Sequencing DNA

Human Whole Genome Sequencing

- Initial Ref Sequence $300 million and took about a decade. (Draft reported in 2001)
- Craig Venter's Genome for ~$10 million. (Pub Oct 2007).
- Yoruban from Nigeria in 8 weeks for $250,000. Approx. 30X coverage. (Pub Nov. 2008)
- Han Chinese in 8 weeks for ~$500,000 at approx. 36X coverage.
- Korean Individual at 27.8X (Pub July 2009).
- Female patient with AML. Sequenced normal and tumor from same patient. 98 full runs on GAI for tumor DNA and 34 full runs for normal skin cell DNA. ~1.5 years to compete both genomes.
- As of January 2012 a human genome can be sequenced for about $5,000 at an average read depth of 30X in 10 days

Applications

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

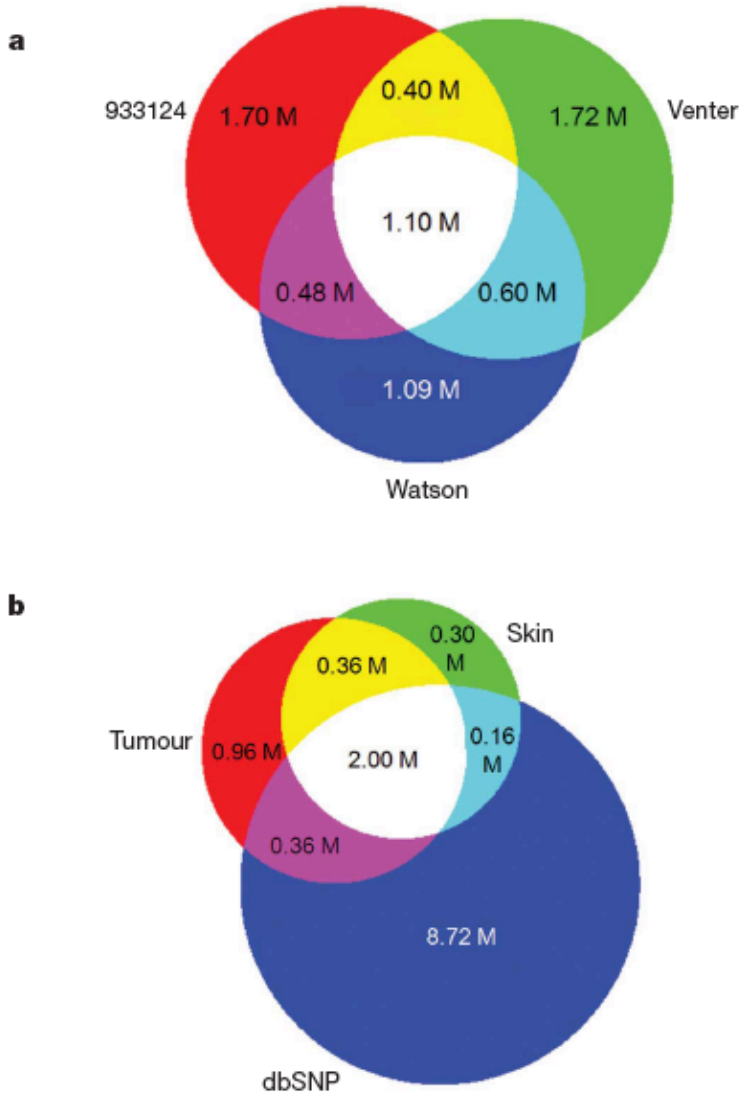
Genomic Sequence of the AML Genome: The Numbers

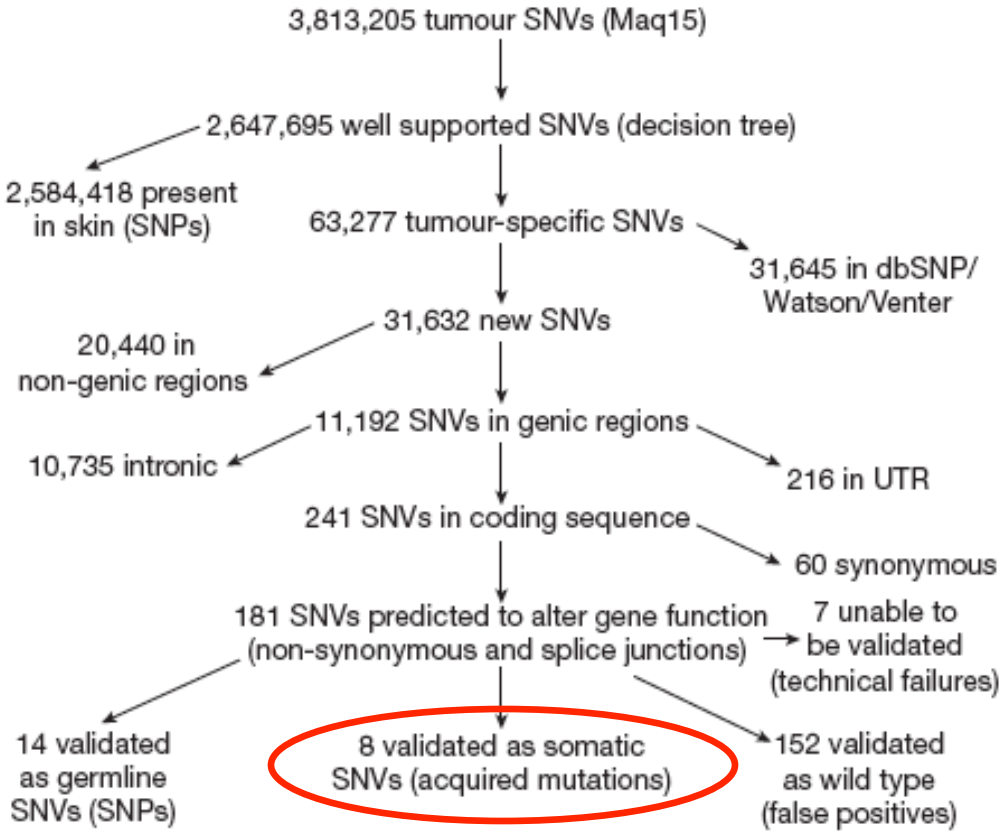
Table 1 | Tumour and skin genome coverage from patient 933124

	Tumour	Skin
Libraries	4	3
Runs	98	34
Reads obtained	5,858,992,064	2,122,836,148
Reads passing quality filter	3,025,923,365	1,228,177,690
Bases passing quality filter	98,184,511,523	41,783,794,834
Reads aligned by Maq	2,729,957,053	1,080,576,680
Reads unaligned by Maq	295,966,312	138,276,594
SNVs detected with respect to hg18 (no Y)	3,811,115	2,918,446
SNVs (chr 1–22) detected with respect to hg18	3,681,968 (100.0%)	2,830,292 (100.0%)
SNVs also present in dbSNP	2,368,458 (64.3%)	2,161,695 (76.4%)
SNVs also present in Venter genome	1,499,010 (40.7%)	1,383,431 (48.9%)
SNVs also present in Watson genome	1,573,435 (42.7%)	1,456,822 (51.5%)
SNVs not in dbSNP/Venter/Watson	1,223,830 (33.2%)	591,131 (20.9%)
SNVs not in dbSNP/Venter/Watson/skin	925,200 (25.1%)	–
HQ SNPs	46,494 (100.0%)	46,572 (100.0%)
HQ SNPs where reference allele is detected	42,419 (91.2%)	38,454 (82.6%)
HQ SNPs where variant allele is detected	43,164 (92.9%)	39,220 (84.2%)
HQ SNPs where both alleles are detected	42,415 (91.2%)	38,454 (82.6%)

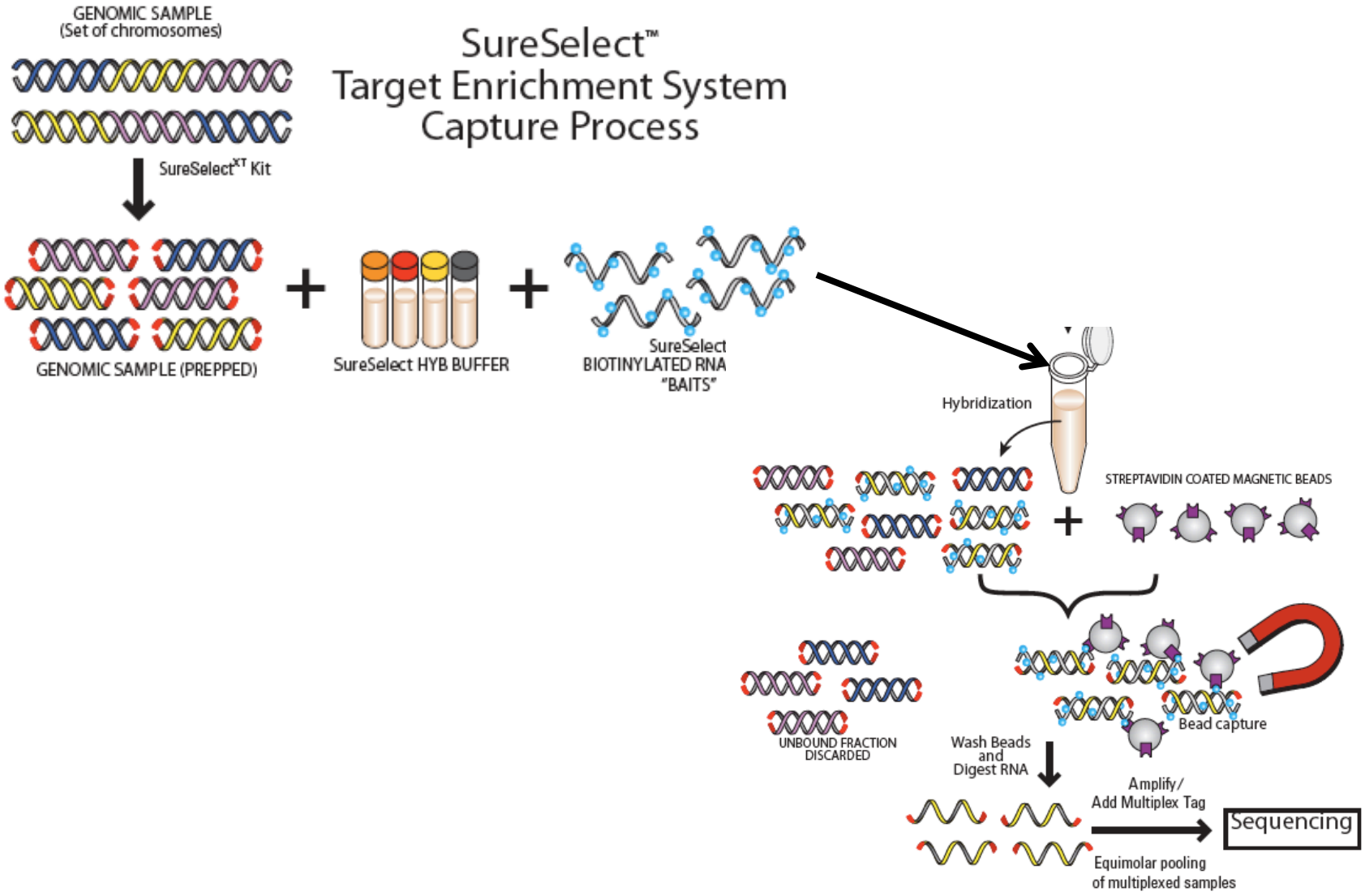
Assessments are shown of the haploid and diploid coverage of the tumour and skin genomes from AML patient 933124.
Chr, chromosome; hg18, human genome version 18; HQ, high quality.

AML:Comparisons





SureSelect Exome Capture

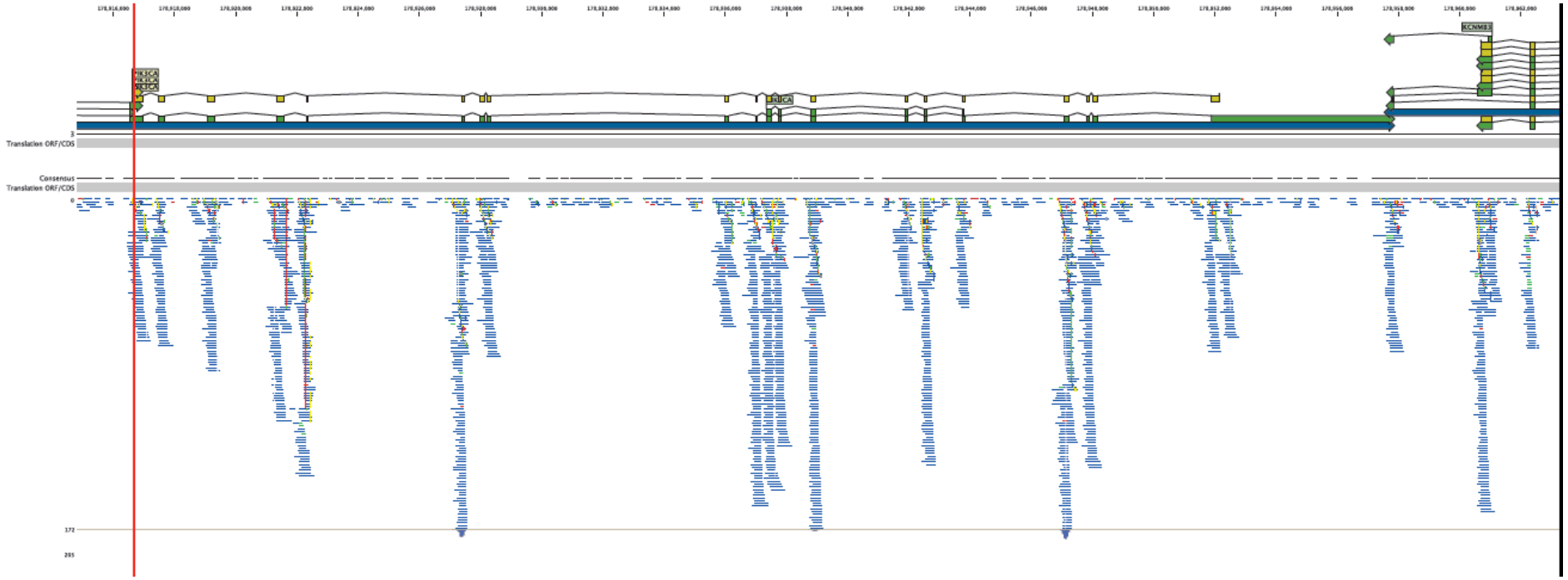


Disease Genes Discovered by Direct Whole Exome Sequencing*

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

Exome Capture-PIK3Ca



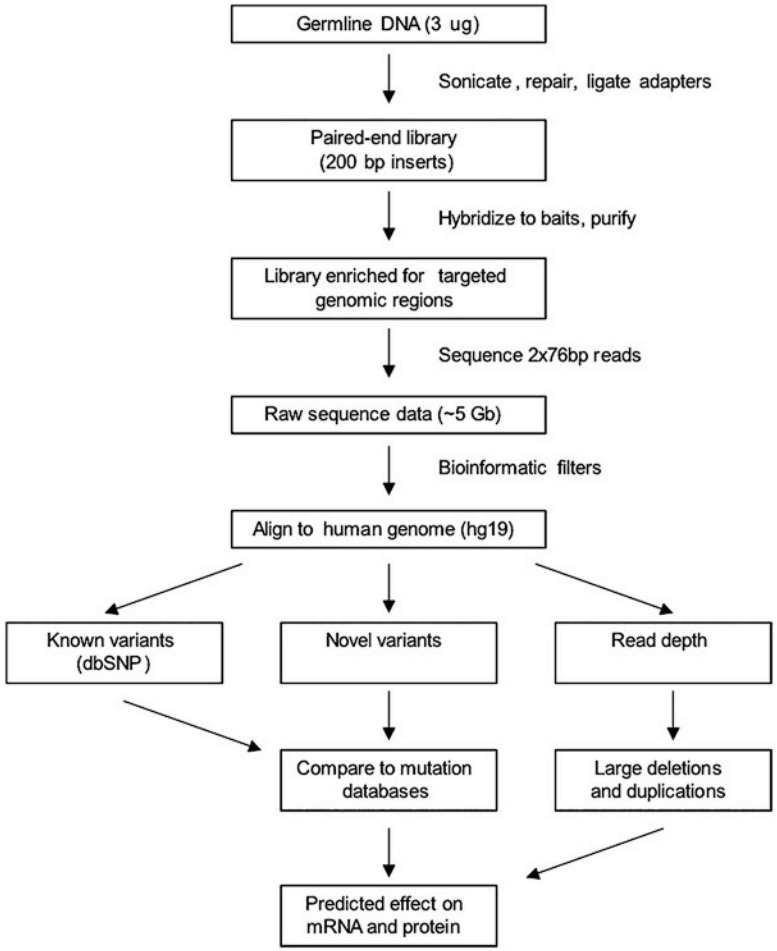
Courtesy of P. Buckhaults

Targeted Re-sequencing

The ability to capture specific sequences in the genome

Microarrays
Long range PCR
Solution capture on Biotin labeled oligos
HaloPlex

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

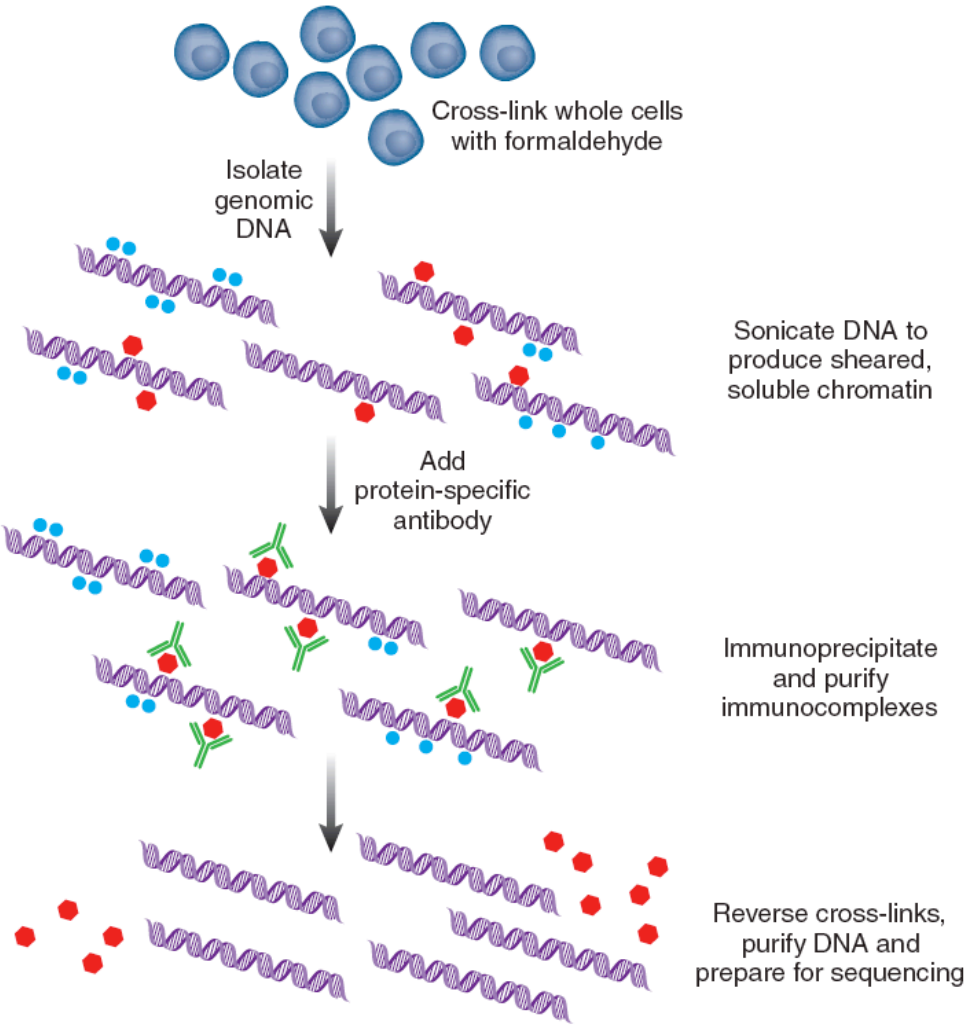


Walsh T et al. PNAS 2010;107:12629-12633

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



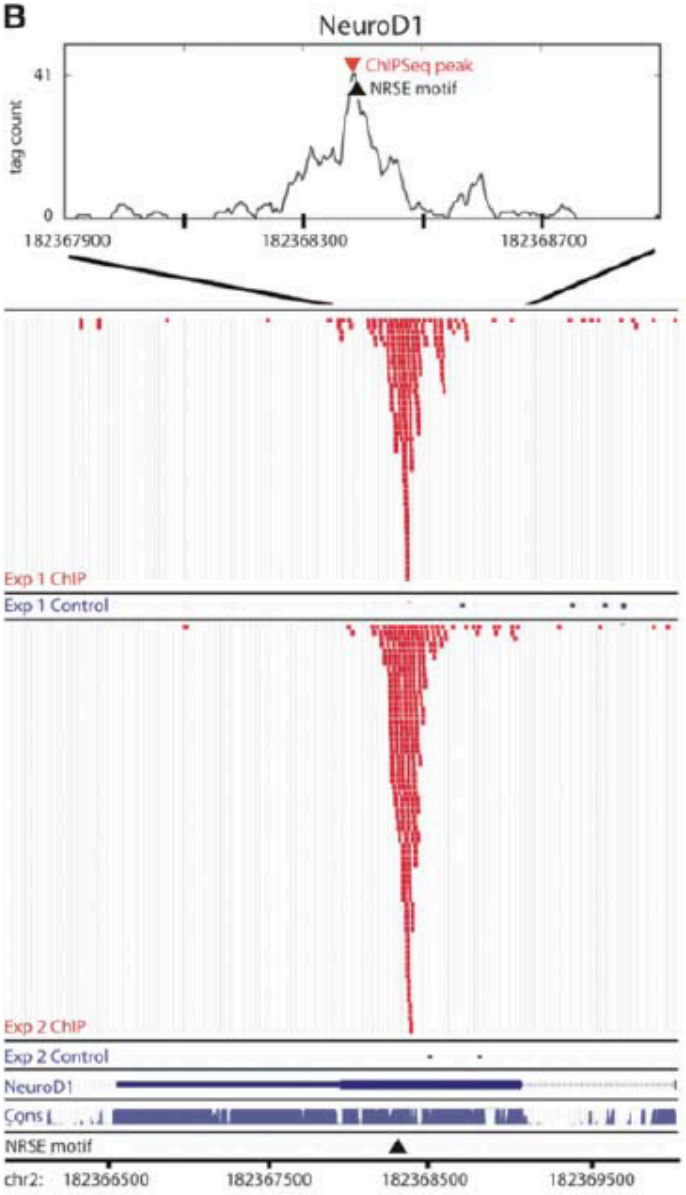
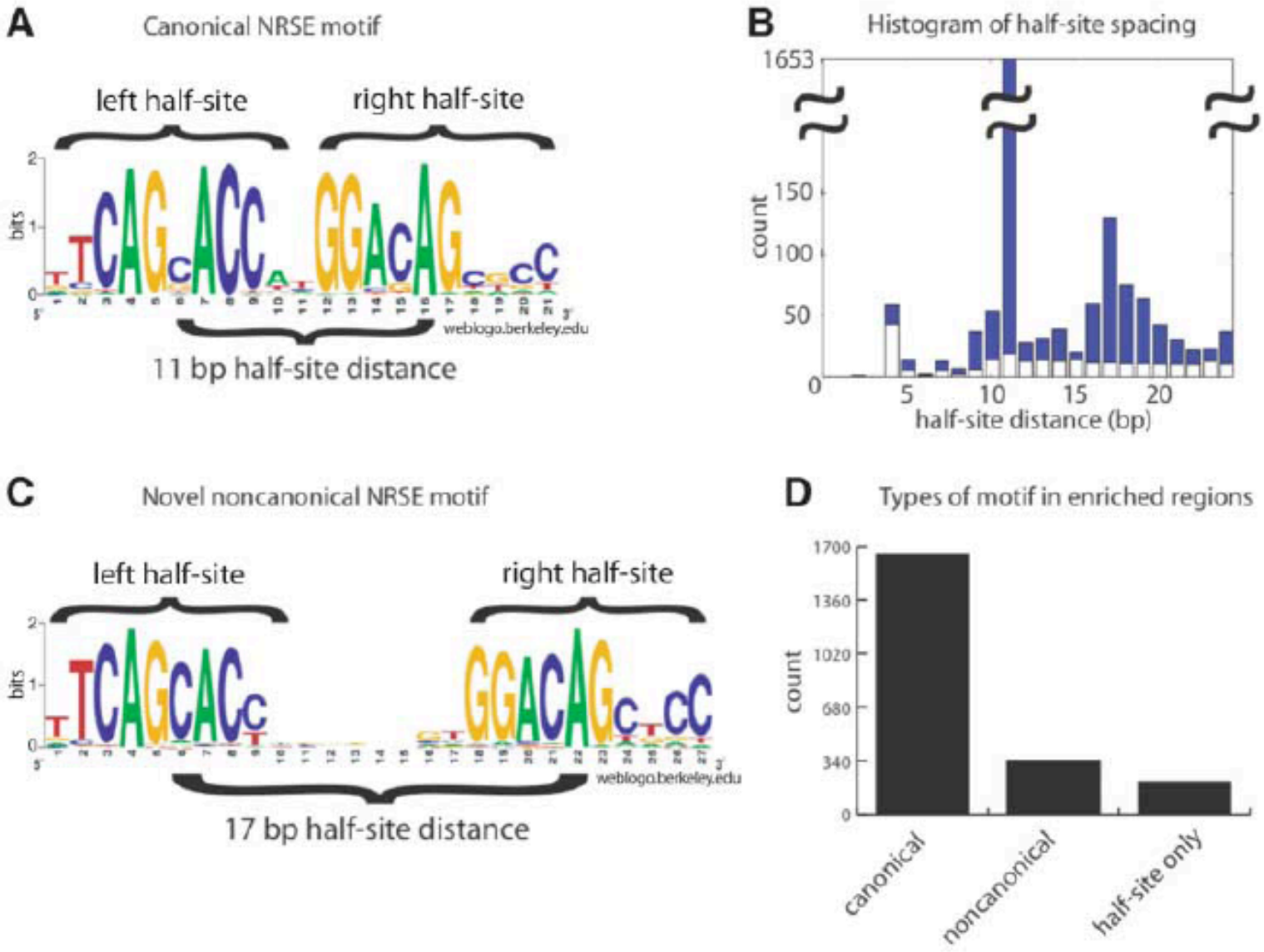


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

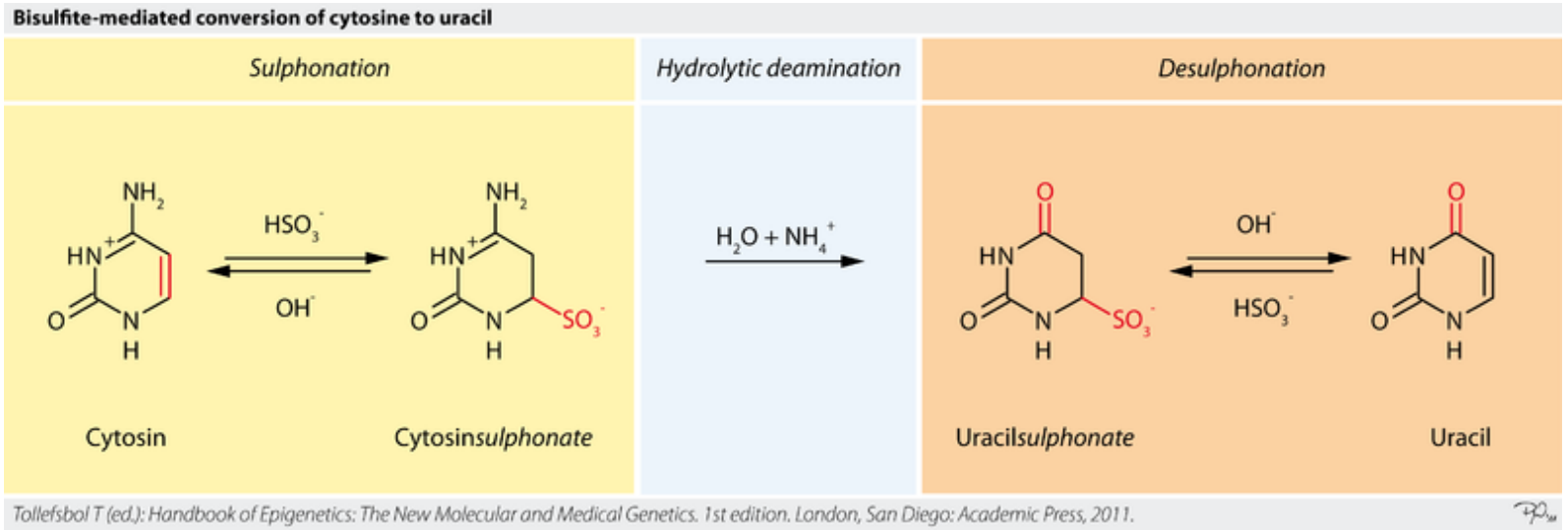
ChIP-Seq



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

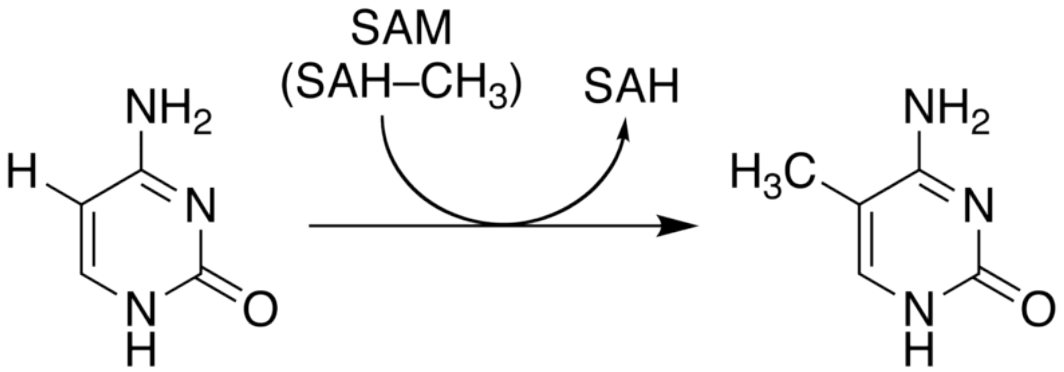
Methylation profiling

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

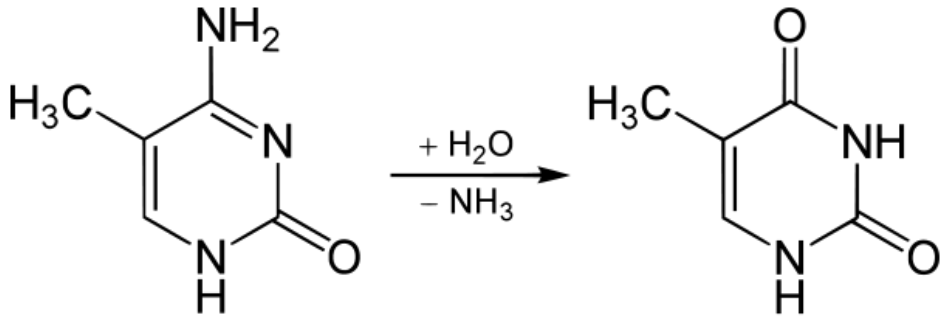


Tollefsbol T (ed.): Handbook of Epigenetics: The New Molecular and Medical Genetics. 1st edition. London, San Diego: Academic Press, 2011.

Cytosine to 5-Methylcytosine to Thymine conversion



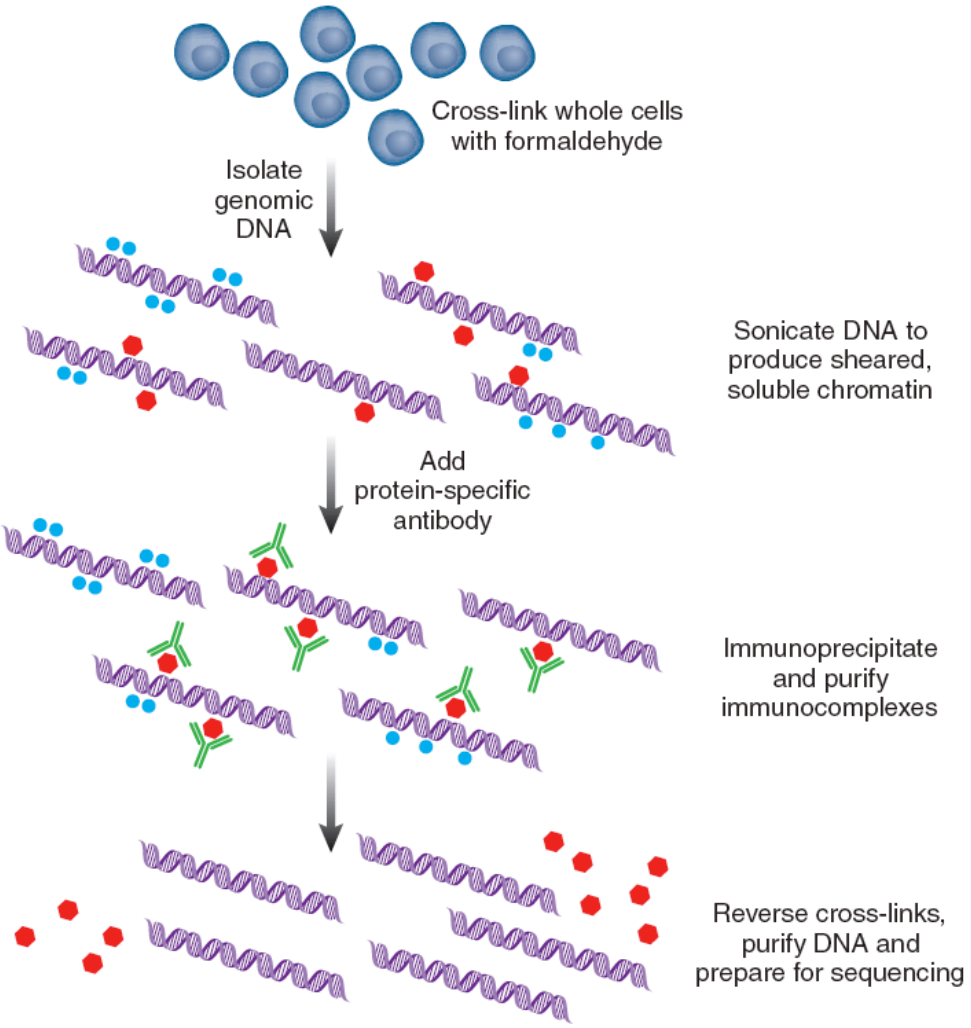
Cytosine 5-Methylcytosine



5-Methylcytosine Thymine

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

MeDIP-Seq



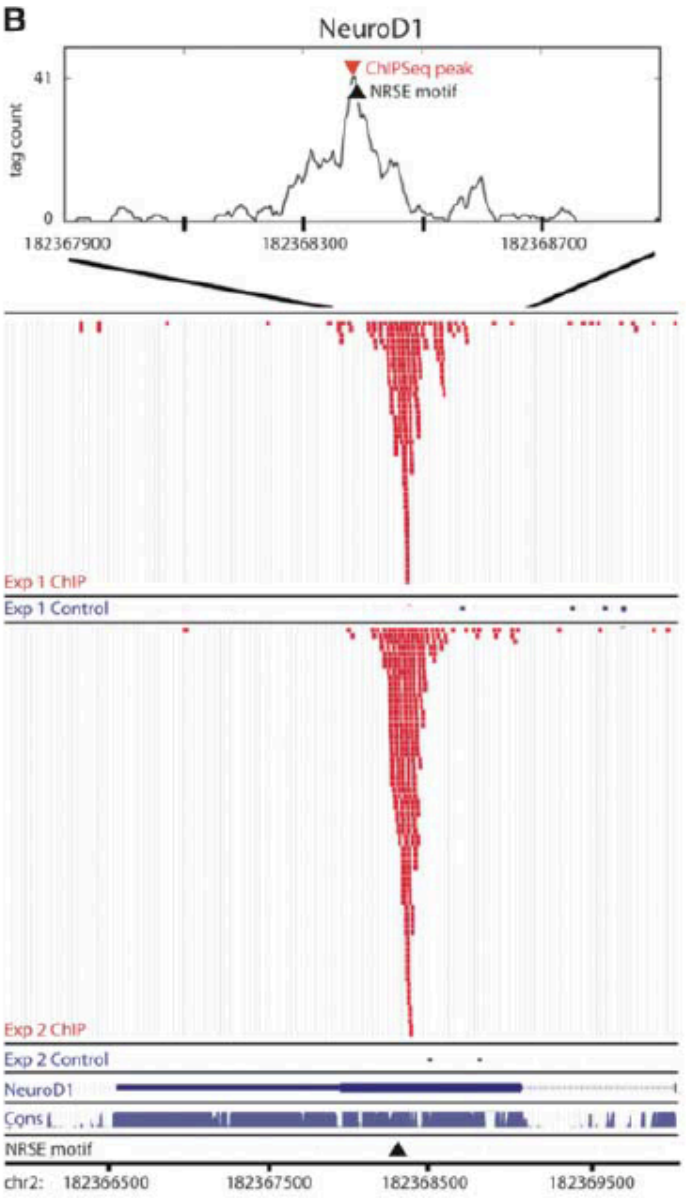


Figure 1 | Workflow of Chip-seq. DNA and proteins are cross-linked and purified; then bound DNA is analyzed by massively parallel short-read sequencing.

ogenic microorganisms that literally share our body space" (Lederberg and McCray 2001). Initial efforts to determine the numbers of microbes in a community and their phylogenetic relationships comprised analyzing the relatively well-conserved 16S rRNA genes in mixtures of organisms (Woese and Fox 1977; Stahl

[1]A complete list of authors and affiliations appears at the end of the paper, before the Acknowledgments section. See also, http://nihroadmap.nih.gov/hmp/members.asp.
[2]Corresponding author.
E-mail jane_peterson@nih.gov; fax (301) 480-2770.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.096651.109. Freely available online through the *Genome Research* Open Access option.

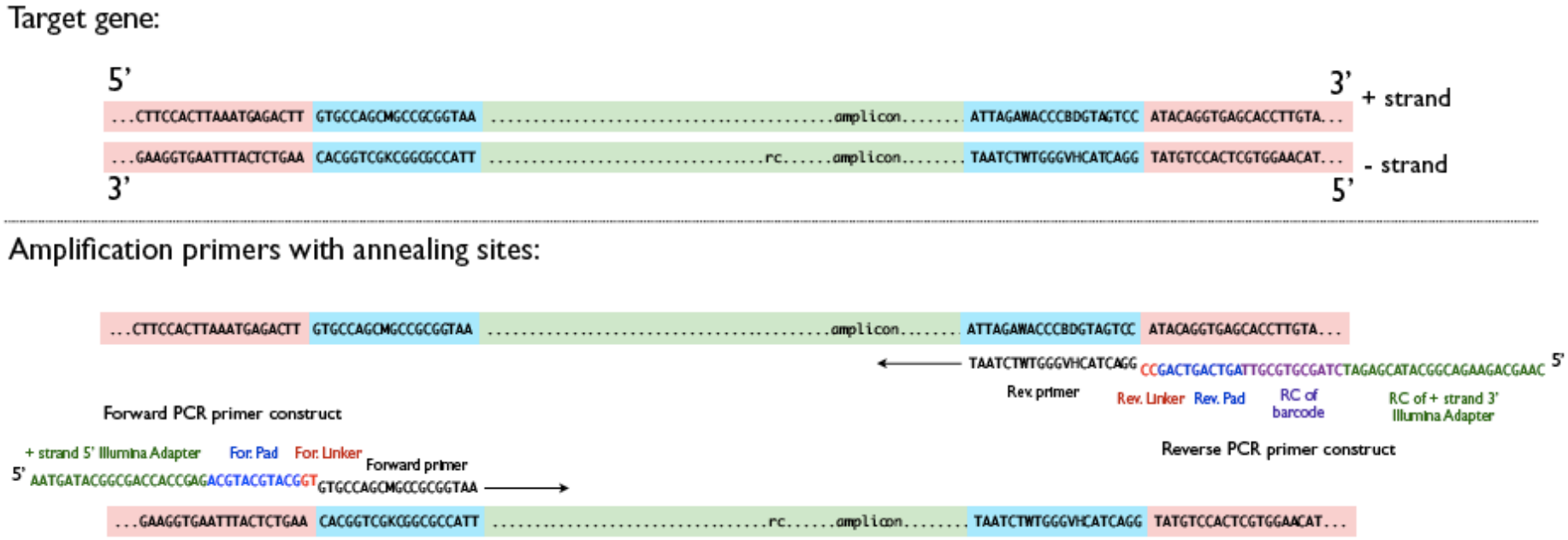
The early studies examining the microbiome stimulated ir undertaking a large-scale investigation of the human i microbiome. An international meeting was held in Par vember 2005 to discuss such an effort. This meeting, host French National Institute for Agricultural Research (IN chaired by Dusko Ehrlich, led to the recommendation th man Intestinal Metagenome Initiative (HIMI) be under define more completely the human intestinal microb health and disease. The meeting attendees also recommer an International Metagenome Consortium be formed together common efforts from around the world to acc the goals of the HIMI (http://human-microbiome.org).

would no longer define the biology at the site as
was done in order to reduce the number of exclu
make it, in the clinicians' opinion, possible to rec
There was concern that recruitment using a prot
volunteers who were "healthy" at each site (as
sample site experts) would have so many exclusi
recruitment would be very slow or impossible.

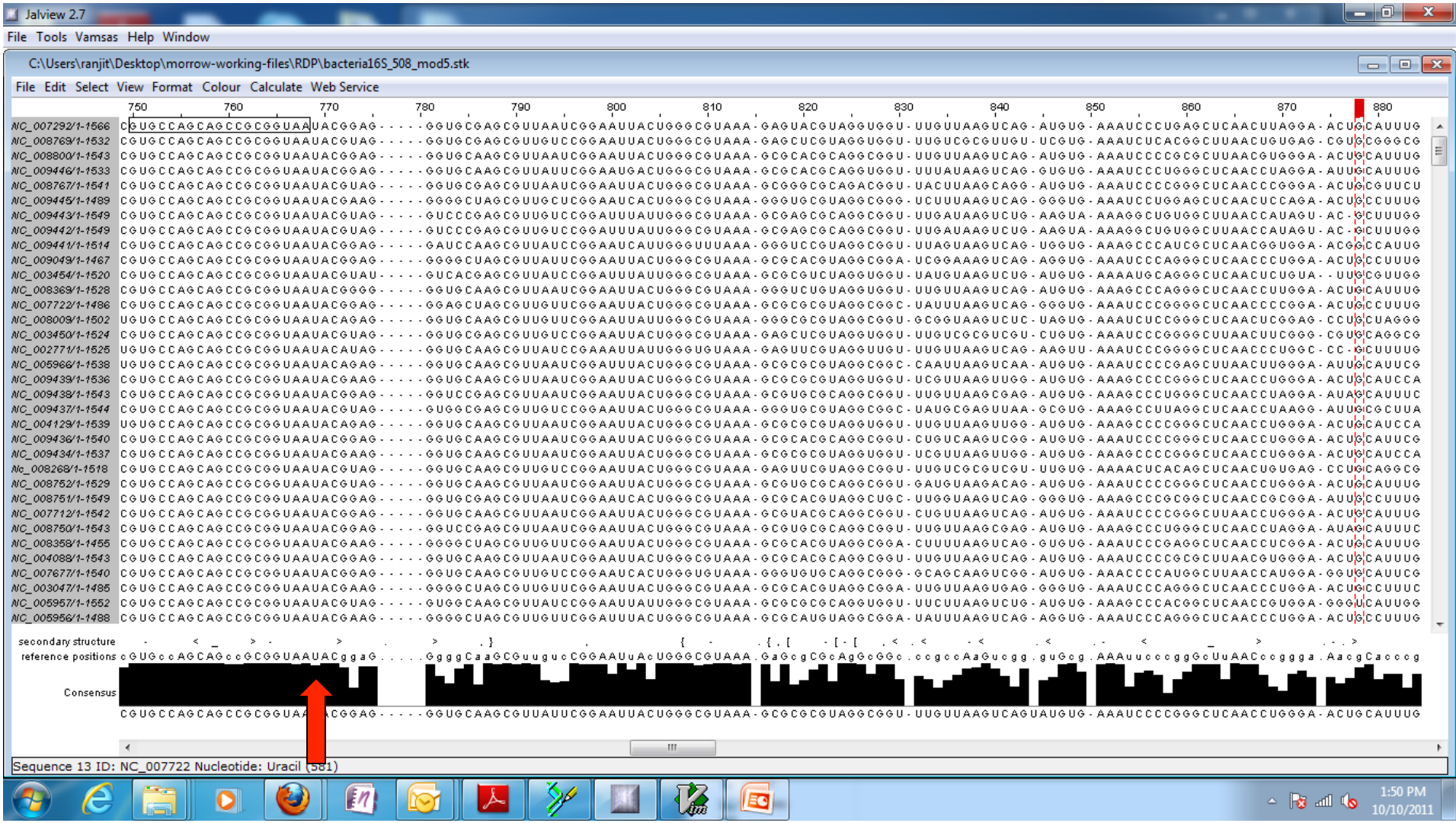
Special attention was paid to the informed cor
that potential sample donors were adequately infc
benefits and risks associated with participation in
resource" project. A template for an informed co
developed and then adapted for use at the twc
sampling took place (Baylor College of Medicine a
University; see http://hmpdacc.org/clinical.html for
Particular attention was given in the consent prc
ing donors about how their privacy would be pr
limitations of the available protections. Donors
that the microbiome data from the study of their sa

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

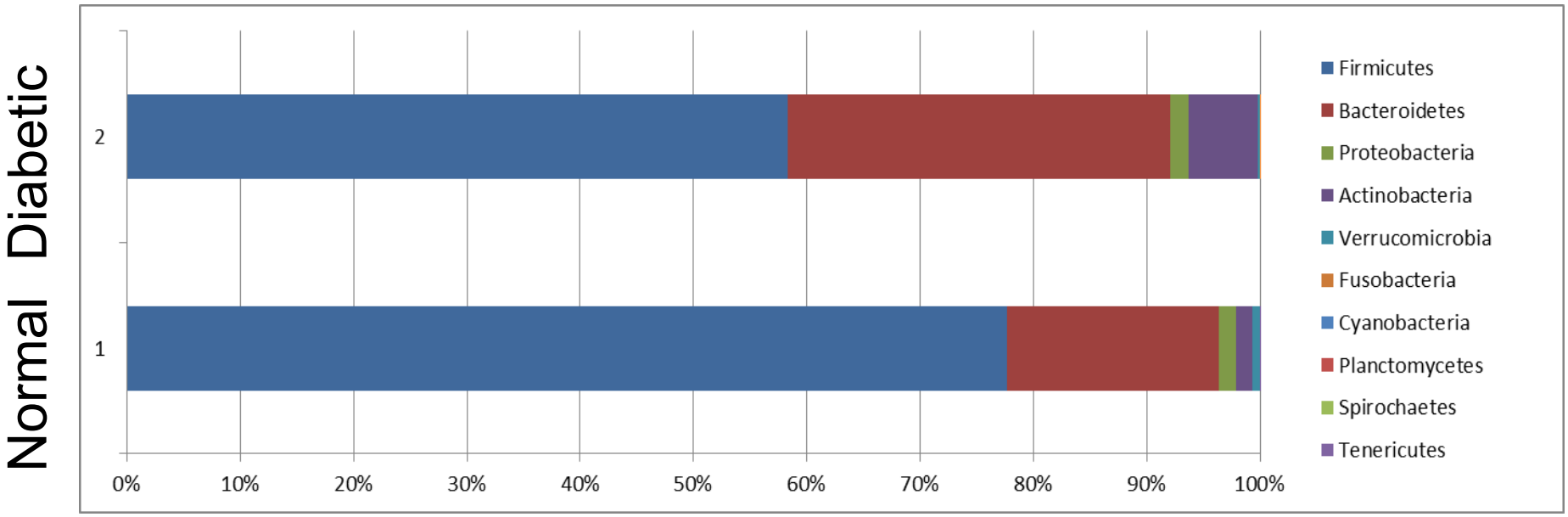
J. Gregory Caporaso[a], Christian L. Lauber[b], William A. Walters[c], Donna Berg-Lyons[b], Catherine A. Lozupone[a], Peter J. Turnbaugh[d], Noah Fierer[b,e], and Rob Knight[a,f,1]



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).