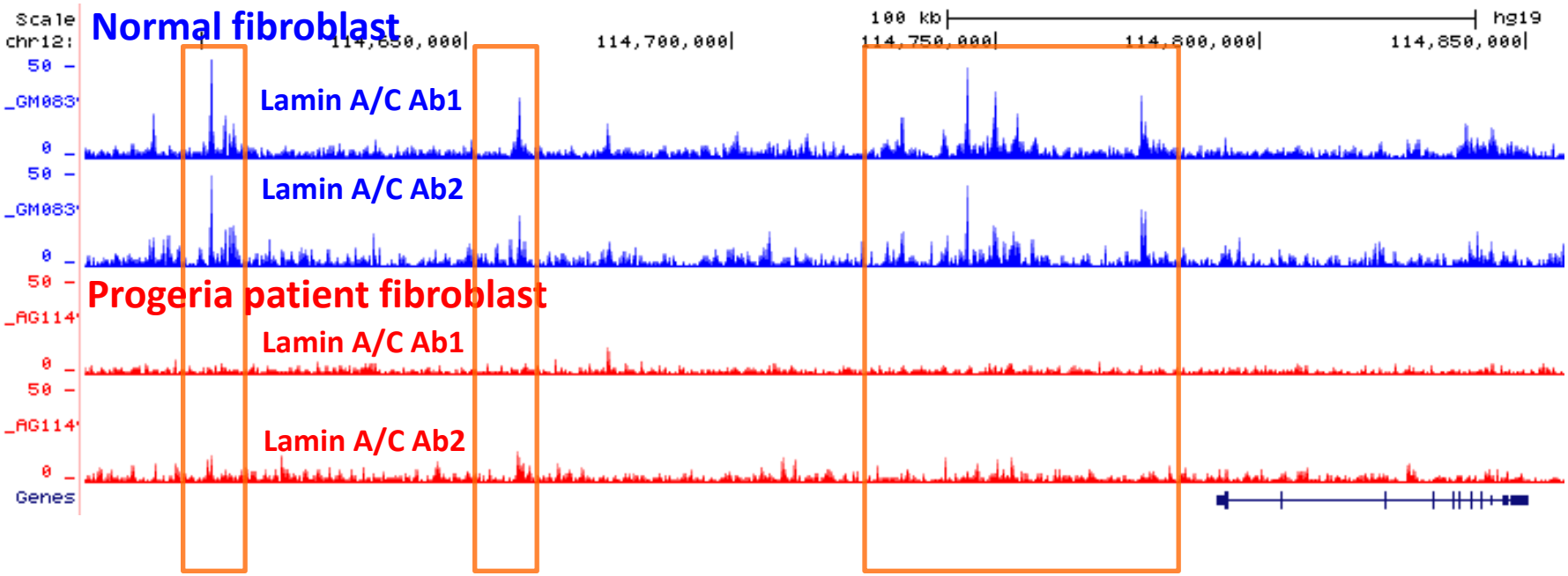
Lamin A/C and lamin B1 are separately localized in the lamina
(Shimi et al., Genes Dev 2008)



The pattern of lamin A/C association is largely maintained in progeria cells



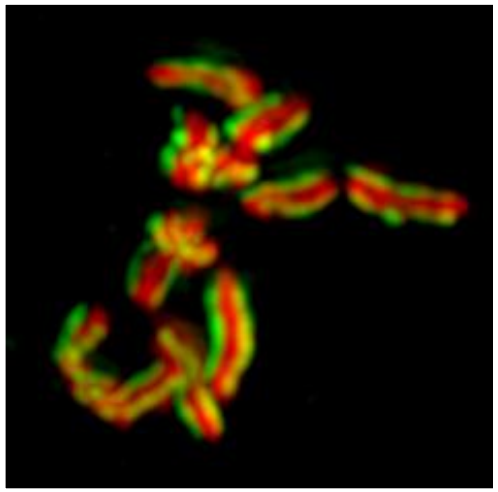
Some regions lose lamin A/C associations in progeria cells



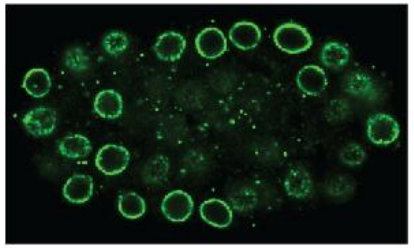
Currently ChIPing from cells treated with Farnesyltransferase Inhibitor (FTI)

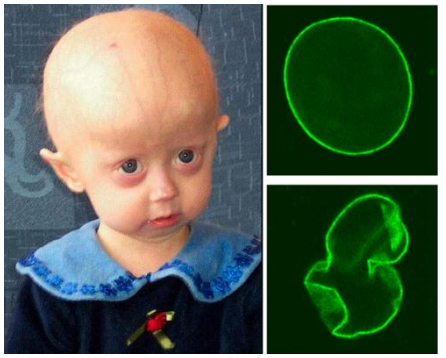
Unique insights gained by modENCODE

(1) Centromere specification



(2) Chromosome-membrane interactions





…and many other discoveries (exon marking by H3K36me3, dosage compensation insights in worms and flies, new functions for H4K20me1, principles of metazoan nucleosome organization…)

Acknowledgements

CENP-A

Reto Gassmann, Andreas Rechtsteiner, Karen W. Yuen, Andrew Muroyama, Francie Barron, Paul Maddox, Joost Monen, Thea Egelhofer, Sevinc Ercan, Karen Oegema, Jason Lieb, Susan Strome, **Arshad Desai**

An Inverse Relationship to Germline Transcription Defines the *C. elegans* Holocentromere in Progeny

Nature. 2012. Apr 8; 484 (7395):4534-7 PMID: 22495302

LEM-2 and Progerin

The Progeria Research Foundation

Kohta Ikegami, Thea Egelhofer, Susan Strome, Jason Lieb

C. elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2.

Genome Biology. 2010;11(12):R120. PMID: 21176223

Worm Chromatin Group Labs

Julie Ahringer
Abby Dernburg
Arshad Desai
Jason Lieb
Susan Strome
X. Shirley Liu
Eran Segal

Fly Chromatin Group Labs

Sally Elgin
Gary Karpen
Mitzi Kuroda
Vince Pirrotta
Peter Park
Dave MacAlpine