illumina®

The Next Generation of Gene Expression Profiling with RNA-Seq

Advancing RNA sequencing and DNA methylation studies with the power of next-generation sequencing.

Table of Contents

Next-Generation Sequencing Accelerates Transcriptomics Research

The Dynamic Landscape of Transcriptomics

Our curiosity and quest for answers has always been the driving force for discovery. As our tools have evolved from the basic light microscope to high-throughput DNA sequencers, so has our understanding of the world around us. Next-generation sequencing (NGS) and NGS-based RNA sequencing (RNA-Seq) are the latest technological advances driving scientists to push beyond the limits of current research methods. As researchers seek to understand how the transcriptome shapes biology, RNA-Seq is becoming one of the most significant and powerful tools in modern science.

A highly sensitive and accurate method for gene expression analysis across the transcriptome, RNA-Seq is providing visibility into previously undetectable changes in gene expression, as well as enabling the characterization of multiple forms of noncoding RNA.[1,2] With RNA-Seq, researchers can detect the fine architecture of the transcriptome, such as transcript isoforms, gene fusions, single nucleotide variants, and other features—without the limitation of prior knowledge.[2,3]

RNA-Seq Drives High-Impact Publications

With these advantages, RNA-Seq is accelerating the pace of research and driving high-impact publications across a broad range of fields, including cancer research, complex disease, and virology (Figure 1). From a recent study in Liberia, where RNA-Seq was used to establish direct evidence of sexual transmission of Ebola,[4] to an immunology study at Harvard University, where single-cell RNA-Seq revealed, for the first time that different immune cells express different isoforms of the same gene,[5] RNA-Seq is breaking new ground. By providing a view into the realm of noncoding transcripts, RNA-Seq is providing a glimpse of their role in complex disease.[6] RNA-Seq has also increased our understanding of agriculturally significant diseases and improved our ability to protect crops and enhance breeding programs.[7,8]

"A near complete snapshot of the *Zea mays* seeding transcriptome revealed from **ultra-deep sequencing.**"

- Sci Rep, 2014

"Long non-coding RNAs: **new players** in cell differentiation & development."

- Nature Reviews Genetics, 2013

"Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprograming."

- Cell Stem Cell, 2015

"**Biased classification of sensory neuron types by large-scale single-cell RNA sequencing.**"

- Nature Neuroscience, 2015

"**Exon-intron circular RNAs regulate transcription in the nucleus.**"

- Nature Structural & Molecular Biology, 2015

Grant Funding and Publication Trends for RNA-Seq Have Never Looked Better

As the impact of RNA-Seq continues to grow, the number of researchers leveraging RNA-Seq is on the rise, as are the number of RNA-Seq–based citations (Figure 1). Data from the National Institutes of Health (NIH) show that publications for RNA-Seq based research have steadily increased from 6 publications in 2008 to nearly 2300 publications in 2015.[9] Furthermore, data from the NIH show that funding for RNA-Seq–based research has steadily increased each year from 2009 to 2015, while funding for array-based research has decreased each year over the same time period (Figure 2).

Since NGS was introduced, substantial advances in sample preparation, chemistry, and instrument output have markedly improved data quality, speed, and affordability (Figure 3). With these trends likely to continue and sequencing improvements still to come, NGS-based RNA-Seq will continue to empower an increasing range of study designs in the near future. The sharp increase in RNA-Seq publications, citations, and grant funding suggests that RNA-Seq may be emerging as the method of choice for transcriptome analysis. Now is the time to leverage the power of next-generation RNA sequencing.

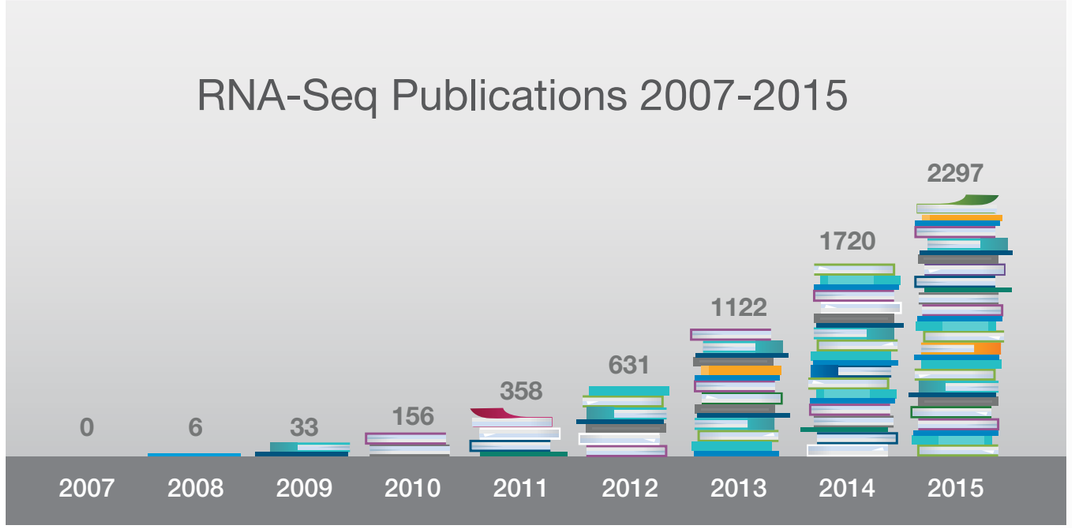


Figure 1: RNA-Seq Publications on the Rise— The number of publications including RNA-Seq data have increased from 6 publications in 2008 to nearly 2300 publications in 2015.[9] Each bar represents the total number of citations returned for each year.

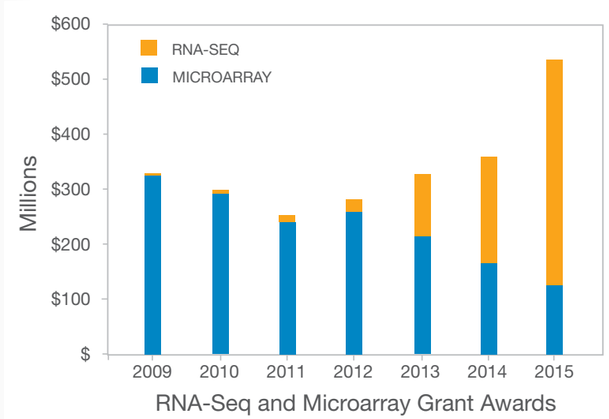


Figure 2: RNA-Seq Grants Outpace Array Grants— NIH grants awarded based on transcriptome research proposals including RNA-Seq have been increasing since 2009 compared to expression arrays.[10]

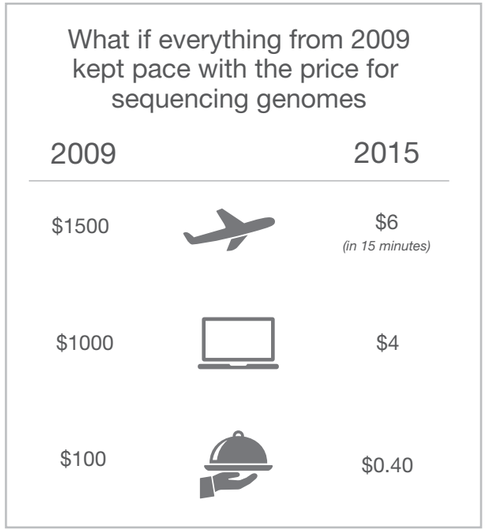


Figure 3: The Dramatic Decrease in the Cost of NGS

The Advantages of RNA Sequencing vs Gene Expression Microarrays

Hypothesis-Free Study Design and Higher Discovery Power

RNA-Seq is a powerful sequencing-based method that enables researchers to break through the inefficiency and expense of legacy technologies such as real-time PCR (RT-PCR) and microarrays. Whether adding RNA-Seq to the current repertoire of research methods or making a complete switch from one method to another, RNA-Seq provides many significant advantages. Because the method does not require predesigned probes, the data sets are unbiased, allowing for hypothesis-free experimental design.[2,3] This type of NGS analysis is a powerful tool for transcript and variant discovery studies that are not possible using traditional microarray-based methods.

Wider Dynamic Range and Higher Sensitivity

Unlike microarrays, which measure continuous probe intensities, RNA-Seq quantifies individual sequence reads aligned to a reference sequence, which produces discreet (digital) read counts.[2] Furthermore, by increasing or decreasing the number of sequencing reads (coverage level or coverage depth), researchers can fine-tune the sensitivity of an experiment to accommodate different study objectives. The digital nature of this process and the ability to control coverage levels supports an extremely broad dynamic range, with absolute rather than relative expression values.[1-3] Assuming 10-50 million mapped reads, the dynamic range of RNA-Seq spans 5 orders of magnitude (>10^5) and is typically several orders of magnitude higher than most array technologies (10^3).[2,11] As a result, RNA-Seq has been shown to detect a higher percentage of differentially expressed genes compared to expression arrays, especially genes with low abundance.[11,12]

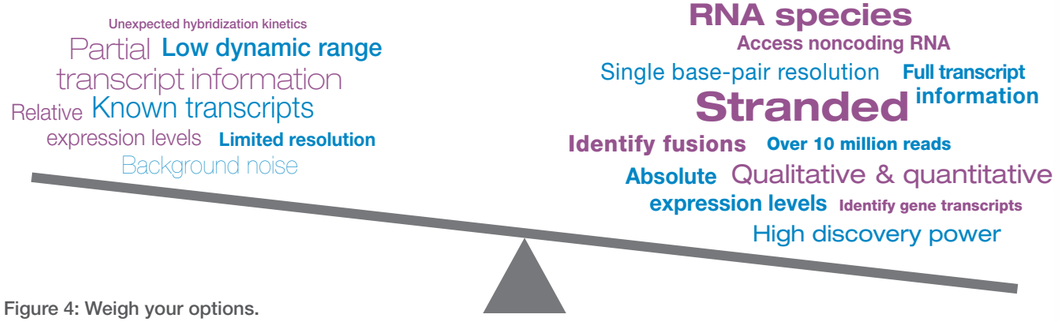


Figure 4: Weigh your options.

Ability to Detect Alternative Splice Sites, Novel Isoforms, and Noncoding RNA

Beyond gene expression profiling, RNA-Seq can identify alternatively spliced isoforms, splice sites, and allele-specific expression—all in a single experiment.[1,2] Furthermore, because RNA-Seq can sequence extremely short fragments and because library prep methods can include or exclude mRNA isolation, it can detect and sequence small RNAs and multiple forms of noncoding RNA, such as small interfering RNA (siRNA), microRNA (miRNA), small nucleolar (snoRNA), and transfer RNA (tRNA).[1,2] The ability to sequence small fragments also enables high-quality data generation with degraded RNA samples, such as formalin-fixed, paraffin-embedded (FFPE) samples.[13] The sequencing data can also be reanalyzed as novel features of the transcriptome are discovered over time. With arrays, the samples would have to be rerun through the entire microarray workflow, from probe design to laboratory work, assuming the original samples were still available.[14] In sum, RNA-Seq offers many advantages over microarrays (Table 1). It provides a unique combination of transcriptome-wide coverage, broad dynamic range, and high sensitivity that can empower researchers to investigate and understand the molecular mechanisms of normal development and disease.

Summary: Advantages of RNA-Seq Over Gene Expression Microarrays

- Provides sensitive, accurate measurement of gene expression at the transcript level
- Generates both qualitative and quantitative data
- Captures splice junctions, fusions, coding, and multiple forms of noncoding RNA such as siRNA, miRNA, snoRNA, and tRNA
- Covers an extremely broad dynamic range
- Delivers superior performance with degraded RNA such as FFPE tissue samples
- Maintains and tracks strand-specific information in the data
- Delivers a more powerful method for discovery applications
- Scales for large studies and high sample numbers

Table 1: Comparison of RNA-Seq technology with expression microarrays

Application	RNA-Seq	Microarray
High run-to-run reproducibility	Yes	Yes
Dynamic range comparable to actual transcript abundances within cells	Yes	No
Able to detect alternative splice sites and novel isoforms	Yes	No
De novo analysis of samples without a reference genome	Yes	No
Re-analyzable data	Yes	No

Gene Expression Studies with RNA-Seq

What is Gene Expression and Why is it Important?

To understand how a seed transforms into a flower or how a starfish regenerates a lost limb, the study of gene expression is required. Gene expression is the process by which the genetic blueprint is translated into functional, biologically active units. Usually, genes encode for proteins, but, as scientists are coming to learn, they encode for a vast array of additional nonprotein-coding elements, such as siRNA, tRNA, and multiple forms of microRNAs.[1] This constellation of RNAs drives essential biological processes, such as transcriptional regulation, cellular differentiation, and, when dysregulated, the development of complex disease.[15] RNA-Seq delivers a high-resolution, base-by-base view of coding and noncoding RNA activity, providing a comprehensive picture of gene expression across the full transcriptome at a specific moment in time (Figure 6). While RT-PCR and expression microarrays can quantify expression levels, a major strength of RNA-Seq lies in its ability to identify novel features of the transcriptome. With RT-PCR and microarrays, only regions of the genome that are known and previously sequenced can be interrogated. With RNA-Seq, the whole transcriptome—including both known and unknown regions—are captured.

Figure 5: See the Big Picture with RNA-Seq.

Gene Expression Applications with RNA-Seq

Differential Expression in Cell Biology and Complex Disease Research

To understand normal cell development and disease mechanisms, researchers frequently investigate differential expression during development, in specific tissues, or in response to varying conditions. RNA-Seq is currently being used to assess gene expression profiles for the study of a number of complex diseases, including cardiomyopathy, multiple sclerosis, and metabolic disorders.[16-18] In a recent study of differential gene expression in individuals with the metabolic disorder nonalcoholic fatty liver disease (NAFLD), 75 miRNAs were identified with 30 upregulated and 45 downregulated miRNAs. Through functional analysis, the researchers identified miRNA gene targets in significant pathways, such as growth factor beta signaling and apoptosis signaling.[18]

Single-cell RNA-Seq has become a powerful tool in stem cell research, immunology, and neurobiology studies to evaluate cell development and differentiation.[19-21] Single-cell RNA-Seq was used in a research study to identify and construct a cellular taxonomy of the visual cortex in adult mice. Based on expression signatures, distinct cell types were identified.[49] Studies designed to test the electrophysiological and axon projection properties of the identified cell types are ongoing.[21]

See how scientists today are using RNA-Seq for gene expression profiling in complex disease research.

Translational profiling identifies a cascade of damage initiated in motor neurons and spreading to glia in mutant *SOD1*-mediated ALS. *Proc Natl Acad Sci.* 2015;112:E6993-7002.

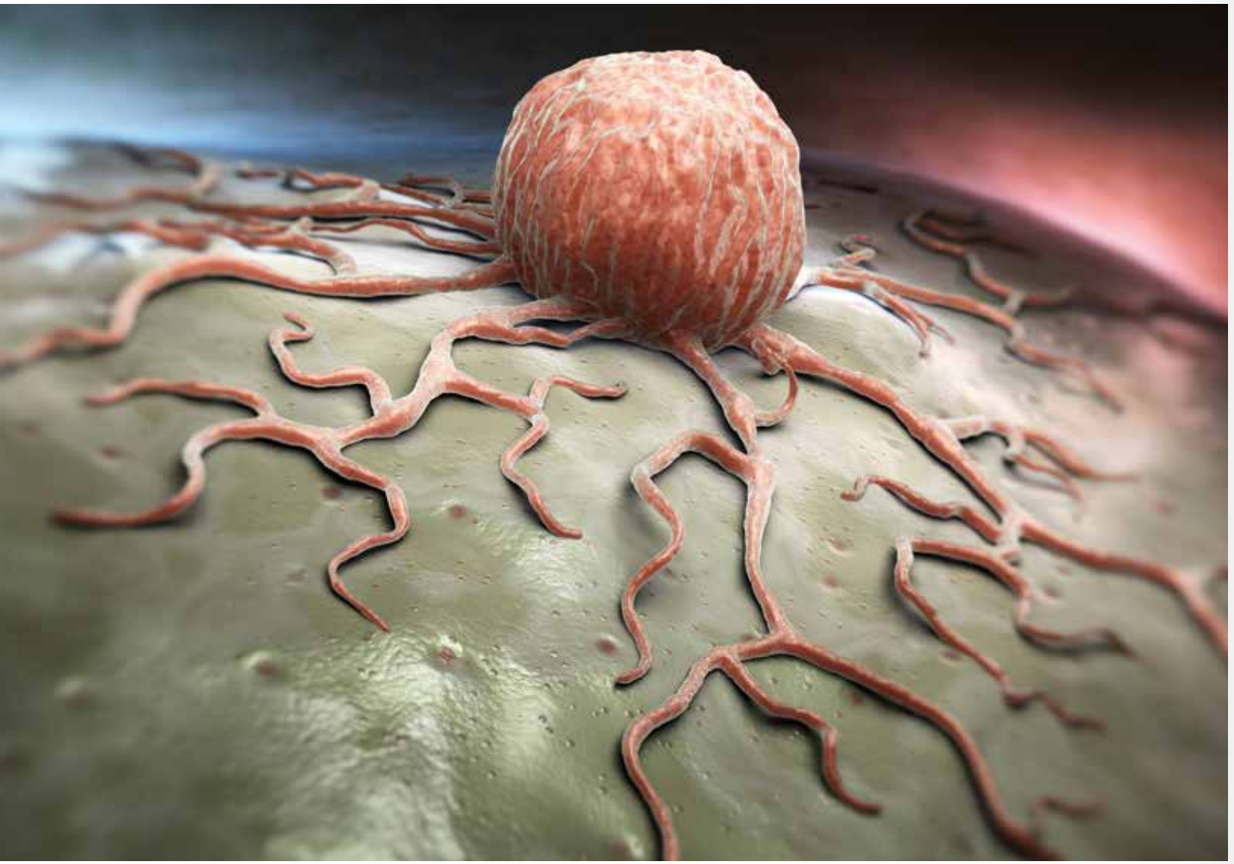
Gene expression profiling of amyotrophic lateral sclerosis (ALS)-causing mutations was performed on samples from the ALS mouse line *SOD1*. Researchers found that pathogenesis involves a temporal cascade of cell type-selective damage initiating in motor neurons. RNA-Seq libraries were prepared using the TruSeq® RNA Library Prep Kit, and sequenced on the HiSeq® System.

Pancreatic ß cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. *Science.* 2015;350:aac4250.

To study the impact of the cell-autonomous clock on pancreatic ß-cells, researchers examined islets from mice with intact or disrupted *BMAL1* expression. Researchers found that oscillation of insulin secretion was synchronized with the expression of genes encoding secretory machinery and insulin regulation. RNA-Seq libraries were prepared with the TruSeq Stranded mRNA Library Prep Kit and sequenced on the NextSeq® 500 and HiSeq Systems.

Depletion of fat-resident Treg cells prevents age-associated insulin resistance. *Nature.* 2015;528:137-41.

Researchers performed comparative gene expression profiling to show that fat-resident regulatory T cells accumulate in adipose tissue as a function of age rather than obesity. RNA-Seq libraries were prepared using the TruSeq RNA Library Prep Kit, and sequenced on the HiSeq System.



Noncoding RNA and Fusion Detection in Cancer Research

A major advantage of RNA-Seq is that it provides a rich view of transcriptome activity well beyond basic abundance measurements. With RNA-Seq, researchers can detect alternative splice sites, characterize noncoding RNA activity, and identify novel gene fusions. Gene fusion detection is particularly significant for cancer research, as 20% of all human tumors have been shown to carry translocations and gene fusions.[22] Furthermore, it is well established that the majority of gene fusions have a significant impact on tumorigenesis and that they have a strong association with morphological phenotype, making them useful as potential diagnostic and prognostic markers.[22] Until recently, the vast majority of gene fusions were detected through cytogenetic characterization, followed by RT–PCR and Sanger sequencing. With the advent of NGS technology and its ability to identify gene fusions, the proportion of gene fusions reported every year is likely to increase rapidly.[22,23]

See how scientists are leveraging RNA-Seq technology in cancer research.

FN1-EGF gene fusions are recurrent in calcifying aponeurotic fibroma. _J Pathol_. 2016;238:502-7.

Researchers identified overexpression of the *EGF* gene due to fusion to a high-expressing *FN1* promoter as a common link in several cases of calcifying aponeurotic fibroma (CAF). This *FN1–EGF* fusion appears to be the main driver mutation in CAF. RNA-Seq libraries were prepared using the TruSeq RNA Access Library Prep Kit and sequenced on the NextSeq 500 System.

Identification of a novel _PARP14-TFE3_ gene fusion from 10-year-old FFPE tissue by RNA-Seq. _Genes Chromosomes Cancer_. 2015 May 29. Epub ahead of print.

RNA-Seq was used to identify a novel *PARP14-TFE3* gene fusion in a 10-year-old FFPE tissue sample that was positive for translocation renal cell carcinoma (RCC). This expands the number of *TFE3* gene fusions known to play a role in RCC. RNA-Seq libraries were prepared using the TruSeq RNA Access Library Prep Kit or the TruSeq RNA Library Preparation Kit and sequenced on the MiSeq® System.

Gene fusion detection in formalin-fixed paraffin-embedded benign fibrous histiocytomas using fluorescence _in situ_ hybridization and RNA sequencing. _Lab Invest_. 2015;95:1071-6.

RNA-Seq was used to evaluate the frequency of *PRKC* gene fusions in 36 FFPE tissue samples with benign fibrous histiocytomas (FH). Researchers conclude that while *PRKC* gene fusions exist in several FH morphologies, the fusions only appear in a minority of cases. RNA-Seq libraries were prepared using the TruSeq RNA Access Library Prep Kit and sequenced on the NextSeq 500 System.

Detection of Emerging Pathogens with RNA-Seq

RNA viruses are ubiquitous in the environment and cause many human diseases such as Influenza, severe acute respiratory syndrome (SARS), Zika, and Ebola. Although many viral diseases are preventable through effective, ongoing vaccine programs, many have no known vaccine and continue to pose significant health risks.[24] Researchers using RNA-Seq had a strong impact on studies investigating the recent Zika virus outbreak in Brazil and the Ebola virus pandemic in West Africa. In 2015, the incidence of microcephaly in Brazil increased by more than 20 times compared to previous years.[25] In 2016, researchers from Rio de Janeiro, Brazil isolated viral RNA from amniotic fluid samples and used RNA-Seq data to show that the Zika virus can cross the placental barrier.[25] This established the strongest direct evidence to date of the association between maternal Zika virus infection and microcephaly in newborns.

In December of 2013, Ebola emerged in Guinea and spread rapidly throughout the West African continent, becoming the largest Ebola pandemic in history.[26] As the incidence of new Ebola cases declined, support efforts turned toward caring for Ebola survivors and their families. Guidelines for safe contact with Ebola survivors changed significantly when a 2015 study revealed that Ebola virus could be sexually transmitted.[4] Direct RNA sequencing of samples from an Ebola virus survivor and the survivor's wife demonstrated that the virus could be transmitted through semen, and that viral particles were present in semen for almost 200 days after the virus had cleared the survivor's bloodstream.[4] This study underscored important new considerations for communities living with Ebola survivors.

See how researchers are using RNA-Seq in the fields of epidemiology and infectious disease.

Determination and therapeutic exploitation of Ebola virus spontaneous mutation frequency. *J Virol*. 2015;90:2345-55.

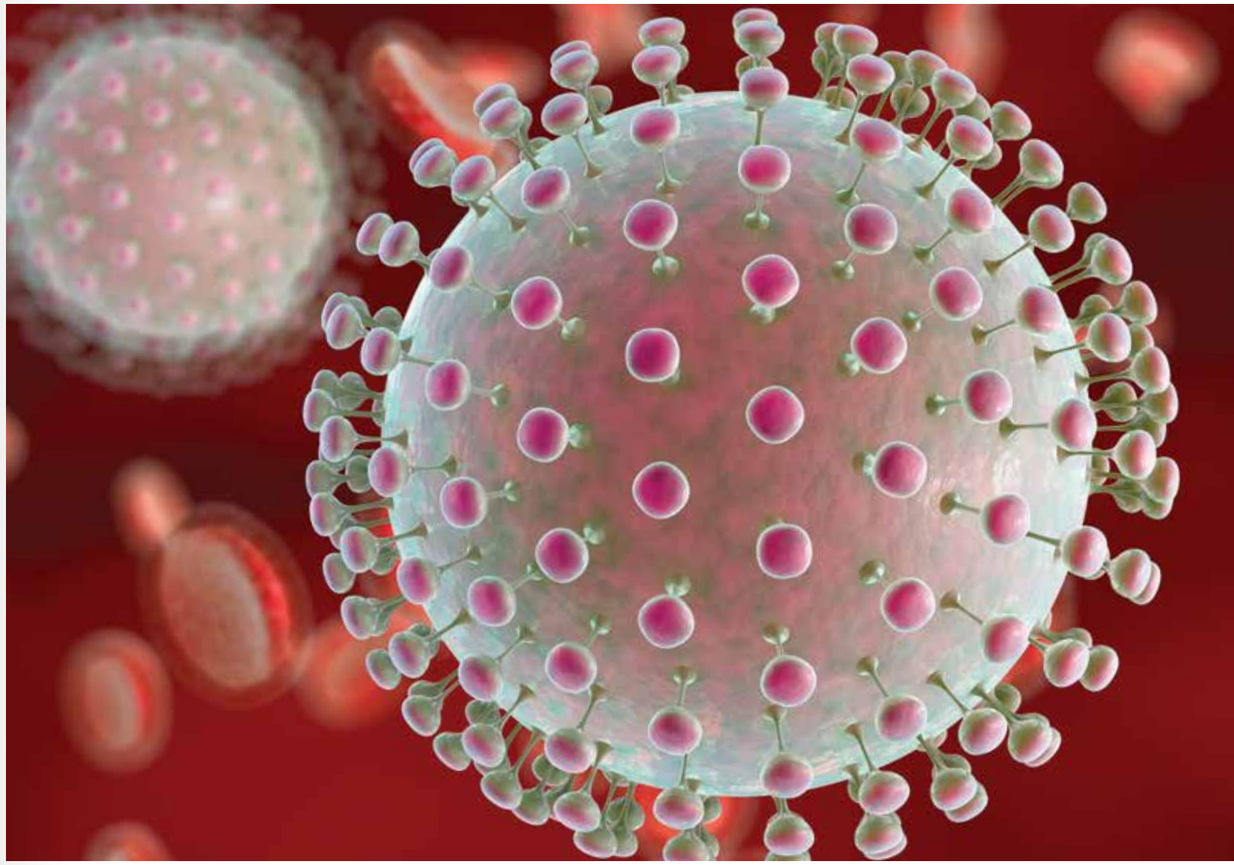
Ultradeep sequencing of a recombinant Ebola virus revealed that while the mutation frequency of the virus is similar to that of other RNA viruses, the genetic diversity was not. This implies that the virus does not tolerate genetic changes well. Taking advantage of this observation, researchers treated mice exposed to the Ebola virus with ribavirin, increasing the rate of viral mutation by 50%. The result was reduced production of infectious virus and up to 75% survival rate of the mice. RNA-Seq libraries were prepared using the TruSeq RNA Library Preparation Kit and sequenced on the MiSeq System.

Molecular evidence of sexual transmission of Ebola virus. *N Engl J Med*. 2015;373(25):2448-54.

Ebola virus genomes were assembled from patient blood samples and a semen sample from a survivor. RNA-Seq demonstrated that the Ebola genomes shared 3 substitutions that were absent from all other Western African Ebola sequences and distinct from the last documented transmission chain in Liberia before this case. This evidence is consistent with sexual transmission of the viral. RNA-Seq libraries were prepared using the TruSeq RNA Access Library Prep Kit or the TruSeq RNA Library Preparation Kit and sequenced on the MiSeq System.

Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis*. 2016;16(6):653-60.

Amniotic fluid samples from 2 pregnant women from Brazil, whose fetuses had been diagnosed with microcephaly, were analyzed using viral metagenomics NGS. The Zika virus genome was identified in the amniotic fluid of both pregnant women, suggesting that the virus can cross the placental barrier. RNA libraries were prepared using the TruSeq Stranded Total RNA LT Library Prep Kit and sequenced on the MiSeq System.



Biomarker and Drug Pathway Discovery with RNA-Seq

Gene expression studies are laying the groundwork for advances in precision medicine by identifying potentially therapeutic biomarkers.[27] Although both expression microarrays and RNA-Seq have been used to develop gene expression-based predictive models, RNA-Seq shows exceptional performance in profiling genes with low expression levels.[27] In addition, NGS analysis has the ability to interrogate a higher number of genetic regions, at single-base resolution. For example, capillary electrophoresis can sequence a single DNA target region per sequencing reaction, and microarrays can interrogate up to 1 million probes per array (Affymetrix GeneChip). In comparison, NGS technology can sequence the whole transcriptome (eg, genes, gene variants, and noncoding transcripts) without the limitation of probe design. For purposes of discovery-related applications or development of future screening methods, the ability to interrogate the whole transcriptome offers a significant advantage.[28]

In the realm of clinical and translational research, RNA-Seq is being used to identify clinically relevant mechanisms of disease, discover predictive biomarkers, and identify responsive subpopulations. In a recent study by the Norvartis Institute, researchers used high-throughput RNA-Seq to screen patient-derived tumor zenografts in an effort to develop predictive models for drug response testing.[29] In this study, over 1000 patient-derived tumor zenograft models were developed from a broad spectrum of common solid tumor tissue types, including liver, lung, breast, skin, lymphoma, and more. The identity of each model was established using single nucleotide polymorphism (SNP) analysis, histology, and RNA-Seq–based expression profiling.

See how scientists are using RNA-Seq for biomarker and drug pathway discovery research.

Transcriptomic variation of pharmacogenes in multiple human tissues and lymphoblastoid cell lines. *Pharmacogenomics J*. 2016 Feb 9. Epub ahead of print.

Expression profiling of 389 genes involved in drug disposition and action was performed across 4 tissues from 139 individuals. Researchers found substantial differences in expression levels and splicing and novel splicing events. RNA-Seq libraries were prepared using the TruSeq RNA Library Prep Kit and sequenced on the HiSeq 2000 System.

Serum miRNA panel as potential biomarkers for chronic hepatitis B with persistently normal alanine aminotransferase. *Clin Chim Acta*. 2015;451:232-9.

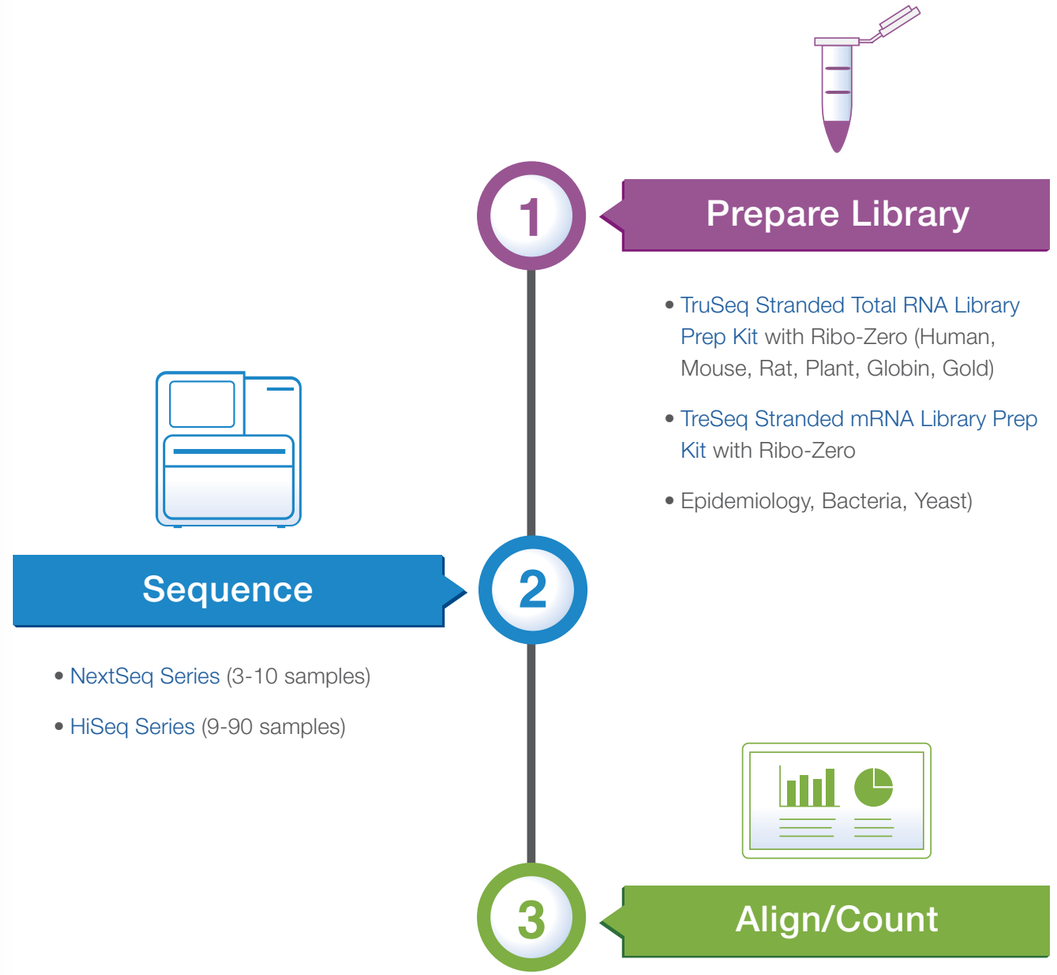
Researchers sequenced circulating miRNAs, discovered novel miRNA biomarkers, and constructed accurate miRNA panels with clinical value for diagnosing hepatitis in all patients. RNA libraries were prepared using the TruSeq Small RNA Library Prep Kit and sequenced on the HiSeq 2000 System.

The identification of circulating miRNA in bovine serum and their potential as novel biomarkers of early mycobacterium avium paratuberculosis infection. *PLoS One*. 2015;10:e0134310.

RNA-Seq was performed in cattle to profile circulating miRNAs. Researchers found novel biomarkers with potential for early diagnosis of infection with the mycobacterium that causes Johne's Disease, a major source of cattle loss. RNA libraries were prepared using the TruSeq Small RNA Library Prep Kit and sequenced on the HiSeq 2500 System.

Illumina Workflows for Gene Expression Studies

Illumina offers fully integrated, RNA-to-data workflows, from initial library preparation to final data analysis. Illumina library prep kits are available for a wide range of RNA-Seq applications including, total RNA-Seq, mRNA-Seq, small RNA-Seq, low-quality samples, and more.



Gene Regulation Studies and DNA Methylation

What is Gene Regulation and Why is it Important?

The regulation of gene expression is a biological process that controls the temporal and spatial expression of gene products, including both mRNA and noncoding RNA transcripts. A wide range of mechanisms are used to increase or decrease gene expression, including binding of regulatory proteins to DNA motifs, binding of RNA polymerase to regulatory elements, and modulation of histone chromatin structure.[30] Determining how protein–DNA interactions regulate gene expression is essential for fully understanding many biological processes and disease states.[31] Chromatin immunoprecipitation sequencing (ChIP-Seq) can leverage NGS to efficiently determine the distribution and abundance of DNA-bound protein targets across the genome at base-pair level resolution.[32]

DNA methylation of cytosine-guanine dinucleotides (CpGs) also plays an important role in connecting the workings of the genetic code to changing environmental factors. It allows cells to acquire and maintain a specialized state and suppresses the expression of viral and nonhost DNA elements.

Aberrant DNA methylation and its impact on gene expression have been implicated in many disease processes, including cancer, neurological disorders, aging, and development.[33,34] High-throughput technologies, such as whole-genome bisulfite sequencing (WGBS), targeted bisulfite sequencing, and methylation microarrays, are powerful tools for investigating the dynamic state of DNA methylation across the genome. This epigenetic information is also highly complementary to DNA sequencing, genotyping, gene expression, and other forms of integrated genomic analysis. A recent study in *Nature Methods* reported the transcriptional and epigenetic methylation analysis of 61 murine stem cells, and found many previously unrecognized associations between methylated regulatory elements and the transcription of key pluripotency genes.[35]

Methylation Sequencing and Microarray Applications

Methylation Microarrays in Epigenome-Wide Association Studies (EWAS) of Obesity and Smoking

Over the past decade, researchers have found significant connections between behavioral epigenetic factors, such as diet, smoking, and exercise, and their effects on complex conditions including obesity and heart disease.[36,37] Because obesity and smoking are major risk factors for comorbidities, such as cancer, cardiovascular disease, and metabolic disorders, and because these complex conditions cannot be understood through genetics alone, researchers are taking a closer look at epigenetic mechanisms.[36] Advances in methylation array technology are making a big impact on the field of epigenetics and are enabling researchers to perform cost-effective whole-genome methylation studies with large sample cohorts. Recent studies using the Infinium® HumanMethylation450 BeadChip have shown that smoking not only alters methylation profiles in adult smokers, but that it also alters methylation profiles in newborns exposed to maternal smoking during pregnancy.[38,39] Another recent set of studies used the same HumanMethylation450 BeadChip to investigate the epigenetic effects of obesity. Researchers at the David Geffen School of Medicine found that obesity accelerates epigenetic aging in human liver tissue, and scientists at the Georgia Prevention Institute found that regions of differentially methylated loci correspond to previous GWAS studies, which have identified genes associated with hypertension, dyslipidemia, and type 2 diabetes in obese subjects.[40,41]

See how scientists today are using methylation arrays in EWAS studies of obesity and smoking

Methylomic aging as a window into the influence of lifestyle: tobacco and alcohol use alter the rate of biological aging. *J Am Geriatr Soc*. 2015;63(12):2519-2525.

Epigenetic profiling was performed on 71 loci previously associated with aging and on loci associated with cigarette and alcohol consumption. Researchers found that while alcohol can have varying effects on aging, smoking has strong negative effects at all levels. Methylation status was interrogated with the Infinium Human Methylation450 BeadChip Kit.

Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. *Environ Res*. 2016;144(Pt A):139-48.

Epigenetic profiling was performed on 26 loci previously associated with prenatal smoking exposure. Researchers found this epigenetic signature is detectable in childhood blood samples and can be used to measure exposure. Genome-wide methylation was measured with the Infinium HumanMethylation450 BeadChip Kit.

Epigenome-wide association study (EWAS) of BMI, BMI change, and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. 2015;24(15):4464-79

Epigenome-wide association study (EWAS) of BMI, BMI change, and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet.* 2015;24(15):4464-79.
Epigenetic profiling was performed on loci associated with elevated adiposity. Researchers found correlation of methylation status and physical markers of obesity. The Infinium HumanMethylation450 BeadChip Kit was used to measure genome-wide methylation in these samples.

Whole-Genome Bisulfite Sequencing in Cancer Research

Over the past decade, epigenetic alterations have been increasingly recognized as important drivers of cancer initiation and progression.[42-44] Whole-genome bisulfite sequencing (WGBS) has enabled the genome-wide mapping of methylation patterns in normal cells and cancer cells, and has broadened our view of the molecular mechanisms affected by altered patterns of methylation. Such mechanisms include hypermethylation, gene silencing, and chromatin remodeling.

Whole-genome methylation mapping of normal and tumor genomes has confirmed that nearly all cancer types show tens to hundreds of genes with abnormal gains in DNA methylation.[43,44] CpG hypermethylation has also been shown to silence DNA repair genes, downregulate tumor suppressor genes, and disrupt the microRNA regulation of oncogenic targets.[45-47] Every year, new studies emerge describing how epigenetic changes help drive the initiation and progression of cancer. These insights are providing a rich source of potential cancer biomarkers and therapeutic opportunities.

See how scientists are using WGBS in their research today.

Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat Biotechnol*. 2015;33:769-74.

Using WGBS, researchers found that the vast majority of reprogrammed cardiomyocytes or neural stem cells obtained from mouse fibroblasts by OSKM-induced transdifferentiation pass through a transient pluripotent state. WGBS libraries were prepared with the EpiGenome/TruSeq DNA Methylation Kit and sequencing was performed on a HiSeq System.

Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep*. 2015;10:1386-97.

This study describes a WGBS assay for DNA methylation mapping in small cell populations or in single cells. The method enables single-cell analysis of DNA methylation in a broad range of biological systems, including stem cell differentiation and cancer. Libraries for WGBS were prepared with the EpiGnome/TruSeq DNA Methylation Kit and sequencing was performed on HiSeq Systems.

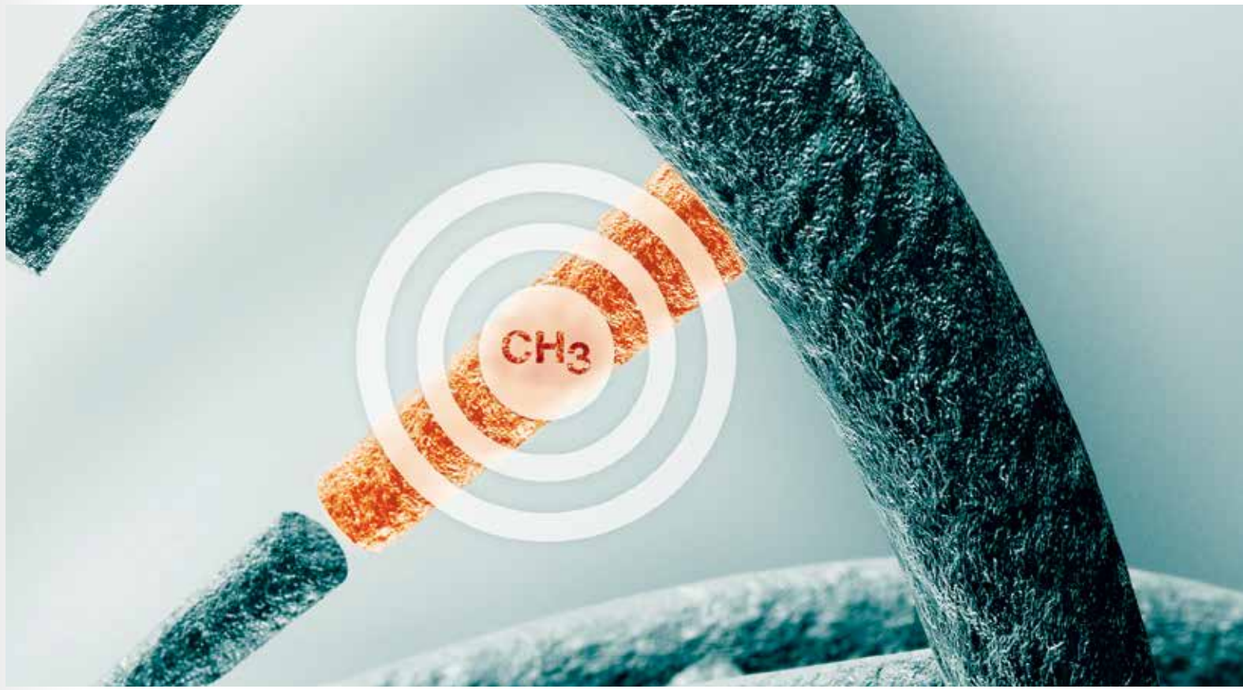
The DNA methylation landscape of Chinese hamster ovary (CHO) DP-12 cells. *J Biotechnol*. 2015;199:38-46.

Using WGBS and gene expression analysis, researchers demonstrate that CHO DP-12 cells exhibit epigenomic profiles similar to cancer and placenta cells, and that these features may be involved in the repression of tissue-specific genes. WGBS libraries were prepared with the EpiGnome/TruSeq DNA Methylation Kit, and sequenced on the HiSeq System.

Targeted Methylation Sequencing in Complex Disease Research

Diseases such as autism, diabetes, and muscular dystrophy are complex multigenic disorders with hundreds of susceptibility loci. Adding another layer to the challenge of understanding these complex conditions is the fact that external factors, such as the environment, diet, and exercise, interact with inherited genetic architecture to mediate disease progression.[48] Compared to WGBS, targeted methylation sequencing offers a cost-effective method for interrogating specific DNA regions of interest. These regions of interest can also be sequenced at high sequencing depth.

Epigenetic mechanisms have been shown to play a significant role in neurodevelopmental disorders, such as autism and Rett syndrome.[48] A recent study in *Molecular Psychiatry*, reported significant correlation between DNA methylation and autistic trait scores in an analysis of monozygotic twins discordant for autism.[49] Metabolic disorders are also influenced by the complex interplay between genetics and environmental factors. For example, nonalcoholic fatty liver disease (NAFLD), a disorder strongly associated with obesity and type 2 diabetes, is known to be influenced by caloric intake as well as genetic predisposition.[50] Targeted methylation sequencing offers a balanced, cost-effective choice that can support both screening and variant discovery applications for advancing breakthroughs in complex disease research.



See how scientists today are leveraging NGS targeted methylation sequencing in complex disease research.

Lactase nonpersistence is directed by DNA variation-dependent epigenetic aging. *Nat Struct Mol Biol*. 2016;23(6):566-73.

Targeted bisulfite sequencing was performed on mouse and human small intestines in the *LCT* and *MCM6* loci. Researchers found that epigenetically controlled regulatory elements account for the differences in lactase mRNA levels. Libraries were prepared with the bisulfite padlock-probe technique and sequencing was performed on a HiSeq 2500 System.

Brain feminization requires active repression of masculinization via DNA methylation. *Nat Neurosci*. 2015;18:690-7.

Gene expression and DNA methylation were measured in the sexually dimorphic preoptic area (POA) of male, female, and masculinized female mice. Researchers found that hormonally mediated decreases in DNA methylation activity result in masculinization of the POA. Bisulfite libraries were sequenced on an Illumina HiSeq System.

Comparison of methyl-capture sequencing vs. Infinium 450K methylation array for methylome analysis in clinical samples. *Epigenetics*. 2016;11:36-48.

Researchers evaluated and compared targeted sequencing and the Infinium HumanMethylation450 BeadChip in terms of coverage, technical variation, and concordance of methylation calls in clinical samples. Targeted Methyl-Seq libraries were sequenced on an Illumina HiSeq System.

Methods for Studying DNA Methylation

DNA Methylation is increasingly recognized as a key contributor to normal physiology, as well as the development of many diseases. Methylation status has been shown to play a role in cancer, obesity, diabetes, and more.[34,51,52] Depending on the goals and size of the study, microarray analysis and NGS are both used to study the effects of DNA methylation on development and disease.

Methylation Arrays

Complex diseases are influenced by a combination of genetic and environmental factors. Genome-wide association studies (GWAS) have played a critical role in the identification of genetic variants related to a broad range of diseases.[53] The emergence of DNA methylation microarrays propelled similar success with EWAS, resulting in the elucidation of novel molecular mechanisms of disease and identification of potential therapeutic biomarkers.[54] While some of the first DNA methylation arrays to enter the market a decade ago had less than 30K CpG sites, current arrays enable the interrogation of over 850K CpG sites[55] across the genome (Figure 6).[55,58-60] Due to their low cost per sample, methylation arrays are a preferred choice for EWAS and other studies with high sample numbers.

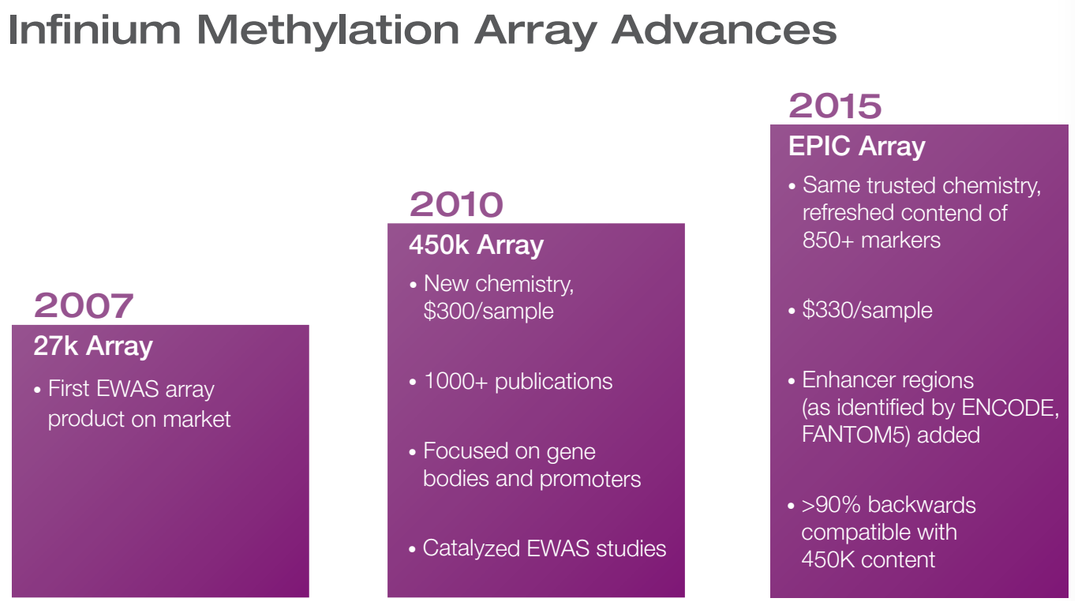


Figure 6: Advances in Methylation Arrays—Over the past decade, methylation arrays have evolved to include more content and a broader range of epigenetic region types.

Whole-Genome Bisulfite Sequencing

WGBS leverages the power of next-generation sequencing to provide a comprehensive view of methylation patterns across the genome. WGBS relies on bisulfite conversion of DNA to detect unmethylated cytosines. During library preparation, bisulfite conversion changes unmethylated cytosines to uracil. Converted bases are identified as thymine in the sequencing data and read counts are used to determine the percentage of methylated cytosines. NGS-based WGBS enables researchers to discover the methylation patterns of CpG (cytosine-guanine dinucleotides), CHH (H represents adenine, cytosine, or thymine), and CHG regions across the entire genome, at single base resolution. Therefore, in comparison to other methylation detection methods such as methylation arrays or targeted bisulfite sequencing, WGBS offers the most powerful and comprehensive approach for discovery-related applications.

Targeted Methylation Sequencing

A third option gaining increasing adoption is targeted methylation sequencing. With the targeted approach, bisulfite conversion is followed by the PCR amplification of specific regions of interest, followed by sequencing. Because only the targeted regions are sequenced, the cost per sample is lower compared to WGBS. Because targeted methylation sequencing only sequences a subset of the genome, it produces more manageable data sets and faster sequencing workflows compared to WGBS. A major benefit of targeted methylation sequencing is that it offers a balanced, cost-effective choice between WGBS and methylation arrays that can support both screening and variant discovery study objectives. Targeted methylation sequencing is ideal for hypothesis testing studies of target regions of interest as well as confirmation of regions identified in EWAS or GWAS studies.[56]

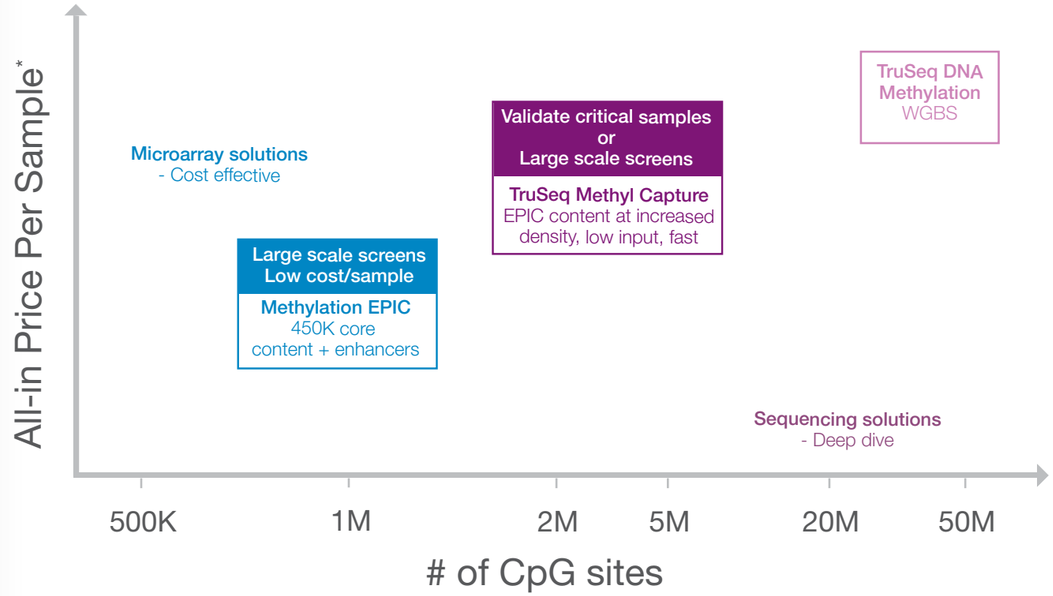


Figure 7: Diverse Methylation Methods— The detection of methylation patterns across the genome can be performed with a number of different methods to support discovery-related and/or screening-related study objectives. *Includes the library prep, cost of sequencer, library prep, indexes, bisulfite conversion, and enrichment reagents.

Illumina Methylation Sequencing Workflow

From library preparation to final data analysis, Illumina provides a fully optimized, integrated workflow for WGBS.

1 Prepare Library

- TruSeq DNA Methylation Kit

2 Sequence

- NextSeq Series (0.5 WGBS samples/flow cell)
- HiSeq 2500 System (2 WGBS samples per flow cell)
- HiSeq 3000/4000 System (2-3 WGBS samples per flow cell)

3 Align/Count

- MethylSeq App - Maps bisulfite-treated sequencing reads to the genome of interest and performs methylation calls using the Bismark algorithm. Bowtie 2, an ultrafast, memory-efficient tool, aligns the read to long reference sequences.
- MethylKit App - Performs differential methylation analysis.
- BaseSpace Data Repository - View example WGBS data generated using the TruSeq DNA Methylation Kit and Illumina sequencing instruments (use the “Methyl-Seq” category filter).

Illumina Methylation Array Workflow

From initial array hybridization to array scanning and final data analysis, Illumina provides a comprehensive, integrated workflow for epigenomics studies using methylation arrays.

1 Prepare Sample

- MethylationEPIC BeadChip - PCR-Free, low sample input (as low as 250 ng), kit using Infinium I and Infinium II assay chemistry technologies.

2 Scan

- iScan® System - Supports high throughput BeadChip processing, scanning hundreds to thousands of samples quickly and accurately.

3 Analyze

- GenomeStudio® Software and Methylation Module - Perform differential methylation analysis and display of chromosomal coordinates, % GC, location in CpG Island, and methylation B values.

Automation Solutions for RNA-Seq

Advances in NGS Library Preparation

While NGS and RNA-Seq are opening the door to new and exciting research breakthroughs, every successful sequencing run requires an equally successful library preparation. Depending on the methods used and the scale of sequencing required, library preparation for RNA-Seq can be time-consuming and create bottlenecks in the sequencing operations workflow. Automated tools can help by minimizing human error, reducing hands-on time, and enabling higher throughput. Previously, only extremely well-funded labs could afford to automate their library preparation operations. Along with the growth in NGS-based research, has come an expansion in the number and variety of automated library prep solutions.

Microfluidics Systems

Microfluidics systems handle liquids in much smaller volumes (often nanoliter to picoliter), typically using mechanical, electrical, or magnetic forces. They cater to different ranges of throughput, from low to high (depending on the application and platform). Some systems are dedicated to 1 type of application or have a narrow focus; others are more flexible and offer a broader set of applications. Multiple microfluidics systems can work together to allow for higher-volume processing and have low- to mid-range capital costs. Microfluidics systems deliver significant advantages over manual library preparation at lower capital costs than most traditional liquid handlers.[57]

Summary: Advantages of Microfluidics Systems Over Manual Library Preparation and Traditional Liquid Handlers

- Reduced hands-on time
- High reproducibility
- Reduced chance of human error
- Low sample input required
- Reduced operations bottlenecks with 24/7 library preparation

Robotic Liquid Handlers

Robotic liquid handling systems are powerful, flexible systems geared towards the greatest demands in high-throughput sample production. Liquid handlers can be programmed for various protocols and allow researchers the flexibility to adjust experimental conditions. Liquid handling robotic systems deliver the highest performance in terms of sample throughput, but also require the highest amount in capital expenditures. While liquid handlers deliver increased reproducibility, reduced hands-on-time, and reduced human error, the challenges with implementation typically include significant capital investment, higher operational expenses, and longer time needed for personnel training.

No limits: Rethinking automated capabilities.

Automation has evolved to an extent that it's no longer safe to make assumptions that certain applications only work with certain types of automated library prep technologies. Take RNA-Seq, for example. Microfluidics hasn't always been considered an option for RNA-Seq library prep, but today it is. For both microfluidics and robotics, sample kits today are simpler and more automation friendly, enabling researchers and clinicians to expand the boundaries of what's possible.

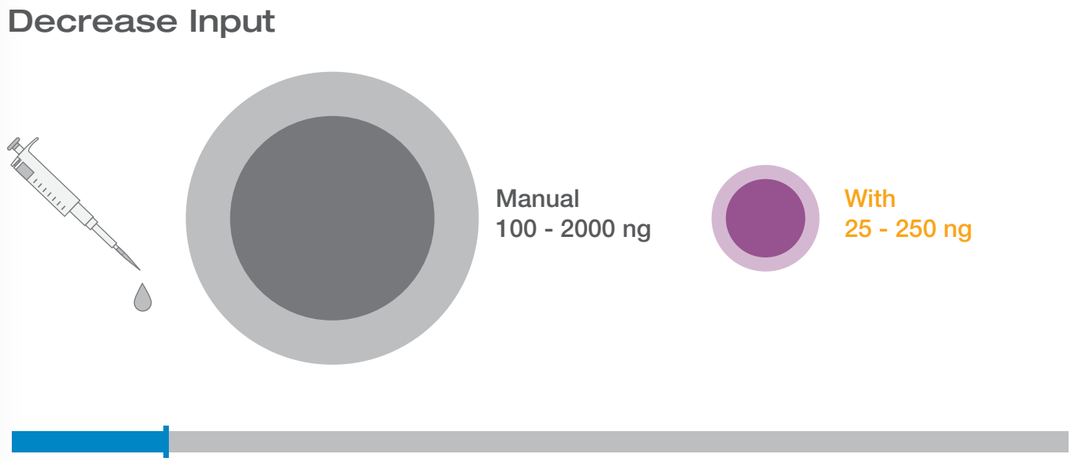


Figure 8: Microfluidic Systems Allow Lower DNA Input Requirements

See how scientists today are leveraging automation solutions for NGS.

Webinar—NeoPrep™ System for RNA-Seq at the SciLifeLab in Stockholm

Max Käller, PhD of the SciLifeLab describes using the NeoPrep System for TruSeq mRNA Library Prep for low-input samples. Data comparing libraries prepared on the NeoPrep System and a liquid handler, shows equivalency and reproducibility on automated platforms.

Summary

Over the past decade, advances in genomics and transcriptomics have led to an improved understanding of complex diseases, cancer biology, and environmental impacts on human health. NGS capabilities have shifted the scope of transcriptomics from the interrogation of a few genes at a time to the profiling of genome-wide expression levels in a single experiment. EWAS support these findings with an additional layer of information, indicating methylation status, and the genetic response to environmental cues. Indeed, one of the most exciting aspects of the genomics revolution is the ability to weave transcriptional, epigenetic, and genetic studies into an integrated view at an unprecedented pace and at an unprecedented scale.

While much progress has been made since the advent of NGS, so much more remains to be explored and discovered. Combining a broad library prep portfolio, high-quality data, and user-friendly analysis apps, Illumina RNA-Seq solutions empower researchers to investigate the molecular mechanisms of human health and disease. Since the introduction of the first sequencing system in 2006, Illumina has been committed to accelerating the pace of research through continuous innovation. Together, in collaboration with scientists from around the world, Illumina works to bring the power of NGS toward a deeper understanding of human biology and toward the promise of advanced precision medicine for future generations.

Learn More

How do I get started with RNA-Seq?

To start planning your RNA-Seq experiment, take advantage of these resources:

- Buyer's Guide: Simple RNA-Seq Workflows
- Buyer's Guide: NGS Systems
- FAQs: RNA-Seq Data Analysis
- Tools: Library Prep Kit Selector
- NeoPrep Library Prep System

What if I need help during a sequencing run or with data analysis?

Whether you have basic data analysis questions that require immediate attention or you have advanced questions requiring in-depth consultations, Illumina can help. Beyond immediate phone and email support, Illumina customer service and support teams provide a full suite of expedient solutions from initial trainings, to instrument support, personalized consultation, and ongoing NGS education. Illumina customer support offerings include:

Illumina Technical Support

Global, 24/5 phone and email support in the Americas, Europe, and Asia-Pacific.

Illumina Technical Support specialists can perform desktop sharing with GoToAssist—a powerful tool for quick identification and diagnosis of issues over the phone with live desktop sharing. For faster case handling, enter your case number at the main phone menu to be routed directly to the Technical Support specialist handling your case.

Illumina University Training

- Instructor-Led Training at your chosen facility
- Instructor-Led Training at an Illumina Training Center
- On-Line courses
- Webinars

Illumina Consulting Services

- Proof-of-Concept Services for instrument and library preparation testing
- Concierge Custom Design Service for design assistance and product optimization
- Illumina Bioinformatics Professional Services for bioinformatics consultation and/or training
- Illumina Genomics IT Consulting Services for genomics IT solutions
- Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ)

Who can I talk to for more information on RNA-Seq?

To speak with an Illumina representative about RNA-Seq solutions, call the Illumina Customer Solutions Center at 1.800.809.4566 (North America) or 1.858.202.4566 (Outside North America) and start planning your RNA-Seq experiments today.

Glossary

coverage level: The average number of sequenced bases that align to each base of the reference DNA. For example, a whole genome sequenced at 30× coverage means that, on average, each base in the genome was sequenced 30 times.

deep sequencing: Sequencing to high coverage levels. For example, WGS is typically performed to 30×–75× coverage while targeted NGS enables sequencing depths of 5000× or higher.

discovery power: In genomics, the ability to identify novel variants.

microfluidics: A technology that works with small volumes of fluids. For library preparation, this is a category of automated systems that can cater to different ranges of throughput, from low to high, depending on the application and platform. Provides high level of consistency and requires fewer steps and less hands-on time than manual methods.

multiplexing: A process where a unique DNA sequence is ligated to fragments within a sequencing library for downstream, *in silico* sorting and identification. Indexes are typically a component of adapters or PCR primers and are ligated to the library fragments during the sequencing library preparation stage. Illumina indexes are typically between 8–12 bp. Libraries with unique indexes can be pooled together, loaded into 1 lane of a sequencing flow cell, and sequenced in the same run. Reads are later identified and sorted via bioinformatic software.

mutation resolution: The size of mutation, in base pairs that a technology is able to detect. For example, karyotyping provides a mutation resolution of 5–10 Mb, while array comparative genomic hybridization provides "higher resolution" by detecting mutations down to 50 kb. NGS techniques provide the highest possible mutation resolution because they can provide single base pair variant detection (detect the presence of a mutation) and nucleotide identification (detect the identity of a mutation).

next-generation sequencing (NGS): A non-Sanger-based high-throughput DNA sequencing technology. Compared to Sanger sequencing, NGS platforms sequence as many as billions of DNA strands in parallel, yielding substantially more throughput and minimizing the need for the fragment-cloning methods that are often used in Sanger sequencing of genomes.

paired-end reads: A strategy involving sequencing of 2 different regions that are located apart from each other on the same DNA fragment. This strategy provides elevated physical coverage and alleviates several limitations of NGS platforms that arise because of their relatively short read length.

quantitative real-time polymerase chain reaction (qPCR): An application that enables the measurement of nucleic acid quantities in samples. The nucleic acid of interest is amplified with the polymerase enzyme. The level of the amplified product accumulation during PCR cycles is measured in real time. These data are used to infer starting nucleic acid quantities.

read: In general terms, a sequence "read" refers to the data string of A,T, C, and G bases corresponding to the sample DNA. With Illumina technology, millions of reads are generated in a single sequencing run. In more specific terms, each cluster on the flow cell produces a single sequencing read. For example, 10,000 clusters on the flow cell would produce 10,000 single reads and 20,000 paired-end reads.

reverse transcription PCR (RT-PCR): An application to measure RNA expression levels using qPCR. RNA starting material is reverse transcribed into complementary DNA (cDNA) by the reverse transcriptase enzyme. Expression levels are usually expressed as a relative value, in comparison to the expression of a reference gene.

robotic liquid handling: A robot that dispenses a selected quantity of reagent, samples, or other liquid to a designated container. Offers consistency and accuracy over manual methods; often used for high-volume samples. Flexible option for labs that lack predictability of applications to be performed.

Sanger sequencing: The sequencing method, also known as capillary electrophoresis sequencing, developed in 1977 by Frederick Sanger. It involves sequencing DNA based on the selective incorporation of chain-terminating dideoxynucleotides by the DNA polymerase enzyme during *in vitro* DNA replication.

sensitivity: In genomics, the ability to detect low-frequency variants, or low-abundance transcripts.

sequencing by synthesis (SBS): A technology that uses 4 fluorescently labeled nucleotides to sequence the tens of millions of clusters on a flow cell surface in parallel. During each sequencing cycle, a single labeled dNTP is added to the nucleic acid chain. The nucleotide label serves as a "reversible terminator" for polymerization: after dNTP incorporation, the fluorescent dye is identified through laser excitation and imaging, then enzymatically cleaved to allow the next round of incorporation. As all 4 reversible terminator-bound dNTPs (A, C, T, G) are present, natural competition minimizes incorporation bias. Base calls are made directly from signal intensity measurements during each cycle, which greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that eliminates sequence-context-specific errors, enabling robust base calling across the genome, including repetitive sequence regions and within homopolymers.

sequencing panel: A subset of genes or target regions identified as regions of interest for a specific area of research.

target region: A specific sequence of the genome, identified as a region of interest, due to possible involvement in or association with biological development, pathogenesis, or other area of study of interest to the investigator. The sequence can be a gene, a gene segment, a gene fusion, a promotor region, part of an intron or exon, or any stretch of sequence of interest to the investigator.

targeted resequencing: A subset of genes or regions of the genome are isolated and selectively enriched or amplified before sequencing. Targeted approaches using NGS allow researchers to focus time, expenses, and data analysis on specific areas of interest. Such targeted analysis can include the exome (the protein-coding portion of the genome), specific genes of interest (custom content), targets within genes, or mitochondrial DNA.

References

1. Ozsolak F, Milos PM. RNA-Sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 2011;12:87-98.
2. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10:57-63.
3. Wilhelm BT, Landry JR. RNA-Seq—quantitative measurement of expression through massively parallel RNA-Sequencing. *Methods.* 2009;48:249-57.
4. Mate SE, Kugelman JR, Nyenswah TG, et al. Molecular Evidence of Sexual Transmission of Ebola Virus. *N Engl J Med.* 2015;25:2448-54.
5. Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 2013;7453:236-40.
6. Fatica A, Bozzoni I. Long noncoding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* 2014;1:7-21.
7. Hamilton JP, Hansey CN, Whitty BR, et al. Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics.* 2011;12:302.
8. Chandriani S, Skewes-Cox P, Zhong W, et al. Identification of a previously undescribed divergent virus from the *Flaviviridae* family in an outbreak of equine serum hepatitis. *Proc Natl Acad Sci.* 2013;15:E1407-15.
9. Publication Data obtained from PubMed (www.ncbi.nlm.nih.gov/pubmed). Accessed March 2016.
10. Grant award data obtained from the NIH website (projectreporter.nih.gov). Accessed March 2016.
11. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One.* 2014;16;9(1):e78644.
12. Wang C, Gong B, Bushel PR, et al. The concordance between RNA-Seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol.* 2014;32:926-932.
13. Illumina (2014) TruSeq RNA Access Library Prep Kit Data Sheet. (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet-truseq-rna-access.pdf).
14. Illumina (2011) RNA-Seq Data Comparison with Gene Expression Microarrays White Paper (www.europeanpharmaceuticalreview.com/wp-content/uploads/Illumina_whitepaper.pdf).
15. Marguerat S, Bähler J. RNA-Seq: from technology to biology. *Cell Mol Life Sci.* 2010;67:569-79.
16. Wang W, Niu Z, Wang Y, et al. Comparative transcriptome analysis of atrial septal defect identifies dysregulated genes during heart septum morphogenesis. *Gene.* 2016;575:303-12.
17. Cao Y, Goods BA, Raddassi K, et al. Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. *Sci Transl Med.* 2015;7:287ra74.
18. Leti F, Malenica I, Doshi M. High-throughput sequencing reveals altered expression of hepatic microRNAs in nonalcoholic fatty liver disease-related fibrosis. *Transl Res.* 2015;166:304-14.
19. Rennert RC, Schäfer R, Bliss T, et al. High-resolution microfluidic single-cell transcriptional profiling reveals clinically relevant subtypes among human stem cell populations commonly utilized in cell-based therapies. *Front Neurol.* 2016;7:41.
20. Bhattacharya B, Puri S, Puri RK, et al. A review of gene expression profiling of human embryonic stem cell lines and their differentiated progeny. *Curr Stem Cell Res Ther.* 2009;4:98-106.
21. Tasic B, Menon V, Nguyen TN, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci.* 2016;19:335-46.
22. Mertens F, Tayebwa J. Evolving techniques for gene fusion detection in soft tissue tumours. *Histopathology.* 2014;64:151-62.
23. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer.* 2007;7:233-45.
24. García-García J, Ramos C. Influenza, an existing public health problem. [Article in Spanish] *Salud Publica Mex.* 2006;48:244-67.
25. Calvet G, Aguiar RS, Melo AS, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis.* 2016;16:653-60.
26. Ebola situation reports. Geneva: World Health Organization, 2015 (apps.who .int/ Ebola/ Ebola-situation-reports). Accessed June 2016.
27. Xu J, Gong B, Wu L, et al. Comprehensive assessments of RNA-Seq by the SEQC consortium: FDA-led efforts advance precision medicine. *Pharmaceutics.* 2016;8(1).
28. Khodakov D, Wang C, Zhang DY. Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Adv Drug Deliv Rev.* 2016 Apr 15. [Epub ahead of print].
29. Gao H, Korn JM, Ferretti S, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med.* 2015;21:1318-25.
30. Yilmaz A, Grotewold E. Components and mechanisms of regulation of gene expression. *Methods Mol Biol.* 2010;674:23-32.
31. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science.* 2007;316:1497-1502.
32. Illumina (2014) TruSeq ChIP Library Prep Kit Data Sheet (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_truseq_chip_sample_prep_kit.pdf).
33. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol.* 2010;28:1057-68.
34. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 2011;12:529-41.
35. Angermueller C, Clark SJ, Lee HJ, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods.* 2016;13:229-32.
36. van Dijk SJ, Molloy PL, Varinli H, et al. Epigenetics and human obesity. *Int J Obes (Lond).* 2015;39:85-97.
37. Singh GB, Sharma R, Khullar M. Epigenetics and diabetic cardiomyopathy *Diabetes Res Clin Pract.* 2011;94:14-21.
38. Joubert BR, Håberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect.* 2012;120:1425-31.
39. Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Falchi M, Suhre K. Association of DNA methylation with age, gender, and smoking in an Arab population. *Clin Epigenetics.* 2015;7:6.
40. Horvath S, Erhart W, Brosch M, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci.* 2014;111:15538-43.
41. Xu X, Su S, Barnes VA, et al. A genome-wide methylation study on obesity: differential variability and differential methylation. *Epigenetics.* 2013;8:522-33.
42. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet.* 2007;8:286-98.
43. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer.* 2011;11:726-34.
44. Jones PA, Baylin SB. The epigenomics of cancer. *Cell.* 2007;128:683-92.
45. Gao D, Herman JG, Guo M. The clinical value of aberrant epigenetic changes of DNA damage repair genes in human cancer. *Oncotarget.* 2016;7:37331-46.
46. Saito Y, Liang G, Egger G, et al. Specific activation of microRNA-127 with downregulation of the protooncogene *BCL6* by chromatin-modifying drugs in human cancer cells. *Cancer Cell.* 2006;9:435-43.
47. Ashida S, Nishimori I, Tanimura M, et al. Effects of *von Hippel-Lindau* gene mutation and methylation status on expression of transmembrane carbonic anhydrases in renal cell carcinoma. *J Cancer Res Clin Oncol.* 2002;128:561-8.
48. Vijayakumar NT, Judy MV. Autism spectrum disorders: Integration of the genome, transcriptome and the environment. *J Neurol Sci.* 2016;364:167-76.
49. Wong CC, Meaburn EL, Ronald A, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry.* 2014;19:495-503.
50. Wilfred de Alwis NM, Day CP. Genes and nonalcoholic fatty liver disease. *Curr Diab Rep.* 2008;8:156-63.

51. Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet.* 2014;383:1990-8.

52. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010;20:440-6.

53. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* 2009;106:9362-7.

54. Paul DS, Beck S. Advances in epigenome-wide association studies for common diseases. *Trends Mol Med.* 2014;20:541-3.

55. Illumina (2015) Infinium MethylationEPIC BeadChip Data Sheet (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/humanmethylationepic-data-sheet-1070-2015-008.pdf).

56. Masser DR, Stanford DR, Freeman WM. Targeted DNA methylation analysis by next-generation sequencing. *J Vis Exp.* 2015;(96).

57. Illumina (2016) NeoPrep Library Prep System Data Sheet (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/neoprep-system-data-sheet-970-2014-004.pdf).

58. Publication Data obtained from PubMed (https://www.ncbi.nlm.nih.gov/gquery/?term=450K). Accessed March 2016.

59. Illumina (2012) DNA Methylation Analysis Data Sheet. (www.illumina.com/Documents/products/datasheets/datasheet_dna_methylation_analysis.pdf).

60. Illumina (2012) Infinium HumanMethylation450 BeadChip Data Sheet (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf).

Industry-leading solutions.

A community of support.

From library prep, arrays, and sequencing to informatics, Illumina genomic solutions empower researchers and clinical researchers across the globe to find the answers they seek.

When you join the Illumina community, you become part of a dynamic scientific movement that includes thousands of researchers and industry thought leaders. Throughout the year, we host user group meetings, symposia, consortia, online forums, and other initiatives—all designed to bring the best minds together to share ideas and advance science.

In addition to on-site training, ongoing support, and phone consults, we offer webinars and courses at various Illumina locations. We're here with all the resources you need to accelerate progress.

Contact your Illumina representative to learn more or visit **www.illumina.com**

A global genomics leader, Illumina provides complete next-generation sequencing workflow solutions to the basic and translational research communities. Illumina technology is responsible for generating more than 90% of the world's sequencing data.* Through collaborative innovation, Illumina is fueling groundbreaking advancements in the fields of oncology, reproductive health, genetic disease, microbiology, agriculture, and forensic science.

*Data calculations on file. Illumina, Inc., 2015.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

Pub. No. 470-2016-003 Current as of 01 November 2016

illumina®