Sequencing Technologies and Applications

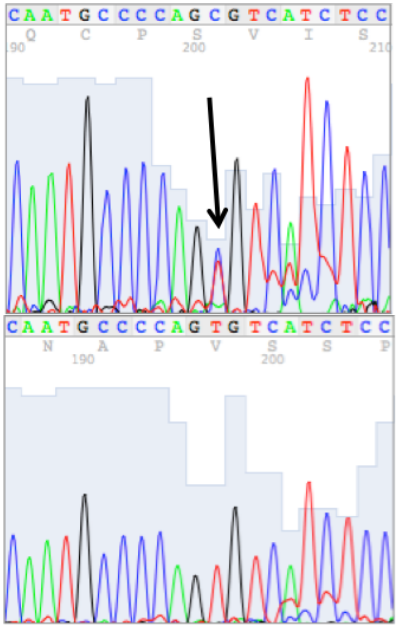
David K. Crossman, Ph.D.
Department of Genetics
Heflin Center for Genomic Sciences

Genotyping Technologies

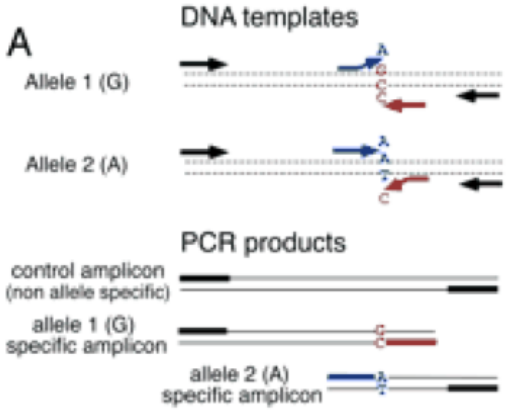
- The technology used depends on the number of variants and number of samples.
- Single to a few SNPs in a small population (≤960)
  - TaqMan
  - Pyrosequencing
  - Allele Specific PCR
- Intermediate # of SNPs, Intermediate population size
  - No good option here (BeadExpress or GoldenGate from Illumina but these are being discontinued)
- Large # of SNPs, large population
  - Infinium from Illumina (up to 5M SNPs per slide)

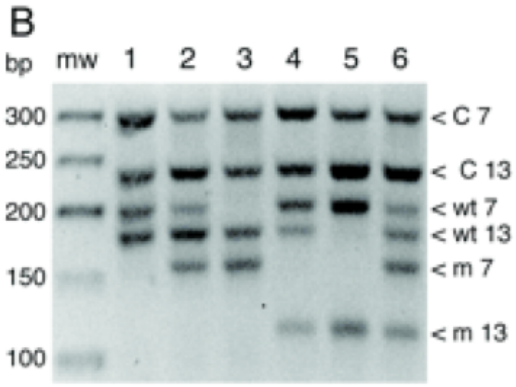
Single SNP Analysis

Direct Sequencing



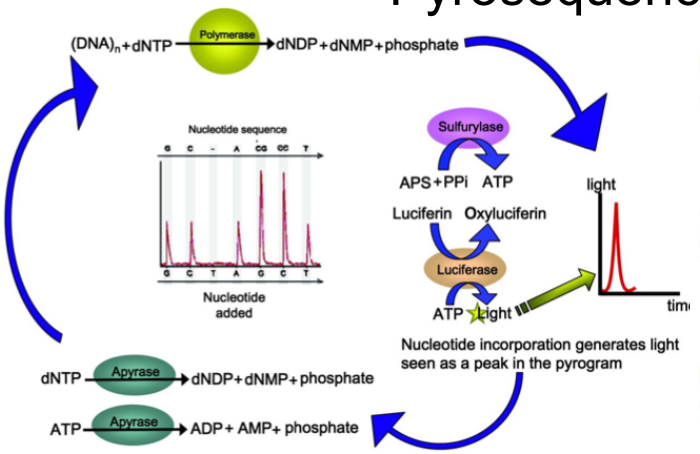
Allele Specific PCR

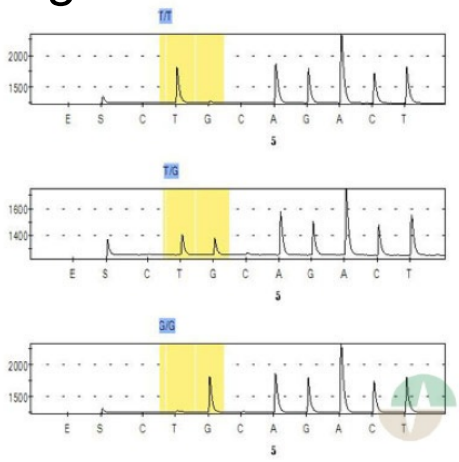




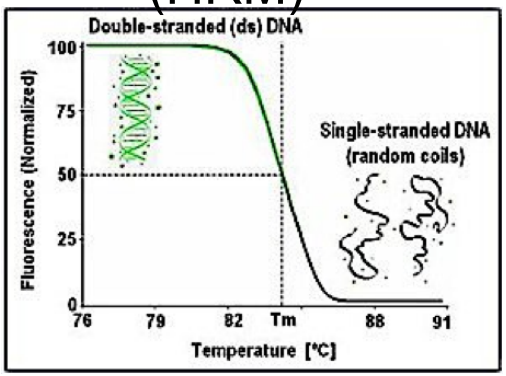
Piccioli, P. et al. Clinical Chem 2006 52:4:739

Pyrosequencing





High Resolution Melting (HRM)



https://dna.utah.edu/Hi-Res/TOP_Hi-Res%20Melting.html

Cycle Sequencing with Dye Termination

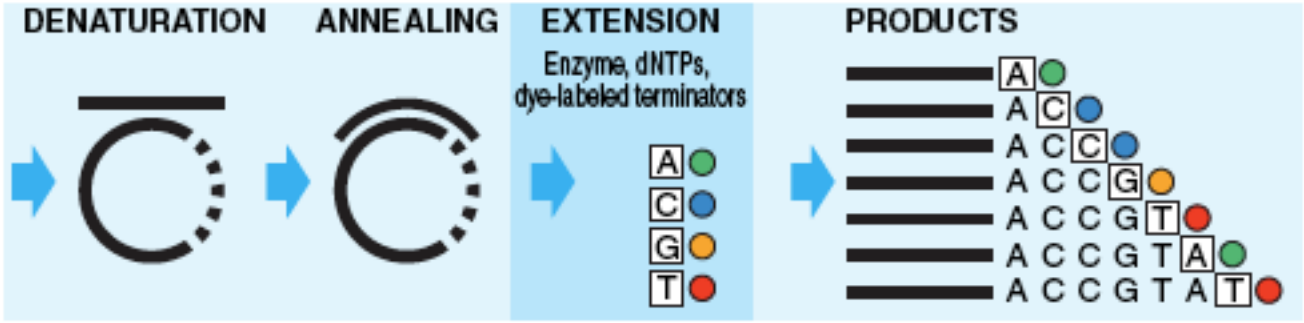
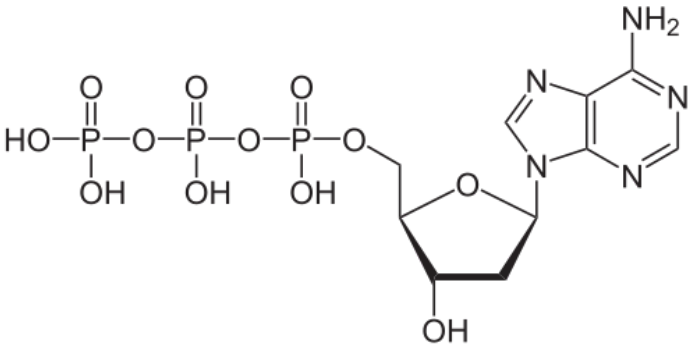
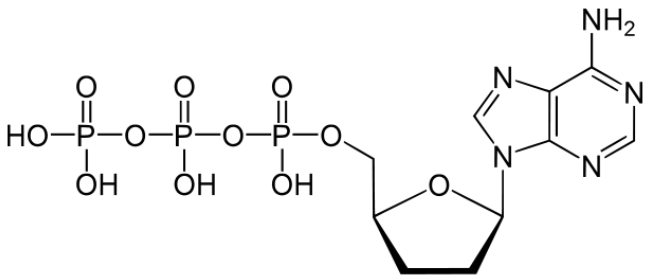
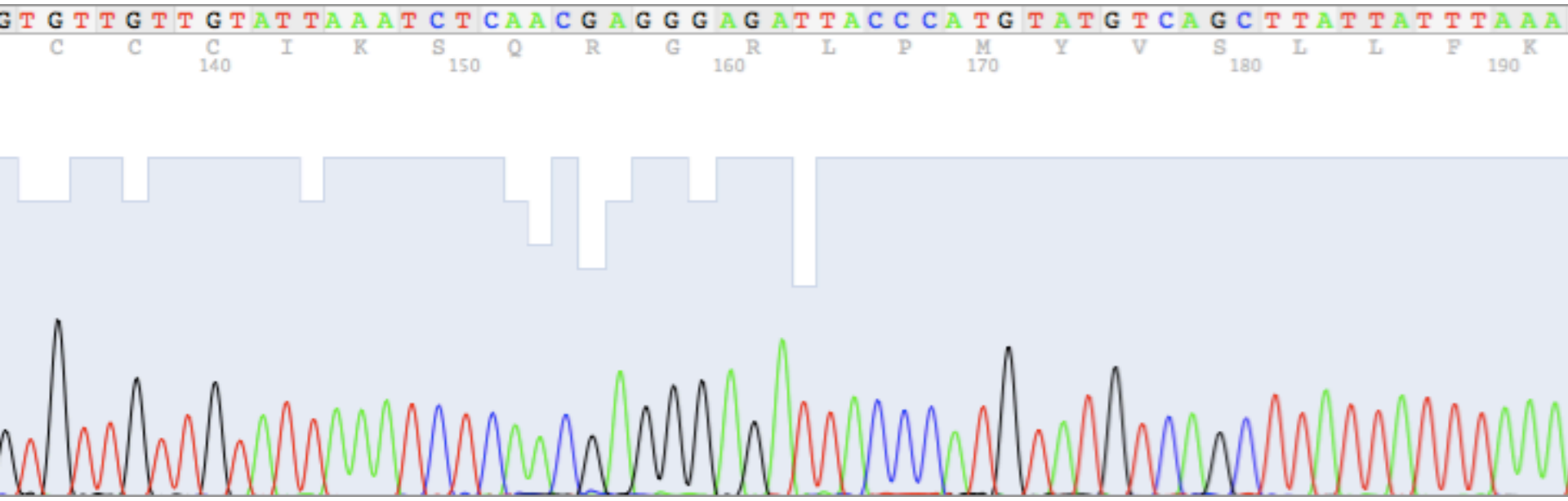


Figure 5 Diagram of dye terminator cycle sequencing






GTGTTGTTGTATTAAATCTCAACGAGGGAGATTACCCATGTATGTCAGCTTATTATTTAAA
C C C I K S Q R G R L P M Y V S L L F K
140 150 160 170 180 190

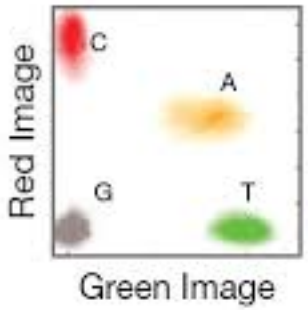
The Next Generation

Illumina Platforms

MiSeq
Single flowell
15-25M reads
25-600bp increments
Higher cost per base sequenced
Single or Paired End

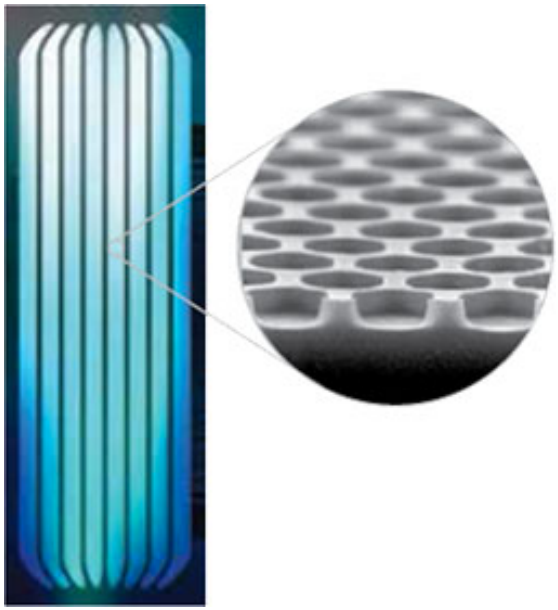
HiSeq2500
Two flowcells
~600billion bases sequenced
50bp-100bp increments
Lower cost per base sequenced
Single reads and Paired end reads
Rapid Runs 26-48hrs

Illumina NextSeq 500

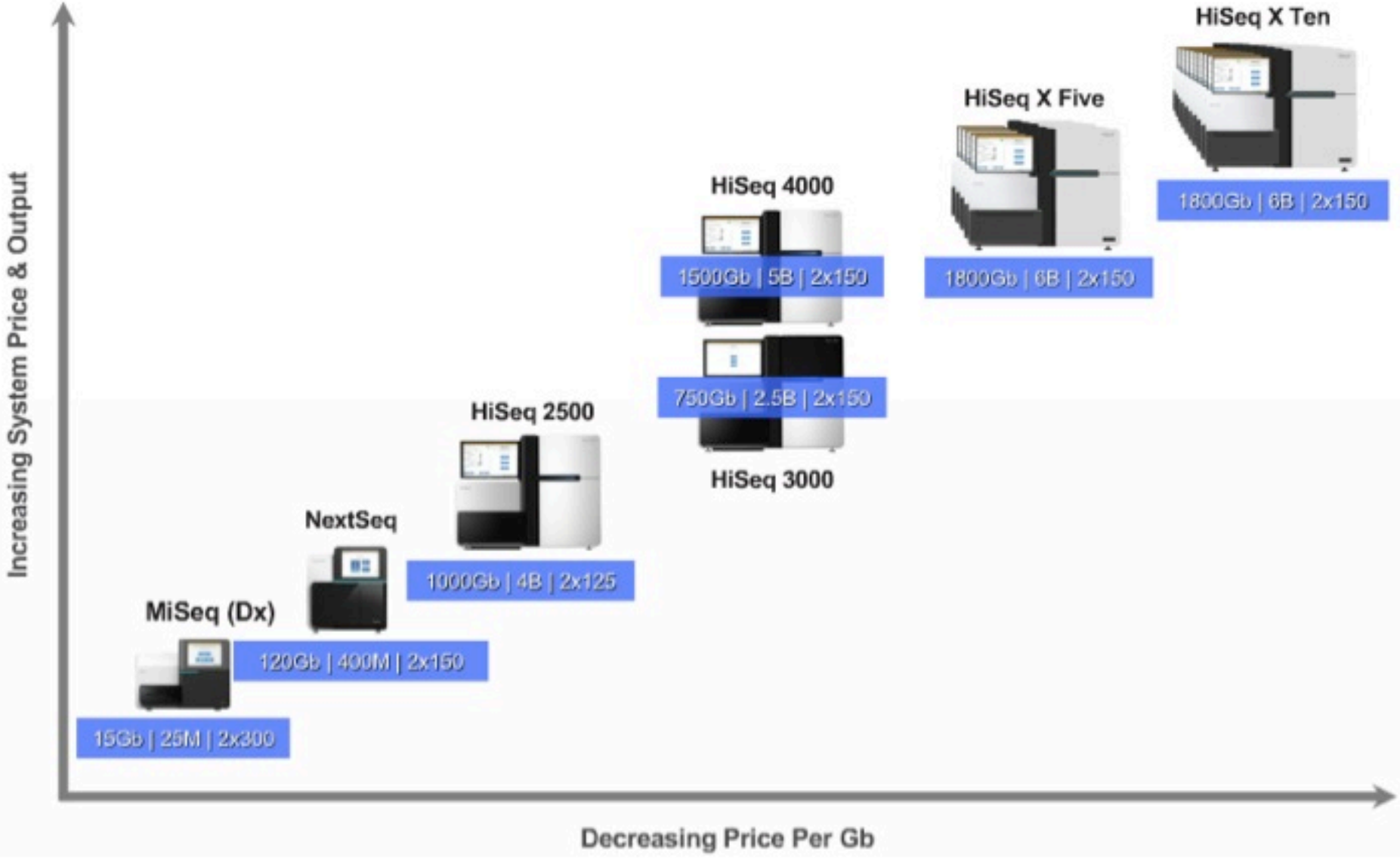


Accelerated detection of all four DNA bases is performed on the NextSeq 500 System using only two images to capture red and green filter wavelength bands. A bases will be present in both images (yellow cluster), C bases in red only, T bases in green only, and G bases in neither.

Illumina HiSeq X Ten



Sequencing Power For Every Scale.

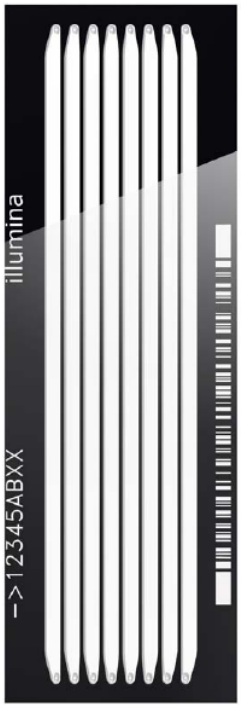


Useful Next-Gen Terms

- Cluster
  - Individual island of DNA molecules representing a single, unique template
- Clusters Passing filter
  - Number of clusters able to be distinguished by the software as individuals
- Fastq
  - DNA Sequence file that is able to be read by downstream analysis applications
- Q-Score
  - A quality score based on the Phred score from Sanger Sequenicng which is the probability a base is incorrect at a give position. Example: Q30 means there is a 1:1000 chance the base is incorrect. Or stated another way it means the base call is 99.9% accurate
- Phasing/Prephasing
  - When the DNA sequencing reaction is either a base ahead or a base behind the majority of the other molecules
- Depth of Coverage
  - The average number of times a base is read within the genome
- Reads
  - Actual sequence

Flowcells through time

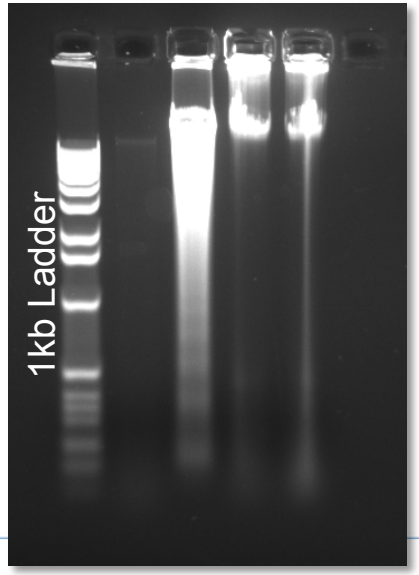
2005 2010 2013

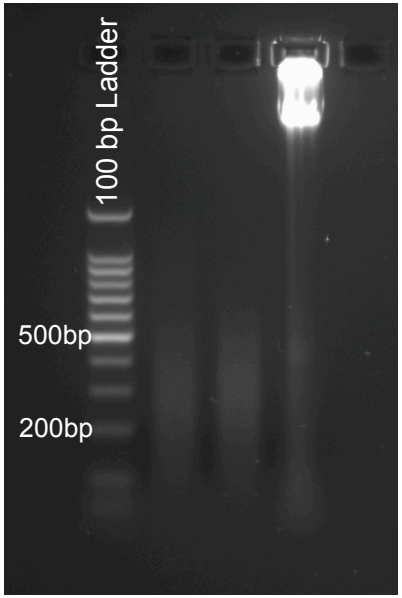


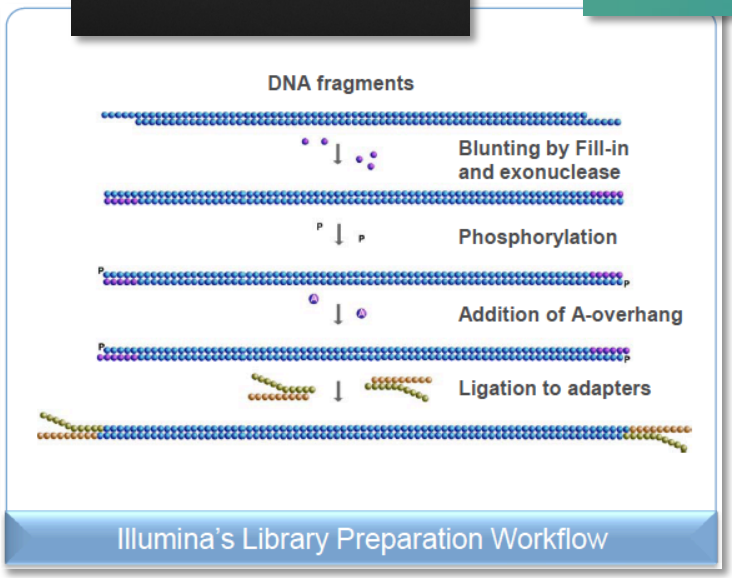
GAIIx HiSeq MiSeq HiSeqX

Not to scale

DNA Library Prep and Flow cell Production

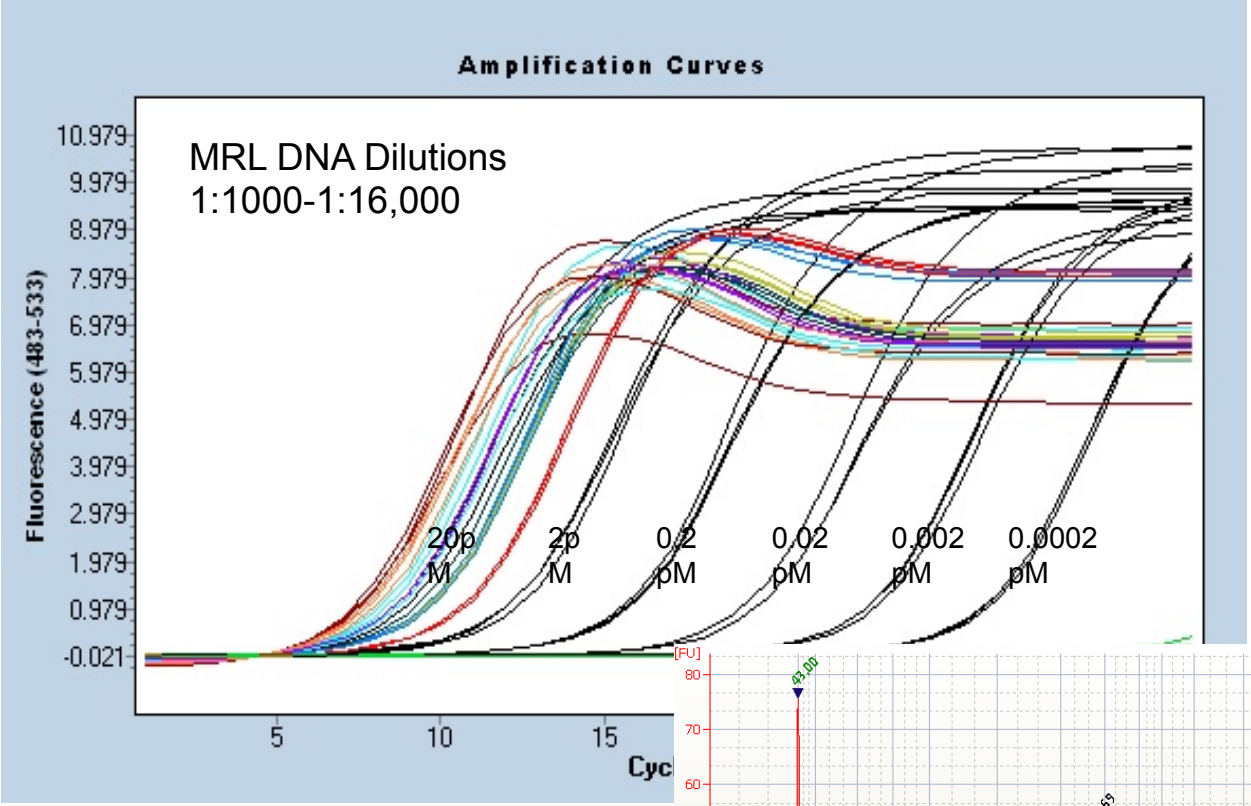


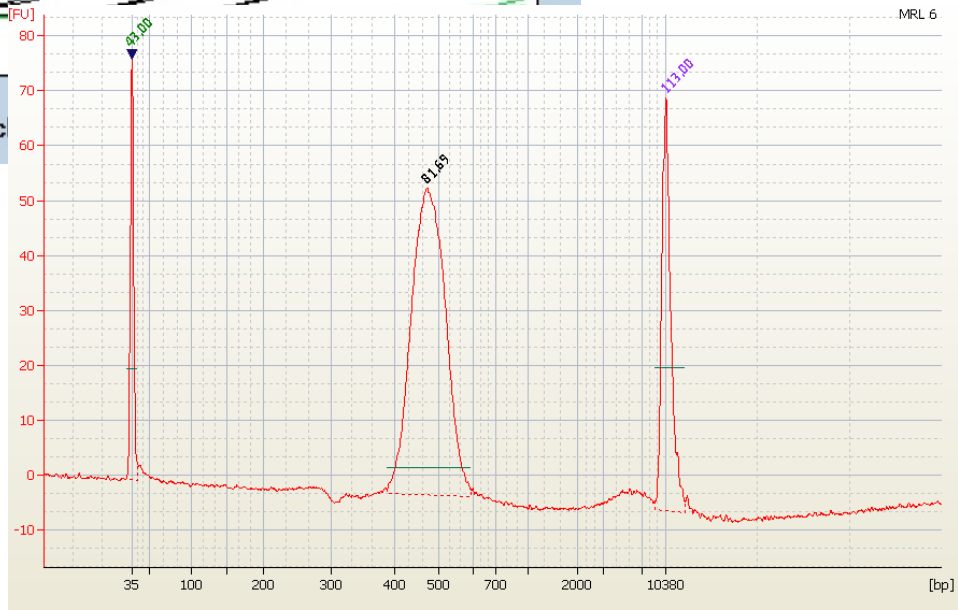




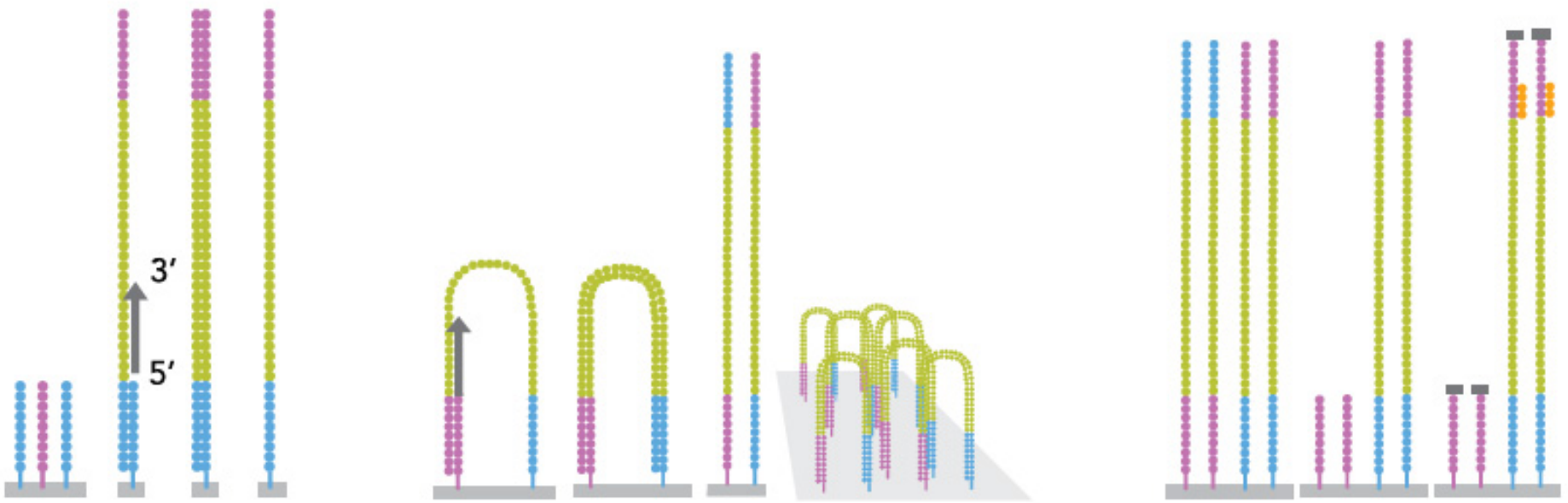
Version 3 HiSeq Flow Cell

Library Assessment and Quantitation

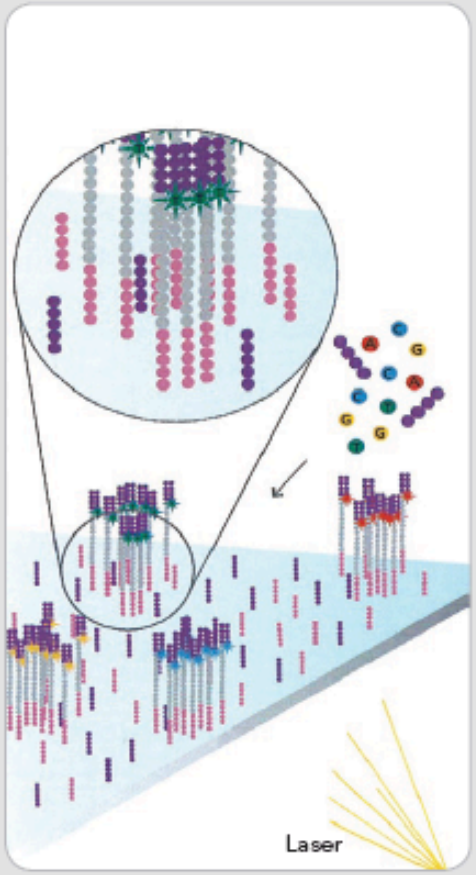




Illumina Cluster Generation

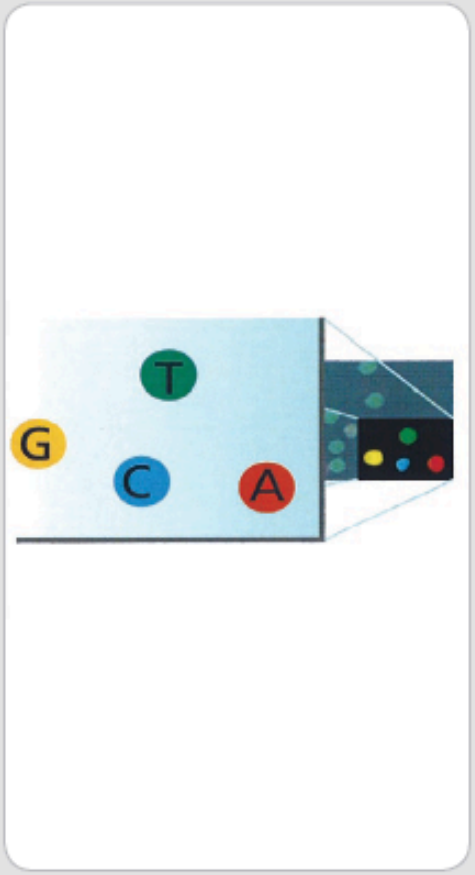


7. DETERMINE FIRST BASE



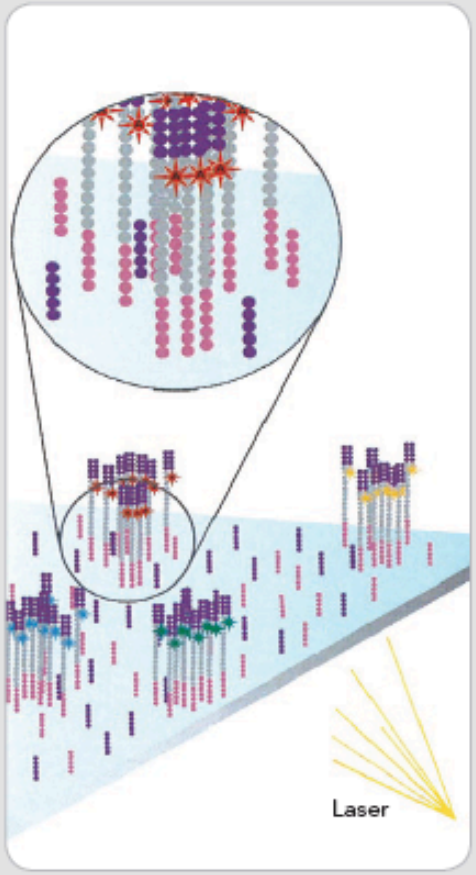
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE



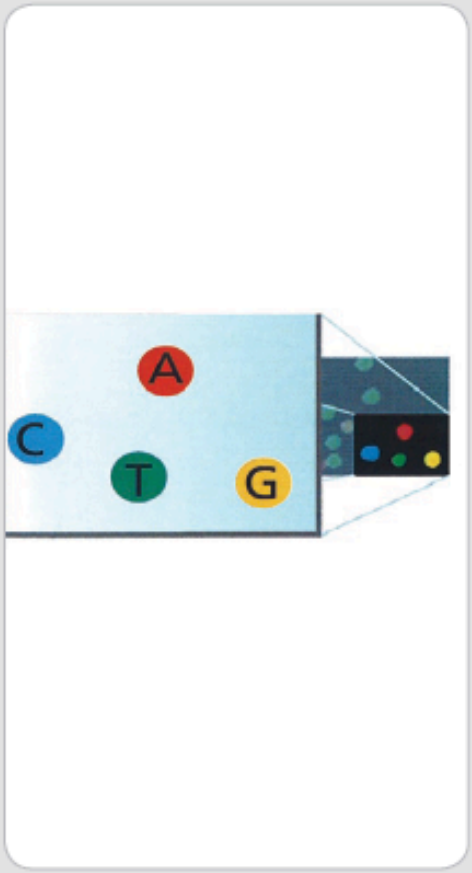
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE



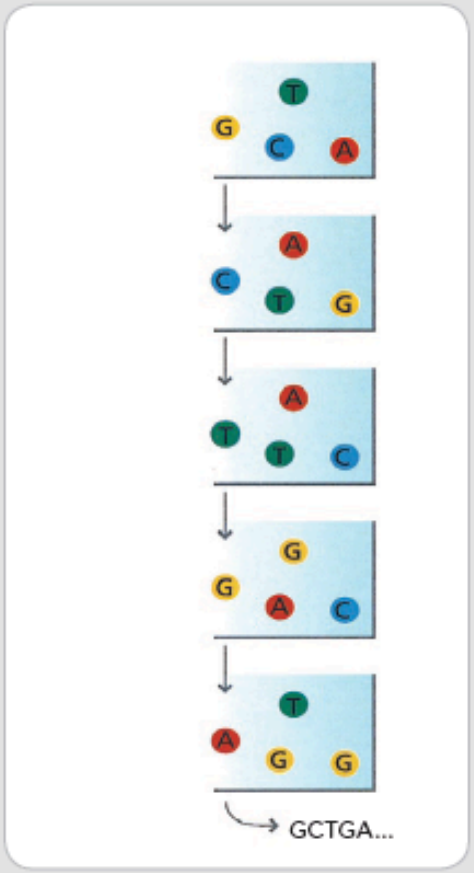
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

10. IMAGE SECOND CHEMISTRY CYCLE



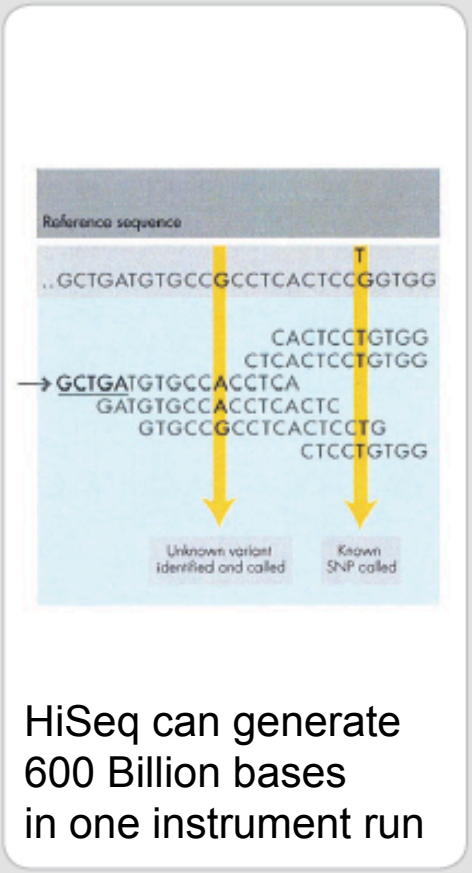
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

11. SEQUENCING OVER MULTIPLE CHEMISTRY CYCLES



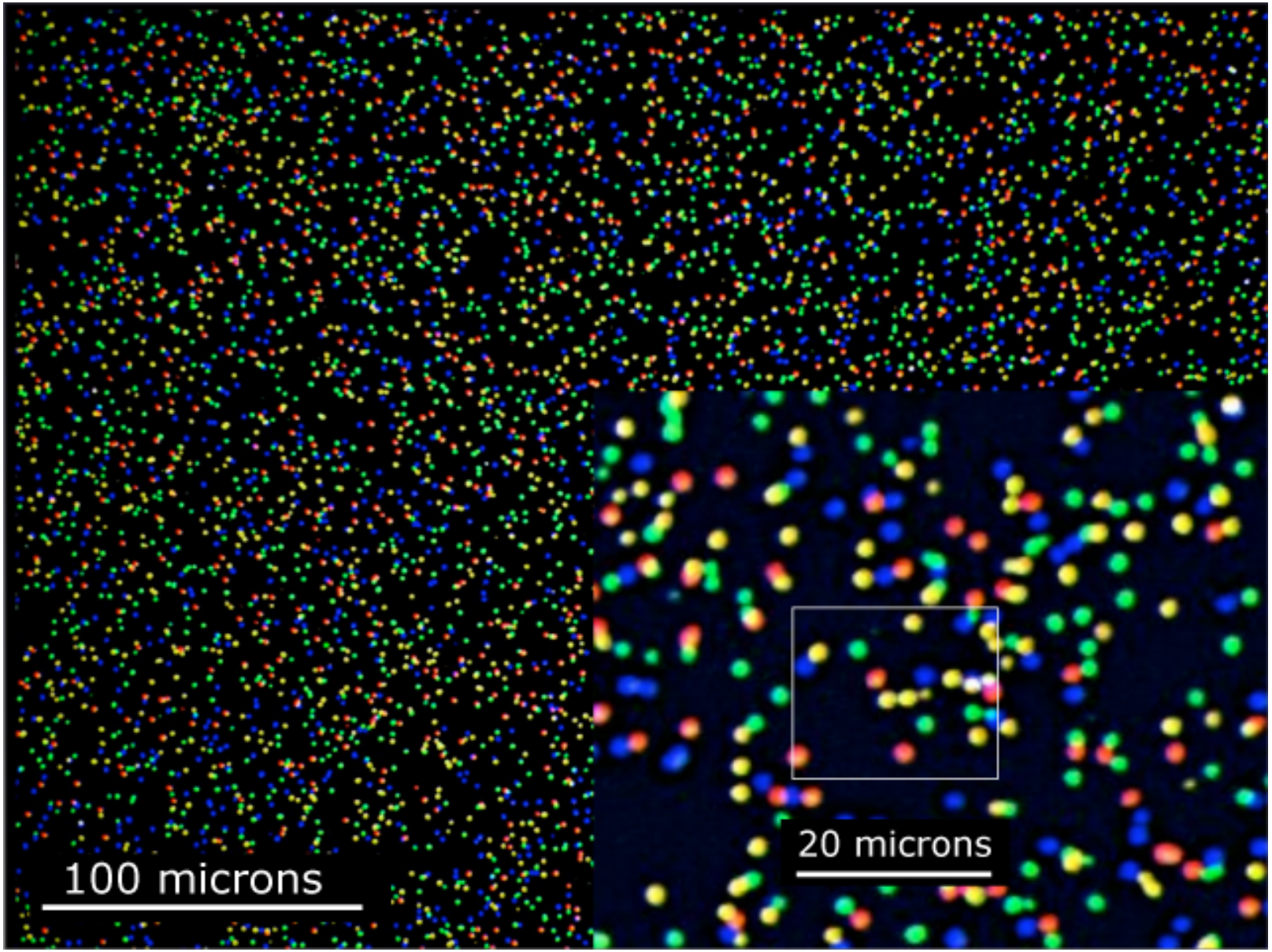
The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

12. ALIGN DATA



HiSeq can generate
600 Billion bases
in one instrument run

The data are aligned and compared to a reference, and sequencing differences are identified.


100 microns
20 microns

Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

Browse | Refresh

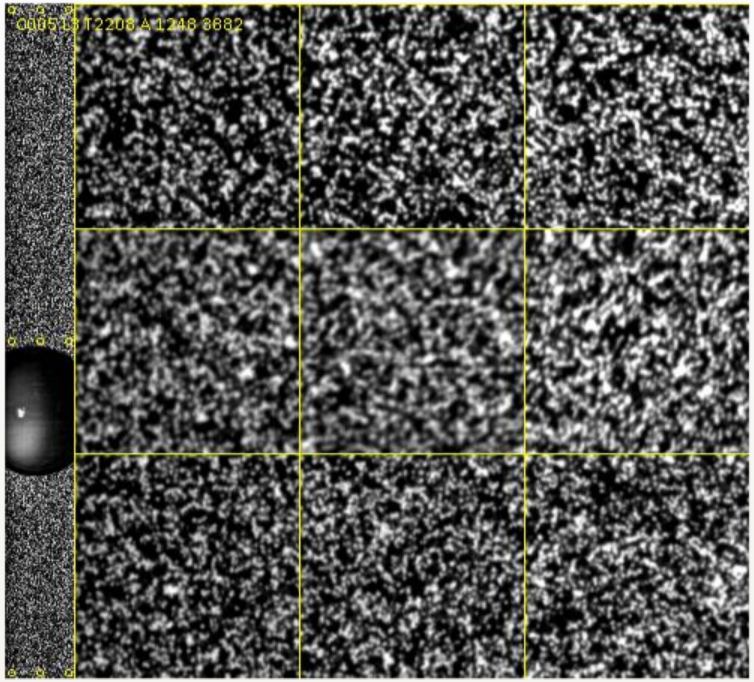
Analysis | Imaging | Summary | Tile Status | Controls

Cycle All | Lane 3 | Surface Bottom | Swath Middle | Section All

Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P9
33148	3	2207	7	100	Bottom	Middle	12/13/201...	1871	2966	931	22
33149	3	2207	7	101	Bottom	Middle	12/13/201...	1840	2929	919	22
33150	3	2207	7	102	Bottom	Middle	12/13/201...	1822	2893	922	22
33151	3	2207	7	103	Bottom	Middle	12/14/201...	1805	2878	906	22
33152	3	2207	7	104	Bottom	Middle	12/14/201...	1802	2876	902	21
33153	3	2207	7	105	Bottom	Middle	12/14/201...	1785	2841	908	21
33154	3	2207	7	106	Bottom	Middle	12/14/201...	1756	2836	874	21
33155	3	2207	7	107	Bottom	Middle	12/14/201...	1749	2813	872	21
33156	3	2207	7	108	Bottom	Middle	12/14/201...	2498	3963	938	25
33913	3	2208	8	1	Bottom	Middle	12/08/201...	2928	4976	1861	35
33914	3	2208	8	2	Bottom	Middle	12/08/201...	3176	4792	1636	41
33915	3	2208	8	3	Bottom	Middle	12/08/201...	3163	4773	1679	38
33916	3	2208	8	4	Bottom	Middle	12/08/201...	3259	4788	1690	34
33917	3	2208	8	5	Bottom	Middle	12/08/201...	2732	4112	1533	28
33918	3	2208	8	6	Bottom	Middle	12/09/201...	3126	4605	1475	33
33919	3	2208	8	7	Bottom	Middle	12/09/201...	2712	4312	1404	38

Rows=41472 | Disp=864 | Sel=1 | Filter

A C G T



start | UAB - The Univer... | runs on 'Isilon Ser... | Illumina Sequenci... | Microsoft PowerP... | Document1 - Micr... | Search Desktop | 3:53 PM

Illumina Sequencing Analysis Viewer 1.7.25 - 111208_SN372_0101_AD0JRMACXX

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

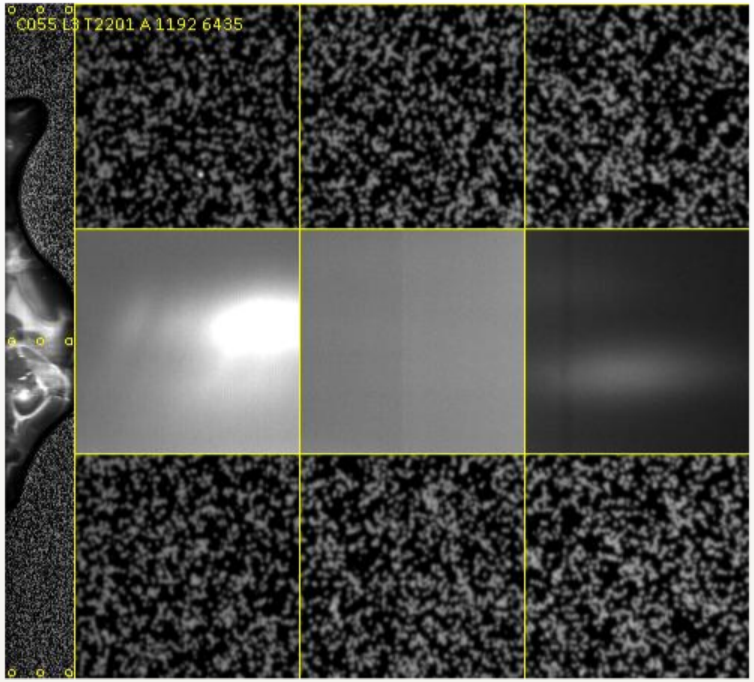
Browse | Refresh

Analysis | Imaging | Summary | Tile Status | Controls

Cycle All | Lane 3 | Surface Bottom | Swath Middle | Section All

A C G T

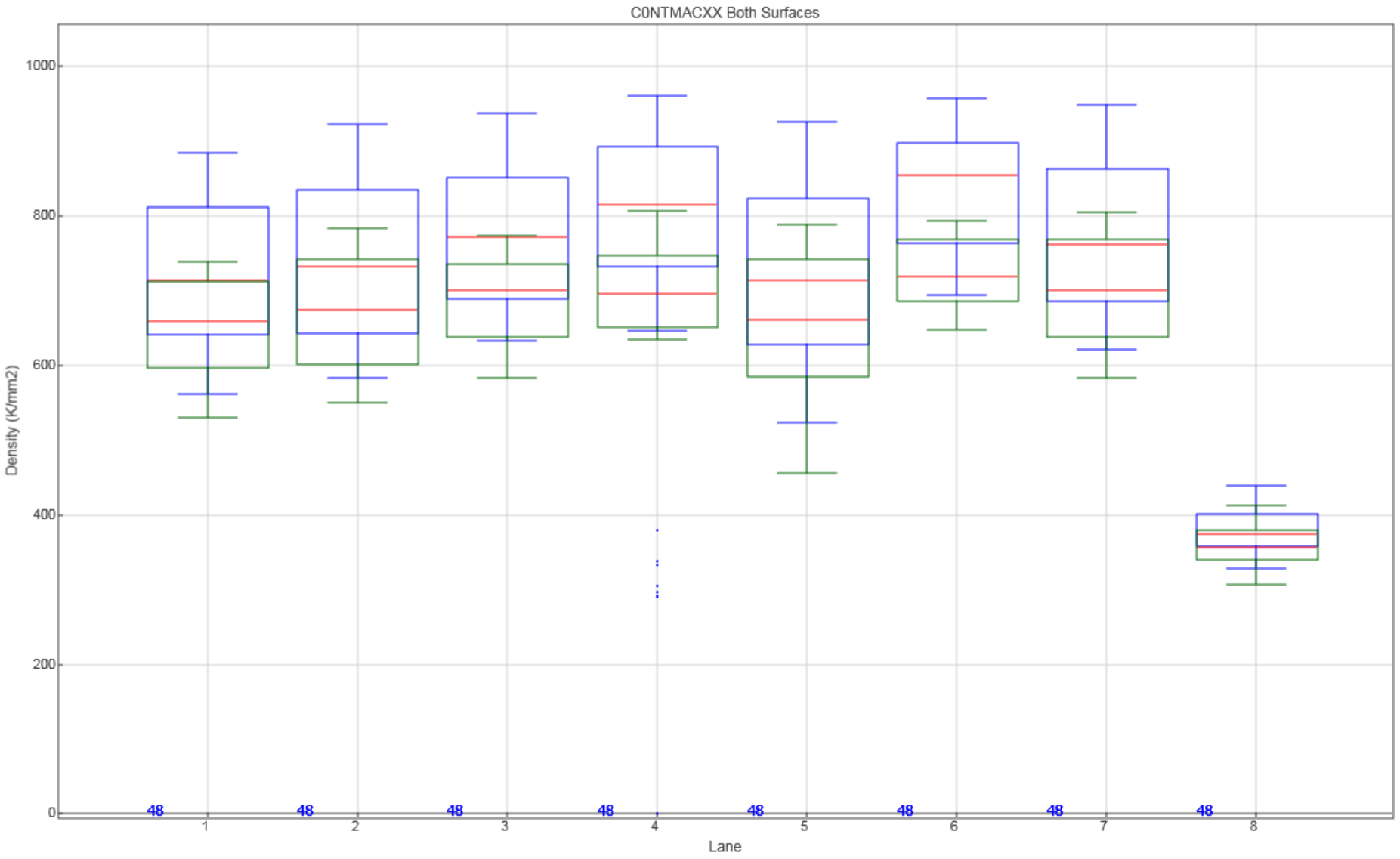
Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P9
27906	3	2201	1	42	Bottom	Middle	12/10/201...	2369	3824	1148	27
27907	3	2201	1	43	Bottom	Middle	12/10/201...	2361	3831	1136	27
27908	3	2201	1	44	Bottom	Middle	12/10/201...	2301	3759	1106	26
27909	3	2201	1	45	Bottom	Middle	12/10/201...	156	218	146	18
27910	3	2201	1	46	Bottom	Middle	12/10/201...	2298	3726	1124	26
27911	3	2201	1	47	Bottom	Middle	12/10/201...	2263	3673	1100	26
27912	3	2201	1	48	Bottom	Middle	12/10/201...	2211	3644	1065	25
27913	3	2201	1	49	Bottom	Middle	12/10/201...	2228	3657	1068	25
27914	3	2201	1	50	Bottom	Middle	12/10/201...	150	206	0	0
27915	3	2201	1	51	Bottom	Middle	12/10/201...	3761	5324	2061	43
27916	3	2201	1	52	Bottom	Middle	12/10/201...	3608	5397	1707	11
27917	3	2201	1	53	Bottom	Middle	12/10/201...	0	0	176	32
27918	3	2201	1	54	Bottom	Middle	12/10/201...	961	4845	167	44
27919	3	2201	1	55	Bottom	Middle	12/10/201...	2538	4054	1430	26
27920	3	2201	1	56	Bottom	Middle	12/10/201...	2813	4531	1552	34
27921	3	2201	1	57	Bottom	Middle	12/10/201...	300	566	0	0



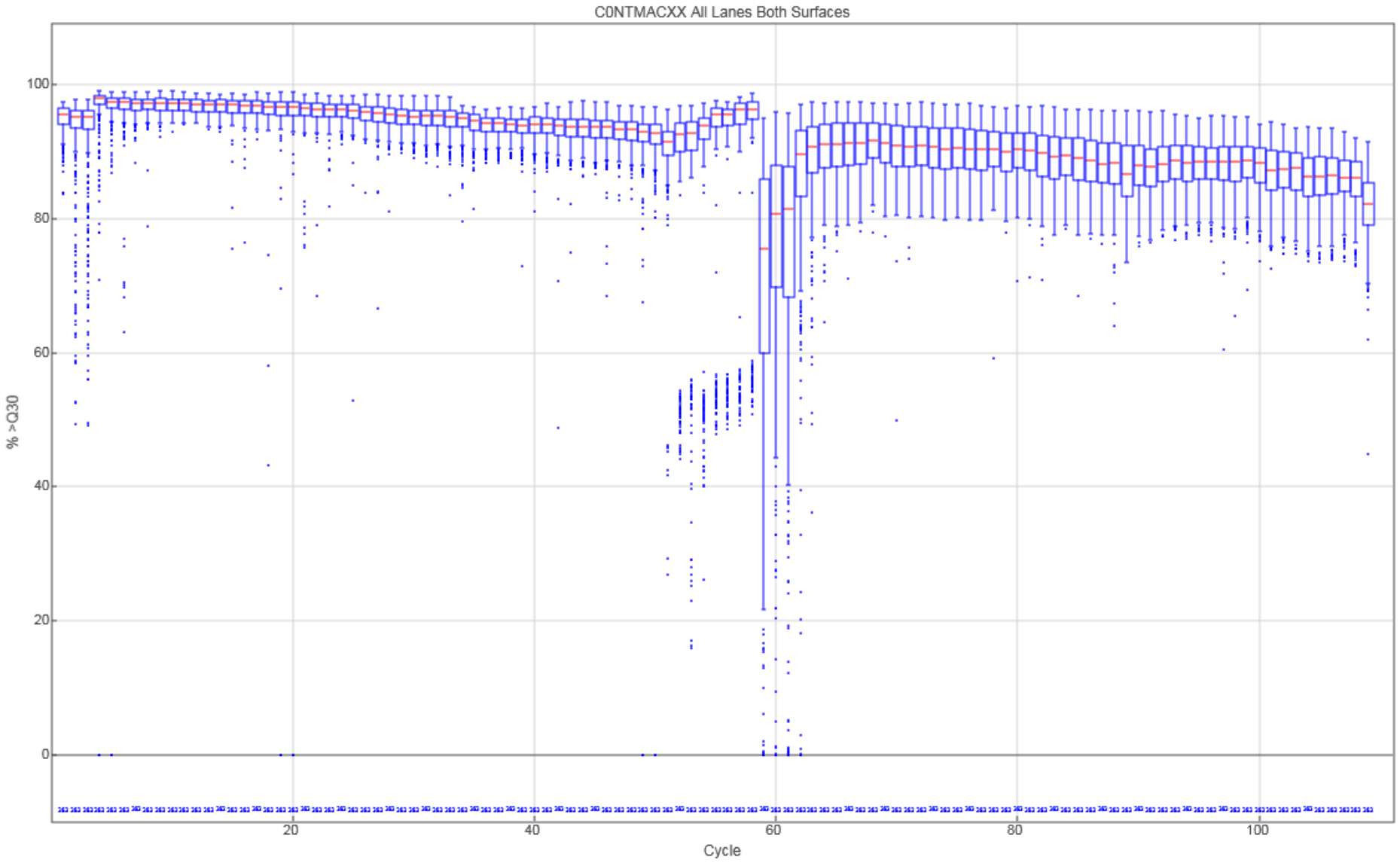
Rows=41472 | Disp=864 | Sel=1 | Filter

start | UAB - The Univer... | runs on 'Isilon Ser... | Illumina Sequenci... | Microsoft PowerP... | Document1 - Micr... | Search Desktop | 3:50 PM

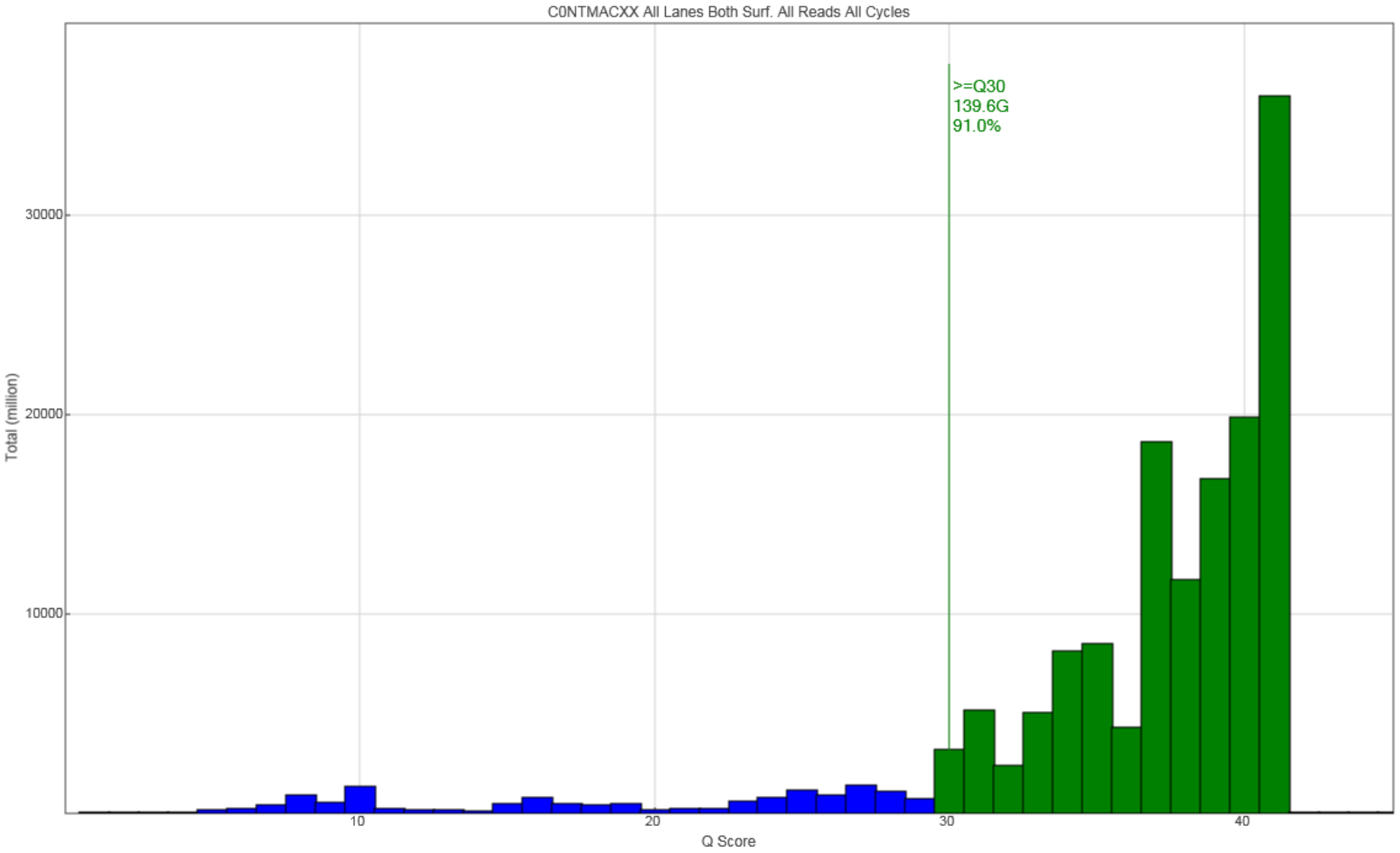
Cluster Density



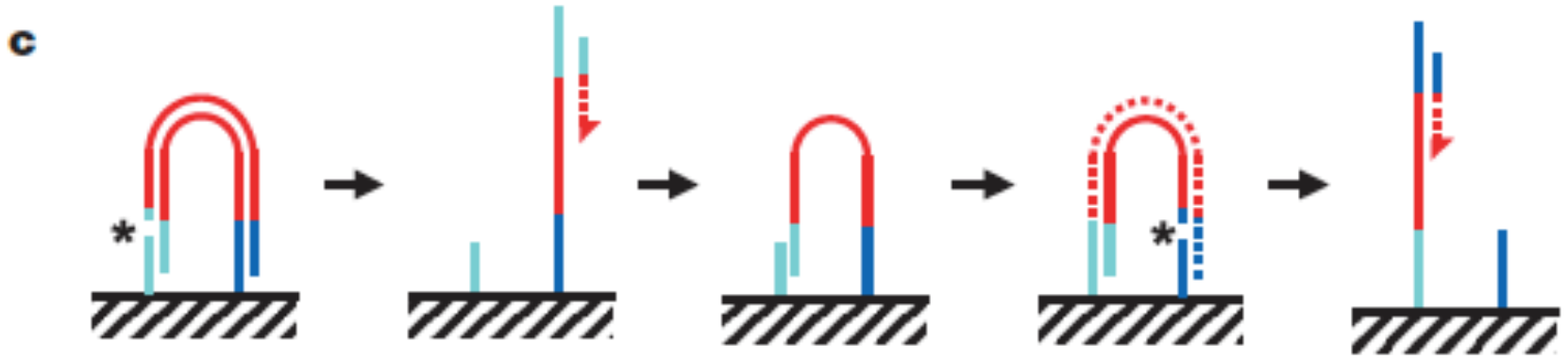
Percent Q30 Scores per cycle for all lanes and both surfaces



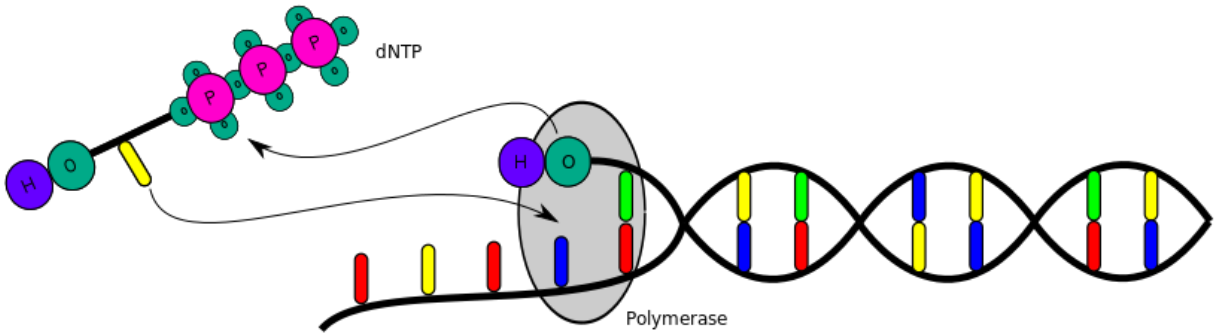
Q Scores


C0NTMACXX All Lanes Both Surf. All Reads All Cycles
>=Q30
139.6G
91.0%
30000
20000
10000
Total (million)
10
20
30
40
Q Score

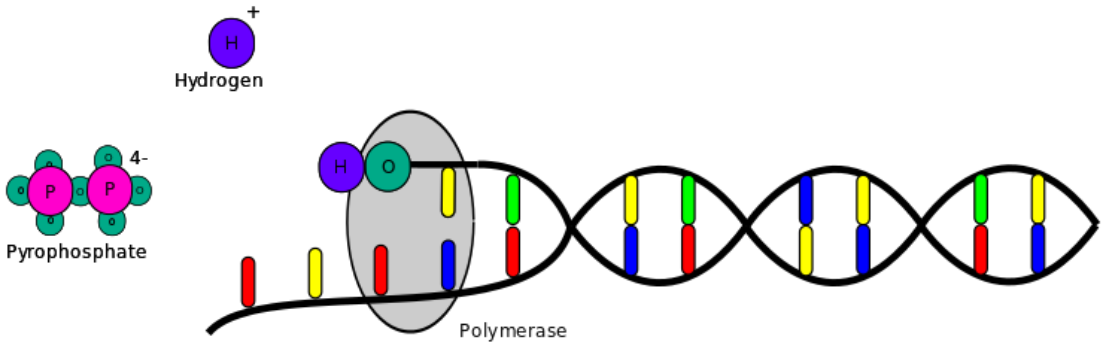
Single End vs. Paired End Sequencing



Ion Semiconductor Sequencing (aka ion Torrent or Proton)



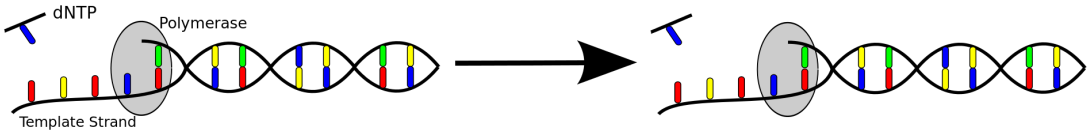
Polymerase integrates a nucleotide.



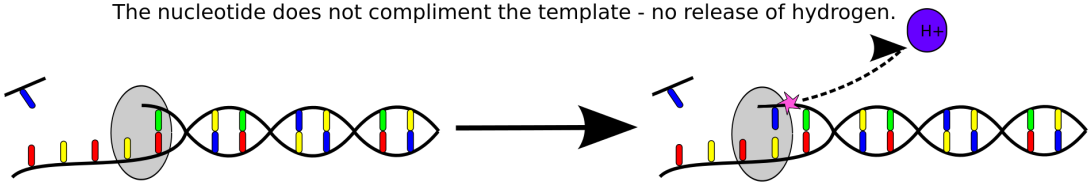
Hydrogen and pyrophosphate are released.

Source:http://en.wikipedia.org/wiki/Ion_semiconductor_sequencing

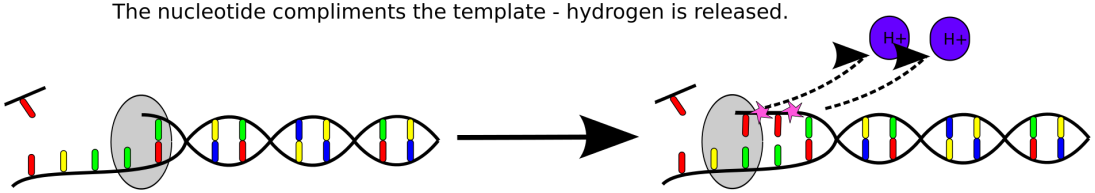
Ion Semiconductor Sequencing



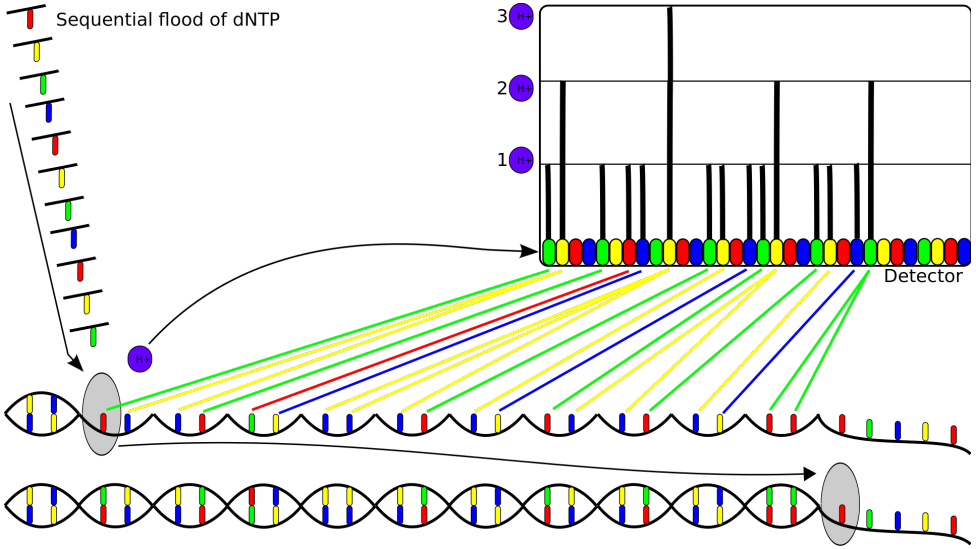
The nucleotide does not compliment the template - no release of hydrogen.



The nucleotide compliments the template - hydrogen is released.

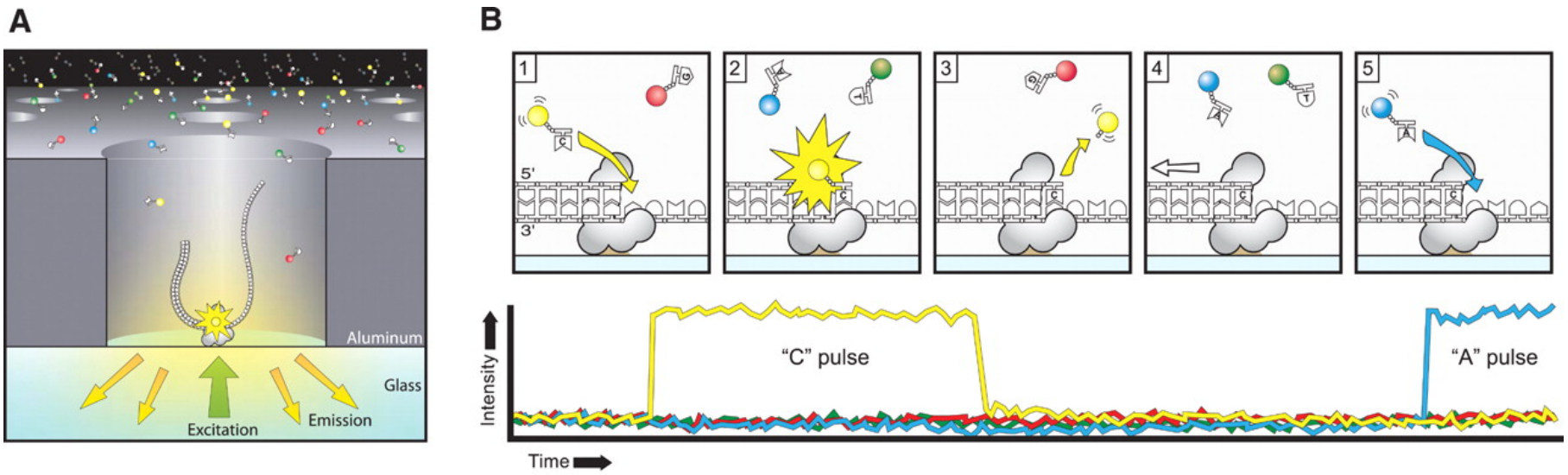


The nucleotide compliments several bases in a row - multiple hydrogen ions are released.



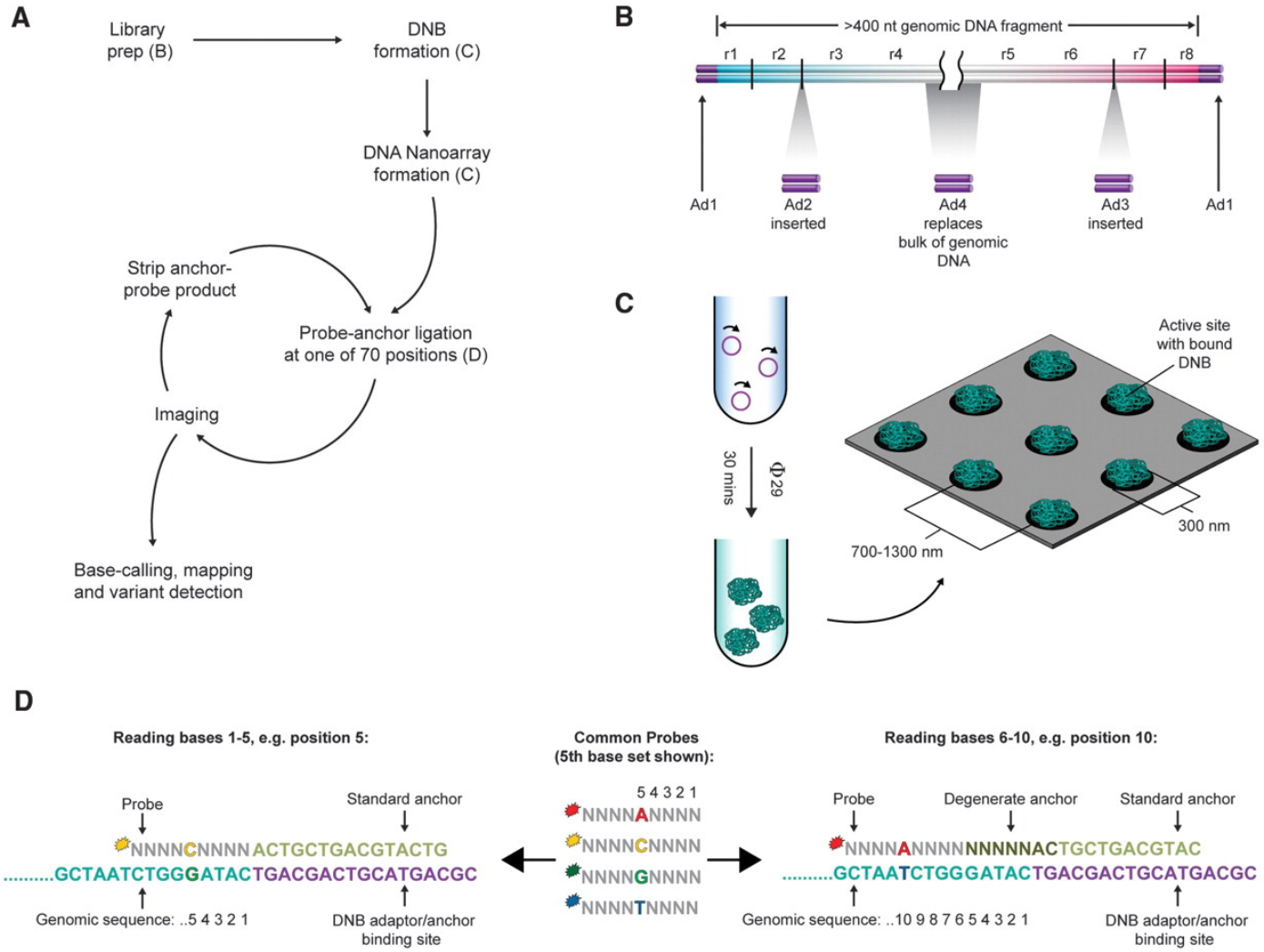
Source:http://en.wikipedia.org/wiki/Ion_semiconductor_sequencing

Pacific Biosciences Technology



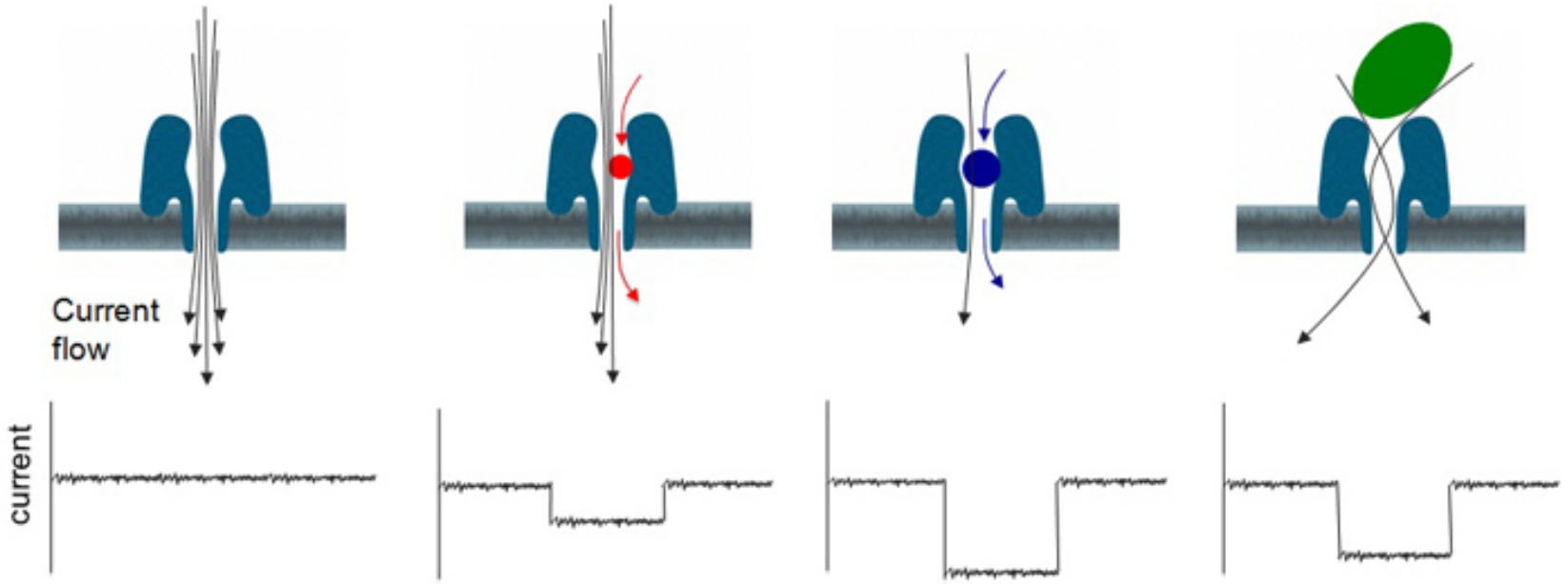
J Eid et al. Science 2009;323:133-138

Complete Genomics Technology



R Drmanac et al. Science 2010;327:78-81

Nanopores



Oxford Nanopore

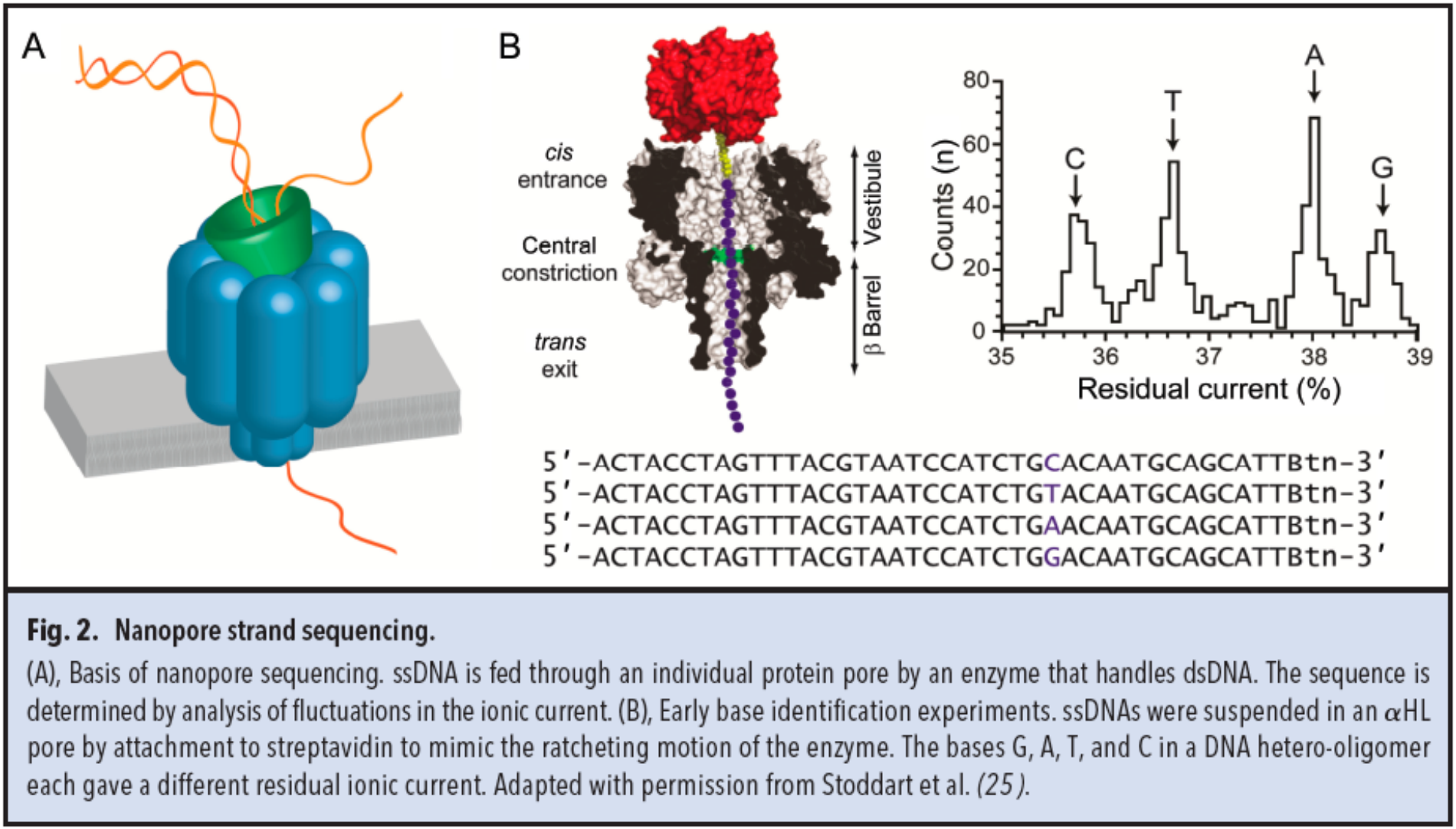
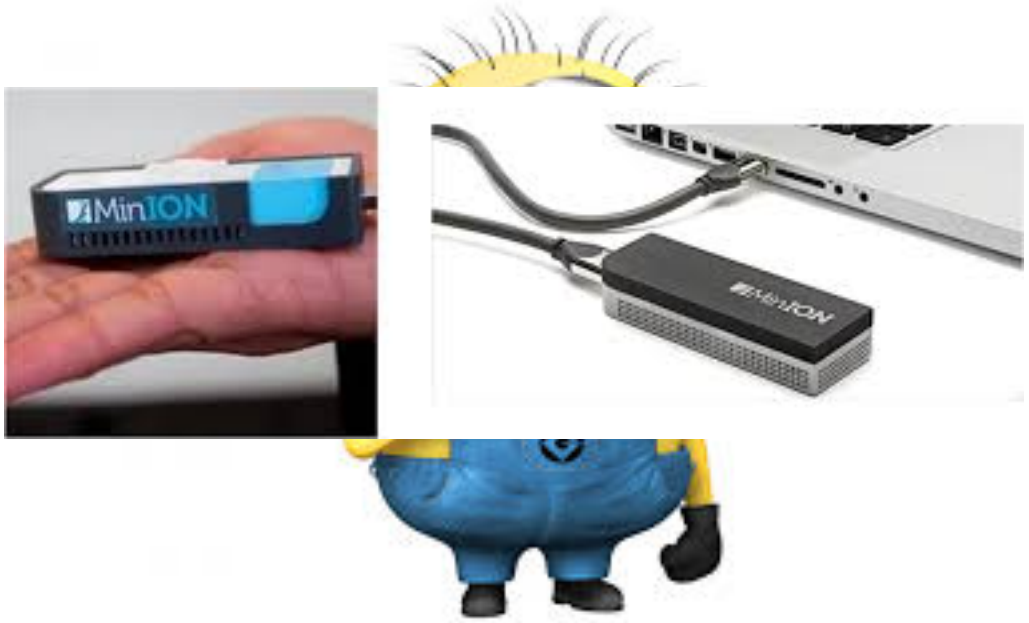


Fig. 2. Nanopore strand sequencing.

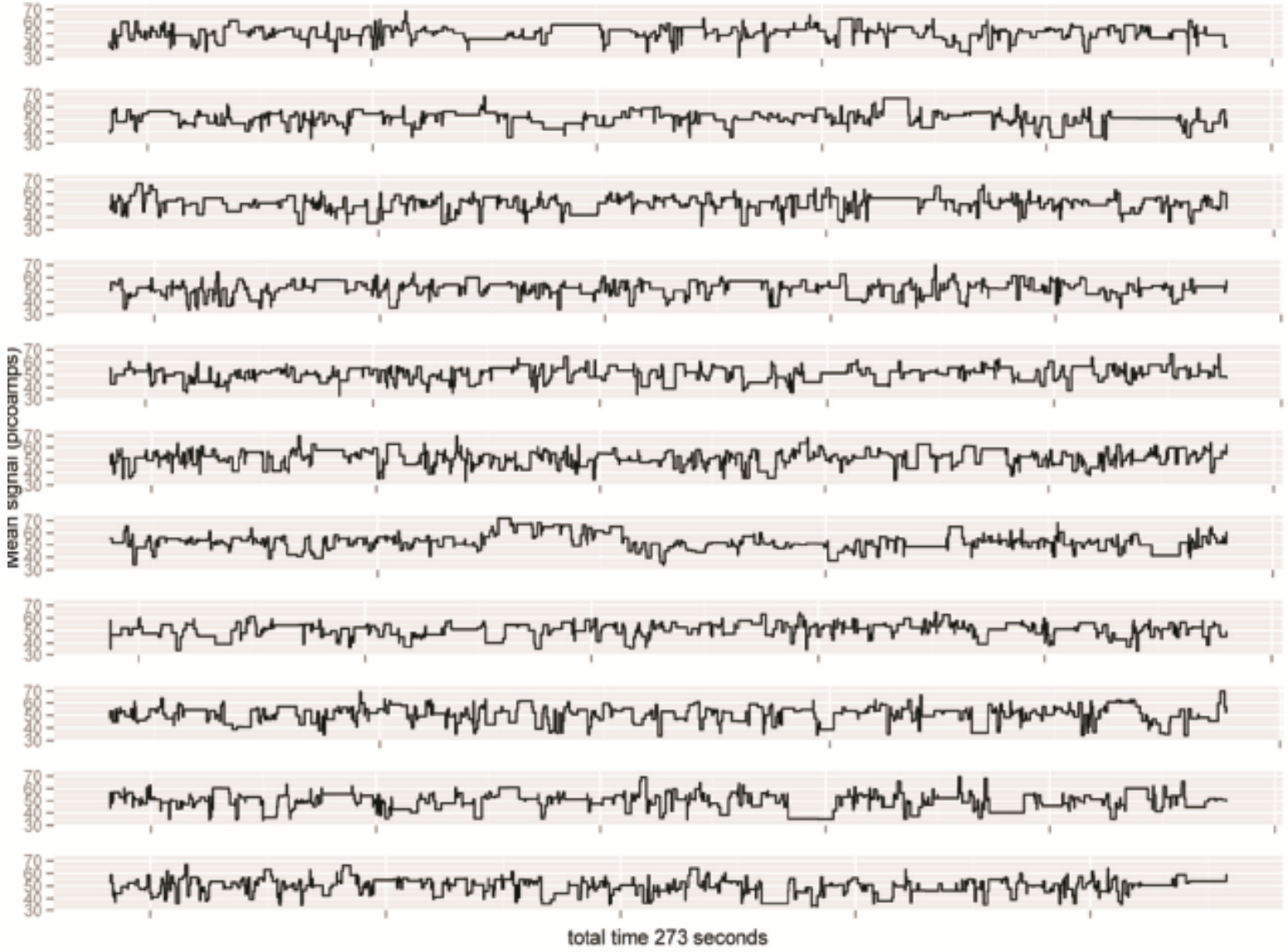
(A), Basis of nanopore sequencing. ssDNA is fed through an individual protein pore by an enzyme that handles dsDNA. The sequence is determined by analysis of fluctuations in the ionic current. (B), Early base identification experiments. ssDNAs were suspended in an αHL pore by attachment to streptavidin to mimic the ratcheting motion of the enzyme. The bases G, A, T, and C in a DNA hetero-oligomer each gave a different residual ionic current. Adapted with permission from Stoddart et al. *(25)*.

Bayley, H. Clinical Chemistry 61:1:25

The MinION

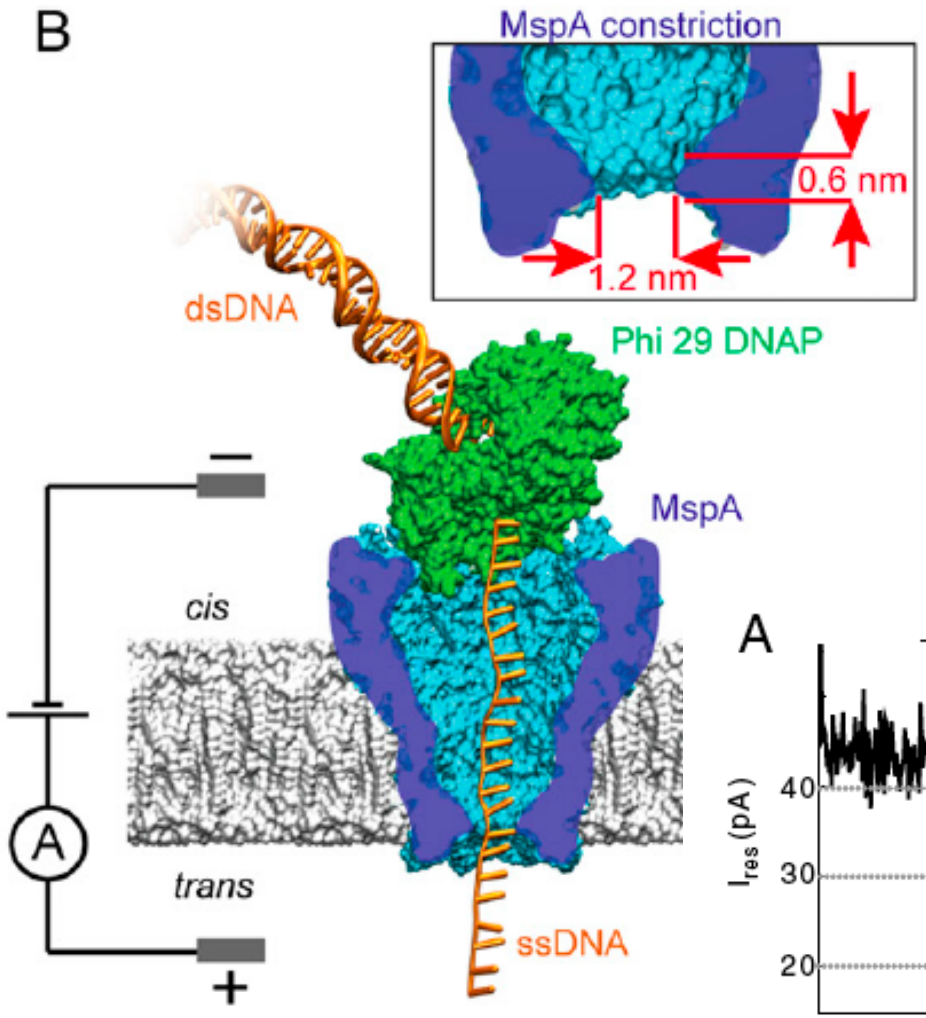


Sequence from the minION

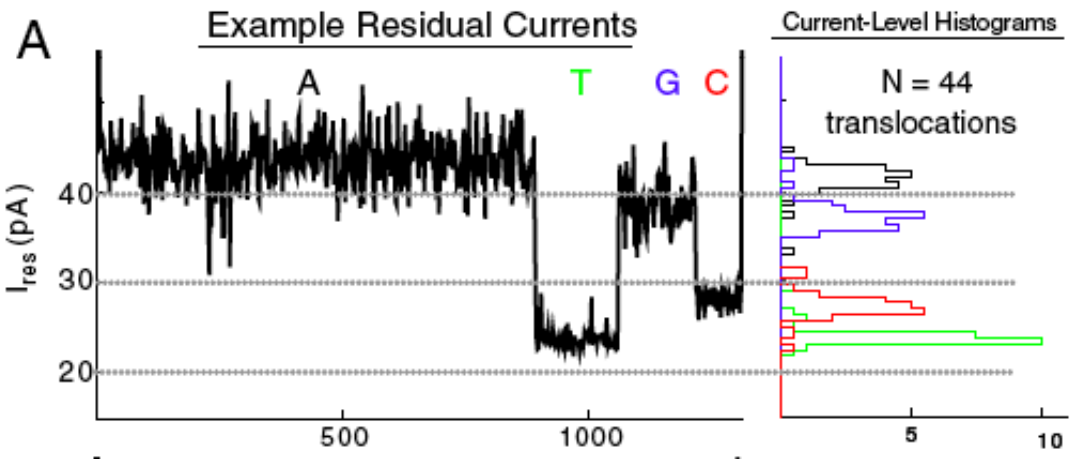


Bayley, H. Clinical Chemistry 61:1:25

MspA Nanopore



Laszlo et al PNAS 110:47:18904



Derrington et al. PNAS 107:37:16060

Sequencing DNA

Human Whole Genome Sequencing

- Initial Ref Sequence $300 million and took about a decade. (Draft reported in 2001)
- Humans sequenced
  - Craig Venter
  - James Watson
  - Yoruban from HapMap
  - Korean (35 individuals published in 2014)
  - Han Chinese
- Broad has released 60K Human Exomes
  - exac.broadinstitute.org
- Broad has “tweeted” it has completed 10K whole human genomes with the Illumina X Ten
- HiSeq2500 High Output--human genome can be sequenced for about $5,000 at an average read depth of 30X in 10 days
- HiSeq2500 Rapid Run— human genome can be sequenced in about 2 days to 30X coverage for ~$4,000.

DNA Sequencing with Next-Generation Technologies

A Draft Sequence of the Neandertal Genome

Richard E. Green,[1]*†‡ Johannes Krause,[1]†§ Adrian W. Briggs,[1]†§ Tomislav Maricic,[1]†§
Udo Stenzel,[1]†§ Martin Kircher,[1]†§ Nick Patterson,[2]†§ Heng Li,[2]† Weiwei Zhai,[3]†‖
Markus Hsi-Yang Fritz,[4]† Nancy F. Hansen,[5]† Eric Y. Durand,[3]† Anna-Sapfo Malaspinas,[3]†
Jeffrey
Hernán
Barbar
Eric S.
Christi
Vladim
Javier
Daniel
Janet K

Sequencing the nuclear genome of the extinct woolly mammoth

Webb Miller[1], Daniela I.
Michael D. Packard[1], Fan
Kerstin Lindblad-Toh[5], Eri
Sharon Sheridan[7], Tom P

Genetic history of an archaic hominin group from Denisova Cave in Siberia

David Reich[1,2]*, Richard E. Green[3,4]*, Martin Kircher[3]*, Johannes Krause[3,5]*, Nick Patterson[2]*, Eric Y. Durand[6]*, Bence Viola[3,7]*
Adrian W. Briggs[1,3], U
Can Alkan[10], Qiaome
Michael Richards[7,13],
Montgomery Slatkin[6]

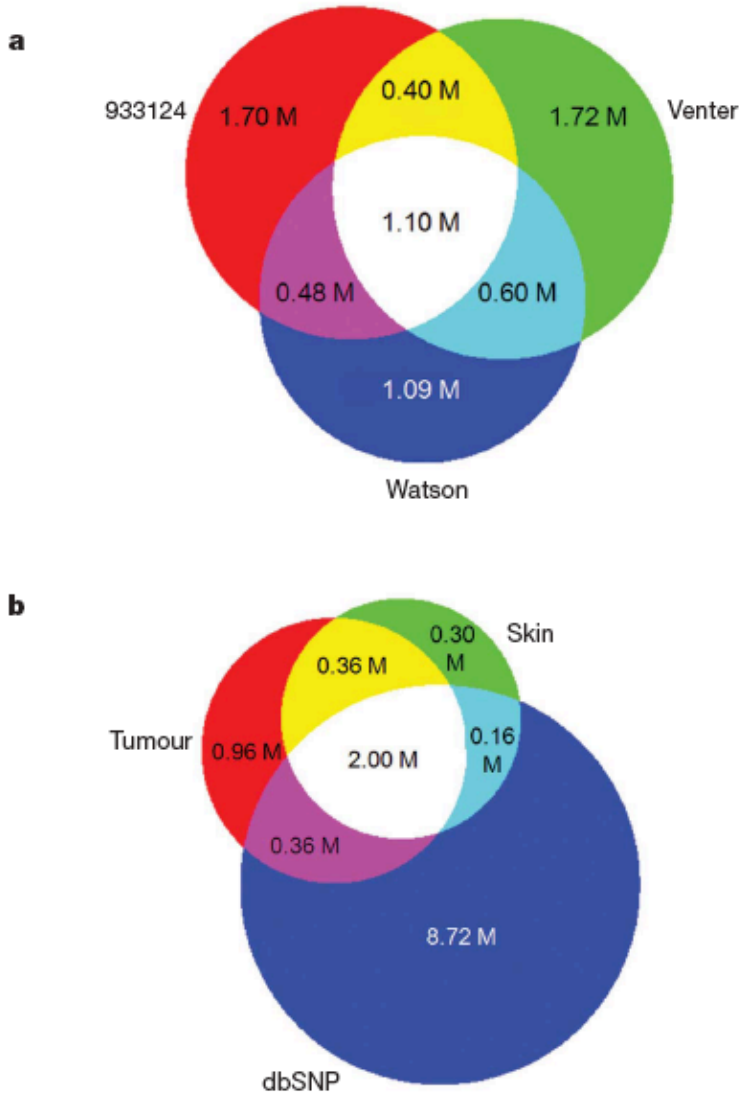
Insights into hominid evolution from the gorilla genome sequence

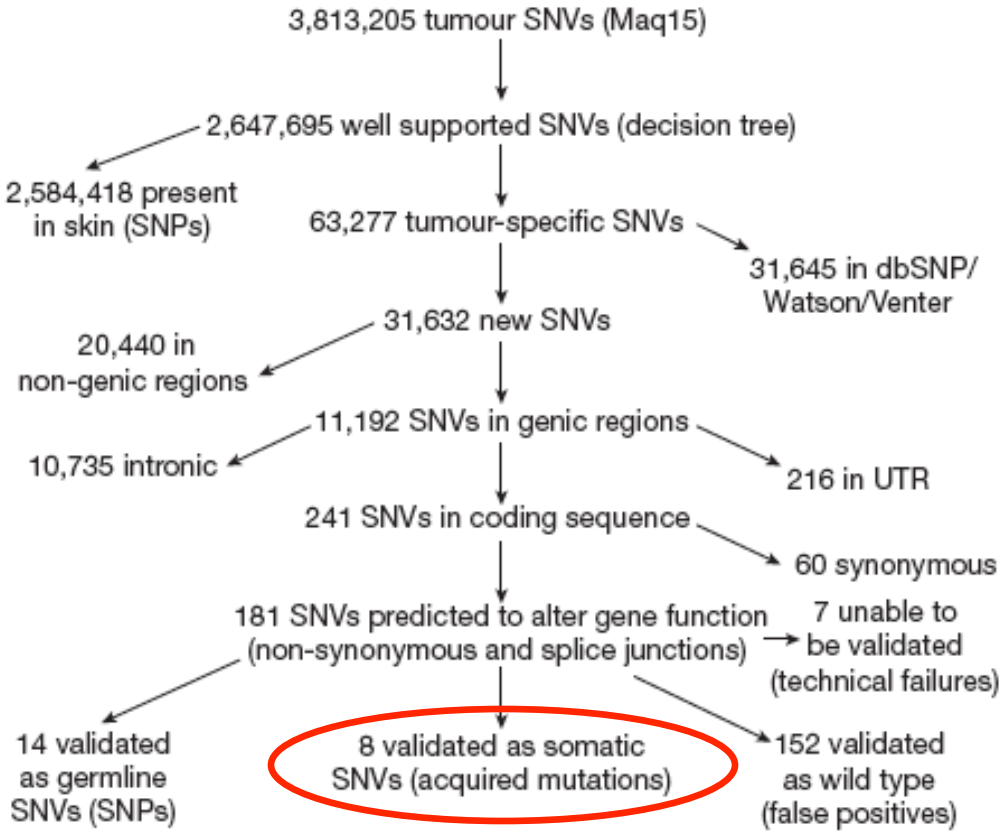
Aylwyn Scally[1], Julien Y. Dutheil[2]†, LaDeana W. Hillier[3], Gregory E. Jordan[4], Ian Goodhead[1]†, Javier Herrero[4], Asger Hobolth[2], Tuuli Lappalainen[5], Thomas Mailund[2], Tomas Marques-Bonet[3,6,7], Shane McCarthy[1], Stephen H. Montgomery[8], Petra C. Schwalie[4], Y. Amy Tang[1], Michelle C. Ward[9,10], Yali Xue[1], Bryndis Yngvadottir[1]†, Can Alkan[3,11], Lars N. Andersen[2], Qasim Ayub[1], Edward V. Ball[12], Kathryn Beal[4], Brenda J. Bradley[8,13], Yuan Chen[1], Chris M. Clee[1], Stephen Fitzgerald[4], Tina A. Graves[14], Yong Gu[1], Paul Heath[1], Andreas Heger[15], Emre Karakoc[3], Anja Kolb-Kokocinski[1], Gavin K. Laird[1], Gerton Lunter[16], Stephen Meader[15], Matthew Mort[12], James C. Mullikin[17], Kasper Munch[2], Timothy D. O'Connor[8], Andrew D. Phillips[12], Javier Prado-Martinez[6], Anthony S. Rogers[1]†, Saba Sajjadian[3], Dominic Schmidt[9,10], Katy Shaw[12], Jared T. Simpson[1], Peter D. Stenson[12], Daniel J. Turner[1]†, Linda Vigilant[18], Albert J. Vilella[4], Weldon Whitener[1], Baoli Zhu[19]†, David N. Cooper[12], Pieter de Jong[19], Emmanouil T. Dermitzakis[5], Evan E. Eichler[3,11], Paul Flicek[4], Nick Goldman[4], Nicholas I. Mundy[8], Zemin Ning[1], Duncan T. Odom[1,9,10], Chris P. Ponting[15], Michael A. Quail[1], Oliver A. Ryder[20], Stephen M. Searle[1], Wesley C. Warren[14], Richard K. Wilson[14], Mikkel H. Schierup[2], Jane Rogers[1]†, Chris Tyler-Smith[1] & Richard Durbin[1]

Applications

- Whole Genome Sequencing
- Exome Sequencing
- Targeted Genomic Sequencing
- Chromatin-IP-Sequencing
- DNAse I Hypersensitivity Sequencing
- Methyl-Seq (RRBS, MeDIP, etc)
- Microbiome Sequencing
- Metagenomics

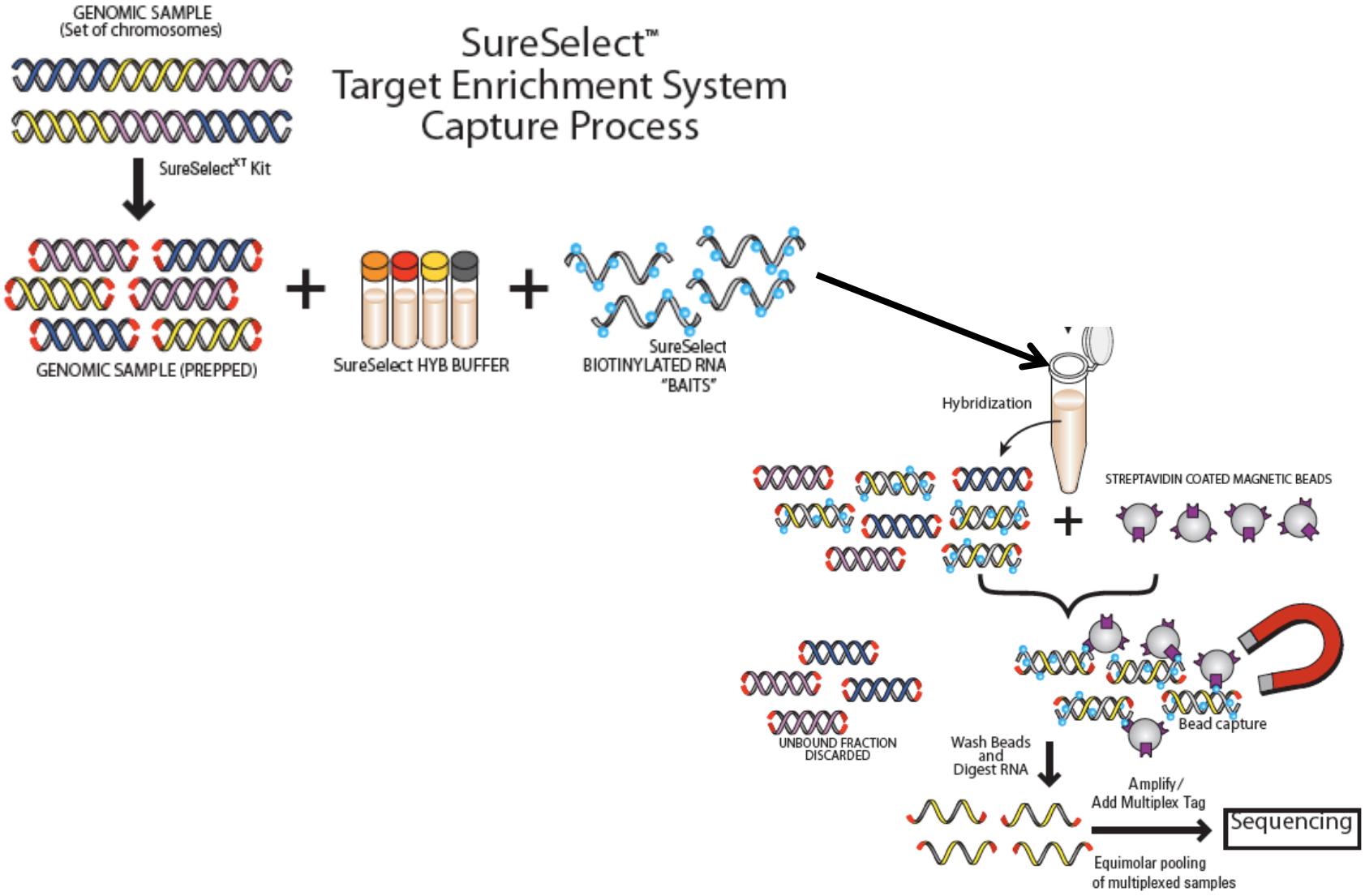
AML:Comparisons





Lei T. et al. 2008. Nature 456:6

SureSelect Exome Capture



Disease Genes Discovered by Direct Whole Exome Sequencing*

Gene Identified	Disease/Syndrome	Reference
MYH3	Freeman-Sheldon Syndrome	Ng SB, et al. 2009. Nature 462
SLC26A3	Bartter Syndrome	Choi M, et al. 2009 PNAS 106(45)
DHODH	Miller Syndrome	Ng SB, et al. 2010 Nat Genet 42(1).
FLVCR2	Fowler Syndrome	Lalonde, E. et al. 2010 Hum Mutat 31(8).
FLNA	Terminal Osseous Dysplasia (TOD)	Sun Y., et al. 2010 Am J. Hum Genet 87(1).
GPSM2	Nonsyndromic Hearling Loss (DFNB82)	Walsh, T. et al. 2010 Am J. Hum Genet 87(1).
HSD17B4	Perrault Syndrome/DBP	Pierce SB, et al. 2010 Am J. Hum Genet 87(2).
MLL2	Kabuki Syndrome	Ng SB, et al. 2010 Nat Genet 42(9).
ABCG5	Hypercholesterolemia	Rios J., et al. 2010 Hum Mol Genet 19(22).
WDR62	Brain Malformations	Bilguvar K, et al. 2010 Nature 467(7312).
PIGV	Hyperphosphatasia Mental Retardation (HPMR)	Krawitz PM, et al. 2010 Nat Genet 42(10)
WDR35	Sensenbrenner Syndrome	Gilissen C, et al. 2010Am J Hum Genet 87(3).
SDCCAG8	Nephromophthisis-related Ciliopathies	Otto EA, et al. 2010 Nat Genet 42(10).
STIM1	Kaposi Sarcoma	Byn M, et al. 2010 J Exp Med 207(11).
SCARF2	Van Den Ende-Gupta Syndrome	Anastasio N. et al. 2010 Am J Hum Genet 87(4).
C20orf54	Brown-Vialetto-Van Laere Syndrome	Green P, et al. 2010 Am J Hum Genet 86(3).
MASP1	Carnevale, Malpuech, OSA and Michels Syndromes	Sirmaci A, at al. 2010 Am J Hum Genet 87(5).
ABCC8	Neonatal Diabetes Mellitus	Bonnefond A, et al. 2010 PLoS One 5(10).
BAP-1	Metastasizing Uveal Melanomas	Harbour JW, et al. 2010 Science Nov 4 Epub.
ACAD9	Complex I Deficiency	Haack TB, et al. 2010 Nat Genet Nov 7 Epub.
DYNC1H1	Mental Retardation	Vissers LELM, et al. 2010 Nat Genet 10.1038/ng.712
RAB39A	Mental Retardation	Vissers LELM, et al. 2010 Nat Genet 10.1038/ng.712
YY1	Mental Retardation	Vissers LELM, et al. 2010 Nat Genet 10.1038/ng.712
DEAF1	Mental Retardation	Vissers LELM, et al. 2010 Nat Genet 10.1038/ng.712

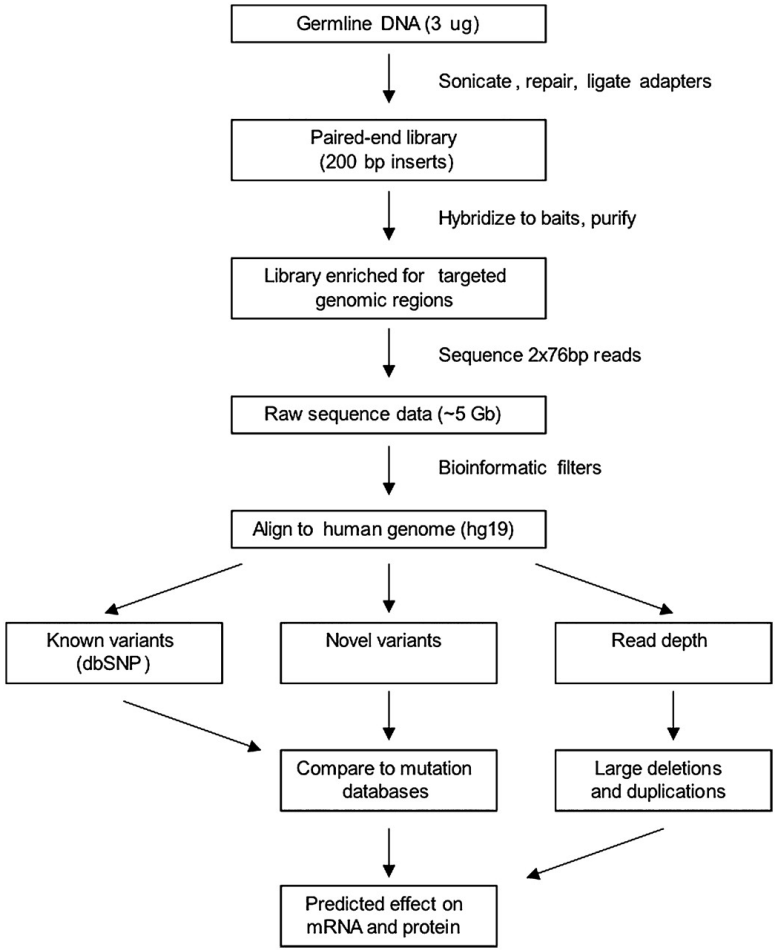
*As of 23 Nov. 2010

Targeted Re-sequencing

The ability to capture specific sequences in the genome

Long range PCR
Multiplex PCR strategies
Solution capture on Biotin labeled oligos
HaloPlex

Genomic Capture of Breast Cancer Relevant Genes Followed by Next-Gen Sequencing.

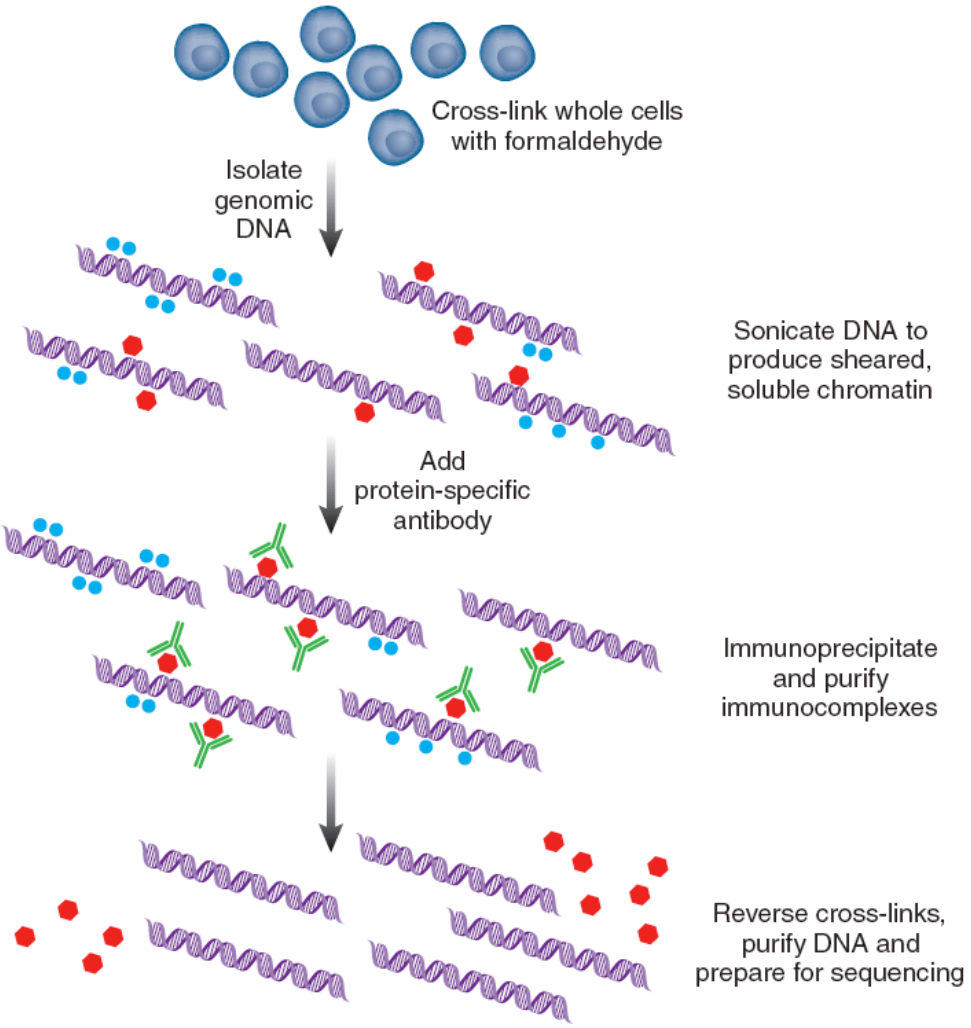


Walsh T et al. PNAS 2010;107:12629-12633

Gene	Chromosome	Start	End
BRCA1	17	41,186,313	41,347,712
BRCA2	13	32,879,617	32,983,809
CHEK2	22	29,073,731	29,147,822
PALB2	16	23,604,483	23,662,678
BRIP1	17	59,759,985	59,940,755
p53	17	7,561,720	7,600,863
PTEN	10	89,613,195	89,738,532
STK11	19	1,195,798	1,238,434
CDH1	16	68,761,195	68,879,444
ATM	11	108,083,559	108,249,826
BARD1	2	215,583,275	215,684,428
MLH1	3	37,024,979	37,102,337
MRE11	11	94,140,467	94,237,040
MSH2	2	47,620,263	47,720,360
MSH6	2	48,000,221	48,044,092
MUTYH	1	45,784,914	45,816,142
NBN	8	90,935,565	91,006,899
PMS1	2	190,638,811	190,752,355
PMS2	7	6,002,870	6,058,737
RAD50	5	131,882,630	131,989,595
RAD51C	17	56,759,963	56,821,692

Johnson et al., Science 316:1497 (2007)

ChIP-Seq



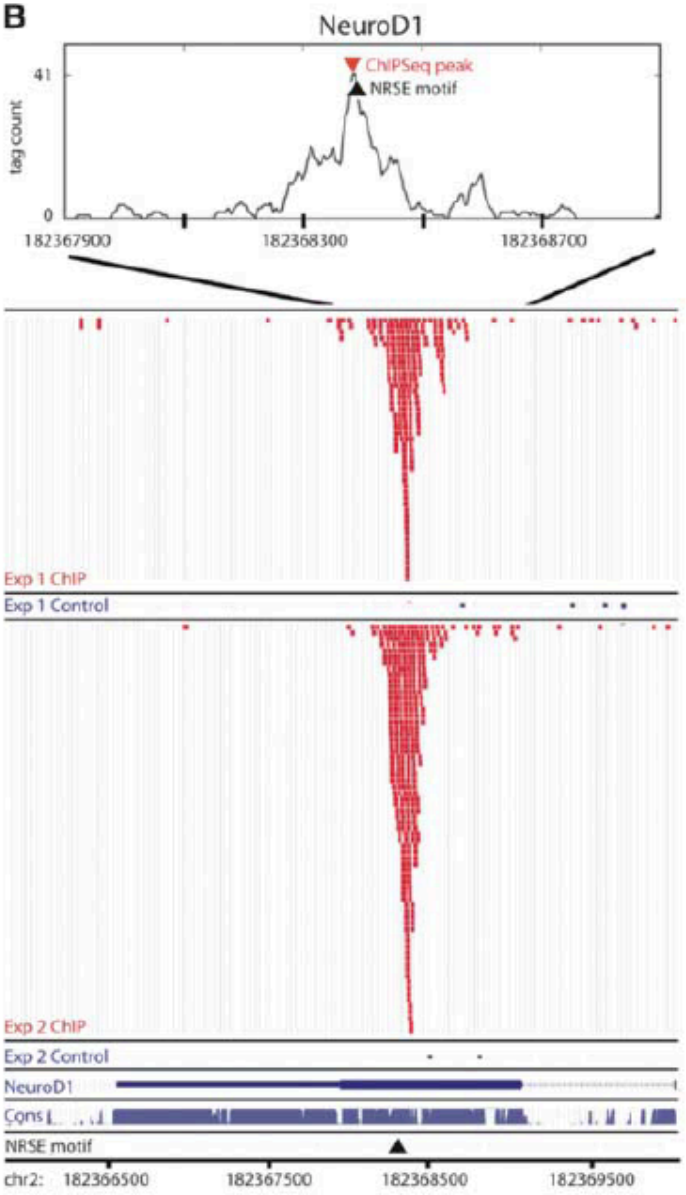
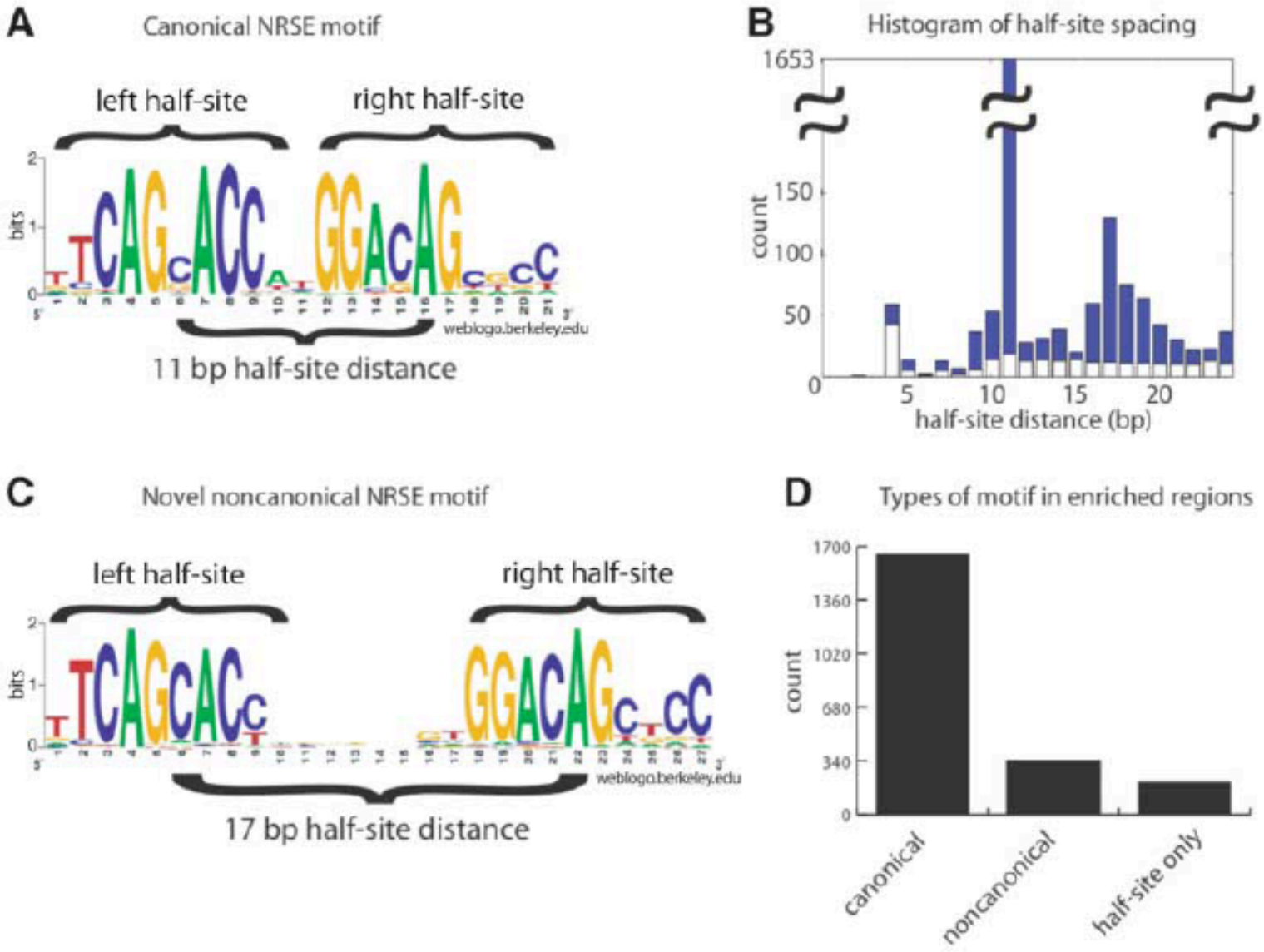


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

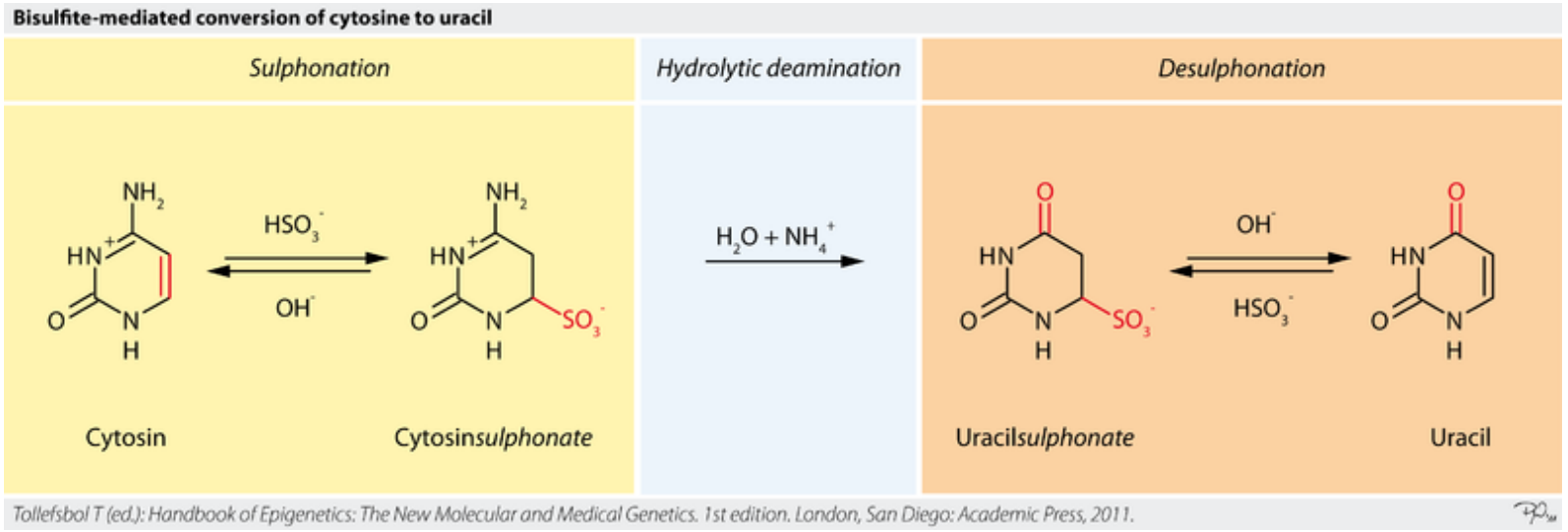
ChIP-Seq



Johnson et al., Science 316:1497 (2007)

Methylation profiling

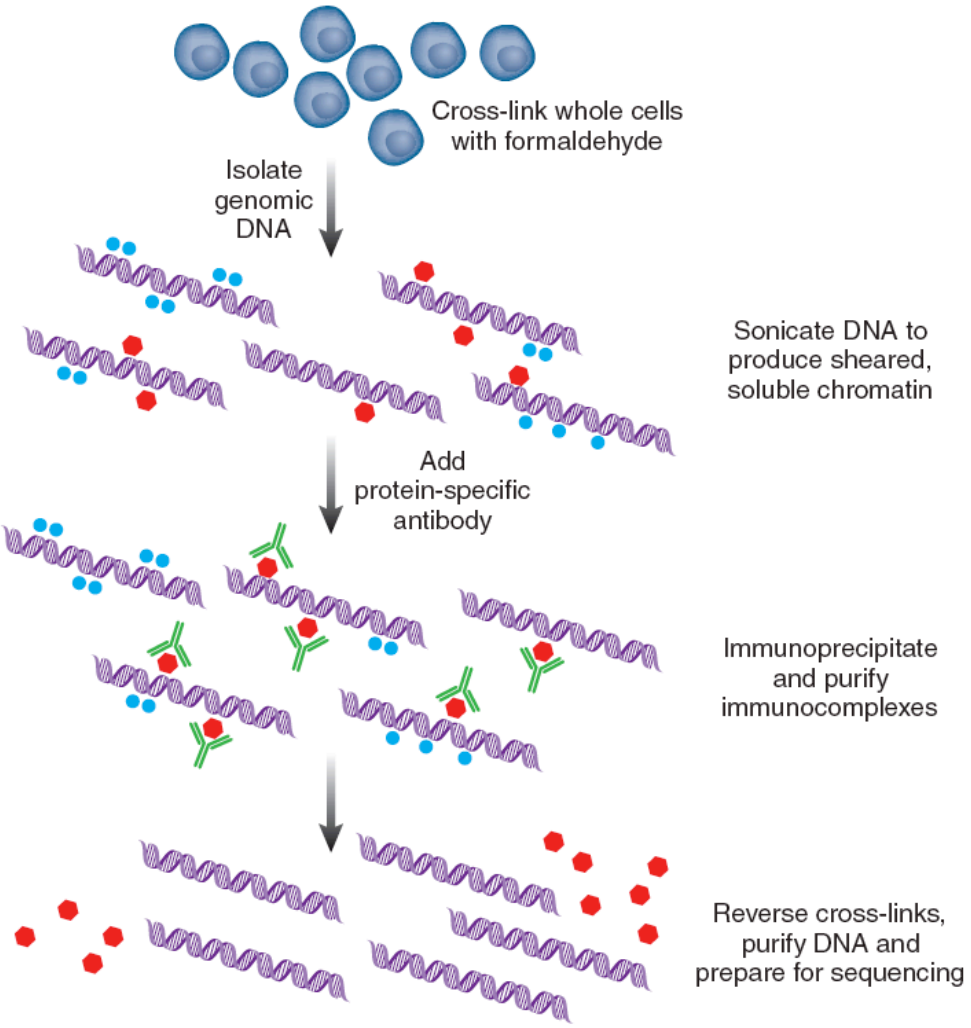
- Whole genome bisulfite sequencing
- MeDIP (Methylated DNA-IP)
- Reduced Representational Bisulfite Sequencing
- Specific Capture methods



Tollefsbol T (ed.): Handbook of Epigenetics: The New Molecular and Medical Genetics. 1st edition. London, San Diego: Academic Press, 2011.

Johnson et al., Science 316:1497 (2007)

MeDIP-Seq



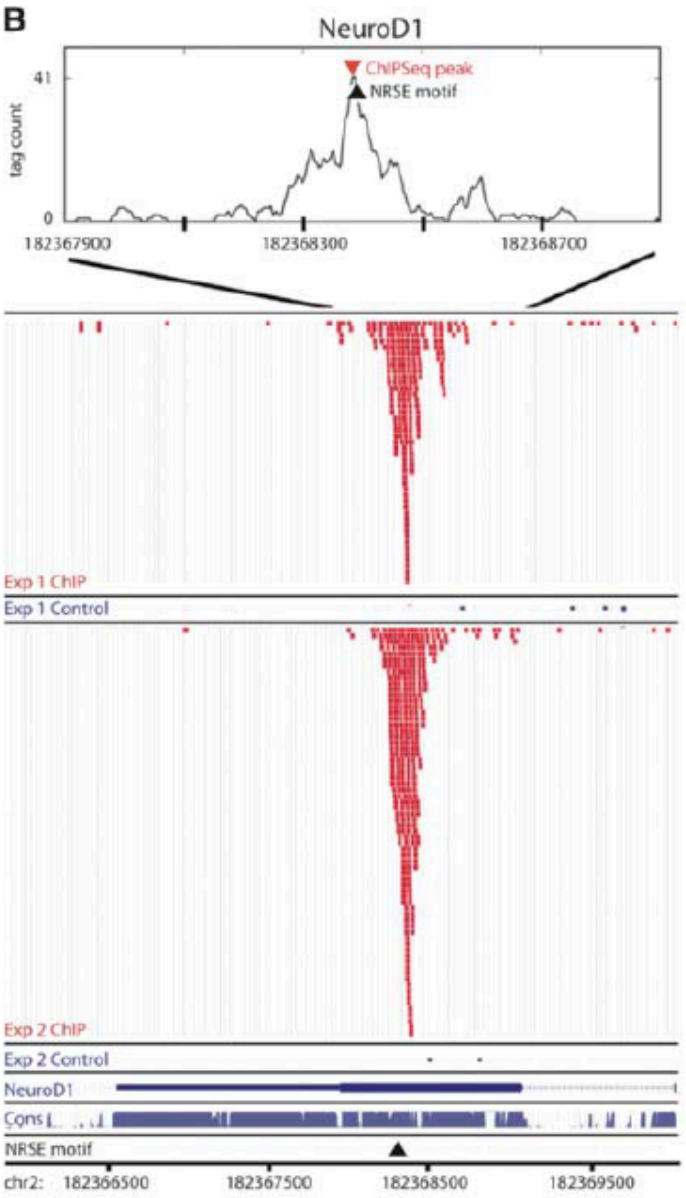
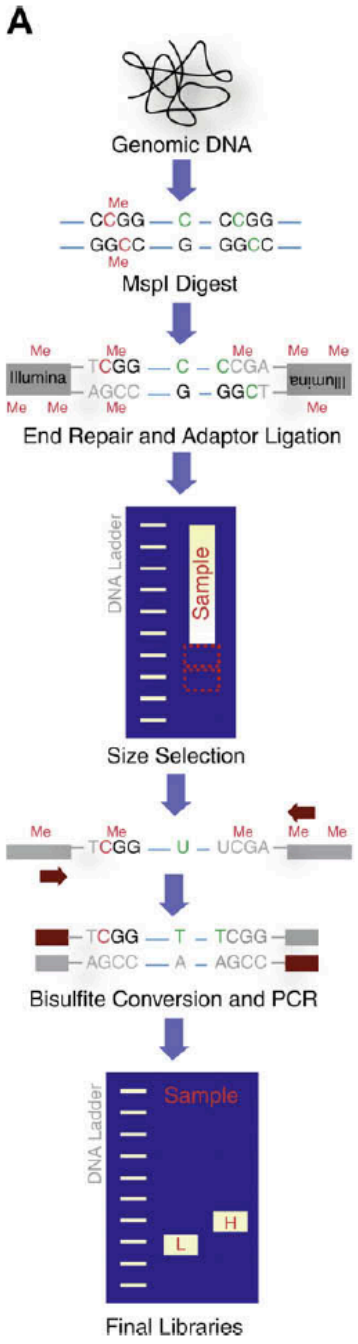
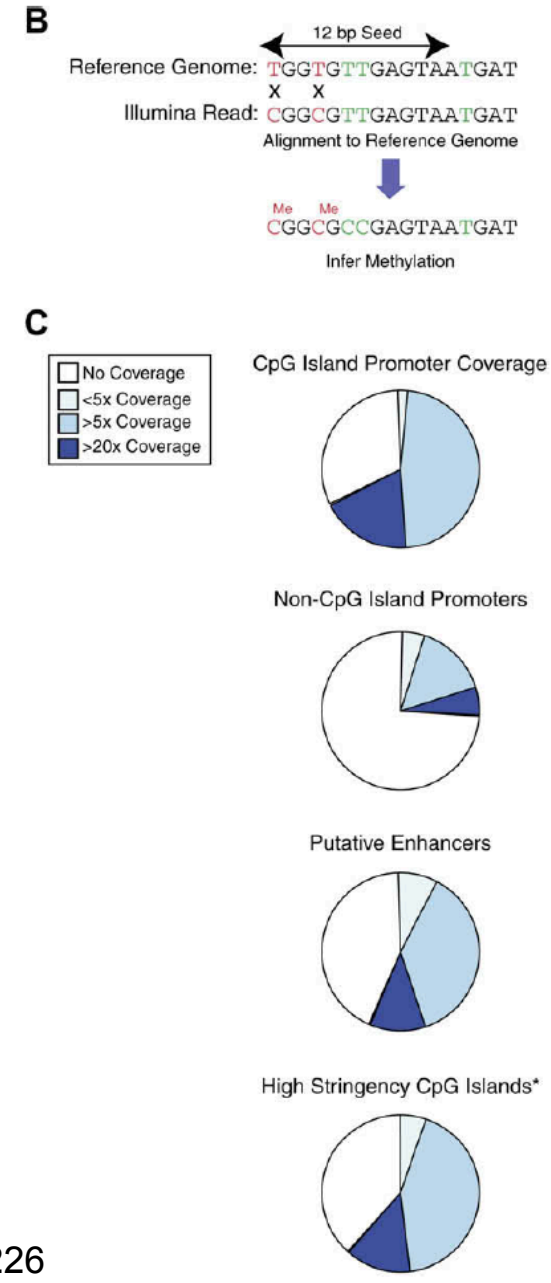


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

RRBS





Smith et al., 2009. Methods 48:226

The NIH Human Microbiome Project

The NIH HMP Working Group[1]

The Human Microbiome Project (HMP), funded as an initiative of the NIH Roadmap for Biomedical Research (http://nihroadmap.nih.gov), is a multi-component community resource. The goals of the HMP are: (1) to take advantage of new, high-throughput technologies to characterize the human microbiome more fully by studying samples from multiple body sites from each of at least 250 "normal" volunteers; (2) to determine whether there are associations between changes in the microbiome and health/disease by studying several different medical conditions; and (3) to provide both a standardized data resource and new technological approaches to enable such studies to be undertaken broadly in the scientific community. The ethical, legal, and social implications of such research are being systematically studied as well. The ultimate objective of the HMP is to demonstrate that there are opportunities to improve human health through monitoring or manipulation of the human microbiome. The history and implementation of this new program are described here.

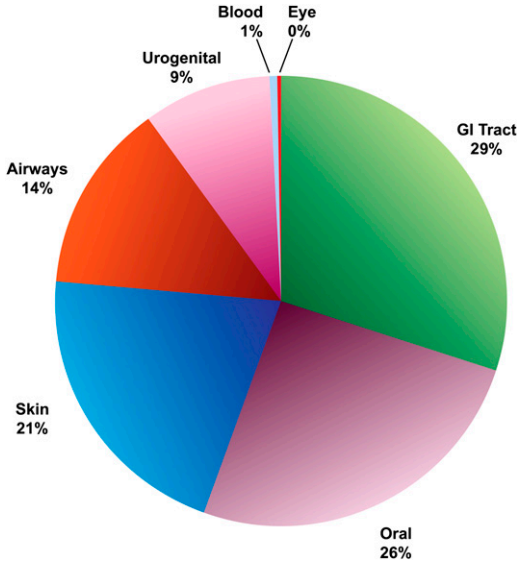
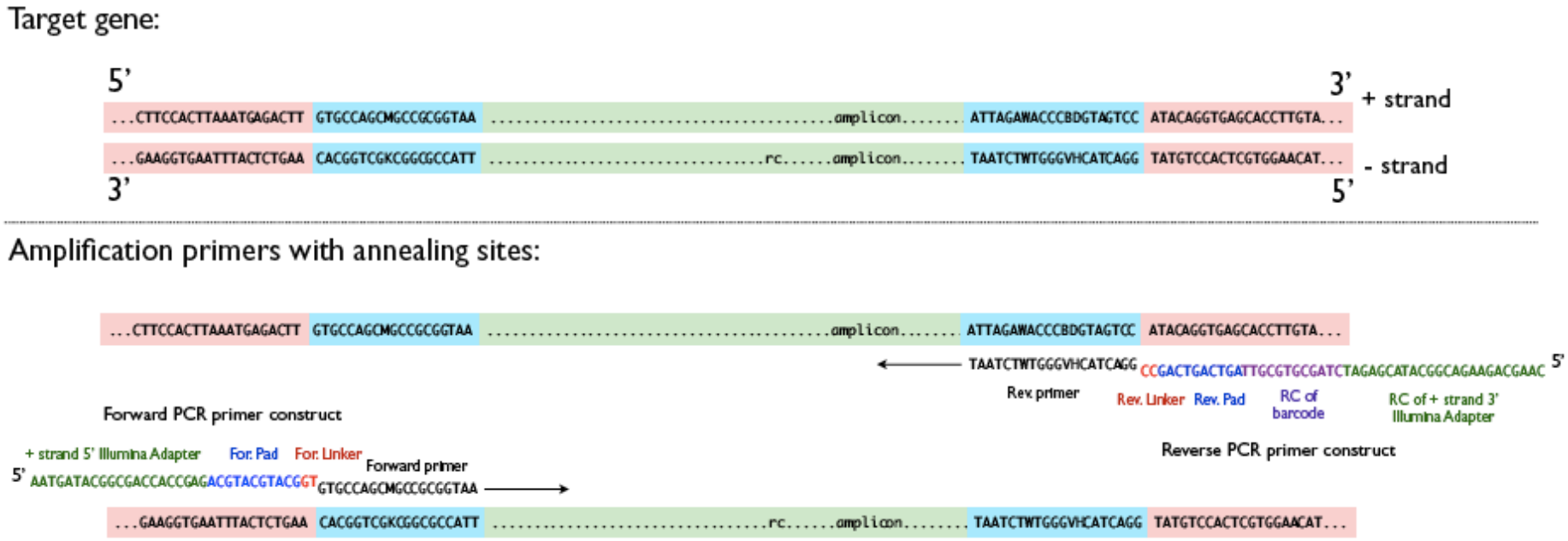


Figure 3. Bacterial distribution by body site. This figure shows the distribution by body site of bacteria that have been sequenced under the HMP or are in the sequencing pipelines.

Peterson J, et al. 2009. Genome Res 19:2317

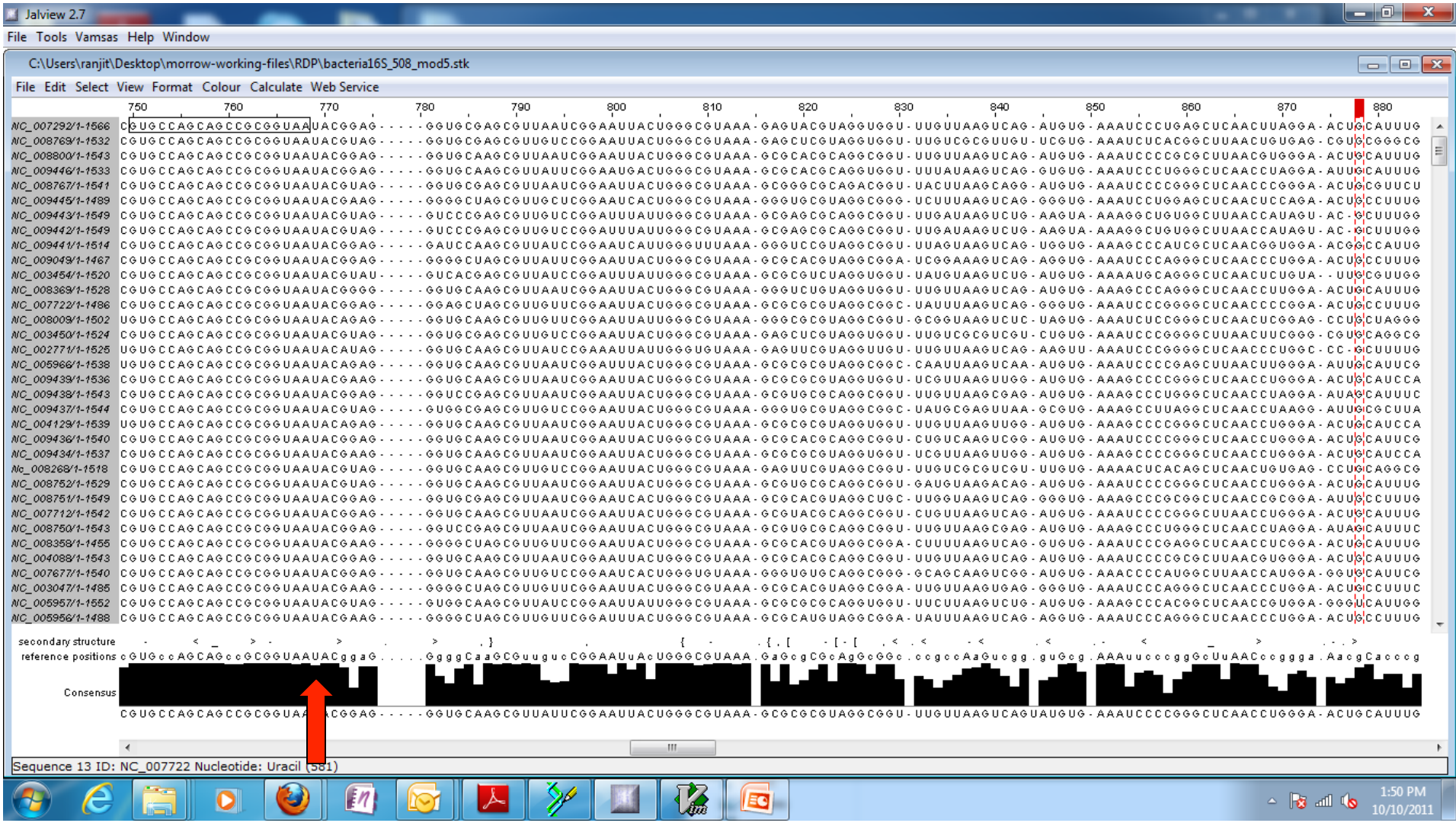
Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample

J. Gregory Caporaso[a], Christian L. Lauber[b], William A. Walters[c], Donna Berg-Lyons[b], Catherine A. Lozupone[a], Peter J. Turnbaugh[d], Noah Fierer[b,e], and Rob Knight[a,f,1]

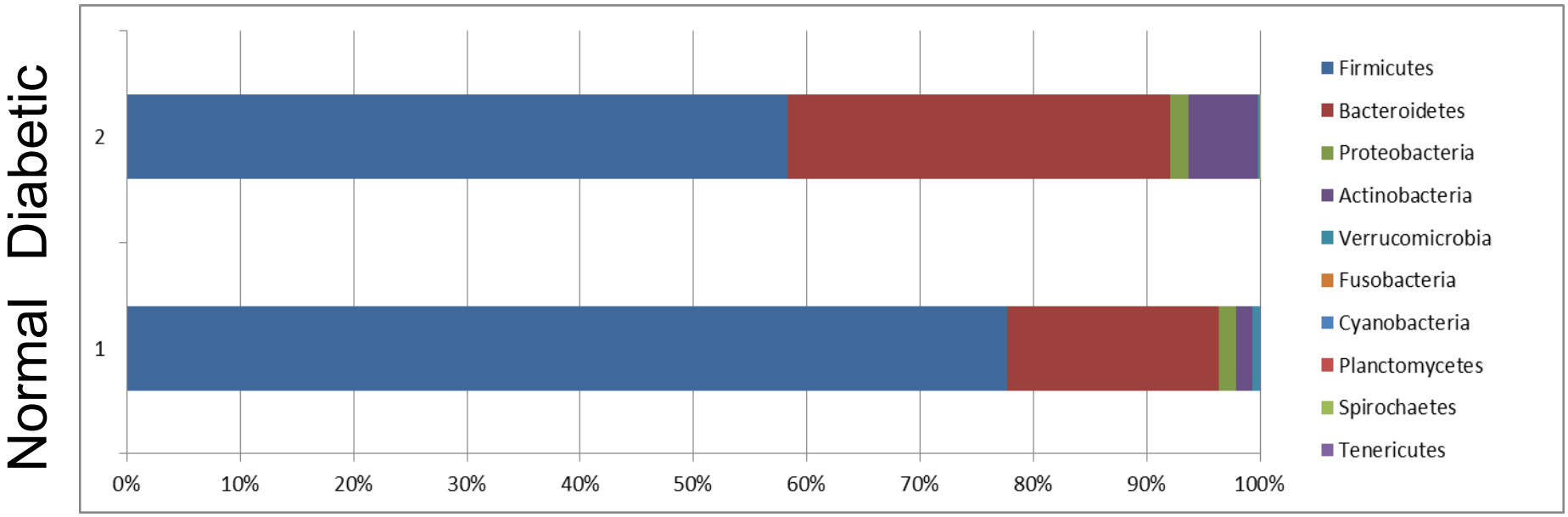


PNAS 2011. 108:supp1:4516

MSA after forward primer



Microbiome at UAB



OPEN ACCESS Freely available online

PLoS one

Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults

Nadja Larsen[1]*, Finn K. Vogensen[1], Frans W. J. van den Berg[1], Dennis Sandris Nielsen[1], Anne Sofie Andreasen[2], Bente K. Pedersen[2], Waleed Abu Al-Soud[3], Søren J. Sørensen[3], Lars H. Hansen[3], Mogens Jakobsen[1]

1 Department of Food Science, University of Copenhagen, Frederiksberg, Denmark, 2 Department of Infectious Diseases and CMRC, University Hospital Rigshospitalet, Copenhagen, Denmark, 3 Department of Biology, University of Copenhagen, Copenhagen, Denmark

The proportions of phylum Firmicutes and class Clostridia were significantly reduced in the diabetic group compared to the control group (P = 0.03).

Sequencing RNA

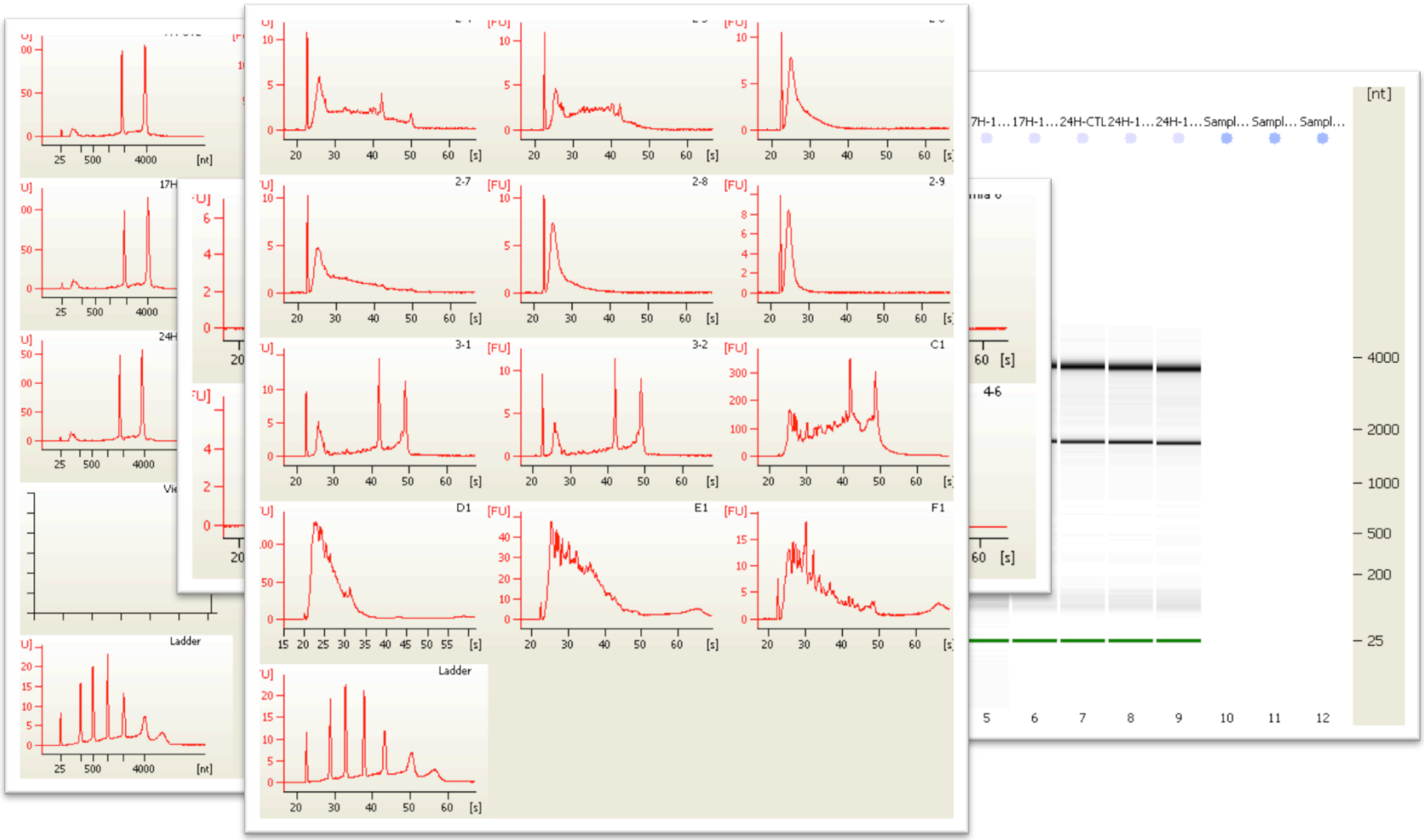
RNA Applications

- mRNA Sequencing (RefSeq, RNASeq)
- microRNA Sequencing
- RNA-IP-Sequencing
- CLIP or HITS-CLIP or PAR-CLIP
- Ribosome Profiling

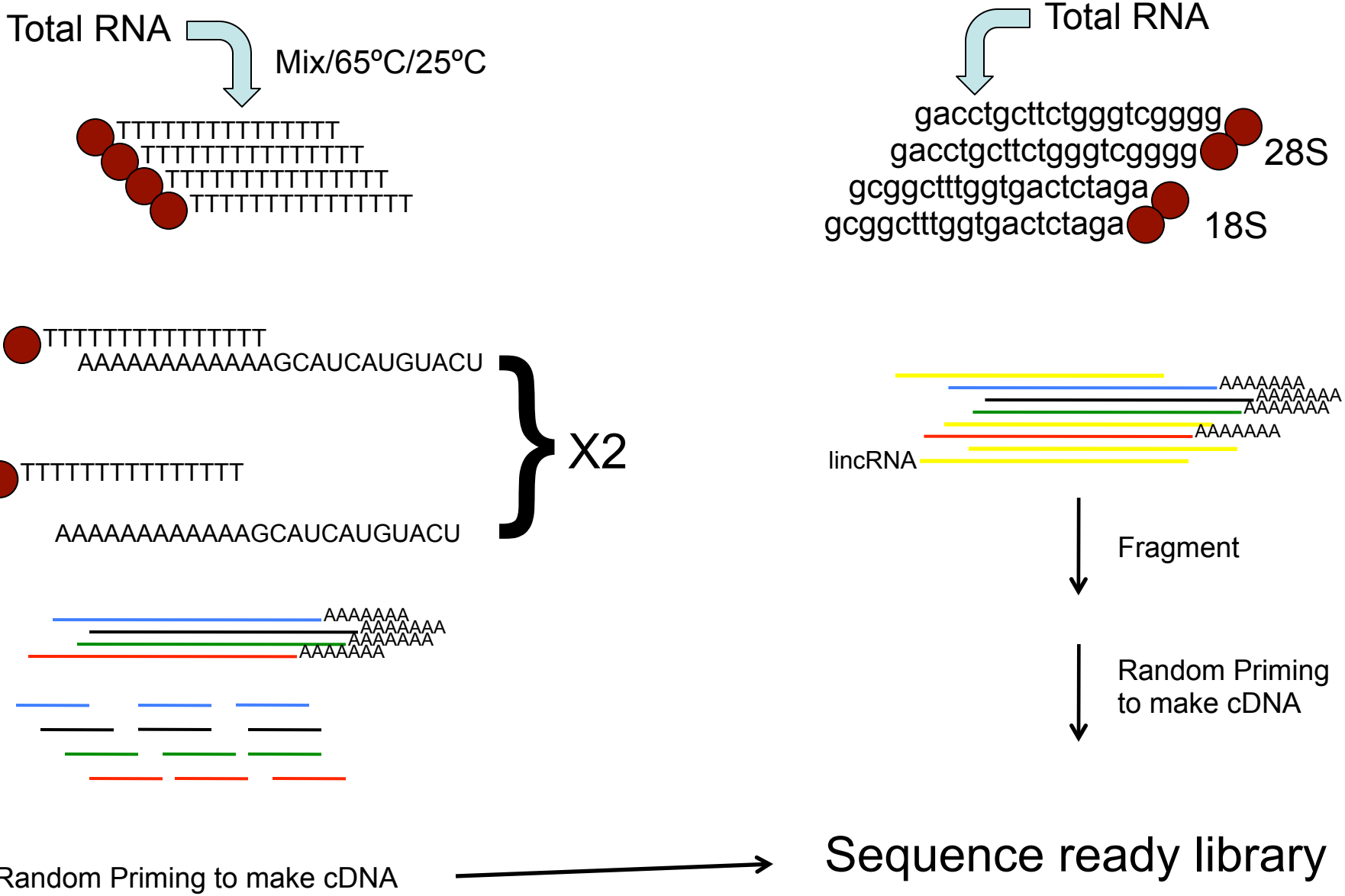
Advantages of RNA-Seq

- Digital gene expression
  - Simply count the number of reads for a given transcript
- Greater dynamic range
- No hybridization bias
- Not dependent on known content
- Generate alternative splice/exon usage
- Identify variants
- Allele Specific Expression
- Identify RNA editing

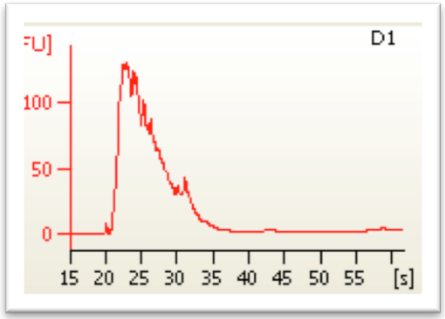
RNA Quality


Ladder
2-7
2-8
2-9
3-1
3-2
C1
D1
E1
F1
Ladder
[nt]
4000
2000
1000
500
200
25

RNA-Seq Library Prep



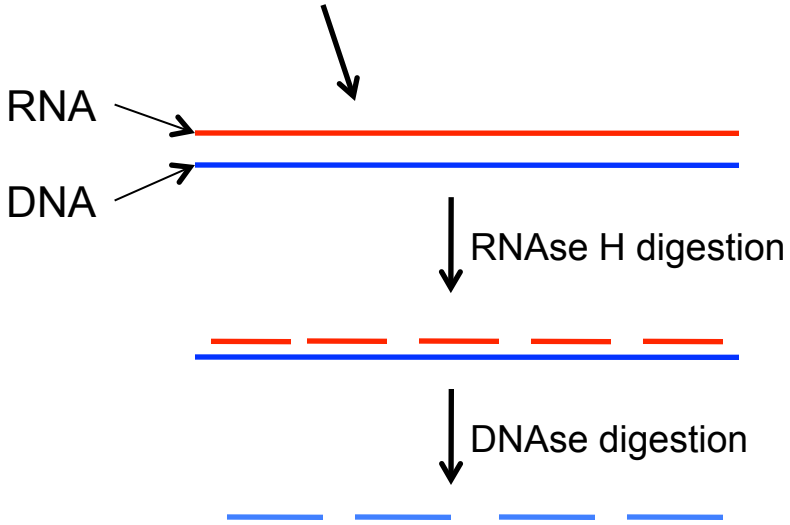
Degraded RNA (FFPE)

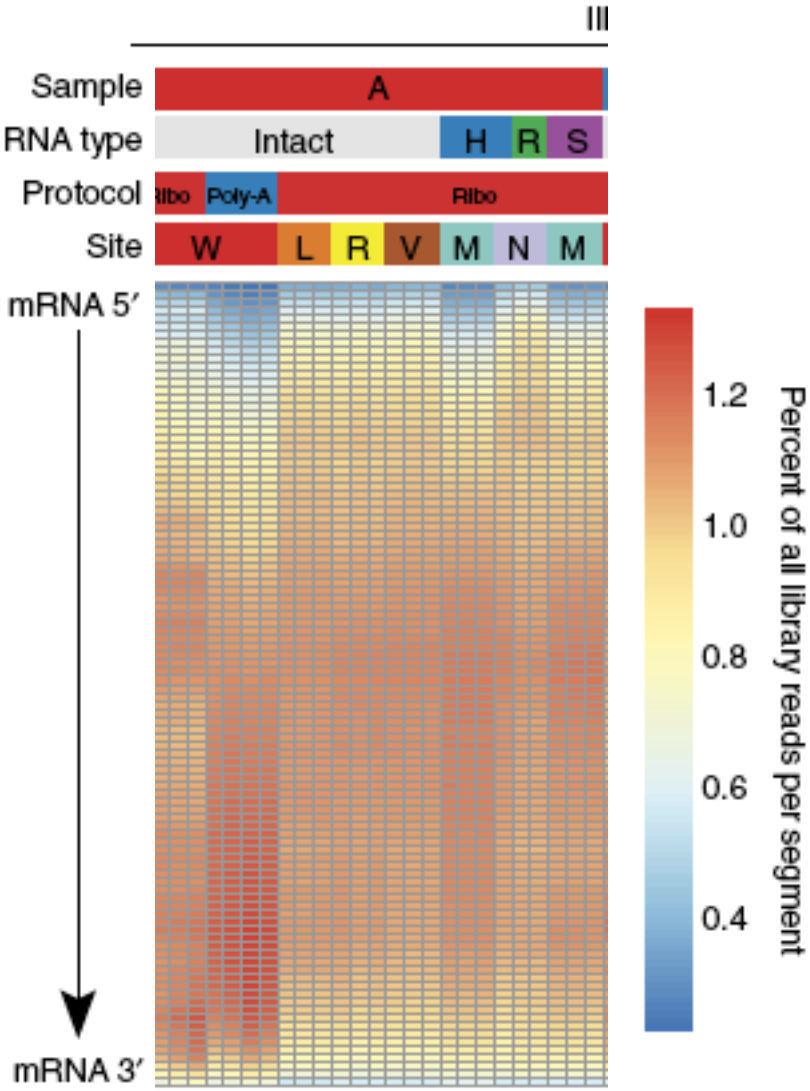


Total RNA

gacctgcttctgggtcgggg
gacctgcttctgggtcgggg 28S

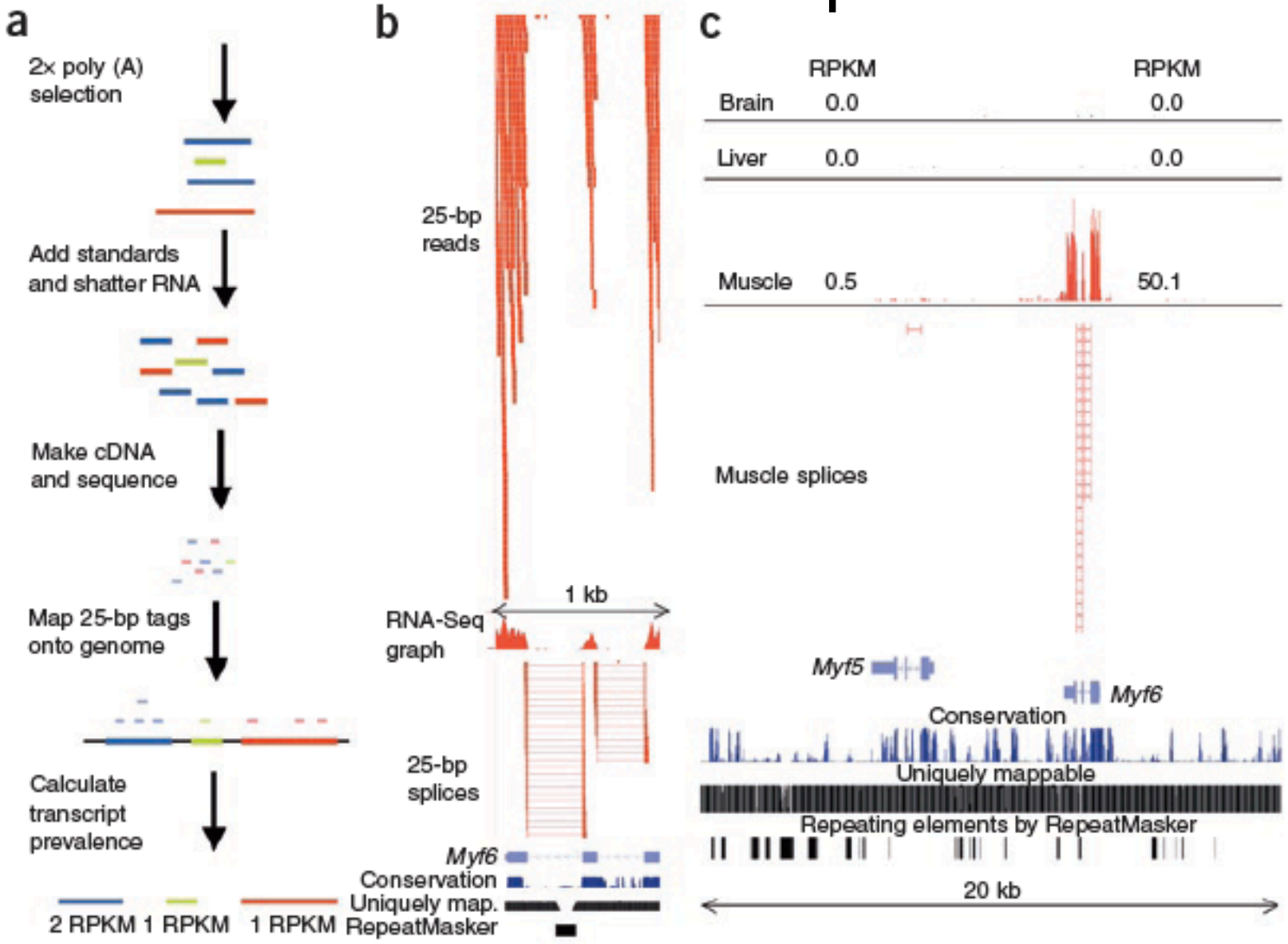
gcggctttggtgactctaga
gcggctttggtgactctaga 18S





From Li et al., 2014 Nat. Biotech.32:915

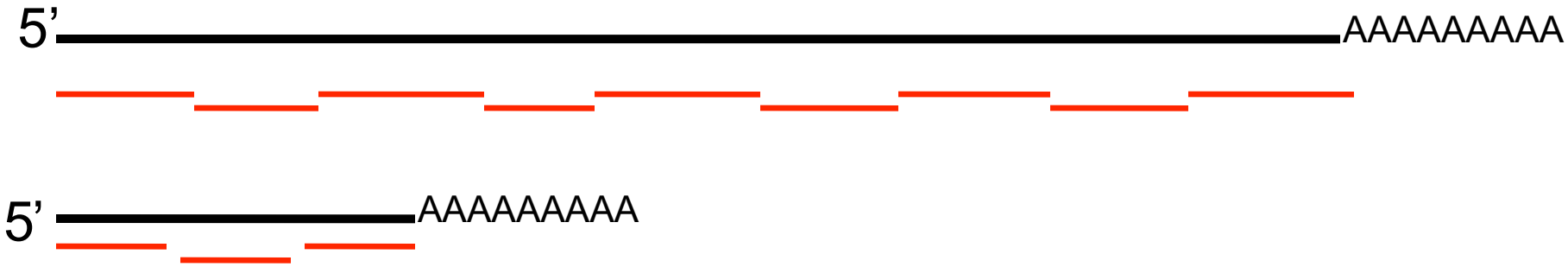
mRNA-Seq



Mortazavi et al., Nature Methods 5:7:621 (2008)

Digital Gene Expression

- Caveat?

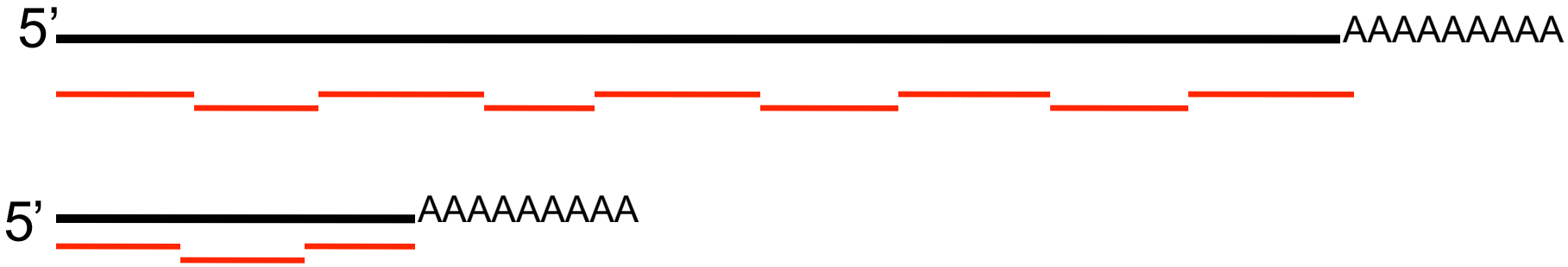


Gene 1 has 9 reads
Gene 2 has 3 reads

Gene 1 would appear to be expressed at 3X the amount of Gene 2

Digital Gene Expression

- Caveat?



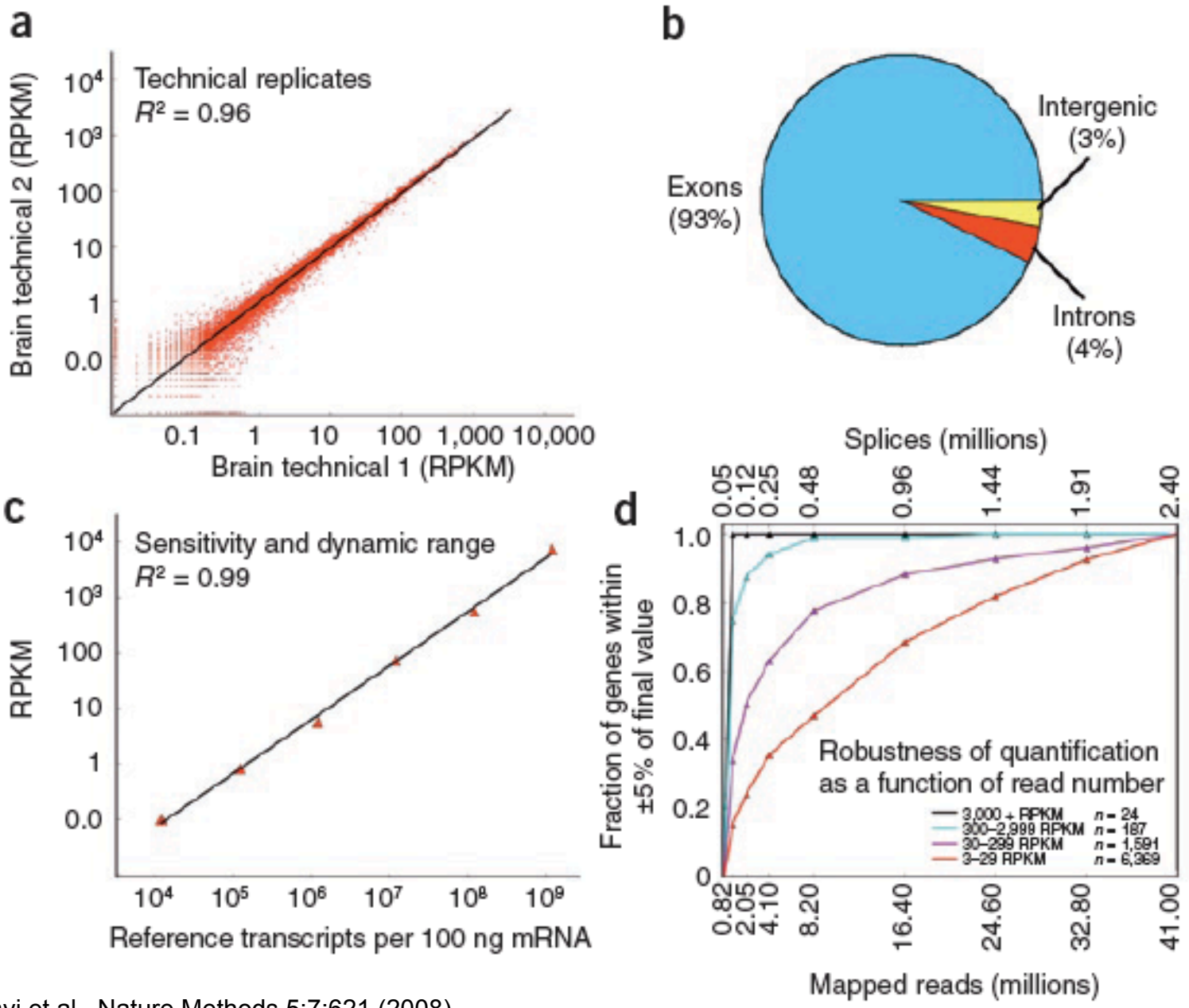
Gene 1 has 9 reads
Gene 2 has 3 reads

Gene 1 is 9kb
Gene 2 is 3kb

Normalize by length
e.g. 9reads/9kb=1

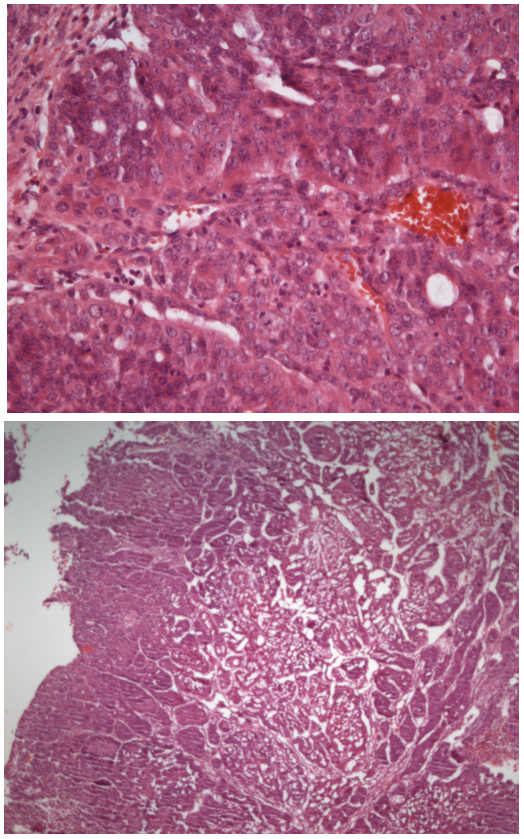
RPKM: reads per kilobase of exon model per million mapped reads

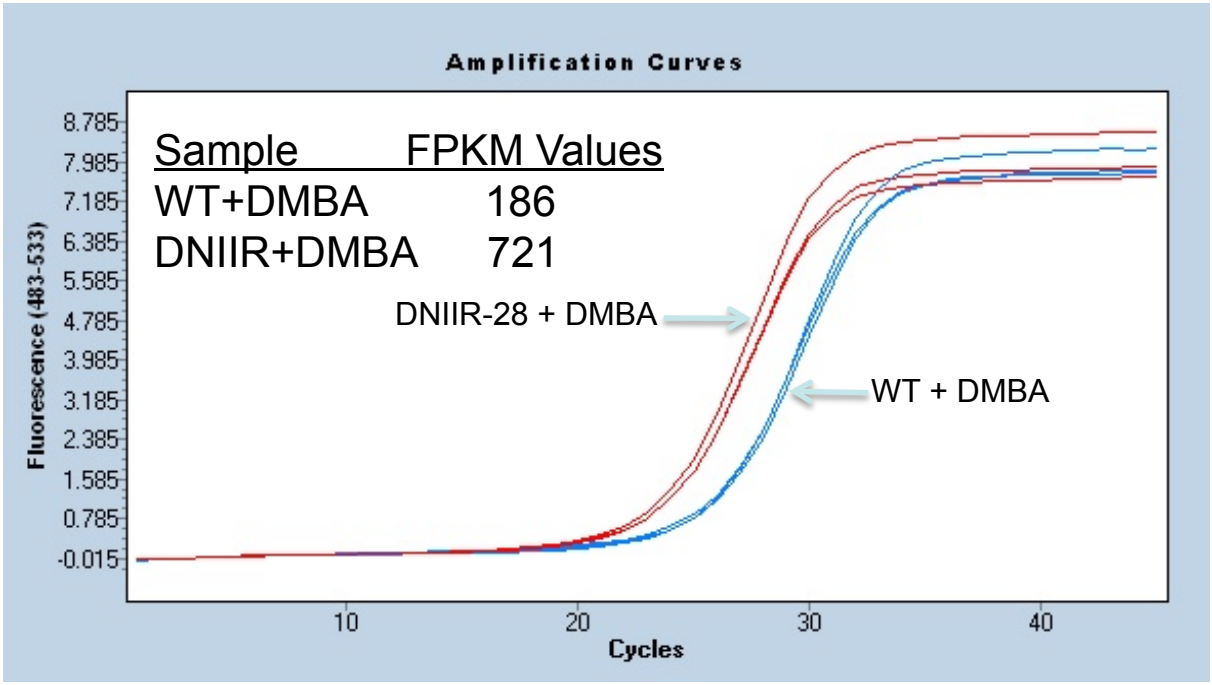
FPKM: fragment of reads per kilobase of exon model per million mapped reads (usually 25bp fragments).



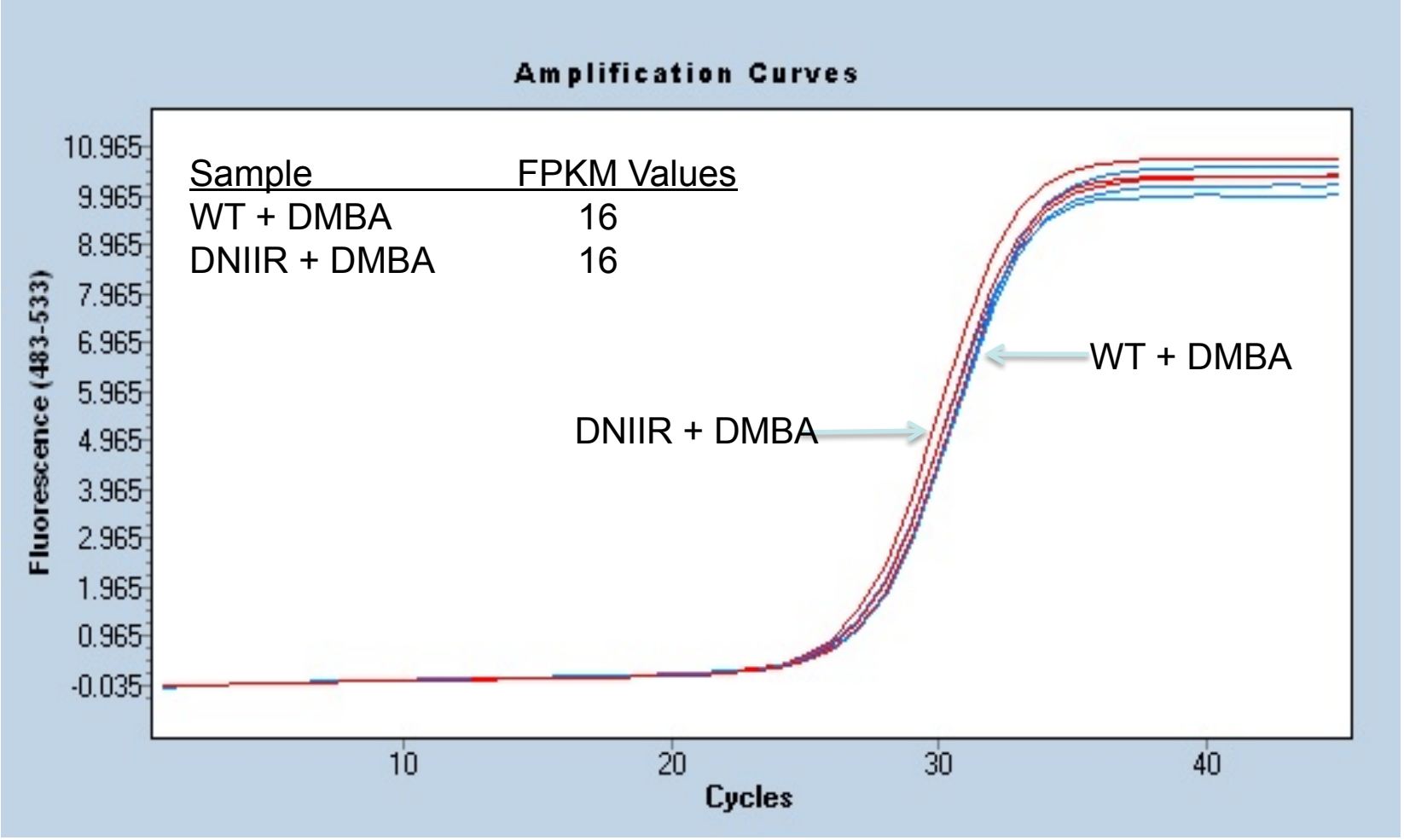
Mortazavi et al., Nature Methods 5:7:621 (2008)

Keratin 8

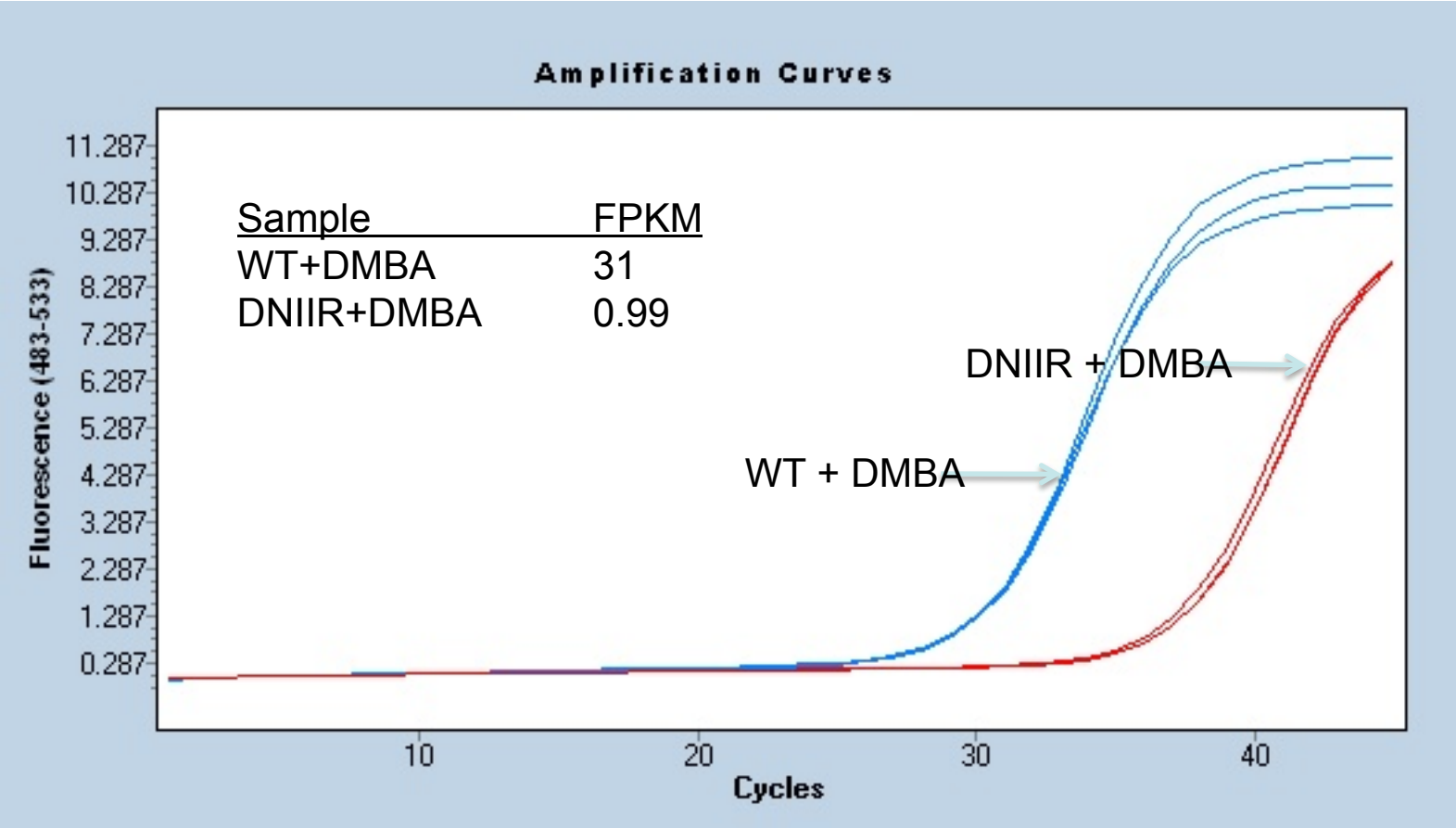



Amplification Curves
Fluorescence (483-533)
8.785
7.985
7.185
6.385
5.585
4.785
3.985
3.185
2.385
1.585
0.785
-0.015
10
20
30
40
Cycles
Sample FPKM Values
WT+DMBA 186
DNIIR+DMBA 721
DNIIR-28 + DMBA
WT + DMBA

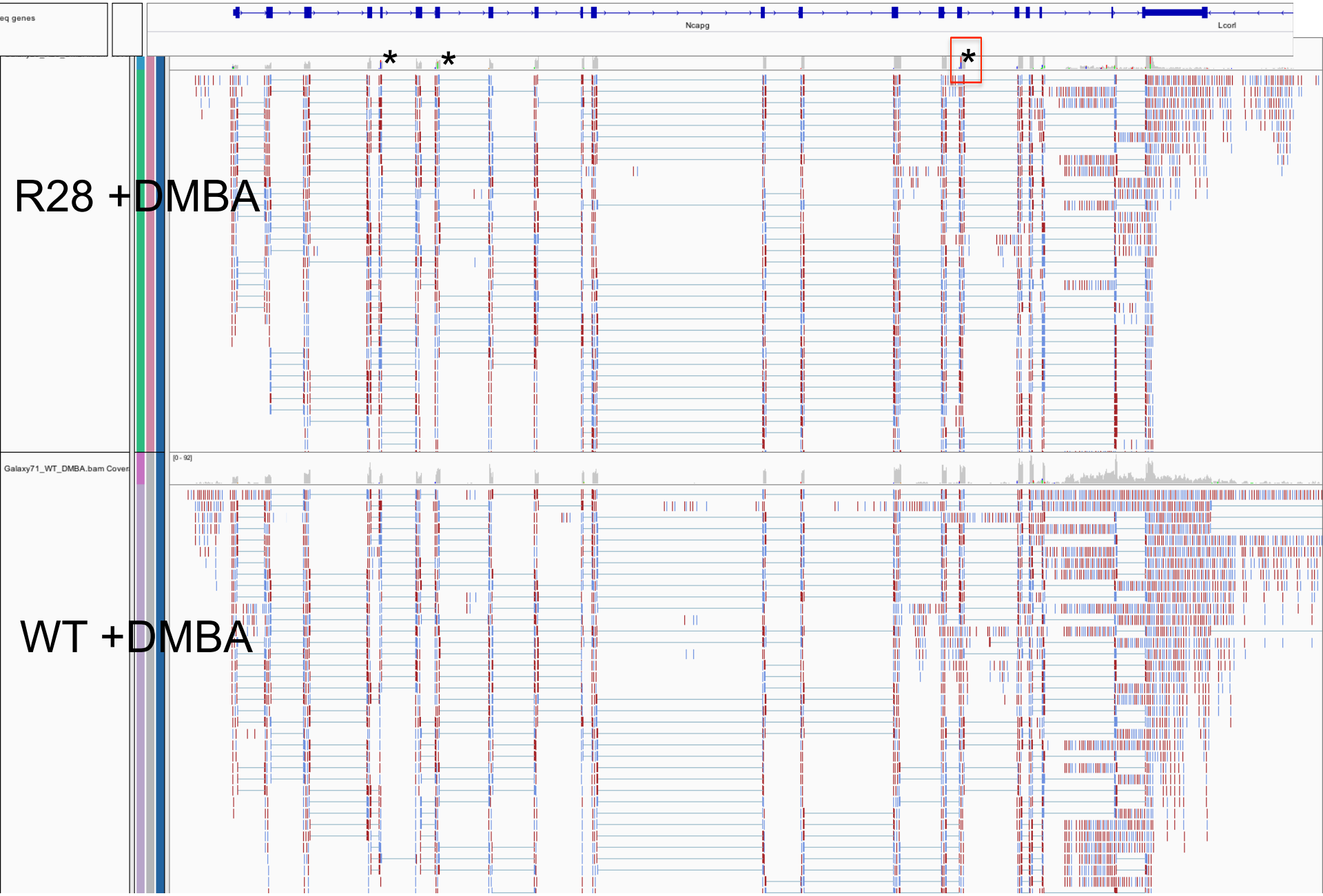
Lipase Maturation Factor 1 (Lmf1)



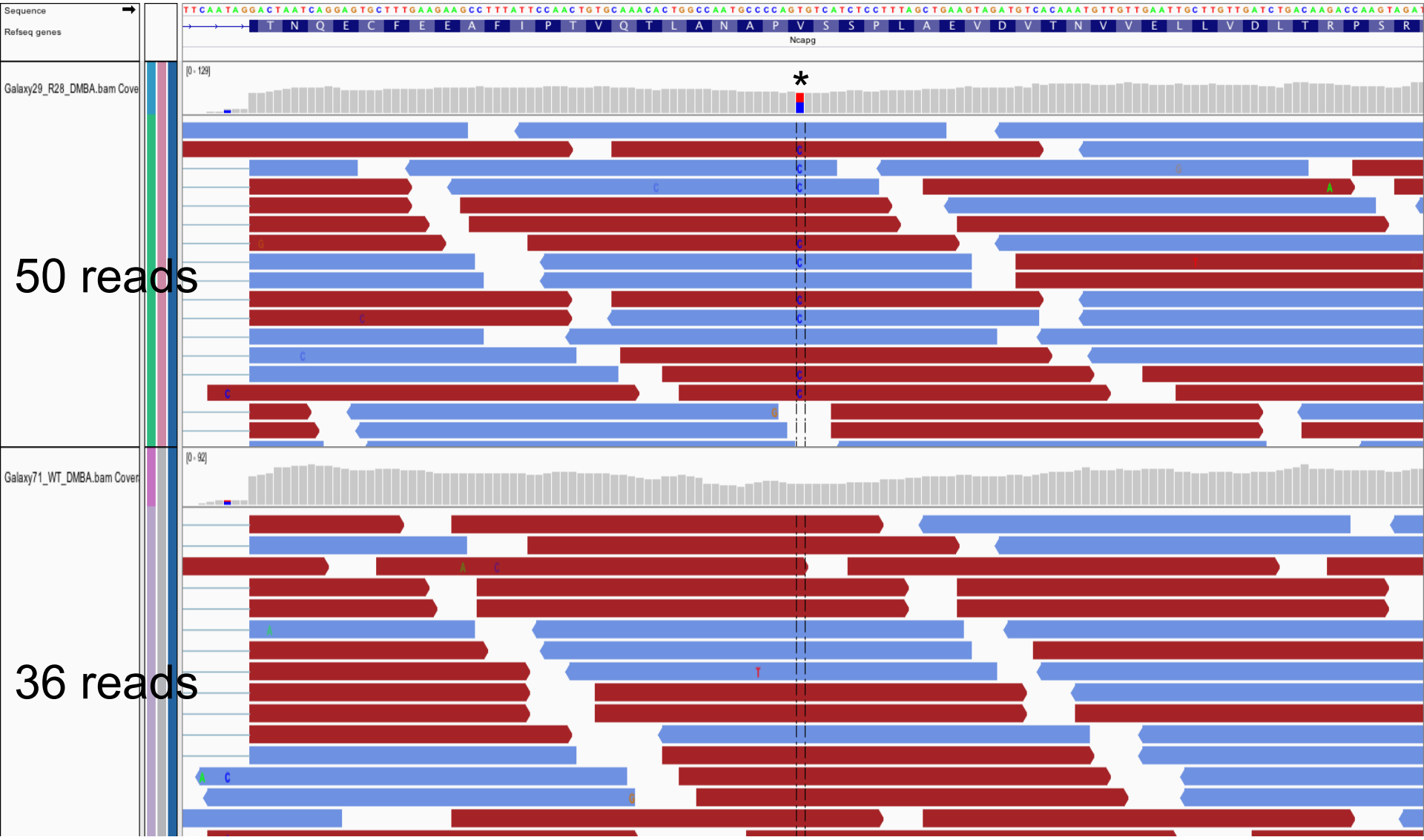
Lysophosphatidic acid receptor 3



Ncapg: Non-SMC condensin I complex, subunit G



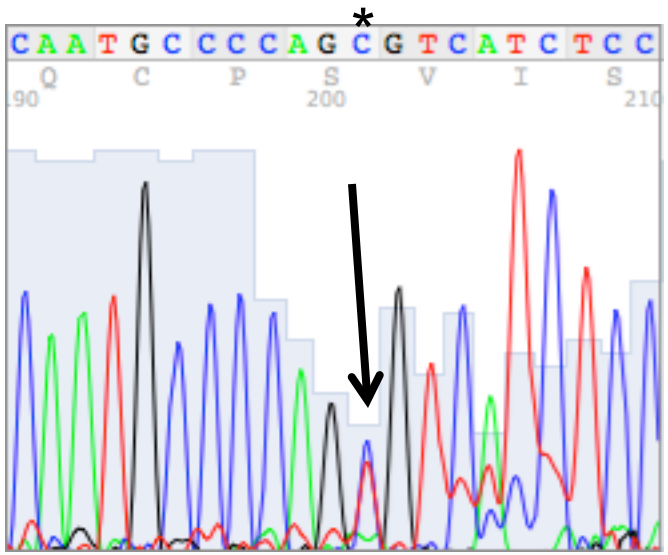
Exon 16 of Ncapg



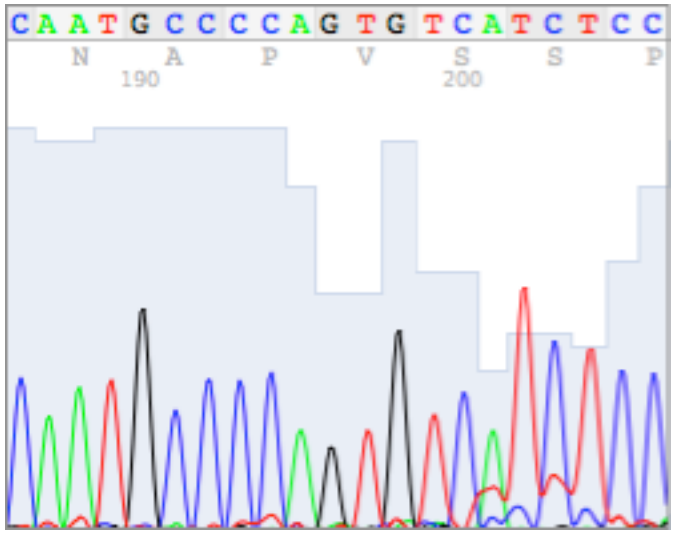
T-C mutation resulting in a Val-Ala change in the protein

Sequence Confirmation of Ncapg mutation

DNIIR

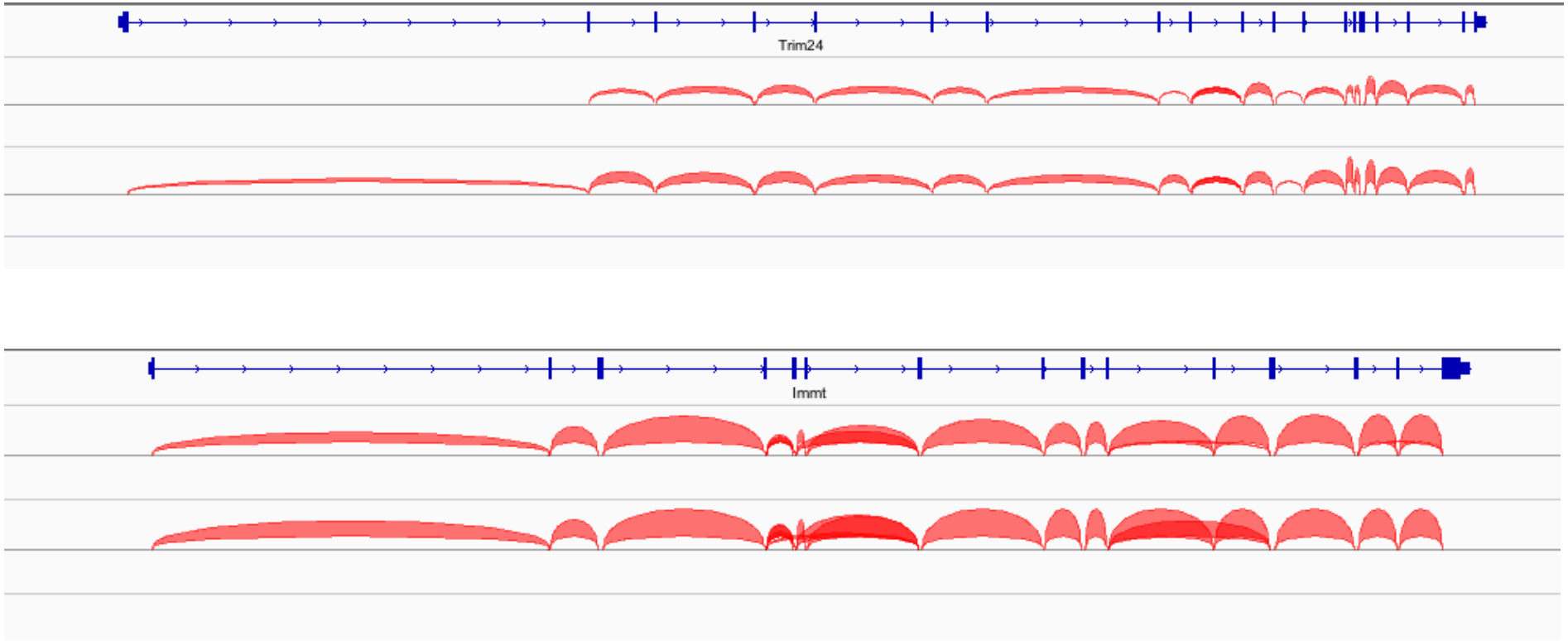


WT

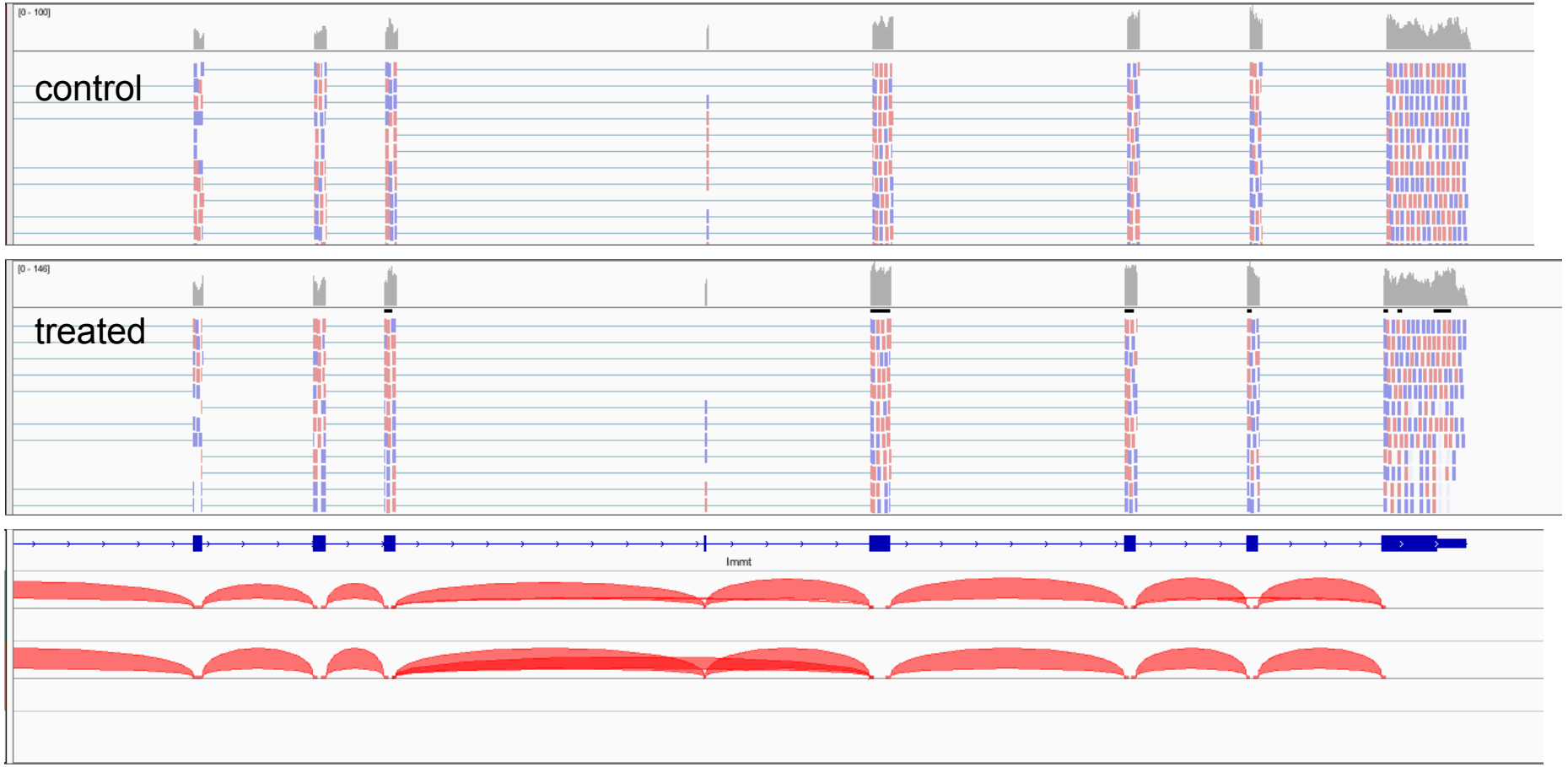


T>C mutation resulting in an Ala>Val change at position aa784 in the protein. The other mutations were a polymorphic T>C change at aa242 and an A>G change at aa347 resulting in a non-synonymous change from Arg>Lys.

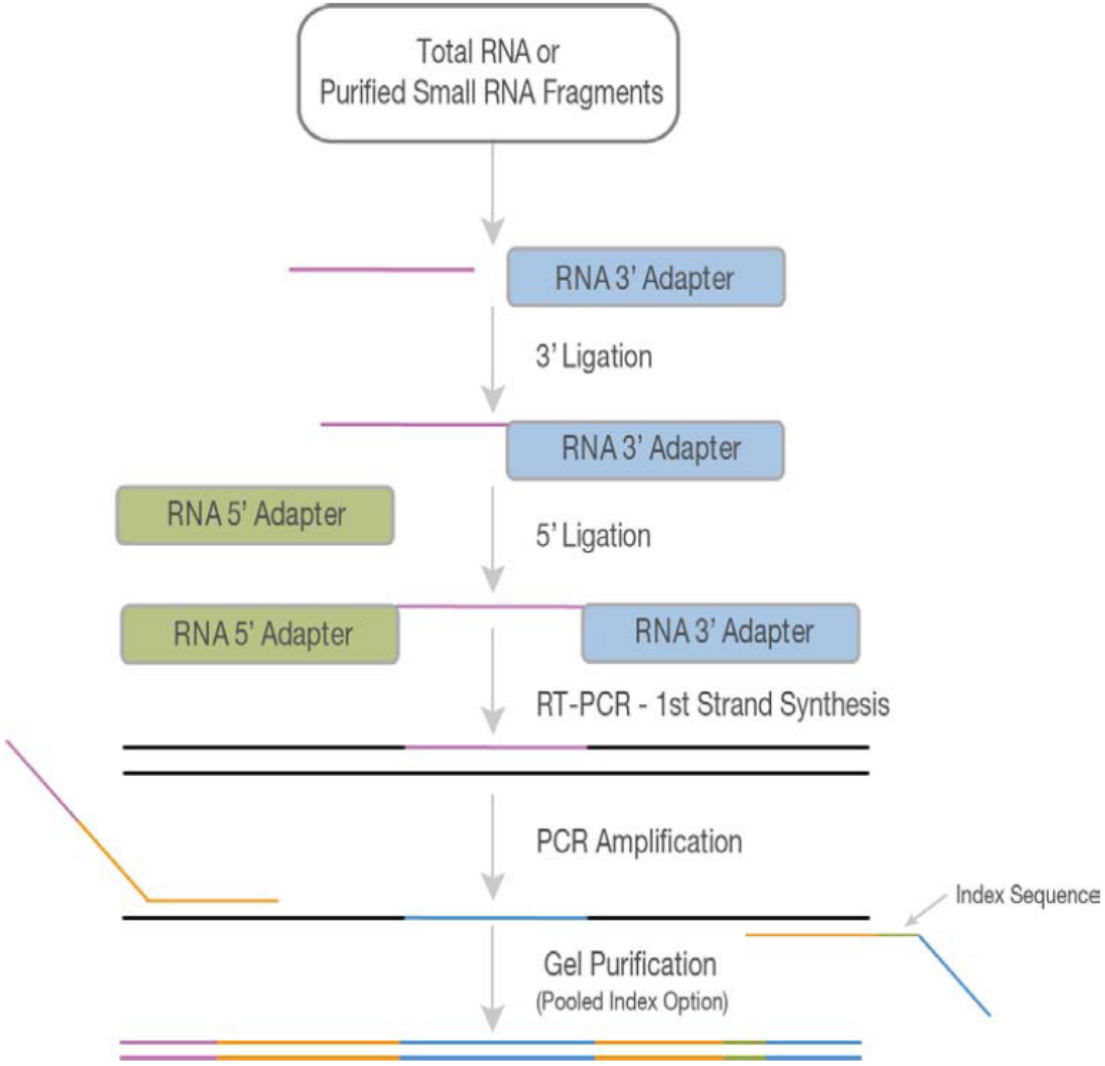
Alternative Exon Usage



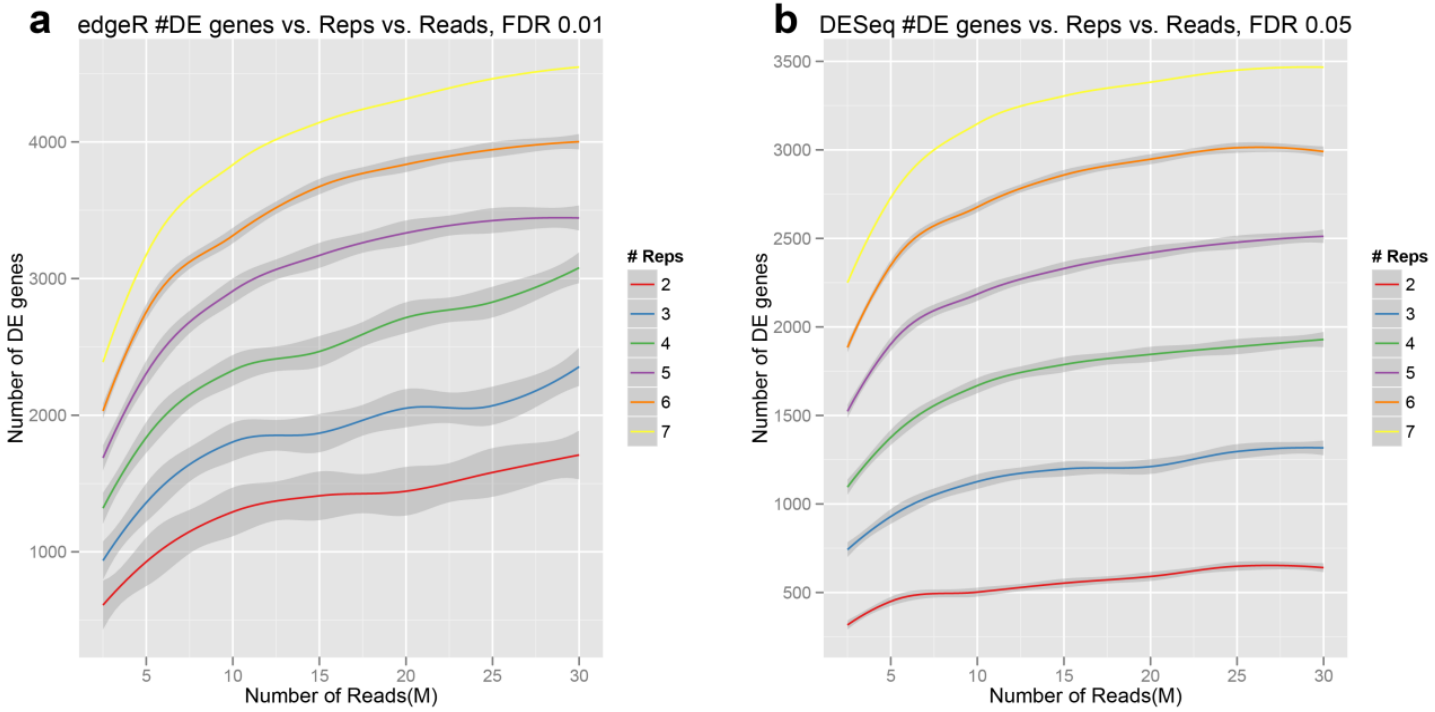
Alternative splicing



microRNA Seq

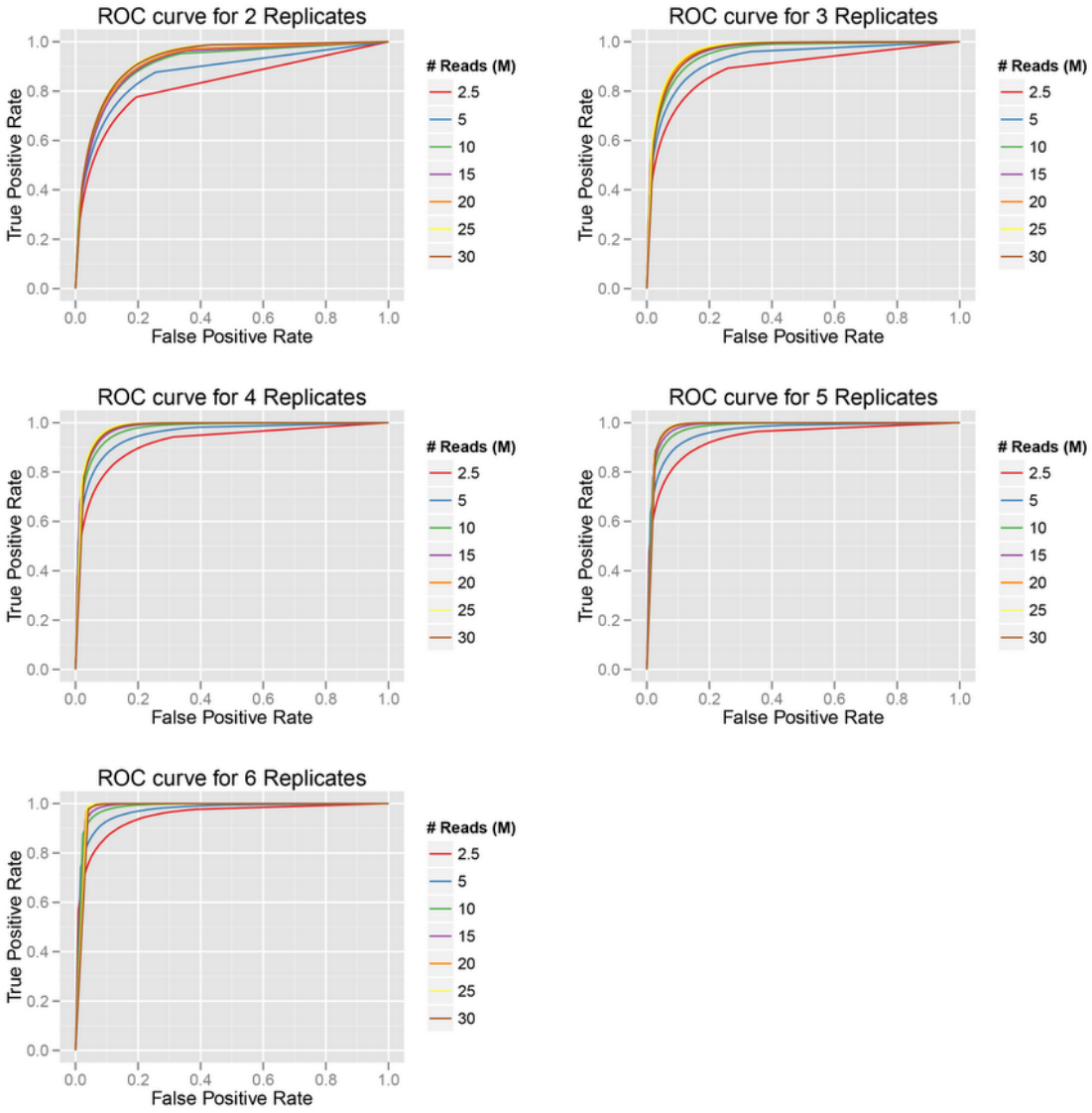


Sequencing Depth v. Biological Replicates



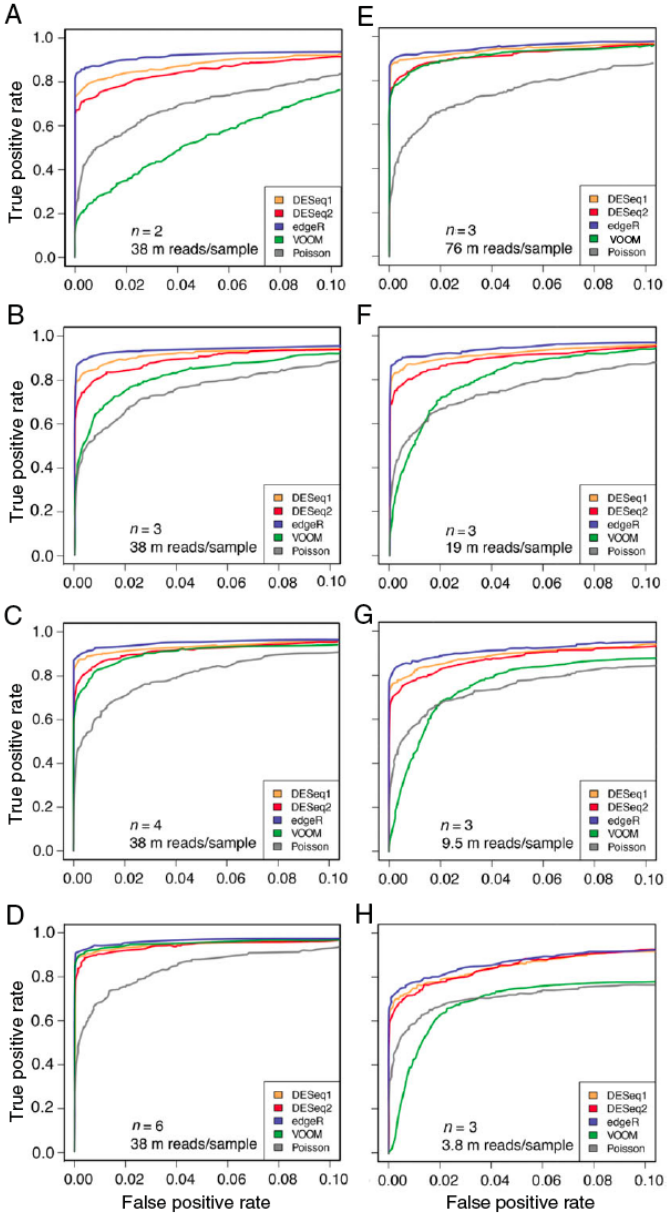
Liu, Y., Zhou, J., and White, KP. 2013. Bioinformatics 30:3:301

ROC curves with Replicates v. Depth



Liu, Y., Zhou, J., and White, KP. 2013. Bioinformatics 30:3:301

Replicates v. Depth with different DE packages



Williams AG, Thomas S., Wyman SL., Holloway AH. 2014.
Curr Prot Human Genet 11.13.1

Recommendations

1. Biological replication is important
   - The more you can afford the better your data

2. Increasing read depth does not substitute for increasing replicates

3. The type of experiment matters
   - Cell lines or inbred strains vs. outbred populations (humans for example).

Summary

- Several different platforms exist utilizing different technologies.
- Generate between 500 million to 600 Billion bases of sequence information per run.
- Several applications including Whole genome sequencing, Targeted genomic seq., ChIP-Seq and mRNA-Seq, among others.
- Data files are very large ≥1Tb of information.
- Personalized medicine via genome sequencing is HERE.