HHS Public Access

Author manuscript

Curr Stem Cell Rep. Author manuscript; available in PMC 2024 June 27.

Published in final edited form as:

Curr Stem Cell Rep. 2023 June ; 9(2): 31–41. doi:10.1007/s40778-023-00227-2.

Transcription factors and splice factors - interconnected regulators of stem cell differentiation

Madison M. Mehlferber,
Department of Biochemistry and Molecular Genetics, University Virginia, Charlottesville, VA 22903

Muge Kuyumcu-Martinez, PhD,
Department of Molecular Physiology and Biological Physics, University of Virginia, School of Medicine, Fontaine Medical Office Building 1, 415 Ray C. Hunt Dr, Charlottesville, VA 22903

Clint L. Miller, PhD,
Department of Public Health Sciences, Department of Biochemistry and Molecular Genetics, and Department of Biomedical Engineering, University of Virginia, Multistory Building, West Complex, 1335 Lee St, Charlottesville, VA 22908, PO Box 800717, Charlottesville, Virginia 22908

Gloria M. Sheynkman, PhD
Department of Molecular Physiology and Biological Physics, Center for Public Health Genomics, UVA Comprehensive Cancer Center, Department of Biochemistry and Molecular Genetics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22903

Summary

Purpose of review: The underlying molecular mechanisms that direct stem cell differentiation into fully functional, mature cells remain an area of ongoing investigation. Cell state is the product of the combinatorial effect of individual factors operating within a coordinated regulatory network. Here, we discuss the contribution of both gene regulatory and splicing regulatory networks in defining stem cell fate during differentiation and the critical role of protein isoforms in this process.

Recent findings: We review recent experimental and computational approaches that characterize gene regulatory networks, splice regulatory networks, and the resulting transcriptome and proteome they mediate during differentiation. Such approaches include long-read RNA sequencing, which has demonstrated high-resolution profiling of mRNA isoforms, and Cas13-based CRISPR, which could make possible high-throughput isoform screening. Collectively, these developments enable systems-level profiling of factors contributing to cell state.

Corresponding author: **Gloria M. Sheynkman, PhD** - gs9yr@virginia.edu.

Conflict of interest: Madison M. Mehlferber, Muge Kuyumcu, Gloria M. Sheynkman declare they have no conflict of interest. Clint L. Miller has received funding support from AstraZeneca for work unrelated to the current study.

Human and Animal Rights and Informed Consent: This article does not contain studies with human or animal subjects performed by any of the authors.

Disclaimer: The funders had no role in the study design, data collection and analysis, decision to publish or preparation of this manuscript.

Summary: Overall, gene and splice regulatory networks are important in defining cell state. The emerging high-throughput systems-level approaches will characterize the gene regulatory network components necessary in driving stem cell differentiation.

Keywords

Alternative Splicing; Stem cells; Development; Regulatory Networks; Transcription Factors; Splice Factors

Stem cells as a powerful system for studying development and disease

Stem cells differentiate into diverse cells - Waddington landscape and molecular patterns

Stem cells can differentiate into any cell type in the body. Initially, human embryos were used to generate embryonic stem cells (hESCs) that are derived from the inner cell mass (ICM) of the blastocyst [1]. Such cells are pluripotent, with the ability to differentiate into any cell of the body given the proper genetic or exogenous factors. Another source of stem cells are terminally differentiated somatic cells, which can be reprogrammed into induced pluripotent cells (iPSCs) by addition of four key transcription factors (TFs)—OCT4, SOX2, c-MYC and KLF4, as shown in the landmark experiment in the Yamanaka lab [2]. iPSCS are largely indistinguishable from embryo-derived hESCs in terms of their genetic, molecular, and phenotypic properties, permitting widespread application of these cells for disease modeling without the ethical issues of human embryo use [3,4].

Stem cells can model the transition from a pluripotent to a differentiated state, which is critical for development of specialized cell types and tissues. Stem cell potency exists in a continuum, with successive cell divisions correlating with narrower differentiation potentials. Such transitions can be thought of as a series of cellular states, which has been analogized by Conrad Waddington as a marble (the “cell”) traveling down a hilly terrain to arrive at a position of the energetically most favorable cellular attractor state (the terminally “differentiated cell”) [5]. Cell states are reflected by global patterns of gene expression, especially transcript and protein molecular expression. Tracking such coordinated molecular expression changes in these interim steps of differentiation can provide insight into the underlying regulatory network logic associated with these cell states, the relationship of which is a critical question for the stem cell field [6,7].

Stem cells are tractable models to link molecular variation to development

Stem cell models provide a portal to study otherwise inaccessible aspects of *in vivo* human development, particularly at the genetic and molecular level. Protocols are now available to direct differentiation of stem cells into hundreds of cell types [8,9]. For example, in cardiovascular development, stem cells can be differentiated into arterial and venous endothelial cells subtypes [10,11], and hematopoietic development have well characterized models [12]. And, it has been shown that hESCs can differentiate into epicardial cells that graft onto damaged heart tissue for repair [13]. More recently, going beyond the constraint of monolayer 2D cell cultures, 3D organoid cultures have been developed for a series of organ types that better recapitulate *in vivo* phenomena such as cell-cell interactions and soluble factor gradients [14].

In addition to hESCs, human iPSCs (hiPSCs) can also be differentiated into several different cell types and are widely used to understand underpinnings of development. For example, in the heart, both hECs and hiPSCs can be differentiated into several different cell types including endothelial cells, endocardial cells, and cardiomyocytes to define molecular drivers of cardiovascular development[15].These differentiated cells can then be cocultured to study the communication between these cells that build up the heart. Furthermore, iPSCs obtained from patients are excellent tools to determine the mechanisms responsible for disease pathogenesis as well as identifying developmental defects that give rise to these complex disease phenotypes. Stem cells also allow temporal analyses of molecular and cellular events that occur during development. Genome edited hiPSCs or hESCs carrying patient specific mutations are used to model cell specific defects that give rise to human diseases well as to perform screens of compounds or drugs for treatment of disease complications[15]. Both hESC and hiPSC mediated stem cell model systems are utilized to identify transcriptional regulatory networks necessary for development and function[16]. With advancements in RNA sequencing and computational methods, post-transcriptional regulatory networks necessary for stem cell differentiation are becoming more appreciated.

Tracing the expression changes of factors associated with differentiation temporally can indicate links between genetic factors and the downstream developmental pathways they regulate [17,18]. Though not fully recapitulating *in vivo* complexity, the journey a stem cell takes during *in vitro* differentiation at least partially recapitulates molecular changes occurring during development and can provide a tractable experimental system with human relevance. The practical benefits include the ability to culture cells *in vitro* to generate sufficient material for high-throughput molecular assays and biochemical and genetic screens. The benefit of human relevance arises from the fact that human stem cells should best recapitulate the repertoire of transcript and protein molecular forms (e.g., isoforms, proteoforms) that are primate or human-specific. Though many genes are conserved between human and model organisms, the molecular details of gene products, such as splicing patterns, tend to diverge greatly [19].

The generation of certain cell types is possible by addition of individual soluble factors or transcriptional regulators [16,20]. In large part this knowledge arose from trial and error or small-scale screens, guided by simple morphogenic or gene expression patterning [21,22], rather than a fundamental understanding of the underlying gene regulatory network that governs cell behavior. Even such factors operate within an interconnected and complex gene regulatory network; and, therefore, an incomplete picture remains of the underlying molecular logic and critical factors that direct differentiation of stem cells into fully functional, mature cells[23]. The field of stem cell systems biology aims to complement focused functional studies, through paradigms that merge high-throughput datasets and computational models.

Gene regulatory networks - TFs mediate pluripotency and differentiation

One type of network that plays a critical role in cell fate decisions are gene regulatory networks (GRNs). A GRN is represented by the set of active transcription factors (TFs) that bind to and regulate their target genes. TFs comprise about 10% of all protein-coding genes

[24]. TFs can serve as an activator that promotes transcription of the gene or a repressor that inhibits, and thus lowers gene transcription. Within a network representation, TFs can be modeled as nodes, from which one or more directed edges point to target genes. The edge has a sign, depending on the activating (+) or repressive (−) function of the TF [25].

TFs networks are characterized by biochemical or genome-wide techniques [26]. The specific DNA sequences to which a TF binds can be assayed using high-throughput approaches[24] such as systematic evolution of ligands by exponential enrichment (SELEX) [24,27], protein binding microarrays (PBMs)[28], or microfluidic chip-based mechanically induced trapping of molecular interactions (MITOMI) [29]. Cell-type specific TF binding within a cellular context can be inferred genome-wide using approaches such as DNAse-Seq[30] or ATAC-Seq[31]. TF binding can be directly mapped to genomic sites with approaches such as Chip-Seq[32] or Cut&Run[33].

TFs as drivers of stem cell fate - GRN knowledge to guide stem cell engineering

TFs can play an outsize role in influencing stem cell fate by modulating gene expression patterns in differentiating cells [34]. To accomplish this in stem cell differentiation, TFs are thought to concurrently repress pluripotent genes and to allow activation of lineage-specific genes [16]. The potent effects of TFs is evident by the fact that *OCT4*, *c-MYC*, *KLF4* and *SOX2* form a core network that can reprogram a somatic cell to an iPSC [2]. Just one or a few TFs can, solely, convert cells to certain lineages. A quintessential example is *MyoD*, which can convert somatic cells into muscle cells [35]. Many other lineage defining TFs exist across the differentiation spectrum, such as *ETV2*, which is necessary and sufficient for converting mesodermal precursors to primordial endothelial cells [36].

About 1,564 TFs have been annotated [37], but for many their role in driving cell fate decisions remain uncharacterized [38]. Recently, this has been tested across most human TFs experimentally [37]. A study employed a library of all human TF open reading frame (ORF) clones to test the role of individual TF expression in differentiating cells [37]. Surprisingly, 241 of the 1,564 TFs tested resulted in a differentiation phenotype across biological replicates. More recently, an even more comprehensive library of human TFs that included all RefSeq annotated ~3,500 TF isoforms were overexpressed in a population of stem cells and, using single cell RNA-seq as a readout—each TFs was profiled in terms of its ability to drive diverse cell-type-specific gene expression programs [39]. For a subset of TFs, the effect of multiple TFs on cell fate were tested, showing both cooperative and antagonistic relationships and reflecting the combinatorial nature of TF activities.

Cellular gene expression is the product of the timing and location of the collective activities of all TFs in a cell. The correlation between TF activity—most commonly, mRNA levels as a proxy—and expression of target genes, either experimentally or through inference via co-expression, is the basis for inferring links between TFs and their downstream targets [40]. These correlations guide computational predictions of TFs responsible for determining cell fates. For example, approaches such as CellNet[41] and Mogrify[42] leverage knowledge of TFs and their downstream targets and pathways responsible for promoting differentiation transitions in order to nominate TFs that would be ideal factors for cell engineering [40,43].

Regulation beyond transcription - alternative splicing regulatory networks

GRNs and the resulting transcriptional output is significantly affected through the process of alternative splicing (AS) [44]. During AS, intronic regions are excised from nascent mRNA while the remaining exon coding regions are ligated together to form distinct mRNA isoforms. Such isoforms can greatly expand the protein functional diversity of the cell [45,46]. The first large-scale RNA-seq datasets of human tissues have revealed that 95% of human genes undergo AS[47], with many tissue-specific AS expression patterns [48].

Splice regulatory networks - the central role of splicing factors

AS is controlled by the splicing regulatory networks (SRN) operative in a cell. The workhorse of SRNs is the spliceosome, a protein machine comprises over 200 proteins [49] and five ribonuclear proteins (U1, U2, U4/U6 and U5) [50,51]. The process of splicing unfolds via the sequential binding of small nuclear ribonucleoproteins (snRNPs) on the pre-mRNA, which eventually catalyze the joining of particular exon junctions [49]. The activity of the spliceosome is regulated by a repertoire of splice factors (SFs), a subfamily of approximately 356[52] RNA binding proteins (RBPs) that bind to sites on pre-mRNA and interacts with components of the spliceosome to enhance or inhibit splicing reactions at certain exons [53–55]. Families of SF exist, such as heterogeneous nuclear ribonucleoproteins (hnRNPs), which tend to inhibit (i.e., silence) splicing [56,57], and serine and arginine rich proteins (SRs), which tend to enhance splicing [51], although the regulatory activity of SFs can be highly dependent on context and position of binding within the pre-mRNA [58]. Many annotated SF functions have been produced by hypothesis-driven studies, and, more recently, systematic efforts, such as through the Encyclopedia of DNA Elements (ENCODE) project have mapped the functional and biophysical networks of 80 SFs in two human cell lines [52].

The emerging role of SFs in mediating pluripotency and differentiation

Just as with TFs, SFs can also act as master regulators of cell fate decisions during stem cell differentiation acting through SRNs [59,60]. For example, *SRSF2* promotes the AS of exon 9 of the *NUMB gene*, creating an isoform of *NUMB* that specifically governs *NOTCH* [61], and thus inducing endothelial cell progenitor cell specification [61]. Muscleblind-like splicing factor (MBNL1) and RNA Binding Fox-2 (RBFOX2) alternatively splice targets that drive iPSCs to mesoderm transitions [62]. RBPs *ESP1*, *ESPR2*, *RBFOX2* and *QKI* coordinate an isoform switch that promotes tissue remodeling from a mesenchymal to epithelial state for kidney development [63]. And, RNA binding motif protein 24 (Rbm24) drives cardiac differentiation programs in human and mouse by AS of genes related to cytoskeletal proteins and ATPase components that promote cardiac development [64]. Splicing regulator QKI is essential to establish splicing networks that control contractile and structural genes in the heart.[65] The repression of PTBP1, another splicing regulator, can convert cardiac fibroblasts into cardiomyocytes or fibroblasts into neuronal cells, indicating its critical role in cell differentiation. [66,67] PTBP1 influences transcriptional networks by regulating transcription factor PBX1 during neuronal differentiation.[68] RBFOX2 is necessary for establishing splicing regulatory networks required for heart and neuronal

development.[69] Overall these examples highlight how SF can serve as master regulators of cellular fate, similar to TFs.

SF networks can be characterized by experimental methods that link RBPs to the target isoforms they regulate. There are several different approaches that include RNA Immunoprecipitation (RIP)[70], and enhanced crosslinking and immunoprecipitation (eCLIP)[71]. Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)[72] One of the most commonly used approaches is CLIP-seq, which links SF to their binding sites within their respective RNA targets transcriptome-wide is crosslinking and immunoprecipitation followed by RNA-sequencing, or CLIP-seq [73,74]. Here, RBPs are subjected to UV irradiation, which cross-links RBPs to the RNA at the site to which they bind. The RBPs, which remain covalently bound to the RNA target, are enriched as an RNA-protein complex, and the population of bound RNA is deeply sequenced and subsequently aligned to the genome to depict locations of the bound RBPs [74], revealing the positions across pre-mRNA and mRNA wherein the RBPs are bound, and presumably active, in a given condition or sample. Another method maps functional relationships between SFs and their targets by experimentally modulating RBP concentrations, such as through siRNA knockdown or overexpression plasmids, followed by measurements of splicing using RNA-seq [52].

Splice factors as regulators of the regulators - TF and SF isoforms

Among the many types of genes that splicing can target, the most marked effects are likely through alternative splicing of potent regulators of stem cell fate. Alternative splicing of such regulators could generate isoforms of the same gene with variable activities - from loss of function to gain in new functions [75]. In other words, splicing can have outsize effects on stem cell fate by “regulating the regulators”.

Splicing influences GRNs by producing TF isoforms that differentially regulate cell fate

A continuum of functional relationships between TF isoforms of the same gene can occur, from isoforms with attenuated, opposite, or tandem functions. The levels and relative stoichiometries of such TF isoforms are directly influenced by AS, and thus AS can influence gene regulatory networks by modifying TF activities through splicing [76,77].

AS can modulate TF functions by production of a sub-functional isoform. For example, AS modulates the activities of *OCT4*. *OCT4* produces at least three major isoforms (OCT4A, OCT4B and OCT4B1)[78,79]. OCT4A is responsible for establishing pluripotency. In contrast, an isoform switch to OCT4B, a sub-functional form of OCT4, results in inhibition of stem cell self-renewal, but may be involved in responses to cell stress [78]. The differences in isoform function may be due to differential inclusion of localization signals in which a nuclear localization signal in OCT4A is absent in OCT4B, reducing its nuclear residence and thus transcriptional activity [80].

AS can modulate TF function by producing an isoform with opposite function. For example, splicing can alter the specificity of the forkhead transcription factor (*FOXP1*) where one isoform promotes pluripotency by directly stimulating expression of the Yamanaka factors,

while another isoform is predominantly expressed in differentiated cells and represses pluripotency genes [81].

And, lastly, AS can modulate TF function in a way in which there is a division of labor between multiple TF isoforms of the same gene. An example can be found in *SALL4*, a member of the spat-like gene family [82]. *SALL4* interacts with both OCT4 and NANOG to regulate pluripotency networks. *SALL4* produces two isoforms, SALL4A and SALL4B, which collaborate to maintain pluripotency networks [83]. SALL4A represses genes associated with differentiation while SALL4B promotes pluripotent gene expression [84]. Interestingly, expression of SALL4B alone is not sufficient to promote the pluripotent state [83].

Analogously to how SFs can modulate activity of TFs, in a similar vein, SFs themselves may also act as upon themselves to modulate their own splicing [60]. Indeed, a pervasive mechanism of SF regulation is a negative feedback loop in which a SF binds to its own pre-mRNA which leads to nonsense-mediated decay products. While further studies work to elucidate these mechanisms as it relates to differentiation, negative regulation has clearly been demonstrated to affect 10–30% of mammalian genes [60], and may play an important role in stem cell SRNs. GRNs and SRNs likely work in concert within stem cells to regulate pathways of pluripotency or differentiation (Figure 1).

Characterization of the transcriptome at isoform-resolution

GRN and SRN programs can influence transcriptome expression. For characterization of the stem cell transcriptome, early efforts employed mid-throughput real-time quantitative PCR (RT-qPCR)[85] assays to quantify pre-selected isoforms panels [79]. More recently, short-read RNA-sequencing (SR RNA-seq) has enabled facile characterization of thousands of annotated and novel splice junctions and exon expression associated with stem cell phenotypes [86–88]. However, the short length of RNA-seq reads limit observation of the entire unambiguous full-length isoform [89,90].

These limitations are addressed with long-read (LR) sequencing platforms, such as through Pacific Biosciences Inc. (PacBio) or Oxford Nanopore Technologies (ONT) [91–94]. In the decade following their introduction, the throughput, affordability, and accuracy of LR sequencing was lower than SR, but with the steady evolution of LR sequencing systems in terms of the chemistries, instrumentation, and computational pipelines (e.g., PacBio's Revio [92], ONT's iSeq100 [95]), accurate transcriptome sequences will likely become accessible at a large depth and breadth. With this higher depth, single cell LR (scLR) methods are being developed for both ONT and PacBio platforms. With ONT as a readout, Barcode identification from Long-reads for AnalyZing single-cell gene Expression (BLAZE) relies on barcodes from ONT long-reads to profile isoforms at single-cell resolution [96]. PacBio has also reached the throughput needed for scLR through development of multiplexed arrays sequencing (MAS-ISO-seq), which uses a concatenation approach to ligate multipole cDNAs into large single molecules that is then sequenced [97]. To infer the functional interrelatedness of isoforms from scLR data, new analysis pipelines have been developed,

such as *acorde*, which analyzes co-expression networks of correlated isoform abundances [98].

Transcriptome and splicing regulation are an intertwined process

AS is co-transcriptional, and thus splicing and transcriptional biochemical processes work in concert within an epigenetic context to drive transcriptional outputs during differentiation [99–101]. Chromatin state has a large influence on splicing [100,102]. Two models have been proposed to describe how chromatin affects splicing outcomes: the kinetic and recruitment model. The kinetic model relates RNA polymerase II transcriptional speed with splice status. “Slow” Pol II increases the time with which an SF is exposed or could bind their cognate RNA binding sites, and thus promotes exon inclusion. The recruitment model focuses on the ability for components such as Pol II C-terminal tail or histone tails to mediate interaction-driven recruitment of SFs onto the nascent pre-mRNA which in turn affects AS [103,104]. Beyond the influence of chromatin and splicing, the biochemical relationships of splicing and transcription are surprisingly intertwined. TFs can bind to and regulate nascent RNA, influencing not only gene expression, but potentially RNA processing such as splicing [101,105,106]. SFs, on the other hand, can influence transcriptional regulation. For example, *RBM20*, through splicing, regulates genes necessary for heart development [107].

Not only are the biochemical mechanisms of transcription and splicing intertwined, but components of the transcriptional (TFs, gene targets) and splicing (SFs, splice targets) networks feed into each other and are involved in cross-regulatory logic, in sometimes unexpected ways. The transcription factor OCT4 upregulates the splice factor *SFRS2*, which regulates the splicing of methyl-CpG-binding protein, *MBD2*, whose isoforms play opposing roles in reprogramming to pluripotency [108]. Upregulation of *SFRS2* increases levels of an isoform of *MBD2* (*MBD2c*), which binds to the promoter of *OCT4* to reinforce the pluripotency core network. Interestingly, loss of either *OCT4* or *SFRS2* activity leads to product of the other isoform of *MBD2* (*MBD2a*), which also binds to the promoter of *OCT4*, but has a different C-terminal domain that silences *OCT4* expression by recruiting the Nucleosome Remodeling and Deacetylation complex (NuRD) [108]. *SRSF2* does not just regulate *MBD2*, but can also change the activity of the transcription factor *FOXP1*, whose isoforms are nearly opposite in their induced stem cell phenotypes; one isoform of *FOXP1* activate genes that promote pluripotency and the other isoform promotes differentiation, a splicing switch that involves exonic changes to the *FOXP1* DNA binding domain [81]. Notably, TFs with zinc finger domains can also bind RNA. Zinc finger domains allow binding to both DNA and RNA. It has been shown that transcription factor GATA4 can bind RNA and regulate alternative splicing networks in the heart.[109]

The functional output of the transcriptome - the stem cell proteome

In large part, the functional effect of the transcriptome manifests through the proteome, making identification of protein expression within the cell equally important. Indeed, transcript and protein abundances are not always highly correlated, with this relationship being affected by several factors including co- and post-translational layers of regulation

[110,111]. Given the importance of splicing networks in stem cell differentiation, approaches to directly measure their functional outputs, or protein isoforms, in stem cells are critical [112,113].

Mass-spectrometry(MS)-based proteomics is a powerful technique for comprehensive general proteome characterization of stem cell states [114–116]. MS has been used to track dynamic changes of proteins over stem cell differentiation, which may not correlate with transcriptional changes. An early work on transcript and protein levels in hESCs demonstrated that up to 50% of changes in protein expression do not have corresponding transcript changes, although some of this non-correlation is attributable to technical variability of the first generation MS instruments [117]. The value of proteomics for discovering important stem cell factors continues to be demonstrated. In a study of young and adult mouse HSCs, a module of proteins were specifically expressed in young mouse HSCs, uncorrelated to the transcript levels [118,119]. Multiplexed isobaric labeling, now a standard approach, allows for measuring protein expression along many more differentiation timepoints [120]. Newer thermal profiling approaches coupled with MS can even assay intrinsic protein stability, and regulated destabilization during early differentiation, such as the ribosomal machinery that exhibits higher stability during differentiation [120].

Though general protein content from genes is readily measured, an ongoing challenge remains the characterization of the final output of gene and splicing regulatory networks: the proteome at isoform resolution. MS characterization of protein isoforms have been applied to histone isoforms and isoforms in mesenchymal stem cells [121,122]. Technical challenges of isoform detection remain, though. In bottom-up MS, proteins are proteolytically digested into short peptides—few peptides uniquely map to an isoform [123], and such peptides are under sampled due to technical issues (MS ionization, charge, etc. [58]). New proteogenomic approaches [90,124–126] can enhance protein detection accuracy and coverage by leveraging matched long-read RNA-seq data to generate a sample specific database of protein isoforms used for MS searching [127], which can provide direct protein level evidence of stable isoform expression [128].

Emerging scalable strategies to causally link isoforms to stem cell phenotypes

Given the critical nature of SRNs in driving cell fates, determining the functional role of their downstream protein isoform products is critical, such as through experimentally testing the effect of knocking out or overexpressing an individual isoform on stem cell phenotypes. Isoforms can be “knocked down” using short interfering RNAs (siRNA) that are designed against regions specific for the target isoform mRNA [129,130]. This approach has been used to modulate splicing patterns of isoforms in cancer cells to functionally decouple gene expression patterns from the individual role of isoforms in driving cell phenotype [131]

To further modulate isoform expression, morpholino oligonucleotides can be used, which are RNA sequences designed to be complementary to the target sequences of RNA. Upon morpholino binding, spliceosome assembly or translation is inhibited through steric hindrance [132,133]. And, the Type II CRISPR-based system Cas13 demonstrates the ability to knock-down RNA isoforms with high specificity and efficiency, with potential to design gRNAs against isoform-specific regions [134,135]. The CRISPR-Cas9 system has been

engineered to modulate expression of individual exons through the paired guide RNAs for alternative exon removal (pgFARM) to enable functional testing of individual exons that may be part of AS pathways [136]. Further application of the CRISPR system has been extended to the CRISPR (Artificial Splicing Factors (CASFx) system which provides the ability to induce AS events onto target regions, to mimic and functionally characterize specific splice isoforms [137].

For similar experimental goals, isoforms can be over-expressed to understand the contribution an isoform has in driving cell fate. Currently, large-scale overexpression screens have been applied to TFs by using a strong promoter to profile individual TF isoforms driving stem cell differentiation[39]. Creation of “ORFeome” libraries have been created to functionally describe and characterize human isoforms [48]. However, large-scale overexpression screens to functionally interrogate the role of isoforms in driving stem cell differentiation have yet to be performed, but similar methods could be applied to elucidate the greater role of isoforms in driving stem cell fate. These screens could complement massively parallel reporter assays for screening candidate functional exonic or cryptic splice variants associated with stem cell mediated traits[138,139]. A range of splicing effects could be evaluated including, intron retention, exon exclusion, 5’ or 3’ UTR usage. More recent high-content cell imaging assays could be combined with these molecular functional screens[140]. By comparing different molecular splicing effects with gene expression and protein abundances and linking these to stem cell phenotypes, we expect these assays to dissect the complex architecture for a range of diseases such as cancers, cardiovascular disease, and autoimmune diseases.

Unresolved questions in the field and future directions

Stem cells are powerful model systems that mirror the processes of human development. Methods to date have focused on linking transcript expression signatures to stem cell phenotype, but emerging methods that combine multiple facets of regulation, such as transcriptional and splicing networks during stem cell differentiation should capture important regulatory programs previously missed. The interrelatedness of transcriptional and splicing networks, during biochemical regulation, as well as between their network components, during stem cell differentiation necessitates a multi-faceted approach to understand.

Funding

This study was funded by grants T32HL007284 to MMM and R35GM142647-02 to GMS. M. N. K-M is partly supported by the following grants: NIH/NHLBI 5R01HL157780-02, NIH/NHLBI 1R01HL150589-01A1, Additional Ventures Single Ventricle Research Fund 1048010, Additional Ventures Expansion Award 988513, and American Heart Association 20TPA35490206. C.L.M. is supported by the following grants: NIH/NHLBI R01HL148239, NIH/NHLBI R01HL164577, and Fondation Leducq “PlaqOmics” 18CVD02.

REFERENCES

Papers of particular interest, published recently, have been highlighted as:

•Of importance

••Of major importance

1. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998;282:1145–7. [PubMed: 9804556]
2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–72. [PubMed: 18035408]
3. Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, et al. Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem Cell Reports. 2016;7:110–25. [PubMed: 27293150]
4. Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. Nat Rev Mol Cell Biol. 2016;17:183–93. [PubMed: 26883003]
5. Creighton H, Waddington CH. The strategy of the genes. AIBS Bull. Oxford University Press (OUP); 1958;8:49.
6. Panina Y, Karagiannis P, Kurtz A, Stacey GN, Fujibuchi W. Human Cell Atlas and cell-type authentication for regenerative medicine. Exp Mol Med. 2020;52:1443–51. [PubMed: 32929224]
7. Macarthur BD, Ma'ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol. 2009;10:672–81. [PubMed: 19738627]
8. Rowe RG, Daley GQ. Induced pluripotent stem cells in disease modelling and drug discovery. Nat Rev Genet. 2019;20:377–88. [PubMed: 30737492]
9. Sharma A, Sances S, Workman MJ, Svendsen CN. Multi-lineage Human iPSC-Derived Platforms for Disease Modeling and Drug Discovery. Cell Stem Cell. 2020;26:309–29. [PubMed: 32142662]
10. Dejana E, Hirschi KK, Simons M. The molecular basis of endothelial cell plasticity. Nat Commun. 2017;8:14361. [PubMed: 28181491]
11. Sriram G, Tan JY, Islam I, Rufaihah AJ, Cao T. Efficient differentiation of human embryonic stem cells to arterial and venous endothelial cells under feeder- and serum-free conditions. Stem Cell Res Ther. 2015;6:261. [PubMed: 26718617]
12. Kennedy M, Awong G, Sturgeon CM, Ditadi A, LaMotte-Mohs R, Zúñiga-Pflücker JC, et al. T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. Cell Rep. 2012;2:1722–35. [PubMed: 23219550]
13. Bargehr J, Ong LP, Colzani M, Davaapil H, Hofsteen P, Bhandari S, et al. Epicardial cells derived from human embryonic stem cells augment cardiomyocyte-driven heart regeneration. Nat Biotechnol. 2019;37:895–906. [PubMed: 31375810]
14. Kim J, Koo B-K, Knoblich JA. Human organoids: model systems for human biology and medicine. Nat Rev Mol Cell Biol. 2020;21:571–84. [PubMed: 32636524]
15. Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nat Rev Drug Discov. Springer Science and Business Media LLC; 2017;16:115–30. [PubMed: 27980341]
16. Yeo J-C, Ng H-H. The transcriptional regulation of pluripotency. Cell Res. 2013;23:20–32. [PubMed: 23229513]
17. Mahla RS. Stem Cells Applications in Regenerative Medicine and Disease Therapeutics. Int J Cell Biol. 2016;2016:6940283. [PubMed: 27516776]
18. Young RA. Control of the embryonic stem cell state. Cell. 2011;144:940–54. [PubMed: 21414485]
19. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420:520–62. [PubMed: 12466850]
20. Lesha E, George H, Zaki MM, Smith CJ, Khoshakhlagh P, Ng AHM. A Survey of Transcription Factors in Cell Fate Control. Methods Mol Biol. 2023;2594:133–41. [PubMed: 36264493]
21. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature. 2008;455:627–32. [PubMed: 18754011]
22. Yamanaka S Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors. Cell Prolif. 2008;41 Suppl 1:51–6. [PubMed: 18181945]

23. Vierbuchen T, Wernig M. Molecular roadblocks for cellular reprogramming. Mol Cell. 2012;47:827–38. [PubMed: 23020854]
24. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;175:598–9. [PubMed: 30290144]
25. Bulyk ML, Marian Walhout AJ. Gene Regulatory Networks [Internet]. Handbook of Systems Biology. 2013. p. 65–88. Available from: 10.1016/b978-0-12-385944-0.00004-6
26. Farnham PJ. Insights from genomic profiling of transcription factors. Nat Rev Genet. 2009;10:605–16. [PubMed: 19668247]
27. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 2010;20:861–73. [PubMed: 20378718]
28. Berger MF, Bulyk ML. Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. Methods Mol Biol. 2006;338:245–60. [PubMed: 16888363]
29. Rockel S, Geertz M, Maerkl SJ. MITOMI: A Microfluidic Platform for In Vitro Characterization of Transcription Factor–DNA Interaction. In: Deplancke B, Gheldof N, editors. Gene Regulatory Networks: Methods and Protocols. Totowa, NJ: Humana Press; 2012. p. 97–114.
30. Galas DJ, Schmitz A. DNAase footprinting a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. Oxford Academic; 1978;5:3157–70. [PubMed: 212715]
31. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21.29.1–21.29.9.
32. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316:1497–502. [PubMed: 17540862]
33. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife [Internet]. 2017;6. Available from: 10.7554/eLife.21856
34. Oh Y, Jang J. Directed Differentiation of Pluripotent Stem Cells by Transcription Factors. Mol Cells. 2019;42:200–9. [PubMed: 30884942]
35. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell. 1987;51:987–1000. [PubMed: 3690668]
36. Aragon JW, Hirschi KK. Endothelial Cell Differentiation and Hemogenic Specification. Cold Spring Harb Perspect Med [Internet]. 2022; Available from: 10.1101/cshperspect.a041164
37. Ng AHM, Khoshakhlagh P, Rojo Arias JE, Pasquini G, Wang K, Swiersy A, et al. A comprehensive library of human transcription factors for cell fate engineering. Nat Biotechnol. 2021;39:510–9. [PubMed: 33257861] ¨¨This study developed a library of human transcription factor open reading frame clones to profile the role of both transcription factor genes and isoforms in promoting cell differentiation. This was one of the first studies to study the role of TFs in defining cell fate at high throughput.
38. Wilkinson AC, Nakauchi H, Göttgens B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. Cell Syst. 2017;5:319–31. [PubMed: 29073372]
39. Joung J, Ma S, Tay T, Geiger-Schuller KR, Kirchgatterer PC, Verdine VK, et al. A transcription factor atlas of directed differentiation. Cell. 2023;186:209–229.e26. [PubMed: 36608654] ¨¨This study used single cell RNA-seq as a readout to profile transcription factor isoforms at a systems-level in its ability to drive cell-specific gene expression programs each TFs was profiled in terms of its ability to drive diverse cell-type-specific gene expression programs.
40. Kinney MA, Vo LT, Frame JM, Barragan J, Conway AJ, Li S, et al. A systems biology pipeline identifies regulatory networks for stem cell engineering. Nat Biotechnol. 2019;37:810–8. [PubMed: 31267104]
41. Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. CellNet: network biology applied to stem cell engineering. Cell. 2014;158:903–15. [PubMed: 25126793]
42. Rackham OJ, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, et al. A predictive computational framework for direct reprogramming between human cell types. Nat Genet. 2016;48:331–5. [PubMed: 26780608]
43. Wells CA, Choi J. Transcriptional Profiling of Stem Cells: Moving from Descriptive to Predictive Paradigms. Stem Cell Reports. 2019;13:237–46. [PubMed: 31412285]

44. Stamm S, Ben-Ari S, Rafalska I, Tang YS, Zhang ZY, Toiber D, et al. Function of alternative splicing. Gene. 2005;344:1–20. [PubMed: 15656968]
45. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. Gene. 2013;514:1–30. [PubMed: 22909801]
46. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010;463:457–63. [PubMed: 20110989]
47. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5. [PubMed: 18978789]
48. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. Cell. 2016;164:805–17. [PubMed: 26871637]
49. Wachtel C, Manley JL. Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Mol Biosyst. 2009;5:311–6. [PubMed: 19396366]
50. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. 2015;84:291–323. [PubMed: 25784052]
51. Ule J, Blencowe BJ. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol Cell. 2019;76:329–45. [PubMed: 31626751]
52. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A large-scale binding and functional map of human RNA-binding proteins. Nature. 2020;583:711–9. [PubMed: 32728246] ¨¨This study characterized on a systems-level the functions of RNA binding proteins(RBPs) providing insight into the dynamic RBP-RNA complexes responsible for driving gene expression programs.
53. Fu X-D, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet. 2014;15:689–701. [PubMed: 25112293]
54. He S, Valkov E, Cheloufi S, Murn J. The nexus between RNA-binding proteins and their effectors. Nat Rev Genet [Internet]. 2022; Available from: 10.1038/s41576-022-00550-0
55. Quattrone A, Dassi E. The Architecture of the Human RNA-Binding Protein Regulatory Network. iScience. 2019;21:706–19. [PubMed: 31733516]
56. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. Hum Genet. 2016;135:851–67. [PubMed: 27215579]
57. Blencowe BJ. Alternative splicing: New insights from global analyses. Cell. 2006;126:37–47. [PubMed: 16839875]
58. Marasco LE, Kornblihtt AR. The physiology of alternative splicing. Nat Rev Mol Cell Biol [Internet]. 2022; Available from: 10.1038/s41580-022-00545-z ¨ This review summarizes the recent findings related to the process of alternative splicing and the complexity of the process with increased focus on the effect splicing induces at the protein level.
59. Wright JE, Ciosk R. RNA-based regulation of pluripotency. Trends Genet. 2013;29:99–107. [PubMed: 23146412]
60. Jangi M, Sharp PA. Building robust transcriptomes with master splicing factors. Cell. 2014;159:487–98. [PubMed: 25417102]
61. Li Y, Wang D, Wang H, Huang X, Wen Y, Wang B, et al. A splicing factor switch controls hematopoietic lineage specification of pluripotent stem cells. EMBO Rep. 2021;22:e50535. [PubMed: 33319461]
62. Venables JP, Lapasset L, Gadea G, Fort P, Klinck R, Irimia M, et al. MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nat Commun. 2013;4:2480. [PubMed: 24048253]
63. Wineberg Y, Kanter I, Ben-Haim N, Pode-Shakked N, Bucris E, Bar-Lev TH, et al. Characterization of alternative mRNA splicing in cultured cell populations representing progressive stages of human fetal kidney development. Sci Rep. 2022;12:19548. [PubMed: 36380228]
64. Zhang T, Lin Y, Liu J, Zhang ZG, Fu W, Guo LY, et al. Rbm24 Regulates Alternative Splicing Switch in Embryonic Stem Cell Cardiac Lineage Differentiation. Stem Cells. 2016;34:1776–89. [PubMed: 26990106]

65. Montañés-Agudo P, Aufiero S, Schepers EN, van der Made I, Cócera-Ortega L, Ernault AC, et al. The RNA-binding protein QKI governs a muscle-specific alternative splicing program that shapes the contractile function of cardiomyocytes. Cardiovasc Res [Internet]. 2023; Available from: 10.1093/cvr/cvad007
66. Keppetipola N, Sharma S, Li Q, Black DL. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. Crit Rev Biochem Mol Biol. 2012;47:360–78. [PubMed: 22655688]
67. Xue Y, Ouyang K, Huang J, Zhou Y, Ouyang H, Li H, et al. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. Cell. 2013;152:82–96. [PubMed: 23313552]
68. Linares AJ, Lin C-H, Damianov A, Adams KL, Novitch BG, Black DL. The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. Elife. 2015;4:e09268. [PubMed: 26705333]
69. Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, et al. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. Genes Dev. 2012;26:445–60. [PubMed: 22357600]
70. Gagliardi M, Matarazzo MR. RIP: RNA Immunoprecipitation. Methods Mol Biol. 2016;1480:73–86. [PubMed: 27659976]
71. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016;13:508–14. [PubMed: 27018577]
72. Danan C, Manickavel S, Hafner M. PAR-CLIP: A Method for Transcriptome-Wide Identification of RNA Binding Protein Interaction Sites. Methods Mol Biol. 2016;1358:153–73. [PubMed: 26463383]
73. Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. Identification of RNA-protein interaction networks using PAR-CLIP. Wiley Interdiscip Rev RNA. 2012;3:159–77. [PubMed: 22213601]
74. Hafner M, Katsantoni M, Köster T, Marks J, Mukherjee J, Staiger D, et al. CLIP and complementary methods. Nature Reviews Methods Primers. Nature Publishing Group; 2021;1:1–23.
75. Castaldi PJ, Abood A, Farber CR, Sheynkman GM. Bridging the splicing gap in human genetics with long-read RNA sequencing: finding the protein isoform drivers of disease. Hum Mol Genet. 2022;31:R123–36. [PubMed: 35960994] ˙This review highlights recent advances provided through long-read sequencing platforms that enable enhanced characterization of splicing quantitative trait loci (sQTL) and the associated protein isoform to link isoforms to disease-relevant protein isoforms to offer solution to treating complex diseases.
76. López AJ. Developmental role of transcription factor isoforms generated by alternative splicing. Dev Biol. 1995;172:396–411. [PubMed: 8612959]
77. Niwa H The principles that govern transcription factor network functions in stem cells. Development [Internet]. 2018;145. Available from: 10.1242/dev.157420
78. Wang X, Dai JW. Concise Review: Isoforms of OCT4 Contribute to the Confusing Diversity in Stem Cell Biology. Stem Cells. 2010;28:885–93. [PubMed: 20333750]
79. Atlasi Y, Mowla SJ, Ziaee SAM, Gokhale PJ, Andrews PW. OCT4 Spliced Variants Are Differentially Expressed in Human Pluripotent and Nonpluripotent Cells [Internet]. Stem Cells. 2008. p. 3068–74. Available from: 10.1634/stemcells.2008-0530 [PubMed: 18787205]
80. Cheong CY, Lufkin T. Alternative Splicing in Self-Renewal of Embryonic Stem Cells. Stem Cells Int [Internet]. 2011; Available from: 10.4061/2011/560261
81. Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, Sung H-K, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell. 2011;147:132–46. [PubMed: 21924763]
82. Tatetsu H, Kong NR, Chong G, Amabile G, Tenen DG, Chai L. SALL4, the missing link between stem cells, development and cancer. Gene. 2016;584:111–9. [PubMed: 26892498]
83. Rao S, Zhen S, Roumiantsev S, McDonald LT, Yuan GC, Orkin SH. Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. Mol Cell Biol. 2010;30:5364–80. [PubMed: 20837710]

84. Chepelev I, Chen X. Alternative splicing switching in stem cell lineages. Front Biol. 2013;8:50–9.
85. Real-time qRT-PCR [Internet]. [cited 2022 Dec 2]. Available from: https://www.ncbi.nlm.nih.gov/probe/docs/techqpcr/
86. Hardwick SA, Joglekar A, Flicek P, Frankish A, Tilgner HU. Getting the Entire Message: Progress in Isoform Sequencing. Front Genet. 2019;10:709. [PubMed: 31475029]
87. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621–8. [PubMed: 18516045]
88. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13. [PubMed: 26813401]
89. Steijger T, Abril JF, Engstrom PG, Kokocinski F, Rgasp Consortium, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013;10:1177–84. [PubMed: 24185837]
90. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;11:1114–25. [PubMed: 25357241]
91. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018;34:666–81. [PubMed: 29941292]
92. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8. [PubMed: 19023044]
93. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13:278–89. [PubMed: 26542840]
94. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. Front Genet. 2019;10:426. [PubMed: 31134132]
95. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17:239. [PubMed: 27887629]
96. You Y, Prawer YDJ, De Paoli-Iseppi R, Hunt CPJ, Parish CL, Shim H, et al. Identification of cell barcodes from long-read single-cell RNA-seq with BLAZE [Internet]. bioRxiv. 2022 [cited 2022 Dec 13]. p. 2022.08.16.504056. Available from: https://www.biorxiv.org/content/10.1101/2022.08.16.504056v1.full
97. Al'Khafaji AM, Smith JT, Garimella KV, Babadi M, Sade-Feldman M, Gatzen M, et al. High-throughput RNA isoform sequencing using programmable cDNA concatenation [Internet]. bioRxiv. 2021 [cited 2022 Dec 13]. p. 2021.10.01.462818. Available from: https://www.biorxiv.org/content/biorxiv/early/2021/10/01/2021.10.01.462818
98. Arzalluz-Luque A, Salguero P, Tarazona S, Conesa A. acorde unravels functionally interpretable networks of isoform co-usage from single cell data. Nat Commun. 2022;13:1828. [PubMed: 35383181]
99. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. 2012;22:1616–25. [PubMed: 22955974]
100. Kosti I, Radivojac P, Mandel-Gutfreund Y. An integrated regulatory network reveals pervasive cross-regulation among transcription and splicing factors. PLoS Comput Biol. 2012;8:e1002603. [PubMed: 22844237]
101. Han H, Braunschweig U, Gonatopoulos-Pournatzis T, Weatheritt RJ, Hirsch CL, Ha KCH, et al. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. Mol Cell. 2017;65:539–553.e7. [PubMed: 28157508]
102. Naftelberg S, Schor IE, Ast G, Kornblihtt AR. Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem. 2015;84:165–98. [PubMed: 26034889]
103. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. Science. 2010;327:996–1000. [PubMed: 20133523]
104. Agirre E, Oldfield AJ, Bellora N, Segelle A, Luco RF. Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. Nat Commun. 2021;12:682. [PubMed: 33514745]

105. Liang Y, Xu H, Cheng T, Fu Y, Huang H, Qian W, et al. Gene activation guided by nascent RNA-bound transcription factors. Nat Commun. 2022;13:7329. [PubMed: 36443367]
106. Oksuz O, Henninger JE, Warneford-Thomson R, Zheng MM, Erb H, Overholt KJ, et al. Transcription factors interact with RNA to regulate genes [Internet]. bioRxiv. 2022 [cited 2022 Dec 7]. p. 2022.09.27.509776. Available from: https://www.biorxiv.org/content/10.1101/2022.09.27.509776v1.full
107. Bertero A, Fields PA, Ramani V, Bonora G, Yardimci GG, Reinecke H, et al. Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat Commun. 2019;10:1538. [PubMed: 30948719]
108. Lu Y, Loh Y-H, Li H, Cesana M, Ficarro SB, Parikh JR, et al. Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. Cell Stem Cell. 2014;15:92–101. [PubMed: 24813856]
109. Zhu L, Choudhary K, Gonzalez-Teran B, Ang Y-S, Thomas R, Stone NR, et al. Transcription Factor GATA4 Regulates Cell Type-Specific Splicing Through Direct Interaction With RNA in Human Induced Pluripotent Stem Cell-Derived Cardiac Progenitors. Circulation. 2022;146:770–87. [PubMed: 35938400]
110. Aydin S, Pham DT, Zhang T, Keele GR, Skelly DA, Pankratz M, et al. Genetic dissection of the pluripotent proteome through multi-omics data integration [Internet]. bioRxiv. 2022 [cited 2022 Dec 12]. p. 2022.04.22.489216. Available from: https://www.biorxiv.org/content/10.1101/2022.04.22.489216v1.abstract
111. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165:535–50. [PubMed: 27104977]
112. Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol. 2013;14:153–65. [PubMed: 23385723]
113. Tress ML, Abascal F, Valencia A. Alternative Splicing May Not Be the Key to Proteome Complexity. Trends Biochem Sci. 2017;42:98–110. [PubMed: 27712956]
114. Lindoso RS, Kasai-Brunswick TH, Monnerat Cahli G, Collino F, Bastos Carvalho A, Campos de Carvalho AC, et al. Proteomics in the World of Induced Pluripotent Stem Cells. Cells [Internet]. 2019;8. Available from: 10.3390/cells8070703 [PubMed: 31861404]
115. Gundry RL, Burridge PW, Boheler KR. Pluripotent stem cell heterogeneity and the evolving role of proteomic technologies in stem cell biology. Proteomics. 2011;11:3947–61. [PubMed: 21834136]
116. Wang J, Trowbridge JJ, Rao S, Orkin SH. Proteomic studies of stem cells. StemBook. Cambridge (MA): Harvard Stem Cell Institute; 2008.
117. van Hoof D, Krijgsveld J, Mummery C. Proteomic analysis of stem cell differentiation and early development. Cold Spring Harb Perspect Biol [Internet]. 2012;4. Available from: 10.1101/cshperspect.a008177
118. Van Hoof D, Heck AJR, Krijgsveld J, Mummery CL. Proteomics and human embryonic stem cells. Stem Cell Res. 2008;1:169–82. [PubMed: 19383398]
119. Zaro BW, Noh JJ, Mascetti VL, Demeter J, George B, Zukowska M, et al. Proteomic analysis of young and old mouse hematopoietic stem cells and their progenitors reveals post-transcriptional regulation in stem cells. Elife [Internet]. 2020;9. Available from: 10.7554/eLife.62210
120. Sabatier P, Beusch CM, Saei AA, Aoun M, Moruzzi N, Coelho A, et al. An integrative proteomics method identifies a regulator of translation during stem cell maintenance and differentiation. Nat Commun. 2021;12:6558. [PubMed: 34772928]
121. Phanstiel D, Brumbaugh J, Berggren WT, Conard K, Feng X, Levenstein ME, et al. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. Proc Natl Acad Sci U S A. 2008;105:4093–8. [PubMed: 18326628]
122. She Y-M, Rosu-Myles M, Walrond L, Cyr TD. Quantification of protein isoforms in mesenchymal stem cells by reductive dimethylation of lysines in intact proteins. Proteomics. 2012;12:369–79. [PubMed: 22144135]
123. Blakeley P, Siepen JA, Lawless C, Hubbard SJ. Investigating protein isoforms via proteomics: a feasibility study. Proteomics. 2010;10:1127–40. [PubMed: 20077415]

124. Parrotta EI, Scalise S, Taverna D, De Angelis MT, Sarro G, Gaspari M, et al. Comprehensive proteogenomic analysis of human embryonic and induced pluripotent stem cells. J Cell Mol Med. 2019;23:5440–53. [PubMed: 31237115]
125. Miller RM, Jordan BT, Mehlferber MM, Jeffery ED, Chatzipantsiou C, Kaur S, et al. Enhanced protein isoform characterization through long-read proteogenomics. Genome Biol. BioMed Central; 2022;23:1–28.
126. Sheynkman GM, Shortreed MR, Cesnik AJ, Smith LM. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. Annual Review of Analytical Chemistry, Vol 9. 2016;9:521–45.
127. Wang X, Slebos RJC, Wang D, Halvey PJ, Tabb DL, Liebler DC, et al. Protein identification using customized protein sequence databases derived from RNA-Seq data. J Proteome Res. 2012;11:1009–17. [PubMed: 22103967]
128. Mehlferber MM, Jeffery ED, Saquing J, Jordan BT, Sheynkman L, Murali M, et al. Characterization of protein isoform diversity in human umbilical vein endothelial cells via long-read proteogenomics. RNA Biol. 2022;19:1228–43. [PubMed: 36457147] ˙This study used long-read sequencing and matched mass-spectrometry data on an endothelial cell type to characterize protein isoform diversity and develop an atlas of novel isoforms of relevance to endothelial phenotypes.
129. Endoh T, Ohtsuki T. Cellular siRNA delivery using cell-penetrating peptides modified for endosomal escape. Adv Drug Deliv Rev. 2009;61:704–9. [PubMed: 19383521]
130. Dana H, Chalbatani GM, Mahmoodzadeh H, Karimloo R, Rezaiean O, Moradzadeh A, et al. Molecular Mechanisms and Biological Functions of siRNA. Int J Biomed Sci. 2017;13:48–57. [PubMed: 28824341]
131. Prinos P, Garneau D, Lucier J-F, Gendron D, Couture S, Boivin M, et al. Alternative splicing of SYK regulates mitosis and cell survival. Nat Struct Mol Biol. 2011;18:673–9. [PubMed: 21552259]
132. Corey DR, Abrams JM. Morpholino antisense oligonucleotides: tools for investigating vertebrate development. Genome Biol. 2001;2:REVIEWS1015.
133. Moulton JD. Using morpholinos to control gene expression. Curr Protoc Nucleic Acid Chem. 2007;Chapter 4:Unit 4.30.
134. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, et al. RNA targeting with CRISPR-Cas13. Nature. 2017;550:280–4. [PubMed: 28976959]
135. Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, et al. RNA editing with CRISPR-Cas13. Science. 2017;358:1019–27. [PubMed: 29070703]
136. Thomas JD, Polaski JT, Feng Q, De Neef EJ, Hoppe ER, McSharry MV, et al. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. Nat Genet. 2020;52:84–94. [PubMed: 31911676]
137. Du M, Jillette N, Zhu JJ, Li S, Cheng AW. CRISPR artificial splicing factors. Nat Commun. 2020;11:2973. [PubMed: 32532987] ¨This study demonstrates the application of CRISPR Artificial Splicing Factors that can recapitulate induction of splicing patterns of target genes.
138. Rhine CL, Neil C, Wang J, Maguire S, Buerer L, Salomon M, et al. Massively parallel reporter assays discover de novo exonic splicing mutants in paralogs of Autism genes. PLoS Genet. 2022;18:e1009884. [PubMed: 35051175]
139. Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, et al. Pathogenic variants that alter protein code often disrupt splicing. Nat Genet. 2017;49:848–55. [PubMed: 28416821]
140. Veschini L, Sailem H, Malani D, Pietiäinen V, Stojiljkovic A, Wiseman E, et al. High-Content Imaging to Phenotype Human Primary and iPSC-Derived Cells. Methods Mol Biol. 2021;2185:423–45. [PubMed: 33165865]

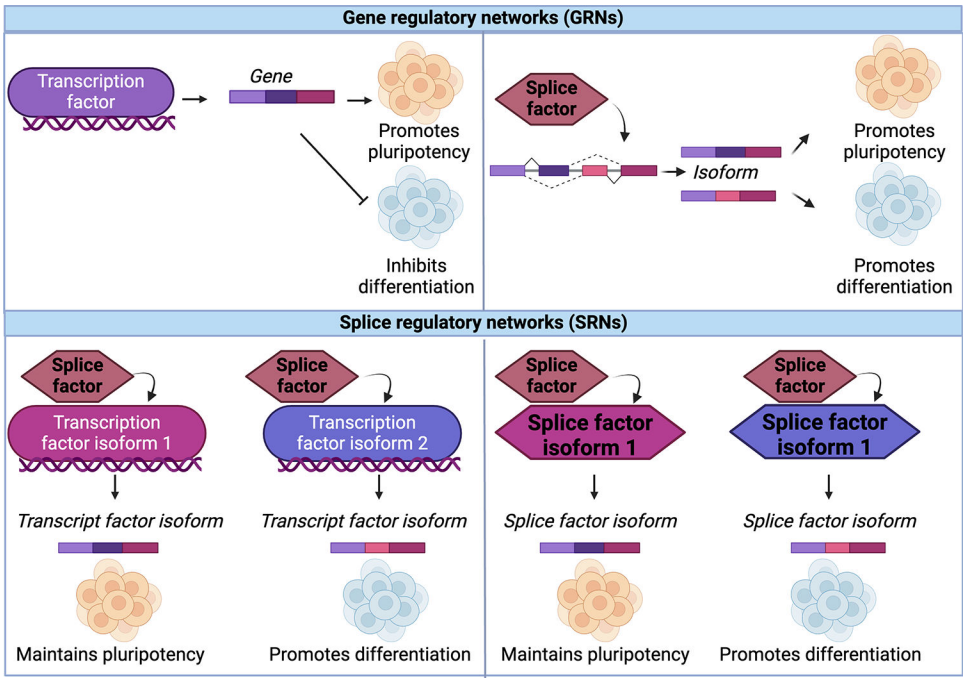

Figure 1.
Demonstration of gene regulatory networks and splice regulatory networks