Understanding your genome: the path to personalized medicine

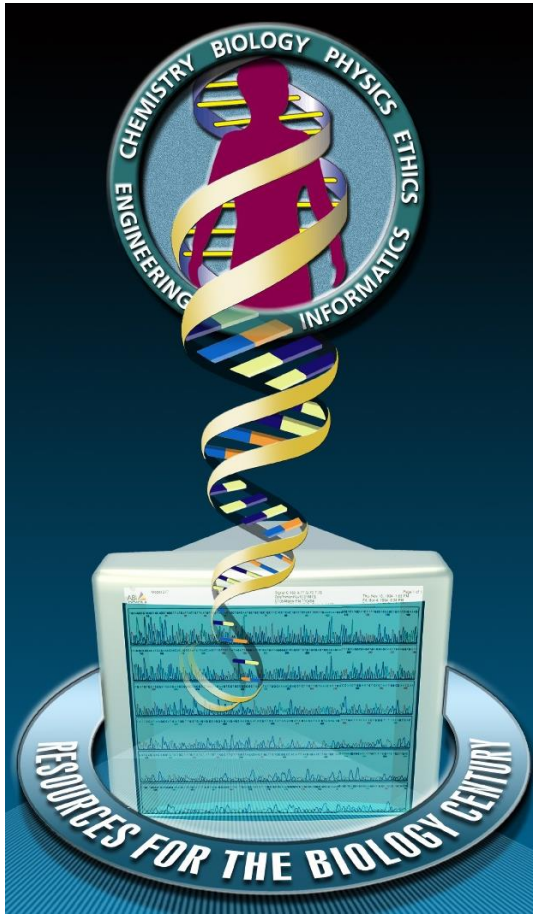
Robin Dowell

BioFrontiers Institute

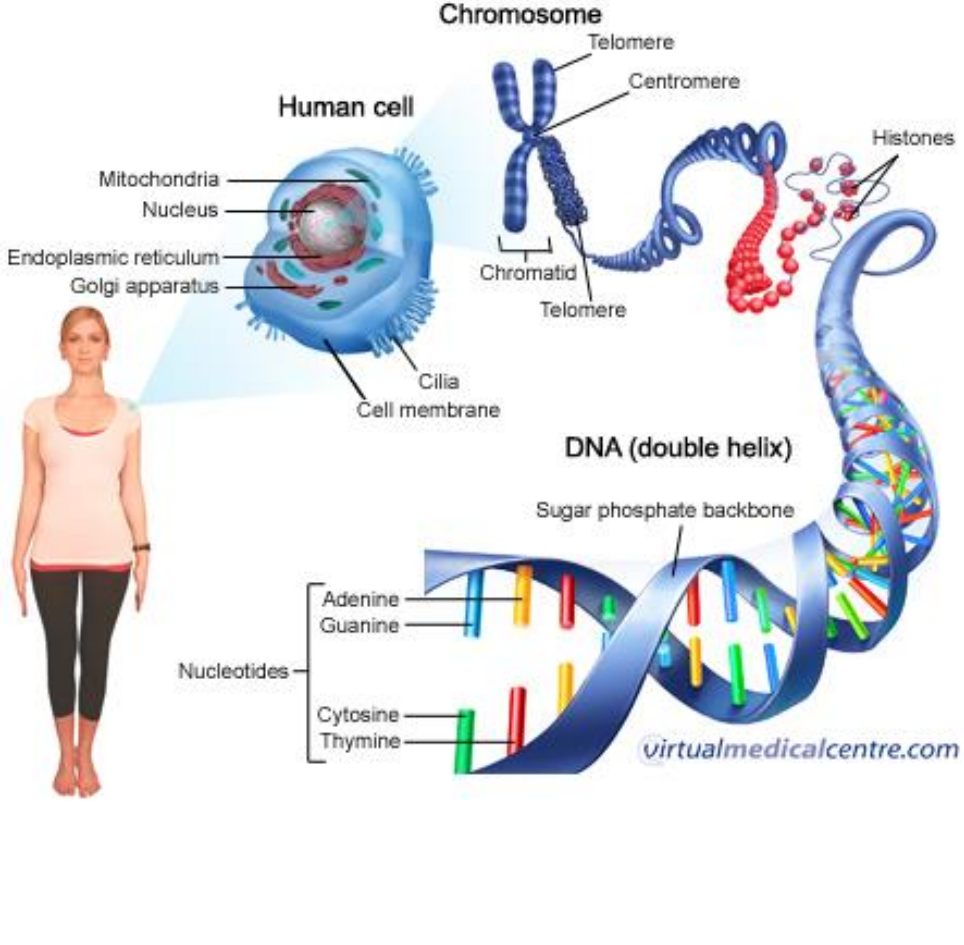
Molecular, Cellular and Developmental Biology

Computer Science

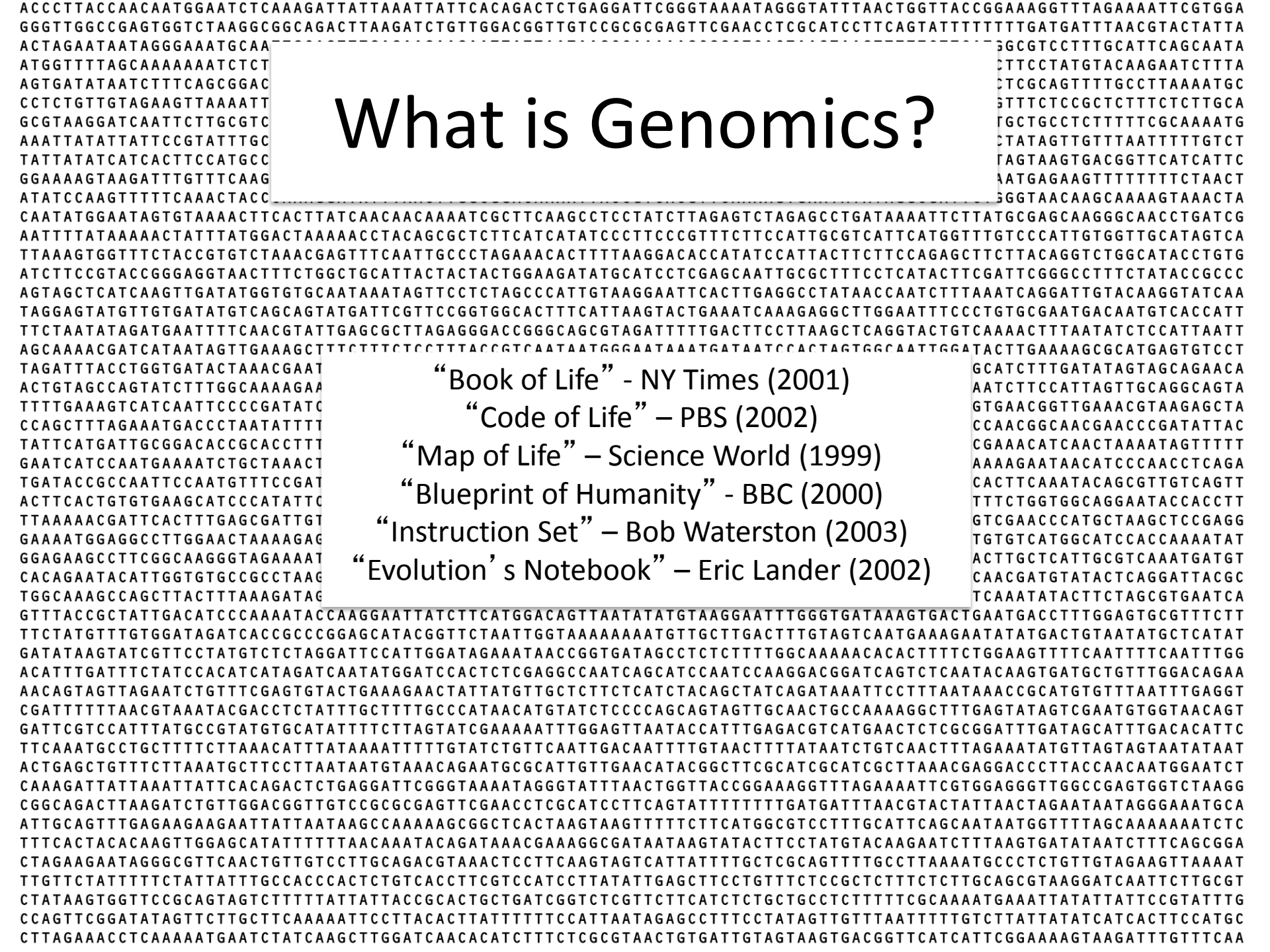
robin.dowell@colorado.edu



http://genomics.energy.gov/

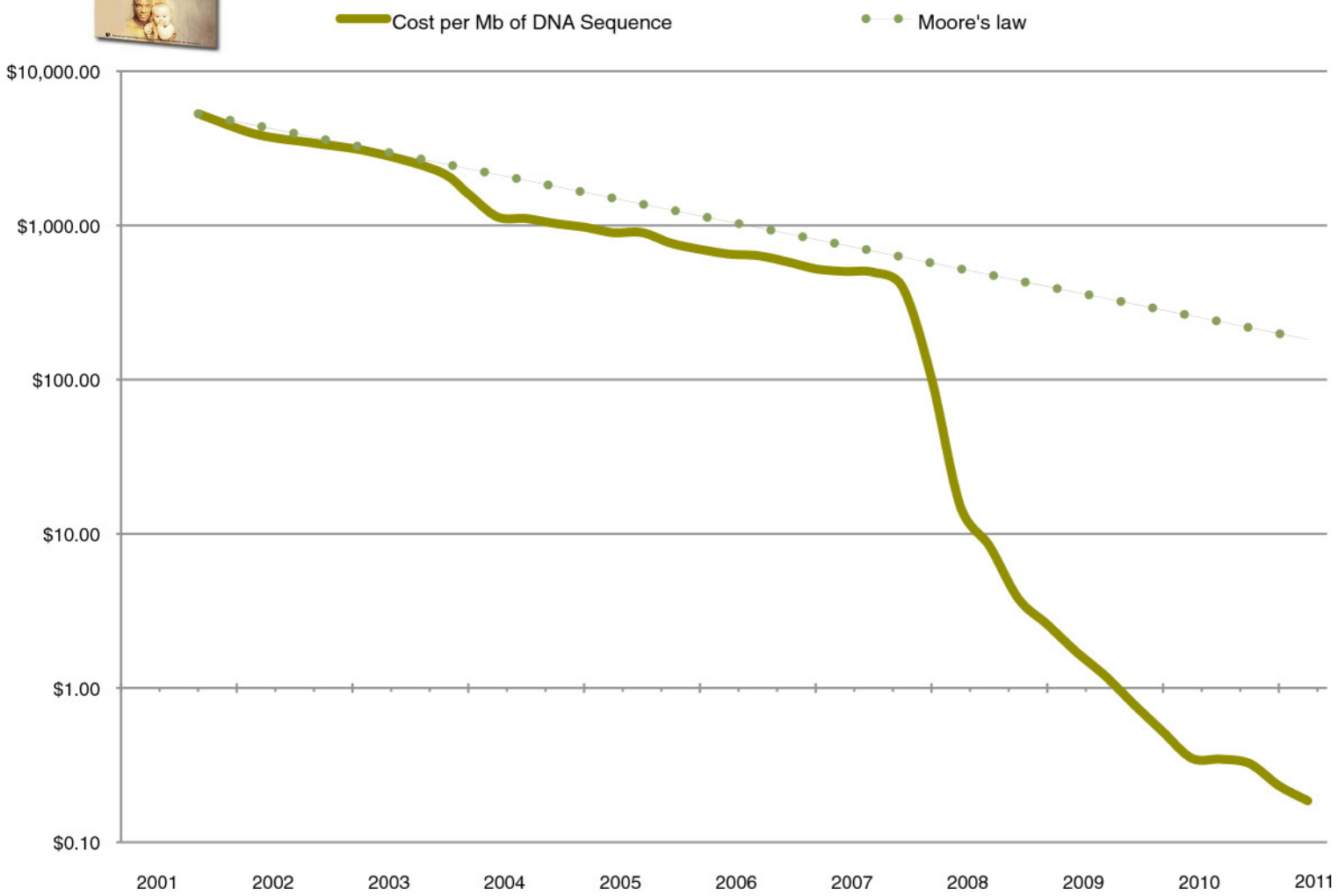


```
ACCCTTACCAACAATGGAATCTCAAAGATTATTAAATTATTCACAGACTCTGAGGATTCGGGTAAAATAGGGTATTTAACTG
GGGTTGGCCGAGTGGTCTAAGGCGGCAGACTTAAGATCTGTTGGACGGTTGTCCGCGCGAGTTCGAACCTCGCATCCTTCAG
ACTAGAATAATAGGGAAATGCAATTGCAGTTTGAGAAGAAGAATTATTAATAAGCCAAAAAGCGGCTCACTAAGTAAGTTTT
ATGGTTTTAGCAAAAAAATCTCTTTCACTACACAAGTTGGAGCATATTTTTTAACAAATACAGATAAACGAAAGGCGATAAT
AGTGATATAATCTTTCAGCGGACTAGAAGAATAGGGCGTTCAACTGTTGTCCTTGCAGACGTAAACTCCTTCAAGTAGTCAT
CCTCTGTTGTAGAAGTTAAAATTTGTTCTATTTTTCTATTATTTGCCACCCACTCTGTCACCTTCGTCCATCCTTATATTGA
GCGTAAGGATCAATTCTTGCGTCTATAAGTGGTTCCGCAGTAGTCTTTTTATTATTACCGCACTGCTGATCGGTCTCGTTCT
AAATTATATTATTCCGTATTTGCCAGTTCGGATATAGTTCTTGCTTCAAAAATTCCTTACACTTATTTTTTCCATTAATAGA
TATTATATCATCACTTCCATGCCTTAGAAACCTCAAAAATGAATCTATCAAGCTTGGATCAACACATCTTTCTCGCGTAACT
GGAAAAGTAAGATTTGTTTCAAGTAAGCTTTGGGGTGCCGTAAACGTACGCCTATTATCAACCATACCGTCTTTTTGTCTAC
ATATCCAAGTTTTTCAAACTACCCAAAGGATATTTAACTTGGCGGACAAAATTACCGTCACGTTCAAAAGTGAATATATAGG
CAATATGGAATAGTGTAAAACTTCACTTATCAACAACAAAATCGCTTCAAGCCTCCTATCTTAGAGTCTAGAGCCTGATAAA
AATTTTATAAAAACTATTTATGGACTAAAAACCTACAGCGCTCTTCATCATATCCCTTCCCGTTTCTTCCATTGCGTCATTC
TTAAAGTGGTTTCTACCGTGTCTAAACGAGTTTCAATTGCCCTAGAAACACTTTTAAGGACACCATATCCATTACTTCTTCC
ATCTTCCGTACCGGGAGGTAACTTTCTGGCTGCATTACTACTACTGGAAGATATGCATCCTCGAGCAATTGCGCTTTCCTCA
AGTAGCTCATCAAGTTGATATGGTGTGCAATAAATAGTTCCTCTAGCCCATTGTAAGGAATTCACTTGAGGCCTATAACCAA
TAGGAGTATGTTGTGATATGTCAGCAGTATGATTCGTTCCGGTGGCACTTTCATTAAGTACTGAAATCAAAGAGGCTTGGAA
TTCTAATATAGATGAATTTTCAACGTATTGAGCGCTTAGAGGGACCGGGCAGCGTAGATTTTTGACTTCCTTAAGCTCAGGT
AGCAAAACGATCATAATAGTTGAAAGCTTTCTTTCTCCTTTACCGTCAATAATGGGAATAAATGATAATCCACTAGTGGCAA
TAGATTTACCTGGTGATACTAAACGAATGTCATTTGCGCCACTGATAGTAACAAATCCAGGTATTGCGATACCCTTGCTTAA
ACTGTAGCCAGTATCTTTGGCAAAAGAATTTATTCCTAGAATTTTTCCTTGATTGTTGGTCTTTGTTGCGGGTAGAATTTAA
TTTTGAAAGTCATCAATTCCCCGATATCCGTTCCACATAGCATCAAAATACTAGAGAACCATCATCTCCATATTCCATGACG
CCAGCTTTAGAAATGACCCTAATATTTTCATTGAAGATTATAGCCGGGCCTCTTCTGTCCAATATTATTAAGGTTCCTTCAA
TATTCATGATTGCGGACACCGCACCTTTTTTAGCATGTATTACCGTACTTGGTATAAAGCCTTGTTTAACGTTCGTAGGCCC
GAATCATCCAATGAAAATCTGCTAAACTTCAATTGCAGTGCATCGCTTTTGGGAAGTTGTCCGTAAAACTTGTTTGTTTCGA
TGATACCGCCAATTCCAATGTTTCCGATGCAAAGGAGATATTTTTTATGGCCAGGTTAGTAGCTCTGTTAAGTACTTTAGCA
ACTTCACTGTGTGAAGCATCCCATATTCTGACTGAACCATTACTATGCCCAGTTAGTAAAGCAGATCTGGTCTCATGCCTTC
TTAAAAACGATTCACTTTGAGCGATTGTCATCATCCCCAGCCAAAGATTTTTTTGAACTGATTGCGCTATACAAGTAGTAAC
GAAAATGGAGGCCTTGGAACTAAAAGAGCCTGCTGGGTATATCAATGTTTCAAGTTCGCCATCTTCTAAAAGTAATAAAATG
GGAGAAGCCTTCGGCAAGGGTAGAAAATTTATTAGTGGAGCTTTGCCAATCACGAAAATAGTTTTTGTTGAACCGGTTTAGC
CACAGAATACATTGGTGTGCCGCCTAAGTCAATCATGGTTATTTCCTGAGGTAAACAGGGATTTTCTGTAGCTTTTGTTGCA
TGGCAAAGCCAGCTTACTTTAAAGATAGCGGGTGTTTCTGTAAAGGAACAGTCTTTCAAGGCAGGATTTGGAAAATTCACAT
GTTTACCGCTATTGACATCCCAAAATACCAAGGAATTATCTTCATGGACAGTTAATATATGTAAGGAATTTGGGTGATAAAG
```

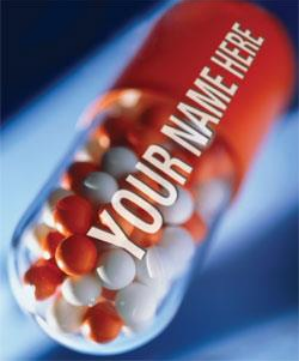

What is Genomics?
“Book of Life” - NY Times (2001)
“Code of Life” – PBS (2002)
“Map of Life” – Science World (1999)
“Blueprint of Humanity” - BBC (2000)
“Instruction Set” – Bob Waterston (2003)
“Evolution’s Notebook” – Eric Lander (2002)


grimmy.com
GENOME PROJECT
WHEN I ASKED WHAT LITTLE GIRLS ARE MADE OF, I WAS HOPING HE WOULD SAY "SUGAR AND SPICE."

Cost of 1Mb of DNA sequencing



Sboner et al. Genome Biology 2011 12:125



What is Personalized Medicine?

"molecules measured in a patient's lab tests can inform decisions about preventing or treating diseases"

Genomics Landscape

Sequencing → 2º Analysis → 3º Analysis → Integration → Applications

Raw Data → Complete Genomes → Genomic Interpretation → Medical Interpretation → Diagnosis, Treatment, Drug Discovery, Screening, Prevention

http://venturebeat.com

Central Dogma

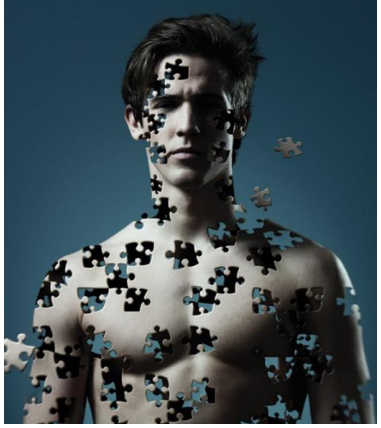
http://www.rsc.org

Proteins in human disease

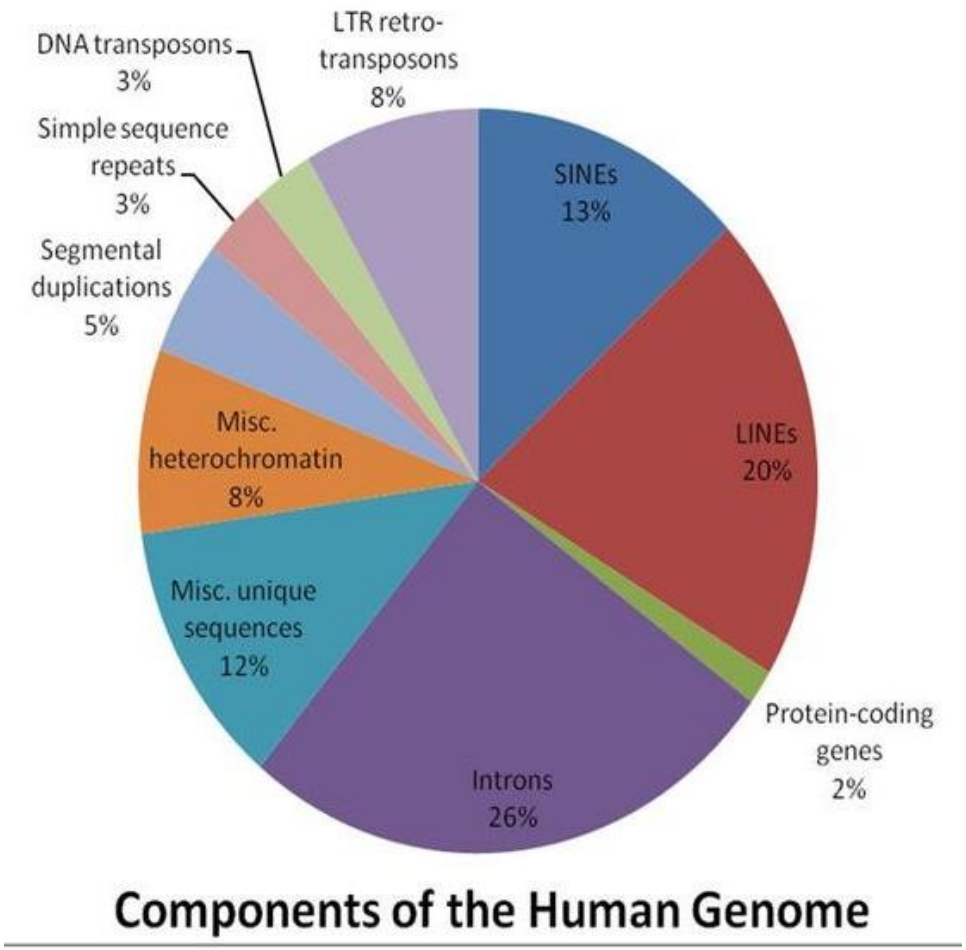
TABLE **22.2**

Examples of Human Disorders Inherited in an Autosomal Dominant Manner

Disorder	Chromosomal Location of Gene	Gene Product	Effects of Disease-Causing Allele
Aniridia	11p	Pax6 transcription factor	An absence of the iris of the eye, leading to visual impairment and sometimes blindness
Achondroplasia	4p	Fibroblast growth factor receptor-3	A common form of dwarfism associated with a defect in the growth of long bones
Marfan syndrome	15q	Fibrillin-1	Tall and thin individuals with abnormalities in the skeletal, ocular, and cardiovascular systems due to a weakening in the elasticity of certain body parts
Osteoporosis	7q	Collagen (type $1_{\alpha 2}$)	Brittle, weakened bones
Familial hypercholesterolemia	19p	LDL receptor	Very high serum levels of low-density lipoprotein (LDL), a predisposing factor in heart disease
Huntington disease	4p	Huntingtin	Neurodegeneration that occurs relatively late in life, usually in middle age
Neurofibromatosis I	17q	Neurofibromin	Individuals may exhibit spots of abnormal pigmentation (café-au-lait spots) and growth of noncancerous tumors in the nervous system



What do we now know about a typical human genome?



Components of the Human Genome

- Nearly ALL genes have alternative isoforms
- 65-80% of the genome is transcribed
- ENCODE claims 80% of genome has some function

<u>COMPARE TWO INDIVIDUALS:</u>

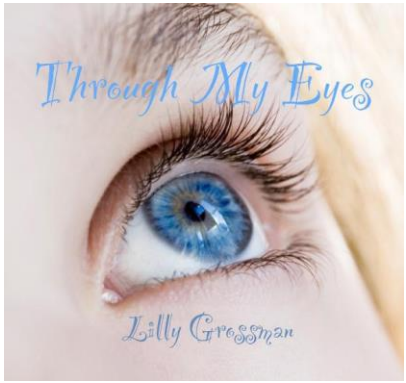
- 99.9% identical
- ~3 million variations (single nucleotide variations, insertions, deletions, copy number variations)

http://www.technologyreview.com/featuredstory/422140/the-human-genome-a-decade-later/

A cool sequencing success story!

Highly recommended reading:

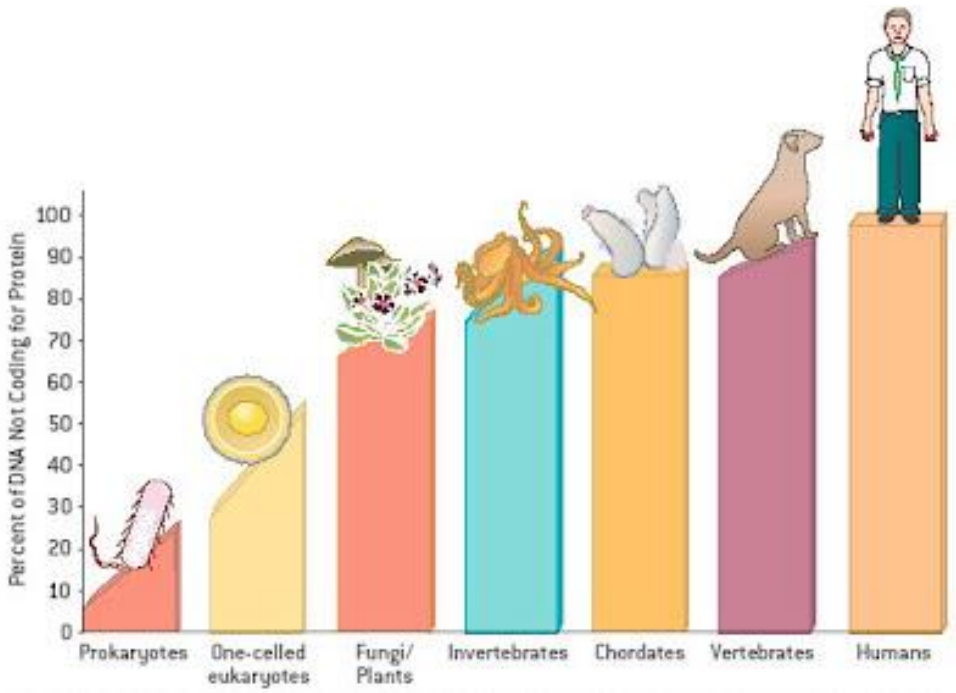
“We Gained Hope.” The Story of Lilly Grossman’s Genome

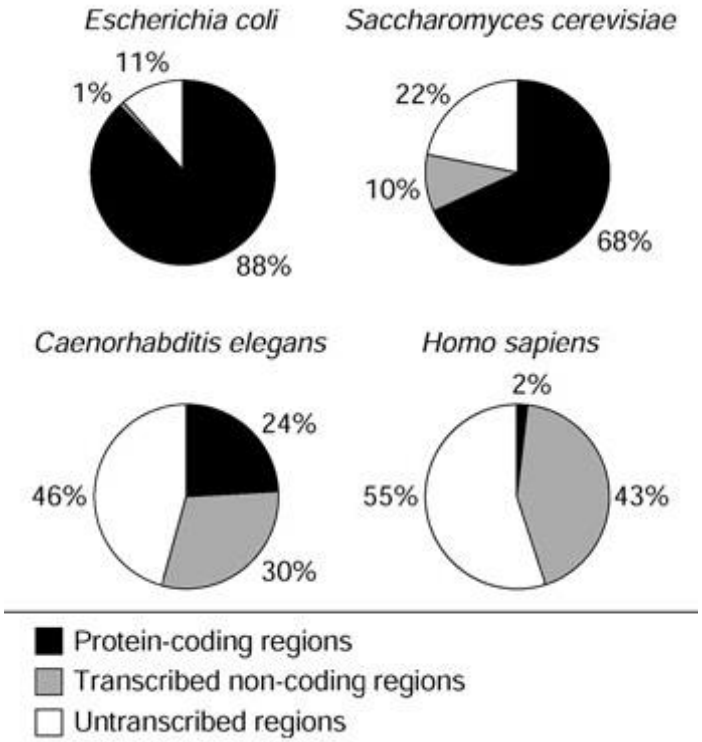


National Geographic’s Blog
“PHENOMENA: Not exactly rocket science”
Posted: Monday March 11, 2013
http://phenomena.nationalgeographic.com/2013/03/11/we-gained-hope-the-story-of-lilly-grossmans-genome/


DID YOU KNOW THAT ONLY 10,000 GENES SEPARATE HUMANS FROM WORMS?
THAT MANY?
GENOME PROJECT
FARSIDE

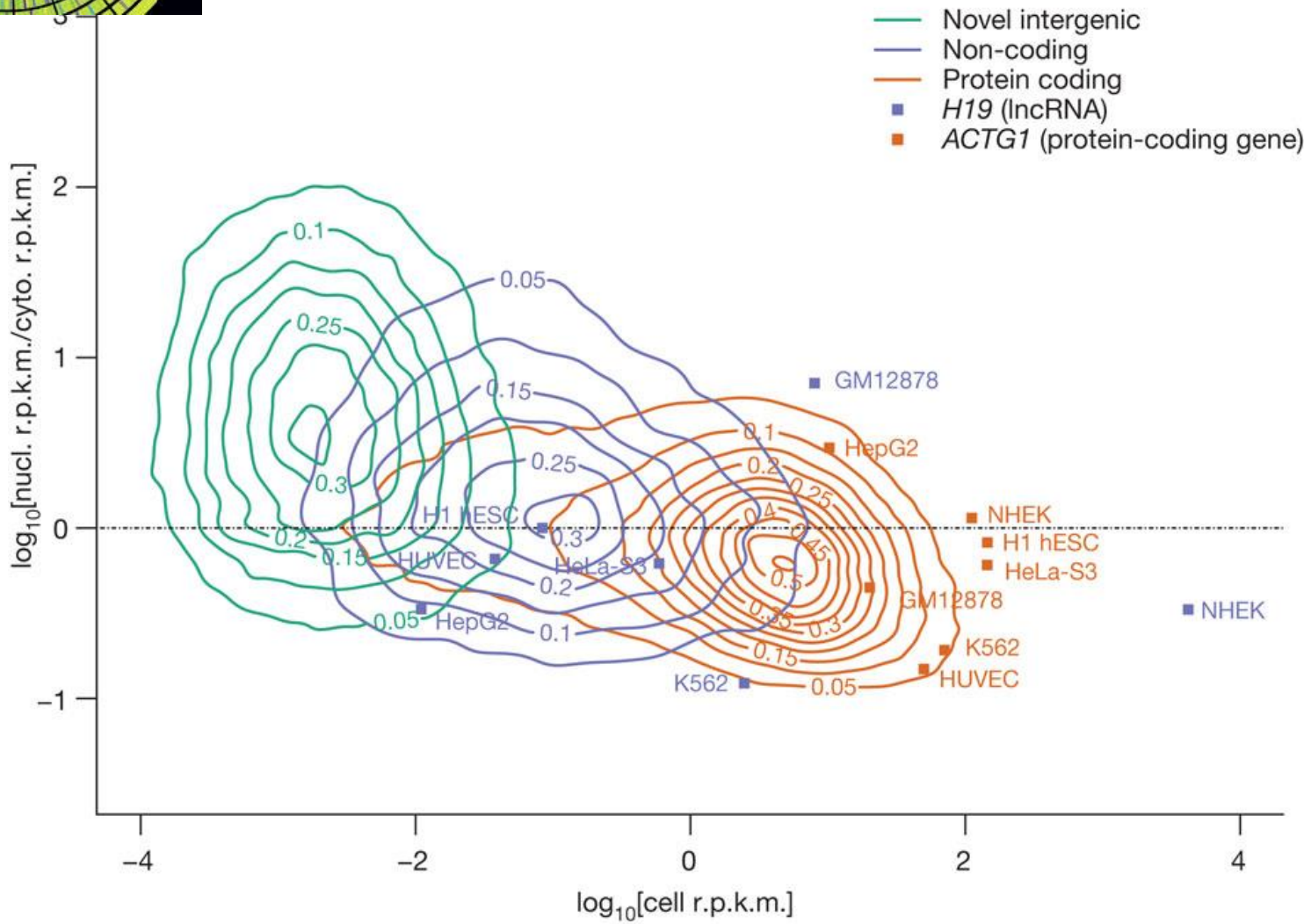
What makes us more than a worm?





John Mattick

What makes us more than a worm?



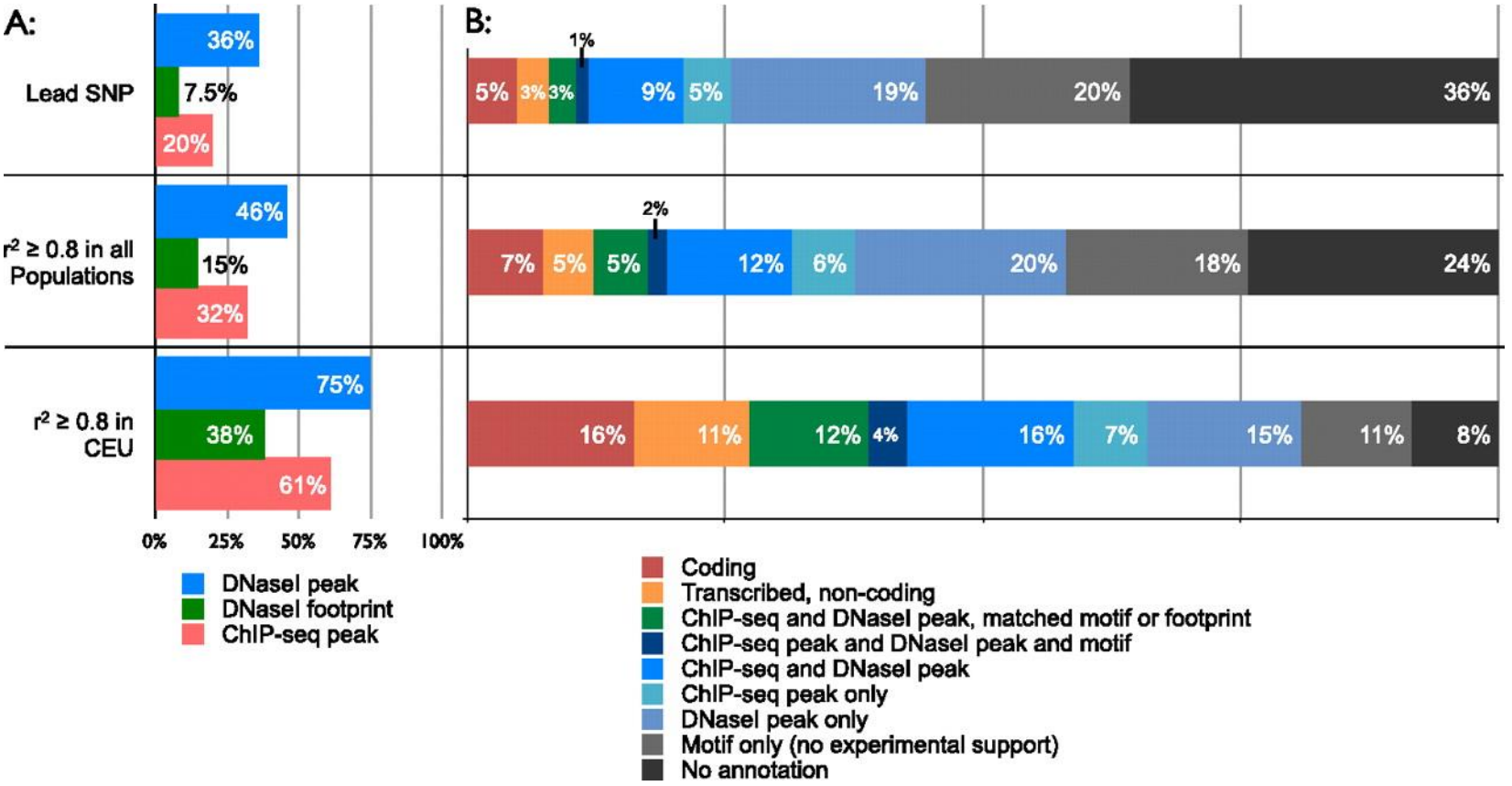
S Djebali *et al. Nature* **489** 101-108 (2012)

We are more than just proteins

DNA

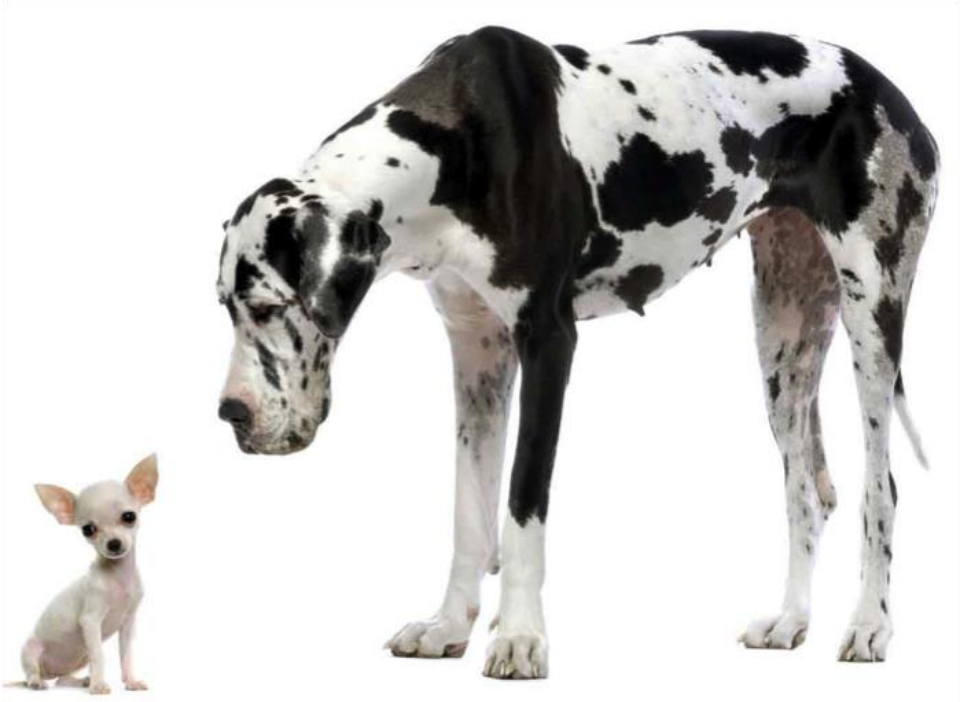
Protein coding gene

Most disease associations are NOT to protein coding regions

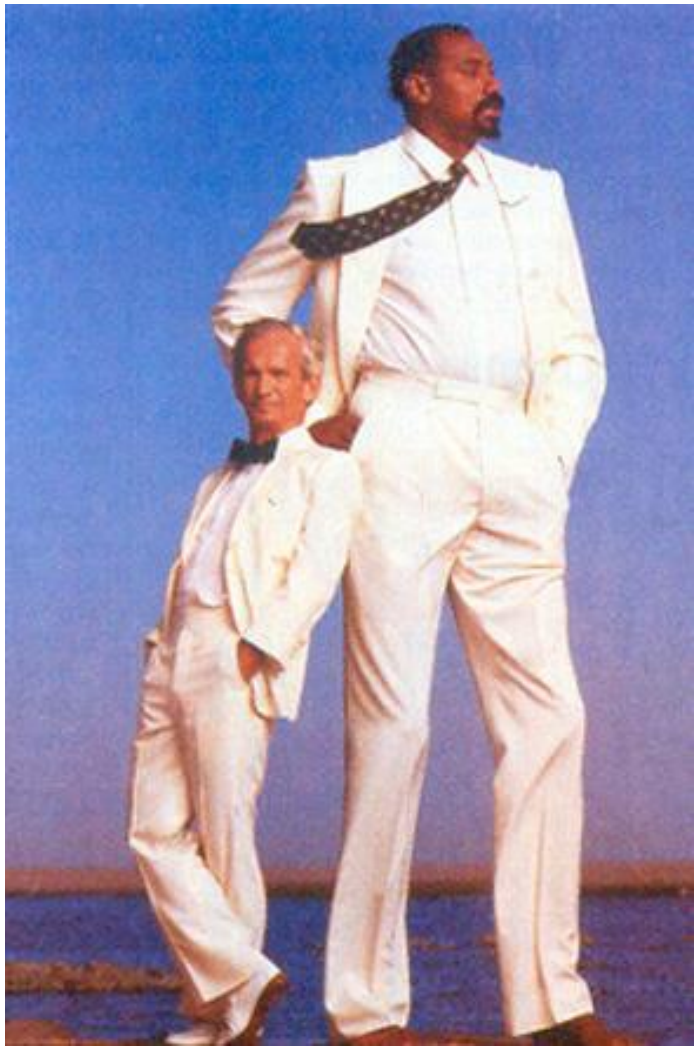


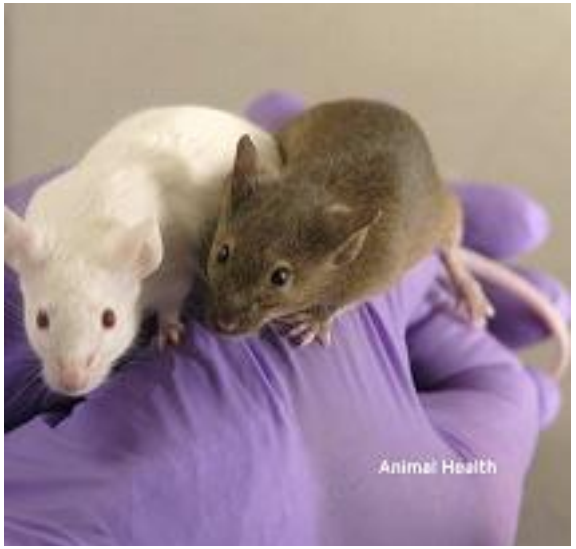
Schaub M A et al. Genome Res. 2012;22:1748-1759

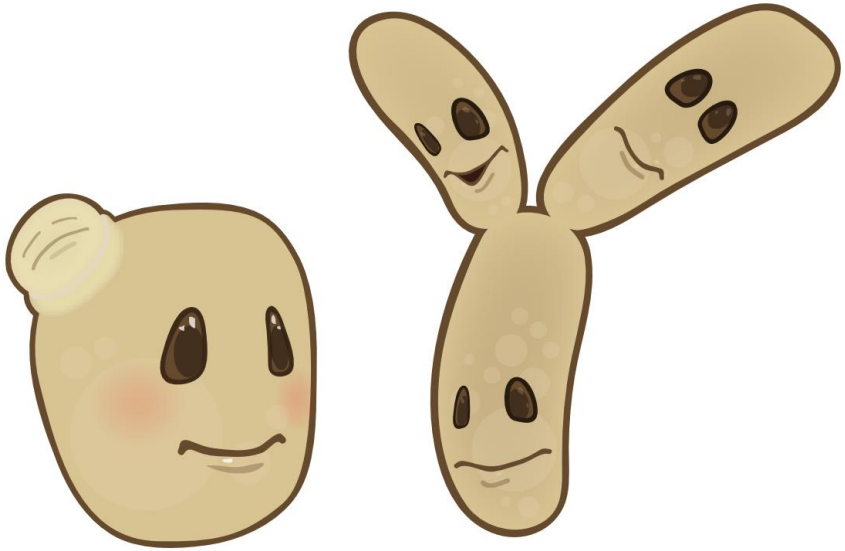
The Dowell Laboratory



Studying Individual Differences

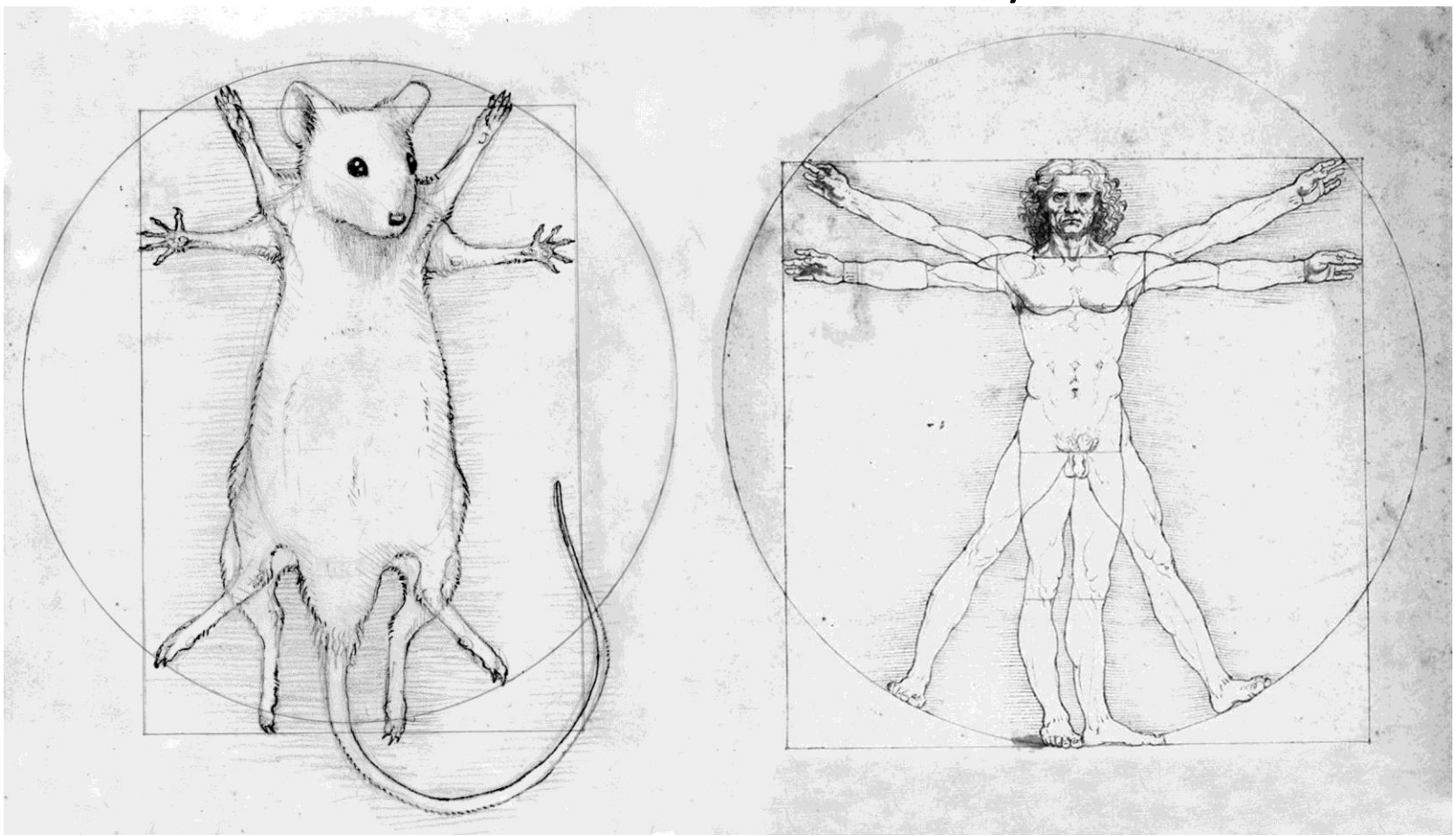


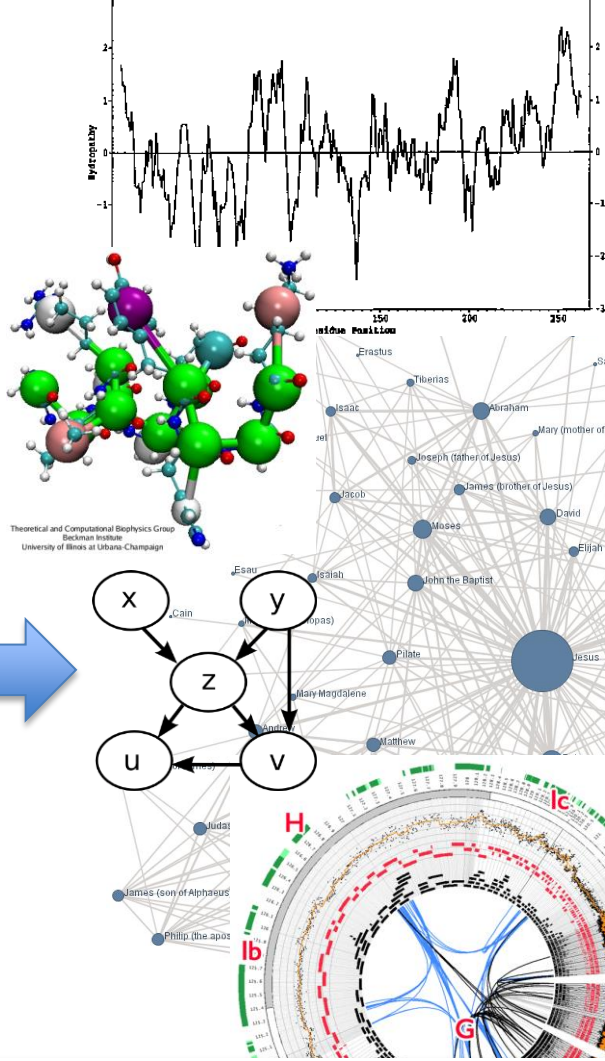




“Nothing in biology makes sense
except in the light of evolution.”

--Theodosius Dobzhansky



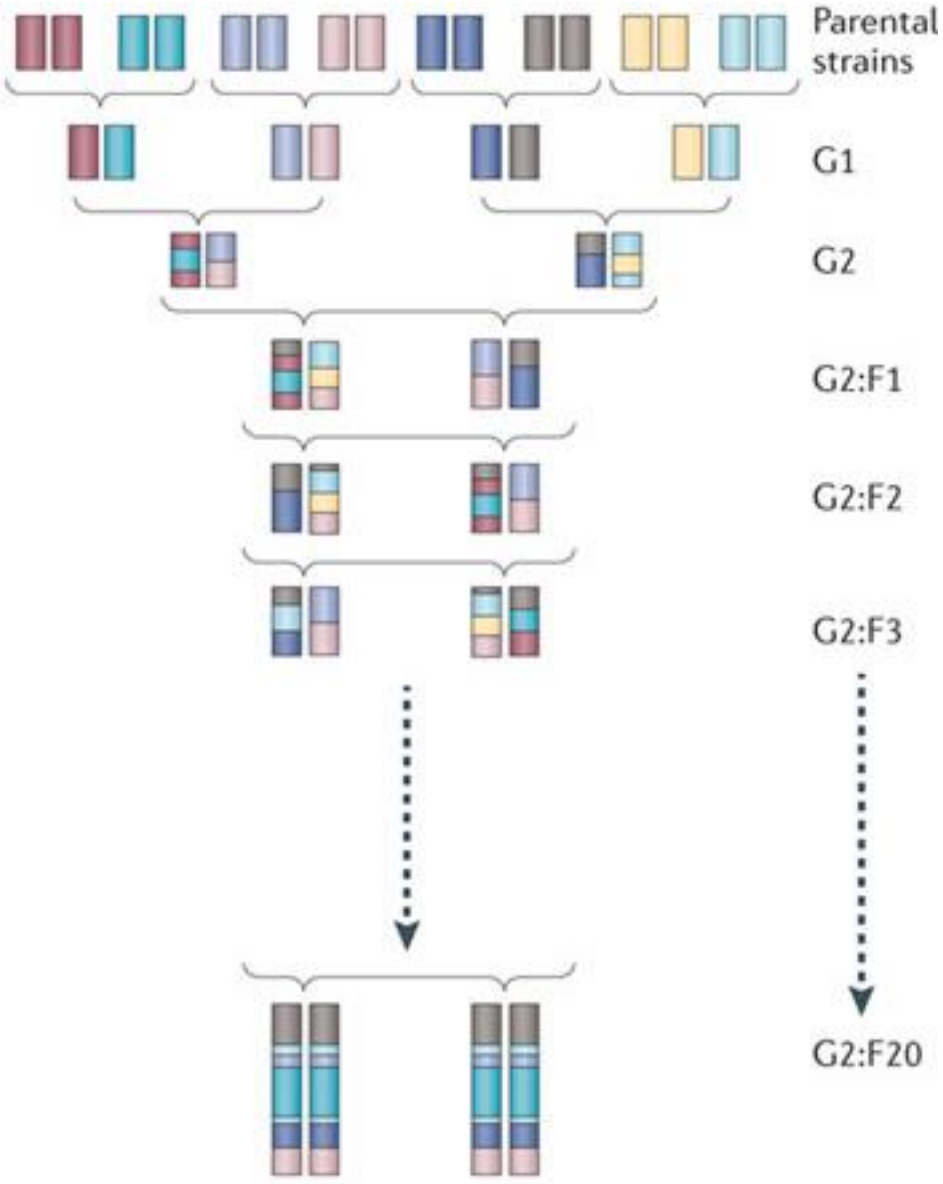


Biological Experiment,
Hypothesis Generation

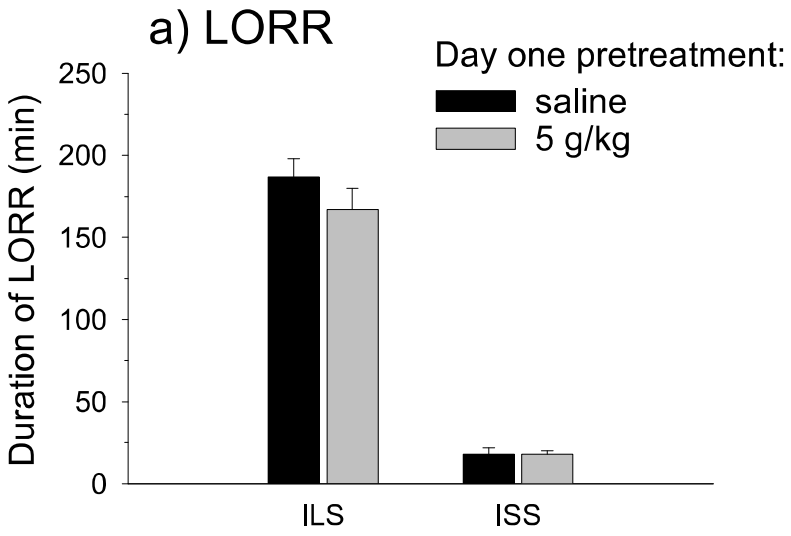
Data Management and
Organization

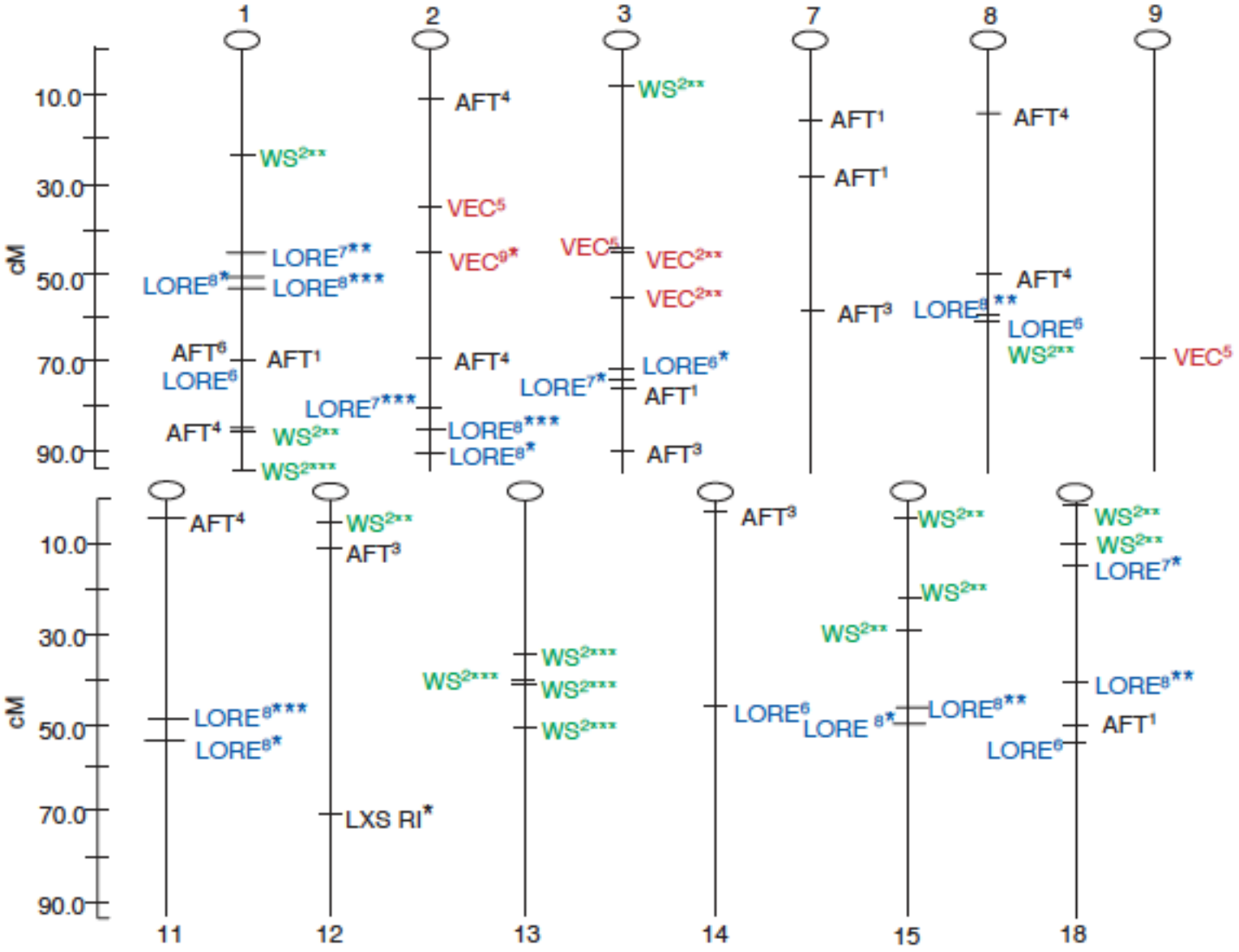
Analysis
Machine Learning,
Modeling,

Case I: Alcohol Tolerance



Recombinant Inbred Mice: LXS strains

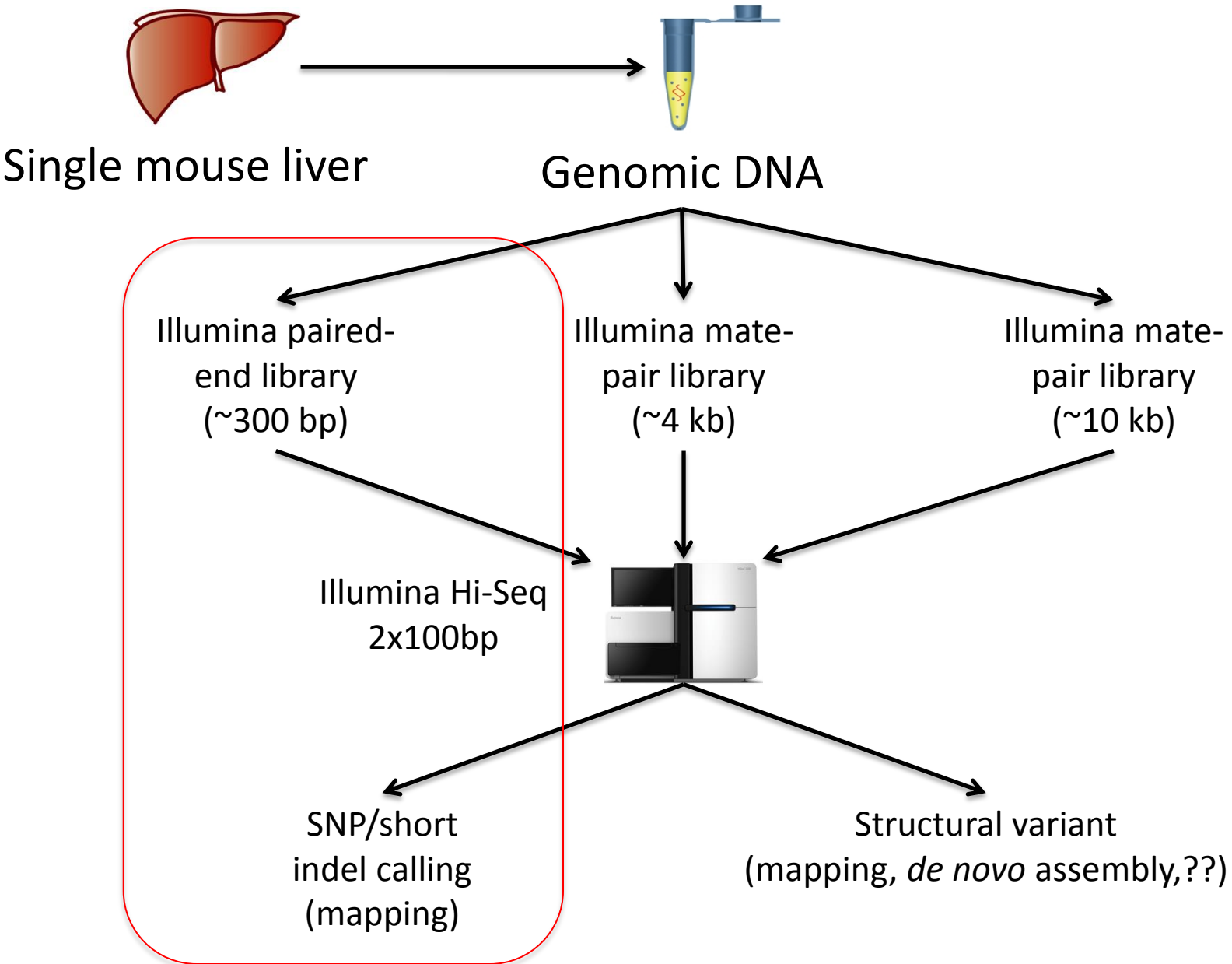




* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$

TRENDS in Genetics

Overall sequencing strategy



Sequencing Data (Hi-Seq 2x100bp)
8 lanes

	ILS	ISS
Total read pairs	427,367,178	439,844,784
Total bases	85,473,435,600	87,968,956,800
Avg Coverage (mm9)	31.3 X	32.3 X

Variation compared to mm9

	# Variants	Percent
Total Variants	9,289,436	100%
snp128	4,547,533	49.0%
indels	1,994,199	21.4%

	BWA	both	bowtie	total
ILS	1,250,824	5,261,670	539,884	7,052,378
ISS	934,491	4,644,440	524,765	6,103,696

Identification of LXS Markers

Total variants: 9,289,436

Removal of variants common to both strains

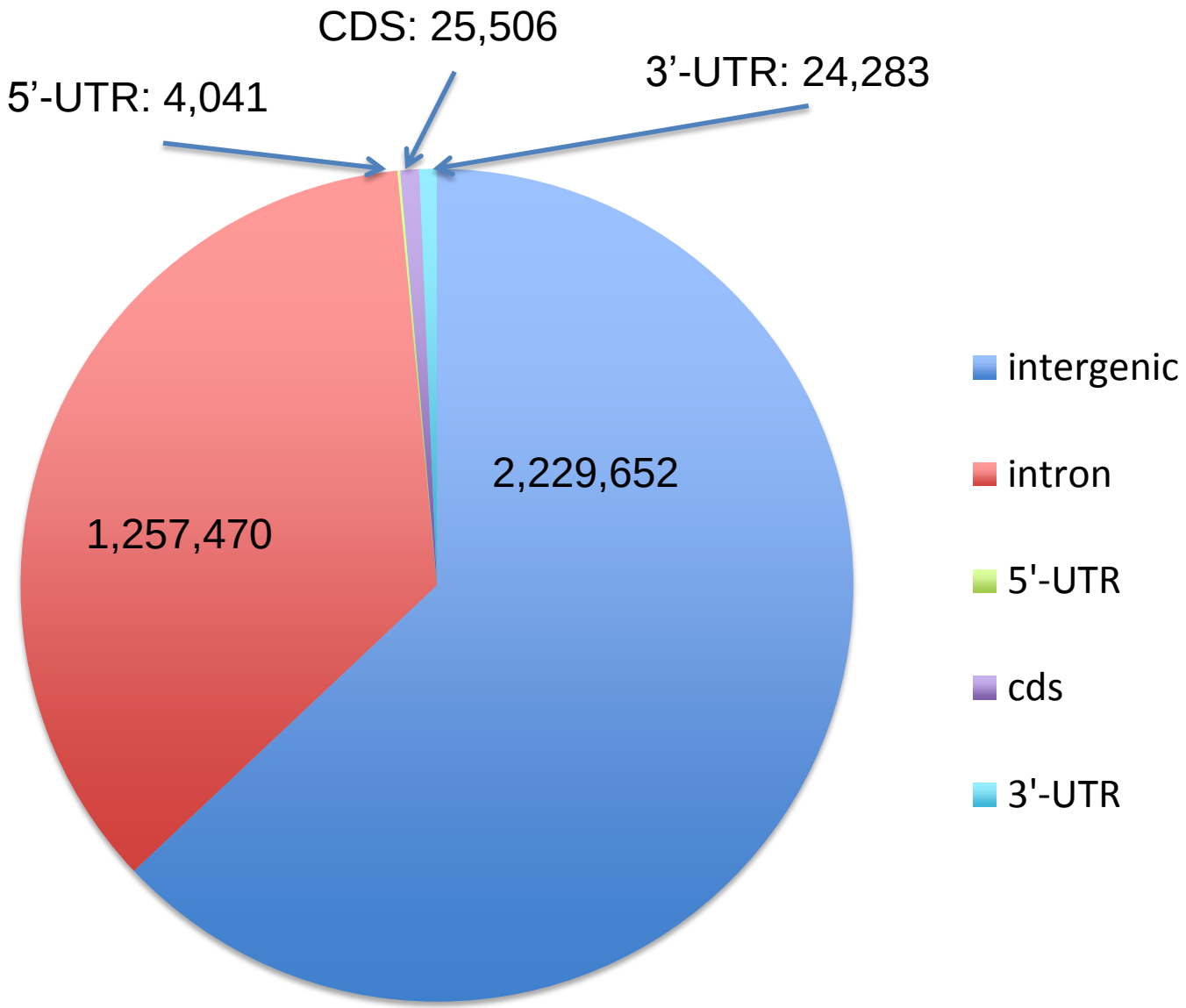
Potential variants: 6,406,493

Removal of ambiguous markers between bwa and bowtie

High quality: 3,537,009

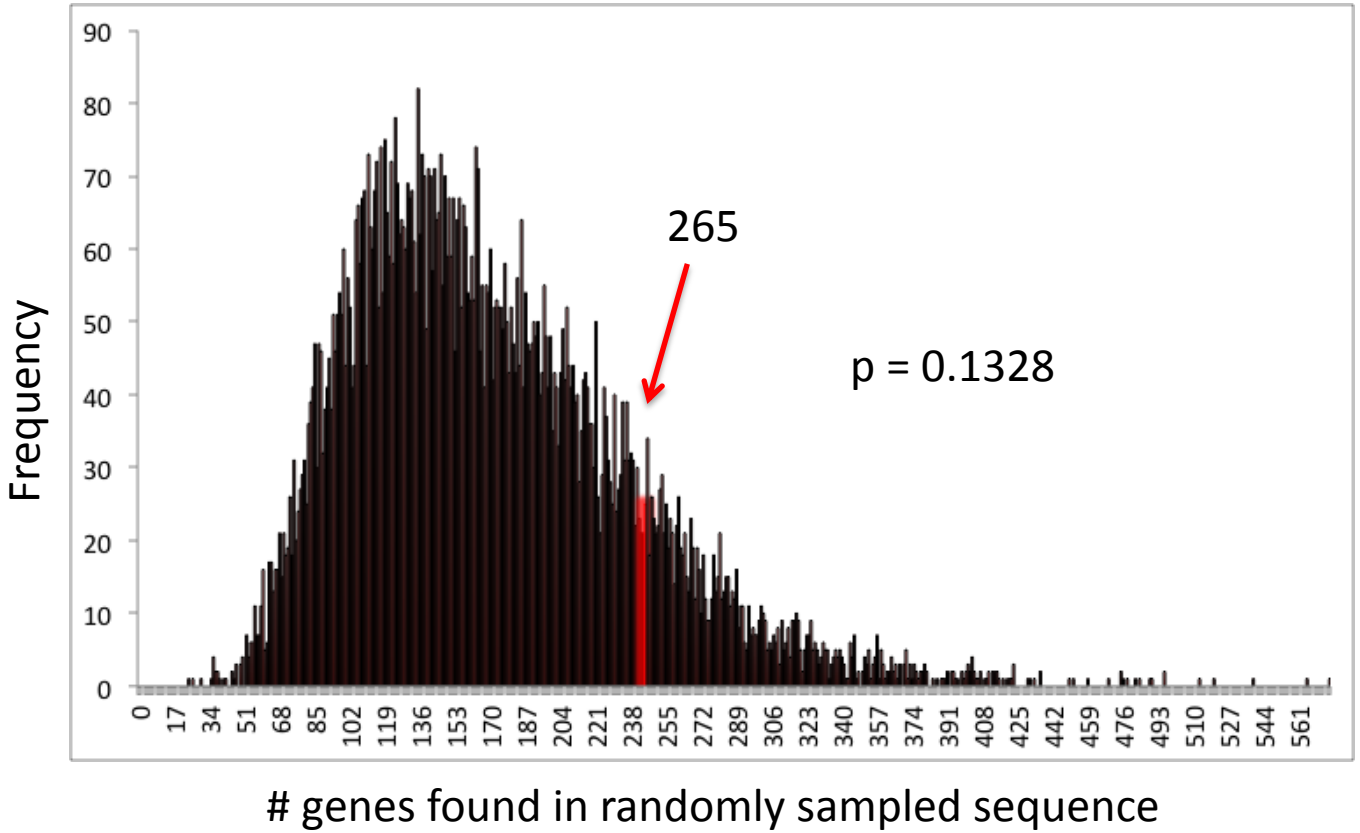
1 marker every ~760 base pairs

Breakdown of marker by refseq

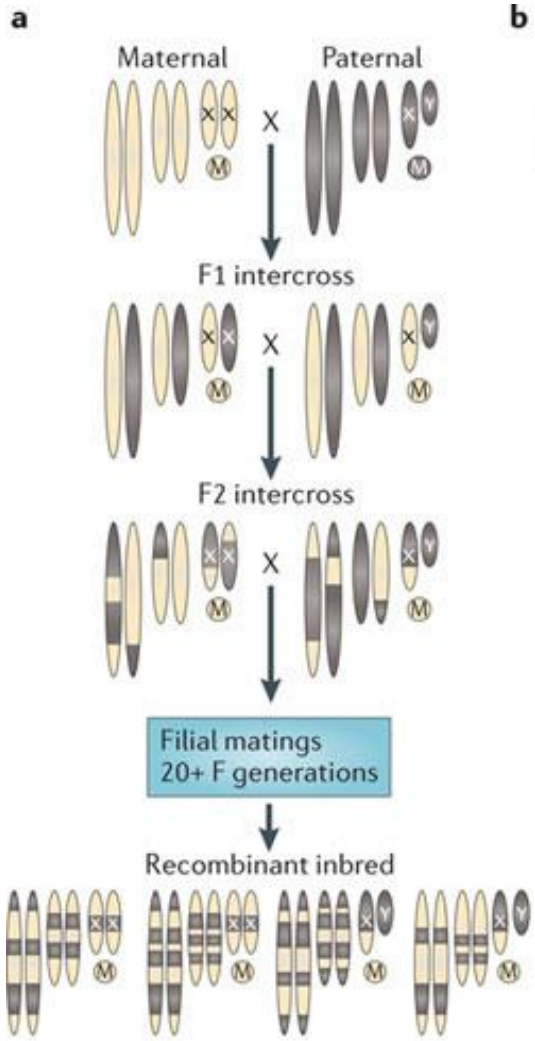


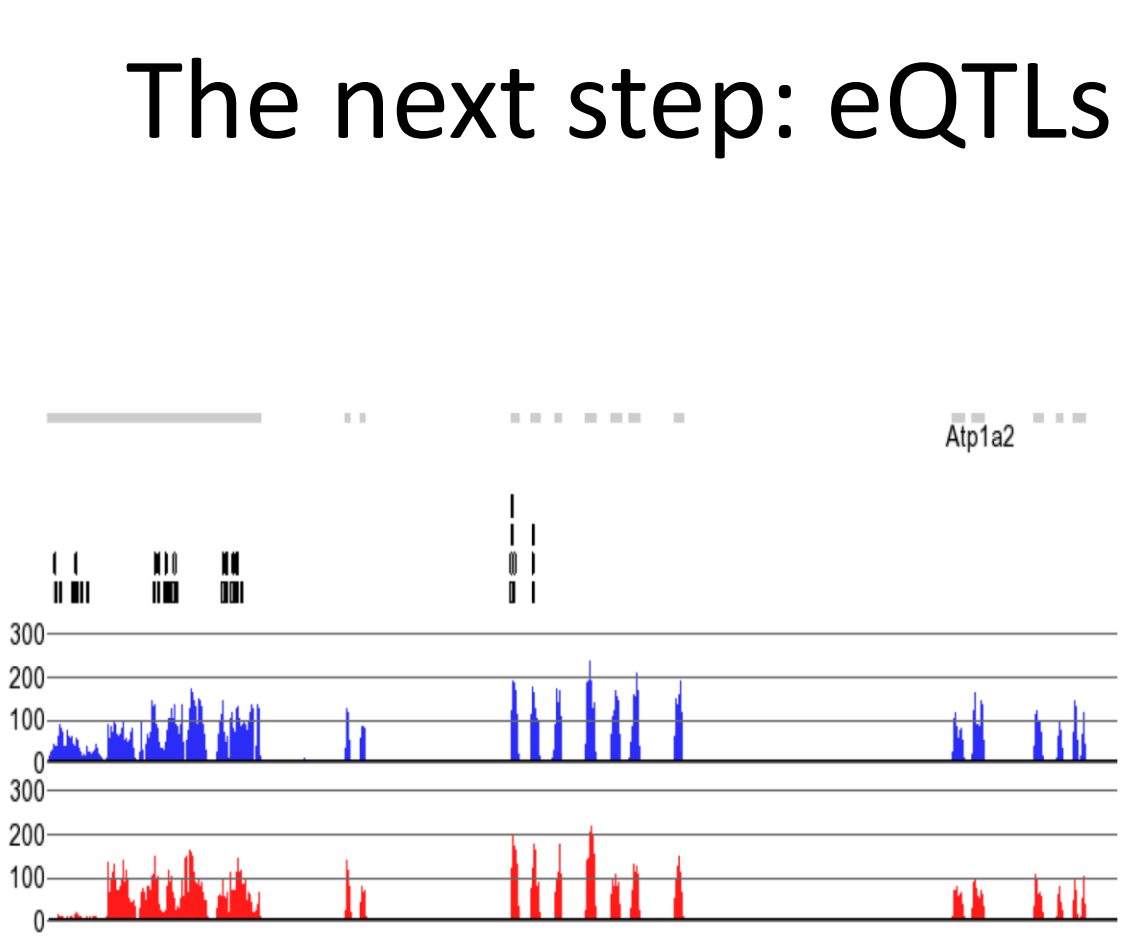
Are protein-altering variations enriched in QTL hotspots?

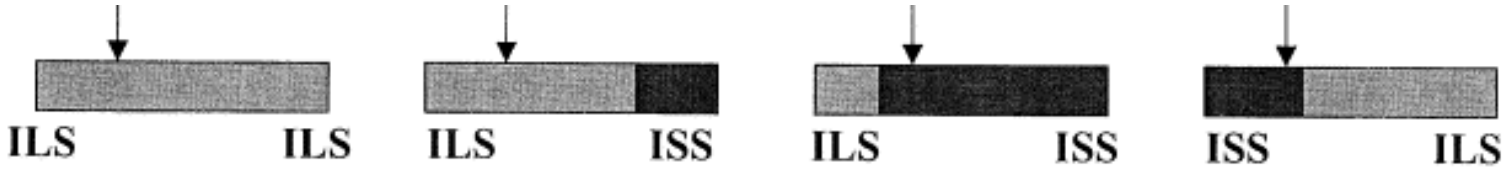
- 3655 genes with protein sequence differences between ILS and ISS
- 265 genes with aa differences found in QTL regions
- 10,000 randomly samples genome with intervals the same size of QTL regions



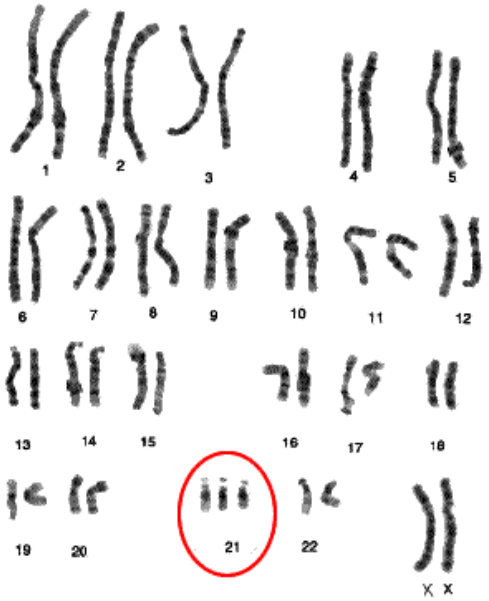
The next step: eQTLs







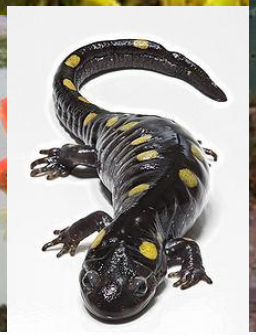
Case II: Increased copy number



Down Syndrome

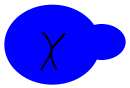
Chris Jepson Photography

Polyploidy and Aneuploidy occur frequently in nature



Rapid evolution of yeast tetraploids in raffinose

Haploid (1N)

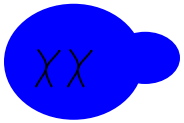


pGAL YFP tADH

OR

pGAL CFP tADH

Diploid (2N)



pGAL YFP tADH

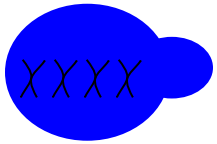
pGAL YFP tADH

OR

pGAL CFP tADH

pGAL CFP tADH

Tetraploid (4N)



pGAL YFP tADH

pGAL YFP tADH

pGAL YFP tADH

pGAL YFP tADH

OR

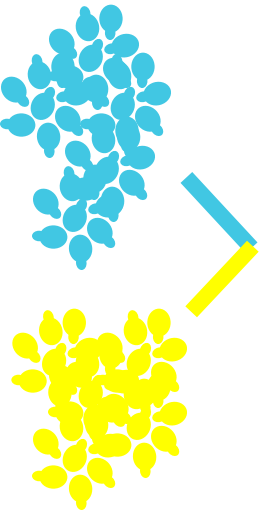
pGAL CFP tADH

pGAL CFP tADH

pGAL CFP tADH

pGAL CFP tADH

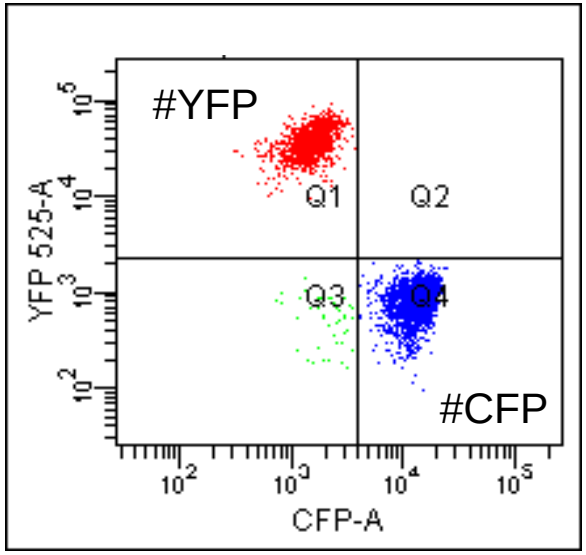
In vitro evolution experimental outline



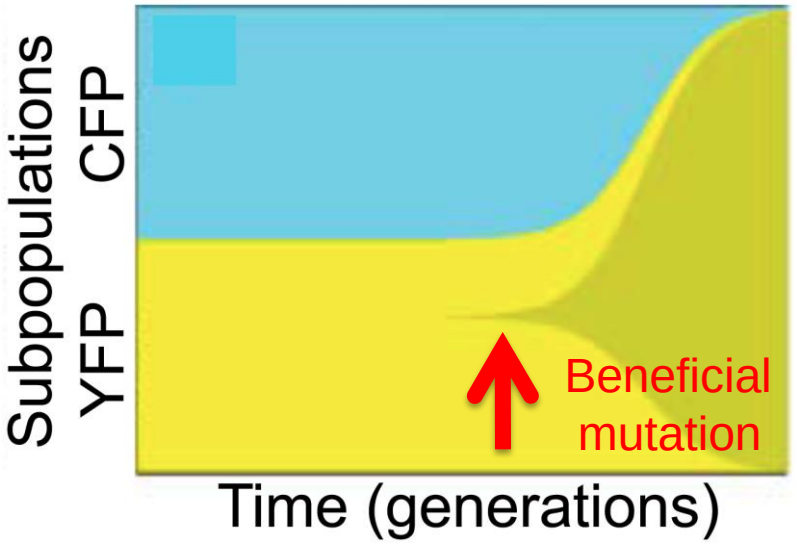
Mix 1:1
CFP:YFP ancestors

Passage every 24 hours,
Flow every 24 hours

Endpoint at 240 generations

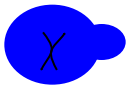


Flow Cytometry

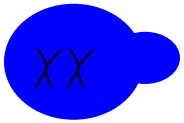


Rapid evolution of yeast tetraploids in raffinose

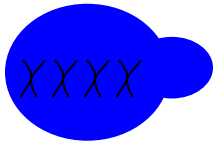
Haploid (1N)

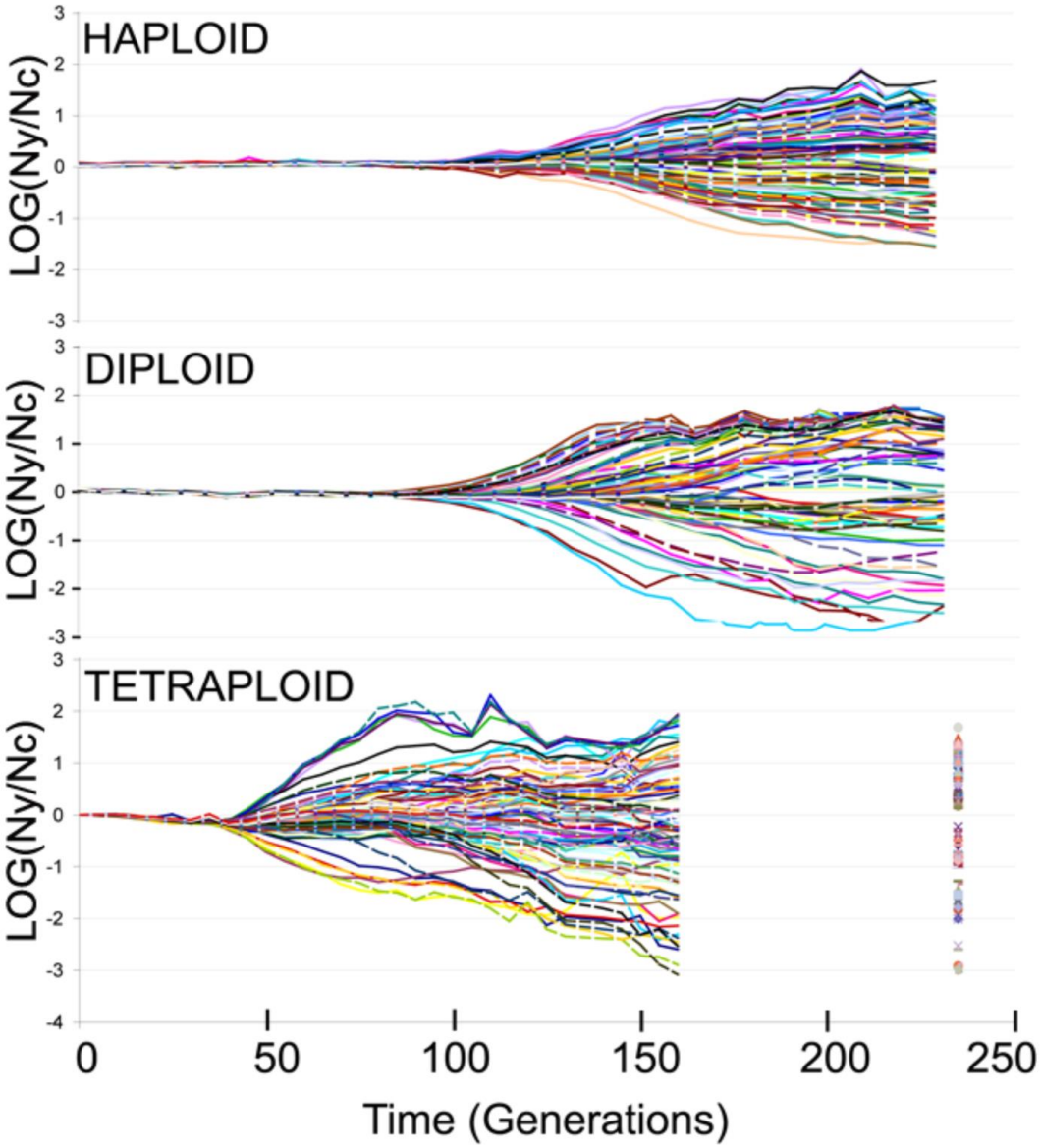


Diploid (2N)



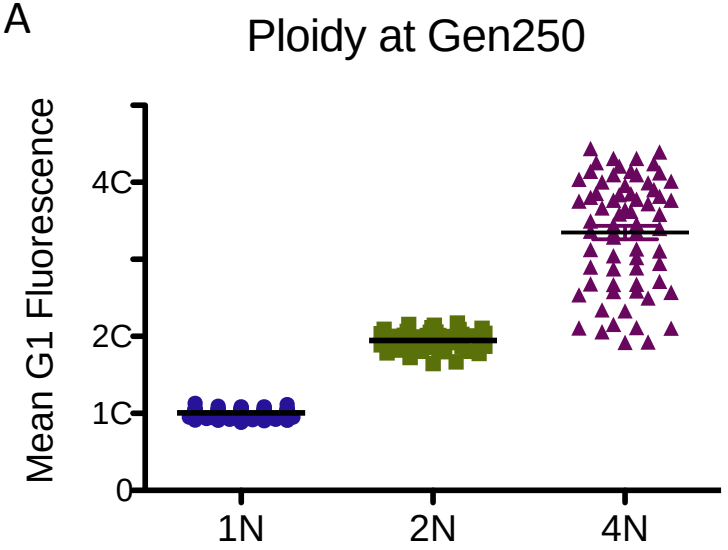
Tetraploid (4N)



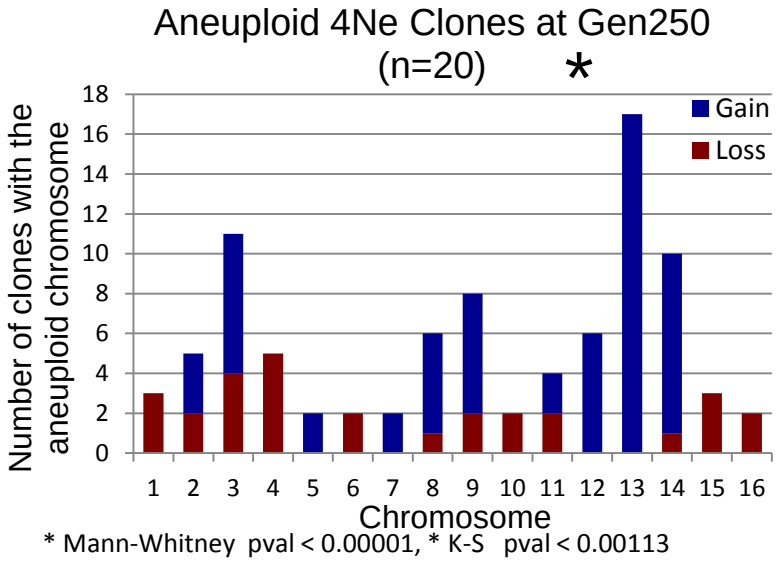


Recurrent aneuploidies occurred during 4N adaptation

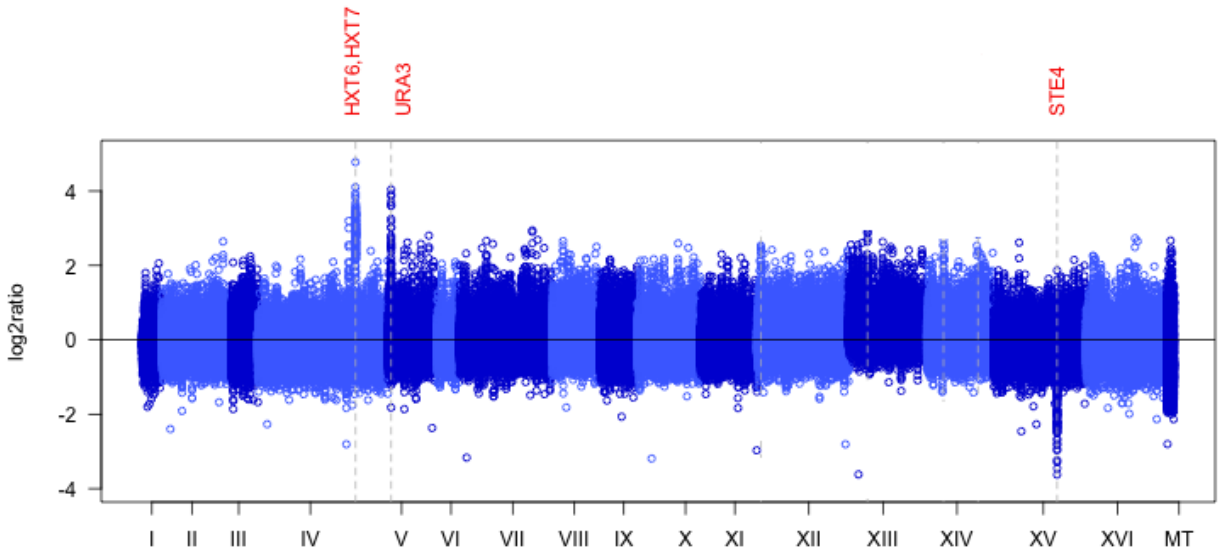
A



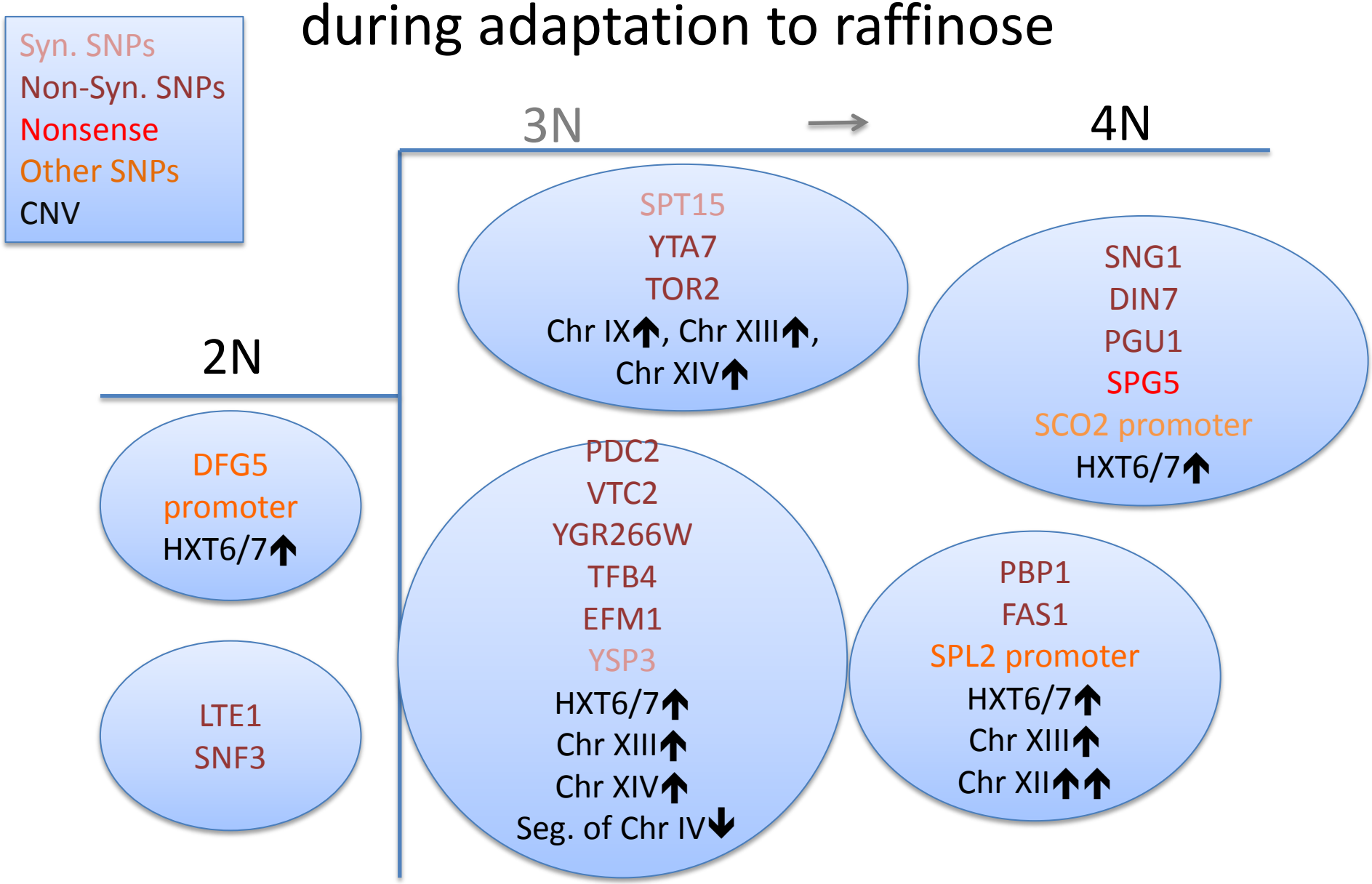
B



C

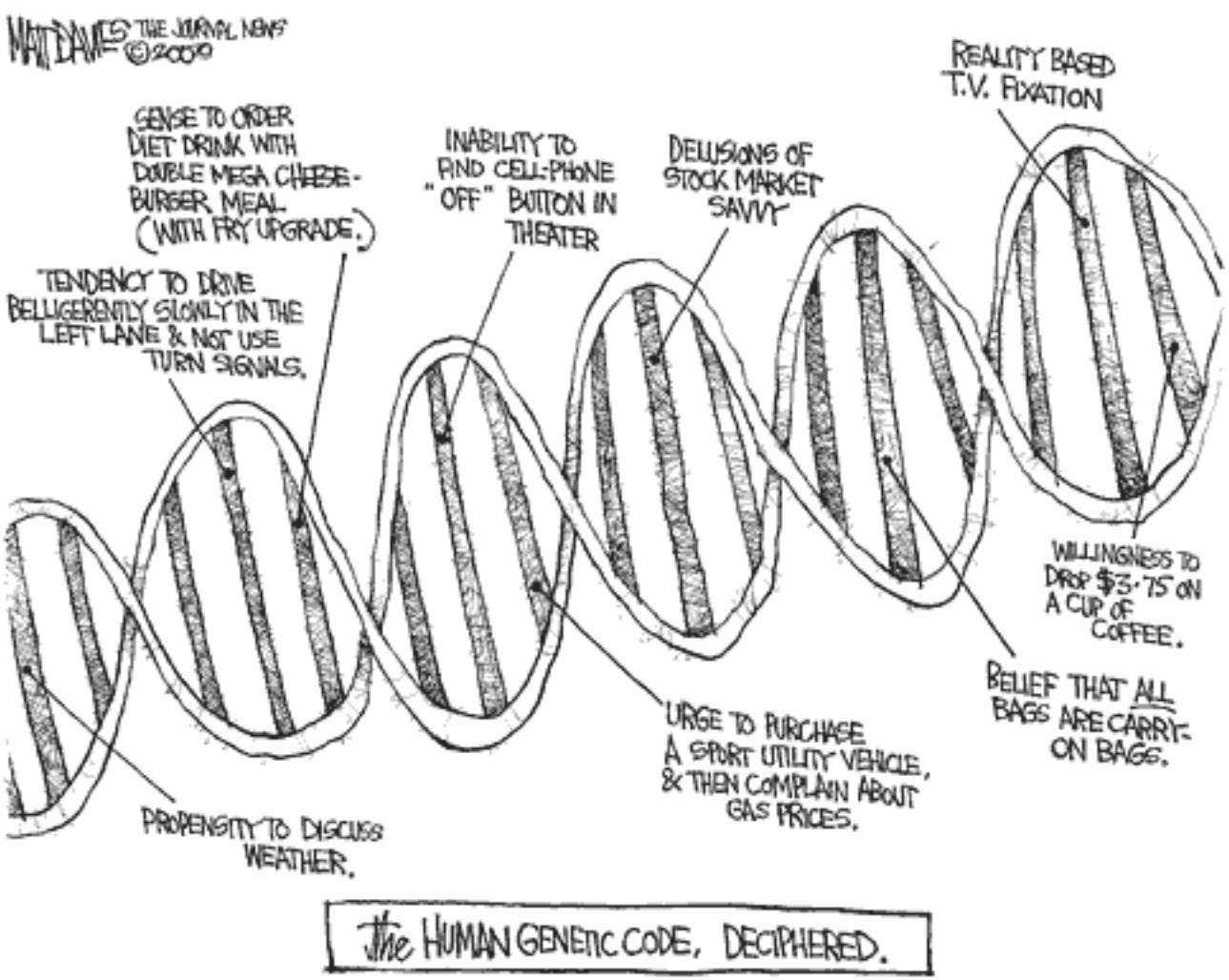


2N and 4N clones acquire genetic mutations during adaptation to raffinose

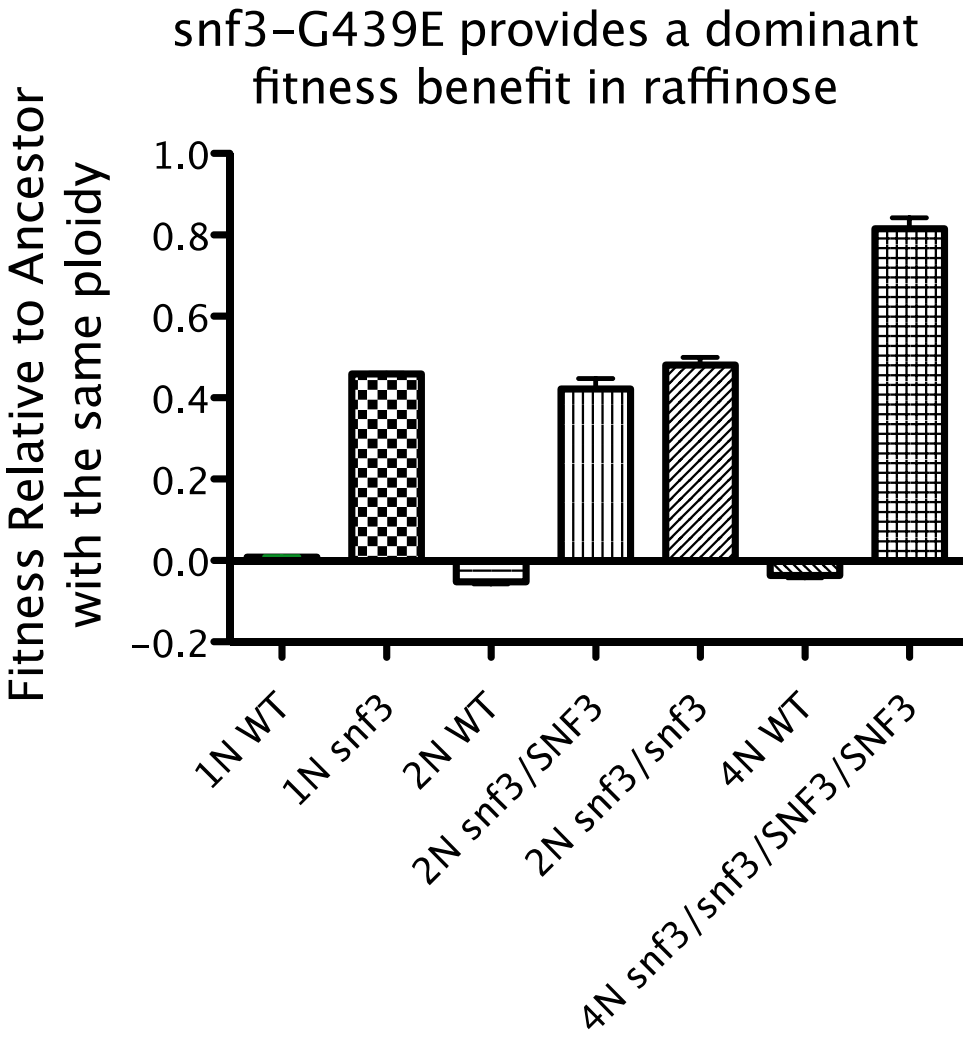


Our work and others suggest the mutation rate is generally higher in cells of higher ploidy.

Still far from “causality”



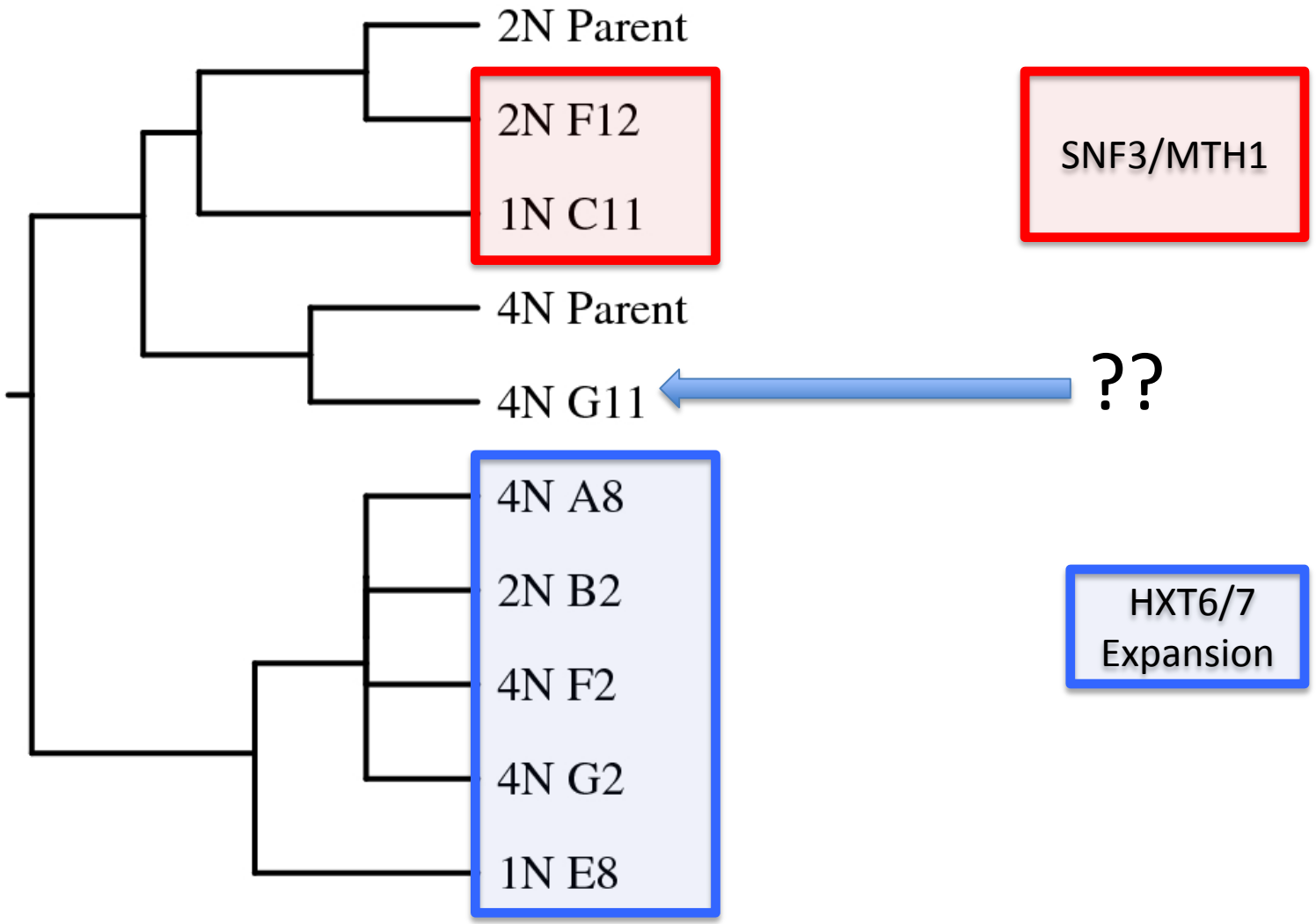
Evidence for dominant beneficial mutations



- *No fitness cost in rich media.*
- *Fitness effect similar to HXT6/7 amplification.*

Vtc2 (vacuolar transport chaperone) mutant data – **suggests** 4N cells also acquire neutral to slightly deleterious mutations

Identify at least two pathways to adaptation!



Yeast teach us a lot about aneuploidy

Identification of Aneuploidy-Tolerating Mutations

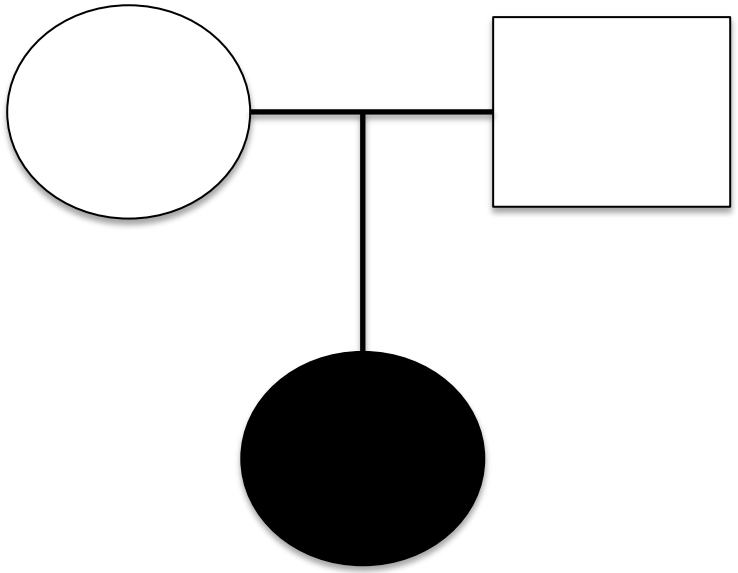
Eduardo M. Torres,[1,2] Noah Dephoure,[3] Amudha Panneerselvam,[1] Cheryl M. Tucker,[4] Charles A. Whittaker,[1] Steven P. Gygi,[3] Maitreya J. Dunham,[5] and Angelika Amon[1,2,*]
[1]David H. Koch Institute for Integrative Cancer Research
[2]Howard Hughes Medical Institute
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Department of Cell Biology, Harvard University Medical School, Boston, MA 02115, USA
[4]Lewis-Sigler Institute, Princeton University, Princeton, NJ 08540, USA
[5]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
*Correspondence: angelika@mit.edu
DOI 10.1016/j.cell.2010.08.038

Aneuploidy Drives Genomic Instability in Yeast

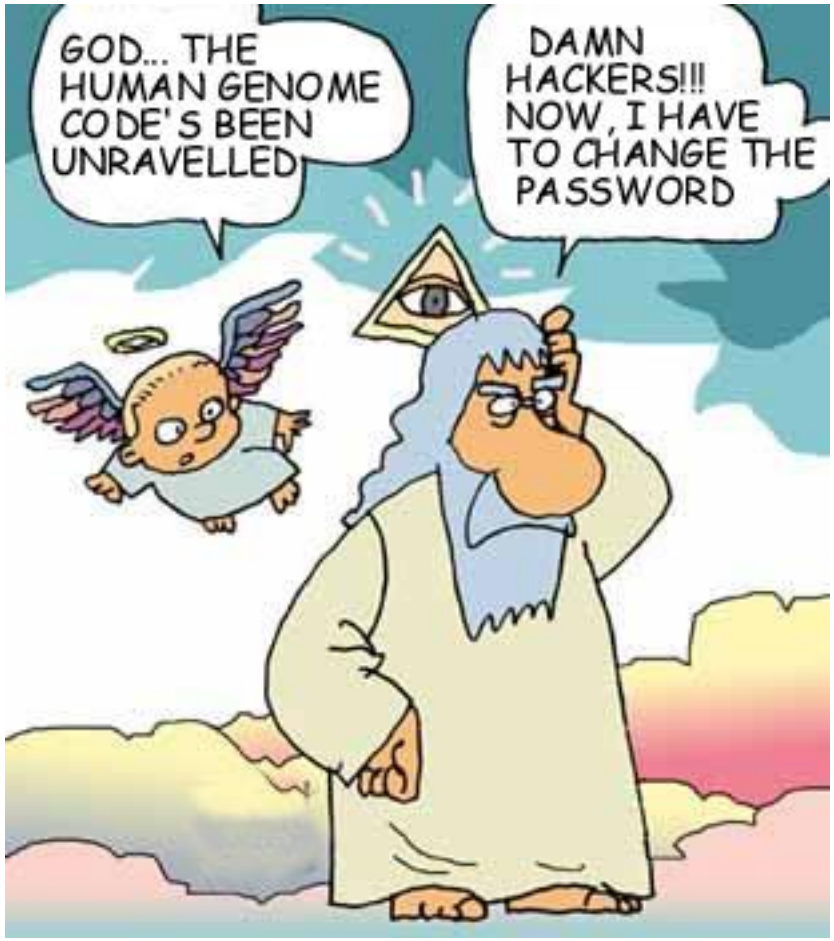
Jason M. Sheltzer,[1] Heidi M. Blank,[1] Sarah J. Pfau,[1] Yoshie Tange,[2] Benson M. George,[1] Timothy J. Humpton,[1] Ilana L. Brito,[3] Yasushi Hiraoka,[2,4] Osami Niwa,[5] Angelika Amon[1*]

Aneuploidy decreases cellular fitness, yet it is also associated with cancer, a disease of enhanced proliferative capacity. To investigate one mechanism by which aneuploidy could contribute to tumorigenesis, we examined the effects of aneuploidy on genomic stability. We analyzed 13 budding yeast strains that carry extra copies of single chromosomes and found that all aneuploid strains exhibited one or more forms of genomic instability. Most strains displayed increased chromosome loss and mitotic recombination, as well as defective DNA damage repair. Aneuploid fission yeast strains also exhibited defects in mitotic recombination. Aneuploidy-induced genomic instability could facilitate the development of genetic alterations that drive malignant growth in cancer.

The next step: Down sequencing



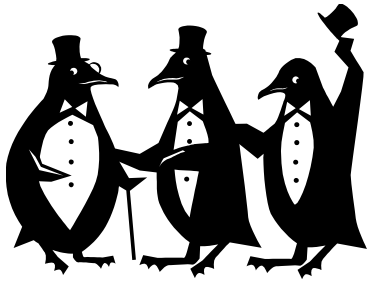
LINDA CRNIC INSTITUTE FOR DOWN SYNDROME


GOD... THE HUMAN GENOME CODE'S BEEN UNRAVELLED
DAMN HACKERS!!! NOW, I HAVE TO CHANGE THE PASSWORD

Acknowledgements

Dowell Laboratory, University of Colorado

Mary Allen, Phil Richmond, Justin Freeman, Aaron Odell, David Knox, Jess Vera, Tim Read, Li Wang, Joe Rokicki, Amber Sorenson, Joey Azofeifa, Anna Lee, Josephina Hendrix



Collaborators:

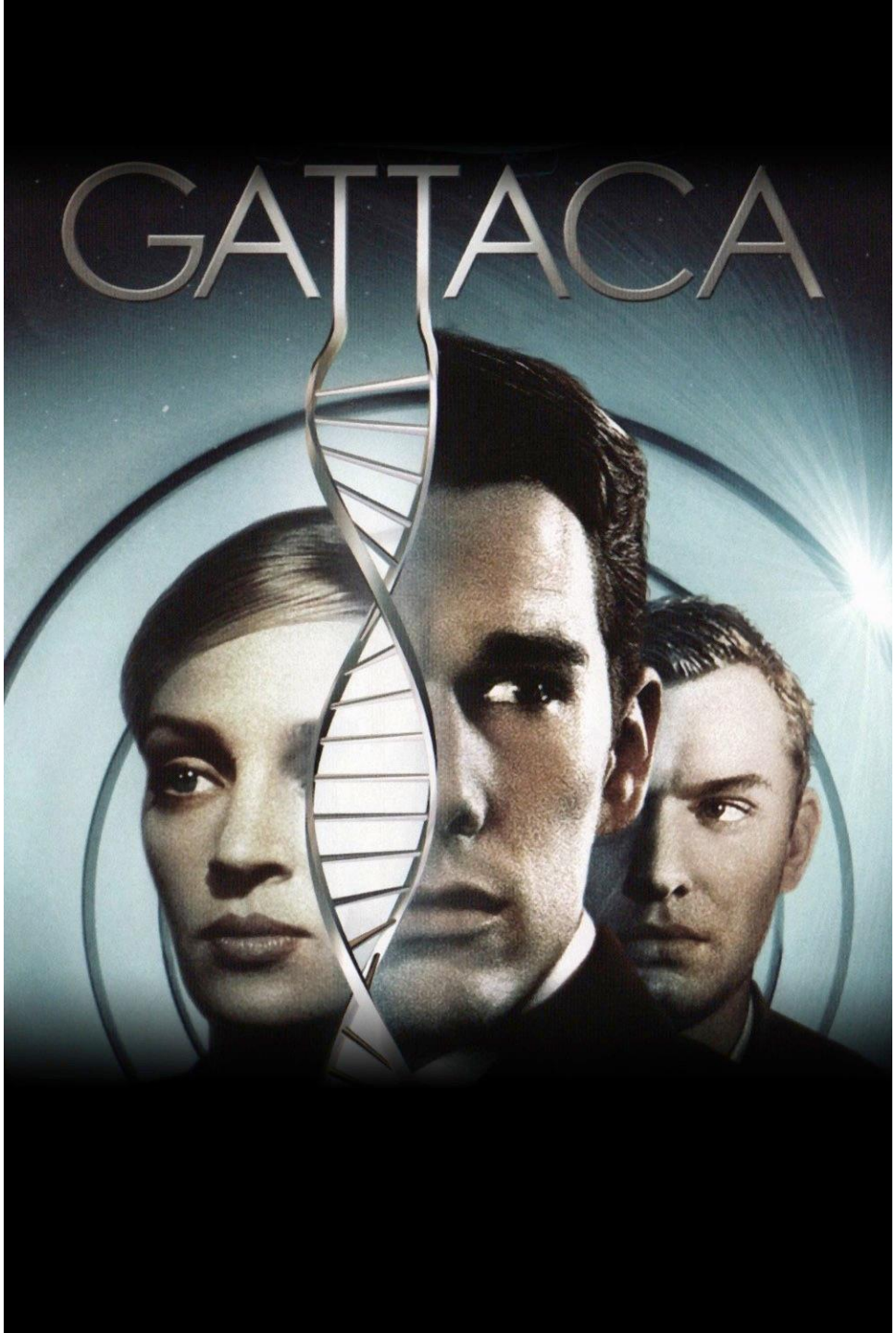
- David Pellman and Anna Selmecki, Dana Farber Cancer Institute
- Joaquin Espinosa, Shelley Copley, Rui Yi, MCDB Univ. of Colorado
- Richard Radcliffe, UC Denver Health Sciences
- Matt Posewitz, Colorado School of Mines

Funding Sources:

- Boettcher Foundation
- Sloan Foundation Fellowship
- Butcher Seed Grant
- Linda Crnic Institute Seed Grant

BOETTCHER FOUNDATION

LINDA CRNIC INSTITUTE FOR DOWN SYNDROME


GATTACA