

# Das Human-Genom und seine sich abzeichnende Dynamik

Matthias Platzer

Genomanalyse  
Leibniz Institut für Altersforschung  
- Fritz-Lipmann Institut (FLI)

A C G T

# Humangenom - Sequenzierung

1. Methode

2. Ergebnisse Biologie

3. Anwendung Medizin

# Humangenom - Sequenzierung

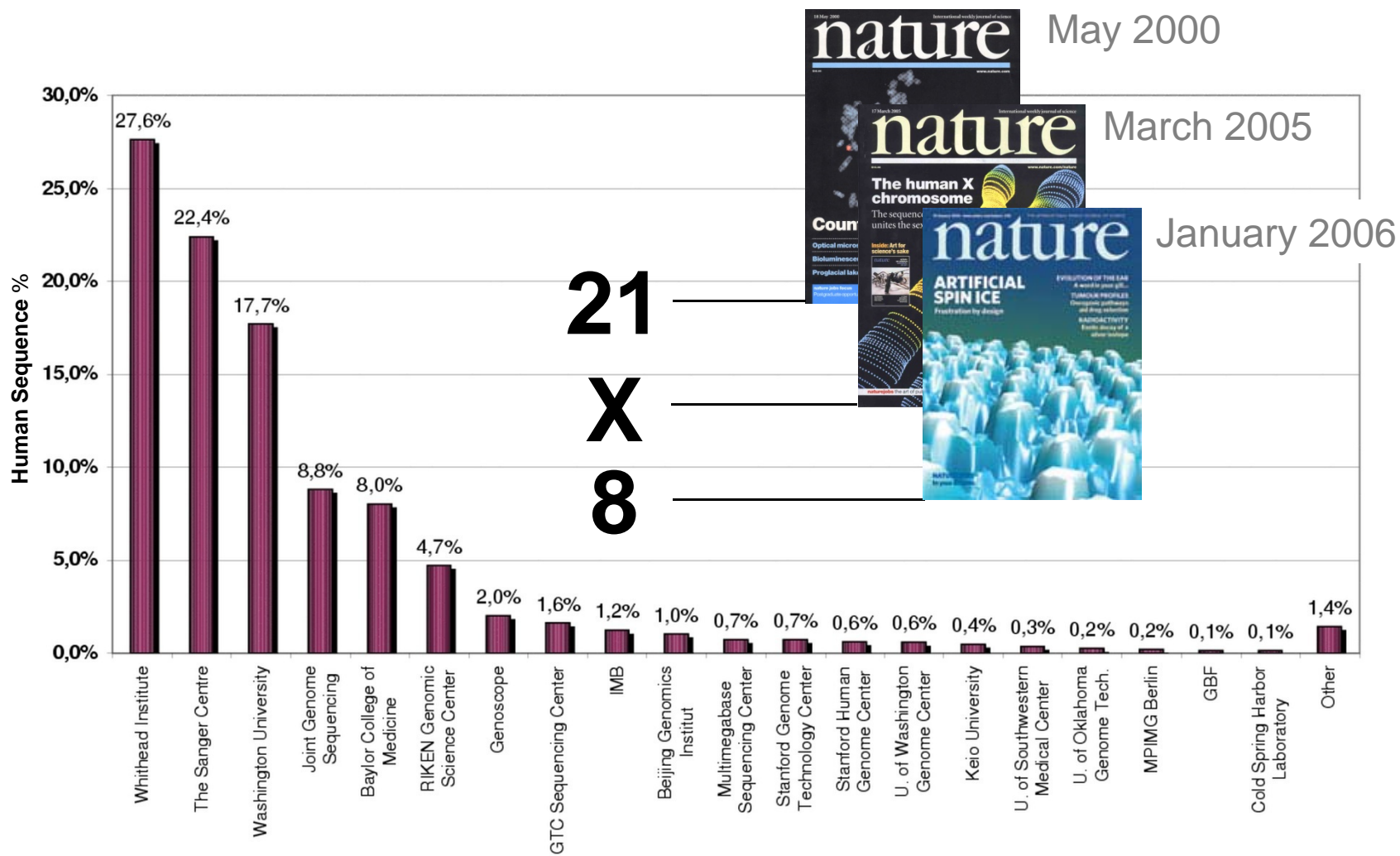
**1. Methode**

**2. Ergebnisse Biologie**

**3. Anwendung Medizin**

# Sequenzierung des Human-Genoms

## Beiträge der Teilnehmer



# Sequenzierung des Human-Genoms

---

## Herausforderungen

### Größe

25x größer als bisher größtes sequenziertes Genom

8x größer als alles bisher Sequenzierte

### Gehalt an sich wiederholenden Sequenzen

erstes wiederholungsreiches Genom, ca.40% erwartet

Ackerschmalwand 11%, Fadenwurm 7%, Fruchtfliege 3%

### Medizinische Relevanz

gesellschaftliche und persönliche Konsequenzen, ethische Brisanz

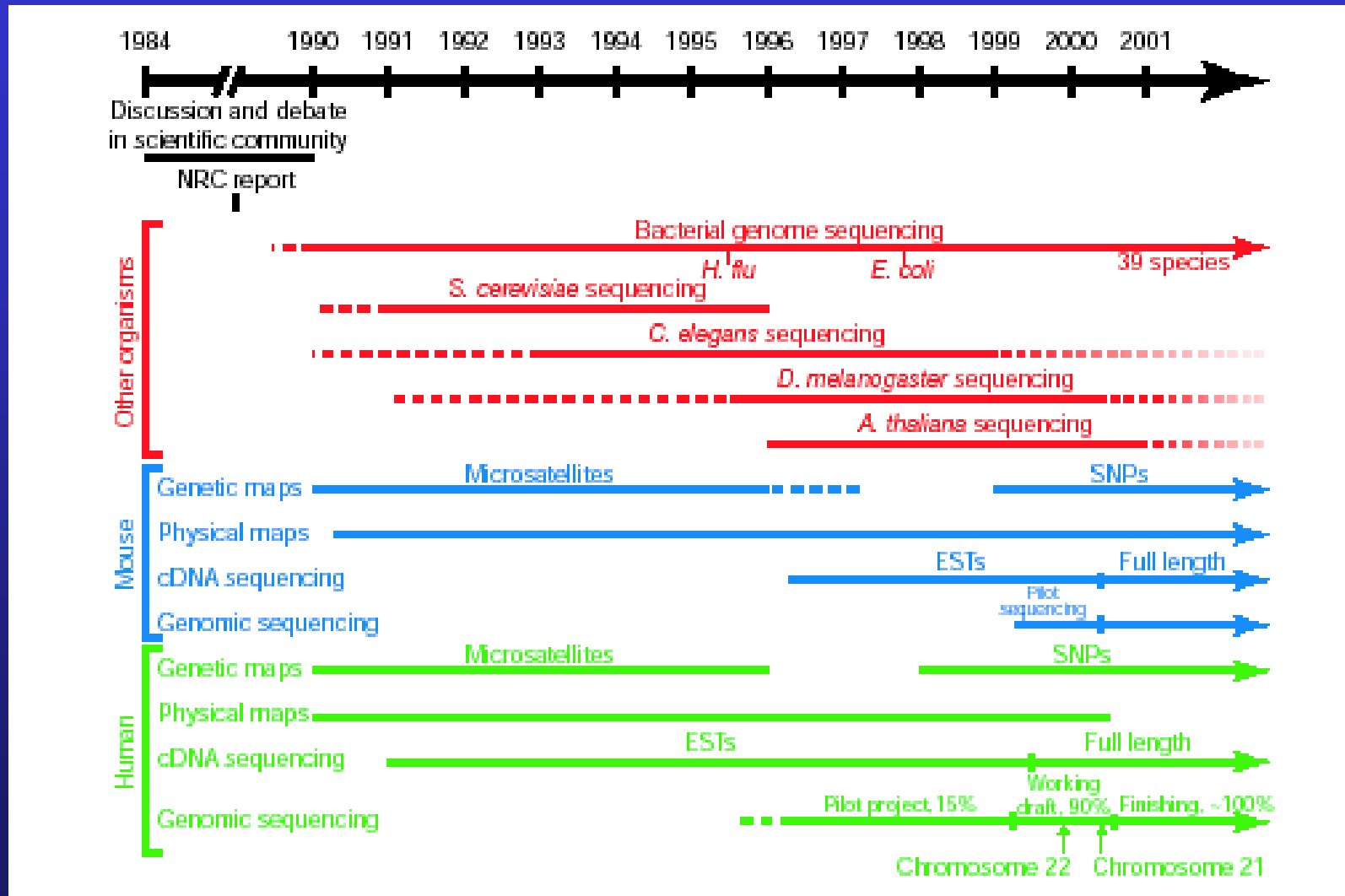
### Internationales Konsortium

20 Zentren aus 6 Ländern

privatwirtschaftliche Konkurrenz 1998-2001

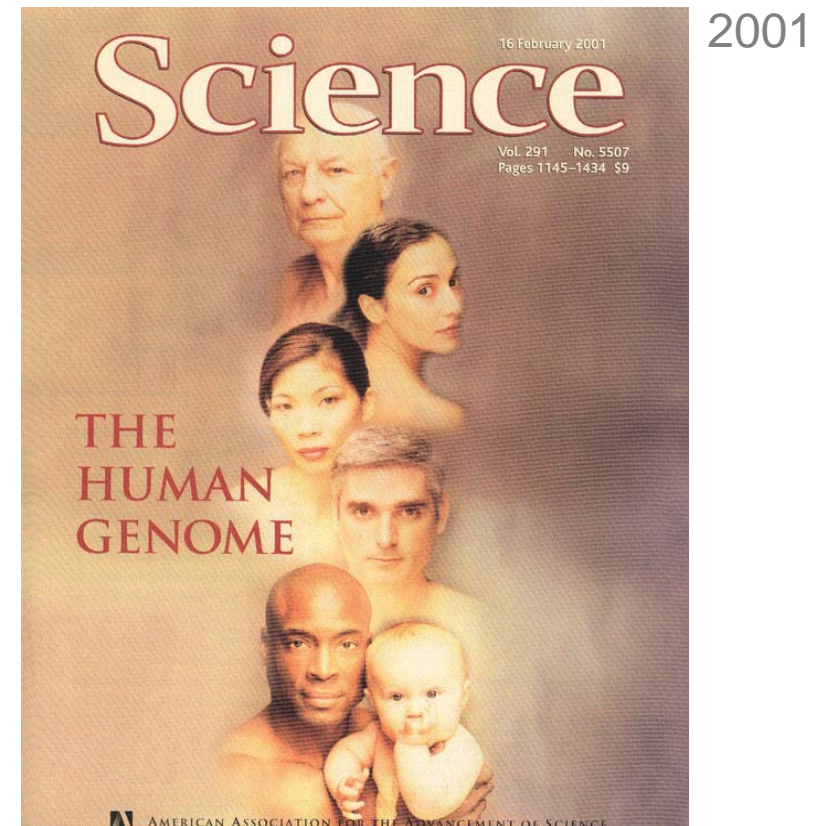
von 15% auf 90% in 15 Monaten (Februar 99 - Juni 2000)

# Sequenzierungen bis 2001



# Sequenzierung des Human-Genoms

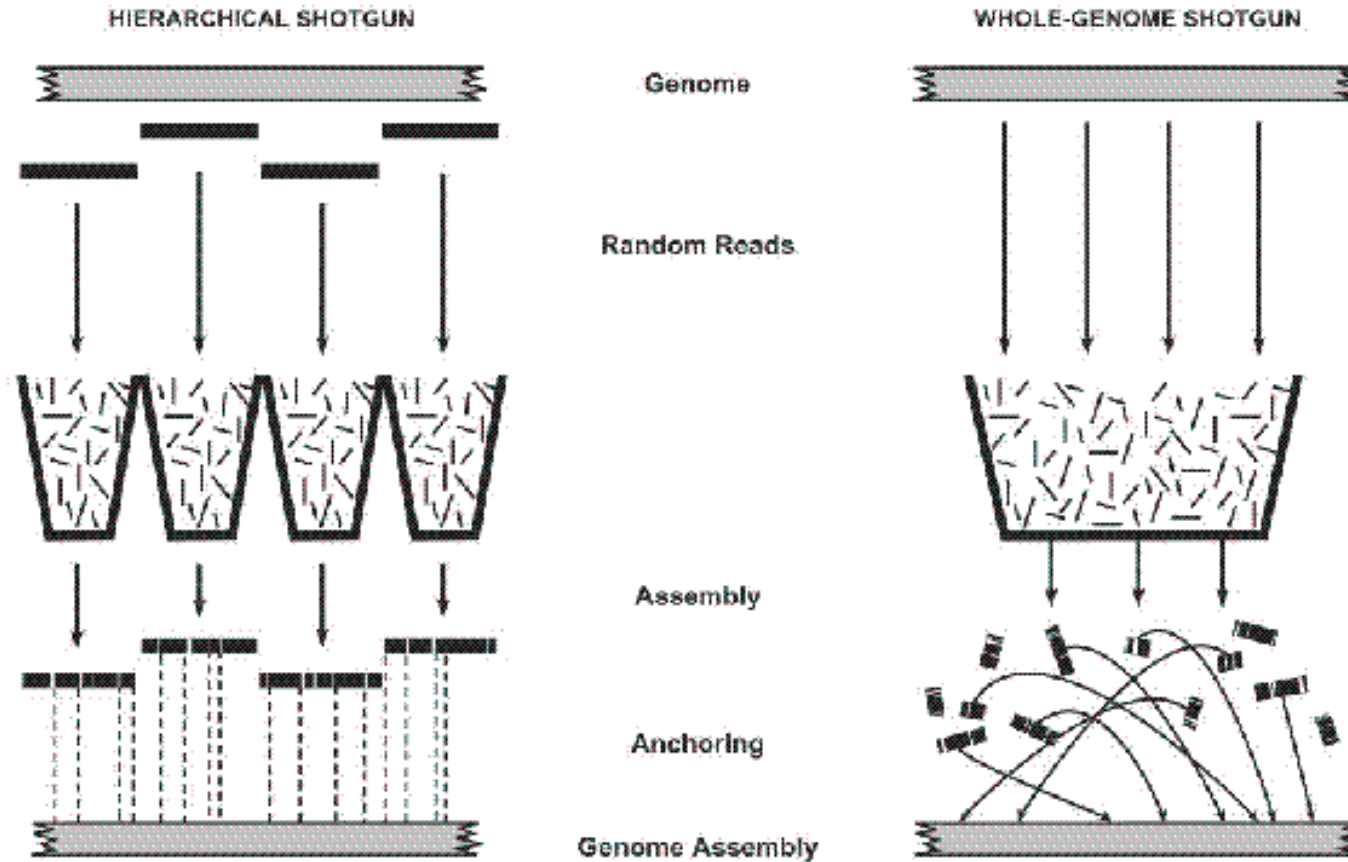
## Publikationen



Internationales akademisches  
Konsortium

# Schrotschuß-Strategien

hierarchisch vs. gesamt-genomisch



2.500 Teile

*Puzzle*

60.000.000 Teile



# 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 population heterogeneity:

**1 SNP/kb**

variation between two individuals:

**1 SNP/10kb**

# Human Genome

Final version October 2004



Initial Seq...



Finishing the euchromatic sequence...

*Academic*

*Private*



The Sequence of ...

		<b>Sequenced Bases</b>	
2.72 Gb	2.85 Gb		2.65 Gb
1,000	283	Clone gaps	54,000
146,000	58	Sequence gaps	116,000
147,000	341	<b>Gaps</b>	170,000

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

# Humangenom - Sequenzierung

1. Methode

2. Ergebnisse Biologie

3. Anwendung Medizin

# Analyse der Human-Genom-Rohfassung

---

## Besonderheiten

### 30.000 -35.000 Gene

erwartet 28.000 - 140.000

Ackerschmalwand 25.700, Fadenwurm 18.300,  
Fruchtfliege 13.300

### komplexere Transkription

60% aller Gene;

mRNAs/Gen: Chr22=2,6, Chr19=3,2

Fadenwurm: 22% aller Gene; 1,3 mRNAs/Gen

### komplexere Proteinarchitektur

keine Neuen Domänen,

"Auflaufen" (accretion) zusätzlicher Domänen  
an den Gen-Enden

### Gentransfer von Prokaryonten

233 Gene mit Ähnlichkeit nur zu Prokaryonten  
davon 133 bei Prokaryoten weit verbreitet

### ungleichmäßige Genverteilung

25% des Genoms sind "genlose Wüste"

Gegensatz zu Ackerschmalwand, Fadenwurm, Fruchtfliege

### 50% repetitive Sequenzen

Ackerschmalwand 11%, Fadenwurm 7%, Fruchtfliege 3%

"fossiles Archiv" der genomischen Evolution  
dramatische Reduktion der Akkumulation  
in den letzten 50 MioJ

SINEs als Symbionten von Genen (Chr19=57%)

### Mutationsraten Frau:Mann 1:2

Vergleich X und Y Chr

### ungleichmäßige SNP-Verteilung

63% aller 5kb-Fragmente enthalten SNP(s)  
interindividueller Unterschied 0.01%

# Analyse der finalen Fassung des Human-Genoms

---

## Besonderheiten

**fast vollständige Sequenz**  
2,85 Mb, 99% des Euchromatins  
genreich, fast alle Lebensprozesse

**nur noch 341 Lücken**  
33 heterochromatisch (Centromere)  
308 euchromatisch (28 Mb),  
50% mit segmentalen Duplikationen

**extrem hohe Genauigkeit**  
99.999%  
ein Fehler auf 100.000 Bausteine

**20.000 -25.000 Gene**  
erwartet 28.000 - 140.000  
Pflanze 25.700, Wurm 18.300, Fliege 13.300

**verbesserte Gen-Vorhersagen**  
58% aller Gene der Rohfassung waren  
fehlerbehaftet

**neue & verlässliche Analysen**  
Pseudogene  
1.183 neue entstandene Gene  
Geruchsrezeptoren, Immunität, Schwangerschaft  
37 'kürzlich gestorbene' Gene  
10 Geruchsrezeptoren

## Charakteristika humaner Gene (Mittelwerte)

Länge Exon	145 bp
Länge Intron	3385 bp
Exonzahl	8,8
5 UTR	300
3 UTR	770
Kodierende Region	1340 bp

# Genetische Variabilität

---

## Einzel-Nukleotid-Polymorphismus (SNP)

ATTCGACGTATTG

ATTCGATGTATTG

↑  
SNP

- in der Regel bi-allelisch
- 12 Mio SNPs genomweit  $1/250$  bp
- 2 Individuen unterscheiden sich in  $\sim 300.000$  SNPs  $1/10.000$  bp
- ca. 5%, d.h. 600.000 SNPs mit Phänotyp (?)  
50-100.000 mit klinische Relevanz (?)

# Genetische Variabilität

---

## Begriffsbestimmung

### Mutationen

Substitution, Insertion, Deletion, Inversion

### Polymorphismen

sind **Mutationen**, die sich mit einer Häufigkeit **>1%** in einer **Population** etabliert haben

Einzel-Nukleotid-Polymorphismen

**SNP** - „single nucleotide“ ~



# Segmentale Duplikationen

---

## Definition

genomische Regionen **>1kb**

Identität **>90%**

## **Human-Genome**

**5.3%** segmental dupliziert

**87%** aller segmentalen Duplikationen **>50 kb**

# Genetische Variabilität

---

## Strukturelle Variationen

### **Polymorphismen >1kb:**

segmentalen Duplikationen

Deletionen

Inversionen

# Genetische Variabilität

---

## Zusammenfassung

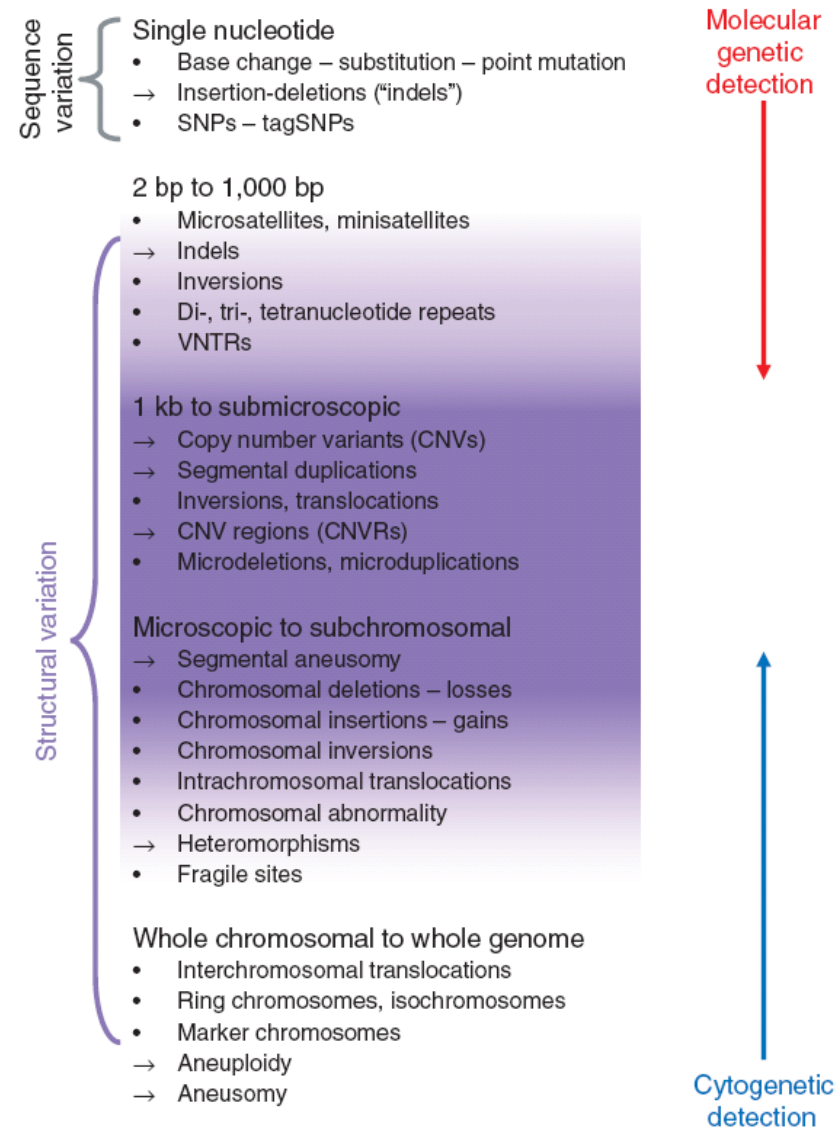
Genomes of any two individuals in the human population **differ more at the structural level** than at the nucleotide sequence level.

Differences between individuals

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

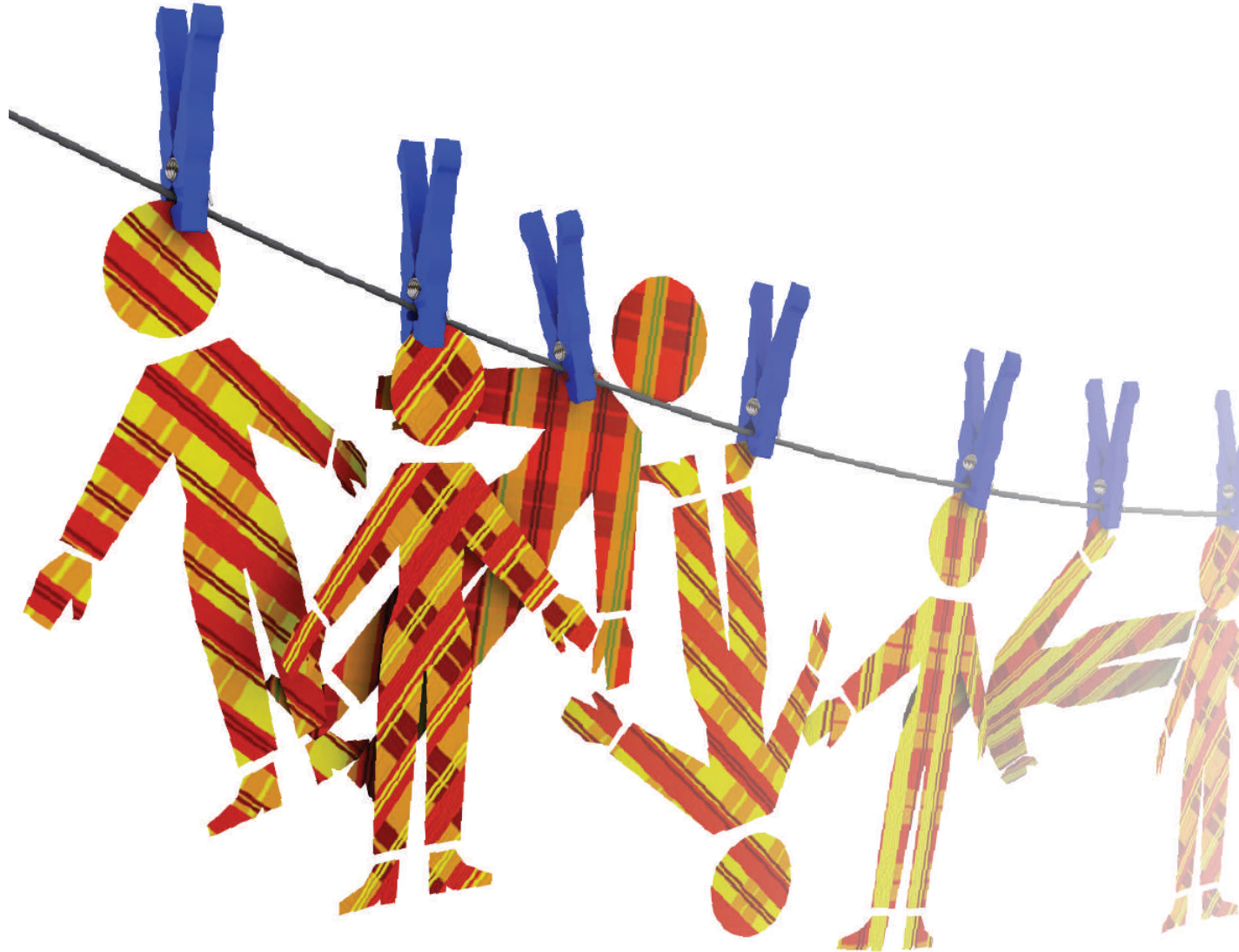
# Genetische Variabilität

## Lexikon



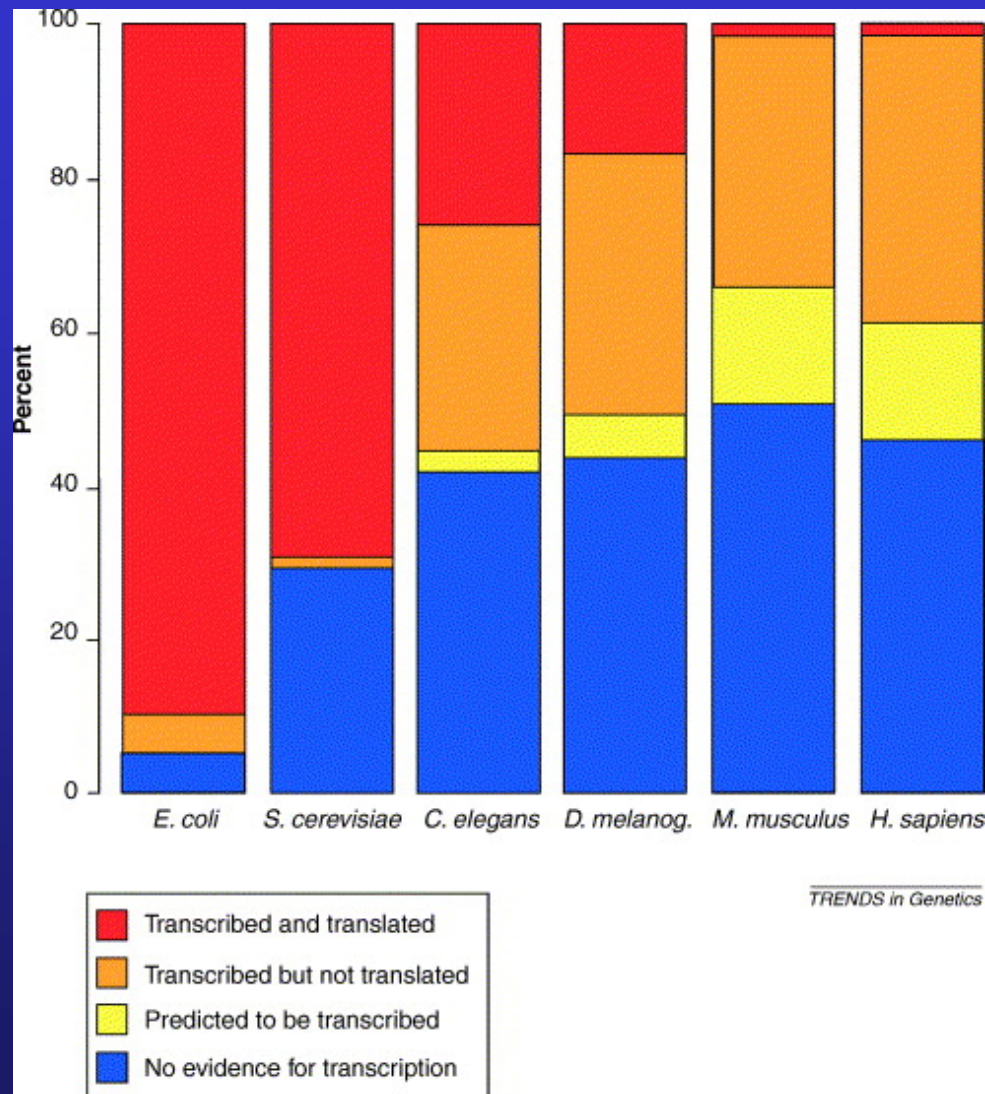
# Genom-Dynamik

Patchwork people ?



*News Feature, Nature 437:1084 (2005)*

# Expansion der nicht kodierenden RNA



Huttenhofer A, Schattner P, Polacek N.  
Non-coding RNAs: hope or hype?  
Trends Genet. 2005 May;21(5):289-97

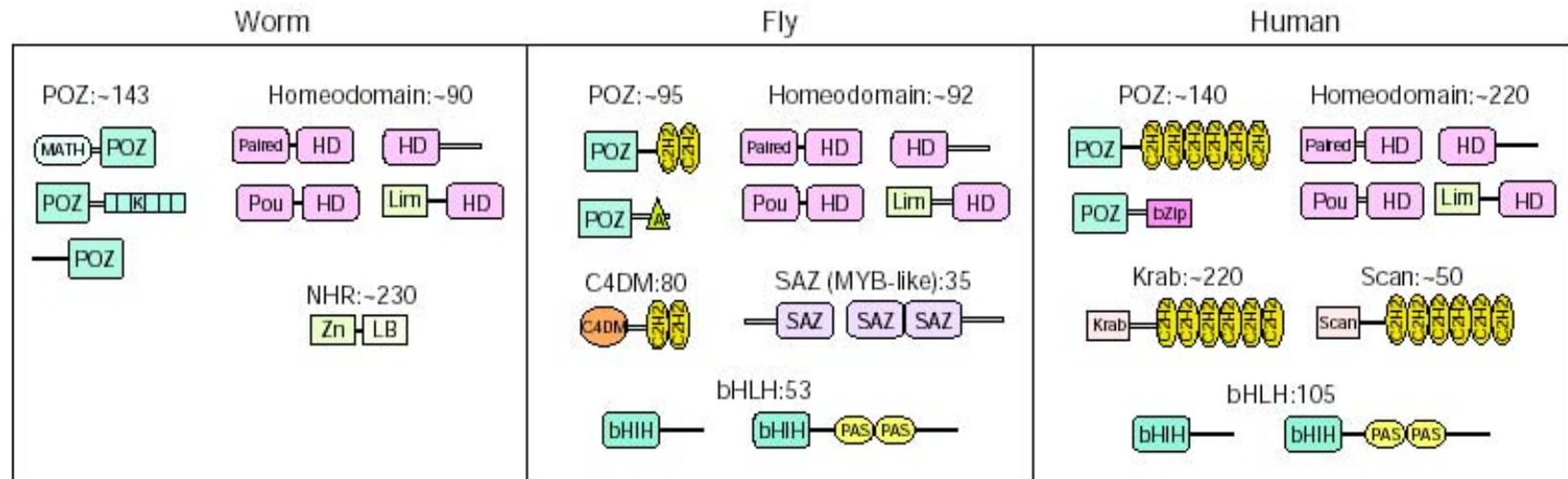
# Typen und Funktionen der nicht kodierenden RNA

Table 1 | **Common ncRNAs and their functions**

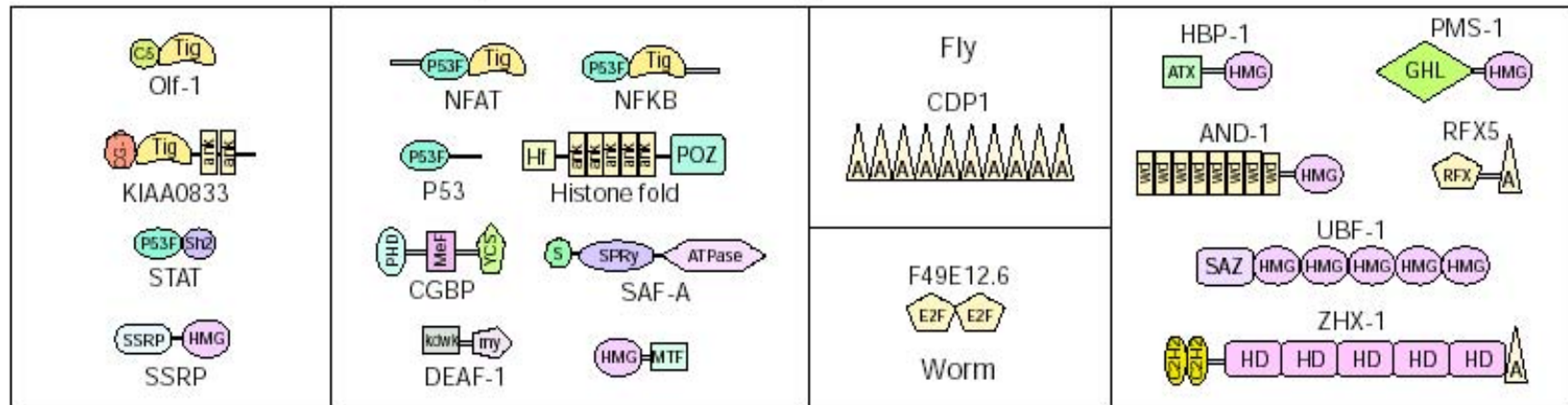
ncRNA type	Description	Function
RNase P	~400 nt long	Cleaves tRNA precursors to result in mature 5' ends; as a catalytic RNA (ribozyme) in bacteria, for example, cleaves tRNA precursors under high monovalent salt conditions in the absence of a protein
miRNA	Small, 21–23 nt long ssRNA	Targets mRNAs for cleavage (in plants) or translation inhibition (in mammals)
siRNA	21–23 nt long ssRNA	Targets mRNAs for cleavage
raRNA	Small, 21–23 nt long ssRNA	Involved in repeat silencing
snoRNA	~50–200 nt long, structured RNA that is localized to the nucleolus	Specifies modification of rRNAs, snRNAs or tRNAs (in Archaea only); C/D box snoRNAs specify 2'-O-methylation of the ribose of a target RNA, H/ACA box snoRNAs specify pseudouridylation
gRNA	Small, ~60 nt long ssRNA, containing a poly U tract at its 3' end (from 5–20 U residues)	Guides U insertions or deletions within mitochondrial pre-mRNAs of certain protozoan organisms, for example, trypanosomes
snRNA	Structured; ~100–300 nt long (in humans)	Guides splicing of pre-mRNAs (for example, U1, U2, U4, U5 and U6 snRNAs)
rRNA	Highly structured; sized between ~120 (5S rRNA) and several thousand nucleotides (18S, 28S rRNAs)	As part of the ribosome it catalyses peptide bond formation (for large rRNA only)
Xist RNA	~17 kb long RNA, which is transcribed from the X chromosome	Involved in X-chromosome inactivation and dosage compensation
tRNA	Highly structured, sized between ~70 and 95 nt	RNA adapter molecules for amino acids; guides amino acids to the ribosome in an mRNA-dependent mode
SRP RNA	Has a rod-like structure, sized ~300 nt (in humans)	Part of the SRP, a ribonucleo-protein complex that is involved in targeting specific proteins to the endoplasmic reticulum for subsequent secretion
6S RNA	Highly structured RNA (~180 nt long in <i>Escherichia coli</i> ), which forms a single hairpin that is found in bacteria	Binds to the $\sigma^{70}$ factor of the RNA polymerase complex, thereby regulating transcription of $\sigma^{70}$ promoters

gRNA, classical guide RNA; miRNA, microRNA; ncRNA, non-protein-coding RNA; rasiRNA, repeat-associated siRNA; rRNA, ribosomal RNA; siRNA, small interfering RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; SRP, signal recognition particle; Xist, X-(inactive)-specific transcript.

# Domänenstruktur von Proteinen



Unique and shared domain organizations in animals



Ancient architectures conserved in all animals

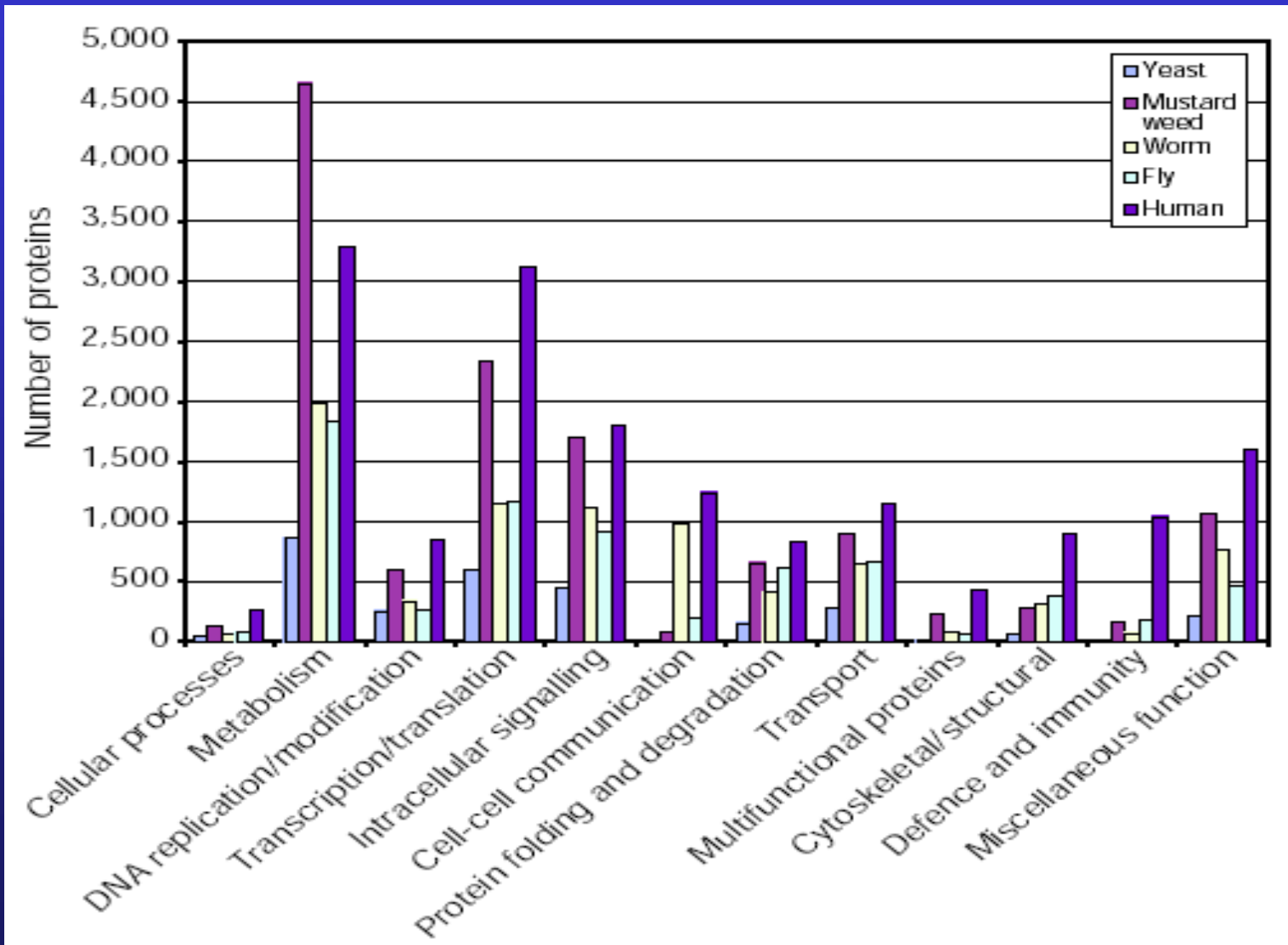
Shared by fly and human

Unique to fly or worm

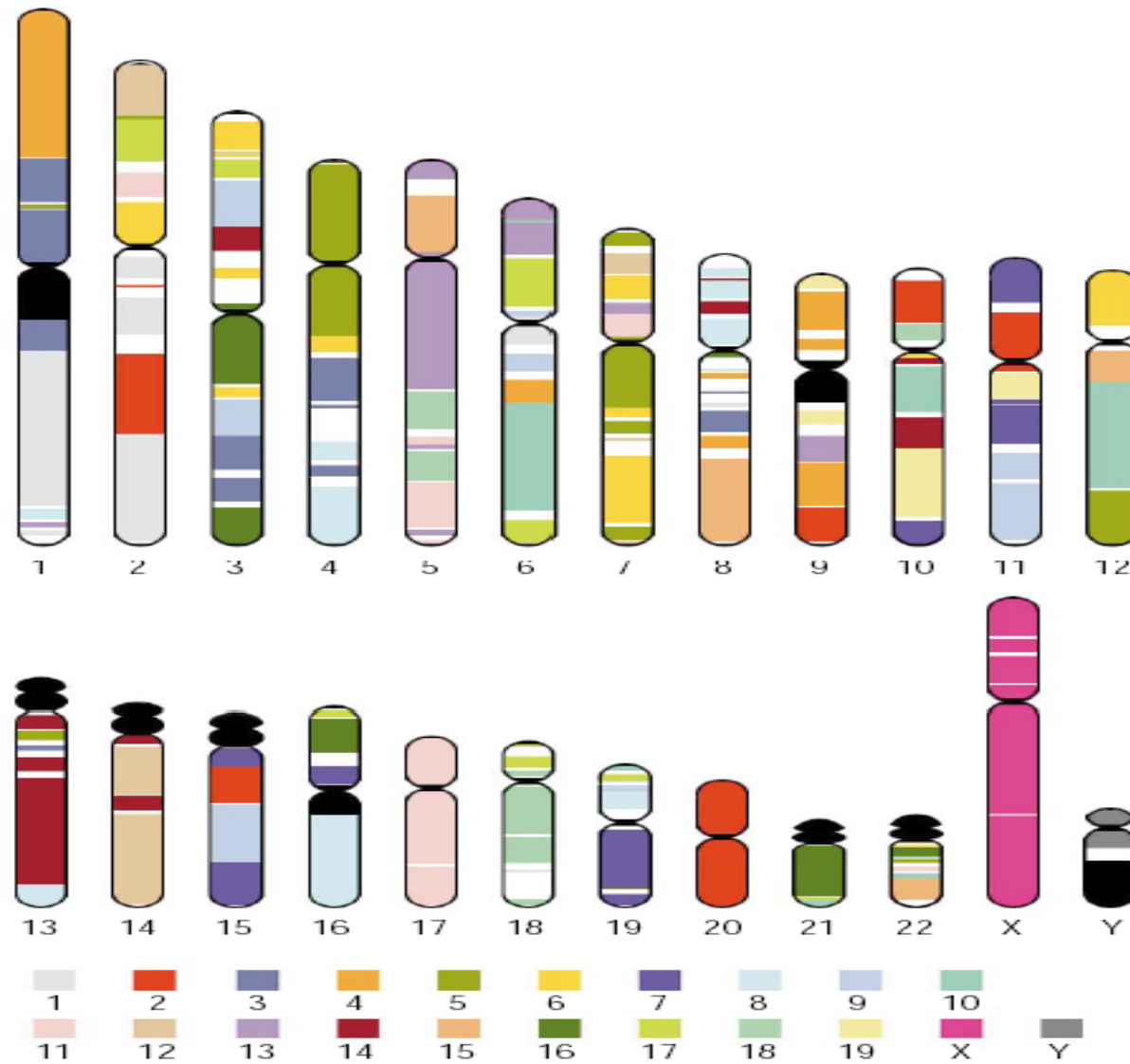
Unique to human



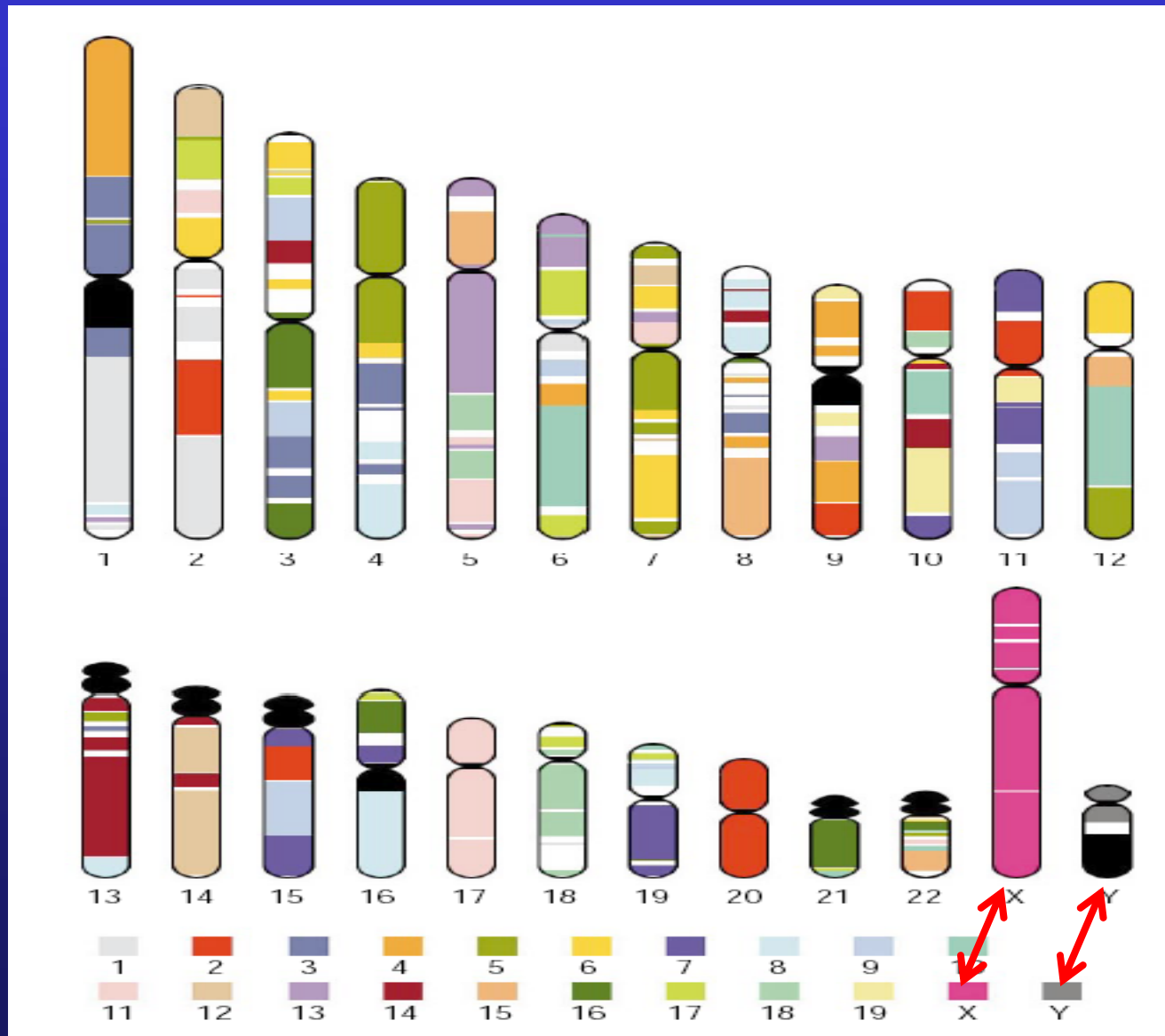
# Gen/Proteinfunktionen



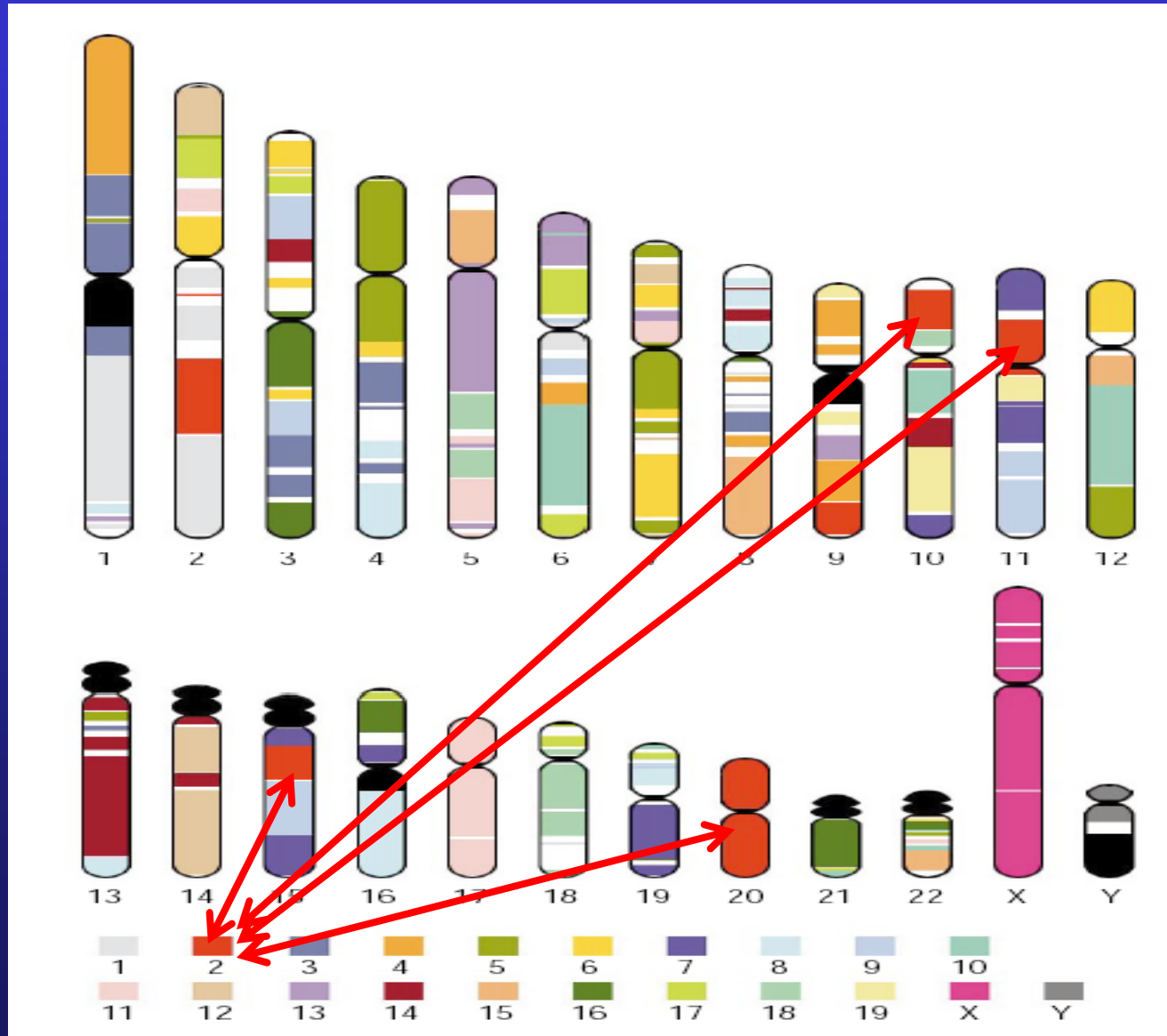
# Chromosomenvergleich Maus-Mensch



# Chromosomenvergleich Maus-Mensch



# Chromosomenvergleich Maus-Mensch



# Humangenom - Sequenzierung

1. Methode

2. Ergebnisse Biologie

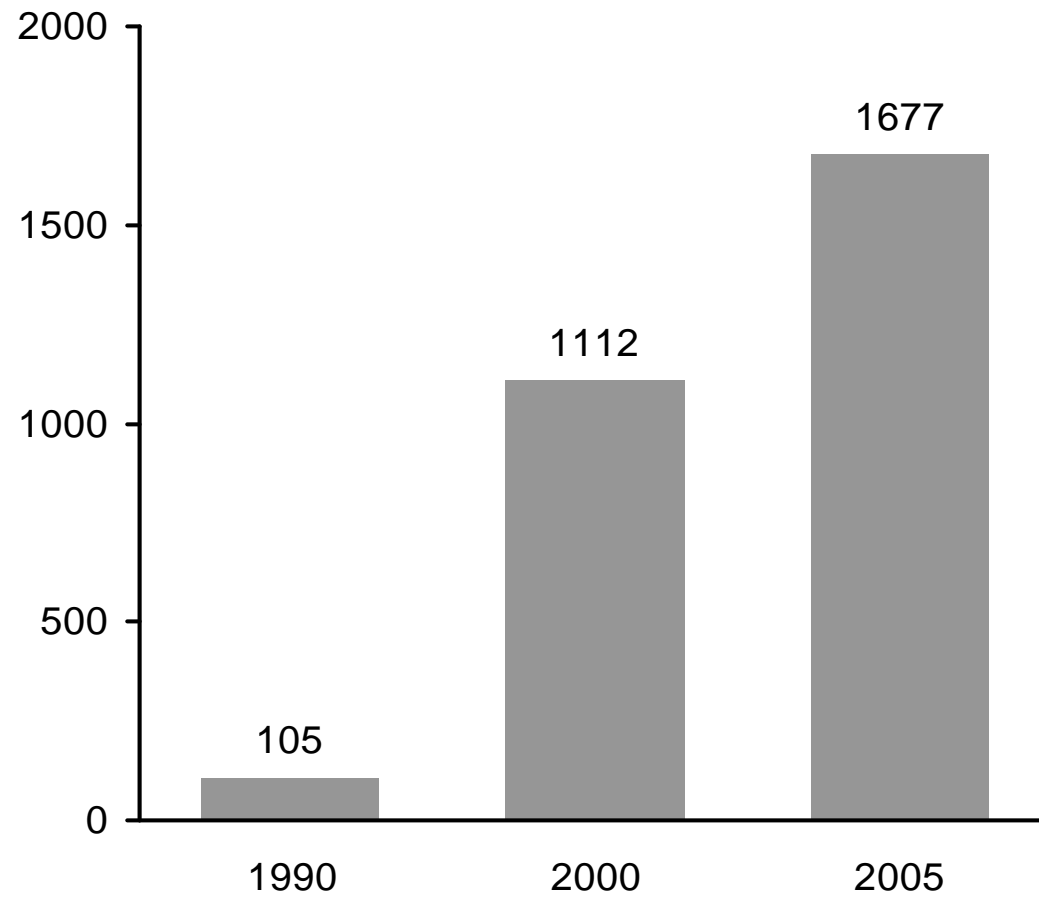
3. Anwendung Medizin

# Diagnose von genetischen Erkrankungen

# „Krankheitsgene“

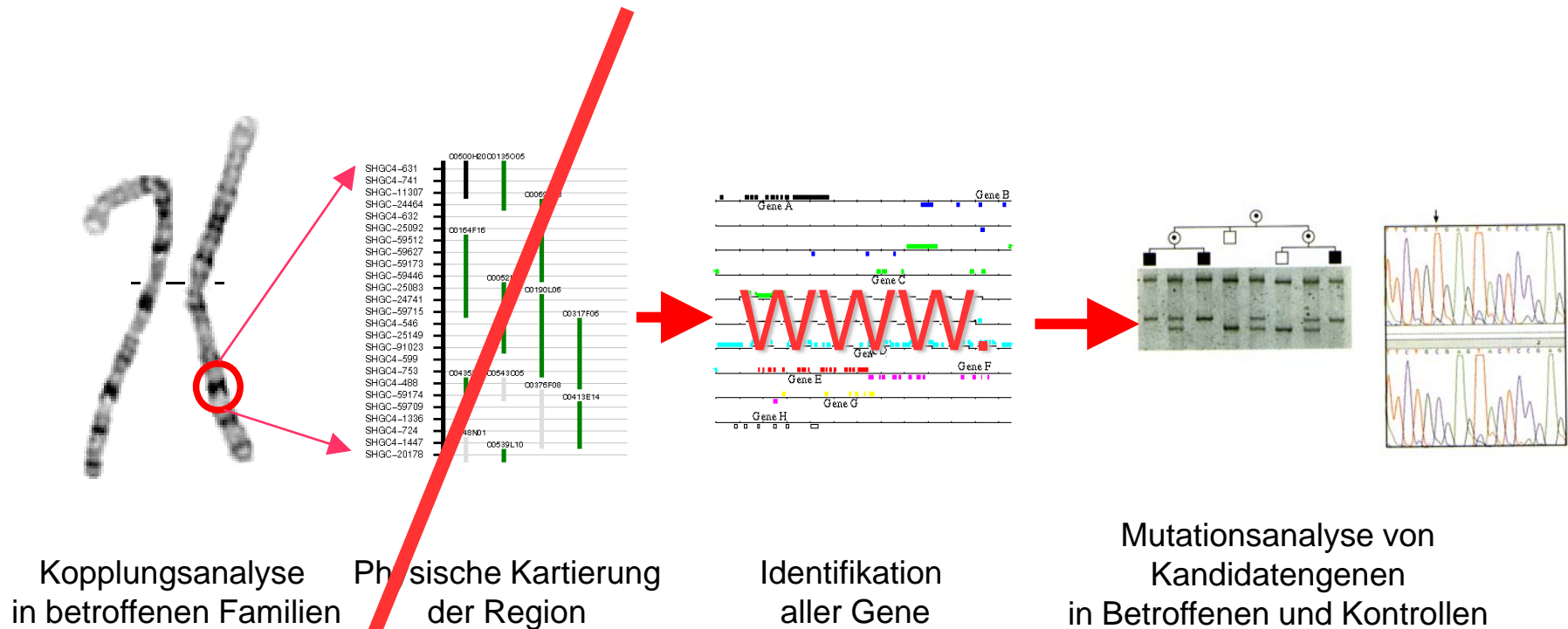
Aufklärungsrate

**Anzahl identifizierter Krankheitsgene**



# Krankheitsgene

## klassische Positionsklonierung





# Lokalisation von Krankheitsgenen

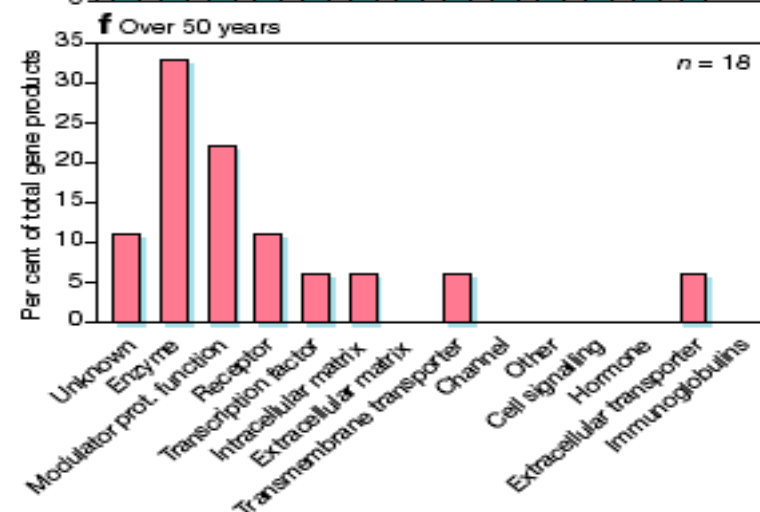
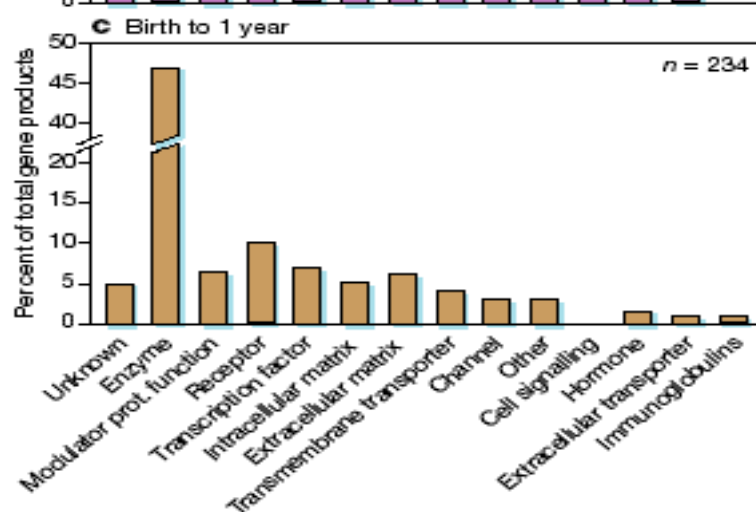
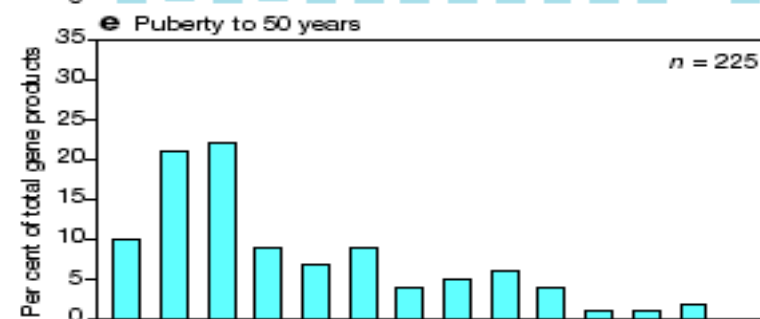
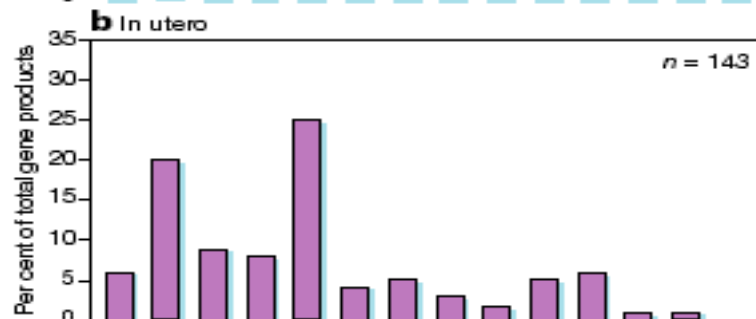
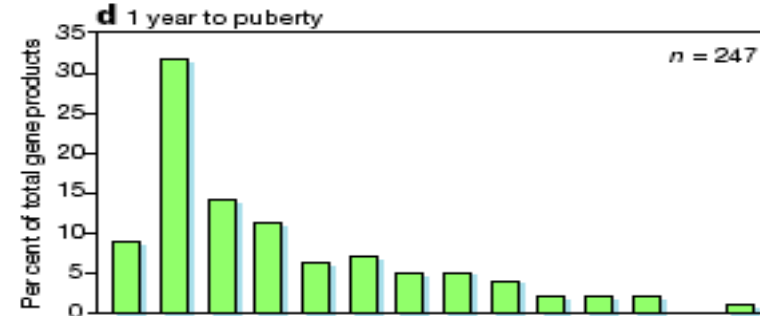
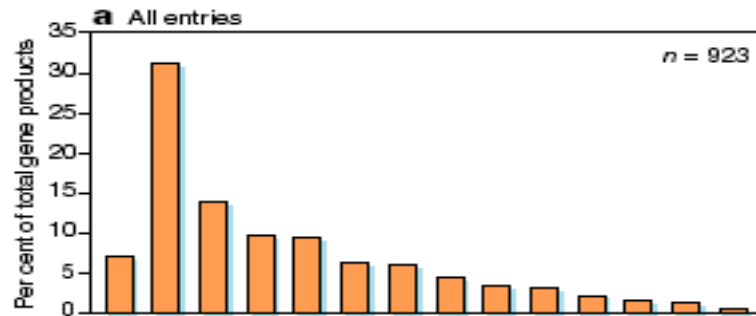
## Human Disease Genes Database

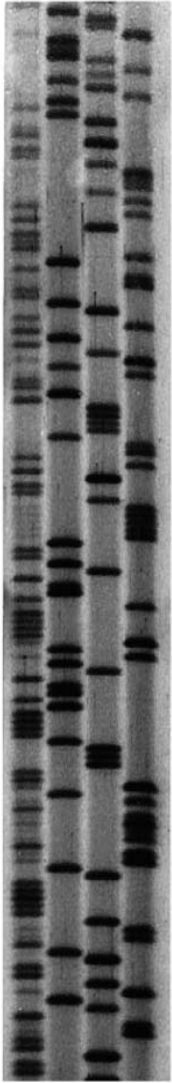
G. Jimenez-Sanchez, B. Childs, D. Valle

Page 1 of 29

MIM #	Locus Name	Disease name	Prot funct	Freq	Mode Inher	Age onset	Life Expect	Malfor
600994	Deafness, autosomal dominant 5	Deafness, autosomal dominant 5 (3)	0			0		0
604392	Arylhydrocarbon-interacting receptor protein-like 1	Leber congenital amaurosis, 604393 (3)	0			1		0
154500	Treacher Collins-Franceschetti syndrome-1 (TREACLE)	Treacher Collins mandibulofacial dysostosis (3)	0			1		1
301200	Amelogenin	Amelogenesis imperfecta (3)	0			1		0
309000	Oculocerebrorenal syndrome of Lowe	Lowe syndrome (3)	0			2		0
300500	Ocular albinism-1, Nettleship-Falls type	Ocular albinism, Nettleship-Falls type (3)	0			2		0
300031	Fragile site, folic acid type, rare, fra(X)(q28)	Mental retardation, X-linked, FRAX type (3)	0			2		0
180090	Retinaldehyde-binding protein-1, cellular	Retinitis pigmentosa, autosomal recessive (3)	0			3		0
604365	Prominin, mouse, homolog-like 1	Retinal degeneration, autosomal recessive, prominin-related (3)	0			3		0
602280	Tubby-like protein-1	Retinitis pigmentosa-14, 600132 (3)	0			3		0
603937	Oxygen-regulated photoreceptor protein-1 (retinitis pigmentosa-1)	Retinitis pigmentosa-1, 180100 (3)	0			3		0
153700	Viteliform macular dystrophy (Best disease)	Macular dystrophy, vitelliform type (3)	0			3		0
601652	Myocilin (trabecular meshwork-induced glucocorticoid response protein)	Glaucoma 1A, primary open angle, juvenile-onset, 137750 (3); Glaucoma 1A, primary open angle, recessive (3)	0			3		0
142640	Histidine-rich glycoprotein	Thrombophilia due to HRG deficiency (3); ?Thrombophilia due to elevated HRG (1)	0			4		0
603904	Integral membrane protein 2B (BRI gene)	Dementia, familial British, 176500 (3); Dementia, familial Danish, 117300 (3)	0			4		0
163890	Synuclein, alpha (non A4 component of amyloid precursor)	Parkinson disease, type 1, 601508 (3)	0			4		0
137290	Membrane component, chromosome 1, surface marker 1 (40kD glycoprotein, identified by monoclonal antibody GA733)	Corneal dystrophy, gelatinous drop-like, 204870 (3)	0			5		0
601517	Ataxin-2	Spinocerebellar ataxia-2, 183090 (3)	0			3		0
162200	Neurofibromin (neurofibromatosis, type I)	Neurofibromatosis, type 1 (3); Watson syndrome, 193520 (3); Leukemia, juvenile myelomonocytic (3)	0			3		0
235200	Hemochromatosis	Hemochromatosis (3); Porphyria variegata, 176200 (3)	0			4		0
240300	Autoimmune regulator	Autoimmune polyglandular disease, type I (3)	0			4		0
173910	Polycystin-2	Polycystic kidney disease, adult, type II (3)	0			4		0
602452	Budding uninhibited by benzimidazoles 1, S. cerevisiae, homolog of (mitotic checkpoint gene BUB1)	Colorectal cancer with chromosomal instability (3)	0			5		0
602117	Necdin	Prader-Willi syndrome, 176270 (3)	0			1		1
219800	Cystinosis	Cystinosis, nephropathic (3)	0			2		0
256731	Ceroid-lipofuscinosis, neuronal-5	Ceroid-lipofuscinosis, neuronal-5, variant late infantile (3)	0			3		0
253800	Fukuyama congenital muscular dystrophy	Muscular dystrophy, Fukuyama congenital (3); ?Walker-Warburg syndrome, 236670 (2)	0			3		1
208900	Ataxia-telangiectasia mutated (includes complementation groups A, C, D, and E)	Ataxia-telangiectasia (3); T-cell prolymphocytic leukemia, sporadic (3); Lymphoma, B-cell non-Hodgkin, somatic (3); (Breast cancer, susceptibility to) (3); Lymphoma, mantle cell (3)	0			3		0
229300	Frataxin	Friedreich ataxia (3); Friedreich ataxia with retained reflexes (2)	0			3		0
156570	5-methyltetrahydrofolate-homocysteine methyltransferase 1	Methylcobalamin deficiency, cbl G type (3)	0			3		0
222300	Wolframin	Wolfram syndrome (3)	0			3		0
256100	Nephronophthisis-1 gene	Nephronophthisis, juvenile (3)	0			3		0
190080	Avian myelocytomatosis viral (v-myc) oncogene homolog	Burkitt lymphoma, 113970 (3)	0			3		0
191100	Tuberous sclerosis-1	Tuberous sclerosis-1 (3)	0			3		0
603009	Dysferlin	Muscular dystrophy, limb-girdle, type 2B, 253601 (3); Myosin myopathy, 254130 (3); Myopathy, distal, with anterior tibial onset (3)	0			4		0

# Krankheitsgene und Funktionen





A C G T

genome.fli-leibniz.de  
Teaching