

Bioinformatik: The Next Generation

Prof. Dr. Caroline Friedel

Lehr- und Forschungseinheit
Bioinformatik



Was ist Bioinformatik?



Theoretische und Praktische Informatik
Statistik, Mathematik
Molekularbiologie, Biochemie, Genetik,
Evolutionbiologie, ...

„Die Bioinformatik ist eine **interdisziplinäre Wissenschaft**,
die **Probleme aus den Lebenswissenschaften** mit
theoretischen computergestützten Methoden löst“

Wikipedia

Informatik → Methoden:

- Algorithmen
- Theoretische Informatik
- Datenbanken
- Softwareentwicklung

Biologie → Anwendungen:

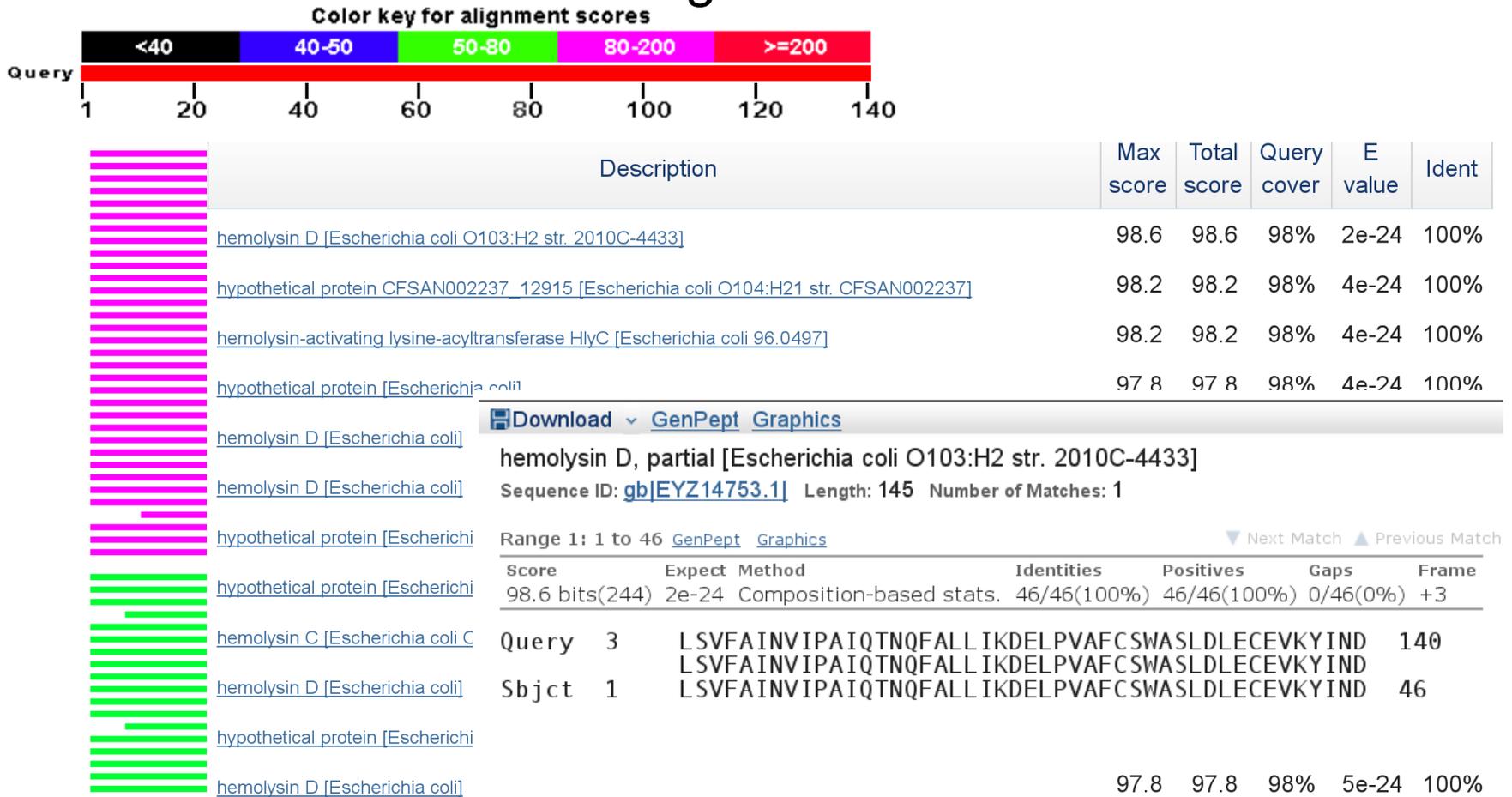
- Genomsequenzierung
- Genexpressionsanalyse
- Biologische Netzwerke
- ...

Der Klassiker: Sequenzsuche

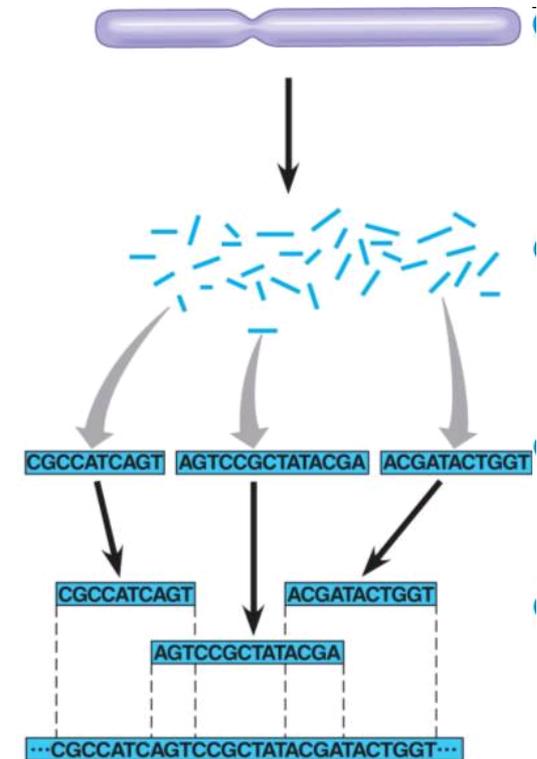


CATTATCTGTTTTTGCTATTAATGTCATACCTGCTATTCAGACAAATCAATTTGCTCTCTTAATAAAGGATG
AGCTTCCTGTAGCTTTTTGTAGCTGGGCCAGTTTAGATCTGGAATGTGAGGTAAAATATATAAATGAT

- **BLAST = Basic Local Alignment Search Tool**



- Erste Anwendungen:
 - Gen- und Proteindatenbanken, Proteinstrukturen, etc.
 - Sequenzanalyse und -vergleiche
- Der Durchbruch:
 - Genomsequenzierung
 - Shotgun sequencing
 - Assemblierung notwendig
 - Menschliches Genom (Celera)
 - > 27 Mio Reads (~550 bp)
 - > 20.000 CPU Stunden
 - 300 Mio \$



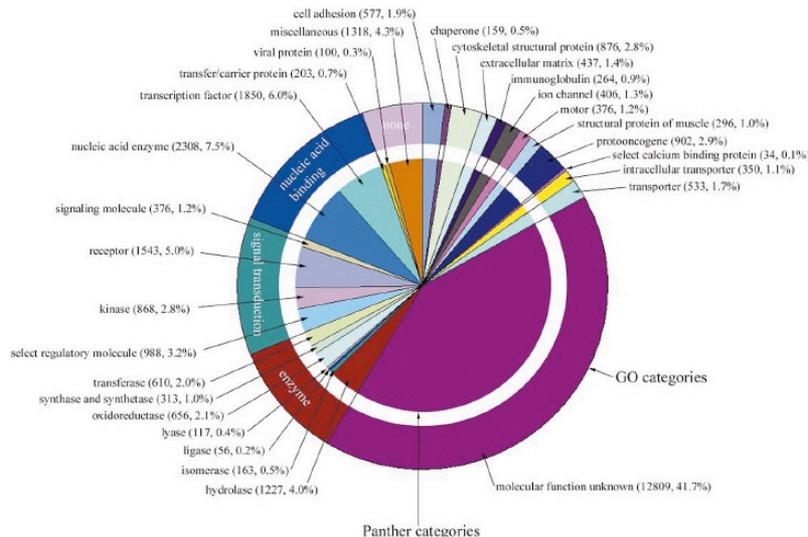
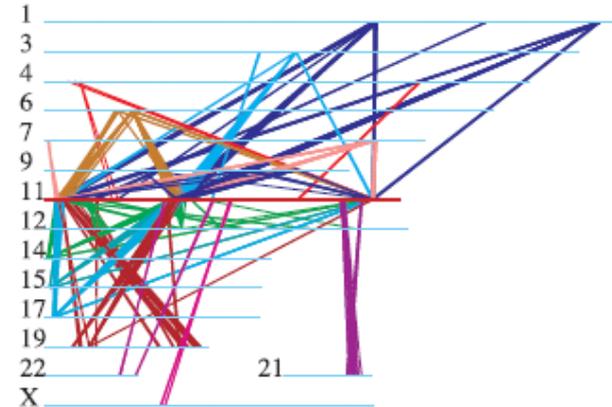
Was noch?



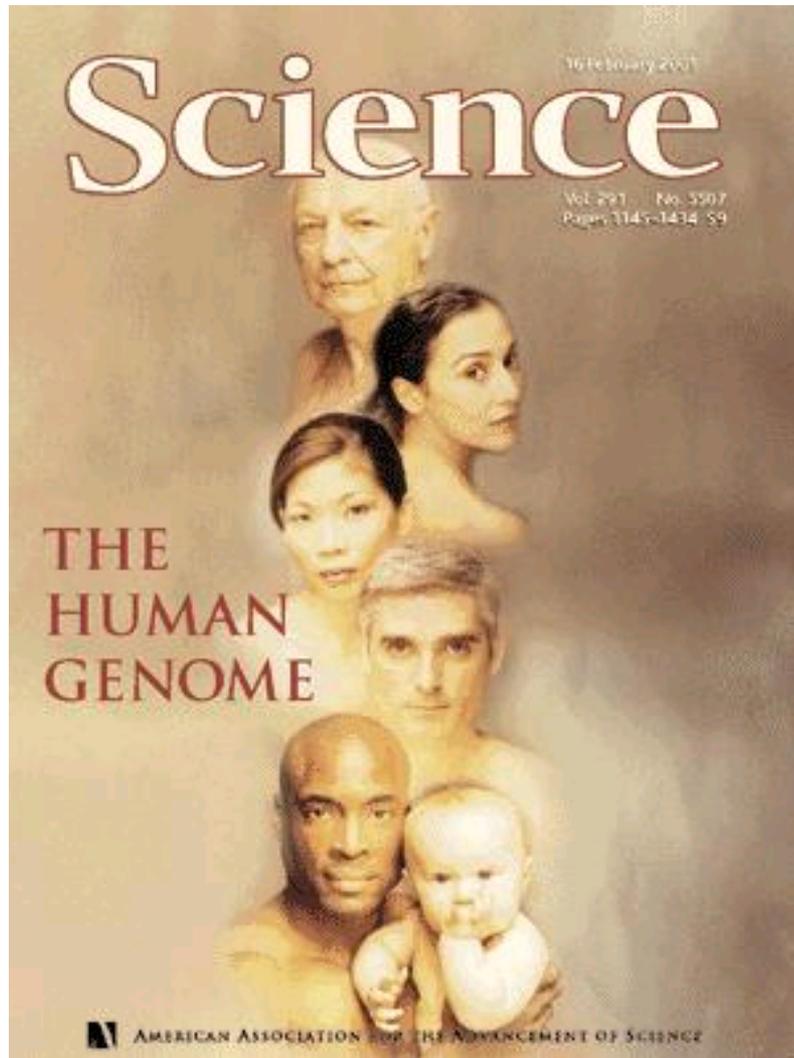
- Gen-Vorhersage und Annotation



- Analyse der Genom-Struktur und Evolution



- Proteinfunktion: Vorhersage und Analyse



15./16. Februar 2001: Die ersten Drafts



The Sequence of the Human Genome

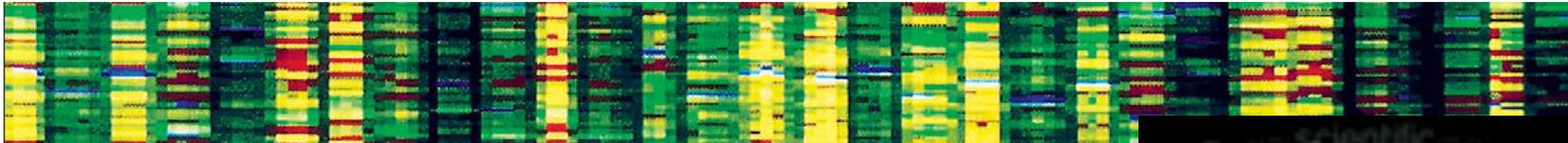
J. Craig Venter et al.



- Abstract
- 1 Sources of DNA and Sequencing Methods
 - 1.1 Library construction and sequencing
 - 1.2 Trace processing
 - 1.3 Quality assessment and control
- 2 Genome Assembly Strategy and Characterization
 - 2.1 Assembly data sets
 - 2.2 Assembly strategies
 - 2.3 Whole-genome assembly
 - 2.4 Compartmentalized shotgun assembly
 - 2.5 Comparison of the WGA and CSA scaffolds
 - 2.6 Mapping scaffolds to the genome
 - 2.7 Assembly and validation analysis
- 3 Gene Prediction and Annotation
 - 3.1 Automated gene annotation
 - 3.2 Otto validation
 - 3.3 Gene number
 - 3.4 Features of human gene transcripts
- 4 Genome Structure
 - 4.1 Cytogenetic maps
 - 4.2 Linkage map
 - 4.3 Correlation between CpG islands and genes
 - 4.4 Genome-wide repetitive elements
- 5 Genome Evolution
 - 5.1 Retrotransposition in the human genome
 - 5.2 Pseudogenes
 - 5.3 Gene duplication in the human genome
 - 5.4 Large-scale duplications
- 6 A Genome-Wide Examination of Sequence Variations
 - 6.1 SNPs found by aligning the Celera consensus to the PFP assembly
 - 6.2 Comparisons to public SNP databases
 - 6.3 Estimation of nucleotide diversity from ascertained SNPs
 - 6.4 Variation in nucleotide diversity across the human genome
 - 6.5 SNPs by genomic class
- 7 An Overview of the Predicted Protein-Coding Genes in the Human Genome
 - 7.1 Molecular functions of predicted human proteins
 - 7.2 Evolutionary conservation of core processes
 - 7.3 Differences between the human genome and other sequenced eukaryotic genomes
- 8 Conclusions



- Ursprüngliche Hoffnung: Genomsequenz verbessert Verständnis von
 - Menschlicher Evolution
 - Ursachen von Krankheiten
 - Wechselspiel zwischen Umwelt und Vererbung
- Stattdessen eher nur mehr Fragen
 - Viel weniger Gene als erwartet: ~20,000 anstatt > 100,000
 - 96-99% Ähnlichkeit zwischen Mensch und Schimpanse. Woher kommt der Unterschied ?
 - Nur wenige Krankheiten durch einzelne Mutationen und Gene verursacht
- Komplexität durch Interaktion zwischen Genom und Umwelt



den Molekularbiologen entziffert – die Bedienungsanleitung dazu bleibt aber ein Rätsel.

Das vertrackte Genom



First Steps on a Long Road

Eric Schadt

Chief Scientific Officer, Pacific Biosciences of California, Menlo Park, CA, USA.

DNA variations and their interaction with diverse environmental pressures are how nature's fundamental forces shape nearly every facet of life. The first human genome sequence, published in 2001, provided a canonical reference from which to understand genome structure, as well as a library of functional units. Before 2001, only a handful of examples of

ome, Part II

celebrating the 10th anniversary of the human genome project, we explore the impacts of genomics on ourselves, as well as

and Laura M. Zahn

and seven gorgeous grants related to my fellow sequencer, a Kalahari Bushman from Windhoek in February 2001, bore an uncanny resemblance to discover that I was and gentle people that ha

THE CITIZEN REVIEW

Initial impact of the human genome

Eric S. Lander¹

The sequence of the human genome has been a decade since its publication, on our understanding of inherited diseases and cancer, and in fulfilling the promise of genomics for medicine.

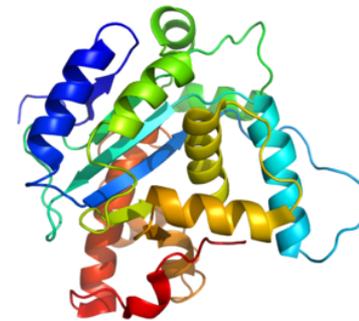
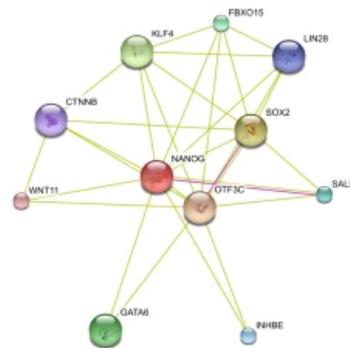
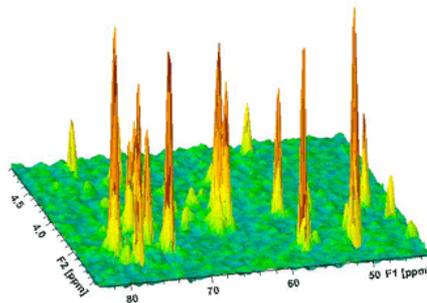
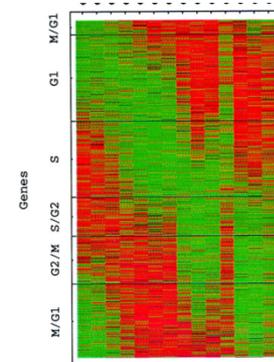
On 15 February 2001, a decade ago this week, *Nature* published a 62-page paper entitled 'Initial sequencing and analysis of the human genome', reporting a first global look at the contents of the human genetic code. The paper¹ marked a milestone in the international Human Genome Project (HGP), a discovery programme conceived in the mid-1980s and launched in 1990. The same week, *Science*

The intensity of interest can be seen in the 2.5 million queries per day to the major genome data servers and in the flowering of a rich field of computational biology.

The greatest impact of genomics has been the ability to investigate biological phenomena in a comprehensive, unbiased, hypothesis-free manner. In basic biology, it has reshaped our view of genome physiology.

This is a monumental month for human genomics as we celebrate 10 years since the publications of the human genome. It is time to reflect on the advances that this endeavor has brought to mankind. The ability for scientists to generate a complete human genome sequence meant that, for the

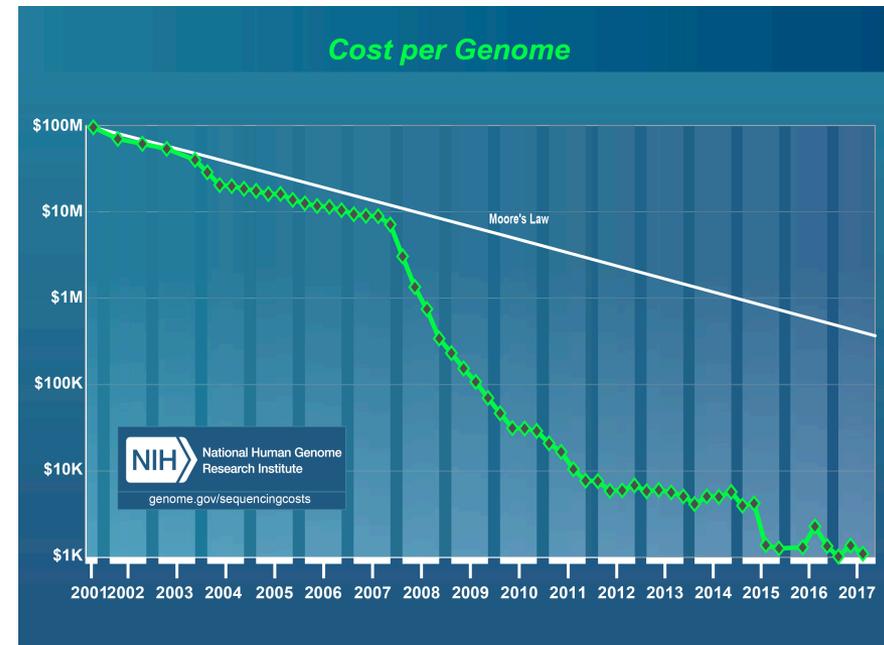
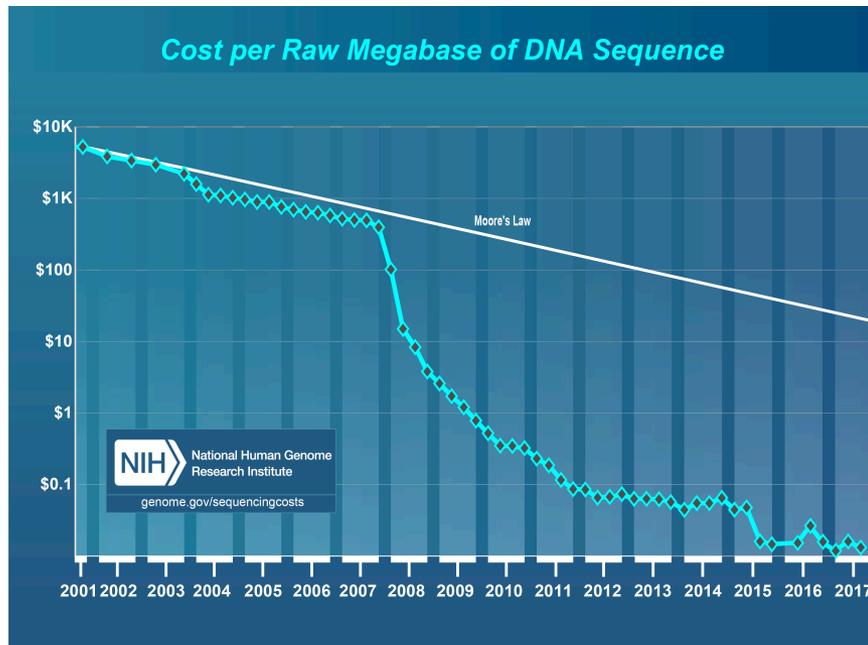
- Omics = Genomweite Studien
- Genomics = Alles was es im Genom gibt
- Transcriptomics = Alle Transkripte / mRNA
- Proteomics = Alle Proteine
- Interactomics = Alle Interaktionen
- Metabolomics = Alle Metabolite
- Kinomics = Alle Kinasen
- ...



Next-Generation-Sequencing: Das Comeback der Sequenzierung



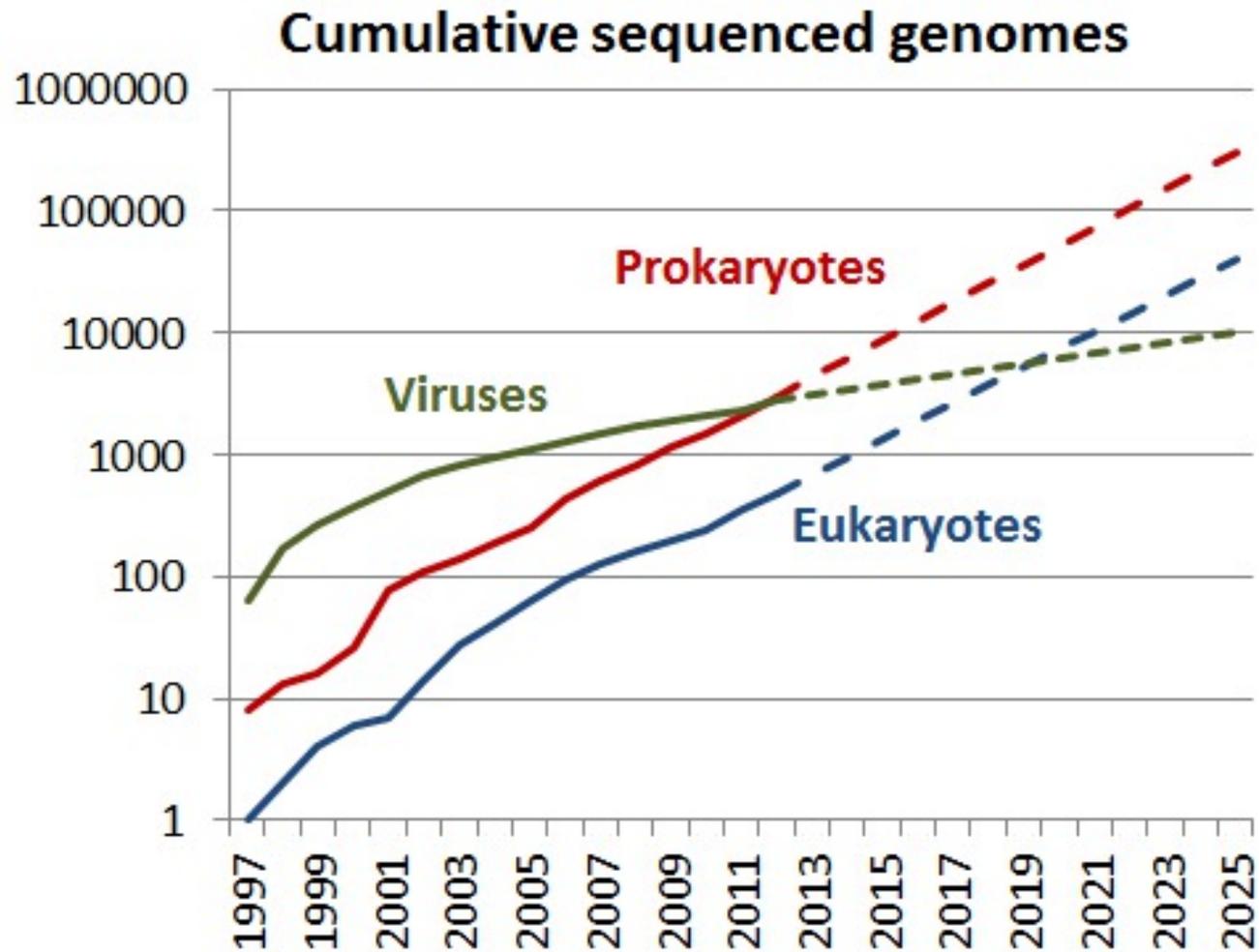
- Schnell und einfach zu bedienen
- Hoher Durchsatz
- ⇒ Sequenzieren wird immer billiger



Next-Generation-Sequencing: Das Comeback der Sequenzierung



Genome über Genome



Genome über Genome

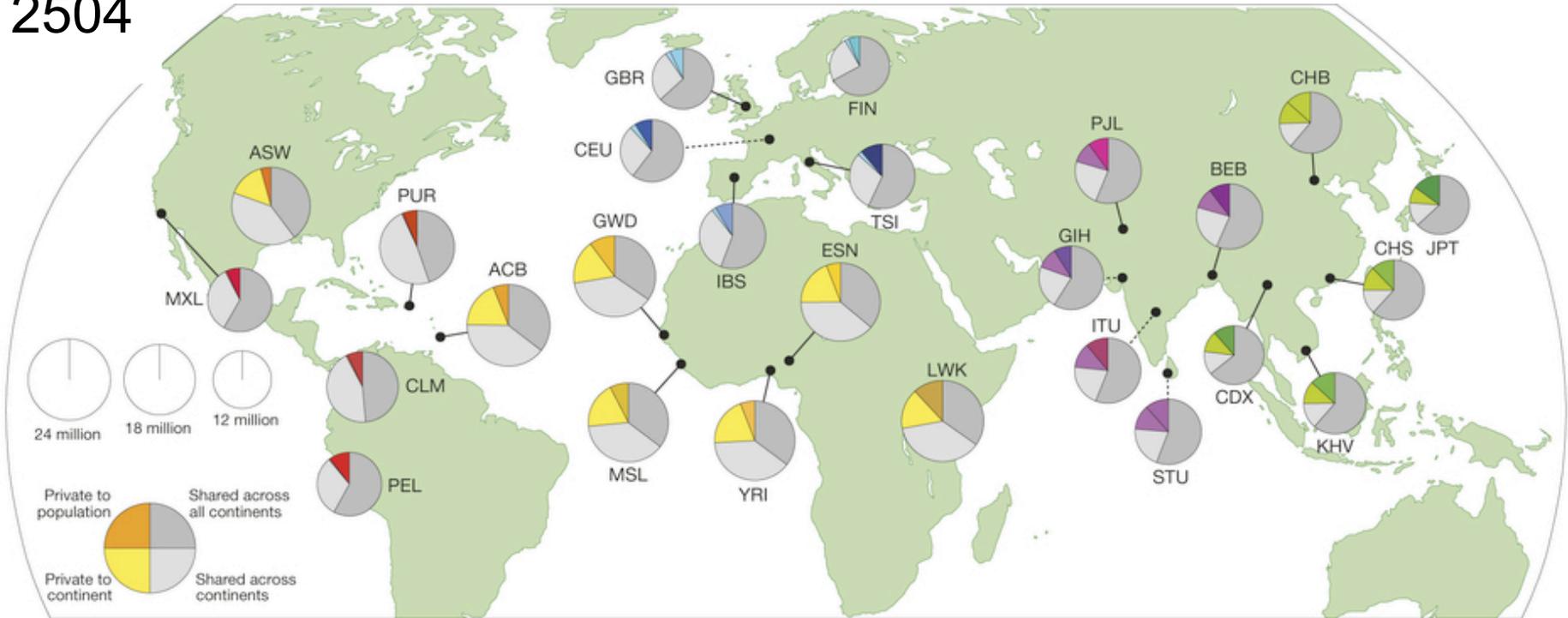


1000 Genomes

A Deep Catalog of Human Genetic Variation



2504



1000\$ Genom



Whole GenomeZ - Whole Genome Sequencing for Advanced Diagnostics (130X + 30X +mtDNA)

€849.00 EUR ~~€1,249.00 EUR~~ You save €400.00 EUR

1	+
	-

ADD TO CART

DESCRIPTION

Whole GenomeZ is a Next Generation Sequencing (NGS) Whole Genome Sequencing with 130X coverage to the Whole Exome at 30X coverage to the rest of the Genome. The Whole GenomeZ is a smart test. as broad as the WGS and as accurate as the WES.



Next-Generation-Sequencing: Das Comeback der Sequenzierung

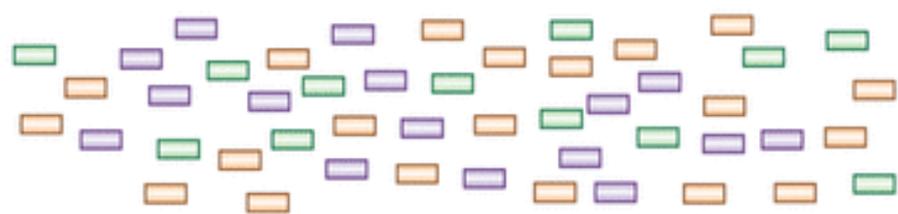


- Nicht mehr nur Genom-Sequenzierung sondern auch
 - RNA-Sequenzierung (RNA-seq)
 - ChIP-seq
 - DNase-seq
 - Ribosome profiling (Ribo-seq)
 - ATAC-seq
 - NET-seq
 - ...

Wozu braucht man die Bioinformatik?



Eingabe: Kurze Sequenzier-Reads



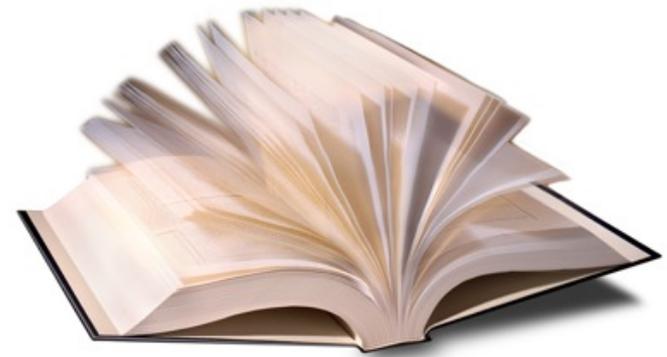
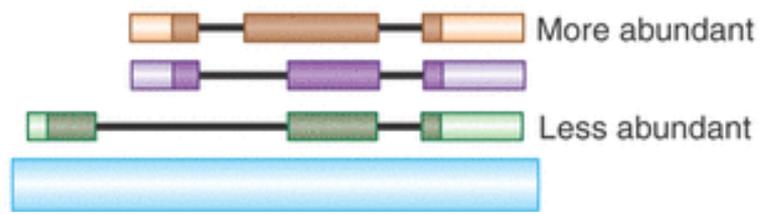
Beispiel: RNA Sequenzierung

Alignment gegen das Genom



Welche mRNAs gibt es?

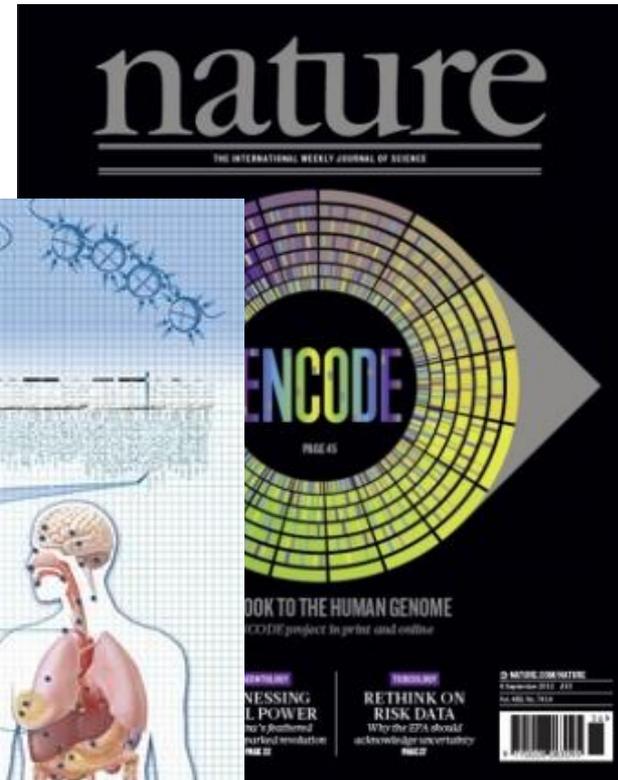
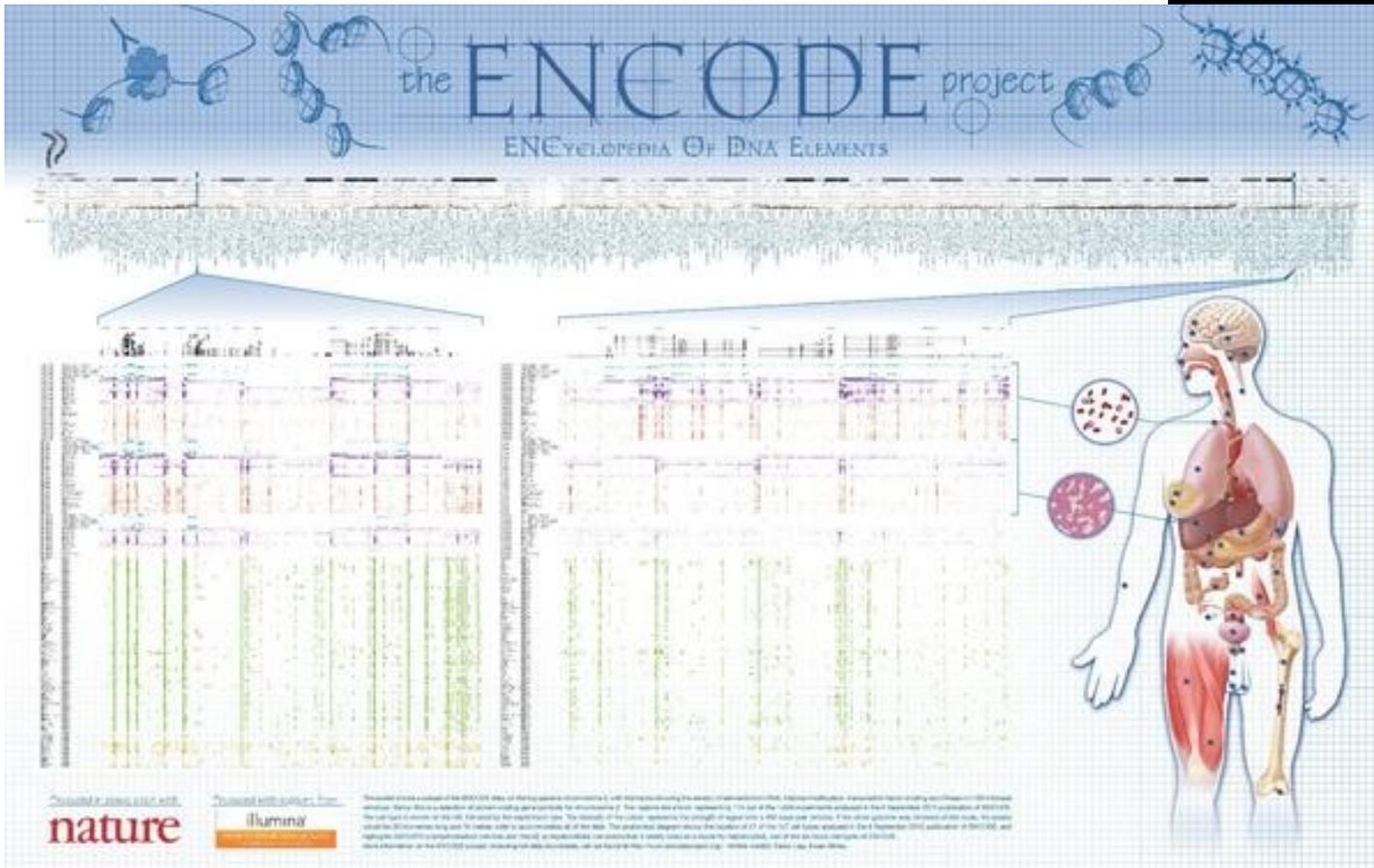
Wie häufig sind sie?



ENCODE



- The ENCODE Project:
ENCyclopedia Of DNA Elements



- 1,600 Arten von Experimenten
- 147 Gewebearten
- >15 Terabyte Rohdaten
- >300 Computerjahre für die Analyse



HiSeq 4000*



NovaSeq 6000**††

Datenmenge pro Run

125–1500 Gb

134–6000 Gb

Dauer

< 1–3.5 days

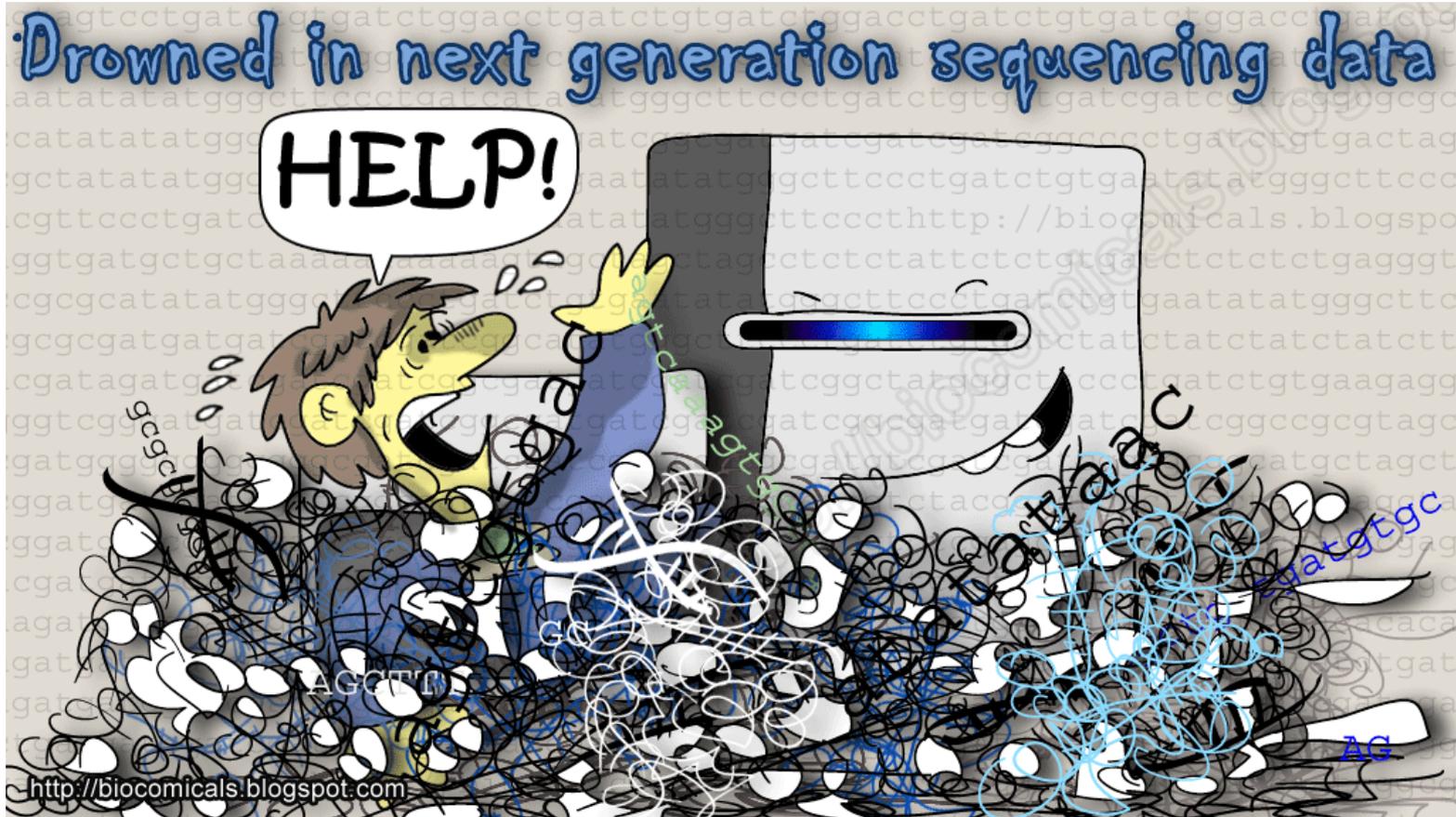
13–44 hr

Anzahl Reads

2.5–5 billion

20 Milliarden

Next-Generation-Sequencing: Das Comeback der Bioinformatik



Ohne **Bioinformatik** geht in der **Molekularbiologie**
und **Medizinischer Forschung** nichts mehr

Wofür braucht man Bioinformatik ?



- Datenverwaltung

- Speichern
- Indizieren
- Integration
- Big Data



Datenbanken

- Datenverarbeitung

- z.B. Assemblierung
- Effiziente Algorithmen
- Künstliche Intelligenz



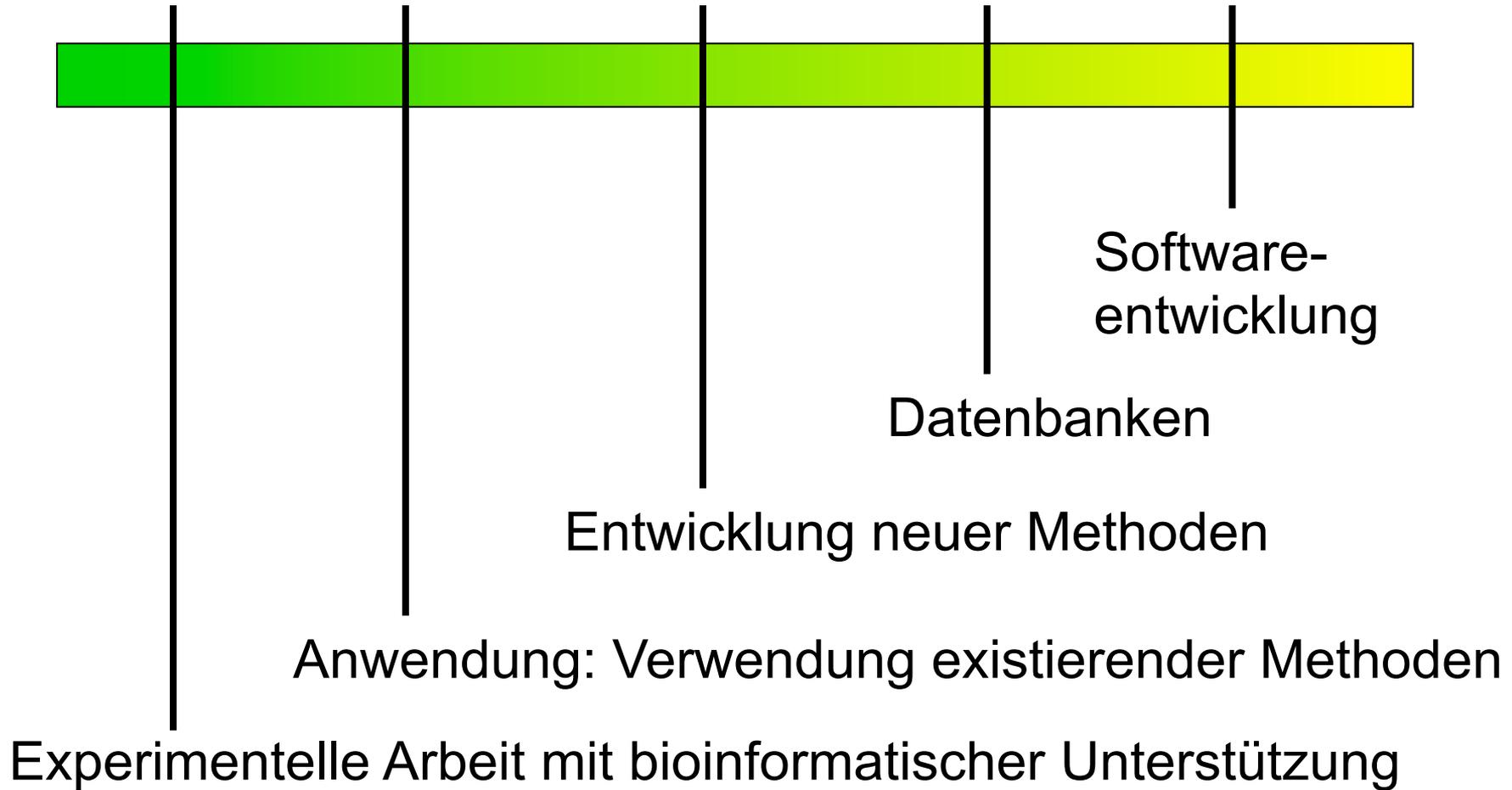
Software / Webserver

- Analyse und Interpretation

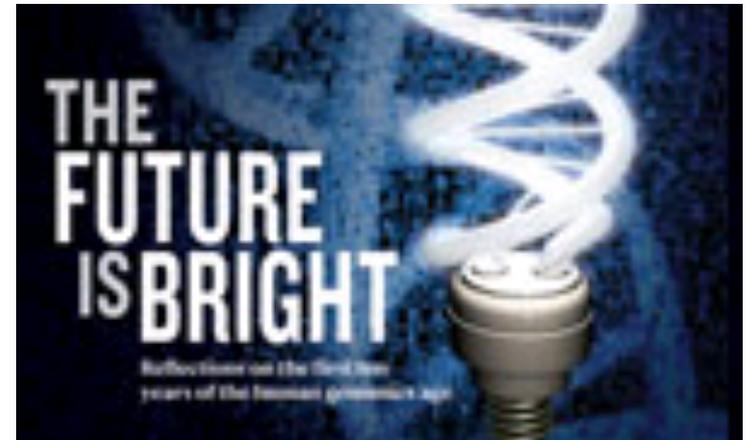
Biologische Erkenntnisse

Biologie

Informatik



- Wissenschaft
- Biotechnologie-Unternehmen:
 - Illumina, Applied Biosystems, Affymetrix, Eurofins, etc.
- Pharmazeutische Forschung:
 - Boehringer-Ingelheim, Bayer Schering, Roche, etc.
- Softwareentwicklung
 - Genomatix, Biomax, etc.
- Data scientist
- Unternehmensberatung



Studienplan Bioinformatik Bachelor

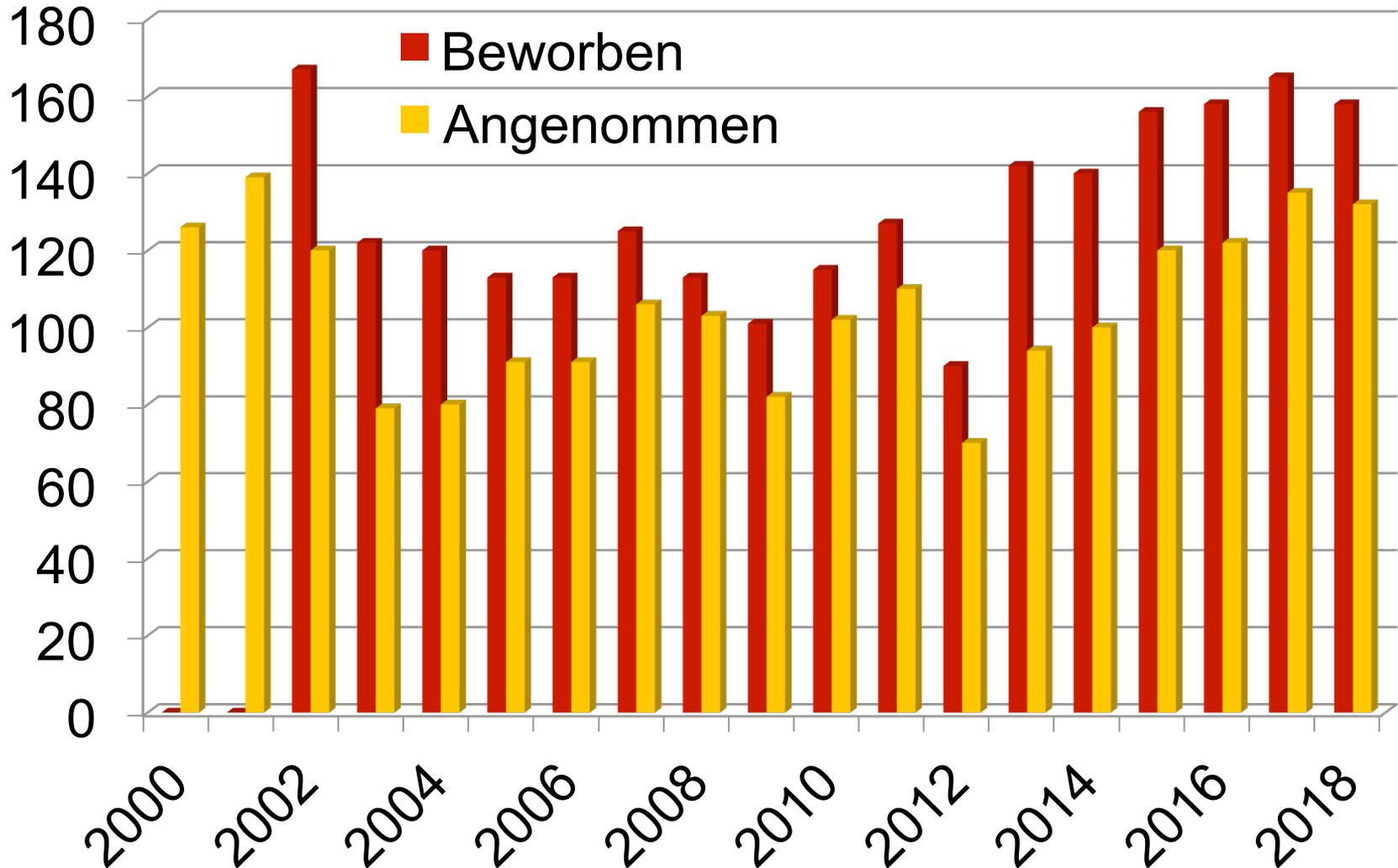


	Informatik	Mathematik, Statistik	Bioinformatik	Biologie, (Bio)chemie
1	Program- mierung	Analysis	Einführung in die Bioinformatik I	Chemie Biologie
2	Algorithmik Bioinformatik Tutorium	Logik & Diskrete Strukturen	Einführung in die Bioinformatik II Problem-Based Learning	Biologie Biochemie Grundlagen
3	Datenbanken	Lineare Algebra	Problem-Based Learning Programmierpraktikum	Biochemie Grundlagen
4	Formale Sprachen & Komplexität	Stochastik & Statistik	Algorithm. Bioinformatik I	Fortgeschr. Biochemie
5			Algorithm. Bioinformatik II Praktikum	Praktikum
6	Wahlpflichtveranstaltungen <small>Prof. Dr. Caroline Friedel, LMU Campustag</small>		Weiterf. Bioinfo., Prakt. Arbeit, Bachelorarbeit	

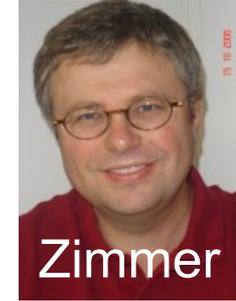


- Eignungsfeststellungsverfahren (EFV)
- Bewerbung bis zum 15. Juli
 - Ausgefüllter Bewerbungsbogen (online)
 - kurzer tabellarischer Lebenslauf
 - Kopie des Abiturzeugnisses
 - Kurzaufsatz, warum man Bioinformatik studieren will
- Auswertung
 - Abiturnote
 - Besonders Fächer: Mathematik, Biologie, Informatik, Chemie, Physik, Deutsch, Englisch
 - ⇒ Annahme oder Auswahlgespräch
- Auswahlgespräch
 - Bewertung: Abiturnote und Ergebnis des Auswahlgesprächs

Annahmemequoten



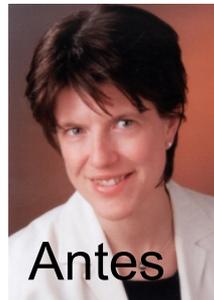
Lehrstühle



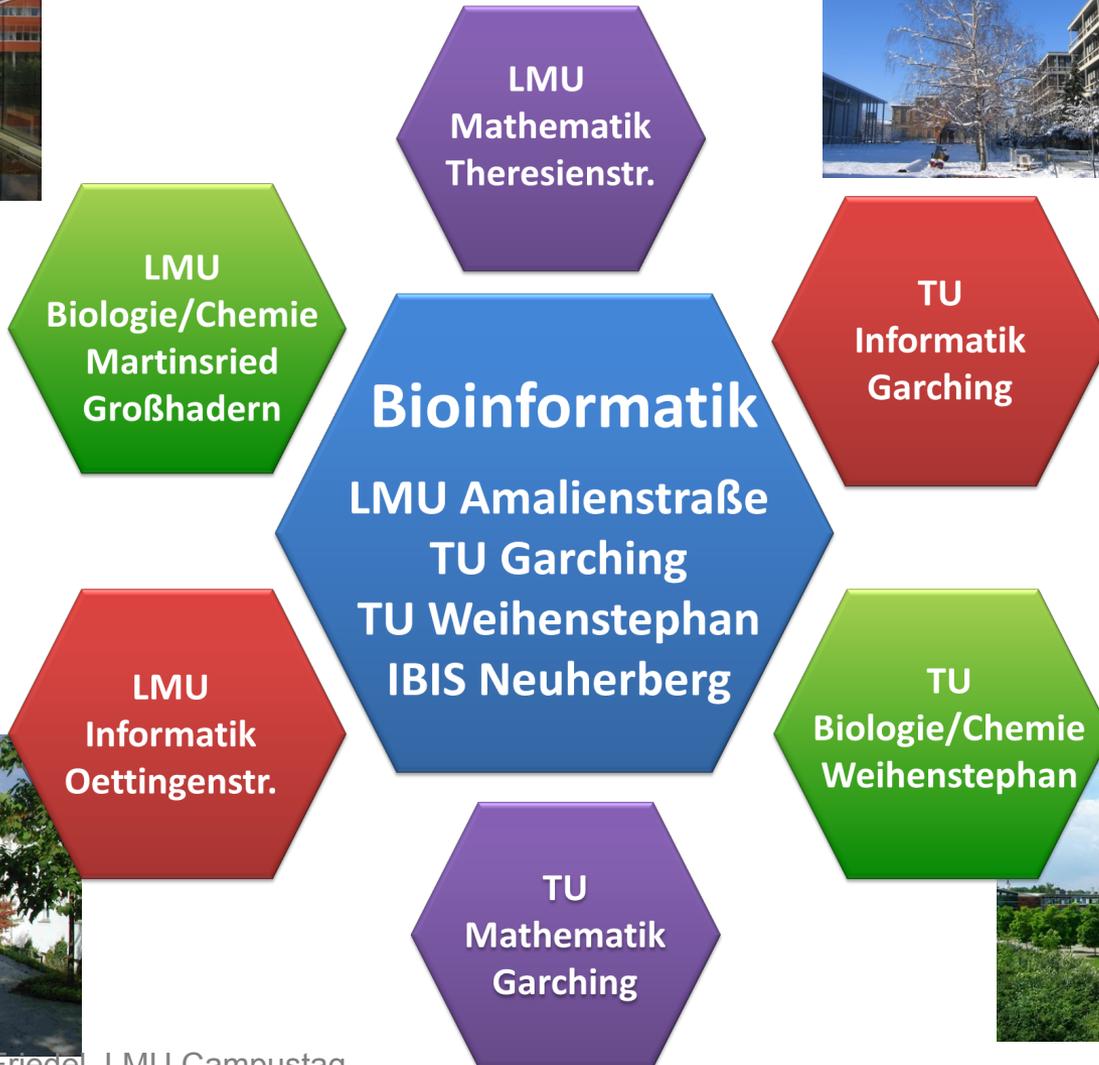
Professoren für Bioinformatik



Assoziierte Gruppen



Standorte



Standorte



LMU
Biologie/Chemie
Martinsried
Großhadern

LMU
Mathematik
Theresienstadt

Bioinformatik



Weihenstephan
Mühlerberg

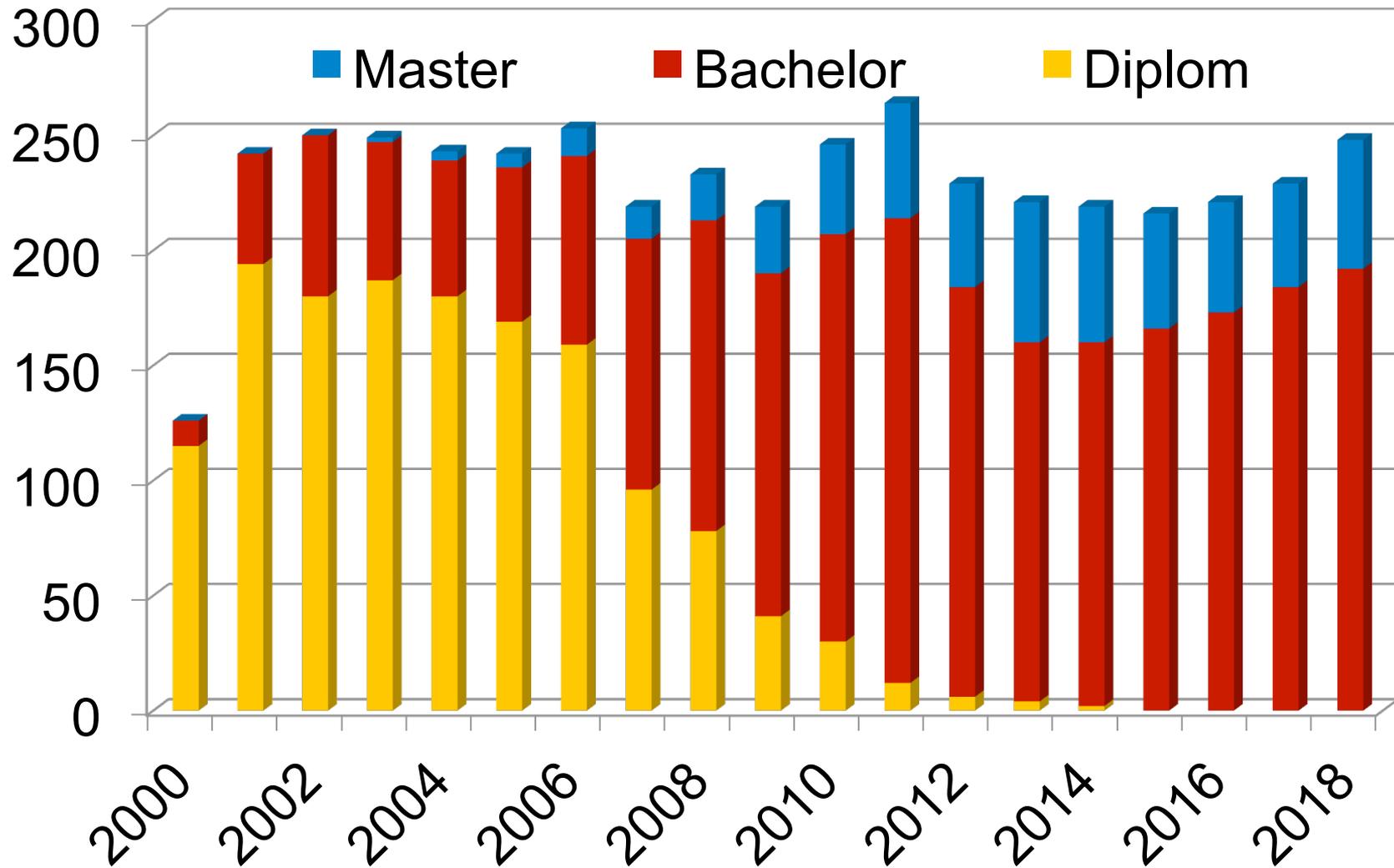
TU
Biologie/Chemie
Weihenstephan



TU
Chemie
Mühlerberg



Studierende





- Web-Seite zum gemeinsamen Studiengang:
 - www.bioinformatik-muenchen.de/
- Web-Seite zum EFV mit Formularen:
 - www.bio.ifi.lmu.de/studium/studiengaenge_bioinformatik/
- Fachschaft Bioinformatik:
 - www.bioinformatik-muenchen.com/
- Informationsstand: Adalberthalle

Noch Fragen ?