

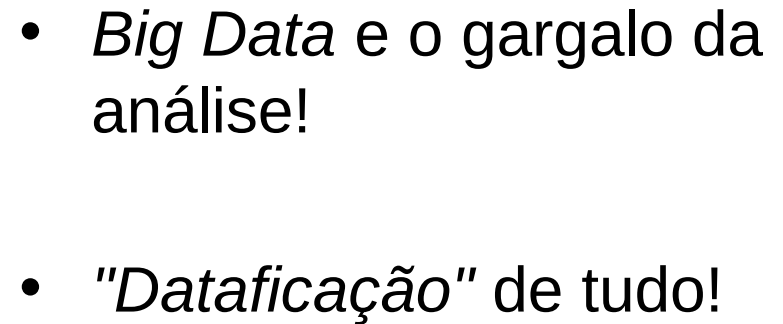
Ciência de Dados Aplicada à Saúde

Código: IAA019

Introdução:

Big Data e Inovação em Saúde

Mauro Castro



< <https://blogs.thomsonreuters.com/> >



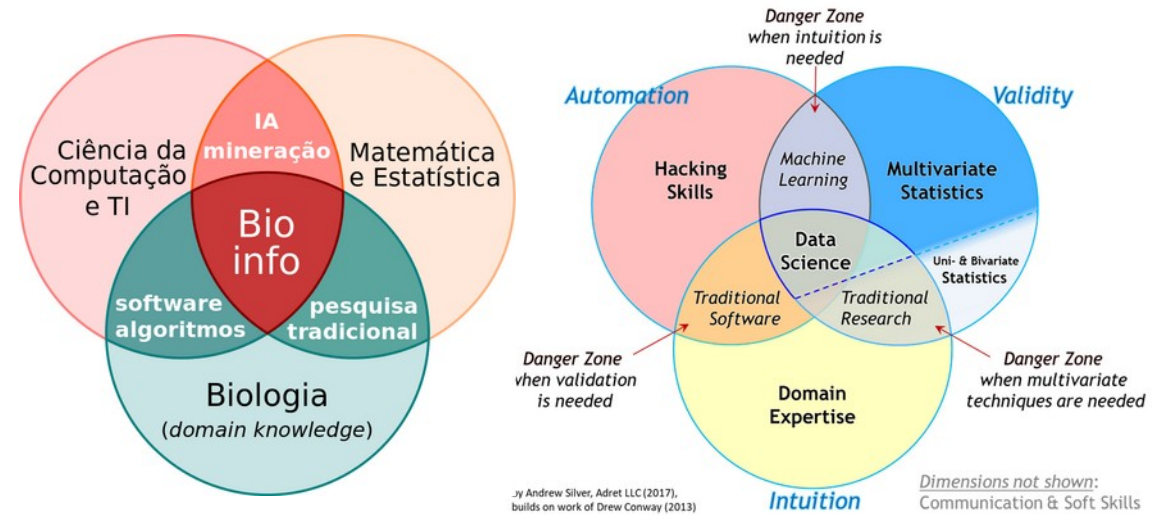
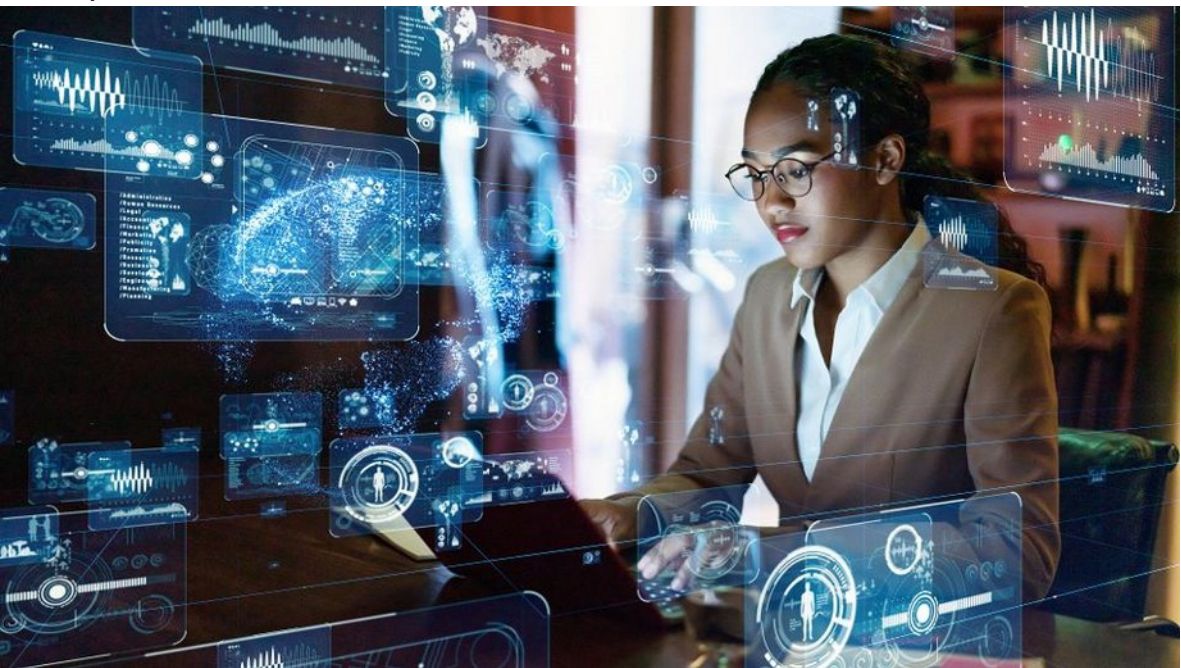
Marr B. **Forbes**, MAR 15, 2016.
< <http://www.forbes.com/> >

- O volume de dados continuará a crescer!
- A carência de pessoal vai expandir.
- *Dados rápidos* e *acionáveis* substituirão o *Big Data*.

Why Data Scientists Aren't Data Engineers.

Walch, K. *Forbes*, Dec 1, 2019.

< <http://www.forbes.com/> >



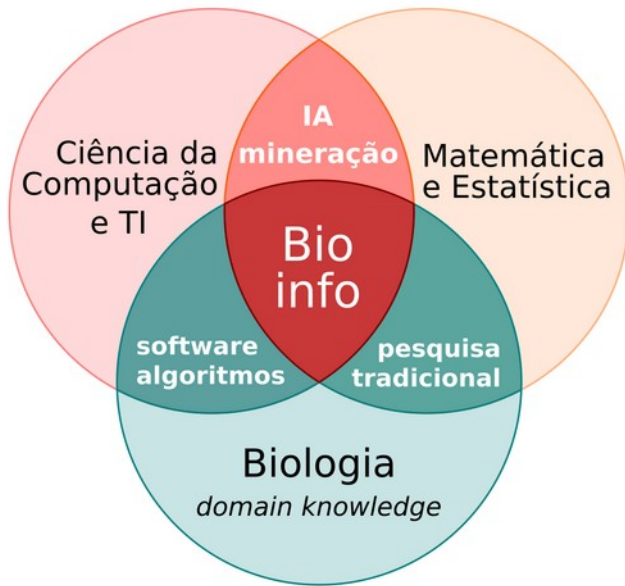
Cientista de Dados	Engenheiro de Dados
Foco	
Cria e testa hipóteses; Analisa e traduz dados “limpos”!	Desenha, constrói e organiza dados a serem processados!
Habilidades	
Estatista e matemática avançadas	Programação avançada
ML / IA	Sistemas e pipelines
Análise Programação Big Data	
Linguagens de Programação e Ferramentas	
R, SAS, Python	Java, Python, Scala...
Formação Básica	
Estatística Probabilidade Matemática Algoritmos	Programação Integração Arquitetura de sistemas

Ciência de Dados e Bioinformática?

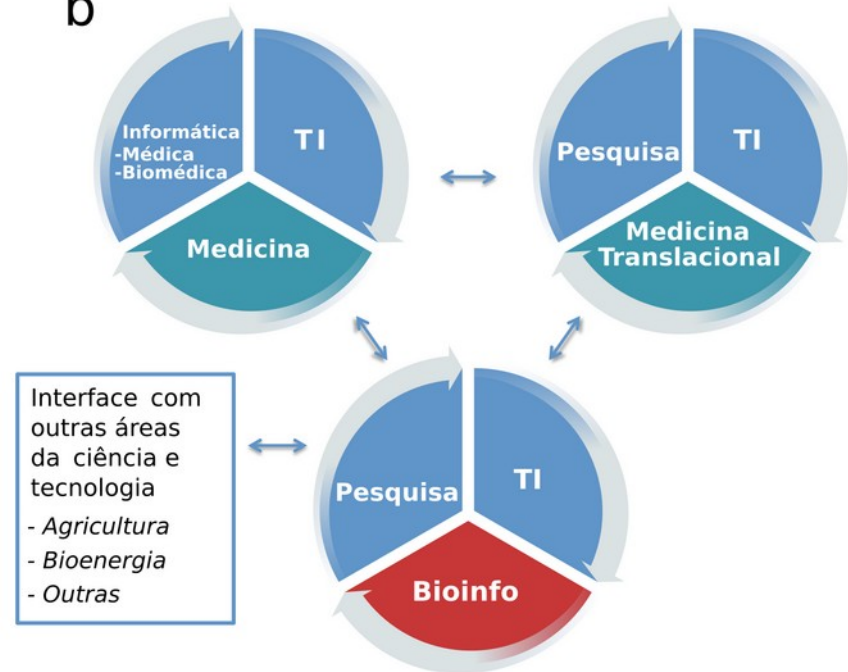


O que é Bioinformática?

a

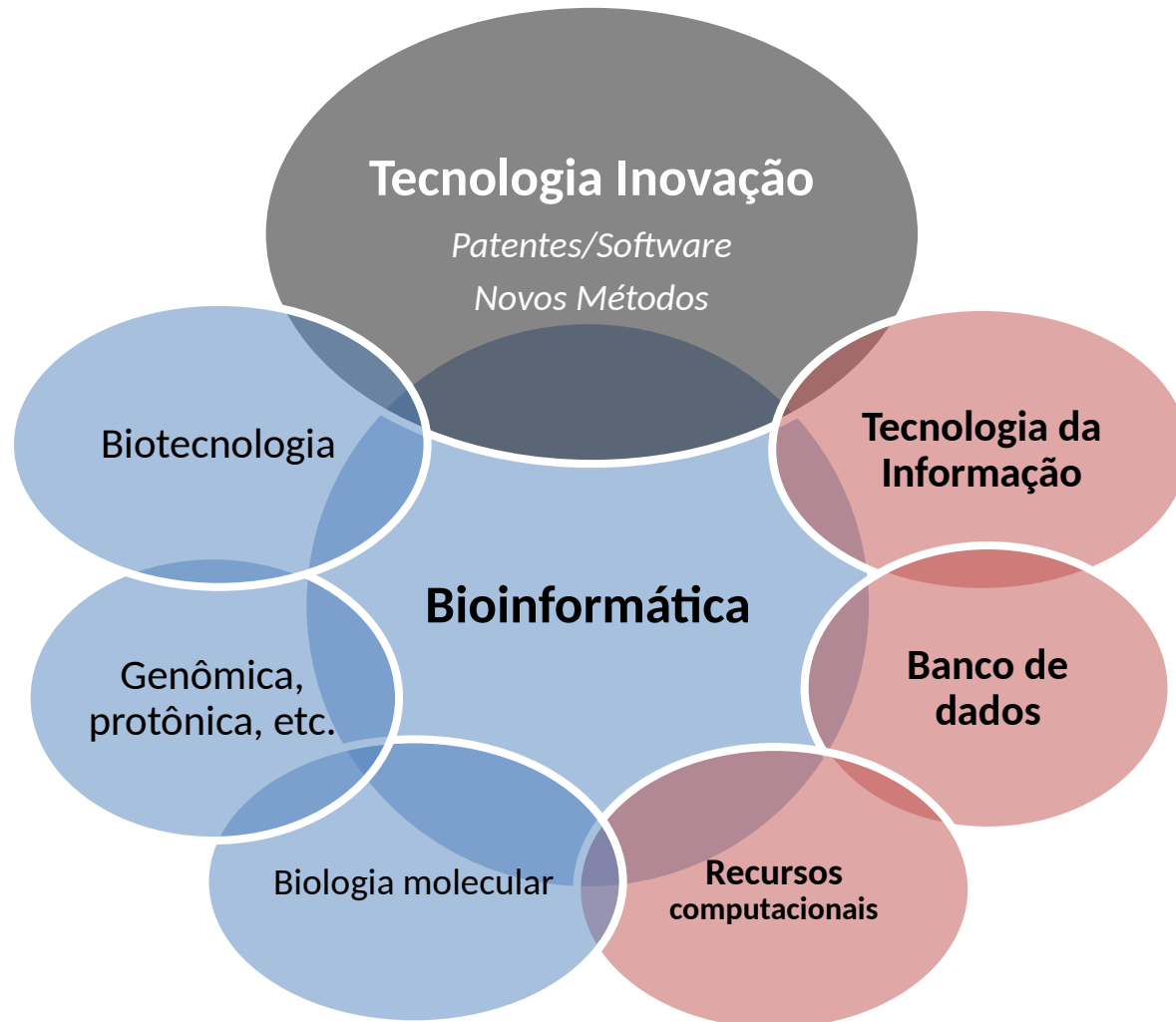


b



<http://www.bioinfo.ufpr.br/guia-do-estudante.html>

O que é Bioinformática?



Competências na área de ciência de dados para análise de dados biológicos

Biologia

Genética
Bioquímica
Evolução
Biotecnologia
Biologia Celular
Biologia Molecular
Biofísica
Genômica
Transcriptômica
Proteômica
Biologia de Sistemas
Outras

Ciência da Computação

Algoritmos e Programação
Banco de Dados
Inteligência Artificial
Aprendizado de Máquina
Outras

Matemática/Estatística

Probabilidade
Estatística
Bioestatística
Matemática biológica
Outras

Outras Ciências, e.g

Física
Química

Outros requisitos, e.g

Seminários
Pesquisa
Estágios

Não-científicos, e.g

Ética
Empreendedorismo
Escrita/comunicação

Welch et al., *PLOS Computational Biology*, Vol. 10, Issue 3, 2014.

Message from ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch^{1*}, Fran Lewitter², Russell Schwartz³, Cath Brooksbank⁴, Predrag Radivojac⁵, Bruno Gaeta⁶, Maria Victoria Schneider⁷

1 School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, United States of America, **2** Bioinformatics and Research Computing, Whitehead Institute, Cambridge, Massachusetts, United States of America, **3** Department of Biological Sciences and School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **4** European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **5** School of Informatics and Computing, Indiana University, Bloomington, Indiana, United States of America, **6** School of Computer Science and Engineering, The University of New South Wales, Sydney, New South Wales, Australia, **7** The Genome Analysis Centre, Norwich Research Park, Norwich, United Kingdom



The skill sets required for success in the field of bioinformatics are considered by several authors: Altman [2] defines five broad areas of competency and lists key technologies; Ranganathan [3] presents highlights from the Workshops on Education

life sciences curricula. Pevzner and Shamir [11] propose that undergraduate biology curricula should contain an additional course, "Algorithmic, Mathematical, and Statistical Concepts in Biology." Wingren and Botstein [12] present a graduate

Computational Biology
Programming
Machine Learning
Data structure
Genetics
Biochemistry
Molecular Biology
Math/Statistics
Probability
Biostatistics
Other Sciences

Bioinformatics User



Leon

Bioinformatics Scientist



Martha

Bioinformatics Engineer

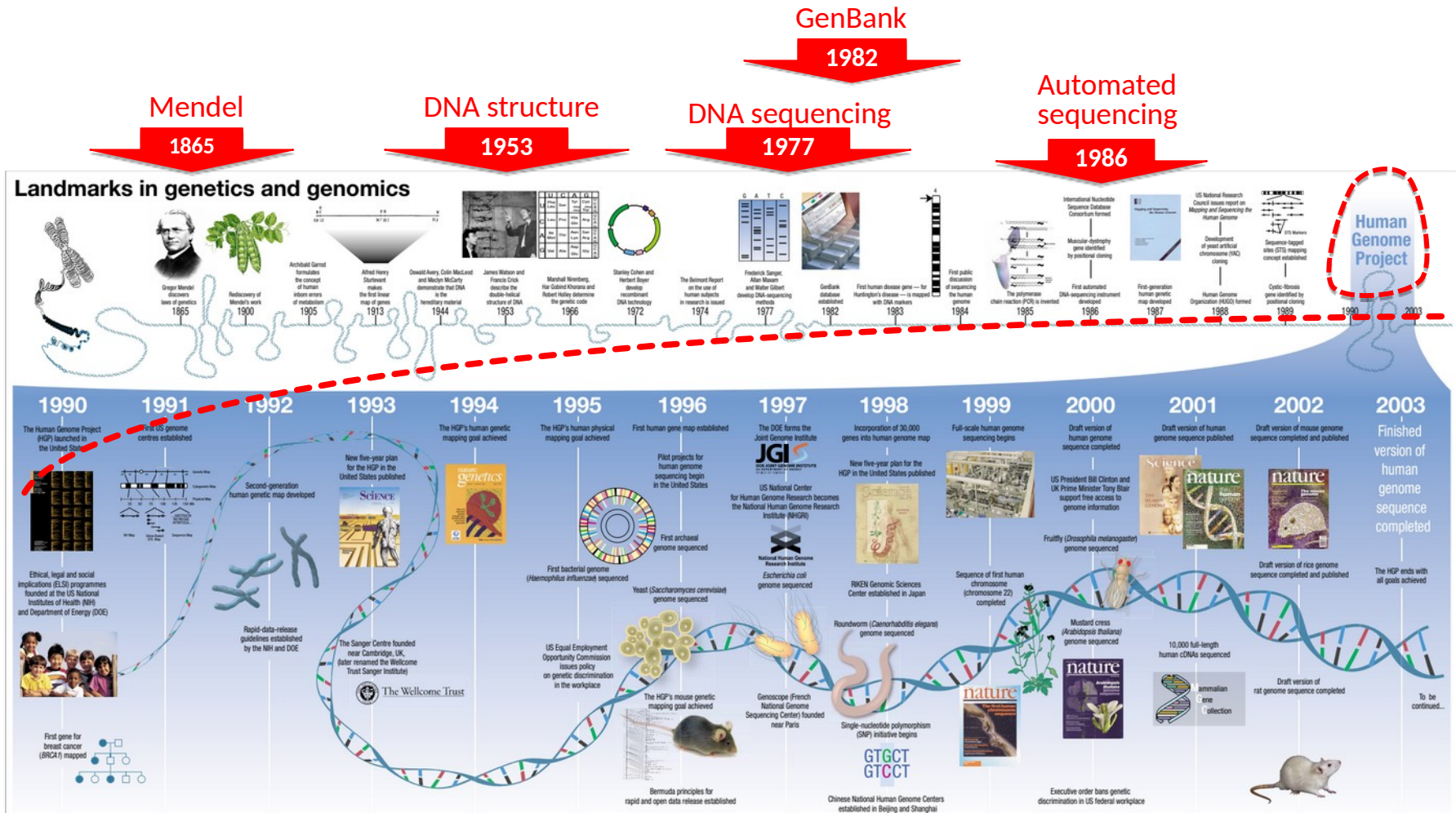


Ivan

Welch et al., *PLOS Computational Biology*, Vol. 10, Issue 3, 2014.

A vision for the future of genomics research

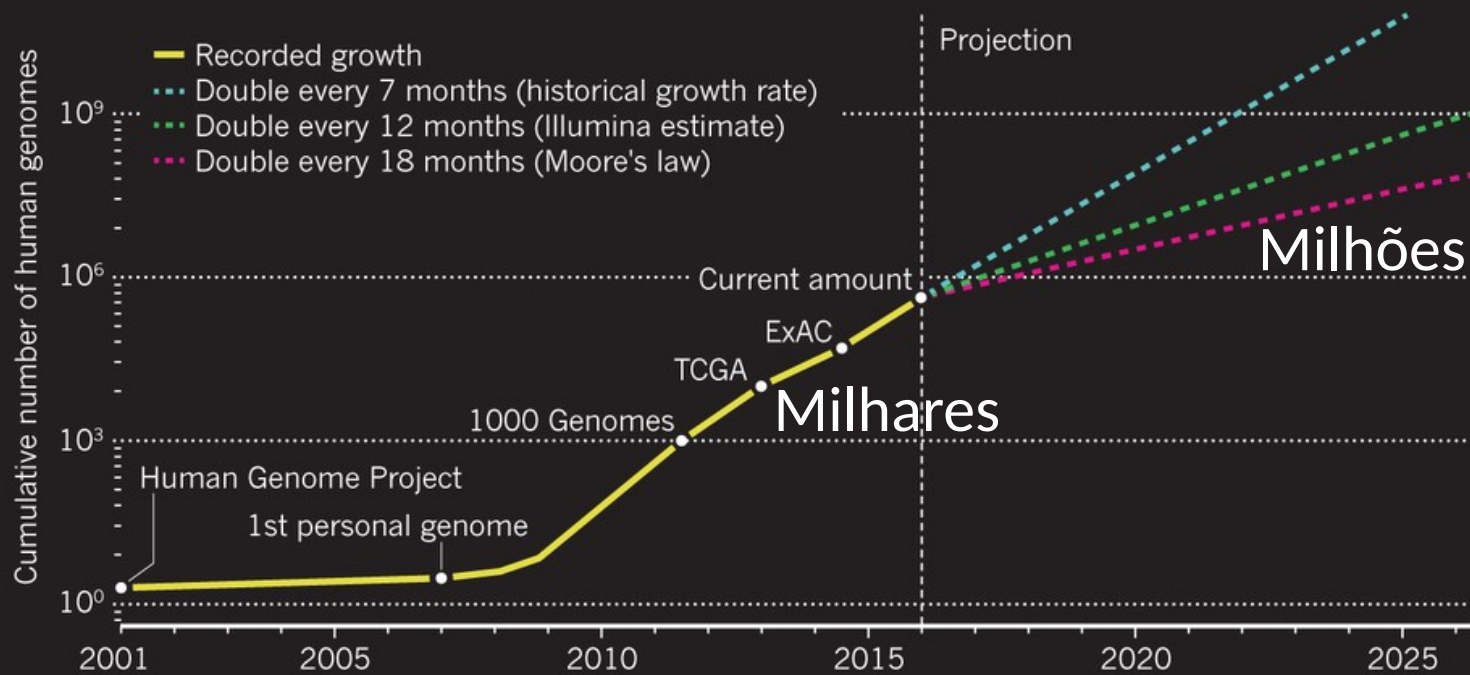
...back then 2003!



Nature 422, 835-847 (24 April 2003)

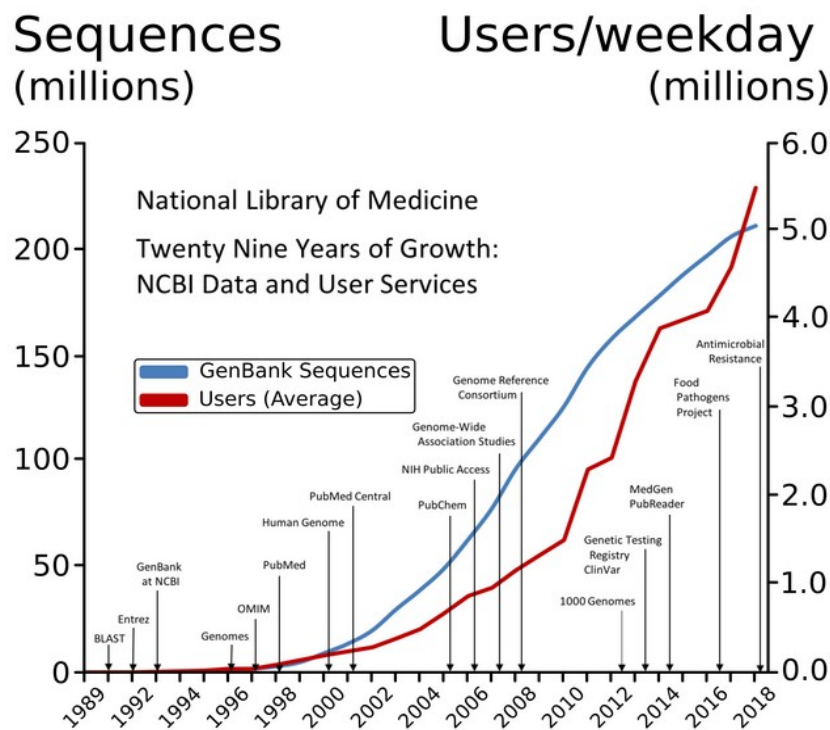
DNA SEQUENCING SOARS

Human genomes are being sequenced at an ever-increasing rate. The 1000 Genomes Project has aggregated hundreds of genomes; The Cancer Genome Atlas (TCGA) has gathered several thousand; and the Exome Aggregation Consortium (ExAC) has sequenced more than 60,000 exomes. Dotted lines show three possible future growth curves.



Nature Vol 527 / November 2015

- *Big Data* no GeneBank
(1989 – 2019)

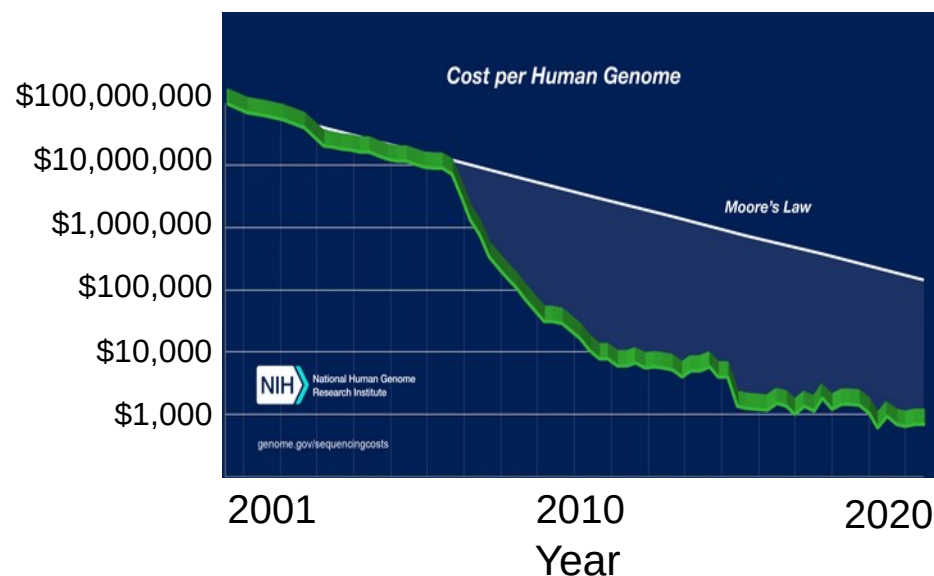


Year (adapted from the HHS/NIH)

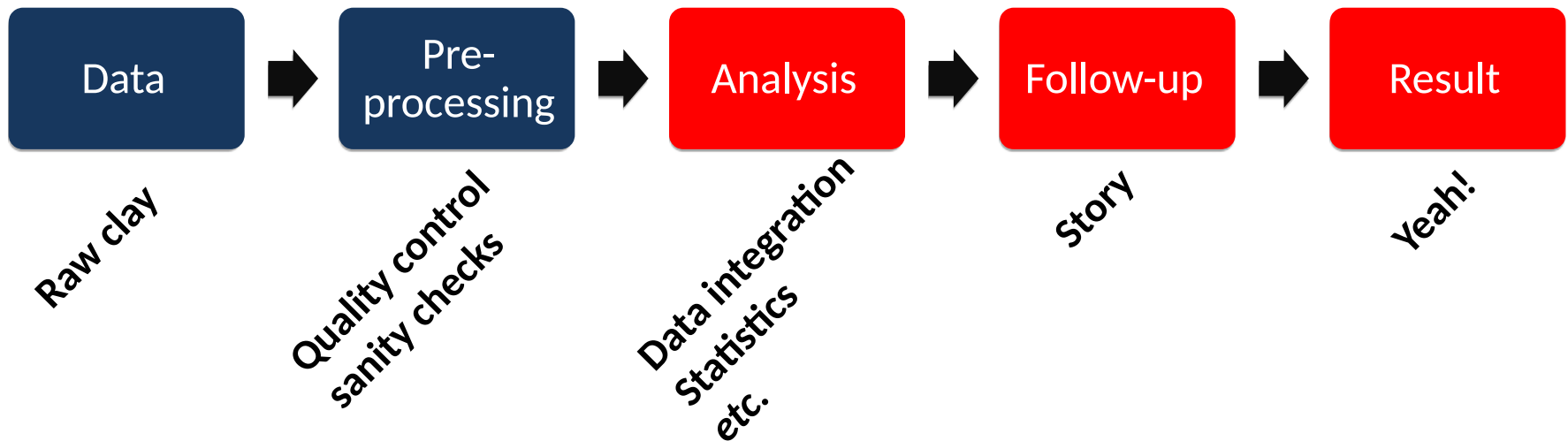
<https://www.nlm.nih.gov/about/2021CJ_NLM.pdf>

Accessed [03 May 2021]

- Estimativa de custo por genoma (2001 – 2020)



Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).
<www.genome.gov/sequencingcostsdata>
Accessed [03 May 2021].



Estamos aqui!

RESEARCH ARTICLE

CANCER

The chromatin accessibility landscape of primary human cancers

M. Ryan Corces^{1*}, Jeffrey M. Granja^{1,2,3*}, Shadi Shams¹, Bryan H. Louie¹, Jose A. Seoane^{2,4,5}, Wanding Zhou⁶, Tiago C. Silva^{7,8}, Clarice Groeneveld⁹, Christopher K. Wong¹⁰, Seung Woo Cho¹, Ansuman T. Satpathy¹, Maxwell R. Mumbach^{1,2}, Katherine A. Hoadley¹¹, A. Gordon Robertson¹², Nathan C. Sheffield¹³, Ina Felau¹⁴, Mauro A. A. Castro⁹, Benjamin P. Berman⁷, Louis M. Staudt¹⁴, Jean C. Zenklusen¹⁴, Peter W. Laird⁶, Christina Curtis^{2,4,5}, The Cancer Genome Atlas Analysis Network[†], William J. Greenleaf^{¶1,2,3,15,16}, Howard Y. Chang^{1,2,17,18}

Corces et al. *Science*, 362(6413):eaav1898, 2018.



RTN

RedeR

Geneplast

RTNsurvival

<http://www.bioconductor.org/>

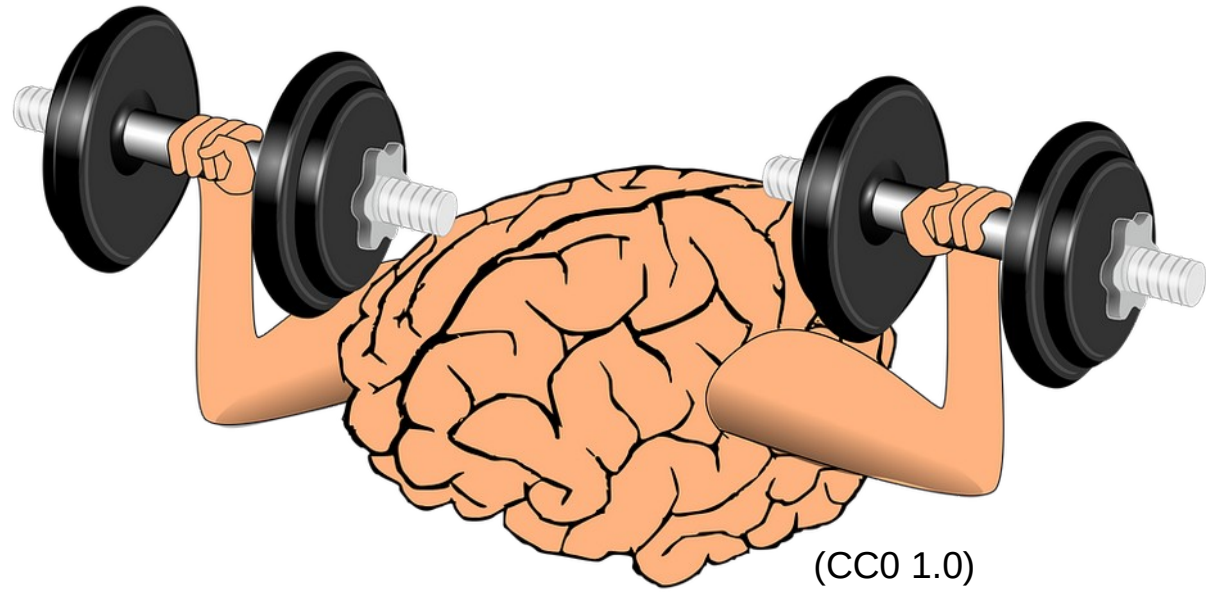
- Ótimas fotos!
Que equipamento
você usou?



- Um cérebro,
um olho e
um dedo.



andrade
etriseviverdehumor.blogspot.com



<https://imgflip.com/i/4yqmqd>

Obrigado!