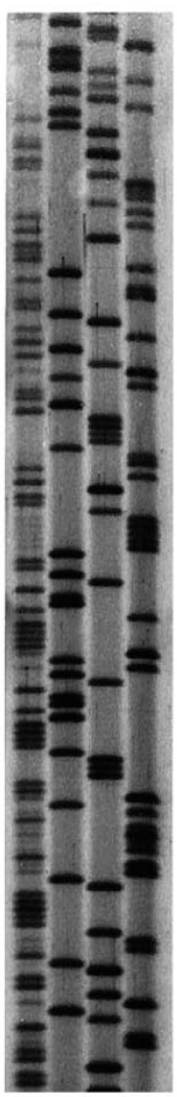
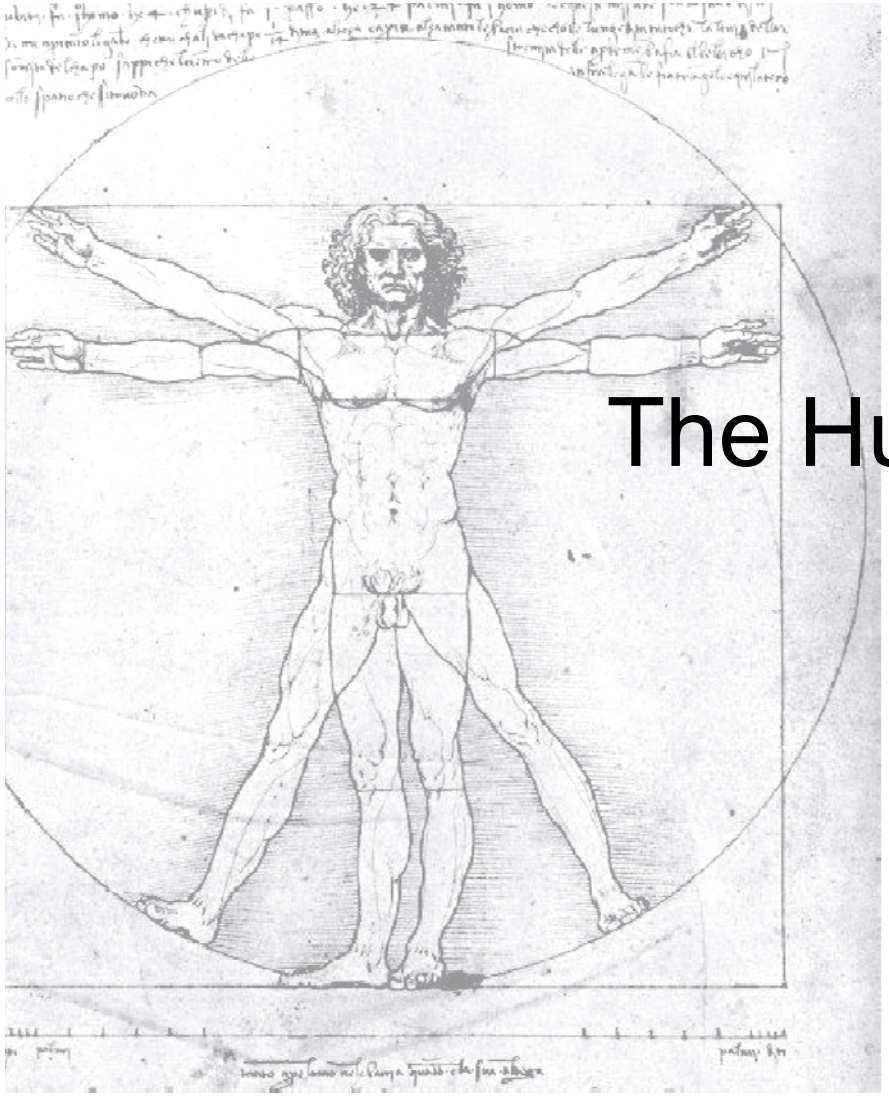




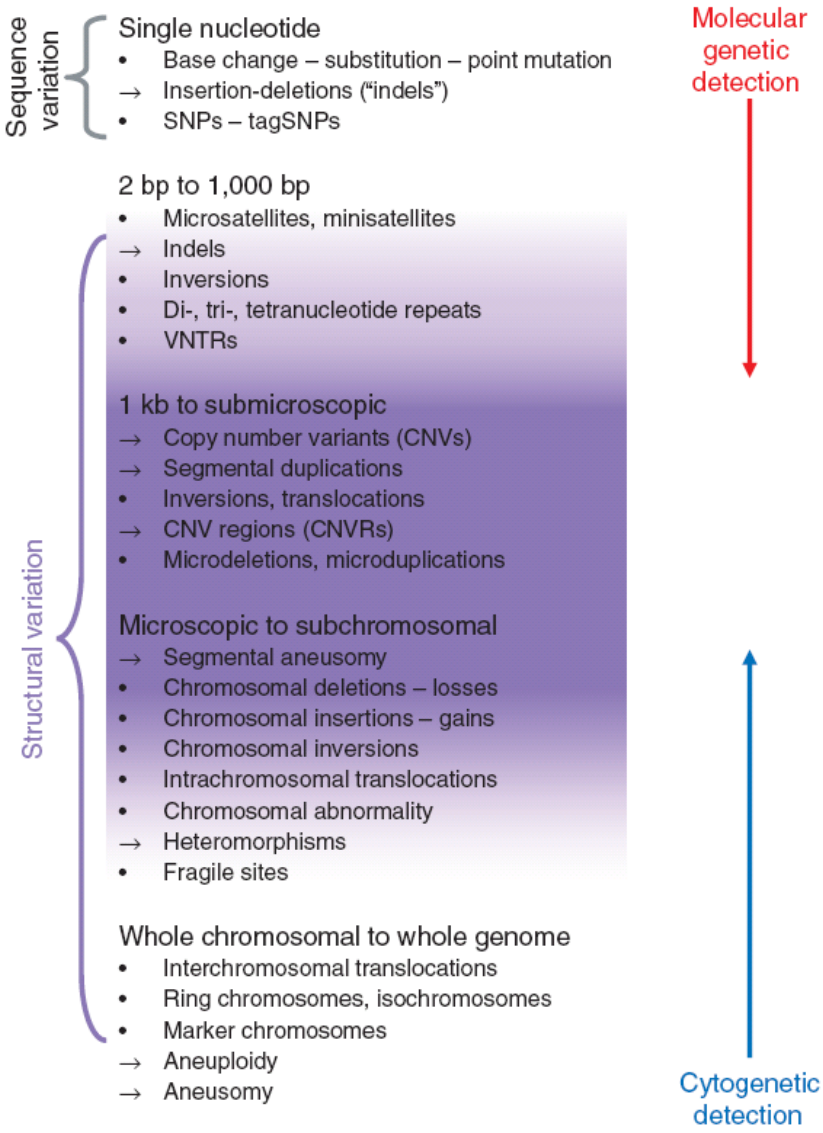
The Human Genome
and its Dynamics

Matthias Platzer

Genome Analysis
Leibniz Institute for Age Research
- Fritz-Lipmann Institute (FLI)

Genetic variation

Lexicon



Scherer et al., Nat Genet Suppl 39:s7 2007

Genetic variation

Terminology

Mutation

= event causing genetic variation

substitution, insertion, deletion, inversion

Polymorphism

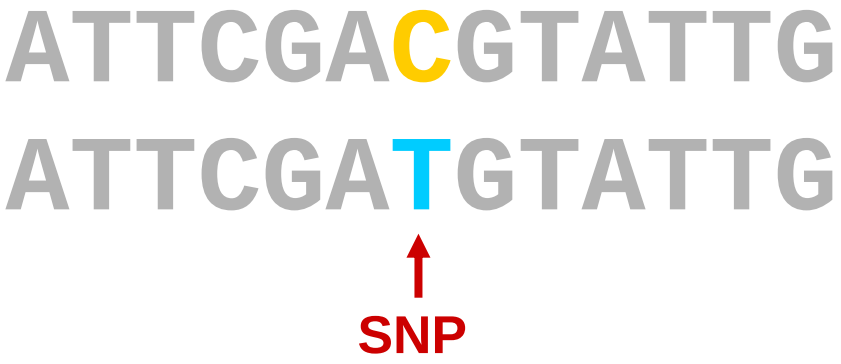
= condition of a variation, when it is established with frequency ≥1% in a population

Mutation *in medical genetics*

= rare variation with a population frequency <1%

Genetic variation

Single Nucleotide Polymorphism (SNP)



- as a rule bi-allelic
- 12 Mio SNPs genom-wide 1/250 bp
- 2 individuals differ in ~300.000 SNPs 1/10.000 bp
- ~5% of SNPs, e.g. 600.000 SNPs with phenotyp (?)
  50-100.000 SNPs with clinical relevance (?)

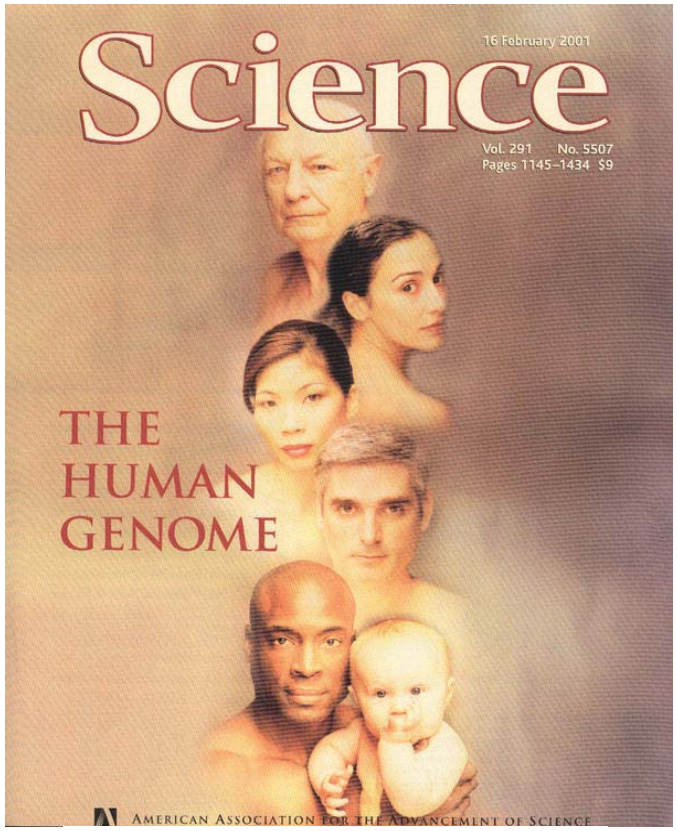
Sequencing of the Human Genome

Publications

2004

2001

International academic consortium



2001

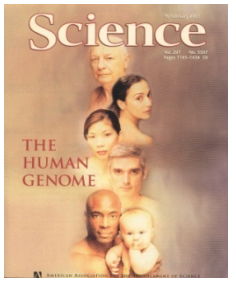
Human Genome

Working Draft versions February 2001

Academic

Initial Sequencing & Analysis…

Private



The Sequence of …

2.72 Gb	**Sequenced Bases**	2.65 Gb
1,000	Clone gaps	54,000
146,000	Sequence gaps	116,000
147.000	**Gaps**	170,000

overall coverage:	**94%**
quality of unfinished data :	< **1 error/10kb** in **91%**

Human Genome

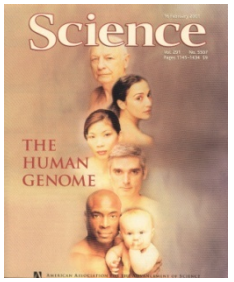
Final version October 2004

Initial Seq…

Academic

Finishing the euchromatic sequence…

Private



The Sequence of …

2.72 Gb	2.85 Gb	**Sequenced Bases**	2.65 Gb
1,000	283	Clone gaps	54,000
146,000	58	Sequence gaps	116,000
147.000	341	**Gaps**	170,000

near-complete sequence:	**99%** of euchromatin
extremely high quality:	**< 1 error/100kb**

Segmental Duplications

Problems of the human reference sequence

~50%

of the

273 interior euchromatic gaps

located in

segmentally duplicated regions

Segmental Duplication

Definition

genomic regions >**1kb**

with nt identity >**90%**

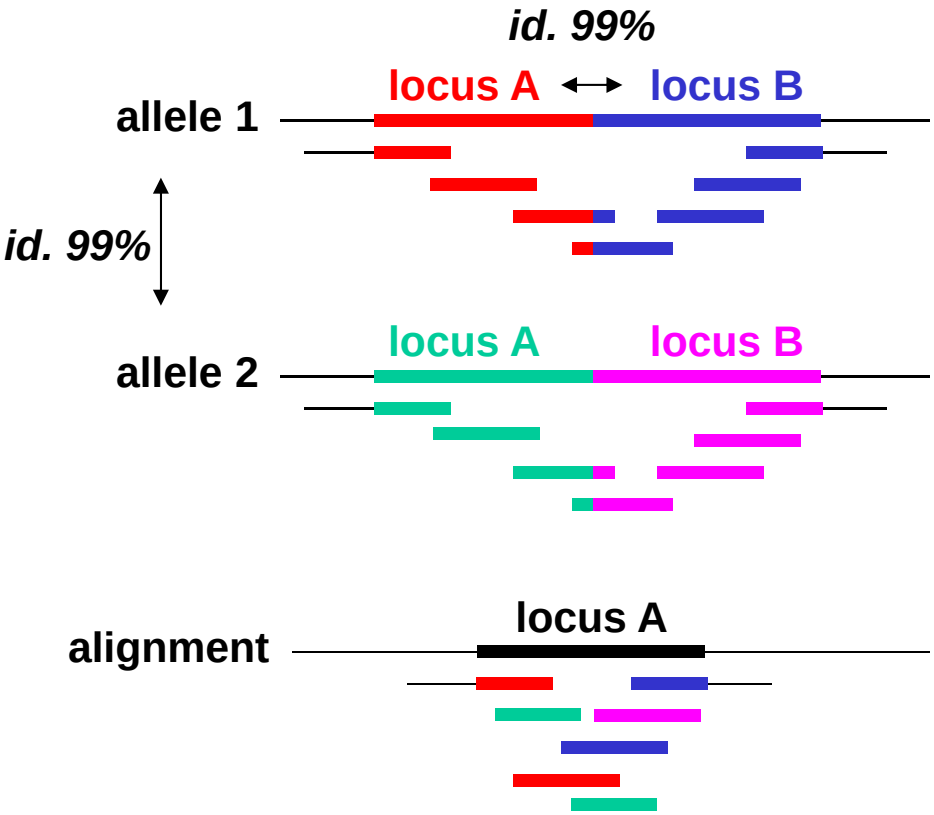
Human genome

5.3% segmentally duplicated

87% of all segmental duplications >**50 kb**

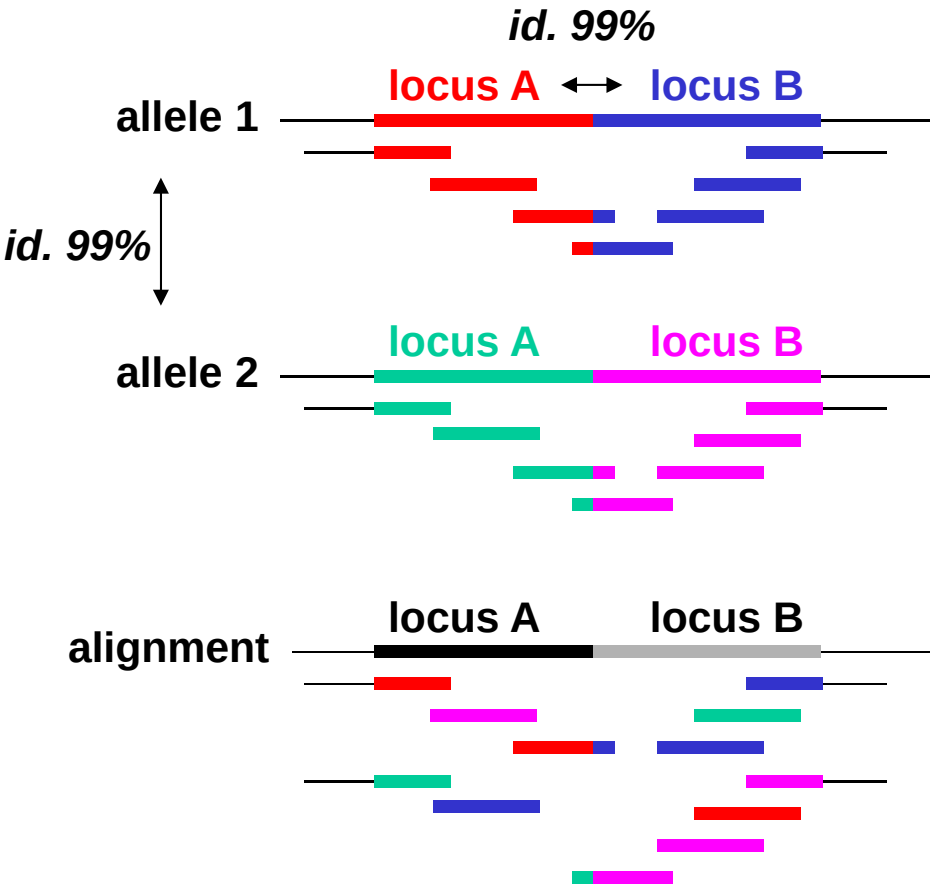
Segmental Tandem Duplications

Assembly problems



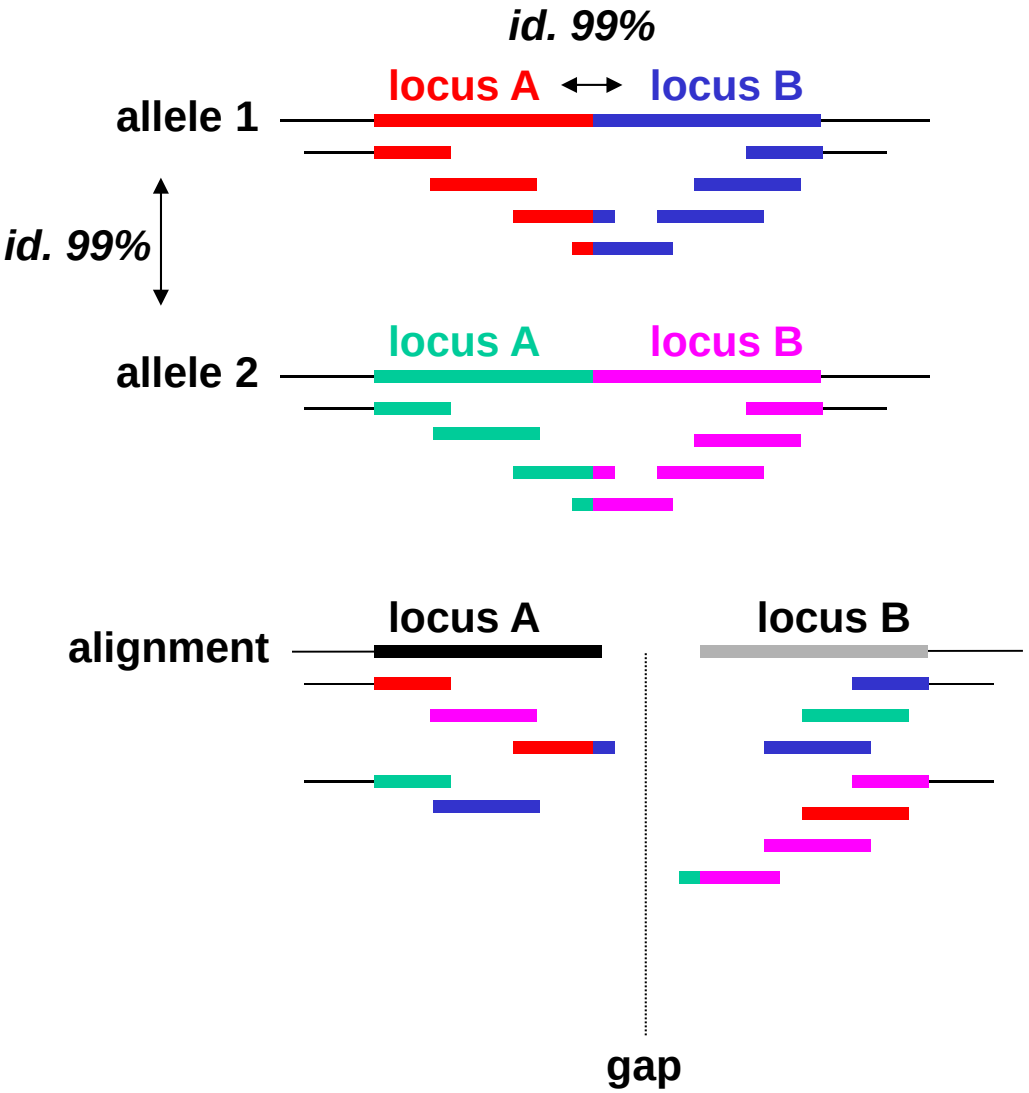
Segmental Tandem Duplications

Assembly problems



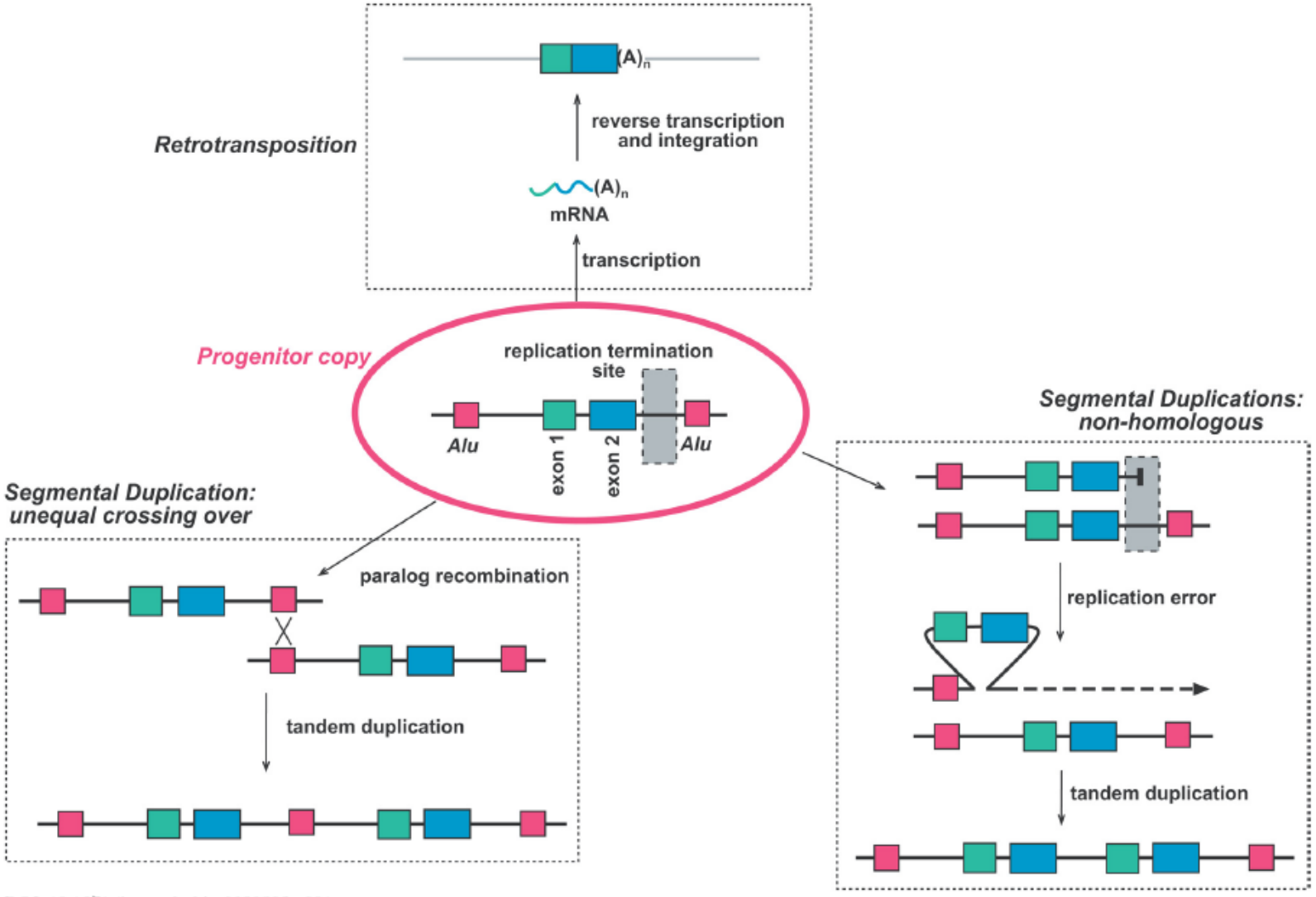
Segmental Tandem Duplications

Assembly problems



Segmental duplications

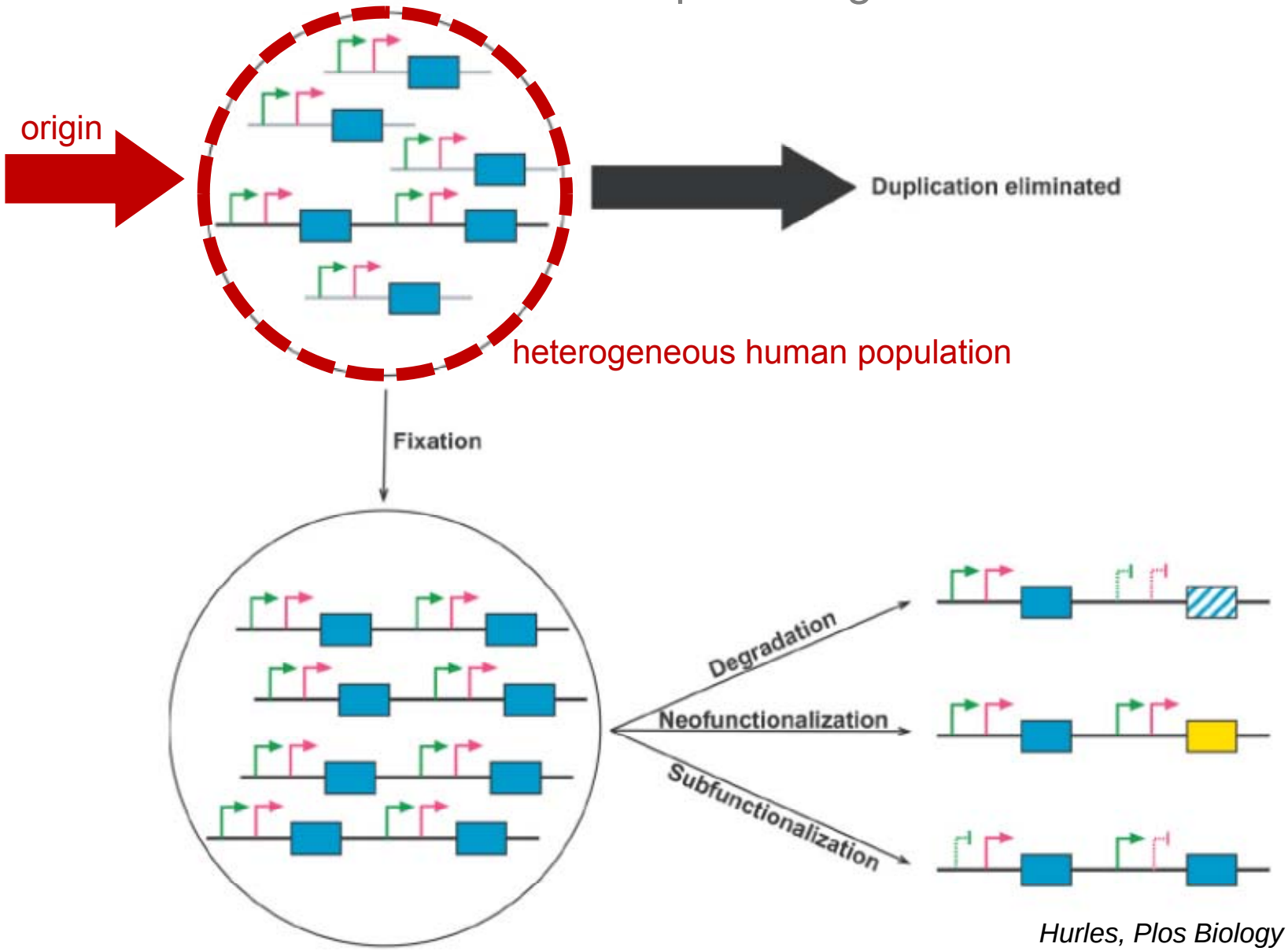
Mechanisms



Hurles, Plos Biology 2:900 (2004)

Segmental duplications

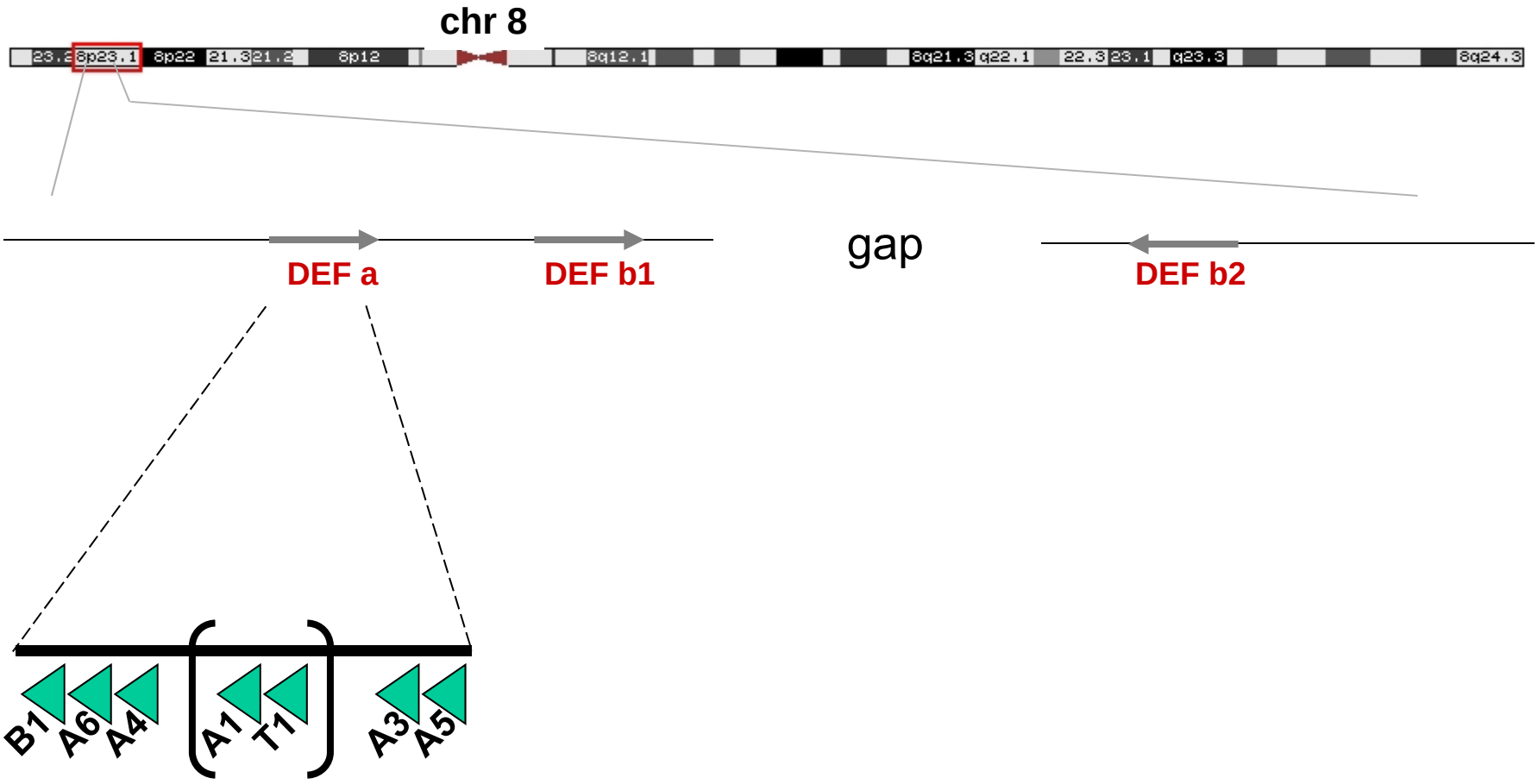
Fate of duplicated genes



Hurles, Plos Biology 2:900 (2004)

DEF cluster at 8p23.1

hg16: 6.3-8.3 Mb



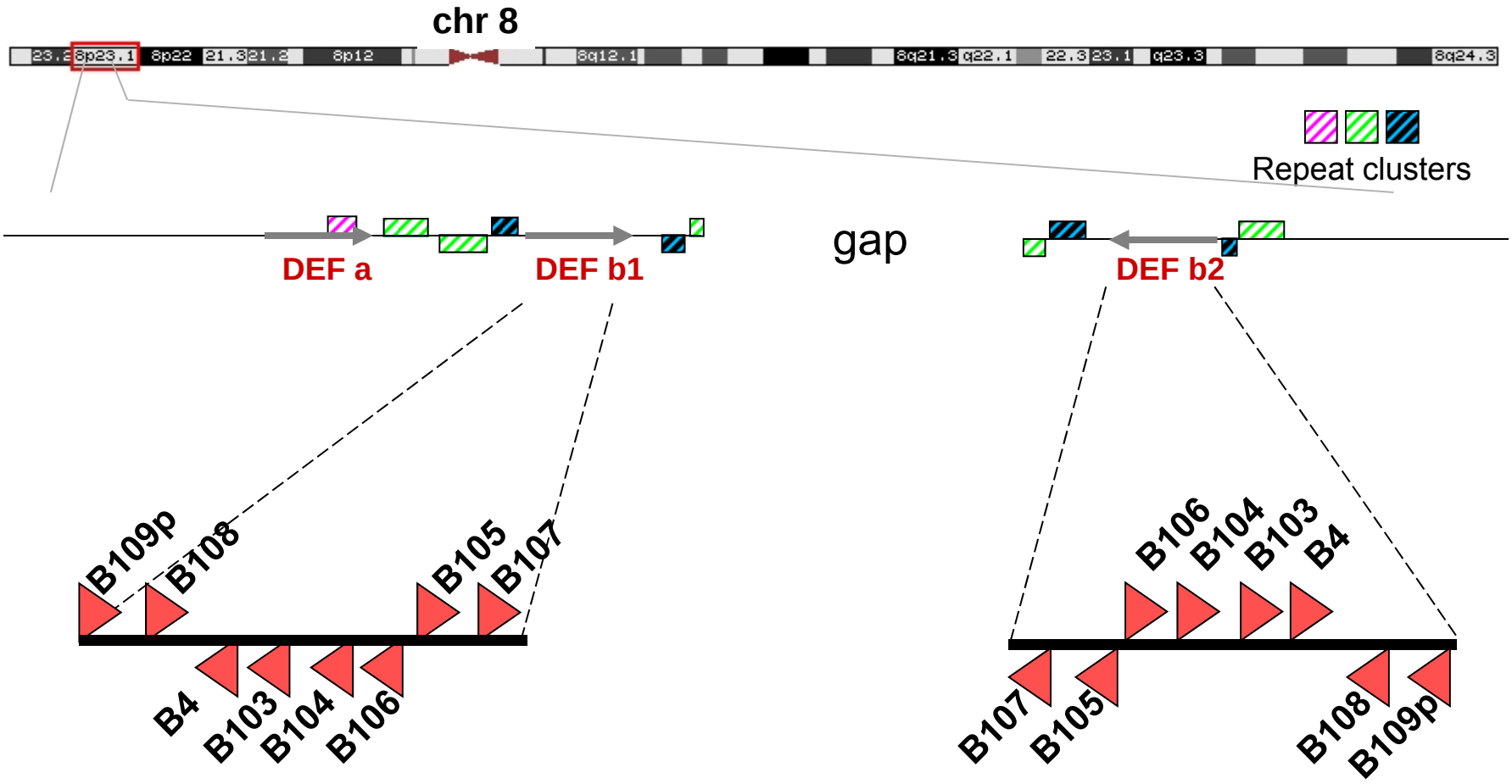
2-13 copies / genome

1-8 copies / chromosome

Taudien et al., BMC Genomics 5:92 (2004)

DEF cluster at 8p23.1

hg16: 6.3-8.3 Mb



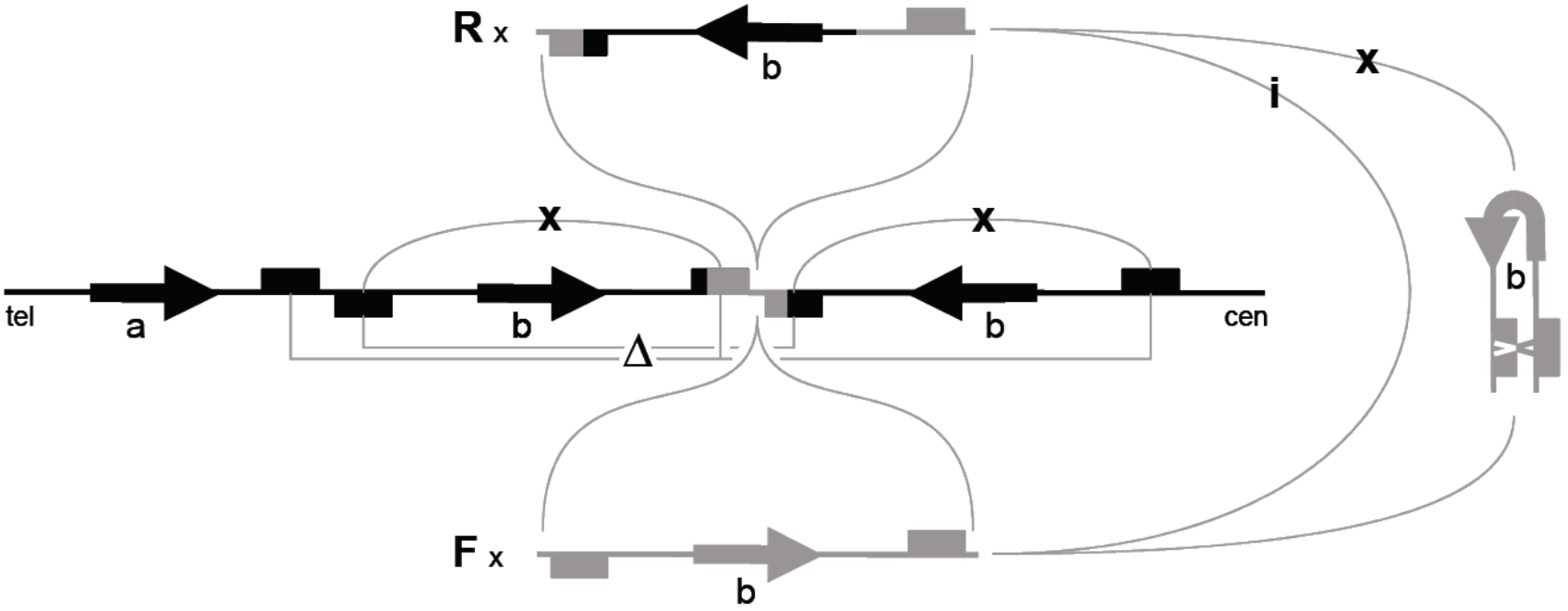
2-12 copies / genome

1-7 copies / chromosome

Taudien et al., BMC Genomics 5:92 (2004)

Genomic variability of 8p23.1 DEF locus

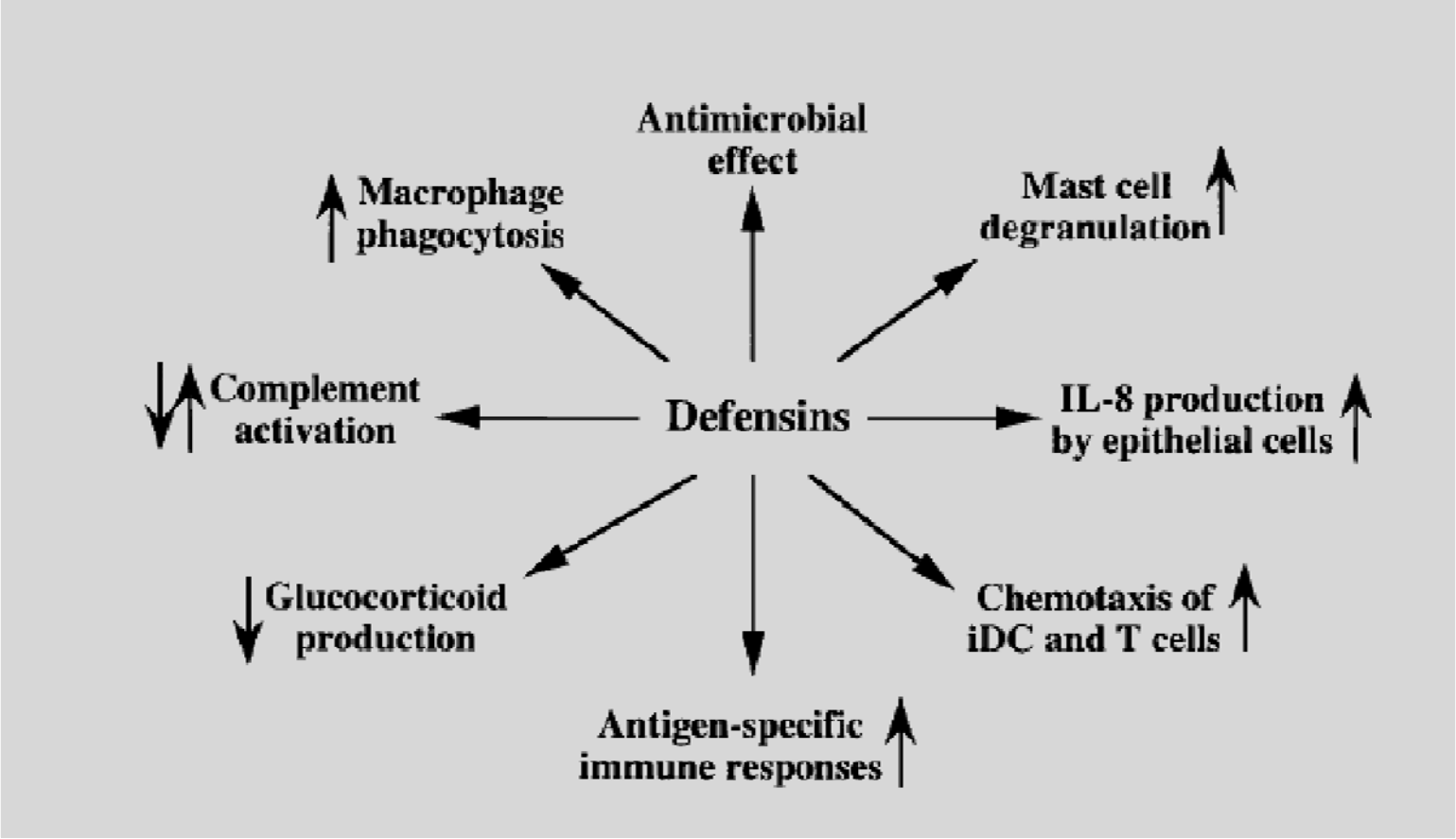
Hypothetical organisation



Taudien et al., BMC Genomics 5:92 (2004)

Defensins (DEF)

Multiple roles



Immunity & Cancer

Yang et al., 2001. Cell. Mol. Life Sci. 58, 978-989

Complex phenotypes / diseases

Structural variations

Gene	Type	Locus	Size (kb)	Phenotype	Copy number
UGT2B17	Deletion	4q13	150	Variable testosterone levels, risk of prostate cancer	0–2
DEFB4	VNTR	8p23.1	20	Colonic Crohn's disease	2–10
FCGR3	Deletion	1q23.3	>5	Glomerulonephritis, systemic lupus erythematosus	0–14
OPN1LW/OPN1MW	VNTR	Xq28	13-15	Red/green colour blindness	0–4/0–7
LPA	VNTR	6q25.3	5.5	Altered coronary heart disease risk	2–38
CCL3L1/CCL4L1	VNTR	17q12	Not known*	Reduced HIV infection; reduced AIDS susceptibilty	0–14
RHD	Deletion	1p36.11	60	Rhesus blood group sensitivity	0–2
CYP2A6	Deletion	19q13.2	7	Altered nicotine metabolism	2–3

*Precise boundaries of the copy-number variant are not known.
VNTR, variable number tandem repeats.

Eichler et al., Nature 447:161 (2007)

Complex phenotypes / diseases

Structural variations

FCGR3 copy number & glomerulonephritis in humans and rats

Nature 439:851 (2006)

Strong association of *de novo* copy number mutations with autism

Science 316: 445 (2007)

Segmental duplications

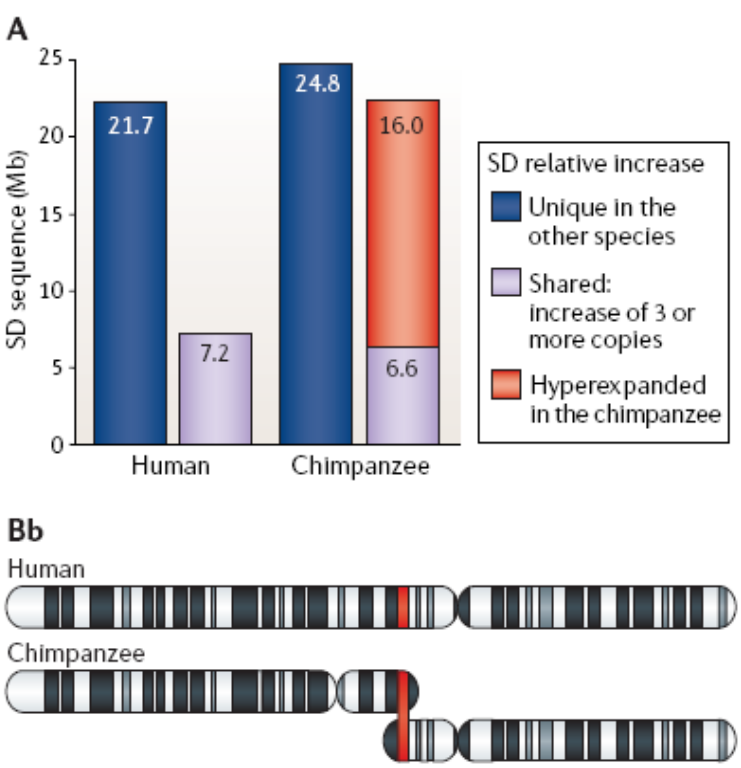
Content of sequenced animal genomes

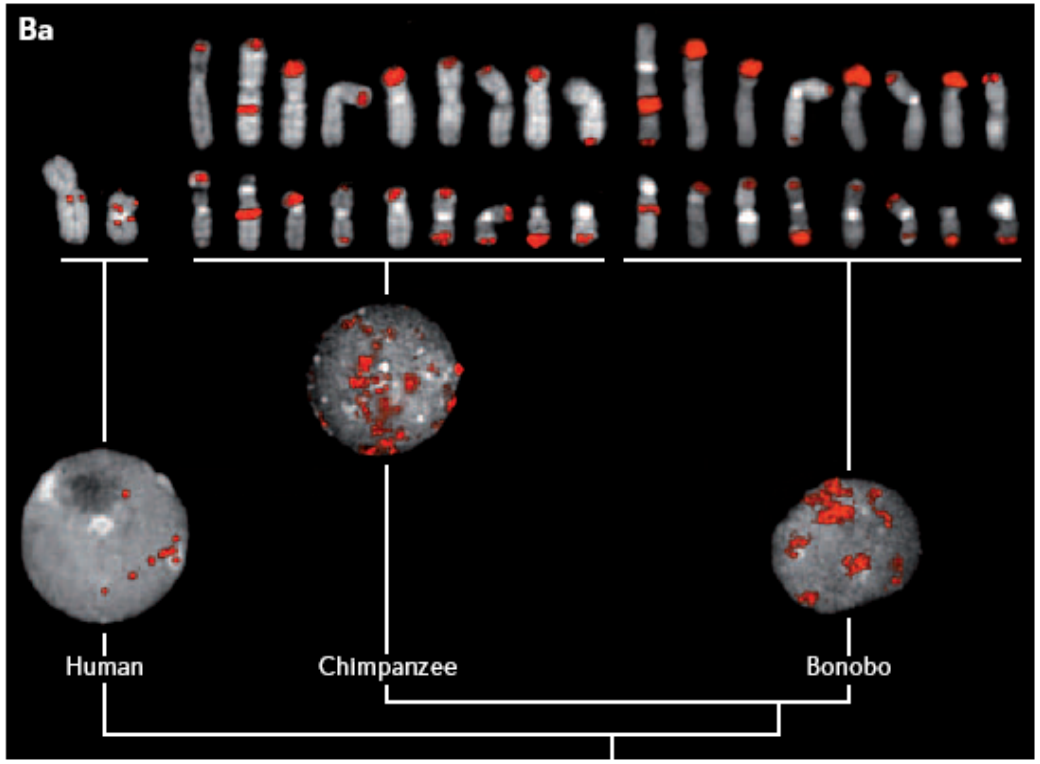
	Caenorhabditis elegans	*Drosophila melanogaster*	Human	Mouse	Rat	Chicken	Chimpanzee*
SDs of >1 kb	4.3%	1.2%	5.2%	2.7%	1.6%	2.7%	N.D.
SDs of >10 kb	0.7%	0.1%	4.5%	2.2%	1.5%	0.3%	N.D.
SDs of >20 kb	N.D.	N.D.	4.0%	1.7%	0.9%	0.0%	~4.8%
Genome size	97	123	2,866	2,506	2,566	1,040	2,866

Bailey & Eichler, Nat Rev Genet 7:552 (2006)

Segmental duplication content of hominoids

Hyperexpansions in chimpanzee

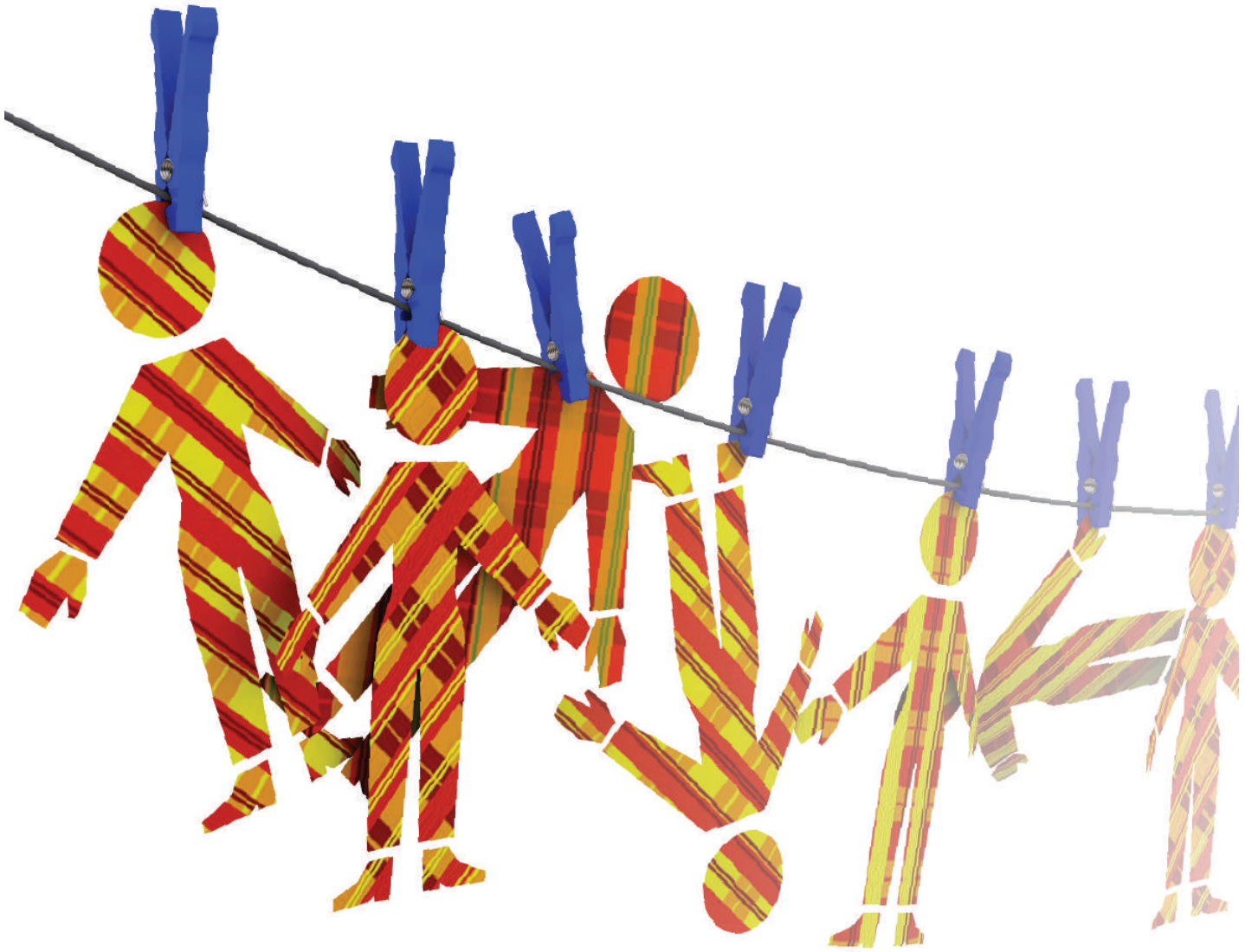




Bailey & Eichler, Nat Rev Genet 7:552 (2006)

Genome Dynamics

Patchwork people ?



News Feature, Nature 437:1084 (2005)

Conclusions

Genomes of any two individuals in the human population differ more at the structural level than at the nucleotide sequence level.

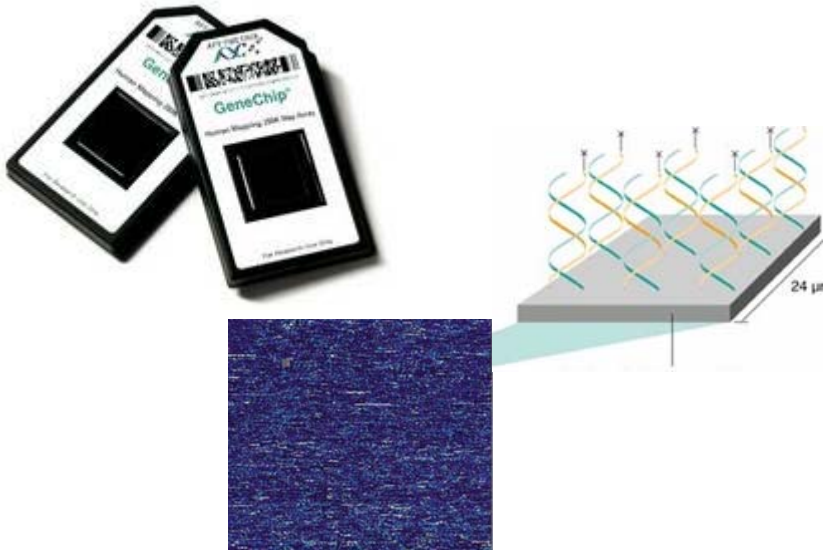
Differences between individuals

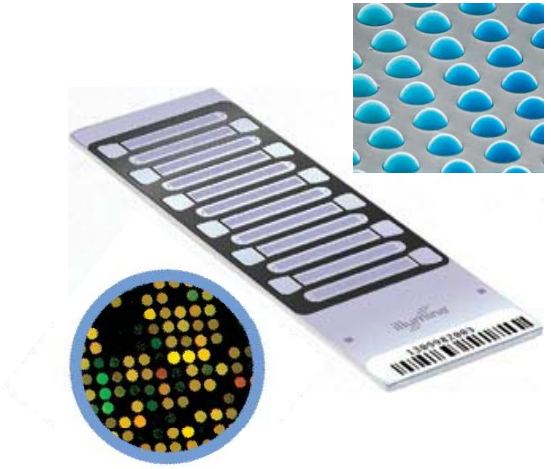
- CNV: >4 Mb >1/800 bp **> 0.12 %**
- SNP: 2.5 Mb 1/1,200 bp **0.08 %**

Sebat, Nat Gen Suppl 39:s3 (2007)

High-throughput SNP genotyping

High-throughput array-based genotyping





Affymetrix

Human SNP Array 6.0

>1.8 million markers

906,600 SNPs

946,000 for CNVs

Illumina

Human 660W-Quad BeadChip

2.6 million markers / four samples

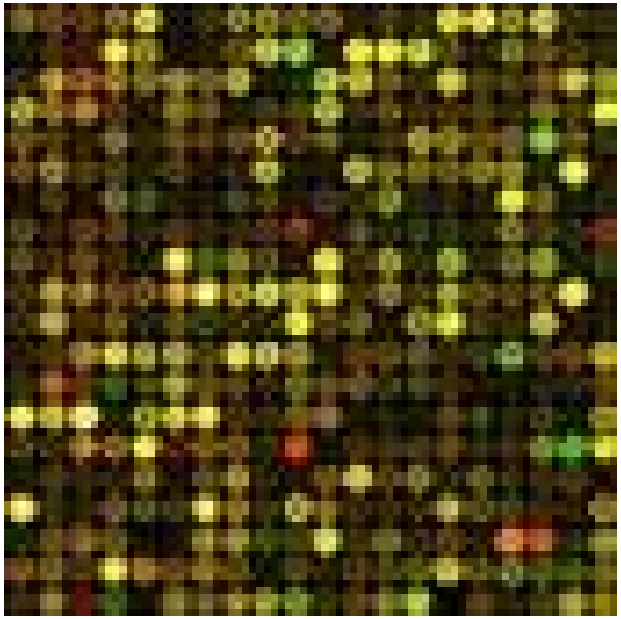
550,000 tag SNPs

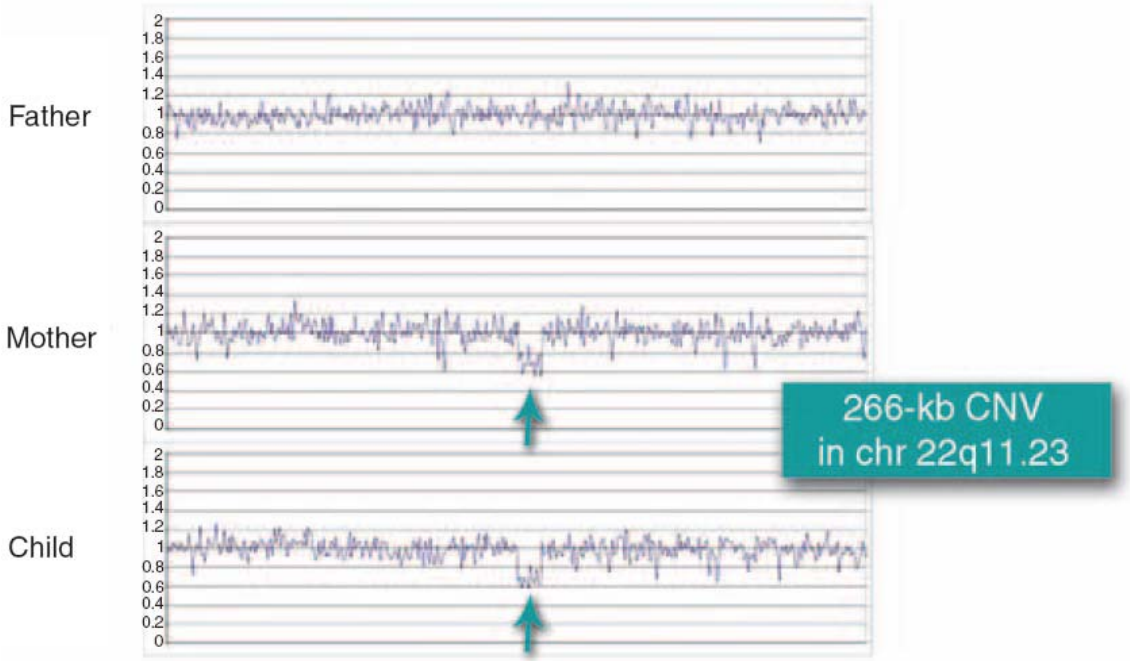
100,000 for CNVs

5,000 common CNVs

Copy number variation (CNV)

Detection by DNA microarrays





- 0.5-2 Mio data points
- comparative hybridization vs. a reference

Lee et al., Nat Genet Suppl 39:s48 (2007)

Genome analysis

DNA sequencing platforms

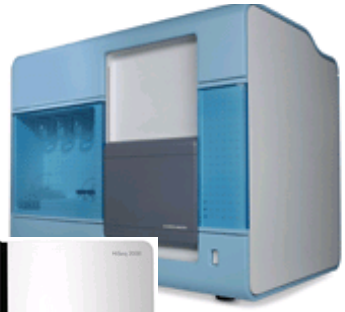
ABI 3730xl
4/2004 & 6/2006

1 Mb/day, 850 nt reads

Roche/454 GS FLX
12/2006

800 Gb/23h, 800 nt reads

Illumina/Solexa GAIIx
12/2008 & 11/2009



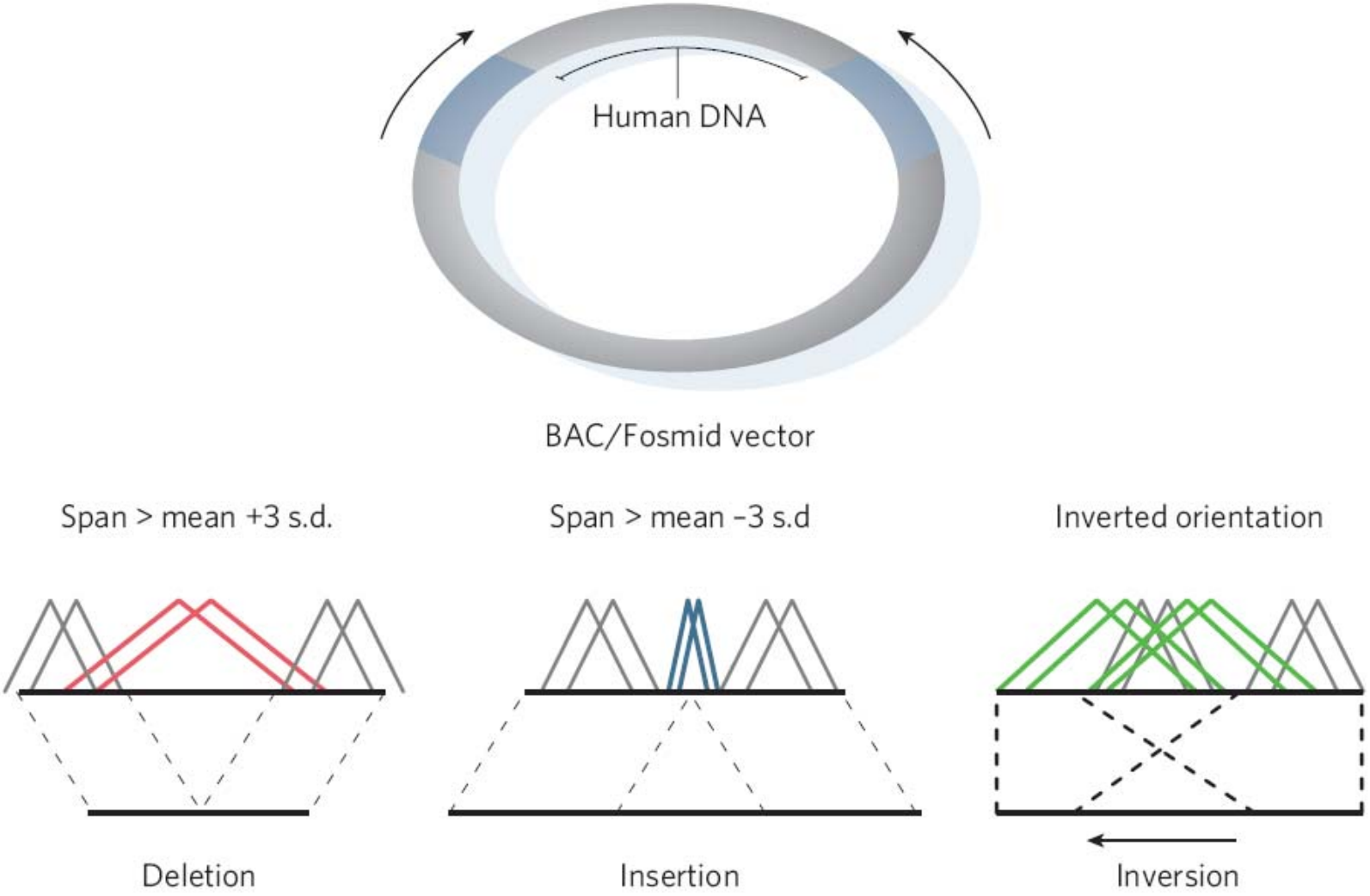
80 Gb/14d, 2x150 nt reads

HiSeq 2000
12/2010

600 Gb/10d, 2x100 nt reads

Completing the map of human genetic variation

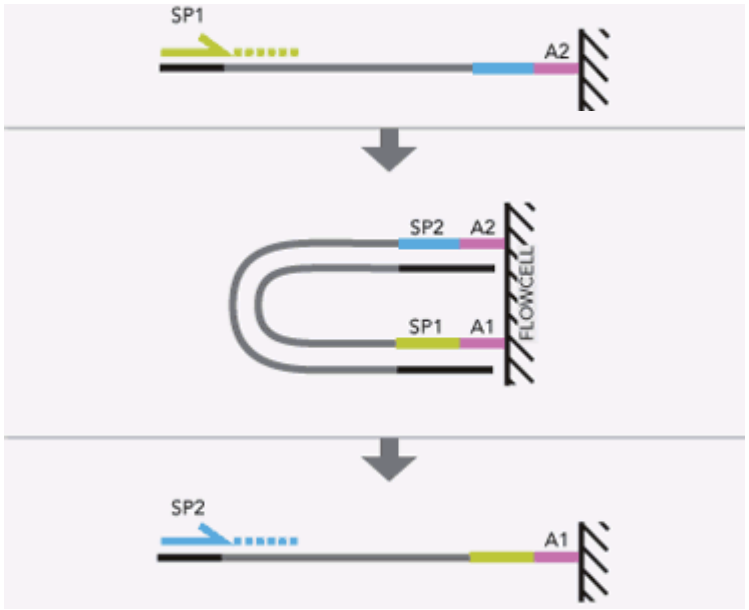
Mapping structural variations

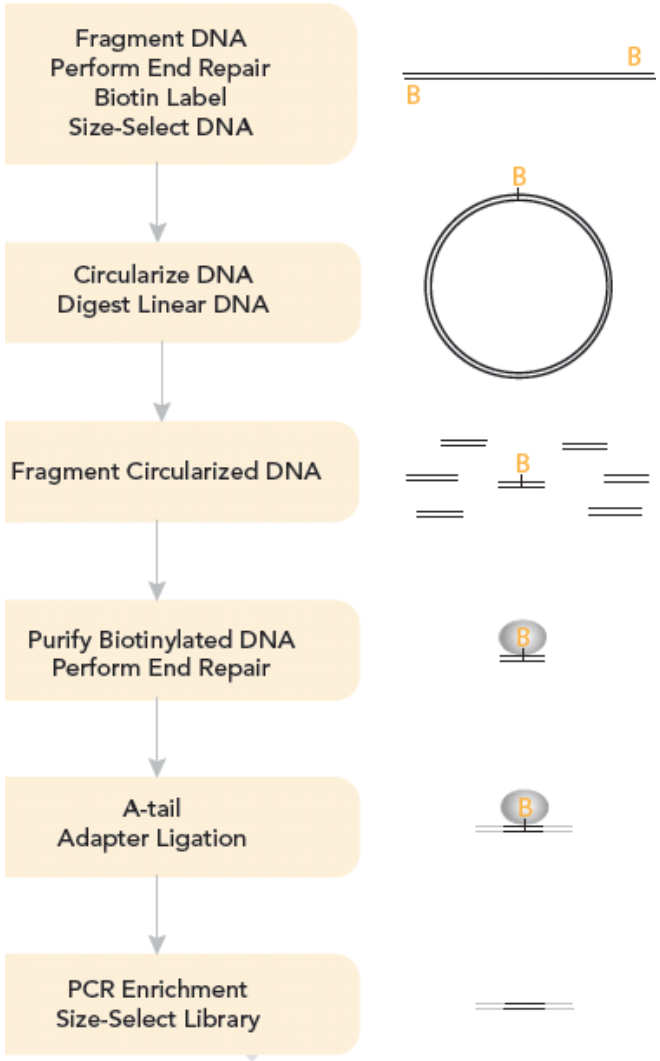


Eichler et al., Nature 447:161 (2007)

Illumina/Solexa

Paired ends & Mate pairs





Human Genome Resequencing

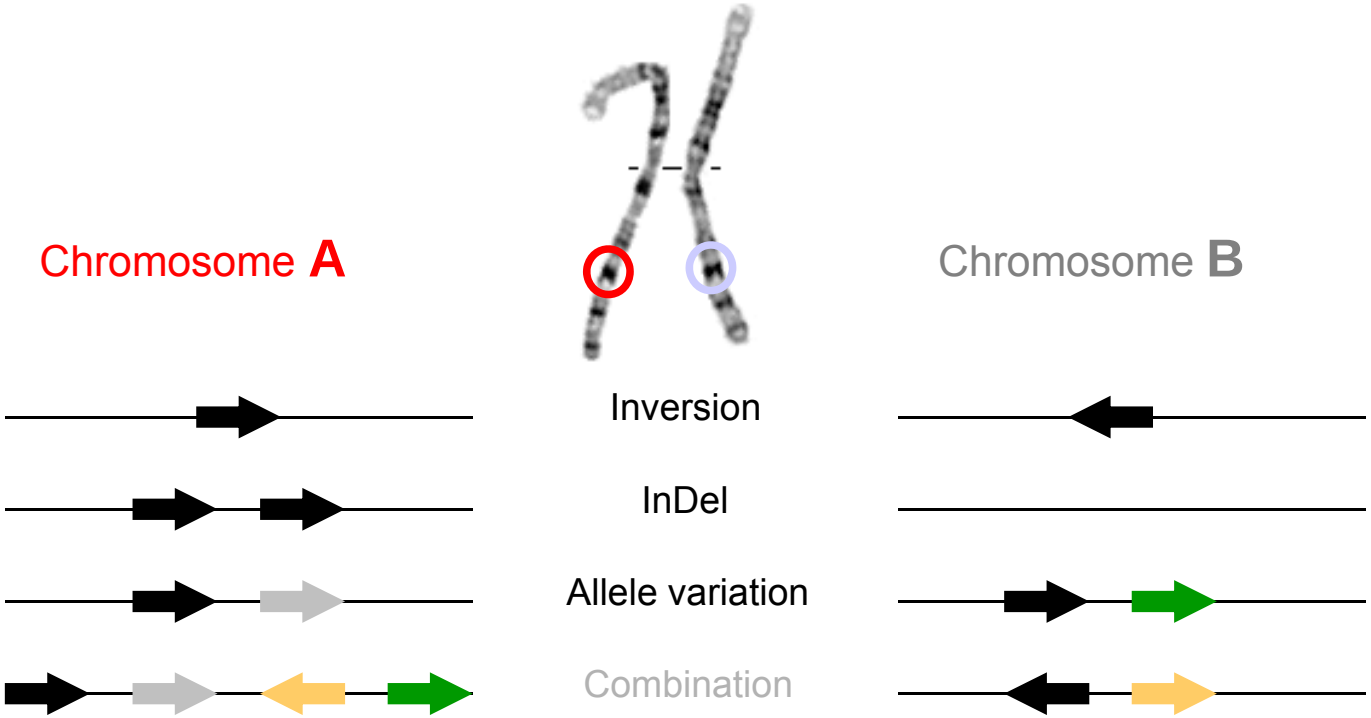
- launched by mid-2010

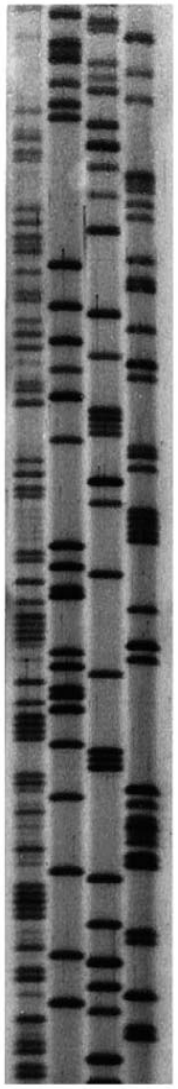
- 2010: ~600 human genomes
- 2011: ~4,000 genomes

- end 2011: 800-1200 genomes/month
- mid 2012: new machines with 6 genomes/day
- end 2012: <3,000 $/genome

Genetic Variability

Structural variations





genome.fli-leibniz.de
Lectures