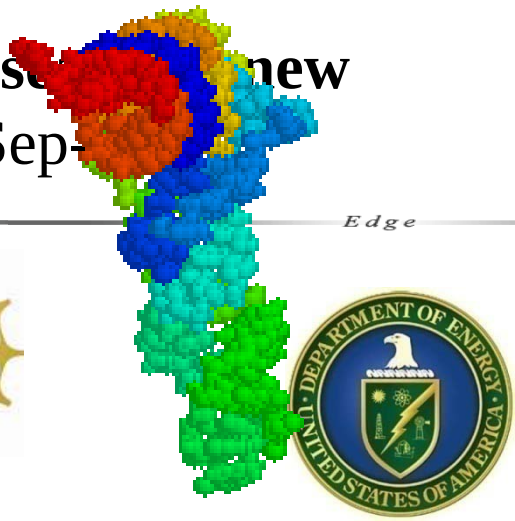
Applying computational & synthetic biology to dis[illegible] new materials IEEE, BE-BMES, GBC/ACM 7 PM 16-Sep-



Edge Edge Edge

Thanks to:

.gov
||
||
.edu
||
||
.org
||
||
.com
||
||

National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov

National Heart Lung and Blood Institute

NSF

genome.gov
National Human Genome Research Institute
National Institutes of Health

DARPA

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

BEIJING INSTITUTE OF GENOMICS
CHINESE ACADEMY OF SCIENCES

VERITAS

MIT

BROAD INSTITUTE

HSCI
HARVARD STEM CELL INSTITUTE

WYSS INSTITUTE

ArmRev.org

PERSONALGENOMES.ORG

Oppenheimer Foundation

LSRF

BILL & MELINDA GATES foundation

PBS

BBC

The Ellison Medical Foundation

molecular systems biology
A new type of online journal from EMBO and Nature Publishing Group

Google

ISILON SYSTEMS

Knome

CORIELL INSTITUTE for Medical Research

LS9, INC.
the renewable petroleum company™

23andMe

Helicos BioSciences Corporation

Azco

JOULE

AFFYMETRIX

Counsyl

illumina

enzymatics

Agilent Technologies

Gen9

RBH

Roche

Complete Genomics

DANAHER MOTION

QTEROS
Biofuels from advanced microbiology

OrbiMed
Healthcare Fund Management

Read = = = = = = = = I/O = = = = = = = Write

Reading & Writing Genomes Goals

- **2nd Generation BI/O: Reading & Writing**
  Engineer cancer- & virus-resistant genomes

- **Personal Genomes – Integration tasks**
  - Personal **G**enomes: **E**nvironments, **T**raits,
  - Stem cells, Microbiome/Immunome
  - Synthesis for Causality (CEGS)

How &why genome engineering for chemicals?

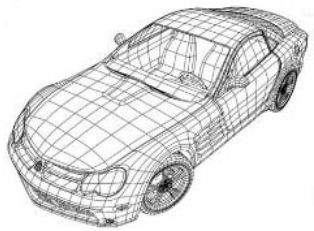
Engineering	# genes	Scale	Application
Genetic	1-2	Plasmid	Protein drugs
Genome	1-2	Chromosome	Mutant models
Metabolic	2-30	Pathway	Chemical production
Code	300-all	Genome	Multi-virus resistance
			Safety, new AA

WYSS INSTITUTE

Engineering Life

Interoperable parts

Hierarchical designs



CAD

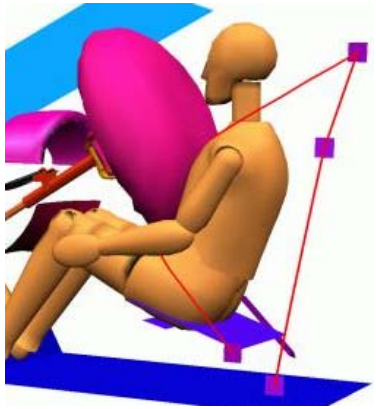
Cost effectiveness

Standards

Isolation

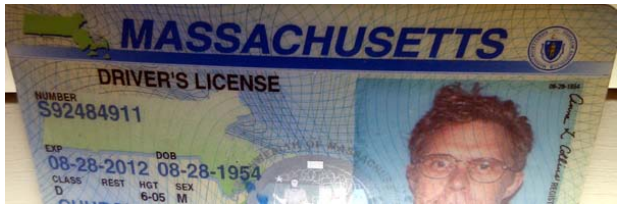
Testing

Redundant systems



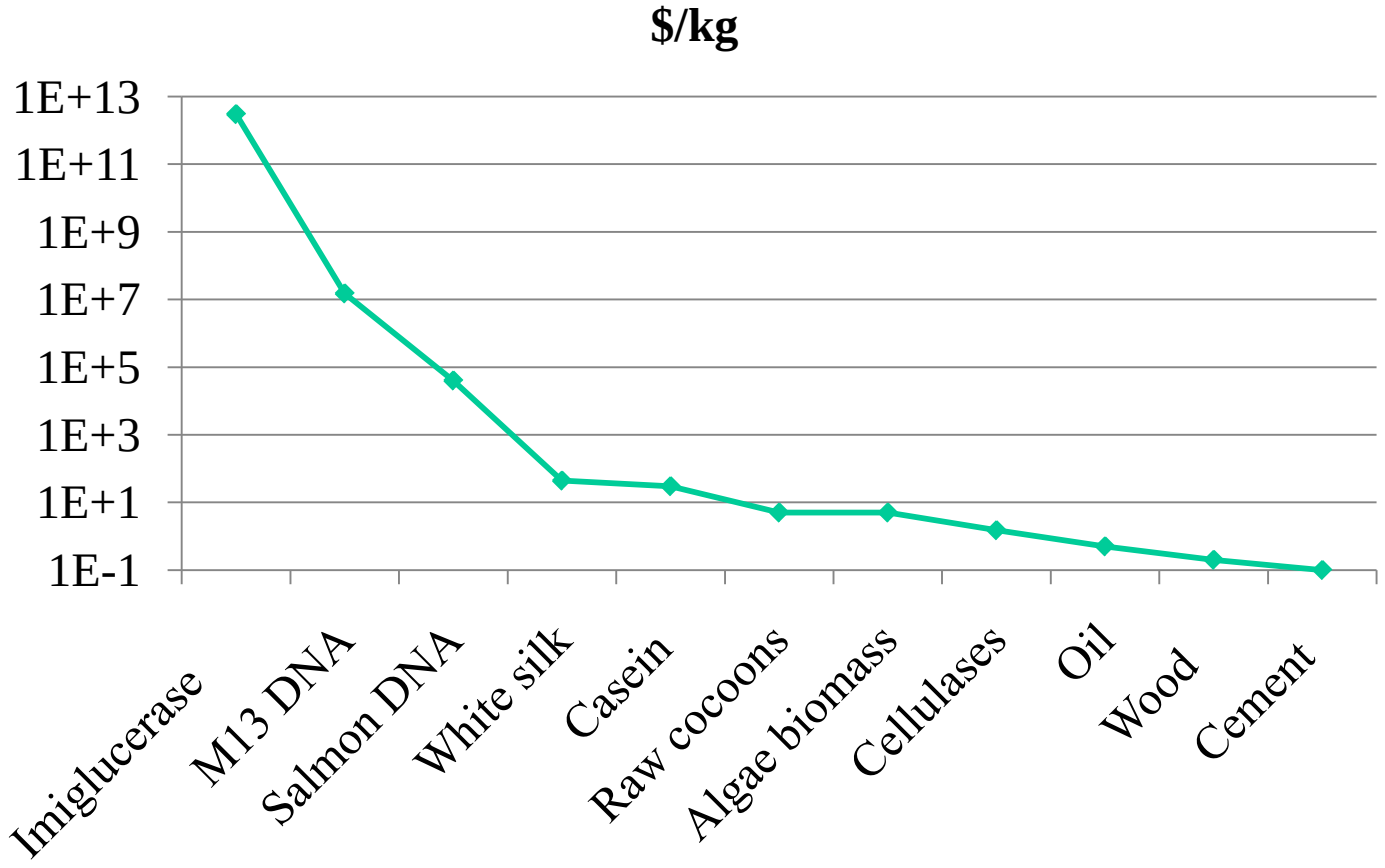
Surveillance

Licensing



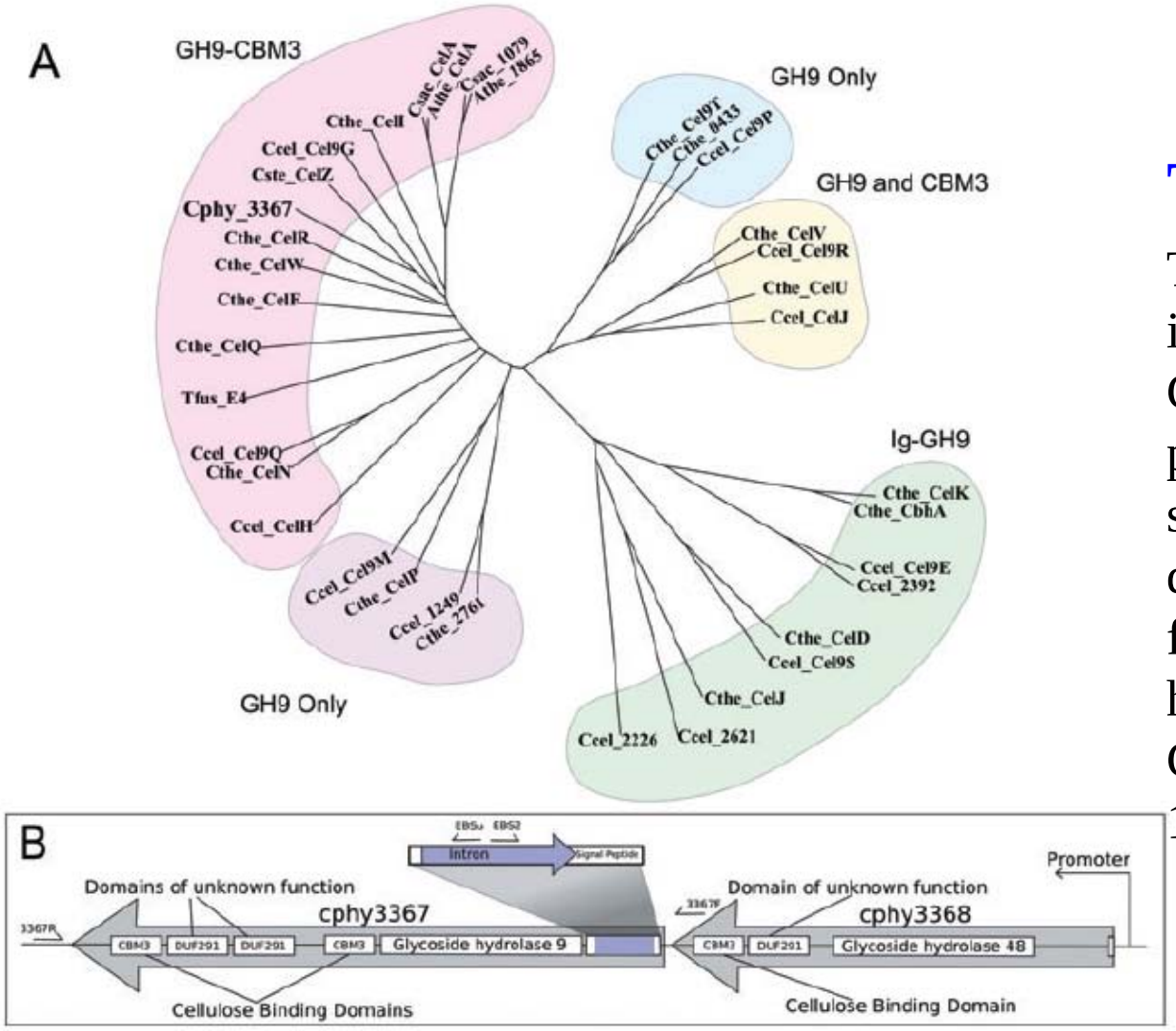
Evolution

How &why genome engineering for chemicals?



WYSS INSTITUTE

Clostridium phytofermentas Cellulase

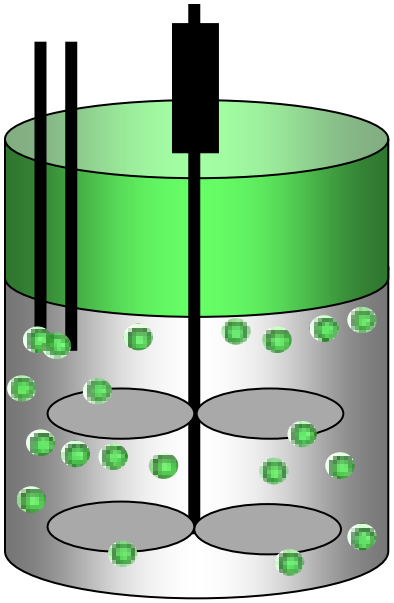


Tolonen et al.

Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367mmi_6890 1..16

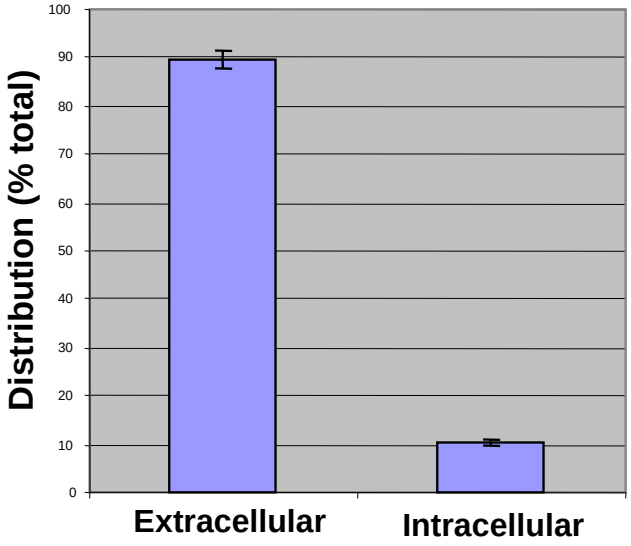
QTEROS
Biofuels from advanced microbiology

Bio-petroleum from microbes flotation -- not distillation

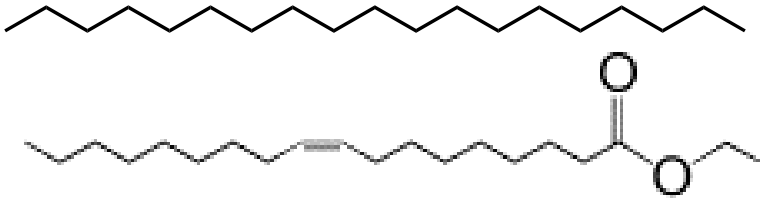


organic

aqueous



Fatty acid derived

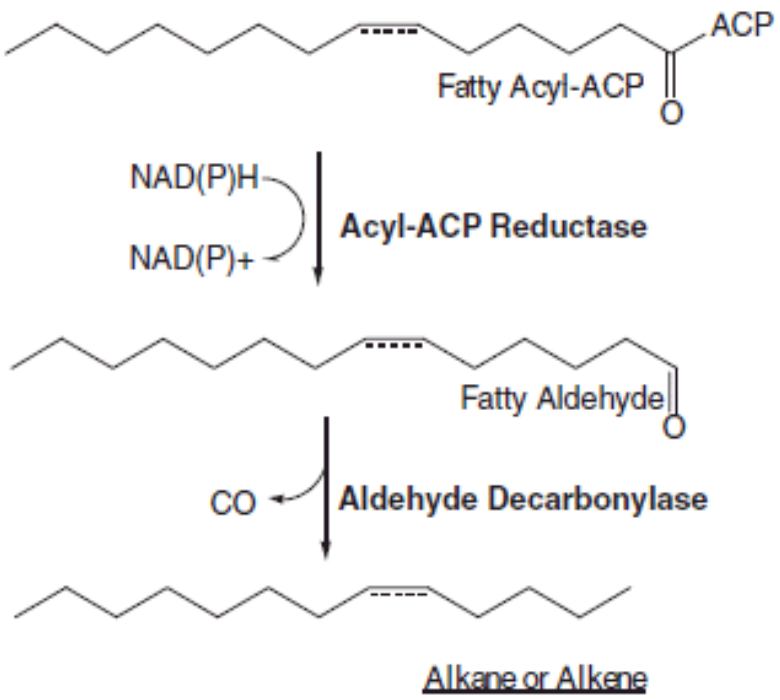


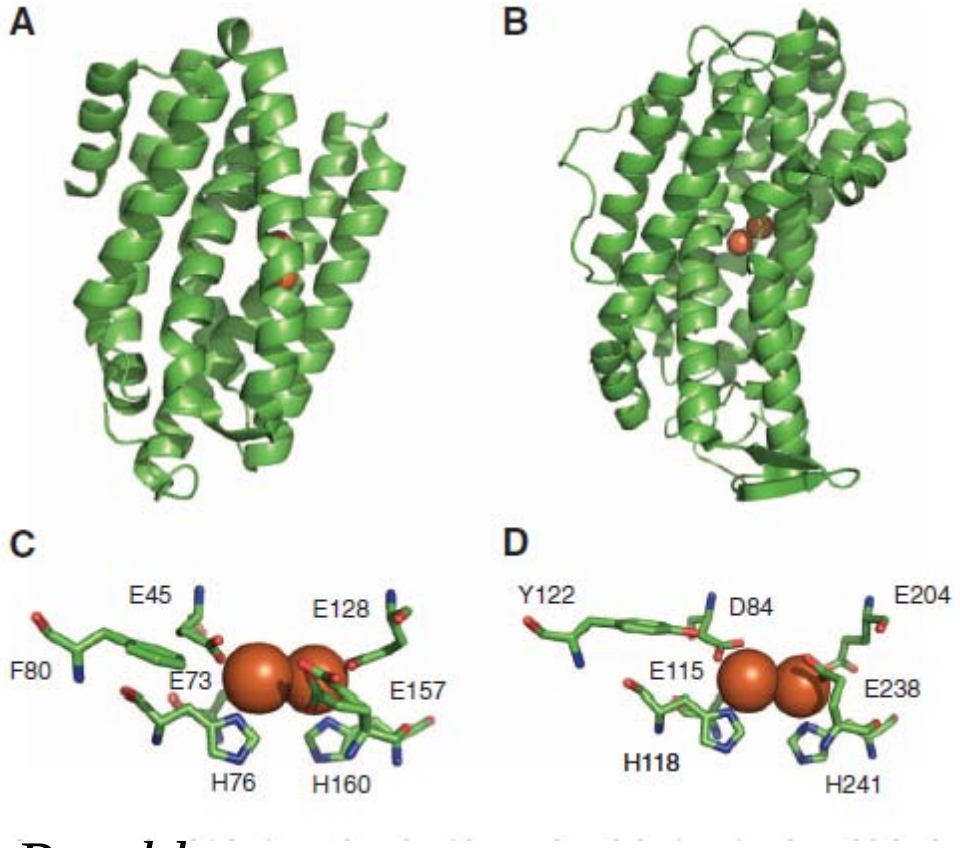
Gasoline & diesel for current engines & infrastructure

LS9, INC.
the renewable petroleum company™

2010 Presidential Green Chemistry Challenge Award

Comparative + Structural Genomics Success





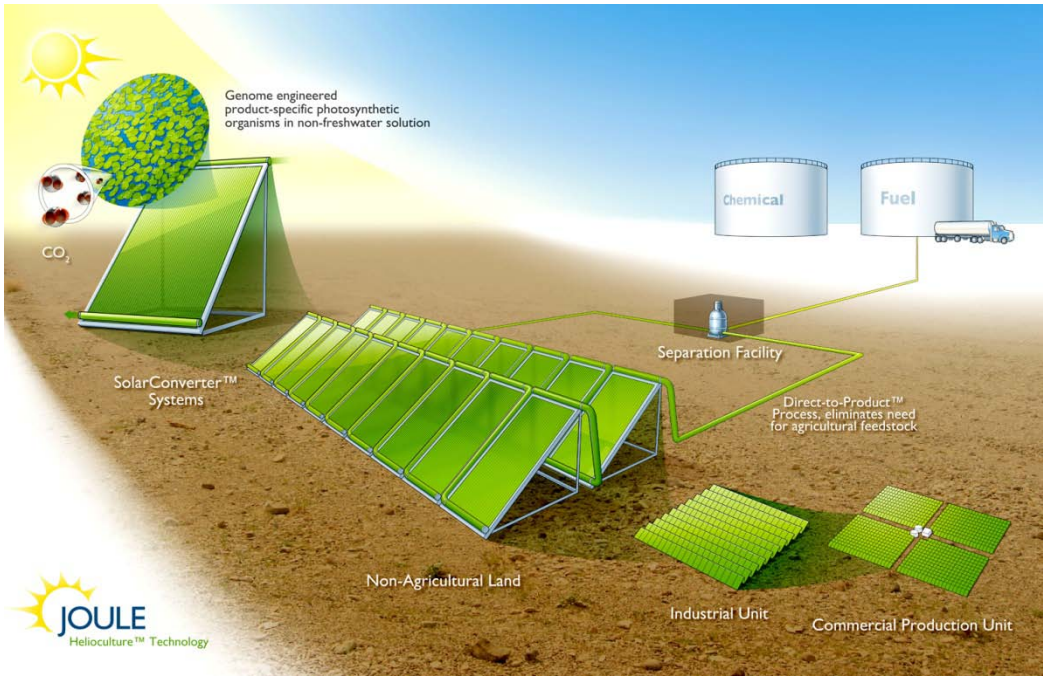
Prochlorococcus Aldehyde decarbonylase

E. coli Ribonucleotide reductase R2

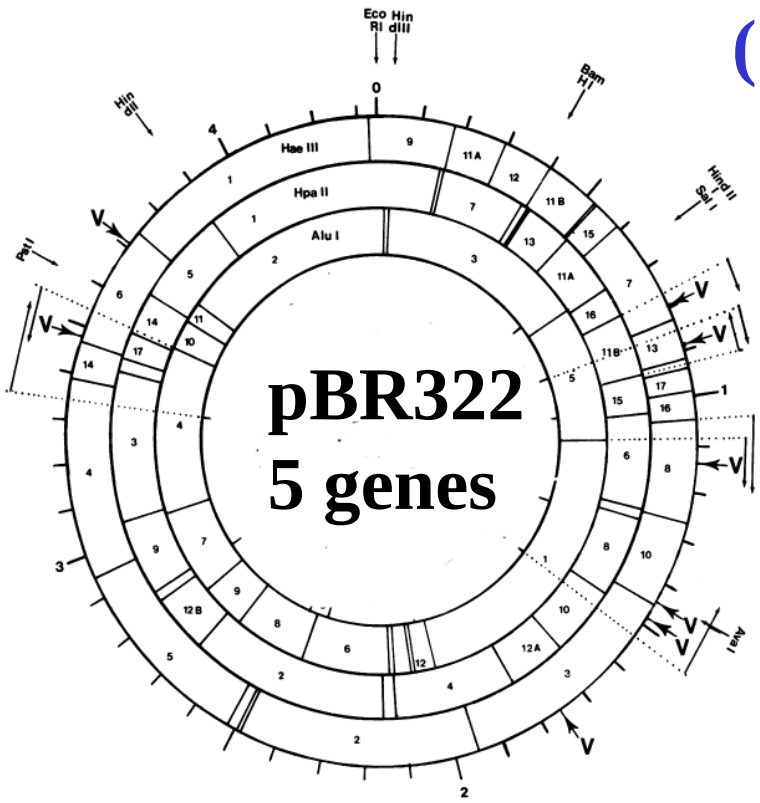
Microbial Biosynthesis of Alkanes.
Schirmer, et al. Science 2010

Joule : fuel from phototrophs

- Achievements:
  - 40% of theoretical maximum productivity of organism
- Ethanol @ 10,000 gal/acre/year (target 25,000)
  - Discovery of unique genes and pathways for hydrocarbon diesel production
  - First-ever secretion of diesel from a phototroph
  - Ability to partition majority of carbon to product via carbon switch



Reading & Writing Genomes:
First semi-synthetic plasmid 1978: $10/b
CGI Human diploid genome 2009: $1500 / 6Gb
(=7 logs/30y mostly since 2005)



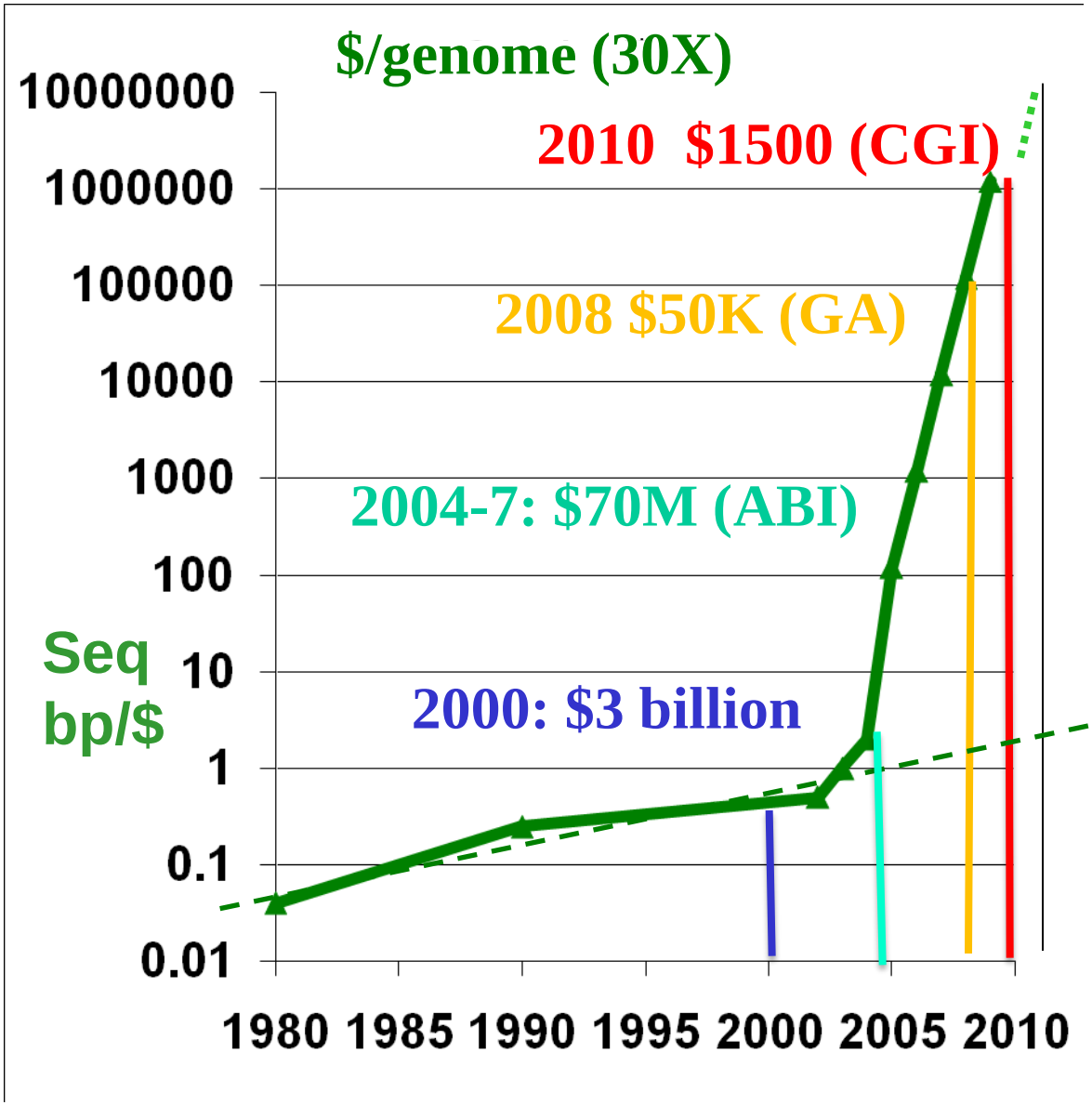
- Human insulin
- Human growth hormone
- Alpha-interferon
- G-CSF
- TPA
- GM-CSF
- Gamma-interferon
- IL-2
- Erythropoietin
- Hepatitis B vaccine

NAR **1978**
Sutcliffe & Church
(BR:Bolivar & Rodriguez)

(Amgen, Biogen, Genentech, etc)

40,000-fold in 4 years

(Moore's law)
1.5x/yr for electronics
vs
10x/yr for DNA Sequencing



>20 years ahead of the 1970-2004 exponential

Moore's law =1.5x/yr vs 10x/yr

1st-generation

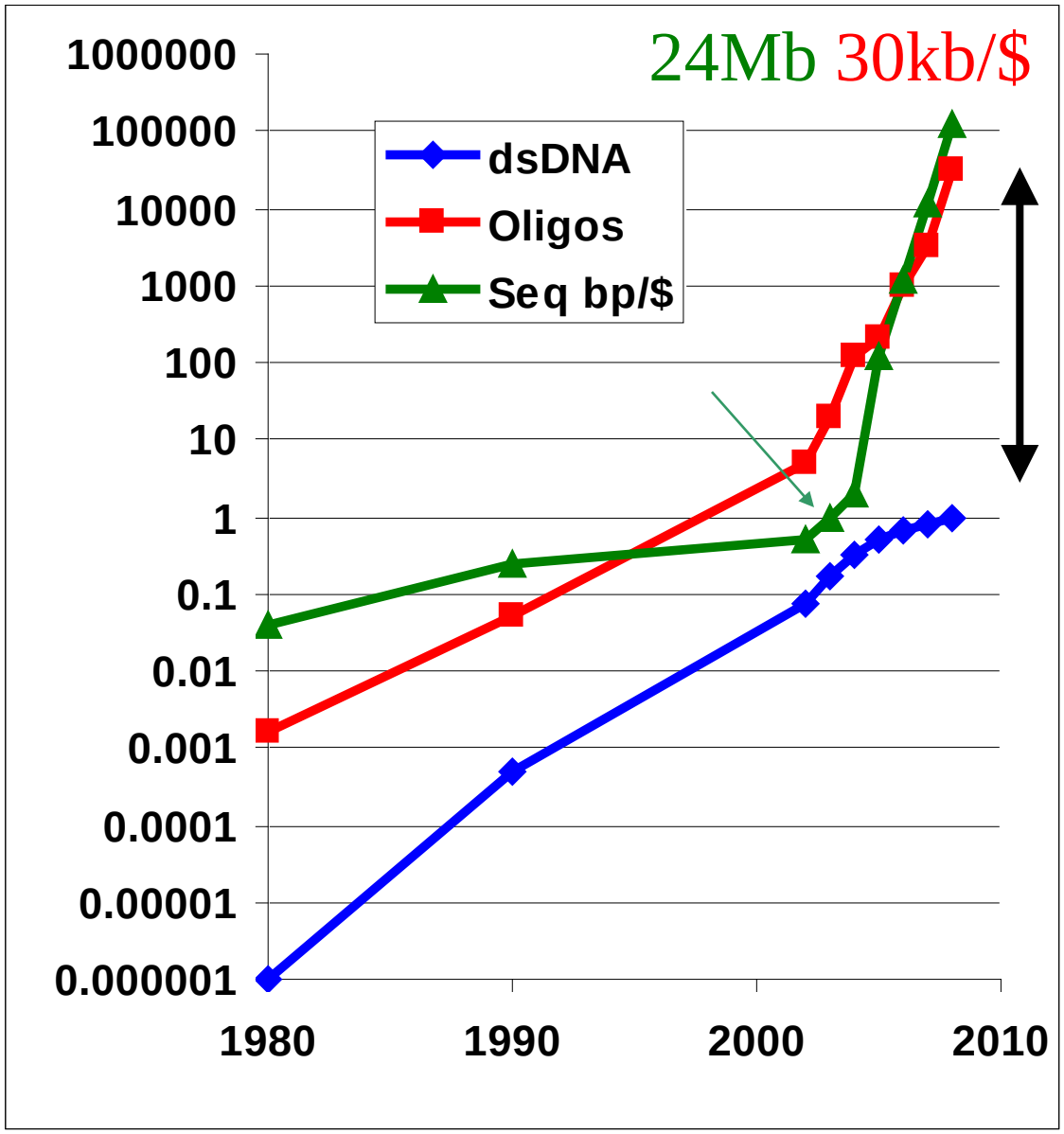
Gene synthesis

vs

2nd-generation

Sequencing

& DNA synthesis



2nd-generation sequencing technologies

1. **Illumina-GA**	**SbP** Fluorescent read-length 2*110 bp
2. **AB-SOLiD**	**SbL** Longest ligation reads
3. **CGI**	**SbL $2000 genome,** rolony grid, 100Kb haplotypes
4. **Polonator**	**SbL/P Open-source,** $170K device, 100Mb haplotypes
5. **Roche-454**	**SbP** Long reads (>0.4 kb)
6. **Helicos**	**SbP-sm** High parallelism & quantitation
7. **Ion Torrent**	**SbP $50K, small device**
8. **Pacific Bio**	**SbP-sm** Long reads (>2.0 kb)
9. **Intelligent Bio**	**SbP** hexagonal grid
10. **GnuBio**	**SbP**-picoliter droplets
11. **Halcyon**	**EM-sm Long reads (>Mb), $100 genome**
12. **Genizon BioSci**	**SbH** in situ sequencing
13. **LightSpeed**	**SbL** 16X higher density, >10X speed
14. **Bionanomatrix**	**SbP-sm** Fluorescent mapping
15. **OxfordNanopore**	**Pore**-protein-**sm** small device
16. **Visigen**	**SbP-sm** Pol <> dNTP FRET
17. **ZS Genetics**	**EM-sm** Iodine labels
18. **Nabsys**	**Pore-SbH-sm** small device
19. **GE Global**	**SbP-sm**
20. **IBM**	**Pore** Si-**sm** small device
21. **Electronic Biosci**	**Pore**-protein-**sm**

2nd-Gen Gene Synthesis: chips

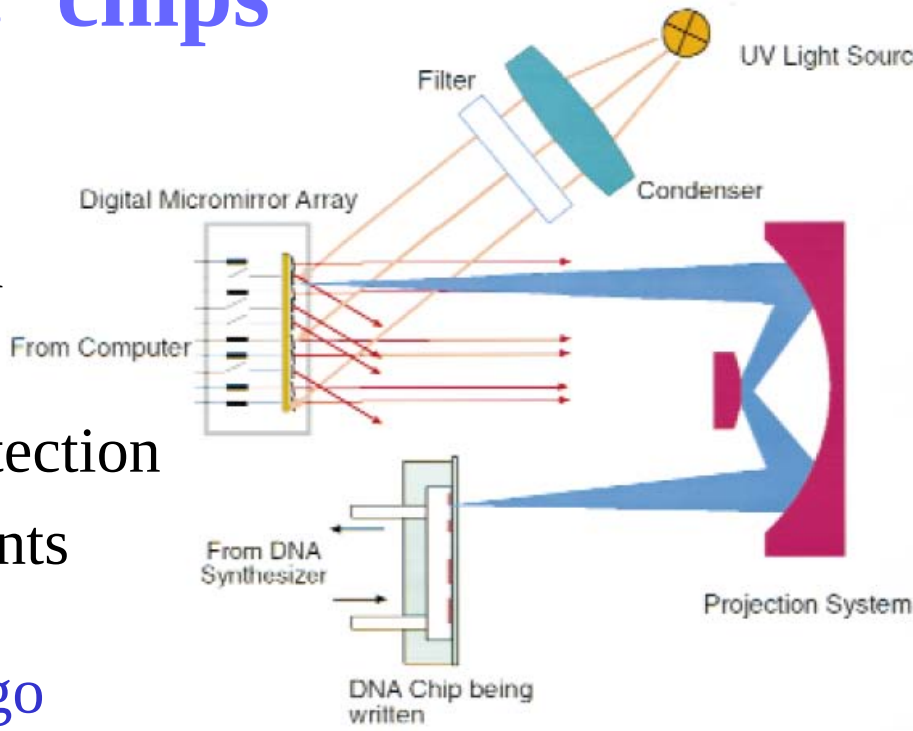
$500 per 15Mbp

8K Xeotron Photo-Generated Acid

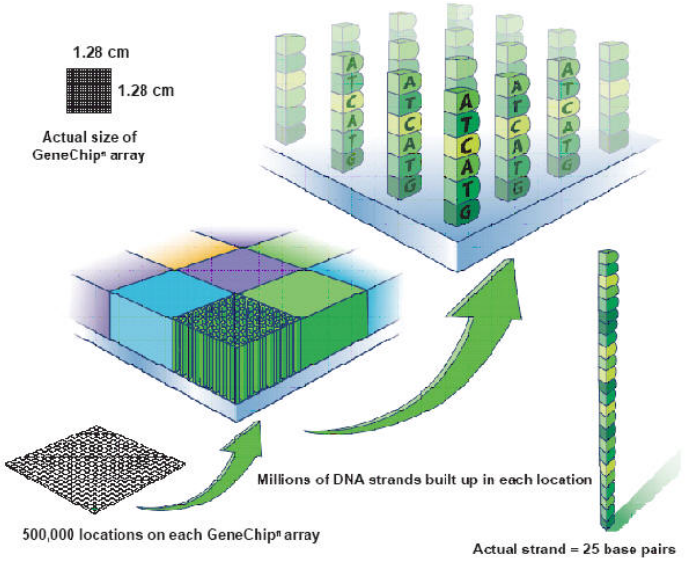
12K Combimatrix Electrolytic

120K Roche, Febit Photolabile 5'protection

244K Agilent Ink-jet standard reagents



4 chip technologies: Amplify oligo pools with flanking universal primers

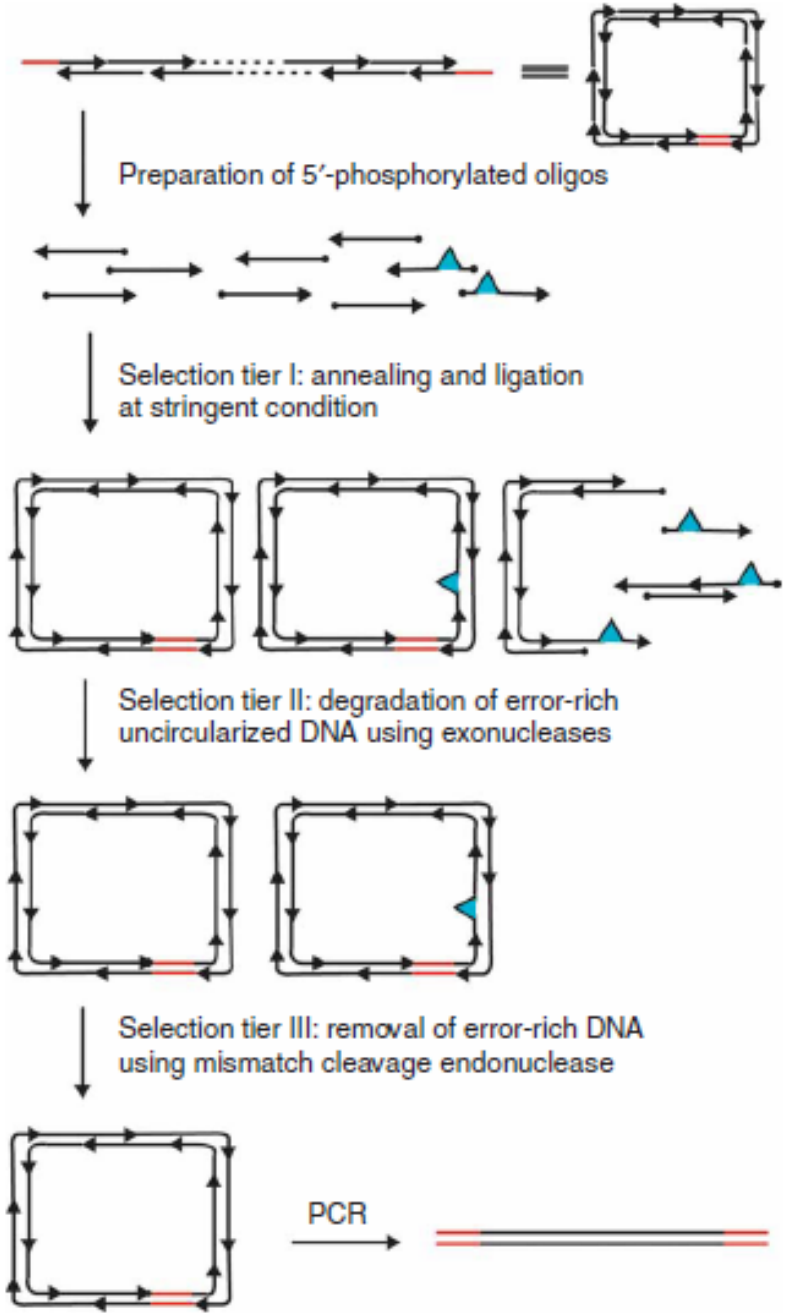


4 Paths to error correction

1.Hyb-Select: Tian et al. 2004 Nature

2. MutS: Carr & Jacobson 2004 NAR

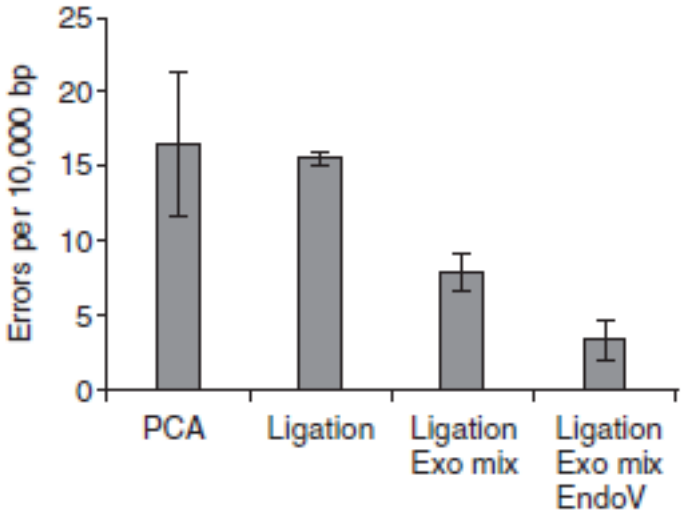
3.MutHLS: Smith & Modrich 1997 PNAS



4 Paths to error correction

#4 : Bang

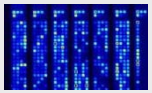
Nat Meth. 2008

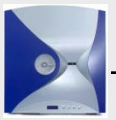


Coalescence of 2nd Generation of DNA reading & writing

"perfect" part

Write



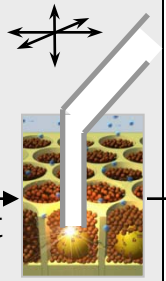


Oligo
elute
amplify

Febit or LC

Read

sort
&
select

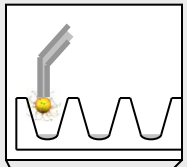


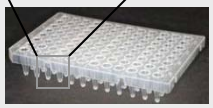
GS FLX
Roche/454

PTP

Select

pick
&
place

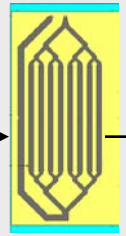




multiwell plate

PCR

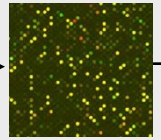
Polonator



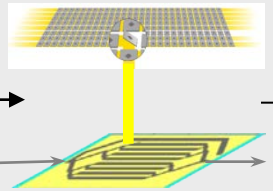
Polonator
flowcell

Agilent OLS

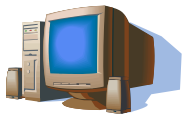
Data



Array
synthesis



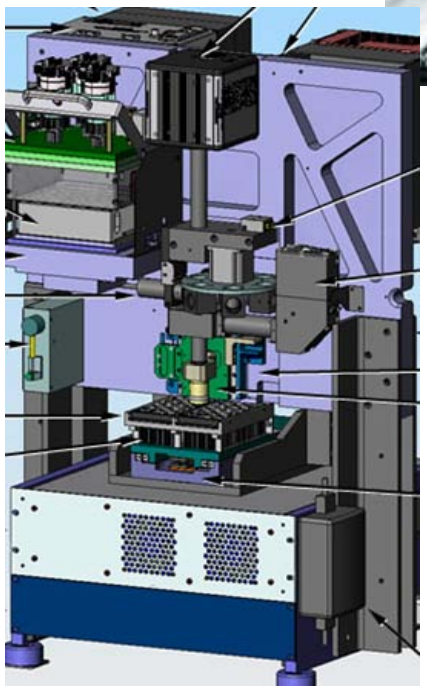
Rolony
Photorelease



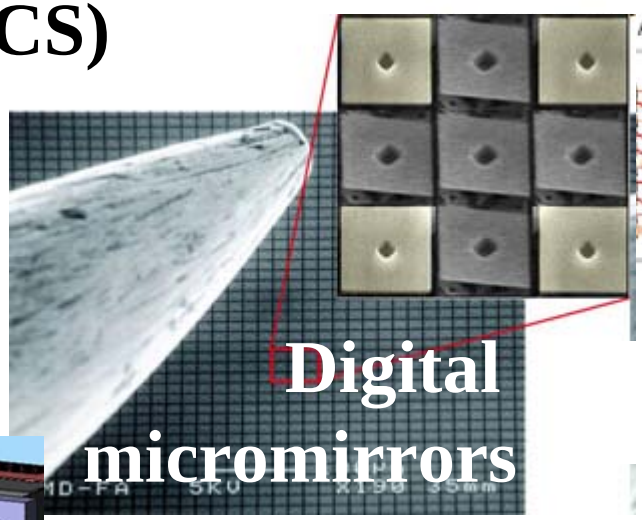
Oligonucleotide-Design

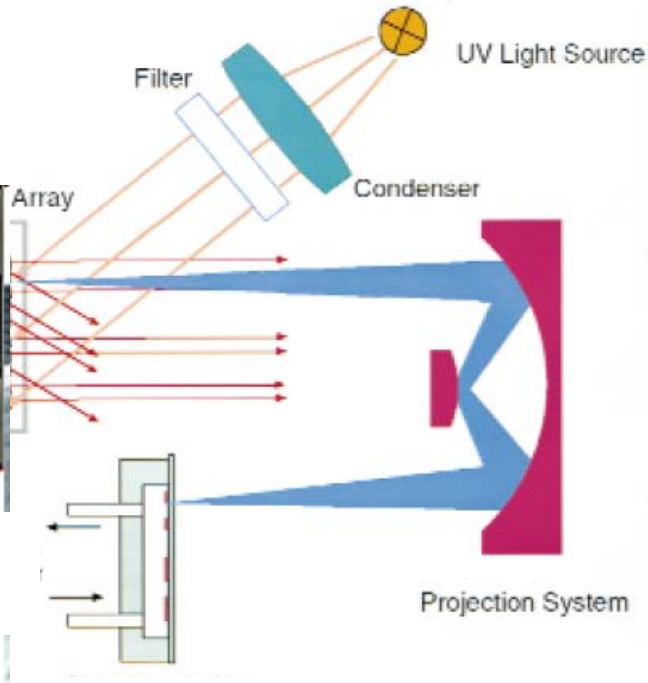
From open-access Sequencer to Bio-Fab

1. Select ‘perfect part’ sequences
2. Device characterization
3. Cell sorting (FACS)



Rich Terry





Flow- cell

billion beads or cells/run

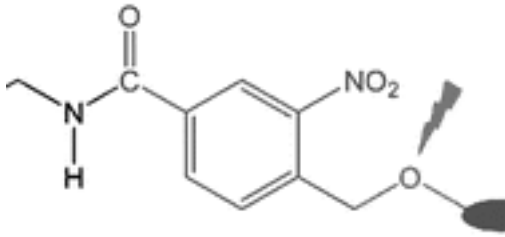
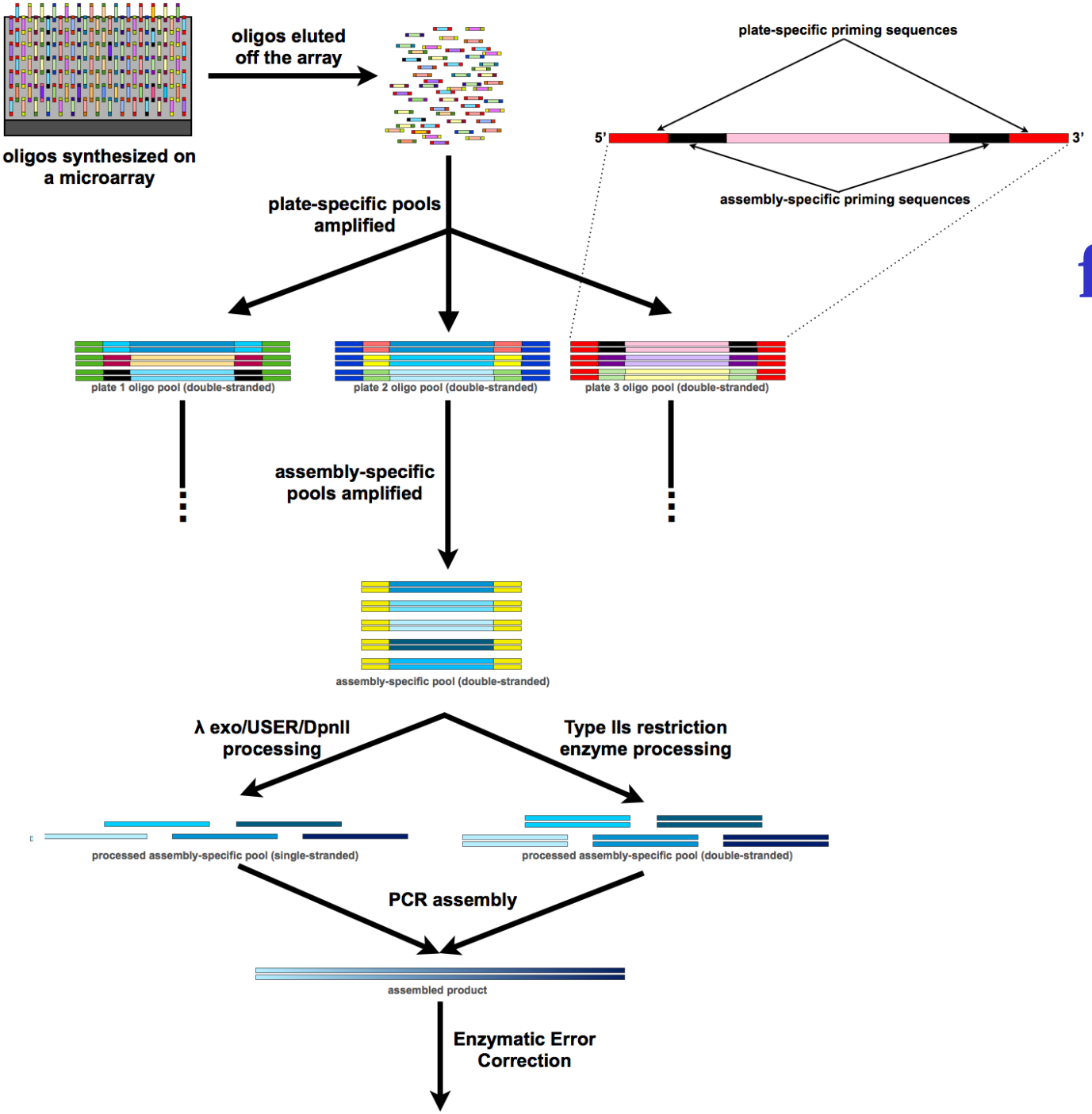


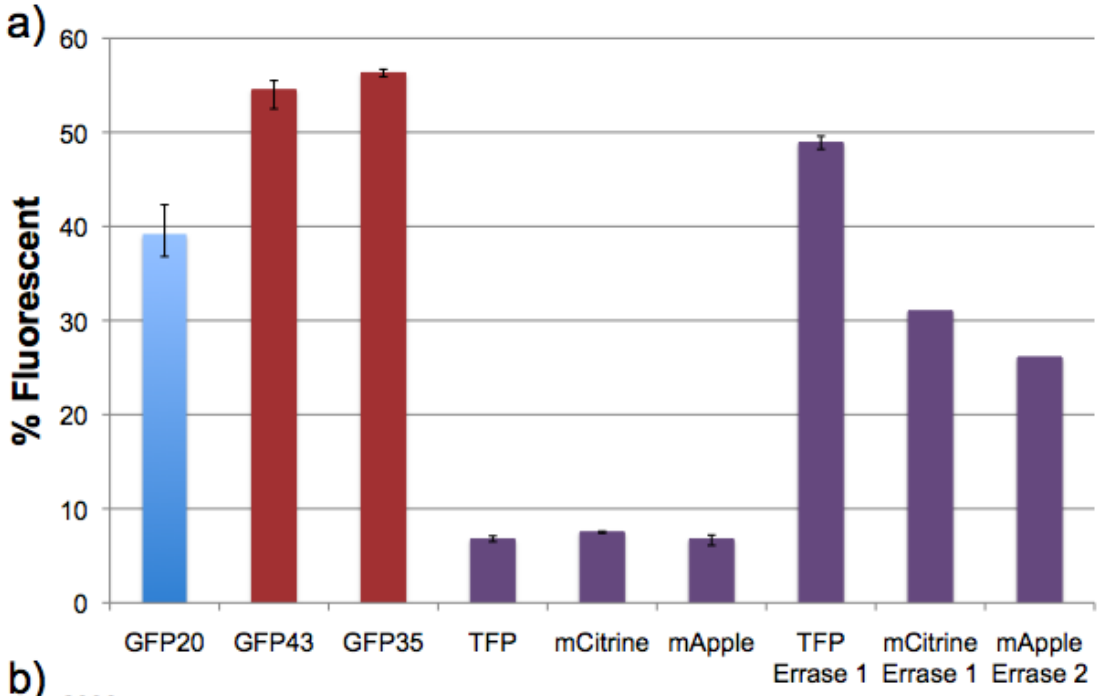
Photo-labile immobilization

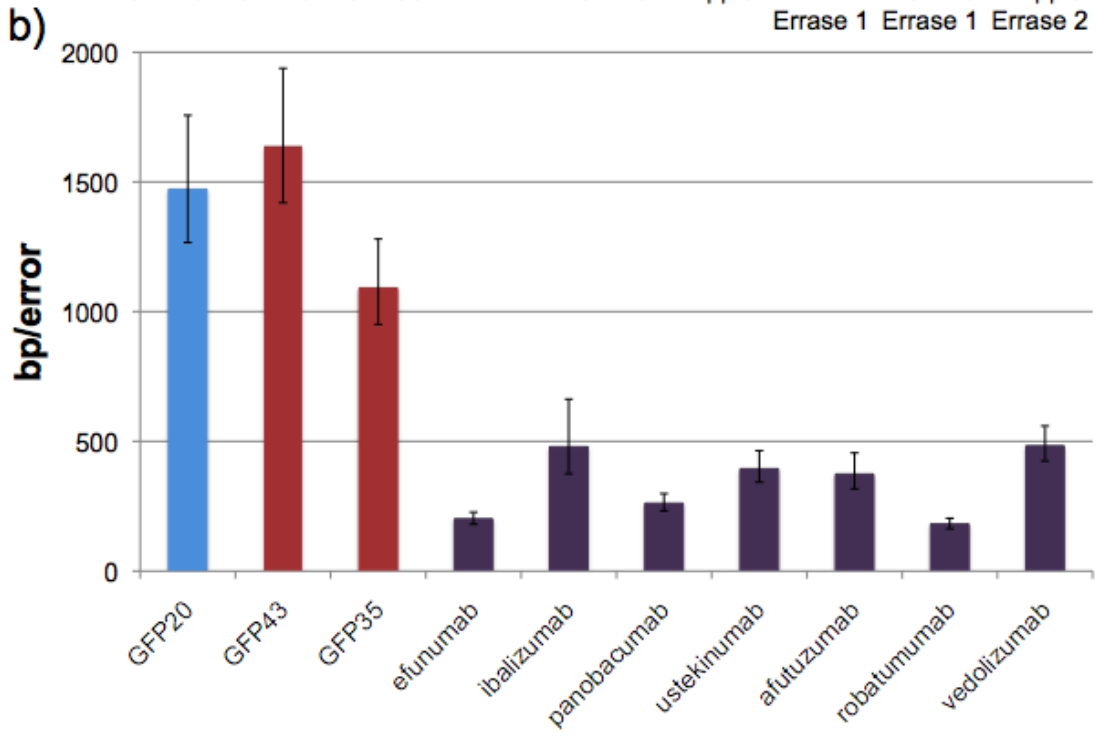
WYSS INSTITUTE



2nd Gen Assembly from 2.6 Mb raw oligos

Kosuri Eroshenko

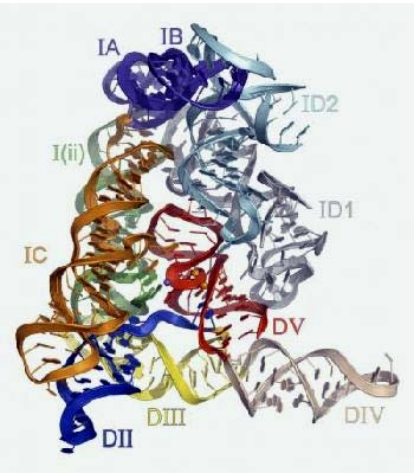




2nd Gen Assembly from 2.6 Mb raw oligos

Kosuri Eroshenko

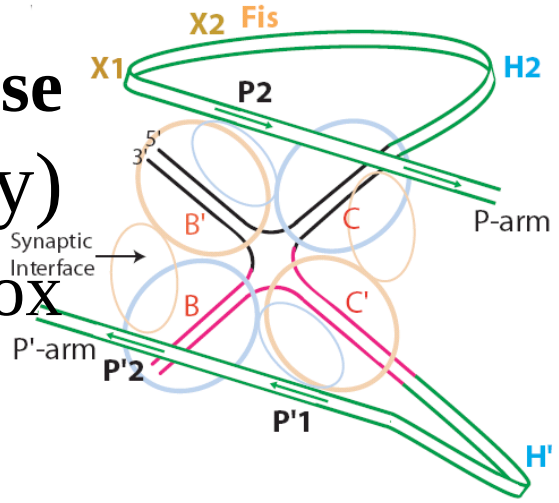
4 Protein/RNA-directed recombination strategies



1. Integrase/recombinase

λ(Gateway)

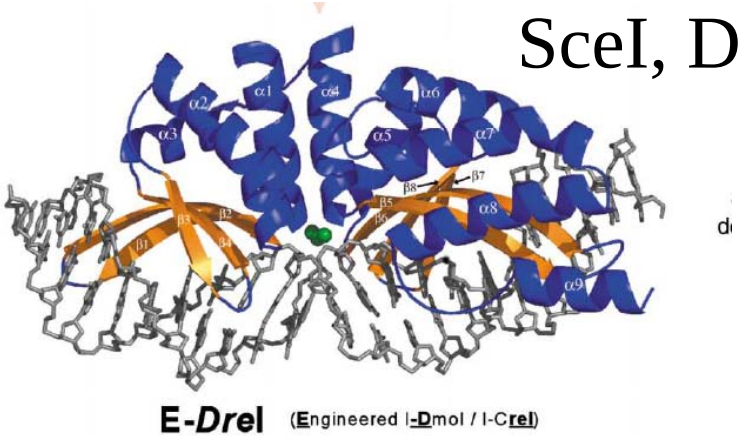
ϕC31, Cre-lox



2. Group II introns

3. Meganucleases

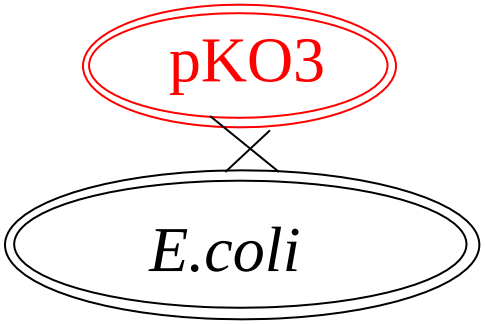
SceI, DreI



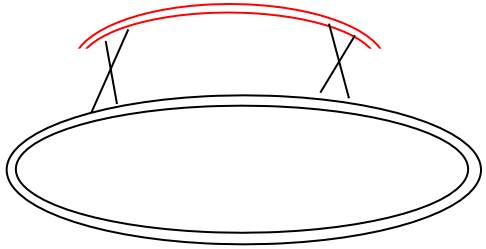
4. Zn Finger Nucleases

2. Pyle et al. 3. Mol Cell. Chevalier, et al. 4. Mol Cell. 2003 Aihara, et al.

4 DNA homology-directed strategies

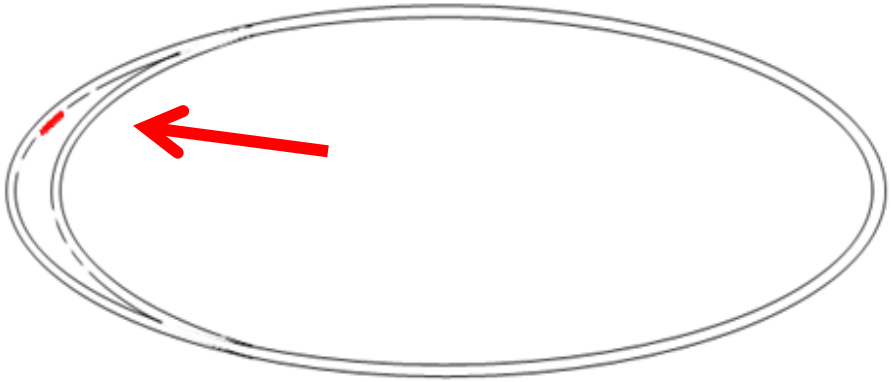


#1: ds-Circle x Circle
2 step recA+ recombination
Select + counterselect
Link et al J. Bact 1997
(Open-access)

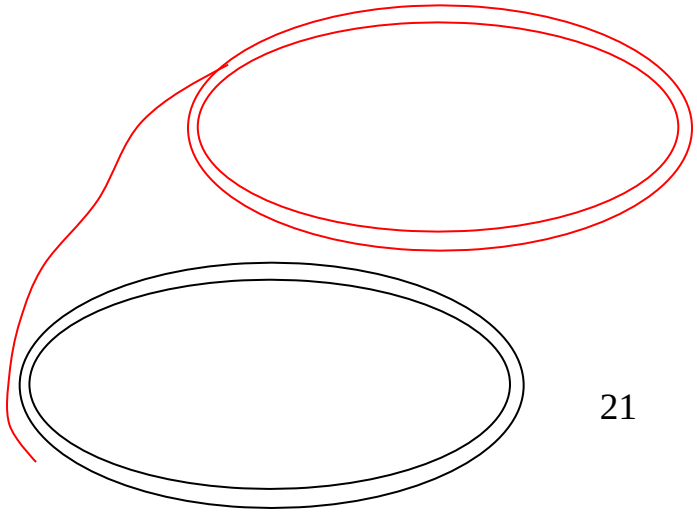


#2: ds-Linear x Circle
1 step 5'>3'exo Reda/E b/T
Select
Zhang et al Nat.Gen 1998 Yu et al. PNAS 2000
(GeneBridges license)

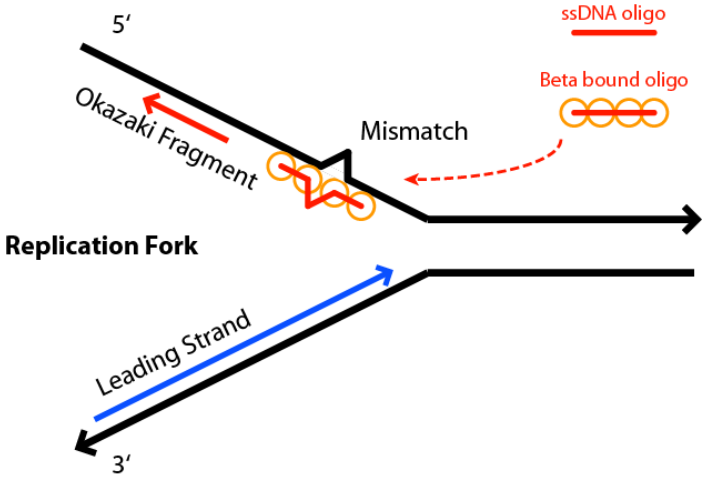
#3: ss-90mer x ds-Circle
Costantino &Court PNAS'03
Wang et al., Nature '09



#4: ss-Mb x ds-Circle conjugation

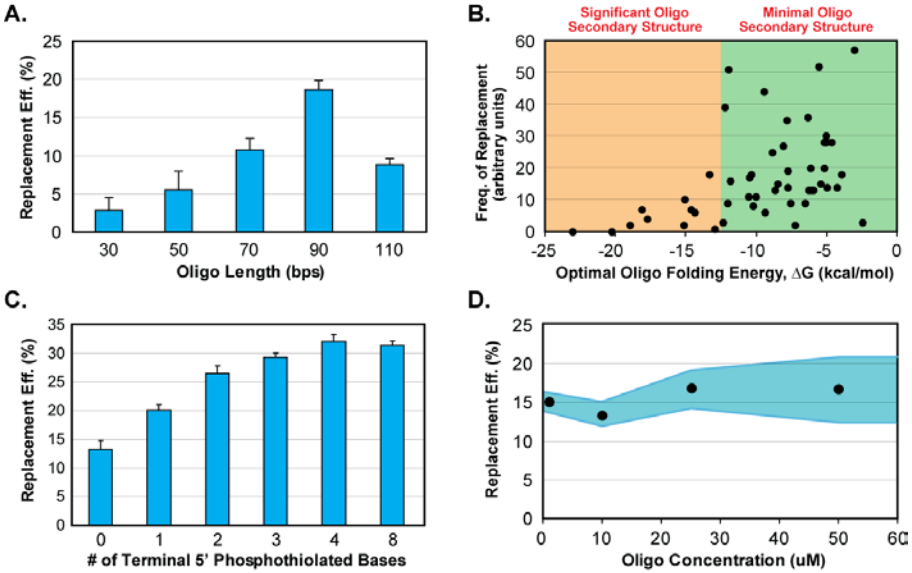


Multiplex Automated Genome Engineering (MAGE)



Allelic Replacement

- Strain: MG1655, ΔmutS, integrated λ-Red
- Highly complex oligo pools for multiplexed multi-loci modifications
- >4 billion bp of targeted genetic variation produced per day



Optimized Parameters

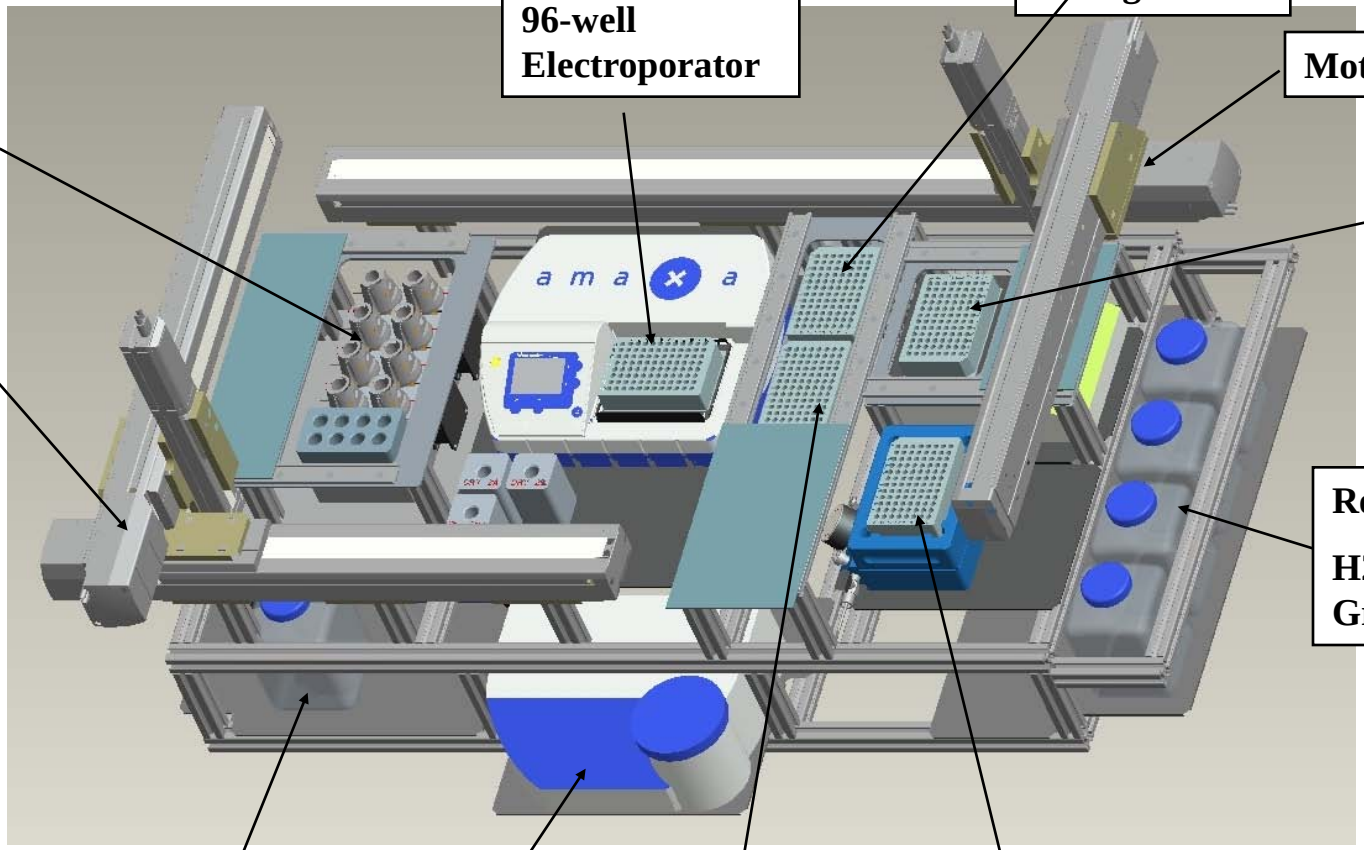
- Oligo length: 90mer
- Oligo 2ndary structure: <12 kcal/mol
- Oligo half-life: 5’ phosphothiol bps
- Oligo conc.: up to 50 uM

- Cycle time: 2 to 2.5 hrs
- up from 1E-4 to 25% efficiency per cycle

Costantino & Court PNAS 2003 ; Wang HH, et al Nature 2009

MAGE System Upgrade Jan-May 2010

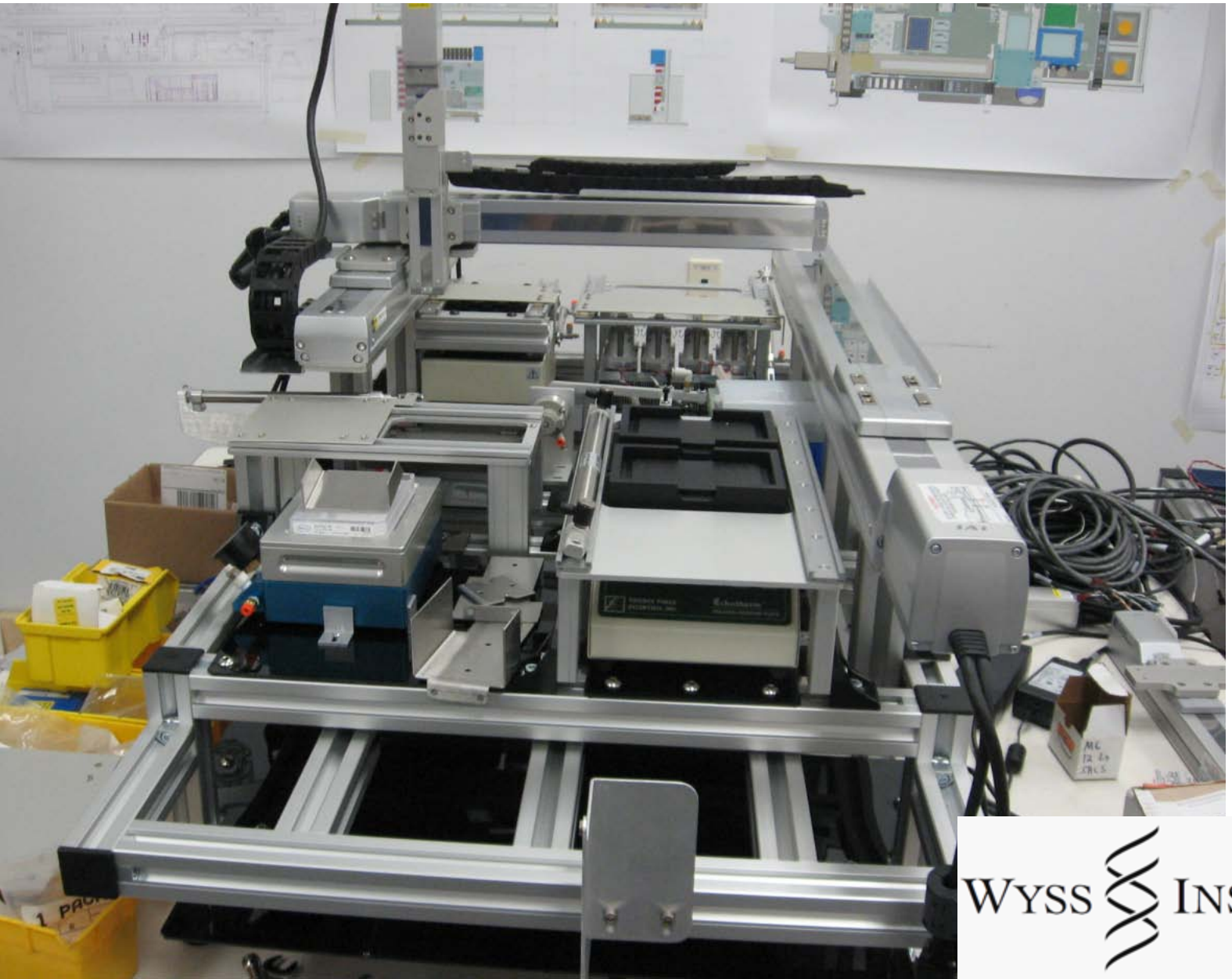
Dimension:
4’ x 3’ x 2.5’



Harris Wang

WYSS INSTITUTE

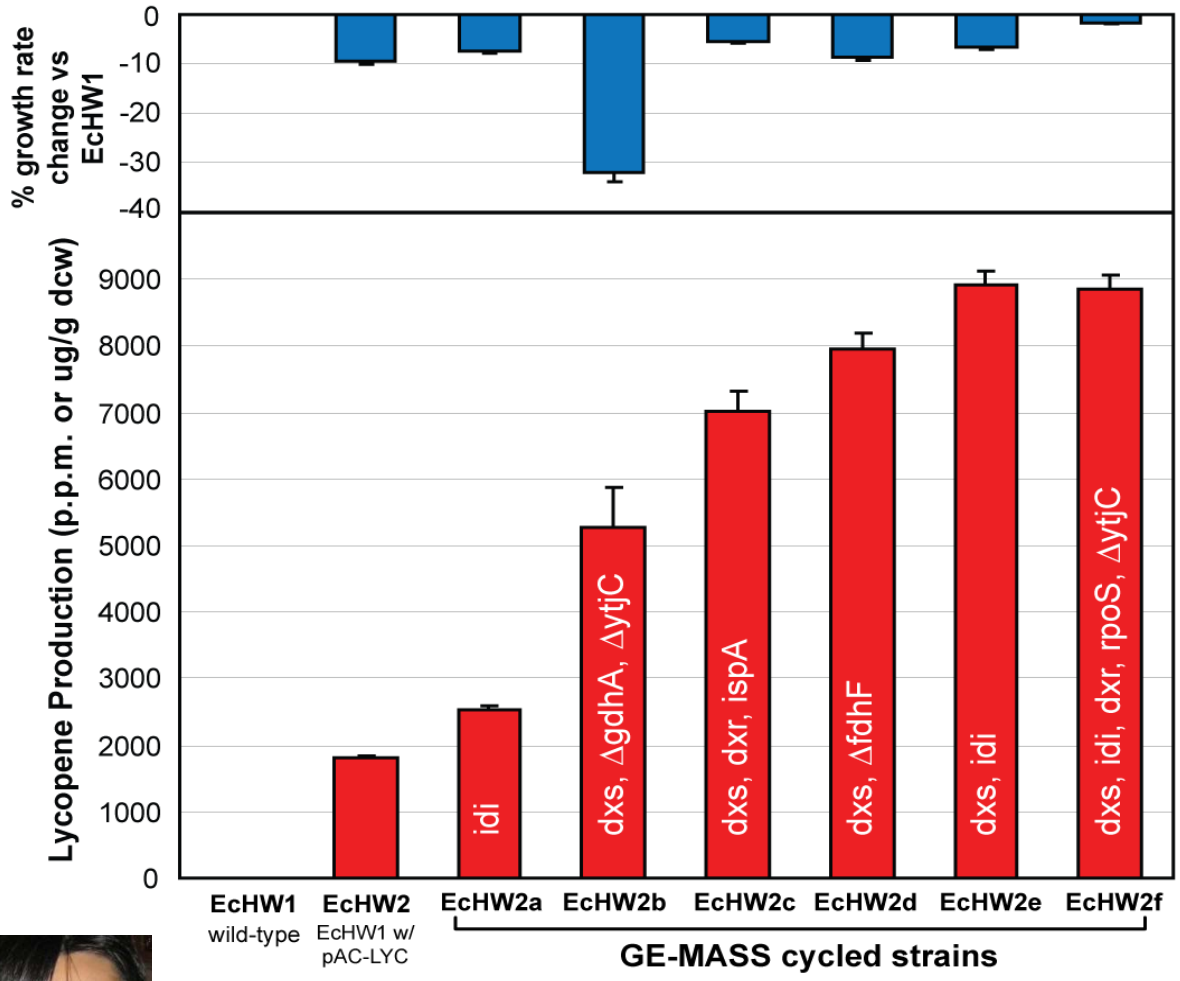
MAGE System Upgrade Sep 2010

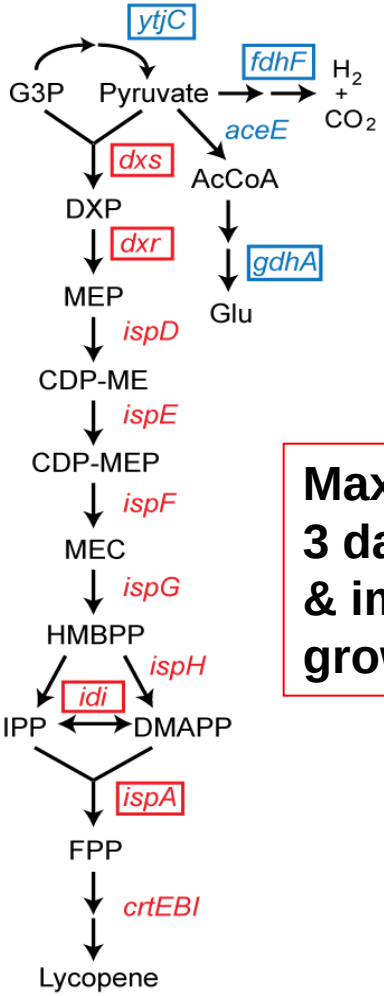


WYSS INSTITUTE

Accelerated Evolution 100K combinations

Lycopene (hydrocarbon): 20 genes up, 4 down, 2 new





Max yields in 3 days & improve growth rates

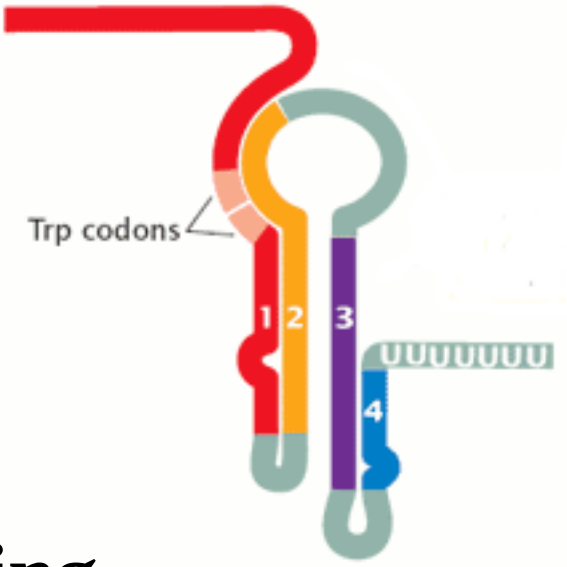
Harris Wang H et al Nature 2009

Existing Sensors (select for new ligands)

55 DNA binding proteins: ada araC arcA argPR carP cpxR crp cspA cynR cysB cytR deoR dnaA dgsA fadR farR fhlA flhCD fnr fruR fur galR gcvA glpR hipB iclR ilvY lacI lexA lrp malT marR melR metJ metR modE nagC narL narP ntrC ompR oxyR pdhR phoB purR rhaS rpoE rpoH rpoN rpoS soxS tetR torR trpR tyrR
12 Riboswitches: Adenine B12 FMN Guanine Glucosamine-6-phosphate Glycine di-GMP Lysine Molybdenum PreQ1 SAM SAH TPP theophylline 3-methylxanthine

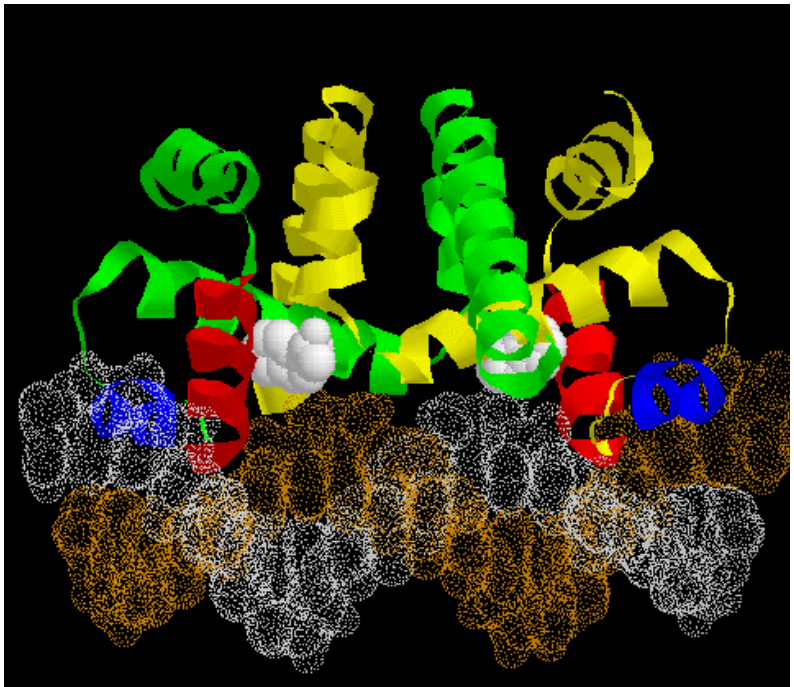
http://pubs.acs.org/doi/abs/10.1021/ja048634j

In vivo coupled biosensors

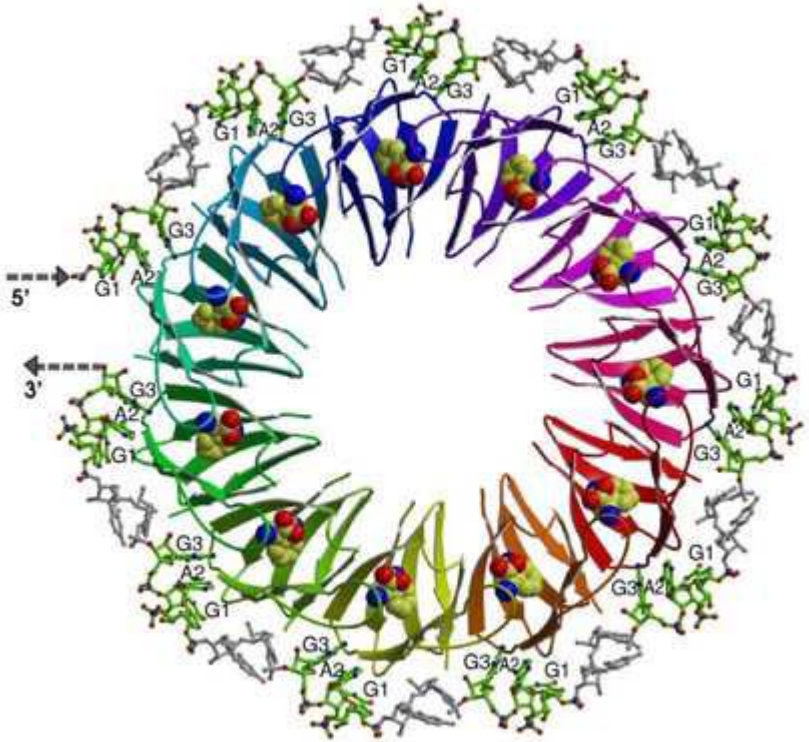


tRNA-ribosome

Repressor ds-DNA

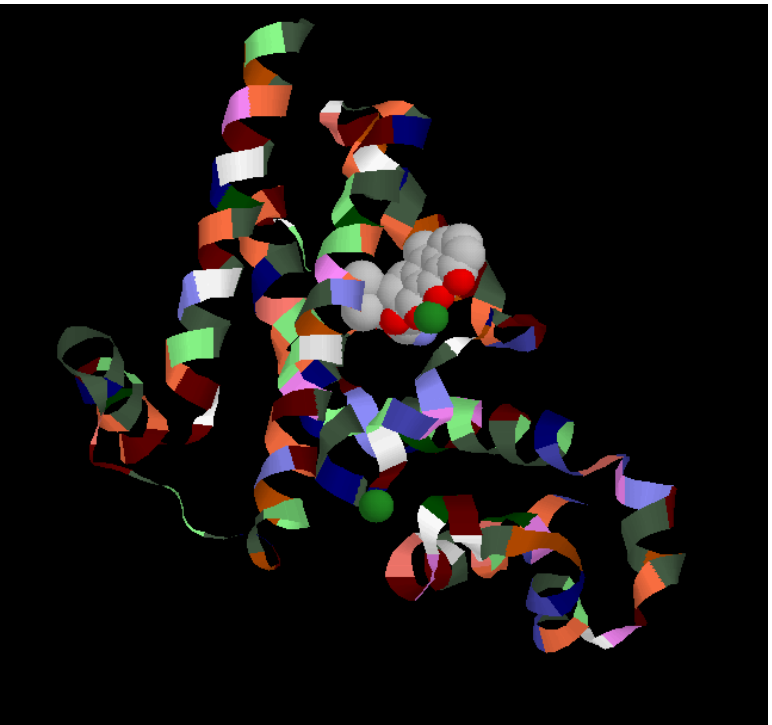


mRNA binding

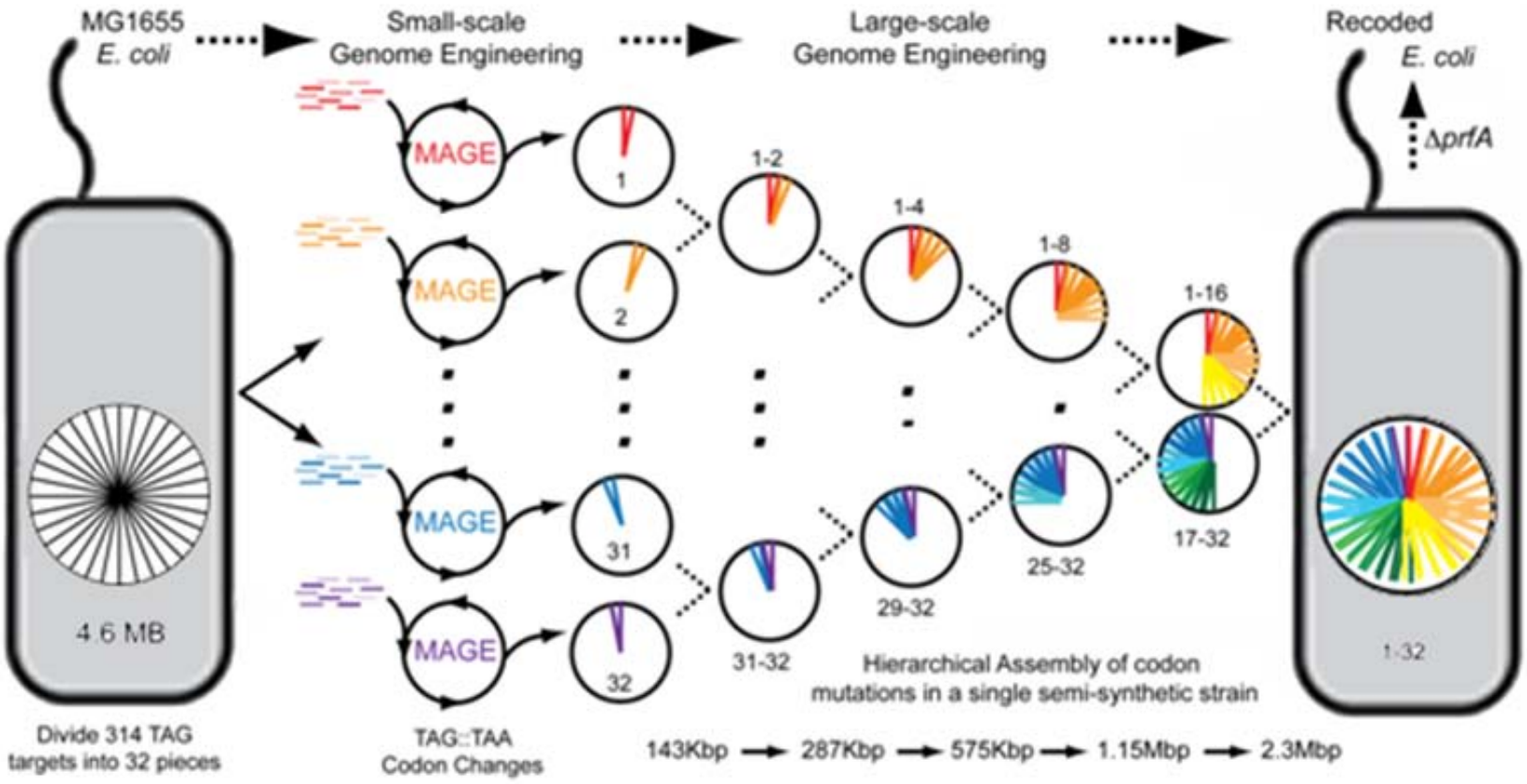


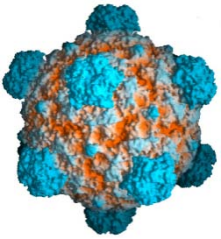
In vivo coupled biosensors

Tetracycline repressor ds-DNA



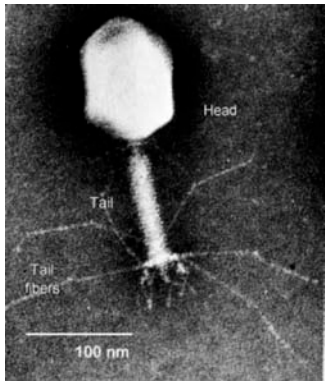
Divide & Conquer Genome Engineering





Multi-virus resistance: stop codons: TAG / total

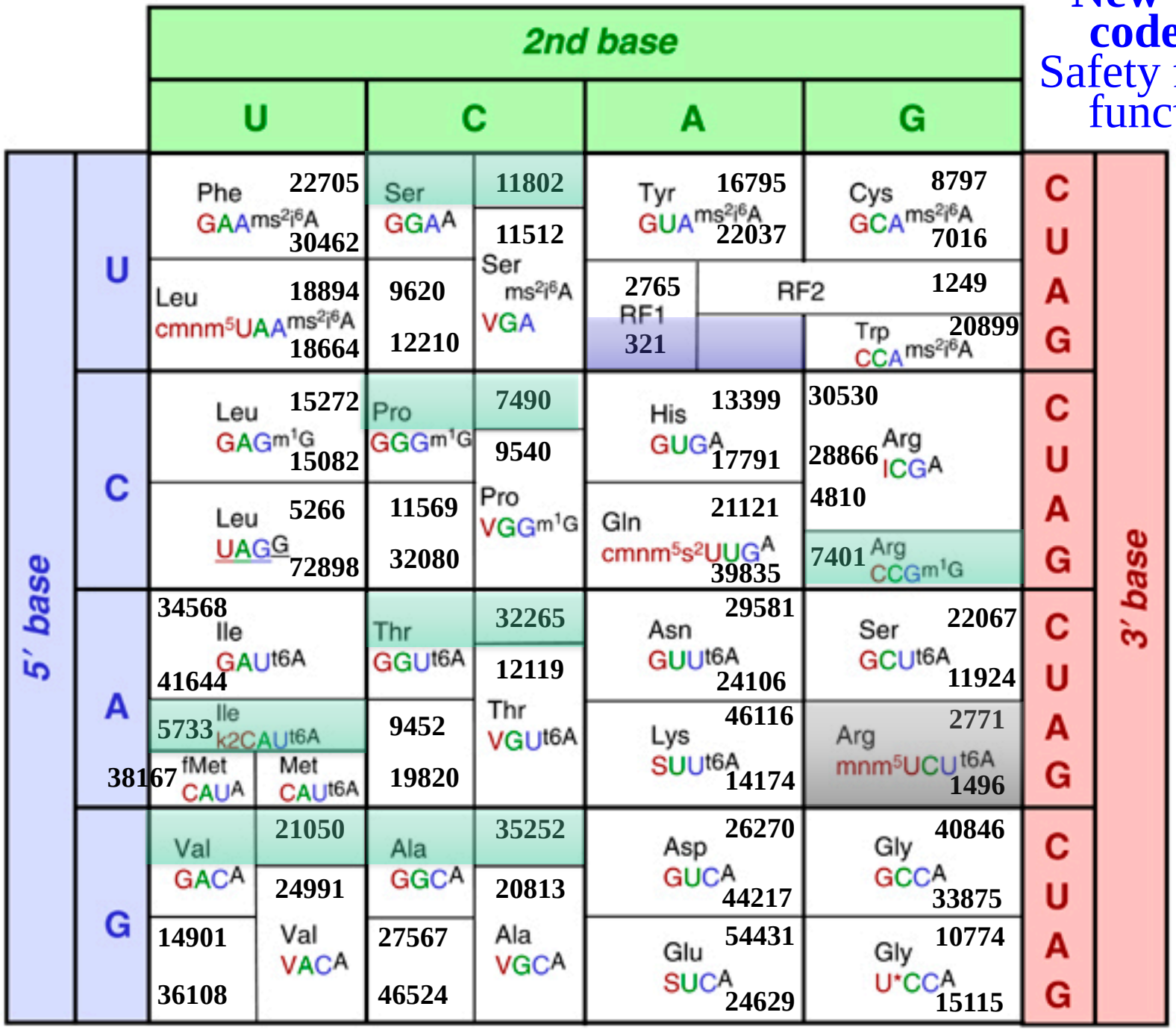
ϕX-174	5,386 b	ss-DNA	**0** / 9
M13	6,407 b	ss-DNA	**1** / 10
MS2	3,569 b	ss-RNA	**2** / 4
T7	39,937 b	ds-DNA	**6** / 60
T4	168,903 b	ds-DNA	**19** / 277
E.coli	4,639,675 b	ds-DNA	**314** /1,360,152



Farren Isaacs et al

ncbi.nlm.nih.gov/nuccore/9626372

56718463 176120924 9627425 29366675 (7 tRNAs: RITSPGL)



New translation code: novel AA

Safety features: no functional DNA exchange

multi-virus resistance

Isaacs
Charalel
Church
Sun
Wang
Carr
Jacobson
Kong
Sterling

Next: Freeing 8/33 tRNAs & 1/3 RF (9/64 codons)

ATA(I), GTC(V), TCC(S), CCC(P), ACC(T), GCC(A), CGG(R), AGR(R), **TAG(-)**

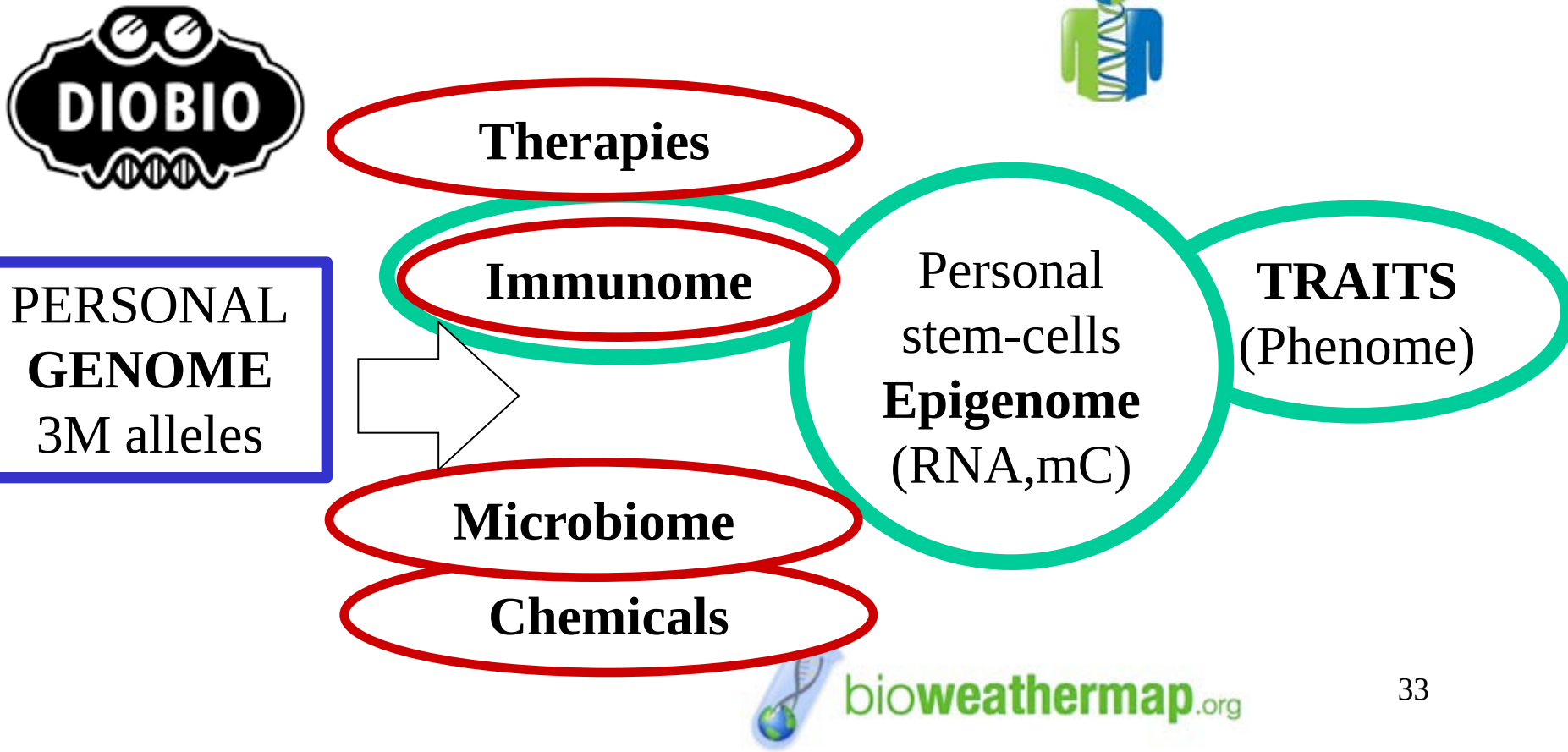
Min # bp changed= 5733 + 21050 + 11802 + 7490 + 32265 + 35382 + 7401 + 2771 + 1496 + **314**

= 125,704 bp (2.7 % of the genome)

Church GM (2009) Safeguarding Biology. Seed 20:84-86.

Genomes Environments Traits

One in a life-time genome + yearly (to daily) tests
Bio-weather map : Allergens, Microbes, Viruses
PersonalGenomes.org



4 Alleles from 4 Genome Sequences

- Primary ciliary dyskinesia (lungs)

- Pyrimidine synthesis (face & limbs)

Logan & Heather Madsen

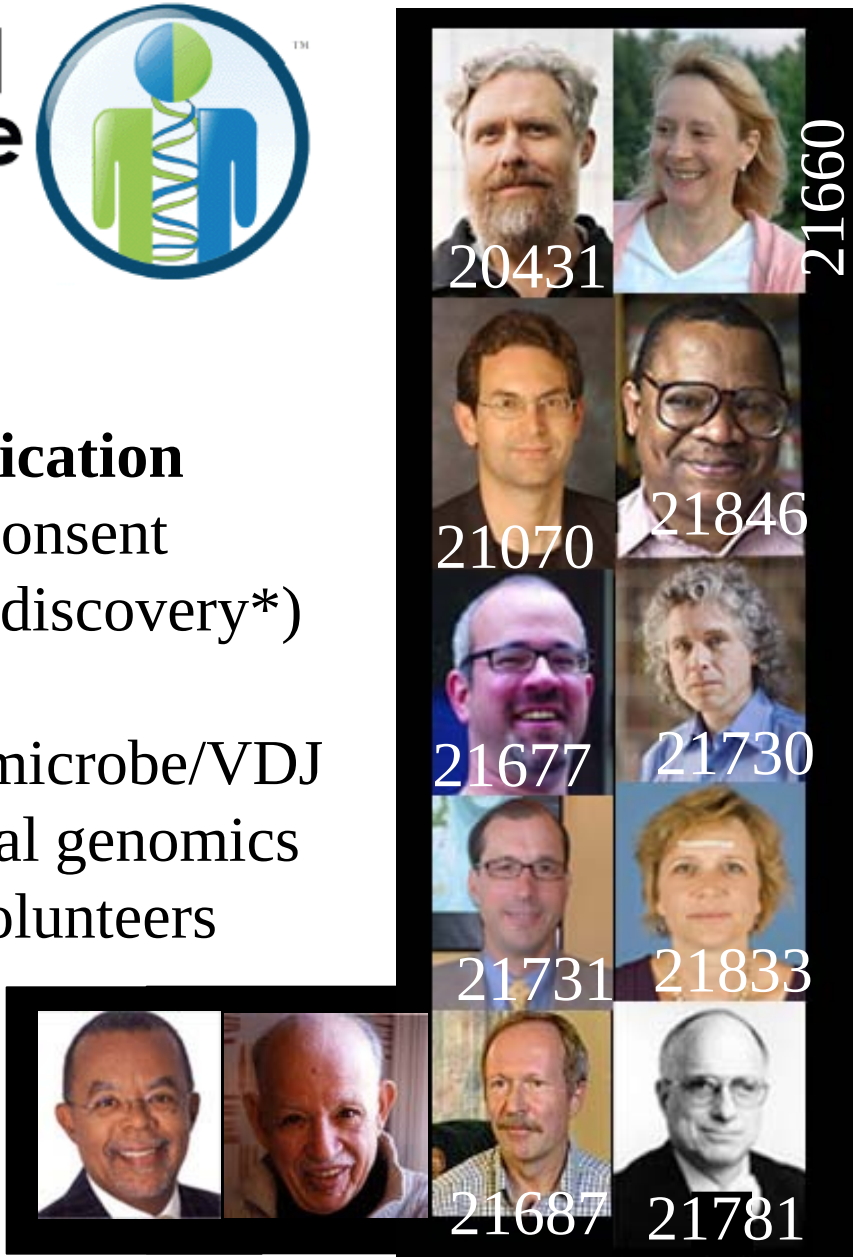
Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. Roach, et al. Science 2010

Exome Sequencing Identifies the Cause of a Mendelian Disorder. Ng et al. Nature Gen. 2010

sltrib.com/health/ci_14648608

Genes
Environments
Traits, cells

Personal Genome Project



1) **First/only open access data**
2) **Avoid over-promising on de-identification**
3) **100% on Exam** to assure informed consent
(*Educate pre-consent rather than post-discovery*)
4) **Genome sequence** + epigenome
5) **Multi-traits**: images, iPS-etc.RNA, microbe/VDJ
6) **Cells available** for personal functional genomics
7) **IRB approval** for 100,000 diverse volunteers
501(c)(3)

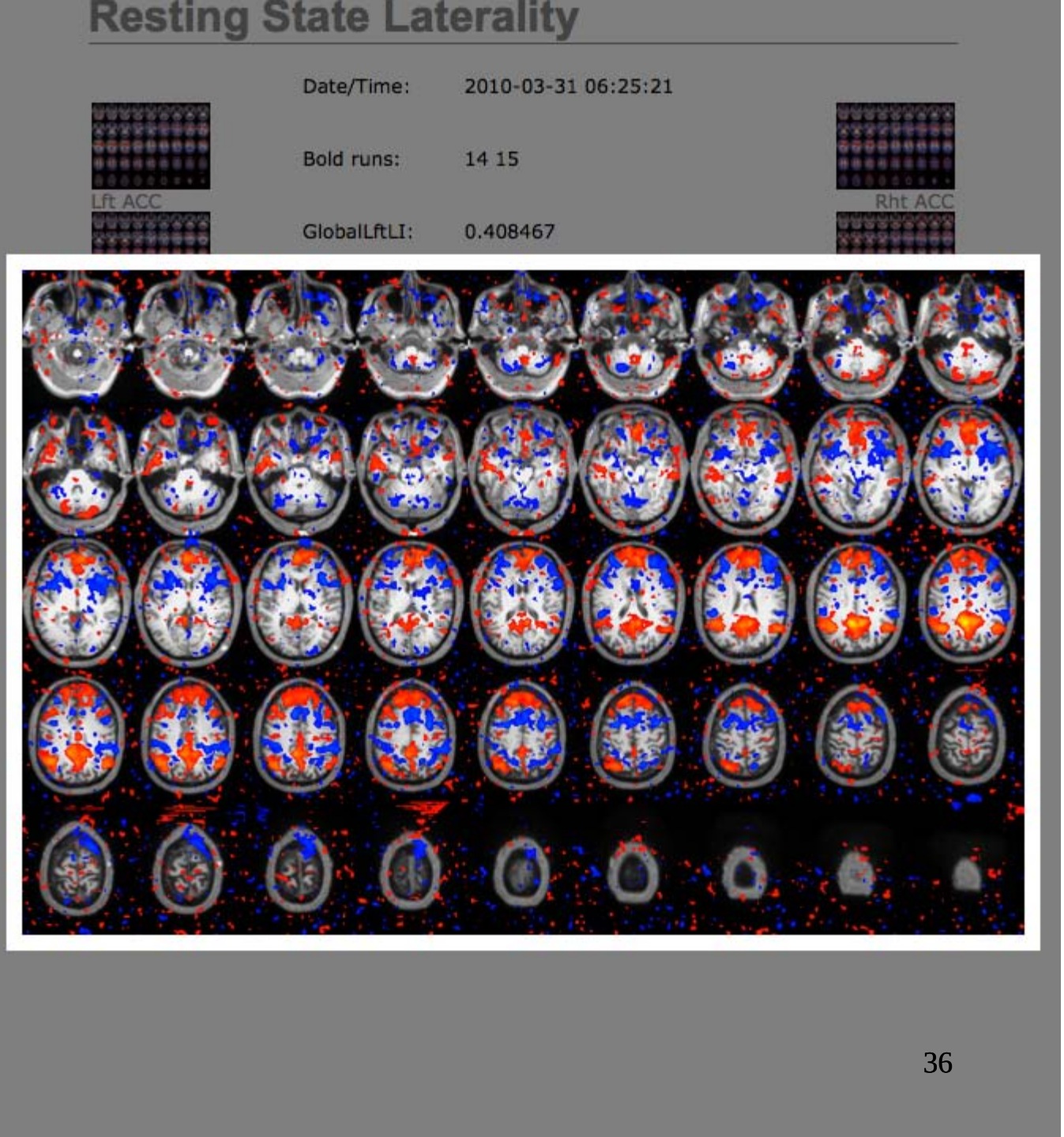
16,000 so far

PGP#1
fMRI

Bruckner

Behavioral & cognitive tests

Nakayama



US clinics quietly embrace whole-genome sequencing

Nature 14-Sep-2010

"If one hospital is doing it, you can be sure others will start, because patients will vote with their feet,"
-- Elizabeth Worthey
HMGC & Children's Hospital of Wisconsin

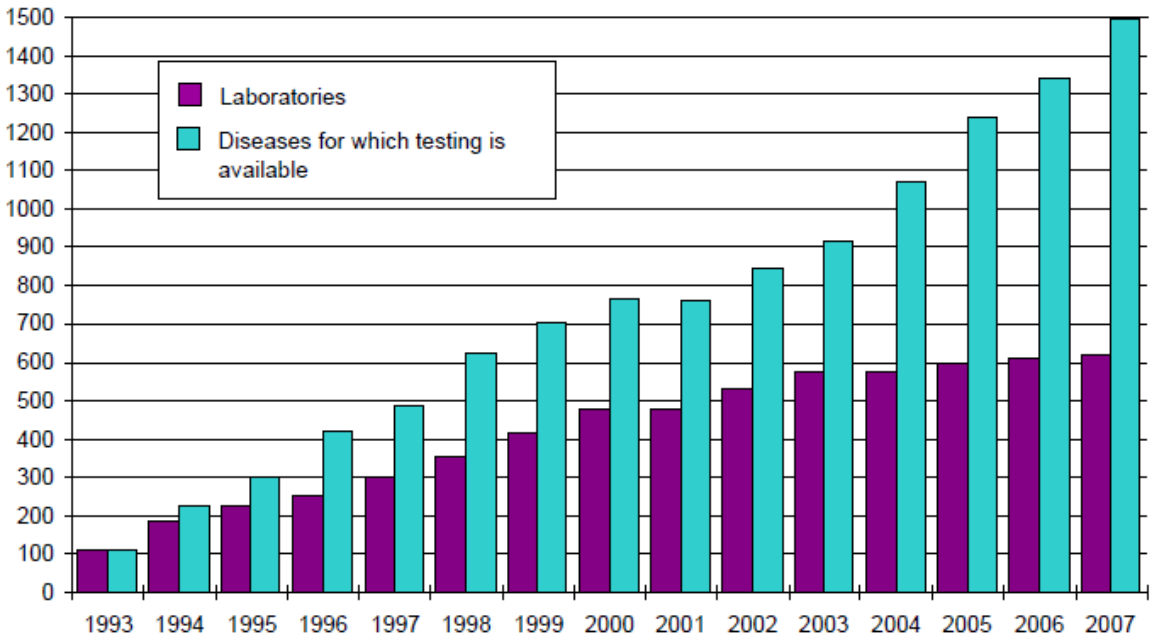
“At the age of 3, he had more than 100 separate surgeries …On the basis of [sequencing], the physician recommended a bone-marrow transplant in June 2010. By mid-July, the child was eating his first meal.”

Medical Genomics: Individually rare collectively common 10%

1540 genes are highly predictive & medically actionable (inherited & cancer) at ~$2K per gene.
****Very few of these are on DTC SNP chips.** Why?**
PKU, Tay Sachs, Cystic Fibrosis, BRCA1/2, etc.

Pharmacogenomic drug/allele combinations: Herceptin, Iressa, ..

Also:
Ancestry, Forensics, Social Networking, Education, Research



Evidence.personalgenomes.org: 32 diploid full&exome: hypertrophic cardiomyopathy allele

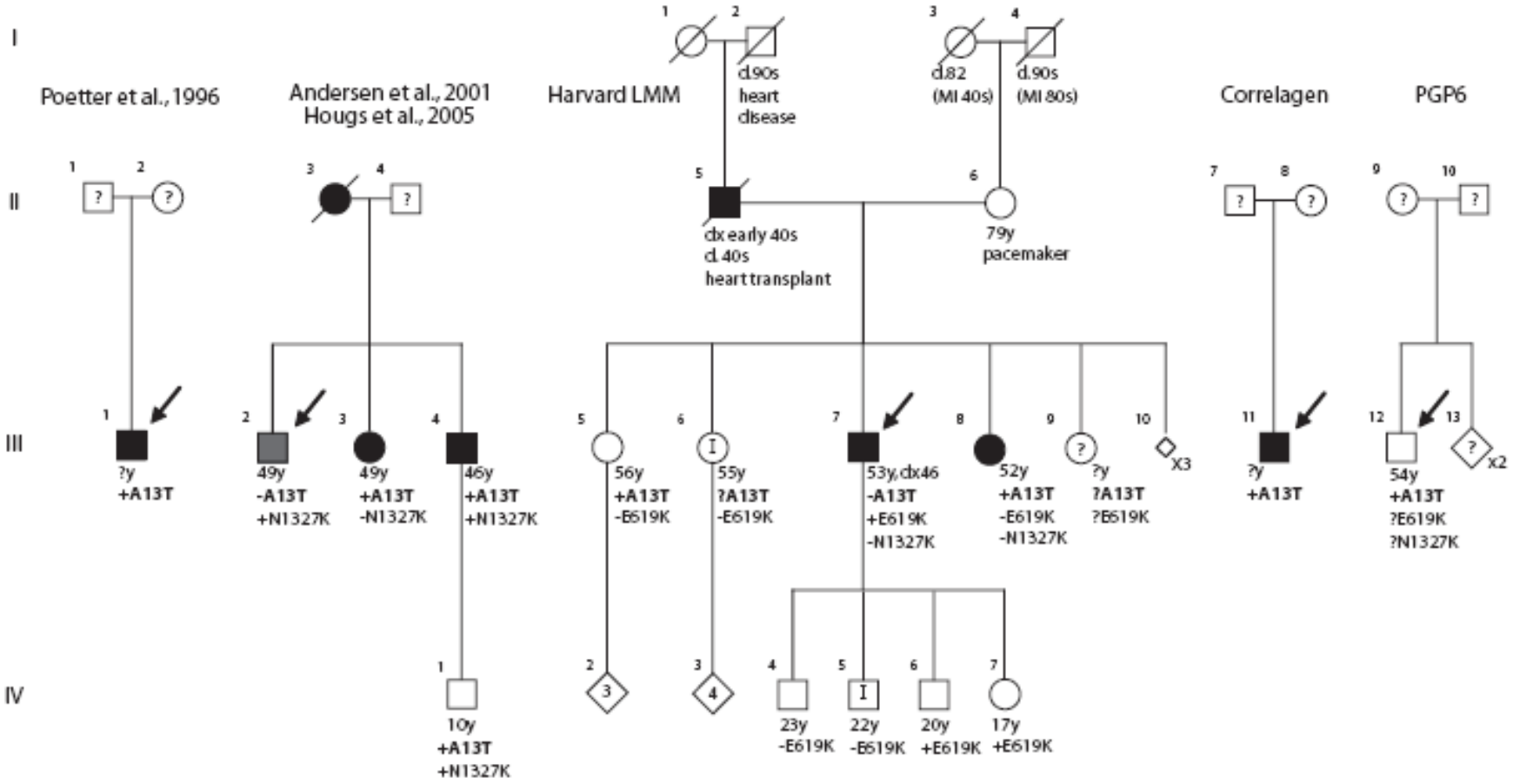
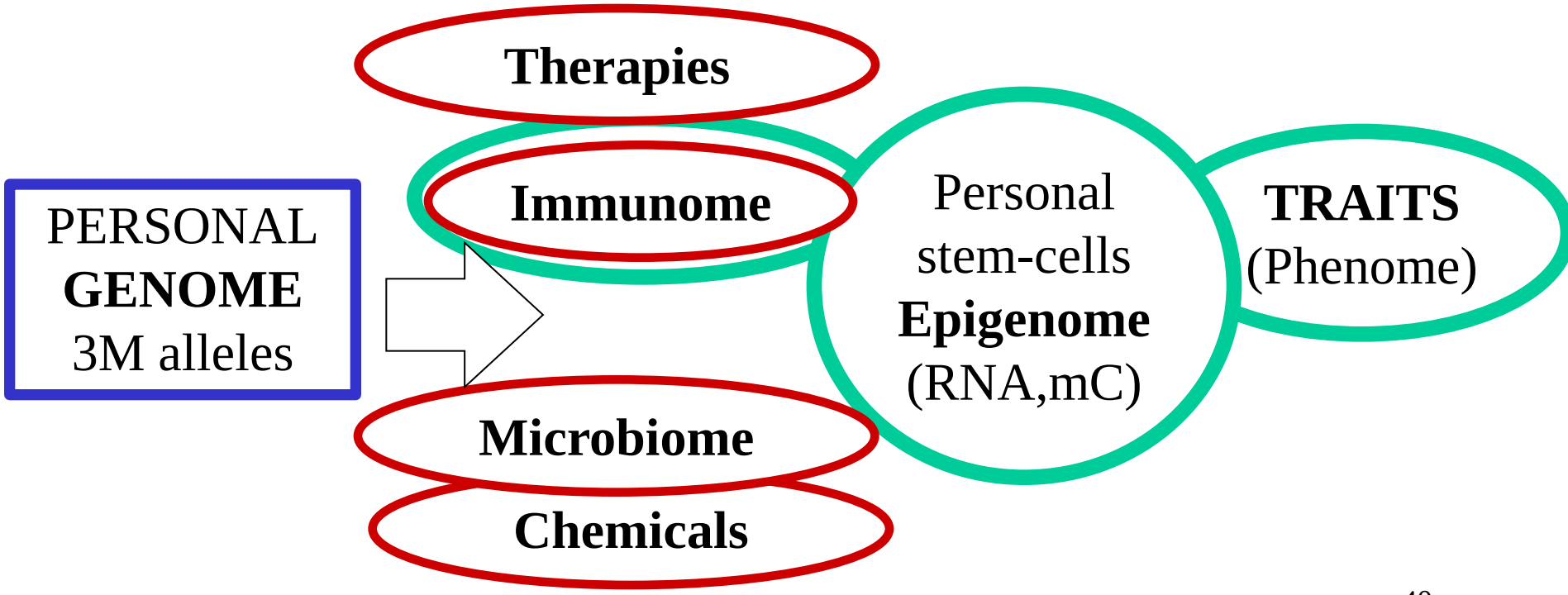


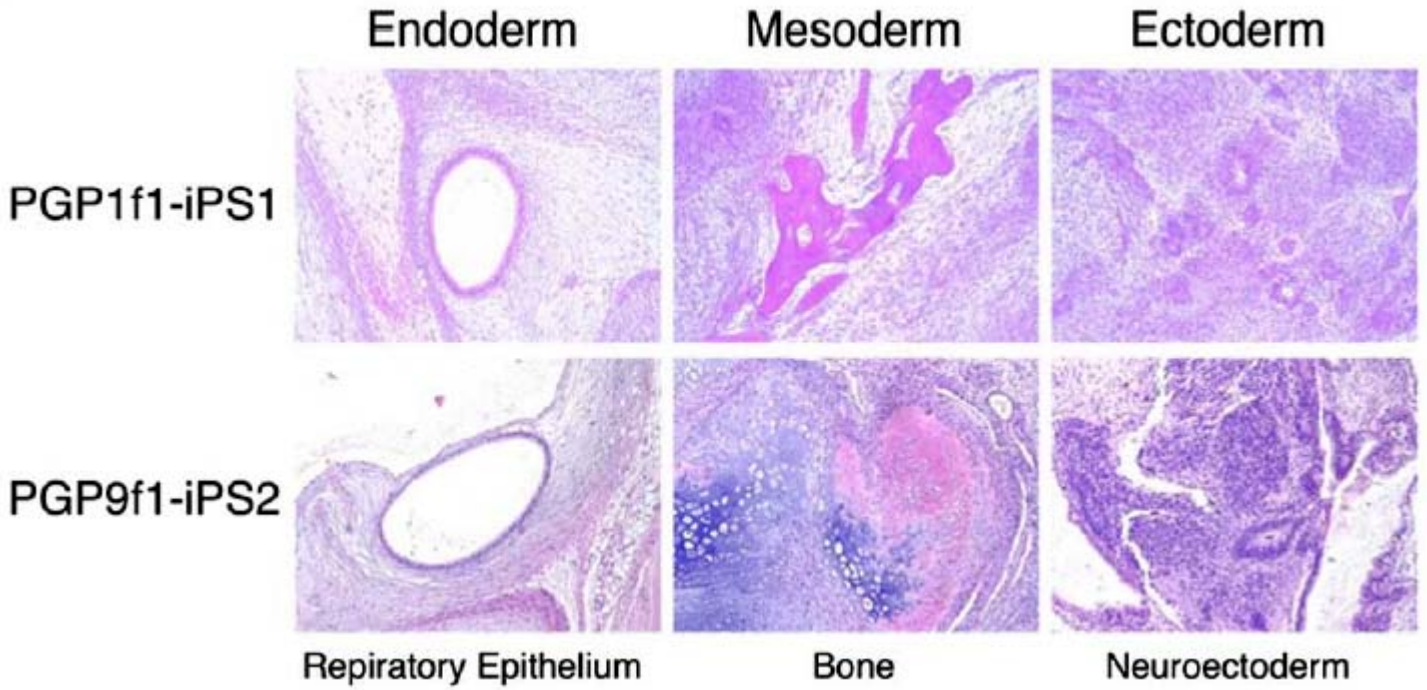
Figure 5.4: **Summary of MYL2 A13T findings from clinical testing.** This variant has been seen in two families and three sporadic cases based on our enquiries to clinical testing laboratories that sequence the MYL2 gene. In one family MYL2 A13T has been seen concurrently with MYBPC3 E619K and in the other family with MYH7 N1327K.

Genomes Environments Traits

One in a life-time genome + yearly (to daily) tests
Bio-weather map : Allergens, Microbes, Viruses
PersonalGenomes.org

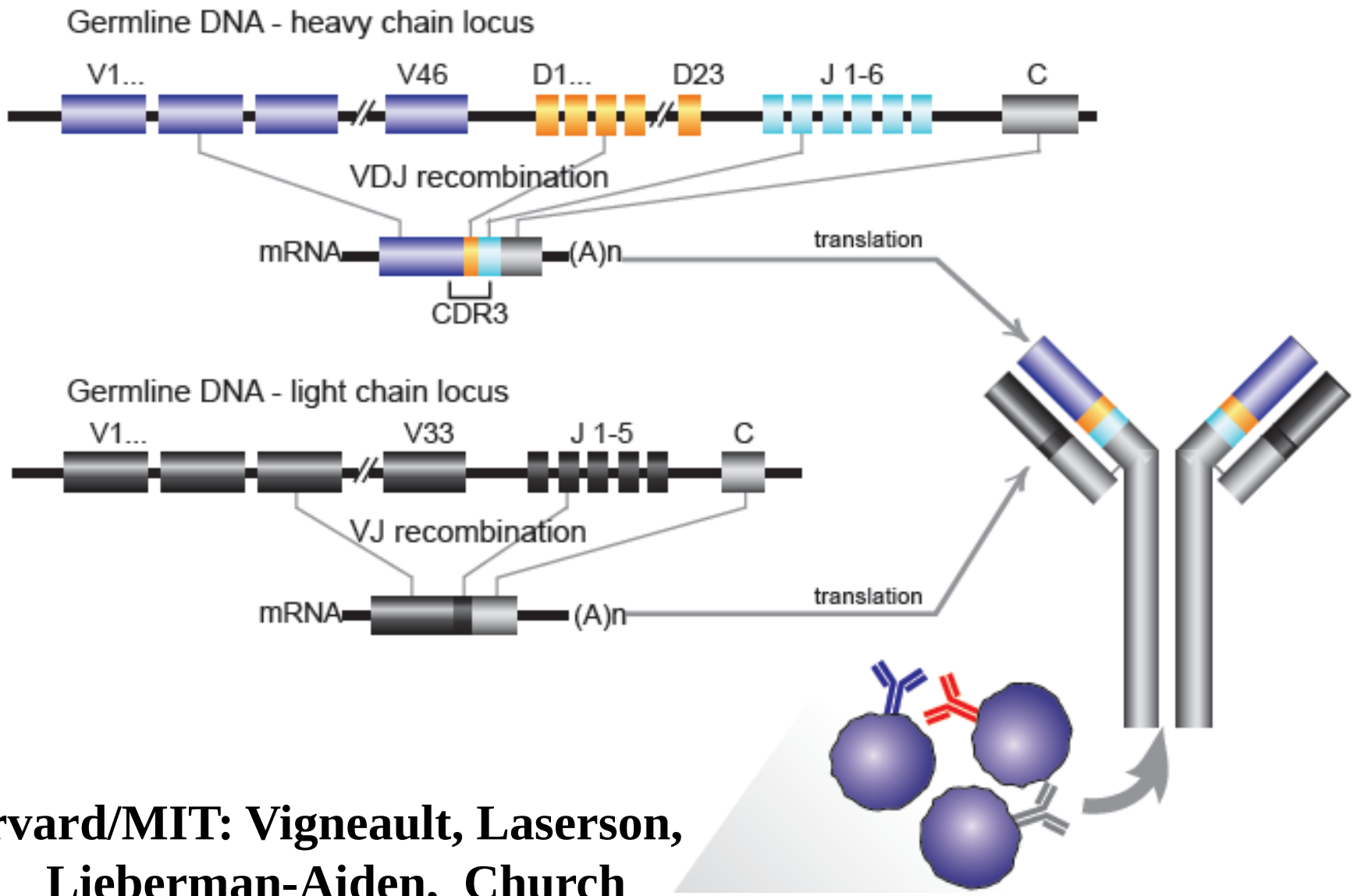


PGP#1 & #9 skin to stem cells to ...



Lee J, Park IH, Gao Y, Li JB, Li Z, Daley G, Zhang K, Church GM (2009) A Robust Approach to Identifying Tissue-specific Gene Expression Regulatory Variants Using Personalized Human Induced Pluripotent Stem Cells. PLoS Genetics Nov 2009

PGP Vaccination Immunome



Harvard/MIT: Vigneault, Laserson, Lieberman-Aiden, Church

Roche: Egholm, Simen

PGP Time Series Vaccine Experiment

Tracking human dynamic response to vaccination to 11 strains:

Hepatitis A+B, Flu A/Brisbane/59/2007 (H1N1)-like, 10/2007 (H3N2)-like, B/Florida/4/2006-like virus

Polio, Yellow fever

Meningococcus

Typhoid, Tetanus

Diptheria, Pertussis

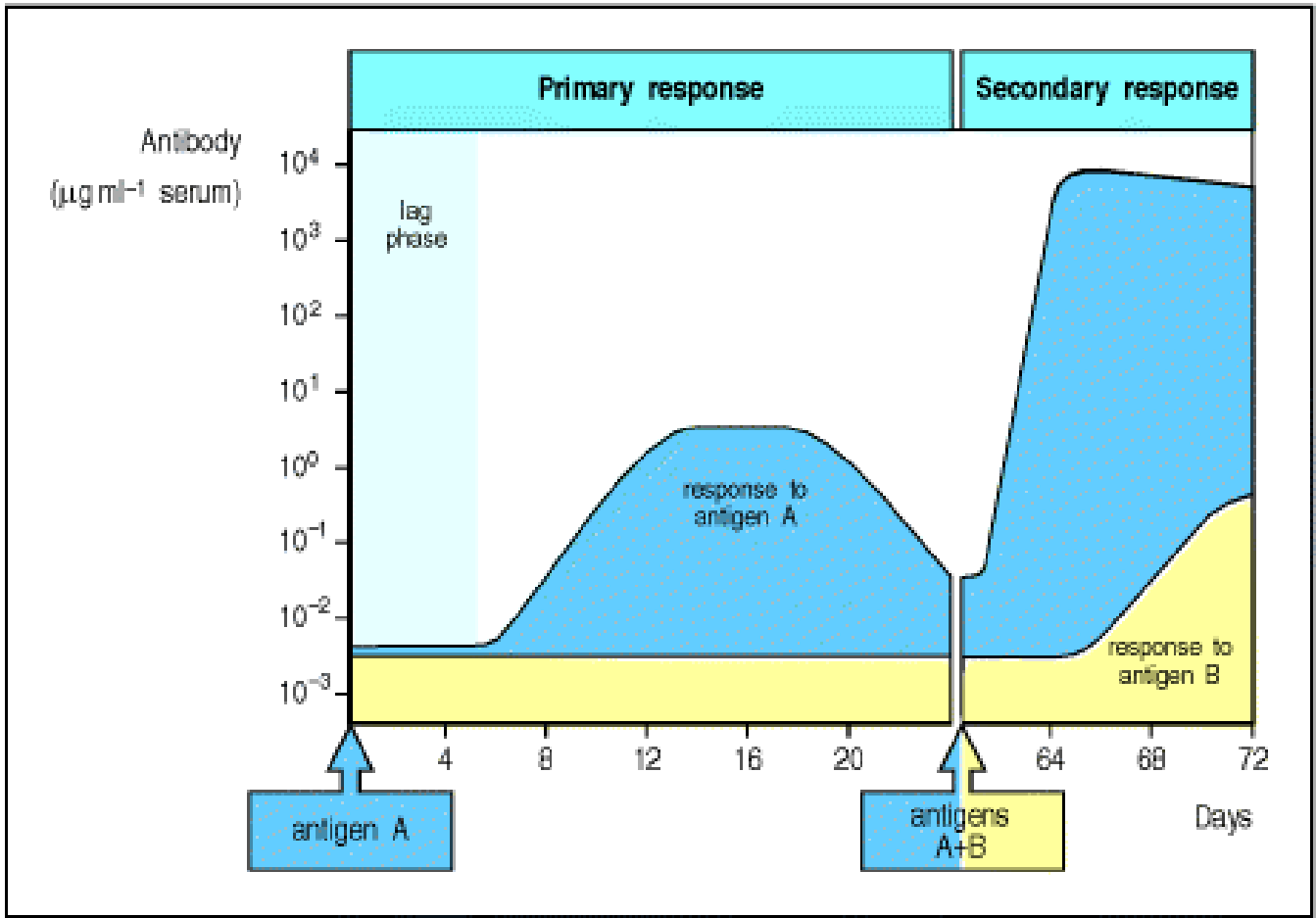
Collect samples at

-14d, 0d,

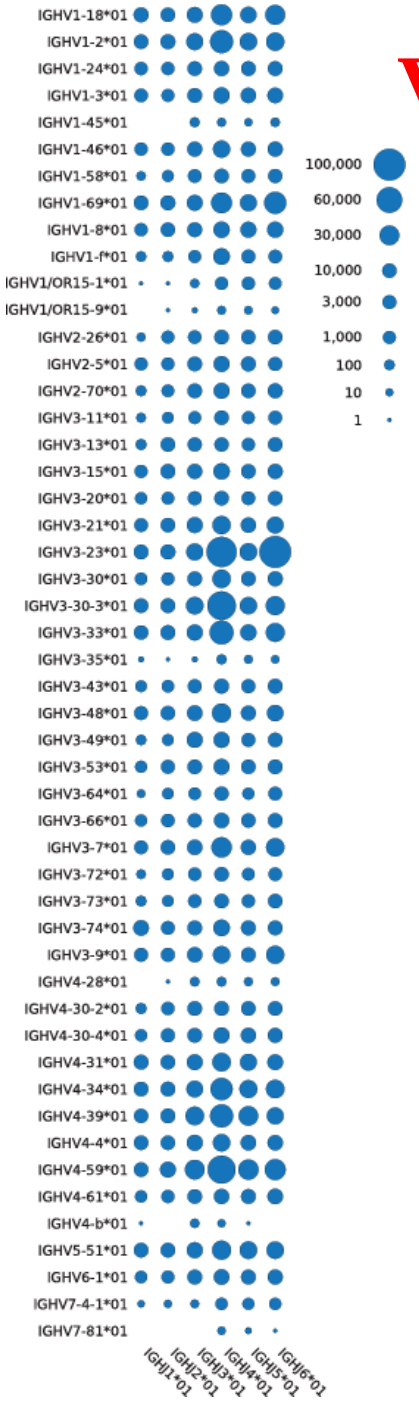
+1d, +3d,

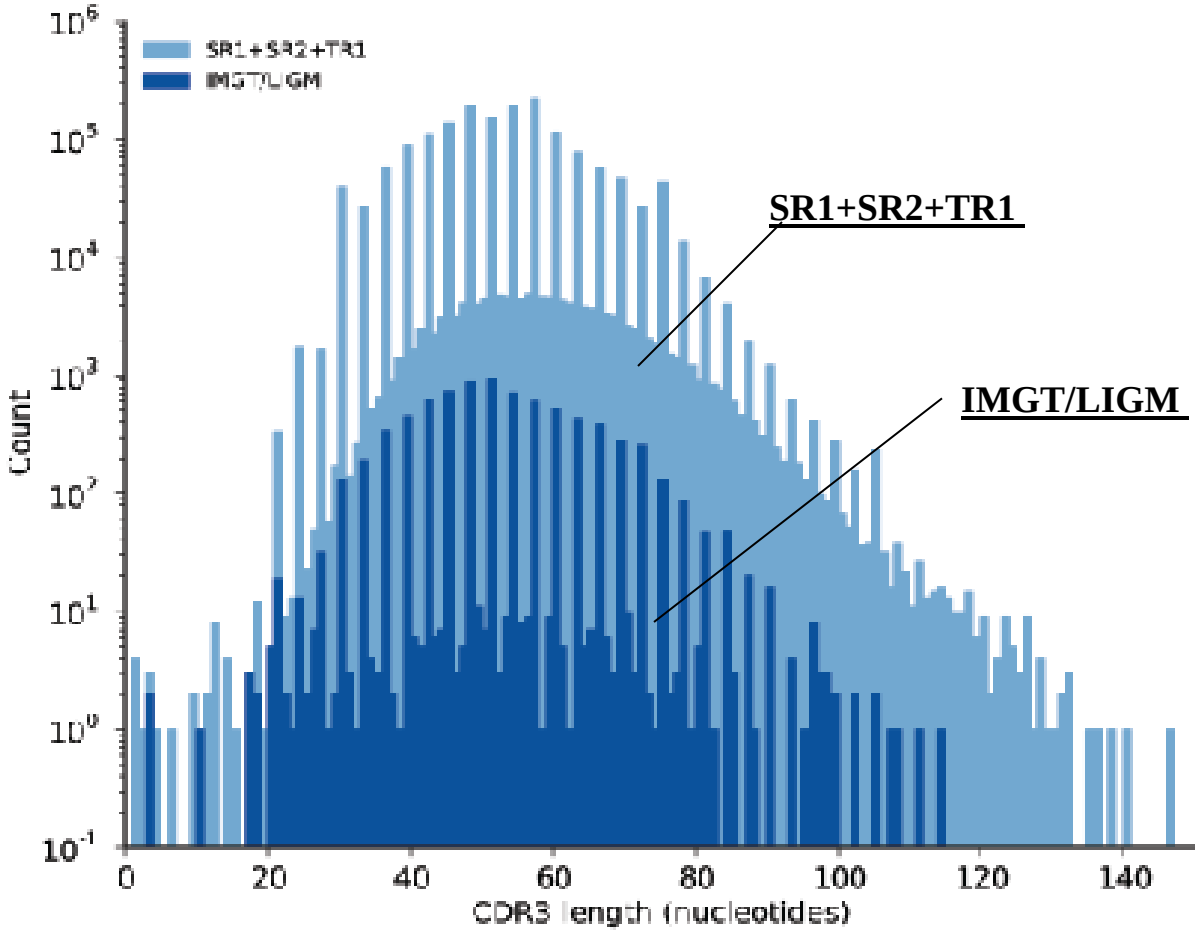
+7d, +14d,

+21d, +28d



V D J usage – CDR3 size distribution



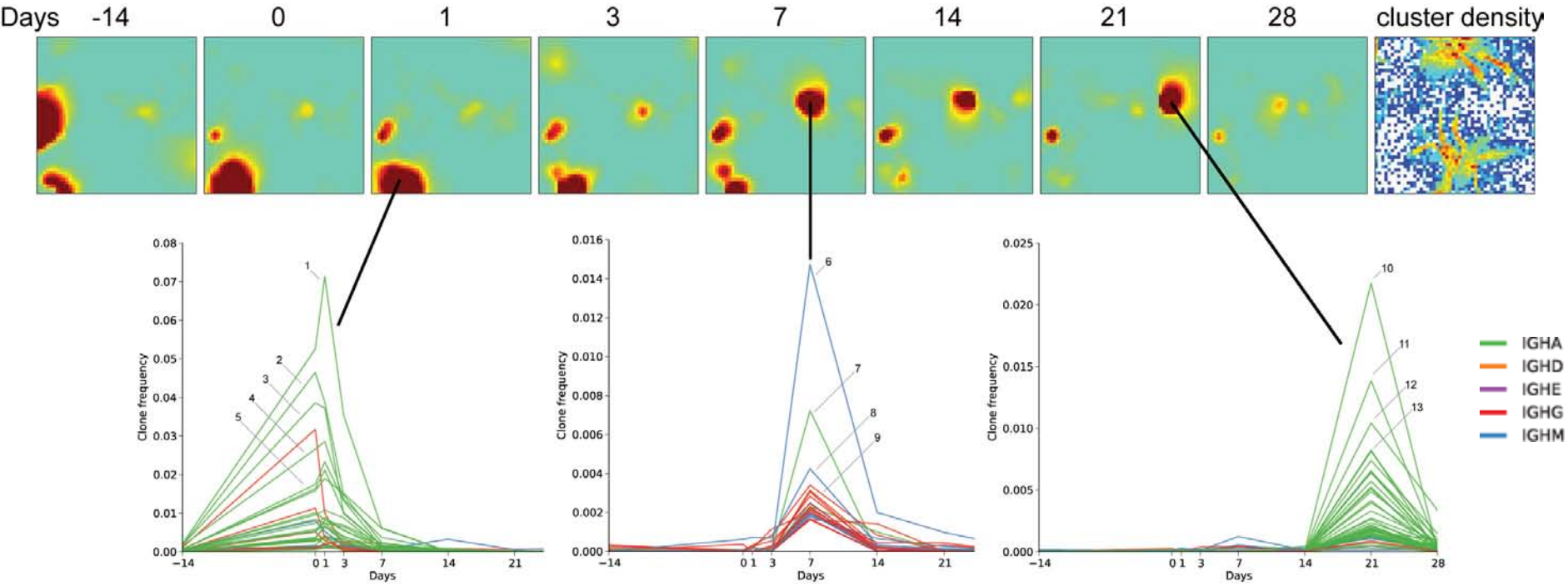


SR1+SR2+TR1

IMGT/LIGM

PGP Vaccination Immunome

Self Organizing Map (SOM) clustering



Reading & Writing Genomes Goals

- **2nd Generation BI/O: Reading & Writing**
  Engineer cancer- & virus-resistant genomes

- **Personal Genomes – Integration tasks**
  - Personal **G**enomes: **E**nvironments, **T**raits,
  - Stem cells, Microbiome/Immunome
  - Synthesis for Causality (CEGS)