



Published in final edited form as:

J Biol Chem. 2007 August 17; 282(33): 23981–23995.

Immediate-Early and Delayed Primary Response Genes are Distinct in Function and Genomic Architecture*

John W. Tullai^{1,§}, Michael E. Schaffer^{2,4,§}, Steven Mullenbrock^{1,§}, Gabriel Sholder¹, Simon Kasif^{2,3}, and Geoffrey M. Cooper^{1,2,†}

¹Department of Biology, Boston University, Boston, MA 02215 U.S.A.

²Bioinformatics Program, Boston University, Boston, MA 02215 U.S.A.

³Department of Biomedical Engineering, Boston University, Boston, MA 02215 U.S.A.

Abstract

The transcriptional program induced by growth factor stimulation is classically described in two stages: the rapid protein synthesis-independent induction of immediate-early genes, followed by the subsequent protein synthesis-dependent induction of secondary response genes. In the current study, we obtained a comprehensive view of this transcriptional program. As expected, we identified both rapid and delayed gene inductions. Surprisingly, however, a large fraction of genes induced with delayed kinetics did not require protein synthesis and therefore represented delayed primary rather than secondary response genes. Of 133 genes induced within 4 hours of growth factor stimulation, 49 (37%) were immediate-early genes, 58 (44%) were delayed primary response genes, and 26 (19%) were secondary response genes. Comparison of immediate-early and delayed primary response genes revealed functional and regulatory differences. Whereas many immediate-early genes encoded transcription factors, transcriptional regulators were not prevalent amongst the delayed primary response genes. The lag in induction of delayed primary response compared to immediate-early mRNAs was due to delays in both transcription initiation and subsequent stages of elongation and processing. Consistent with increased abundance of RNA polymerase II at their promoters, immediate-early genes were characterized by over-representation of transcription factor binding sites and high affinity TATA boxes. Immediate-early genes also had short primary transcripts with few exons, whereas delayed primary response genes more closely resembled other genes in the genome. These findings suggest that genomic features of immediate-early genes, in contrast to the delayed primary response genes, are selected for rapid induction, consistent with their regulatory functions.

The binding of growth factors to cell surface receptors leads to the activation of signaling pathways that ultimately control cell proliferation, differentiation and survival. The critical targets of these signaling cascades include transcription factors, and many of the changes in cell behavior resulting from growth factor stimulation are due to altered programs of gene expression. The canonical model of a highly ordered program of gene expression induced by growth factor stimulation is the coordinate regulation of primary and secondary response genes. The initial transcriptional response to growth factor stimulation is the induction of approximately 100 primary response genes (1,2). Induction of these genes does not require *de novo* protein synthesis and is therefore mediated by preexisting transcription factors. Most of the characterized primary response genes (termed immediate-early genes) are maximally

*This work was supported by NIH grant RO1 CA18689 to GMC and ITR-048715 and NHGRI grant R33 HG002850 to SK.

⁴Current address: Pfizer Inc., Research Technology Center, 620 Memorial Drive, Cambridge, MA 02139

[†]To whom correspondence should be addressed: Boston University, Department of Biology, 5 Cummington Street, Boston, MA 02215. Tel.: 617-353-8735; Fax: 617-353-8484; E-mail: gmcooper@bu.edu.

[§]These authors contributed equally to this study

induced within 30 minutes of growth factor stimulation, although a few examples of primary response genes that are induced with slower kinetics have been described (3-8).

Many of the well-characterized primary response genes encode transcription factors, which regulate downstream secondary response genes as part of a larger transcriptional program (1, 2). Secondary response genes are induced later than immediate-early genes and their induction is distinct from that of primary response genes in requiring *de novo* protein synthesis. Thus, the generally accepted model of growth factor-induced gene expression has two major components: the initial induction of primary response (immediate-early) genes, followed by a compulsory delay allowing translation of their mRNAs to produce the transcription factors that then induce the secondary response genes.

In the current study, we employed global expression profiling to analyze the temporal program of transcriptional alterations induced by growth factor stimulation of human cells. As expected, we identified distinct patterns of rapid and delayed gene inductions. Surprisingly, however, we observed that a large fraction of delayed inductions did not require protein synthesis, and therefore represented delayed induction of primary response genes rather than induction of secondary response genes. These results suggested that the transcriptional program induced by growth factor stimulation involved not only the induction of immediate-early and secondary response genes, but also the induction of a large group of delayed primary response genes that had previously been unrecognized.

The delayed primary response genes differed from immediate-early genes in both their functions and genomic architecture. Whereas many immediate-early genes encode transcription factors, transcriptional regulators were not prevalent amongst the delayed primary response genes. Rapid transcriptional induction of immediate-early genes was associated with several unique characteristics of these genes, including over-representation of shared transcription factor binding sites in upstream sequences of this gene set, high affinity TATA boxes in their core promoters, and short primary transcripts with few exons. In all of these features, delayed primary response genes more closely resembled other genes in the genome. These findings distinguish immediate-early from delayed primary response genes in terms of both function and transcriptional regulation, and suggest that immediate-early genes may have been selected for rapid induction based on their functions as transcriptional regulators. In contrast, the slower induction of both delayed primary and secondary response genes is consistent with their activities as effectors rather than mediators of growth factor signaling.

EXPERIMENTAL PROCEDURES

Cell Culture and RNA Extraction

T98G human glioblastoma cells were grown in Minimal Essential Medium (Invitrogen) supplemented with fetal calf serum (10%). Cells were rendered quiescent by incubation in serum-free medium for 72 h, and either left unstimulated, or stimulated for the indicated times with human platelet-derived growth factor (PDGF)¹-BB (50 ng/ml) (Sigma), epidermal growth factor (EGF) (30 ng/mL) (Calbiochem) or 20% fetal calf serum. When called for, cycloheximide (10 µg/ml, a concentration that inhibits protein synthesis >90% in T98G cells) was added 30 min prior to PDGF addition. Total RNA for real-time reverse transcription polymerase chain reactions (RT-PCR) for microarray validations and heteronuclear RNA

¹The abbreviations used are: PDGF, platelet-derived growth factor; EGF, epidermal growth factor; RT-PCR, reverse transcription polymerase chain reaction; hnRNA, heteronuclear RNA; IEG, immediate-early gene; D-PRG, delayed primary response gene; SRG, secondary response gene; CHX, cycloheximide; SRF, serum response factor; AP-1, activator protein-1; STAT, signal transducer and activator of transcription; CREB, cyclic AMP response element binding protein; BRE, TFIIB recognition element; Inr, initiator; MTE, motif 10 element; DPE, downstream core promoter element; MED-1, multiple start site element downstream; CAGE, cap analysis of gene expression; pol II, RNA polymerase II;

(hnRNA) analysis was extracted with TRIzol reagent (Invitrogen). Following ethanol precipitation, total RNA was applied to an RNeasy column (Qiagen) for further purification and treated with DNase according to the manufacturers' protocols. RNA for microarray experiments was extracted with TRIzol reagent (Invitrogen) followed by poly(A)⁺ RNA isolation with an Oligotex mRNA Midi Kit (Qiagen) according to each manufacturer's protocol.

Oligonucleotide Array Spotting

Microarrays were fabricated by resuspending 21,329 70-mer oligonucleotides from Operon's Human Genome Array-Ready Oligo Set Version 2.0 in 3X SSC to a final 30 μ M concentration, and spotted onto amino-silane coated GAPS II slides (Corning) with an OmniGrid Accent microarrayer (GeneMachines). After drying, the slides were post-processed according to the oligonucleotide manufacturer's protocol. Briefly, to promote spot uniformity, the microarrays were rehydrated with nuclease-free water and snap-dried on a 100°C hot plate. The slides were UV crosslinked with 65 mJ of energy and shaken for 20 min in a blocking solution of 1-methyl-2-pyrrolidinone, 171 mM succinic anhydride, and 43 mM sodium borate. Finally, the slides were washed successively in water and 95% ethanol, then centrifuged at 800 rpm for 5 min to dry.

Microarray Sample Preparation, Hybridization, and Image Analysis

Starting with 100 ng of poly(A)⁺ RNA, one round of RNA amplification was performed with the MessageAmp aRNA Amplification Kit (Ambion) using a 4:1 amino-allyl UTP:UTP ratio for aRNA incorporation. For each sample, 8 μ g of aRNA was coupled to N-hydroxysuccinimidyl esters of cyanine-3 or cyanine-5 (Amersham). Following clean-up, treated and untreated aRNA samples with opposing cyanine labels were combined, concentrated, and treated with a fragmentation reagent (Ambion) according to the manufacturer's protocol. For each slide, 4 μ g of both treated and untreated cyanine-labeled aRNA samples were combined with a hybridization buffer (2.3X SSC, 18 mM HEPES, 0.2 mg/ml BSA, 0.6 mg/ml poly(A), 0.2% SDS), heat denatured for 3 minutes at 95°C, and applied to microarrays under a LifterSlip coverslip (Erie Scientific). The slides were placed in a hybridization chamber (Dietech) and incubated in a 63°C water bath for 16 hr. Following hybridization, the slides were successively washed in 0.6X SSC with 0.025% SDS, 0.05X SSC, and water, then dried by centrifugation at 1000 rpm for 3 min. The microarrays were scanned with an Axon 4000B scanner and adaptive spot segmentation performed with GenePix Pro software (version 5.0) (Axon Instruments). For each treated sample, three independent replicate microarray experiments were performed.

Microarray Data Analysis

Triplicate dye-swap, background-subtracted median intensity values were used as input to the LIMMA analysis package (9) in Bioconductor (10), and average LOESS-corrected log₂ ratios were used to estimate differential gene expression. For the PDGF-treated samples, genes with positive log₂ ratios greater than or equal to 1 (2-fold) relative to untreated samples and FDR-corrected (11) moderated *t*-test *p*-values less than 0.01 were considered differentially expressed. Additional microarray dye-swap experiments were performed to identify genes with PDGF-stimulated inductions that were independent of new protein synthesis. RNA was extracted from cells treated with cycloheximide for 30 min followed by 2 or 4 hr PDGF treatments. For each of three replicates, two microarray experiments were performed with different reference samples. The first compared cycloheximide and PDGF-treated samples to untreated samples, while the other compared cycloheximide and PDGF-treated samples to PDGF-treated samples.

Real-time RT-PCR for Microarray Validations and Heteronuclear RNA Measurements

Reverse transcription of 0.5 µg of total RNA was performed in 50 µl using SYBR green RT-PCR reagents and random hexamer primers (Applied Biosystems) as recommended by the manufacturer. Following a 95°C incubation for 10 minutes, forty cycles of PCR (95°C/15s; 60°C/1m), were then performed on an ABI Prism 7900HT Sequence Detection System with 0.5 µl of the RT reaction, 100 nM PCR primers (Supplementary Table 1) and SYBR Green PCR Master Mix in 5 µl reactions. Threshold cycles (C_T) for three replicate reactions were determined using Sequence Detection System software (version 2.2.2) and relative transcript abundance calculated following normalization with a GAPDH PCR amplicon. Amplification of only a single species was verified by a dissociation curve for each reaction.

Chromatin Immunoprecipitation

Chromatin immunoprecipitations were performed as previously described (12), with modifications. Chromatin was immunoprecipitated overnight at 4°C using 6.25 µg/ml anti pol II antibody (N-20) (Santa Cruz Biotechnology, sc-899). Protein A agarose beads were washed successively in low salt wash, high salt wash, LiCl wash, and twice in 1×TE. Immunoprecipitated chromatin was quantified with real-time PCR using primers designed in proximity to the transcription start site as annotated in Entrez Gene (see Supplementary Table 1 for primer sequences).

Gene Feature and Genomic Sequence Data

Unless otherwise noted, all gene and transcript annotations, including genomic positions of transcription initiation sites and exon/intron boundaries for 23,969 human RefSeq transcripts, were obtained from the Entrez Gene database, corresponding to human genome build version 36.1 (April 3, 2006, ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/ASN_BINARY/Mammalia/Homo_sapiens.ags.gz) (13). Model transcripts (RefSeq accession numbers with 'XM_' prefix) and transcripts mapped to alternate human contig assemblies (RefSeq accession numbers with 'AC_' prefix) were not included in these analyses. For the core promoter analysis and splice site characterization, genomic sequence data was extracted from assembled RefSeq chromosome sequences for human genome build version 36.1 (ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens).

Gene Ontology Analysis

Gene Ontology (GO) terms were obtained from the Entrez Gene database (June 28, 2006, <ftp://ftp.ncbi.nlm.nih.gov/gene/gene2go>) for all human genes, and transitive closure of each term relationship was extracted from the daily GO build (June 28, 2006, http://archive.godatabase.org/latest-termdb/go_daily-termdb-tables.tar.gz) (14). Functional enrichment of co-expressed gene sets was determined with a one-tailed Fisher's exact test (15) by comparing the frequency of each term and all ancestors terms against the expected frequency from all annotated genes on the microarray. Only genes with at least one GO annotation were included in the analysis.

Transcription Factor Binding Site Analysis

Over-representation of transcription factor binding sites in the upstream regions of immediate-early and delayed primary response genes was analyzed as previously described (12,16), using the program Tractor (Schaffer, et al., manuscript in preparation) with 588 vertebrate transcription factor binding site matrices from TRANSFAC Professional (version 11.1) (17) and "minSUM" Match thresholds (18). For each matrix, the predicted site frequencies per gene for both the immediate-early and delayed primary response gene sets were compared to the site frequencies per gene observed in the upstream regions of 350 background genes using a permutation test. These background genes were randomly selected from genes expressed, but

not induced by PDGF, on the microarrays. Two independent analyses were performed. The first used only human sequences for predictions and the second considered only sites predicted within the same position of a human-dog-mouse multiple sequence alignment of each upstream region. Sequences and MULTIZ alignments were selected from the University of California Santa Cruz's Genome Browser (human, dog, and mouse versions hg18, mm8, canFam2, respectively) (19). For both analyses, the results were filtered to only include matrices that predicted, on average, less than 1 site per kb in background sequences and that detected sites upstream of at least 10% of the gene set being tested. *P*-values were adjusted with a false-discovery rate (FDR) correction (11). Only those matrices meeting the criteria of less than 1 hit per kb of upstream sequence in the background set and at least one hit in 10% of the test genes were considered in the correction.

Core Promoter Analysis

The promoters for immediate-early and delayed primary response genes were scanned using the Match algorithm with no score thresholds (18) and position-specific scoring matrices (PSSMs) representing six core promoter elements (Supplementary Fig. 1). The regions -48 to -21, -55 to -5, -13 to +15, +7 to +38, +17 to +43, and +89 to +177 relative to the transcription initiation site were scanned for the TFIIB recognition element (BRE) (20), TATA box, initiator (Inr), motif 10 element (MTE) (21), downstream core promoter element (DPE) (22), and multiple start site element downstream (MED-1) (23), respectively. For each core promoter element, the highest scoring position within each window on the forward strand was recorded for each transcript. Because some genes encode multiple transcripts, the maximum scores among all transcripts were determined for each human gene in the Entrez Gene database. Assessment of the biological significance of TATA prediction scores was performed as previously described (12). To identify at least 95% of sequences in three classes of TATA binding sites previously defined (24), 'TATAAA', 'TAAATA', and 'TATATA', a threshold of 0.7 was selected. This threshold was then used to identify the genome-wide frequency of TATA boxes predicted between -55 and -5 upstream of 23,969 human RefSeq transcripts.

Cap Analysis of Gene Expression Tag Classification

Human data for cap analysis of gene expression (CAGE) tag clusters and their attributes (25) were obtained from http://gerg01.gsc.riken.jp/cage_analysis/export/hg17prmtr/ and the classification of tag clusters was obtained from <ftp://fantom.gsc.riken.jp/FANTOM3/TC/>. The data were combined into a single relational database and for each RefSeq identifier, a representative transcription start site with maximal tag frequency and the associated class identifier were extracted. Carninci et al. (25) limited the cluster annotation to those with at least 100 tags, therefore, only 5,755 unique RefSeq transcripts were associated with a particular class. Classification information was available for 18 of 46 immediate-early genes, 21 of 50 delayed primary response genes, and 6 of 17 secondary response genes. To test for enrichment of each class in the primary response gene sets, one-sided Fisher's exact test *p*-values were calculated.

RESULTS

Kinetics of Primary Response Gene Induction

Microarray analysis was used to measure changes in gene expression following growth factor stimulation of quiescent human T98G cells, which can be reversibly arrested in the G₀ state by growth factor deprivation (26,27). Cells were rendered quiescent by serum deprivation, and then stimulated to re-enter the cell cycle by treatment with PDGF for 0.5, 2, and 4 hours. To distinguish primary from secondary response genes, transcript levels were also determined following 2 and 4 hour PDGF treatment in the presence of the protein synthesis inhibitor, cycloheximide.

The data for all genes that were induced greater than 2-fold ($p < 0.01$) are presented as a heat diagram in Fig. 1 (microarray data are presented in Supplementary Table 2). Of a total of 133 induced genes, 49 were induced >2-fold by 0.5 hours, characteristic of immediate-early genes. This group of genes included several well-known immediate-early genes, such as *FOS*, *FOSB*, *JUN*, *NR4A1*, *NR4A2*, and *MCL1*. In addition, a number of these genes were super-induced in the presence of cycloheximide, as previously observed for immediate-early genes. In contrast, a total of 84 genes were induced >2-fold only after 2-4 hours of PDGF treatment. The initial inductions of 26 of these genes were inhibited at least 50% by cycloheximide, as expected for secondary response genes that require *de novo* protein synthesis for transcription. These genes included well-characterized secondary response genes, such as *MMP3* (1) and *MMP13* (28). Surprisingly, induction of the remaining 58 genes was not blocked by cycloheximide, even though significant induction of these mRNAs required 2-4 hours of PDGF treatment. It is noteworthy that the number of primary response genes exhibiting these delayed kinetics of induction (delayed primary response genes) exceeded both the number of immediate-early genes and secondary response genes induced in these experiments.

The induction kinetics of several representative genes were analyzed using real-time RT-PCR with a finer resolution time course, following 0.5, 1, 2, 3, 4, 5, and 6 hours of PDGF treatment (Fig. 2). Consistent with the current microarray results and previous studies (16,29), two well-characterized primary response genes, *FOS* and *MCL1*, exhibited rapid but transient inductions that peaked at 0.5 hours following PDGF treatment (Fig. 2A). *MMP3*, a known secondary response gene, was not significantly induced until 3 hours of PDGF treatment and was blocked by cycloheximide, confirming the array results (Fig. 2B). In contrast to *FOS* and *MCL1*, five delayed primary response genes, *VCL*, *PLOD2*, *DKK1*, *SOD2*, and *CCND1*, demonstrated slower inductions, reaching maximal mRNA levels between 1 and 4 hours following PDGF stimulation (Fig. 2C). Consistent with the microarray results, induction of these genes was not blocked by cycloheximide, confirming their classification as delayed primary response genes. Additionally, to validate the array results, transcript levels of a total of 19 genes following 0.5, 2, and 4 hours of PDGF treatment, in the presence and absence of cycloheximide, were independently tested using quantitative real-time RT-PCR, the results of which confirmed the microarray data (Supplementary Table 3).

To determine if representative genes were induced with similar kinetics in response to mitogens other than PDGF, quiescent T98G cells were alternatively stimulated with EGF or serum in the presence or absence of cycloheximide (Fig. 3). Consistent with the results obtained with PDGF, *FOS* and *MCL1* were induced as immediate-early genes, *MMP3* as a secondary response gene, and *VCL*, *PLOD2*, *DKK1*, *SOD2*, and *CCND1* as delayed primary response genes following both serum and EGF treatment.

Delayed Primary Response Genes are Functionally Distinct from Immediate-Early Genes

To gain insight into possible functional differences, the immediate-early, delayed primary response, and secondary response genes were compared using the Gene Ontology (GO) database. Functional enrichment of GO terms was assessed by analysis of the frequency of GO terms in each set of genes compared to the expected frequency in all annotated genes on the array. The Molecular Function and Cellular Component GO terms that were significantly enriched ($p < 0.01$) and identified at least 10% of the genes in each group are summarized in Table 1. The immediate-early gene set was highly enriched in Molecular Function terms related to transcriptional regulation, with “DNA binding” and “transcription factor activity” among the most frequently represented categories (Table 1A). These functions were not significantly enriched in either the delayed primary response or secondary response genes (Table 1A). Similarly, the Cellular Component term “nucleus” was highly enriched in the immediate-early genes, but not in the delayed primary response or secondary response genes (Table 1B). These

findings are consistent with the recognized role of immediate-early genes as encoding transcription factors that then regulate secondary response genes. However, they also suggest distinct functions for the delayed primary response genes, as well as for secondary response genes, compared to the immediate-early genes.

Analysis of Promoters and Upstream Regions

The differing kinetics of induction of immediate-early and delayed primary response genes could result from a variety of factors, alone or in combination, including differences in transcription initiation, elongation, pre-mRNA processing, or mRNA stability. We therefore used a combination of computational and experimental approaches to compare several properties of these groups of genes. Initially, we explored the possibility of differences in the upstream regions of the immediate-early and delayed primary response genes that might be the cause of their distinct kinetics of induction. Co-regulated genes often share similar transcription factor binding sites, and groups of genes demonstrating different kinetics of induction might be expected to be under differential transcriptional control. Upstream regions of immediate-early and delayed primary response genes were therefore analyzed separately for over-represented transcription factor binding sites compared to a background set of genes that were expressed in T98G cells but not induced by PDGF stimulation (12,16). Sequences corresponding to 1, 3, and 5 kb upstream of each human gene, as well as the corresponding orthologous murine and canine sequences, were analyzed with the Match program using 588 vertebrate matrices from TRANSFAC Professional (v11.1) and a scoring threshold to minimize the sum of false negative and false positive (minSUM) hits. Analysis of human sequences alone identified matrices representing 4 transcription factors, serum response factor (SRF), nuclear factor kappa B (NF- κ B, represented by the V\$CREL_01 and V\$NFKAPPAB65_01 matrices), PAX-3 and early growth response (KROX) transcription factors, as significantly over-represented in upstream sequences of the immediate-early genes (Table 2A; the entire list can be found in Supplementary Table 4). In contrast, upstream regions of the set of delayed primary response genes lacked over-represented binding site matrices for these or other transcription factors.

The analysis was extended with phylogenetic footprinting to identify over-represented binding sites that were conserved in orthologous genomic regions of dog and mouse. The statistical analysis was performed with the same background sequence set, but only sites predicted in all three organisms at the same position of a multiple sequence alignment were scored. The top ten conserved transcription factor binding site matrices with the most significant *p*-values are shown in Table 2B (the entire output can be found in Supplementary Table 4). As expected (and as was also observed in the human-only analysis), conserved binding sites for known regulators, including SRF, NF- κ B, cyclic AMP response element-binding protein (CREB) and activator protein-1 (AP-1), were significantly overrepresented in the upstream regions of immediate-early genes. Notably, neither these nor other transcription factor binding site matrices were significantly over-represented in the set of upstream regions of the delayed response genes. Similar results were obtained when scanning 1 kb, 3 kb or 5 kb upstream regions.

The core promoter sequences were also examined for possible differences in binding sites for general transcription factors. The core promoter includes the TATA box, the TFIIB recognition element (BRE) (20), the initiator (Inr), the motif 10 element (MTE) (21), the downstream core promoter element (DPE), frequently found in TATA-less promoters (22), and the multiple start site element downstream (MED-1) (23) (Fig. 4A). Core promoter regions of immediate-early and delayed primary response genes were compared by analysis of the promoter sequences of each gene set near the expected positions for six core promoter elements (Fig. 4A). The highest scoring subsequences within these windows were determined using Match with frequency

matrices representing each promoter element (Supplementary Fig. 1). For each element, the distributions of scores for the immediate-early and delayed primary response genes were compared to one another and to a genome-wide score distribution with the Wilcoxon rank sum test (Fig. 4B).

The results indicate a significant difference in the TATA scores for the immediate-early genes ($p=5.1\times 10^{-7}$) relative to scores for 18,191 human genes in the Entrez Gene database, while significant differences in scores for the other core promoter elements were not observed. Furthermore, immediate-early gene TATA scores were significantly higher than those for the delayed primary response genes ($p=2.0\times 10^{-3}$) (Fig. 4B). In contrast, TATA scores of the delayed primary response genes did not differ significantly from all genes in the Entrez Gene database (TATA scores for all individual genes can be found in Supplementary Table 5).

The distributions of TATA scores for the entire Entrez Gene database, immediate-early genes, and delayed primary response genes are plotted in Fig. 4C. A threshold score of 0.7, which identifies more than 95% of sequences bound by human TATA binding protein (hTBP) *in vitro* (24), was used to define a functional TATA box (12). When applied genome-wide, this threshold identified 22% of genes with at least one transcript containing a TATA box, a figure similar to other estimates of TATA box prevalence (30). Using this threshold, 27 of 46 (59%) immediate-early and 17 of 50 (34%) delayed primary response genes contained a TATA box. The TATA box prevalence in immediate-early genes differed significantly ($p=1.0\times 10^{-7}$ by one-sided Fisher's Exact Test) from the Entrez Gene database.

In a survey of the human and mouse transcriptomes, Carninci et al. (25) experimentally identified several classes of transcription start site signatures using cap analysis of gene expression (CAGE). Transcription start sites of the single peak (SP) class were enriched in TATA boxes (25), so we also compared the immediate-early and delayed primary response genes according to the classes of transcription start sites identified by CAGE. Transcription start sites were divided into single peak (SP), broad (BR), bimodal/multimodal (MU), or broad with dominant peak (PB) classes. A significant bias for the SP class was found in the immediate-early gene set ($p=4.4\times 10^{-7}$), whereas the delayed primary response genes showed only a moderate enrichment for the same class ($p=7.5\times 10^{-3}$) (Fig. 4D, and listed individually in Supplementary Table 5). None of the other CAGE classes were significantly enriched in either primary response gene set. The SP class represents transcripts with a single, well-defined transcription start site. While both primary response gene classes show enrichment of the SP class, nearly all of the annotated immediate-early genes (15 of 18) were designated as SP, whereas only about half (11 of 21) of the delayed primary response genes had such an annotation. These results indicate that the immediate-early genes may have a greater tendency to initiate transcription from a well-defined initiation site than delayed primary response genes or other genes in the database. This is also consistent with the observed enrichment of TATA boxes in immediate-early gene promoters.

RNA Polymerase II Occupancy at Promoters of Immediate-Early and Delayed Primary Response Genes

Identification of distinct transcription factor binding site enrichment and TATA box abundance upstream of the immediate-early genes suggests that the lag in delayed primary response gene expression may result from slower transcription initiation rates, which could be a consequence of RNA polymerase II (pol II) abundance and/or recruitment at target gene promoters. To investigate this further, pol II binding to promoters of immediate-early and delayed primary response genes was investigated by chromatin immunoprecipitation (ChIP) analysis.

Quiescent T98G cells were treated with PDGF for 0-4 hrs and subjected to pol II ChIP, using an antibody against the N-terminus of pol II so that recognition of pol II was not affected by

modifications of its carboxy-terminal domain (CTD). Pol II occupancy was examined at the transcription start sites of 11 immediate-early (Fig. 5A) and 19 delayed primary response genes (Fig. 5B). All of these genes had pol II occupancy above that observed at the non-transcribed β -globin gene (Fig. 5), as well as several other negative control genes (data not shown). For the majority of genes (73% of immediate-early genes and 68% of delayed primary response genes), pol II occupancy did not change upon PDGF stimulation, suggesting a post-polymerase recruitment mechanism may be responsible for gene induction. Preloaded pol II at transcription start sites is not unprecedented, as *FOS* and *MYC* are well-established examples of genes with a paused polymerase in their proximal promoter regions in unstimulated cells (31).

For the 3 immediate-early genes that exhibited an increase (>1.75 fold) in pol II promoter occupancy upon PDGF treatment, all 3 had maximum pol II occupancy at 0.5 hour of PDGF treatment, coincident with their mRNA inductions (Fig. 5A). For the delayed primary response genes that had increased pol II occupancy upon PDGF treatment, 3 of 6 (*DKK1*, *DDX21*, and *ESDN*) had peak pol II occupancy after 2 hours of PDGF treatment (Fig. 5B), consistent with the possibility that delayed recruitment of pol II may play a part in their delayed mRNA inductions. However, the other 3 of these 6 delayed primary response genes (*VCL*, *TGFB2*, and *EPHA2*) demonstrated peak pol II occupancy after only 0.5 hour of PDGF treatment, similar to what was observed for some immediate-early genes and much earlier than their mRNA inductions. This suggests that the delay in mRNA induction for these genes occurs after the recruitment of pol II.

Although pol II recruitment does not appear to be a major factor resulting in delayed mRNA inductions, there was a clear difference in pol II occupancy between the immediate-early and delayed primary response gene groups. For both untreated cells (Fig. 5C) and for the timepoint at which maximum pol II occupancy was observed (Fig. 5D), the immediate-early genes had significantly higher pol II occupancy than the delayed primary response genes ($p=0.026$ for untreated cells and $p=0.0017$ at the time of maximum pol II occupancy), possibly correlating with the differences in promoters and transcription start sites between these two groups of genes.

Analysis of hnRNA Transcription

Since $\sim 70\%$ of the genes tested by ChIP did not demonstrate a change in pol II occupancy upon PDGF treatment, it is possible that transcriptional changes are not responsible for the observed mRNA inductions. To test this, hnRNA levels were measured with real-time RT-PCR using 5'-biased intron-specific primers (32) for 23 delayed primary response genes following PDGF stimulation (Supplementary Table 6). For all of the genes tested, hnRNA levels increased ≥ 2 -fold and were similar to or greater than the corresponding mRNA inductions upon PDGF treatment. Thus, although pol II occupancy at these genes is not altered, their transcription is induced upon PDGF treatment.

The kinetics of hnRNA synthesis for 12 representative delayed primary response genes and two immediate-early genes (*FOS* and *MCL1*) are presented in Fig. 6A. The hnRNA levels for *FOS* and *MCL1* (Fig. 6A panel 1) peaked at 15 minutes of PDGF treatment, corresponding to maximum mRNA induction at 30 minutes. Kinetics of hnRNA induction varied between different delayed primary response genes, which were categorized into 3 groups with examples shown in Fig. 6A panels 2-4. For some of these genes, the accumulation of hnRNA and mRNA in the same experiment is compared in Fig. 6B.

Seven of the delayed primary response genes (30%) exhibited a clear delay in hnRNA synthesis, not reaching peak hnRNA levels until 1-3 hours of PDGF treatment (Fig. 6A, panel 4). For these genes, the delay in transcription of hnRNA was consistent with their delayed

mRNA expression (Fig. 2 and Fig. 6B). Two of these genes, *DKK1* and *ESDN*, also have delayed pol II recruitment with peak occupancy at 2 hours (see Fig. 5B).

In contrast, 5 delayed primary response genes (22%), including *VCL*, responded with transcriptional kinetics similar to immediate-early genes, with peak induction of hnRNA at 15 minutes of PDGF treatment (Fig. 6A, panel 2). These kinetics of *VCL* hnRNA synthesis are in agreement with the pol II ChIP results, which revealed maximum pol II occupancy at the *VCL* promoter following 30 minutes of PDGF treatment (Fig. 5B). The hnRNA levels of these genes increased much earlier than their mRNA levels (Fig. 2 and Fig. 6B), indicating that mechanisms following the initiation of transcription and the start of productive elongation are resulting in delayed mRNA induction.

A third group of 11 delayed primary response genes (48%) (Fig. 6A, panel 3) showed intermediate delays in transcription of their hnRNAs. These genes are less clearly defined in terms of kinetics of transcription, although it is likely that steps both preceding and following the start of productive elongation may play a role in their delayed mRNA inductions.

Primary Transcript Feature Analysis

The above results indicated that although delays in transcription initiation and/or the start of productive elongation contributed to the lag in expression of some delayed primary response genes, other differences in transcription or processing also contributed to the delay in mRNA formation. We therefore analyzed other features that might affect rates of transcription or mRNA processing, including primary transcript length and intron/exon structure.

Because processing of pre-mRNA is a potential rate limiting step in gene expression, variations in 5' donor and 3' acceptor splice sites could indicate a general difference in splicing efficiency between the classes of primary response genes. We therefore compared the 5' and 3' splice site nucleotide compositions for the immediate-early and delayed primary response genes. However, there was no significant difference between the splice site characteristics of these groups of genes (Supplementary Fig. 2). Next, the primary transcript length and exon frequency distributions for immediate-early and delayed primary response genes were compared to each other and to a distribution of all genes in the Entrez Gene database (Fig. 7). The analysis indicated a significant difference in both the primary transcript length ($p=4.2\times 10^{-8}$) and exon frequency ($p=1.4\times 10^{-4}$) distributions of immediate-early genes relative to the genome-wide distributions (Fig. 7) (see Supplementary Table 5 for individual genes). In contrast, no significant differences were noted when these features of delayed primary response genes were compared to the genome-wide distribution. Furthermore, the immediate-early primary transcripts were significantly shorter than those of the delayed primary response genes (on average, ~19 kb versus ~58 kb, respectively, $p=2.5\times 10^{-9}$) and contained significantly fewer exons (on average, 5.8 versus 10.4, respectively, $p=1.4\times 10^{-4}$). These results suggested that, in addition to other gene features, the observed lag in mature mRNA induction of some delayed primary response genes may be related to both primary transcript length and exon frequency.

DISCUSSION

In this study, we have undertaken a comprehensive global analysis of the time course of gene induction following growth factor stimulation of quiescent human cells. As expected, we identified both rapid and delayed gene inductions resulting from PDGF stimulation. Forty-nine genes were induced within 30 minutes of stimulation, as expected for immediate-early genes, whereas 84 genes required 2-4 hours of PDGF stimulation for maximum induction. Surprisingly, we found that the majority of the genes induced with delayed kinetics (58/84) were primary response genes, since their induction was not inhibited by cycloheximide. The transcriptional program induced by growth factor stimulation thus involved three distinct

classes of genes: immediate-early genes, delayed primary response genes, and secondary response genes, which accounted for approximately 37%, 44% and 19% of the genes induced within 4 hours of PDGF stimulation, respectively. Similar kinetics of induction of representative delayed primary response genes were observed in response to the alternative mitogens EGF and serum, suggesting that their induction kinetics are not PDGF-specific event. Examples of delayed primary response genes have been observed by others in primary human fibroblasts (3,6,29), rat arterial smooth muscle cells (4) and mouse 3T3 cells (5,7,8), but the large number of primary response genes we found to be induced with such delayed kinetics was unexpected, suggesting a more complex regulatory landscape in mammalian cells.

Transcriptional programs are often represented as gene networks, where products of expressed genes activate or repress secondary downstream gene targets. Many analyses assume temporal regulation according to the canonical immediate-early/secondary response gene paradigm to infer protein-gene interactions from correlations in gene expression data (33). By highlighting the unexpectedly high incidence of delayed primary response genes, our results have broad implications for analyses that infer regulatory interactions from temporal correlations in gene expression. Since many genes that are induced with a significant lag after growth factor stimulation are still primary response genes, it cannot be assumed that temporally delayed gene expression requires the prior induction of upstream transcriptional regulators.

Because delayed primary response genes represented a major component of the transcriptional response to growth factor stimulation, we used both computational and experimental tools to elucidate the properties of this group of genes. We first sought to determine whether the delayed primary response genes shared similar functions with the immediate-early genes. Therefore, the immediate-early and delayed primary response genes' functional classifications were compared using the Gene Ontology (GO) database. The immediate-early genes were enriched in Molecular Function terms related to transcriptional regulation. This corresponded well with their recognized role as transcriptional effectors in the induction of secondary response genes. In contrast, the delayed primary response genes were not enriched in functions related to transcriptional regulation and had no significant functional overlap with the immediate-early genes. These comparisons suggest that the products of immediate-early genes may have unique functions in regulating the transcriptional response to growth stimulation, while the delayed primary and secondary response genes may function as effectors of this transcriptional program. In this regard, it is noteworthy that cyclin D1 was initially described as a secondary response gene in macrophages, whose induction linked cell cycle proliferation to growth factor stimulation (34). However, cyclin D1 behaved as a delayed primary response gene in the present study, as well as in 3T3 cells (8) and human fibroblasts (6).

We also examined the basis for the distinct kinetics of induction of immediate-early and delayed primary response gene mRNAs. Analysis of hnRNA demonstrated that both immediate-early and delayed primary response genes were induced at the transcriptional level. The hnRNAs of immediate-early genes were rapidly induced, coincident with the rapid inductions of their mRNAs. The lag in induction of a number of delayed primary response mRNAs appeared to result from either a delay in transcription initiation or the start of productive elongation, as suggested by the delayed inductions of their hnRNAs. In contrast, hnRNAs of other delayed primary response genes were rapidly induced, suggesting that the lag in mRNA induction resulted from delays in subsequent stages of transcriptional elongation or processing. These differences between the kinetics of induction of immediate-early and delayed primary response gene mRNAs appear to be associated with a combination of factors, including the over-representation of upstream binding sites for shared transcription factors, core promoter elements, gene length, and exon frequency.

Computational comparisons revealed striking differences in the prevalence of predicted binding sites for shared transcription factors in the upstream regions of immediate-early and delayed primary response genes. Binding sites for several known regulators, including SRF, AP-1, CREB, KROX and NF- κ B, were over-represented in the upstream regions of immediate-early genes compared to other genes that were expressed in T98G cells but not induced by PDGF. In contrast, binding sites for either these or other transcription factors were not significantly over-represented upstream of the delayed primary response genes. The absence of predicted binding site enrichment upstream of the delayed primary response genes may indicate that, whereas immediate-early genes are activated by a shared set of transcription factors, the delayed primary response genes are controlled by a more diverse set of regulators, which would not be identified as over-represented in the gene set. Alternatively, it is possible that delayed primary response genes contain fewer clusters of transcription factor binding sites near their promoters than immediate-early genes, or that the transcription factor binding sites upstream of delayed primary response genes are lower affinity sites than those upstream of immediate-early genes, since lower affinity sites that are divergent from the binding site matrix might not be scored in the computational analysis. Both of these factors could reduce the affinity of transcription factor binding to the promoter regions of delayed primary response genes, correspondingly reducing their rates of transcriptional activation.

The core promoters of the immediate-early genes also differed from those of the delayed primary response genes. In particular, promoters of the immediate-early genes contained higher affinity TATA boxes than those of the delayed primary response genes. Similarly, the prevalence of TATA boxes in the promoters of immediate-early genes (59%) was significantly higher than in the promoters of delayed primary response genes (34%) or in all genes in the genome (22%). This may have important implications in transcription initiation, with higher affinity TATA boxes conferring greater transcriptional activity on the promoters of immediate-early genes. Reinforcing the notion that the immediate-early genes have stronger, more defined initiation is the demonstration that these genes also have a significant bias for the SP, or single peak, promoter class defined by CAGE analysis (25). Moreover, because some components of the transcription initiation complex, including TBP, remain bound to DNA following pol II promoter clearance, the stability of these factors may modulate the transcription reinitiation rate. Thus, high scoring TATA boxes present in immediate-early promoters may represent higher affinity TBP binding sites that confer rapid reinitiation (35). Indeed, previous work demonstrated instability of TBP-TATA interactions following the first round of transcription (36) and non-canonical TATA box sequences diminish binding of TFIIA (37), a general transcription factor that is thought to stabilize the TBP-TATA complex (38).

The differences in both upstream transcription factor binding sites and core promoters are also consistent with differences in the binding of RNA polymerase II to the promoter regions of immediate-early and delayed primary response genes. Chromatin immunoprecipitation indicated that pol II was bound to the promoters of both immediate-early and delayed primary response genes in unstimulated cells, and that pol II occupancy increased on the promoters of about one-third of the genes in both groups following growth factor stimulation. Thus, transcriptional induction of the majority of immediate-early and delayed primary response genes may result from the start of productive elongation by a paused polymerase, rather than by recruitment of pol II to the preinitiation complex. These findings are consistent with previous demonstrations of paused polymerases near the transcription start sites of immediate-early genes, including *FOS* and *MYC* (31), as well as with global analyses that have detected preinitiation complexes at the promoters of many non-transcribed genes in human cells (39). Importantly, however, the amount of pol II bound to the promoters of immediate-early genes was significantly greater than that bound to the promoters of delayed primary response genes. These differences in pol II occupancy highlight a key distinction between the immediate-early and delayed primary response gene promoters. Together with the differences in both upstream

transcription factor binding sites and TATA boxes, these findings point to transcription initiation, and perhaps reinitiation, as one of the primary mechanisms for rapid responses of immediate-early genes to growth factor stimulation relative to the delayed primary response genes.

Our analysis also revealed significant differences between the immediate-early and delayed primary response genes in both primary transcript lengths and exon frequencies. The immediate-early genes tend to be shorter and contain fewer exons than the delayed primary response genes, which are similar in length and exon frequency to other genes in the genome. These transcript features may contribute significantly to the lag in mRNA expression of delayed primary response genes, particularly for those genes that displayed a rapid induction of transcription, as detected by hnRNA. *VCL* provides an extreme example of the possible effect of primary transcript length and exon frequency on kinetics of mRNA expression. Analysis of hnRNA established that transcription of *VCL* was rapidly initiated, similar to immediate-early genes, such as *FOS* and *MCL-1*. Consistent with its rapid transcriptional induction, SRF has been reported to be a key inducer of *VCL* (40). However, the accumulation of *VCL* mRNA was delayed by 2-3 hours compared to the hnRNA. This lag in mature *VCL* mRNA production may be explained by the 122 kb primary transcript length, which is more than six times the average immediate-early gene primary transcript length, and the presence of 22 exons, which is almost four times the average number of immediate-early gene exons. At the other extreme, *DKK* (another delayed primary response gene) has a primary transcript of only 3.3 kb containing 4 exons, comparable to that of the shortest immediate-early genes. In contrast to *VCL*, transcriptional induction of *DKK* is delayed for 2-3 hours after growth factor stimulation, coincident with increased pol II occupancy at its promoter. *DKK* may therefore represent an example of a gene whose delayed induction results primarily from a lag in pol II recruitment and transcription initiation.

Multiple differences between immediate-early and delayed primary response genes thus appear to contribute to the distinct kinetics of induction of their mRNAs. The immediate-early genes are characterized by over-representation of binding sites for several transcription factors in their upstream regions, promoters with high affinity TATA boxes, and short primary transcripts containing relatively few exons. In all of these respects, the delayed primary response genes are similar to other genes in the genome. Additional features, such as chromatin structure, may also distinguish immediate-early from delayed primary response genes, as has been reported for genes displaying rapid versus delayed inductions in response to other stimuli (41-43).

To determine whether these characteristics of immediate-early genes were consistent in other cell types, we analyzed the features of immediate-early genes induced by the mitogenic stimuli EGF in HeLa cells and serum in MCF10A cells (normal human breast epithelial cells) in published data sets (44). As in T98G cells, the immediate-early genes induced in both HeLa and MCF10A cells showed an over-representation of transcription factor binding sites, including sites for SRF, AP1, CREB, KROX and NF- κ B, that were conserved in mouse and dog (Supplementary Table 7; for complete Transfac output, see Supplementary Table 8). Likewise, immediate-early genes in HeLa and MCF10A cells had significantly higher TATA scores, lower exon frequencies and shorter transcript lengths as compared to the genome as a whole (Supplementary Figures 3-4). Thus, the immediate-early genes induced in T98G, HeLa, and MCF10A cells by three different mitogens share common characteristics of genomic organization.

The multiple features associated with rapid induction of immediate-early genes may have been selected for based on the functions of immediate-early gene products as transcriptional regulators that mediate subsequent alterations in gene expression in response to growth factor stimulation. The rapid induction of immediate-early genes might be expected to play an

important role in achieving a robust cellular response to extracellular signals. In contrast, the lag in induction of both delayed primary and secondary response genes is consistent with the apparent functions of these genes as effectors rather than mediators of growth factor signaling. Thus, immediate-early genes are not only characterized by a lack of requirement for new protein synthesis prior to their transcriptional induction; they also possess distinct genomic features that may have been selected to confer rapid inducibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Ulla Hansen for helpful comments and discussion of the manuscript and to Scott Schaus, Erin Eastwood, Melissa Dominguez and Jolyon Terragni for microarray printing assistance.

REFERENCES

- Herschman HR. *Ann. Rev. Biochem* 1991;60:281–319. [PubMed: 1883198]
- Winkles JA. *Prog Nucleic Acid Res Mol Biol* 1998;58:41–78. [PubMed: 9308363]
- Burger C, Wick M, Brusselbach S, Muller R. *J Cell Sci* 1994;107:241–252. [PubMed: 7513713]
- Dixon BS, Sharma RV, Dennis MJ. *J Biol Chem* 1996;271:13324–13332. [PubMed: 8662783]
- Freter RR, Alberta JA, Hwang GY, Wrentmore AL, Stiles CD. *J Biol Chem* 1996;271:17417–17424. [PubMed: 8663287]
- Sewing A, Burger C, Brusselbach S, Schalk C, Lucibello FC, Muller R. *J Cell Sci* 1993;104:545–555. [PubMed: 8389378]
- Williams M, Lyu MS, Yang YL, Lin EP, Dunbrack R, Birren B, Cunningham J, Hunter K. *Genomics* 1999;55:327–334. [PubMed: 10049588]
- Winston JT, Pledger WJ. *Mol Biol Cell* 1993;4:1133–1144. [PubMed: 8305735]
- Smyth, GK. Limma: linear models for microarray data. In: Gentleman, R.; Carey, V.; Dudoit, S.; Irizarry, R.; Huber, W., editors. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer; New York: 2005.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. *Genome Biol* 2004;5:R80. [PubMed: 15461798]
- Benjamini Y, Hochberg Y. *J. R. Statist. Soc* 1995;57:289–300.
- Tullai JW, Chen J, Schaffer ME, Kamenetsky E, Kasif S, Cooper GM. *J Biol Chem*. 2007In press
- Maglott D, Ostell J, Pruitt KD, Tatusova T. *Nucleic Acids Res* 2005;33:D54–58. [PubMed: 15608257]
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. *Nat Genet* 2000;25:25–29. [PubMed: 10802651]
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. *Nat Genet* 1999;22:281–285. [PubMed: 10391217]
- Tullai JW, Schaffer ME, Mullenbrock S, Kasif S, Cooper GM. *J Biol Chem* 2004;279:20167–20177. [PubMed: 14769801]
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. *Nucleic Acids Res* 2006;34:D108–110. [PubMed: 16381825]
- Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. *Nucleic Acids Res* 2003;31:3576–3579. [PubMed: 12824369]
- Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR,

- Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. *Nucleic Acids Res* 2006;34:D590–598. [PubMed: 16381938]
20. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. *Genes Dev* 1998;12:34–44. [PubMed: 9420329]
21. Ohler U, Liao GC, Niemann H, Rubin GM. *Genome Biol* 2002;3:RESEARCH0087
22. Burke TW, Kadonaga JT. *Genes Dev* 1996;10:711–724. [PubMed: 8598298]
23. Ince TA, Scotto KW. *J Biol Chem* 1995;270:30249–30252. [PubMed: 8530439]
24. Hardenbol P, Wang JC, Van Dyke MW. *Nucleic Acids Res* 1997;25:3339–3344. [PubMed: 9241250]
25. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. *Nature Genetics* 2006;38:626–635. [PubMed: 16645617]
26. Stein GH. *J. Cell. Physiol* 1979;99:43–54. [PubMed: 222778]
27. Takahashi Y, Rayman JB, Dynlacht BD. *Genes. Dev* 2000;14:804–816. [PubMed: 10766737]
28. Selvamurugan N, Fung Z, Partridge NC. *FEBS Lett* 2002;532:31–35. [PubMed: 12459458]
29. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. *Science* 1999;283:83–87. [PubMed: 9872747]
30. Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr. *Genome Biol* 2005;6:R33. [PubMed: 15833120]
31. Saunders A, Core LJ, Lis JT. *Nat Rev Mol Cell Biol* 2006;7:557–567. [PubMed: 16936696]
32. Lipson KE, Baserga R. *Proc Natl Acad Sci U S A* 1989;86:9774–9777. [PubMed: 2481317]
33. Bar-Joseph Z. *Bioinformatics* 2004;20:2493–2503. [PubMed: 15130923]
34. Matsushime H, Roussel MF, Ashmun RA, Sherr CJ. *Cell* 1991;65:701–713. [PubMed: 1827757]
35. Yean D, Gralla J. *Mol Cell Biol* 1997;17:3809–3816. [PubMed: 9199314]
36. Fan X, Shi H, Adelman K, Lis JT. *Proc Natl Acad Sci U S A* 2004;101:6934–6939. [PubMed: 15103022]
37. Stewart JJ, Stargell LA. *J Biol Chem* 2001;276:30078–30084. [PubMed: 11402056]
38. Liu Q, Gabriel SE, Roinick KL, Ward RD, Arndt KM. *Mol Cell Biol* 1999;19:8673–8685. [PubMed: 10567590]
39. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. *Nature* 2005;436:876–880. [PubMed: 15988478]
40. Gineitis D, Treisman R. *J. Biol. Chem* 2001;276:24531–24539. [PubMed: 11342553]
41. Natoli G, Sacconi S, Bosisio D, Marazzi I. *Nat Immunol* 2005;6:439–445. [PubMed: 15843800]
42. Ramirez-Carrozzi VR, Nazarian AA, Li CC, Gore SL, Sridharan R, Imbalzano AN, Smale ST. *Genes and Development* 2006;20:282–296. [PubMed: 16452502]
43. Sacconi S, Pantano S, Natoli G. *J Exp Med* 2001;193:1351–1359. [PubMed: 11413190]
44. Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, Amariglio N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y. *Nat Genet* 2007;39:503–512. [PubMed: 17322878]
45. Jin VX, Singer GA, Agosto-Perez FJ, Liyanarachchi S, Davuluri RV. *BMC Bioinformatics* 2006;7:114. [PubMed: 16522199]

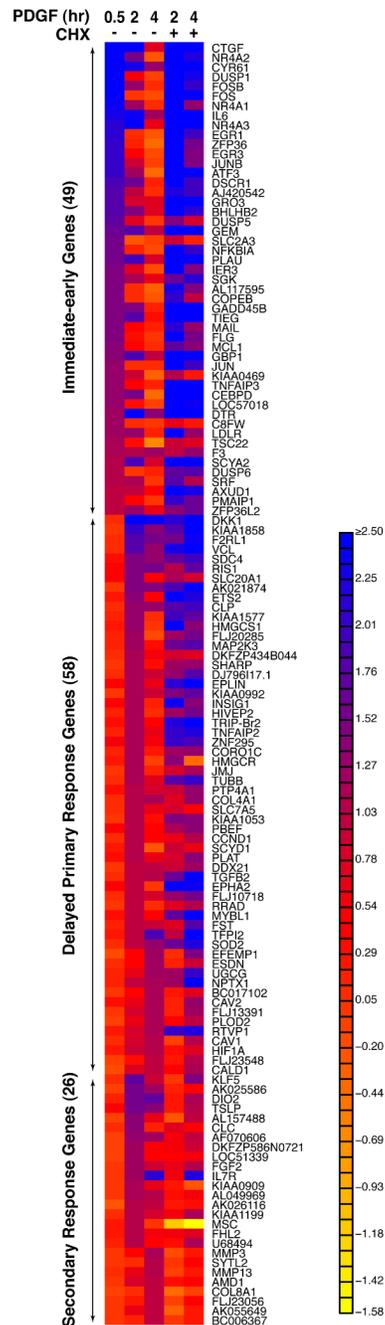


Fig 1. Genes induced by PDGF in quiescent T98G cells

T98G cells were rendered quiescent by serum starvation for 72 hours and then stimulated by treatment with PDGF for 0.5, 2 or 4 hours. Where indicated, cells were preincubated with cycloheximide (CHX) for 30 minutes before PDGF treatment. Results of triplicate microarray analyses are presented as a heat diagram illustrating average \log_2 ratios for genes that were induced greater than 2-fold ($p \leq 0.01$) by PDGF. Genes are arranged by the earliest time point at which the PDGF induction exceeded 2-fold ($p \leq 0.01$) and classifications are based on the inhibition of PDGF inductions by cycloheximide. Of the 133 genes induced by PDGF, 49 were rapidly induced by 30 min and were considered immediate-early genes. The remaining genes were separated into two groups of 58 delayed primary response genes, which are induced

independently of translation, and 26 secondary response genes, whose transcription requires new protein synthesis.

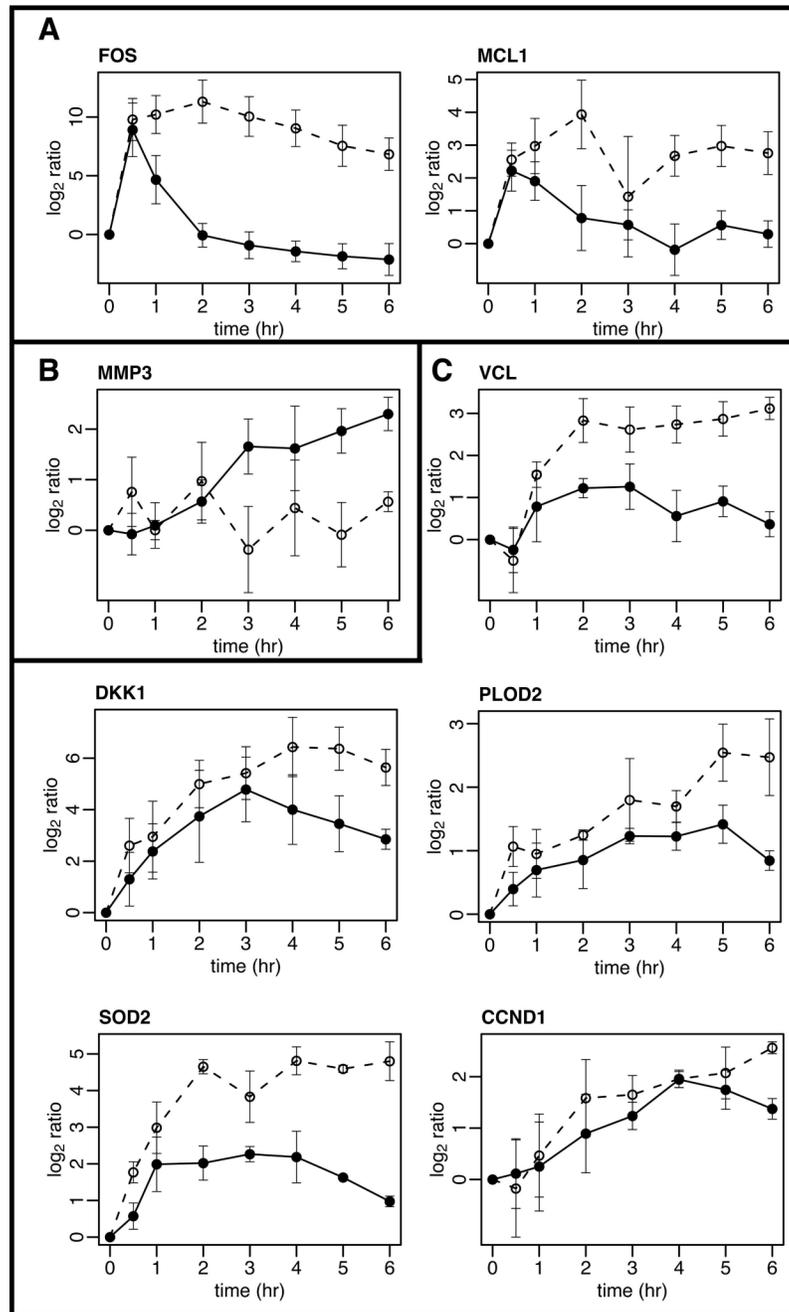


Fig 2. Real-time RT-PCR analysis of gene induction

Quiescent T98G cells were treated with PDGF and RNA was extracted following 0.5, 1, 2, 3, 4, 5, and 6 hr of treatment (solid lines). In addition, cells were treated with cycloheximide 30 min prior to PDGF stimulation (dashed lines). Induction of representative genes was quantitated by real-time RT-PCR relative to untreated cells. **A.** immediate-early genes, *FOS* and *MCL1*, **B.** secondary response gene, *MMP3*, **C.** delayed primary response genes, *VCL*, *PLOD2*, *DKK1*, *SOD2*, and *CCND1*. Data are mean \pm standard error (SEM) of three independent cultures.

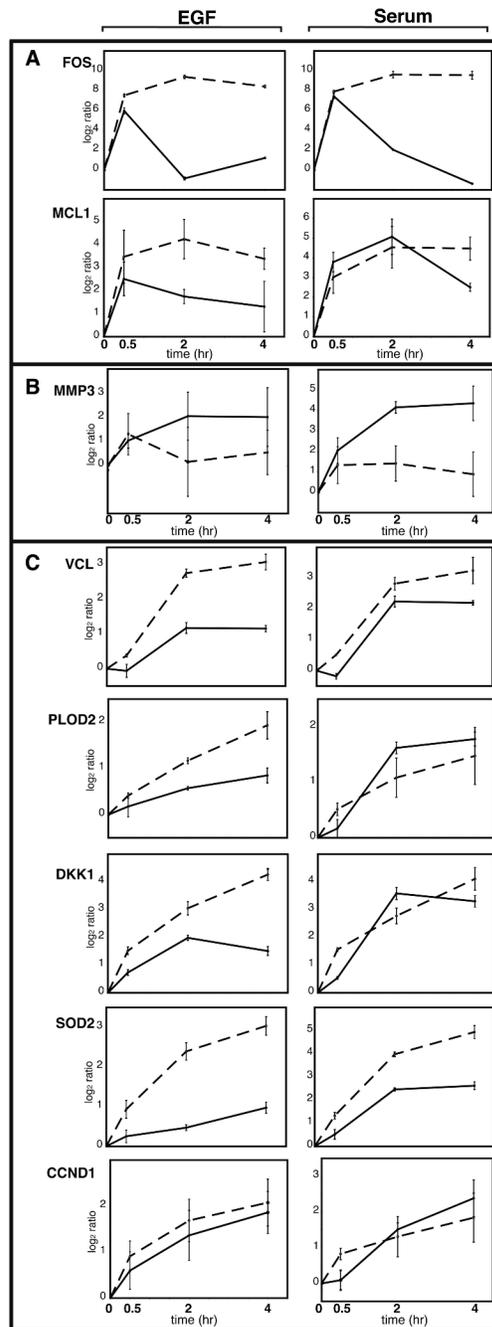


Fig 3. Kinetics of gene induction by EGF and serum

Quiescent T98G cells were stimulated with EGF or 20% serum and RNA was extracted following 0.5, 2, and 4 hr of treatment (solid lines). In addition, cells were treated with cycloheximide 30 min prior to EGF or serum stimulation (dashed lines). Induction of representative genes was quantitated by real-time RT-PCR relative to untreated cells. A. immediate-early genes, *FOS* and *MCL1*, B. secondary response gene, *MMP3*, C. delayed primary response genes, *VCL*, *PLOD2*, *DKK1*, *SOD2*, and *CCND1*. Data are mean \pm SEM of three independent cultures.

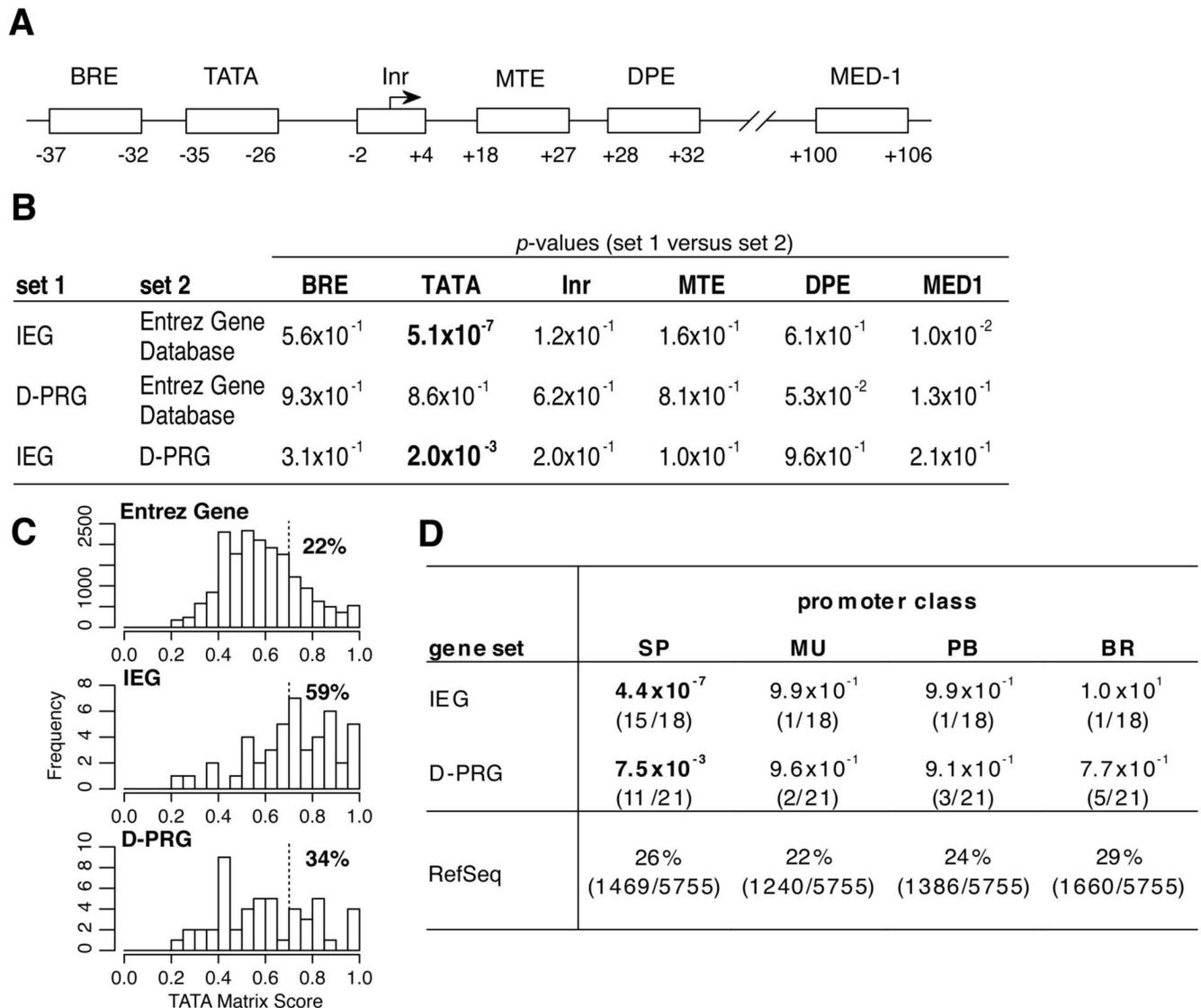


Fig 4. Analysis of core promoter elements and transcription start sites

A. Diagram of six core promoter elements with positions indicated relative to the transcription initiation site (+1). Figure adapted (45) to include MED-1 (23). B. Comparisons of core promoter elements. Matrix score distributions for each core promoter element in immediate-early gene (IEG) and delayed primary response gene (D-PRG) promoters were compared to each other and to all genes in the Entrez Gene database. *P*-values reflect score distributions between the two groups indicated as assessed by the Wilcoxon rank sum test (two-sided for IEGs or D-PRGs versus Entrez Gene database, one-sided for IEGs versus D-PRGs). Significant *p*-values are indicated in bold. C. Histograms of TATA score distributions. The fraction of genes in each group with matrix scores greater than 0.7, reflecting functional TATA boxes, is indicated in the upper-right of each panel. D. CAGE tag classifications. Summary of one-sided Fisher's exact test *p*-values reflecting enrichment of the four mutually exclusive promoter classes described by Carninci et al. (25) for IEGs and D-PRGs. The four promoter classes, defined by the shape of the CAGE tag distribution, include single peak (SP), bimodal/multimodal (MU), broad with dominant peak (PB), or broad (BR). The fractions of immediate-

early genes, delayed primary response genes, and of the 5755 RefSeq transcripts annotated with these classes are in parentheses. Values in bold indicate p -values ≤ 0.01 .

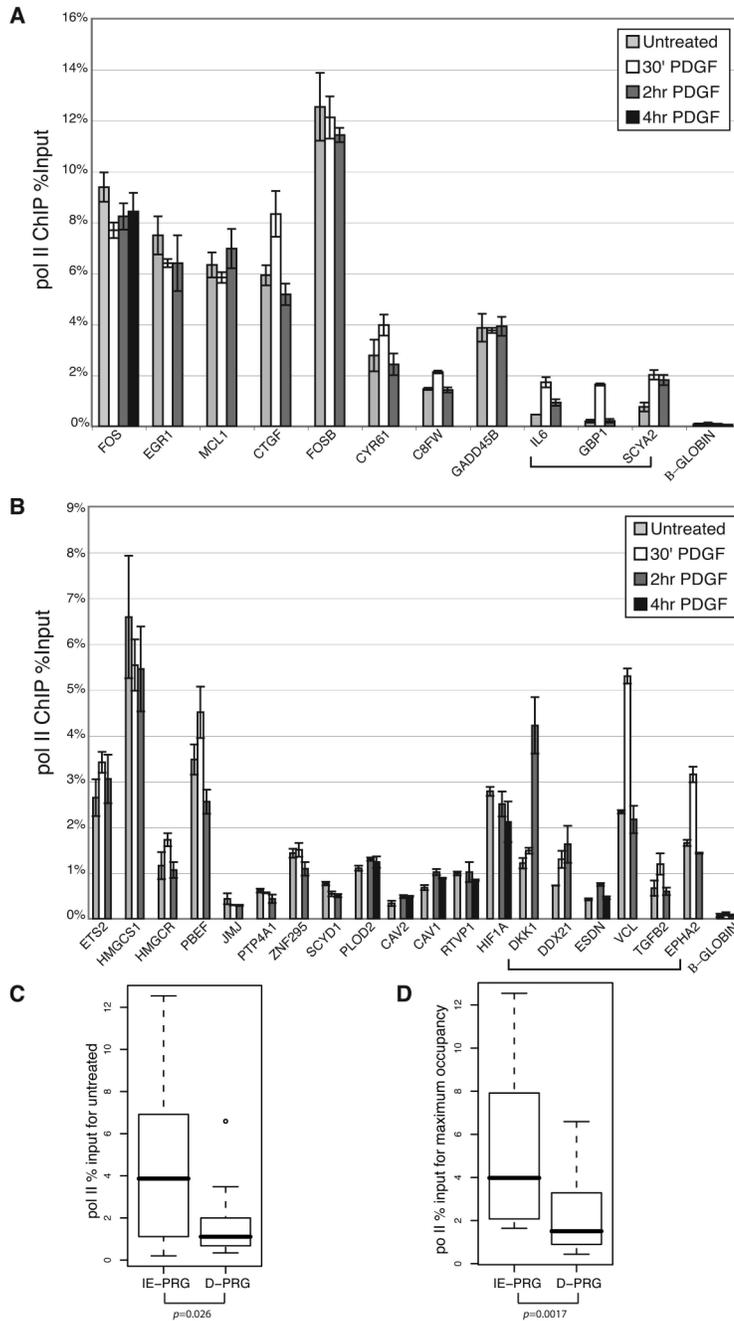


Fig 5. RNA polymerase II occupancy at the promoter regions of immediate-early and delayed primary response genes
 RNA polymerase II ChIP assays were conducted on quiescent T98G cells and on cells that were treated for 0.5, 2, and 4 hr with PDGF. ChIP primers were designed at the transcription start sites of the corresponding genes and immunoprecipitated material was plotted as a percentage of input \pm SEM. Results are averages of 2-4 determinations. Data for immediate-early genes are presented in *panel A* and for delayed primary response genes in *panel B*. Brackets indicate genes whose pol II occupancy was induced >1.75 fold upon PDGF treatment. *B-globin* was used as the negative control. Other negative controls including *ACTC*, *MYOD1*, and *MYOG* (all muscle-specific genes) yielded similar results (not shown).

Data for pol II occupancy at immediate-early compared to delayed response gene promoters is presented as box plots for untreated cells (*panel C*) and at the time of maximum pol II occupancy (*panel D*). *P*-values were derived using the Wilcoxon rank sum test.

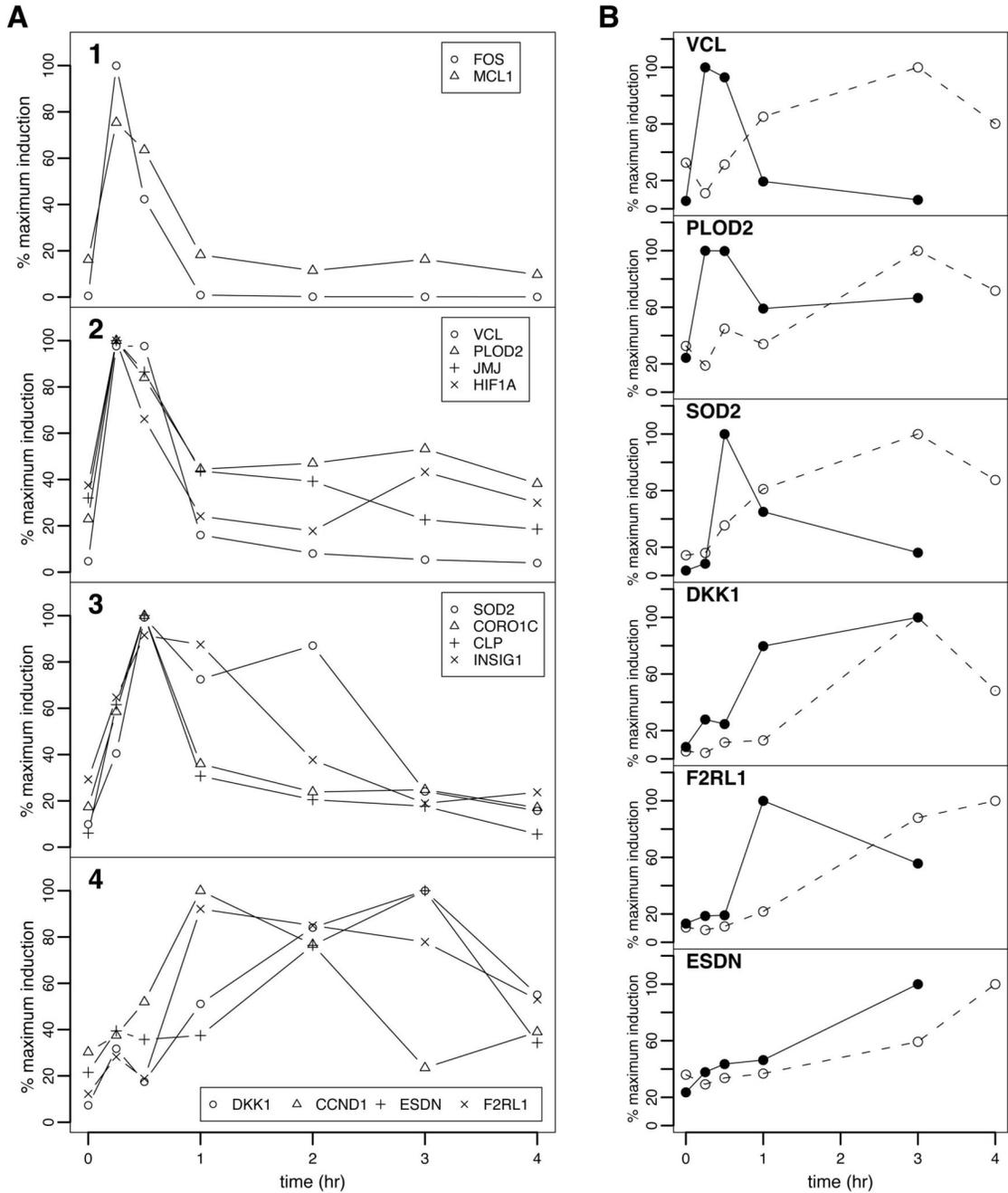


Fig 6. Kinetics of hnRNA synthesis

A. Total RNA was extracted following 0, 0.25, 0.5, 1, 2, 3, and 4 hr of PDGF treatment of quiescent T98G cells and analyzed by real-time RT-PCR using intron primers to detect hnRNA. Data are presented as percentage of maximum induction for representative immediate-early genes (panel 1) and delayed primary response genes with rapid, intermediate, and delayed inductions of hnRNA (panels 2-4, respectively), and are averages of 2-3 time courses (error bars are not indicated). B. Cells were treated and RNA was harvested as described in A. Shown are the results from a single RNA isolation in order to directly compare the kinetics of hnRNA (solid lines) and mRNA (dashed lines) synthesis.

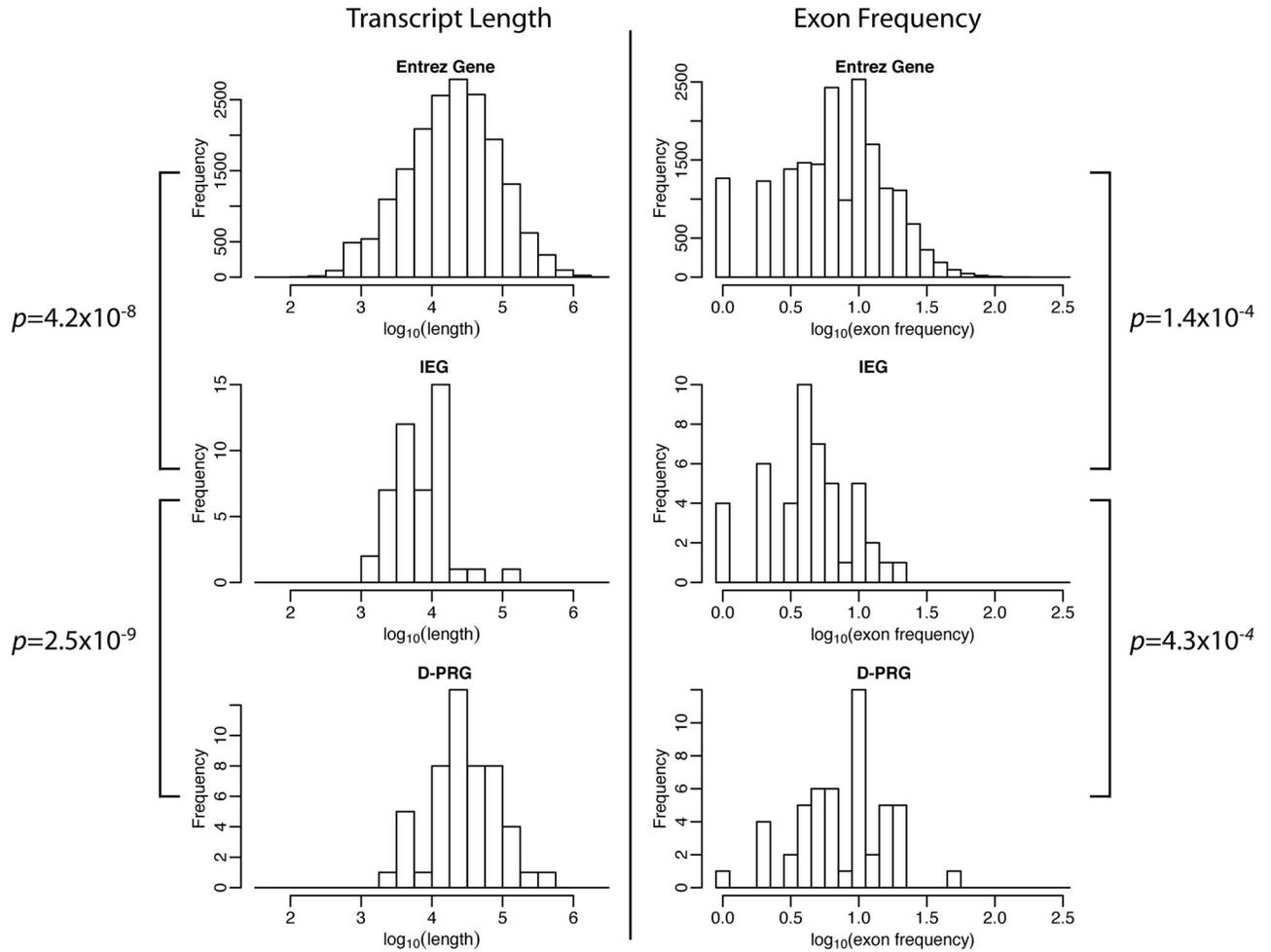


Fig 7. Transcript length and exon frequency distributions for immediate-early and delayed primary response genes

Histograms illustrate the distribution of the minimum transcript lengths and minimum exon frequency distributions across all Entrez Gene transcripts, immediate-early genes, and delayed primary response genes. P -values were calculated by the Wilcoxon rank sum test (two-sided for comparison of either D-PRG or IEG to Entrez Gene, and one-sided for comparison of IEG to D-PRG). The differences between the delayed primary response genes and Entrez Gene were not significant for either exon frequency or transcript length. Analysis of maximum transcript lengths and maximum exon frequency distributions yielded similar results.

Table 1

Summary of Gene Ontology (GO) term analysis

All GO terms in the Molecular Function (A) and Cellular Component (B) categories that were enriched in immediate-early genes (IEG), delayed primary response genes (D-PRG), or secondary response genes (SRG) ($p < 0.01$) and were represented in at least 10% of the genes are summarized (entries with significant p -values are indicated in bold). The fraction of genes with each GO term annotation is listed in the 'genes' column for each gene set.

	IEG		D-PRG		SRG	
	%genes	p -value	%genes	p -value	%genes	p -value
A. Molecular Function GO Summary						
DNA binding (GO:0003677)	42%	2.0×10^{-5}	14%	6.2×10^{-1}	7.7%	8.8×10^{-1}
nucleic acid binding (GO:0003676)	42%	2.4×10^{-3}	17%	8.4×10^{-1}	7.7%	9.6×10^{-1}
transcription regulator activity (GO:0030528)	37%	1.5×10^{-6}	7.1%	8.0×10^{-1}	15%	3.7×10^{-1}
transcription factor activity (GO:0003700)	35%	6.1×10^{-8}	4.8%	7.9×10^{-1}	0.0%	-
sequence-specific DNA binding (GO:0043565)	23%	1.1×10^{-6}	2.4%	7.6×10^{-1}	0.0%	-
protein dimerization activity (GO:0046983)	16%	7.3×10^{-6}	7.1%	3.5×10^{-2}	0.0%	-
cytoskeletal protein binding (GO:0008092)	0.0%	-	14%	1.4×10^{-3}	7.7%	3.3×10^{-1}
actin binding (GO:0003779)	0.0%	-	12%	1.7×10^{-3}	7.7%	2.4×10^{-1}
metalloendopeptidase activity (GO:0004222)	0.0%	-	0.0%	-	15%	4.8×10^{-3}
B. Cellular Component GO Summary						
nucleus (GO:0005634)	62%	1.2×10^{-4}	23%	9.3×10^{-1}	27%	7.5×10^{-1}
extracellular space (GO:0005615)	15%	3.8×10^{-3}	7.5%	2.1×10^{-1}	27%	2.2×10^{-3}
extracellular region (GO:0005576)	18%	5.9×10^{-2}	28%	6.5×10^{-4}	33%	8.5×10^{-3}
extracellular region part (GO:0044421)	15%	2.5×10^{-2}	15%	2.8×10^{-2}	33%	1.3×10^{-3}
extracellular matrix (sensu Metazoa) (GO:0005578)	2.6%	6.1×10^{-1}	7.5%	6.8×10^{-2}	20%	4.8×10^{-3}

Table 2

Over-represented transcription factor binding sites upstream of immediate-early and delayed primary response genes

One, three and five kb upstream regions were examined for over-represented transcription factor binding sites using TRANSFAC matrices. The complete TRANSFAC output prior to filtering can be found in Supplementary Table 4. For each gene set the table columns indicate the FDR corrected *p*-values. **A.** Transcription factor binding site predictions in human sequences with permutation *p*-values ≤ 0.05 (bold). **B.** The top ten ranking transcription factor binding site predictions with permutation *p*-values ≤ 0.01 (bold) conserved in human, dog and mouse. IEG, immediate-early genes; D-PRG, delayed primary response genes; NA, not applicable, referring to no occurrence or not meeting the criteria of $>10\%$ of the gene set hit and on average less than 1 hit per kb in the background set. The 1 kb conserved list contains 15 matrices, as the 8th through 15th-ranked positions shared the same *p*-value. The 3 kb conserved list contains 12 matrices, as the 10th through the 12th-ranked positions shared the same *p*-value.

A. Human Only			
5kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
3kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
V\$SRF_Q6	4.0×10^{-2}	NA	V\$NFKAPPAB65_01
V\$NFKAPPAB65_01	4.0×10^{-2}	>0.8	V\$KROX_Q6
V\$CREL_01	4.9×10^{-2}	>0.8	V\$SRF_Q4
B. Human-Dog-Mouse Conservation			
5kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
3kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
V\$IK3_01	4.4×10^{-6}	>0.8	V\$NFKAPPAB65_01
V\$AP1_Q4	3.0×10^{-5}	NA	V\$NFKB_C
V\$NFKAPPAB65_01	3.0×10^{-5}	>0.8	V\$SRF_C
V\$CREB_Q2	5.2×10^{-5}	>0.8	V\$CREL_01
V\$CREB_01	5.2×10^{-5}	>0.8	V\$NFKAPPAB_01
V\$SRF_C	8.6×10^{-5}	NA	V\$SRF_Q4
V\$AP1_Q4_01	1.3×10^{-4}	NA	V\$AP1_Q6_01
V\$ATF4_Q2	1.4×10^{-4}	>0.8	V\$SRF_01
V\$TCF11_01	1.4×10^{-4}	>0.8	V\$ATF4_Q2
V\$AP1_Q2_01	1.4×10^{-4}	>0.8	V\$CREB_01
V\$AP1F1_Q2	1.4×10^{-4}	>0.8	V\$IK3_01
V\$CREL_01	1.4×10^{-4}	NA	V\$SRF_Q6
V\$NRF2_Q4	1.4×10^{-4}	>0.8	
V\$CREB_Q2	1.4×10^{-4}	>0.8	
V\$CREB_Q3	1.4×10^{-4}	>0.8	
5kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
V\$NFKAPPAB65_01	9.6×10^{-9}	>0.8	V\$NFKAPPAB65_01
V\$NFKB_C	4.7×10^{-7}	NA	V\$NFKB_C
V\$SRF_C	9.7×10^{-7}	NA	V\$CREL_01
V\$CREL_01	1.2×10^{-6}	>0.8	V\$NFKAPPAB_01
V\$NFKAPPAB_01	1.5×10^{-6}	NA	V\$AP1_Q4_01
V\$SRF_Q4	8.2×10^{-6}	NA	V\$AP1_Q6_01
V\$AP1_Q6_01	1.0×10^{-5}	>0.8	V\$SRF_01
V\$SRF_01	1.0×10^{-5}	NA	V\$AP1_01
V\$ATF4_Q2	1.0×10^{-5}	>0.8	V\$AP1_Q6
V\$TCF11_01	1.0×10^{-5}	>0.8	V\$SRF_C
V\$AP1_Q2_01	1.7×10^{-5}	>0.8	
V\$AP1F1_Q2	1.7×10^{-5}	>0.8	
V\$CREL_01	1.7×10^{-5}	>0.8	
V\$NRF2_Q4	1.7×10^{-5}	NA	
V\$CREB_Q2	1.7×10^{-5}	>0.8	
V\$CREB_Q3	1.7×10^{-5}	>0.8	
5kb			
matrix	IEG <i>p</i> -value	D-PRG <i>p</i> -value	matrix
V\$PAX3_B	2.0×10^{-2}	>0.8	V\$PAX3_B
V\$KROX_Q6	2.1×10^{-2}	>0.8	V\$KROX_Q6
V\$SRF_Q4	2.1×10^{-2}	>0.8	V\$SRF_Q4