Genome Sequencer FLX System

Genome Sequencer FLX System

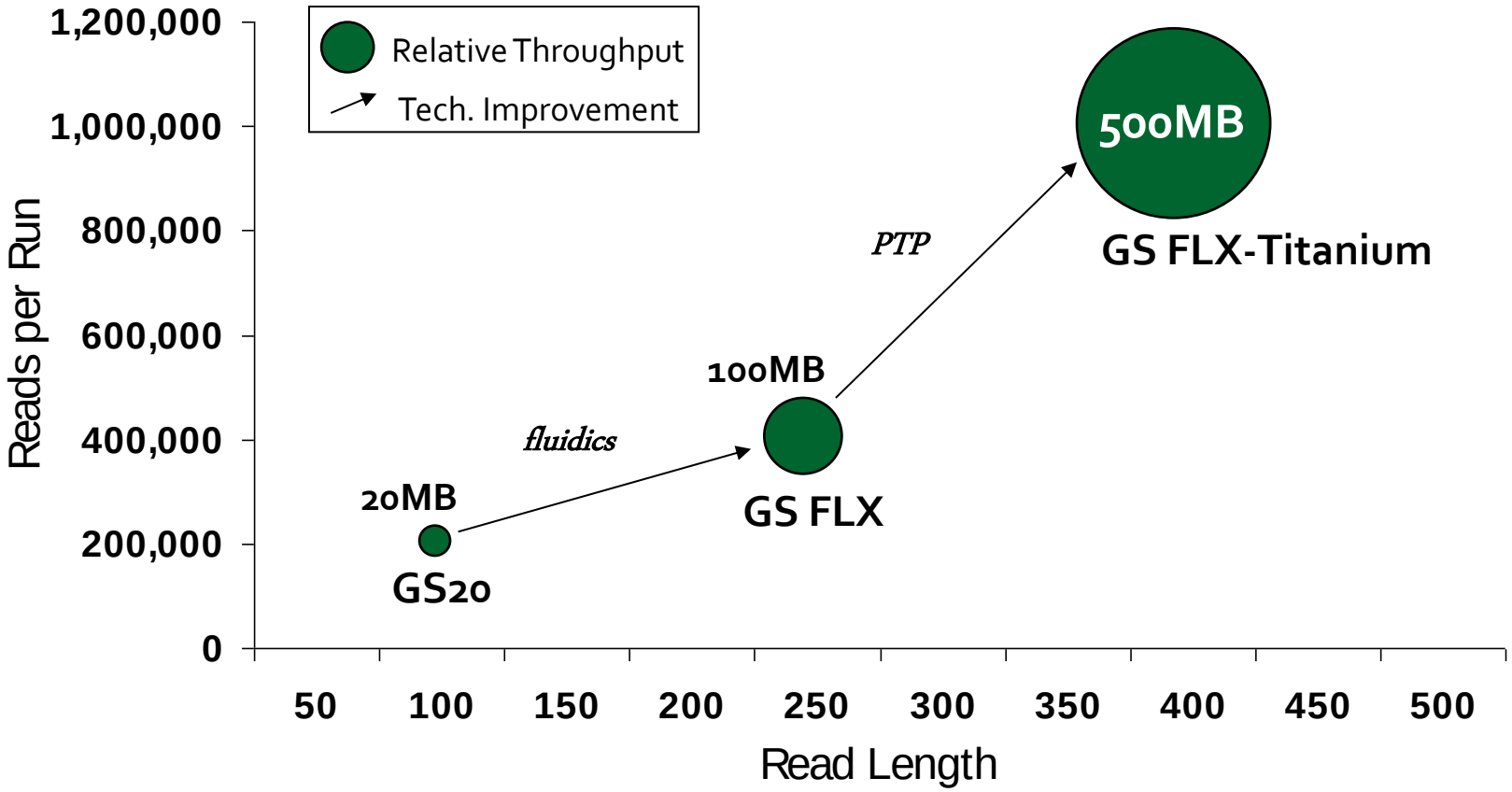
- Overview of the Genome Sequencer FLX System
- Genome Sequencer FLX System Performance
- Genome Sequencer FLX System Workflow
- Genome Sequencer FLX System Data
- Flexibility
- Applications
- The Future

GS Evolution

GS Evolution	Genome Sequencer 20	Genome Sequencer FLX	Genome Sequencer FLX Titanium
Read length	100 bases	> 200 bases	> 400 bp
# of clonal reads	>200,000	> 400,000	> 2,000,000
System throughput /8hr shift	20-30mb	100mb	0,5 - 1gb
Cost	$6,000-9,000.00	$3,000-9,000	$10,000 ?
Accuracy	99.99%	99.99%	99.99%

Evolution of 454 Sequencing

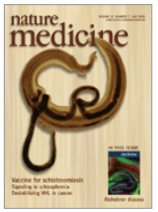
Increasing reads/run, read length, throughput

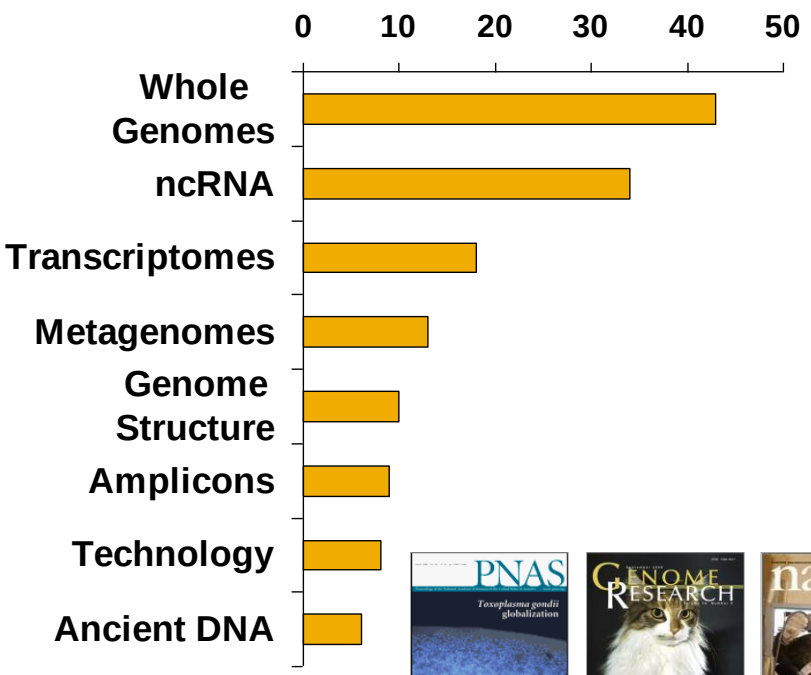


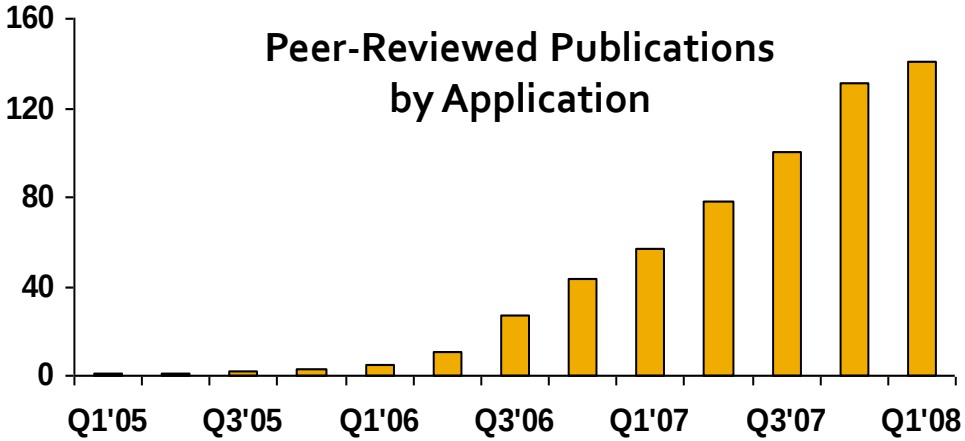
- **Obtain more comprehensive data from your genomic samples**
  - Generate 400,000 reads per run with 250-base reads at 99.5% accuracy
- **Reduce your cost per result**
  - Ultra-high throughput significantly reduces cost compared to traditional sequencing
  - Long, highly accurate reads and lack of bias deliver confident, quality results
  - Pool 12 samples per region, up to 192 samples per run, for low cost per sample
- **Expand your project capabilities**
  - Utilize long shot-gun reads or 3K Long-Tag Paired End reads.
- **Publish Faster**
  - Sequence 100 million high-quality filtered bases per 7.5 hour run and accelerate time to result with easy to use analysis tools

Customer Success with the GS FLX

Turning Sequencing into Publications





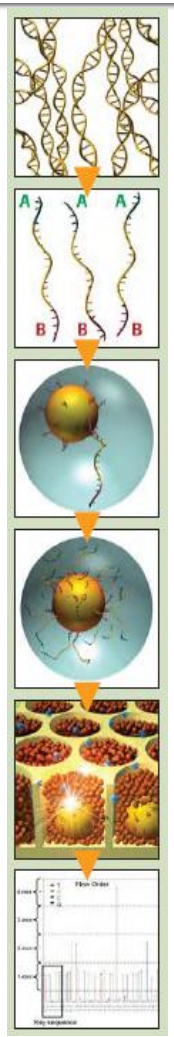


GS FLX System Performance

Genome Sequencer FLX System	
Throughput	100 million high-quality, filter-passed bases per run, 200 million bases per day (1 billion bases per day Titanium)
Run Time	7.5 hours
Read Length	250 bases (>400-base reads Titanium)
Reads per Run	400,000 high-quality reads
Single-Base Accuracy	>99.5% single-read accuracy over 250 bases
Sample Input Requirement	As little as 100 ng of DNA
Data	Trace data accepted by NCBI since 2005
Computing Requirements	Single Linux server
Robustness	No complex optics or lasers; long-life reagents

Sequencing Workflow

One fragment = One bead = One read



Sample Input and Fragmentation

*Generation of small **DNA fragments** via nebulization*

Library Preparation

*Ligation of A/B-Adaptors flanking **single-stranded DNA fragments***

One Fragment = One Bead

*Emulsification of beads and fragments in water-in-oil **microreactors***

emPCR (emulsion PCR) Amplification

***Clonal amplification** of fragments bound to beads in microreactors*

One Bead = One Read

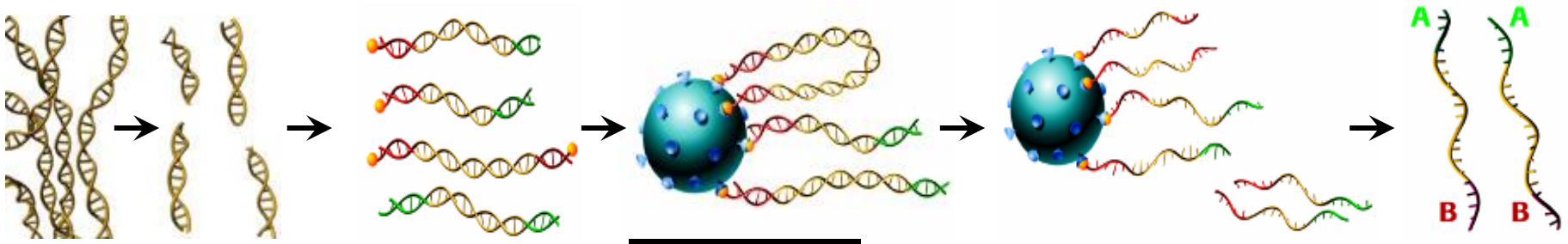
Sequencing by synthesis

Data Analysis

Flowgrams, basecalling, analysis tools

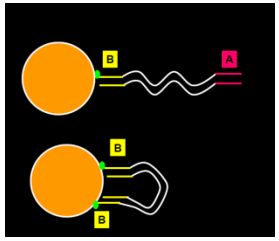
Sequencing Workflow

Library Preparation



Genome fragmented by nebulization

sstDNA library created with adaptors



A/B fragments selected using streptavidin-biotin purification

DNA library preparation

Hands on time 4.0 h
Total time 4.5 h

emPCR

Hands on time 4.0 h
Total time 8.0 h

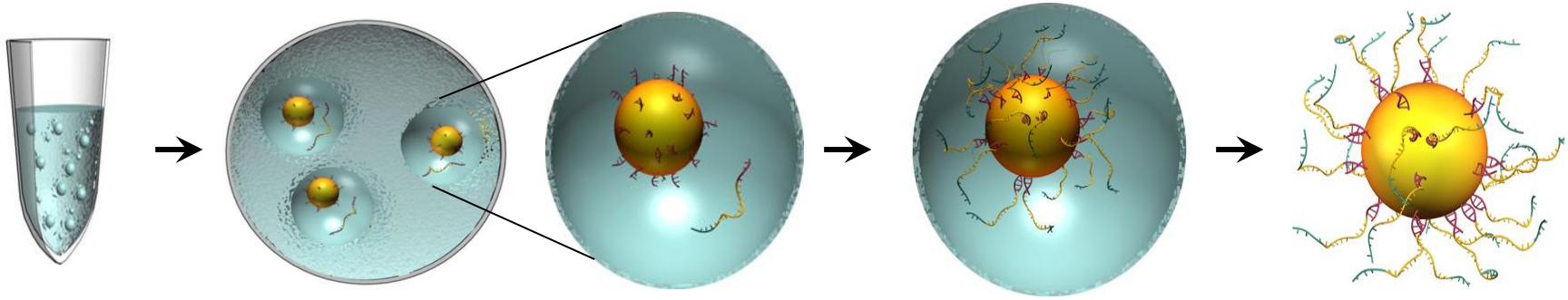
Sequencing

Hands on time 0.5 h
Total time 7.5 h

www.roche-applied-science.com

Sequencing Workflow

Emulsion PCR



Anneal sstDNA to an excess of DNA Capture Beads

Emulsify beads and PCR reagents in water-in-oil microreactors

Clonal amplification occurs inside microreactors

Break microreactors, enrich for DNA-positive beads

DNA library preparation

Hands on time 4.0 h
Total time 4.5 h

emPCR

Hands on time 4.0 h
Total time 8.0 h

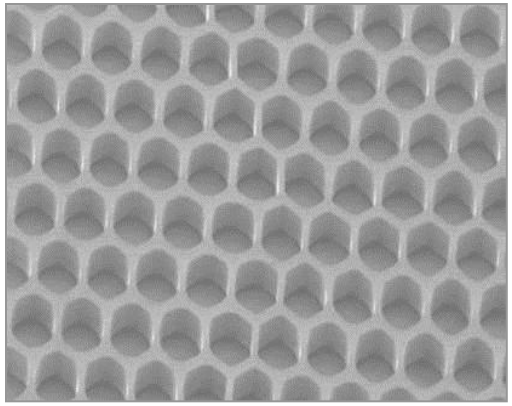
Sequencing

Hands on time 0.5 h
Total time 7.5 h

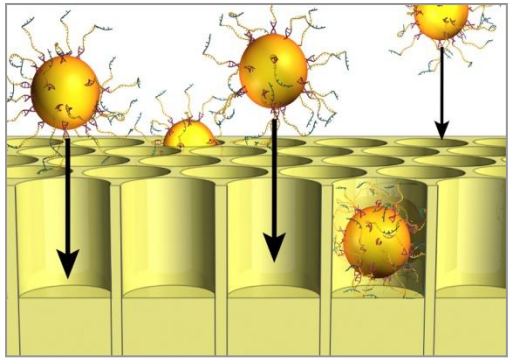
www.roche-applied-science.com

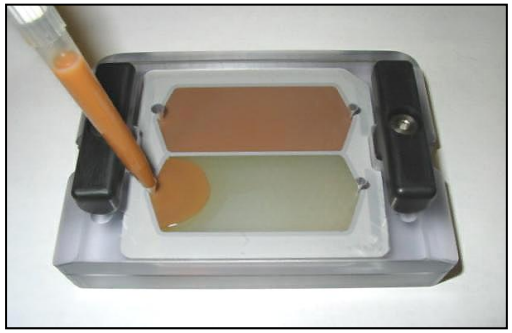
Sequencing Workflow

Loading of PicoTiterPlate (PTP) Device

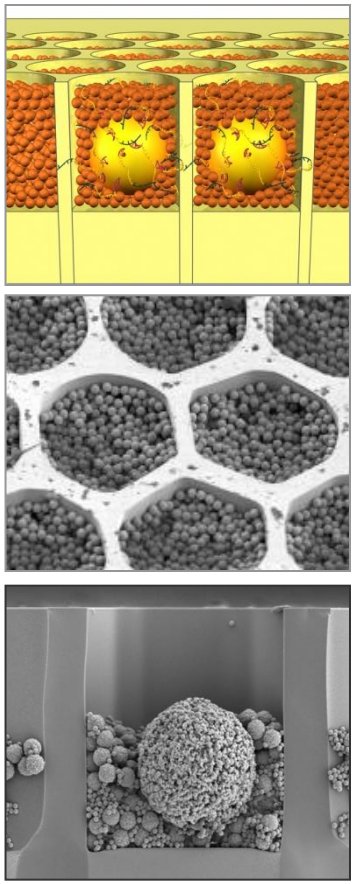


- 1.6 milion wells per PTP
- Well diameter: average of 44 µm
- Each well is only able to accept a single DNA bead.





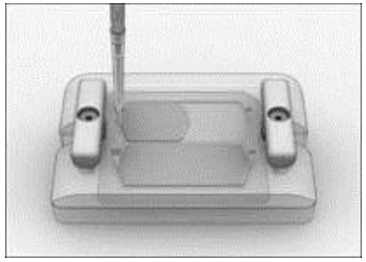
- DNA capture beads and enzyme beads deposited onto PTP



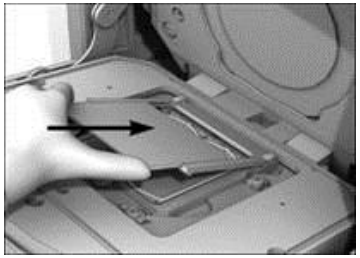
www.roche-applied-science.com

Sequencing Workflow

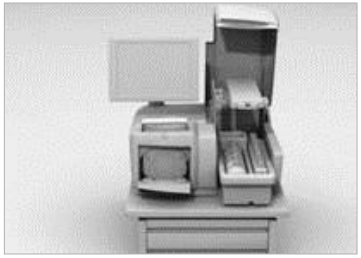
Instrument Loading



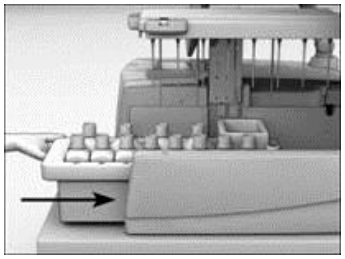
Load beads onto PicoTiterPlate device



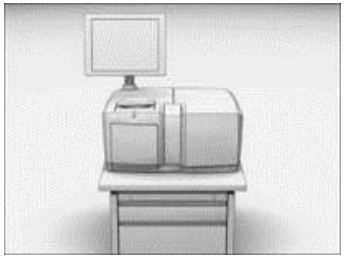
Load PicoTiterPlate device on instrument



Open instrument



Load sequencing reagents

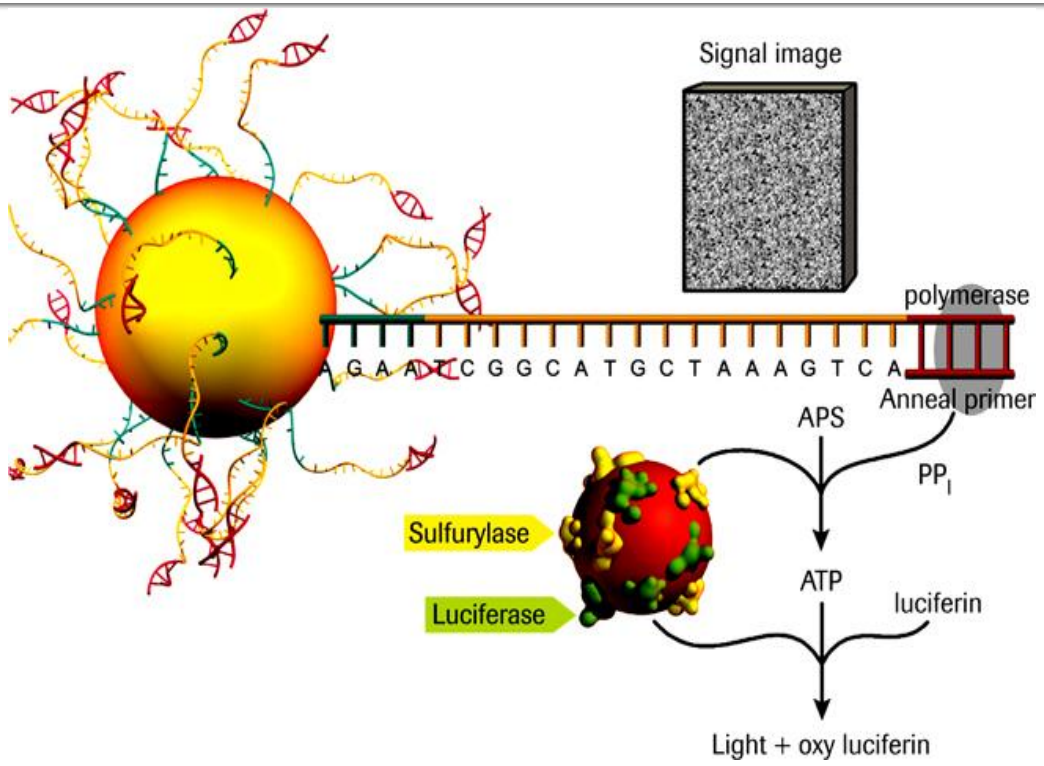


Press START for Sequencing!

www.roche-applied-science.com

Sequencing Workflow

Sequencing by Synthesis



- Bases (TACG) are flowed sequentially and always in the same order (100 times for a large GS FLX run) across the PicoTiterPlate device during a sequencing run.
- A nucleotide complementary to the template strand generates a light signal.
- The light signal is recorded by the CCD camera.
- The signal strength is proportional to the number of nucleotides incorporated.

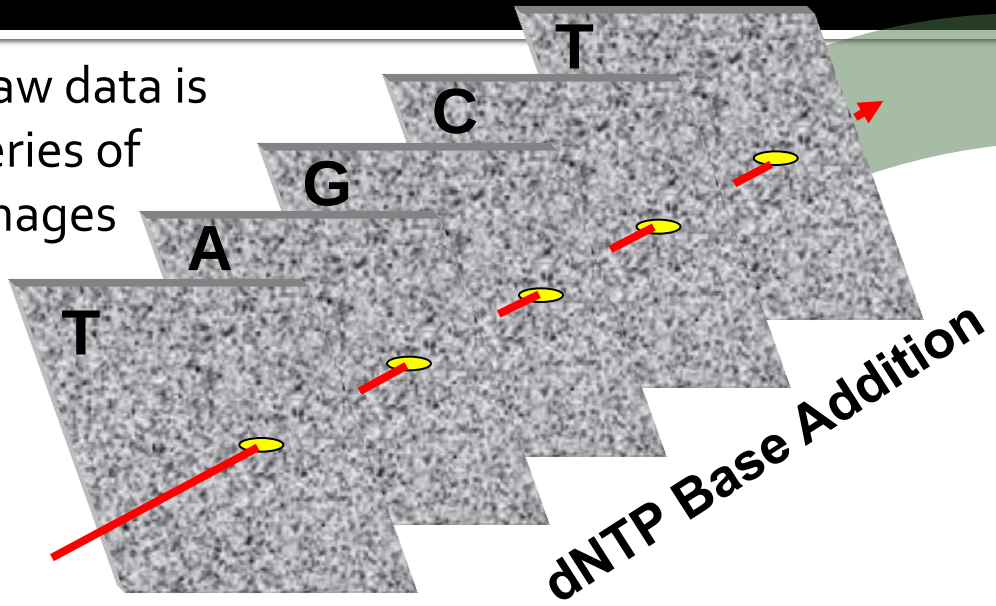
DNA library preparation		emPCR		Sequencing	
Hands on time	4.0 h	Hands on time	4.0 h	**Hands on time**	**0.5 h**
Total time	4.5 h	Total time	8.0 h	**Total time**	**7.5 h**

www.roche-applied-science.com

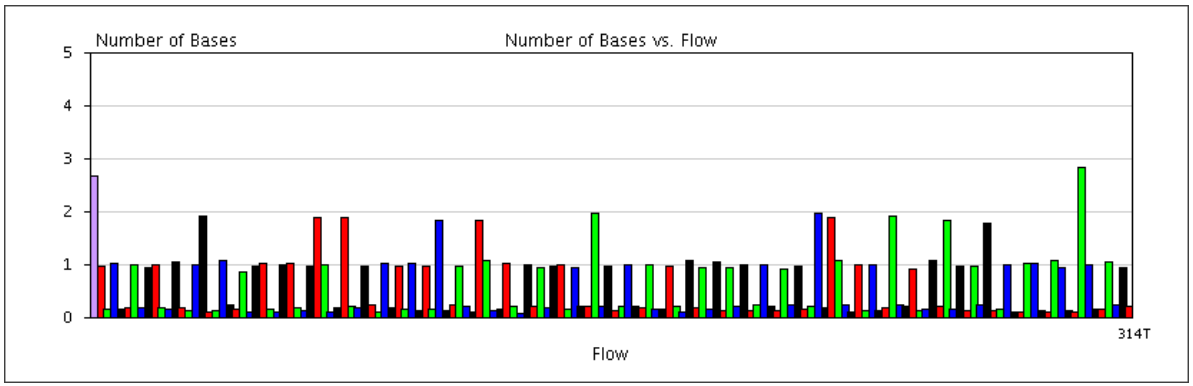
GS FLX Data

Image Processing Overview

1. Raw data is series of images



2. Each well's data extracted, quantized and normalized

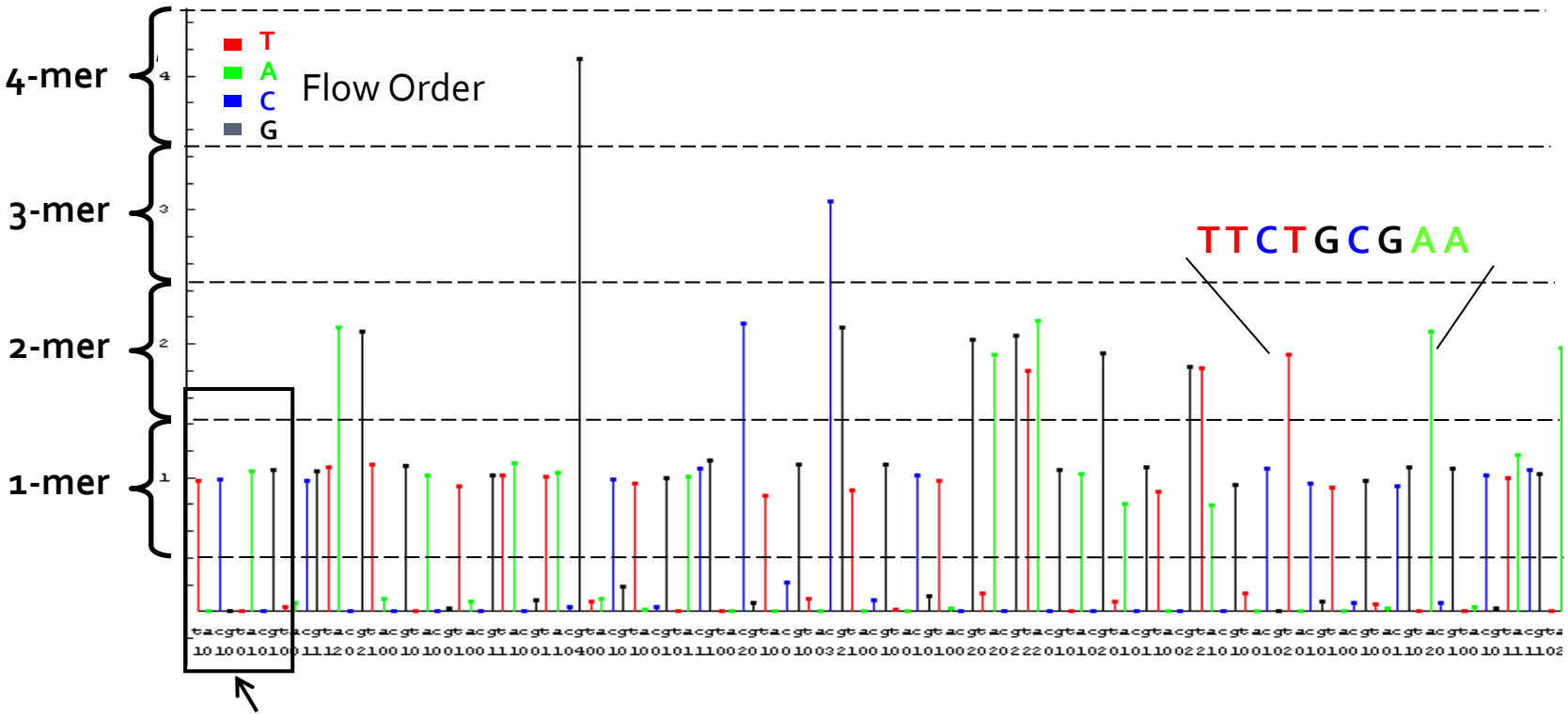


3. Read data converted into "flowgrams"

www.roche-applied-science.com

GS FLX Data

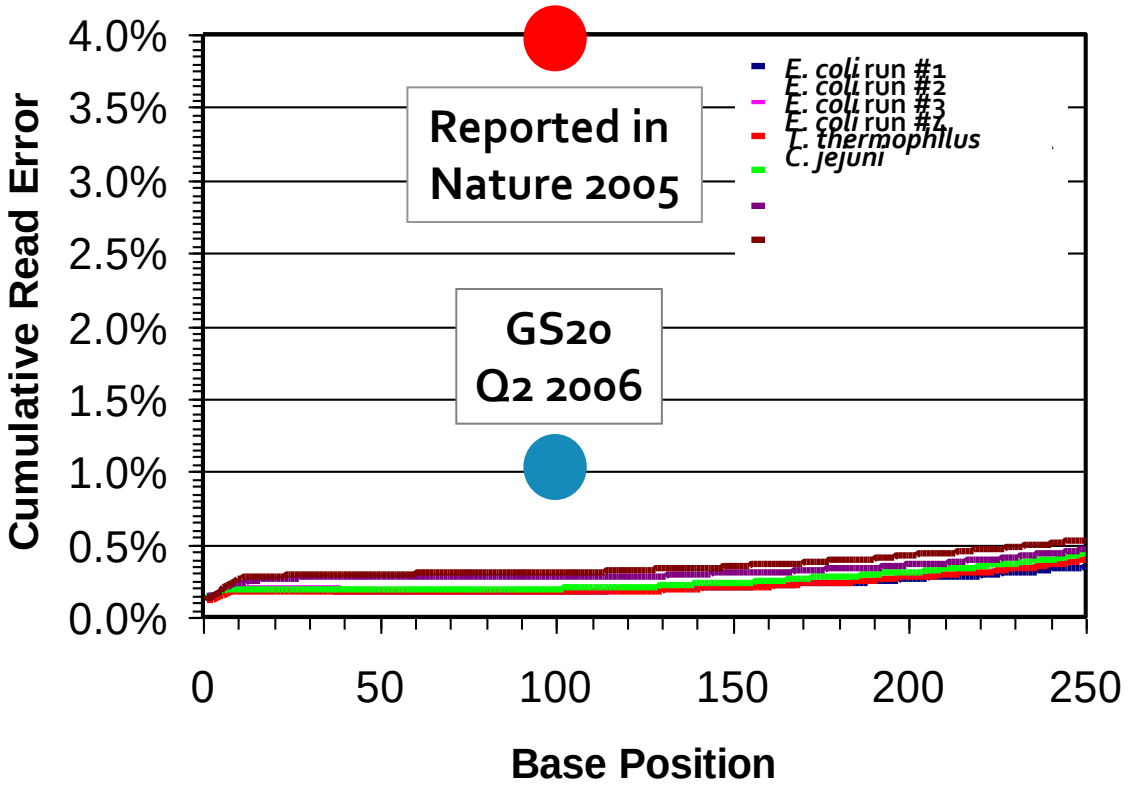
Flowgram Generation



www.roche-applied-science.com

GS FLX Data

Single-Read Accuracy: cumulative read error

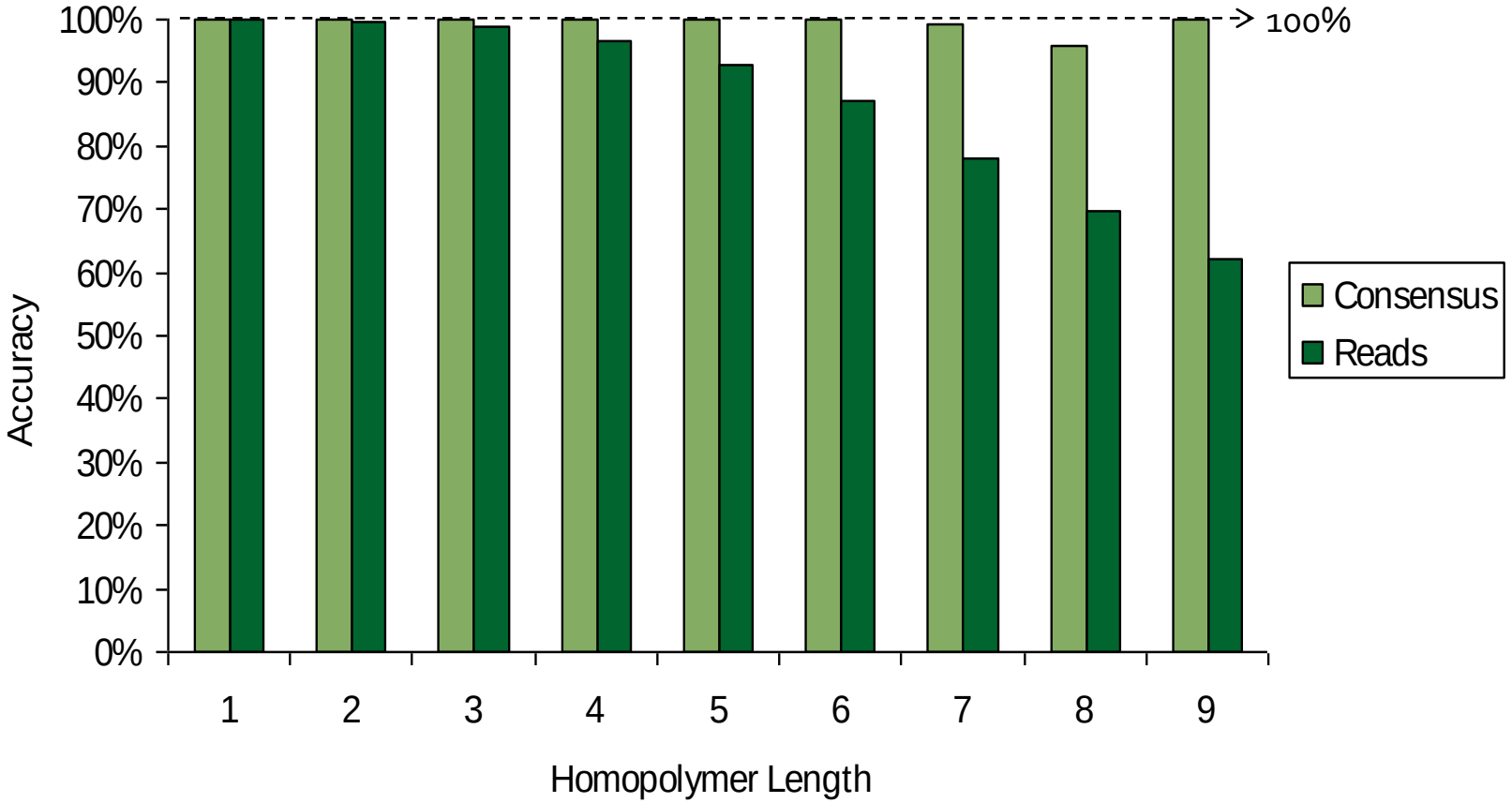


- Single-Read Accuracy > 99.5% (including all indel errors)
- Ignoring indels (homopolymers) the individual read accuracy is ~99.996%

www.roche-applied-science.com

GS FLX Data

Homopolymers in E. coli: Individual Reads versus Consensus



www.roche-applied-science.com

GS FLX System Flexibility

Sequencing Kit Formats

Achieve optimization of cost, coverage, and time to result for your specific research needs

Sequencing Kit	Reads per Run	Average Read Length	Run Time	Optimal For
GS LR70	400,000	250 bases	7.5 hours	High-throughput / multi-sample applications
GS LR25	70,000	250 bases	7.5 hours	Small projects, application development, feasibility testing
GS SR70	400,000	100 bases	4.5 hours	Short-read applications

GS FLX System Flexibility

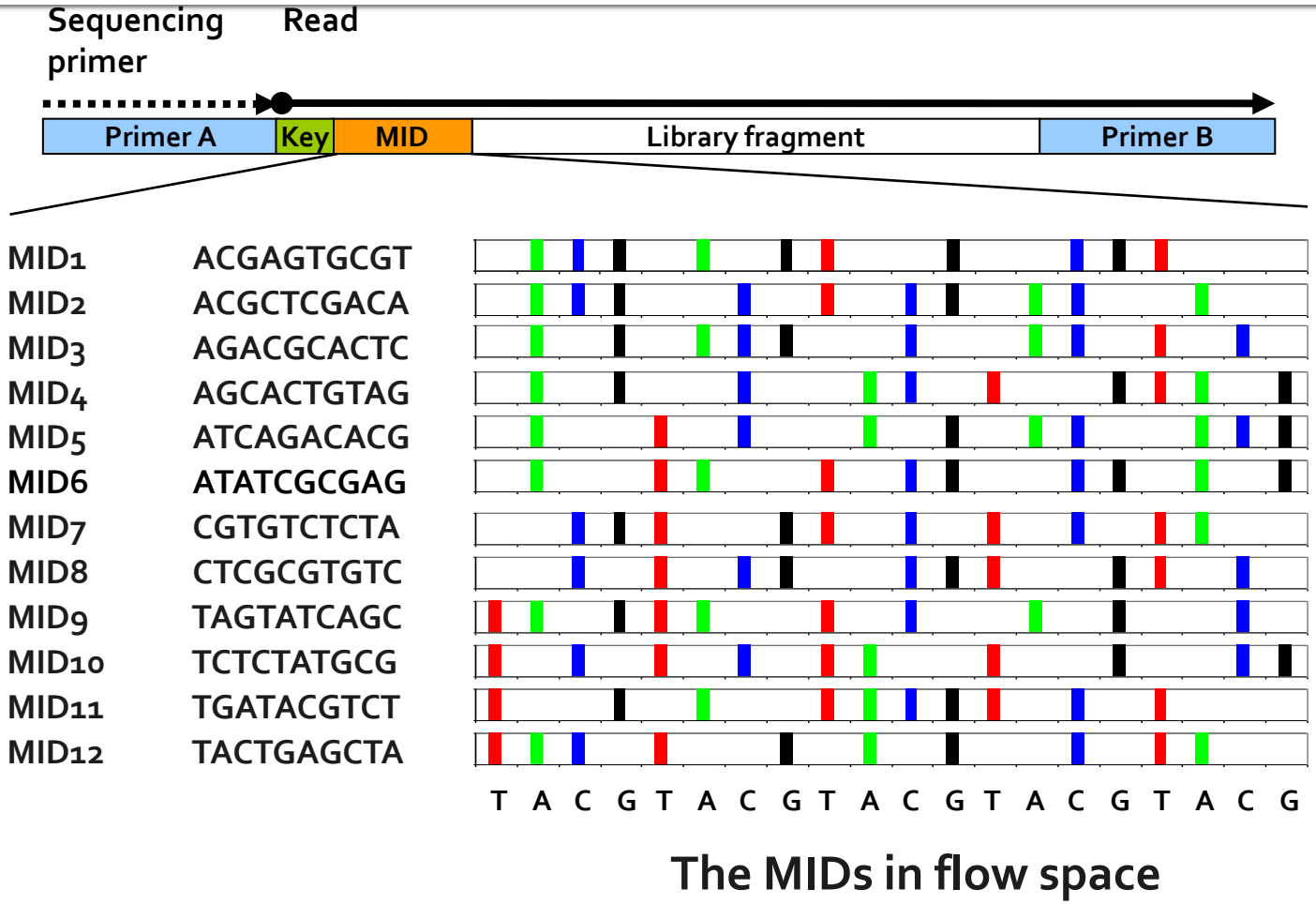
Multiplexing: Physical Separation of Samples

Tailor sample number and throughput to your specific needs

Gasket Format (regions)	GS Sequencing Kit Compatibility	Number of Reads per Region
1	GS LR70 and GS SR70	400,000
	GS LR25	70,000
2	GS LR70 and GS SR70	200,000
4	GS LR70 and GS SR70	70,000
	GS LR25	12,000
8	GS LR70 and GS SR70	30,000
16	GS LR70 and GS SR70	12,000

GS FLX System Flexibility

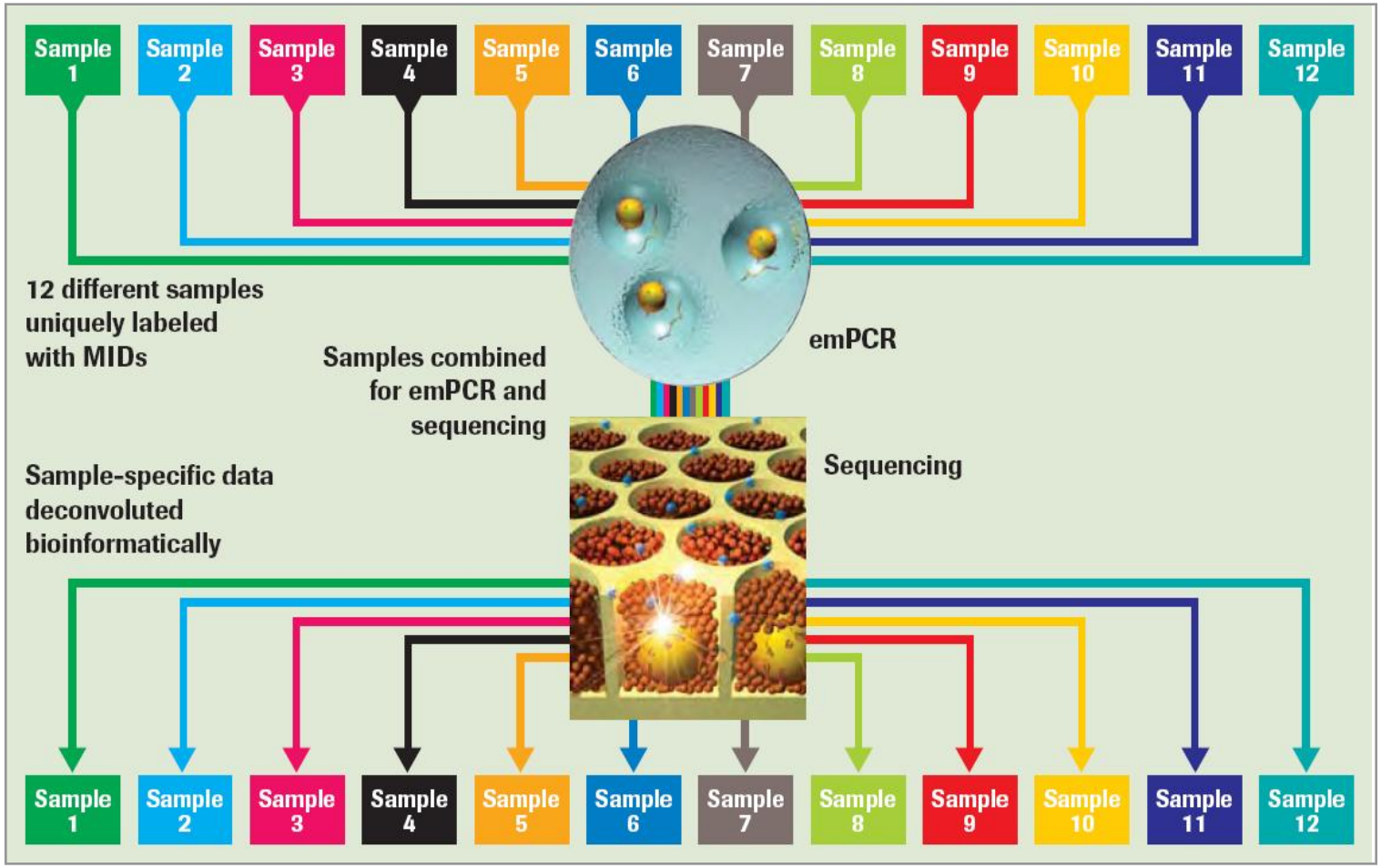
Multiplex Identifiers (MIDs)



www.roche-applied-science.com

GS FLX System Flexibility

Multiplex Identifiers (MIDs)

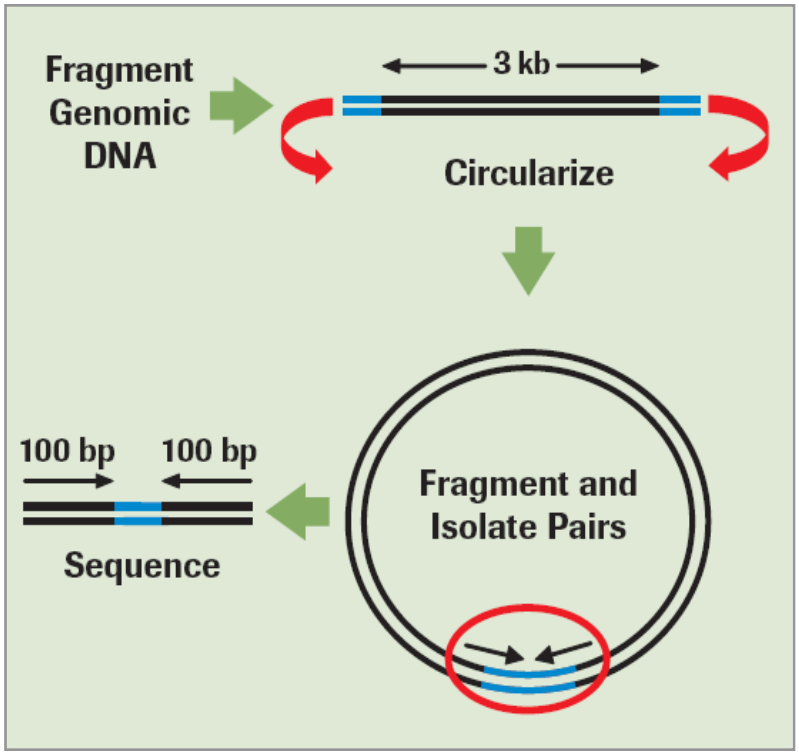


www.roche-applied-science.com

GS FLX System Flexibility

3K Long-Tag Paired End Reads

- New Protocol for sequencing of 250 bases from each end of a 3,000 base span on a single read
- Use for
  - De novo assemblies of complex genomes
  - Identify structural variations at high resolution



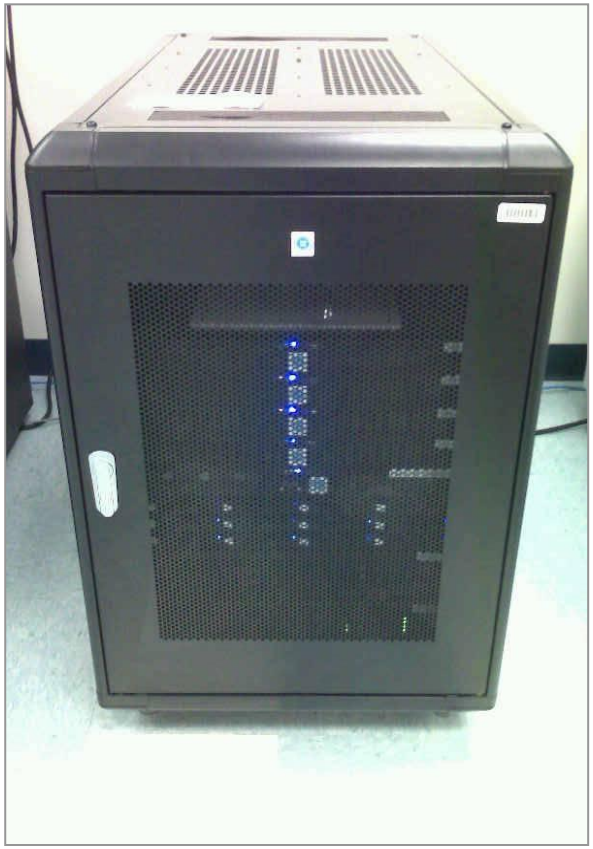
www.roche-applied-science.com

Data Analysis Server for Titanium Cluster

Fast data analysis with a one box solution

For an out-of-the-box solution we have qualified a provider who will supply an integrated system. This purchase is made through Roche and is delivered as a one box solution.

Server specifications are available for a do-it-yourself option.



GS FLX Data Analysis Tools

Analyze data without a complex IT infrastructure

Obtain results from your sequence data quickly and affordably with the powerful suite of analysis tools provided with the Genome Sequencer FLX System

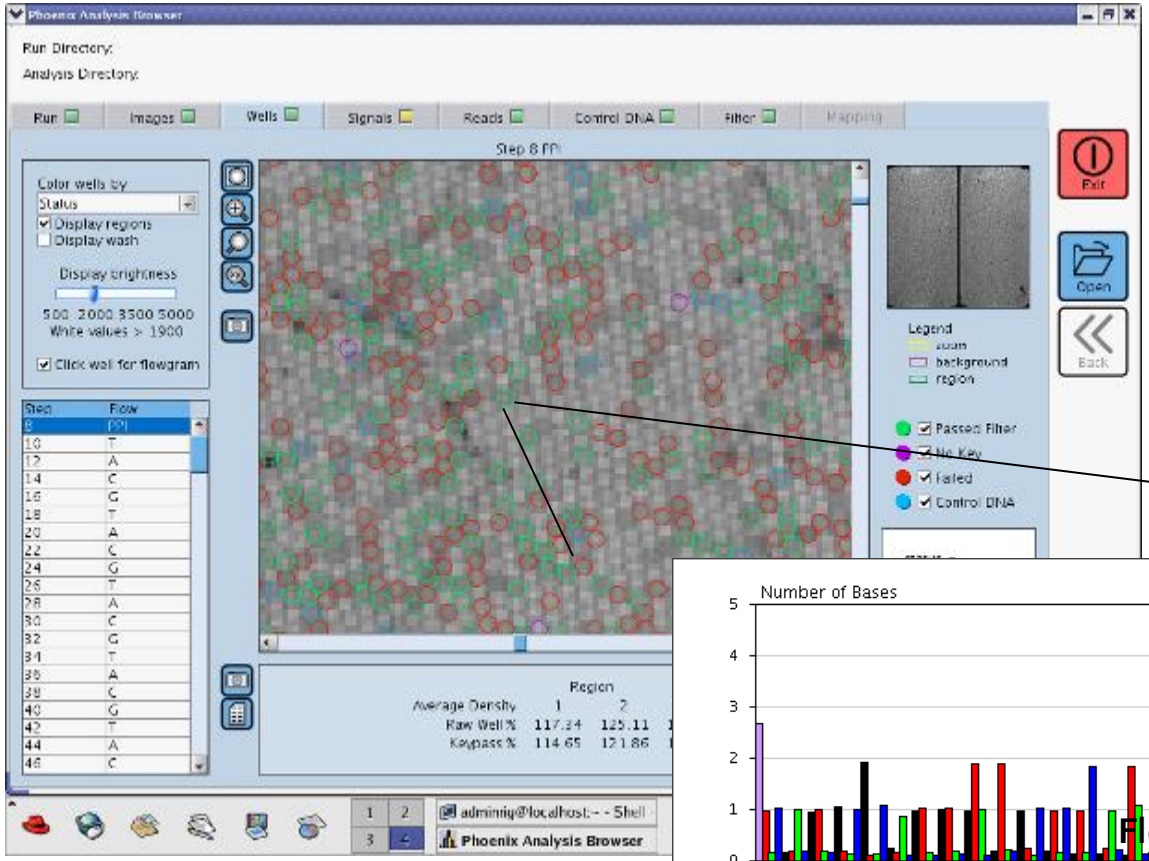
Data Management, Processing, and Analysis Features	Supplied with the GS FLX System
User-friendly GUI	Yes
Integration with LIMS	Yes
Bar-code scanning	Yes
PHRED quality scores*	Yes
Trace data accepted by GenBank	Yes
Hardware	
Single-node computing	Yes
Manageable data storage	Yes
Software Tools	
De novo assembly tool	Yes
Mapping tool	Yes
Amplicon analysis tool	Yes
Third-party LIMS	Yes
Third-party options	Yes

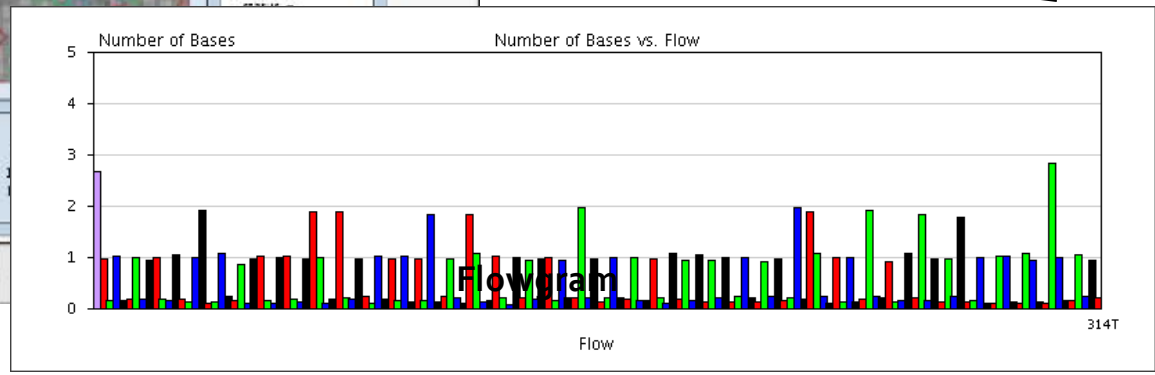
www.roche-applied-science.com

* PHRED equivalent score has been developed in conjunction with The Broad Institute.

GS FLX Data Analysis Tools

GS Run Browser Software





www.roche-applied-science.com

GS FLX Data Analysis Tools

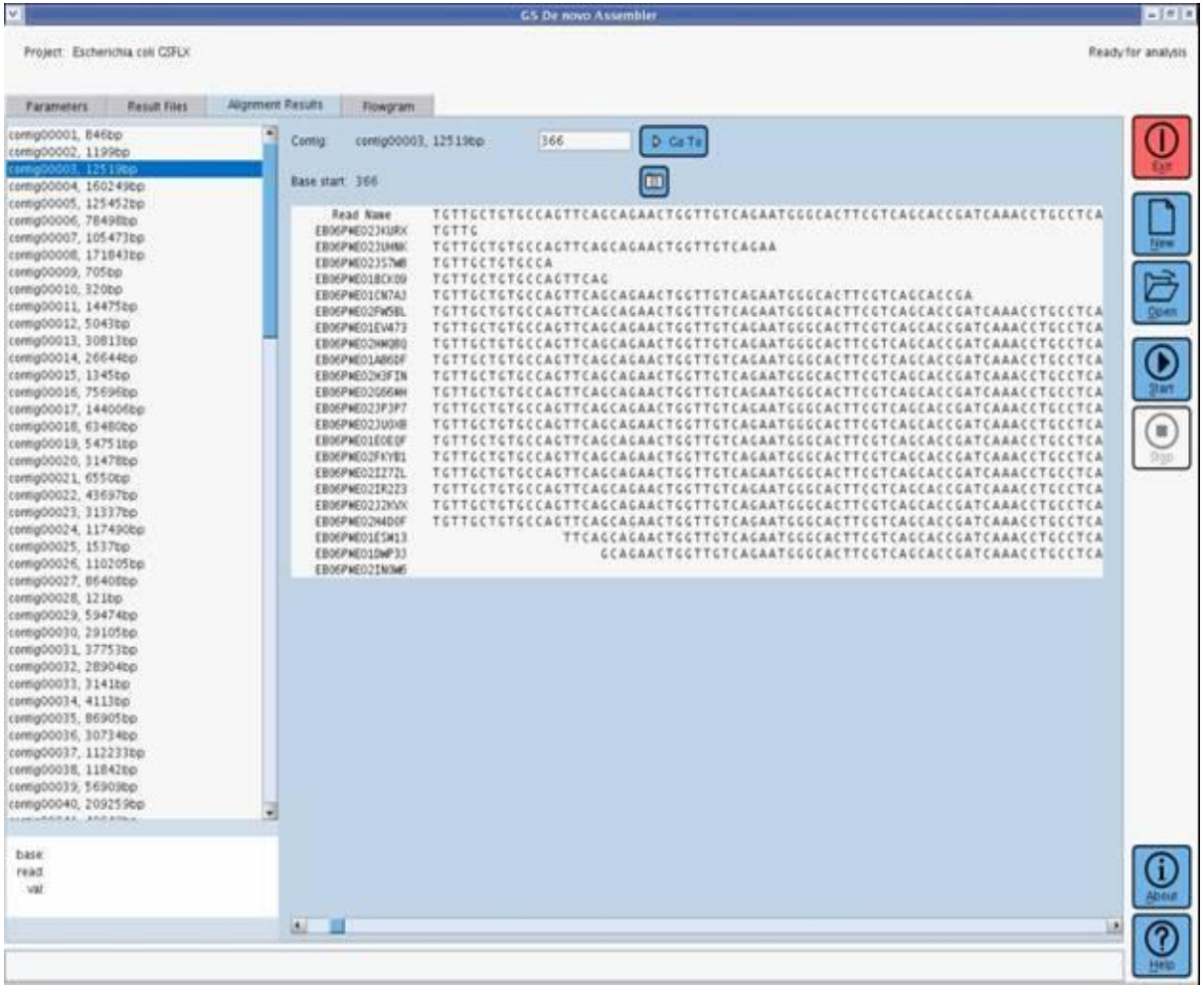
GS de novo Assembler Software

- Perform whole genome shotgun assembly of genomes with or without paired-end data.
- Order contigs into scaffolds using available paired-end reads.
- Assemble genomes up to 50 megabases in size.
- Select the read files you want to assemble by browsing in the GUI.
- View the assembly and browse the individual reads that make up the multiple

Example Data: *de novo* sequencing and assembly of complete genomes				
	E. coli K-12	*C. jejuni*	*T. thermophilus*	*S. cerevisiae*
Genome Size (bases)	4,639,675	1,641,481	2,127,575	12,495,682
% GC content	50	30	69	38
Oversampling	20x	20x	20x	21x
Number of Contigs	113	33	65	488
Number of Scaffolds	11	4	7	85
Total Coverage	97.62%	97.54%	98.04%	93.09%
Consensus Accuracy	99.999%	99.998%	99.998%	99.979%

GS FLX Data Analysis Tools

GS de novo Assembler Software



GS FLX Data Analysis Tools

GS Reference Mapper Software

- Map reads to any reference genome and generate a consensus sequence.
- Easily view all differences compared to the reference sequence with automatic output to separate files.
  - Insertions- inserted blocks of up to 50 bases
  - Deletions- deleted blocks of up to 50 bases
  - SNPs
- Quickly identify high-confidence differences compared to the reference genome, which are singled out into a separate file.
- Compare genomes up to three gigabases in size.
- Choose files by browsing in the GUI.

Example Data: Whole genome resequencing and mutation detection				
	E. coli K-12	*C. jejuni*	*T. thermophilus*	*S. cerevisiae*
Genome Size (bases)	4,639,675	1,641,481	2,127,575	12,495,682
% GC content	50	30	69	38
Number of Runs	1	½	½	½
Number of Reads Mapped	397,418	169,560	196,557	176,576
Total Coverage	98.52%	98.43%	99.09%	97.29%
Consensus Accuracy	100.000%	99.999%	99.998%	99.998%

GS FLX Data Analysis Tools

GS Amplicon Variant Analyzer Software

- Automatically compute the alignment of reads from amplicon-based samples against a reference sequence.
- Quickly identify variants and their respective frequencies in large pools of data
  - Screen for known variants
  - Discover unknown variants
  - Identify haplotypes
- Detect low-frequency (<1%) variants

The GS Amplicon Variant Analyzer Software can also be used for the following studies:

Resequencing

- Analyze disease-associated regions
- Discover SNPs, insertions, and deletions on a population level
- Detect rare somatic mutations
- Perform viral subtyping

Epigenetics

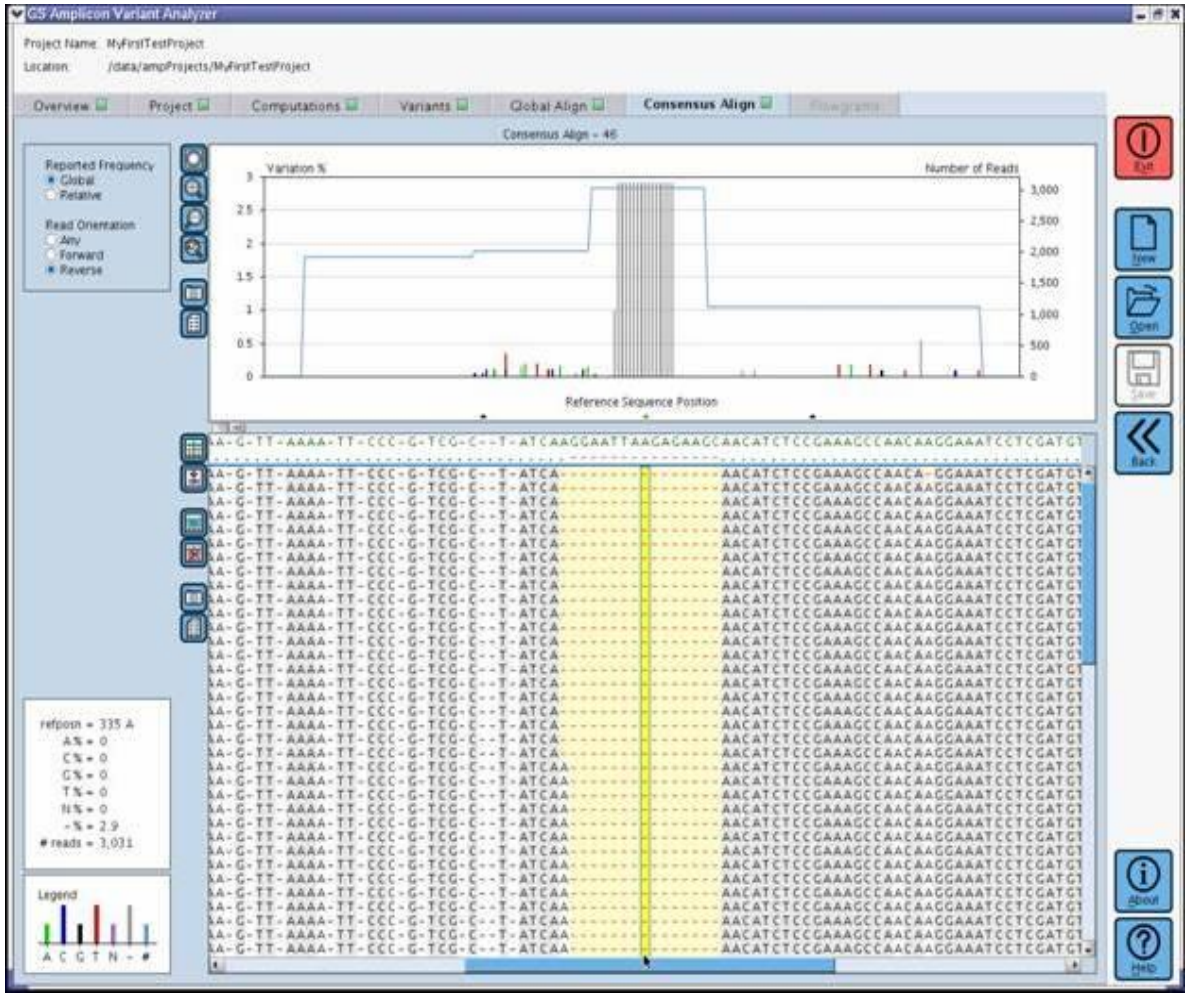
- Analyze DNA methylation patterns

Metagenomics and genetic diversity

- Sequence 16S RNA

GS FLX Data Analysis Tools

GS Amplicon Variant Analyzer Software



Sequencing Technologies Analysis

- The apparent price per base advantage of micro-read teachnologies in re-sequencing applications shrinks dramatically after adjusting for
  - Loss of raw reads when mapped against the reference
  - Additional over-sequencing required due to lower single read accuracy
  - Fully loaded costs (including required bioinformatics labor and IT maintenance)
- No amount of additional over-sequencing can provide equivalent quality to 454 Sequencing due to micro-read technologies'
  - Shorter read length
  - Greater bias
  - Limitation to SNP identification, versus detection of indels (large and small) and larger scale genome variation
- Price advantage per equivalent base will decisively shift to 454 Sequencing with the XLR HD kit format, which will run on the current GS FLX instrument

Total Cost Matters

The Real Cost per Run

When calculating the real cost per result, all cost must be considered

- Reagents - Labor and bioinformatics - Instrument cost

- Using few long reads compared to many micro-reads the bioinformatics requirement per run is much lower with 454 Sequencing (for the average user, 3X more informatics time is required per Illumina run)
- The Instrument cost must be considered when considering the total cost of the result produced by that instrument

Site Preparation

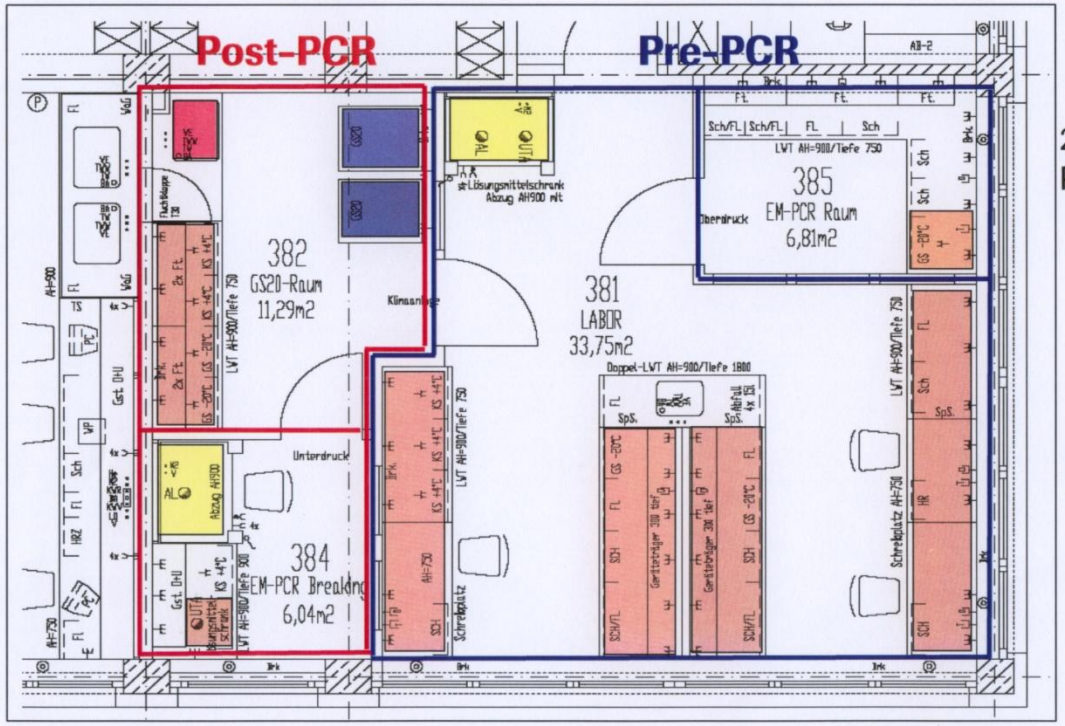
GS FLX Site Preparation

Example Lab Set Up ("Usual" Throughput)

Roche

4. Instrument

3. Amplicon Handling



2. emPCR Preparation

1. Library Preparation

www.roche-applied-science.com

Equipment

1. Bioanalyzer or Nanodrop
2. Nebulizator
3. RT-PCR (LightCycler)
4. PCR
5. Laminar flow (P2)
6. Mixer for emulsion prep.
7. N2
8. Centrifuge for plates (specific model)

Total cost ~500.000