

Sequencing Technologies and Applications

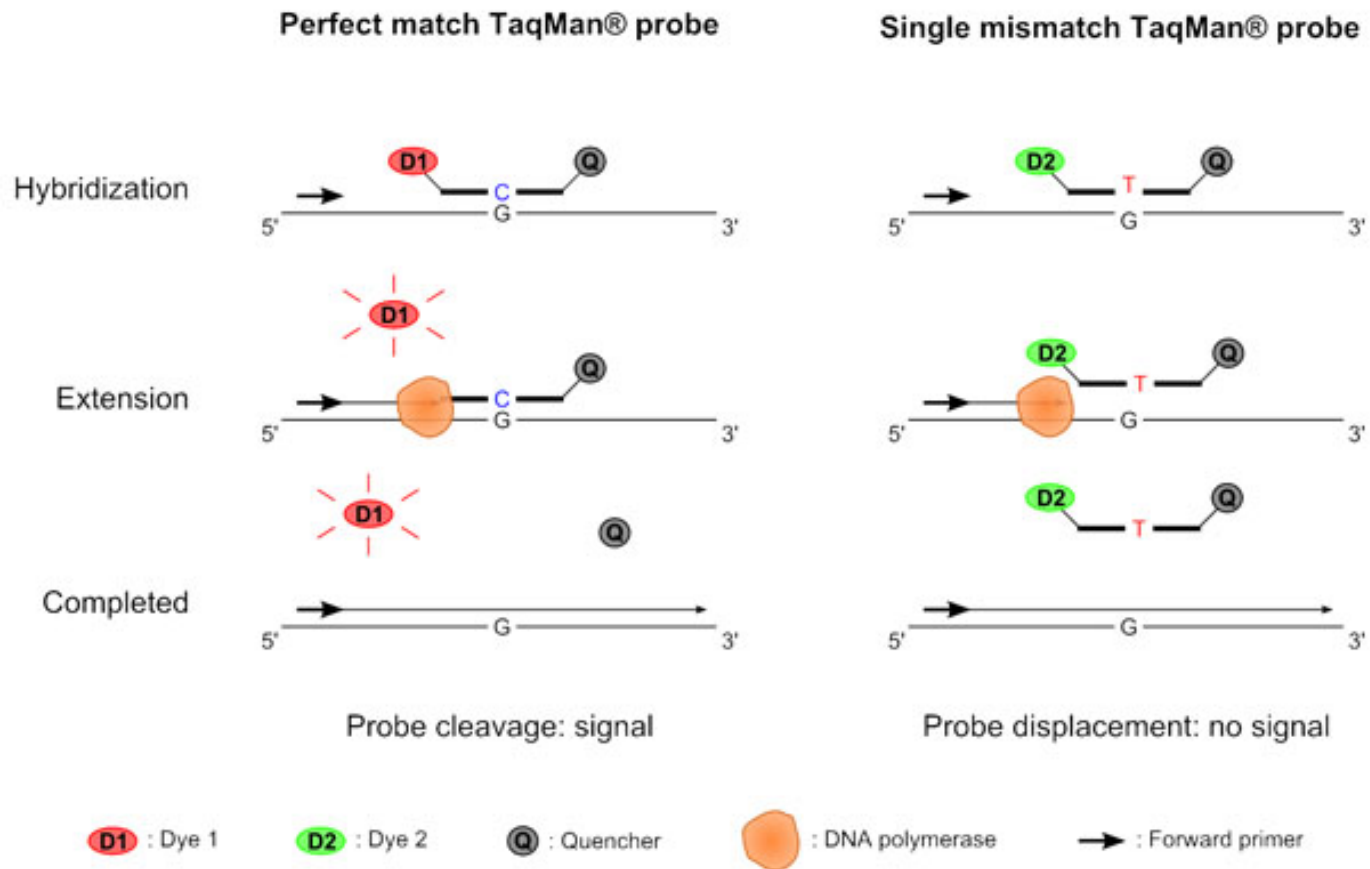
Michael R. Crowley, Ph.D.
Department of Genetics
Heflin Center for Genomic Sciences

Genotyping Technologies

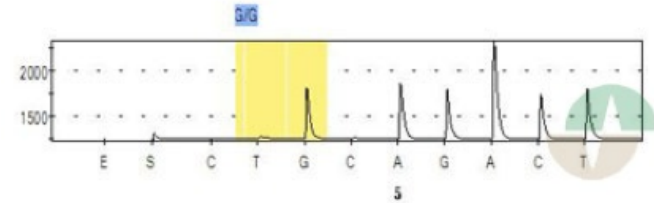
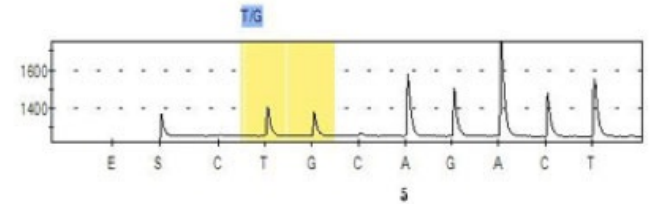
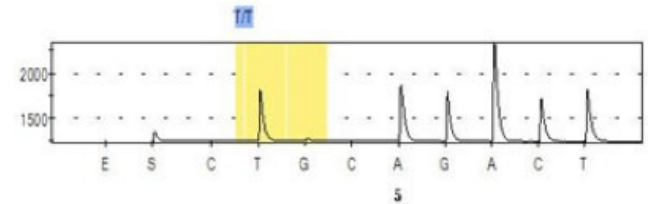
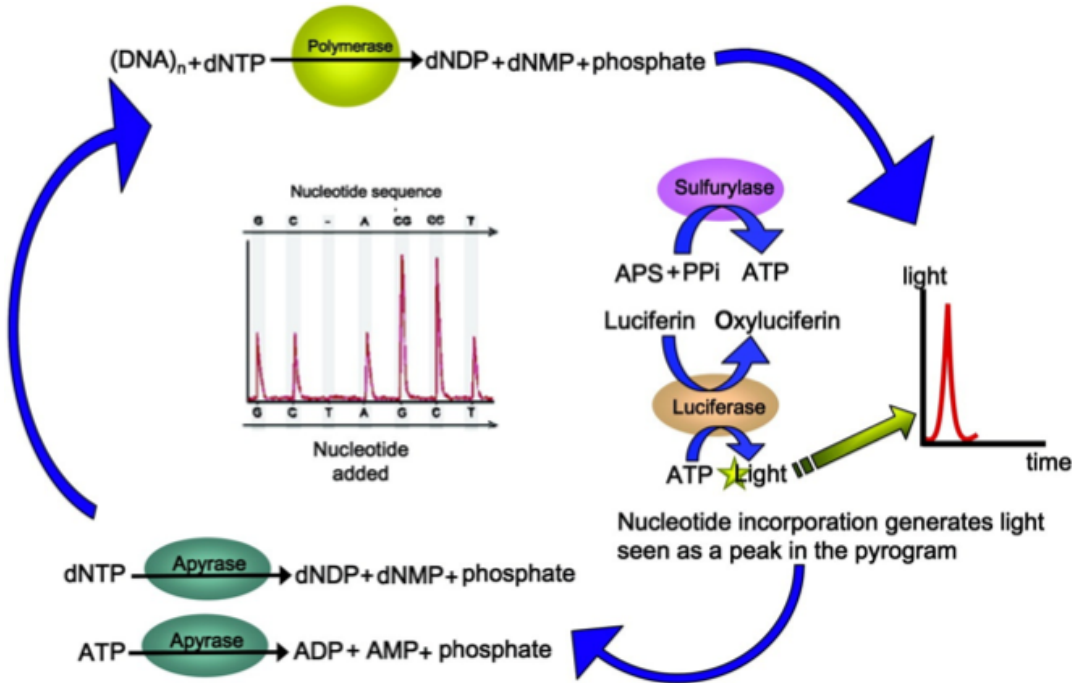
- The technology used depends on the number of variants and number of samples.
- Single to a few SNPs in a small population (≤ 960)
 - TaqMan
 - Pyrosequencing
 - Allele Specific PCR
 - Direct sequencing
 - High Resolution melting
- Intermediate # of SNPs with intermediate population size—no good options anymore.
- Large # of SNPs, large population
 - Infinium from Illumina (up to 5M SNPs per slide)

Single SNP Analysis

TaqMan Probes

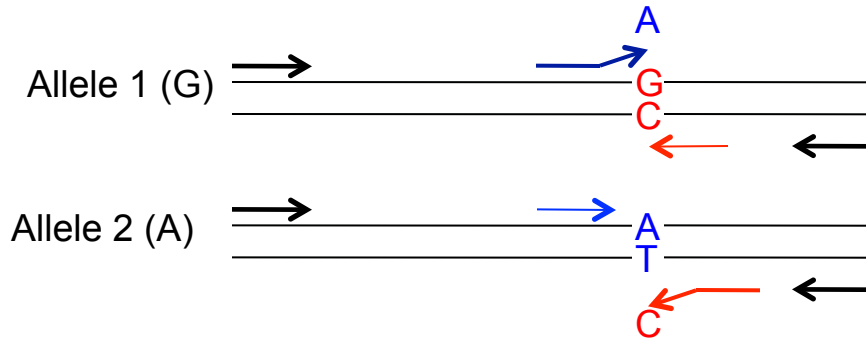


Pyrosequencing

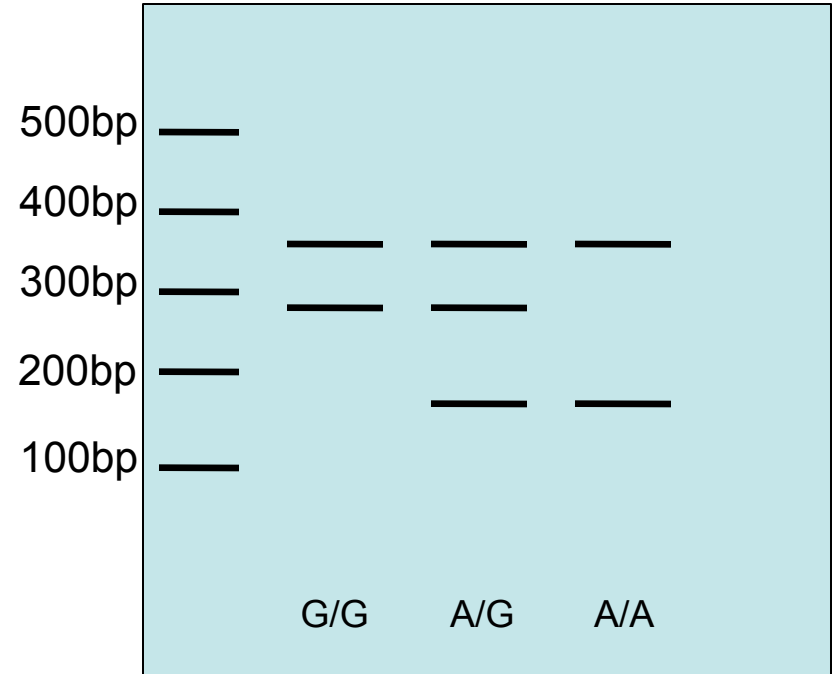
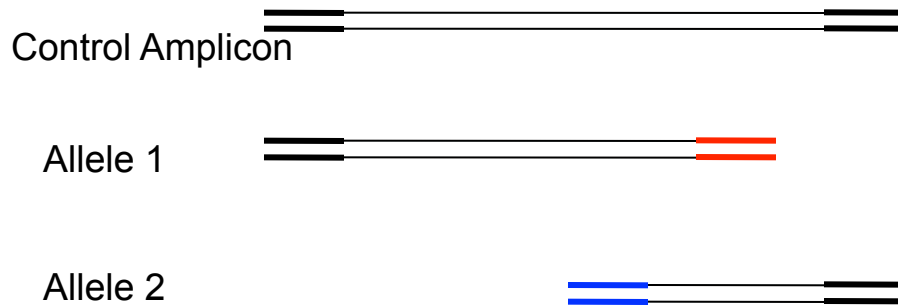


Allele Specific PCR

DNA Templates



PCR Products



Cycle Sequencing with Chain Termination

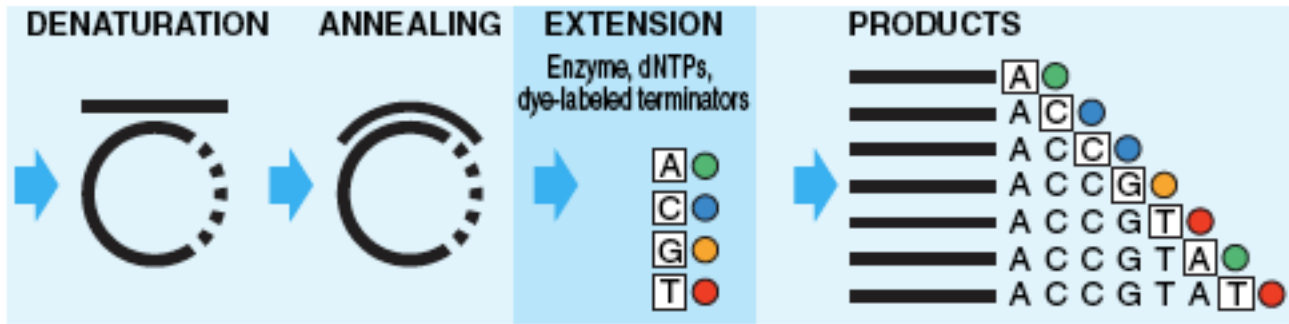
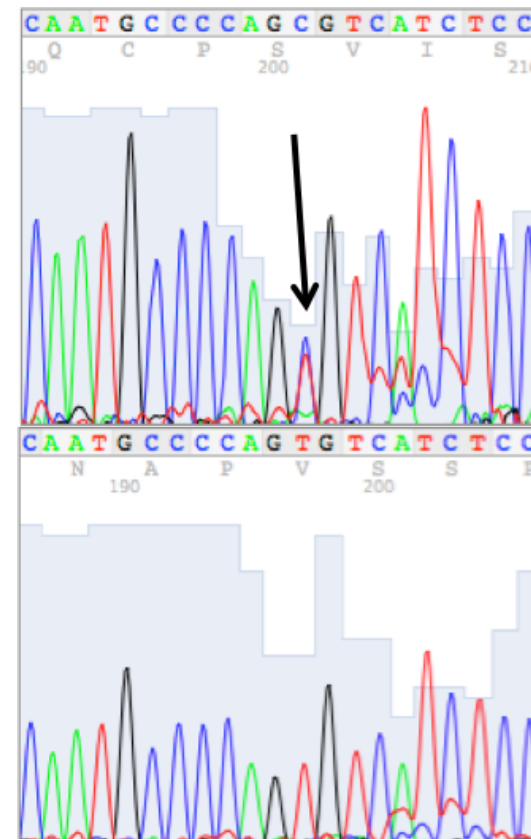
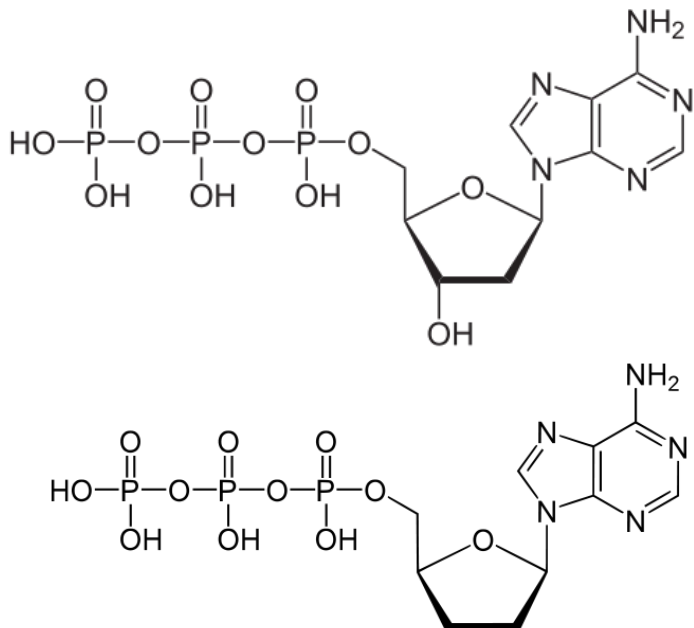
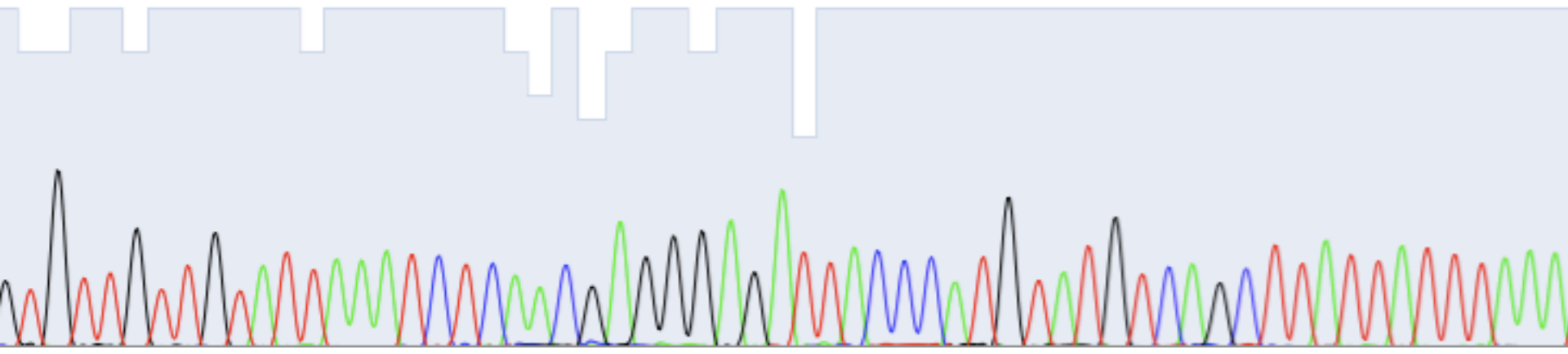


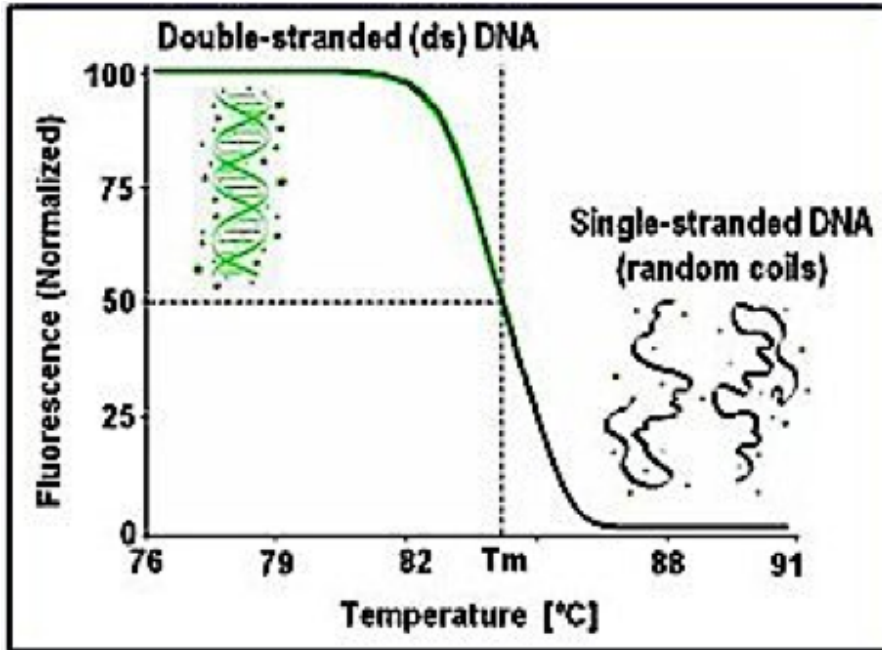
Figure 5 Diagram of dye terminator cycle sequencing



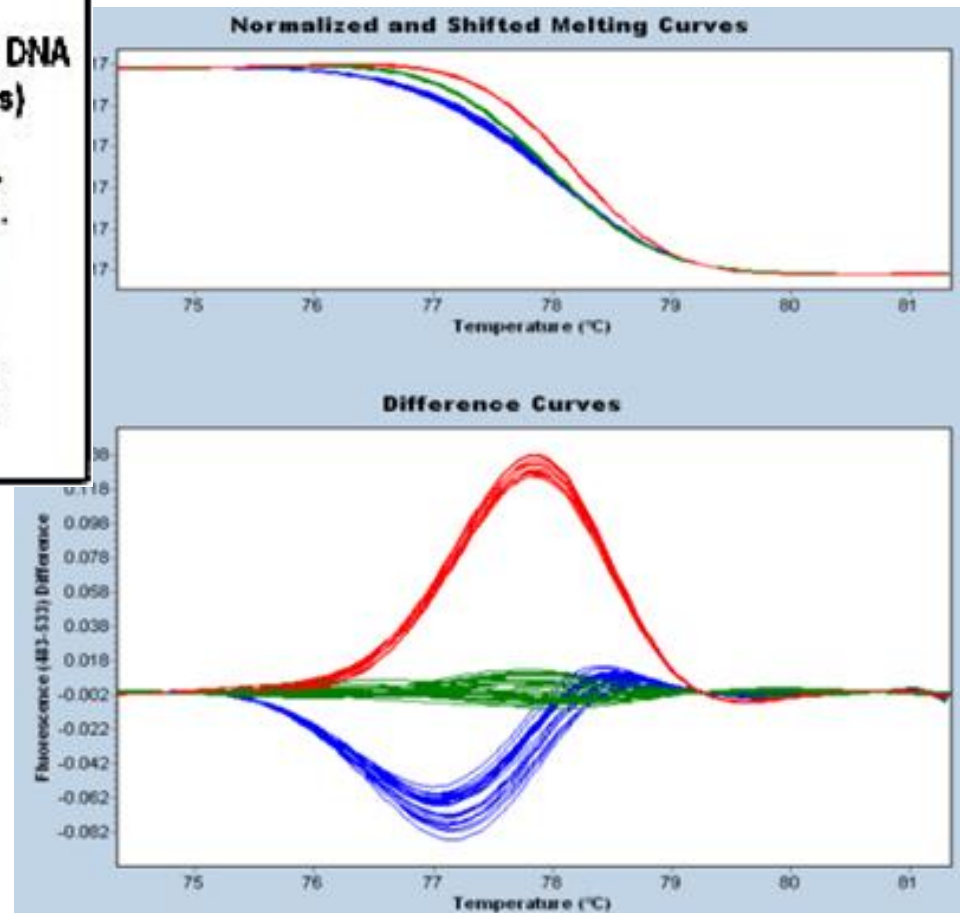
G T G T T G T T G T A T T A A A T C T C A A C G A G G G A G A T T A C C C A T G T A T G T C A G C T T A T T A T T T A A A
C C C I K S 150 Q R G 160 R L P M 170 Y V S 180 L L F 190 K



High Resolution Melting



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



The Next Generation

Illumina Platforms

- UAB Stem Cell Institute



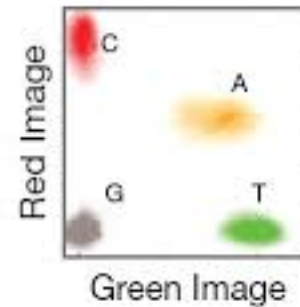
(

MiSeq
 Single flowcell
 36bp increments
 25-600bp per base seq
 Single read and Paired end reads



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



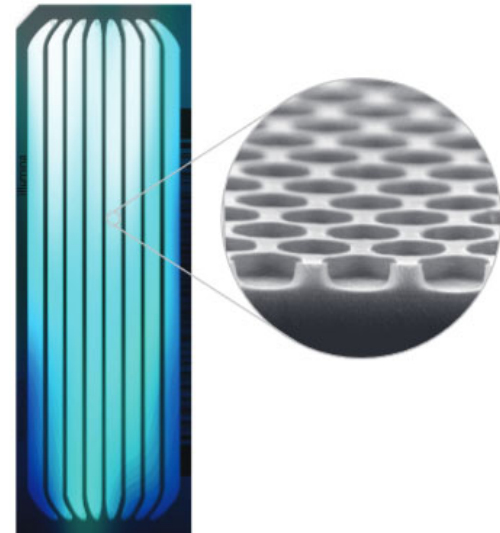
ONE SYSTEM, TWO OUTPUT MODES

High-Output	Mid-Output
Up to 120 Gb 400M clusters FF 1 x 75 bp to 2 x 150 bp	Up to 40 Gb 130M clusters FF 2 x 75 bp to 2 x 150 bp
30x genome	2-3 exomes
6-12 exomes RNA-Seq	2-4 samples RNA-Seq
20 GEX profiles MSPT	6-36 panels

51 © 2018 Illumina Inc. All Rights Reserved.

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.

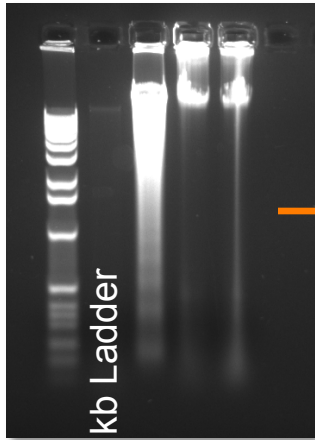
Illumina HiSeq X Ten (or 5)



Useful Next-Gen Terms

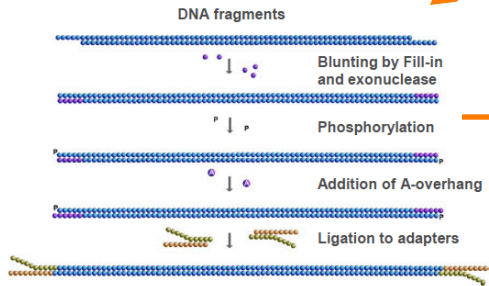
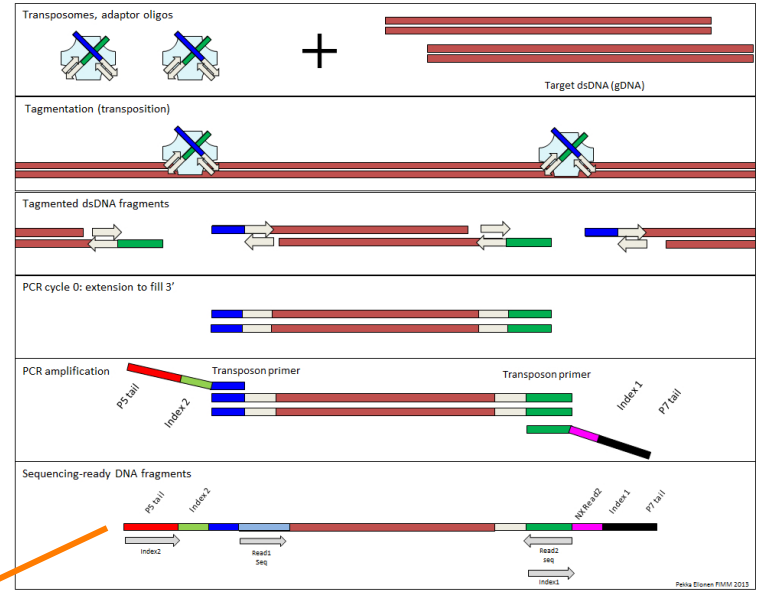
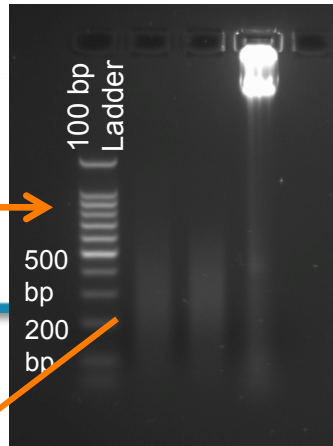
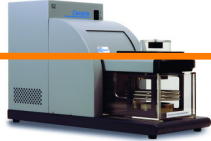
- Cluster
 - Individual island of DNA molecules representing a single, unique template
- Clusters Passing filter
 - Number of clusters able to be distinguished by the software as individuals
- Fastq
 - DNA Sequence file that is able to be read by downstream analysis applications
- Q-Score
 - A quality score based on the Phred score from Sanger Sequencing 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

DNA Library Prep and Flow cell Production



S-series

- Manual
- Single

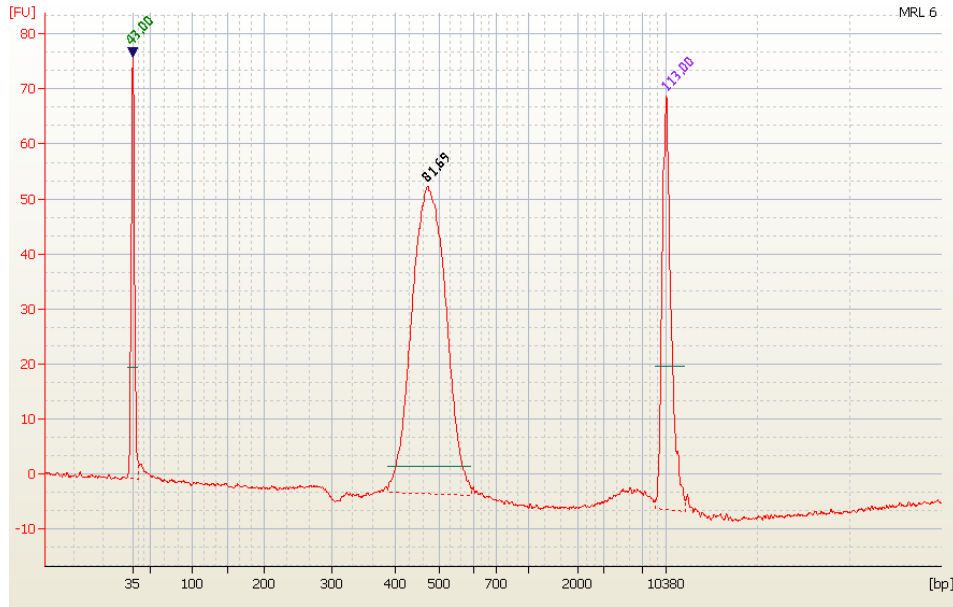


Illumina's Library Preparation Workflow

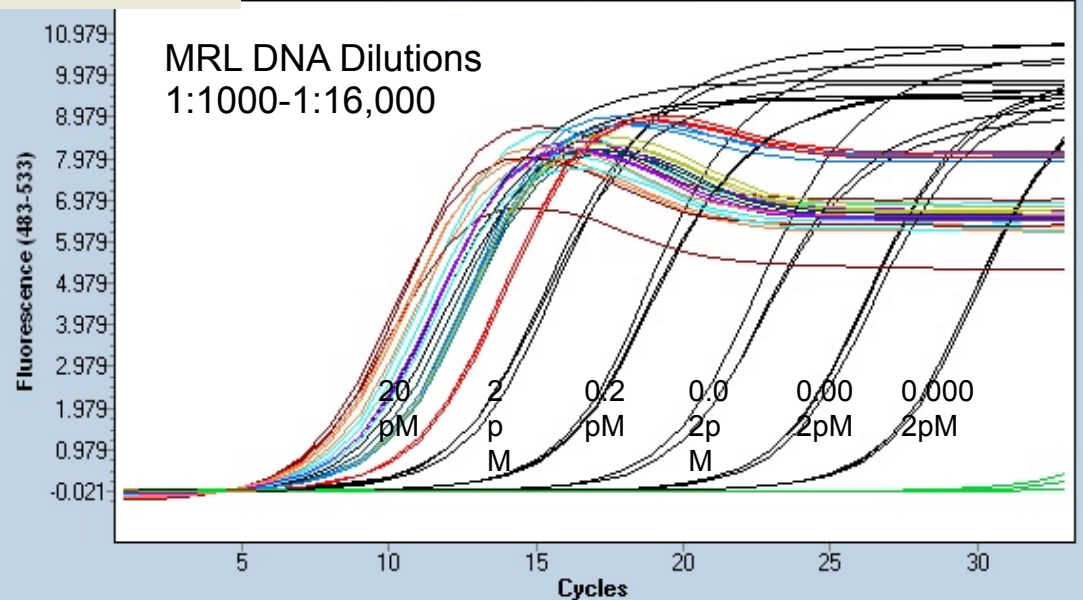


Version 3
HiSeq Flow Cell

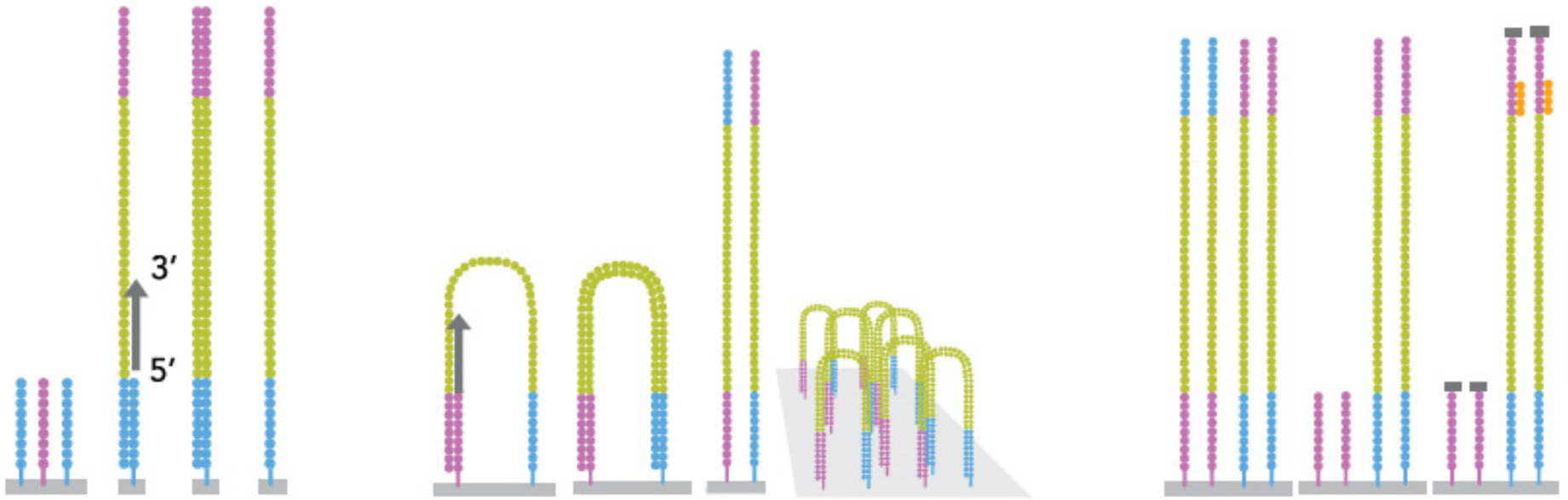
Library Assessment and Quantitation



Amplification Curves

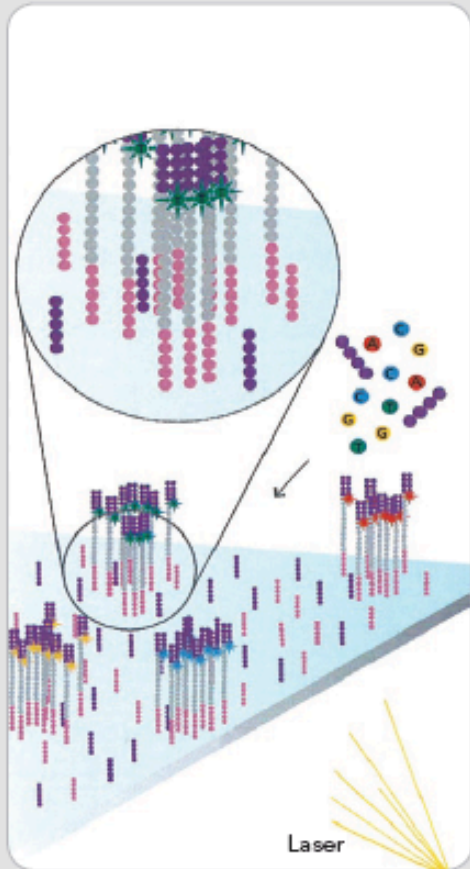


Illumina Cluster Generation



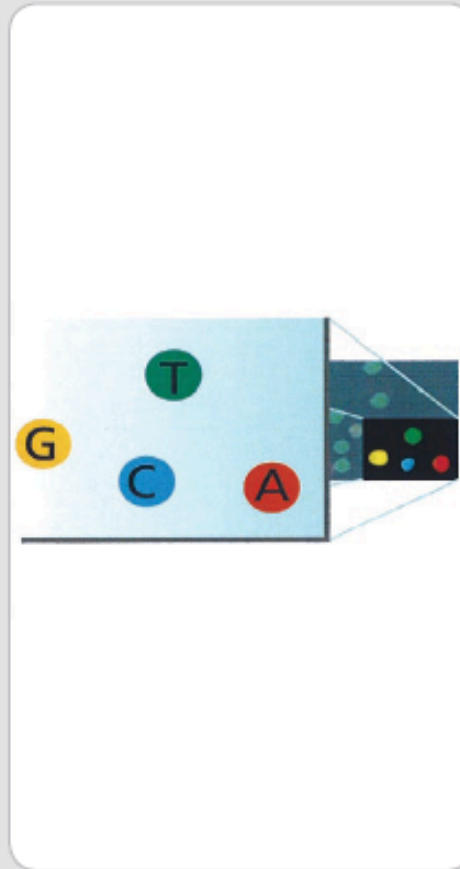
Occurs on the cBot or onboard the instrument

7. DETERMINE FIRST BASE



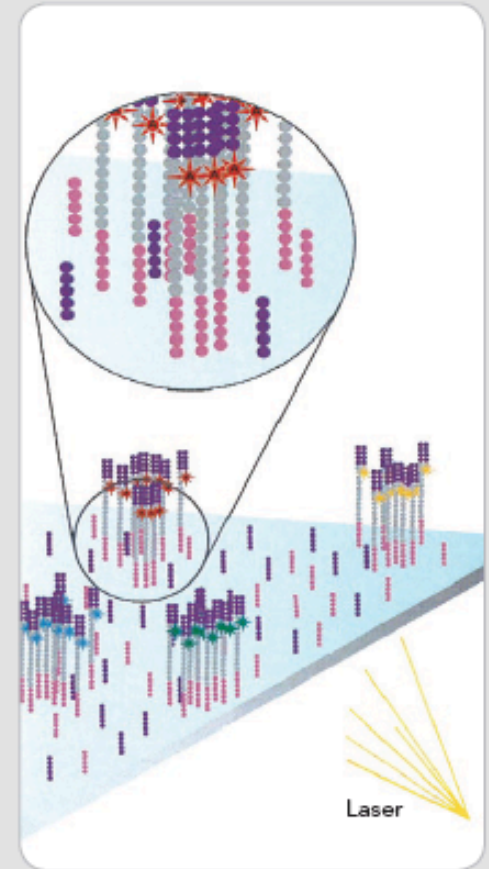
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE



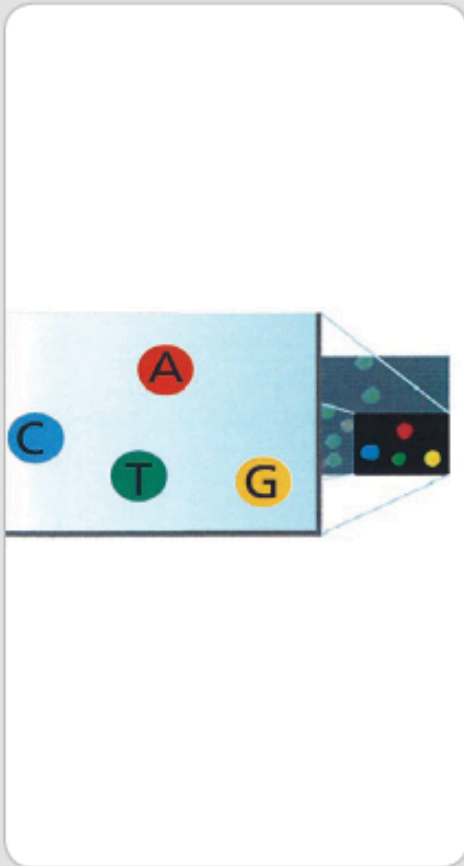
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

9. DETERMINE SECOND BASE



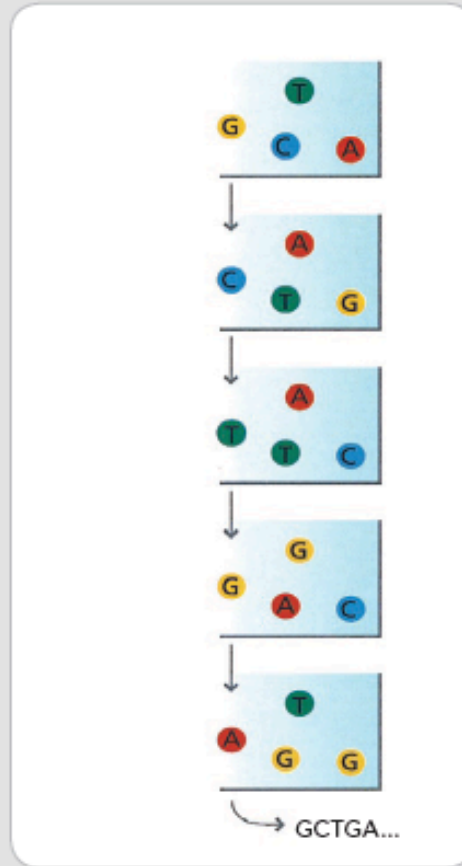
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

10. IMAGE SECOND CHEMISTRY CYCLE



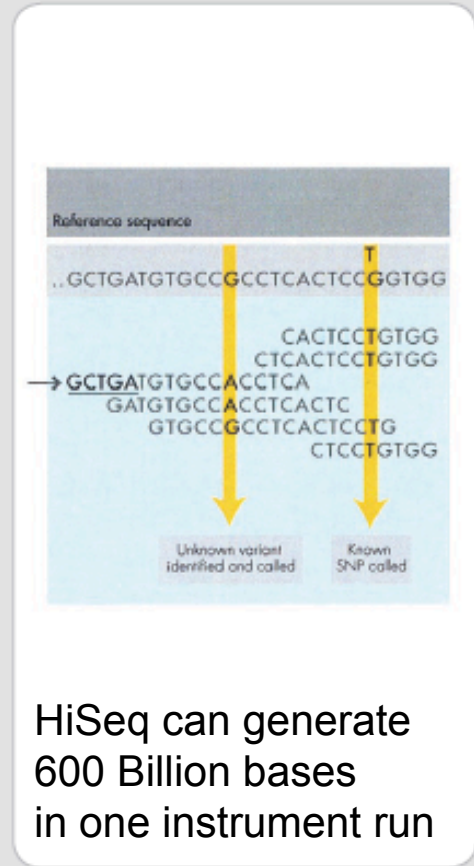
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

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

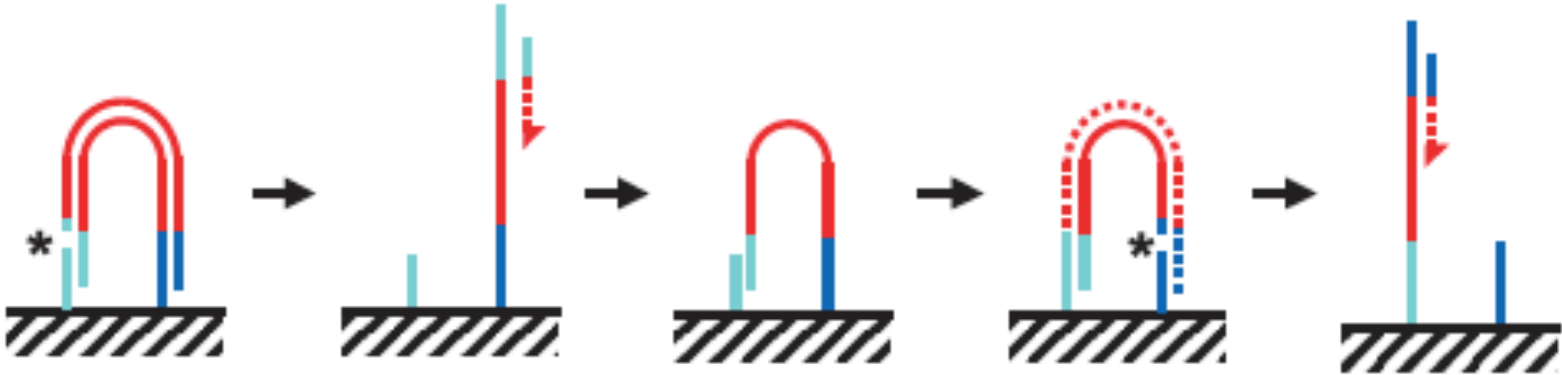


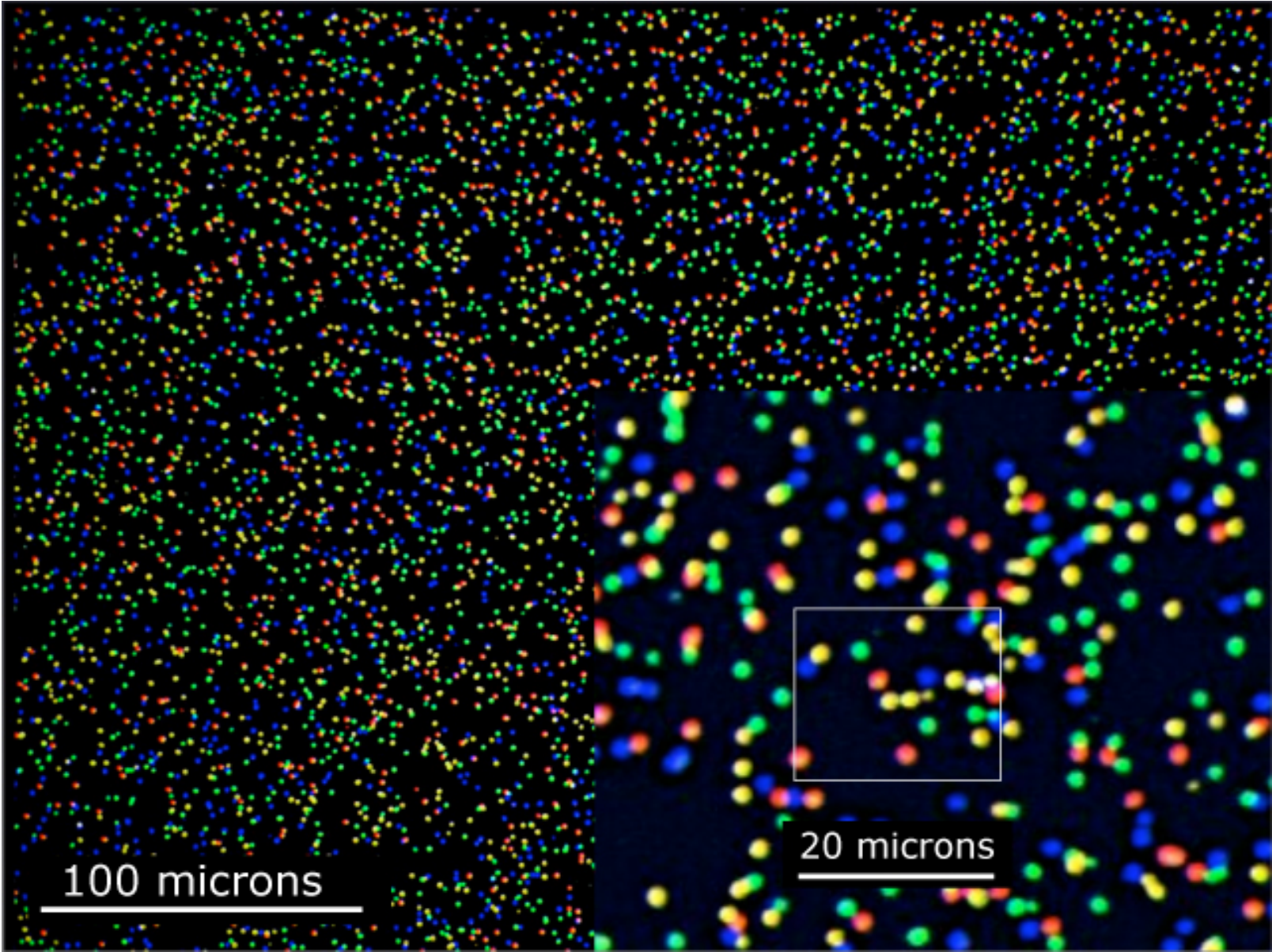
The data are aligned and compared to a reference, and sequencing differences are identified.

HiSeq can generate
600 Billion bases
in one instrument run

Single End vs. Paired End Sequencing

c





Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

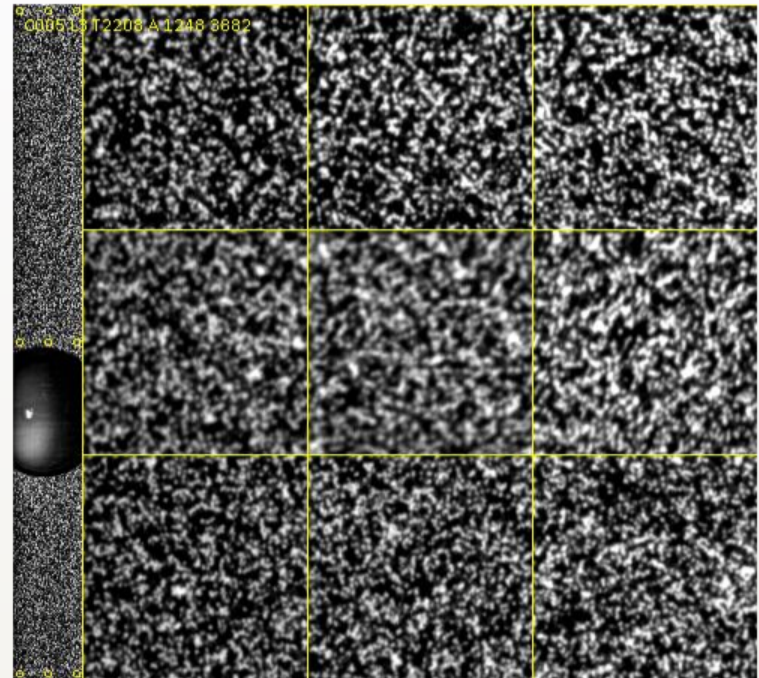
Browse **Refresh**

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	P90 T
33147	3	2207	7	99	Bottom	Middle	12/13/201...	1077	2909	932	22
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	36
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	36



Rows=41472 Disp=864 Sel=1 Filter

Sequencing Analysis Viewer

Run Folder: Y:\111208_SN372_0101_AD0JRMACXX

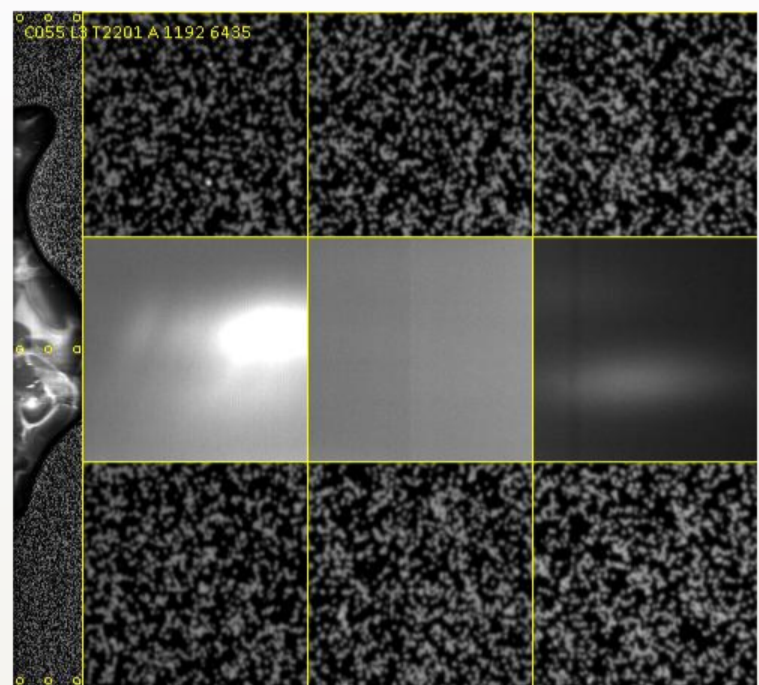
Browse **Refresh**

Analysis **Imaging** Summary Tile Status Controls

Cycle All Lane 3 Surface Bottom Swath Middle Section All

A C G T

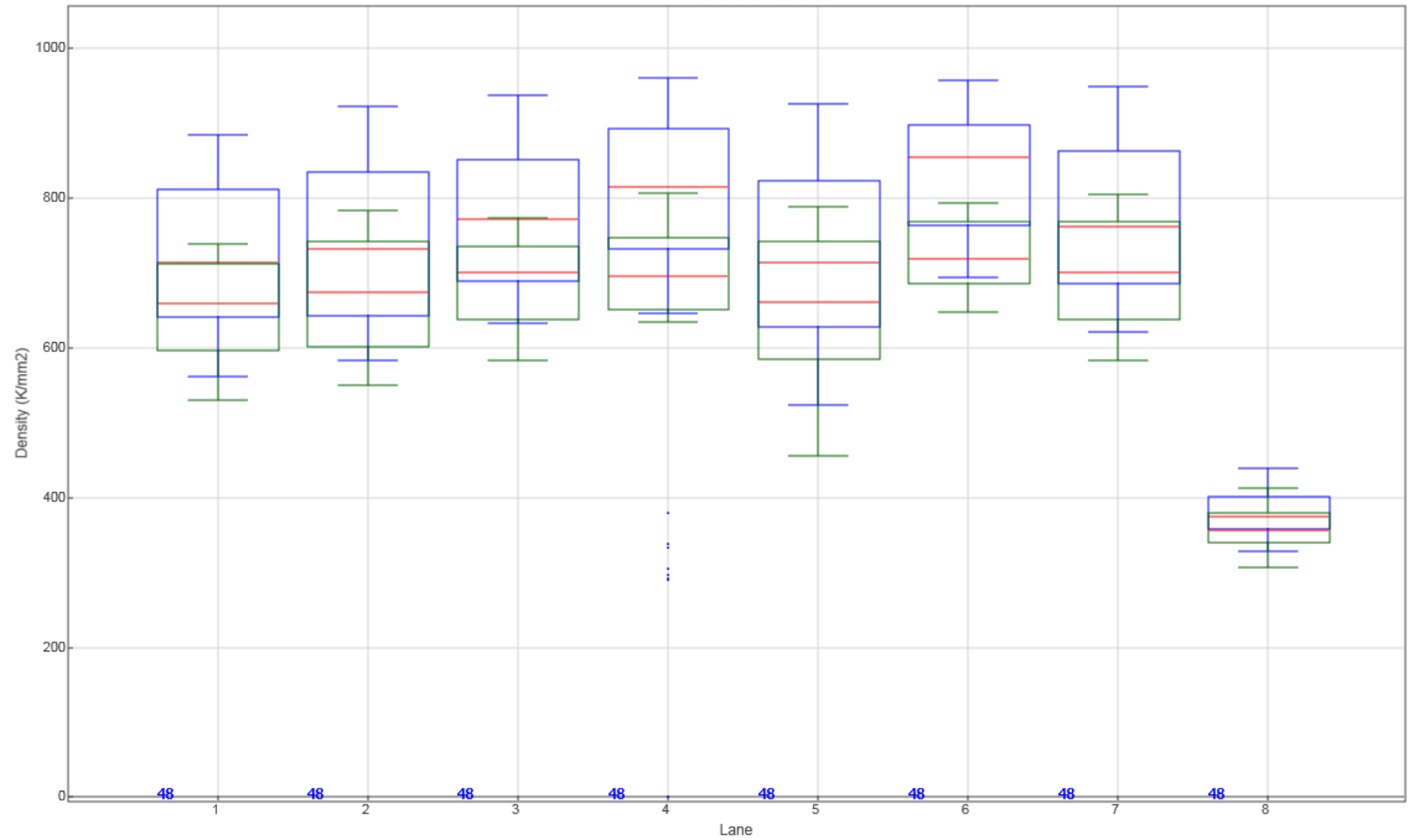
Index	Lane	Tile	Section	Cycle	Surface	Swath	Time	P90 A	P90 C	P90 G	P90 T
27905	3	2201	1	41	Bottom	Middle	12/10/201...	2368	3639	1103	26
27906	3	2201	1	42	Bottom	Middle	12/10/201...	2369	3824	1148	27
27907	3	2201	1	43	Bottom	Middle	12/10/201...	2361	3831	1136	27
27908	3	2201	1	44	Bottom	Middle	12/10/201...	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
27922	3	2201	1	58	Bottom	Middle	12/10/201...	0	0	0	0



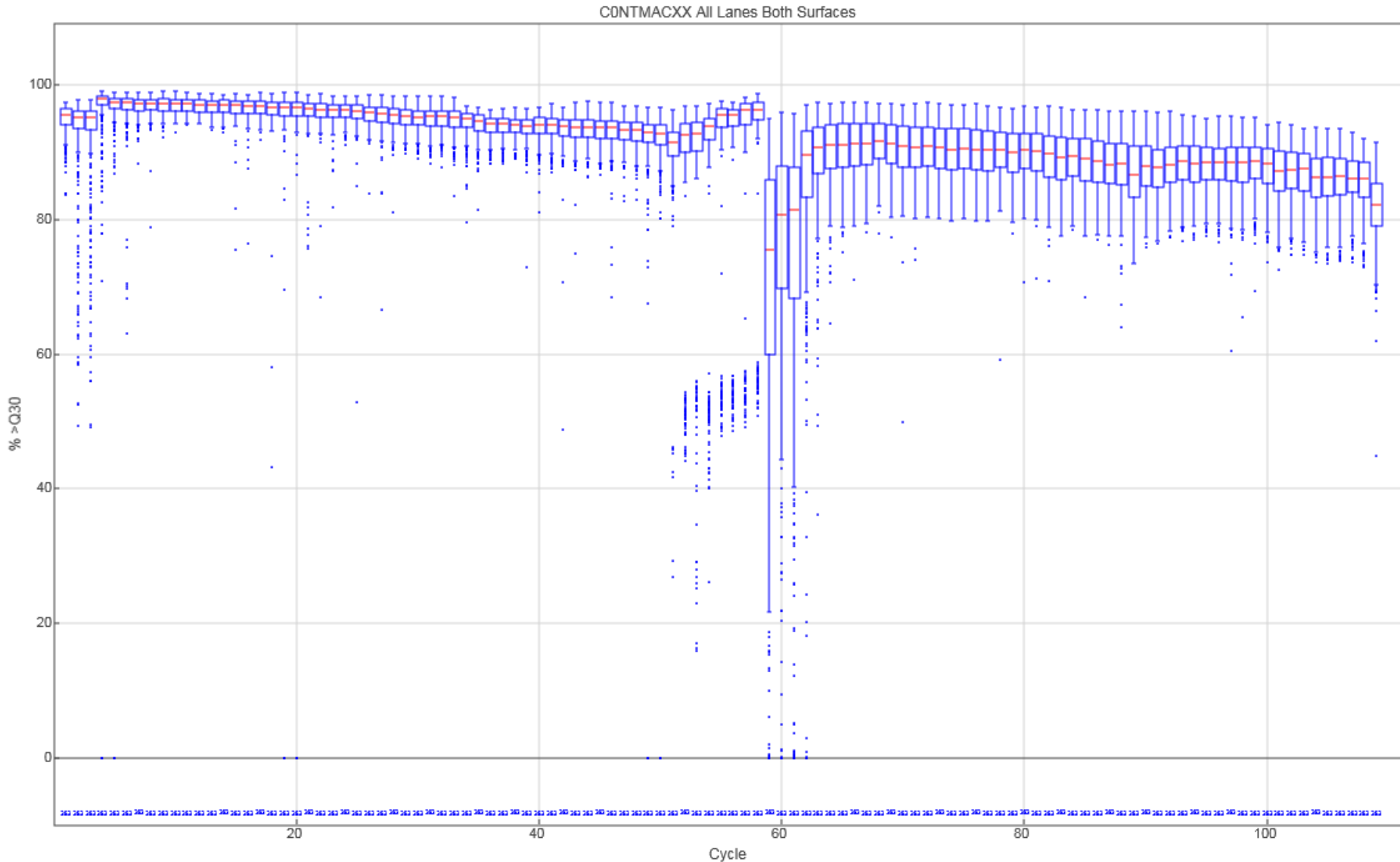
Rows=41472 Disp=864 Sel=1 Filter

Cluster Density

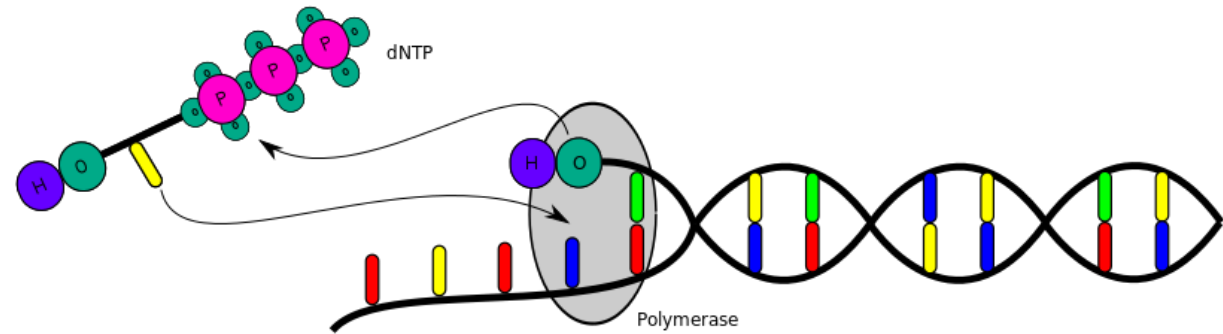
CONTMACXX Both Surfaces



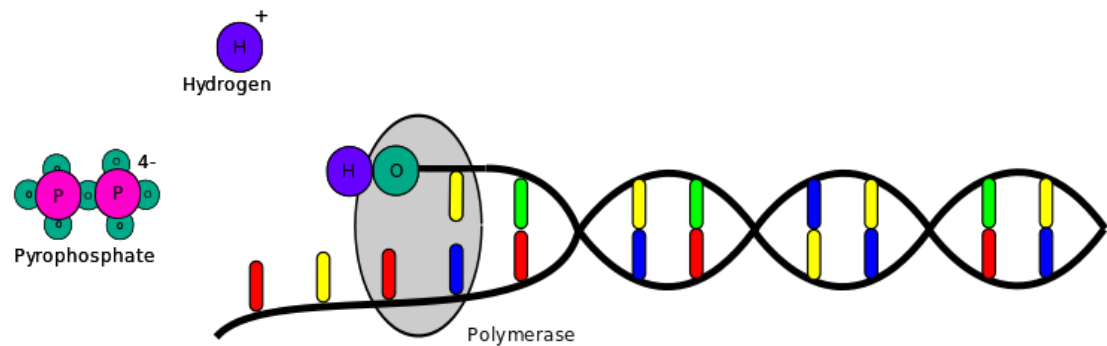
Percent Q30 Scores per cycle for all lanes and both surfaces



Ion Semiconductor Sequencing (aka ion Torrent or Proton)

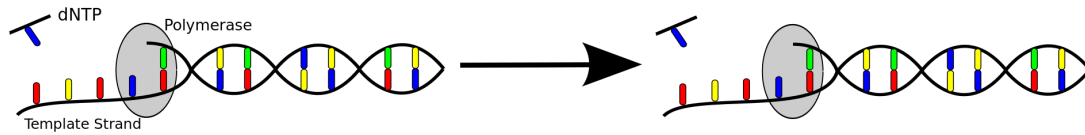


Polymerase integrates a nucleotide.

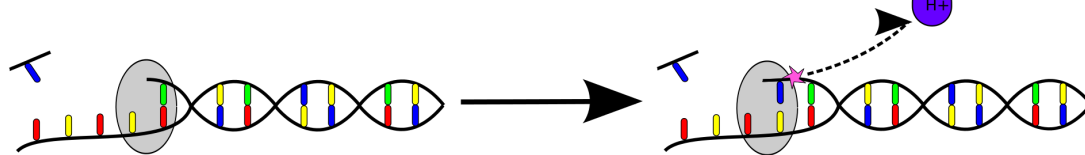


Hydrogen and pyrophosphate are released.

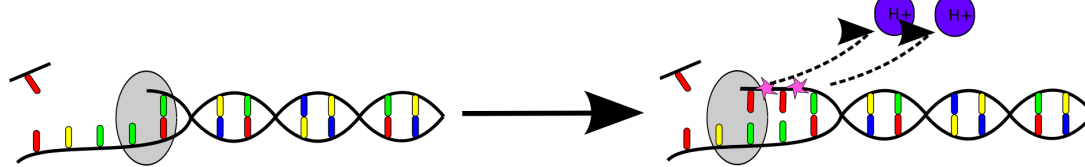
Ion Semiconductor Sequencing



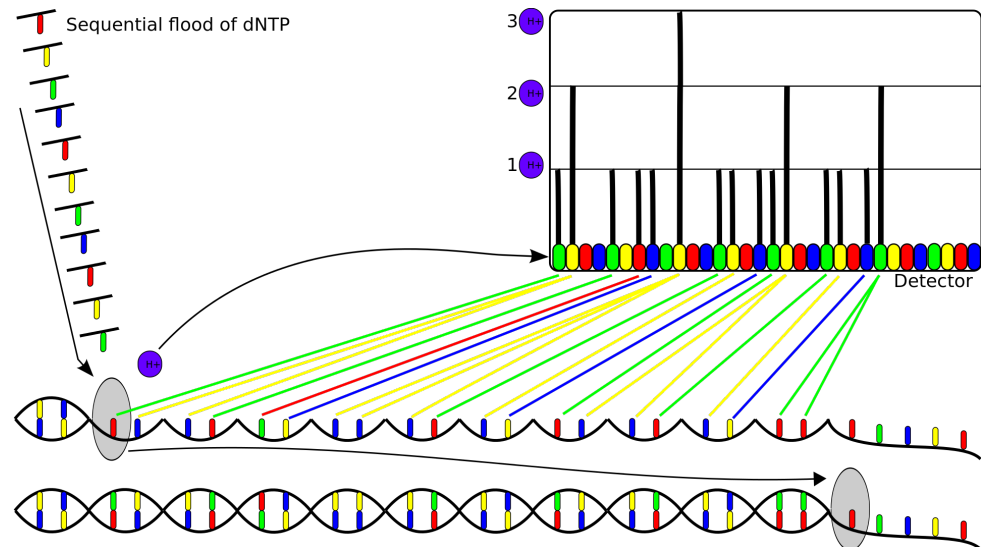
The nucleotide does not complement the template - no release of hydrogen.



The nucleotide complements the template - hydrogen is released.

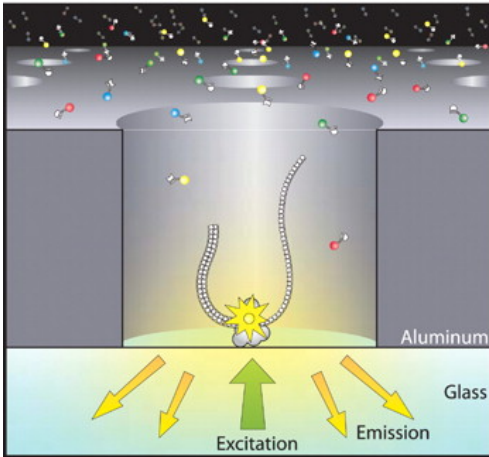


The nucleotide complements several bases in a row - multiple hydrogen ions are released.

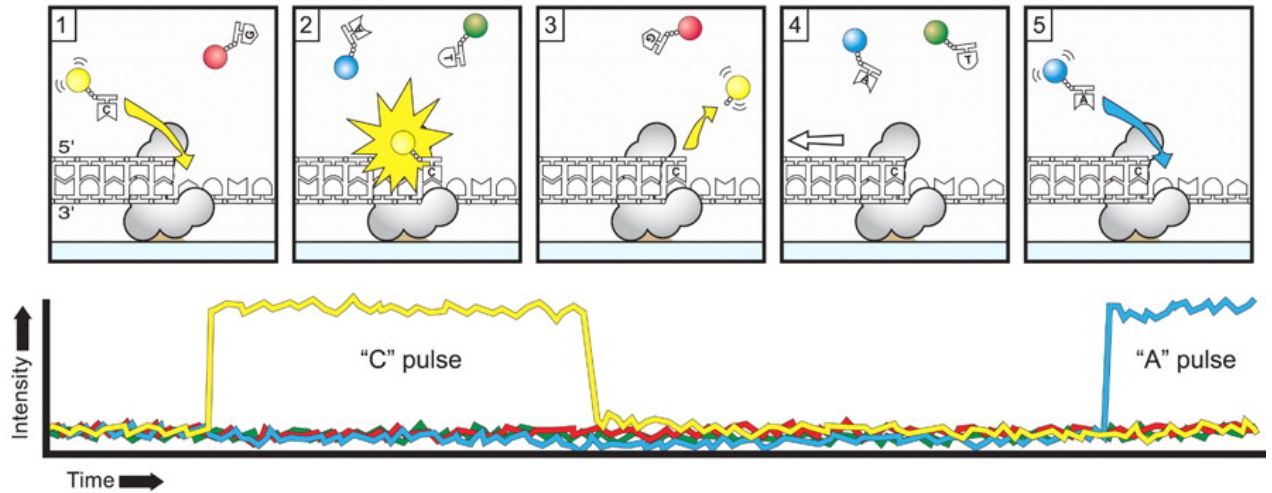


Pacific Biosciences Technology

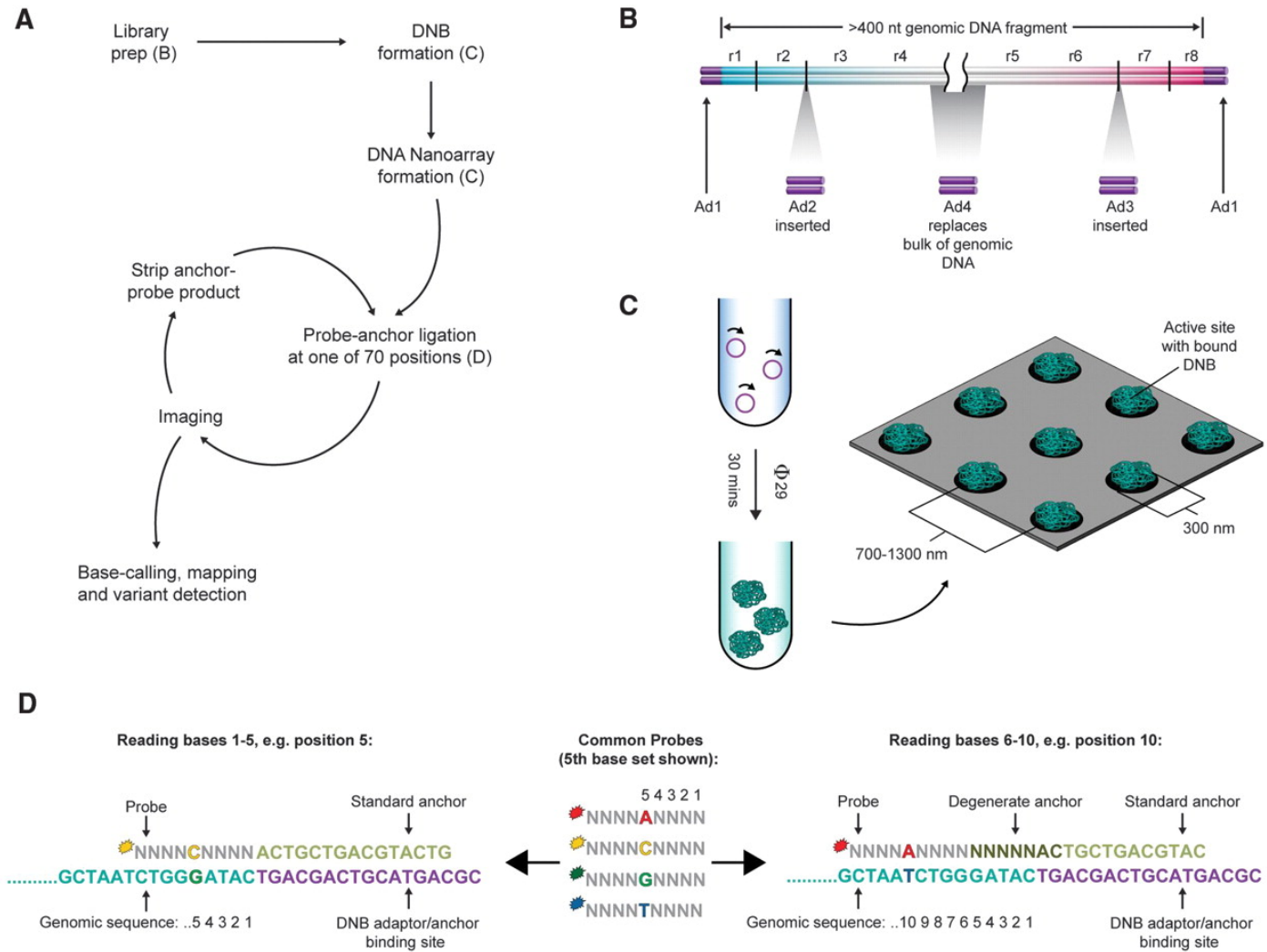
A



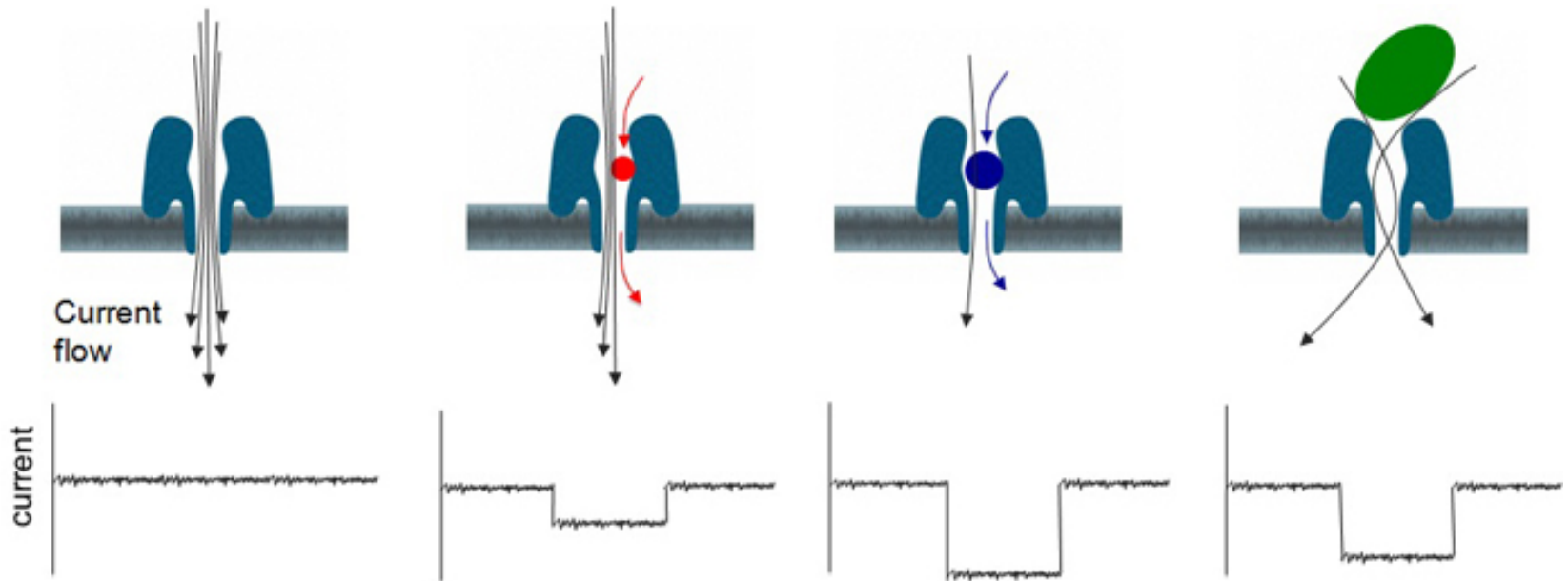
B



Complete Genomics Technology



Nanopores



Oxford Nanopore

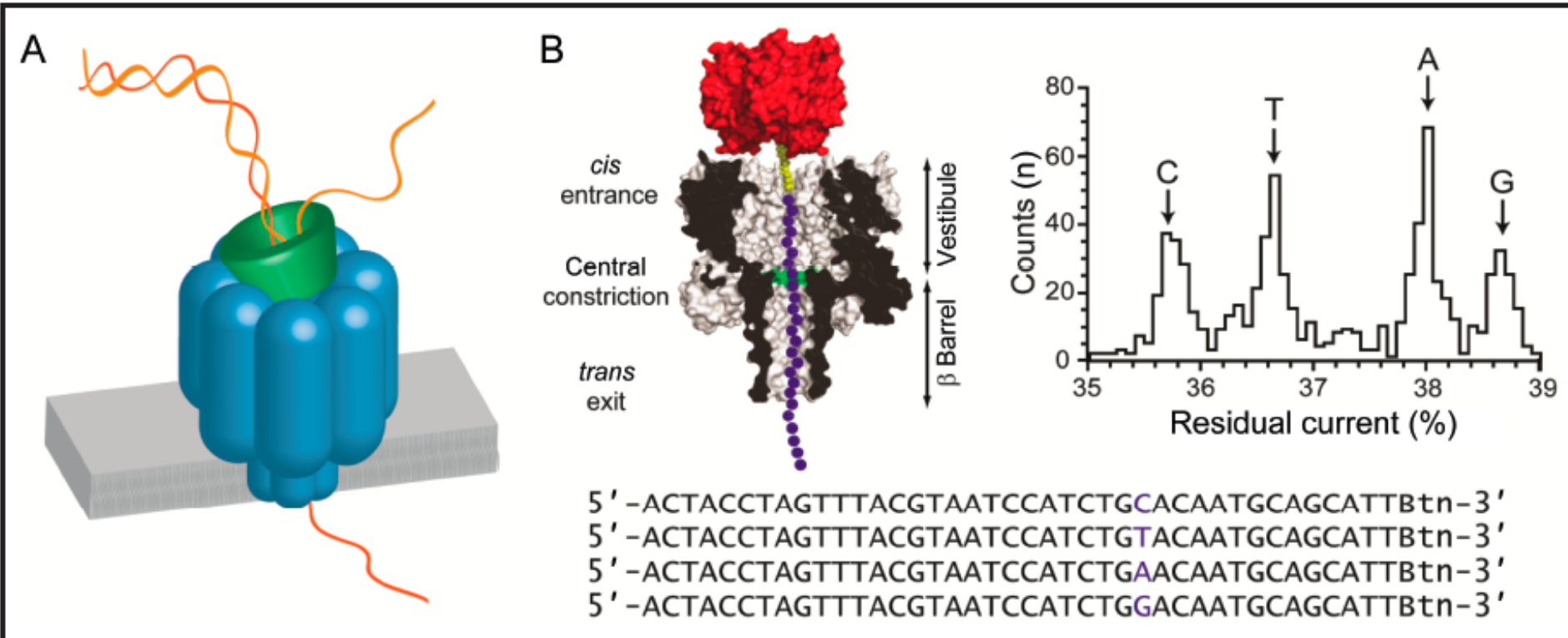
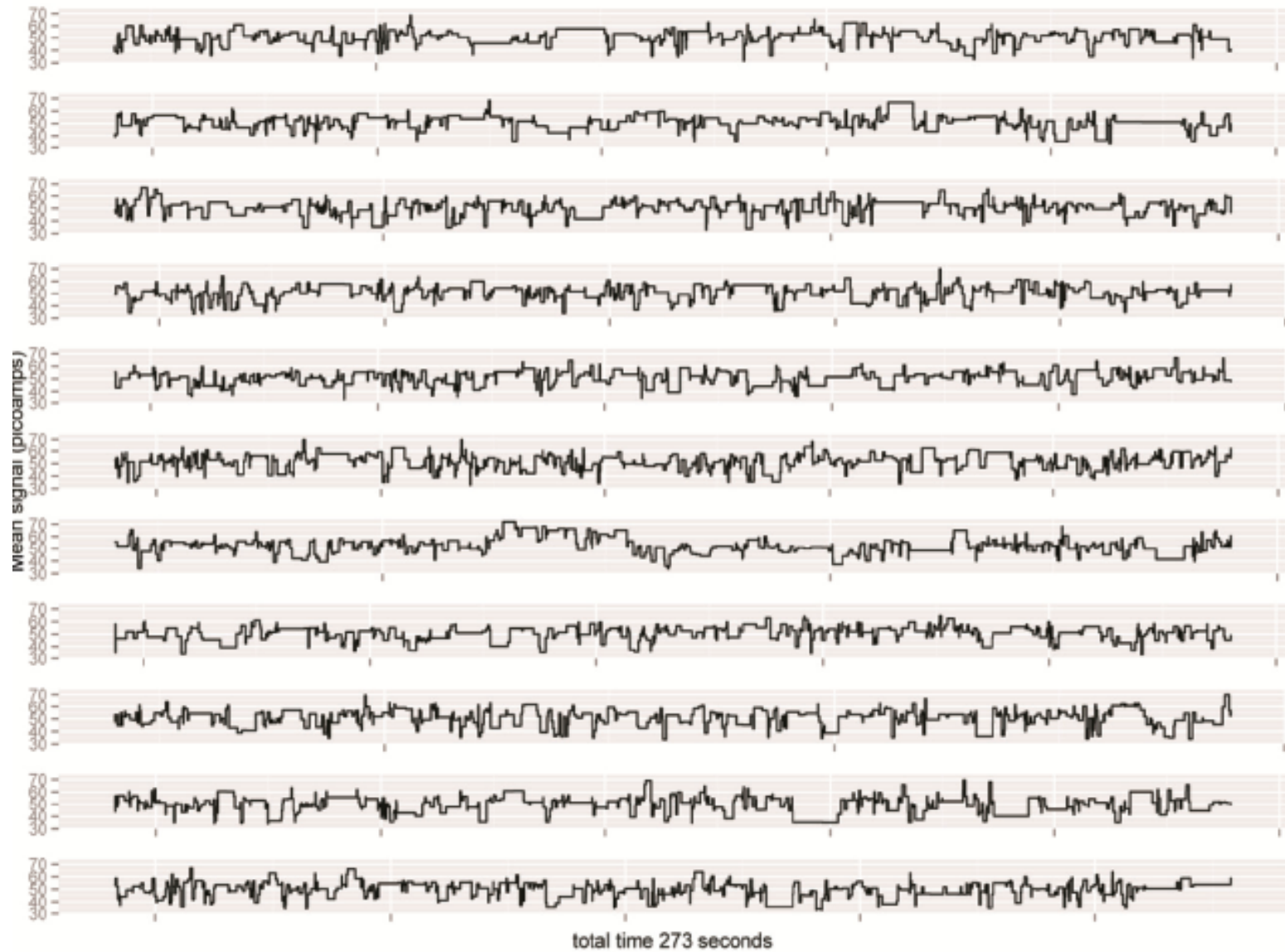


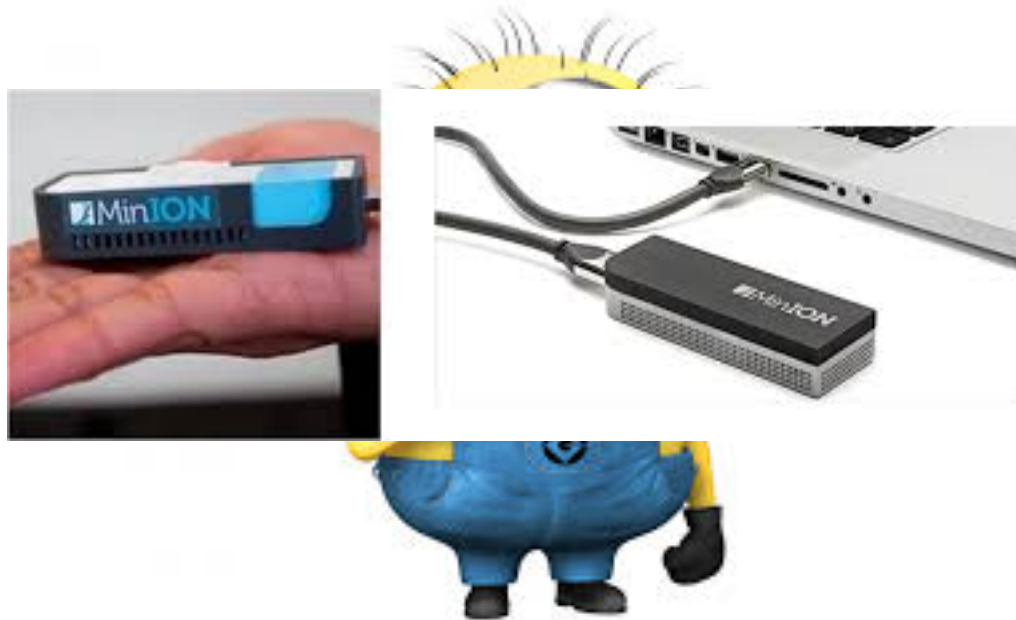
Fig. 2. Nanopore strand sequencing.

(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 α HL 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).

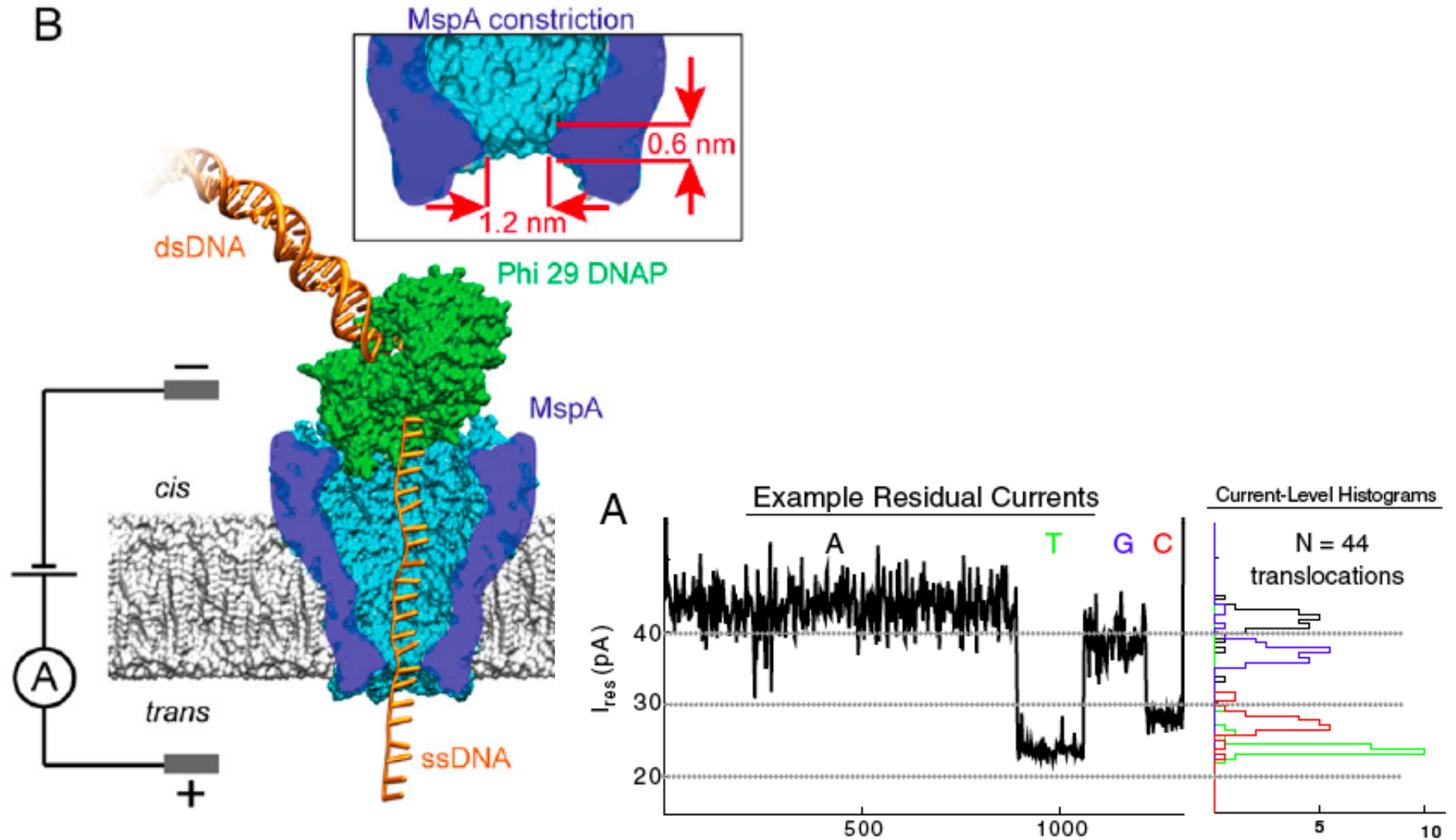
Sequence from the minION



The MinION



MspA Nanopore



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.

-Seq

- ChIP-Seq
- Exome-Seq
- RAD-Seq or Restriction site-associated DNA Sequencing.
- RRL-Seq or Reduced Representational Libraries
- STARR-Seq
- 3C, 4C and 5C-Seq
- cfDNA-Seq
- Strand-Seq

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 M. Leach,^{6†} et al.^{7,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000}



Sequencing the nuclear genome of the extinct woolly mammoth

Hernán
Barbar
Eric S.
Christi
Vladim
Javier F
Daniel
Janet K

Webb Miller¹, Daniela I. I.
Michael D. Packard¹, Fan
Kerstin Lindblad-Toh⁵, Eri
Sharon Sheridan⁷, 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*}, Rence Viola^{3,7*},

Adrian W. Briggs^{1,3}, U
Can Alkan¹⁰, Qiaomei
Michael Richards^{7,13},
Montgomery Slatkin⁶

Insights into hominid evolution from the gorilla genome sequence



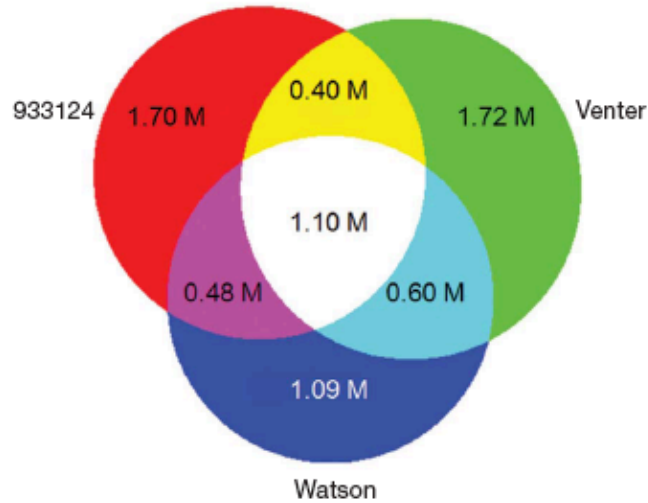
Aylwyn Scally¹, Julien Y. Duthéil^{2†}, LaDeana W. Hillier³, Gregory E. Jordan⁴, Ian Goodhead^{1†}, Javier Herrero⁴, Asger Hobolth², Tuuli Lappalainen⁵, Thomas Mailund², Tomas Marques-Bonet^{3,6,7}, Shane McCarthy¹, Stephen H. Montgomery⁸, Petra C. Schwalie⁴, Y. Amy Tang¹, Michelle C. Ward^{9,10}, Yali Xue¹, Bryndis Yngvadottir^{1†}, Can Alkan^{3,11}, Lars N. Andersen², Qasim Ayub¹, Edward V. Ball¹², Kathryn Beal⁴, Brenda J. Bradley^{8,13}, Yuan Chen¹, Chris M. Clee¹, Stephen Fitzgerald⁴, Tina A. Graves¹⁴, Yong Gu¹, Paul Heath¹, Andreas Heger¹⁵, Emre Karakoc³, Anja Kolb-Kokocinski¹, Gavin K. Laird¹, Gerton Lunter¹⁶, Stephen Meader¹⁵, Matthew Mort¹², James C. Mullikin¹⁷, Kasper Munch², Timothy D. O'Connor⁸, Andrew D. Phillips¹², Javier Prado-Martinez⁹, Anthony S. Rogers^{1†}, Saba Sajadian³, Dominic Schmidt^{9,10}, Katy Shaw¹², Jared T. Simpson¹, Peter D. Stenson¹², Daniel J. Turner^{1†}, Linda Vigilant¹⁸, Albert J. Vilella⁴, Weldon Whitener¹, Baoli Zhu^{19†}, David N. Cooper¹², Pieter de Jong¹⁹, Emmanouil T. Dermitzakis⁵, Evan E. Eichler^{3,11}, Paul Flicek⁴, Nick Goldman⁴, Nicholas I. Mundy⁸, Zemin Ning¹, Duncan T. Odom^{1,9,10}, Chris P. Ponting¹⁵, Michael A. Quail¹, Oliver A. Ryder²⁰, Stephen M. Searle¹, Wesley C. Warren¹⁴, Richard K. Wilson¹⁴, Mikkel H. Schierup², Jane Rogers^{1†}, Chris Tyler-Smith¹ & Richard Durbin¹

Applications

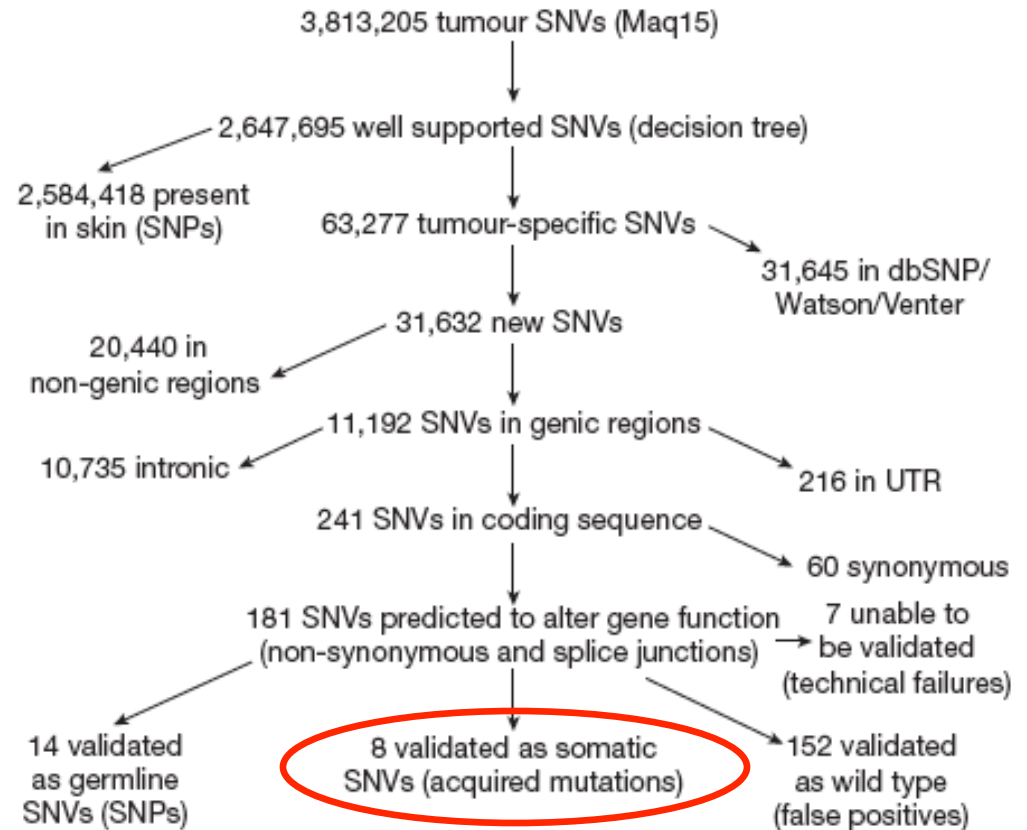
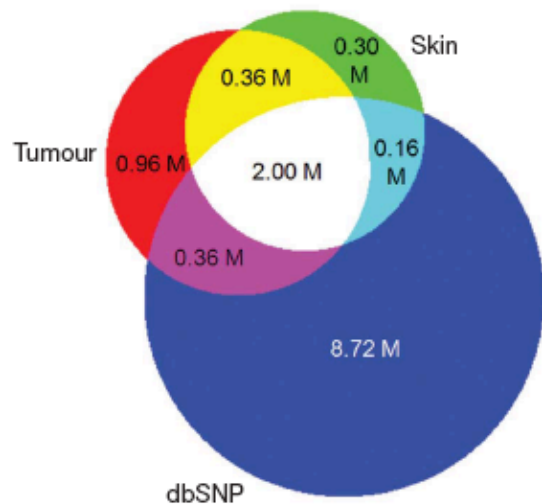
- Whole Genome Sequencing
- Exome Sequencing
- Targeted Genomic Sequencing
- Chromatin-IP-Sequencing
- DNase I Hypersensitivity Sequencing
- Methyl-Seq (RRBS, MeDIP, etc)
- Microbiome Sequencing
- Metagenomics

AML: Comparisons

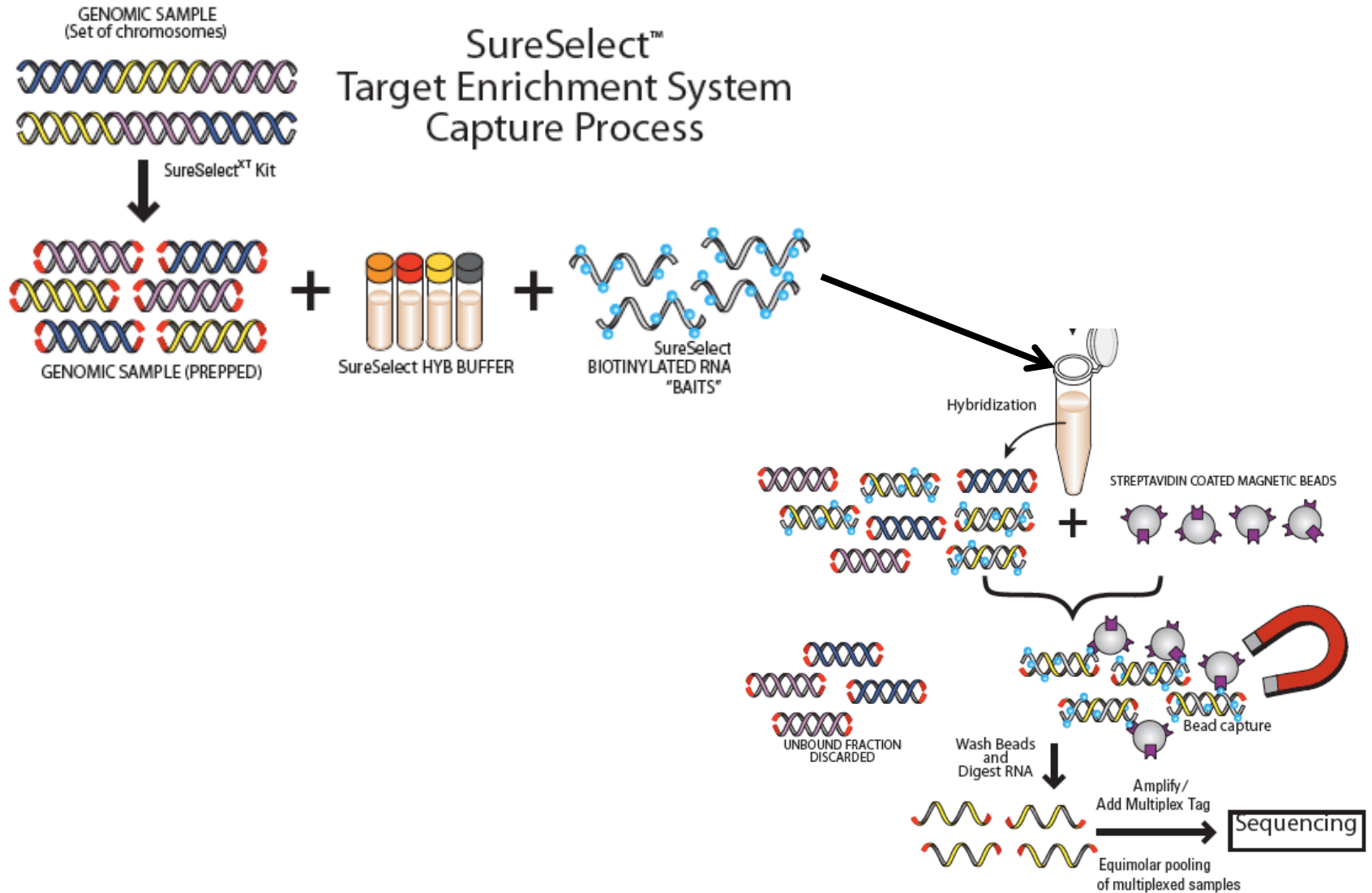
a



b



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 Hearing 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

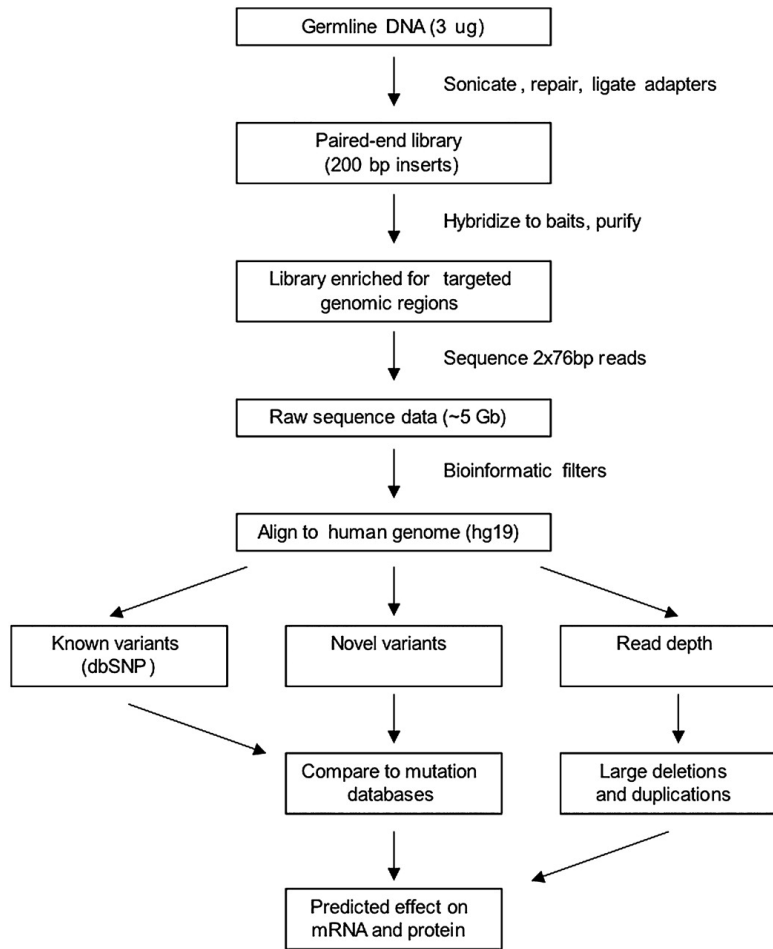
Long range PCR

Multiplex PCR strategies

Solution capture on Biotin labeled oligos

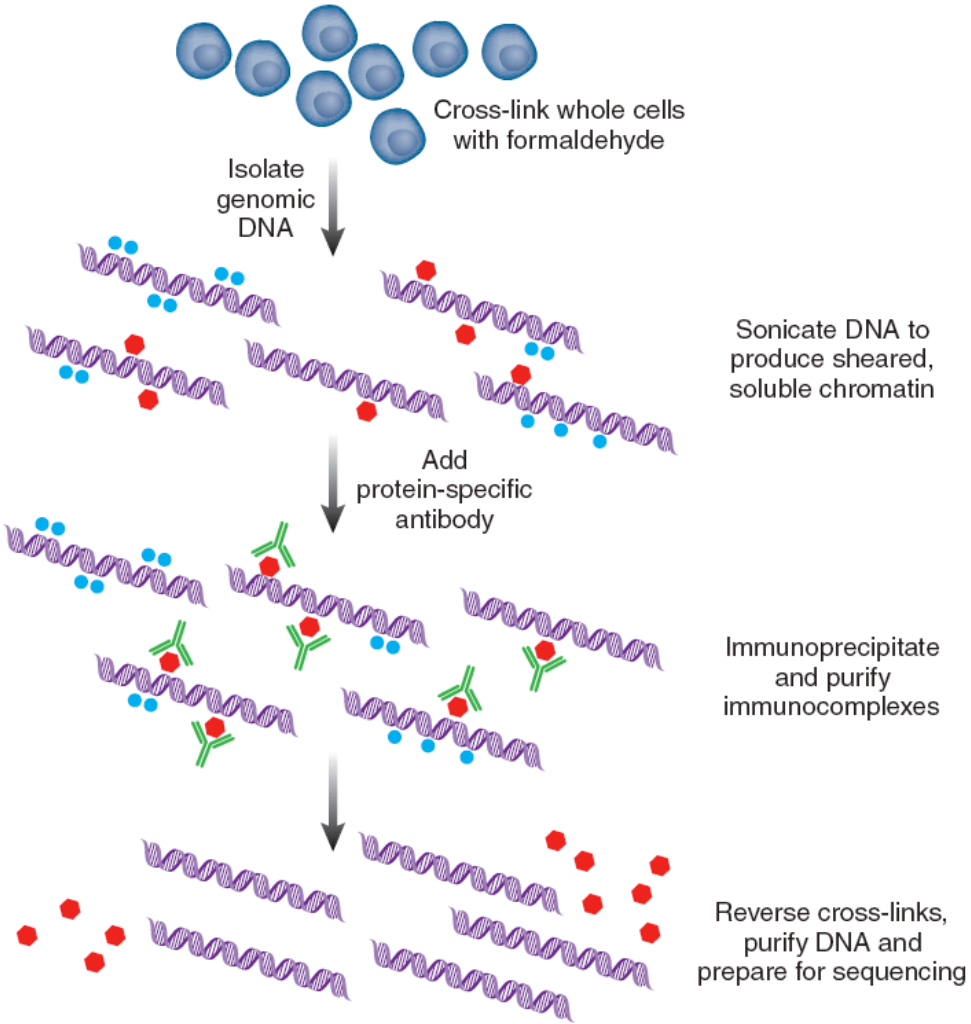
HaloPlex

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



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

ChIP-Seq



Sonicate DNA to produce sheared, soluble chromatin

Immunoprecipitate and purify immunocomplexes

Reverse cross-links, purify DNA and prepare for sequencing

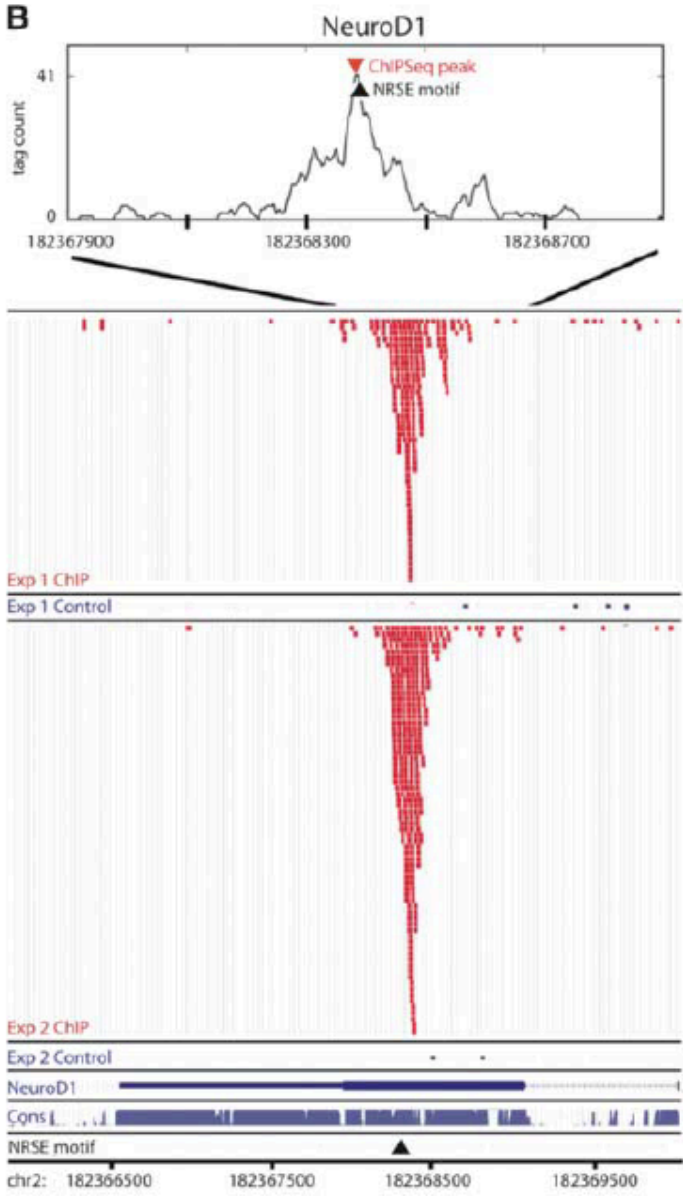
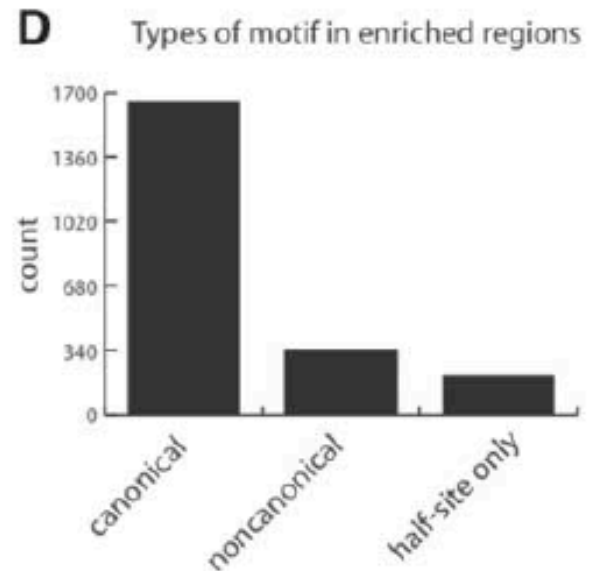
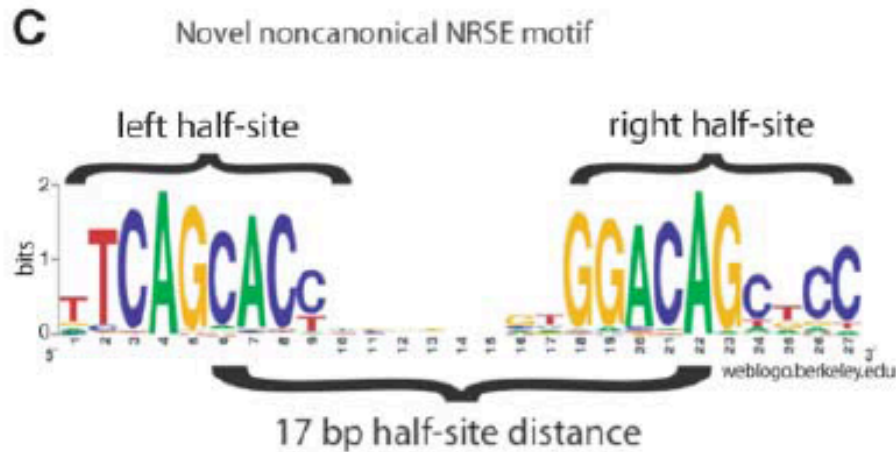
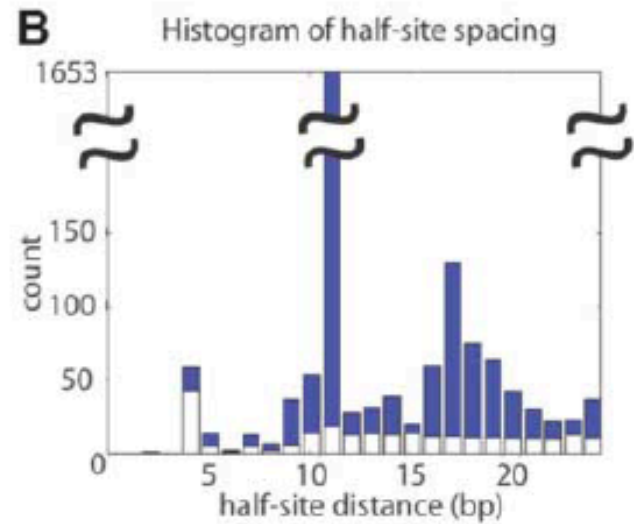
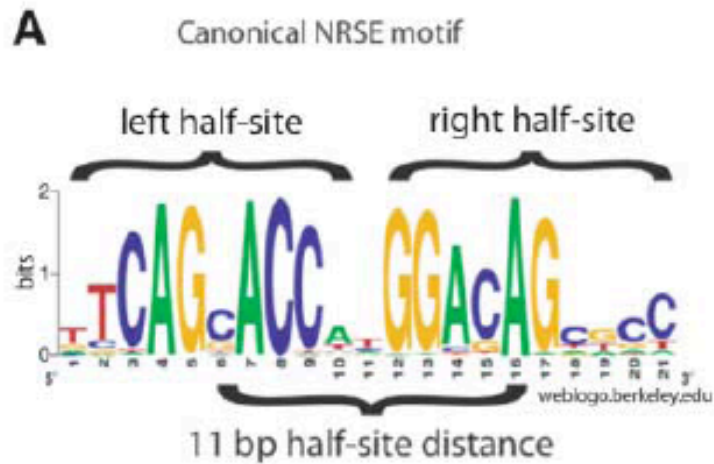


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

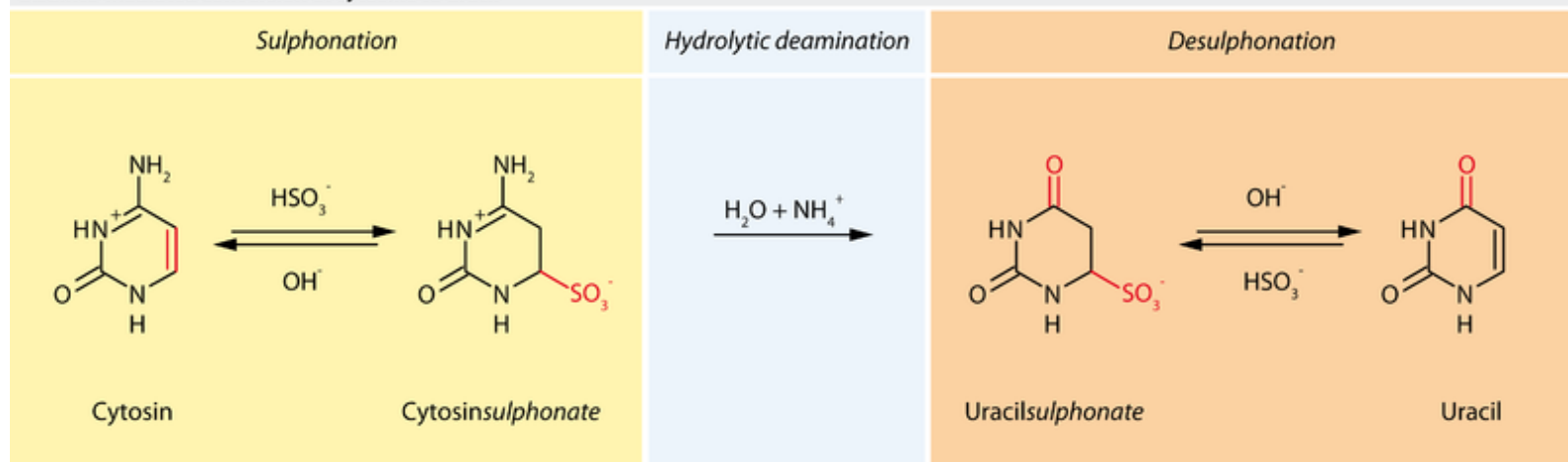
ChIP-Seq



Methylation profiling

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

Bisulfite-mediated conversion of cytosine to uracil



MeDIP-Seq

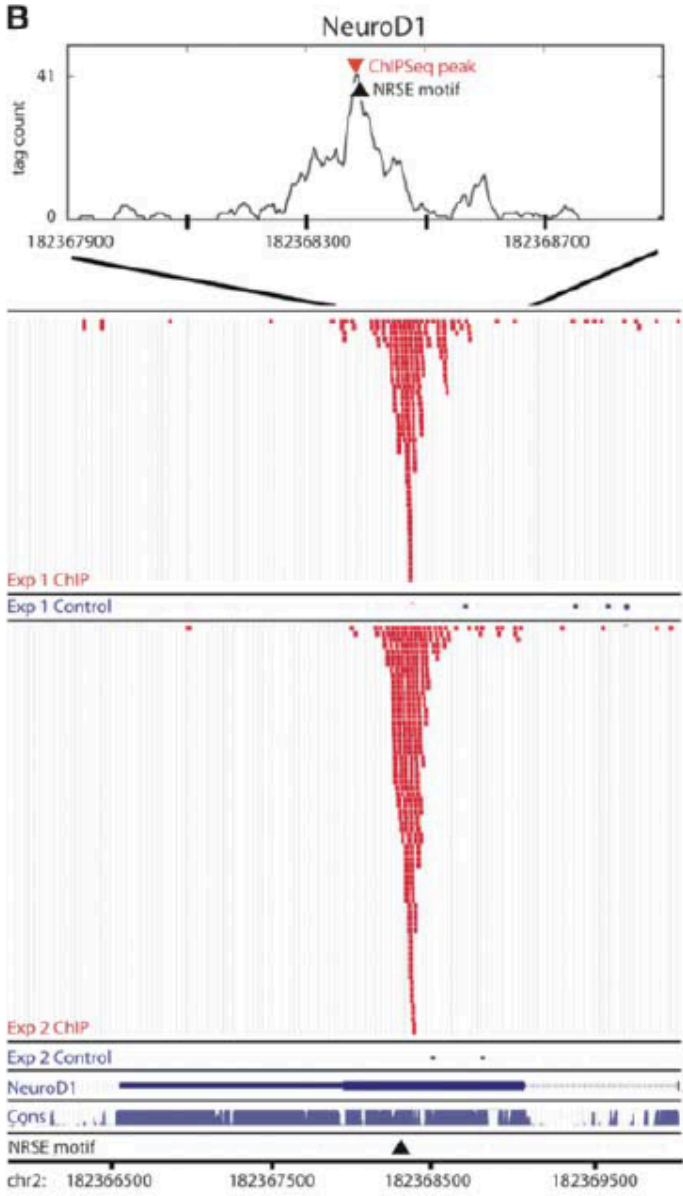
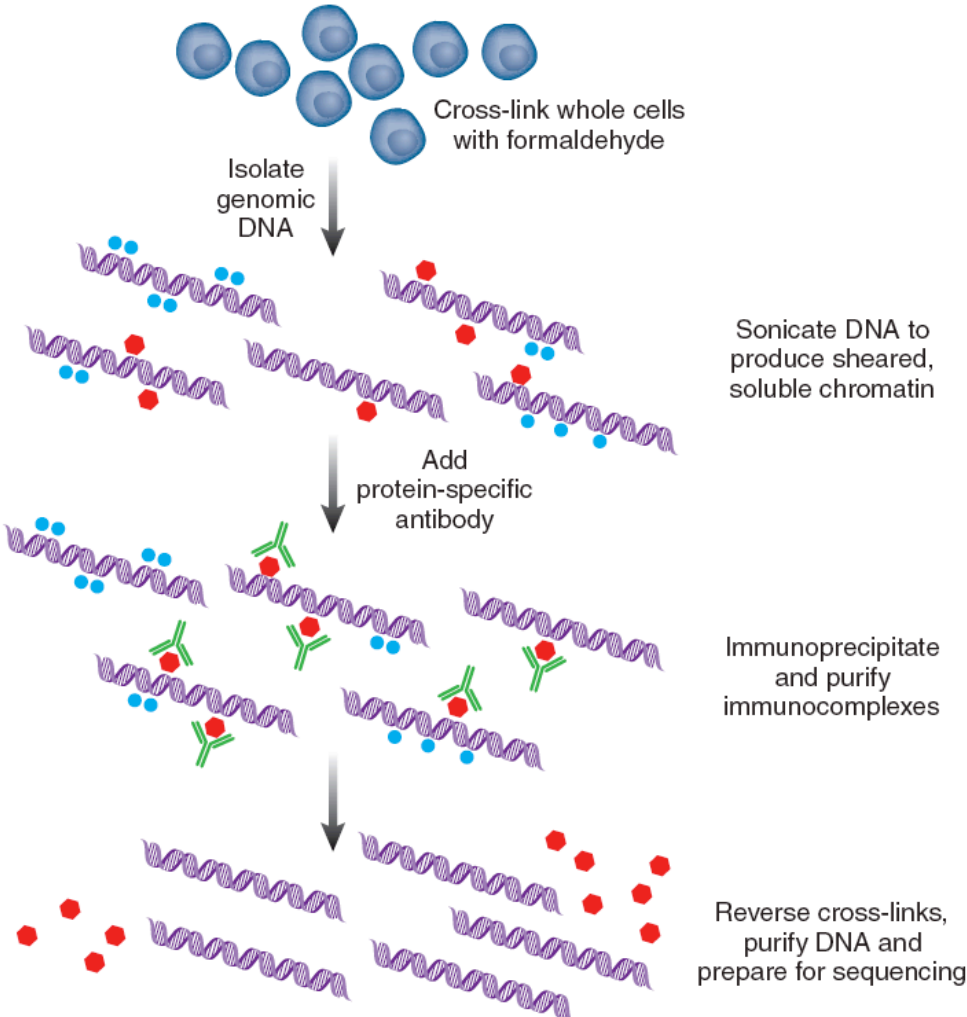
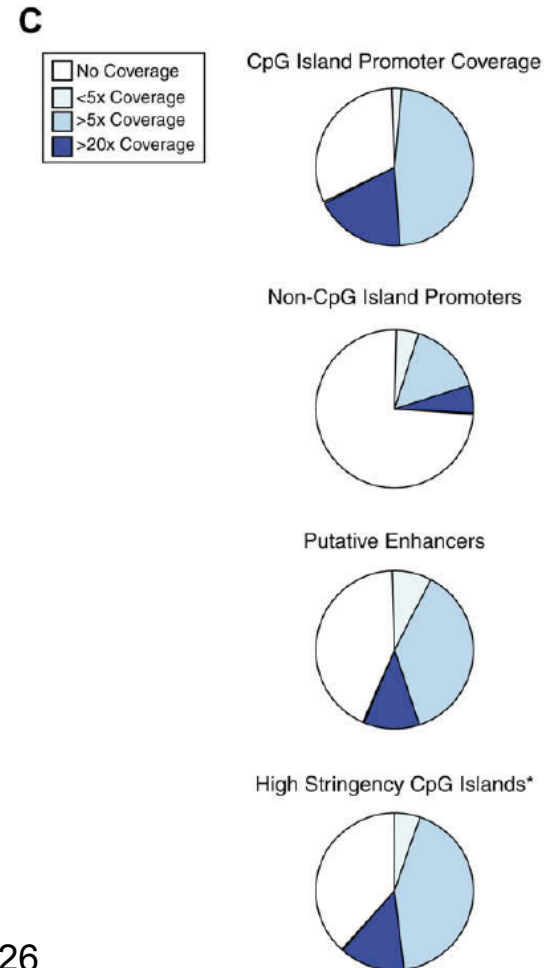
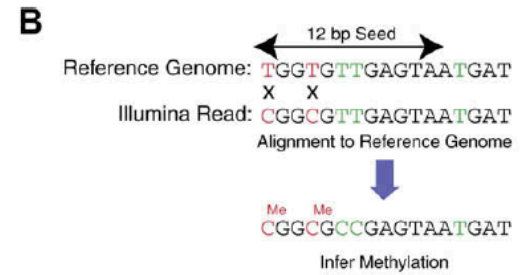
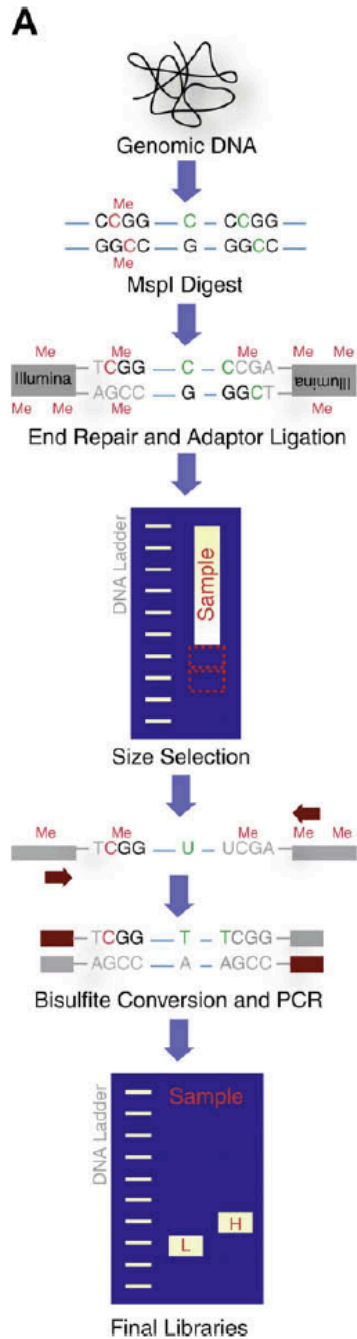


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

RRBS



The NIH Human Microbiome Project

The NIH HMP Working Group¹

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.**

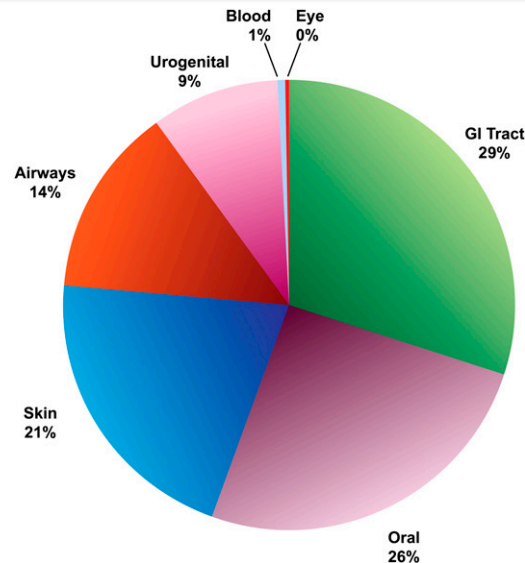
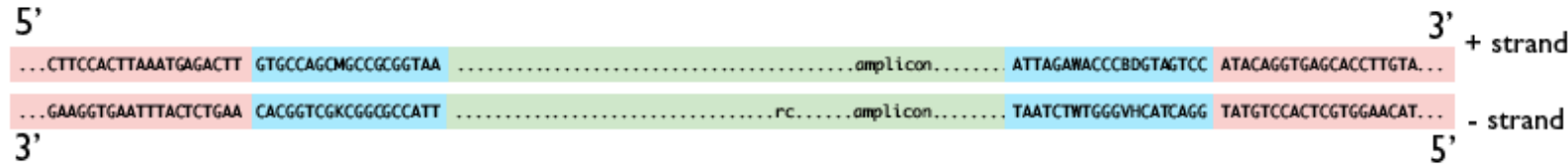


Figure 3. Bacterial distribution by body site. This figure shows the distribution by body site of bacteria that have been sequenced under the HMP or are in the sequencing pipelines.

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}

Target gene:



Amplification primers with annealing sites:



MSA after forward primer

Jalview 2.7
File Tools Vamsas Help Window

C:\Users\ranjit\Desktop\morrow-working-files\RDP\bacterial16S_508_mod5.stk

File Edit Select View Format Colour Calculate Web Service

750 760 770 780 790 800 810 820 830 840 850 860 870 880

NC_007292/1-1566 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C G A G C G U U A A U C G G A A U U A C U G G G C G U A A A . G A G U A C G U A G G U G G U . U U G U U A A G U C A G . A U G U G . A A A U C C C G U A G C U C A A C U U A G G A . A C U G C A U U U G
NC_008769/1-1532 C G U G C C A G C A G C C G C G G U A A U A C G U A G G G U G C G A G C G U U G U C C G G A A U U A C U G G G C G U A A A . G A G C U C G U A G G U G G U . U U G U C G C G U U G U . U C G U G . A A A U C U C A C G G C U U A A C U G U G A G . C G U G C G G G C G
NC_008800/1-1543 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A A U C G G A A U U A C U G G G C G U A A A . G C G C A C G C A G G C G G U . U U G U U A A G U C A G . A U G U G . A A A U C C C G C G C U U A A C G U G G G A . A C U G C A U U U G
NC_009446/1-1533 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A U C C G G A A U A G A C U G G G C G U A A A . G C G C A C G C A G G U G G U . U U U A U A A G U C A G . G U G U G . A A A U C C U G G G C U C A A C C U A G G A . A U U G C A U U U G
NC_008767/1-1541 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C G A G C G U U A U C C G G A A U U A C U G G G C G U A A A . G C G G G C G C A G A C G G U . U A C U U A A A G C A G G . A U G U G . A A A U C C C G G G C U C A A C C C G G G A . A C U G C G U U C U
NC_009445/1-1489 C G U G C C A G C A G C C G C G G U A A U A C G A A G G G G G C U A G C G U U G C U C G A A A C U G G G C G U A A A . G G G U G C G U A G G C G G G . U C U U U A A G U C A G . G G G U G . A A A U C C U G G A G C U A A C C U C A G A . A C U G C C U U U G
NC_009443/1-1549 C G U G C C A G C A G C C G C G G U A A U A C G U A G G U C C G A G C G U U G U C C G G A U U U A U G G G C G U A A A . G C G A G C G C A G G C G G U . U U G A U A A G U C U G . A A G U A . A A A G G C U G U G G C U U A A C C A U A G U . A C . G C U U U G G
NC_009442/1-1549 C G U G C C A G C A G C C G C G G U A A U A C G U A G G U C C G A G C G U U G U C C G G A U U U A U G G G C G U A A A . G C G A G C G C A G G C G G U . U U G A U A A G U C U G . A A G U A . A A A G G C U G U G G C U U A A C C A U A G U . A C . G C U U U G G
NC_009441/1-1514 C G U G C C A G C A G C C G C G G U A A U A C G G A G G A U C C A A G C G U U A U C C G G A A U C A U U G G G C U U A A A . G G G U C C G U A G G C G G U . U U A G U A A G U C A G . U G G U G . A A A G C C A U C G C U C A A C C U G G G A . A C G C C A U U G
NC_009049/1-1467 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G G G C U A G C G U U A U C C G G A A U U A C U G G G C G U A A A . G C G C A C G U A G G C G G A . U C G G A A A G U C A G . A G G U G . A A A U C C A G G G C U C A A C C C U G G A . A C U G C C U U U G
NC_003454/1-1520 C G U G C C A G C A G C C G C G G U A A U A C G U A U G U C A C G A G C G U U A U C C G G A U U U A U G G G C G U A A A . G C G C U C U A G G G C G G U . U A U G U A A G U C U G . A U G U G . A A A U G C A G G G C U C A A C C U C U G A . U U G C U U G G
NC_008369/1-1528 C G U G C C A G C A G C C G C G G U A A U A C G G G G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G G G U C U G U A G G U G G U . U U G U U A A G U C A G . A U G U G . A A A G C C A G G G C U C A A C C U U G G A . A C U G C A U U U G
NC_007722/1-1486 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G A G C U A G C G U U G U U C C G A A U U A C U G G G C G U A A A . G C G C G C U A G G C G G C . U A U U A A A G U C A G . G G G U G . A A A U C C C G G G G C U C A A C C C G G A . A C U G C C U U U G
NC_008009/1-1502 U G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A U C C G G A A U U A U G G G C U U A A A . G G G U C C G U A G G C G G U . U U A G U A A G U C U C . U A G U G . A A A U C C C G G G C U C A A C C U C G G A . C C U G C U A G G G
NC_003450/1-1524 C G U G C C A G C A G C C G C G G U A A U A C G U A G G G U G C G A G C G U U G U C C G G A A U U A C U G G G C G U A A A . G A G C U C G U A G G U G G U . U U G U C G C G U C G U . C U G U G . A A A U C C C G G G C U U A A C U U C G G G . C G U G C A G G C G
NC_002771/1-1525 U G U G C C A G C A G C C G C G G U A A U A C A U A G G G U G C A A G C G U U A U C C G A A A U U A U G G G C U U A A . G A G U U C G U A G G U U G . U U G U U A A G U C A G . A A G U U . A A A U C C C G G G C U C A A C C C U G G C . C C . G C U U U U G
NC_005966/1-1508 U G U G C C A G C A G C C G C G G U A A U A C A G A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G C G C U A G G C G G C . C A A U U A A G U C A A . A U G U G . A A A U C C C G A G C U U A A C U U G G G A . A U U G C A U U C G
NC_009439/1-1536 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A A U C G G A A U U A C U G G G C G U A A A . G C G C G C U A G G U G G U . U C G U U A A G U U G G . A U G U G . A A A G C C C G G G C U C A A C C U G G G A . A C U G C A U C C A
NC_009438/1-1543 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U C C A A G C G U U A U C C G A A U U A C U G G G C G U A A A . G C G U C G C A G G C G G U . U U G U U A A G C A G . A U G U G . A A A G C C U G G G C U C A A C C U G G A . A U A G C A U U U C
NC_009437/1-1544 C G U G C C A G C A G C C G C G G U A A U A C G U A G U G U G C G A G C G U U G U C C G G A A U U A C U G G G C G U A A A . G G G U C G U A G G C G G C . U A U G C G A G U U A A . G C G U G . A A A G C C U U A G G G C U C A A C C U A A G G . A U U G C G C U U A
NC_004129/1-1539 U G U G C C A G C A G C C G C G G U A A U A C A G A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G C G C U A G G U G G U . U U G U U A A G U U G G . A U G U G . A A A G C C C G G G C U C A A C C U G G G A . A C U G C A U C C A
NC_009436/1-1540 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A A U C G G A A U U A C U G G G C G U A A A . G C G C A G C G U A G G C G G U . C U G U C A A G U C G G . A U G U G . A A A U C C C G G G C U C A A C C U G G G A . A C U G C A U U C G
NC_009434/1-1537 C G U G C C A G C A G C C G C G G U A A U A C G A A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G C G C U A G G U G G U . U C G U U A A G U U G G . A U G U G . A A A G C C C G G G C U C A A C C U G G G A . A C U G C A U C C A
NC_008268/1-1518 C G U G C C A G C A G C C G C G G U A A U A C G U A G G G U G C A A G C G U U A U C C G G A A U U A C U G G G C G U A A A . G A G U U C G U A G G C G G U . U U G U C G C G U C G U . U U G U G . A A A A C U C A C A G C U C A A C U G U G A G . C C U G C A G G C G
NC_008752/1-1529 C G U G C C A G C A G C C G C G G U A A U A C G U A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G U G C G C A G G C G G U . G A U G U A A G A C A G . A U G U G . A A A U C C C G G G C U C A A C C U G G G A . A C U G C A U U U G
NC_008751/1-1549 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C G A G C G U U A A U C C G A A U A C U G G G C G U A A A . G C G C A C G U A G G C U G C . U U G G U A A G U C A G . G G G U G . A A A G C C C G G G C U C A A C C G C G G A . A U U G C C U U U G
NC_007712/1-1542 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G U C A A G C G G U . C U G U U A A G U C A G . A U G U G . A A A U C C C G G G C U U A A C C U G G G A . A C U G C A U U U G
NC_008750/1-1543 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U C C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G U G C G A G G C G G U . U U G U U A A G C A G . A U G U G . A A A G C C C G G G C U C A A C C U A G G A . A U A G C A U U U C
NC_008359/1-1455 C G U G C C A G C A G C C G C G G U A A U A C G A A G G G G G C U A G C G U U G U C C G G A A U U A C U G G G C G U A A A . G C G C A C G U A G G C G G A . C U U U U A A G U C A G . G U G U G . A A A U C C G A G G G C U C A A C C U C G G A . A C U G C A U U U G
NC_004088/1-1543 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A A U C C G A A U U A C U G G G C G U A A A . G C G C A C G C A G G C G G U . U U G U U A A G U C A G . A U G U G . A A A U C C C G C G C U U A A C G U G G G A . A C U G C A U U U G
NC_007677/1-1540 C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U G U C C G G A A U A C U G G G U G U A A A . G G G U G U G C A G G C G G G . G C A G C A A G U C G G . A U G U G . A A A C C C A U G G C U U A A C C A U G G A . G G U G C A U U C G
NC_003047/1-1485 C G U G C C A G C A G C C G C G G U A A U A C G A A G G G G G C U A G C G U U G U U C C G G A A U U A C U G G G C G U A A A . G C G C A C G U A G G C G G A . U U G U U A A G U G A G . G G G U G . A A A U C C A G G G C U C A A C C C U G G A . A C U G C C U U U C
NC_005957/1-1552 C G U G C C A G C A G C C G C G G U A A U A C G U A G G U G G C A A G C G U U A U C C G G A A U U A U U G G G C G U A A A . G C G C G C G A G G U G G U . U U C U U A A G U C U G . A U G U G . A A A G C C A C G G C U C A A C C U G G A . G G G U C A U U G G
NC_005956/1-1488 C G U G C C A G C A G C C G C G G U A A U A C G A A G G G G G C U A G C G U U G U U C C G G A A U U A C U G G G C G U A A A . G C G C A U G U A G G C G G A . U A U U U A A G U C A G . A G G U G . A A A U C C A G G G C U C A A C C C U G G A . A C U G C C U U U G

secondary structure
reference positions C G U G C C A G C A G C C G C G G U A A U A C G G A G G g g g C a a g C g u u g u c C G G A A U U A c U G G G C G U A A A . G a G g C g c A g G g G g c . c g g c c A a g u c g g . g u G c g . A A A u u c c g g g G c U u A A C c c g g g a . A a c G C a c c g

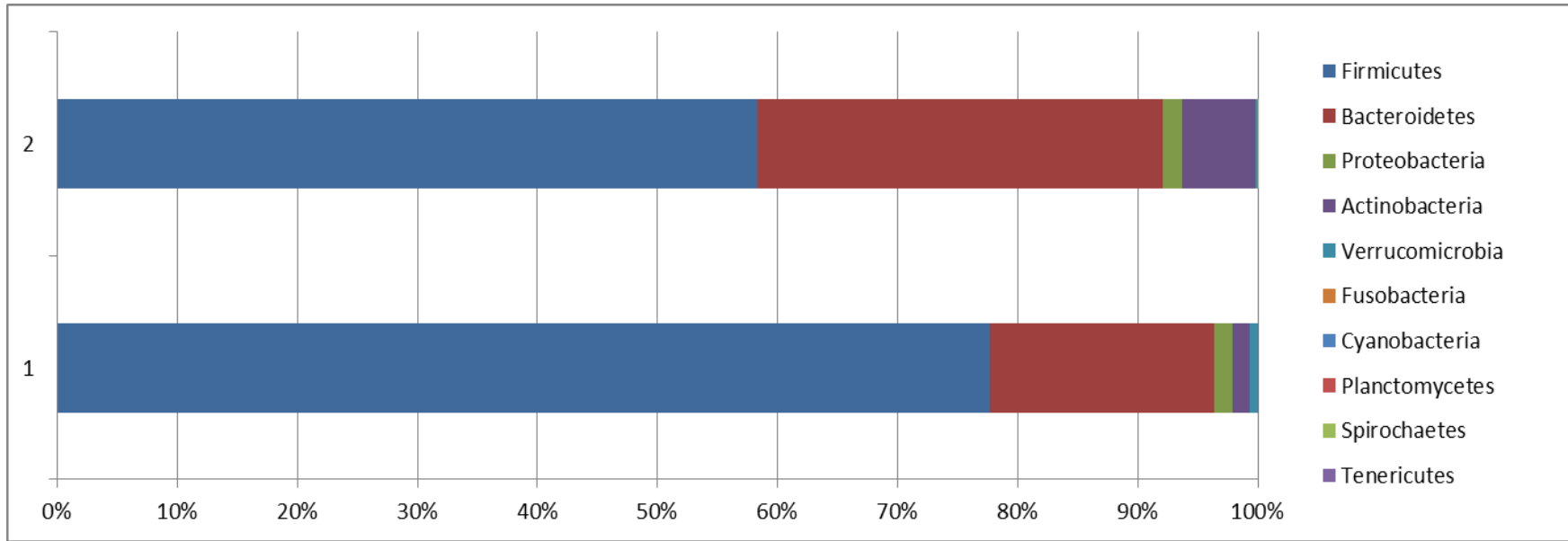
Consensus C G U G C C A G C A G C C G C G G U A A U A C G G A G G G U G C A A G C G U U A U C C G A A U U A C U G G G C G U A A A . G C G C G C U A G G C G G U . U U G U U A A G U C A G U A U G U G . A A A U C C C G G G C U C A A C C U G G G A . A C U G C A U U U G

Sequence 13 ID: NC_007722 Nucleotide: Uracil (561)

1:50 PM
10/10/2011

Microbiome at UAB

Normal Diabetic



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. Vogensen¹, Frans W. J. van den Berg¹, Dennis Sandris Nielsen¹, Anne Sofie Andreassen², Bente K. Pedersen², Waleed Abu Al-Soud³, Søren J. Sørensen³, Lars H. Hansen³, Mogens Jakobsen¹

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

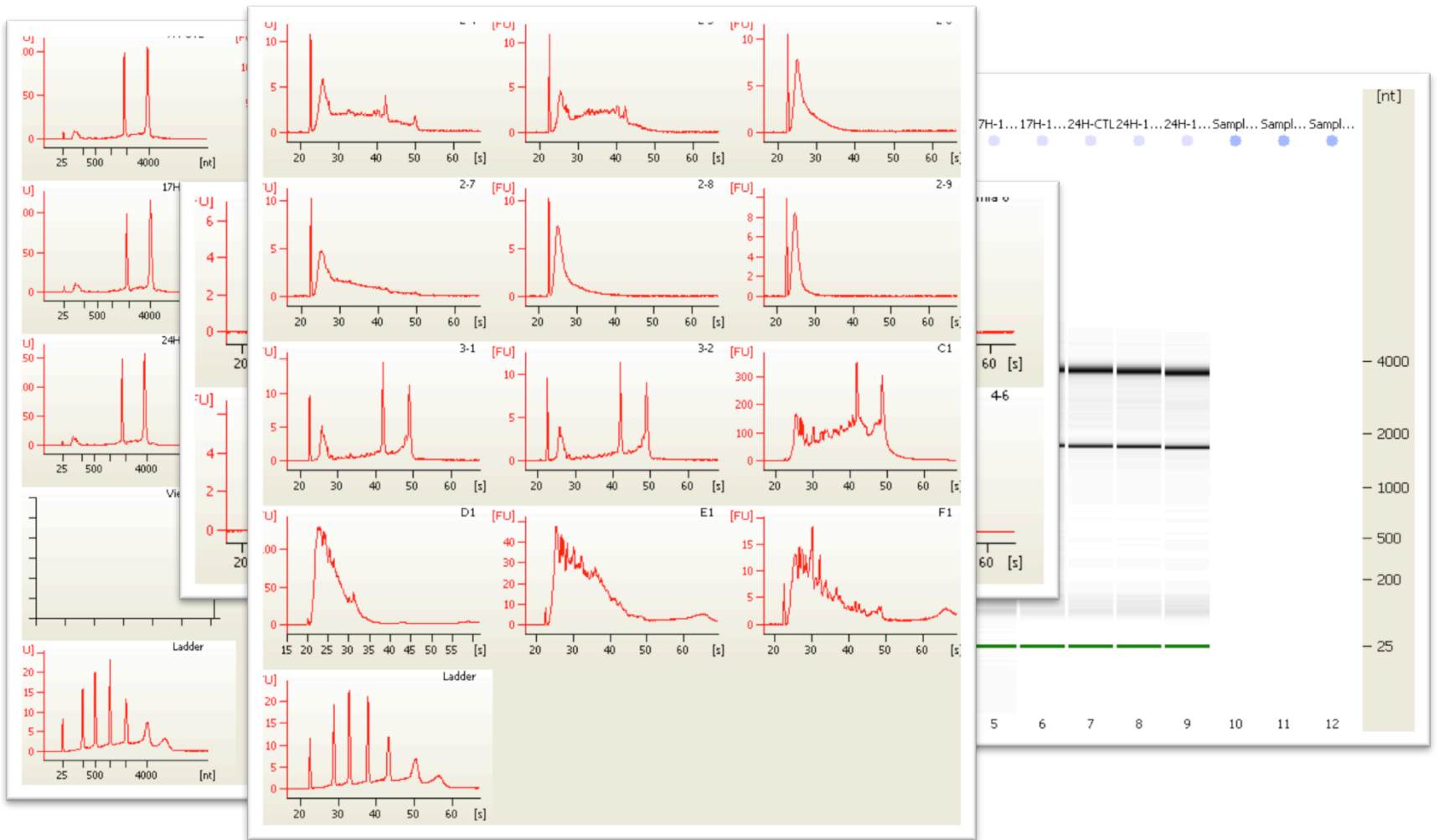
RNA Applications

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

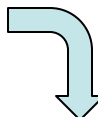
Advantages of RNA-Seq

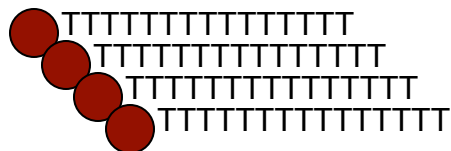
- Digital gene expression
 - Simply count the number of reads for a given transcript
- Greater dynamic range
- No hybridization bias
- Not dependent on known content
- Generate alternative splice/exon usage
- Identify variants
- Allele Specific Expression
- Identify RNA editing

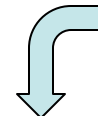




RNA Quality

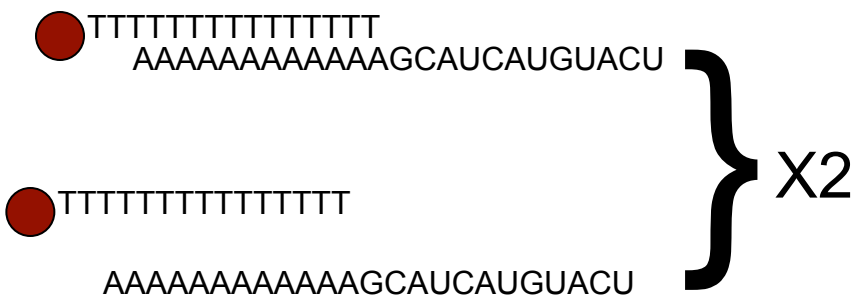




RNA-Seq Library Prep

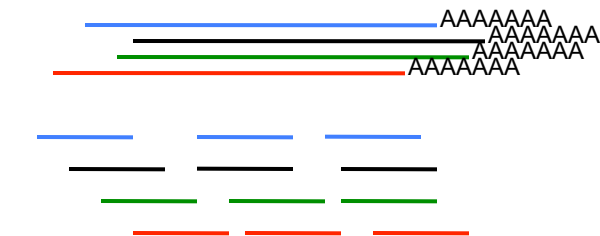
Total RNA  Mix/65°C/25°C



Total RNA 
gacctgcttctgggtcgggg 
gacctgcttctgggtcgggg  28S
gcggcttgggtgactctaga 
gcggcttgggtgactctaga  18S



Fragment 
Random Priming to make cDNA 

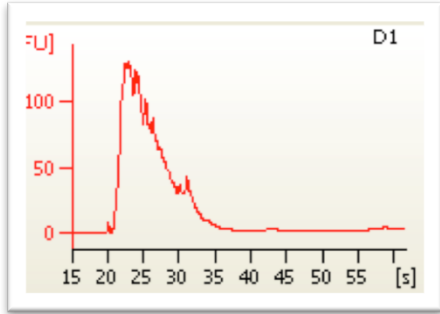


Random Priming to make cDNA



Sequence ready library

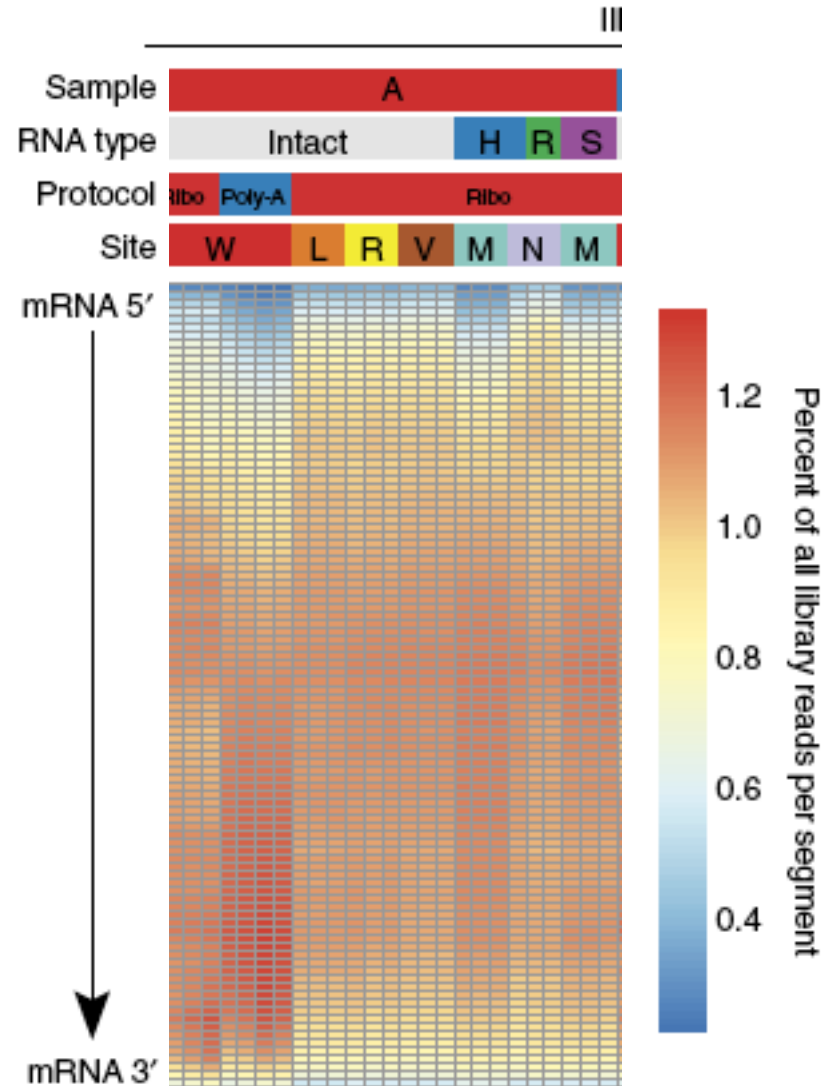
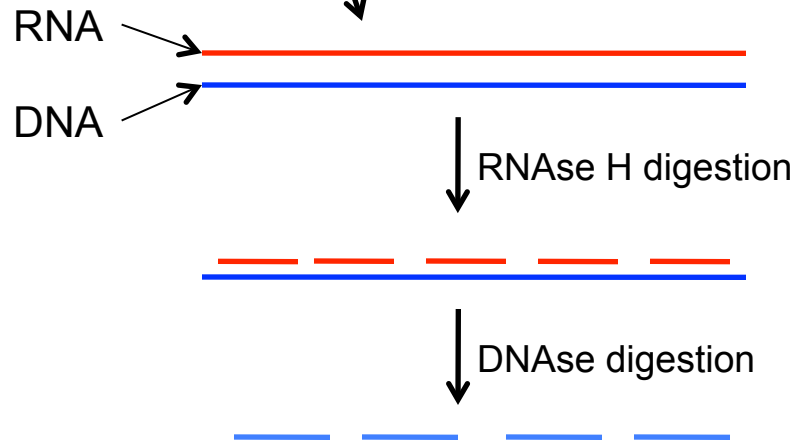
Degraded RNA (FFPE)



Total RNA

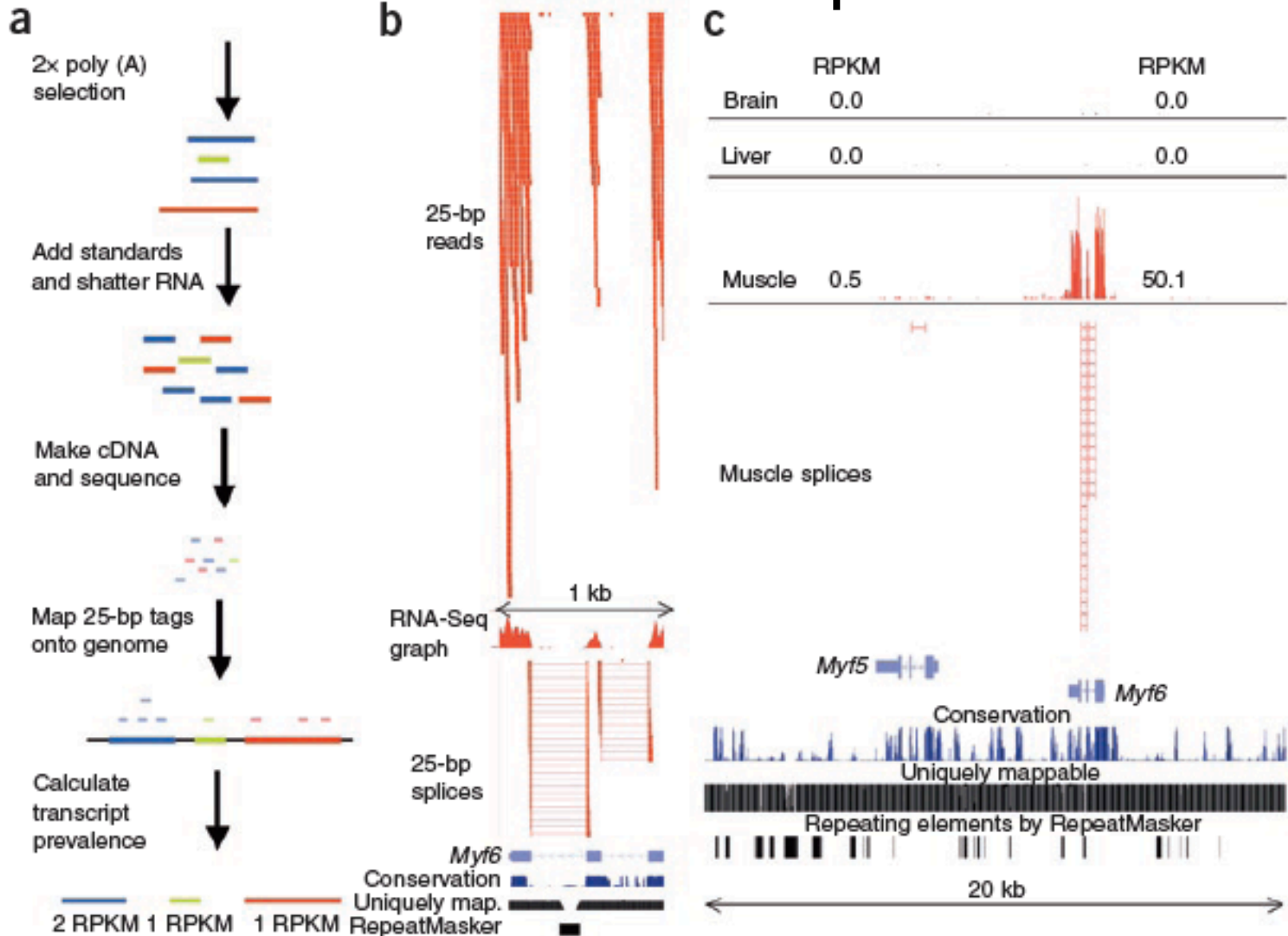
gacctgcttctgggtcgggg
gacctgcttctgggtcgggg 28S

gcggctttggtgactctaga
gcggctttggtgactctaga 18S



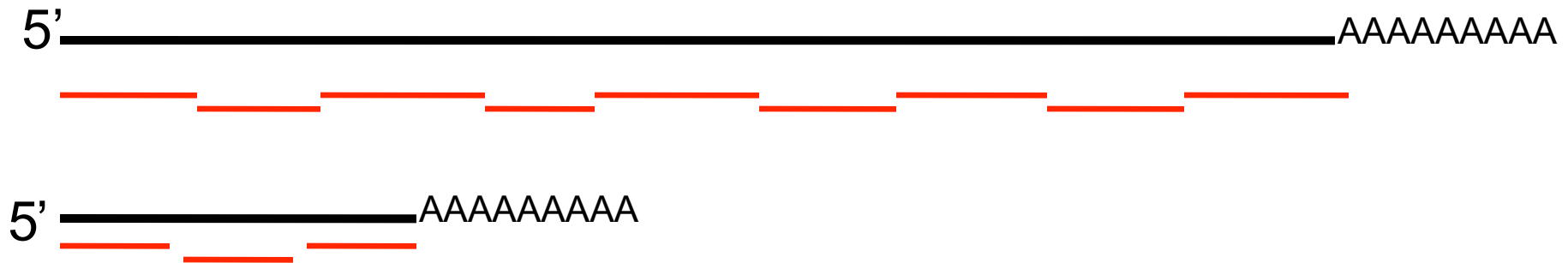
From Li et al., 2014 Nat. Biotech.32:915

mRNA-Seq



Digital Gene Expression

- Caveat?

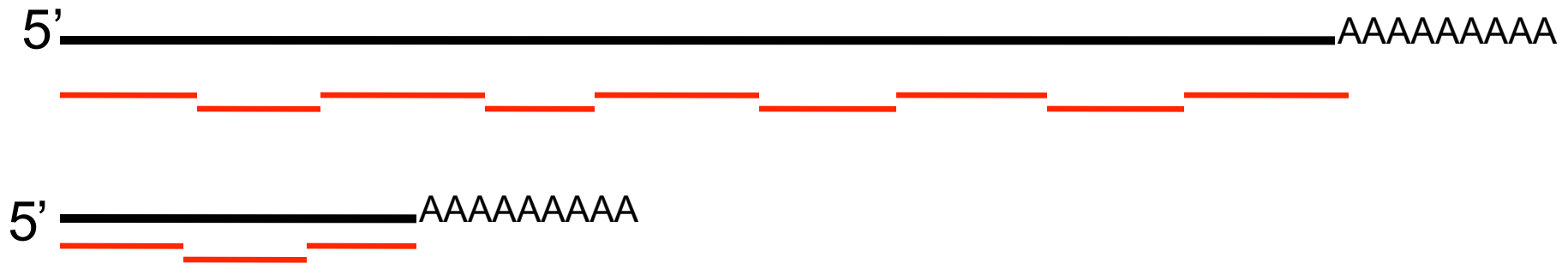


Gene 1 has 9 reads
Gene 2 has 3 reads

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

Digital Gene Expression

- Caveat?



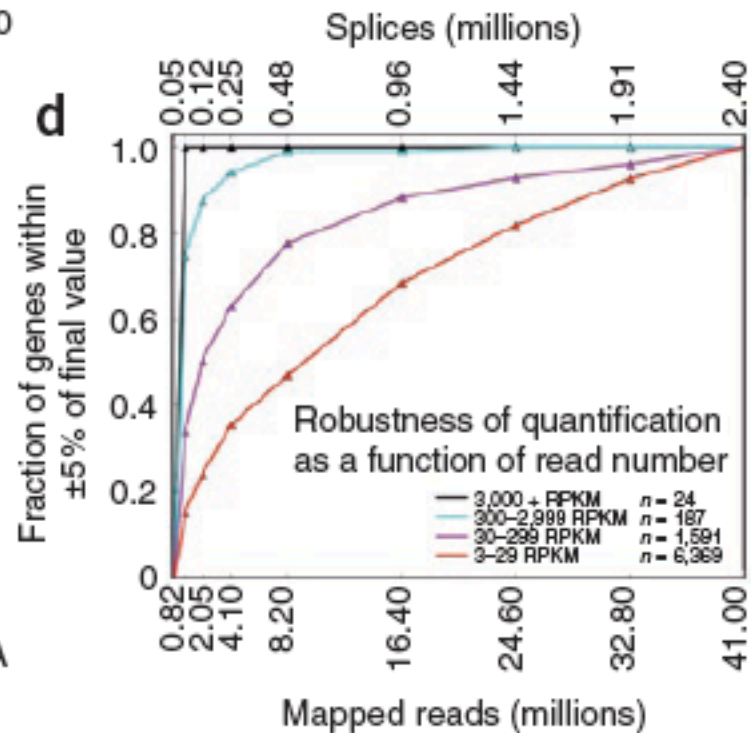
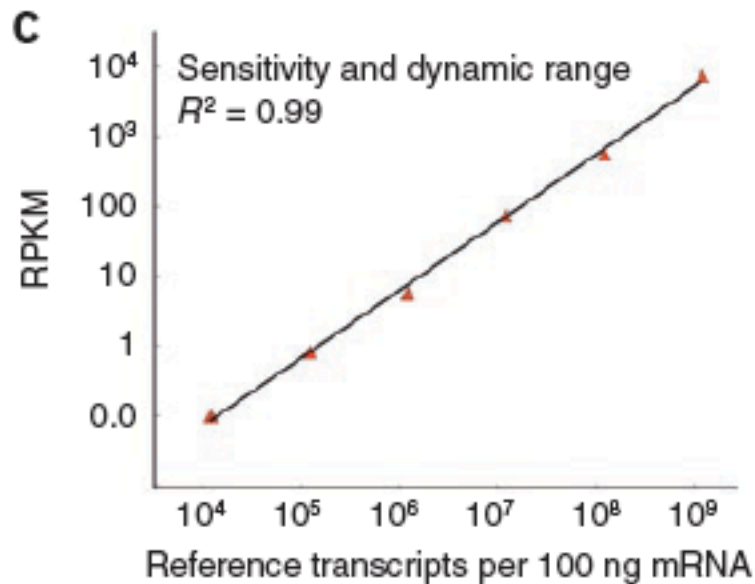
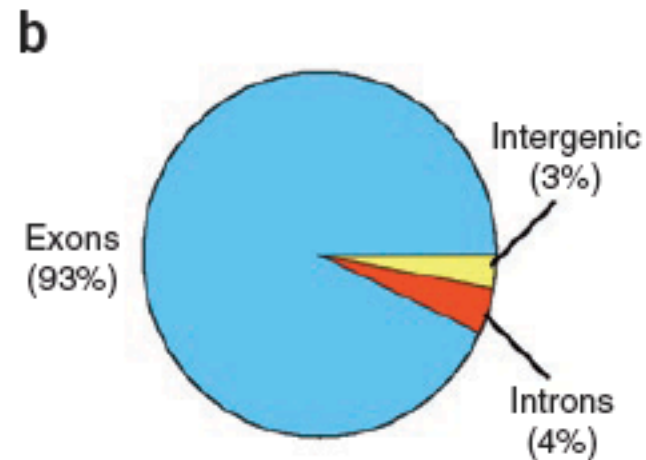
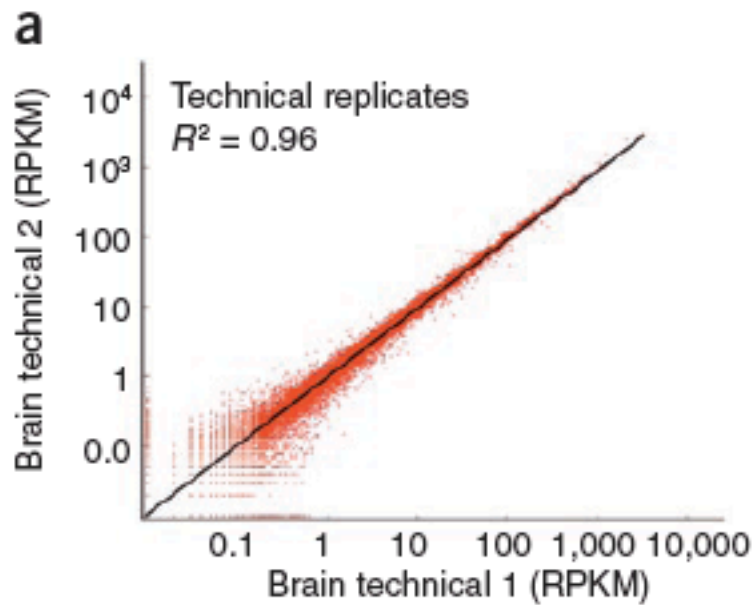
Gene 1 has 9 reads
Gene 2 has 3 reads

Gene 1 is 9kb
Gene 2 is 3kb

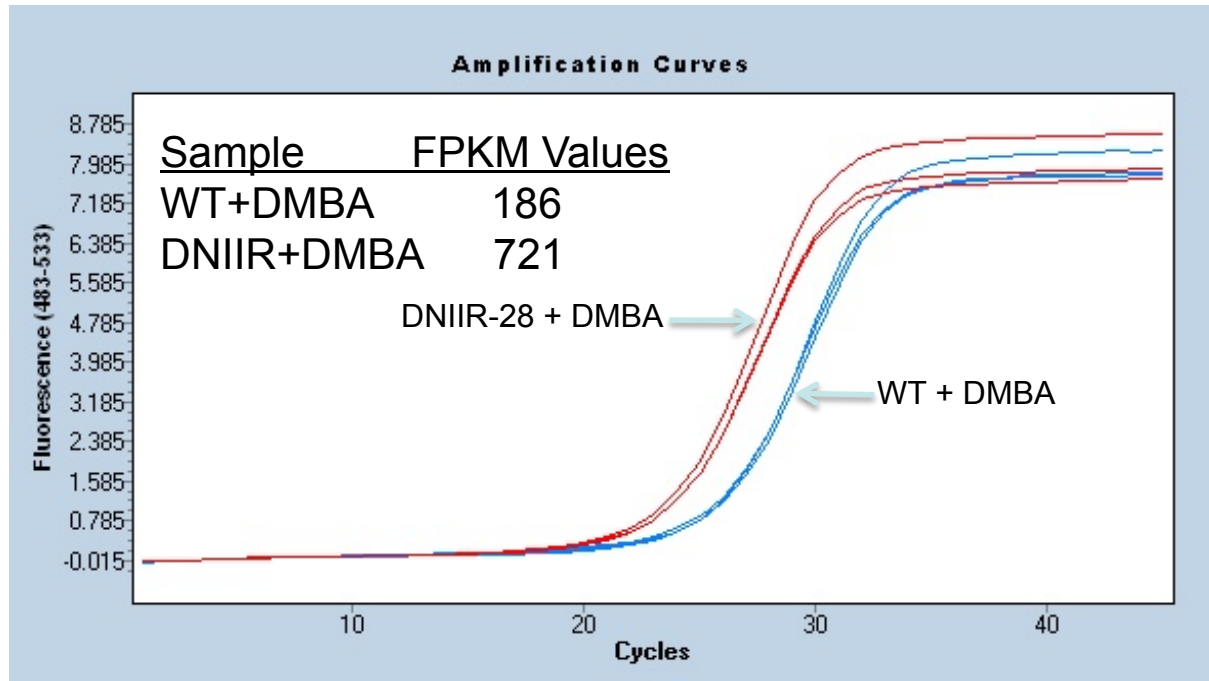
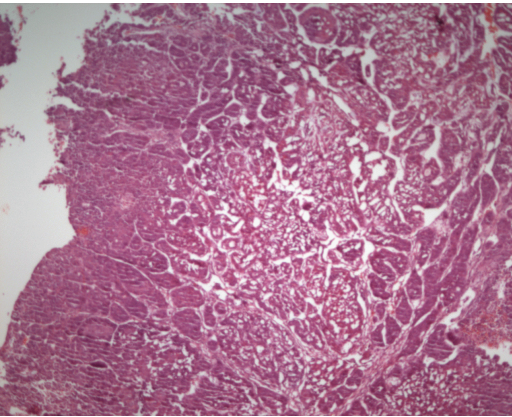
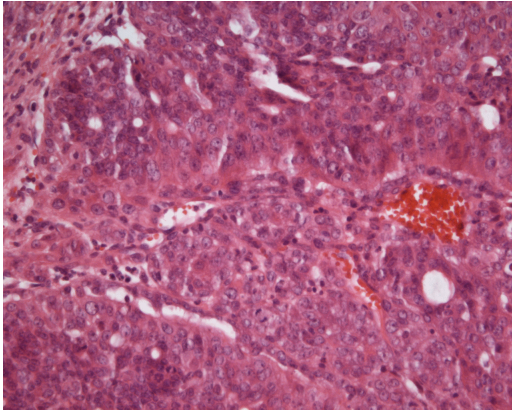
Normalize by length
e.g. $9\text{reads}/9\text{kb}=1$

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

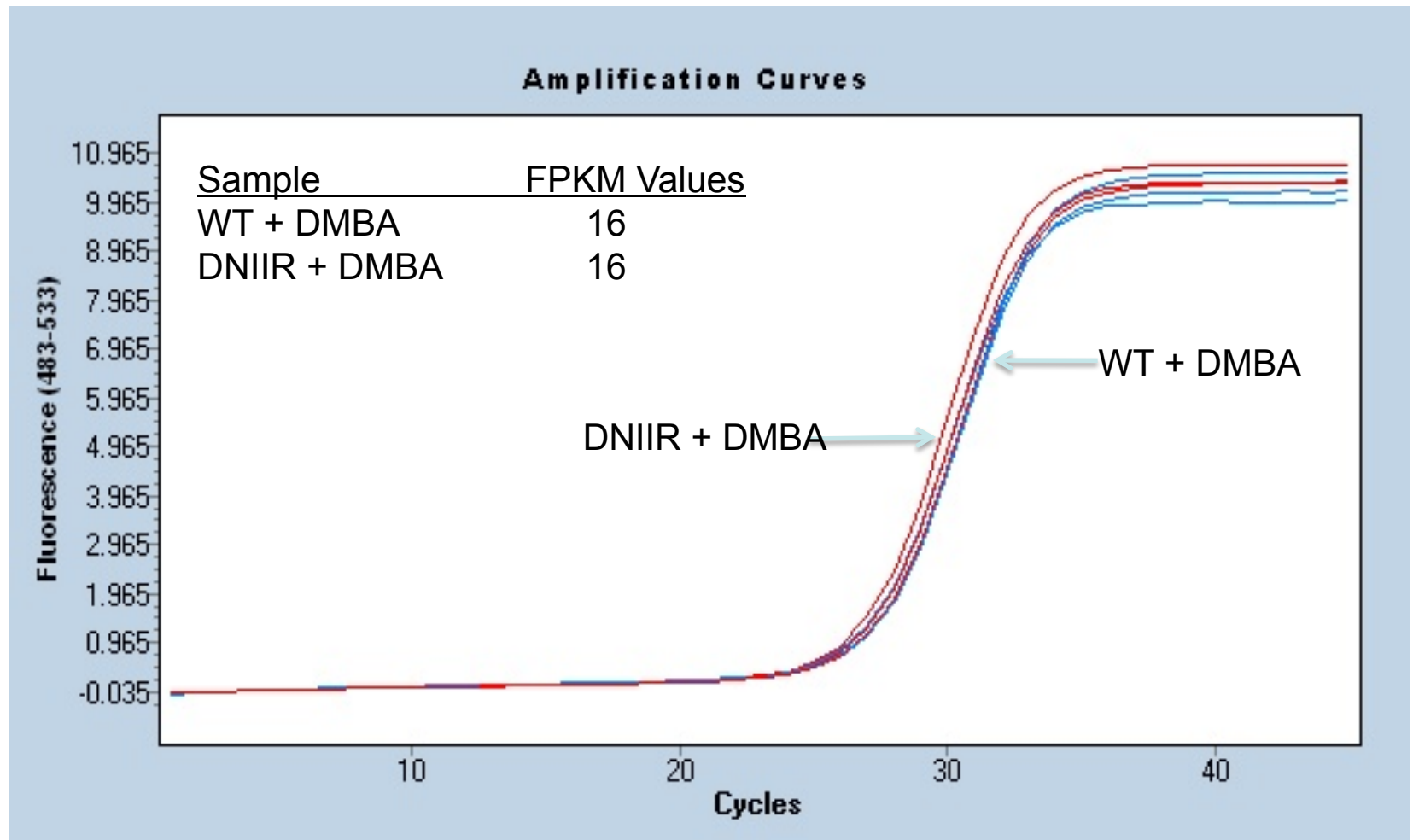
FPKM: fragment of reads per kilobase of exon model per million mapped reads (usually 25bp fragments).



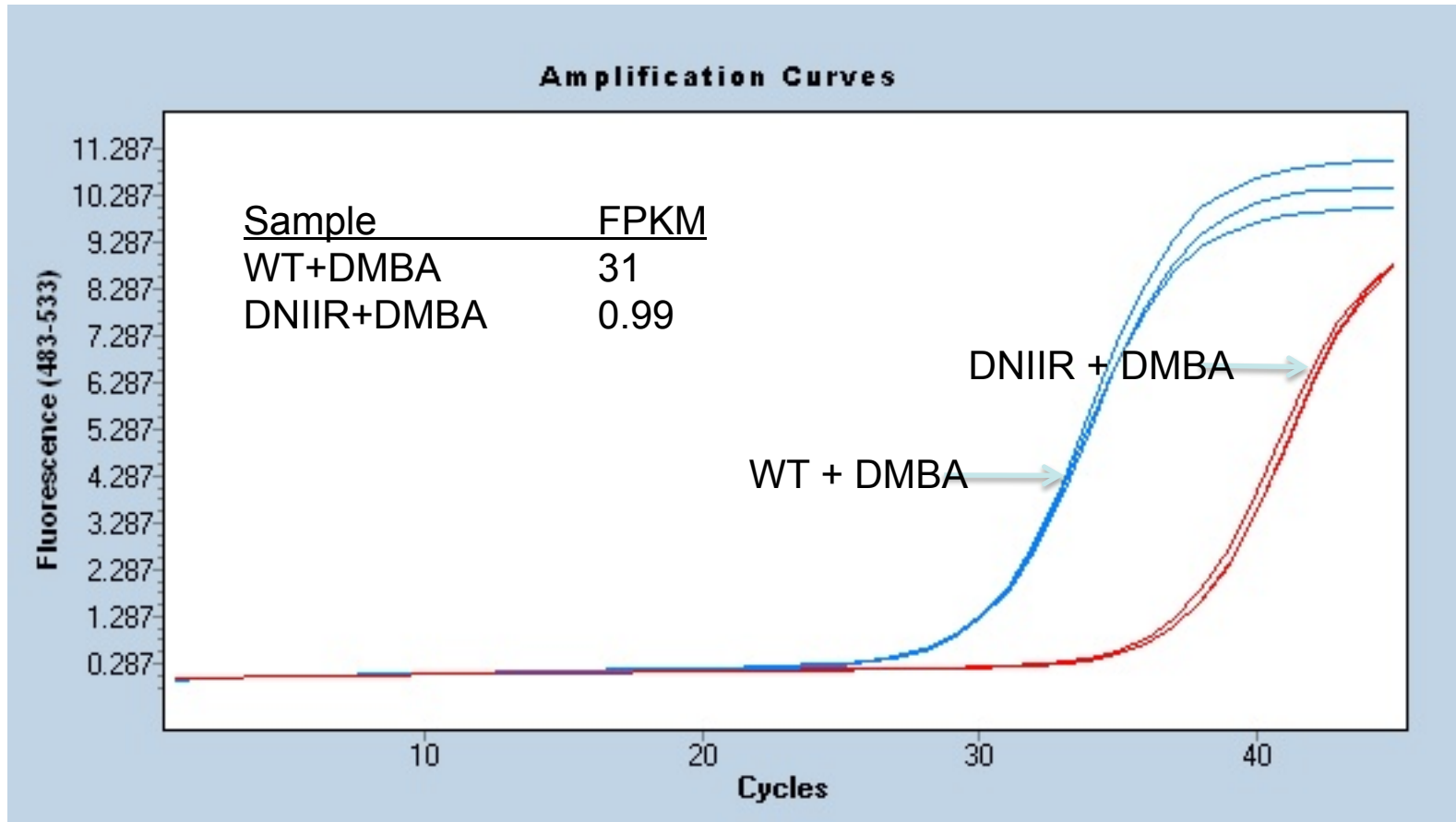
Keratin 8



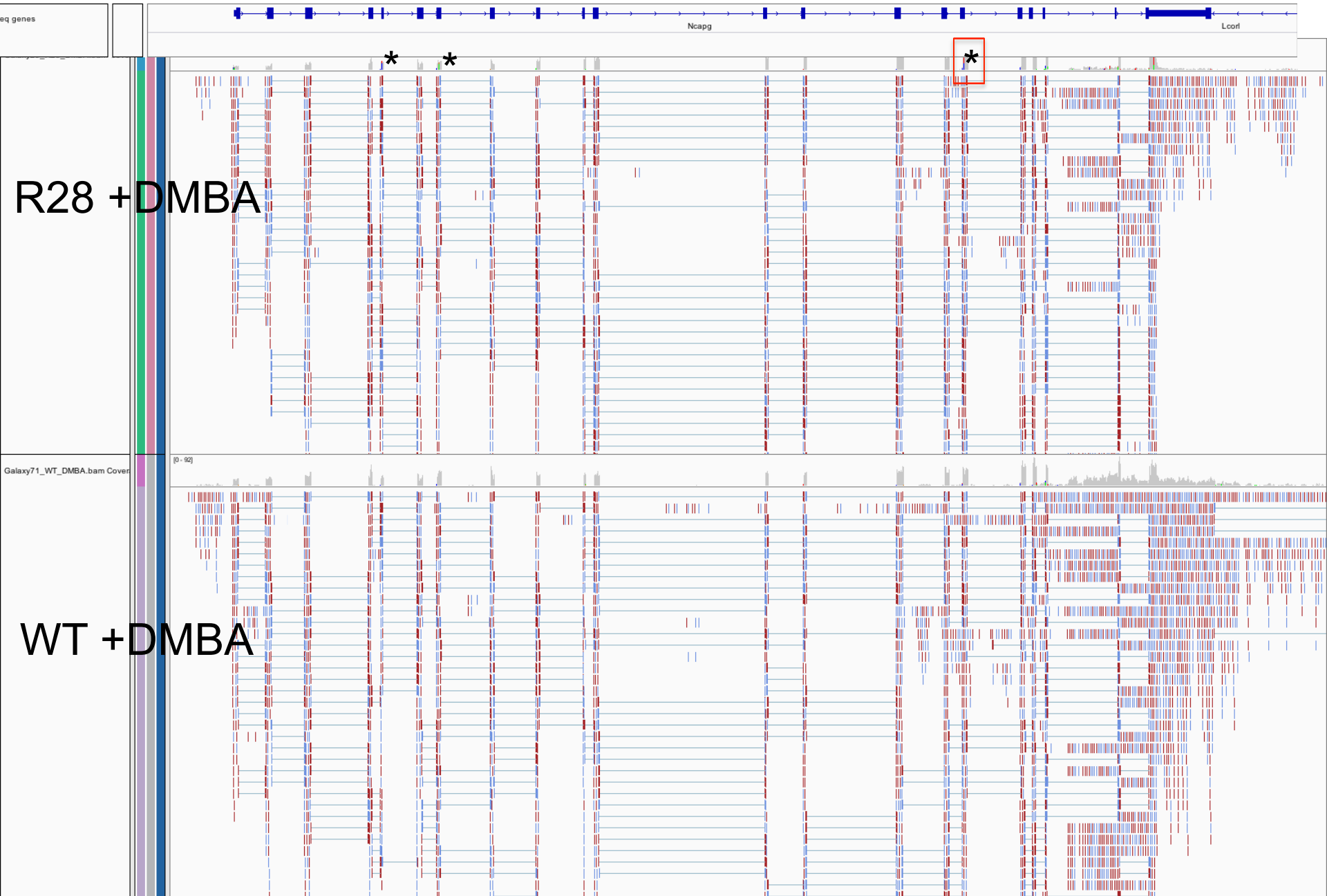
Lipase Maturation Factor 1 (Lmf1)



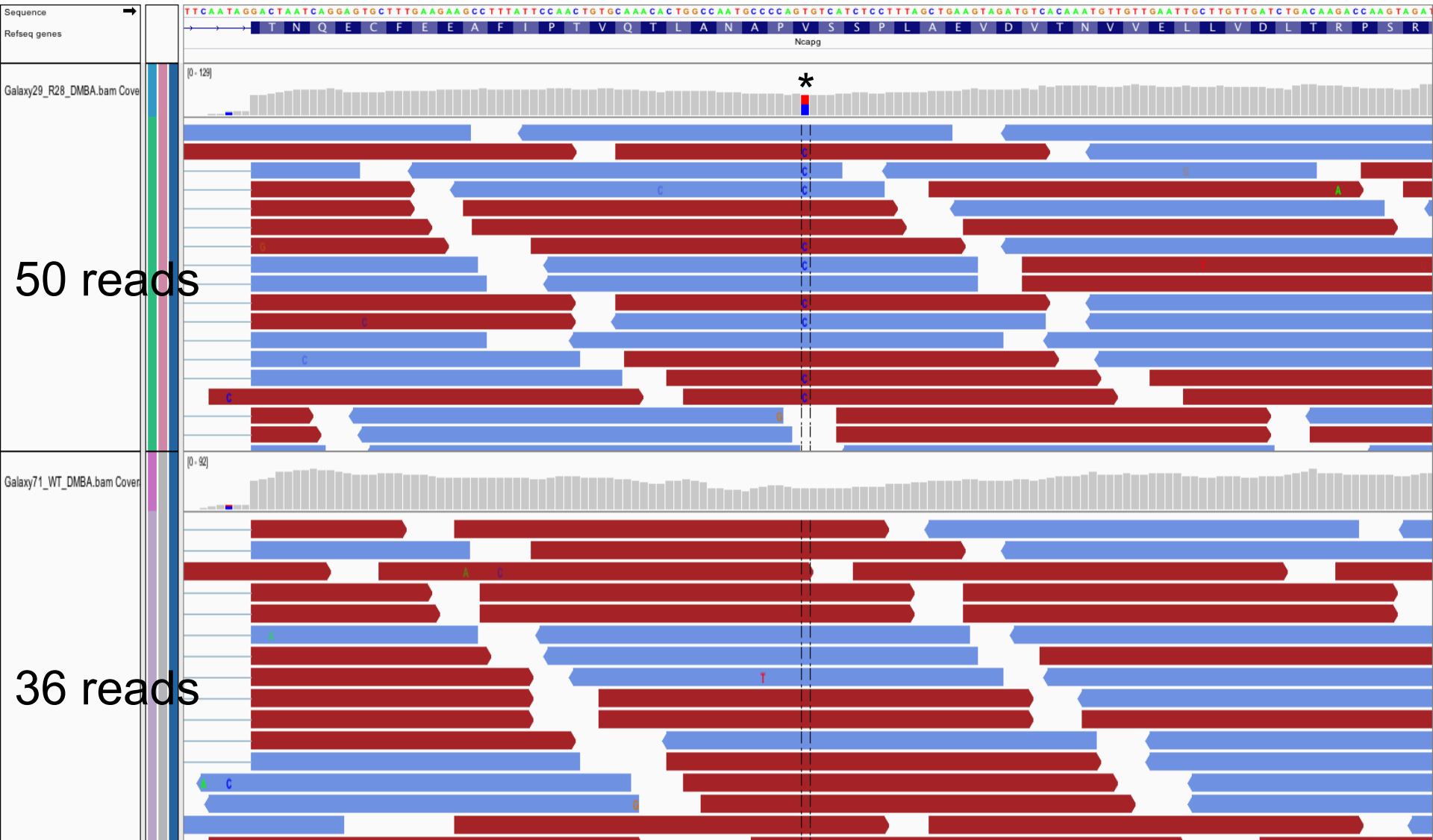
Lysophosphatidic acid receptor 3



Ncapg: Non-SMC condensin I complex, subunit G

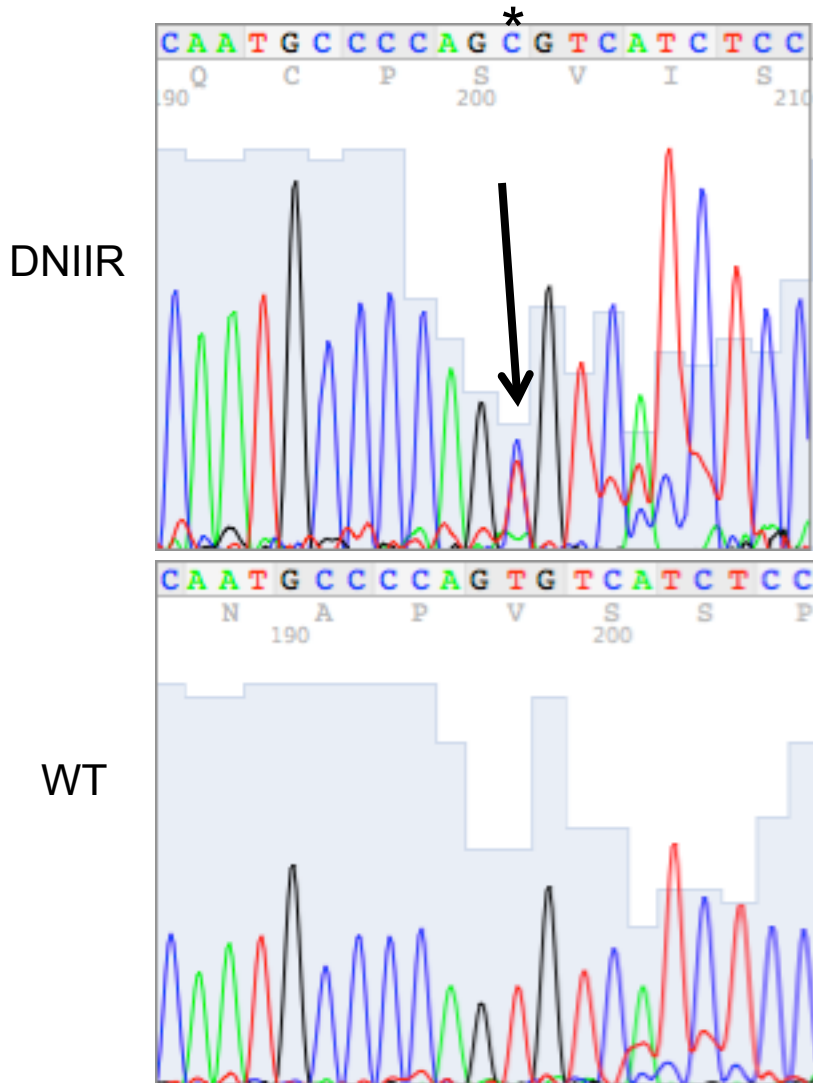


Exon 16 of Ncapg



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

Sequence Confirmation of Ncapg mutation

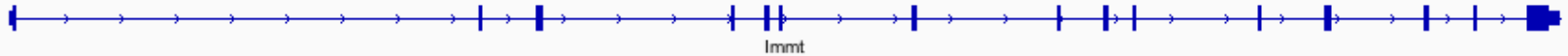


T>C mutation resulting in an Ala>Val change at position aa784 in the protein. The other mutations were a polymorphic T>C change at aa242 and an A>G change at aa347 resulting in a non-synonymous change from Arg>Lys.

Alternative Exon Usage



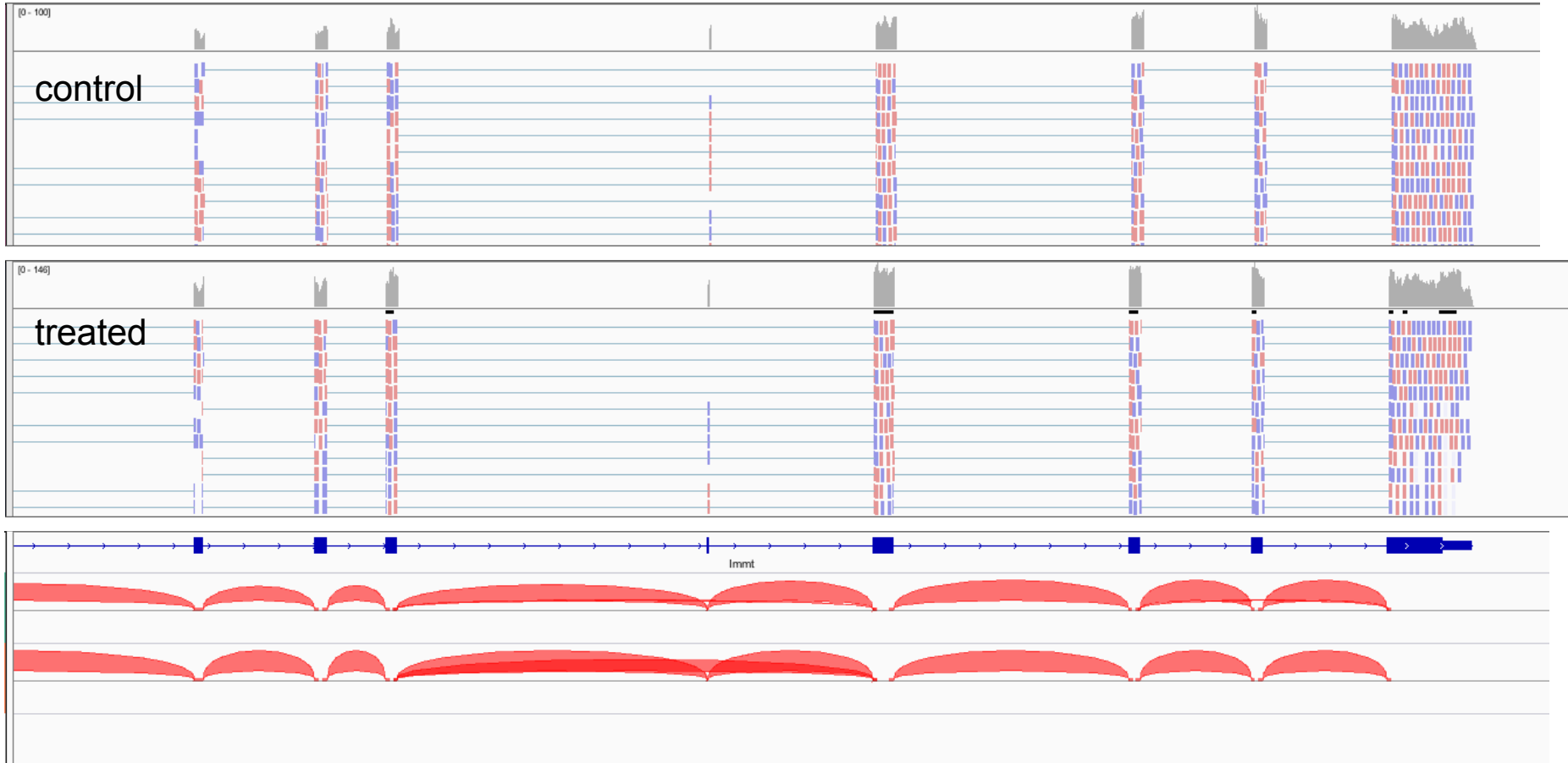
Trim24



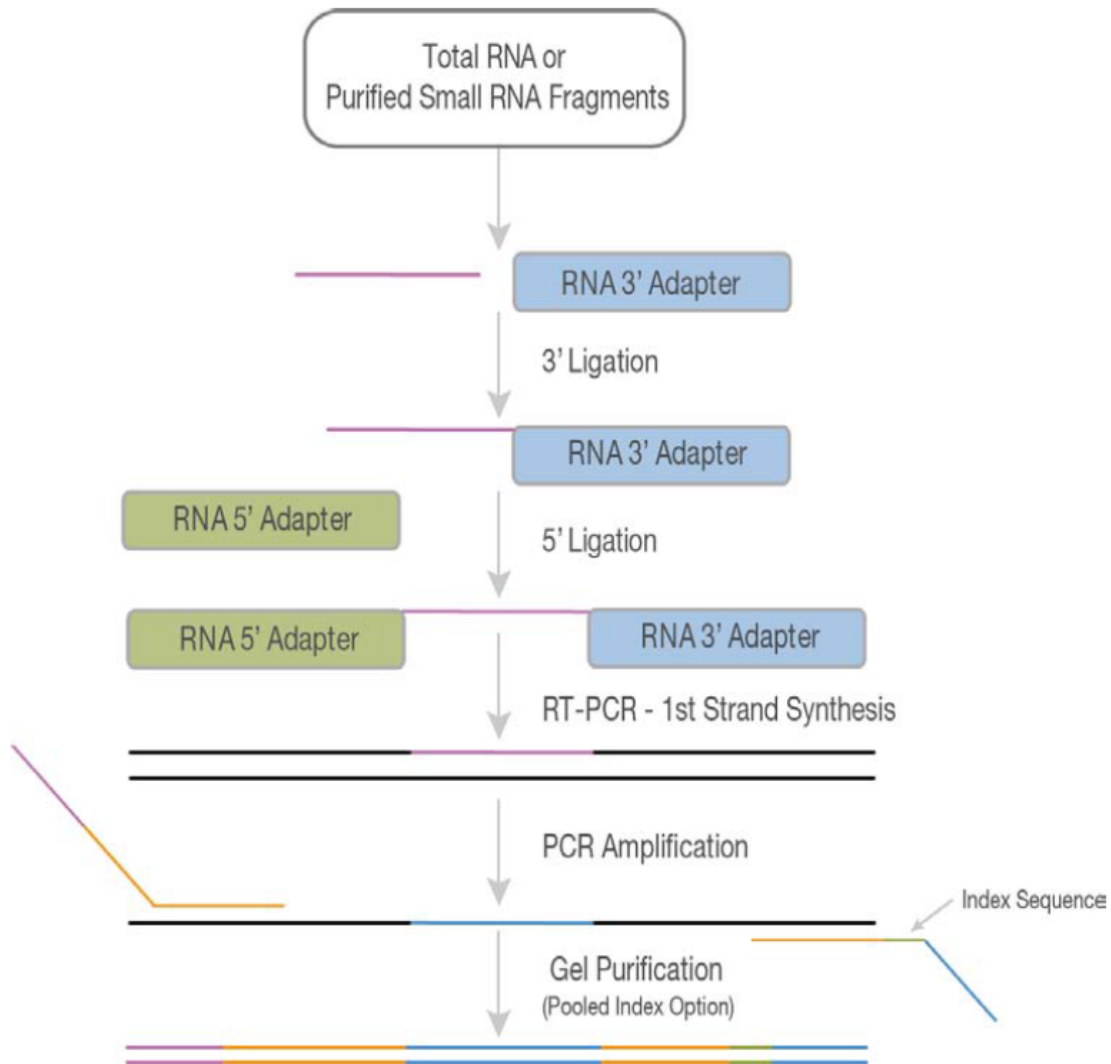
Immt



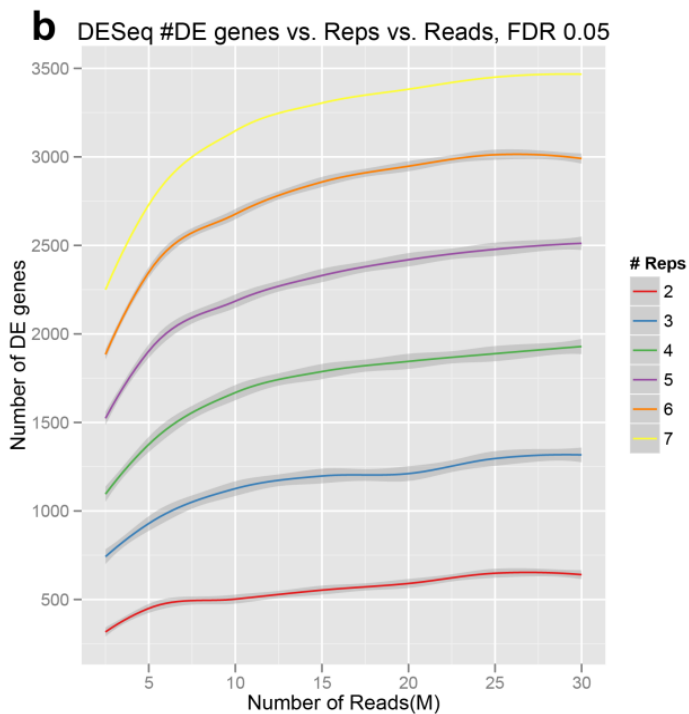
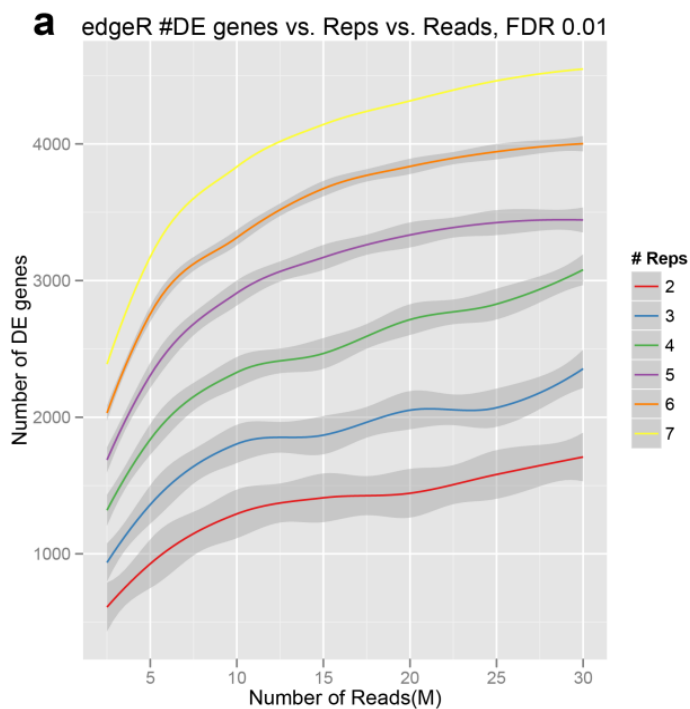
Alternative splicing



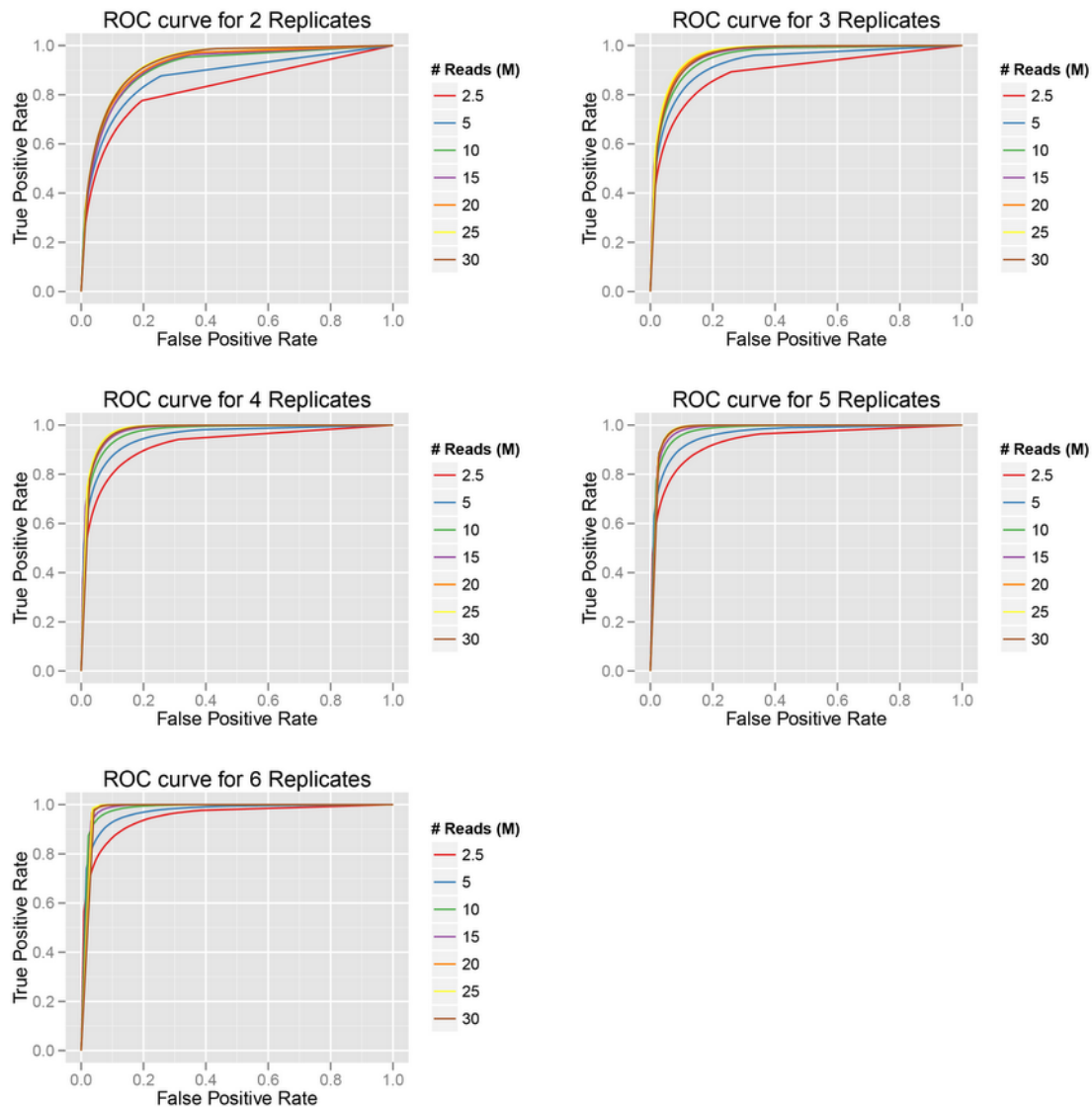
microRNA Seq



Sequencing Depth v. Biological Replicates

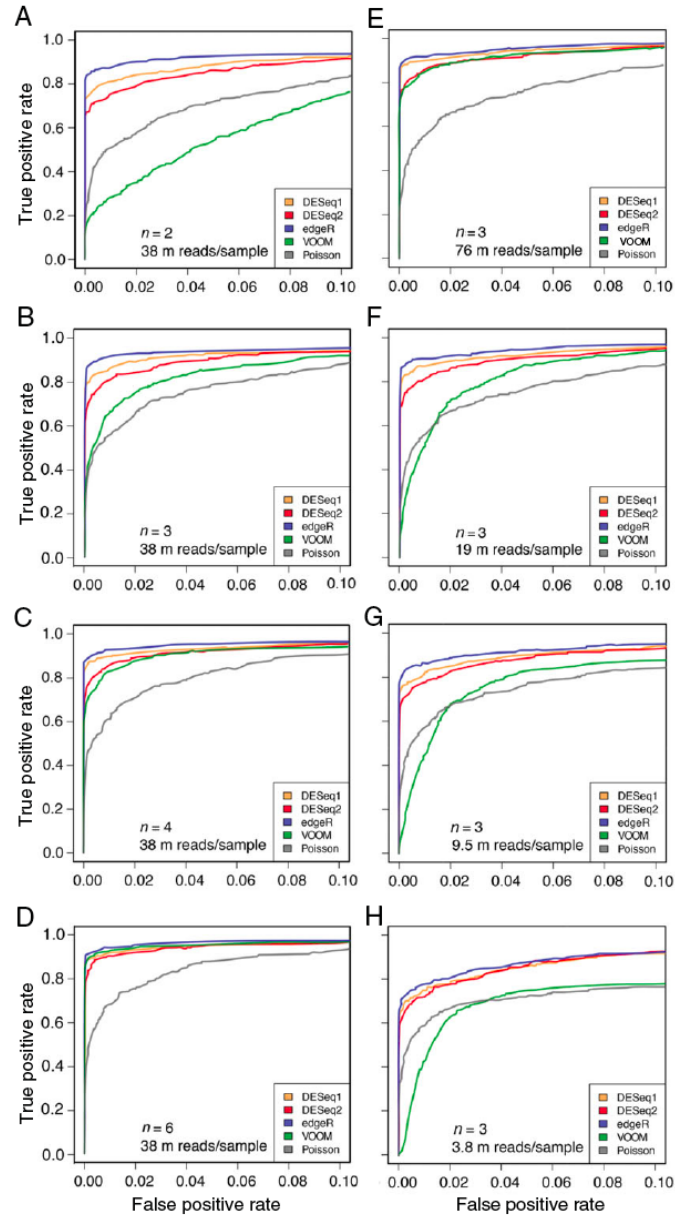


ROC curves with Replicates v. Depth



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 ≥ 1 Tb of information.
- Personalized medicine via genome sequencing is **HERE**.