Part libraries for complex synthetic biology

Dr Tom Ellis

January 2019

Lecture Content

- Beyond BioBricks: scaling parts
- Promoter design & libraries
- Part design with the RBS Calculator
- Dealing with local sequence context
- Terminator libraries & design
- Transcription factor libraries
- CRISPRi repressors

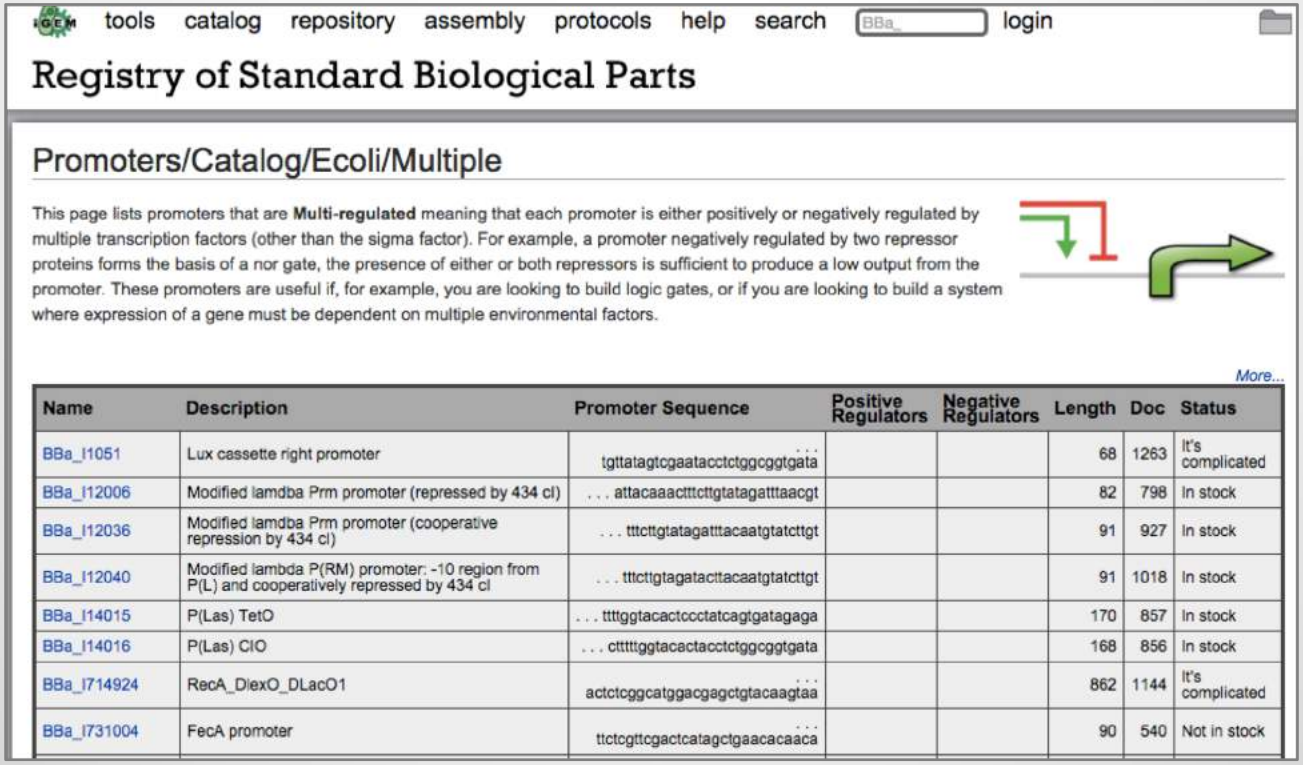
Learning Objectives

- To explain methods to generate part libraries
- To understand how small RNA parts can be designed using mathematical methods
- To introduce the issue of context dependency
- To show solutions to scaling regulators
- To introduce CRISPRi methods

The iGEM Parts Registry

http://parts.igem.org/Main_Page

- 20000+ BioBrick-formatted parts
- But questionable quality and data unreliable



tools catalog repository assembly protocols help search BBa_ login

Registry of Standard Biological Parts

Promoters/Catalog/Ecoli/Multiple

This page lists promoters that are **Multi-regulated** meaning that each promoter is either positively or negatively regulated by multiple transcription factors (other than the sigma factor). For example, a promoter negatively regulated by two repressor proteins forms the basis of a nor gate, the presence of either or both repressors is sufficient to produce a low output from the promoter. These promoters are useful if, for example, you are looking to build logic gates, or if you are looking to build a system where expression of a gene must be dependent on multiple environmental factors.

More...

Name	Description	Promoter Sequence	Positive Regulators	Negative Regulators	Length	Doc	Status
BBa_I1051	Lux cassette right promoter	...tgttatagtcgaatacctctggcggtgata			68	1263	It's complicated
BBa_I12006	Modified lamdba Prm promoter (repressed by 434 cI)	...attacaaactttcttgtatagatttaacgt			82	798	In stock
BBa_I12036	Modified lamdba Prm promoter (cooperative repression by 434 cI)	...tttcttgtatagatttacaatgtatcttgt			91	927	In stock
BBa_I12040	Modified lambda P(RM) promoter: -10 region from P(L) and cooperatively repressed by 434 cI	...tttcttgtagatacttacaatgtatcttgt			91	1018	In stock
BBa_I14015	P(Las) TetO	...ttttggtacactcoctatcagtgatagaga			170	857	In stock
BBa_I14016	P(Las) CIO	...ctttttggtacactacctctggcggtgata			168	856	In stock
BBa_I714924	RecA_DiexO_DLacO1	...actctcggcatggacgagctgtacaagtaa			862	1144	It's complicated
BBa_I731004	FecA promoter	...ttctcgttcgactcatagctgaacacaaca			90	540	Not in stock

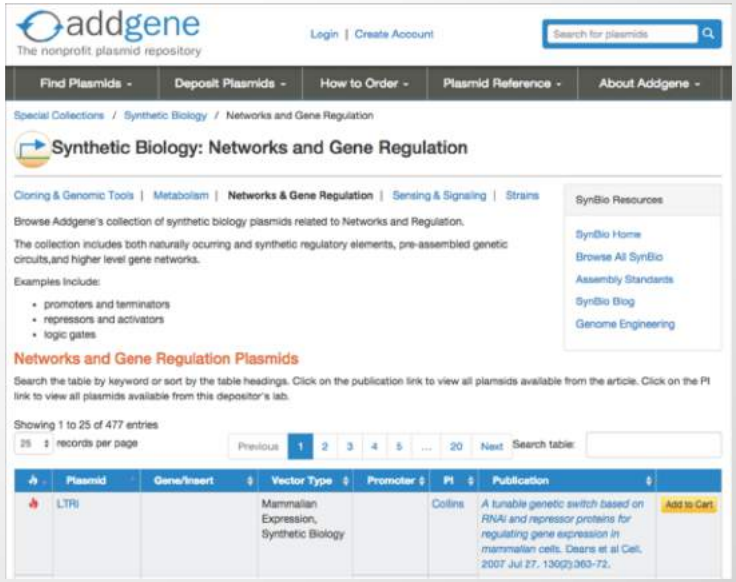
SynBERC, Addgene & BIOFAB

The alternative to iGEM is professional registries:

- SynBERC: a registry for some US synthetic biology
- Addgene: company to aid sharing published plasmids
- BIOFAB: a US effort to make professional parts

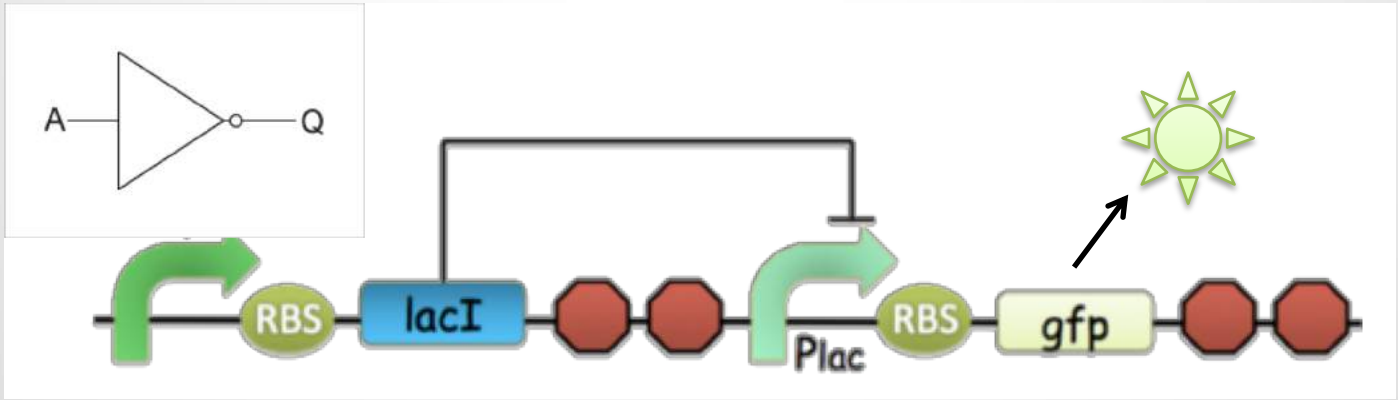
synberc
Building the Future with Biology

biofab



Our example: Inverter Network

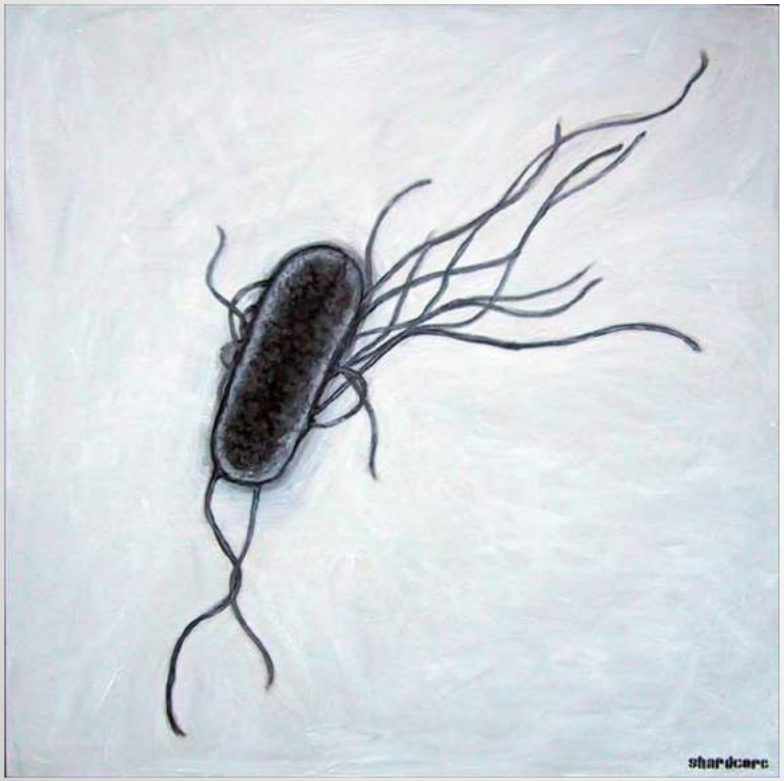
- Constitutive expression of a repressor that shuts off a downstream promoter
- Basic 'wire' device in synthetic gene networks
- Logic function: NOT



- **Scaling: why can't we make 100 working inverters?**

Bacterial systems: *E. coli*

- **Unless stated all content in this lecture is for parts and devices that work in *E. coli***



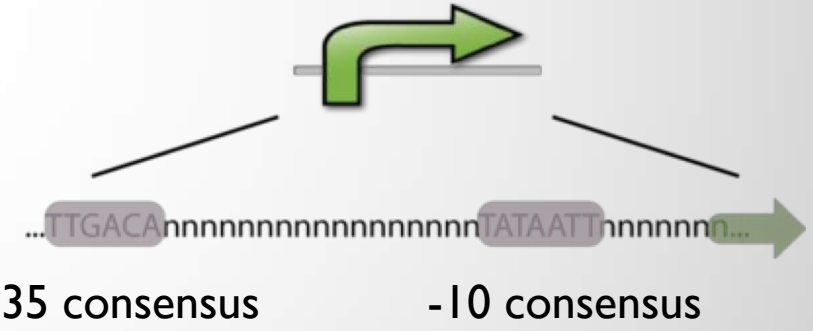
Parts and devices made for *E. coli* may not work in other bacteria and almost always don't work in eukaryotes like yeast and mammalian cells

… and *vice versa*

Constitutive Promoters

- Promoter design is a straightforward example of how to make a library of parts
- For constitutive *E. coli* promoters there are 2 main methods:

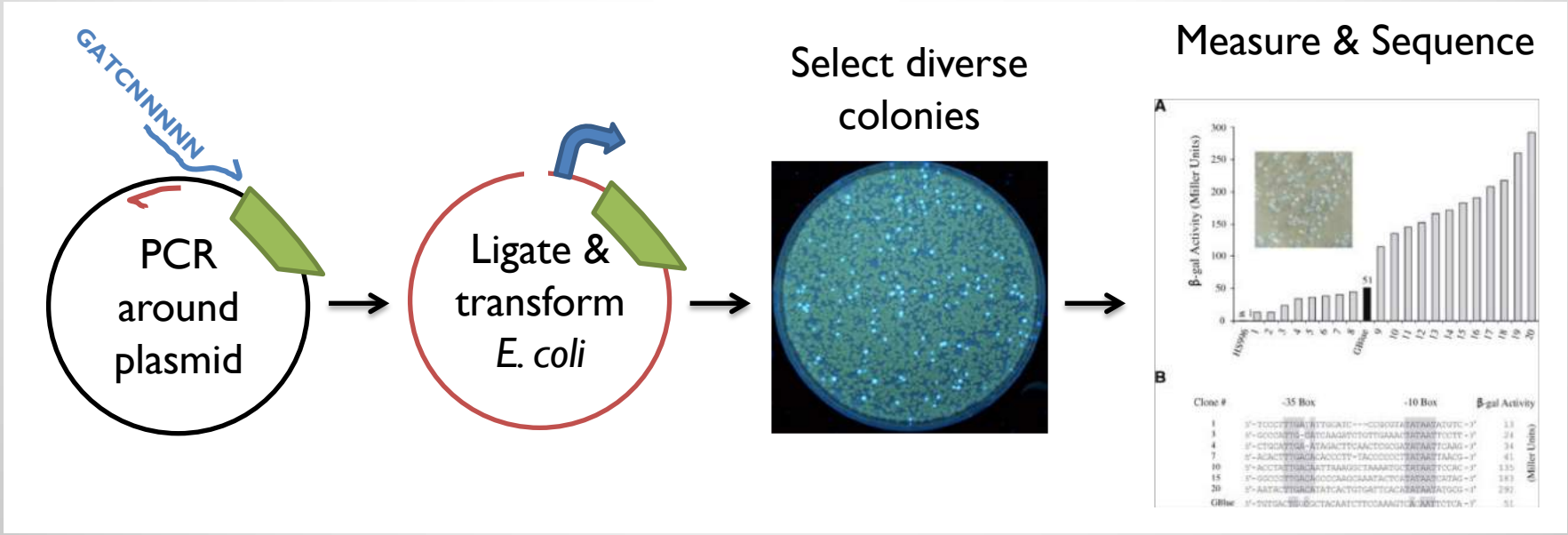
(a) Conservative mutation of consensus sequences



(b) Liberal mutation of sequences between the consensus sites using 'N' bases 'Synthetic Promoter Library' method: Jensen & Hammer 1998

Constitutive Promoter Libraries

- Constitutive *E. coli* promoters are short enough to be encoded on a primer
- Library can be made in a few days at low cost



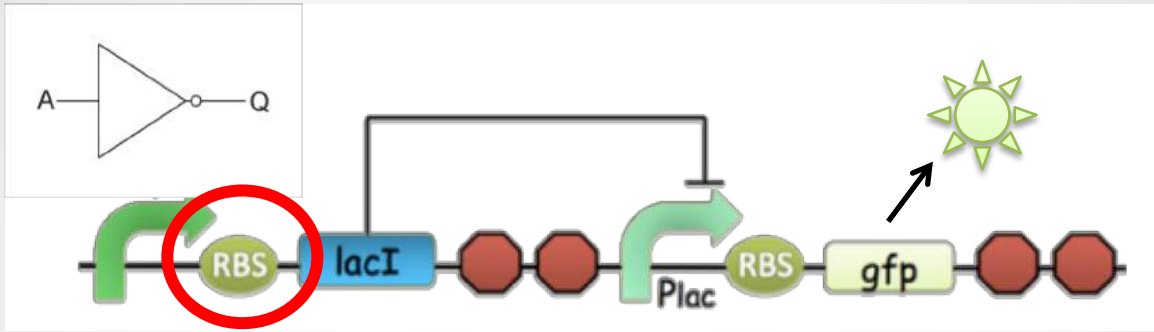
Anderson Promoter Library

- Example of promoter library made by method #1
- http://parts.igem.org/Promoters/Catalog/Anderson

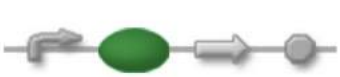
Anderson promoter collection

Identifier	Sequence[a]	Measured Strength[b]
BBa_J23119	ttgacagctagctcagtcctaggtataatgctagc	n/a
BBa_J23100	ttgacggctagctcagtcctaggtacagtgctagc	1
BBa_J23101	tttacagctagctcagtcctaggtattatgctagc	0.70
BBa_J23102	ttgacagctagctcagtcctaggtactgtgctagc	0.86
BBa_J23103	ctgatagctagctcagtcctagggattatgctagc	0.01
BBa_J23104	ttgacagctagctcagtcctaggtattgtgctagc	0.72
BBa_J23105	tttacggctagctcagtcctaggtactatgctagc	0.24
BBa_J23106	tttacggctagctcagtcctaggtatagtgctagc	0.47
BBa_J23107	tttacggctagctcagcctaggtattatgctagc	0.36
BBa_J23108	ctgacagctagctcagtcctaggtataatgctagc	0.51
BBa_J23109	tttacagctagctcagtcctagggactgtgctagc	0.04
BBa_J23110	tttacggctagctcagtcctaggtacaatgctagc	0.33
BBa_J23111	ttgacggctagctcagtcctaggtatagtgctagc	0.58
BBa_J23112	ctgatagctagctcagtcctagggattatgctagc	0.00
BBa_J23113	ctgatggctagctcagtcctagggattatgctagc	0.01
BBa_J23114	tttatggctagctcagtcctaggtacaatgctagc	0.10
BBa_J23115	tttatagctagctcagccttggtacaatgctagc	0.15
BBa_J23116	ttgacagctagctcagtcctagggactatgctagc	0.16
BBa_J23117	ttgacagctagctcagtcctagggattgtgctagc	0.06
BBa_J23118	ttgacggctagctcagtcctaggtattgtgctagc	0.56

Ribosome Binding Sites



Ribosome Binding Sites

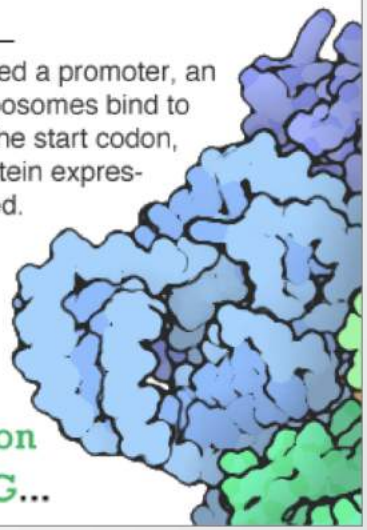


To make a protein in a cell you need a promoter, an RBS, a CDS and a terminator. Ribosomes bind to the RBS and begin translation at the start codon, which is ATG on the registry. Your choice of RBS will affect protein expression levels. RBSs are small so can be synthesized or assembled. We have have about 150 RBS parts on the Registry. The most used is BBa_E0034. There are many RBS collections on the Registry:

Anderson Collection | By expression level | *E. coli* | Eukaryotic

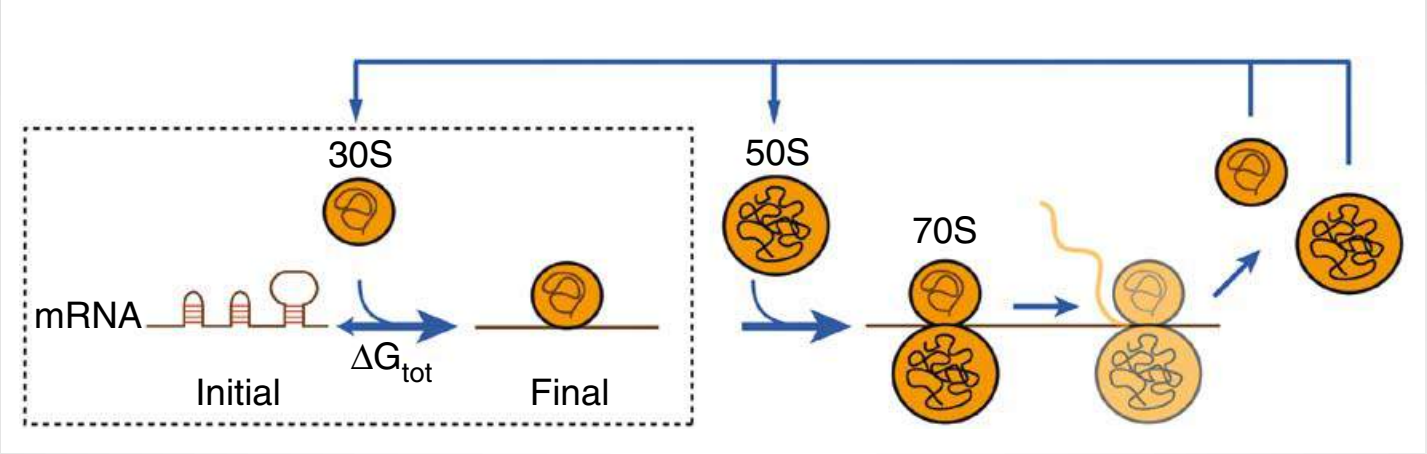
Promoter Ribosome binding site Start Codon

...TCTAGAGAAAGANNNGANNNACTAGATG...



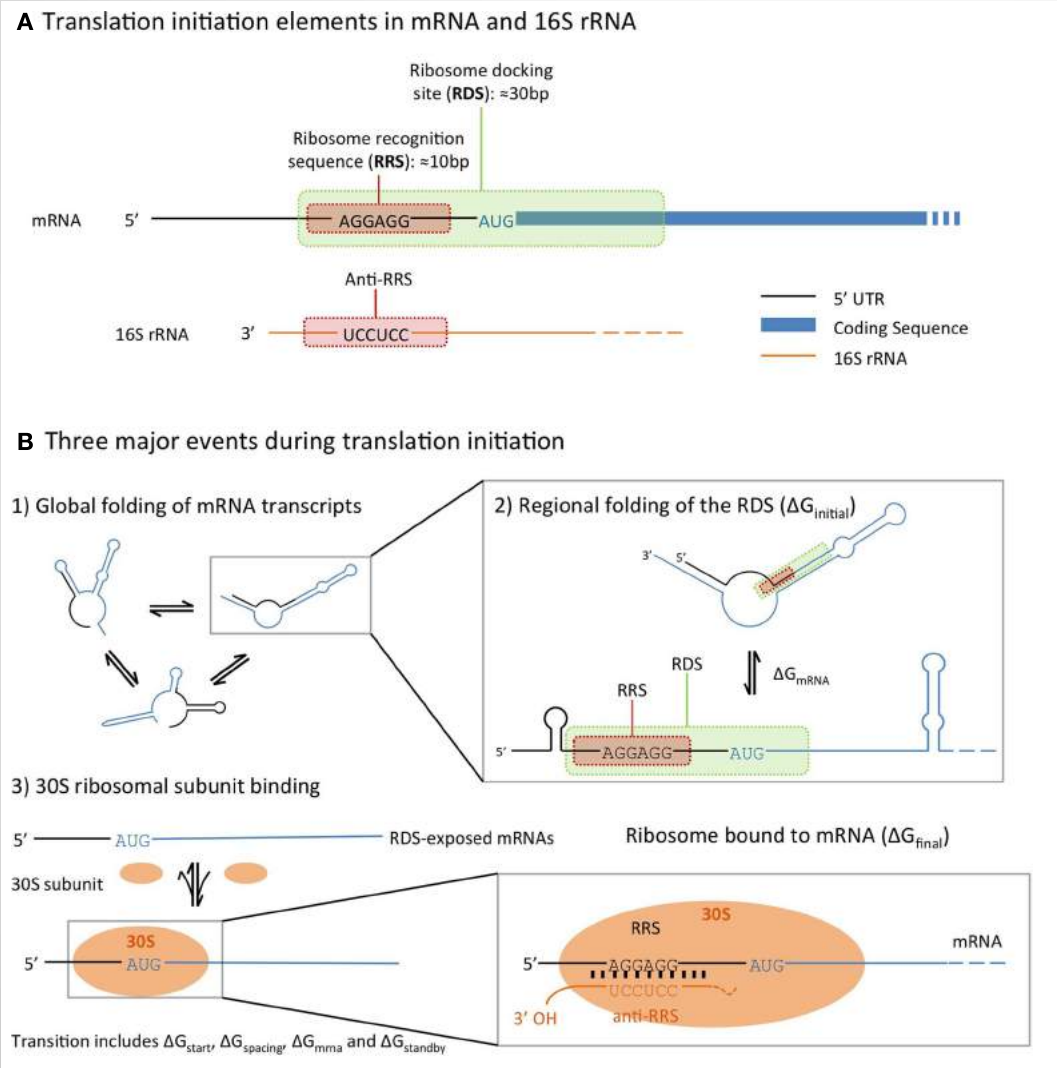
A part so small that possibly it could be designed….

Designing Ribosome Binding Sites



- Rate of translation largely determined by rate of translation initiation
- Initiation rate determined by interactions between RBS sequence on the mRNA and the part of the16S rRNA within the ribosome
- *Therefore rate is largely determined by RNA:RNA interactions*

RNA:RNA interactions at RBS



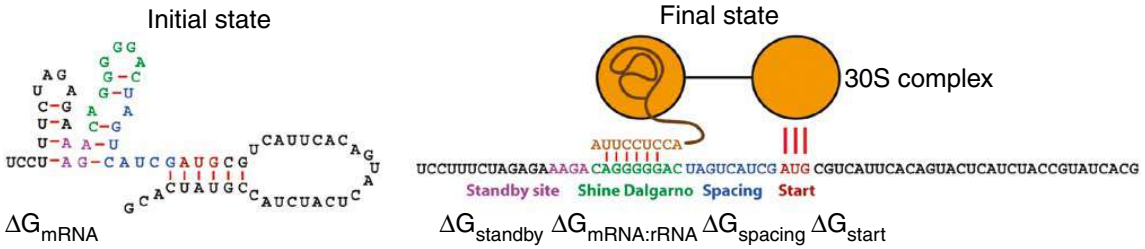
Rate of initiation at RBS can be estimated by calculating the rate of binding between mRNA and rRNA sequences

Gibbs free energy (ΔG) can be calculated for any DNA or RNA sequence because we know the energy of base-pairing

To do this use NUPACK
http://www.nupack.org/

Ribosome Binding Calculator

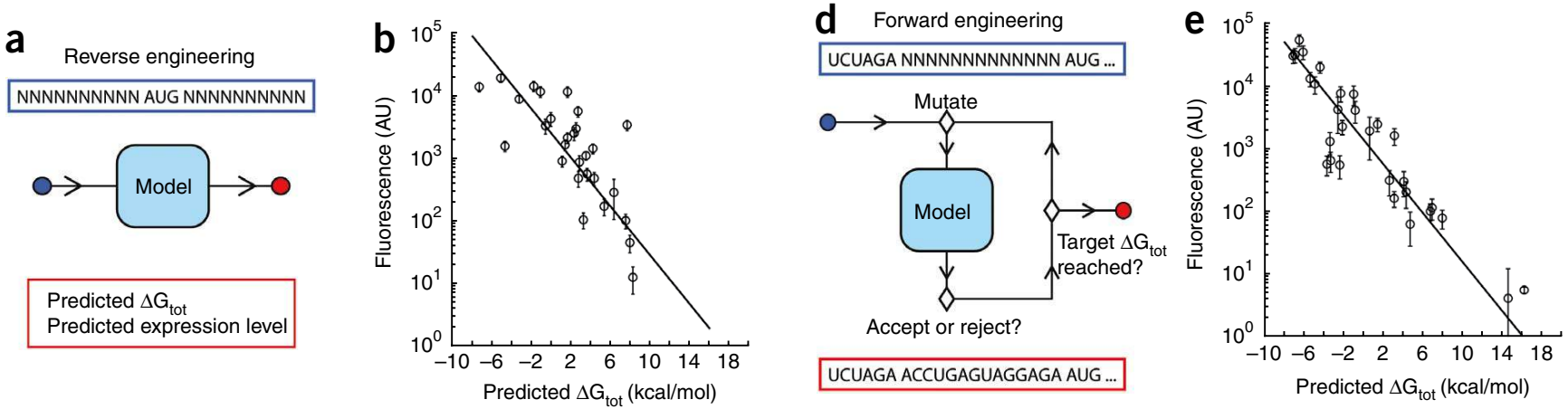
Salis *et al.* Nature Biotech 2009



Free energy calculated for any RBS sequence

Prediction of rate of translation from an RBS

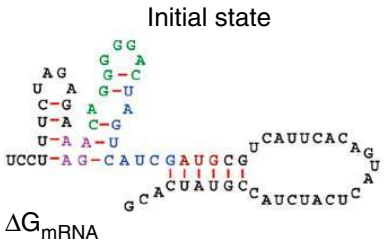
$$\Delta G_{tot} = \Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby} - \Delta G_{mRNA}$$



https://salis.psu.edu/software/

* sort of works 50% of the time

CONTEXT: RBS is not an isolated part



ΔG mRNA can only be calculated if RNA sequence upstream and downstream of RBS is included (>50 nt)

So sequence either side of RBS part influences the RBS

This feature is known as '***Context Dependency***'
i.e. the RBS strength is dependent on the local sequence context

Upstream: (promoter) and scar

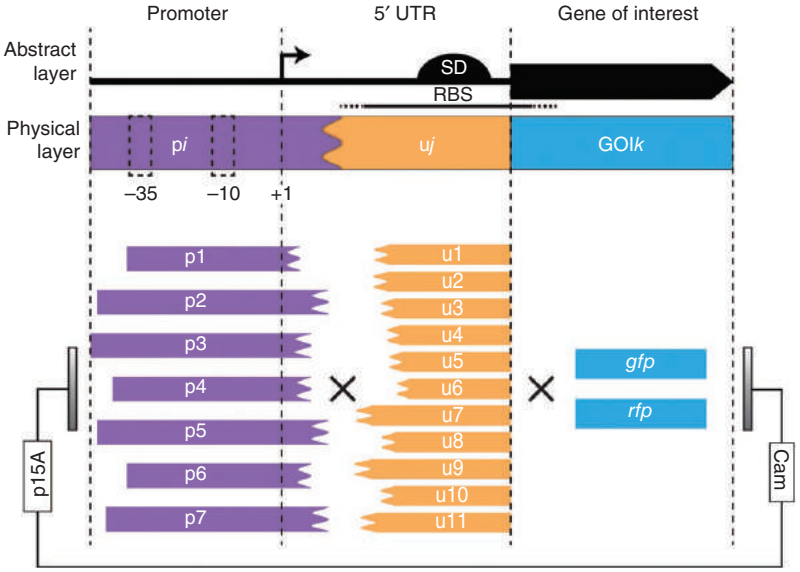
Downstream: gene CDS

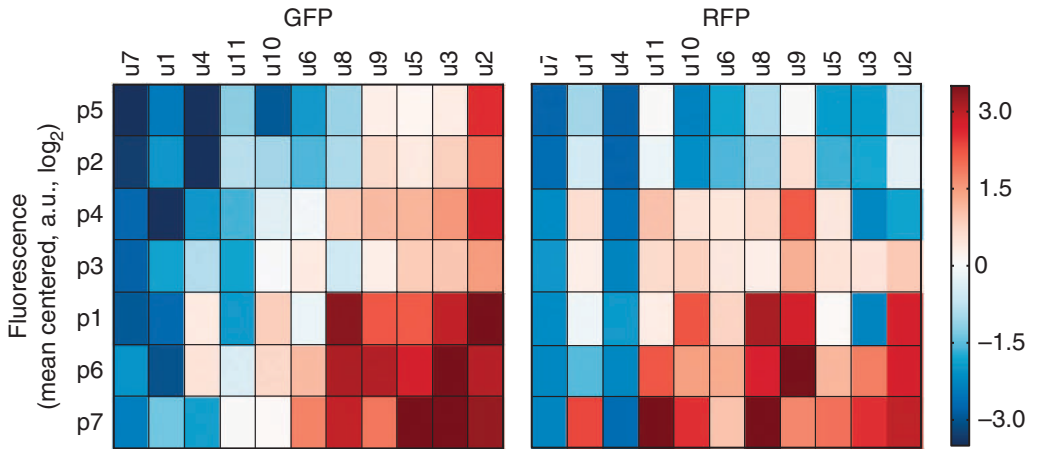
So… RBS strength <u>will</u> change when put in front of a different CDS

CONTEXT: a problem for synthetic biology

- Small DNA parts are like words affected by the surrounding sentence: e.g. "please set the table with a set of plates"

- Combining many different promoters, RBS and CDS parts doesn't lead to predictable gene expression output





Mutalik *et al.* Nature Methods 2013a

Alleviating Context

- To overcome context you either have to:

(a) Understand enough so you can predict its effect

  - RBS Calculator uses this approach: models effect of upstream and downstream part sequence on RBS part

(b) Use parts that remove context (i.e. insulators)

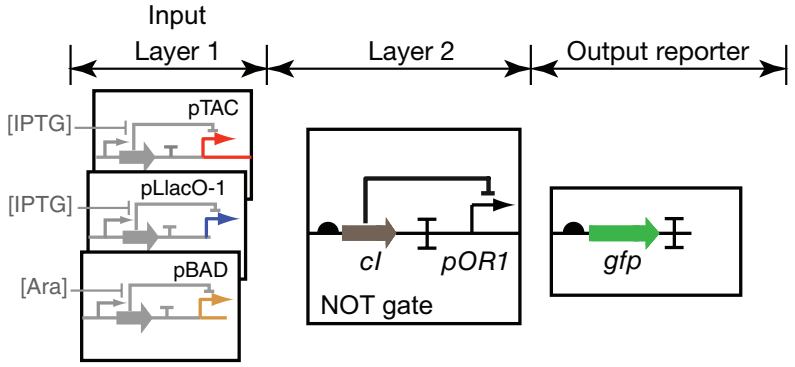
  - Three methods were developed for this…

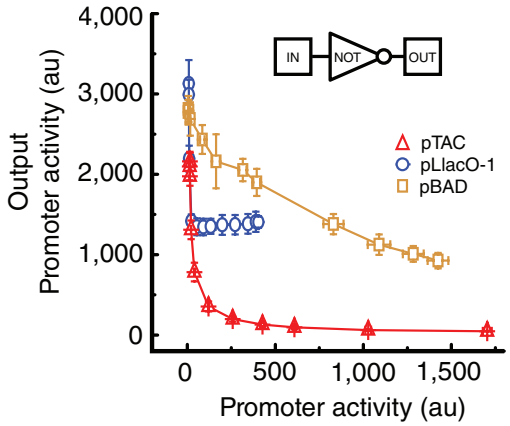
Alleviating Context: RiboJ Method

… but many important promoters don't end at the +1 site

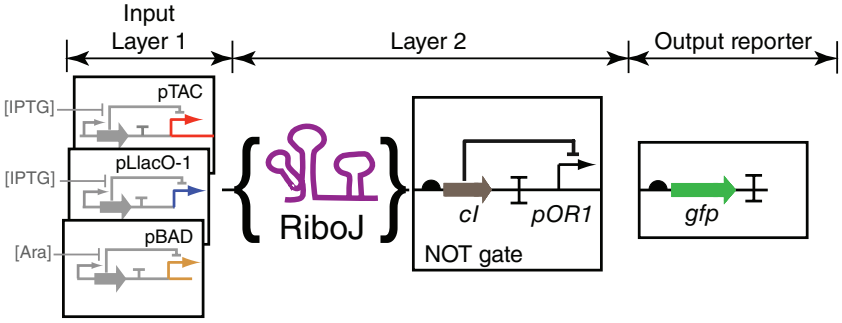
Regulated promoters often have sequence after +1 that get transcribed

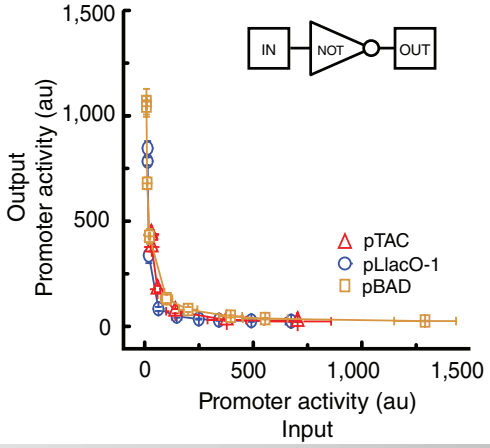
Different upstream promoters encode different 5' bases of mRNA & alter RBS strength of CDS





Solution: add a self-cleaving **ribozyme** part (RiboJ) between Promoter & RBS to leave a clean 5' UTR





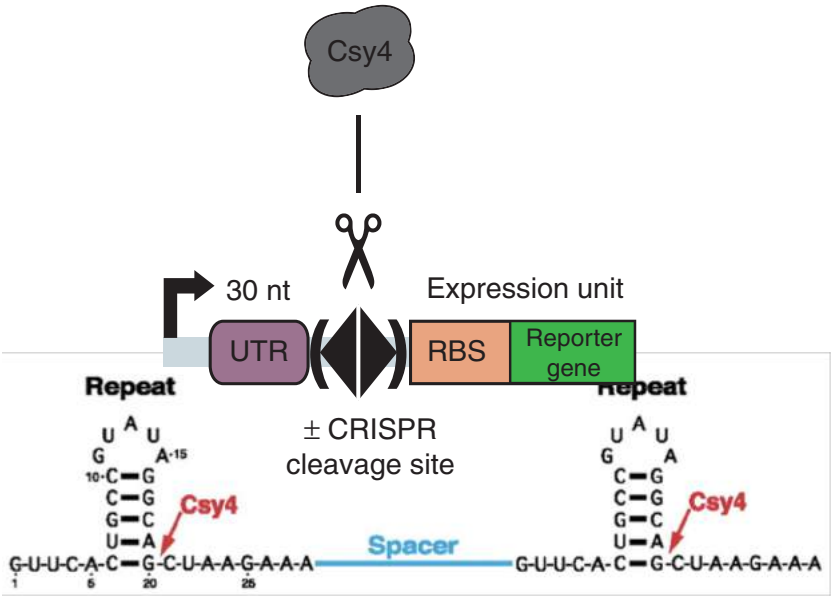
Lou, Stanton *et al.* Nature Biotech 2012

Alleviating Context: Csy4 Method

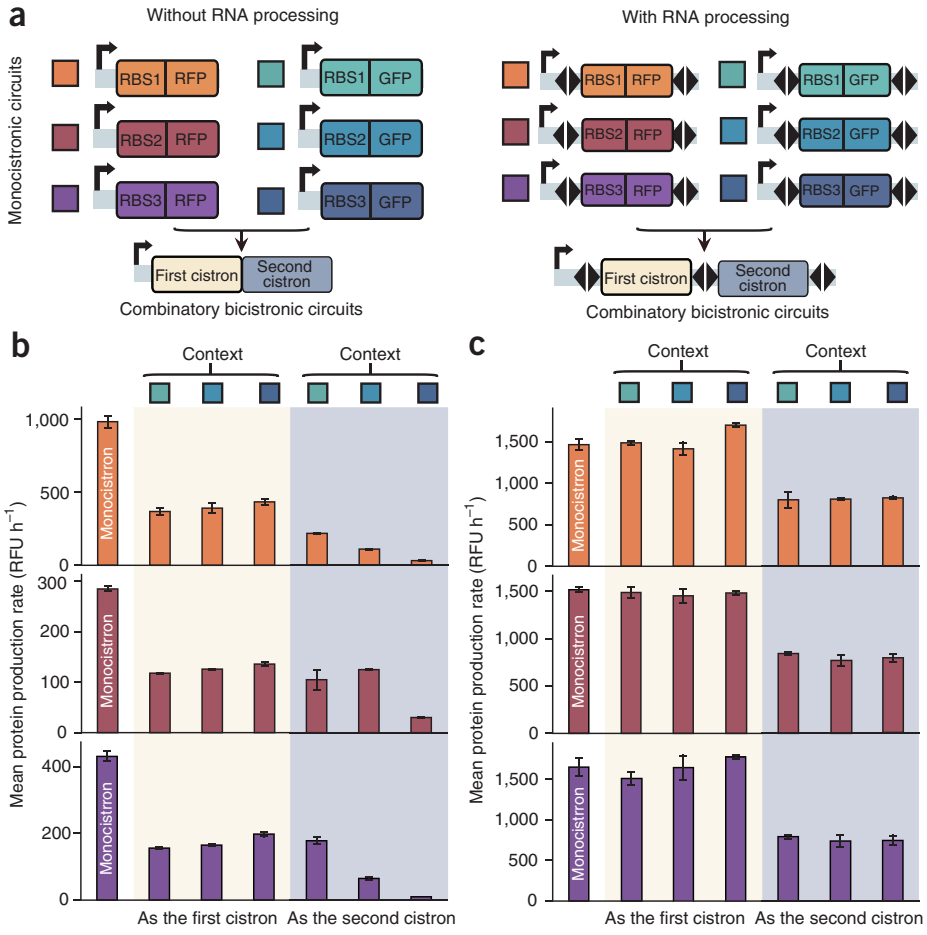
… but it would be a cumbersome to put RiboJ parts everywhere

Instead we can cut the mRNA using sequence-specific enzyme **Csy4**

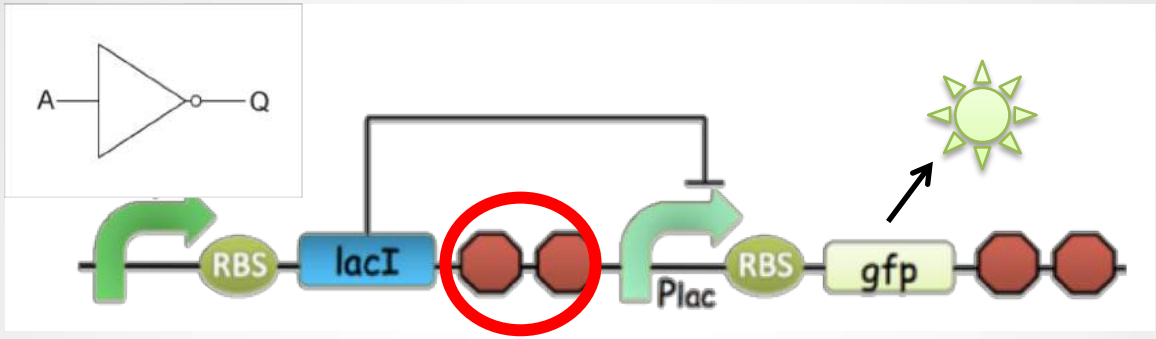
Csy4 enzyme expressed in *E. coli* will cut a short specific RNA sequence that makes a hairpin fold



Qi, Haurwitz *et al.* Nature Biotech 2012



Transcription Terminators



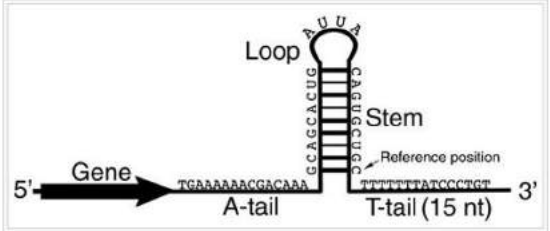
Terminators

< Back to Catalog

Terminators are genetic parts that usually occur at the end of a gene or operon and cause transcription to stop. In prokaryotes, terminators usually fall into two categories (1) rho-independent terminators and (2) rho-dependent terminators.

Rho-independent terminators are generally composed of palindromic sequence that forms a stem loop rich in G-C base pairs followed by several T bases. The conventional model of transcriptional termination is that the stem loop causes RNA polymerase to pause and transcription of the poly-A tail causes the RNA:DNA duplex to unwind and dissociate from RNA polymerase.

All the *E. coli* terminators in the Registry are rho-independent terminators. Rho-dependent terminators are not included, because rho-dependent terminators are not specified by sequence.



A rho-independent transcriptional terminator. Image courtesy of Kingsford *et al.*.

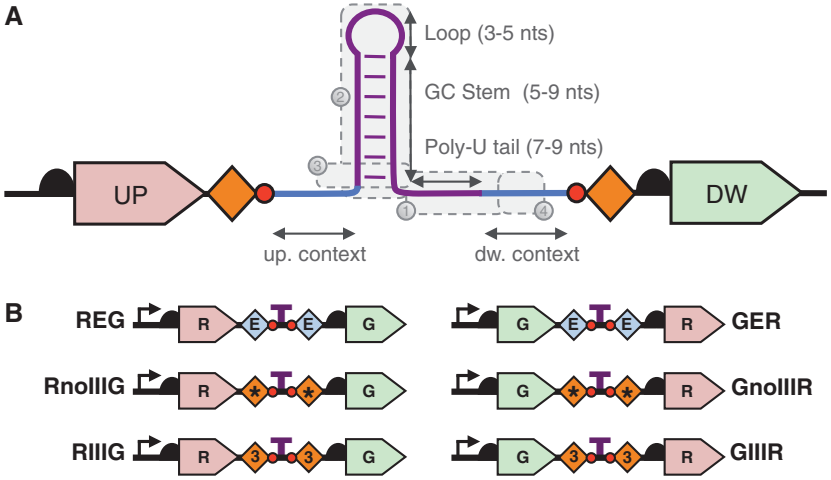
Catalog: Are you looking for a terminator to use? The registry has a collection of bacterial, yeast, and eukaryotic terminators.

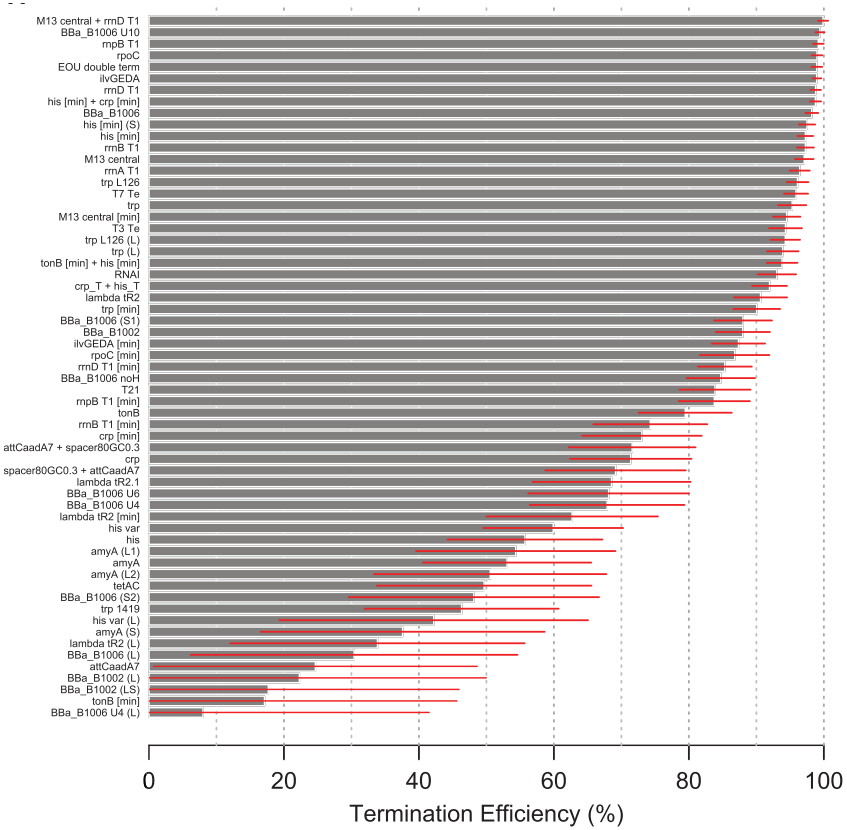
- Terminators: boring parts but a challenge for cloning
- Not a good idea to repeatedly use the same one

Transcription Terminator Libraries I

BIOFAB designed and characterised 100s of terminators & used data to model further design

- Terminator is an RNA stem/loop
- Measure by placing between reporter genes





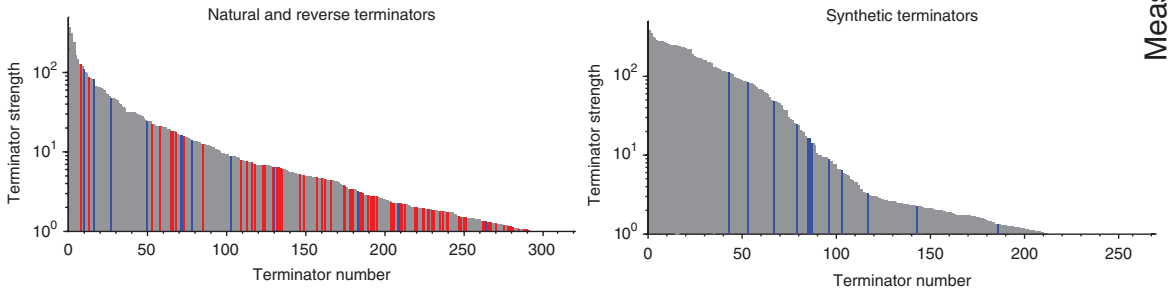
Cambray *et al.* Nucleic Acids Research 2013

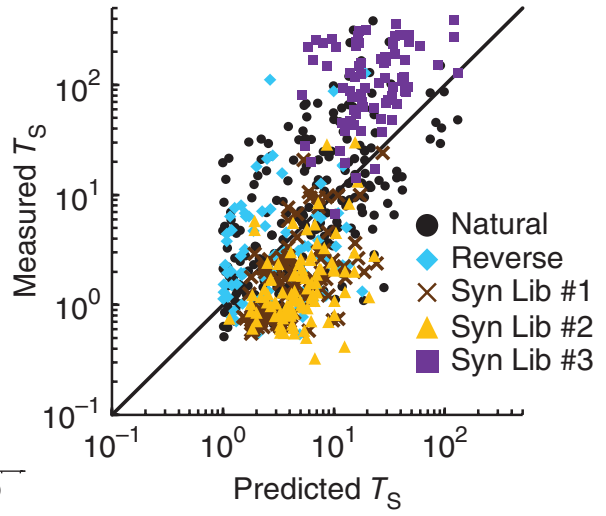
Transcription Terminator Libraries 2

Voigt lab characterised 500+ terminators and used data to make a 'terminator calculator'

- Measured natural and designed terminators
- Derived a biophysical model of RNA folding to explain efficiency
- Equation is 'somewhat' predictive

$$T_S = 1 + \frac{1}{B_1 e^{\beta_1 \Delta G_L} + B_4 e^{\beta_4 (\Delta G_B + \Delta G_A - \Delta G_U)} (1 + B_1 e^{\beta_1 \Delta G_L})}$$





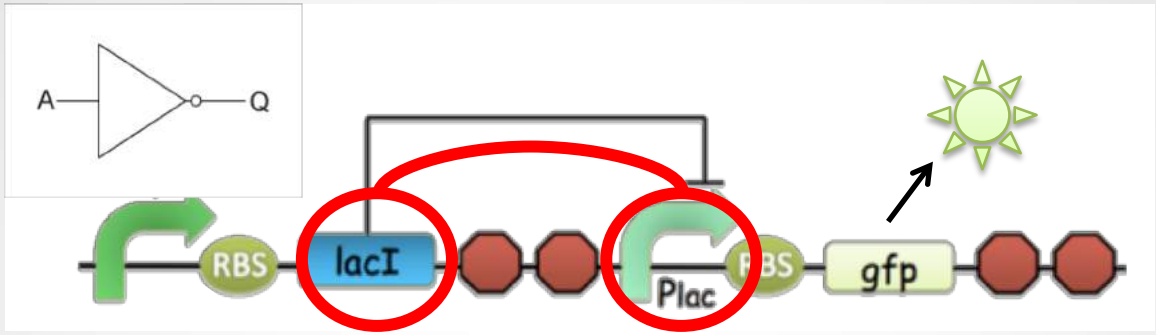
Chen, Lui, Nielsen *et al*. Nature Methods 2013

We're halfway there…

- Constitutive Promoter Libraries
- RBS Designs and RBS Libraries
- Tricks to account for 'context' effects
- Terminator Libraries

- Time to now get a bit more complex...

Regulators & Regulated Promoters



These are the key pairs of parts that enable logic

TetR and pTet (-)
LacI and pLac (-)
LuxR and pLux (+)

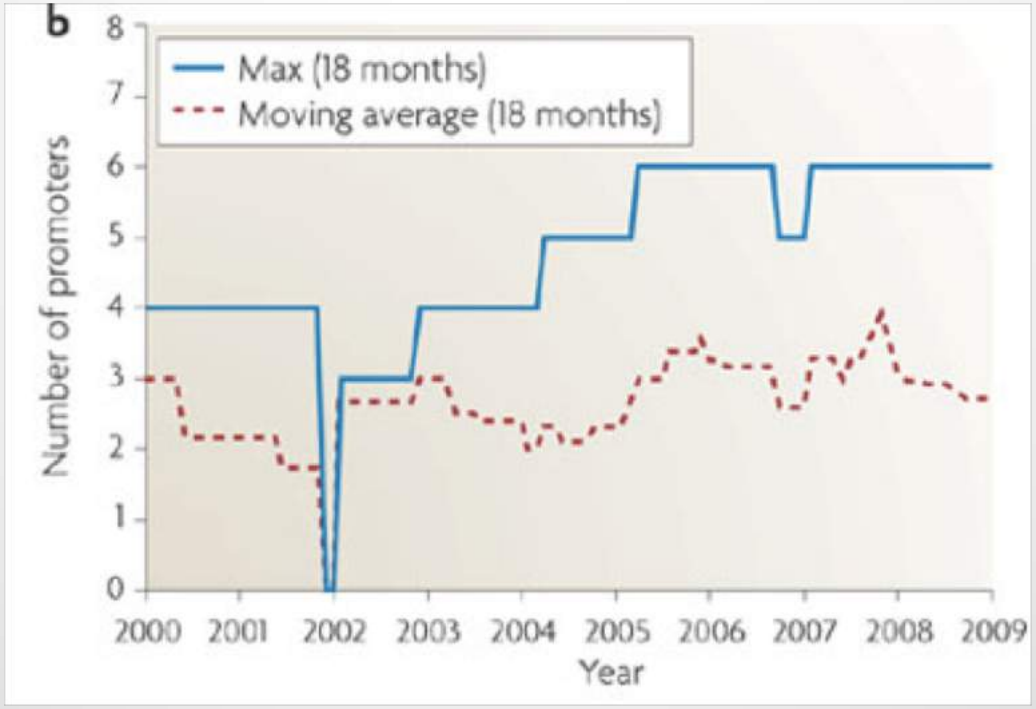
AraC and pAraBAD (+/-)
cI and pOR1(-)
OmpR-P and pOmpC (+)

Classic transcription factor/promoter pairs behave differently

For scalable logic we need hundreds of predictable pairs

Regulators & Regulated Promoters

Complexity of devices can't increase without a large ***orthogonal set*** of predictable regulators and promoter pairs



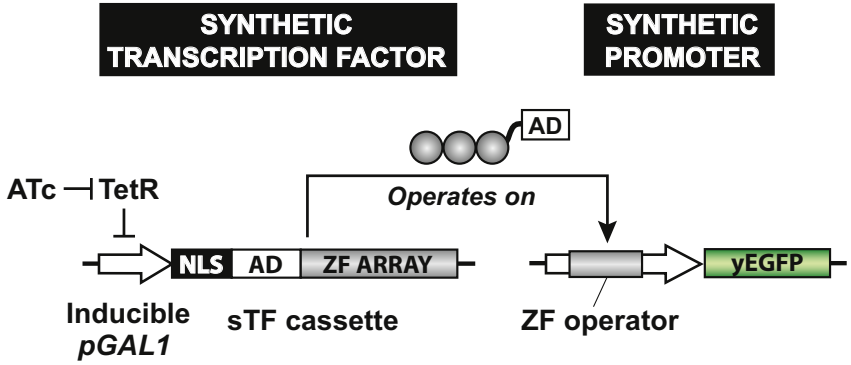
Purnick & Weiss. Nature Reviews Mol Cell Biol. 2009

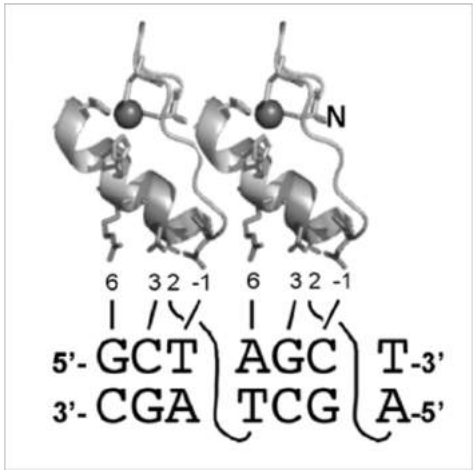
1: Modular transcription factors: Zinc Finger Proteins

Zinc Finger Transcription Factors can be designed to recognise different promoters

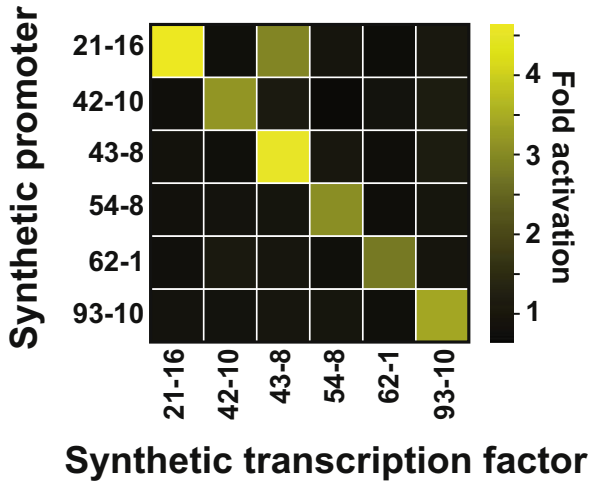
Khalil *et al.* made a great set of synthetic TFs paired to synthetic promoters

…but they don't work in *E. coli* (only yeast)





	ZINC FINGER RESIDUES			SYNTHETIC PROMOTER OPERATORS		
sTF	Finger 1	Finger 2	Finger 3	EcoRI	Binding sequence	BamHI
13-6	TNQKLEV	VRHNLQR	QHPNLTR	GAATTC	a GAA GAT GGT g	GGATTC
14-3	APSKLDR	LGENLRR	DGGNLGR	GAATTC	g GAC GAC GGC a	GGATTC
21-16	RNFILQR	QGGNLVR	QQTGLNV	GAATTC	a TTA GAA GTG a	GGATTC
36-4	GRQALDR	DKANLTR	QRNNLGR	GAATTC	c GAA GAC GCT g	GGATTC
37-12	RNFILQR	DRANLRR	RHDQLTR	GAATTC	t GAG GAC GTG t	GGATTC
42-10	TGQILDR	VAHSLKR	DPSNLRR	GAATTC	a GAC GCT GCT c	GGATTC
43-8	RQDRLDR	QKEHLAG	RRDNLNR	GAATTC	a GAG TGA GGA c	GGATTC
54-8	NKTDLGR	RRDMLRR	RMDHLAG	GAATTC	a TGG GTG GCA t	GGATTC
55-1	DESTLRR	MKHHLGR	RSDHLSL	GAATTC	c TGG GGT GCC c	GGATTC
62-1	TGQRLRI	QNQNLAR	DKSVLAR	GAATTC	g GCC GAA GAT a	GGATTC
92-1	DSPTLRR	QRSSLVR	ERGNLTR	GAATTC	a GAT GTA GCC t	GGATTC
93-10	APSKLKR	HKSSLTR	QRNALSG	GAATTC	c TTT GTT GGC a	GGATTC
97-4	RQSNLSR	RNEHLVL	QKTGLRV	GAATTC	a TTA TGG GAG a	GGATTC
129-3	TAAVLTR	DRANLTR	RIDKLGD	GAATTC	c GGG GAC GTC a	GGATTC
150-4	KGERLVR	RMDNLST	RKDALNR	GAATTC	g GTG TAG GGG t	GGATTC
151-1	IPNHLAR	QSAHLKR	QDVSLVR	GAATTC	t GCA GGA GGT g	GGATTC
158-2	DKTKLRV	VRHNLTR	QSTSLQR	GAATTC	t GTA GAT GGA g	GGATTC
172-5	MKNTLTR	RQEHLVR	QKPHLSR	GAATTC	a GGA GGG GCT c	GGATTC
173-3	SAQALAR	QQTNLAR	VGSNLTR	GAATTC	a GAT GAA GCT g	GGATTC



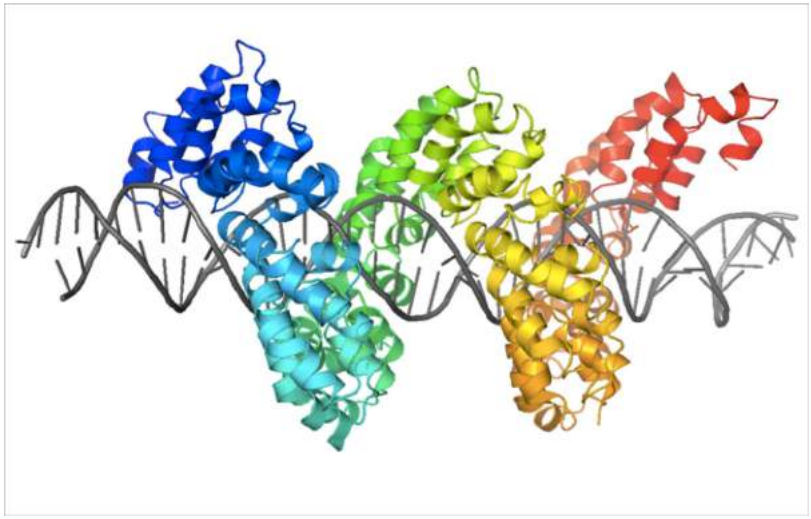
Khalil, Lu, *et al.* Cell. 2012

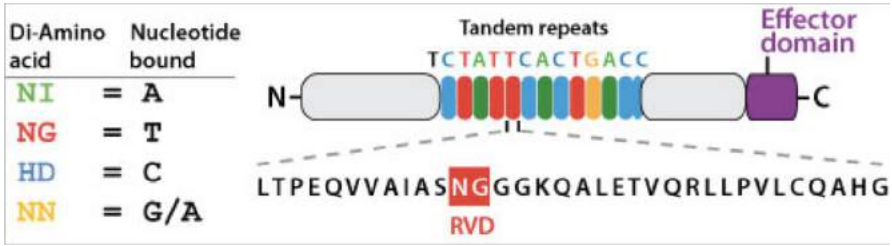
2: Modular transcription factors: TAL Effectors

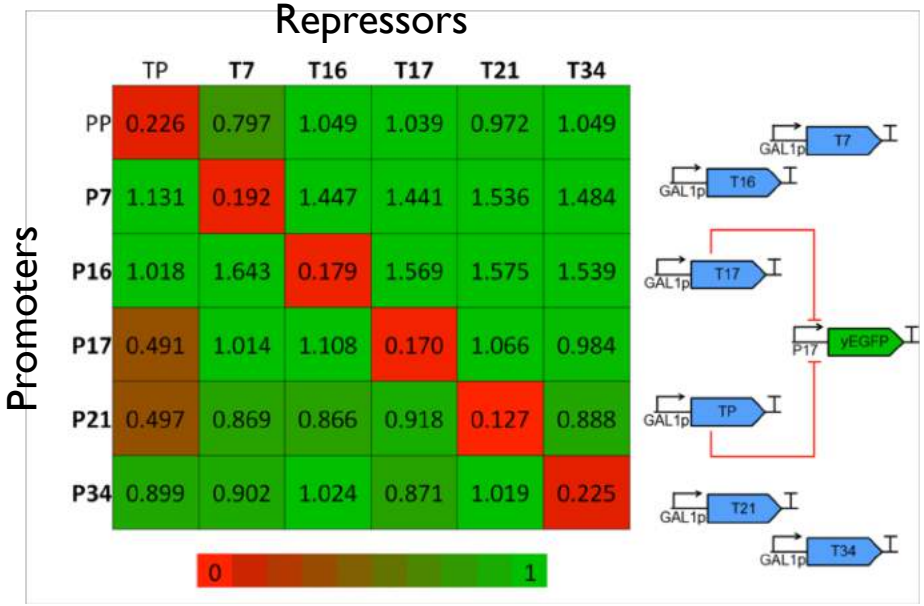
TAL Effector Proteins can be designed to bind specifically to any DNA sequence

Blount *et al.* made a great set of synthetic TFs paired to synthetic promoters

…but they don't work in *E. coli (again!)*



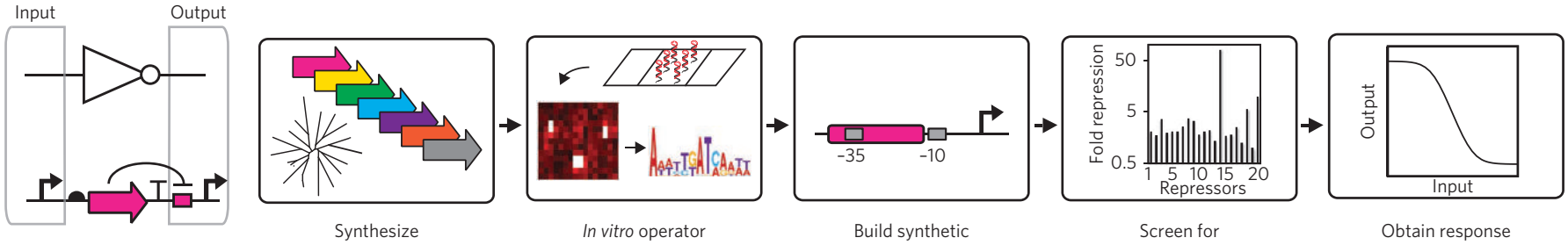




Blount *et al. Unpublished Data*

3: 'Part Mining' for Orthogonal Regulators

To get an orthogonal set of repressors and promoter pairs for *E. coli* Stanton *et al.* 'mined' DNA diversity from microbe genome sequences



Found proteins similar to TetR from sequence databases, synthesised them and worked out what DNA sequence they bind

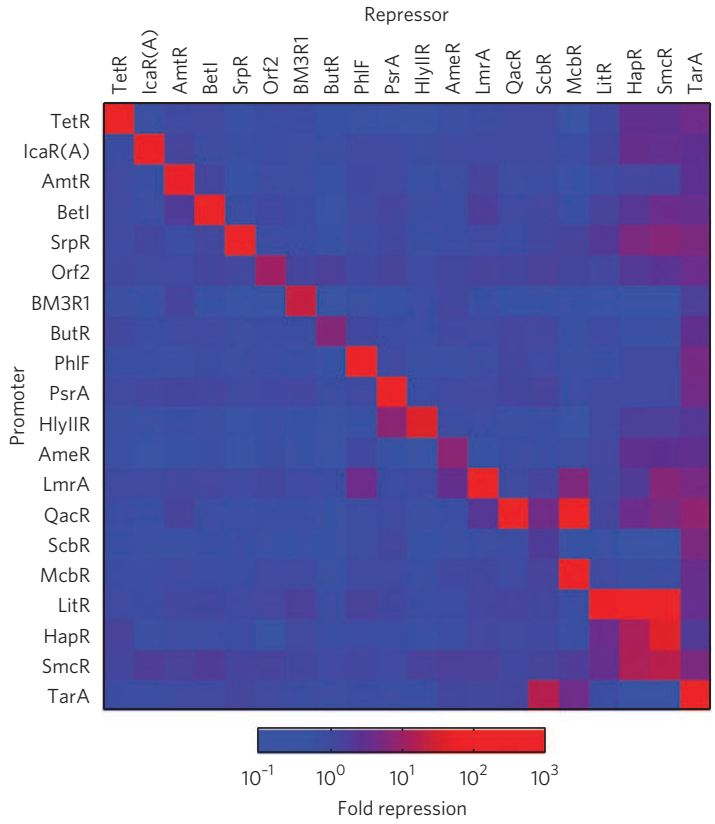
AmeR	gatAGTGACAAAC	TTGACA	ACTCATCACTtcctagg	tataat	gctagctactagag**aaagaggagaaa**t
AmtR	gattcgttaccaa	ttgaca	gTTTCTATCGATCTATA	GATAAT	gctagctactagag**aaagaggagaaa**t
BetI	gattcgttaccaa	ttgaca	ATTGATTGGACGTTCAA	TATAAt	gctagctactagag**aaagaggagaaa**t
BM3R1	gattcgttaccaa	ttgaCG	GAATGAACGTTCATTCC	Gataat	gctagctactagag**aaagaggagaaa**t
ButR	gattcGTGTCACT	TTGACA	GCAGTGTCACtcctagg	tataat	gctagctactagag**aaagaggagaaa**t
HapR	gattcgttaccaa	ttgaca	gctagctcTTATTGATT	TTTAAT	CAAATAAtactagag**aaagaggagaaa**
HlyIIR	gattcgttaccaa	ttgacA	TATTTAAAATTCTTGTT	TAAAat	gctagctactagag**aaagaggagaaa**t
IcaR	gattcgttaccaa	ttgaca	aTTCACCTACCTTTCGT	TAGGTT	AGGTTGTtactagag**aaagaggagaaa**
LitR	gattcgttaccaa	tTGACA	AATTTATAAATTGTCAg	tataat	gctagctactagag**aaagaggagaaa**t
LmrA	gattcgttaccaa	ttgaca	actggtggtcgaatcaa	GATAAT	AGACCAGTCACTATATTTtactagag**a**
McbR	gattcgttaccaa	ttgaca	ATAGAAAGATCTGTCTA	tataat	gctagctactagag**aaagaggagaaa**t
Orf2	gattcgttaccaa	ttgaca	CTAACTGCTGTTCAGTT	AGGTTg	ctagc**aaagaggagaaa**tactagatgg
PhlF	gattcgttaccaa	ttgacA	TGATACGAAACGTACCG	TATCGT	TAAGGTtactagag**aaagaggagaaa**t
PsrA	GGAACAAACGTTTGA	TTGAca	gctagctcagtcctagg	tataat	gctagctactagag**aaagaggagaaa**t
QacR	gattcgttaccaa	ttgaca	gctagctcagtcctaCT	TTAGTA	TAGAGACTGAGCGGTCGGTCTATAtac
ScbR	gattcgttaccaa	ttgaca	gctagctATCATACCGC	TATAAT	GGTATGTTtactagag**aaagaggagaa**
SmcR	gattcgttaccaa	ttgaca	TTATTGATAAATCTGCG	TAAAAT	gctagctactagag**aaagaggagaaa**t
SrpR	gattcgttaccaa	ttgaca	gctagctcagtcctagg	tATATA	CATACATGCTTGTTTGTTTGTAAACta
TarA	gattcgttaccaa	ttgaca	gctAAACATACCGTGTG	GTATGT	TCtagctactagag**aaagaggagaaa**t
TetR	tcagtgatagaga	ttgaca	TCCCTATCAGTGATAGA	tataat	gagcactactagag**aaagaggagaaa**t

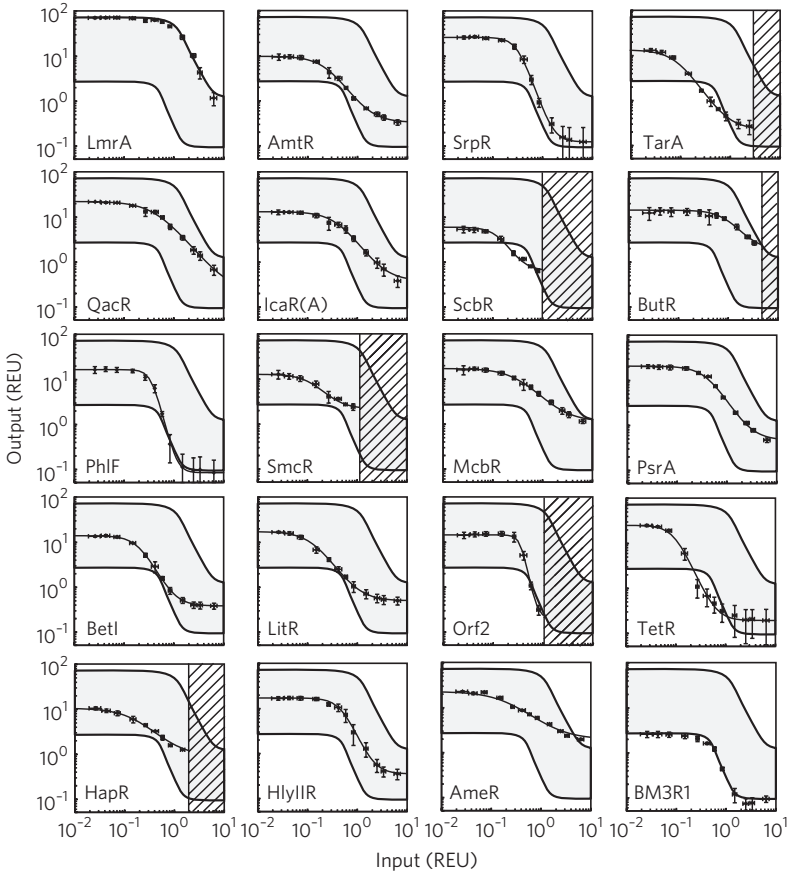
Stanton *et al.* Nature Chem Biol 2014

3: 'Part Mining' for Orthogonal Regulators

73 TetR-family repressors identified, 16 show strong specific repression

Project yielded 16 orthogonal NOT gates (TetR plus 15 new ones)





Stanton *et al.* Nature Chem Biol 2014

CRISPR – a 'game-changer'

- CRISPR = Clustered Regularly Interspaced Short Palindromic Repeats

- Bacterial immune system where RNA sequences related to phages are made and guide a DNA-cutting enzyme (e.g. Cas9) to cut any DNA that matches the 'guide' RNA sequence

- Together CRISPR guide RNAs and Cas9 cut DNA

CRISPR interference (CRISPRi)

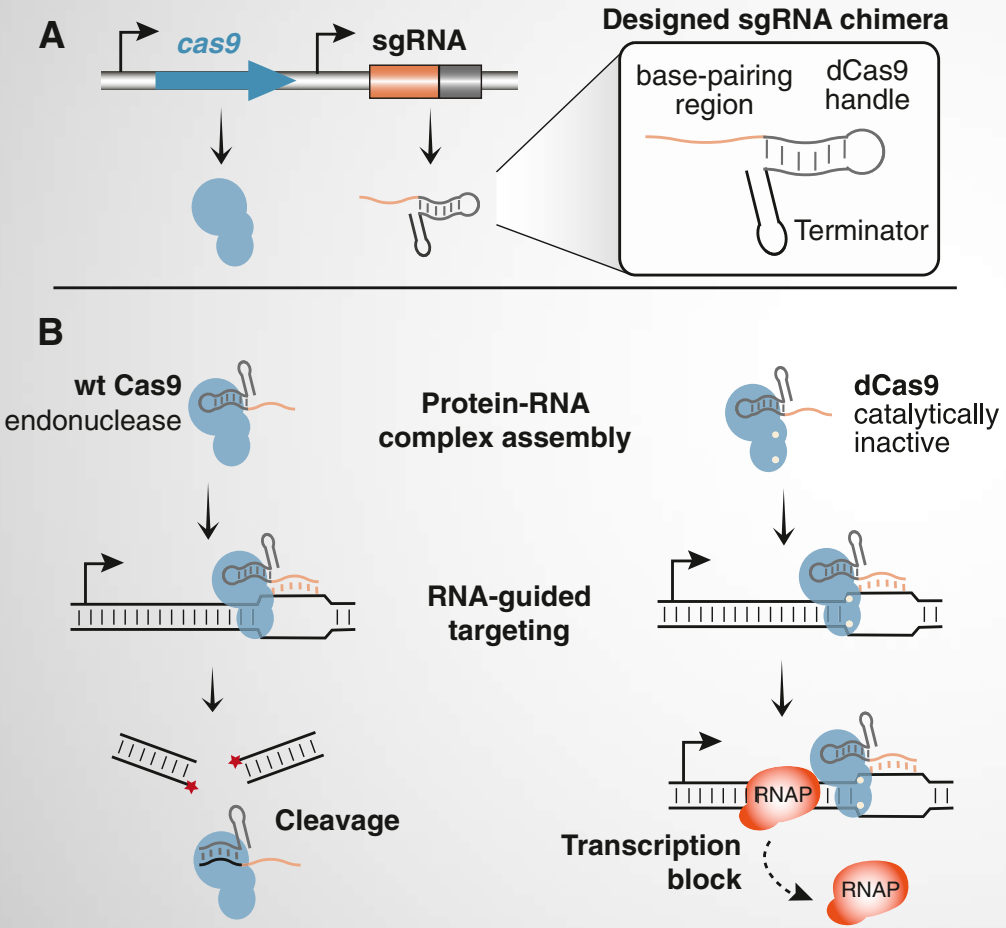
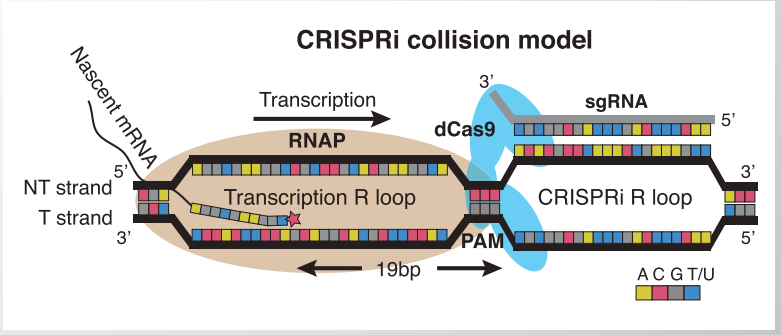


Figure 1. Design of the CRISPR Interference System

CRISPR/Cas is an RNA-guided DNA cutting system

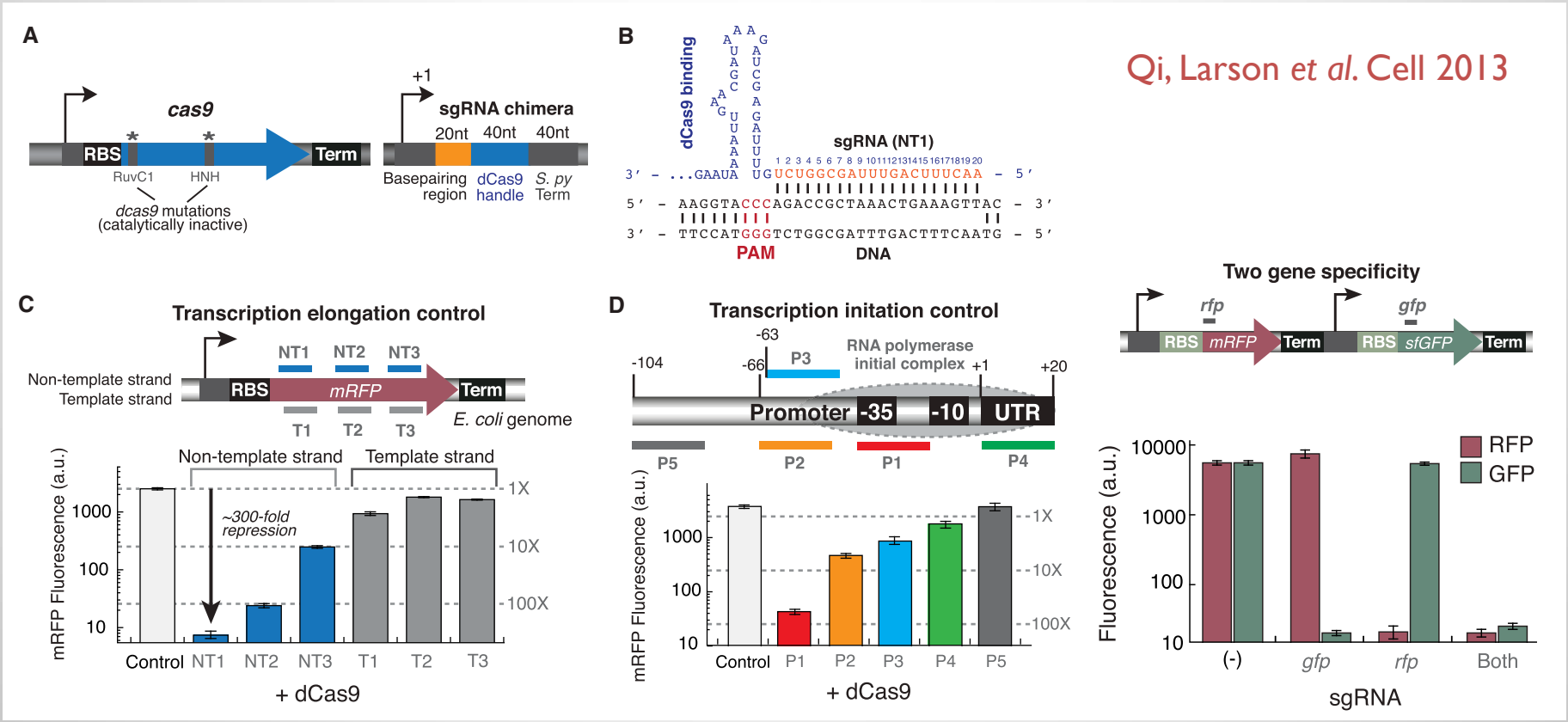
Cas9 enzyme binds to a synthetic guide RNA (sgRNA) which matches a DNA target

dCas9 is deactivated Cas9: mutation means it binds DNA & represses instead of cutting

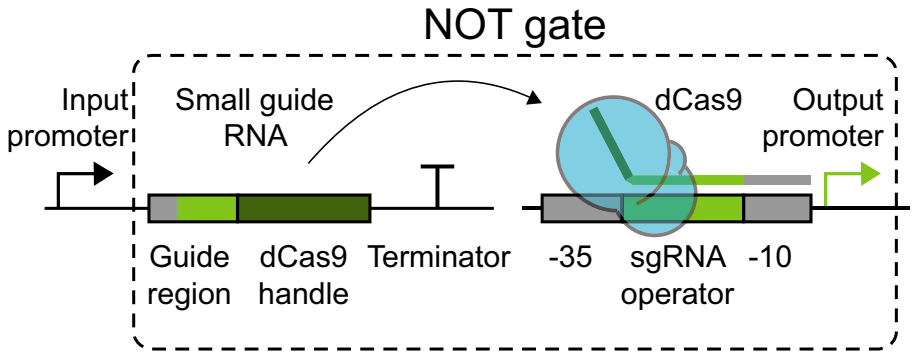


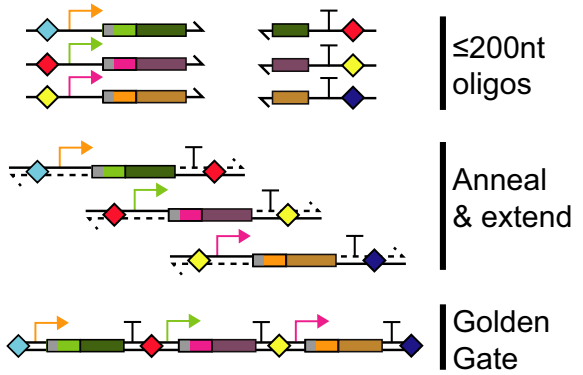
Guiding dCas9 for CRISPRi

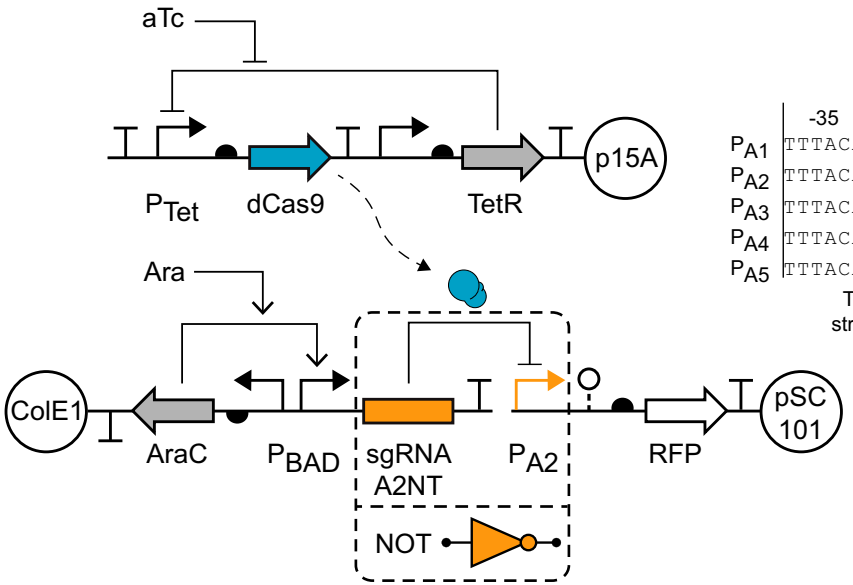
Sequence-specific repression can be seen when guide RNAs target dCas9 to:
(a) elongating non-template strand or (b) core promoter template strand

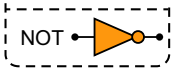


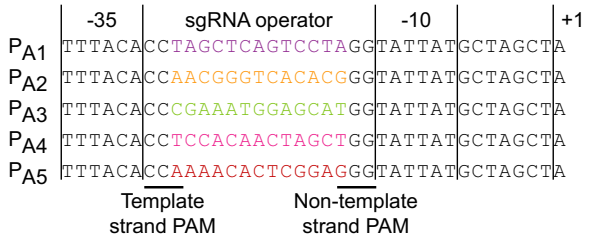
Scalable Regulation with CRISPRi

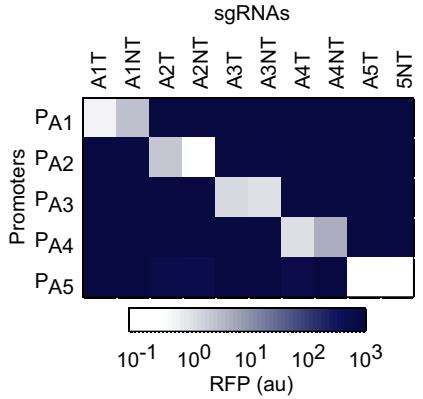








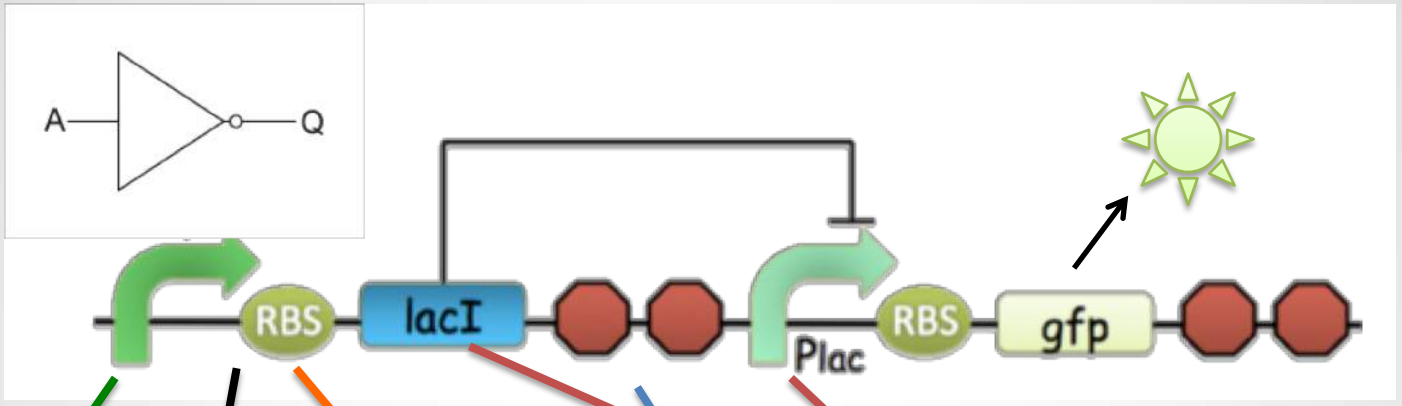




CRISPRi gives an orthogonal set of regulators and promoter pairs

Nielsen and Voigt. Molecular Systems Biology 2014

Scalable Inverter Networks



- 100+ constitutive promoters
- RiboJ part or Csy4 sites to remove 5' context
- RBS custom design using RBS Calculator
- BCD method gives 50+ modular RBS parts
- 16 TetR-related TF-promoter pairs\
- ZF-TFs/ TALES in yeast
- 10000+ dCas9 options
- 600+ measured terminators
- Two terminator design models

Summary

- We now have 100s of parts for each key position in genetic networks
- Short parts can be (somewhat) designed
- Context issues present a challenge
- 'Mining' for parts yields orthogonal pairs
- We now have enough parts to rationally make dozens of different inverters in *E. coli* and Yeast
- CRISPRi is making life even easier

Example Exam Questions

- How do you make and characterise promoter and terminator libraries?
- How does the RBS Calculator work?
- What is context dependency and how it can be tackled?
- Why are orthogonal regulators important and give examples of these?

Key References

Original Research

The RBS Calculator
Salis *et al*. Nature Biotech 2009

RiboJ Insulators
Lou, Stanton *et al*. Nature Biotech 2012

Set of TetR-family logic gates
Stanton *et al*. Nature Chem Biol 2014

Key References

REVIEWS

Principles of Gene Circuit Design
Brophy and Voigt. Nature Methods 2014

How different RBS Calculators work
Reeve *et al*. Frontiers in Bioengineering & Biotechnology 2014

Context dependency
Cardinale and Arkin. Biotechnology Journal 2012

CRISPRi development
Qi, Larson *et al*. Cell 2013

Other References

BIOFAB Context Assessment
Mutalik *et al*. Nature Methods 2013a

Csy4 Insulators
Qi, Haurwitz *et al*. Nature Biotech 2012

BIOFAB Terminators
Cambray *et al*. Nucleic Acids Research 2013

Voigt Lab Terminators
Chen, Lui, Nielsen *et al*. Nature Methods 2013

Zinc Finger Transcription Factors for yeast
Khalil, Lu, *et al*. Cell. 2012