Next-Generation Sequencing – NGS core facility, principles of NGS platforms, applications, and experimental considerations

基因體定序 - NGS核心設施簡介、應用、及實驗設計之考量

Mei-Yeh Lu 呂美曄
Associate Research Specialist,
High Throughput Genomics Core
Biodiversity Research Center, Academia Sinica

LSL, 2017/4/25

Outlines

Part I: DNA Sequencing & NGS Core

- Evolution of sequencing technologies
- NGS platforms and comparisons

Part II: Principles of common NGS applications

- DNA
- RNA

Part III: Prep works

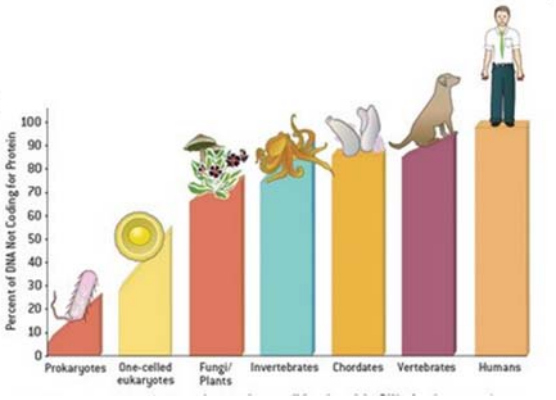
- Sample requirement and considerations
- Submission to NGS Core

Part IV: NGS Data & Resources

- Data types and QC
- Public resources

HT Genomics Core in Academia Sinica

- 2008:Established for biofuel pl
- 2013: Promoted as NGS servic
- Internal & collaborative projec
  - Pathogenic bacteria
  - Pathogenic/medicinal fungi
  - worms
  - Insects
  - Evo-devo: avian species
  - C3/C4 plants
  - Marine animals



Need Super Sequencers!!

http://www.ashg.org/education/everyone_1.shtml

NGS Service at BRC core

- Current NGS lineup:
  - 3 Illumina (HiSeq2500 *2, MiSeq)
  - Roche 454 GS+
  - PacBio Sequel
- **SOPs:** established for various NGS applications for both sequencing platforms
- Provides consultation on:
  - Project's need
  - suitable NGS experiment design
  - Sample preparation
  - Cost analysis

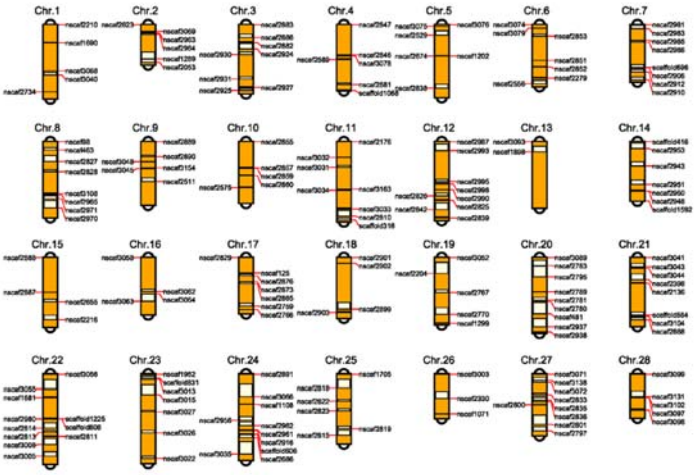
http://ngs.biodiv.tw/NGSCore/

I. DNA Sequencing

- Evolution of sequencing technologies
- NGS platforms and comparisons

Genome

- a full haploid set of chromosomes with all its gene; the total genetic constitution of a cell or organism.



- **Nuclear chromosome**
- **Plasmid (bacterial)**
- **Viral**

Plastids:

- **Mitochondrial**
- **Chloroplast**

In general, genome size increases with organism's complexity.

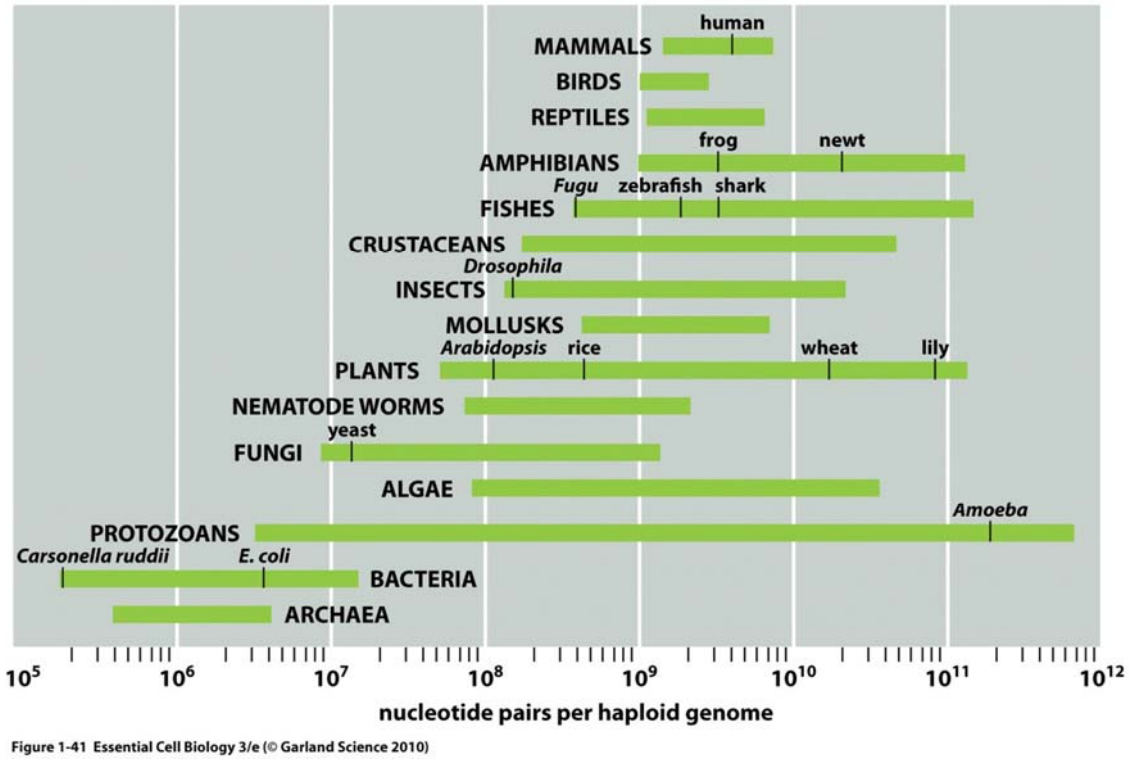
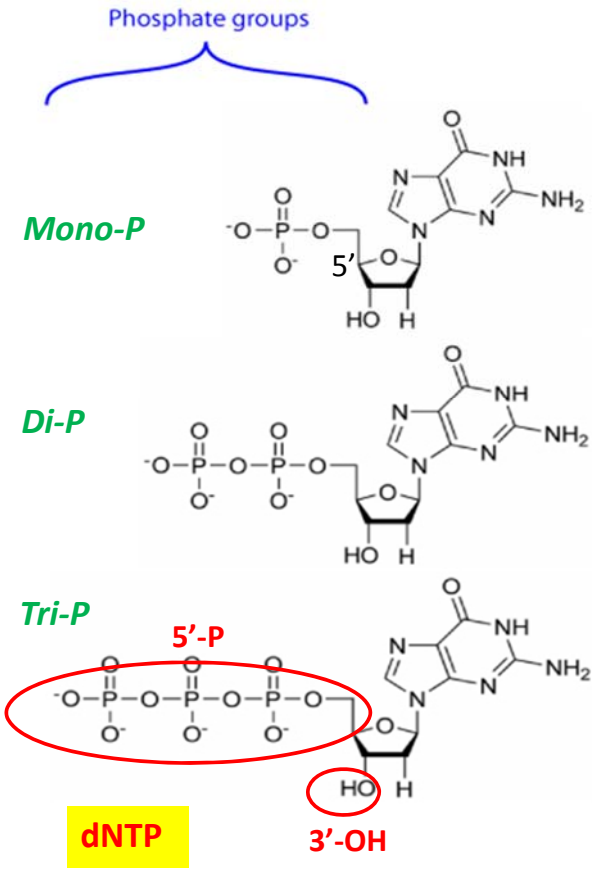
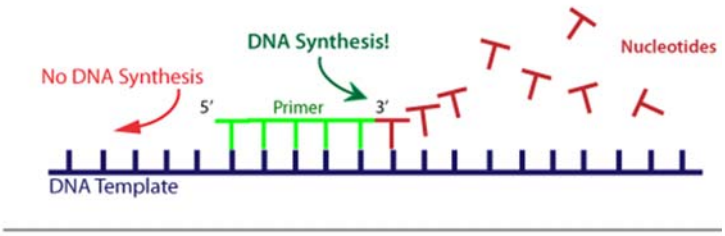
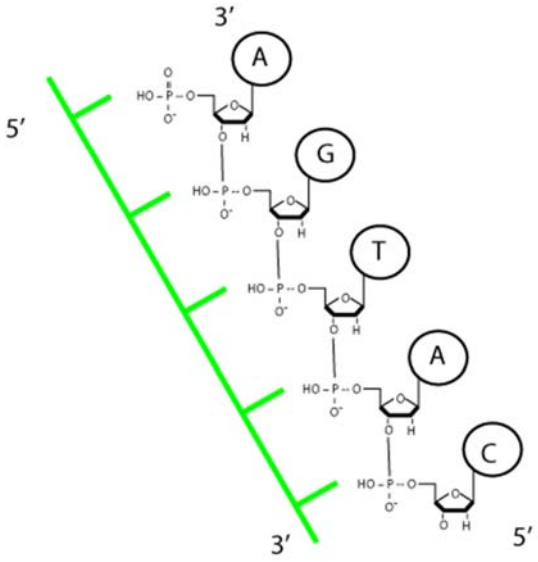


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

DNA synthesis - polymerization

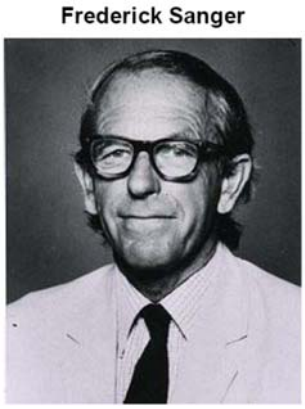


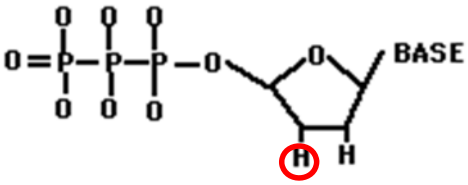




http://www.visionlearning.com/en/library/Biology/2/DNA-III/180

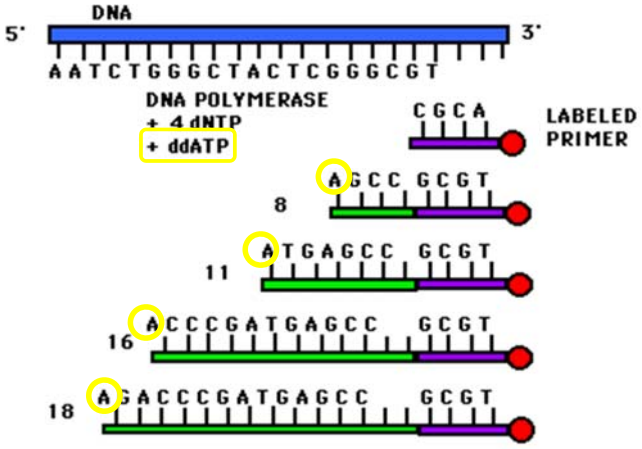
Sanger Seq.: chain termination w/ dideoxy nucleotide

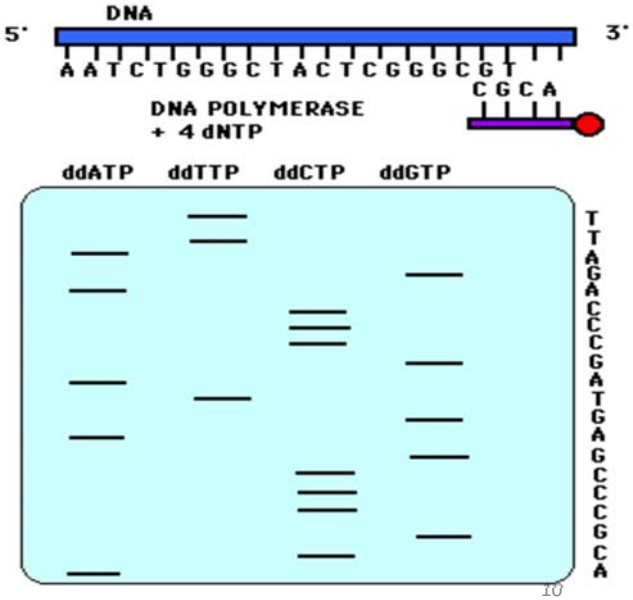




dideoxy nucleotide

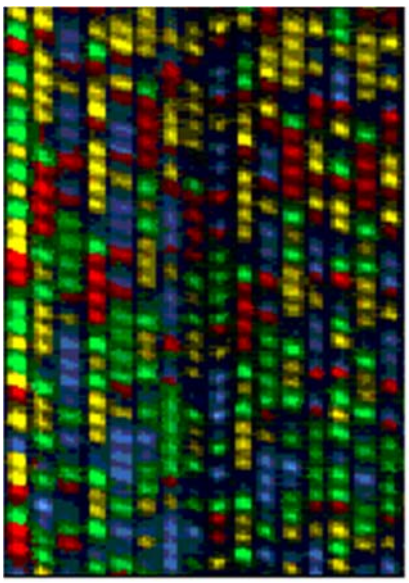
1977

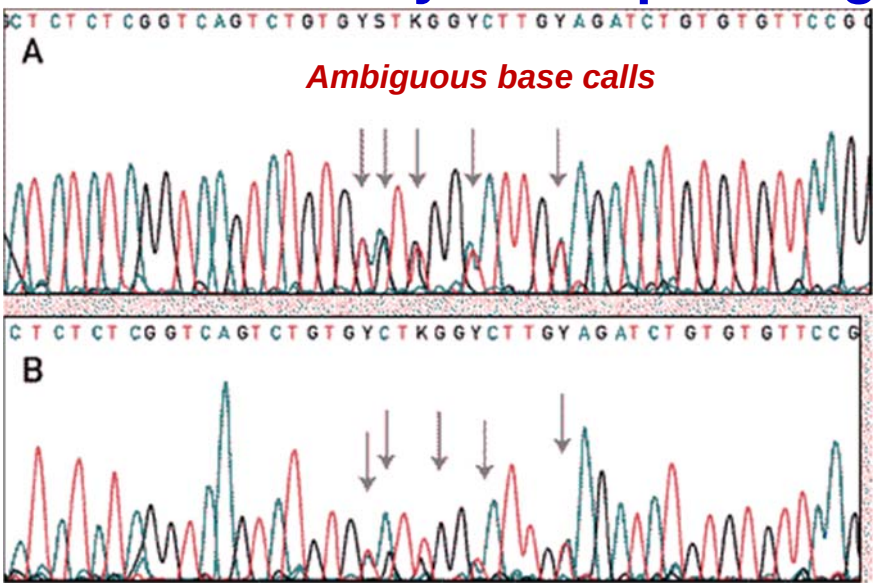


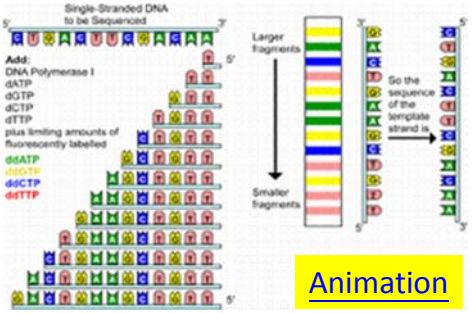


http://www.bio.davidson.edu/Bio111/seq.html

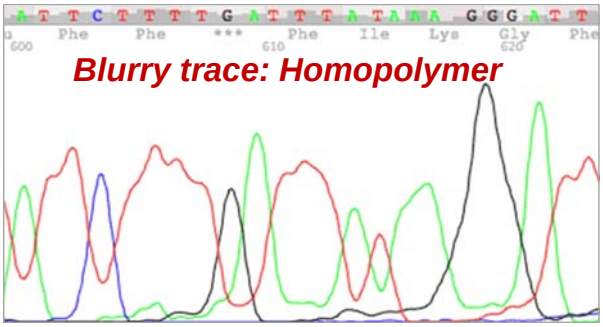
Fluorescent Dye: Terminator Cycle Sequencing







Animation

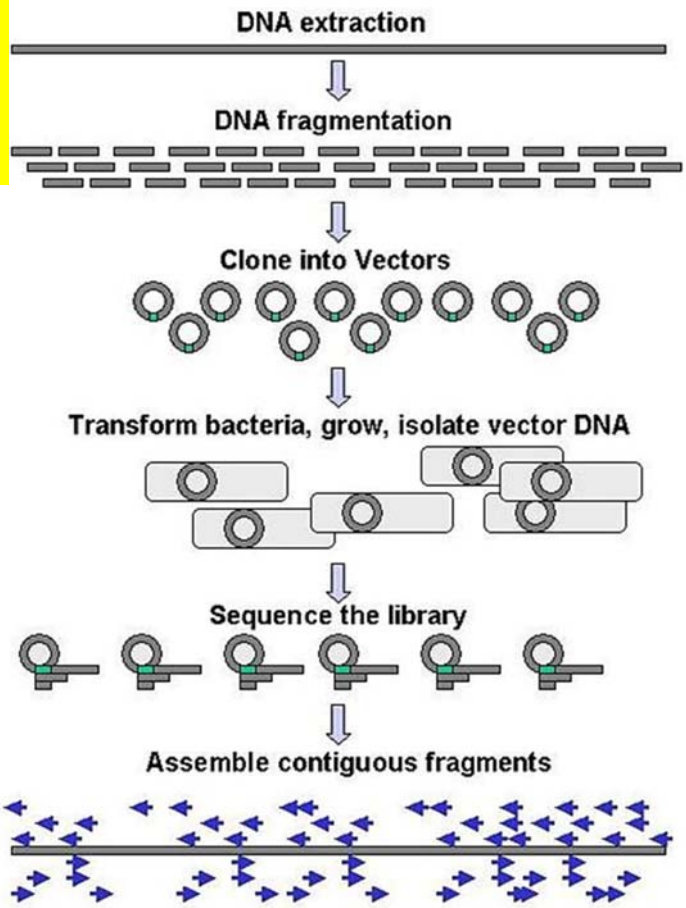


Genome Sequencing: Hierarchical cloning

BAC: 100-200kb

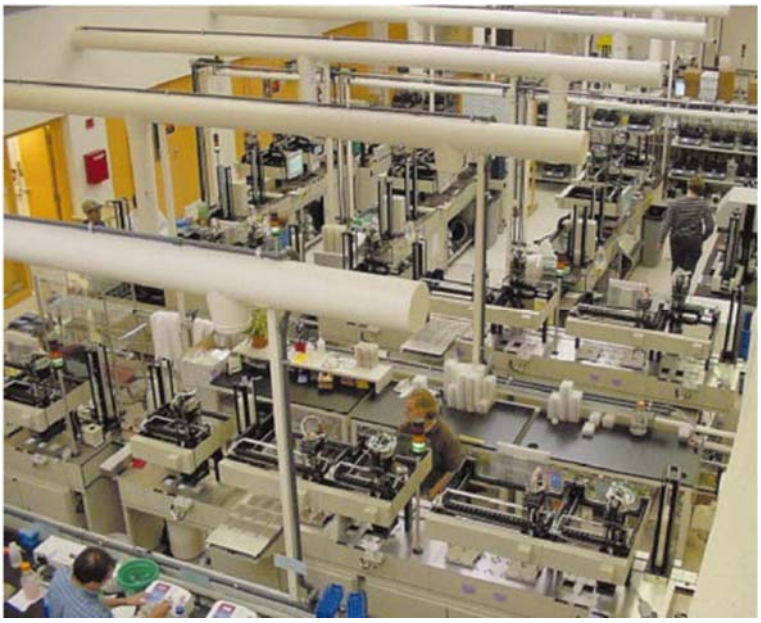
Fosmid: 30-40 kb

Plasmid: 1-10kb



Large scale Cappilary Sequencing

Library factory - Whitehead Institute



Sequencing factory - Sanger Institute

Nature 409, 860-921(15 February 2001)

Next generation Sequencing

- Improvements in:
  - Enzymes
  - Chemistry
  - Image analysis
  - increased sequencing capabilities.

- Potentials to dramatically accelerate biological and biomedical research
  - *by enabling the comprehensive analysis of genomes, transcriptomes and interactomes,*
  - *by tending to become inexpensive, routine and widespread, rather than requiring very costly production-scale efforts.*

From: Introduction to NGS.

NGS – massive parallel sequencing

Current Popular platforms:

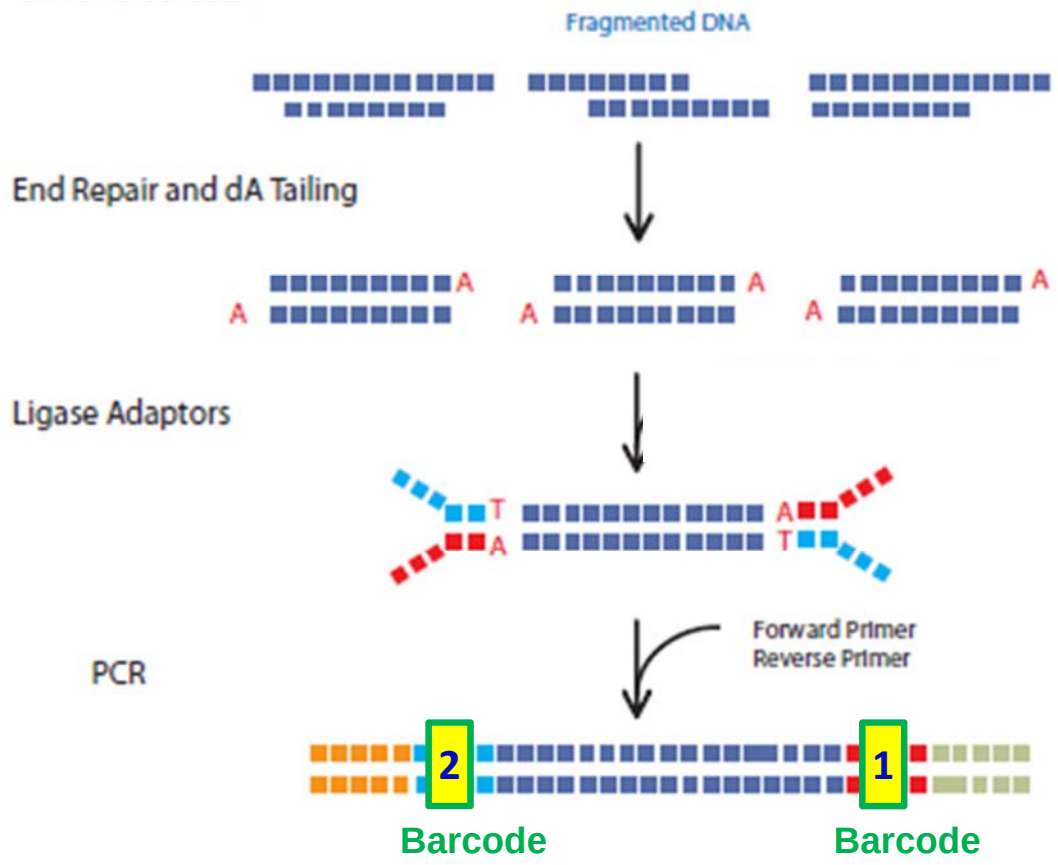
- **2nd-Gen: clonal amplification**
  - Roche 454: GS FLX, , 454 Jr., 454 XL+, 454 Jr.
  - Illumina: GA, Miseq, HiSeq, NovaSeq
  - Life Technologies: SOLiD, Ion Torrent, Ion Proton

- **3rd-Gen: single molecule sequencing**
  - Pacific Biosciences: PacBio RS II, Sequel
  - Oxford Nanopore Technologies

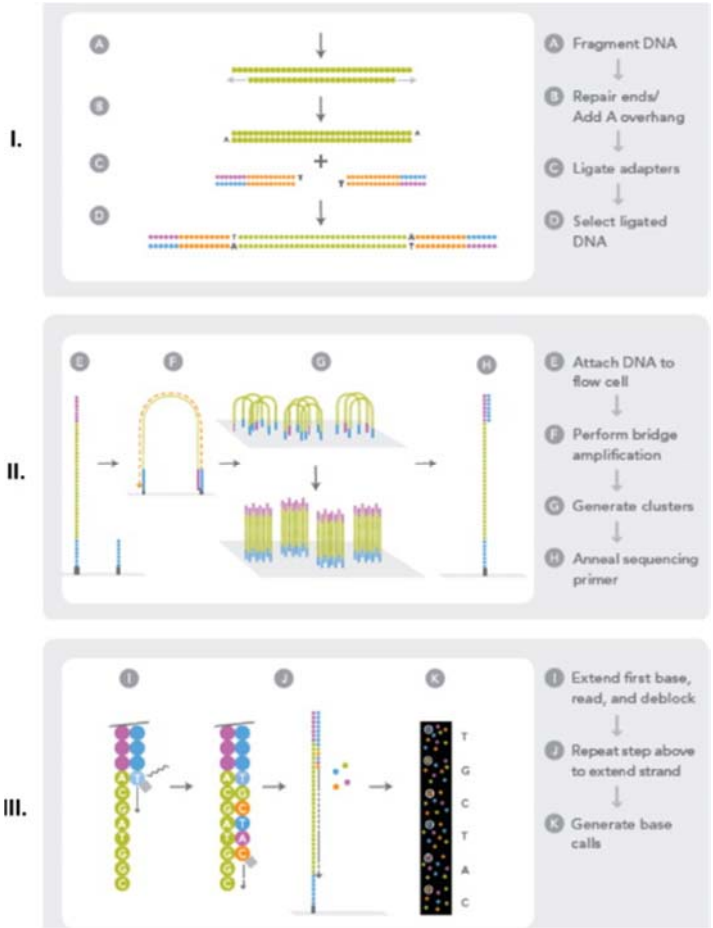
Advances in Sequencing Technologies

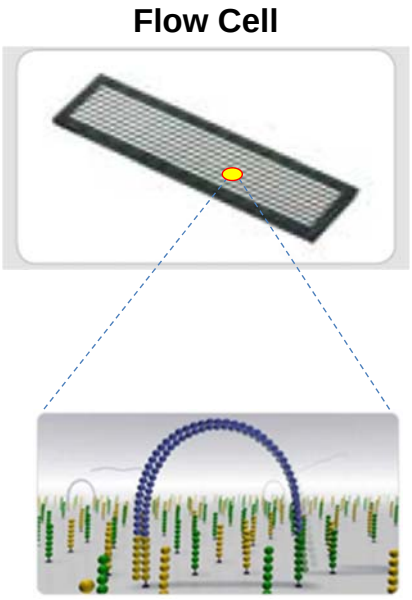
https://www.slideshare.net/AustralianBioinformatics/introduction-to-nextgeneration

NGS Library Preparation Workflow

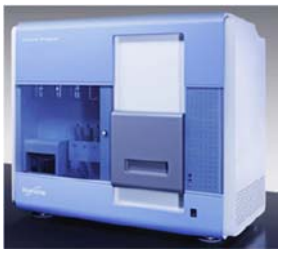


Illumina/Solexa: Cyclic Reversible Terminator





Illumina – Flow cell imaging

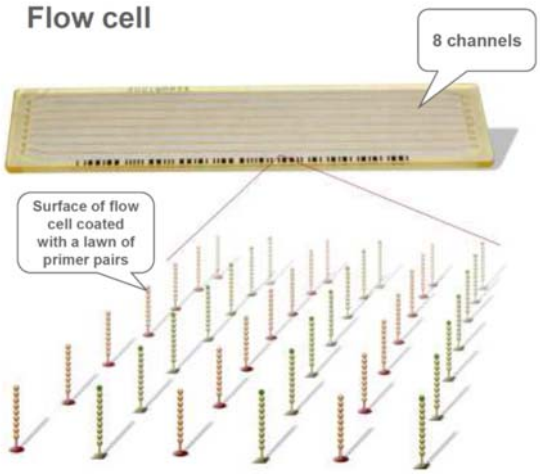


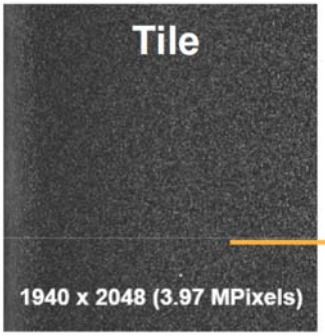
GA IIx

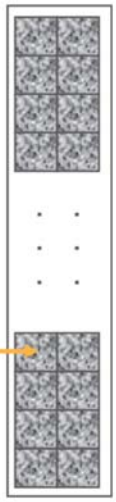
HiSeq 2500
(HT*8 / Rapid*2)

NextSeq 500

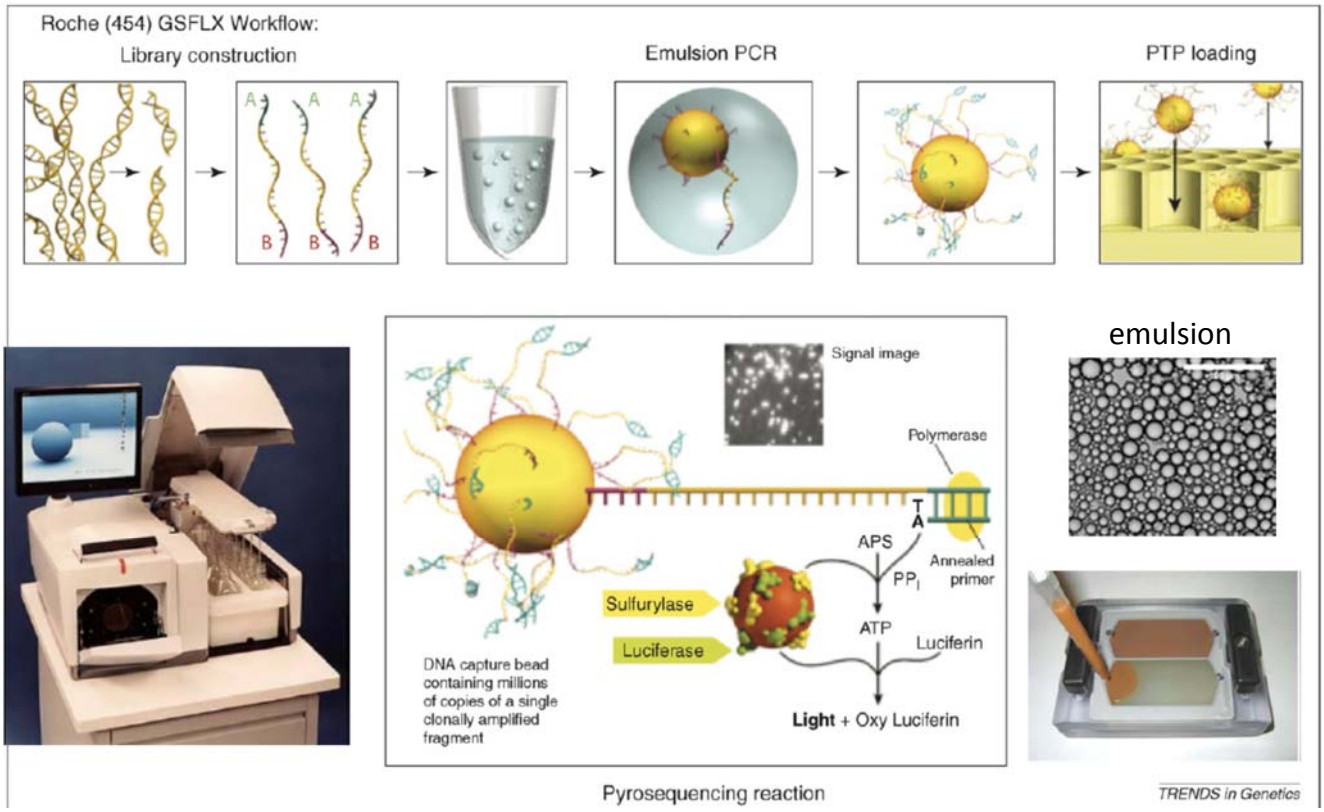
MiSeq v2



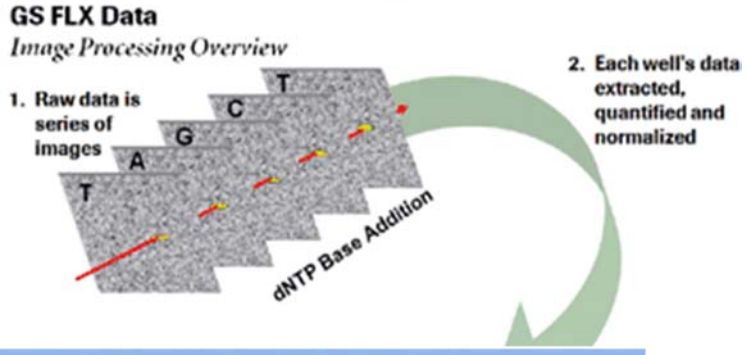


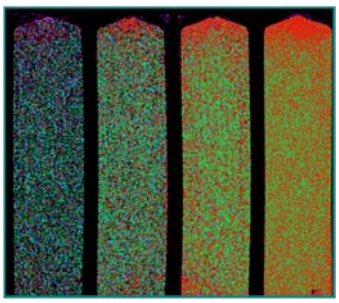


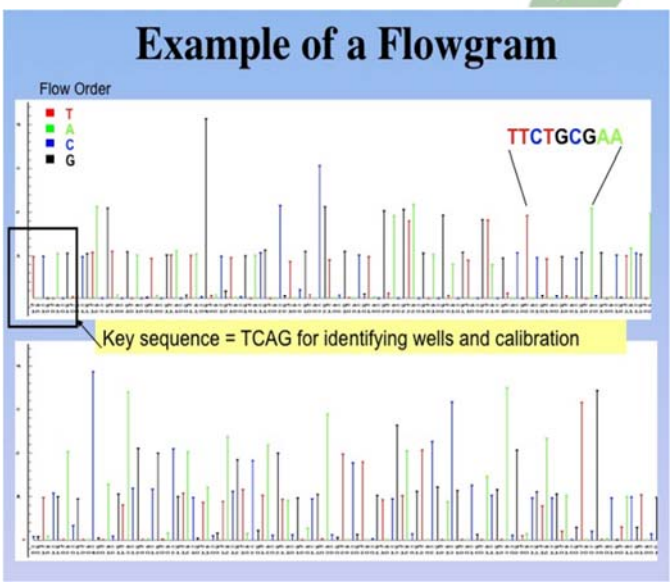
454: emPCR & pyrosequencing

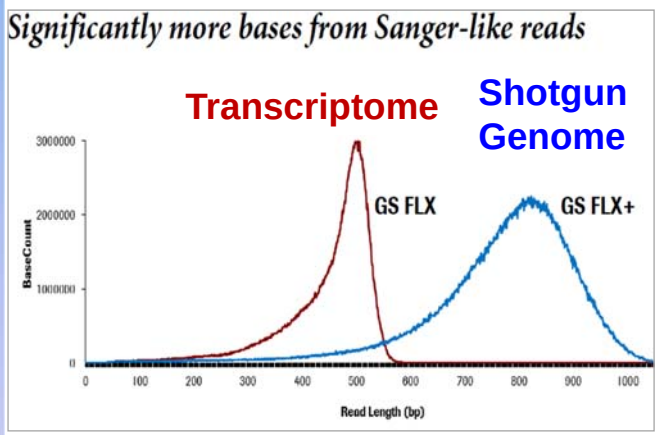


454 flowgram and read length profile

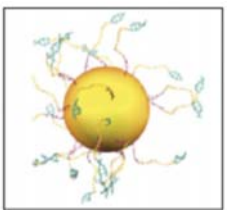






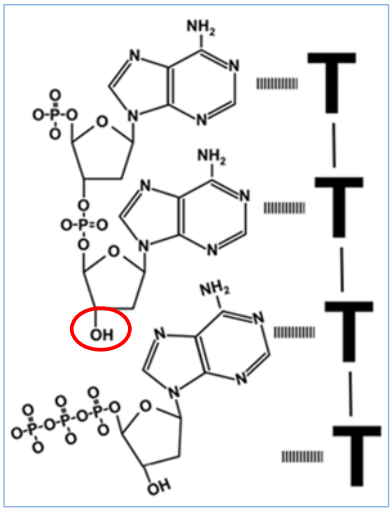


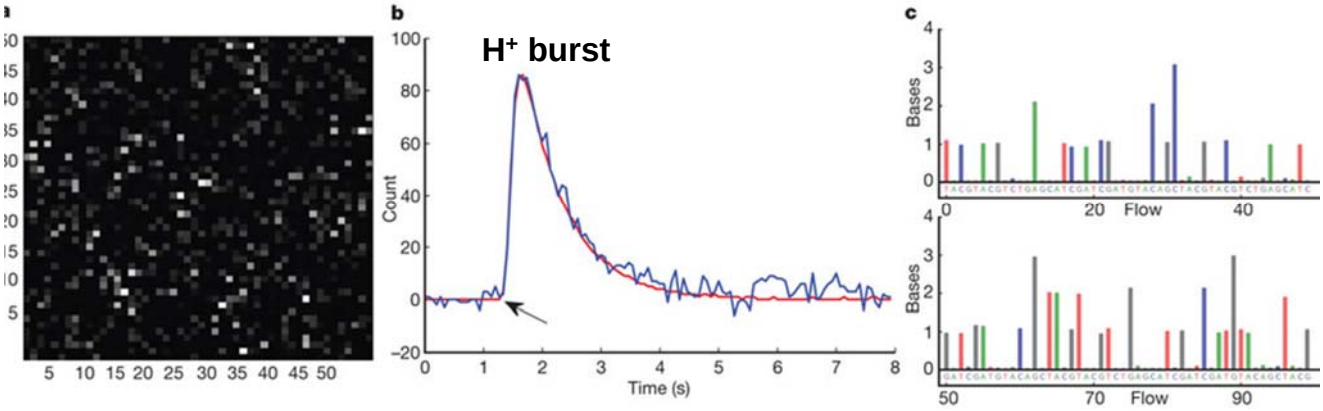
Ion Torrent/Proton:
Sensing bulk release of H^+



emPCR

Semi-conductor

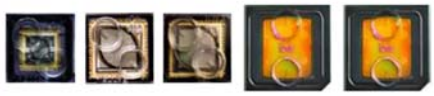


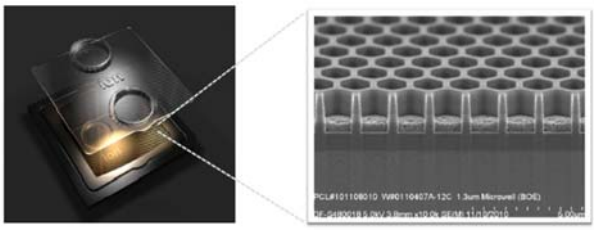


JM Rothberg *et al. Nature* **475**, 348-352 (2011) doi:10.1038/nature10242

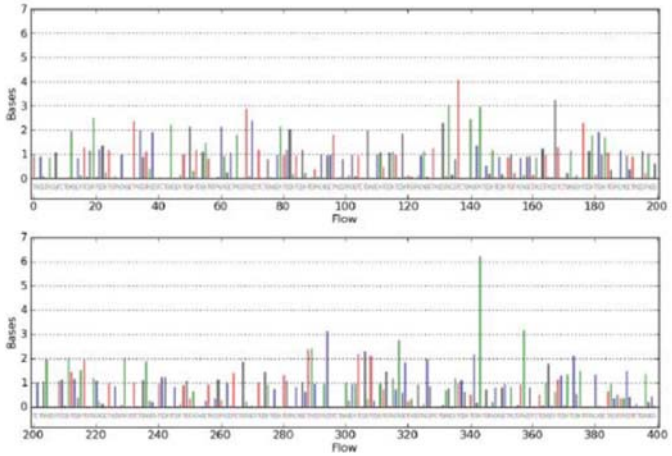
Life Technologies – proton sensing

PGM™ for genes.
Proton™ for genomes.
Sequencing for all.



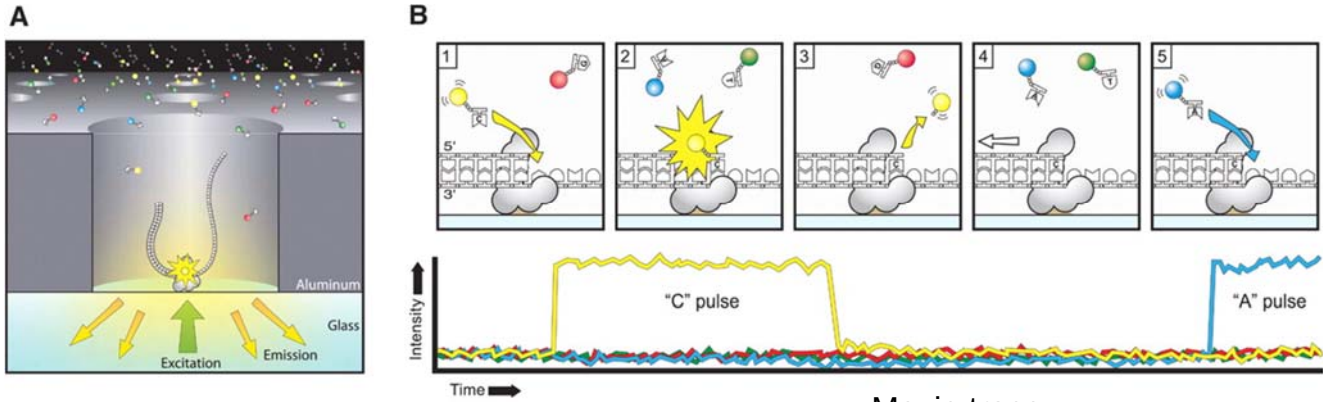


Chip: 台積電

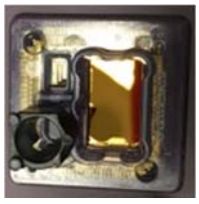


PacBio: 3rd-Gen SMRT Sequencing

- Single Molecular Real Time (SMRT) real-time technology
- ZMW (zero-mode waveguides), a 100-nm hole with DNA/Polymerase complex immobilized at the bottom; recording fluorescence released from P-dNTP upon incorporation



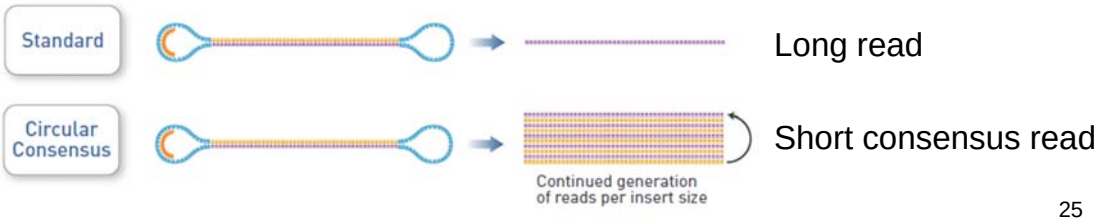
Movie trace



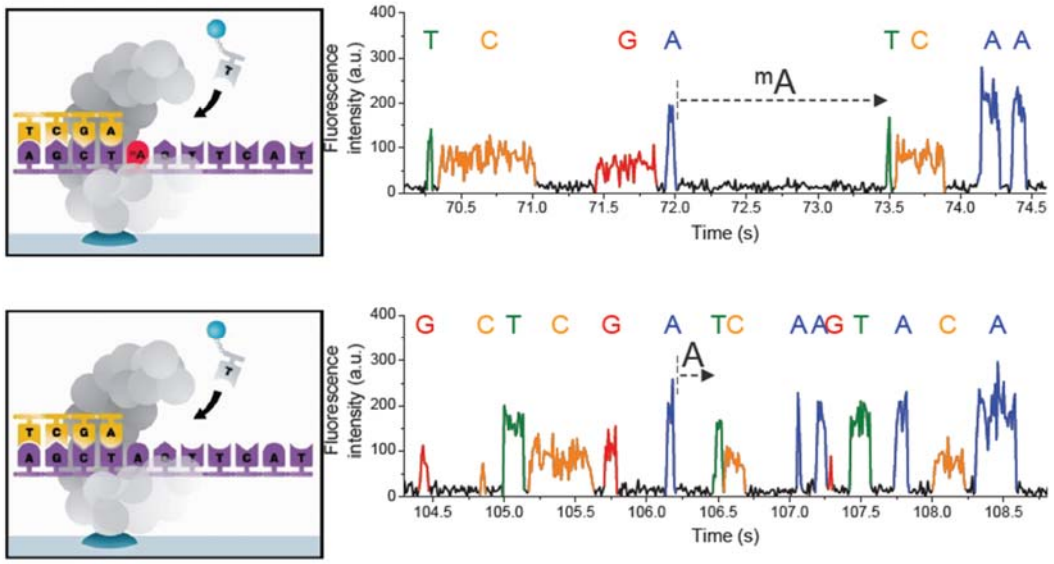
Signal Processing and Base Calling

Converting pulses of light into DNA bases and kinetic measures

Read length: avg. 7-10 kb, up to 20kb
Throughput: 3-5Gb
Accuracy: 85% (1X) to 99.99% (30X)

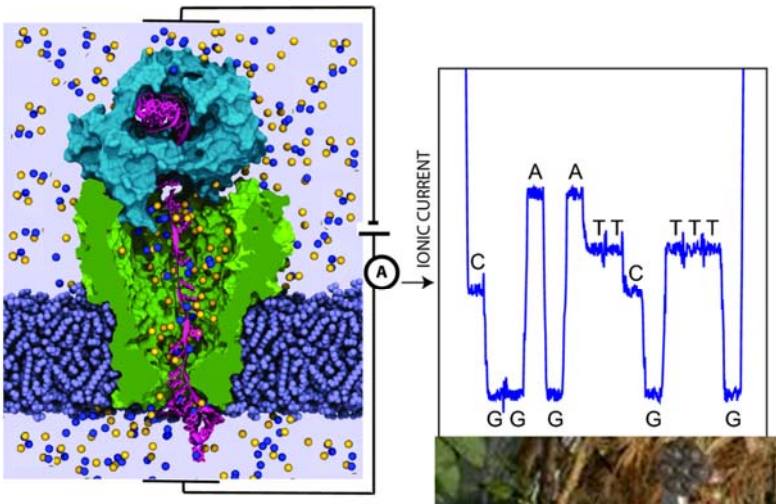


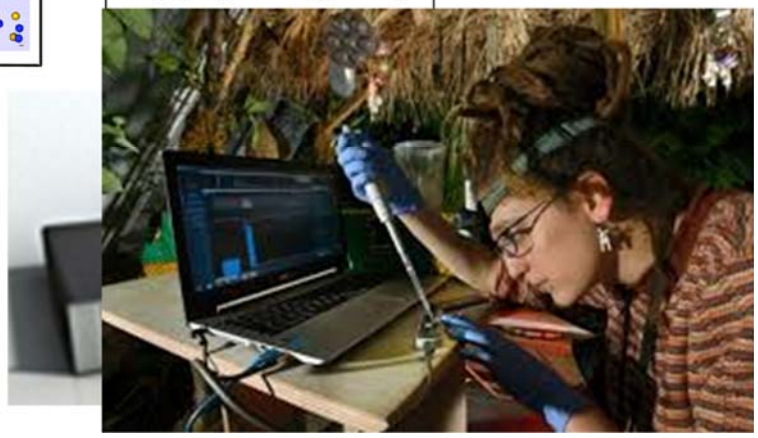
Key Feature: Kinetic Information



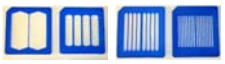
- Differentiation between modified and non-modified bases
  - Epigenetics, DNA damage, New, novel modifications
- Direct observation (*e.g.* no bisulfite)

NanoPore Sequencing Technology

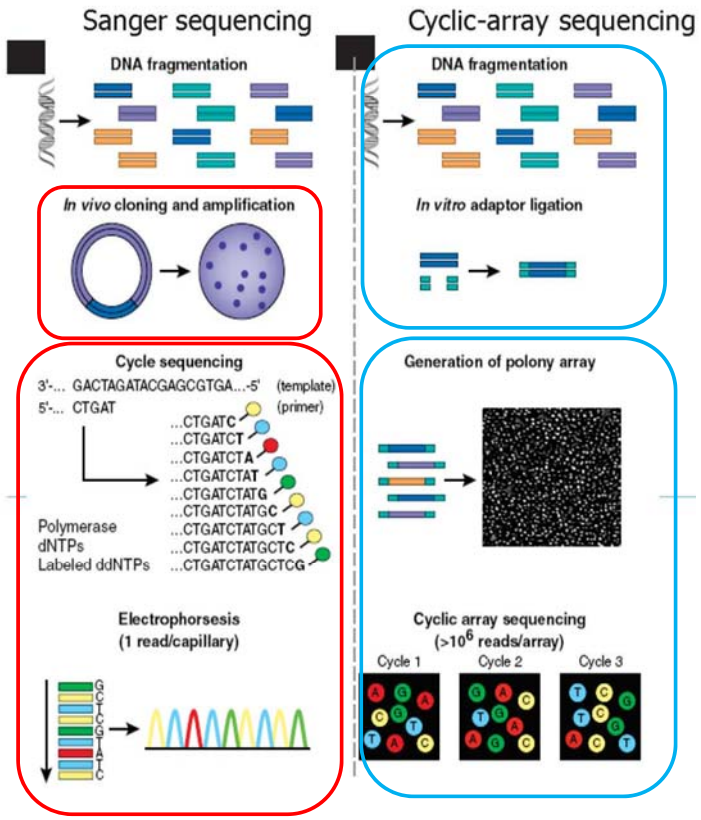




NGS Platforms & Features

	Roche 454 GS +	Illumina HiSeq 2500 (*2)	Illumina MiSeq	PacBio Sequel
Chemistry	Pyrosequencing	Cyclic reversible terminator		Real-time single molecule polymerization
Chip format				
Output/run	500-850 Mb	HT mode: 1 Tb Rapid mode: 150 Gb	up to 15 Gb	Current: 3-5 Gb Future: 7-10 Gb
Read length	400-1000 nt	PE 50-250 nt	PE 50-300 nt	1-15 kb
# Fragments /lane	1 M reads	150-180 M (Rapid) 200-250M (HT)	12-15 M (v2) 20-25M (v3)	350-500 K / SMRT cell
Data quality	> 99%; homopolymeric errors; tolerate high GC%	> 99.9%; Tolerate homopolymer; sensitive to high GC	> 99.9%; Tolerate homopolymer; sensitive to high GC	Raw ~ 85%; CCS ~ 99.9%; homopolymeric errors; tolerate high GC%
Application	De novo genome & transcriptome; Long amplicons	Highest throughput De novo assembly; Re-sequencing	De novo assembly; Re-sequencing; amplicon	Genome assembly; structural variation; phasing; Iso-Seq

Next-generation DNA sequencing



Advantages:

- adaptor-mediated library construction
- Clonal amplification to enhance signal intensity

- No bacterial cloning, colony picking, chr. Walking

- Array-based sequencing

- Massive parallel sequencing

- Much cheaper per *output unit*

http://ueb.ir.vhebron.net/NGS

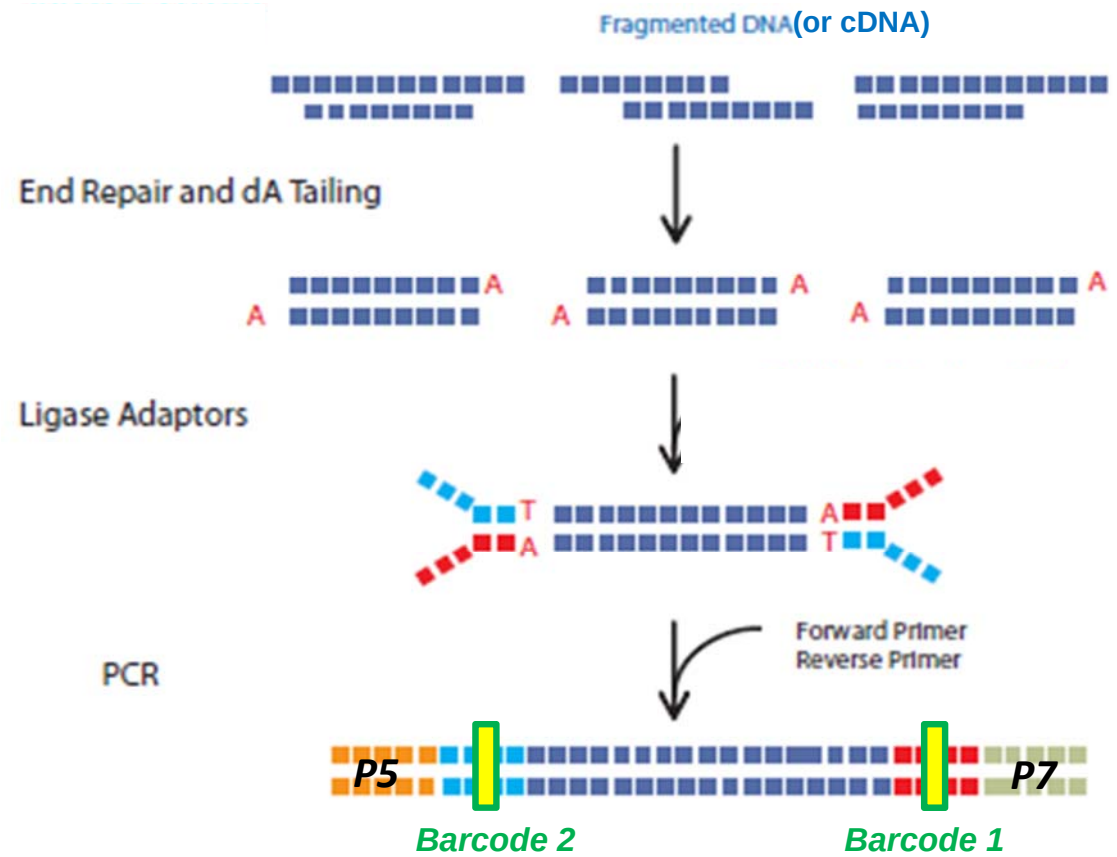
II. Principles of common NGS applications

- **DNA applications**
- **RNA applications**

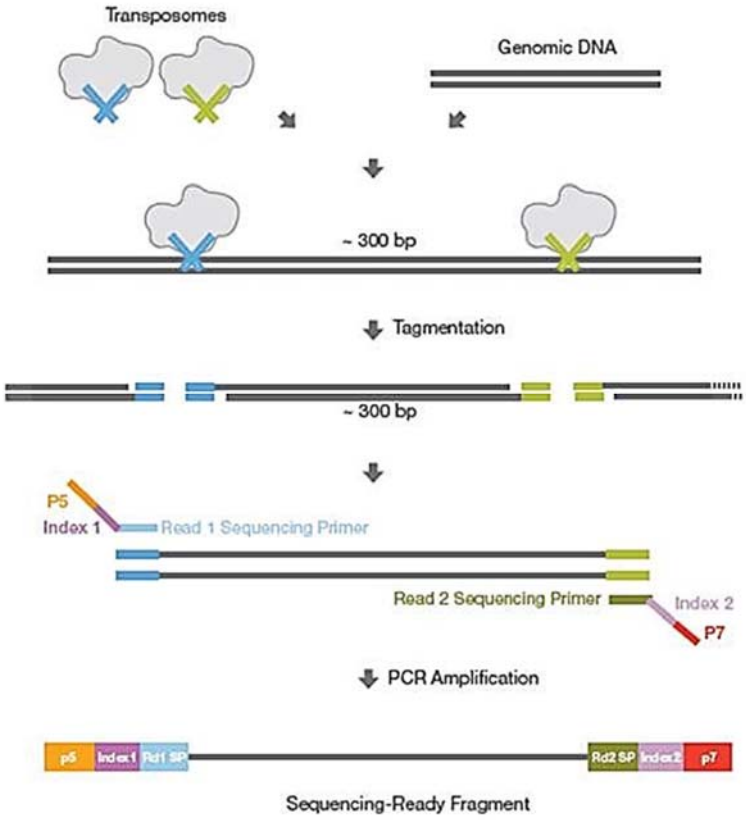
Genome Sequencing

- *De novo*
- *Re-sequencing*

1. Shotgun gDNA Library Preparation



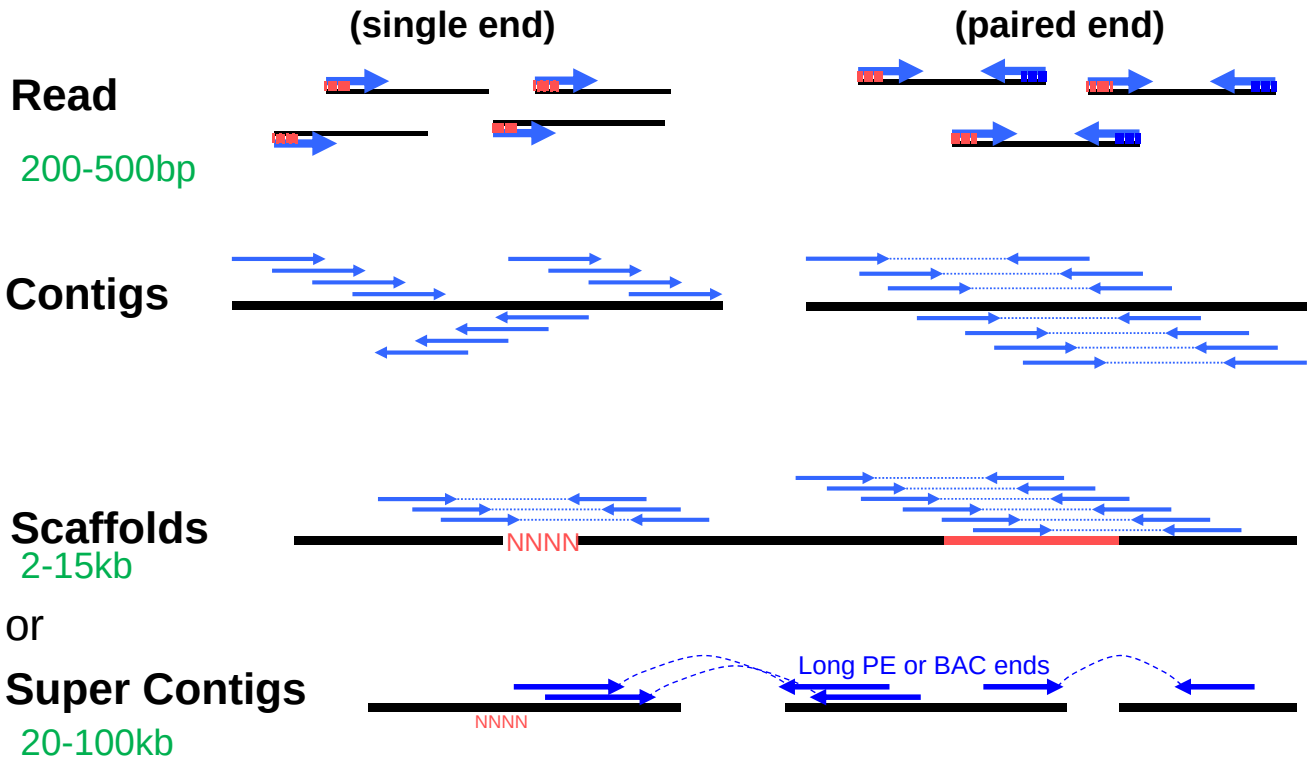
2. Tn tagging - Nextera Library prep



1. **TemplateDNA + transposome complex (contain adaptor)**
2. **Tagmentation breaks DNA and add adaptor to ends**
3. **PCR amplification to engineer barcode and sequencing primers**

http://www.gtbiotech.com.tw/products/Nextera_XT_DNA.asp

Hierarchical Genome Assembly



De novo Genome Sequencing

- **Local assembly: with size selection**
  - shotgun PE: *end-overlapping*
  - 200~600bp insert; PE2*150~300bp, 50-200X coverage
- **Scaffolding:**
  - Mate-pair: 2~15kb (size selected), 20-50X
  - 454 long PE: 20~30kb jump
- **Bridging** over small gaps: **LIPE**
  - Shotgun 1~1.6kb insert
- **Gap filling**: PacBio, NanoPore long Reads
  - 1~30 kb

Re-sequencing: Variant detection

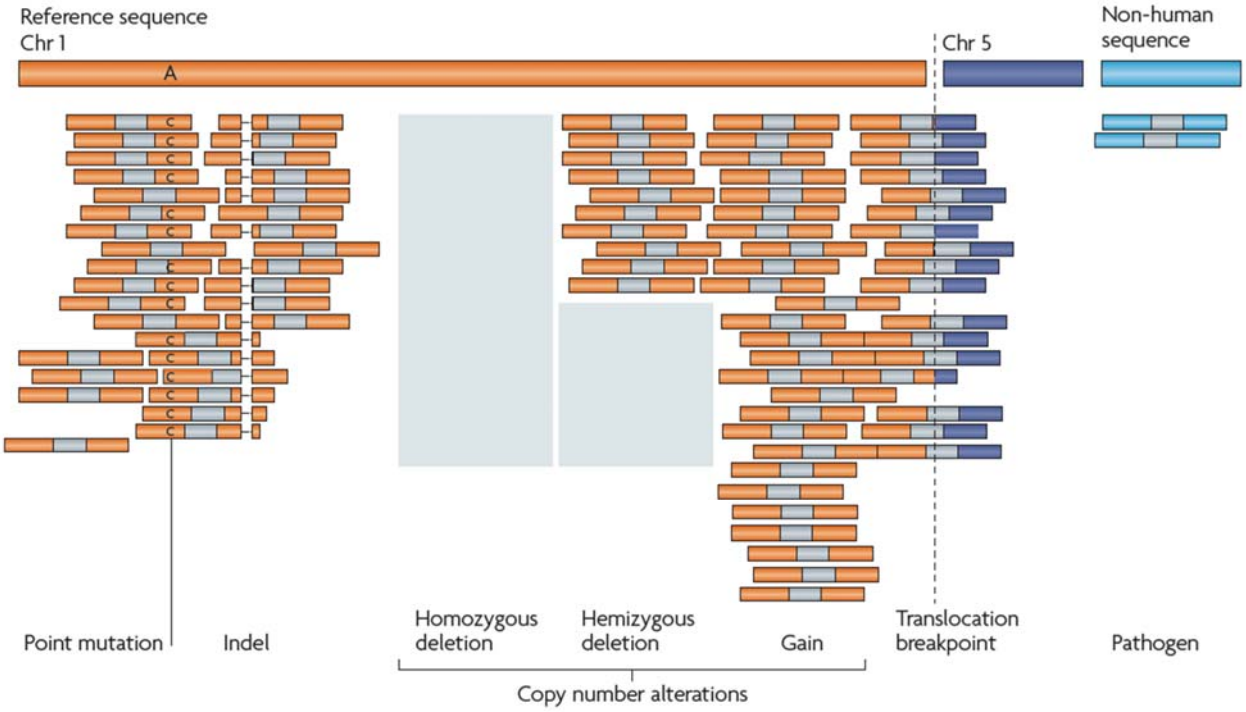


Figure 3 | **Types of genome alterations that can be detected by second-generation sequencing.** Sequenced

Meyerson et al. NRG 2010

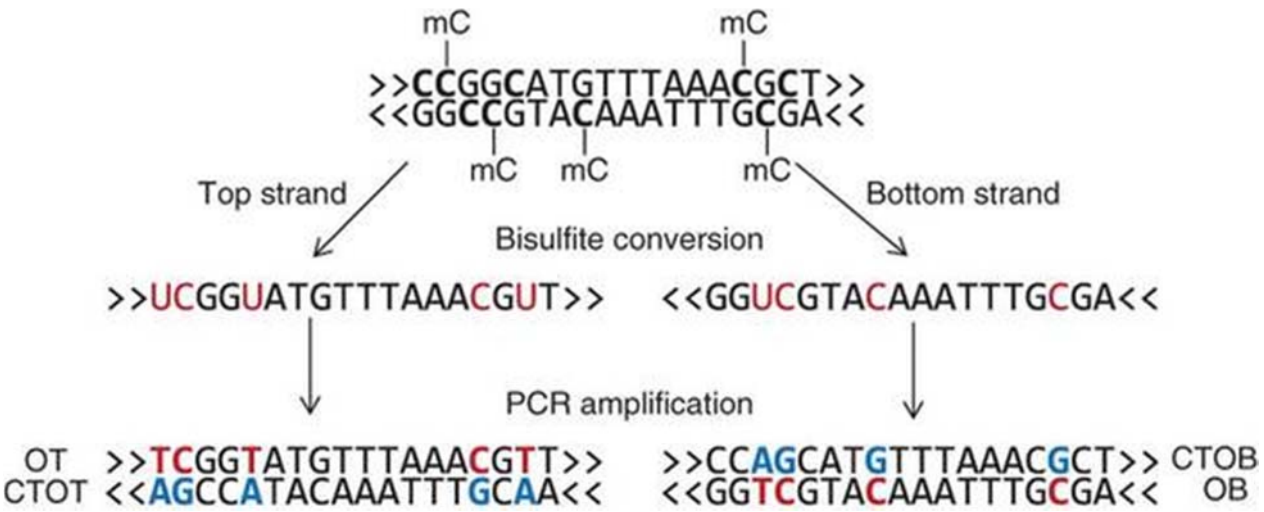
Whole Genome Re-Sequencing

- **Variant calling: SNPs & short INDELs**
  - shotgun PE: SR or PE
  - Read length: 100~150bp
  - >30X coverage per haploid genome
- **Genome re-arrangement: longer INDELs and break point**
  - Gel-size selection
  - PE2*150~300; 30-50X coverage
- **Copy number variation:**
  - PE2*100-150; >30X coverage

Epigenetic sequencing

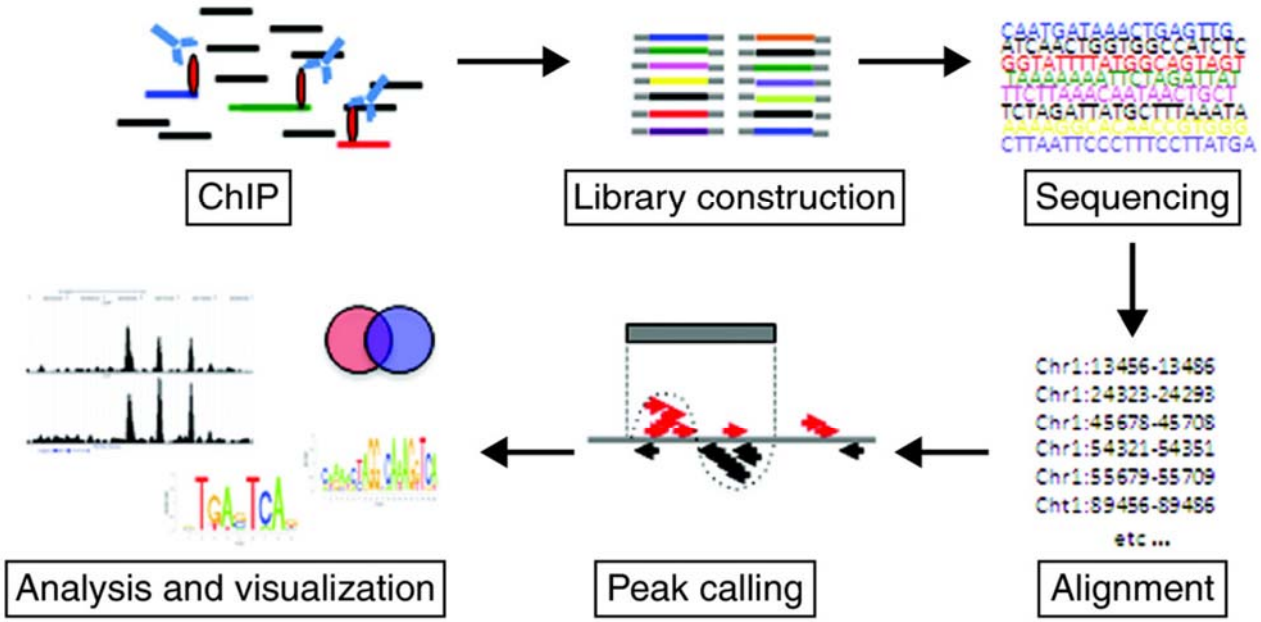
- ***Bisulfite-seq***
- ***ChIP-seq***
- ***Histone-IP***

Effect of bisulfite treatment of DNA

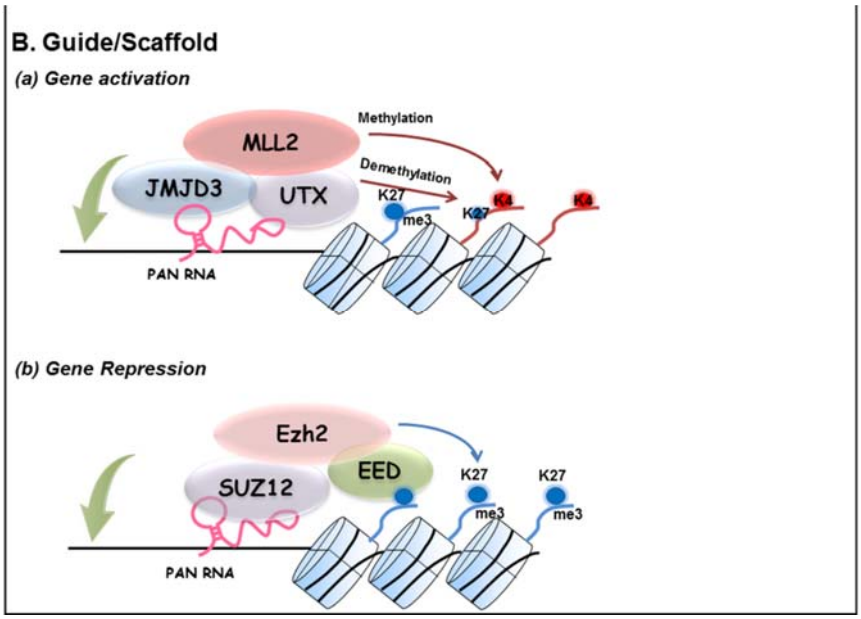


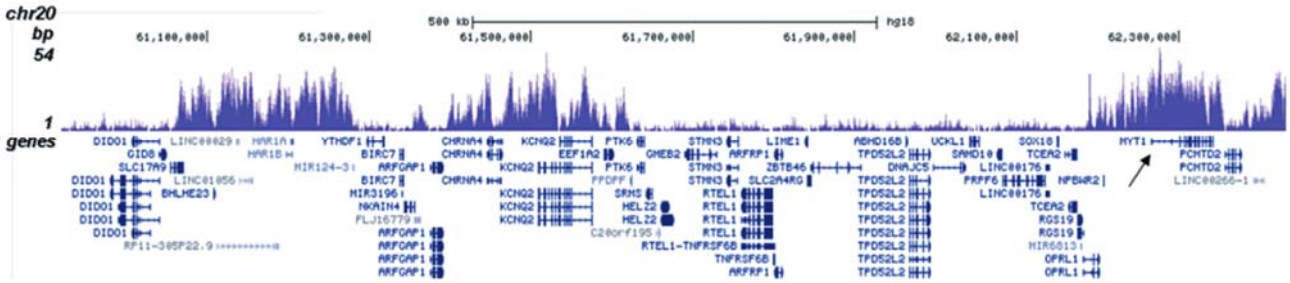
http://www.nature.com/nmeth/journal/v9/n2/full/nmeth.1828.html

ChIP-seq procedure



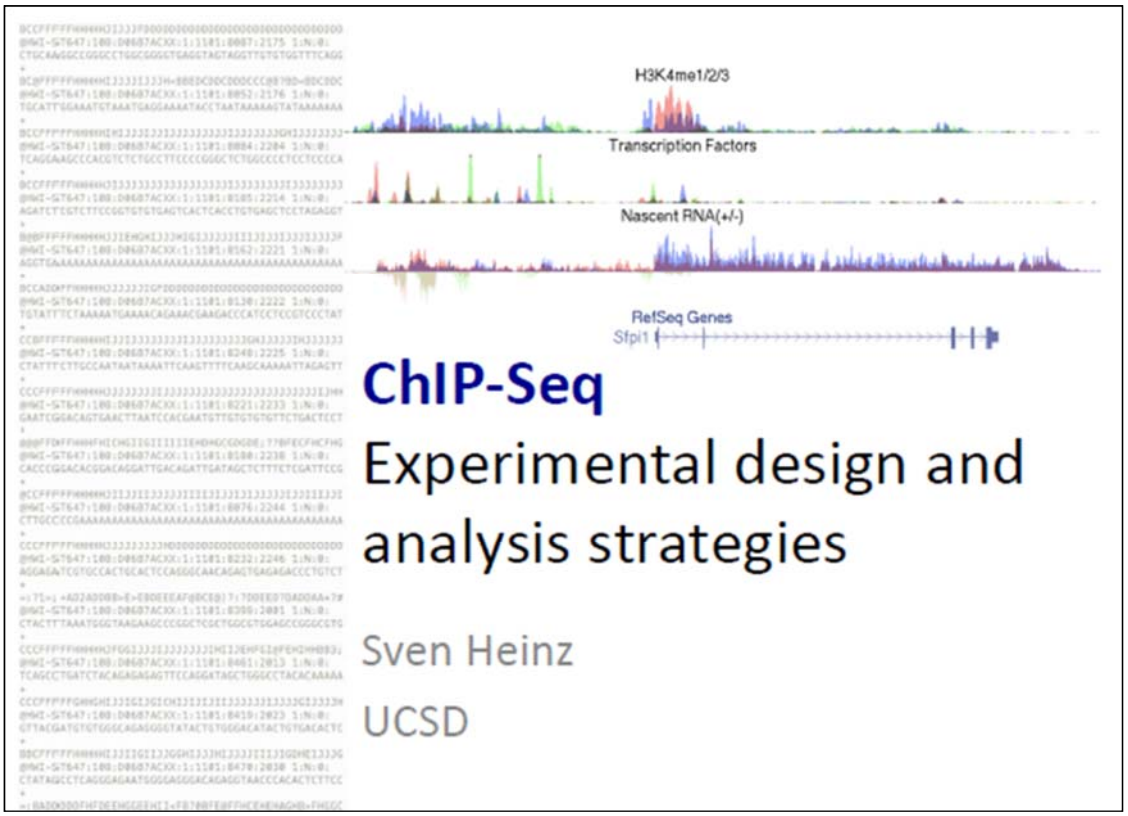
Anti-K27me3





http://www.mdpi.com/1999-4915/6/11/4165/htm

ChIP-seq Experimental Design



http://pharmacology.ucsd.edu/graduate/courseinfo/BIOM231-SP13_8_ChIPseq.pdf

ChIP-Seq considerations

1. **Sample requirement:**
   - Qubit assay: 10-100 ng (eg. 15 ng in 40 ul)
   - Purity: OD 260/280 ratio at 1.7-1.9
   - Smear pattern: BioAnalyzer HS DNA, distribution at 0.2-0.8kb, major~250bp
   - Optional: qPCR confirmation of target genes
2. **Controls:**
   - Negative: Mock (no primary Ab), Total INP, un-induced
   - Positive: primary Ab (ChIP-grade) from independent sources
3. **Sequencing format:**
   - SE or PE; read length 50-100nt
   - barcode 12-48-plex

Sequencing and data QC

- **Sequencing Depth and Barcoding:**
  - 10-20 million reads can reveal can provide reasonable for a ChIP-Seq sample
  - May require more reads to identify rare binding events
  - SR: normal ChIP-Seq; with unique mapping provides quantitative detection
  - PE or long read: to increase mapping precision, or for Allele-specific ChIP-seq

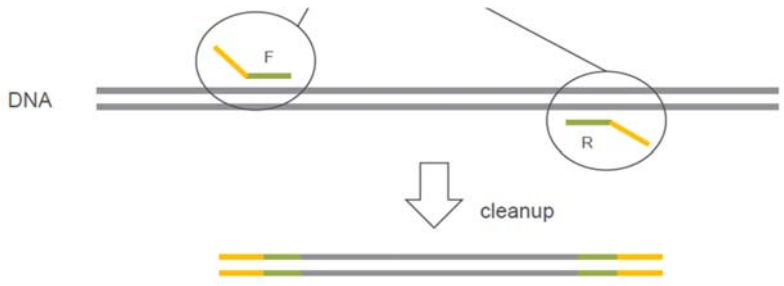
- **Data Quality Control:**
  - low dupreads
  - mapping to both strands
  - similar GC% distribution of whole genome (low GC bias)
  - data visualization in Genome Browser or IGV

Amplicon sequencing

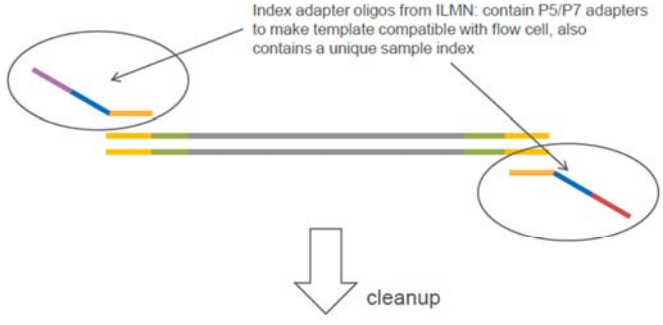
1. *Two-step PCR*
2. *Nextera XT tagmentation*

Amplicon library: 2-step PCR

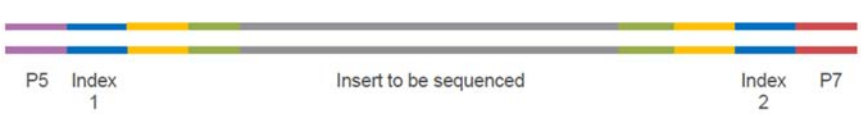
1st round: Amplify ROI



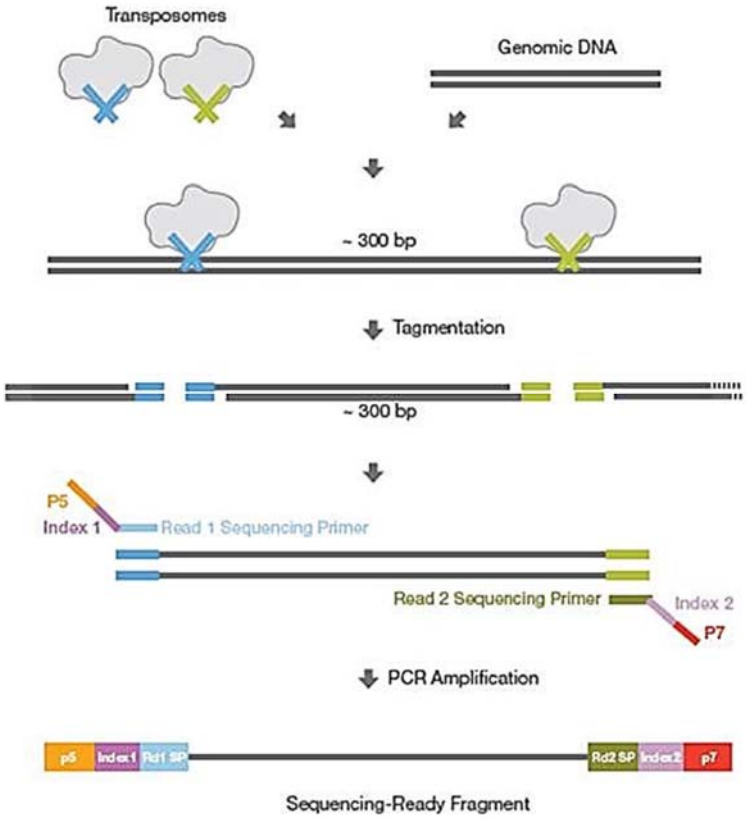
2nd round: add indices + adapters



Best: 400-600bp



Nextera Library prep - Tn tagging



1. TemplateDNA + transposome complex (contain adaptor)
2. Tagmentation breaks DNA and add adaptor to ends
3. PCR amplification to engineer barcode and sequencing primers

http://www.gtbiotech.com.tw/products/Nextera_XT_DNA.asp

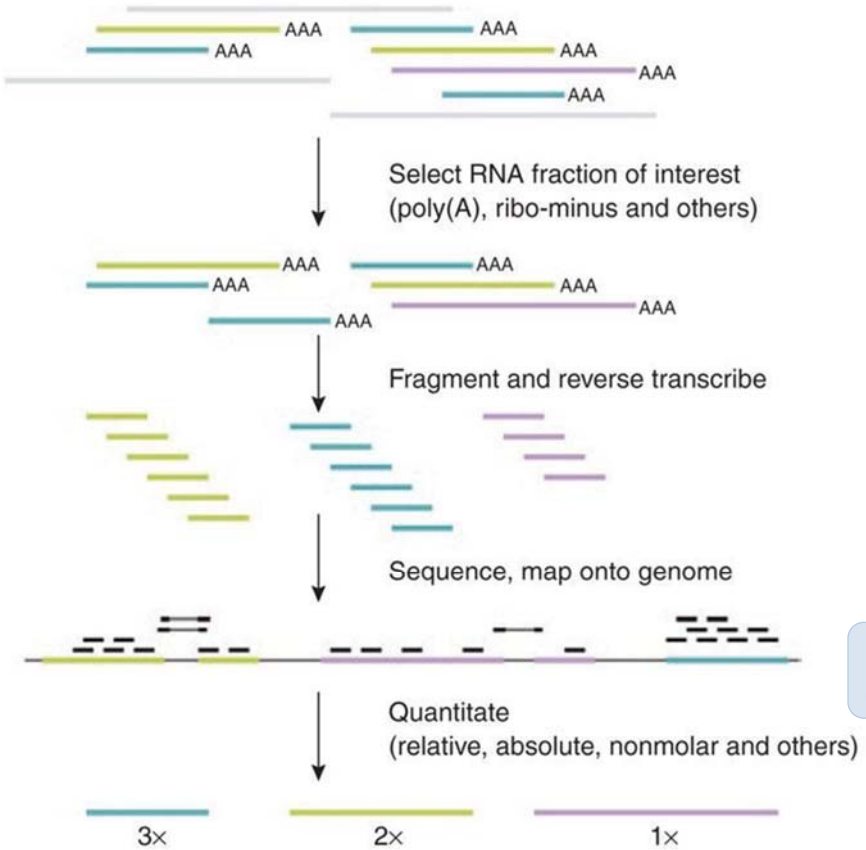
Transcriptome sequencing

- *mRNA vs stranded*
- *smRNA, non-coding RNA*

RNA-seq: considerations

RNA extraction mRNA enrich.	**Total lysate ppt vs Column cleanup?** **Removal of abundant RNA?** ***(polyA+, rRNA depletion, DSN, …)***
cDNA synthesis & library prep	**ds-cDNA vs Stranded-specific?** ***(overlapping genes? non-coding RNA?)***
Sequencing; coverage	**Read format & length** ***PE:** de novo, large genome, PE150~200* ***SR:** re-seq, small genome, SR100* ***miRNA, degradome:** SR50*

Transcriptome profiling: RNA-seq



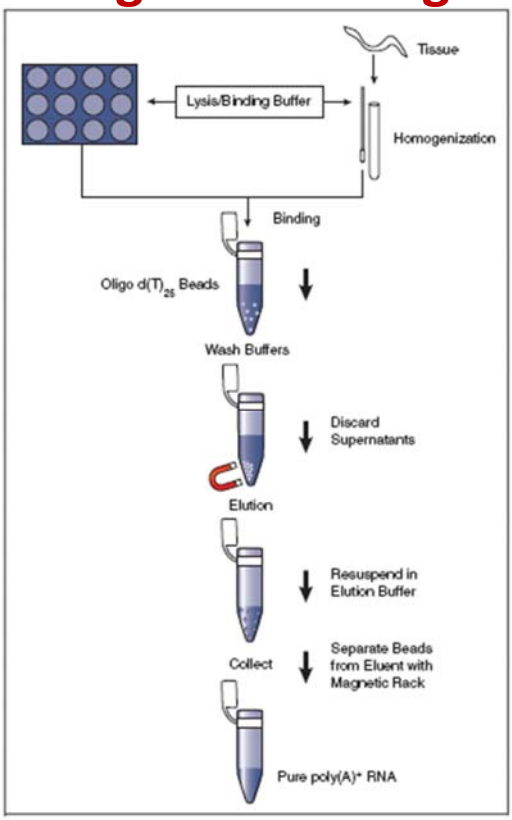
RPKM (FPKM)

$$\frac{\text{Mapped Reads}}{\text{Gene length (Kb) x total reads (M)}}$$

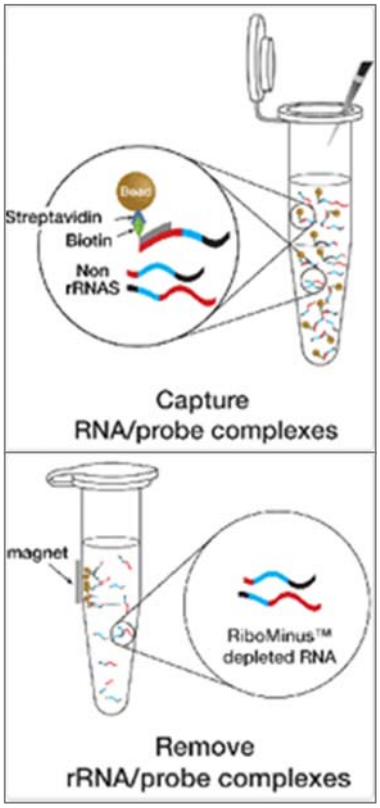
Nature Methods 6, S22 - S32 (2009)

mRNA enrichment

Oligo-dT binding

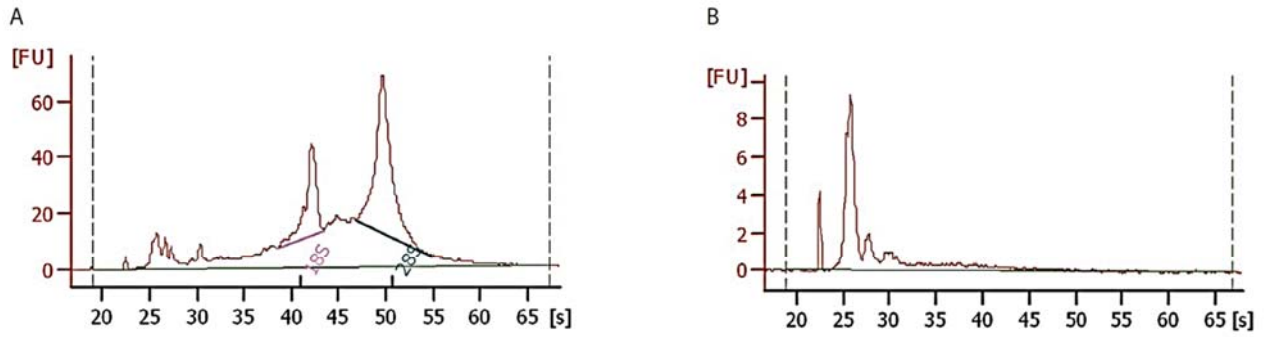


rRNA removal



rRNA Removal from Eukaryotic RNA

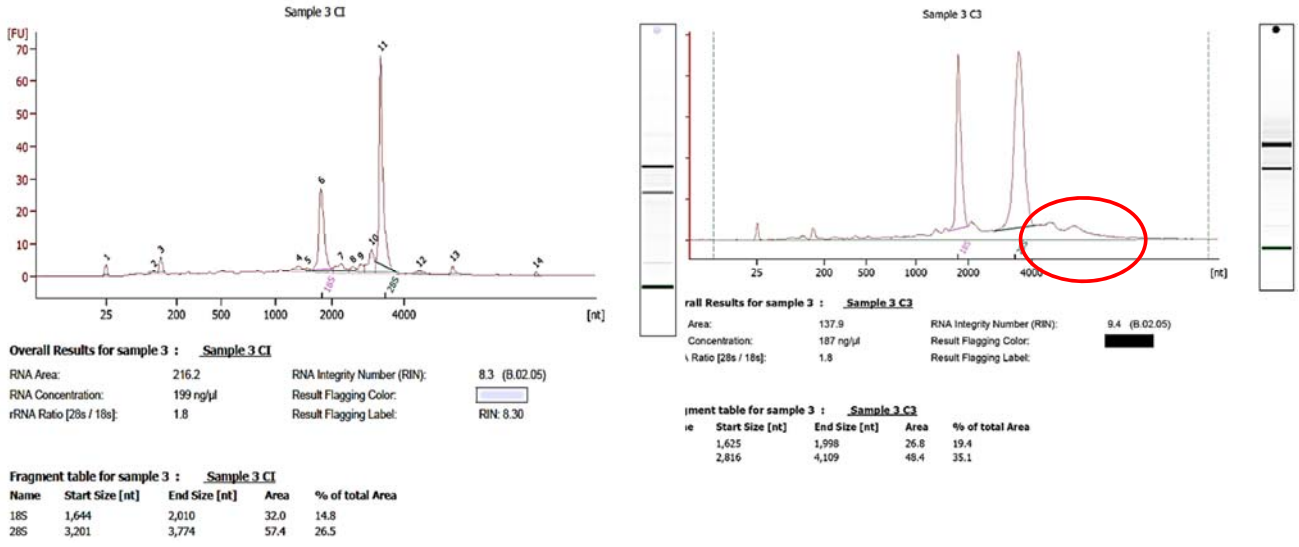
Figure 2. Removal of rRNAs from total RNA.



Removal of ribosomal RNAs from total RNA. Total RNA from MCF-7 cells was examined using a nanochip on a Bioanlayzer before (**A**) and after (**B**) treatment with the Ribo-Zero Gold Kit.

rRNA depletion works for degraded RNAs, but require deeper sequencing.

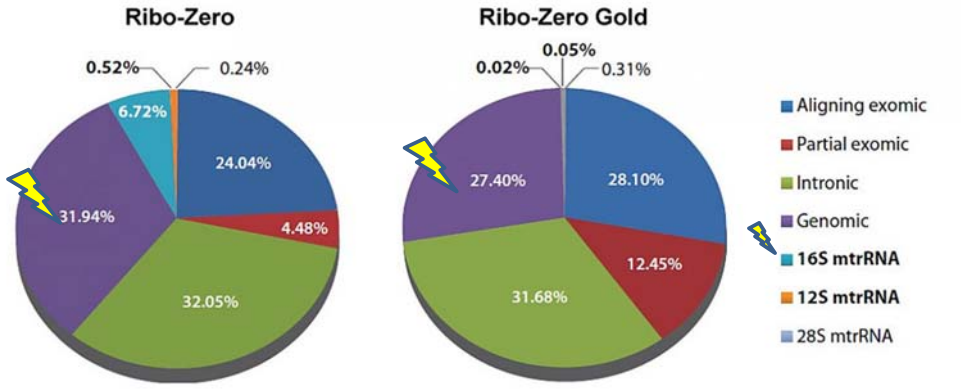
DNA contamination in RNA sample



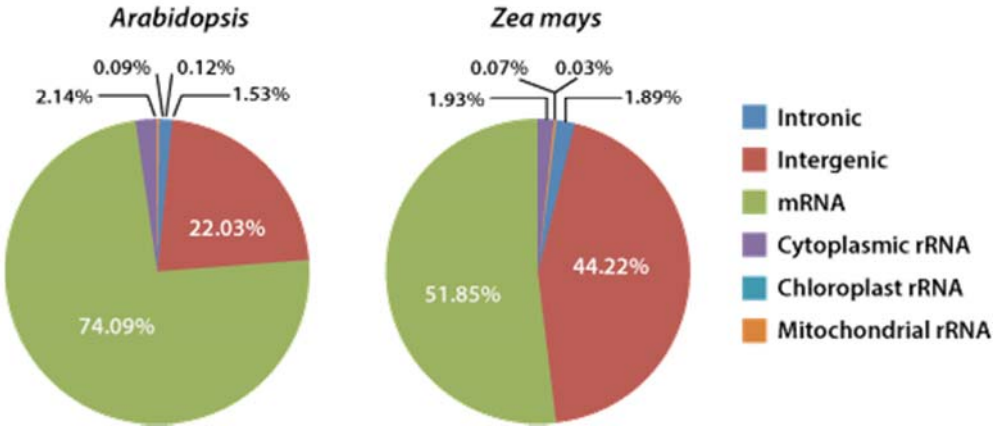
Have seen contamination up to 30%!!

rRNA removal by RiboZero depletion

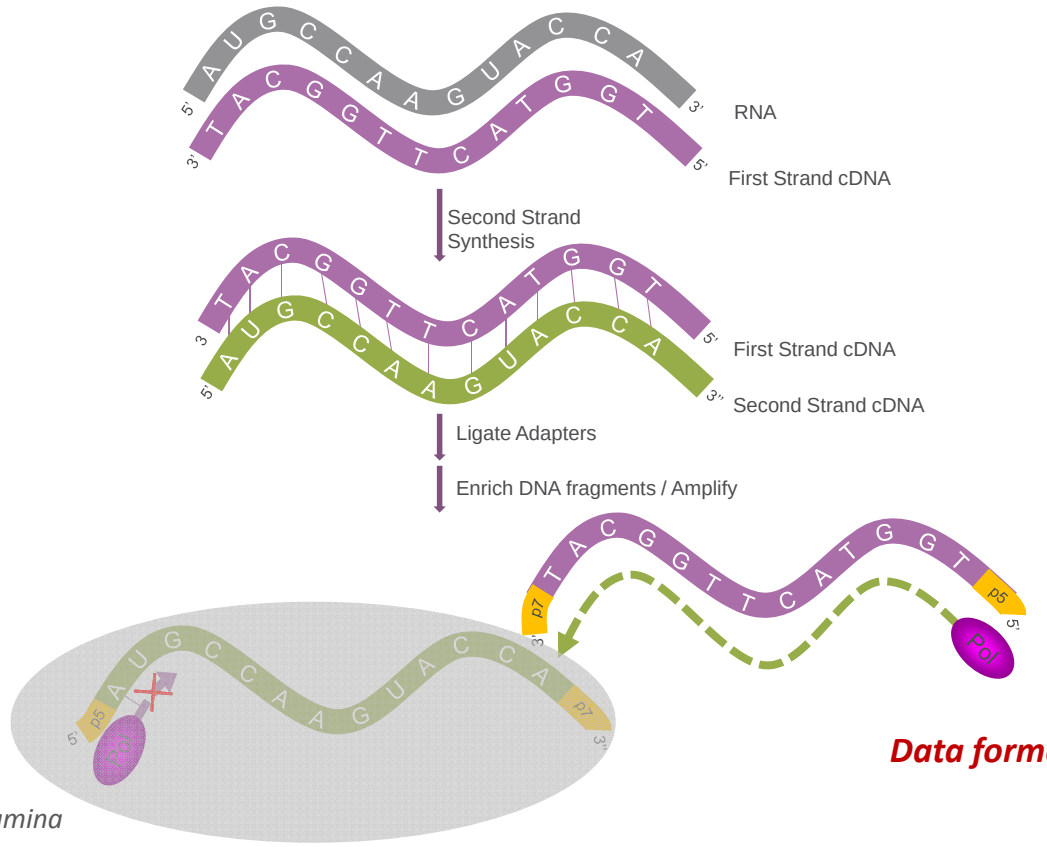
Oligo version



Species divergence



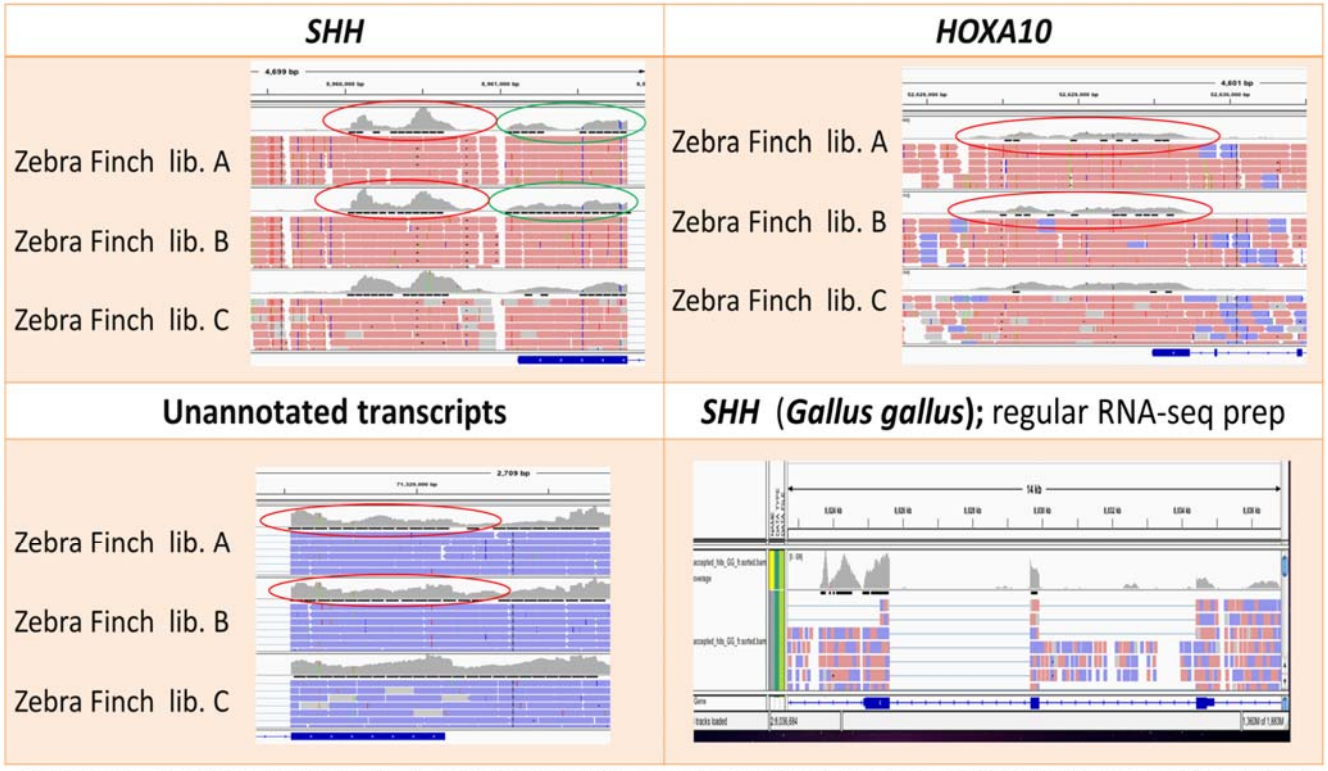
Strand-specific RNA-seq prep



Data format: RF

Source: Illumina

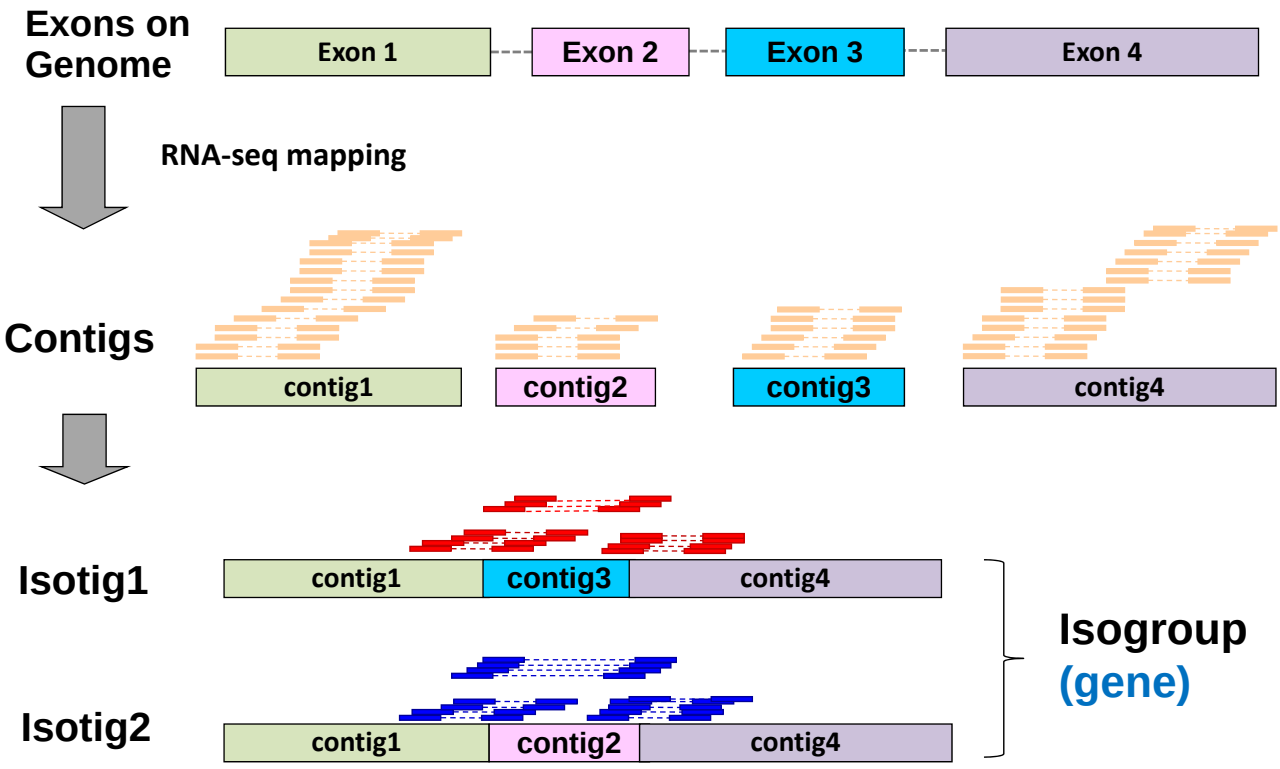
Stranded RNA-seq: >90% strand-specificity



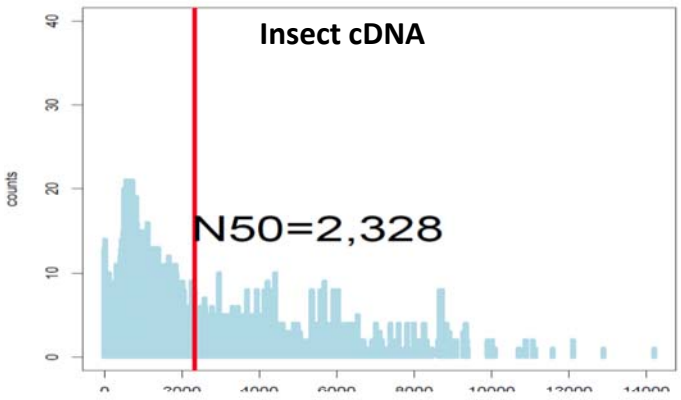
*Genome viewer: IGV genome viewer (version 2.2), plus strand(Blue), minus strand(Red), no paired end info (could be plus or minus strand, Gra
non-stranded protocol, the amount of blue alignments are almost equal to red alignment

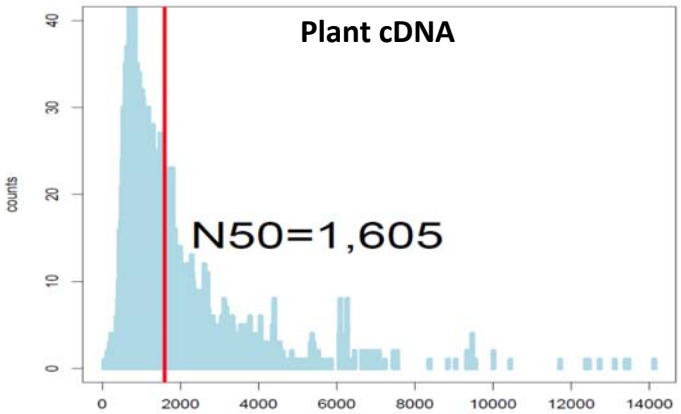
Courtesy of Chih-kuan Chen

Transcriptome assembly



454 Isotig histogram and assembly stats

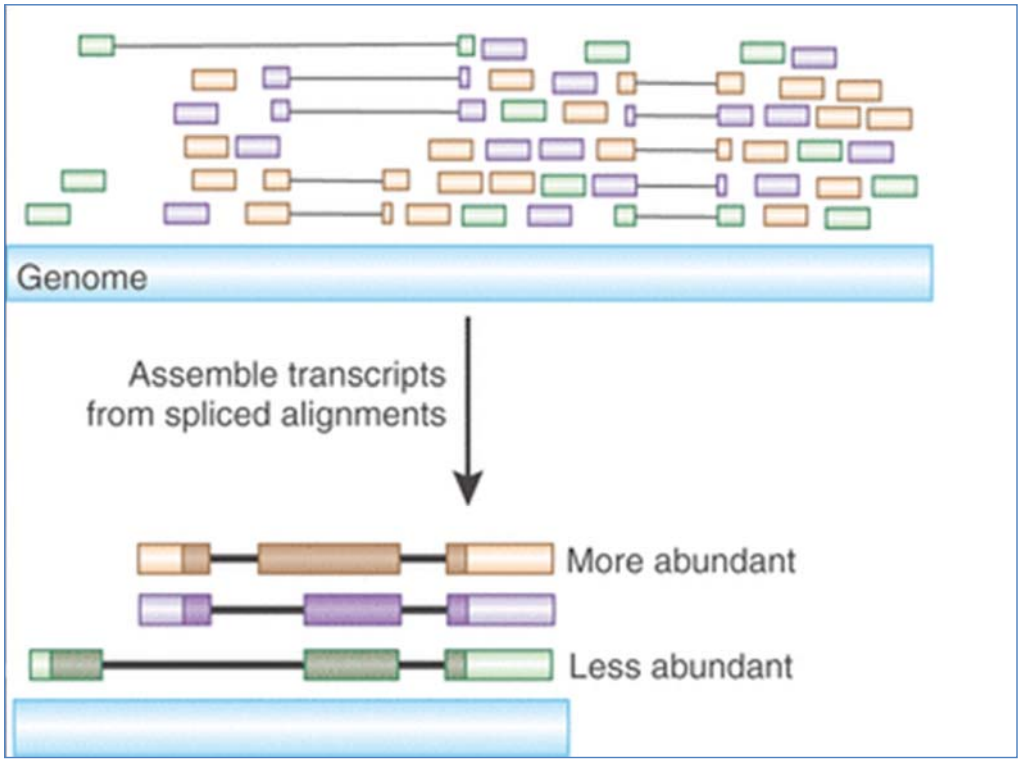


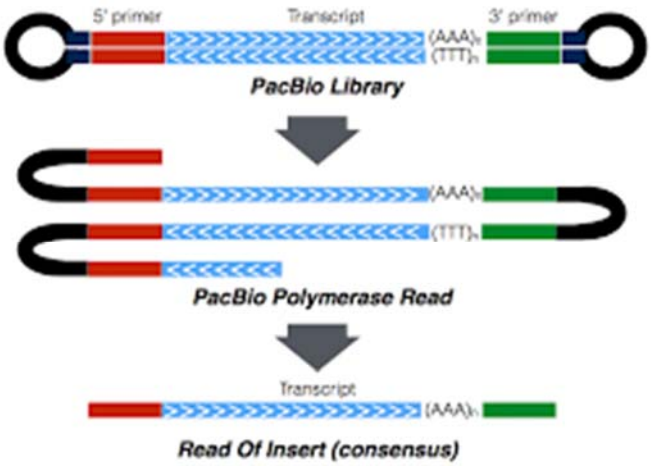


Newbler 3.0 assembly	Insect cDNA	Plant cDNA
total # of Reads	994,094	1,704,310
total # of Bases	428,812,259	726,697,350
isotig Metrics (Transcript)		
# Isotigs	16,983	31,349
# Bases	26,766,462	43,294,186
Avg IsotigSize	1,576	1,381
N50 IsotigSize	2,328	1,605
Largest IsotigSize	14,210	14,135
isogroup Metrics (Gene)		
# Isogroups	12,843	20,939
avgContigCnt	1.6	1.9
largestContigCnt	250	464
avgIsotigCnt	1.3	1.5
largestIsotigCnt	72	64
# with OneIsotig	10,903	14,657

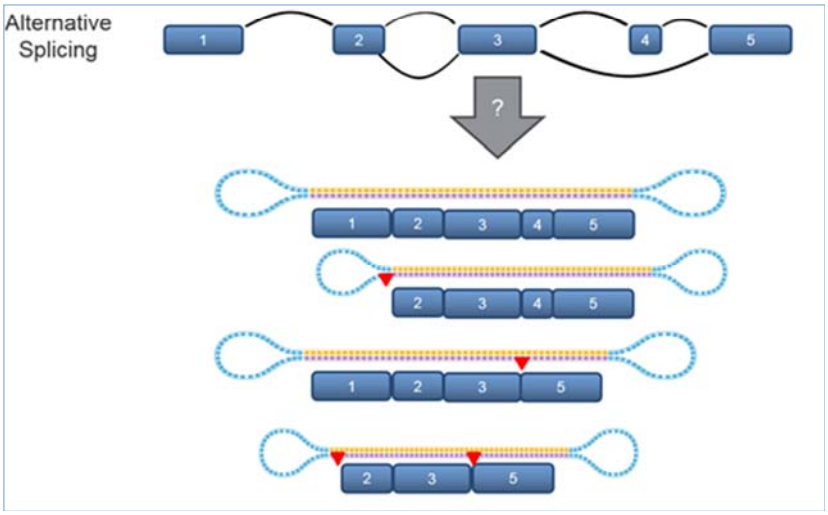
By: Jin Tang, Yuchuan Teng

TopHat and Cufflinks





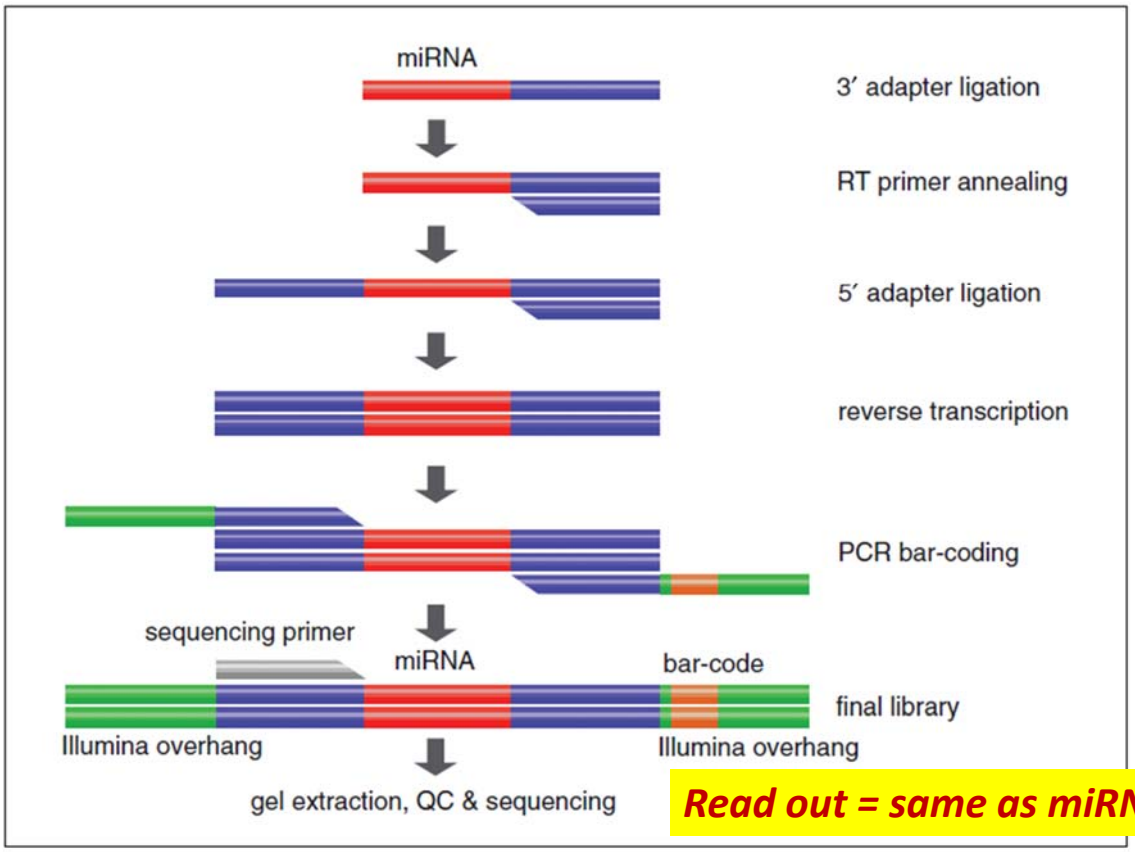
PacBio: Iso-Seq



Transcriptome Sequencing

- **Sample:**
  - mRNA to pair with miRNA or lincRNA study?
- **mRNA enrichment method**
  - Oligo-dT vs Ribo-depeletion
- **Controls and Biological replicates**
- **Time course?**
- **Re-sequencing (mapping) vs de novo assembly**
- **NGS:**
  - Shotgun: SR or PE
  - PE2*100~200bp
  - Coverage depends on need for detection sensitivity
- **Splicing variants? Fusion junction?**
  - Gel-size selection
  - PE2*150~300

smRNA library prep - Directional



Read out = same as miRNA

Curr Protoc Hum Genet. 2012 Apr;Chapter 11:Unit 11.12.1-10.

Summary of NGS Project Design

1. **Application type:**
   - NGS platform, sequencing format
   - data scale/coverage depth
   - Sequencing bias?
2. **Sample nature:**
   - Ploidity, plastids, pure/meta sample
   - Genome size, GC%, repeats
3. **Experimental Design:**
   - Biological replicate (n=3?)
   - Control set?
4. **NGS bioinformatic algorithms**
   - For long or short reads? SR vs PE?
   - Computational demand

III. NGS Core & Prep works

- Sample requirement
- Considerations
- Submission to NGS Core

High Throughput Genomics Core at BRCAS

Welcome | Publications | Instruments | Get Started! | Services & Charges | Documentation | Contact & Location | FAQs | Calendar

Documentation

Home / Documentation

http://ngs.biodiv.tw/NGSCore/documentation/

Forms

Sequencing Application Form for Illumina	Download
Sequencing Application Form for PacBio Sequel	Download
Sequencing Application Form for Roche 454	Download
Sample Submission Form	Download

Guides

NGS Sample Preparation Guide	Download
Guidelines for Chloroform Purification (PacBio)	Download
16S Metagenomic Library Prep Guide (Illumina)	Download
Info for mRNA-Seq and smRNA-Seq	Download
Chip-Seq Q&A	Download

Core News

- Launching of the 3rd-Gen sequencing service of PacBio Sequel system
- Promotional Discount on HiSeq
- 本核心實驗室之儀器使用效益獲得肯定
- Web server (Pydio) has been disabled

Lecture & Seminar

- 2016-03-24

Related Web Links

	Application Type
(S-B)	TruSeq Methylation DNA (Bisulfite Seq)
(S-Ca)	TruSeq Paired-End DNA, Gel free
(S-Cb)	TruSeq Paired-End DNA, Gel plus
(S-D)	TruSeq Long Insert Paired-End DNA (1-2Kb)
(S-E)	TruSeq Small RNA (miRNA)
(S-G)	Degradome
(S-I)	TruSeq ChIP DNA (ChIP Seq)
(S-Ma)	Nextera Mate-Pair DNA, Gel free (no size selection)
(S-Mb)	Nextera Mate-Pair DNA, Gel plus (4 sizes, 1-15Kb)
(S-Mc)	Nextera Mate-Pair DNA, Gel plus (6 sizes, 1-40Kb)
(S-N)	TruSeq Synthetic Long-Read DNA
(S-P)	Indexing PCR (2nd Step PCR)
(S-Q)	Nextera DNA
(S-R)	Nextera XT DNA
(S-S)	TruSeq Stranded RNA, Poly-A
(S-T)	TruSeq Stranded RNA, Ribo-zero
(S-Wa)	Low-Input Stranded RNA, Poly-A (Clonetech)
(S-Wb)	Low-Input Stranded RNA, Ribo-zero (Clonetech)
(S-Wc)	Low-Input Stranded RNA, aRNA (Epicentre)
(S-Xa)	Human Exome Capture (Agilent)
(S-Xb)	Human Exome Capture (SeqCap)

Illumina Hiseq 2500 Sequen Services

HiSeq High-Throughput Mode (8 lanes

	Mode Type (200-250M clusters per lane)
(HTSR100)	HiSeq HT v4, Single Read 100
(HTPE125)	HiSeq HT v4, Paired End 2*125

HiSeq Rapid Mode (2 lanes per chip)

	Mode Type (150-180M clusters per lane)
(RpSR50)	HiSeq Rapid v2, Single Read 50
(RpSR100)	HiSeq Rapid v2, Single Read 100
(RpPE100)	HiSeq Rapid v2, Paired End 2*100
(RpPE150)	HiSeq Rapid v2, Paired End 2*150
(RpPE200)	HiSeq Rapid v2, Paired End 2*200
(RpPE250)	HiSeq Rapid v2, Paired End 2*250

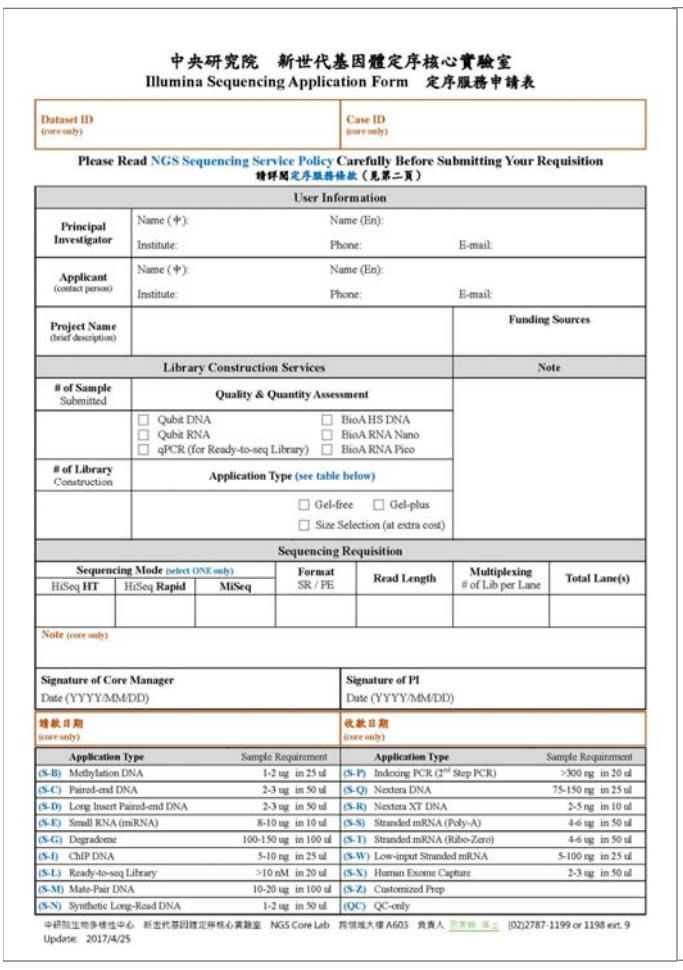
Illumina MiSeq Sequencing S

	Mode Type (10-15M clusters in v2; 20-25M clusters in v3)
(MS50)	MiSeq 50 cycles v2, PE2*25 or SR50
(MS150)	MiSeq 150 cycles v3, PE2*75 or SR150
(MS300)	MiSeq 300 cycles v2, PE2*150 or SR300
(MS500)	MiSeq 500 cycles v2, PE2*250
(MS600)	MiSeq 600 cycles v3, PE2*300

NGS Service Workflow

Inquire NGS service	Contact Core Manager to inquire service application of interest
Project Consultation	Project discussion regarding sample organism, sequence type, and scale of data required
Sample quality pre-check	Prepare Sample Information Form and email to Manager for visual quality inspection
Submit sample and forms	Send approved samples and forms to core, signed in by Core Manager for project establishment
QC test by Core	Check DNA/RNA concentration, purity, and quality
Library construction	Application-specific procedure; PCR amplification
Library QC	Check library size range and concentration; qPCR normalization
Sequencing	Application dependent sequencing (SR vs PE; read length of choice)
Data preparation	Demultiplexing; Read folder generation, summary file (optional: assembly or mapping analysis upon request)
Return data to user	Transfer data/files by data retrieval server (Pydio), portable HD

http://140.109.28.108/NGSCore/?page_id=40



中央研究院　新世代基因體定序核心實驗室
Illumina Sequencing Application Form　定序服務申請表

Dataset ID (core only)	Case ID (core only)

Please Read NGS Sequencing Service Policy Carefully Before Submitting Your Requisition
請詳閱定序服務條款（見第二頁）

User Information			
Principal Investigator	Name (中):	Name (En):	
	Institute:	Phone:	E-mail:
Applicant (contact person)	Name (中):	Name (En):	
	Institute:	Phone:	E-mail:
Project Name (brief description)			Funding Sources

Library Construction Services		Note
# of Sample Submitted	Quality & Quantity Assessment	
	☐ Qubit DNA ☐ BioA HS DNA ☐ Qubit RNA ☐ BioA RNA Nano ☐ qPCR (for Ready-to-seq Library) ☐ BioA RNA Pico	
# of Library Construction	Application Type (see table below)	
	☐ Gel-free ☐ Gel-plus ☐ Size Selection (at extra cost)	

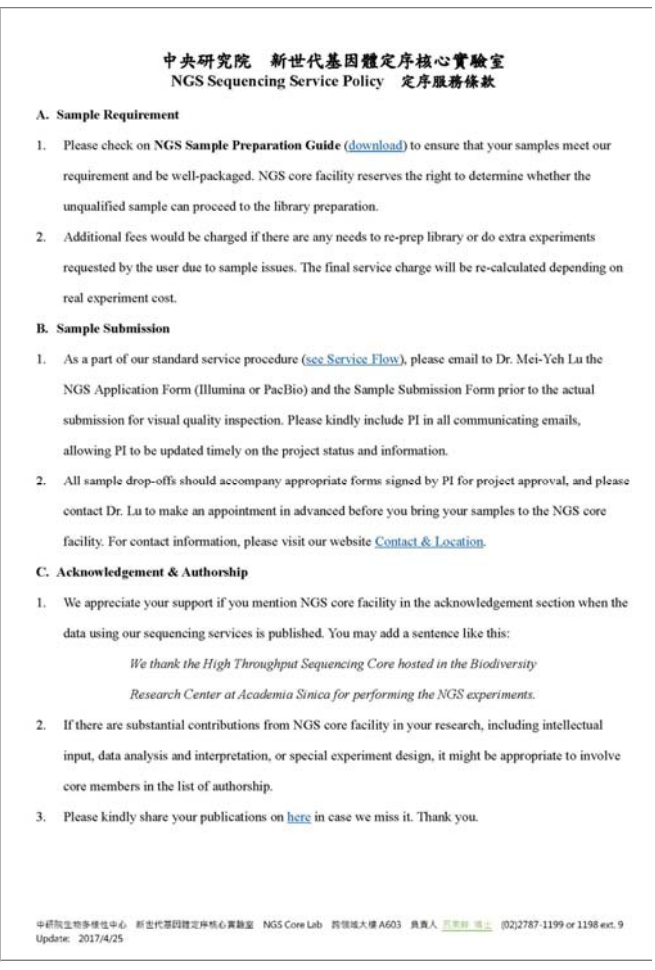
Sequencing Requisition						
Sequencing Mode (select ONE only)			Format	Read Length	Multiplexing	Total Lane(s)
HiSeq HT	HiSeq Rapid	MiSeq	SR / PE		# of Lib per Lane	

Note (core only)

Signature of Core Manager	Signature of PI
Date (YYYY/MM/DD)	Date (YYYY/MM/DD)
接收日期 (core only)	收款日期 (core only)

Application Type	Sample Requirement	Application Type	Sample Requirement
(S-B) Methylation DNA	1-2 ug in 25 ul	(S-P) Indexing PCR (2nd Step PCR)	>300 ng in 20 ul
(S-C) Paired-end DNA	2-3 ug in 50 ul	(S-Q) Nextera DNA	75-150 ng in 25 ul
(S-D) Long Insert Paired-end DNA	2-3 ug in 50 ul	(S-R) Nextera XT DNA	2-5 ng in 10 ul
(S-E) Small RNA (miRNA)	8-10 ug in 10 ul	(S-S) Stranded mRNA (Poly-A)	4-6 ug in 50 ul
(S-G) Degradome	100-150 ug in 100 ul	(S-T) Stranded mRNA (Ribo-Zero)	4-6 ug in 50 ul
(S-I) ChIP DNA	5-10 ng in 25 ul	(S-W) Low-input Stranded mRNA	5-100 ng in 25 ul
(S-L) Ready-to-seq Library	>10 nM in 20 ul	(S-X) Human Exome Capture	2-3 ug in 50 ul
(S-M) Mate-Pair DNA	10-20 ug in 100 ul	(S-Z) Customized Prep	
(S-N) Synthetic Long-Read DNA	1-2 ug in 50 ul	(QC) QC-only	

中研院生物多樣性中心　新世代基因體定序核心實驗室　NGS Core Lab　跨領域大樓 A603　負責人 [illegible]　(02)2787-1199 or 1198 ext. 9
Update: 2017/4/25



中央研究院　新世代基因體定序核心實驗室
NGS Sequencing Service Policy　定序服務條款

A. Sample Requirement

1. Please check on **NGS Sample Preparation Guide** (download) to ensure that your samples meet our requirement and be well-packaged. NGS core facility reserves the right to determine whether the unqualified sample can proceed to the library preparation.
2. Additional fees would be charged if there are any needs to re-prep library or do extra experiments requested by the user due to sample issues. The final service charge will be re-calculated depending on real experiment cost.

B. Sample Submission

1. As a part of our standard service procedure (see Service Flow), please email to Dr. Mei-Yeh Lu the NGS Application Form (Illumina or PacBio) and the Sample Submission Form prior to the actual submission for visual quality inspection. Please kindly include PI in all communicating emails, allowing PI to be updated timely on the project status and information.
2. All sample drop-offs should accompany appropriate forms signed by PI for project approval, and please contact Dr. Lu to make an appointment in advanced before you bring your samples to the NGS core facility. For contact information, please visit our website Contact & Location.

C. Acknowledgement & Authorship

1. We appreciate your support if you mention NGS core facility in the acknowledgement section when the data using our sequencing services is published. You may add a sentence like this:

   We thank the High Throughput Sequencing Core hosted in the Biodiversity Research Center at Academia Sinica for performing the NGS experiments.

2. If there are substantial contributions from NGS core facility in your research, including intellectual input, data analysis and interpretation, or special experiment design, it might be appropriate to involve core members in the list of authorship.
3. Please kindly share your publications on here in case we miss it. Thank you.

中研院生物多樣性中心　新世代基因體定序核心實驗室　NGS Core Lab　跨領域大樓 A603　負責人 [illegible]　(02)2787-1199 or 1198 ext. 9
Update: 2017/4/25

中央研究院　新世代基因體定序核心實驗室
NGS Sample Submission Form　樣品明細表

Submission Date	Case ID (core only)
Application Type	

1. Please refer to **NGS Sample Preparation Guide** for the labeling and packing, and ensure the quantity and quality of your samples fulfill all the requirements.
2. Please attach **Gel Image (required)** or **BioAnalyzer Traces (optional)** at the end of the NGS Sample Submission Form.
3. **Sample Name** must be **the same** as the labeling on the tube cap.

Organism or Species		Sample Type: ☐ genomic DNA ☐ amplicon ☐ total RNA ☐ ChIP DNA ☐ cDNA ☐ Ready-to-seq Library ☐ plasmid DNA ☐ mRNA ☐ _____
Genome Size or DNA Length		
Purification Method		Enzyme Treatment & Usage: ☐ DNase ☐ RNase ☐ RNase Inhibiter　Dissolved in: ☐ H_2O ☐ EB ☐ _____

	Sample Name (tube labeling)	(optional) Qubit (ng/ul)	Nano Drop (ng/ul)	Vol. (ul)	Amt. (ug)	OD 260/280	OD 260/230	(optional) rRNA Ratio	(optional) RIN	Notes	Sample ID (core only)
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											

中研院生物多樣性中心　新世代基因體定序核心實驗室　NGS Core Lab　跨領域大樓 A603　負責人 呂英銘 博士　(02)2787-1199 or 1198 ext. 9
Update: 2017/4/25

中央研究院　新世代基因體定序核心實驗室
Supplemental Information of NGS Samples

Gel Image (required)
1. **Post-run staining** only.
2. Please indicate **sample no.** & **major marker sizes** (ladder should cover at least 0.1-10kb).

a. Gel percentage: ____% of ☐ TAE or ☐ TBE agarose gel
b. Run condition: ____voltage for ____min
c. Loading amount: ____ng of sample per lane; ____ng of ladder per lane

中央研究院　新世代基因體定序核心實驗室
Supplemental Information of NGS Samples

BioAnalyzer Electropherogram (optional)
1. Please arrange images of BioA traces according to the sample number on the sample submission form.
2. **BioA DNA** ☐ High Sensitive ☐ 1000 ☐ 7500 ☐ 12000; or **BioA RNA 6000** ☐ Nano ☐ Pico

Sample # 1	Sample # 2
Sample # 3	**Sample # 4**

高通量定序核心實驗室樣品送件流程

1. 填妥樣品送件單，E-mail至NGS Core負責人信箱，樣品確認沒問題即可送件
2. 參考NGS Core樣品送件須知，將樣品備妥
3. 預約定序樣品送件時間，即可進行送件
4. QC通過後將開始進行收費
5. 定序完成後寄送通知信件，即可上網下載定序資料

Sample Submission

1. Prior project communication; email files for preview

2. Submission by appointment:
   a. Samples: Provide samples in low-bind tubes
   b. transferred on ice/dry ice/liquid N2
   c. Submit along with NGS application form & Sample form
   d. Attach gel images and/or BioAnalyzer traces for quality check
   e. Make appointment in advance

General Sample Requirements

DNA:

- RNase-treated and purified
- Submission amount > 3X of library input
- High purity (NanoDrop ratios, BioA, gel)
- Long integrity (>23~48kb)
- Low in inhibitors and contaminants (eg. EDTA, CTAB)
- Concentration: 200-800 ng/ul

RNA:

- DNase-treated and purified
- Submission amount >2X of library input
- High purity (NanoDrop ratios, BioA, gel)
- High in rRNA ratio, RIN
- Low in inhibitors and contaminants
- Concentration:
  - mRNA 200-1000 ng/ul
  - smRNA: 500-2000 ng/ul

Auxiliary equipments

Qubit
Fluorometer

BioAnalyzer
(up to 11 samples)

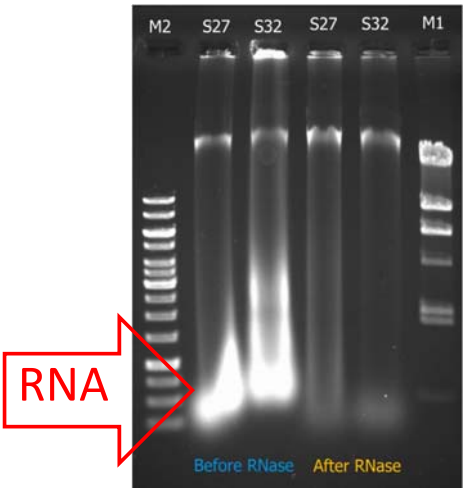
Fragment Analyzer
(up to 96well plates*3)

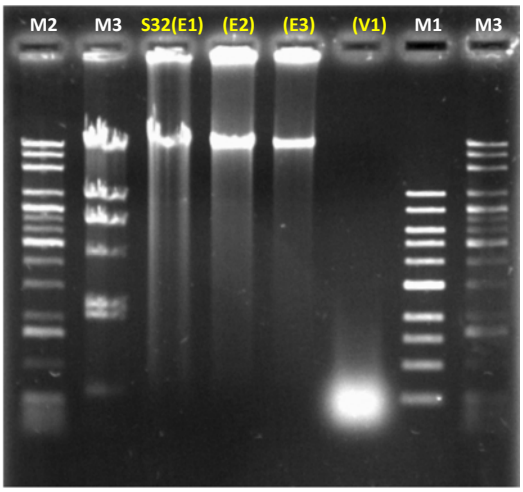
Covaris
(DNA shearing)
0.2~10kb

BluePippin
(gel size selection)

LC480 qPCR

Genomic DNA QC

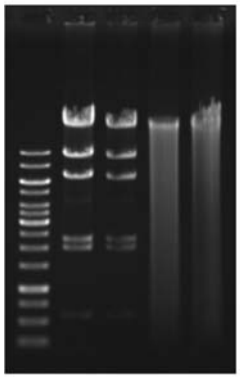




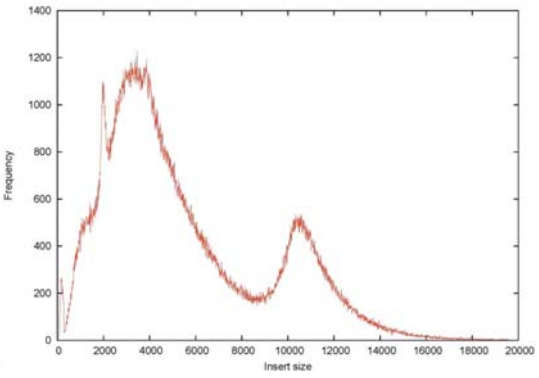
	OD 260/280	OD 260/230	NanoDrop (ng/uL)	Qubit DNA (ng/uL)	RNA Carry over (NanoDrop/Qubit)
S32-original	2.04	2.05	2250.8	56.0	40.19 X
S32_V1	2.16	2.52	1207.40	7.29	165.62 X
S32_E1	1.77	0.94	394.50	131.00	3.01 X
S32_E2	1.67	0.82	45.06	14.10	3.20 X
S32_E3	1.75	0.75	11.48	4.49	2.56 X

Fungal genome MP – 1~20kb

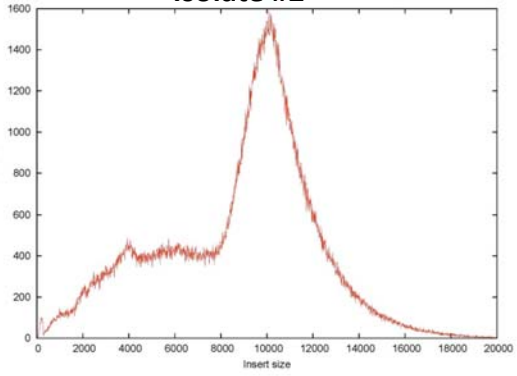
Assembly version	Assembly size (bp)	num. scaffolds	average (kb)	largest scaff (kb)	N50 (kb)	N50 (n)	N90 (kb)	N90 (n)	valid nuc	Ns
Hong Kong version	31,259,288	1,434	22	407	79	122	12	484	31,259,288	0
NTU version	~30Mb	NA	NA	NA	~50kb	NA	NA	NA	NA	NA
BRC v1 (PE+3kb ; sequenced in Japan)	33,490,075	2,555	13	2,449	421	18	48	94	33,052,597	437,478
BRC v2 isolate 1 (PE + 5 MP libraries up to 20kb)	31,254,640	97	322	3,610	2,385	6	650	16	30,893,103	361,537
BRC v2 isolate 2 (PE + 5 MP libraries up to 20kb)	**31,591,454**	**107**	**295**	**5,891**	**3,404**	**4**	**1,061**	**10**	**31,091,726**	**499,728**



Isolate #1



Isolate #2



RNA QC

BioAnalyzer RNA ladder

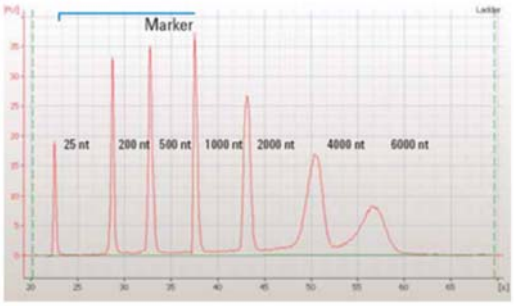
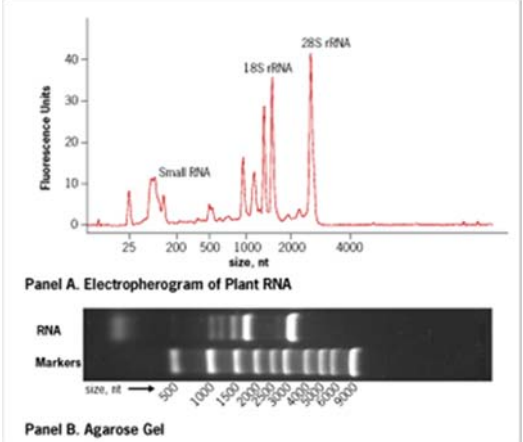
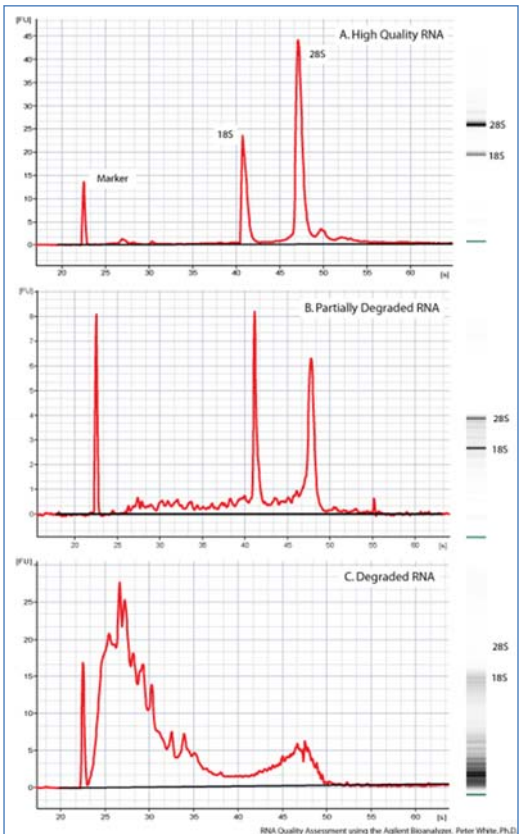


Figure 1 RNA 6000 Nano ladder

Plant total RNA

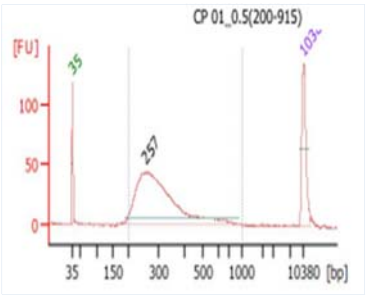


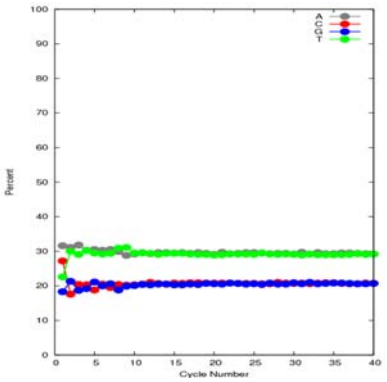
Human RNA – various degradation



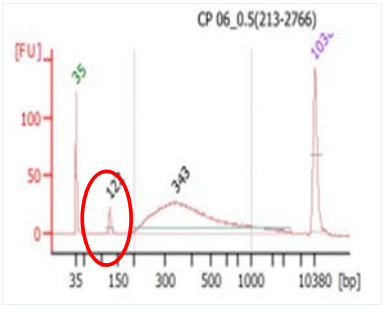
Low Input Sample – biased sequencing

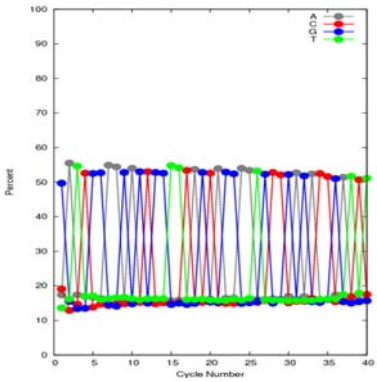
Normal input- Little bias



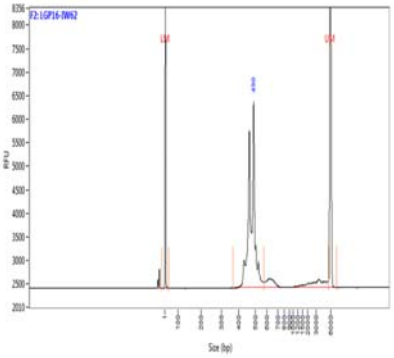


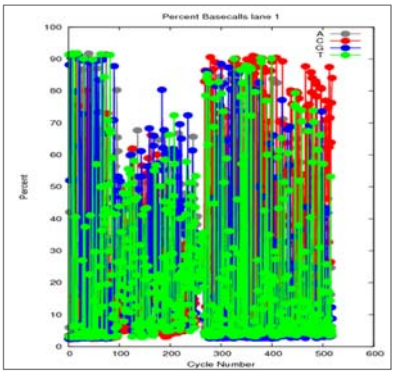
Low input - adaptor bias





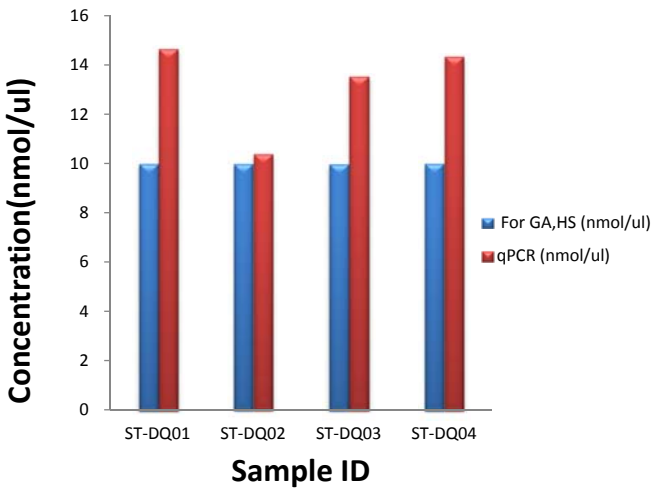
Amcplion – extreme bias

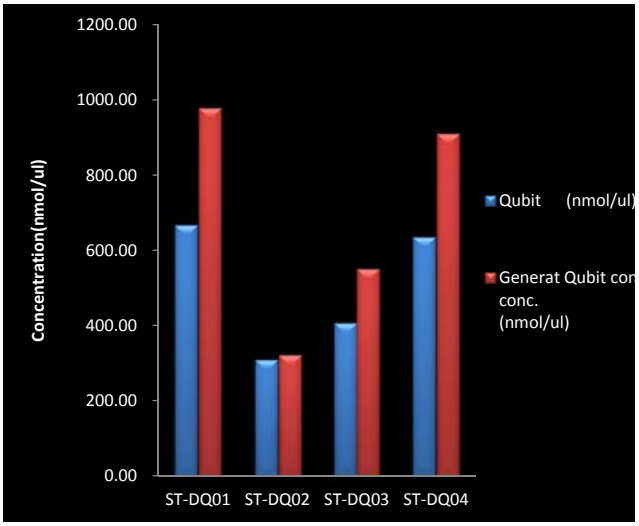




Library molar normalization – NGS qPCR

Sample	bp	Qubit (ng/ul)	Qubit (nmol/ul)	For GA,HS (nmol/ul)	qPCR (nmol/ul)	Ratio of qPCR/Qubit	Generat Qubit conc.to qPCR conc. (nmol/ul)
ST-DQ01	321	141.5	667.90	10	14.66	1.47	**979.14**
ST-DQ02	359	73	308.10	10	10.40	1.04	**320.42**
ST-DQ03	376	101	407.00	10	13.55	1.36	**551.49**
ST-DQ04	361	151.5	635.90	10	14.33	1.43	**911.24**



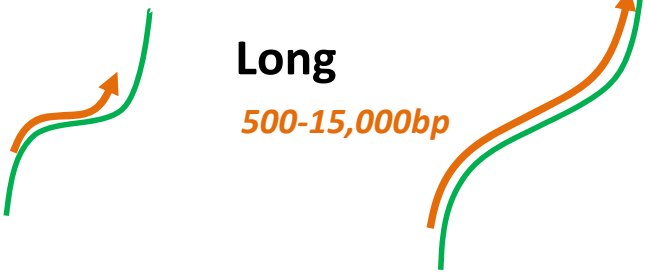


IV. NGS Data & Resources

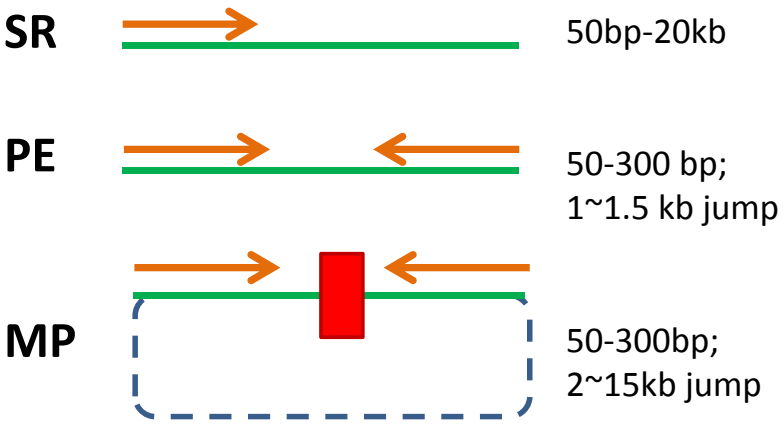
- Data types and QC
- Public resources

Types and Characteristics of NGS Reads

- **Read length:** **Short** *50-300bp* **Long** *500-15,000bp*

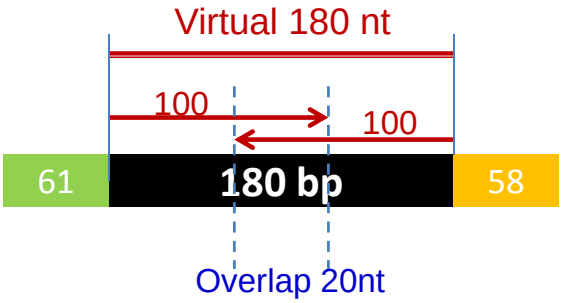


- **Read types:**

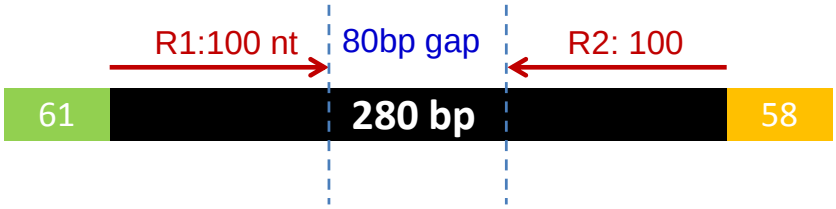


Insert size vs Library Fragment Size

300-bp fragment:
Ends overlapped 20 bp



400-bp fragment:
Ends gapped by 80 bp



Illumina Read – fastQ file

Index sequence
no control
Y/N: failing PF or not
Sequence header *Machine ID, FC ID* *Lane ID* *Read1 or Read2*

```
@HWI-D00368:32:H8R31ADXX:2:1101:2034:2140 1:N:0:CAGATC
TTTGNCGAGAACTGGAATTGAACCAATATTTAAGTCTTACAAGGAATTCGTTTTAAC
+
@@@F#2ADFDHHHJJJJJGHHIIJIIJJJIJGGJHEIIJIJIIJIIJJJIJJJJIGI
```

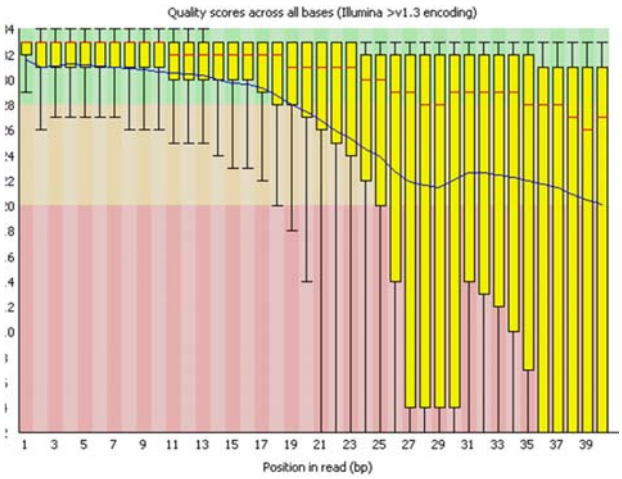
Q-score header

Seq. performance assessment – Base Q

Phred quality scores Q: logarithmically related to error probabilities

P by $Q = [-10 * \log_{10}(P)]$

Phred Score Q	Error probability	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%



FastQC Report

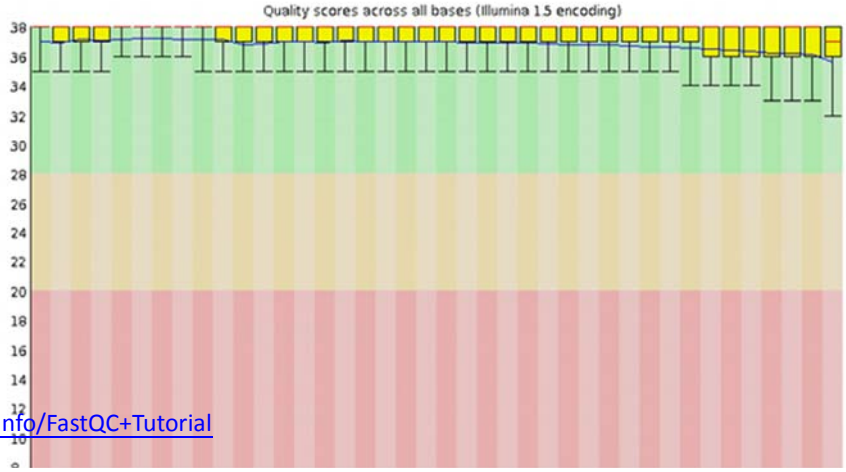
Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Kmer Content

Basic Statistics

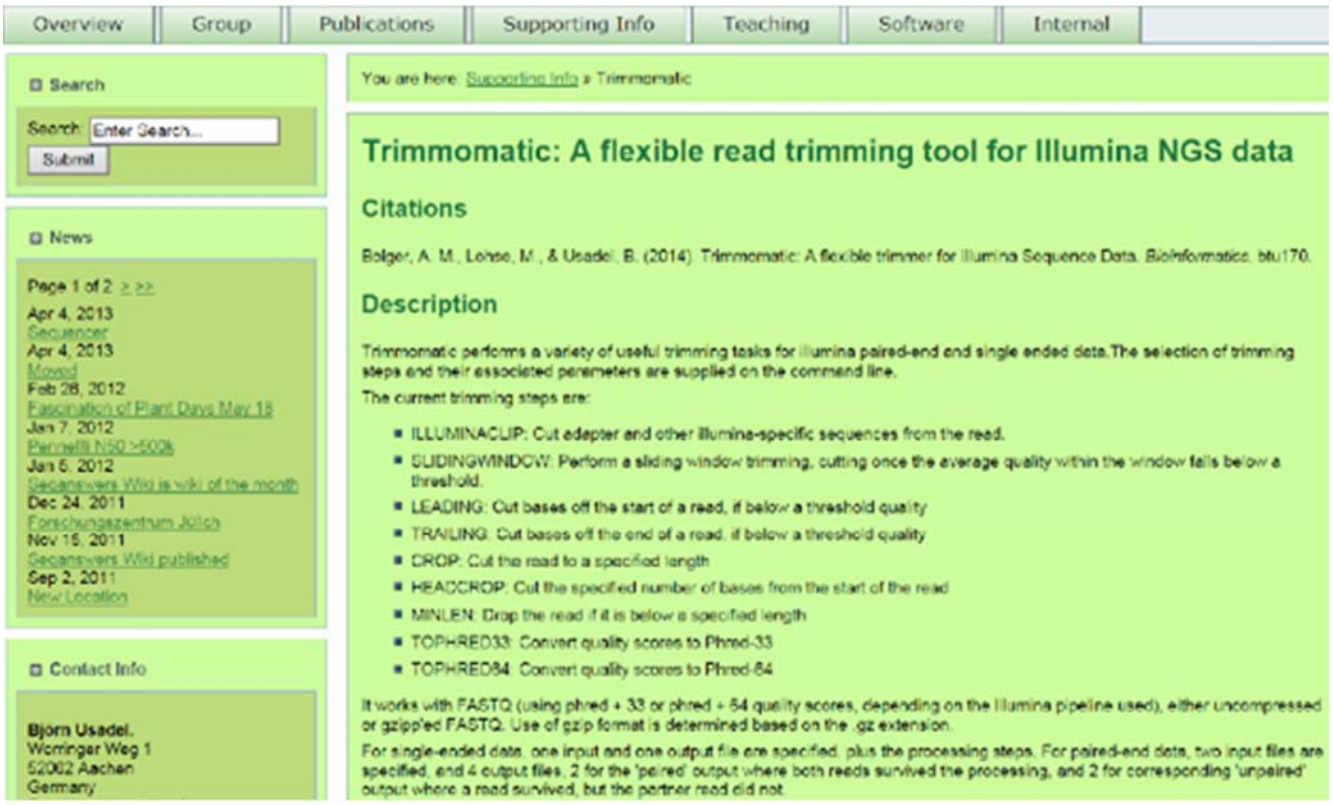
Measure	Value
Filename	good_sequence_short.fastq
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Filtered Sequences	0
Sequence length	40
%GC	45

Per base sequence quality



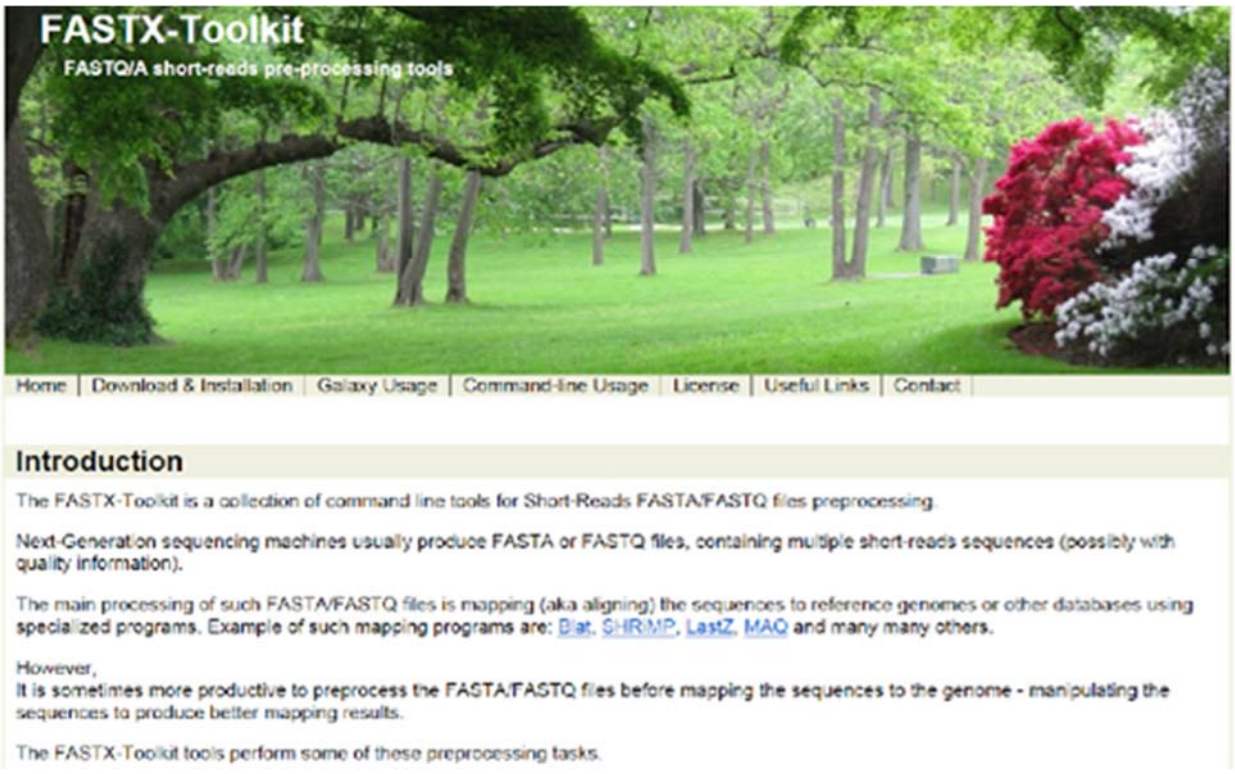
https://wiki.hpcc.msu.edu/display/Bioinfo/FastQC+Tutorial

Data Trimming - Trimmomatic



http://www.usadellab.org/cms/?page=trimmomatic

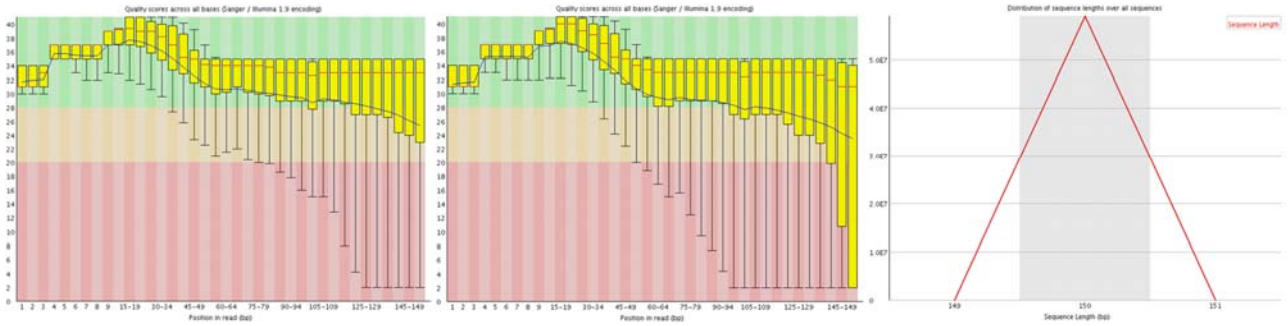
Data pre-processing: eg. FastX-toolkit



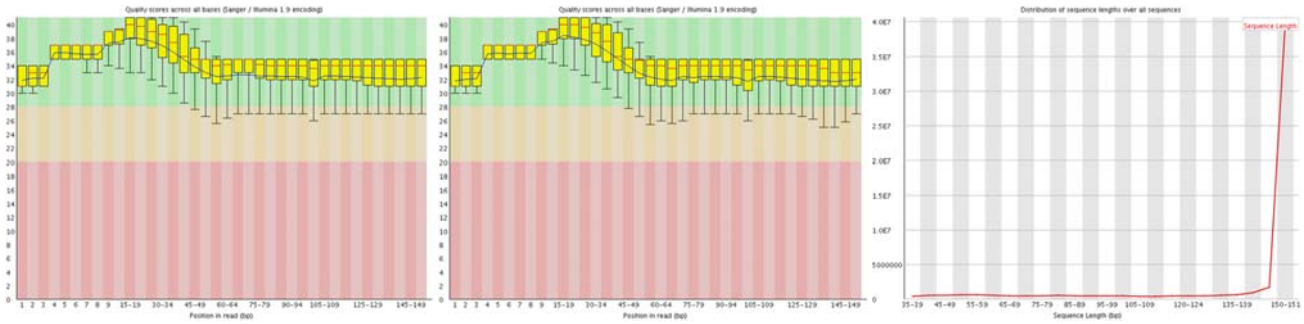
http://hannonlab.cshl.edu/fastx_toolkit/

Adapter Trimming Result of HiSeq genomic PE Reads

PE 2*151nt (raw; R1 & R2)

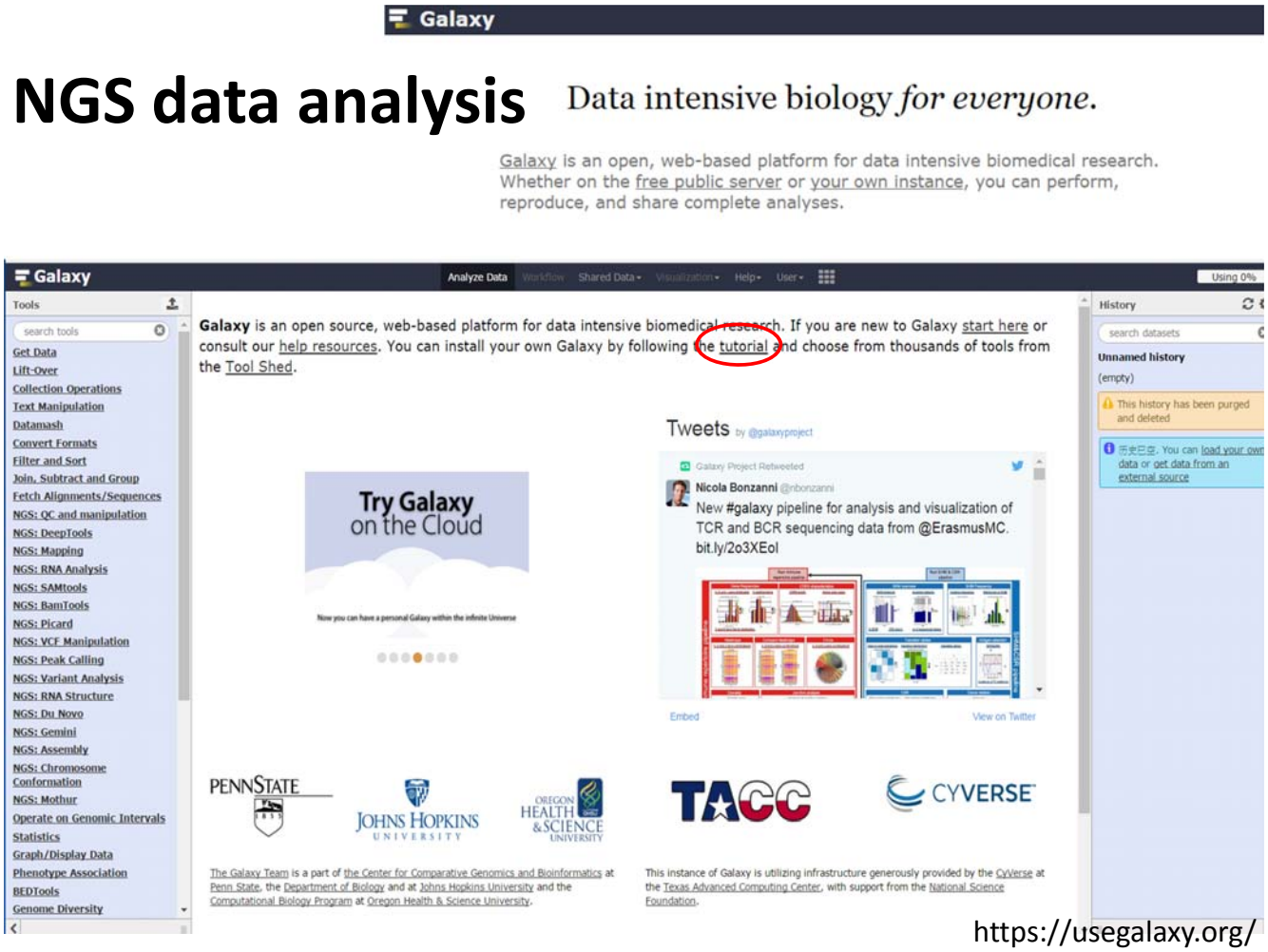


Reads after adapter trimming by Trimmomatic



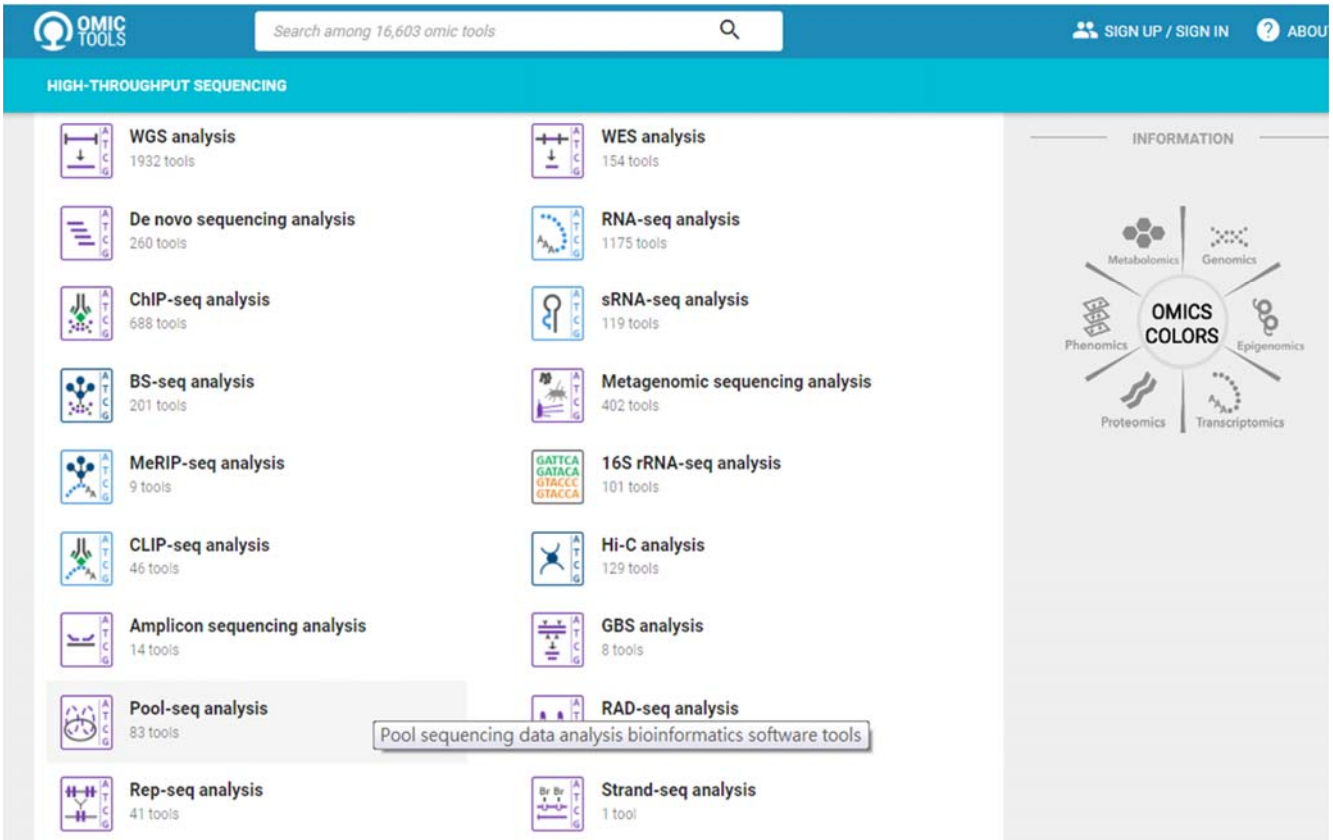
Public Resources

NGS data analysis

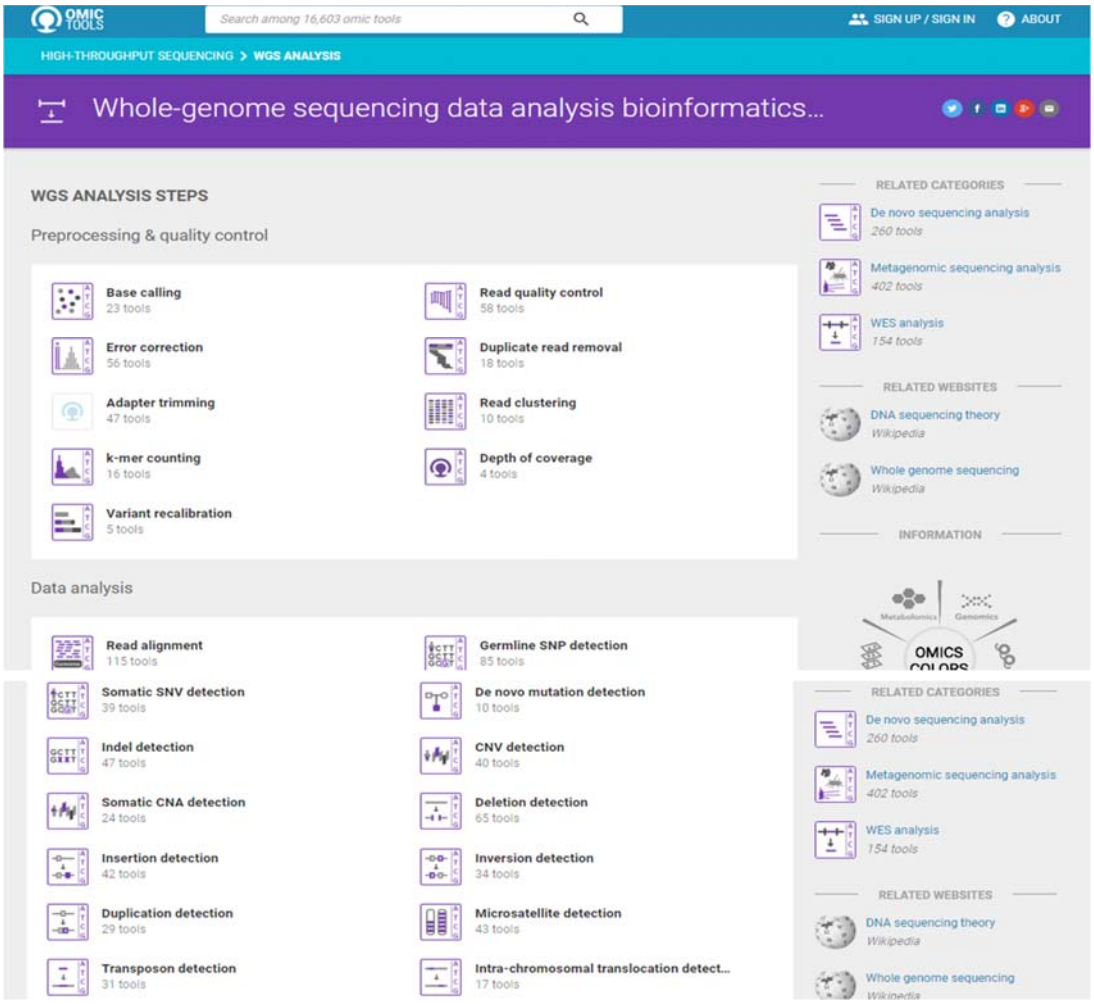


https://usegalaxy.org/

OMICS Tools: High Throughput Sequencing

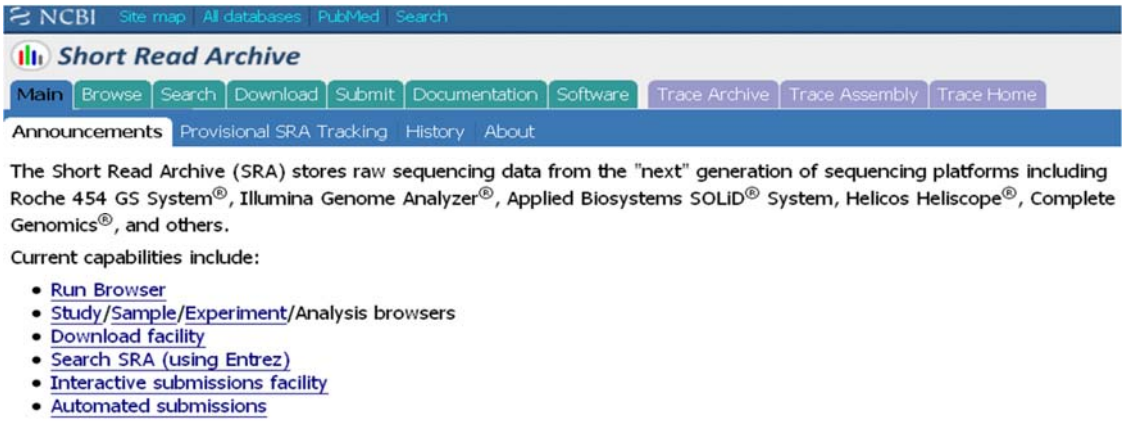


http://omictools.com/sga-s1518.html

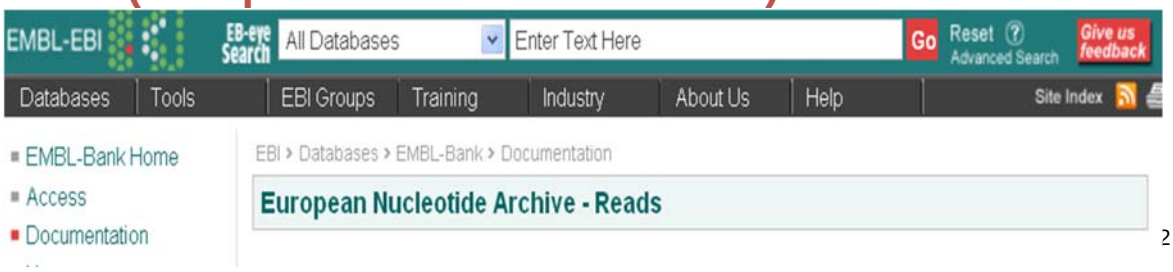


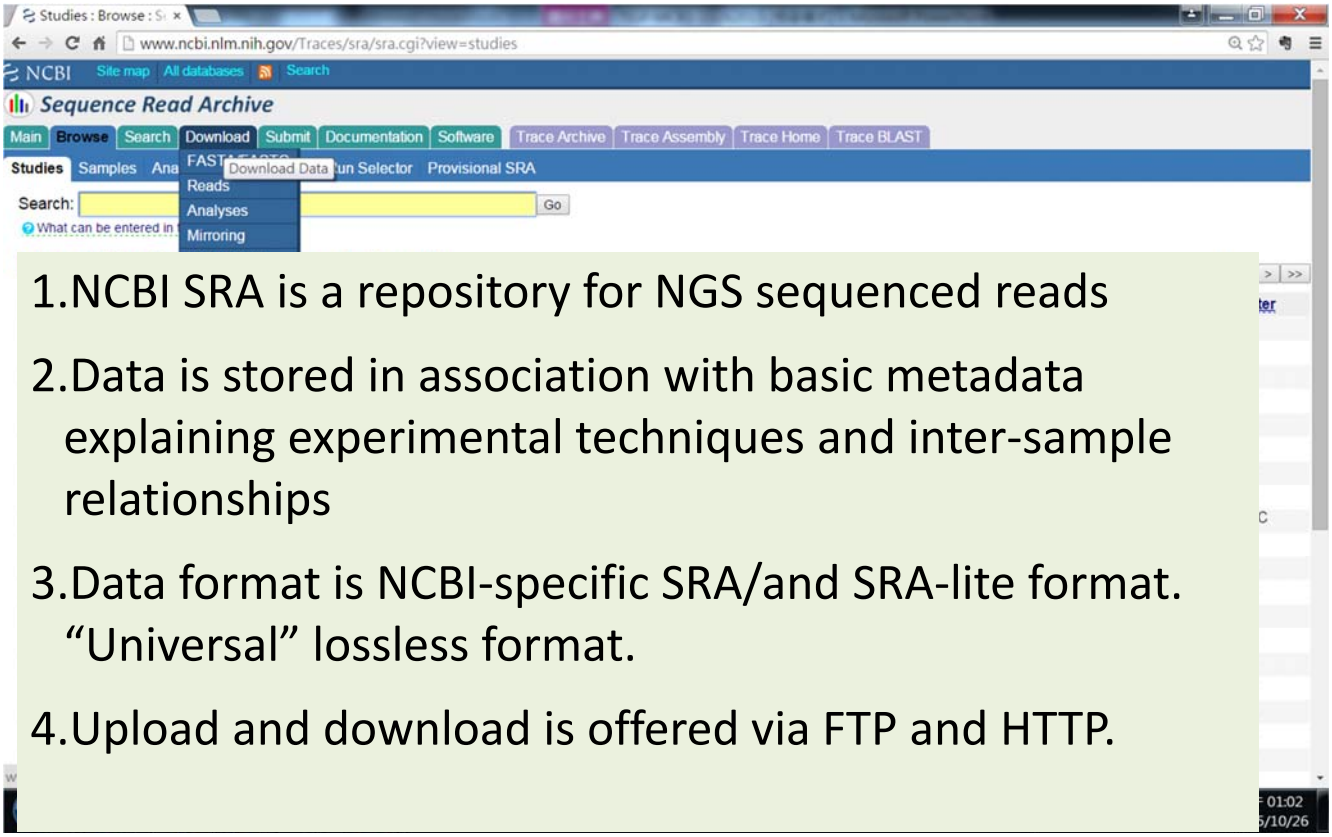
Public NGS Databases

NIH Short Read Archive

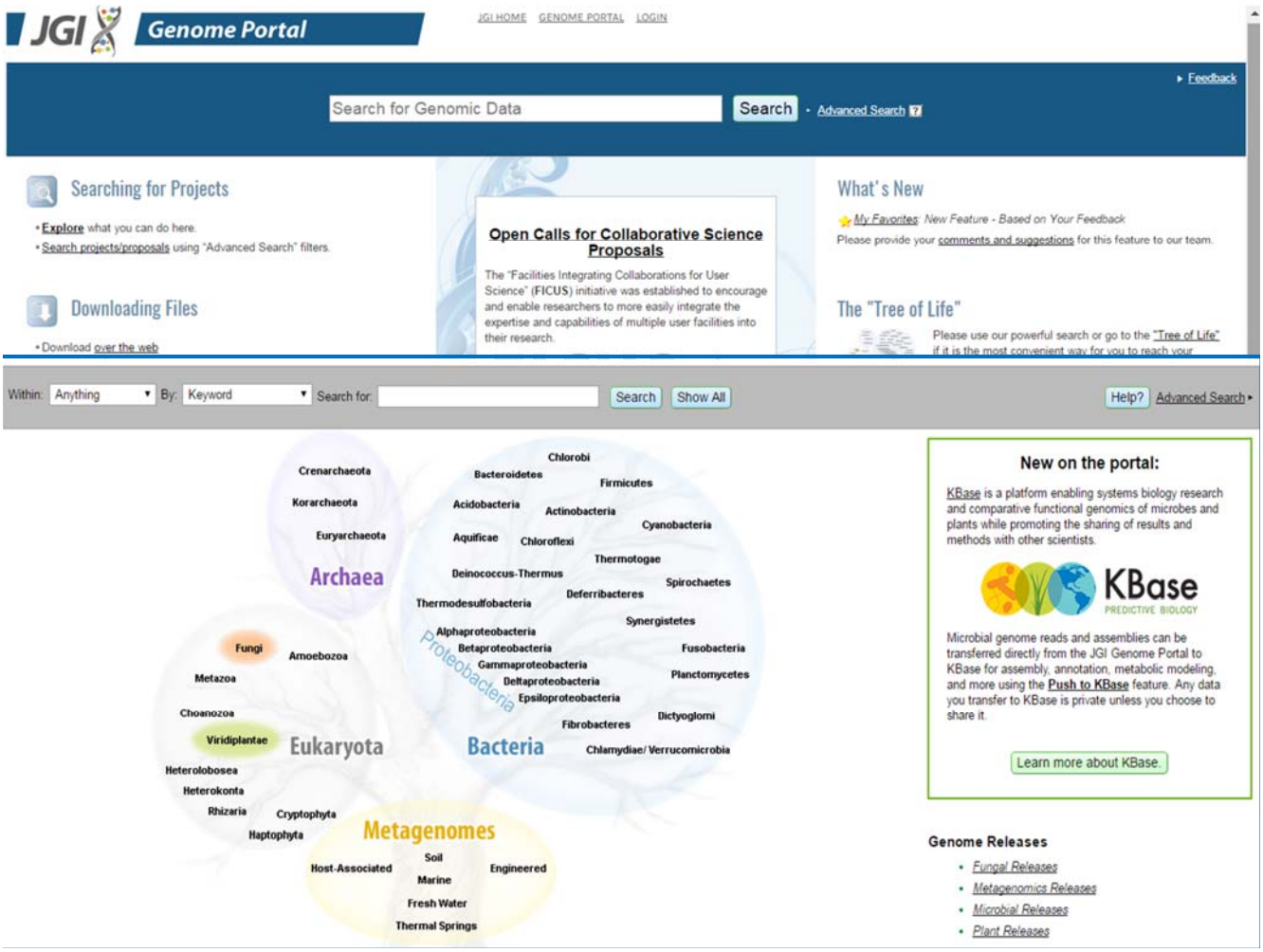


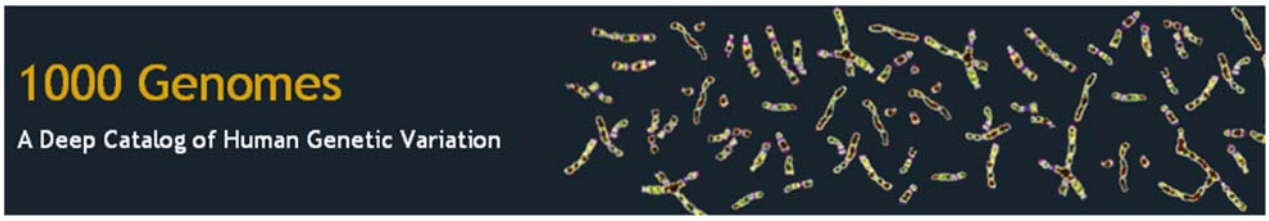
ERA (European Short Read Archive)





1. NCBI SRA is a repository for NGS sequenced reads
2. Data is stored in association with basic metadata explaining experimental techniques and inter-sample relationships
3. Data format is NCBI-specific SRA/and SRA-lite format. “Universal” lossless format.
4. Upload and download is offered via FTP and HTTP.





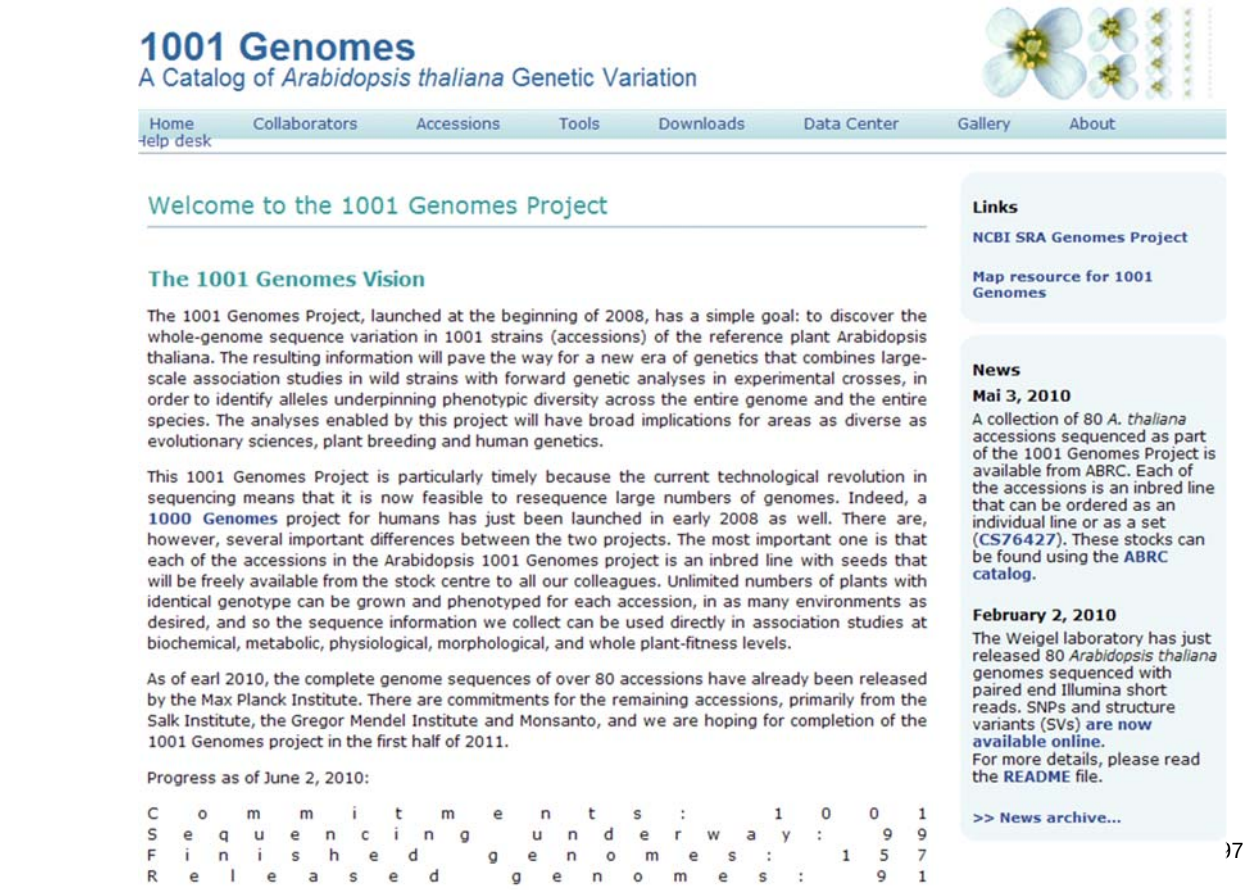
International consortium announces the 1000 Genomes Project.

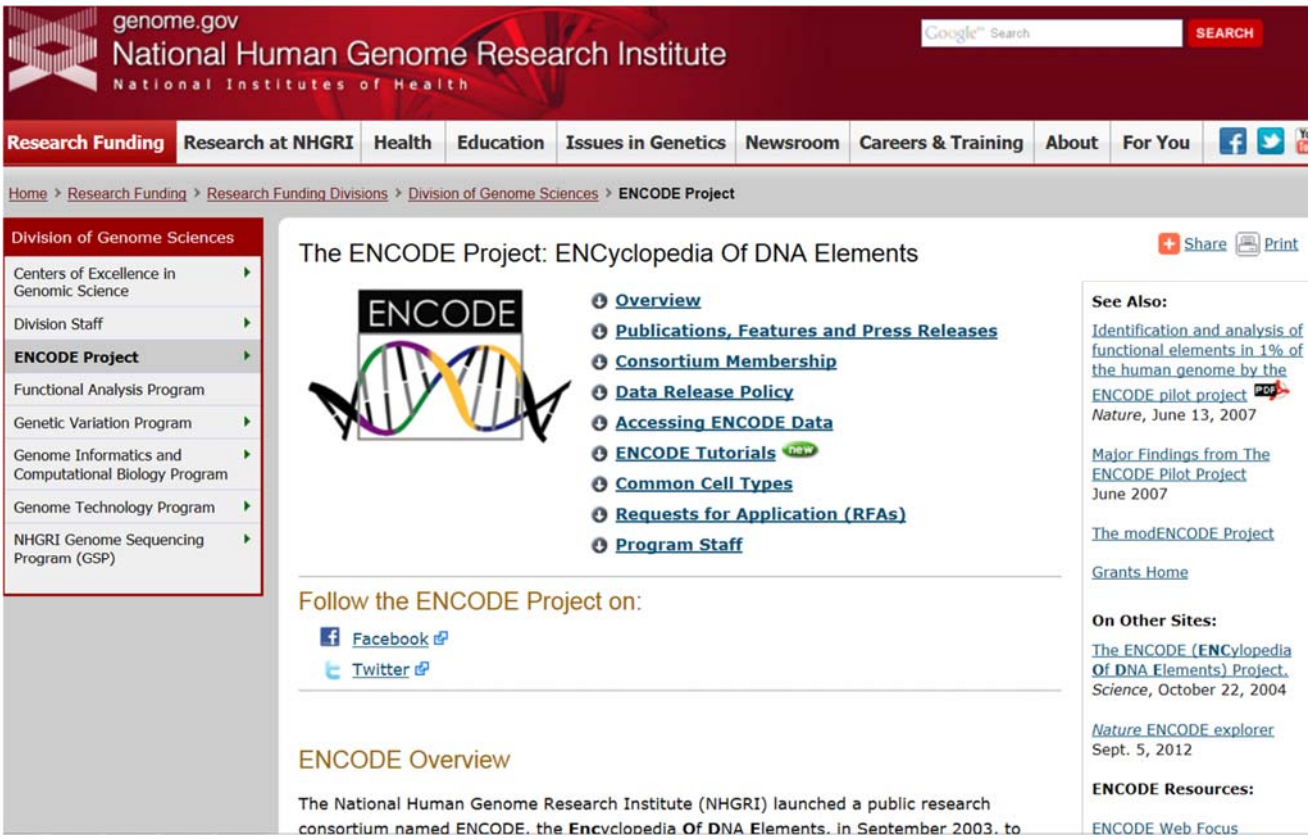
Any two people have 99% identical DNA.

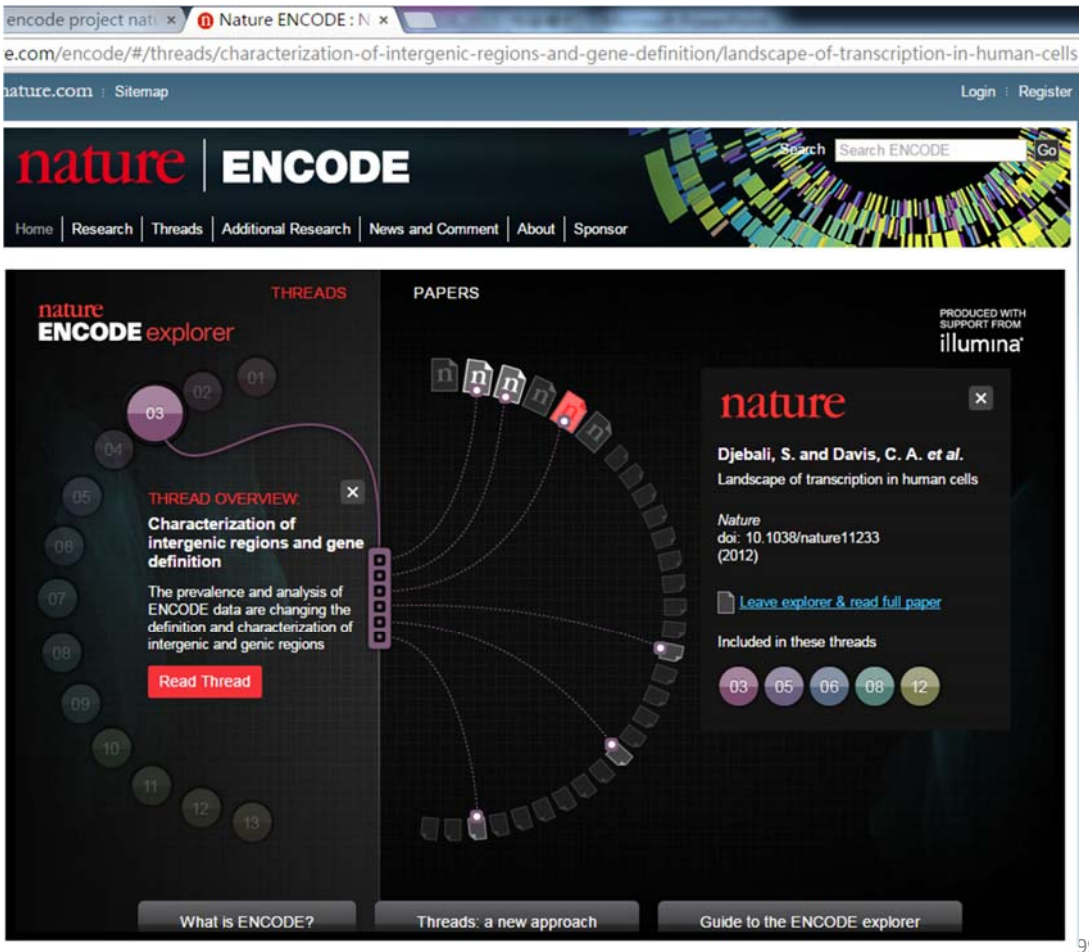
http://www.wired.com/wiredscience/2008/01/the-genome-in-h/

Public release date: 22-Jan-2008

Genome sequences are important







http://www.nature.com/encode/

NGS Reviews

Next-generation DNA sequencing

Jay Shendure[1] & Hanlee Ji[2]

***Nature Biotechnology* 26**, 1135 - 11

ADVANCES IN WOUND CARE

Mary Ann Liebert, Inc. publishers

Journals | Search | Alerts

Adv Wound Care (New Rochelle). 2015 Jan 1; 4(1): 50–58. PMCID: PMC4281878

doi: 10.1089/wound.2014.0542

Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research

Brendan P. Hodkinson and Elizabeth A. Grice*

Author information ▶ Article notes ▶ Copyright and License information ▶

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Sequencing technologies – the next generation

*Michael L. Metzker**‡*

***Nature Review Genetics* 11,** 31-46 (2010)

nature REVIEWS GENETICS

nature.com ▸ journal home ▸ archive ▸ issue ▸ review ▸ full text

NATURE REVIEWS GENETICS | REVIEW

ARTICLE SERIES: Applications of next-generation sequencing

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin, John D. McPherson & W. Richard McCombie

Affiliations | Corresponding author

Nature Reviews Genetics 17, 333–351 (2016) | doi:10.1038/nrg.2016.49

Published online 17 May 2016

NIH Public Access

Author Manuscript

J Genet Genomics. Author manuscript; available in PMC 2011 April 13.

Published in final edited form as:

J Genet Genomics. 2011 March 20; 38(3): 95–109. doi:10.1016/j.jgg.2011.02.003.

The impact of next-generation sequencing on genomi

Jun Zhang[a,b,*], **Rod Chiodini**[c], **Ahmed Badr**[a], and **Genfa Zhang**[d]

[a] COE for Neurosciences, Department of Anesthesiology, Texas Tech University Center El Paso, TX 79905, USA

Video Clips

- Sanger Sequencing of DNA [HD Animation]
  - https://www.youtube.com/watch?v=nudG0r9zL2M
- Pyro Sequencing
  - https://www.youtube.com/watch?v=nFfgWGFe0aA
- Illumina Sequencing Technology
  - https://www.youtube.com/watch?v=womKfikWlxM
- Ion Torrent™ next-gen sequencing technology
  - https://www.youtube.com/watch?v=WYBzbxIfuKs
- Single Molecule Real Time Sequencing - Pacific Biosciences
  - https://www.youtube.com/watch?v=v8p4ph2MAvI
- Oxford Nanopore Technologies
  - https://www.youtube.com/watch?v=3UHw22hBpAk
- Next-Generation Sequencing Technologies - Elaine Mardis (2014)
  - https://www.youtube.com/watch?v=6Is3W7JkFp8

- PCR (Polymerase Chain Reaction)
  - https://www.youtube.com/watch?v=iQsu3Kz9NYo
- Polymerase Chain Reaction [HD Animation]
  - https://www.youtube.com/watch?v=0HCWmD7Mv8U

Applications of Next-Generation Sequencing