



**Universiteit  
Leiden**  
The Netherlands

## **In silico and wet lab approaches to study transcriptional regulation**

Hestand, M.S.

### **Citation**

Hestand, M. S. (2010, June 29). *In silico and wet lab approaches to study transcriptional regulation*. Retrieved from <https://hdl.handle.net/1887/15753>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/15753>

**Note:** To cite this publication please use the final published version (if applicable).

## Chapter 5

# Genome-Wide Assessment of Differential Roles for p300 and CBP in Transcription Regulation

Yolande F.M. Ramos<sup>1†</sup>, Matthew S. Hestand<sup>2,3†</sup>, Matty Verlaan<sup>1</sup>,  
Elise Krabbendam<sup>1</sup>, Yavuz Ariyurek<sup>2</sup>, Michiel van Galen<sup>2</sup>, Hans van Dam<sup>1</sup>,  
Gert-Jan B. van Ommen<sup>3</sup>, Johan T. den Dunnen<sup>2,3</sup>, Alt Zantema<sup>1‡</sup>,  
Peter A.C. 't Hoen<sup>3‡</sup>

<sup>1</sup>Department of Molecular Cell Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands.

<sup>2</sup>Leiden Genome Technology Center, Leiden University Medical Center, Postzone S4-0P, PO Box 9600, 2300 RC Leiden, The Netherlands.

<sup>3</sup>Department of Human and Clinical Genetics, Leiden University Medical Center, Postzone S4-0P, PO Box 9600, 2300 RC Leiden, The Netherlands.

<sup>†</sup>Both authors contributed equally to the work presented.

<sup>‡</sup>Both authors contributed equally to the work presented

*Nucleic Acids Res.* 2010 Apr 30. [Epub ahead of print]  
Parts of this manuscript have been adapted to more appropriately fit this thesis.

## 5.1 Abstract

Despite high levels of homology, transcription coactivators p300 and CREB binding protein (CBP) are both indispensable during embryogenesis. They are known to largely regulate the same genes. To identify genes preferentially regulated by p300 or CBP, we performed an extensive genome-wide survey using ChIP-seq on cell-cycle synchronized cells. We found that 57% of the tags were within genes or proximal promoters, with an overall preference for binding to transcription start and end sites. The heterogeneous binding patterns possibly reflect the divergent roles of CBP and p300 in transcriptional regulation. Most of the 16,103 genes were bound by both CBP and p300. However, after stimulation 89 and 1944 genes were preferentially bound by CBP or p300, respectively. Target genes were found to be primarily involved in the regulation of metabolic and developmental processes, and transcription, with CBP showing a stronger preference than p300 for genes active in negative regulation of transcription. Analysis of transcription factor binding sites suggest that CBP and p300 have many partners in common, but AP-1 and Serum Response Factor (SRF) appear to be more prominent in CBP-specific sequences, whereas AP-2 and SP1 are enriched in p300-specific targets. Taken together, our findings further elucidate the distinct roles of coactivators p300 and CBP in transcriptional regulation.

## 5.2 Introduction

The primary mechanism to control cellular processes, such as proliferation and differentiation, is by regulation of gene expression (reviewed in (104; 105; 106)). Gene expression is a highly coordinated process that results in the synthesis of messenger RNA after recruitment of histone modifying factors, the pre-initiation complex, and transcription factors (TFs) to regulatory regions of the chromatin. The histone modifications that take place during this process, including methylation and acetylation, play a critical role in gene regulation, and defects have been implicated in many pathological conditions from cancer to autoimmune diseases (107; 108; 109). Recently, chromatin immunoprecipitation (ChIP) has been extensively applied in combination with high-throughput sequencing to map genome-wide chromatin modification profiles in human T cells (110; 111) and in mouse ES cells (112). Binding sites of the insulator binding protein CTCF (110), RNA pol II (113; 110) and several TFs (114; 115; 116; 35) have also been mapped. The acetylation profile in primary human T cells was further investigated by determining the binding of several histone deacetylases (117) and histone acetyltransferases (HATs) including p300. Binding of p300 was found both at genes and at intergenic DNase hypersensitive sites, consistent with binding to enhancers, found in other p300 ChIP-sequencing experiments (118; 119).

The HAT p300 and its family member CREB-binding protein (CBP) are transcription coactivators for a broad range of genes involved in multiple cellular processes such as proliferation, differentiation, apoptosis, and DNA repair (reviewed in (120; 121)). In addition, a number of studies suggested the involvement of p300 and CBP in pathological disorders such as the Rubinstein-Taybi Syndrome (reviewed in (122)) and the development of cancer (reviewed in (123)). Originally, CBP was identified through its association with the phosphorylated TF CREB (124), but CBP and p300 also interact with many other TFs, such as cJun (125), p53 (19), and MyoD (126). Apart from the transcriptional regulation through acetylation of histones and other factors, p300 and CBP can also act as a bridge or as a scaffold between upstream TFs and the basal transcription machinery.

A crucial role for both p300 and CBP in development was shown in mice with a homozygous deletion of either gene (*Ep300* and *Crebbp* for the proteins p300 and CBP) resulting in embryonic lethality at a very early stage (20; 21). Interestingly, the double heterozygous *Ep300*<sup>+/-</sup>/*Crebbp*<sup>+/-</sup> mice also die *in utero* (20), indicating that a fine-tuned balance in the expression of both proteins is needed to ensure the normal development. From phenotypic changes in the knock-out mice it is indicated that p300 and CBP have different functions, which has been further illustrated in additional *in vivo* studies (127; 128; 129). A comparison between the acetyltransferase domains of p300 and CBP showed that they differ structurally (130). In part, this might contribute to their functional differences. However, the current detailed mechanism of action of p300 and CBP and the differences between these transcription coactivators is not clear.

In contrast to the *in vivo* situation, most studies with cells cultured *ex vivo* show similar functions for p300 and CBP, and only limited differential roles for p300 and CBP have been described (reviewed in (120)). To obtain a better insight into genes regulated by the general transcription coactivators p300 or CBP next-generation sequencing of ChIP genomic fragments (ChIP-seq) (35) was performed. ChIP-seq and

ChIP-on-microarray (ChIP-chip) have high correspondence in results, but ChIP-seq offers the advantages of requiring less input material, potential to identify binding sites with low affinity, not being limited to target regions (*i.e.* probes on a microarray), not having hybridization errors and it is less costly for whole genome analysis (35). In this study, we used the glioblastoma cell line T98G. T98G cells can easily be synchronized by serum-deprivation and reintroduced into the cell cycle upon stimulation with serum and TPA. Previously, RNA pol II ChIP was performed in growth factor stimulated T98G cells (131), and this showed that 30 minutes upon growth factor stimulation occupancy of the polymerase at the promoters of immediate early genes was maximal. We observed that maximal occupancy of p300 and CBP at promoters of immediate early genes was also around 30 min (Y.F.M.R., unpublished results).

We show p300 and CBP binding to the chromatin in quiescent and stimulated cells, and alterations in their binding to a large number of genes after stimulation. In most cases there is overlap between regions bound by p300 and CBP, but we also identified distinct regions of binding, indicating specific targets for each of these acetyltransferases. Bound regions were analyzed genome-wide for their position relative to genes and were found to have a preference for transcription start sites (TSSs) and transcript ends. Interestingly, functional classification of target genes suggests that CBP is more involved in the regulation of transcription inhibition than p300. A list of TFs that might be involved in the transcription regulation of the identified genes together with p300 and/or CBP was obtained by searching for enriched TF binding sites (TFBSs) in the bound regions. Results show previously established binding partners, and suggest differences for p300 and CBP in their preferences for TFs.

## 5.3 Materials and Methods

### 5.3.1 Cell Culture, ChIP, qPCR, and Sequencing

Human glioblastoma T98G cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin (100  $\mu\text{g}/\text{ml}$ ) and streptomycin (100  $\mu\text{g}/\text{ml}$ ). Prior to stimulation with serum (20%) and tetradecanoyl phorbol acetate (TPA 100 ng/ml; Sigma), cells were serum starved for 23 days (DMEM supplemented with 0.1% FBS).

For sequencing, chromatin was isolated from serum-starved cells (T0) and from cells stimulated for 30 min with serum and TPA (T30). Chromatin from T30 samples was prepared in duplicate, each being used for individual ChIPs, sequencing and downstream analysis. In addition, for more time-point specific data (analyzed only by ChIP and quantitative PCR) we isolated chromatin at 0, 2, 5, 15, 30, 60, 90, 120 and 360 min following stimulation. Chromatin was prepared and ChIPs were carried out as previously described, including fragmentation by sonication (132) (fragment size 500 bp). Immunoprecipitations were performed for p300 using the p300-(2) antibody produced in our lab (125), and for CBP with a commercially available antibody (A22 from Santa Cruz).

For Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) analysis, RNA was isolated using the SV Total RNA isolation System (Promega Corporation Benelux), according to the manufacturers' protocol, and first-strand cDNA synthesis was performed using 1  $\mu\text{g}$  of RNA and ImProm II reverse transcriptase (Promega Corporation

Benelux).

Quantitative PCR for ChIP and for cDNA samples was carried out using the Applied Biosystems 7900HT Fast Real-Time PCR System with SYBR Green PCR Master Mix (Applied Biosystems Europe). Primers were designed using the Primer Express program from Applied Biosystems (for sequences of primers see Additional Table 5.1). Efficiency of the ChIP is presented as percentage of the input. Expression levels of the genes as determined by quantitative RT-PCR were normalized to *GAPDH*, and fold induction was calculated with reference to the untreated samples ( $t = 0$  minutes).

For ChIP-seq all samples were prepared with Illumina's DNA sampleprep Kit (FC-102-1001) according to the manufacturer's protocol. Single ends of each sample were then sequenced on a single lane of the Illumina Genome Analyzer (GAI for samples CBP T0 and T30-1 and p300 T0 and T30-1, GAI for samples CBP T30-2 and p300 T30-2) for 36 cycles.

### Illumina Genome Analyzer Sequencing Analysis

Sequencing results were run through the standard Illumina GAPipeline (v1.0 for GAI runs and v1.3 for GAI runs) to convert images to reads (unaligned sequences produced by the Illumina Genome Analyzer) and edit for quality (FIRECREST, Bustard and GERALD). A general overview of the entire ChIP-seq analysis is provided in Additional Figure 5.1A. The reads were then trimmed to the first 32 bp to remove lower quality base calls at the 3'-end of the read. These were then run through the developing GAPSS\_R ([www.lgtc.nl/GAPSS](http://www.lgtc.nl/GAPSS)) pipeline. This pipeline took the reads, removed the first base pair (often low quality compared to other 5' nucleotides), converted to FASTA format, aligned to the human reference genome (NCBI build 36) with Rmap v0.41 (40), permitting up to two mismatches, and exported tags (the term for aligned reads) into region files (merging adjacent nucleotides with at least one aligned read into one region, followed by compressing those regions within 100 bp into one (based on a range of compression sizes, see below and Additional Figures 5.1B and 5.2)). The pipeline also created wiggle files (viewable in the UCSC genome browser (103)). These tracks had positions with only a single read removed, in order to create more manageable files.

All unedited wiggle files were concatenated to one with custom Perl scripts and converted to a region file (a range of compression windows (20, 50, 100, 150 and 200 bp) were used) with GAPSS\_R scripts. The compression windows account for small gaps in the genomic sequences covered, such as the result of non-unique genomic sequences (Rmap does not map to these). An appropriate compression size is hard to determine, considering a bigger window results in less regions (Additional Figure 5.2) and therefore specificity, but covers larger genomic repeats. We settled on a window of 100 bp to retain a large number of regions, while at the same time accounting for small repetitive elements. This consensus region file had the number of tags from the individual region files mapped to it with a custom Perl script. To make data more manageable and reduce background or very low affinity binding we removed regions with <6 tags (total over all samples). To further reduce the noise only regions with at least 1 tag/million reads aligned (18.1 tags across all total samples) were evaluated. Without applying this threshold performance was poorer, as addressed in the results.

To annotate regions we downloaded from Ensembl 54 Biomart (3; 4) for all genes (with an HGNC ID) the chromosomal location, strand, gene start, gene end, transcript start, transcript end and gene ID. These were loaded into a custom mysql database that was queried to annotate regions for overlap with genes (including flanking 1 kb). We also annotated for distance to the nearest TSSs and transcript ends. Histograms were plotted in the statistical language R to visualize the distance to TSSs and transcript ends.

### 5.3.2 Statistical Analysis

The statistical language R was used to evaluate reproducibility and overlap across samples and to determine genes with differential TF binding across different conditions. To be able to compare data across samples, samples were scaled to the average total number of tags per condition/coactivator. A square root transformation was applied before calculating the reproducibility and comparability across samples. This was to stabilize the variance, inherent to the counting process, over the entire intensity range (133), and to spread the data points better over the intensity range (Additional Figure 5.3). After this, to give a better estimation of the comparability of the data from the different samples Pearson's correlations were calculated in R. This was done on all regions with abundance  $>1$  tag/million tags and a square root transformation applied before calculating the correlations. The Pearson's correlations on the linear scale were slightly lower.

Subsequently, data were summarized at the gene level by adding all tags within a gene or its 1 kb flanking regions. To determine the genes different between conditions/coactivators Fisher's exact p-values were calculated in R. For each individual gene, a two-by-two table was created containing the number of tags for this gene in condition 1 and condition 2 and the total number of tags in condition 1 and condition 2. We then applied the method of Benjamini and Hochberg to correct for multiple testing.

### 5.3.3 Functional Classification

A list of 250 genes, identified as most significantly different between the time points for each coactivator (T30/T0 with adjusted p-value  $<0.001$ ), was uploaded in DAVID 2008 (89; 134) for functional enrichment analysis. To obtain a general impression of the types of processes in which CBP and p300 are involved, functional annotation charts were generated for the Gene Ontology (GO) term GOTERM\_BP\_ALL (91; 92) using a human background.

In addition, significantly different genes at T30 were divided into two groups where either CBP or p300 binding was higher. From these groups, the 250 genes most significantly different were uploaded in DAVID 2008 for functional enrichment analysis. Individual GO-terms with a p-value  $<0.001$  are shown for genes with higher CBP or p300 binding.

### 5.3.4 CORE\_TF Analysis for TF Partners of p300/CBP

We took the same significant gene sets as from the functional analysis and retrieved the most substantially sequenced region (most number of tags in this particular re-

gion) for these genes. These regions were extended at both sides to a final length of 2 kb and sequences retrieved with Ensembl Perl API. As a background set, we retrieved 3000 random genes' TSSs from Ensembl Biomart that were located on chromosomes 1–22, X and Y and retrieved the sequences  $\pm 1$  kb from these TSSs. The regions based on significantly different genes were entered into CORE.TF (76) as experimental sequences and the random TSS sequences were entered as background regions. We evaluated enrichment of TFBSs (defined as TRANSFAC (51) position weight matrices) in the experimental sequences using the most stringent Match setting (55; 51) to minimize false positives. P-values representing the significance of over-representation were calculated with a binomial test.

## 5.4 Results

### 5.4.1 Initial Sequencing Analysis

Stringent regulation of gene expression is fundamental to control cellular processes such as proliferation and differentiation. The general coactivators p300 and CBP play an important role in the regulation of gene transcription by virtue of their acetyltransferase activity. We set out to determine and compare genes regulated by p300 and CBP. Chromatin was isolated from serum-starved (T0) and from stimulated (T30, done in duplo) human glioblastoma cells and ChIP-seq performed using CBP- and p300-specific antibodies.

Sequence files generated by the Illumina GAPipeline were submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/Traces/sra>: SRS009476, SRS009457, SRS009477, SRS009478, SRS009479 and SRS009480) (135). The reads passing quality control were mapped to the human reference genome and adjacent tags were joined into regions (Table 5.1). We also have made sequencing data available as UCSC hg18 viewable wiggle tracks (excluding positions with only one tag aligned, Additional Files 5.1-5.6).

### 5.4.2 Preferential Binding in Genes and Promoters

Without applying a threshold of 1 tag/million tags, we found low overlap of identified regions in the replicated samples indicating that regions with low abundance represent noise (data not shown). With the threshold of 1 tag/million tags, we found a high consistency in the identified regions between all samples (47.96 and 47.43% overlap between CBP and p300 at T0 and T30, respectively; Table 5.2). Concordantly, the reproducibility between biological replicates was high (Pearsons correlation: 0.77 and 0.87 for CBP and p300, respectively). A similarly high correlation was found across the different samples (Pearsons correlation 0.81 on average between all time points and coactivators; Additional Table 5.2), indicating relatively minor differences in the distribution of p300 and CBP binding sites across the genome. In subsequent analyses, datasets of the T30 biological replicates were summed and treated as one sample, which provided us with high-quality results.

To study the biological implications of our data, we annotated the regions obtained from sequencing with Ensembl and found that the sequenced regions covered 16,103 annotated genes in total. When looking at conditions and coactivators independently

Table 5.1: Sequencing Results

CBP				
t(min)	# reads	# aligned	% aligned	# regions*
T0	5498759	4018590	73	713141
T30 <sup>1</sup>	6389605	4849826	76	889781
T30 <sup>2</sup>	6047530	5001204	83	851988
p300				
t(min)	# reads	# aligned	% aligned	# regions*
T0	6327413	5086340	80	841029
T30 <sup>1</sup>	6446269	5156450	80	802627
T30 <sup>2</sup>	6065594	5124836	84	684222

The total number of reads, reads aligned, percentage aligned and number of regions created (\*after compressing regions within 100 bp into one and excluding regions composed of only a single tag) for each condition (T0: quiescent cells and T30: 30 min after growth factor stimulation) and for each transcription coactivator (CBP or p300). For T30-independent biological replicates were sequenced as indicated by <sup>1</sup> and <sup>2</sup>.

Table 5.2: Region overlap

	CBP T0	CBP T30 <sup>2</sup>	CBP T30 <sup>1</sup>	p300 T0	p300 T30 <sup>2</sup>	p300 T30 <sup>1</sup>
CBP T0	267562	129089	133493	140245	120092	134799
CBP T30 <sup>2</sup>		315020	143160	152520	136405	148669
CBP T30 <sup>1</sup>			322556	151519	136685	150180
p300 T0				322354	138933	158708
p300 T30 <sup>2</sup>					267804	139592
p300 T30 <sup>1</sup>						304880

The number of regions, after applying thresholds (> 1 tag/million tags), overlapping between conditions (T0 and T30) and coactivators (CBP and p300). For T30-independent biological replicates were sequenced as indicated by <sup>1</sup> and <sup>2</sup>.

there were 16,045, 16,075, 15,684, and 15,996 genes identified as bound by CBP at T0 and T30, and by p300 at T0 and T30, respectively. We observed similar percentages of tags in genes and their 1 kb flanking regions in all samples (57.08, 57.10, 57.30 and 59.93% for CBP-T0, CBP-T30, p300-T0 and p300-T30, respectively). Therefore, both CBP and p300 appear to be needed to maintain basal levels of expression in quiescent cells as well as to activate or repress transcription after serum stimulation.

Previous studies have focused on the binding of p300 to enhancers (118; 117). First, we evaluated the distance for all regions bound by CBP or p300 to the nearest TSS and transcript end (polyadenylation site). We found that genome-wide, 57% of all tags could be annotated to genes (1 kb) and a clear preference for TSSs and transcript ends was observed (Figure 5.1A and B). There were no apparent differences between the profiles of CBP and p300 (data not shown). Also, different from what has been shown before for most TFs or histone modification maps, p300 and CBP show three distinct patterns of binding, including a distinguished peak (binding to a specific site like the TSS, *e.g.* *ZNF688*; Figure 5.1C), binding across the gene with

no clear preference for a specific region (*e.g.* *EGR1*; Figure 5.1D), and so-called U-shaped binding (binding across the gene with a bias toward the TSS and transcript end, *e.g.* *DUSP1*; Figure 5.1E).

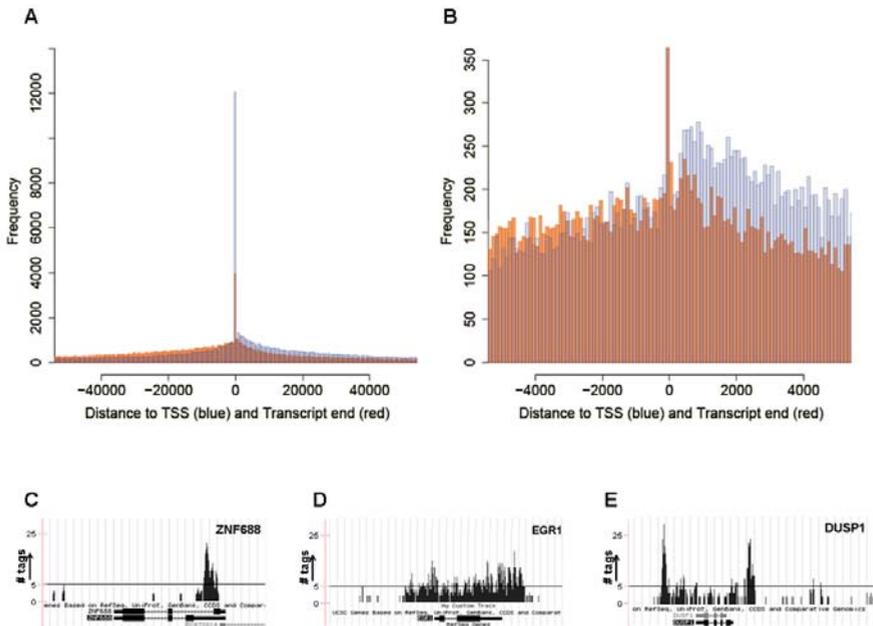


Figure 5.1: Histogram for the compilation of ChIP-seq regions showing the frequency of the distance from the localization of a sequenced region to the nearest transcription start site (blue) and transcript end (red) (full plot in (A), zoomed in (B)), which indicates a preference for binding to TSSs and transcript ends (color figure available at <http://nar.oxfordjournals.org/cgi/content-nw/full/gkq184v1/F1>). Representative examples of the different types of binding are shown as custom tracks on the UCSC genome browser: binding to a specific site resulting in a peak (C), binding across the gene (D), and U-shaped binding, with binding across the gene with preference for both TSS and transcript end (E). The y-axis indicates the number of tags aligned at each position in the genome. The black line in Figure 5.1C-E indicates a value of 5 tags in the custom tracks.

### 5.4.3 Differential Binding by CBP and p300

With most data corresponding to a genic region, we focused our following analyses to genes, and on those regions within 1 kb upstream of TSSs and 1 kb downstream transcript ends, (16,103 genes across all four samples). Since we were especially interested in genes that were preferentially regulated by p300 or CBP during entry in the cell cycle, a Fishers exact test was performed to determine statistically significant differences in the total number of tags localized to a certain gene in different conditions

(between time points or between coactivators) studied.

Despite high overlap in regions bound by CBP and p300 in quiescent and in stimulated cells (Table 5.2), there was also a considerable number of quantitative changes in CBP and p300 binding upon stimulation. Significant differences between p300 and CBP binding was found for 120 and for 1611 genes at a false discovery rate of 0.1% at T0 and T30, respectively (Figure 5.2A). At a false discovery rate of 1% this was 256 and 2502 at T0 and T30, respectively (Additional Table 5.3). From the genes differentially bound by p300 and CBP in quiescent cells (T0), only 25 did not have significantly different binding upon growth factor stimulation (Figure 5.2A). These results indicate very high overlap in genes bound by p300 as well as by CBP in the quiescent state and a divergence of the roles of CBP and p300 mainly during periods of activated transcription. Analysis of the 250 genes that were most significantly different in our data, showed that for the majority p300 binding was higher than CBP binding (191 and 227 of 250 genes, for T0 and T30, respectively).

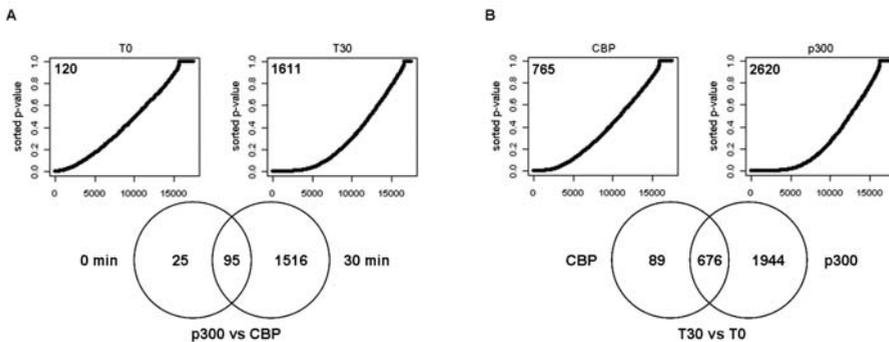


Figure 5.2: Genes differentially bound by CBP and p300 (A) and between time points (B). P-values (Fishers exact test) for the indicated comparisons were sorted in rising order and plotted (Upper panels). Under the null hypothesis of no significant differences, this would give a straight line on the diagonal. However, as becomes evident by the curve shape there is a bias towards low p-values. The number of genes with significant differences between conditions are indicated in the graphs (false discovery rate of 0.1%). Venn diagrams (Lower panels) demonstrate the number of significantly different bound genes, as shown in the plots above, that overlap between time points (A) or coactivators (B).

When comparing between time points, we found 765 genes differentially bound by CBP and 2620 genes differentially bound by p300 (Figure 5.2B). Of the 250 genes, which were most significantly different between time points, the majority (209 and 155 for CBP and p300, respectively) demonstrated higher binding for both, p300 and CBP, at T30 when compared to T0. In addition, the majority of genes with changed binding of CBP after stimulation also demonstrated difference in binding by p300 (676 out of 765 and 2620 genes, respectively; Figure 5.2B). The apparently higher number of genes with significant changes in p300 binding is likely due to the higher efficiency of the p300 antibody causing better signal-to-noise ratios and higher sensitivity in the detection of quantitative changes in binding profiles (see below). The high level of

overlap between coactivators can explain the restricted number of differences found thus far in functions of p300 and CBP.

We present a full list of genes bound by CBP and p300 at T0 and T30 in Additional File 5.7. Table 5.3 lists the 10 genes for which levels of binding differ most significantly between the four samples. Among the genes with strongest CBP and p300 binding and most significantly different between T30 and T0 are many immediate-early genes that are bound by both p300 and CBP (Table 3; *e.g.* *ATF3*, *FOSB*, and *DUSP1*).

Table 5.3:

CBP T30 vs T0			p300 T30 vs T0		
gene	p-value	ratio	gene	p-value	ratio
CATSPER3	0	2.58	THBS1	$6.13 \times 10^{-251}$	6.07
ATF3	$5.58 \times 10^{-101}$	4.51	ATF3	$1.50 \times 10^{-240}$	5.83
TRIP13	$4.63 \times 10^{-94}$	11.07	FOSB	$2.22 \times 10^{-213}$	14.03
CYR61	$4.15 \times 10^{-80}$	6.43	CYR61	$1.27 \times 10^{-187}$	6.33
FOSB	$5.04 \times 10^{-75}$	10.03	EGR1	$1.66 \times 10^{-178}$	14.11
SMAD3	$8.18 \times 10^{-72}$	2.09	TPM1	$2.25 \times 10^{-175}$	4.81
TMEM49	$4.60 \times 10^{-71}$	2.32	DUSP1	$2.05 \times 10^{-147}$	6.81
MYH9	$2.15 \times 10^{-69}$	2.39	MYH9	$8.27 \times 10^{-144}$	2.84
CRISPLD2	$1.45 \times 10^{-67}$	2.89	NR4A1	$4.74 \times 10^{-143}$	13.16
THBS1	$1.32 \times 10^{-66}$	4.15	CRISPLD2	$1.45 \times 10^{-140}$	4.04
T0 p300 vs CBP			T30 p300 vs CBP		
gene	p-value	ratio	gene	p-value	ratio
CXXC1	$8.28 \times 10^{-229}$	22.71	CXXC1	$3.06 \times 10^{-43}$	20.32
AKT1S1	$1.39 \times 10^{-192}$	6.03	MKKS	$1.66 \times 10^{-41}$	3.87
FBXL19	$9.96 \times 10^{-172}$	5.56	CATSPER3	$1.07 \times 10^{-30}$	1.44
MKKS	$1.67 \times 10^{-154}$	4.73	AKT1S1	$6.27 \times 10^{-24}$	4.46
C3orf19	$2.39 \times 10^{-135}$	7.93	FAM40A	$1.94 \times 10^{-22}$	4.21
BSCL2	$3.52 \times 10^{-130}$	8.25	FBXL19	$1.09 \times 10^{-21}$	3.27
THBS1	$5.46 \times 10^{-120}$	2.1	ZNF350	$1.09 \times 10^{-21}$	9.14
MADCAM1	$1.24 \times 10^{-112}$	7.85	METTL3	$1.50 \times 10^{-19}$	13.47
ZNF175	$1.01 \times 10^{-103}$	17.86	MADCAM1	$1.30 \times 10^{-18}$	7.2
C1orf174	$1.17 \times 10^{-102}$	9.84	C1orf174	$1.84 \times 10^{-17}$	7.25

Top 10 genes that are most significantly different between time points and coactivators according to the p-values of the Fishers exact test. The ratio shows the quantitative difference in binding as expressed by the number of tags between the two samples that are compared: from T30 and T0 (upper half of the table) and from p300 and CBP (lower half of the table).

#### 5.4.4 Validation

To validate our results and to refine the temporal resolution of the experiment, genes were selected to further characterize with ChIP and quantitative PCR in a time-course from 0 to 360 min following stimulation with serum and TPA. The genes included genes bound by both CBP and p300 and genes unique to one of the coactivators, and spanned a wide range of significance values (Figure 5.3E). In general, the recov-

ery obtained (as a percentage of the input) for CBP is lower than for p300 (Figure 5.3A-D), consistent with the generally lower number of tags for CBP in each region of the ChIP-seq experiment (significant quantitative correlation between results of the qPCR and ChIP-seq experiment for the genes presented here are shown in Additional Figure 5.4). The qPCR results also confirm the differential binding across time points established with ChIP-seq analysis for all genes analyzed (Figure 5.3A-D and Additional Figure 5.5), and demonstrate that for most genes the temporal binding pattern is comparable between CBP (black bars) and p300 (white bars). This is true for the increased binding to the promoter of *CTGF*, as well as for the decreased binding to the promoter of *ZNF608* in stimulated cells compared to unstimulated cells (Figure 5.3A and B). Binding to the promoter of *CDK5* differs for p300 and CBP (Figure 5.3C). Binding of p300 is increased in time with a maximum at 60 min post-stimulation, while there is hardly any change in the binding of CBP. These results correlate with the statistical analysis, that demonstrated significant changes between p300 and CBP at T30, and a significant increase in p300 but not CBP binding between T30 and T0 (Figure 5.3E). The binding of p300 and CBP to the *SERPINE1* gene increased significantly over time (p-value of  $1.88 \times 10^{-17}$  for CBP T30 versus T0 and  $1.59 \times 10^{-24}$  for p300 T30 versus T0). Inspection of the wiggle track (Figure 5.3D) revealed that p300 and CBP bound mainly to the 3'-UTR and to a lesser extent to the region around the TSS of the *SERPINE1* gene. Also, the small increase observed around the TSS could be confirmed by qPCR. The wiggle file for *SERPINE1* also shows a stronger binding >2 kb upstream of the TSS (Figure 5.3D). The interaction to this putative enhancer region and the change upon stimulation was also confirmed by qPCR of ChIP samples (Additional Figure 5.5J).

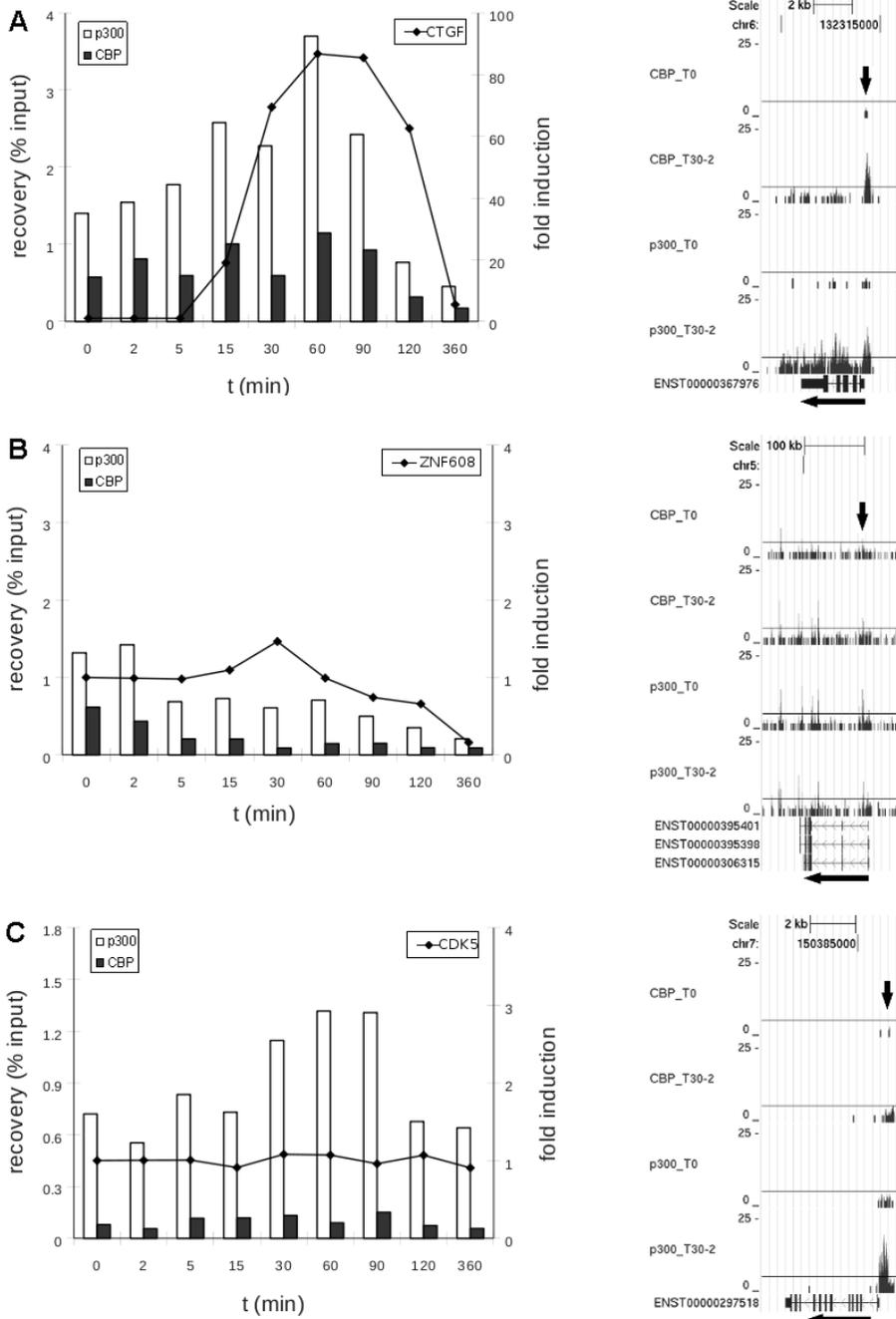


Figure 5.3 (continues on next page)

To evaluate whether changes in p300 and CBP binding also affected gene expression, we performed quantitative RT-PCR for the genes *CTGF*, *ZNF608*, *CDK5*, and *SERPINE1* (Figure 5.3A-D: the line in the graphs shows fold induction in the time course). For the three genes with increased binding, two (*CTGF* and *SERPINE1*)

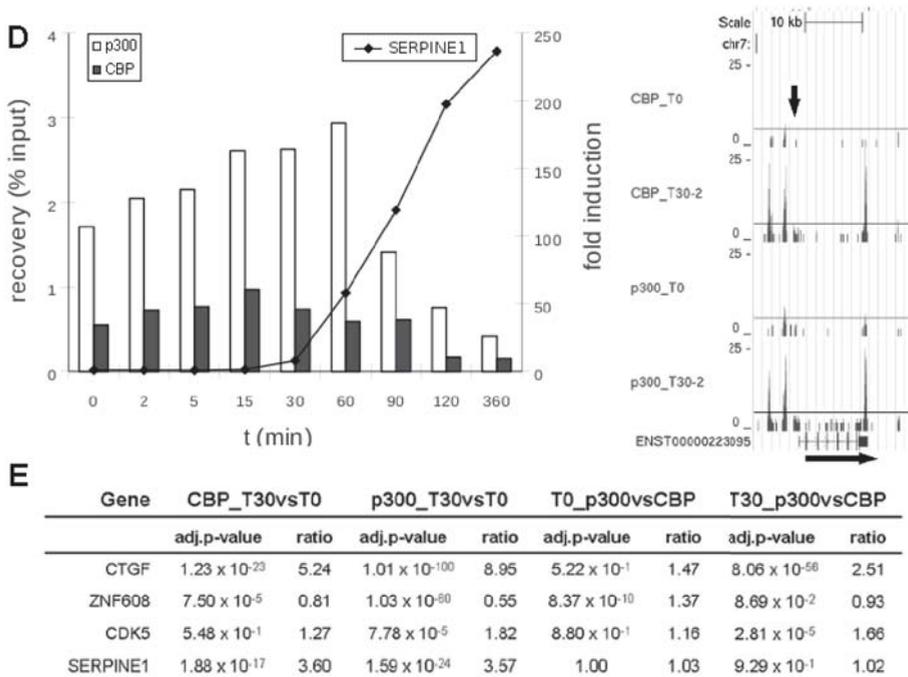


Figure 5.3: (continued from previous page): ChIP-analysis for time-course experiment (0, 2, 5, 15, 30, 60, 90, 120 and 360 min after stimulation of serum-starved T98G cells with serum and TPA). Shown are graphs for qPCR results (x-axis: time in minutes; left y-axis: ChIP recovery in percentages of the input; right y-axis: fold induction for the RT-qPCR with reference to the untreated samples ( $t = 0$  min)) and screen-shots from custom tracks of the UCSC genome browser for the ChIP-seq data (T0 and T30 only) for *CTGF* (A), *ZNF608* (B), *CDK5* (C), and *SERPINE1* (D). White bars: p300 ChIP; black bars: CBP ChIP; RT-qPCR data are indicated as dots, interconnected; arrows in the screen-shots indicate the position of the PCR-amplicon. Also indicated for these genes is the adjusted P-value and the ratio difference between time-points (T30 versus T0) and coactivators (p300 versus CBP) of the total number of reads along the whole gene, plus 1 kb up- or downstream from all ChIP-seq data (E).

show increased expression. The gene *ZNF608* shows a decrease in expression over time, consistent with decreased binding of p300/CBP. *CDK5* did not show any differences in expression. This is consistent with the uniform levels of CBP binding over time, but not with the increased binding of p300. Most likely, for *CDK5* and possibly also for other genes binding of p300/CBP is not sufficient to induce the expression but other factors that play a critical role are also required. Obviously, gene expression is a complex process and highly variable between genes, so only detailed studies can unravel the role of specific factors.

### 5.4.5 Biological Processes Coordinated by p300 and CBP

To get an impression of the biological implications of p300 and CBP binding, we clustered genes regulated by CBP and p300 into functional pathways. We used DAVID 2008 (89; 134) to classify the 250 genes most significantly differing between time points (for both CBP and p300). The analysis (p-value < 0.001) shows that CBP as well as p300 are mainly involved in transcription regulation of genes controlling developmental processes and metabolic processes (such as *NR4A*, *CRISPLD2*, *CRIM1*, *CYCLIN-L1*, and *PER1*) and of genes coding for proteins that control gene expression (such as *ATF3*, *FOSB*, *SP3* and *HES1*; see Additional File 5.8). Next, using DAVID 2008 we wanted to specify in more detail whether certain groups of genes were preferentially bound by CBP or by p300. Remarkably, in the cluster of genes regulating transcription, those with significantly higher CBP than p300 binding are involved in negative regulation of transcription (Table 5.4, and Additional File 5.8). Another interesting observation for genes preferentially bound by CBP is the presence of clusters related to signal transcription/cell communication. In the list obtained for higher levels of p300, mainly clusters relate to transcription and metabolic processes are found.

### 5.4.6 Analysis of ChIP-seq Regions for Consensus Transcription Factor Binding Sites

CBP and p300 do not bind DNA directly, but regulate by binding to many different protein partners. Therefore, to identify (DNA-binding) partners of p300/CBP, we looked for enrichment of TFBSs in and around the regions bound by CBP and/or p300 in the 250 genes that differ most significantly between time points (the same genes that were used for DAVID analysis). We found a significant over-representation of AP-1, CREB, NFKB and SRF binding sites in the gene regions bound by both CBP and p300 (Table 5.5 and Additional File 5.9), which are known to be regulated by CBP and/or p300 (136). As mentioned before, there is more binding of p300 and CBP to the chromatin at T30 after stimulation. Therefore, enrichment of the TFBSs in our sequences likely reflects increased binding of these factors upon growth factor stimulation.

We also compared genes significantly different between coactivators at T30 using the same lists of 250 genes as used for the functional classification. CREB and YY1 are significantly enriched in both gene sets (Table 5.5; all results are presented in Additional File 5.9). However, CBP binding was found to correlate more with AP-1 and SRF binding partners than p300, whereas p300 binding was more correlated to AP-2, E2F and SP1-binding. These results indicate that CBP and p300 share some, but not all, regulatory partners.

## 5.5 Discussion

Transcription coactivators CBP and p300 share high levels of homology and, in many cases, the same regulatory regions are targeted for transcription regulation. This is in contrast with the fact that both proteins are indispensable during embryogenesis. To investigate which genes are regulated, and whether there is a difference in those

Table 5.4: Functional classification for genes bound by CBP or p300

T30 p300 higher than CBP				Enrichment	
ID <sub>(GO:#)</sub>	Term	Count	%		
0010467	gene expression	81	38.76	$2.57 \times 10^{-12}$	2.06
0044237	cellular metabolic process	127	60.77	$3.75 \times 10^{-10}$	1.44
0008152	metabolic process	135	64.59	$5.27 \times 10^{-10}$	1.38
0044238	primary metabolic process	126	60.29	$1.36 \times 10^{-9}$	1.42
0043170	macromolecule metabolic process	111	53.11	$8.79 \times 10^{-8}$	1.44
0006350	transcription	56	26.79	$4.10 \times 10^{-7}$	1.95
0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	72	34.45	$1.41 \times 10^{-6}$	1.66
0045449	regulation of transcription	53	25.36	$1.82 \times 10^{-6}$	1.91
0016070	RNA metabolic process	58	27.75	$3.07 \times 10^{-6}$	1.8
0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	53	25.36	$3.64 \times 10^{-6}$	1.87
0010468	regulation of gene expression	54	25.84	$4.93 \times 10^{-6}$	1.83
0031323	regulation of cellular metabolic process	54	25.84	$1.62 \times 10^{-5}$	1.76
0043283	biopolymer metabolic process	84	40.19	$1.78 \times 10^{-5}$	1.47
0019222	regulation of metabolic process	55	26.32	$2.08 \times 10^{-5}$	1.73
0006351	transcription, DNA-dependent	48	22.97	$3.29 \times 10^{-5}$	1.81
0032774	RNA biosynthetic process	48	22.97	$3.39 \times 10^{-5}$	1.81
0006355	regulation of transcription, DNA-dependent	46	22.01	$8.66 \times 10^{-5}$	1.77
0006979	response to oxidative stress	7	3.35	$5.49 \times 10^{-4}$	6.83
0050794	regulation of cellular process	67	32.06	$8.46 \times 10^{-4}$	1.42
T30 CBP higher than p300				Enrichment	
ID <sub>(GO:#)</sub>	Term	Count	%		
0051056	regulation of small GTPase mediated signal transduction	18	7.06	$8.70 \times 10^{-9}$	5.97
0046578	regulation of Ras protein signal transduction	13	5.10	$5.62 \times 10^{-6}$	5.36
0007242	intracellular signaling cascade	38	14.90	$2.47 \times 10^{-5}$	2.06
0009966	regulation of signal transduction	21	8.24	$2.49 \times 10^{-5}$	2.98
0007154	cell communication	77	30.20	$3.31 \times 10^{-5}$	1.52
0007165	signal transduction	71	27.84	$6.04 \times 10^{-5}$	1.54
0007265	Ras protein signal transduction	13	5.10	$9.42 \times 10^{-5}$	4.03
0007264	small GTPase mediated signal transduction	18	7.06	$1.32 \times 10^{-4}$	2.94
0007399	nervous system development	23	9.02	$2.26 \times 10^{-4}$	2.4
0016481	negative regulation of transcription	13	5.10	$3.31 \times 10^{-4}$	3.52
0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	13	5.10	$7.31 \times 10^{-4}$	3.22

Significantly enriched GO categories for genes that show higher binding of CBP or p300 at 30 min after stimulation with TPA and serum (ID: GO-category-number, term: description of the GO category; count: number of significant genes in this GO category; %: percentage of significant genes in this GO category; p-value: statistical significance of the GO category (p-value from hypergeometric test for over-representation); enrichment: fold enrichment of significant genes compared to the background.

regulated by p300 and by CBP upon growth factor stimulation a genome-wide screen was performed in T98G cells. Although there is a high concordance between binding targets of p300 and CBP, and both seem to regulate the same biological pathways, we have identified significant differences in the levels and targets of binding. These differences include the diversity in the regulation of genes involved in transcription, and in cell death and cell adhesion. In addition, regulatory regions of these genes showed significant differences in binding sites of other TFs and TF families such as

Table 5.5: Enrichment for transcription factor binding sites in CBP/p300 bound sequences

TFBS	T30 vs T0		T30		
	CBP	p300	TFBS	CBP>p300	p300>CBP
AP-1	0	0	AP-1	0	$2.83 \times 10^{-2}$
CREB	$8.21 \times 10^{-5}$	$1.84 \times 10^{-5}$	AP-2	$9.69 \times 10^{-1}$	$7.30 \times 10^{-9}$
NFKB	$4.55 \times 10^{-7}$	$1.49 \times 10^{-4}$	CREB	$2.96 \times 10^{-4}$	$7.60 \times 10^{-9}$
SRF	$3.41 \times 10^{-7}$	$1.51 \times 10^{-5}$	E2F	$2.39 \times 10^{-1}$	0
			SP1	$9.98 \times 10^{-1}$	$2.21 \times 10^{-7}$
			SRF	$8.81 \times 10^{-4}$	$2.12 \times 10^{-1}$
			YY1	$2.55 \times 10^{-6}$	0

TFBSs with the most significant p-values for enrichment in regions bound by CBP and p300 at 0 and 30 min after stimulation with serum and TPA.

AP-1, AP-2, SP1, E2F and SRF.

It is well established that p300/CBP associate to both enhancers and TSSs. Previous studies have focused on the enhancer-binding of p300 (119; 118; 117). Although we also found examples of enhancer-binding, over 57% of all tags are within genes or proximal promoters (+/-1 kb), and genome-wide we find that binding is primarily located around TSSs and to some extent also to transcript ends (Figure 5.1A and B). Therefore, we chose to focus our analysis of CBP/p300 in relation to genic regions. In all, we found 16,103 genes bound by CBP or p300 at T0 or T30, with over 97.4% of genes bound by both coactivators at both time points.

When analyzing the binding of CBP/p300 to genes, we did not only observe distinct regions of binding. There was a high variety in binding patterns for both coactivators. This includes binding to a clear and distinct region (*e.g.* to the TSS; referred to as peak), binding across the gene, or a combination of more prominent binding around the TSS and the transcript end, as well as binding across the gene (in the text referred to as U-shaped binding; Figure 5.1C-E).

At present, the mechanisms that determine the diverse binding patterns remain to be established. Possibly, it is dependent on the way p300/CBP regulate transcription of a particular gene. Both, p300 and CBP can bind to specific TFs, and this might result in a distinguished peak around the TSS and transcript end. In addition, p300 and CBP regulate chromatin structure via the acetylation of histones, thereby making the chromatin more prone to be targeted for transcription. This might account for binding (to the histones) across the gene. Binding across a gene was previously described to occur also by protein kinases (137). Chow *et al.* (137) propose that in this way the kinases may contribute to transcription initiation and elongation, or processes such as 5' capping, and splicing. Binding to both the TSS and transcript end has previously been observed for RNA polymerase II (138). Interaction between CBP/p300 and RNA polymerase II (139) may explain the presence of similar ChIP-seq patterns for these acetyltransferases. Enrichment at the TSS might correlate to the longer time needed for transcription initiation compared to transcript elongation. The peak at the transcript end might correlate to widespread transcription of antisense transcripts (140), a phenomena that is particularly prominent in the 3'-end of genes (36).

Our ChIP-seq data are from arrested cells and from cells 30 min after stimulation. Therefore, over-representation of genes required early in the cell cycle was expected at T30. The number of reads correlates roughly to the binding affinity of proteins for that region and immediate-early genes are among the genes with the highest number of tags. Analysis of a number of these genes with quantitative RT-PCR (Figure 5.3A, Additional Figure 5.5, and data not shown), also showed increase in gene expression for immediate-early genes. Our data suggest that at T30 CBP and p300 are more intimately involved in the regulation of transcriptional activation of immediate-early genes compared to other groups of genes. Consistently, Tullai *et al.* (131) previously published microarray data on gene expression of serum-starved T98G cells upon growth stimulation. We found that from 49 immediate-early genes that were identified, 36 demonstrated significantly increased binding by CBP and p300 at T30 compared to T0 in our analysis (3 out of 49 could not be identified in Ensembl).

With the time-course experiment, most genes that were analyzed show maximal binding between 30 and 60 min after stimulation. The time-course experiment confirms high accuracy of our data since all genes tested, although different levels of significance (from  $2 \times 10^{-147}$  to  $9 \times 10^{-1}$ ) and variable ratios of difference (from 1 to 9) were chosen, confirm binding of the coactivators and changes in time.

Binding of p300 and CBP to the chromatin occurs through the interaction with TFs. To obtain more insight in transcription regulatory complexes bound by p300 and/or CBP, we set out to identify possible partners of CBP and p300 for the genes identified in our experiment. Therefore, we analyzed for the enrichment of TFBSs. When looking at genes with significant binding at T30 for each coactivator, we found some examples of TFBSs that were found to be specific only for CBP or for p300. For example, AP-1 and SRF binding sites were significantly enriched in CBP bound regions, while AP-2, E2F and SP1 binding sites were more abundant in p300 bound regions. This may represent TFs that are regulated during the cell cycle, in most cases, solely by CBP or p300 and contribute to their unique functions.

We observed overlap in enrichment of TFBSs for proteins such as YY1 and CREB. Interestingly, YY1 is known to contribute to cell-cycle regulation and can serve both, as a transcriptional repressor and an activator (141). Also, YY1 is known to interact with p300/CBP, as well as with other TFs identified in this study (AP-1, AP-2, NFKB, E2, SP1 and CREB) (141; 142; 143). Our functional classification suggested that CBP is more associated with transcriptional repression, whereas p300 is more associated with transcriptional activation. It could be speculated that YY1 is a putative partner involved in this functional difference between p300 and CBP, while the p300-YY1 complex might activate transcription *in vivo*, the CBP-YY1 complex might account for transcriptional repression.

In the future, it would be valuable to perform ChIP-seq in the same cell line and conditions with antibodies for the coactivator specific TFs in this study (AP-1, AP-2, SP1, E2F and SRF, and YY1). This will confirm whether genome-wide CBP/p300 and their specific regulatory partners cooperate, and it will help to further elucidate their role in cell-cycle control. In addition, ChIP-seq with antibodies specific to open chromatin states will be helpful to unravel the mechanisms leading to the diverse binding patterns.

## 5.6 Conclusion

Transcription coactivators CBP and p300 share high levels of homology and, in many cases, the same regulatory regions are targeted for transcription regulation. This is in contrast with the fact that both proteins are indispensable during embryogenesis. To investigate which genes are regulated, and whether there is a difference in those regulated by p300 and by CBP upon growth factor stimulation a genome-wide screen was performed in T98G cells. Although there is a high concordance between binding targets of p300 and CBP and both seem to regulate the same biological pathways, we have identified significant differences in the levels and targets of binding. In addition, regulatory regions of target genes also showed significant differences in TFBSs of other TFs such as AP-1, AP-2, SP1, E2F and SRF.

Besides the differences in targets of p300 and CBP, we identified various binding-patterns that potentially correlate with different types of transcription regulation by p300 and CBP. Most interestingly, we observed a so-called U-shaped binding with high levels of p300/CBP at both, TSS and transcript end. Possibly, the acetyltransferases contribute to other processes such as transcription elongation and reverse transcription. Taken together, our data contribute to the improvement of our knowledge of processes that regulate gene expression by the transcription coactivators p300 and CBP, and confirm that regulation by these coactivators is not identical.

## 5.7 Funding

Centre for Medical Systems Biology within the framework of the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO); Center for Biomedical Genetics (in the Netherlands).

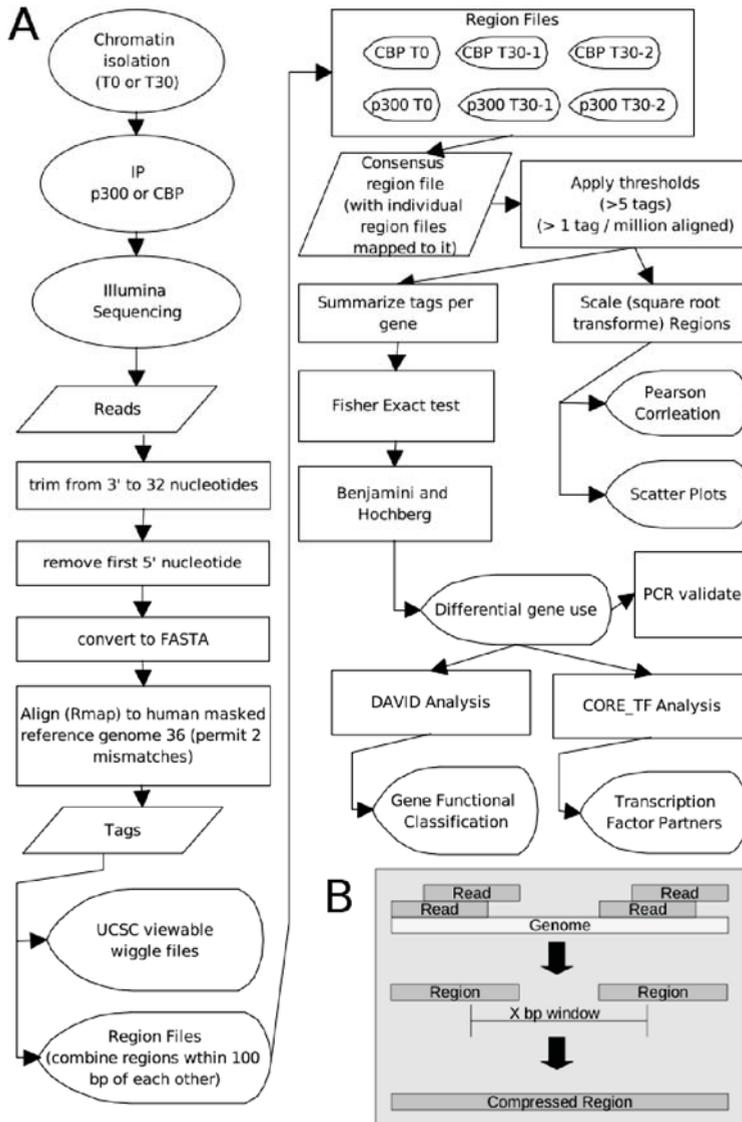
*Conflict of interest statement.* None declared.

## 5.8 Acknowledgements

We wish to thank Michel P. Villerius for his computational support and Dorien JM Peters for her review and comments about the manuscript.

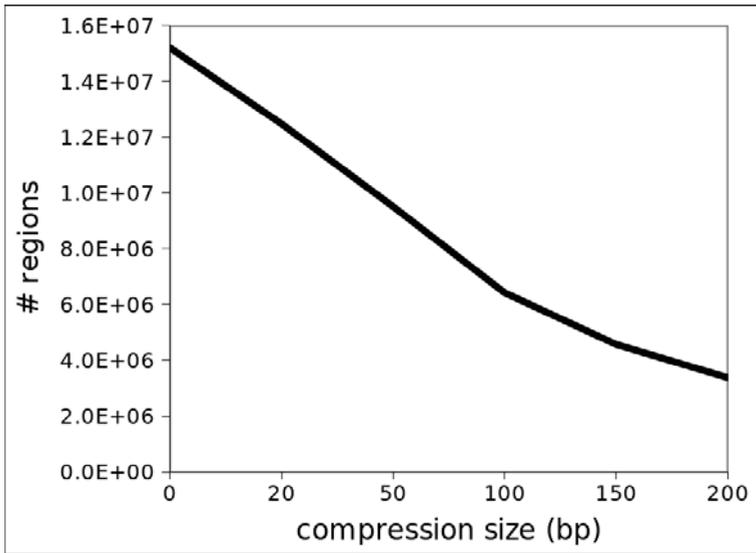
## 5.9 Additional Files

Additional Figure 5.1



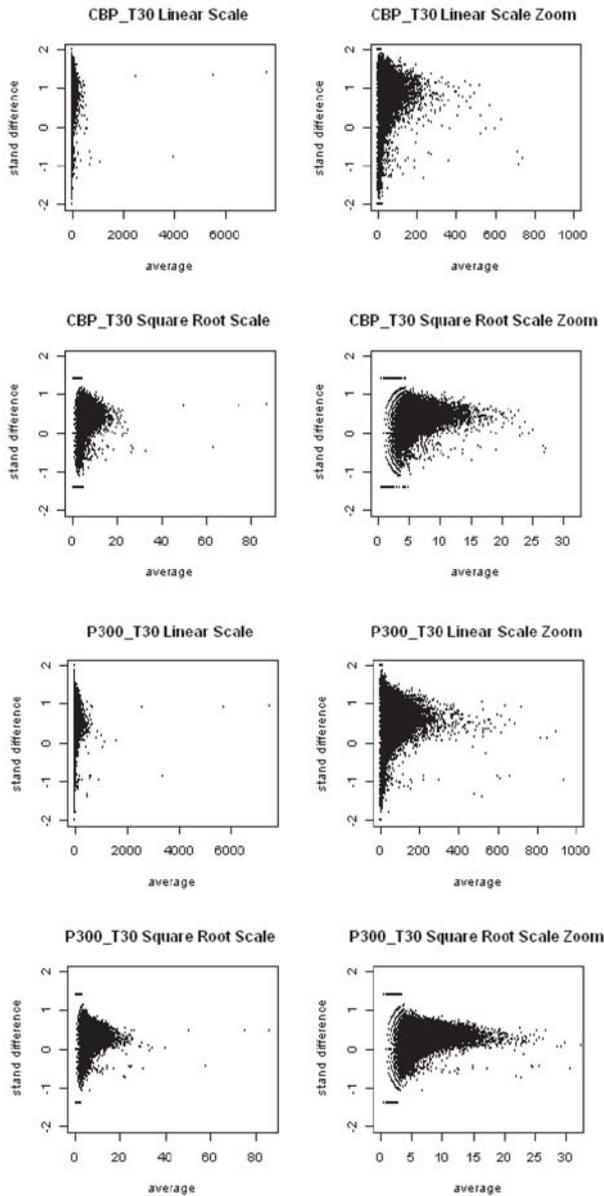
Flow chart of the experimental set-up (A). Also demonstrated is the creation of regions from aligned reads and compressing regions within a window of X bps (B).

Additional Figure 5.2



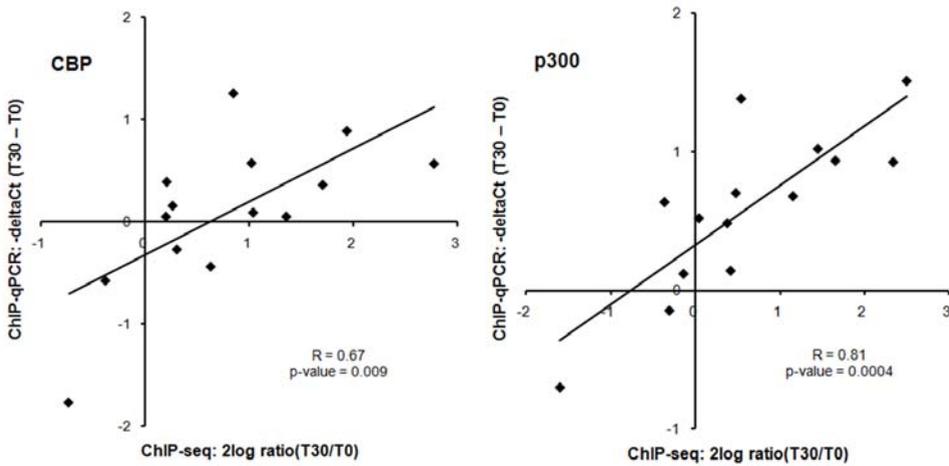
To demonstrate the effect of compressing regions with variable window sizes (x-axis) we plotted the number of regions obtained (y-axis).

## Additional Figure 5.3



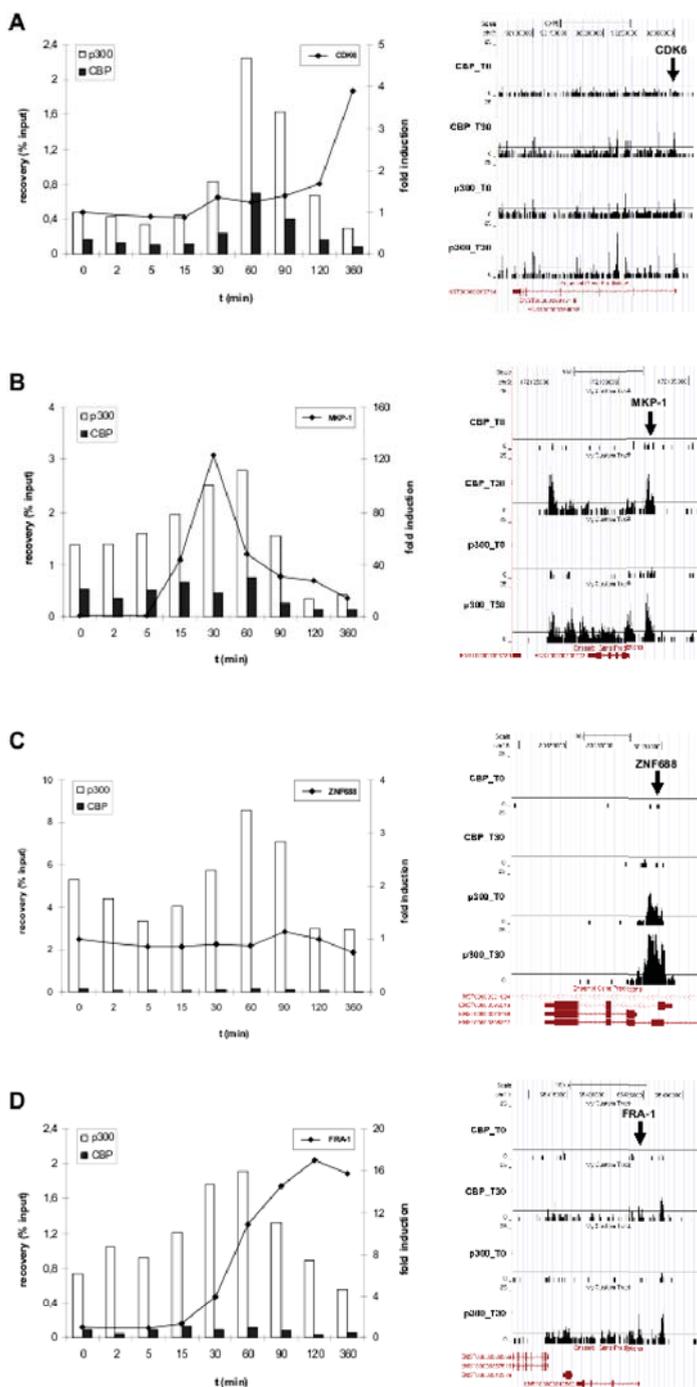
To demonstrate the variance stabilizing property of the square root transformation, we plotted the standardized difference (=difference divided by the mean) of the tags per region in the two biological replicates for CBP\_T30 (A) and P300\_T30 (B) against the average number of tags per region for those samples. Top panels are on the linear scale. Lower panels are on the square root scale. Left panels show the entire range of tags; right panels zoom in on the majority of regions with lower number of tags. The plot shows that the variance is much more homogeneously distributed on the square root scale.

Additional Figure 5.4



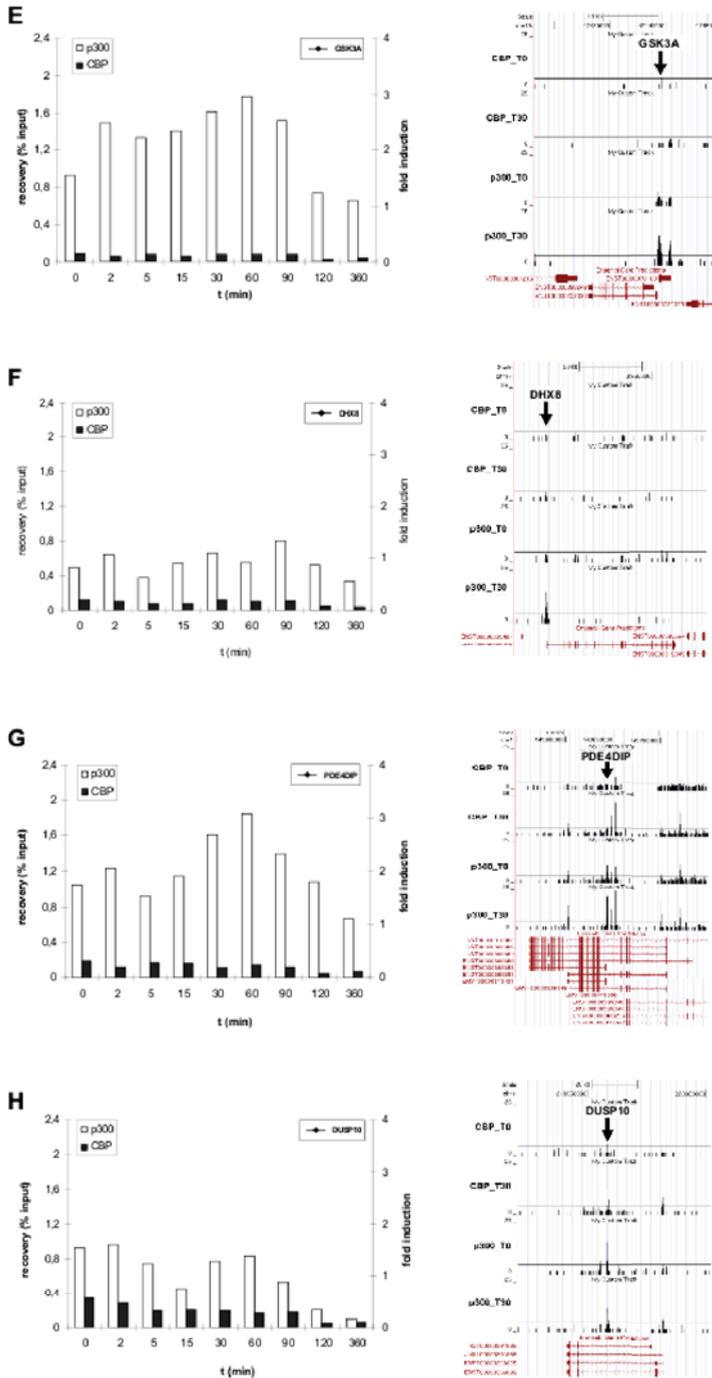
Correlation between ChIP-seq and ChIP-qPCR data for CBP (left) and p300 (right). For all genomic regions validated by qPCR (*CDK5*, *CDK6*, *CTGF*, *DHX8*, *DUSP10*, *FRA1*, *GSK3A*, *MKP1*, *PDE4DIP*, *SERPINE1* (TSS), *SERPINE1* (2kb upstream), *TAF15*, *ZNF608*, and *ZNF688*), we plotted the T30/T0 ratio obtained from the ChIP-Seq experiments (x-axis) against the T30/T0 ratio obtained from the ChIP-qPCR experiments (y-axis). Since deltaCt values plotted for the qPCR experiments reflect 2log differences in binding, also the ratio of the number of sequences from the ChIP-Seq experiments were 2log transformed. We counted the number of sequences aligned to each bp in the region spanned by the PCR primers +/- the average fragment length of 500 nucleotides. This was done because only the starts of fragments were sequenced and aligned, and these may fall outside of the PCR region, despite the presence of the PCR region in the fragment. We plotted delta Ct (y-axis) to provide a positive correlation since lower Ct values represent more binding, and higher Ct values less binding. The Ct values for time point T0 and T30 were obtained after fitting of a second-order polynomial through the Ct values of the time course from 0 to 90 minutes to improve the precision. The Pearson correlation coefficient and the p-value representing the significance of the correlation are given.

## Additional Figure 5.5A-D



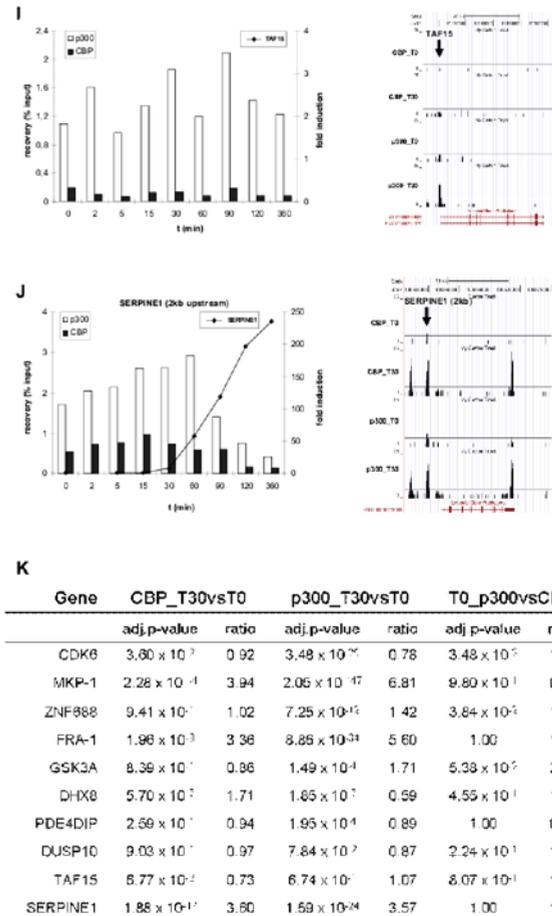
continued on next page

## Additional Figure 5.5E-H



continued on next page

## Additional Figure 5.5I-K



(continued from previous page): ChIP-analysis for time course experiment (0, 2, 5, 15, 30, 60, 90, 120, and 360 minutes after stimulation of growth arrested T98G cells with serum and TPA). Shown are graphs for qPCR results (x-axis: time in minutes; left y-axis: ChIP recovery in percentages of the input; right y-axis: fold induction

for the RT-qPCR with reference to the untreated samples ( $t=0$  minutes)) and screen-shots from custom tracks of the UCSC genome browser for the ChIP-seq results (T0 and T30 only) for *CDK6* (A), *MKP-1* (B), *ZNF688* (C), *FRA-1* (D), *GSK3A* (E), *DHX8* (F), *PDE4DIP* (G), *DUSP10* (H), *TAF15* (I), and *SERPINE1* (J; 2 kb upstream of the TSS). Arrows indicate the position of the PCR-amplicon.

Also indicated for these genes is the adj. p-value and the ratio difference between time points (T30 versus T0) and coactivators (p300 versus CBP) of the total number of sequences from all ChIP-seq data (K). (White bars: p300 ChIP; black bars: CBP

ChIP; line: fold induction of RT-qPCR; arrows in the screen-shots indicate the position of the PCR-amplicon. Also indicated for these genes is the adjusted p-value and ratio difference between time-points (T30 versus T0) and coactivators (p300 versus CBP) of the total number of reads from all ChIP-seq data (K).

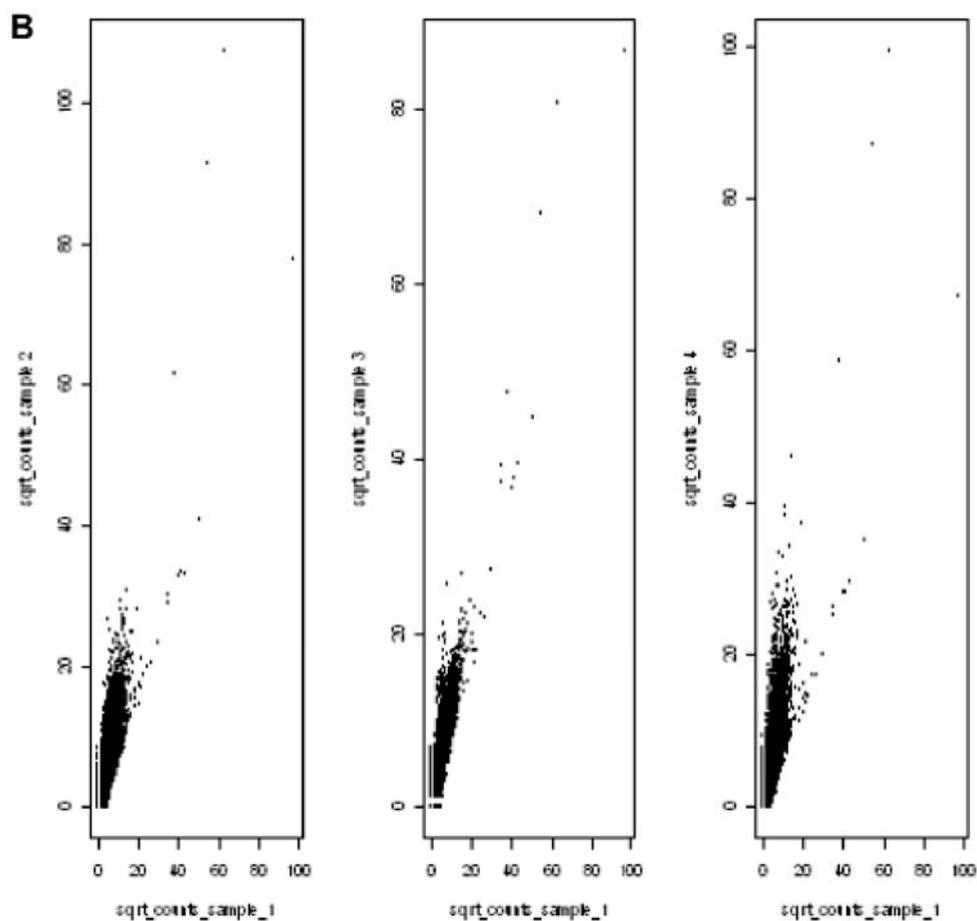
Additional Table 5.1

<b>A</b>		<b>B</b>	
primer:	sequence:	primer:	sequence:
ChIP CDK5 Fw	5'-TGCATTCTGGAACGCGTAGTC-3'	RT CDK5 up	5'-GCGACAAGAAGCTGACTTTGG-3'
ChIP CDK5 Rv	5'-TTATGTGCTCCACGCCCTCT-3'	RT CDK5 dn	5'-CGAGGTCACCAATTGCAACTGT-3'
ChIP CDK6 Fw	5'-CCATGCTTTTCCCAGATGGA-3'	RT CDK6 up	5'-CTGGCATTCAATCGTTGGC-3'
ChIP CDK6 Rv	5'-GAGTAAGAAGCTGCTGTCTCCCA-3'	RT CDK6 dn	5'-TACACCGATGGAAGAACTGGG-3'
ChIP CTGF Fw	5'-TGAGTTGATGAGGCAGGAAGG-3'	RT CTGF up	5'-TACCAATGACAACGCCCTCTG-3'
ChIP CTGF Rv	5'-CACAAACAGGGACATTCCTCG-3'	RT CTGF dn	5'-ACGGATGCACCTTTTTGCC-3'
ChIP DHX8 Fw	5'-GCGGAAGAAGCTGCAAACTC-3'	RT FRA-1 up	5'-CTCTTCTGTGATCCACCCAA-3'
ChIP DHX8 Rv	5'-CCCAAGTGATTGTCCAGCTCA-3'	RT FRA-1 dn	5'-TGGGTAAGGTGGCACCTTCTG-3'
ChIP DUSP10 Fw	5'-TCCCTCCTTTAATCCGCTC-3'	RT MKP-1 up	5'-CTTTGCTGTCTCGACCAATG-3'
ChIP DUSP10 Rv	5'-GGCTCCACCTTTTTGCTCA-3'	RT MKP-1 dn	5'-CTGGCCAAATGTTCTGCCT-3'
ChIP FRA-1 Fw	5'-AAGGCCAGTGGAAAGACCTCA-3'	RT SERPINE1 up	5'-TGGCCCATGAAAAGGACTGT-3'
ChIP FRA-1 Rv	5'-CCGTTTCTGCTCCACAAAA-3'	RT SERPINE1 dn	5'-AGAGAACCTGGGAATGACCGA-3'
ChIP GSK3A Fw	5'-TCAGGTCTGGACTATTCCCA-3'	RT ZNF608 up	5'-CGTGGAGGAACAAAACGTACG-3'
ChIP GSK3A Rv	5'-GTTGACGTATCCTCCCAT-3'	RT ZNF608 dn	5'-GGTACTCACAAAACCTGGGA-3'
ChIP MKP-1 FW	5'-CTTTGCTGTCTCGACCAATG-3'	RT ZNF688 up	5'-TTGCGTATTGCGGGTACAGAG-3'
ChIP MKP-1 RV	5'-CTGCGCAAATGTTCTGCCT-3'	RT ZNF688 dn	5'-AAGAGATGAGGGCTGTTGG-3'
ChIP PDE4DIP Fw	5'-CTTAGCAGATGAAAAGCGCTG-3'		
ChIP PDE4DIP Rv	5'-AAATCCGGTTGCACACCTG-3'		
ChIP SERPINE1 Fw	5'-AAAAGCAGGCAACGTGAGCT-3'		
ChIP SERPINE1 Rv	5'-TGACCCAAAAAGCCTAGGACC-3'		
ChIP SERPINE 1 (2kb) Fw	5'-TCAGTGTGGTTGCCCTTGGTA-3'		
ChIP SERPINE (2kb) 1 Rv	5'-GGAAGCGTCGGATGTTTGT-3'		
ChIP TAF15 Fw	5'-CCTCTTTCGTTTCTAACC GC-3'		
ChIP TAF15 Rv	5'-AGTACCAATGCCACGATCACG-3'		
ChIP ZNF608 Fw	5'-TGACTAGCACAGCCGCACCTT-3'		
ChIP ZNF608 Rv	5'-TGTGCTATTTTCCCTTCG-3'		
ChIP ZNF688 Fw	5'-TTGCGTATTGCGGGTACAGAG-3'		
ChIP ZNF688 Rv	5'-GAAAGATGTAAGGGCCGAAG-3'		

Sequence of primers used for qPCR analysis of the ChIPs (A) and for RT-qPCR (B).

Additional Table 5.2

A) Pearson						
Correlation	CBP T0	CBP T30 <sup>1</sup>	CBP T30 <sup>2</sup>	p300 T0	p300 T30 <sup>1</sup>	p300 T30 <sup>2</sup>
CBP T0	1	0.6698245	0.7215659	0.7663312	0.660192	0.722461
CBP T30 <sup>1</sup>		1	0.7654689	0.740022	0.869117	0.85208
CBP T30 <sup>2</sup>			1	0.7202088	0.773906	0.805657
p300 T0				1	0.753338	0.808835
p300 T30 <sup>1</sup>					1	0.869657
p300 T30 <sup>2</sup>						1



Pearson correlation between the different samples (A) and scatter plots (B).

**Additional Table 5.3**

Fisher Test	adj.p-value <0.01	adj.p-value <0.001	difference $\geq 5x$
CBP T30vsT0	1231	765	11
p300 T30vsT0	3730	2620	44
T0 p300 vs CBP	256	120	22
T30 p300 vs CBP	2502	1611	42

Number of genes significantly different between the different samples for adjuvant p-values  $<0.01$  and  $<0.001$ , as determined with Fisher Exact Test, and the number of genes with p-values  $<0.001$ , with a difference of at least 5 times between the samples.

**Additional Files 5.1 to 5.6**

Wiggle files for CBP T0, CBP T30<sup>1</sup>, CBP T30<sup>2</sup>, p300 T0, p300 T30<sup>1</sup>, and p300 T30<sup>2</sup>, created as described in Materials and Methods and viewable in the UCSC genome browser ((103) (single reads were removed) are available at [nar.oxfordjournals.org/cgi/content/full/gkq184/DC1](http://nar.oxfordjournals.org/cgi/content/full/gkq184/DC1).

**Additional File 5.7**

Genes annotated for p300 and CBP ChIP-seq in quiescent (T0) and in growth factor stimulated (T30) cells. Sheet 1 shows the 1,6103 genes that were identified (as explained in Sheet 2, the Ensemble gene ID, the sum of the number of tags sequenced and the number after normalization for gene length, the ratio between the different samples, the p-value, and the adjusted p-value for the ratios are shown). Available at [http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009\\_R2\\_supplemental\\_file\\_7.xls](http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009_R2_supplemental_file_7.xls).

**Additional File 5.8**

Functional classification performed with DAVID 2008 Functional Annotation (<http://david.abcc.ncifcrf.gov/home.jsp>) of 250 genes that differed most significantly at T30 versus T0 for binding by CBP (worksheet CBP\_T30vsT0), by p300 (worksheet p300.T30vsT0), and of 250 genes where CBP binding is significantly higher than p300 binding (worksheet T30.CBPhigherthanp300) or where p300 binding is significantly higher than CBP binding (worksheet T30.p300higherthanCBP). Available at [http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009\\_R2\\_supplemental\\_file\\_8.xls](http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009_R2_supplemental_file_8.xls).

**Additional File 5.9**

Full list of Transcription Factor binding sites that were found to be enriched for T30 versus T0, and for p300 versus CBP, as analyzed with CORE\_TF (76). Available at [http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009\\_R2\\_supplemental\\_file\\_9.xls](http://nar.oxfordjournals.org/content/vol0/issue2010/images/data/gkq184/DC1/NAR-02256-X-2009_R2_supplemental_file_9.xls).

