Bioinformatics is Like A Band-aid

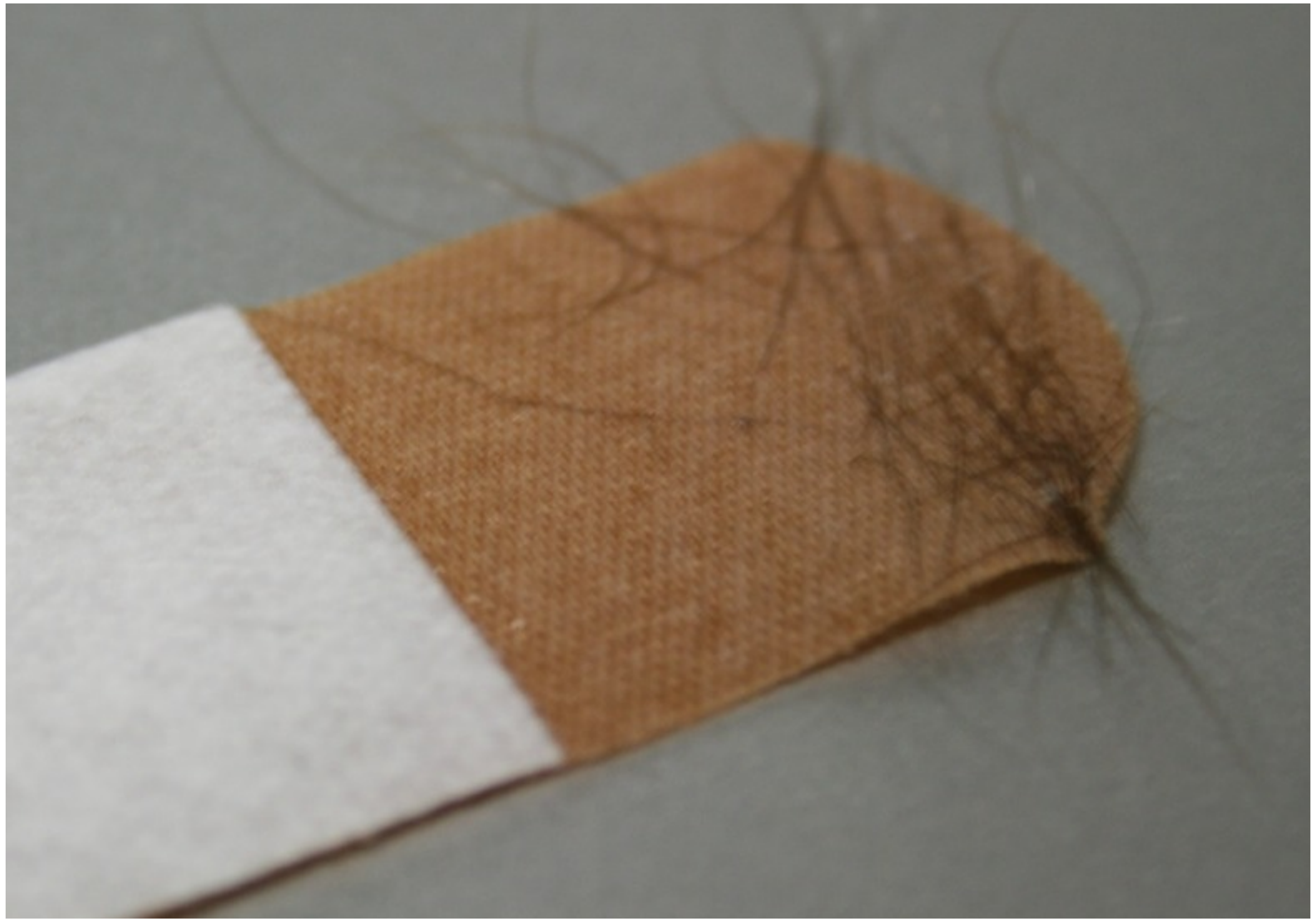
A. Malcolm Campbell

DAVIDSON

University of Georgia
January 11, 2013

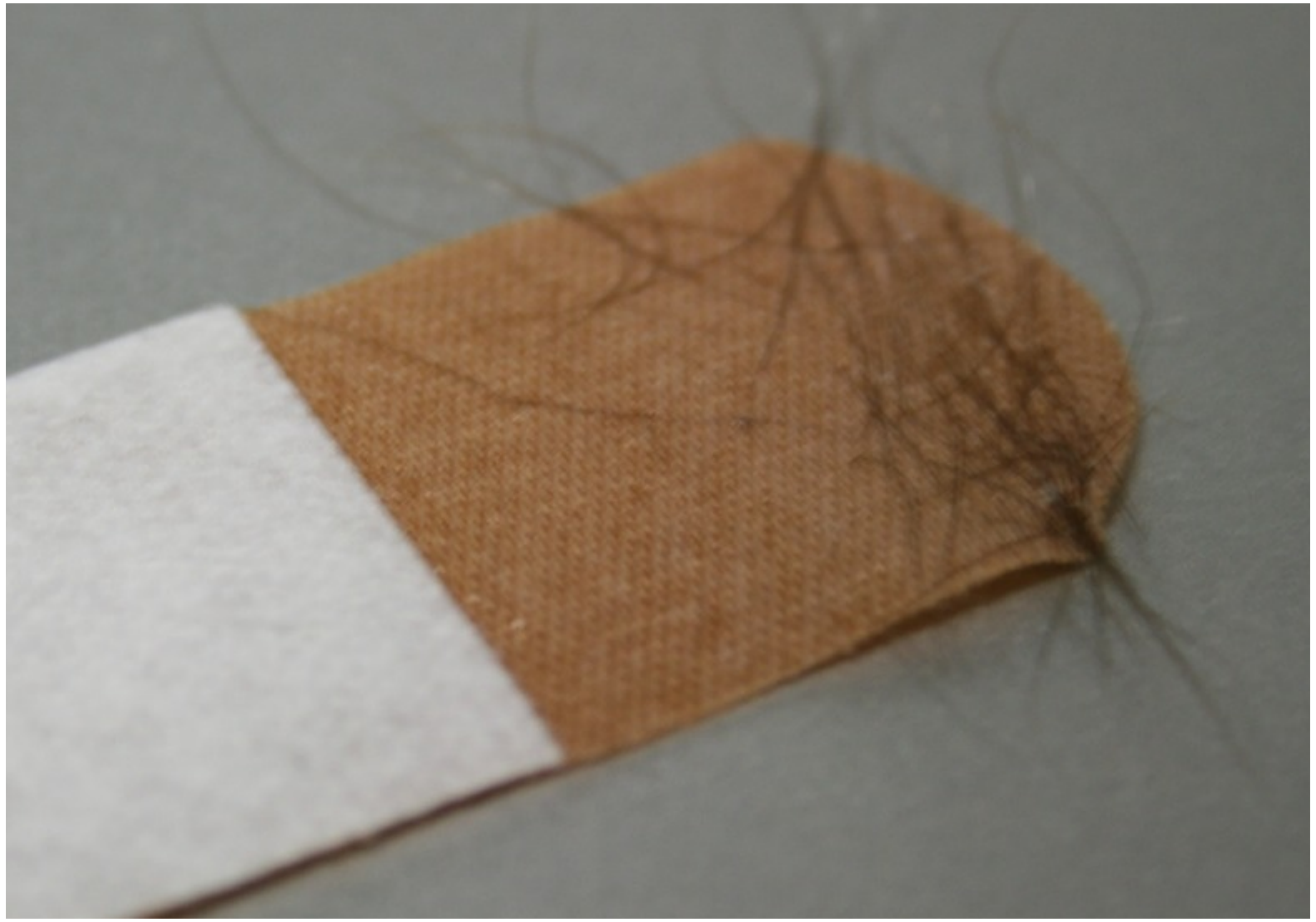
Two Ways to Remove Band-aids

Two Ways to Remove Band-aids

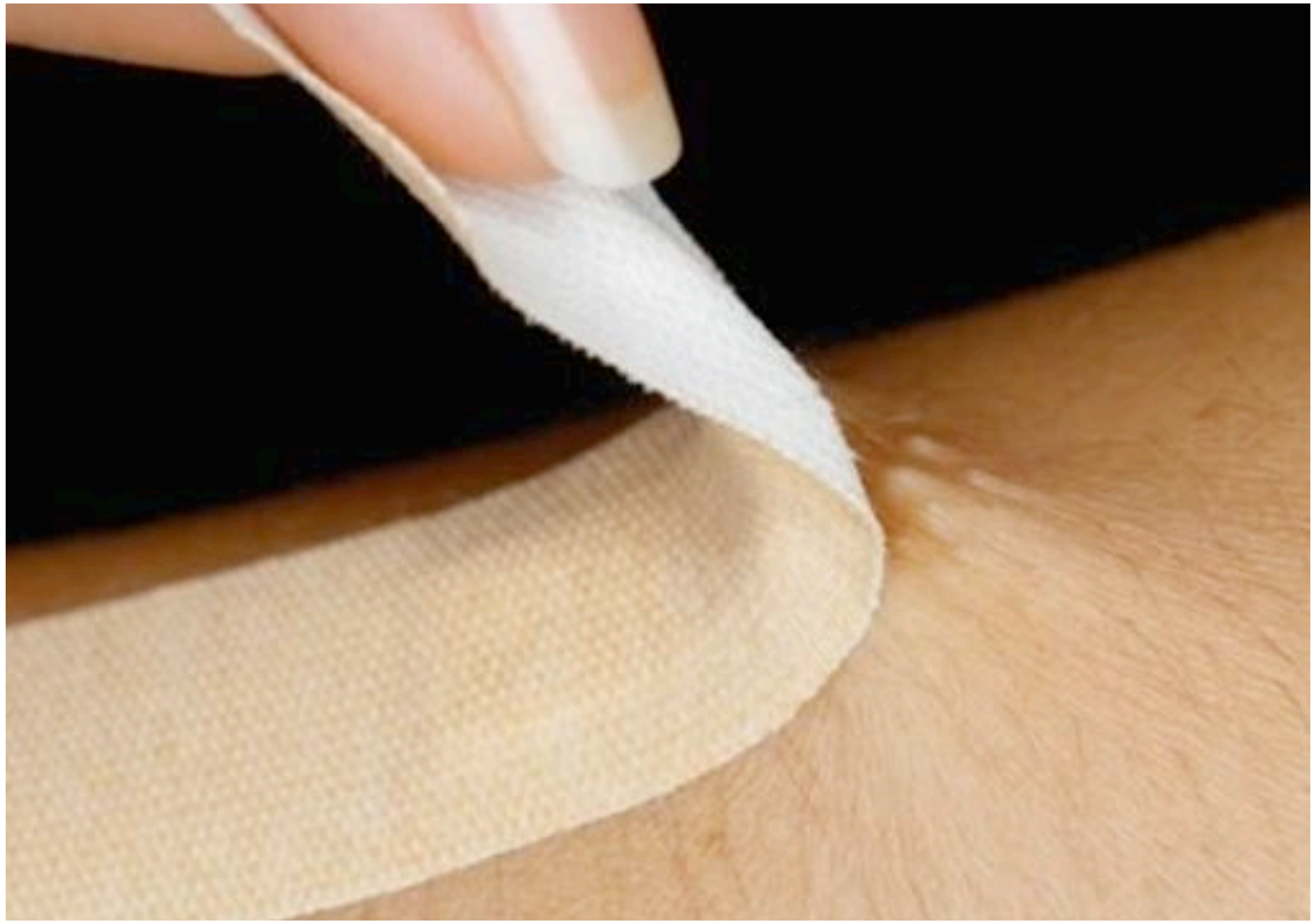


fast yank

Two Ways to Remove Band-aids



fast yank



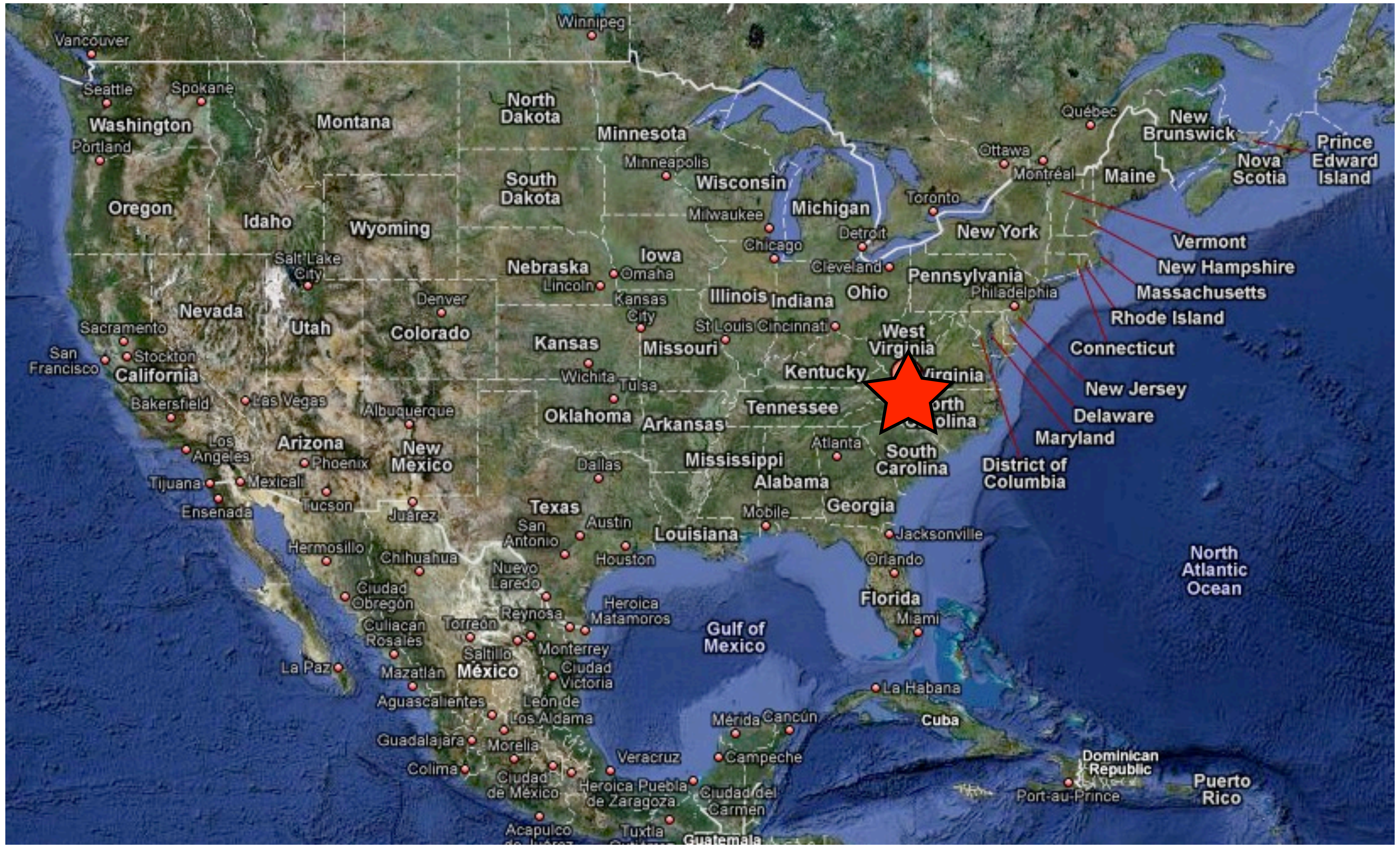
slow pull

Outline of Presentation

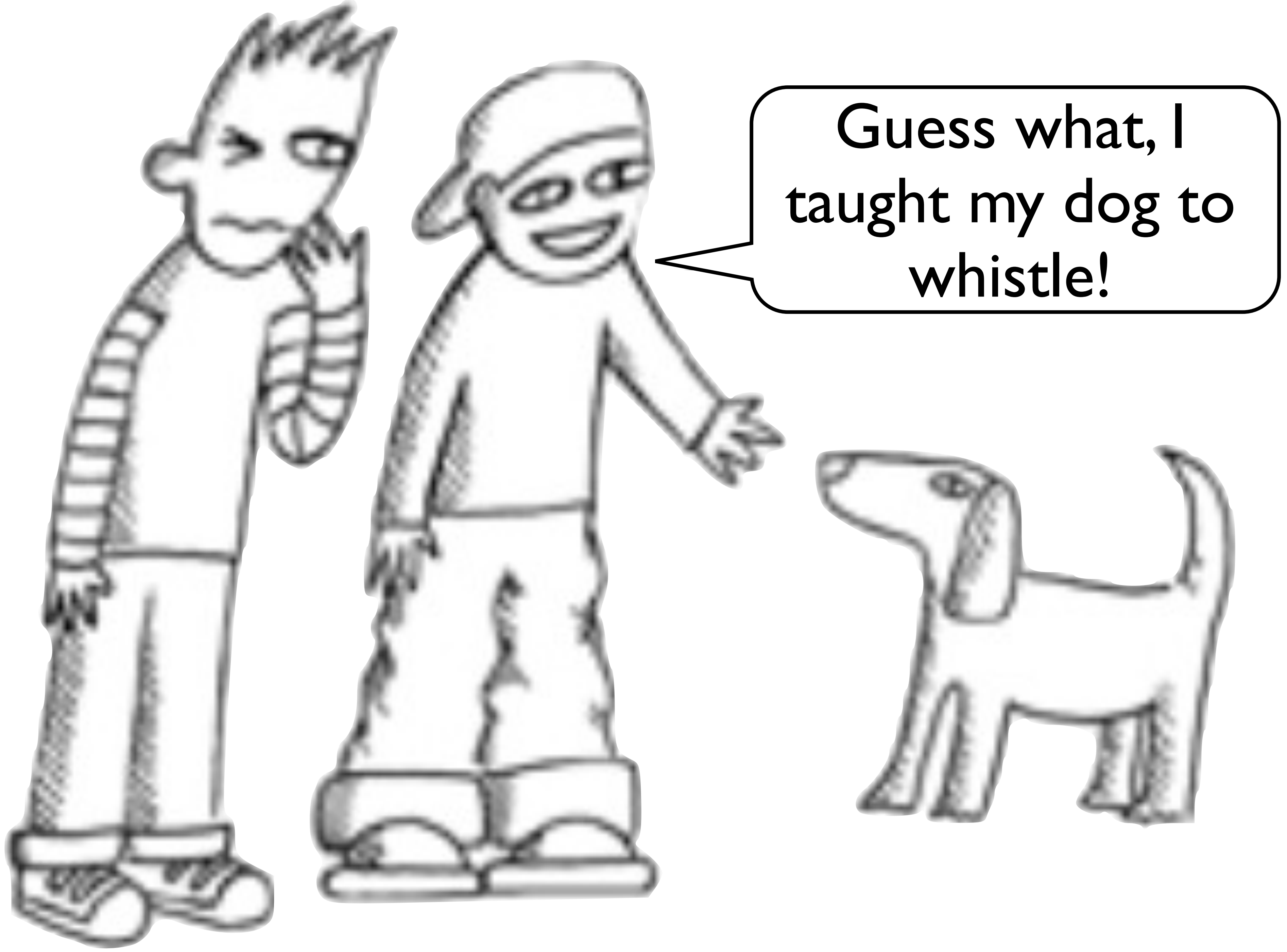
1. My Definition of Bioinformatics
2. Jump Right In (rip it off)
3. Gradual Introduction (pull slowly)
4. Student Outcomes and Data
5. Discussion and Exchange of Ideas

Davidson College

Davidson, NC USA
Liberal Arts College
2,000 Undergraduates
48 States
36 Countries

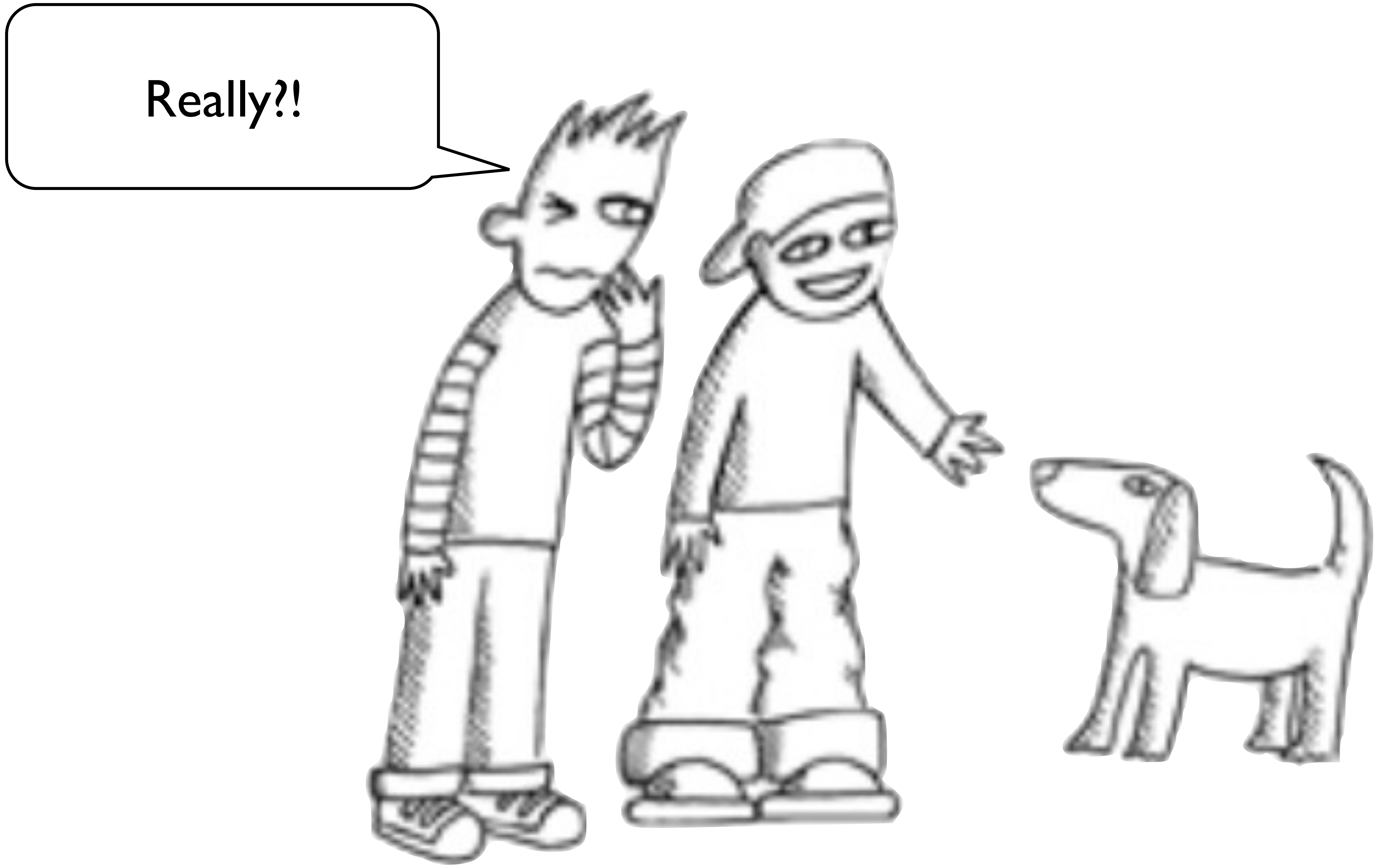


Teaching *vs* Learning

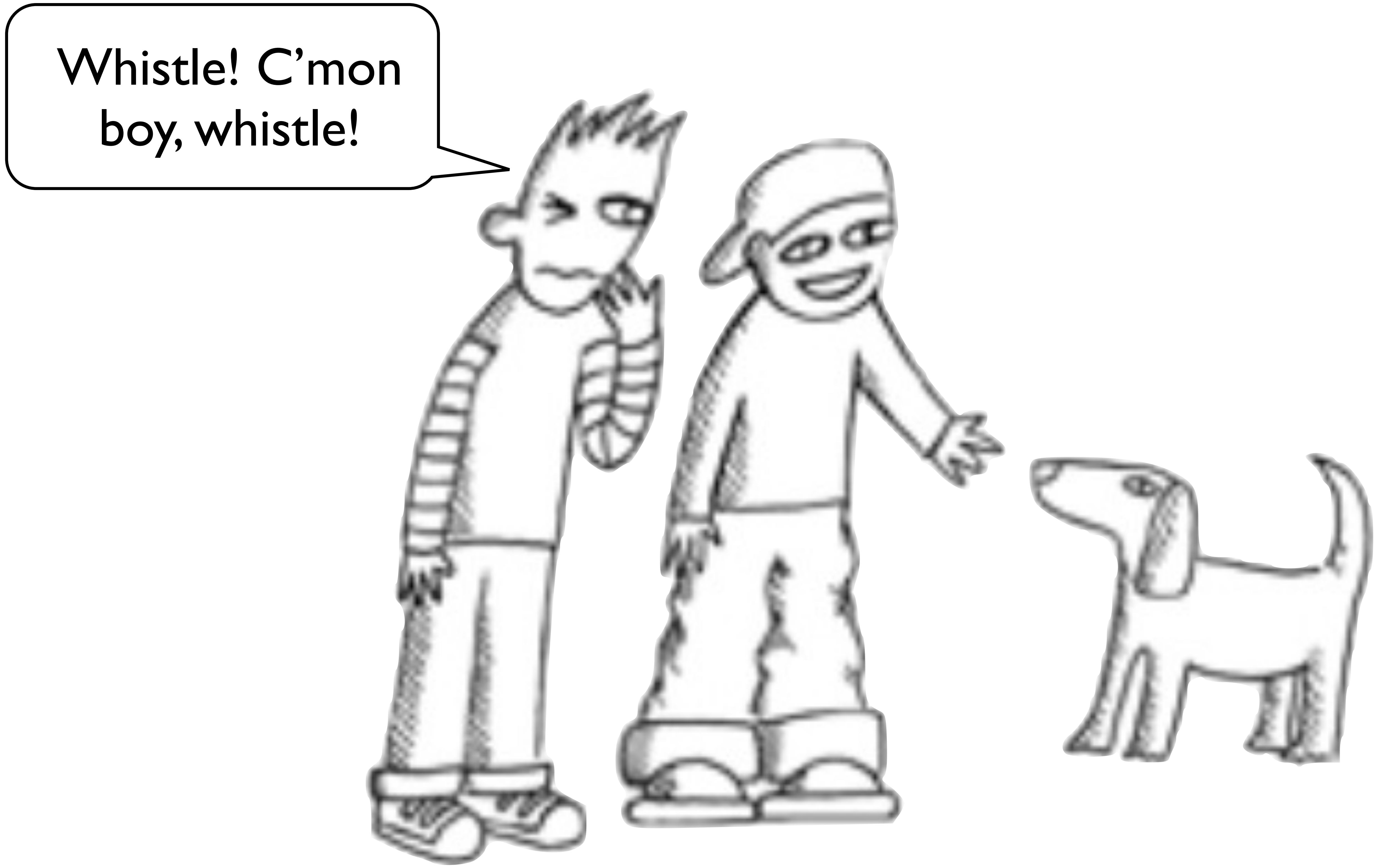


http://thegoodpeople.se/blog/?p=208

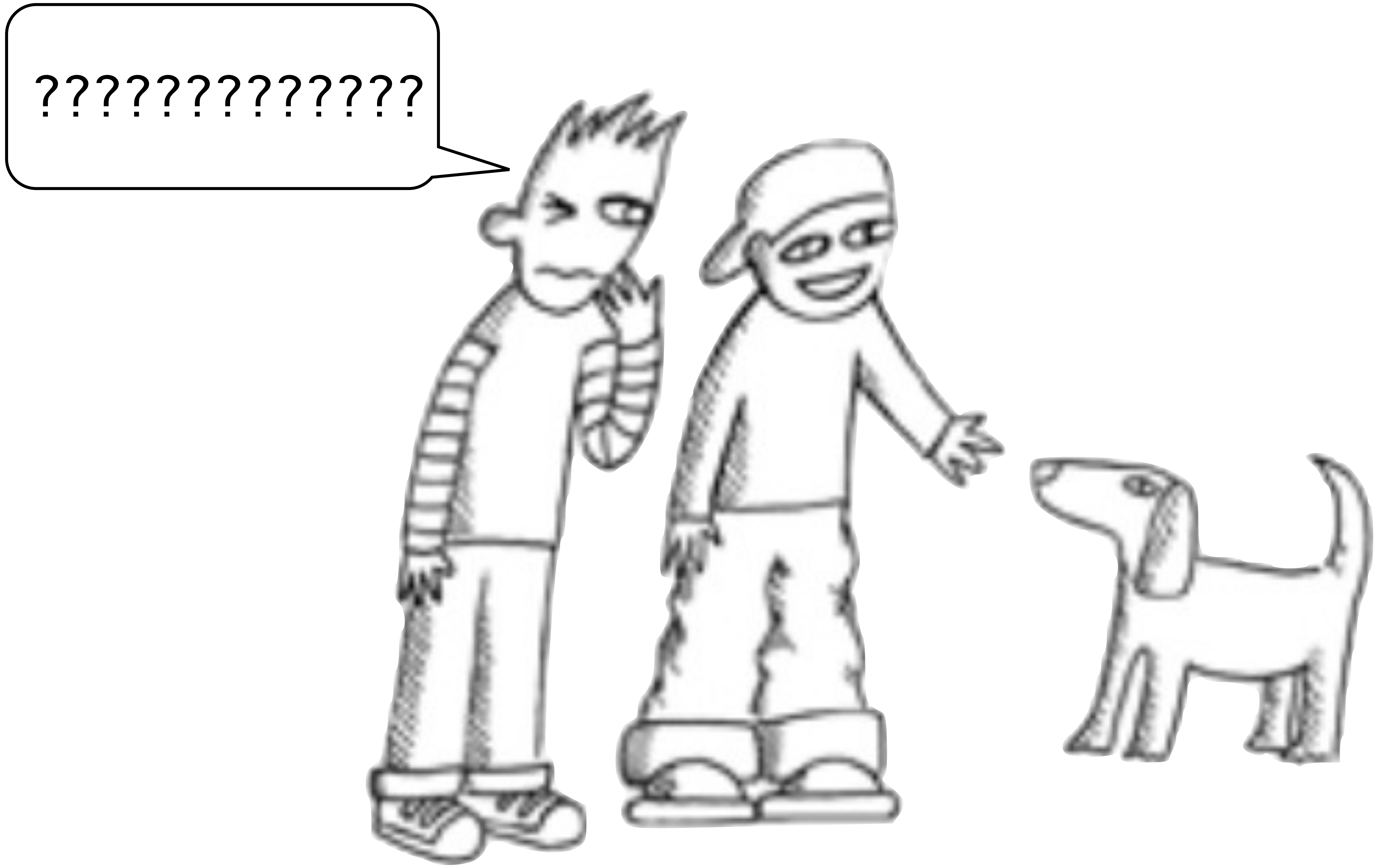
Teaching *vs* Learning



Teaching *vs* Learning



Teaching *vs* Learning



Teaching *vs* Learning

Teaching *vs* Learning

Genomics, Proteomics & Bioinformatics

mostly juniors and seniors
majority biology majors
+
psychology, neuroscience, math

Genomics, Proteomics & Bioinformatics

Students should be able to:

- interpret real genomics papers
- use data to answer questions
- mine research databases
- connect science to real world problems

Genomics, Proteomics & Bioinformatics

Sequencing

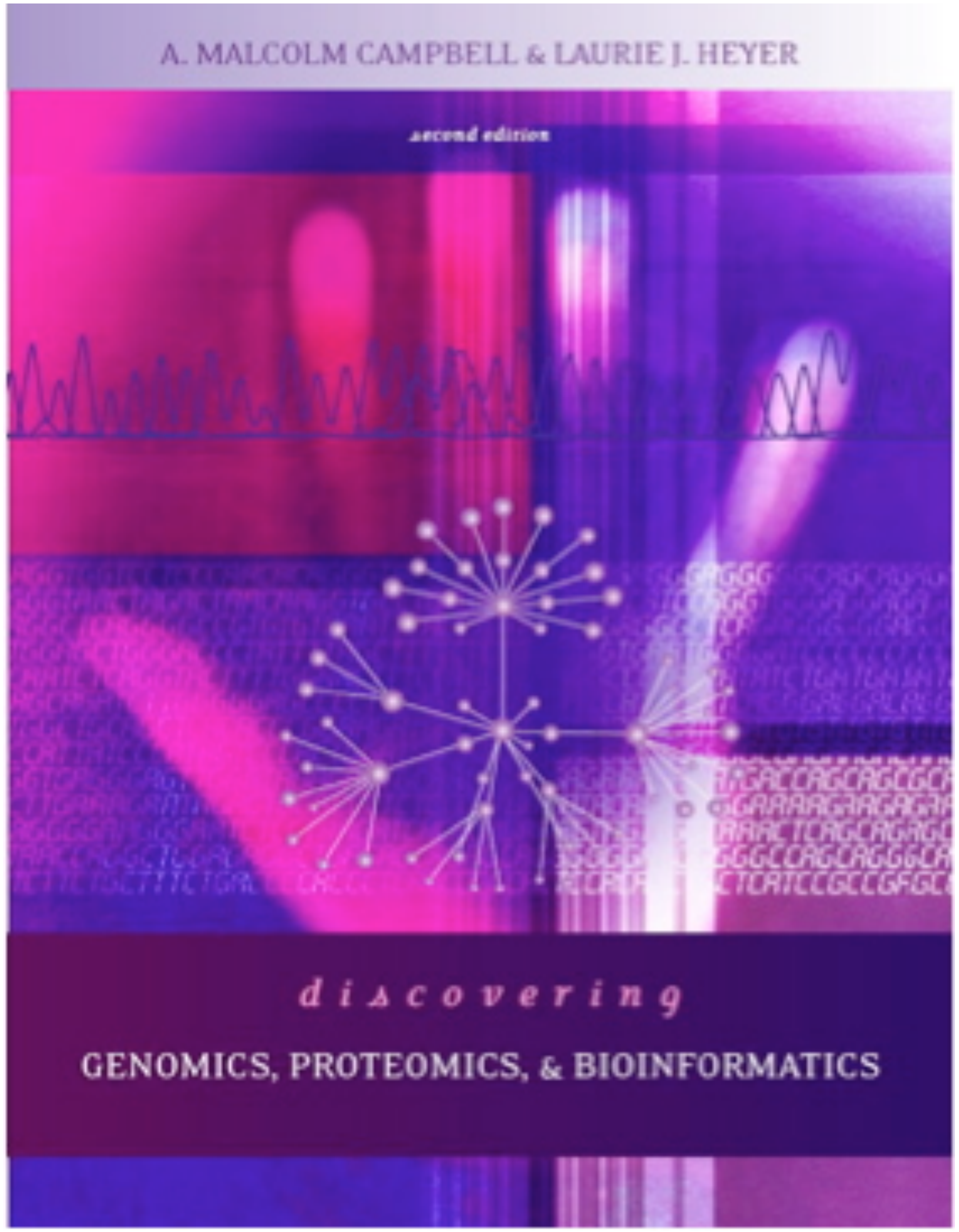
Epigenomics

Variation

Transcriptome

Proteome

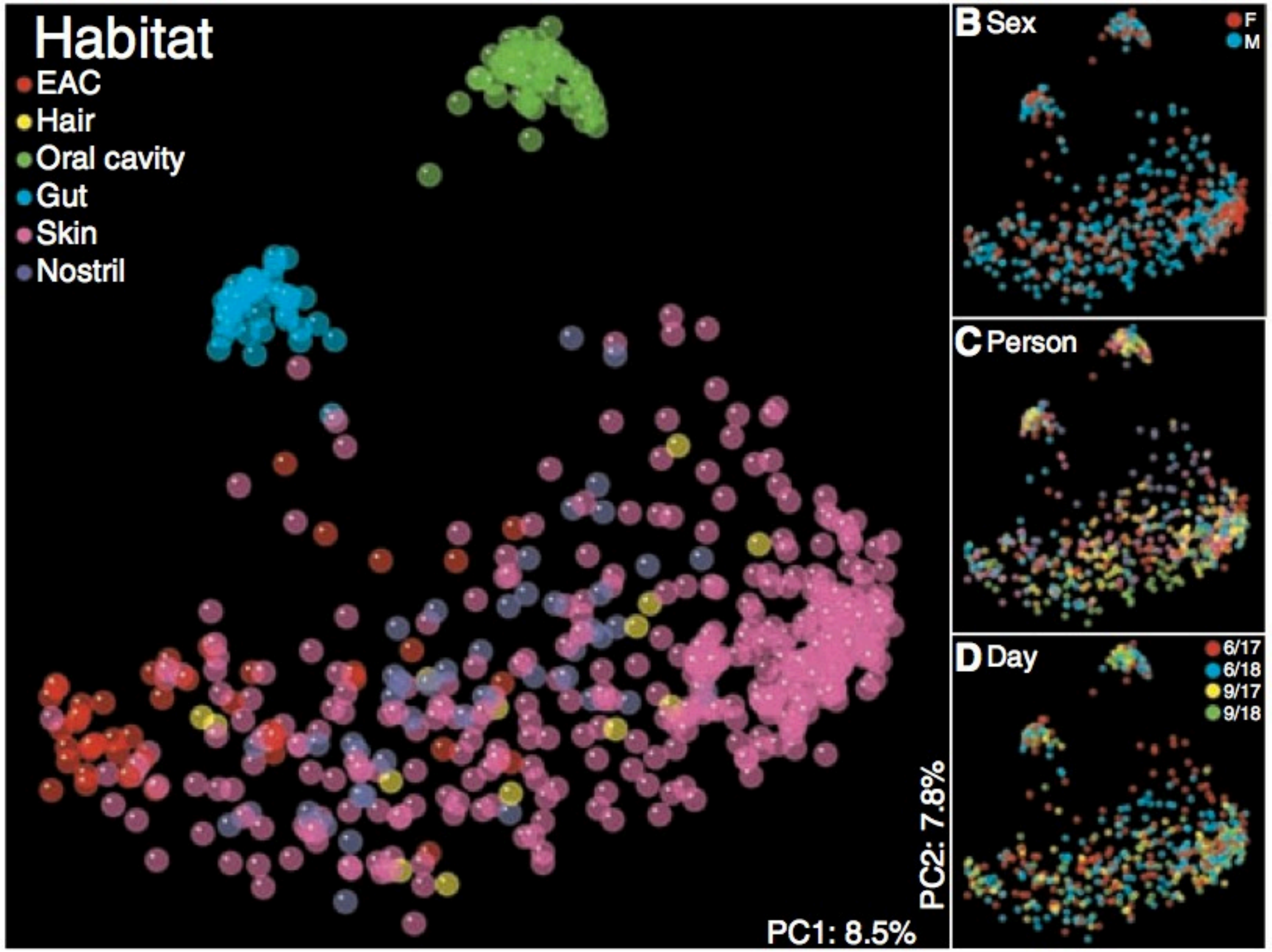
Systems Biology



Genomics, Proteomics & Bioinformatics

Read Papers

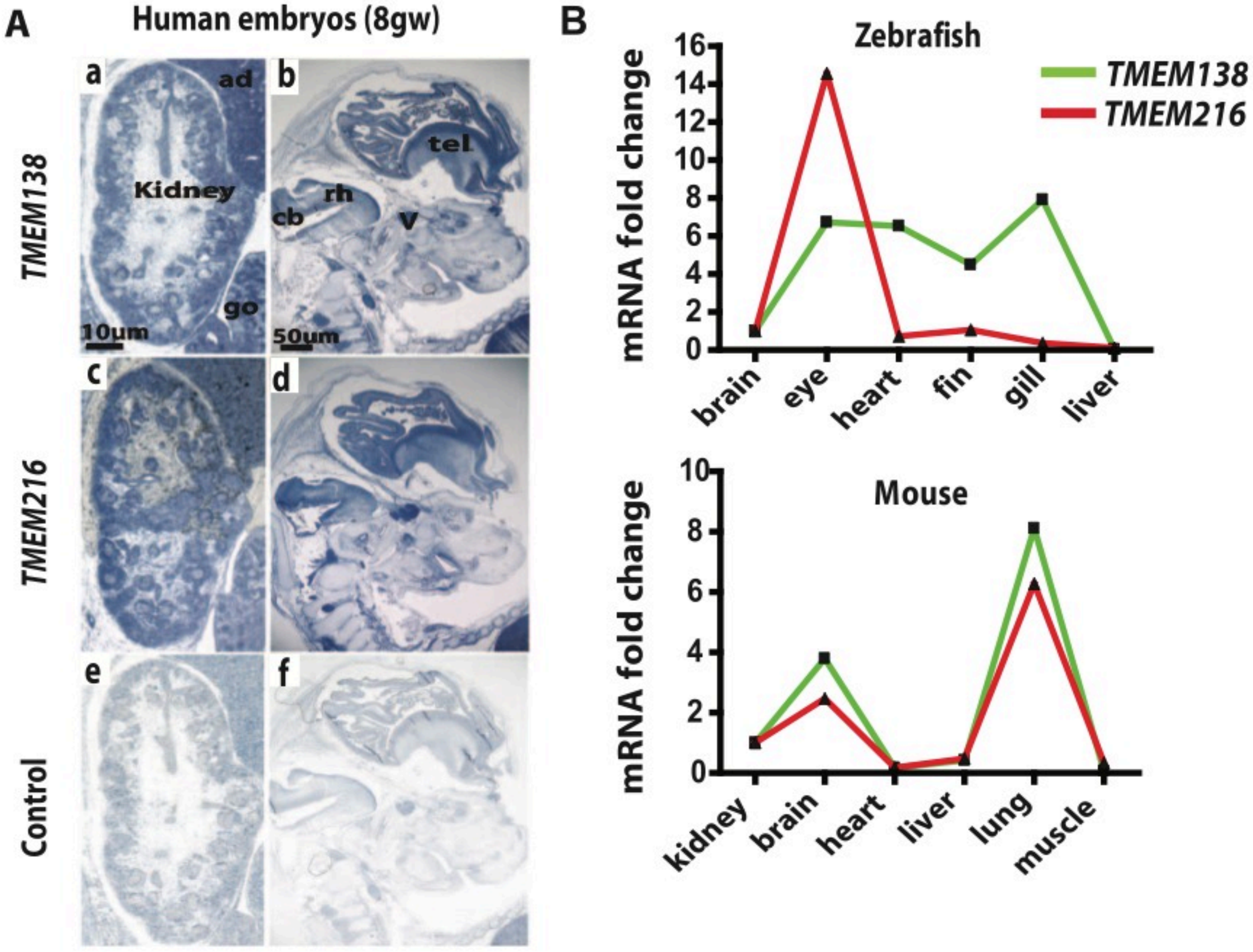
Bacterial Community Variation in Human Body Habitats Across Space and Time



Genomics, Proteomics & Bioinformatics

Web Pages

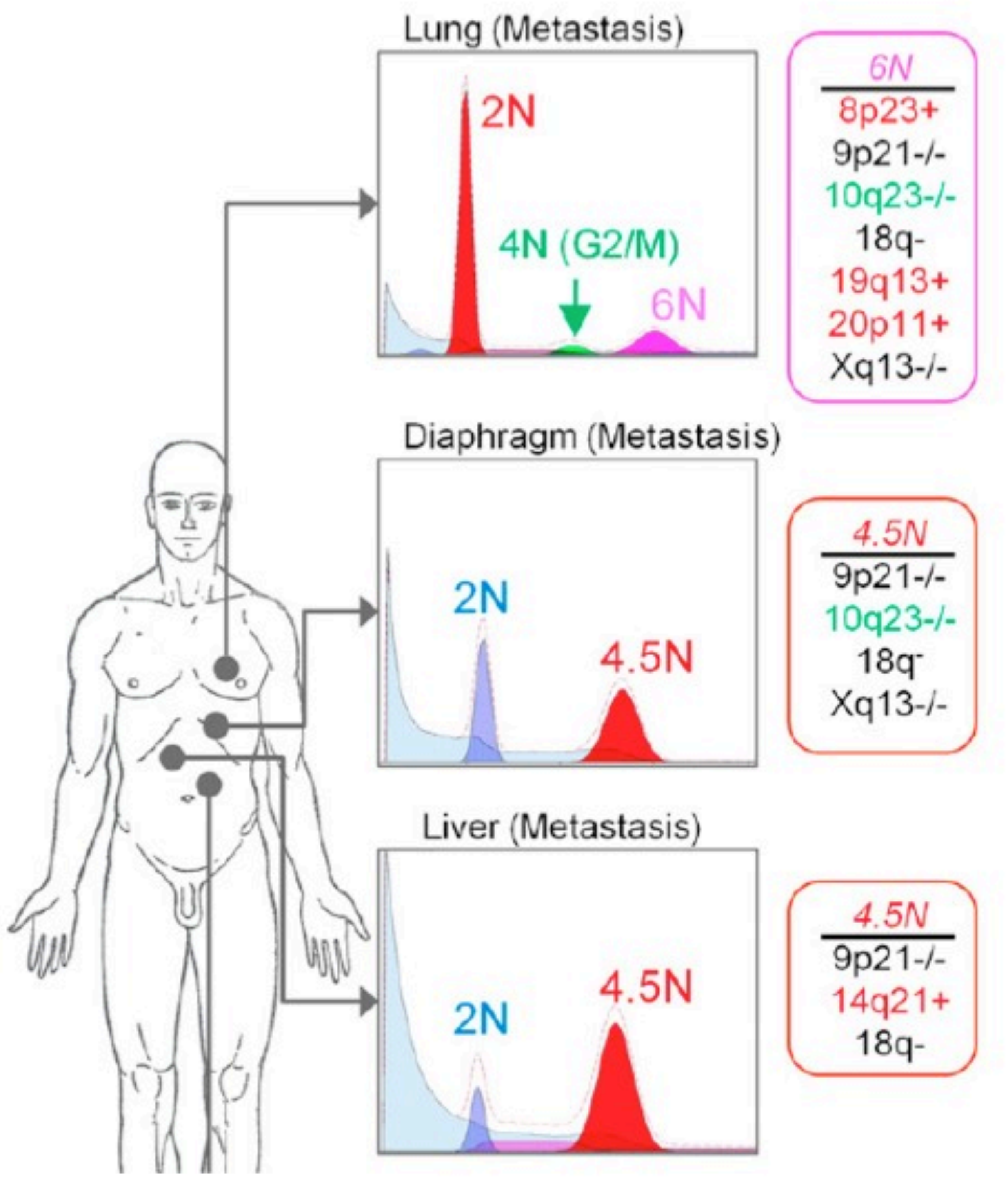
Evolutionarily Assembled cis-Regulatory Module at a Human Ciliopathy Locus

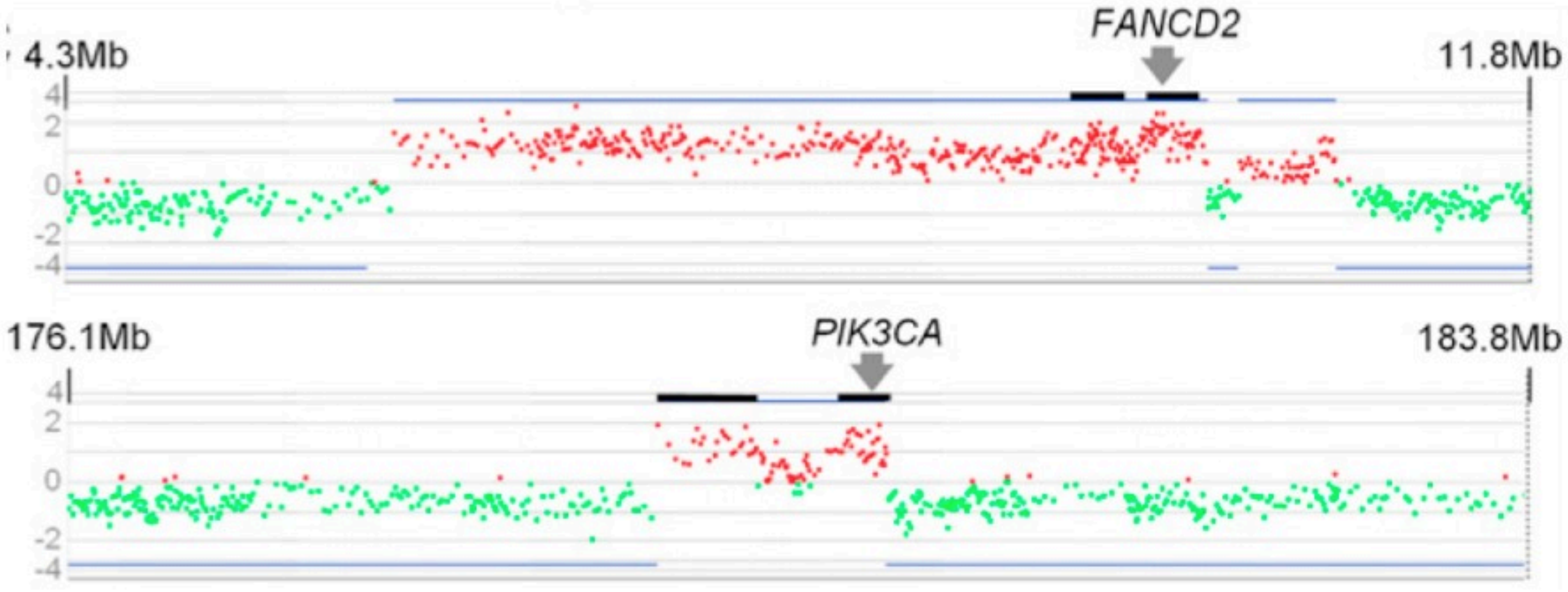


Genomics, Proteomics & Bioinformatics

Exams

Based on what you have learned this semester, what procedure would you recommend in the form of ideal, personalized cancer treatment of this patient (hospice and similar recommendations are not a valid answer for this test, though it might be in real life). *Limit your answer to a maximum of 3 sentences*.





No one *gives* you an education.

If you want one, you have to take it.

John Taylor Gotto

http://nccueagles.yuku.com/topic/6441#.T1o5-pjufqE

Laboratory Methods in Genomics

mostly juniors and seniors
majority biology & math majors

Laboratory Methods in Genomics

Students should be able to:

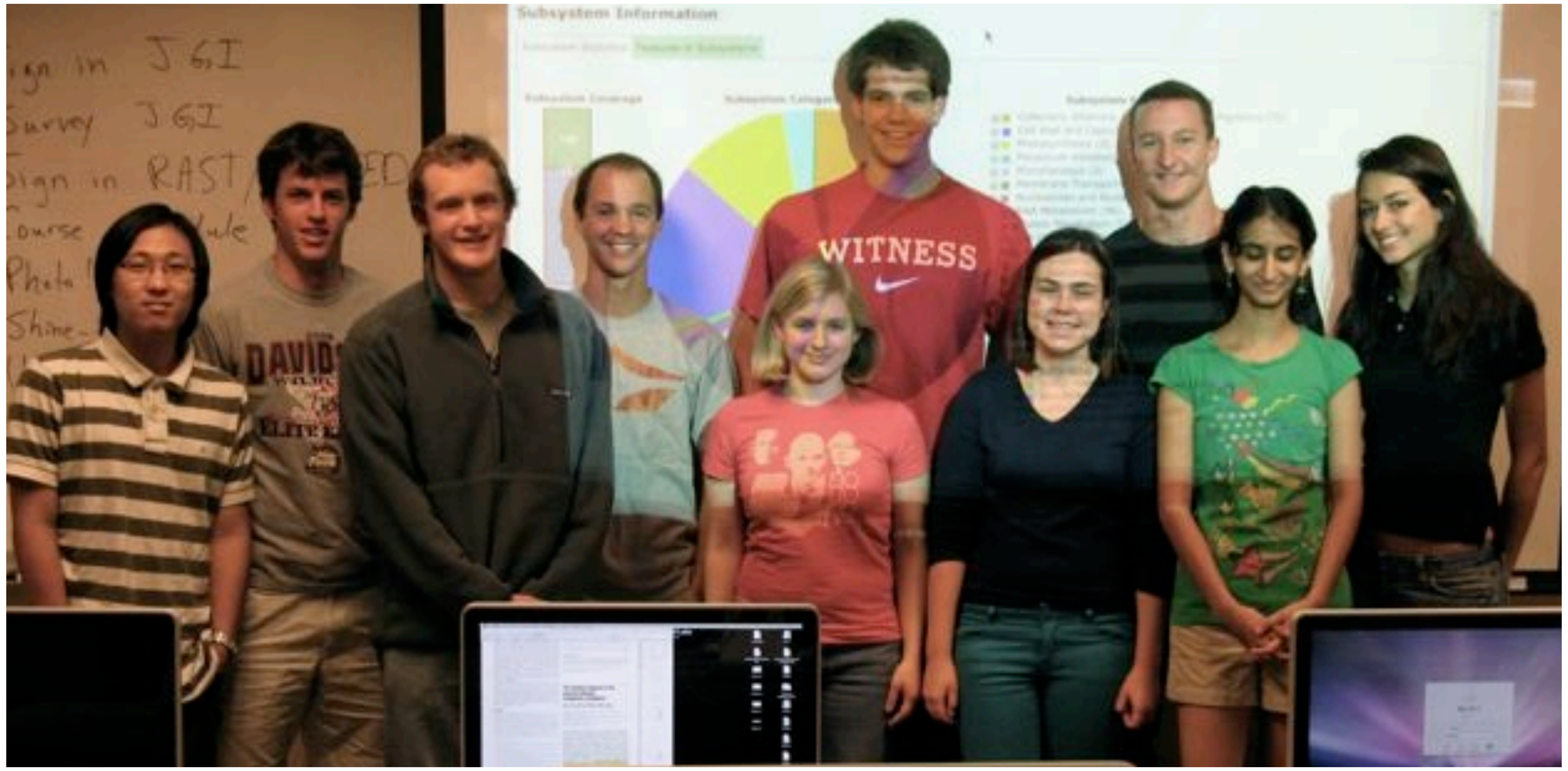
- do real genome research
- evaluate ambiguities
- use databases to annotate genes
- determine pathway completeness
- communicate to wider audience

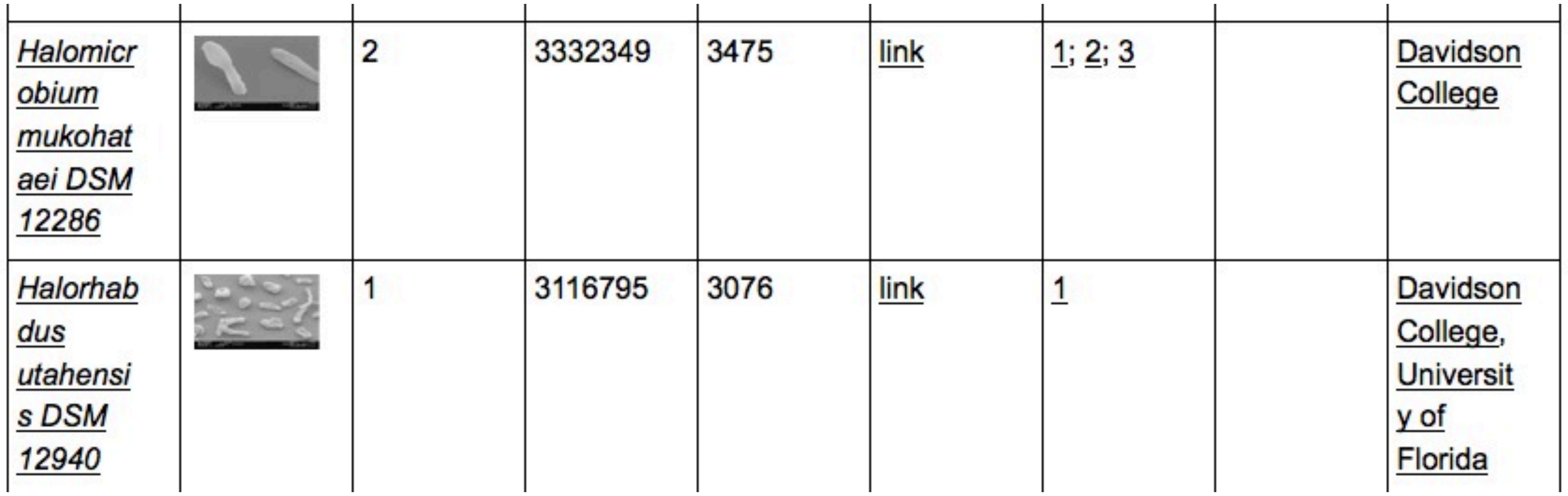
Laboratory Methods in Genomics

2008

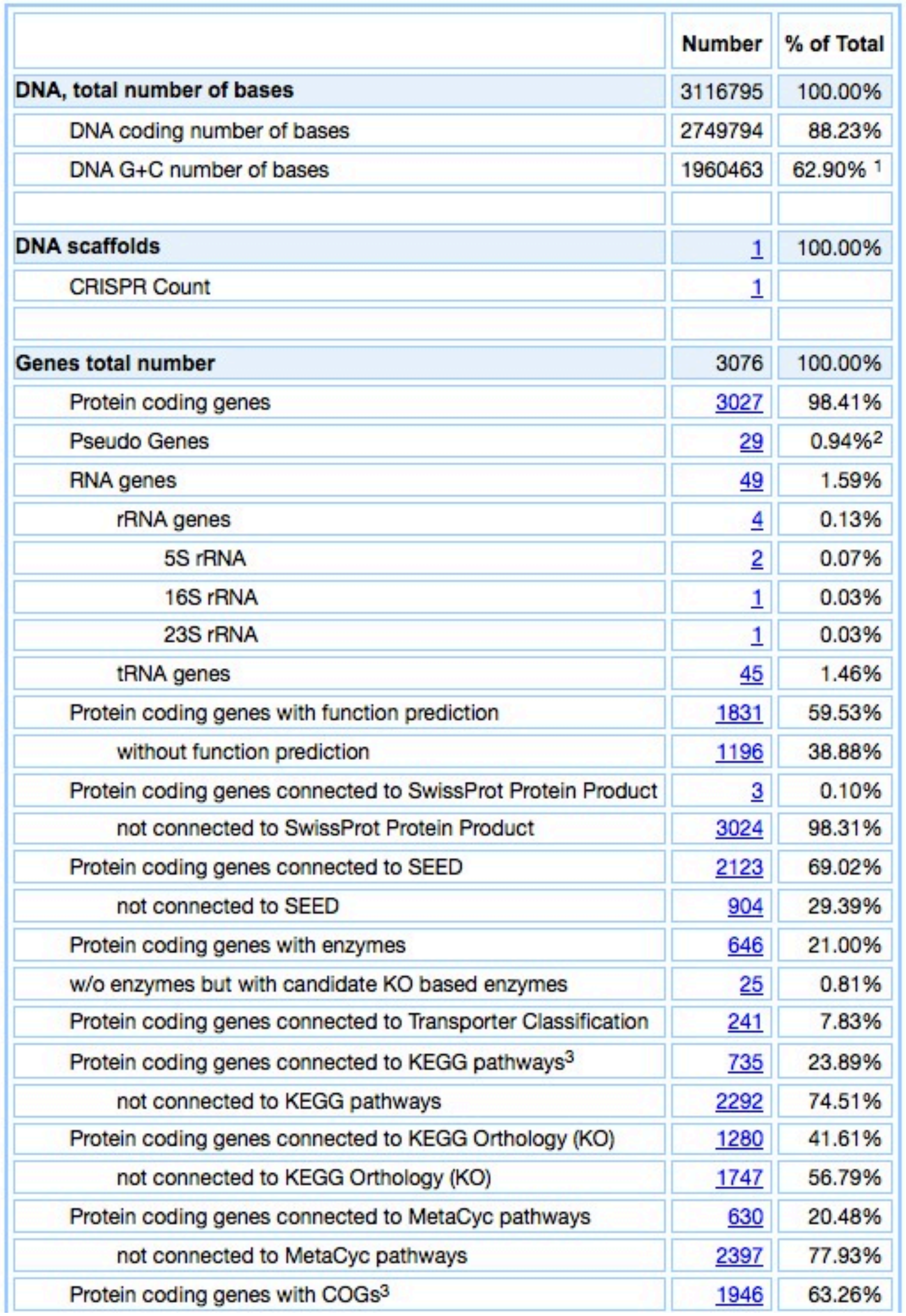
JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

DOE Joint Genome Institute
Enabling Advances in Bioenergy & Environmental Research





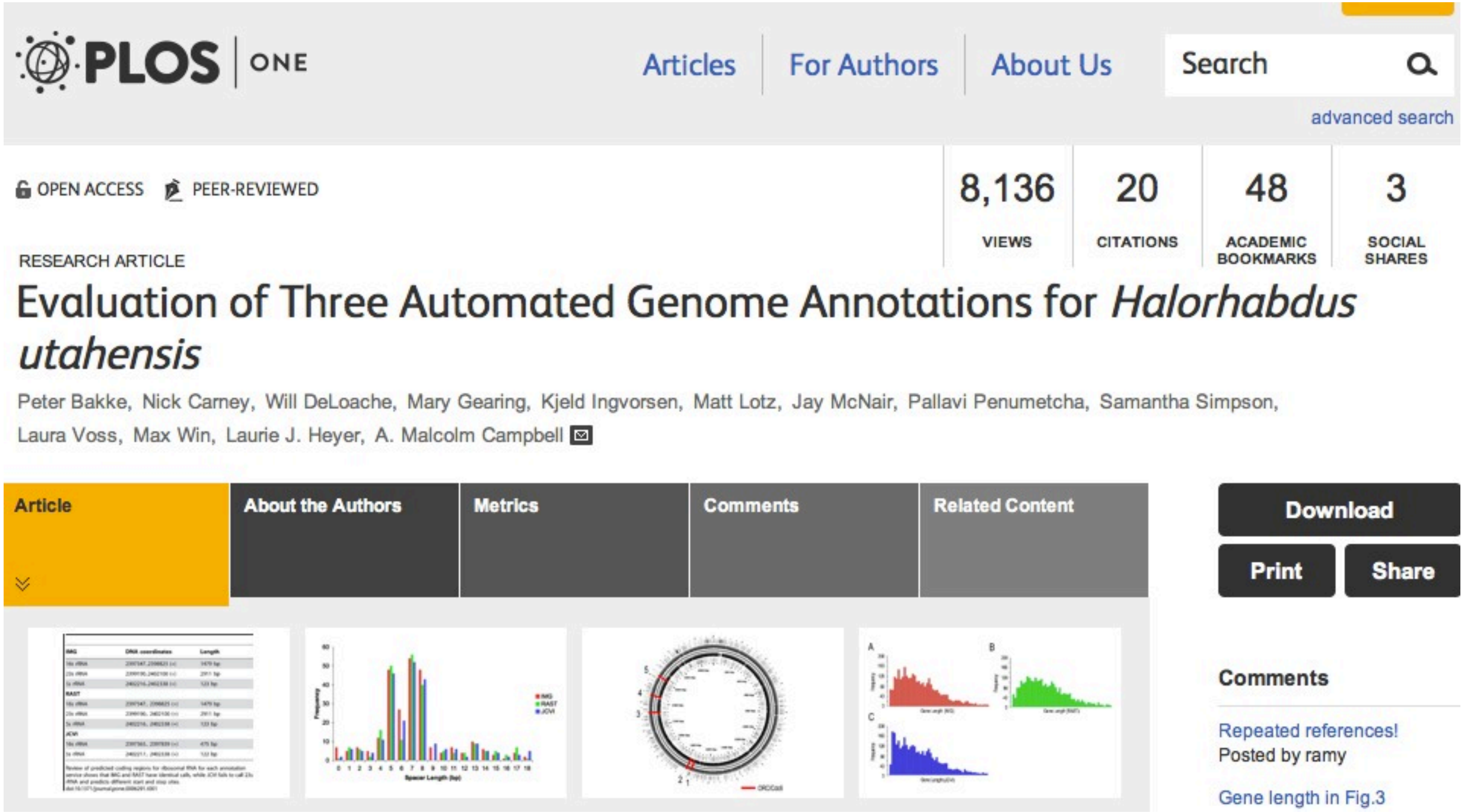
Halomicrobium mukohataei DSM 12286		2	3332349	3475	link	1; 2; 3		Davidson College
Halorhabdus utahensis DSM 12940		1	3116795	3076	link	1		Davidson College, University of Florida



	Number	% of Total
DNA, total number of bases	3116795	100.00%
DNA coding number of bases	2749794	88.23%
DNA G+C number of bases	1960463	62.90% [1]
DNA scaffolds	1	100.00%
CRISPR Count	1	
Genes total number	3076	100.00%
Protein coding genes	3027	98.41%
Pseudo Genes	29	0.94%[2]
RNA genes	49	1.59%
rRNA genes	4	0.13%
5S rRNA	2	0.07%
16S rRNA	1	0.03%
23S rRNA	1	0.03%
tRNA genes	45	1.46%
Protein coding genes with function prediction	1831	59.53%
without function prediction	1196	38.88%
Protein coding genes connected to SwissProt Protein Product	3	0.10%
not connected to SwissProt Protein Product	3024	98.31%
Protein coding genes connected to SEED	2123	69.02%
not connected to SEED	904	29.39%
Protein coding genes with enzymes	646	21.00%
w/o enzymes but with candidate KO based enzymes	25	0.81%
Protein coding genes connected to Transporter Classification	241	7.83%
Protein coding genes connected to KEGG pathways[3]	735	23.89%
not connected to KEGG pathways	2292	74.51%
Protein coding genes connected to KEGG Orthology (KO)	1280	41.61%
not connected to KEGG Orthology (KO)	1747	56.79%
Protein coding genes connected to MetaCyc pathways	630	20.48%
not connected to MetaCyc pathways	2397	77.93%
Protein coding genes with COGs[3]	1946	63.26%

Laboratory Methods in Genomics

2008



8,136	20	48	3
VIEWS	CITATIONS	ACADEMIC BOOKMARKS	SOCIAL SHARES

RESEARCH ARTICLE

Evaluation of Three Automated Genome Annotations for *Halorhabdus utahensis*

Peter Bakke, Nick Carney, Will DeLoache, Mary Gearing, Kjeld Ingvorsen, Matt Lotz, Jay McNair, Pallavi Penumetcha, Samantha Simpson, Laura Voss, Max Win, Laurie J. Heyer, A. Malcolm Campbell ✉

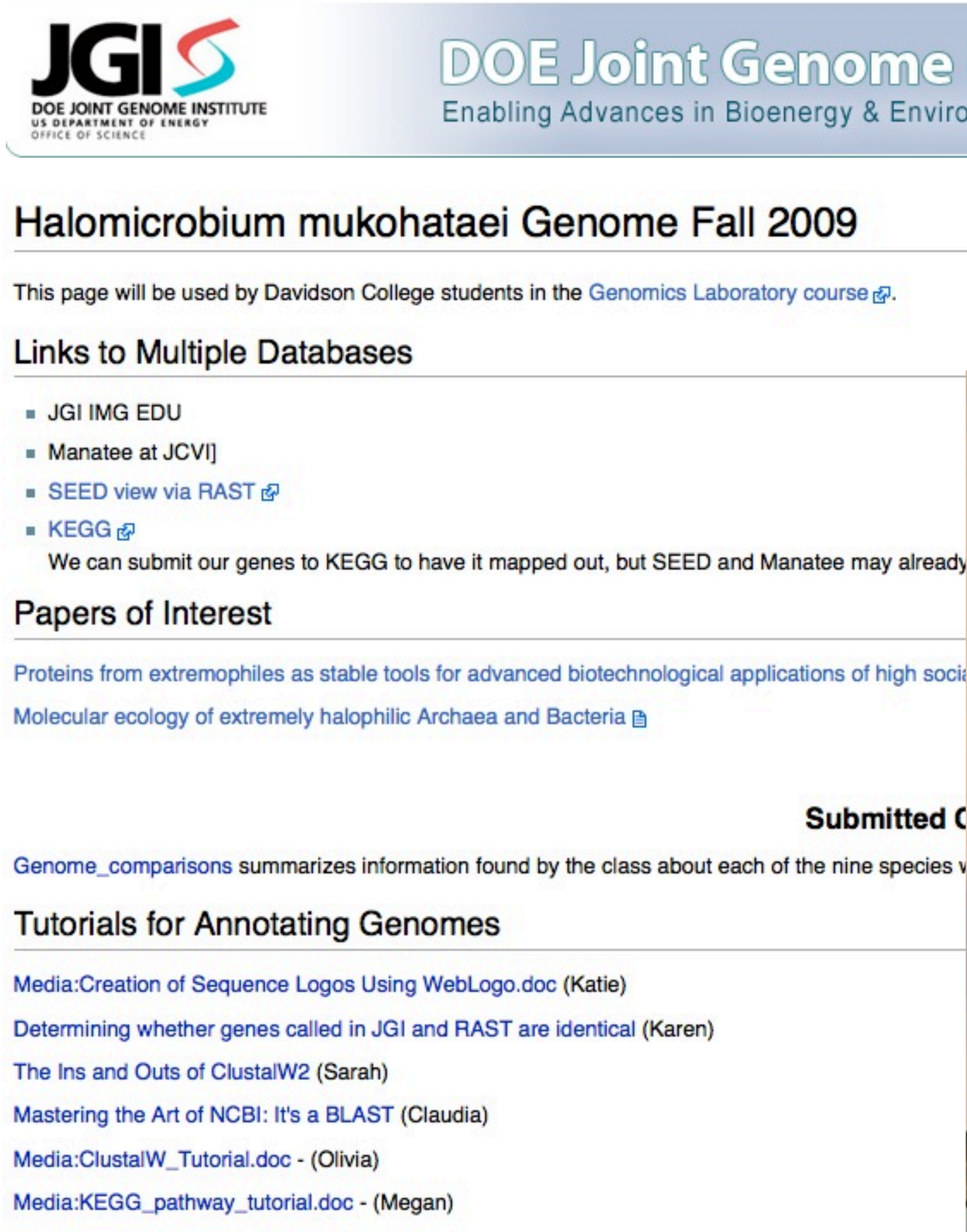
Comments

Repeated references!
Posted by ramy

Gene length in Fig.3

Laboratory Methods in Genomics

2009



JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

DOE Joint Genome Institute

Enabling Advances in Bioenergy & Environmental Research

Halomicrobium mukohataei Genome Fall 2009

This page will be used by Davidson College students in the Genomics Laboratory course.

Links to Multiple Databases

- JGI IMG EDU
- Manatee at JCVI]
- SEED view via RAST
- KEGG
  We can submit our genes to KEGG to have it mapped out, but SEED and Manatee may already

Papers of Interest

Proteins from extremophiles as stable tools for advanced biotechnological applications of high soci

Molecular ecology of extremely halophilic Archaea and Bacteria

Submitted C

Genome_comparisons summarizes information found by the class about each of the nine species v

Tutorials for Annotating Genomes

Media:Creation of Sequence Logos Using WebLogo.doc (Katie)

Determining whether genes called in JGI and RAST are identical (Karen)

The Ins and Outs of ClustalW2 (Sarah)

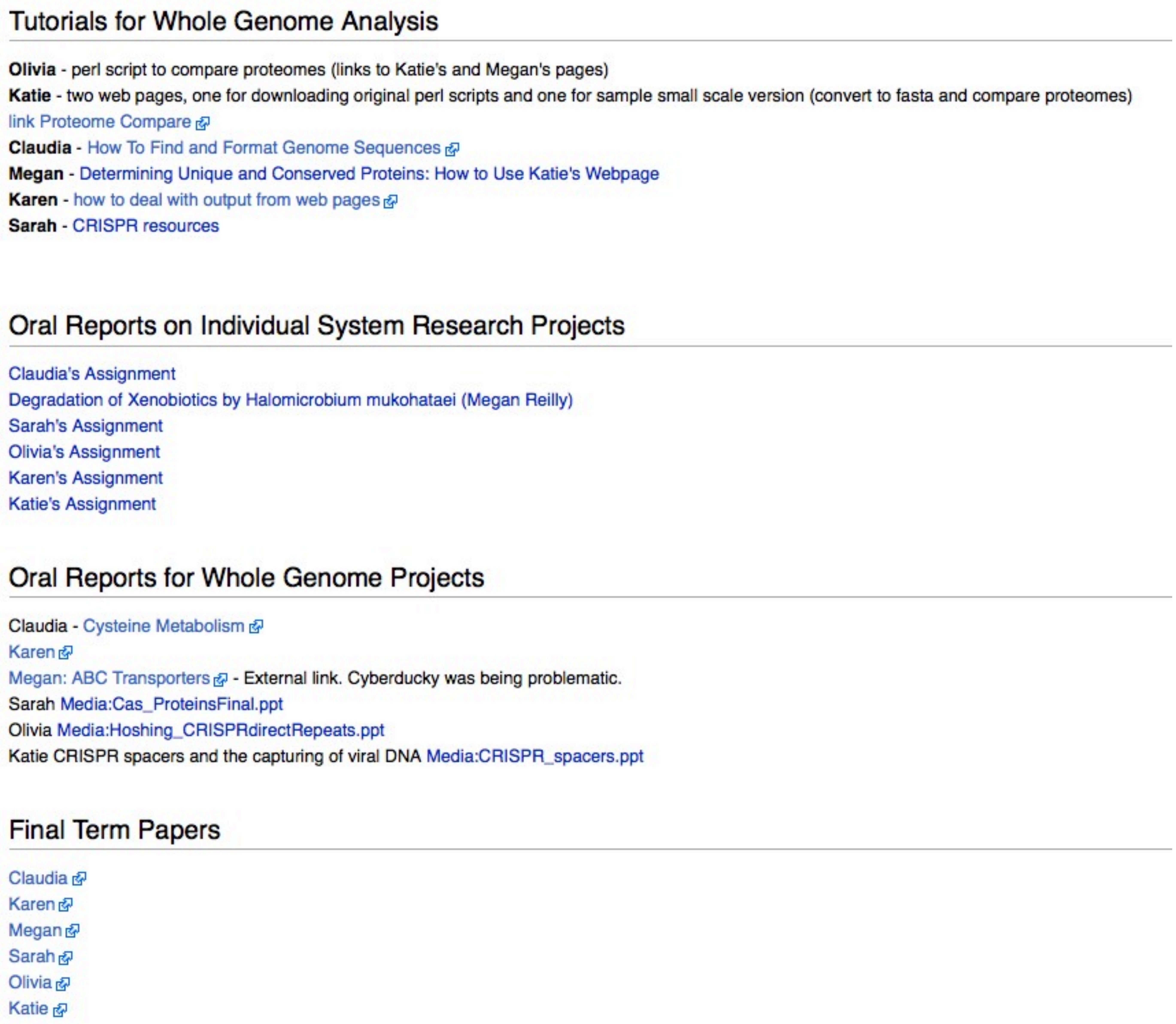
Mastering the Art of NCBI: It's a BLAST (Claudia)

Media:ClustalW_Tutorial.doc - (Olivia)

Media:KEGG_pathway_tutorial.doc - (Megan)

Laboratory Methods in Genomics

2009



Tutorials for Whole Genome Analysis

Olivia - perl script to compare proteomes (links to Katie's and Megan's pages)
Katie - two web pages, one for downloading original perl scripts and one for sample small scale version (convert to fasta and compare proteomes)
link Proteome Compare
Claudia - How To Find and Format Genome Sequences
Megan - Determining Unique and Conserved Proteins: How to Use Katie's Webpage
Karen - how to deal with output from web pages
Sarah - CRISPR resources

Oral Reports on Individual System Research Projects

Claudia's Assignment
Degradation of Xenobiotics by Halomicrobium mukohataei (Megan Reilly)
Sarah's Assignment
Olivia's Assignment
Karen's Assignment
Katie's Assignment

Oral Reports for Whole Genome Projects

Claudia - Cysteine Metabolism
Karen
Megan: ABC Transporters - External link. Cyberducky was being problematic.
Sarah Media:Cas_ProteinsFinal.ppt
Olivia Media:Hoshing_CRISPRdirectRepeats.ppt
Katie CRISPR spacers and the capturing of viral DNA Media:CRISPR_spacers.ppt

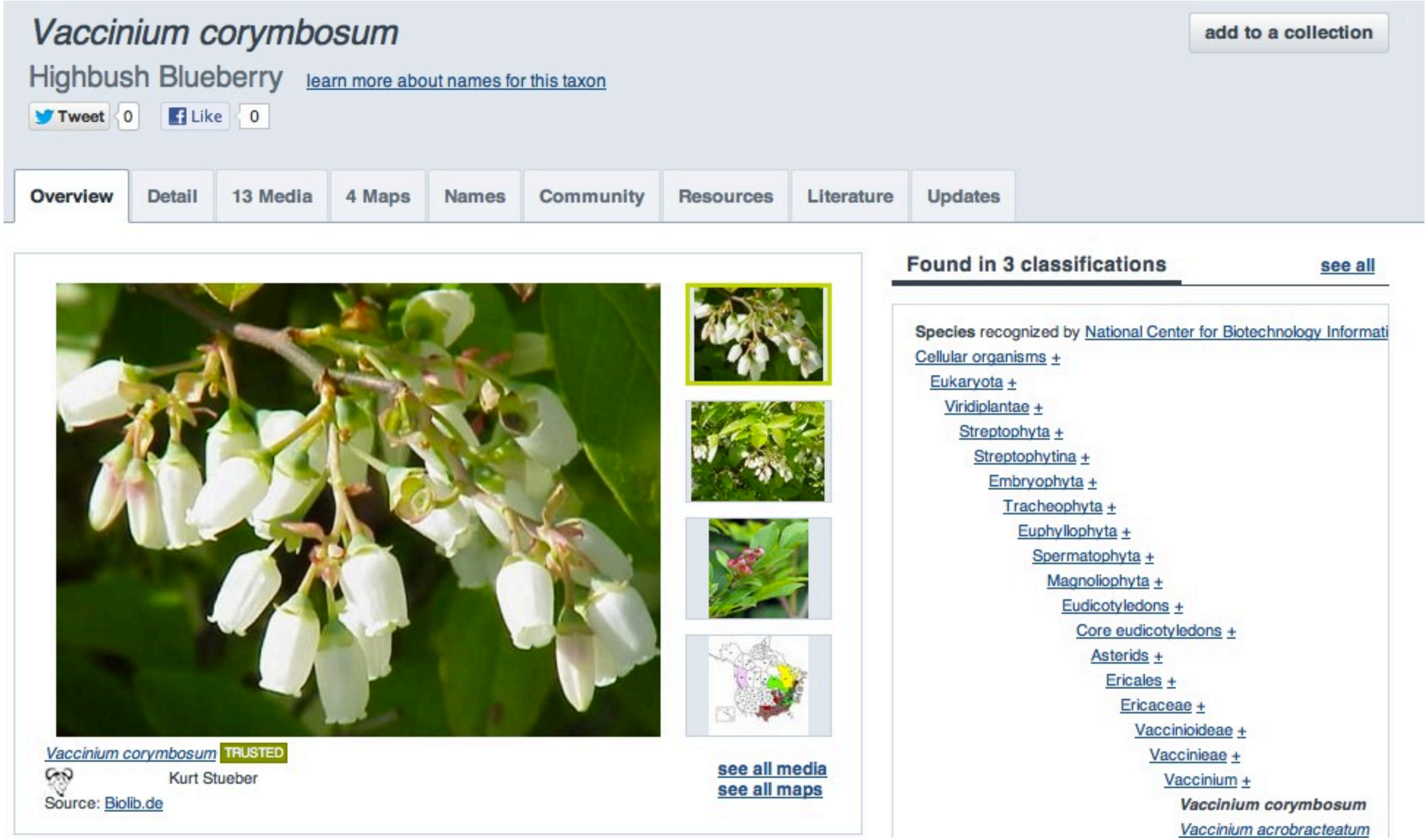
Final Term Papers

Claudia
Karen
Megan
Sarah
Olivia
Katie

Laboratory Methods in Genomics

2011

2012



Vaccinium corymbosum

Highbush Blueberry learn more about names for this taxon

Vaccinium corymbosum

Kurt Stueber

Source: Biolib.de

Found in 3 classifications

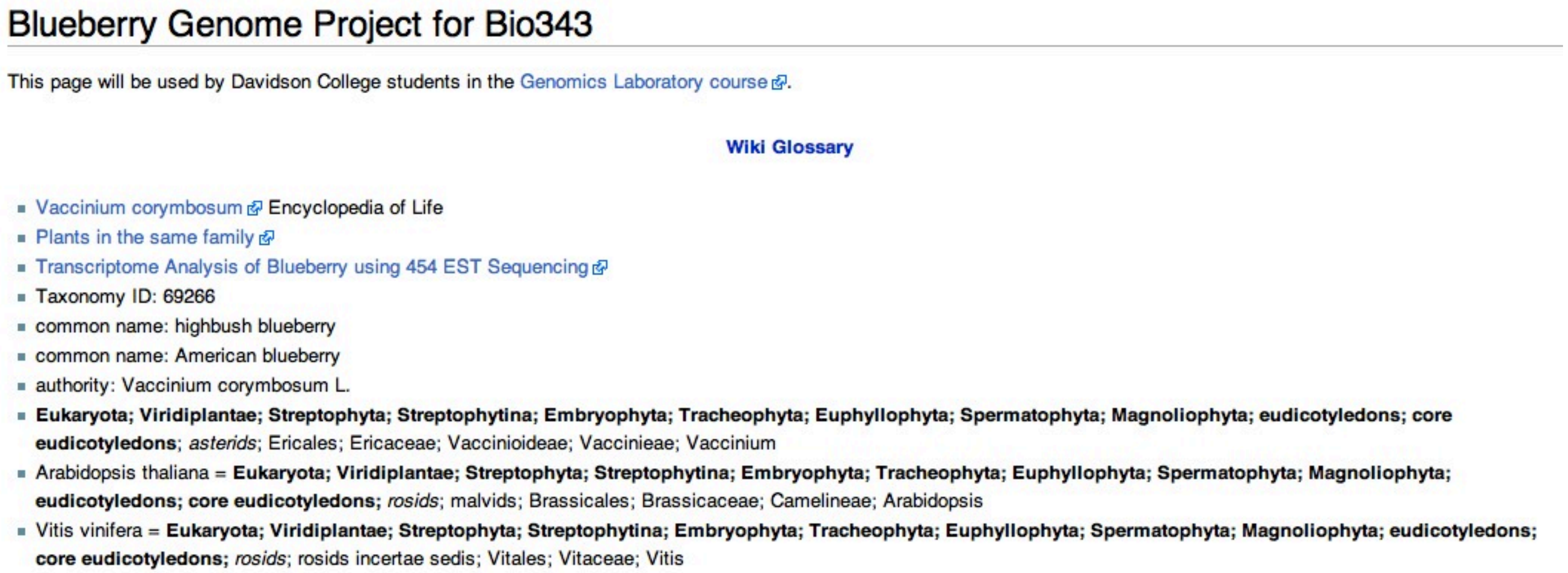
Species recognized by National Center for Biotechnology Informati

- Cellular organisms +
  - Eukaryota +
    - Viridiplantae +
      - Streptophyta +
        - Streptophytina +
          - Embryophyta +
            - Tracheophyta +
              - Euphyllophyta +
                - Spermatophyta +
                  - Magnoliophyta +
                    - Eudicotyledons +
                      - Core eudicotyledons +
                        - Asterids +
                          - Ericales +
                            - Ericaceae +
                              - Vaccinioideae +
                                - Vaccinieae +
                                  - Vaccinium +
                                    - ***Vaccinium corymbosum***
                                    - *Vaccinium acrobracteatum*

Laboratory Methods in Genomics

2011

2012



Blueberry Genome Project for Bio343

This page will be used by Davidson College students in the Genomics Laboratory course.

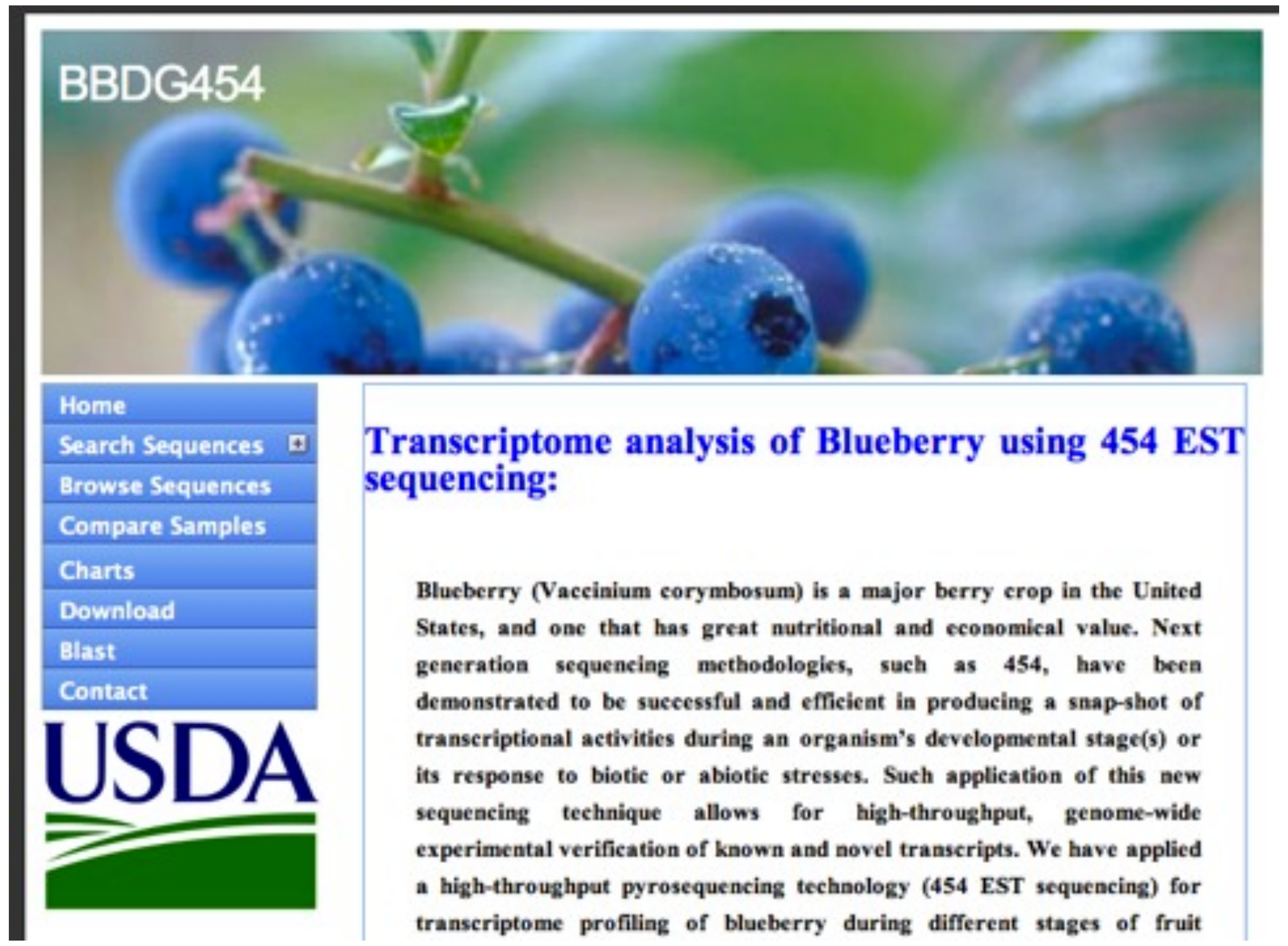
Wiki Glossary

- Vaccinium corymbosum Encyclopedia of Life
- Plants in the same family
- Transcriptome Analysis of Blueberry using 454 EST Sequencing
- Taxonomy ID: 69266
- common name: highbush blueberry
- common name: American blueberry
- authority: Vaccinium corymbosum L.
- **Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons**; *asterids*; Ericales; Ericaceae; Vaccinioideae; Vaccinieae; Vaccinium
- Arabidopsis thaliana = **Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;** *rosids*; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis
- Vitis vinifera = **Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;** *rosids*; rosids incertae sedis; Vitales; Vitaceae; Vitis

Laboratory Methods in Genomics

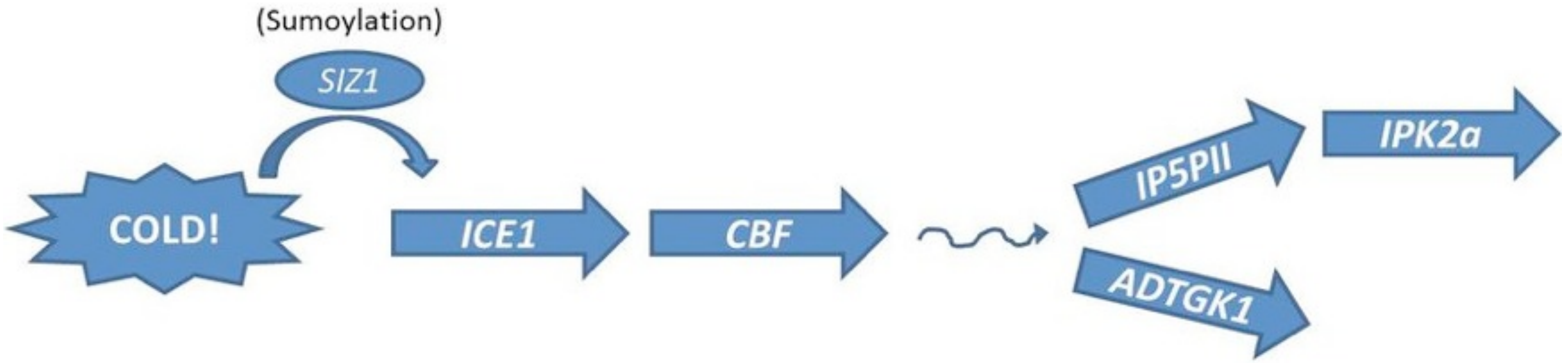
2011

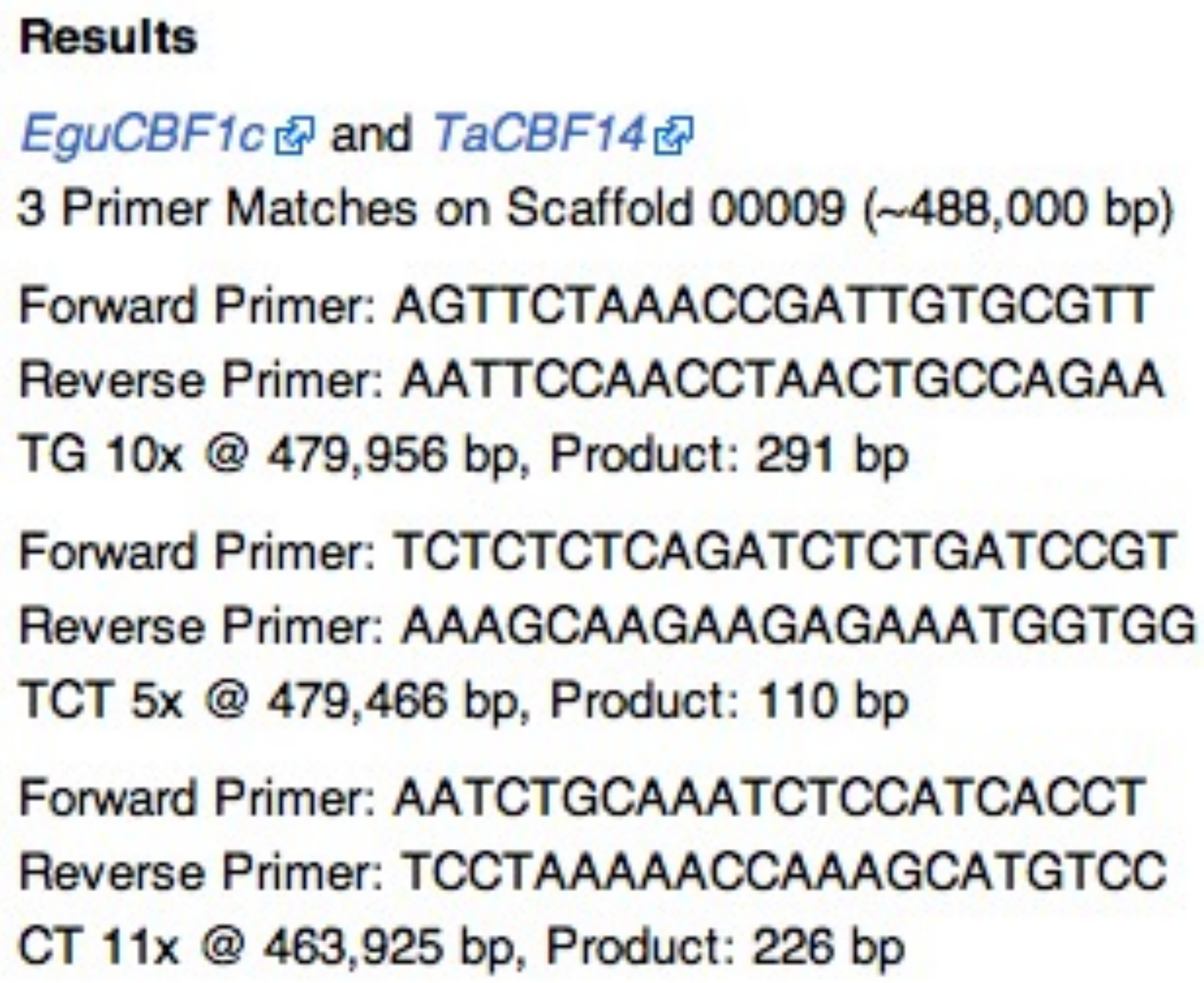
2012



Transcriptome analysis of Blueberry using 454 EST sequencing:

Blueberry (Vaccinium corymbosum) is a major berry crop in the United States, and one that has great nutritional and economical value. Next generation sequencing methodologies, such as 454, have been demonstrated to be successful and efficient in producing a snap-shot of transcriptional activities during an organism's developmental stage(s) or its response to biotic or abiotic stresses. Such application of this new sequencing technique allows for high-throughput, genome-wide experimental verification of known and novel transcripts. We have applied a high-throughput pyrosequencing technology (454 EST sequencing) for transcriptome profiling of blueberry during different stages of fruit





Results

EguCBF1c and *TaCBF14*

3 Primer Matches on Scaffold 00009 (~488,000 bp)

Forward Primer: AGTTCTAAACCGATTGTGCGTT
Reverse Primer: AATTCCAACCTAACTGCCAGAA
TG 10x @ 479,956 bp, Product: 291 bp

Forward Primer: TCTCTCTCAGATCTCTGATCCGT
Reverse Primer: AAAGCAAGAAGAGAAATGGTGG
TCT 5x @ 479,466 bp, Product: 110 bp

Forward Primer: AATCTGCAAATCTCCATCACCT
Reverse Primer: TCCTAAAAACCAAAGCATGTCC
CT 11x @ 463,925 bp, Product: 226 bp

Education is the only industry where customers never complain when they get less product for their money.

Bioinformatics

Dr. Laurie Heyer
(Math Dept.)

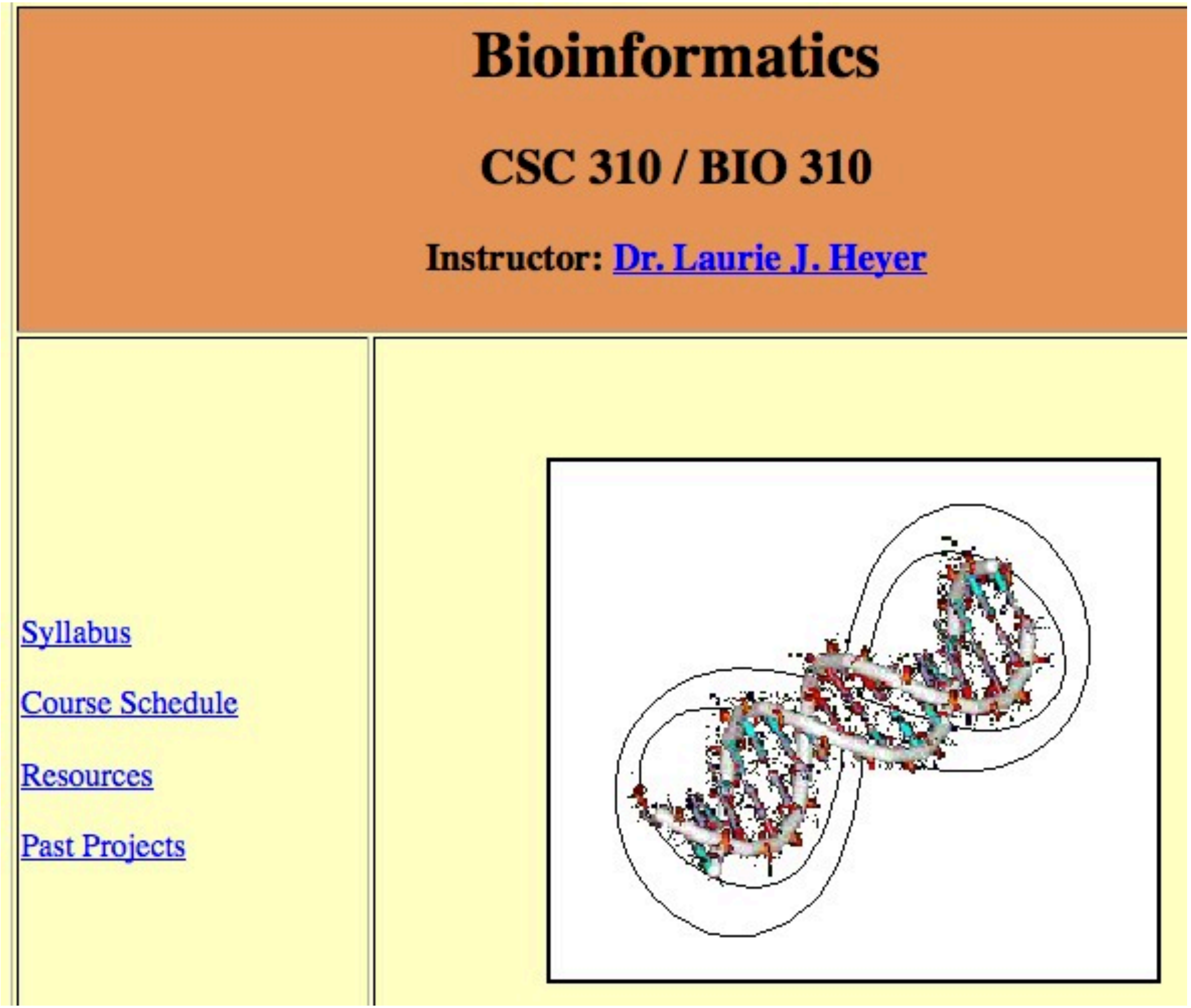
mostly juniors and seniors
50% math majors + 50% biology majors

Bioinformatics

Students should be able to:

- write computer code to do tasks
- work with real data
- collaborate with different major
- produce functional online tools

Bioinformatics



Bioinformatics

CSC 310 / BIO 310

Instructor: Dr. Laurie J. Heyer

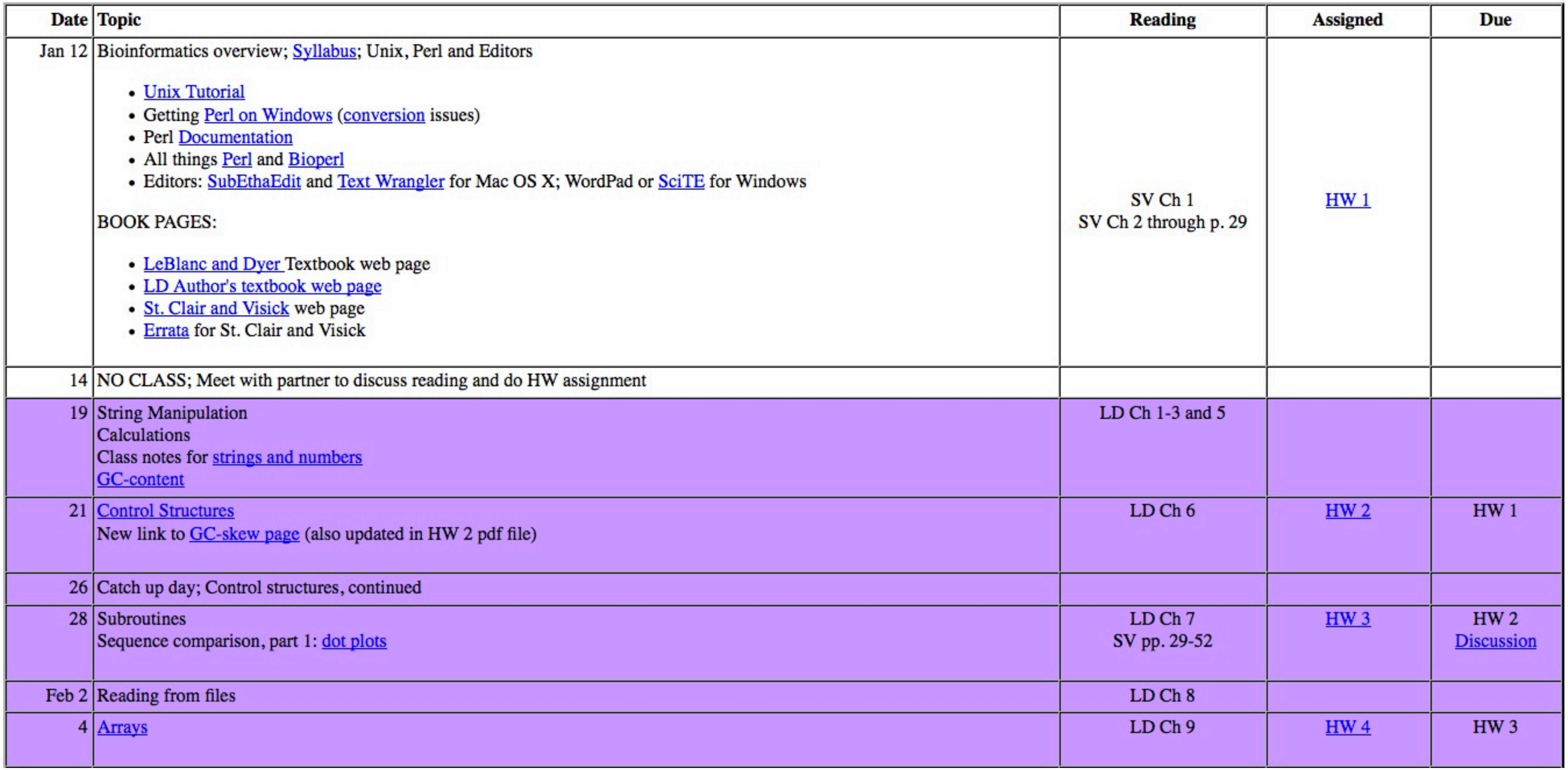
Syllabus

Course Schedule

Resources

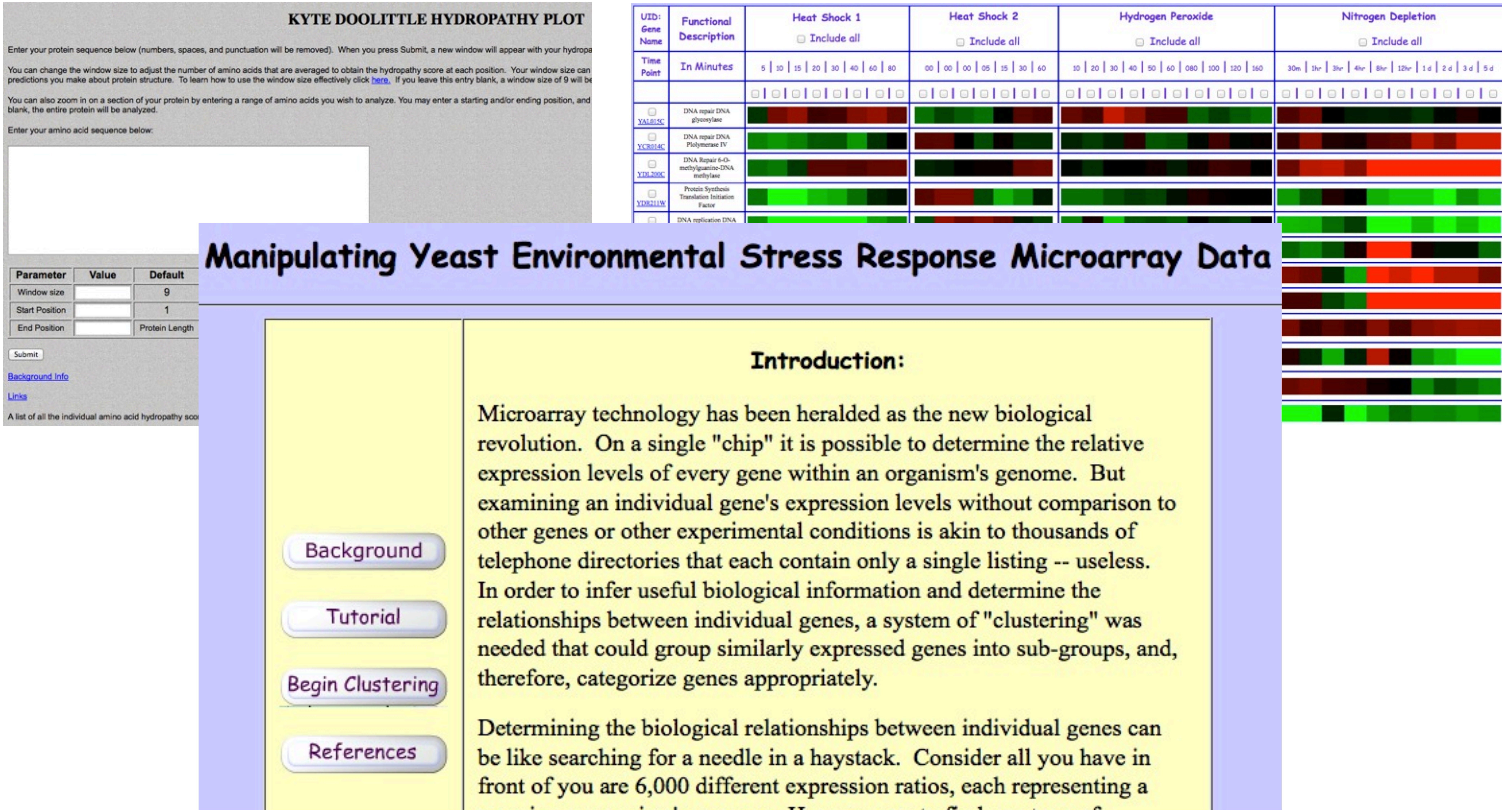
Past Projects

Bioinformatics



Date	Topic	Reading	Assigned	Due
Jan 12	Bioinformatics overview; Syllabus; Unix, Perl and Editors • Unix Tutorial • Getting Perl on Windows (conversion issues) • Perl Documentation • All things Perl and Bioperl • Editors: SubEthaEdit and Text Wrangler for Mac OS X; WordPad or SciTE for Windows BOOK PAGES: • LeBlanc and Dyer Textbook web page • LD Author's textbook web page • St. Clair and Visick web page • Errata for St. Clair and Visick	SV Ch 1 SV Ch 2 through p. 29	HW 1	
14	NO CLASS; Meet with partner to discuss reading and do HW assignment			
19	String Manipulation Calculations Class notes for strings and numbers GC-content	LD Ch 1-3 and 5		
21	Control Structures New link to GC-skew page (also updated in HW 2 pdf file)	LD Ch 6	HW 2	HW 1
26	Catch up day; Control structures, continued			
28	Subroutines Sequence comparison, part 1: dot plots	LD Ch 7 SV pp. 29-52	HW 3	HW 2 Discussion
Feb 2	Reading from files	LD Ch 8		
4	Arrays	LD Ch 9	HW 4	HW 3

Bioinformatics



KYTE DOOLITTLE HYDROPATHY PLOT

Enter your protein sequence below (numbers, spaces, and punctuation will be removed). When you press Submit, a new window will appear with your hydropa

You can change the window size to adjust the number of amino acids that are averaged to obtain the hydropathy score at each position. Your window size can predictions you make about protein structure. To learn how to use the window size effectively click here. If you leave this entry blank, a window size of 9 will be

You can also zoom in on a section of your protein by entering a range of amino acids you wish to analyze. You may enter a starting and/or ending position, and blank, the entire protein will be analyzed.

Enter your amino acid sequence below:

Parameter	Value	Default
Window size		9
Start Position		1
End Position		Protein Length

A list of all the individual amino acid hydropathy sco

UID: Gene Name	Functional Description	Heat Shock 1 ☐ Include all	Heat Shock 2 ☐ Include all	Hydrogen Peroxide ☐ Include all	Nitrogen Depletion ☐ Include all
Time Point	In Minutes	5 10 15 20 30 40 60 80	00 00 00 05 15 30 60	10 20 30 40 50 60 080 100 120 160	30m 1hr 3hr 4hr 8hr 12hr 1 d 2 d 3 d 5 d
☐ YAL015C	DNA repair DNA glycosylase				
☐ YCR014C	DNA repair DNA Plolymerase IV				
☐ YDL200C	DNA Repair 6-O-methylguanine-DNA methylase				
☐ YDR211W	Protein Synthesis Translation Initiation Factor				
☐	DNA replication DNA				

Manipulating Yeast Environmental Stress Response Microarray Data

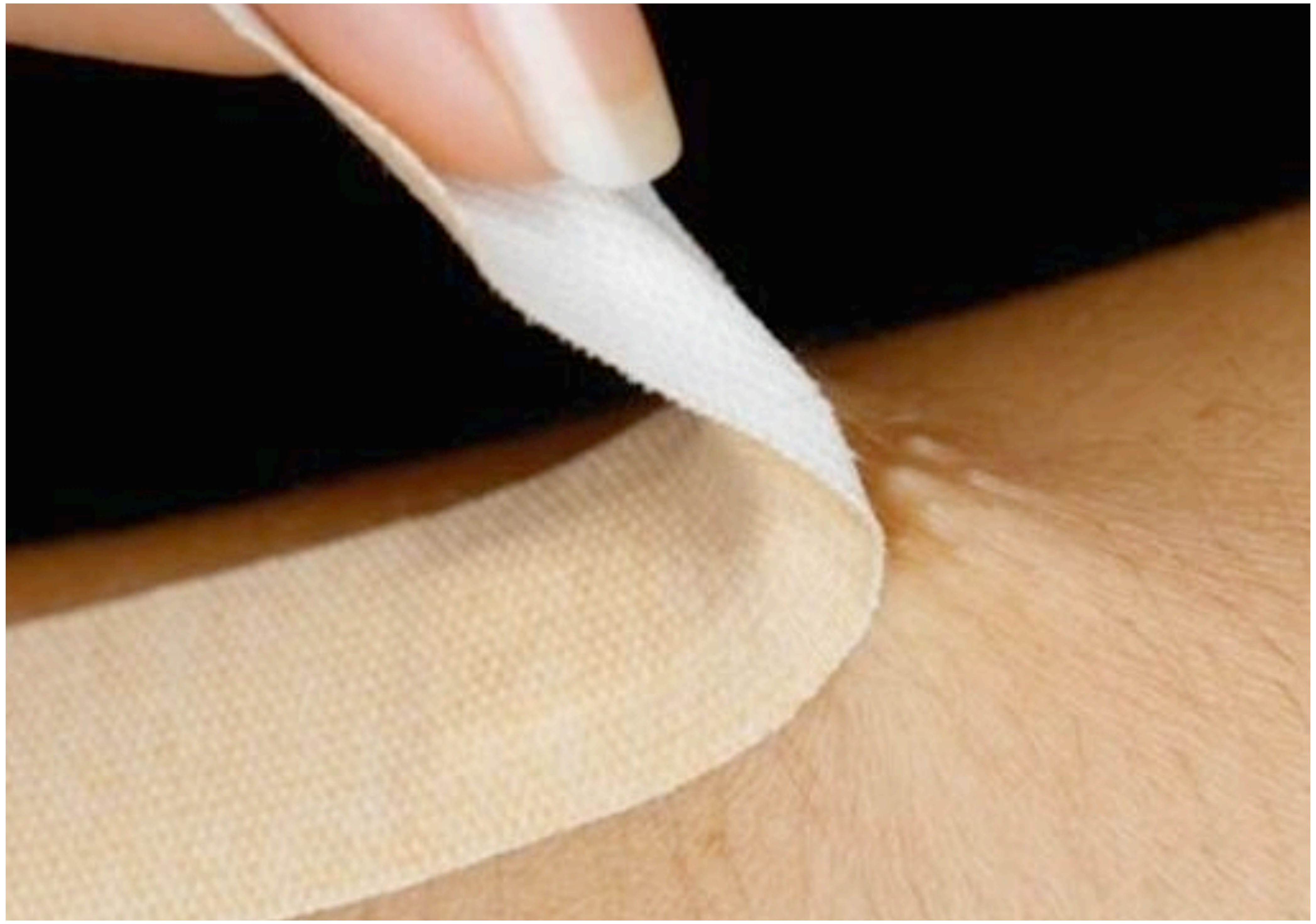
Introduction:

Microarray technology has been heralded as the new biological revolution. On a single "chip" it is possible to determine the relative expression levels of every gene within an organism's genome. But examining an individual gene's expression levels without comparison to other genes or other experimental conditions is akin to thousands of telephone directories that each contain only a single listing -- useless. In order to infer useful biological information and determine the relationships between individual genes, a system of "clustering" was needed that could group similarly expressed genes into sub-groups, and, therefore, categorize genes appropriately.

Determining the biological relationships between individual genes can be like searching for a needle in a haystack. Consider all you have in front of you are 6,000 different expression ratios, each representing a

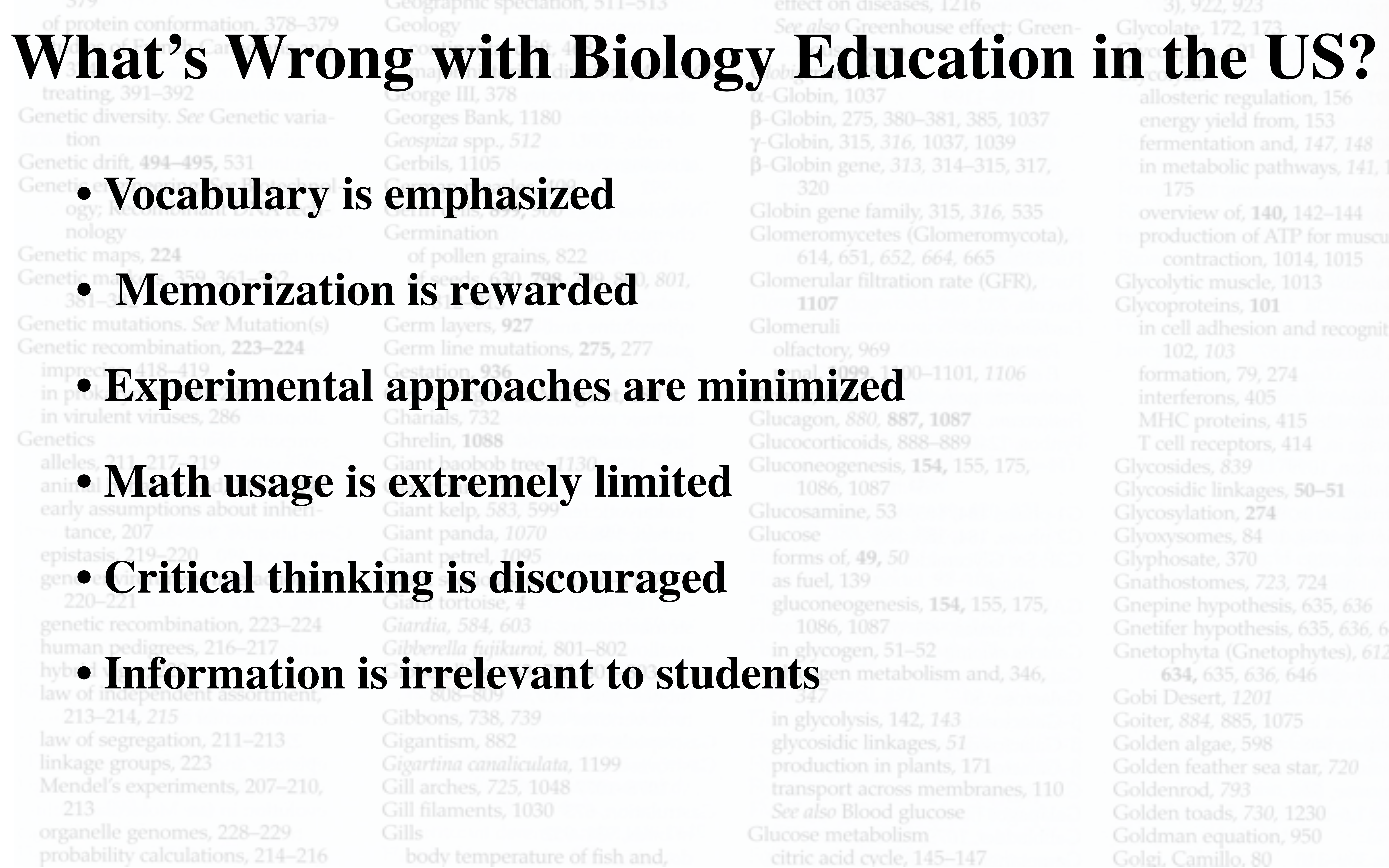
Introductory Biology

Integrating Concepts in Biology



A. Malcolm Campbell, Laurie J. Heyer
and Christopher J. Paradise

What’s Wrong with Biology Education in the US?

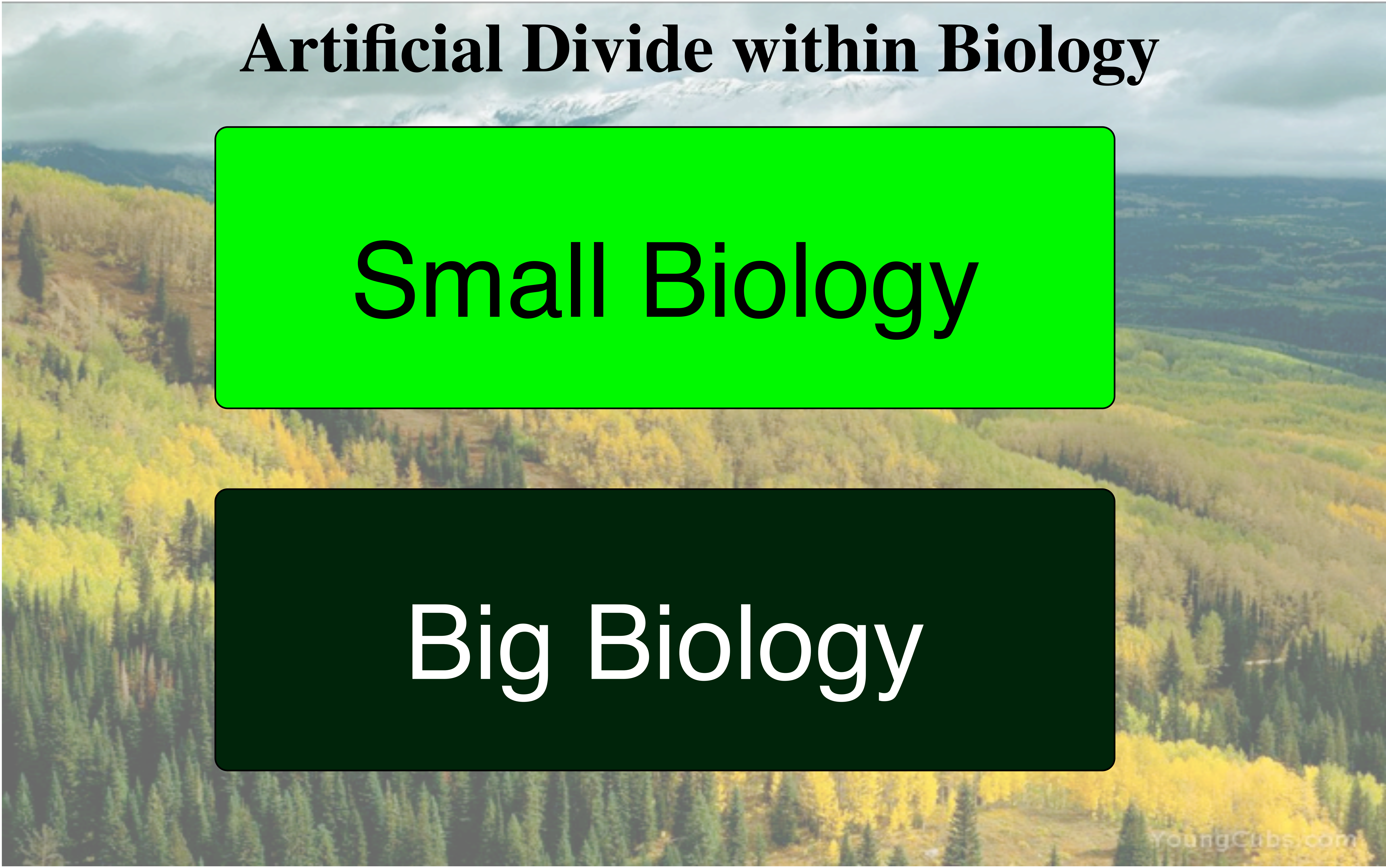


- Vocabulary is emphasized
- Memorization is rewarded
- Experimental approaches are minimized
- Math usage is extremely limited
- Critical thinking is discouraged
- Information is irrelevant to students

Present information and data…

… in the context of the big picture.

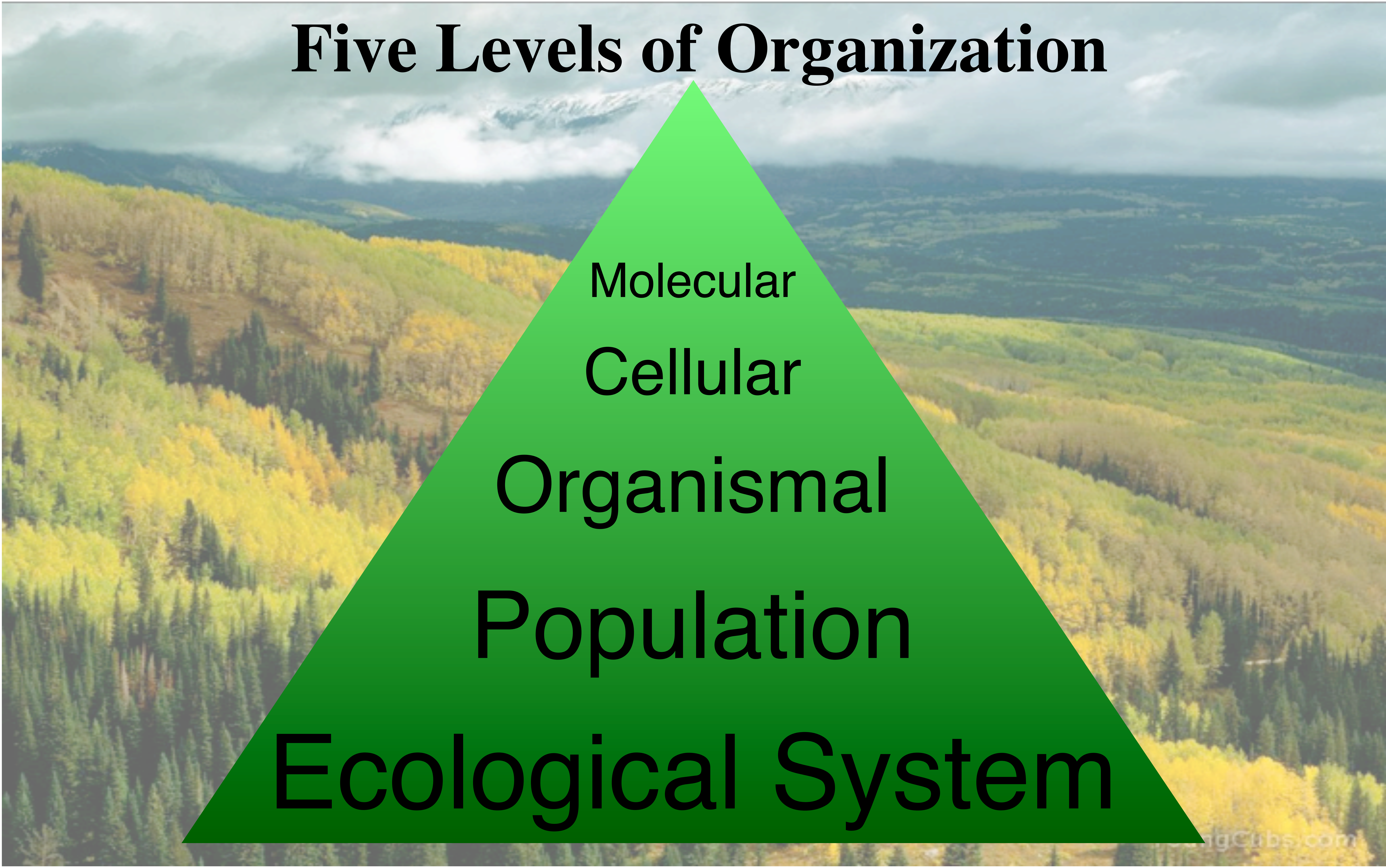
Artificial Divide within Biology



Small Biology

Big Biology

Five Levels of Organization



Molecular

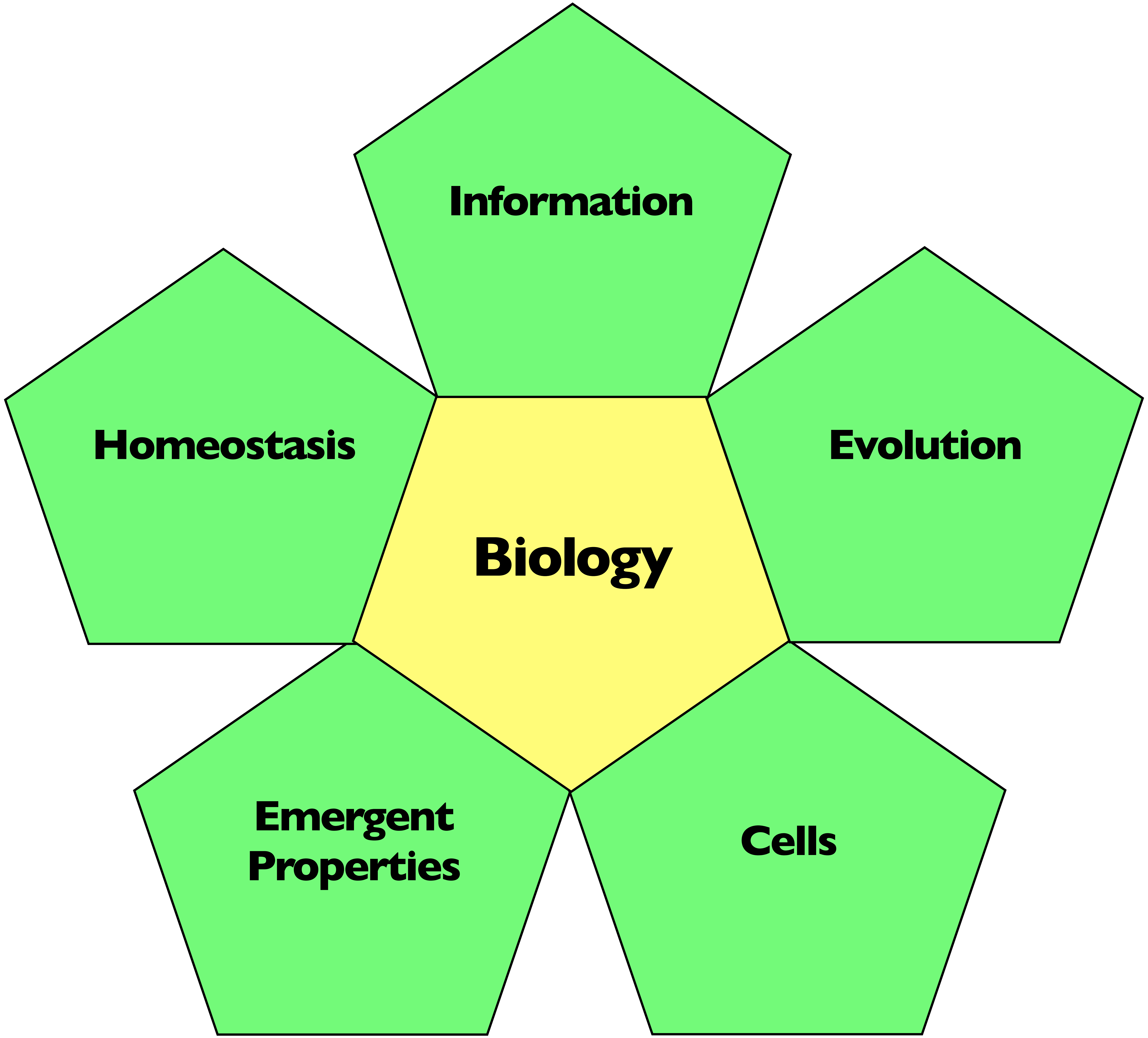
Cellular

Organismal

Population

Ecological System

Five Big Ideas of Biology



Information

Homeostasis

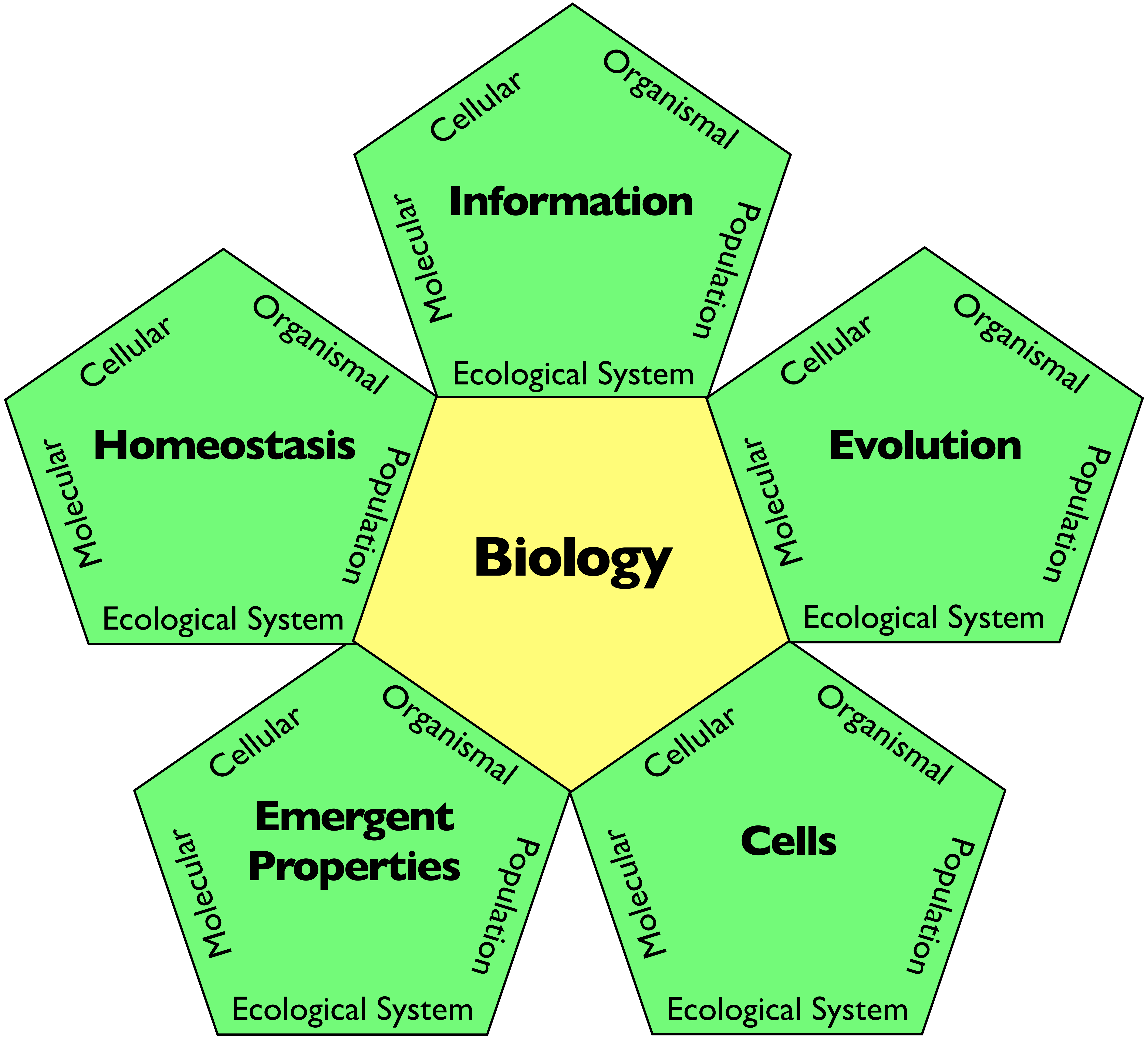
Evolution

Biology

Emergent Properties

Cells

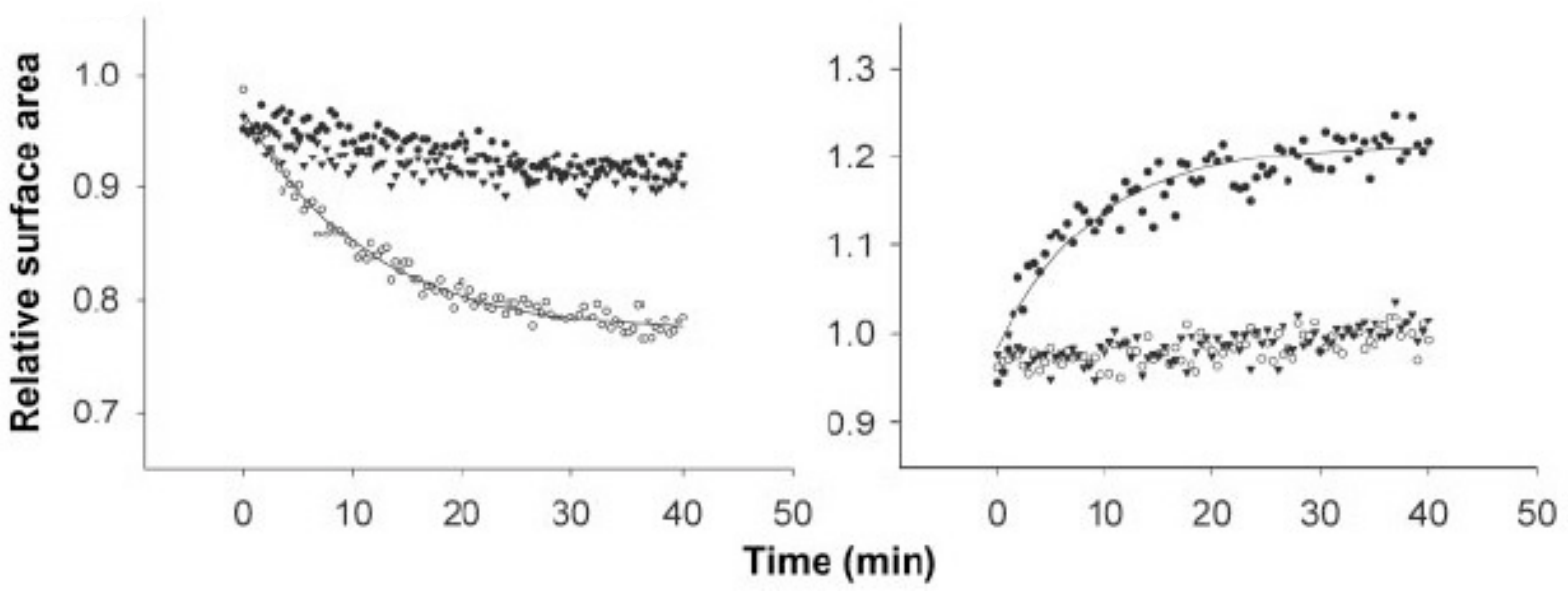
Five by Five Matrix of Biology



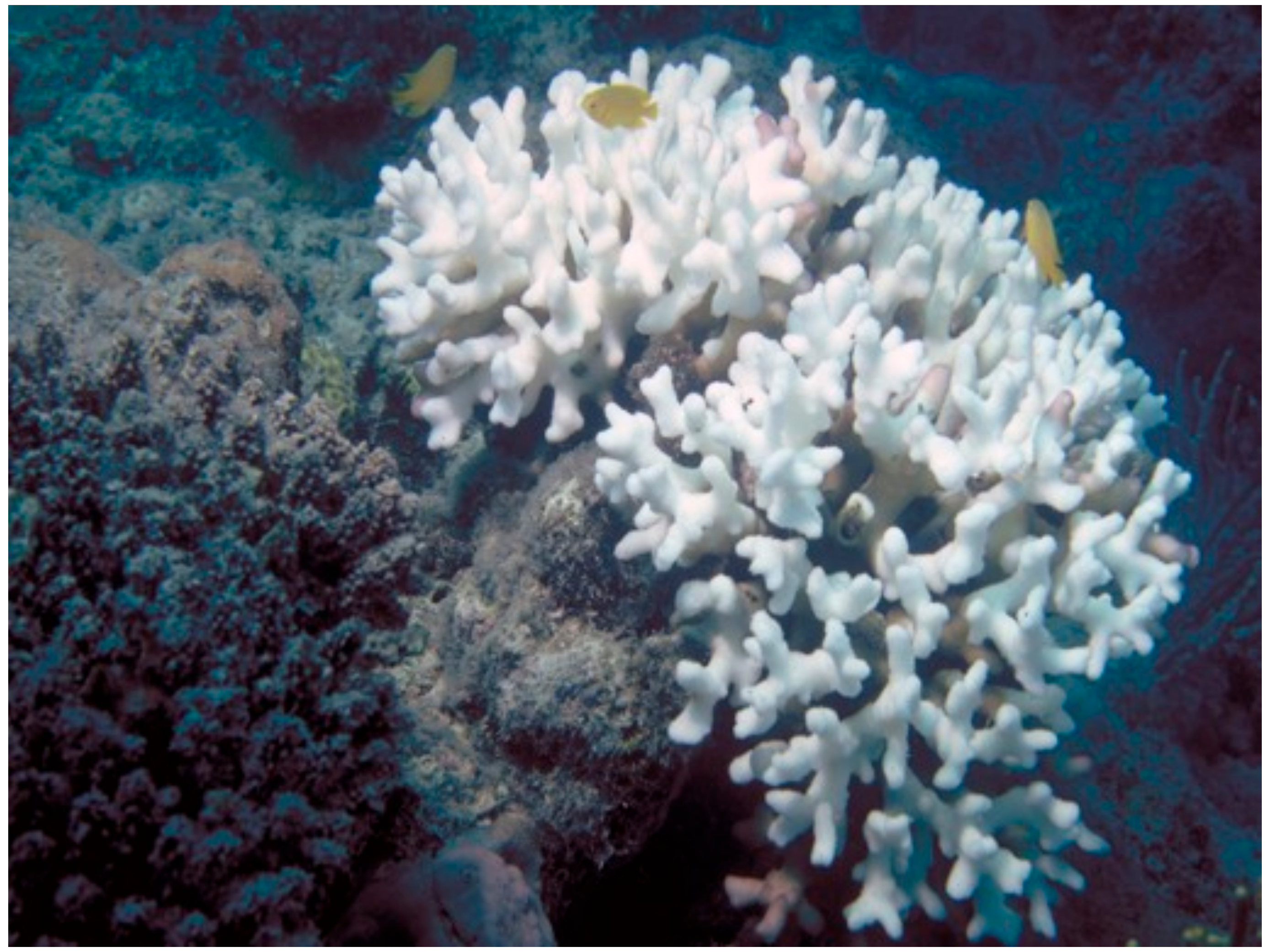
BioMath Explorations

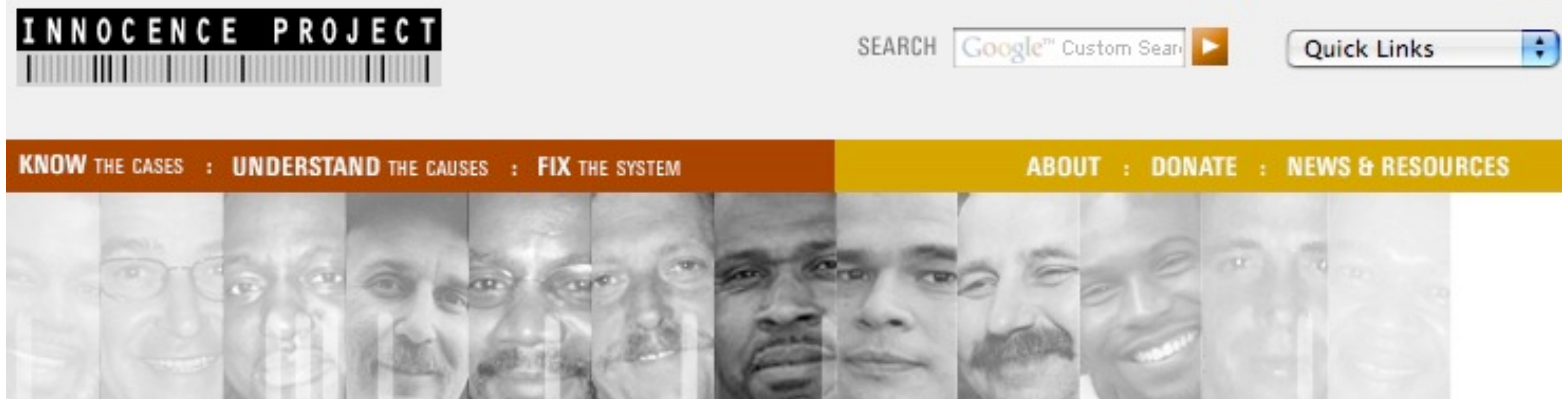
BioMath Exploration 6.3

How can you fit exponential curves to data?



Ethical, Legal and Social Implications





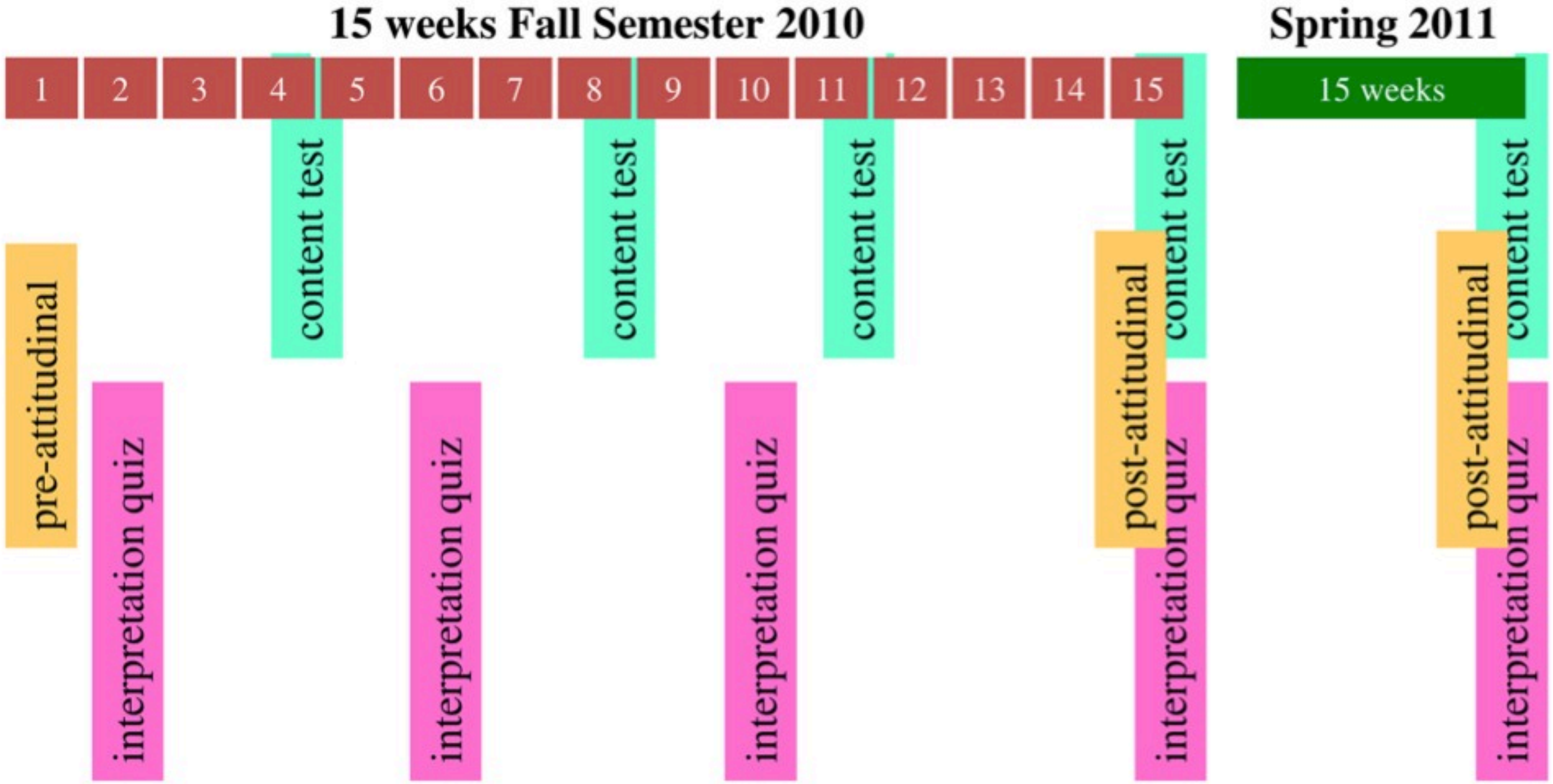
Are religion and evolution compatible?

Is science possible if you are uncertain about what is true?

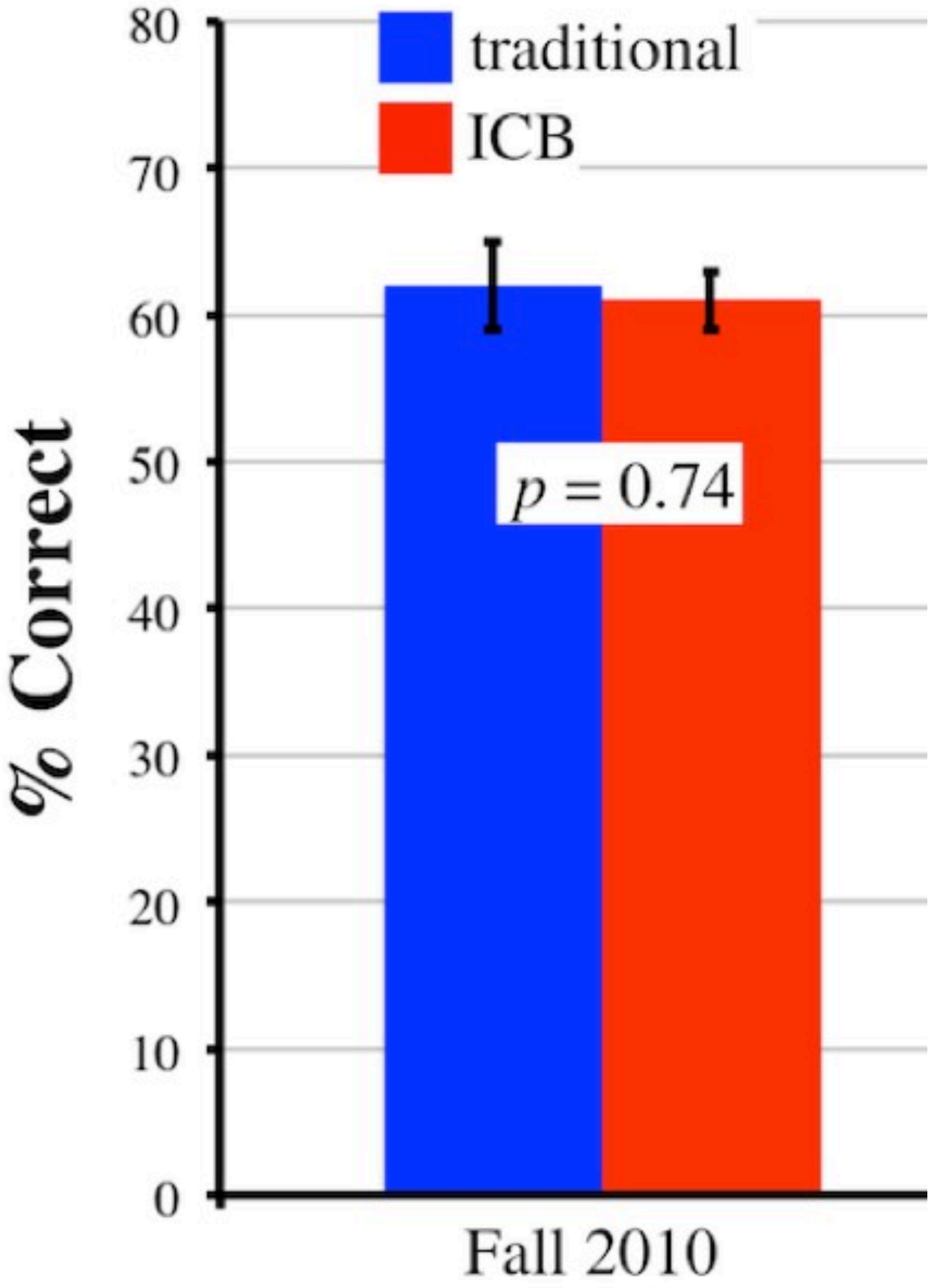
Does basic biology have any impact on the real world?

Who owns your DNA?

Student Outcomes and Data

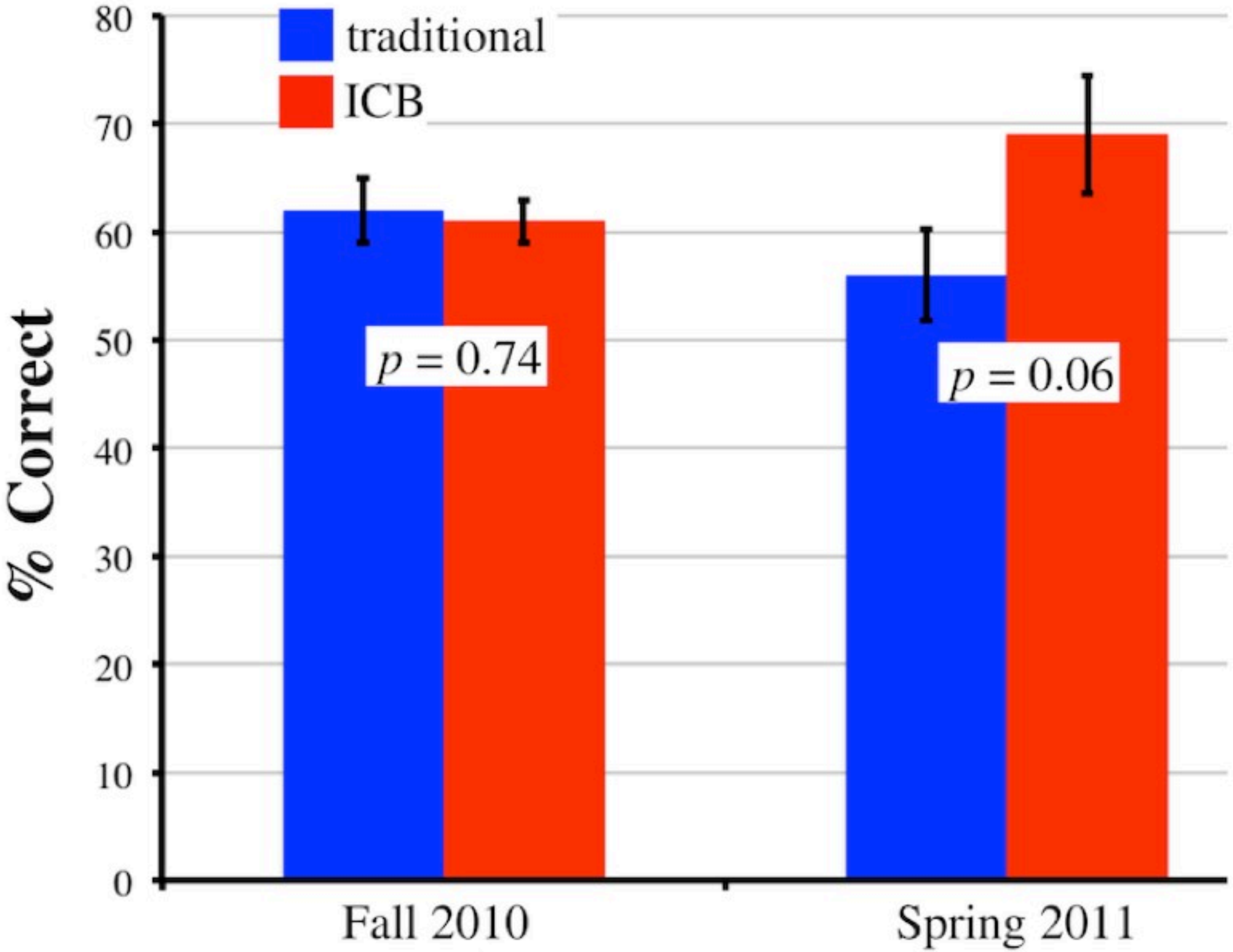


Do *ICB* students learn less than others?



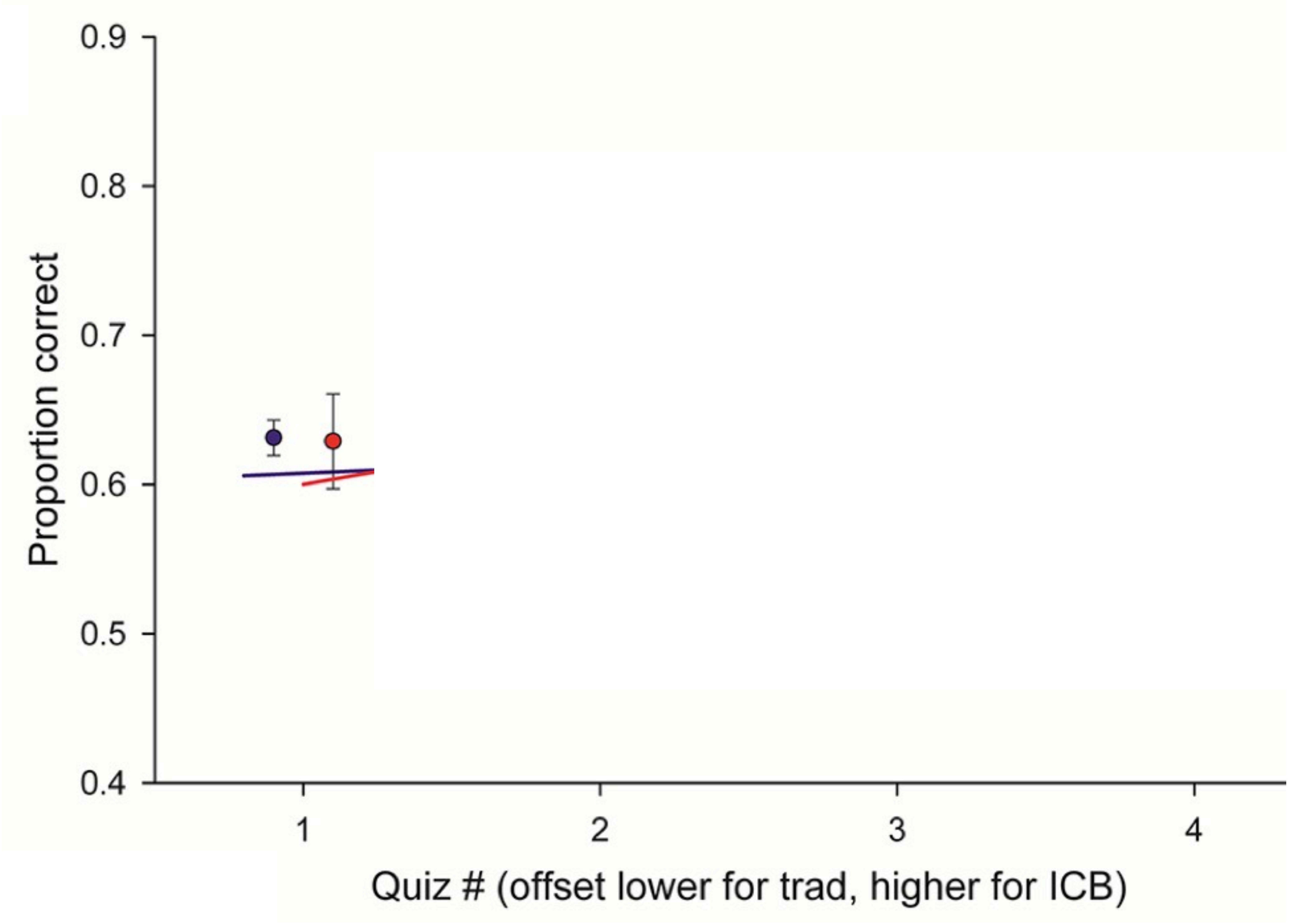
no

Do *ICB* students learn **more** than others?



yes?

Can *ICB* students interpret data better?



yes

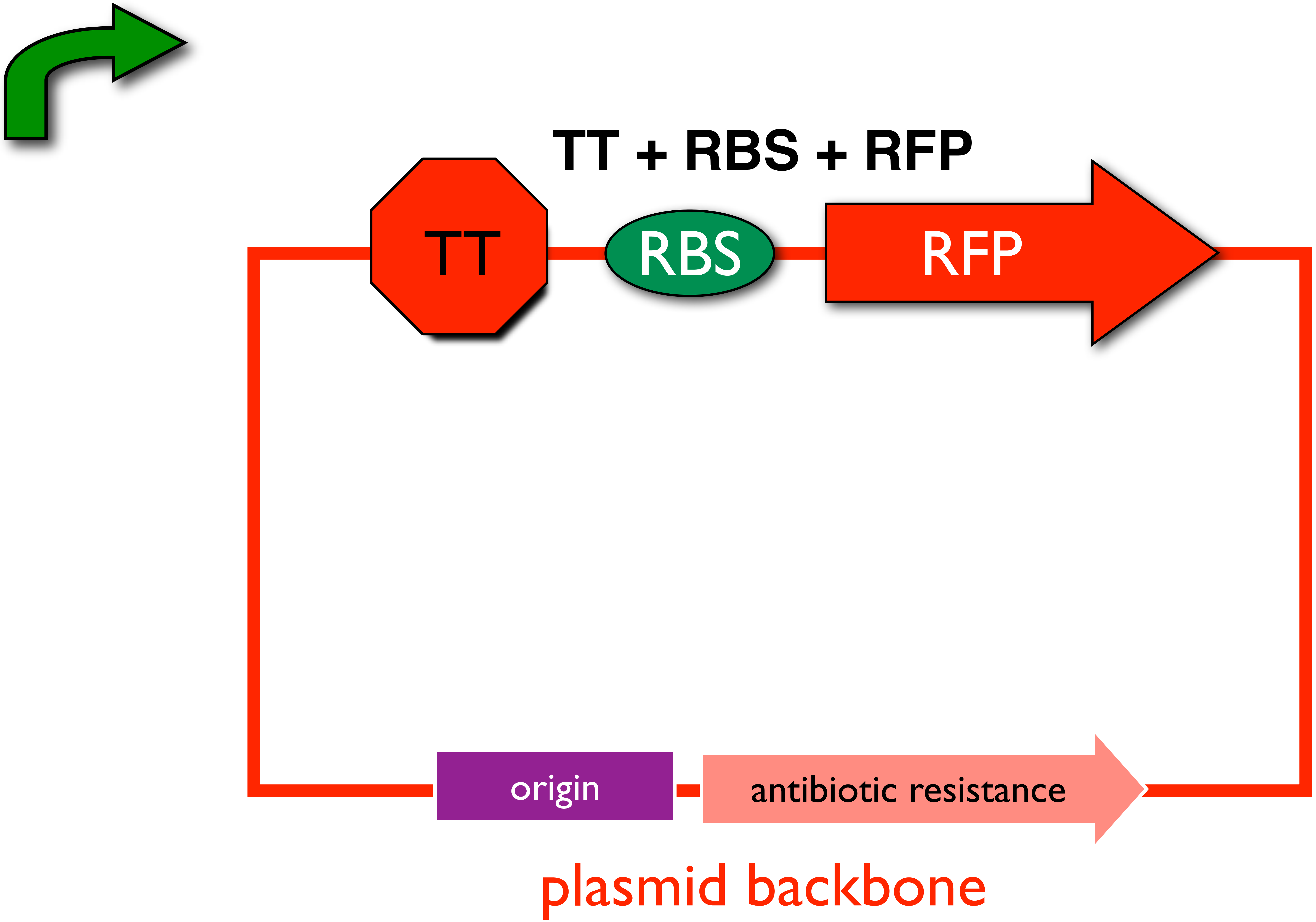
Can intro bio students do real research and characterize a new promoter?

What is Synthetic Biology?

Implementation of engineering principles and mathematical modeling to the design and construction of biological parts, devices, and systems with applications in energy, medicine, and technology.

www.bio.davidson.edu/projects/gcat/Synthetic/What_Is_SynBio.html

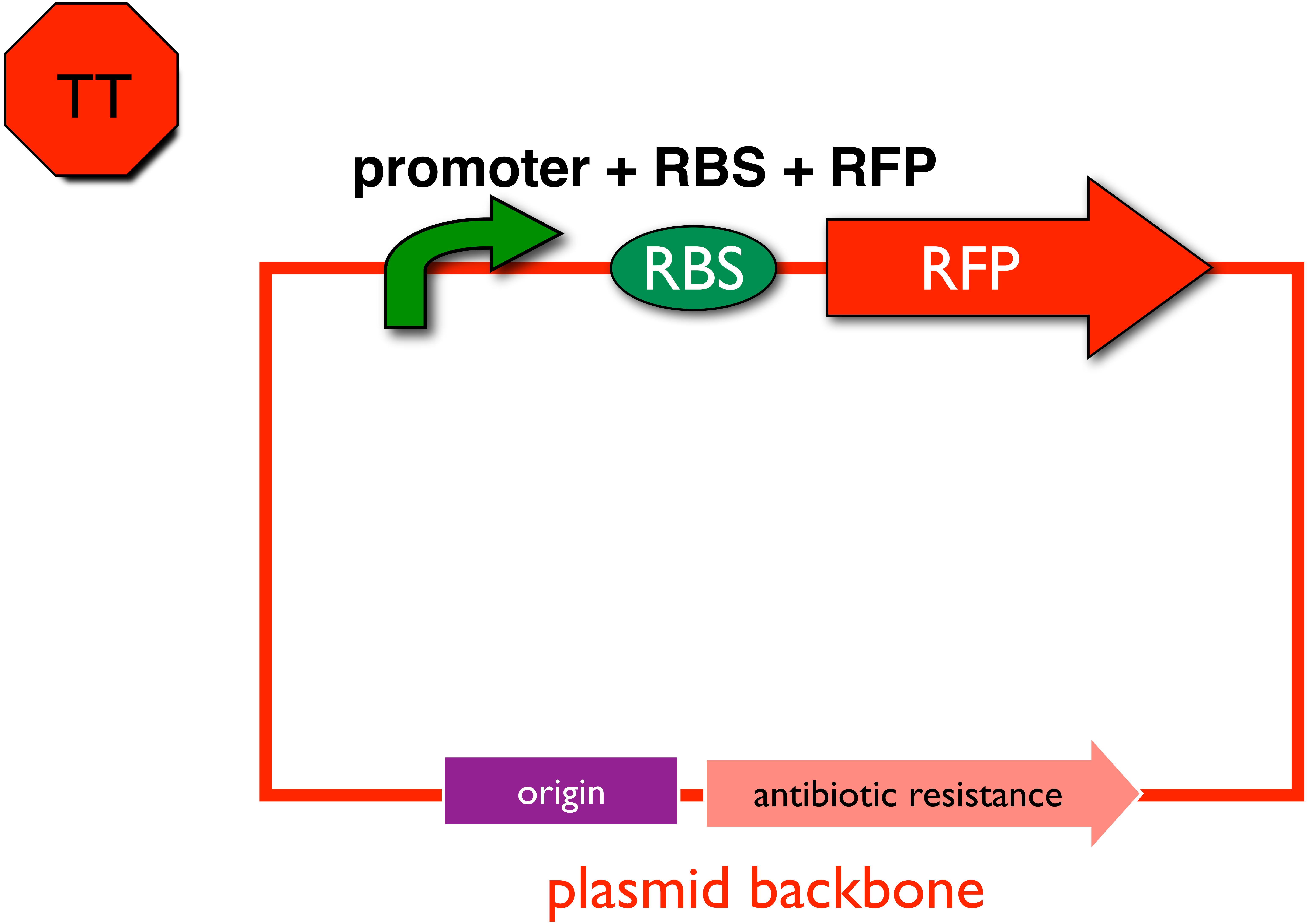
Golden Gate Assembly Method



TT + RBS + RFP

plasmid backbone

Golden Gate Assembly Method



Eco RI

GAATTC
CTTAAG

palindrome

type II

Bsa I

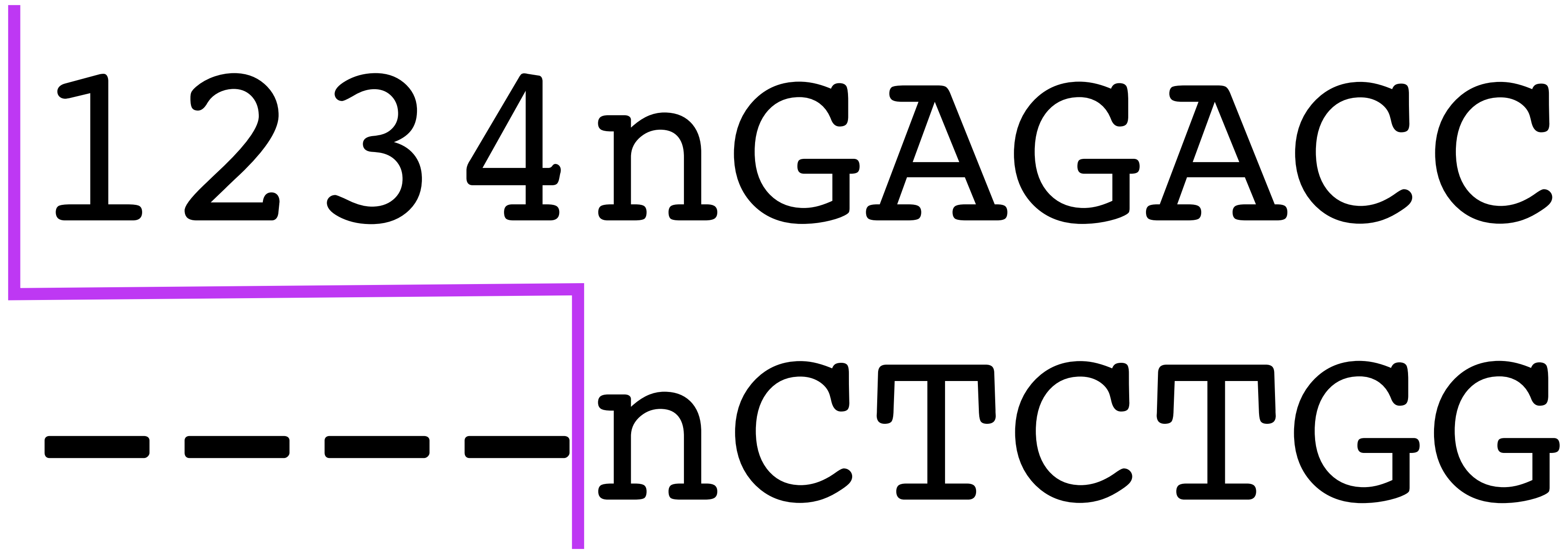
GAGACC

CTCTGG

not a palindrome

type IIs

Bsa I



```
1234nGAGACC
----nCTCTGG
```

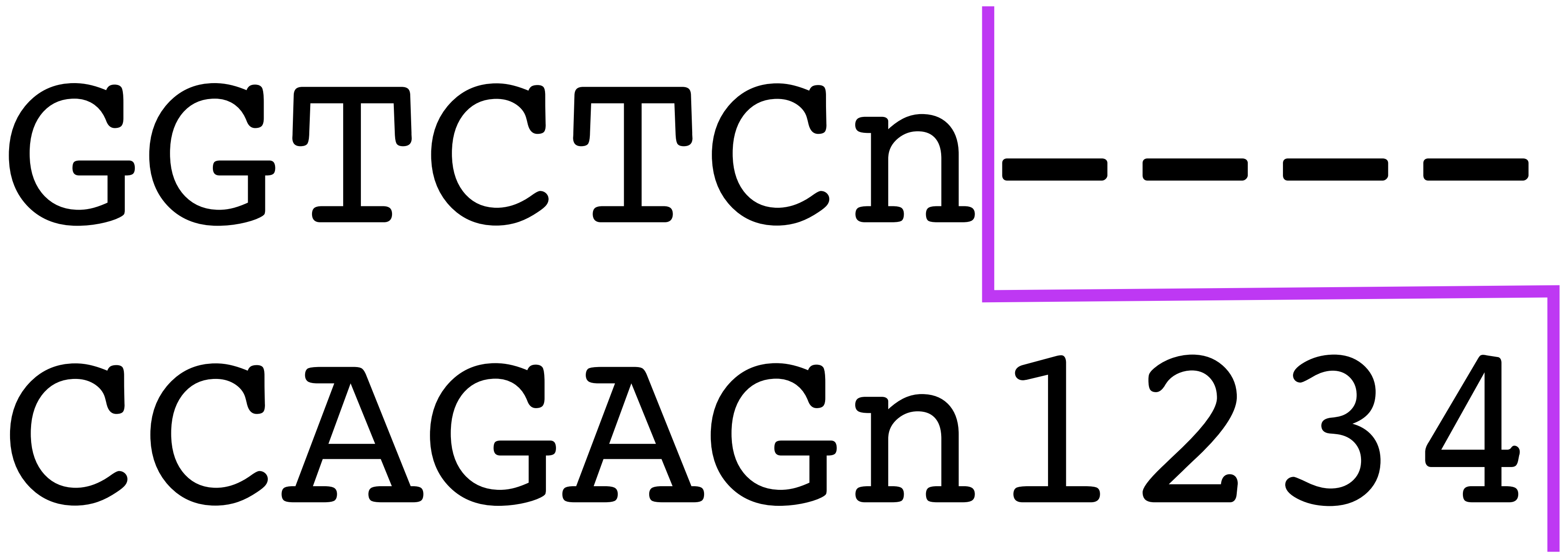
type IIs

Bsa I

```
        1234nGAGACC
----        nCTCTGG
```

type IIs

Bsa I



GGTCTCn----

CCAGAGn1234

type IIs

Bsa I

```
GGTCTCn                    ----
CCAGAGn1234
```

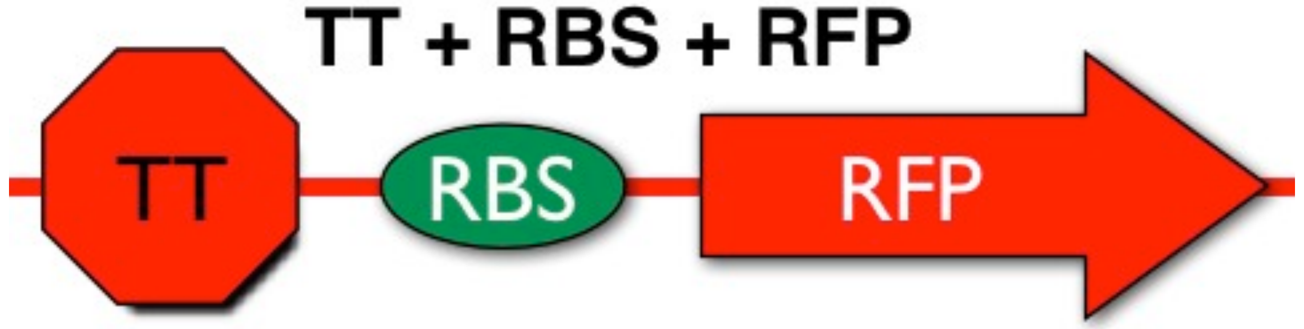
type IIs

Bsa I

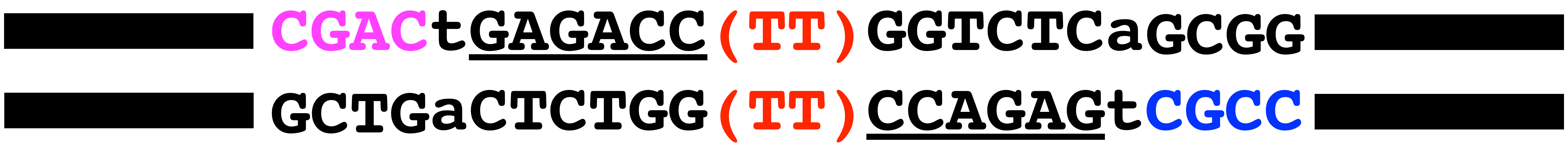
CGACtGAGACC(TT)GGTCTCaGCGG

GCTGaCTCTGG(TT)CCAGAGtCGCC

Bsa I

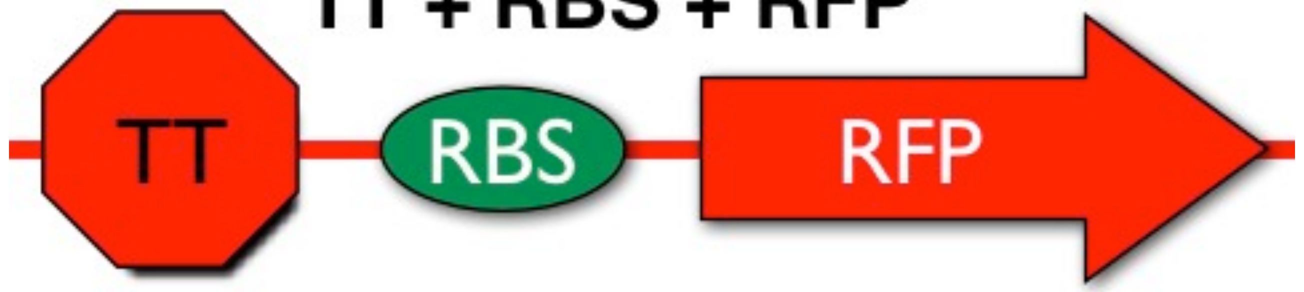


CGACtGAGACC(TT)GGTCTCa
aCTCTGG(TT)CCAGAGtCGCC



CGACtGAGACC(TT)GGTCTCaGCGG

GCTGaCTCTGG(TT)CCAGAGtCGCC


TT + RBS + RFP

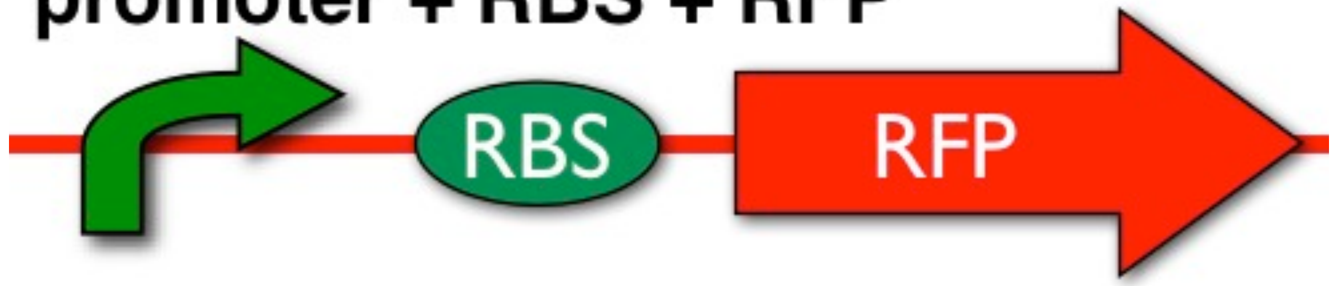
CGACtGAGACC(TT)GGTCTCa
aCTCTGG(TT)CCAGAGtCGCC

CGAC(promoter)
(promoter)CGCC

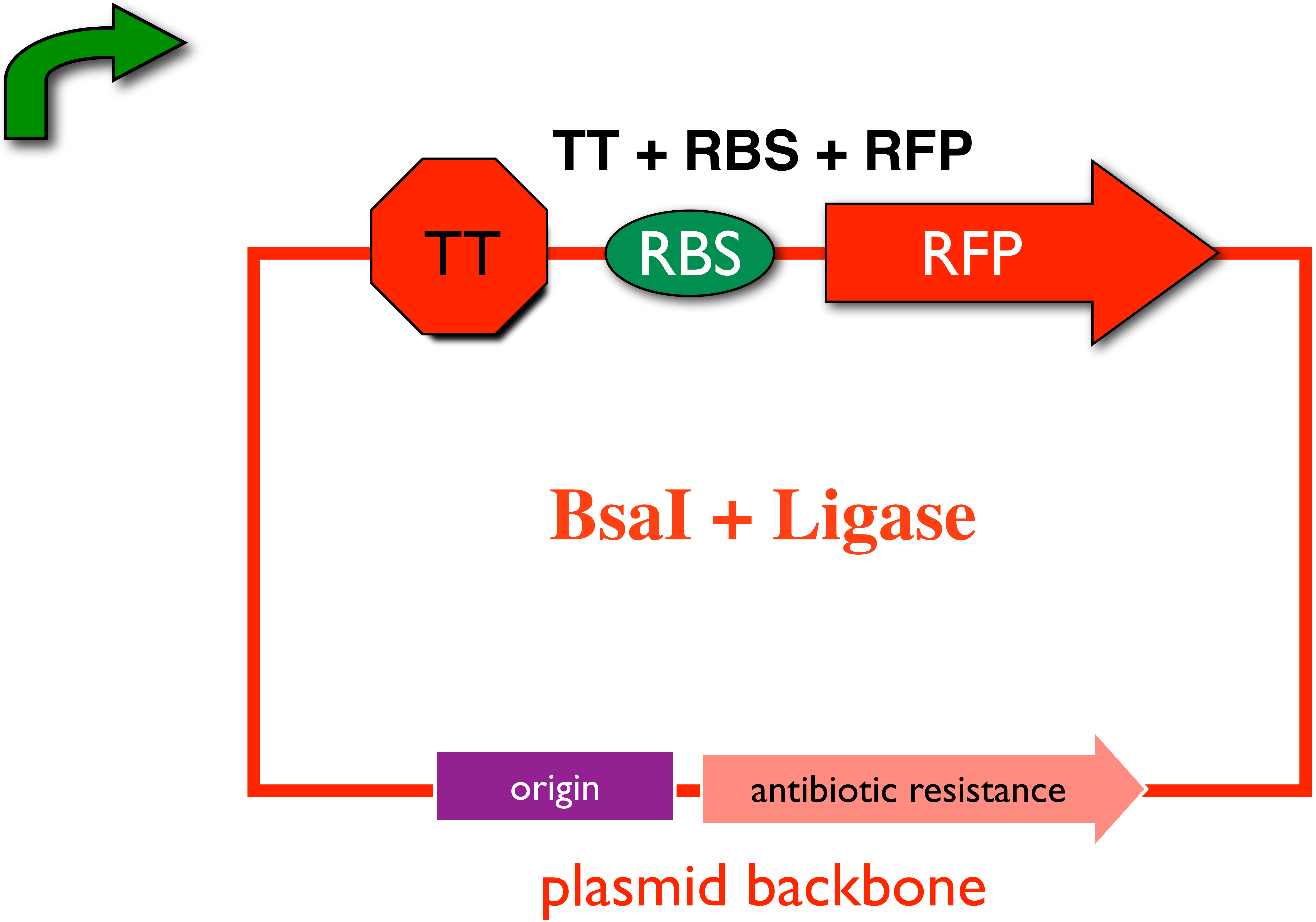
CGACtGAGACC(TT)GGTCTCa
aCTCTGG(TT)CCAGAGtCGCC


CGAC(promoter)GCGG
GCTG(promoter)CGCC

promoter + RBS + RFP



GGA Ligation Method

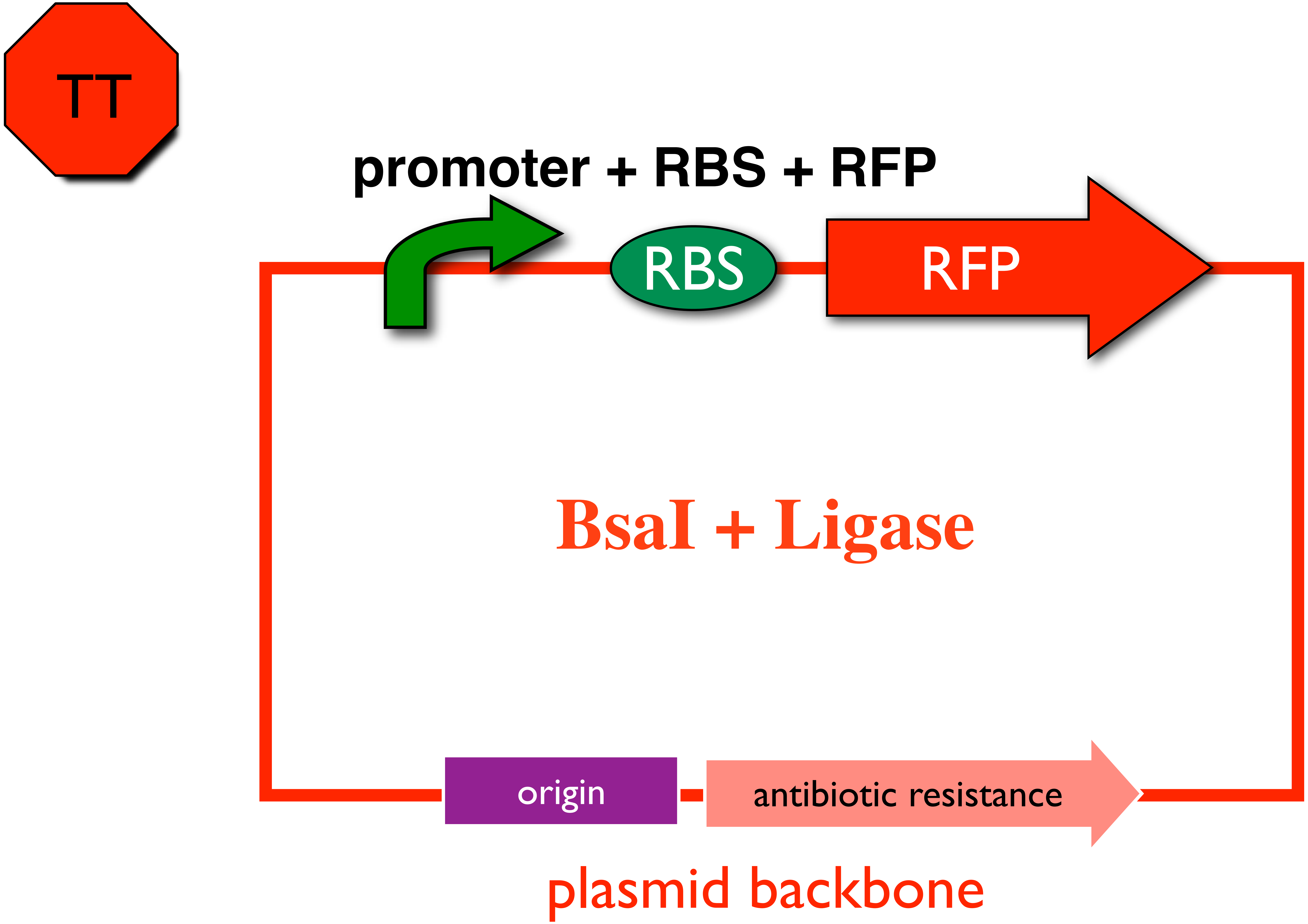


TT + RBS + RFP

BsaI + Ligase

plasmid backbone

GGA Ligation Method

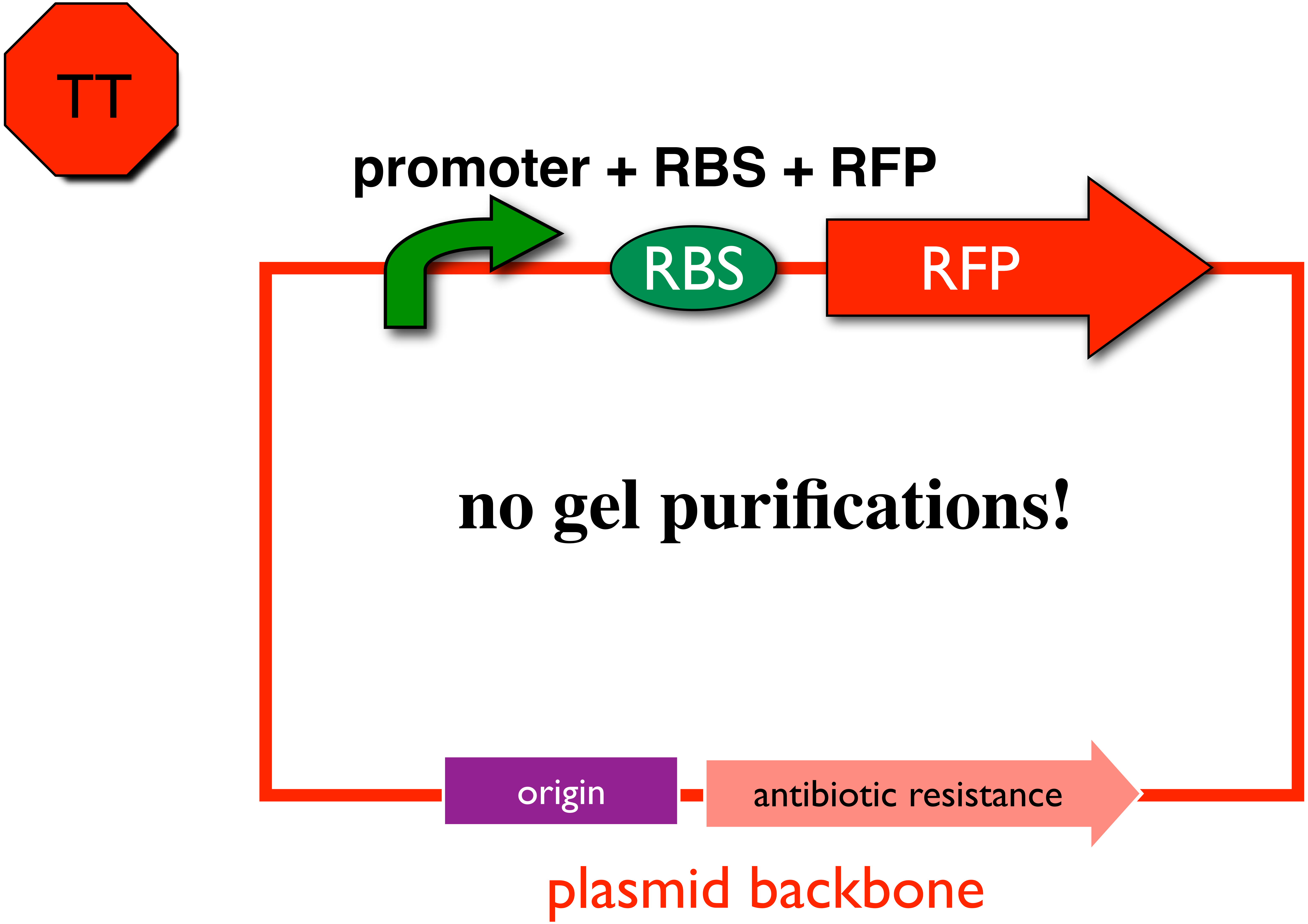


promoter + RBS + RFP

BsaI + Ligase

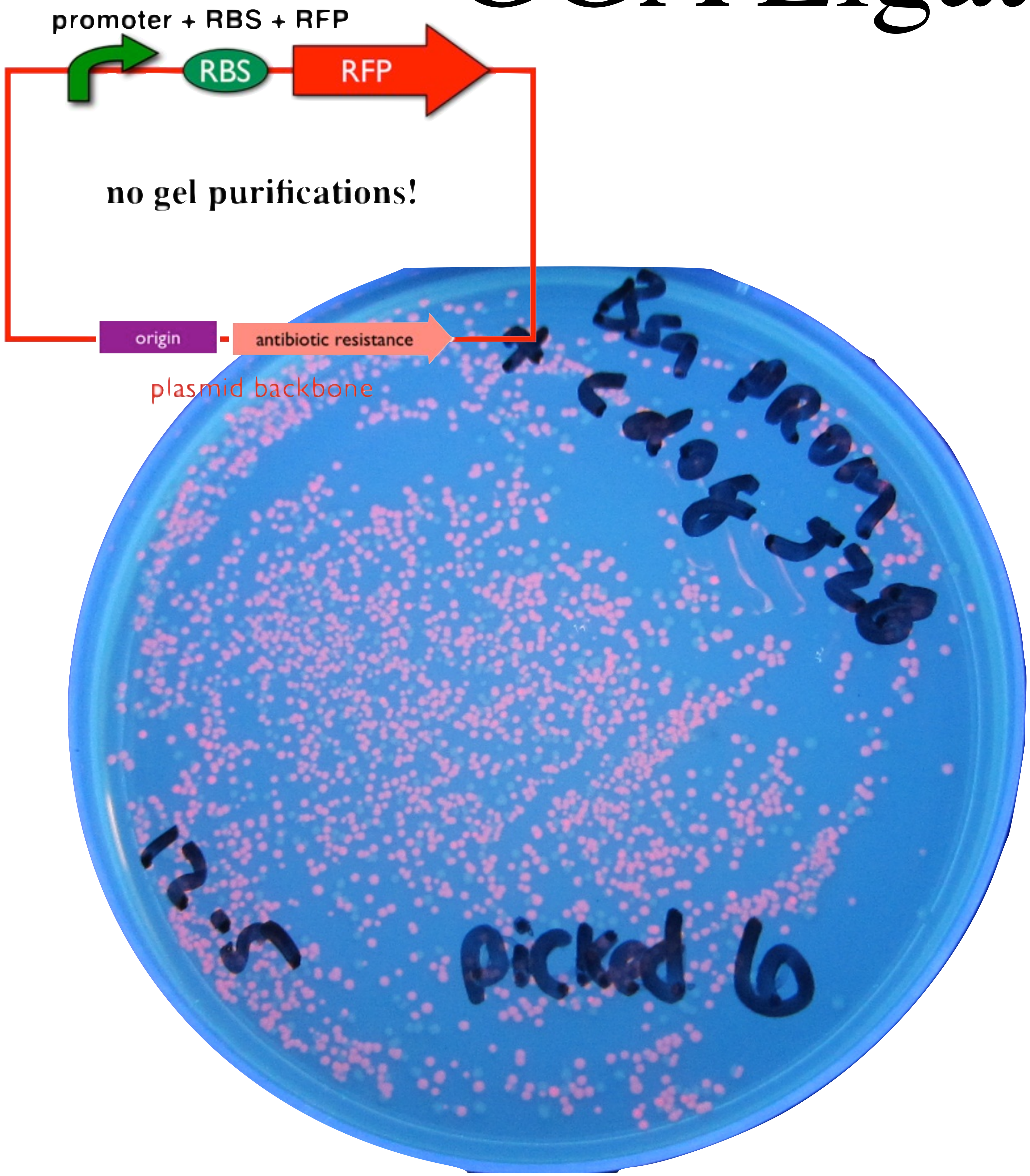
plasmid backbone

GGA Ligation Method



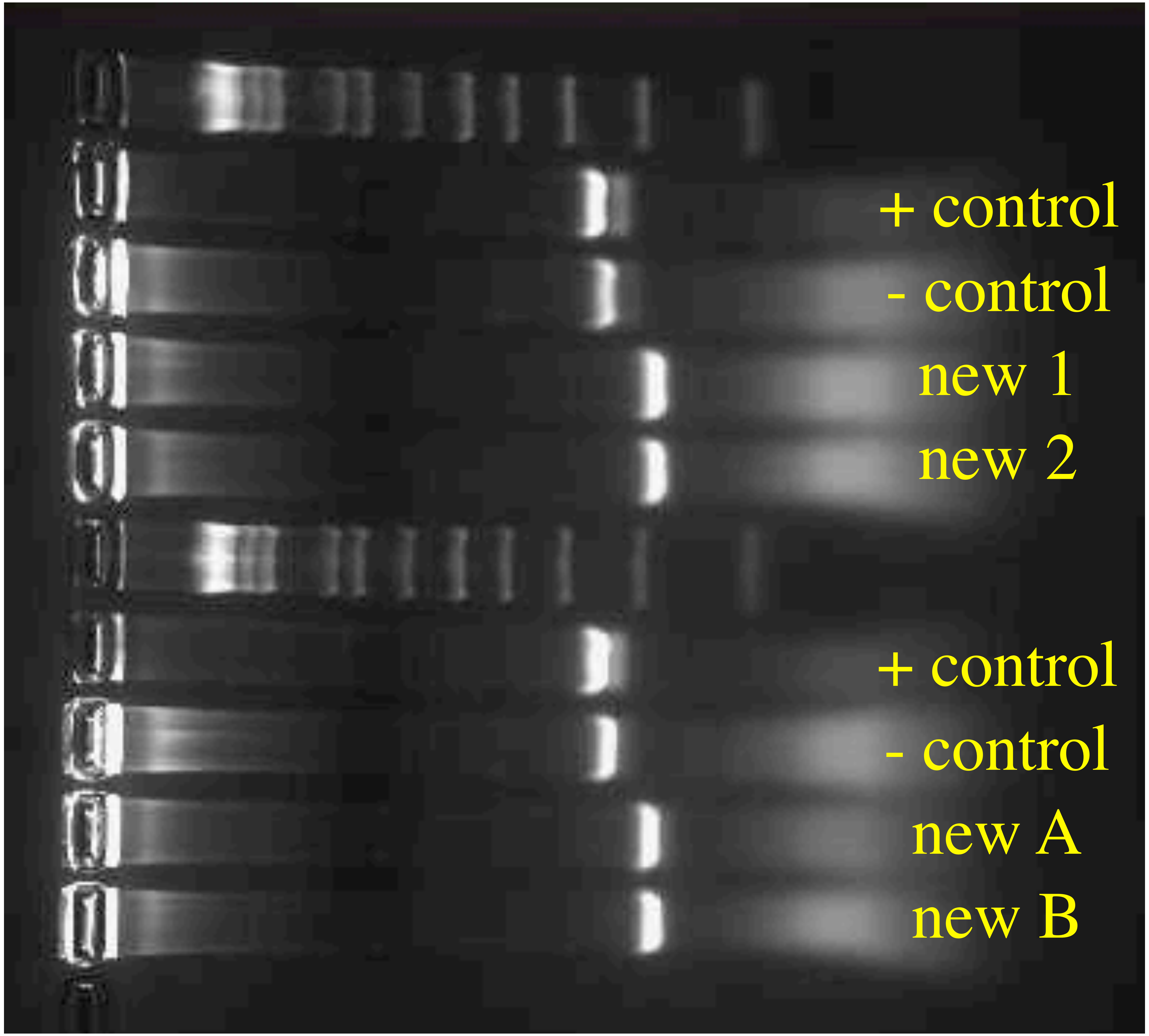
no gel purifications!

GGA Ligation Method



no gel purifications!

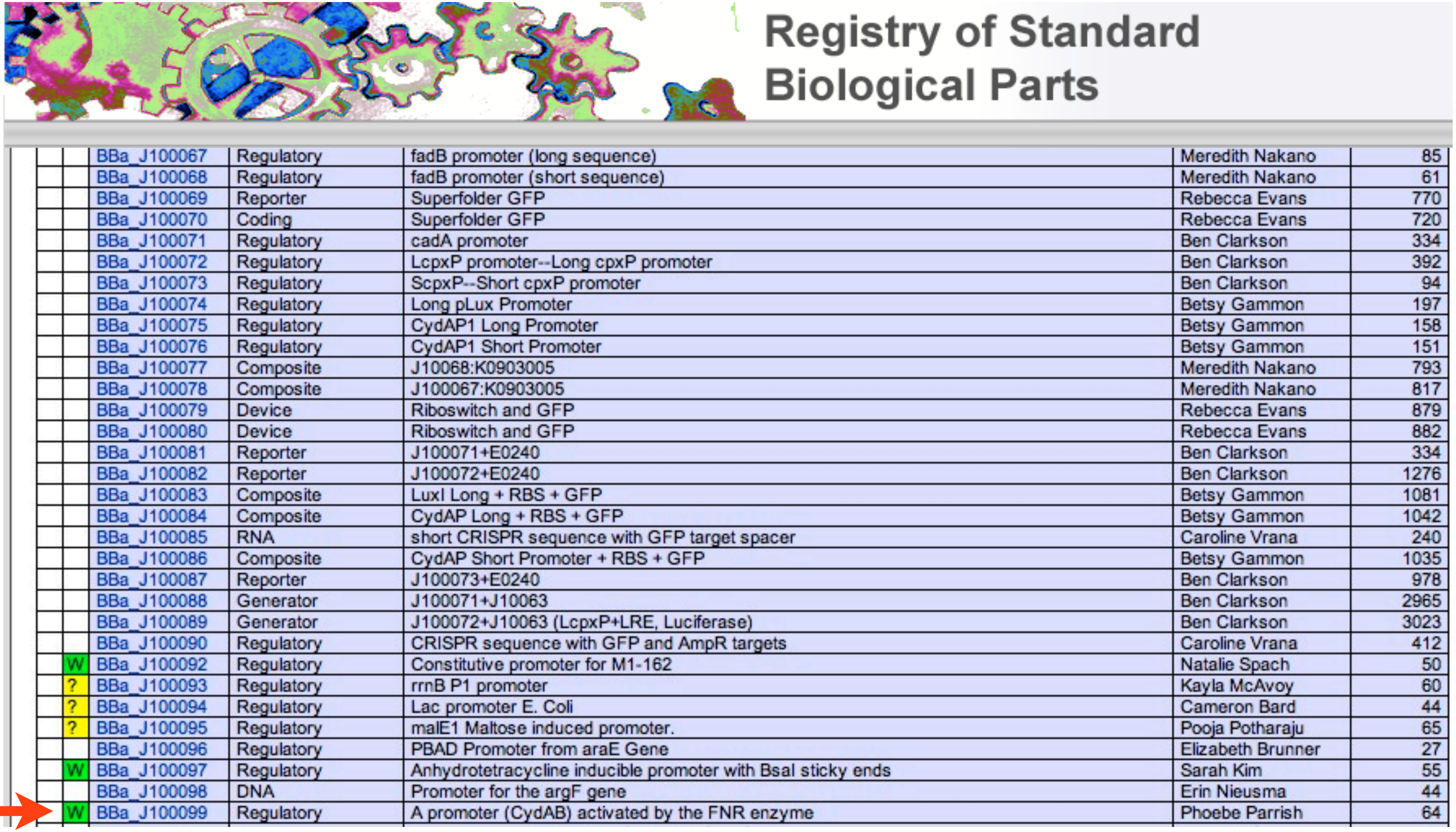
Student Sample, September 2012



+ control
- control
new 1
new 2

+ control
- control
new A
new B

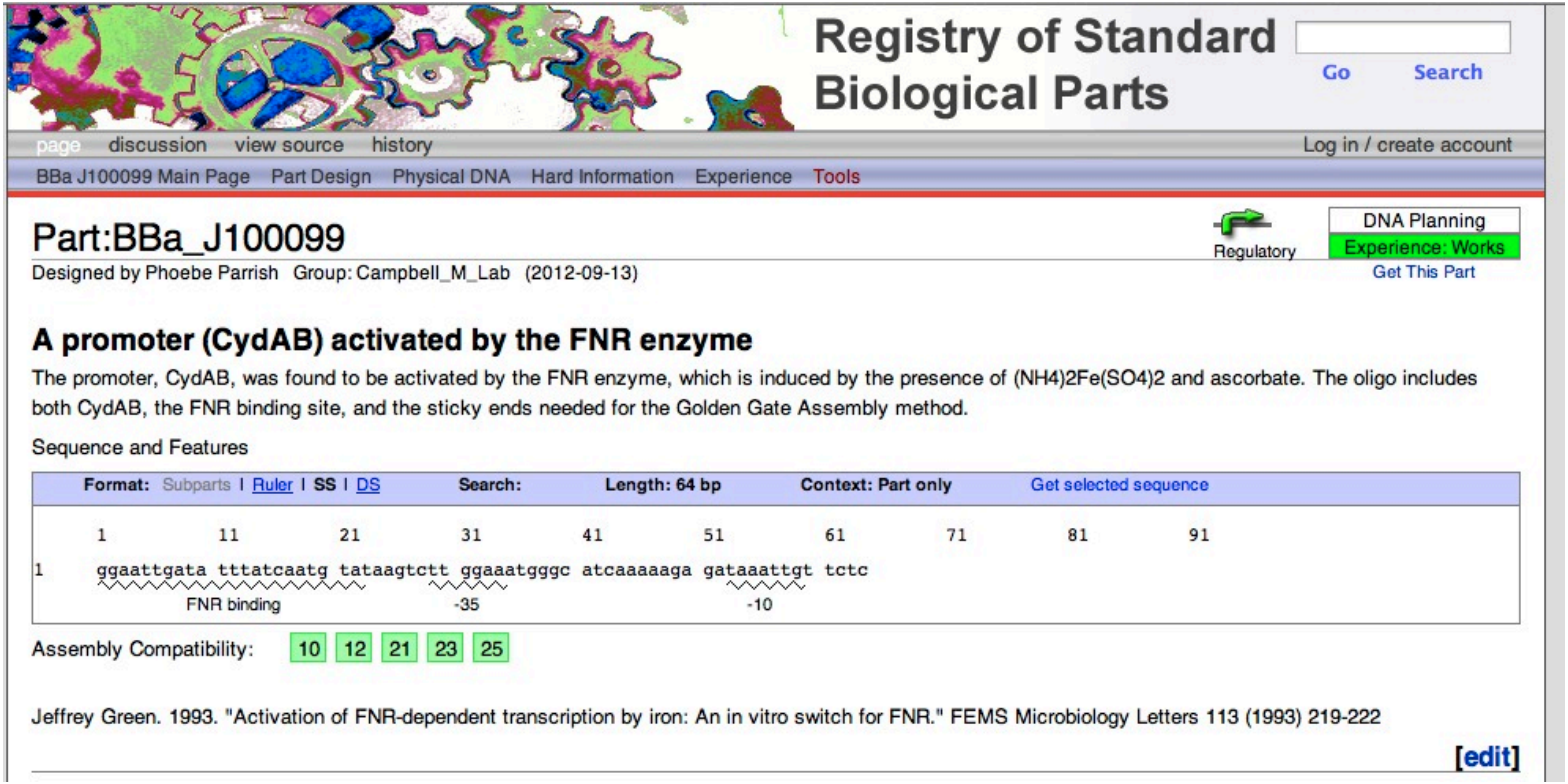
Student Sample, September 2012



Registry of Standard Biological Parts

		BBa_J100067	Regulatory	fadB promoter (long sequence)	Meredith Nakano	85
		BBa_J100068	Regulatory	fadB promoter (short sequence)	Meredith Nakano	61
		BBa_J100069	Reporter	Superfolder GFP	Rebecca Evans	770
		BBa_J100070	Coding	Superfolder GFP	Rebecca Evans	720
		BBa_J100071	Regulatory	cadA promoter	Ben Clarkson	334
		BBa_J100072	Regulatory	LcpxP promoter--Long cpxP promoter	Ben Clarkson	392
		BBa_J100073	Regulatory	ScpxP--Short cpxP promoter	Ben Clarkson	94
		BBa_J100074	Regulatory	Long pLux Promoter	Betsy Gammon	197
		BBa_J100075	Regulatory	CydAP1 Long Promoter	Betsy Gammon	158
		BBa_J100076	Regulatory	CydAP1 Short Promoter	Betsy Gammon	151
		BBa_J100077	Composite	J10068:K0903005	Meredith Nakano	793
		BBa_J100078	Composite	J100067:K0903005	Meredith Nakano	817
		BBa_J100079	Device	Riboswitch and GFP	Rebecca Evans	879
		BBa_J100080	Device	Riboswitch and GFP	Rebecca Evans	882
		BBa_J100081	Reporter	J100071+E0240	Ben Clarkson	334
		BBa_J100082	Reporter	J100072+E0240	Ben Clarkson	1276
		BBa_J100083	Composite	LuxI Long + RBS + GFP	Betsy Gammon	1081
		BBa_J100084	Composite	CydAP Long + RBS + GFP	Betsy Gammon	1042
		BBa_J100085	RNA	short CRISPR sequence with GFP target spacer	Caroline Vrana	240
		BBa_J100086	Composite	CydAP Short Promoter + RBS + GFP	Betsy Gammon	1035
		BBa_J100087	Reporter	J100073+E0240	Ben Clarkson	978
		BBa_J100088	Generator	J100071+J10063	Ben Clarkson	2965
		BBa_J100089	Generator	J100072+J10063 (LcpxP+LRE, Luciferase)	Ben Clarkson	3023
		BBa_J100090	Regulatory	CRISPR sequence with GFP and AmpR targets	Caroline Vrana	412
	W	BBa_J100092	Regulatory	Constitutive promoter for M1-162	Natalie Spach	50
	?	BBa_J100093	Regulatory	rrnB P1 promoter	Kayla McAvoy	60
	?	BBa_J100094	Regulatory	Lac promoter E. Coli	Cameron Bard	44
	?	BBa_J100095	Regulatory	malE1 Maltose induced promoter.	Pooja Potharaju	65
		BBa_J100096	Regulatory	PBAD Promoter from araE Gene	Elizabeth Brunner	27
	W	BBa_J100097	Regulatory	Anhydrotetracycline inducible promoter with BsaI sticky ends	Sarah Kim	55
		BBa_J100098	DNA	Promoter for the argF gene	Erin Nieusma	44
	W	BBa_J100099	Regulatory	A promoter (CydAB) activated by the FNR enzyme	Phoebe Parrish	64

Student Sample, September 2012



Part:BBa_J100099

Designed by Phoebe Parrish Group: Campbell_M_Lab (2012-09-13)

A promoter (CydAB) activated by the FNR enzyme

The promoter, CydAB, was found to be activated by the FNR enzyme, which is induced by the presence of $(NH_4)_2Fe(SO_4)_2$ and ascorbate. The oligo includes both CydAB, the FNR binding site, and the sticky ends needed for the Golden Gate Assembly method.

Sequence and Features

Format: Subparts | Ruler | **SS** | DS **Search:** **Length: 64 bp** **Context: Part only** Get selected sequence

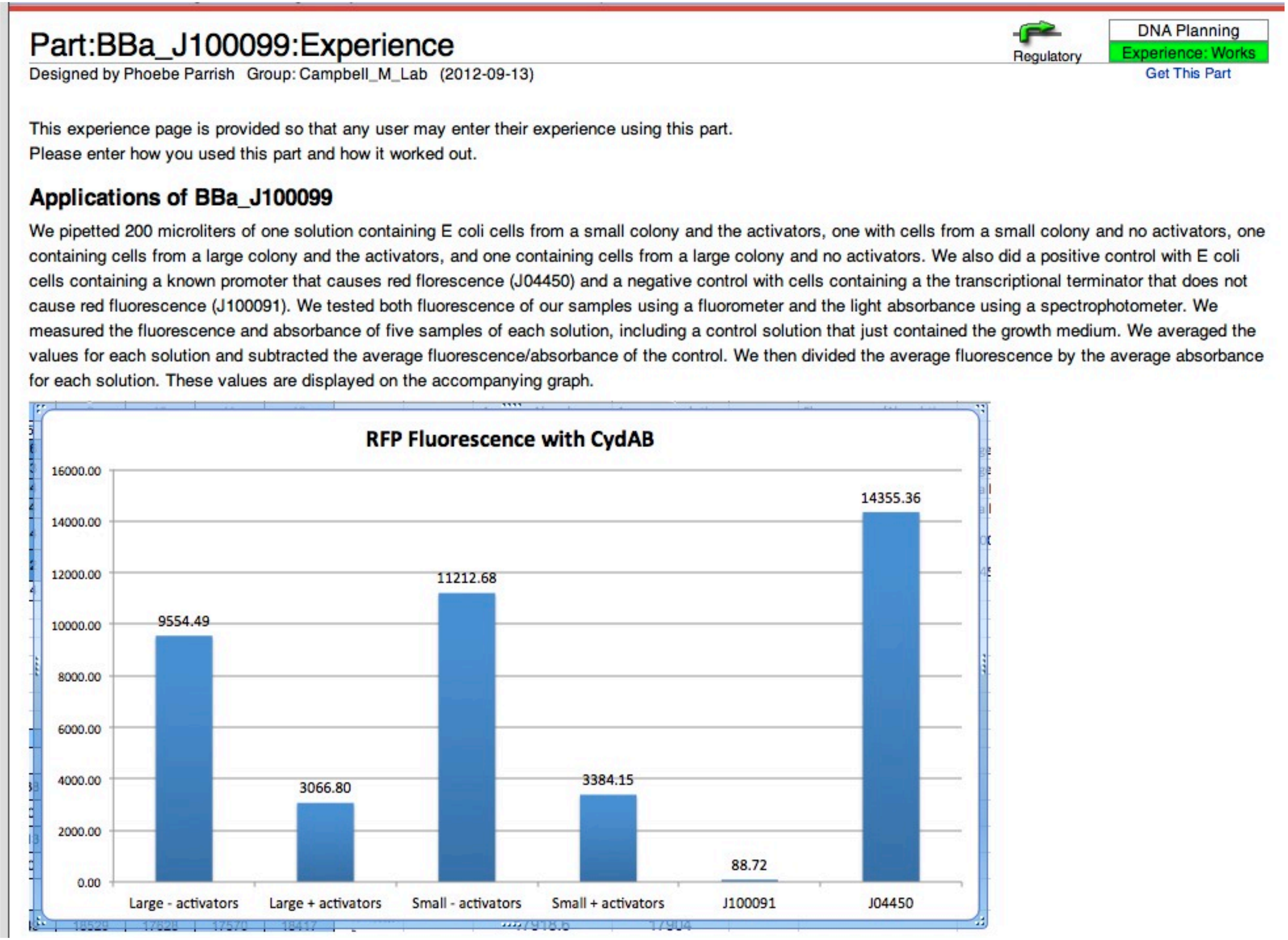
```
        1          11         21         31         41         51         61         71         81         91

1       ggaattgata tttatcaatg tataagtctt ggaaatgggc atcaaaaaga gataaattgt tctc
                  FNR binding                -35                       -10
```

Assembly Compatibility: 10 12 21 23 25

Jeffrey Green. 1993. "Activation of FNR-dependent transcription by iron: An in vitro switch for FNR." FEMS Microbiology Letters 113 (1993) 219-222

Student Sample, September 2012



Part:BBa_J100099:Experience

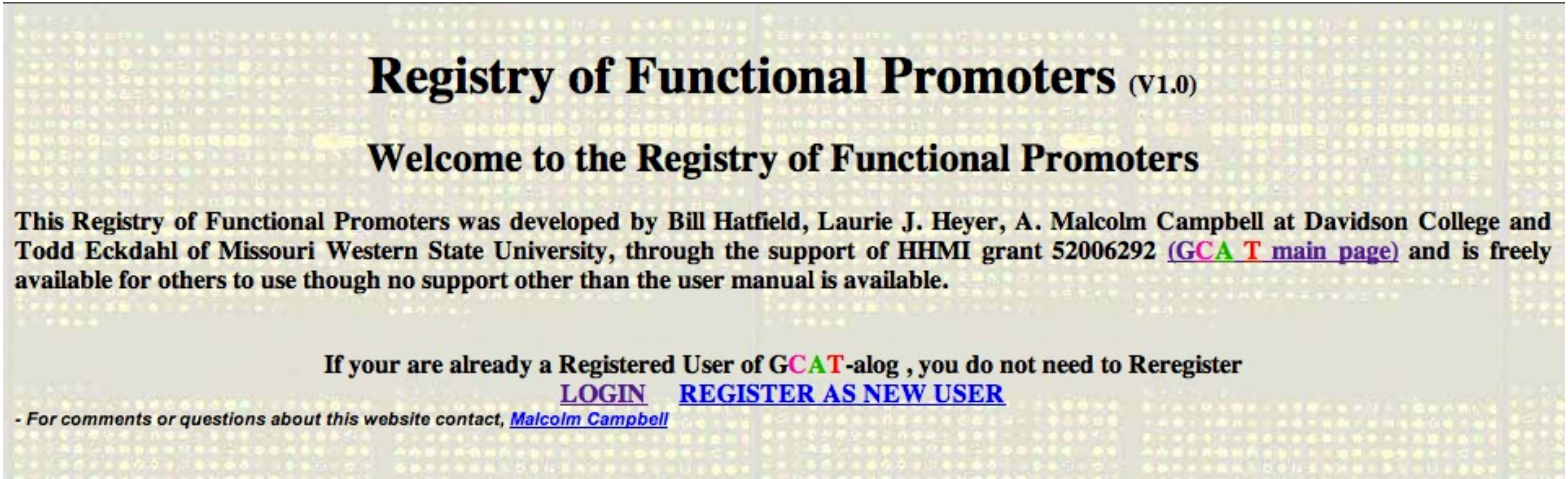
Designed by Phoebe Parrish   Group: Campbell_M_Lab   (2012-09-13)

This experience page is provided so that any user may enter their experience using this part.
Please enter how you used this part and how it worked out.

Applications of BBa_J100099

We pipetted 200 microliters of one solution containing E coli cells from a small colony and the activators, one with cells from a small colony and no activators, one containing cells from a large colony and the activators, and one containing cells from a large colony and no activators. We also did a positive control with E coli cells containing a known promoter that causes red florescence (J04450) and a negative control with cells containing a the transcriptional terminator that does not cause red fluorescence (J100091). We tested both fluorescence of our samples using a fluorometer and the light absorbance using a spectrophotometer. We measured the fluorescence and absorbance of five samples of each solution, including a control solution that just contained the growth medium. We averaged the values for each solution and subtracted the average fluorescence/absorbance of the control. We then divided the average fluorescence by the average absorbance for each solution. These values are displayed on the accompanying graph.

Registry of Functional Promoters (RFP)



Registry of Functional Promoters (V1.0)

Welcome to the Registry of Functional Promoters

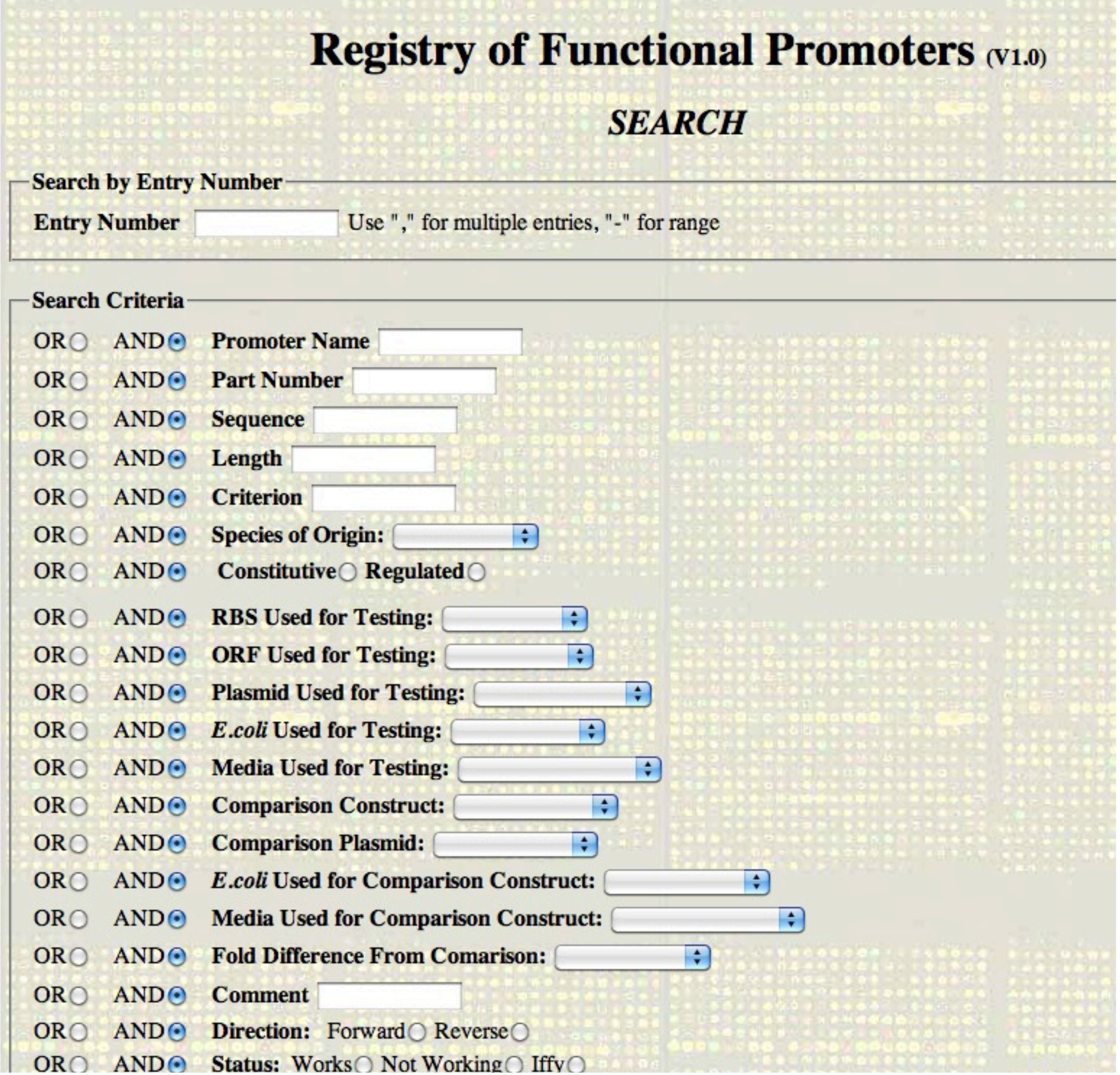
This Registry of Functional Promoters was developed by Bill Hatfield, Laurie J. Heyer, A. Malcolm Campbell at Davidson College and Todd Eckdahl of Missouri Western State University, through the support of HHMI grant 52006292 (GCA T main page) and is freely available for others to use though no support other than the user manual is available.

If your are already a Registered User of GCAT-alog , you do not need to Reregister

- For comments or questions about this website contact, Malcolm Campbell

gcat.davidson.edu/RFP/

Registry of Functional Promoters (RFP)



Registry of Functional Promoters (V1.0)

SEARCH

Search by Entry Number

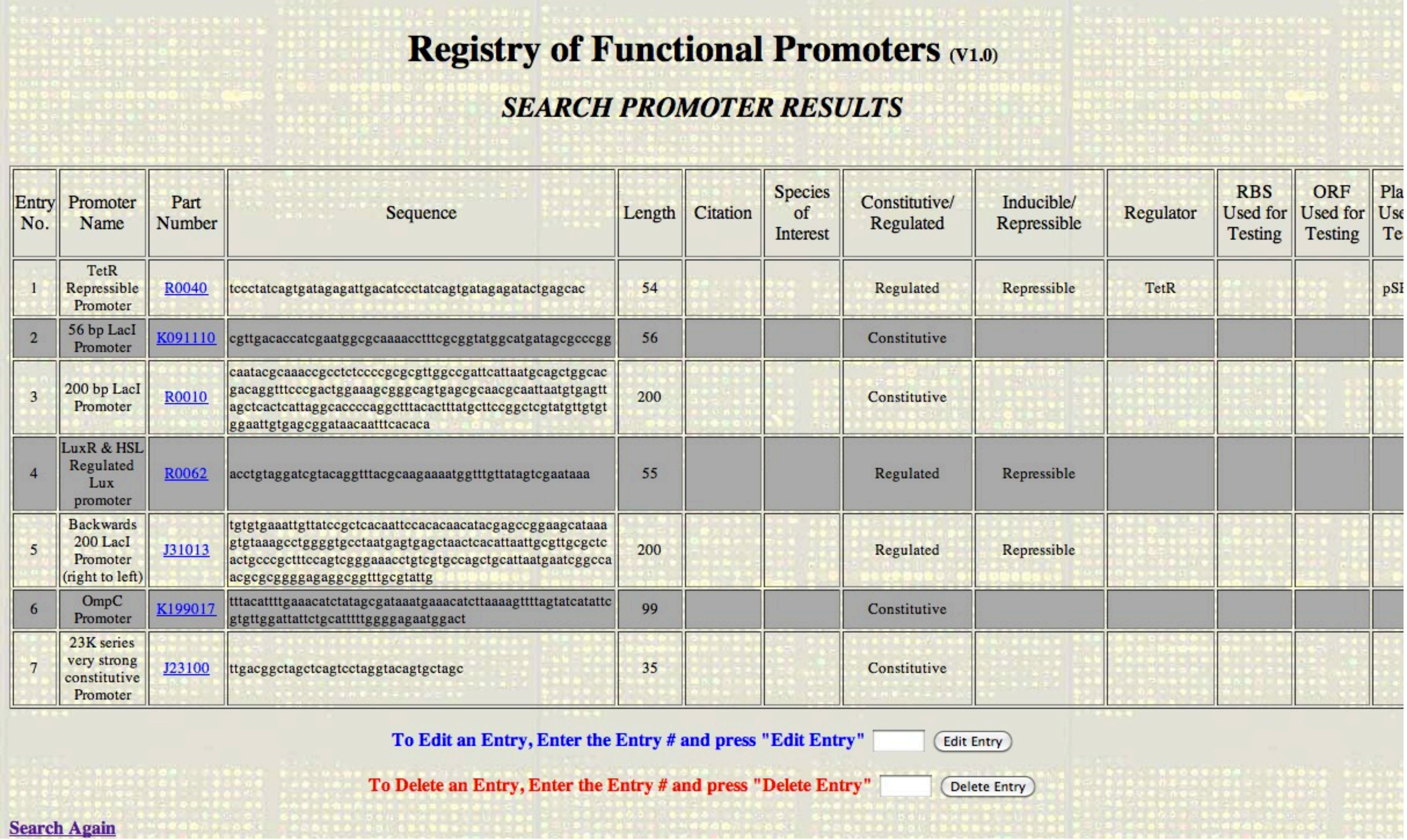
Entry Number Use "," for multiple entries, "-" for range

Search Criteria

OR○ AND● **Promoter Name**
OR○ AND● **Part Number**
OR○ AND● **Sequence**
OR○ AND● **Length**
OR○ AND● **Criterion**
OR○ AND● **Species of Origin:**
OR○ AND● **Constitutive**○ **Regulated**○
OR○ AND● **RBS Used for Testing:**
OR○ AND● **ORF Used for Testing:**
OR○ AND● **Plasmid Used for Testing:**
OR○ AND● ***E.coli* Used for Testing:**
OR○ AND● **Media Used for Testing:**
OR○ AND● **Comparison Construct:**
OR○ AND● **Comparison Plasmid:**
OR○ AND● ***E.coli* Used for Comparison Construct:**
OR○ AND● **Media Used for Comparison Construct:**
OR○ AND● **Fold Difference From Comarison:**
OR○ AND● **Comment**
OR○ AND● **Direction:** Forward○ Reverse○
OR○ AND● **Status:** Works○ Not Working○ Iffy○

gcat.davidson.edu/RFP/

Registry of Functional Promoters (RFP)



Registry of Functional Promoters (V1.0)

SEARCH PROMOTER RESULTS

Entry No.	Promoter Name	Part Number	Sequence	Length	Citation	Species of Interest	Constitutive/ Regulated	Inducible/ Repressible	Regulator	RBS Used for Testing	ORF Used for Testing	Pla Us Te
1	TetR Repressible Promoter	R0040	tccctatcagtgatagagattgacatccctatcagtgatagagatactgagcac	54			Regulated	Repressible	TetR			pSI
2	56 bp LacI Promoter	K091110	cgttgacaccatcgaatggcgcaaaacctttcgcggtatggcatgatagcgcccgg	56			Constitutive					
3	200 bp LacI Promoter	R0010	caatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcac gacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagtt agctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgt ggaattgtgagcggataacaatttcacaca	200			Constitutive					
4	LuxR & HSL Regulated Lux promoter	R0062	acctgtaggatcgtacaggtttacgcaagaaaatggtttgttatagtcgaataaa	55			Regulated	Repressible				
5	Backwards 200 LacI Promoter (right to left)	J31013	tgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaa gtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctc actgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcca acgcgcggggagaggcggtttgcgtattg	200			Regulated	Repressible				
6	OmpC Promoter	K199017	tttacattttgaaacatctatagcgataaatgaaacatcttaaaagttttagtatcatattc gtgttggattattctgcattttggggagaatggact	99			Constitutive					
7	23K series very strong constitutive Promoter	J23100	ttgacggctagctcagtcctaggtacagtgctagc	35			Constitutive					

To Edit an Entry, Enter the Entry # and press "Edit Entry" ☐ Edit Entry

To Delete an Entry, Enter the Entry # and press "Delete Entry" ☐ Delete Entry

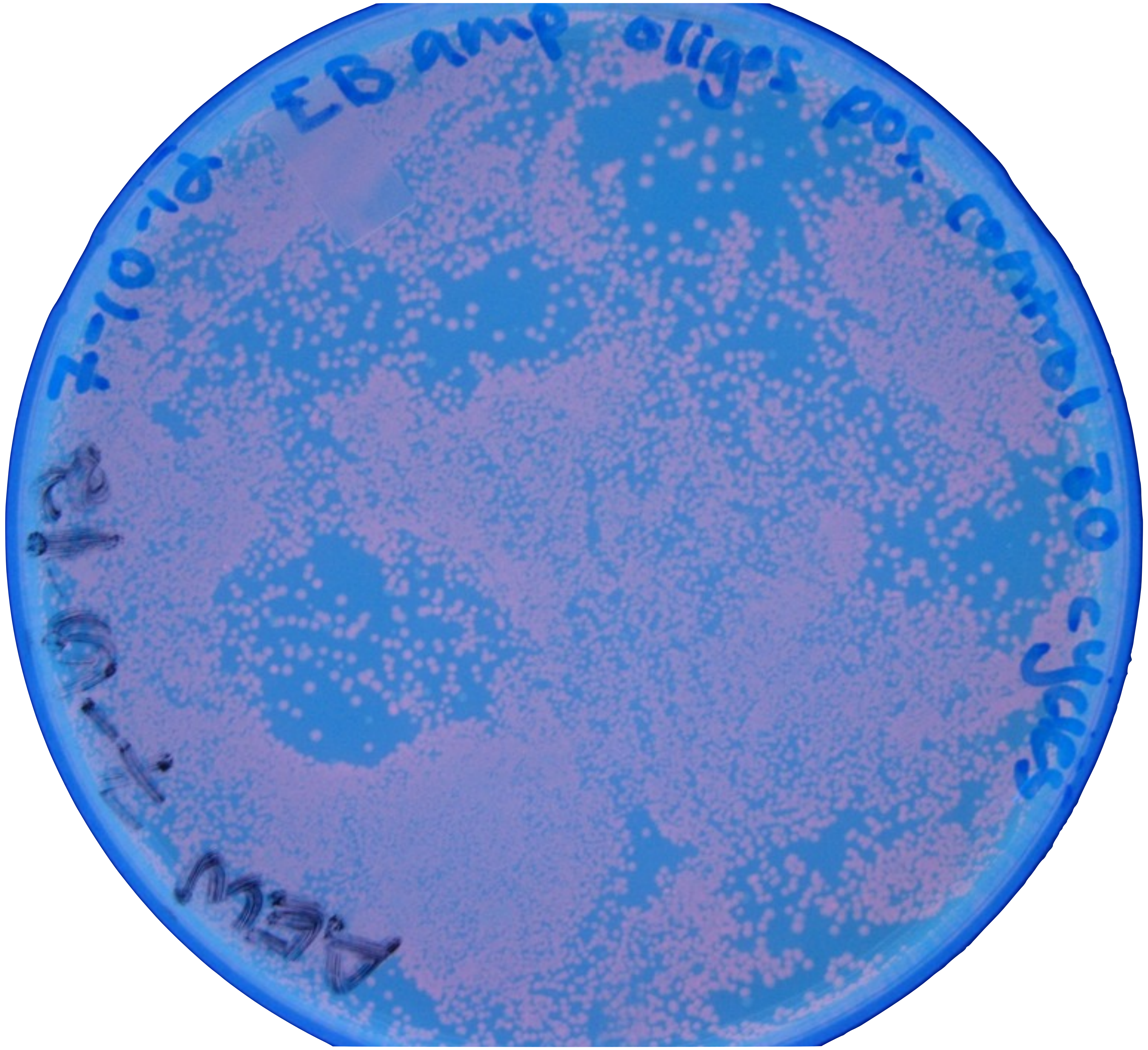
Search Again

gcat.davidson.edu/RFP/

Testing Known Promoters: Ptac

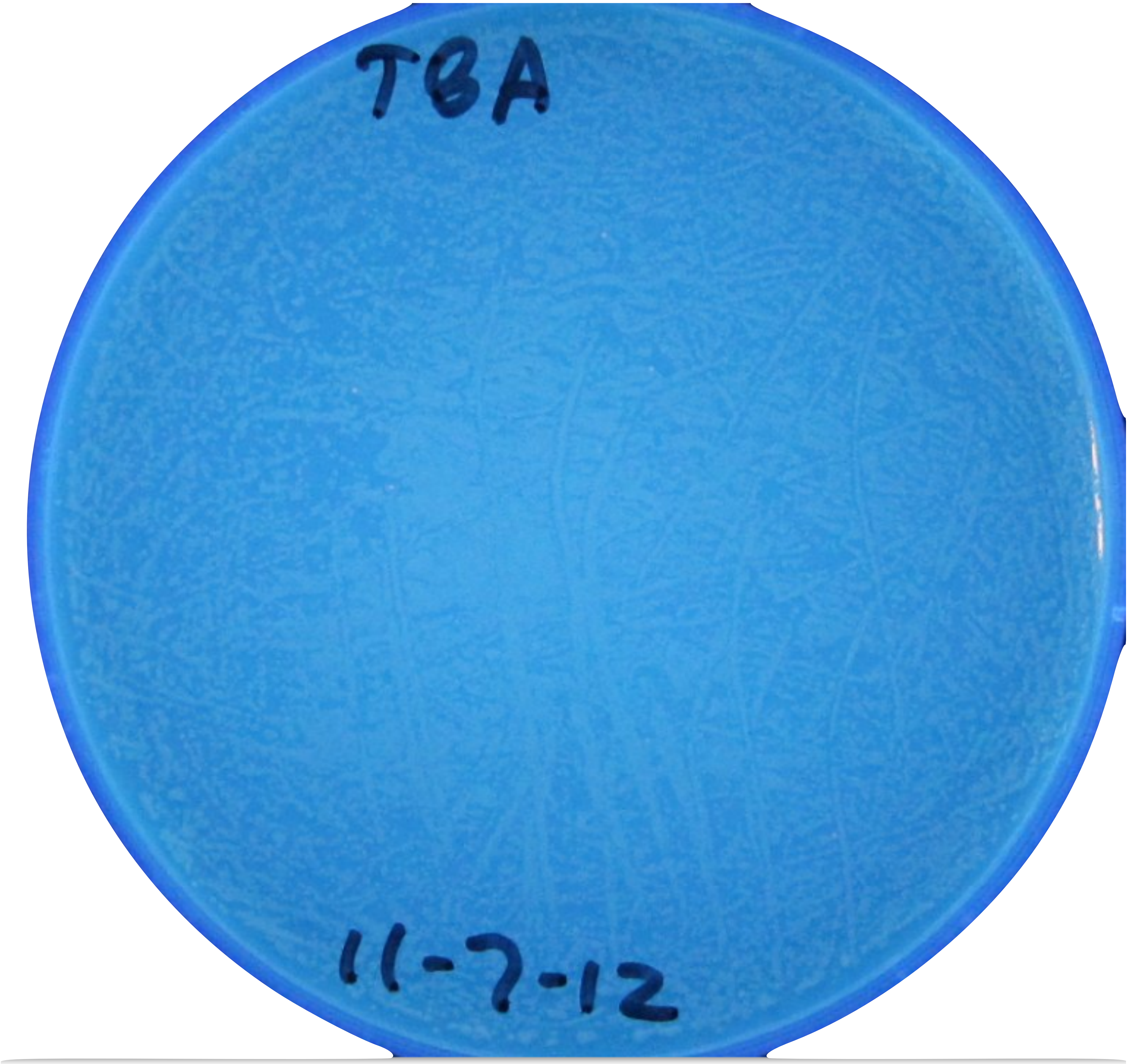
-35 -10

5′ CGACGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGA 3′
3′ CTCGACAACTGTTAATTAGTAGCCGAGCATATTACACACCTCGCC 5′



Student Sample, November 2012

```
                -35    ATAA (deleted)   -10
5’ CGACGAGCTGTTGACA----ATCATCGGCTCGTATAATGTGTGGA        3’
3‘          CTCGACAACTGT----TAGTAGCCGAGCATATTACACACCTCGCC 5’
```



Student Sample, November 2012

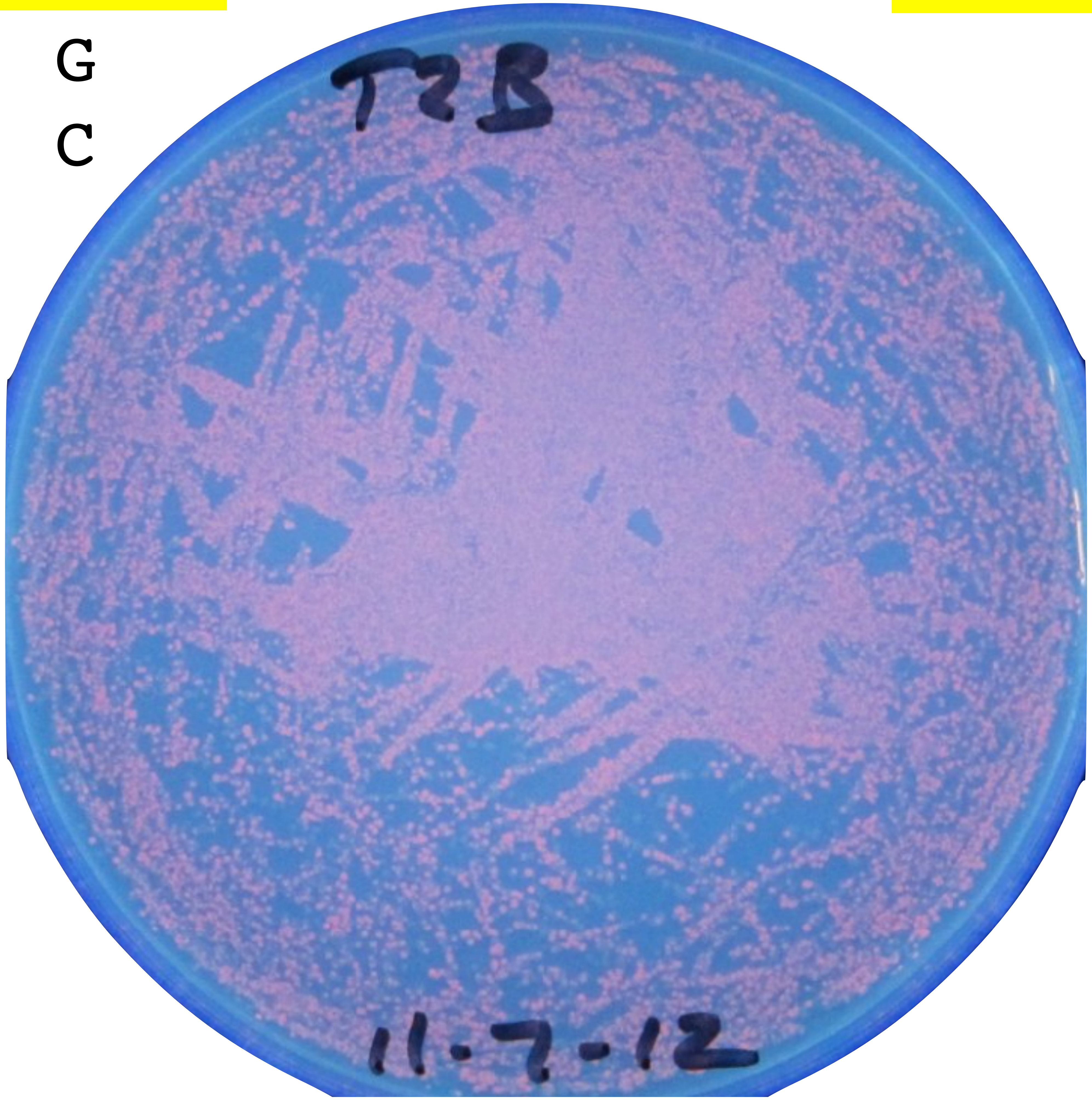
-35 -10

5′ CGACGAGCTGTTtACAATTAATCATCGGCTCGTATAATGTGTGGA 3′

3′ CTCGACAAaTGTTAATTAGTAGCCGAGCATATTACACACCTCGCC 5′

G

C



Undergraduate Summer Research:

the SAT problem

Undergraduates Design Bacterial Computers

A. Malcolm Campbell
Biology and **GCAT**

Laurie J. Heyer
Mathematics and **GCAT**

DAVIDSON

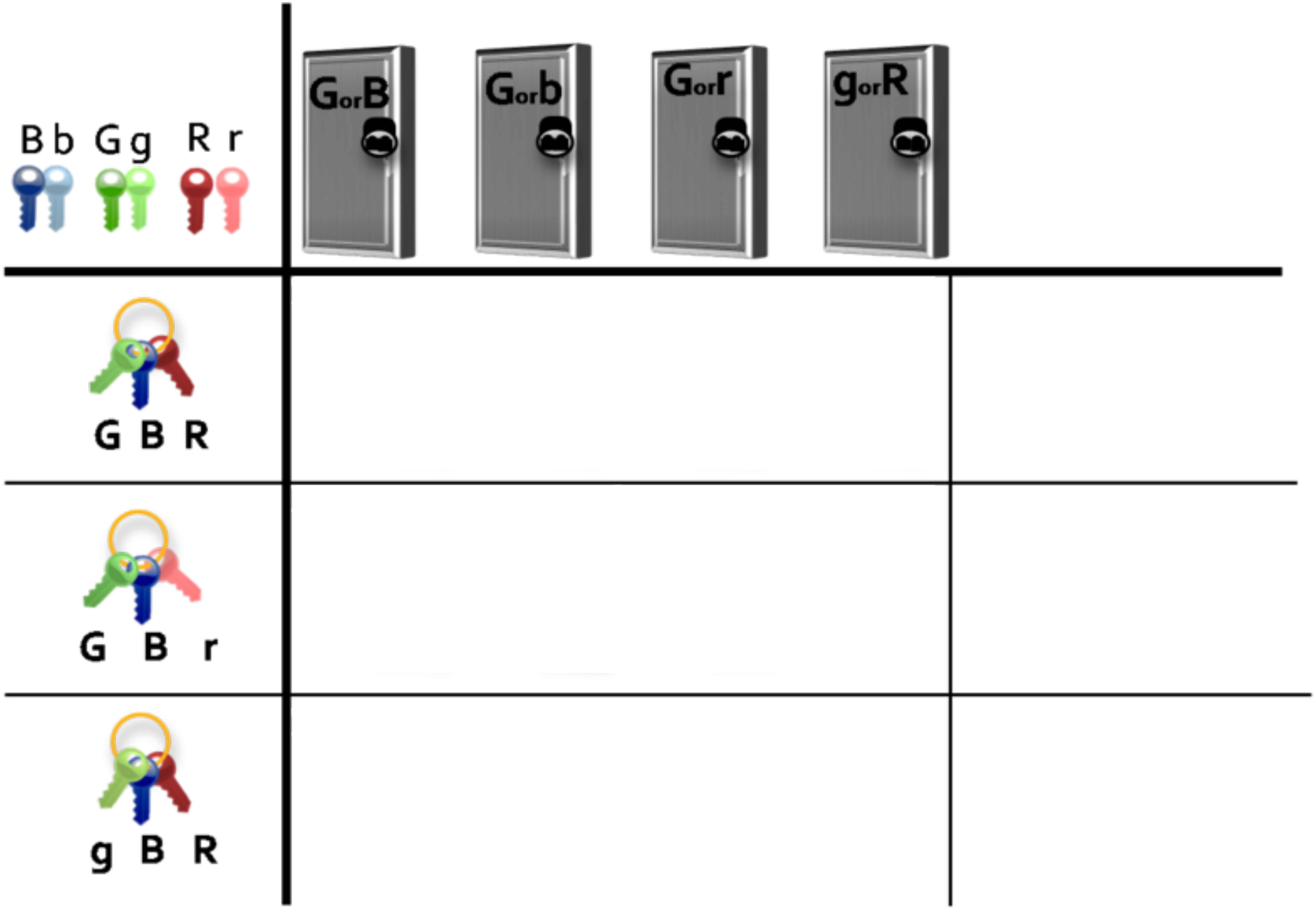
Todd T. Eckdahl
Biology and **GCAT**

Jeff L. Poet
Mathematics and **GCAT**

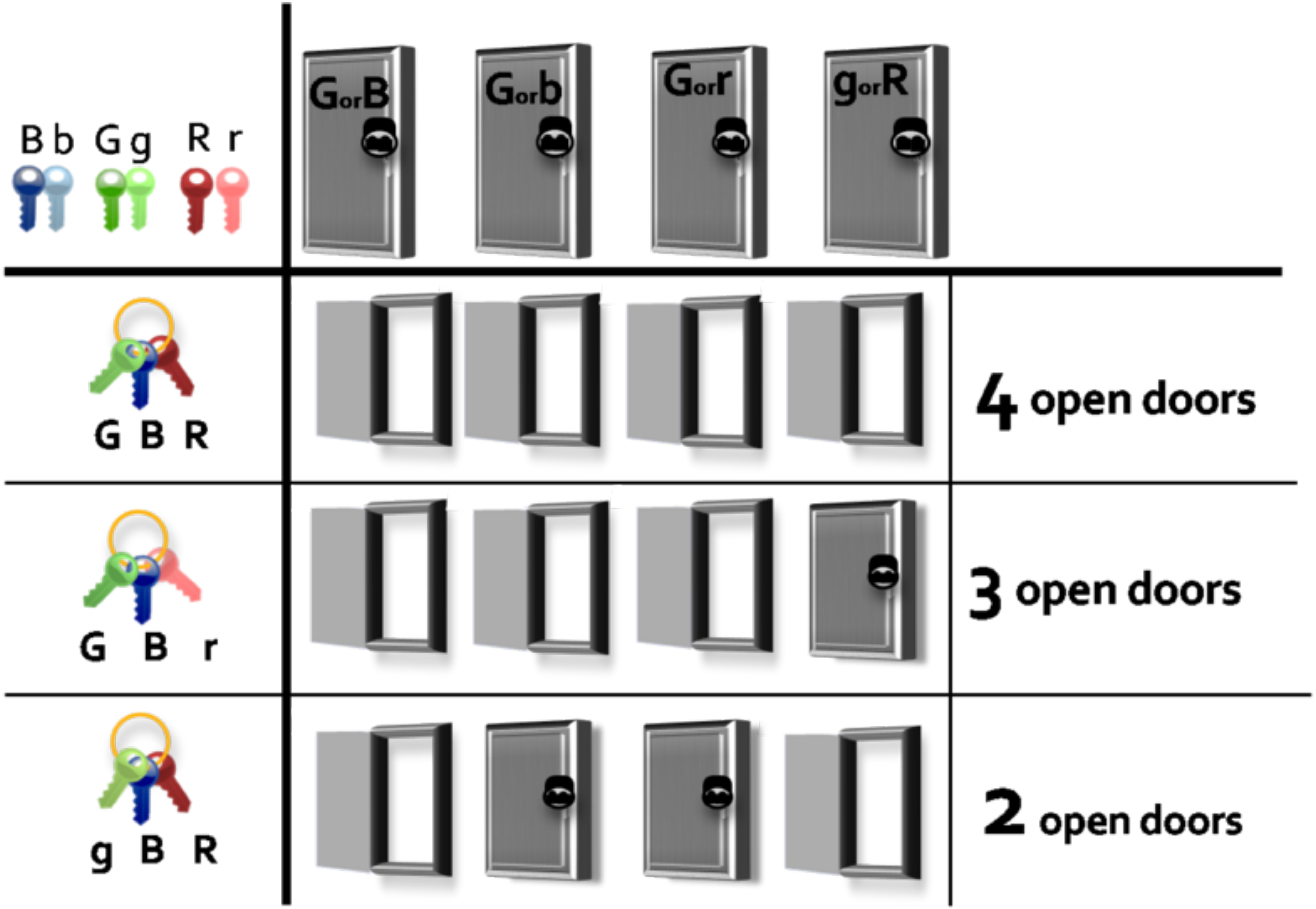
Discover gold WESTERN

Define the SATisfiability Problem

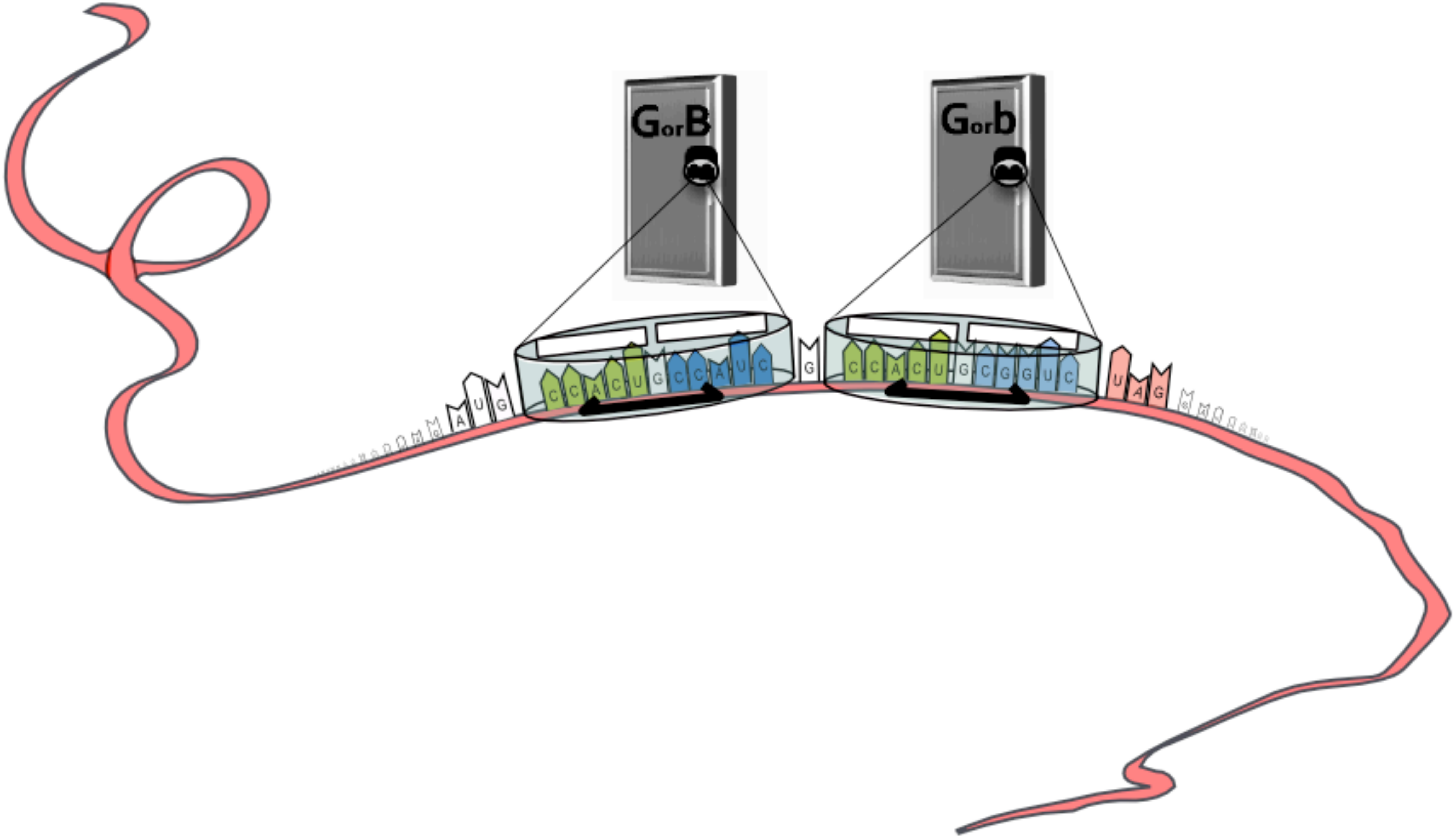
Define the SATisfiability Problem



Define the SATisfiability Problem



Converting Math to Biology



Central Dogma

DNA

atgccctactcactacctatagcgcat

↓ transcription

mRNA

aug ccc uac uca cua ccu aua ccg cau

↓ translation

Protein

M P Y S H P I P H

Lock: Frameshift Mutation

DNA
atgccctactcactacctatagcgcat

transcription

mRNA
aug ccc uac uca cua ccu aua ccg cau

translation

Protein

M P Y S H P I P H

DNA
atgccc**TC**tactcactacctatagcgcat

mRNA
aug ccc UCu acu cac uac cua uac cgc au

Protein

M P S T H Y H Y R

Unlock: Frameshift Suppression

DNA
atgccctactcactacctatagcgcat

↓

mRNA
aug ccc uac uca cua ccu aua ccg cau

↓

Protein

M P Y S H P I P H

DNA
atgccc**TC**tactcactacctatagcgcat

↓

5 base suppressor tRNA

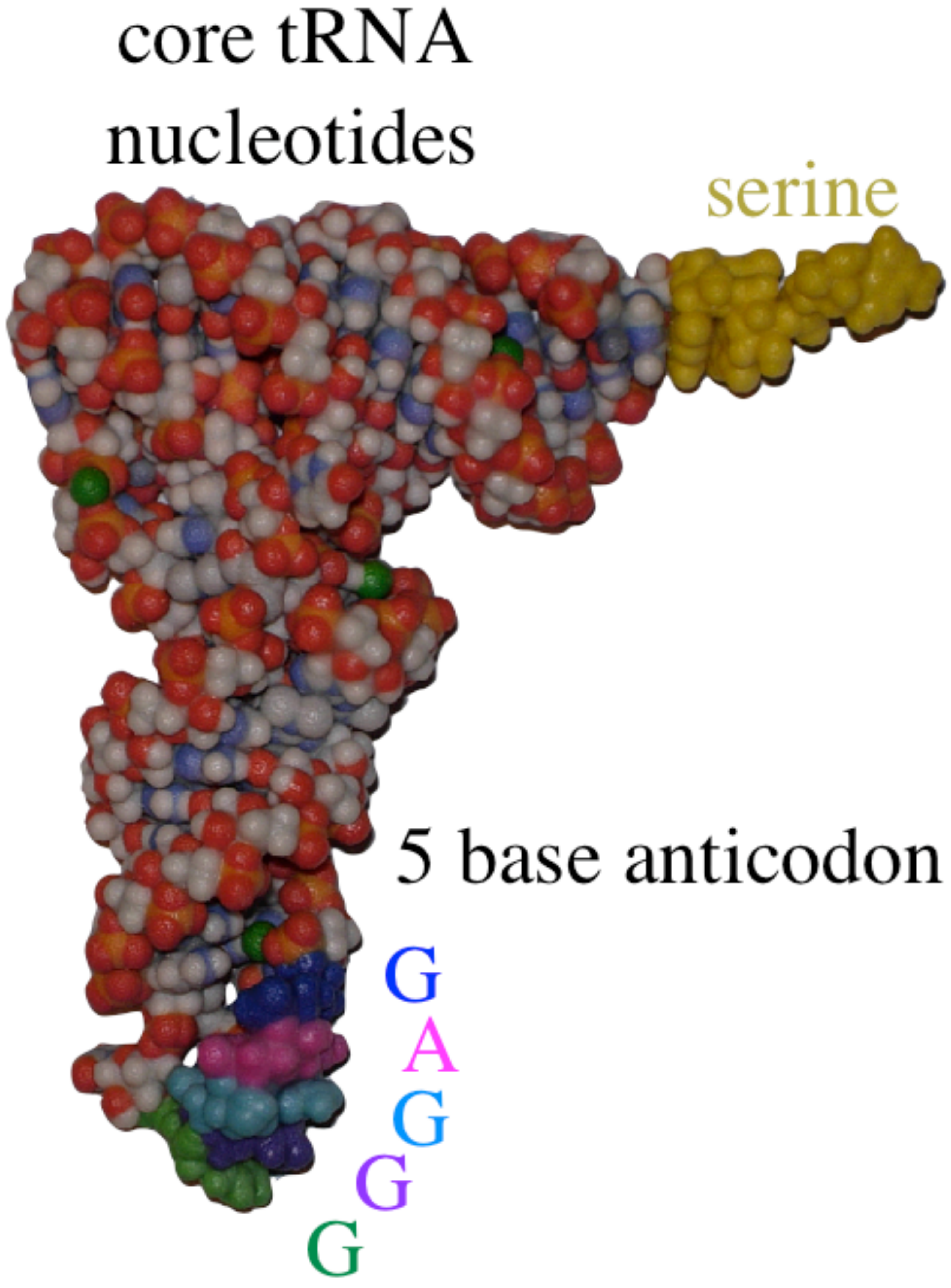
aug cccUC uac uca cua ccu aua ccg cau

↓

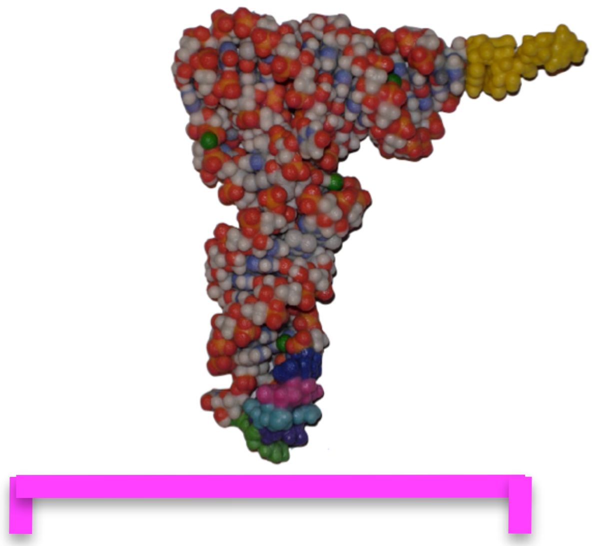
Protein

M S Y S H P I P H

Key: suppressor tRNA

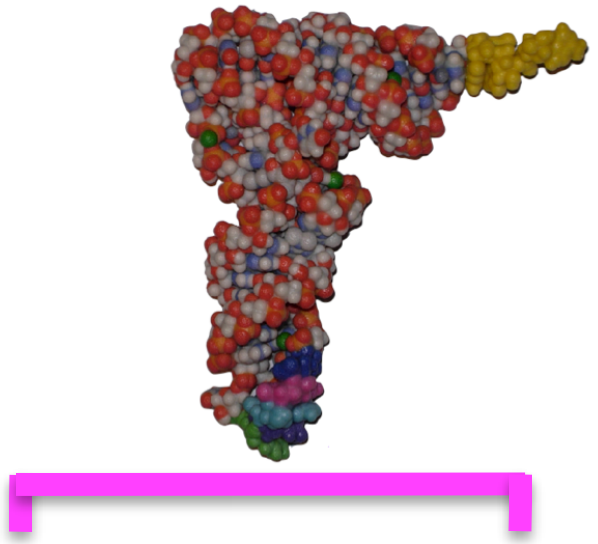


Coding 2-SAT Clause



G ATG NNNNNN gNN NNN → satisfied

Coding 2-SAT Clause

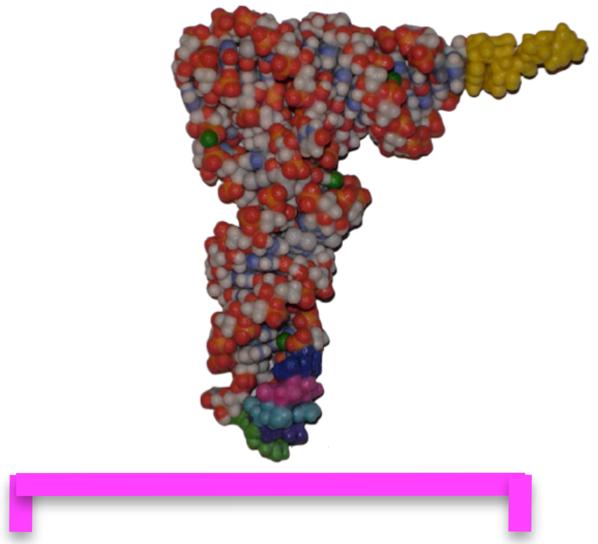


G ATG NNNNNN gNN NNN → satisfied

OR

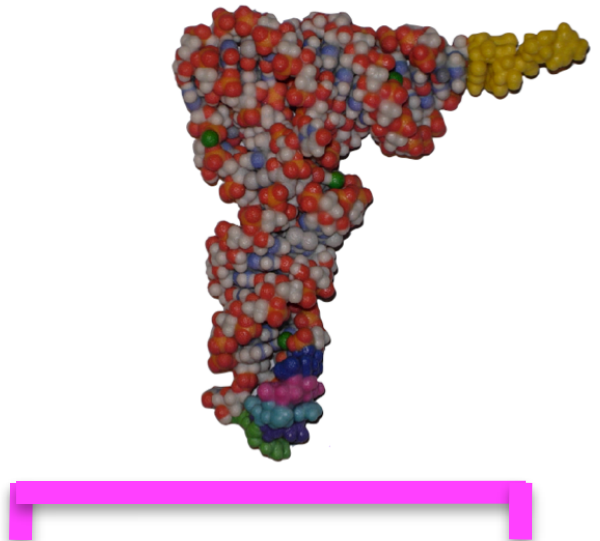
B ATG NNN NNg NNNNNN → satisfied

Coding 2-SAT Clause



G ATG NNNNNN gNN NNN → satisfied

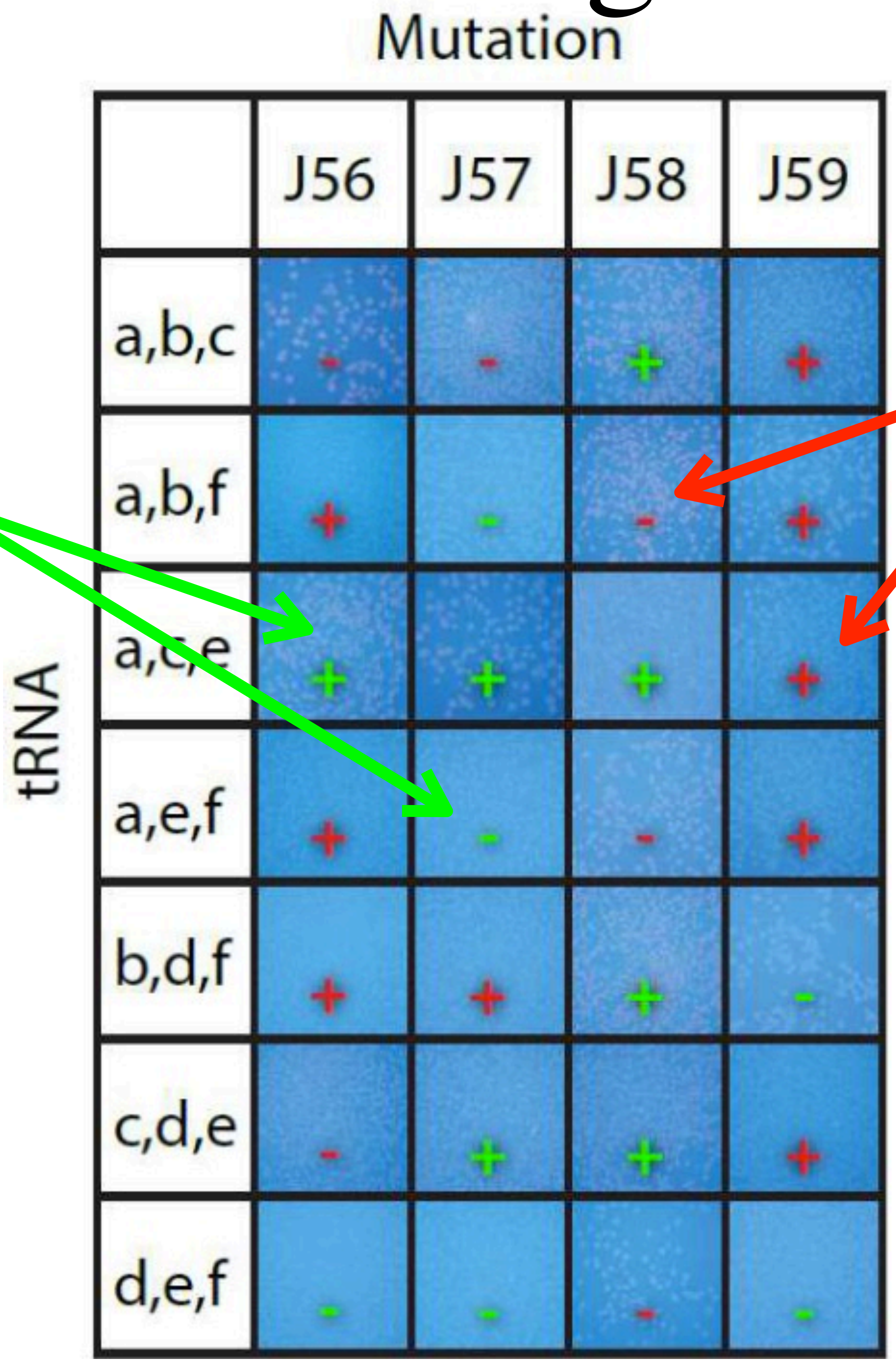
OR



B ATG NNN NNg NNNNNN → satisfied

ATG NNN NNg NNN 🛑 → no satisfaction

Confounding Results



Mutation

tRNA	J56	J57	J58	J59
a,b,c	-	-	+	+
a,b,f	+	-	-	+
a,c,e	+	+	+	+
a,e,f	+	-	-	+
b,d,f	+	+	+	-
c,d,e	-	+	+	+
d,e,f	-	-	-	-

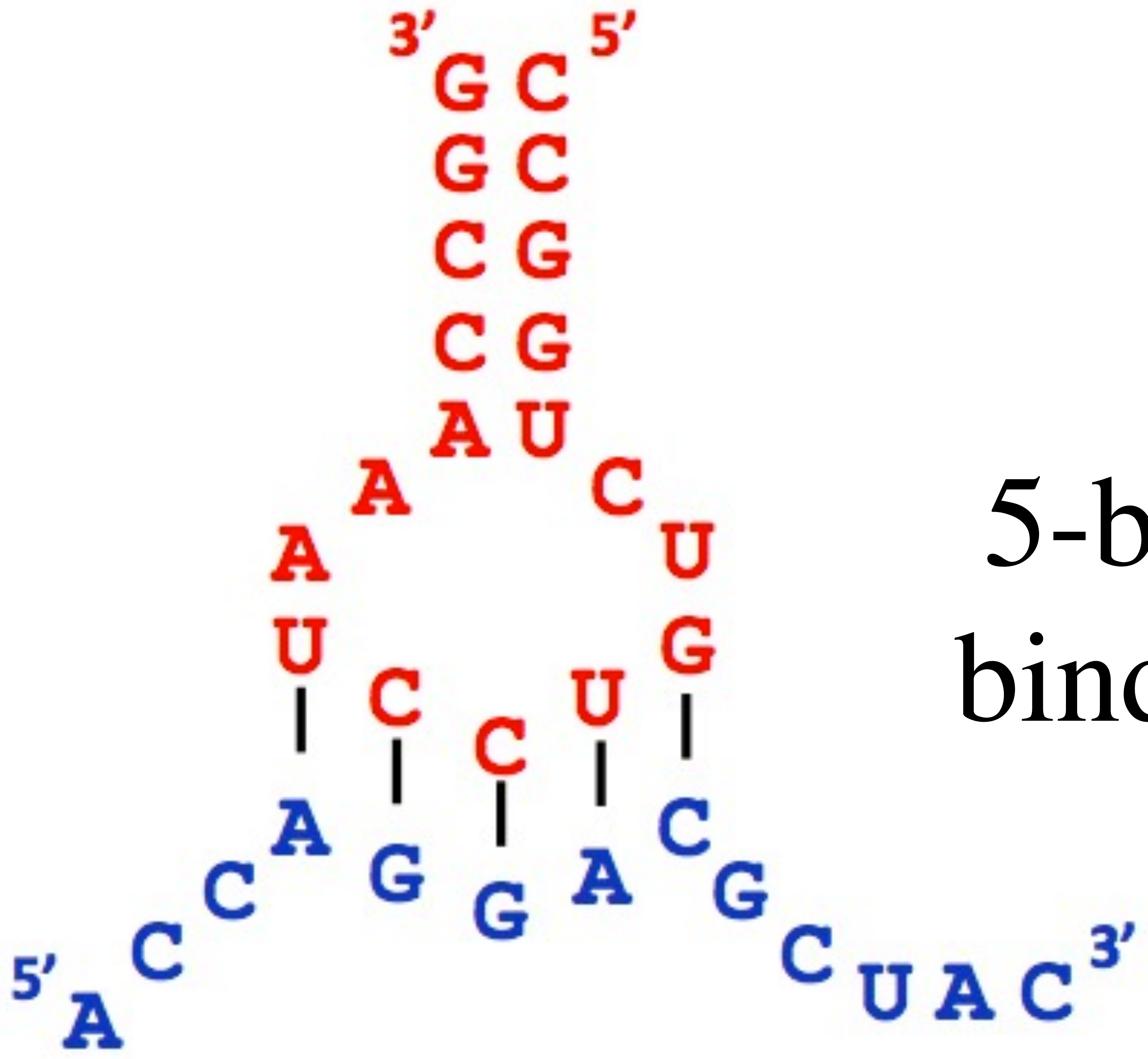
matched expectation

contradicted expectation

Published Mechanism

tRNA

mRNA

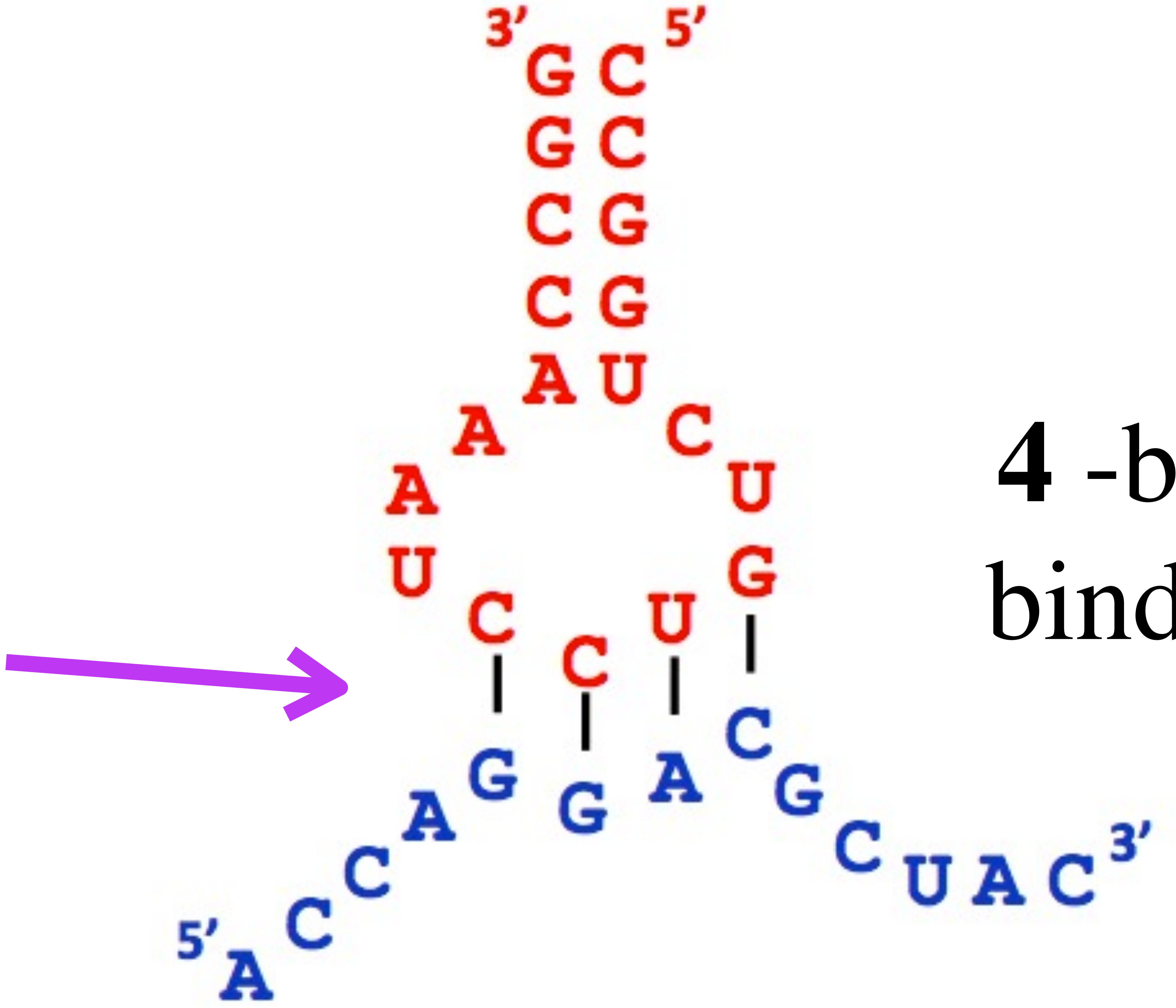


5-base binding

New Insight

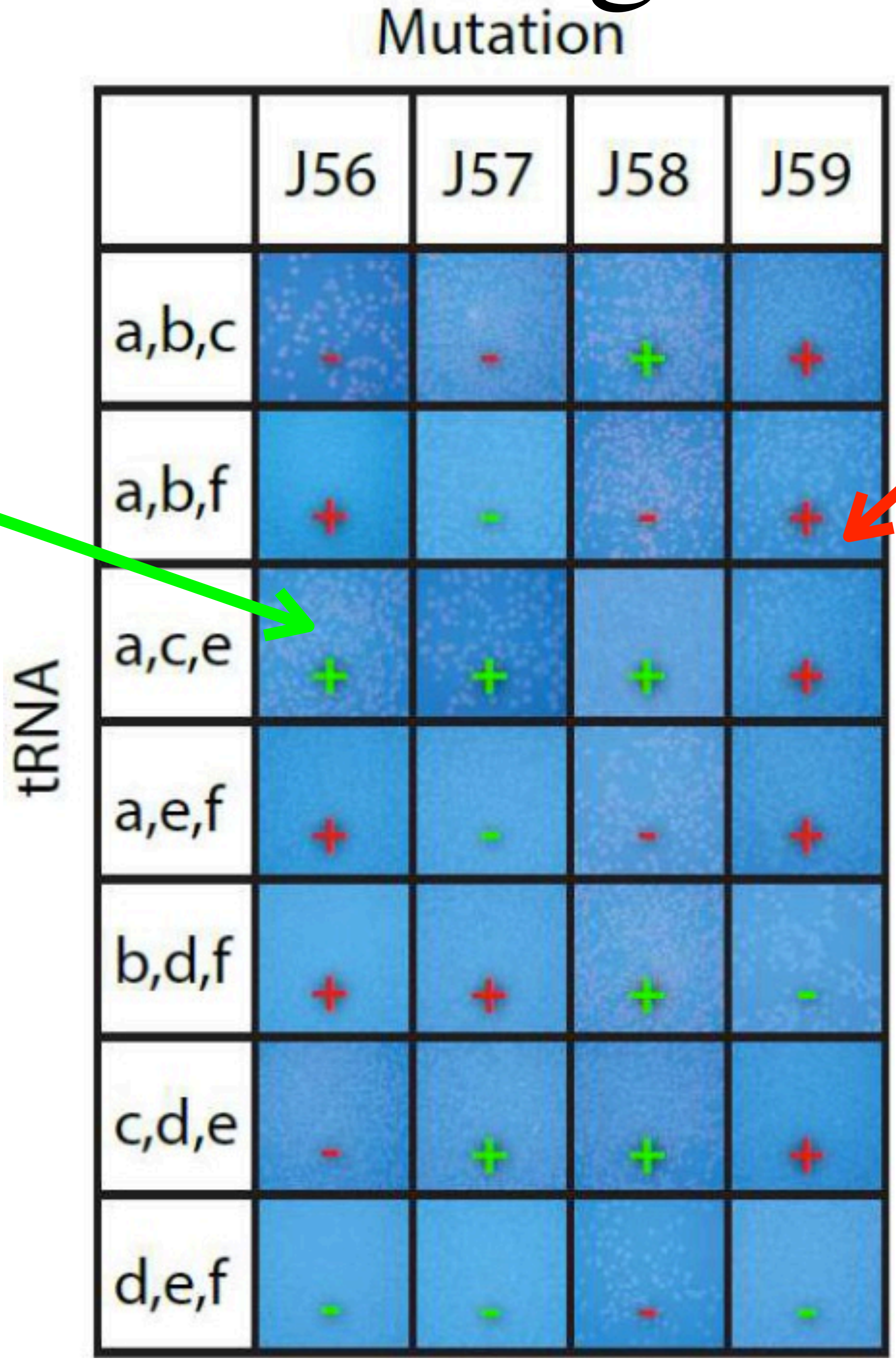
tRNA

mRNA



4 -base binding

Confounding Results

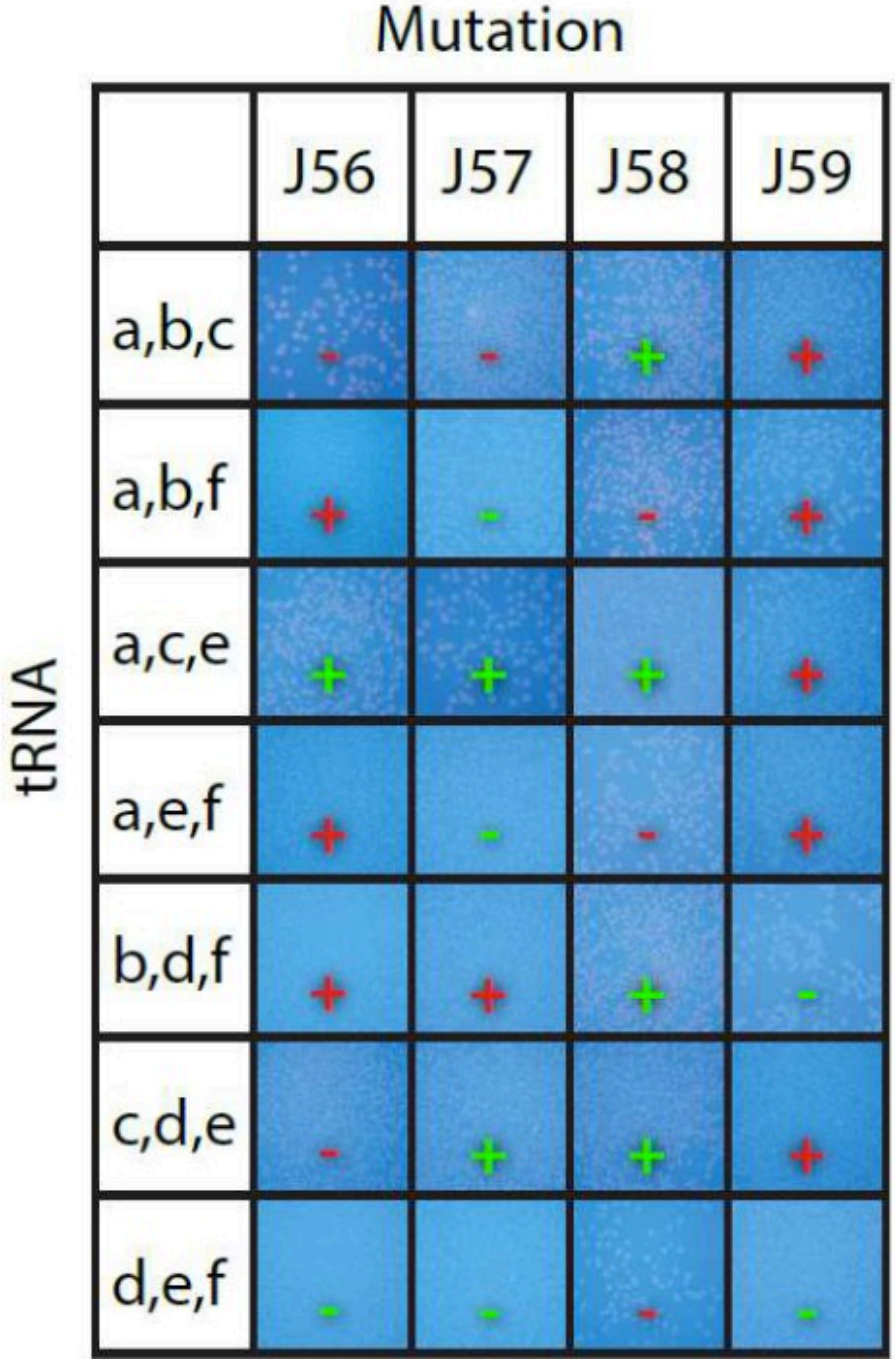


tRNA \ Mutation	J56	J57	J58	J59
a,b,c	-	-	+	+
a,b,f	+	-	-	+
a,c,e	+	+	+	+
a,e,f	+	-	-	+
b,d,f	+	+	+	-
c,d,e	-	+	+	+
d,e,f	-	-	-	-

matched expectation

contradicted expectation

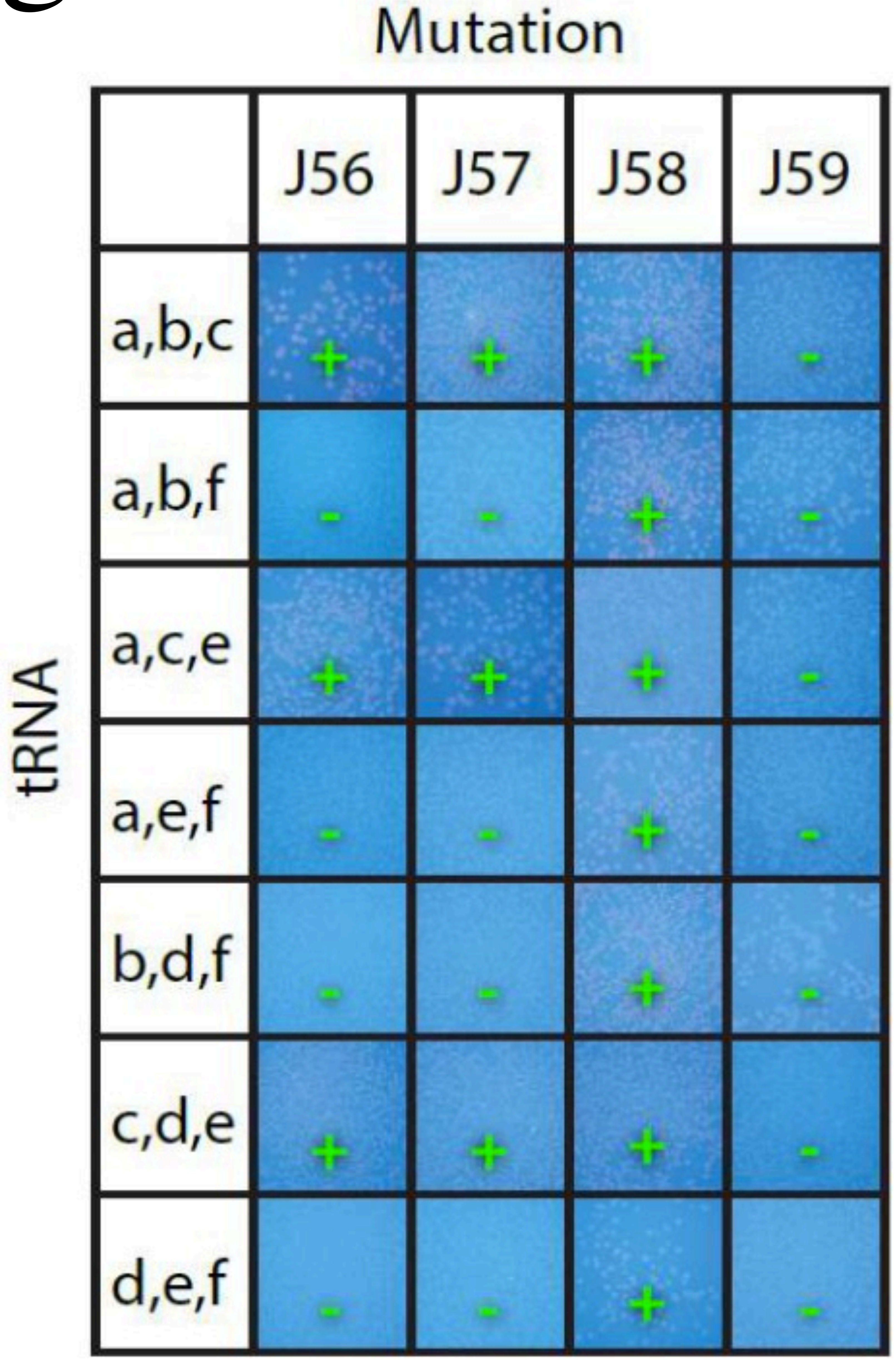
Confounding Results



Mutation

tRNA

	J56	J57	J58	J59
a,b,c	-	-	+	+
a,b,f	+	-	-	+
a,c,e	+	+	+	+
a,e,f	+	-	-	+
b,d,f	+	+	+	-
c,d,e	-	+	+	+
d,e,f	-	-	-	-



Mutation

tRNA

	J56	J57	J58	J59
a,b,c	+	+	+	-
a,b,f	-	-	+	-
a,c,e	+	+	+	-
a,e,f	-	-	+	-
b,d,f	-	-	+	-
c,d,e	+	+	+	-
d,e,f	-	-	+	-

Published Sept., 2012

Journal Article Synopsis

IBC 2012, vol. 4, article no. 10, pp. 1-12 | doi: 10.4051/ibc.2012.4.3.0010

view 1227 | download 172 | rating 7.2 | comment 0

Reports on negative result (Synthetic biology, Biological computation/Database, Biomathematics/Mathematical Biology and Medicine)

Open Access, Open Review

Bacterial Logic Devices Reveal Unexpected Behavior of Frameshift Suppressor tRNAs

Eric M. Sawyer[1,2], Cody Barta[2], Romina Clemente[1], Michel Conn[2], Clif Davis[2], Catherine Doyle[1], Mary Gearing[1], Olivia Ho-Shing[1], Alyndria Mooney[1,3], Jerrad Morton[2], Shamita Punjabi[1], Ashley Schnoor[4], Siya Sun[4], Shashank Suresh[5], Bryce Szczepanik[2], D. Leland Taylor[1], Annie Temmink[5], William Vernon[2], A. Malcolm Campbell[1], Laurie J. Heyer[5], Jeffrey L. Poet[4] and Todd Eckdahl[2,*]

[1]Department of Biology, Davidson College, Davidson, NC 28035
[2]Department of Biology, Missouri Western State University, St. Joseph, MO 64507
[3]Department of Biology, University of Arkansas at Pine Bluff, Pine Bluff, AR 71601
[4]Department of Computer Science, Math and Physics, Missouri Western State University, St. Joseph, MO 64507
[5]Department of Mathematics, Davidson College, Davidson, NC 28035

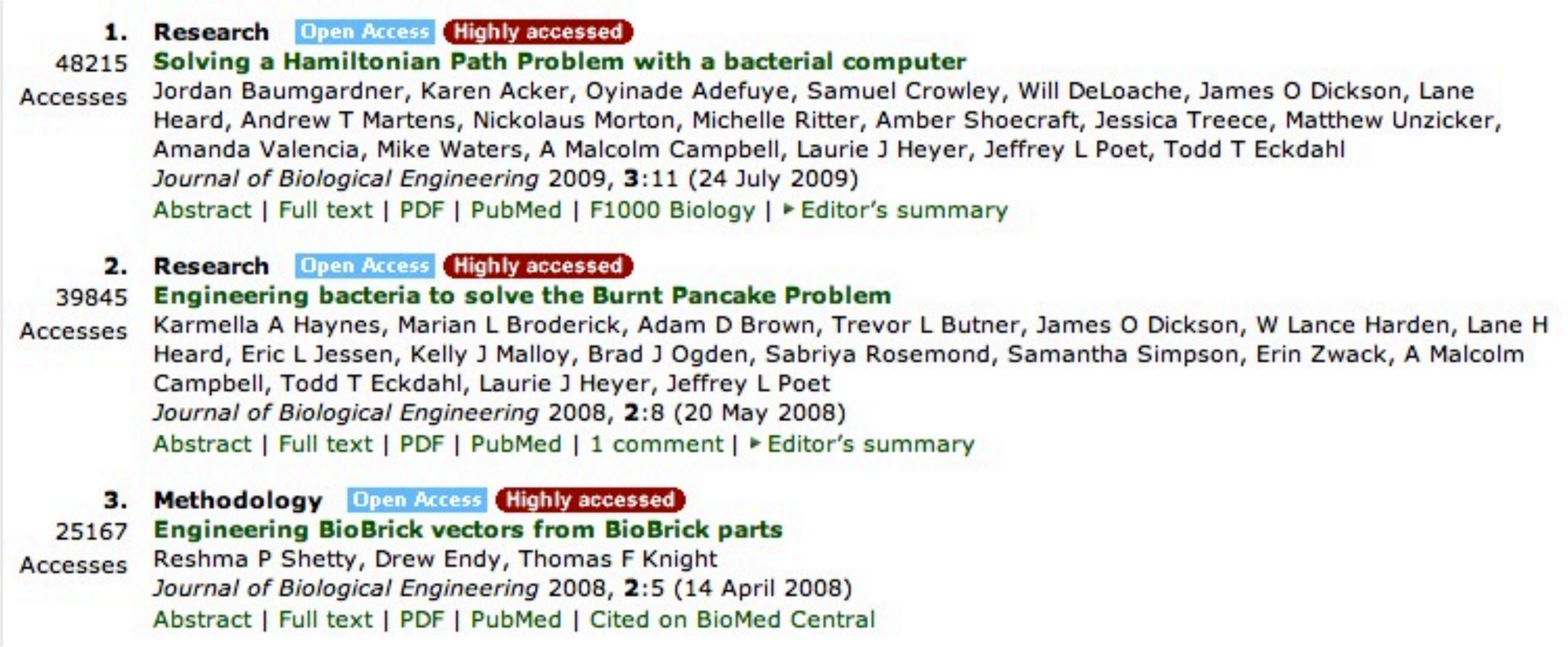
18 undergraduate coauthors

SynBio Papers Published

JOURNAL OF BIOLOGICAL ENGINEERING

27 undergraduate coauthors

Top 10 most accessed articles of all time

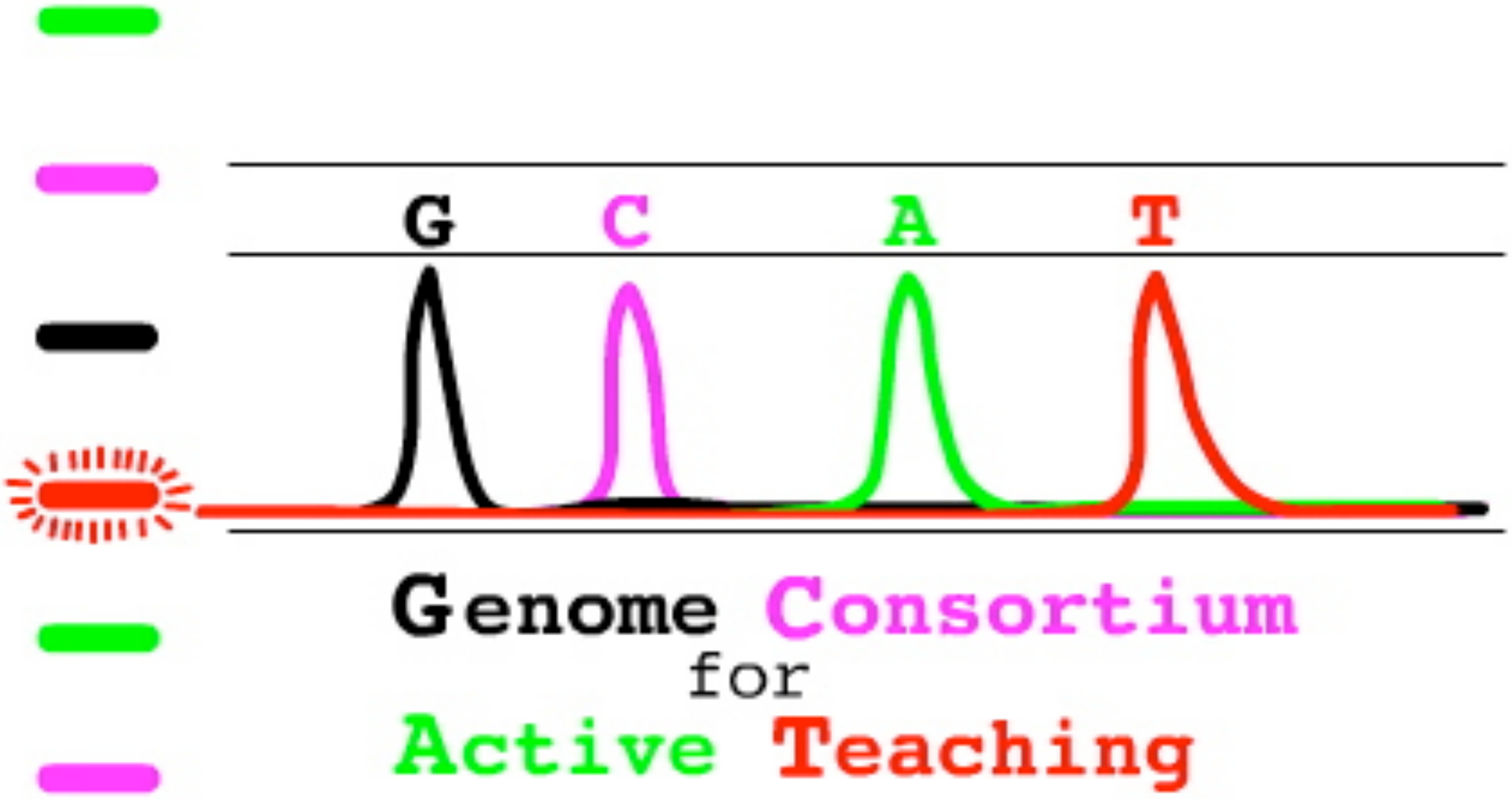


1. **Research** Open Access Highly accessed
48215 Accesses
Solving a Hamiltonian Path Problem with a bacterial computer
Jordan Baumgardner, Karen Acker, Oyinade Adefuye, Samuel Crowley, Will DeLoache, James O Dickson, Lane Heard, Andrew T Martens, Nickolaus Morton, Michelle Ritter, Amber Shoecraft, Jessica Treece, Matthew Unzicker, Amanda Valencia, Mike Waters, A Malcolm Campbell, Laurie J Heyer, Jeffrey L Poet, Todd T Eckdahl
Journal of Biological Engineering 2009, **3**:11 (24 July 2009)
Abstract | Full text | PDF | PubMed | F1000 Biology | ▸ Editor's summary

2. **Research** Open Access Highly accessed
39845 Accesses
Engineering bacteria to solve the Burnt Pancake Problem
Karmella A Haynes, Marian L Broderick, Adam D Brown, Trevor L Butner, James O Dickson, W Lance Harden, Lane H Heard, Eric L Jessen, Kelly J Malloy, Brad J Ogden, Sabriya Rosemond, Samantha Simpson, Erin Zwack, A Malcolm Campbell, Todd T Eckdahl, Laurie J Heyer, Jeffrey L Poet
Journal of Biological Engineering 2008, **2**:8 (20 May 2008)
Abstract | Full text | PDF | PubMed | 1 comment | ▸ Editor's summary

3. **Methodology** Open Access Highly accessed
25167 Accesses
Engineering BioBrick vectors from BioBrick parts
Reshma P Shetty, Drew Endy, Thomas F Knight
Journal of Biological Engineering 2008, **2**:5 (14 April 2008)
Abstract | Full text | PDF | PubMed | Cited on BioMed Central

12 Year Collaboration
Three Countries



www.bio.davidson.edu/GCAT

GCAT

Faculty Development

NSF funded

GCAT

Faculty Development

9 workshops, 360 faculty

GCAT

Faculty Development

360 faculty

X 100 students each

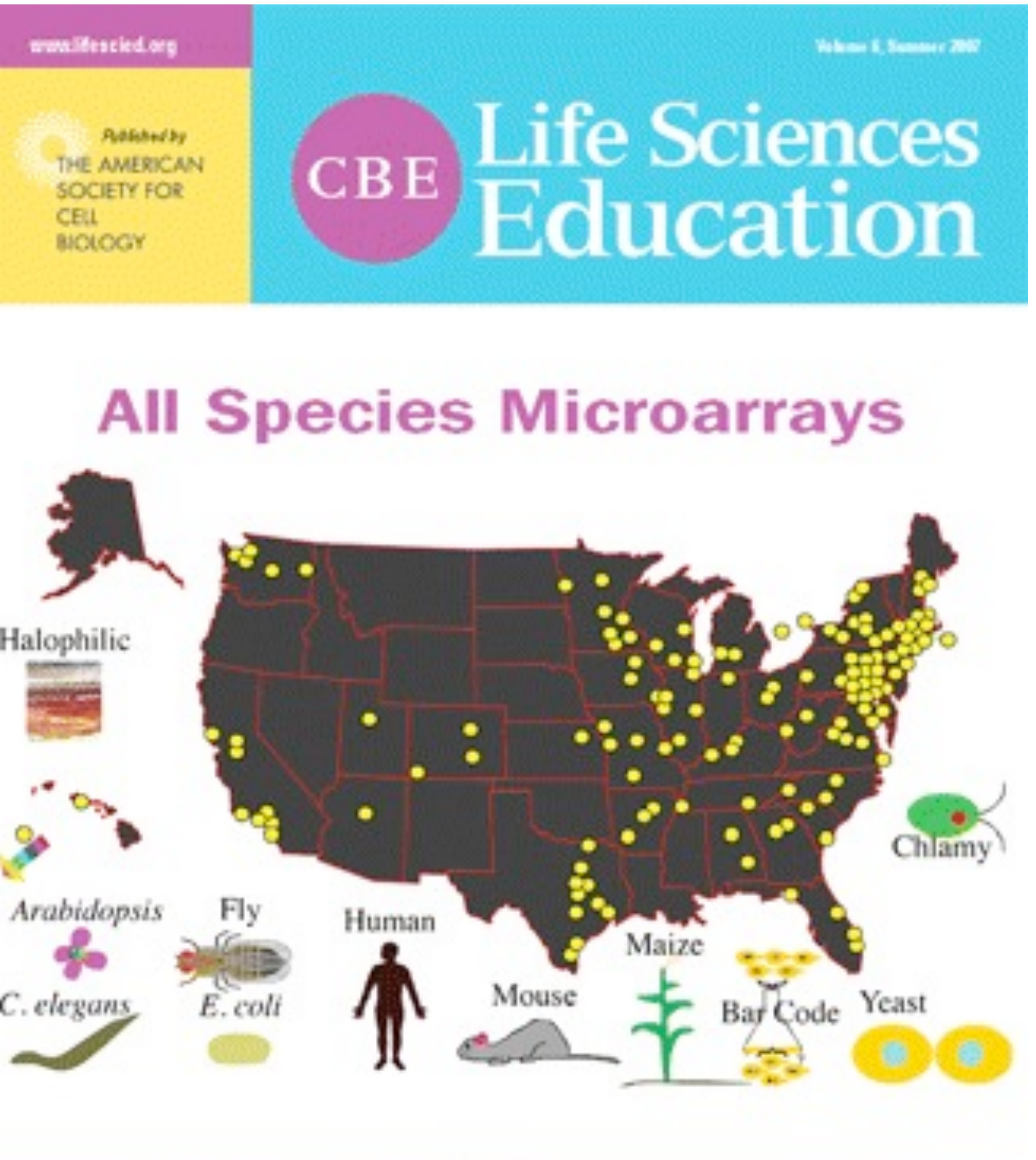
36,000 students/year

Faculty Appreciate **GCAT** Resources

	Mean	SD
Use microarray technology without **GCAT**	1.5	0.75
Online **GCAT** protocols useful	4.4	0.69
The **GCAT** -Listserv helpful	4.2	1.0
GCAT network significant factor	4.2	0.79
Positive experience using **GCAT**	4.6	0.60
I would use **GCAT** again in the future	4.7	0.63

1 = strongly disagree
5 = strongly agree

GCAT



www.lifescied.org

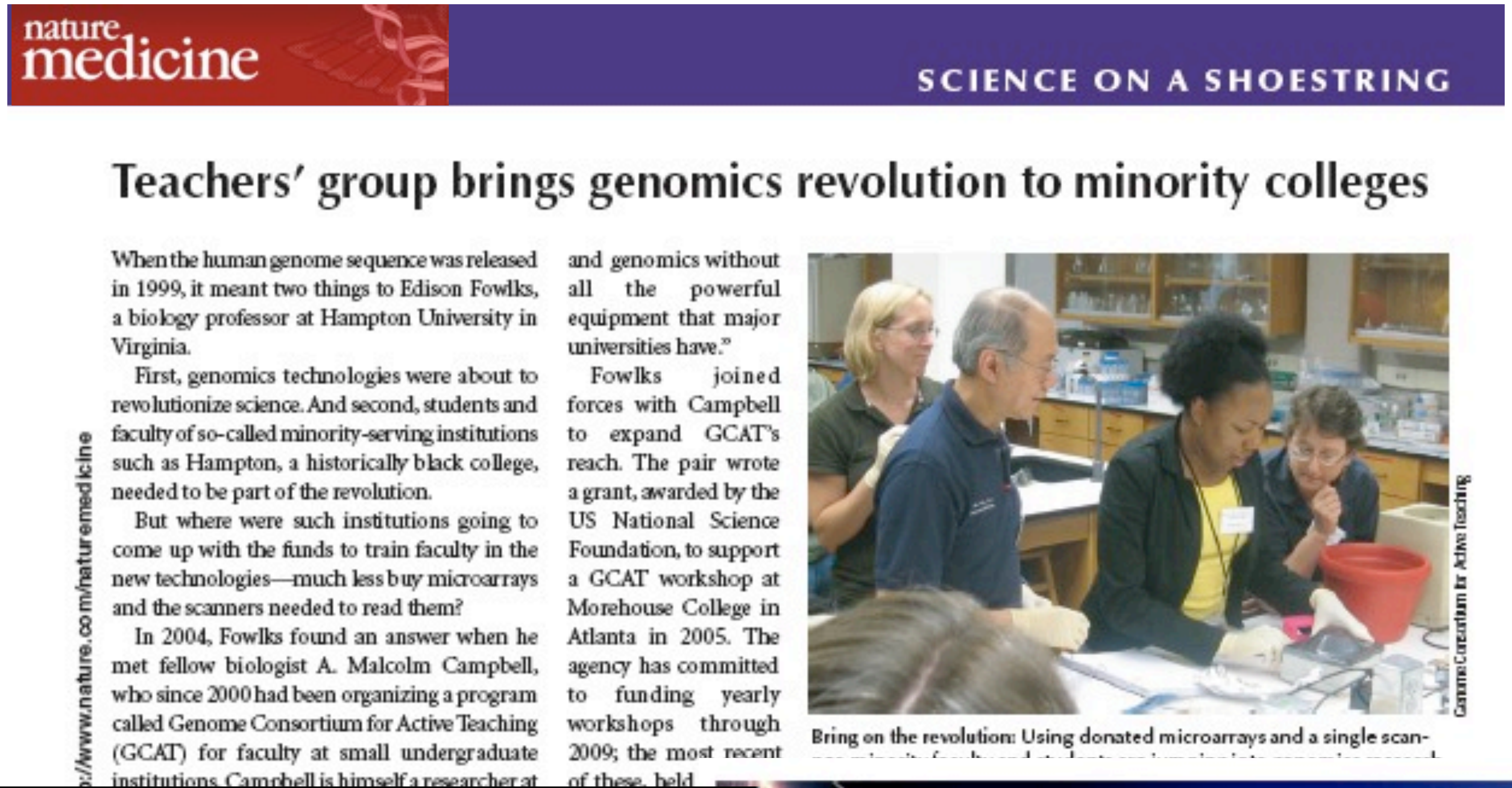
Published by THE AMERICAN SOCIETY FOR CELL BIOLOGY

Volume 6, Summer 2007

CBE Life Sciences Education

All Species Microarrays

GCAT: Genome Conso
for Active Tea



nature medicine

SCIENCE ON A SHOESTRING

Teachers' group brings genomics revolution to minority colleges

When the human genome sequence was released in 1999, it meant two things to Edison Fowlks, a biology professor at Hampton University in Virginia.

First, genomics technologies were about to revolutionize science. And second, students and faculty of so-called minority-serving institutions such as Hampton, a historically black college, needed to be part of the revolution.

But where were such institutions going to come up with the funds to train faculty in the new technologies—much less buy microarrays and the scanners needed to read them?

In 2004, Fowlks found an answer when he met fellow biologist A. Malcolm Campbell, who since 2000 had been organizing a program called Genome Consortium for Active Teaching (GCAT) for faculty at small undergraduate institutions. Campbell is himself a researcher at

and genomics without all the powerful equipment that major universities have."

Fowlks joined forces with Campbell to expand GCAT's reach. The pair wrote a grant, awarded by the US National Science Foundation, to support a GCAT workshop at Morehouse College in Atlanta in 2005. The agency has committed to funding yearly workshops through 2009; the most recent of these, held

Bring on the revolution: Using donated microarrays and a single scan-

Genome Consortium for Active Teaching

://www.nature.com/naturemedicine

EDUCATIONFORUM

COLLABORATIVE PROGRAMS

Genome Consortium for Active Teaching (GCAT)

A. Malcolm Campbell,[1,2*†] Todd T. Eckdahl,[2,4] Edison Fowlks,[2,5] Laurie J. Heyer,[2,3] Laura L. Mays Hoopes,[2,6] Mary Lee Ledbetter,[2,7] Anne G. Rosenwald[2,8]

A supportive network of scientists and faculty brings sophisticated microarray experiments to the undergraduate lab and classroom.

Biological research has been transformed in recent years by substantial advances in efficient data accumulation. The transcription output for every gene in a genome now can be measured in an afternoon; before it might have taken years. However, the recent advances in technology have yet to be incorporated into many biology classrooms (1). Most undergraduates are taught the same way their instructors were taught, which seldom reflects leading-edge research practices. Training faculty in the latest research methods is not well supported on most campuses (2). Worse yet, when students with outdated undergraduate science experiences become primary and secondary school teachers, they condemn future generations to inadequate preparation for college. Today's teachers may also neglect the more quantitative aspects and increased interdisciplinary involvement of modern biology (3–5). Educational options that reflect quantitative, interdisciplinary, and technological trends would provide students with experiences that mirror today's scholarship.

We have developed the Genome Consortium for Active Teaching (GCAT) (6) to engage undergraduates in genomics experimental design and data analysis. GCAT faculty use DNA microarrays to bring the excitement of interdisciplinary research to students. Students discover the importance of quantitative data analysis, and the faculty are reinvigorated by the opportunity to learn new technology.

GCAT in the lab. Undergraduates prepare samples and scan microarrays as part of their research at Davidson College.

Origins of GCAT

GCAT was formed in 1999 with the intent of bringing genomics into undergraduate curricula, primarily through student research (7, 8). Leading scientists donated materials and equipment. Undergraduates designed and performed experiments (see photograph above), mailed their microarrays for scanning, and then downloaded and analyzed their data (9).

Two limiting factors, long-term scanner access and a growing appetite for microarrays, were addressed by grant support and further donations from scientists (10–12). GCAT thus grew in size and expertise. GCAT supports free access to information and results through its Web site (6) and a listserv of more than 200 subscribers.

GCAT projects replaced student laboratory methods less prevalent in today's research, such as cloning and sequencing a gene and Northern blotting.

Rapid Growth

GCAT is committed to enabling any institution to adopt the use of microarrays in its undergraduate curriculum at affordable prices. To date, about 5000 undergraduates from 120 schools have used about 3400 microarrays. For the 2005–2006 academic year, GCAT provided more than 750 microarrays of nine plant, animal, and microbial species to students on 64 different campuses (6, 9). Tested protocols and teaching aids are available from GCAT. Continued grant support (11) covers the cost of microarrays.

Schools pay a nominal fee to GCAT for microarrays and scanning. Students produce and hybridize their own probes. Other than the scanners, only standard molecular biology equipment is required; the software is free. The summer workshop costs, which are currently covered by grant support, are about $2300 per participant.

The number of interested faculty continues to grow. Although this enthusiasm is more a measure of the importance of the microarray method in molecular biology today than of GCAT itself, it also serves as a testament to GCAT's user-friendly format.

GCAT faculty use the microarrays in various ways. Some analyze existing data sets, such as the yeast diauxic shift data (13) that shows how yeast switch from one metabolic route to another. Other faculty members offer courses in which students collect their own microarray data. Students have studied the effects of environmental conditions on growth, aging in yeast, chromatin structure, and the cellular side effects of chemotherapy (6). Microarrays offer a view of the connections between different pathways in a cell in ways that are hidden by many other methods. For example, one student project looked for expression changes in DNA replication mutants and found cell wall assembly changes, thus linking cytokinesis to mitosis.

Dissemination Through Faculty Development

GCAT has sponsored data generation (wet lab) and data analysis (dry lab) workshops in various settings (14). Wet and dry lab sessions work best when they run 2 and 3 days, respectively. Participants learn data analysis using MAGIC Tool freeware (15). MAGIC Tool works on any computer platform and is designed to enhance student understanding of

---

[1]Department of Biology, [2]Genome Consortium for Active Teaching, [3]Department of Mathematics, Davidson

ncemag.org SCIENCE VOL 311 24 FEBRUARY 2006 1103
Published by AAAS

JMBE

Journal of Microbiology & Biology Education

May 2007
Volume 8

In This Issue:

Microarrays and Data Analysis
David Kushner

Introductory Biology Discussion Group Evaluation
Marcy Peteroy-Kelly

Online Versus Onsite Bioinformatics Instruction
Kristina Obom and Patrick Cummings

Microbial Mats as Educational Tools
Carlos Rios-Velazquez, Lilliam Casillas-Martinez, and Pieter T. Visscher

Published by the:
AMERICAN SOCIETY FOR MICROBIOLOGY

www.MicrobeLibrary.org

NSF Current

Monthly Highlights of Research and Education Sponsored by the National Science Foundation

In This Issue: Latest Plant Genome Awards • Teachers Jockey Genes
RNA Plays Novel Role • New Nanotechnology Centers • NSF's 2005 Facility Plan

October 2005

Awards 19 New Plant Genome Research Projects

genomes of economically important plants are often large and lex, but through in-depth studies scientists uncover information that be translated into new and improved agricultural products and ices.

National Science Foundation (NSF) made 19 new awards totaling million in the eighth year of its Plant Genome Research Program P). The two- to five- year awards, ranging from $622,000 to $7.7 n, fund research and tools to reveal information in the genomes of omically important crop plants, such as wheat and soybean, as well as crease understanding of the genetics underlying plant processes ding disease resistance, flavor development, seed growth and wood tion.

NSF's press release and the list of 2005 PGRP awards for more nation.

NSF made 19 new awards in the eighth year of its Plant Genome Research Program. The awards will support genomics research in major crop plants such as soybean and also in trees including the loblolly pine and poplar. Credit: N. Rager Fuller, NSF.

uctors from Minority-Serving Institutions Learn to Teach Microarray Technology

Microarray technology, one of the hottest techniques in biological research, simultaneously measures the expression levels of tens of thousands of genes. Performing DNA microarray experiments and analyzing the mounds of resulting data are generally thought to be beyond the reach of all but a small number of undergraduates working in top research labs. However, the Genome Consortium for Active Teaching (GCAT), composed of faculty from over 120 primarily undergraduate institutions, has allowed over 4,000 undergraduates to conduct research using DNA microarrays.

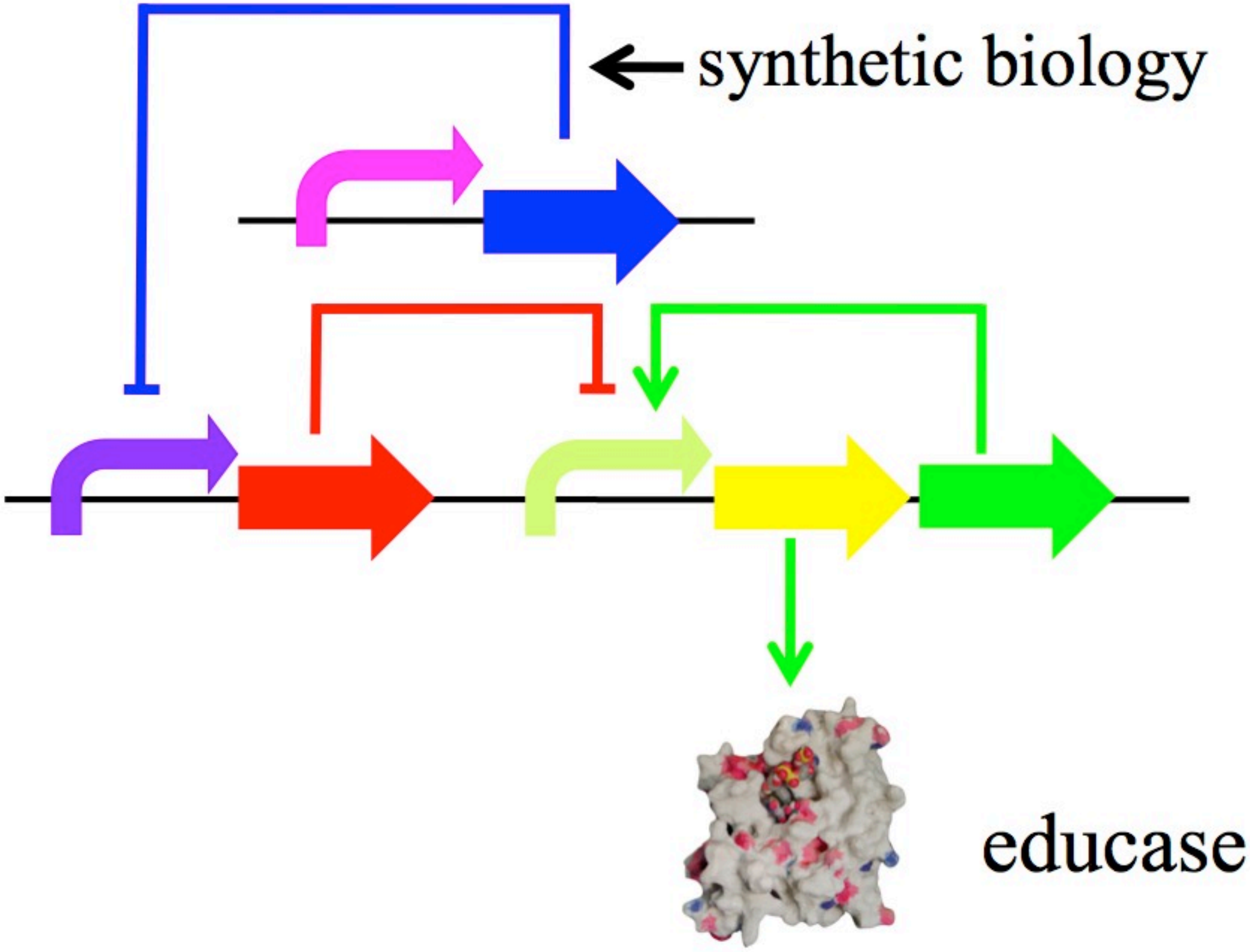
Research grants funded

Published basic research

Published pedagogy

Teaching awards

GCAT SynBio

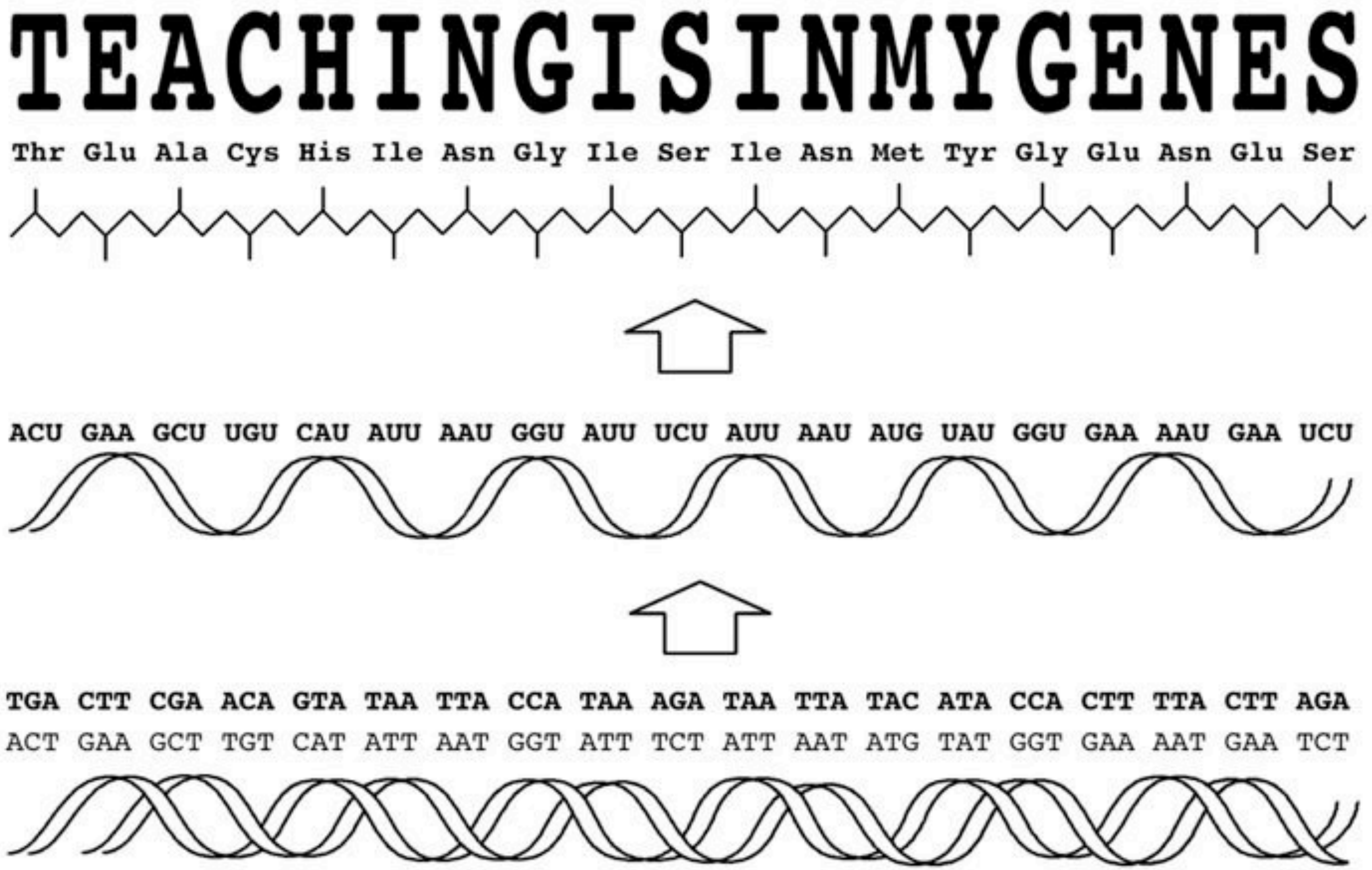


GCAT SynBio Faculty Workshops

15 pairs of faculty

1 Bio + 1 Other

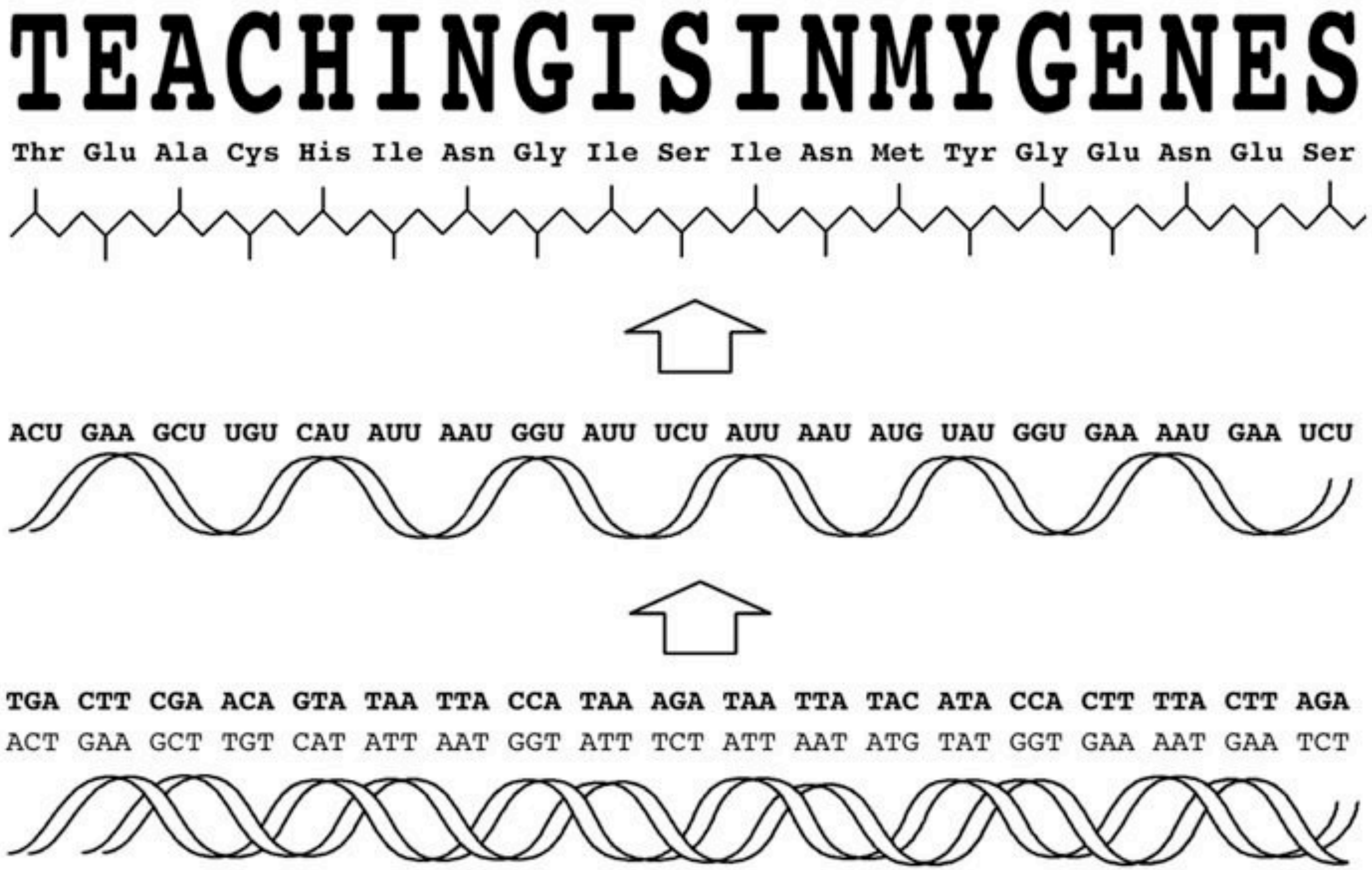
Summers 2010-2014



GCAT SynBio Faculty Workshops

30 faculty X 5 years X 100 students each

45,000 undergraduates over 5 years



NATIONAL SCIENCE FOUNDATION

HHMI
HOWARD HUGHES
MEDICAL INSTITUTE

The scenery only changes for the lead dog.

The scenery only changes for the lead dog.

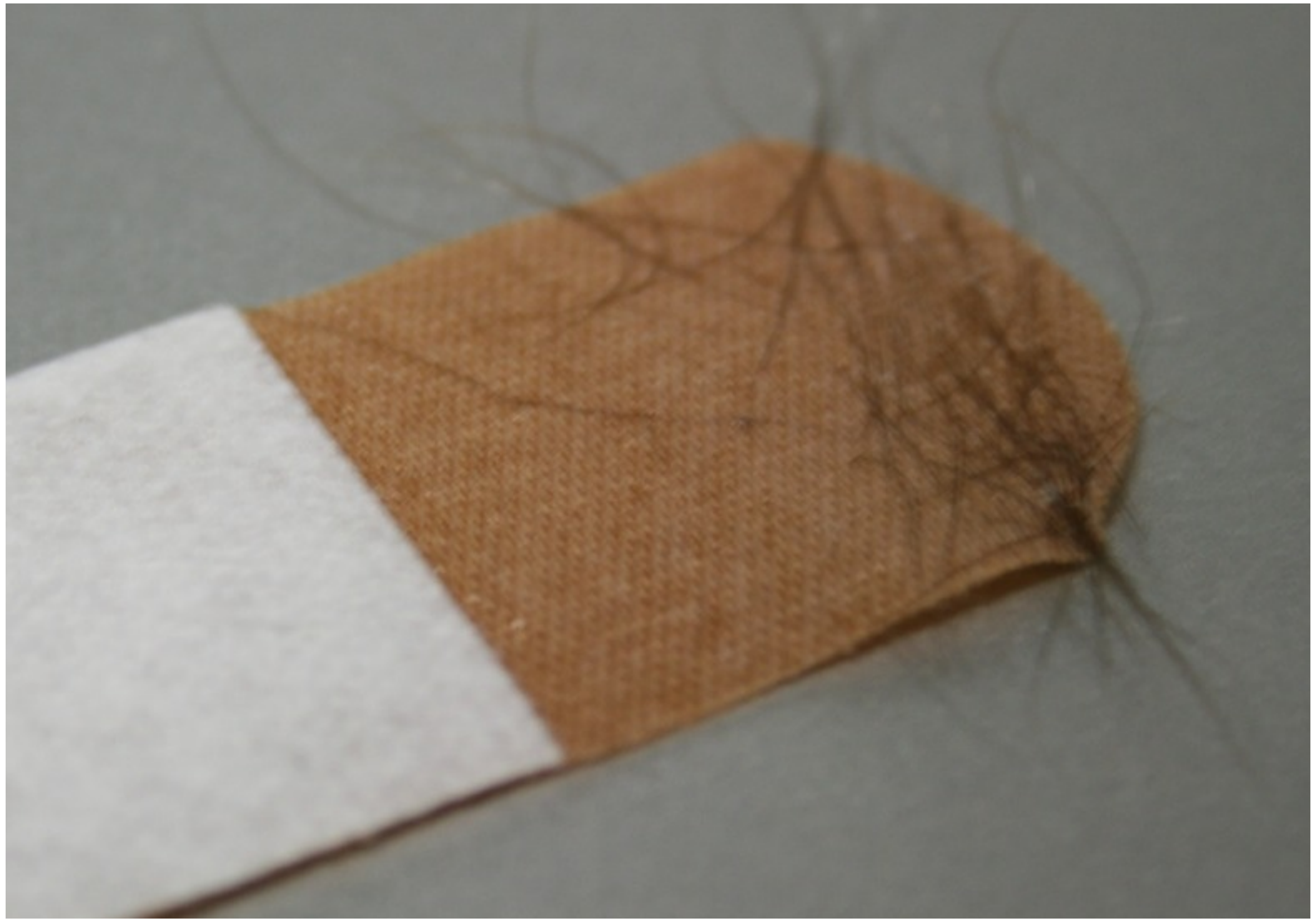
Collaborators

Faculty: Laurie Heyer, Jeff Poet, Todd Eckdahl, Karmella Haynes, Pat Sellers, Mark Barsoum

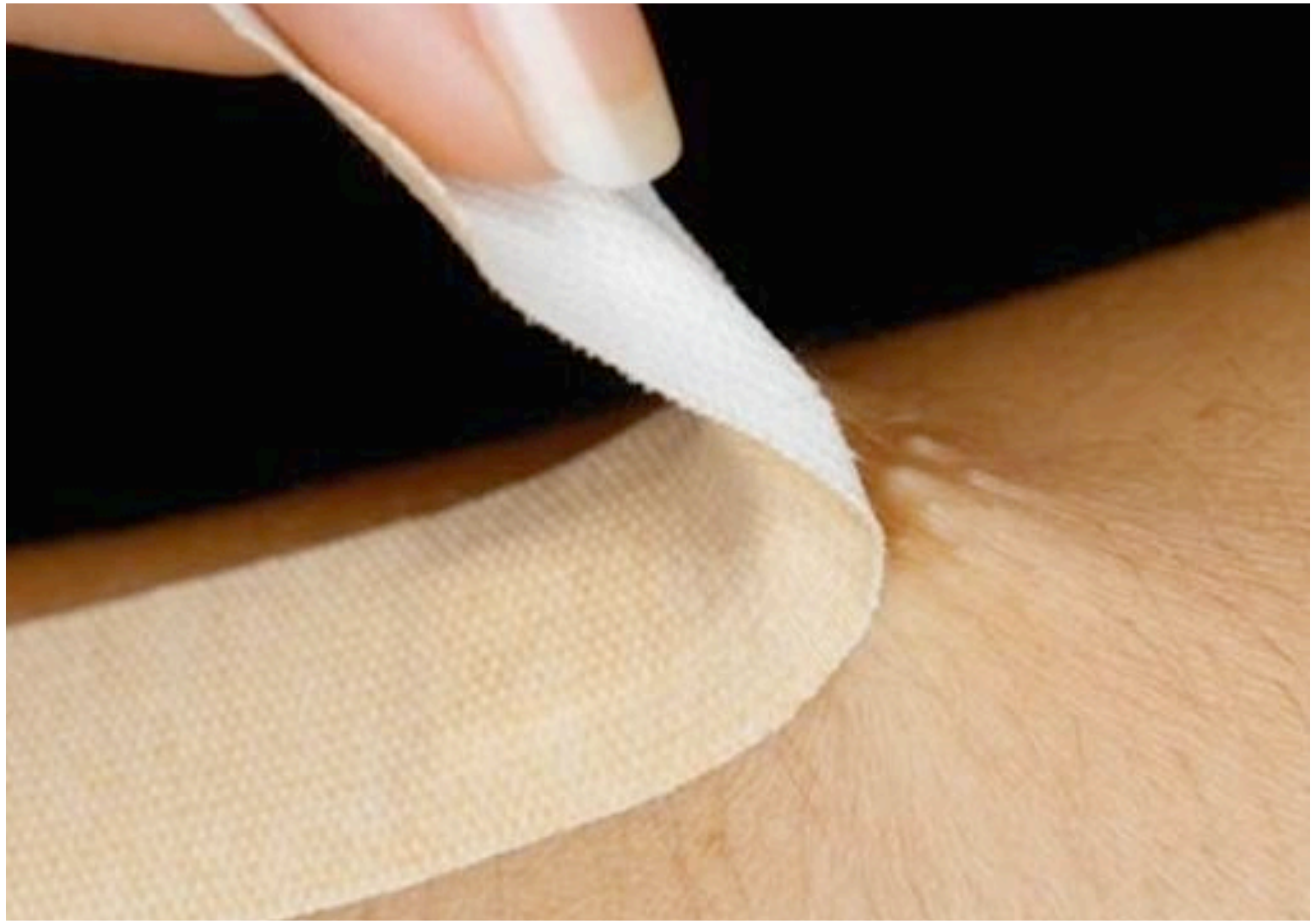
Students: Annie Wacker, Andrew Lantz, Tucker Whitesides, Ben Clarkson, Becca Evans, Betsy Gammon, Meredith Nakano, Caroline Vrana, Jonah Galeota-Sprung, Julia Fearington, Lilly Wilson, Pooja Potharaju, James Harden, Catherine Doyle, Duke DeLoache, Anvi Raina, Jamela Peterson, Stephen Streb, Linda Kleist, Katie Richeson, Steph Meador, Tom Shuman, Tori Rinker, Eugene Shiu, Nitya Rao, Keila Alfred, Romina Clemente, A.J. Grant, Mary Gearing, Kin Lau, Olivia Ho-Shing, Shamita Punjabi, Eric Sawyer, Shashank Suresh, Leland Taylor, Annie Temmink, Alyndria Thompson, Oyinade Adefuye, Will DeLoache, Jim Dickson, Andrew Martens, Amber Shoecraft, Mike Waters, Karen Acker, Bruce Henschen, Lance Harden, Sabriya Rosemond, Samantha Simpson, Erin Zwack, Kelly Davis, James Barron, Will DeLoache, Erin Feeney, Kristi Muscalino, Madeline Parra, Pallavi Penumetcha, Karlesha Roland, Max Win, Kristen DeCelle, Matt Gemberling, Oscar Hernandez, Andrew Drysdale, Mac Cowell, Nick Cain, Tamar Odel, and Jackie Ryan.

The Duke Endowment, NSF, HHMI
Genome Consortium for Active Teaching (**GCAT**)
Davidson College James G. Martin Genomics Program
MWSU SGA, Foundation & Summer Research Institute

How should you teach bioinformatics?

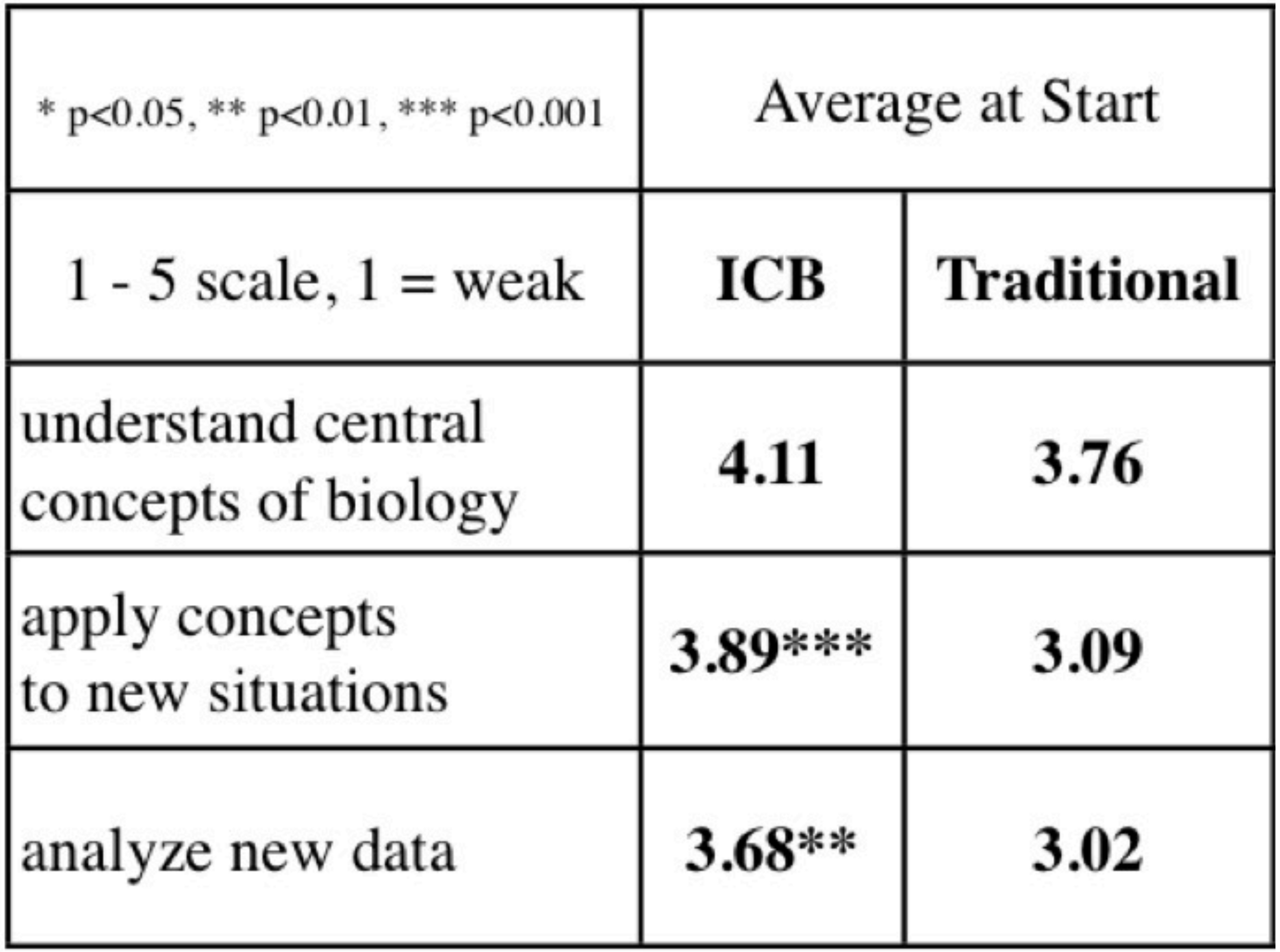


fast yank



slow pull

Are *ICB* students overconfident?

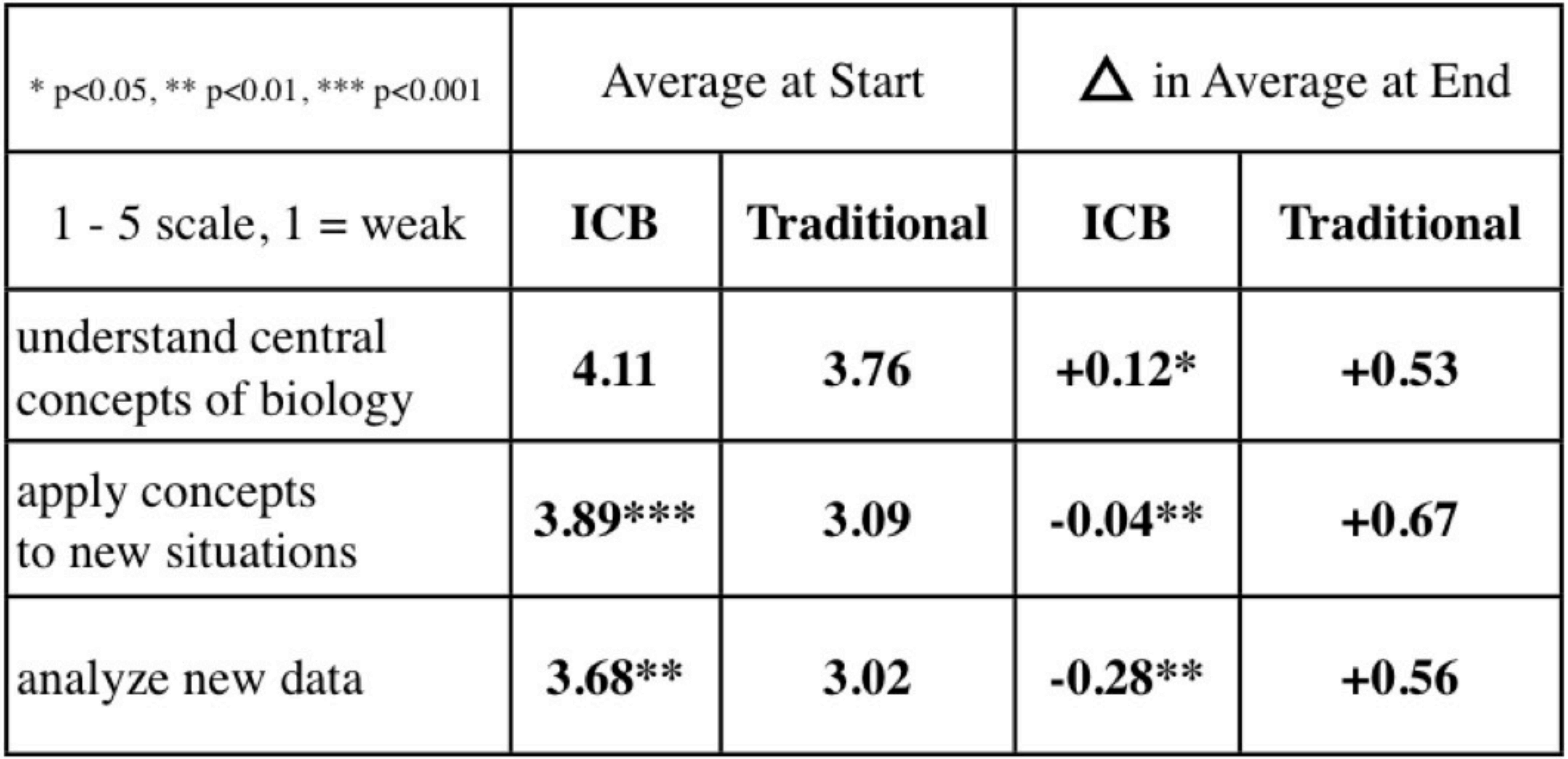


* p<0.05, ** p<0.01, *** p<0.001	Average at Start	
1 - 5 scale, 1 = weak	**ICB**	**Traditional**
understand central concepts of biology	**4.11**	**3.76**
apply concepts to new situations	**3.89*** **	**3.09**
analyze new data	**3.68** **	**3.02**

yes?

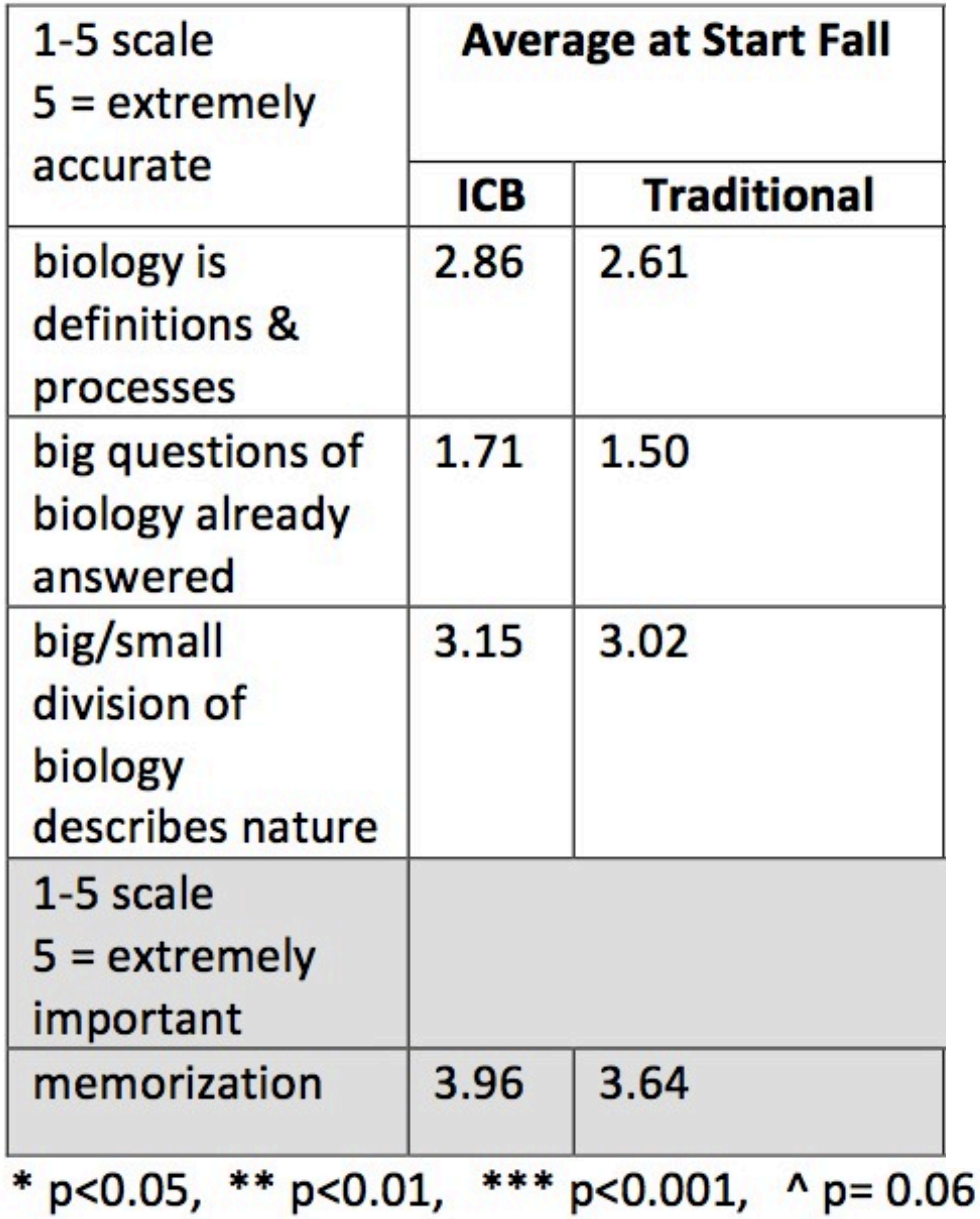
Are *ICB* students overconfident?

less so



* $p<0.05$, ** $p<0.01$, *** $p<0.001$	Average at Start		Δ in Average at End	
1 - 5 scale, 1 = weak	**ICB**	**Traditional**	**ICB**	**Traditional**
understand central concepts of biology	**4.11**	**3.76**	**+0.12***	**+0.53**
apply concepts to new situations	**3.89*** **	**3.09**	**-0.04****	**+0.67**
analyze new data	**3.68****	**3.02**	**-0.28****	**+0.56**

Do *ICB* students see biology differently?

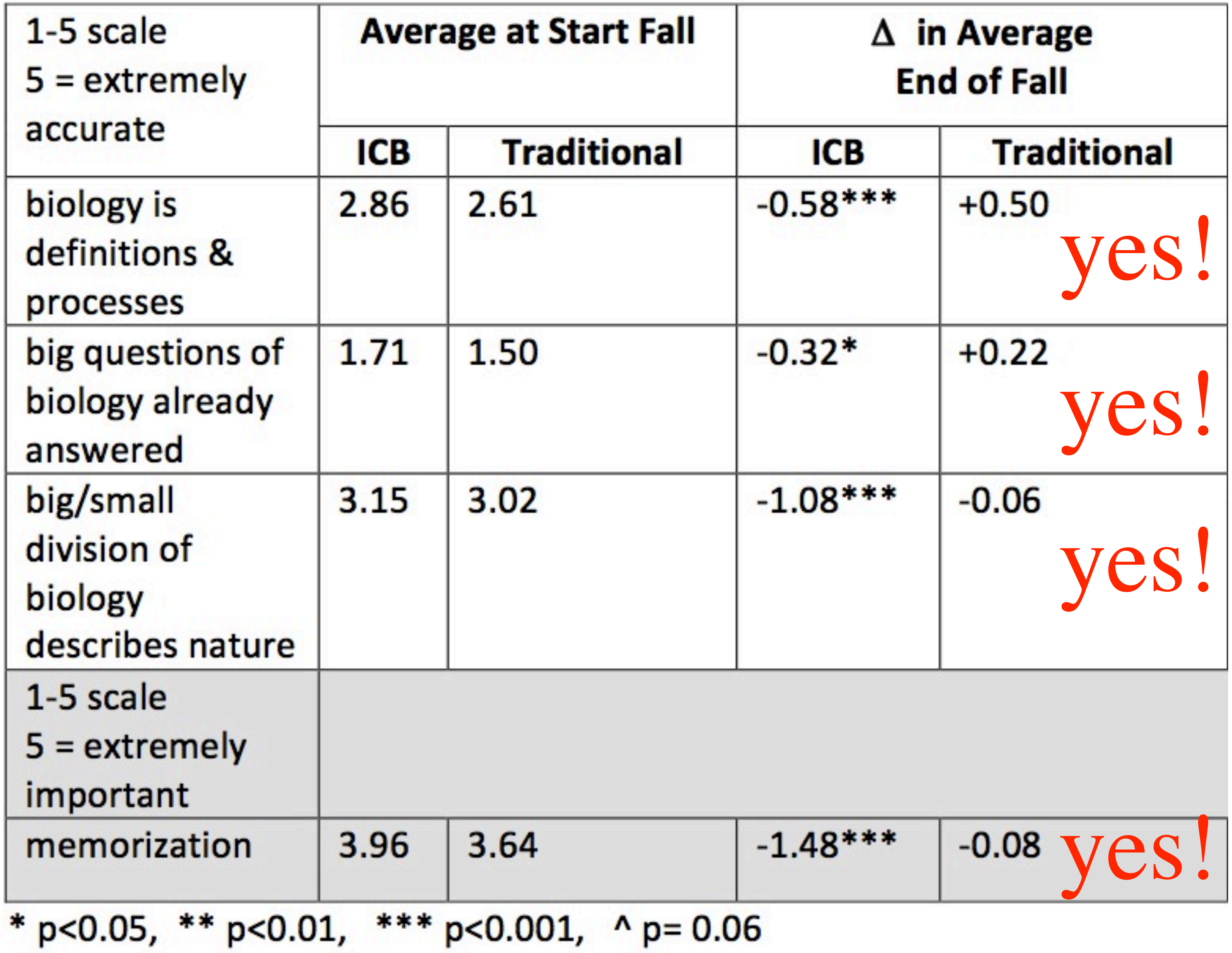


1-5 scale 5 = extremely accurate	Average at Start Fall	
	ICB	**Traditional**
biology is definitions & processes	2.86	2.61
big questions of biology already answered	1.71	1.50
big/small division of biology describes nature	3.15	3.02
1-5 scale 5 = extremely important		
memorization	3.96	3.64

* $p<0.05$, ** $p<0.01$, *** $p<0.001$, ^ $p= 0.06$

no

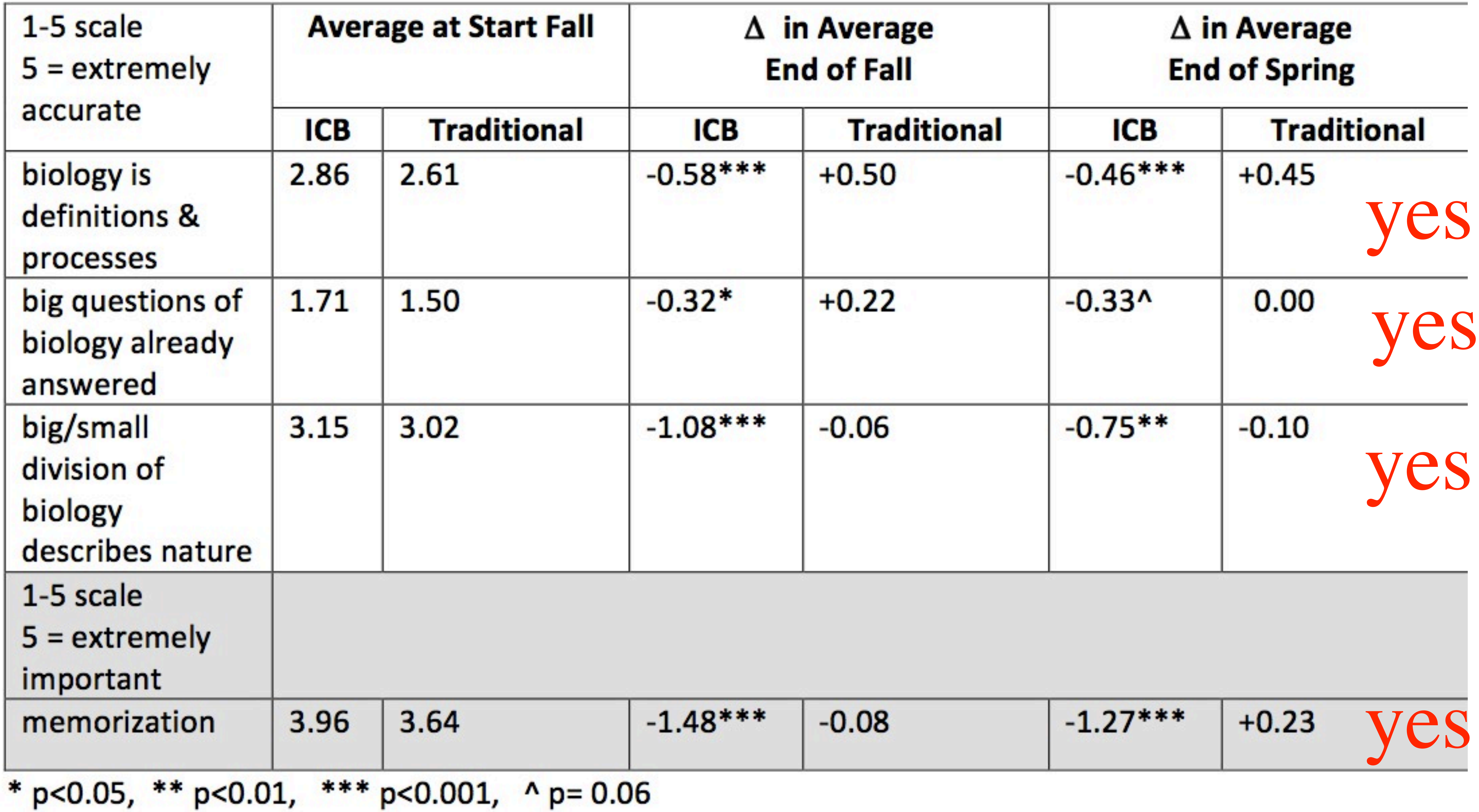
Do *ICB* students see biology differently?



1-5 scale 5 = extremely accurate	**Average at Start Fall**		**Δ in Average End of Fall**	
	ICB	**Traditional**	**ICB**	**Traditional**
biology is definitions & processes	2.86	2.61	-0.58***	+0.50 yes!
big questions of biology already answered	1.71	1.50	-0.32*	+0.22 yes!
big/small division of biology describes nature	3.15	3.02	-1.08***	-0.06 yes!
1-5 scale 5 = extremely important				
memorization	3.96	3.64	-1.48***	-0.08 yes!

* p<0.05, ** p<0.01, *** p<0.001, ^ p= 0.06

Do *ICB* students see biology differently?



1-5 scale 5 = extremely accurate	Average at Start Fall		Δ in Average End of Fall		Δ in Average End of Spring	
	ICB	Traditional	ICB	Traditional	ICB	Traditional
biology is definitions & processes	2.86	2.61	-0.58***	+0.50	-0.46***	+0.45
big questions of biology already answered	1.71	1.50	-0.32*	+0.22	-0.33^	0.00
big/small division of biology describes nature	3.15	3.02	-1.08***	-0.06	-0.75**	-0.10
1-5 scale 5 = extremely important						
memorization	3.96	3.64	-1.48***	-0.08	-1.27***	+0.23

* p<0.05, ** p<0.01, *** p<0.001, ^ p= 0.06

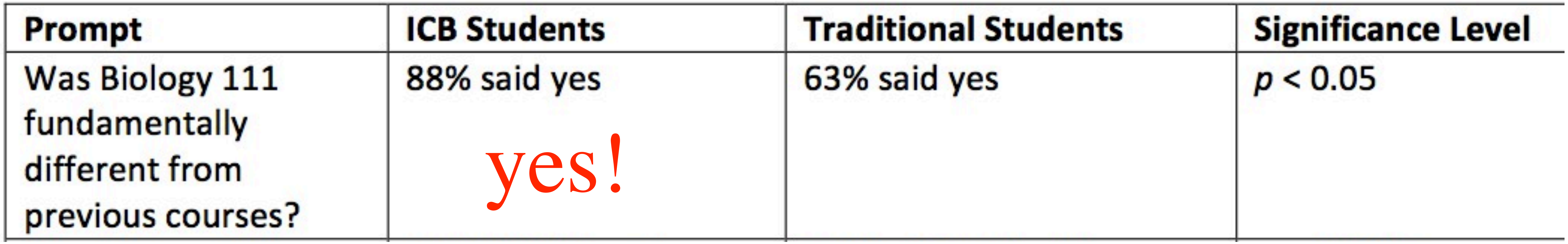
yes!

yes?

yes!

yes!

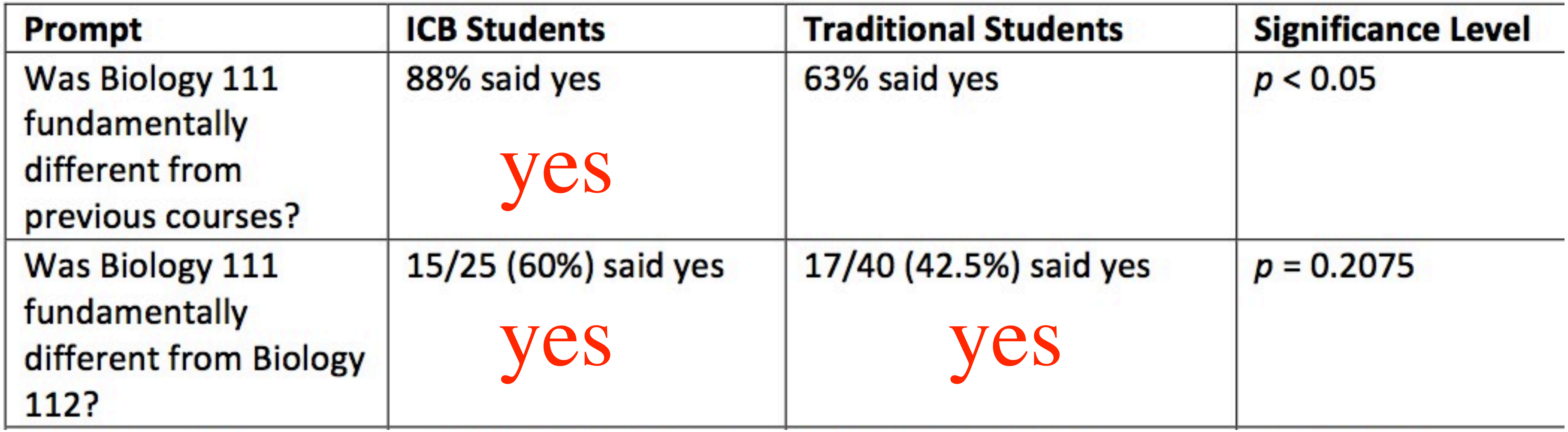
Do *ICB* students see difference in courses?



Prompt	**ICB Students**	**Traditional Students**	**Significance Level**
Was Biology 111 fundamentally different from previous courses?	88% said yes	63% said yes	$p < 0.05$

yes!

Do *ICB* students see difference in courses?



Prompt	ICB Students	Traditional Students	Significance Level
Was Biology 111 fundamentally different from previous courses?	88% said yes yes	63% said yes	$p < 0.05$
Was Biology 111 fundamentally different from Biology 112?	15/25 (60%) said yes yes	17/40 (42.5%) said yes yes	$p = 0.2075$

Do *ICB* students see difference in courses?

Prompt	**ICB Students**	**Traditional Students**	**Significance Level**
Was Biology 111 fundamentally different from previous courses?	88% said yes yes	63% said yes	$p < 0.05$
Was Biology 111 fundamentally different from Biology 112?	15/25 (60%) said yes yes	17/40 (42.5%) said yes yes	$p = 0.2075$
For those who answered yes above, did Biology 112 require more memorization than Biology 111?	12/15 (80%) said yes yes!	2/17 (12%) said yes	$p = 0.0002$