Telescripting is a fundamental cotranscriptional gene regulation process that relies on U1 snRNP (U1) to suppress premature 3′-end cleavage and polyadenylation (PCPA) in RNA polymerase II (Pol II) transcripts, which is necessary for full-length transcription of thousands of protein-coding (pre-mRNAs) and long noncoding (lncRNA) genes. Like U1 role in splicing, telescripting requires U1 snRNA base-pairing with nascent transcripts. Inhibition of U1 base-pairing with U1 snRNA antisense morpholino oligonucleotide (U1 AMO) mimics widespread PCPA from cryptic polyadenylation signals (PASs) in human tissues, including PCPA in introns and last exons’ 3′-untranslated regions (3′ UTRs). U1 telescripting–PCPA balance changes generate diverse RNAs depending on where in a gene it occurs. Long genes are highly U1-telescripting-dependent because of PASs in introns compared to short genes. Enrichment of cell cycle control, differentiation, and developmental functions in long genes, compared to housekeeping and acute cell stress response genes in short genes, reveals a gene size–function relationship in mammalian genomes. This polarization increased in metazoan evolution by previously unexplained intron expansion, suggesting that U1 telescripting could shift global gene expression priorities. We show that modulating U1 availability can profoundly alter cell phenotype, such as cancer cell migration and invasion, underscoring the critical role of U1 homeostasis and suggesting it as a potential target for therapies. We describe a complex of U1 with cleavage and polyadenylation factors that silences PASs in introns and 3′ UTR, which gives insights into U1 telescripting mechanism and transcription elongation regulation.

U1 snRNP comprises a single noncoding small nuclear RNA (164 nt in human) and 10 proteins (U1-specific U1-70K, U1A, U1C, and seven Sm proteins common to spliceosomal snRNPs). RNA sequencing (RNA-seq) from cells transfected with U1 snRNA antisense morpholino oligonucleotide (U1 AMO) titrated to mask all or nearly all U1 snRNA 5′ sequence (“high U1 AMO”), compared to control nonspecific AMO (cAMO), showed unfamiliar and striking “Z patterns” in thousands of genes, consisting of RNAs extending several kilobases (typically, ∼1–3 kb and up to tens of kilobases) from the transcription start site (TSS) that end abruptly in an intron (Fig. 1). At their ends these RNAs had 3′-poly(A)s specified by canonical PAS hexamers, AAUAAA and variants thereof, indistinguishable from classical PASs at the ends of full-length pre-mRNAs and lncRNAs (Proudfoot and Brownlee 1976; Tian and Manley 2017). The widespread PCPA elicited by high U1 AMO revealed that U1 is a PCPA suppressor. Importantly, detection of 3′-poly(A) tags at U1 AMO-induced PCPA positions in normal tissues in human and other organisms (Derti et al. 2012; Oh et al. 2017) demonstrated that PCPA and U1 telescripting are physiological processes, and U1 AMO is a useful tool to study it.

As expected, U1 AMO also caused widespread splicing inhibition, evident in intron retention in many introns that are not PCPAd. However, U1 telescripting appears to be additional and separable from U1’s role in splicing. As PCPA frequently occurs in introns well before their 3′ss is transcribed, PCPA is not secondary to splicing inhibition (of the same intron) and U1 telescripting is U1-specific and may not require other spliceosomal snRNPs (U2,
U4, U5, U6, or the minor snRNPs U11, U12, U4atac, U6atac) (Kaida et al. 2010). Although telescripting and splicing require U1 base-pairing to the nascent transcript and U1 bound at 5′ss can function in both, telescripting can also be supplied by U1 base-paired to sequences that cannot function as 5′ss, giving U1 telescripting more potential sites (Berg et al. 2012).

Pol II chromatin immunoprecipitation-sequencing (ChIP-seq) maps complemented the nascent RNA maps (Fig. 1; Gilmour and Lis 1984; Oh et al. 2017), indicating that U1 telescripting acts cotranscriptionally, as opposed to post-transcriptionally (e.g., by reprocessing from full-length transcripts). Like CPA at normal 3′ends (the ultimate PASs), Pol II ChIP-seq tapered off within a few kilobases from PCPA locations, a gradual post-3′-end transcription termination zone (TZ) (Fong et al. 2015) consistent with the torpedo termination model (Connelly and Manley 1988; Proudfoot 2016). Pol II occupancy downstream from TZs through the rest of the gene was sharply reduced or eliminated. Notably, U1 AMO did not inhibit transcription initiation (and promoter-proximal pause Pol II release into elongation); rather, transcription continued to stream into genes and frequently increased up to PCPA points, evident by high pol II occupancy from the TSSs to PCPA points in U1 AMO compared to control. Thus, full-length Pol II transcription elongation in most genes in vertebrates depends on averting PCPA termination at PAS checkpoints, which U1 telescripting provides.

**BIOLOGICAL ROLES OF U1 TELESCRIPTING**

Technically, PAS selection in genes that have more than one can be classified as alternative polyadenylation (APA). Although APA, involving choice among tandem PASs in terminal exons 3′-untranslated regions (3′UTRs), and a few instances of APA in gene bodies had been known previously (Wang et al. 2008; Hartmann and Válcarcel 2009; Di Giammartino et al. 2011), the PCPA-U1 telescripting phenomenon as a major transcription control was unexpected.

**Transcriptome Control and Proteome Diversification**

RNA-seq and PCPA maps developed from Z-pattern-detecting algorithms and nongenomic 3′-poly(A) reads at various U1 AMO doses helped identify key telescripting features and its biological roles. PCPA in a gene can be complete, effectively shutting down full-length transcription or partial (in some genes even at high U1 AMO), affecting a fraction of a gene’s transcripts (e.g., RAB7A and E2F3; Fig. 1), and can occur at multiple PASs in the same gene. Any PCPA is at the expense of full-length transcription; however, it can potentially generate diversity of RNAs and proteins. Figure 2 illustrates various outcomes of PCPA at different locations that U1 telescripting can regulate. The relative abundance of the different isoforms can be manipulated with U1 AMO dose (and correspondingly decreases available U1), which increases usage of TSS-proximal PAS with higher U1 AMO, consistent with a cotranscriptional process whereby the first PAS encountered by Pol II that is unprotected by U1 is PCPAed. Pervasive PCPA at multiple cryptic PASs in introns counters the textbook view that PASs function primarily in the 3′UTR.

PCPA in the 1–3 kb from the first 5′ss (which is <0.5 kb from the TSS) in intron1 makes polyadenylated RNAs that are generally rapidly degraded by nuclear exosomes.
These RNAs are difficult to detect by standard total RNA-seq but are readily detected if exosomes are inhibited or with high U1 AMO, which results in massive PCPA that overwhelms exosomes. PCPA detection is also enhanced by selective RNA-seq of nascent RNAs. This, and 3′-poly(A) tags in tissues, show TSS-proximal PCPA, indicating that U1 telescripting is a normal source of transcription attrition (Oh et al. 2017; So et al. 2019). Stealthy PCPA likely play a significant role in gene expression down-regulation. Conditions in which such drastic PCPA occurs in nature have not been described, although we envision toxins, viruses, and other pathogens evolved mechanisms to inhibit U1 telescripting as means to shut off host mRNA synthesis.

In a few cases, TSS-proximal PCPAed RNAs are relatively stable and have a translation open reading frame (ORF) extending from exon1 into intron1, where PCPA occurs. These can be functional mRNAs that encode the same amino terminus as the full-length mRNA, but a different carboxyl terminus and function, such as epidermal growth factor receptor (EGFR) and related receptors (Vorlová et al. 2011).

**Promoter Directionality**

TSS-proximal PCPA occurs naturally in noncoding upstream antisense RNAs (uaRNAs [Almada et al. 2013], also known as PROMPTs [Preker et al. 2008; Ntini et al. 2013]) transcribed from divergent Pol II promoters in many genes (Fig. 2). uaRNAs tend to have high ratios of PASs to U1 binding sites, causing their PCPA and rapid elimination (Almada et al. 2013; Ntini et al. 2013). In contrast, over a similar distance from the TSS (1–3 kb), RNA in the sense direction have high ratios of U1 base-pairing sites to PASs. Preferential uaRNA pruning contributes to promoter directionality in the sense, protein coding transcripts thereby enhancing production of mRNAs. These observations highlight the general role of PCPA and telescripting in shaping the transcriptome.

**Short mRNA Isoform Switching, Cell Stimulation, and Oncogenicity**

Despite U1’s high abundance (~1,000,000 per cell in HeLa [Baserga and Steitz 1993]), even small changes in available U1 levels have profound effects that are not readily detected in high U1 AMO because of toxicity and drastic PCPA closer to TSS. Low U1 AMO (masking <15% of U1), in addition to low level PCPA, causes widespread 3′ UTR shortening (increased usage of proximal PAS in the last exon) and shifts to shorter mRNA isoforms (Berg et al. 2012), that are hallmarks of and play an important role in stimulated states in immune cells and neurons, cell proliferation, and cancer (Niibori et al. 2007; Flavell et al. 2008; Sandberg et al. 2008; Mayr and Bartel 2009; Lianoglou et al. 2013; Xia et al. 2014; de Morree et al. 2019). Although it maintains mRNA’s full-length coding sequence, 3′ UTR shortening removes binding sites for RBPs and miRNAs that regulate translation, stability and localization thereby altering these functions.

Low U1 AMO dose-dependently mimics short mRNA isoform switching because of intronic PCPA induced by transient transcription up-regulation in stimulated neurons (Berg et al. 2012). For example, PCPA in an intron in homer-1, which encodes a synaptogenesis scaffold protein,
creates a short (carboxy-terminal deleted) isoform with antagonistic function that buffers overstimulation and prevents epilepsy (Niibori et al. 2007). Conversely, U1 overexpression suppressed mRNA isoform switching in stimulated neurons (Berg et al. 2012). Short isoform switching in stimulated neurons tracked with the transient burst in transcription, during which absolute U1 levels did not decrease, which could create transient telerecording deficit because of increased competition for U1 (Berg et al. 2012). U1 levels inevitably lag behind rapidly up-regulated transcription because U1 biogenesis entails elaborate RNP assembly by the SMN complex (So et al. 2017). These findings suggest that transient U1 shortage (for telerecording) is a built-in aspect of acute response. The CPA factors (CPAFs) and splicing factors (SFs) may be similarly affected. It remains to be determined how the critical CPAF–telerecording balance is regulated.

Down-regulation of two CPAFs, CFIm25 and CFIm68, elicits widespread 3′ UTR shortening and is oncogenic (Masamha et al. 2014; Masamha and Wagner 2017). Interestingly, low U1 AMO also dose-dependently increased cancer cells’ migration and invasion in vitro by up to 500% (Oh et al. 2020). U1 overexpression had the opposite effect. U2 AMO and high U1 AMO were toxic in <24 h. In addition to 3′ UTR length changes (generally shortening in low U1 AMO and lengthening in U1 overexpression), numerous transcriptome changes that could contribute to the altered phenotype were observed, including alternative splicing, and mRNA expression levels of proto-oncogenes and tumor suppressors. These findings reveal an unexpected role for U1 homeostasis (available U1 and its activity) in relation to oncogenic and activated cell states and suggest U1 as a potential target for their modulation (Oh et al. 2020). Recent reports of oncogenic mutations in U1 5′ sequence in cancer patients (Shuai et al. 2019; Suzuki et al. 2019) support this notion.

**Selective Telerecording Dependence of Long Introns: A New Layer of Gene Regulation**

Nearly 1,000 human genes show no evidence of PCPA at any U1 AMO dose (Oh et al. 2017). The PCPAed and non-PCPAed genes have strikingly different lengths (medians of 39 kb vs. 14.2 kb, respectively; overall expressed genes 22.8 kb) derived almost entirely from intron length. Thus, U1 telerecording is selectively required for full-length transcription of long genes, an unprecedented gene regulation mechanism based on gene length. Remarkably, at high U1 AMO, many small PCPA-resistant genes (median 6.8 kb) were up-regulated, spliced robustly, and produced more mRNA and protein (up to fivefold) (Oh et al. 2017). Gene ontology revealed that short PCPA-insensitive genes are enriched in primary response genes that are induced during acute cell stimulation and necessary to enhance cell survival and adapt to adverse environmental changes, including MYC (proto-oncogene), CYR61 (chemo-resistance), and GADD45B. These primary response functions are underrepresented in long genes, which are instead more highly expressed in differentiated tissues, and enriched in specialized functions. Tumor suppressors, DNA damage responsive, neuronal, and developmental genes are among the longest (Bertagnolli et al. 2013; Gabel et al. 2015), which makes them highly susceptible to PCPA. This surprising gene size–function relationship creates a stratified system that depends on U1 telerecording and can be actuated to rapidly shift expression priorities by cotrancriptional PCPA.

Tracking the gene groups in divergent metazoans, showed that gene–size function polarization increased by selective intron expansion in evolution, suggesting it was beneficial. We proposed that transient cotrancriptional PCPA in long genes decreases competition for transcription and RNA processing factors that normally limit small genes’ expression, boosting their up-regulation in acute phase and facilitating shifts in polarized priorities (e.g., growth vs. differentiation) (Oh et al. 2017; Venters et al. 2019).

**MECHANISM**

**A Complex of U1 with CPAFs Suppresses PASs in Introns**

A general U1 telerecording model (Berg et al. 2012; Venters et al. 2019) incorporating the observations described above and in publications cited therein proposed that CPAFs bind nascent Pol II transcripts cotrancriptionally, aided by interactions with Pol II carboxy-terminal domain (CTD) and various RNA-binding proteins (RBPs). This recruitment is initiated early on near the TSS (Dantonel et al. 1997; McCracken et al. 1997; Calvo and Manley 2003) and could leave the first PAS in the RNA vulnerable to PCPA. U1 base-paired at the first 5′ss, which is enhanced by the 5′-cap binding complex (CBC), prevents PCPA from downstream PASs nearby in the intron. However, the effective range of U1 telerecording seems to be limited (<1 kb, possibly 2–3 kb). Thus, U1 base-paired to 5′ss would be insufficient to suppress PASs in longer introns (as in most mammalian genes), requiring additional U1 binding near PASs in the intron, which may explain the need for extra U1. Despite 1:1 stoichiometry with other snRNPs in U2 spliceosomes (U2, U4, U5, U6), U1 is several-fold more abundant than U4 and U6 in human cells (Baserga and Steitz 1993). Short introns have fewer PASs, stochastically, and kinetic competition from splicing lessens PCPA likelihood. Except for the 3′ UTR, exons are short (median ~145 nt in humans) and may be protected by bound splicing factors.

To understand the molecular mechanism of U1 telerecording, it is required to know where U1 and CPAFs are bound on nascent transcripts relative to PCPA locations and to identify potential interactions between U1 and CPAFs. High degeneracy of U1 base-pairing and numerous PAS variants required their binding sites to be determined experimentally. These objectives were addressed by minimal RNA–protein and protein–protein cross-linking (XL) in cells with formaldehyde, followed by RNase digestion (to <150 nt), and high stringency parallel immunopurifications (IPs) with multiple antibodies to capture U1 and CPAF interactions. We have applied this procedure
to the U1-specific proteins, several CPAFs, other snRNPs, and splicing factors in control and U1 AMO–transfected cells and used mass spectrometry (MS) and RNA-seq to define the XLIPs’ protein composition and stoichiometry and map RNA fragments bound to them (So et al. 2019). This procedure is broadly applicable for comprehensive RNP definition in cells.

U1 and CPAF XLIPs–RNA-seq mapped to expected positions, including 5′ss and PASs in terminal exons, respectively. Additionally, U1 and CPAF XLIPs coincided with PCPA locations in introns (Fig. 3). The compositions and arrangement of U1 proteins and CPAFs and the sequences they bind in U1 snRNA and PASs, respectively, are shown in Figure 4. XLIP-seq of the U1-specific proteins and CPAFs of the three major CPA subunits (CFIm25, Fip1, and CstF64; CFIm, CPSF, and CstF, respectively; Fig. 4) coincided in the same peaks (in a 100-nt window). In contrast, essential splicing factors (e.g., SF3B1) were absent from the same peaks, but present at splice sites. XLIPs-MS uncovered a novel complex comprising U1 and CPAFs (U1–CPAFs), that binds and suppresses PASs in introns and 3′ UTRs (Figs. 3 and 4). U1–CPAFs are distinct from U1 complexes with spliceosomes. These observations suggested that U1 telescripting is mediated by direct binding to CPAFs at PASs, ruling out several alternative scenarios. U1–CPAFs as the telescripting complex can explain early studies showing that U1 prevents PAS usage in HIV-1 transcripts (Ashe et al. 1995, 1997) by U1′s base-pairing at cryptic 5′ss in the 5′ long terminal repeat (5′ LTR), and in bovine papilloma virus pre-mRNA (Gunderson et al. 1998).

Comparison of XLIPs from U1 AMO and cAMO and purification of U1–CPAFs in active/uninhibited state with biotin-labeled U1 AMO transfected into cells (triggering PCPA), showed that U1 AMO remodels, but does not disrupt, U1–CPAFs (So et al. 2019). At least two changes in

![Figure 3. Peaks of U1 and CPAF XLIPs RNA-seq colocalized at PCPA locations in introns. Genome browser views of representative PCPA genes are shown with XLIPs, poly(A)-seq (So et al. 2019), and iCLIP (Yao et al. 2012) (e.g., RAB7A and EXT1) from HeLa cells transfected with either control or U1 AMO. The major PCPA points, identified from U1 AMO-induced 5-min EU-labeled, oligo(dT)-selected RNA-seq are indicated with arrows. U1C, SF3B1, and CPAF (CFIm25, Fip1, and CstF64) XLIPs and UV iCLIP-seq (CstF64) coincide (100-nt window) with PCPA points.](image-url)
U1–CPAFs induced by U1 AMO help us understand U1 telescripting: loss of CPA-inhibitory interactions and gain of CPA-stimulating factor, represented by U1A and CFIm68, respectively. U1-free U1A and U1A bound and in specific contexts in the last exon, can inhibit CPA by binding to poly(A) polymerase (PAP) (Boelens et al. 1993; Gunderson et al. 1994; Klein Gunnewiek et al. 2000; Phillips and Gunderson 2003; Workman et al. 2014). U1A associations with CPAFs were abolished by U1 AMO, which is compatible with a role for U1A in telescripting in the context of U1–CPAFs. In contrast, U1C and U1-70K’s binding to CPAFs were not disrupted with U1 AMO and may explain why U1–CPAFs remain largely intact despite loss of U1 base-pairing. U1 AMO replaced CFIm59 binding to CFIm25 with CFIm68, CFIm68/CFIm25 is CPA-stimulating, but CFIm59/CFIm25 is not (Rüegsegger et al. 1998; Kim et al. 2010; Zhu et al. 2018). Thus, U1 is a CPA-regulating subunit of the U1–CPAFs that is held together and suppresses PASs by multiple protein–protein interactions that remain to be fully defined.

Additional U1–CPAFs associations detected in the XLIPs included the CBC/ARS2 (Giacometti et al. 2017), exosomes, transcription and nuclear export complex (TREX) (Silla et al. 2018) and other transcription factors, chromatin remodelers, and CDK12, known to phosphorylate Pol II CTD (Fig. 4). Recent studies have shown that U1 AMO induces Pol II pausing and PCPA at the first nucleosome barrier, generally in the first part of long introns (Chiu et al. 2018). These barriers thus serve U1-controlled transcription checkpoints requiring chromatin remodelers and CDK12 to get through. U1s bound at the first 5’ss and possibly additional sites upstream of the intronic checkpoints suppress PAS clusters at these nucleosome barriers, thereby regulating Pol II flow. Interestingly, CDK12 is frequently mutated in cancer and contributes to BRCAness (loss of DNA damage repair, DDR) (Cancer Genome Atlas Research Network 2011; Abeshouse et al. 2015; Quereda et al. 2019). CDK12 knockdown or its inhibition with THZ531, have also been recently shown to cause PCPA (Dubbury et al. 2018; Krajewska et al. 2019). Thus, U1 and CDK12 are both necessary for telescripting, and THZ531 provides an additional tool for future studies.

CONCLUSION

Studies on U1 telescripting add to U1’s established role in splicing, placing it at the center of gene expression regulation, in both transcription and RNA processing. It is a remarkable range of functions for one small RNP, and consequently changes in U1 homeostasis have profound impacts on cell survival and adaptation to stimuli and cell behavior. The ability to modulate these biological processes experimentally, such as with U1 AMO, should have many applications, including potential therapies. Future research on U1 telescripting mechanism and U1 abundance regulation will help realize these opportunities.

ACKNOWLEDGMENTS

We are grateful to the members of our laboratory for helpful discussions. This work was supported by the National Institutes of Health (NIH) (R01 GM112923 to G.D.). G.D. is an Investigator of the Howard Hughes Medical Institute.
