Published in final edited form as:

J Bioinform Comput Biol. 2011 April ; 9(2): 269–282.

COMPARING MULTIPLE ChIP-SEQUENCING EXPERIMENTS

HATICE GULCIN OZER^{*,†,§}, YI-WEN HUANG^{‡,¶}, JIEJUN WU^{*,‡,||}, JEFFREY D. PARVIN^{*,†,**},
TIM HUI-MING HUANG^{‡,††}, and KUN HUANG^{*,†,‡‡}

^{*}Department of Biomedical Informatics, The Ohio State University, Columbus, OH, 43210, USA

[†]The Ohio State University CCC Biomedical Informatics Shared Resource, The Ohio State University, Columbus, OH, 43210, USA

[‡]The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH, 43210, USA

Abstract

New high-throughput sequencing technologies can generate millions of short sequences in a single experiment. As the size of the data increases, comparison of multiple experiments on different cell lines under different experimental conditions becomes a big challenge. In this paper, we investigate ways to compare multiple ChIP-sequencing experiments. We specifically studied epigenetic regulation of breast cancer and the effect of estrogen using 50 ChIP-sequencing data from Illumina Genome Analyzer II. First, we evaluate the correlation among different experiments focusing on the total number of reads in transcribed and promoter regions of the genome. Then, we adopt the method that is used to identify the most stable genes in RT-PCR experiments to understand background signal across all of the experiments and to identify the most variable transcribed and promoter regions of the genome. We observed that the most variable genes for transcribed regions and promoter regions are very distinct. Gene ontology and function enrichment analysis on these most variable genes demonstrate the biological relevance of the results. In this study, we present a method that can effectively select differential regions of the genome based on protein-binding profiles over multiple experiments using real data points without any normalization among the samples.

Keywords

Next generation sequencing; ChIP sequencing; comparison; normalization

1. Introduction

During the past few years, with the rapid advancement in the next generation sequencing (NGS) technology, and related techniques such as chromatin immunoprecipitation (ChIP),

[§]gulcin.ozero@osumc.edu

[¶]yi-wen.huang@osumc.edu

^{||}wu.403@osu.edu

^{**}jeffrey.parvin@osumc.edu

^{††}tim.huang@osumc.edu

^{‡‡}kun.huang@osumc.edu

researchers can investigate the protein–DNA binding relationship at high resolution using the ChIP-sequencing (ChIP-seq) method. In a recent study, we have generated a large collection of ChIP-seq data for different cell lines and samples as well as for different proteins. The goal of this study is to understand the epigenetic regulation of breast cancer and the effect of estrogen. So far we have generated 46 lanes of ChIP-seq data from the Illumina Genome Analyzer II (GAII) for five different proteins in eight different samples (including the same samples under different conditions) with a total of more than 11 giga bases. The proteins include RNA polymerase II (Pol II), estrogen receptor α (ER α), and epigenetic markers including H3K4-dimethylation (H3K4me2), H3K9-dimethylation (H3K9me2), and H3K27-trimethylation (H3K27me3).

The accumulation of this large amount of data brings up an important question. How can we explore the difference of protein-binding profiles among different cell types and multiple proteins? In another words, can we identify the genes with significantly different binding profiles for multiple proteins and cell types to differentiate different samples? To answer these questions, usually we need to develop effective comparison methods such as data normalization between the data. However, since the data were generated from samples with distinctively different phenotypes and different proteins, normalization is not feasible or even meaningful. Instead, we need to resort to other approaches. In this paper, we adopted a method previously used for gene expression analysis in microarray and quantitative real-time PCR (qPCR) experiments. This approach allowed us to identify genes that show differentiated protein-binding profiles across multiple samples.

An advantage of this method is that it circumvented the process of normalization among the samples and is robust to experimental variations. Our results paved the way to carry out a whole genome screening in the future work.

2. Methods

2.1. Dataset

ChIP-seq experiments are conducted on ten different samples with five different pull-down proteins. Table 1 summarizes the types of samples and cell lines and pull-down proteins. The samples include mammosphere from human breast cancer tissue, MCF10A (a non-tumorigenic human mammary epithelial cell line), MCF7 (a human breast cancer epithelial cell line), and OHT (a tamoxifen-resistant cell line derived from MCF7). In addition, we also have data for two hepatoma (liver cancer) cell lines PLC5 and Huh7.

The treatments on the samples include AKT gene knockout (in MCF10A) and application of E2 (also known as 17-Estradiol, to mammosphere and MCF7 and OHT cell lines). For each sample, ChIP-seq experiments focusing on different proteins have been carried out. The proteins include RNA polymerase II (Pol II), estrogen receptor α (ER α), and histone marks including H3K4-dimethylation (H3K4me2), H3K9-dimethylation (H3K9me2), and H3K27-trimethylation (H3K27me3).

2.2. Sequence alignment and annotation

We used Illumina's Eland mapping algorithm to map sequence reads to Human reference genome hg18. For every experiment we focused on transcribed regions and promoter regions. We extracted 20,998 transcribed region coordinates from RefGene database.¹ Then, we defined coordinates of promoter regions as ± 5 K bases of transcription start sites (TSS).

Based on these region coordinates, we calculated total number of sequence reads that are aligned to each region across all of the samples. Here we will refer to total number sequence reads in a defined region as *region count* (g). Using these region counts, we built two matrices 20,998 (regions) by 50 (experiments) for transcribed regions and promoter regions separately. Out of 20,998 regions, we selected the ones that have 50 or more region counts at least in one of the experiments. That resulted in 14,462 by 50 and 11,808 by 50 matrices of transcribed regions and promoter regions, respectively.

2.3. Comparison of multiple ChIP-Seq experiments

When the experiments are clustered based on these region counts, we observed that pull-down protein and the cell line drives the clustering in general. Moreover, region count profiles on transcribed regions and promoter regions end up with different characteristics. Region count profile is more cell-line specific for transcribed regions, while it is more pull-down-protein specific for promoter regions.

Figures 1 and 2 depict correlation and clustering of the experiments based on transcribed region counts and promoter region counts, respectively. This analysis demonstrates that region count profiles of the experiments with the same pull-down protein on different cell lines or the same cell lines with different treatment options are very similar. On the other hand, region count profiles of experiments with certain pull-down proteins are very different.

For example, correlation within MCF7 and OHT, MCF10 and mammosphere cell lines is quite strong regardless of the pull-down protein and the treatment in transcribed regions (Fig. 1), while correlation within pull-down proteins is stronger than the correlation within cell line and treatment in promoter regions (Fig. 1). Besides, correlations between experiments with H3K9 Me2 and H3K27 Me3 (Fig. 2) are very strong, while correlations between experiments H3K4 Me2 and H3K27 Me3 (Figs. 1 and 2), Pol II and H3K27 Me3 (Fig. 2), and Pol II and H3K9 Me2 (Figs. 1 and 2) are very poor.

2.4. Selecting features based on protein binding profiles over multiple samples

Since these experiments are done on different cell lines and with different pull-down proteins, background signal is not uniform across these experiments. We need a method to describe background variation across all experiments.

Vandesompele *et al.* (2002) introduced a gene stability measure to identify most stably expressed control genes in RT-PCR experiments. They defined gene stability measure M as the average pairwise variation between any particular gene and all other genes. Variation between two genes is calculated as standard deviation of log-transformed gene expression

ratios. A small M value means expression of this gene is quite stable across all samples, while a large M value means this gene varies a lot across all samples.²

We adopted this method to determine the least stable regions across all experiments as opposed to the original intent of the method to identify stable controls. When the regions were ranked by their M values, the small M values indicate the regions that do not vary in the samples or in response to experimental treatment in our 50 ChIP-seq experiments. This helped us to understand background variation and determine the most variable regions across all experiments.

Out of n regions in total, for every combination of two regions j and k and in experiment i , log2-transformed ratios of region counts g_{ij} and g_{ik} are calculated. Array G_{jk} of m elements consists of these ratios across m experiments [Eq. (1)]. Pairwise variation is calculated as the standard deviation of the G_{jk} elements [Eq. (2)]. The region-stability measure M_j for a region j is defined as the mean of all pairwise variations, S_{jk} [Eq. (3)].

$$G_{jk} = \left\{ \log_2 \left(\frac{g_{1j}}{g_{1k}} \right), \log_2 \left(\frac{g_{2j}}{g_{2k}} \right), \log_2 \left(\frac{g_{ij}}{g_{ik}} \right) \right\} = \left\{ \log_2 \left(\frac{g_{ij}}{g_{ik}} \right) \right\}_{i \rightarrow m}, \quad (1)$$

$$S_{jk} = \text{std}(G_{jk}), \quad (2)$$

$$M_j = \frac{\sum_{k=1}^n S_{jk}}{n-1}. \quad (3)$$

Out of 20,998 transcribed genes regions listed in RefGene database, we selected 14,462 transcribed regions and 11,808 promoter regions that have 50 or more sequence reads on them at least in one of the experiments. Then we calculated M values for each region for both sets. We implemented M value calculations in R statistical data analysis language. Calculations are done on the computing cluster that consists of 72 computing nodes with dual quad core AMD Opteron 2378 processors. We submitted ~100 regions per job and each job is completed in five to six hours.

3. Results

3.1. Calculation of M values

First, we investigate the range of M value and its implications. Figure 3 shows the histogram of M values for 14,462 transcribed regions and 11,808 promoter regions. M values range between 1 and 2.9, and it is less than 2 for 97% and 94% of the transcribed and promoter regions, respectively. This indicates that there is a background pattern that most of the regions follow. Small M value means counts for this region are quite stable across all of the experiments. On the other hand, large M value means counts for this region are quite different than the background pattern.

3.2. Gene ontology and function enrichment analysis

For gene ontology and function enrichment analysis we focused on top 200 transcribed regions and top 200 promoter regions as the most variable regions across all of the experiments (see supplementary Tables 1 and 2). Then, counts for these selected regions are used to cluster genes and experiments. The heatmap in Fig. 3 shows clustering of these top 200 most variable regions across 50 ChIP-seq experiments.

We performed functional analysis on the genes corresponding to the most variable 200 regions across 50 ChIP-seq experiments using Ingenuity Pathway Analysis (IPA).³ Tables 2 and 3 list the IPA function and disease annotations, associated p -values (right-tailed Fisher's exact test) and the total number of genes in that category. For the genes with the most variable regions in the transcribed regions, the highly enriched biological processes include cancer, cell death, and cell growth. For the genes with the most variable regions in the promoter regions, the highly enriched biological processes focus on tissue repairing, activation of various types of cells — in particular lymphocytes, tissue development, and liver disease.

4. Conclusion and Discussion

There are a number of normalization methods introduced to compare pairs of sequencing experiments or series of the same type of sequencing experiments.^{4–9} However, it is not feasible to apply these methods to compare tens, hundreds of sequencing experiments that are conducted on different samples with different pull-down proteins since the assumption on the same background level and noise model does not hold for different proteins. Our method provides a means to integrate ChIP-seq data from multiple experiments with different proteins, which can be of great importance in characterizing the regulome of the cells.

We observed that sample type and pull-down protein are the major factors that determine the whole genome profile of the ChIP-seq datasets. As shown in Figs. 1 and 2, experiments are tightly clustered by pull-down protein and/or cell line. For example, two hepatoma cell lines PLC5 and Huh7 are clustered with a non-tumorigenic human mammary epithelial cell line MCF10A with H3K27me3 epigenetic marker. In this case, H3K27me3 determines the background signal. On the other hand, human epithelial cell lines of MCF7 (breast cancer) and MCF10A (non-tumorigenic mammary) do not cluster by H3K9me3 but tightly cluster by sample cell line. These observations confirm that the whole genome profile of different cell lines and pull-down proteins are not the same. In addition, different regions of the genome have different profiles across the same set of experiments. Therefore, we need a method to compare multiple ChIP-seq datasets without changing the observed data.

In this study, we introduced a new method to compare multiple ChIP-seq experiments without normalization. While this method was originally developed for analyzing qRT-PCR data, which usually involve a small number of genes, the notion of analyzing patterns across multiple samples nevertheless is applicable to the large-scale genome-wide analysis. Specially, for every region, we calculated the average pairwise variation between itself and

all other regions as the stability measure, M value. By ranking all of the regions by M values, we identified the 200 most variable regions across 50 ChIP-seq experiments.

When we investigate 200 most variable transcribed and promoter regions, we observed that there are only 16 genes shared by these two regions. Furthermore, out of 411 and 752 regions with an M value larger than two in transcribed and promoter regions respectively, there are only 73 genes shared by the two regions. This observation demonstrates that there are substantial differences in enrichment of these two regions across the same set of experiments.

The functional and disease enrichment analyses on genes associated with the top 200 regions in each analysis demonstrate the biological relevance of the results. Genes associated with top 200 transcribed regions are highly enriched with several cancer-related functional groups such as cancer (74 genes) and breast cancer (36 genes), which is consistent with the fact that most of the cell lines and samples are from cancer tissues. In addition, cell death-related biological processes are highly enriched such as cell death (57 genes) and apoptosis (50 genes). Since the samples contain cell lines that are immortal and human tissues samples (mammosphere) that are not immortal, these genes may highlight such difference between the tissues. Finally, it is noticeable that the top functional groups consist of the most fundamental biological processes, including protein synthesis, cell death, cell cycle (interphase), and general cancer-related functions. In contrast, the highly enriched functional groups for the genes with the top 200 promoter regions focus more on cell/tissue types-specific functions as well as cell interaction processes. For instance, the samples contain two different tissues — breast tissue (MCF7, MCF10A, mammosphere, OHT) and liver tissue (PLC5 and Huh7). Out of the 200 genes, four of them are identified as liver disease genes, which implies that these genes can be applied to differentiate the two tissues. In addition, several enriched functional groups are related to tissue repairing, induction of immune cells (e.g. T lymphocytes), and connective tissue degradation. Given the complicated roles of tumor microenvironment in cancer development and its existence in mammosphere samples, we can explore more about the roles of the tumor microenvironment cells (e.g. immune cells and blood cells) and their interactions with the tumor.

This approach can also be used to evaluate the quality of series of a experiments. For example, MCF7 and OHT samples for ER α protein do not cluster with the mammosphere samples for the same protein. For these particular samples, we know that MCF7 and OHT samples for ER α protein had some problems in sample preparation.

In conclusion, we introduced a method to effectively compare multiple ChIP-seq experiments without changing the real observations. The proposed method does not require normality or homogeneity of variance for the data points.² We identified the least stable gene regions across a large set of experiments using the stability measure M . Our results confirm the effectiveness of the method and necessitate whole genome screening as the future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project is partially support by NCI grants (U54CA113001 and R01CA141090), and the PhRMA Foundation.

References

1. RefGene Database. Available at: <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/refGene.txt.gz>
2. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002; 3:research0034.1–0034.11. [PubMed: 12184808]
3. Ingenuity Pathway Analysis, IPA. Available at: <http://www.ingenuity.com>
4. Taslim C, Wu J, Yan P, Singer G, Parvin J, Huang T, Lin S, Huang K. Comparative study on ChIP-seq data: Normalization and binding pattern characterization. *Bioinformatics.* 2009; 25(18):2334–2340. [PubMed: 19561022]
5. Balwierz PJ, Carninci P, Daub CO, Kawai J, Hayashizaki Y, Van Belle W, Beisel C, van Nimwegen E. Methods for analyzing deep sequencing expression data: Constructing the human and mouse promoterome with deepCAGE data. *Genome Biol.* 2009; 10:R79. [PubMed: 19624849]
6. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods.* 2008; 5:613–619. [PubMed: 18516046]
7. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628. [PubMed: 18516045]
8. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keefe S, Haas S, Vingron M, Lehrach H, Yaspo ML. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science.* 2008; 321:956–960. [PubMed: 18599741]
9. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010; 11:R25. [PubMed: 20196867]

Biographies

Hatice Gulcin Ozer received her B.Sc. degree in Computer Engineering from Bogazici University, Turkey, her M.Sc. degree in Biophysics from Gaziantep University, Turkey, and her Ph.D. degree in Biophysics from The Ohio State University, USA, in 2000, 2003, and 2008 respectively.

She is currently a postdoctoral researcher in the Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA.

Yi-Wen Huang received his B.S. and M.S degrees from National Chung-Hsing University, Taiwan. He obtained his Ph.D. from The Ohio State University at 2006. Subsequently, he stayed at The Ohio State University Comprehensive Cancer Center as a post-doctoral researcher. Currently, he is a research scientist at The Ohio State University. The long-term research interests of Dr. Huang are to understand how epigenetics, especially DNA methylation, contributes to cancer development.

Jiejun Wu received his B.Sc. degree in Biology and Genetics from Peking University, China, and his Ph.D. degree in Biochemistry from The Ohio State University, USA, in 1983 and 1989 respectively. He is currently a research scientist at the Ohio State University.

Jeffrey Parvin received a Ph.D. in Microbiology and an M.D. from Mt. Sinai School of Medicine in 1989. For his Ph.D., Dr. Parvin studied the evolution of influenza viruses, and he developed methods for modifying the genome of the virus. From 1989 to 1994, Dr. Parvin was a postdoctoral fellow in the laboratory of Phillip A. Sharp at MIT, where he focused on the biochemistry of gene expression. From 1995 to 2007, Dr. Parvin was an Assistant Professor and then an Associate Professor at Harvard Medical School and based in the Brigham and Women's Hospital. While at Harvard Medical School, Dr. Parvin

continued the study of the biochemistry of gene expression and developed an interest in the breast cancer-associated tumor suppressor, BRCA1. Analysis of BRCA1 function in normal breast cells took the Parvin lab in new directions, including continuing studies on RNA transcription, but also DNA damage repair, function of ubiquitin ligases, and the regulation of centrosomes. Fellows in the Parvin lab developed the only known cell-free biochemical assays regulated by the BRCA1 enzymatic activity. Informatics methods of studying BRCA1 were begun in collaboration with the laboratory of Marc Vidal, and BRCA1 interacting proteins were inferred from gene expression data, and subsequently biologically confirmed. In 2007, Dr. Parvin joined the Department of Biomedical Informatics and the Cancer Center at Ohio State University, where he is starting the Systems Biology division. Continuing efforts in the Parvin lab are focused on using informatics methods to find proteins that participate in the same pathways as does BRCA1. In 2008, Dr. Parvin became the Director of the Biomedical Informatics Shared Resource at the OSU Comprehensive Cancer Center, and from 2008 to 2010, he was the interim-chair of the Biomedical Informatics Department.

Tim Hui-Ming Huang received his B.Sc. degree in Biology from Tunghai University, Taiwan, and his Ph.D. degree in Genetics from University of California in 1980 and 1989 respectively. Currently he is Professor in Department of Molecular Virology, Immunology and Medical Genetics–Human Cancer Genetics, The Ohio State University, Columbus, Ohio, USA. His research interest is focused around epigenetic alterations in breast and ovarian cancers.

Kun Huang received two B.S. degrees in Biology and Computer Science from Tsinghua University, Beijing, China, in 1996. In addition, he received the M.S. degree in physiology in 1998, the M.S. degree in Electrical Engineering in 2000, the M.S. degree in Mathematics in 2002, and the Ph.D. degree in Electrical and Computer Engineering in 2004, all from the University of Illinois at Urbana-Champaign. Currently he is an Associate Professor in the Department of Biomedical Informatics, The Ohio State University (OSU), Columbus, Ohio,

USA and the Co-Director of the OSU Comprehensive Cancer Center Biomedical Informatics Shared Resources. His research interests include systems biology, bioinformatics, biomedical imaging analysis, computer vision, and machine learning.

Fig. 1. Heatmap of experiment clustering based on correlation of region counts in transcribed regions. Pearson correlation coefficient is calculated for every sample pair using read counts in selected 14,462 regions.

Fig. 2. Heatmap of experiment clustering based on correlation of region counts in promoter region. Pearson correlation coefficient is calculated for every sample pair using read counts in selected 11,808 regions.

Fig. 3. Histogram of the M values of 14,462 and 11,808 transcribed and promoter regions of RefGene database across 50 ChIP-seq experiments, respectively.

Fig. 4. Clustering of top 200 most variable genes and 50 ChIP-seq experiments for transcribed regions. The distance function for hierarchical clustering is 1-correlation.

Fig. 5. Clustering of top 200 most variable genes and 50 ChIP-seq experiments for promoter regions. The distance function for hierarchical clustering is 1-correlation.

Table 1
Total of 50 ChIP-seq experiments are done on 10 different samples with 5 pull-down proteins.

Samples	Protein					Total number of lanes from Illumina GAT
	Pol II	ERα	H3K4me2	H3K9me2	H3K27me3	
mannosphere	2	1				3
mannosphere+E2	2	1				3
MCF10A w/AKT knockout				4	2	6
MCF10A				4	2	6
MCF7	2	2	2	4		10
MCF7+E2	2	2				4
OHT	2	2	2		4	10
OHT+E2	2	2				4
PLC5					2	2
Huh7					2	2
Total	12	10	4	12	12	50

Table 2

Top 25 IPA function and disease annotation terms associated with the top 200 transcribed regions.

Function annotation	p-value	# Genes
Neoplasia	3.72E-13	75
Cancer	6.47E-13	74
Tumorigenesis	1.95E-12	77
Breast cancer	5.52E-12	36
Synthesis of protein	1.18E-11	23
Cell death of fibroblast cell lines	7.08E-11	22
Biosynthesis of protein	9.76E-11	20
Cell death of cell lines	1.30E-09	43
Apoptosis of fibroblast cell lines	5.09E-09	17
Apoptosis of cell lines	9.19E-09	37
Cell death of tumor cell lines	1.06E-08	35
Carcinoma	1.33E-08	40
Cell death	1.71E-08	57
Apoptosis of tumor cell lines	3.06E-08	31
Apoptosis of eukaryotic cells	4.95E-08	44
Apoptosis	5.00E-08	50
Non-Hodgkin lymphoma	5.20E-08	13
Non-Hodgkin's disease	5.46E-08	14
Lymphomagenesis	8.61E-08	15
Cell death of eukaryotic cells	9.00E-08	49
Lung cancer	1.48E-07	21
Interphase	1.87E-07	21
Lymphoma	1.94E-07	14
Growth of cell lines	2.44E-07	29
Growth of fibroblast cell lines	2.62E-07	13

Table 3

Top 25 IPA function and disease annotation terms associated with the top 200 promoter regions.

Function annotation	<i>p</i> -value	# Genes
Influx of lymphocytes	2.77E-06	3
Repair of tissue	7.68E-06	4
Activation of myeloid dendritic cells	2.37E-05	3
Multiple sclerosis	3.39E-05	10
Myocarditis	5.62E-05	4
Induction of eukaryotic cells	6.87E-05	6
Influx of cells	7.32E-05	5
Detachment of corneocytes	7.91E-05	2
Inflammation of granuloma	7.91E-05	2
Influx of T lymphocytes	2.36E-04	2
Proliferation of sympathetic neuron	2.36E-04	2
Repair of epithelial tissue	2.36E-04	2
Accumulation of phagocytes	2.71E-04	6
Induction of normal cells	3.26E-04	5
Endocrine system disorder	4.64E-04	42
Activation of gamma-delta T lymphocytes	4.69E-04	2
Chemoattraction of antigen presenting cells	5.15E-04	3
Myocarditis of mice	5.15E-04	3
Feeding	5.49E-04	6
Degradation of connective tissue	6.07E-04	3
Ovarian cancer	6.52E-04	12
Developmental process of peripheral blood mononuclear leukocytes	7.10E-04	3
Mobilization of phagocytes	7.10E-04	3
Nonalcoholic fatty liver disease	7.10E-04	3
Influx of leukocytes	7.41E-04	4