

REVIEW

Opportunities for RNA sequencing in physiology: from big data to understanding homeostasis and heterogeneity

 Jeremy W. Prokop,^{1,2} Stephanie M. Bilinovich,^{2,3} Ember Tokarski,⁴ Sangeetha Vishweswaraiah,^{1,5} Sophie VanderWeele,^{2,6} Akansha S. Das,² Surya B. Chhetri,⁷ Alexander Dao,² Sanjana Arora,¹ Austin Goodyke,¹ Katie L. Buelow,¹ Mason Westgate,¹ Elizabeth A. VanSickle,¹ Claudia J. Edell,⁸ Lance N. Benson,⁸ Daniel B. Campbell,² Caleb P. Bupp,^{2,9} Amanda Holsworth,^{2,10} Nicholas L. Hartog,^{2,10} Jena M. Krueger,^{2,11} Marcos Cordoba,^{2,12} Matthew Sims,¹³ Maximiliano A. Tamae Kakazu,^{2,14} Angela M. Peraino,¹⁴ Stewart F. Graham,^{1,5} Tim Triche Jr.,⁴ Elora Hussain,¹⁵ Mara L. Leimanis-Laurens,^{2,15} Connie M. Krawczyk,¹⁶ Jennifer S. Pollock,⁸ and Surender Rajasekaran^{1,2,15}

¹Corewell Health Research Institute, Corewell Health, Grand Rapids, Michigan, United States; ²College of Human Medicine, Michigan State University, Grand Rapids, Michigan, United States; ³Department of Chemistry, Grand Valley State University, Allendale, Michigan, United States; ⁴Center for Epigenetics, Van Andel Institute, Grand Rapids, Michigan, United States; ⁵Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, Michigan, United States; ⁶Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States; ⁷Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, United States; ⁸Section of Cardio-Renal Physiology and Medicine, Division of Nephrology, Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States; ⁹Medical Genetics, Corewell Health, Grand Rapids, Michigan, United States; ¹⁰Allergy and Immunology, Corewell Health, Grand Rapids, Michigan, United States; ¹¹Department of Neurology, Helen DeVos Children's Hospital, Corewell Health, Grand Rapids, Michigan, United States; ¹²Department of Obstetrics, Gynecology and Women's Health, Corewell Health, Grand Rapids, Michigan, United States; ¹³Section of Infectious Diseases and International Medicine, Department of Internal Medicine, Beaumont Royal Oak, Royal Oak, Michigan, United States; ¹⁴Division of Pulmonary and Critical Care, Corewell Health, Grand Rapids, Michigan, United States; ¹⁵Pediatric Intensive Care Unit, Helen DeVos Children's Hospital, Corewell Health, Grand Rapids, Michigan, United States; and ¹⁶Department of Metabolism and Nutritional Programming, Van Andel Institute, Grand Rapids, Michigan, United States

Abstract

The quantity of physiological data has grown exponentially, yielding insights into mechanisms of phenotypic and disease pathways. Among the powerful tools for physiological omics is the study of RNA, where broad sequencing of RNA leads to hypothesis generation and testing while providing observational discovery. Emphasis has been placed on RNA molecules that code for proteins, even though they represent a minority of total RNA. Diverse sequencing methods have rapidly expanded the identification of non-protein-coding molecules, including nonsense-mediated decay and long non-coding RNAs (lncRNA), which now represent the most diverse class of RNA. Increasing attention needs to be paid to the data processing of RNA sequencing to interpret transcript-level mapping data in the context of protein biology, as many protein-coding genes have diverse noncoding transcripts. Over the past several years, single-cell and spatial transcriptomics have yielded unprecedented insights into cellular, tissue, and organ physiology. Building on these advancements, bulk RNA sequencing tools have begun producing robust deconvolution methods that enhance the analysis of human genes, the detection of foreign RNA from bacteria and viruses, and provide deep insights into complex immunological events, such as B- and T-cell recombination. Over a million RNA-sequencing datasets have been generated, providing resources for data scientists to reprocess data and expand larger databases. From model organisms to complex human diseases, RNA sequencing resources continue to transform our knowledge of the complexity of personalized disease insights. Observational science is at the core of physiology, and growth of RNA sequencing represents a significant tool for physiologists.

heterogeneous response; homeostasis; physiological omics; single-cell RNA sequencing; spatial transcriptomics



Correspondence: J. W. Prokop (jprokop54@gmail.com).
Submitted 13 October 2025 / Revised 13 October 2025 / Accepted 24 October 2025



INTRODUCTION

The Greek word *physiologos* translates to “the study of nature,” suggesting that a physiologist is an observer who interprets. Physiology, the science of how living things function, was born from observational sciences. Hippocrates’ studies of the body’s humors that form the foundations of homeostasis and the search for cause and effect led to an appreciation for the complex systems within living species.

Stemming from the growth of anatomical knowledge and investigations into diseases, insights elevated through observation of biological systems, giving rise to testable hypotheses and model organism research that became experimental physiology (1–3). The observations were powered by advanced tools of recording instruments (blood pressure, heart rate, electrophysiology, and lung capacity) that enabled measurements to be processed for statistical power (4). Similar to the early days of physiology, expanding tools such as DNA and RNA sequencing, along with other omics, enabled a new age of observational science that significantly enhanced our ability to form unbiased hypotheses. With the launch of the journal *Physiological Genomics* in 1999 by the American Physiological Society, a new era of omic-centric physiology was born (5).

Of the RNA within a cell, 98% of transcriptional output is noncoding and represents many undefined physiological mechanisms in cell regulation, structural biology, RNA interference, cosuppression, transgene silencing, imprinting, and methylation (6). Although the human genome codes for 19,433 protein-coding genes, there are 507,365 uniquely spliced RNA sequences from 78,691 unique genes known as of 2025 and 278,326 transcripts from 78,275 genes in mice (7), representing a significant volume of underexplored biology. These numbers do not account for the unique splicing that can result in individuals due to genetic variants, the complex genome rearrangements within the acquired immune system, and the millions of transcripts from species found within the biomass of humans (8). Although multiple genes have been extensively established for involvement within physiology, many of the human and mouse transcripts remain understudied, with the majority of transcripts with one or fewer publications addressing their biology, especially those RNA molecules that do not code for a protein (9).

Two hallmark concepts taught in physiology are homeostasis and heterogeneity (10–12). The emerging tools of RNA sequencing continue to advance our knowledge of both concepts, where the tools enable broad mapping while giving fine details to develop testable hypotheses at the cellular and spatial levels (8). The emergence of precision medicine, which can pair phenotypes with genomics, is potentiated by RNA sequencing and other omics to address clinical insights, which has expanded our knowledge of homeostasis and heterogeneity.

RNA homeostasis coordinates volume, microRNAs, cell fate, and senescence at the cellular level through gene regulation and RNA turnover (13–15). These studies are being increasingly powered by advancing tools of single-cell and spatial RNA sequencing (16–18). Imbalance of RNA homeostasis has been suggested in states such as cancer (19), inflammatory (20), stress (21), aging (22), and cardiovascular

disease (23, 24), to name a few. A central aspect of physiology is the integration of many phenotypes into simple logistical mechanisms, where RNA sequencing elevates our insights and connects diverse phenotypes through homeostatic circuits.

Human genome diversity has gained an appreciation for modifying disease pathology, where genome sequencing tools have enabled these studies (25). Within areas such as cystic fibrosis, the same genetic variant can manifest with highly diverse phenotypes, where factors from genetic modifiers to infection status can modulate nearly every disease factor (26). This has given rise to fields such as response genetics, where scientists can study complex genetics by environmental factor interactions (27–29). These environmental factors are not addressed by genome sequencing alone but require additional omic metrics. In this review article, we will highlight how RNA sequencing can embrace both homeostasis and heterogeneity.

As the cost of these sequencing tools has decreased, it has opened the door to the idea that simple hypothesis testing with RNA sequencing outperforms single RNA molecule tracking with real-time PCR. This has enabled hypothesis testing to be paired with broader, unbiased observational sciences that can reframe scientific discovery if proper analysis is given. As the number of tools for RNA sequencing has grown, it has unfortunately also created challenges in identifying the right tool for the right project, as some of the tools have higher expenses that limit their broader scaling, yet they provide cellular or spatial resolution of knowledge. As most sequencing machines have a high price tag, this has led to the concentration on sequencing centers and cores, facilitating higher volumes and lower costs. Although this has been beneficial in bringing these tools into more laboratories with a lower entry point, it has also decreased the number of experiments where the sequencing chemistry itself is variable, thereby producing less variability in types of discoveries. Labs often rely on one method over others for the simplicity of working with cores, reducing the options for optimizing the right tools for the right questions.

Within this review, we discuss the current state of tools for RNA sequencing, examples of usages within physiology, and the areas that are likely to continue powering physiologists. The article aims to provide a primer to scientists and trainees new to the field while guiding future work into rapidly emerging areas. The goal is to provide a community resource to match the RNA sequencing tools to projects and assist the community in advancing the field of physiology.

THE HUMAN GENOME TRANSITION TO TRANSCRIPT DISCOVERY

The pilot human genome, completed in 2003–2004, enabled the new era of sequencing-based discoveries, with 99% of the genome mapped, enabling searching for RNA coding genes (30). The genome was narrowed to approximately 20,000–25,000 protein-coding genes based on bioinformatic filtering of open reading frames. The sequencing of genomes was powered by next-generation sequencing, DNA fragmentation, and DNA sequencing by polymerase fluorescent reactions imaged on high-resolution cameras (31, 32).

With the advancements of the next-generation platforms, researchers paired cDNA synthesis from RNA as input into the sequencing to map the coded RNA from specimens (33). Taking all RNA and making cDNA results in most sequencing for ribosomal RNA (rRNA), given that it represents approximately 80% of RNA volume within human cells. Researchers primarily focused on oligo-dT-based capture to remove the rRNA and capture the polyadenylation (polyA)-enriched protein-coding messenger RNA (mRNA) class (34). The ENCYclopedia Of DNA Elements (ENCODE) (35) and GENCODE initiatives (36), launched shortly after the draft genome completion, organized the mapping of RNA molecules and built gene regulation insights. In 2005–2006, the initial Gencode release started to develop maps of genes (37). By 2012, the GENCODE release mapped 20,687 protein-coding and 9,640 long non-coding RNA (lncRNA) loci (36).

The early RNA sequencing studies were focused on which DNA elements were transcribed into RNA (37), the mapping of transcriptional start sites within the genome based on 5' untranslated region (UTR) extensions of RNA (38), how those RNAs get spliced (39), and how RNA molecules can be fused within cancer (40). Further techniques to remove the rRNA through degradation (ribo-reduced) enabled the cDNA conversion of a broader class of RNA molecules (41) while also capturing bacterial RNA that is not polyA-modified (42). By 2021, the GENCODE reference mapped protein-coding genes shrunk to 19,954, while the lncRNA scaled to 17,957 loci (43).

The original next-generation sequencing technologies were limited in understanding splicing complexity due to the short-read chemistry, where fragments are sequenced for 50–300 bases at a time. Thus, connecting splicing sites more than a few hundred bases apart was challenging to perform through bioinformatic technologies. Advancing technologies, such as single molecular PacBio and protein pore-based nanopore, were able to generate long reads of RNA that vastly enhanced our knowledge of splicing and non-coding RNA (44, 45). By 2025, the GENCODE map showed stability of protein-coding genes (19,433) while there has been a massive amplification of known non-coding RNA (lncRNA 35,899) (46).

As the transcript and gene maps were refined, additional methods to capture RNA were developed. These included capturing RNA within specific compartments such as the nucleus (47–49) or bound to ribosomes in active translation (50–52). Cells could be sorted through various platforms, and the RNA barcoded to open the door for single-cell RNA sequencing (53–55). Histology sections could be placed on barcoded surfaces, imaged, and sequenced to develop spatial transcriptomics (56–58). More recently, technological advances in nanopore chemistries have removed the need for cDNA conversion, with the possibility of directly sequencing RNA molecules that include hundreds of chemical RNA-base modifications (59, 60). Often overlooked for the advancement of RNA sequencing are the incredible commercial advancements in RNA stabilization reagents, novel collection methods, and RNA purification systems that have powered the collections (61). As the past two decades have highlighted, the technological advancements for RNA sequencing have refined our knowledge of the transcript landscape and enabled powerful tools for physiological characterizations.

THE HUMAN AND MOUSE TRANSCRIPTOMES

The 49th GENCODE version of the human and 38th mouse transcriptomes were released in September of 2025 (Fig. 1) (46). The human release contained 78,691 total genes with 507,365 known spliced transcripts. The mouse release contained 78,275 total genes with 278,326 known spliced transcripts. This release showed the most considerable growth of non-coding RNA transcripts of any previous GENCODE version, representing technological advances in targeted capturing and long-read sequencing platforms. The lower number of transcripts within the mouse likely reflects less utilization of long-read platforms to generate broad tissue data, which have recently been applied to human samples. Based on quasi-alignment filtering (62), we present the transcripts from diverse biotypes for genes that are highly concordant in human and mouse (Fig. 1A), whereas transcript biology shows more divergence between the species (Fig. 1B). It is worth noting that many transcripts can be classified into multiple biotype categories. We have utilized the Gencode single annotation events for simplicity. In the text below, we do not focus on the well-reviewed protein-coding transcripts, which others have extensively reviewed (8, 63–65).

Long Noncoding RNA

The largest group of known transcripts comes from lncRNA, representing 45.6%/46.1% (human/mouse) of all genes and 37.7%/56.0% of transcripts. The noncoding RNA molecules are the most diverse of any biotype and represent the largest biomass within cells, where ribosomal RNA accounts for around 80% of the total biomass. A few lncRNA molecules, such as XIST, have been shown to have pivotal roles in processes such as X-chromosome inactivation (XCI; 66). In many RNA sequencing experiments of physiological processes and diseases, lncRNA molecules are differentially expressed (67). Human and model organism biology, including obesity (68), cancer (69), homeostasis (70), and cardiovascular (71), have shown the involvement of lncRNA in phenotypic variability. The advent of CRISPR-based technologies is quickly identifying lncRNA functional mechanisms, such as the recent work linking the lncRNA *KILLR* in breast cancer and the regulation of DNA replication/repair processes (72). The physiological processes for lncRNA continue to grow to include gene regulation, structural biology, posttranscriptional control, splicing, and posttranslational modifications (73–75).

Deciphering the molecular function of lncRNA has been challenging due to the lack of molecular biology knowledge and species specificity, to produce limited bioinformatic tools for processing the thousands of lncRNA gene sequences for function (76–78). Many genome or phenome-wide association (GWAS/PheWAS) studies have fine-mapped genetic variants found within lncRNA. Yet, these variants have not been functionally characterized (79, 80). lncRNA likely represents a wide array of mechanisms that have yet to be discovered (81), suggesting an important area for future research within physiology.

Retained Intron RNA

Retained intron transcripts represent around 6.7% of all known transcripts. These transcripts are often from protein-

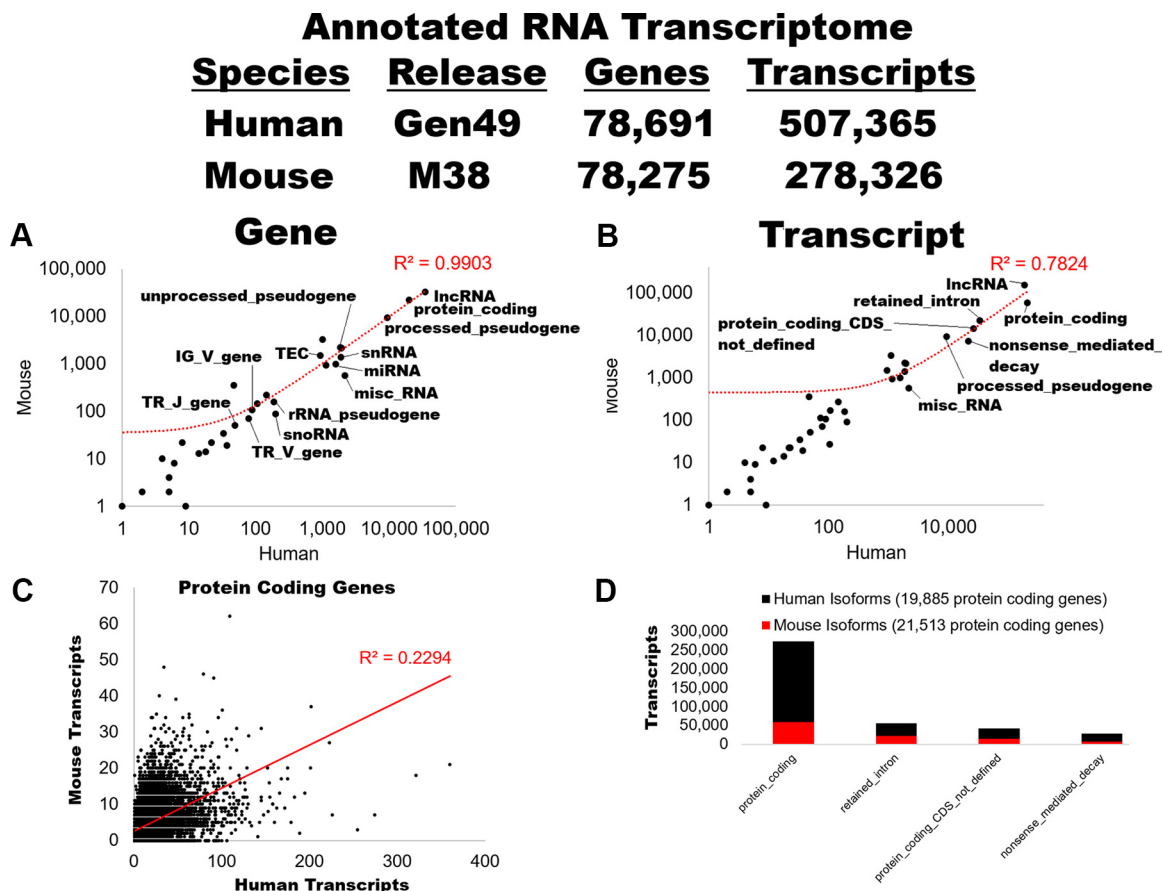


Figure 1. Current knowledge of human and mouse RNA biotypes. **A** and **B**: the number of each biotype in the reference Gencode transcripts for human (Gen49, x-axis) and mouse (M38, y-axis) for genes (**A**) and transcripts (**B**). **C**: number of transcripts per protein coding gene for humans (x-axis) and mice (y-axis). **D**: the number of transcripts of diverse biotypes for genes with human and mice protein-coding transcripts.

coding genes that get processed and withhold an intron or part of an intron into the final mature product, resulting in the loss of the open reading frame sequence of the protein, which often results in controlled degradation (82). However, alterations in the degradation pathways of retained intron transcripts can elevate their presence in cells (83). These transcripts can be challenging to detect reliably in sequencing and bioinformatics (84, 85). The biological roles of retained intron transcripts have been growing, where they are critical for neuron differentiation (86), Alzheimer's disease (87), aging (88), hypoxia (89), activation of CD4 T cells (90), and cardiovascular diseases (91).

Nonsense-Mediated Decay RNA

Nonsense-mediated decay (NMD) transcripts represent around 3%–5% of all known transcripts and have an early stop codon before the last splicing site of the transcript (92, 93). This results in the accumulation of proteins bound to the RNA that are not removed by ribosomes in translation, resulting in a controlled degradation of the RNA (94–96). NMD transcripts are well-studied in rare diseases. In many cases of splice or nonsense variants resulting in early stop codons, the NMD degradation results in haploinsufficiency of the RNA, where transcript abundance can be compared with the normal allele, resulting in an allelic imbalance in expression (97, 98). In some cases of rare disease splice or

nonsense variants, NMD does not degrade the early termination transcript, resulting in a partial protein that can produce autosomal dominant negative phenotypes (99, 100). An example is the recent discovery of a *RABL3* early stop codon that avoids NMD regulation for familial pancreatic ductal adenocarcinoma through a 36 amino acid peptide-based disruption of KRAS biology (101).

Various biological processes, such as development, cancer, infections, stress response, T-cell receptor selection, and aging, are known to alter the NMD degradation processes (102–106). When NMD is inhibited, the protein products of NMD transcripts can result in partial proteins with either de novo function relative to the full-length protein-coding genes or in dominant negative forms of biological competition for the full-length proteins (107). The nature of environmental regulation of NMD produces a rapid sensing switch within cells, where the NMD transcripts are continually made, but their accumulation through NMD inhibition results in rapid cellular changes. Endoplasmic reticulum (ER) stress is a well-studied environmental control mechanism (108). Factors such as elevated protein-folding-derived ER stress have been shown to activate quality control mechanisms that include the suppression of NMD degradation and altered transcripts of genes such as *UPF1* (109, 110).

The balance of preventing deleterious early stop codons from impacting physiology in rare diseases and the environmental

sensing mechanisms makes NMD transcripts important for future research (111). Advancing treatments, including exon skipping and splice site-altering oligonucleotides, hold incredible promise for NMD-based genetic treatments (112).

Alternative Transcripts and Splicing

The human and mouse transcriptomes have multiple ways to splice a gene together through the differential inclusion of exons in the mature RNA. In the 2025 Gencode release for biotypes of RNA, lncRNA, and protein coding have the most transcripts per gene in mice and humans (Fig. 1B). This is followed by retained-intron and nonsense-mediated decay RNA transcripts. Breakdown of each protein-coding gene shows low correlation between the number of transcripts annotated in the human and mouse, suggesting that splicing biology knowledge or biological roles differ between the two species (Fig. 1C). Short-read sequencing technologies powered the detection of different transcripts at the exon-exon junction detections, limited by the distance of multiple junction sites to know the actual final RNA sequences (39, 113, 114). Genes such as *KRAS* (115, 116), *RAC1* (117), *FGFR2* (118), and *Titin* (119) were well documented for how changes to the protein from splicing altered physiological processes.

The advent of long-read sequencing paired with high depth of sequence coverage has enabled more extensive full transcript maps of gene splicing (120, 121). Integration of both tools with large-scale expression databases enables more systematic maps of splicing, such as the small GTPase family of proteins, including *KRAS* and *RAC1*, that have been shown to have a broad protein-altering landscape with phenotypic functions (122). There is an urgent need for basic protein characterizations of splice variants (123). Knowledge of the alternative splicing, including the presence of cryptic splice sites less often used (124, 125), has given rise to the use of antisense oligonucleotides to control splicing around genetic variants in numerous genes such as *SCN1A* (126, 127). The FDA has authorized oligonucleotides that can alter the splicing of genes such as *SMN2* to counter the *SMN1* genetic variants in spinal muscular atrophy (SMA) (128).

The knowledge of the complexity of the spliceosome has significantly increased in the past few years, yielding insights into the many genetic and environmental factors that can regulate the alternative splicing of transcripts (129–133). In many cases, the alternative splicing results in unproductive transcripts regulated by mechanisms such as NMD (134). For example, mutant p53 can result in alternative splicing of multiple pancreatic ductal adenocarcinoma oncogenes (135). Broad splicing changes have been identified within mouse models of ischemic reperfusion injury (91).

There is a complex interaction of splicing factors with polyadenylation machinery as a polyA signal is encountered by the RNA polymerase II (136, 137). A delicate balance of splicing factors' with polyadenylation factors' binding (U1 snRNP, SRSF3, SRSF7, CPEB4, CELF2, FUS, HNRNPC, PABPN1), along with the elongation rates and secondary structure of transcripts, can influence the detection of the polyA signal sites within RNA polymerase II transcription to influence the dynamics of exon inclusion and where polyA tail addition ends transcription (137–140). For many transcripts, variable

sites in the 3'UTR can be used for the generation of polyA addition and transcript elongation termination, resulting in the same protein but with changes of the 3'UTR regulation of transcripts, including well-known roles in RNA half-life (141). Physiological examples of this include cold exposure dynamics in mitochondrial proton uncoupling in brown fat (142), cell cycle progression regulated by polo (143), and axon-specific targeting of mRNAs in neurons (144), or in many steps of the circadian clocks (145). In other cases, the inclusion of an early exon with a polyA signal site detected by the polyA machinery can result in alternative transcripts, protein isoforms, or even nonprotein coding outcomes such as rapidly degraded transcripts of non-stop decay that can often be impacted in physiological conditions such as cancer (138, 146).

Splicing is not the only biological process that can contribute to differential exon inclusions. Mechanisms for altering either end of a transcript, such as alternative transcriptional start sites and 3'UTR readthrough, can change the processed RNA. The alternative ends of transcripts drive a high diversity of RNA across human tissues (147). The deep sequencing of the 5'UTR as part of GENCODE, ENCODE, and FANTOM consortiums resulted in the identification of alternative first exons of genes, suggesting that within many genes of the genome, there were more than one loci that transcription can start (38, 148–150).

One of the most well-characterized alternative transcriptional start sites in physiology is the *SHROOM3* gene (151). A significant GWAS loci for chronic kidney disease falls near *SHROOM3*, which was annotated for years as intronic within the *SHROOM3* gene (152). However, *SHROOM3* utilizes a secondary transcriptional start site within the kidney podocytes that loops with the upstream GWAS loci (151). CRISPR-based mutation of the CKD risk variant within cell lines directly altered gene expression levels of the secondary start site, and the shorter protein isoform was shown to recover kidney function in morpholino-based knockdown with human mRNA recovery in zebrafish (151). This example is a valuable lesson to carefully annotate noncoding variants relative to RNA transcripts to account for the complete insights into transcript diversity.

Variability in the 3' end of RNA molecules contains occasional read-through, where the normal 3'UTR gets spliced to an exon that retains an open reading frame. Leptin receptor (*LEPR*) shows the evolutionary selection of usage for RNA read-through, where in vertebrates, the *LEPR* inserted 3' the endospain gene (*LEPROT*), utilizing the *LEPROT* broadly expressed promoter to drive ubiquitous expression profile of *LEPR*. One of the most biologically active RNA read-through examples is that of the insulin (*INS*)-*IGF2* fusion transcript, where the read-through of *INS* allows the splicing to join with *IGF2* (153). The *IGF2* 5'UTR uses an alternative open reading frame, resulting in a protein broadly expressed in the pancreas and liver, with human variants further disrupting the stop codon location (154). Variants through GWAS connect this *INS-IGF2* transcript to antibody production against Insulin that is highly associated with type 1 diabetes (155).

Care in Processing Protein-Coding RNA Transcripts

Many studies performing bioinformatics of RNA sequencing sum transcript level mapping into gene annotations,

adding all transcripts and protein isoforms into a single gene event. This summation makes gene ontology easier but ignores the biological complexity of transcripts. For genes coding for multiple protein sequences based on alternative exon splicing, a gene expression summation calculation loses the capacity for understanding the transcript-level biology. Further complicating this summation is that of the current protein-coding genes of humans and mice, many contain additional non-protein-coding transcripts, such as retained introns and NMD (Fig. 1D). As those transcripts map overlapping segments with the protein-coding isoforms, they can significantly impact the interpretation of protein levels when summing transcripts. We recommend analyzing RNA sequencing data at the transcript level, with biotype-level summation if necessary, per gene to account for these confounding factors.

SAMPLE ISOLATION AND SEQUENCING PLATFORM COMPARISONS

Over the last decade, the rapid growth of sequencing and the availability of reference genomes have led to the expansion of sample isolation and sequencing methods for analyzing RNA molecules. This has resulted in a set of tools for physiologists, each with benefits relative to the others for sample isolation, RNA preparation, and sequencing of RNA (Fig. 2). Below is a description of each approach/platform, its benefits, and the current challenges, based on the state of knowledge as of 2025.

Sample Isolation

Bulk extractions.

Biomarker discoveries and knowledge advancements in RNA-seq have been driven by the bulk extraction of RNA from whole tissues. Any whole organ or biofluid (blood, saliva, sputum, buccal swab, urine) collected can be lysed

and RNA isolated. As RNA isolation became simplified through commercial kits, the speed and efficiency of RNA isolation decreased the likelihood of changes in transcriptomic signatures that can occur in more laborious and time-consuming isolation methods. Moreover, many of the components in between cells that contain RNA (bacteria, exosomes, or cell-free RNA) are also captured in bulk extractions, as presented later in USAGE OF RNA SEQUENCING. Bulk extractions blend the many cells that make up tissues and biofluids, creating uncertainty about where signals are emerging and decreasing the ability to map low-cell-number biological contributions. As our knowledge of diverser cell types has expanded, it has created computational approaches for cell type deconvolution of bulk analyses (156, 157). Several computational techniques have been described for dissecting cellular content directly from expression profiles of mixture samples and deconvolving RNA admixture into their component cell types without physical dissociation, antibodies, or living material (158). However, these methods still suffer from signal-to-noise dynamics, uncertainty of the ground truth of cell-to-gene levels, and are often difficult to interpret in light of the limited response biology known at the cell level (159).

Sorted cell.

Although the bulk analysis has given broad insights into physiology, the development of single-cell resolution of RNA has given more power to the diverse cell types within specimens. Beginning from the mechanical separation of cells into wells using techniques such as surface capture or flow sorting into high-throughput plates, followed by cDNA preparation, the techniques of single-cell RNA sequencing have now massively parallelized the isolation and preparation of thousands of cells within each sample (55, 160). Early separation of cells from tissues would require extensive time to dissociate material, isolate different cell types, and then perform RNA isolation and sequencing. This yielded early

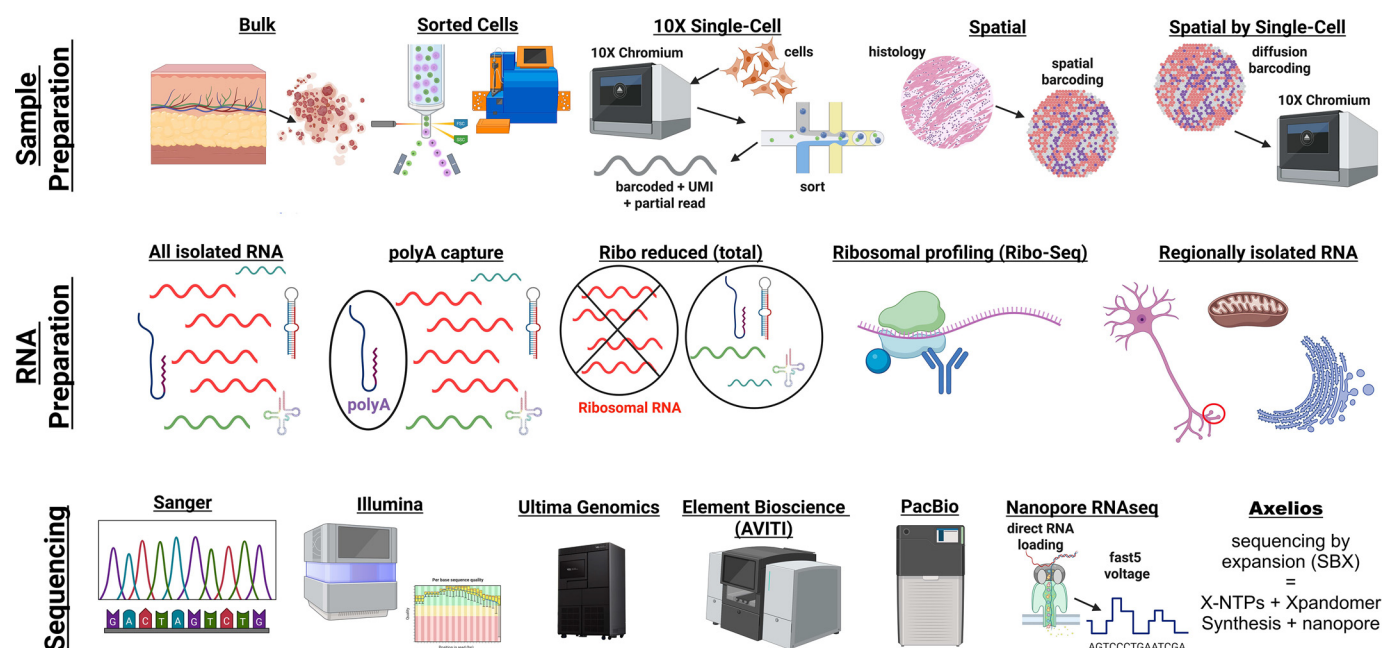


Figure 2. Current state of RNA sequencing techniques. Figure created with a licensed version of BioRender.com.

and powerful insights into contributions of different cell types, but was limited to already existing knowledge of what cell groups were, as there was a need to isolate each cell type with known markers. Also, the bulk assessment of any cell group removed much of the heterogeneous dynamics, such as cell cycle and feedback loops, being hampered by many of the same challenges found within bulk sample isolations.

Platforms such as the 10× Genomics chromium controller utilize microfluidics to separate cells or nuclei with labeled beads to apply barcodes and unique molecular identifiers (UMIs) to each cell's RNA, followed by sequencing and deconvolution. As the technique has taken off in applications, the bioinformatic pipelines have simplified the complex analysis, elevating the resources of limited sequencing depth and cell identifications (54, 161). The density of cells sampled enables single experiment landscapes of cell dynamics that facilitate the complex discovery of heterogeneous and homeostasis biology; yet, this complexity has been challenging to understand with the volume of data (162–164).

Single-cell RNA sequencing has revolutionized the identification of cellular landscapes, from cell cycle dynamics to subclusters of functional cells (53). Cardiovascular systems (165), blood cells (166), developmental biology (167), angiogenesis (168), respiratory (169), cancer (170), to nearly every aspect of physiology (171) single-cell RNA sequencing is expanding our knowledge. Compared with other platforms, single-cell RNA sequencing allows exploration of cell heterogeneity at a single-cell resolution, fate mapping, and lineage tracing without genetic labeling, identifying new or rare cell populations based on transcriptomic signatures, and tracking transcriptomic changes in individual cells during disease progression, treatment, and recovery (172). The tools of single-cell are especially powerful for the study of diverse model organisms and environmental factors (173). One of the most expansive usages of single-cell technologies is within the immune system, opening the door for exploring the diverse cell types of innate and acquired activation while observing the combination of chains within the immune repertoire (174–176).

In oncology, single-cell sequencing has led to a better understanding of complex tumor cell behavior, evolution of the cancer, and the identification of novel biomarkers that can be used to develop precision therapies. One significant benefit of single-cell RNA sequencing is improved analysis of multifaceted tumor microenvironments and inferences of cell-cell communication that can assist in creating immunotherapy strategies, particularly overcoming immune resistance (177). More recently, single-cell RNA sequencing has been integrated with other sequencing methods such as ATAC-seq for gene regulation (178, 179) to electrophysiology (180) and patch clamping (181) techniques to decipher cellular physiology of gene expression.

Even though single-cell RNA sequencing holds incredible promise, it has several limitations. Most notable is the cost. Although the cost of sequencing each cell type has been reduced to acceptable levels, the platform is still significantly more expensive than bulk sequencing due to the number of cells needing analysis (172). Cells must be handled longer than bulk approaches, increasing the chances that cellular states are altered. Dissociating and sorting cells takes time, which can modify the expression profiles of cells, yielding

some nonbiological outcomes. In addition, many cells are impacted by flow-sorting or contain multiple nuclei, such as neutrophils, skeletal muscle, and cardiomyocytes, making it difficult to interpret the full spectrum of cellular/nuclear heterogeneity. Newer strategies are emerging that enable non-fluidic barcoding, lowering the cost per cell.

One of the major challenges with single-cell sequencing is the depth of sequencing needed to quantify expression differences between individual cells. Most laboratories utilize depths of approximately 50,000 reads per cell, elevating to millions of reads when profiling thousands of cells in an experiment (50,000 reads/cell × 10,000 cells = 500 million reads), which is sufficient to identify most cell types with clustering analysis (182). If the focus of a project is rare cell types or differential expression analysis, more robust expression may be warranted (~100,000 reads/cell). If the focus is on transcript splicing dynamics, even deeper sequencing is suggested (~500,000 to 1 million reads per cell), quickly elevating the cost to staggering levels relative to bulk RNA workflows. When addressing genes expressed <20 transcripts per million (TPM) in bulk analysis (20 TPM = 1 transcript for every 50,000 reads sequenced), it is often suggested that deeper sequencing is required, but rarely used throughout the field (183). When low sequencing depth is used for each cell, a value of zero has uncertain results and can lead to biased interpretations when reads do show up. Although imputation methods can resolve the technical versus biological zero low-abundance issues, they can also present biases in discovery with overinterpretation or bias of the zero values (184).

Another limitation is that single-cell RNA sequencing only provides static RNA expression profiles, providing limited temporal information and requiring repeat assays to get profiles at different points, such as disease progression and pre- and posttreatment. Pseudotime analysis, a computational tool that ranks potential dynamic processes in cells based on the heterogeneity of transcriptional expression levels, is commonly used in single-cell RNA sequence analysis to create a trajectory mimicking how cells change over time. Continued advancements in algorithms, computational tools, and multiomics can reduce the limitations (185). Finally, the nature of 5' or 3' short-read sequencing, often used in single-cell does not resolve full transcript dynamics, making many of the above discussions of biotype and transcript discoveries challenging (186). The pairing of single-cell sequencing with long-read chemistries promises to remove some transcript limitations but further elevates the cost per cell for data, given the required number of reads per cell to accurately measure transcript differences (187, 188). Overall, single-cell RNA sequencing advances our knowledge of cellular physiology and will continue growing to overcome many limitations.

Spatial sample processing.

Although single-cell RNA sequencing advances cellular discovery, spatial RNA sequencing promises to refine our tissue physiology and disease knowledge (189, 190). The complex anatomy of cells within any biological tissue has long been appreciated through histology, and dating back to 1987, groups have advanced the mapping of genes into spatial discoveries (191).

The rapidly emerging technologies to transition histology into RNA sequencing have powered spatial transcriptomics growth. Most notably, utilizing spatial sequencing with specialized computational tools such as reference single-cell RNA sequence atlases, cell2location, cytoSPACE, ST deconvolve, and Boyers space enhances cellular resolution, transcript coverage, and alignment of independent samples (189). For example, surfaces can be labeled with barcodes, and the histology sections laid down on that surface to facilitate spatial barcoding that can be deconvoluted post-sequencing. The density of these barcodes has elevated to reach a sub-single-cell level map of expression. These advancing technologies have yielded unparalleled insights into physiology's spatial heterogeneity and homeostasis (192). They have also validated the identity of new cell types discovered by single-cell sequencing, expanding cell-cell or neighborhood relationships within functional tissue units, and generated cellular and molecular atlases of healthy organs through the work of groups such as Human Biomolecular Atlas Program and Human Cell Atlas (193). The application to heart (194), kidneys (193, 195), skeletal muscle (196), endocrine organs (197), skin (198), brain (199, 200), and lung (201), has grown basic knowledge of healthy tissue with more recent applications to disease pathologies.

Similar to all techniques, spatial transcriptomics has multiple limitations, most notably the cost per sample processed. The limited sequencing depth of spatial platforms can also make the depth of cell heterogeneity difficult. Pairing single-cell RNA sequencing enables a higher sequencing depth for cells matched to histological locations of spatial transcriptomics. This shows that technique integrations hold the greatest depth of knowledge advancement but with significant cost limitations (189). One of the significant limitations of spatial transcriptomics is the need for cells to fit into histological slides, where many tissues create challenges. Similarly, artificial intelligence digital pathology tools such as FUSION cannot perform the same interpretation as manual work by pathologists in providing histological correlation to spatial transcriptomics and cannot provide single-cell resolution. As a result, the time and cost prohibit the process of having pathologists manually annotate spatial transcriptomics data for accurate results (193). The three-dimensional nature of tissues also elevates the cost of spatial transcriptomics with the need for many slice analyses to recreate the maps of whole tissues, limiting the work to well-funded consortium initiatives.

Emerging tools are combining spatial mapping with single-cell technologies. Tools such as Curio Trekker use spatial barcodes that are released from the surface to diffuse into and label the nucleus that is closest (202). When the nuclei are dissociated, it enables single-cell sequencing that can be taken back to the spatial barcode locations.

Most spatial transcriptomics has been performed on mRNA with limited accounting for the noncoding biotypes of protein-coding genes (58). Similar to single-cell sequencing, the technological pairing of spatial with long-read sequencing holds the promise to move to transcript-level discoveries (194). Still, it also significantly elevates the cost of an already expensive technique. Spatial transcriptomics, although still limited in its infancy, most likely holds the greatest potential of all RNA sequencing platforms for the

field of anatomy and physiology and will likely bring critical advancements to the characterization of many underexplored genes of the human genome. As new spatial platforms such as the Aviti24 Teton cytoprofile technology emerge, we will also see the fusion of multiomics, where proteins can be detected using barcode-labeled antibody binding sequencing linked to RNA sequencing within the same spatial samples.

RNA Preparation

Bulk RNA.

Sample preparation is not the only part of the isolation process that alters the discovery potential in RNA sequencing. The type of RNA molecules isolated also has significant impact on how data are interpreted. The isolation of all the RNA within the cell, followed by sequencing, results in the majority of data to be that of rRNA. The resulting data from bulk RNA isolation produces only a fraction of usable insights and requires deeper sequencing to find valuable signals. Thus, there is a need to isolate RNA types or to remove the rRNA to identify biologically meaningful information.

PolyAdenylation capture.

Given that eukaryotic protein-coding RNA is polyadenylated, it is possible to capture these specific RNA molecules using the polyA signal to remove the rRNA. This can be accomplished through the use of oligo dT sequences to initiate cDNA amplification, through the use of labeled beads enabling mRNA capture, or in some cases the engagement into protein pores in platforms such as nanopore (203) (where bulk RNA can be used as starting material). It should be noted that prokaryotic species RNA lacks polyA tails and thus cannot be captured with this method. The capture of polyA mRNA also removes most of the other noncoding RNA biotypes in addition to rRNA. Although our knowledge of these other RNA biotypes has been slower to discovery than mRNA, as already addressed, their increasingly recognized importance has shifted the RNA sequencing field to pivot more toward other isolation methods, even as polyA isolations are a far cheaper method of removing rRNA.

Ribosomal reduction.

To remove the rRNA, which accounts for 80%–90% of cellular RNA, techniques have emerged to deplete rRNA through selective hybridization (204). The remaining mRNA and the diverse non-coding RNA can then be sequenced on any sequencing chemistry. These ribosomal (ribo) reduction strategies are of high use in prokaryotic species that lack polyA mRNA signalling. Although ribo reduced depletion methods produce robust insights for biotypes of RNA, the isolation cost significantly more than polyA isolations. Newer generation, often species-specific, rRNA enzymatic degradation methods are beginning to reduce the costs of ribo reduction (205, 206).

Other profiling techniques.

Several other RNA isolation methods outside of poly(A) capture and ribo reduction have also been developed. Isolation of RNA from organelle-specific locations, such as mitochondria, endoplasmic reticulum, or the nucleus, has emerged (207, 208). Ribo-Seq offers a capture of the active engagement

of RNA with the ribosome for translation (50). The capture of ribosome-engaged mRNA also enables kinetic insights into the dynamics of translation (209, 210), with the techniques now moving into single-cell resolution (211). Neurons have the ability to transport mRNA far from the nucleus to individual synapse regions (212). Neurites have one of the most complex localization systems, where protein localization is primarily driven by the mRNA localization, insights made possible through unique neurite RNA isolation methods (213). Interactions of the RNA 3'UTR elements with proteins such as UNK drive the RNA to the neurites (214). Together, these isolation techniques expand RNA-seq beyond abundance toward understanding localization, translation kinetics, and regulation within complex cellular microenvironments.

Sequencing Platforms

Sanger.

The sequencing of DNA has long been done through fragments of ~1,000 bases, starting with Sanger and automated Sanger platforms and advancing into next-generation platforms. As RNA sequencing increased, many genes were first discovered through the isolation, cloning, and Sanger sequencing of cDNA. Although Sanger sequencing can produce fragments of larger than 1,000 bases with high fidelity, the extreme cost per base relative to other platforms often pushes these platforms more into validation methods or low complexity assays with a few readouts (215).

Short-read sequencing.

The short-read platforms were primarily driven by Illumina-based sequencing, with newer companies such as Element Biosciences' AVITI platform and Ultima Genomics UG 100 elevating a new competitive sequencing market. The cost of these machines can represent a significant investment for laboratories, where many institutes and commercial vendors offer sequencing through cores and services to bring the technology to more laboratories. It should be noted that the cost per base of sequencing significantly decreases with the larger volume machines, where pooling strategies are implemented in large volume centers to reduce per-experiment costs.

The foundation of all short-read platforms is the fragmentation of nucleotides, cDNA in the case of RNA sequencing, followed by polymerase addition of labeled nucleotides and some form of imaging or size separation. The sequencers of short-read platforms can take upwards of a day to return any data during sequencing. Short-read sequencing can be performed as either single-end or paired-end. In single-end, one side of the fragmented cDNA is sequenced, whereas in paired-end, both are sequenced. Although single-end sequencing is cheaper and takes half the time to complete, paired-end sequencing is preferred by most laboratories, given the enhanced nature of the reads to confirm splice sites.

Long-read sequencing.

The use of long-read sequencing platforms has revolutionized the detection of gene-level transcripts. The two commonly used long-read sequencing platforms are PacBio and nanopore. These long-read techniques have an increased cost per sample relative to short-read platforms, with further increase in cost when paired with single-cell or spatial RNA

sequencing workflows. Newer improvements, such as using unique molecular identifiers (UMIs), have increased the accuracy of these long-read platforms by generating consensus around the multiple reads of each UMI, but these approaches increase cost (216). Long-read sequencing has been especially valuable for under-characterized or novel model organism genomics and transcriptomics, where it can generate reference maps with less bioinformatic needs (217, 218). This is particularly important in the rapidly evolving field of metagenomics (219, 220).

PacBio, which sequences cDNA using real-time polymerase tracking known as single-molecule real-time sequencing, has a more considerable capital cost for the sequencer but provides higher quality/accuracy sequencing outputs. The circularization of cDNA constructs enables read corrections through repeated sequencing of each fragment, which can be used for additional single-cell preparations (221). Nanopore, which can sequence cDNA or direct RNA through protein pores to track subtle voltage changes, offers the lowest capital sequencing costs (<\$1,000) but has lower sequencing accuracy. Both of these long-read platforms offer full-length transcript sequencing that facilitates the analysis of all transcript splice sites per read, enhancing the identification of splicing complements greater than the few-hundred-base limit of short-read technologies. Similar to short-read technologies, both platforms require additional methods to give cellular or spatial insights. The platforms are also more challenging to handle batching, with the number of reads generated per run far below the short-read high-density sequencing platforms, decreasing the number of samples barcoded into a single sequencing run.

For the nanopore platform, with the laboratory addition of polyA tags onto the RNA and ribo-reduction, it is possible to sequence noncoding RNA. In addition, nanopore is currently the fastest platform for data generation, as the voltage analysis of each read can be rapidly converted into base calls, generating real-time sequencing insights that accumulate over time. In addition, nanopore offers the ability to directly run RNA molecules through the protein pores, thereby reducing the time to data generation without requiring cDNA conversion (222). The direct RNA bases can also be processed for the role of RNA base modifications (223, 224), similar to DNA methylation, but where RNA is known to have dozens more modification types than DNA (225–227). Nanopore's rapid and direct data generation makes the platform ideal for future scaling into clinical specimen analysis, generating much excitement for applications in physiological genomics. For example, nanopore offers return of that data that could be used when the need for data is exigent, such as in critically ill and presurgical patients.

Although much of the sequencing market has been dominated by Illumina short-read, these long-read technologies continue to mature. Newer possibilities, such as sequencing by expansion (SBX) chemistry of Roche (228), are completely redesigning the strategies of sequencing. SBX utilizes novel expandable nucleotides (X-NTP) bases that can be added to synthesis using Xpandomer methods, creating bulkier bases that are rapidly and accurately sequenced at ultra-low cost through nanopore structures. As we enter the new era of a competitive sequencing market, there is considerable hope that competition will drive innovation and cost

reductions for RNA sequencing, thereby bringing many of these tools to a wider range of laboratories.

RNA SEQUENCING RESOURCES

Publicly Available RNA Sequencing Data

Millions of samples have been processed as the above RNA sequencing platforms have been utilized for the past two decades. Many of these datasets are available within public databases, enabling scientists to reprocess data as tools and resources grow. For example, as discussed in THE HUMAN GENOME TRANSITION TO TRANSCRIPT DISCOVERY, as the Gencode knowledge has increased over time, reprocessing older data can provide insights into transcripts not annotated within older Gencode releases. The National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) contains the largest deposit of raw sequencing data. As of late 2024, the species with the most datasets available is SARS-CoV-2 (Fig. 3A), an RNA virus sequenced

extensively during the 2020 pandemic (229). Human RNA sequencing datasets are approaching 2 million available runs, with mice also above a million datasets. This is followed by a rapid decrease in data for other model organisms, such as plants, flies, rats, and fish (Fig. 3A).

Most sequencing has been on the Illumina platform for the 12 million runs of RNA sequencing available, with a rapid decrease to PacBio and nanopore-based long-read datasets (Fig. 3B). It should be noted that human and mouse long-read datasets are far below other organisms, such as SARS-CoV-2, for long-read sequences. For humans, Illumina datasets available for public download as the FASTQ file and paired-end, there are over one million sequencing runs from 27,397 projects (known as BioProjects within NCBI) at the end of 2024 (Fig. 3C). These projects contain a broad range of sample and sequencing depths. The sequencing depth of any experiment represents the number of reads generated per sample, where more reads increase the confidence of data mapped to a reference as coverage of unique splice sites increases, especially for RNA molecules with lower expression levels. If analyzing

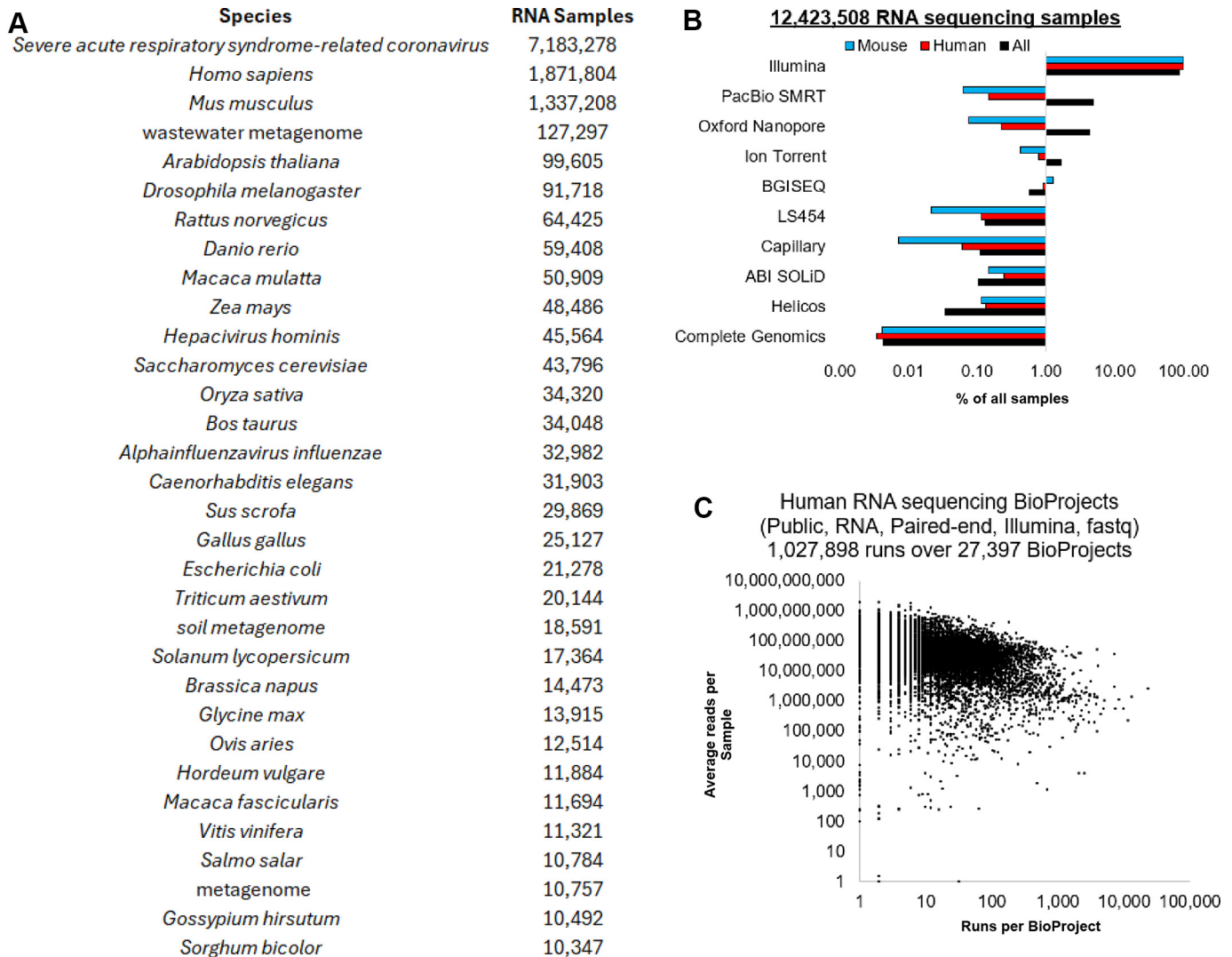


Figure 3. Data for RNA sequencing available in the NCBI SRA. A: number of RNA sequencing runs per species. B: the chemistry types of RNA sequencing for NCBI SRA runs. C: the number of runs within a BioProject (x-axis) and the average depth of reads (y-axis) per sample of RNA sequencing datasets. All data was pulled on November 11, 2024.

gene transcripts, splicing, the immune repertoire, or environmental factors (bacteria or viruses), samples should be sequenced to a higher depth. When interested in quantifying human mRNA, many groups will euthanize the sequencing depth to elevate samples and stretch the cost of sequencing.

Many large consortia deposit vast resources of raw RNA sequencing datasets such as ENCODE (BioProject PRJNA30709), the noncoding RNA atlas (PRJNA576920), and the Precision Medicine Research consortium (PRJEB58625). In some cases, the RNA sequencing data is protected under the database of Genotypes and Phenotypes (dbGaP), where Institutional Review Board (IRB) and safety plans are required before data can be obtained to protect human specimens from reidentification using the raw reads.

The depositing of data for single-cell datasets is unstructured in the current NCBI SRA, with some groups deconvoluting data before submission, resulting in a large number of runs with low depth of coverage, such as BioProject PRJNA591860 and PRJNA638640. In other cases, runs with deep sequencing can often be single-cell datasets deposited without a breakdown of each cell's expression, such as BioProject PRJNA551745 and PRJNA788872. Groups working with any NCBI SRA RNA sequencing data are advised to read as much information as possible on the breakdown of runs and sequencing depth before working with the data. This is also why including detailed metadata when depositing samples into the NCBI SRA is so important.

In many data analysis projects with already generated RNA sequencing datasets, laboratories must account for batch effects and the experimental clustering of data, including experimental design, RNA extractions, and sequencing methods. Batch effects can be normalized with various computational methods, but none of these methods are perfect, and care should always be taken when comparing quantitative data across BioProjects (230–232). In data analysis from multiple BioProjects, selecting desired data quality, sequencing depth to reflect the type of molecules of interest, and a larger volume of samples to enhance discovery is normal.

A significant limitation of reprocessing NCBI SRA data is the need for large data downloads and extensive storage. One of the methods for rapidly identifying desired reads within an RNA sequencing data set is using the SRA BLAST tools (233), where the SRA database can be selected from BLAST search settings. This tool can have utility, for example, in searching unique splice sites from a few select SRA runs placed in the SRA accession. If someone wishes to process full alignments of SRA data, the SRA toolkit simplifies data downloads and parsing of paired-end reads.

To mitigate the issues of massive data downloads for RNA sequencing analysis, the NCBI has built an SRA in the cloud strategy that enables researchers to utilize Amazon or Google cloud approaches to move buckets of data to remote computing. These cloud strategies are suggested for optimal outcomes for projects where thousands of RNA sequencing datasets will be processed. As the utilization of RNA sequencing continues to scale throughout physiology, it is essential to collaborate in data processing to better optimize hypotheses and experimental designs in the context of the extensive RNA sequencing resources available.

Model Organisms

The laboratory mouse (*Mus musculus*) is the most studied model organism with RNA sequencing. By the end of 2024, the laboratory mouse had 1,349,165 runs within the NCBI SRA, 769,291 as paired-end, and 579,874 as single-end. Almost all these datasets are from Illumina short-read platforms, with around 2,000 samples for long-read technologies.

The mouse genome database contains a large amount of curated resources on RNA sequencing. These include curated and normalized RNA sequencing datasets, developmental biology insights, and differential expression profiles in experiments (234). One of the primary reasons the mouse has been so well explored for RNA sequencing is the many genetic models used for physiological characterizations, including consomic/congenic mapping, recombinant inbred, heterogenous stock, CRISPR editing, and conditional tissue-specific knockout methods. The GeneNetwork tools have curated many of the RNA sequencing resources of commonly used genetic models of laboratory mice, creating an easy-to-search tool for any gene symbol (235).

The zebrafish (*Danio rerio*), fly (*Drosophila melanogaster*), rat (*Rattus norvegicus*), and other organisms all have RNA sequencing datasets available and are curated in their respective genome databases (Table 1). Although these species have thousands of datasets, it should be noted that the transcriptome references have still lagged behind that of the human and mouse annotations based on the limited sampling. These transcriptomes often lack extensive noncoding RNA transcripts, and the protein-coding isoforms are often underdeveloped for diverse tissues and splicing complexity. In a recent study from our team, we demonstrated a 30%–40% difference in alignment for samples of the right or left ventricle relative to the leaflets of the valves within the hearts of sheep (242), highlighting the major limitations in diverse species RNA assessments without establishing de novo transcriptome annotations.

Table 1. Model organism databases with RNA sequencing data and resources

Model Organism	Database	Website	Citations
Mouse	MGI: Mouse Genome Informatics	https://informatics.jax.org/	(236)
Rat	RGD: Rat Genome Database	https://rgd.mcw.edu/	(237)
Zebrafish	ZFIN: Zebrafish Information Network	http://zfin.org/	(238)
Worm/Nematode	WormBase	https://wormbase.org/	(239)
Frog	Xenbase	https://www.xenbase.org/	(240)
Pig	Pig Genome Database	https://www.ensembl.org/Sus_scrofa	
Dog	Dog Genome Database	https://www.ensembl.org/Canis_lupus_familiaris	
Sheep	Sheep Genome Database	https://www.ensembl.org/Ovis_aries	
Cat	Cat Genome Database	https://www.ensembl.org/Felis_catus	
Cross-species	Alliance of Genome Resources	https://www.alliancegenome.org/	(241)

Many other model organisms have had their genomes sequenced and paired with RNA sequencing to annotate transcriptomes available within the NCBI genome resources. Translational research studies often leverage large animal models to improve translation into human disease insights, and while databases do exist for many of these large animal models, such as pig (*Sus scrofa*) and sheep (*Ovis aries*), annotation is much more limited, and breeding is much less controlled. Care should be taken when working with any model organism to address whether the reference resources included tissues selected for experiments. Otherwise, it is advised to consider assembling a de novo transcriptome using tools like Trinity (243) or performing long-read sequencing for better transcript maps. There is incredible potential for expanding RNA sequencing to diverse animal model physiology, as our team has shown in using RNA sequencing to uncover unique gene expression in the tricuspid valve due to pulmonary hypertension (244), discover the leptin sequence in birds (245) or to define genes on the rat sex chromosomes (246, 247).

Variant to Gene Tools

As millions of human RNA sequencing data points are available, integration with other physiological tools has expanded (Fig. 4). One of the most valuable resources has been linking genomic data to gene expression (variant to gene, Fig. 4). Over a million human genomes have been sequenced, informing population structure and rare variants (248–250). The gnomAD database, as of version 3 in 2024, contains 730,947 exomes and 76,215 whole genome sequences. This population structure and frequency of variants have been linked to large-scale RNA sequencing analysis of 54 non-diseased tissues from approximately 1,000 organ donors to build the genotype-tissue expression (GTEx) consortium (251, 252). Genetic variants have been associated with expression differences within each tissue, known as expression quantitative trait loci (eQTLs) or splicing quantitative trait loci (sQTLs) differences. The vast resources of GTEx can be studied through an online browser for genes or variants, providing integrated statistical analysis of *P* values and fold change for any tissue that can be compared with population allele frequencies (variant to gene axes, Fig. 4). Multiple ongoing consortia are moving eQTL analysis into single-cell RNA sequencing data integrations.

The expansion of ENCODE datasets can further refine eQTL variants into biological mechanisms, particularly for gene regulation (253). Simplified analysis can be performed using genome annotation states such as ChromHMM (254) and simplified scoring tools such as CADD (255). Care must be taken when linking any quantitative trait loci to mechanisms of gene regulation, as the inheritance of multiple variants within linkage disequilibrium (LD) complicates trait analyses, often requiring narrowing multiple potential causal variants for eQTLs (151). Polygenic priority score is a gene prioritization method that leverages GWAS summary statistics and incorporates data from bulk and single-cell expression datasets, curated biological pathways, and predicted protein-protein interactions (256).

Gene to Protein Tools

Broad human RNA sequencing profiles can also be linked to protein and spatial biology by integrating expression with

proteomic techniques such as immunohistochemistry, immunofluorescence, and subcellular localization analyses (Gene to Protein, Fig. 4) (257). The vastest resource for linking these datasets is the Human Protein Atlas (HPA) (258, 259), consisting of thousands of highly curated and validated antibody studies to link genes to proteomic resources. These online tools have grown to include brain analyses (260), functional classes of immune cells (261), secreted proteins (262), single-cell datasets (263), The Cancer Genome Atlas biomarkers (264), and commonly used cell lines for tissue culture (265). These HPA resources are valuable tools for physiologists to link gene expression to thousands of proteins.

Gene to Cell Tools

As single-cell and spatial platforms have emerged as powerful cellular and tissue resources, they have also resulted in organized consortia to pool resources into valuable databases (Gene to Cell, Fig. 4). The most complex of human tissues, the brain, highlights the incredible growth of these consortia. The Allen Institute's investments in organizing brain data have given rise to the resources of the Allen Brain Map (266, 267). From complex circuits to connectivity, the brain resources have been merged with RNA sequencing resources to inform mouse and human developmental biology and disease genetic architecture. The larger consortia outside the Allen Brain Atlas have furthered these tools into a systematic map of the human brain connectome that was heavily influenced by RNA sequencing discoveries (268, 269).

The Chan Zuckerberg Initiative (CZI) has heavily invested in community data integrations for single-cell and spatial profiles of nearly every human organ (Gene to Cell, Fig. 4). Their CellxGene tools comprise over 100 million cells from >1,700 datasets for differential expression, spatial maps, cell developmental trajectories, and annotated cell expression (270). The integrated analysis of these datasets has elucidated nearly a thousand functional cell types based on the expression of unique genes that can be integrated into any existing or future single-cell RNA sequencing experiment to facilitate cell annotations.

Gene to Trait Tools

The most crucial aspect of any physiologist's research program is the exploration of phenotypes and disease traits, where RNA sequencing can be integrated with the scaled exploration (Gene to Trait, Fig. 4). The gap in genotype to phenotype is the major obstacle in advancing RNA signals as biomarkers in clinical decision-making. There have been attempts, mainly in oncology, to advance phenotype to genes through RNA profiles to identify genes linked to cellular transformation (271).

The first knockout mouse model altered the immune-related gene $\beta 2$ -microglobulin to drive a deficiency in CD8 T cells (272). Since then, thousands of laboratories have knocked out genes in species including mice, rats, and zebrafish, followed by phenotyping. With higher throughput approaches to mouse knockout studies, thousands of genes can be linked to high-density programs such as the International Mouse Phenotyping Consortium (IMPC). With over 100,000 phenotypic statistical tests from nearly 10,000 gene knockouts, the IMPC also provides knockout replacement strategies to

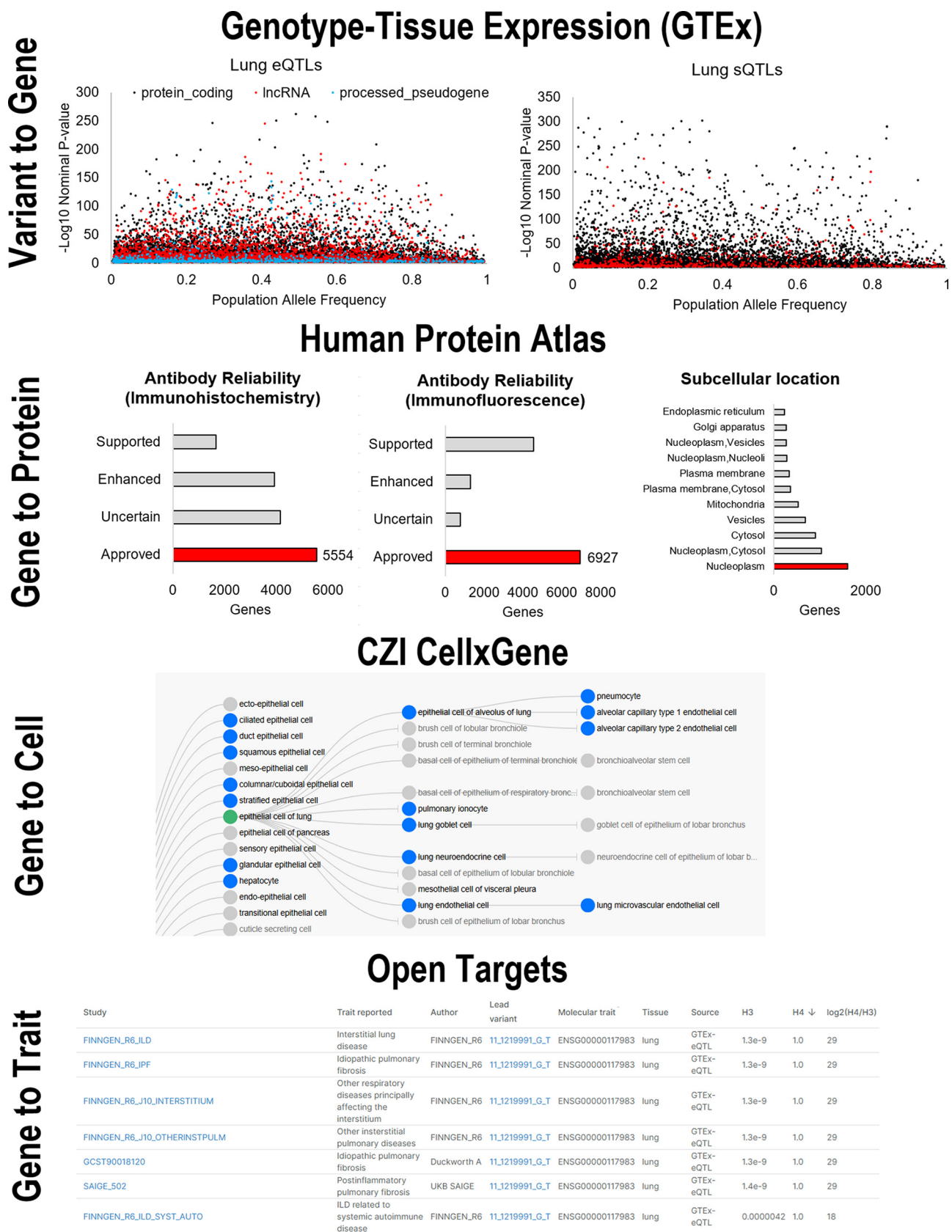


Figure 4. Databases of RNA sequencing to physiological mechanisms.

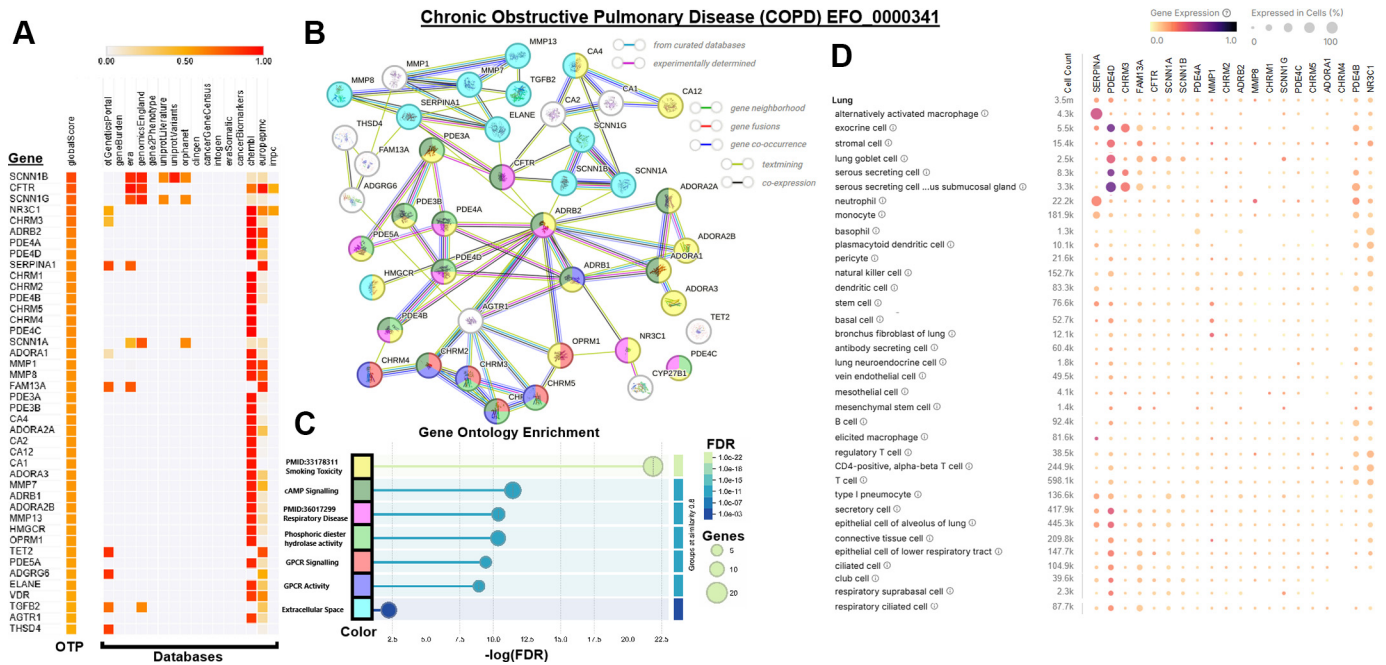


Figure 5. Gene to Trait analysis for COPD. A: heatmap of Open Targets Platform (OTP) scores >0.5 extracted for chronic obstructive pulmonary disease (EFO_0000341, the unique id for this medical phenotype term). B: STRING-based network analysis of A genes for human knowledge of protein interactions. C: Gene Ontology enrichment terms for B. D: CZI CellxGene single cell expression for the top 20 genes from A.

explore gene expression during development (273). The rat genome database also provides hundreds of gene knockout and genetic breeding schemes for phenotypic insights (274). Sometimes, the scientific approach neglects what the knocked-out gene does not affect rather than what it does, something that RNA sequencing could further elucidate if paired with approaches such as the IMPC.

Human genetic variants have traditionally been linked to physiological traits through either rare disease genetic mapping or Genome/Phenome-wide association studies (GWAS/PheWAS) (9, 275). RNAseq and genetic variant insights can be connected to various cancers through the COSMIC database (276). The GWAS catalog started the large-scale curation of thousands of studies for hundreds of biological traits (277). The UK biobank has furthered this work by colocating the variant to gene tools for each GWAS/PheWAS-associated loci (Open Targets Genetics, OTG) (278). These tools allow users to search phenotypic traits, which returns a list of loci that link eQTL/sQTL data from variant linkage disequilibrium (LD) blocks to genes. The data analysis of OTG further integrates multiple gene-centric databases with the GWAS/PheWAS data into a curated list of genes to disease (Open Targets Platform, OTP) (279). For example, it is possible to extract the highly associated genes for chronic obstructive pulmonary disease (COPD) from a dozen data structures into a simple tab-separated file from OTP (Fig. 5A). These genes can quickly be processed for pathway analysis using tools such as STRING (280) (Fig. 5B) and gene ontology enrichment (Fig. 5C) to establish physiological mechanisms for disease. Further dissection using the CZI CellxGene tools can resolve the functional cells within the lung tissue, each gene may work through (Fig. 5D). This workflow has available data for thousands of physiological traits and biological questions.

A recent study using an ancestry GWAS/PheWAS approach identified new putative causal variants, genes, and pathways, some targeted by existing drugs (281). As the All of Us data structure grows for nearly a million genomes linked to medical insights, further discoveries linking RNA sequencing to biological traits will be made. GWAS and single-cell RNA sequencing can substantially improve our understanding of disease heterogeneity and the identification of COPD (or other disease) endotypes and subpopulations (282).

Current Challenges within RNA Data

The massive expansion of RNA sequencing in physiology is not without continuing challenges. As already discussed, there is a need to define the physiology of noncoding RNA, where new techniques and the expansion of CRISPR-based technologies hold promise but require extensive new funding initiatives. More simple challenges within the data can also be seen. For example, existing RNA sequencing resources for large animal models more closely resembling human physiology. Models such as pigs, sheep, and cattle should be further explored for bulk tissue, single-cell, and spatial analysis through strategically invested research funds.

One of the significant challenges in physiology is the analysis of sex and gender as variables in phenotypes. Datasets such as GTEx (Fig. 6A) and the CZI atlas (Fig. 6, B and C) highlight the imbalance in samples collected between annotated sexes. These imbalances can make it very difficult to dissect molecular contributions to sex and gender, driving often over-simplified views of what sex and gender are (284). Within sex differences, one of the needed significant expansion areas for RNAseq is the understanding of expression for genes from the sex chromosomes, where the broad tissue

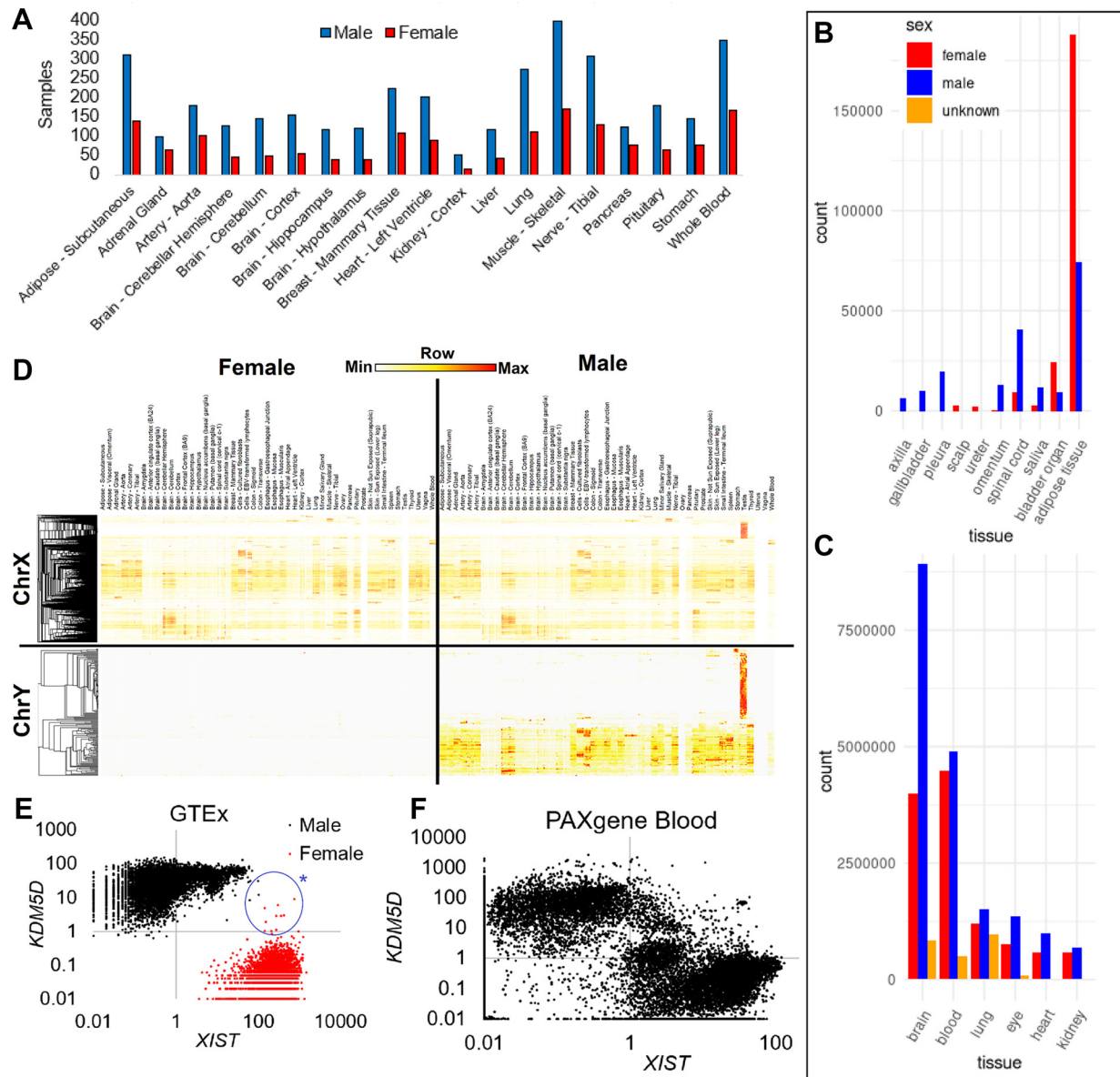


Figure 6. Sex biases of sample collections within large RNA sequencing databases. **A:** the number of samples within GTEx tissue RNA sequencing for male (blue) or female (red). **B:** ten human tissues in the CZI CellxGene Census have the largest imbalance by sex (excluding sex-specific organs). Five samples were from only one sex, and the remainder had a ratio greater than 2:1. **C:** tissues with more than 1 million cells in the CZI CellxGene census show a consistent bias toward male samples, particularly in brain tissue with additional complexity due to the presence of cells with no sex listed as notable in the lung. **D:** heat map of the ChrX or ChrY transcripts from GTEx in males and females. Each tissue is averaged over multiple samples. Each row shows a z-score with red being the highest tissue for that transcript. Dendrograms represent one minus Pearson's correlations. **E** and **F:** *KDM5D* of ChrY (y-axis) and *XIST* of ChrX (x-axis) expression levels from 17,382 GTEx samples over 54 tissues (**E**) and 16,243 blood PAXgene tube samples (**F**) processed from our previous work (283). CZI, Chan Zuckerberg Initiative; GTEx, Genotype-Tissue Expression.

atlas of GTEx shows that some genes have ubiquitous expression while others are enriched to specific tissues, such as the reproductive organs (Fig. 6D). RNA sequencing has been critical to understanding sex differences by establishing relationships between gene dosage (copy number) and transcript abundance of genes on the sex chromosomes (285, 286). Expression from the Y-chromosome (ChrY) genes has remained one of the most understudied aspects of sex differences (287), where GTEx data suggests many genes show tissue specificity (Fig. 6D, bottom right) yet lack physiological characterization.

X-chromosome inactivation (XCI) and the escape of XCI by various genes has developed at the single-cell level using variant tracking (288, 289), emerging to show the heterogeneous nature of variant influences on sex-linked phenotypes. More investments into these methods need to be explored. XCI dynamics are often assumed to be an efficient mechanism, as is the presence of ChrY genes. Yet, plotting the ubiquitously expressed ChrY gene *KDM5D* relative to the *XIST* gene for known sexes of samples from GTEx suggests high variability with the presence of ChrY genes and *XIST* elevation (Fig. 6E). Further processing of

16,243 blood PAXgene tube samples from our previous work (283) also suggests the same trend observed for *KDM5D* and *XIST* expression (Fig. 6F).

Sex, as defined by RNAseq, is not binary, as we show in Fig. 6. Sex chromosome genes can have different states that may reflect concepts of microchimeric sex chromosome genes and cells during pregnancy (290), genetic loss, or gain of sex chromosomes and their genes for intersex disorders (291), or the occasional XCI variability between tissues and individuals (292). RNA sequencing is critical to further refine the nature of XCI imbalance and dosage of sex chromosome genes that can cause female disease for primarily annotated male diseases due to a single ChrX inactivation throughout all cells (293). Gender has been even less explored based on gene expression, as has the understanding of hormone replacement therapies. Recent work has started to show the value of these explorations of RNA sequencing for gender (294). Using transcriptomics, the individual's sex can be confirmed retrospectively to determine if there are differences in the often self-reported gender and sex in publicly available sample sets.

The influence of rare genetic variants on RNA is another poorly explored area of physiology. Rare diseases have vastly enhanced our knowledge of phenotypic diversity, yet how these variants alter gene expression, splicing, and noncoding RNA remains underexplored. Rare disease patient-centric RNA sequencing could elucidate many genotype-phenotype associations but is challenged by the obtainable tissue. As patients with rare diseases are spread across the globe, the process for obtaining biospecimens, especially for pediatric patients, can vary widely. Some patients may not have access to clinical laboratories that can collect, process, and store samples needed for RNA sequencing. Growth of induced pluripotent stem cells (iPSCs) from these patients may enable broader development insights when paired with RNA sequencing, but these initiatives require more investments and biobanking initiatives. In addition, the complex treatment dynamics of gene therapies for these patients would be paired well with RNA sequencing explorations for immune alterations and outcome biomarkers (128).

RNA sequencing does provide several ethical challenges that need consideration. The extraction of variants within RNA sequencing data is associated with several issues. The analysis of variants from existing data needs to be performed carefully, as not all patients consented to variant insights and the risks that can be associated with them. From disease risk variants to population structure, many patients were not informed of the risks these analyses may entail. It remains uncertain how many variants within each RNA sequencing data set may potentially reidentify a sample, especially when combined with deposited metadata. As IRB offices and projects continue to attempt linkage of RNA sequencing to metadata of samples, it also continues to complicate the potential for reidentification of any specimen, where these risks need to be carefully considered in the study design and consenting process. Although science has advanced significantly by increasing physiological insights linked to each sample, patient protection must be accounted for in already generated data and future experiments. Given the more challenging nature of maintaining data under the protection of dbGaP or locally maintained repositories (such as REDCap),

this challenge should never be underestimated when moving into human RNA sequencing.

■ USAGE OF RNA SEQUENCING

The growth of RNA sequencing has shown promise in revolutionizing precision medicine and other applications in physiology. To conclude this review, we aim to highlight several of our first-hand experiences with new directions for RNA sequencing.

Multidimensional Deconvolution of RNA

Our team has been working to develop and implement precision transcriptomics. Building on our experiences in whole genome sequencing rare diseases, we began exploring what can be found in RNA sequencing human cohorts that might contribute to disease insights. Our first explorations were centered within critical care units of the hospital, especially the pediatric or neonatal intensive care unit (PICU/NICU). Our work utilized a blood collection system, PAXgene, that facilitated the easier clinical collection and stabilization of all RNA within the blood specimens. This was processed with paired-end total RNA sequencing (ribosomal and globin reduced) on the Illumina short-read platforms.

Three projects/cohorts were started as case versus control statistical designs but morphed into precision discoveries out of exploring the complex heterogeneous data of clinical samples. Two additional studies were then completed for more precision transcriptomic insights to make five cohorts of RNA sequencing (Fig. 7). Figure 7 shows the multiple areas of mapping knowledge that include diverse species, host gene responses, and the immune repertoire based on acquired immune response. Throughout these five studies of analysis, our team identified tools to elevate the mapping of RNA reads to enhance discovery, where we settled on a set of tools able to identify foreign RNA (295), human transcript alignments (62), and the complex immune repertoire (296). As Fig. 7 shows, these studies produced batch effects in their data mapping but provided statistical power within each cohort to make discoveries for individual samples.

Our first study was that of PICU patients with multiple organ dysfunction syndrome (MODS) paired with sedation controls matched to age and sex. The MODS phenotype, defined as two or more organs in failure, proved too heterogeneous for routine statistical mapping tools to resolve significance relative to controls (297). However, a subset of the MODS patients required extracorporeal membrane oxygenation, where statistics were powered to identify risk expression profiles associated with various histone genes and neutrophil extracellular trap (NETosis) production (298).

Our second study of patient samples focused on the NICU/PICU to address respiratory syncytial virus (RSV) relative to sedation controls in young children (299). Within the RNA sequencing data, we found that patients with RSV had significantly elevated immune system genes, including interferon and cytokine signaling. The expression of several sex chromosome genes was also significantly altered in RSV, suggesting reasons why males tend to have higher rates of hospitalization in children.

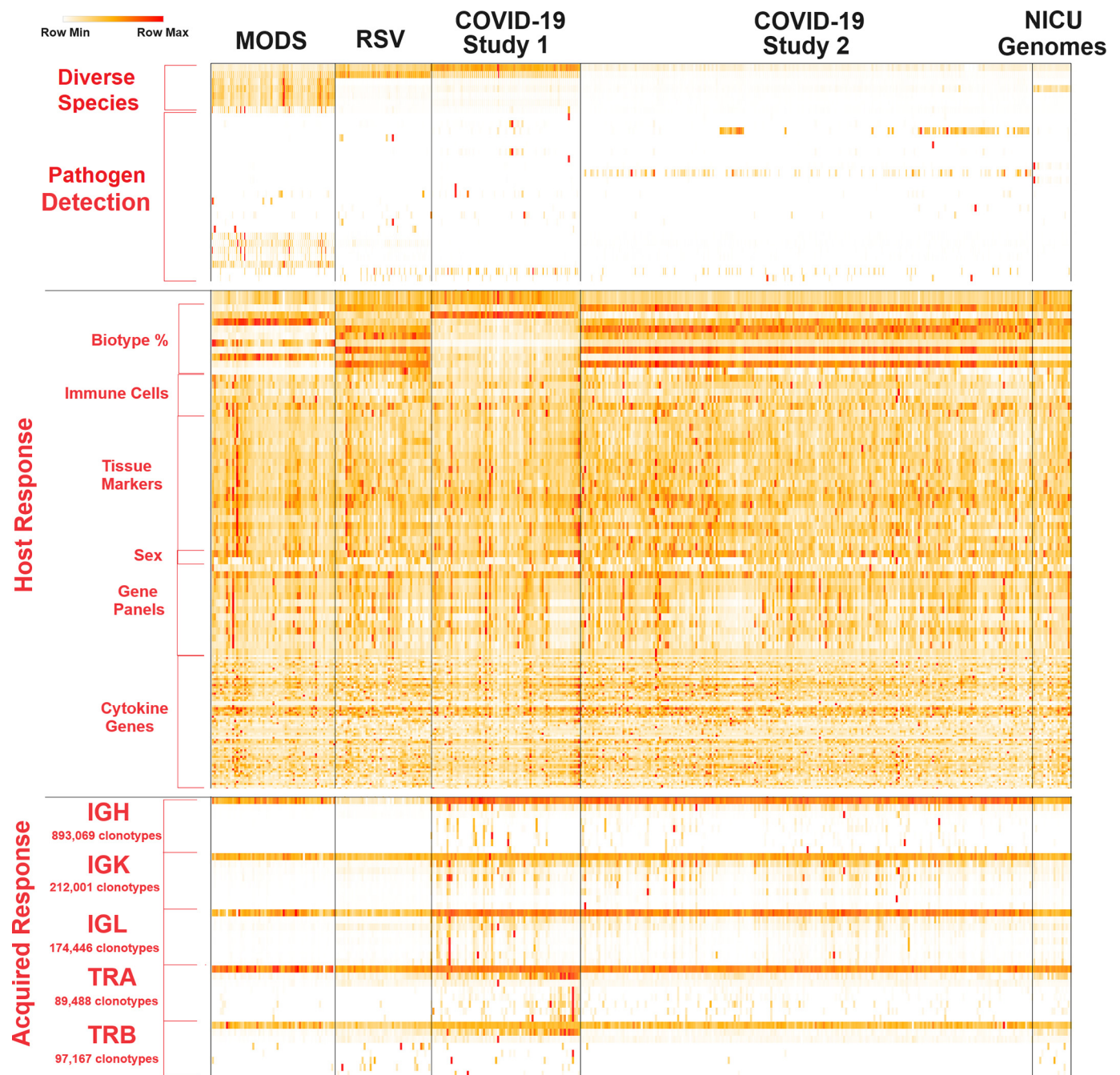


Figure 7. Corewell health internal patient RNA sequencing mapped for multidimensional insights from PAXgene tubes in human ribosomal reduced short-read, paired-end RNA sequencing. The heatmap is shown as a z-score colored per row, with the lowest in white and the highest in red.

The third blood study addressed adult hospitalizations early in COVID-19 relative to healthy adults without COVID-19 (300). The study included 11 cases that resulted in death, with 25 COVID-19 patients who survived. It was performed before any clinical trials were started on COVID-19 vaccines. The statistical analysis relative to controls elucidated increased gene signatures of neutrophil mechanisms, secretory granules, and NETosis with decreased signatures of lymphocytes. However, when mapping the data for each patient to those genes and pathways and performing unsupervised sample clustering, it was also realized that two subclusters within the patients

suggested different activations or different temporal states. This highlights one of our team's most valuable lessons in clinical RNA sequencing utility: traditional statistical models are often built on binary comparisons and outcomes. Disease mechanisms can be heterogeneous and require more than one answer. It is always suggested that mapping data from significance be studied within individual samples to explore their heterogeneity, for example, by performing unsupervised clustering. This suggests that even in COVID-19, understanding the immune response could hold potential for stratifying clinical trials based on the immune states (301).

Toward Precision, Patient-Level Transcriptomics

All three of the above studies showed heterogeneous individuals, where we realized the power of studying heterogeneous biology within RNA sequencing. As the heatmap of Fig. 7 shows, within many of our cohorts, there are unique patients with strong red bands in nearly every category of data mapping, suggesting that patients can have unique elevations. The real challenge is understanding what signatures are real biological insights relative to the background noise of the data. Instead of performing case versus control statistical testing, we refocused our statistical approaches to outlier mapping. We used Z-score outliers of human genes per patient, matching the genes to enriched gene ontology terms to identify patient-altered biological pathways. These pathways were then compared with the additional mapping data, such as non-human species and the antibody or T-cell receptors of the immune repertoire. As with rare disease whole genome sequencing (9), our team hypothesizes that linking biological heterogeneity through this type of precision transcriptomics may guide better future diagnoses and treatment plans for each patient.

We first showed the value of understanding infection mapping using RNA sequencing within the MODS cohort (297). Patients with MODS had three time points of blood collection to represent early care to the recovery period. RNA sequencing data analysis was blinded to any clinical data to avoid bias in the detection of signals. Our team showed several patients within the cohort with outlier detections of *Streptococcus* species, all later confirmed to have clinically detected strep infections. The RNA sequencing tools provided subspecies of insights, whereas the clinical data confirmed a broad Strep infection. Within one of the patients, we showed a later blood sample to have outlier levels of *Staphylococcus aureus*, where later in the clinical data correlations, we confirmed a hospital-acquired staph infection was present in that patient.

One of the unique mappings in a patient with MODS was read to *Morbillivirus*, a known cause of measles, which was present at two time points in that patient. This patient presented none of the symptoms of measles; thus, the signal was thought to be noise in RNA sequencing. Yet every other sample processed in the MODS and the other cohorts had zero reads aligned to *Morbillivirus*, suggesting this was not noise of the technique. From the RNA reads aligned to *Morbillivirus*, we further extracted the reads, reassembled them into a construct at the sequence level, and a BLAST search of the base level sequence identified high homology to the measles vaccine. In the clinical note follow-up due to the read detection, it was confirmed that the patient received a measles vaccine a few days before hospitalization with a bacterial infection.

This work highlights the potential of bacterial and viral detections of RNA sequencing that can be combined with other mapped data from RNA sequencing to revolutionize our knowledge of the complex physiology of infections. To highlight the potential of heterogeneous physiological responses within individuals, we highlight one of the patients with MODS, where we discovered a process we coined virally induced genetics (Fig. 8). The patient, in their teens, presented symptoms of an infection that progressed into MODS

that included hemophagocytic lymphohistiocytosis (HLH) due to a clinically confirmed Epstein–Barr virus (EBV), which was detectable at in the RNA sequencing data. The patient's signatures showed unique outlier genes connected to EBV response and signs of altered RNA processes such as RNA surveillance and NMD.

A rapid clinical exome of the patient found variants for a possible autosomal recessive rare disease in the *RNASEH2B* gene, but the clinical presentation did not match that rare disease. One of the variants they inherited in *RNASEH2B* from their dad was predicted to be benign. The variant they inherited from their mom was a splice change with unknown transcript and protein outcomes. Making sense of a splice variant outcome on a protein requires tools such as RNA sequencing to determine the new splice change that occurs. The patient's RNA sequencing confirmed the presence of a new splice that caused a frameshift and early truncation of the *RNASEH2B* protein. However, using the variants, we could also track in the RNA sequencing the splice variant (maternally inherited) usage was highest at the height of their EBV infection, with recovery showing a higher use of their *RNASEH2B* gene inherited from dad when they recovered. This suggested that NMD could regulate the splice variant and early stop codon frameshift but that something altered NMD during the EBV infection. This aligns with observations that EBV can suppress NMD to replicate, which we suspect resulted in a rare autosomal dominant form of the disease phenotype in the patient only during their viral infection of the blood cells (275).

Further processing the immune repertoire, we could find an antibody signature elevated later in the patient's RNA sequencing. There was suspicion that the antibody signature could potentially be a risk factor for autoimmune disease based on the patient's transient rare disease, HLH, and NETosis. What makes the immune repertoire so powerful, over something like antibody serological insights, is the ability to get sequence-level acquired response data that can be matched to other RNA sequencing samples. With the patient's antibody profile, we could search for blood PAXgene data from the NCBI SRA (283) to find similar datasets that link to other disease phenotypes. Nearly all of the matched samples were from infection patients or those with autoimmunity such as lupus, further suggesting the patient may be at risk for autoimmunity. A 2-year follow-up of the patient showed many signs of these autoimmune issues that were being treated clinically following their PICU EBV complication. This ability to link acute infections to long-term disease risks is one of the most promising aspects of precision transcriptomics.

As the patient in Fig. 8 shows, COVID-19 and the SARS-CoV-2 infection seem to suggest shared responses within antibodies of the immune repertoire. Our COVID-19 study also showed through precision transcriptomics that SARS-CoV-2 is a potent suppressor of immunity in many patients, which can result in the reactivation of dormant viruses such as EBV, shingles, or Torque teno virus, as observed through outlier mapping (300). Other patients had bacterial signatures as outliers for species such as *Pseudomonas*, which are normal flora that can be elevated in immune-suppressed individuals. This suggests that COVID-19, through immune suppression, can elevate patient-specific clinical secondary

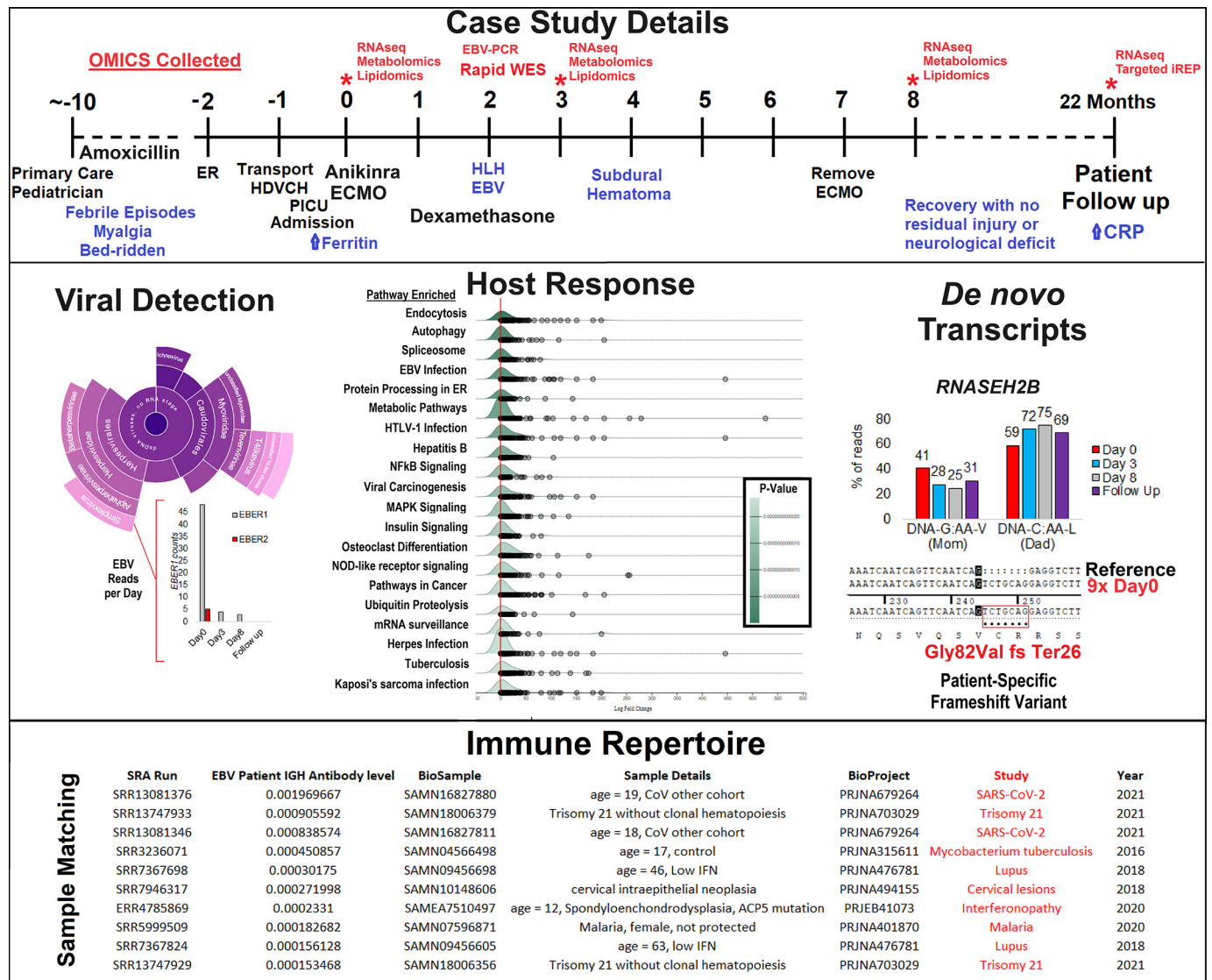


Figure 8. A case study of multiple organ dysfunction that highlights the role of environmental regulation of nonsense-mediated decay transcripts and physiological outcomes. *Top:* shows the clinical course with omic datasets collected in red. *Middle:* shows the patient mapping data from viral detection of EBV, outlier gene ontology enrichment terms relative to the larger cohort of patients, and the allelic imbalance of the RNASEH2B gene as determined by RNA sequencing. *Bottom:* shows the antibody sequences extracted from RNA sequencing matched to thousands of samples within the NCBI SRA. EBV, Epstein-Barr virus; NCBI SRA, National Center for Biotechnology Information Sequence Read Archive. *Time points of omics collection.

infections, so traditional cohort-level mapping approaches would not be significantly powered to measure individual species. Our team has rapidly been expanding these precision transcriptomic tools into other clinical areas, such as inborn errors of immunity, pregnancy complications, rare diseases, traumatic brain injury, and gene therapy, which have the potential for complex interactions of genetics, immune system dynamics, and infections on medical and physiological questions.

Not only can these tools be applied to new sample collections, but they can also be applied to older data within the NCBI SRA. For example, using multidimensional mapping tools, we processed 8,274 RNAseq samples from 132 different experiments for skin samples (302). These tools elucidated many novel discoveries in microbiome-host dynamics, such as the presence of *Bacillus megaterium* correlated with

antibody production in patients with hidradenitis suppurativa. A few powerful computers could synergize skin data into sex differences, antibody profiles across data, novel viruses in humans (equine infectious anemia virus), and elucidate the methods that produce the best mapping of specific data that might be desirable to future studies. This work was led by a medical student passionate about dermatology, fitting the computational analysis during the rigorous medical school training. The potential to pair these mapping tools with artificial intelligence (AI) and the broader resources of the NCBI SRA holds incredible promise to reach a new era of precision medicine.

Integration with Other Omics

RNAseq has revolutionized transcriptomics by enabling high-resolution analysis of gene expression, splicing events,

and noncoding RNA profiles (303). However, the complexity of biological systems often necessitates a multiomics approach to capture the full spectrum of molecular interactions (304). Integrating RNAseq with other omics layers, such as genomics, epigenomics, proteomics, metabolomics, and spatial omics, provides a more comprehensive understanding of cellular physiology, disease mechanisms, and therapeutic targets (305). For instance, the combination of RNAseq with whole genome or exome sequencing enables linking genomic variants, such as single-nucleotide polymorphisms (SNPs) and indels, to gene expression profiles, facilitating the identification of eQTLs, sQTLs, allele-specific expression in disease pathogenesis, and the impact of rare genetic variants on transcriptome regulation (251). As is currently being unraveled in the case of cancer (306), that >90% of cancers are initiated by environmental exposures (exposome), leading to genetic changes in the subsequent proliferating cancer cells. In response to the altered cancerous cell, cancer-specific metabolism, which is genetically reprogrammed, facilitates and sustains this new cancerous cellular growth. These three “omics” work together to create a new hostile environment for the host. Current science is unraveling these complex interactions in an integrated and holistic approach, which gives new hope for prevention and treatment strategies.

Similarly, integrating RNAseq with epigenomics, which focuses on DNA methylation, histone modifications, and chromatin accessibility, provides insights into transcriptional regulation when combined with techniques like ATAC-seq, bisulfite sequencing, ChIP-seq, and nanopore direct sequencing (307–309). This approach allows researchers to identify regulatory elements controlling gene expression, link epigenetic modifications to differential expression and splicing, and explore gene regulation in developmental and disease contexts (308, 310).

Further integration of RNAseq with proteomics addresses the imperfect correlation between transcriptome data and protein abundance due to posttranscriptional regulation (311, 312). By combining transcriptomics with proteomics techniques such as mass spectrometry and protein arrays, researchers can link gene expression to protein abundance and activity, uncover translational control mechanisms and posttranslational modifications, and study protein-protein interaction networks influenced by transcriptomic changes (313, 314). Similarly, pairing RNAseq with metabolomics, which profiles small molecules and metabolites via methods like NMR, UPLC-MS, and GC-MS (315), highlights gene-to-phenotype relationships and uncovers metabolic pathways regulated at the transcriptional level (316, 317). This integration is particularly valuable for understanding metabolic flux (318), disease biomarker discovery, and exploring the molecular underpinnings of conditions such as cancer, neurodegeneration, and metabolic disorders (318, 319).

In addition, spatial omics, including spatial transcriptomics and proteomics, merge RNA expression data with spatial context to map cell type-specific gene expression within tissues, study tumor microenvironments, and identify spatial coexpression patterns in developmental biology (320). Single-cell omics further enhances RNAseq by combining it with single-cell ATAC-seq, proteomics, or spatial data, allowing deep characterization of cellular heterogeneity,

elucidation of developmental trajectories, and mapping of cellular ecosystems in cancer, immune responses, and tissue regeneration (185, 321, 322).

Despite these advances, multiomics data integration presents challenges, including harmonization across platforms with differing resolution and scale, handling batch effects and variability, and interpreting biological relevance within complex datasets (323, 324). Additional practical challenges for the multiomics community include data acquisition, integration (machine learning, unsupervised, supervised, deep learning), analysis, and storage, which require specialized equipment, trained personnel, and abundant financial resources (325, 326). These factors impede the practical and timely clinical translation of this work to the bedside. Additional challenges include a lack of standardization of sample collection, missing data and samples, batch effects, and data complexity, to mention a few. However, some of these barriers are being explored and overcome in the wearable device arena, which offers real-time, continuous treatment and environmental exposure monitoring (327).

Temporal-Rhythmic Expression Dynamics

Circadian rhythms are critical for maintaining physiological homeostasis, where much of our knowledge of circadian rhythms comes from a balance of transcriptomics and protein assessments. This section highlights the power of following techniques like transcriptomics over temporal analyses. Every tissue and cell in the body has a circadian rhythm that will respond to Zeitgebers or time givers. Zeitgebers include light, food intake, and physical activity (328). The suprachiasmatic nucleus (SCN) in the hypothalamus of the brain responds to light cues from the retina, entraining or driving the central circadian clock as well as regulation of body temperature (329), melatonin secretion (330), and signaling to peripheral tissue clocks through autonomic innervation (331). Many studies have shown that aberrant light exposure leads to negative health consequences due to shifts in the circadian phase and difficulty entraining the circadian rhythm (332–334). Multiple recent studies indicated that night-shift workers are at an increased risk of cardiovascular disease due to circadian disruption (335–337). Due to light-induced circadian disruption, many physiological consequences arise, including sleep/wake cycle misalignment and metabolic dysfunction (333, 334, 338).

Peripheral clocks are complex, as each tissue has its clock and will respond differently to different Zeitgebers. The liver and adipose tissue rely on food intake cues for normal tissue rhythmicity (339–342). Mistimed eating can lead to metabolic dysfunction or glucose intolerance (343, 344), increasing the likelihood of obesity and its associated comorbidities (345, 346). The kidney and vasculature also benefit from adequately timed eating, as demonstrated by interventions that alter the timing of feeding (347–349). The Pollock group (coauthors of paper) found that mistimed feeding of healthy food in mice (food available during the light/inactive phase only) completely reversed the blood pressure rhythms, with blood pressure peaks during the light phase and dips during the dark phase, as well as increasing aortic stiffness. Interestingly,

these mice maintained a normal diurnal rhythm of sodium excretion (350). Sutton et al. (351) showed that prediabetic obese men on a 6-h time-restricted feeding (food intake from 8:00 AM to 2:00 PM) intervention had significantly reduced blood pressure compared with controls with a 12–15 h feeding window. The Pollock laboratory has also recently shown that time-restricted feeding intervention in an animal model of diet-induced obesity significantly reduced renal fibrosis and aortic stiffness, highlighting the importance of timing food intake for cardiorenal function (352). Exercise affects circadian rhythms by altering the phase of melatonin to improve the sleep/wake cycle (353, 354) and diurnal fluctuations of muscle strength (355). Exercise has also been found to improve blood pressure rhythms with increased dipping at night (356). These reports further highlight a connection between maintaining circadian rhythms through food cues and/or physical activity, critical to understanding whole body physiology.

RNA transcription is also circadian-dependent (357, 358). Rhythmic RNA expression can occur in up to 80% of the total RNA transcripts in specific tissues (359–361). Menet et al. (357) elegantly showed rhythmic gene expression patterns in nascent RNA and mRNA throughout a 24-h period. Studies have also shown that the recruitment of RNA polymerase II to promoter sites occurs on a diurnal rhythm, along with histone methylation, H3K4me3, and H3K36me3, which are associated with the transcription start site and gene splicing (362). RNA transcription in eukaryotes requires DNA-dependent RNA polymerases that, in turn, utilize transcription factors binding at promoter regions to enhance the assembly of (or repress) the initiation complex (363, 364). Thus, small changes in transcription factor expression (365) and/or nuclear localization (366) can result in dramatic changes in RNA expression (365) and cellular fate (366) by regulating the assembly and transcriptional capacity of the holoenzyme. RNA sequencing studies with application to physiology are susceptible to the time at which the samples were initially obtained, and even 2-h sample-acquisition differences between groups can result in the false attribution of differentially expressed genes to treatment rather than properly to time-of-day (359, 360).

About 30% of all mammalian genes have oscillating expression over a 24-h period. These rhythmic genes regulate the circadian transcription-translational feedback loop, such as *Arntl* (aka *Bmal1*), *Clock*, *Per1*, *Per2*, *Cry1* and *Cry2*, and many non-circadian-associated genes (367–369). The central molecular clock proteins, *Bmal1* and *Clock*, are transcription factors (370, 371) that bind E-Box DNA motifs to drive the expression of thousands of genes either indirectly by increasing chromatin accessibility for the binding of other transcription factors or directly by recruiting RNA polymerase (358, 372–374). *Bmal1* is sensitive to regulation by other proteins (370, 371) and pharmacological interventions (375–377) to drive a phase-shift (delaying or accelerating) in *Bmal1* expression. It is then conceivable that differentially expressed genes identified from RNA-sequencing studies may amount to phase shifts in *Bmal1* or other circadian-relevant transcription factor expression. Thus, RNA-sequencing studies must report the time and conditions at which samples were obtained, and for increased scientific rigor, more than one time-of-day endpoint may need to be incorporated.

There are still many questions about why and how the circadian regulation of specific genes affects physiological function. Of note, the capacity of CLOCK:BMAL1 to bind to E-Box motifs is moderated by tissue and cell-specific chromatin accessibility (373) and local DNA-shape and histone modifications (372), indicating clock protein regulation of rhythmic RNA expression is distinct to each tissue. Tissue-specific and rhythmic regulation of chromatin accessibility results in a transcriptional landscape that is unique and variable for each cell type based on time of day, potentially leading to time-sensitive and clinically relevant tissue-specific responses to intervention and/or challenge, whether by pharmacological intervention or some other insult (360). For example, we know that circadian expression of genes in the kidney plays a critical role in electrolyte homeostasis (378), and dysregulation of these renal transcriptomic rhythms contributes to a blunted capacity to respond to environmental changes and increased renal injury upon hypertensive challenge (378–380). Circadian regulation in the liver controls lipid metabolism (381), detoxification (382), and nutrient uptake (383). Studies in rodents and humans show that exercise affects the rhythmicity of core circadian genes and genes associated with stress, like *Herpud1* (384, 385), expressed explicitly in the skeletal muscle. Microarray analysis of the left ventricle of the heart 8 h post-induction of myocardial infarction indicated significant differences in the inflammatory response, metabolic pathways, and transcription regulation based on whether the infarct was initiated during the beginning of the mouse inactive (lights-on) or active (lights-off) period (386). A similar time-of-day sensitive difference in RNA transcription has also been observed in the human immunologic response to hypoxia (387) and in the responsiveness of various cancers to chemotherapeutics (388, 389).

These observations suggest studies interrogating the transcriptomic response to perturbation may be sensitive to the time the stressor or challenge began. Therefore, considering not just the time at which the sample was obtained but also the time at which the challenge was initiated may prove critical for future RNA sequencing studies to improve pathway identification. As the access and utilization of RNA sequencing technology expand in physiology, physiologically relevant parameters, such as when the experimental intervention took place and when the sample was obtained, will become increasingly critical for precise pathway analysis and for enabling a systematic and consistent comparison between published studies. These metadata are rarely present in previous RNA sequencing experiments, and this represents one of the most likely factors for heterogeneous insights within experiments, especially as the tools are applied to precision medicine approaches.

Direct RNA Sequencing

One of the most unique up-and-coming RNA sequencing platforms is the nanopore direct method, which removes the need for cDNA generation by directly sequencing the RNA molecules through protein pores. Our team has applied this approach to two laboratory-controlled cell culture experiments, highlighting the potential of the tools for exploring the signaling dynamics of RNA base modifications in brain and kidney physiology.

Using induced pluripotent stem cell (iPSC) derived mini-brains, we combined short-read RNA sequencing, single-cell RNA sequencing, and nanopore direct RNA sequencing for the role of diesel particulate matter (DPM) exposure on human brain development (390). Short-read RNA sequencing showed that gene expression within the mini-brains was altered with DPM treatment, with the more expensive single-cell RNA sequencing then showing that the mini-brains had 19 cell types with significant gene expression changes in several cell clusters. Genes altered in expression were associated with oxidative phosphorylation. The nanopore direct RNA sequencing was able to compare the raw voltage data of control or DPM exposure mini-brains to detect subtle shifts that are indicative of RNA base modifications (Fig. 9, left top).

Interestingly, many of the genes where a modification occurred showed multiple sites within the transcript to be modified (Fig. 9, left middle). The genes where these modifications occurred were not random but showed changes to RNA transcripts significantly enriched to mitochondrial function and oxidative phosphorylation (Fig. 9, left bottom).

A second study linked hypoxia (1% oxygen) exposure within confluent (nonproliferative), sheer-stress-treated (391) human renal proximal tubule epithelial cells (RPTEC) to signal biology (122). Using short-read RNA sequencing, we established a sheer stress model that, with additional immunofluorescence, showed how hypoxia alters cilia physiology within a cell culture model (391). RNA sequencing also established that this model can capture a transient state of

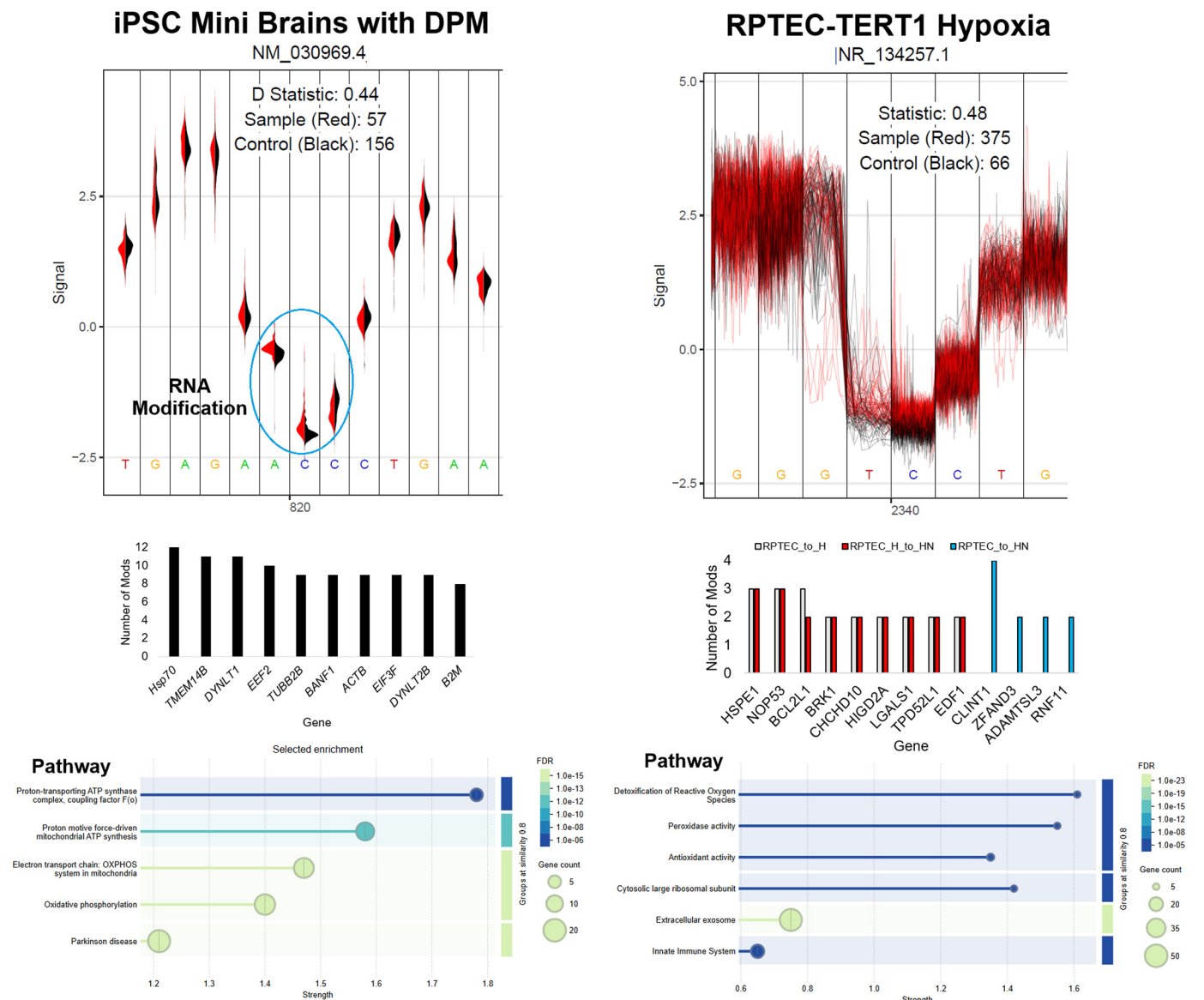


Figure 9. Nanopore-based direct RNAseq elucidates RNA modification signal transduction due to environmental alterations. *Left:* shows data from iPSC-derived mini brains exposed to diesel particulate matter (DPM). *Right:* show data from RPTEC-TERT1 kidney cells exposed to 1% hypoxia. *Top:* show representative voltage plots used to map modification sites with the mini-brain shown with violin representation and the kidney plot shown as raw voltage overlay plots. *Middle:* show top genes with the number of identified modifications within the transcripts. The kidney panel shows the differential modification detections over the three labeled experiments. *Bottom:* show pathways enriched for modified genes based on STRING analysis. H, hypoxic exposure; HN, hypoxic to normoxic exposures; iPSC, induced pluripotent stem cell.

epithelial cells for epithelial-mesenchymal transition, similar to how hypoxia can induce cancer metastasis in epithelial cells.

Using this RPTEC hypoxia model, we applied direct nanopore RNA sequencing, which identified RNA modification sites (Fig. 9, right top) (122). We used three cell groups and performed three comparisons of modifications for this study: 1) control cells versus 48 h of hypoxia exposure (RPTEC_to_H), 2) 48 h of hypoxia exposure versus cells exposed for 48 h hypoxia followed by 48 h with normal oxygen (RPTEC_H_to_HN), and 3) control to the reoxygenated cells (RPTEC_to_HN). This work also suggested that RNA base modifications occurred multiple times within multiple transcripts (Fig. 9, right middle). Interestingly, the reoxygenation of the cells for 48 h seemed to reverse most modification sites, with some modifications that were stable for longer, which we hypothesize is based on RNA half-life differences. The genes that showed modification due to hypoxia were not random but enriched for pathways such as reactive oxygen species and innate immunity, both well established in hypoxia response. One of the top modified transcripts was *RHOA*, a well-established control gene for cancer metastasis (122).

Having two models where case versus control direct nanopore-based statistics suggest targeted modifications to genes involved in physiological pathways already established to logical outcomes of the models suggests that RNA base modifications may have functions in signal transductions. This could be similar to protein phosphorylation, which revolutionized our insights into signal transduction (392, 393). RNA base modifications are well established to control RNA transcript turnover and stability (227). This could have rapid signal potential through RNA to protein translation rates that could take environmental signals and metabolic control of enzymes to manifest amplification to cellular physiology. If true, direct RNA sequencing with nanopore could also revolutionize our signal transduction knowledge over the coming years if applied to broad physiological experiments, especially those where control and treatment can be compared.

Artificial Intelligence and Machine Learning

As the dimensions of data from RNA sequencing increase to include splicing dynamics, RNA biotype insights, foreign RNA (such as bacteria and virus), the immune repertoire, cell-specificity, tissue spatial dynamics, cyclical expression dynamics, RNA modifications, disease-centered profiles, sex/gender differences, and the many other aspects of RNA biology, the ability to understand the complex system exceeds what scientists can comprehend. There is a growing need for machine learning and artificial intelligence tools to help organize and interpret the complex biology within RNA. Early in this phase, the data integrations and models often focused on single factors such as expression profiles (394), immunotherapies (395), single-cell atlases (270, 396, 397), and spatial transcriptome deconvolution (398–400). One of the exciting emerging areas of RNA and artificial intelligence is the rapid growth of RNA structure predictions that have long challenged our knowledge of how RNA functions within a cell (401, 402). As discussed throughout this review, if we imagine an AI system that can inform physiology around

homeostasis and heterogeneity, there must be modeling that incorporates the many dimensions of RNA knowledge addressed throughout this review.

Interventions with RNA

As our knowledge of RNA has expanded, it has opened the door for therapeutic interventions. Delivery of RNA and RNA-regulating molecules is rapidly emerging therapeutically to manipulate gene expression (403), targeting established genes in infectious diseases, cancers, the immune system, and Mendelian rare diseases (128). This means that the scientific results of short-read, long-read, single-cell, and spatial RNA sequencing detection of RNA involved in disease transformation can be targeted with synthesized mRNA (to elevate), antisense RNA (to reduce), or CRISPR-mediated expression regulation (elevate or reduce). Multiple of these molecules are now FDA-approved and have direct clinical usage, highlighting the future of RNA-based therapies (128).

RNA modifications, including methylation, an epigenetic modification that is very different from the mechanisms that govern DNA methylation, have an impact on the half-life of therapeutic RNA. Protein regulation of the RNA methylation includes “writers,” “erasers,” and “readers” that respectively deposit, remove, and recognize methylated RNA. RNA methylation at many levels, particularly N6-methyladenosine (m6A), 5-methylcytosine (m5C), N3-methylcytosine (m3C), N1-methyladenosine (m1A), and N7-methylguanosine (m7G), has been suggested as disease therapeutic targets (404). More work is needed in this rapidly emerging area on how best to control synthesized RNA therapies.

Delivery of the RNA molecules can be based on viral packaging or lipid-based particles with some target specificity. The specificity can be further selected using the expression construct coded within the delivery system based on cell-specific promoters/enhancers driven by extensive single-cell level expression integrations (405). Many of the current clinical trials of gene therapies were based on bulk tissue-specific RNA expression datasets, where we anticipate a new era of cell-level expression control of gene therapy systems to reduce off-target profiles and complications (128, 406). Spatial transcriptomics applied to disease will yield single-cell, region-specific insights that allow even further refining of disease-specific expression control (407). One of the most significant advancements in RNA, which elevated to the Nobel Prize, was the understanding of how RNA modifications contribute to the half-life of RNA (408–410), which elevated the ability to develop mRNA-based vaccines during COVID-19 (411).

Developing drug targets needs relative abundance and knowledge of the function of RNA transcripts paired with knowledge of protein abundance and outcomes, emphasizing expression dynamics like circadian rhythms that are often overlooked. These systems are not always consistent between humans and model organisms such as rodents, thus requiring more human systems for developing RNA-centered therapies. The further development of human wet laboratory models, such as induced pluripotent stem cells (iPSCs) and other ex vitro models, holds promise in validating RNA therapies before human clinical trials. These human-centered RNA models could enable the expansion of

heterogeneity insights that often lead to failed clinical trials and expanding lists of adverse outcomes following FDA approvals (128). It is also critical to elevate programs in biomarker assessments for RNA gene therapies even after they have been FDA-approved, using the RNA sequencing tools addressed throughout this article. This is critical given the often low sample size of clinical trials that cannot capture the broad heterogeneous responses within cells to the RNA and the immune response to the packaging material (128). The advancing AI tools that integrate the broader RNA insights are likely to inform more on the heterogeneous risks of RNA therapies and to develop more homeostasis controls that optimize the “Goldilocks” needs of physiological control without adverse outcomes.

The introduction of machine learning and artificial intelligence to large volumes and diverse types of RNA data can guide many processes of RNA-based therapies from delivery, cell-specific expression, disease-centric control, and fine-tuning the system to identify the Goldilocks of expression to have just the right amount of RNA function for physiological needs while accounting for cyclical dynamics and feedback loops. As data increases and computational tools are developed to integrate better RNA sequencing platforms mentioned earlier, our ability to develop a toolbox of resources in gene-based therapies grows. This toolbox of RNA can likely enable lower-cost, faster-moving therapy opportunities that reflect the alignment of physiological heterogeneity and homeostasis that define precision medicine in the future.

CONCLUSIONS

RNA sequencing continues to expand in usage and methods. The next decade will likely see new technologies and techniques for processing RNA, with significant advancements in cell physiology mechanisms. The emerging market competition in sequencing is likely to drive innovation and cost reductions that will facilitate more RNA sequencing into physiology projects. The tools will be applied to broad physiological systems and model organisms, while their use for processing clinical samples will also elevate. Our knowledge of disease will be enhanced as the tools are applied. The challenges of selecting which tool to use for which samples will not resolve soon, with the trade-off for depth of knowledge relative to costs getting more difficult as costs decrease on platforms such as single-cell and spatial. The depth of data generated will also get more challenging, requiring laboratories to understand existing data to inform new experimental designs. Like the roots of all physiology, observational discoveries will be plentiful for laboratories utilizing RNA sequencing, better designing hypotheses built around the massive volumes of RNA insights, and the depth of diverse samples already available. More importantly, every physiologist will now be powered with tools that can better resolve the complexities of homeostasis and heterogeneity of disease, enabling the future of precision medicine that was promised with the completion of the human genome.

GRANTS

This work was supported by funds from the American Heart Association 23SCISA1144887 (to J.W.P.), the National Institutes of

Health Grant R01AI171984 (to J.W.P., T.T., and C.M.K.), and the Corewell Health Foundation (to J.W.P. and S.R.).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

J.W.P.: Conceptualization, Formal analysis, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing; S.M.B.: Formal analysis, Visualization, Writing – original draft, Writing – review & editing; E.T.: Formal analysis, Visualization, Writing – original draft, Writing – review & editing; S.V.: Writing – original draft, Writing – review & editing; S.V.W.: Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing; A.S.D.: Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing; S.B.C.: Formal analysis, Writing – original draft, Writing – review & editing; A.D.: Writing – original draft, Writing – review & editing; S.A.: Writing – original draft, Writing – review & editing; A.G.: Writing – original draft, Writing – review & editing; K.L.B.: Writing – original draft, Writing – review & editing; M.W.: Writing – original draft, Writing – review & editing; E.A.V.: Writing – original draft, Writing – review & editing; C.J.E.: Writing – original draft, Writing – review & editing; L.N.B.: Writing – original draft, Writing – review & editing; D.B.C.: Formal analysis, Writing – original draft, Writing – review & editing; C.P.B.: Formal analysis, Writing – original draft, Writing – review & editing; A.H.: Writing – original draft, Writing – review & editing; N.L.H.: Formal analysis, Writing – original draft, Writing – review & editing; J.M.K.: Formal analysis, Writing – original draft, Writing – review & editing; M.C.: Formal analysis, Writing – original draft, Writing – review & editing; M.S.: Formal analysis, Writing – original draft, Writing – review & editing; M.A.T.K.: Formal analysis, Writing – original draft, Writing – review & editing; A.M.P.: Formal analysis, Writing – original draft, Writing – review & editing; S.F.G.: Writing – original draft, Writing – review & editing; T.T.: Formal analysis, Writing – original draft, Writing – review & editing; E.H.: Formal analysis, Writing – original draft, Writing – review & editing; M.L.L.-L.: Formal analysis, Writing – original draft, Writing – review & editing; C.M.K.: Formal analysis, Writing – original draft, Writing – review & editing; J.S.P.: Conceptualization, Supervision, Writing – original draft, Writing – review & editing; S.R.: Conceptualization, Formal analysis, Supervision, Writing – original draft, Writing – review & editing.

REFERENCES

1. Franco NH. Animal experiments in biomedical research: a historical perspective. *Animals (Basel)* 3: 238–273, 2013. doi:10.3390/ani3010238.
2. Westerhof N. A short history of physiology. *Acta Physiol (Oxf)* 202: 601–603, 2011. doi:10.1111/j.1748-1716.2011.02286.x.
3. Karasszon D. From the “physiologus” to physiology. *Acta Physiol Hung* 100: 246–250, 2013. doi:10.1556/APhysiol.100.2013.2.11.
4. Borell M. Instrumentation and the rise of modern physiology. *Science & Technology Studies* 5: 53–62, 1987.
5. Carroll RG, Frank M, Ra’anan A, Matyas ML. APS: 125 years of progress of physiology as a scientific discipline and a profession. *Adv Physiol Educ* 37: 1–9, 2013. doi:10.1152/advan.00146.2012.
6. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* 2: 986–991, 2001. doi:10.1093/embo-reports/kve230.
7. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 47: D766–D773, 2019. doi:10.1093/nar/gky955.

8. **Ozsolak F, Milos PM.** RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12: 87–98, 2011. doi:10.1038/nrg2934.
9. **Prokop JW, May T, Strong K, Bilinovich SM, Bupp C, Rajasekaran S, Worthey EA, Lazar J.** Genome sequencing in the clinic: the past, present, and future of genomic medicine. *Physiol Genomics* 50: 563–579, 2018. doi:10.1152/physiolgenomics.00046.2018.
10. **Rubin H.** The significance of biological heterogeneity. *Cancer Metastasis Rev* 9: 1–20, 1990. doi:10.1007/BF00047585.
11. **Bassingthwaite JB.** Physiological heterogeneity: fractals link determinism and randomness in structures and functions. *News Physiol Sci* 3: 5–10, 1988. doi:10.1152/physiologyonline.1988.3.1.5.
12. **Sun YH, Wu YL, Liao BY.** Phenotypic heterogeneity in human genetic diseases: ultrasensitivity-mediated threshold effects as a unifying molecular mechanism. *J Biomed Sci* 30: 58, 2023. doi:10.1186/s12929-023-00959-7.
13. **Han X, Xing L, Hong Y, Zhang X, Hao B, Lu JY, Huang M, Wang Z, Ma S, Zhan G, Li T, Hao X, Tao Y, Li G, Zhou S, Zheng Z, Shao W, Zeng Y, Ma D, Zhang W, Xie Z, Deng H, Yan J, Deng W, Shen X.** Nuclear RNA homeostasis promotes systems-level coordination of cell fate and senescence. *Cell Stem Cell* 31: 694–716.e11, 2024. doi:10.1016/j.stem.2024.03.015.
14. **Strange K.** Cellular volume homeostasis. *Adv Physiol Educ* 28: 155–159, 2004. doi:10.1152/advan.00034.2004.
15. **Reichholf B, Herzog VA, Fasching N, Manzenreither RA, Sowemimo I, Ameres SL.** Time-resolved small RNA sequencing unravels the molecular principles of MicroRNA homeostasis. *Mol Cell* 75: 756–768.e7, 2019. doi:10.1016/j.molcel.2019.06.018.
16. **Wakabayashi T, Naito H.** Cellular heterogeneity and stem cells of vascular endothelial cells in blood vessel formation and homeostasis: insights from single-cell RNA sequencing. *Front Cell Dev Biol* 11: 1146399, 2023. doi:10.3389/fcell.2023.1146399.
17. **Mahata B, Zhang X, Kolodziejczyk AA, Proserpio V, Haim-Vilimovsky L, Taylor AE, Hebenstreit D, Dingler FA, Moignard V, Göttgens B, Arit W, McKenzie ANJ, Teichmann SA.** Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Rep* 7: 1130–1142, 2014. doi:10.1016/j.celrep.2014.04.011.
18. **Liu Z, Ruter DL, Quigley K, Tanke NT, Jiang Y, Bautch VL.** Single-cell RNA sequencing reveals endothelial cell transcriptome heterogeneity under homeostatic laminar flow. *Arterioscler Thromb Vasc Biol* 41: 2575–2584, 2021. doi:10.1161/ATVBAHA.121.316797.
19. **Gallardo-Dodd CJ, Kutter C.** The regulatory landscape of interacting RNA and protein pools in cellular homeostasis and cancer. *Hum Genomics* 18: 109, 2024. doi:10.1186/s40246-024-00678-6.
20. **Meizlish ML, Franklin RA, Zhou X, Medzhitov R.** Tissue homeostasis and inflammation. *Annu Rev Immunol* 39: 557–581, 2021. doi:10.1146/annurev-immunol-061020-053734.
21. **Chovatiya R, Medzhitov R.** Stress, inflammation, and defense of homeostasis. *Mol Cell* 54: 281–288, 2014. doi:10.1016/j.molcel.2014.03.030.
22. **Hartl FU.** Cellular homeostasis and aging. *Annu Rev Biochem* 85: 1–4, 2016. doi:10.1146/annurev-biochem-011116-110806.
23. **Neppl RL, Wang DZ.** The myriad essential roles of microRNAs in cardiovascular homeostasis and disease. *Genes Dis* 1: 18–39, 2014. doi:10.1016/j.gendis.2014.06.003.
24. **Ounzain S, Crippa S, Pedrazzini T.** Small and long non-coding RNAs in cardiac homeostasis and regeneration. *Biochim Biophys Acta* 1833: 923–933, 2013. doi:10.1016/j.bbamcr.2012.08.010.
25. **McClellan J, King MC.** Genetic heterogeneity in human disease. *Cell* 141: 210–217, 2010. doi:10.1016/j.cell.2010.03.032.
26. **Sanders M, Lawlor MJM, Li X, Schuen JN, Millard SL, Zhang X, Buck L, Grysko B, Uhl KL, Hinds D, Stenger CL, Morris M, Lamb N, Levy H, Bupp C, Prokop JW.** Genomic, transcriptomic, and protein landscape profile of CFTR and cystic fibrosis. *Hum Genet* 140: 423–439, 2021. doi:10.1007/s00439-020-02211-w.
27. **Ising M, Holsboer F.** Genetics of stress response and stress-related disorders. *Dialogues Clin Neurosci* 8: 433–444, 2006. doi:10.31887/DCNS.2006.8.4/missing.
28. **Ma ZS.** Heterogeneity-disease relationship in the human microbiome-associated diseases. *FEMS Microbiol Ecol* 96: fiae093, 2020. doi:10.1093/femsec/fiae093.
29. **Casanova JL, Abel L.** The human genetic determinism of life-threatening infectious diseases: genetic heterogeneity and physiological homogeneity? *Hum Genet* 139: 681–694, 2020. doi:10.1007/s00439-020-02184-w.
30. **International Human Genome Sequencing Consortium.** Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945, 2004. doi:10.1038/nature03001.
31. **Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M.** Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364, 2012. doi:10.1155/2012/251364.
32. **Goodwin S, McPherson JD, McCombie WR.** Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17: 333–351, 2016. doi:10.1038/nrg.2016.49.
33. **Wang Z, Gerstein M, Snyder M.** RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57–63, 2009. doi:10.1038/nrg2484.
34. **Chu Y, Corey DR.** RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* 22: 271–274, 2012. doi:10.1089/nat.2012.0367.
35. **ENCODE Project Consortium.** The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306: 636–640, 2004. doi:10.1126/science.1105136.
36. **Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al.** GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760–1774, 2012. doi:10.1101/gr.135350.111.
37. **Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Lagarde J, Gilbert JG, Storey R, Swarbreck D, Rossier C, Ucla C, Hubbard T, Antonarakis SE, Guigo R.** GENCODE: producing a reference annotation for ENCODE. *Genome Biol* 7, Suppl 1: S4.1–S4.9, 2006. doi:10.1186/gb-2006-7-s1-s4.
38. **Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al.** Gateways to the FANTOM₅ promoter level mammalian expression atlas. *Genome Biol* 16: 22, 2015. doi:10.1186/s13059-014-0560-6.
39. **Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, MacLeod JN, Chiang DY, Prins JF, Liu J.** MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* 38: e178, 2010. doi:10.1093/nar/gkq622.
40. **Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM.** Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458: 97–101, 2009. doi:10.1038/nature07638.
41. **Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM.** Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* 15: 419, 2014. doi:10.1186/1471-2164-15-419.
42. **Wahl A, Huptas C, Neuhaus K.** Comparison of rRNA depletion methods for efficient bacterial mRNA sequencing. *Sci Rep* 12: 5765, 2022. doi:10.1038/s41598-022-09710-y.
43. **Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al.** GENCODE 2021. *Nucleic Acids Res* 49: D916–D923, 2021. doi:10.1093/nar/gkaa1087.
44. **Ament IH, DeBruyne N, Wang F, Lin L.** Long-read RNA sequencing: a transformative technology for exploring transcriptome complexity in human diseases. *Mol Ther* 33: 883–894, 2025. doi:10.1016/j.ymthe.2024.11.025.
45. **Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Guil Q.** Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 21: 30, 2020. doi:10.1186/s13059-020-1935-5.
46. **Mudge JM, Carbonell-Sala S, Diekhans M, Martinez JG, Hunt T, Jungreis I, et al.** GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* 53: D966–D975, 2025. doi:10.1093/nar/gkae1078.
47. **Conrad T, Albrecht AS, de Melo Costa VR, Sauer S, Meierhofer D, Ørom UA.** Serial interactome capture of the human cell nucleus. *Nat Commun* 7: 11212, 2016. doi:10.1038/ncomms11212.
48. **Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS.** Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat Protoc* 9: 989–1009, 2014. doi:10.1038/nprot.2014.058.
49. **Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, Araújo-Bravo MJ, Lee J, Fishman M, Robbins GE, Lin X, Venepally P, Badger JH, Galbraith DW, Gage FH, Lasken RS.** RNA-sequencing from single nuclei. *Proc*

- Natl Acad Sci USA* 110: 19802–19807, 2013. doi:10.1073/pnas.1319700110.
50. Calviello L, Ohler U. Beyond read-counts: ribo-seq data analysis to understand the functions of the transcriptome. *Trends Genet* 33: 728–744, 2017. doi:10.1016/j.tig.2017.08.003.
51. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, Stevanovic S, Zimmer R, Dölken L. Improved Ribo-seq enables identification of cryptic translation events. *Nat Methods* 15: 363–366, 2018. doi:10.1038/nmeth.4631.
52. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, Couso JP. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife* 3: e03528, 2014. doi:10.7554/eLife.03528.
53. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. *Mol Cell* 58: 610–620, 2015. doi:10.1016/j.molcel.2015.04.005.
54. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 15: e8746, 2019. doi:10.15252/msb.20188746.
55. Ziegenhain C, Vieth B, Parekh S, Reinis B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 65: 631–643.e4, 2017. doi:10.1016/j.molcel.2017.01.023.
56. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol* 41: 773–782, 2023. doi:10.1038/s41587-022-01448-2.
57. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. *Genome Med* 14: 68, 2022. doi:10.1186/s13073-022-01075-1.
58. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder P, Bergmann O, Lundberg J, Frisén J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353: 78–82, 2016. doi:10.1126/science.aaf2403.
59. Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. *Nat Methods* 19: 1160–1164, 2022. doi:10.1038/s41592-022-01633-w.
60. Garalde DR, Snell EA, Jachimowicz D, Sipsos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, Jordan M, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace EJ, Jayasinghe L, Wright C, Blasco J, Young S, Brocklebank D, Juul S, Clarke J, Heron AJ, Turner DJ. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15: 201–206, 2018. doi:10.1038/nmeth.4577.
61. Zarei M. Advances in point-of-care technologies for molecular diagnostics. *Biosens Bioelectron* 98: 494–506, 2017. doi:10.1016/j.bios.2017.07.024.
62. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419, 2017. doi:10.1038/nmeth.4197.
63. Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* 34: 77–137, 2000. doi:10.1146/annurev.genet.34.1.77.
64. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci USA* 104: 19428–19433, 2007. doi:10.1073/pnas.0709013104.
65. Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48: 249–257, 2009. doi:10.1016/j.ymeth.2009.03.016.
66. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. *Nature* 379: 131–137, 1996. doi:10.1038/379131a0.
67. Melissari MT, Grote P. Roles for long non-coding RNAs in physiology and disease. *Pflugers Arch* 468: 945–958, 2016. doi:10.1007/s00424-016-1804-y.
68. Dayal Aggarwal D, Mishra P, Yadav G, Mitra S, Patel Y, Singh M, Sahu RK, Sharma V. Decoding the connection between lncRNA and obesity: perspective from humans and Drosophila. *Heliyon* 10: e35327, 2024. doi:10.1016/j.heliyon.2024.e35327.
69. Jin H, Du W, Huang W, Yan J, Tang Q, Chen Y, Zou Z. lncRNA and breast cancer: progress from identifying mechanisms to challenges and opportunities of clinical treatment. *Mol Ther Nucleic Acids* 25: 613–637, 2021. doi:10.1016/j.omtn.2021.08.005.
70. Scholda J, Nguyen TTA, Kopp F. Long noncoding RNAs as versatile molecular regulators of cellular stress response and homeostasis. *Hum Genet* 143: 813–829, 2024. doi:10.1007/s00439-023-02604-7.
71. Kohlmaier A, Holdt LM, Teupser D. Long noncoding RNAs in cardiovascular disease. *Curr Opin Cardiol* 38: 179–192, 2023. doi:10.1097/HCO.0000000000001041.
72. Wang L, Bitar M, Lu X, Jacquelin S, Nair S, Sivakumaran H, Hillman KM, Kaufmann S, Ziegman R, Casciello F, Gowda H, Rosenbluh J, Edwards SL, French JD. CRISPR-Cas13d screens identify KILR, a breast cancer risk-associated lncRNA that regulates DNA replication and repair. *Mol Cancer* 23: 101, 2024. doi:10.1186/s12943-024-02021-y.
73. Sun M, Kraus WL. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocr Rev* 36: 25–64, 2015. doi:10.1210/er.2014-1034.
74. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA* 5: 17, 2019. doi:10.3390/ncrna5010017.
75. Yan B, Wang Z. Long noncoding RNA: its physiological and pathological roles. *DNA Cell Biol* 31, Suppl 1: S34–S41, 2012. doi:10.1089/dna.2011.1544.
76. Leone S, Santoro R. Challenges in the analysis of long noncoding RNA functionality. *FEBS Lett* 590: 2342–2353, 2016. doi:10.1002/1873-3468.12308.
77. Chen H, Shan G. The physiological function of long-noncoding RNAs. *Noncoding RNA Res* 5: 178–184, 2020. doi:10.1016/j.ncrna.2020.09.003.
78. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, Chen R, Dean C, Dinger ME, Fitzgerald KA, Gingeras TR, Guttman M, Hirose T, Huarte M, Johnson R, Kanduri C, Kapranov P, Lawrence JB, Lee JT, Mendell JT, Mercer TR, Moore KJ, Nakagawa S, Rinn JL, Spector DL, Ulitsky I, Wan Y, Wilusz JE, Wu M. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* 24: 430–447, 2023. doi:10.1038/s41580-022-00566-8.
79. Giral H, Landmesser U, Kratzer A. Into the wild: GWAS exploration of non-coding RNAs. *Front Cardiovasc Med* 5: 181, 2018. doi:10.3389/fcvm.2018.00181.
80. Wu K, Bu F, Wu Y, Zhang G, Wang X, He S, Liu MF, Chen R, Yuan H. Exploring noncoding variants in genetic diseases: from detection to functional insights. *J Genet Genomics* 51: 111–132, 2024. doi:10.1016/j.jgg.2024.01.001.
81. Chen LL, Kim VN. Small and long non-coding RNAs: past, present, and future. *Cell* 187: 6451–6485, 2024. doi:10.1016/j.cell.2024.10.024.
82. Gordon JM, Phizicky DV, Neugebauer KM. Nuclear mechanisms of gene expression control: pre-mRNA splicing as a life or death decision. *Curr Opin Genet Dev* 67: 67–76, 2021. doi:10.1016/j.gde.2020.11.002.
83. Shamnas VM, Singh A, Kumar A, Mishra GP, Sinha SK. Exons: offering new roles to retained introns-the novel regulators of protein diversity and utility. *AoB Plants* 16: plae014, 2024. doi:10.1093/aobpla/plae014.
84. David JK, Maden SK, Wood MA, Thompson RF, Nellore A. Retained introns in long RNA-seq reads are not reliably detected in sample-matched short reads. *Genome Biol* 23: 240, 2022. doi:10.1186/s13059-022-02789-6.
85. Broseus L, Ritchie W. Challenges in detecting and quantifying intron retention from next generation sequencing data. *Comput Struct Biotechnol J* 18: 501–508, 2020. doi:10.1016/j.csbj.2020.02.010.
86. González-Iglesias A, Arcas A, Domingo-Muelas A, Mancini E, Galcerán J, Valcárcel J, Farías I, Nieto MA. Intron retention tightly regulates the stemness/differentiation switch in the adult neurogenic niche. *Nat Commun* 15: 2837, 2024 [Erratum in *Nat Commun* 15: 3458, 2024]. doi:10.1038/s41467-024-47092-z.
87. Ruiz-Gabarre D, Vallés-Saiz L, Carnero-Espejo A, Ferrer I, Hernández F, García-Escudero R, Ávila J, García-Escudero V. Intron retention as a productive mechanism in human MAPT: RNA species generated by retention of intron 3. *EBioMedicine* 100: 104953, 2024. doi:10.1016/j.ebiom.2023.104953.

88. Pabis K, Barardo D, Sirbu O, Selvarajoo K, Gruber J, Kennedy BK. A concerted increase in readthrough and intron retention drives transposon expression during aging and senescence. *eLife* 12: RP87811, 2024. doi:10.7554/eLife.87811.
89. Northrup V, Perez LJ, Edgett BA, Karakach T, Simpson JA, Brunt KR. Intron retention is a mechanism of erythropoietin regulation in brain cell models. *Gene* 898: 148099, 2024. doi:10.1016/j.gene.2023.148099.
90. Zheng JT, Lin CX, Fang ZY, Li HD. Intron retention as a mode for RNA-Seq data analysis. *Front Genet* 11: 586, 2020. doi:10.3389/fgene.2020.00586.
91. Xia B, Shen J, Zhang H, Chen S, Zhang X, Song M, Wang J. The alternative splicing landscape of infarcted mouse heart identifies isoform level therapeutic targets. *Sci Data* 11: 1154, 2024. doi:10.1038/s41597-024-03998-3.
92. Lejeune F. Nonsense-mediated mRNA decay at the crossroads of many cellular pathways. *BMB Rep* 50: 175–185, 2017. doi:10.5483/BMBRep.2017.50.4.015.
93. Lejeune F. Nonsense-mediated mRNA decay, a finely regulated mechanism. *Biomedicines* 10: 141, 2022. doi:10.3390/biomedicines10010141.
94. Powers KT, Szeto JA, Schaffitzel C. New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes. *Curr Opin Struct Biol* 65: 110–118, 2020. doi:10.1016/j.sbi.2020.06.011.
95. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 96: 307–310, 1999. doi:10.1016/S0092-8674(00)80542-5.
96. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76: 51–74, 2007. doi:10.1146/annurev.biochem.76.050106.093909.
97. Teran NA, Nachun DC, Eulalio T, Ferraro NM, Smail C, Rivas MA, Montgomery SB. Nonsense-mediated decay is highly stable across individuals and tissues. *Am J Hum Genet* 108: 1401–1408, 2021. doi:10.1016/j.ajhg.2021.06.008.
98. Petric Howe M, Patani R. Nonsense-mediated mRNA decay in neuronal physiology and neurodegeneration. *Trends Neurosci* 46: 879–892, 2023. doi:10.1016/j.tins.2023.07.001.
99. Torene RI, Guillen Sacoto MJ, Millan F, Zhang Z, McGee S, Oetjens M, Heise E, Chong K, Sidlow R, O'Grady L, Sahai I, Martin CL, Ledbetter DH, Myers SM, Mitchell KJ, Retterer K. Systematic analysis of variants escaping nonsense-mediated decay uncovers candidate Mendelian diseases. *Am J Hum Genet* 111: 70–81, 2024. doi:10.1016/j.ajhg.2023.11.007.
100. Dyle MC, Kolakada D, Cortazar MA, Jagannathan S. How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdiscip Rev RNA* 11: e1560, 2020. doi:10.1002/wrna.1560.
101. Nissim S, Leshchiner I, Mancias JD, Greenblatt MB, Maertens O, Cassa CA, et al. Mutations in RABL3 alter KRAS prenylation and are associated with hereditary pancreatic cancer. *Nat Genet* 51: 1308–1314, 2019. doi:10.1038/s41588-019-0475-y.
102. Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean JA, Wilkinson MF. An alternative branch of the nonsense-mediated decay pathway. *EMBO J* 26: 1820–1830, 2007. doi:10.1038/sj.emboj.7601628.
103. Behera A, Panigrahi GK, Sahoo A. Nonsense-mediated mRNA decay in human health and diseases: current understanding, regulatory mechanisms and future perspectives. *Mol Biotechnol* 67: 3374–3390, 2025. doi:10.1007/s12033-024-01267-7.
104. Ma Z, Sharma R, Rogers AN. Physiological consequences of nonsense-mediated decay and its role in adaptive responses. *Biomedicines* 12: 1110, 2024. doi:10.3390/biomedicines12051110.
105. Patro AK, Panigrahi GK, Majumder S, Das R, Sahoo A. Nonsense-mediated mRNA decay: physiological significance, mechanistic insights and future implications. *Pathol Res Pract* 264: 155677, 2024. doi:10.1016/j.prp.2024.155677.
106. Lambert JM, Ashi MO, Srouf N, Delpy L, Saulière J. Mechanisms and regulation of nonsense-mediated mRNA decay and nonsense-associated altered splicing in lymphocytes. *Int J Mol Sci* 21: 1335, 2020. doi:10.3390/ijms21041335.
107. Frischmeyer PA, Dietz HC. Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet* 8: 1893–1900, 1999. doi:10.1093/hmg/8.10.1893.
108. Longman D, Jackson-Jones KA, Maslon MM, Murphy LC, Young RS, Stoddart JJ, Hug N, Taylor MS, Papadopoulos DK, Cáceres JF. Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum. *Genes Dev* 34: 1075–1088, 2020. doi:10.1101/gad.338061.120.
109. Ottens F, Efstathiou S, Hoppe T. Cutting through the stress: RNA decay pathways at the endoplasmic reticulum. *Trends Cell Biol* 34: 1056–1068, 2024. doi:10.1016/j.tcb.2023.11.003.
110. Mailliot J, Vivoli-Vega M, Schaffitzel C. No-nonsense: insights into the functional interplay of nonsense-mediated mRNA decay factors. *Biochem J* 479: 973–993, 2022. doi:10.1042/BCJ20210556.
111. Miller JN, Pearce DA. Nonsense-mediated decay in genetic disease: friend or foe? *Mutat Res Rev Mutat Res* 762: 52–64, 2014. doi:10.1016/j.mrrev.2014.05.001.
112. McMahon M, Maquat LE. Exploring the therapeutic potential of modulating nonsense-mediated mRNA decay. *RNA* 31: 333–348, 2025. doi:10.1261/rna.080334.124.
113. Baralle D, Buratti E. RNA splicing in human disease and in the clinic. *Clin Sci (Lond)* 131: 355–368, 2017. doi:10.1042/CS20160211.
114. Li YI, Knowles DA, Humphrey J, Barreira AN, Dickinson SP, Im HK, Pritchard JK. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* 50: 151–158, 2018. doi:10.1038/s41588-017-0004-9.
115. Seth R, Crook S, Ibrahim S, Fadhil W, Jackson D, Ilyas M. Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer. *Gut* 58: 1234–1241, 2009. doi:10.1136/gut.2008.159137.
116. Nuevo-Tapióles C, Philips MR. The role of KRAS splice variants in cancer biology. *Front Cell Dev Biol* 10: 1033348, 2022. doi:10.3389/fcell.2022.1033348.
117. Koehn OJ, Lorimer E, Unger B, Harris R, Das AS, Suazo KF, Auger SA, Distefano MD, Prokop JW, Williams CL. GTPase splice variants RAC1 and RAC1B display isoform-specific differences in localization, prenylation, and interaction with the chaperone protein SmgGDS. *J Biol Chem* 299: 104698, 2023. doi:10.1016/j.jbc.2023.104698.
118. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* 33: 591–601, 2009. doi:10.1016/j.molcel.2009.01.025.
119. Guo W, Bharmal SJ, Esbona K, Greaser ML. Titin diversity—alternative splicing gone wild. *J Biomed Biotechnol* 2010: 753675, 2010. doi:10.1155/2010/753675.
120. Joglekar A, Hu W, Zhang B, Narykov O, Diekhans M, Marrocco J, Balacco J, Ndhlovu LC, Milner TA, Fedrigo O, Jarvis ED, Sheynkman G, Korkin D, Ross ME, Tilgner HU. Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. *Nat Neurosci* 27: 1051–1063, 2024. doi:10.1038/s41593-024-01616-4.
121. Aguzzoli Heberle B, Brandon JA, Page ML, Nations KA, Dikobe KI, White BJ, Gordon LA, Fox GA, Wydenworth ME, Doyle PH, Williams BA, Fox EJ, Shantaraman A, Ryten M, Goodwin S, Ghiban E, Wappel R, Mavruk-Eskipehlivan S, Miller JB, Seyfried NT, Nelson PT, Fryer JD, Ebbert MTW. Mapping medically relevant RNA isoform diversity in the aged human frontal cortex with deep long-read RNA-seq. *Nat Biotechnol* 43: 635–646, 2025. doi:10.1038/s41587-024-02245-9.
122. Das AS, Sherry EC, Vaughan RM, Henderson ML, Zieba J, Uhl KL, Koehn O, Bupp CP, Rajasekaran S, Li X, Chhetri SB, Nissim S, Williams CL, Prokop JW. The complex, dynamic SpliceOme of the small GTPase transcripts altered by technique, sex, genetics, tissue specificity, and RNA base editing. *Front Cell Dev Biol* 10: 1033695, 2022. doi:10.3389/fcell.2022.1033695.
123. Prokop JW, Lazar J, Crapitto G, Smith DC, Worthey EA, Jacob HJ. Molecular modeling in the age of clinical genomics, the enterprise of the next generation. *J Mol Model* 23: 75, 2017. doi:10.1007/s00894-017-3258-3.
124. Nelson KK, Green MR. Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proc Natl Acad Sci USA* 87: 6253–6257, 1990. doi:10.1073/pnas.87.16.6253.
125. Kapustin Y, Chan E, Sarkar R, Wong F, Vorechovsky I, Winston RM, Tatusova T, Dibb NJ. Cryptic splice sites and split genes. *Nucleic Acids Res* 39: 5837–5844, 2011. doi:10.1093/nar/gkr203.

126. **Lauffer MC, van Roon-Mom W, Aartsma-Rus A; N = 1 Collaborative.** Possibilities and limitations of antisense oligonucleotide therapies for the treatment of monogenic disorders. *Commun Med (Lond)* 4: 6, 2024. doi:10.1038/s43856-023-00419-1.
127. **Sparber P, Sharova M, Davydenko K, Pyankov D, Filatova A, Skoblov M.** Deciphering the impact of coding and non-coding SCN1A gene variants on RNA splicing. *Brain* 147: 1278–1293, 2024. doi:10.1093/brain/awad383.
128. **Henderson ML, Zieba JK, Li X, Campbell DB, Williams MR, Vogt DL, Bupp CP, Edgerly YM, Rajasekaran S, Hartog NL, Prokop JW, Krueger JM.** Gene therapy for genetic syndromes: understanding the current state to guide future care. *BioTech (Basel)* 13: 1, 2024. doi:10.3390/biotech13010001.
129. **Witten JT, Ule J.** Understanding splicing regulation through RNA splicing maps. *Trends Genet* 27: 89–97, 2011. doi:10.1016/j.tig.2010.12.001.
130. **Zhan X, Lu Y, Shi Y.** Molecular basis for the activation of human spliceosome. *Nat Commun* 15: 6348, 2024. doi:10.1038/s41467-024-50785-0.
131. **Vorländer MK, Rothe P, Kleinfeld J, Cormack ED, Veleti L, Riabov-Bassat D, Fin L, Phillips AW, Cochella L, Plaschka C.** Mechanism for the initiation of spliceosome disassembly. *Nature* 632: 443–450, 2024. doi:10.1038/s41586-024-07741-1.
132. **Rogalska ME, Mancini E, Bonnal S, Gohr A, Dunyak BM, Arecco N, Smith PG, Vaillancourt FH, Valcárcel J.** Transcriptome-wide splicing network reveals specialized regulatory functions of the core spliceosome. *Science* 386: 551–560, 2024. doi:10.1126/science.adn8105.
133. **Zhang W, Zhang X, Zhan X, Bai R, Lei J, Yan C, Shi Y.** Structural insights into human exon-defined spliceosome prior to activation. *Cell Res* 34: 428–439, 2024. doi:10.1038/s41422-024-00949-w.
134. **Fair B, Buen Abad Najjar CF, Zhao J, Lozano S, Reilly A, Mossian G, Staley JP, Wang J, Li Yi.** Global impact of unproductive splicing on human gene expression. *Nat Genet* 56: 1851–1861, 2024. doi:10.1038/s41588-024-01872-x.
135. **Escobar-Hoyos LF, Penson A, Kannan R, Cho H, Pan CH, Singh RK, et al.** Altered RNA splicing by mutant p53 activates oncogenic RAS signaling in pancreatic cancer. *Cancer Cell* 38: 198–211.e8, 2020. doi:10.1016/j.ccell.2020.05.010.
136. **Colgan DF, Manley JL.** Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11: 2755–2766, 1997. doi:10.1101/gad.11.21.2755.
137. **Mitschka S, Mayr C.** Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* 23: 779–796, 2022. doi:10.1038/s41580-022-00507-5.
138. **Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G.** U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468: 664–668, 2010. doi:10.1038/nature09479.
139. **Dubbury SJ, Boutz PL, Sharp PA.** CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 564: 141–145, 2018. doi:10.1038/s41586-018-0758-y.
140. **Yu M, Yang W, Ni T, Tang Z, Nakadai T, Zhu J, Roeder RG.** RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* 350: 1383–1386, 2015. doi:10.1126/science.aad2338.
141. **Lee SH, Mayr C.** Gain of additional BIRC3 protein functions through 3'-UTR-mediated protein complex formation. *Mol Cell* 74: 701–712.e9, 2019. doi:10.1016/j.molcel.2019.03.006.
142. **Chen HF, Hsu CM, Huang YS.** CPEB2-dependent translation of long 3'-UTR Ucp1 mRNA promotes thermogenesis in brown adipose tissue. *EMBO J* 37: e99071, 2018. doi:10.15252/embj.201899071.
143. **Berkovits BD, Mayr C.** Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522: 363–367, 2015. doi:10.1038/nature14321.
144. **Terenzio M, Koley S, Samra N, Rishal I, Zhao Q, Sahoo PK, Urisman A, Marvaldi L, Osés-Prieto JA, Forester C, Gomes C, Kalinski AL, Di Pizio A, Doron-Mandel E, Perry RB, Koppel I, Twiss JL, Burlingame AL, Fainzilber M.** Locally translated mTOR controls axonal local translation in nerve injury. *Science* 359: 1416–1421, 2018. doi:10.1126/science.aan1053.
145. **Torres M, Becquet D, Franc JL, François-Bellan AM.** Circadian processes in the RNA life cycle. *Wiley Interdiscip Rev RNA* 9: e1467, 2018. doi:10.1002/wrna.1467.
146. **Lee SH, Singh I, Tisdale S, Abdel-Wahab O, Leslie CS, Mayr C.** Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 561: 127–131, 2018. doi:10.1038/s41586-018-0465-8.
147. **Reyes A, Huber W.** Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res* 46: 582–592, 2018. doi:10.1093/nar/gkx1165.
148. **Abugessaisa I, Noguchi S, Carninci P, Kasukawa T.** The FANTOM5 computation ecosystem: genomic information hub for promoters and active enhancers. *Methods Mol Biol* 1611: 199–217, 2017. doi:10.1007/978-1-4939-7015-5_15.
149. **Rojas-Duran MF, Gilbert WV.** Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* 18: 2299–2305, 2012. doi:10.1261/rna.035865.112.
150. **Qin Z, Stoilov P, Zhang X, Xing Y.** SEASTAR: systematic evaluation of alternative transcription start sites in RNA. *Nucleic Acids Res* 46: e45, 2018. doi:10.1093/nar/gky053.
151. **Prokop JW, Yeo NC, Ottmann C, Chhetri SB, Florus KL, Ross EJ, Sosonkina N, Link BA, Freedman BI, Coppola CJ, McDermott-Roe C, Leysen S, Milroy LG, Meijer FA, Geurts AM, Rauscher FJ, Ramaker R, Flister MJ, Jacob HJ, Mendenhall EM, Lazar J.** Characterization of coding/noncoding variants for SHROOM3 in patients with CKD. *J Am Soc Nephrol* 29: 1525–1535, 2018. doi:10.1681/ASN.2017080856.
152. **Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al.** A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* 51: 957–972, 2019. doi:10.1038/s41588-019-0407-x.
153. **Kanatsuna N, Taneera J, Vaziri-Sani F, Wierup N, Larsson HE, Delli A, Skärstrand H, Balhuizen A, Bennet H, Steiner DF, Törn C, Fex M, Lernmark Å.** Autoimmunity against INS-IGF2 protein expressed in human pancreatic islets. *J Biol Chem* 288: 29013–29023, 2013. doi:10.1074/jbc.M113.478222.
154. **Cook TW, Wilstermann AM, Mitchell JT, Arnold NE, Rajasekaran S, Bupp CP, Prokop JW.** Understanding insulin in the age of precision medicine and big data: under-explored nature of genomics. *Biomolecules* 13: 257, 2023. doi:10.3390/biom13020257.
155. **Inshaw JRJ, Sidore C, Cucca F, Stefana MI, Crouch DJM, McCarthy MI, Mahajan A, Todd JA.** Analysis of overlapping genetic association in type 1 and type 2 diabetes. *Diabetologia* 64: 1342–1347, 2021. doi:10.1007/s00125-021-05428-0.
156. **Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, Ranson M, Ashford B.** Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. *Brief Bioinform* 22: bbab259, 2021. doi:10.1093/bib/bbab259.
157. **Im Y, Kim Y.** A comprehensive overview of RNA deconvolution methods and their application. *Mol Cells* 46: 99–105, 2023. doi:10.14348/molcells.2023.2178.
158. **Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA.** Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37: 773–782, 2019. doi:10.1038/s41587-019-0114-2.
159. **Garmire LX, Li Y, Huang Q, Xu C, Teichmann SA, Kaminski N, Pellegrini M, Nguyen Q, Teschendorff AE.** Challenges and perspectives in computational deconvolution of genomics data. *Nat Methods* 21: 391–400, 2024. doi:10.1038/s41592-023-02166-6.
160. **Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y.** Single-cell RNA sequencing technologies and applications: a brief overview. *Clin Transl Med* 12: e694, 2022. doi:10.1002/ctm2.694.
161. **Hwang B, Lee JH, Bang D.** Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 50: 1–14, 2018 [Erratum in *Exp Mol Med* 53: 1005, 2021]. doi:10.1038/s12276-018-0071-8.
162. **Choi YH, Kim JK.** Dissecting cellular heterogeneity using single-cell RNA sequencing. *Mol Cells* 42: 189–199, 2019. doi:10.14348/molcells.2019.2446.
163. **Ke M, Elshenawy B, Sheldon H, Arora A, Buffa FM.** Single cell RNA-sequencing: a powerful yet still challenging technology to study cellular heterogeneity. *Bioessays* 44: e2200084, 2022. doi:10.1002/bies.202200084.
164. **Gupta RK, Kuznicki J.** Biological and medical importance of cellular heterogeneity deciphered by single-cell RNA sequencing. *Cells* 9: 1751, 2020. doi:10.3390/cells9081751.

165. Khan SU, Huang Y, Ali H, Ali I, Ahmad S, Khan SU, Hussain T, Ullah M, Lu K. Single-cell RNA sequencing (scRNA-seq): advances and challenges for cardiovascular diseases (CVDs). *Curr Probl Cardiol* 49: 102202, 2024. doi:10.1016/j.cpcardiol.2023.102202.
166. Jain V, Yang WH, Wu J, Roback JD, Gregory SG, Chi JT. Single cell RNA-seq analysis of human red cells. *Front Physiol* 13: 828700, 2022. doi:10.3389/fphys.2022.828700.
167. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol* 14: 479–492, 2018. doi:10.1038/s41581-018-0021-7.
168. Rohlenova K, Goveia J, García-Caballero M, Subramanian A, Kalucka J, Treps L, et al. Single-cell RNA sequencing maps endothelial metabolic plasticity in pathological angiogenesis. *Cell Metab* 31: 862–877.e14, 2020. doi:10.1016/j.cmet.2020.03.009.
169. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Kuo SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587: 619–625, 2020. doi:10.1038/s41586-020-2922-4.
170. Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, Warren A, Rodman C, Roth JA, Bender SA, Kumar B, Rocco JW, Fernandes PACM, Mader CC, Keren-Shaul H, Plotnikov A, Barr H, Tsherniak A, Rozenblatt-Rosen O, Krizhanovsky V, Puram SV, Regev A, Tirosh I. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 52: 1208–1218, 2020. doi:10.1038/s41588-020-00726-6.
171. Taheri-Araghi S, Brown SD, Sauls JT, McIntosh DB, Jun S. Single-cell physiology. *Annu Rev Biophys* 44: 123–142, 2015. doi:10.1146/annurev-biophys-060414-034236.
172. Chang X, Zheng Y, Xu K. Single-cell RNA sequencing: technological progress and biomedical application in cancer research. *Mol Biotechnol* 66: 1497–1519, 2024. doi:10.1007/s12033-023-00777-0.
173. Collí-Dulá RC, Papatheodorou I. Single-cell RNA sequencing offers opportunities to explore the depth of physiology, adaptation, and biochemistry in non-model organisms exposed to pollution. *Comp Biochem Physiol Part D Genomics Proteomics* 52: 101339, 2024. doi:10.1016/j.cbd.2024.101339.
174. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. *Science* 358: 58–63, 2017. doi:10.1126/science.aan6828.
175. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 18: 35–45, 2018. doi:10.1038/nri.2017.76.
176. Dash P, McClaren JL, Oguin TH, Rothwell W, Todd B, Morris MY, Becksfort J, Reynolds C, Brown SA, Doherty PC, Thomas PG. Paired analysis of TCR α and TCR β chains at the single-cell level in mice. *J Clin Invest* 121: 288–295, 2011. doi:10.1172/JCI44752.
177. Jia Q, Chu H, Jin Z, Long H, Zhu B. High-throughput single-cell sequencing in cancer research. *Signal Transduct Target Ther* 7: 145, 2022. doi:10.1038/s41392-022-00990-4.
178. Jia G, Preussner J, Chen X, Guenther S, Yuan X, Yekelchik M, Kuenne C, Looso M, Zhou Y, Teichmann S, Braun T. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat Commun* 9: 4877, 2018. doi:10.1038/s41467-018-07307-6.
179. Kim SS, Truong B, Jagadeesh K, Dey KK, Shen AZ, Raychaudhuri S, Kellis M, Price AL. Leveraging single-cell ATAC-seq and RNA-seq to identify disease-critical fetal and adult brain cell types. *Nat Commun* 15: 563, 2024. doi:10.1038/s41467-024-44742-0.
180. Camunas-Soler J. Integrating single-cell transcriptomics with cellular phenotypes: cell morphology, Ca²⁺ imaging and electrophysiology. *Biophys Rev* 16: 89–107, 2024. doi:10.1007/s12551-023-01174-2.
181. Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, Yilmaz M, Reimer J, Shen S, Bethge M, Tolias KF, Sandberg R, Tolias AS. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* 34: 199–203, 2016. doi:10.1038/nbt.3445.
182. Zhang MJ, Ntranos V, Tse D. Determining sequencing depth in a single-cell RNA-seq experiment. *Nat Commun* 11: 774, 2020. doi:10.1038/s41467-020-14482-y.
183. Baccarella A, Williams CR, Parrish JZ, Kim CC. Empirical assessment of the impact of sample number and read depth on RNA-seq analysis workflow performance. *BMC Bioinformatics* 19: 423, 2018. doi:10.1186/s12859-018-2445-2.
184. Linderman GC, Zhao J, Roulis M, Bielecki P, Flavell RA, Nadler B, Kluger Y. Zero-preserving imputation of single-cell RNA-seq data. *Nat Commun* 13: 192, 2022. doi:10.1038/s41467-021-27729-z.
185. Wu X, Yang X, Dai Y, Zhao Z, Zhu J, Guo H, Yang R. Single-cell sequencing to multi-omics: technologies and applications. *Biomark Res* 12: 110, 2024. doi:10.1186/s40364-024-00643-4.
186. Arzalluz-Luque Á, Conesa A. Single-cell RNAseq for the study of isoforms-how is that possible? *Genome Biol* 19: 110, 2018. doi:10.1186/s13059-018-1496-z.
187. Kabza M, Ritter A, Byrne A, Sereti K, Le D, Stephenson W, Sterne-Weiler T. Accurate long-read transcript discovery and quantification at single-cell, pseudo-bulk and bulk resolution with isosceles. *Nat Commun* 15: 7316, 2024. doi:10.1038/s41467-024-51584-3.
188. Huang S, Shi W, Li S, Fan Q, Yang C, Cao J, Wu L. Advanced sequencing-based high-throughput and long-read single-cell transcriptome analysis. *Lab Chip* 24: 2601–2621, 2024. doi:10.1039/d4lc00105b.
189. Gulati GS, D'Silva JP, Liu Y, Wang L, Newman AM. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat Rev Mol Cell Biol* 26: 11–31, 2025. doi:10.1038/s41580-024-00768-2.
190. Conrad T, Altmüller J. Single cell- and spatial 'Omics revolutionize physiology. *Acta Physiol (Oxf)* 235: e13848, 2022. doi:10.1111/apha.13848.
191. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 19: 534–546, 2022 [Erratum in *Nat Methods* 19: 628, 2022]. doi:10.1038/s41592-022-01409-2.
192. Moor AE, Itzkovitz S. Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr Opin Biotechnol* 46: 126–133, 2017. doi:10.1016/j.copbio.2017.02.004.
193. Jain S, Eadon MT. Spatial transcriptomics in health and disease. *Nat Rev Nephrol* 20: 659–671, 2024. doi:10.1038/s41581-024-00841-1.
194. Boileau E, Li X, Naarmann-de Vries IS, Becker C, Casper R, Altmüller J, Leuschner F, Dieterich C. Full-length spatial transcriptomics reveals the unexplored isoform diversity of the myocardium post-MI. *Front Genet* 13: 912572, 2022. doi:10.3389/fgene.2022.912572.
195. Dixon EE, Wu H, Sulvarán-Guel E, Guo J, Humphreys BD. Spatially resolved transcriptomics and the kidney: many opportunities. *Kidney Int* 102: 482–491, 2022. doi:10.1016/j.kint.2022.06.011.
196. Coulis G, Jaime D, Guerrero-Juarez C, Kastenschmidt JM, Farahat PK, Nguyen Q, et al. Single-cell and spatial transcriptomics identify a macrophage population associated with skeletal muscle fibrosis. *Sci Adv* 9: eadd9984, 2023. doi:10.1126/sciadv.add9984.
197. Matsumoto R, Yamamoto T. Single-cell and spatial transcriptomics in endocrine research. *Endocr J* 71: 101–118, 2024. doi:10.1507/endocrj.EJ23-0457.
198. Schepps S, Xu J, Yang H, Mandel J, Mehta J, Tolotta J, Baker N, Tekmen V, Nikbakht N, Fortina P, Fuentes I, LaFleur B, Cho RJ, South AP. Skin in the game: a review of single-cell and spatial transcriptomics in dermatological research. *Clin Chem Lab Med* 62: 1880–1891, 2024. doi:10.1515/cclm-2023-1245.
199. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* 358: 64–69, 2017. doi:10.1126/science.aan6827.
200. Piwecka M, Rajewsky N, Rybak-Wolf A. Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nat Rev Neurol* 19: 346–362, 2023. doi:10.1038/s41582-023-00809-y.
201. Megas S, Wilbrey-Clark A, Maartens A, Teichmann SA, Meyer KB. Spatial transcriptomics of the respiratory system. *Annu Rev Physiol* 87: 447–470, 2025. doi:10.1146/annurev-physiol-022724-105144.
202. You Y, Fu Y, Li L, Zhang Z, Jia S, Lu S, Ren W, Liu Y, Xu Y, Liu X, Jiang F, Peng G, Sampath Kumar A, Ritchie ME, Liu X, Tian L. Systematic comparison of sequencing-based spatial transcriptomic methods. *Nat Methods* 21: 1743–1754, 2024. doi:10.1038/s41592-024-02325-3.
203. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J, Sadowski N, Holmes N, de Jesus JG, Jones KL, Soulette CM, Snutch TP, Loman N, Paten B, Loose M, Simpson JT, Olsen HE, Brooks AN, Akeson M, Timp W. Nanopore native RNA sequencing of a human poly(A)

- transcriptome. *Nat Methods* 16: 1297–1305, 2019 [Erratum in *Nat Methods* 17: 114, 2020]. doi:10.1038/s41592-019-0617-2.
204. O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol* 4: Unit 4.19, 2013. doi:10.1002/0471142727.mb0419s103.
205. Choe D, Szubin R, Poudel S, Sastry A, Song Y, Lee Y, Cho S, Palsson B, Cho BK. RiboRid: a low cost, advanced, and ultra-efficient method to remove ribosomal RNA for bacterial transcriptomics. *PLoS Genet* 17: e1009821, 2021. doi:10.1371/journal.pgen.1009821.
206. Singh A, Xue A, Tai J, Mbadugha F, Obi P, Mascarenhas R, Tyagi A, Siena A, Chen YG. A scalable and cost-efficient rRNA depletion approach to enrich RNAs for molecular biology investigations. *RNA* 30: 728–738, 2024. doi:10.1261/rna.079761.123.
207. Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY. Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* 178: 473–490.e26, 2019. doi:10.1016/j.cell.2019.05.027.
208. Pearce SF, Cipullo M, Chung B, Brierley I, Rorbach J. Mitochondrial translation from human cell culture: a high resolution view of mitochondrial translation. *Methods Mol Biol* 2192: 183–196, 2021. doi:10.1007/978-1-0716-0834-0_14.
209. Schott J, Reitter S, Lindner D, Grosser J, Bruer M, Shenoy A, Geiger T, Mathes A, Dobrev G, Stoecklin G. Publisher Correction: Nascent Ribo-Seq measures ribosomal loading time and reveals kinetic impact on ribosome density. *Nat Methods* 20: 316, 2023. doi:10.1038/s41592-023-01765-7.
210. Xiong Z, Xu K, Lin Z, Kong F, Wang Q, Quan Y, Sha QQ, Li F, Zou Z, Liu L, Ji S, Chen Y, Zhang H, Fang J, Yu G, Liu B, Wang L, Wang H, Deng H, Yang X, Fan HY, Li L, Xie W. Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol* 24: 968–980, 2022. doi:10.1038/s41556-022-00928-6.
211. VanInsberghe M, van den Berg J, Andersson-Rolf A, Clevers H, van Oudenaarden A. Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature* 597: 561–565, 2021. doi:10.1038/s41586-021-03887-4.
212. Steward O, Banker GA. Getting the message from the gene to the synapse: sorting and intracellular transport of RNA in neurons. *Trends Neurosci* 15: 180–186, 1992. doi:10.1016/0166-2236(92)90170-d.
213. Zappulo A, van den Bruck D, Ciolli Mattioli C, Franke V, Imami K, McShane E, Moreno-Estelles M, Calviello L, Filipchuk A, Peguero-Sanchez E, Müller T, Woehler A, Birchmeier C, Merino E, Rajewsky N, Ohler U, Mazzoni EO, Selbach M, Akalin A, Cherkulaeva M. RNA localization is a key determinant of neurite-enriched proteome. *Nat Commun* 8: 583, 2017. doi:10.1038/s41467-017-00690-6.
214. Arora A, Castro-Gutierrez R, Moffatt C, Eletto D, Becker R, Brown M, Moor AE, Russ HA, Taliaferro JM. High-throughput identification of RNA localization elements in neuronal cells. *Nucleic Acids Res* 50: 10626–10642, 2022. doi:10.1093/nar/gkac763.
215. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, Clement T, Toohey-Kurth K. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Invest* 32: 767–775, 2020. doi:10.1177/1040638720905833.
216. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R, Albertsen M. High-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or PacBio sequencing. *Nat Methods* 18: 165–169, 2021. doi:10.1038/s41592-020-01041-y.
217. Murigneux V, Rai SK, Furtado A, Bruxner TJC, Tian W, Harliwong I, Wei H, Yang B, Ye Q, Anderson E, Mao Q, Drmanac R, Wang O, Peters BA, Xu M, Wu P, Topp B, Coin LJM, Henry RJ. Comparison of long-read methods for sequencing and assembly of a plant genome. *Gigascience* 9: g1aa146, 2020. doi:10.1093/gigascience/g1aa146.
218. Zhang T, Zhou J, Gao W, Jia Y, Wei Y, Wang G. Complex genome assembly based on long-read sequencing. *Brief Bioinform* 23: bbac305, 2022. doi:10.1093/bib/bbac305.
219. Frank JA, Pan Y, Tooming-Klunderud A, Eijssink VGH, McHardy AC, Nederbragt AJ, Pope PB. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Sci Rep* 6: 25373, 2016. doi:10.1038/srep25373.
220. Xie H, Yang C, Sun Y, Igarashi Y, Jin T, Luo F. PacBio long reads improve metagenomic assemblies, gene catalogs, and genome binning. *Front Genet* 11: 516269, 2020. doi:10.3389/fgene.2020.516269.
221. Shi ZX, Chen ZC, Zhong JY, Hu KH, Zheng YF, Chen Y, Xie SQ, Bo XC, Luo F, Tang C, Xiao CL, Liu YZ. High-throughput and high-accuracy single-cell RNA isoform analysis using PacBio circular consensus sequencing. *Nat Commun* 14: 2631, 2023. doi:10.1038/s41467-023-38324-9.
222. Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. *BMC Genomics* 25: 528, 2024. doi:10.1186/s12864-024-10440-w.
223. Tan L, Guo Z, Wang X, Kim DY, Li R. Utilization of nanopore direct RNA sequencing to analyze viral RNA modifications. *mSystems* 9: e0116323, 2024. doi:10.1128/msystems.01163-23.
224. Maestri S, Furlan M, Mulrone L, Coscujuela Tarrero L, Ugolini C, Dalla Pozza F, Leonardi T, Birney E, Nicassio F, Pelizzola M. Benchmarking of computational methods for m6A profiling with nanopore direct RNA sequencing. *Brief Bioinform* 25: bbae001, 2024. doi:10.1093/bib/bbae001.
225. Harcourt EM, Kietrys AM, Kool ET. Chemical and structural effects of base modifications in messenger RNA. *Nature* 541: 339–346, 2017. doi:10.1038/nature21351.
226. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* 39: D195–D201, 2011. doi:10.1093/nar/gkq1028.
227. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA modifications in gene expression regulation. *Cell* 169: 1187–1200, 2017. doi:10.1016/j.cell.2017.05.045.
228. Kokoris M, McRuer R, Nabavi M, Jacobs A, Prindle M, Cech C, et al. Sequencing by Expansion (SBX) – a novel, high-throughput single-molecule sequencing technology (Preprint). *bioRxiv* 2025.02.19.639056, 2025. doi:10.1101/2025.02.19.639056.
229. Prokop JW, Alberta S, Witteveen-Lane M, Pell S, Farag HA, Bhargava D, Vaughan RM, Frisch A, Bauss J, Bhatti H, Arora S, Subrahmanya C, Pearson D, Goodyke A, Westgate M, Cook TW, Mitchell JT, Zieba J, Sims MD, Underwood A, Hassouna H, Rajasekaran S, Tamae Kakazu MA, Chesla D, Olivero R, Caulfield AJ. SARS-CoV-2 genotyping highlights the challenges in spike protein drift independent of other essential proteins. *Microorganisms* 12: 1863, 2024. doi:10.3390/microorganisms12091863.
230. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, Chen J. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 21: 12, 2020. doi:10.1186/s13059-019-1850-9.
231. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res* 42: e161, 2014. doi:10.1093/nar/gku864.
232. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2: lqaa078, 2020. doi:10.1093/nargab/lqaa078.
233. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 34: D173–D180, 2006. doi:10.1093/nar/gkj158.
234. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE; Mouse Genome Database Group. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res* 47: D801–D806, 2019. doi:10.1093/nar/gky1056.
235. Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork: a tool-box for systems genetics. *Methods Mol Biol* 1488: 75–120, 2017. doi:10.1007/978-1-4939-6427-7_4.
236. Baldarelli RM, Smith CL, Ringwald M, Richardson JE, Bult CJ; Mouse Genome Informatics Group. Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* 227: iyae031, 2024. doi:10.1093/genetics/iyae031.
237. Smith JR, Tutaj MA, Thota J, Lamers L, Gibson AC, Kundurthi A, Gollapally VR, Brodie KC, Zacher S, Laulederkind SJF, Hayman GT, Wang SJ, Tutaj M, Kaldunski ML, VEDI M, Demos WM, De Pons JL, Dwinell MR, Kwitek AE. Standardized pipelines support

- and facilitate integration of diverse datasets at the Rat Genome Database. *Database (Oxford)* 2025: baee132, 2025. doi:10.1093/database/baee132.
238. Bradford YM, Van Slyke CE, Muyskens JB, Tseng WC, Howe DG, Fashena D, Martin R, Paddock H, Pich C, Ramachandran S, Ruzicka L, Singer A, Taylor R, Westerfield M. ZFIN updates to support zebrafish environmental exposure data. *Genetics* 229: iyaf021, 2025. doi:10.1093/genetics/iyaf021.
239. Sternberg PW, Van Auken K, Wang Q, Wright A, Yook K, Zarowiecki M, Arnaboldi V, Becerra A, Brown S, Cain S, Chan J, Chen WJ, Cho J, Davis P, Diamantakis S, Dyer S, Grigoriadis D, Grove CA, Harris T, Howe K, Kishore R, Lee R, Longden I, Luypaert M, Müller HM, Nuin P, Quinton-Tulloch M, Raciti D, Schedl T, Schindelman G, Stein L. WormBase 2024: status and transitioning to Alliance infrastructure. *Genetics* 227: iyae050, 2024. doi:10.1093/genetics/iyae050.
240. Fisher M, James-Zorn C, Ponferrada V, Bell AJ, Sundararaj N, Segerdell E, Chaturvedi P, Bayyari N, Chu S, Pells T, Lotay V, Agalakov S, Wang DZ, Arshinoff BI, Foley S, Karimi K, Vize PD, Zorn AM. Xenbase: key features and resources of the Xenopus model organism knowledgebase. *Genetics* 224: iyad018, 2023. doi:10.1093/genetics/iyad018.
241. Alliance of Genome Resources Consortium. Updates to the Alliance of Genome Resources central infrastructure. *Genetics* 227: iyae049, 2024. doi:10.1093/genetics/iyae049.
242. Goodyke A, Gaweda B, Piekarska M, Arora S, Westgate M, Loyaga-Rendon R, Jani M, Rausch MK, Aguirre A, Prokop JW, Timek TA. The tricuspid valve is transcriptionally active during prolonged pressure overload, right-sided heart failure, and valve regurgitation (Preprint). *bioRxiv* 2025.07.21.666043, 2025. doi:10.1101/2025.07.21.666043.
243. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8: 1494–1512, 2013. doi:10.1038/nprot.2013.084.
244. Gaweda B, Goodyke A, Prokop J, Arora S, Iwasieczko A, Piekarska M, Zagorski J, Widenka K, Rausch MK, Aguirre A, Timek TA. Differential gene expression of leaflet tissue in chronic ovine functional tricuspid regurgitation. *Eur J Cardiothorac Surg* ezae218, 2024. doi:10.1093/ejcts/ezae218.
245. Prokop JW, Schmidt C, Gasper D, Duff RJ, Milsted A, Ohkubo T, Ball HC, Shawkey MD, Mays HL, Cogburn LA, Londraville RL. Discovery of the elusive leptin in birds: identification of several “missing links” in the evolution of leptin and its receptor. *PLoS One* 9: e92751, 2014. doi:10.1371/journal.pone.0092751.
246. Prokop JW, Chhetri SB, van Veen JE, Chen X, Underwood AC, Uhl K, Dwinell MR, Geurts AM, Correa SM, Arnold AP. Transcriptional analysis of the multiple Sry genes and developmental program at the onset of testis differentiation in the rat. *Biol Sex Differ* 11: 28, 2020. doi:10.1186/s13293-020-00305-8.
247. Prokop JW, Tsaih S-W, Faber AB, Boehme S, Underwood AC, Troyer S, Playl L, Milsted A, Turner ME, Ely D, Martins AS, Tutaj M, Lazar J, Dwinell MR, Jacob HJ. The phenotypic impact of the male-specific region of chromosome-Y in inbred mating: the role of genetic variants and gene duplications in multiple inbred rat strains. *Biol Sex Differ* 7: 10, 2016. doi:10.1186/s13293-016-0064-z.
248. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, Rehm HL, MacArthur DG, O'Donnell-Luria A; Genome Aggregation Database Consortium. Variant interpretation using population databases: lessons from gnomAD. *Hum Mutat* 43: 1012–1030, 2022. doi:10.1002/humu.24309.
249. Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet* 21: 448, 2020. doi:10.1038/s41576-020-0255-7.
250. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 625: 92–100, 2024 [Erratum in *Nature* 626: E1, 2024]. doi:10.1038/s41586-023-06045-0.
251. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369: 1318–1330, 2020. doi:10.1126/science.aaz1776.
252. Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group., Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration & Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis & Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group. Genetic effects on gene expression across human tissues. *Nature* 550: 204–213, 2017. doi:10.1038/nature24277.
253. Snyder MP, Gingeras TR, Moore JE, Weng Z, Gerstein MB, Ren B, Hardison RC, Stamatoyannopoulos JA, Graveley BR, Feingold EA, Pazin MJ, Pagan M, Gilchrist DA, Hitz BC, Cherry JM, Bernstein BE, Mendenhall EM, Zerbino DR, Frankish A, Flicek P, Myers RM; ENCODE Project Consortium. Perspectives on ENCODE. *Nature* 583: 693–698, 2020 [Erratum in *Nature* 605: E4, 2022]. doi:10.1038/s41586-020-2449-8.
254. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330, 2015. doi:10.1038/nature14248.
255. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47: D886–D894, 2019. doi:10.1093/nar/gky1016.
256. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, Patwardhan TA, Kanai M, Nasser J, Fulco CP, Tashman KC, Aguet F, Li T, Ordovas-Montanes J, Smillie CS, Biton M, Shalek AK, Ananthakrishnan AN, Xavier RJ, Regev A, Gupta RM, Lage K, Ardlie KG, Hirschhorn JN, Lander ES, Engreitz JM, Finucane HK. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat Genet* 55: 1267–1276, 2023. doi:10.1038/s41588-023-01443-6.
257. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. *Science* 356: eaal3321, 2017. doi:10.1126/science.aal3321.
258. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Björling L, Ponten F. Towards a knowledge-based human protein atlas. *Nat Biotechnol* 28: 1248–1250, 2010. doi:10.1038/nbt1210-1248.
259. Uhlen M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* 347: 1260419, 2015. doi:10.1126/science.1260419.
260. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, Oksvold P, Edfors F, Limiszewska A, Hikmet F, Huang J, Du Y, Lin L, Dong Z, Yang L, Liu X, Jiang H, Xu X, Wang J, Yang H, Bolund L, Mardinoglu A, Zhang C, von Feilitzen K, Lindskog C, Pontén F, Luo Y, Hökfelt T, Uhlen M, Mulder J. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* 367: eaay5947, 2020. doi:10.1126/science.aay5947.
261. Uhlen M, Karlsson MJ, Zhong W, Tebani A, Pou C, Mikes J, Lakshmikanth T, Forsström B, Edfors F, Odeberg J, Mardinoglu A, Zhang C, von Feilitzen K, Mulder J, Sjöstedt E, Hober A, Oksvold P, Zwahlen M, Pontén F, Lindskog C, Sivertsson Å, Fagerberg L, Brodin P. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* 366: eaax9198, 2019. doi:10.1126/science.aax9198.
262. Uhlen M, Karlsson MJ, Hober A, Svensson A-S, Scheffell J, Kotol D, et al. The human secretome. *Sci Signal* 12: eaaz0274, 2019. doi:10.1126/scisignal.aaz0274.
263. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, Sjöstedt E, Butler L, Odeberg J, Dusart P, Edfors F, Oksvold P, von Feilitzen K, Zwahlen M, Arif M, Altay O, Li X, Özcan M, Mardinoglu A, Fagerberg L, Mulder J, Luo Y, Pontén F, Uhlen M, Lindskog C. A single-cell type transcriptomics map of human tissues. *Sci Adv* 7: eabh2169, 2021. doi:10.1126/sciadv.abh2169.

264. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, Benfais R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Mücke P, Lindskog C, Mardinoglu A, Ponten F. A pathology atlas of the human cancer transcriptome. *Science* 357: ean2507, 2017. doi:10.1126/science.aan2507.
265. Jin H, Zhang C, Zwahlen M, von Feilitzen K, Karlsson M, Shi M, Yuan M, Song X, Li X, Yang H, Turkez H, Fagerberg L, Uhlén M, Mardinoglu A. Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nat Commun* 14: 5417, 2023. doi:10.1038/s41467-023-41132-w.
266. Jones AR, Overly CC, Sunkin SM. The Allen Brain Atlas: 5 years and beyond. *Nat Rev Neurosci* 10: 821–828, 2009. doi:10.1038/nrn2722.
267. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, Hawrylycz M, Dang C. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* 41: D996–D1008, 2013. doi:10.1093/nar/gks1042.
268. Feng G, Wang Y, Huang W, Chen H, Cheng J, Shu N. Spatial and temporal pattern of structure-function coupling of human brain connectome with development. *eLife* 13: RP93325, 2024. doi:10.7554/eLife.93325.
269. Zhang XH, Anderson KM, Dong HM, Chopra S, Dhamala E, Emami PS, Gerstein MB, Margulies DS, Holmes AJ. The cell-type underpinnings of the human functional cortical connectome. *Nat Neurosci* 28: 150–160, 2025. doi:10.1038/s41593-024-01812-2.
270. Abdulla S, Aevertmann B, Assis P, Badajoz S, Bell SM, Bezzi E, et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res* 53: D886–D900, 2025. doi:10.1093/nar/gkae1142.
271. Darwiche R, Struhl K. Pheno-RNA, a method to associate genes with a specific phenotype, identifies genes linked to cellular transformation. *Proc Natl Acad Sci USA* 117: 28925–28929, 2020. doi:10.1073/pnas.2014165117.
272. Koller BH, Marrack P, Kappler JW, Smithies O. Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8⁺ T cells. *Science* 248: 1227–1230, 1990. doi:10.1126/science.2112266.
273. Groza T, Gomez FL, Mashhadi HH, Muñoz-Fuentes V, Gunes O, Wilson R, Cacheiro P, Frost A, Kesikvali-Bond P, Vardal B, McCoy A, Cheng TK, Santos L, Wells S, Smedley D, Mallon AM, Parkinson H. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res* 51: D1038–D1045, 2023. doi:10.1093/nar/gkac972.
274. Vedi M, Smith JR, Thomas Hayman G, Tutaj M, Brodie KC, De Pons JL, Demos WM, Gibson AC, Kaldunski ML, Lamers L, Laulederkind SJF, Thota J, Thorat K, Tutaj MA, Wang S-J, Zacher S, Dwinell MR, Kwitek AE. 2022 updates to the Rat Genome Database: a Findable, Accessible, Interoperable, and Reusable (FAIR) resource. *Genetics* 224: iyad042, 2023. doi:10.1093/genetics/iyad042.
275. Prokop JW, Jdanov V, Savage L, Morris M, Lamb N, VanSickle E, Stenger CL, Rajasekaran S, Bupp CP. Computational and experimental analysis of genetic variants. *Compr Physiol* 12: 3303–3336, 2022. doi:10.1002/cphy.c210012.
276. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 39: D945–D950, 2011. doi:10.1093/nar/gkq929.
277. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington ZM, Welter D, Burdett T, Hindorf L, Flicek P, Cunningham F, Parkinson H. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 45: D896–D901, 2017. doi:10.1093/nar/gkw1133.
278. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res* 49: D1311–D1320, 2021. doi:10.1093/nar/gkaa840.
279. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, Carmona M, Faulconbridge A, Hercules A, McAuley E, Miranda A, Peat G, Spitzer M, Barrett J, Hulcoop DG, Papa E, Koscielny G, Dunham I. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res* 47: D1056–D1065, 2019. doi:10.1093/nar/gky1133.
280. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45: D362–D368, 2017. doi:10.1093/nar/gkw937.
281. Shrine N, Izquierdo AG, Chen J, Packer R, Hall RJ, Guyatt AL, et al. Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nat Genet* 55: 410–422, 2023 [Erratum in *Nat Genet* 55: 1778–1779, 2023]. doi:10.1038/s41588-023-01314-0.
282. Castaldi PJ, Sauler M. Molecular characterization of the distal lung: novel insights from chronic obstructive pulmonary disease omics. *Am J Respir Crit Care Med* 210: 147–154, 2024. doi:10.1164/rccm.202310-1972PP.
283. Bauss J, Morris M, Shankar R, Olivero R, Buck LN, Stenger CL, Hinds D, Mills J, Eby A, Zagorski JW, Smith C, Cline S, Hartog NL, Chen B, Harris J, Carcillo JA, Rajasekaran S, Bupp CP, Prokop JW. CCR5 and biological complexity: the need for data integration and educational materials to address genetic/biological reductionism at the interface of ethical, legal, and social implications. *Front Immunol* 12: 790041, 2021. doi:10.3389/fimmu.2021.790041.
284. Prokop JW. Biological sex: why we should listen to genetics and lessons from race to guide precision medicine. *Physiol Genomics* 57: 598–602, 2025. doi:10.1152/physiolgenomics.00065.2025.
285. San Roman AK, Godfrey AK, Skaletsky H, Bellott DW, Groff AF, Harris HL, Blanton LV, Hughes JF, Brown L, Phou S, Buscetta A, Kruszka P, Banks N, Dutra A, Pak E, Lasutschinkow PC, Keen C, Davis SM, Tartaglia NR, Samango-Sprouse C, Muenke M, Page DC. The human inactive X chromosome modulates expression of the active X chromosome. *Cell Genom* 3: 100259, 2023. doi:10.1016/j.xgen.2023.100259.
286. Blanton LV, San Roman AK, Wood G, Buscetta A, Banks N, Skaletsky H, Godfrey AK, Pham TT, Hughes JF, Brown LG, Kruszka P, Lin AE, Kastner DL, Muenke M, Page DC. Stable and robust Xi and Y transcripts drive cell-type-specific autosomal and Xa responses in vivo and in vitro in four human cell types. *Cell Genom* 4: 100628, 2024. doi:10.1016/j.xgen.2024.100628.
287. Prokop JW, Deschepper CF. Chromosome Y genetic variants: impact in animal models and on human disease. *Physiol Genomics* 47: 525–537, 2015. doi:10.1152/physiolgenomics.00074.2015.
288. Tomofuji Y, Edahiro R, Sonehara K, Shirai Y, Kock KH, Wang QS, et al. Quantification of escape from X chromosome inactivation with single-cell omics data reveals heterogeneity across cell types and tissues. *Cell Genom* 4: 100625, 2024. doi:10.1016/j.xgen.2024.100625.
289. Gylemo B, Bensberg M, Hennings V, Lundqvist C, Camponeschi A, Goldmann D, Zhang H, Selimović-Pašić A, Lentini A, Ekwall O, Nestor CE. A landscape of X-inactivation during human T cell development. *Nat Commun* 15: 10527, 2024. doi:10.1038/s41467-024-54110-7.
290. Bianchi DW, Zickwolf GK, Weil GJ, Sylvester S, DeMaria MA. Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc Natl Acad Sci USA* 93: 705–708, 1996. doi:10.1073/pnas.93.2.705.
291. Berglund A, Stochholm K, Gravholt CH. The epidemiology of sex chromosome abnormalities. *Am J Med Genet C Semin Med Genet* 184: 202–215, 2020. doi:10.1002/ajmg.c.31805.
292. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Lappalainen T, Regev A, Ardlie KG, Hacohen N, MacArthur DG. Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz. Landscape of X chromosome inactivation across human tissues. *Nature* 550: 244–248, 2017 [Erratum in *Nature* 555: 274, 2018]. doi:10.1038/nature24265.
293. Morris O, Morris M, Jobe S, Bhargava D, Krueger JM, Arora S, Prokop JW, Stenger C. Genomic landscape of chromosome X factor VIII: from hemophilia A in males to risk variants in females. *Genes (Basel)* 15: 1522, 2024. doi:10.3390/genes15121522.

294. Harris RM, Whitfield T, Blanton LV, Skaletsky H, Blumen K, Hyland P, McDermott E, Summers K, Hughes JF, Jackson E, Teglas P, Liu B, Chan YM, Page DC. Independent effects of testosterone, estradiol, and sex chromosomes on gene expression in immune cells of trans- and cisgender individuals (Preprint). *bioRxiv* 2024.10.08.617275, 2024. doi:10.1101/2024.10.08.617275.
295. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20: 257, 2019. doi:10.1186/s13059-019-1891-0.
296. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 12: 380–381, 2015. doi:10.1038/nmeth.3364.
297. Prokop JW, Shankar R, Gupta R, Leimanis ML, Nedveck D, Uhl K, et al. Virus-induced genetics revealed by multidimensional precision medicine transcriptional workflow applicable to COVID-19. *Physiol Genomics* 52: 255–268, 2020. doi:10.1152/physiolgenomics.00045.2020.
298. Shankar R, Leimanis ML, Newbury PA, Liu K, Xing J, Nedveck D, Kort EJ, Prokop JW, Zhou G, Bachmann AS, Chen B, Rajasekaran S. Gene expression signatures identify paediatric patients with multiple organ dysfunction who require advanced life support in the intensive care unit. *EBioMedicine* 62: 103122, 2020. doi:10.1016/j.ebiom.2020.103122.
299. Gupta R, Leimanis ML, Adams M, Bachmann AS, Uhl KL, Bupp CP, Hartog NL, Kort EJ, Olivero R, Comstock SS, Sanfilippo DJ, Lunt SY, Prokop JW, Rajasekaran S. Balancing precision versus cohort transcriptomic analysis of acute and recovery phase of viral bronchiolitis. *Am J Physiol Lung Cell Mol Physiol* 320: L1147–L1157, 2021. doi:10.1152/ajplung.00440.2020.
300. Prokop JW, Hartog NL, Chesla D, Faber W, Love CP, Karam R, Abualkheir N, Feldmann B, Teng L, McBride T, Leimanis ML, English BK, Holsworth A, Frisch A, Bauss J, Kalpage N, Derbedrossian A, Pinti RM, Hale N, Mills J, Eby A, VanSickle EA, Pageau SC, Shankar R, Chen B, Carcillo JA, Sanfilippo D, Olivero R, Bupp CP, Rajasekaran S. High-density blood transcriptomics reveals precision immune signatures of SARS-CoV-2 infection in hospitalized individuals. *Front Immunol* 12: 694243, 2021. doi:10.3389/fimmu.2021.694243.
301. Hartog N, Faber W, Frisch A, Bauss J, Bupp CP, Rajasekaran S, Prokop JW. SARS-CoV-2 infection: molecular mechanisms of severe outcomes to suggest therapeutics. *Expert Rev Proteomics* 18: 105–118, 2021. doi:10.1080/14789450.2021.1908894.
302. Bhargava D, Labadie A, Hanson-Rios-Stutz RL, Goodyke A, Moses EM, Das AS, Vanderweele S, Lemon JV, Cook TW, Pearson D, Redinger JM, Caulfield AJ, Olivero R, Foster K, Ashack K, Rajasekaran S, Bupp CP, Triche TJ, Krawczyk CM, Chesla D, Sims MD, Hartog NL, Prokop JW. From castaways to discoveries: unveiling treasures in skin rnaseq using a novel multidimensional data processing workflow including infection-host dynamics. *Physiol Genomics* 57: 343–356, 2025. doi:10.1152/physiolgenomics.00093.2024.
303. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* 17: 257–271, 2016. doi:10.1038/nrg.2016.10.
304. Gutierrez Reyes CD, Alejo-Jacuinde G, Perez Sanchez B, Chavez Reyes J, Onigbinde S, Mogut D, Hernández-Jasso I, Calderón-Vallejo D, Quintanar JL, Mechref Y. Multi omics applications in biological systems. *Curr Issues Mol Biol* 46: 5777–5793, 2024. doi:10.3390/cimb46060345.
305. Liu X, Peng T, Xu M, Lin S, Hu B, Chu T, Liu B, Xu Y, Ding W, Li L, Cao C, Wu P. Spatial multi-omics: deciphering technological landscape of integration of multi-omics and its applications. *J Hematol Oncol* 17: 72, 2024. doi:10.1186/s13045-024-01596-9.
306. Wishart D. Metabolomics and the multi-omics view of cancer. *Metabolites* 12: 154, 2022. doi:10.3390/metabo12020154.
307. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc* 17: 1518–1552, 2022. doi:10.1038/s41596-022-00692-9.
308. Guo J, Grow EJ, Yi C, Mlcochova H, Maher GJ, Lindskog C, Murphy PJ, Wike CL, Carrell DT, Goriely A, Hotaling JM, Cairns BR. Chromatin and single-cell RNA-Seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* 21: 533–546.e6, 2017. doi:10.1016/j.stem.2017.09.003.
309. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 14: 411–413, 2017. doi:10.1038/nmeth.4189.
310. Doane AS, Elemento O. Regulatory elements in molecular networks. *Wiley Interdiscip Rev Syst Biol Med* 9: 10.1002/wsbm.1374, 2017. doi:10.1002/wsbm.1374.
311. Ponomarenko EA, Krasnov GS, Kiseleva OI, Kryukova PA, Arzumanyan VA, Dolgalev GV, Ilgisonis EV, Lisitsa AV, Poverennaya EV. Workability of mRNA sequencing for predicting protein abundance. *Genes (Basel)* 14: 2065, 2023. doi:10.3390/genes14112065.
312. Wang X, Liu Q, Zhang B. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* 14: 2676–2687, 2014. doi:10.1002/pmic.201400184.
313. Rozanova S, Barkovits K, Nikolov M, Schmidt C, Urlaub H, Marcus K. Quantitative mass spectrometry-based proteomics: an overview. *Methods Mol Biol* 2228: 85–116, 2021. doi:10.1007/978-1-0716-1024-4_8.
314. Coarfa C, Grimm SL, Rajapakshe K, Perera D, Lu HY, Wang X, Christensen KR, Mo Q, Edwards DP, Huang S. Reverse-phase protein array: technology, application, data processing, and integration. *J Biomol Tech* 32: 15–29, 2021. doi:10.7171/jbt.21-3202-001.
315. Amberg A, Riefke B, Schlotterbeck G, Ross A, Senn H, Dieterle F, Keck M. NMR and MS methods for metabolomics. *Methods Mol Biol* 1641: 229–258, 2017. doi:10.1007/978-1-4939-7172-5_13.
316. Chen W, Guo W, Li Y, Chen M. Integrative analysis of metabolomics and transcriptomics to uncover biomarkers in sepsis. *Sci Rep* 14: 9676, 2024. doi:10.1038/s41598-024-59400-0.
317. Patt A, Siddiqui J, Zhang B, Mathé E. Integration of metabolomics and transcriptomics to identify gene-metabolite relationships specific to phenotype. *Methods Mol Biol* 1928: 441–468, 2019. doi:10.1007/978-1-4939-9027-6_23.
318. Huang Y, Mohanty V, Dede M, Tsai K, Daher M, Li L, Rezvani K, Chen K. Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlex. *Nat Commun* 14: 4883, 2023. doi:10.1038/s41467-023-40457-w.
319. Abdik E, Çakır T. Transcriptome-based biomarker prediction for Parkinson's disease using genome-scale metabolic modeling. *Sci Rep* 14: 585, 2024. doi:10.1038/s41598-023-51034-y.
320. Takano Y, Suzuki J, Nomura K, Fujii G, Zenkoh J, Kawai H, Kuze Y, Kashima Y, Nagasawa S, Nakamura Y, Kojima M, Tsuchihara K, Seki M, Kanai A, Matsubara D, Kohno T, Noguchi M, Nakaya A, Tsuboi M, Ishii G, Suzuki Y, Suzuki A. Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat Commun* 15: 10637, 2024. doi:10.1038/s41467-024-54671-7.
321. Fan J, Lu F, Qin T, Peng W, Zhuang X, Li Y, Hou X, Fang Z, Yang Y, Guo E, Yang B, Li X, Fu Y, Kang X, Wu Z, Han L, Mills GB, Ma X, Li K, Wu P, Ma D, Chen G, Sun C. Multiomic analysis of cervical squamous cell carcinoma identifies cellular ecosystems with biological and clinical relevance. *Nat Genet* 55: 2175–2188, 2023. doi:10.1038/s41588-023-01570-0.
322. Chu LX, Wang WJ, Gu XP, Wu P, Gao C, Zhang Q, Wu J, Jiang DW, Huang JQ, Ying XW, Shen JM, Jiang Y, Luo LH, Xu JP, Ying YB, Chen HM, Fang A, Feng ZY, An SH, Li XK, Wang ZG. Spatiotemporal multi-omics: exploring molecular landscapes in aging and regenerative medicine. *Mil Med Res* 11: 31, 2024. doi:10.1186/s40779-024-00537-4.
323. Ma A, McDermaid A, Xu J, Chang Y, Ma Q. Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol* 38: 1007–1022, 2020. doi:10.1016/j.tibtech.2020.02.013.
324. Santiago-Rodríguez TM, Hollister EB. Multi-omic data integration: a review of concepts, considerations, and approaches. *Semin Perinatol* 45: 151456, 2021. doi:10.1016/j.semperi.2021.151456.
325. Babu M, Snyder M. Multi-omics profiling for health. *Mol Cell Proteomics* 22: 100561, 2023. doi:10.1016/j.mcpro.2023.100561.
326. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Brief Bioinform* 23: bbab454, 2022. doi:10.1093/bib/bbab454.

327. Gambhir SS, Ge TJ, Vermesh O, Spitler R, Gold GE. Continuous health monitoring: an opportunity for precision health. *Sci Transl Med* 13: eabe5383, 2021. doi:10.1126/scitranslmed.abe5383.
328. Heyde I, Oster H. Differentiating external zeitgeber impact on peripheral circadian clock resetting. *Sci Rep* 9: 20114, 2019. doi:10.1038/s41598-019-56323-z.
329. Ruby NF, Dark J, Burns DE, Heller HC, Zucker I. The suprachiasmatic nucleus is essential for circadian body temperature rhythms in hibernating ground squirrels. *J Neurosci* 22: 357–364, 2002. doi:10.1523/JNEUROSCI.22-01-00357.2002.
330. Schwartz MD, Wotus C, Liu T, Friesen WO, Borjigin J, Oda GA, de la Iglesia HO. Dissociation of circadian and light inhibition of melatonin release through forced desynchronization in the rat. *Proc Natl Acad Sci USA* 106: 17540–17545, 2009. doi:10.1073/pnas.0906382106.
331. Vujovic N, Davidson AJ, Menaker M. Sympathetic input modulates, but does not determine, phase of peripheral circadian oscillators. *Am J Physiol Regul Integr Comp Physiol* 295: R355–R360, 2008. doi:10.1152/ajpregu.00498.2007.
332. LeGates TA, Altimus CM, Wang H, Lee HK, Yang S, Zhao H, Kirkwood A, Weber ET, Hattar S. Aberrant light directly impairs mood and learning through melanopsin-expressing neurons. *Nature* 491: 594–598, 2012. doi:10.1038/nature11673.
333. Fan X, Chen D, Wang Y, Tan Y, Zhao H, Zeng J, Li Y, Guo X, Qiu H, Gu Y. Light intensity alters the effects of light-induced circadian disruption on glucose and lipid metabolism in mice. *Am J Physiol Endocrinol Physiol* 322: E1–E9, 2022. doi:10.1152/ajpendo.00025.2021.
334. Woelders T, Beersma DGM, Gordijn MCM, Hut RA, Wams EJ. Daily light exposure patterns reveal phase and period of the human circadian clock. *J Biol Rhythms* 32: 274–286, 2017. doi:10.1177/0748730417696787.
335. Toffoli B, Tonon F, Giudici F, Ferretti T, Ghirigato E, Contessa M, Francica M, Candido R, Puato M, Grillo A, Fabris B, Bernardi S. Preliminary study on the effect of a night shift on blood pressure and clock gene expression. *Int J Mol Sci* 24: 9309, 2023. doi:10.3390/ijms24119309.
336. Wang N, Sun Y, Zhang H, Wang B, Chen C, Wang Y, Chen J, Tan X, Zhang J, Xia F, Qi L, Lu Y. Long-term night shift work is associated with the risk of atrial fibrillation and coronary heart disease. *Eur Heart J* 42: 4180–4188, 2021. doi:10.1093/eurheartj/ehab505.
337. Yang L, Luo Y, He L, Yin J, Li T, Liu S, Li D, Cheng X, Bai Y. Shift work and the risk of cardiometabolic multimorbidity among patients with hypertension: a prospective cohort study of UK Biobank. *J Am Heart Assoc* 11: e025936, 2022. doi:10.1161/JAHA.122.025936.
338. West AC, Smith L, Ray DW, Loudon ASI, Brown TM, Bechtold DA. Misalignment with the external light environment drives metabolic and cardiac dysfunction. *Nat Commun* 8: 417, 2017. doi:10.1038/s41467-017-00462-2.
339. Tuvia N, Pivovarova-Ramich O, Murahovschi V, Lück S, Grudziecki A, Ost AC, Kruse M, Nikiforova VJ, Osterhoff M, Gottmann P, Gögebakan Ö, Sticht C, Gretz N, Schupp M, Schürmann A, Rudovich N, Pfeiffer AFH, Kramer A. Insulin directly regulates the circadian clock in adipose tissue. *Diabetes* 70: 1985–1999, 2021. doi:10.2337/db20-0910.
340. Greenwell BJ, Trott AJ, Beytebiere JR, Pao S, Bosley A, Beach E, Finegan P, Hernandez C, Menet JS. Rhythmic food intake drives rhythmic gene expression more potently than the hepatic circadian clock in mice. *Cell Rep* 27: 649–657.e5, 2019. doi:10.1016/j.celrep.2019.03.064.
341. Martín-Reyes F, Ho-Plagaro A, Rodríguez-Díaz C, Lopez-Gómez C, García-Serrano S, de los Reyes DR, Gonzalo M, Fernández-García JC, Montiel-Casado C, Fernández-Aguilar JL, Fernández JR, García-Fuentes E, Rodríguez-Pacheco F. Oleic acid regulates the circadian rhythm of adipose tissue in obesity. *Pharmacol Res* 187: 106579, 2023. doi:10.1016/j.phrs.2022.106579.
342. Manella G, Sabath E, Aviram R, Dandavate V, Ezagouri S, Golik M, Adamovich Y, Asher G. The liver-clock coordinates rhythmicity of peripheral tissues in response to feeding. *Nat Metab* 3: 829–842, 2021. doi:10.1038/s42255-021-00395-7.
343. Acosta-Rodríguez VA, Rijo-Ferreira F, van Rosmalen L, Izumo M, Park N, Joseph C, Hepler C, Thorne AK, Stubblefield J, Bass J, Green CB, Takahashi JS. Misaligned feeding uncouples daily rhythms within brown adipose tissue and between peripheral clocks. *Cell Rep* 43: 114523, 2024. doi:10.1016/j.celrep.2024.114523.
344. Chamorro R, Basfi-Fer K, Sepúlveda B, Fariás R, Rojas P, Carrasco F, Codoceo J, Inostroza J, Ruz M. Meal timing across the day modulates daily energy intake in adult patients with type 2 diabetes. *Eur J Clin Nutr* 76: 1470–1477, 2022. doi:10.1038/s41430-022-01128-z.
345. Bo S, Musso G, Beccuti G, Fadda M, Fedele D, Gambino R, Gentile L, Durazzo M, Ghigo E, Cassader M. Consuming more of daily caloric intake at dinner predisposes to obesity. A 6-year population-based prospective cohort study. *PLoS One* 9: e108467, 2014. doi:10.1371/journal.pone.0108467.
346. Aqeel MM, Guo J, Lin L, Gelfand SB, Delp EJ, Bhadra A, Richards EA, Hennessy E, Eicher-Miller HA. Temporal dietary patterns are associated with obesity in US adults. *J Nutr* 150: 3259–3268, 2020. doi:10.1093/jn/nxaa287.
347. Jang KW, Kim YS, Kim MJ, Kim SR, Lee DW, Lee SB, Kim IY. Time-restricted feeding protects against cisplatin-induced acute kidney injury in mice. *Kidney Res Clin Pract* 43: 444–456, 2024. doi:10.23876/j.krcp.23.351.
348. Azemi AK, Siti-Sarah AR, Mokhtar SS, Rasool AHG. Time-restricted feeding improved vascular endothelial function in a high-fat diet-induced obesity rat model. *Vet Sci* 9: 217, 2022. doi:10.3390/vetsci9050217.
349. Lao BN, Luo JH, Xu XY, Fu LZ, Tang F, Ouyang WW, Xu XZ, Wei MT, Xiao BJ, Chen LY, Wu YF, Liu XS. Time-restricted feeding's effect on overweight and obese patients with chronic kidney disease stages 3–4: a prospective non-randomized control pilot study. *Front Endocrinol (Lausanne)* 14: 1096093, 2023. doi:10.3389/fendo.2023.1096093.
350. Zhang D, Colson JC, Jin C, Becker BK, Rhoads MK, Pati P, Neder TH, King MA, Valcin JA, Tao B, Kasztan M, Paul JR, Bailey SM, Pollock JS, Gamble KL, Pollock DM. Timing of food intake drives the circadian rhythm of blood pressure. *Function (Oxf)* 2: zqaa034, 2021. doi:10.1093/function/zqaa034.
351. Sutton EF, Beyl R, Early KS, Cefalu WT, Ravussin E, Peterson CM. Early time-restricted feeding improves insulin sensitivity, blood pressure, and oxidative stress even without weight loss in men with pre-diabetes. *Cell Metab* 27: 1212–1221.e3, 2018. doi:10.1016/j.cmet.2018.04.010.
352. Fehrenbach D, Edell C, Smart C, Wassenaar J, Nguyen B, Rhoads M, Pollock D, Pollock J, Madhur M. Time restricted feeding reverses western diet-induced disruptions in diurnal blood pressure rhythms and aortic media thickening. *Physiology* 38: 5733127, 2023. doi:10.1152/physiol.2023.38.S1.5733127.
353. Buxton OM, Lee CW, L'Hermite-Baleriaux M, Turek FW, Van Cauter E. Exercise elicits phase shifts and acute alterations of melatonin that vary with circadian phase. *Am J Physiol Regul Integr Comp Physiol* 284: R714–R724, 2003. doi:10.1152/ajpregu.00355.2002.
354. Miyazaki T, Hashimoto S, Masubuchi S, Honma S, Honma KI. Phase-advance shifts of human circadian pacemaker are accelerated by daytime physical exercise. *Am J Physiol Regul Integr Comp Physiol* 281: R197–R205, 2001. doi:10.1152/ajpregu.2001.281.1.R197.
355. Chtourou H, Driss T, Souissi S, Gam A, Chaouachi A, Souissi N. The effect of strength training at the same time of the day on the diurnal fluctuations of muscular anaerobic performances. *J Strength Cond Res* 26: 217–225, 2012. doi:10.1519/JSC.0b013e31821d5e8d.
356. Fairbrother K, Cartner B, Alley JR, Curry CD, Dickinson DL, Morris DM, Collier SR. Effects of exercise timing on sleep architecture and nocturnal blood pressure in prehypertensives. *Vasc Health Risk Manag* 10: 691–698, 2014. doi:10.2147/VHRM.S73688.
357. Menet JS, Rodríguez J, Abruzzi KC, Rosbash M. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1: e00011, 2012. doi:10.7554/eLife.00011.
358. Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338: 349–354, 2012. doi:10.1126/science.1226339.
359. Mure LS, Le HD, Benegiamo G, Chang MW, Rios L, Jillani N, Ngotho M, Kariuki T, Dkhissi-Benyahya O, Cooper HM, Panda S. Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* 359: eaao0318, 2018. doi:10.1126/science.aao0318.
360. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology

- and medicine. *Proc Natl Acad Sci USA* 111: 16219–16224, 2014. doi:10.1073/pnas.1408886111.
361. Ruben MD, Wu G, Smith DF, Schmidt RE, Francey LJ, Lee YY, Anafi RC, Hogenesch JB. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Sci Transl Med* 10: eaat8806, 2018. doi:10.1126/scitranslmed.aat8806.
 362. Le Martelot G, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B, Herr W, Deplancke B, Schibler U, Rougemont J, Guex N, Hernandez N, Naef F; CyclIX Consortium. Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. *PLoS Biol* 10: e1001442, 2012. doi:10.1371/journal.pbio.1001442.
 363. Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev* 114: 3203–3223, 2014. doi:10.1021/cr400730x.
 364. Jacobs RQ, Schneider DA. Transcription elongation mechanisms of RNA polymerases I, II, and III and their therapeutic implications. *J Biol Chem* 300: 105737, 2024. doi:10.1016/j.jbc.2024.105737.
 365. Hecker N, Seemann SE, Silahatoglu A, Ruzzo WL, Gorodkin J. Associating transcription factors and conserved RNA structures with gene regulation in the human brain. *Sci Rep* 7: 5776, 2017. doi:10.1038/s41598-017-06200-4.
 366. Ly QK, Nguyen MT, Ngo THP, Lee W. Essential role of cortactin in myogenic differentiation: regulating actin dynamics and myocardin-related transcription factor A-serum response factor (MRTFA-SRF) signaling. *Int J Mol Sci* 25: 13564, 2024. doi:10.3390/ijms252413564.
 367. Littleton ES, Childress ML, Gosting ML, Jackson AN, Kojima S. Genome-wide correlation analysis to identify amplitude regulators of circadian transcriptome output. *Sci Rep* 10: 21839, 2020. doi:10.1038/s41598-020-78851-9.
 368. Ovacki MA, Sukumaran S, Almon RR, DuBois DC, Jusko WJ, Androulakis IP. Circadian signatures in rat liver: from gene expression to pathways. *BMC Bioinformatics* 11: 540, 2010. doi:10.1186/1471-2105-11-540.
 369. Desai VG, Moland CL, Branham WS, Delongchamp RR, Fang H, Duffy PH, Peterson CA, Beggs ML, Fuscoe JC. Changes in expression level of genes as a function of time of day in the liver of rats. *Mutat Res* 549: 115–129, 2004. doi:10.1016/j.mrfmmm.2003.11.016.
 370. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet* 18: 164–179, 2017. doi:10.1038/nrg.2016.150.
 371. Rijo-Ferreira F, Takahashi JS. Genomics of circadian rhythms in health and disease. *Genome Med* 11: 82, 2019. doi:10.1186/s13073-019-0704-0.
 372. Marri D, Filipovic D, Kana O, Tischkau S, Bhattacharya S. Prediction of mammalian tissue-specific CLOCK-BMAL1 binding to E-box DNA motifs. *Sci Rep* 13: 7742, 2023. doi:10.1038/s41598-023-34115-w.
 373. Menet JS, Pescatore S, Rosbash M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev* 28: 8–13, 2014. doi:10.1101/gad.228536.113.
 374. Beytebierre JR, Trott AJ, Greenwell BJ, Osborne CA, Vitet H, Spence J, Yoo S-H, Chen Z, Takahashi JS, Ghaffari N, Menet JS. Tissue-specific BMAL1 cisomes reveal that rhythmic transcription is associated with rhythmic enhancer-enhancer interactions. *Genes Dev* 33: 294–309, 2019. doi:10.1101/gad.322198.118.
 375. Li Z, Fu B, Wei A, Wu Y, Huang M, Zhang E, Cui B, Wang B, Peng H. d-Glucosamine induces circadian phase delay by promoting BMAL1 degradation through AMPK/mTOR pathway. *Life Sci* 325: 121765, 2023. doi:10.1016/j.lfs.2023.121765.
 376. Zuo Y, Hou Y, Wang Y, Yuan L, Cheng L, Zhang T. Circadian misalignment impairs oligodendrocyte myelination via Bmal1 overexpression leading to anxiety and depression-like behaviors. *J Pineal Res* 76: e12935, 2024. doi:10.1111/jpi.12935.
 377. Chhe K, Hegde MS, Taylor SR, Farkas ME. Circadian effects of melatonin receptor-targeting molecules in vitro. *Int J Mol Sci* 25: 13508, 2024. doi:10.3390/ijms252413508.
 378. Douma LG, Costello HM, Crislip GR, Cheng KY, Lynch IJ, Juffre A, Barral D, Masten S, Roig E, Beguiristain K, Li W, Bratanatawira P, Wingo CS, Gumz ML. Kidney-specific KO of the circadian clock protein PER1 alters renal Na⁺ handling, aldosterone levels, and kidney/adrenal gene expression. *Am J Physiol Renal Physiol* 322: F449–F459, 2022. doi:10.1152/ajprenal.00385.2021.
 379. Zietara A, Spires DR, Juffre A, Costello HM, Crislip GR, Douma LG, Levchenko V, Dissanayake LV, Klemens CA, Nikolaienko O, Geurts AM, Gumz ML, Staruschenko A. Knockout of the circadian clock protein PER1 (Period1) exacerbates hypertension and increases kidney injury in Dahl salt-sensitive rats. *Hypertension* 79: 2519–2529, 2022. doi:10.1161/HYPERTENSIONAHA.122.19316.
 380. Douma LG, Crislip GR, Cheng KY, Barral D, Masten S, Holzworth M, Roig E, Glasford K, Beguiristain K, Li W, Bratanatawira P, Lynch IJ, Cain BD, Wingo CS, Gumz ML. Knockout of the circadian clock protein PER1 results in sex-dependent alterations of ET-1 production in mice in response to a high-salt diet plus mineralocorticoid treatment. *Can J Physiol Pharmacol* 98: 579–586, 2020. doi:10.1139/cjpp-2019-0688.
 381. Gachon F, Leuenberger N, Claudel T, Gos P, Joffe C, Fleury Olela F, de Mollerat Du Jeu X, Wahli W, Schibler U. Proline- and acidic amino acid-rich basic leucine zipper proteins modulate peroxisome proliferator-activated receptor alpha (PPARalpha) activity. *Proc Natl Acad Sci USA* 108: 4794–4799, 2011. doi:10.1073/pnas.1002862108.
 382. Yang X, Downes M, Yu RT, Bookout AL, He W, Straume M, Mangelsdorf DJ, Evans RM. Nuclear receptor expression links the circadian clock to metabolism. *Cell* 126: 801–810, 2006. doi:10.1016/j.cell.2006.06.050.
 383. Lamia KA, Sachdeva UM, DiTacchio L, Williams EC, Alvarez JG, Egan DF, Vasquez DS, Juguilon H, Panda S, Shaw RJ, Thompson CB, Evans RM. AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* 326: 437–440, 2009. doi:10.1126/science.1172156.
 384. Zamboni AC, McDearmon EL, Salomonis N, Vranizan KM, Johansen KL, Adey D, Takahashi JS, Schambelan M, Conklin BR. Time- and exercise-dependent gene regulation in human skeletal muscle. *Genome Biol* 4: R61, 2003. doi:10.1186/gb-2003-4-10-r61.
 385. Harmsen J-F, Kotte M, Habets I, Bosschee F, Frenken K, Jorgensen JA, de Kam S, Moonen-Kornips E, Cissen J, Doligkeit D, van de Weijer T, Erazo-Tapia E, Buitinga M, Hoeks J, Schrauwen P. Exercise training modifies skeletal muscle clock gene expression but not 24-hour rhythmicity in substrate metabolism of men with insulin resistance. *J Physiol* 602: 6417–6433, 2024. doi:10.1113/JP285523.
 386. Bennardo M, Alibhai F, Tsimakouridze E, Chinnappareddy N, Podobed P, Reitz C, Pyle WG, Simpson J, Martino TA. Day-night dependence of gene expression and inflammatory responses in the remodeling murine heart post-myocardial infarction. *Am J Physiol Regul Integr Comp Physiol* 311: R1243–R1254, 2016. doi:10.1152/ajpregu.00200.2016.
 387. Manella G, Ezagouri S, Champigneulle B, Gaucher J, Mendelson M, Lemarie E, Stauffer E, Pichon A, Howe CA, Doutreleau S, Golik M, Verges S, Asher G. The human blood transcriptome exhibits time-of-day-dependent response to hypoxia: lessons from the highest city in the world. *Cell Rep* 40: 111213, 2022. doi:10.1016/j.celrep.2022.111213.
 388. Lee Y, Fong SY, Shon J, Zhang SL, Brooks R, Lahens NF, Chen D, Dang CV, Field JM, Sehgal A. Time-of-day specificity of anticancer drugs may be mediated by circadian regulation of the cell cycle. *Sci Adv* 7: eabd2645, 2021. doi:10.1126/sciadv.abd2645.
 389. Ector C, Schmal C, Didier J, De Landtsheer S, Finger AM, Müller-Marquardt F, Schulte JH, Sauter T, Keilholz U, Herzel H, Kramer A, Granada AE. Time-of-day effects of cancer drugs revealed by high-throughput deep phenotyping. *Nat Commun* 15: 7205, 2024. doi:10.1038/s41467-024-51611-3.
 390. Bilinovich SM, Uhl KL, Lewis K, Soehnlen X, Williams M, Vogt D, Prokop JW, Campbell DB. Integrated RNA sequencing reveals epigenetic impacts of diesel particulate matter exposure in human cerebral organoids. *Dev Neurosci* 42: 195–207, 2020. doi:10.1159/000513536.
 391. Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal AW, Kim Y, Kyle PB, Attipoe E, Johnson AC, Uhl KL, Sirpilla OL, Jahanbakhsh S, Robinson M, Levy S, Valdar W, Garrett MR, Solberg Woods LC. Sept8/SEPTIN8 involvement in cellular structure and kidney damage is identified by genetic mapping and a novel human tubule hypoxic model. *Sci Rep* 11: 2071, 2021. doi:10.1038/s41598-021-81550-8.

392. **Graves JD, Krebs EG.** Protein phosphorylation and signal transduction. *Pharmacol Ther* 82: 111–121, 1999. doi:10.1016/s0163-7258(98)00056-4.
393. **Pawson T, Scott JD.** Protein phosphorylation in signaling—50 years and counting. *Trends Biochem Sci* 30: 286–290, 2005. doi:10.1016/j.tibs.2005.04.013.
394. **Pandey D, Onkara Perumal P.** A scoping review on deep learning for next-generation RNA-Seq. data analysis. *Funct Integr Genomics* 23: 134, 2023. doi:10.1007/s10142-023-01064-6.
395. **Li B.** Personalized immunotherapy of patients: defining by single-cell RNA-seq with artificial intelligence. *Med Res Arch* 11: 10.18103/mra.v11i8.4293, 2023. doi:10.18103/mra.v11i8.4293.
396. **Madadi Y, Monavarfeshani A, Chen H, Stamer WD, Williams RW, Yousefi S.** Artificial intelligence models for cell type and subtype identification based on single-cell RNA sequencing data in vision science. *IEEE/ACM Trans Comput Biol Bioinform* 20: 2837–2852, 2023. doi:10.1109/TCBB.2023.3284795.
397. **Quake SR.** A decade of molecular cell atlases. *Trends Genet* 38: 805–810, 2022. doi:10.1016/j.tig.2022.01.004.
398. **Song Q, Su J.** DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief Bioinform* 22: bbab414, 2021. doi:10.1093/bib/bbaa414.
399. **Lu L, Ono N, Welch JD.** Linking transcriptome and morphology in bone cells at cellular resolution with generative AI. *J Bone Miner Res* 40: 20–26, 2024. doi:10.1093/jbmr/zjae151.
400. **Li Y, Stanojevic S, Garmire LX.** Emerging artificial intelligence applications in spatial transcriptomics analysis. *Comput Struct Biotechnol J* 20: 2895–2908, 2022. doi:10.1016/j.csbj.2022.05.056.
401. **Zhang J, Lang M, Zhou Y, Zhang Y.** Predicting RNA structures and functions by artificial intelligence (Preprint). *Trends Genet* S0168–9525(23)00229-9, 2023. doi:10.1016/j.tig.2023.10.001.
402. **Sato K, Hamada M.** Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery. *Brief Bioinform* 24: bbab186, 2023. doi:10.1093/bib/bbad186.
403. **Qin S, Tang X, Chen Y, Chen K, Fan N, Xiao W, Zheng Q, Li G, Teng Y, Wu M, Song X.** mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Signal Transduct Target Ther* 7: 166, 2022. doi:10.1038/s41392-022-01007-w.
404. **Chen Y, Jiang Z, Yang Y, Zhang C, Liu H, Wan J.** The functions and mechanisms of post-translational modification in protein regulators of RNA methylation: current status and future perspectives. *Int J Biol Macromol* 253: 126773, 2023. doi:10.1016/j.ijbiomac.2023.126773.
405. **Paunovska K, Loughrey D, Dahlman JE.** Drug delivery systems for RNA therapeutics. *Nat Rev Genet* 23: 265–280, 2022. doi:10.1038/s41576-021-00439-4.
406. **Zhao S, Hong CKY, Myers CA, Granas DM, White MA, Corbo JC, Cohen BA.** A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. *Nat Genet* 55: 346–354, 2023. doi:10.1038/s41588-022-01278-7.
407. **Bravo González-Blas C, Quan XJ, Duran-Romaña R, Taskiran II, Koldere D, Davie K, Christiaens V, Makhzami S, Hulselmans G, de Waegeneer M, Mauduit D, Poovathingal S, Aibar S, Aerts S.** Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol Syst Biol* 16: e9438, 2020. doi:10.15252/msb.20209438.
408. **Sahin U, Karikó K, Türeci Ö.** mRNA-based therapeutics—developing a new class of drugs. *Nat Rev Drug Discov* 13: 759–780, 2014. doi:10.1038/nrd4278.
409. **Weissman D, Karikó K.** mRNA: fulfilling the promise of gene therapy. *Mol Ther* 23: 1416–1417, 2015. doi:10.1038/mt.2015.138.
410. **Pardi N, Hogan MJ, Porter FW, Weissman D.** mRNA vaccines - a new era in vaccinology. *Nat Rev Drug Discov* 17: 261–279, 2018. doi:10.1038/nrd.2017.243.
411. **Nance KD, Meier JL.** Modifications in an emergency: the role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent Sci* 7: 748–756, 2021. doi:10.1021/acscentsci.1c00197.