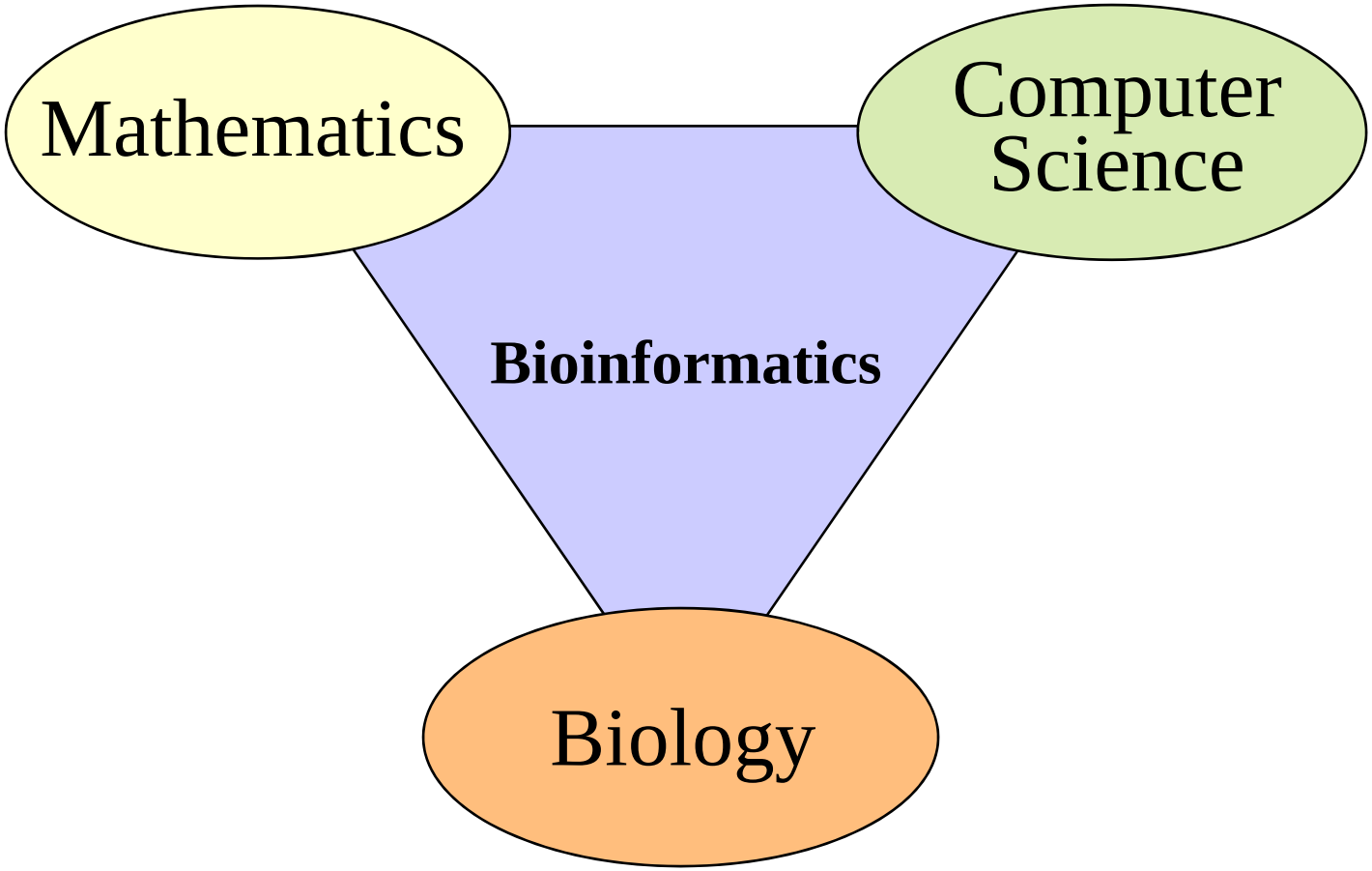
DAVIDSON

Bringing Computation to Life

Laurie J. Heyer
L.R. King Assoc. Professor of Mathematics
Davidson College


Mathematics
Computer Science
Bioinformatics
Biology

Outline

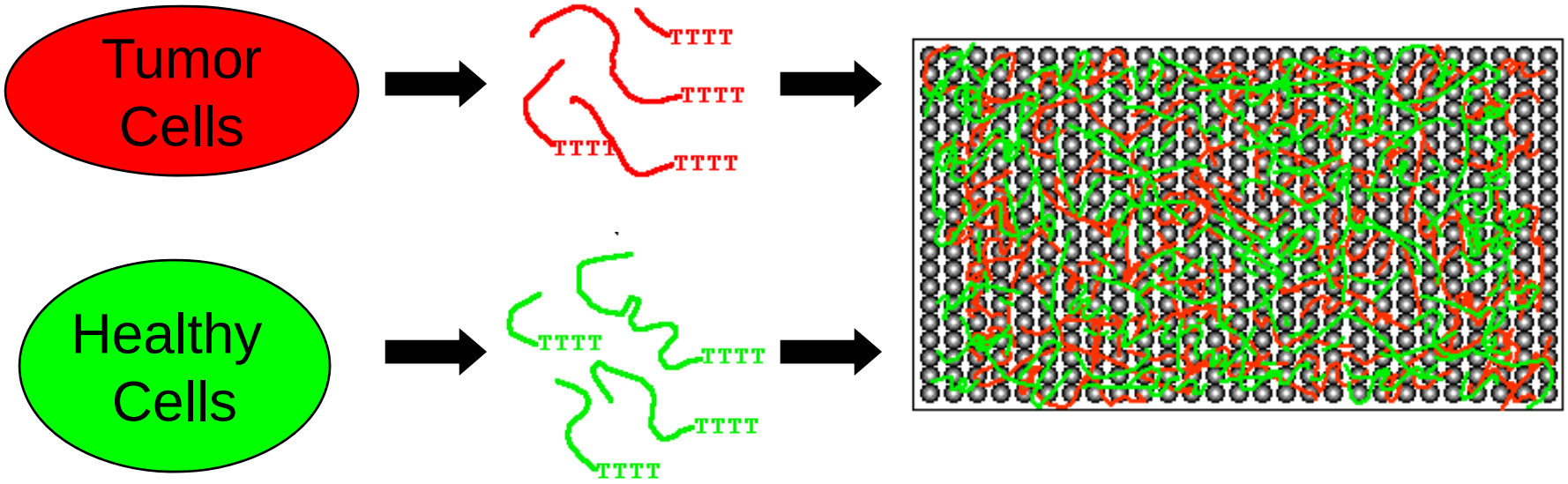
- Computing in the Post-Genomic Era
  - Microarray Data Analysis
- “Living Hardware”
  - Computing with bacteria using principles of synthetic biology

Human Genome Project

- 3.2 billion basepairs
  ~ 3 gigabytes
  - human only
  - no annotation
- DNA in GenBank
  ~ 56,037,734,462 basepairs
- Sequence comparison

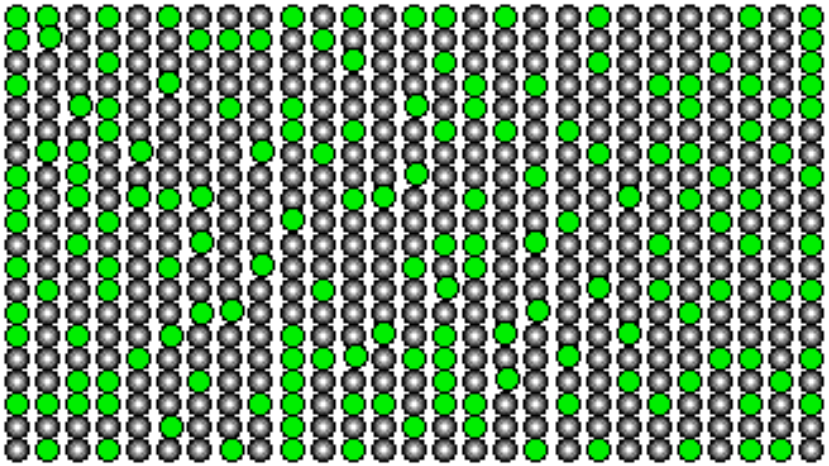
  n^2 space
  ~ 1 exabyte (1 million terabytes)
  to compare two human sized genomes

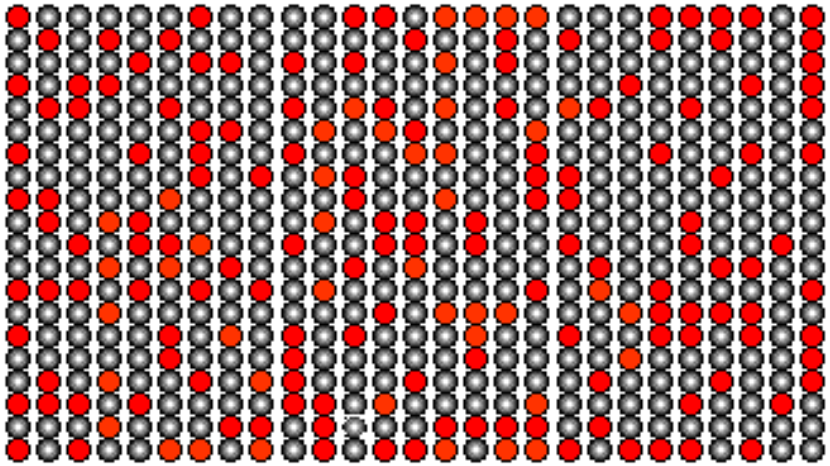
Computation in the Post-Genomic Era: Microarray Experiments



1. Isolate RNA from cells
2. Fluorescently label RNA
3. Hybridize to microarray

Scanning Microarrays



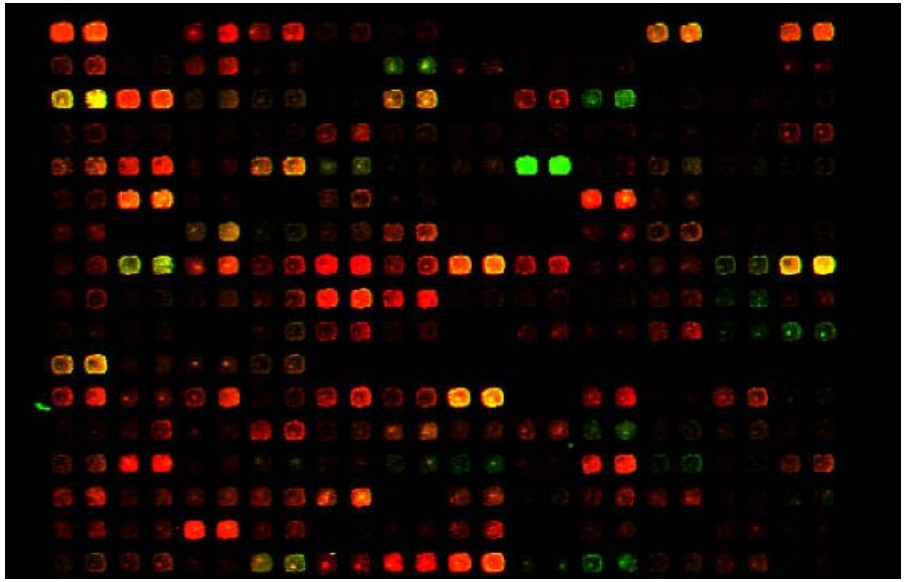


2 TIFF images
x 20 MB / TIFF image

---

40 MB image data

One microarray =



17,000 genes
1 Red:Green ratio / gene
x 4 bytes / ratio

---

66 KB ratio data

One Microarray Study =

3 microarrays / patient
x 69 patients

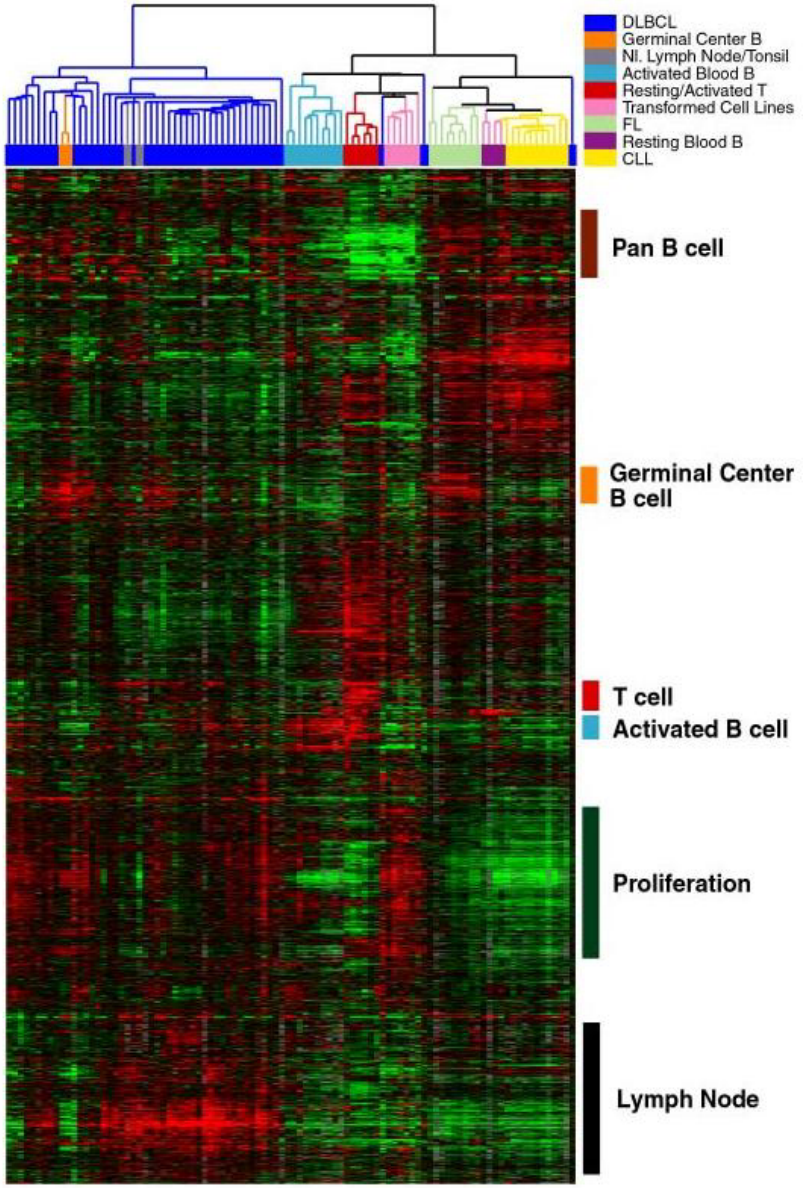
---

8 GB image data
4 MB ratio data

$_{17,000}C_2$ (= 145 million)
pairs of genes
x 4 bytes / correlation

---

550 MB similarity data



MicroArray Genome Imaging and Clustering Tool

Image Analysis (Segmentation)

- Locate spot (signal) pixels
- Measure intensity of signal and background in each channel
- Compute ratio

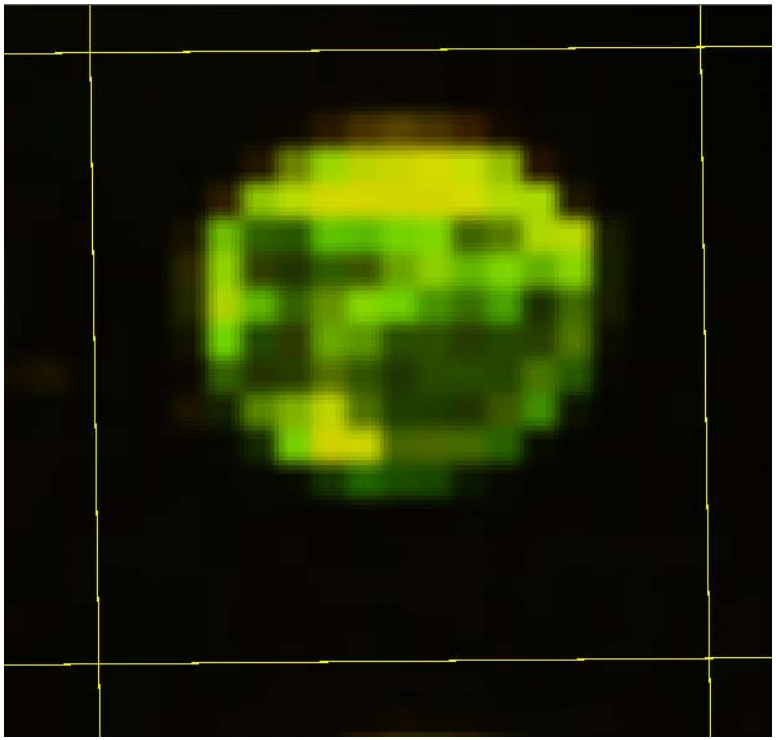
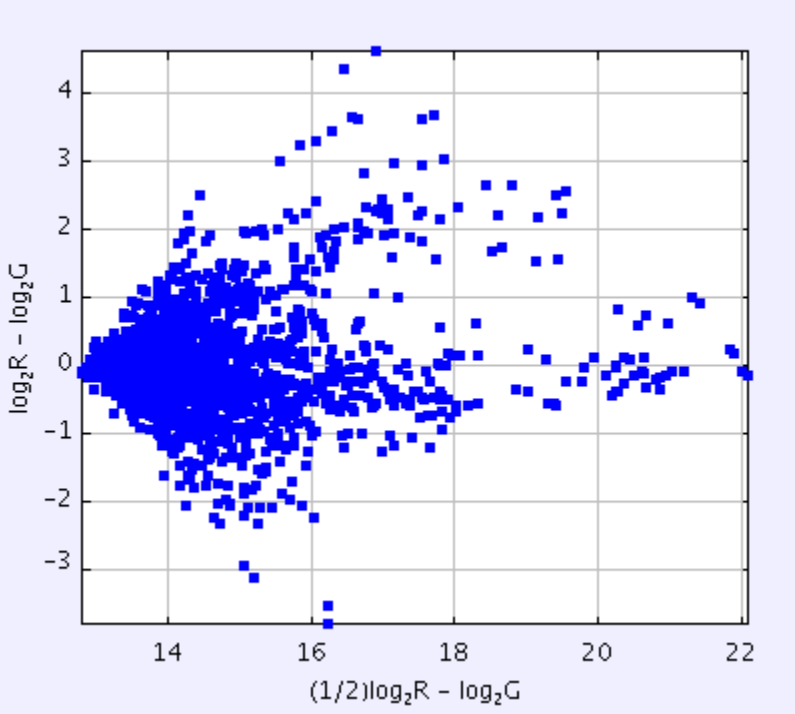
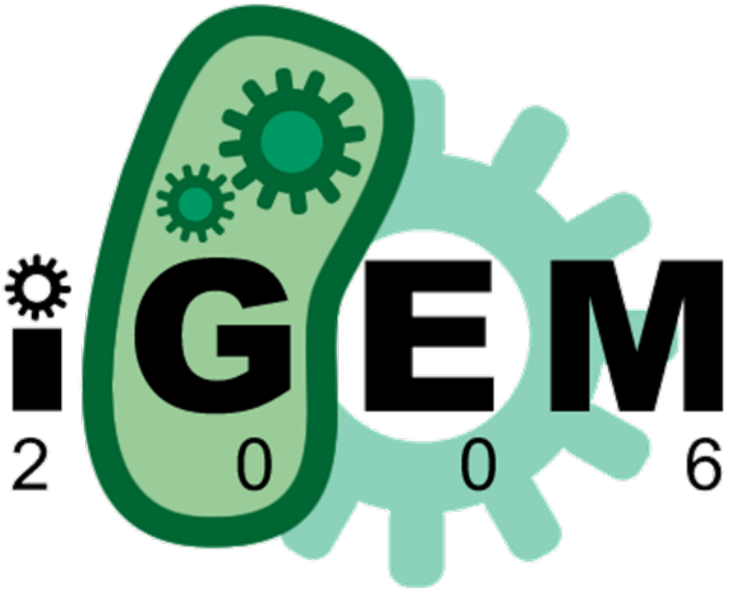


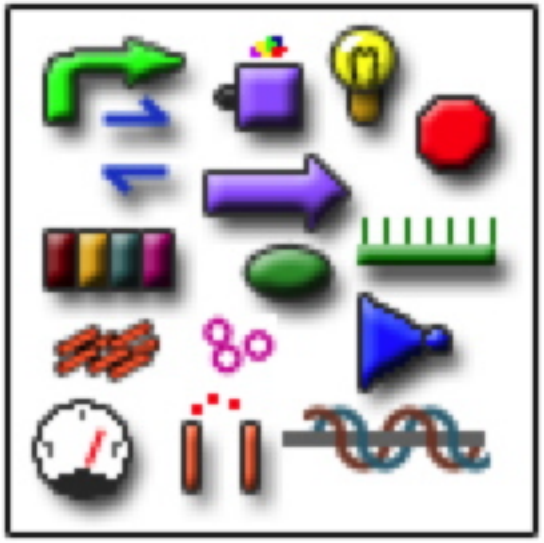
Image Quality Assessment

- Assess intensity-dependent effects
- Normalization methods
  - median shift
  - LOWESS

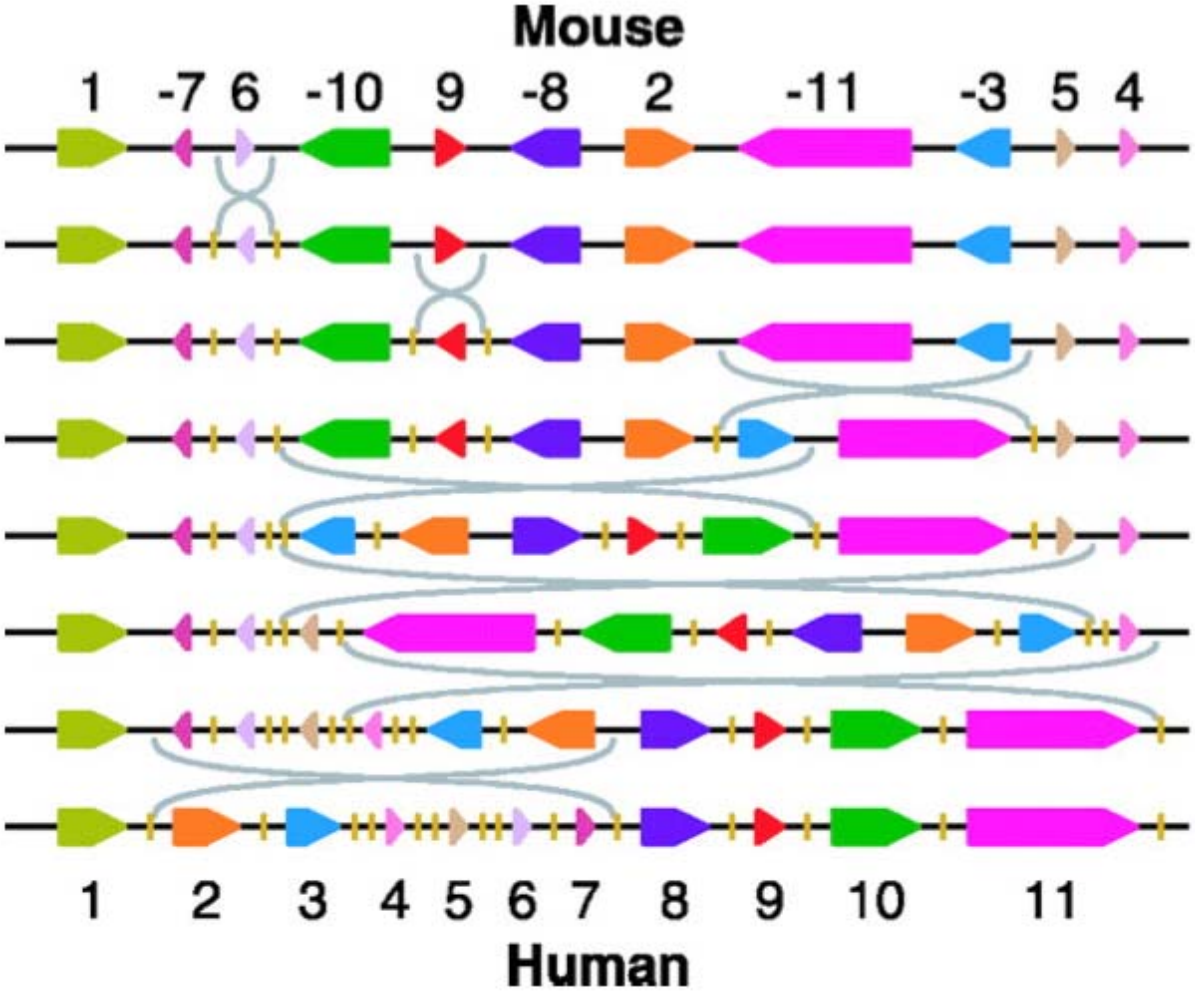


What is synthetic biology?





Genome Rearrangement

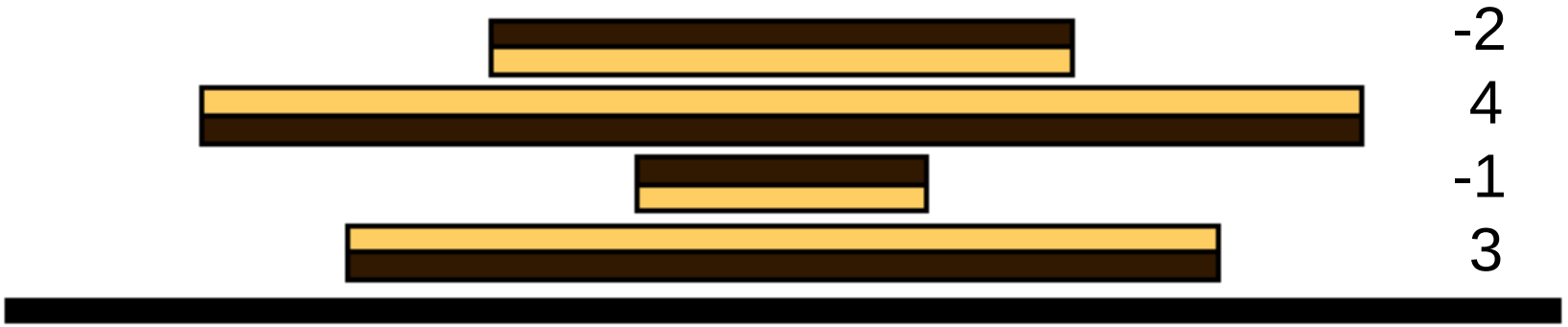


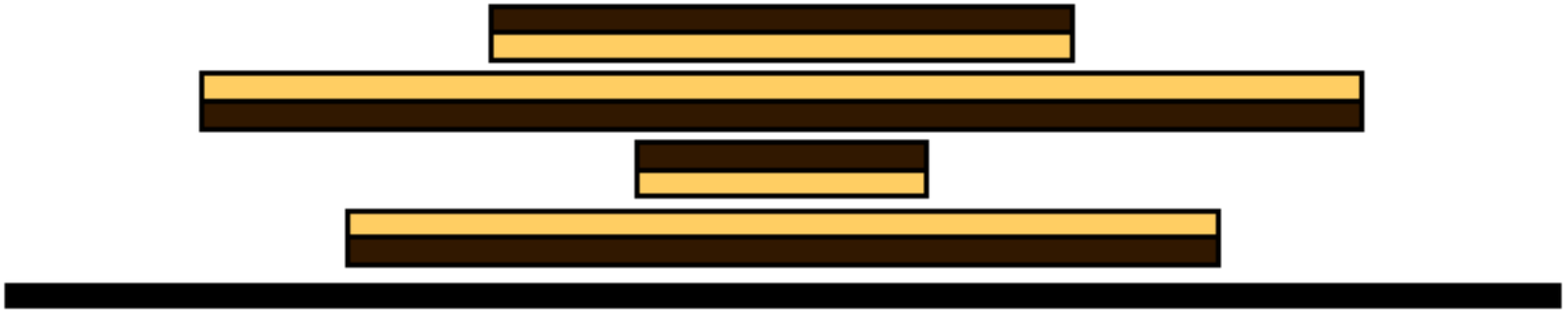
The (Original) Pancake Problem

- Given a stack of pancakes of varying sizes
- Grab some number from top of stack and invert
- Continue until stack is sorted by size (largest on bottom, smallest on top)
- What is the minimum number of flips?
- What is the worst case stack?

— Harry Dweighter, (a.k.a. Jacob Goodman), American Mathematical Monthly (1975

The Two-Spatula Burnt Pancake Problem





Movie Recap

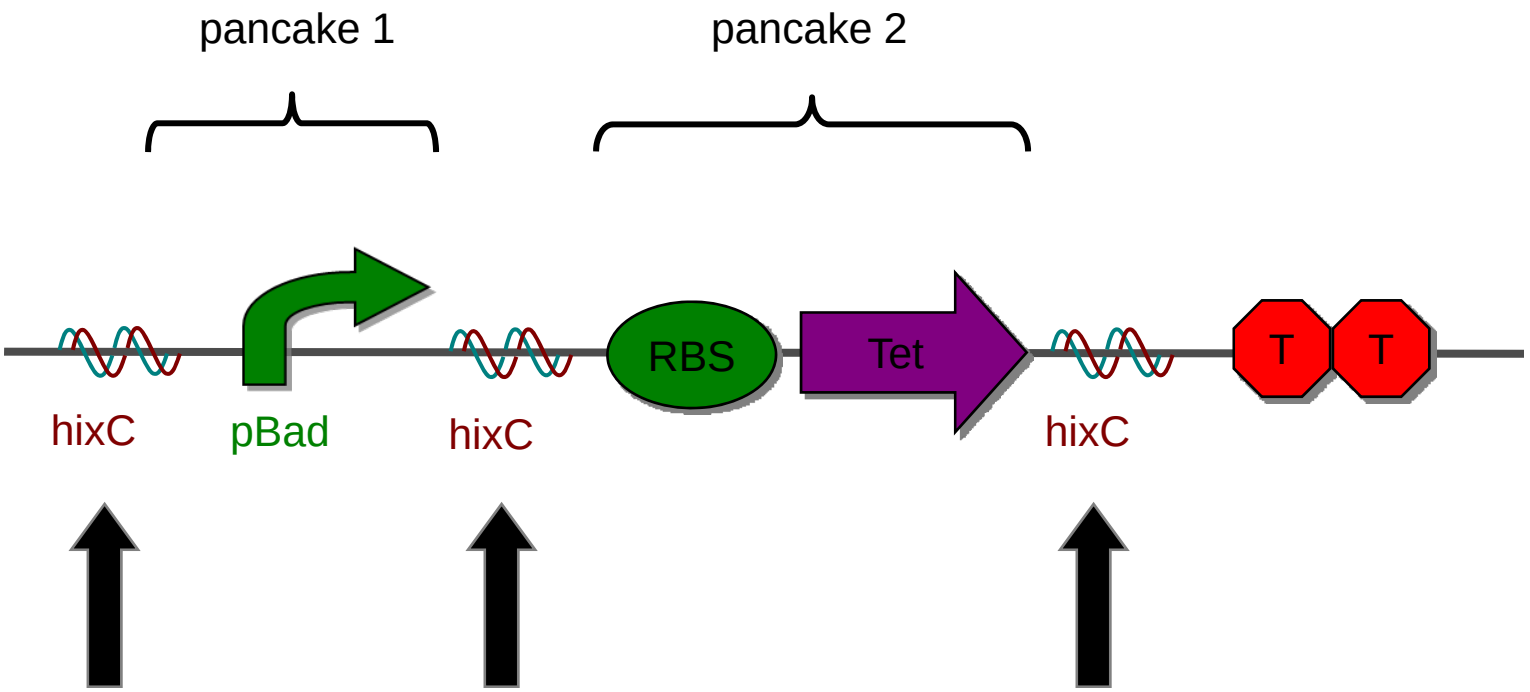
-2 4 -1 3

1 -4 2 3

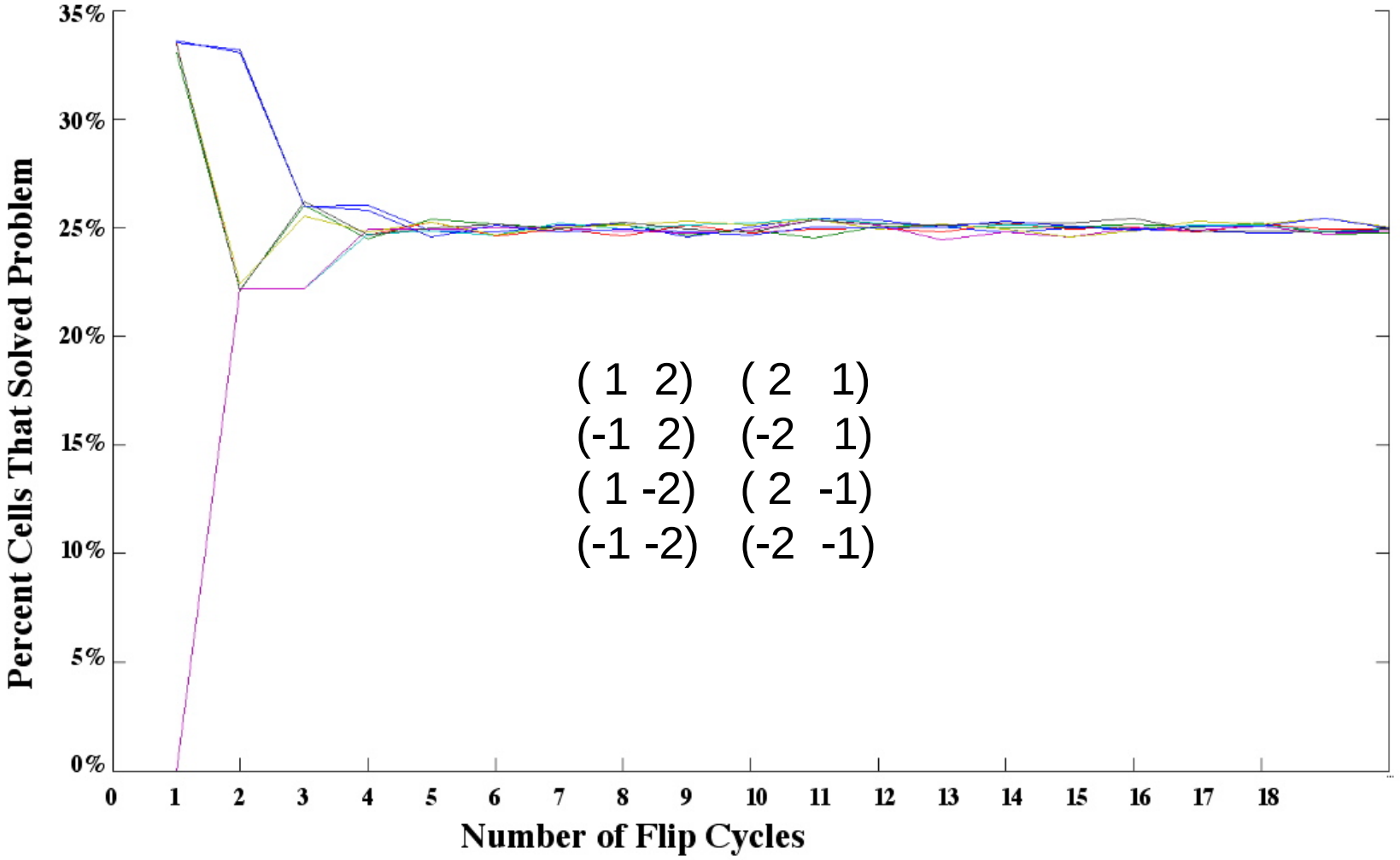
1 -3 -2 4

1 2 3 4

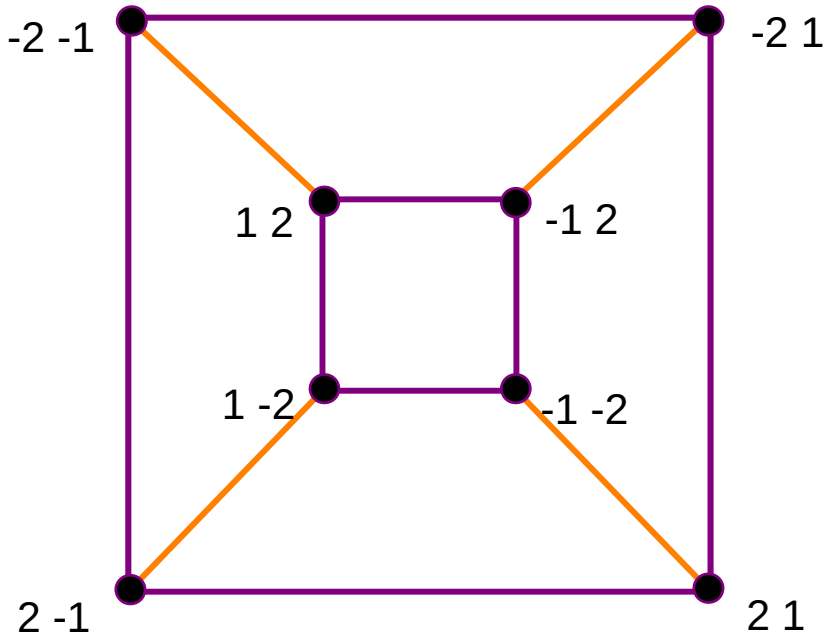
Building Biological Pancakes



Two Pancake Simulations



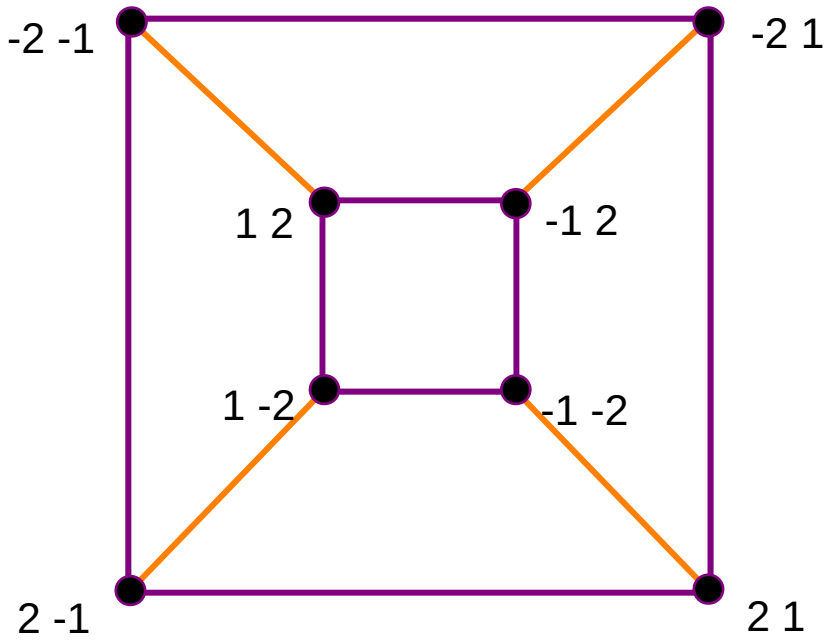
Adjacency Matrix



P_2: 2 Burnt Pancakes
8 vertices, each with degree 3

A	1 2	-1 2	-1 -2	1 -2	-2 -1	-2 1	2 1	2 -1
1 2	0	1	0	1	1	0	0	0
-1 2	1	0	1	0	0	1	0	0
-1 -2	0	1	0	1	0	0	1	0
1 -2	1	0	1	0	0	0	0	1
-2 -1	1	0	0	0	0	1	0	1
-2 1	0	1	0	0	1	0	1	0
2 1	0	0	1	0	0	1	0	1
2 -1	0	0	0	1	1	0	1	0

Paths of Length 3

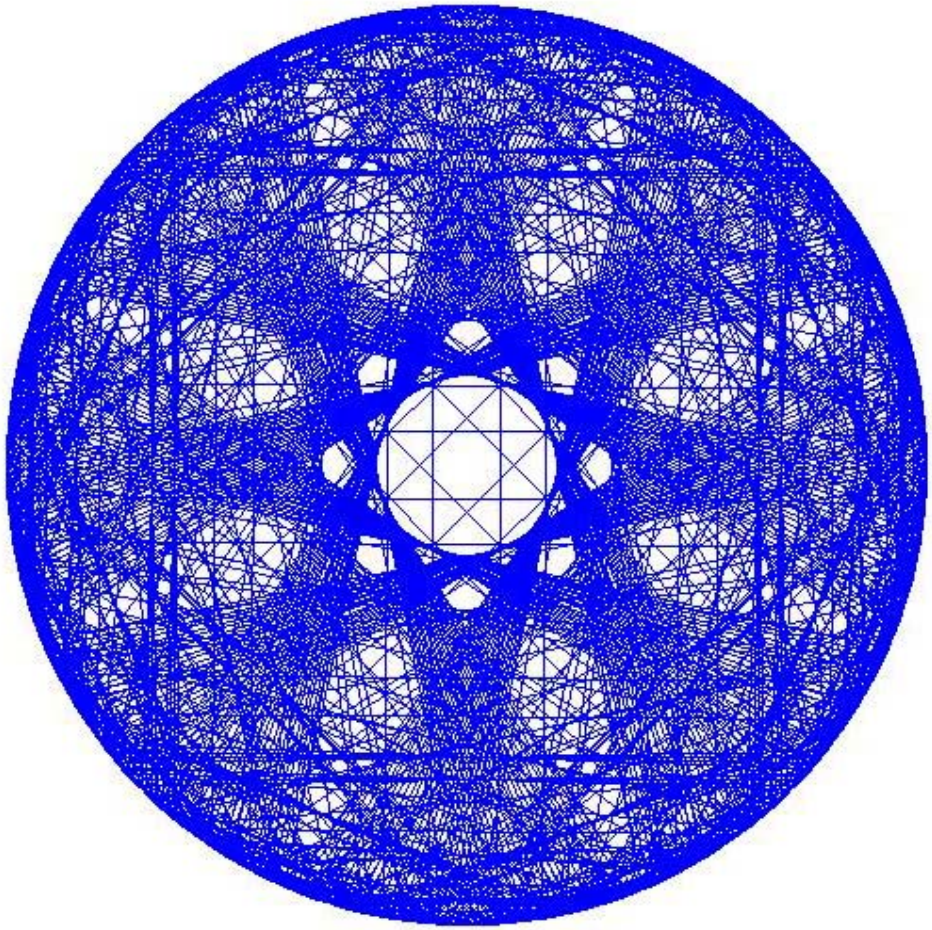


SR_2: 2 Burnt Pancakes
8 vertices, each with degree 3

Each path taken with probability 1/27

A^3	1 2	-1 2	-1 -2	1 -2	-2 -1	-2 1	2 1	2 -1
1 2	0	7	0	7	7	0	6	0
-1 2	7	0	7	0	0	7	0	6
-1 -2	0	7	0	7	6	0	7	0
1 -2	7	0	7	0	0	6	0	7
-2 -1	7	0	6	0	0	7	0	7
-2 1	0	7	0	6	7	0	7	0
2 1	6	0	7	0	0	7	0	7
2 -1	0	6	0	7	7	0	7	0

Bigger Stacks of Pancakes



Acknowledgements

- A. Malcolm Campbell
- GCAT
- Nicholas Dovidio, Michael Gordon, Gavin Taylor, Mac Cowell, David Moskowitz, Danielle Choi, Parul Karnik, Adam Abele
- Lance Harden, Erin Zwack, Sabriya Rosemond, Samantha Simpson, Karmella Haynes, Missouri Western iGEM team
- Phillip Compeau
- HHMI, NSF, Davidson College