

# Bioinformatik: The Next Generation

Prof. Dr. Caroline Friedel

Lehr- und Forschungseinheit  
Bioinformatik



# Was ist Bioinformatik?



Politik Wirtschaft Panorama Sport München Bayern Kultur Wissen Digital Chancen Reise Auto Stil mehr...



Bioinformatik



Finden



Ressort ▼

Typ ▼

Quelle ▼

Datum ▼

14.01.2016 | 11:34



Molekularbiologie

## Forscher untersuchen "dunkle Materie" im Menschen

SZ • Wissen • Molekularbiologie

... Abschnitt" innerhalb eines Moleküls. Mit dieser bioinformatischen Methode bleibt zwar das Dilemma bestehen, ... Forscherinnen,

Andrea Schafferhans vom Lehrstuhl für Bioinformatik der Technischen Universität München. Auch der ... Von *Katrin Blawat*

27.11.2015 | 10:40

Food-Hacker

## Veggie-Burger sollen bald nach Fleisch schmecken

SZ • Wissen • Ernährung

... Mayo ohne Ei produziert wird, arbeiten dafür Bioinformatiker mit Küchenchefs zusammen. Gründer Josh Tetrick ... Von *Kathrin Burger*

11.11.2015 | 10:38



Neues EU-Gremium

## Mathematiker und Physiker sollen EU-Kommission beraten

SZ • Wissen • EU-Gremium

... erin Elvira Fortunato sowie der polnische Bioinformatiker Janusz Bujnicki. Die "hochrangige Gruppe wissenschaftlicher ... Von *Kai Kupferschmidt*

01.10.2015 | 15:03



Wissenschaft

## Daten frei im Internet: Erbgut von 2500 Menschen entziffert

dpa

... funktionelle Konsequenzen haben", erläuterte Bioinformatiker Oliver Stegle, der an einem Außeninstitut des ...

# 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:

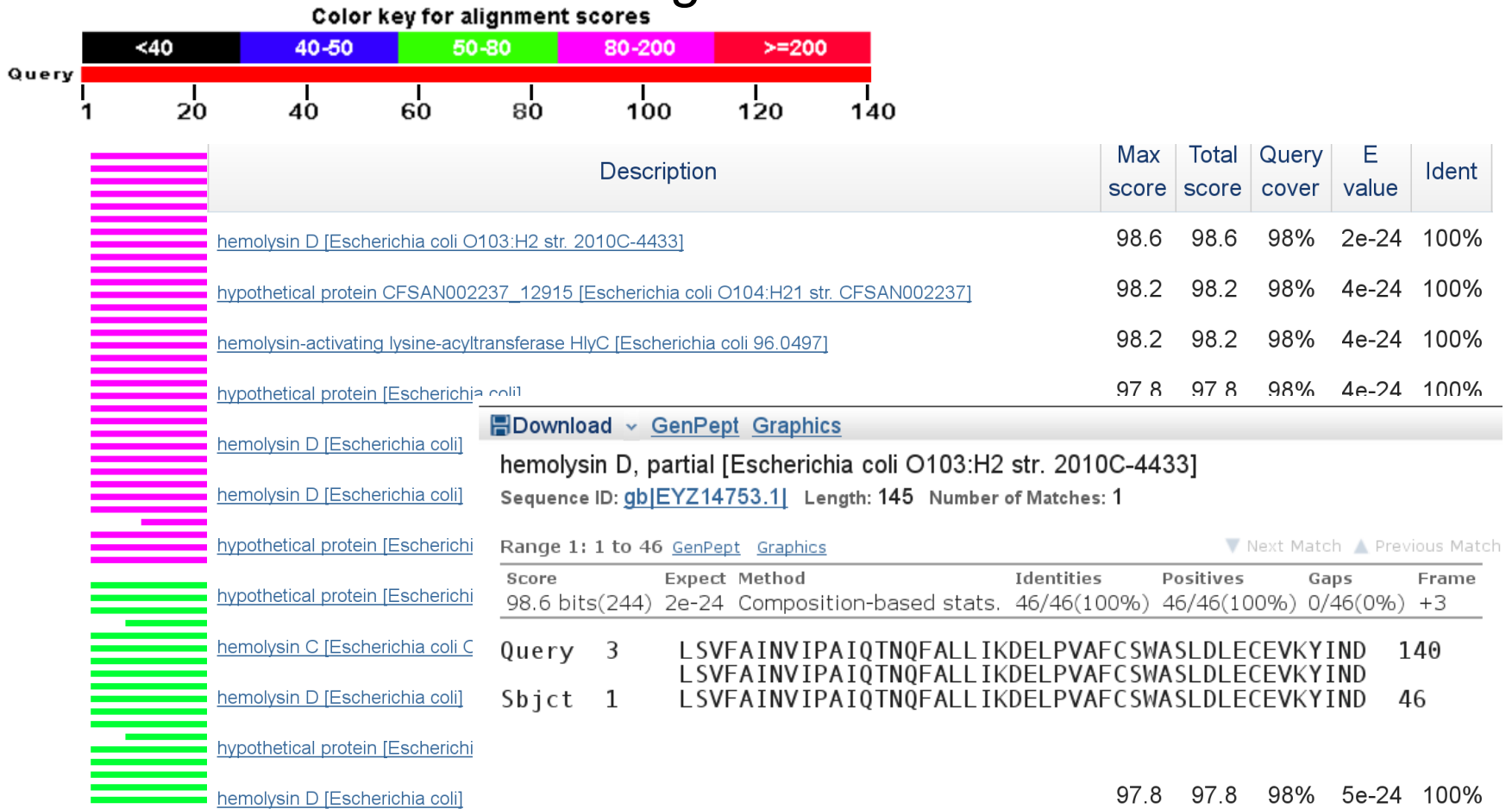
- Genomsequenzierung
- Genexpressionsanalyse
- Biologische Netzwerke
- ...

# Der Klassiker: Sequenzsuche

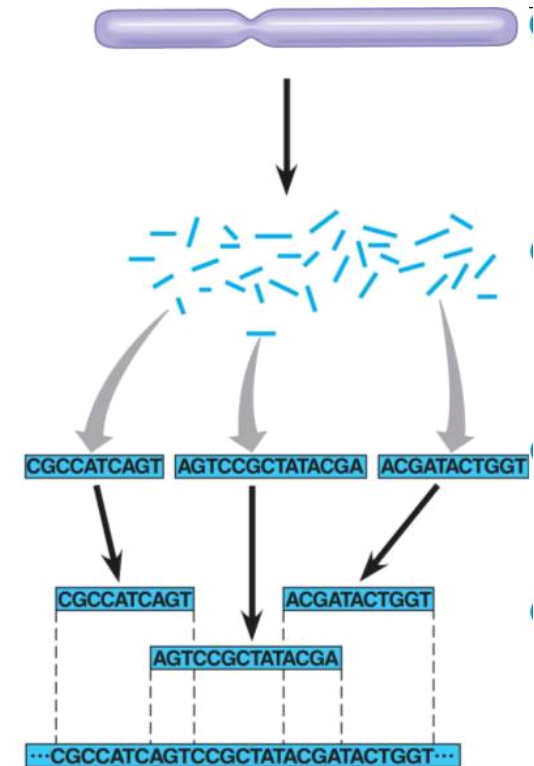


CATTATCTGTTTTTGCTATTAATGTCATACCTGCTATTCAGACAAATCAATTTGCTCTCTTAATAAAGGATG  
AGCTTCCTGTAGCTTTTTGTAGCTGGGCCAGTTTAGATCTGGAATGTGAGGTAAAATATATAAATGAT

- **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?

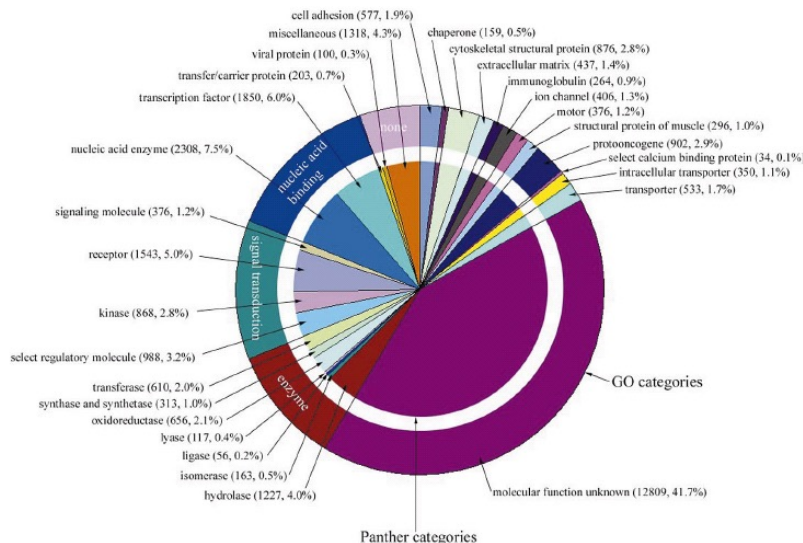
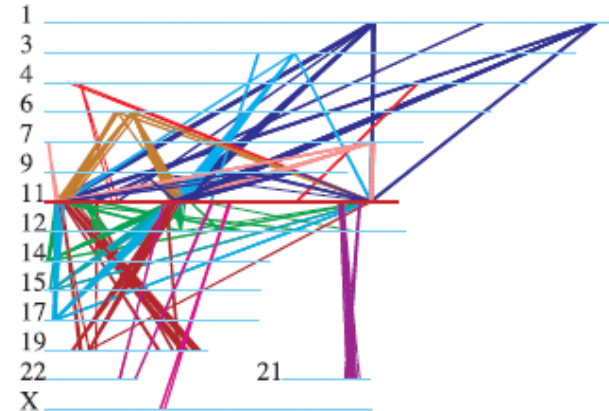


- Gen-Vorhersage und Annotation

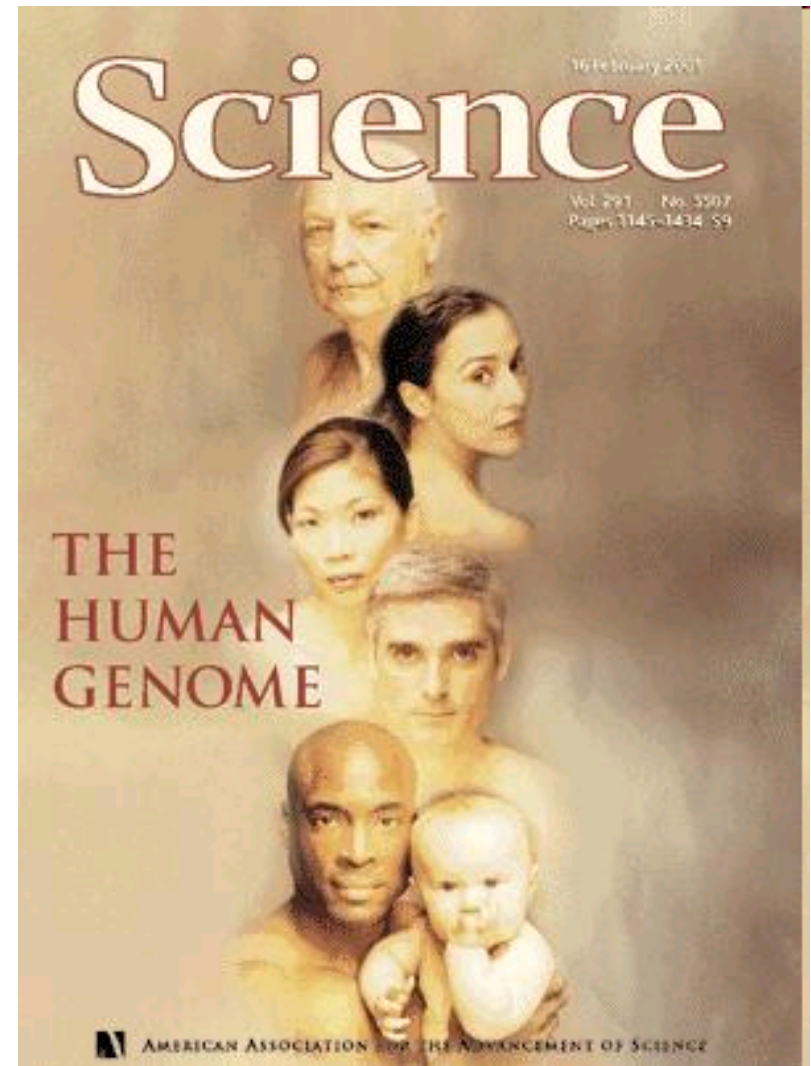
ACGGACGGCTATATACGACGGGACTACACGGT



- Analyse der Genom-Struktur und Evolution



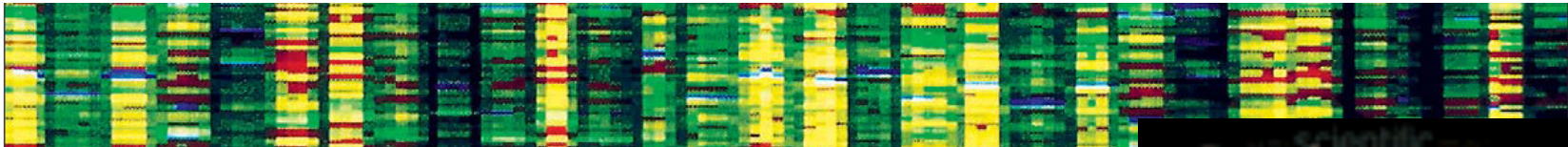
- Proteinfunktion: Vorhersage und Analyse





- 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





...en 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

THE EDITORIAL REVIEW

### Initial impact of the human genome

Eric S. Lander<sup>1</sup>

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 paper<sup>1</sup> 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.

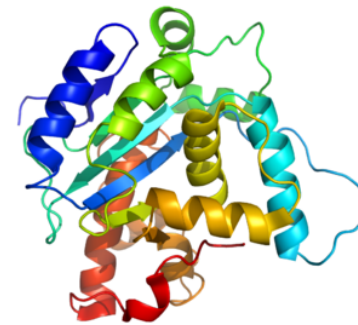
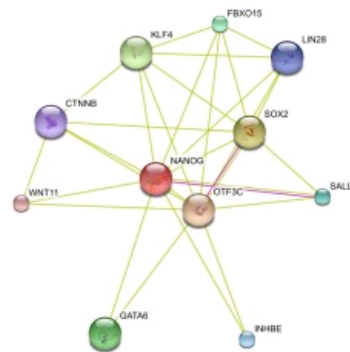
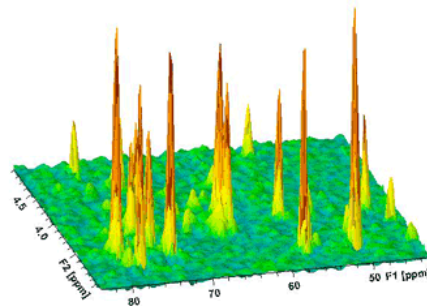
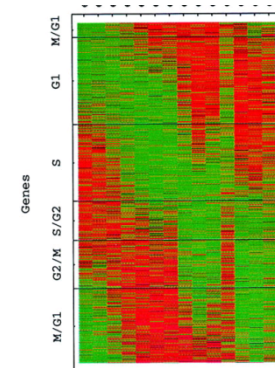
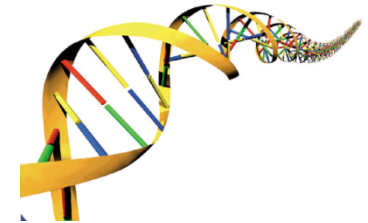
This is a monumental month for human genomics as we celebrate 10 years since the publications of the human genome. It is time to reflect on the advances that this endeavor has brought to mankind. The ability for scientists to generate a complete human genome sequence meant that, for the

### ...ome, Part II

...celebrating the 10th anniversary of the human genome sequence. We explore the impacts of genomics on ourselves, as well as on other species, and Laura M. Zahn

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

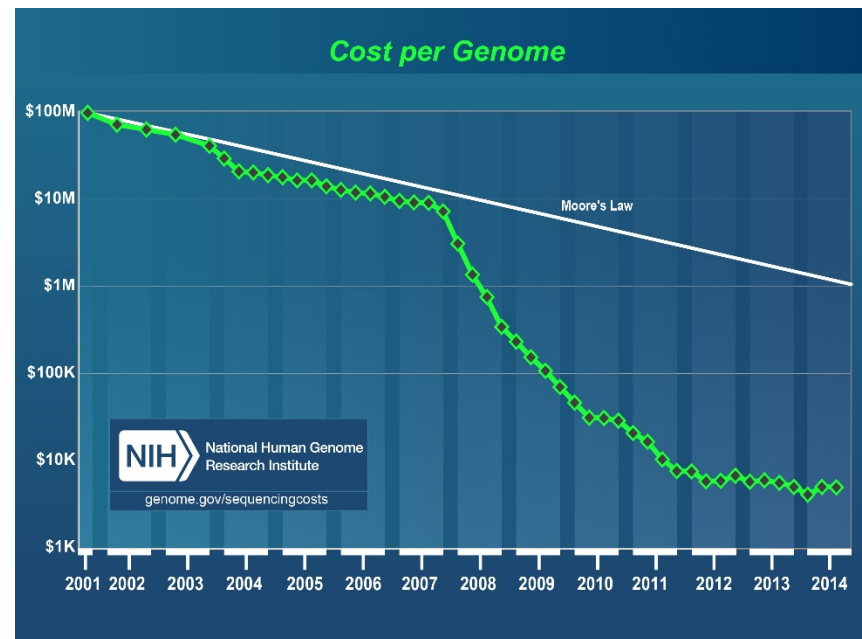
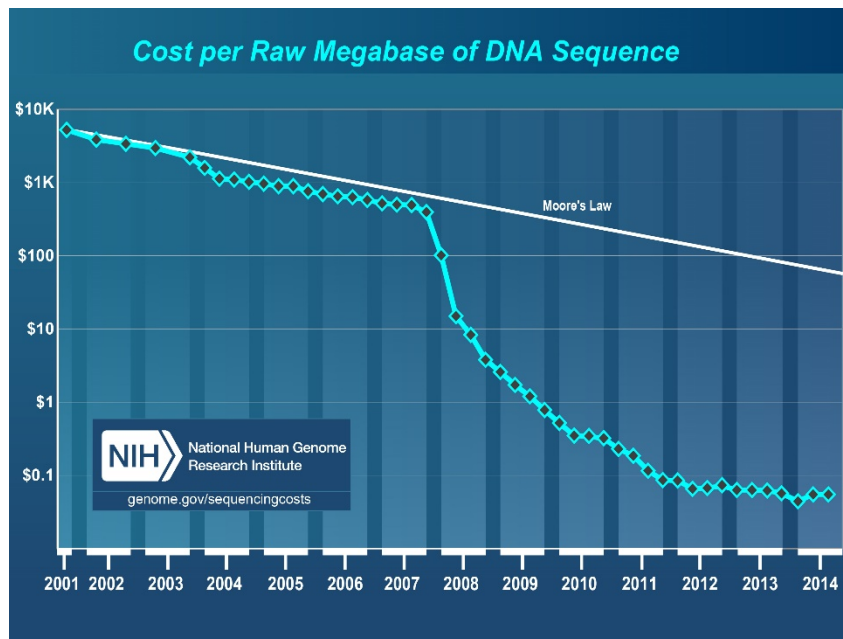
- Omics = Genomweite Studien
- Genomics = Alles was es im Genom gibt
- Transcriptomics = Alle Transkripte / mRNA
- Proteomics = Alle Proteine
- Interactomics = Alle Interaktionen
- Metabolomics = Alle Metabolite
- Kinomics = Alle Kinasen
- ...



# Next-Generation-Sequencing: Das Comeback der Sequenzierung



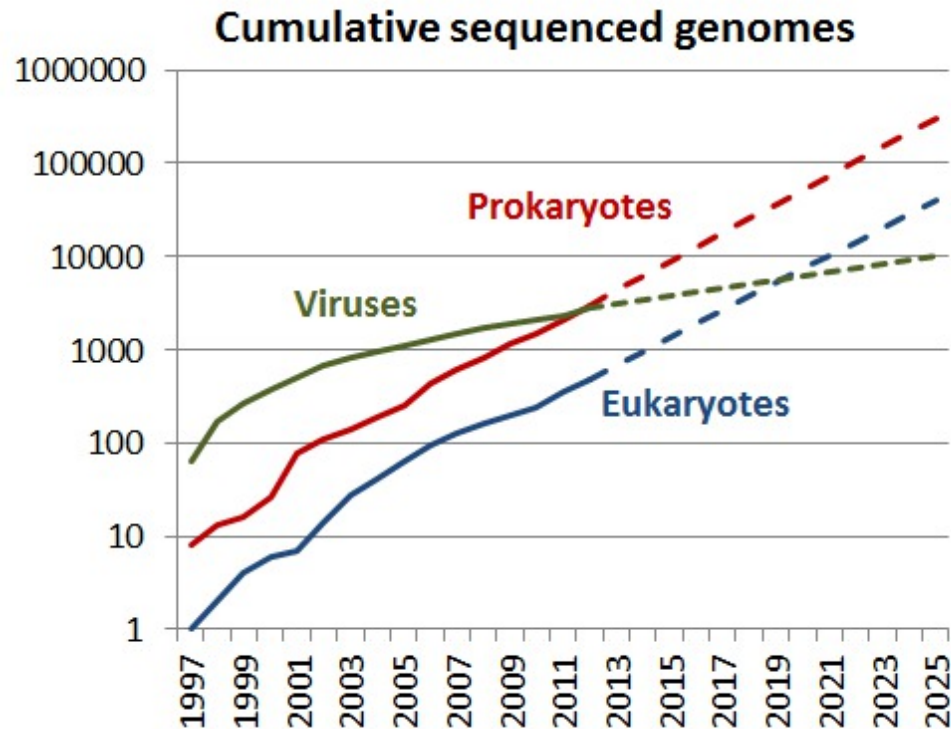
- Schnell und einfach zu bedienen
- Hoher Durchsatz
- ⇒ Sequenzieren wird immer billiger



# Genome über Genome



- Eukaryoten: 1842 (2015)
- Prokaryoten: 30684 (2015)
- Viren: 4465 (2015)

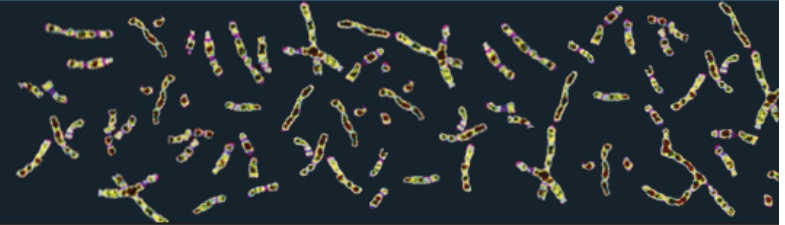


# Genome über Genome

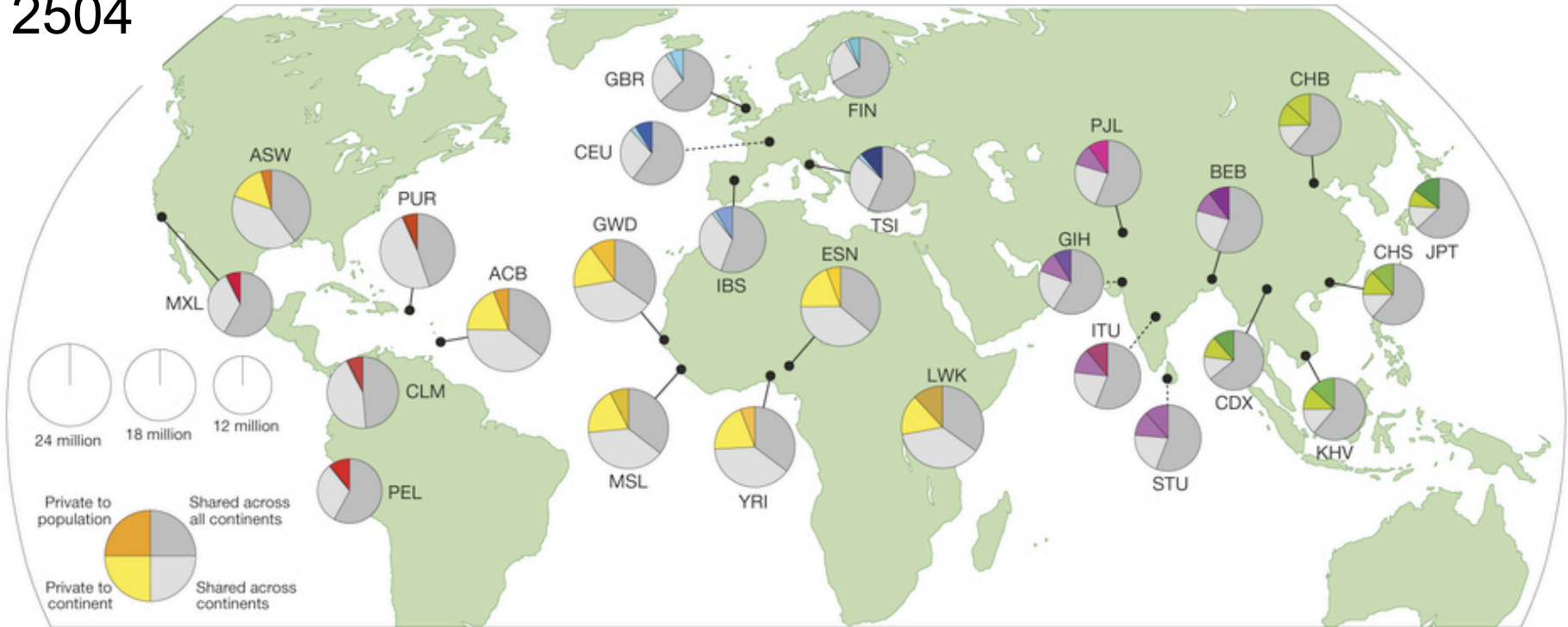


## 1000 Genomes

A Deep Catalog of Human Genetic Variation



2504





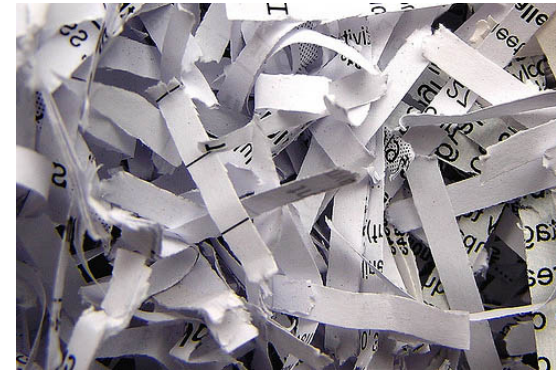
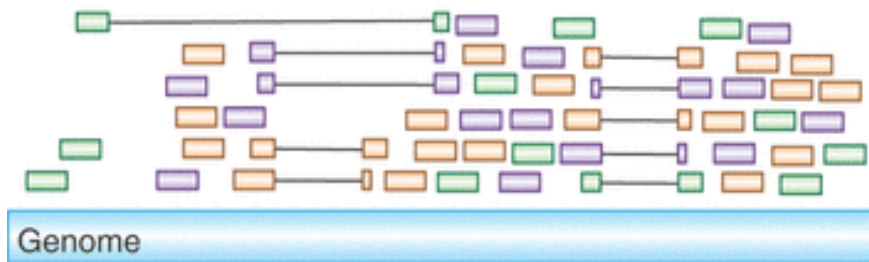
# Beispiel-Fragestellung: RNA Sequenzierung (RNA-seq)



Eingabe: Kurze  
Sequenzier-  
Reads

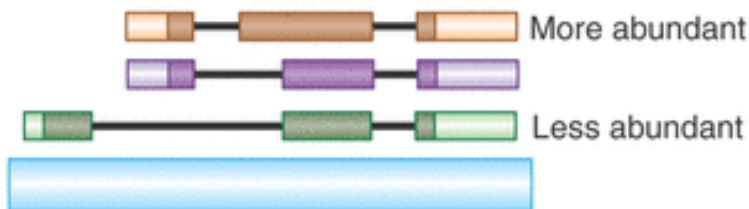


Alignment gegen das  
Genom



Welche mRNAs  
gibt es?

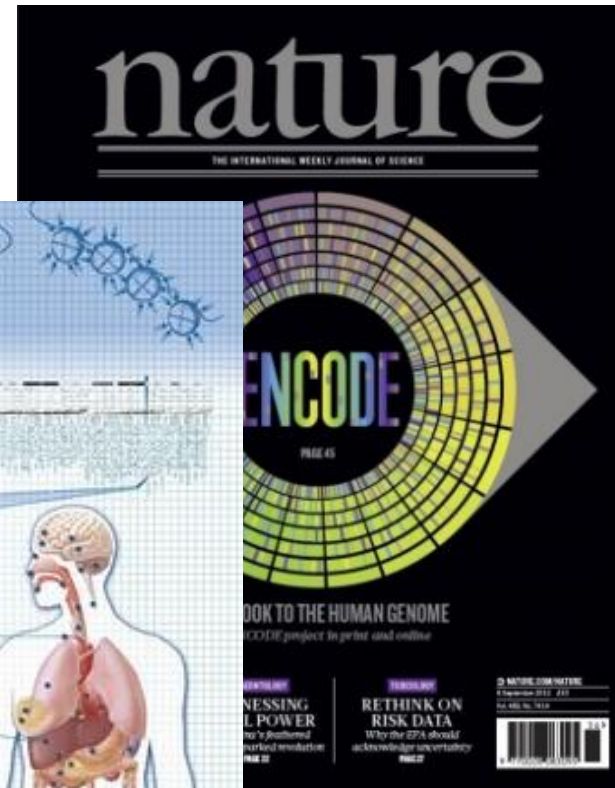
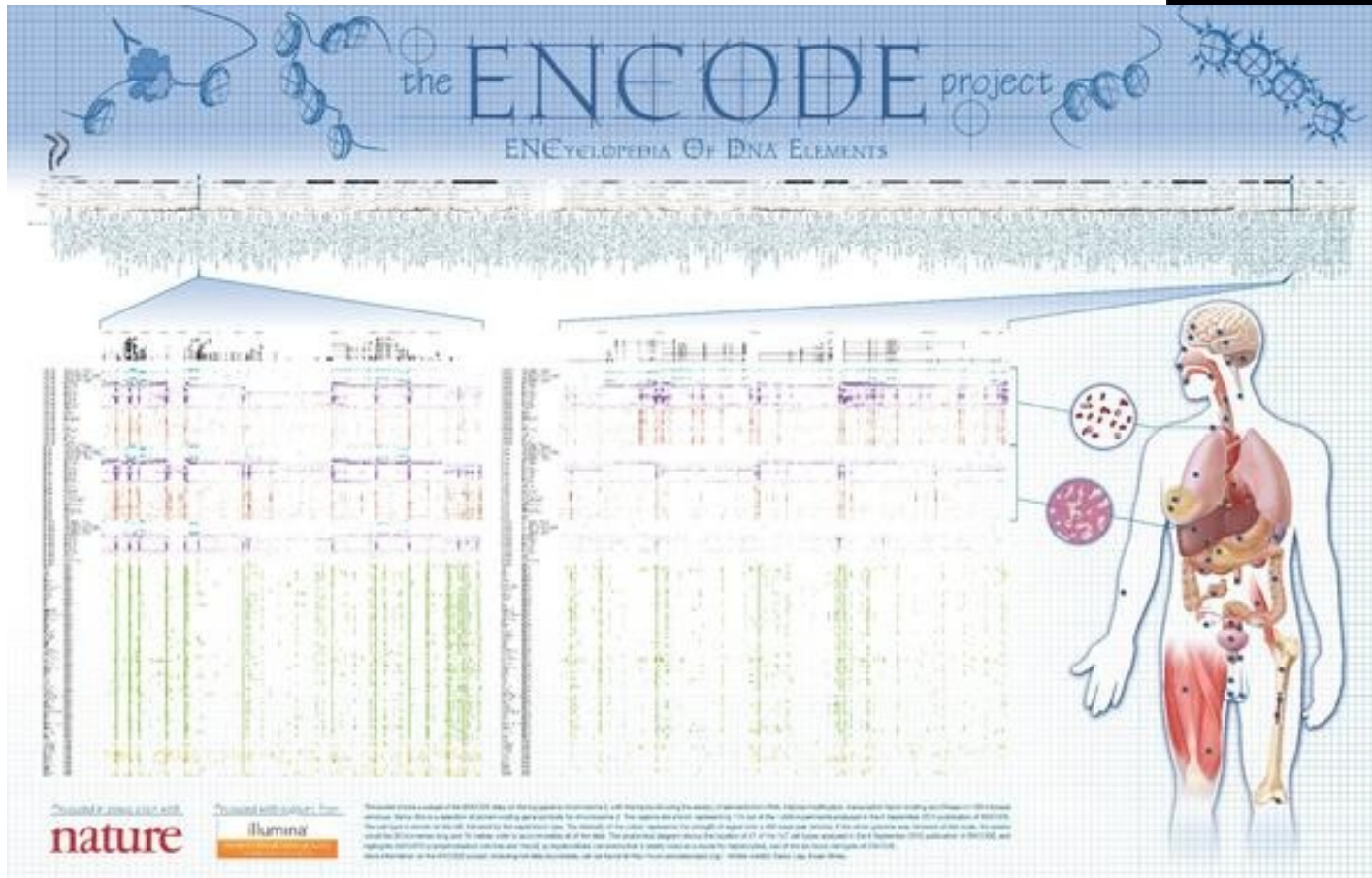
Wie häufig  
sind sie?



# ENCODE



- The ENCODE Project:  
ENCyclopedia Of DNA Elements







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



**HiSeq 4000**

125-1500 Gb

<1-3.5 days



**HiSeq X Five\***

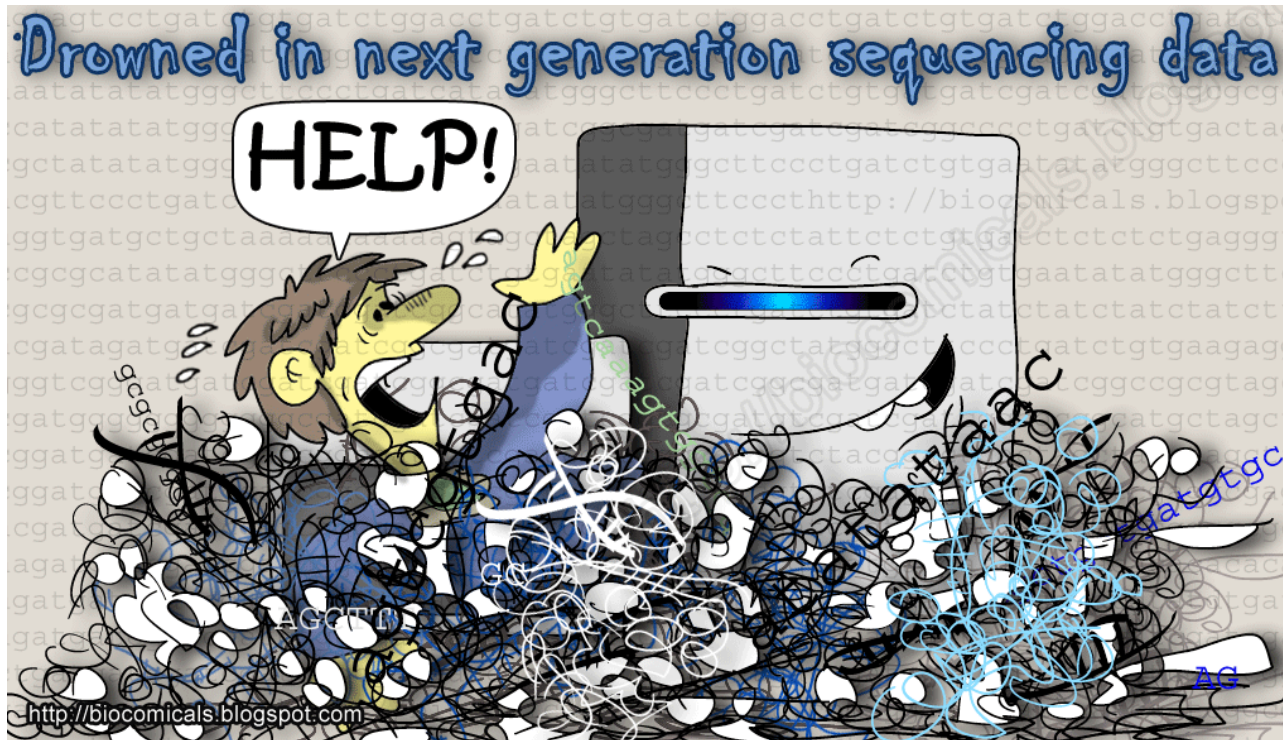
900-1800 Gb

<3 days

Datenmenge  
pro Run

Dauer

# Next-Generation-Sequencing: Das Comeback der Bioinformatik



Ohne Bioinformatik geht in der  
(experimentellen) Biologie nichts mehr

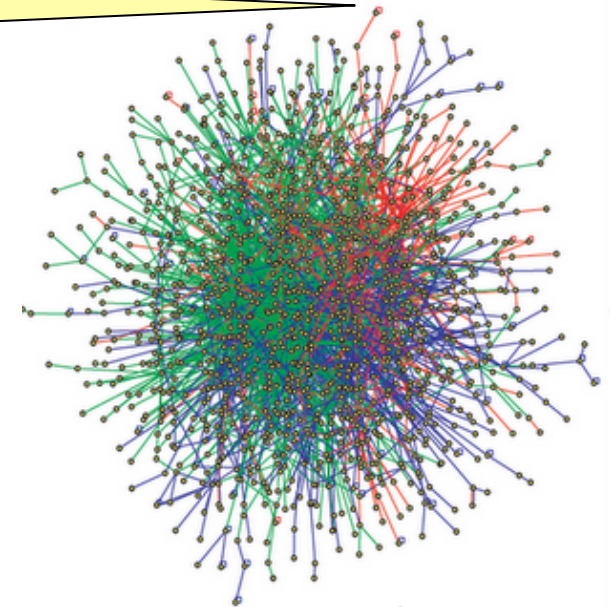
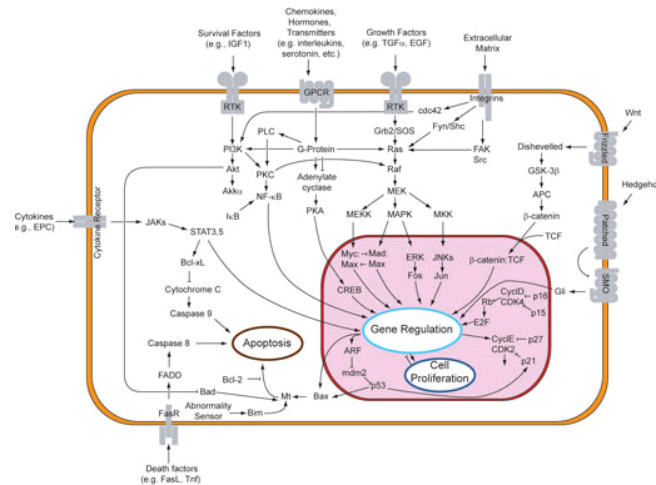
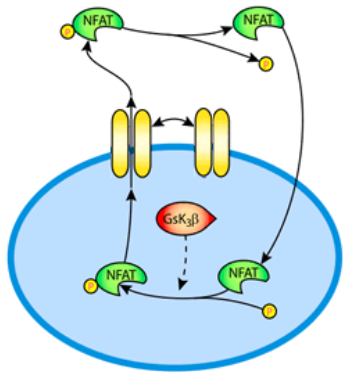


# Ziel: Von Daten zur Systembiologie



- Möglichst vollständige Beschreibung eines biologischen Systems

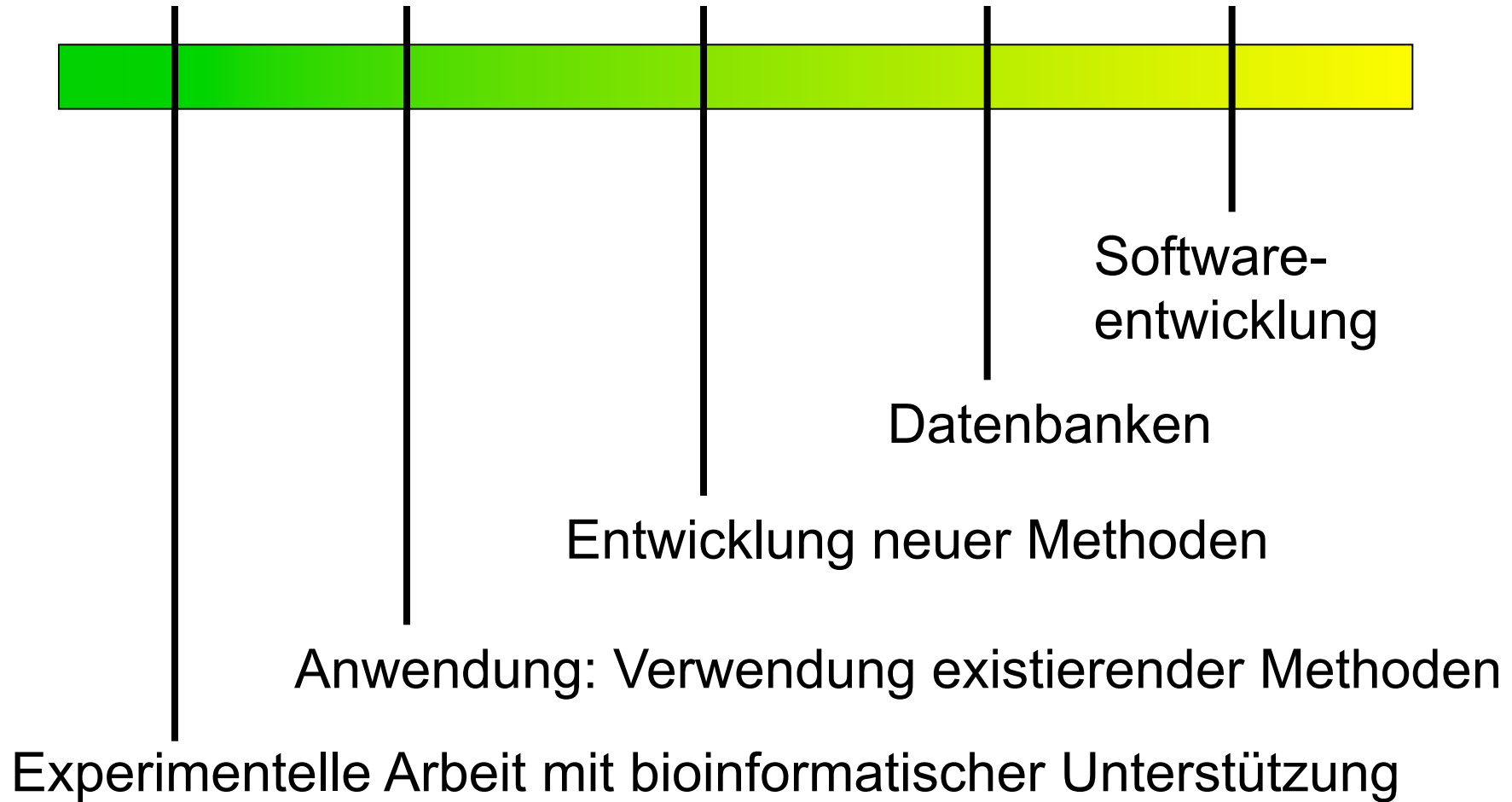
Detailgrad



Größe des Systems

Biologie

Informatik



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



# Studienplan Bioinformatik Bachelor



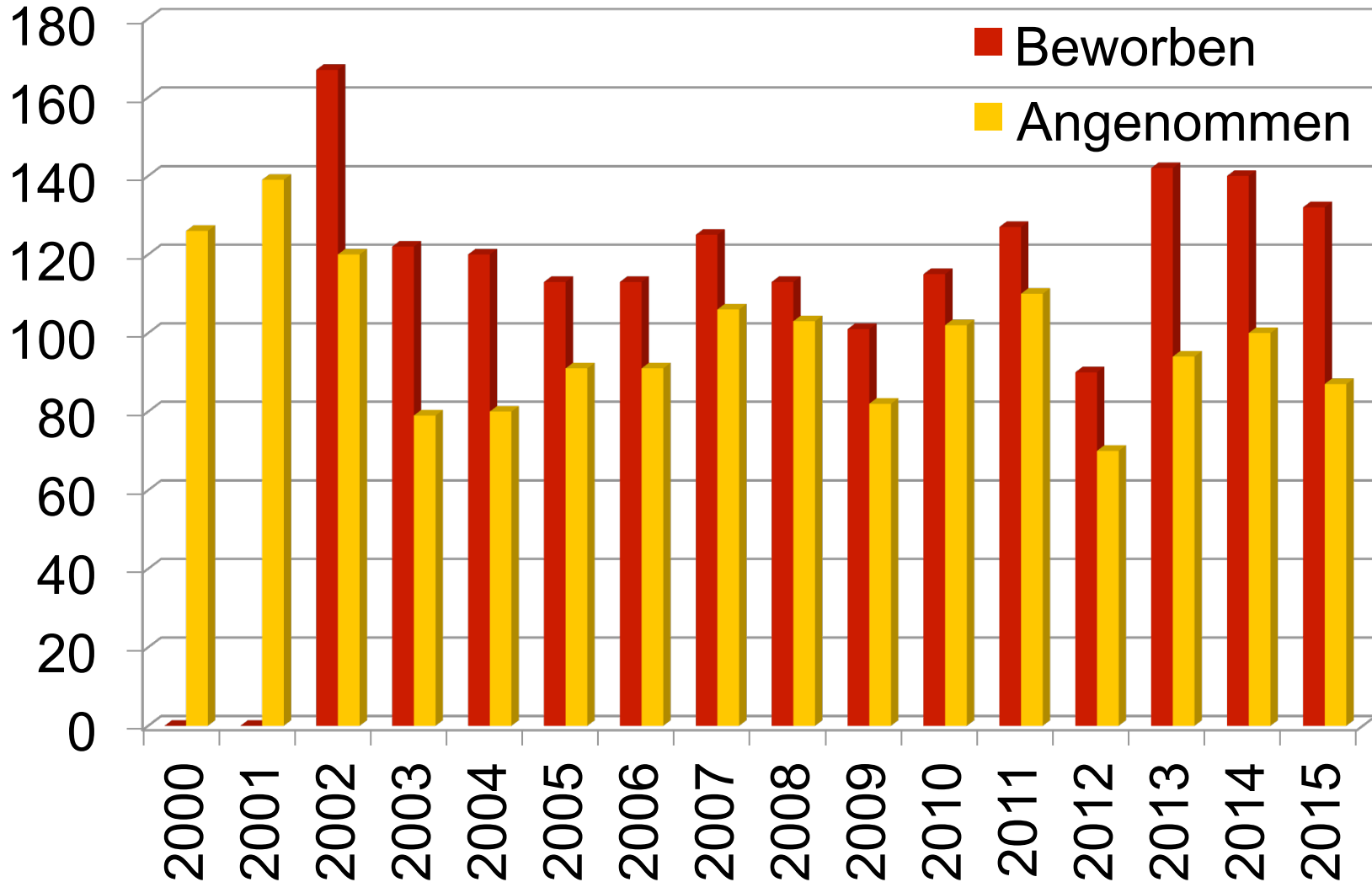
	Informatik	Mathematik	Bioinformatik	Biologie, Chemie
1	Programmierung	Analysis	Einführung in die Bioinformatik I, Tutorium	Biologie I Chemie I
2	Algorithmik	Logik & Diskrete Strukturen	Einführung in die Bioinformatik II, Proseminar	Biologie II Chemie II Biochemie I
3		Lineare Algebra	Programmier-Praktikum	Biochemie II, Praktikum
4	Theoretische Informatik	Stochastik & Statistik	Algorithm. Bioinfo. I, Bioinfo.-Ressourcen	
5	Wahlpflichtveranstaltungen		Algorithm. Bioinfo. II, Hauptseminar, Praktikum	
6			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

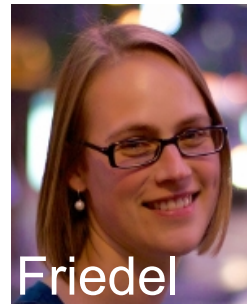
# Annahmequoten



## Lehrstühle



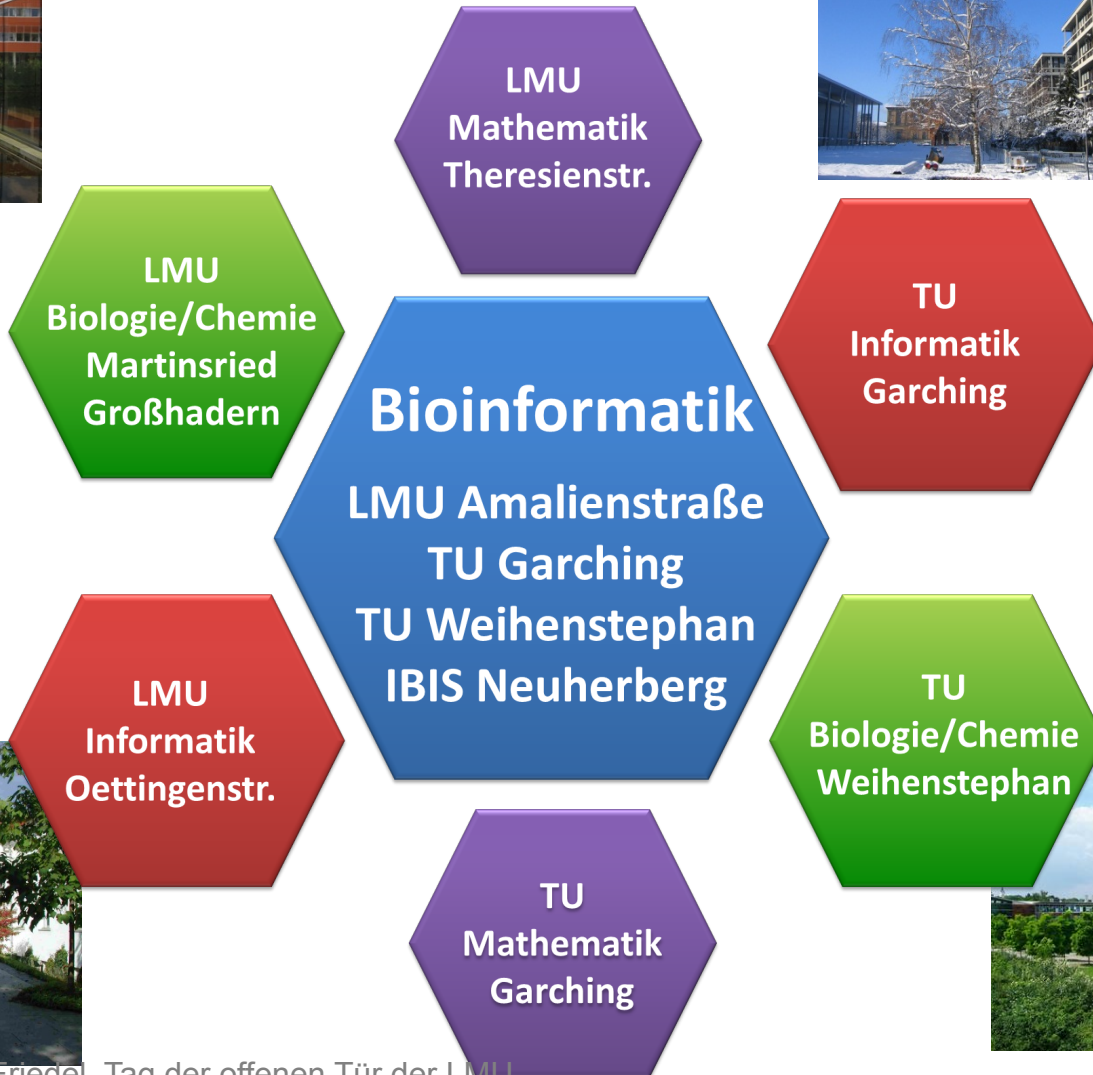
## Professoren für Bioinformatik



## Assoziierte Gruppen



# Standorte



# Standorte



LMU  
Biologie/Chemie  
Martinsried  
Großhadern

LMU  
Mathematik  
Theresienstadt

Bioinformatik



Weihenstephan  
Mühlerberg

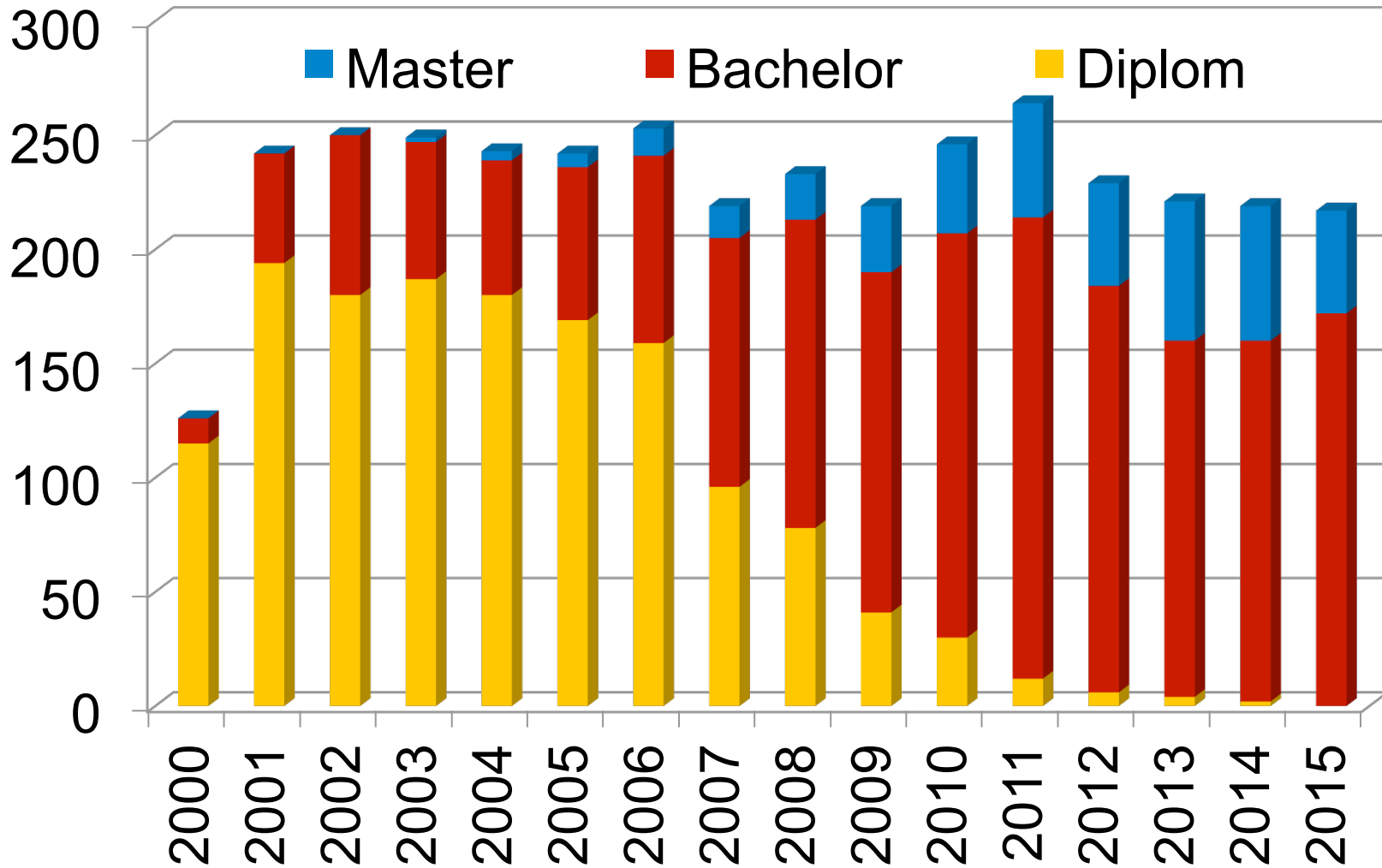
TU  
Biologie/Chemie  
Weihenstephan



TU  
Chemie  
München



# Studierende





- Web-Seite zum gemeinsamen Studiengang:
  - [www.bioinformatik-muenchen.de/](http://www.bioinformatik-muenchen.de/)
- Web-Seite zum EFV mit Formularen:
  - [www.bio.ifi.lmu.de/EFV/](http://www.bio.ifi.lmu.de/EFV/)
- Fachschaft Bioinformatik:
  - [www.bioinformatik-muenchen.com/](http://www.bioinformatik-muenchen.com/)
- Facebook-Seite
  - <http://www.facebook.com/Bioinformatik.Muenchen>
- Informationsstand: HS A 120

Noch Fragen ?