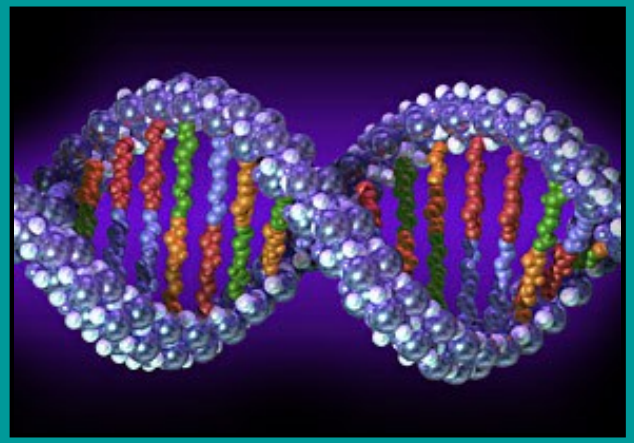
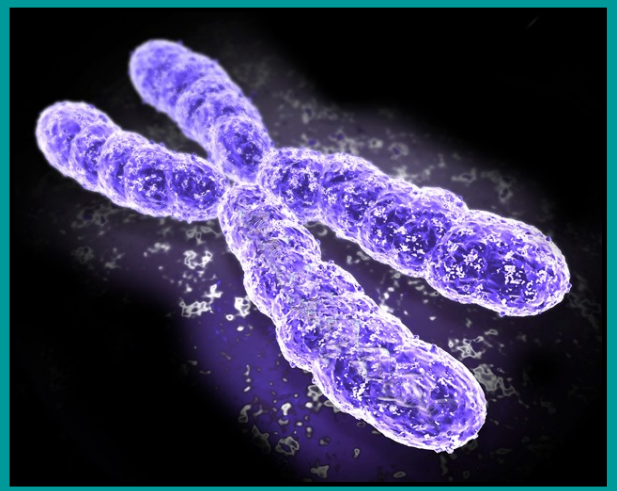
Genome-wide Mapping of Chromosomal Proteins in Drosophila

Genome Sequence & Gene Expression



QuickTime™ and a YUV420 codec decompressor are needed to see this picture.



Chromosome Inheritance & Genome Stability

Chromatin states & Nuclear Organization

The Players

Antibodies, Chromatin Preps, ChIP-Mapping, Validation

- **Gary Karpen - LBNL and UC Berkeley**
- **Sarah Elgin - Washington University**
- **Mitzi Kuroda - Harvard Medical School**
- **Vince Pirrotta - Rutgers University**

Bioinformatics / Computational Analysis

- **Peter Park - Harvard Medical School**
- **Peter Kharchenko**

Which Proteins ?

~100 total

Histone modifications and variants (~50)

'heterochromatin', silenced genes

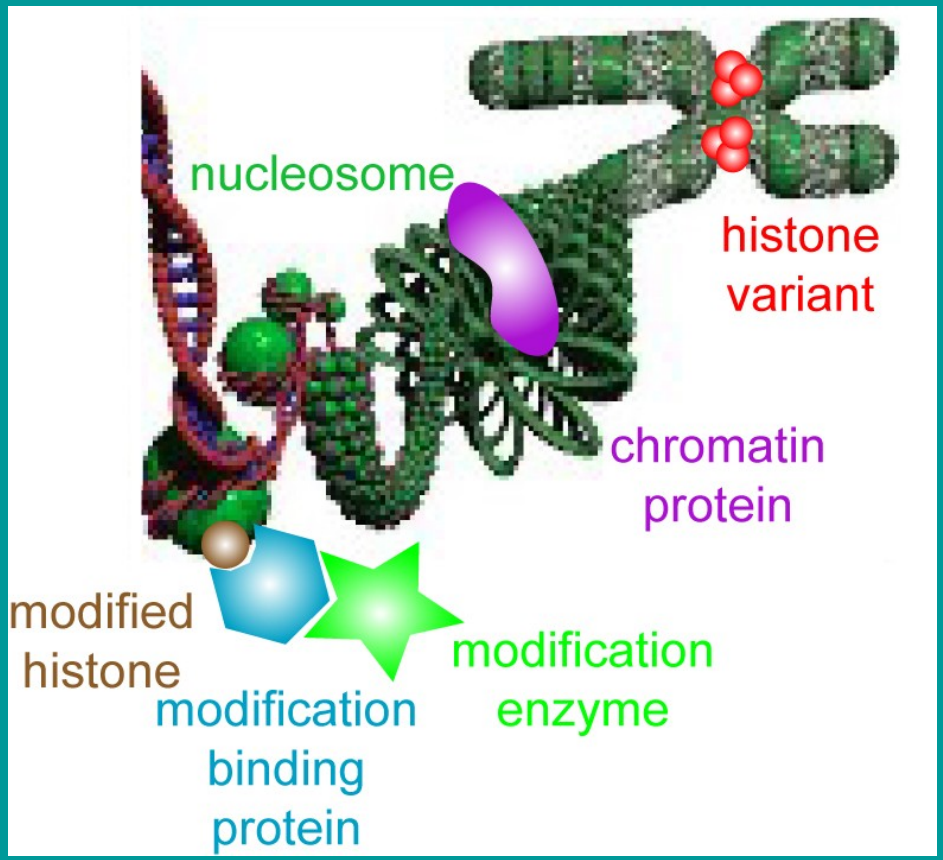
- H3K9me,me2,me3
- H3K27me,me2,me3
- H2Av

'euchromatin', active genes

- H3K4me,me2,me3
- H3 and H4 K acetylations (limited)
- H2Av

Histone modifying enzymes (~25)

- Su(var)3-9, G9a, dSET DB1 (H3K9me)
- TRX, TRR, ASH1 (H3K4me)
- E(z) (H3K27me)
- LSD1 demethylase (H3K4me)
- JMJD2A, UTX demethylases (H3K9me, H3K27me)



Which Proteins ?

Chromosomal proteins (~25)

Histone modification binding

- Pc (H3K27me)
- HP1 (H3K9me)
- H2Av

Chromatin remodeling

- SWI/SNF
- NURF
- ACF

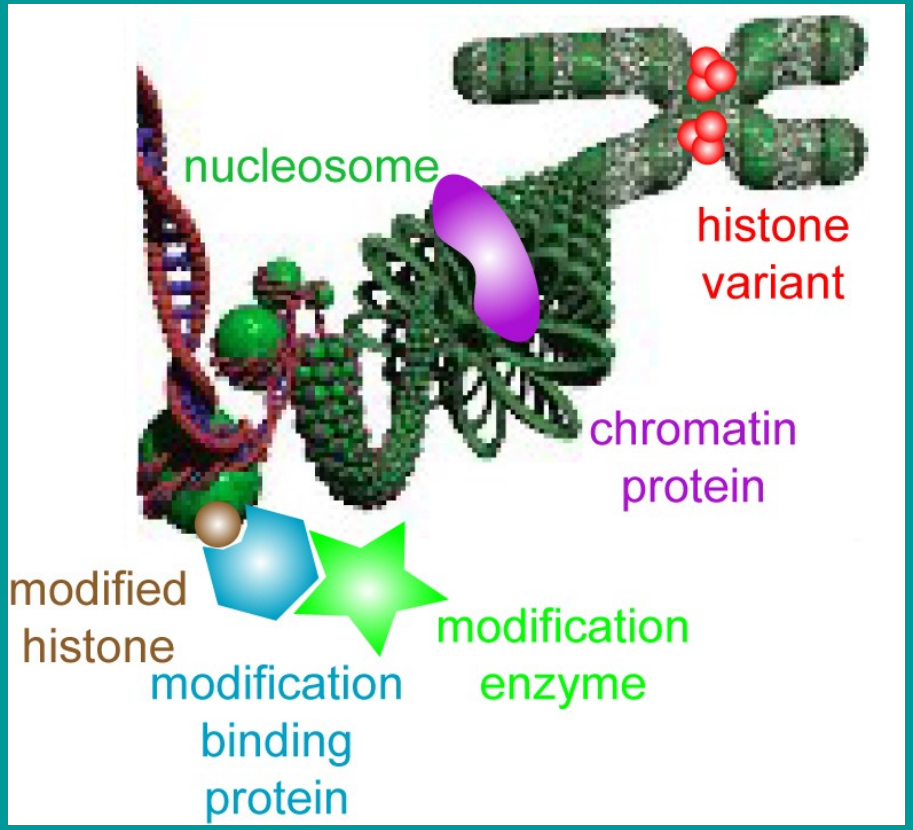
Insulators

- Su(Hw)
- Mod(mdg)4
- CTCF

Nuclear lamina and pores

- Lamin C
- Lamin B receptor
- Nup 153

3D architecture



'Epigenetic Sets'

H3K9me, me2, me3: modifications

Su(var)3-9, G9a, dSET DB1: methylases

JMJD2A: demethylase

HP1: binds H3K9me2 and 3, Su(var)3-9

More details

Poster # 767B

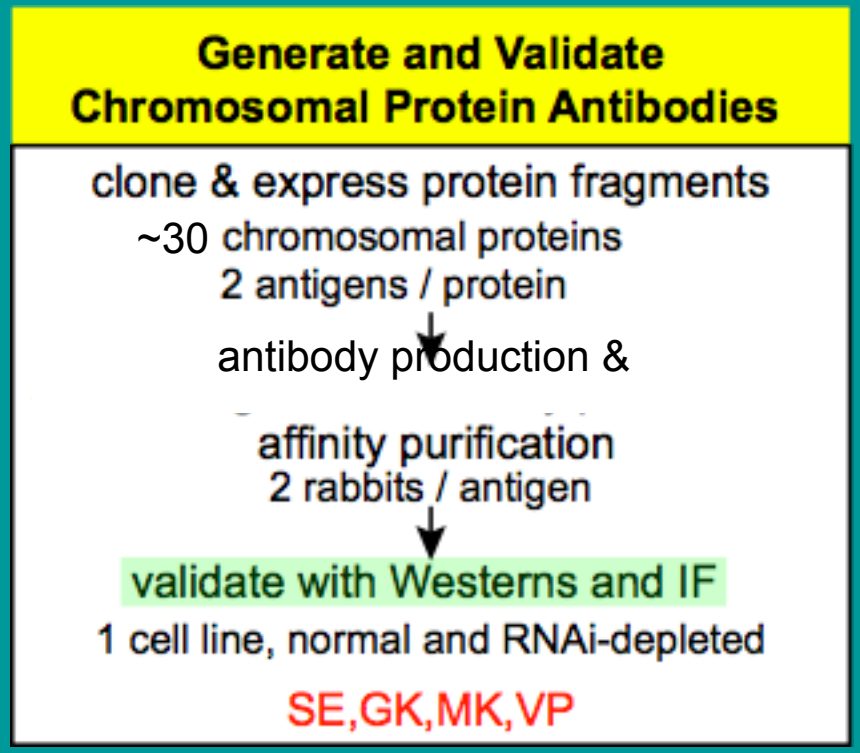
General Approach: ChIP

Antibodies

purchase histone modification/ variant-specific antibodies

Validate these (mass spectrometry)

obtain existing antibodies for other chromosomal proteins (~20)



make antibodies for other ~30 chromosomal proteins

two Abs: against two protein fragments per protein

Donate good antibodies

Chromatin Preparation

3 cell lines (MK, VP)

- S2 (embryonic)
- Cl.8+ (larval wing disc)
- MLDmBG3-c2 (neuronal)

2 embryo stages (SE)

- 2-4 hrs (blastoderm, relatively uniform)
- 14-16 hrs (highly determined/differentiated)

Target: map 100 proteins in these 5 sources

150 proteins/modifications on lists

initially will just use 2 cell lines (S2 & Cl.8)
to identify which proteins show interesting patterns,
which redundant/less interesting (e.g. histone acetylations)

Differentiayed tissues

1) salivary glands-direct correlation with protein mapping on polytenes

2) third instar wing discs- undifferentiated and diploid, show specific patterns of gene silencing and activation, and of interest for comparisons with Cl.8+ cells

3) adult heads-terminally-differentiated diploid cells

4) germ-line cells- likely to be the most undifferentiated cells in the organism, could identify a baseline state for chromatin

Data Collection

Array Hybridization and Data Collection

hybridization
1 array / ChIP replicate
(2 arrays per protein)

↓

scanning

↓

deposit data on FTP site

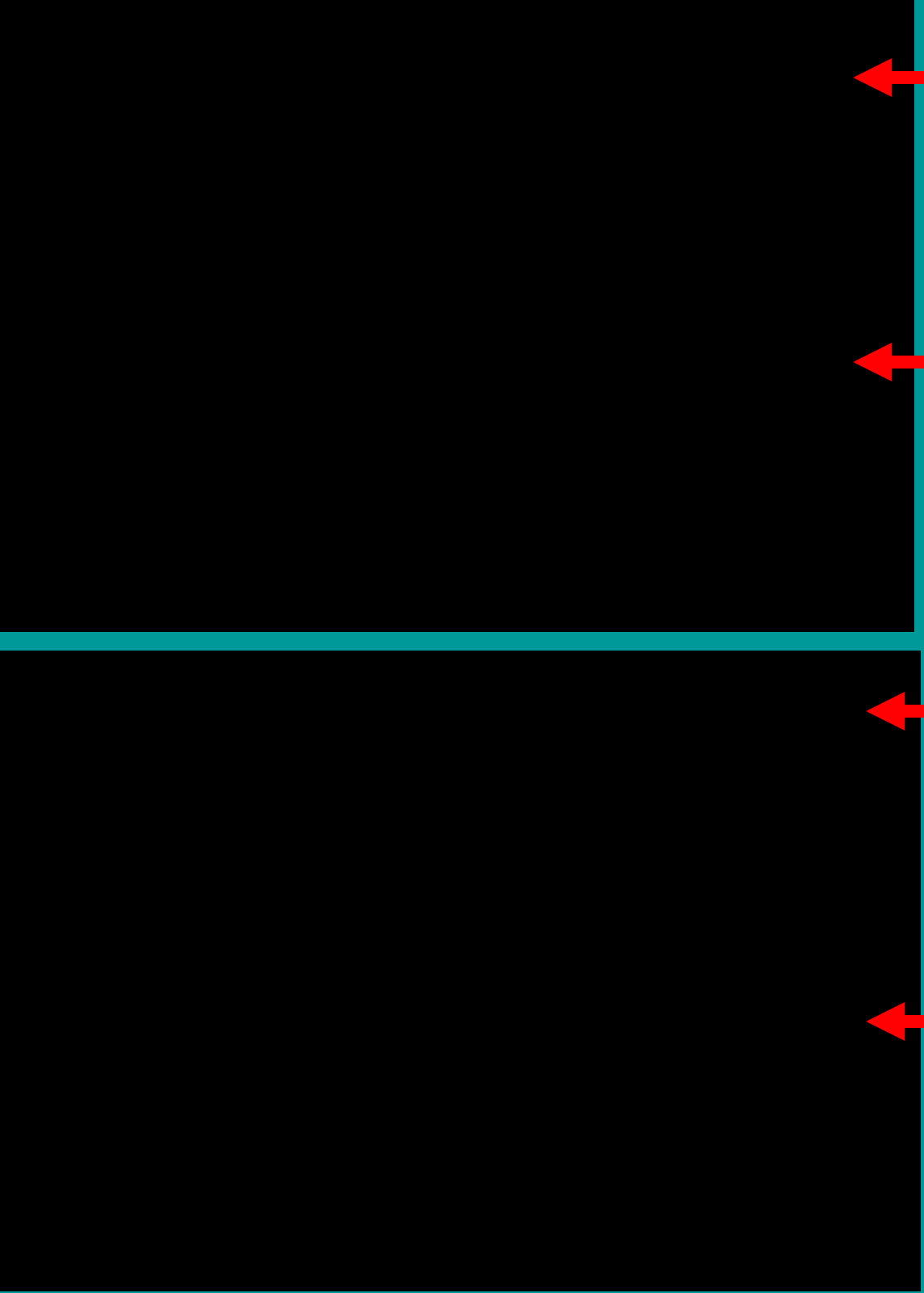
VP

Affymetrix Genome Tiling Arrays

Direct sequencing
Solexa ?? ABI??

better coverage (e.g. satellites)
more quantitative?
cheaper?
availability?

Rutgers Facility



ChIP/seq

may offer very significant advantages in resolution and dynamic range

Kuroda & Park

Data Analysis

statistical analysis of data
quality control
normalization

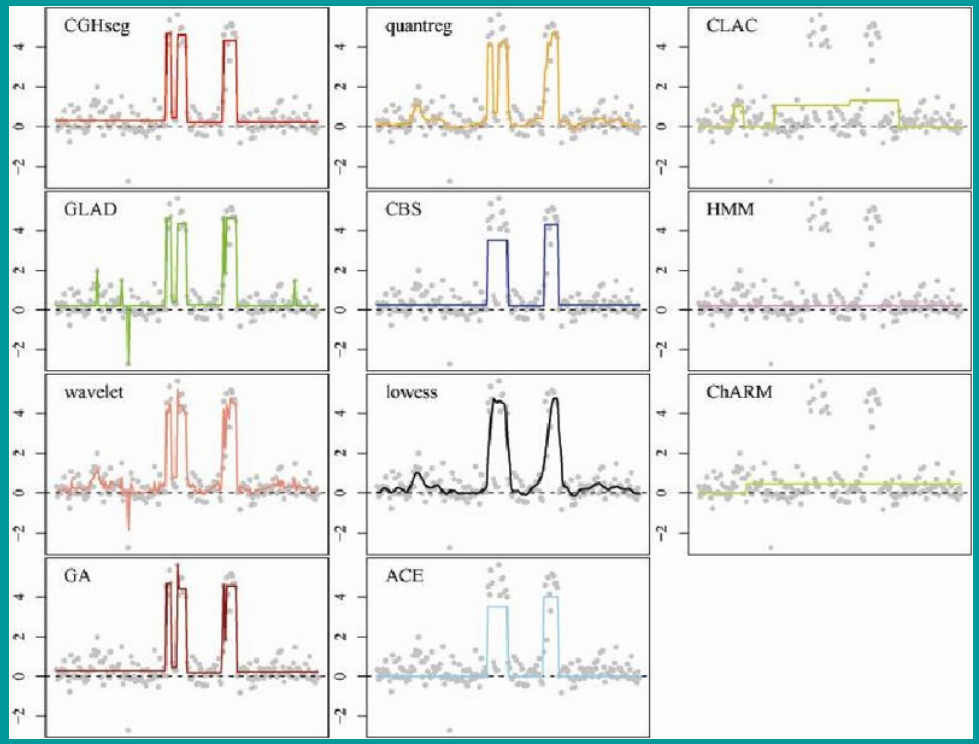
↓

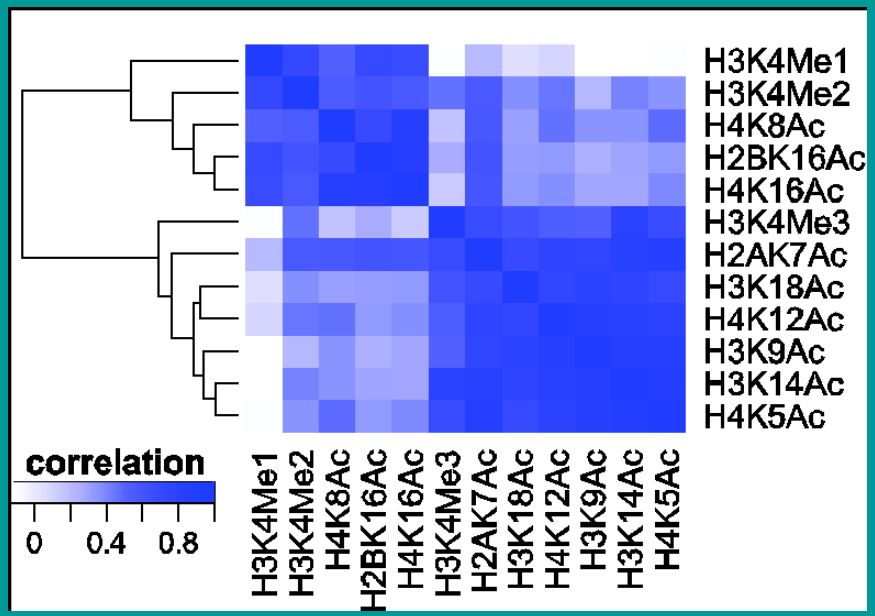
identify peaks and broad domains

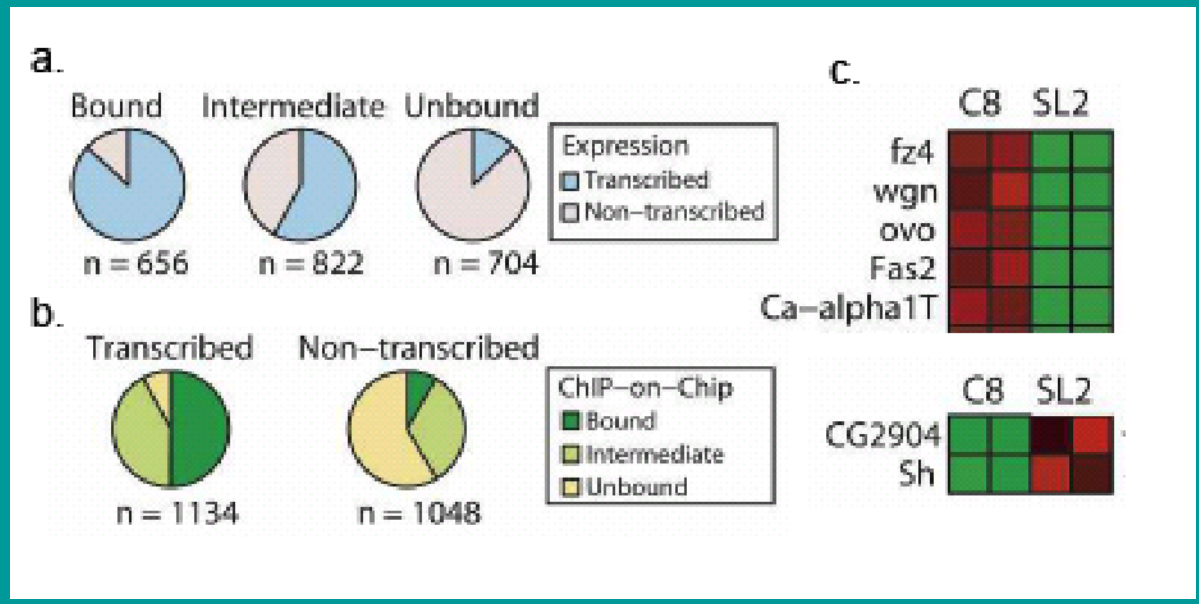
↓

bioinformatic comparisons
compare tissues
combinatorial patterns
compare to transcriptome data
cluster analysis

PP







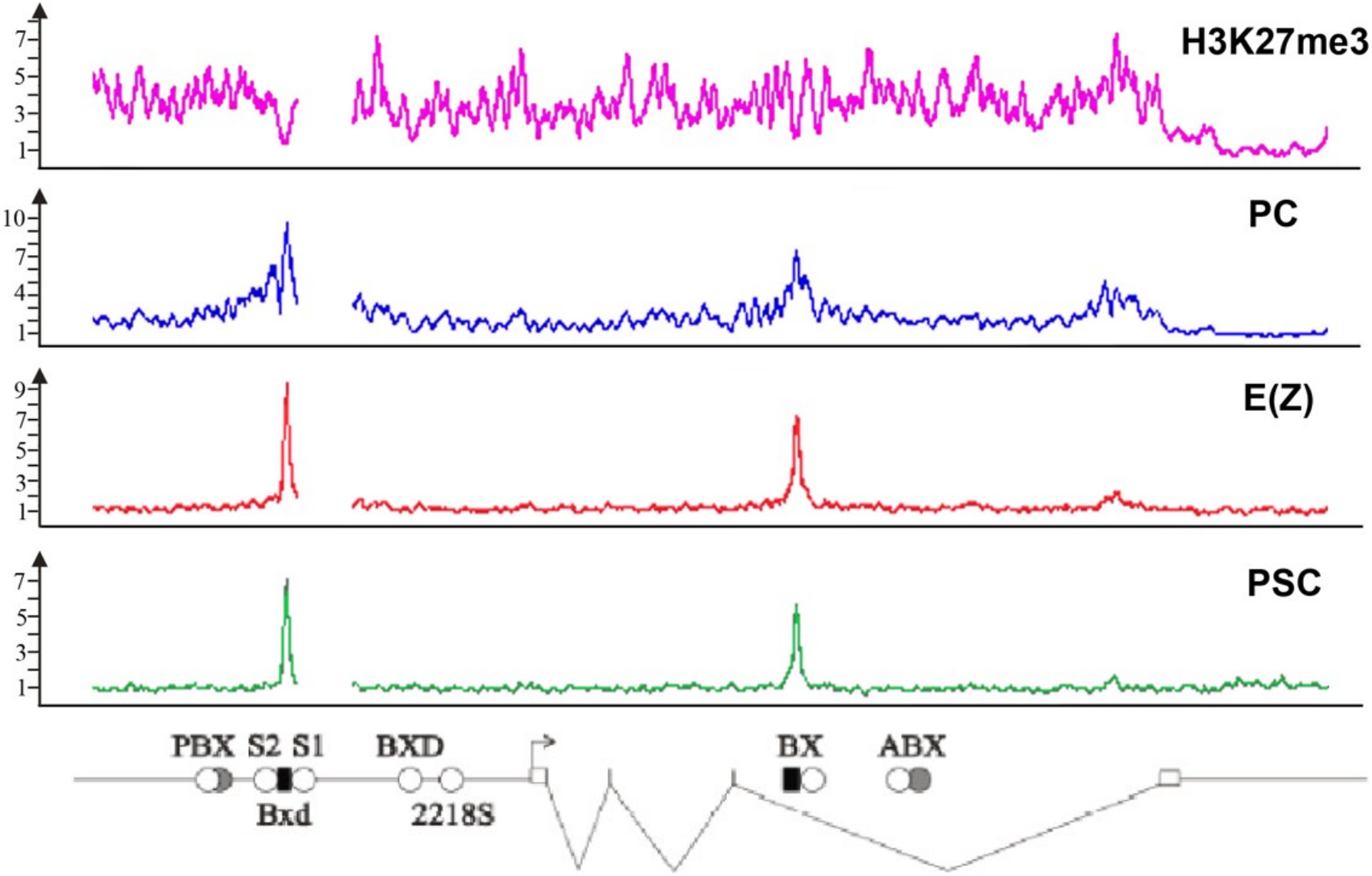
Data Submission

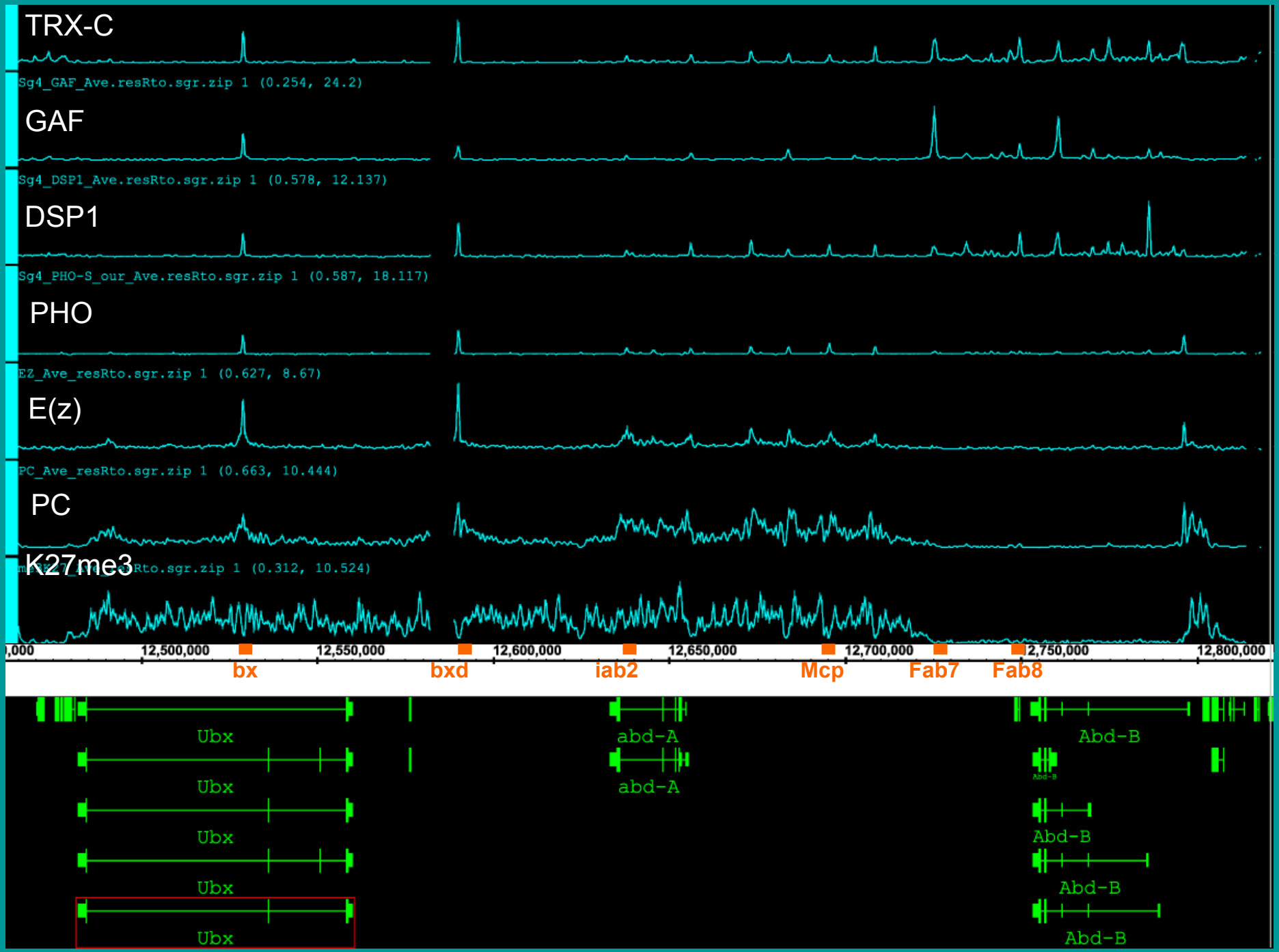
submit all data to DCC once verified, update after validation

all antibodies will be made available to the ‘public’

PcG proteins and histone modifications

at the *Ubx* gene



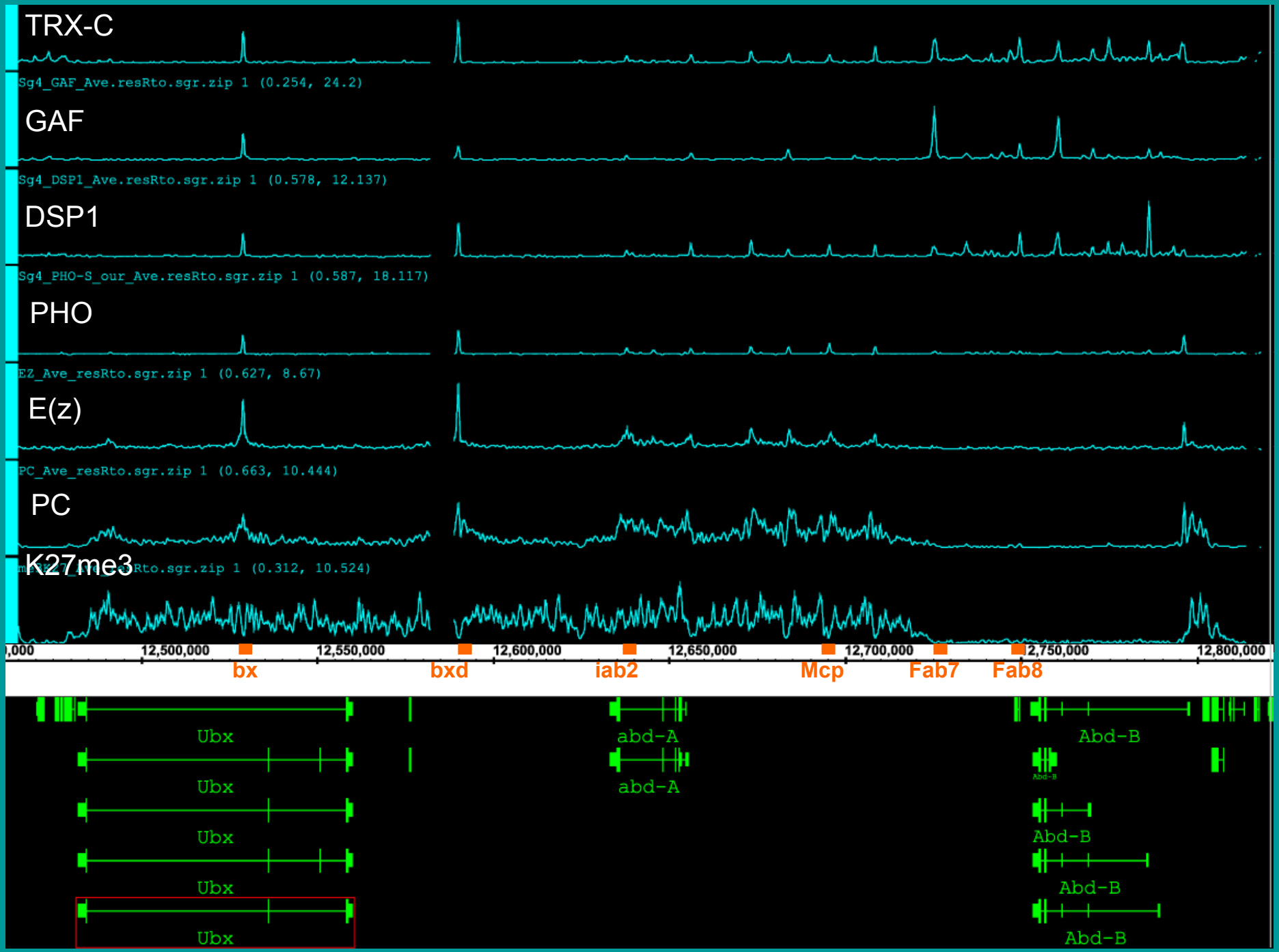


Yuri Schwartz & Tanya Kahn

Binding sites

for DNA binding proteins with known connection to PcG function

protein	"traditional" binding consensus	Microarray-based binding consensus
GAF	GAGAG	pair of GAGAG spaced from 1 to 150bp
PHO	GCCAT	GCNGCCATttt
DSP1	GAAAA	GTTGTT, GTGCGT
PSQ	GAGAG	gtgGTGACCG, CGATA

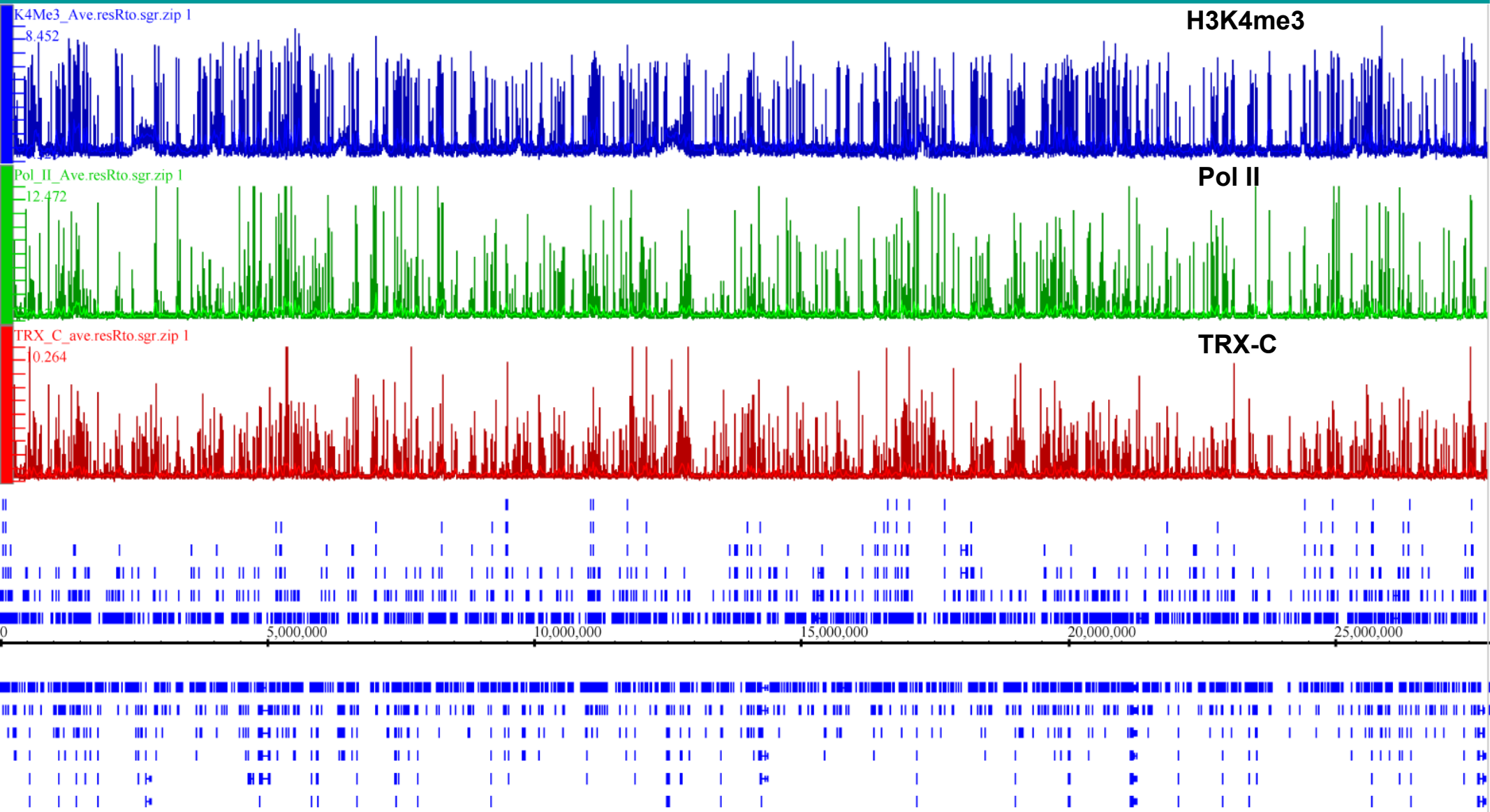


Yuri Schwartz & Tanya Kahn

91%
computationally defined PREs in Sg4
show prominent binding of TRX-C

TRX-C binds to thousands of active genes

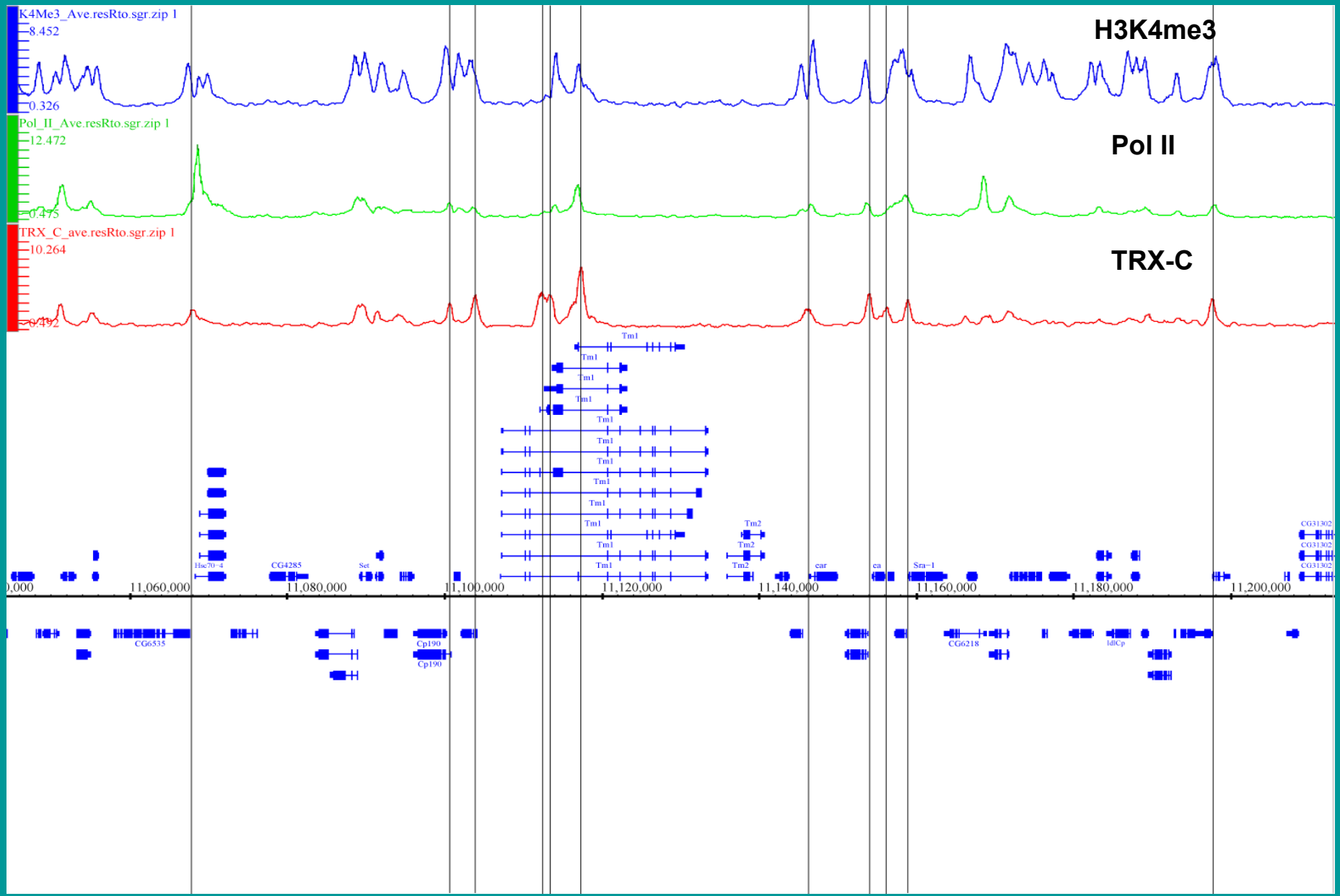
(3R chromosome in Sg4 cells)



Yuri Schwartz & Tanya Kahn

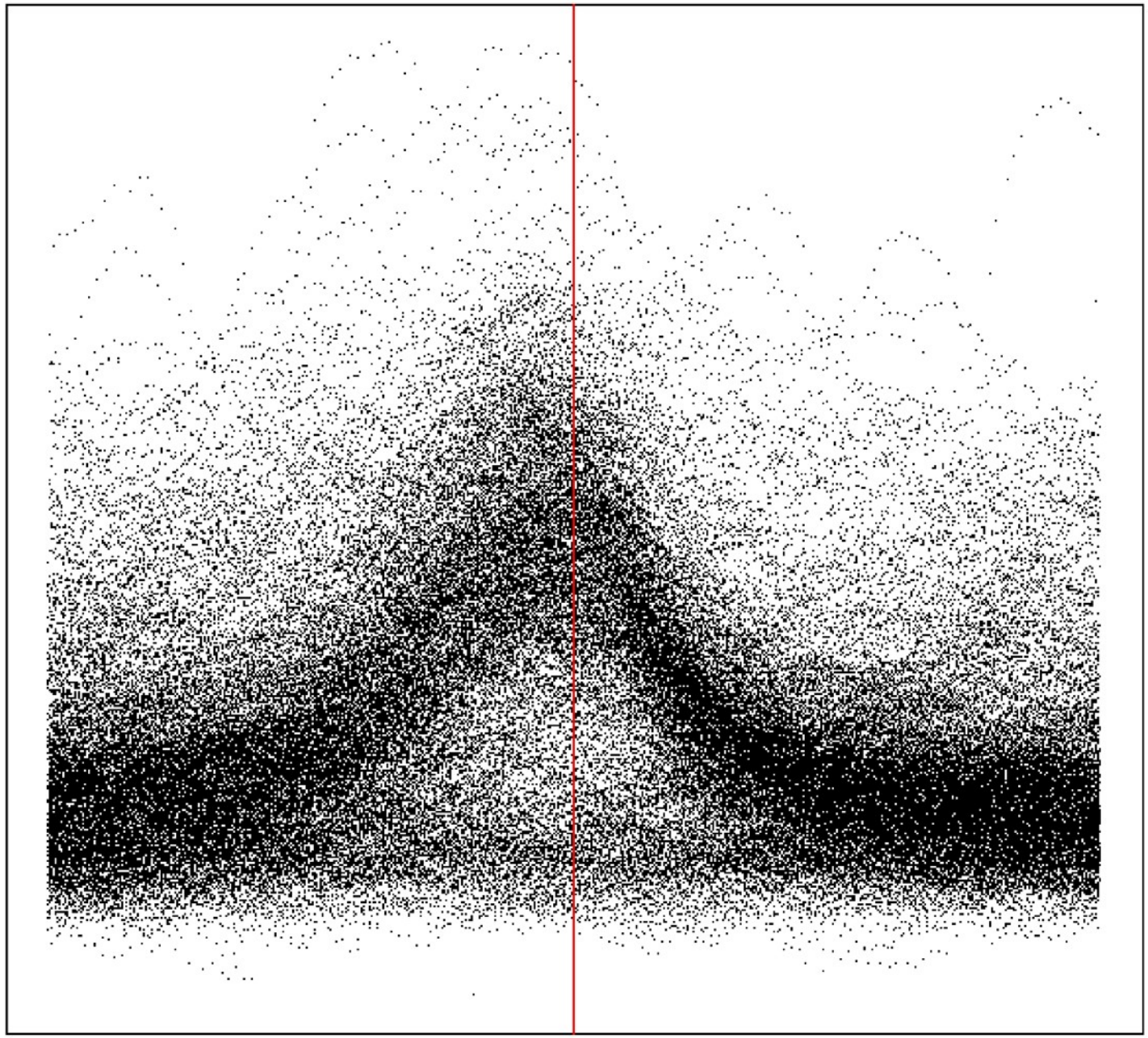
TRX-C binds to thousands of active genes

(fragment of 3R chromosome)



Yuri Schwartz & Tanya Kahn

TRX at TSS

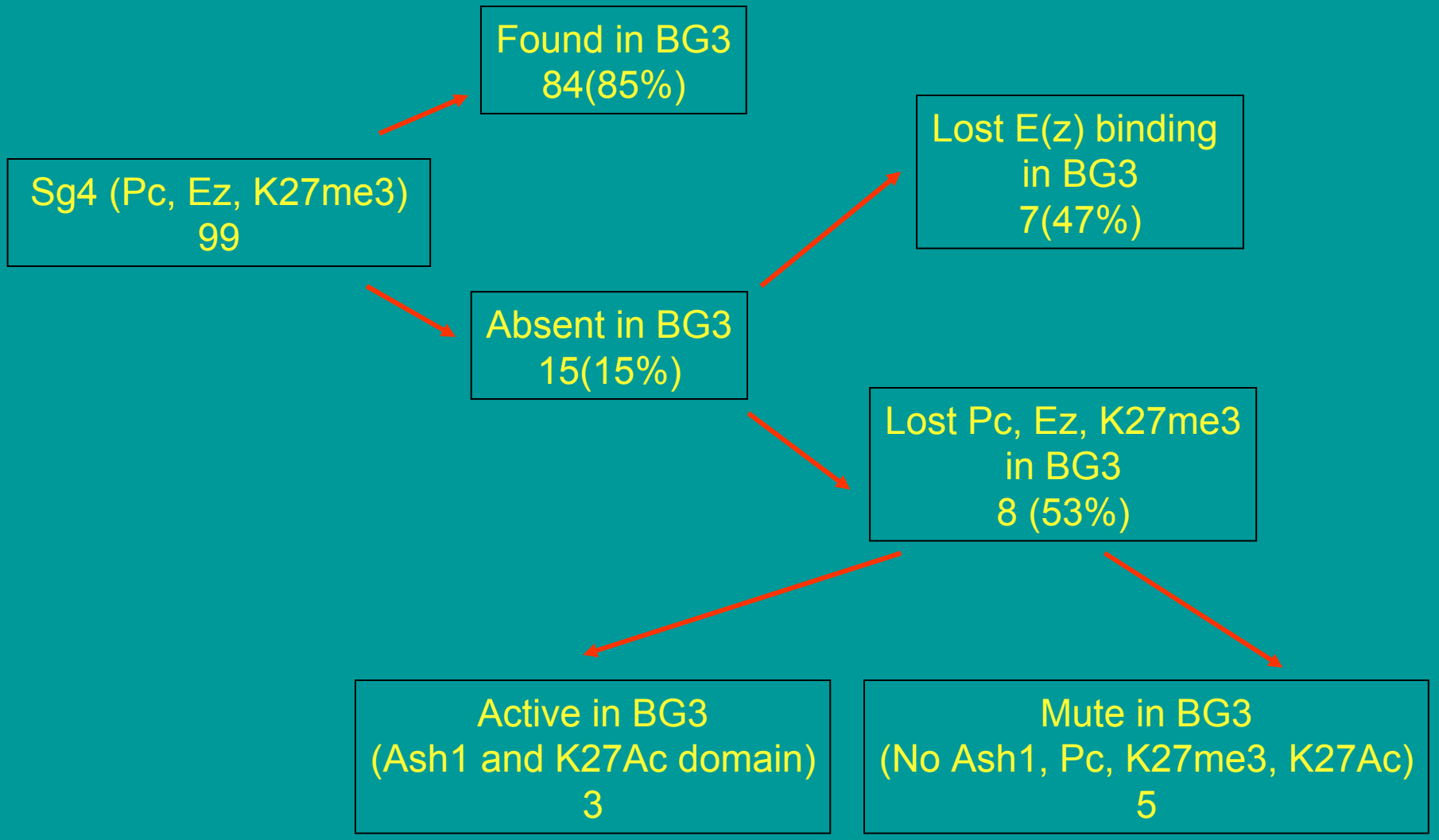


Alternative chromatin states of PcG target genes

Cell lines: Sg4 – embryonic origin
BG3c2 – neural origin
D23 – wing imaginal disc

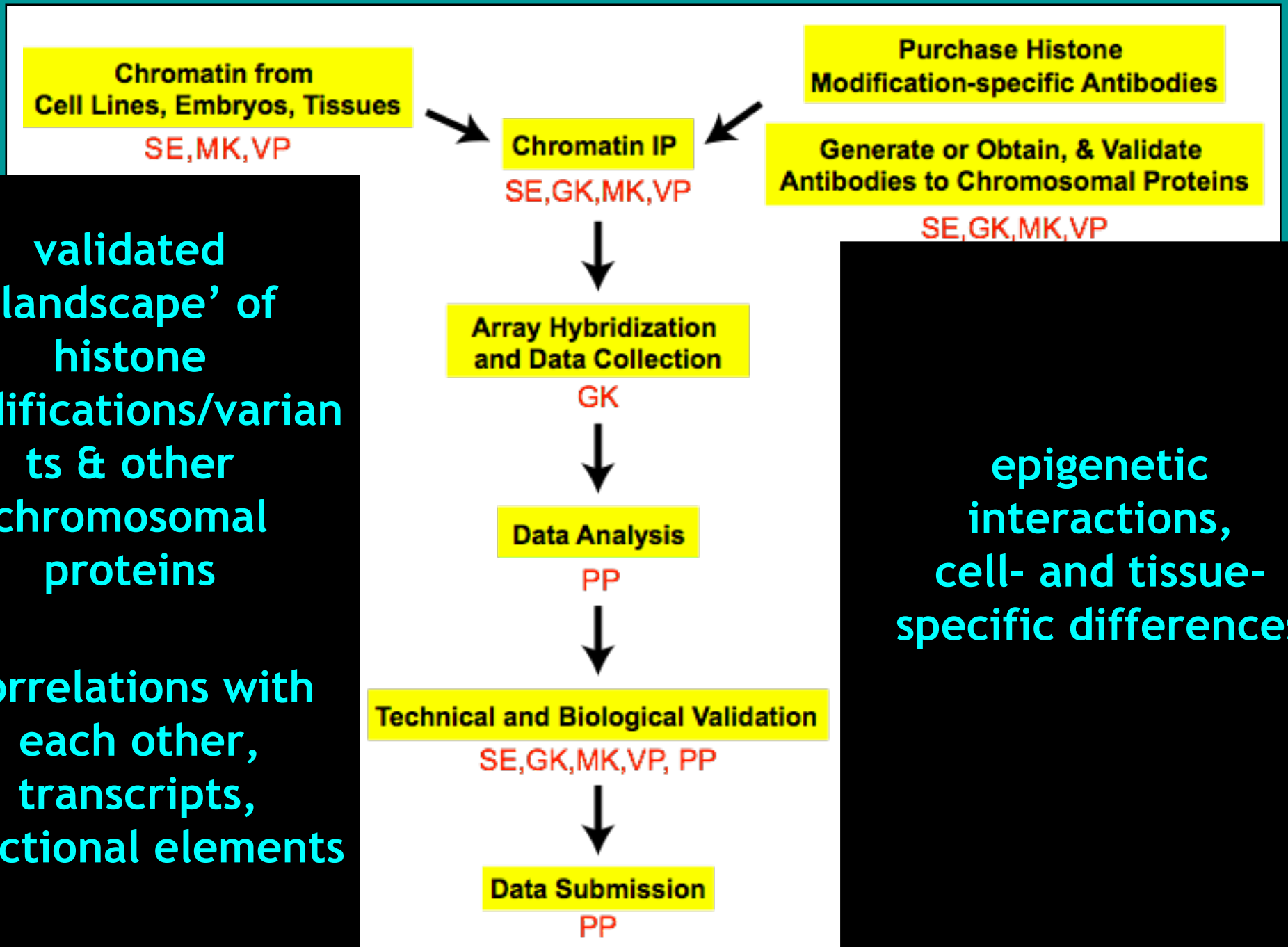
Proteins: Pc, Ez, (Psc for Sg4 only), H3K27me3
Trx-C, Trx-N, Ash1, PolII,
H3K4me3, H3K27Ac

Overlapping between semicomplete PcG sites



Yuri Schwartz & Tanya Kahn

General Plan and Expectations



validated 'landscape' of histone modifications/variants & other chromosomal proteins

correlations with each other, transcripts, functional elements

epigenetic interactions, cell- and tissue-specific differences

General Approach: ChIP-array, ChIP-chip