

Illumina Sequencing Services with Genpathway ChIP Assay and Data Analysis Services

Illumina and Genpathway services combine to offer a complete ChIP sequencing solution for unbiased genome-wide discovery of transcription factor binding sites in nearly any genome.

INTRODUCTION

The combination of Genpathway FactorPath ChIP (chromatin immunoprecipitation) and Illumina high-throughput sequencing enables genome-wide discovery of transcription factor binding sites. Used together, the services provided by the two companies start with cells and provide fully analyzed results to researchers. Customers can use the two services to access Illumina's high-throughput sequencing technology and Genpathway's ChIP and genome-wide data analysis expertise to discover binding sites of interest.

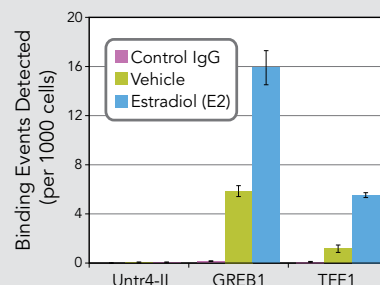
Here, we describe the analysis of SRC-3 as an example factor, beginning with sample preparation and ending with comprehensive data analysis.

CHIP SEQUENCING FOR GENOME-WIDE COVERAGE

The identification and quantitation of transcription factor and co-regulator binding is highly valuable for understanding gene regulation and pathways. This information is relevant to a wide range of applications from basic disease research to compound mechanism of action and pharmacogenomics. The powerful combination of Genpathway's FactorPath ChIP assays and proprietary software and data analysis with Illumina DNA Sequencing Services provides investigators an unsurpassed level of information on transcription factor binding and gene regulation¹⁻³.

Genpathway's FactorPath assay is a high-quality ChIP-based assay for essentially any DNA-associated factor that uses optimized and proprietary procedures to provide maximal sensitivity and minimal background. When used in combination with Illumina Sequencing, significantly greater coverage of the genome is achieved due to direct identification of millions of non-repetitive ChIP DNA fragments. Essentially any genome (with at least some genomic sequence available) can be studied

FIGURE 1: FACTORPATH QUERY VALIDATION OF CHIP



FactorPath Query Assay of three ChIP DNA samples shows significant SRC-3 binding in MCF7 cells ChIPed with the SRC-3 antibody (vehicle and E2) and increased binding with E2 treatment. Untr4-II is a negative control region on chromosome 4.

with FactorPath ChIP and Illumina Sequencing Services.

Illumina's massively parallel sequencing technology results in millions of ChIP fragment sequences aligned to the genome. Genpathway's software analyzes these sequence alignments to provide comprehensive results on the locations, metrics, and genomic annotations for fragment clusters over an assigned threshold. As a result of the inherent depth of reads, each assay identifies hundreds to thousands of binding sites across the entire genome and quantifies differences in binding levels between sites and samples. Data generated with this assay include both known and novel factor binding sites located near or within genes or in unannotated regions of the genome.

CO-ACTIVATOR SRC-3 BINDING SITES IN MCF7 CELLS

To demonstrate the results of the complete FactorPath ChIP and Illumina Sequencing Services, binding sites for the nuclear receptor co-activator SRC-3 were detected across the genome of MCF7 breast cancer cells. Further,

TABLE 1: SIZE DISTRIBUTION OF INTERVALS

	VEHICLE	E2	INPUT	IgG
Fragments Analyzed	1,583,097	1,790,922	1,681,562	619,834
Intervals	2,221	2,966	388	135
PEAK FRAGMENT DENSITY				
> 50	86	210	4	7
41–50	55	134	1	8
31–40	146	254	19	4
26–30	134	234	26	8
21–25	300	378	45	10
15–20	835	1,025	135	45
11–14	665	731	158	53

differential binding levels between estradiol (E2)-treated and vehicle-treated cells were determined.

Chromatin was isolated and immunoprecipitated (ChIPed) using an antibody against SRC-3. Controls included E2-treated chromatin ChIPed with a control IgG and input DNA (not ChIPed). To confirm that appropriate SRC-3-bound DNA was obtained with high quality (good signal and low background binding), Genpathway's qPCR-based FactorPath Query assay was performed. As expected, known estrogen receptor (ER)/SRC-3 binding sites located 5' of the TFF1 and GREB1 genes exhibited significant levels of SRC-3 binding in the appropriate ChIP samples (Figure 1). SRC-3 binding was significantly induced in the E2-treated cells relative to vehicle treatment. The ChIP and input samples were then amplified using standardized procedures, and another FactorPath Query Assay was performed before sequencing with the Illumina Genome Analyzer.

GENOME-WIDE BINDING SITE DETECTION

Sequencing each of the four DNA samples resulted in 2.6–3.9 million fragments uniquely mapped to the genome. Duplicate fragments were excluded from further analysis (Table 1). The complete data package included with this service involves many analysis steps and an overview of SRC-3 binding analysis is described here.

For analysis by Genpathway's software, the genome is divided into 32-nucleotide bins and the fragments in each bin are counted. Regions containing bins with fragment counts over the significance threshold (11 for this assay) are referred to as intervals. In this assay, bins within intervals had 11 to 243 fragments each. The greatest number of intervals was seen in the SRC-3 ChIP sample from E2-treated cells (Table 1). Intervals in the E2-treated cells contained more fragments and higher peak fragment densities (for each interval, the number of fragments in the peak bin), indicating stronger binding.

Interval locations were analyzed in relation to current gene annotation (Table 2). Consistent with previously performed SRC-3 FactorPath assays, relatively few (13–15%) intervals were located in typical promoter regions (within 7.5kb upstream or 2.5kb downstream of the transcription start site, TSS) and even fewer (4–5%) were within 500bp of a gene TSS.

By assigning fragments to bins, additional metrics can be calculated, such as average fragment count per bin across the intervals (avg value) and fragment number in the peak bin (peak value). Differential binding strengths for the same location between samples are indicated by differences in the various metrics. These metrics are used to compare across samples using a parameter called active regions. Active regions are any genomic regions that contain intervals. By definition, intervals that overlap (in the same or different samples) are contained within the same active region. Selected columns from the SRC-3 active regions spreadsheet are shown in Table 3.

TABLE 2: LOCATION OF INTERVALS RELATIVE TO GENES

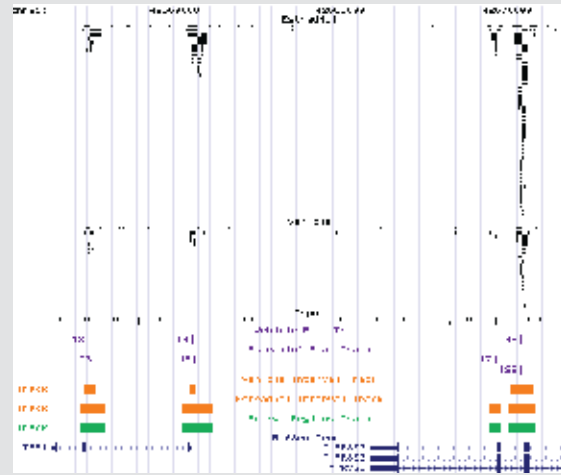
INTERVAL LOCATION	VEHICLE	VEHICLE %	E2	E2 %
Within 10kb of a gene	1,232	55.5%	1,691	57.0%
Not within 10kb of a gene	989	44.5%	1,275	43.0%
In promoters (-7.5/+2.5kb of TSS)	323	14.5%	383	12.9%
Within 500bp of TSS	118	5.3%	116	3.9%
Total	2,221	100%	2,966	100%

The TFF1-TMPRSS3 gene region on chromosome 21 is a typical interval identified by this assay (Figure 2). The genome browser view of the region shows the fragment clusters, the intervals for the vehicle- and E2-treated cells, and the active regions. In this region, three relatively large clusters can be seen at locations 3kb downstream from the TFF1 TSS, within the TFF1 promoter, and inside the TMPRSS3 gene (both genes are transcribed from right to left). Two of the three sites exhibited induction of SRC-3 binding after E2 treatment. In addition, a smaller induced interval is present 3' of the major interval in the TMPRSS3 gene. This binding site was only suggested previously from ChIP-on-chip data as a shoulder of the major peak. These results demonstrate the higher resolution attainable by ChIP Sequencing due to the massive quantity of ChIP fragments sequenced directly with Illumina Sequencing technology.

INTERVAL VALIDATION

Representative intervals identified by FactorPath ChIP and Illumina Sequencing Services were validated as being true SRC-3 binding sites using Genpathway's FactorPath Query assay. Of 83 genomic regions with peak fragment densities of 0 to more than 40 fragments (62 intervals had peak fragment densities > 11) tested with this highly

FIGURE 2: BINDING EVENT DISPLAY



Locations and relative binding levels for SRC-3 in and around the TFF1 and TMPRSS3 genes in MCF7 cells treated with E2 (top) or vehicle (second row). Sequenced and aligned DNA fragments for three of the four samples are shown at the top (black tick marks). Intervals and active regions determined by Genpathway's analysis software are shown (orange and green bars, respectively). SRC-3 binding is clearly induced by E2 treatment at three of the four sites. Input DNA showed background levels only (third row).

TABLE 3: EXAMPLES OF ACTIVE REGIONS ON CHROMOSOME 1 SHOWING SELECTED COLUMNS IN DATA ANALYSIS

CHR	START	LENGTH	VEHICLE AVG VALUE	E2 AVG VALUE	VEHICLE PEAK VALUE	E2 PEAK VALUE	GENE LIST	DISTANCE TO START	POSITION*	VEHICLE INTERVAL PRESENT	E2 INTERVAL PRESENT
1	200,348,388	665	6.5	38.8	10	67	GPR37L1	-9,932	up	0	1
1	200,367,908	793	5.5	18.6	9	28	GPR37L1, ARL8A	9652, 12185	down, down	0	1
1	200,570,948	633	3.5	25.6	5	37	UBE2T	6,443	in gene	0	1
1	200,985,092	345	11.2	7.7	14	9	JARID1B	58,908	in gene	1	0
1	201,324,772	953	6.5	38.0	13	61	MYOG, ADORA1	-3459, -1157	up, up	0	1
1	201,552,228	633	17.6	7.2	23	10	BTG2	11,257	down	1	0
1	202,555,140	345	13.2	4.0	15	5	PLEKHA6	40,355	in gene	1	0
1	202,616,100	633	16.4	24.1	22	37				1	1
1	202,652,932	377	12.2	4.3	14	7	PPP1R15B, PIK3C2B	-5578, 72977	up, down	1	0
1	205,143,428	697	8.9	45.5	16	76	IL24, FAIM3	6364, 18190	in gene, down	1	1

*up = upstream, down = downstream

TABLE 4: VALIDATION OF SRC-3 BINDING WITH FACTORPATH QUERY

PEAK FRAGMENT DENSITY	INTERVALS EXAMINED	FOLD BINDING AT SPECIFIC SITE VS. NEGATIVE REGION			
		> 5x	3-5x	2-3x	1-2x
> 40	17	17	0	0	0
31-40	15	15	0	0	0
21-30	12	10	0	2	0
15-20	10	9	1	0	0
11-14	8	6	2	0	0
5-10	11	5	1	5	0
0-4	10	1	0	2	7
TOTAL	83	63	4	9	7

quantitative assay, 60 intervals validated as strong (> 5x over background with FactorPath Query) or moderate (3-5x over background) SRC-3 binding sites (Table 4). In fact, 51 of the 54 intervals with peak fragment densities greater than 15 were found to be strong binding sites. At the lower end, 6 of the 11 intervals with low peak fragment densities (5-10) exhibited at least moderate binding over background in the FactorPath Query assay.

QUANTITATIVE BINDING LEVEL ANALYSIS

Because of the large number of ChIP fragments sequenced by the Illumina Genome Analyzer, quantitative information about factor binding levels can be obtained. Examples of intervals with differential numbers of fragments that were tested by FactorPath Query are illustrated in Table 5. All sites tested that showed differential numbers of fragments (at least 2-fold difference) were confirmed as differential in SRC-3 binding by FactorPath Query.

GENOME-WIDE FACTOR BINDING ANALYSIS SERVICE

The Genpathway FactorPath ChIP and Illumina Sequencing Services provide a powerful genome-wide approach for identifying and quantifying binding sites for nearly any transcription factor in any organism. The combined services include cell or tissue processing, antibody selection and qualification, ChIP using Genpathway's optimized procedures for greater sensitivity and specificity, ChIP DNA amplification, QC steps throughout the assay, high-throughput parallel sequencing using the Illumina Genome Analyzer, and complete data analysis using Genpathway's proprietary software analysis tools.

TABLE 5: VALIDATION OF FACTORPATH CHIP SEQUENCING BINDING QUANTIFICATION

GENE	BINDING RATIO BY CHIP-SEQ	BINDING RATIO BY FPQ	E2 EFFECT
TFF1 (prom)	3.4	2.76	Induction
GREB1	3.6	6.23	Induction
TMPRSS3	2.9	4.31	Induction
ARL3	6.9	3.60	Induction
NAV2	2.8	2.77	Induction
TFF1 (gene)	1.3	1.09	None
THADA (1)	1.1	1.00	None
THADA (2)	0.7	0.92	None
F5	1.4	1.20	None

Uniquely mapped fragments located near repetitive sequences and repeat masked on tiling arrays are included in the ChIP sequencing analysis. This results in increased genomic coverage and more comprehensive results. Researchers will enjoy other advantages including competitive pricing, Genpathway's standard QC steps, efficient turnaround of results, and technical support for investigators at every level. Combining Genpathway's years of ChIP expertise and data analysis with Illumina's high-throughput sequencing provides researchers a faster path to results and publication by not having to set up and troubleshoot ChIP in their own lab.

SERVICE PROJECT INITIATION

For more information or to discuss initiating a project, please contact Genpathway technical support at 1.858.457.3250 or visit www.genpathway.com.

REFERENCES

- (1) Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129: 823-837.
- (2) Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* 316: 1497-1502.
- (3) Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, et al. (2007) Genome-Wide Profiles of Stat1 DNA Association Using Chromatin Immunoprecipitation and Massively Parallel Sequencing. *Nature Methods* 4: 651-657.

Illumina, Inc.

Customer Solutions

9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY

© 2008 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadPress, VeraCode, IntelliHyb, iSelect, CSPro, iScan, and GenomeStudio are registered trademarks or trademarks of Illumina. Genpathway, FactorPath, and The CHIP Assay Experts are registered trademarks or trademarks of Genpathway, Inc. All other brands and names contained herein are the property of their respective owners.

Pub. No. 770-2008-006 Current as of 22 April 2008

