Optimizing Cluster Densities

August 12, 2014
Call in number:
Conference code:

illumina

Goals

- What contributes to cluster density
- How to diagnose clustering issues
- How to achieve optimal cluster density

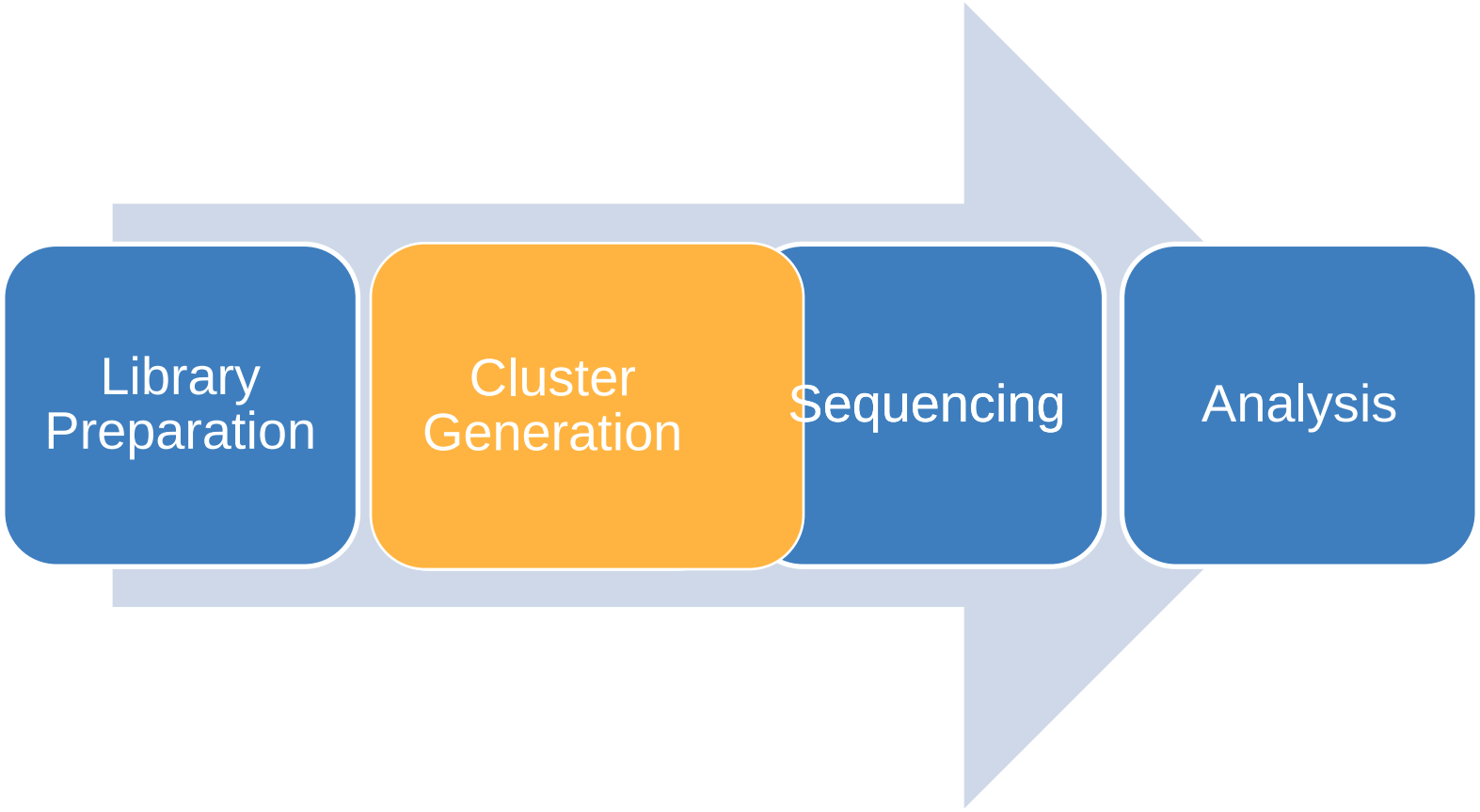
Major Topics of Discussion

This presentation focus on the HiSeq and MiSeq systems for demonstrations, but general concepts apply to our other sequencing systems.

1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

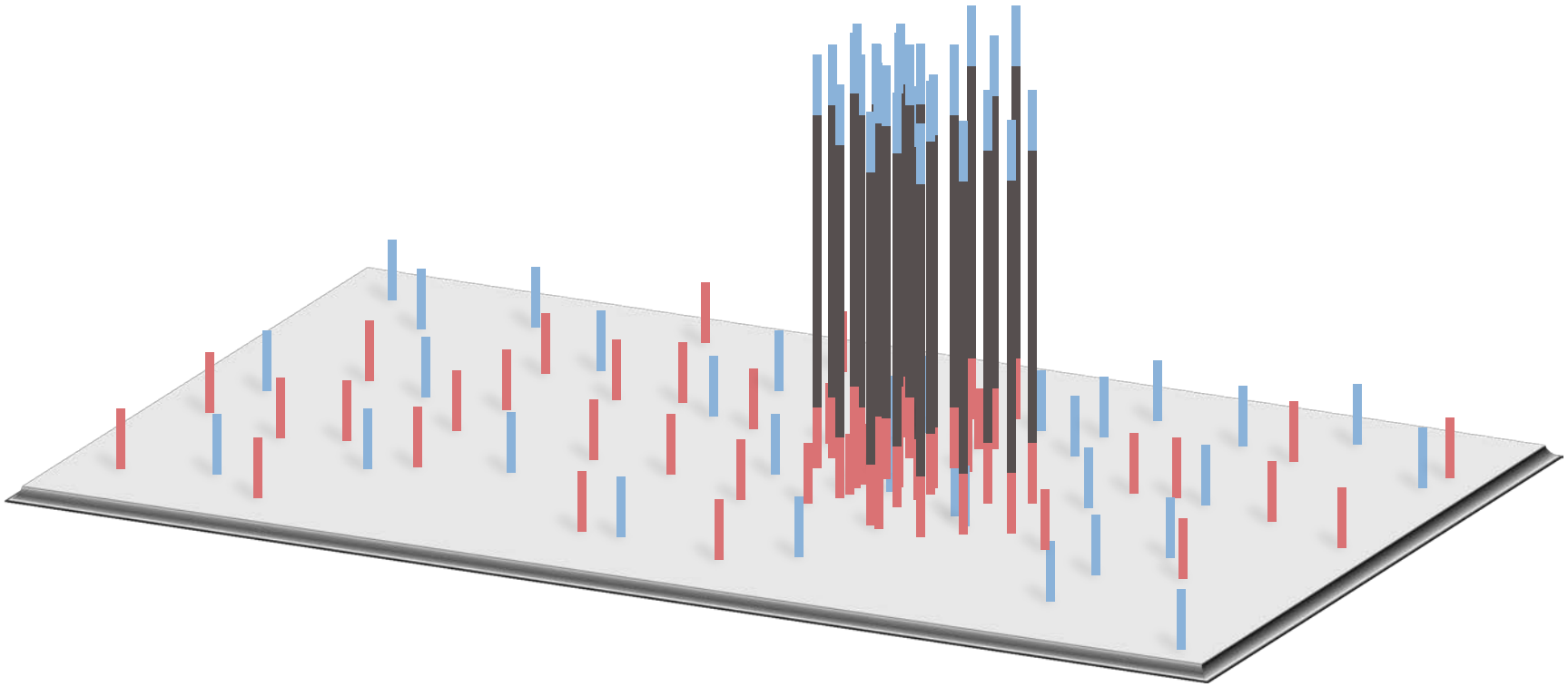
1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

What is a Cluster?



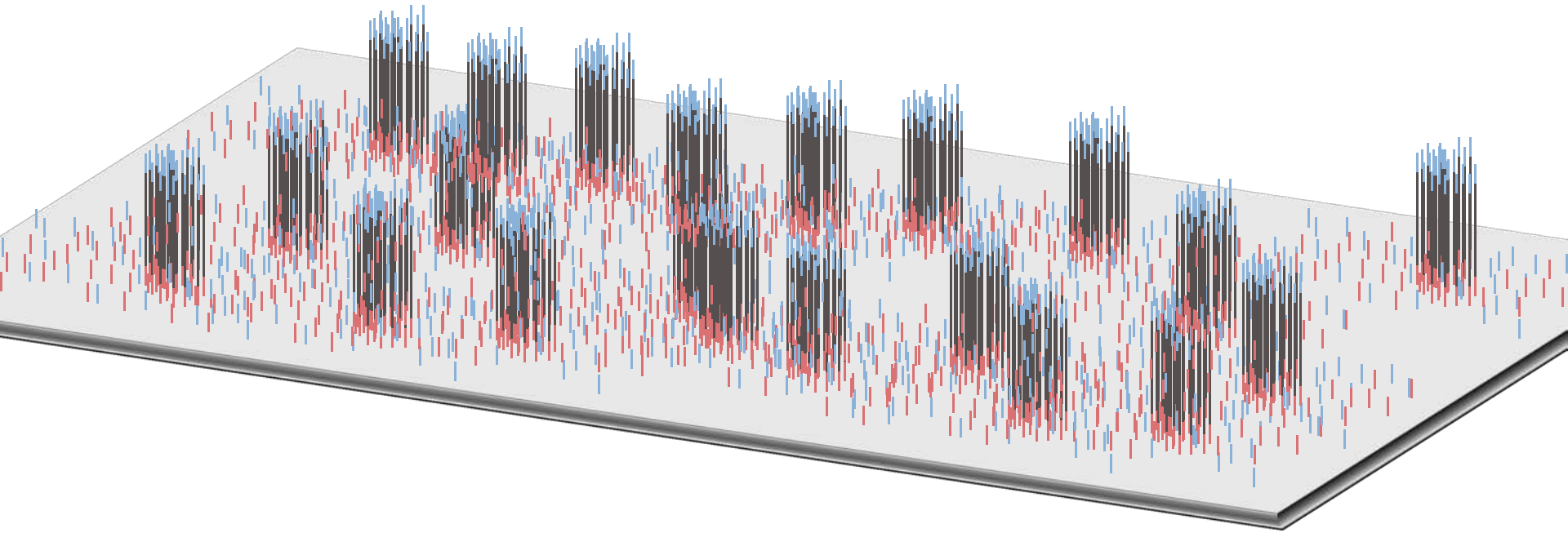
What is a Cluster?

- Many copies of a single fragment from a library that has been clonally amplified
- Covalently bound to flow fell
- Amplified through bridge amplification



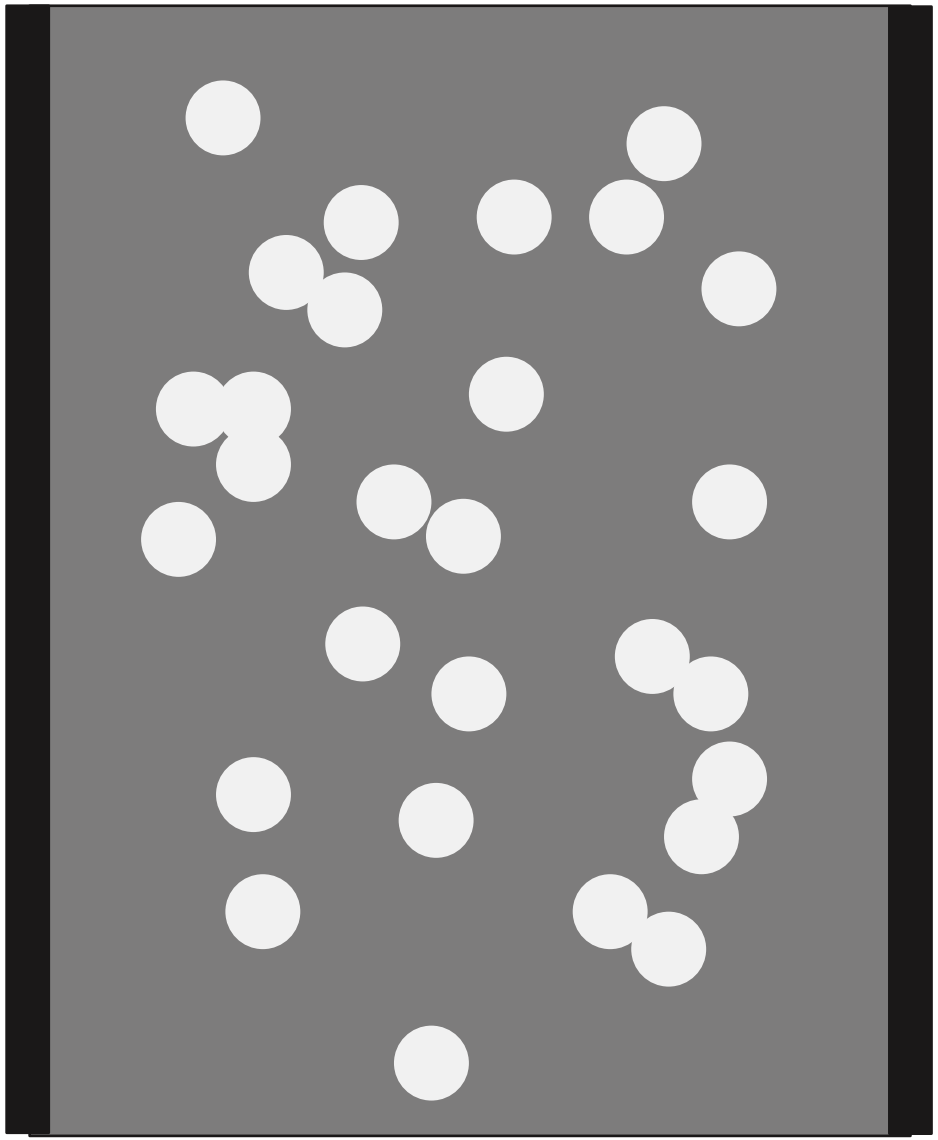
What is a Cluster?

- Many clusters are created across the flow cell
- Massively parallel sequencing



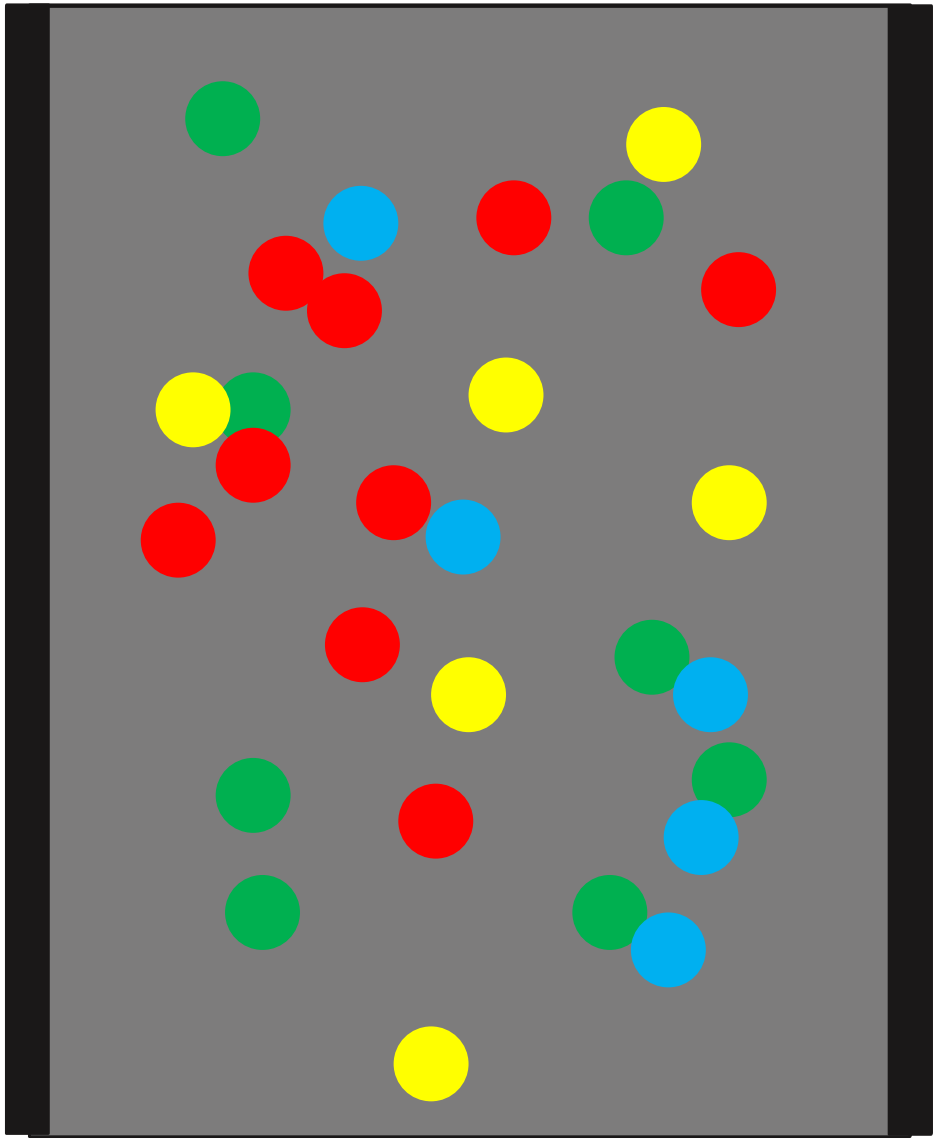
Clustering

- Clusters form during cluster generation
- Clusters do not change in number or location during the run
- During the Paired End Turnaround, size can increase



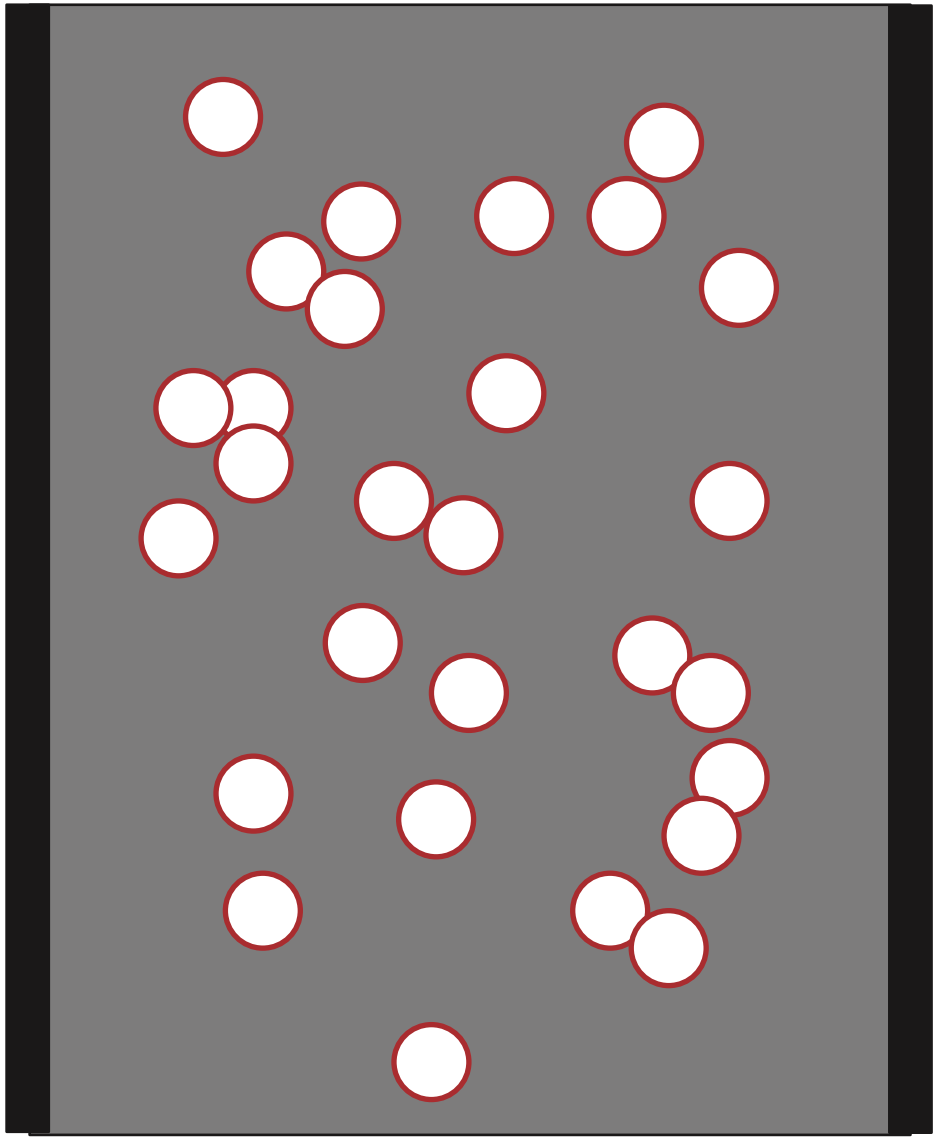
Clustering

- Clusters form during cluster generation
- Clusters do not change in number or location during the run
- During the Paired End Turnaround, size can increase
- Incorporation mix flows over the flowcell, adding a fluorescent terminator tagged nucleotide to the next spot in the sequencing strand

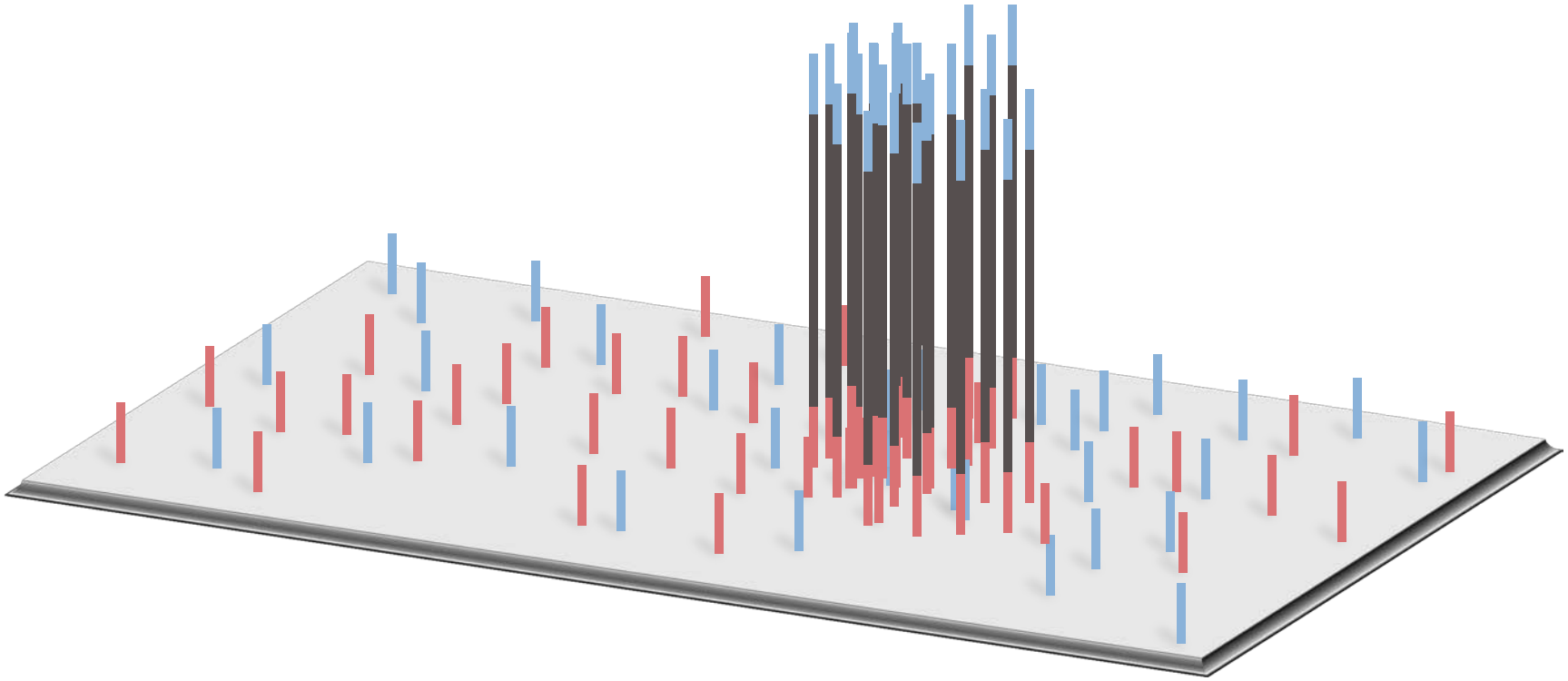


Clustering

- Clusters form during cluster generation
- Clusters do not change in number or location during the run
- During the Paired End Turnaround, size can increase
- Incorporation mix flows over the flowcell, adding a fluorescent terminator tagged nucleotide to the next spot in the sequencing strand
- Cluster map is built in first few cycles of read 1 and is used for the entire run

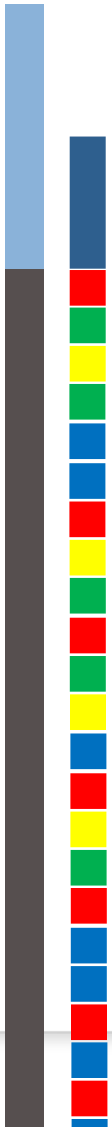


How are clusters imaged?

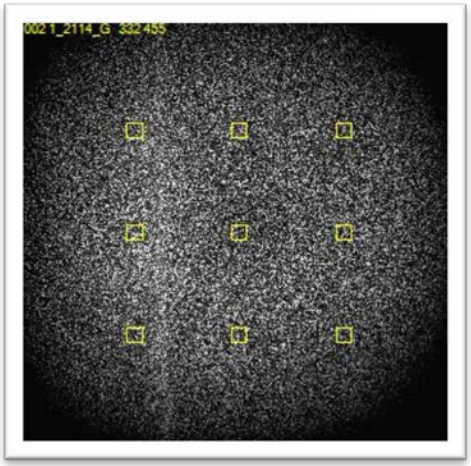


illumina

How are clusters imaged?

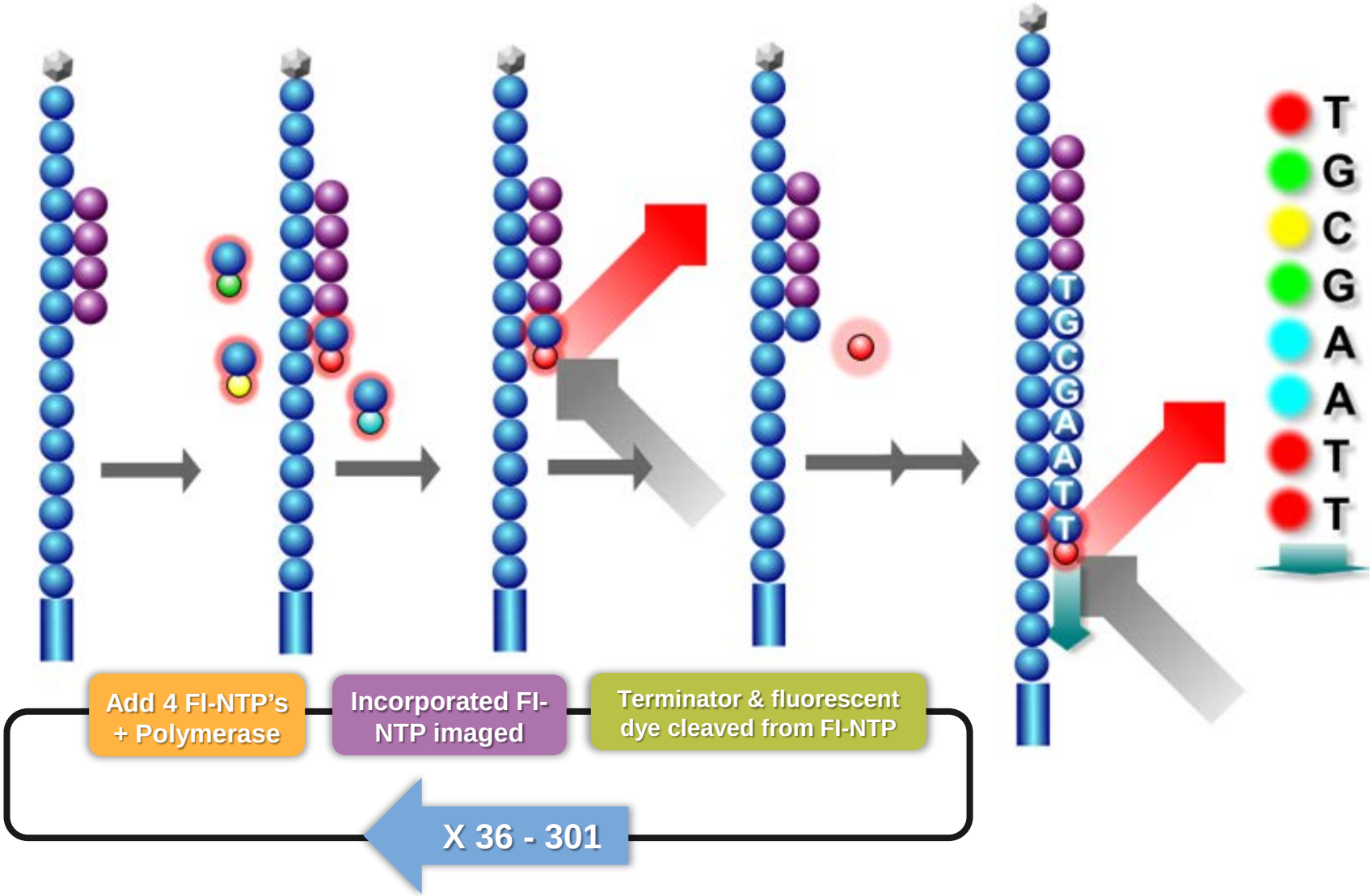


- Clusters can only be seen and mapped if the R1 primer attaches properly
- After 1st base incorporation, imaging takes place



- If no clusters are found, the instrument will not proceed with Sequencing By Synthesis (SBS)

SBS Chemistry

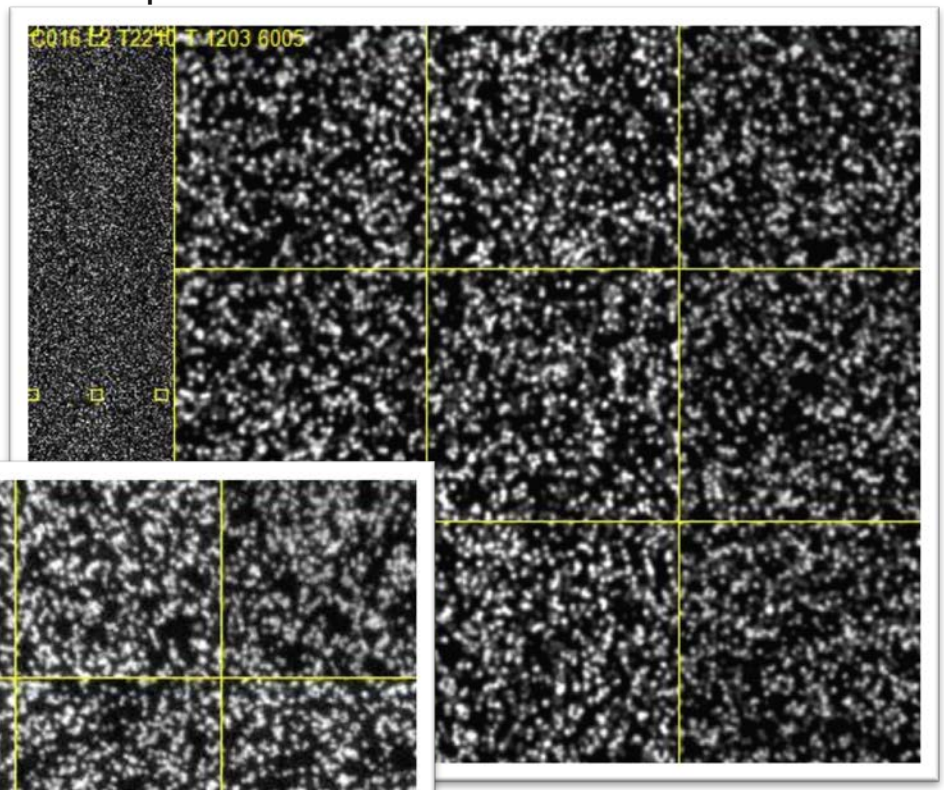


illumina

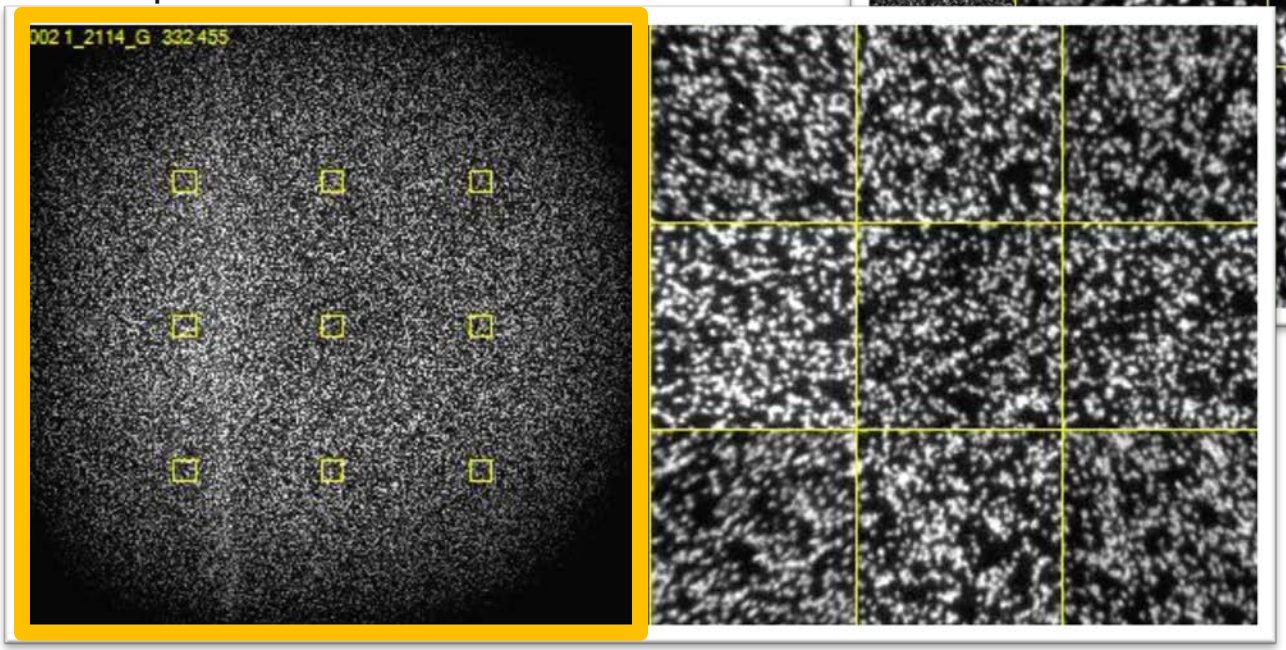
What is a Cluster - Thumbnails

- Thumbnails provide diagnostic information for run issues
- Not a good way to evaluate run quality

HiSeq Thumbnail



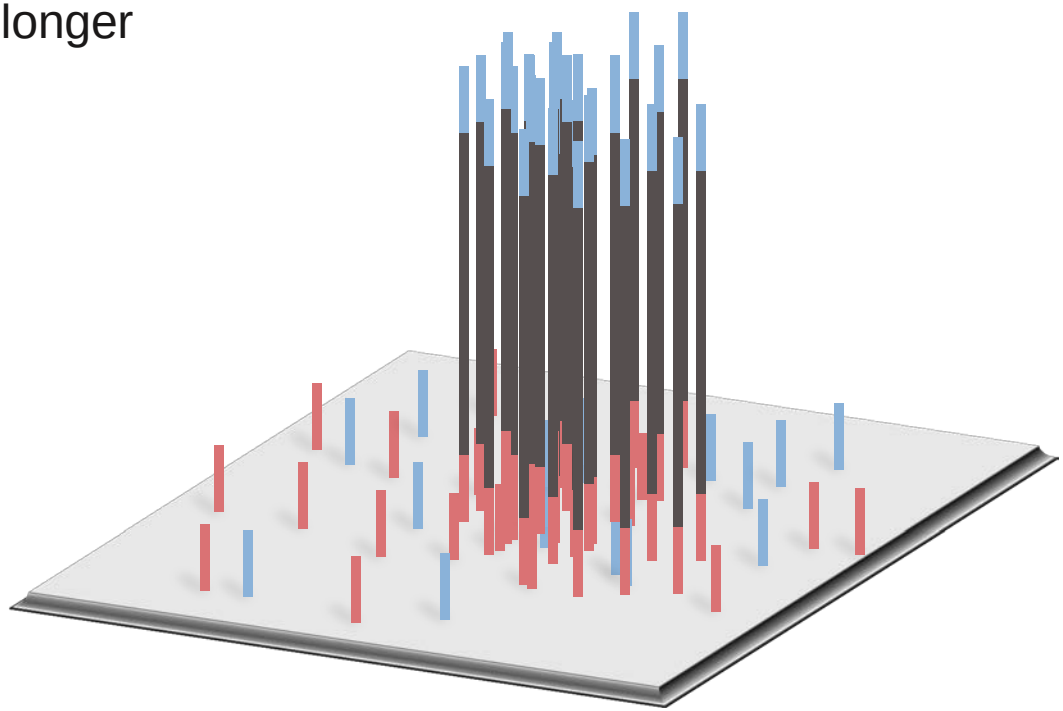
MiSeq Thumbnail



1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

Factors that contribute to clustering efficiency

- Fragment length
  - Short fragments cluster more efficiently than longer ones – there will be more clusters.
  - ~200bp is shorter, ~800bp is longer
  - Similar to most exponential amplifications
- Cluster generation process
  - cBot, HiSeq, MiSeq, and NextSeq user slightly different cluster generation recipes
- NaOH
- Sample-specific factors

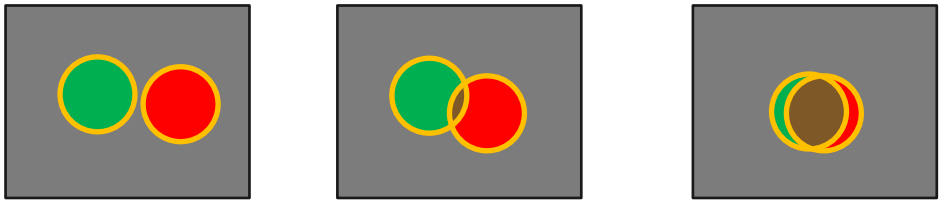


Four Major Topics of Discussion

1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

Clusters vs Clusters PF (Passing Filter)

- The “Filter” prevents multiple template signal or low quality clusters



- Clusters PF is generated from the first 25 cycles
- Only clusters that pass filter will be counted towards run yield
- Clusters passing filter = Single end reads

MiSeq
Focused power. Speed and simplicity for targeted and small genome sequencing.

NextSeq 500
Flexible power. Speed and simplicity for everyday genomics.

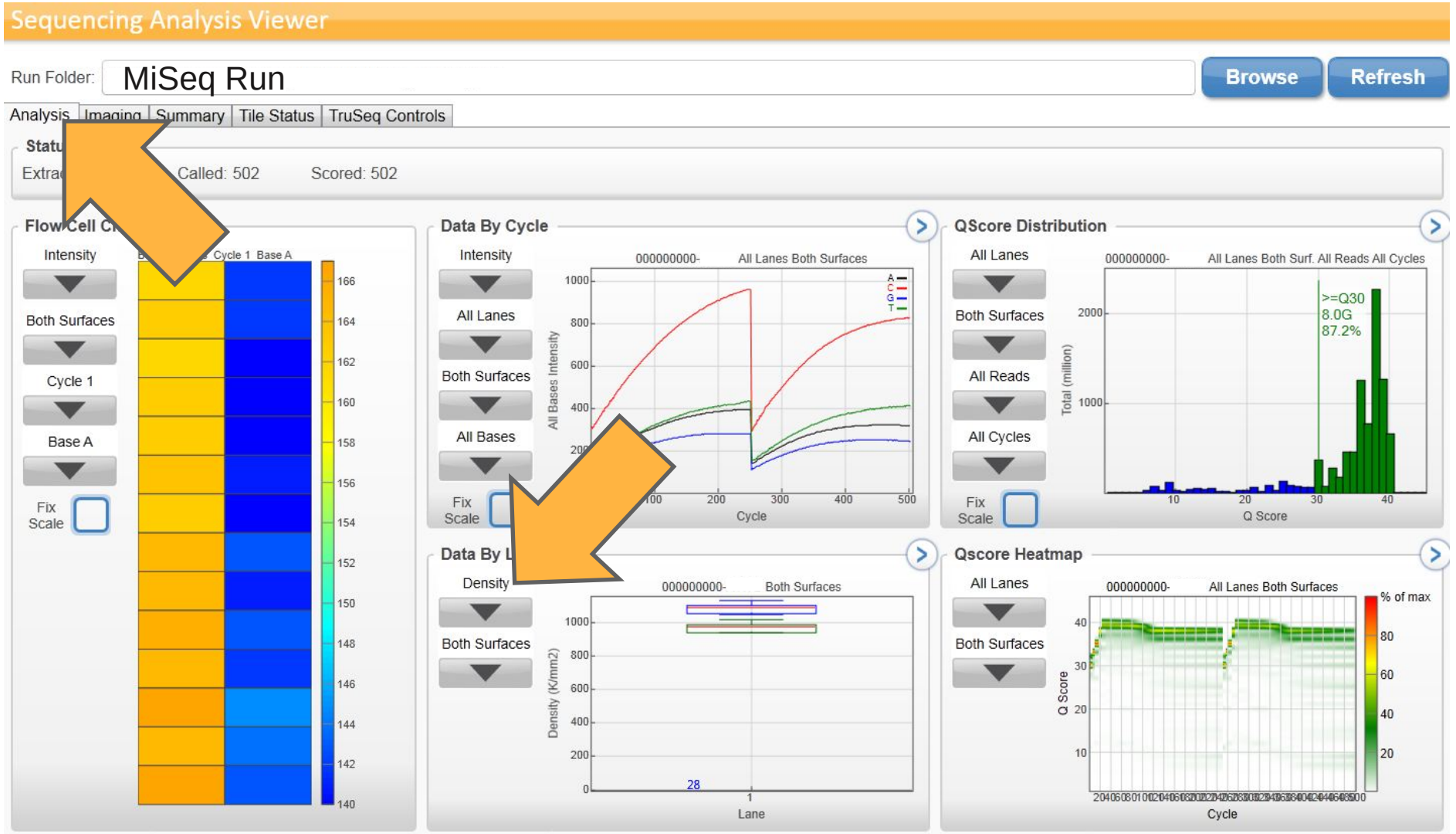
HiSeq 2500
Production power. Power and efficiency for large-scale genomics.

HiSeq X*
Population power. $1,000 human genome and extreme throughput for population-scale sequencing.

Reads per flow cell†	25 Million‡	130 Million	400 Million	300 Million	2 Billion	3 Billion

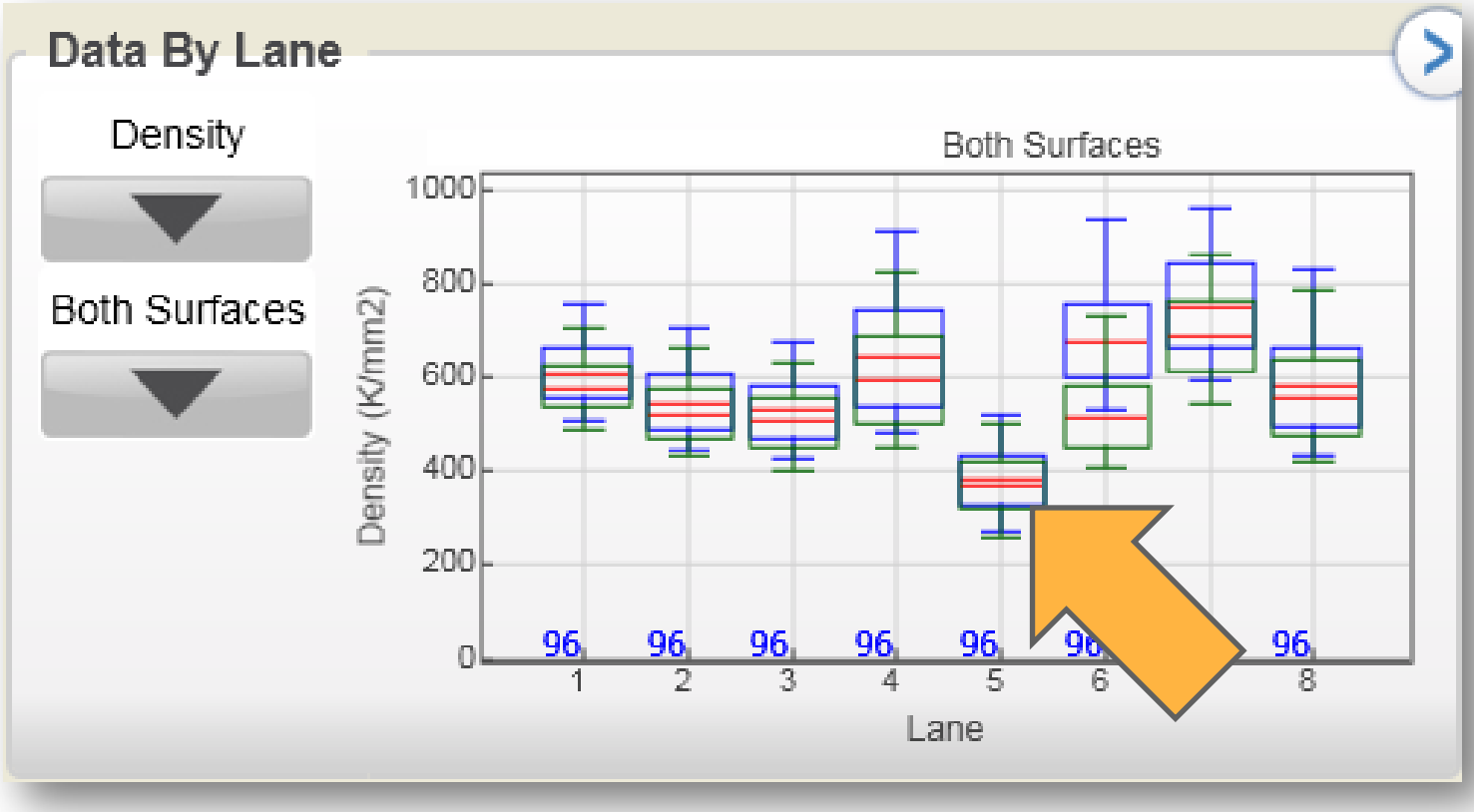
illumina

Viewing %PF Score – Sequence Analysis Viewer (SAV) – Summary Tab



illumina

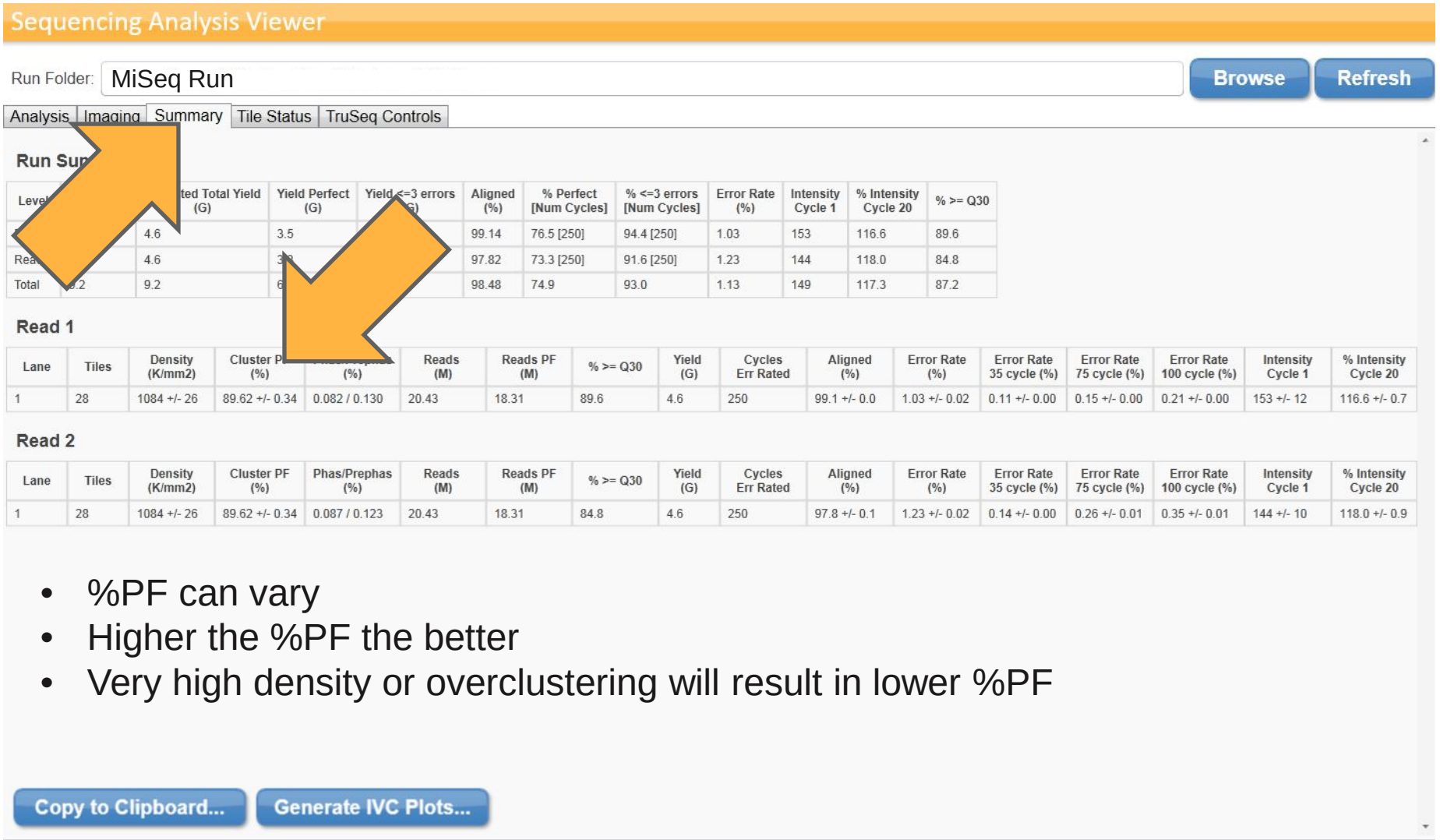
Data By Lane: Density –Total and %PF



Box Plots	
Blue	Raw Cluster Counts
Green	Clusters passing filter
	Median

illumina

Viewing %PF Score – Sequence Analysis Viewer (SAV) – Summary Tab



- %PF can vary
- Higher the %PF the better
- Very high density or overclustering will result in lower %PF

illumina®

- What is a cluster?
- Factors that contribute to clustering efficiency
- Clusters vs Clusters PF (Passing Filter)
- Common clustering issues
- How to get good cluster numbers
- Resources

Common clustering issues

- **Overclustering**
  - Loaded too much library
    - Quantification
    - Compensation for library efficiency variation
    - Low diversity libraries cannot reach same density as normal libraries
- Underclustering
  - Loaded too little library
  - NaOH
  - Poor library

- **Overclustering will result in less or no data**

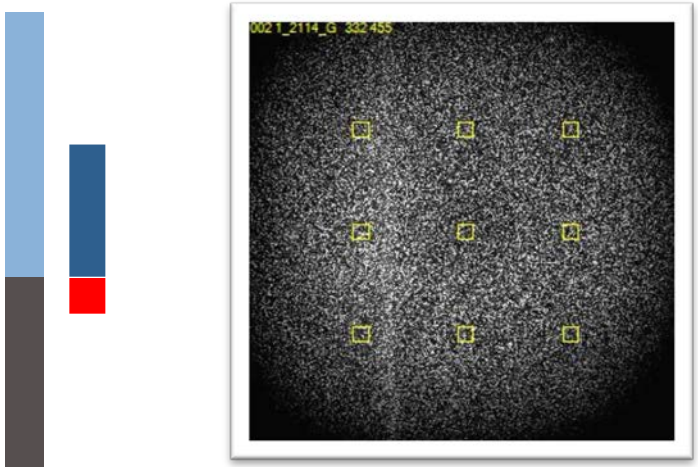
Stick to cluster density range

Platform	Optimal Raw Cluster Density (CD)
GAIIx (v5 cluster kit, >SCS 2.9)	700-800K clusters/mm^2
HiSeq 2000/2500 High Output (v3)	750-850K clusters/mm^2
HiSeq 2500 High Output (v4)	950- 1150K clusters/mm^2
HiSeq 2500 Rapid Run	850-1000K clusters/mm^2
MiSeq v2 Reagents	1000-1200K clusters/mm^2
MiSeq v3 Reagents	1200-1400K clusters/mm^2
NextSeq 500	160-180K clusters/mm^2

- Raw metrics are given as K clusters/mm2 (1,000 clusters per millimeter squared)
- Metrics are for high diversity samples

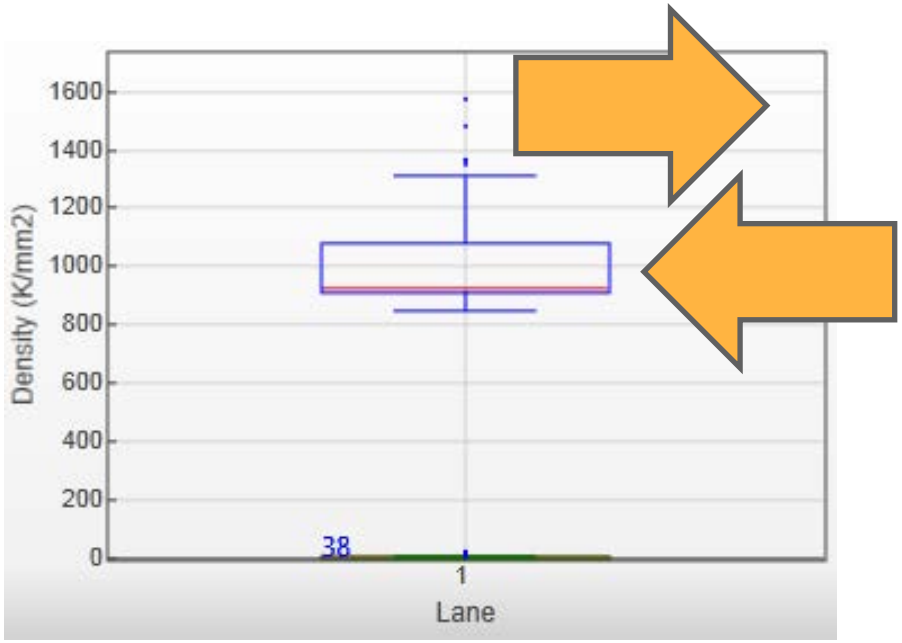
Common clustering issues

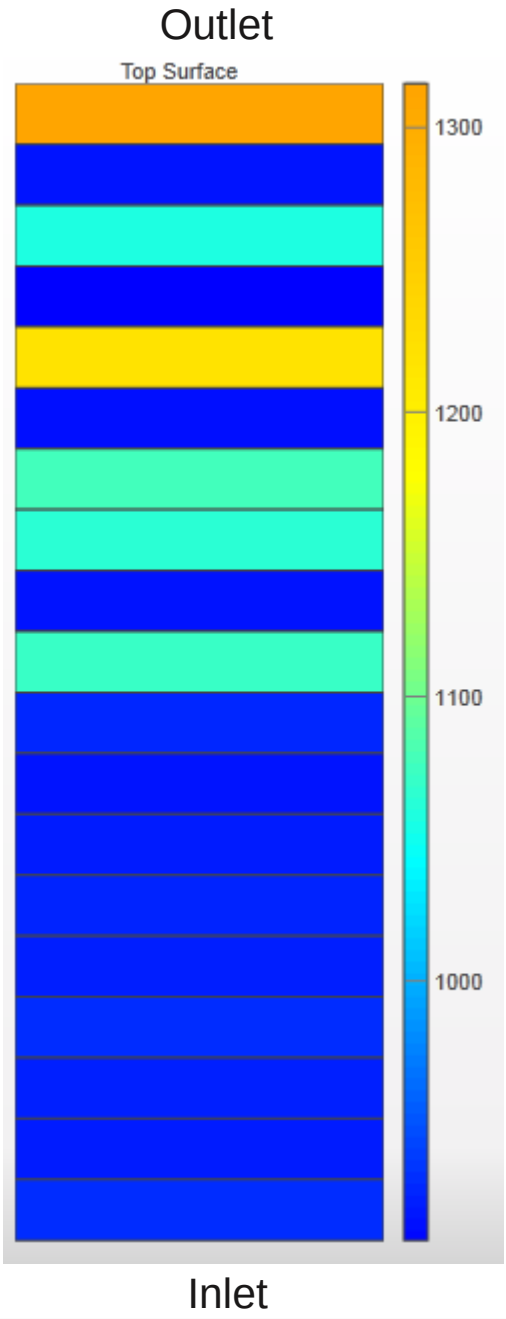
- Inconsistent clustering - Can be inconsistent from run to run or across the flow cell
  - Library issues
    - Non-standard libraries may not be optimized for the platform
- Primer binding
  - Not true “clustering issue” but can manifest similarly



Overclustering

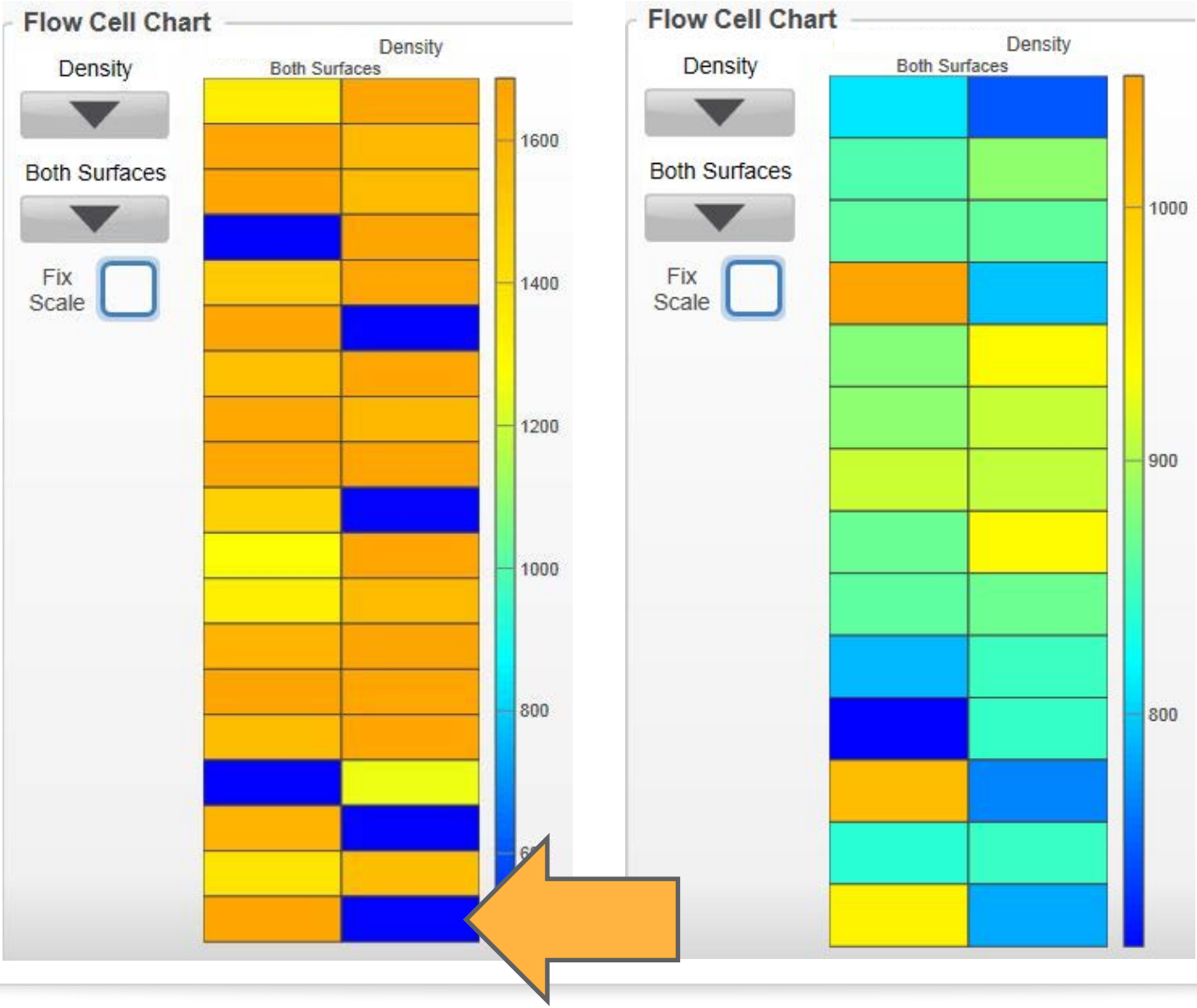
- Can manifest in multiple ways:
  - Really high cluster numbers
  - Really low cluster numbers
  - Drop in intensity after cycle 5
  - Drop in intensity/quality after Paired End Turnaround





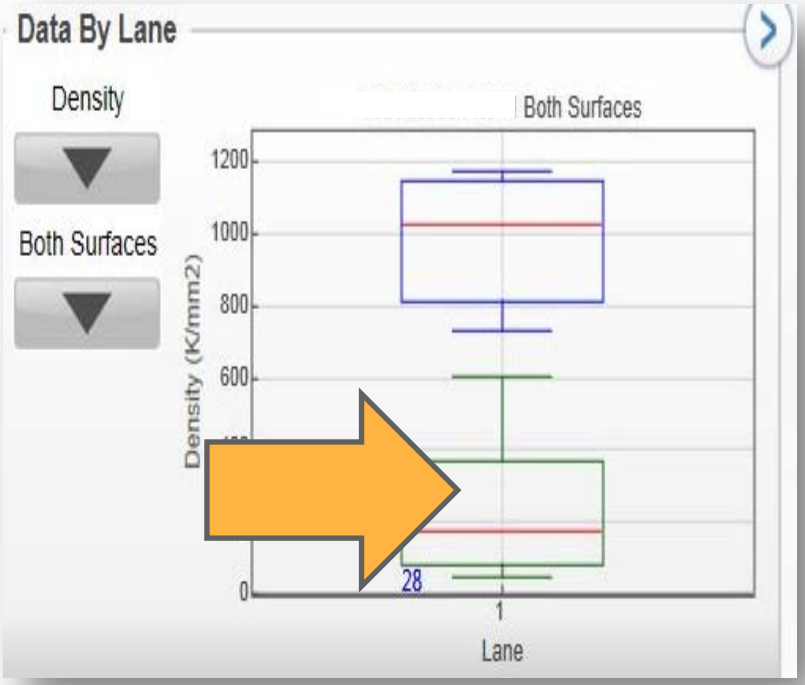
illumina

Over clustering – inconsistent cluster density



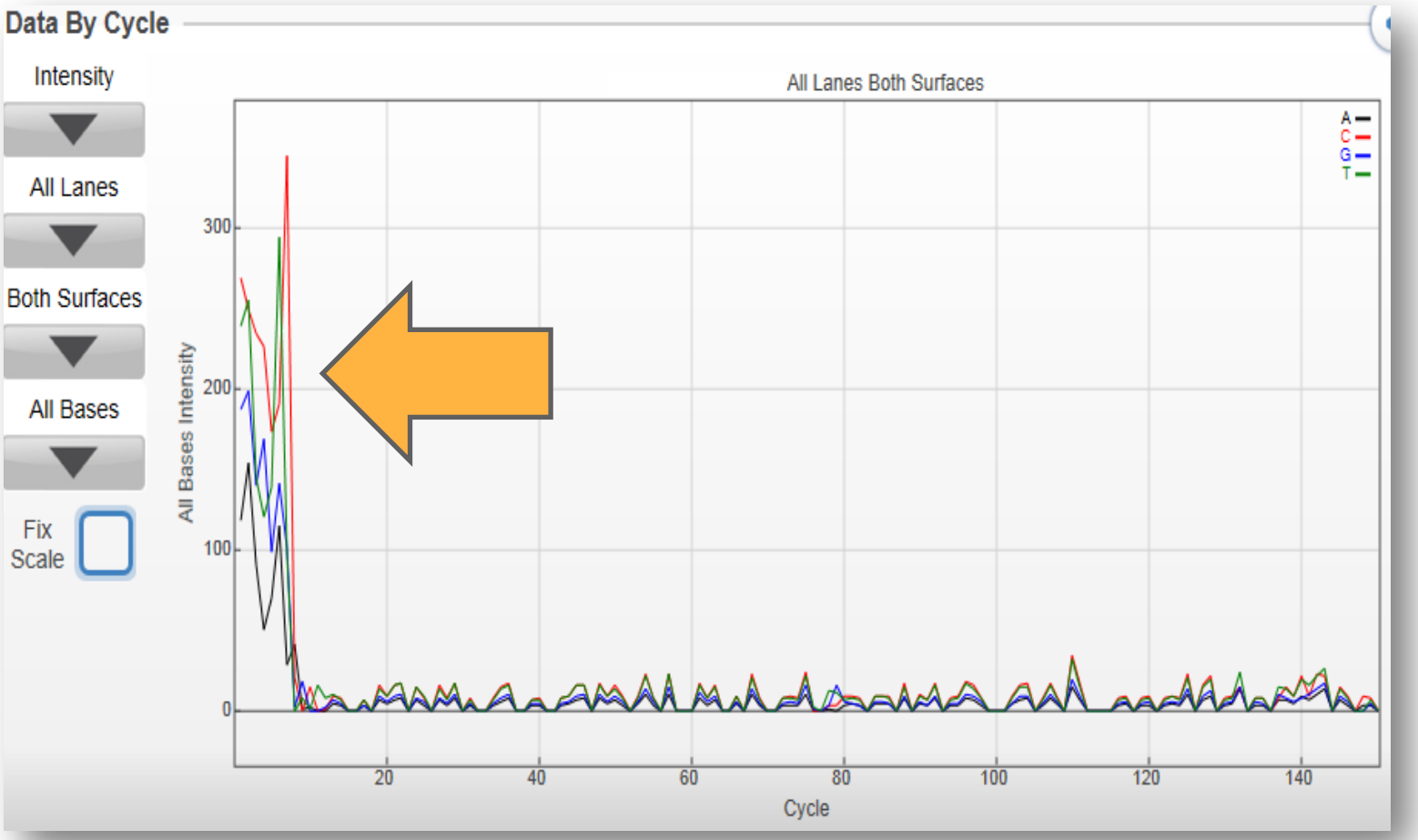
illumina®

Over clustering – Low %Pass Filter



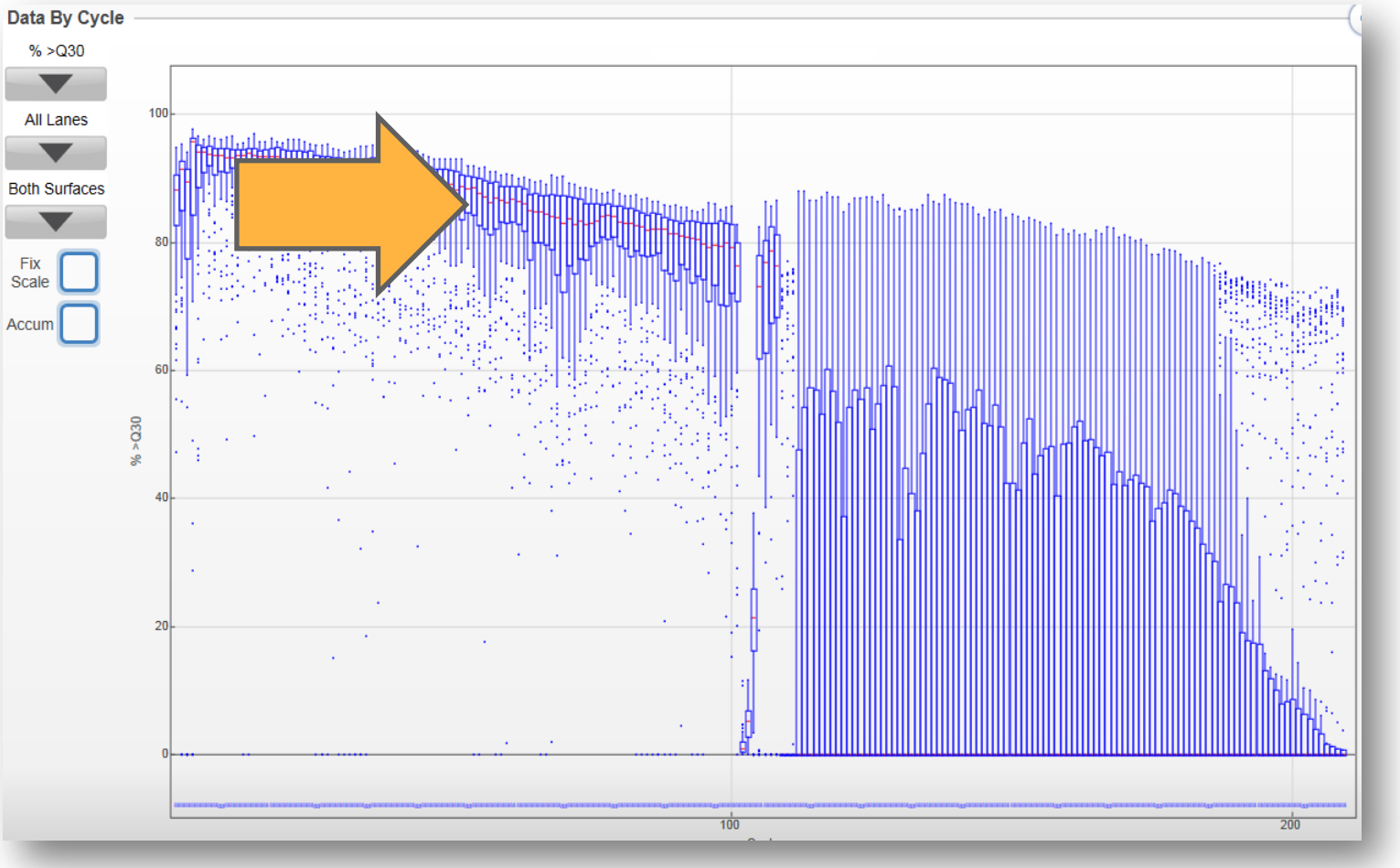
illumina

Over clustering - Drop in intensity after cycle 5



illumina

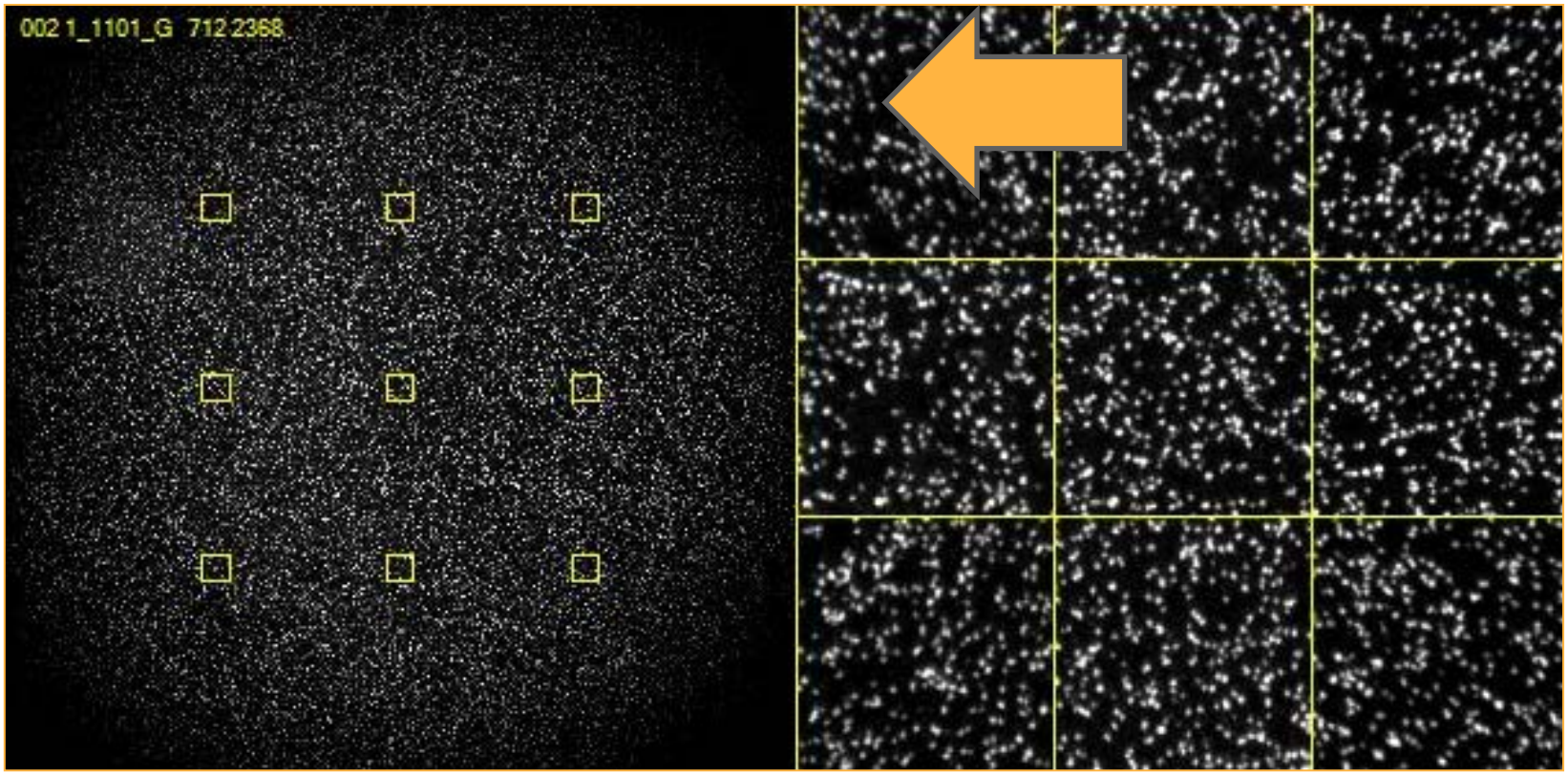
Over clustering - Drop in intensity/quality after Paired End Turnaround



illumina®

Cluster Density: Thumbnail Images

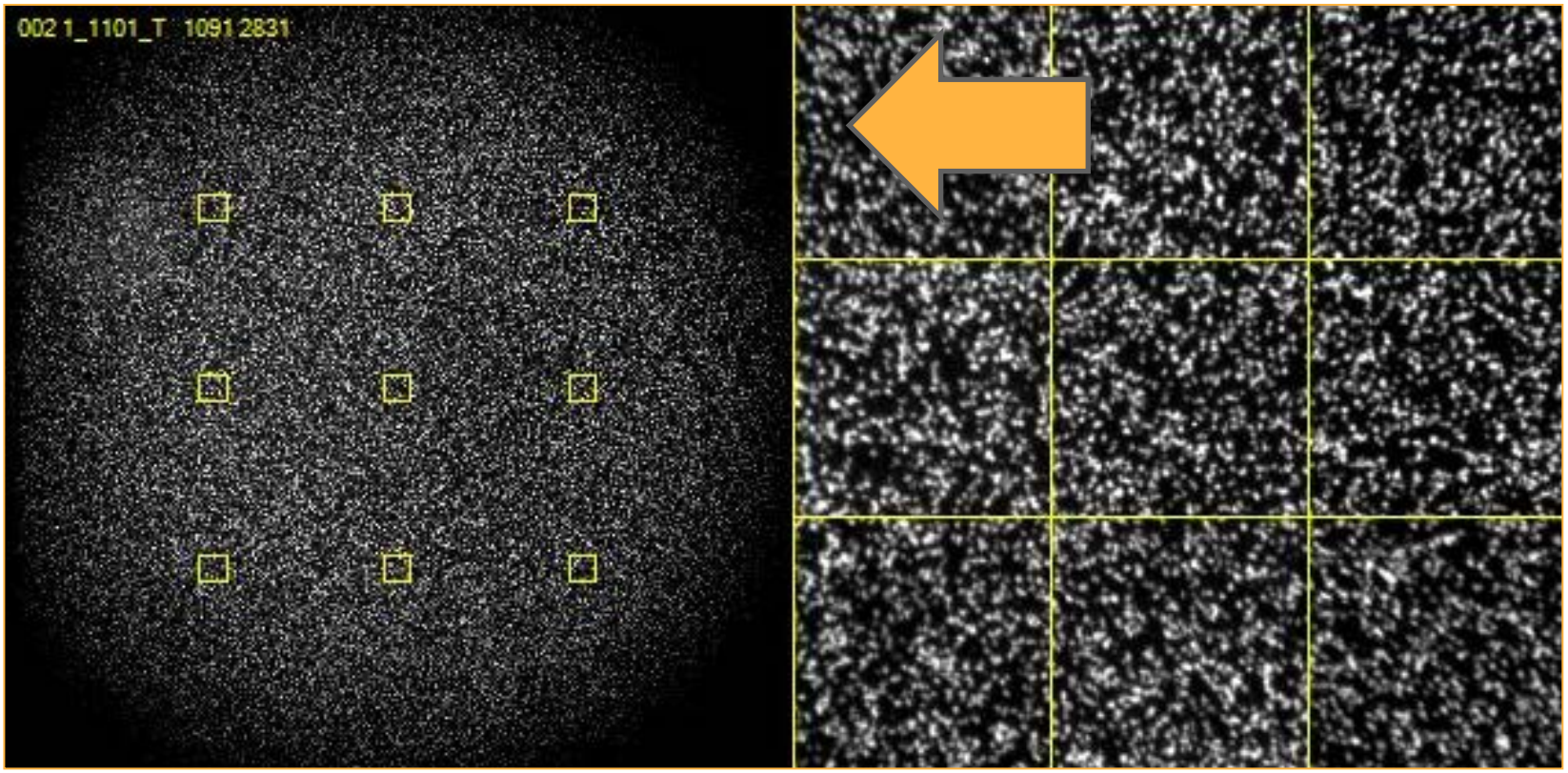
Note: Cluster density appearance in thumbnails will vary on different platforms



Normal/low density – MiSeq v2

Cluster Density: Thumbnail Images

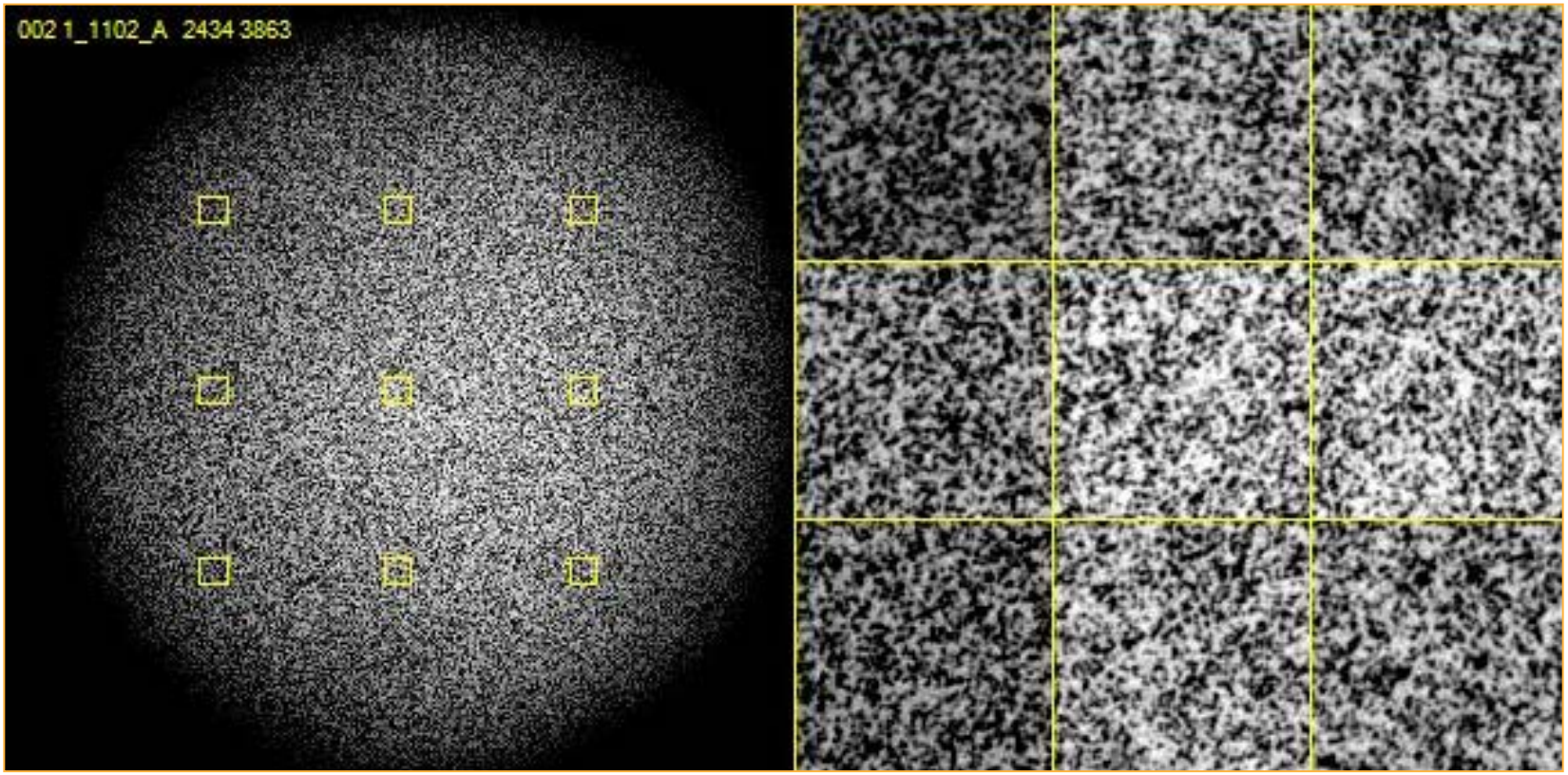
Still white spots on a black background



Highest recommended density – MiSeq (v2)

Cluster Density: Thumbnail Images

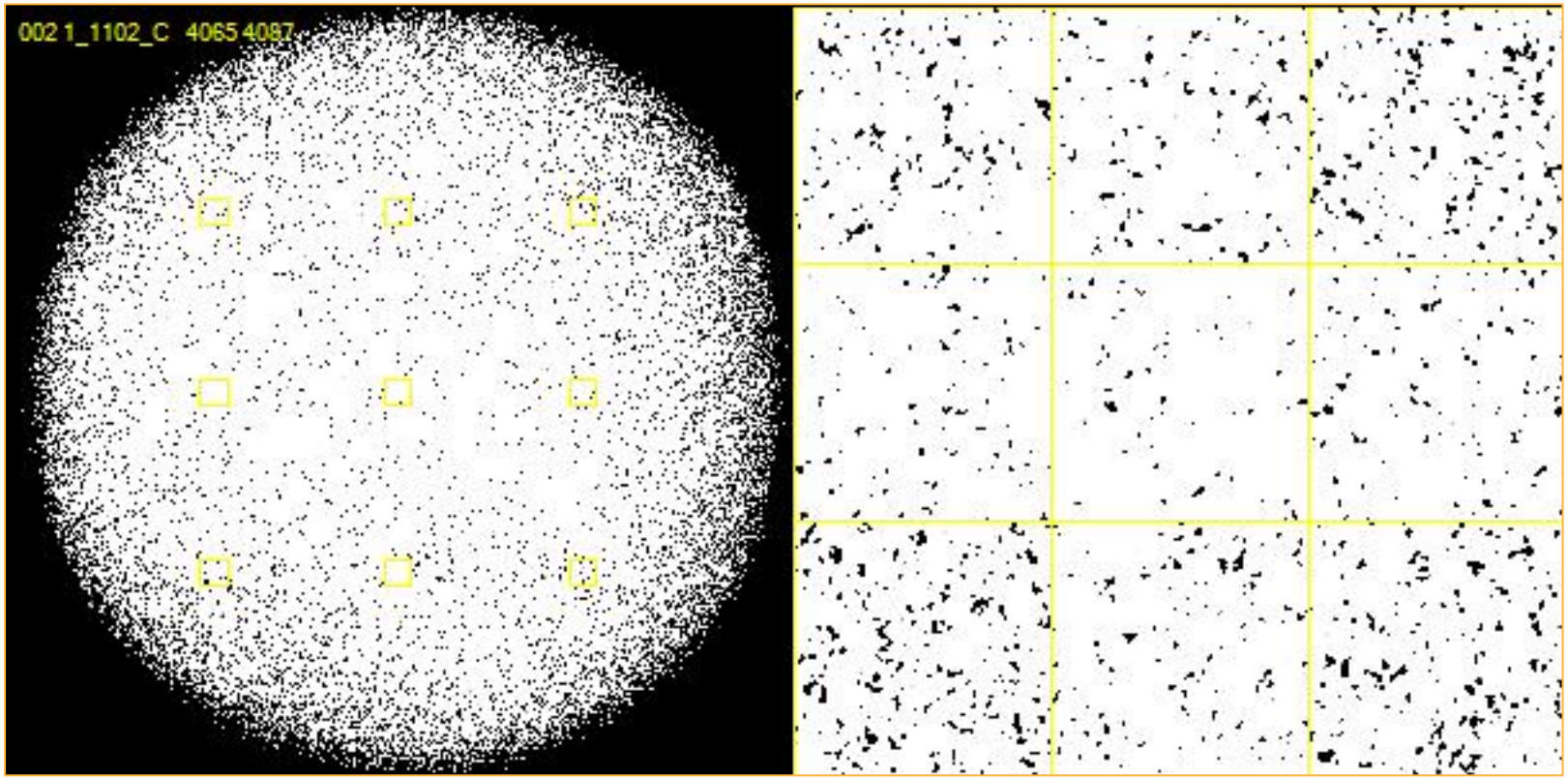
No clear white spots – more of a haze



Over-clustered! – MiSeq (v2)

Cluster Density: Thumbnail Images

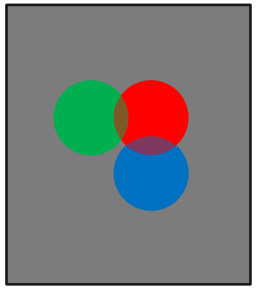
Can manifest as clustering failure

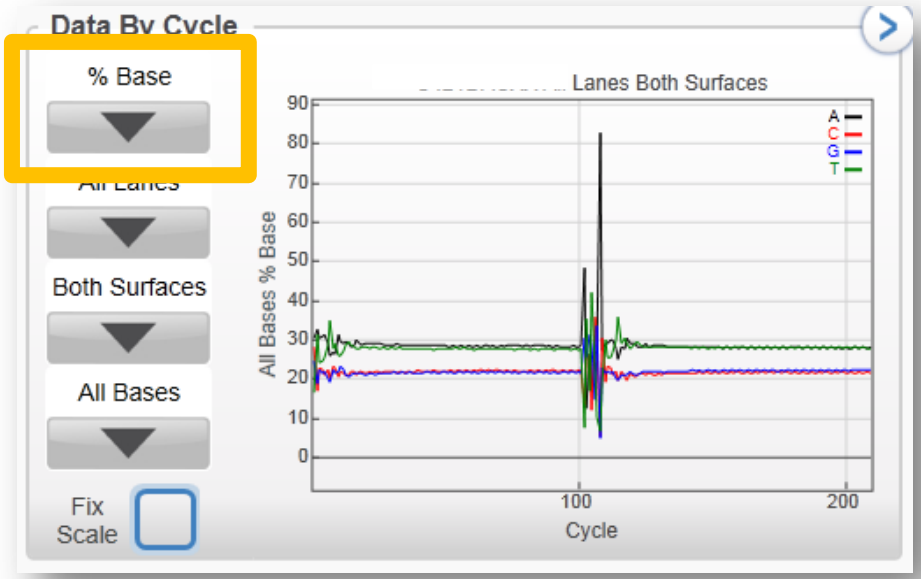


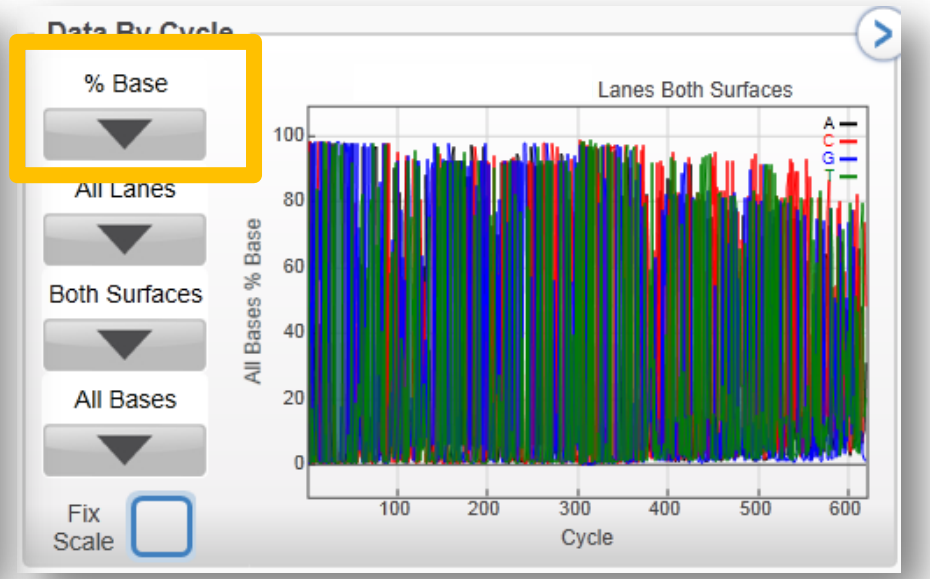
Saturated! (Cannot measure CD accurately)

Diversity

- Cluster density specifications are for high diversity libraries
- Low diversity samples do not use random fragmentation during library preparation
- Most standard Illumina library preparation methods have good diversity



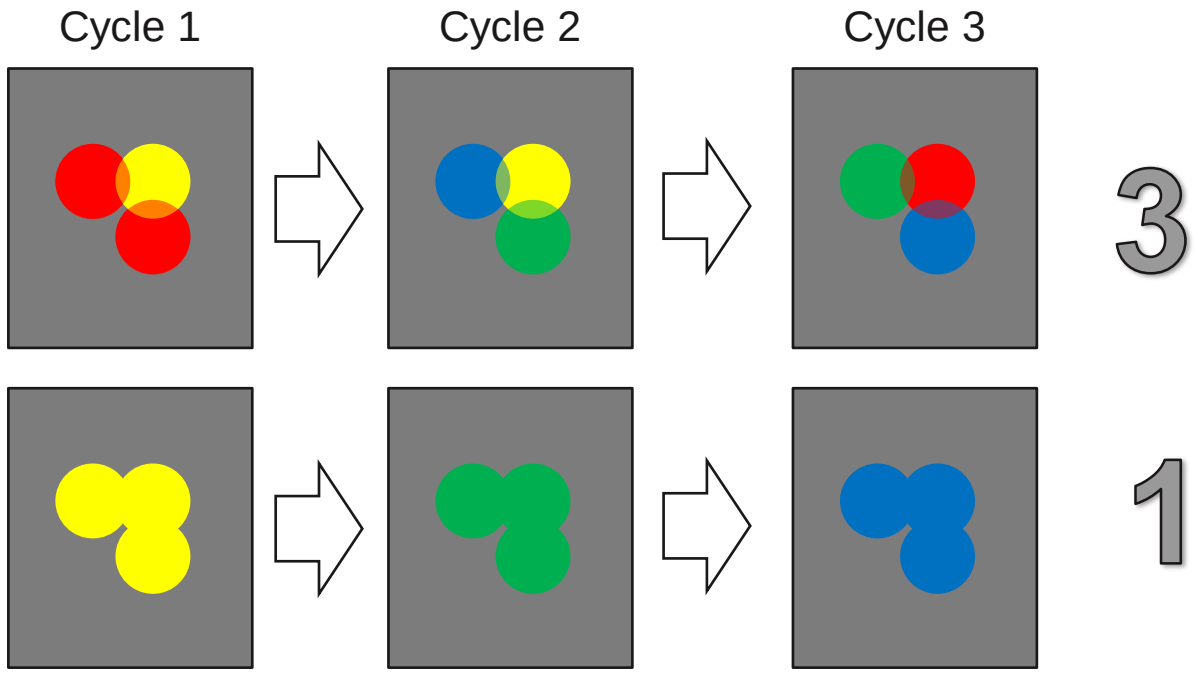




illumina

Diversity

- Low diversity samples cannot reach the same cluster densities
- Clusters that do not have different signals during cluster mapping will be called as one cluster which will fail chastity or have poor quality later in the run



- **When running low diversity libraries, keep densities low and PhiX spike in high**

1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

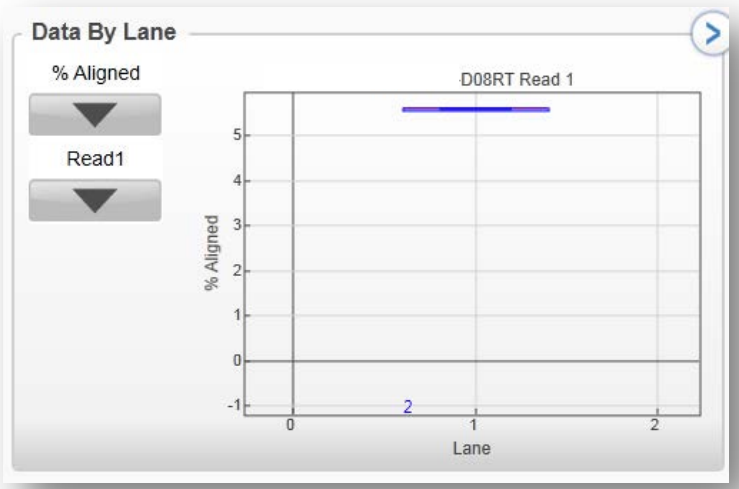
How to get good cluster numbers

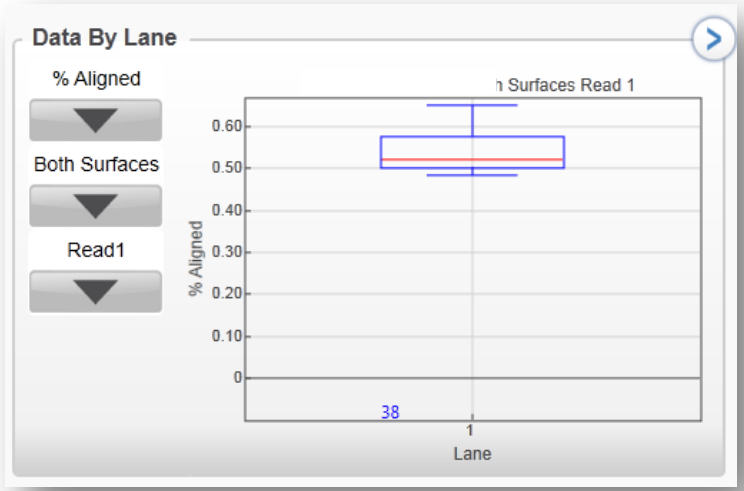
- Good quantification!
  - Use the quantification method for your specific library preparation kit
- Empirical adjustment in small increments
  - Start low and increase until you are in the recommended range, then stay there
- Repeat targeting bottom up for new library types
  - A new library method can have very different clustering efficiencies
  - When running a new library, lower cluster density to avoid over clustering
- Loading concentration to cluster density is not always a linear relationship

Sample Prep Method	Quality Check on Bioanalyzer	Recommended Quant Method	Comments
TruSeq DNA (PCR-Free, Nano)	✓	qPCR	
Nextera DNA	✓	Qubit	
Nextera XT		None	Built-in normalization step Final library is single stranded
Nextera Mate Pair	✓	qPCR	Can validate library size by gel or Bioanalzyer
TruSeq Synthetic Long Reads	✓	qPCR	
Nextera Rapid Capture	✓	Qubit	No carrier DNA present
TruSight panels	✓	Qubit	
TruSight Tumor	✓	None	Built-in normalization step
TruSight Myeloid		None	Built-in normalization step Final library is single stranded
TruSeq Amplicon (Custom, Cancer)		None	Built-in normalization step Final library is single stranded
TruSeq RNA (v2, Access, Stranded)	✓	qPCR	
TruSeq Targeted RNA Expression	✓	None	Built-in normalization step based on TruSeq Targeted RNA Calculator
TruSeq Small RNA	✓	Bioanalyzer	Reduced quant precision needed due to low concentration libraries and short read lengths
TruSeq ChIP-Seq	✓	qPCR	

How to get good cluster numbers

- Compare PhiX alignment with PhiX spiked in
  - PhiX is automatically aligned and parsed out
  - Helpful in determining approximate clustering efficiency compared to PhiX
  - %Aligned = PhiX Aligned
  - %Aligned should be close to % spiked in





  - Can help troubleshoot between quantification concerns and clustering efficiency efficiency concerns

Goals

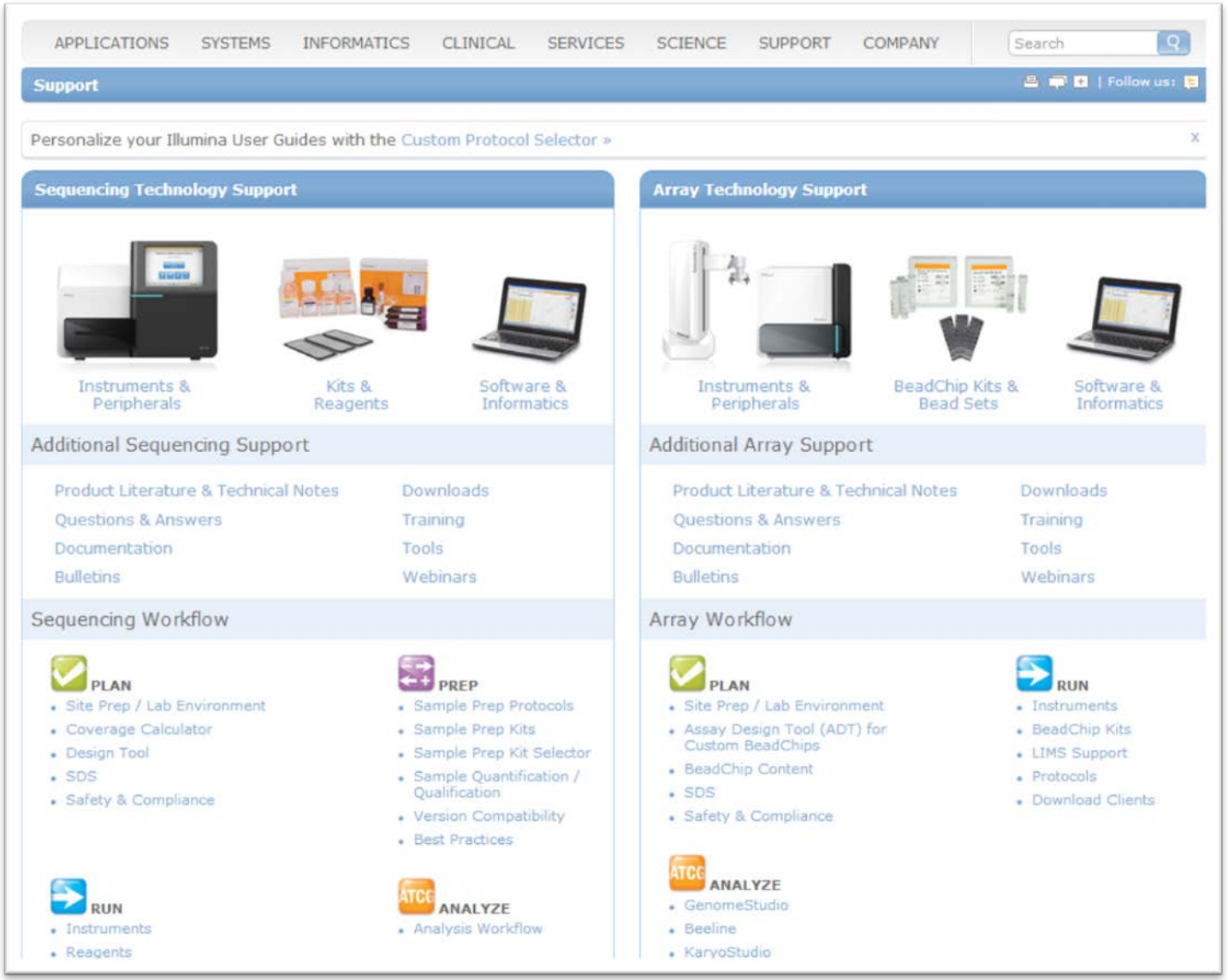
- What contributes to cluster density
  - Loading concentration
  - Library length
- How to diagnose clustering issues
  - Overclustering
  - Underclustering

Underclustering is preferable to over overclustering

- How to achieve optimal cluster density
  - Good quantification
  - Target optimal density by increasing in small amounts

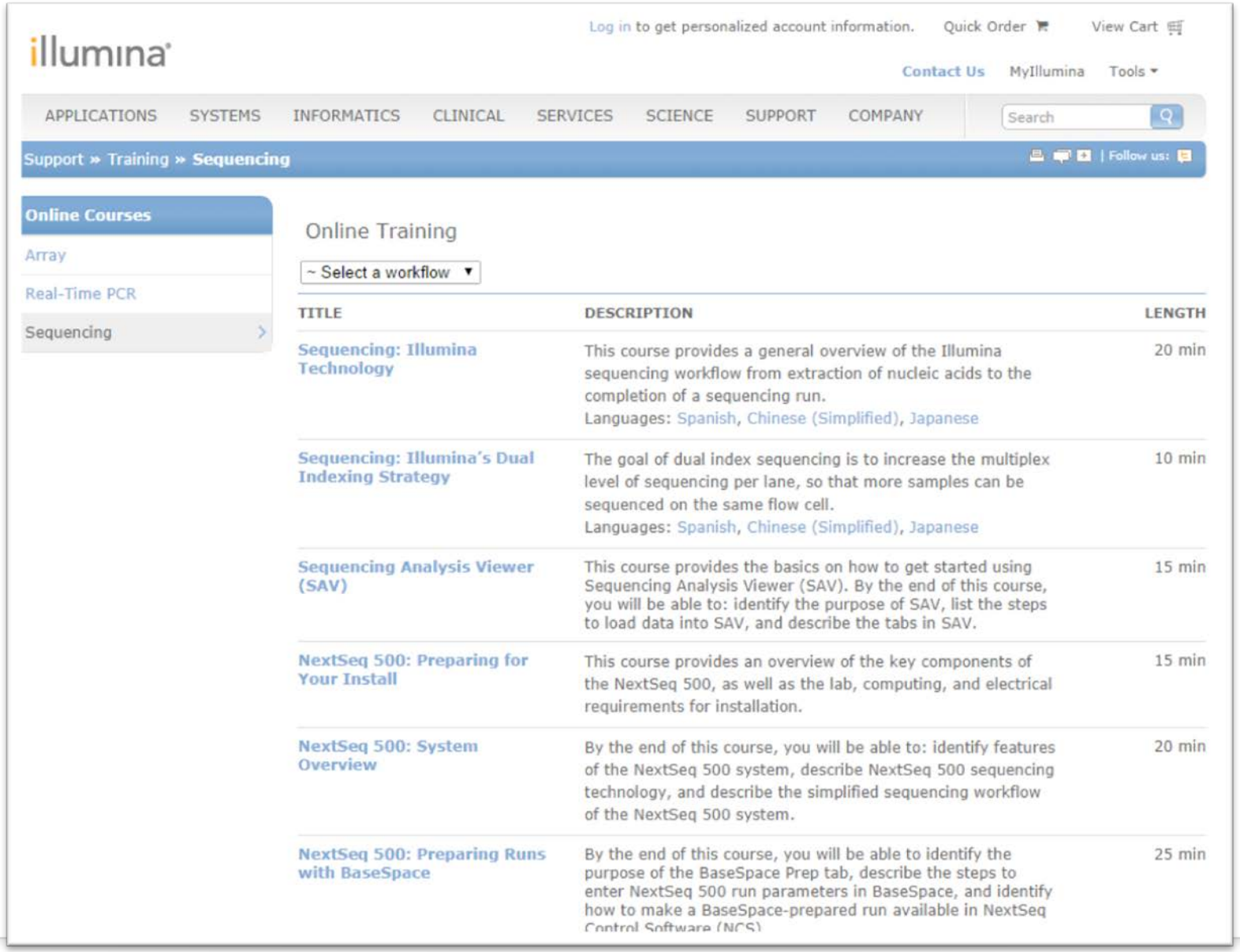
1. What is a cluster?
2. Factors that contribute to clustering efficiency
3. Clusters vs Clusters PF (Passing Filter)
4. Common clustering issues
5. How to get good cluster numbers
6. Resources

Resources – Support Pages



illumina

Resources – Training Videos



Resources - Bulletins

Five Things You Can Do to Improve the Quality and Quantity of Your Sequencing Data

Dec 18 2013

By using these five basic strategies, you can n
Illumina sequencing platform.

Keep your cluster density within Illumina
recommended cluster densities is very importa
Exceeding the recommendation may have neg
output. You can find cluster density specificatio

Add PhiX a
designed to
provide erro
here for the
kit, a positiv
recommend

Verify you
of the prope
Bioanalyzer
library can b

Double che
Illumina libr
need the Tru
always mak

Follow bes
denatured c
concentratio
sequencing

Considerations for Library Migration Between Illumina Sequencing Platforms

Oct 22 2013

While each of Illumina's sequencing platforms and modes use the same Illumina technology to obtain sequencing data, there are slight variations in the chemistry utilized on different sequencing platforms. All libraries constructed with currently available Illumina sample prep kits are compatible with all Illumina sequencing platforms.

ring between different
platforms when working with

red-end flow cells are
Seq flow cells are paired-
d libraries before running on

mperature during on board
Bot, which is utilized for
rature of 60°C.

rm deblocking at 60°C and
for both of these steps.

emistry, onboard cluster
ter at approximately
loaded on a GA or HiSeq High
deration when migrating

Cluster Density Specifications for Illumina Sequencing Platforms

May 19 2014

Cluster density is an important factor in data quality and yield. The following table includes a list of optimal raw cluster density specifications for balanced libraries (such as PhiX):

Platform	Mode/Reagents	Optimal Raw Cluster Density
GAIIx	SCS 2.8 or higher, TruSeq Cluster v5 reagents	700–800 K/mm²
HiSeq	High output, TruSeq v3 reagents	750K–850 K/mm²
HiSeq	High output, HiSeq v4 reagents (requires upgrade)	950-1050 K/mm²
HiSeq	Rapid run, TruSeq Rapid v1 reagents	850–1000 K/mm²

Other Resources

- Illumina Support Pages
- Relevant Support Pages
  - PhiX Control v3 Support Pages
  - Preparing Libraries for Sequencing on the MiSeq
  - Denaturing and Diluting Libraries for the HiSeq and GAIIx
  - Sequencing Library qPCR Quantification Guide
- Training Videos
- Support Bulletins
  - Cluster Density Specifications for Illumina Sequencing Platforms
  - Library denaturation considerations for MiSeq V3 sequencing runs
  - Considerations for Library when Switching Between Illumina Sequencing Platforms
  - Five Things You Can Do to Improve the Quality and Quantity of Your Sequencing Data
- Sequencing Analysis Viewer (SAV) Guide & Download