

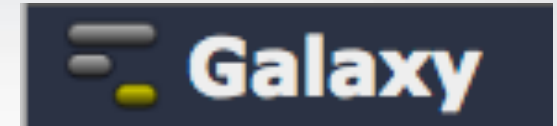
# Sequencing for All™

Ion Torrent™ next-generation  
sequencing solutions

Mike Lelivelt, Ph.D  
Director Bioinformatics  
Ion Torrent

*life*  
technologies  
A Thermo Fisher Scientific Brand

# 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.

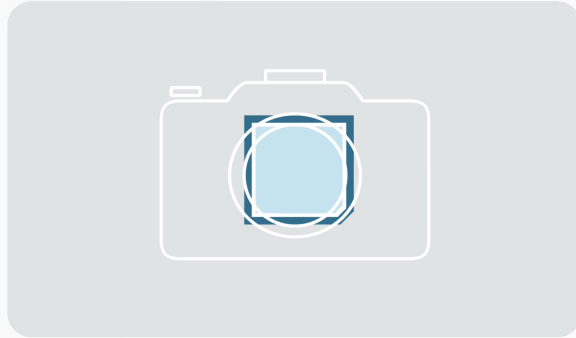


*life*  
technologies

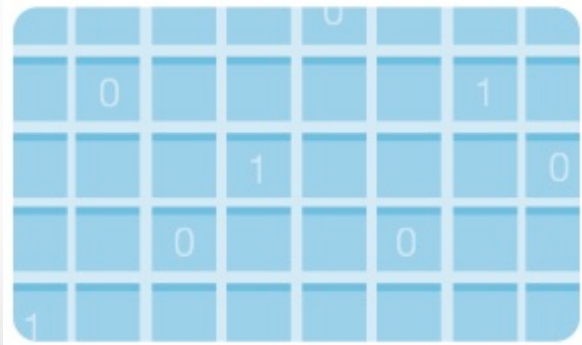
# Ion Torrent™ semiconductor sequencing

*Uses a process similar to that used in a digital camera*

**Digital camera chip**



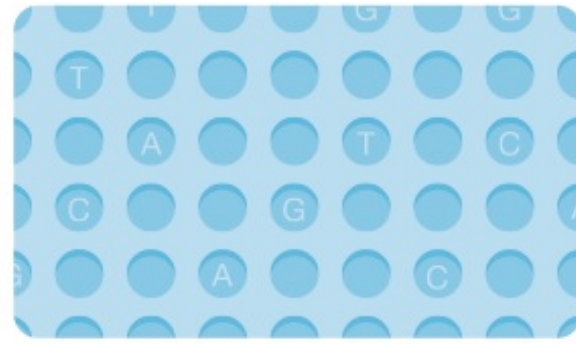
Covered in millions of pixels that convert light to digital information



**Ion Torrent™ sequencing chip**



Millions of wells covering those pixels that convert chemical into digital information

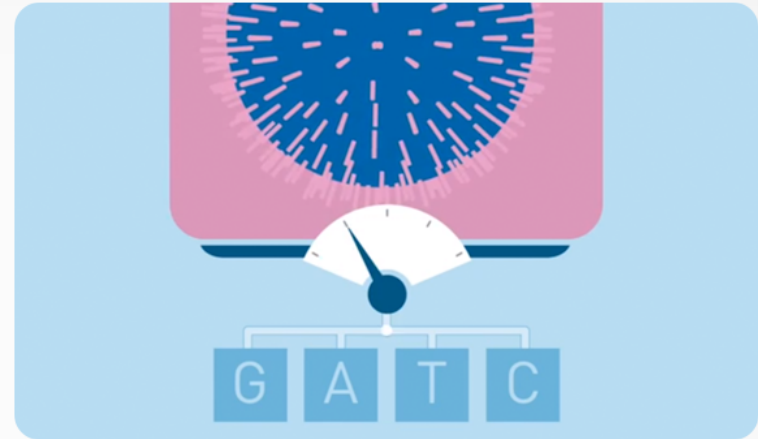


# Ion Torrent™ semiconductor sequencing

*The world's smallest pH meter*



Whenever a nucleotide is incorporated into a single strand of DNA, a hydrogen ion is released changing the pH of the surrounding solution

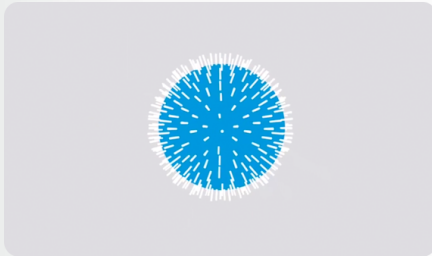


Ion Torrent semiconductor sequencing chip is able to detect this change in pH and convert it directly to base calls

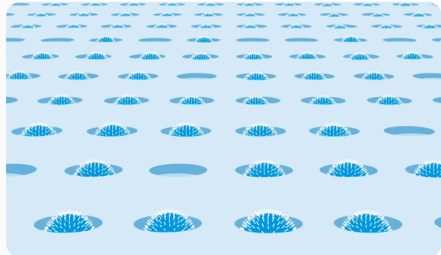
# Ion Torrent™ semiconductor sequencing

## *The sequencing process*

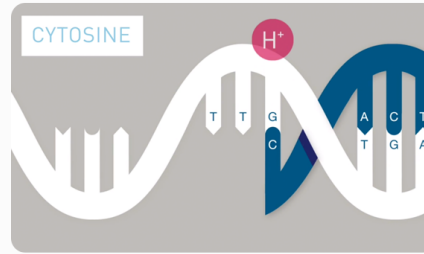
Copy DNA



Load chip



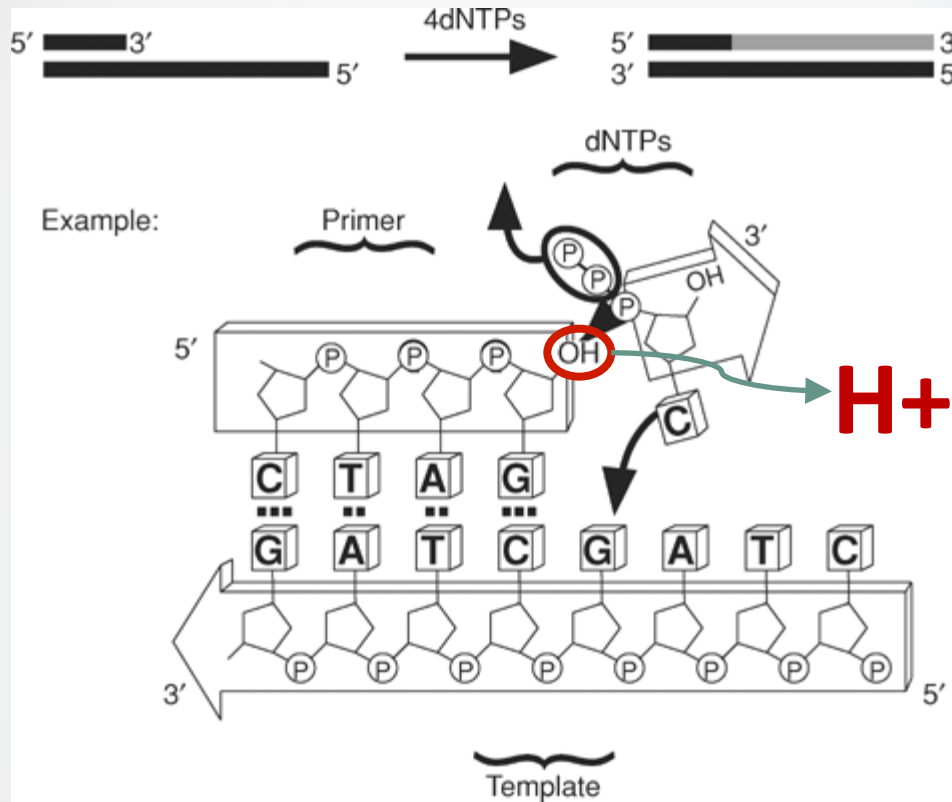
Incorporate nucleotide



Detect and call



# Simple Natural Chemistry



## Eliminate source error:

- Modified bases
- Fluorescent bases
- Laser detection

## Eliminate read length limitations:

- Unnatural bases
- Protect/de-protect
- Slow cycle time

# Agenda

## ✓ Ion Chef & Accuracy Update

HiQ Universal Kits for PGM & Preview of P1v4 performance

## ✓ Ion Proton™ System

Exomes, Transcriptomes

P11 Chip update



## ✓ Ion TrueMates Library Kit

New Long Mate Pair Kit

## ✓ Ion Bioinformatics

Torrent Suite & Ion Reporter Software

ion torrent  
Sequencing for all.™

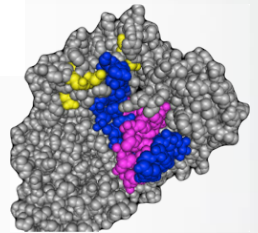
life  
technologies



# Focus Areas for Ion Torrent



Workflow



Accuracy

*life*  
technologies



# Simplified Workflow with the Ion Chef™ System



**Ion Chef™ System**

## Simple to use

- Fully automated template prep AND chip loading
- Simple reagent and consumables loading
- Minimizes potential sources of variability

## High throughput

- Processes 2 chips and multiple samples within hours

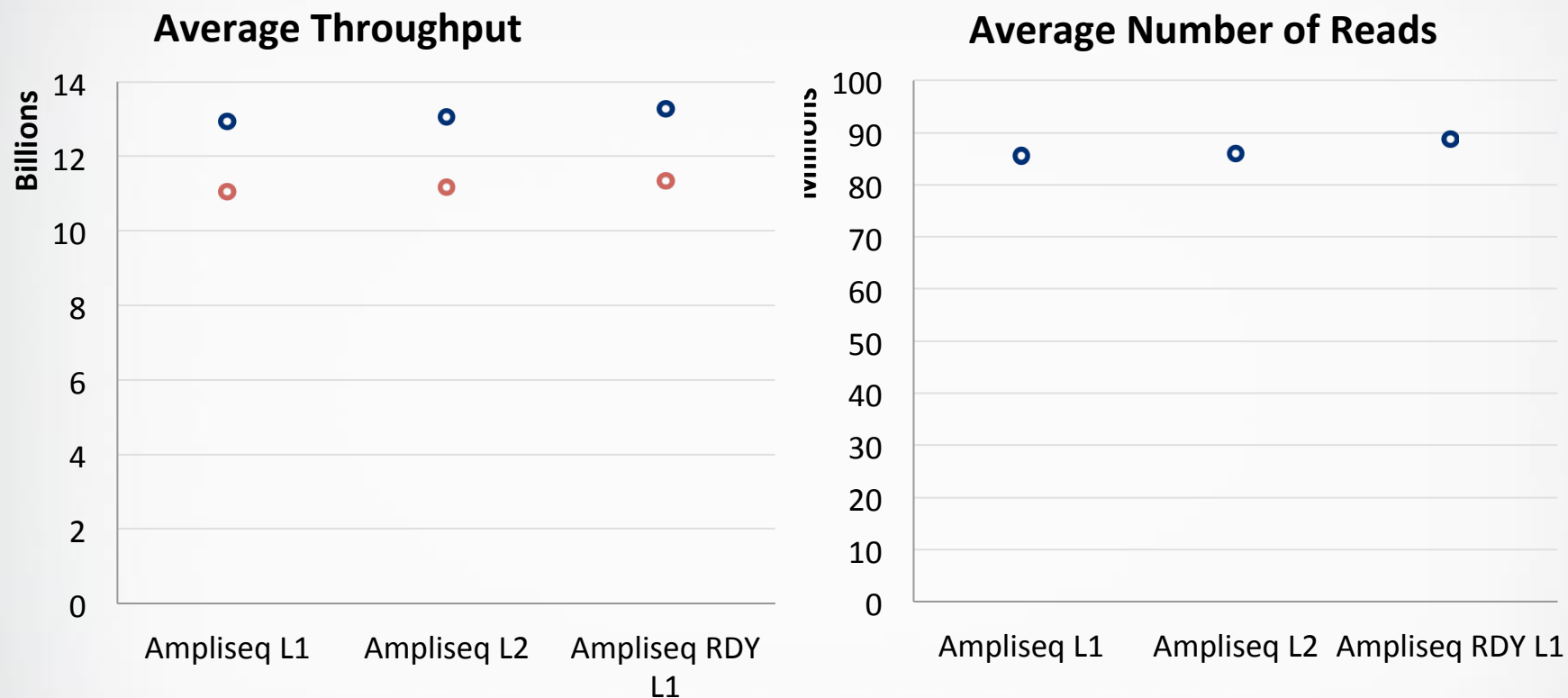
## Flexible

- Supports Ion PGM™ and Ion Proton™\* systems

\*Ion Proton™ System supported in Q2



# Ion Chef™ for Proton PI Chips – Validation Runs



- **Average of 122 Proton runs in validation.**

● AQ17    ● AQ20

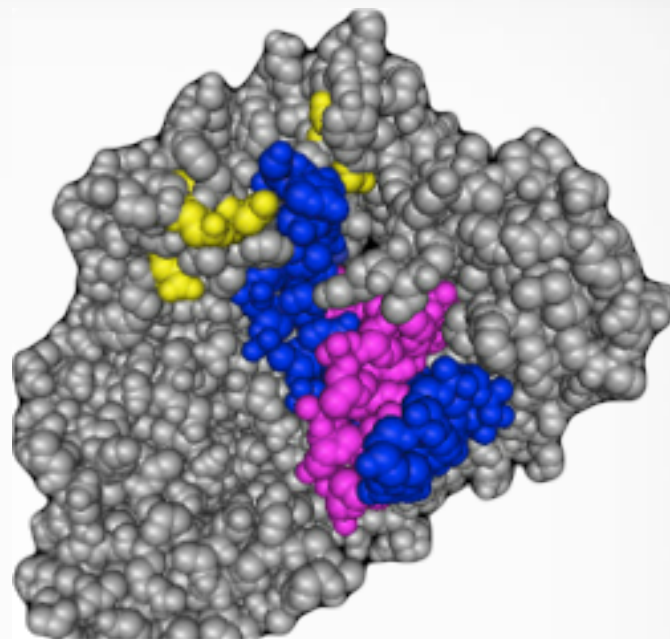


# HiQ & Proton Update

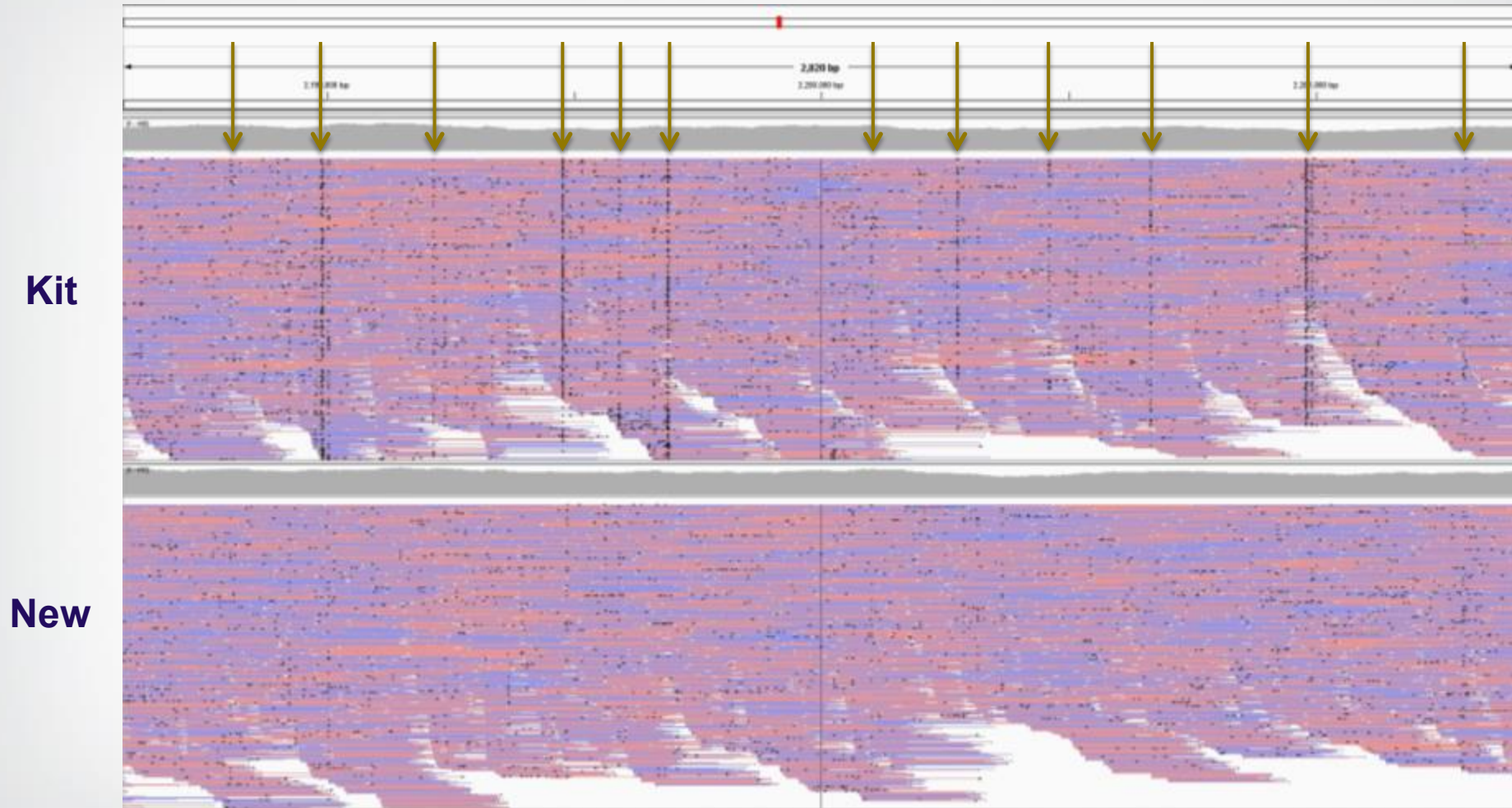


# Approach to Achieving Higher Accuracy

- Performed molecular evolution screen to identify high accuracy DNA polymerases
- Screened >10,000 enzyme variants
- Identified >10 high accuracy candidates
- Yielded new “Hi-Q™ Chemistry”
- Observed significant reduction in insertion and deletion errors

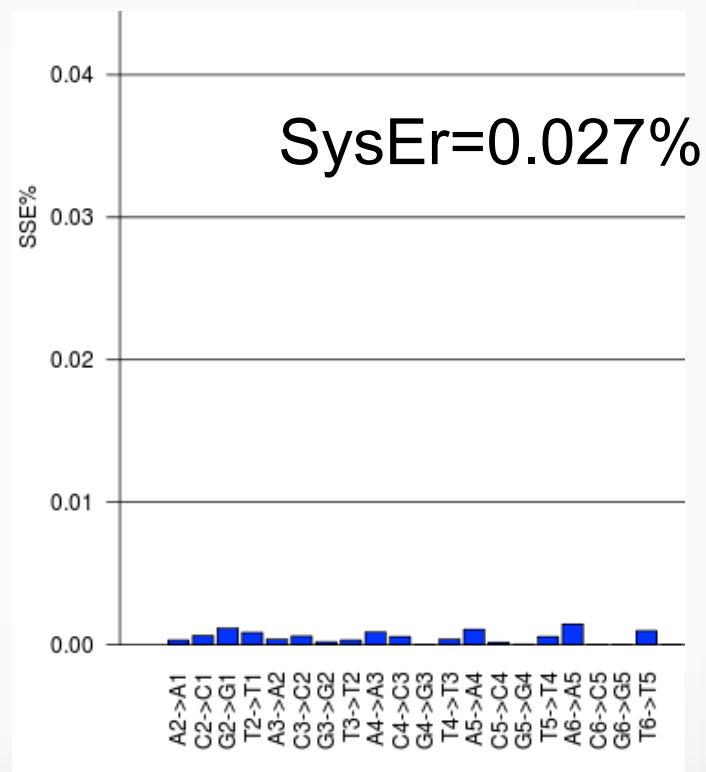
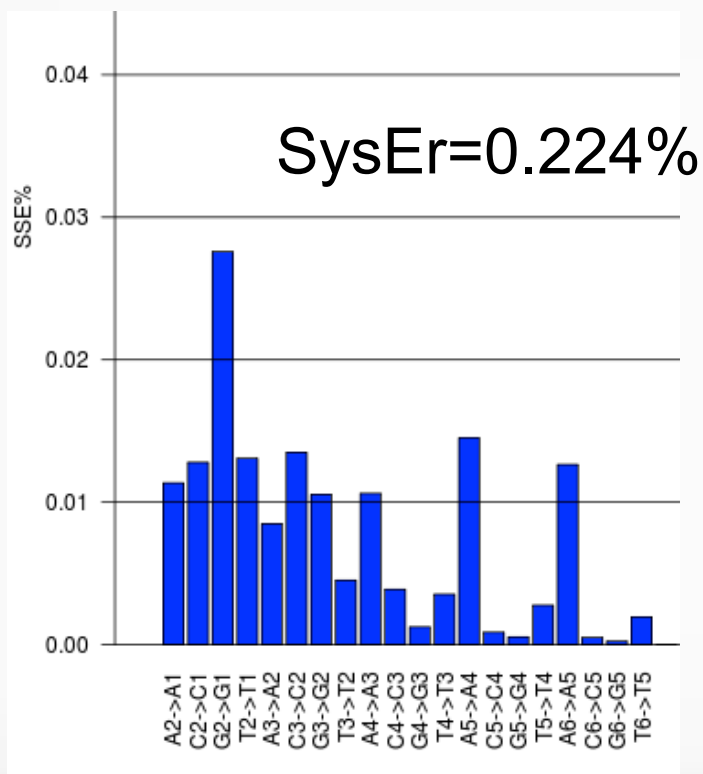


# HiQ - Dramatic Improvement as seen in IGV



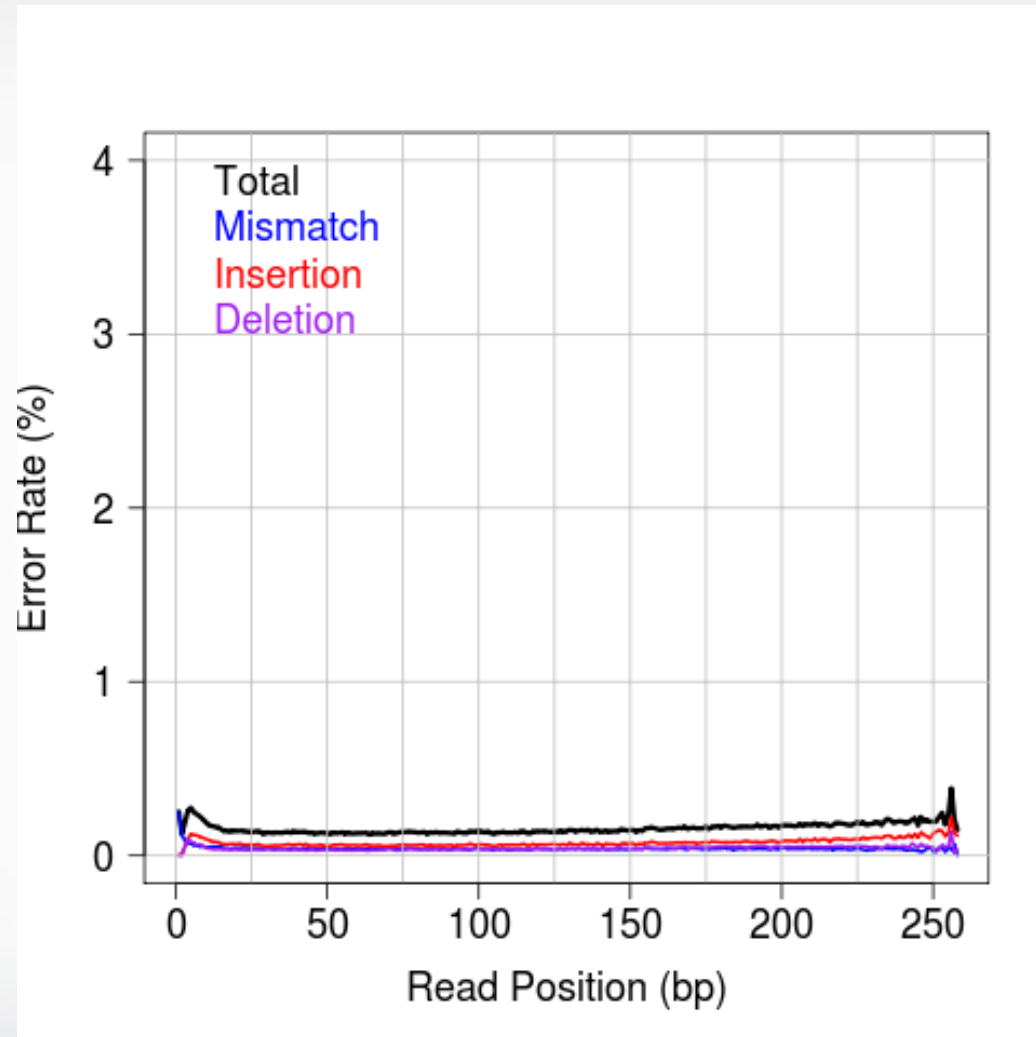
*mfe*  
technologies

# HiQ - Systematic Errors Reduced by 90%



# Hi-Q™ Enzyme

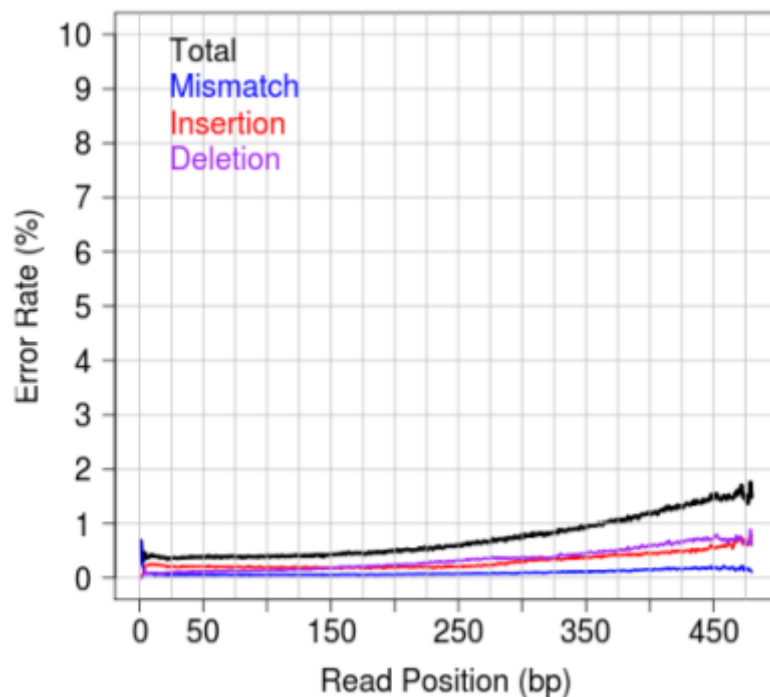
## Achieved Accuracy of 99.9% on 300bp Templates



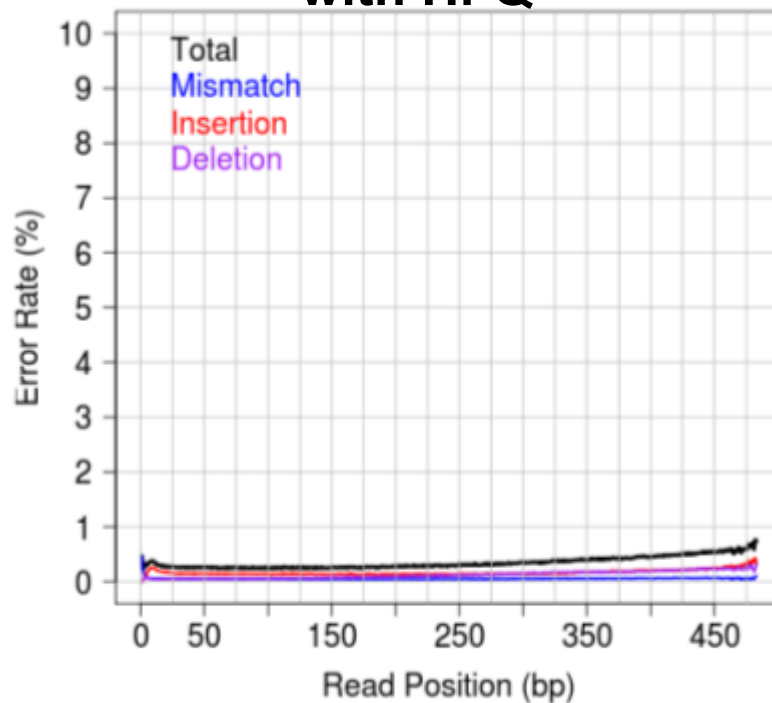


# Hi-Q™ Improves Accuracy of a 500bp Template

Ion PGM™ 400 bp kit

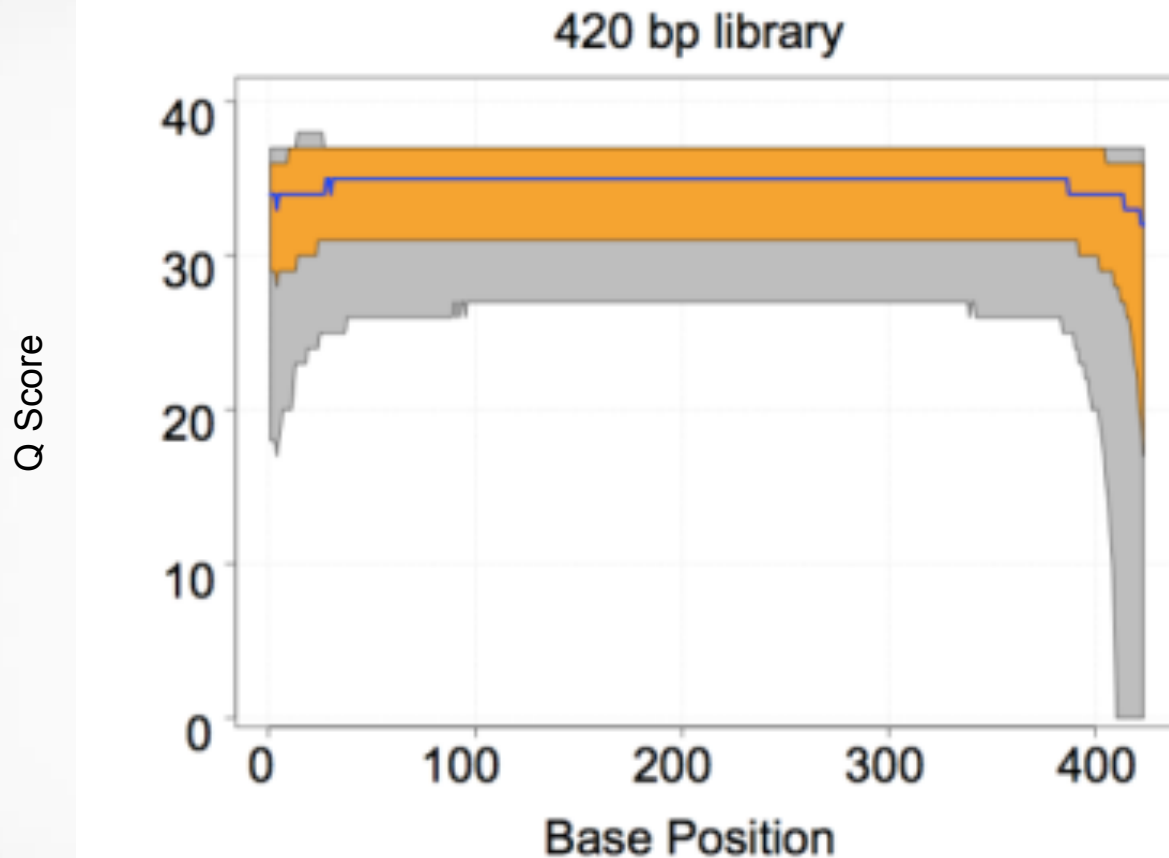


Ion PGM™ Universal Kit  
with Hi-Q™



# Hi-Q™ Enzyme: High Percentage of bases >Q30

All bases >Q30 in range 1-400bp  
Ion PGM System



Data available at Ion Community

<http://ioncommunity.lifetechnologies.com/docs/DOC-8251>

life  
technologies

# Ion Proton P1v3 Sensitivity Jumps to 92%

*Use the right variant caller for Ion data. NOT GATK!  
Ion data is very competitive with ILMN.*

## Our new comparison: target region only

Data	Pipeline	GIB Sensitivity	GIB Specificity	Ti/Tv	SNPs	Indels	Novel %
Illumina(150X)	BWA+Freebayes-Prep	96.88%	100.00%	2.612	37,031 (94.48%)	2,126 (5.42%)	1,599 (4.08%)
Ion (30X)	IonT-30x-Tmap+Gatk_HC-Prep	42.01%	99.96%	3.119	13,625 (36.70%)	23,494 (63.29%)	22,207 (59.82%)
Ion (30X)	Tmap+Gatk_UG	77.33%	99.73%	2.983	18,924 (12.31%)	134,756 (87.69%)	128,514 (83.62%)
Ion (30X)	TMAP+FREEBAYES+Prep	82.42%	99.92%	2.79	20,553 (35.86%)	36,748 (64.12%)	35,120 (61.28%)
Ion (123X)	TMAP+TVC+Prep	92.53%	100.00%	2.532	48,292 (95.34%)	2,320 (4.58%)	817 (1.61%)

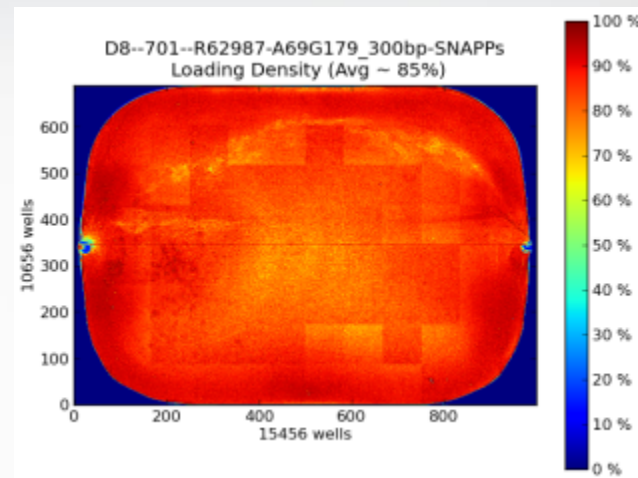
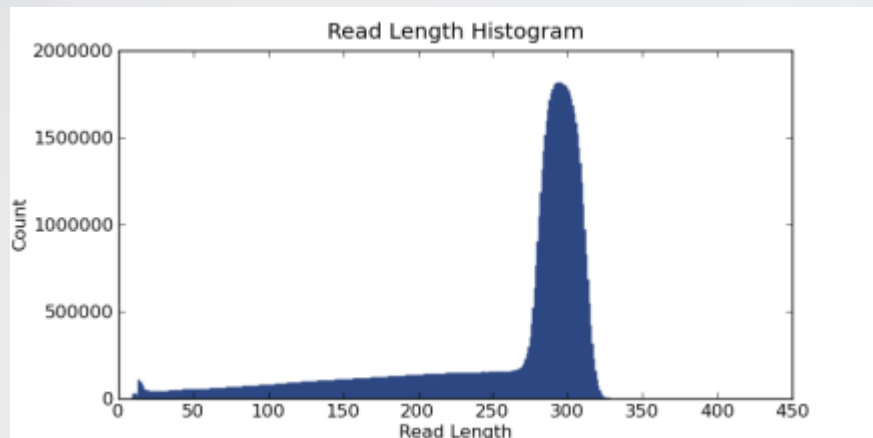
## Original NIST paper comparison (Fig 3d and Supp Fig29-32): target + off target

Data	Pipeline	GIB Sensitivity	GIB Specificity	Ti/Tv	SNPs	Indels	Novel %
Illumina(150X)	BWA+Freebayes-Prep	96.88%	100.00%	1.965	377,501 (93.07%)	27,559 (6.79%)	39,969 (9.85%)
Ion (30X)	IonT-30x-Tmap+Gatk_HC-Prep	42.01%	99.96%	2.828	21,747 (41.38%)	30,801 (58.61%)	28,765 (54.74%)
Illumina(150X)	Novoalign+Freebayes-Prep	96.31%	100.00%	2.061	287,744 (92.22%)	23,872 (7.65%)	18,492 (5.93%)
Illumina(150X)	Novoalign+Samtools-Prep	96.91%	100.00%	1.94	380,122 (92.71%)	29,871 (7.29%)	37,526 (9.15%)

New Ion Exome Data with P1v3 and TSv4.0

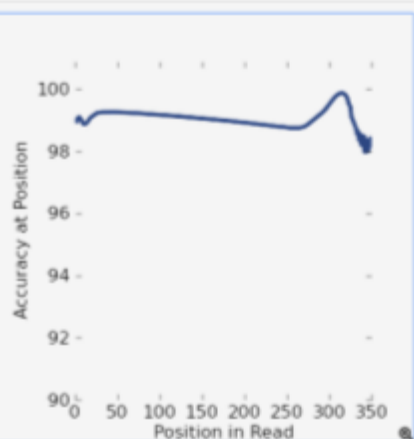


# Preview of Ion PI™ v4 Chemistry Development



**99.1%**

Mean Raw Accuracy 1x



**19.8 G**

AQ17 Total Bases

## Alignment Quality

	AQ17	AQ20	Perfect
Total Number of Bases [bp]	19.8 G	17.9 G	12.2 G
Mean Length [bp]	246	230	166
Longest Alignment [bp]	416	416	394
Mean Coverage Depth [x]	6.4	5.8	4.0

- 99.1% raw read accuracy
- 230 bp MRL AQ20
- 19.8 Gb total bases
- PI v4 chip innovation + Hi-Q™ Chemistry

life  
technologies

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

# Ion PII™ Chip: A 660 Megapixel Proton Video Camera

## Contrast to the Ion PI™ Chip:

- 4x as many wells and transistors for much higher throughput
- Twice the data rate to maximize data analysis rates and minimize analysis times
- Higher signal to noise to maximize quality of signal output



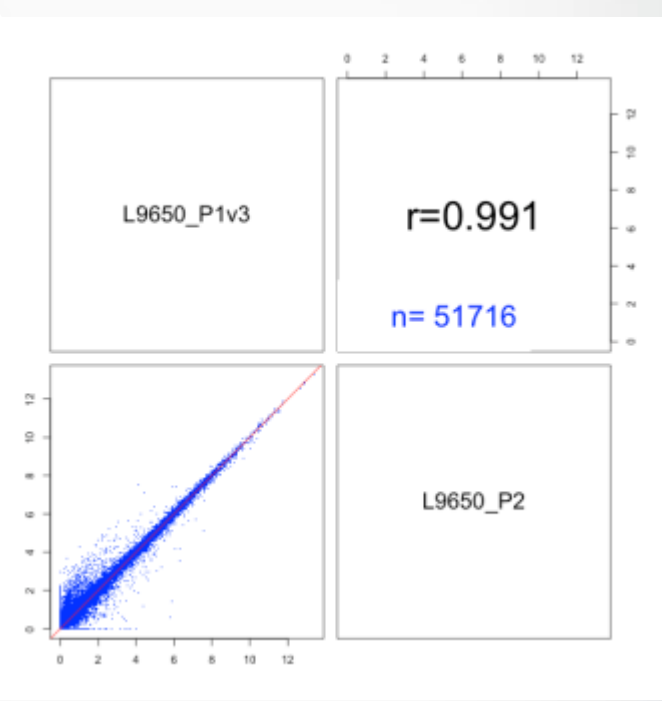
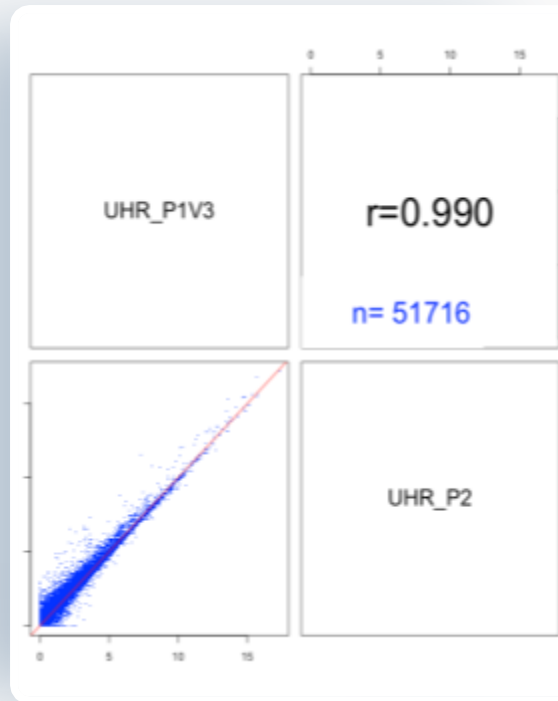


# Ion PII™ Chip Transcriptomes

Internal HBR Library

Internal UHR Library

Customer Library

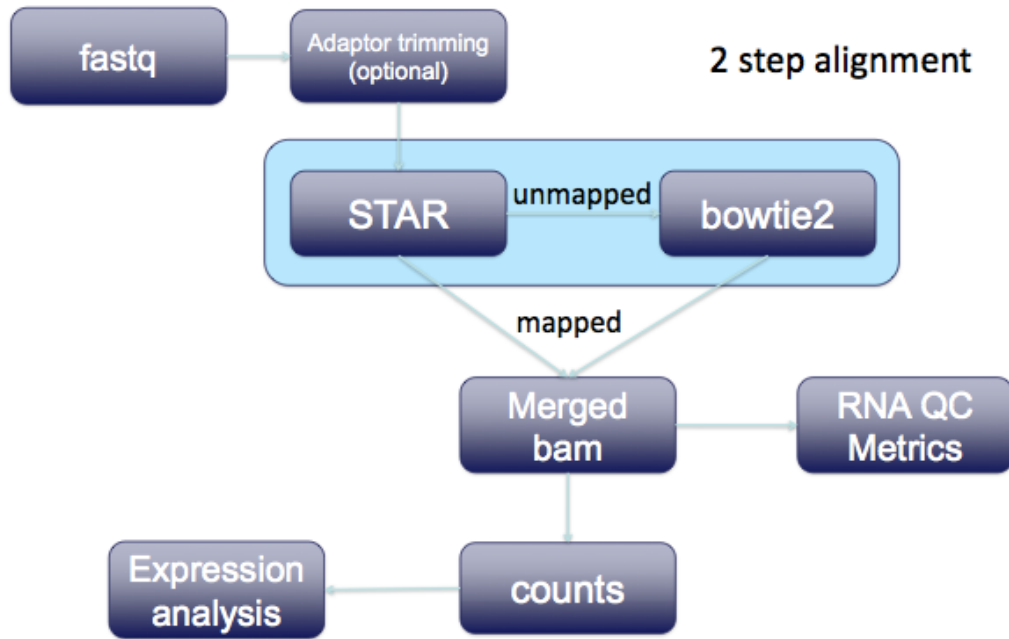




# Want to analyze your Ion RNA-seq data in Galaxy?

## RESULTS

Figure 1. Two-step alignment workflow



Two-step alignment workflow. The reads are aligned to STAR (or TopHat2). The reads will be either aligned or un-aligned in this process. The **un-aligned** reads are then aligned with bowtie2 (in local mode). The aligned bam from first step and second step is then merged to generate final aligned bam. This final bam is used for further downstream analysis.

Visit Ion  
Community for  
detailed info on  
methods

The screenshot shows the Ion Community website interface. The top navigation bar includes 'Home', 'Applications', 'Datasets', 'Protocols', and 'Sys'. A search bar is located on the right. Below the navigation bar, there is a section titled 'More documents in' with a dropdown menu set to 'Proton'. The main content area displays a document titled 'Two Step Alignment Method for Ion Proton™ Transcriptome Data' (Version 4). The document is attributed to 'Lifetech Administrator' and 'Mary Budagyan'. A 'Required:' section lists the following tools and their corresponding URLs:

Tool	URL
input.fastq	The fastq file from Proton run
cutadapt	<a href="http://code.google.com/p/cutadapt/">http://code.google.com/p/cutadapt/</a>
tophat2	<a href="http://tophat.cbcb.umd.edu/">http://tophat.cbcb.umd.edu/</a>
bam2fastq	<a href="http://www.hudsonalpha.org/gslinformation/software/bam2fastq">http://www.hudsonalpha.org/gslinformation/software/bam2fastq</a>
bowtie2	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
picard	<a href="http://picard.sourceforge.net/">http://picard.sourceforge.net/</a>

<http://ioncommunity.lifetechnologies.com/docs/DOC-7062>



# Sample Data is Freely Available at the Ion Community



The screenshot shows the Ion Community website interface. At the top is the 'life technologies' logo and the 'Ion Community' title. A navigation bar includes links for Home, Applications, Datasets (which is highlighted), Protocols, and Systems. Below this is a 'Datasets By Chip' section with a dropdown menu set to 'Datasets'. A sub-navigation bar includes 'Overview' (highlighted), 'Content', 'People', 'Subspaces and Projects', and 'Reports'. The main content area is titled 'HIGHLIGHTED DATASETS' and features a table of datasets. Above the table is a row of chip type filters: 314, 318, 316, 314v2, PIV2 (selected), 318v2, and 316v2. The table lists five datasets with their names, application types, chip types, and posting dates.

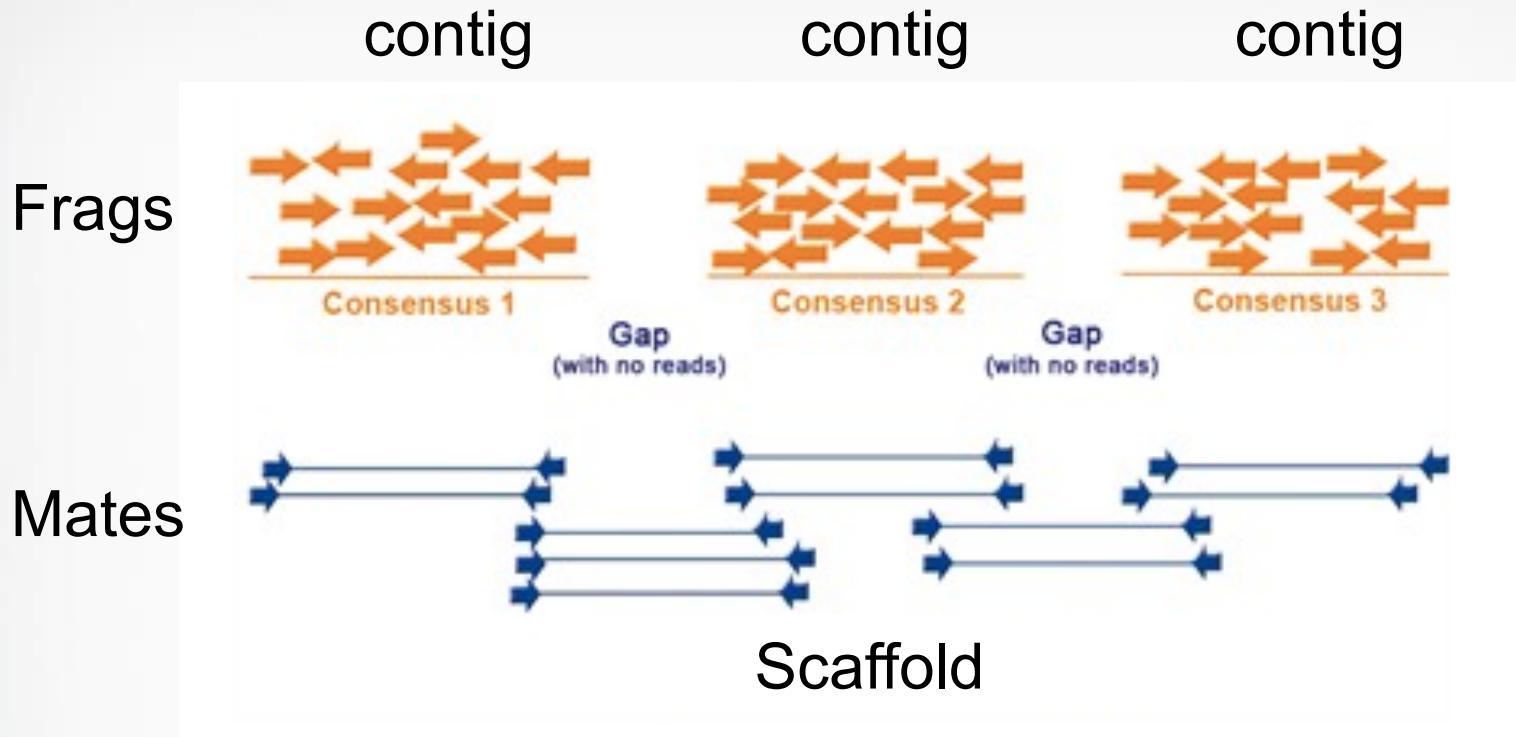
Name	Application Type	Chip Type	Date Posted
<a href="#">Ion Proton Whole Transcriptome RNA-Seq using Breast Cancer Cell Lines</a>	RNA-seq	PIv2	September 12, 2013
<a href="#">Ion RNA AmpliSeq 1200 Gene Breast Cancer Research Panel from Proton using Cell Lines</a>	RNA-seq	PIv2	August 30, 2013
<a href="#">Ion TargetSeq Exome on Ion Proton System using Ion PI Chip v2</a>	Exome	PIv2	August 5, 2013
<a href="#">Ion Proton Breast Tumor Tissue Whole Transcriptome RNA-Seq Datasets</a>	RNA-seq	PIv2	July 26, 2013
<a href="#">Ion AmpliSeq Exome on Ion Proton System using Ion PI Chip v2</a>	Exome	PIv2	June 30, 2013



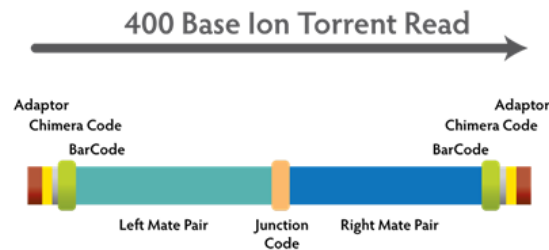
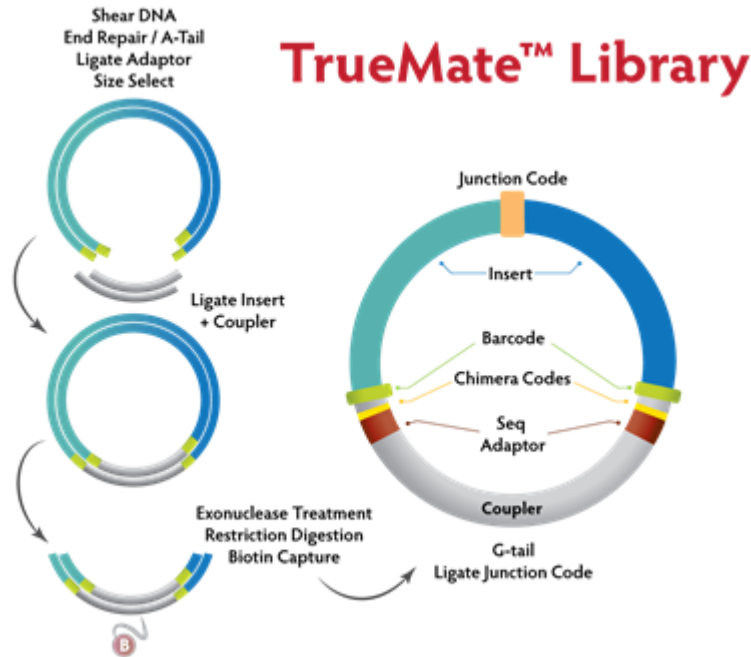
## Ion True Mates – Long Mate Pair Library Kit



# Long Mates Bridge Gaps from Short Reads in Assembly



# Ion TrueMate Libraries – One read gives two ends

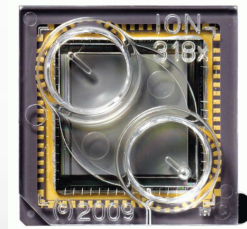


## Key Performance Metrics:

- Construction of “true” mate-pair libraries (not paired-end reads)
- Produces LMP libraries with 2-8 kb inserts now, 10-50 kb later
- gDNA input from 2-10ug total
- <8 kb protocol does not require the use of agarose gels (“gel-free”)
- Chimera code enrich for true mates
- 3 day workflow w 6 ½ hrs hands on

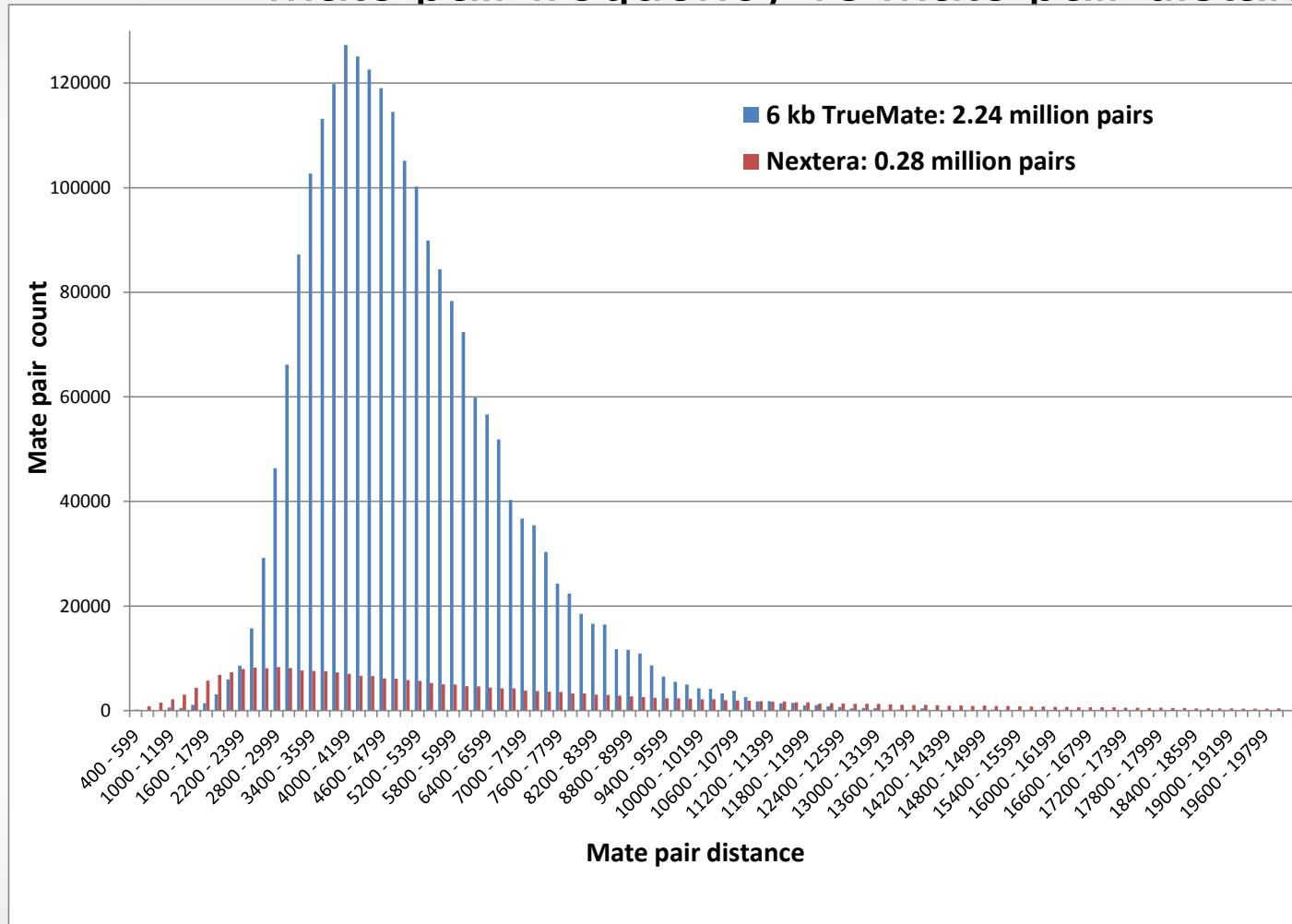
# Ion True Mates K12 Runs Data & Mapping

Library:	6kb gTUBE DAV-308	8kb gTUBE BUT-874	10kb gTUBE RAT-314	8kb Mega BEL-842
Raw reads	5,749,629	7,057,911	6,184,059	6,909,356
Mates	3,385,544 (59%)	3,777,430 (54%)	3,911,402 (63%)	4,106,400 (59%)
True mates	3,181,604 (94% of mates)	3,559,833 (94% of mates)	3,824,755 (98% of mates)	3,985,628 (97% of mates)
Chimera	203,940 (6%)	217,597 (6%)	86,647 (2%)	120,772 (3%)
Split reads (mate reads)	2,667,486 (84%) (5,334,972)	2,933,787 (82%) (5,867,574)	3,206,630 (84%)	3,288,275 (83%)



# Now which mate pair library would you rather use?

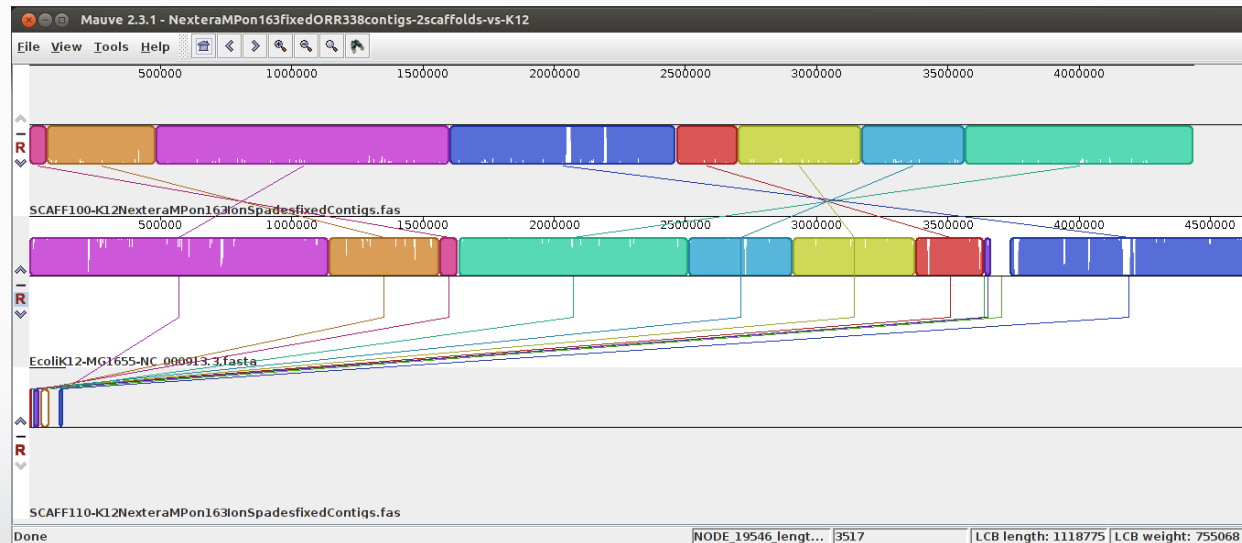
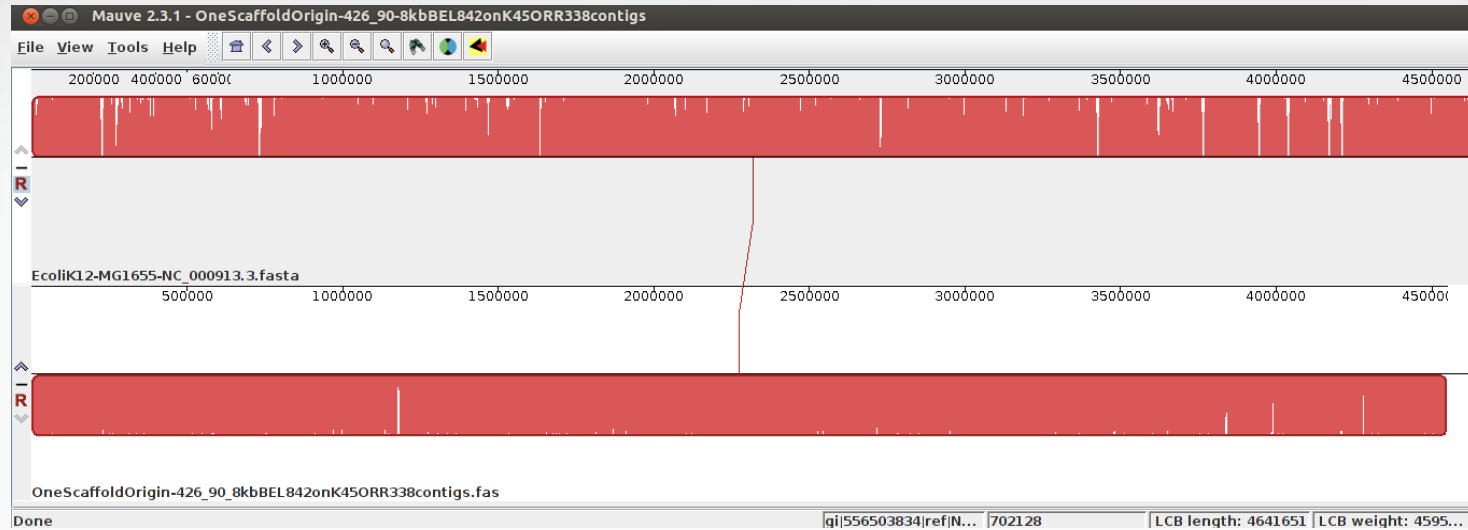
Mate pair frequency vs Mate pair distance



► TrueMate means tighter size distribution *and* more pairs!



# Ion Gives One Scaffold for K12. Illumina gives 8!

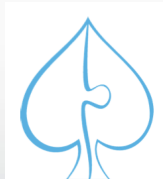


# Mate pair consistency comparison - DNASTar

	<u>8kb TrueMates</u>	<u>Nextera</u>
Total mate pair reads:	6,566,706	2,724,966
Total pairs:	3,283,353	1,362,483
Unassembled pairs:	972,775 (30% of total)	1,122,230 ( <u>82%</u> of total)
Assembled pairs:	2,310,578	240,253
Pairs consistent:	1,761,158 (76%)	167,136 (69%)
<b>Pairs inconsistent:</b>	<b>14,863 (&gt;1%)</b>	<b>16,245 (10%)</b>
thus pairs used for scaffolding =	544,557	56,872

# Long Mates Improve Assemblies Using SPAdes v3.1

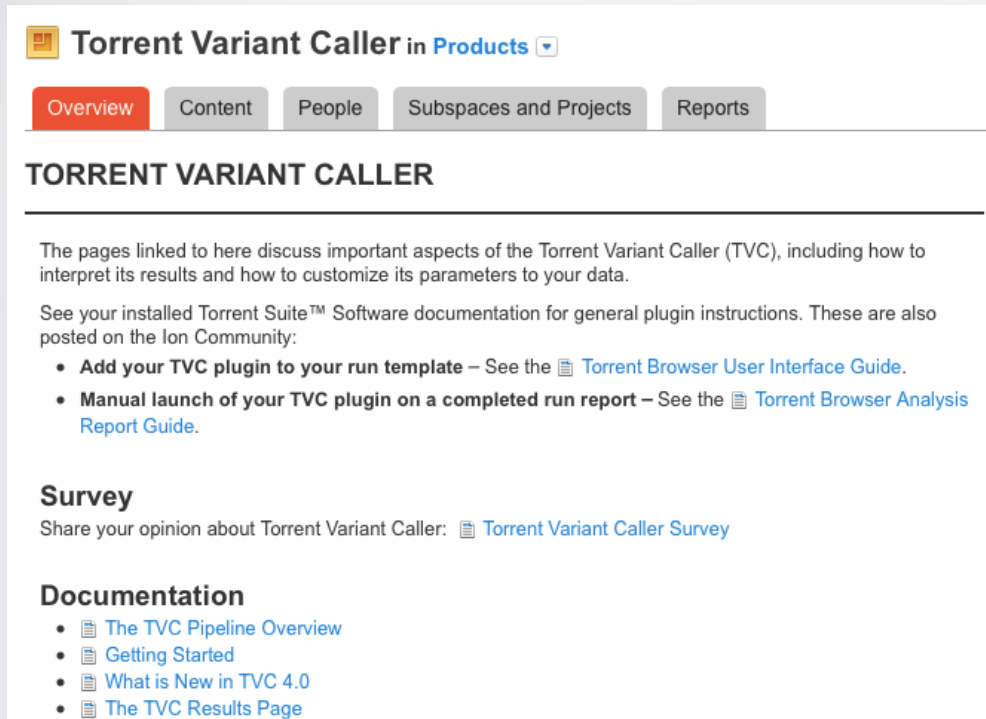
	PGM only	PGM + PacBio	PGM+Ilmn MP	PGM+MP+PacBio
Contigs > 500 bp	87	6	17	4
Largest Contig	284347	1476698	1255741	2563916
Total Length	4569274	4629059	4603307	4628981
NG50	117697	1375866	1187040	2563916
NG75	64475	1005145	651436	1186006
Misassemblies*	3 (1)	8 (1)	10 (3)	10 (3)
Genome fraction (%)	98.42	99.79	99.04	99.79
Indels per 100 kbp	5.13	6.57	5.16	6.72




# Ion Informatics Update



# Torrent Variant Caller – Available Stand Alone



 **Torrent Variant Caller** in [Products](#) ▾

[Overview](#) [Content](#) [People](#) [Subspaces and Projects](#) [Reports](#)

---

## TORRENT VARIANT CALLER

---

The pages linked to here discuss important aspects of the Torrent Variant Caller (TVC), including how to interpret its results and how to customize its parameters to your data.

See your installed Torrent Suite™ Software documentation for general plugin instructions. These are also posted on the Ion Community:

- **Add your TVC plugin to your run template** – See the [Torrent Browser User Interface Guide](#).
- **Manual launch of your TVC plugin on a completed run report** – See the [Torrent Browser Analysis Report Guide](#).

**Survey**  
Share your opinion about Torrent Variant Caller: [Torrent Variant Caller Survey](#)

**Documentation**

- [The TVC Pipeline Overview](#)
- [Getting Started](#)
- [What is New in TVC 4.0](#)
- [The TVC Results Page](#)

Supports Ubuntu 10.04.

- Compile successfully on
- Debian 7 (Wheezy) and newer
  - Ubuntu 10.04, 11.10, 12.04, 13.10, 14.0
  - CentOS 6.4
  - CentOS 5.9

Both mapping and variant calling can now run outside of Torrent Suite on your own compute cluster or in Galaxy

# Torrent Server Virtual Machine for TSS 4.2

- Torrent Server Virtual Machine (TS-VM) is a way for customers to test-drive a newer version of TS without having to upgrade their version of Torrent Suite.
- Some key features:
  - Run Virtual Machine on customer's own Torrent Server
  - Re-analyze existing data without copying any files.
  - Run new plugin-ins on existing Reports.
  - TS-VM is possible for TS 4.0, but command line setup is required.

# TS-VM Analysis Options

- Re-analysis options depend on what File Types are Imported into the TS-VM.

File Type Imported	Re-analysis Possible
Report Files (BAM & FASTQ)	Selected Plugins
Basecaller Input (WELLS)	Basecalling and Alignment
Signal Processing Input (DAT)	Signal Processing

- Re-analyze from Wells
  - User has to Import 'Basecaller Input' category for desired Run.
  - No need to "Copy" files.
- Re-run a Plug-in
  - User has to Import 'Report Files' category for the desired Run.
  - The selected Report shows up in the TS-VM and can be displayed – and plugins executed.

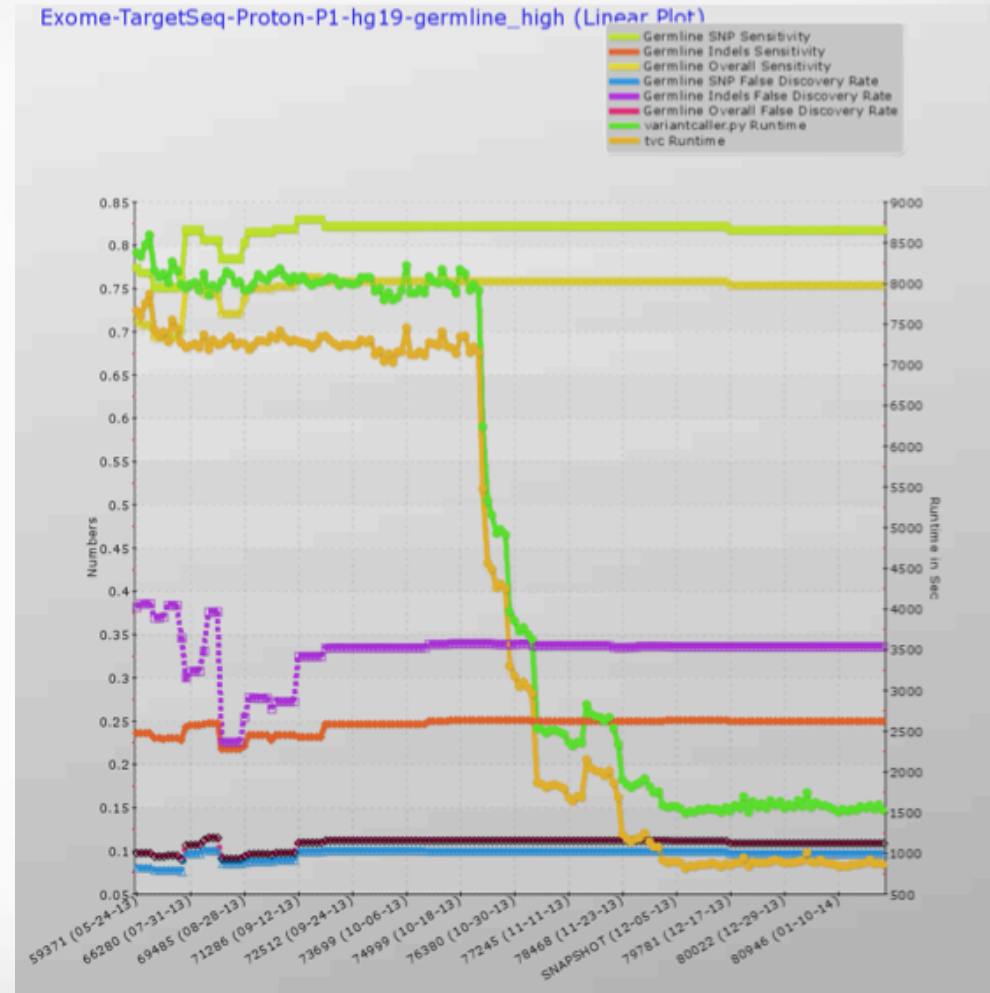


# 12 fold speed boost in TVC v4.2

12 fold speed boost by integrating the **candidate generation** and **evaluation** steps and thereby reducing the rework involved in loading BAM files.

In addition, the primer trimming step will also be optimized and integrated within TVC.

Further improvements will come from integrating and optimizing long index assembly step.



# Ion Reporter™ Software

Simple workflows for one, two, or three samples



## **Single sample analysis**

What is different from a reference genome?



## **Paired analysis**

How do the variants differ between these two samples?



## **Somatic paired analysis**

What variants exist in a tumor but not the germline sample?

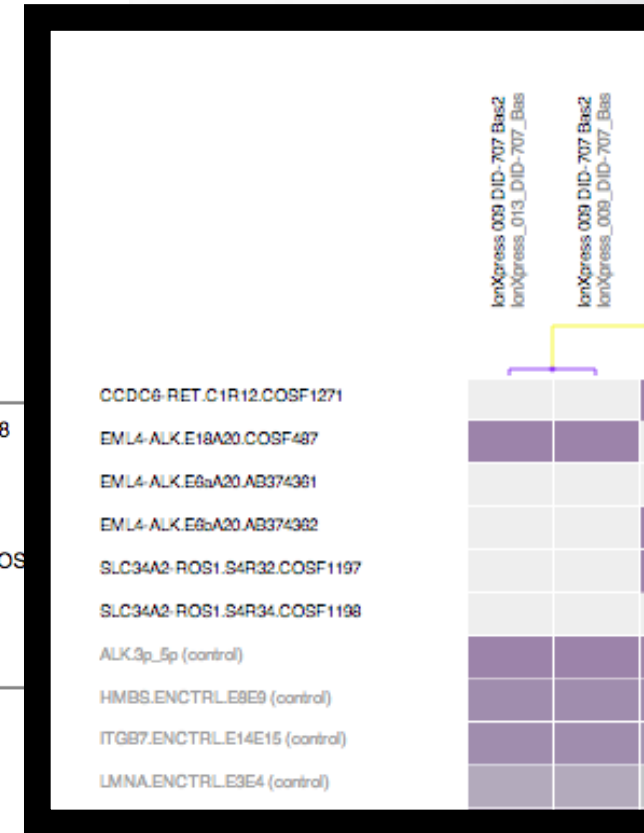
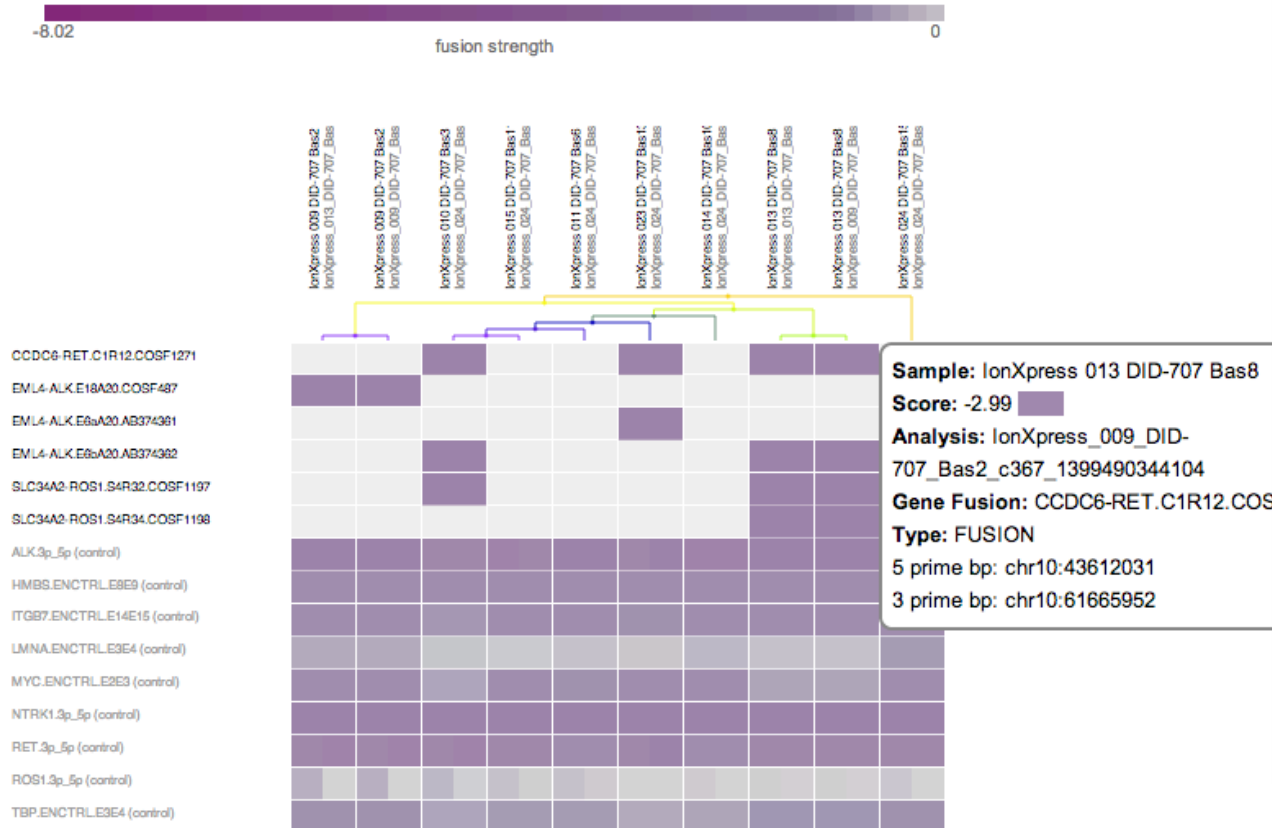


## **Trio exome analysis for inherited diseases**

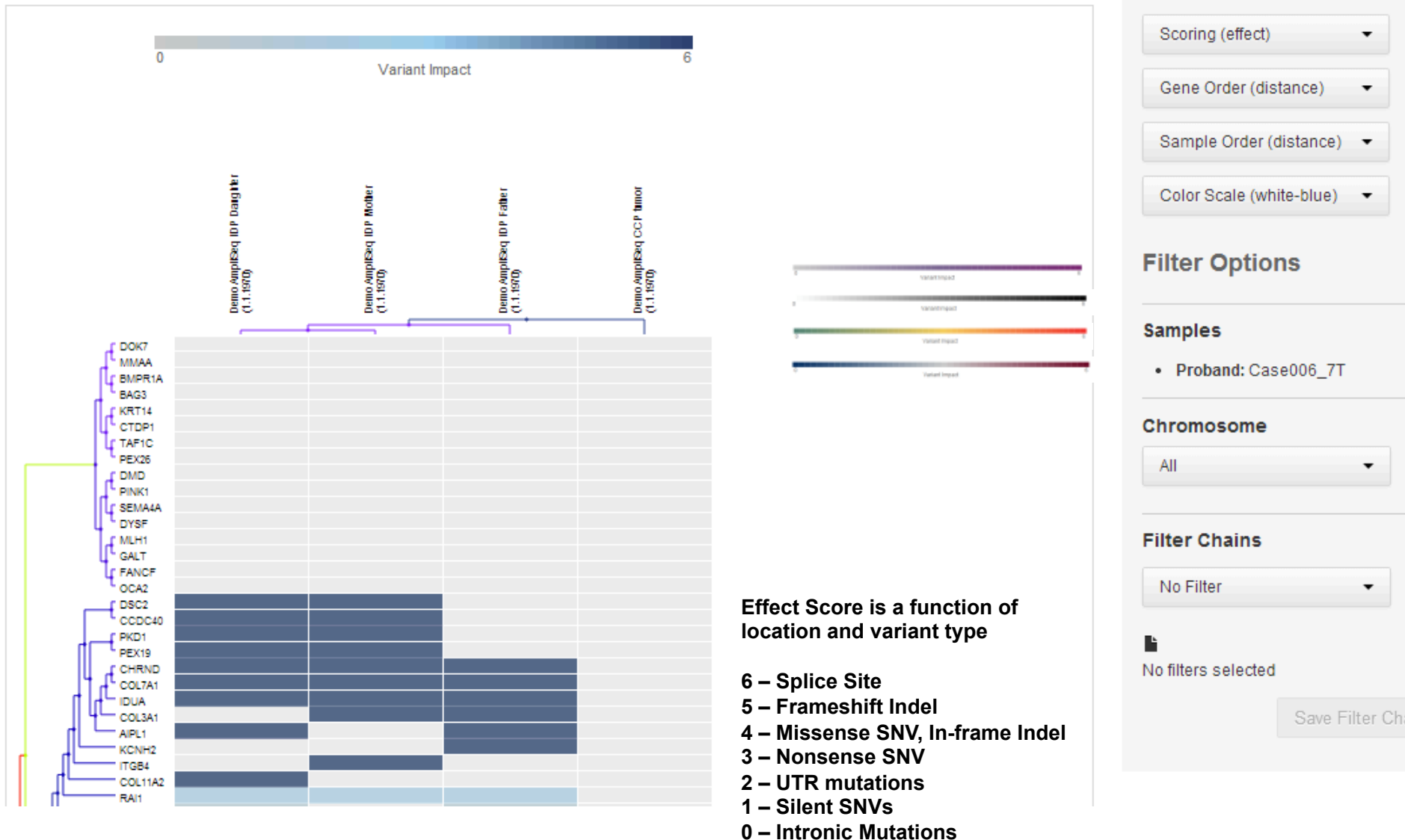
What variants are shared/different across related individuals?

# IR 4.2: Multi-Sample Visualization (Heatmaps)

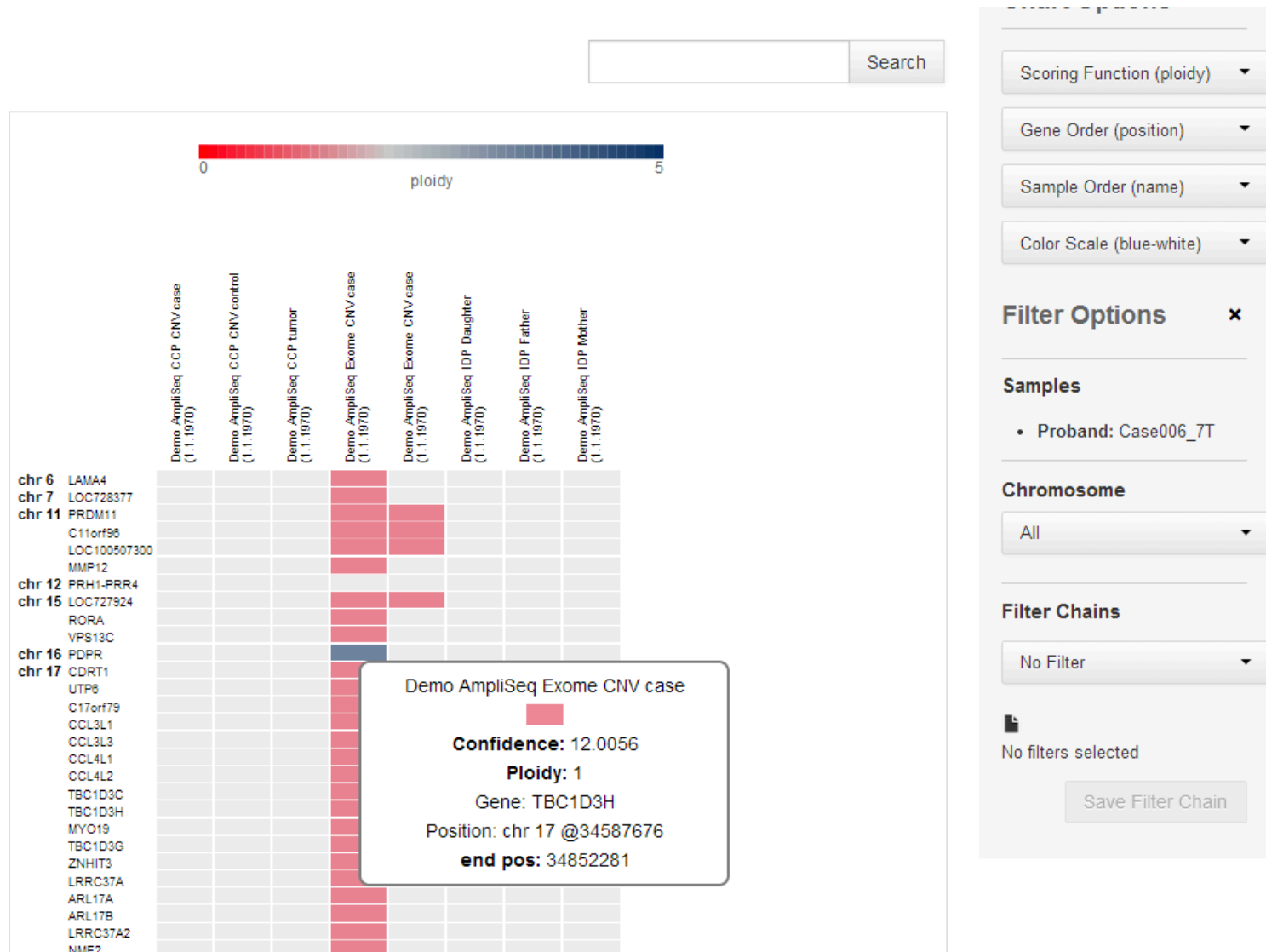
[Variants Table](#)
[Variant Impact](#)
[Gene Fusions](#)
[CNV Heat Map](#)



## Multi-Sample Visualization: Variant Comparison



# Multi-Sample Visualization: CNV



# Sequencing for All™

[lifetechnologies.com/iontorrent](https://lifetechnologies.com/iontorrent)

*life*  
technologies

A Thermo Fisher Scientific Brand