MiSeq™ Personal Sequencing System

Sequencing made accessible

Brent Anderson, MS
Regional Genomics Specialist
A Sequencer for Every Need. Every Budget. Every Lab.

Redefining the trajectory of sequencing.


Two proven technologies. One powerful platform.

The most widely cited platform, now at half the price.

My Samples. My Study. MiSeq.

HiSeq 2000  HiSeq 1000  HiScanSQ  GA\textsubscript{IIx}  MiSeq
Unparalleled Publications Ramp

Fastest NGS publication rate, >1850 peer reviewed publications

Cumulative Illumina Peer Reviewed Sequencing Publications
From HiSeq to MiSeq

HiSeq:
- High yield
- High throughput
- High data quality

MiSeq:
- Fast run time
- Low cost device
- Same data quality
  (lower table shows equivalent runs)

<table>
<thead>
<tr>
<th></th>
<th>HiSeq</th>
<th>MiSeq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads</td>
<td>2 x 100</td>
<td>2 x 150</td>
</tr>
<tr>
<td>Yield</td>
<td>600 Gb</td>
<td>&gt;1.5 Gb</td>
</tr>
<tr>
<td>Run time</td>
<td>10 days</td>
<td>&lt;1 day</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HiSeq</th>
<th>MiSeq1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>603,000</td>
<td>562,000</td>
</tr>
<tr>
<td>PF (%)</td>
<td>94.0</td>
<td>93.9</td>
</tr>
<tr>
<td>Perfect reads (%)</td>
<td>83</td>
<td>83.8</td>
</tr>
<tr>
<td>Bases &gt;Q30 (%)</td>
<td>92.8</td>
<td>94.4</td>
</tr>
</tbody>
</table>

(>99.9% accuracy)
Spanning the Spectrum of Needs

Yield (per run) | 10Mb | 100Mb | 1Gb | 10Gb | 100Gb | 1Tb
---|---|---|---|---|---|---
Cost (per genome) | $50k | $50k | $20k | $5k | $1k
Time (per run) | 4 hr | 1 day | 10 days | 14 days

Applications
- **Targeted Tests**
- **Whole Genomes**
MiSeq System
MiSeq System
Proven Pedigree – Bench top Friendly

Single Instrument Workflow
Meet the MiSeq Instrument
My Samples. My Study. MiSeq

Included On-Instrument:
Cluster Generation (amplification)
Paired-End Fluidics
Computing for Primary and Secondary Analysis
MiSeq Instrument
Next-Gen Made Simple: Load & Go

DESIGNED FOR THE WAY YOU WANT TO WORK
Preloaded single use reagent cartridge
Positive consumables tracking
Auto flow cell positioning
Walkaway automation
MiSeq Workflow - Analyze
*On-Instrument Alignment & Variant Calling*

**SIMPLE & INTUITIVE DATA ANALYSIS**

- Analysis complete in <2 hours
- Base calling, quality values & variant reporting
- Output in industry standard formats
- Real time run monitoring
MiSeq Workflow - Run
More Innovation = Faster Results

SBS cycle times (2007-HiSeq)

- 31 min to 21 min (chemistry)
- 47 min to 25 min (imaging)
- 78 min to 46 min (total)

MiSeq Innovations

- Further chemistry improvements
  - Reagents, formulations, protocols
- Optimized fluidics
  - flow-rate, volumes, flow cell
- Scan less area
  - New flow cell
MiSeq Software Workflow

*Intuitive and Automated*

**Experimental Setup**

**Instrument Control**

**Data Analysis**
MiSeq Experimental Setup

Guided User Procedure

Offline wizard to help organize samples and experiment design
Creates set of instructions for the instrument on a per run basis
No knowledge of bioinformatics required
MiSeq Instrument Control Software

Fast and Efficient Run Setup

Oversized buttons make on-screen navigation easy

Starting a MiSeq run only takes a few clicks

Short video clips demonstrate proper loading of flow cell and reagents

MiSeq automatically performs bioinformatics analysis to produce pre-canned reports

No bioinformatics skills or intervention needed
MiSeq Workflow

*Prep, Run, Analyze*

- **Library Prep**: 1.5 hours (15 minutes hands-on)
- **Amplification thru Sequencing**: 4.5 hours (20 minutes hands-on)
- **Alignment and Variant Calling**: 2 hours (fully automated)

Sample-to-Answer in as Little as Eight Hours*

*1x36bp run – 3 hr sequencing*
Epicentre Nextera Technology for Library Prep

Single Tube, Rapid Library Prep

SIMPLE, FAST LIBRARY PREP IN LESS THAN 2 HOURS
Closed tube DNA fragmentation
Transposon-mediated library preparation
Ultra-low input requirements (50 ng)

ENABLES A RANGE OF CE AND NGS APPLICATIONS

VALIDATED BY LEADING RESEARCHERS
MiSeq System Applications
A New Era In Amplicon Sequencing

WHAT IF IN A SINGLE WEEK YOU COULD...
Sequence >150,000 amplicons
Sequence >384 samples
Automated analysis to call variants for all samples

Discover
• Variants including SNPs, SV and somatic mutations

Validate
• Follow up after whole genome and targeted resequencing, and genome wide association studies

Screen
• Population genetics
• Large-scale disease studies

Sequence >150,000 amplicons
Sequence >384 samples
Automated analysis to call variants for all samples
TruSeq Custom Amplicon Sequencing

Highly multiplexed sample and target amplicon resequencing for MiSeq

UNPRECEDENTED AMPLICON AND SAMPLE MULTIPLEXING

Easily prepare 1-192 samples targeting 48-384 amplicons

REVOLUTIONARY WORKFLOW

Go from DNA to data in 2 days, including analysis report
Only <8 hrs from DNA to MiSeq-ready library
Up to 384 amplicons per sample; 96 samples per plate
Integrated barcodes for easy sample indexing

START QUICKLY – EASILY SCALE

Use standard lab equipment, no microfluidics needed

POWERFUL CUSTOMIZATION

Easy to use online design tool to personalize content
Rapid turnaround of designs and oligos
MiSeq - Comparison to CE Sequencing

Example: TruSeq Custom Amplicon with 96 Samples x 96 Targets

---

**CE**

Dispense, PCR, Cleanup

- 96 plates

Quant

- 96 plates

Dye-terminator sequencing, cleanup

- 3730xl
- 2-3 weeks
- ~$3/target

---

**TruSeq Amplicon**

One plate of gDNA+oligos for all 96 amplicons

- 96-well plate

Assay Biochemistry

- **TruSeq Custom Amplicon**
  - Highly multiplexed
  - Integrated sample barcoding
  - Standard lab equipment

Pool Libraries and Sequence

- MiSeq
- 1 day
- <$0.75/target
MiSeq Applications
Application Flexibility for Every Lab

CAPILLARY ELECTROPHORESIS AND NEXT-GEN APPLICATIONS ON A SINGLE INSTRUMENT

FLEXIBLE AND SCALABLE AMPLICON SEQUENCING ASSAYS

SEQUENCE BACTERIAL OR VIRAL GENOMES IN ANY LAB

MAKE SCREENING ROUTINE
Infectious diseases
Vaccine QC
Clone checking
Library QC
Tissue typing

CHIP-SEQ OR SMALL RNA SEQUENCING
MiSeq Applications
Amplicon Sequencing

- Multiplexed amplicon sequencing of KRAS and BRAF in a single workday
- 165,000x average coverage of 4 amplicons
- All basecalls concordant with previous whole genome sequencing data
MiSeq Amplicon Workflow

- An example amplicon dataset (KRAS/BRAF single exons) was produced from a total of 6 samples:
  - NA18507 human gDNA control
  - DNA extracted from 5 FFPE tissue samples (tumour and normal samples)
  - Each sample was amplified out of the source DNA using locus specific primers also containing Illumina sequencing adaptors, the products were then pooled
- Clustering was followed by automatically by 2x77 bp sequencing - completely covering each amplicon
Preparing amplicons from FFPE sample and sequencing on the MiSeq is fast and straightforward.

- **C→G Variant detected in 27.5% of reads in FFPE Rectal tumor sample**
  - Same variant only detected in 1.1% of reads in normal rectal FFPE sample

**Data Quality**
- Raw read accuracy 99.8%
- Total Base calls >Q30, 94%

<table>
<thead>
<tr>
<th>Sample</th>
<th>Coverage</th>
<th>Noise (%)</th>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA18507 Control</td>
<td>179,630</td>
<td>0.21</td>
<td>None</td>
</tr>
<tr>
<td>Rectal Tumor</td>
<td>177,392</td>
<td>0.24</td>
<td>27.5% C→G</td>
</tr>
<tr>
<td>Rectal Normal</td>
<td>178,687</td>
<td>0.27</td>
<td>1.1% C→G</td>
</tr>
<tr>
<td>Gastric Tumor</td>
<td>176,530</td>
<td>0.23</td>
<td>None</td>
</tr>
<tr>
<td>Gastric Normal</td>
<td>161,856</td>
<td>0.28</td>
<td>None</td>
</tr>
<tr>
<td>Average</td>
<td>170,983</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>
MiSeq Applications
Small Genome Resequencing

- Resequencing of 5.2Mb *B. cereus* in a single workday
- 5.4 million reads yielded 175Mb of data which aligned to ATCC10987 with mismatch rate of 0.06%
- >99% of genome with average coverage of 30x
MiSeq Applications

Small Genome de novo Sequencing

- Economical Small Genome Sequencing of 2.8Mb MRSA
- Yielded *de novo* 2.78Mb assembly with N50 of 72.9kb and max contig of 247kb
- Reads aligned to 2.8 Mb EMRSA15 reference with average mismatch rate of 0.1%

*Source: phil.cdc.gov*
MiSeq Applications

Plasmid Sequencing

- Sequencing 8 kb pET11a-PRO217 plasmid and four mutated versions in a single workday
- De novo assembly yielded single 8 kb contig for each plasmid
- 100% concordance with previous Sanger sequencing results confirming mutations
MiSeq Instrument
Next-Gen Sequencing Just Got Personal

- Input
  - Nanogram amounts of nucleic acid

- Library Preparation
  - As short as 90 minutes

- Output
  - >1.5Gb

- Read Length
  - Up to 2x150bp

- Raw Accuracy
  - >85% bases >Q30 at 2x150bp

- Sample-to-Answer
  - As fast as a single working day

- Paired End Capability
  - On instrument

- Amplification Capability
  - On instrument in 1 hour

- Computer Hardware/Data Analysis
  - On instrument

- Instrument Price
  - $125K (all inclusive)

- Price per run
  - $600 - $950

- Price per Mb
  - $0.50
MiSeq
*The Next-Step in Next-Gen*

Proven SBS reversible terminator chemistry
No incremental sample prep equipment:
Revolutionary workflow – speed and simplicity
New era of user experience
Unmatched personal NGS output & speed

Questions?
Questions
MiSeq *E. coli* K12 MG1655 vs. Ion Torrent *E. coli* (various strains)
Analysis Summary

- **Illumina MiSeq**
  - *E. Coli* K12 MG1655 (1 run)

- **Three sets of Ion Torrent data have recently become publicly available for performance analysis**
  - *E. Coli* K-12 DH10B sequenced by EdgeBio (6 runs on 314 chip)
  - *E. Coli* O104:H4 (TY-2482) sequenced by BGI (7 runs on 314 chip)
  - *E. Coli* O104:H4 (LB226692) sequenced by LIFE Technologies (8 runs on 314 chip)

- **Run Performance Analysis**
  - Data loaded into FastQC
  - FastQC output files loaded into Excel to allow plotting multiple samples on same plot

- **Coverage Analysis**
  - EdgeBio *E. Coli* K-12 DH10B (aligned with TMAP 0.0.19) BAM file available
  - Illumina *E. Coli* K12 MG1655 (aligned with ELAND) BAM file available
  - BAM alignment files loaded into standard Illumina TruSeq coverage comparison script
  - BAM alignment files loaded into Picard
## MiSeq Yields >70× More Output Per Run

<table>
<thead>
<tr>
<th>Source</th>
<th>Instrument</th>
<th># of Runs</th>
<th>E. Coli Strain</th>
<th>Average Q Score / Run</th>
<th>Average Yield / Run</th>
<th>Average Depth / Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina</td>
<td>MiSeq</td>
<td>1</td>
<td>K12 MG1655</td>
<td>31</td>
<td>1.7 GB</td>
<td>393x</td>
</tr>
<tr>
<td>EdgeBio</td>
<td>Ion Torrent PGM (314)</td>
<td>6</td>
<td>K12 DH10B</td>
<td>19</td>
<td>24 MB</td>
<td>5x</td>
</tr>
<tr>
<td>BGI</td>
<td>Ion Torrent PGM (314)</td>
<td>7</td>
<td>O104:H4</td>
<td>18</td>
<td>11 MB</td>
<td>2x</td>
</tr>
<tr>
<td>LIFE Technologies</td>
<td>Ion Torrent PGM (314)</td>
<td>8</td>
<td>O104:H4</td>
<td>17</td>
<td>15 MB</td>
<td>3x</td>
</tr>
</tbody>
</table>

**Sequence More Strains Per Run, NOT More Runs Per Strain**

The Illumina MiSeq data represents what the forthcoming system is capable of: 21 runs across 3 sites including LIFE Technologies yield a reasonable comparison of PGM performance.
Positional Q Score Comparison MiSeq vs. Ion Torrent

E. coli Data

High Quality Bases at the End of a MiSeq Read

MiSeq: http://www.illumina.com/miseq
Life Technologies: http://www.iontorrent.com/applications-speed/
Positional Q Score Comparison MiSeq vs. Ion Torrent

E. coli Data – Multiple Runs

High Quality Bases at the End of a MiSeq Read
MiSeq Yields a High Depth of Coverage

Note: x- and y- axes are at different scale

MiSeq

Mean mapped coverage = 362.01, IQR = 61
% of reads have a Mapped Coverage depth of 15 or less = 0.89
% genome covered = 98.19

Ion Torrent

Mean mapped coverage = 5.05, IQR = 4
% of reads have a Mapped Coverage depth of 15 or less = 99.96
% genome covered = 92.68

Sequence More Strains Per Run, NOT More Runs Per Strain
At Equivalent Coverage, MiSeq Yields a More Even Coverage Distribution

MiSeq down-sampled

Ion Torrent
Data Sources

► Raw Data (FASTQ)
  – ILMN E. Coli K12 MG1655 05/27/2011  
    ▪ http://www.illumina.com/miseq
  – BGI E. coli O104:H4 06/02/2011  
  – EdgeBio E coli K-12 DH10B 05/18/2011  
  – LIFE E coli O104:H4 06/08/2011  
    ▪ http://www.iontorrent.com/applications-speed/

► Aligned Data
  – EdgeBio E coli K-12 DH10B 05/18/2011  
  – ILMN E. Coli K12 MG1655 05/27/2011  
    ▪ http://www.illumina.com/miseq

► Additional E.coli Data Sequenced on MiSeq
  – http://www.hpa-bioinformatics.org.uk/lgp/genomes