



SMRT Informatics Developers Conference

Kevin Corcoran
SVP Market Development

FIND MEANING IN COMPLEXITY

Thank You!

NIST

**National Institute of
Standards and Technology**
U.S. Department of Commerce



Genome in a Bottle

- Connect to the **NIST-Visitor**
- Username: **SMR05INF05**
- Password: **yCt5)8xamSU^**



You can tweet unless the speaker says “No Tweeting”

Please respect the speaker’s Tweeting policy.



ISMB
2010
BOSTON



18th Annual International Conference on
Intelligent Systems for Molecular Biology

SIGS AND TUTORIALS
July 9-10
CONFERENCE
July 11-13



An Official Conference of the
International Society for
Computational Biology

Pacific Biosciences Launches Software Developer's Network

July 09, 2010 07:30 AM Eastern Daylight Time

BOSTON--(BUSINESS WIRE)--As part of its commitment to introducing third generation DNA sequencing technology to the market, Pacific Biosciences today announced the launch of a software developer's network – the PacBio DevNet – at the Eighteenth International Conference on Intelligent Systems for Molecular Biology (ISMB 2010).

Throughout the development of the company's Single Molecule Real Time (SMRT™) technology platform, Pacific Biosciences has been working closely with members of the informatics community to develop and define standards for working with single molecule sequence data. Now, as the company prepares for the commercial launch of the PacBio RS, it is launching a more formal program to support the needs of the informatics community.

"We look forward to contributing to a robust analytical ecosystem that allows more scientists to exploit the new possibilities enabled by this technology."

The PacBio DevNet was created to support the ecosystem of academic informatics developers, life scientists, and independent software vendors interested in creating tools to work with PacBio's third generation sequencing data. Interested parties can sign up for the PacBio DevNet at www.pacbiodevnet.com, a hub for data sets, source code for algorithms, application programming interfaces (APIs), conversion tools to industry standard formats, and documentation related to SMRT sequencing.

Eric Schadt, Ph.D., Chief Scientific Officer for Pacific Biosciences commented: "Single Molecule Real Time sequencing introduces entirely new dimensions to data, such as a time component, that are unlike anything the bioinformatics community has encountered to this point. Therefore, in addition to a strong internal focus on informatics development, we are committed to supporting third-party software development and facilitating the rapid adoption of this new data type into the scientific community where the really exciting 'big science' can begin to happen."

At the ISMB conference, PacBio scientists and collaborators will present results from some of their informatics development efforts to date, including new algorithms tailored to the unique characteristics of the SMRT data such as its long reads. Pacific Biosciences has also developed a suite of data management and analysis software tools that mimic the granularity, scalability and functionality of the PacBio RS. These informatics solutions are designed to efficiently integrate with the user's LIMS system, making them accessible not only to high-end informatics researchers, but also to biologists and clinical researchers.



Pacific Biosciences DevNet

DevNet supports analysis of sequence data produced from the Pacific Biosciences DNA sequencing platform. Open source software, data, documentation and tips for PacBio sequencing can be found here.

News and Features

SMRT Analysis v2.3.0 is now released. It includes improvements for Long Amplicon Analysis, Iso-Seq™, Quiver and CCS.

Together with performance enhancements, these advances boost accuracy of results, speed up analysis, and provide more options for analyzing amplicon pools of mixed sizes such as full-length HLA Class I and II genes.

SMRT® Analysis v2.3.0 includes:

- Greatly improved performance and robustness across all protocols
- Analysis calibration for the new P6 Polymerase with C4 Sequencing Chemistry
- New intuitive user interface for Iso-Seq™
- Analysis support for a greater range and mix of amplicon sizes (3-9 kb) such as pooling full-length HLA class I and class II genes

DevNet Users:

- Analysis protocol for Full Length Viral DNA – Minor Variant – HIV Use

Contents for Developers

Share, learn about, discover and explore software and data for SMRT Sequencing.

Tutorials

Iso-Seq transcriptome analysis, Bioinformatics, *De novo* assembly, Base Modification, Transcripts, Barcoding

Datasets: Ecoli, Human cDNA, Human 10x, and others

Discussion: SMRT Analysis issues, SEQanswers

File format specifications

base file, alignment, metadata, statistics, run definition

Software

PacBio Software

Compatible third-party software



Datasets

John Harting edited this page on Apr 21 · 93 revisions

Here are some sample datasets so that you can explore PacBio® sequence data as well as file types generated by the PacBio *RS* and SMRT® Analysis.

PacBio RS II P6/C4 Sequencing Chemistry

- [Caenorhabditis elegans 40x Sequence Coverage](#)
- [HLA Multiplexed GenDx Amplicons HLA-A, -B, -C, -DQB1, and -DRB1](#)
- [E. coli size selected 20kb library](#)

PacBio RS II P5/C3 Sequencing Chemistry

- [Human 54x Sequence Coverage](#)
- [Human 10x Sequence Coverage](#)
- [Human microbiome mock community shotgun sequencing data](#)
- [Arabidopsis Sequence and Assembly \(*Arabidopsis thaliana* Ler-0\)](#)
- [Arabidopsis Sequence and Assembly \(*Arabidopsis lyrata*\)](#)
- [Drosophila Sequence and Assembly](#)
- [Neurospora Crassa \(Fungus\) Genome, Epigenome, and Transcriptome](#)

Release of First PacBio Large Genome Dataset



The Assemblathon
@assemblathon

Follow

PacBio parrot data is now available for Assemblathon 2: korflab.ucdavis.edu/Datasets/Assem See the README here: korflab.ucdavis.edu/Datasets/Assem

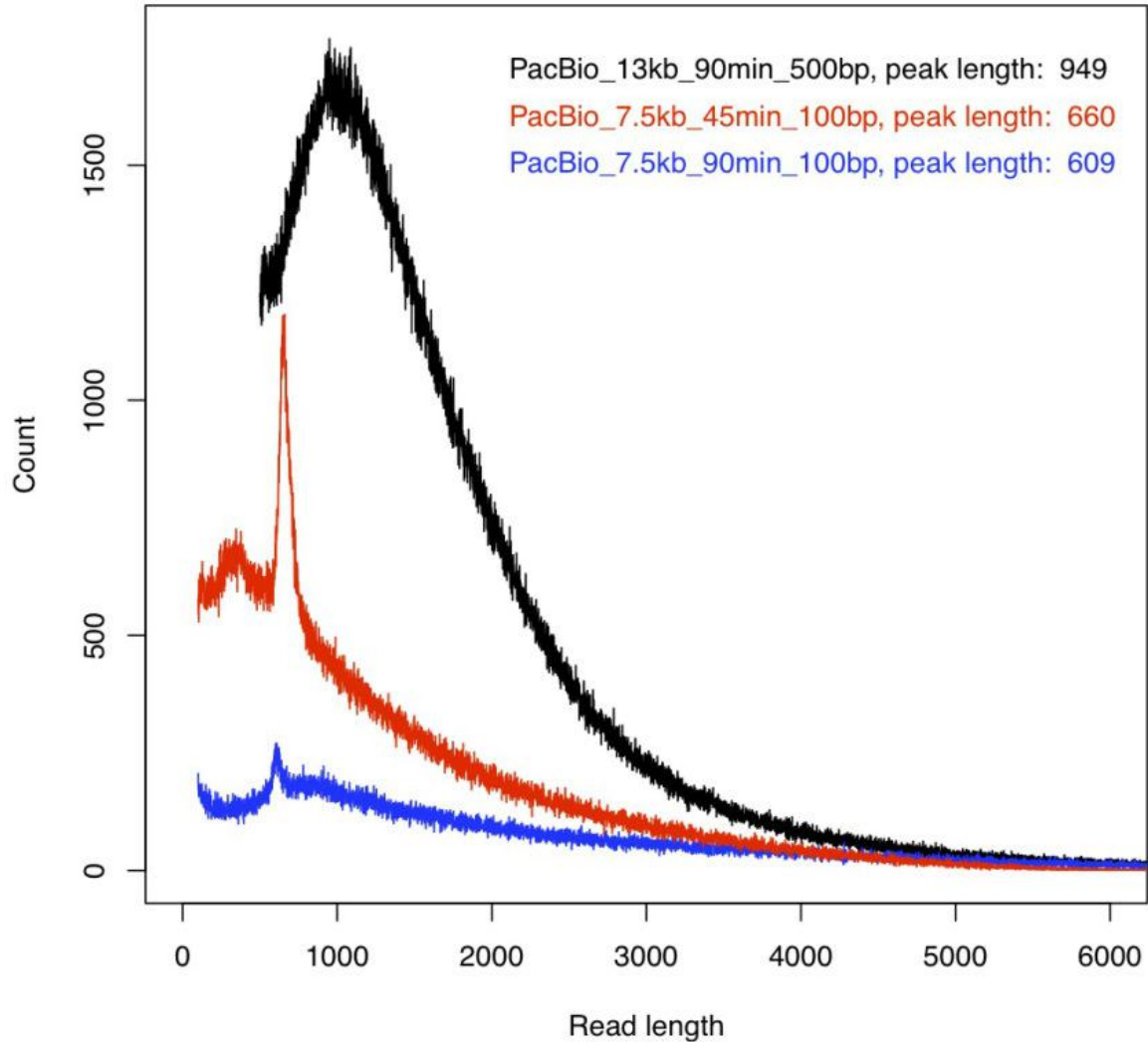
RETWEETS
4



10:48 AM - 2 Jul 2011



Early Chemistry - 2011



First Community Developed Software for PacBio Data

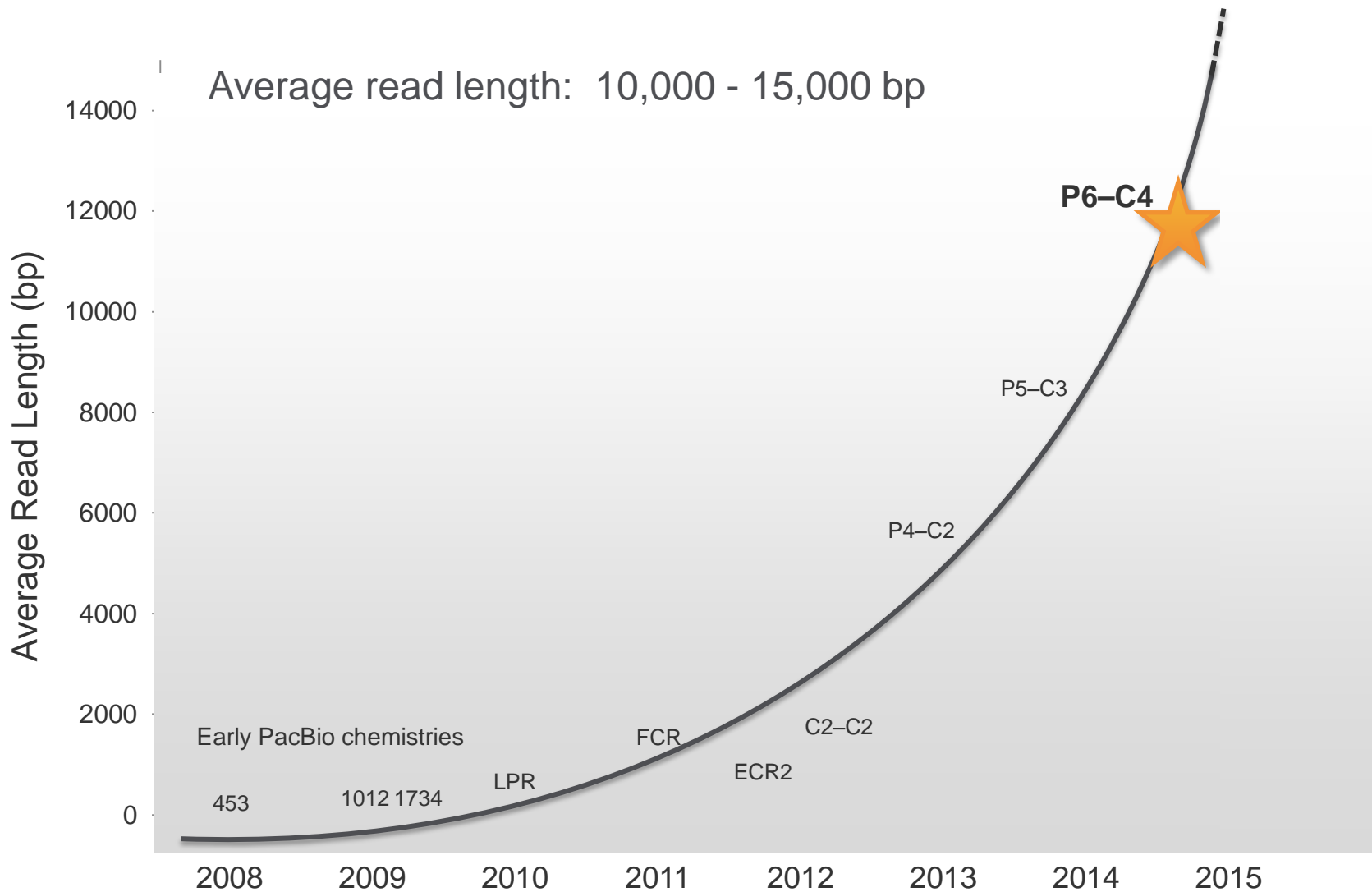
PBcR Pipeline -> MHAP



GATK



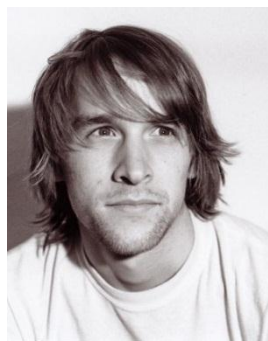
Current Sequencing Chemistry P6-C4



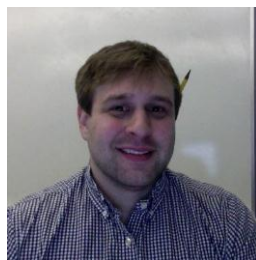
More Community Tools Were Developed



PacMonSTR



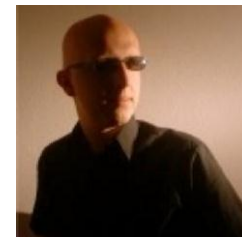
PBJelly
PB Honey



PB Honey



Dazzler
Celera Assembler



Mira



Allpaths-LG



AgIn



Blasr
SMRT-SV



AgIn



Lumpy



Lumpy



IsoSeq Tools



MatchAnnot



LSC



Delly



MHAP



FinishSC

Technology Improvements

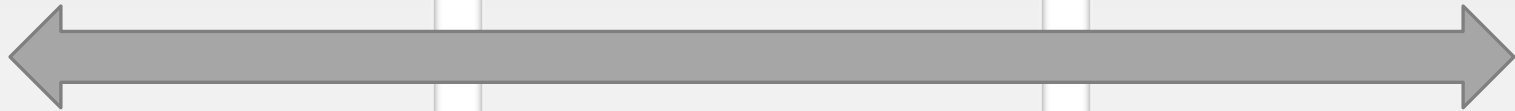
Sample Preparation



Sequencing



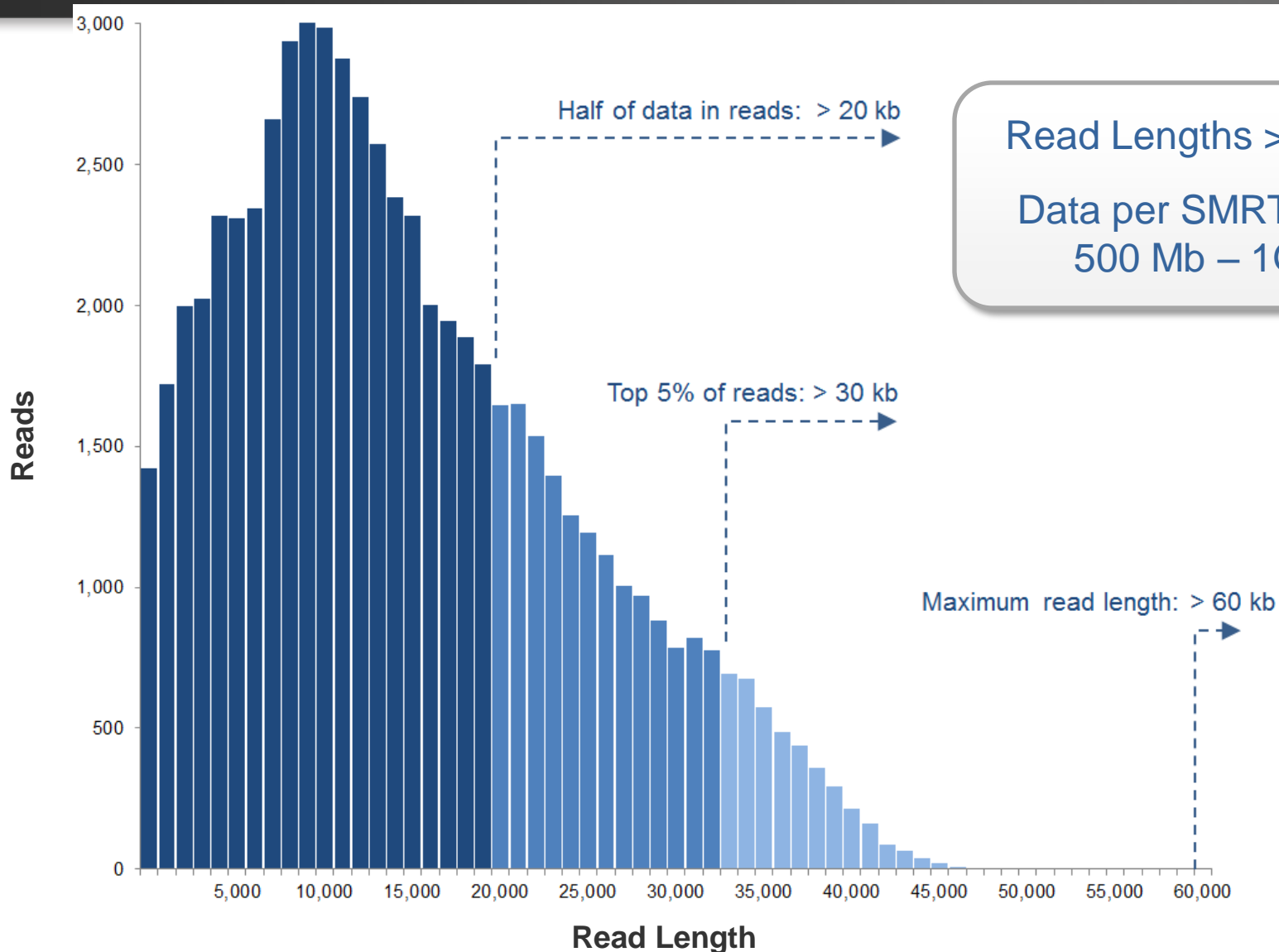
Data Analysis



6 Hour Movies

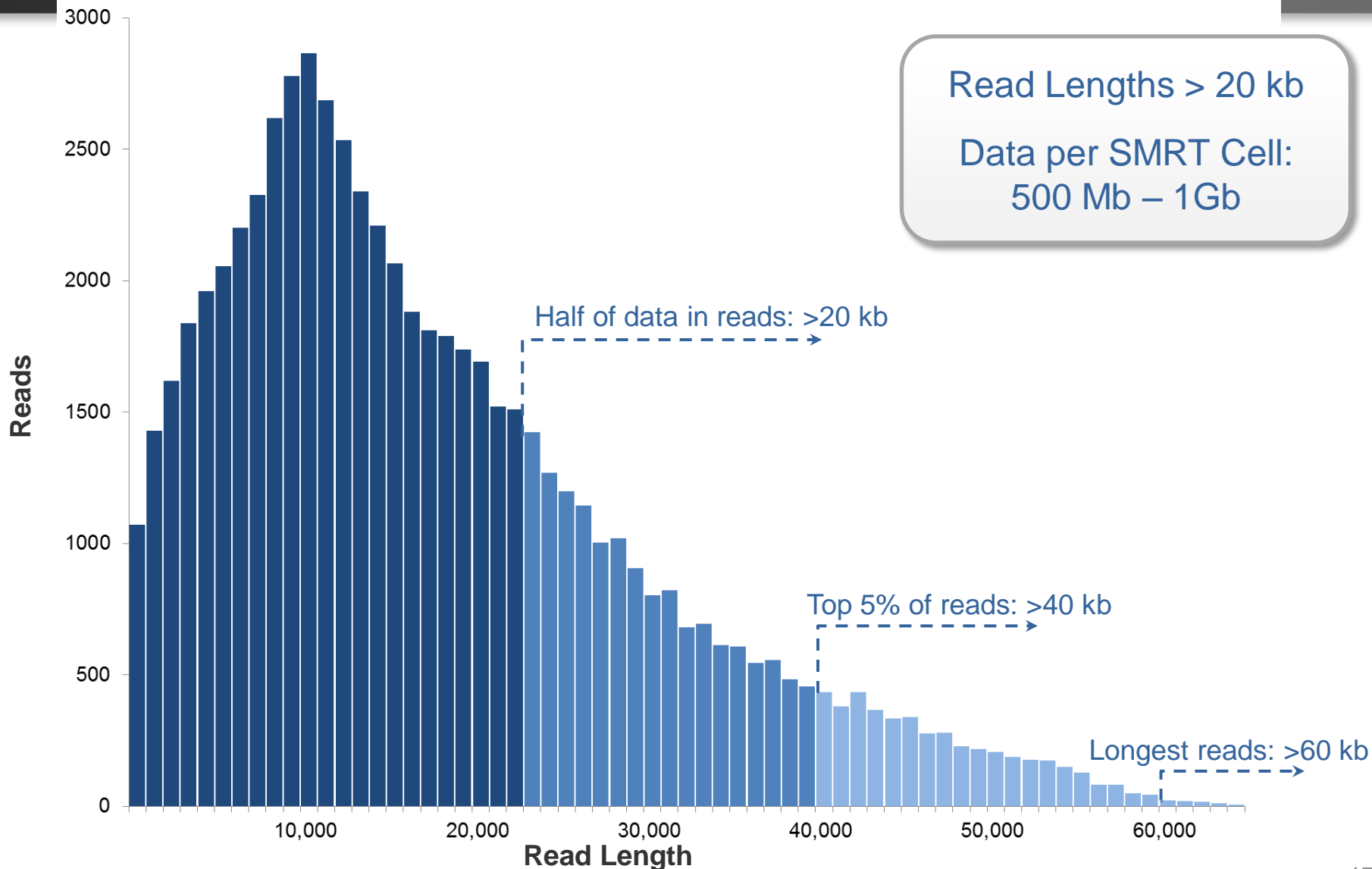


Long Read Lengths with P6-C4 & 4 Hour Movies



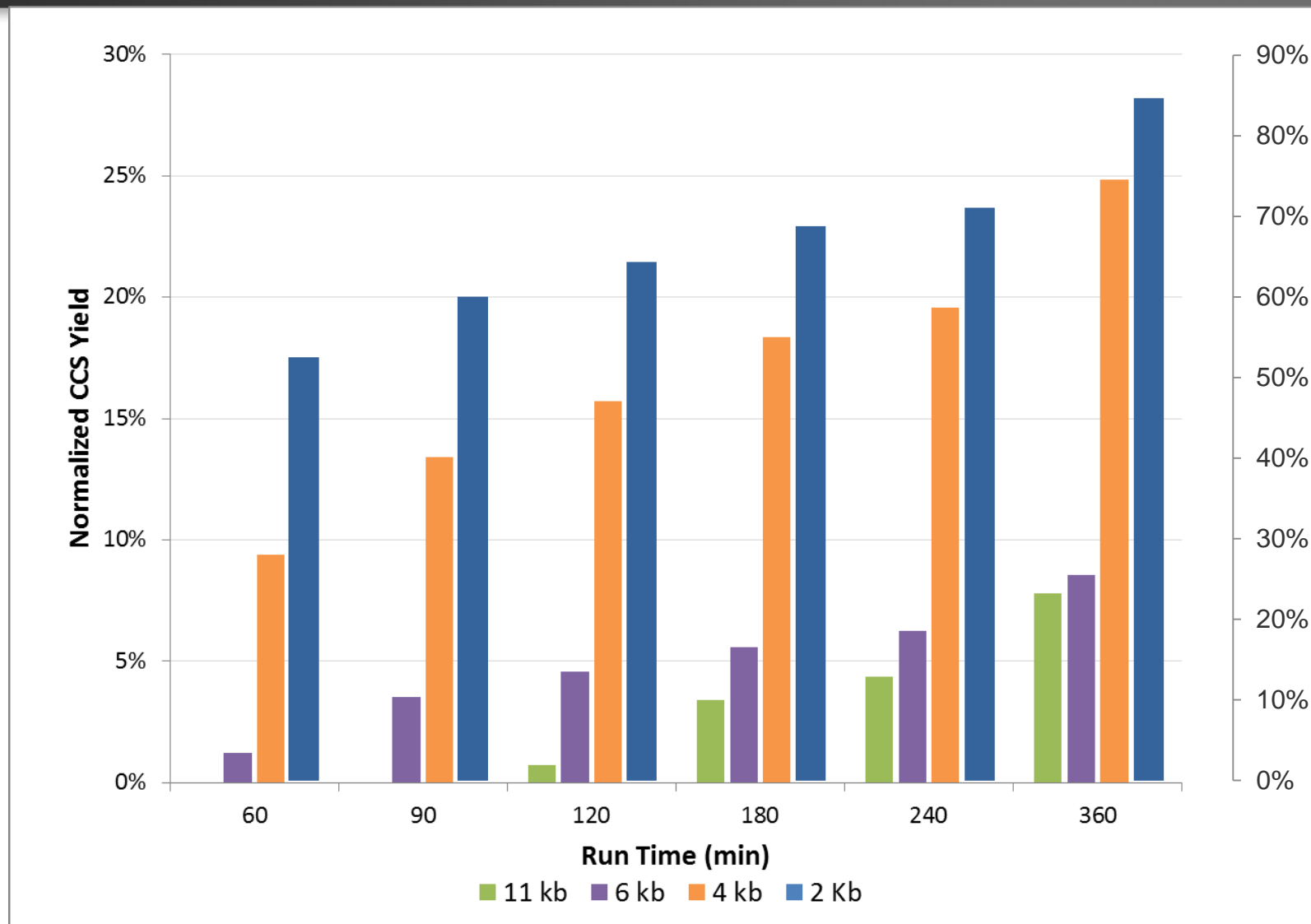
Data generated with 20 kb size-selected human library using 4 hour movies with P6-C4 chemistry, analyzed with SMRT® Analysis v 2.3. Each SMRT Cell generates ~55,000 reads.

Long Read Lengths with P6-C4 & 6 Hour Movies

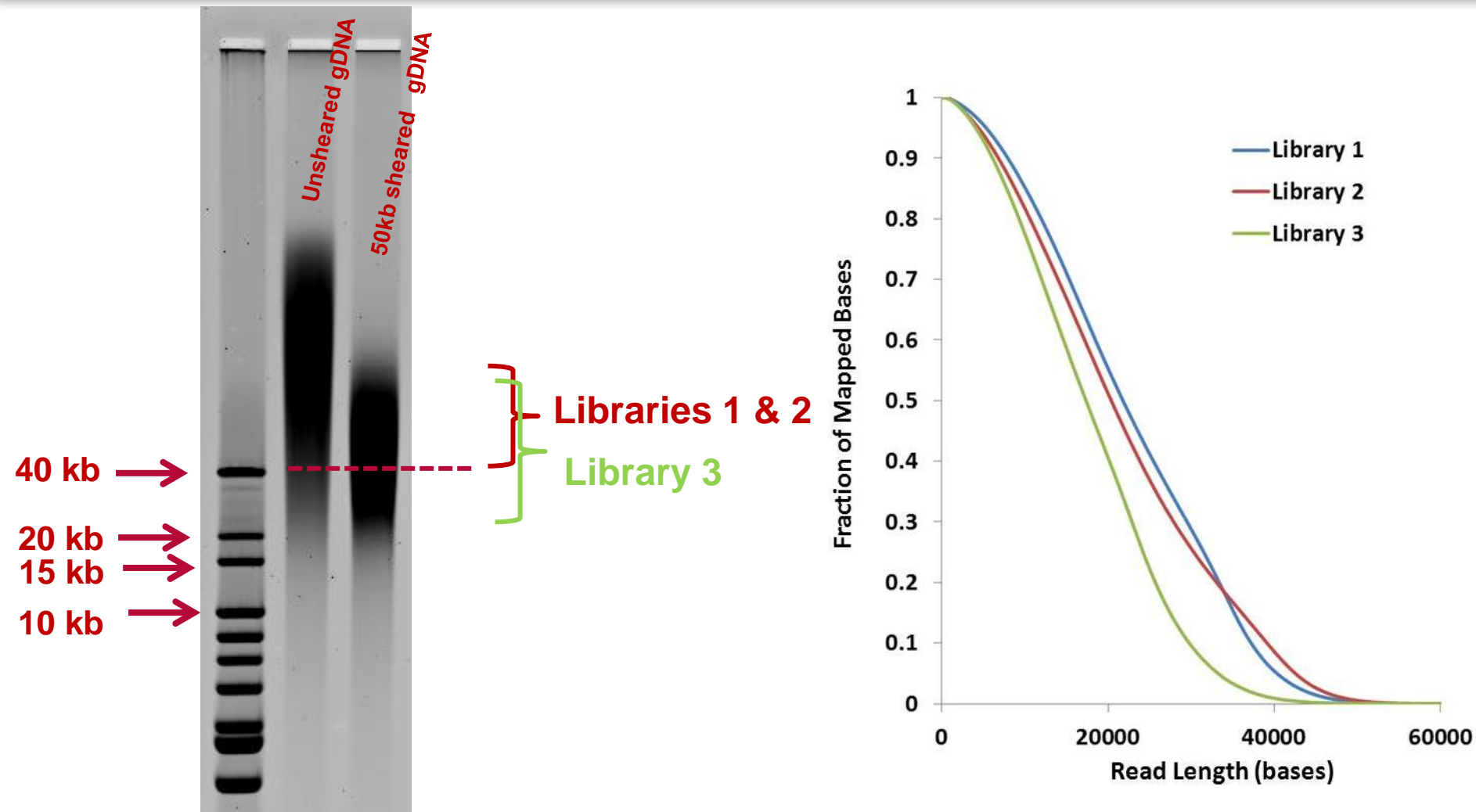


Data generated with 20 kb size-selected human library using 6 hour movies with P6-C4 chemistry, analyzed with SMRT® Analysis v 2.3. Each SMRT Cell generates ~55,000 reads.

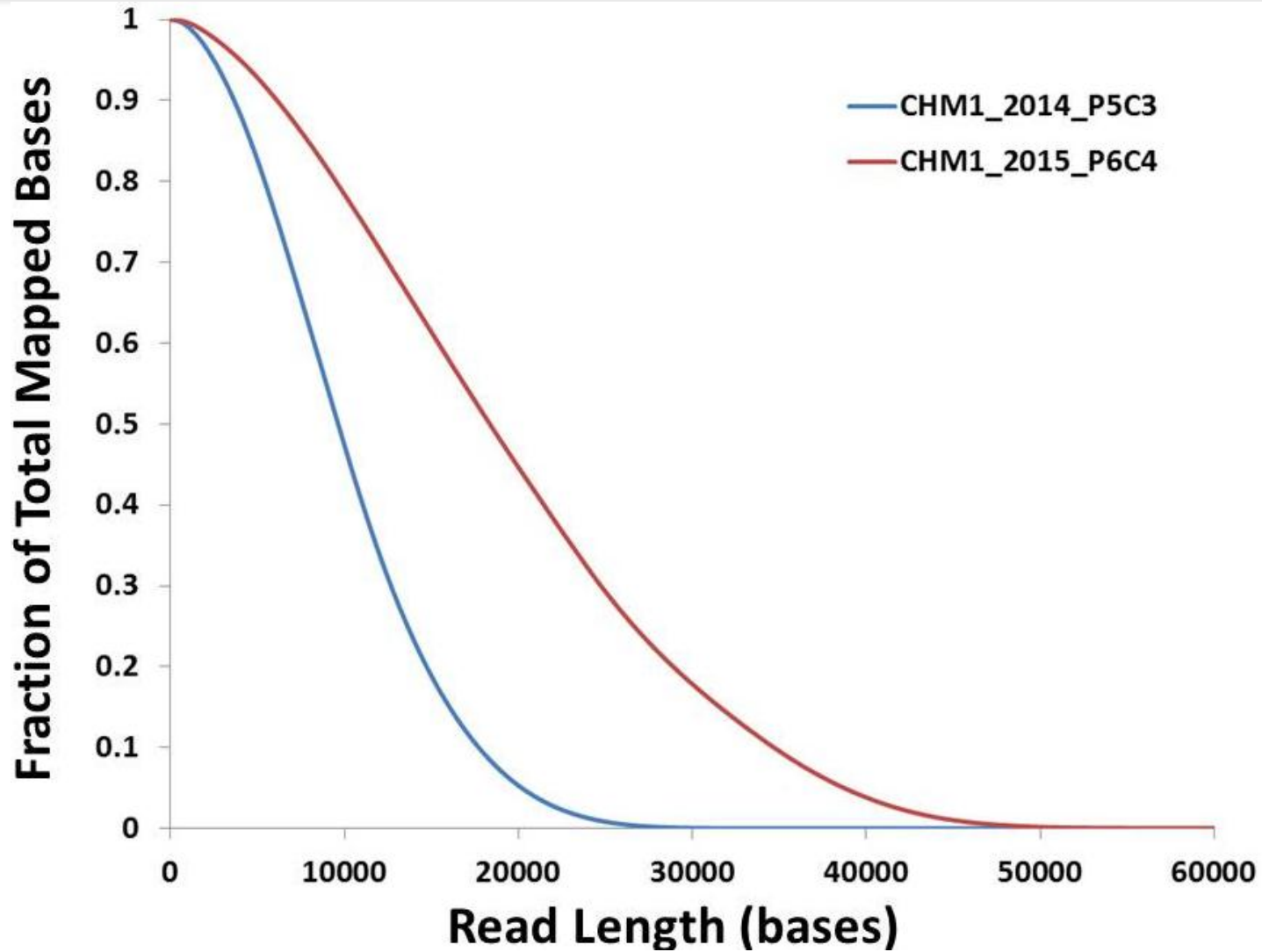
CCS Yield



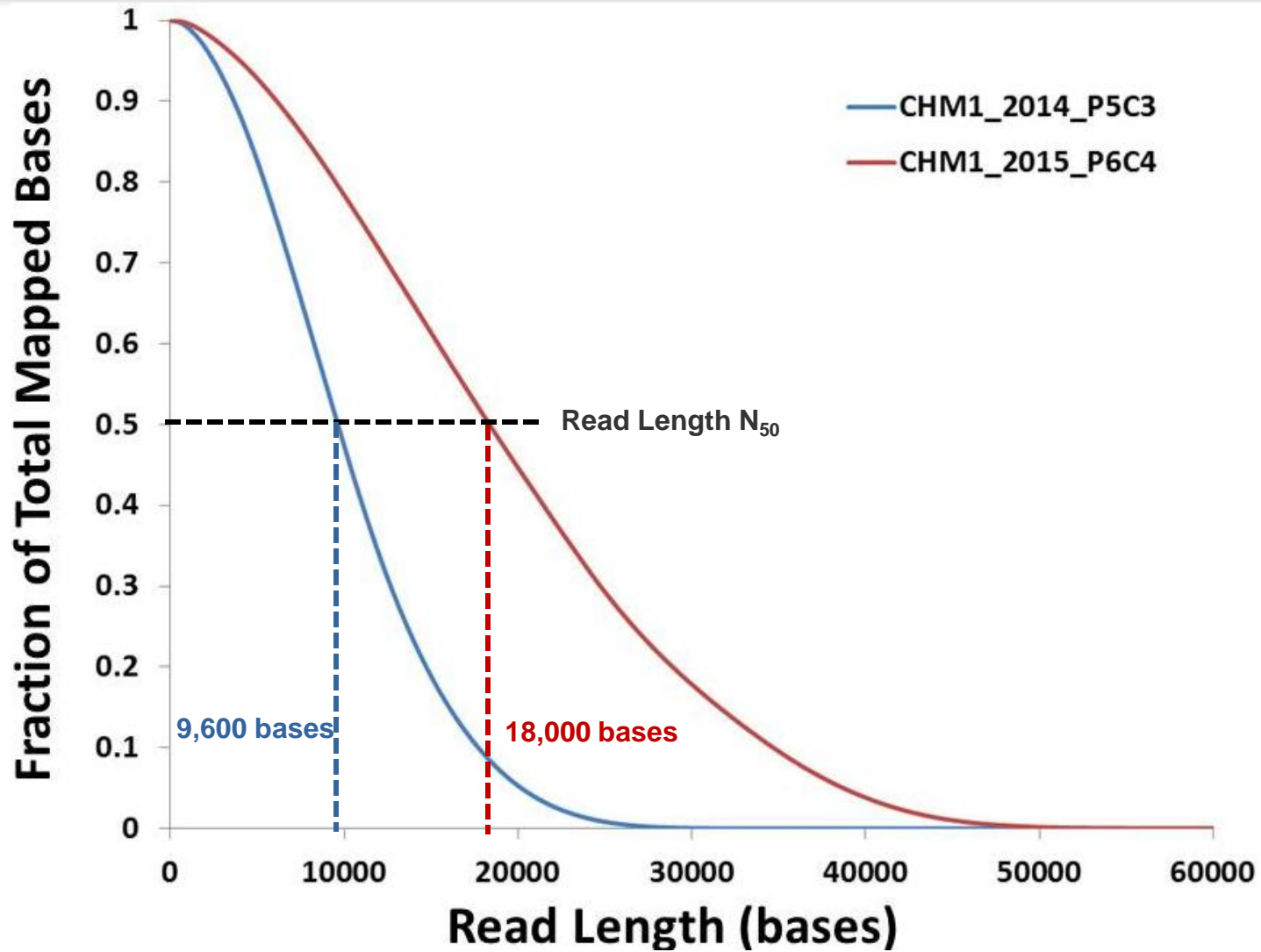
Moving Beyond 20kB Libraries



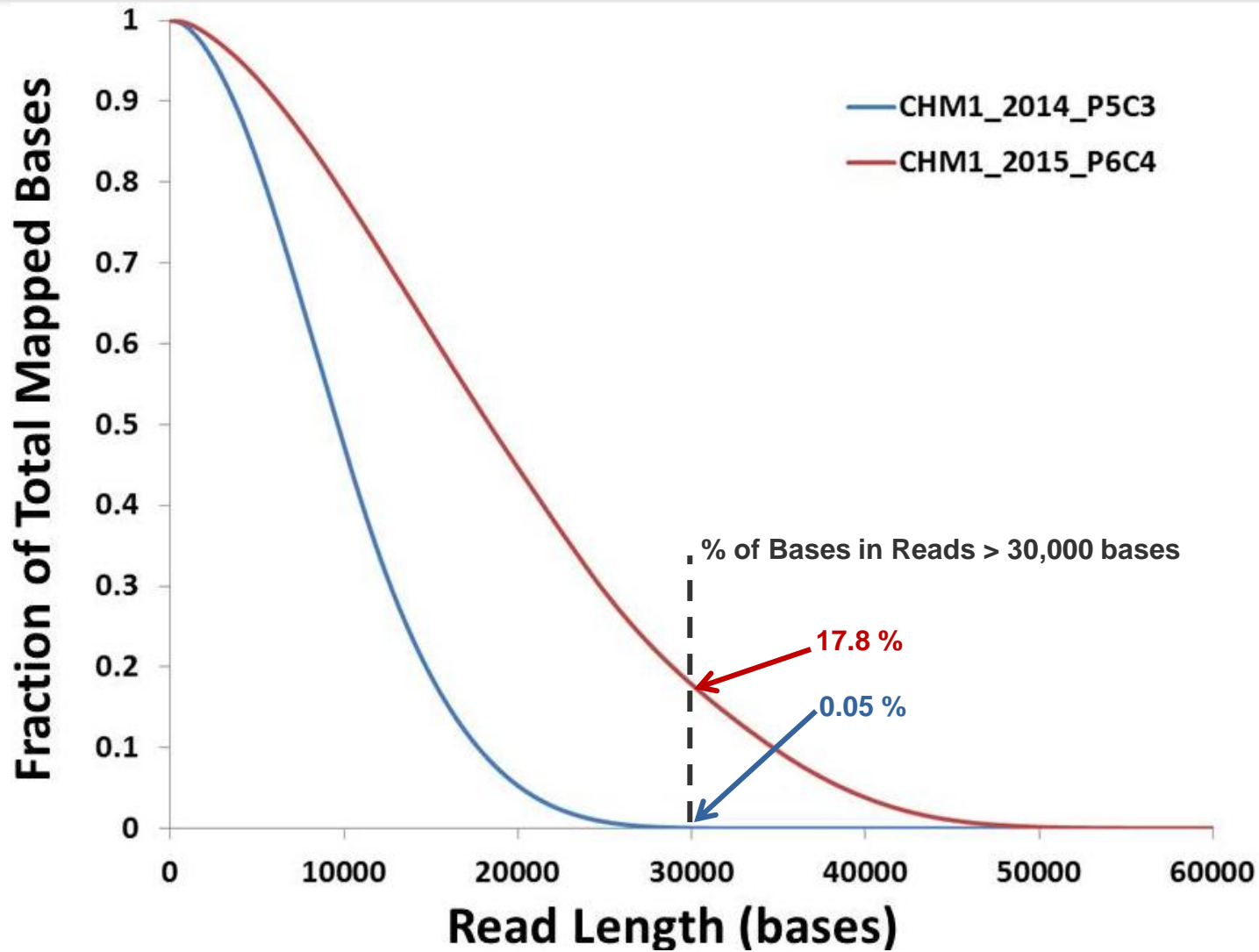
More Bases in Longer Reads for Assembly



More Bases in Longer Reads for Assembly

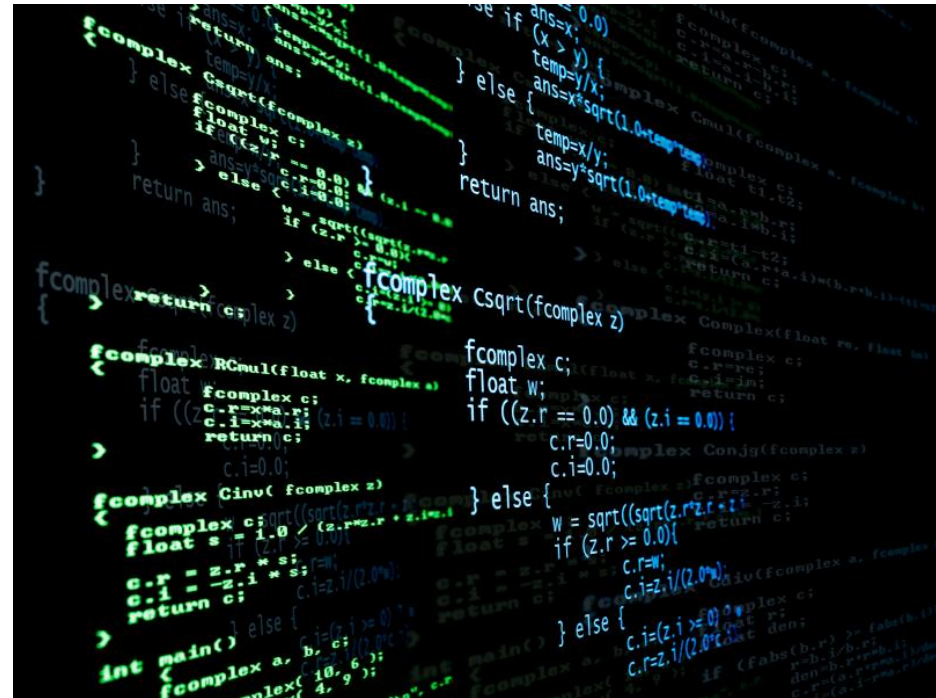


More Bases in Longer Reads for Assembly



Continuous Collaboration with Informatics Community

- SMRT Informatics Developers Conference – 2016
 - In ~ 6 months –2016, place TBD
 - Future data releases - customers, PacBio, collaborations
- Follow up meetings for each application area
- Google groups for information exchange





Search for groups or messages



Groups

CREATE GROUP

My groups

Home

Starred

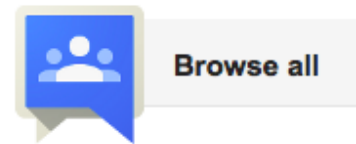
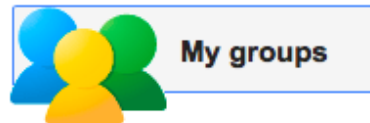
▼ Favorites

Click on a group's star icon to add it to your favorites

▼ Recently viewed

Low Cost Tools

[Privacy](#) - [Terms of Service](#)



All of your discussions in one place

Organize with favorites and folders, choose to follow along via email, and quickly find unread posts.

Express yourself

Use rich-text editing to customize your posts with fonts, colors, and images.

People power discussions

Use photos, nicknames, and automatic translations to share your thoughts with the world.

Speed matters

Keyboard shortcuts and a streamlined design mean you won't spend time waiting to read and get involved. Press

Google groups:

- SMRT_denovo
- SMRT_sv
- SMRT_isodeq
- SMRT_kinetics

Thank you

National Institute of
Standards and Technology

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce



Genome in a Bottle