

Data Science for High-Throughput Sequencing

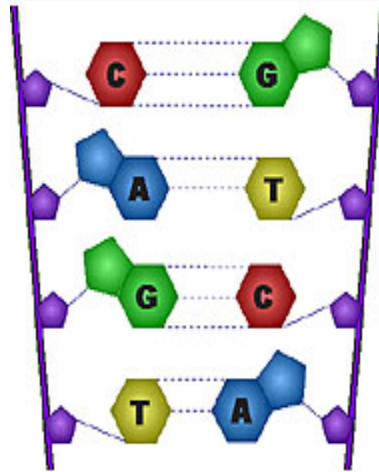
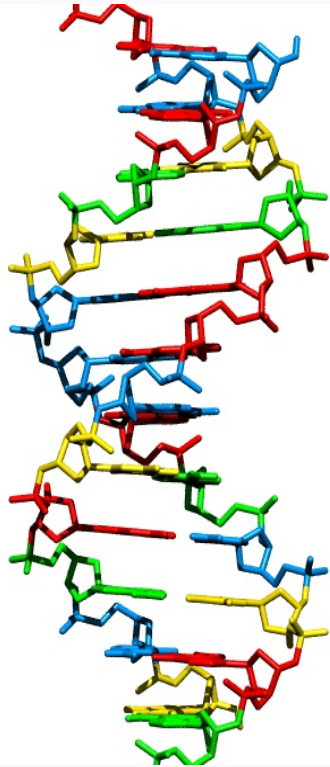
Lecture 1

Instructor:

David Tse

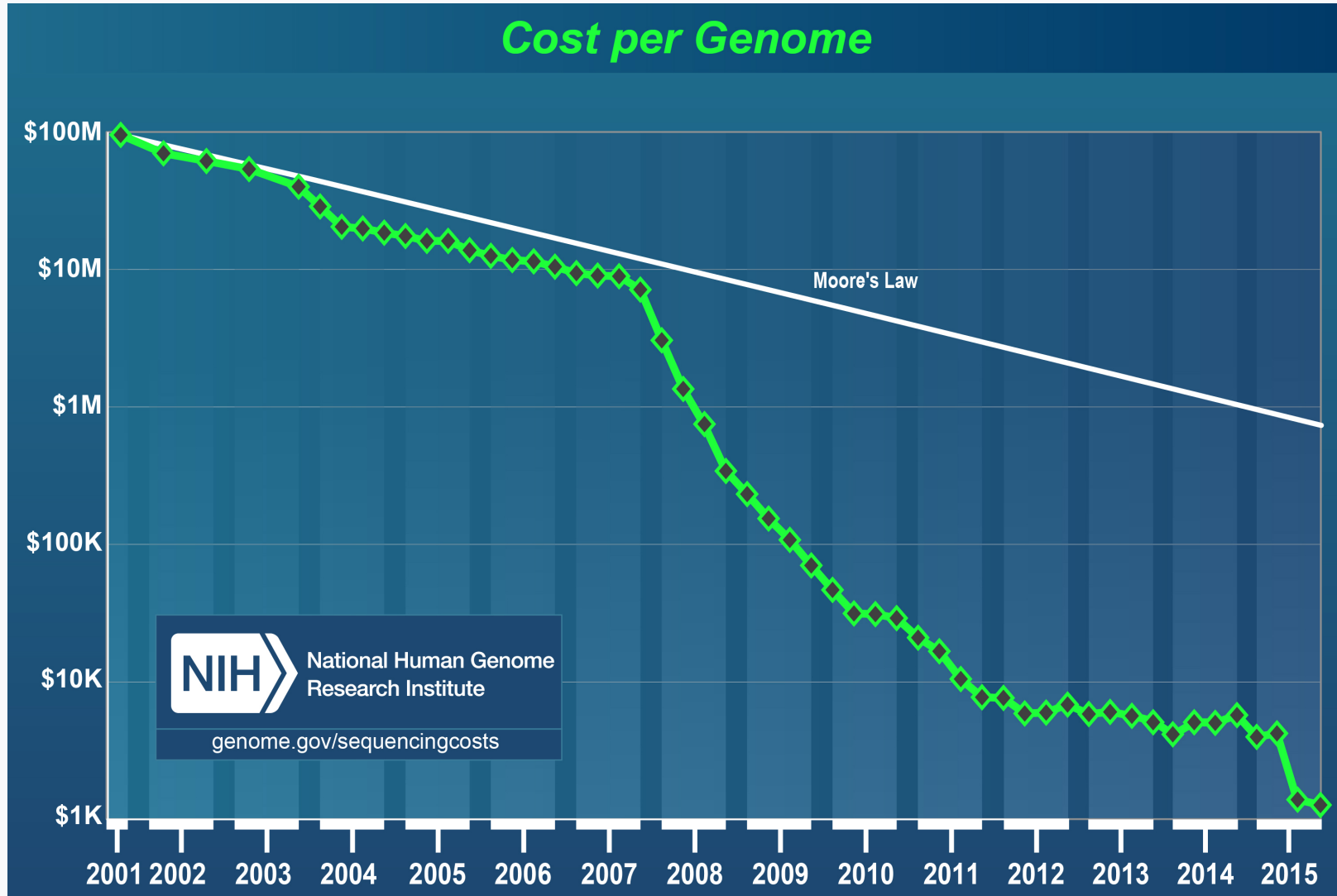
dntse@stanford.edu

The Genome

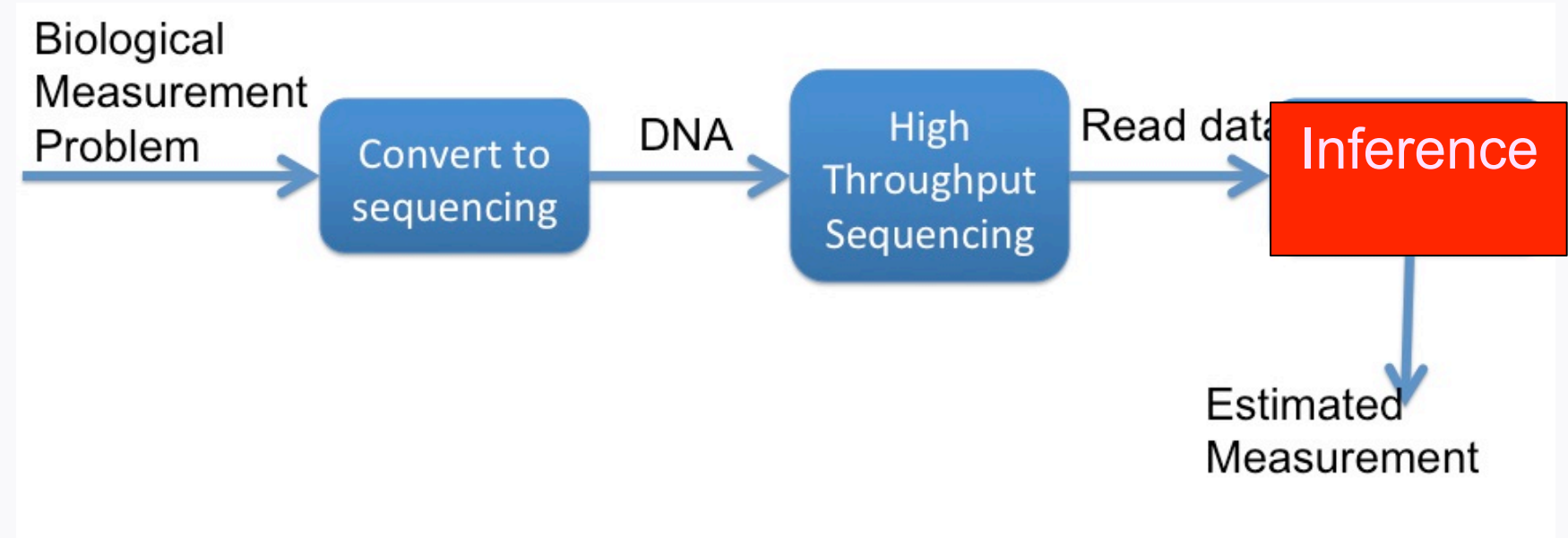


...ACGTGACTGAGGACCGTG
CGACTGAGACTGACTGGGT
CTAGCTAGACTACGTTTTA
TATATATATACGTCGTCGT
ACTGATGACTAGATTACAG
ACTGATTTAGATACCTGAC
TGATTTTAAAAAATATT...

High-throughput sequencing revolution



High-throughput sequencing: Microscope in the big data era



Genomic variations, 3-D structures, transcription, translation, protein interaction, etc.

High-throughput sequencing assays

Courtesy: Lior Pachter

- **dsRNA-Seq:** Qi Zheng et al., "[Genome-Wide Double-Stranded RNA Sequencing Reveals the Functional Significance of Base-Paired RNAs in Arabidopsis](#)," *PLoS Genet* 6, no. 9 (September 30, 2010): e1001141, doi:10.1371/journal.pgen.1001141.
- **FRAG-Seq:** Jason G. Underwood et al., "[FragSeq: Transcriptome-wide RNA Structure Probing Using High-throughput Sequencing](#)," *Nature Methods* 7, no. 12 (December 2010): 995–1001, doi:10.1038/nmeth.1529.
- **SHAPE-Seq:** (a) Julius B. Lucks et al., "[Multiplexed RNA Structure Characterization with Selective 2'-hydroxyl Acylation Analyzed by Primer Extension Sequencing \(SHAPE-Seq\)](#)," *Proceedings of the National Academy of Sciences* 108, no. 27 (July 5, 2011): 11063–11068, doi:10.1073/pnas.1106501108.
(b) Sharon Aviran et al., "[Modeling and Automation of Sequencing-based Characterization of RNA Structure](#)," *Proceedings of the National Academy of Sciences* (June 3, 2011), doi:10.1073/pnas.1106541108.
- **PARTE-Seq:** Yue Wan et al., "[Genome-wide Measurement of RNA Folding Energies](#)," *Molecular Cell* 48, no. 2 (October 26, 2012): 169–181, doi:10.1016/j.molcel.2012.08.008.
- **PARS-Seq:** Michael Kertesz et al., "[Genome-wide Measurement of RNA Secondary Structure in Yeast](#)," *Nature* 467, no. 7311 (September 2, 2010): 103–107, doi:10.1038/nature09322.
- **Structure-Seq:** Yiliang Ding et al., "[In Vivo Genome-wide Profiling of RNA Secondary Structure Reveals Novel Regulatory Features](#)," *Nature* advance online publication (November 24, 2013), doi:10.1038/nature12756.
- **DMS-Seq:** Silvi Rouskin et al., "[Genome-wide Probing of RNA Structure Reveals Active Unfolding of mRNA Structures in Vivo](#)," *Nature* advance online publication (December 15, 2013), doi:10.1038/nature12894.
- Viral RNA
- **Cir-Seq:** Ashley Acevedo, Leonid Brodsky, and Raul Andino, "[Mutational and Fitness Landscapes of an RNA Virus Revealed through Population Sequencing](#)," *Nature* 505, no. 7485 (January 30, 2014): 686–690, doi:10.1038/nature12861.
- DNA
- **Dup-Seq:** Schmitt, Michael W., Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, and Lawrence A. Loeb. "[Detection of Ultra-rare Mutations by Next-generation Sequencing](#)." *Proceedings of the National Academy of Sciences* 109, no. 36 (September 4, 2012): 14508–14513. doi:10.1073/pnas.1208715109.
- **IMS-MDA-Seq:** Helena M. B. Seth-Smith et al., "[Generating Whole Bacterial Genome Sequences of Low-abundance Species from Complex Samples with IMS-MDA](#)," *Nature Protocols* 8, no. 12 (December 2013): 2404–2412, doi:10.1038/nprot.2013.147.
- Chromatin structure, accessibility and nucleosome positioning
- **Nucleo-Seq:** Anton Valouev et al., "[Determinants of Nucleosome Organization in Primary Human Cells](#)," *Nature* 474, no. 7352 (June 23, 2011): 516–520, doi:10.1038/nature10002.
- **DNase-Seq:** Gregory E. Crawford et al., "[Genome-wide Mapping of DNase Hypersensitive Sites Using Massively Parallel Signature Sequencing \(MPSS\)](#)," *Genome Research* 16, no. 1 (January 1, 2006): 123–131, doi:10.1101/gr.4074106.
- **DNaseI-Seq:** Jay R. Hesselberth et al., "[Global Mapping of protein-DNA Interactions in Vivo by Digital Genomic Footprinting](#)," *Nature Methods* 6, no. 4 (April 2009): 283–289, doi:10.1038/nmeth.1313.
- **Sono-Seq:** Raymond K. Auerbach et al., "[Mapping Accessible Chromatin Regions Using Sono-Seq](#)," *Proceedings of the National Academy of Sciences* 106, no. 35 (September 1, 2009): 14926–14931, doi:10.1073/pnas.0905442106.
- **Hi-C-Seq:** Erez Lieberman-Aiden et al., "[Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome](#)," *Science* 326, no. 5950 (October 9, 2009): 289–293, doi:10.1126/science.1181369.

- **ChIA-PET-Seq:** Melissa J. Fullwood et al., “[An Oestrogen-receptor- \$\alpha\$ -bound Human Chromatin Interactome](#),” *Nature* 462, no. 7269 (November 5, 2009): 58–64, doi:10.1038/nature08497.
- **FAIRE-Seq:** Hironori Waki et al., “[Global Mapping of Cell Type–Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NFI Family in Adipocyte Differentiation](#),” *PLoS Genet* 7, no. 10 (October 20, 2011): e1002311,
- **NOME-Seq:** Theresa K. Kelly et al., “[Genome-wide Mapping of Nucleosome Positioning and DNA Methylation Within Individual DNA Molecules](#),” *Genome Research* 22, no. 12 (December 1, 2012): 2497–2506, doi:10.1101/gr.143008.112.
- **ATAC-Seq:** Jason D. Buenrostro et al., “[Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-binding Proteins and Nucleosome Position](#),” *Nature Methods* advance online publication (October 6, 2013), doi:10.1038/nmeth.2688.
- Genome variation
- **RAD-Seq:** Nathan A. Baird et al., “[Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers](#),” *PLoS ONE* 3, no. 10 (October 13, 2008): e3376, doi:10.1371/journal.pone.0003376.
- **Freq-Seq:** Lon M. Chubiz et al., “[FREQ-Seq: A Rapid, Cost-Effective, Sequencing-Based Method to Determine Allele Frequencies Directly from Mixed Populations](#),” *PLoS ONE* 7, no. 10 (October 31, 2012): e47959, doi:10.1371/journal.pone.0047959.
- **CNV-Seq:** Chao Xie and Martti T. Tammi, “[CNV-seq, a New Method to Detect Copy Number Variation Using High-throughput Sequencing](#),” *BMC Bioinformatics* 10, no. 1 (March 6, 2009): 80, doi:10.1186/1471-2105-10-80.
- **Novel-Seq:** Iman Hajirasouliha et al., “[Detection and Characterization of Novel Sequence Insertions Using Paired-end Next-generation Sequencing](#),” *Bioinformatics* 26, no. 10 (May 15, 2010): 1277–1283, doi:10.1093/bioinformatics/btq152.
- **TAm-Seq:** Tim Forshew et al., “[Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA](#),” *Science Translational Medicine* 4, no. 136 (May 30, 2012): 136ra68, doi:10.1126/scitranslmed.3003726.
- DNA replication
- **Repli-Seq:** R. Scott Hansen et al., “[Sequencing Newly Replicated DNA Reveals Widespread Plasticity in Human Replication Timing](#),” *Proceedings of the National Academy of Sciences* 107, no. 1 (January 5, 2010): 139–144, doi:10.1073/pnas.0912402107
- **ARS-Seq:** Ivan Liachko et al., “[High-resolution Mapping, Characterization, and Optimization of Autonomously Replicating Sequences in Yeast](#),” *Genome Research* 23, no. 4 (April 1, 2013): 698–704, doi:10.1101/gr.144659.112.
- **Sort-Seq:** Carolin A. Müller et al., “[The Dynamics of Genome Replication Using Deep Sequencing](#),” *Nucleic Acids Research* (October 1, 2013): gkt878, doi:10.1093/nar/gkt878.
- **Pool-Seq:** Robert Kofler, Andrea J. Betancourt, and Christian Schlötterer, “[Sequencing of Pooled DNA Samples \(Pool-Seq\) Uncovers Complex Dynamics of Transposable Element Insertions in *Drosophila Melanogaster*](#),” *PLoS Genet* 8, no. 1 (January 26, 2012): e1002487, doi:10.1371/journal.pgen.1002487.
- Replication
- **Bubble-Seq:** Larry D. Mesner et al., “[Bubble-seq Analysis of the Human Genome Reveals Distinct Chromatin-mediated Mechanisms for Regulating Early- and Late-firing Origins](#),” *Genome Research* (July 16, 2013), doi:10.1101/gr.155218.113.

- **RNA-Seq:** Ali Mortazavi et al., "[Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq](#)," *Nature Methods* 5, no. 7 (July 2008): 621–628, doi:10.1038/nmeth.1226.
- **GRO-Seq:** Leighton J. Core, Joshua J. Waterfall, and John T. Lis, "[Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters](#)," *Science* 322, no. 5909 (December 19, 2008): 1845–1848, doi:10.1126/science.1162228.
- **Quartz-Seq:** Yohei Sasagawa et al., "[Quartz-Seq: a Highly Reproducible and Sensitive Single-cell RNA-Seq Reveals Non-genetic Gene Expression Heterogeneity](#)," *Genome Biology* 14, no. 4 (April 17, 2013): R31, doi:10.1186/gb-2013-14-4-r31.
- **CAGE-Seq:** Hazuki Takahashi et al., "[5' End-centered Expression Profiling Using Cap-analysis Gene Expression and Next-generation Sequencing](#)," *Nature Protocols* 7, no. 3 (March 2012): 542–561, doi:10.1038/nprot.2012.005.
- **Nascent-Seq:** Joseph Rodriguez, Jerome S. Menet, and Michael Rosbash, "[Nascent-Seq Indicates Widespread Cotranscriptional RNA Editing in Drosophila](#)," *Molecular Cell* 47, no. 1 (July 13, 2012): 27–37, doi:10.1016/j.molcel.2012.05.002.
- **Precapture RNA-Seq:** Tim R. Mercer et al., "[Targeted RNA Sequencing Reveals the Deep Complexity of the Human Transcriptome](#)," *Nature Biotechnology* 30, no. 1 (January 2012): 99–104, doi:10.1038/nbt.2024.
- **Cel-Seq:** Tamar Hashimshony et al., "[CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification](#)," *Cell Reports* 2, no. 3 (September 27, 2012): 666–673, doi:10.1016/j.celrep.2012.08.003.
- **3P-Seq:** Calvin H. Jan et al., "[Formation, Regulation and Evolution of Caenorhabditis Elegans 3'UTRs](#)," *Nature* 469, no. 7328 (January 6, 2011): 97–101, doi:10.1038/nature09616.
- **NET-Seq:** L. Stirling Churchman and Jonathan S. Weissman, "[Nascent Transcript Sequencing Visualizes Transcription at Nucleotide Resolution](#)," *Nature* 469, no. 7330 (January 20, 2011): 368–373, doi:10.1038/nature09652.
- **SS3-Seq:** Oh Kyu Yoon and Rachel B. Brem, "[Noncanonical Transcript Forms in Yeast and Their Regulation During Environmental Stress](#)," *RNA* 16, no. 6 (June 1, 2010): 1256–1267, doi:10.1261/rna.2038810.
- **FRT-Seq:** Lira Mamanova et al., "[FRT-seq: Amplification-free, Strand-specific Transcriptome Sequencing](#)," *Nature Methods* 7, no. 2 (February 2010): 130–132, doi:10.1038/nmeth.1417.
- **3-Seq:** Andrew H. Beck et al., "[3'-End Sequencing for Expression Quantification \(3SEQ\) from Archival Tumor Samples](#)," *PLoS ONE* 5, no. 1 (January 19, 2010): e8768, doi:10.1371/journal.pone.0008768.
- **PRO-Seq:** Hojoong Kwak et al., "[Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing](#)," *Science* 339, no. 6122 (February 22, 2013): 950–953, doi:10.1126/science.1229386.
- **Bru-Seq:** Artur Veloso et al., "[Genome-Wide Transcriptional Effects of the Anti-Cancer Agent Camptothecin](#)," *PLoS ONE* 8, no. 10 (October 23, 2013): e78190, doi:10.1371/journal.pone.0078190.
- **TIF-Seq:** Vicent Pelechano, Wu Wei, and Lars M. Steinmetz, "[Extensive Transcriptional Heterogeneity Revealed by Isoform Profiling](#)," *Nature* 497, no. 7447 (May 2, 2013): 127–131, doi:10.1038/nature12121.
- **3'-Seq:** Steve Lianoglou et al., "[Ubiquitously Transcribed Genes Use Alternative Polyadenylation to Achieve Tissue-specific Expression](#)," *Genes & Development* 27, no. 21 (November 1, 2013): 2380–2396, doi:10.1101/gad.229328.113.
- **TIVA-Seq:** Ditte Lovatt et al., "[Transcriptome in Vivo Analysis \(TIVA\) of Spatially Defined Single Cells in Live Tissue](#)," *Nature Methods* 11, no. 2 (February 2014): 190–196, doi:10.1038/nmeth.2804.
- **Smart-Seq:** Simone Picelli et al., "[Full-length RNA-seq from Single Cells Using Smart-seq2](#)," *Nature Protocols* 9, no. 1 (January 2014): 171–181, doi:10.1038/nprot.2014.006.

- **PAS-Seq:** Peter J. Shepard et al., "[Complex and Dynamic Landscape of RNA Polyadenylation Revealed by PAS-Seq](#)," *RNA* 17, no. 4 (April 1, 2011): 761–772, doi:10.1261/rna.2581711.
- **PAL-Seq:** Alexander O. Subtelny et al., "[Poly\(A\)-tail Profiling Reveals an Embryonic Switch in Translational Control](#)," *Nature* advance online publication (January 29, 2014), doi:10.1038/nature13007.
- Translation
- **Ribo-Seq:** Nicholas T. Ingolia et al., "[Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling](#)," *Science* 324, no. 5924 (April 10, 2009): 218–223, doi:10.1126/science.1168978.
- **Frac-Seq:** Timothy Sterne-Weiler et al., "[Frac-seq Reveals Isoform-specific Recruitment to Polyribosomes](#)," *Genome Research* (June 19, 2013), doi:10.1101/gr.148585.112.
- **GTI-Seq:** Ji Wan and Shu-Bing Qian, "[TISdb: a Database for Alternative Translation Initiation in Mammalian Cells](#)," *Nucleic Acids Research* (November 6, 2013), doi:10.1093/nar/gkt1085.
- TFBS and Enhancer activity
- **SELEX-Seq:** Matthew Slattery et al., "[Cofactor Binding Evokes Latent Differences in DNA Binding Specificity Between Hox Proteins](#)," *Cell* 147, no. 6 (December 9, 2011): 1270–1282, doi:10.1016/j.cell.2011.10.053.
- **CRE-Seq:** Jamie C. Kwasnieski et al., "[Complex Effects of Nucleotide Variants in a Mammalian Cis-regulatory Element](#)," *Proceedings of the National Academy of Sciences* 109, no. 47 (November 20, 2012): 19498–19503, doi:10.1073/pnas.1210678109.
- **STARR-Seq:** Cosmas D. Arnold et al., "[Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq](#)," *Science* 339, no. 6123 (March 1, 2013): 1074–1077, doi:10.1126/science.1232542.
- **SRE-Seq:** Robin P. Smith et al., "[Massively Parallel Decoding of Mammalian Regulatory Sequences Supports a Flexible Organizational Model](#)," *Nature Genetics* 45, no. 9 (September 2013): 1021–1028, doi:10.1038/ng.2713.
- **HITS-KIN-Seq:** Ulf-Peter Guenther et al., "[Hidden Specificity in an Apparently Nonspecific RNA-binding Protein](#)," *Nature* 502, no. 7471 (October 17, 2013): 385–388, doi:10.1038/nature12543.
- **Capture-C-Seq:** Jim R. Hughes et al., "[Analysis of Hundreds of Cis-regulatory Landscapes at High Resolution in a Single, High-throughput Experiment](#)," *Nature Genetics* 46, no. 2 (February 2014): 205–212, doi:10.1038/ng.2871.
- RNA-RNA interaction
- **CLASH-Seq:** Aleksandra Helwak et al., "[Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding](#)," *Cell* 153, no. 3 (April 2013): 654–665, doi:10.1016/j.cell.2013.03.043.
- RNA-DNA binding
- **ChIRP-Seq:** Ci Chu et al., "[Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions](#)," *Molecular Cell* 44, no. 4 (November 18, 2011): 667–678, doi:10.1016/j.molcel.2011.08.027.
- **CHART-Seq:** Matthew D. Simon et al., "[The Genomic Binding Sites of a Noncoding RNA](#)," *Proceedings of the National Academy of Sciences* 108, no. 51 (December 20, 2011): 20497–20502, doi:10.1073/pnas.1113536108.
- **RAP-Seq:** Jesse M. Engreitz et al., "[The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome](#)," *Science* 341, no. 6147 (August 16, 2013): 1237973, doi:10.1126/science.1237973.

- **RIP-Seq:** Ci Chu et al., "[Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions](#)," *Molecular Cell* 44, no. 4 (November 18, 2011): 667–678, doi:10.1016/j.molcel.2011.08.027.
- **PAR-Clip-Seq:** Markus Hafner et al., "[Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP](#)," *Cell* 141, no. 1 (April 2, 2010): 129–141, doi:10.1016/j.cell.2010.03.009.
- **iCLIP-Seq:** Julian König et al., "[iCLIP Reveals the Function of hnRNP Particles in Splicing at Individual Nucleotide Resolution](#)," *Nature Structural & Molecular Biology* 17, no. 7 (July 2010): 909–915, doi:10.1038/nsmb.1838.
- **PTB-Seq:** Yuanchao Xue et al., "[Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping](#)," *Molecular Cell* 36, no. 6 (December 24, 2009): 996–1006, doi:10.1016/j.molcel.2009.12.003.
- Protein-DNA binding
- **ChIP-Seq:** David S. Johnson et al., "[Genome-Wide Mapping of in Vivo Protein-DNA Interactions](#)," *Science* 316, no. 5830 (June 8, 2007): 1497–1502, doi:10.1126/science.1141319.
- **ChIP-Seq:** Tarjei S. Mikkelsen et al., "[Genome-wide Maps of Chromatin State in Pluripotent and Lineage-committed Cells](#)," *Nature* 448, no. 7153 (August 2, 2007): 553–560, doi:10.1038/nature06008.
- **HiTS-Flip-Seq:** Razvan Nutiu et al., "[Direct Measurement of DNA Affinity Landscapes on a High-throughput Sequencing Instrument](#)," *Nature Biotechnology* 29, no. 7 (July 2011): 659–664, doi:10.1038/nbt.1882.
- **Chip-exo-Seq:** Ho Sung Rhee and B. Franklin Pugh, "[Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution](#)," *Cell* 147, no. 6 (December 9, 2011): 1408–1419, doi:10.1016/j.cell.2011.11.013.
- **PB-Seq:** Michael J. Guertin et al., "[Accurate Prediction of Inducible Transcription Factor Binding Intensities In Vivo](#)," *PLoS Genet* 8, no. 3 (March 29, 2012): e1002610, doi:10.1371/journal.pgen.1002610.
- **AHT-ChIP-Seq:** Sarah Aldridge et al., "[AHT-ChIP-seq: a Completely Automated Robotic Protocol for High-throughput Chromatin Immunoprecipitation](#)," *Genome Biology* 14, no. 11 (November 7, 2013): R124, doi:10.1186/gb-2013-14-11-r124.
- Protein-protein interaction
- **PDZ-Seq:** Andreas Ernst et al., "[Coevolution of PDZ Domain–ligand Interactions Analyzed by High-throughput Phage Display and Deep Sequencing](#)," *Molecular BioSystems* 6, no. 10 (2010): 1782, doi:10.1039/c0mb00061b.
- Small molecule-protein interaction
- **PD-Seq:** Daniel Arango et al., "[Molecular Basis for the Action of a Dietary Flavonoid Revealed by the Comprehensive Identification of Apigenin Human Targets](#)," *Proceedings of the National Academy of Sciences* 110, no. 24 (June 11, 2013): E2153–E2162, doi:10.1073/pnas.1303726110.
- Small molecule-DNA interaction
- **Chem-Seq:** Lars Anders et al., "[Genome-wide Localization of Small Molecules](#)," *Nature Biotechnology* 32, no. 1 (January 2014): 92–96, doi:10.1038/nbt.2776.
- Methylation
- **CAB-Seq:** Xingyu Lu et al., "Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA," *Journal of the American Chemical Society* 135, no. 25 (June 26, 2013): 9315–9317, doi:10.1021/ja4044856.
- **HELP-Seq:** Mayumi Oda et al., "[High-resolution Genome-wide Cytosine Methylation Profiling with Simultaneous Copy Number Analysis and Optimization for Limited Cell Numbers](#)," *Nucleic Acids Research* 37, no. 12 (July 1, 2009): 3829–3839, doi:10.1093/nar/gkp260.
- **TAB-Seq:** Miao Yu et al., "Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome," *Cell* 149, no. 6 (June 8, 2012): 1368–1380, doi:10.1016/j.cell.2012.04.027.

- **TAmC-Seq:** Liang Zhang et al., “Tet-mediated Covalent Labelling of 5-methylcytosine for Its Genome-wide Detection and Sequencing,” *Nature Communications* 4 (February 26, 2013): 1517, doi:10.1038/ncomms2527.
- **fcAB-Seq:** Chun-Xiao Song et al., “Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming,” *Cell* 153, no. 3 (April 25, 2013): 678–691, doi:10.1016/j.cell.2013.04.001.
- **MeDIP-Seq:** Thomas A. Down et al., “[A Bayesian Deconvolution Strategy for Immunoprecipitation-based DNA Methylome Analysis](#),” *Nature Biotechnology* 26, no. 7 (July 2008): 779–785, doi:10.1038/nbt1414.
- **Methyl-Seq:** Alayne L. Brunner et al., “[Distinct DNA Methylation Patterns Characterize Differentiated Human Embryonic Stem Cells and Developing Human Fetal Liver](#),” *Genome Research* 19, no. 6 (June 1, 2009): 1044–1056, doi:10.1101/gr.088773.108.
- **oxBS-Seq:** Michael J. Booth et al., “[Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution](#),” *Science* 336, no. 6083 (May 18, 2012): 934–937, doi:10.1126/science.1220671.
- **RBBS-Seq:** Zachary D. Smith et al., “[High-throughput Bisulfite Sequencing in Mammalian Genomes](#),” *Methods* 48, no. 3 (July 2009): 226–232, doi:10.1016/j.jymeth.2009.05.003.
- **BS-Seq:** Ryan Lister et al., “[Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences](#),” *Nature* 462, no. 7271 (November 19, 2009): 315–322, doi:10.1038/nature08514.
- **BisChIP-Seq:** Aaron L. Statham et al., “[Bisulfite Sequencing of Chromatin Immunoprecipitated DNA \(BisChIP-seq\) Directly Informs Methylation Status of Histone-modified DNA](#),” *Genome Research* 22, no. 6 (June 1, 2012): 1120–1127, doi:10.1101/gr.132076.111.
- Phenotyping
- **Bar-Seq:** Andrew M. Smith et al., “[Quantitative Phenotyping via Deep Barcode Sequencing](#),” *Genome Research* (July 21, 2009), doi:10.1101/gr.093955.109.
- **TraDI-Seq:** Gemma C. Langridge et al., “[Simultaneous Assay of Every Salmonella Typhi Gene Using One Million Transposon Mutants](#),” *Genome Research* (October 13, 2009), doi:10.1101/gr.097097.109.
- **Tn-Seq:** Tim van Opijnen, Kip L. Bodi, and Andrew Camilli, “[Tn-seq: High-throughput Parallel Sequencing for Fitness and Genetic Interaction Studies in Microorganisms](#),” *Nature Methods* 6, no. 10 (October 2009): 767–772, doi:10.1038/nmeth.1377.
- **IN-Seq:** Andrew L. Goodman et al., “[Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat](#),” *Cell Host & Microbe* 6, no. 3 (September 17, 2009): 279–289, doi:10.1016/j.chom.2009.08.003.
- **Immuno-Seq:** Harlan S. Robins et al., “[Comprehensive Assessment of T-cell Receptor B-chain Diversity in Aβ T Cells](#),” *Blood* 114, no. 19 (November 5, 2009): 4099–4107, doi:10.1182/blood-2009-04-217604.
- **mutARS-Seq:** Ivan Liachko et al., “[High-resolution Mapping, Characterization, and Optimization of Autonomously Replicating Sequences in Yeast](#),” *Genome Research* 23, no. 4 (April 1, 2013): 698–704, doi:10.1101/gr.144659.112.
- **Ig-Seq:** Vollmers, Christopher, Rene V. Sit, Joshua A. Weinstein, Cornelia L. Dekker, and Stephen R. Quake. “[Genetic Measurement of Memory B-cell Recall Using Antibody Repertoire Sequencing](#)” *Proceedings of the National Academy of Sciences* 110, no. 33 (August 13, 2013): 13463–13468. doi:10.1073/pnas.1312146110.
- **Ig-seq:** Busse, Christian E., Irina Czogiel, Peter Braun, Peter F. Arndt, and Hedda Wardemann. “[Single-cell Based High-throughput Sequencing of Full-length Immunoglobulin Heavy and Light Chain Genes](#).” *European Journal of Immunology* (2013): n/a–n/a. doi:10.1002/eji.201343917.
- **Ren-Seq:** Florian Jupe et al., “[Resistance Gene Enrichment Sequencing \(RenSeq\) Enables Reannotation of the NB-LRR Gene Family from Sequenced Plant Genomes and Rapid Mapping of Resistance Loci in Segregating Populations](#),” *The Plant Journal* 76, no. 3 (2013): 530–544, doi:10.1111/tpj.12307.
- **Mu-Seq:** Donald R. McCarty et al., “[Mu-seq: Sequence-Based Mapping and Identification of Transposon Induced Mutations](#),” *PLoS ONE* 8, no. 10 (October 23, 2013): e77172, doi:10.1371/journal.pone.0077172.
- **Stable-Seq:** Ikin Kim et al., “[High-throughput Analysis of in Vivo Protein Stability](#),” *Molecular & Cellular Proteomics: MCP* 12, no. 11 (November 2013): 3370–3378. doi:10.1074/mcp.O113.031708.


Applications: basic science

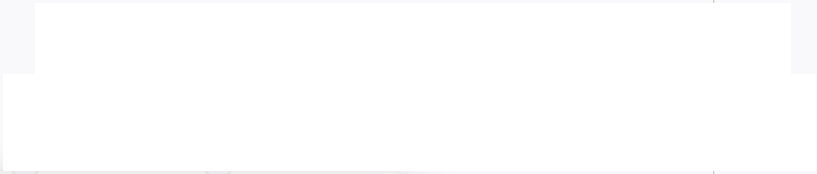
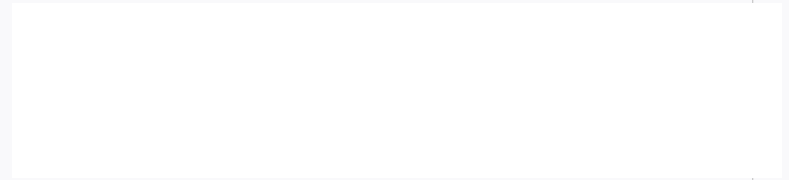
- Structure: sequence, genes, 3-D shape
- Function: DNA-RNA-protein interactions
- Evolution: history and population dynamics

Applications: medicine

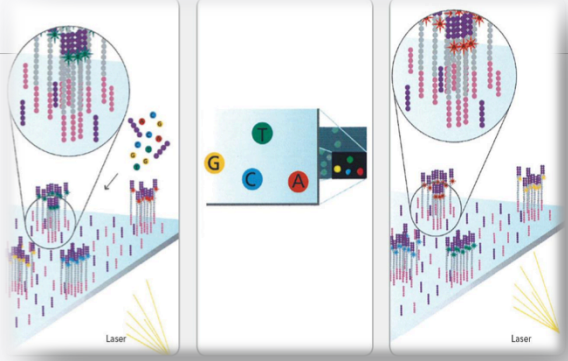
- Genotype → phenotype
- Cancer
- Drug prediction
- Pathogen detection, pre-natal testing

The Data: shotgun sequencing

 The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.



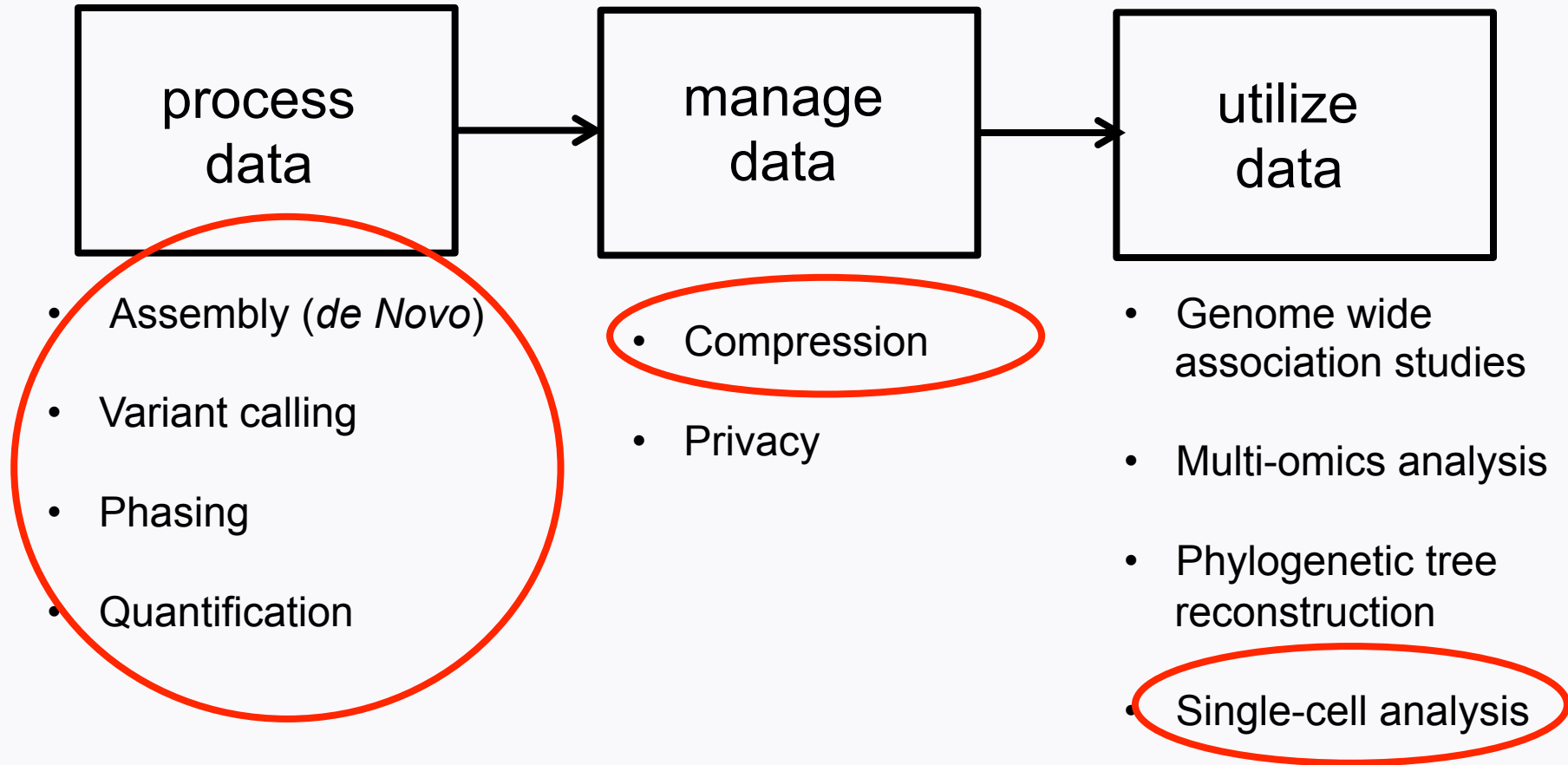
read



Technologies

Sequencer	Sanger 3730xl	454 GS	Ion Torrent	SOLiDv4	Illumina HiSeq 2000	Pac Bio
Mechanism	Dideoxy chain termination	Pyrosequencing	Detection of hydrogen ion	Ligation and two-base coding	Reversible Nucleotides	Single molecule real time
Read length	400-900 bp	700 bp	~400 bp	50 + 50 bp	100 bp PE	>10000 bp
Error Rate	0.001%	0.1%	2%	0.1%	2%	10-15%
Output data (per run)	100 KB	1 GB	100 GB	100 GB	1 TB	10 GB

Data science of high-throughput sequencing



Tools

- Combinatorial algorithms
- Information theory
- Signal processing
- Machine learning

Audience

- Theoretically inclined students interested in biological applications and exposure to real data.
- Computational biology students interested in tools to develop new methods.