HHS Public Access

Author manuscript

Immunol Rev. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Immunol Rev. 2022 January ; 305(1): 9–28. doi:10.1111/imr.13033.

What sequencing technologies can teach us about innate immunity

Mays Mohammed Salih, **Susan Carpenter**

Department of Molecular Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA 95064

Summary

For years we have taken a reductionist approach to understanding gene regulation through the study of one gene in one cell at a time. While this approach has been fruitful it is laborious and fails to provide a global picture of what is occurring in complex situations involving tightly coordinated immune responses. The emergence of whole genome techniques provides a system level view of a response and can provide a plethora of information on events occurring in a cell from gene expression changes to splicing changes and chemical modifications. As with any technology this often results in more questions than answers, but this wealth of knowledge is providing us with an unprecedented view of what occurs inside our cells during an immune response. In this review we will discuss the current RNA-sequencing technologies and what they are helping us learn about the innate immune system.

Keywords

Next Generation Sequencing; Innate immunity; Long non-coding RNAs; Long-Read Sequencing; RNA modifications; CRISPR

Introduction

Over the last four decades there have been many technological advances in high throughput approaches to study gene expression from DNA microarrays to the development of next generation sequencing (NGS) [1]. NGS provides a wealth of knowledge in terms of biological processes from profiling of gene expression changes, to the identification of genetic variants including single nucleotide polymorphism (SNPs), to the study of splicing and chemical modifications. These tools are making a strong impact on a number of fields of research and here we will focus on what they have taught us so far and what they could be used for in the future in relation to regulation within the innate immune system.

Our innate immune system provides one of the first lines of defense against infection; It serves as a rapid response involving transient activation of inflammation [2]. This is essential

Correspondence to: Susan Carpenter, Ph.D., Associate Professor of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Mail Stop: 225 Sinsheimer Labs, 1156 High St, Santa Cruz, CA 95064, Phone: 8315028155, sucarpen@ucsc.edu; Mays Mohammed Salih, PhD Candidate, Carpenter Lab, Molecular, Cellular and Developmental Biology, University of California, Santa Cruz, 1156 High St, Santa Cruz CA, 95064, mmohamm7@ucsc.edu.

to maintaining homeostasis, but if left unchecked can become chronic and result in a host of inflammatory or autoimmune conditions [3]. Understanding the molecular mechanisms that govern inflammation and drive inflammatory and autoimmune diseases have presented a long-standing challenge due to the combination of genetic and environmental factors in addition to the complexity of the pathways involved [3], making it difficult to develop cures or even new drugs for therapeutic intervention. Since the development of NGS it has been used extensively to study inflammatory diseases from efforts such as large genome-wide association studies (GWAS) designed to identify possible disease-causing genes, to the identification of variants as well as studying altered gene expression programs [4]. In addition to the complexities surrounding protein regulation in the immune system, NGS has also unveiled the presence of 1000s of non-coding genes. It is now appreciated that the majority of any given genome is transcribed and yet only 3% is protein coding[5]. Figuring out the functional and biological significance of these transcripts in relation to innate immunity is only beginning. Finally, NGS has brought a renewed focus on the importance of post-transcriptional events such as splicing, RNA editing and RNA modifications during an immune response. As expected, inflammatory and autoimmune diseases are complicated; each cell type can be involved to varying degrees in the pathogenesis of any given disease driving the need for the development of single cell sequencing technology. It can be daunting to think about how we unravel such complexity in the immune response. Here we will review what NGS has allowed us to glimpse in terms of regulation within innate immunity. There is still a lot for us to learn and with the speed at which these technologies are developing we continue to get one step closer to producing better therapeutics with the long-term goal of eventually curing inflammatory and autoimmune diseases.

Evolution of RNA-sequencing

DNA sequencing has evolved rapidly from first generation Sanger sequencing to the so-called next generation sequencing (NGS) which includes second generation short-read sequencing to the more recent third generation long-read sequencing technologies [6,7,8]. NGS was quickly adopted as a tool to profile the transcriptome by isolating RNA and converting it to cDNA for sequencing (RNA-sequencing or RNA-seq) as an attractive alternative to microarray technology. RNA-seq possesses a number of advantages compared to microarrays including the detection of novel sequences, broad dynamic range, high specificity and sensitivity capable of picking up low abundance transcripts [9,10].

The first high-throughput sequencing platform appeared in 2005 [11] and was followed by multiple NGS platforms, the most common of which is the Illumina-based sequencing technology. Rapid growth in NGS use was prompted by its application in the whole genome sequencing project (WGS) and continued to grow as an essential tool due to its biomedical applications, its use in epidemiological studies of infectious diseases, surveillance of foodborne illnesses and viral diversity studies [12,13,14]. Massively parallel NGS technology or commonly known as “deep sequencing” refers to sequencing a genomic region multiple times, sometimes hundreds or even thousands of times (referred to as coverage) allowing for the detection of rare clonal cells, or microbes comprising as little as 1% of the original sample [15]. Advances in sequencing depth and error reduction elevated the field of biomedical discovery from studying individual genes in order to discover disease variants to

whole genome studies. The most commonly used sequencing platforms (second generation technologies) are generally divided into two categories: MiSeq or MiniSeq platforms which are relatively cheap and provide low to medium throughput, while HiSeq, NovaSeq or NextSeq are more expensive but provide high throughput (Table 1). All second-generation technologies provide fragmented short reads that require subsequent genome assembly. The low-cost high throughput sequencing technologies allowed for a deeper and more thorough understanding of genetic variation and complexity and allowed us an unprecedented view into novel transcripts and epigenetic regulation. These advances push us closer to personalized medicine where a patient's genome can be readily sequenced to try and detect disease associated variants. It currently takes less than a few days and costs about $1000 to sequence a human genome [16]. The cost will continue to decrease and soon patients will not just have their genomic DNA sequenced but also their transcriptome to obtain information on post-transcriptional regulatory events that could be dysregulated in a diseased state.

Large scale initiatives that utilized deep sequencing to study genetic variation have shown their effectiveness in covering >91% of the human genome with high confidence and resulted in the discovery of about 150 million SNPs in the coding and non-coding parts of the genome [17]. Thanks to recent advances in sequencing technology, as well as computational pipelines we now have a close to complete reference genome. The first reference genome (GRCh37) published in 2001 covered 90% of the human genome with 15,000 gaps, representing sequences from 13 donors constructed into a mosaic haploid genome [18,19]. The reference genome is now in its 20th rendition; GRCh38 published in 2013 with merely 738 unclosed gaps [20] with continual advances in NGS holding the promise of closing these gaps in the next iterations. GRCh38 remains limited because it represents genetic sequences from a few individuals and doesn't begin to cover the complex genetic variability especially in regions with high allelic diversity such as the major histocompatibility complex (MHC). Also, it fails to represent regions where haplotypes are represented in similar frequencies in different populations [19]. This paved the way for initiatives such as the "1000 genomes project" which was completed in 2015 and reconstructed the genomes of 2,504 individuals from 26 populations, it characterized over 88 million variants along with 3.6 million short insertions/deletions (indels) covering huge population diversity using multiple sequencing technologies [21]. While this work was a heroic undertaking, we need to expand these efforts if we are to appreciate the full genomic diversity of populations across the globe. There are populations that were never even sampled and others such as those of European descent that have been oversampled which has been reviewed in depth in [22]. Abi-Rached *et al.*, highlight the shortcomings of the 1000s genomes project when it comes to understanding the complexity of the immune system. The human leukocyte antigen (HLA) region is a highly polymorphic and well-studied region that encodes MHC molecules; however, data from the 1000 genomes project failed to detect over 70% of rare and 20% of common HLA variants [23]. There are serious immune conditions such as sickle cell anemia as well as autoinflammatory conditions such as systemic lupus erythematosus (SLE) that disproportionately impact African Americans compared to those of European descent [24]. Yet, in GWAS studies, Hispanic, African American and indigenous people continue to be under sampled. We need much more inclusive data if we are to fully appreciate how genetic diversity and genome plasticity contribute to disease states [22].

NGS provided the depth necessary to detect novel sequences and transcripts, such as long noncoding RNAs (lncRNAs), which we will discuss in depth later. It also allows for the detection of splice isoforms of the same gene, thus providing information about alternative splicing at a specific loci under treatment conditions, as well as alternative promoter usage and premature termination [25]. However, as Illumina short read sequencing arose to become the gold standard in genome profiling, progress continued to be made in developing newer tools to overcome short read technologies shortcomings. Some of these shortcomings include short read length (<300bp), which makes it difficult to detect structural variation, its size bias due to PCR amplification, insensitivity to highly repetitive or GC rich regions or homologous elements (Fig.1A) [26]. These limitations have contributed to failed attempts at understanding or identifying causal mutations and or dysregulated pathways in patients suffering from complex inflammatory diseases.

Exploring the dark matter of the genome

One of the biggest discoveries emerging from next gen RNA-sequencing studies was the fact that the majority of the genome is actively transcribed, yet only a small percentage <3% is translated into protein [5]. The next question was, and continues to be, what is all this RNA doing and is it biologically active? The largest group of non-coding RNA produced in the genome are lncRNAs (Fig. 2). LncRNAs are defined as transcripts >200bp in length with no protein-coding potential. LncRNAs exhibit low sequence conservation despite some stability in genetic loci conservation [31]. While there are nearly 18,000 lncRNA transcripts annotated in the human genome (Fig. 2) [5], the majority of them remain unstudied and their function remains uncharacterized. While these transcripts were first thought to be transcriptional noise it was clear early on that these are dynamically regulated regions and their exact functions are only beginning to be uncovered [32]. LncRNAs are often classified based on their orientation or site of transcription relative to their neighboring protein coding gene such as antisense, intronic and intergenic lncRNAs. There are also lncRNAs emerging from enhancer regions (eRNAs) as well as from promoters [33]. The largest group of lncRNAs are intergenic meaning they lie between two protein coding genes and contain their own independent promoters [34]. For a thorough review on lncRNA biogenesis and their many modes of post transcriptional regulation we recommend a recent review by Statello et al. [35].

LncRNAs can mediate regulation of genes through a wide variety of mechanisms which can broadly be broken into two categories of *cis* or *trans* regulators (Fig. 3, 1 and 2). LncRNAs functioning in *cis*, means they regulate their neighboring genes on the same allele. This form of regulation is perhaps not surprising given the fact that many lncRNAs display a similar expression pattern as their neighboring protein coding genes [35]. Interestingly, this regulation can be independent of the transcript itself and instead rely on recruitment of the transcriptional and splicing machinery to the neighboring locus (Fig. 3, 1.A) [36]. LncRNAs can also regulate neighboring genes in *cis* by acting as enhancer lncRNAs (e-lncRNAs) where they recruit mediators and co-activators to the locus and facilitate coordinate activation through chromatin looping between the enhancer and promoter of the neighboring gene (Fig. 3, 1.B) [37,38,39,40,41]. An example of *cis* regulation is the innate immune regulatory lncRNA, *Rroid*, that directly interacts with the promoter of its neighboring gene *Id2* in innate lymphoid cells (ILCs), promoting chromatin accessibility and deposition of STAT5

at the promoter of *Id2* prompting the cells to commit to an ILC fate [42]. LncRNAs can also repress expression of neighboring genes in *cis* (Fig. 3, 1.C) including *Morrbid*, which recruits PCR2 complex to *Bcl2l11* promoter through chromatin looping and allows PCR2 to deposit methyl tags at *Bcl2l11* promoter suppressing its expression [43,44]. Through this interaction with the promoter of pro-apoptotic *BcL2l11*, *Morrbid* is able to tightly regulate the survival of neutrophils, eosinophils and classical monocytes in response to pro-survival signals by cytokines, thus balancing an appropriate protective immune response against the deleterious consequences of prolonged activation.

LncRNAs can regulate genes on a different allele or different chromosome in *trans*. LncRNAs can function in *trans* through interactions with RNA binding proteins (RBPs) to regulate splicing or stability of transcripts (Fig. 3, 2.A and B) [45,46,47]. Alternatively, it can occur by binding of a lncRNA to mRNA transcripts through base pairing to promote or suppress stability and translation (Fig.3, 2.C and D) [48,49,50]. Others function by sequestering suppressors from the gene promoter to allow transcription factor binding and subsequent gene expression (Fig.3, 2.E) [51,52]. Most of the lncRNAs described to date in the innate immune system regulate genes in *trans*.

LncRNAs and innate immunity

LncRNAs play various roles in biological processes, including splicing [53], protein localization [54] and cellular proliferation [55,56]. LncRNAs are highly cell type specific in their expression patterns which makes them attractive as disease biomarkers for [57]. This is something that could be a particularly attractive area of investigation for autoinflammatory conditions such as arthritis and SLE which are notoriously difficult to diagnose quickly in the clinic.

Over the last decade there has been a significant increase in the number of lncRNAs being characterized to function in various ways within the immune system, from immune cell development to gene regulation. We cannot cover in depth all the lncRNAs identified to function in the immune system in this review and therefore we direct readers to the following reviews for in depth analysis of each lncRNA and its specific role in the innate immune system, reviewed in [58,59,60,61]. Here we will focus on lncRNAs that show some common or unique mechanisms of action within the immune system (Table 2).
One of the first long intergenic non-coding RNAs (lincRNAs) identified in the immune system is *lincRNA-Cox2*; It was first described to be induced ~1000 fold following lipopolysaccharide (LPS) activation by Guttman et. al, [62]. They utilized the chromatin signatures of active promoters (trimethylation of lysine 4 on histone 3, H3K4me3) and active transcription (trimethylation of lysine 36 on histone 3 H3K36me3) and performed chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq) to capture all actively transcribed genes including lncRNAs. Since this study we and others have shown that *lincRNA-Cox2* is a highly inflammatory inducible gene that functions broadly to regulate immune genes during the innate immune response [63,64,65,66,67,68]. Interestingly, using multiple genetic mice models we found that *lincRNA-Cox2* can function both in *cis* where it regulates the critical immune gene *Ptgs2 (Cox2)* through an enhancer RNA mechanism as well as in *trans* to regulate a wide variety of immune genes in macrophages

[67,68]. We found that *LincRNA-Cox2* can function in *trans* to negatively regulate basal expression of interferon stimulated genes through interactions with hnRNPA2/B1 and hnRNPA/B (Fig. 3, 2.B) [63]. Interestingly, many lncRNAs that function within the immune system appear to do so through interactions with various hnRNP proteins. TNFα and HNRNPL related immunoregulatory lincRNA (*THRIL*) was found to regulate expression of the *TNFA* gene through binding to HNRNPL forming a complex that binds to TNFA promoter region (Figure 3, 2.B) [69]. *LincRNA-EPS* represses immune response genes by associating with chromatin in the nucleus to create a heterochromatin (repressive) environment. It binds to HNRNPL through a specialized motif at 3' end forming a complex that represses immune gene expression (Fig.3, 2.B) [70].

P50 associated *Cox2* extragenic RNA (*PACER*) is another lncRNA that functions to regulate *Ptgs2* (also known as *Cox2*). *PACER* functions by sequestering the P50 repressive complex of NF-kB away from the *Ptgs2* promoter allowing recruitment of the active dimers of NF-kB and RNA pol II initiation complex to promote the activation of *Ptgs2* (Fig.3, 2.D) [51]. In a similar mechanism the lncRNA *NEAT1* was found to sequester the *IL8* repressor, splicing factor proline/glutamine rich (SFPQ) from the promoter into a heterochromatin structure "paraspeckle" leading to the transcriptional activation of *IL8* in response to viral infection or toll-like receptor 3 (TLR3) activation (Fig.3, 2.D) [52].

IL1B-eRNA is an example of an e-lncRNA in the immune system which acts as an enhancer for *IL1B* gene through binding to the PU.1 transcription factor and the *IL1B* promoter activating *IL1B* gene expression in response to an inflammatory stimulus (Fig.3, 1.B) [71,72].

While GWAS studies have mostly been utilized to study various SNPs arising in protein coding genes it is clear that over 90% of all SNPs lie within the non-coding space in the genome [73]. *Lnc13* was identified in a study by Castellanos-Rubio *et al.*, where they showed that the Celiac Disease associated SNP, rs917997, lies within this locus [74]. *Lnc13* regulates inflammatory genes and mediates its function *via* hnRNPD [74]. They showed that the SNP disrupts the RNA-protein interaction making the lncRNA dysfunctional.

LncRNAs can function within the cytoplasm or the nucleus to mediate their effects on immune genes. *Lethe* is a predominantly nuclear lncRNA that is involved in the negative feedback loop of the NF-κB pathway through direct binding to Rela and inhibiting its interaction and activation of genes within the nucleus [75]. *Lnc-DC* is localized to the cytoplasm where it directly binds to signal transducer and activator of transcription 3 (STAT3) to promote its phosphorylation and induce dendritic cell differentiation [76]. Recently we characterized the cytoplasmically localized lncRNA, gastric adenocarcinoma predictive long intergenic noncoding RNA (*GAPLINC*), as a conserved lncRNA that functions as a negative regulator of the inflammatory response in human and murine macrophages. *Gaplinc* KO mice are resistant to LPS induced endotoxic shock and mechanistically *GAPLINC* appears to function within the cytoplasm to control expression levels of NF-kB (Rela) and limit its localization to the cytoplasm during homeostasis.

Emergence of long read sequencing technologies:

As we continue to study the transcriptome, we need tools capable of capturing layers of genetic complexity that contribute to gene regulation including detection of strandedness, DNA and RNA modifications and splice variants. In addition, PCR amplification steps involved in the majority of library preparation protocols introduce biases such as large duplicate portions and uneven distribution of read coverage across targeted sequences [78]. In recent years, Pacific Biosciences (PacBio) and Oxford Nanopore (ONT) have developed multiple new sequencing techniques capable of producing continuous reads longer than 10kb in length directly from DNA or RNA which are helping researchers to answer complex biological questions (Fig 1 B and C). These approaches allow for *de novo* transcript assembly which means less reliance on the often error prone reference genome and without the need for PCR thus eliminating PCR bias.

Pacific Biosciences long read technology (PacBio):

The core technology emerging from PacBio is single molecule, real-time (SMRT) sequencing, where DNA is directly used to produce reads with read length around 10Kb. The initial technology was developed with low accuracy of 70–90% compared to illumina NGS accuracy of >99%, but increasing read accuracy was possible through read to read correction, despite being computationally intensive [79]. PacBio relies on a circular DNA template SMRTbell composed of a double stranded DNA insert flanked by two single stranded hairpin adapters on both ends (Fig.1 B) [26]. DNA polymerase is attached and the complex is read through a SMRT cell where DNA polymerase adds fluorescently labeled dNTPs and allows for base by base readout of the template [26].

Recent development in PacBio technology allowed for enhanced accuracy (>99%) through the development of high fidelity (HiFi) reads using circular consensus sequencing (CCS) [27]. The consensus sequence results from repeated passes of DNA polymerase through the template resulting in multiple error-prone subreads. Collectively, these subreads lead to a highly accurate consensus sequence with a high confidence that any detected variability is due to biological variants rather than sequencing errors [27]. In addition to long read length (average of 20kb) and high accuracy (>99%), PacBio technology provides uniform coverage across the template due to elimination of the amplification step; it can also sequence through regions that are inaccessible to Illumina due to high GC content, complexity, and repetition to achieve unambiguous mapping [26].

Oxford Nanopore Technologies (ONT):

Oxford Nanopore Technologies uses a linear DNA molecule attached to a sequence adapter loaded with a motor protein (Fig. 1 C) [26]. The motor protein feeds the DNA molecule through a nanopore embedded in a synthetic membrane, as the negatively charged DNA strand travels through the pore, individual bases cause a disruption in the current allowing calling of individual bases in real time [26]. ONT can generate continuous reads exceeding megabases in length, surpassing PacBio read length with a wide range of base calling accuracy [80].

ONT is the leading developer of direct RNA sequencing. This is a significant advance as it eliminates the cDNA synthesis step thus reducing errors associated with the reverse transcription step and allows for direct detection of modifications such as N6-methyladenosine (m^6A) and 5-methylcytosine (5-mC) [81] which we will cover in more detail in later sections [82]. Direct RNA sequencing produced a comparable number of reads that aligned to the transcriptome when compared to Illumina (79%) and slightly less than what was recorded for the cDNA long read dataset (90%). Both cDNA and direct RNA nanopore sequencing yield reads that are similar in length [83]. In addition, direct RNA sequencing displayed less of a bias towards transcript length and GC content than Illumina thus allowing for a more uniform coverage across the transcriptome [83]. While ONT platforms like the MinION are able to produce more than one million reads per run, there were concerns initially with the higher error rate associated with it [84]. In recent years researchers have applied many protocols and computational changes in an effort to increase accuracy. An example is the 2D sequencing protocol that involves ligating the template and the complementary strand of DNA using a hairpin; this enables both strands to pass through the pore and produce a more accurate consensus sequence [30]. A more recent approach, ID^2, involves sequencing both strands without the need for physical ligation and yields a high accuracy consensus of ~97% [30]. The Rolling Circle Amplification to Concatemeric Consensus (R2C2) method was developed by the Vollmers lab at UCSC leading to an increase in read accuracy as well as providing more comprehensive and quantitative analysis of RNA transcript isoforms [85]. R2C2 relies on introducing 8bp splints to both ends of the reverse transcribed cDNA, only full length-cDNA is then circularized and amplified using rolling circle amplification (RCA) [85]. Using this protocol and sequencing on a MinION generated more than 400,000 reads with base calling accuracy of 94%, covering the whole cDNA molecule [85]. In recent studies utilizing R2C2 and enhanced computational base calling programs the accuracy is now reaching 99.45% which rivals all short-read approaches [86], with a substantially enhanced ability to resolve transcript isoforms and avoid ambiguous mapping. Another advantage provided by R2C2 is the extremely low concentration input requirement (50ng); which enables accurate, high throughput sequencing from highly limited samples such as patient biopsies or blood samples.

Advantages to using long read sequencing to study Innate Immunity

While short read RNA-seq has provided us with vast new insights into gene regulation during inflammation we still do not have a complete picture of the key players and mechanisms involved in these complex processes. In this section we will outline the ways that long read technology can help expand our understanding of innate immunity, from better understanding what is being made within the genomes of immune cells, to gaining a picture of the post transcriptional regulatory changes that occur to the genes produced following inflammatory activation.

Construction of accurate and complete genomes:

While the majority of this review is covering the use of RNA-seq technologies to understand the immune system it is worth noting that we rely on the reference genome to interpret our RNA-seq data. Currently the reference genome is less than ideal for the reasons described

earlier. However, there are a number of ways in which researchers have been combining short and long read data to improve accuracy of the reference genome. Short read and long read technologies have their pros and cons. Short-read data is extremely accurate with high depth while long-read data is typically less accurate with shallow depth (Fig.1). However, several studies now pair the less accurate long read data with the highly accurate short read data to achieve maximum base calling accuracy [87,88]. Combining the two technologies means it is possible to obtain correctly mapped genomes covering highly repetitive regions. A recent study has applied this approach to datasets from monocytes and peripheral blood mononuclear cells (PBMCs) to characterize variations in 8 different immune system genomic loci [88]. They were able to construct a *de novo* assembly of the human leukocyte antigen, immunoglobulins, T cell receptors, and killer-cell immunoglobulin-like receptors. This study demonstrates the utility of accurate long read sequencing data in studying complex immune regulation loci and to aid in the discovery of novel structural variants in these regions which we will discuss in more depth later.

Despite continuous efforts and technological advances in NGS, gaps are still present in the latest human genome assembly GRCh38, these gaps are mainly associated with highly repetitive regions. These regions are often found around the centromeres, but they can now be resolved with the use of long accurate reads that resolve the entire region in one continuous read [89]. Ultralong nanopore sequencing provides an unprecedented advantage of being able to sequence an entire chromosome telomere to telomere with high base calling accuracy that will enable the gaps in the latest human genome assembly GRCh38 to be filled in. In a recent study, Miga *et al.*, performed the first high-coverage ultra-long-read nanopore sequencing that resulted in the first complete assembly of the human X chromosome [90]. In this study, they combined ONT ultra-long-reads with complementary technologies such as PacBio sequencing and high coverage Illumina sequencing for quality improvement and validation to sequence the hydatidiform mole CHM13 genome, which is a type of haploid organism and therefore a useful tool that has been used previously to assist with filling in gaps in diploid genomes. This approach allowed them to construct an assembly totaling 2.9Mb with half of the genome contained in continuous sequences with a continuity that exceeds GRCh38. Through the whole genome assembly in combination with polishing techniques, they were able to manually assemble the complete, gapless X chromosome *de novo* with accuracy that exceeds 99.99%. The continuing advancements in long read sequencing and analysis pipelines open a path for constructing the complete accurate human genome. This will improve mapping accuracy for techniques that rely on mapping to the reference genome such as RNA-seq, ChIP-seq and ATAC-seq.

One clear disadvantage to the hybrid approach just described is the fact that it requires generating different libraries and using multiple sequencing platforms in order to generate a complete genome. Instead, one approach used to assemble bacterial genomes utilizes “consensus polishing” where a subset of the longest reads from long read dataset such as PacBio are utilized as input for the assembly [91]. Hierarchical genome-assembly process (HGAP) assembler developed by Cin *et al.*, [92] uses the entire dataset from one library that consists of both long and shorter read fragments to correct errors in the longest reads and produce an accurate highly polished assembly. While this approach has only been

utilized for bacteria so far one can imagine its usefulness in assisting with the assemblies of eukaryotic genomes [91,92].

Understanding splicing and its role in immune response regulation

Splicing is the process of intron removal from a pre-mRNA transcript to produce a mature mRNA. Alternative splicing takes a number of forms including use of alternative start sites, alternative polyadenylation, exclusion or skipping exons, retention of introns, use of alternative start or final exons or use of alternative 5' or 3' splice sites. Alternative splicing (AS) is a highly regulated process enabling a single gene to produce multiple isoforms, thus increasing the complexity of gene function and the proteome [93,94,95,96]. Transcriptome profiling performed on many immune cell lineages uncovered that AS affects >60% of expressed genes and that B and T cells differed in AS events of genes that were similarly expressed in both lineages, indicating cell type specificity in the splicing process [97]. A study by *Pai et al.*, [95] used short reads NGS (SR-NGS) to study mRNA processing changes in macrophages in response to bacterial infections (*Salmonella typhimurium* and *Listeria monocytogenes*) and their impact on the overall immune response. They reported that 6–10% of genes switch their dominant isoform post infection with high enrichment for genes involved in the immune response. Interestingly, they reported that 47% of genes that displayed differential isoform usage were not differentially expressed following infection, which highlights the importance of studying isoform usage in addition to gene expression changes. They found an overall tendency to include skipped exons and to shorten 3'UTRs post infection. The shorter 3'UTRs eliminated target immune associated microRNA (miRNA) binding regions which inhibited their binding and allowed the transcript to escape repression. Their data indicates that the observed splicing changes could be carried out by heterogeneous nuclear ribonucleoproteins and serine and arginine-rich proteins (hnRNPs and SRs), which belong to splicing gene families and show an increase in expression post infection, in addition to themselves being subject to alternative splicing [95].

While short read data can be useful to study isoform usage its major disadvantage is that the reads might not always capture all spliced junctions. Paired end 150bp reads are useful but long read data is a game changer for the isoform profiling field. Advances in long read sequencing allowed for the detection of novel isoforms and quantification of isoform expression under different conditions. Long reads that exceed 1kb in length eliminated the need to assemble short read data to construct isoforms, allowing for a more accurate detection of isoforms and discovery of novel ones. In order to better understand isoform expression and isoform level changes in immune cells in response to inflammatory stimuli, we utilized ONT technology combined with R2C2 to generate an isoform level transcriptome atlas of macrophage activation (IAMA) following activation with a variety of inflammatory stimuli [86]. Using this method, we were able to generate 14,961,450 R2C2 reads at a median length 942nt across multiple ONT MinION flow cells. In addition, we were able to achieve an unprecedented increase in base calling accuracy from 97.9 to 99.45%. This enabled us to identify 29,637 high confidence isoforms; they included at least one isoform for 69% of the genes that are differentially expressed in any condition, and one isoform for 80% of the genes that were differentially expressed in all conditions. Of the total number of isoforms, 19%, were novel with annotated splice sites in

unannotated configurations (novel in catalog, NIC) and 7% were isoforms that use at least one unannotated splice site (novel not in catalog, NNC). This data is a reference database for researchers interested in knowing the exact isoform of their gene of interest that is expressed in a primary macrophage at baseline or following inflammatory stimulation. We find this data particularly useful for the study of lncRNAs as it can accurately map their full-length sequence without the need for laborious techniques such as 5'−3' RACE.

A combination of short and long read data is being utilized to help obtain a more accurate picture of the isoforms being produced following splicing. This was utilized by a study that performed Iso-seq (full length isoform sequencing through PacBio long read sequencing) in addition to illumina sequencing to study transcriptional diversity in whole blood samples [98]. They detected 57 isoforms at 42 loci that do not overlap with any GENCODE transcripts (unannotated) and are missed by short read sequencing [98]. We recently utilized both illumina short-read and ONT long-read data to study alternative splicing events in human and murine macrophages following inflammatory activation with LPS [99]. We showed that alternative first exon usage is the dominant splicing event, making up 50% of all events found in human and mouse macrophages following inflammatory activation. As mentioned earlier, it is known that the reference genome and current annotated transcriptome assemblies are incomplete. Therefore, we utilized the program full-length alternative isoform analysis of RNA (FLAIR) [100,101] to combine our long-read data with our short-read data to generate a new reference transcript from which to perform our splicing analysis. This approach enabled the identification of 95 novel alternative first exon (AFE) events in response to LPS, 50% of which were not differentially expressed at the RNA level following stimulation, again highlighting the importance of considering alternative splicing as a key regulatory mechanism during an immune response. We also discovered a novel isoform of cytosolic dsDNA sensor, Aim2, that is induced by an inflammatory stimulus, where an alternative first exon is used through alternative splicing and a new transcription start site (TSS). This novel inflammatory driven isoform is myeloid specific and is shown to be less efficiently translated when compared to the canonical form because this novel isoform possesses an iron specific translational mechanism through an iron-responsive element in its 5′UTR. These studies highlight the power of combining both short and long read sequencing to understand the splicing landscape of immune cells.

Applying these approaches in disease samples could help provide much needed insights into mechanisms of dysregulation. One of the most important takeaways from all the splicing studies is that many dominant isoform changes seen following inflammation are in proteins where they are not necessarily differentially regulated at the RNA level. These can be missed if one simply focuses on the top most up or down regulated genes. Instead by studying the splice sites it is possible to uncover disease specific isoforms of genes that could be missed by only focusing on differential expression approaches.

Identification of Structural Variation

Long-read sequencing is a reliable tool to study native and disease associated structural variations like copy number variations (CNVs), duplications, translocations and inversions [102]. These variations have been difficult to identify, due to their small size (could be as

small as 50bp) leading to inaccurate mapping when using short reads. Long reads enable more accurate mapping and a better understanding of the genetic architecture of the region in healthy and disease conditions. Studies have reported on the advances achieved by using new CCS technology when compared to short read sequencing, where they sequenced the well-characterized human HG002/NA24385 genome and obtained precision and recall rates of ~99.91% for single-nucleotide variants (SNVs), 95.98% for insertions and deletions <50 bp (indels) and 95.99% for structural variants [103]. Another study that utilized a SMRT based technology to sequence HX1 was able to construct an assembly that fills 28% of gaps in the reference genome GRCh38 and discover HX-1 specific sequence that has not yet been reported [98].

Studying RNA modifications in innate immunity

A-to-I editing

A-to-I editing is a form of post-transcriptional modification that occurs in all classes of eukaryotic RNA (mRNA, tRNA, rRNA and ncRNA) and is considered the most widespread RNA modification in mammals [104]. It involves the chemical change of an adenosine residue to an inosine by adenosine deaminase that acts on RNA (ADAR) [105], which in turn is recognized as Guanosine (G) by both translational and splicing machinery [106]. It serves as an essential mechanism for the immune system to differentiate between self and non-self where A-to-I editing occurs in endogenous double stranded (dsRNA) allowing it to avoid detection by the cytosolic dsRNA receptor MDA5 [107]. Studies on ADAR1 knockout mice show that loss of ADAR1 results in increased expression of type I IFNs as reviewed by Wang *et al.*, [108]. ADAR expression is inducible following activation with TNFA, IFNG or LPS in myoblasts [109], T cells and macrophages [110] indicating a possible regulatory role for A-to-I editing during the inflammatory response. When A-to-I editing occurs in coding regions it can lead to changes in protein sequence, however, A-to-I editing is most common in noncoding regions such as introns and UTRs [111], which can result in nuclear retention, degradation, alternative splicing, and translation regulation of the mRNA. Dysregulation of editing has been implicated in various inflammatory and autoimmune diseases [112,113,114].

Interestingly, many novel editing sites in lncRNAs have been recorded in glioblastomas (brain cancer) [115]. LncRNAs can form secondary folds that generate dsRNA making them substrates for ADAR editing [111]. Editing of lncRNAs can result in nuclear retention or degradation acting as a negative feedback mechanism to regulate lncRNA function. Since ADAR binds to double stranded regions, it could bind a lncRNA and inhibit another RNA binding protein from forming an interaction thereby impacting lncRNA function in both an editing dependent and independent manner. Considering the role that A-to-I plays in the immune system highlights the importance of being able to detect and quantify these changes accurately. Sanger sequencing was able to detect edited sites [116] and revealed that both edited and unedited transcripts can be expressed in the same tissue, and the ratio between the two can vary by tissue type and developmental stage. Next, high throughput NGS proved to be capable of detecting A-to-I editing, however, concerns around NGS biases persisted, including ambiguous mapping, sequencing error and genomic SNPs that could lead to inaccurate identification of edits [117]. Therefore, some considerations need to

be taken when designing an experiment to estimate differential editing between samples and treatments; sequencing depth and coverage are essential for accurate identification of edited sites. In addition, edits residing in repetitive elements exhibit low abundance and require ultra-high coverage for reliable detection and quantification [118]. Improvement in edit identification accuracy can be achieved by increasing sequencing depth, preferentially employing protocols for strand specific-paired end read sequencing [117]. This also highlights the need for a complete and accurate reference genome that represents the vast genetic diversity of the population to enable accurate mapping of detected edits and avoid errors due to sequencing or mapping artifacts. Combining high throughput short with long reads can be extremely beneficial for thorough A-to-I edits detection. Short reads provide the required accuracy to detect edits, while long read sequencing provides high mappability power and ability to map complex repetitive regions [119]. Nanopore direct RNA sequencing provides an added advantage by revealing the complexity of mRNA modification in full-length single molecule reads while avoiding the bias associated with PCR and reverse transcription steps [120]. In addition, selecting appropriate bioinformatics pipelines is essential for the analysis of NGS data to ensure the correct identification of edited sites, currently available analysis workflows are discussed in detail by Diroma et. al, [117].

Methylation

While there are ~170 chemical modifications that can occur on RNA, the most abundant internal modification in mRNA is N6-methyladenosine (M^6A) [121], accounting for approximately 50% of methylated ribonucleotides [122]. These modifications are involved in shaping the fate of the transcript and regulating many aspects of RNA metabolism including transcription, splicing, export, translation and stability [123]. Studying the importance of these modifications has created the field of “epitranscriptomics” [124], which has been greatly fueled by the development of NGS allowing for rapid identification and evaluation of these modifications. M^6A is a reversible modification, which ignited interest in its dynamics and the features that help regulate it [125]. It was discovered that the same position might only be modified in a fraction of transcripts, serving as further indication that m^6A could possibly serve a regulatory role in many biological processes [126].

There are three methyltransferases, methyltransferase-like 3 (METTL3), methyltransferase-like 14 (METTL14) and Wilms’ tumor 1-associating protein (WTAP) primarily responsible for shaping the m^6A RNA landscape by transferring a methyl group to the N-6 position of the adenosine base [127,128,129]. These tags are removed by demethylases “erasers” such as alpha-ketoglutarate-dependent dioxygenase fat mass and obesity-associated protein (FTO) [130] and RNA demethylase ALKBH5 [131].

More recently modifications and their roles in regulating lncRNAs have emerged. LncRNAs can function as decoys and scaffolds, which depend on the structure of the RNA and therefore a single modification like m^6A could improve or eradicate these RNA-protein interactions. One of the best studied lncRNAs is X-inactive specific transcript (*XIST*) whose job is to mediate silencing on the inactive X in females. Recently it has been reported that m^6A modifications within XIST are critical to its function [132]. Patil *et al.*, reported that YTH domain containing 1 (YTHDC1) recognizes m^6A sites on *XIST* and is required for

its function such that artificial tethering of this protein to XIST can rescue silencing in the absence of m^{6}A marks [132]. m^{6}A modifications within the lncRNA *THOR* are read by YTHDF1 and YTHDF2 and these interactions help stabilize the oncogenic lncRNA [133].

LncRNAs are also known to play an important role in the stress response; highly repetitive satellite III (HSATIII) l lncRNAs function in forming the stress bodies nSBs in response to thermal stress [134]. These GGAAU rich lncRNAs sequester serine and arginine rich splicing factors (SRSFs) during thermal stress to suppress splicing of hundreds of introns. In addition, it was found that methylation of the same GGAAU motifs sequesters m^{6}A reader proteins such as YTHDC1 to repress M^{6}A dependent splicing during the thermal recovery phase [135]. Thus, lncRNAs that constitute nSBs serve as gene regulation hubs by serving a dual function as molecular sponges for RNA splicing proteins and m^{6}A readers to regulate intron splicing events during thermal stress responses.

Recent studies have pointed to RNA modification involvement in immune regulation. A recent study investigated the involvement of m^{6}A in the inflammatory cycle of dental pulp disease [136]. They demonstrated that METTL3 depletion resulted in a decrease in inflammatory cytokine expression as well as a decrease in the phosphorylation of IKKα/β, p65 and IκBα in the NF-κB signaling pathway in addition to p38, ERK and JNK in the MAPK signaling pathway in dental pulp cells when treated with LPS. This was facilitated by an increase in production of the myeloid differentiation primary response 88 (MyD88) splice variant (MyD88S). MyD88S exerts a negative effect on TLR signaling pathways and limits the duration of innate immune activation. A separate study that utilized a pooled CRISPR screen approach showed similar results where they demonstrated that METTL3-deficient macrophages exhibited reduced TNFα production upon LPS stimulation and showed that METTL3 KO mice display susceptibility to bacterial infections and faster tumor growth [137]. METTL3 depletion resulted in loss of m^{6}A modifications on the TLR4 negative regulator *Irakm* slowing down its degradation and resulting in suppression of TLR4 activation.

M^{6}A has been implicated in viral propagation and antiviral immunity by impacting transcript stability. A study showed that viral infection of cells depleted of m^{6}A writer METTL3 or reader YTHDF2 resulted in an induction of interferon-stimulated gene production which suppressed viral propagation [138].

While NGS has rapidly increased the ability to study RNA modifications there are some limitations to the approaches described thus far. These include the need for the RT step that erases the modifications and renders them indistinguishable from regular RNA bases. Also, the RNA species of interest (mRNA and lncRNAs) are of low abundance which makes sequencing sensitivity a concern, in addition to limited computational tools capable of reliably distinguishing modified bases in sequencing data [125]. Most of the approaches to study m^{6}A involve use of antibodies to enable m^{6}A immunoprecipitation and sequencing, such as m^{6}A-seq and MeRIP-seq [125,139]. M^{6}A-seq was used to map the modification on the human and mouse transcriptome to better understand conservation and dynamic changes in response to treatments [139]. M^{6}A was shown to be most prevalent around stop codons near 3'UTRs [140] and within long exons and that this pattern is evolutionarily conserved, pointing

to a possible functional regulatory role [139]. There are some concerns of non-specific binding of the m^{6}A antibodies to other methyl tags [141,142]. For a review and resource on m^{6}A mapping we direct the readers to the following study by Xiang et. al [139].

Direct RNA sequencing

ONT provides the only library preparation to date to probe native RNA, without the need for an RT step, at single nucleotide resolution and with long reads. This approach is capable of detecting changes in the current as the nucleic acid travels through the pore with a sensitivity that allows for direct RNA bases calling with and without modifications (Fig. 1, C). A variety of computational programs were developed to read the modified bases such as ELIGOS [143], MINES [144] and EpiNano [145]. MINES (M^{6}A Identification using Nanopore Sequencing), was able to identify >40,000 m^{6}A sites at single base and isoform level resolution in primary human epithelial cell line [144]. Using direct RNA sequencing and meRIP-seq revealed that absence of m^{6}A either by silencing METTL3 or YTHDC1 resulted in an overall decrease in late viral RNAs, viral proteins and infectious progeny of Adenovirus [146]. This decrease was mainly a result of the decrease in late splicing efficiency indicating that m^{6}A regulates splicing of viral transcripts.

Single cell RNA sequencing technologies (scRNA-seq)

Any given cell population within the immune system whether it is macrophages, dendritic cells, neutrophils etc, are all heterogeneous populations. While bulk sequencing provides a systems level view of gene regulation within a given population of cells it fails to describe the variety of responses between individual cells following activation. Single cell RNA-seq (scRNA-seq) allows for the identification and classification of new cell types based on their gene expression profiles as well as providing insights into how cells within a population respond to a given stimulus [147,148,149] Shalek *et al.*, produced two of the earliest single cell studies examining the responses of bone marrow derived dendritic cells (BMDCs) to TLR ligand activation [149,150]. They demonstrated bimodal activity in both gene expression as well as splicing in BMDCs following stimulation with LPS [149]. There are a number of reasons why heterogeneity might be evident in a seemingly homogenous population of cells such as differences in cell state including cell cycle, stochastic gene expression differences to name a few. Shalek et. al, showed that there appears to be different maturity states across the BMDCs, and this is accompanied by distinct splicing patterns where on a population level many isoforms of a gene are identified, but there is a dominant isoform expressed in one cell compared to another [149]. They also noted that distinct precocious cells exist that produce large amounts of interferon early in the immune response and could impact neighboring cells in a paracrine manner through the secretion of IFNs [150].

ScRNA-seq has enabled the discovery of distinct classes of human dendritic cells as well as innate lymphoid cells [151,152,153,154,155]. ScRNA-seq has also provided insights into macrophage populations as well as their ability to fight infections including *Salmonella* [156,157]. Avraham *et al.*, showed that variation in host cell responses could be related to differences in bacterial factors within the invading bacteria [156]. Saliba *et al.*, took a similar approach and noted interestingly that cells harboring non proliferating bacteria

are in an M1 pro-inflammatory state while cells containing proliferating bacteria are in an anti-inflammatory (M2) state suggesting the bacteria can alter polarization states of host macrophages [157]. The power of scRNA-seq provides insights into the intricacies of immune responses and how much they vary even within what was initially thought to be a homogeneous population of cells.

The human cell atlas project is a consortium wide effect involving scientists across a number of disciplines coming together to map all the cells of the human body [158]. They will utilize scRNA-seq in addition to techniques such as Mass Cytometry, epigenome sequencing and in situ approaches to provide a complete picture of the active molecular pathways present in healthy cells. The hope is that by understanding what is occurring in healthy cells and tissues it will provide the framework needed to understand what goes wrong during a diseased state. There are similar efforts underway in Europe by the Lifetime initiative that are utilizing single cell approaches to better understand complex diseases as well as trying to dissect individual cells' response to treatment [159].

While these large consortium efforts are exciting and will undoubtedly provide us with enormous amounts of data, there are however limitations to this technology. It is costly and this limits the number of single cells you can study at a given time. This technology is rapidly evolving and with that cost will continue to go down. Kasmia *et al.*, provide a review that covers the scRNA-seq pipelines, as well as the pros and cons of the various techniques [160]. 10X genomics is one the most common scRNA-seq pipelines and involves mapping the 5' or 3' ends of polyA transcripts and so this approach only gives a glimpse into gene expression and not any information on the whole body of the gene. It lacks isoform and sequence variation information and also fails to capture genes that are not polyadenylated. The average number of reads from any given cell from a chromium platform averages ~10,000 reads, therefore one only gets to study the most abundant genes expressed in a given cell. There are platforms available for in depth analysis of single cells that are capable of getting up to 1 million reads such as the C1 platform, however, this greatly limits the number of cells that can be studied in a single experiment [160]. New library prep methods including R2C2 which was mentioned earlier can be utilized to generate single cell libraries that can be sequenced either using illumina or nanopore sequencing. Volden *et al.*, profiled 3000 peripheral immune cells using R2C2 and were capable of clustering them into their cell types (T, B cell, monocytes etc.) based on their gene expression profiles [161]. This is a powerful advance to single cell sequencing as it provides isoform level transcriptomes in addition to gene expression profiles of immune cells. Another disadvantage of all the sequencing technologies discussed thus far is that they require lysing of cells to extract RNA and therefore only provide a snapshot of what was happening at the time the cells were lysed. Live-seq is a new innovative approach designed to allow for sampling of single cells during a live immune response [162]. It can act as a recorder and allow researchers to evaluate the immune response in the same cell overtime. This technology was used to study macrophages sampled both at baseline and following LPS stimulation over time. They concluded that baseline levels of the protein NFKBIA (IκBα) and cell cycle state as the major determinants of the observed phenotypic changes [162]. As with all the technologies mentioned here scRNA-seq will undoubtedly continue to develop and become cheaper which will allow for more widespread use. With this will come more information which will

be useful in deciphering the complex immune responses that occur in healthy and diseased cells.

High-throughput Functional Characterization of genes: CRISPR

It is clear that deep sequencing approaches provide us with this unprecedented view of what is being produced from the genome, but it fails to provide insights into the biological relevance of all the transcription that is occurring. In order to make the most of all the emerging genomic sequencing it is necessary to establish rapid functional characterization pipelines. The development of CRISPR-Cas9 has helped revolutionize the field of functional genomics by offering a tool from which we can rapidly functionally characterize genes in our systems of interest. Cas9 is a deoxyribose nuclease (DNase) that can be specifically targeted to genomic regions *via* a guide RNA (gRNA) [163,164]. Classical use of CRISPR in its enzymatically active form is ideal for removing protein coding genes and determining a phenotype. Targeting of Cas9 to such region results in a blunt double-stranded DNA break that is repaired by the imprecise Non-Homologous End-Joining (NHEJ) DNA repair pathway, leading to small deletions or insertions that disrupt the open reading frame and therefore result in loss of the protein. The simplicity of the guide RNA cloning system makes it amenable to high throughput approaches meaning that hundreds to thousands of proteins can be studied in a pooled fashion in any biological context. This is an attractive pairing to high throughput screening in which you identify all the interesting proteins that are turned on or off in your system of choice and now you want to know which of those proteins are actually important in your biology of choice. This has been utilized by a number of groups to try and better understand the genes involved in the immune system.

Genome wide screens have been performed to identify new regulators of TNF, TLR3 signaling as well as the NLRP3 inflammasome [165,166,167]. There have been many pooled CRISPR screens performed to try and better understand host-viral interactions which is reviewed in [168]. Perturb-seq is an approach in which CRISPR screening is combined with single cell sequencing readouts allowing for both target identification and mechanistic insights in one experiment. Dixit *et al.*, used this approach to target 24 transcription factors in bone marrow derived dendritic cells and reconstruct the complex interplay between positive and negative regulators within the LPS signaling pathway [169].

We recently performed a pooled high throughput screen in macrophages where we targeted all annotated protein coding genes, microRNAs as well as targeting 3’UTRs of known essential genes. In addition to the inflammatory screen we performed a viability screen which identified all genes required for viability including macrophage specific viability genes such as IRF8 [170]. We also provided insights into new regulatory elements present in the 3’UTRs of essential genes. We made use of our recently developed NF-kB-GFP reporter system [66] and identified 115 novel regulators of NF-kB as well as showing that TNF can act as a negative regulator of the pathway in a cell intrinsic manner [170]. The majority of the pooled based screens have been performed in cell lines, but the technology is now being utilized to move towards *in vivo* screening as well as screening primary human cells. Lafleur et. al, have developed CHIME: CHimeric IMmune Editing using CRISPR in the bone marrow to study gene expression *in vivo* [171]. They performed a pooled *in vivo*

screen targeting 21 genes (using 110 sgRNAs) specific for T cell biology and identified the Protein Tyrosine Phosphatase Non-Receptor Type 2 (*Ptpn2*) as a negative regulator of CD8+ T cell-mediated responses to LCMV infection.

The Marson lab at UCSF have been pioneering ways to utilize high throughput CRISPR approaches to knockout or knockin genes in primary human T cells with the view to gaining insights into the molecular mechanisms in healthy and diseased T as well as developing tools for therapeutics use in the future [172,173]. They have shown how amenable their tools are to other immune cells by demonstrating their ability to knockout genes in primary CD14 monocytes, cells that are typically genetically intractable [174].

While enzymatically active Cas9 is powerful for targeting proteins, it is not so easy to employ to interrogate the function of lncRNAs that do not contain ORFs. Instead CRISPRi has been effectively utilized to target non-coding regions of the genome. CRISPRi involves a catalytically inactivated version of Cas9 fused to the KRAB (Krüppel associated box) chromatin-silencing domain which when targeted to the transcription start of a gene induces heterochromatin formation and silencing [175,176].

There are many versions of the CRISPRi system that have been used to study lncRNAs and they have been reviewed here [177]. More recently the system has been utilized to perform high throughput screens to rapidly determine which lncRNAs are important for viability. Liu *et al.*, [178], employed a CRISPRi platform targeting 16,401 lncRNAs in seven different cell lines including human transformed and induced pluripotent stem cells (iPSC) lines. They identified 499 lncRNAs required for cellular growth with cell type specificity, confirming that lncRNAs serve cell type specific functions [178]. However, it is important to note that there are some technical challenges to the CRISPRi system. Gilbert *et al.*, showed that there is an ideal guide RNA targeting window of ~ −500 to +500 nucleotides surrounding the transcription start site [179]. Therefore, in order to design a library to target all lncRNAs or protein coding genes, it is essential to know exactly where the start sites are. It is well appreciated that lncRNAs in particular are poorly annotated and even in our limited experience using screening tools we have found that many proteins are also incorrectly annotated [180,181,182]. Some of this comes down to the nature of the reference genome utilized for the design of the guide RNAs. Since many lncRNAs are cell type specific and even protein coding genes can have alternative start sites that are cell specific [183,184], it means that it is necessary to obtain sequencing data from your cell type of interest prior to designing a library.

While there are clearly many advantages to the use of CRISPR for functional genetics from the speed at which it can be carried out into the vast number of genes that can be interrogated at once, these are expensive and time-consuming approaches. Screens only work if the read out is amenable to a high throughput system which means these approaches are somewhat limited in the scope of biology, they can provide insights to. However, these functional technologies are continuing to evolve. We need continued innovation in high throughput functional assays if we wish to make sense of all the sequencing data that is being generated.

Combining CRISPR with long read sequencing

This is an incredibly powerful approach for studying disease causing mutations. A recent study used CRISPR-Cas9 to cut out the known oncogenes BRCA1 and 2 and combined this with long read sequencing to identify new structural variants [185]. They studied a family with a history of breast cancer that had negative results for mutations by traditional sequencing methods including whole exome sequencing. They identified a retrotransposon insertion which resulted in the formation of a pseudoexon in the *BRCA1* message and introduced a premature truncation [185]. There are many inflammatory diseases for which whole exome sequencing has produced underwhelming results. Perhaps if instead, we focused on possible disease associated genes and sequenced them in individual patients using this CRISPR-long read approach we might uncover the mechanism at play.

Conclusions and future directions

Since there are so many sequencing techniques and platforms available, it is important to decide exactly what question you want to answer for any given experiment. It is also important to recognise and utilize the wealth of data that comes from any of these experiments. Many RNA-seq experiments are undertaken to perform differential expression analysis and while this technique is indeed useful for this it is also incredibly expensive and there is so much more you can obtain from the data. These experiments capture not just protein coding genes but also many noncoding RNAs. LncRNAs are known to be more cell type specific in their expression levels compared to proteins so if someone is interested in looking for unique signatures in diseased versus healthy states it is worth analysing the data for lncRNAs as well as proteins. As mentioned, splicing is a key regulatory mechanism in any immune response, and it also changes in diseased conditions. Many isoforms being used after inflammation are in genes that don't show differential expression levels, instead they display isoform switching and this should not be overlooked especially when studying diseased conditions. We feel that the future lies with long read technology as it offers a wealth of information from isoform identification to RNA modifications. The major focus to date has been only on m^6A, but there remain 169 other modifications that have not been examined to any great extent. A recent consortium has been established and is led by Prof. Angela Brooks called The Long-read RNA-seq Genome Annotation Assessment Project (LRGASP). They have tasked researchers with comparing library preparation protocols as well as computational approaches in order to help set a standard for long read data capture experiments (https://www.gencodegenes.org/pages/LRGASP/). Long read data will provide a much clearer picture of the splicing and modification landscapes that will allow us to better appreciate disease specific isoforms or modified genes and could allow for more targeted approaches for therapeutic intervention for inflammatory and autoimmune conditions.

Acknowledgements

S.C is supported by R01 AI148413 from NIAID and R35GM137801 from NIGMS. M.M.S is partially supported by funding from the Tobacco related disease research program grant number 27IP-0017HC. We would like to thank Elektra Robinson, Dr. Miten Jain, Samaneh Kiany and Dr. Cristina Flores-Arenas for their critical reading of this manuscript. Figures 1 and 3 were generated using Biorender.com.

References:

1. Roh SW, Abell GCJ, Kim K-H, Nam Y-D, Bae J-W. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. Trends Biotechnol 2010;28(6):291–299. doi:10.1016/j.tibtech.2010.03.001 [PubMed: 20381183]
2. Monaco F, Robbins J. Defective thyroglobulin synthesis in an experimental rat thyroid tumor. Lack of membrane-bound sialyltransferase activity. J Biol Chem 1973;248(7):2328–2336. [PubMed: 4698223]
3. Chen L, Deng H, Cui H, et al. Inflammatory responses and inflammation-associated diseases in organs. Oncotarget 2018;9(6):7204–7218. doi:10.18632/oncotarget.23208 [PubMed: 29467962]
4. Petersen B-S, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. BMC Genet 2017;18(1):14. doi:10.1186/s12863-017-0479-5 [PubMed: 28193154]
5. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22(9):1775–1789. doi:10.1101/gr.132159.111 [PubMed: 22955988]
6. Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006;16(6):545–552. doi:10.1016/j.gde.2006.10.009 [PubMed: 17055251]
7. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17(6):333–351. doi:10.1038/nrg.2016.49 [PubMed: 27184599]
8. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol 2018;122(1):e59. doi:10.1002/cpmb.59 [PubMed: 29851291]
9. Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. Microarray and its applications. J Pharm Bioallied Sci 2012;4(Suppl 2):S310–312. doi:10.4103/0975-7406.100283 [PubMed: 23066278]
10. Meera Krishna B, Khan MA, Khan ST. Next-Generation Sequencing (NGS) Platforms: An Exciting Era of Genome Sequence Analysis. In: Tripathi V, Kumar P, Tripathi P, Kishore A, Kamle M, eds. Microbial Genomics in Sustainable Agroecosystems Springer Singapore; 2019:89–109. doi:10.1007/978-981-32-9860-6_6
11. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437(7057):376–380. doi:10.1038/nature03959 [PubMed: 16056220]
12. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 2018;24(4):335–341. doi:10.1016/j.cmi.2017.10.013 [PubMed: 29074157]
13. Lewis T, Loman NJ, Bingle L, et al. High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. J Hosp Infect 2010;75(1):37–41. doi:10.1016/j.jhin.2010.01.012 [PubMed: 20299126]
14. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol 2012;3:329. doi:10.3389/fmicb.2012.00329 [PubMed: 22973268]
15. Goldman D, Domschke K. Making sense of deep sequencing. Int J Neuropsychopharmacol 2014;17(10):1717–1725. doi:10.1017/S1461145714000789 [PubMed: 24925306]
16. Hayden EC. Technology: The $1,000 genome. Nature 2014;507(7492):294–295. doi:10.1038/507294a [PubMed: 24646979]
17. Telenti A, Pierce LCT, Biggs WH, et al. Deep sequencing of 10,000 human genomes. Proc Natl Acad Sci U S A 2016;113(42):11901–11906. doi:10.1073/pnas.1613365113 [PubMed: 27702888]
18. Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J 2020;18:9–19. doi:10.1016/j.csbj.2019.11.002 [PubMed: 31890139]
19. E pluribus unum. Nat Methods 2010;7(5):331. doi:10.1038/nmeth0510-331 [PubMed: 20440876]
20. Zhao T, Duan Z, Genchev GZ, Lu H. Closing Human Reference Genome Gaps: Identifying and Characterizing Gap-Closing Sequences. G3 (Bethesda) 2020;10(8):2801–2809. doi:10.1534/g3.120.401280 [PubMed: 32532800]

21. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526(7571):68–74. doi:10.1038/nature15393 [PubMed: 26432245]
22. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016;538(7624):161–164. doi:10.1038/538161a [PubMed: 27734877]
23. Abi-Rached L, Gouret P, Yeh J-H, et al. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS One 2018;13(10):e0206512. doi:10.1371/journal.pone.0206512 [PubMed: 30365549]
24. Ramos PS, Shedlock AM, Langefeld CD. Genetics of autoimmune diseases: insights from population genetics. J Hum Genet 2015;60(11):657–664. doi:10.1038/jhg.2015.94 [PubMed: 26223182]
25. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. Bioinform Biol Insights 2015;9(Suppl 1):29–46. doi:10.4137/BBI.S28991 [PubMed: 26609224]
26. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet 2020;21(10):597–614. doi:10.1038/s41576-020-0236-x [PubMed: 32504078]
27. Hon T, Mars K, Young G, et al. Highly accurate long-read HiFi sequencing data for five complex genomes. Sci Data 2020;7(1):399. doi:10.1038/s41597-020-00743-4 [PubMed: 33203859]
28. Payne A, Holmes N, Rakyan V, Loose M. Whale Watching with BulkVis: A Graphical Viewer for Oxford Nanopore Bulk Fast5 Files. Genomics; 2018. doi:10.1101/312256
29. Tyler AD, Mataseje L, Urfano CJ, et al. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Sci Rep 2018;8(1):10931. doi:10.1038/s41598-018-29334-5 [PubMed: 30026559]
30. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol 2018;19(1):90. doi:10.1186/s13059-018-1462-9 [PubMed: 30005597]
31. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 2011;147(7):1537–1550. doi:10.1016/j.cell.2011.11.055 [PubMed: 22196729]
32. Hon C-C, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature 2017;543(7644):199–204. doi:10.1038/nature21374 [PubMed: 28241135]
33. Wu H, Yang L, Chen L-L. The Diversity of Long Noncoding RNAs and Their Generation. Trends Genet 2017;33(8):540–552. doi:10.1016/j.tig.2017.05.004 [PubMed: 28629949]
34. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 2018;19(3):143–157. doi:10.1038/nrm.2017.104 [PubMed: 29138516]
35. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 2021;22(2):96–118. doi:10.1038/s41580-020-00315-9 [PubMed: 33353982]
36. Engreitz JM, Haines JE, Perez EM, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature 2016;539(7629):452–455. doi:10.1038/nature20149 [PubMed: 27783602]
37. Kim T-K, Hemberg M, Gray JM. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol 2015;7(1):a018622. doi:10.1101/cshperspect.a018622 [PubMed: 25561718]
38. Lai F, Orom UA, Cesaroni M, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 2013;494(7438):497–501. doi:10.1038/nature11884 [PubMed: 23417068]
39. Xiang J-F, Yin Q-F, Chen T, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. Cell Res 2014;24(5):513–531. doi:10.1038/cr.2014.35 [PubMed: 24662484]
40. Ballarino M, Cipriano A, Tita R, et al. Deficiency in the nuclear long noncoding RNA Charme causes myogenic defects and heart remodeling in mice. EMBO J 2018;37(18):e99697. doi:10.15252/embj.201899697 [PubMed: 30177572]

41. Petermann F, Pękowska A, Johnson CA, et al. The Magnitude of IFN-γ Responses Is Fine-Tuned by DNA Architecture and the Non-coding Transcript of Ifng-as1. Mol Cell 2019;75(6):1229–1242.e5. doi:10.1016/j.molcel.2019.06.025 [PubMed: 31377117]
42. Mowel WK, McCright SJ, Kotzin JJ, et al. Group 1 Innate Lymphoid Cell Lineage Identity Is Determined by a cis-Regulatory Element Marked by a Long Non-coding RNA. Immunity 2017;47(3):435–449.e8. doi:10.1016/j.immuni.2017.08.012 [PubMed: 28930659]
43. Kotzin JJ, Spencer SP, McCright SJ, et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. Nature 2016;537(7619):239–243. doi:10.1038/nature19346 [PubMed: 27525555]
44. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev Genet 2020;21(2):102–117. doi:10.1038/s41576-019-0184-5 [PubMed: 31729473]
45. Yap K, Mukhina S, Zhang G, Tan JSC, Ong HS, Makeyev EV. A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. Mol Cell 2018;72(3):525–540.e13. doi:10.1016/j.molcel.2018.08.041 [PubMed: 30318443]
46. Lee S, Kopp F, Chang T-C, et al. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell 2016;164(1–2):69–80. doi:10.1016/j.cell.2015.12.017 [PubMed: 26724866]
47. Tichon A, Perry RB-T, Stojic L, Ulitsky I. SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA. Genes Dev 2018;32(1):70–78. doi:10.1101/gad.309138.117 [PubMed: 29386330]
48. Kretz M, Siprashvili Z, Chu C, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 2013;493(7431):231–235. doi:10.1038/nature11661 [PubMed: 23201690]
49. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011;470(7333):284–288. doi:10.1038/nature09701 [PubMed: 21307942]
50. Carrieri C, Cimatti L, Biagioli M, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 2012;491(7424):454–457. doi:10.1038/nature11508 [PubMed: 23064229]
51. Krawczyk M, Emerson BM. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes. Elife 2014;3:e01776. doi:10.7554/eLife.01776 [PubMed: 24843008]
52. Imamura K, Imamachi N, Akizuki G, et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol Cell 2014;53(3):393–406. doi:10.1016/j.molcel.2014.01.009 [PubMed: 24507715]
53. Gonzalez I, Munita R, Agirre E, et al. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol 2015;22(5):370–376. doi:10.1038/nsmb.3005 [PubMed: 25849144]
54. Munschauer M, Nguyen CT, Sirokman K, et al. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. Nature 2018;561(7721):132–136. doi:10.1038/s41586-018-0453-z [PubMed: 30150775]
55. Li XL, Pongor L, Tang W, et al. A small protein encoded by a putative lncRNA regulates apoptosis and tumorigenicity in human colorectal cancer cells. Elife 2020;9:e53734. doi:10.7554/eLife.53734 [PubMed: 33112233]
56. Rossi M, Bucci G, Rizzotto D, et al. LncRNA EPR controls epithelial proliferation by coordinating Cdkn1a transcription and mRNA decay response to TGF-β. Nat Commun 2019;10(1):1969. doi:10.1038/s41467-019-09754-1 [PubMed: 31036808]
57. Zou Y, Xu H. Involvement of long noncoding RNAs in the pathogenesis of autoimmune diseases. J Transl Autoimmun 2020;3:100044. doi:10.1016/j.jtauto.2020.100044 [PubMed: 32743525]
58. Atianand MK, Caffrey DR, Fitzgerald KA. Immunobiology of Long Noncoding RNAs. Annu Rev Immunol 2017;35:177–198. doi:10.1146/annurev-immunol-041015-055459 [PubMed: 28125358]

59. Mowel WK, Kotzin JJ, McCright SJ, Neal VD, Henao-Mejia J. Control of Immune Cell Homeostasis and Function by lncRNAs. Trends Immunol 2018;39(1):55–69. doi:10.1016/j.it.2017.08.009 [PubMed: 28919048]
60. Walther K, Schulte LN. The role of lncRNAs in innate immunity and inflammation. RNA Biol 2021;18(5):587–603. doi:10.1080/15476286.2020.1845505 [PubMed: 33138685]
61. Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. Nat Immunol 2017;18(9):962–972. doi:10.1038/ni.3771 [PubMed: 28829444]
62. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009;458(7235):223–227. doi:10.1038/nature07672 [PubMed: 19182780]
63. Carpenter S, Aiello D, Atianand MK, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 2013;341(6147):789–792. doi:10.1126/science.1240925 [PubMed: 23907535]
64. Hu G, Gong A-Y, Wang Y, et al. LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-Mediated Chromatin Remodeling. J Immunol 2016;196(6):2799–2808. doi:10.4049/jimmunol.1502146 [PubMed: 26880762]
65. Tong Q, Gong A-Y, Zhang X-T, et al. LincRNA-Cox2 modulates TNF-α-induced transcription of Il12b gene in intestinal epithelial cells through regulation of Mi-2/NuRD-mediated epigenetic histone modifications. FASEB J 2016;30(3):1187–1197. doi:10.1096/fj.15-279166 [PubMed: 26578685]
66. Covarrubias S, Robinson EK, Shapleigh B, et al. CRISPR/Cas-based screening of long non-coding RNAs (lncRNAs) in macrophages with an NF-κB reporter. J Biol Chem 2017;292(51):20911–20920. doi:10.1074/jbc.M117.799155 [PubMed: 29051223]
67. Elling R, Robinson EK, Shapleigh B, et al. Genetic Models Reveal cis and trans Immune-Regulatory Activities for lincRNA-Cox2. Cell Rep 2018;25(6):1511–1524.e6. doi:10.1016/j.celrep.2018.10.027 [PubMed: 30404006]
68. Robinson EK, Worthington AK, Poscablo DM, et al. LincRNA-Cox2 Functions to Regulate Inflammation in Alveolar Macrophages during Acute Lung Injury. Immunology; 2021. doi:10.1101/2021.07.15.452529
69. Li Z, Chao T-C, Chang K-Y, et al. The long noncoding RNA THRIL regulates TNFα expression through its interaction with hnRNPL. Proc Natl Acad Sci U S A 2014;111(3):1002–1007. doi:10.1073/pnas.1313768111 [PubMed: 24371310]
70. Atianand MK, Hu W, Satpathy AT, et al. A Long Noncoding RNA lincRNA-EPS Acts as a Transcriptional Brake to Restrain Inflammation. Cell 2016;165(7):1672–1685. doi:10.1016/j.cell.2016.05.075 [PubMed: 27315481]
71. Ha S-D, Cho W, DeKoter RP, Kim SO. The transcription factor PU.1 mediates enhancer-promoter looping that is required for IL-1β eRNA and mRNA transcription in mouse melanoma and macrophage cell lines. J Biol Chem 2019;294(46):17487–17500. doi:10.1074/jbc.RA119.010149 [PubMed: 31586032]
72. IIott NE, Heward JA, Roux B, et al. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun 2014;5:3979. doi:10.1038/ncomms4979 [PubMed: 24909122]
73. Kumar V, Westra H-J, Karjalainen J, et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013;9(1):e1003201. doi:10.1371/journal.pgen.1003201 [PubMed: 23341781]
74. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, et al. A long noncoding RNA associated with susceptibility to celiac disease. Science 2016;352(6281):91–95. doi:10.1126/science.aad0467 [PubMed: 27034373]
75. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. Elife 2013;2:e00762. doi:10.7554/eLife.00762 [PubMed: 23898399]
76. Wang P, Xue Y, Han Y, et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science 2014;344(6181):310–313. doi:10.1126/science.1251456 [PubMed: 24744378]

77. Vollmers AC, Covarrubias S, Kuang D, et al. A conserved long noncoding RNA, GAPLINC, modulates the immune response during endotoxic shock. Proc Natl Acad Sci U S A 2021;118(7):e2016648118. doi:10.1073/pnas.2016648118 [PubMed: 33568531]

78. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods 2009;6(4):291–295. doi:10.1038/nmeth.1311 [PubMed: 19287394]

79. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020;21(1):30. doi:10.1186/s13059-020-1935-5 [PubMed: 32033565]

80. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 2018;36(4):338–345. doi:10.1038/nbt.4060 [PubMed: 29431738]

81. Liu H, Begik O, Lucas MC, et al. Accurate detection of m6A RNA modifications in native RNA sequences. Nat Commun 2019;10(1):4079. doi:10.1038/s41467-019-11713-9 [PubMed: 31501426]

82. Soneson C, Yao Y, Bratus-Neuenschwander A, Patrignani A, Robinson MD, Hussain S. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. Nat Commun 2019;10(1):3359. doi:10.1038/s41467-019-11272-z [PubMed: 31366910]

83. Garalde DR, Snell EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 2018;15(3):201–206. doi:10.1038/nmeth.4577 [PubMed: 29334379]

84. Byrne A, Beaudin AE, Olsen HE, et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. Nat Commun 2017;8:16027. doi:10.1038/ncomms16027 [PubMed: 28722025]

85. Volden R, Palmer T, Byrne A, et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. Proc Natl Acad Sci U S A 2018;115(39):9726–9731. doi:10.1073/pnas.1806447115 [PubMed: 30201725]

86. Vollmers AC, Mekonen HE, Campos S, Carpenter S, Vollmers C. Generation of an isoform-level transcriptome atlas of macrophage activation. J Biol Chem 2021;296:100784. doi:10.1016/j.jbc.2021.100784 [PubMed: 34000296]

87. Singh M, Al-Eryani G, Carswell S, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun 2019;10(1):3120. doi:10.1038/s41467-019-11049-4 [PubMed: 31311926]

88. Zhang J-Y, Roberts H, Flores DSC, et al. Using de novo assembly to identify structural variation of eight complex immune system gene regions. PLoS Comput Biol 2021;17(8):e1009254. doi:10.1371/journal.pcbi.1009254 [PubMed: 34343164]

89. Staden R A strategy of DNA sequencing employing computer programs. Nucleic Acids Res 1979;6(7):2601–2610. doi:10.1093/nar/6.7.2601 [PubMed: 461197]

90. Miga KH, Koren S, Rhie A, et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature 2020;585(7823):79–84. doi:10.1038/s41586-020-2547-7 [PubMed: 32663838]

91. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods 2015;12(8):733–735. doi:10.1038/nmeth.3444 [PubMed: 26076426]

92. Chin C-S, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10(6):563–569. doi:10.1038/nmeth.2474 [PubMed: 23644548]

93. Boudreault S, Martenon-Brodeur C, Caron M, et al. Global Profiling of the Cellular Alternative RNA Splicing Landscape during Virus-Host Interactions. PLoS One 2016;11(9):e0161914. doi:10.1371/journal.pone.0161914 [PubMed: 27598998]

94. Ivanov P, Anderson P. Post-transcriptional regulatory networks in immunity. Immunol Rev 2013;253(1):253–272. doi:10.1111/imr.12051 [PubMed: 23550651]

95. Pai AA, Baharian G, Pagé Sabourin A, et al. Widespread Shortening of 3' Untranslated Regions and Increased Exon Inclusion Are Evolutionarily Conserved Features of Innate Immune Responses

to Infection. PLoS Genet 2016;12(9):e1006338. doi:10.1371/journal.pgen.1006338 [PubMed: 27690314]

96. Wang Y, Liu J, Huang BO, et al. Mechanism of alternative splicing and its regulation. Biomed Rep 2015;3(2):152–158. doi:10.3892/br.2014.407 [PubMed: 25798239]
97. Ergun A, Doran G, Costello JC, et al. Differential splicing across immune system lineages. Proc Natl Acad Sci U S A 2013;110(35):14324–14329. doi:10.1073/pnas.1311839110 [PubMed: 23934048]
98. Shi L, Guo Y, Dong C, et al. Long-read sequencing and de novo assembly of a Chinese genome. Nat Commun 2016;7:12065. doi:10.1038/ncomms12065 [PubMed: 27356984]
99. Robinson EK, Jagannatha P, Covarrubias S, et al. Inflammation drives alternative first exon usage to regulate immune genes including a novel iron-regulated isoform of Aim2. Elife 2021;10:e69431. doi:10.7554/eLife.69431 [PubMed: 34047695]
100. Tang AD, Soulette CM, van Baren MJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat Commun 2020;11(1):1438. doi:10.1038/s41467-020-15171-6 [PubMed: 32188845]
101. Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat Methods 2019;16(12):1297–1305. doi:10.1038/s41592-019-0617-2 [PubMed: 31740818]
102. De Coster W, Van Broeckhoven C. Newest Methods for Detecting Structural Variations. Trends Biotechnol 2019;37(9):973–982. doi:10.1016/j.tibtech.2019.02.003 [PubMed: 30902345]
103. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol 2019;37(10):1155–1162. doi:10.1038/s41587-019-0217-9 [PubMed: 31406327]
104. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 2010;79:321–349. doi:10.1146/annurev-biochem-060208-105251 [PubMed: 20192758]
105. Singh M Dysregulated A to I RNA editing and non-coding RNAs in neurodegeneration. Front Genet 2012;3:326. doi:10.3389/fgene.2012.00326 [PubMed: 23346095]
106. Grammeltvedt R, Berg T. Influence of insulin on lysosomal activity and urea production in isolated parenchymal cells from rat liver. Hoppe Seylers Z Physiol Chem 1976;357(7):977–981. doi:10.1515/bchm2.1976.357.2.977 [PubMed: 1033139]
107. Yu Z, Chen T, Cao X. RNA editing by ADAR1 marks dsRNA as “self.” Cell Res 2015;25(12):1283–1284. doi:10.1038/cr.2015.135 [PubMed: 26596786]
108. Wang Q, Li X, Qi R, Billiar T. RNA Editing, ADAR1, and the Innate Immune Response. Genes (Basel) 2017;8(1):E41. doi:10.3390/genes8010041 [PubMed: 28106799]
109. Meltzer M, Long K, Nie Y, Gupta M, Yang J, Montano M. The RNA editor gene ADAR1 is induced in myoblasts by inflammatory ligands and buffers stress response. Clin Transl Sci 2010;3(3):73–80. doi:10.1111/j.1752-8062.2010.00199.x [PubMed: 20590675]
110. Yang J-H, Luo X, Nie Y, et al. Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. Immunology 2003;109(1):15–23. doi:10.1046/j.1365-2567.2003.01598.x [PubMed: 12709013]
111. Yang Y, Zhou X, Jin Y. ADAR-mediated RNA editing in non-coding RNA sequences. Sci China Life Sci 2013;56(10):944–952. doi:10.1007/s11427-013-4546-5 [PubMed: 24008387]
112. Gallo A RNA editing enters the limelight in cancer. Nat Med 2013;19(2):130–131. doi:10.1038/nm.3072 [PubMed: 23389604]
113. Shallev L, Kopel E, Feiglin A, et al. Decreased A-to-I RNA editing as a source of keratinocytes’ dsRNA in psoriasis. RNA 2018;24(6):828–840. doi:10.1261/rna.064659.117 [PubMed: 29592874]
114. Vlachogiannis NI, Gatsiou A, Silvestris DA, et al. Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. J Autoimmun 2020;106:102329. doi:10.1016/j.jaut.2019.102329 [PubMed: 31493964]
115. Silvestris DA, Scopa C, Hanchi S, Locatelli F, Gallo A. De Novo A-to-I RNA Editing Discovery in lncRNA. Cancers (Basel) 2020;12(10):E2959. doi:10.3390/cancers12102959 [PubMed: 33066171]

116. Ramaswami G, Li JB. Identification of human RNA editing sites: A historical perspective. Methods 2016;107:42–47. doi:10.1016/j.ymeth.2016.05.011 [PubMed: 27208508]
117. Diroma MA, Ciaccia L, Pesole G, Picardi E. Elucidating the editome: bioinformatics approaches for RNA editing detection. Brief Bioinform 2019;20(2):436–447. doi:10.1093/bib/bbx129 [PubMed: 29040360]
118. Lo Giudice C, Silvestris DA, Roth SH, et al. Quantifying RNA Editing in Deep Transcriptome Datasets. Front Genet 2020;11:194. doi:10.3389/fgene.2020.00194 [PubMed: 32211029]
119. Berbers B, Saltykova A, Garcia-Graells C, et al. Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus. Sci Rep 2020;10(1):4310. doi:10.1038/s41598-020-61158-0 [PubMed: 32152350]
120. Maitra RD, Kim J, Dunbar WB. Recent advances in nanopore sequencing. Electrophoresis 2012;33(23):3418–3428. doi:10.1002/elps.201200272 [PubMed: 23138639]
121. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell 2017;169(7):1187–1200. doi:10.1016/j.cell.2017.05.045 [PubMed: 28622506]
122. Wei CM, Gershowitz A, Moss B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. Cell 1975;4(4):379–386. doi:10.1016/0092-8674(75)90158-0 [PubMed: 164293]
123. Shi H, Wei J, He C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. Mol Cell 2019;74(4):640–650. doi:10.1016/j.molcel.2019.04.025 [PubMed: 31100245]
124. Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA stability. Exp Mol Med 2020;52(3):400–408. doi:10.1038/s12276-020-0407-z [PubMed: 32210357]
125. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods 2016;14(1):23–31. doi:10.1038/nmeth.4110 [PubMed: 28032622]
126. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. Proc Natl Acad Sci U S A 1984;81(18):5667–5671. doi:10.1073/pnas.81.18.5667 [PubMed: 6592581]
127. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA 1997;3(11):1233–1247. [PubMed: 9409616]
128. Liu J, Yue Y, Han D, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol 2014;10(2):93–95. doi:10.1038/nchembio.1432 [PubMed: 24316715]
129. Ping X-L, Sun B-F, Wang L, et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res 2014;24(2):177–189. doi:10.1038/cr.2014.3 [PubMed: 24407421]
130. Jia G, Fu Y, Zhao X, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 2011;7(12):885–887. doi:10.1038/nchembio.687 [PubMed: 22002720]
131. Zheng G, Dahl JA, Niu Y, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 2013;49(1):18–29. doi:10.1016/j.molcel.2012.10.015 [PubMed: 23177736]
132. Patil DP, Chen C-K, Pickering BF, et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature 2016;537(7620):369–373. doi:10.1038/nature19342 [PubMed: 27602518]
133. Liu H, Xu Y, Yao B, Sui T, Lai L, Li Z. A novel N6-methyladenosine (m6A)-dependent fate decision for the lncRNA THOR. Cell Death Dis 2020;11(8):613. doi:10.1038/s41419-020-02833-y [PubMed: 32792482]
134. Ninomiya K, Adachi S, Natsume T, et al. LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. EMBO J 2020;39(3):e102729. doi:10.15252/embj.2019102729 [PubMed: 31782550]

135. Ninomiya K, Iwakiri J, Aly MK, et al. m6 A modification of HSATIII lncRNAs regulates temperature-dependent splicing. EMBO J 2021;40(15):e107976. doi:10.15252/embj.2021107976 [PubMed: 34184765]
136. Feng Z, Li Q, Meng R, Yi B, Xu Q. METTL3 regulates alternative splicing of MyD88 upon the lipopolysaccharide-induced inflammatory response in human dental pulp cells. J Cell Mol Med 2018;22(5):2558–2568. doi:10.1111/jcmm.13491 [PubMed: 29502358]
137. Tong J, Wang X, Liu Y, et al. Pooled CRISPR screening identifies m6A as a positive regulator of macrophage activation. Sci Adv 2021;7(18):eabd4742. doi:10.1126/sciadv.abd4742 [PubMed: 33910903]
138. Winkler R, Gillis E, Lasman L, et al. m6A modification controls the innate immune response to infection by targeting type I interferons. Nat Immunol 2019;20(2):173–182. doi:10.1038/s41590-018-0275-z [PubMed: 30559377]
139. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 2012;485(7397):201–206. doi:10.1038/nature11112 [PubMed: 22575960]
140. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 2012;149(7):1635–1646. doi:10.1016/j.cell.2012.05.003 [PubMed: 22608085]
141. Kane SE, Beemon K. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. Mol Cell Biol 1985;5(9):2298–2306. doi:10.1128/mcb.5.9.2298-2306.1985 [PubMed: 3016525]
142. Zeng Y, Wang S, Gao S, et al. Refined RIP-seq protocol for epitranscriptome analysis with low input materials. PLoS Biol 2018;16(9):e2006092. doi:10.1371/journal.pbio.2006092 [PubMed: 30212448]
143. Jenjaroenpun P, Wongsurawat T, Wadley TD, et al. Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res 2021;49(2):e7. doi:10.1093/nar/gkaa620 [PubMed: 32710622]
144. Lorenz DA, Sathe S, Einstein JM, Yeo GW. Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. RNA 2020;26(1):19–28. doi:10.1261/rna.072785.119 [PubMed: 31624092]
145. Liu H, Begik O, Novoa EM. EpiNano: Detection of m6A RNA Modifications Using Oxford Nanopore Direct RNA Sequencing. Methods Mol Biol 2021;2298:31–52. doi:10.1007/978-1-0716-1374-0_3 [PubMed: 34085237]
146. Price AM, Hayer KE, McIntyre ABR, et al. Direct RNA sequencing reveals m6A modifications on adenovirus RNA are necessary for efficient splicing. Nat Commun 2020;11(1):6016. doi:10.1038/s41467-020-19787-6 [PubMed: 33243990]
147. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol 2018;18(1):35–45. doi:10.1038/nri.2017.76 [PubMed: 28787399]
148. Yang KY, Ku M, Lui KO. Single-cell transcriptomics uncover distinct innate and adaptive cell subsets during tissue homeostasis and regeneration. J Leukoc Biol 2020;108(5):1593–1602. doi:10.1002/JLB.6MR0720-131R [PubMed: 33070367]
149. Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013;498(7453):236–240. doi:10.1038/nature12172 [PubMed: 23685454]
150. Shalek AK, Satija R, Shuga J, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 2014;510(7505):363–369. doi:10.1038/nature13437 [PubMed: 24919153]
151. Breton G, Lee J, Zhou YJ, et al. Circulating precursors of human CD1c+ and CD141+ dendritic cells. J Exp Med 2015;212(3):401–413. doi:10.1084/jem.20141441 [PubMed: 25687281]
152. Villani A-C, Satija R, Reynolds G, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017;356(6335):eaah4573. doi:10.1126/science.aah4573 [PubMed: 28428369]

153. See P, Dutertre C-A, Chen J, et al. Mapping the human DC lineage through the integration of high-dimensional techniques. Science 2017;356(6342):eaag3009. doi:10.1126/science.aag3009 [PubMed: 28473638]

154. Björklund ÅK, Forkel M, Picelli S, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol 2016;17(4):451–460. doi:10.1038/ni.3368 [PubMed: 26878113]

155. Yu Y, Tsang JCH, Wang C, et al. Single-cell RNA-seq identifies a PD-1hi ILC progenitor and defines its development pathway. Nature 2016;539(7627):102–106. doi:10.1038/nature20105 [PubMed: 27749818]

156. Avraham R, Haseley N, Brown D, et al. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell 2015;162(6):1309–1321. doi:10.1016/j.cell.2015.08.027 [PubMed: 26343579]

157. Saliba A-E, Li L, Westermann AJ, et al. Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular Salmonella. Nat Microbiol 2016;2:16206. doi:10.1038/nmicrobiol.2016.206 [PubMed: 27841856]

158. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. Elife 2017;6:e27041. doi:10.7554/eLife.27041 [PubMed: 29206104]

159. Rajewsky N, Almouzni G, Gorski SA, et al. LifeTime and improving European healthcare through cell-based interceptive medicine. Nature 2020;587(7834):377–386. doi:10.1038/s41586-020-2715-9 [PubMed: 32894860]

160. Kashima Y, Sakamoto Y, Kaneko K, Seki M, Suzuki Y, Suzuki A. Single-cell sequencing techniques from individual to multiomics analyses. Exp Mol Med 2020;52(9):1419–1427. doi:10.1038/s12276-020-00499-2 [PubMed: 32929221]

161. Volden R, Vollmers C. Highly Multiplexed Single-Cell Full-Length CDNA Sequencing of Human Immune Cells with 10X Genomics and R2C2. Genomics; 2020. doi:10.1101/2020.01.10.902361

162. Chen W, Guillaume-Gentil O, Dainese R, et al. Genome-Wide Molecular Recording Using Live-Seq. Molecular Biology; 2021. doi:10.1101/2021.03.24.436752

163. Hochstrasser ML, Doudna JA. Cutting it close: CRISPR-associated endoribonuclease structure and function. Trends Biochem Sci 2015;40(1):58–66. doi:10.1016/j.tibs.2014.10.007 [PubMed: 25468820]

164. Sternberg SH, Doudna JA. Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell 2015;58(4):568–574. doi:10.1016/j.molcel.2015.02.032 [PubMed: 26000842]

165. Parnas O, Jovanovic M, Eisenhaure TM, et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell 2015;162(3):675–686. doi:10.1016/j.cell.2015.06.059 [PubMed: 26189680]

166. Zablocki-Thomas L, Menzies SA, Lehner PJ, Manel N, Benaroch P. A genome-wide CRISPR screen identifies regulation factors of the TLR3 signalling pathway. Innate Immun 2020;26(6):459–472. doi:10.1177/1753425920915507 [PubMed: 32248720]

167. Schmid-Burgk JL, Chauhan D, Schmidt T, et al. A Genome-wide CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Screen Identifies NEK7 as an Essential Component of NLRP3 Inflammasome Activation. J Biol Chem 2016;291(1):103–109. doi:10.1074/jbc.C115.700492 [PubMed: 26553871]

168. Krey K, Babnis AW, Pichlmair A. System-Based Approaches to Delineate the Antiviral Innate Immune Landscape. Viruses 2020;12(10):E1196. doi:10.3390/v12101196 [PubMed: 33096788]

169. Dixit A, Parnas O, Li B, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 2016;167(7):1853–1866.e17. doi:10.1016/j.cell.2016.11.038 [PubMed: 27984732]

170. Covarrubias S, Vollmers AC, Capili A, et al. High-Throughput CRISPR Screening Identifies Genes Involved in Macrophage Viability and Inflammatory Pathways. Cell Rep 2020;33(13):108541. doi:10.1016/j.celrep.2020.108541 [PubMed: 33378675]

171. LaFleur MW, Nguyen TH, Coxe MA, et al. A CRISPR-Cas9 delivery system for in vivo screening of genes in the immune system. Nat Commun 2019;10(1):1668. doi:10.1038/s41467-019-09656-2 [PubMed: 30971695]

172. Schumann K, Raju SS, Lauber M, et al. Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. Nat Immunol 2020;21(11):1456–1466. doi:10.1038/s41590-020-0784-4 [PubMed: 32989329]

173. Roth TL, Li PJ, Blaeschke F, et al. Pooled Knockin Targeting for Genome Engineering of Cellular Immunotherapies. Cell 2020;181(3):728–744.e21. doi:10.1016/j.cell.2020.03.039 [PubMed: 32302591]

174. Hiatt J, Cavero DA, McGregor MJ, et al. Efficient generation of isogenic primary human myeloid cells using CRISPR-Cas9 ribonucleoproteins. Cell Rep 2021;35(6):109105. doi:10.1016/j.celrep.2021.109105 [PubMed: 33979618]

175. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc 2013;8(11):2180–2196. doi:10.1038/nprot.2013.132 [PubMed: 24136345]

176. Ying Y, Yang X, Zhao K, et al. The Krüppel-associated box repressor domain induces reversible and irreversible regulation of endogenous mouse genes by mediating different chromatin states. Nucleic Acids Res 2015;43(3):1549–1561. doi:10.1093/nar/gkv016 [PubMed: 25609696]

177. Phelan JD, Staudt LM. CRISPR-based technology to silence the expression of IncRNAs. Proc Natl Acad Sci U S A 2020;117(15):8225–8227. doi:10.1073/pnas.2003702117 [PubMed: 32245807]

178. Liu SJ, Horlbeck MA, Cho SW, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 2017;355(6320):aah7111. doi:10.1126/science.aah7111 [PubMed: 27980086]

179. Gilbert LA, Larson MH, Morsut L, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 2013;154(2):442–451. doi:10.1016/j.cell.2013.06.044 [PubMed: 23849981]

180. Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet 2018;19(9):535–548. doi:10.1038/s41576-018-0017-y [PubMed: 29795125]

181. Salzberg SL. Next-generation genome annotation: we still struggle to get it right. Genome Biol 2019;20(1):92. doi:10.1186/s13059-019-1715-2 [PubMed: 31097009]

182. Boettcher M, Covarrubias S, Biton A, et al. Tracing cellular heterogeneity in pooled genetic screens via multi-level barcoding. BMC Genomics 2019;20(1):107. doi:10.1186/s12864-019-5480-0 [PubMed: 30727954]

183. Bertomeu T, Coulombe-Huntington J, Chatr-Aryamontri A, et al. A High-Resolution Genome-Wide CRISPR/Cas9 Viability Screen Reveals Structural Features and Contextual Diversity of the Human Cell-Essential Proteome. Mol Cell Biol 2018;38(1):e00302–17. doi:10.1128/MCB.00302-17

184. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res 2018;46(2):582–592. doi:10.1093/nar/gkx1165 [PubMed: 29202200]

185. Walsh T, Casadei S, Munson KM, et al. CRISPR-Cas9/long-read sequencing approach to identify cryptic mutations in BRCA1 and other tumour suppressor genes. J Med Genet Published online October 15, 2020:jmedgenet-2020–107320. doi:10.1136/jmedgenet-2020-107320

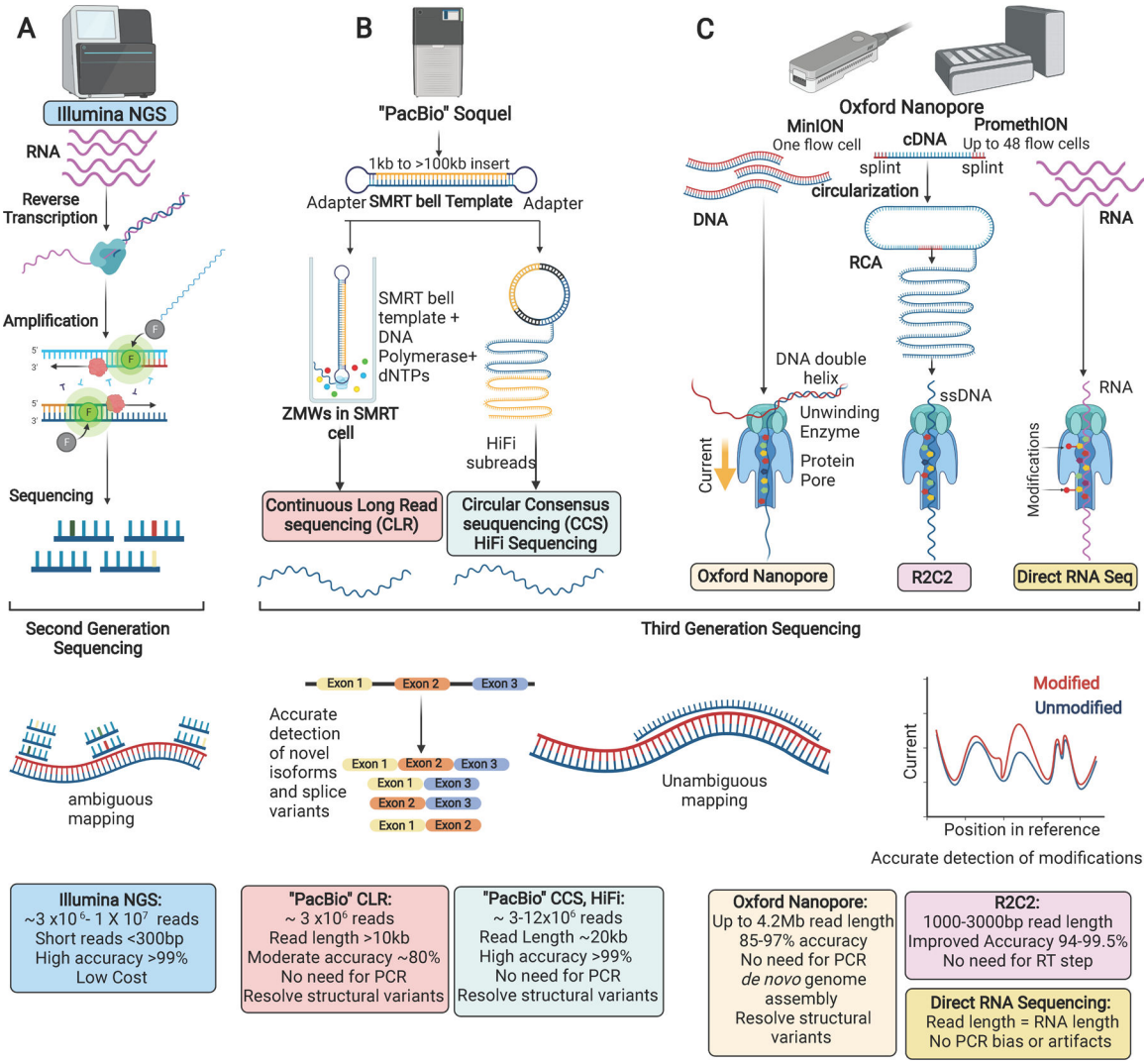

Figure 1. Evolution of RNA sequencing technologies.

A. Illumina NGS involves RT and amplification steps prior to sequencing. It is high throughput as it can yield anywhere from 3 million to 1 billion short read fragments <300bp in length. It currently provides highest read accuracy >99% in combination with low run cost. However, the short-read length results in ambiguous mapping and inability to resolve genomic variants and GC rich regions. **B.** Long read sequencing using PacBio technology generates up to 3 million long reads and relies on a circular DNA SMRTbell template composed of a double stranded DNA insert flanked by two single stranded hairpin adapters on both ends. DNA polymerase is attached and the complex is read in a zero mode waveguides (ZMW) SMRT cell where DNA polymerase adds fluorescently labeled dNTPs and allows for base by base readout of the template. The original technology relied on continuous long read (CLR) sequencing that yields reads longer than 10Kb with moderate accuracy ~80%. To increase read accuracy circular consensus sequencing (CCS) was introduced, it produces high fidelity (HiFi) reads through repeated passes of DNA polymerase through the template resulting in multiple error-prone subread, which when compiled produces a highly accurate consensus sequence. It yields a higher number of

reads compared to CLR with longer read length and accuracy that exceeds 99%. Both technologies eliminate the need for PCR, and produce reads long enough to detect variants, novel splice isoforms and can be mapped accurately. **C.** Oxford Nanopore Technologies (ONT) uses a linear DNA molecule attached to a sequence adapter loaded with a motor protein that pushes the DNA molecule through a nanopore. As the negatively charged DNA strand travels through the pore, individual bases cause a disruption in the current allowing us to call individual bases in real time. Many platforms were developed such as the MinION and the PromethION that differ in number of flow cells and subsequently in read number output. ONT can generate reads exceeding 1Mb in length with variable base calling accuracy 85–97% which is dependent on the protocol used to generate the reads and the computational program used for base calling. Nevertheless, it generates reads long enough for unambiguous mapping, allows for *de novo* genome assembly and detection of structural variants as well as novel isoforms. The Rolling Circle Amplification to Concatemeric Consensus (R2C2) method relies on introducing 8bp splints to both ends of the reverse transcribed cDNA, only full length-cDNA is then circularized and amplified using rolling circle amplification (RCA). It is used to increase accuracy - can produce >94% read accuracy- and to increase resolution of RNA transcript isoforms. Additionally, ONT allows for direct RNA sequencing which eliminates the need for the RT and PCR steps and allows for detection of native isoforms and RNA modifications.

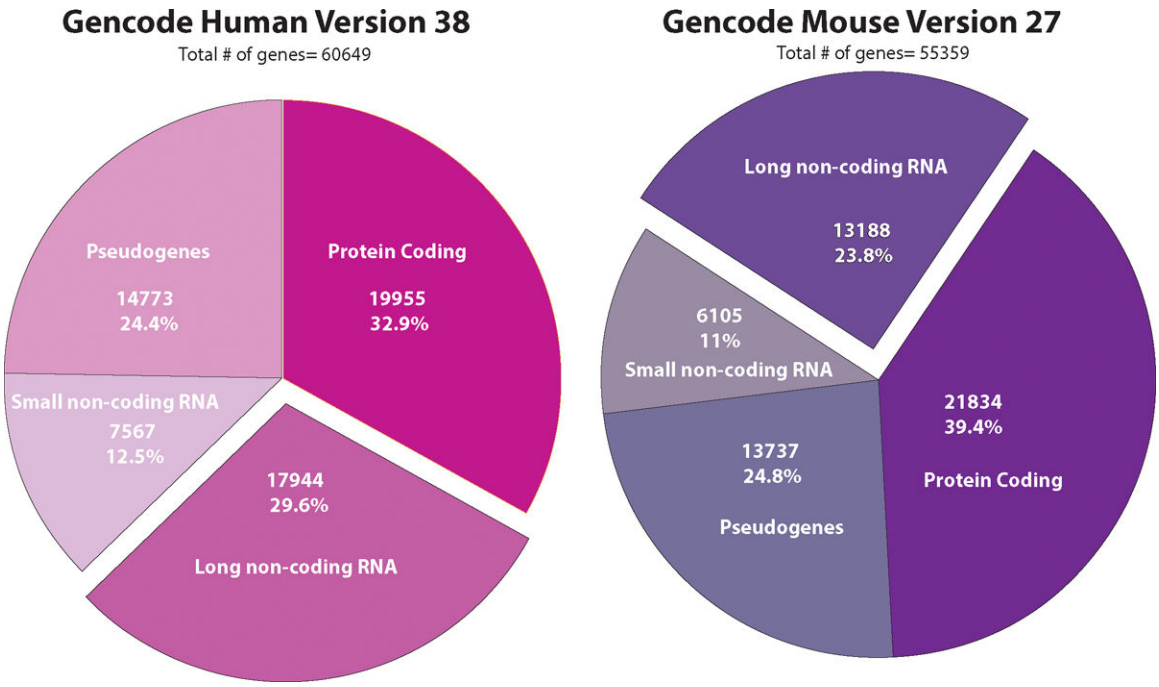

Figure 2. Annotated transcripts in Gencode based on the most recent human and mouse release.
Pie Charts showing statistics of annotated genes in human GRCh38.p13 (version 38) and mouse GRCm39 (release M27).

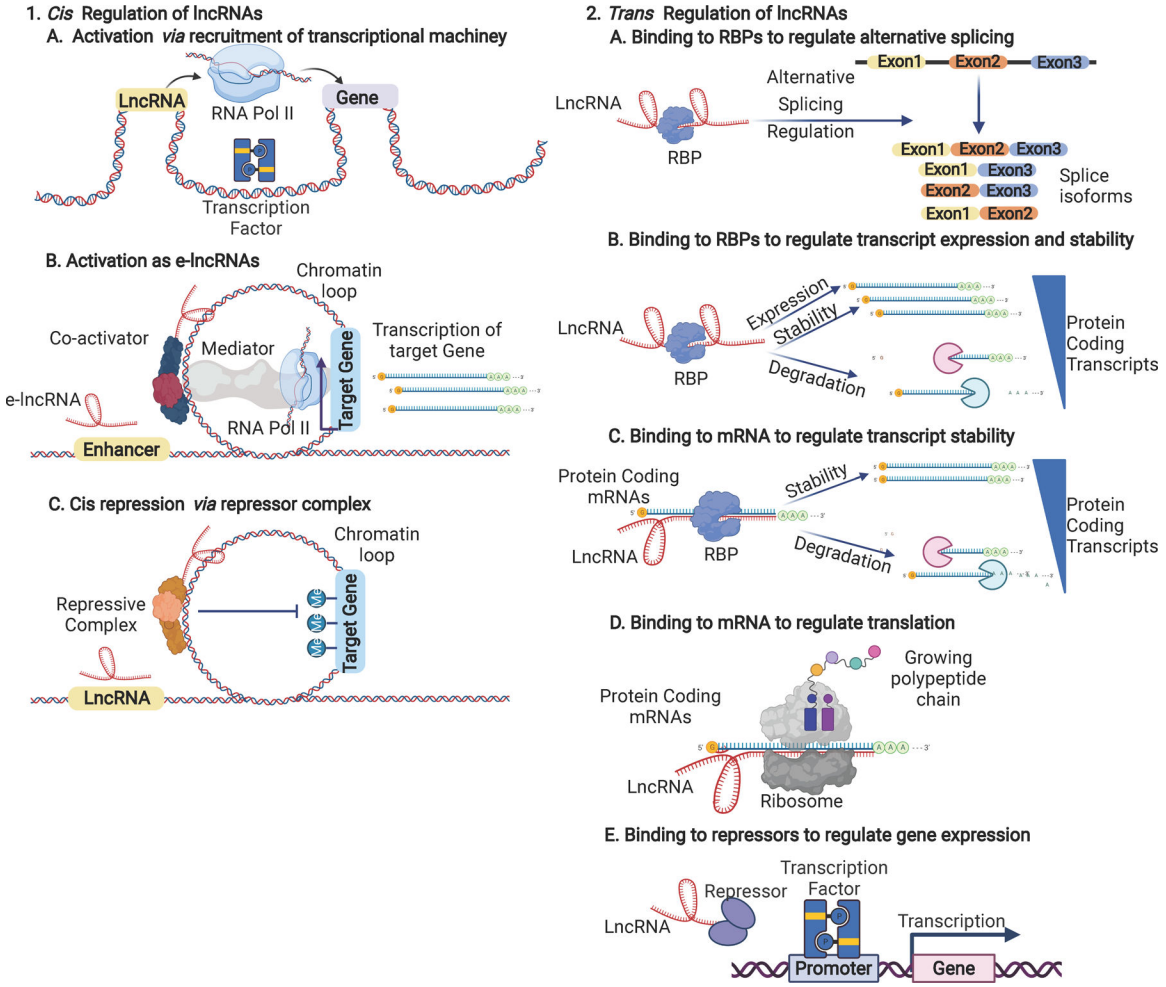

Figure 3. Modes of gene regulation by lncRNAs.

1.LncRNAs can regulate neighboring genes in *cis* through a variety of mechanisms. **A:** They can activate gene expression through recruitment of the transcriptional and splicing machinery to the neighboring locus. **B:** They can activate gene expression by acting as enhancer RNAs (e-lncRNAs) where they recruit mediators and co-activators to locus and facilitate coordinate activation through chromatin looping between the enhancer and promoter of the neighboring gene. **C:** They can repress neighboring gene expression by recruiting a repressive complex to the neighboring gene promoter through chromatin looping, leading to methyl tag deposition and gene expression inhibition. 2. LncRNAs can regulate genes on a different allele or different chromosome in *trans* through a variety of mechanisms. **A.** Through interactions with RNA binding proteins (RBPs) to regulate alternative splicing of the RBP target transcript. **B.** Through interactions with RBPs to regulate target transcript expression, stability or degradation. **C.** It can occur through binding of a lncRNA directly to mRNA transcripts through base pairing impacting transcript stability and degradation. **D.** Binding of a lncRNA directly to mRNA transcripts through base pairing impacting recruitment of the polysome and transcript translation. **E.** LncRNAs can sequester

suppressors from the gene promoter to allow transcription factor binding and subsequent gene expression.

Table 1.

Developments of evolving sequencing technologies with variable throughput, accuracy and cost.

Platform	Throughput	Accuracy	Number of Reads per Run	Maximum Output	Read length	Run Time	Price
Sanger	Low	99.99% ‡	1	--	400–600bp [12]	4 hrs	$3/reaction †
Ilumina MiSeq †	Low	>96% §	25 million	15Gb	1x36bp 2x300bp [12]	4–55hrs	$24/Mb [12]
MiniSeq †	Low	>80% high quality bases §	25 million	7.5Gb	1x75bp 2x150bp [12]	4–24hrs	$24/Mb [12]
Illumina HiSeq 4000 RapidRun §	High	>88% bases with high quality scores §	300 million	250GB 750GB 1500GB	1x50bp 2x75bp 2x150bp	1–3.5 days	$1400–2000/ Lane †
NovaSeq 6000 †	High	>94% bases with high quality scores §	650 – 800M 2 – 2.5B	400 Gb 750 Gb	2 x 250bp 2 x 150bp	38hr 44hr	$8,578/Flow cell $5,590/ Lane
NextSeq 550 †	High	>80% bases with high quality scores §	400 million	120Gb	1x75bp 2x150bp	12–30hr	20$/Gb †
PacBio Soquel II ᵀᵀ (HiFi) [26,27]	High	>99.9%	300K – 12M per SMRT cell	30–50Gb per SMRT cell ᵀᵀ	~ 10–25Kb	10 hr per SMRT cell	>2000$/flow cell
ONT Minion (DNA) ¶, [28]	High	85–97%[29,30]	4–10 million	20–50Gb	Up to 4.2Mb	Up to 72hr	425–900$/ flow cell
ONT Promethion 24, 48 ¶ (DNA)	High	85–97%	4–10 million	Up to 245Gb	Up to 4.2Mb	Up to 72hr	Up to 2000$/ flow cell

† Data from Genewiz website

‡ Data from Thermofisher website

§ Data from Illumina Website

¶ Data from Nanopore website

ᵀᵀ PacBio website

Table 2.

lncRNAs involved in inflammatory gene expression regulation.

LncRNA	Function	Mode of regulation	Localization	Source
Rroid	Promotes ILC proliferation	*Cis* through direct interaction with neighboring gene promoter	Nuclear	42
Morrbid	Represses *Bcl2l11* expression	*Cis* Recruits PCR2 which deposits methyl tags on *Bcl2l11* promoter	Nuclear	43
LincRNA-Cox2	- Required for *ptgs2* expression - Regulates expression of critical immune response genes	- Enhancer RNA mechanism - Inhibits expression of ISGs through interactions with hnRNPA/B and A2/B1 in trans - trans activation via unknown mechanism	Nuclear and Cytoplasmic	67
Lnc13	Regulates inflammation and is dysregulated in Celiac disease	*Trans* regulation of immune genes through interactions with hnRNPD	Nuclear	74
Lethe	Negative feedback of NF-κB pathway	*Trans* via direct binding to Rela	Nuclear	75
THRIL	Induction of TNFA expression	*Trans* via direct binding to HNRNPL	Nuclear	69
linRNA-EPS	Suppresses immune gene expression	*Trans*; Binds to chromatin and Hnrnpl to form complexes	Nuclear	70
Lnc-DC	Dendritic cell differentiation	*Trans* via direct binding to STAT3	Cytoplasm	76
GAPLINC	Inhibits basal activation of NF-kB	Functions in *trans* to regulate to regulate basal levels of NF-kB and limit its localization to the cytoplasm	Cytoplasm	77
PACER	Promotes *Ptgs/Cox2* expression	Cis; sequesters p50 away from *Ptgs2* promoter allowing for its expression	Nuclear	51
NEAT1	Promotes *IL8* expression	Trans; sequesters the SFPQ repressor complex to paraspeckles	Nuclear	52
IL-1B eRNA	Promotes *IL1B* expression	Binds to PU.1 transcription factor and *IL1B* promoter to activate transcription	Nuclear	71,72