### Ion Torrent: Open, Accessible, Enabling

Mike Lelivelt, Ph.D., Director of Bioinformatics

The content provided herein may relate to products that have not been officially released and is subject to change without notice. ion torrent ◊★△○×□+≈ by *life* technologies"

Guanin

## Why Ion Torrent loves Galaxy?



- How genomics software works...
- Algos get wrapped into workflows that kick out pretty pictures. User turns knob. Create subset. Repeat.
- Genomics will impact diagnostic medicine.
- But there's a whole lot more discovery to undertake.
- The future of genomics in medicine is dependent on Galaxy.





## Galaxy Development Challenge

- Build a Galaxy transfer plug-in. Win 500 points on Ion Community (good towards an iPad), plus a iPod Nano.
- Gain the admiration of your Galaxy colleagues.
- Plus we'll give another iPad away at the end of this talk.





## Simple Natural Chemistry



Eliminate source error:

- Modified bases
- Fluorescent bases
- Laser detection

Eliminate read length limitations:

- Unnatural bases
- Protect/de-protect

by *life* technologies<sup>™</sup>

Slow cycle time

Sequence is determined by measuring hydrogen ions released (1 per base added per DNA strand) during 2nd strand synthesis when complementary base (A, C, G or T) are sequentially incorporated by DNA polymerase.

## **Transformative Technology**

Scalable Semiconductor Technology Plus Simple Chemistry





## **Torrent Browser runs on Torrent Server**

Local compute and storage with an integrated web interface



- Torrent Server hardware appliance
- Torrent Browser easy web access to Ion data
- Plugins for secondary analysis e.g. variant calling

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic us  $A \circ X \Box + \mathcal{Z}$ by *life* technologies" Data Flows Directly to Ion Reporter Torrent Circuit and Desktop Solutions also Enable Other Research Applications



The content provided herein may relate to products that have not been officially released and is subject to change without notice.  $\wedge \star \wedge \circ \times \Box + \sim$ 

by *life* technologies<sup>™</sup>

### **Torrent Suite Data Analysis Flow**



by *life* technologies<sup>™</sup>

### Proton Data Processing On/Off Instrument



ion torrent ◊ ★ △ ○ × □ + ≈ by *life* technologies"

### Torrent pipeline – Approx. data sizes\*



### Proton pipeline – Approx. data sizes\*



Note: Files are in "native" format

ion torrent ◊ ★ △ ○ × □ + ≈ by *life* technologies"

### Proton<sup>™</sup> Torrent Server Specifications

- A single, free-standing tower computer appliance that includes Torrent Suite Software v3.0 for signal processing, base calling, alignment, and variant calling
- Stores >50 Ion Proton™ I Chip runs







The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## Efficient data processing on Torrent Server



Processing times for Proton 1 chips will be hours, not days.





Ion PGM" Sequencer

## **Torrent Suite Virtual Machine**

- Virtual machines are available for use in development
  - Penguin Computing: <u>http://torrentsuite.iontorrent.com</u>
  - Amazon AMI: search "ion-torrent" or AMI-2c64c745
  - Download to run locally









### Torrent Suite SDK





## Ion Community

Learn, Collaborate, Contribute

### **Ion Community**



- Over 11,00 members
- Over 100 new members each week
- Source code
- Datasets

### Run RecognitION



- Rank your runs against other users
- Win free chips for monthly best run
- Win \$5k for new records

### Torrent Browser Plugin Store



- Share your plugins with the community and earn points towards prizes
- Over 25 plugins currently available for FREE



## Ion PGM<sup>™</sup> Performance Update



### Accuracy A track record of continuous improvement



### Nothing Gets Better Faster



### High Data Quality May 2012 launch 4.6Mb EHEC E. coli Read Length:

#### **Quality Score:** E. coli Ion Xpress Plus Fragment Library 700 600 500 400 300 200 100 0 30 +0 17 20 **BASE QUALITY (Q) BEA-638** ~250X Coverage Most Common Read Length 255 bp 99.6% Mean Raw Accuracy (1X)

BASES (Mb)

E. coli Ion Xpress Plus Fragment Library



#### **Raw Accuracy:**

READS

PER BASE ACCURACY

E. coli Ion Xpress Plus Fragment Library



by *life* technologies<sup>™</sup>

## AmpliSeq<sup>™</sup> Highly Consistent Results

A summary of seven validation runs performed in early June 2012.

Comprehensive Cancer Panel	Standard DNA (N=4)	FFPE DNA (N=3)
Coverage Uniformity	93.72 ± 0.16%	93.4 ± 1.3%
On Target Reads	94.30 ± 0.63%	95.4 ± 5.5%
Raw Accuracy	99.42 ± 0.08%	98.8 ± 0.8%
Depth of Coverage	350.53 ± 64.51x	342 ± 92x

- Coverage Uniformity—bases covered at 20% of the mean coverage
- On-target bases (Specificity) Bases that mapped to target regions out of total bases per run
- Depth of Coverage Average coverage over target region from a single 318<sup>™</sup> Chip



### Accurate 300 base Reads – Q3 2012



by *life* technologies<sup>™</sup>

### Accurate 400 base Reads – Q4 2012



by *life* technologies<sup>™</sup>

## **Homopolymer Performance**

Indel and homopolymer accuracy has improved ~450% since July 2011

	Indels per 100 bp
July 2011 Kits (Loman <i>et al</i> )	1.44
May 2012 Kits	0.32



**Homopolymer Length** 

Indels per 100bp. FASTQ files for Ion Torrent (1) run in the *Nature Biotechnology* paper were mapped with bwa using the same parameters as Loman et al and the indels\_per\_100 stat in Supplementary Table1 was calculated using the same script referenced in the publication (

https://github.com/nickloman/benchtop-sequencing-comparison/blob/ master/scripts/read\_bam.py) as well as the same reference Trimming and Filtering: Read filtering was done using a sliding window of size 10 moving from 5' to 3'. When the average QV of the window dropped below 15, the 3' end of the read was clipped. HP accuracy: BAM files were parsed using a C++ program alignStats, available as a binary in Torrent Suite, with the following parameters: alignStats -i <br/> <br/> -p 1

The resulting Default.sam.parsed files were used for homopolymer accuracy calculations. Accuracy is defined as 1 minus the total number of errors divided by numbers of bases and plotted against homopolymer length.

ion torrent <sup>◊</sup> ★ △ ○ × □ + ≈ <sup>by life technologies™</sup>

### Improving Homopolymer Calling

- NA12878 Sample Runs
- KER\_919
- FLO\_530
- FLO\_550

Total number of HP Variant (size > 3) Annotated in CCP 409 Panel = 14. Total positions analyzed:  $\sim$ 21000 homopolymer positions of length greater than or equal to 3 were analyzed using the model.

Sample	ТР	FN	FP
KER_919	13	1	3
FLO_530	13	1	3
FLO_550	13	1	4



Confidential and Proprietary—DO NOT DUPLICATE

### Detect 5% SNP in Ion AmpliSeq<sup>™</sup> Cancer Panel

- Detected rare 5% variants in Ion AmpliSeq<sup>™</sup> Cancer Panel
  - 90:10 mix of two known DNA samples (NA12878:NA19240)
  - Average coverage 1500x
- Sensitivity 83%
  - detecting 5 out of 6 true positives at 5% frequency
- Consensus Accuracy --99.98%



Detect rare 5% SNP in a heterogeneous population

ion torrent ◊ ★ △ ○ × □ + ≈ by *life* technologies"

### Ion Proton<sup>™</sup> System Overview



### Ion Semiconductor Sequencing Unprecedented Scalability Since 2011



been officially released and is subject to change without notice.

by *life* technologies<sup>™</sup>

### \$500 Exome and \$1,000 Genome Sequencing in as little as 2 hours on the Benchtop



### Ion Proton™ I Chip

2 human exomes Up to 10 Gb 165 million wells Up to 200bp reads 2-4 hour runs \$1,000 per run

### Ion Proton<sup>™</sup> II Chip

1 human genome Up to 20X 660 million wells Up to 200bp reads 2-4 hour runs \$1,000 per run



ion torrent <sup>∧</sup> ★ △ ○ × □ + ≈ by *life* technologies"

The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## Harnessing a Decade of Moore's Law in One Leap...





ion torrent

 $\wedge \star \wedge \circ \times \Box + \approx$ 

by *life* technologies<sup>™</sup>

Nature 475, 348-352 (21 July 2011)

## Ion Proton<sup>™</sup> I Chip: 165 million wells (>100-fold more wells than Ion 314<sup>™</sup> Chip)

The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## ...While Seamlessly Scaling Ion's PostLight™ Sequencing Chemistry...



### Same Signal, Same Speed

 Proton Pre-Release Data
 Not Final Commercial Performance

Internally generated R&D data shown The content provided herein may relate to products that have not been officially released and is subject to change without notice. ion torrent ∧ ★ △ ○ × □ + ≈ by *life* technologies"

### ...Resulting in the Rapid Development of the Ion Proton<sup>™</sup> System over the Past 6 Months



 Proton Pre-Release Data
 Not Final Commercial Performance

Internally generated R&D data shown The content provided herein may relate to products that have not been officially released and is subject to change without notice. ion torrent ◊ ★ △ ○ × □ + ≈ by life technologies"

### Ion Proton<sup>™</sup> System The Benchtop Genome Center



- Runs Ion Proton<sup>™</sup> I and II Chips
- Benchtop system containing state-of-the-art electronics to support the highest throughput
  - Dual 8-core Intel® Xeon® Sandy Bridge processors
  - 128 GB of RAM
  - Altera® Stratix V FPGA
  - NVIDIA GPU
  - 11 TB of hybrid (SSD&HDD) storage
  - Ubuntu® 11.1 operating system
- List price: \$224K system (USD, sequencer & server)

The content provided herein may relate to products that have not been officially released and is subject to change without notice.

### Hardware Differences



- Instrument:
  - AD on Instrument
  - Low Compute PPC
  - Reader Board:
    - Older generation FPGAs
  - No HW Accelleration for Analysis
- Torrent Server
  - Supports up to 2 PGMs
  - 12 Intel Cores
  - 1 Nvidia C2050 GPU
  - 48 G Memory



- Instrument:
  - AD on Ion Chip
  - Dual Socket Xeon Server
  - Reader Board:
    - Latest Stratix V FPGAs
  - NVidia GPU Acceleration
- Torrent Server
  - Supports Single Proton
  - 16 Intel Cores
  - 2 NVidia C2075 GPUs
  - 128 G Memory

### Single Day Workflow...



The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## ...With the Fastest Sequencing Run Times

### Sequencer Throughput Comparison (Gb per hr)

		• Exome	and genome s	sequencing i	n hours
	Proton II 10 Gb/hr	<ul> <li>Proton I throughput comparable to industry standard</li> </ul>			
		<ul> <li>Proton I improve</li> </ul>	I will be ~3X f ments to othe	aster than pr er platforms	ojected
		2500 4.4 Gb/hr			
	Proton I 2.5 Gb/hr	2000 2.3 Gb/hr	318 0.3 Gb/hr	2X150 0.1 Gb/hr	_
	lon Proton™	Illumina HiSeq™	lon PGM™	Illumina MiSeq™	1

The content provided herein may relate to products that have not been officially released and is subject to change without notice. ion torrent

 $\delta \star \Delta O \times \Box + \tilde{z}$ 

by *life* technologies<sup>™</sup>

### Exome Workflow and Data on the Ion Proton<sup>™</sup> System



### Ion Proton<sup>™</sup> System Exome Workflow A Complete, End-to-End Exome Solution





The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## TargetSeq<sup>™</sup> Enrichment Uses a Simple **Probe Capture Workflow**



## Ion TargetSeq<sup>™</sup> Exome Kit with Ion Proton<sup>™</sup> Sequencing Yields High Quality Exomes...

Chip Loading	Count	Percentage	
Total Addressable Wells	164,822,784		
Wells with ISPs	139,574,799	85%	
Live ISPs	135,946,926	97%	
Test Fragment ISPs	39,283	<1%	
Library ISPs	135,907,643	100%	

Throughput (Gb)	Mappable	Perfect
Total Number of Bases	7.0 Gb	5.2 Gb

Throughput (Reads)	Mappable	Perfect
# of Reads ≥50 bp	99,615,132	49,126,825
# of Reads ≥100 bp	88,236,276	21,837,494
# of Reads ≥150 bp	584,960	2,654
Data Quality	Avg. Raw Read	@ 100 bp

99%

CDG-94-r23244-4 Loading Density (Avg ~ 85%)



Modal (most common) Read Length:125 bp

 Proton Pre-Release Data
 Not Final Commercial Performance

Accuracy

Internally generated R&D data shown The content provided herein may relate to products that have not been officially released and is subject to change without notice.

99%



## ...with Excellent Coverage Uniformity and SNP Concordance



Coverage

# called SNPs	% SNPs in dbSNP135	# called Hets	% Hets in dbSNP135	# called Homs	% Homs in dbSNP135
29,598	98.0%	17,579	97.2%	12,019	99.3%

40
 40
 41
 42
 43
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 44
 4

Internally generated R&D data shown The content provided herein may relate to products that have not been officially released and is subject to change without notice. 

### Transcriptome Workflow on the Ion Proton<sup>™</sup> System



### Transcriptome Workflow on the Ion Proton™ A Rapid, End-to-End Transcriptome Solution





The content provided herein may relate to products that have not been officially released and is subject to change without notice.

### Whole Transcriptome Library Solution Ion Total RNA-Seq Kit + RiboMinus™ Eukaryote Kit v2

### Ion Total RNA-Seq Kit v2

•Creates whole transcriptome (WT) libraries

•Maintains strand orientation

•Minimizes bias and error

•Fast, simple, workflow (no gel purification)

•Low sample input requirements: 100ng of total RNA or 1-500ng of mRNA

### <u>RiboMinus™ Eukaryote Kit v2</u>

Improved probe design targeting multiple rRNA types

- <10% mapping to cytoplasmic and mtRNA
- •1h workflow (optional mag bead purification)
- •Low and standard input kits:
- 10ng-1µg for low; 1-5µg for standard input
- •Comparable pricing to current RiboMinus kits





## More Information is Available

Learn more about how to analyze Ion RNA-Seq Data

 White paper is available on the Ion Community that explains more details on processing Ion RNA-seq data

traditional platforms for measuring gene expression in

a cell type or tissue [1,2]. Furthermore, with increased

in cancer [4], and microRNAs [5] (miRNA) are possible.

sequencing depth, other more specialized analyses such as

identification of novel splice variants [3], transcript fusions



- 2. Quality control/assessment
- 3. Mapping to a reference genome and or transcripts
- 4. Counting mapped reads
- 5. Generation of global mapping statistics



### lifetechnologies.com/ioncommunity

#### ion torrent <sup>()</sup> ★ △ ○ × □ + ≈ <sup>()</sup> by *life* technologies"

## QA/QC for ERCC

Quickly generate a QA / QC plot for samples using ERCC spike-ins

ERCC\_Analysis

### Dose response plot for any run using ERCC

### ERCC Report



by *life* technologies<sup>™</sup>



## Ion RNA-Seq Data Analysis Workflow

Leverages tools and software already familiar to microarray users



Avadis NGS 🔅 Partek

Microarray User-Friendly Analysis tools:

- Partek® and GeneSpring® based tools
- Full compatibility with NGS and microarray data
- Both tools take FASTQ files from PGM, align the sequences and provide an extensive set of options for data visualization, statistical tools and annotation tools
- Comprehensive data analysis tools to reduce your time to publication





### Thanks – Let's eat & drink. Repeat.



### Just one more thing...

- Right now, let's give away an iPad 2.
- Right here.
- Right now.
- Then we drink.





# Ion torrent $\delta \star \Delta O \times \Box + \sim$ by *life* technologies™

All products mentioned in this presentation are for Research Use Only, not intended for any animal or human therapeutic or diagnostic use.