SYNTHETIC BIOLOGY AND DATA-DRIVEN SYNTHETIC BIOLOGY FOR PERSONALIZED MEDICINE AND CLEAN ENERGY

Russell Hanson

OSCON Data July 25-27, 2011

Portland, OR

TheCureIsNow.org

Outline

- What is synthetic biology?
- How does synthetic biology relate to genomic science/personalized medicine
- “Open source” tools in science and bioengineering/synbio
- Bio-Computer aided design (AutoCAD, etc.)
- Mesoscopic Physics
- Diff Eqs and parameter optimizations

Outline, cont.

- IP Issues in personalized medicine
- Pink Army and co-op, open-source biopharma
- Non-profit clinical research organizations (CRO's)
- Disease-management companies, broken business model
- *Big Data* in electronic medical records, personal genomics, biostatistics, hospitals, community health centers

OSCON Data: This may not be a talk about a social web app and the database technologies used for that app, but truly, what would a social web, etc. app be without the life or biology on that social graph

What is life?

What is artificial life?

And, incidentally, what is consciousness?

Igor Aleksander (1995)

- Brain as state machine
- Inner neuron partitioning
- Conscious and unconscious states
- Perceptual learning and memory
- Prediction
- Self-Awareness
- Representation and meaning
- Learning utterances
- Learning language
- Will
- Instinct
- Emotion

Bernard Baars (1988)

- Definition and context setting
- Adaptation and learning
- Editing
- Flagging and debugging
- Recruiting and control
- Decision-making (executive function)
- Analogy forming-function
- Metacognitive and self-monitoring function
- Autoprogramming and self-maintenance function
- Definitional and context-setting function

from “When The Turing Test Is Not Enough,” by George Dvorsky

We are open source

Life is open source

We should have excellent tools to manage our source (code)

OpenWetWare.org

BioBricks Foundation

iGEM

Synthetic Biology
based on standard parts

Go Search

IGEM/Learn About

What is the iGEM Competition?

The International Genetically Engineered Machine competition (iGEM) is the premiere undergraduate Synthetic Biology competition. Student teams are given a kit of biological parts at the beginning of the summer from the Registry of Standard Biological Parts. Working at their own schools over the summer, they use these parts and new parts of their own design to build biological systems and operate them in living cells. This project design and competition format is an exceptionally motivating and effective teaching method.

iGEM Competition History

iGEM began in January of 2003 with a month-long course at MIT during their Independent Activities Period (IAP). The students designed biological systems to make cells blink. This design course grew to a summer competition with 5 teams in 2004, 13 teams in 2005 - the first year that the competition grew internationally - 32 teams in 2006, 54 teams in 2007, 84 teams in 2008, 112 teams in 2009, and 130 teams in 2010. Projects ranged from a rainbow of pigmented bacteria, to banana and wintergreen smelling bacteria, an arsenic biosensor, Bactoblood, and buoyant bacteria.

Previous competition websites:

iGEM 2011
iGEM 2010
iGEM 2009
iGEM 2008
iGEM 2007
iGEM 2006
iGEM 2005
iGEM 2004
iGEM Competitions Summary

Can simple biological systems be built from standard, interchangeable parts and operated in living cells? Or is biology so complicated that every case is unique?

NextGen sequencing

Table 1. Manufacturer's specifications for instrument configuration and production of single end sequences from a single flow cell

Platform	Method	Template prep	Starting DNA (μg)	Instrument configuration	Throughput statistic	Data per run (Gbp)	Reagent cost per run ($)[a]	Run time
454 GS-FLX	Pyrosequencing	Emulsion PCR	3–5	Single picotiter plate, partitionable into 8 lanes	238-bp read[b]	0.1	8500	7.5h
Illumina 1G	Four-color SBS with reversible terminators[c]	Bridge PCR	0.1–1	Single flow cell, partitionable into 8 lanes	35-bp read	1.3	3000	3 d
ABI SOLiD	Oligonucleotide ligation with two-base, four-color encoding	Emulsion PCR	0.1–20	Independently controlled dual-flow cells, each partitionable into 8 lanes	35-bp reads, mapped to reference sequence allowing up to three mismatches	4	3400	7 d
Helicos HeliScope	Single-color SBS with virtual terminators	Not applicable	Not available	Single 25-lane flow cell	30-bp read	7.5	18,000	14 d

[a]Reagent costs are list prices.
[b]Average read length for a typical whole-genome library, using long read kit.
[c](SBS) Sequencing by synthesis.

Holt A. et al., Genome Res. 2008

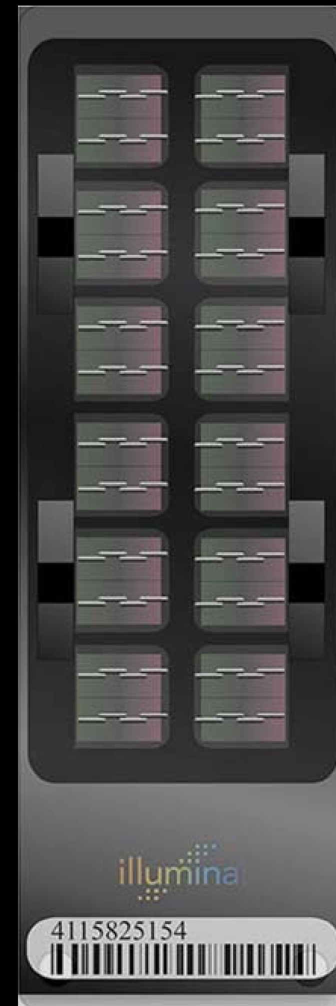

454


Solexa


SOLiD


HeliScope

SNP Chips



Affymetrix 500K probe chip ~$450USD

Single nucleotide polymorphism (SNP) and gene expression arrays

- Copy number profile
- Gene expression/over expression levels in a tissue
- Major copy proportion
- Machine learning on SNPs, SVM, Bayesian networks, Markov chains
- Biomarker identification, verification, and development. Biomarker for autism, “lack of empathy”

Where are we?

- How much do we know about cellular processes, regulation, oncogenes, protein-protein interactions, microbiome interactions?
- How much do we know about brain structure-function relationships, gene activation in neurons, memory, eletrophysiology, neuronal cognition, brainmaps from *C. elegans,* to *Mus musculus*, to *Homo sapiens*?

How to Access the Human Genome (and other sequenced genomes)

- ftp://ftp.ncbi.nih.gov

Index of ftp://ftp.ncbi.nih.gov/genbank/genomes

Up to higher level directory

A thaliana	10/17/2003	0:00:00
Anopheles gambiae	5/7/2002	0:00:00
Bacteria	4/7/2004	18:28:00
C elegans	6/14/2002	0:00:00
D melanogaster	10/19/2000	0:00:00
H sapiens	4/15/2004	0:23:00
Leptospira interrogans serovar Copenhageni	3/22/2004	17:40:00
MITOCHONDRIA	11/2/1999	0:00:00
M musculus	5/12/2002	0:00:00
P falciparum	5/11/1999	0:00:00
Plasmodium falciparum	10/11/2002	0:00:00
README		
README OLD		
R norvegicus		
S cerevisiae		

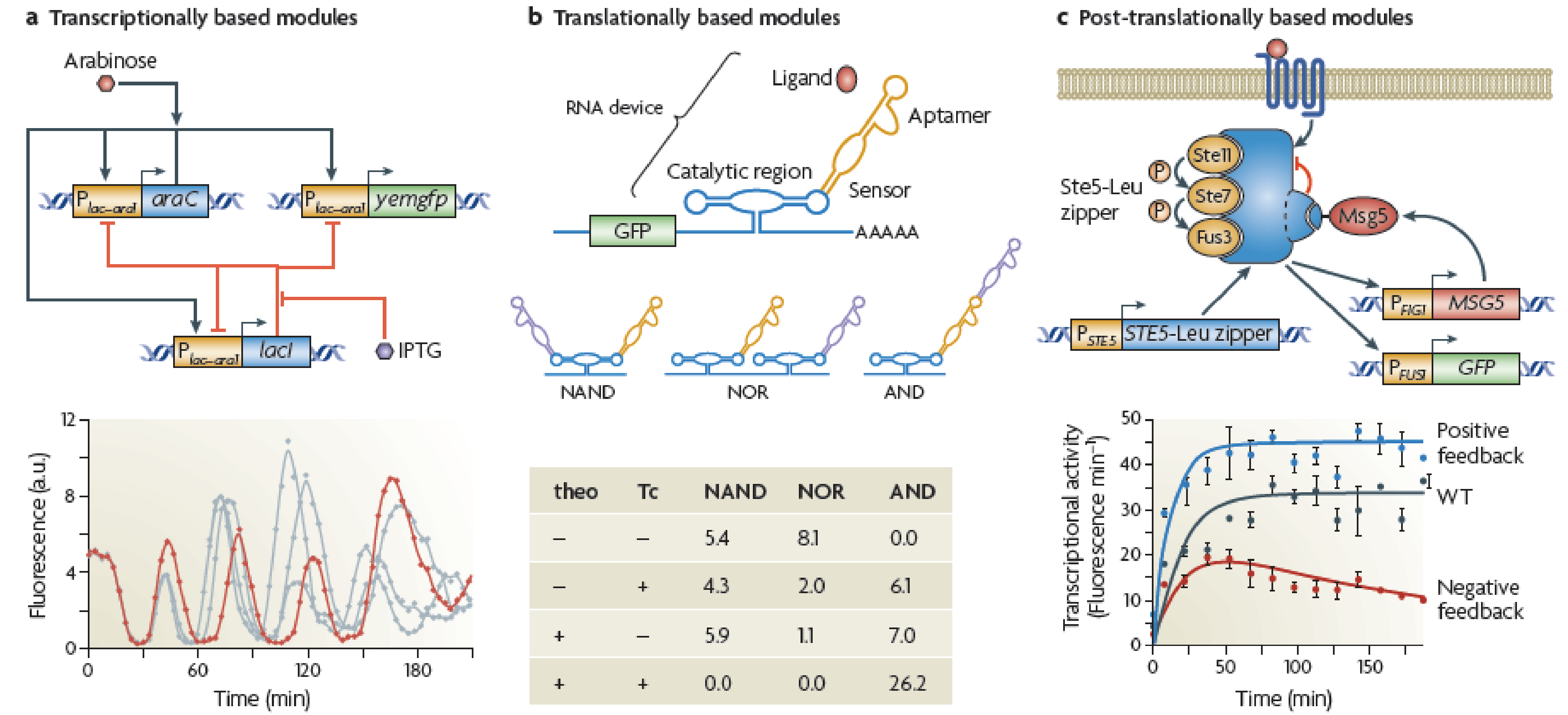
```
hs_phs0.fna.gz Survey sequence (appr
hs_phs1.fna.gz Unordered contigs (ea
hs_phs2.fna.gz Ordered contigs (each
hs_phs3.fna.gz Finished sequence
```

Index of ftp://ftp.ncbi.nih.gov/genbank/genomes/H_sapiens

Up to higher level directory

README	2 KB	10/17/2003	0:00:00
hs phase0.fna.gz	94408 KB	4/14/2004	21:36:00
hs phase1.fna.gz	606370 KB	4/14/2004	22:35:00
hs phase2.fna.gz	48487 KB	4/14/2004	22:39:00
hs phase3.fna.gz	1090520 KB	4/15/2004	0:23:00

“The second wave of synthetic biology: from modules to systems”



theo	Tc	NAND	NOR	AND
–	–	5.4	8.1	0.0
–	+	4.3	2.0	6.1
+	–	5.9	1.1	7.0
+	+	0.0	0.0	26.2

Figure 1 | **Modules based on transcriptional, translational and post-translational control. a** | The dual-feedback

“The second wave of synthetic biology: from modules to systems,” Purnick PEM, Weiss R, Nature Reviews Molecular Cell Biology, Vol 10 Iss 6 P410-422 Jun 2009

Using increasing abstraction, performance characteristics, datasheets -- like integrated circuits (IC's)

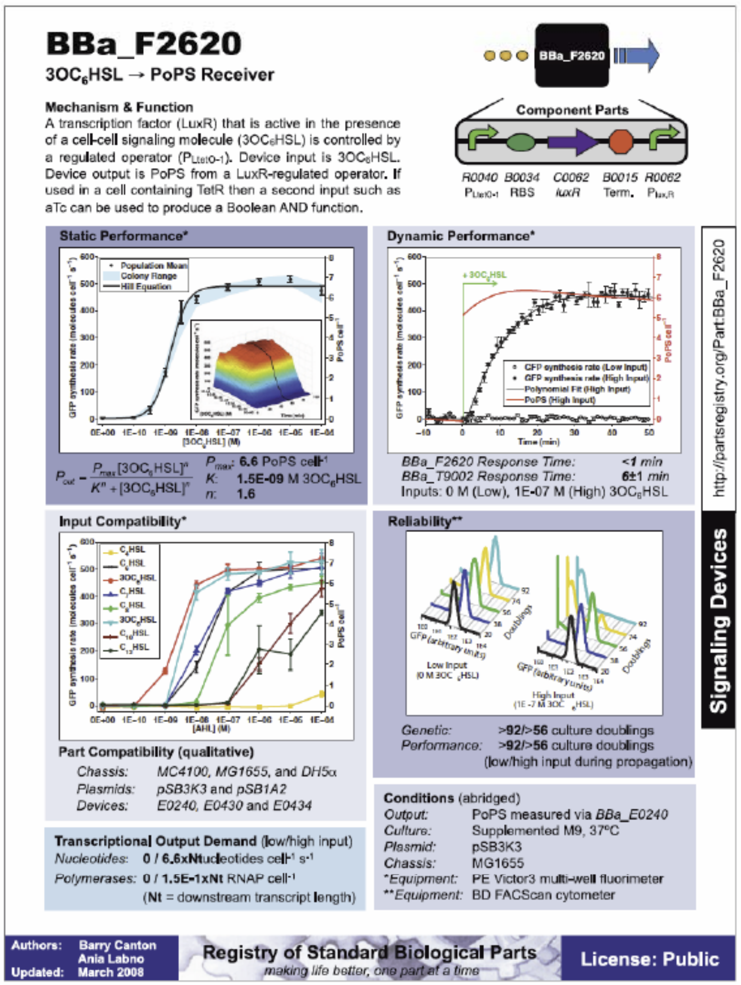
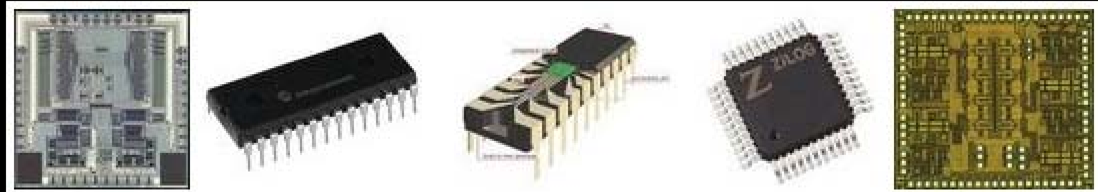


Figure 3 A prototypical 'datasheet' that summarizes current knowledge of the behavior of the receiver BBa_F2620. The datasheet, which includes a general description and a summary of relevant performance characteristics, is designed to support rapid reuse of the device. The description of the receiver is also available in electronic format[23]. A glossary for the datasheet is provided in Box 3.



"… represent quantitative characteristics of biological devices in the form of a **datasheet** and demonstrate it on a device composed of BioBrick parts. (PartsRegistry.org)"

(Canton, Labno, Endy 2008 Refinement and standardization of synthetic biological parts and devices.)

Graphical layout, design and ‘programming’, *CellDesigner*

tutorial at: http://celldesigner.org/~funa/CellDesignerTutorial2007.pdf

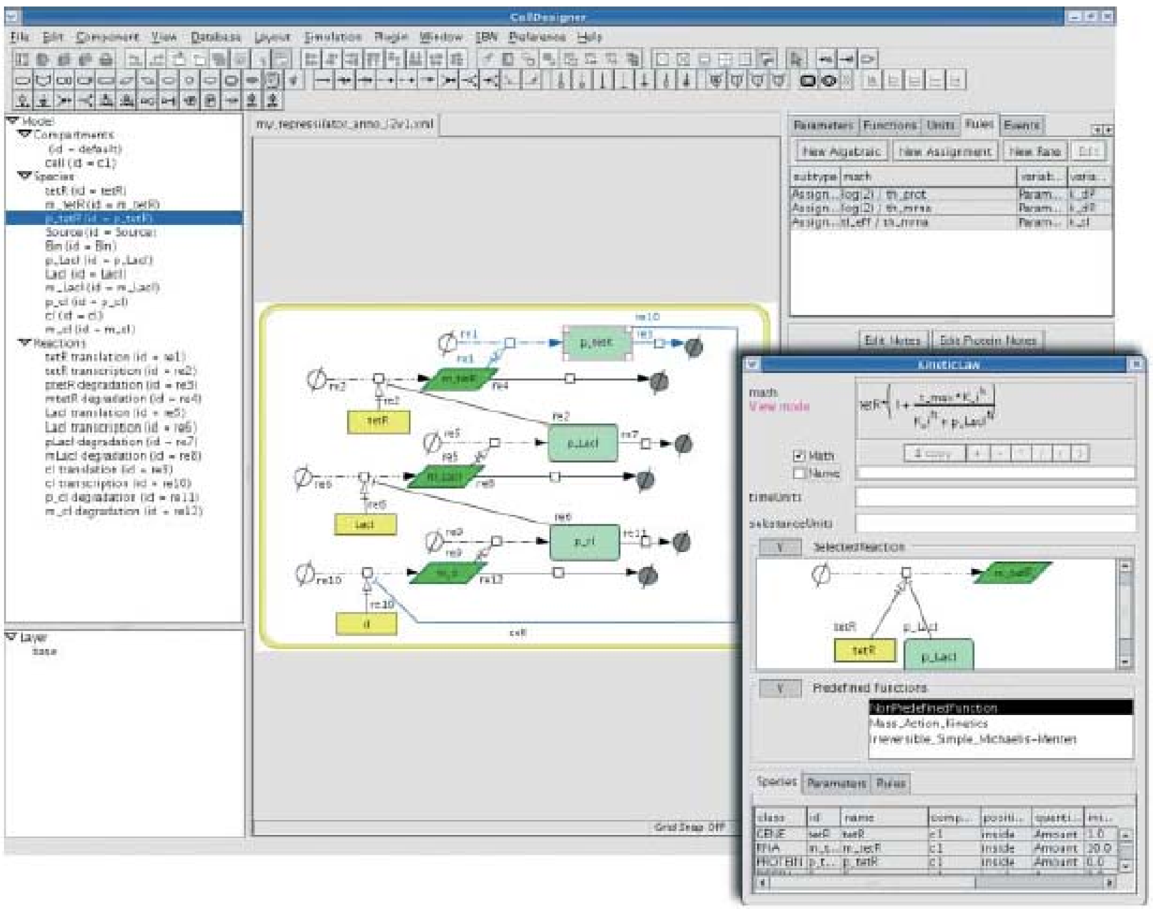
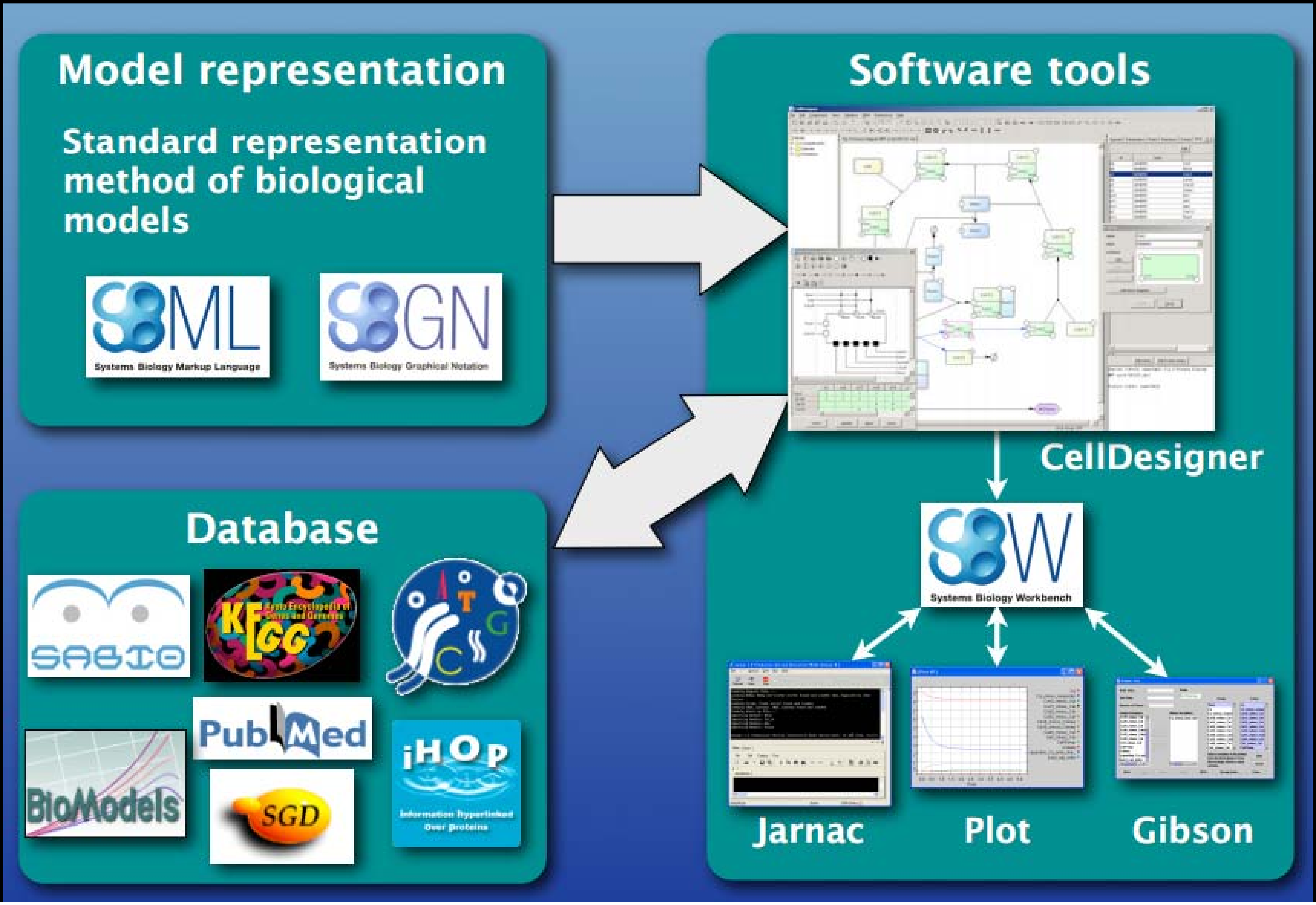


Figure 3. The CELLDESIGNER interface provides an ample selection of graphical elements and editing options to design biochemical and mathematical models. The repressilator is represented in CELLDESIGNER's graphical notation, a derivative of SBGN. The raised window is the editing form for the kinetic law of a single reaction.

Graphical layout, design and 'programming', *CellDesigner 2*



Nanobots to treat NK leukemia cells

DNA scaffolds and
cell-surface receptors

Software, etc. for DNA
origami, scaffold design

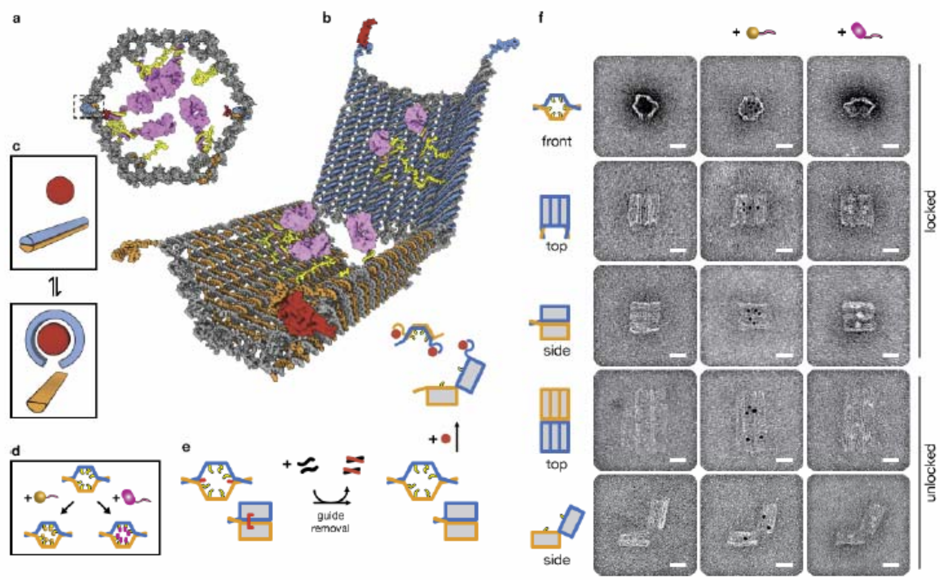


Figure 1 — Design and TEM analysis of an aptamer-gated DNA nanorobot. The device transitions from its closed state (a) to open (b) when aptamer-based locks are displaced by binding to an antigen key (c). Payloads such as gold nanoparticles and antibody fragments (d) can be loaded. Removable guide staples (e) aid in initial folding to high yield. Electron micrographs show the nanorobot with different payloads and conformations. Scale bars 20 nm.

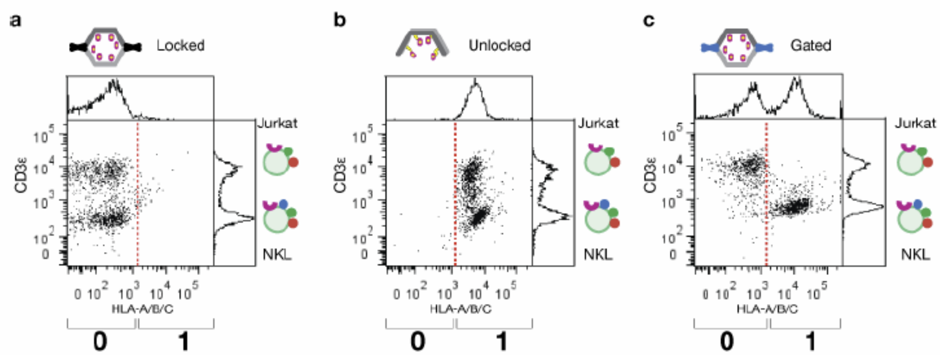


Figure 2 — Flow-cytometric analysis of nanorobot selectivity in a complex mixture. (a) NKL and Jurkat cells were mixed, labeled with FITC-anti-human CD3ε and incubated with robots loaded with APC-anti-human HLA A/B/C Fab' for 30 min. Locked robots remain inactive. (b) Unlocked robots react with both cell populations. (c), Gated robots react only with the cell population expressing the proper key.

23andMe, NextGen sequencing and SynBio get together

- 23andMe version 3 uses Illumina HumanOmniExpress BeadChip SNP chips ($250/sample)
- Genome Wide Association Studies
- --> At $250/sample, genomic personalization is very realizable

HumanOmniExpress BeadChip

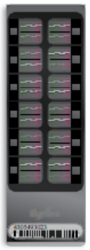
Process more samples to discover novel disease and trait associations, with optimized tag SNP content and industry-leading pricing.

OVERVIEW

The HumanOmniExpress (OmniExpress) BeadChip delivers superior power for genome-wide association studies (GWAS), providing high sample throughput and comprehensive genomic content at the industry's best price. Using the proven iScan System, this twelve-sample BeadChip offers unrivaled throughput of thousands of samples per week—the ideal solution for processing the greatest number of samples within a given budget. Optimized tag SNP content from all three HapMap phases has been strategically selected to capture the greatest amount of variation and drive the discovery of novel associations with traits and diseases.

The OmniExpress BeadChip includes convenient kit packaging, a streamlined PCR-free protocol, and integrated analysis software to provide a comprehensive DNA analysis solution. With the highest data quality and best content, including full support of copy number variation (CNV) applications, this powerful genotyping tool allows you to make more meaningful discoveries and publish faster.

FIGURE 1: TWELVE-SAMPLE OMNIEXPRESS BEADCHIP



The OmniExpress BeadChip's twelve-sample format supports rapid, cost-effective studies.

PRELIMINARY PRODUCT INFORMATION

PARAMETERS*	OMNIEXPRESS BEADCHIP
SNP loci per sample	> 700,000
Samples per BeadChip	12
DNA input	200 ng per sample
Instrument support	iScan
Array processing throughput†	> 1,400 samples per week
Assay	Infinium® HD
Publications using the assay	> 300+
Content source	HapMap Phases 1, 2, and 3
Genomic coverage*	0.91 (CEU) / 0.91 (JPT+CHB) / 0.68 (YRI) at $r^2 > 0.8$
Marker spacing for CNV discovery	~4.0 kb mean / ~2.0 kb median
Software support	Full support for genotype calls and CNV analysis within GenomeStudio®
Data quality	• Call frequency: > 99% average • Reproducibility: > 99.9%
Estimated scan time	5 minutes per sample on the iScan system
List price ‡	$250 per sample (USD)

* Estimated before final Gentrain; final numbers will vary.
† Estimate assumes one iScan system, one AutoLoader2, one Tecan Robot, and a five-day work week.
‡ Represents pricing for North America—contact your local sales representative for regional pricing.

The OmniExpress BeadChip is also the basis of the 2010 GWAS product roadmap, which will provide researches favorable pricing and future access to five million variants per sample, including broad coverage of rare variants identified by the 1000 Genomes Project.

ADDITIONAL INFORMATION

For additional information on the Omni family of microarrays, visit www.illumina.com/gwas.

Illumina, Inc.
Customer Solutions
9885 Towne Centre Drive
San Diego, CA 92121 USA
1.800.809.4566 toll-free
1.858.202.4566 tel
techsupport@illumina.com
www.illumina.com

More new CAD/BioCAD software tools

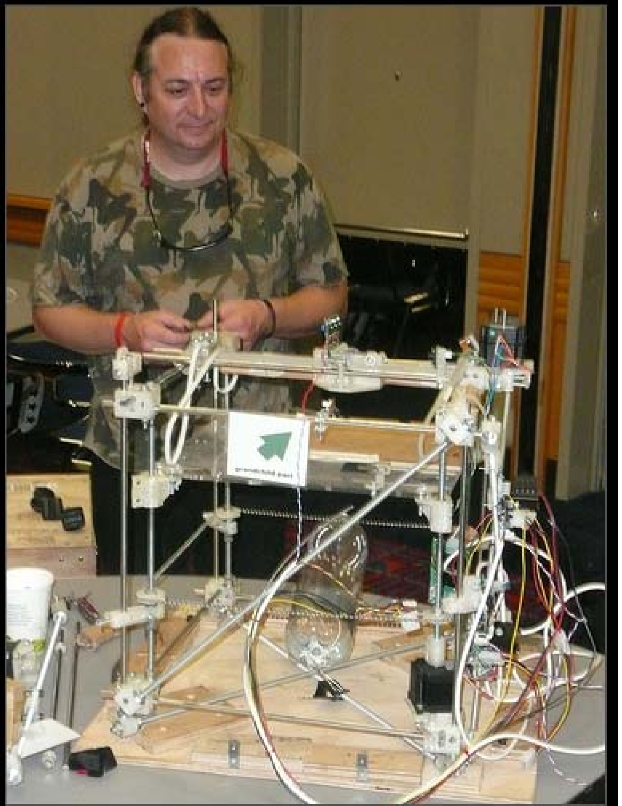
- ‘Matter compiler’
- Thingiverse, cheap 3d printers, laser cutters, Shapeways, personal fab technology, makerbot, etc.
- AutoDesk, EugeneCAD,

ClothoCAD

- RepRap – exponentially

scalable matter compiler

/extruding 3D printer



Project 365, Day 203: RepRap Matter Compiler at OScon08 by FallenPegasus

- Movie at: http://www.molecularmovies.com/movies/berry_apoptosis.html

Biological “simulators”

- NAMD – and other molecular dynamics, etc.
- Mixed Quantum Mechanics/Molecular Mechanics (QM/MM/...)
- Biological pathway/chemical reaction simulators/optimizers
- Many projects on gene network analysis, stochastic simulation, etc.

Design of bio/chemical devices from (1) first principles and functional abstractions and (2) from data

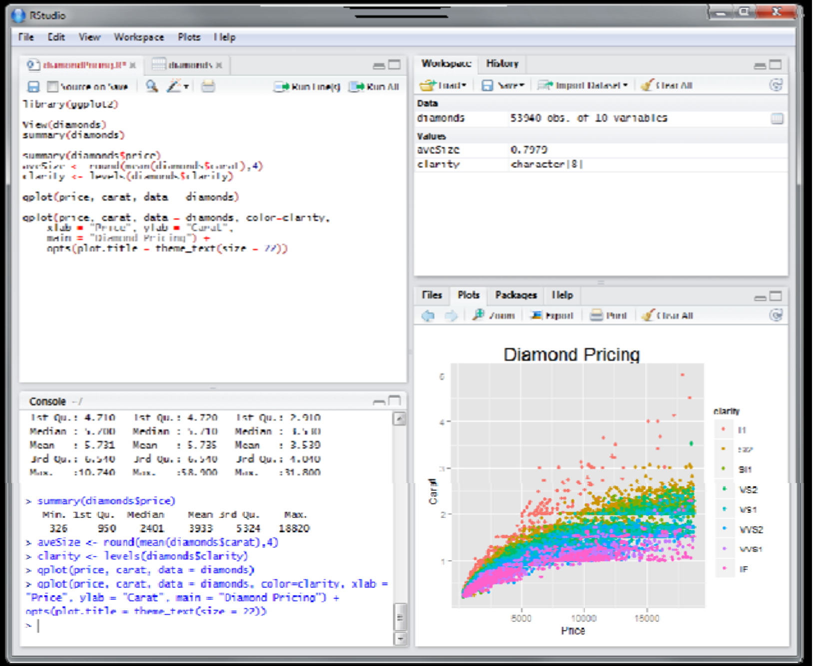
Data

Bio/chem/phys first principles and functional abstractions

How and where these data are used, or made usable

- Randomized clinical trials
- Biomarker development, genomic tests in CLIA certified labs
- Licensing to other companies for marketing and development

Tools that are used to analyze these data



- R (!), BioConductor
- Biostatistics, Matlab
- Python
- Chi-square tests,

permutation tests, Monte Carlo resampling

- Parametric statistics, nonparametric statistics, generative statistics
- Statistical learning theory, online learning

Approval of processes for cellular therapies/treatments

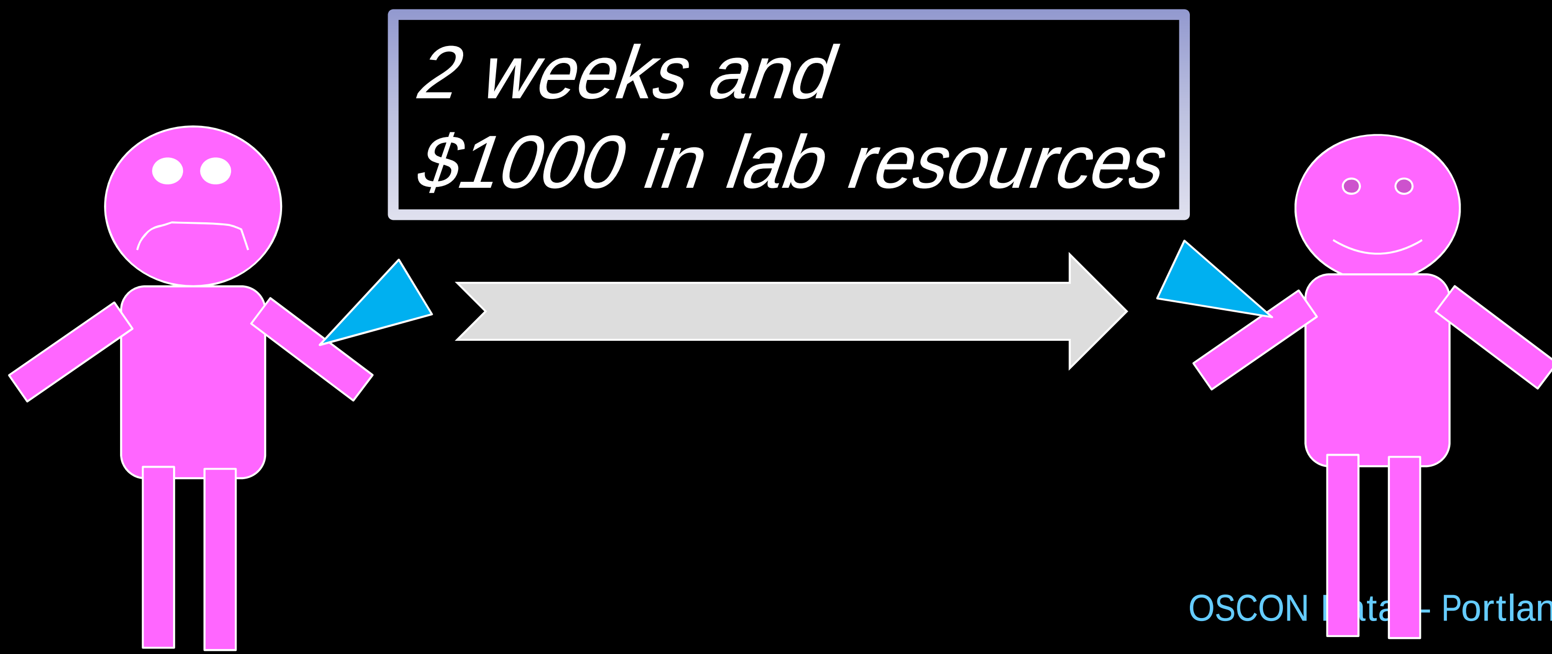
- FDA approval for process, not for drug/biologic/device

"…progress on the development of a new experimental MCL treatment called immunotherapy, in which a patient's own immune cells are collected from the bloodstream before ASCT, treated in the laboratory to have antitumor activity, and reintroduced into the body after ASCT. The goal of this immunotherapy is to reduce the likelihood of relapse after transplantation. A major challenge of immunotherapy is getting the transferred immune cells to persist in the body after the transfer. In preclinical studies, Dr. Jensen and his colleagues found that using a certain subset of immune cells called central memory T-cells increased the likelihood that the transferred T-cells will persist."

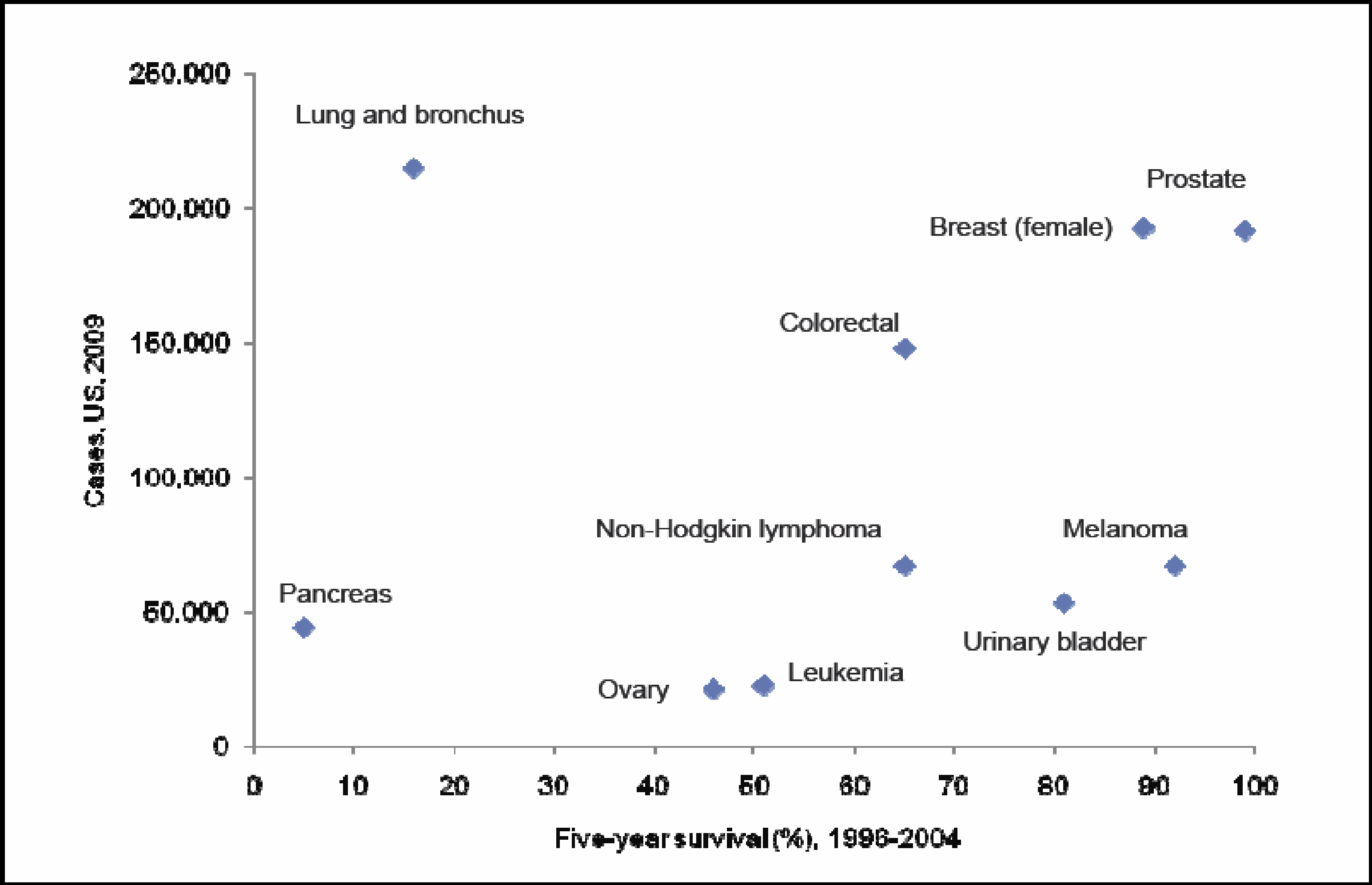
Report from the Lymphoma Research Foundation, Spring 2010, Volume 8, Number 1

Pink Army – a biopharma co-op

- You own your own cells, therapies based off your own cells should be *yours* in the co-op.
- If a biopharma company cures your disease, *they lose a customer*, who is cured.



Cancer 5-year survival rates, US



Division of Cancer Control and Population Sciences, 2008

Non-profit clinical research organizations (CRO's)

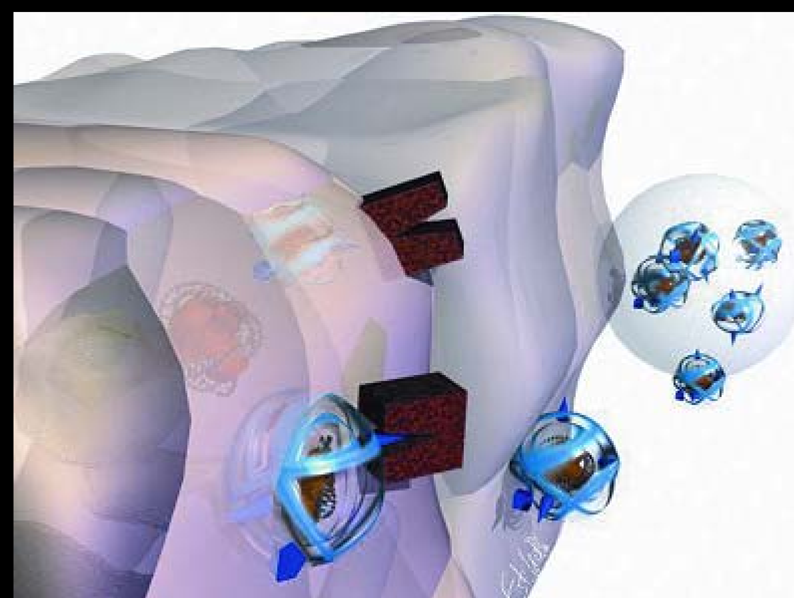
- Randomized trials
- A melanoma trial with 1,000 patients costs $60M in CRO expenses
- A 10,000 patient chronic obstructive pulmonary disease (COPD) costs $240-$250M in CRO expenses

More medical foci of SynBio

- Oncolytic viruses
- RNAi for HER2 in breast cancer (also an iGEM team project)
- Designed, logic, and ‘programmed’
- Not *computational devices* but use logic and/or procedures

Using data to engineer therapies or cures for genetic diseases

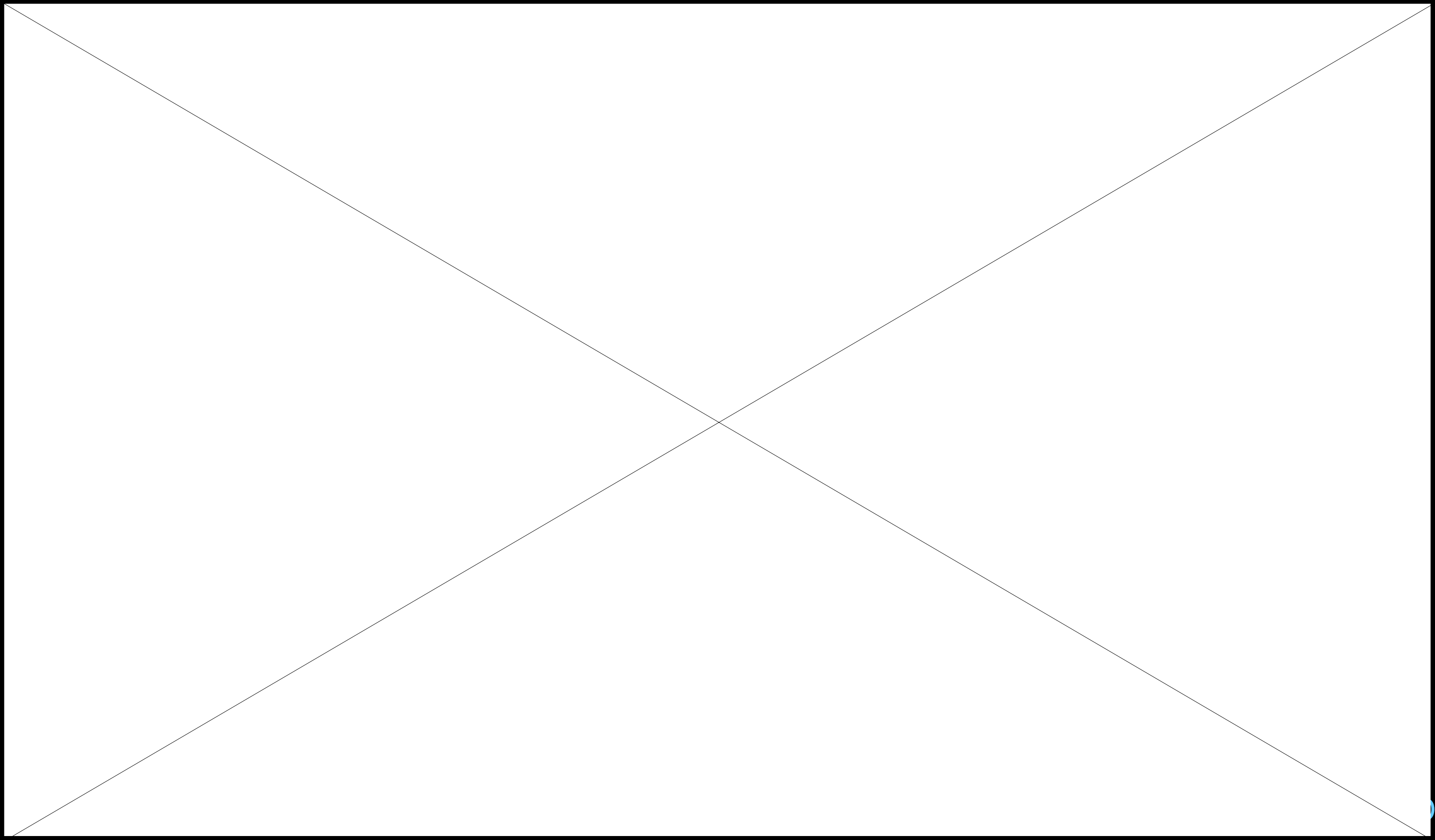
```
IF ( RNA AND RNA AND RNA AND RNA AND RNA )

THEN
   RELEASE PROTEIN INITIATING APOPTOSIS IN THE TUMOR CELL
```



More biological nanobots

- Tetrahedral robots, made from biological or organic substrate: 12Tet Tetrahedral Rover Robot

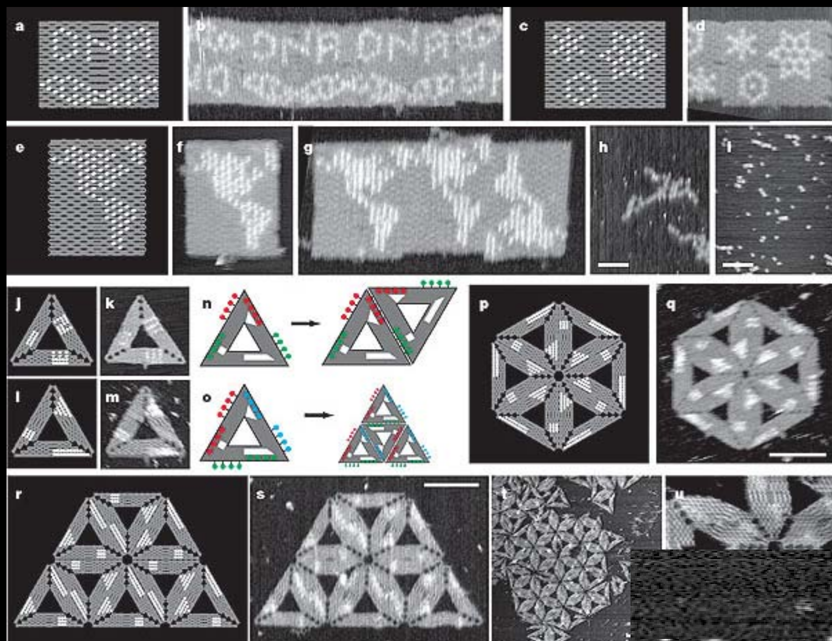
  http://www.youtube.com/watch?v=txZMLS7YD6Q

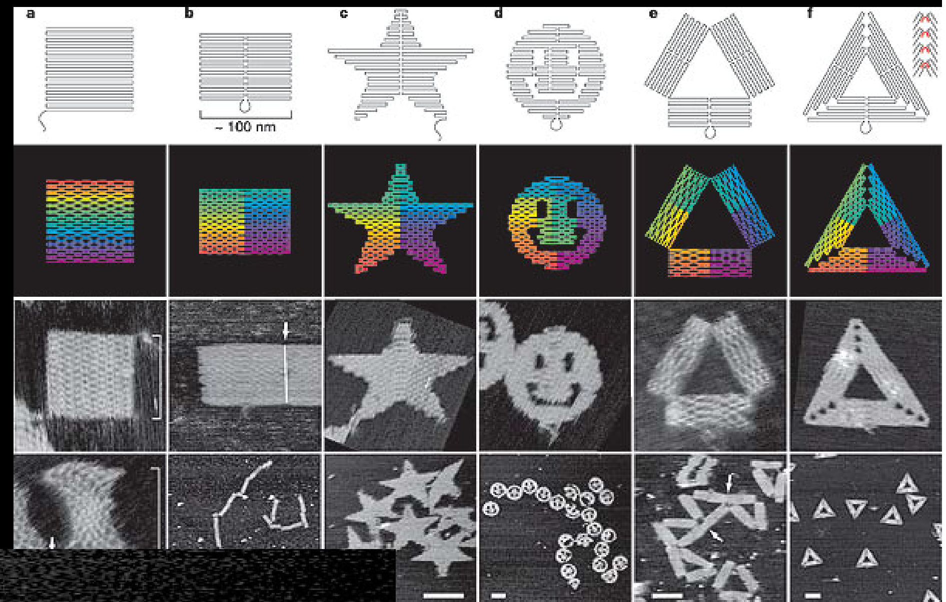


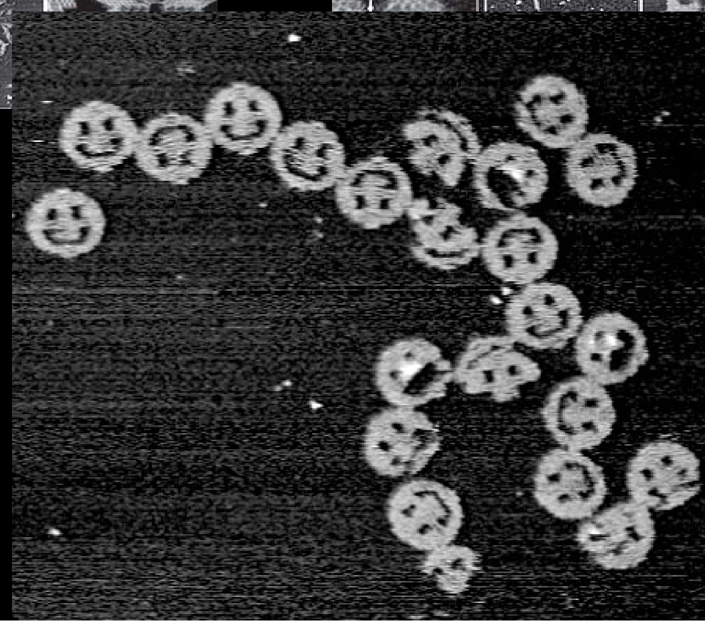
OR

More biological nanobots, 2

DNA Origami







CON Data – Portland, OR

Energy, biofules, and bioreactors

‘The oil multinational, despised by green activists for its support of scientists skeptical of climate change, plans to invest $600 million (£370 million) in a joint venture with Mr Venter’s company, Synthetic Genomics. The thinking is that algae is a more efficient source of fuel than conventional biofuels, such as ethanol made from corn or sugar cane and biodiesel made from wheat or palm oil. Algae can be processed into fuels similar to petrol and diesel — and it consumes carbon dioxide as it grows.

According to Exxon, the yield of biofuel from algae is 2,000 gallons per acre, more than three times that of biodiesel from palm oil and eight times the ethanol yield from corn. It is also believed to have the edge over other biofuels because of its suitability for use as a jet fuel. “The real challenge to creating a viable, next-generation biofuel is the ability to produce it in large volumes,” Mr Venter said.’

via http://www.timesonline.co.uk/tol/news/environment/article6710846.ece

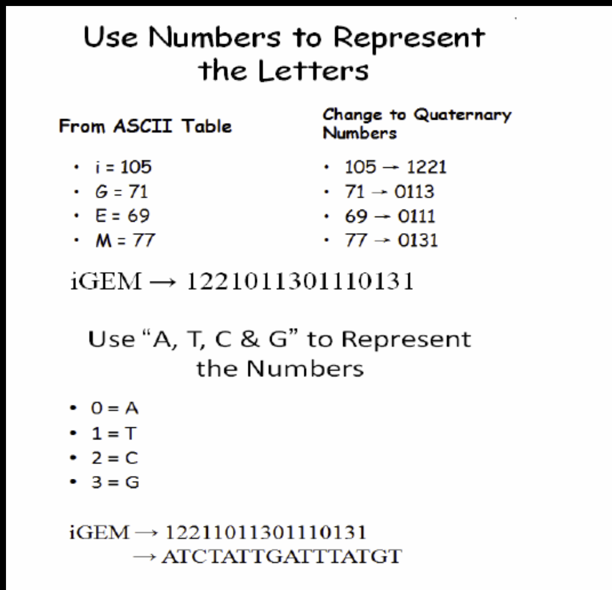
Biological light sources

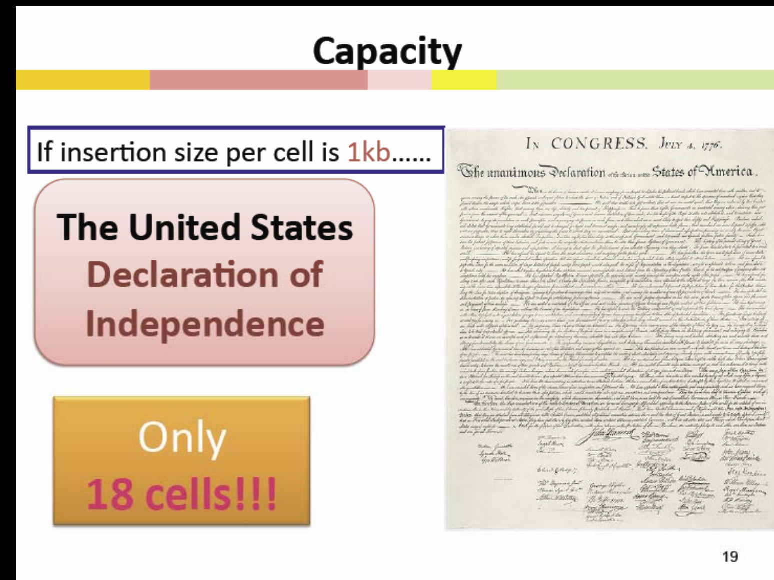
E.glowli: a bioluminescent future

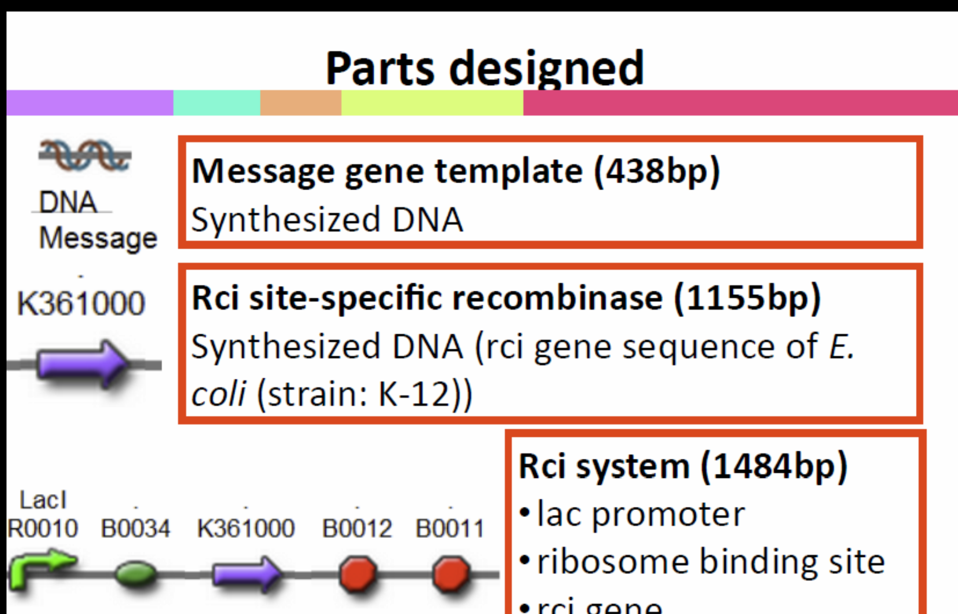
U Cambridge, iGEM 2010 team

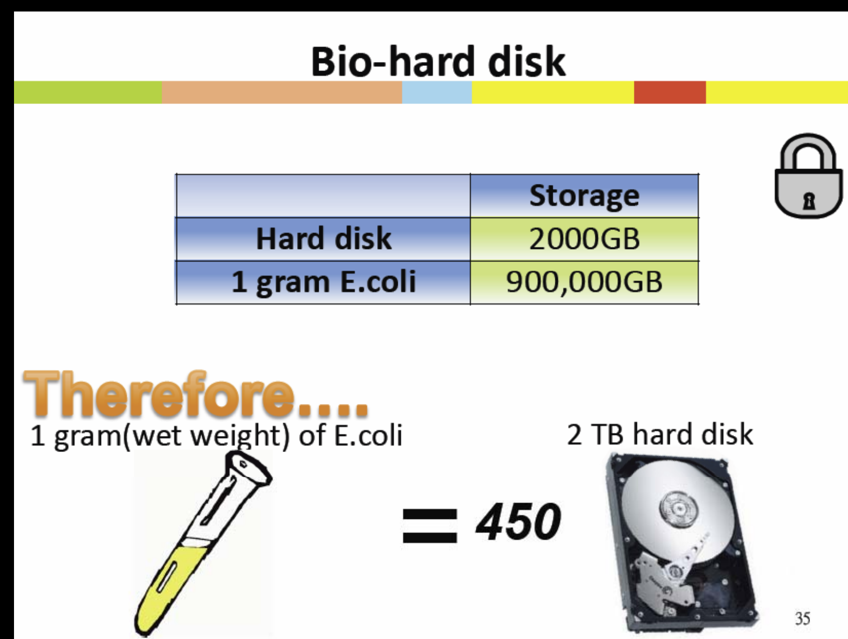
Biological memory

CUHK, iGEM 2010 team





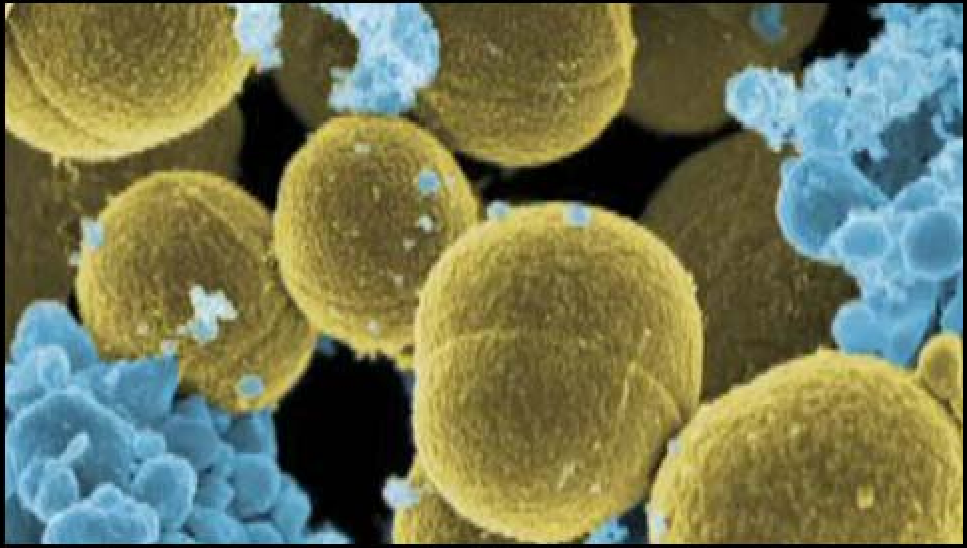




BIG BIOLOGICAL DATA!!

Microbiome Engineering

- Use the microbiome to produce therapies for existing diseases in controlled, continuous ways: insulin for diabetics, interleukin 10 for Crohn’s disease, etc.
- Highly competitive environment: microbe needs orthogonal edge to survive



And so forth... and biology

- PolySilicon and biology
- SynBio in chemicals industry – cleaning, refinement, scalability, sensors, etc.

Conclusions

- Wide ranging implications to existing problems in health, engineering, energy, manufacture
- Particularly for human healthcare, genomic research on diseases, data-aided design and testing
- New manufacture processes for new biologically-designed constructs

http://www.linkedin.com/in/russellhanson
http://twitter.com/russell_hanson
http://www.russellhanson.com/