

# SEQanswers.com: the next generation sequencing community

Eric C. Olivares, Ph.D.  
eric@seqanswers.com  
Twitter: @SEQanswers  
May 25, 2011

Bay Area Bioinformatics Meetup

# Overview

- Introduction to me
- Introduction to SEQanswers.com
- General library preparation
- Sequencing technologies
- Future of the site
- Questions

# Who am I?



- 1998 UC Davis, BS, Genetics
- 2003 Stanford, PhD, Genetics
  - “Phage integrases for gene therapy: from concept to applications”
- 2003 Founding scientist of startup
- 2005 Invitrogen R&D
  - PureLink™ Genomic DNA
  - RiboMinus™ system
- 2007 Founded SEQanswers.com
- 2008 Pacific Biosciences R&D
  - Sample prep

# A note about perceived conflicts...

- I work at and own stock in Pacific Biosciences
  - Sponsors of this session
  - Didn't edit my talk
- I also run SEQanswers
  - Advertisers...thanks!
- Call bias if you see it, I'm human



# Why SEQanswers.com?

- 2007, Nothing Existed
- Community is key
  - News
  - Opinions
  - Technology discussions
  - Protocol/experience sharing
  - Software tools
- Local is good, worldwide is better
- Hobbies demand forums!



# What is SEQanswers.com?

1. Forum style information resource
2. Wiki organizing more defined information
  - Software tools
  - Sequencing service providers

# Forum style community information resource

- 30 subforums
- Bioinformatics
- Supplier specific forums (ILMN, 454, etc)
- Application specific forums (Sample Prep, RNA-seq, etc)
- Literature Watch
- Events/Conferences
- Jobs forums



Forum	Last Post	Threads	Posts
<b>Introductions</b> (1 Viewing) New here? Stop in and introduce yourself. Where you are, what you work on, etc.	<a href="#">I'm a total newbie and would...</a> by <a href="#">zhacker</a> Today 07:59 AM	558	1,217
<b>General</b> (2 Viewing) Any topic/question that does not fit into the subcategories below. If you're unsure of where to put something, ask in here!	<a href="#">UCSC data in SQTITE</a> by <a href="#">TDT</a> Today 01:41 PM	406	1,678
<b>Core Facilities</b> Dealing with customer samples, data, and the challenges that come with both?	<a href="#">qPCR's utility in nextgen...</a> by <a href="#">monad</a> 05-12-2011 10:51 AM	13	63
<b>Literature Watch</b> (2 Viewing) Discussion of any scientific study related to high content or next generation genomics. Whole genome association, metagenomics, digital gene expression, etc.	<a href="#">Field guide to...</a> by <a href="#">ECO</a> Today 07:46 PM	962	1,176
<b>Events / Conferences</b> (1 Viewing) Relevant events or scientific conferences that may be of interest to the community. (Mgmt reserves right to limit commercial content without approval.)	<a href="#">Workshop on Comparative...</a> by <a href="#">shandley</a> Yesterday 02:09 PM	140	393
<b>Bioinformatics</b> if TOPIC == 'computers' and SEARCHED_ALREADY == True : continue			
<b>Bioinformatics</b> (29 Viewing) Discussion of next-gen sequencing related bioinformatics: resources, algorithms, open source efforts, etc	<a href="#">Bowtie mapping quality in SAM</a> by <a href="#">Hunny</a> Today 09:52 PM	4,402	22,411
<b>Jobs Forums</b> Need one? Need to fill one? Post in the appropriate place below			
<b>Industry Jobs!</b> (1 Viewing) Job listings at companies that sponsor SEQanswers.com! Interested in posting a job? Contact admin@ for access.	<a href="#">Sr. IT Analyst, SAP Finance R...</a> by <a href="#">Pacific Biosciences</a> Today 04:39 PM	150	157
<b>Academic/Non-Profit Jobs</b> (2 Viewing) Post and look for academic, non-profit, or other non-industrial jobs here.	<a href="#">how to become a dual...</a> by <a href="#">banarise</a> Today 01:25 PM	214	252

# The power of community...

- Software troubleshooting at its best...
  - “SEQanswers boasts a number of authors of major second generation sequencing analysis packages, including Bowtie, Tophat, BFAST and Samtools, as regular contributors.” –Keith Robison, OmicsOmics.blogspot.com  
<http://omicsomics.blogspot.com/2011/05/forums-open-beats-closed-hands-down.html>
- Bowtie = Ben Langmead (u=1129) 194 posts
- Tophat = Cole Trapnell (u=1564) 183 posts
- BFAST = Nils Homer (u=1586) 895 posts (moderator!)
- Samtools = Heng Li (u=129) 449 posts
- Many, many others...



# 14 appearances in the literature...



SEQanswers

[Home Page](#)  
[Forums](#)

wiki navigation

[Main page](#)  
[Recent changes](#)  
[Random page](#)  
[Help](#)

Software

[Software hub](#)  
[Browse software](#)  
[Software list](#)

Toolbox

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Printable version](#)  
[Permanent link](#)  
[Browse properties](#)

Page **Discussion** [Read](#) [Edit](#) [View history](#)  [Go](#) [Search](#)

## Papers Referencing SEQanswers

(Redirected from [The Greatest Papers in the World](#))

This is a list of publications that have referenced SEQanswers.

This list also exists [on the forums](#) as well.

Year	Publication	Lead Author	Title	PubMed	Publication Link
2008	Nature Biotechnology	Shendure	Next-generation DNA sequencing	<a href="http://www.ncbi.nlm.nih.gov/pubmed/18846087">http://www.ncbi.nlm.nih.gov/pubmed/18846087</a>	<a href="http://www.nature.com/nbt/journal/v26/n10/abs/nbt1486.html">http://www.nature.com/nbt/journal/v26/n10/abs/nbt1486.html</a>
2009	Briefings in Bioinformatics	Horner	Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing	<a href="http://www.ncbi.nlm.nih.gov/pubmed/19864250">http://www.ncbi.nlm.nih.gov/pubmed/19864250</a>	<a href="http://bib.oxfordjournals.org/cgi/content/abstract/11/2/181">http://bib.oxfordjournals.org/cgi/content/abstract/11/2/181</a>
2009	Nature Biotechnology	Trapnell	How to map billions of short reads onto genomes.	<a href="http://www.ncbi.nlm.nih.gov/pubmed/19430453">http://www.ncbi.nlm.nih.gov/pubmed/19430453</a>	<a href="http://www.nature.com/nbt/journal/v27/n5/abs/nbt0509-455.html">http://www.nature.com/nbt/journal/v27/n5/abs/nbt0509-455.html</a>
2009	Nature Methods	McPherson	Next-generation gap	<a href="http://www.ncbi.nlm.nih.gov/pubmed/19844227">http://www.ncbi.nlm.nih.gov/pubmed/19844227</a>	<a href="http://www.nature.com/hmeth/journal/v6/n11s/full/hmeth.1268.html">http://www.nature.com/hmeth/journal/v6/n11s/full/hmeth.1268.html</a>
2009	Nature Methods	Flicek	Sense from sequence reads: methods for alignment and assembly	<a href="http://www.ncbi.nlm.nih.gov/pubmed/19844229">http://www.ncbi.nlm.nih.gov/pubmed/19844229</a>	<a href="http://www.nature.com/hmeth/journal/v6/n11s/full/hmeth.1376.html">http://www.nature.com/hmeth/journal/v6/n11s/full/hmeth.1376.html</a>

# SEQwiki

- Organizing definable information
  - Software tools
  - Sequencing service providers
- Community edited




**Immense gratitude owed to Dan Bolser,  
post doc @ University of Dundee, Scotland**

# Forums are terrible at some things

	Sticky: <a href="#">Software packages for next gen sequence analysis</a> (1 2 3)		12-25-2009 06:45 PM by <a href="#">ECO</a>	<a href="#">236</a>	162,503	
---	--	---	---	---------------------	---------	---

01-23-2008, 11:19 PM #1

[sci\\_guy](#)  
Member  
  
Join Date: Jan 2008  
Location: Sydney  
Posts: 76

 **Software packages for next gen sequence analysis**

28 Dec 2009: This thread has been closed. Please see our [wiki software portal](#) for information about each of these packages.

**A reasonably thorough table of next-gen-seq software available in the commercial and public domain**

**Integrated solutions**

- \* [CLCbio Genomics Workbench](#) - *de novo* and reference assembly of Sanger, Roche FLX, Illumina, Helicos, and SOLiD data. Commercial next-gen-seq software that extends the CLCbio Main Workbench software. Includes SNP detection, ChIP-seq, browser and other features. Commercial. Windows, Mac OS X and Linux.
- \* [Galaxy](#) - Galaxy = interactive and reproducible genomics. A job webportal.
- \* [Genomatix](#) - Integrated Solutions for Next Generation Sequencing data analysis.
- \* [JMP Genomics](#) - Next gen visualization and statistics tool from SAS. They are [working with NCGR](#) to refine this tool and produce others.
- \* [NextGENe](#) - *de novo* and reference assembly of Illumina, SOLiD and Roche FLX data. Uses a novel Condensation Assembly Tool approach where reads are joined via "anchors" into mini-contigs before assembly. Includes SNP detection, ChIP-seq, browser and other features. Commercial. Win or MacOS.
- \* [SeqMan Genome Analyser](#) - Software for Next Generation sequence assembly of Illumina, Roche FLX and Sanger data integrating with Lasergene Sequence Analysis software for additional analysis and visualization capabilities. Can use a hybrid templated/de novo approach. Commercial. Win or Mac OS X.
- \* [SHORE](#) - SHORE, for Short Read, is a mapping and analysis pipeline for short DNA sequences produced on a Illumina Genome Analyzer. A suite created by the 1001 Genomes project. Source for POSIX.
- \* [SlimSearch](#) - Fledgling commercial product.

**Align/Assemble to a reference**

- \* [BFAST](#) - Blat-like Fast Accurate Search Tool. Written by Nils Homer, Stanley F. Nelson and Barry Merriman at UCLA.
- \* [Bowtie](#) - Ultrafast, memory-efficient short read aligner. It aligns short DNA sequences (reads) to the human genome at a rate of 25 million reads per hour on a typical workstation with 2 gigabytes of memory. Uses a Burrows-Wheeler-Transformed (BWT) index. [Link to discussion thread here](#). Written by Ben Langmead and Cole Trapnell. Linux, Windows, and Mac OS X.
- \* [BWA](#) - Heng Lee's BWT Alignment program - a progression from Maq. BWA is a fast light-weighted tool that aligns short sequences to a sequence database, such as the human reference genome. By default, BWA finds an alignment within edit distance 2 to the query sequence. C++ source.
- \* [ELAND](#) - Efficient Large-Scale Alignment of Nucleotide Databases. Whole genome alignments to a reference genome. Written by Illumina author Anthony J. Cox for the Solexa 1G machine.
- \* [Exonerate](#) - Various forms of pairwise alignment (including Smith-Waterman-Gotoh) of DNA/protein against a reference. Authors are Guy St C Slater and Ewan Birney from EMBL. C for POSIX.

# SEQwiki Software Hub

<http://seqanswers.com/wiki/Software>



SEQanswers

[Home Page](#)  
[Forums](#)

wiki navigation

[Main page](#)  
[Recent changes](#)  
[Random page](#)  
[Help](#)

Software

[Software hub](#)  
[Browse software](#)  
[Software list](#)

Toolbox

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Printable version](#)  
[Permanent link](#)

Page [Discussion](#)

Read [Edit](#) [View history](#)

[Go](#) [Search](#)

## Software

*This page is the hub for all the software described in the wiki*

**FACT:** As of 21:03 Tuesday 24 May 2011 SEQwiki contains pages for 431 bioinformatics applications! [Click here for the full list.](#)

[Click here for examples of other 'queries'.](#)

There are:

- 431 bioinformatics applications,
- 336 references, and
- 507 URLs.

## Below you can...

[\[edit\]](#)

- [Browse software.](#)
- [Search for software.](#)
- [Add a software package...](#)

## Add a new bioinformatics application to the wiki

[\[edit\]](#)

After [searching](#), enter the name of the application that you want to add and click 'Add or edit'.

[Add or edit](#)

**Note:** to edit an existing package, you can enter it's name above and click 'Add or edit'.

# Yup, still terrible...

12-16-2008, 07:59 PM
#1

**ECO**  
--Site Admin--

Join Date: Oct 2007  
Location: SF Bay Area, CA, USA  
Posts: 824

**List of Next Gen Sequencing Service Providers**

Let's start one! I've got some in a list but figured we could all do it quicker, so I'm opening this one up to everyone!

29 Dec 08: Some great work by everyone, I've started to compile it below. Please let me know if you have comments, corrections, or additions to anything. Everything is listed in region-company alphabetical order to remain fair.

21 May 09: Many additions thanks to everyone who posted here, and to mtleaks for compiling them for me!

28 Oct 09: Many more additions, added states and alphabetized the US. Sincerest apologies for even \*insinuating\* that NZ was part of the Australian continent. Oceania FTW.

9 May 10: Updated. Data ported to Wiki!

See and edit the same data in the wiki:  
[Next Gen Sequencing Service Providers](#)

Region	Country	Company	Homepage	Instruments
Asia	China	Beijing Genomics Institute	<a href="http://www.genomics.org.cn/en/index.php">http://www.genomics.org.cn/en/index.php</a>	454 / Illumina (*only* 30 GAs) / SOLiD
	India	Genotypic	<a href="http://genotypic.co.in/newgen/">http://genotypic.co.in/newgen/</a>	454 / Helicos / Illumina / SOLiD
	India	Geschickten Solutions	<a href="http://www.geschickten.com/">http://www.geschickten.com/</a>	Informatics
	India	Ocimum Biosolutions	<a href="http://www.ocimumbio.com/web/">http://www.ocimumbio.com/web/</a>	454 / Illumina / Informatics / SOLiD
	India	Xcelris Labs	<a href="http://www.xcelrislabs.com/">http://www.xcelrislabs.com/</a>	454 / SOLiD
	Japan	Riken Genesis	<a href="http://www.rikengenesis.jp/en/index.html">http://www.rikengenesis.jp/en/index.html</a>	??
	Korea	Macrogen	<a href="http://www.macrogen.com/">http://www.macrogen.com/</a>	454 / Illumina / SOLiD
	Singapore	AIT Biotech	<a href="http://www.aitbiotech.com/HTPSequencing.htm">http://www.aitbiotech.com/HTPSequencing.htm</a>	SOLiD
	Taiwan	MissionBio	<a href="http://www.missionbio.com.tw/">http://www.missionbio.com.tw/</a>	SOLiD
Oceania	Australia	AGRF	<a href="http://www.agrf.org.au/NEW-SERVICE-i...eneration.html">http://www.agrf.org.au/NEW-SERVICE-i...eneration.html</a>	Illumina
	Australia	Geneworks	<a href="http://www.geneworks.com.au/Content.aspx?p=84">http://www.geneworks.com.au/Content.aspx?p=84</a>	Illumina
	Australia	Micromon (Department of Microbiology, Monash University, Victoria)	<a href="http://dna.med.monash.edu.au/">http://dna.med.monash.edu.au/</a>	Illumina / Informatics
	New Zealand (definitely not Australia)	Allan Wilson Centre Genome Service at Massey University	<a href="http://www.allanwilsoncentre.ac.nz/AWCGSintro.htm">http://www.allanwilsoncentre.ac.nz/AWCGSintro.htm</a>	Illumina
	New Zealand (definitely not Australia)	University of Otago	<a href="http://sequence.otago.ac.nz/index.html">http://sequence.otago.ac.nz/index.html</a>	454 / SOLiD

# SEQwiki NGS Service Providers

[http://seqanswers.com/wiki/Next Gen Sequencing Service Provider](http://seqanswers.com/wiki/Next_Gen_Sequencing_Service_Provider)



SEQanswers

[Home Page](#)  
[Forums](#)

wiki navigation

[Main page](#)  
[Recent changes](#)  
[Random page](#)  
[Help](#)

Software

[Software hub](#)  
[Browse software](#)  
[Software list](#)

Toolbox

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Printable version](#)  
[Permanent link](#)  
[Browse properties](#)

Page [Discussion](#)

[Read](#)

[Edit](#)

[View history](#)

[Go](#)

[Search](#)

## Next Gen Sequencing Service Provider

This is an attempt to compile all known next generation sequencing service providers around the world. It originated in [this thread](#) in the [Service Providers Forum](#).

A Google Map version of all next gen machines can be found [here](#).

Use 'facets' to narrow results.

[Add \(or edit\) a service provider](#)

[TILES](#) • [TABLE](#)

sorted by: [Company](#); then by... ☒ grouped as sorted

### AGOWA - LGC

Region Europe  
Country UK  
Homepage <http://www.lgc.co.uk/pdf/Next%20gen%20sequencing%20flyer%20web.pdf>  
Service 454

### AGRF

Region Oceania  
Country Australia  
Homepage <http://www.agrf.org.au/NEW-SERVICE-illumina-GA-II-Next-Generation.html>  
Service Illumina

### AIT Biotech

Region Asia  
Country Singapore  
Homepage <http://www.aitbiotech.com/HTPSequencing.htm>  
Service SOLiD

### Agencourt

#### Service

3 3730xl  
59 454  
1 454 and  
Bioinformatics  
1 AB 3730XL Sanger  
Sequencing

#### Region

13 Asia  
1 Brussels  
44 Europe  
48 North America  
6 Oceania  
1 South America

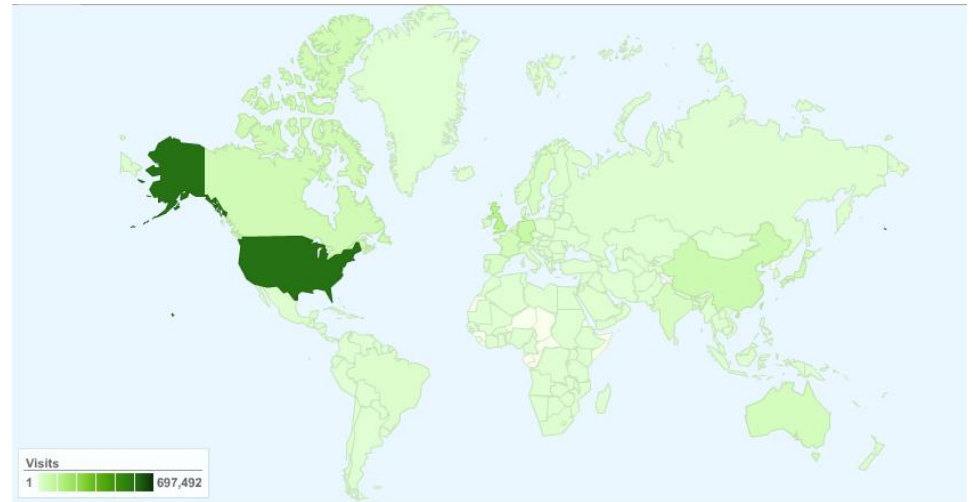
#### Country

4 Australia  
1 Austria  
3 Belgium  
1 Brazil  
3 Canada  
1 China



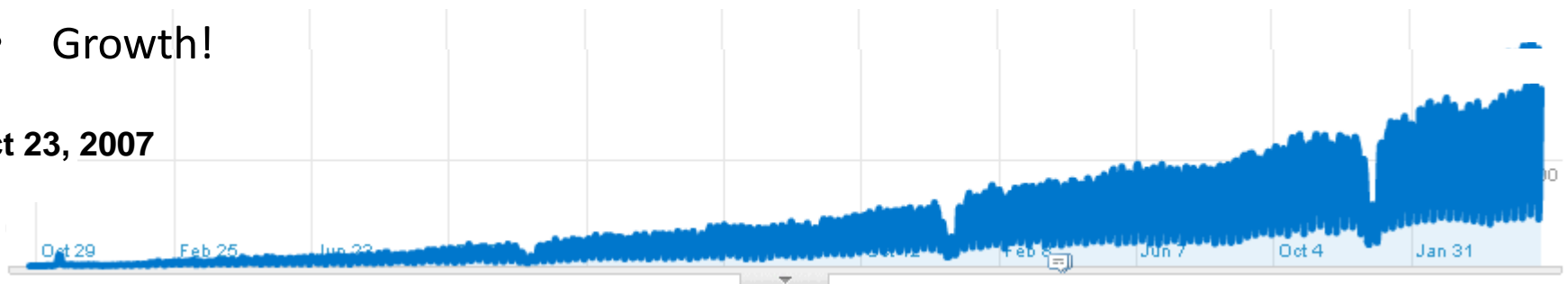
# Traffic

- 1000s of unique worldwide visitors per day
- Hundreds of thousands of pageviews per month
- As of 5/24:
  - Threads: 9,425
  - Posts: 39,963
  - Members: 17,258



- Growth!

Oct 23, 2007



# Library Prep Basics

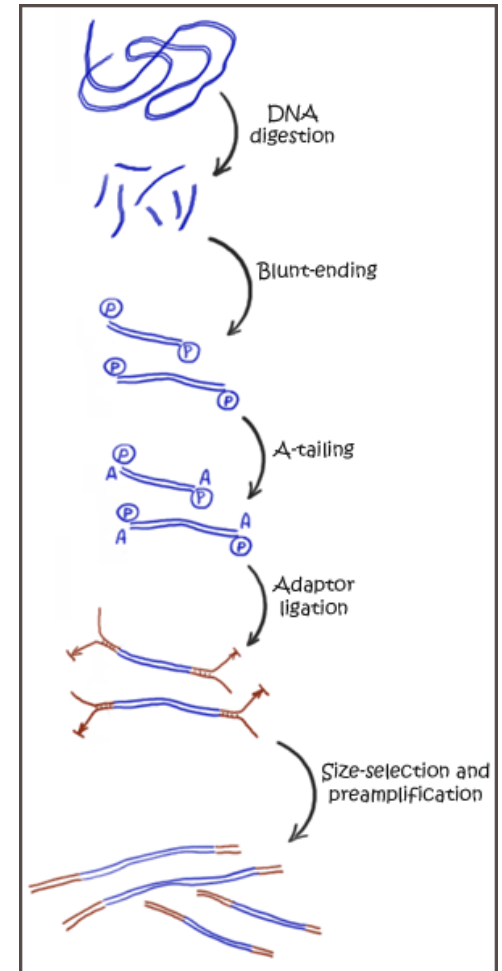


# Library Prep Basics

- Library:
  - Representative pool of DNA fragments from sample of interest
- Ideal library:
  - Tight size distribution of fragments
  - Bounded by a known anchor for priming
  - Bias-free
    - Amplification
    - Shearing
    - Molecular biology (heat, denaturants, etc)

# Generic Fragment Library Workflow

- Fragment DNA
  - Various physical or enzymatic methods
- Repair Ends
  - Fix staggered ends and add 5' phosphates
- Ligate Adapters
  - Add small universal oligonucleotides
- Purify/Size Select
  - Remove non- or partially-ligated products
- Amplify
  - Clustering (Illumina)
  - Emulsion (454, SOLiD, Ion Torrent)



Illumina example from:  
<http://seq.molbiol.ru/>

# Library considerations for Bioinformaticians

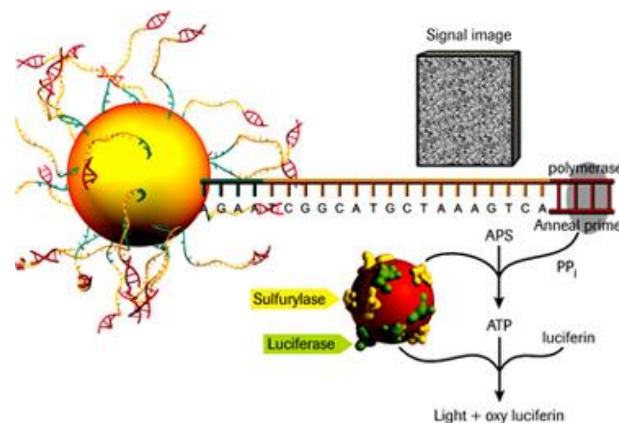
Problem	Symptom
GC bias	Uneven coverage, poor variant calls
Short readlengths	Mapping (repeats, structural variants, etc)
PCR duplicates	Artefactual variant calls
Chimeric inserts	Biologically irrelevant mappings
Low yield	Signal to noise, poor basecall quality
Wet-lab to biofx handoff	Incorrect data (Barcode assignment)
Contamination / Wrong Reference	Unaligned data

# Sequencing Technologies

What's unique?

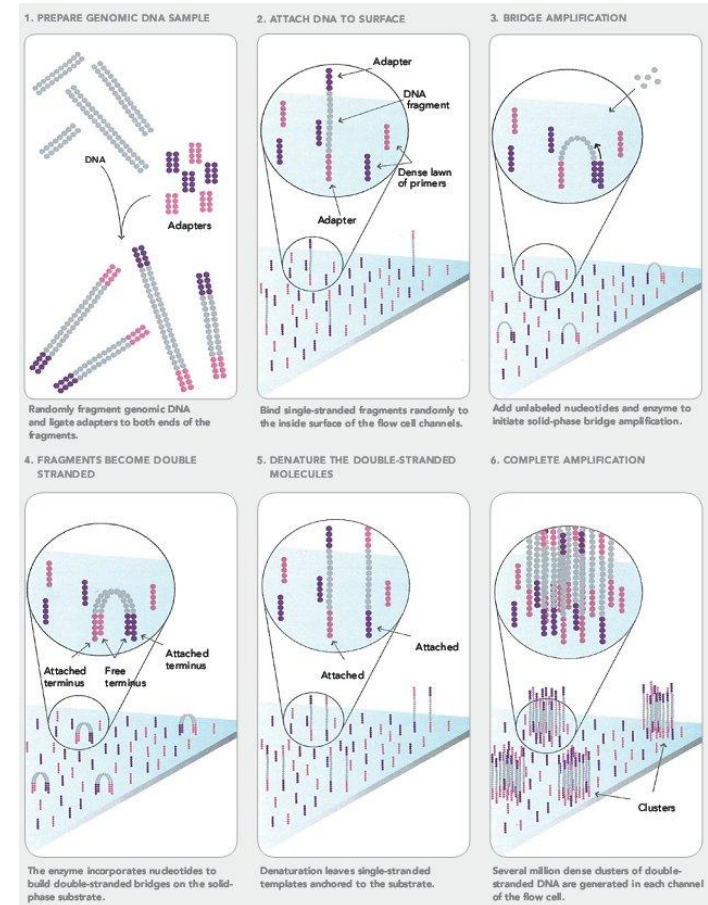
# 454: Pyrosequencing

- Pyrosequencing:
  - Release of pyrophosphate
  - Detection by enzymatic cascade which generates light
- Template beads held in PicoTiterPlate
- Chemistry creates homopolymer predominant error model
- Read lengths of 400 – 700bp



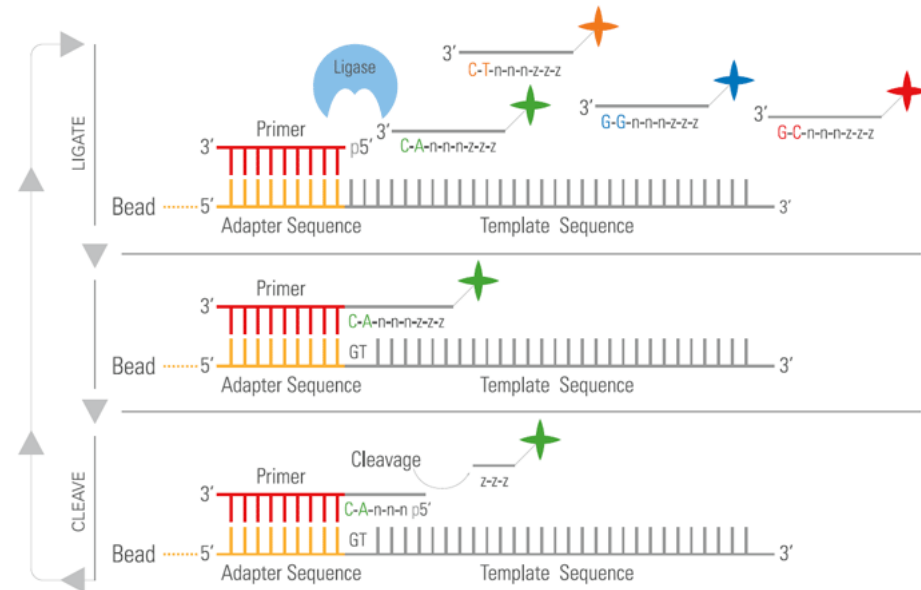
# Illumina: Seq by Synthesis

- Sequencing by synthesis
  - Incorporate single labeled base
  - Image
  - Deprotect
  - Repeat
- Solid-phase immobilized clusters ('colonies') of templates on slide
- Readlengths of up to 150bp



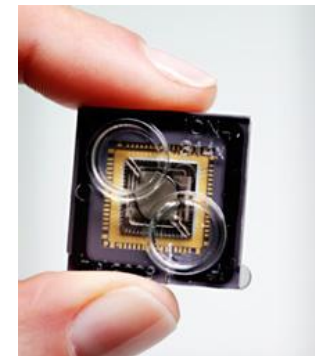
# SOLiD: Seq by Ligation

- Sequential ligation of labeled probes
- Beads immobilized on glass slide
- Unique read property “colorspace”
- Interesting algorithmic challenges and market impacts



# Ion Torrent PGM

- Founded by Jonathan Rothberg (454!)
- Direct electrical detection of protons released upon incorporation
- Beads immobilized in disposable parallelized microchip pH meter
- Very inexpensive
- Fast run times
- Readlengths <100bp





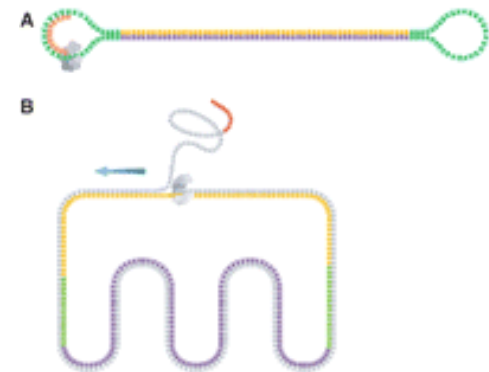
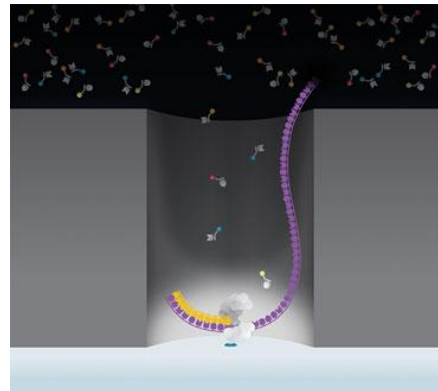
# Helicos: true SMS

- First commercialized single molecule sequencer
- TIRF microscopy of sequencing by synthesis
- Readlengths of 25bp
- First de-commercialized single molecule sequencer
  - Just kidding. Mostly.



# PacBio: single-molecule real-time

- Video of polymerases incorporating labeled nucleotides
- Optical confinement achieved with zero-mode waveguides
- Single molecule raw error rates higher than ensemble calls
- Unique template configuration
  - “SMRTbell™ template”
  - Enables circular consensus
- Readlengths ...



# Future of the site

- Q&A format (?)
  - Less personal but more information rich
  - <http://biostar.stackexchange.com> (1300 users and growing)
- Wiki expansion
- Recruit bloggers
- Conference
- Community science
- Find help
  - Designer
  - Wiki guru
  - PHP/Django/CMS/backend
  - Legal – official policy for information sharing
  - Opinionated bloggers

# Acknowledgements

- Every member
- Volunteer moderators
  - Nils Homer and Lex Nederbragt
- Wiki Contributors, Dan Bolser
- Sponsors
- PacBio for being supportive
- Family...late nights and weekends spent on interweb