Statistical reconstruction of yeast nuclear organisation

Tarn Duong

G5 Computational Imaging and Modelling (Christophe Zimmer)

Cell biology and Infection Department day
16 October 2008

Introduction

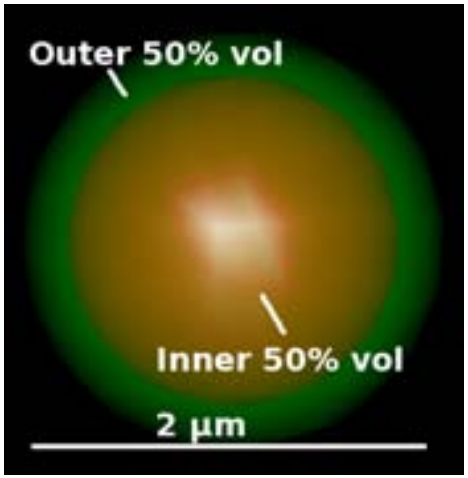
- Nuclear organisation = spatial organisation of genome inside nucleus
- Important for nuclear function: transcription, DNA repair and replication

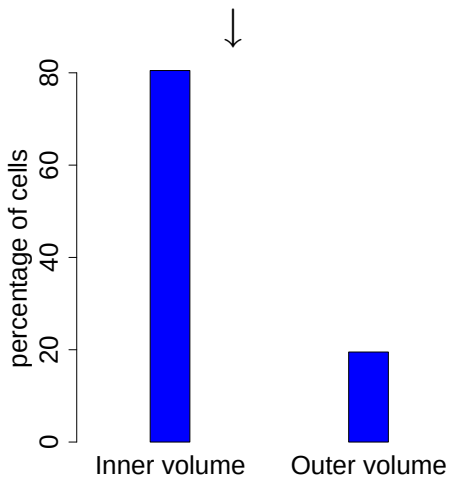
Introduction

- Nuclear organisation = spatial organisation of genome inside nucleus
- Important for nuclear function: transcription, DNA repair and replication
- PTR 218 ‘Functional analysis of gene location and dynamics through quantitative imaging’
  - Labs of U. Nehrbass, B. Dujon and C. Zimmer
- In *Saccharomyces cerevisiae* yeast, e.g.
  - GAL1 gene moves to nuclear periphery during transcription (Cabal et al., *Nature*, 2006)
  - Genes near telomeres (chromosomal extremities) at the nuclear periphery
    - tend to be silenced (Hediger et al., *Current Biol.*, 2002)
    - have highest DNA repair efficiency (Thérizols et al., *JCB*, 2005)
- but detailed nuclear organisation in eukaryotic cells (including yeast) is largely unknown

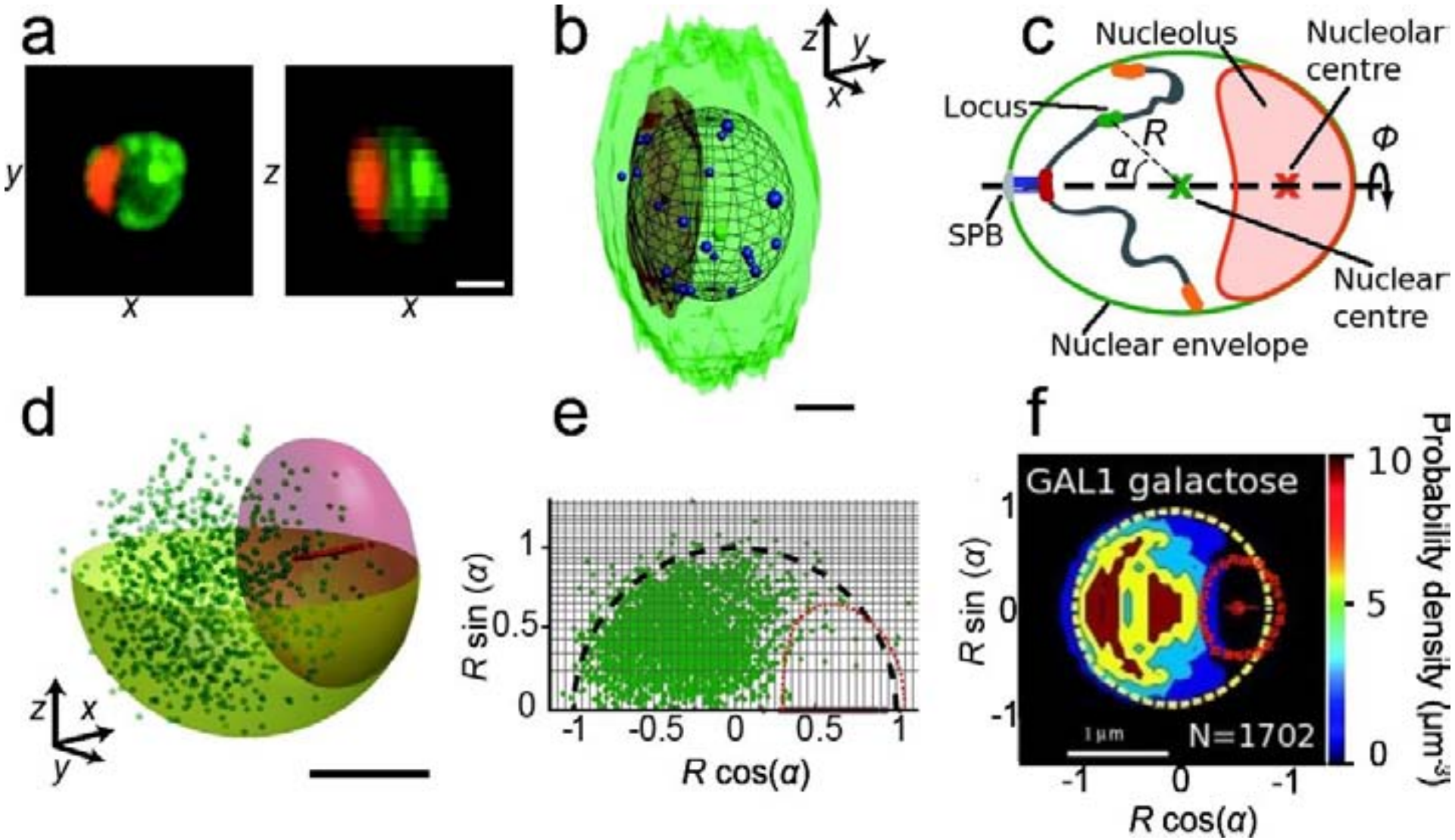
Current state of statistical description of nuclear organisation

- Chromatin has random components of motion (Heun et al., *Science*, 2001)
  - statistical descriptions required
- binary classification of distance to nuclear periphery
- low resolution
- diffraction limit for optical microscopes: $\sim 0.25\mu$m laterally, $\sim 0.5\mu$m axially



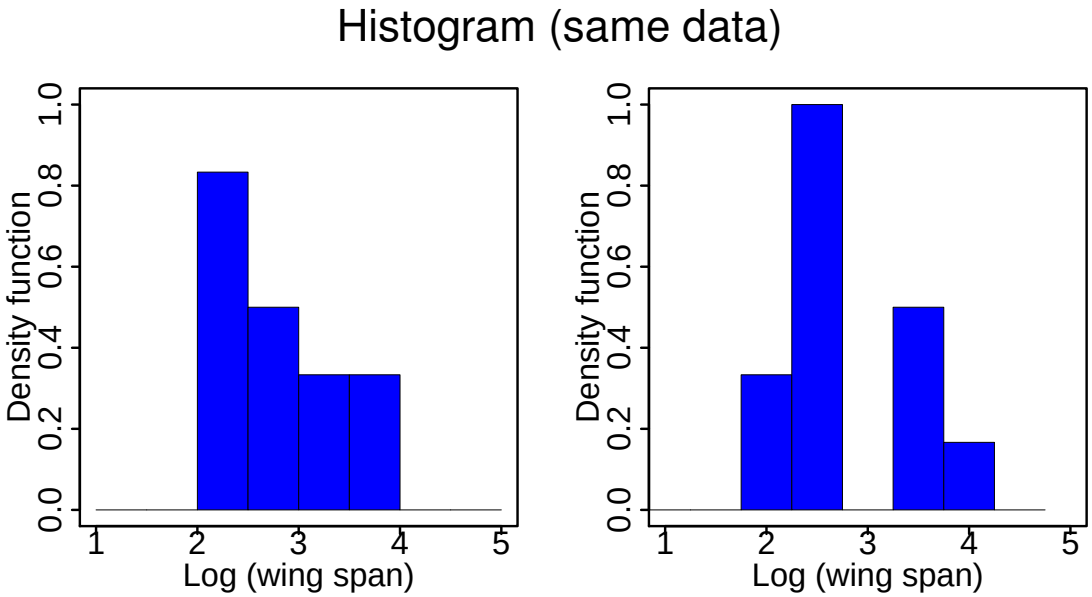


From 3-d microscope images to 2-d locus maps



Ref: (Berger et al., *Nature Meth.*, 2008, In press)

Histograms vs kernel estimators

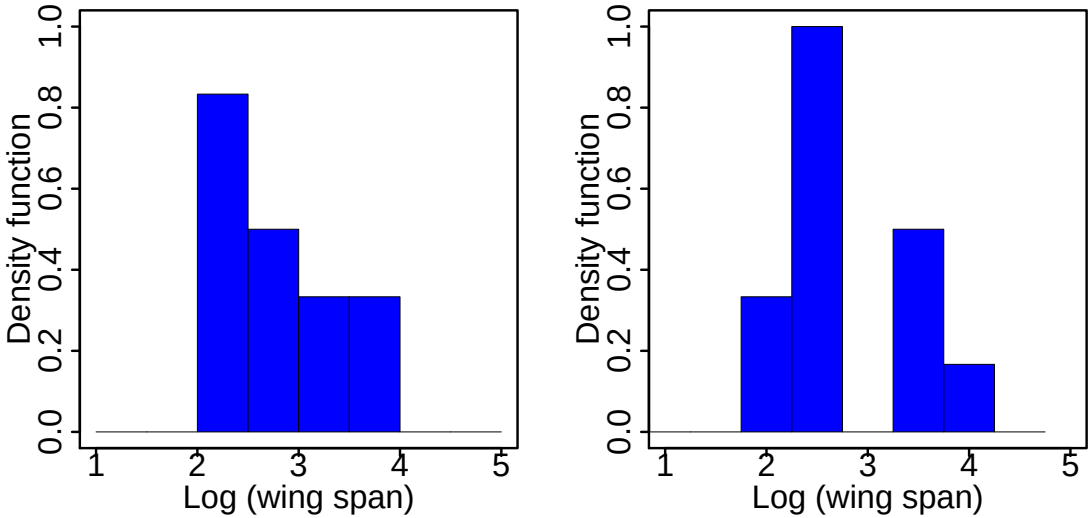


$$\hat{f}_{\text{Hist}}(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \in \text{bin}(x)\}$$

arbitrary placement of bin end points
unrealistic jump discontinuities

Histograms vs kernel estimators

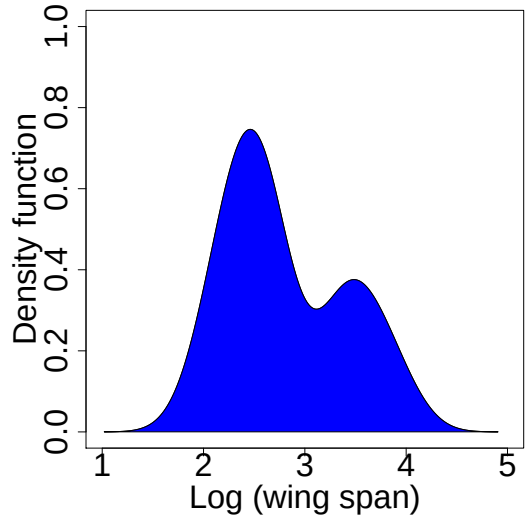
Histogram (same data)



$$\hat{f}_{\text{Hist}}(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \in \text{bin}(x)\}$$

arbitrary placement of bin end points
unrealistic jump discontinuities

Kernel estimator



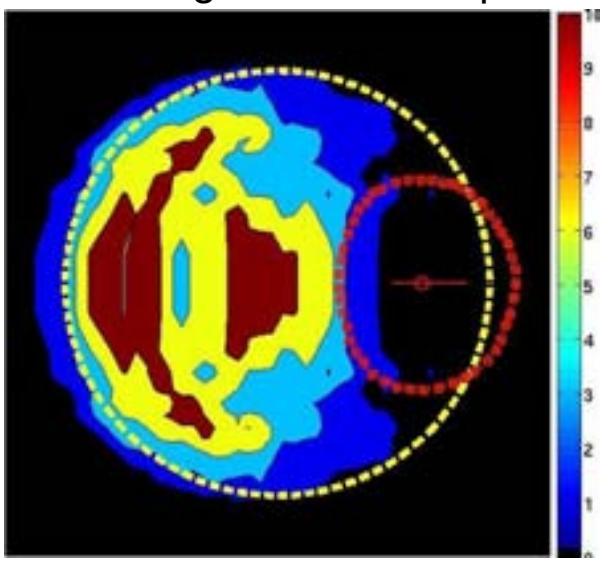
$$\hat{f}_{\text{Kern}}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i)$$

no end points required
no jump discontinuities
faster convergence in prob.

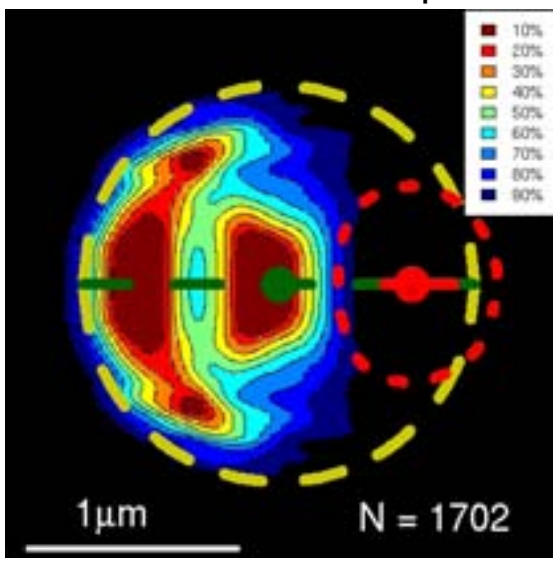
2-d chromosomal locus map

- GAL1 gene under galactose conditions, $n = 1702$ cells
- Visually similar locus maps
- Both are high resolution ($\lesssim 150$nm)
- but kernel smoother map is more statistically rigorous

histogram locus map



kernel locus map



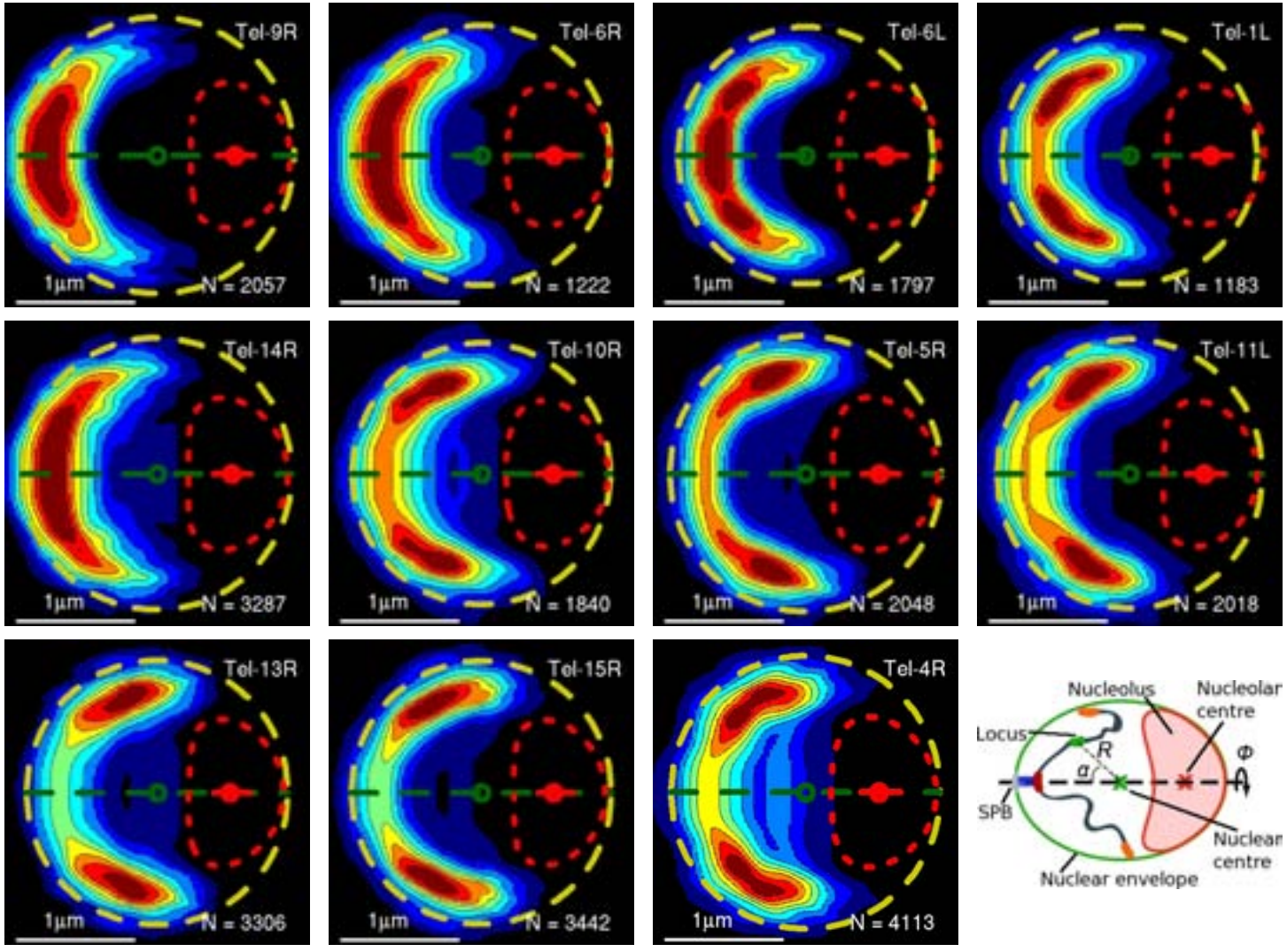
Focusing on telomeres (1)

- Existing evidence that 32 yeast telomeres preferentially localise at the nuclear periphery and form 4 to 5 clusters
  - highly non-uniform localisation inside nucleus
- Ideal candidates for investigating spatial location and nuclear function
  - Working hypothesis: proximity of telomeres directly related to their recombination efficiency (Gotta et al., *JCB*, 1996)

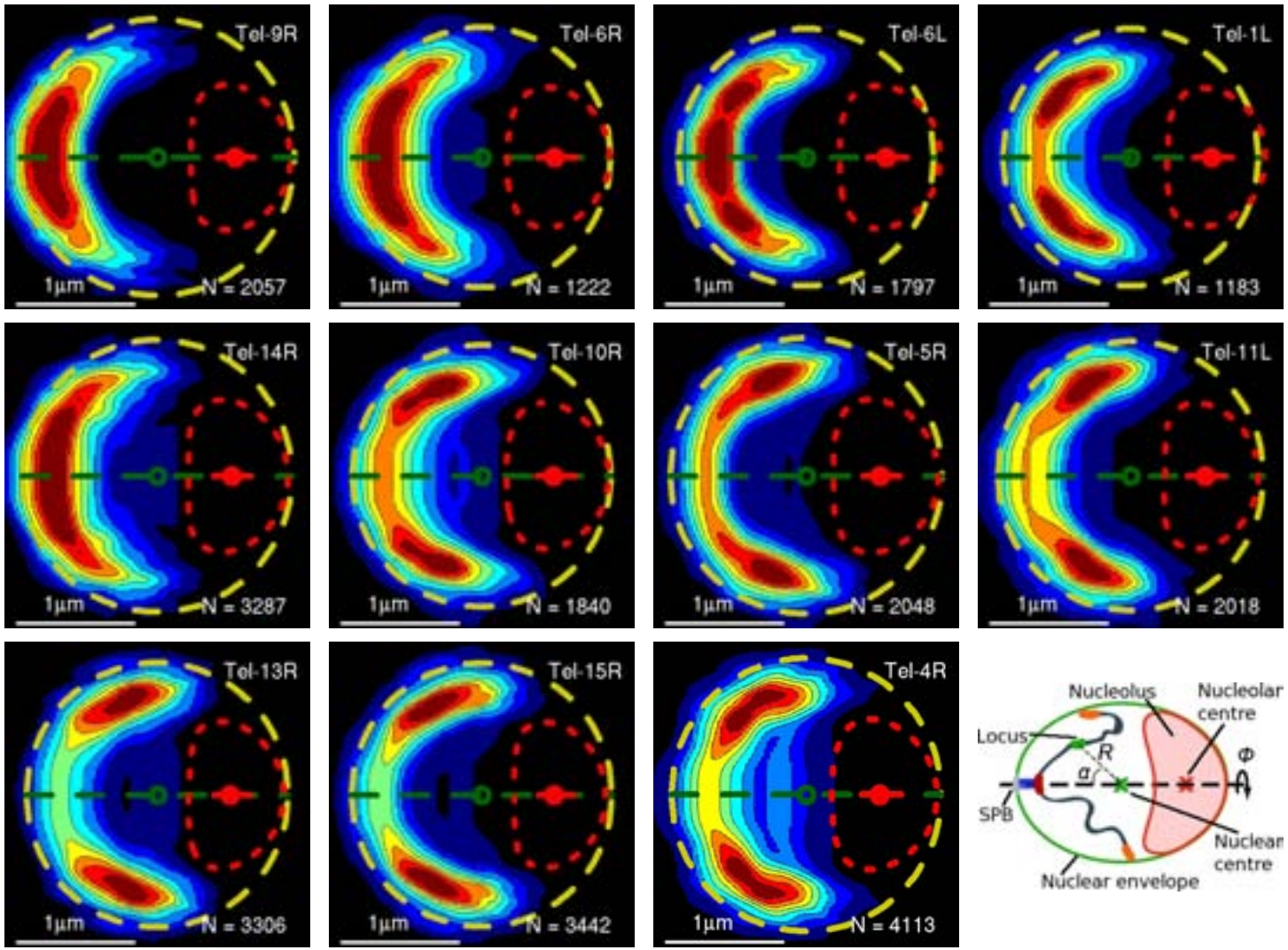
Focusing on telomeres (1)

- Existing evidence that 32 yeast telomeres preferentially localise at the nuclear periphery and form 4 to 5 clusters
  - highly non-uniform localisation inside nucleus
- Ideal candidates for investigating spatial location and nuclear function
  - Working hypothesis: proximity of telomeres directly related to their recombination efficiency (Gotta et al., *JCB*, 1996)
- 2-d locus maps reveal localisation of single locus wrt nuclear landmarks
- but require localisation of telomeres wrt each other

Focusing on telomeres (2)

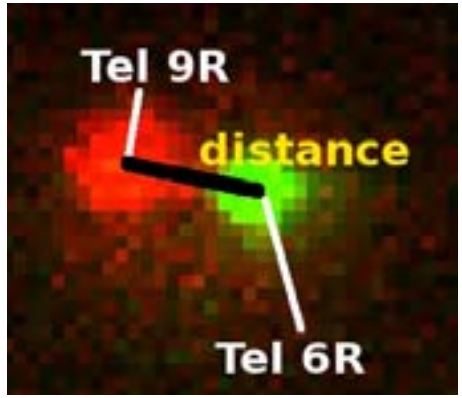


Focusing on telomeres (2)

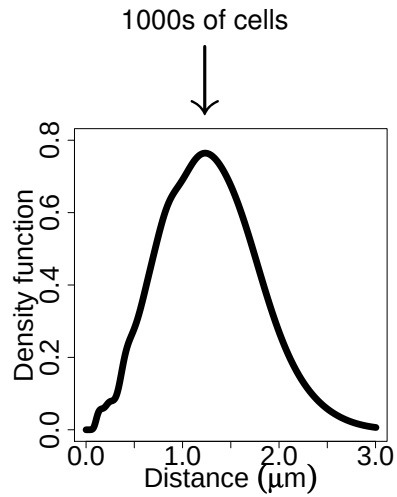


Caveat: Overlapping 2-d maps does NOT imply colocalisation

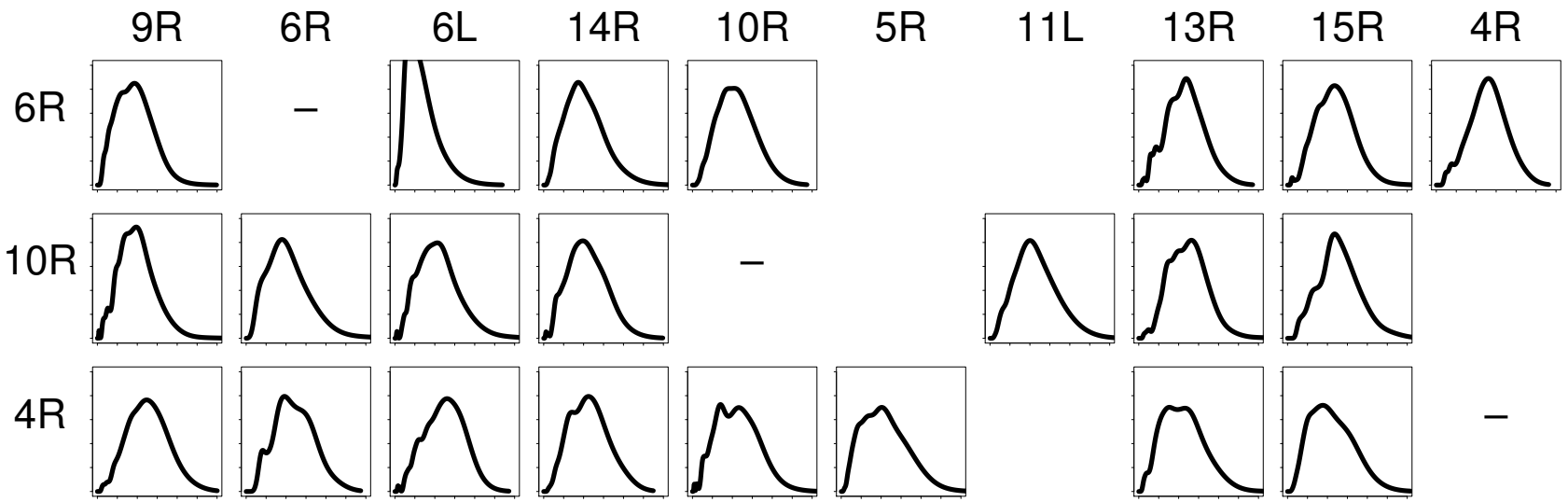
Telomere-telomere pairwise distances (1)



- Distance between two telomeres
- Nuclear landmarks unable to be tagged concurrently (only 2 colours available)
- NB: different experiments to those for chromosomal loci

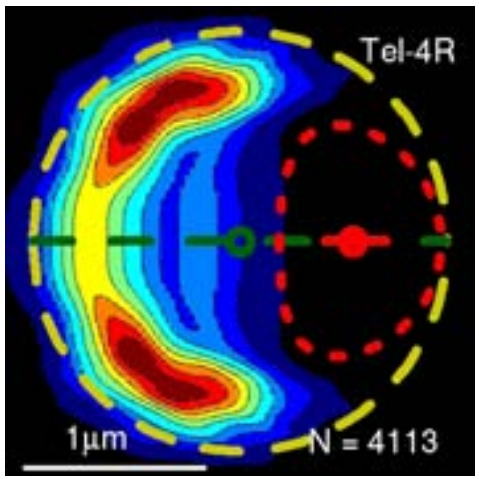


Telomere-telomere pairwise distances (2)

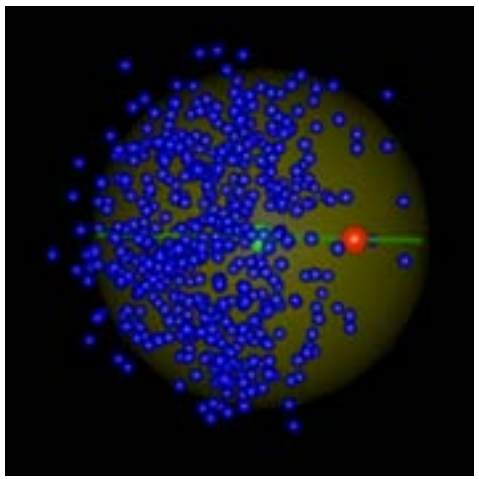


Simple 3-d model (1)

- Limitations of 2-d maps
  - Rotation angle about the central axis (green line) not known
- Assume uniform angles on $[0°, 360°]$

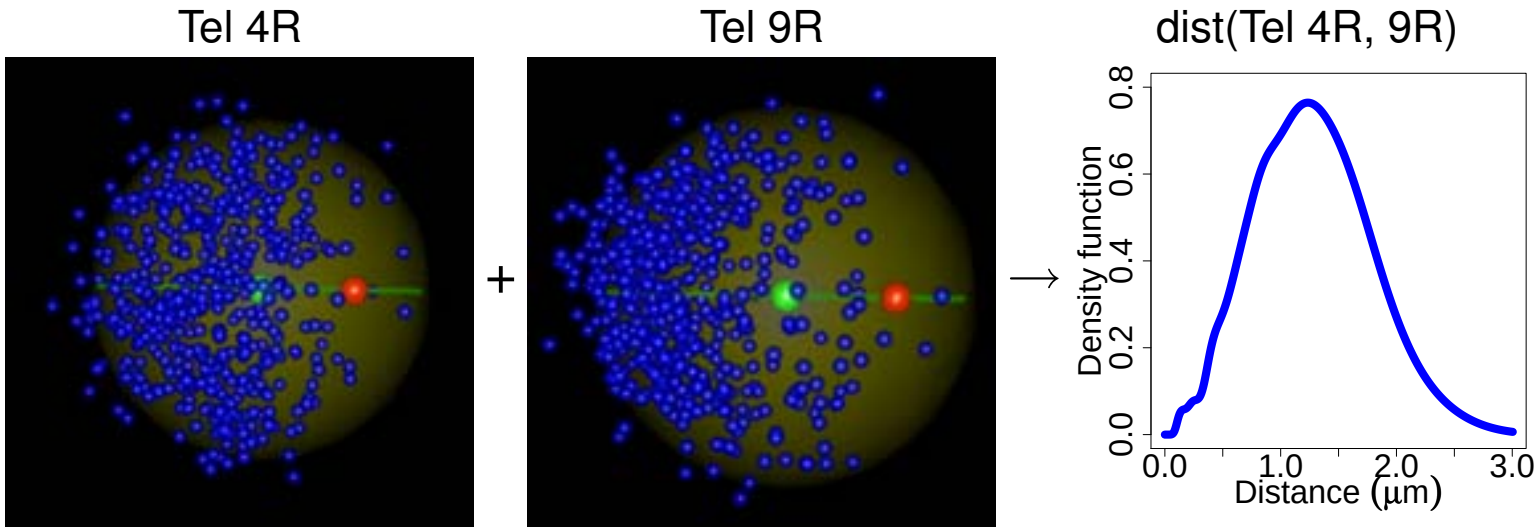


$+\text{Unif}[0°, 360°] \rightarrow$



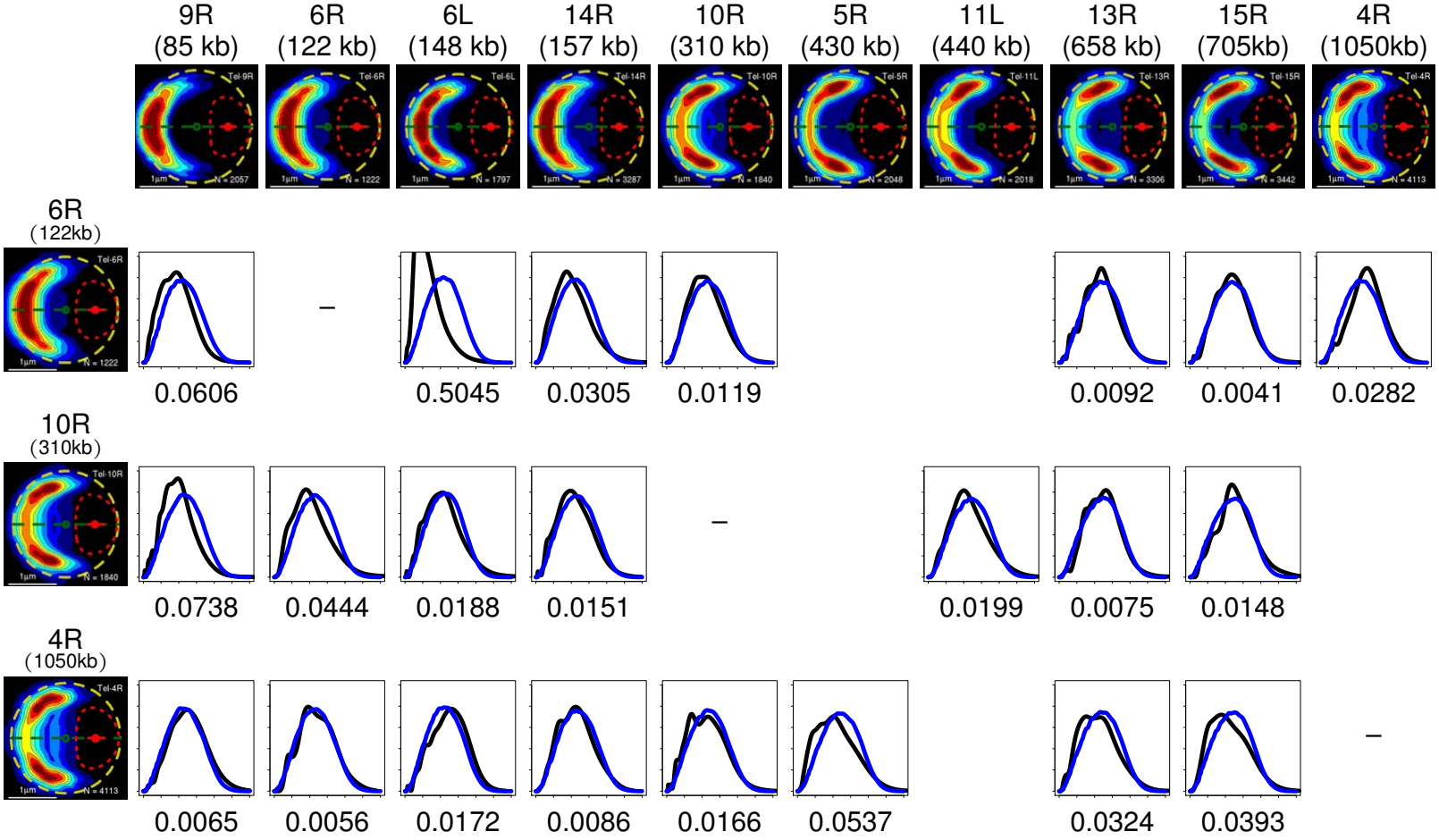
Simple 3-d model (2)

- No direct 3-d validation check of uniformity assumption
- Indirect validation via distance between pairs of telomeres
- With extra assumption of statistical independence



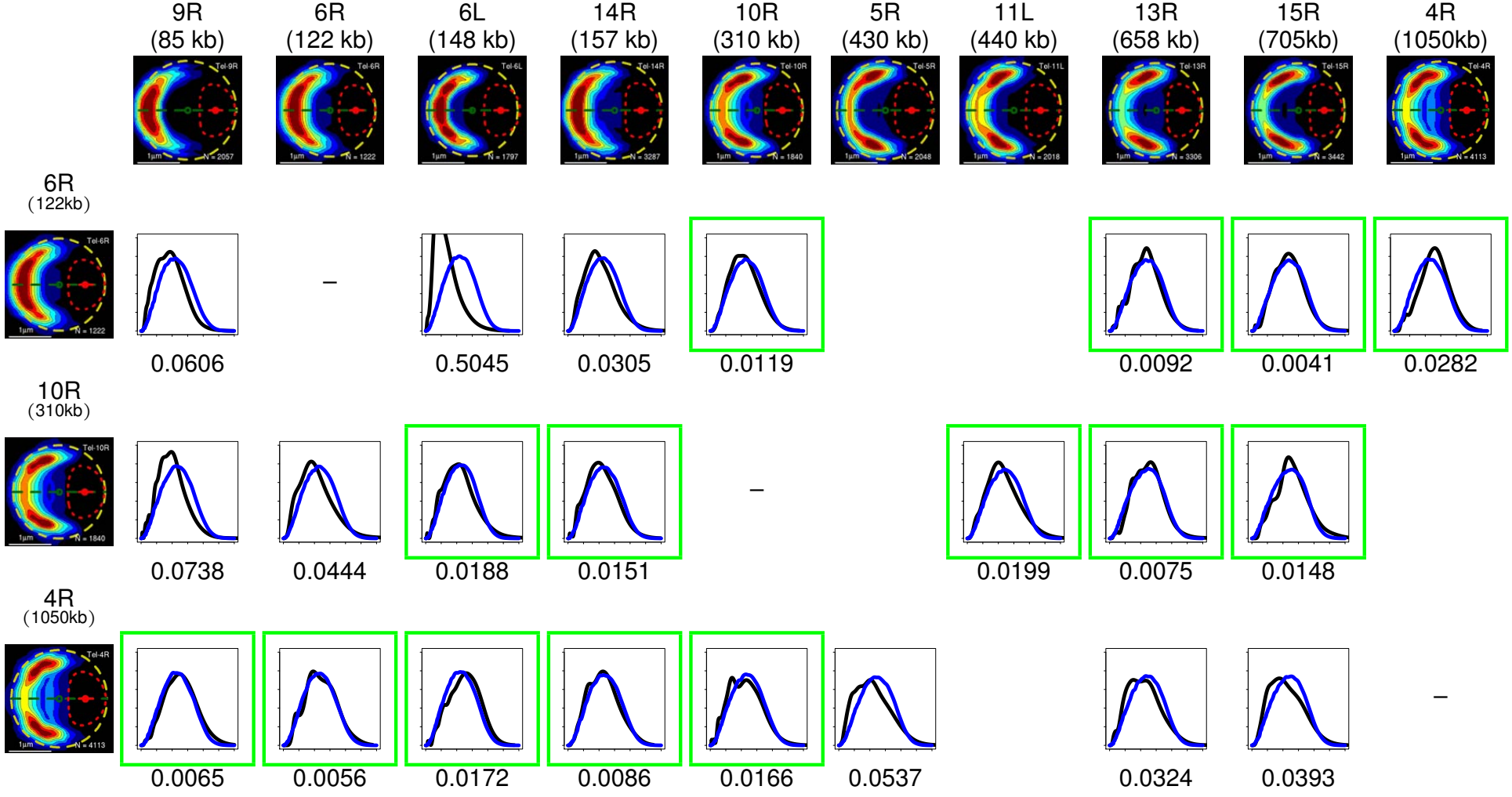
Checking the assumptions of the simple 3-d model

Measured distances Predicted distances (uniform angles + independence)



Checking the assumptions of the simple 3-d model

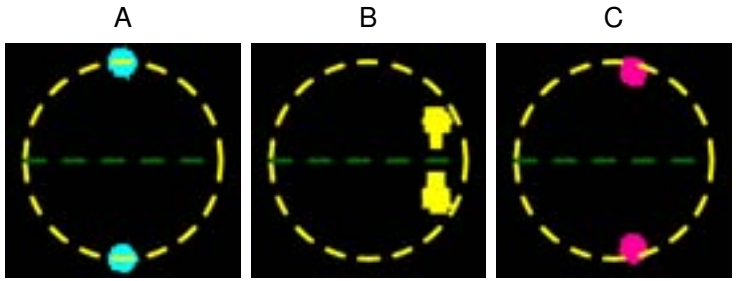
Measured distances Predicted distances (uniform angles + independence) Threshold = $\int (f_{4R6R} - f_{6R4R})^2 = 0.0322$



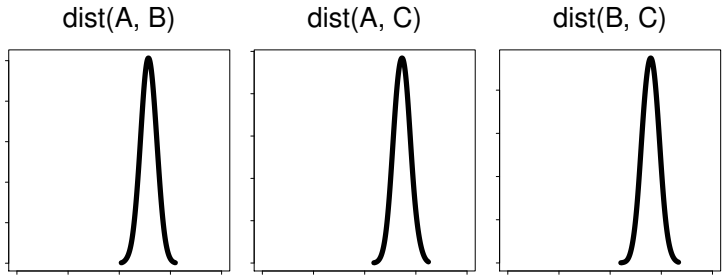
Uniform angles NOT satisfied in all cases

Data-based reconstruction (synthetic example)

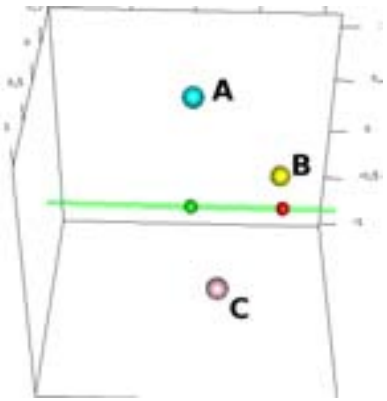
2-d locus maps



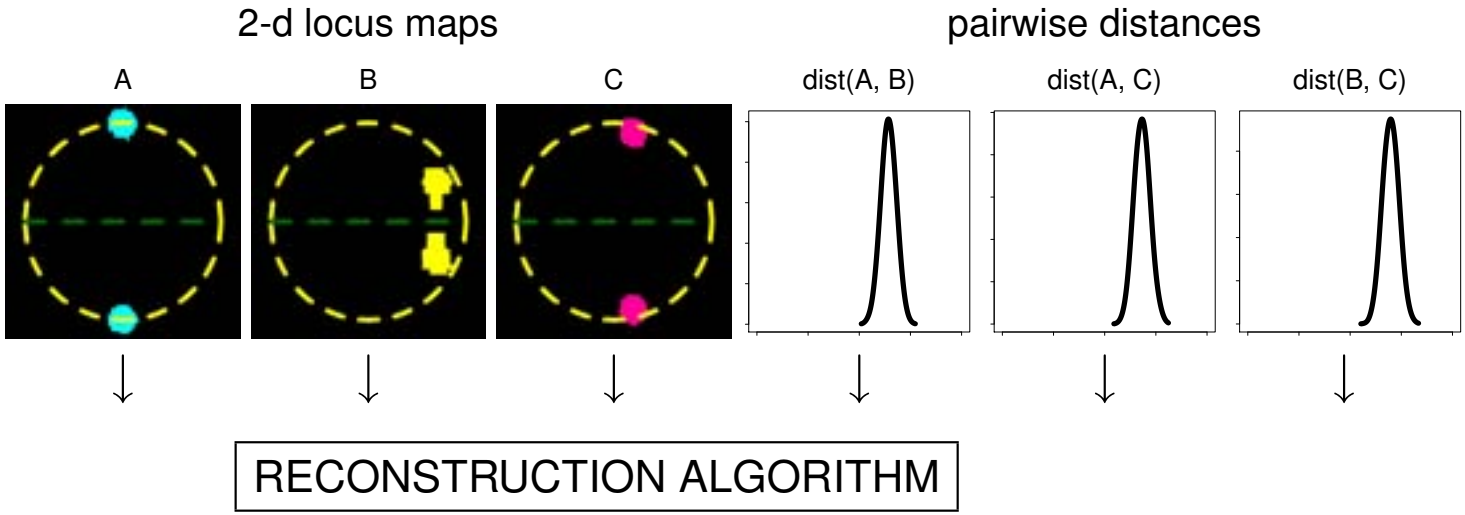
pairwise distances



True median locations

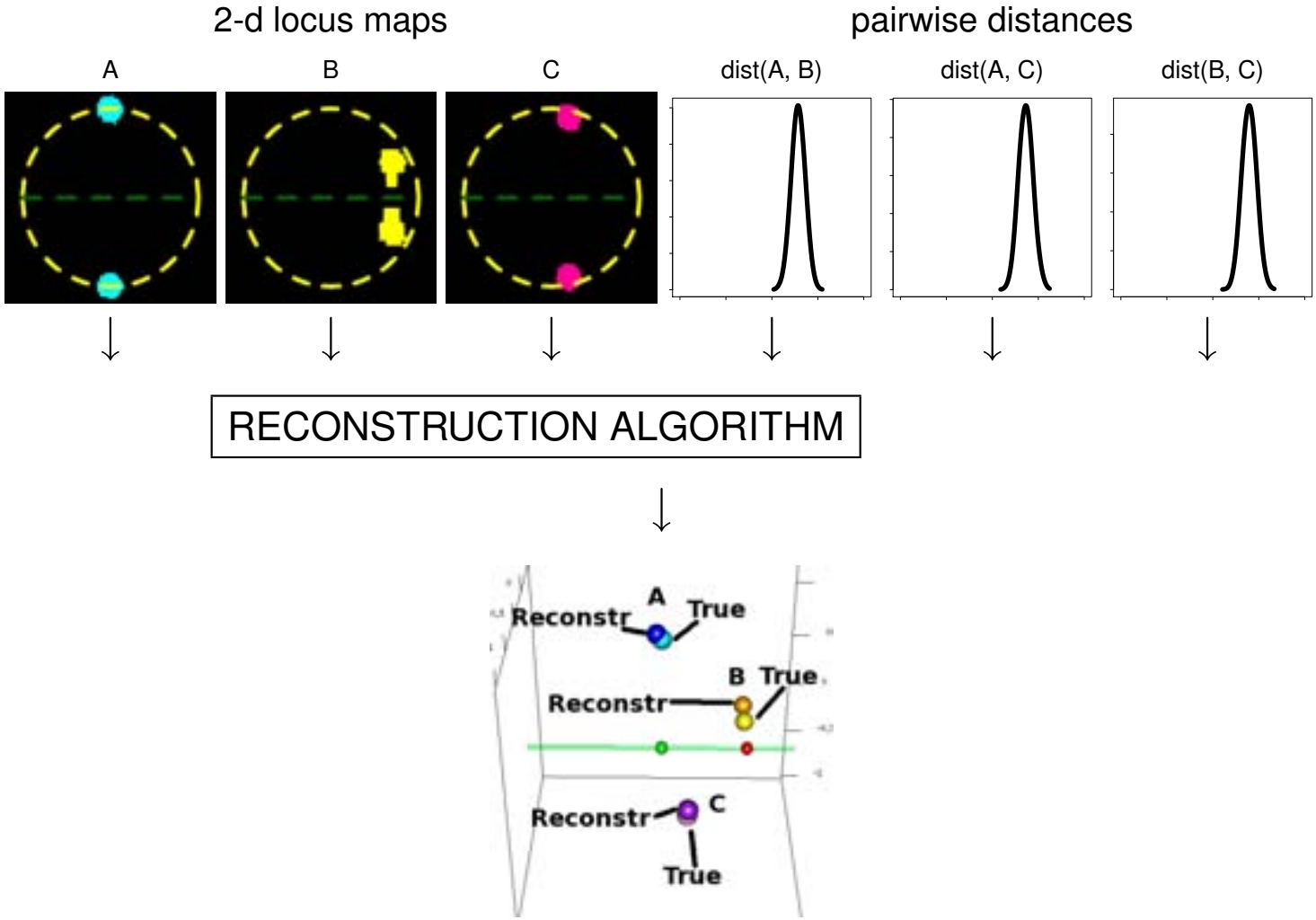


Data-based reconstruction (synthetic example)



- All six inputs (3 locus maps, 3 distances) are each drawn from different cell populations
- Match pairs of locus maps with corr. pairwise distances
- Simulation from non-parametric distributions

Data-based reconstruction (synthetic example)



Next steps

- Thoroughly test reconstruction algorithm on more simulation settings
- Apply to real telomeres data

Summary and future directions

Yeast nuclear organisation revealed in more details with:

- 2-d chromosomal kernel locus map
  - high resolution i.e. not limited by microscope diffraction
  - statistically rigorous
- 3-d reconstruction given 2-d locus maps and pairwise distances
  - work in progress
  - ultimate goal to reconstruct 3-d location of complete genome
  - connections to physical models of genome

Acknowledgements

PTR 218

- Unité de Génétique des Levures, Institut Pasteur
  - P. Thérizols, E. Fabre
- Unité de Biologie Cellulaire du Noyau, Institut Pasteur
  - A.B. Berger, G.G. Cabal, F. Feuerbach
- Laboratoire de Biologie Moléculaire des Eucaryotes, CNRS, Université de Toulouse
  - O. Gadal
- Plate-forme d'Imagerie Dynamique, Institut Pasteur
- G5 Imagerie et Modélisation, Institut Pasteur
  - C. Zimmer, M. Lelek