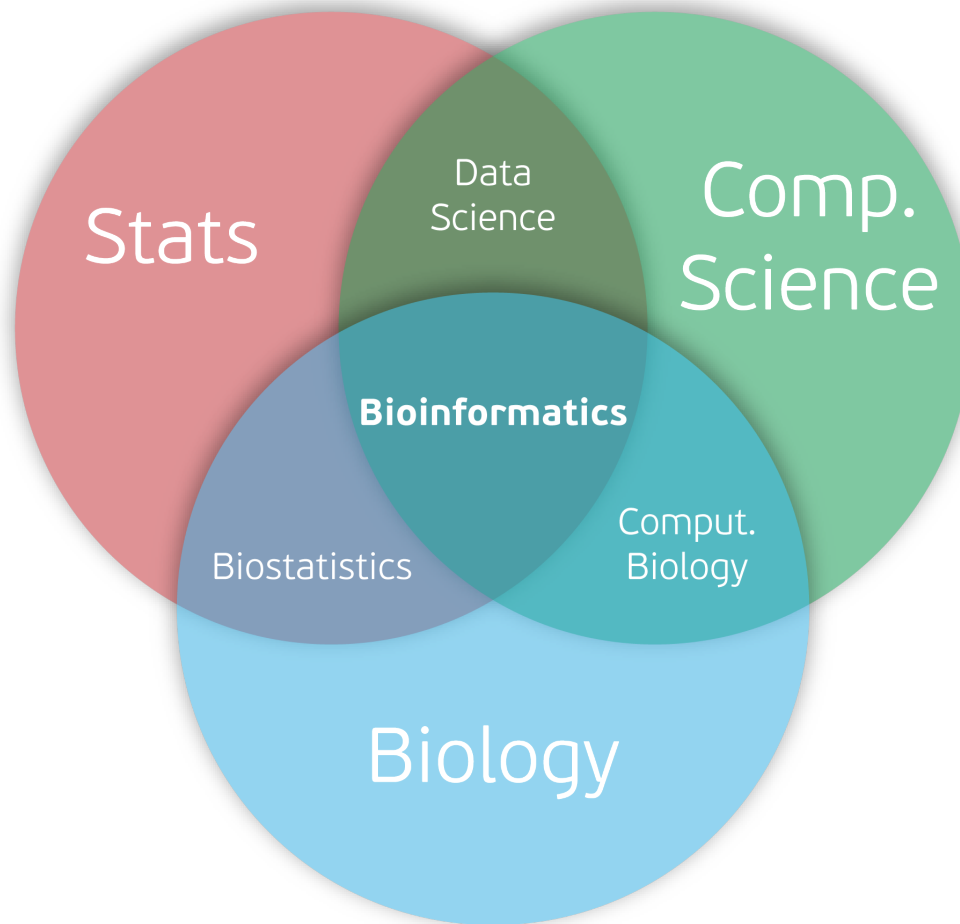




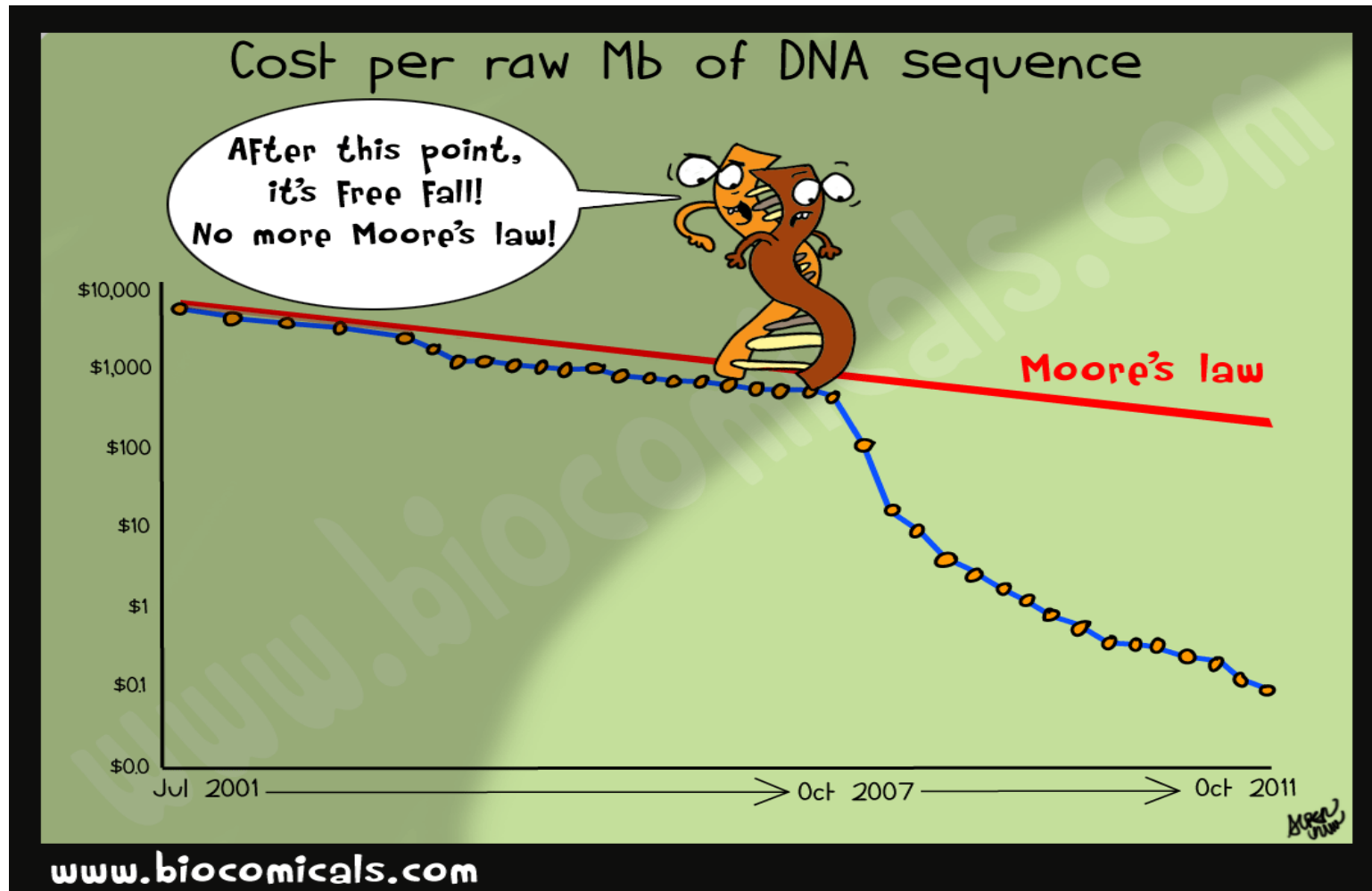
Computational Biology and Bioinformatics

McGill CB2
(Macdonald campus)

What is bioinformatics?



Cost of generated data

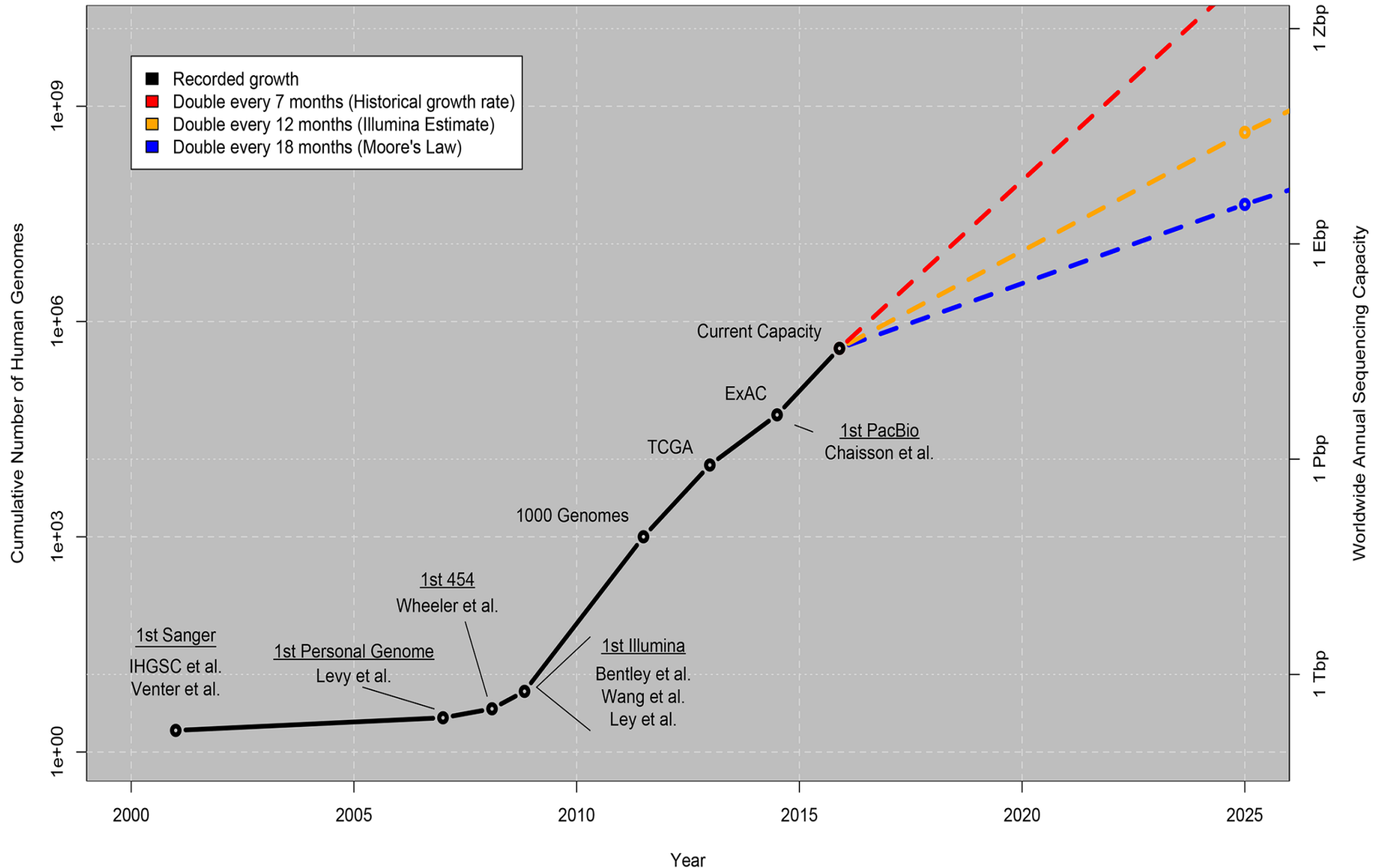


Moore's Law describes a long-term trend in the computer hardware industry that involves the doubling of '**compute power**' every two years.

Technology improvements that 'keep up' with Moore's Law are widely regarded to be doing exceedingly well, making it useful for comparison.

Growth of DNA sequencing




Growth of DNA Sequencing






STORAGE LIMITS

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today's storage technology.

	 Hard disk	 Flash memory	 Bacterial DNA
Read-write speed (μ s per bit)	> ~3,000–5,000	~100	<100
Data retention (years)	> >10	>10	>100
Power usage (watts per gigabyte)	> ~0.04	~0.01–0.04	<10 ⁻¹⁰
Data density (bits per cm ³)	> ~10 ¹³	~10 ¹⁶	~10 ¹⁹

WEIGHT OF DNA NEEDED TO STORE WORLD'S DATA


~1 kg

©nature

<http://www.nature.com/news/how-dna-could-store-all-the-world-s-data-1.20496>

EDITORIALS



Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.



**We are research
parasites!**

A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as “research parasites.”

Importance of groups like this

- We can discuss multidisciplinary topics (topics as diverse as members of this group)
- Help out each other
- Know about conferences and workshops, latest softwares and research articles
- Keep up with developments in the field
- Sharpen our communication and analytical skills
- Don't be the "lonely bioinformatician"

Objectives

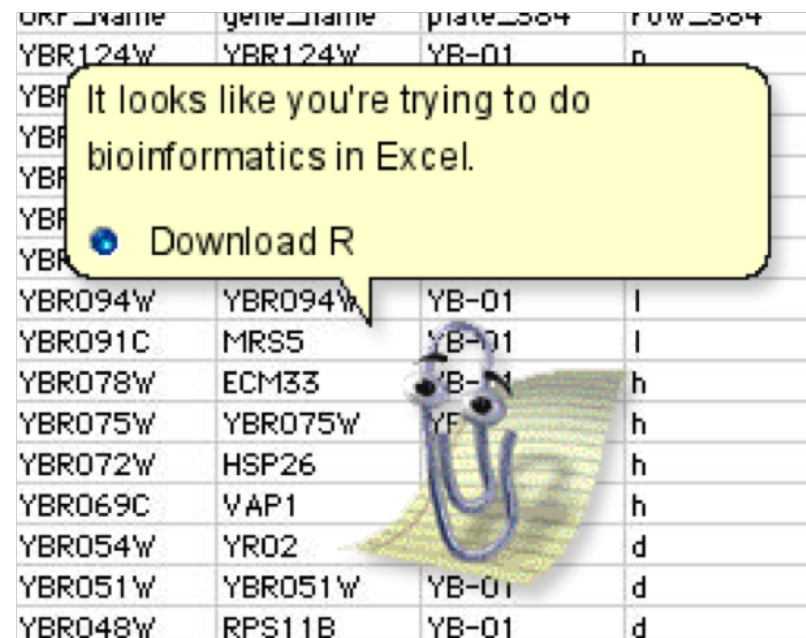
- Education
- Papers presentation (like a journal club)
 - Introduction to field
 - Introduction to topic
 - Methodology of paper (main focus)
- Presentation for a conference/ meeting
- Tutorials of packages or softwares (members are free to volunteer to lead on topics of their choosing)
- Discussion

Meeting

- Once in a week
- Doodle poll to decide time
- Place: to be decided
- Please step forward to present
- Invite your friends/ colleagues/ professors
(maybe also from downtown) to give a talk

How to do bioinformatics?

There is no **Microsoft Office** for **Bioinformatics**!



The image shows a screenshot of a Microsoft Excel spreadsheet. A yellow tooltip with a blue border is overlaid on the spreadsheet, containing the text: "It looks like you're trying to do bioinformatics in Excel." Below this text is a blue circular icon with a white 'd' and the text "Download R". A blue paperclip icon is also visible, attached to a yellow sticky note that is partially covering the spreadsheet cells. The spreadsheet has four columns: "ORF_name", "gene_name", "plate_name", and "row_name". The rows contain various gene identifiers and plate names.

ORF_name	gene_name	plate_name	row_name
YBR124W	YBR124W	YB-01	n
YBR124W	YBR124W	YB-01	n
YBR124W	YBR124W	YB-01	n
YBR124W	YBR124W	YB-01	n
YBR124W	YBR124W	YB-01	n
YBR124W	YBR124W	YB-01	n
YBR094W	YBR094W	YB-01	l
YBR091C	MRS5	YB-01	l
YBR078W	ECM33	YB-01	h
YBR075W	YBR075W	YB-01	h
YBR072W	HSP26	YB-01	h
YBR069C	VAP1	YB-01	h
YBR054W	YR02	YB-01	d
YBR051W	YBR051W	YB-01	d
YBR048W	RPS11B	YB-01	d

<https://www.biostars.org/p/16049/>

<http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages>

Help?

- **Google it!**
- **Social media:**
- **Biostars:**
- **Stackoverflow:**
- **Seqanswers:**
- **MonBUG:**



- <https://www.biostars.org/p/142494/>

**Any
suggestions?**



Thank you!

