L1.2

- L1.2.1 - Human Genome variation
- L1.2.2 - Chromosomes

HGP 2003 → Post-genomics

Genetics

Comparative (phylogenetic conservation indicates conserved function)

Human Genetic Variation (1000 Human Genomes - HapMap)

GWAS – Genome variations – phenotype correlation

Gene expression and phenotype

Functional Genomics (ENCODE – FANTOM)

Epigenomics: CpG methylation

Histone modifications (PTMs)

Chromatin status

Protein-DNA mapping (e.g. transcription factors)

Transcriptomics: Coding and noncoding RNAs

Human Genome Project

Human genetic variation

Genetic analysis of diseases

Functional annotation of the Human Genome (and others, e.g. mouse)

The Encyclopedia of DNA Elements (ENCODE) + other similar projects (e.g. FANTOM)

The idea was to obtain functional information for every single nucleotide of the human genome

Started in 2000 using automated Sanger sequencing on 1% human genome (ca. 30 Mb), completed in 2006

With the advent of Next Generation Sequencing Technology, first draft completed in 2012

Genetics

Individual genomes display **variants**

SNP – single nucleotide polymorphisms

Indels – insertions and deletions

CNV – copy number variations

TE – transposable elements number and position

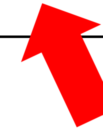
Variants are associated to more or less evident **phenotypes**

Some variants are clearly associated to specific **pathologies**.

Other variants are associated only weakly with a phenotype but require other variants (often in other loci) to become significantly associated (combinatorial association).

Projects are under way to describe all variants associated to risk of disease (GWAS: Genome Wide Association Studies)

16 | VOL.5 NO.1 | JANUARY 2008 | NATURE METHODS



SPECIAL FEATURE | COMMENTARY

Next-generation sequencing transforms today's biology

Stephan C Schuster

A new generation of non-Sanger-based sequencing technologies has delivered on its promise of sequencing DNA at unprecedented speed, thereby enabling impressive scientific achievements and novel biological applications. However, before stepping into the limelight, next-generation sequencing had to overcome the inertia of a field that relied on Sanger-sequencing for 30 years.

NGS

Fragment the DNA (or RNA) to be sequenced in smaller pieces

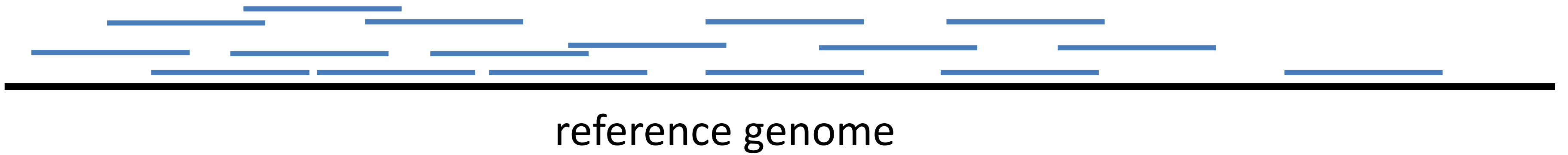
Physically separate the fragments

Highly-parallel sequencing of fragments, high-throughput

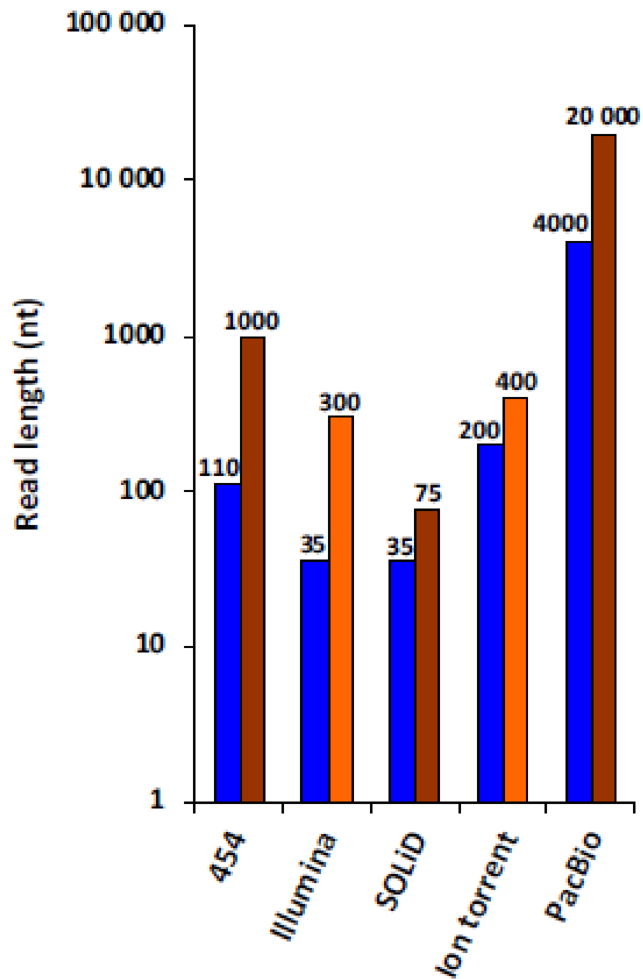
No cloning step required

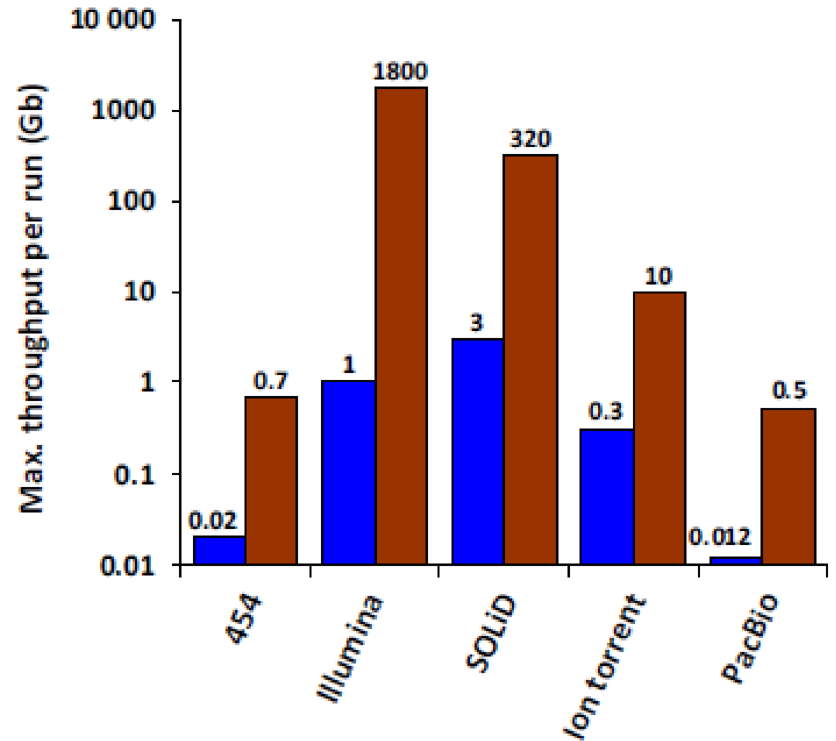
NGS sequencing produces hundreds of millions of short «reads» per run

Reads are mapped to the reference genome



reference genome

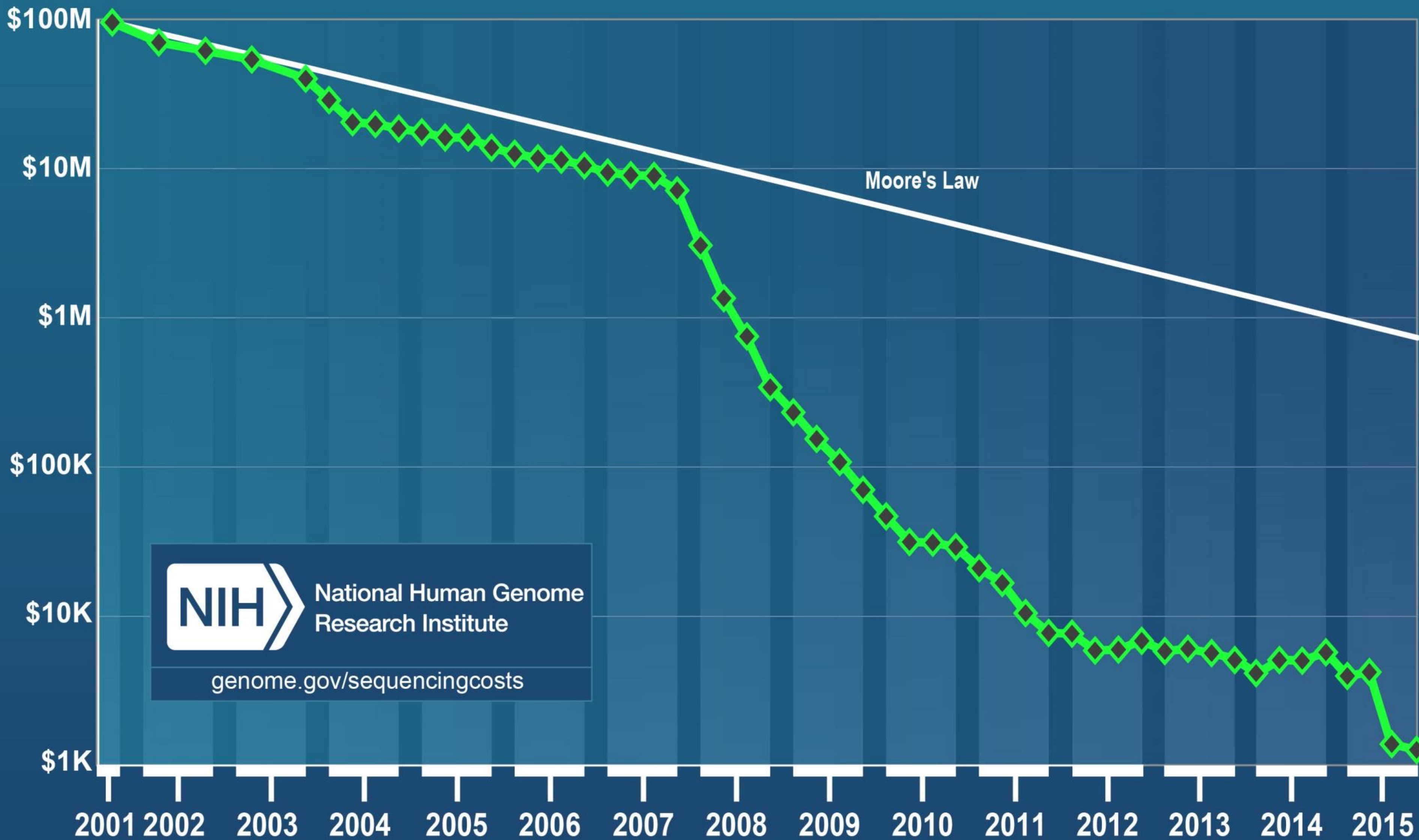




In blue the first version of the instruments

From Van Dijk et al., 2014 (Textbook)

Cost per Genome



1000 Human Genomes, HapMap project
Describing variations among genomes of individuals

GWAS
Genome-wide association studies
Variations (SNPs, CNV, indels) studied in individuals as related to the occurence of a phenotype (pathology, risks, other features)

TCGA – The Cance Genome Atlas
Sequencing of tumor cell DNA to evidence mutations occurring in tumors.

The 1000 Genomes Project

http://www.internationalgenome.org/

Started immediately after the HGP but it was dramatically accelerated by introduction of NGS

2010 report

ARTICLE

doi:10.1038/nature09534

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

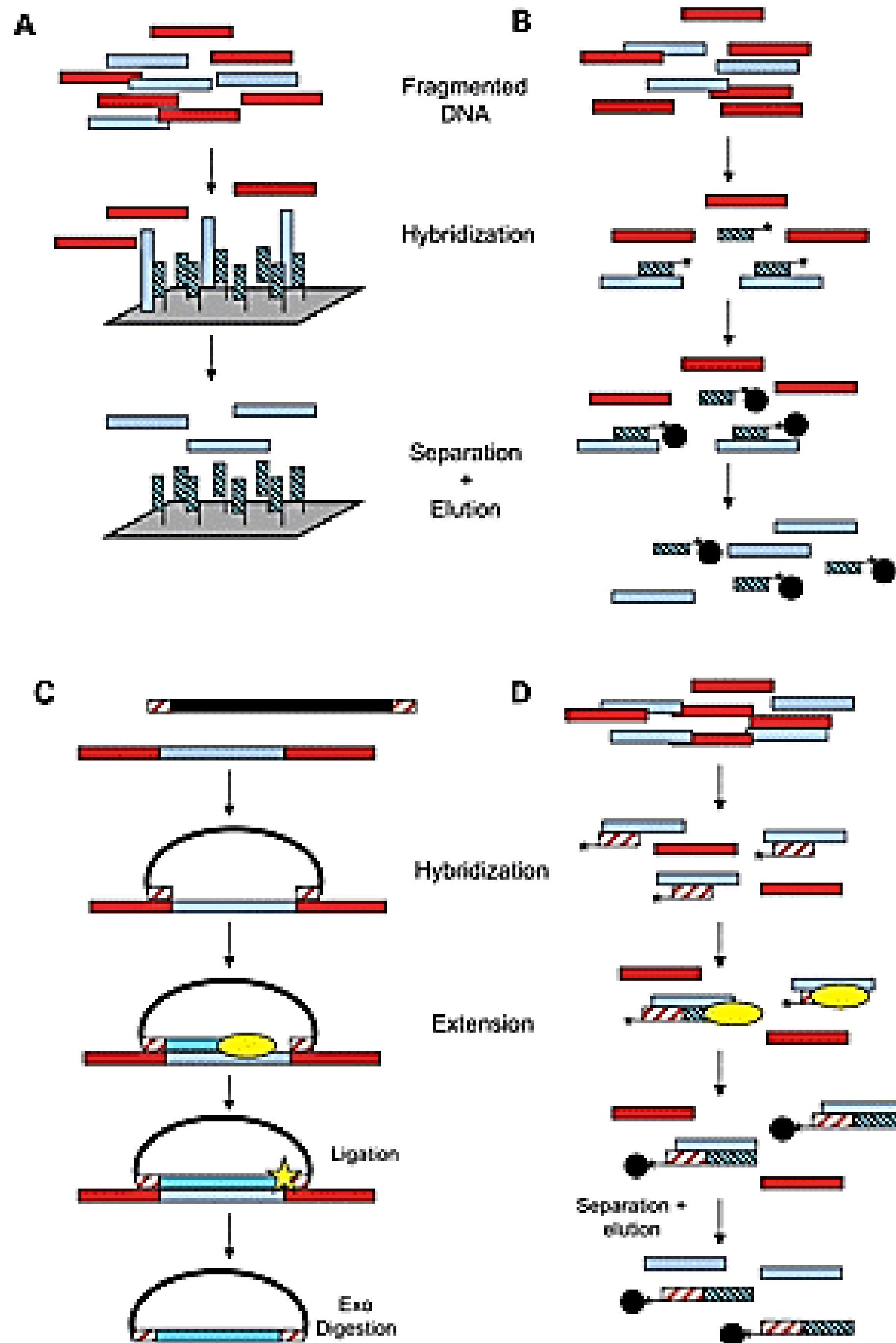
The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother-father-child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of *de novo* germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

Exome sequencing

Due to elevated costs, many studies were limited to the «**exome**»
Exome is the set of sequences that make up all known mRNAs.

Requires enrichment of exon sequences from a genomic DNA. This is obtained using different methods, as exemplified in these schemes.

From: Teer and Mullikin, 2010. Hum Mol Genet. 9(R2):R145-51



ARTICLE

OPEN

doi:10.1038/nature15393

A global reference for human genetic variation

The 1000 Genomes Project Consortium*

The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. We characterized a broad spectrum of genetic variation, in total over 88 million variants (84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all phased onto high-quality haplotypes. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

WGS
exome
µarray genotyping

Abstract

The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 26 populations using a combination of **low-coverage whole-genome sequencing**, **deep exome sequencing**, and **dense microarray genotyping**. We characterized a broad spectrum of genetic variation, in total over 88 million variants (84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all **phased onto high-quality haplotypes**. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

68 | NATURE | VOL 526 | 1 OCTOBER 2015

Note:

orange buttons mean something that you should search for and contribute to in the Methodological Wiki on the Moodle site

Student activities

wikis, databases, background, methodology, Forum

Students Wiki on methodology: Group choice

1

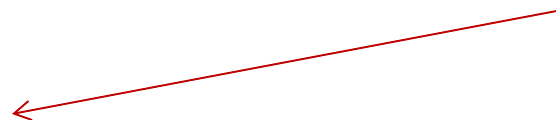
Select which subject you want to contribute to.

2

How to use the Wiki

3

Student Wiki on methodology



nature

68 | NATURE | VOL 526 | 1 OCTOBER 2015

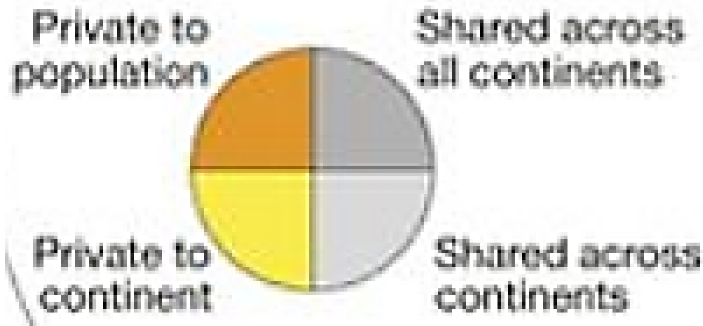
Population sampling.

a

GBR FIN CEU TSI IBS PJL CHB BEB JPT CHS GIH ITU STU CDX KHV ASW PUR MXL ACB GWD ESN MSL YRI LWK CLM PEL

24 million 18 million 12 million

Private to population
Shared across all continents
Private to continent
Shared across continents



Graphs sow the fraction of variants that are either shared among all Humans or private to continents or populations.

[The 1000 Genomes Project Consortium]

A typical genome

We find that a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites (Fig. 1b and Table 1).
Although **>99.9% of variants consist of SNPs and short indels**, **structural variants affect more bases**: the typical genome contains an estimated **2,100 to 2,500 structural variants** (∼1,000 large deletions, ∼160 copy-number variants, ∼915 Alu insertions, ∼128 L1 insertions, ∼51 SVA insertions, ∼4 NUMTs, and ∼10 inversions), affecting ∼20 million bases of sequence.

NUMT=nuclear mitochondrial DNA segment

The 1000 Genomes Project Consortium

68 | NATURE | VOL 526 | 1 OCTOBER 2015

The majority of variants in the data set are **rare**: ∼64 million autosomal variants have a frequency <0.5%, ∼12 million have a frequency between 0.5% and 5%, and only ∼8 million have a frequency >5%

Nevertheless, the majority of variants observed in a single genome are common: just 40,000 to 200,000 of the variants in a typical genome (1–4%) have a frequency <0.5%

Genome Browser for variants, with data from 1000HGP is available:

https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/

Comparative genomics.

https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/

Is there any relationship between the evolutionary degree (organismal complexity) and genome size ?

Textbook G

Leading Edge
Review

Cell

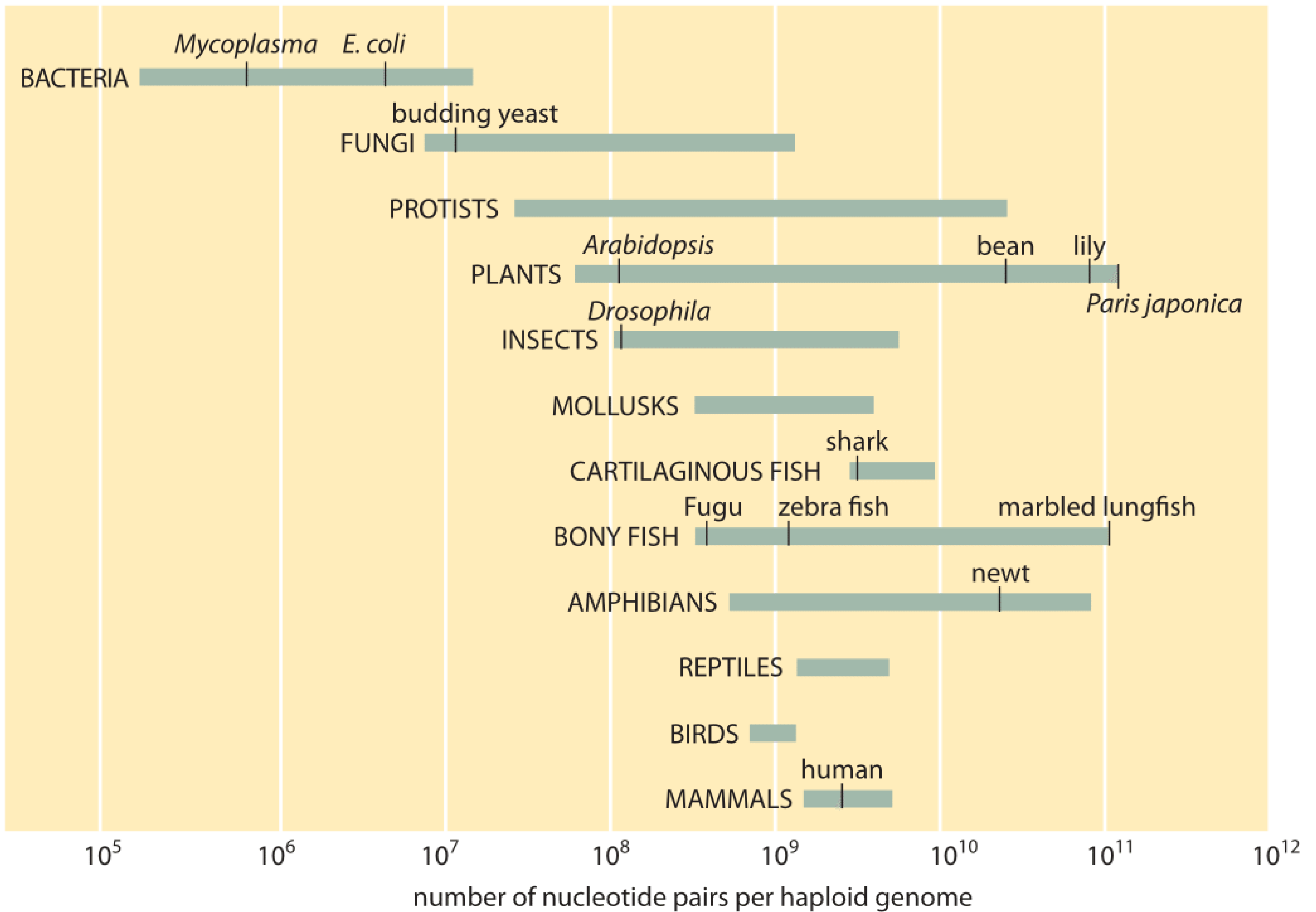
Looping Back to Leap Forward: Transcription Enters a New Era

Michael Levine,[1,*] Claudia Cattoglio,[1,2] and Robert Tjian[1,2,*]
[1]Department of Molecular and Cell Biology
[2]Howard Hughes Medical Institute, CIRM Center of Excellence, Li Ka Shing Center for Biomedical and Health Sciences
University of California, Berkeley, Berkeley, CA 94707, USA
*Correspondence: mlevine@berkeley.edu (M.L.), jmlim@uclink4.berkeley.edu (R.T.)
http://dx.doi.org/10.1016/j.cell.2014.02.009

Comparative genome analyses reveal that organismal complexity scales not with gene number but with gene regulation. Recent efforts indicate that the human genome likely contains hundreds of thousands of enhancers, with a typical gene embedded in a milieu of tens of enhancers. Proliferation of *cis*-regulatory DNAs is accompanied by increased complexity and functional diversification of transcriptional machineries recognizing distal enhancers and core promoters and by the high-order spatial organization of genetic elements. We review progress in unraveling one of the outstanding mysteries of modern biology: the dynamic communication of remote enhancers with target promoters in the specification of cellular identity.

Comparative genome analyses reveal that organismal complexity scales not with gene number but with gene regulation. Recent efforts indicate that the human genome likely contains hundreds of thousands of enhancers, with a typical gene embedded in a milieu of tens of enhancers. Proliferation of cis-regulatory DNAs is accompanied by increased complexity and functional diversification of transcriptional machineries recognizing distal enhancers and core promoters and by the high-order spatial organization of genetic elements. We review progress in unraveling one of the outstanding mysteries of modern biology: the dynamic communication of remote enhancers with target promoters in the specification of cellular identity.

organismal complexity scales not with gene number but with gene regulation


mostly coding DNA
mostly non-coding DNA
Mycoplasma
E. coli
BACTERIA
budding yeast
FUNGI
PROTISTS
Arabidopsis
bean
lily
PLANTS
Paris japonica
Drosophila
INSECTS
MOLLUSKS
shark
CARTILAGINOUS FISH
Fugu
zebra fish
marbled lungfish
BONY FISH
newt
AMPHIBIANS
REPTILES
BIRDS
human
MAMMALS
10^5
10^6
10^7
10^8
10^9
10^{10}
10^{11}
10^{12}
number of nucleotide pairs per haploid genome

For those of you interested in Transposable Elements biology:

Current Opinion in Genetics and Development, April 2018 issue:

https://www.sciencedirect.com/journal/current-opinion-in-genetics-and-development/vol/49/suppl/C

L1.1.2

Chromosomes and Nuclear organization

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Cell Biology

Nuclear organization: taking a position on gene expression

Pamela K Geyer, Michael W Vitalini and Lori L Wallrath

Textbook

Eukaryotic genomes are divided into chromosomes that occupy defined regions or territories within the nucleus. These chromosome territories (CTs) are arranged based on the transcriptional activity and chromatin landscape of domains. In general, transcriptionally silent domains reside at the nuclear periphery, whereas active domains locate within the interior. Changes in nuclear position are observed for stress-induced and developmentally regulated tissue-specific genes. Upon activation, these genes move away from a CT to inter-chromosomal space containing nuclear bodies enriched in gene expression machinery. Gene activation is not always accompanied by movement, as positioning is dictated by many determinants, including gene structure and the local genomic environment. Collectively, tissue-specific nuclear organization results from a culmination of inputs that result in proper transcriptional regulation.

Address
University of Iowa, Iowa City, IA 52242, United States

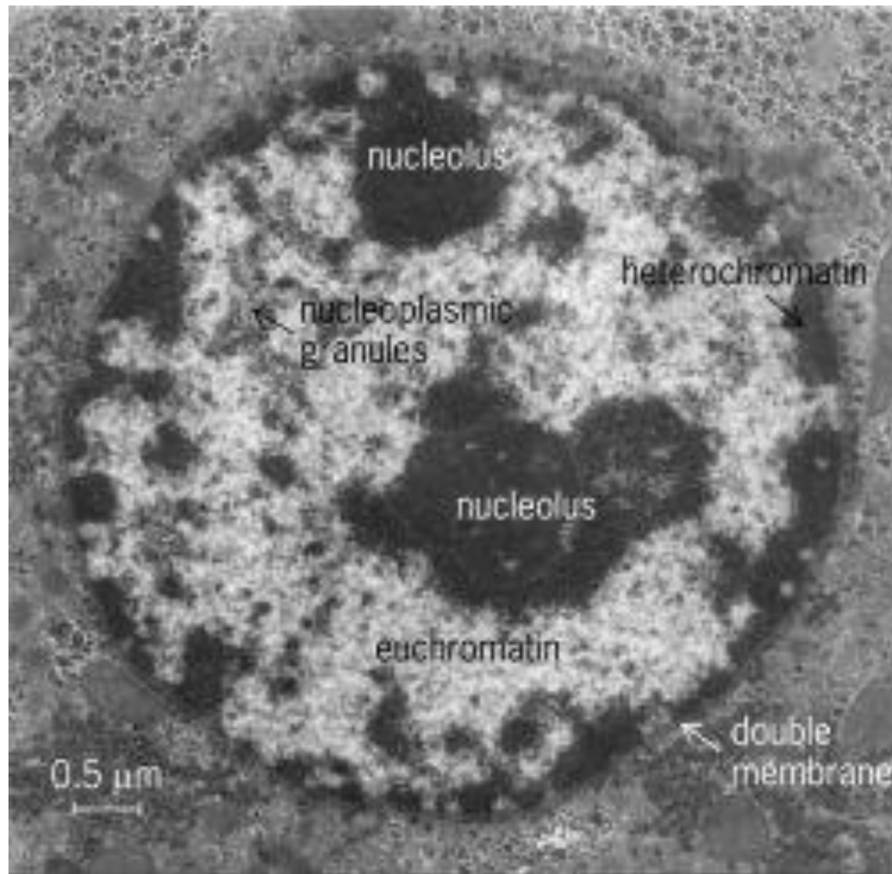
Corresponding author: Wallrath, Lori L (lori-wallrath@uiowa.edu)

Current Opinion in Cell Biology 2011, **23**:354–359

chemical cross-linking and massive parallel sequencing to define genome-wide relationships [3–5]. Results from these studies suggest that the genome is arranged as inter-digitated CTs rather than randomly inter-twined chromosomes [6]. Emerging from these investigations is a picture of the nucleus as an ordered organelle; the consequences of this organization are just being realized.

Nuclear organization during differentiation

Studies have linked nuclear organization to cellular differentiation. Cultured pluripotent mouse embryonic stem (ES) cells possess dispersed chromatin with limited compaction. Upon differentiation, they show changes in chromatin structure that include large-scale compaction of genomic domains [7]. Consistent with these findings, embryonic development proceeds from a single cell embryo possessing a 'featureless' nucleus with dispersed chromatin, to differentiated cells possessing nuclei with peripherally located compact chromatin domains [8]. Interestingly, an extended and dispersed chromatin meshwork was identified in the eight-cell epiblast, reminiscent of nuclear structures defined in cultured ES cells. In contrast to the 'open' chromatin structure in the epiblast nuclei,



A rat liver cell nucleus at Transmission E.M.

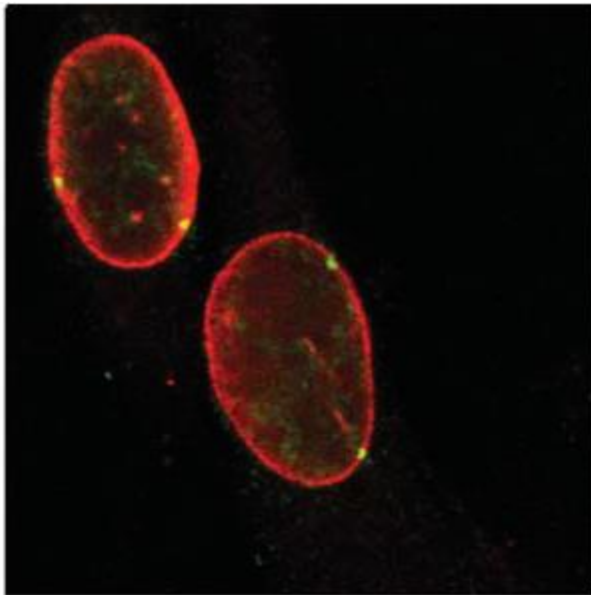
Eukaryotic genomes are divided into chromosomes that occupy defined regions or territories within the nucleus. These chromosome territories (CTs) are arranged based on the transcriptional activity and chromatin landscape of domains. In general, transcriptionally silent domains reside at the nuclear periphery, whereas active domains locate within the interior. Changes in nuclear position are observed for stress-induced and developmentally regulated tissue-specific genes. Upon activation, these genes move away from a CT to inter-chromosomal space containing nuclear bodies enriched in gene expression machinery. Gene activation is not always accompanied by movement, as positioning is dictated by many determinants, including gene structure and the local genomic environment. Collectively, tissue-specific nuclear organization results from a culmination of inputs that result in proper transcriptional regulation.

Gene-poor regions and **silenced genes** are frequently found at the **nuclear periphery** (same as heterochromatin)

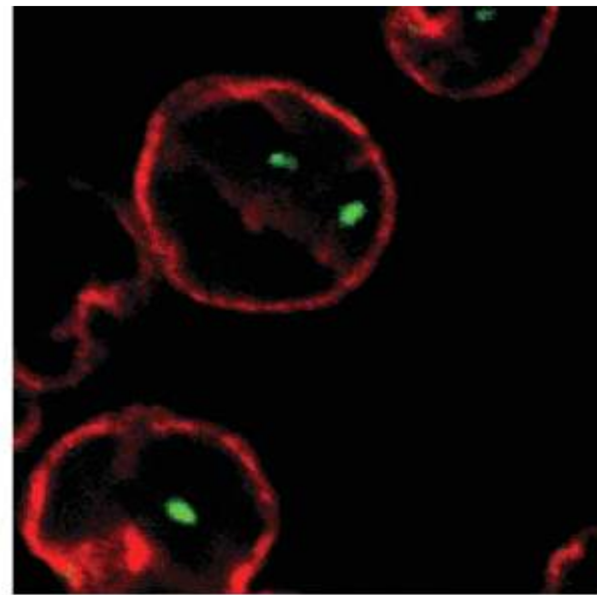
How do we know this ?

3D DNA immunoFISH of the **Igh** loci

What is this ? Go to NCBI-Gene database



fibroblast



Pro-B cell

Immuno-FISH

From: Reddy et al., 2008 (Research paper 1)

Several new concepts:

Chromosomal territories

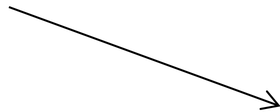
Chromatin landscapes

Transcriptionally silent domains

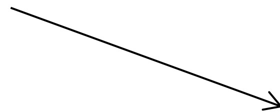
Transcriptionally active domains

Changes in nuclear position (induced genes)

Input



tissue specific nuclear organization



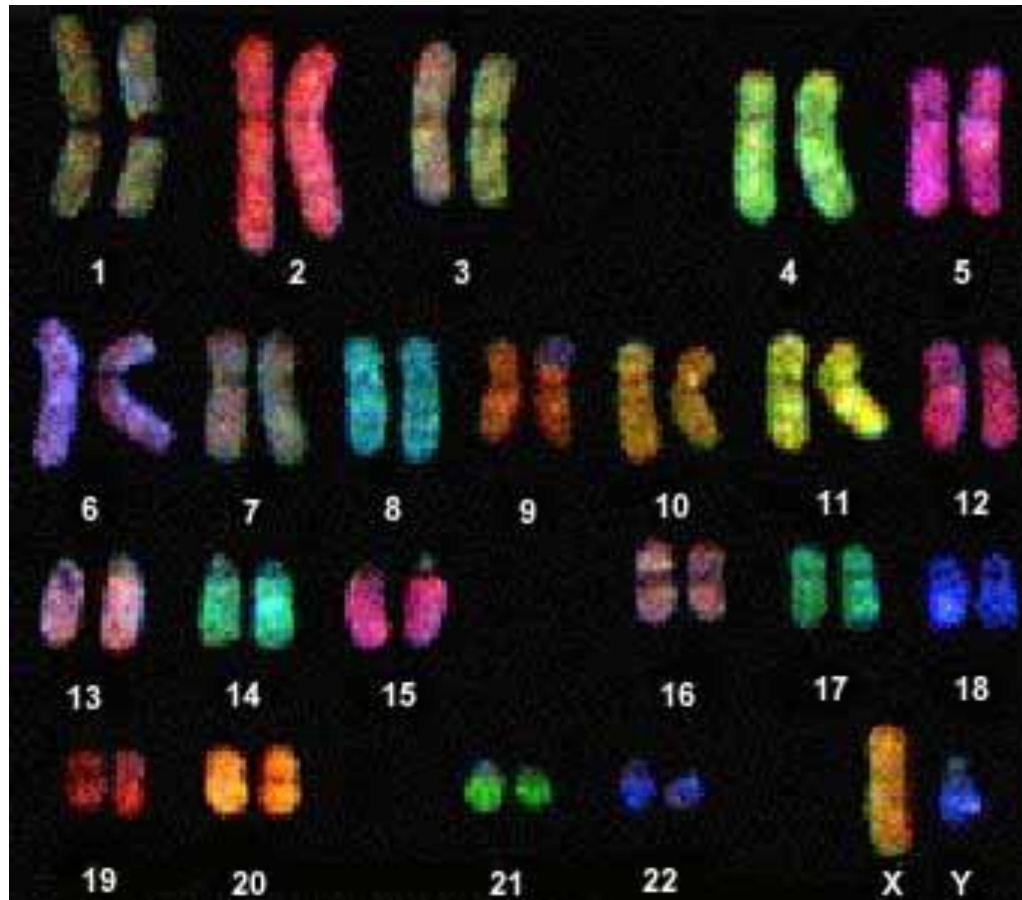
proper transcriptional regulation

Chromosomal territories (CT)

Individual chromosomes can be «painted» and maintain separate territories in interphase nuclei

Chromosome painting techniques can visualize individual chromosomes both at the metaphase and interphase

Individual «chromosome painting» is a **FISH technique**

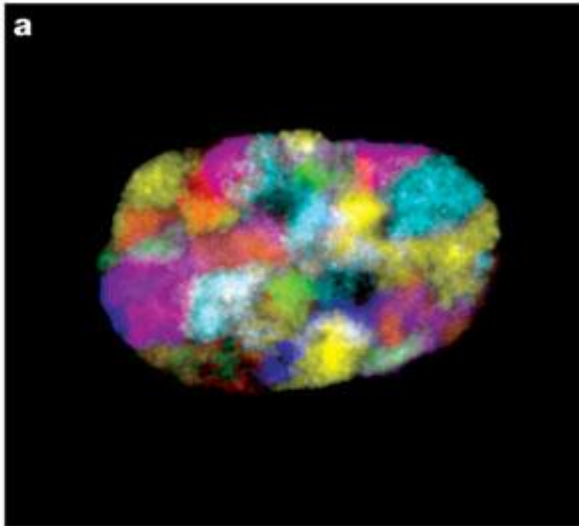


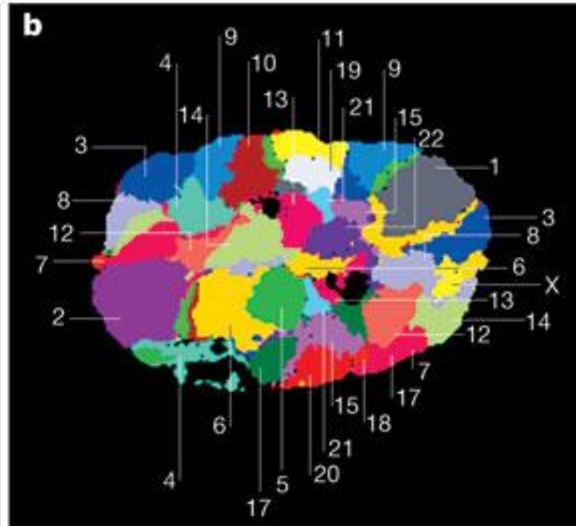
Metaphase human chromosomes

Results from these studies suggest that the genome is arranged as inter-digitated CTs rather than randomly inter-twined chromosomes [6]. Emerging from these investigations is a picture of the nucleus as an ordered organelle; the consequences of this organization are just being realized.

Interdigitated, inter-twined or inter-twingled CTs ?

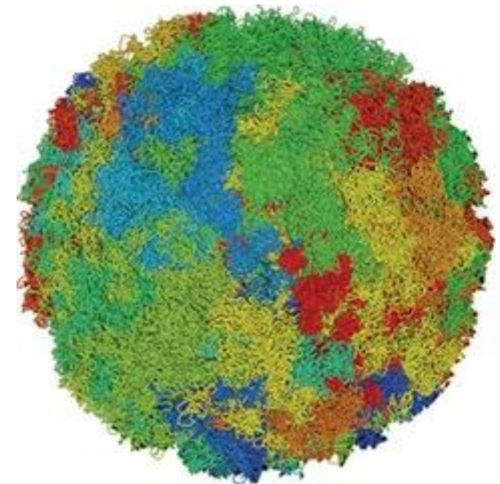
CT = chromosomal territory





Copyright © 2005 Nature Publishing Group
Nature Reviews | Genetics

Mapping CTs in cell nucleus



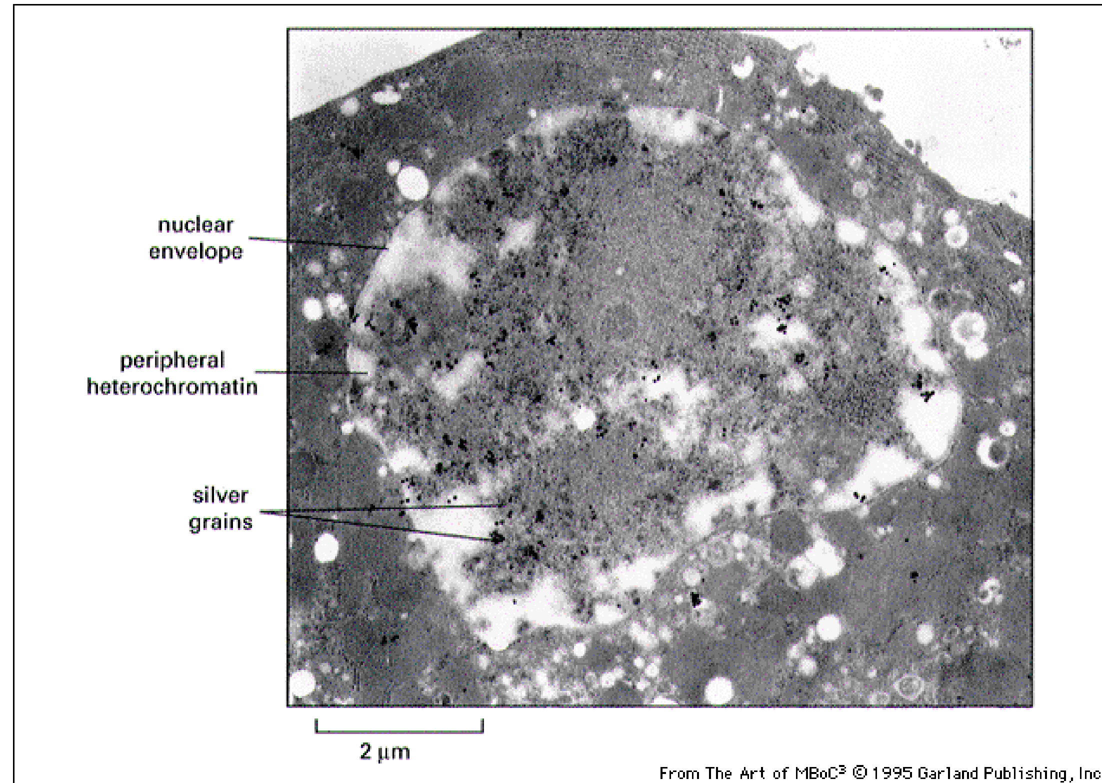
One interesting question: if heterochromatin and CT are established during differentiation...

Does heterochromatin i.e. nuclear organization,

differ

from one cell type to another?

This is quite old concept, indeed….



Macroscopic

http://www.google.it/search?q=nucleus+electron+microscopy&hl=it&client=firefox-a&hs=rE0&rls=org.mozilla:it:official&prmd=imvns&source=lnms&tbm=isch&ei=IJ1dT-SSF9HoOYWWlfYM&sa=X&oi=mode_link&ct=mode&cd=2&ved=0CBQQ_AUoAQ&biw=1330&bih=647

Another interesting collection of E.M. images of different nuclei is found **here**

Electron Microscopic Atlas of cells, tissues and organs in the internet

Deutsche Version
dieser Seite

Editor:
Dr. med.
H. Jastrow

Every attempt was made to provide correct information and labelling, however any liability for eventual errors or incompleteness is rejected!

Do macroscopic differences reflect different (macro)-molecular organization ?

Open Access

Research

Tissue-specific spatial organization of genomes

Luis A Parada[*], Philip G McQueen[†] and Tom Misteli[*]

Addresses: [*]National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. [†]Mathematical and Statistical Laboratory, Division of Computational Biology, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA.

Correspondence: Tom Misteli. E-mail: mistelit@mail.nih.gov

Published: 21 June 2004

Genome ***Biology*** 2004, **5:**R44

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/7/R44

Received: 21 April 2004
Revised: 24 May 2004
Accepted: 25 May 2004

© 2004 Parada *et al.*; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Genomes are organized *in vivo* in the form of chromosomes. Each chromosome occupies a distinct nuclear subvolume in the form of a chromosome territory. The spatial positioning of chromosomes within the interphase nucleus is often nonrandom. It is unclear whether the nonrandom spatial arrangement of chromosomes is conserved among tissues or whether spatial genome organization is tissue-specific.

Results: Using two-dimensional and three-dimensional fluorescence *in situ* hybridization we have carried out a systematic analysis of the spatial positioning of a subset of mouse chromosomes in several tissues. We show that chromosomes exhibit tissue-specific organization. Chromosomes are distributed tissue-specifically with respect to their position relative to the center of the nucleus and also relative to each other. Subsets of chromosomes form distinct types of spatial clusters in different tissues and the relative distance between chromosome pairs varies among tissues.

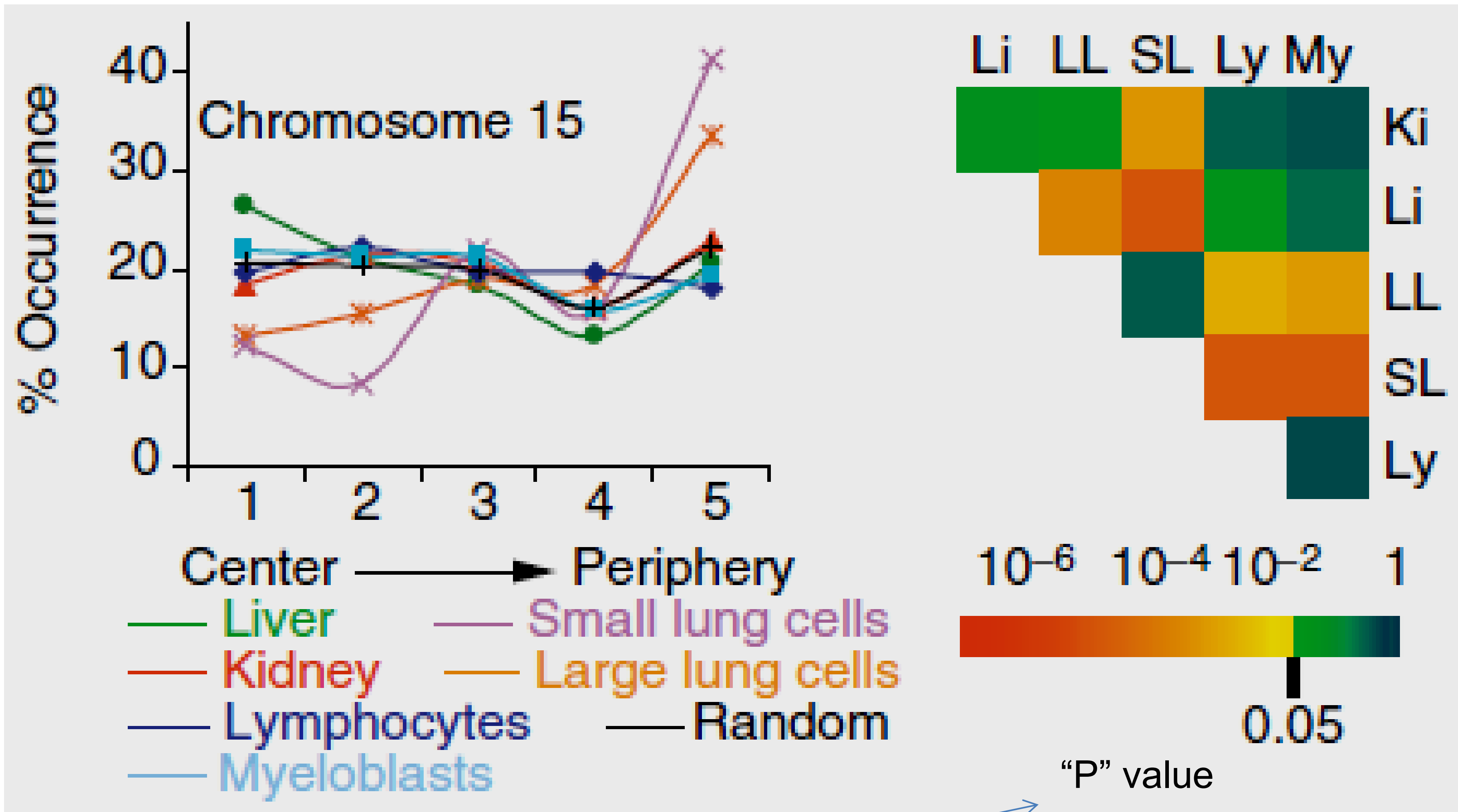
Consistent with the notion that nonrandom spatial proximity is functionally relevant in determining the outcome of chromosome translocation events, we find a correlation between tissue-specific spatial proximity and tissue-specific translocation prevalence.

2D and 3D FISH

Introducing nuclear localization of interphase chromosomes

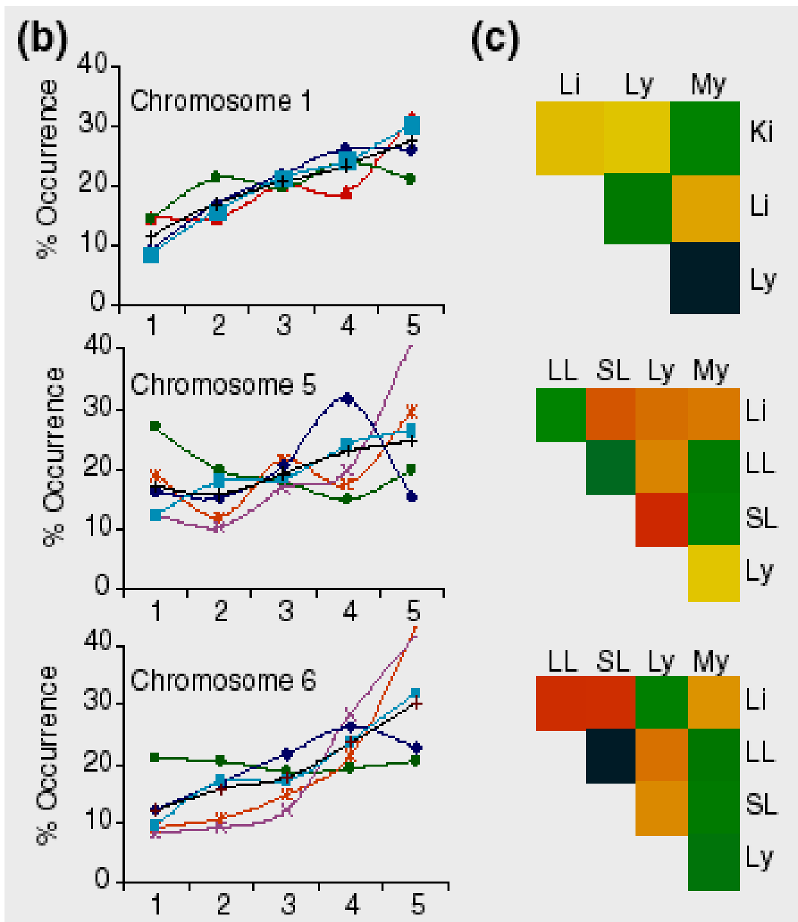
Chromosome 5 painting

From Parada et al., 2004, Genome Biol. 5:R44.

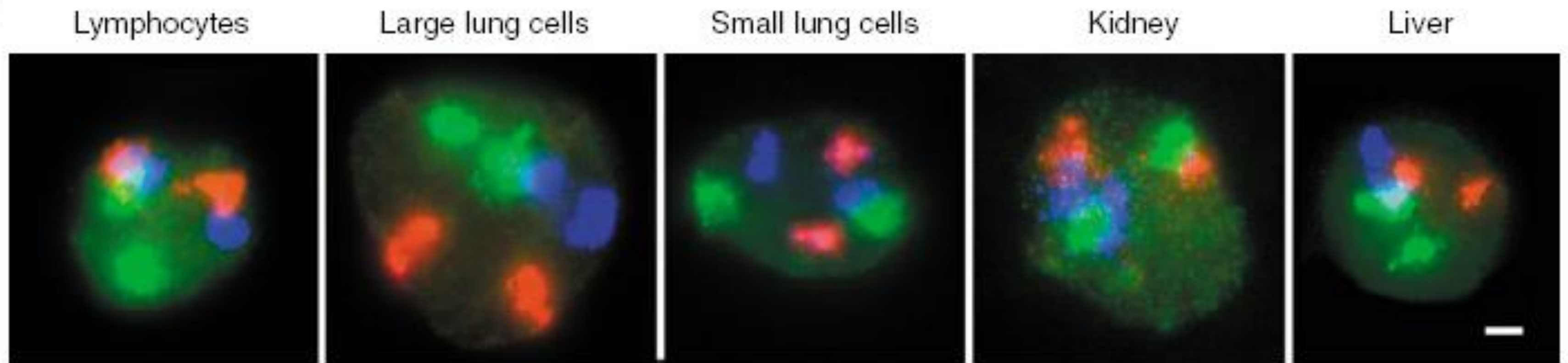


«P-value»

From Parada et al., 2004, Genome Biol. 5:R44.



From Parada et al., 2004, Genome Biol. 5:R44.



Tissue-specific **relative positioning** of chromosomes 12, 14 and 15

From Parada et al., 2004, Genome Biol. 5:R44.

Couple of chromosomes

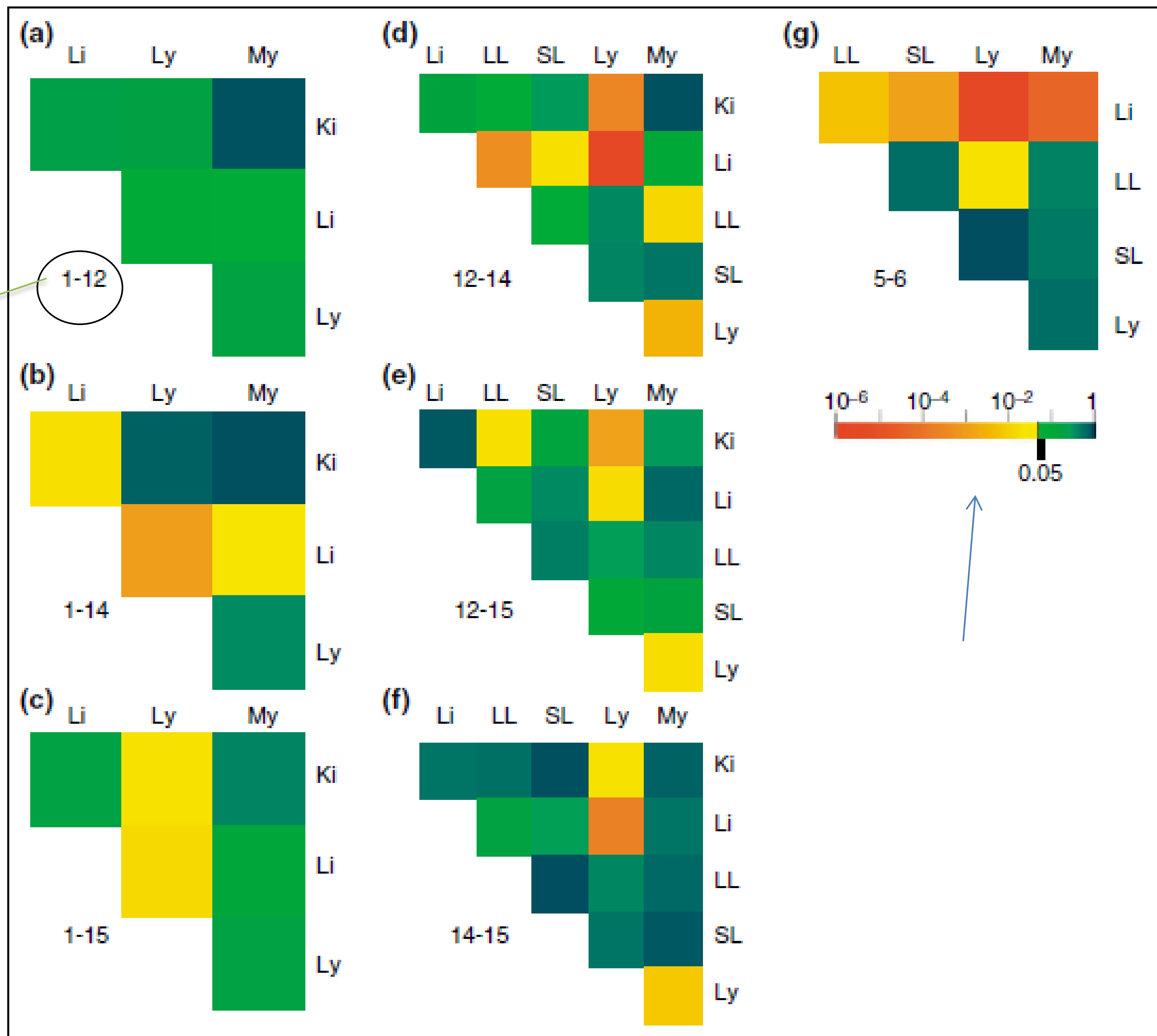


Figure 2

Tissue-specific distances between chromosomes. Average minimum separations between the most proximal pairs of nonhomologous chromosomes were compared pairwise using the Kolmogorov-Smirnov test. **(a)** Chromosome pair 1-12; **(b)** 1-14; **(c)** 1-15; **(d)** 12-14; **(e)** 12-15; **(f)** 14-15; **(g)** 5-6. *p*-values < 0.05 (yellow/red) were considered significant. Abbreviations as in Figure 1. Between 41 and 180 cells were analyzed per tissue.

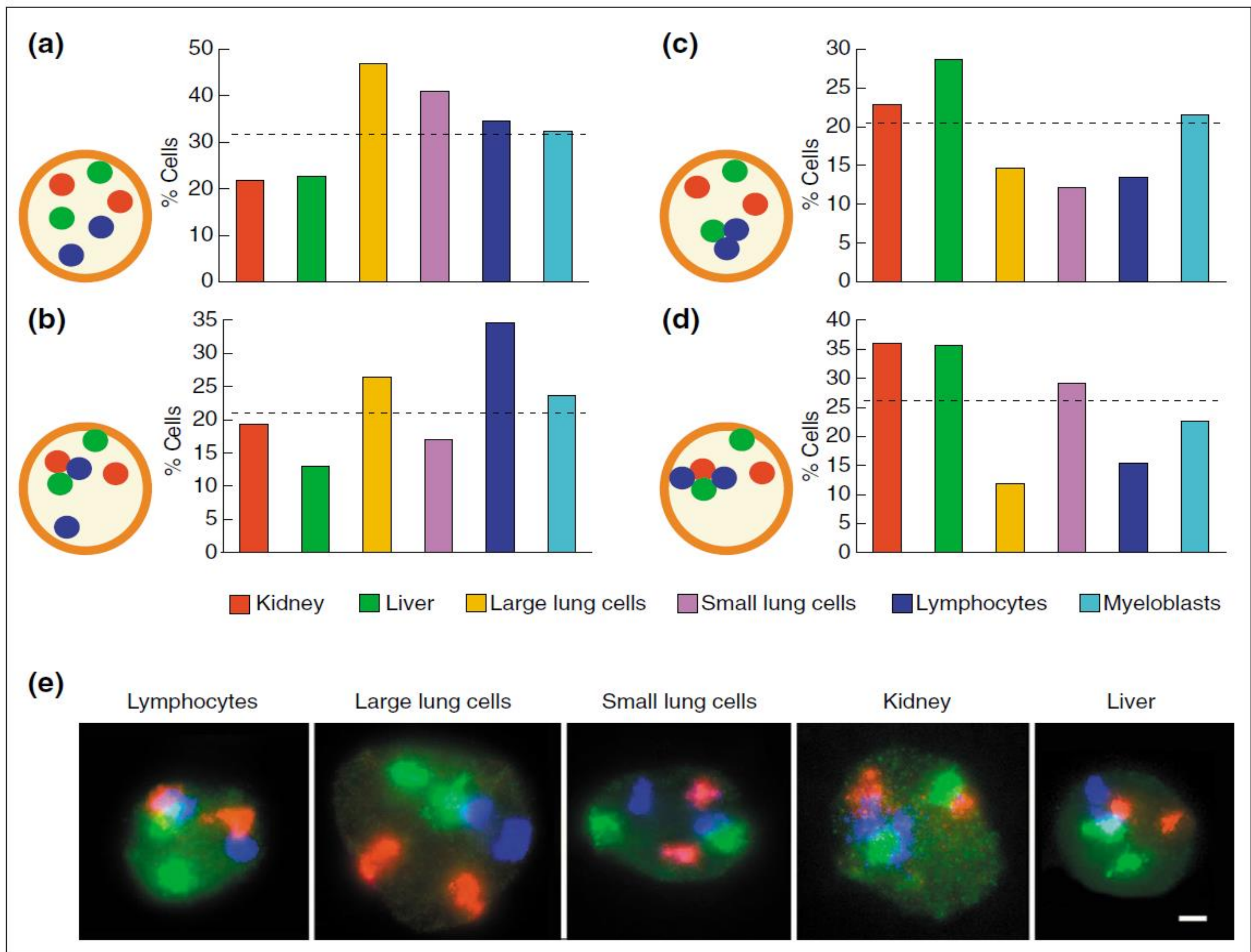


Figure 3
Tissue-specific relative positioning of chromosomes 12, 14 and 15. Quantitation of triplet cluster formation by determining for each tissue type the percentage of cells containing **(a)** no 12-14-15 triplet clusters, **(b)** a single triplet cluster of exactly one chromosome 12, 14 and 15, **(c)** a single cluster of a pair of homologues and one additional chromosome, or **(d)** a cluster of homologues and more than one additional chromosome. Expected values based on random distribution of chromosomes are indicated by a dashed line. Between 41 and 180 cells were analyzed per tissue. **(e)** FISH analysis of different cell types for chromosome 12 (red), 14 (blue), and 15 (green). Distinct preferential cluster types are found in different cell types. Scale bar, 1.8 μm.

Take-home messages:

- Interphase chromosomes occupy discrete “territories” within the nucleus
- Position of interphase chromosomes is cell-specific
- Relative positioning is also cell type-specific

Next Lesson (Wednesday):

1. Reddy’s paper (Your Research Paper No. 1)

2. Start «Chromatin structure»

You should read (at least start reading) your textbook on this subject:

Bannister Kouzarides 2011 PTMs rev