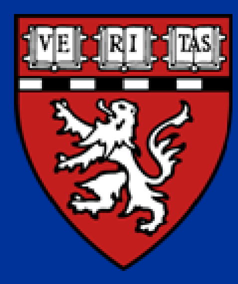
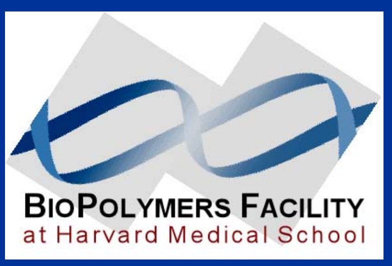




Next Generation Sequencing
Illumina GA Workflow
Sample Prep Details

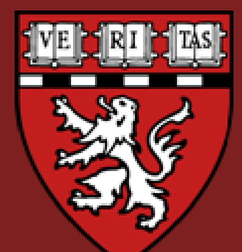
Robert Steen

Biopolymers Facility
Department of Genetics
Harvard Medical School

ABRF 2010

"I am sorry if this smacks of an investigator, hair on fire, running into your lab needing data yesterday for a grant submission…."

Jay Fox



Biopolymers Core Facility

Our Customers:

(over past 12 months)

5 Harvard Institutions
34 Non-Harvard Academic Institutions
6 Private Companies

* 438 Laboratories
* 1,611 Users

Our Platforms and Services:

* Affymetrix
* Agilent Bioanalyzer 2100
* Applied Biosystems 3730xl
* Applied Biosystems 7900HT
* Applied Biosystems SOLiD V3
* Axon GenePix 4000B
* Hamilton StarPlus Robot
* Two - Illumina Genome Analyzer II
* Luminex LS200
* NanoDrop ND-8000
* QIAcube Robot
* Tecan Genios

* DNA / RNA Analysis
* DNA / RNA Preps & Purification
* DNA Sequencing
* Gene Expression
* Micro RNA Analysis
* Next-Generation DNA Sequencing
    * ChIP-Seq / RIP-Seq
    * mRNA-Seq
    * small RNA-Seq
    * Exome Sequencing
    * Whole Genome Sequencing
* Protein Interaction Assays
* Quantitative PCR
* SNP / Microsatellite Genotyping

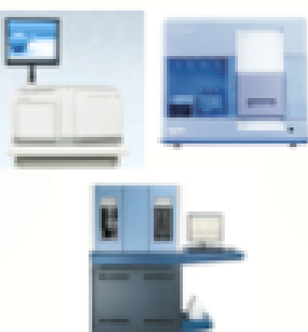
Applications

Next Gen-Omics

Genome Resequencing

mRNA Tag Profiling

Methylation Analysis



Small RNA Identification

Functional Elements (ChIP-Seq, DNAse-Seq)

Transcriptome Sequencing

Slide Source: Nature Web

Applications In Our Facility

92	Flow Cells (~ 12 months)
736	Lanes
373	Samples

Lanes Run by Application

97	ChIP-Seq
68	Digital Gene Expression
278	Genomic (Single Read And Paired End of various types)
123	PhiX Genomic Control
24	Small RNA
114	Whole Transcriptome

NGS Workflow Summary

- **LIMS & Website**
- Information Presentation
- Sample Ordering
- Sample Tracking
- Data Management
- Invoicing

- **Sample Processing / Library Prep**
- Primary Sample QC
- Library Sample Prep Protocols
- Library QC

- **Sequencing Run**
- Pre-Cluster QC
- Cluster Generation
- Sequencing
- On Instrument QC

- **Data Analysis & Distribution**
- Primary SCS 2.6 & RTA
- Pipeline OLB V1.6
- Internal Transfer, FTP, FEDEX HD

Laboratory Information Management System (LIMS) and Your Website

Do Your Customers Understand the Technology?

- New Customer or Someone New to NGS
  - Inquiry Form
  - Consultation -- This is often VERY important!

- Existing Customer or Knowledgeable User
  - Straight to Sample Submission

More Information is Better

Biopolymers Facility
NRB0088
617-432-7481
nextgen@genome.med.harvard.edu
BPF Quote #

PO/Billing Code:

Samples will not be put into queue until billing information is received.

Sample Owner ______________________
PI Name/Lab ______________________
Organization ______________________
Phone ______________________
email ______________________
BPF User ID ______________________

If submitting a prepared library, please indicate:

Preparation Details

Sample Assignment (for BPF use)		Sample name (as labeled)	Organism	Is your sample (check one): Prepared Library	RNA/DNA for lib prep	Library Type*	Indexed/ Barcoded Library (Y/N)	Vol. (ul)	Conc. if known (ng/ul)	Adaptors† Single Read	Paired End	Library size (bp)
(example)	EX	Wild Type 14	*Mus musculus*	[X]	[ ]	*Genomic*	N	20	~50	[X]	[ ]	200
	1			[ ]	[ ]					[ ]	[ ]	
	2			[ ]	[ ]					[ ]	[ ]	
	3			[ ]	[ ]					[ ]	[ ]	
	4			[ ]	[ ]					[ ]	[ ]	
	5			[ ]	[ ]					[ ]	[ ]	
	6			[ ]	[ ]					[ ]	[ ]	
	7			[ ]	[ ]					[ ]	[ ]	

Run Details

	Run/Flowcell Type (check one): Single Read	Paired End	# of lanes	# of cycles (36/72)	Primer Name Genomic or specify if custom
EX	[X]	[ ]	1	36	genomic
1	[ ]	[ ]			
2	[ ]	[ ]			
3	[ ]	[ ]			
4	[ ]	[ ]			
5	[ ]	[ ]			
6	[ ]	[ ]			
7	[ ]	[ ]			

Additional Comments:

Please note any expected sequence bias.

*Library Types: smRNA, Genomic, mRNA-seq, ChIP-seq Digital Gene Expression (NlaIII or DpnII)

† Paired End adapters are compatible with Single Read and Paired End flowcells

Delivered by: ______________________ Date: ____________

(BPF) Order #
(BPF) File Info:

Received by (BPF Staff member): ______________________ Date (BPF): ____________ Freezer (BPF): ____________

Basic Sample Prep Outline

Isolate DNA / RNA (In our case customer always does this)

Fragment (Covaris, Bioruptor, Nebulize, Enzymes, Epicentre Transposase, etc.)

End Repair and Add “A” (Most cases)

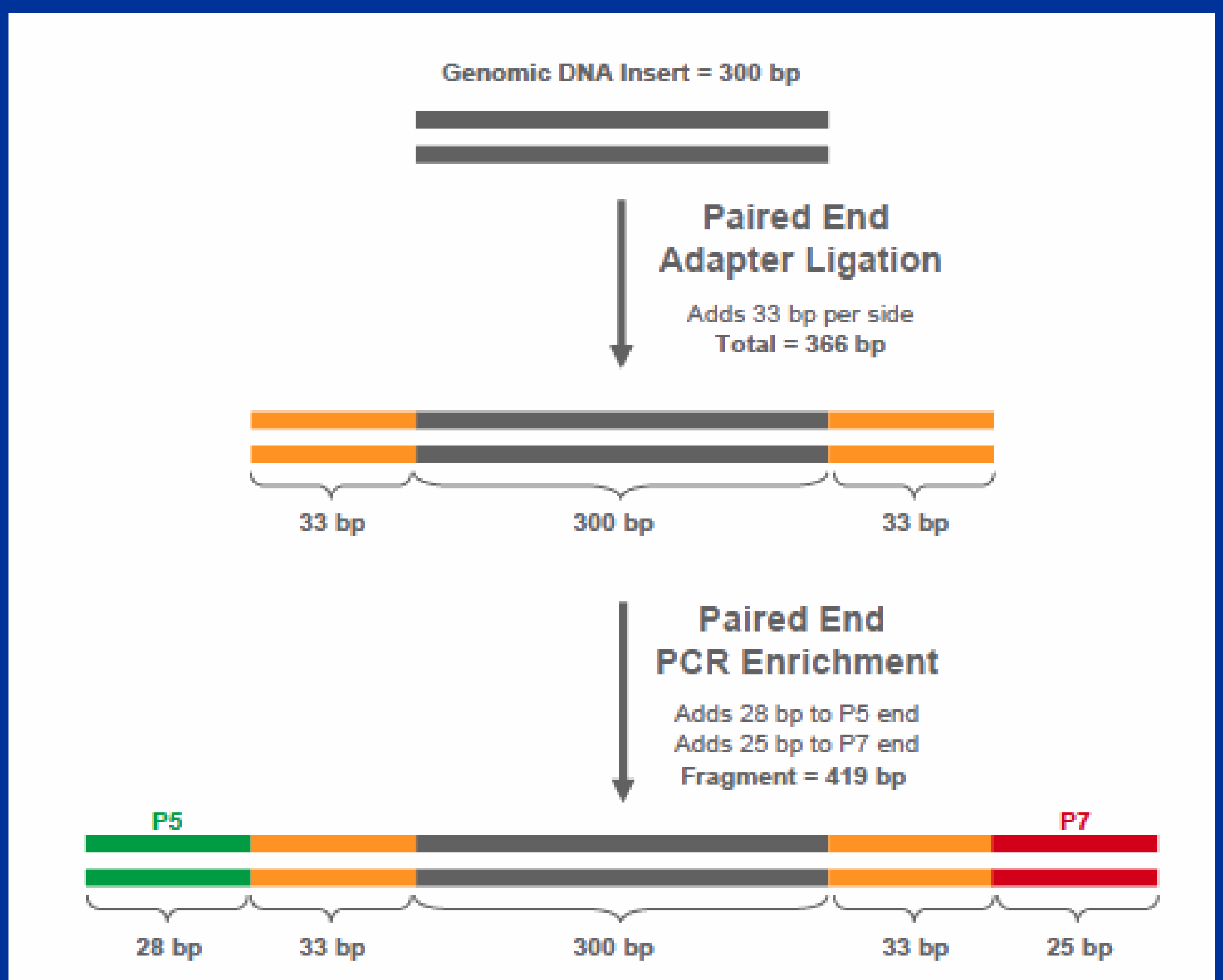
Ligate Adaptors (Most cases)

Size Select / Purify (Traditional Gels, Columns, E-Gel System, SPRI Beads)

Amplify (No PCR Libraries | NATURE METHODS | VOL.6 NO.4 | APRIL 2009)

Purify (Columns, E-Gel)

Basic Sample Prep Diagram



Slide Source: Broad / Illumina

Important Garbage In = Garbage Out

Start with Good Quality DNA / RNA
Sufficient Quantity is Important
Require QC on Bioanalyzer Before Starting

Bad RNA

[FU]
40
30
20
10
0
25 200 500 1000 2000 4000 [nt]

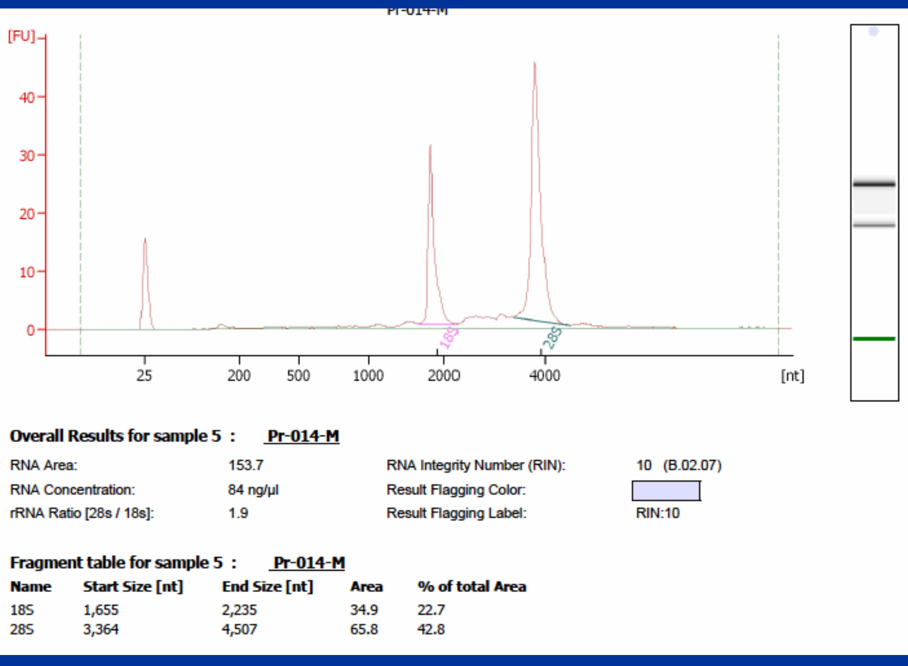
Overall Results for sample 6 : Shubha Control

RNA Area:	319.1	RNA Integrity Number (RIN):	3.8 (B.02.07, Anomaly Threshold(s) manually adapted)
RNA Concentration:	140 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN: 3.80

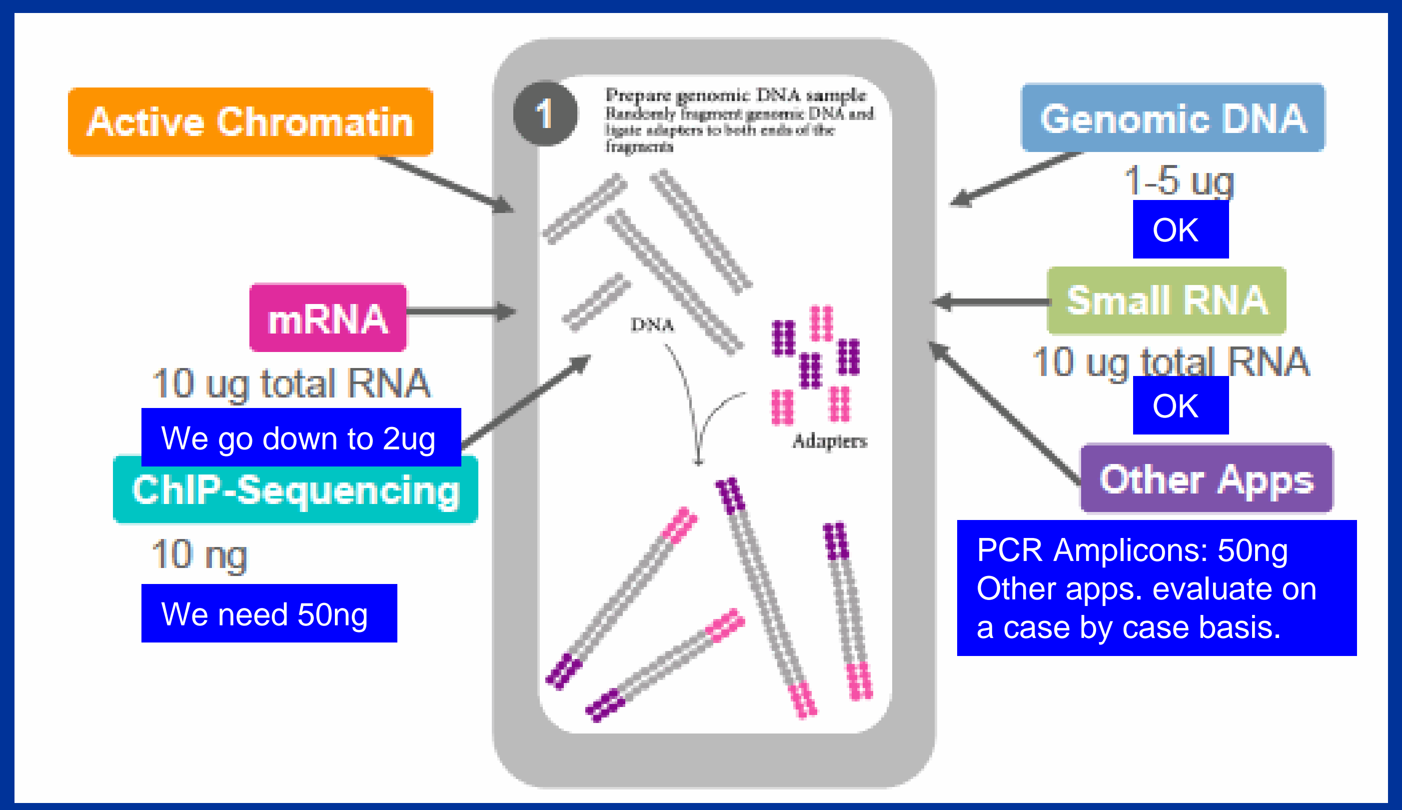
Fragment table for sample 6 : Shubha Control

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,659	1,814	17.3	5.4

Good RNA



Stick to Your Lab's Requirements



Slide Source: Broad / Illumina

Fragmenting

We use Covaris S2 with mini-tube holder.

Precise and reproducible shearing.

Know what you are working with; Bioanalyze all samples.

Isolated ChIP DNA

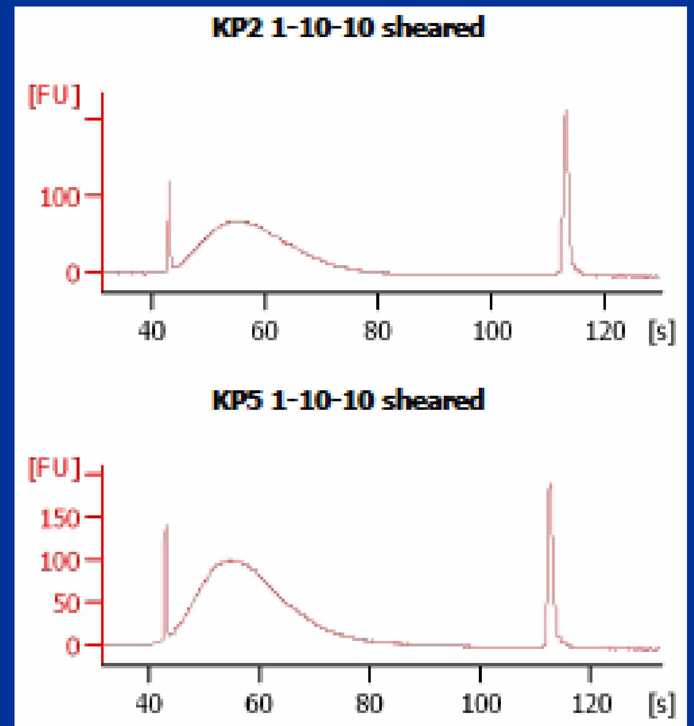
KP2 1-10-10

[FU]
150
100
50
0
40 60 80 100 120 [s]

KP5 1-10-10

[FU]
150
100
50
0
40 60 80 100 120 [s]

After Shearing



Note About Kits

Our facility currently uses:

-Illumina kit for RNA-Seq library preps.

-NEB kit for Genomic, ChIP and Similar Preps.

Our customers often create their own kits from list of enzymes.

Important: Adaptor sequences matter (obviously).

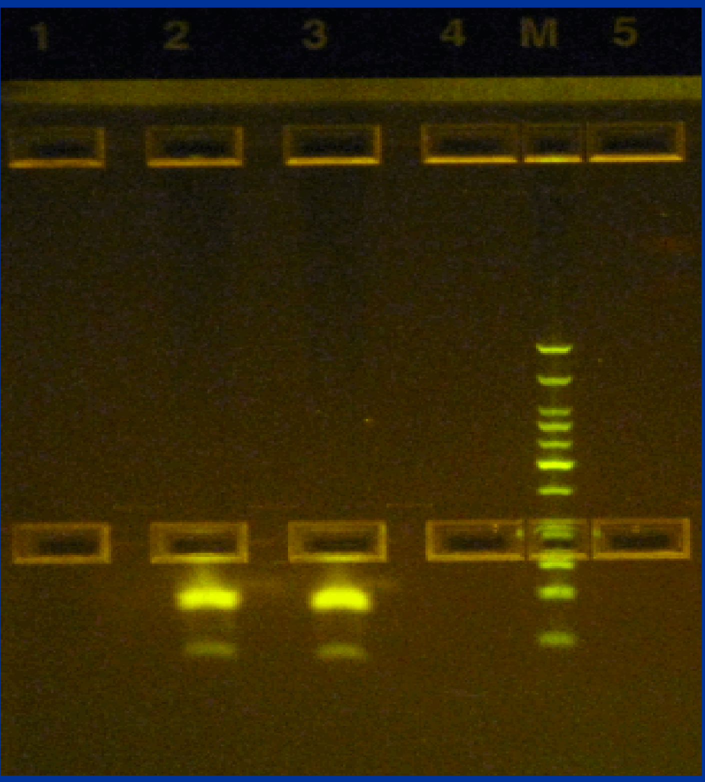
Our lab only makes paired-end libraries now and we ask
our users to do the same.

Very Important: Discuss barcoding / indexing.
-Potential for biasing reads (may or may not matter)
-Difficulty calling image data.
-Base balance your indexes or use a spike in such as PhiX.

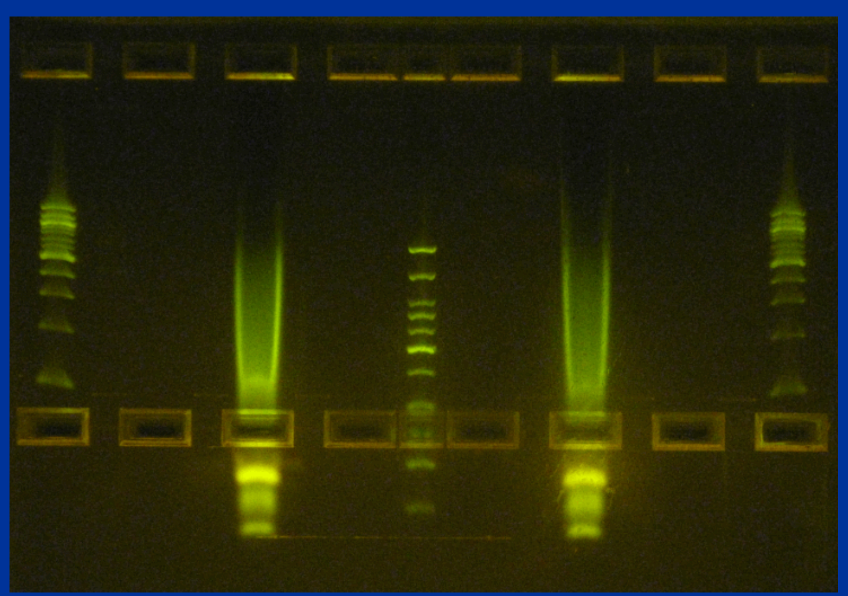
Size Selection

Invitrogen's E-Gel system with ibase.
-Very Fast ~15 mins.
-High recovery 50% to 80%.
-Very easy to collect multiple fractions (we collect 3-5).
-Also use in place of final column to clean up and verify.

ChIP DNA



Genomic DNA



Amplification

Important not to over-amplify.
Easy to introduce bias.

Currently using 10 to 18 cycles
Project / Sample dependent.

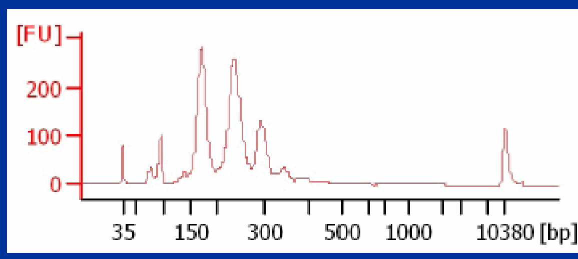
In place of QiaQuick PCR Purification Kit we use E-Gel.
Allows for efficient clean up, high recovery and library verification.

Would like to implement no-PCR libraries in coming year.

Library QC Part One

Agilent Bioanalyzer High Sensitivity DNA Chip.

-Library must be proper "size and shape".
-Must have a minimum concentration.



JL1

JL2

JL3

NS11.2 1:10

MZ30

MZ31

Library QC Part Two

SYBR QPCR Assay

-Assay based on original design from Stuart Levine at MIT.
-Use BioA concentration to make initial 10nM Stock.
-Make Std Curve with Illumina PhiX Library (new lots need characterizing).
-Dilute curves and stds. 1:1000.
-Use KAPA BIOSYSTEMS SYBR Fast qPCR Kit.
-Run on AB 7900HT in 384 well plate.
-Based on CT derived concentrations, adjust libraries for clustering.

```
SYBR_F-AATGATACGGCGACCACCGA
     5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT..........AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG-3'
     3'-TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA..........TCTAGCCTTCTCGAGCATACGGCAGAAGACGAAC-5'
                                                                                           AGCATACGGCAGAAGACGAAC-SYBR_R
```

-Suggest using robot to reduce variability in std. curve and libraries.
-Investigating KAPA BIOSYSTEMS Library Quant Kit.

Cluster and Sequence

Follow routine maintenance for Cluster Station and GA!

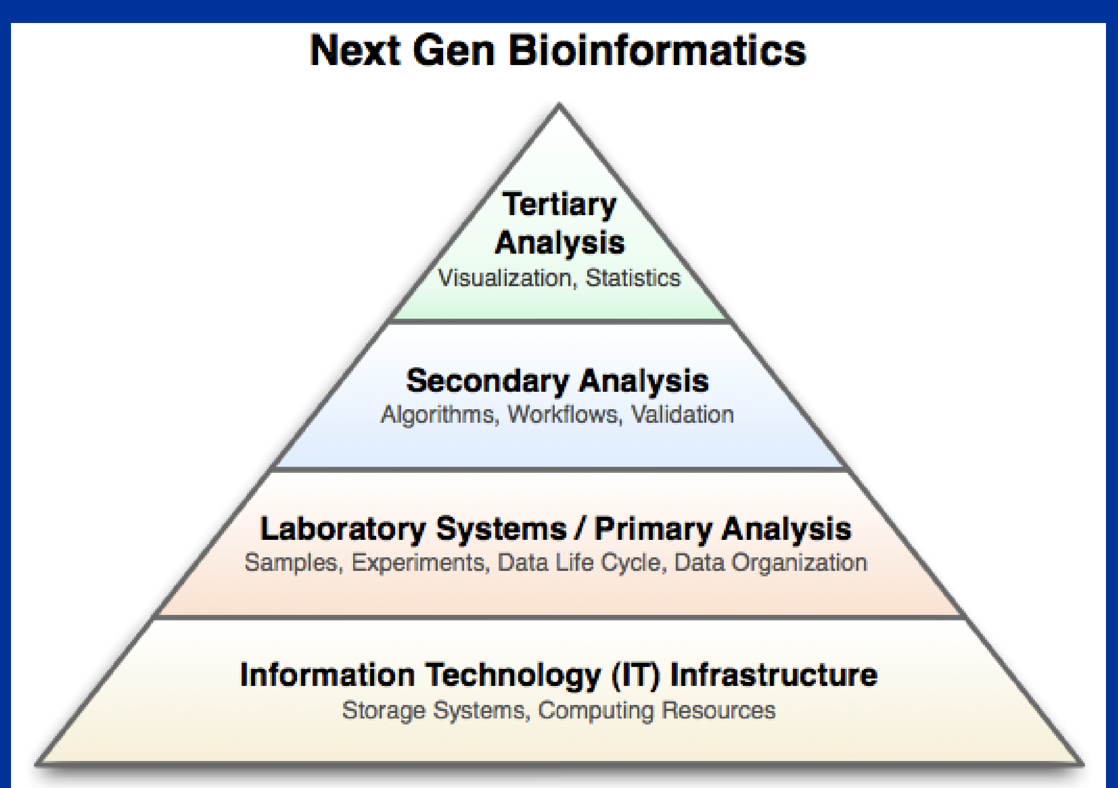
Always, do fluidics line checks on Cluster Station and GA!

Use fresh reagents and track lot #'s.
For longer runs, pre-mix cycle seq mix to reduce run variability

First base report is a good indicator of run quality but will not match final run analysis metrics exactly and that is okay.

Most often successful trouble shooting step when there are no or poor intensity clusters is to re-hyb the seq primer.

What Will Your Core Provide?



In Our Lab:

No detailed Bioinformatics provided.

We provide the middle of the pyramid.

We use Institutional Infrastructure for IT components.

Slide Image Source: Geospiza

Pipeline Notes

Run SCS2.6 and RTA to determine real time
QC Metrics for Run.

HOWEVER

All runs are re-analyzed on HMS Orchestra cluster.
(HMS maintained cluster with 1,000's of cpu's.)
Specifically: Seven IBM blades, 8 cores and 32G RAM each.)

WHY?

Seen an increased use of indexes and odd adaptors by users.
Not base balancing the indexes / adaptors.

Significantly increase pass filter cluster yield by using
ten cycles for cluster identification and specifying which
base to start on. This is too compute intensive for RTA.

Thank You!