Cell

Resource

A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells

Nasun Hah,[1,2,5] Charles G. Danko,[1,3,5] Leighton Core,[1] Joshua J. Waterfall,[1] Adam Siepel,[3] John T. Lis,[1,2] and W. Lee Kraus[1,2,4,*]

[1]Department of Molecular Biology and Genetics
[2]Graduate Field of Biochemistry, Molecular and Cell Biology
[3]Department of Biological Statistics and Computational Biology
Cornell University, Ithaca, NY 14853, USA
[4]Cecil H. and Ida Green Center for Reproductive Biology Sciences and Division of Basic Research, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[5]These authors contributed equally to this work
*Correspondence: lee.kraus@utsouthwestern.edu
DOI 10.1016/j.cell.2011.03.042

SUMMARY

We report the immediate effects of estrogen signaling on the transcriptome of breast cancer cells using global run-on and sequencing (GRO-seq). The data were analyzed using a new bioinformatic approach that allowed us to identify transcripts directly from the GRO-seq data. We found that estrogen signaling directly regulates a strikingly large fraction of the transcriptome in a rapid, robust, and unexpectedly transient manner. In addition to protein-coding genes, estrogen regulates the distribution and activity of all three RNA polymerases and virtually every class of noncoding RNA that has been described to date. We also identified a large number of previously undetected estrogen-regulated intergenic transcripts, many of which are found proximal to estrogen receptor binding sites. Collectively, our results provide the most comprehensive measurement of the primary and immediate estrogen effects to date and a resource for understanding rapid signal-dependent transcription in other systems.

INTRODUCTION

The steroid hormone estrogen, acting through estrogen receptors (ERs), plays key roles in a variety of fundamental developmental and physiological processes, as well as many disease states (Deroo and Korach, 2006). Mammals express two ER isoforms, ERα and ERβ, which exhibit distinct tissue-specific expression patterns and biological roles (Deroo and Korach, 2006; Warner et al., 1999). ERs function primarily as nuclear transcription factors, which dimerize upon binding of the natural ligand, 17β-estradiol (E2), and act as potent regulators of gene expression. ERα binds to > 10,000 sites across the genome and acts to (1) promote the recruitment of coregulators that mediate posttranslational modification of histones or other transcription factors and (2) regulate the binding or activity of the RNA polymerase II (Pol II) transcriptional machinery, ultimately altering the transcriptome in estrogen-responsive cells (Acevedo and Kraus, 2004; Cheung and Kraus, 2010; Ruhl and Kraus, 2009).

Previous studies analyzing steady-state gene expression patterns in the presence and absence of E2 have failed to reveal a consistent view of the estrogen-regulated gene set. In particular, the use of expression microarrays has produced discrepancies in the numbers of estrogen-regulated genes in the widely used ERα-positive MCF-7 human breast cancer cell line, ranging from 100 to 1500 (Cheung and Kraus, 2010; Kininis and Kraus, 2008). In addition, genomic ChIP analyses of ERα and Pol II have not produced a clear picture of the estrogen-regulated gene set either. This is due, in part, to the difficulty in assigning ERα binding events to specific gene regulatory outcomes (Carroll et al., 2006; Welboren et al., 2009). Another limitation of these analyses is that they have focused on the effects of estrogen signaling on Pol II transcription, without considering potential effects on Pol I and Pol III.

A fundamental weakness that is inherent in monitoring estrogen-dependent gene expression by assessing changes in mature mRNA is that longer treatments are required to allow time for mRNA accumulation (~3–24 hr). This time allows the accumulation of transcripts from primary ERα target genes but also leads to a host of secondary transcriptional effects that are not directly mediated by ERα. To address these concerns, preliminary attempts to define the immediate transcriptional effects of estrogen signaling using the translation inhibitor cycloheximide indicated that only 20%– 30% of the genes showing changes in expression are primary targets (Lin et al., 2004). Using cycloheximide to infer primary estrogen target genes is problematic, however, because (1) cycloheximide does not inhibit the effects of noncoding regulatory RNAs on gene expression, which is becoming widely recognized as an important mechanism underlying the regulation of many genes (Krol et al., 2010), and (2) the levels of steady-state mRNA depend not only on

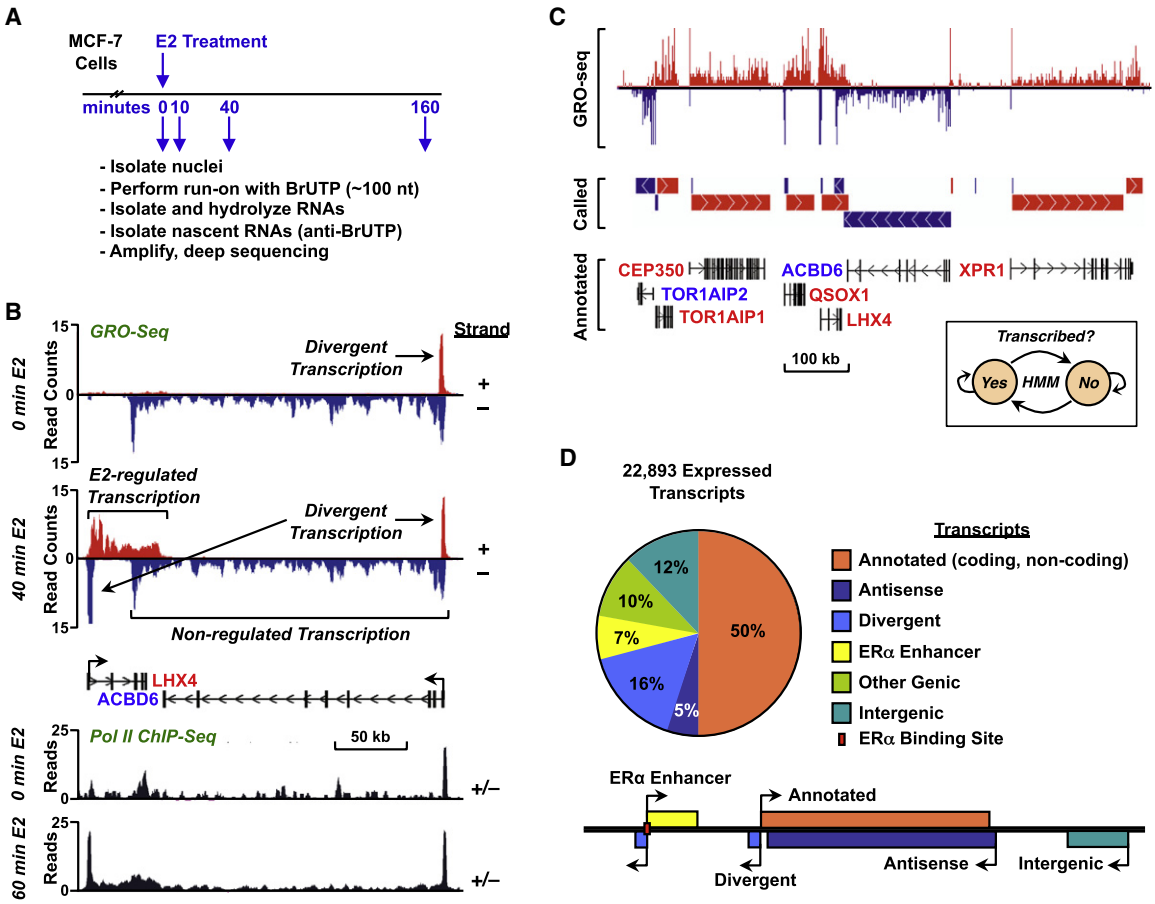

Figure 1. GRO-Seq Provides a Detailed View of the E2-Regulated Transcriptome in MCF-7 Cells

(A) Overview of the experimental set-up for GRO-seq analysis using MCF-7 cells.

(B) Genome browser view for a specific locus showing GRO-seq (top) and Pol II ChIP-seq (bottom) data illustrating the features of transcription and the effects of estrogen treatment.

(C) De novo detection of transcripts using GRO-seq data (top) and an HMM (inset). Called transcripts (middle) match well to RefSeq annotations (bottom).

(D) Classification of transcripts based on the annotation filter (Figures S1E and S1F).

See also Figure S1.

transcriptional regulation by E2, but also on the rates of elongation, pre-mRNA processing, and mRNA degradation (Widelitz et al., 1987). Due to these factors, it is clear that a new approach is required to conclusively identify primary estrogen target genes.

Here, we used global nuclear run-on and sequencing (GRO-seq) (Core et al., 2008) to identify the immediate effects of estrogen signaling on the entire transcriptome in MCF-7 cells. GRO-seq is a direct sequencing method that provides a "map" of the position and orientation of all engaged RNA polymerases across the genome at extremely high resolution, providing a direct measure of transcription. Using GRO-seq in combination with a bioinformatic approach based on hidden Markov models (HMMs), we determined all (i.e., both annotated and unannotated) genomic regions in MCF-7 cells that are transcribed by Pols I, II, and III. In addition, we identified the primary transcriptional targets of E2 signaling by focusing on short treatments (i.e., 0, 10, and 40 min) prior to the activation of secondary targets. Our unique approach has revealed many unexpected features of E2-regulation, providing the most comprehensive measurement of the primary and immediate effects of E2 signaling to date. Our results provide a model and resource for understanding rapid signal-dependent transcription in other systems.

RESULTS AND DISCUSSION

Generation of GRO-Seq Libraries from Estrogen-Treated MCF-7 Cells

To investigate the immediate effects of estrogen on the transcriptome of human cells, we treated estrogen-deprived ERα-positive MCF-7 human breast cancer cells with a short time course of 17β-estradiol (E2) (0, 10, 40, and 160 min) (Figure 1A). The estrogen-deprived MCF-7 cells continued to grow actively (Figure S1A available online), and the population of cells showed a normal distribution through the cell cycle (Figure S1B). Nuclei were isolated from two biological replicates of the E2-treated MCF-7 cells and subjected to the GRO-seq procedure to generate ~100 bp libraries representing nascent RNAs,

which were sequenced using an Illumina Genome Analyzer (Figure 1A). Short-reads were aligned to the human reference genome (hg18, NCBI36), including autosomes, the X chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1). Approximately 13 to 17 million reads were uniquely mapped to the genome for each treatment condition, and the biological replicates for each time point were highly correlated (average correlation coefficient = 0.98) (Figure S1C). GRO-seq returns data from all three RNA polymerases (Pols I, II, and III). To validate whether the reads mapping to the supposed loci transcribed by Pols I, II, and III were correlated with the activities of each individual RNA polymerase, we carried out filter binding assays with combinations of polymerase inhibitors to isolate each polymerase. As expected, the activities detected by the filter binding assays were comparable to GRO-seq product fraction, with a slight underrepresentation of the apparent fraction of Pol I transcripts by GRO-seq due to an enrichment of positions that are not mappable in the repetitive rDNA sequences (Figure S1D).

Figure 1B (top) shows a representative histogram of read counts versus genomic position for a locus containing the *LHX4* and *ACBD6* genes. Key features of the data set are illustrated in this representation, including strand-specific transcription, divergent transcription near transcription start sites (TSSs), and robust E2-dependent induction for some genes (e.g., *LHX4*). These features are not readily apparent in ChIP-seq data from the same region (Figure 1B, bottom).

Unbiased Assignment of GRO-Seq Reads to Specific Transcripts

To determine the effects of E2 on the entire transcriptome (i.e., annotated and unannotated; coding and noncoding), we developed an unbiased approach for calling transcripts using a two-state HMM. The model takes as input information about read counts across the genome and subsequently divides the genome into two states representing "transcribed" and "nontranscribed" regions (Figure 1C, inset; see Supplemental Information for additional details). An example of the input and output of this algorithm for a gene-rich region of the genome is shown in Figure 1C. The top panel shows the raw sequence read counts for the GRO-seq data, the middle panel shows the predicted transcripts, and the bottom panel shows the RefSeq annotations.

To evaluate the robustness of our approach, we compared our predicted transcript calls to existing annotations when these were available (see Supplemental Information for details). First, we determined whether our predictions reflect entire transcripts, as opposed to breaking each gene up into a series of smaller units. Then, we determined whether our approach can accurately identify nontranscribed intervals between neighboring but distinct gene annotations. We found that 90% of transcribed annotated genes overlap with exactly one transcript and that 82% of called transcripts overlapping an annotated gene do so with exactly one annotation. Together, these results suggest that our HMM-based transcript calls largely recapitulate public annotations. In many cases, our transcript calls provided new or more refined information about TSSs, 5′ exons, and transcription termination sites than was available in existing databases.

Using our algorithm, we assigned the genomic reads into 22,893 transcripts at one or more points during the E2 treatment time course, covering ~27% of the MCF-7 genome.

Transcripts called by the HMM were divided using a heuristic approach into six distinct, nonoverlapping classes, which describe the best classification of each transcript given currently available annotations and other information (Figures S1E and S1F; see Extended Experimental Procedures for additional details). The six classes of transcripts that we defined are illustrated in Figure 1D and include: (1) annotated genic and noncoding RNA transcripts, (2) antisense (genic) transcripts, (3) divergent transcripts, (4) ERα enhancer transcripts, which likely correspond to the recently described enhancer RNAs (Kim et al., 2010), (5) other transcripts falling into annotated regions but poorly matching the annotation, and (6) completely unannotated, intergenic transcription. Although each transcript is assigned to only one of these six classes, within each class, multiple annotations could be applied, allowing the accurate annotation of miRNA genes that fall inside of the introns of protein-coding genes. We found that 50.1% of the called transcripts map to previously annotated genes or noncoding RNAs, 5.2% map to antisense transcripts, 16.4% map to divergent transcripts, 6.8% map to ERα binding enhancers, and 12.1% are entirely unannotated intergenic transcripts (Figure 1D).

Extensive Estrogen-Dependent Changes in the MCF-7 Transcriptome

We determined which of the 22,893 transcripts change in response to E2 using a recently described model-based approach (Robinson et al., 2010) that detects changes beyond the global level of variation (Figure S2A; see Experimental Procedures for details). We focused our analysis on a 12 kb window at the 5′ end of each transcript, as we expect to observe changes during the first 10 min in this window that will not yet have spread to the 3′ end of longer transcripts. Surprisingly, we found that transcription of an unexpectedly large fraction (~26%) of the MCF-7 transcriptome is altered (up- or downregulated relative to the control/untreated condition) upon E2 treatment for at least one point in the time course (Figure 2A; comparisons are relative to the untreated condition). Large fractions of the genome are regulated even for the short treatments used in our experiments, strongly suggesting that these are direct actions of ERα. For example, at 10 min of E2 treatment, almost 10% of the MCF-7 cell transcriptome was significantly regulated at a false discovery rate of 0.1% (Figure 2B). Another surprising finding concerns the dynamics of regulation for up- and downregulated transcripts. Through 40 min of E2 treatment, the time point at which the largest number of transcripts were regulated in our analyses, roughly equal numbers were upregulated and downregulated, but by 160 min ~75% of the transcripts were downregulated (Figure 2B). Those transcripts showing regulation at 10 or 40 min represent the most comprehensive and accurate definition of the immediate transcriptional targets of the estrogen-signaling pathway described to date.

Next, we examined the regulation of the different classes of transcripts in greater detail. Annotated protein-coding and functional RNA transcripts as a group, as well as those unannotated

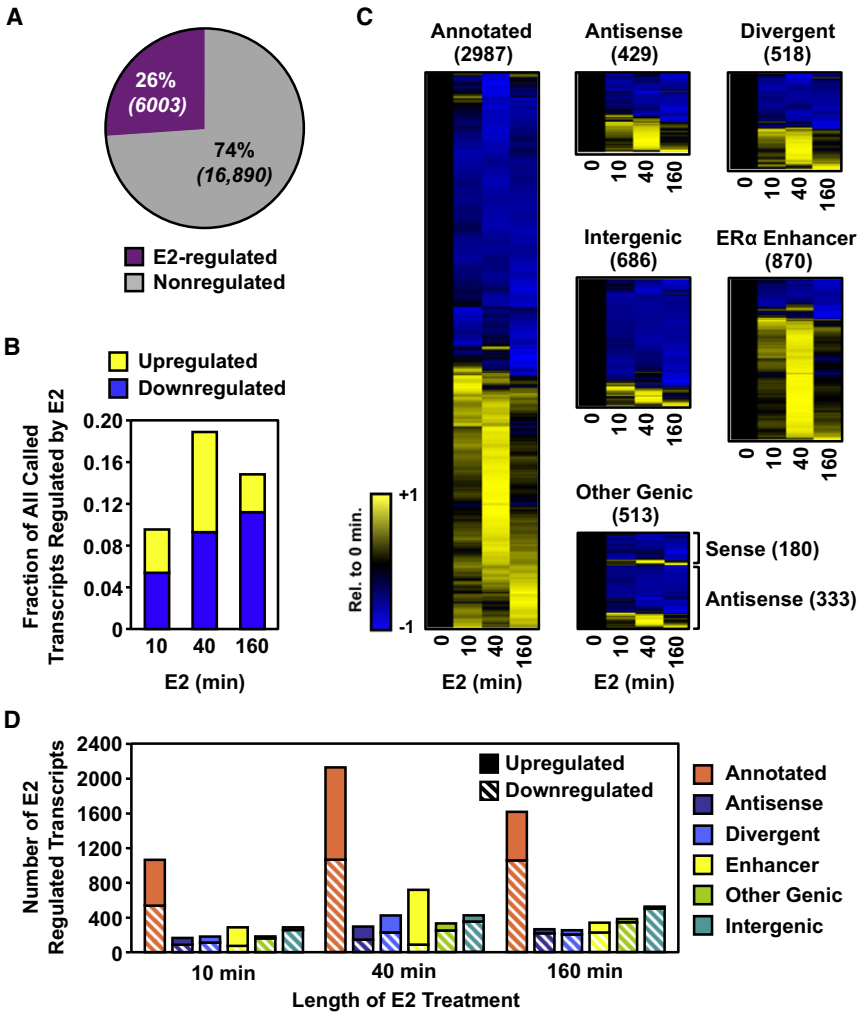

Figure 2. A Large Fraction of the MCF-7 Transcriptome Is Regulated by Estrogen

(A) The fraction of all transcripts that are regulated by E2 at any time point.

(B) The fraction of all transcripts that are up- or downregulated by E2 at the time point shown.

(C) Heatmap representations of time-dependent regulation by E2 for each transcript class. Values are centered and scaled to the 0 min time point.

(D) The fraction of each class of transcript that is up- or downregulated by E2 at each time point. See also Figure S2.

transcripts with possible roles in gene regulation (e.g., divergent and antisense), had approximately equal numbers of upregulated and downregulated transcripts at 40 min (Figures 2C and 2D). In contrast, the ERα enhancer transcripts were predominately upregulated, whereas the intergenic transcripts were predominantly downregulated. Together, these results suggest a coordinated transcriptional response in which E2 signaling directs the transcriptional machinery from intergenic regions to those more critical to the estrogen response. In addition, they give a fundamentally different view of estrogen-regulated gene expression than has been obtained using expression microarrays, especially with respect to the timing, magnitude, and extent of regulation.

Regulation of Unannotated Noncoding Transcripts by Estrogen: Divergent, Antisense, and Intergenic Transcripts

Our GRO-seq data revealed extensive estrogen regulation of a large set of unannotated noncoding transcripts, including divergent, antisense, and intergenic transcripts. Although the functions of these transcripts are largely unknown, their regulation by E2 suggests a role in estrogen-dependent transcriptional responses. The production and accumulation of divergent transcripts were first documented in recent studies using high-throughput genome-wide sequencing approaches with human fibroblasts (Core et al., 2008) and mouse embryonic stem cells (Seila et al., 2008). Divergent transcripts are transcribed in the opposite direction from primary transcripts at the promoters of transcribed genes and are also produced at enhancers (e.g., eRNAs; Kim et al., 2010) and other unannotated regions that are transcribed. The function of divergent transcripts is unknown, but their production has been suggested to promote an open chromatin architecture at promoters through the generation of a nucleosome-free region or negative superhelical tension (Core et al., 2008; Seila et al., 2008, 2009). We identified 518 divergent transcripts associated with the promoters of protein-coding genes, enhancers, and other unannotated transcribed regions that are regulated by E2 for at least one time point (FDR q value < 0.001). Using these annotations, we tested whether production of a given E2-regulated divergent transcript correlates with the synthesis of the corresponding primary transcript. To do so, we tested 844 primary/divergent transcript pairs for which either the divergent, primary, or both transcripts were regulated by E2 for at least one time point. As shown in Figure S2B (left), E2-dependent changes in divergent transcription were strongly correlated with E2-dependent changes in the corresponding primary transcripts (Pearson correlation: 0.744; $p < 2.2 \times 10^{-16}$). This result is consistent with a role for divergent transcription in facilitating E2-dependent transcription of the corresponding primary transcript.

Although not well characterized, antisense transcription has been shown to have roles in the degradation of corresponding sense transcripts (Katayama et al., 2005; Werner et al., 2009), as well as gene silencing at the chromatin level (Liu et al., 2010; Morris et al., 2008). Of 1197 transcripts annotated as

antisense to a protein-coding transcript, we identified 429 that are regulated by E2 (FDR q value < 0.001) (Figure S2C). As with the divergent transcripts, we determined whether production of a given E2-regulated antisense transcript correlates with the synthesis of the corresponding primary transcript. Based on 582 sense/antisense transcript pairs, we found a remarkably high correlation between genes and their antisense transcripts (Pearson correlation: 0.654; $p < 2.2 \times 10^{-16}$) (Figure S2B, right). This is particularly surprising given that, unlike divergent transcripts, antisense transcripts do not share a proximal promoter with the sense transcript, although promoter-promoter contact through genomic looping might allow for coordinated transcriptional responses. If antisense transcripts play a role in the degradation of the sense transcript, as has been suggested previously, then their E2-dependent production may provide a "built-in" means of attenuating the steady-state levels of a select set of estrogen-regulated transcripts.

We also identified 2761 transcripts that have no specific relation to previous genome annotations. Of these, 686 were regulated by E2 for at least one time point. Interestingly, the vast majority of these E2-regulated intergenic transcripts are downregulated by E2 treatment (Figure 2D). The function of these transcripts is unknown. Some may represent currently unannotated protein-coding transcripts or functional RNAs. Ascribing a function to these RNAs and determining their relative stability in the steady-state cellular RNA pool will require additional studies. Their downregulation by E2, however, suggests a link to the estrogen signaling program. Perhaps they act to antagonize E2-dependent transcriptional responses and must be shut down to achieve a full estrogen response. Alternatively, their antagonism by E2 may be a passive effect of RNA polymerases being diverted to bona fide transcriptional targets of the estrogen-signaling pathway, as suggested previously (Carroll et al., 2006).

Rapid, Extensive, and Transient Regulation of Protein-Coding Transcripts by Estrogen

Numerous studies have examined the steady-state regulation of protein-coding transcripts by E2 using expression microarrays (Cheung and Kraus, 2010; Kininis and Kraus, 2008). Given the sensitivity of our approach for detecting immediate transcriptional changes in response to short E2 treatments, we extracted and examined the protein-coding transcripts in our GRO-seq data for comparison. We focused on annotations in the RefSeq database because this set is among the most comprehensive collection of transcripts and has extensive and well-documented overlap with expression microarrays. As noted above, we used read counts in a 12 kb window at the 5′ end of each annotation and determined regulation by E2 using the edgeR package, filtering for a false discovery rate of 0.1%.

Using this approach, we detected a total of 3098 protein-coding transcripts whose levels changed relative to the control (untreated) condition at one or more of the points in the E2 treatment time course. In total, these transcripts represent ~15% of all genes annotated in RefSeq (~33% of 9337 expressed genes) that are responsive to E2. This is a considerably larger number of genes than was detected previously at 1 or 3 hr of E2 treatment using expression microarrays (Cheung and Kraus, 2010; Kininis and Kraus, 2008; Figure S3A). Surprisingly, we found ~1000 genes total to be up- or downregulated after only 10 min of E2 treatment. We used hierarchical clustering to define four classes of genes sharing similar patterns of regulation, including a class of rapidly downregulated genes and three classes of genes with maximal transcription at the three E2 treatment time points (10, 40, or 160 min) (Figures 3A and 3B). The downregulated class was the largest, comprising ~50% of the E2-regulated protein-coding transcripts. The majority of genes in this class were rapidly downregulated (by 10 min, on average) and tended (with a few exceptions) to stay downregulated throughout the time course. Upregulated genes with maximal transcription at 40 min were the second largest class, comprising ~34% of the E2-regulated protein-coding transcripts. Although the time course of induction or repression varied among the four classes, the magnitude of response did not differ between the classes (Figure 3C). Interestingly, the genes in the "10 minute max" and "40 minute max" classes returned, on average, to the basal levels of transcription by the end of the E2 treatment time course (Figure 3B), highlighting the rapid and transient nature of the transcriptional response for the majority of the upregulated genes.

Biologically relevant changes in transcription should be accompanied, in most cases, by similar changes in the steady-state level of the corresponding mRNA. We tested this expectation using both genomic and gene-specific comparisons. First, we compared fold changes in primary transcription that were detected using our GRO-seq data to fold changes at the level of steady-state mRNA (3 or 12 hr of E2 treatment) from published expression microarray data for MCF-7 cells. For the subset of genes that we observed to be regulated by GRO-seq, we found that the strongest correlations were between either the 40 or 160 min GRO-seq time points and the 3 hr microarray time point (Figures S3B and S3C). Note, however, that there are many more genes detected as E2 regulated by GRO-seq than by expression microarray analyses (Figure S3A). If we limited the analysis to only genes that change in the microarray analysis (FDR corrected q value < 0.05), we see an even higher correlation between GRO-seq and microarray data (Figure 3D; Spearman's correlation: 0.75). This analysis suggests that the early actions of E2 are almost all mediated at the level of transcription and that E2 does not affect RNA stability or degradation rate directly. These results provide a first indication that transcription, as determined by GRO-seq, is propagated to changes in the steady-state levels of the corresponding mRNAs.

Next, we randomly selected a set of 10 to 20 genes for each of the four classes (54 genes total) and measured the relative steady-state levels of mRNA from each gene over a 6 hr time course of E2 treatment using RT-qPCR. In general, the changes in transcription measured by GRO-seq were reflected in corresponding changes in the steady-state mRNA levels measured by RT-qPCR (Figure 3E and Figure S4). In almost all cases, we observed a delay of ~1–3 hr between the peak fold changes measured by GRO-seq and RT-qPCR. This delay reflects the time necessary for changes in Pol II (measured at the 5′ end in GRO-seq) to reach the 3′ end of the gene and for mRNA to accumulate (or degrade) by a detectable level. As with the comparisons to the microarray expression data, these results indicate that changes in transcription are efficiently translated into

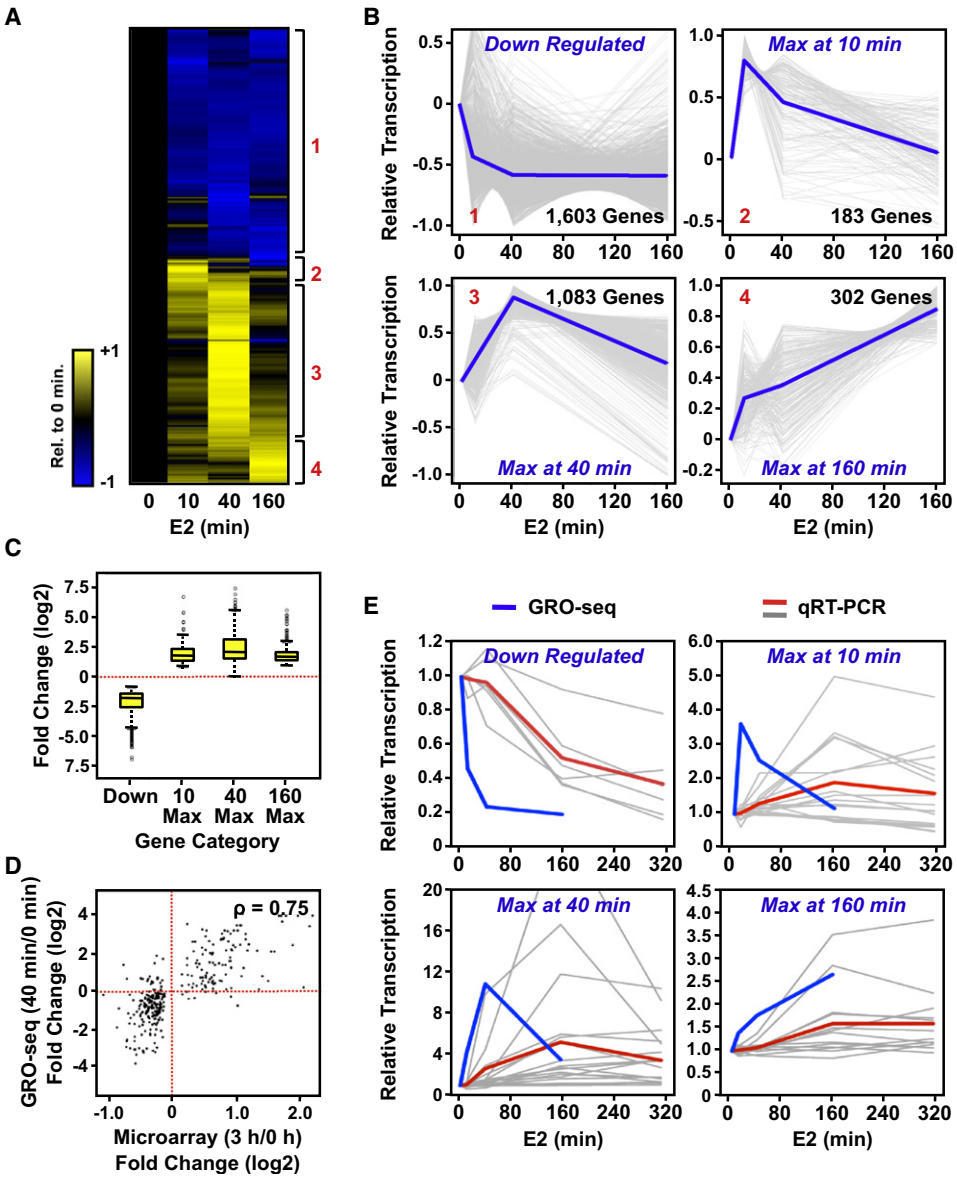

Figure 3. GRO-Seq Identifies Four Distinct Classes of E2-Regulated RefSeq Genes

(A) Heatmap of the time course of E2-dependent regulation of RefSeq genes. Red numbers indicate the four different classes of regulation.

(B) Centered-scaled traces showing the regulation of the four distinct classes of E2 regulation. Gray lines represent GRO-seq data for individual genes, and blue lines represent the mean of the individual traces.

(C) Box and whiskers plot showing the E2-dependent fold change for genes in each of the four classes.

(D) Correlation between fold changes measured by GRO-seq and expression microarrays for genes that show a change in the microarray analyses.

(E) Comparison of GRO-seq data to mRNA expression measured by RT-qPCR. Blue lines represent the mean of the GRO-seq data for the genes analyzed. Gray lines represent RT-qPCR data for individual genes, and red lines indicate the mean.

See also Figure S3 and Table S1.

changes in the steady-state levels of the corresponding mRNAs. The correspondence was strongest for the downregulated and the 40 min max GRO-seq classes (>80% of genes assayed showed corresponding changes) and weaker for the 10 min max and 160 min max classes (~50% of genes assayed showed corresponding changes). The discrepancies between transcription and steady-state mRNA levels may be due to inherent instability of certain nascent transcripts, which prevents them from generating mature transcripts. Alternatively, they may reflect active posttranscriptional regulation of specific transcripts (e.g., by miRNAs; see below). Interestingly, we identified a number of cases for each GRO-seq time point in which E2-dependent changes in transcription were accompanied by corresponding changes in the levels of the cognate protein,

including the 10 min max group (e.g., KRT19, MYC, and VDR; Figures S3D and S3E).

Gene ontology (GO) analyses of the four classes of genes revealed a similar pattern of enrichment in gene ontological categories for the downregulated and 40 min max classes (Tables S1A and S1C), which differ from or add to those derived previously from microarray expression analyses (Carroll et al., 2006; Frasor et al., 2003). Specifically, there was a significant enrichment in GO terms related to transcription, nucleic acid metabolism, cell surface receptor, and G protein-coupled signaling. The fact that the same GO terms but different genes are regulated in both the major up- and downregulated classes suggests a switch from one cellular signaling program (e.g., serum response) to another (i.e., estrogen signaling); each pathway may require the same functional categories of genes but use a distinct set of genes within each category. Interestingly, the 160 min max class was significantly enriched in GO terms related to ribosome biogenesis, translation, and protein synthesis (discussed and elaborated below) (Table S1D), whereas a very modest enrichment of GO terms was observed for the 10 min max class (Table S1B).

Together, our results show that the transcriptional response to estrogen signaling for protein-coding genes (and other classes of transcripts, as well; see below) is rapid, extensive, and transient. This represents a different view of the estrogen response than has been provided by microarray expression studies, which have suggested a continually increasing set of regulated genes in response to E2 treatment, many of which are likely to be secondary or tertiary effects (Figure S3A).

Pol II Dynamics in Response to E2

Because the transcriptional response for protein-coding genes to estrogen signaling was rapid and transient, we explored the dynamics of Pol II at the promoters of the four classes defined in the hierarchical clustering analysis. We performed metagene analyses across the promoter regions of each class from −4 kb to +4 kb for each treatment time point (Figure 4A). The peak of reads in the immediate vicinity of TSS indicates the presence, on average, of engaged Pol II before and after E2 treatment. The decrease (or increase) of reads in the downstream region indicates the downregulation (or upregulation) of transcription in response to E2. This presentation of the GRO-seq results highlights the following: on average, (1) loading of Pol II at the TSSs of upregulated genes increases in response to E2 treatment, (2) divergent transcription of the upregulated genes increases in response to E2 treatment, (3) downregulation affects primarily Pol II in the gene bodies, and (4) loading of Pol II at the TSSs and divergent transcription largely follow the Pol II response in the body of the gene.

The increase in Pol II loading at the TSS in response to E2 suggests that Pol II loads more rapidly than it escapes into the body of the gene for these classes of E2-regulated genes. This is especially evident between the 10 and 40 min time points for the 40 min max genes and between the 40 and 160 min time points for the 160 min max genes, for which we see increased Pol II loading at the earlier time point followed by an appreciable increase in Pol II in the body of the gene at the later time point. This "delayed" pattern of loading and escape is perhaps unexpected for the 160 min max genes, as the pausing of Pol II in the promoter proximal region is thought to allow rapid activation of transcription in response to cellular signaling (Lis, 1998). Alternatively, such a response fits well with a recent suggestion that pausing of Pol II in the promoter proximal region allows synchronous gene activation (Boettiger and Levine, 2009).

The dynamics of Pol II can also be clearly observed in examples from specific up- and downregulated genes (Figures 4B and 4C and Figure S4). With E2 upregulated genes, the leading edge of a Pol II wave was observed traveling into the gene body upon E2 treatment (Figure 4B). In contrast, with E2 downregulated genes, the lagging edge of a Pol II wave was observed as the polymerases were cleared from the TSS (Figure 4C). The results from our GRO-seq analysis have provided an unprecedented view of the Pol II dynamics in response to a sustained signal.

Regulation of miRNA Gene Transcription by Estrogen: Parallels to the Regulation of Protein-Coding Genes

Our GRO-seq approach also provides considerable information regarding the transcriptional regulation of primary microRNA transcripts. MicroRNAs (miRNAs) are ~22 nt noncoding regulatory RNAs that mediate posttranscriptional regulation of gene expression by inhibiting the translation or promoting the degradation of target mRNAs. miRNA precursor transcripts (pri-miRNAs) are generated by Pol II, or in some cases Pol III, either as part of a "host" gene in which they are embedded or from an intergenic region using their own promoter (Krol et al., 2010). Using our GRO-seq data set, we explored the regulation of pri-miRNA gene transcription by E2. We unambiguously identified 322 expressed miRNA-containing transcripts in our data set based on miRBase ver. 14. Of these, 119 (~37%) were regulated by E2 during at least one time point (FDR q value < 0.001). Regulated pri-miRNAs included some previously published estrogen-regulated miRNAs, including mir-181a, mir-181b, and mir-21. Overall, the pattern of regulation depicted in the heatmap shown in Figure 5A mirrors that observed for the protein-coding transcripts (i.e., approximately half upregulated and half downregulated), which is consistent with a large fraction being processed from protein-coding transcripts. Examples of the transcriptional response of specific pri-miRNAs are shown in Figure 5B. The primary transcript of both examples is considerably larger than the processed miRNA. Therefore, as with the protein-coding genes, the leading (or lagging) edge of the polymerase wave can be seen during the transcriptional response of the upregulated (or downregulated) genes. Together, these results suggest that the transcription of pri-miRNA genes is regulated by E2 in a similar pattern and with similar kinetics as protein-coding genes.

Next, we determined whether estrogen stimulation involves a coordinated response between pri-miRNA transcripts and the protein-coding genes that they ultimately regulate. For this analysis, we reasoned that the subset pri-miRNAs undergoing long-lasting and relatively large regulatory changes are the most likely to be reflected as changes in processed, mature miRNA. Therefore, we focused on 47 of the 119 (~40%) regulated pri-miRNA transcripts that show more than 3-fold up- or downregulation. These 47 robustly E2-regulated pri-miRNAs potentially target ~2700 mRNAs according to the TargetScan database (Grimson et al., 2007; Lewis et al., 2005), or ~12.8% of RefSeq annotated mRNAs.

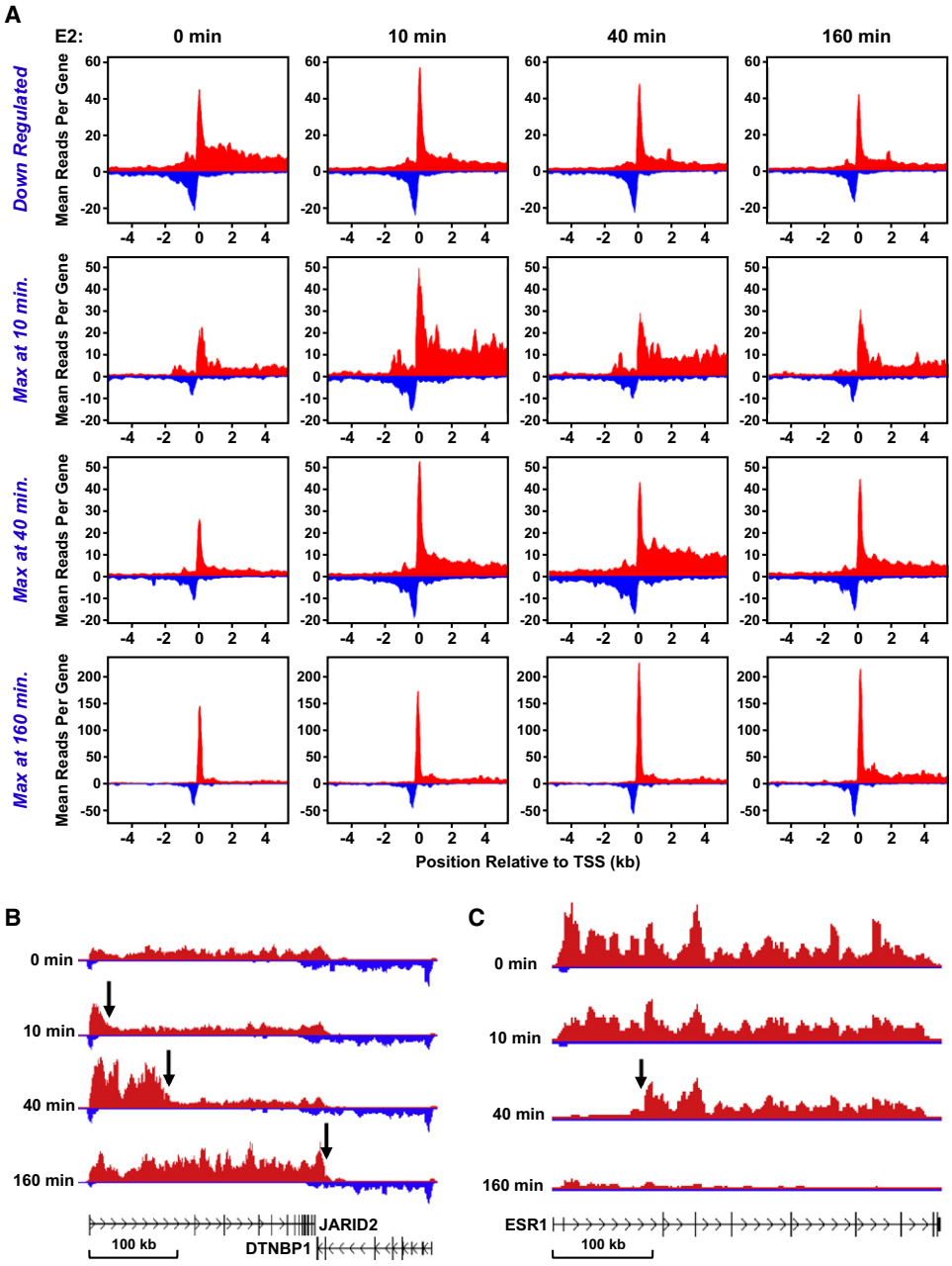

Figure 4. GRO-Seq Reveals the Dynamics of E2-Dependent Transcription

(A) Metagene representations showing the average profile of GRO-seq sequence reads near and at the TSSs of RefSeq genes in each of the four classes during the E2 treatment time course.

(B and C) Gene-specific views of the leading (B) and lagging (C) edges of a Pol II "wave" shown for the upregulated gene *JARID2* (B) and the downregulated gene *ESR1* (C), respectively, during the E2 treatment time course.

See also Figure S4.

Interestingly, as shown in Figure 5C, MCF-7 cells express a larger fraction of the ~2700 target mRNAs than expected, such that 16.6% of expressed genes are targets of these miRNAs ($p = 3.7 \times 10^{-14}$; Fisher's exact test). This enrichment is consistent with an integrated regulatory program between the miRNAs expressed in a cell and the corresponding mRNA targets, consistent with previous suggestions (Farh et al., 2005). Importantly, the subset of genes regulated by E2 is enriched even further over those genes that are expressed by the cell, such that 18.6% of E2-regulated mRNAs are targets of E2-regulated pri-miRNAs ($p = 0.03$) (Figure 5C). Moreover, this pattern of enrichment was also discovered when selecting a smaller set of miRNAs that are > 5-fold regulated by E2 ($p = 0.02$) or taking all miRNA transcripts

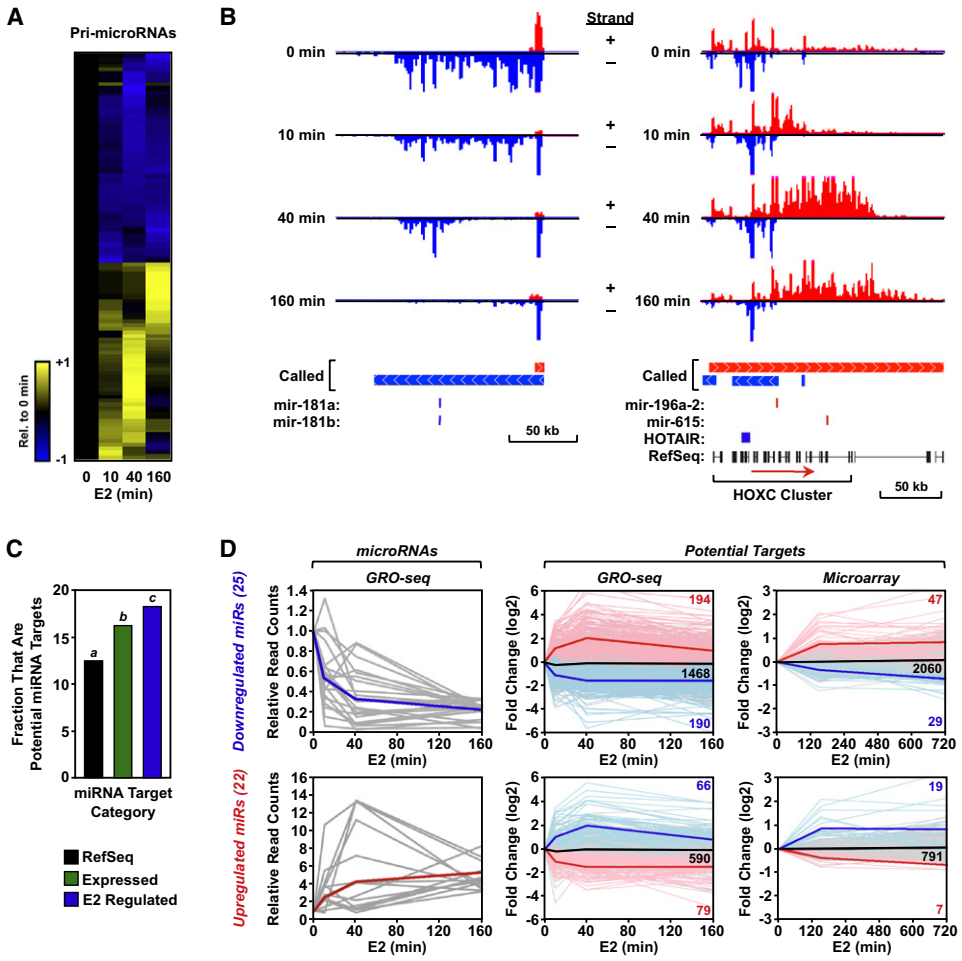

Figure 5. E2 Regulates the Transcription of Primary miRNA Genes

(A) Heatmap of the time course of E2-dependent regulation of primary miRNA transcripts.

(B) Gene-specific examples of downregulated (left) and upregulated (right) primary miRNA genes. Called transcripts and annotations are shown. In the right panel, the "+ strand" called transcript (red) is actually the number of smaller called transcripts that, at the resolution used to represent this region, appear as one transcript.

(C) Fraction of the specified subset of annotated genes that are predicted to be targets of an E2-regulated miRNA based on TargetScan. Bars with different superscripts are significantly different by Fisher's exact test ($p = 3.7 \times 10^{-14}$ for a/b; $p = 0.03$ for b/c; $p = 1.8 \times 10^{-13}$ for a/c).

(D) (Left) GRO-seq data for pri-miRNA transcripts that are upregulated (bottom) or downregulated (top) $\geq$ 3-fold by E2. Gray lines, data for individual genes; blue lines, average for all genes. (Middle and right) GRO-seq (middle) and expression microarray (right) data for all of the potential targets of miRNAs encoded by the pri-miRNA transcripts shown in the left panels. Faded red, black, and blue lines, data for individual upregulated, unregulated, and downregulated genes, respectively (the counts for each type are listed). Bold red, black, and blue lines, averages for all upregulated, unregulated, and downregulated genes, respectively.

See also Figure S5.

regardless of their fold change ($p = 0.003$), indicating that our results are robust to the threshold chosen for the analysis. We found no evidence that E2 specifically coordinates the transcriptional regulation of pri-miRNAs with the direction (i.e., up or down) of regulation of their potential target mRNAs, either by GRO-seq (Figure 5D, middle) or by expression microarrays (Figure 5D, right). In fact, we found evidence for both coordinated and compensatory regulation (Figure 5D; see Figure S5 for a detailed explanation). Together, these results suggest an integrated regulatory program for E2-regulated transcription of pri-miRNA transcripts and the mRNAs targeted by the mature miRNAs.

Dramatic Upregulation of the Protein Biosynthetic Machinery by Estrogen Signaling

Because our GO analyses showed enrichment in genes with a primary biological function in protein biosynthesis, we asked whether E2 signaling has a broader effect on the protein biosynthetic machinery. GRO-seq provides a measure of all three eukaryotic polymerases; we therefore extracted and analyzed the data for changes in the 45S rRNA (RNA Pol I) and tRNAs (Pol III) annotated in the rnaGene track in the UCSC genome browser. Our analysis revealed that the transcription of Pol I and Pol III transcripts shows a similar pattern of regulation by E2: (1) an initial burst at 10 min, (2) a slight decrease at 40 min,

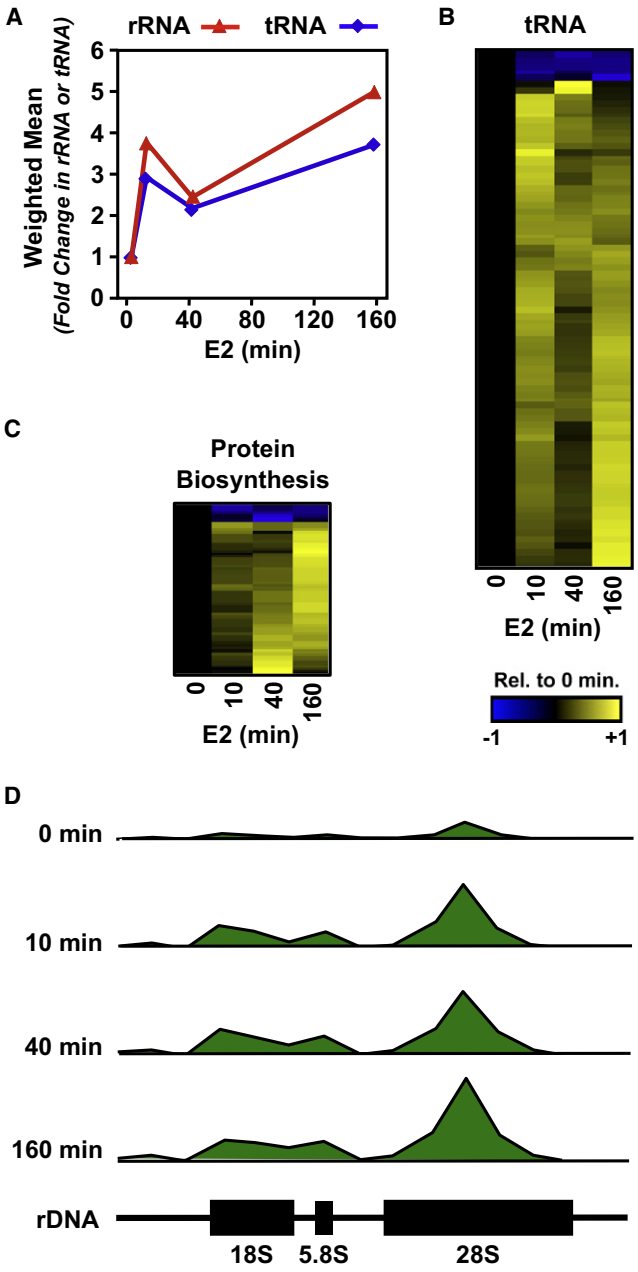

Figure 6. E2 Regulates Transcription by Pol I and Pol III

(A) E2-dependent fold change in the transcription of the 45S rDNA (Pol I) and tRNA (Pol III) genes.

(B and C) Heatmap of the time course of E2-dependent regulation of tRNA transcripts (B) or protein-coding transcripts encoding genes with a biological function or cellular compartment related to the synthesis, metabolism, or function of mature tRNAs or rRNAs (C).

(D) GRO-seq reads mapped to the human rDNA gene (GenBank U13369.1) shown in 1 kb bins relative to the genome location during the time course of E2 treatment.

See also Figure S6.

and (3) a maximal increase at 160 min (Figures 6A and 6D). These rapid effects are indicative of a primary, rather than secondary, transcriptional response to estrogen signaling.

For individual tRNA genes, changes were strongly biased toward upregulation, with the transcription of > 90% of the tRNA genes showing upregulation (Figure 6B). Furthermore, this regulation unambiguously affects 158 of the 486 functional annotated tRNA genes (32%) in at least one of the time points. If the cell is indeed regulating tRNA genes in order to facilitate an increase in translation, one may expect that all 20 amino acids will be upregulated. Indeed, we found that, of the 158 upregulated tRNA genes, at least one tRNA gene coding for each of the 20 amino acids is represented ($p = 0.0012$; Fisher's exact test) (Figure S6A). In addition to the 20 primary amino acids, we also found the tRNA coding for the amino acid variant selenocysteine, which is thought to play a role in antioxidant activity and hormone biosynthesis (Stadtman, 1996), to be regulated by E2. Because each three-letter combination of codons is represented multiple times in the 486 annotated tRNA genes, we also asked whether E2 regulates a larger fraction of the 64 possible codon combinations than expected by chance. Indeed, we find that 64% of the 64 codon combinations are unambiguously regulated by E2, which is more than expected based on our ability to call 32% of tRNA genes as regulated ($p = 0.0027$; Fisher's exact test). These results demonstrate that the observed changes in the protein biosynthetic machinery are applied in a robust and coordinated manner across amino acid and codon variations.

We also conducted a more focused analysis of protein-coding genes with functions or cellular localization suggesting a role in protein biosynthesis (e.g., ribosome biogenesis, tRNA aminoacetylation, etc.; see Figure S6B for all GO terms used). As we observed for tRNA genes, protein-coding genes represented in these groups are strongly biased toward upregulation (Figure 6C). As suggested by the GO analysis above, these genes are strongly enriched in the 160 min max class ($p = 6.7 \times 10^{-13}$; Fisher's exact test), suggesting that these are sustained effects that translate the widespread changes observed in the cellular transcriptome to the proteome.

Taken together, these results demonstrate a potent effect of estrogen signaling on the protein biosynthetic machinery, which fits well with the known mitogenic effects of E2 on MCF-7 cells. In addition, they highlight the fact that estrogen signaling has strong, immediate, and likely direct effects on transcription by all three RNA polymerases, not just Pol II. Upregulation of the protein biosynthetic machinery is likely a means by which the estrogen-signaling pathway prepares the cell for translation of the protein-coding transcripts that are newly synthesized in response to estrogen signaling.

Relationship of ERα-Binding Sites to Primary Estrogen Target Genes

Although most ERα-binding sites are located distal to the promoters of protein-coding genes, a small but highly significant enrichment of ERα-binding sites has been observed in the proximal promoters of upregulated genes (Carroll et al., 2005, 2006), consistent with a direct role of ERα in mediating their regulation. Because our GRO-seq data reflect the direct transcriptional output of the cell and because our shorter treatment times make it unlikely that we will detect secondary changes in transcription, we reasoned that we should observe that a larger fraction of the genes that are regulated by GRO-seq are near

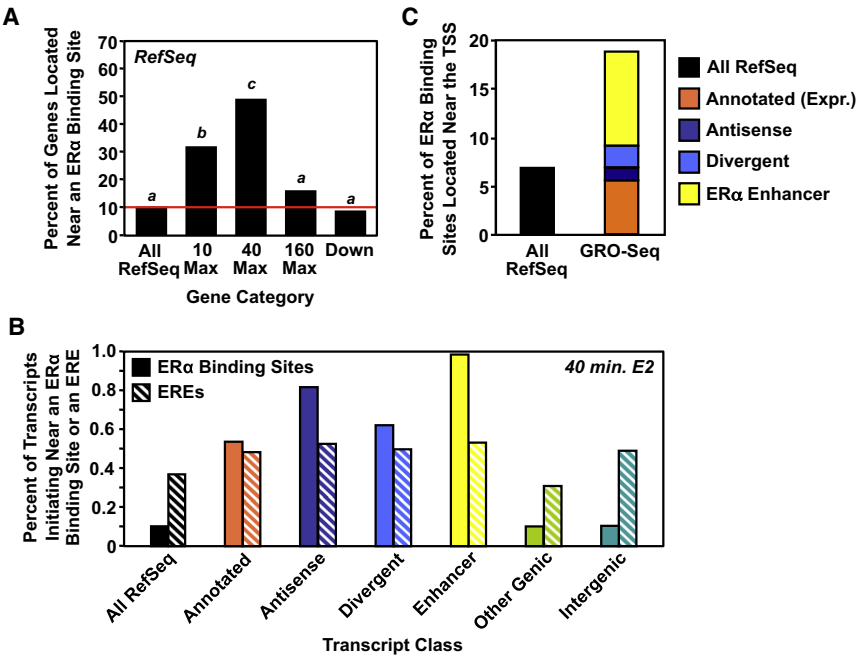

Figure 7. ERα-Binding Sites Are Enriched in the Promoters of Primary E2 Target Genes

(A) The fraction of the specified subset of RefSeq genes with an ERα-binding site found within 10 kb of the TSS. Bars with different superscripts are significantly different by Fisher's exact test ($p < 1.2 \times 10^{-12}$).

(B) The fraction of the specified transcript class defined at 40 min. E2 treatment that initiates near an ERα-binding site or an ERE.

(C) The fraction of ERα-binding sites found within 1 kb of either all well-annotated RefSeq genes or the specified subset of de novo transcript annotations determined by GRO-seq analysis.

ERα-binding sites. To test this hypothesis, we used existing ERα ChIP-seq data (Welboren et al., 2009) to determine the fraction of E2-regulated RefSeq genes with a proximal ERα-binding site (<10 kb to the transcription start site). Indeed, we found that 46% of genes upregulated by E2 at shorter time points (i.e., 10 and 40 min) contain an ERα-binding site within 10 kb of the transcription start site.

Interestingly, when we analyzed the four classes of RefSeq genes (i.e., 10, 40, 160 min max, and downregulated) separately, we found striking differences in binding site enrichment between these classes (Figure 7A). In particular, almost half of the genes in the 40 min max class are located within 10 kb of an ERα-binding site, a striking enrichment over the ~10% found for RefSeq genes in general ($p < 2.2 \times 10^{-16}$; Fisher's exact test). Genes in the 10 min max class are also substantially enriched for proximal ERα-binding sites (33%; $p = 1.2 \times 10^{-12}$). Upregulated genes that peak after 160 min have a lower level of enrichment that is not statistically significant (12%; $p = 0.24$), suggesting that a substantial fraction of this subset of genes reflects secondary effects. Conversely, downregulated genes were slightly less likely than average to be located within 10 kb of an ERα-binding site (8%; $p = 0.01$). This observation strongly suggests that E2 mediates up- and downregulation by different mechanisms and that immediate upregulated genes tend to be the direct genomic targets of ERα. Those E2-regulated genes that do not have a proximal ERα-binding site may be regulated by (1) other promoter-proximally bound transcription factors acting as endpoints of membrane-initiated E2-signaling pathways or (2) looping from distal ERα enhancers to the promoters.

Looking more broadly across the transcript classes, we found that the sets defined at 40 min of E2 treatment show a greater enrichment of both ERα-binding sites and EREs than the sets defined at the other time points. Interestingly, whereas the percentage of transcripts initiating near a bioinformatically defined estrogen response element (ERE) is not greatly enriched compared to all RefSeq transcripts and is relatively constant across the transcript classes (i.e., ~30%–50%), the percentage of transcripts initiating near an experimentally defined ERα-binding site varies considerably (Figure 7B). We observed the greatest enrichment of ERα-binding sites, compared to all RefSeq, near the initiation sites for annotated, antisense, divergent, and enhancer transcripts, suggesting similar modes of E2-dependent regulation as were observed for the protein-coding transcripts (Figure 7B).

We next determined the fraction of all ERα-binding sites that map within the proximal promoter (<1 kb) for each class of transcript defined in our GRO-seq analysis (i.e., looking from an ERα-binding site-centric view, as opposed to the transcript-centric view above). We found that ~18% of all ERα-binding sites fall near transcripts detected using our HMM in MCF-7 cells (Figure 7C). This includes ~5%–6% of ERα-binding sites near transcripts matching annotated genes that were specifically found to be expressed in MCF-7 cells using our approach (Figure 7C, orange bar), as well as an additional ~12% of ERα-binding sites found in the proximal promoters of genes producing transcripts that are not currently annotated in public databases (i.e., antisense, divergent, and enhancer transcripts). Though this finding still suggests that long-range enhancer-promoter interactions play a pivotal role in actions of ERα, as suggested previously (Fullwood et al., 2009; Pan et al., 2008; Theodorou and Carroll, 2010), it demonstrates a 3- to 4-fold increase in the fraction of ERα-binding sites that are located near TSSs.

Collectively, our results provide a new view of signal-dependent transcription events that suggest new questions and new ways of thinking about specific aspects of the transcriptional response.

EXPERIMENTAL PROCEDURES

Additional details about the experimental procedures can be found in the Supplemental Information.

Cell Culture

MCF-7 cells were maintained and propagated as described previously (Kininis et al., 2009).

Generation and Analysis of GRO-Seq Libraries

GRO-seq was performed as described previously (Core et al., 2008), with limited modifications. The data are available from the NCBI's Gene Expression Omnibus (accession number GSE27463), and the scripts are available upon request from the corresponding author.

Generation of GRO-Seq Libraries

Libraries were generated from two biological replicates of MCF-7 cells grown in estrogen-free medium and treated with 100 nM E2 as indicated. The libraries were sequenced using an Illumina Genome Analyzer.

Transcript Calling and Annotation

Short-reads were aligned to the human reference genome (hg18, NCBI36), including autosomes, X chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1) using SOAP2 (Li et al., 2009). A two-state hidden Markov model (HMM) (Durbin et al., 1998) was used to call transcripts, which were then divided into six distinct, nonoverlapping classes, which are intended to describe the function of each transcript. Annotations were made using the decision tree outlined in Figure S1E and based on a set of definitions (Figure S1F).

Determining Estrogen Regulation of Called Transcripts

E2-dependent changes in gene expression were detected using the edgeR package (v.1.4.1) (Robinson et al., 2010). For each GRO-seq time point, reads were counted in a window at the 5′ end of each transcript (+1 to +13 kb). Transcripts that change between the vehicle control and the 10, 40, or 160 min time points were collected for analysis if they met a false discovery rate (FDR) corrected q value threshold ($q < 0.001$), corresponding to an ~0.1% false discovery rate under the edgeR modeling assumptions.

Clustering, Time Course, and Classification of Temporal Profiles

We selected all genes with an FDR corrected q value of 0.001 at any point during the time course for inclusion in the temporal analysis. Computations were performed in the statistical package R, using the same pipeline that we described previously (Danko and Pertsov, 2009).

Additional Genomic Analyses

In addition to the analyses described above, we performed a set of more focused analyses, as described below. Unless otherwise noted, all computations were performed in R.

Gene Ontology Analyses

Gene ontology analyses were performed using GoStat (http://gostat.wehi.edu.au/; Beissbarth and Speed, 2004). All expressed genes were used as a background set to analyze GO terms for each class ($p < 0.05$).

Protein Biosynthesis-Associated Protein-Coding Genes

Protein-coding genes with a primary biological function or cellular compartment associated with the ribosome were identified using the Gene Ontology (GO) website (http://www.geneontology.org/) (Figure S6B).

Comparing E2-Induced Changes in Transcripts Called by GRO-Seq to Changes Observed by Expression Microarrays and Pol II ChIP-Seq

Raw CEL files from existing microarray data sets collected using the Affymetrix U133 platform were analyzed together using a previously described pipeline (Danko and Pertsov, 2009). Normalized microarray data were compared to read counts mapping to the +1 to +13 kb window of genes regulated by E2 during at least one point in the GRO-seq time course.

MicroRNA Analyses

We identified E2-regulated primary transcripts from our HMM transcript prediction algorithm that contain known miRNAs as described above. Each of these E2-regulated pri-miRNAs was associated with its regulatory targets using the TargetScan database (Lewis et al., 2005). Additional analyses were performed as described in the Supplemental Information.

Comparing the GRO-Seq Results to Known ERα-Binding Sites

For the 10,205 ERα-binding sites defined by Welboren et al. (2009), we calculated: (1) the fraction of genes in a particular class that are found within 10 kb of an ERα-binding site (Figure 7A) or (2) the fraction of ERα-binding sites mapping to within 1 kb, 5 kb, or 10 kb from the 5′ end of the nearest transcript identified de novo using the HMM described above or in a public database (Figure 7B).

Correlations between Primary Transcripts and Antisense/Divergent Transcripts

Transcripts corresponding to sense/antisense or sense/divergent pairs were collected, and the reads were counted and analyzed using R.

Metagene Analyses

We used metagene representations to illustrate the distribution of reads near a "typical" transcription start site. Mathematically, we defined a metagene as specified in the Supplemental Information.

RT-qPCR Gene Expression Analyses

Changes in the steady-state levels of the E2-regulated genes were analyzed by RT-qPCR, as previously described (Kininis et al., 2009). The fold expression changes were normalized to *GAPDH* as an internal standard.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at doi:10.1016/j.cell.2011.03.042.

ACKNOWLEDGMENTS

We thank Andre Martins for helpful insights and suggestions and Xin Luo and Shrikanth Gadad for critical comments on this manuscript. This work was supported by an NIH training award (T32HD052471) and a postdoctoral fellowship from the PhRMA Foundation to C.G.D., grants from the NIH (GM25232 and HG04845) to J.T.L., and a grant from the NIH/NIDDK (DK058110) to W.L.K.

Received: November 16, 2010
Revised: February 7, 2011
Accepted: March 24, 2011
Published online: May 5, 2011

REFERENCES

Acevedo, M.L., and Kraus, W.L. (2004). Transcriptional activation by nuclear receptors. Essays Biochem. *40*, 73–88.

Beissbarth, T., and Speed, T.P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics *20*, 1464–1465.

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471–473.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. *38*, 1289–1297.

Cheung, E., and Kraus, W.L. (2010). Genomic analyses of hormone signaling and gene regulation. Annu. Rev. Physiol. *72*, 191–218.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Danko, C.G., and Pertsov, A.M. (2009). Identification of gene co-regulatory modules and associated cis-elements involved in degenerative heart disease. BMC Med. Genomics *2*, 31.

Deroo, B.J., and Korach, K.S. (2006). Estrogen receptors and human disease. J. Clin. Invest. *116*, 561–570.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids (Cambridge, UK: Cambridge University Press).

Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817–1821.

Frasor, J., Danes, J.M., Komm, B., Chang, K.C., Lyttle, C.R., and Katzenellenbogen, B.S. (2003). Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology *144*, 4562–4574.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al; RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium. (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564–1566.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kininis, M., Isaacs, G.D., Core, L.J., Hah, N., and Kraus, W.L. (2009). Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol. Cell. Biol. *29*, 1123–1133.

Kininis, M., and Kraus, W.L. (2008). A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. Nucl. Recept. Signal. *6*, e005.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. *11*, 597–610.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics *25*, 1966–1967.

Lin, C.Y., Ström, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A., et al. (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. Genome Biol. *5*, R66.

Lis, J. (1998). Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb. Symp. Quant. Biol. *63*, 347–356.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science *327*, 94–97.

Morris, K.V., Santoso, S., Turner, A.M., Pastori, C., and Hawkins, P.G. (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet. *4*, e1000258.

Pan, Y.F., Wansa, K.D., Liu, M.H., Zhao, B., Hong, S.Z., Tan, P.Y., Lim, K.S., Bourque, G., Liu, E.T., and Cheung, E. (2008). Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. J. Biol. Chem. *283*, 32977–32988.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ruhl, D.D., and Kraus, W.L. (2009). Chapter 5 biochemical analyses of nuclear receptor-dependent transcription with chromatin templates. Prog. Mol. Biol. Transl. Sci. *87*, 137–192.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle *8*, 2557–2564.

Stadtman, T.C. (1996). Selenocysteine. Annu. Rev. Biochem. *65*, 83–100.

Theodorou, V., and Carroll, J.S. (2010). Estrogen receptor action in three dimensions - looping the loop. Breast Cancer Res. *12*, 303.

Warner, M., Nilsson, S., and Gustafsson, J.A. (1999). The estrogen receptor family. Curr. Opin. Obstet. Gynecol. *11*, 249–254.

Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J. *28*, 1418–1428.

Werner, A., Carlile, M., and Swan, D. (2009). What do natural antisense transcripts regulate? RNA Biol. *6*, 43–48.

Widelitz, R.B., Duffy, J.J., and Gerner, E.W. (1987). Accumulation of heat shock protein 70 RNA and its relationship to protein synthesis after heat shock in mammalian cells. Exp. Cell Res. *168*, 539–545.