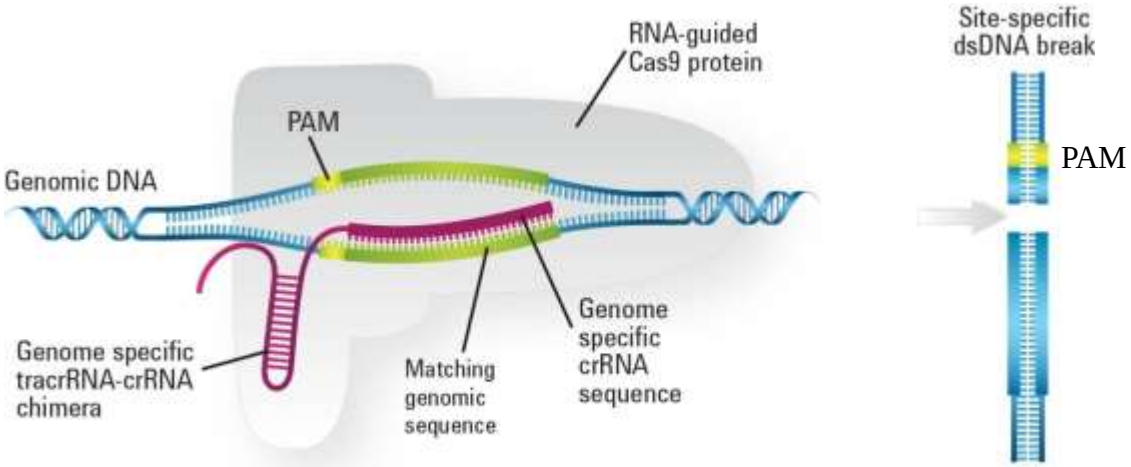
Pooled CRISPR guide RNA libraries for functional genomics screening: Do you know what's in your library?

Peter Sheffield

R&D Scientific Program Director

Agilent Technologies

INTRODUCTION
CRISPR: A Programmable Tool for Genome Engineering



- CRISPR = Clustered Regularly Interspaced Short Palindromic Repeats
- A multi-component system consisting of an RNA component and a CRISPR-associated protein (Cas)
    - RNA (CRISPR) sequence defines target location
    - CRISPR-associated protein (Cas) determines functional modification made at target location

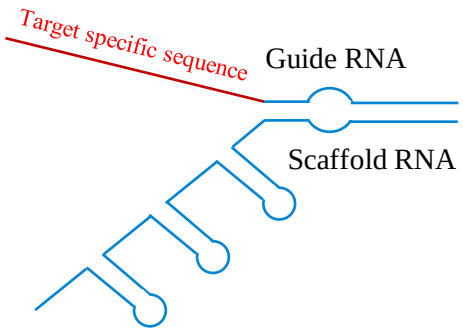
INTRODUCTION
CRISPR: A Programmable Tool for Genome Engineering

Part 1 - CRISPR-associated protein (Cas)

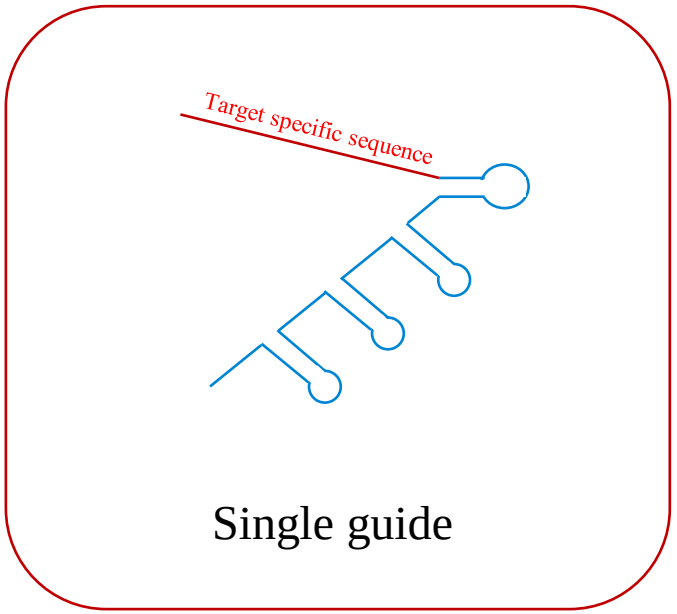
- CRISPR-associated protein (Cas) determines functional modification made at target location
  - wtCas9 - induces double stranded breaks - gene function knock-out
  - dCas9 - 'Dead' becomes a DNA targeting protein – CRISPRa/i
  - nCas9 - 'Nicking' induces single stranded nicks at target location, either top or bottom strand depending on Mutation

Part 2 - RNA (CRISPR) sequence defines target location

- Two RNAs
- Single guide RNA



Two guides



Single guide

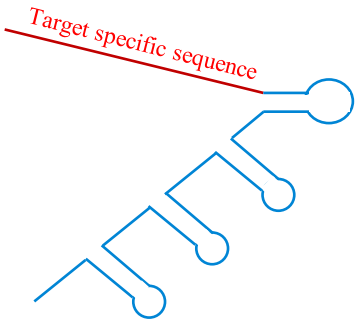
Single Guide CRISPR RNAs

Can any sequence be used as a CRISPR?

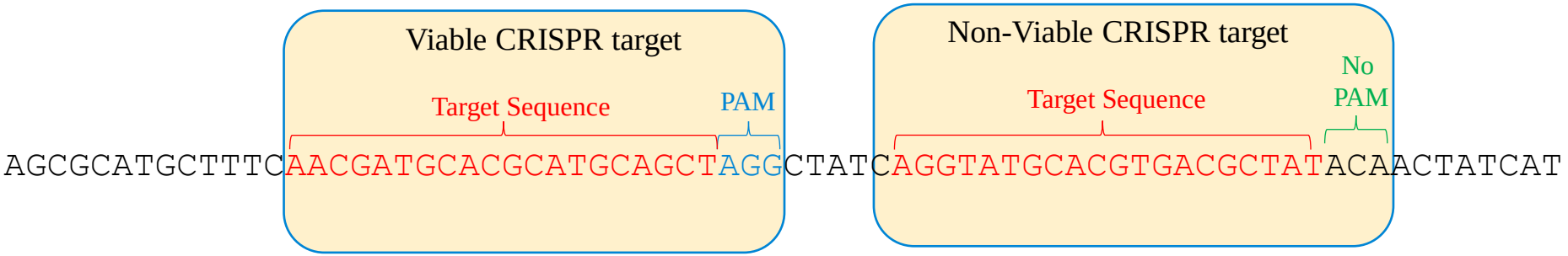
Yes - as long as it is immediately upstream of a PAM site!

What is a PAM site?

PAM = Protospacer Adjacent Motif is a 2-6 base pair DNA sequence immediately adjacent to the DNA sequence targeted by the Cas9 protein, for *S. pyogenes* PAM = NGG



Single guide



Principles of Designing a Genome Wide CRISPR library
It's a numbers game!

Human Genome = 3 x 10^9 bp

Cas9 PAM = NGG = once every ~40 bp = ~150 Million PAM si… strands)

Too many to Screen!

How do you filter the CRISPRs?

- Target certain areas of the genome
  - Exonic regions
  - Intronic regions
  - Regulatory regions
  - Gene clusters
  - Gene Families

Good CRISPR versus Bad CRISPR

- Not all CRISPRs are equal
  - Algorithms to score
    - Doench[1, 2]
    - Zhang[3] (Broad)

1. "Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation" John G Doench et al. Nature Biotechnology 32, 1262–1267 (2014)
2. "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9" John G Doench et al. Nature Biotechnology 34, 184–191 (2016)
3. "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells" Ophir Shalem, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei Mikkelson, Dirk Heck, Benjamin L. Ebert, David E. Root, John G. Doench, and Feng Zhang. Science. 2014 Jan 3; 343(6166): 84–87.

Example: GeCKO
Genome-scale CRISPR-Cas9 Knock-Out

Designed by The Broad Institute (Zhang)

Human Genome – 150 Million CRISPRs

Criteria	Targets	GeCKO Library Design				
		n	CRISPRs	A	n	B
Coding Regions	~19,050	6	~114,300	~62,000	3	~53,000
miRNA	~1,860	4	~7,200	~3,600	3	~3,600
Non-targeting Controls	1,000	1	1,000	1,000	1	1000
Total	**~21,800**	**~123,000**		**~66,000**		**~57,000**

Pooled CRISPR Guide Libraries

Genome Wide Modification Experiments

Using a pooled CRISPR library, containing many thousands of CRISPR RNAs, multiple CRISPR/Cas reactions can be performed simultaneously on a genome wide scale.

Gene Knock-Out:

- Introduces non-sense mutations into the reading frame of any gene in the genome via Non-homologous end joining (NHEJ), a repair pathway that joins two broken ends together.
- NHEJ does not require a homologous template for repair and typically leads to the introduction of small insertions/deletions at the break site often creating frame-shifts that knockout gene function.

CRISPRa/i:

- a = gene Activation
- i = gene Inhibition

- Targets CRISPR sites within the regulatory regions of genes and using dCas9 fusions to recruit transcriptional modification machinery that either activates or inhibits gene expression

All three methods especially useful for identifying biological relevant pathways, new drug targets for disease models

Designing a Genome Wide CRISPR library: Factors to consider

CRISPR Experiment

- Gene Knock-Out
  - Preferentially target the 5’ exons over the 3’ exons.
    - Increases chances of creating a Null mutant.
  - Target just the genes you are interested in knocking out
- CRISPRa/i
  - Target the regulatory regions upstream of coding regions.
  - Target only the gene families of interested

Determining *n* CRISPR/target

- Increasing the number of CRISPRs/target builds redundancy into the design which increase the likelihood of having an affect on your target.
- Increasing the number of CRISPRs/target can also help in the analysis of the Next Gen Sequencing data by adding mathematical confidence that the changes seen in the library distribution pre- and post- data sets are statistically relevant.

Designing a Genome Wide CRISPR library: Factors to consider

Minimize bias in Design

- No CRISPR duplicates
- Ideally all CRISPR sequences in design should be unique (Ratio 1=1)
- The presence of pseudogenes and splice variants can create instances where some CRISPR sequences are duplicated. These duplications introduce bias into a design (Ratio 1≠1)
  - Either COMPLETELY remove all CRISPRs that are Duplicated and replace with a different CRISPR that is not duplicated

    or
  - Eliminate all duplicates but leave one copy in design
- Both methods maintain CRISPR ratio of 1=1.
- The first method could result in the replacement of a good CRISPR with a less desirably one.
- However, the second method means you have CRISPR sequences that target multiple loci in the genome. This could lead to analysis complications downstream of your experimental data.

How do you get all the CRISPR machinery into the target cell?

For a CRISPR/Cas event to occur you need the following to be present in the cell at the same time

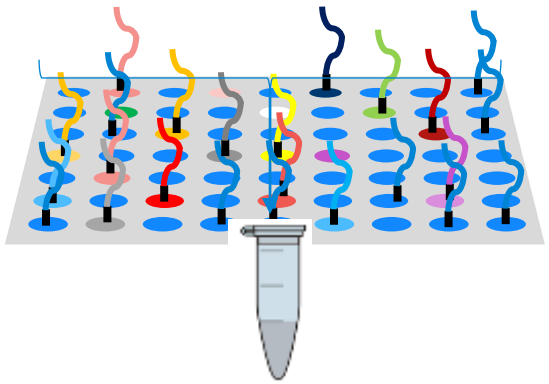
- CRISPR guide RNA
- Cas9 protein

Delivery Methods

- Pooled CRISPR libraries are best delivered via a DNA vector delivery method, e.g. Lentivirus
  - Lentiviral vectors are amenable to molecular biology manipulation
  - Lentiviral particles transfect many different cell lines, non-dividing cell
  - Stably integrate into the genome
  - Perfect for expression of small sgRNA's
- Cas9 protein can be delivered via multiple methods
  - Direct transfection of active protein
  - Via a DNA expression vector (viral or non-viral)
  - Direct transfection of mRNA

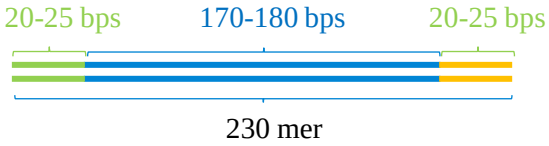
- Delivery of the CRISPR sgRNAs and the Cas9 protein can be done simultaneously or sequentially
- Cas9 modifications can be made in parallel on a genome-wide scale

Synthesizing a CRISPR Oligo library:



Oligo libraries:

- Agilent can print up to 240,000 user-definable sequences/chip
- Sequences up to 230 nt
- Pooled into a single tube/library



- Libraries are synthesized as ssDNA
- Libraries usually need to be amplified by PCR
- PCR acts as an enrichment step by amplifying the full length oligos
- CRISPR libraries are an ideal size for Oligo synthesis (<140 nt)

Library Construction Considerations

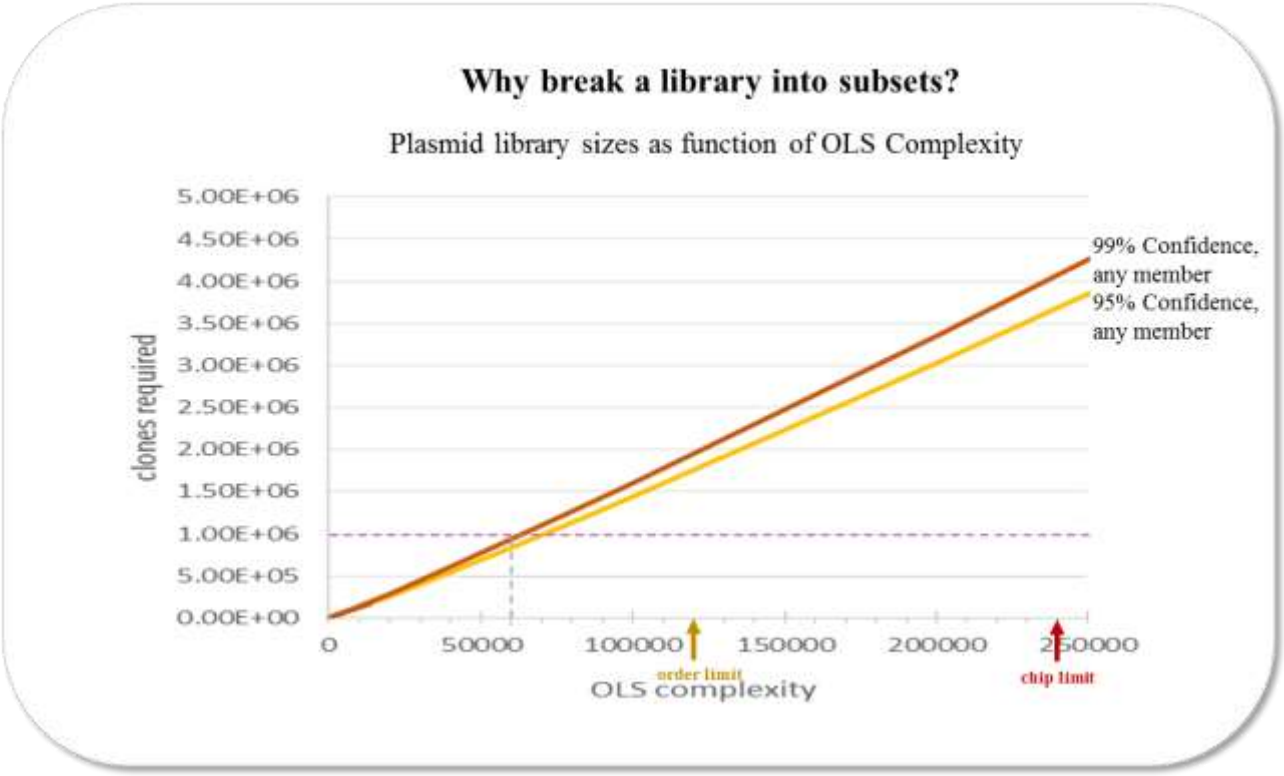
Example case: GeCKo library for genome-wide Cas9 mediated gene knockout

CRISPR library

GeCKo v2 libraries

species	human	mouse
genes targeted	19,050	20,611
targeting constructs/gene	6	6
miRNA targeted	1,864	1,175
targeting constructs/miRNA	4	4
control (nontargeting) sgRNAs	1,000	1,000
total sgRNA constructs	123,411	130,209
set A	66,172	
set B	57,239	
non-redundant set A	64,580	
non-redundant set B	56,869	
total non-redundant	121,449	

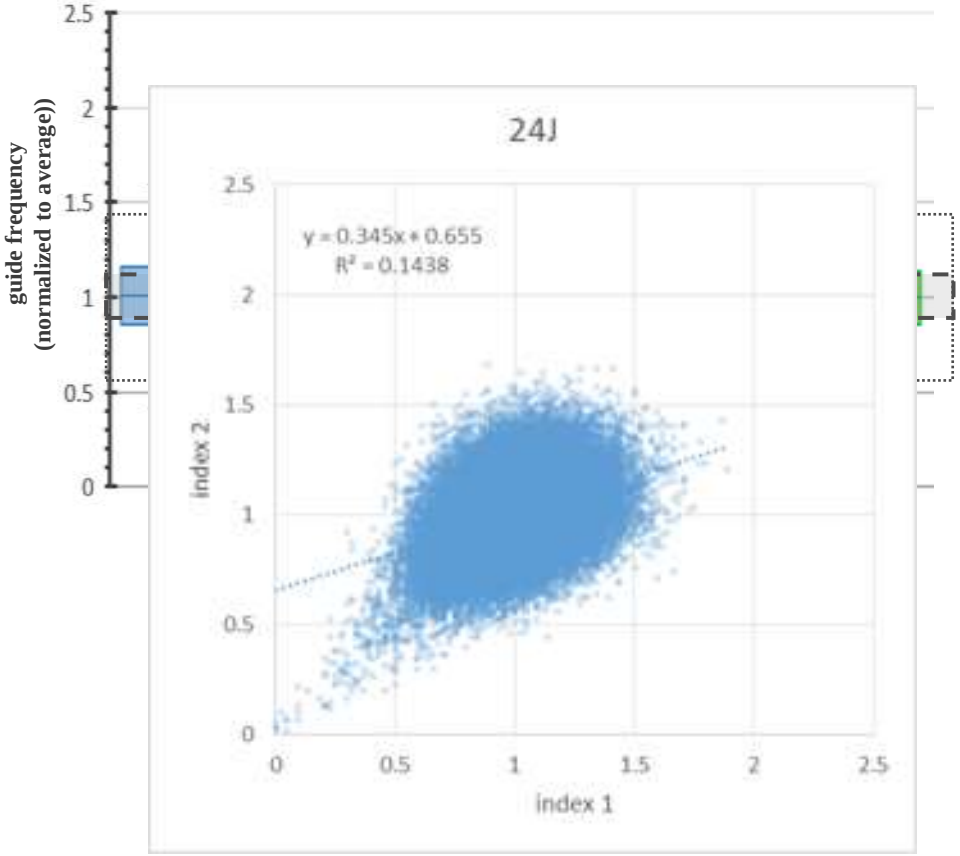
Sanjana et al., *Nature Methods* **11**: 783-784 (2014)

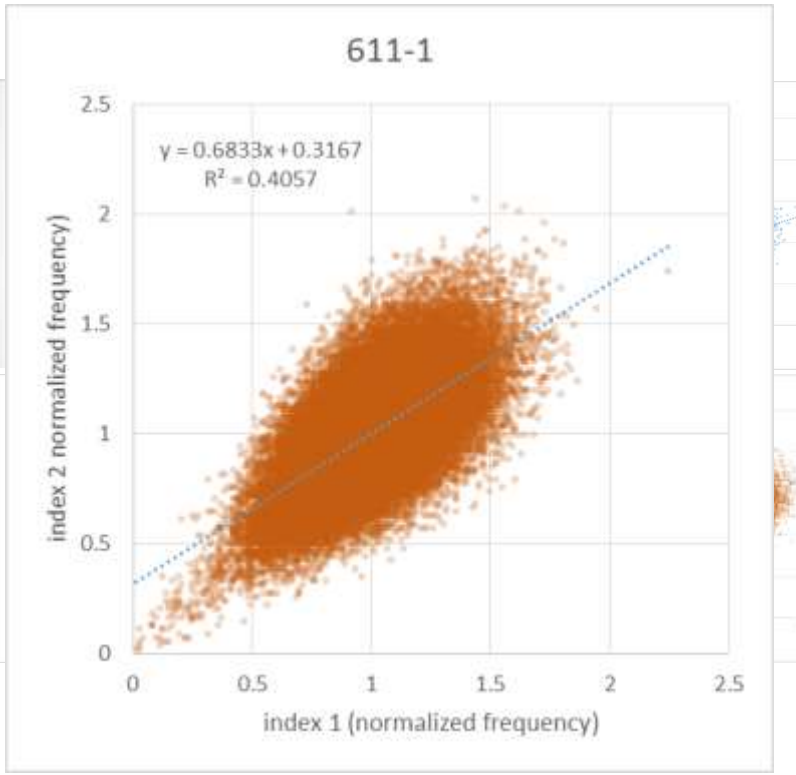


Quality of Oligo libraries: representation of library members

Representation of guides analyzed using NGS

box whisker plots of guide representation in oligo libraries





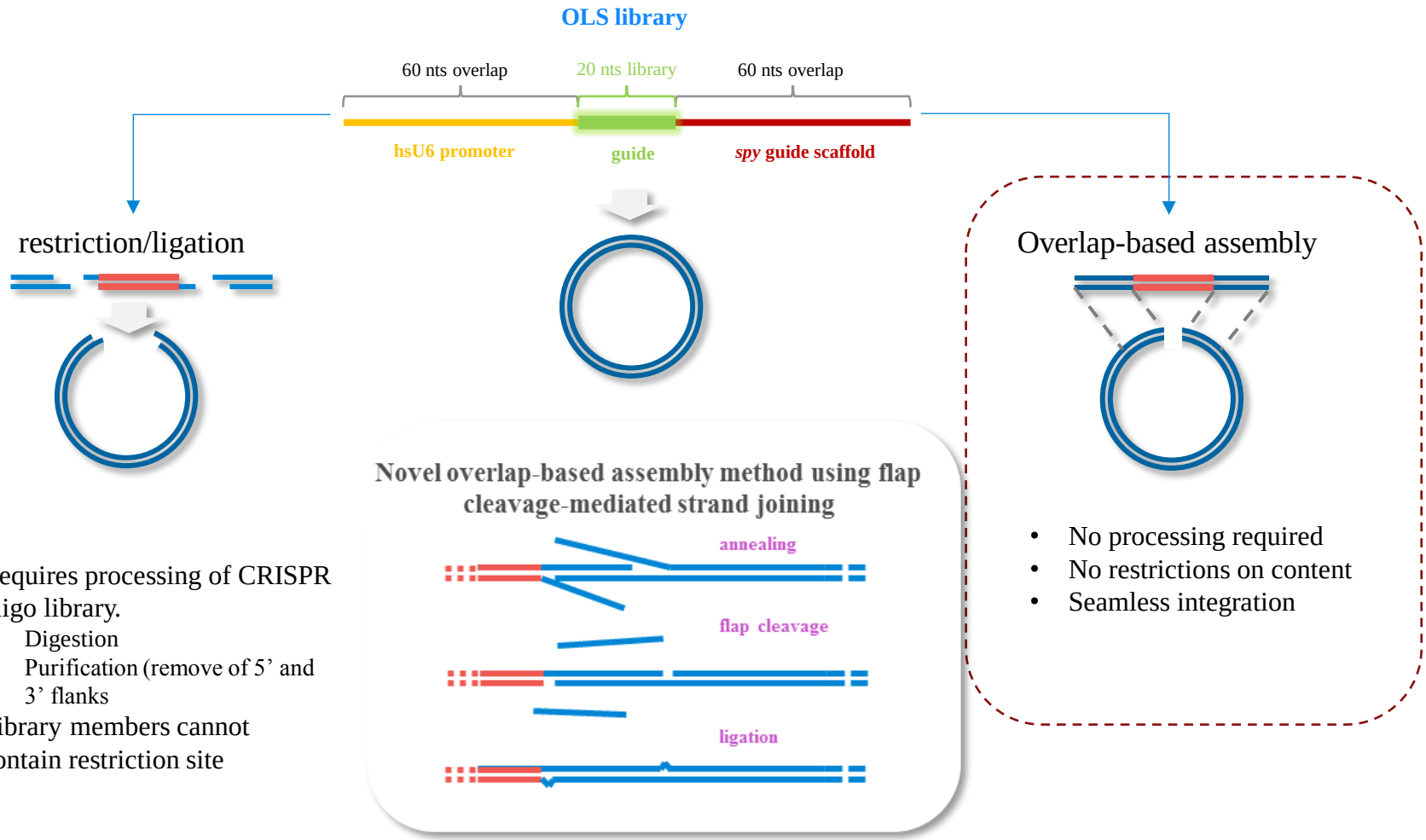
Synthesis Protocol 1 Synthesis Protocol 2

of the same CRISPR design

Library Construction Considerations

- Each step in the process can introduce bias into a library
- Each method needs to be as *neutral* as possible
- Each step needs to be controlled and monitored
- The Cloning and Library Expansion steps have the greatest potential to introduce bias into a library

Cloning an CRISPR oligo library into a plasmid vector



- Requires processing of CRISPR oligo library.
  - Digestion
  - Purification (remove of 5' and 3' flanks
- Library members cannot contain restriction site

- No processing required
- No restrictions on content
- Seamless integration

SureVector Cloning

Constructing plasmid libraries from amplified OLS libraries

Agilent's SureVector Sure Guide Cloning Kit and workflow

Assembly of plasmid library → **Transformation of bacterial host** → **Selection and expansion** → **Analysis by sequencing**

4 x 20 µl reactions:
75 ng vector
6.5 ng OLS library
4.5-fold molar excess of OLS library

95°C, 1 minute
95°C, 20 sec
60°C, 90 sec
65°C, 60 sec
8 cycles

23 minutes programmed
25-30 minutes real time

Reaction Clean up

Electroporation of Electro Ten-blue cells

Select in 3D soft agar matrix

Limited growth (2-3 days) at 30°C
Followed by 4-6 hours growth in liquid media prior to plasmid isolation

NGS sequencing:
- **distribution**

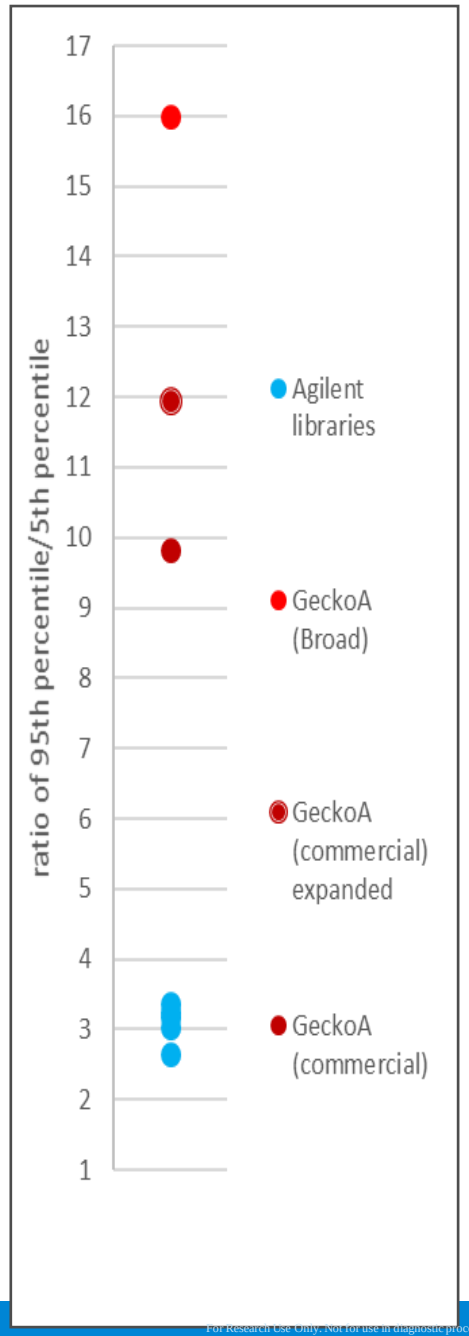
45 minutes,
15 minutes hands on

90 minutes
45 minutes hands on

2 days

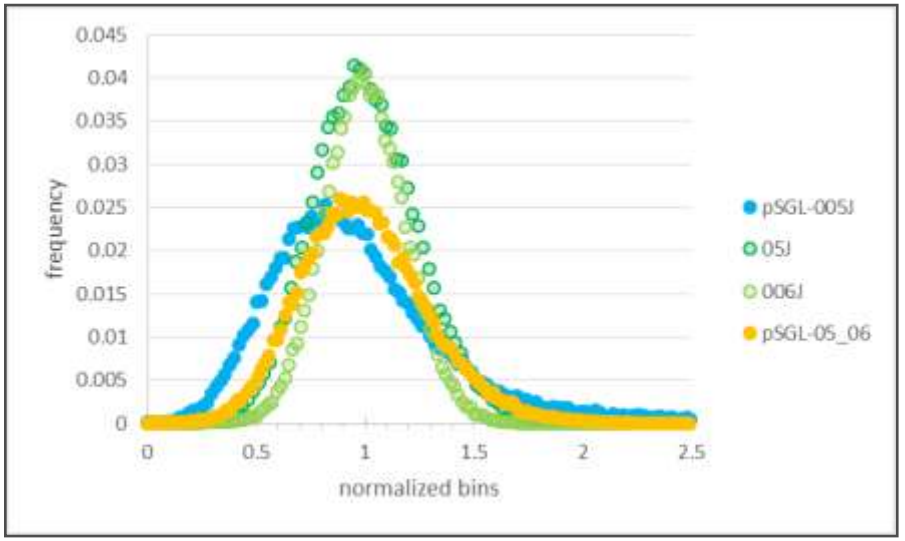
$$\text{Ratio } 95^{th}/5^{th} \text{ Percentile} = \frac{\text{Frequency of } 95^{th} \text{ Percentile}}{\text{Frequency of } 5^{th} \text{ Percentile}}$$

The lower the $95^{th}/5^{th}$ Percentile ratio the better the library



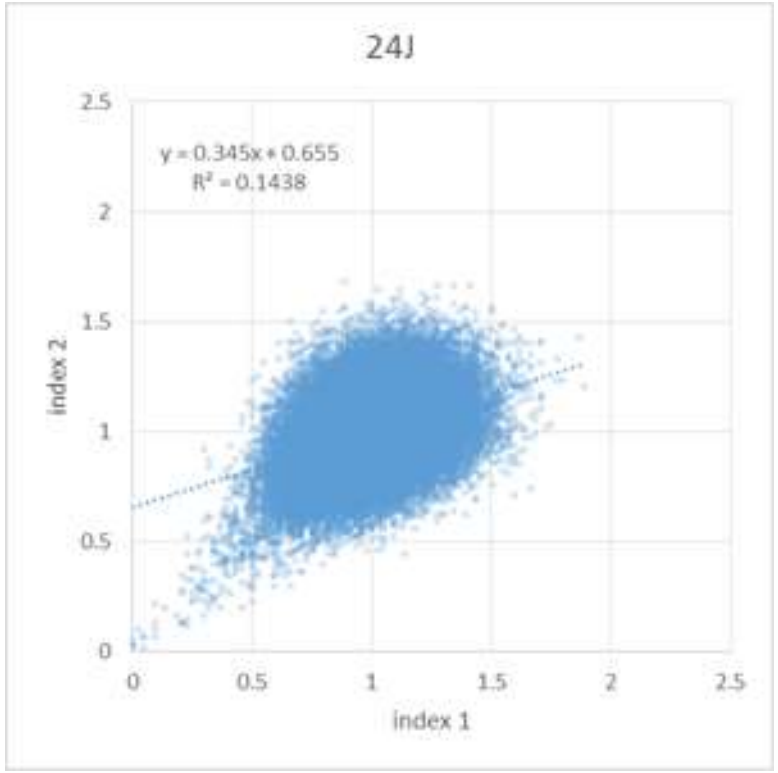
Methods to Generate highest quality plasmid libraries from OLS designs

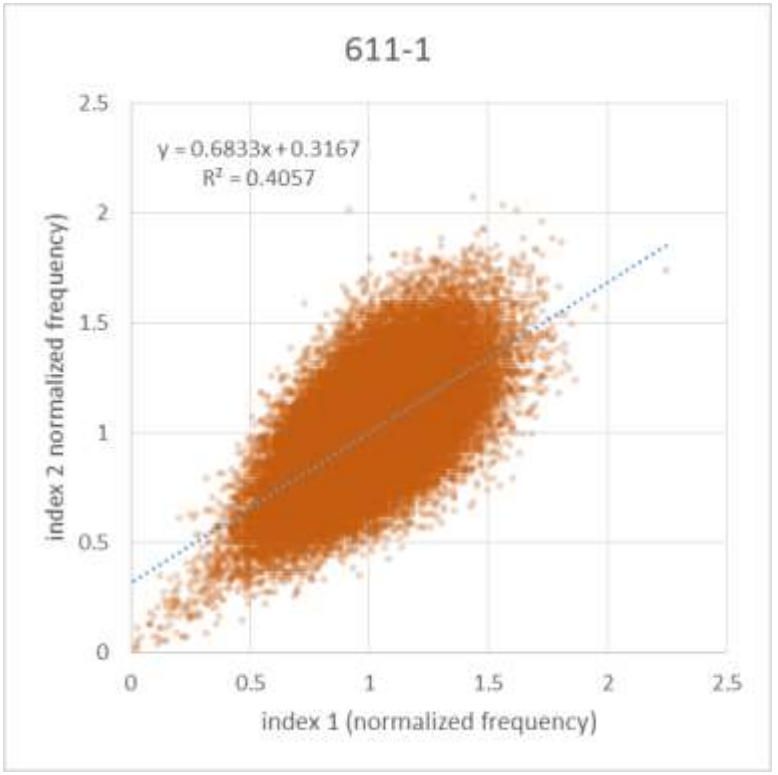
- Start with highest quality OLS libraries
- combine (+) strand and (-) strand designs
- control growth conditions



		pSGL-05J	pSGL-05/06J
strand		P1	Combined
90% range		4.632	2.800
80% range		3.000	2.200
60% range		2.069	1.659
reads/guide		47.6	55.4

Quality of Oligo libraries: representation of library members



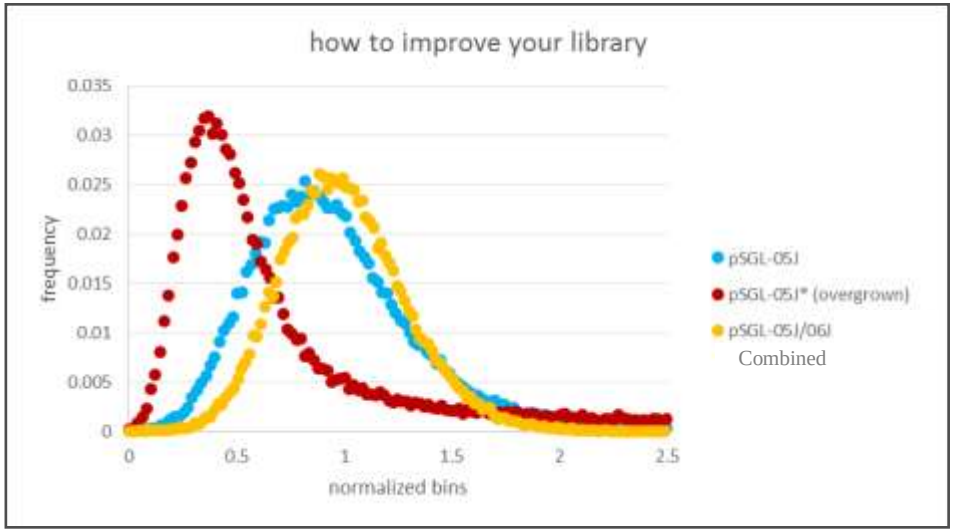


Synthesis Protocol 1 Synthesis Protocol 2

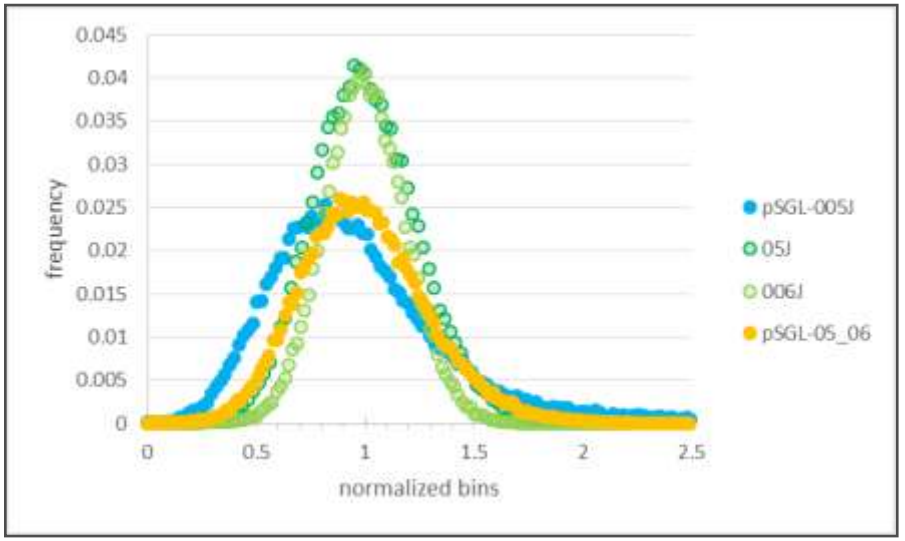
of the same CRISPR design

Methods to Generate highest quality plasmid libraries from OLS designs

- Start with highest quality OLS libraries
- combine strands generated by different synthesis protocols (P1 and P2)
- control growth conditions



Uncontrolled growth of bacterial cultures containing CRISPR libraries introduces bias into the library

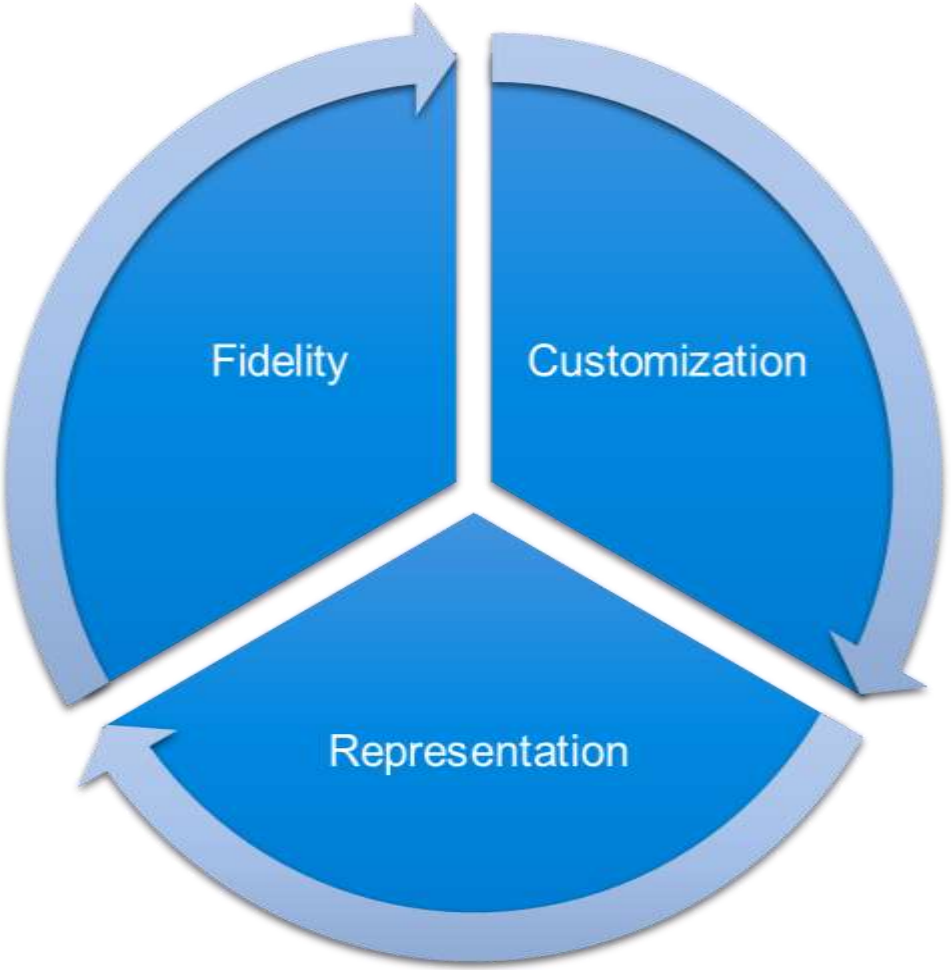


	pSGL-05J	pSGL-05/06J
strand	P1	Combined
90% range	4.632	2.800
80% range	3.000	2.200
60% range	2.069	1.659
reads/guide	47.6	55.4

So what makes a Good CRISPR Library

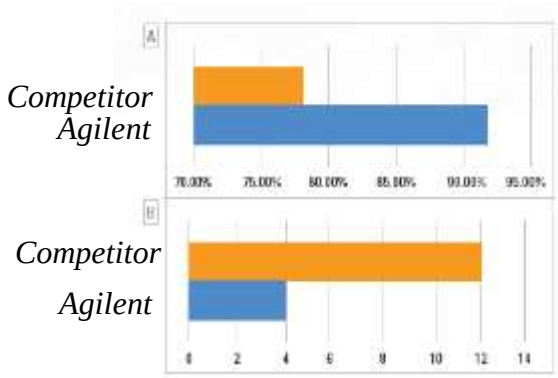
Three critical aspects that make a good pooled CRISPR Library

- Fidelity
- Customization
- Representation



Fidelity:
Are my CRISPR Sequences correct?

- **A high fidelity oligo library**



CRISPR Libraries

- **More correct constructs**
- 92% vs. 77%

- **Fewer avg. errors/kb**
- 4 vs. 12

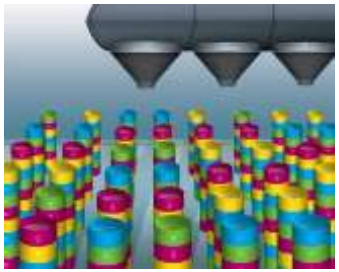
- **Fewer errors** per kbase pair improves library quality
- **Reduces screening time** and false negatives

Customization:
Any CRISPR Sequence you need

Easy Customization

End-user has complete control

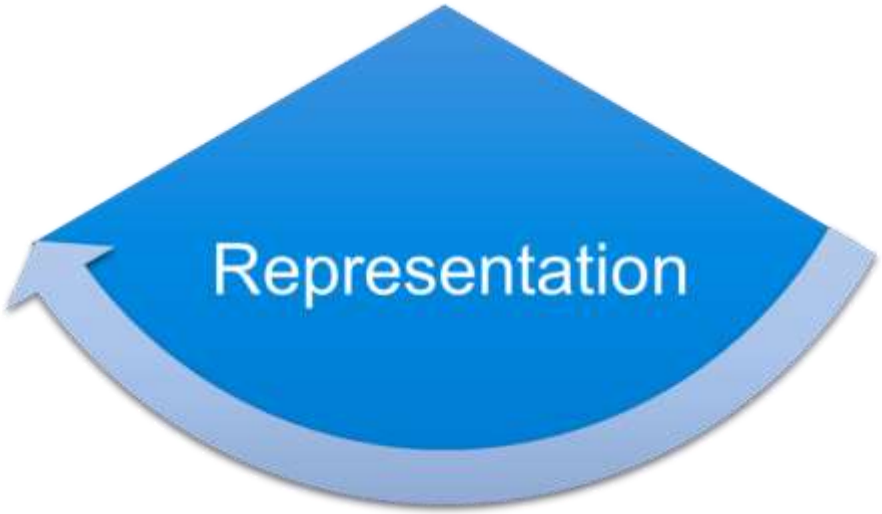
- Any sequence
- Any species
- Any application



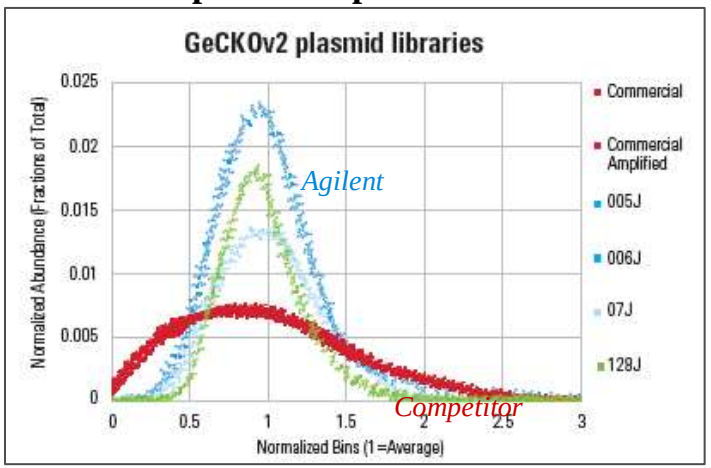
Agilent provides CRISPR Libraries

- Ready-to-clone
- Ready-to-amplify libraries for total flexibility

Representation:
Are all my guides present?



Improved Representation



- SureVector CRISPR Cloning Kit
- Fewer missed guides
- Less under-represented population

	Libraries	Missed Guides	90th/10th percentile	95th/5th percentile	99.5th/0.5th percentile
Agilent	pSGL-007J	1	2.32	3.04	6.72
	pSGL-006J	1	2.47	3.38	11.06
	pSGL-005J	1	2.38	3.19	9.83
Competitors	pSGL-128J-dc	1	1.99	2.64	8.30
	GeCKO (Broad)	?	8.73	16.00	NA
	GeCKO (Competitor)	39	5.29	9.83	68.40
	GeCKO (Competitor) expanded	204	6.00	11.95	333.00

- Plasmid libraries with 90/10 ratios < 3
- Find all the hits in your screen, not just the over-represented ones

Now You have a high Quality CRISPR Library, What Now?

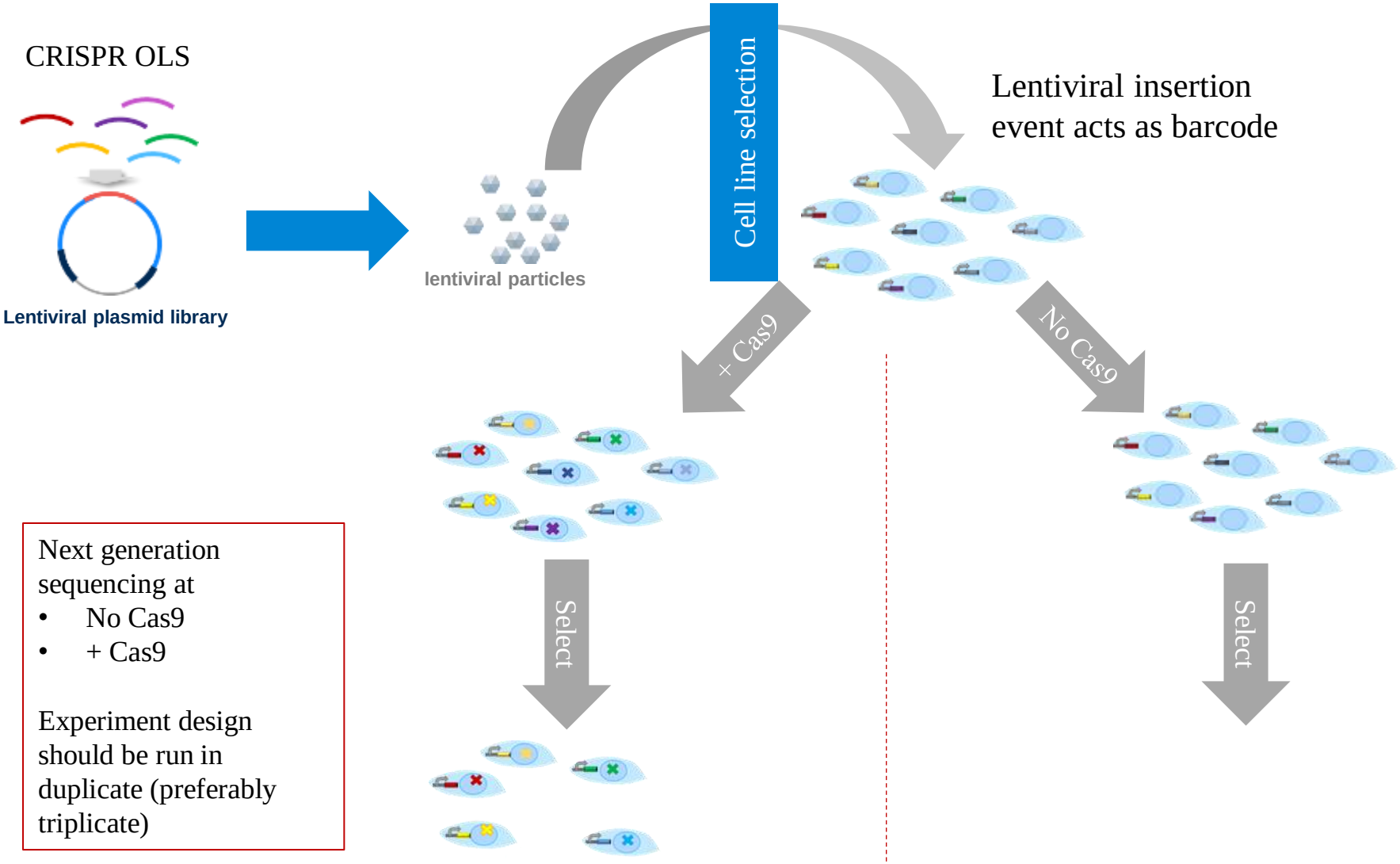
Genome Wide Modification Experiments

Using a pooled CRISPR library, containing many thousands of CRISPR RNAs, multiple CRISPR/Cas reactions can be performed simultaneously on a genome wide scale.

- Gene Knock-Out
- CRISPRa/i

- Experimental Design Decisions
  - Determine Cell line
  - Number of Replicates
  - Selection protocol
  - Next Gen Sequencing protocol
  - Data Analysis pipeline

Functional screening using CRISPR/CAS



Next generation sequencing at
- No Cas9
- + Cas9

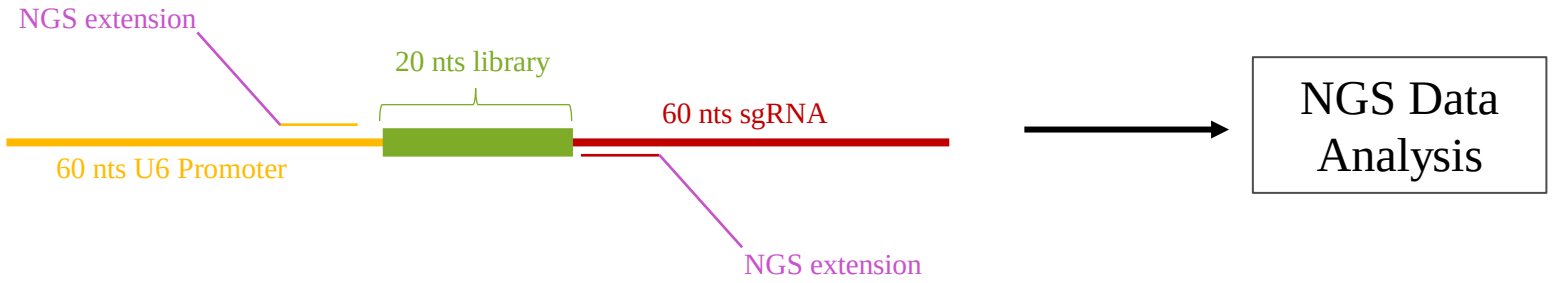
Experiment design should be run in duplicate (preferably triplicate)

NGS of CRISPR Oligo libraries, plasmid libraries and transduced cell lines

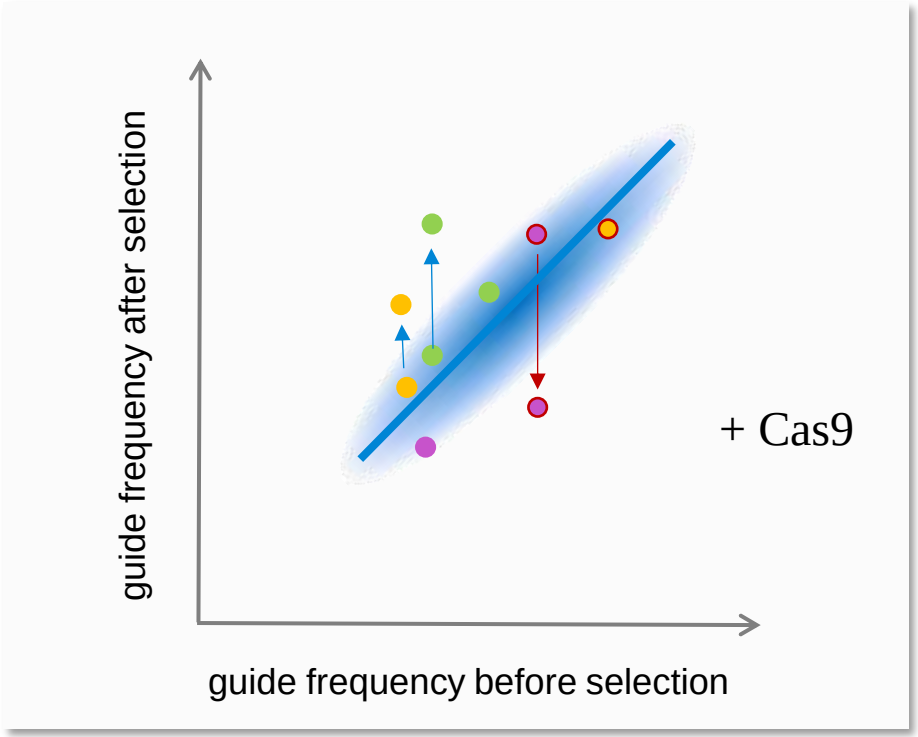
This section is present in all CRISPR library entities Oligo library, Plasmid and transduced cell line and can be amplified using universal primers

This region of the library can be isolated via PCR

The resultant amplicon can then be used in your preferred NGS workflow.



Analysis of Next Gen Sequence Data from a CRISPR/CAS Experiment



Statistical analysis packages

- RIGER
- HiT Select
  Diaz et al., *NAR* (2014)
- MAGeCK
  Li et al., *Genome Biology* **15**:554 (2014)
- casTLE
  Morges et al., *Nature Biotech* ePub April 11th (2016)

No Cas9 the library should not show too much bias towards any one CRISPR

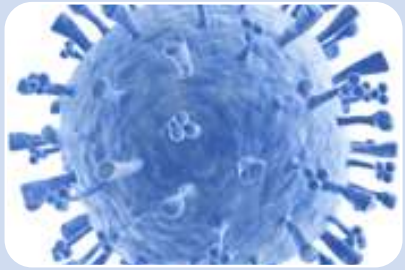
All libraries have an inherent amount of background variation (noise / bias)

With Cas9 the library should start to show bias towards certain CRISPRs due to selection

However, a bad library will have more noise/bias which makes it harder to identify the significant CRISPRS. This requires your screen to encompass a greater number of cells

Agilent SureGuide Products

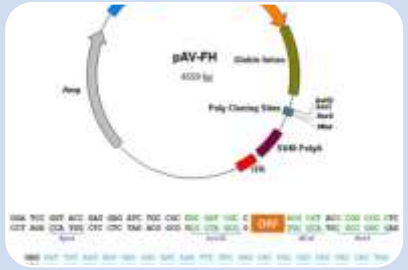
Ready-to-Package



Catalog gRNA Libraries

- Plasmid Library
- GeCKOv2
- Human and Mouse
- Cloned into lentivirus vector with hU6 promoter

Ready-to-Clone



Custom Human & Mouse gRNA Libraries

- Pre-amplified oligo library
- User defined subset or designed
- Mammalian systems
- Compatible with SureVector cloning

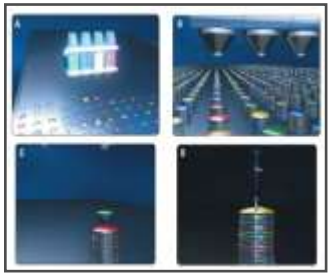
Ready-to-Amplify

Custom User Defined Libraries

- Unamplified oligo pool
- Any species, any cloning method
- Entirely custom by user design

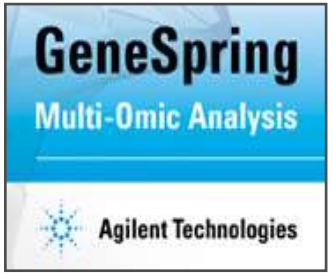
Agilent's Commitment to Genome Engineering

- Agilent has expertise in manufacturing market-leading high fidelity, complex, ultra-long **oligonucleotide libraries,** making market leading microarrays and SureSelect NGS solutions.



- Agilent offers a portfolio of **specialized cloning Kits (SureVector), competent cells, and high fidelity PCR enzymes** as well as market-leading **QuikChange site directed mutagenesis** technology – developed with expertise from Stratagene (now part of Agilent).

- **Broad portfolio of analytical separation and measurement tools** (e.g., LC-/GC-MS, NMR), **liquid handling automation and informatics solutions**.



Leading Provider of Measurement Tools & Solutions






Learn more: www.agilent.com/genomics/SureGuide

Acknowledgements:

La Jolla R&D
Peter Sheffield
Carsten Carstens
Katie Felts
Sarah Johns
Vivian Zhang

Santa Clara Marketing
Ben Borgo
Caroline Tsou