

FactorPath™ ChIP-Seq and ChIP-on-chip

Comprehensive discovery of transcription factor binding sites across whole genomes

FactorPath™ is a powerful assay for discovering, verifying and quantifying the binding of transcription factors and co-factors across the genome. FactorPath ChIP-Seq and ChIP-on-chip are based on chromatin immunoprecipitation (ChIP) using procedures optimized by Active Motif that detect transcription factor binding with maximum sensitivity and minimal background. Active Motif's FactorPath assays include initial qualification of the antibody (or use of a qualified antibody), chromatin preparation and quantitation, as well as qPCR quality control steps after ChIP and after amplification. For ChIP-Seq, the amplified libraries are sequenced, generating millions of 36 bp tags that can be uniquely mapped to the genome, thus revealing enrichment at sites of transcription factor binding. For ChIP-on-chip, the amplified ChIP DNA is labeled and hybridized to one or more Affymetrix GeneChip® arrays. Arrays are typically Human or Mouse Tiling 2.0R Arrays or Promoter 1.0R Arrays, each of which contains non-repetitive tiling probes of 25 nucleotides in length with 10-base pair gaps. Active Motif's internal research and development has shown that whole-genome data generated from these two platforms are largely overlapping. Although the data provided in this document was generated using ChIP-on-chip, ChIP-Seq data would be presented in the same manner and be expected to give similar results.

Discovery of Binding Sites for SRC-3

FactorPath ChIP-Seq and ChIP-on-chip enable investigators to identify transcription factor binding sites comprehensively across the genome, as illustrated by ChIP studies targeting SRC-3, a co-activator of several nuclear receptors including estrogen receptor (ER). In a representative assay, MCF-7 breast cancer cells were treated with estradiol for 1 hour and the resulting chromatin was immunoprecipitated using an antibody specific for SRC-3. Levels of SRC-3 binding were first validated with FactorPath qPCR (quantitative PCR) using primers that amplify known ER binding sites located 5' of the TFF1 and GREB1 genes. As shown in Figure 1, a significantly higher amount (~100-fold) of DNA containing SRC-3 binding sites was observed in the SRC-3 ChIP DNA vs. the control IgG ChIP DNA. Following amplification of the ChIP DNA in preparation for array hybridization, a similar level of enrichment was maintained (Figure 1).

The amplified DNAs were labeled using the GeneChip® WT Double-Stranded DNA Terminal Labeling Kit and hybridized to human tiling arrays, then washed and scanned. Raw data from the scans was analyzed using Affymetrix® Tiling Analysis Software (TAS) and the results were visualized using the Integrated Genome Browser (IGB). Active Motif's proprietary ChIP Analysis Software was used to compile the data into tables that contain information about binding site locations, binding intensities and associated gene and genomic annotations.

SRC-3 Binding Sites are Found Throughout the Genome

Visualization of the hybridized arrays in IGB showed numerous peaks representing putative SRC-3 sites, as illustrated for duplicate arrays in Figure 2. In this example, three peaks inside and between the TFF1 and TMPRSS3 genes are clearly distinguished. Further analysis using Active Motif's software resulted in a comprehensive list of genomic locations. An excerpt of this list

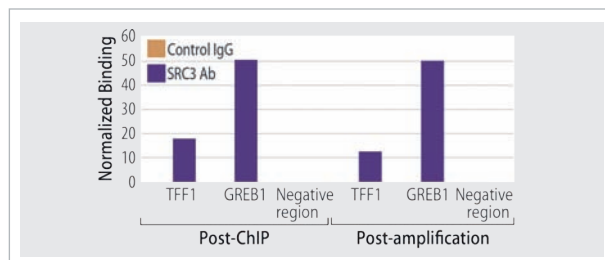


Figure 1. FactorPath qPCR of SRC-3 and control IgG ChIP DNA. PCR primers that amplify known binding sites, TFF1 and GREB1, show enrichment in estradiol-treated MCF-7 cells both after ChIP and after amplification of ChIP DNA.

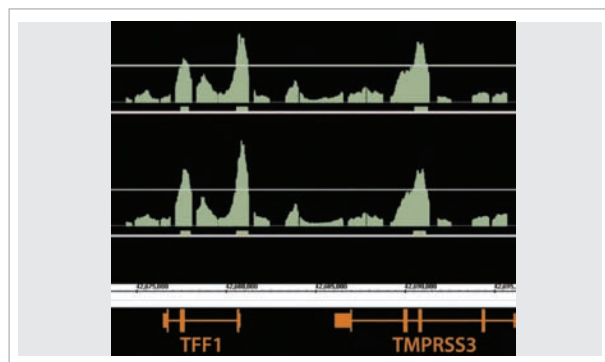


Figure 2. Following Array hybridization data is visualized using the Integrated Genome Browser. Zoomed-in regions show SRC-binding sites near the TFF1 transcription start site, inside the TFF1 gene, and inside the TMPRSS3 gene. Genes are transcribed from right to left.

is shown in Table 1. Review of the putative SRC-3 binding sites represented by peaks on the arrays showed that essentially all previously known SRC-3 binding sites were identified, as well as a large number of novel sites.

Overall results obtained from hybridization to Arrays B, C, and D of the GeneChip® Human Tiling Array Set are listed in Tables 2 and 3. Interestingly, a significant number of sites (almost 30%) were in unannotated regions of the genome. It is also striking that many of the SRC-3 binding sites were located within and downstream from genes. Only 4.3% of the sites were located within 5 kb upstream of known genes.

Validation of Tiling Array Peaks using qPCR

A high percentage of the peaks were verified as true SRC-3 binding sites using Active Motif's FactorPath qPCR assay. As illustrated in Figure 3 and Table 4, 100% of large-medium sized peaks were confirmed as true SRC-3 binding sites. More specifically, nearly all large peaks confirmed as strong binding sites ($\geq 5X$ over background), while most of the medium peaks confirmed as either strong or moderate binding sites for SRC-3 (Table 4).

Identification of Binding Sites for p53

FactorPath ChIP-Seq and ChIP-on-chip can be used to identify binding sites for any factor as long as an antibody that works in ChIP can be identified. To demonstrate the usefulness of this assay with a different factor, chromatin from untreated, proliferating MCF-7 cells was immunoprecipitated using an antibody against p53. Following hybridization to an Affymetrix GeneChip® Human Tiling 2.0 Array, a number of moderate-to-large peaks were observed in genomic areas corresponding to known p53 binding sites as well as novel sites (see Figure 4 for a representative novel site). These peaks were confirmed as true p53 binding sites using FactorPath qPCR.

Genome-wide Discovery of Factor Binding Sites with High Sensitivity

FactorPath ChIP-Seq and ChIP-on-chip are comprehensive services that enable investigators to identify transcription factor binding sites across whole genomes. These services include chromatin preparation and quantitation, ChIP, qPCR of positive control regions, amplification, qPCR of amplified ChIP DNA to ensure that enriched regions are still represented in the amplified sample, Next-Gen sequencing or Affymetrix array hybridization, and data analysis. Active Motif's ability to scan across the entire genome is important because increasing evidence, including that presented in this Application Note, shows that factor binding sites are located largely in regions outside of classic promoters.

Table 1. Examples of SRC-3-specific array peaks on chromosome 19.

Interval start	Interval end	Gene	Distance to gene start	Distance to gene end	Location
1132744	1133240	KIAA0963	-7981	-74599	Upstream
2486540	2486829	GNG7	110449	-24611	In gene
14815918	14816464	OR7A10	-2775	-3704	Upstream
15460283	15460958	PGLYRP2	-9646	-20495	Upstream
15947884	15948282	LOC126537	22555	3333	Downstream
16434947	16436450	CALR3	31553	14438	Downstream
34854025	34854662	PLEKHF1	6222	-4179	In gene
35865774	35865953	None*			
36463389	36464254	ZNF537	-2239	-6561	Upstream
37545573	37546208	ZNF507	111726	-20256	In gene

* no known gene within 20 kb

Table 2. SRC-3-specific peaks identified on three tiling arrays.

Tiling Array Type	Number of peaks
B (chromosomes 2, 9, 19)	371
C (chromosomes 3, 21, 22, X)	307
D (chromosomes 4, 15, 18, 20)	474
Extrapolated to whole human genome	2688

Table 3. Peak locations relative to genes.

Peaks on Array B	Number	Percent
Upstream (-5 kb)	16	4.3%
Upstream (-5 to -20 kb)	49	13.2%
In gene	174	46.9%
Downstream (+20 kb)	26	7.0%
No gene	106	28.6%
TOTAL	371	100.0%

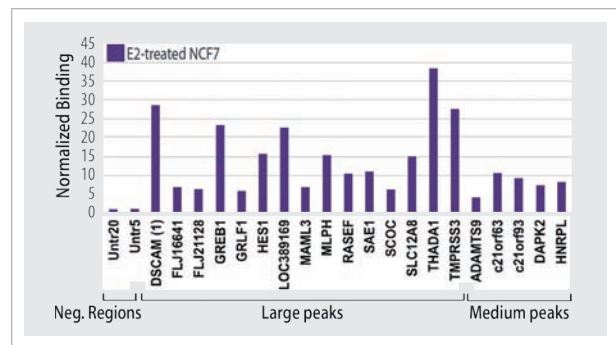


Figure 3. Validation by FactorPath qPCR that peaks observed on the tiling arrays represent SRC-3 binding sites. Untr5 and Untr20 are negative control regions that show the level of background signal.

Overview of Active Motif Services

Active Motif services have wide utility in the discovery and development of targeted drugs, novel diagnostic biomarkers, and the identification of causal disease and pharmacological pathways.

We offer a complete suite of epigenetic services for understanding gene regulation and genetic pathways:

- **FactorPath™** for discovery of transcription factor binding sites
- **TranscriptionPath™** for discovery of genomic DNA sequences undergoing active transcription in cells
- **MethylPath™** and **HistonePath™** for discovery and analysis of epigenetic events such as DNA methylation and histone modification
- **Bisulfite Sequencing** for determining the methylation status of cytosines at specific CpGs
- **Antibody Qualification** to validate those antibodies that are suitable for use in ChIP, ChIP-Seq and ChIP-on-chip

All of our services include expert consultation on experimental design, quality controls at multiple steps and comprehensive data analysis. Please contact us with any questions you might have about initiating a project in your cell system.

Table 4. Validation of SRC-3 binding by FactorPath qPCR.

TAS p-value Peaks	# sites (%)	Fold over background				
		5X tested	3-5X (Strong)	2-3X (Moderate)	1-2X (Weak)	(None)
Large	> 99.9	20	19	1	0	0
Medium	99.8-99.9	27	20	5	2	0
Small	< 99.8	14	0	0	7	7

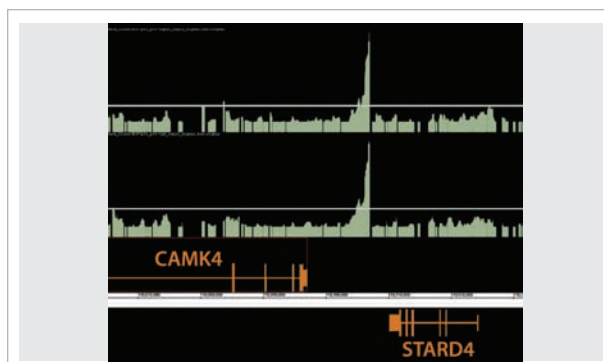


Figure 4. Array peaks on IGB indicating a p53 binding site between the 3' ends of genes CAMK4 and STARD4 (transcribed left to right and right to left, respectively). The two green rows represent duplicate hybridized arrays.