Introduction to Trinity RNA-Seq

Berlin, June 2017

Brian Haas
Broad Institute

Physalia
Courses

Welcome to the Berlin 2017 Trinity Workshop Wiki!

Day	Time	Activities
Monday, June 12	morning	Workshop Introduction
		Exploring the Computational Infrastructure
	afternoon	Unix command-line review
		Data overview and setup
		Using FASTQC and Trimmomatic
Tuesday, June 13	morning	Trinity de novo transcriptome assembly
		Expression quantification
	afternoon	Quality assessment for assembly
		QC samples and replicates
Wednesday, June 14	morning	Statistical methods for differential expression analysis
	afternoon	Transcript clustering and expression profiling
Thursday, June 15	morning	Methods for functional annotation
	afternoon	Trinotate and TrinotateWeb
	afternoon	Functional enrichment analysis
Friday, June 16	morning	Review and custom data analyses

https://github.com/trinityrnaseq/BerlinTrinityWorkshop2017/wiki

Generating RNA-Seq: *How to Choose?*

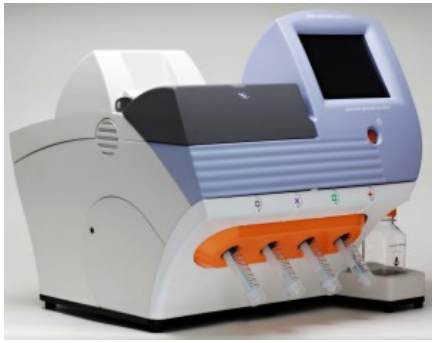
Many different instruments hit the scene in the last decade

Illumina

454

SOLiD

Helicos



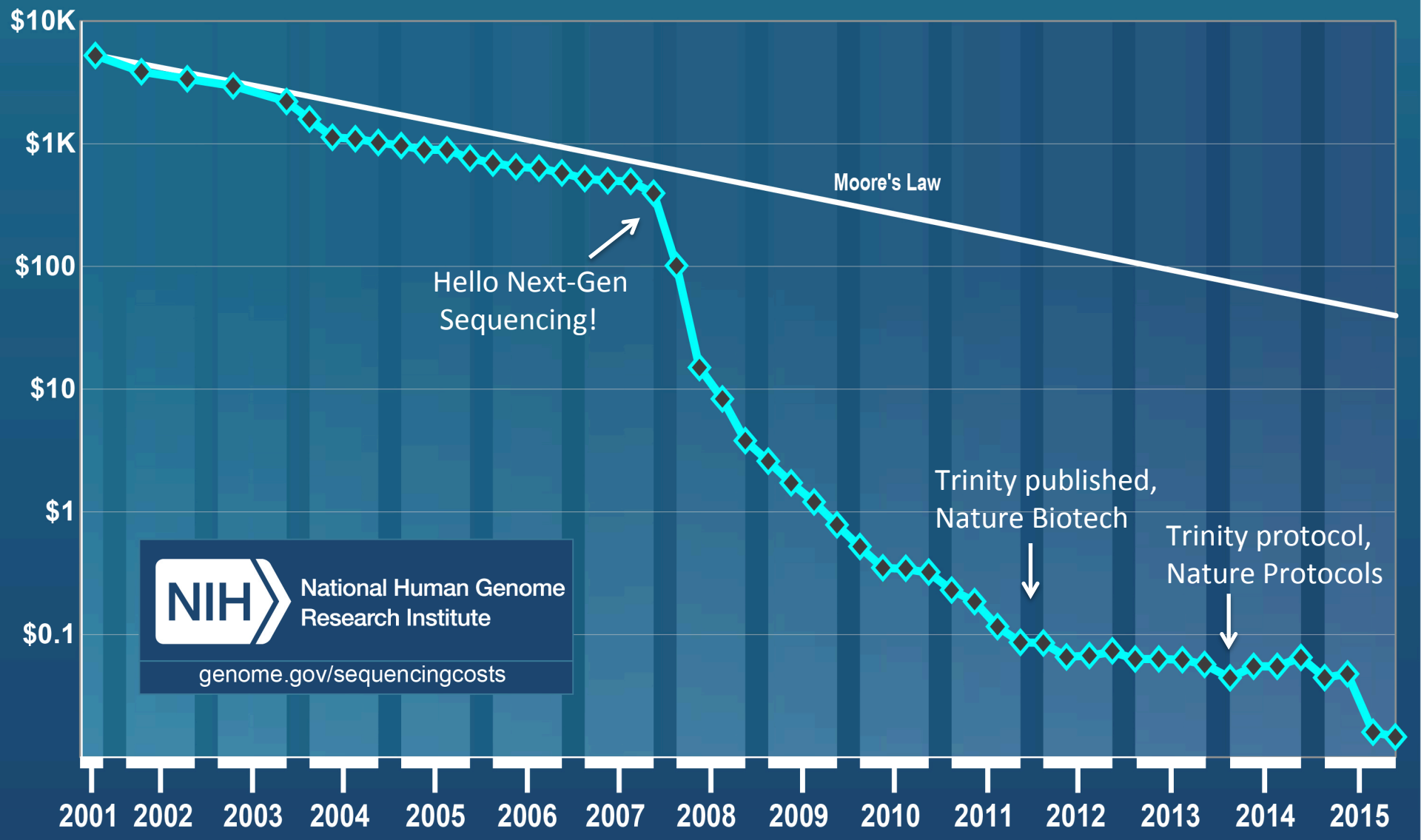
Ion Torrent

Pacific Biosciences

Oxford Nanopore

Slide courtesy of Joshua Levin, Broad Institute.

Cost per Raw Megabase of DNA Sequence

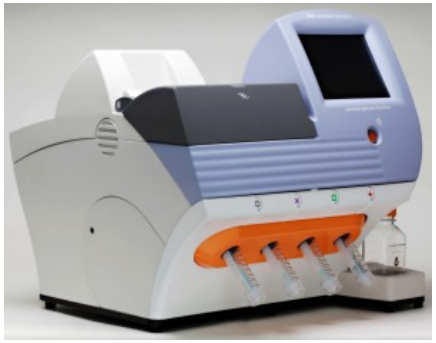


From https://www.genome.gov/sequencingcostsdata/

RNA-Seq: *How to Choose?*

Illumina

Helicos



Ion Torrent

"What I especially like about this baby is this little drawer where I can keep my lunch."

Oxford Nanopore

Slide courtesy of Joshua Levin, Broad Institute.

Generating RNA-Seq: *How to Choose?*

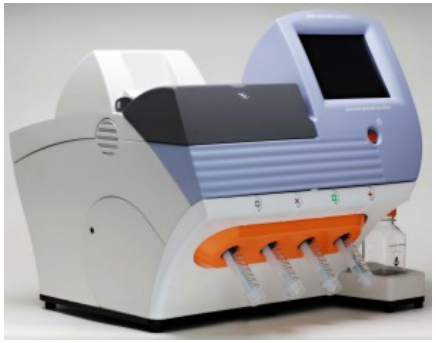
Popular choices for RNA-Seq today

Illumina

454

SOLiD

Helicos



Ion Torrent

Pacific Biosciences

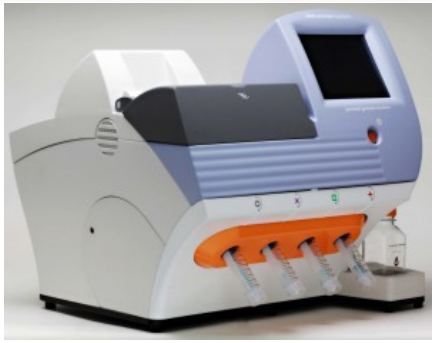
Oxford Nanopore

Generating RNA-Seq: *How to Choose?*

Popular choices for RNA-Seq today

[Current RNA-Seq workhorse]

Illumina



Ion Torrent

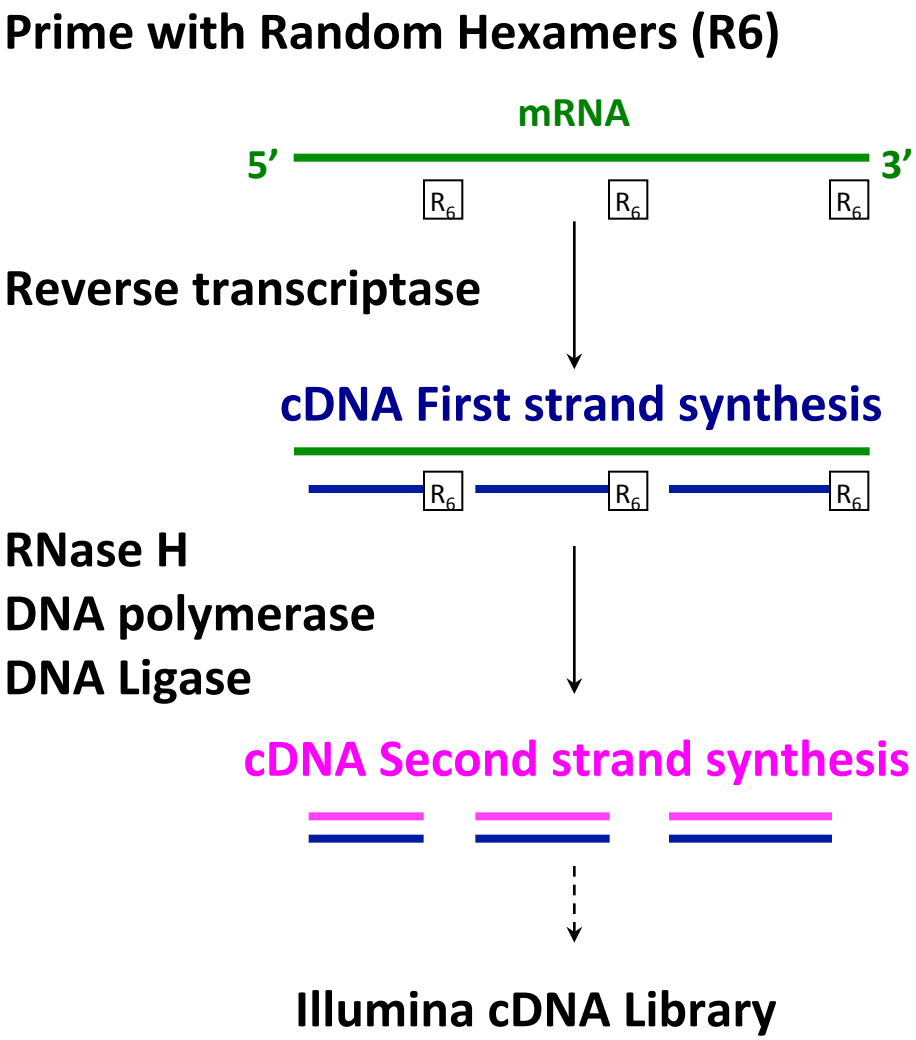
[Full-length single molecule sequencing]

Pacific Biosciences

[Newly emerging technology for full-length single molecule sequencing]

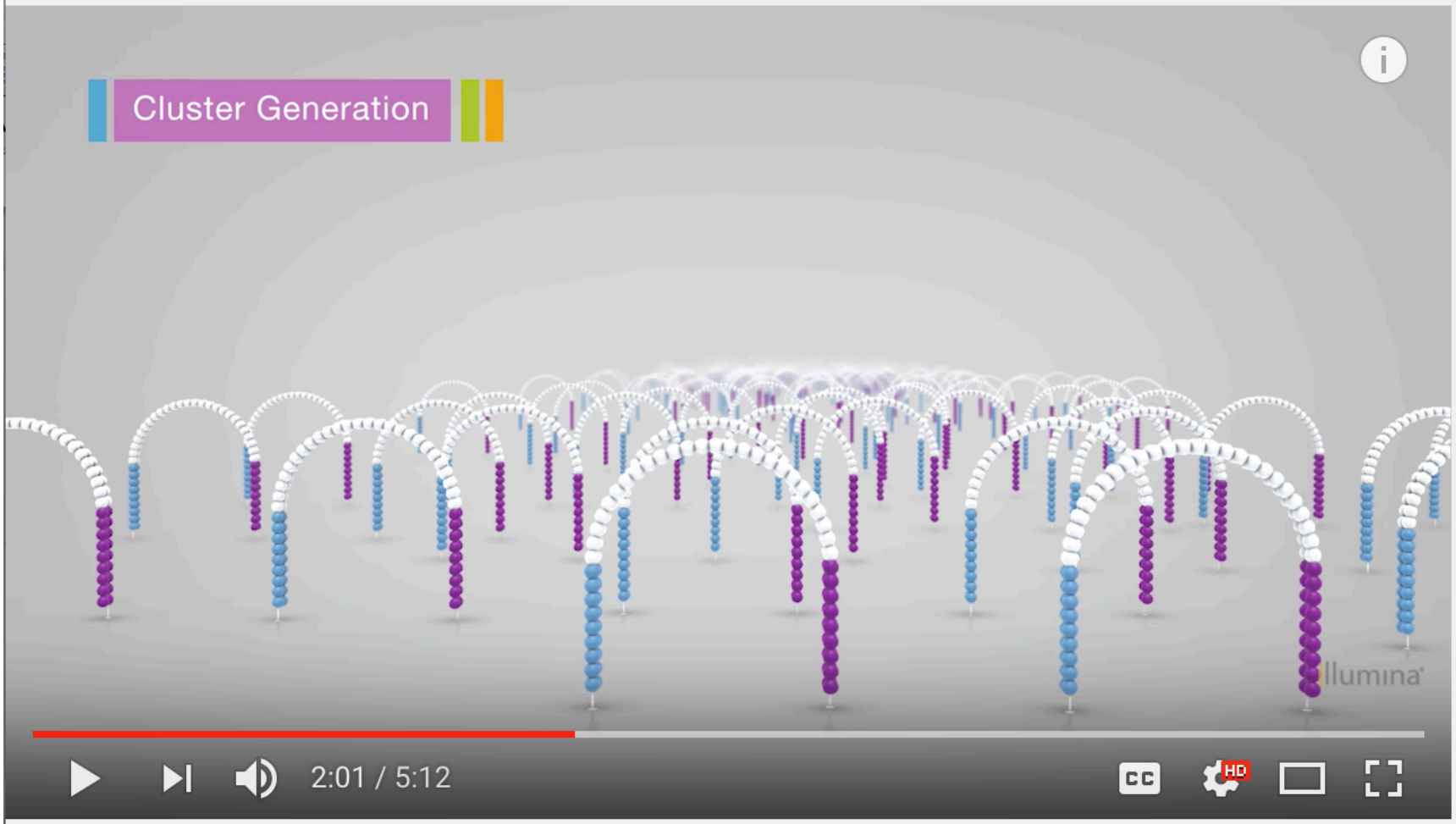
Oxford Nanopore

RNA-Seq: How do we make cDNA?



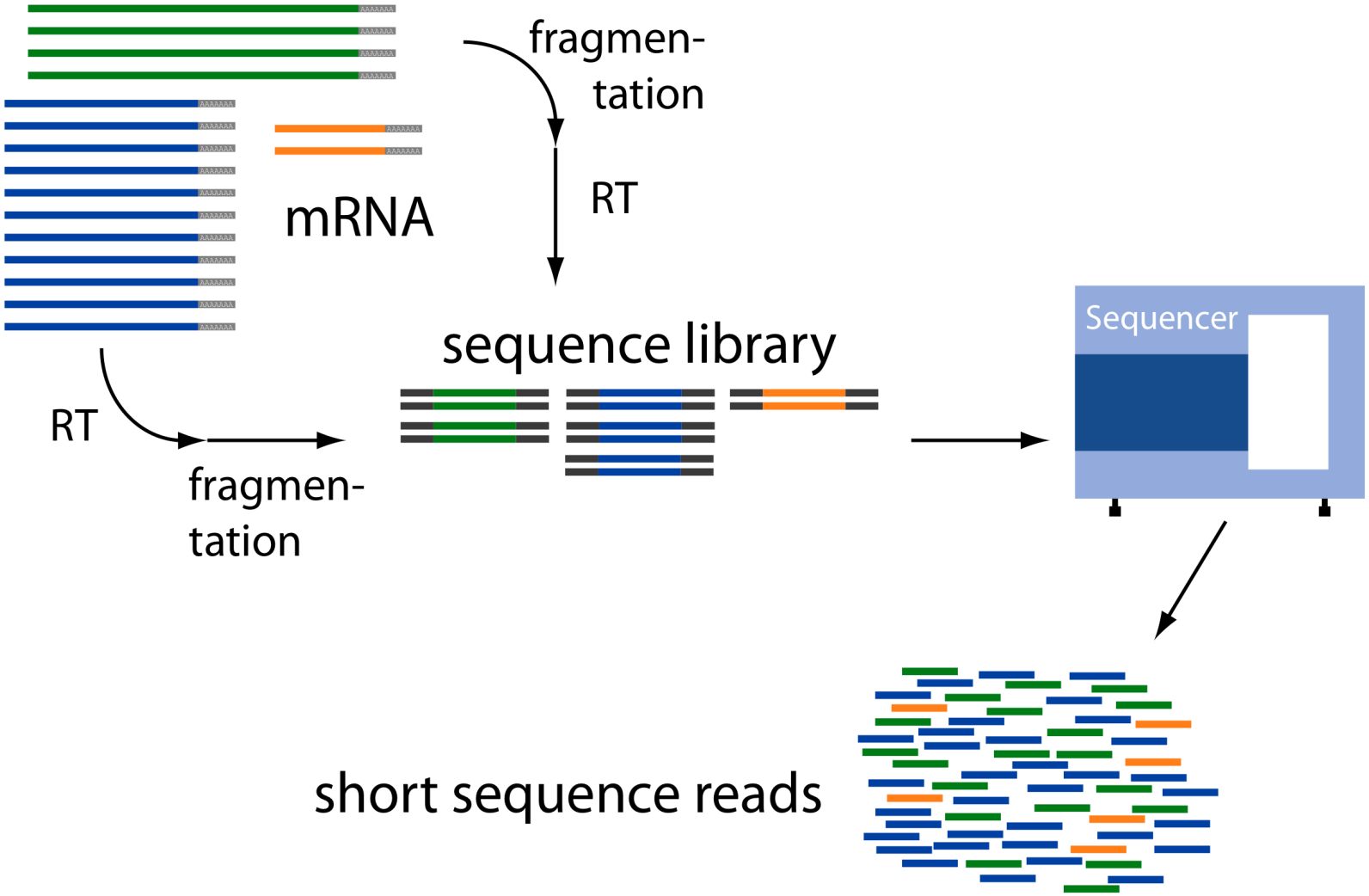
Slide courtesy of Joshua Levin, Broad Institute.

Illumina Sequencing by Synthesis



Illumina Inc https://www.youtube.com/watch?v=fCd6B5HRaZ8

Overview of RNA-Seq



From: http://www2.fml.tuebingen.mpg.de/raetsch/members/research/transcriptomics.html

Common Data Formats for RNA-Seq

FASTA format:

```
>61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
```

FASTQ format:

```
@61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
+
ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBC?CCCCCCCCC@@CACCCCCA
```

Read

Quality values

Interpreting Base Quality Values

Phred_Quality_Value = AsciiEncodedQual('B') – 33 = 30

Phred_Quality_Value = -10 * $\log_{10}$(Pwrong('T'))

Pwrong('T') = 10^(30/-10) = 10^-3 = 0.001

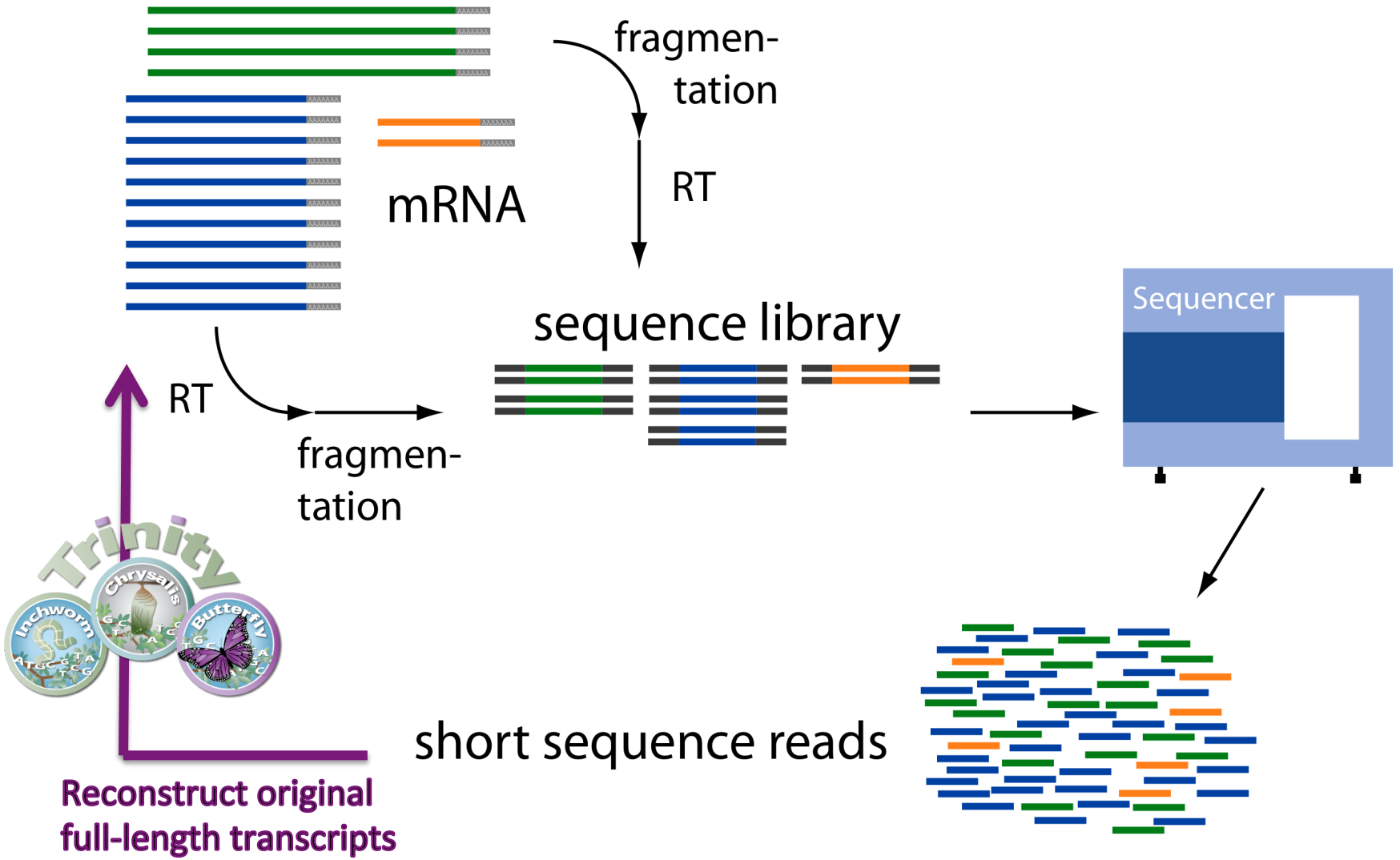
Paired-end Sequences

Two FastQ files, read name indicates left (/1) or right (/2) read of paired-end

```
@61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
+
ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBC?CCCCCCCCC@@CACCCCCA
```

```
@61DFRAAXX100204:1:100:10494:3070/2
CTCAAATGGTTAATTCTCAGGCTGCAAATATTCGTTCAGGATGGAAGAACA
+
C<CCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBCCCC
```

Overview of RNA-Seq



From: http://www2.fml.tuebingen.mpg.de/raetsch/members/research/transcriptomics.html

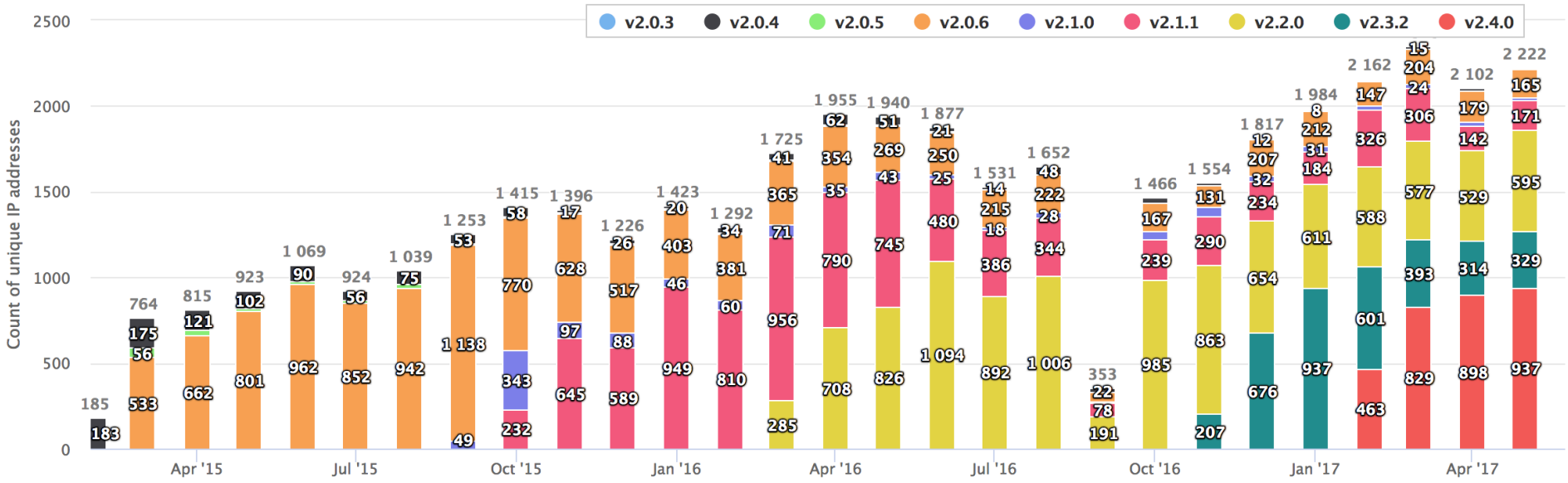
The Ever-Growing Trinity User Community

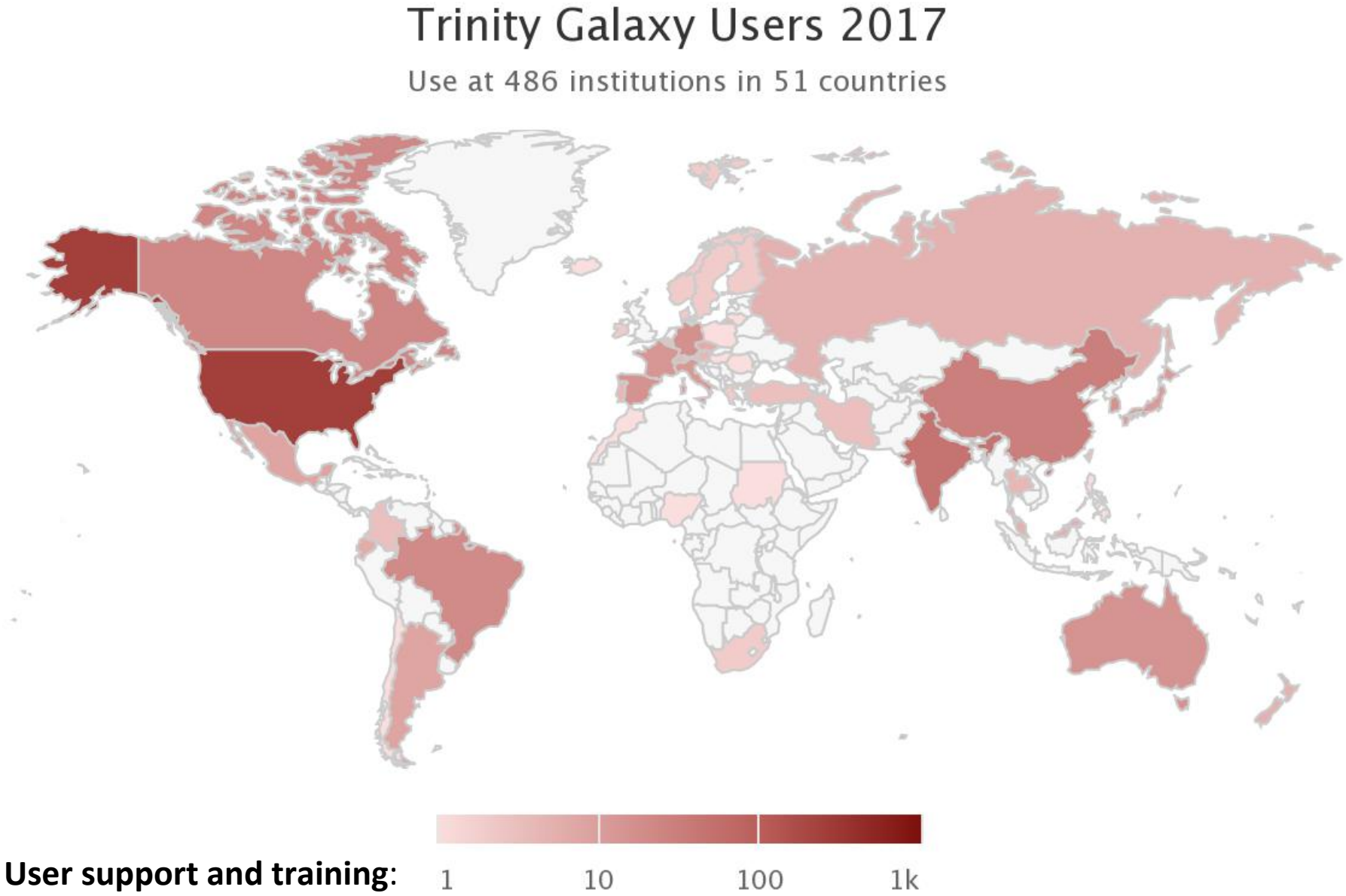
- ~2k unique users per month
- >4k literature citations (~20% cancer community)
- Open Source software development contributions from the Trinity community.

GitHub

http://trinityrnaseq.github.io

Trinity Usage Tracked by Unique IP Address





User support and training:

- Google group for community interaction and support.
- Extensive documentation, user guides, tutorials and protocols
- Demo and training videos
- On-site training workshops

RNA-Seq De novo Assembly Using Trinity

Quick Guide for the Impatient

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Download Trinity here.

Build Trinity by typing 'make' in the base installation directory.

Assemble RNA-Seq data like so:

```
Trinity --seqType fq --left reads_1.fq --right reads_2.fq --CPU 6 --max_memory 20G
```

Find assembled transcripts as: 'trinity_out_dir/Trinity.fasta'

Use the documentation links in the right-sidebar to navigate this documentation, and contact our Google group for technical support.

nature protocols

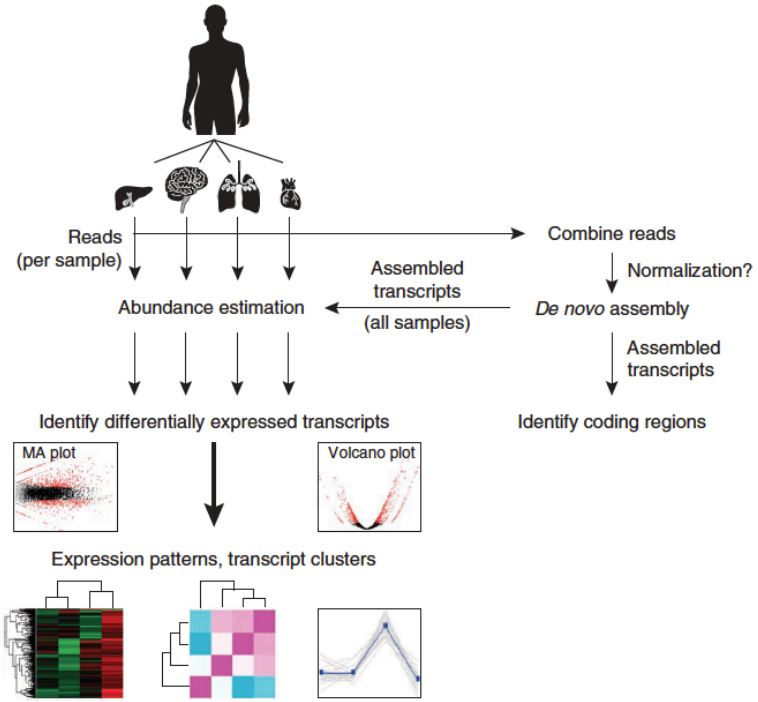
NATURE PROTOCOLS | PROTOCOL

De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis

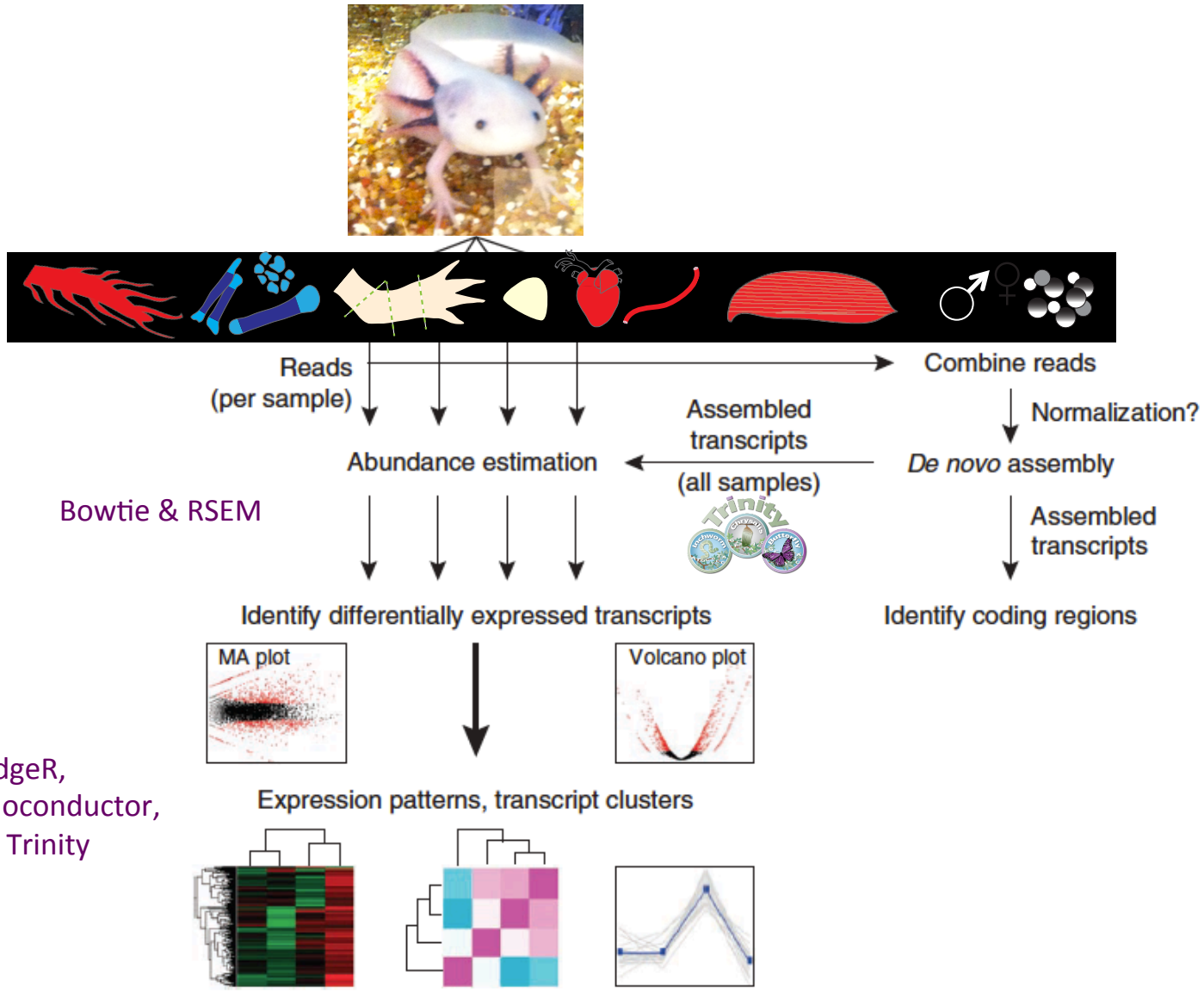
Brian J Haas, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, David Eccles, Bo Li, Matthias Lieber, Matthew D MacManes, Michael Ott, Joshua Orvis, Nathalie Pochet, Francesco Strozzi, Nathan Weeks, Rick Westerman, Thomas William, Colin N Dewey, Robert Henschel, Richard D LeDuc, Nir Friedman & Aviv Regev

Affiliations | Contributions | Corresponding authors

Nature Protocols **8**, 1494–1512 (2013) | doi:10.1038/nprot.2013.084
Published online 11 July 2013



Framework for De novo Transcriptome Assembly and Analysis



1.3 Billion
Total Reads

86 Million
Normalized Reads

Bowtie & RSEM

EdgeR,
Bioconductor,
& Trinity

Example Applications of the Trinity RNA-Seq Protocol

Resource

A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors

Donald M. Bryant[1,6], Kimberly Johnson[1,6], Tia DiTommaso[1], Timothy Tickle[2], Matthew Brian Couger[3], Duygu Payzin-Dogru[1], Tae J. Lee[1], Nicholas D. Leigh[1], Tzu-Hsing Kuo[1], Francis G. Davis[1], Joel Bateman[1], Sevara Bryant[1], Anna R. Guzikowski[1], Stephanie L. Tsai[4], Steven Coyne[1], William W. Ye[1], Robert M. Freeman Jr.[5], Leonid Peshkin[5], Clifford J. Tabin[4], Aviv Regev[2], Brian J. Haas[2], , , Jessica L. Whited[1,7],

Original Article

Loggerhead sea turtle embryos (*Caretta caretta*) regulate expression of stress response and developmental genes when exposed to a biologically realistic heat stress

Blair P. Bentley, Brian J. Haas, Jamie N. Tedeschi, Oliver Berry

Got RNA-Seq?

Run Trinity