

So you want to do Bioinformatics?

René Warren M.Sc. Biochemistry

Bioinformatics Coordinator @



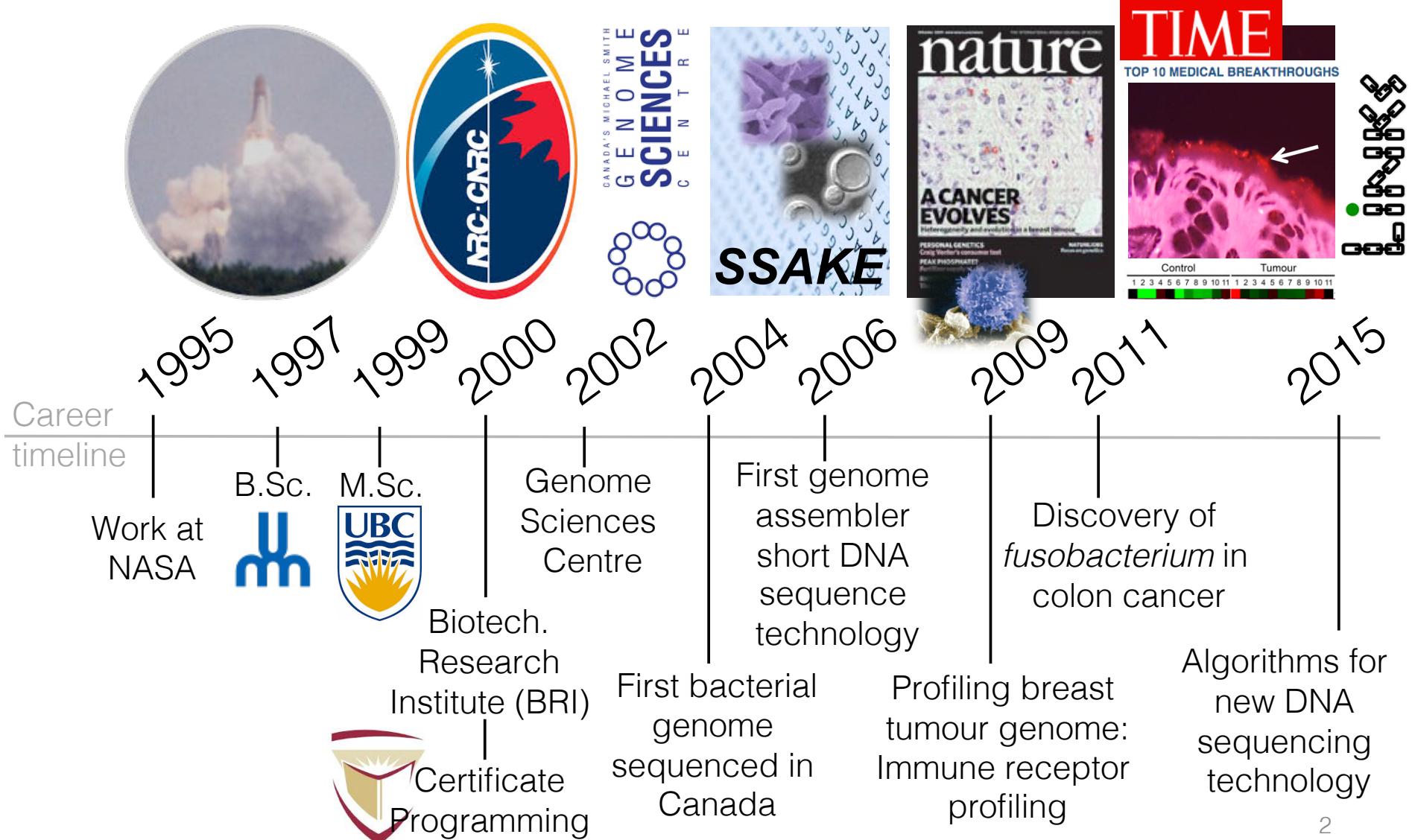
CHAD S. MICHAEL SMITH
**GENOME
SCIENCES
CENTRE**
UNIVERSITY OF TORONTO

SBN Expo, February 25th, 2015

Photo: R. Warren

About me

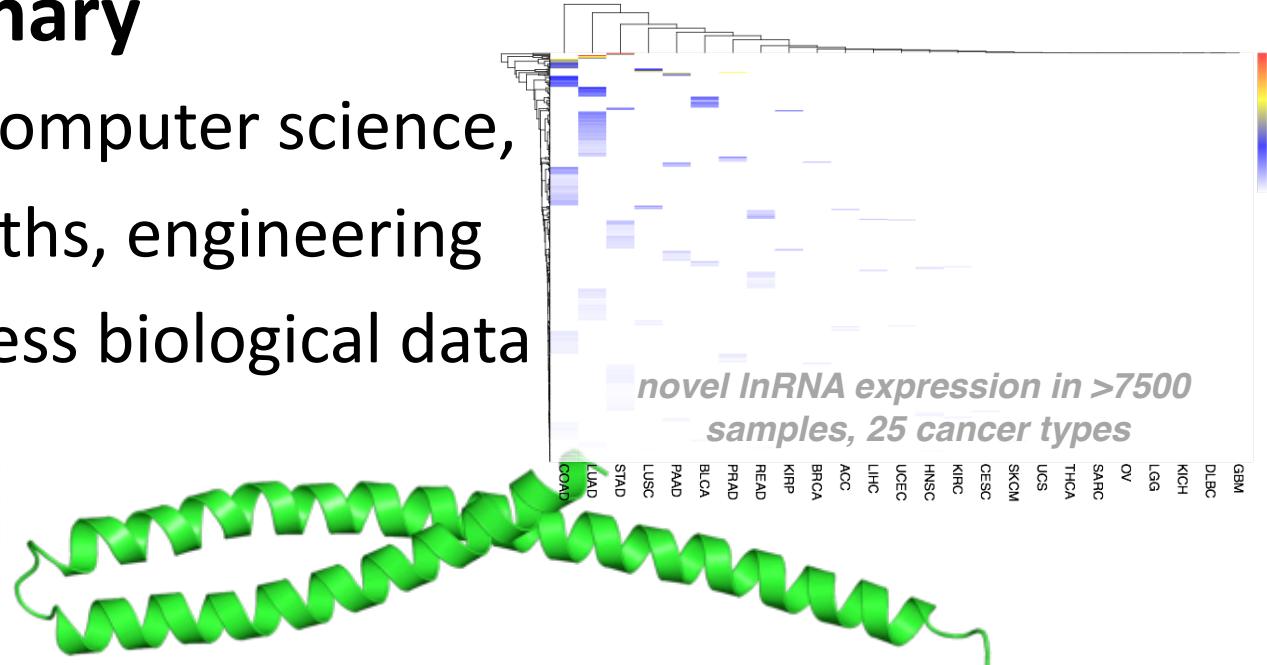
renewarren.ca



What is Bioinformatics ?

- **Interdisciplinary**

- Combines computer science,
statistics, maths, engineering
- Study/process biological data



- **Umbrella term**

- *Bioinformatics* (pipeline development, run software)
- *Computational biology* (algorithm development)

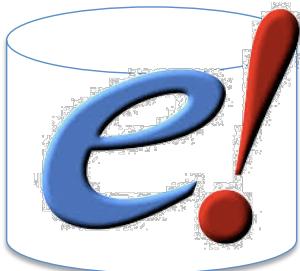
Bioinformatics Work

Support

Research

Build Infrastructure

- Databases
- Pipelines
- Systems



Run Software

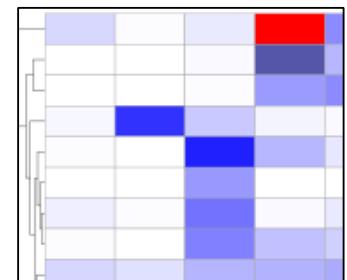
- 3rd party tools
- Queries



Technology Development

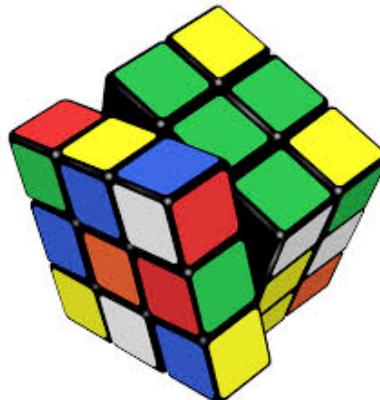
- Algorithm design
- Implementation

```
>>> myseq = Seq("AGTACACTGGT")
>>> myseq
Seq('AGTACACTGGT', Alphabet)
>>> print myseq
AGTACACTGGT
>>> myseq.transcribe()
Seq('AGUACACUGGU', RNAAlphabet)
>>> myseq.translate()
Translating sequence...
```



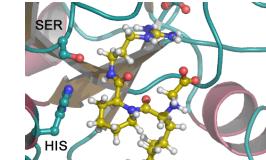
Why it's Cool ?

- Young
 - Largely untapped: opportunities, potential
 - Personal computers history only $\frac{1}{4}$ century old
 - Only 15 years since human genome sequenced
- Discoveries to be made
- Technologies to be built
- Impact on medicine: Diagnostics, Prognostics, Treatment
- Challenging



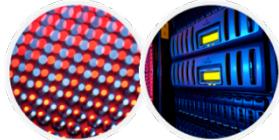
Does it have a Future ?

- Vast needs



- Genomics, Transcriptomics, Proteomics, Drug/Vaccine design
Clinical gene panel, Lab management systems

- Improved technology > decreased cost > analysis bottleneck



- Compute infrastructure can't keep up with [genomics] data
 - Petabyte storage and growing

- Not a one size-fits-all

- Need efficient algorithms, data structure, know-how

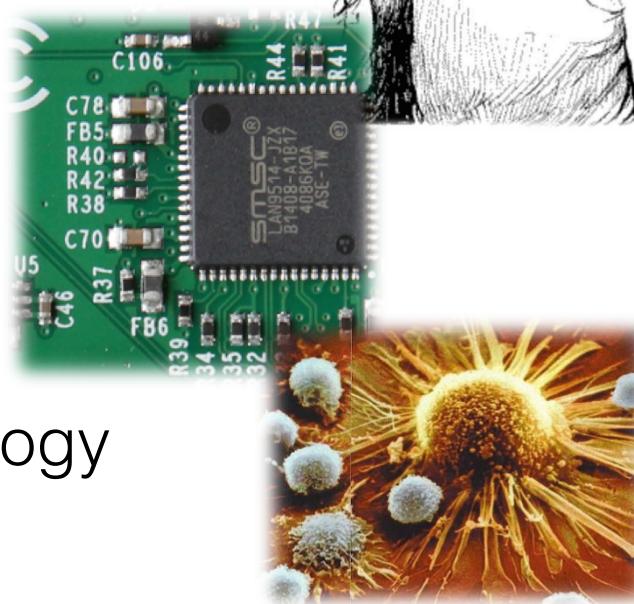
- Era of personalized [precision] medicine

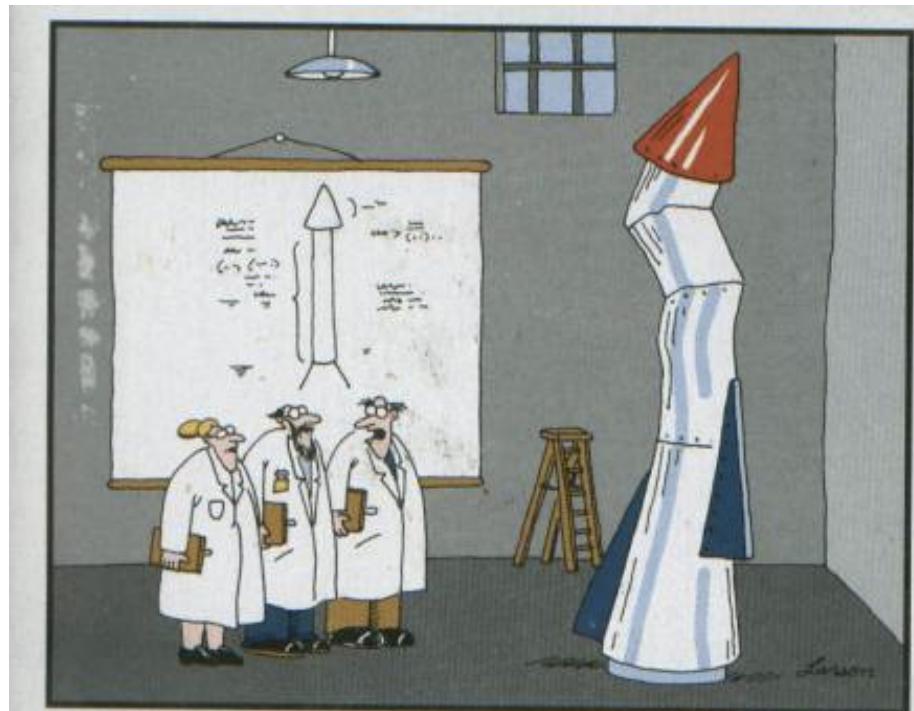
- Uptake in the clinic, not just research

Children's Hospital LA Commits \$50M to Expand Center for Personalized Medicine
Feb 12, 2015 | a GenomeWeb staff reporter
NEW YORK (GenomeWeb) – Children's Hospital Los Angeles said today that it will commit \$50 million in institutional funding to expand its Center for Personalized Medicine.

What is Needed ?

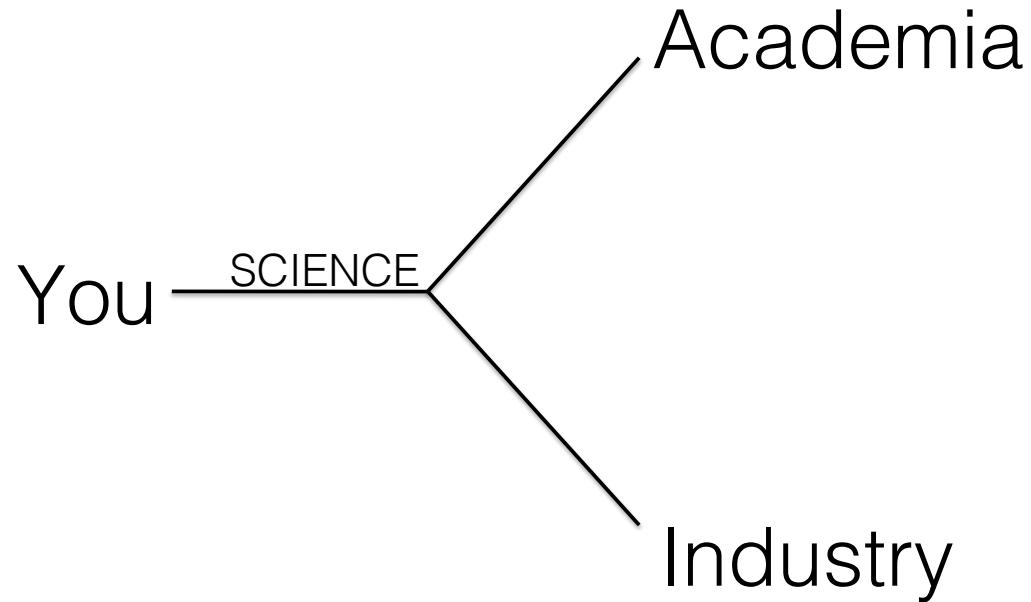
- Passion for Science
- Problem-Solving Skills
- Programming Base
 - Scripting
 - Unix OS
 - Web
 - Database
- Interest in Biology



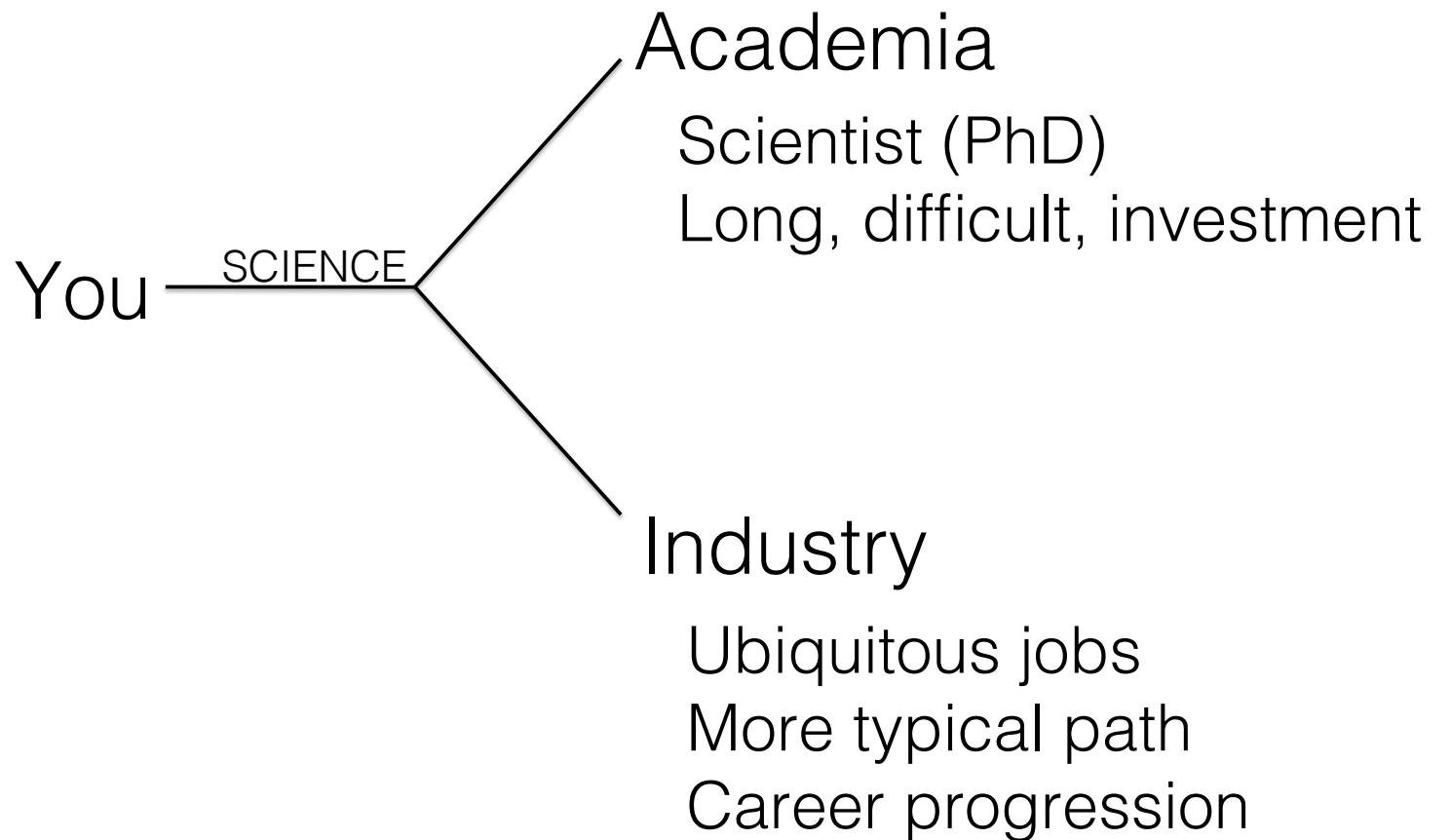


**"It's time we face reality, my friends. ...
We're not exactly rocket scientists."**

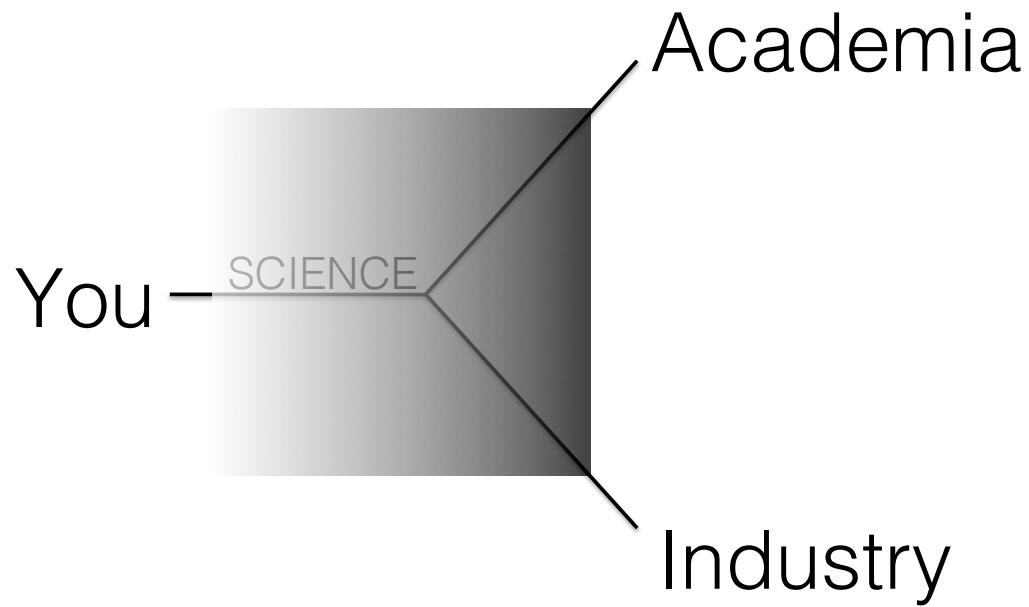
Career Path Paradigm



Career Path Paradigm



Off the Beaten Path



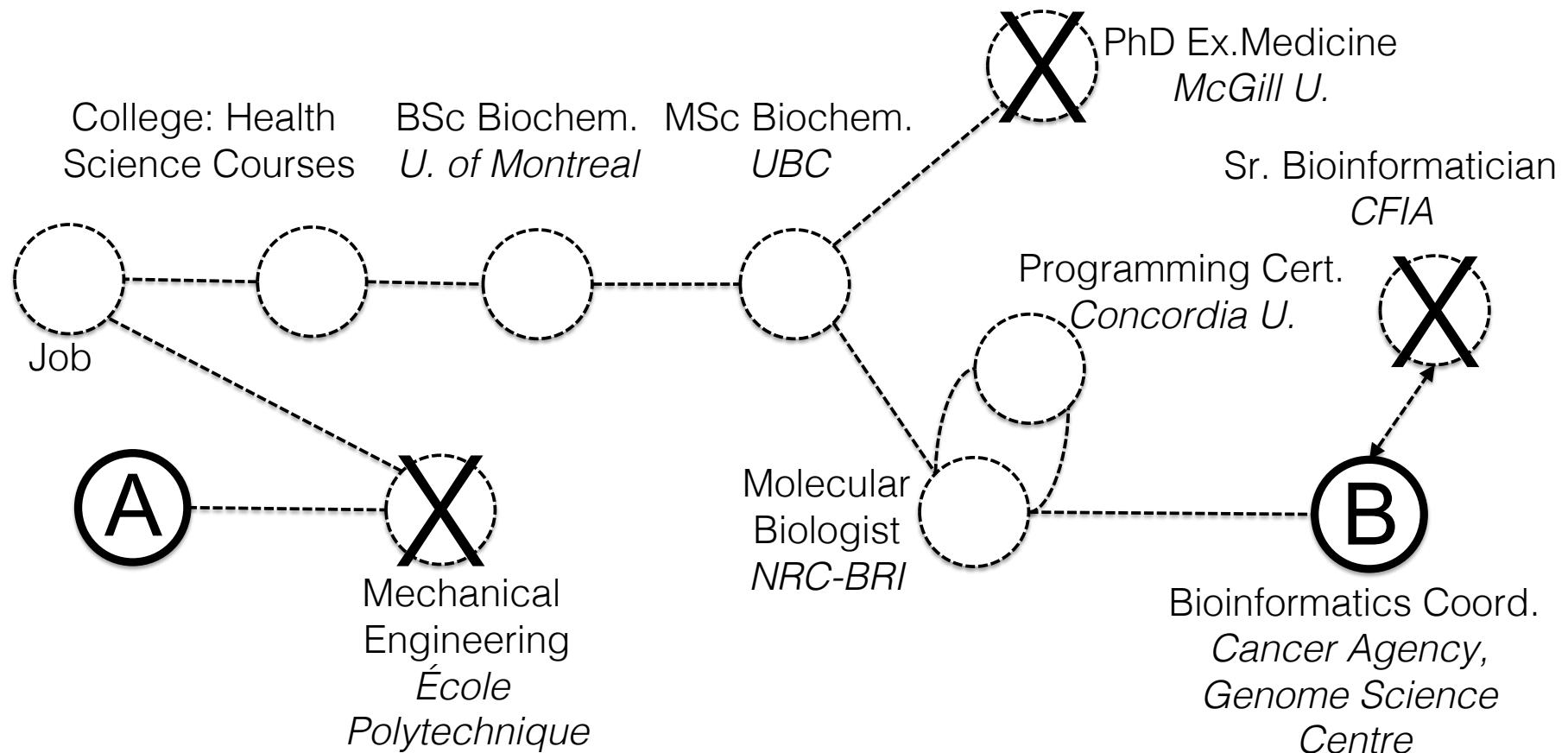
You ————— School ————— Career

- Path to career not unicorns & rainbows
- Experiences, interactions shape careers
 - Field not mature growing up

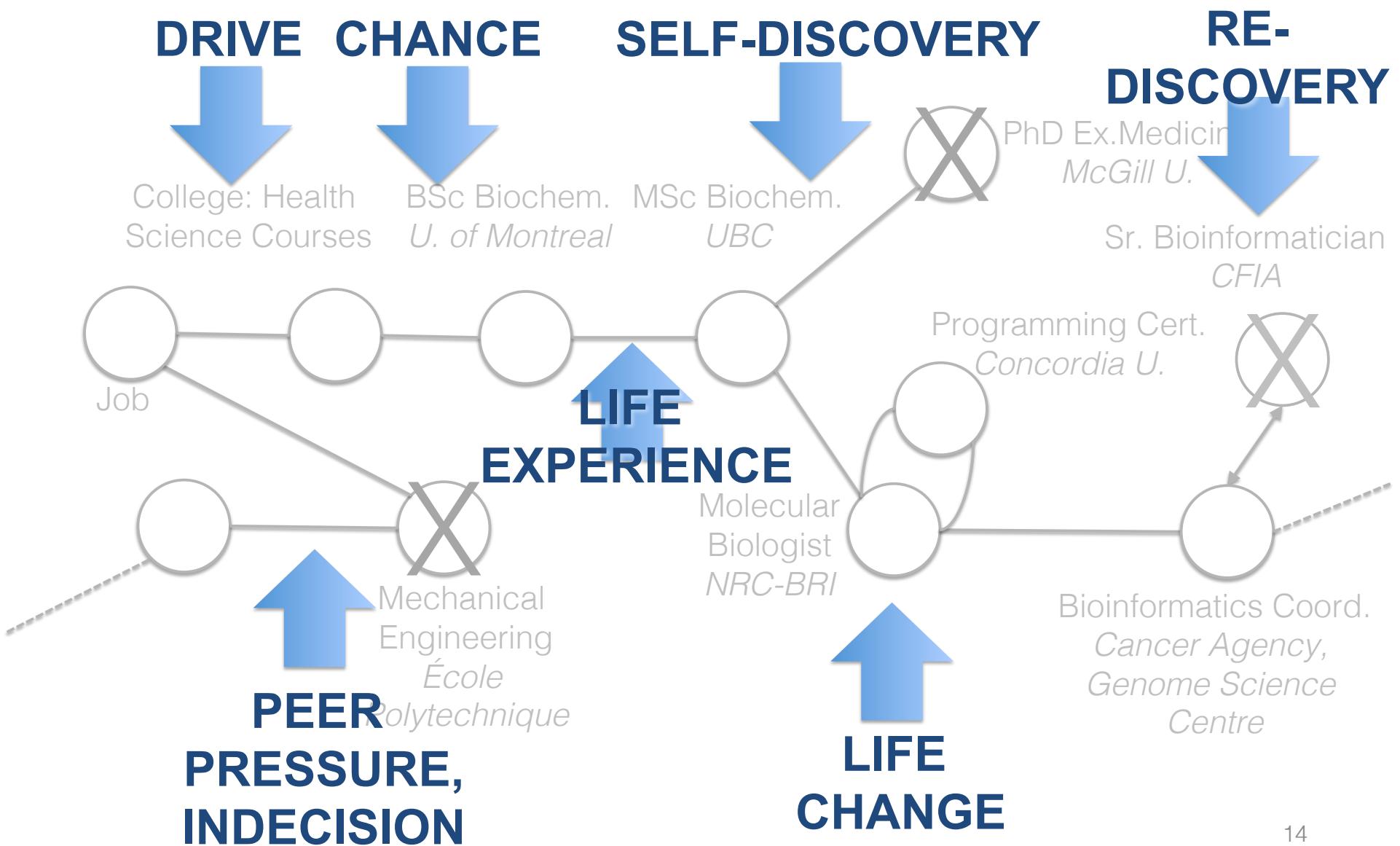


"I have not failed. I've just found 10,000 ways that won't work"

-Thomas A. Edison



It's the Journey, not Destination



Geologist? Vet? Architect? Engineer?



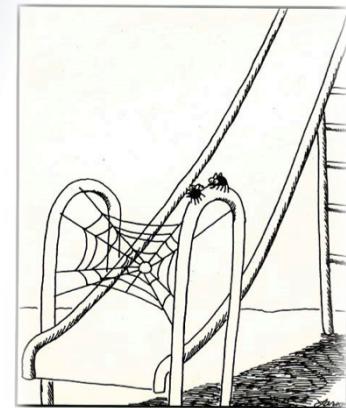
"I'm not lazy, I'm networking."

Network

Chance



Interests Aptitudes Opportunity



"If we pull this off, we'll eat like kings."

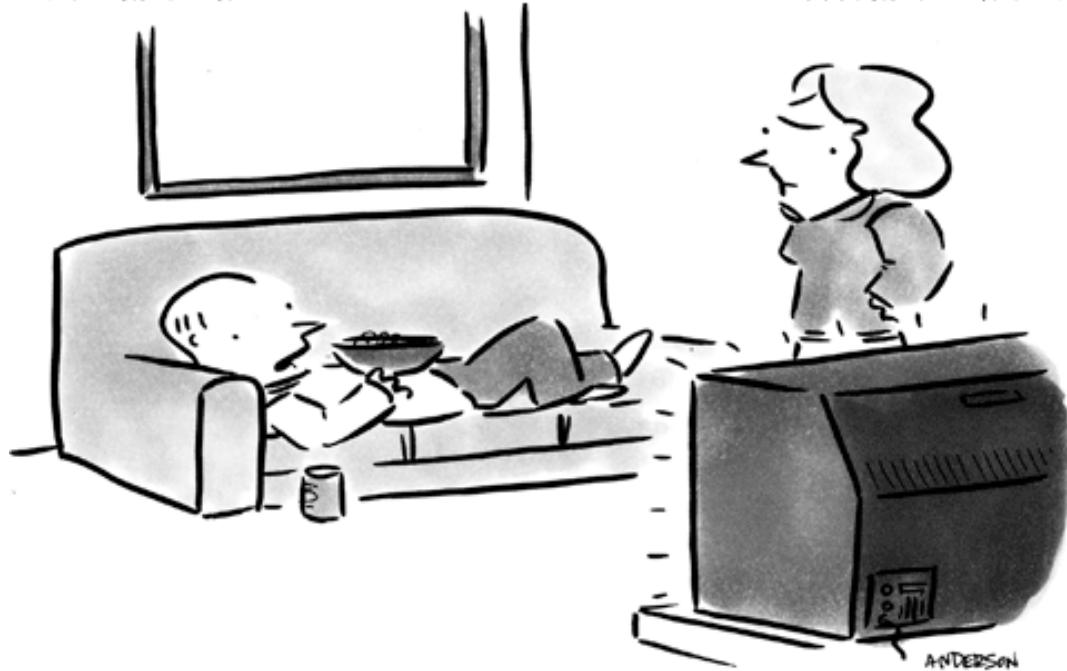


Chance



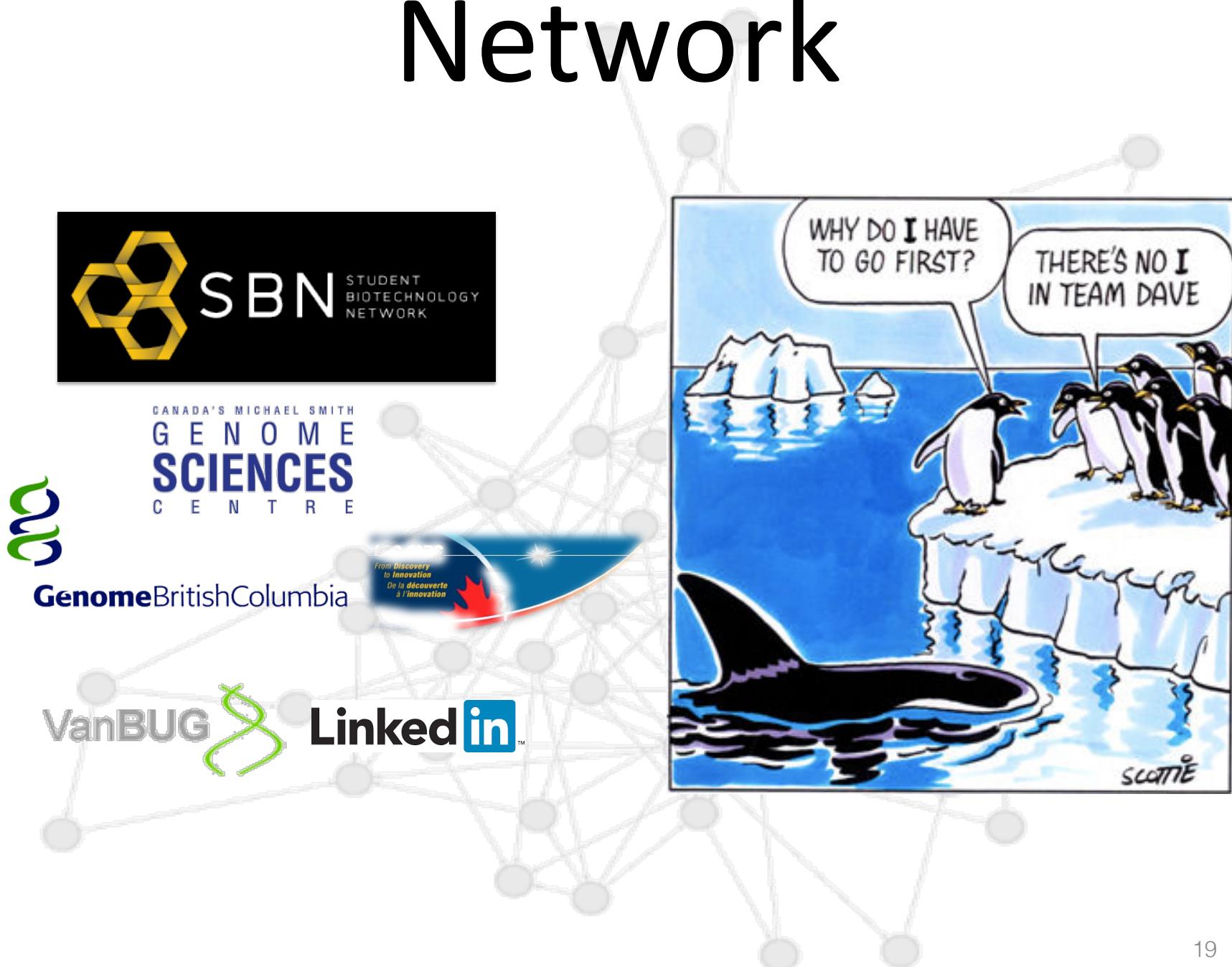
© MARK ANDERSON

WWW.ANDERZOONS.COM



"I'm not lazy, I'm networking."

Network





Interests

- Nature / Rocks
 - Collecting fossils
 - Polishing
- Building “Stuff”
- Breaking “Stuff”
 - Opening RC cars, radios;
putting back with Crazy Glue



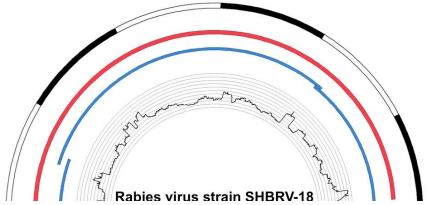
- Programming
 - “Basic” (Bill Gates) on *Superboard II*, and *Apple II* clone (1981)
- Fascination with space
- Academic standings
 - Weak high-school biology
 - Strong math/physics/comp. science



Superboard, OHIO Scientific, 1980



www.jolyon.co.uk



Aptitudes

2000
2003
2006
2009
2011
2013
2015

Cumate gene-switch engineered from *Pseudomonas*

Genome sequences of *Rhodococcus* & *Cryptococcus*

SAM: automate genome assembly & annotation

FASSI: improve large genome draft with maps

SSAKE: first genome assembler w\ short sequences

Synthetic biology – rebuilding *H.influenzae* in *E.coli*

Sequence profiling immune receptors

Cancer epitope survey COSMIC/TCGA

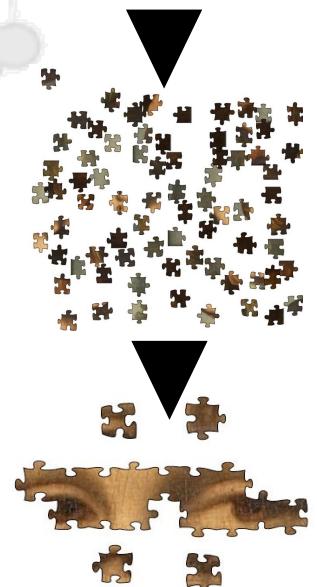
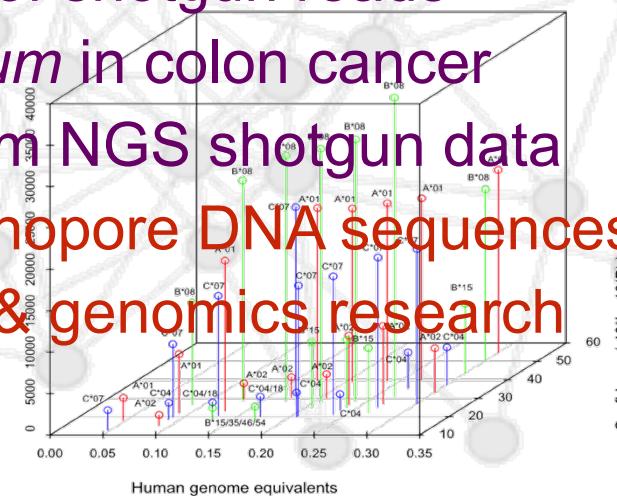
TASR: targeted assembly of shotgun reads

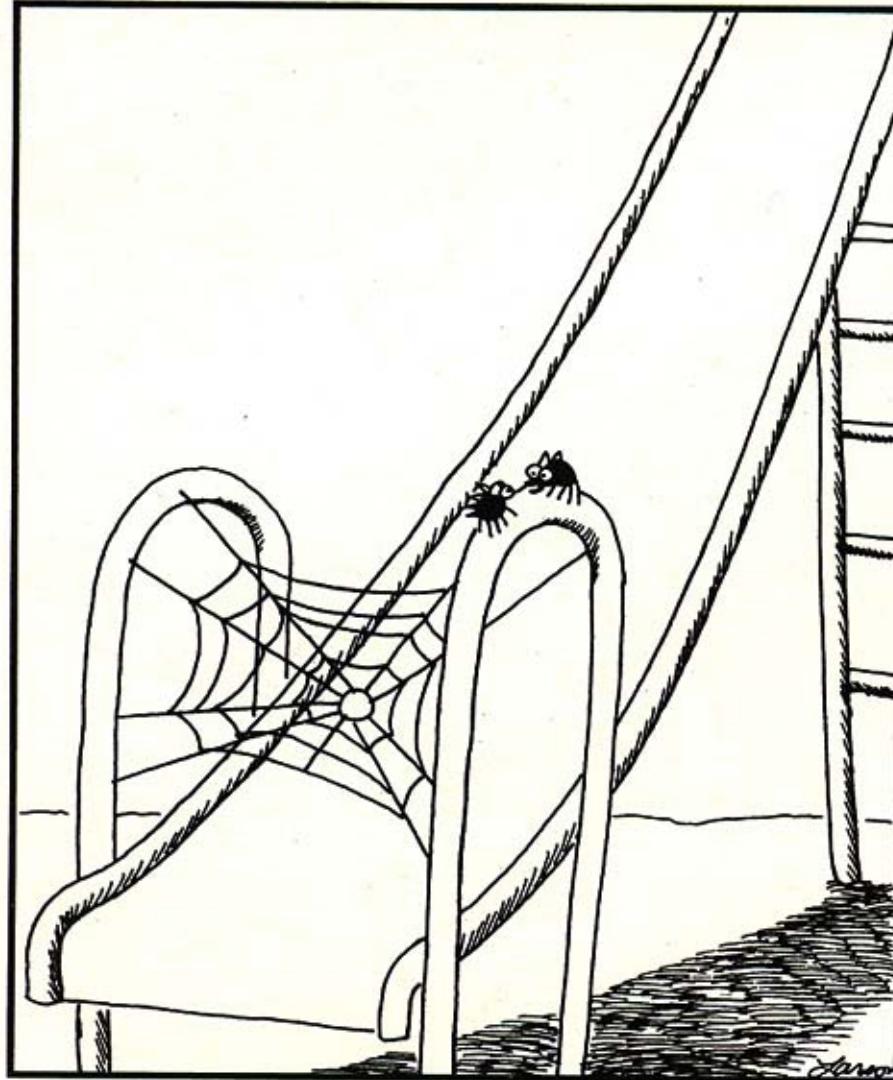
Discovery of *Fusobacterium* in colon cancer

HLAminer: HLA typing from NGS shotgun data

LINKS : Scaffolding w\ Nanopore DNA sequences

Coordinate bioinformatics & genomics research





"If we pull this off, we'll eat like kings."



Opportunity

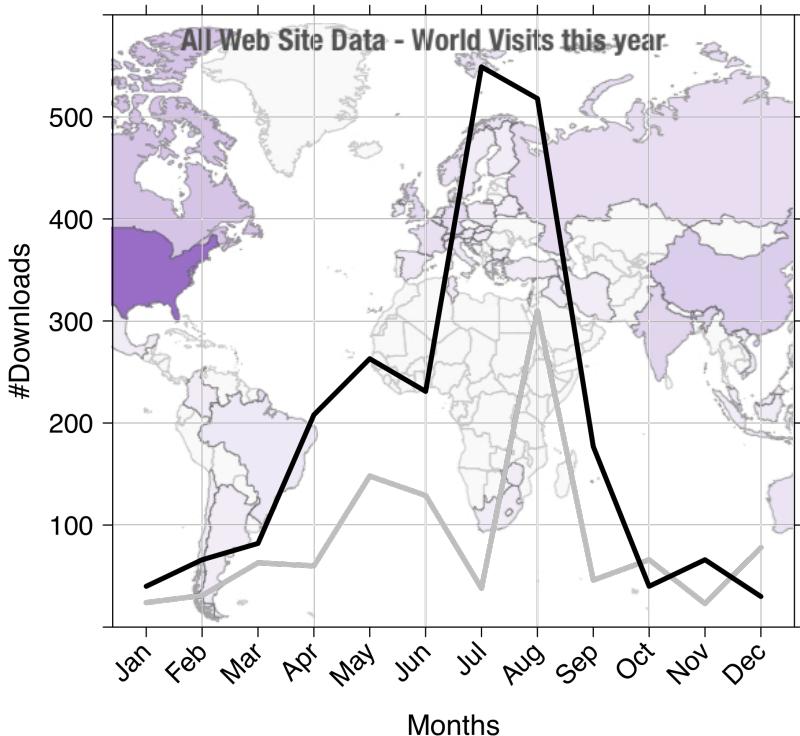
Technology > Challenges > Innovation

Bioinformatics

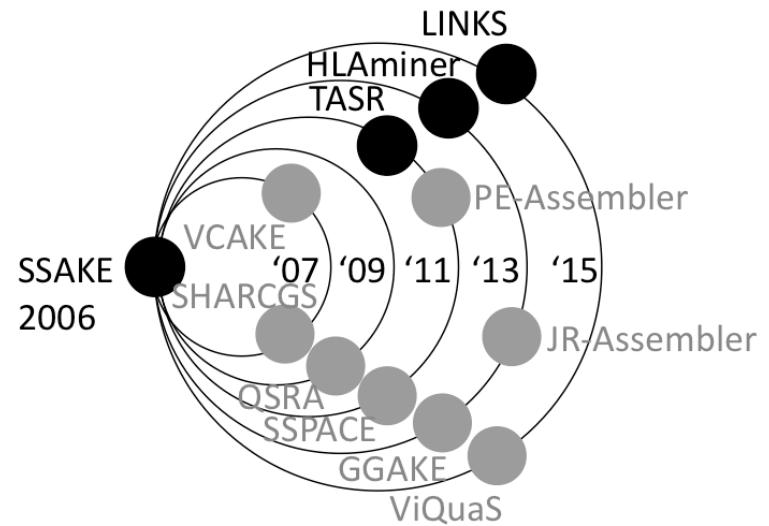
Assembling millions of short DNA sequences using SSAKE

René L. Warren*, Granger G. Sutton¹, Steven J. M. Jones and
Robert A. Holt

2013 2014



- Genome assembly possible with tiny 25 nt DNA sequences
- Algorithm: core of many tools



- Engine of clinical HLA diagnostics at:

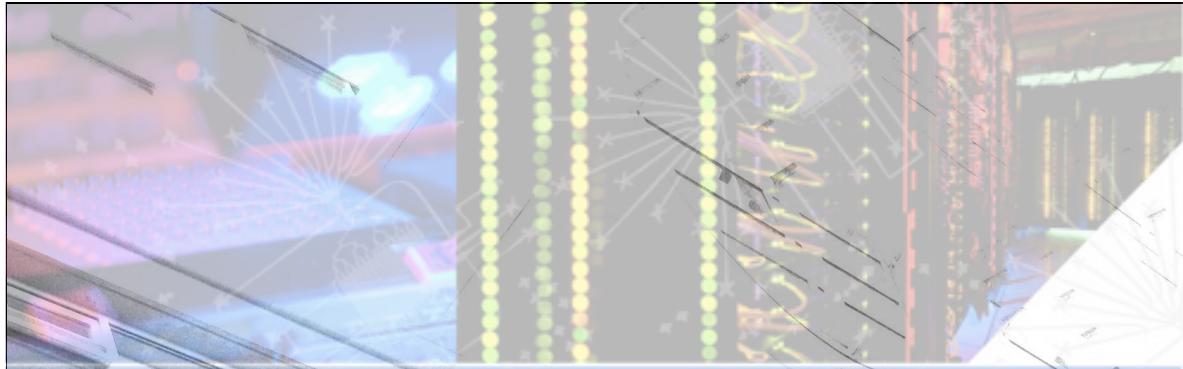


In Closing

- Few A → B perfect paths
 - “The greatest glory in living lies not in never falling, but in rising every time we fall.” — Nelson Mandela
- Be true to yourself & *have fun !*
- Challenges > Opportunities > Notoriety > Self-worth > Fullfillment : *say “YES” to the Now!*
- **Like solving problems in biology, using engineering concepts & interested in ‘puters:**



Bioinformatics may be for you!



CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

A leading international centre for genomics and bioinformatics research committed to advancing knowledge of cancer-related diseases, improving human health through disease prevention, diagnosis and therapeutic approaches, and realizing social and economic benefits of genomics research.

54 TERABASES SEQUENCED • A HUMAN GENOME EVERY 17 MINUTES • HIGH PERFORMANCE COMPUTING



AFFILIATIONS BC Cancer Agency • BC Cancer Foundation • Genome BC • Genome Canada • Simon Fraser University • University of British Columbia • Génome Sciences Institute

Thank you! Merci!

**Robert Warren & Lise Pelletier
Bob Molday (UBC)
Alaka Mullick (NRC)
GSC folks**



GenomeBritishColumbia



CIHR IRSC
Canadian Institutes of
Health Research
Instituts de recherche
en santé du Canada

FRSQ
Québec
Fonds de la
recherche en santé
du Québec



THE
UNIVERSITY OF
BRITISH
COLUMBIA



GenomeCanada

BCCA CANCER RESEARCH CENTRE

Photo: R. Warren

