Optimizing Cluster Densities

August 12, 2014 Call in number: Conference code:



IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.





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



- 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?



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



What is a Cluster - Thumbnails

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

MiSeq Thumbnail



HiSeq Thumbnail





- 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







- 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

şi	MiSeg Focused power, Speed and mplicity for targeted and small genome sequencing.	NextSe Flexible powe simplicity for genon	q 500 r. Speed and r everyday nics.	HiSe Production po efficiency fo geno	g 2500 wer. Power and rr large-scale mics.	HiSeg 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		



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

Sequencing Analysis Viewer



Data By Lane: Density –Total and %PF



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

Sequencing Analysis Viewer

Run Fo	lder: N	liSeq Rı	un												Bro	wse	Refresh
Analysis	Imagir	na Summa	ry Tile S	Status Trus	Seq Controls												
Run S	Sup																*
Level		ted To (G)	otal Yield	Yield Perfect (G)	Yield <=3 errors	Aligned (%)	% Perfect [Num Cycles]	% <=3 errors [Num Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q3	60				
<		4.6		3.5		99.14	76.5 [250]	94.4 [250]	1.03	153	116.6	89.6					
Rea		4.6		*		97.82	73.3 [250]	91.6 [250]	1.23	144	118.0	84.8					
Total	0.2	9.2		6		98.48	74.9	93.0	1.13	149	117.3	87.2					
Read	1																
Lane	Tiles	Density (K/mm2)	Cluster (%)	P (%	(M)	Rea	ndsPF (M) %>⊧	= Q30 Yield (G)	Cycles Err Rate	d Ali	gned En %)	ror Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	28	1084 +/- 26	89.62 +/-	0.34 0.082/	0.130 20.43	18.31	89.6	4.6	250	99.1 -	⊧/- 0.0 1.03	3 +/- 0.02	0.11 +/- 0.00	0.15 +/- 0.00	0.21 +/- 0.00	153 +/- 1 2	116.6 +/- 0.7
Read	2																
Lane	Tiles	Density (K/mm2)	Cluster (%)	PF Phas/Pi	rephas Reads	Rea	ndsPF (M) %>	= Q30 Yield (G)	Cycles Err Rate	d Ali	gned En %)	ror Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20

• %PF can vary

28

• Higher the %PF the better

1084 +/- 26 89.62 +/- 0.34 0.087 / 0.123

20.43

Very high density or overclustering will result in lower %PF

84.8

4.6

250

97.8 +/- 0.1

1.23 +/- 0.02 0.14 +/- 0.00 0.26 +/- 0.01 0.35 +/- 0.01

18.31

144 +/- 10

118.0 +/- 0.9



- 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





Over clustering – inconsistent cluster density



Over clustering – Low %Pass Filter



Over clustering - Drop in intensity after cycle 5



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





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



Normal/low density – MiSeq v2



Still white spots on a black background



Highest recommended density – MiSeq (v2)



No clear white spots - more of a haze



Over-clustered! – MiSeq (v2)



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





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





- 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)	\checkmark	qPCR	
Nextera DNA	\checkmark	Qubit	
Nextera XT		None	Built-in normalization step Final library is single stranded
Nextera Mate Pair	\checkmark	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	\checkmark	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)	\checkmark	qPCR	
TruSeq Targeted RNA Expression	\checkmark	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	\checkmark	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 is preferable to over overclustering

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



- 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



Resources – Training Videos

illumına [,]				Log in	to get person	alized account	information. Qu	uick Order 🐂 t Us MyIllumina	View Cart ∰ Tools ▼		
APPLICATIONS	SYSTEMS	INFORMATICS	CLINICAL	SERVICES	SCIENCE	SUPPORT	COMPANY	Search	Q		
Support » Training >	» Sequencin	9						8.98	🛿 Follow us: 📜		
Online Courses		Online Trai	ning flow T								
Real-Time PCR		TITLE	DESC	RIPTION				LENGTH			
Sequencing		Sequencing: I Technology	This c seque compl Langu	This course provides a general overview of the Illumina sequencing workflow from extraction of nucleic acids to the completion of a sequencing run. Languages: Spanish, Chinese (Simplified), Japanese							
		Sequencing: Illumina's Dual Indexing Strategy			The goal of dual index sequencing is to increase the multiplex level of sequencing per lane, so that more samples can be sequenced on the same flow cell. Languages: Spanish, Chinese (Simplified), Japanese						
		(SAV) This course provides the basics on how to get started using Sequencing Analysis Viewer (SAV). By the end of this course, you will be able to: identify the purpose of SAV, list the steps to load data into SAV, and describe the tabs in SAV.							15 min		
		NextSeq 500: Your Install	Preparing fo	This c the No requir	ourse provide extSeq 500, a rements for in	oonents of and electrical	15 min				
		NextSeq 500: System Overview		By the of the techno of the	By the end of this course, you will be able to: identify features of the NextSeq 500 system, describe NextSeq 500 sequencing technology, and describe the simplified sequencing workflow of the NextSeq 500 system.						
		NextSeq 500: Preparing Runs with BaseSpace			By the end of this course, you will be able to identify the purpose of the BaseSpace Prep tab, describe the steps to enter NextSeq 500 run parameters in BaseSpace, and identify how to make a BaseSpace-prepared run available in NextSeq Control Software (NCS)						



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			Considerations for Library Migration Between Illumina Sequencing Platforms								
Illumina sequ	encing platform	n.	Oct 22 2013								
Keep your of recommende Exceeding the output. You of	d cluster density d cluster densit e recommendat an find cluster o	y within Illumina ies is very importa ion may have neg density specificatio	While each of Illumina's s sequencing data, there an All libraries constructed w Illumina sequencing platfo	equencing platforms and modes us re slight variations in the chemistry vith currently available Illumina sar orms.	use the same Illumina technology to obtain ry utilized on different sequencing platforms. ample prep kits are compatible with all						
Add PhiX a	Cluster Der	nsity Specificatio	ns for Illumina Sequer	ncing Platforms	ring between different platforms when working with						
provide erro	May 19 2014										
here for the	Cluster densit	v is an important fa	ctor in data quality and vie	d. The following table includes a lis	ist of						
recommend	optimal raw c	luster density specif	Seq flow cells are paired- id libraries before running on								
Verify you											
Bioanalyzer					mperature during on board						
library can t	Platform	tform Mode/Reagents Option x SCS 2.8 or higher, TruSeq Cluster v5 reagents reagents		Optimal Raw Cluster Density	rature of 60°C.						
Double che Illumina libr need the Tri always mak	GAIIx			700-800 K/mm²	orm deblocking at 60°C and for both of these steps.						
Follow bes denatured c concentratio	HiSeq	High output,	TruSeq v3 reagents	750K-850 K/mm ²	emistry, onboard cluster ter at approximately loaded on a GA or HiSeq Hig deration when migrating	٦					
sequencing	HiSeq	High output, HiSe u	eq v4 reagents (requires upgrade)	950-1050 K/mm ²							
-	HiSeq	Rapid run, Tru	Seg Rapid v1 reagents	850-1000 K/mm ²	:11						

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
 - <u>Considerations for Library when Switching Between Illumina Sequencing</u> <u>Platforms</u>
 - Five Things You Can Do to Improve the Quality and Quantity of Your Sequencing Data
 - Sequencing Analysis Viewer (SAV) Guide & Download