GENE EXPRESSION SIGNATURES OF HUMAN PRIMARY MONOCYTES FROM HEALTHY INDIVIDUALS AND XLA PATIENTS USING DEEP RNA SEQUENCING ANALYSIS

HODA MIRSAFIAN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

GENE EXPRESSION SIGNATURES OF HUMAN PRIMARY MONOCYTES FROM HEALTHY INDIVIDUALS AND XLA PATIENTS USING DEEP RNA SEQUENCING ANALYSIS

HODA MIRSAFIAN

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: HODA MIRSAFIAN

Registration/Matric No: SHC120084

Name of Degree: Doctor of Philosophy (Ph.D.)

Title of Thesis ("this Work"): Gene expression signatures of human primary monocytes from healthy individuals and XLA patients using deep RNA sequencing analysis

Field of Study: Bioinformatics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: Dr. Saharuddin Bin Mohamad

Designation: Supervisor

Witness's Signature Date:

Name: Prof. Dr. Amir Feisal Merican Bin Aljunid Merican

Designation: Supervisor

ABSTRACT

Monocytes are essential cells of the innate immune system. They play important roles in the initiation and declaration of inflammation, generally through release of inflammatory cytokines, ROS (Reactive Oxygen Species) during phagocytosis and the activation of adaptive immune system. In this thesis, the transcriptome of primary monocytes from 6 healthy subjects and 3 patients with X-linked agammaglobulinemia (XLA), one of the inherited form of Primary immunodeficiency diseases (PIDs), were sequenced using deep polyadenylated (Poly(A)+) paired-end RNA sequencing (RNA-Seq) technique. The gene expression profiling was conducted on both healthy and disease RNA-Seq datasets. Approximately 1.3 billion reads were generated from healthy subject's RNA-Seq datasets. Using this datasets, the expression of 17,657 genes (including 11,644 protein-coding, 3,515 non-coding, and 2,498 pseudogenes) and 81,419 transcripts (including 70,457 annotated transcripts and 4,935 novel transcripts) were profiled from healthy subjects. The sequencing also generated approximately 477 million reads from XLA patients' samples which lead to the profile of 17,510 genes (including 11,788 protein-coding, genes 3,681 non-coding genes, and 2,041 pseudogenes) and 62,367 transcripts (including 58,136 annotated and 4,231 novel transcripts). A comparative study was conducted on gene expression profiles of 3 healthy male and 3 healthy female subjects to look into possible gender differences in expression patterns of immune-related genes. The results revealed that the innate immune-related genes are not equally expressed in primary monocytes of healthy male and female which indicated the disparity in innate immune response based on gender. Furthermore, the RNA-Seq datasets of the 6 healthy subjects were integrated with public domain RNA-Seq datasets of human monocytes to construct the gene reference catalogue of primary monocytes from healthy state monocytes. The long non-coding RNAs (lncRNAs) expression patterns analysis in monocytes was also conducted using these datasets which led to identification of several novel long intergenic

non-coding RNAs (lincRNAs) that have not been previously reported in monocytes. A comparative study was performed on gene expression profiles of XLA patients and healthy male subjects. The analysis detected several innate immune-related genes which are differentially expressed between XLA patients and healthy subjects, suggesting impaired immune function of monocytes and increased of susceptibility to apoptosis in monocytes of XLA patients. The results also showed the significant changes in lncRNAs expression patterns in primary monocytes of XLA patients compared to healthy subjects which may play roles in regulating the cell cycle and apoptosis in primary monocytes of XLA patients. The high-resolution genome-wide transcriptome expression profile of primary monocytes present in this study would provide a better understanding of monocytes characterization and function in healthy and XLA states. It also facilitates the detailed analyses of innate immune system abnormalities and novel pathomechanism concerning XLA.

ABSTRAK

Monosit adalah sel-sel penting dalam sistem imun manusia. Monosit memainkan peranan penting dalam permulaan dan perkembangan keradangan, secara amnya melalui pembebasan sitokin radang, ROS (Spesies Oksigen Reaktif) semasa fagositosis dan pengaktifan sistem imun adaptif. Dalam tesis ini, transkriptom monosit daripada 6 orang individu yang sihat dan 3 orang pesakit yang mengidap penyakit X-dikaitkan agammaglobulinemia (XLA), salah satu bentuk kes yang diwarisi oleh penyakit Kurang Daya Tahan Primer (PID), telah diprofilkan dengan menggunakan teknik poliadenilasi (Poly $(A)^{+}$) hujung berpasangan RNA penjujukan (RNA-Seq) secara mendalam. Kajian terhadap ekspresi gen-gen telah dijalankan ke atas kedua-dua set data RNA-Seq, individu sihat dan pesakit. Kira-kira 1.3 bilion bacaan dijana daripada set data RNA-Seq individu sihat. Menggunakan set data ini, ekspresi 17,657 gen (termasuk 11,644 protein-coding, 3,515 bukan pengekodan, dan 2,498 pseudogen) dan 81,419 transkrip (termasuk 70,457 transkrip beranotasi dan 4935 transkrip novel) telah diprofil daripada individu sihat. Penjujukan RNA ini juga telah menghasilkan kira-kira 477,000,000 bacaan daripada set data pesakit XLA yang diprofilkan iaitu: 17,510 gen (termasuk 11,788 protein-coding, 3,681 gen bukan pengekodan, dan 2,041 pseudogen) dan 62,367 transkrip (termasuk 58,136 beranotasi dan 4,231 transkrip novel). Kajian perbandingan telah dijalankan ke atas profil ekspresi gen 3 lelaki dan 3 perempuan yang sihat untuk mengetahui peranan perbezaan jantina dalam pola ekspresi gen-gen yang berkaitan dengan sistem imun. Hasil kajian menunjukkan bahawa ekspresi gen-gen sistem imun adalah tidak sama dalam monosit lelaki dan perempuan yang sihat, mencadangkan bahawa perbezaan jantina memainkan peranan dalam tindak balas imun secara semula jadi. Tambahan pula, set data RNA-Seq daripada 6 individu sihat telah diintegrasikan dengan set data RNA-Seq bagi membina katalog rujukan gen monosit delam individu yang sihat. Analisis corak ekspresi gen RNA panjang bukan pergekodan

(lncRNAs) dalam monosit juga dijalankan dengan menggunakan set data ini, yang membawa kepada pengenalan beberapa transkip novel intergenic bukan pengekodan (lincRNAs) yang belum dilaporkan sebelum ini. Selain itu, satu lagi kajian perbandingan telah dilakukan ke atas profil ekspresi gen pesakit XLA dengan individu lelaki yang sihat. Analisis ini mengesan beberapa gen sistem imun yang terekspres pada kadar yang berbeza antara pesakit XLA dengan individu lelaki yang sihat, menunjukkan fungsi monosit dalam sistem imun terjejas dan peningkatan kecenderungan apoptosis dalam monosit pesakit XLA. Keputusan ini juga menunjukkan perubahan ketara dalam corak lncRNAs dalam monosit utama pesakit XLA berbanding lelaki yang sihat yang boleh memainkan peranan dalam mengawal kitaran sel dan apoptosis dalam monosit utama pesakit XLA. Resolusi keseluruhan profil transkriptom monosit utama di dalam kajian ini akan memberi pemahaman yang lebih baik kepada pencirian monosit dan fungsi di dalam keadaan yang sihat dan berpenyakit. Ia juga memudahkan analisis secara terperinci mengenai keabnormalan sistem imun dan mechanisma patolog novel berkaitan penyakit XLA.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my supervisors Dr. Saharuddin Bin Mohamad and Prof. Dr. Amir Fesial Merican Bin Aljunid Merican for their continuous support of my Ph.D study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of my research and writing of this thesis. It has been pleasure to work under them and the experience in one that I am ever grateful for.

My sincere thanks also goes to Dr. Adiratna Mat Ripen, head of Primary Immunodeficiency Unit, Allergy and Immunology Research Centre, IMR, for her kind support, and guidance for performing laboratory experiments of this study.

I would also like to thank Dr. Masita binti Arip, head of Allergy and Immunology Research Centre, IMR, for allowing the usage of equipment and flexible schedule needed for the completion of this project. Also I thank Ms Chear Chai Teng, the staff of Primary Immunodeficiency unit, for her valuable assistance in performing screening tests and providing clinical notes of the patients.

I wish to thank to Mr Chang Lee Wei, research officer of CRYSTAL center, University of Malaya, and Dr. Thamilvaani Manaharan for their inspiration, guidance and advice. The financial support from University of Malaya in form of IPPP Postgraduate Grant and High Impact Research (HIR) Grant are also greatly acknowledged.

Last but not list, I would like to express my special thanks to my parents for their unflagging love and unconditional support through my life and my studies. I owe them everything and wish I could show them just how much I love and appreciate them. I am also very grateful to my brother who was always willing to help me and encourage me in whatever I wanted to do. I also thank to my friends, Elham Kalfi, Niloofar Alavi, Maryam Yazdi and Negar Havazade, who supported my emotionally and encourage me during my Phd study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

adjp	Adjusted the p-value
BLASTN	Basic Local Alignment Search Tool for Nucleotide
Bp	Base pair
BSA	Bovine Serum Albumin
cDNA	Complementary Deoxyribonucleic Acid
CPAT	Coding Potential Assessment Tool
DAVID	Database for Annotation, Visualization and Integrated Discovery
DE	Differentially Expressed
EDTA	Ethylenediaminetetraacetic acid
eRNAs	Enhancer RNAs
FITC	Fluorescein Isothiocyanate
FPKM	Fragments Per kilobases of Exon Per Million Fragments Mapped
GO	Gene Ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
Ig	Immunoglobulin
KEGG	Kyoto Encyclopedia of Genes and Genomes
lincRNAs	long intergenic non-coding RNAs
lncRNAs	long non-coding RNAs
miRNAs	microRNAs
ml	Milliliter
mRNAs	Messenger RNAs
ng	Nanogram
NGS	Next Generation Sequencing Technologies
ORFs	Open Reading Frames
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate-buffered Saline Buffer
PIDs	Primary Immunodeficiency Diseases
$Poly(A)^-$	Non-Polyadenylate
$Poly(A)^+$	Polyadenylated
qRT-PCR	Quantitative Reverse Transcription Polymerase Chain Reaction
RNA-Seq	RNA sequencing
RT-PCR	Reverse Transcription Polymerase Chain Reaction
siRNA	small interfering RNAs
snoRNAs	small nucleolar RNAs
TFs	Transcription Factors
WebGestalt	WEB-based GeneSet Analysis Toolkit
XLA	X-linked Agammaglobulinemia
β-ME	β-Mercaptoethanol
µg	Microgram
µl	Microliter

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background

The immune system consists of different cell types which protect the body against various illnesses by perceiving and reacting towards antigens leading to expression of distinct gene expression profiles. The immune system is classified into two categories: innate immune system (comprising myeloid cells; monocytes and macrophages, neutrophils, eosinophils basophils, mast cells and natural killer (NK) cells) and adaptive immune system (comprising lymphoid cells; B cells, and various types of T cells). Monocytes are key elements of the innate immune system which become the first line of defense response against pathogens (Janeway, 2001). They are mononuclear cells and play central role in innate immune-mediated processes including clearance of microbial infections and cellular debris, secretion of immunoregulatory bioactive factors including interleukins, interferons, chemokines and growth factors and control of cancer progression (Kraft-Terry & Gendelman, 2011).

Primary immunodeficiencies (PIDs) are disorders in which specific component of the immune system is either absent or does not function properly. PIDs are caused by mutations in a particular gene or several genes, which may result in the defects in the innate immunity or/and adaptive immunity of the body (McCusker & Warrington, 2011; Schroeder, Schroeder, & Sheikh, 2004). X-linked agammaglobulinemia (XLA) is one of the genetic form of PIDs. It is a rare disease influencing males in roughly 1/200,000 live births (Vihinen et al., 1999). XLA is caused by mutations in the *BTK* (Bruton Tyrosine Kinase) (Vetrie et al., 1993) which resulted in the defects of development and maturation of B cell within the bone marrow and a considerable decrease or complete absence of mature B cells in peripheral blood. Due to the absence of mature B cells, XLA patients have significantly low levels of all major serum immunoglobulins and consequently,

would be subjected to serve chronic bacterial infections (Noordzij et al., 2002; Ochs & Smith, 1996). So far, the literature has indicated that *BTK* is crucial for development and differentiation of the B cells (Lopez-Herrera et al., 2014; Maas & Hendriks, 2001; Middendorp, Dingjan, Maas, Dahlenborg, & Hendriks, 2003). However, recently, it has been reported that *BTK* is also involved in the regulation of other cell types, such as neutrophils (Honda et al., 2012), NK cells (Bao et al., 2012), and monocytes/macrophages (Koprulu & Ellmeier, 2009). Previous report proposed the impaired phagocytosis in monocytes of XLA patients due to *BTK* deficiency (Amoras, Kanegane, Miyawaki, & Vilela, 2003).

In living organism, DNA encodes all the information that are essential for each single cell function. Cells can dynamically access and translate specific information through gene expression by selectively switching on and off particular genes. In the selected genes, the information encoded are transcribed into RNA molecules, which consequently can be translated into proteins or can be directly used to control gene expression (Finotello & Di Camillo, 2015). The transcriptome is a set of all RNA transcripts existing in a cell or tissue at a certain point of time under specific conditions (Sirri, Urcuqui-Inchima, Roussel, & Hernandez-Verdun, 2008). Thus transcriptome study is essential to identify the current state of the cells and fundamental pathogenic mechanisms of diseases. In addition, differential gene expression study facilitates the comparison of gene expression profiles from different cells and conditions to characterize genes that are responsible in the determination of phenotypes. For example, the comparison of healthy versus diseased cells can present new insights on genetic aspects involved in pathology (Finotello & Di Camillo, 2015). Previously, microarray has been the most important and commonly used method for transcriptome analysis (Baldi & Hatfield, 2002), however recently, high-throughput RNA sequencing (RNA-Seq) has become a powerful alternative approach for transcriptome studies. RNA-Seq able to

qualitatively and quantitatively explore RNA molecules including messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), microRNAs (miRNAs) and small interfering RNAs (siRNA) (Dong & Chen, 2013).

During the period of this study, in the PID unit of the Allergy and Immunology Research Centre (AIRC), Institute for Medical Research, Malaysia, 11 male subjects with no circulating B cells, low serum immunoglobulin isotypes and downregulation of *BTK* expression in their monocytes were diagnosed as XLA. Molecular genetic tests have been performed on all patients from unrelated families which revealed *BTK* gene mutations in all patients. Novel *BTK* invariant splice site mutations were identified in one of the patients (Chear et al., 2013). In this thesis, deep RNA-Seq analysis was performed on primary monocytes from 3 selected XLA patients as well as 6 healthy subjects to generate a genome-wide transcriptome profile of primary monocytes under healthy and XLA disease states. A comparative analysis has been conducted on transcriptomes of XLA patients and healthy subjects to study the differential gene expression patterns and variations between two datasets. Furthermore, the RNA-Seq datasets from 6 healthy subjects were integrated with other public domain RNA-Seq datasets for human monocytes to generate the comprehensive gene reference catalogue of human monocytes. The workflow of this study is depicted in Figure 1.1. This study would provide important insights the gene expression patterns in primary monocytes under healthy and XLA disease states and offers a new direction for the physiopathology of XLA.

1.2 Hypothesis

BTK deficiency in XLA patients leads to defect in B cell development and function. However, *BTK* is also known to be expressed in monocytes. Due to *BTK* deficiency in XLA patients, it is hypothesized that the transcriptome expression of

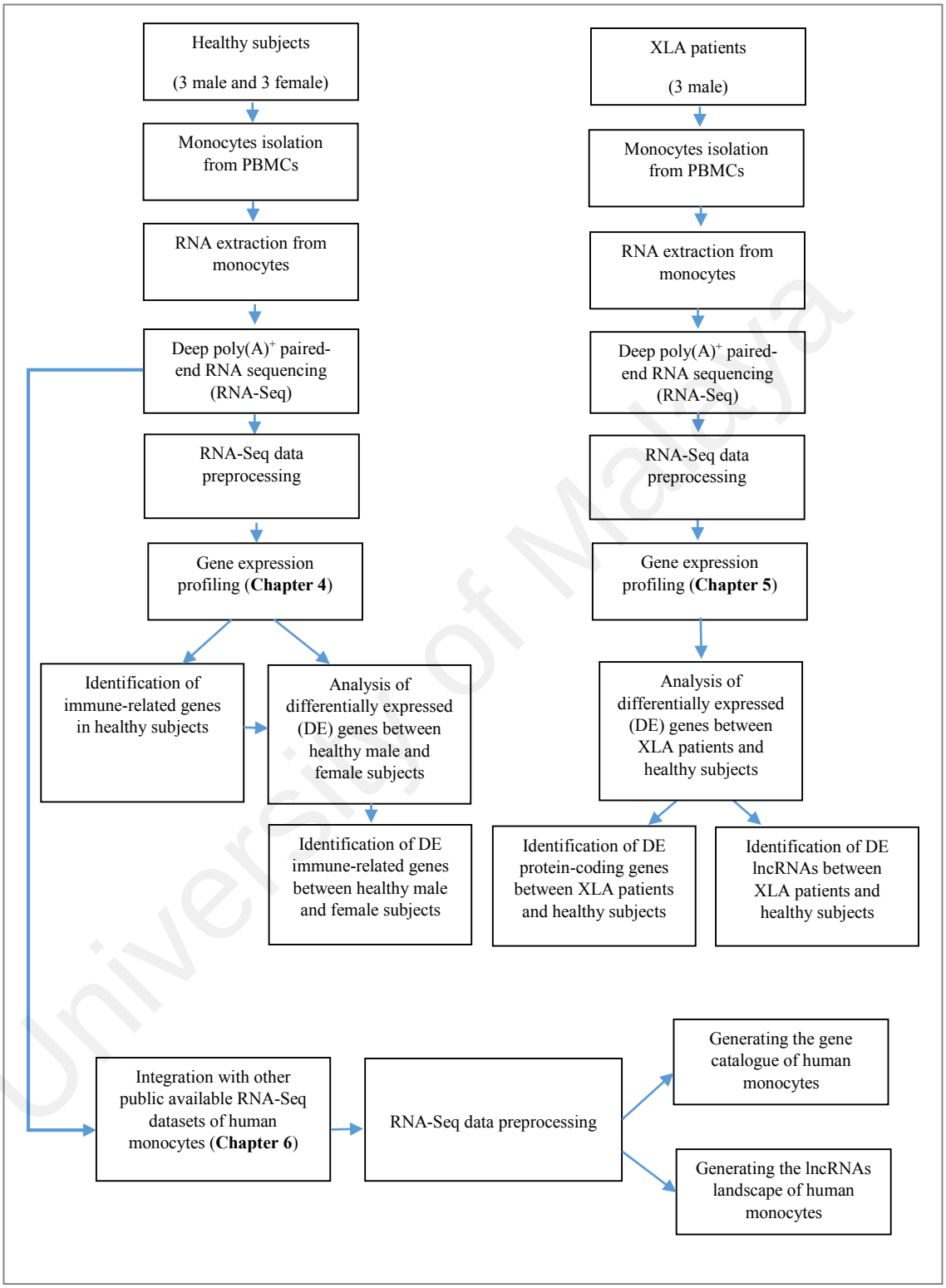

Figure 1.1: Schematic representation of the workflow used in this study.

primary monocytes in XLA patients would differ from the one expressed in healthy subjects that would reflect the impaired function of monocytes of XLA patients.

1.3 Research questions

There are four research questions based on the mapping and quantification of the transcriptome of monocytes using deep RNA-Seq technology involved in this thesis:

1. What is the transcriptome profile of the primary monocytes of healthy subjects?
2. What is the transcriptome profile of the primary monocytes of XLA patients?
3. Are there any differences in transcriptome of primary monocytes between XLA patients and healthy subjects?
4. Which biological pathways in primary monocytes of XLA patients are affected by *BTK* deficiency?

1.4 Objectives

To answer the research questions, four objectives were set up:

1. To profile the gene expression of human primary monocyte from healthy male and female subjects.
2. To profile the gene expression of human primary monocyte from XLA patients.
3. To identify the gene expression profile variation of human primary monocyte between XLA patients and healthy male subjects.
4. To determine the biological pathway of primary monocytes that affected by *BTK* deficiency.

1.5 Thesis organization

The contents of this thesis are organized into several chapters which are: Chapter 1: Introduction, Chapter 2: Literature review, Chapter 3: Methodology, Chapters 4, 5 and

6: Results and Discussion, Chapter 7: General Discussion, and Conclusions. Chapter 1 contains the overall introduction to the research concerned and the objectives of the study. Chapter 2 provides a general introduction on immune system and innate immune system, monocytes development and functions, PIDs disease followed by description of XLA, the roles of *BTK* in innate immune system, and RNA-Seq technology. Chapter 3 describes the research design and methodology, including blood sample collection from healthy subjects and XLA patients, monocyte isolation, RNA extraction from purified monocytes, RNA sequencing and bioinformatics data analysis. Chapters 4, 5, and 6 describe all the findings of this study. Also the relationship between results in this study and the one reported by others discussed. Chapter 7 includes summarizes the all findings and points out limitations of the current work, and also outlines directions for future research.

CHAPTER 2: LITRERATURE REVIEW

2.1 Immune system

The immune system's main function is to defend the body against diseases and infections. The immune system is comprised of different cell types, tissues and organs as well as secreted compounds in which all interact to identify and eliminate numerous pathogenic microbes and toxins in the body. There are two parts of the immune system: innate and adaptive (Turnbaugh et al., 2007). The components of innate and adaptive immunity are summarized in Table 2.1. The innate immune system is non-specific and exhibit a robust immune response, cellular and humoral processes. On the other hand, the adaptive immune system shows a highly miscellaneous range of antigen-specific recognition receptors that facilitate detection and removal of pathogens. Furthermore, the adaptive immune warrants adapted immune reactions and long-lasting immunological memory against recurrent infection (Dunkelberger & Song, 2010).

The innate immune response is the first line of human body's defense which reacts quickly to any infectious agent and offers the primary phase of an actual defense (Medzhitov & Janeway, 1997; Zimmerman, 2012). This system identifies, destroys and delivers antigens to the subsequent lymphoid tissue. Host molecules mediate recognition of pathogens through the *PRRs* (Pattern Recognition Receptors) which identify viral and microbial components. *PRRs* are expressed on the surface of the cells and in intracellular compartments, or secreted into the blood stream and tissue fluids (Abbas, Lichtman, & Pillai, 2012; Tizard, 2013). *PAMPs* (Pathogen-associated Molecular Patterns) are molecular components of pathogen that are recognized and bind to the *PRRs*. *TLRs* (Toll-like Receptors) are a major group of *PRRs* that have important role in recognition of a wide range of *PAMPs*, leading to activation of the immune responses (Kuby, Kindt, Goldsby, & Osborne, 2007).

Table 2.1: Components of the human immune system (adapted from Benito-Martin, Di Giannatale, Ceder, & Peinado, 2015).

Immunity	Molecules	Molecules
Innate	Monocytes Macrophages Dendritic cells Natural killer (NK) cells Neutrophils Mast cells Basophils Eosinophils	Cytokines Chemokines Complement
Adaptive	T cells: Hellper T ($CD4^+$) Killer T ($CD8^+$) Memory T Suppresser T B cells	Cytokines Antibodies

TLRs recognize the *PAMPs* at the cellular surface or endosomal membranes. Upon binding of *PAMPs*, *TLRs* transfer signals into the intracellular environment via adapter proteins such as *TRIF* (Toll like Receptor-domain-containing adapter-inducing interferon-β) (Yamamoto, 2003), *MAL* (Mal T-Cell Differentiation Protein) (Fitzgerald et al., 2001) and *MyD88* (Myeloid Differentiation Primary Response 88) (Medzhitov & Janeway, 1997; Sun & Ding, 2006) which induces *NFκB* (Nuclear Factor Kappa B Subunit) signaling and the MAP kinase pathway and consequently, secretion of pro-inflammatory cytokines and co-stimulatory molecules. (Janeway & Medzhitov, 2002; Piras & Selvarajoo, 2014; Tizard, 2013) (Figure 2.1).

RLRs (RIG-I-like Receptors; Retinoic Acid-Inducible Gene-I-like Receptors) and *NLRs* (NOD-like Receptors; Nucleotide-Binding Oligomerization Domain-like Receptors) are other members of *PRRs*, which are cytosolic detection systems for intracellular *PAMPs* (Kawai & Akira, 2009). *RLRs* belong to the DExD/H box helicases protein family including probable ATP-dependent RNA RIG-I (Schlee, 2013). Their main role is in the recognition of viruses through binding to *PAMPs* motifs within RNA ligands that accrue during virus infection. This interaction activates signaling pathway that induce production of *IFNs* (Interferons) and proinflammatory cytokines (Gack, 2014; Weber et al., 2013) (Figure 2.1). *NLRs* comprise a large family of intracellular *PRRs*, such as *NOD1* (Nucleotide Binding Oligomerization Domain Containing 1), *NOD2* (Nucleotide Binding Oligomerization Domain Containing 2) and *NALP3* (Pyrin Domain Containing 3). *NOD1*, *NOD2* detect the intracellular cell products of bacteria, while *NALP3* responds to several stimuli to form a multi-protein complex named *NALP3* inflammasome, which stimulates the release of the *IL-1* (Interleukin 1) family of cytokines (Kawai & Akira, 2009; Shayakhmetov, 2010) (Figure 2.1).

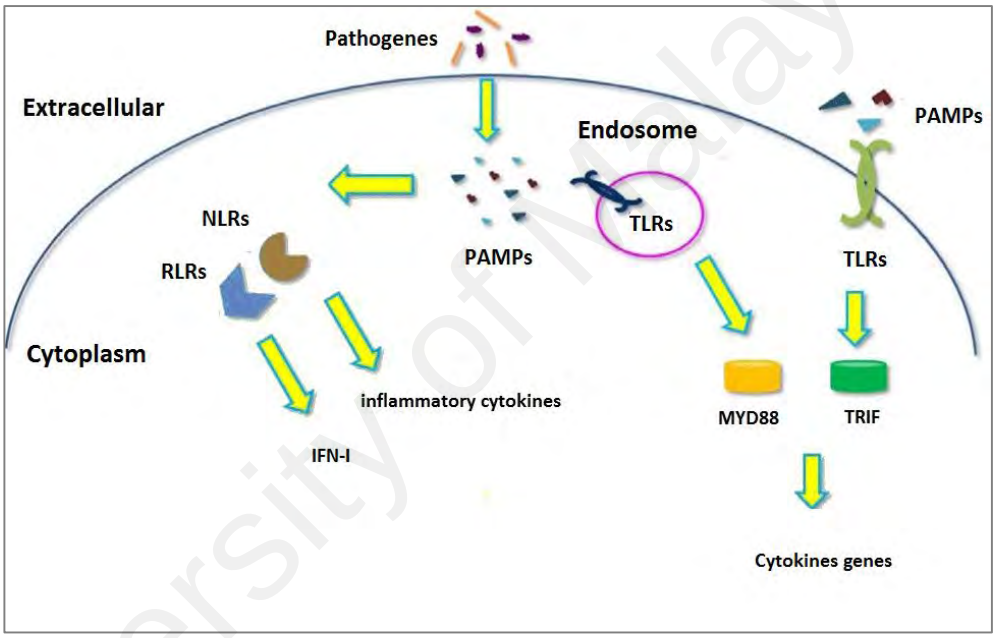

Figure 2.1: Three classes of pattern recognition receptors (*TLRs*, *RLRs*, and *NLRs*) with their roles in inducing host antiviral responses (adapted from Reddy, 2014).

2.1.1 Monocytes

Monocytes are important components of the innate immune system and large circulating leukocytes of the myeloid family. They have vital roles in the initiation and resolution of inflammation primarily via phagocytosis, generation of ROS (reactive oxygen species), activation of the acquired immune system and release of inflammatory cytokines. Monocytes constitute 5 to 10% of the entire white blood cells in human body (Bijl, 2006).

2.1.1.1 Monocytes development and functions

Monocytes are derived from hematopoietic stem cells in the bone marrow that proliferate and differentiate via several commitment phases. Hematopoietic stem cells generate multipotent common myeloid progenitor cells (CMPs), that, differentiate into granulocyte-monocyte progenitor cells (GMPs) and then to monocyte–dendritic progenitor cells (MDPs). MDPs produce monocytes and committed dendritic progenitors cells (CDPs) in the bone marrow (Geissmann et al., 2010) (Figure 2.2).

Monocytes enter the blood stream, where they circulate for 1 to 3 days, and apparently mature during circulation (Ziegler-Heitbrock, 2000). In the blood, monocytes act as the first defense line against invading pathogens. Upon inflammatory conditions and tissue damage, monocytes enter the into the tissue, and differentiate into tissue macrophages or dendritic cells (Auffray, Sieweke, & Geissmann, 2009; Varol, Yona, & Jung, 2009; Yona & Jung, 2010). Based on the condition of the tissue environment, monocyte differentiated into macrophages that acquire tissue specificity, such as microglia cells in the brain, splenic macrophages, Kupffer cells in the liver, osteoclasts in the bone or alveolar macrophages in the lung. Under inflammatory conditions, they may also differentiate into myeloid dendritic cells (mDCs) (Geissmann et al., 2010). Therefore, their plasticity to differentiate into several cell types shows their potential to

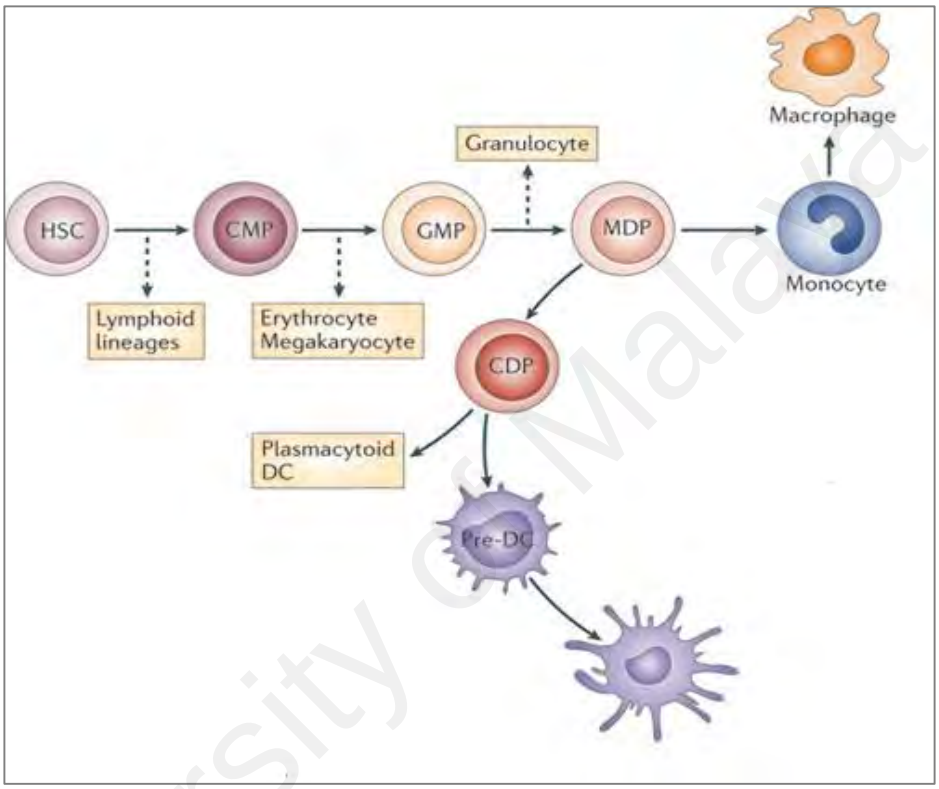

Figure 2.2: The developmental pathway of monocytes from hematopoietic stem cells (adapted from Chow, Brown, & Merad, 2011). Common myeloid progenitor cells (CMPs) generated from hematopoietic stem cells, which then differentiated into granulocyte-monocyte progenitor cells (GMPs) and monocyte–dendritic progenitor cells (MDPs). MDPs can give rise to monocytes and committed dendritic progenitors cells (CDPs) in the bone marrow.

participate in a wide variety of cellular processes (Geissmann et al., 2010). Because macrophages mainly perform phagocytosis and mDCs' major function is to process antigens and present them to T-cells, monocytes are assumed to be between the adaptive and innate immunity (Geissmann et al., 2010).

2.1.1.2 Monocytes phenotypic heterogeneity

Monocytes are heterogeneous and can be classified into three subgroups based on the expression levels of antigenic markers CD16 (Fcγ Receptor III) and CD14 (a receptor for bacterial lipopolysaccharide). The classes are: (i) "Classical" monocytes ($CD14^{++}CD16^{-}$) which expressed high levels of CD14 without expressing CD16 and accounting for 90−95% of monocytes in the bloodstream. (ii) "Non-classical" monocytes ($CD14^{+}CD16^{++}$) which have low expression of CD14 and high expression of CD16, and (iii) "intermediate" monocytes ($CD14^{++}CD16^{+}$) which expressed both CD14 and CD16 markers (Martinez, 2009; Passlick, Flieger, & Ziegler-Heitbrock, 1989; Ziegler-Heitbrock et al., 2010). CD14 is necessary for identifying bacterial lipopolysaccharide (LPS) existing in Gram-negative bacteria. It is acts as *PRRs* that receives LPS from LPS-binding protein. CD14-LPS work together with several *TLRs*, such as *TLR2*/*TLR6*, *TLR2*/*TLR1* and *TLR4-MD2* (Myeloid Differentiation Factor-2) to stimulate the endotoxin cellular response (Ziegler-Heitbrock et al., 2010). CD16 is a relatively low-affinity receptor for the Fc portion of IgG antibodies in complex with their antigens, which induces the monocytes to engage in antibody-antigen complexes by phagocytosis and eliminate them from the circulation (Ziegler-Heitbrock et al., 2010).

The monocytes subsets are different in terms of chemokine receptor expression, phagocytic activity and tissue distribution in steady state or during inflammation. The classical monocytes express high levels of *CD62L* (CD62 Antigen-Like Family Member L) and *CCR2* (C-C Motif Chemokine Receptor 2) and low level of *CX3CR1* (C-X3-C

Motif Chemokine Receptor 1) (Figure 2.3). Since the monocytes mainly performing phagocytosis, they display high peroxidase activity, and make low levels of *TNFα* (Tumor Necrosis Factor Alpha) and high levels of *IL-10* (Interleukin 10) in reaction to LPS (Cros et al., 2010; Frederic Geissmann, Jung, & Littman, 2003). Classical monocytes favorably express genes contributing to coagulation, angiogenesis and wound healing (Wong et al., 2011). The intermediate monocytes have inflammatory role and low peroxidase activity, but higher capacity to make and release *TNFα* and *IL-1β* (Interleukin 1 Beta) in response to LPS (Cros et al., 2010). Intermediate monocytes are linked to T cell activation and antigen presentation by gene signature (Wong et al., 2011). Intermediate and classical monocytes are tethered during inflammation, and enter into the tissue through interaction of *CCR5* (C-C Motif Chemokine Receptor 5)/*CCL5* (C-C Motif Chemokine Ligand 5) and/or *CCR2*/*CCL2* (C-C Motif Chemokine Ligand 2) in a *VLA1* (Very Late Activation Antigen-1)/*VCAM1* (Vascular Cell Adhesion Molecule 1) dependent way. The non-classical monocytes patrol blood vessels for damage and act via interaction of *CX3CR1*/*CCL3* (C-C Motif Chemokine Ligand 3) complementary pair through the Leu-CAM (Leukocyte Adhesion Molecules) family integrin *LFA-1* (Lymphocyte Functional Antigen-1)/*ICAM1* (Intercellular Adhesion Molecule 1)-dependent way. *IL-1β*, and *TNFα* are released by this subset in reaction to DNA, RNA particles, associating with the pathological role in autoimmune disease, like rheumatoid arthritis (Wong et al., 2011) (Figure 2.3).

2.2 Gender and immune response

The immune function and response to the pathogen are affected by various factors, modulators and challenges. One of the important factor is gender (Lozano et al., 2012; Oertelt-Prigione, 2012). Generally, mortality and morbidity rates are lower in females than males throughout life (Lozano et al., 2012). Increased severity of infectious diseases

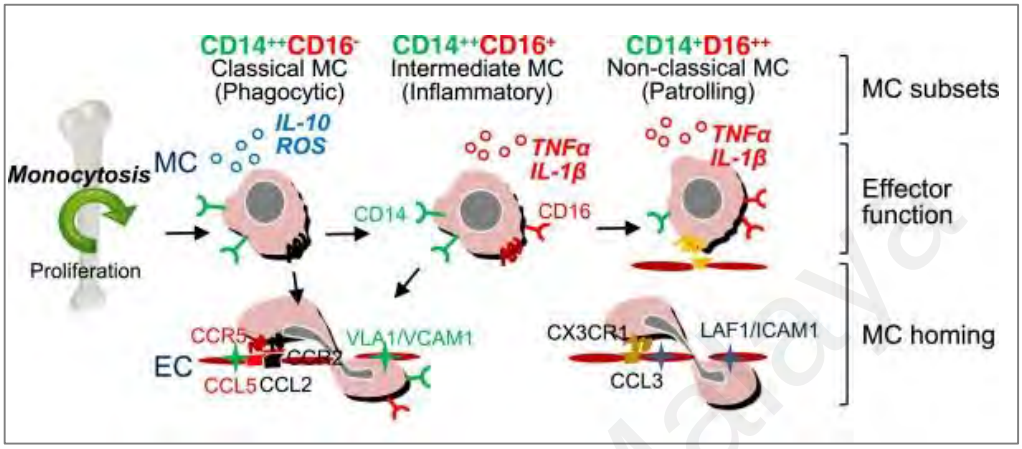

Figure 2.3: Differentiation and functions of human monocytes subsets. Under healthy state, classical monocytes leave the bone marrow, circulate in the blood stream and can differentiate into intermediate monocytes and non-classical monocytes in circulation. Classical monocytes have high phagocytosis and antimicrobial capacities and secrete *ROS* (Reactive Oxygen Species) and *IL-10* (Interleukin 10) upon LPS stimulus. Upon inflammatory stimulation, the intermediate and non-classical monocytes secrete inflammatory cytokines, *TNFα* and *IL-1β*. During inflammation, classical and intermediate monocytes are bound together and enter the tissue through interaction of complementary pair *CCR2*/*CCL2* or/and *CCR5*/*CCL5* in a *VLA1*/*VCAM1* dependent manner (adapted from Yang, Zhang, Yu, Yang, & Wang, 2014).

and susceptibility for males are the causes of this uneven distribution during infancy and childhood (Anker, 2007). It is more likely that male humans and mice have more frequent and more severe parasitic, bacterial, fungal, and viral infections than females (Klein, 2000). Yet, females are more susceptible to robust immune reactions to antigenic challenges, including vaccination and infection (Klein, Jedlicka, & Pekosz, 2010). The differential gene expression study on $CD4^+$ T cells from gut mucosal samples of healthy subjects indicated that females had higher levels of inflammation-associated gene expression as well as $CD8^+$ and $CD4^+$ T cells activation (Sankaran-Walters et al., 2013). Enumeration of lymphocyte subset in blood showed that female had higher number of B cells than male (Abdullah et al., 2012) Furthermore, Fan et al. (2014) recently stated the presence of gender differences in global gene expression of B cells, which are attributed to estrogen. Even though female shows better immune responses that may lead to faster infection clearance, immune-mediated pathology may also develop in them (Meier et al., 2009). For instance, chances of death from *H5N1* (Avian Influenza Virus) is 2 to 6 times more in female, partially because of intensified immune responses (Klein, 2012). Also, female are more susceptible to many inflammatory and autoimmune diseases than male. Female constitute 80% of individuals with autoimmunediseases (Voskuhl, 2011), while male may die from all malignant cancers 1.6 times more than female (Cook, McGlynn, Devesa, Freedman, & Anderson, 2011). The quantity and activity of cells related to innate immunity are different between male and female. The macrophages and neutrophil's phagocytic activity is lower in males than females (Klein, 2004). After antigenic or parasitic stimulation, females produce and release more *NO* (Nitric Oxide), *TXB2* (Thromboxane B2) and *PGE2* (Prostaglandin E2) than males (Barna, Komatsu, Bi, & Reiss, 1996). However, other studies show that following trauma, concentrations of plasma from many pro-inflammatory cytokines, including *TNFα* and *IL-6* (Interleukin 6) are higher in male (Diodato, Knöferl, Schwacha, Bland, & Chaudry, 2001). Also, NK

cells are major first defense line against parasites. Lower NK cells activity has been reported in female in the luteal phase of menstrual cycle and those with regular menstrual cycles in comparison to male (Souza et al., 2001). Studies on mice revealed that oestradiol is able to decrease the NK cells' quantity as well as activity (Hanna & Schneider, 1983). In female, antigen-presenting cells (APCs) are more effective in providing peptides than in male (Weinstein, Ran, & Segal, 1984). In female mice, the expression of MHC class II on microglia, endothelial cells and astrocytes is improved after infection of the central nervous system than male ones. Moreover, studies show that in female mice, macrophages express greater level of *p38*/*MAPK14* (Mitogen-Activated Protein Kinase 14) and *MyD88*; thereby higher activation following LPS challenge compared to male (Barna, Komatsu, Bi, & Reiss, 1996).

This variability in immune responses between gender might be attributed to the genetic, sex hormones (Bhatia, Sekhon, & Kaur, 2014) and gender variant behaviors (Muenchhoff & Goulder, 2014). The genetic variance resulted in the presence of different genes on the X and Y chromosomes. The X chromosome is one of the main differences of immune responses between genders (Candore et al., 2010). The X chromosome encodes over 1100 genes, which is 10 times higher than the Y chromosome. Immune function is regulated by many genes on the X chromosome and they also play significant part in controlling gender differences in the development of immune-related diseases (Libert, Dejager, & Pinheiro, 2010). These immune-related genes code several proteins *PRPs* (such as *TLR7* and *TLR8*), transcriptional factors [such as *FOXP3* (Forkhead Box P3)], cytokine receptors (such as *IL2RG* (Interleukin 2 Receptor Subunit Gamma) and *IL13RA2* (Interleukin 13 Receptor Subunit Alpha 2) (Fish, 2008). Since female cells hold two copies of the X chromosome, their cells inactivate X chromosome to avoid double quantity of the encoded genes, resulting in cellular mosaicism. Therefore, female cells express approximately 50% of X-encoded genes from the paternal X

chromosome, and 50% from the maternal X chromosome (Lahn, Pearson, & Jegalian, 2001). Consequently, male exhibit higher prevalence of X-linked immune-deficiencies. Nevertheless, female develop autoimmune diseases more often, perhaps due to differences in effects of sex chromosome genes and gonadal hormones (Libert et al., 2010; Voskuhl, 2011).

Sex hormones are categorized as progesterone, estrogens (mostly 17 β-estradiol in the ovarian cycle) and androgens (primarily testosterone) (Bhatia et al., 2014). It is recognized that sex hormones control immune response via the relation with particular hormone receptors expressed by immune cells, and have also a significant role in controlling the onset/continuation of autoimmune diseases (Ortona, Delunardo, Maselli, Pierdominici, & Malorni, 2015). Normally, steroid hormones play an opposite part in the immune response since androgens and progesterone function as natural immunosuppressants and estrogen works as enhancer of humoral immunity. Particularly, androgens, progesterone and estrogens are seen both in female and male, but at different levels. Also, their effects is contingent on the type of target immune cell and their concentration levels (Ortona et al., 2015).

Furthermore, cultural and social factors are significant factors in susceptibility to disease, because gender has effect on exposure patterns to infections and cures (Anker, 2007). In some societies, for example, mortality rates of measles are greater in female than male since it is more likely that girls stay at home, so increasing their contact with sick siblings as well as risk of infection (Fish, 2008). World Health Organization (WHO), has also reported crucial gendered differences in access to healthcare, which can influence the care levels given to male and female. For instance, a study in Kolkata, India, showed higher chance of rehydration and receiving qualified health care in male with diarrhea than female (Anker, 2007).

2.3 Primary immunodeficiency disease (PIDs)

A heterogeneous group of diseases called the primary immunodeficiency diseases (PIDs) which caused by congenital defects in the growth and maturation of immune cells. This group may include adaptive immunity defects, with the involvement of one or both of T and B cells. The innate immune system may also be affected by PIDs. The disorders in most cases of PIDs are resulted from single gene mutation; yet, some of these disorders comprised more than one gene mutations (McCusker & Warrington, 2011; Schroeder et al., 2004). Higher susceptibility to certain infectious pathogens is the major characteristic of patients with PIDs. There is correlation between the type of immunological defect and the pathogen type.

So far, over 300 varieties of PIDs have been identified and categorized into 8 groups includes combined immunodeficiencies, well-defined syndromes with immunodeficiency, mainly antibody deficiencies, congenital phagocytes defects, diseases of immune dysregulation, innate immunity defects, auto-inflammatory disorders, and complement deficits (Bousfiha et al., 2015; Picard et al., 2015). About 50% of the PIDs are related to insufficient or defective production of antibody, resulted from very low quantities of B cells producing antibody or B cells that have malfunction, leading to insufficient production of antigen-specific antibodies. Septicemias with bacteria as well as recurrent pulmonary and sinus infections are characteristics of these disorders. To date, over 30 types of antibody deficiency have been identified (Picard et al., 2015) (Table 2.2). X-linked agammaglobulinemia (XLA) is the most severe deficiency, characterized by a low level or absence of mature B cells or antibody-secreting plasma cells (Picard et al., 2015).

Table 2.2: Classification of predominantly antibody deficiencies (adapted from Picard et al., 2015).

Diseases
Severe reduction in all serum immunoglobulin isotypes with profoundly decreased or absent B cells 1. BTK deficiency 2. μ heavy chain deficiency 3. I5 deficiency 4. Igα deficiency 5. Igβ deficiency 6. BLNK deficiency 7. PI3KR1 deficiency 8. E47 transcription factor deficiency 9. Thymoma with immunodeficiency
Severe reduction in at least 2 serum immunoglobulin isotypes with normal or low number of B cells 1. Common variable immuno-deficiency disorders 2. CD19 deficiency 3. CD81 deficiency 4. CD20 deficiency 5. CD21 deficiency 6. TACI deficiency 7. BAFF receptor deficiency 8. TWEAK deficiency 9. NFKB2 deficiency 10. MOGS deficiency 11. TRNT1 deficiency 12. TTC37 deficiency
Severe reduction in serum IgG and IgA with normal/elevated IgM and normal numbers of B cells 1. AID deficiency 2. UNG deficiency 3. INO80 4. MSH6
Isotype or light chain deficiencies with generally normal numbers of B cells 1. Activated PI3K-δ 2. PI3KR1 loss of function 3. Ig heavy chain mutations and deletions 4. IGKC deficiency 5. Isolated IgG subclass deficiency 6. IgA with IgG subclass deficiency 7. Specific antibody deficiency with normal Ig concentrations and normal numbers 8. of B cells 9. Transient hypogammaglobulinemia of infancy with normal numbers of B cells 10. CARD 11 gain of function

2.3.1 X-linked Agammaglobulinemia (XLA)

XLA was the foremost described human immunodeficiency for which the laboratory and clinical findings dictated effective treatment. Bruton (1952) reported the case of an 8 year-old boy who had recurrent infections and no obvious gamma globulin fraction was observed in analysis of serum by protein electrophoresis (Bruton, 1952). The subject was cured with monthly intramuscular injections of human gamma globulin with considerable clinical progress. While this case had no family history, the subsequent cases which almost were males, showed a similar clinical phenotype. This suggested that Bruton's agammaglobulinemia was inherited in an X-linked pattern and females that have a *BTK* mutant on one of their X chromosomes are carriers for XLA (Elphinston & Wickes, 1956; O'Brien & Sereda, 1956).

2.3.1.1 Molecular basis of XLA

In 1993, it was reported that XLA is caused by mutations in the *BTK* (Bruton's tyrosine kinase) (Vetrie et al., 1993). The *BTK* is situated at band Xq21.3 to Xq22 at the long arm of the X chromosome, spanning 37.5 kb with 19 exons making 659 amino acids (Gaspar & Kinnon, 2001; Vetrie et al., 1993). *BTK* is a member of the Tec family of tyrosine kinases, which also consist of *BMX* (BMX Non-Receptor Tyrosine Kinase), *TEC* (Tec Protein Tyrosine Kinase), *RIK* (RS2-interacting KH Domain Protein) and *ITK* (IL2 Inducible T-Cell Kinase) (Mano, 1999). The *BTK* protein is important for Pre-B cells maturation. *BTK*'s main function is to promote expansion of pre-B cell from preB1 to preB2. Over 800 mutations have been identified in *BTK* (Mohamed et al., 2009). Because of the mutation, development of B cell is affected, which results in significant reduction of mature B lymphocytes levels (< 1%) in the peripheral blood flow. Thus, plasma cells are not developed and subsequently levels of all immunoglobulins classes are markedly low with almost no humoral responses (Noordzij et al., 2002; Ochs & Smith, 1996).

Moreover, the sizes of tonsils and lymph nodes (having high number of B cells) are reduced. Nevertheless, the number and function of T cell not affected (Suri, Rawat, & Singh, 2016). *BTK* also contributes to the control of other innate immune cells, e.g. NK cells (Bao et al., 2012), neutrophils (Honda et al., 2012) and monocytes/macrophages (Koprulu & Ellmeier, 2009).

2.3.1.2 Clinical manifestations in patients with XLA

Recurrent bacterial lower and upper infections of the respiratory tract are the most common sign of XLA. Normally patients with XLA will experience recurrent sino-pulmonary infections, such as sinusitis, otitis media, gastrointestinal and pneumonia infections as well as bronchitis. However, the incidence of these symptoms is variable (Conley & Howard, 2002; Conley, Rohrer, & Minegishi, 2000; Plebani et al., 2002). Usually, encapsulated pyogenic bacteria such as *Streptococcus Pyogenes*, *Streptococcus Pneumoniae*, *Pseudomonas Species* and *Hemophilus Influenzae type B* cause the infections (Lederman & Winkelstein, 1985). Signs of chronic and recurrent sino-pulmonary infections are digital clubbing, postnasal discharge, bronchiectasis and tympanic membrane perforation. The most important clinical signs in confirming the diagnosis are absent or atrophied tonsils and lymph nodes. *Mycoplasma* infections result in joint, urogenital and respiratory infections (Roifman et al., 1986). Gastrointestinal tract infections are also common in these patients. (Chusid, Coleman, & Dunne, 1987; Kerstens et al., 1992). The XLA patients can well tolerate most of the viral infections during childhood because of undamaged function of T cell. However, they are vulnerable to certain enteroviruses such as *Coxsackie*, *Poliovirus* and *Echovirus* virus.

The vaccine-associated polio virus infection has been reported in a child with XLA, who had received oral polio vaccine (Sarpong, Skolnick, Ochs, Futatani, & Winkelstein, 2002). Chronic meningoencephalitis can be caused by *enteroviral* infections

that may lead to subtle neuroregression and turn to full blown neurologic damage and coma (Halliday, Winkelstein, & Webster, 2003). *Enteroviral* infection of skin and muscle can sometimes be mistaken for dermatomyositis-like syndrome, showing peripheral edema and erythematous rash (Rudge et al., 1996). It can even lead to chronic hepatitis exhibiting rash, fever and high hepatic enzymes. Some rare signs are alopecia, glomerulonephritis, von Recklinghausen disease and amyloidosis. About 18% of children with XLA, particularly those with acute infection, experience neutropenia (Jacobs, Guajardo, & Anderson, 2008).

Those with XLA have a lower chance of inflammatory or autoimmune disease than other major immune deficiency diseases. A new web-based survey showed that 69% of patients stated at least once and 53% of them reported several inflammatory symptoms. Yet, only 28% of the participants were officially diagnosed with an inflammatory disease (Hernandez-Trujillo et al., 2014). Moreover, a significant number of patients with XLA showed symptoms of inflammatory bowel disease, arthritis or other inflammatory problems, and these issues were more common than in able-bodied people. Progressive encephalopathy with unidentified etiology has been reported in individuals with XLA, even in patients getting a long-term immunoglobulin replacement therapy (Sag et al., 2014). In the initial phases, subtle cognitive impairment happens together with disorder in the frontal lobe functions. Moreover, some patients suffer from movement disorders, which gradually progress and finally causes severe physical and cognitive disability and may be even lethal. Secondary autoimmunity to irregulated immune reactions is deemed as a pathogenetic mechanism, yet has not been completely clarified (Lee et al., 2010).

2.3.1.3 Diagnostic tests for XLA

A patient is suspected with XLA if having marked reduction in circulating B cells, all serum immunoglobulin levels and liability to recurrent infections with bacterial and

enteroviruses (Lee et al., 2010). Although, family history, physical examination and clinical history may help in diagnosing PIDs, cellular and molecular biology examinations are essential to confirm the diagnosis. As primary screening, the levels of serum immunoglobulins (IgM, IgG and IgA) is measured. The low levels of immunoglobulins would suggest antibody deficits. Then the patients' peripheral blood is subjected to lymphocytes subsets enumeration through flow cytometry. If the lymphocytes subsets enumeration of $CD19^+$ B is below 2%, it would suggest the possibility that the patient is sufficiency from XLA disease (Lee et al., 2010). Furthermore, to confirm the diagnosis of XLA, the genetic examination of *BTK* need to be conducted. Science the patients are having the low number of B cells or in some cases totally lacking of B cells (Tao, Boyd, Gonye, Malone, & Schwaber, 2000), other type of cells that express *BTK* are usually used to study the *BTK* gene expression. Monocyte are reported to express *BTK* and monocytes cell numbers is found to be not affected in XLA patients. Since, the mutant *BTK* expression is low or not expressed in monocytes (Futatani et al., 1998), determination of *BTK* expression in monocytes using flow cytometry examination was developed to ease the diagnosis of XLA (Futatani et al., 1998). The flow cytometric method was reported to succesfully diagnose XLA from *BTK* expression in monocytes (Kanegane et al., 2001; López-Granados, Pérez de Diego, Ferreira Cerdán, Fontán Casariego, & García Rodríguez, 2005). However, some of the individuals with missense mutations in *BTK* expression, do express *BTK* in their B cells and monocytes at the same level with healthy subjects, but the *BTK* protein is functionally defective (Pérez de Diego et al., 2008). Due to these circumstances, genetic examination is needed to support the cellular based examination of *BTK* to confirm the diagnosis of XLA in patients (Ameratunga, Woon, Neas, & Love, 2010; Chear et al, 2013; Hashimoto et al., 1996; Vorechovský et al., 1995; Zhang et al., 2010).

2.3.1.4 Clinical management of XLA

To manage the XLA, immunoglobulin replacement therapy is done through subcutaneous or intravenous route (Gaspar & Kinnon, 2001). Subcutaneous and intravenous routes are safe and have similar efficacy for the deterrence and management of infections as well as early observation for disease problems (Maarschalk-Ellerbroek, Hoepelman, & Ellerbroek, 2011). A study reported that some patients are still vulnerable to infections of respiratory tract, particularly pneumonia in spite of immunoglobulin replacement therapy (Plebani et al., 2002). It can be attributed to inadequate residual serum IgG or deficiency of IgG transport to immunize the mucosal surface site. Chronic infections such as bronchiectasis and chronic sinusitis are effectively treated by a combination of antibiotic prophylaxis and immunoglobulin replacement therapy (Fried & Bonilla, 2009). Other methods such as transplantation of hematopoietic stem cell and lentiviral mediated gene therapy have been introduced to manage XLA, but the risks associated with these procedures can affect quality of life of the patients (Conley et al., 2005; Hendriks, Bredius, Pike-Overzet, & Staal, 2011). Thus, it is important to make an accurate diagnosis before implementing risky actions.

2.3.1.5 BTK and the innate immune system

BTK is mainly expressed in B cells. Many genetic and biochemical studies have shown significant function of *BTK* in B lymphocytes and discussed how *BTK* controls development, differentiation and activation of B cell (Lopez-Herrera et al., 2014; Maas & Hendriks, 2001; Middendorp et al., 2003). In recent years, growing evidence showed that *BTK* is also involved in regulating innate immune functions (Honda et al., 2012; Marron, Martinez-Gallo, Yu, & Cunningham-Rundles, 2012; Marron, Rohr, Martinez-Gallo, Yu, & Cunningham-Rundles, 2010; Sochorova et al., 2007). The significance of *BTK* for macrophage function was first seen in X-linked immunodeficient (XID) mice

infected with microfilaria (Mukhopadhyay et al., 2002). The study reported late microfilaria clearance together with low levels of *IL-12* (Interleukin 12A), *IL-1* and *TNF* production as well as decrease in *NO* production in XID mice (Mukhopadhyay et al., 2002). Similarly, Schmidt and collogues showed that in primary macrophages, *BTK* is activated by *TLR4* and is essential for normal *TLR*-induced *IL-10* production in various populations of macrophage. It has been also demonstrated that, *BTK* plays a crucial role in initiating *TLR3* signaling and in *BTK* deficient macrophages (Schmidt, Thieu, Mann, Ahyi, & Kaplan, 2006). In the absence of *BTK*, *TLR3*-induced *PI3K* (Phosphoinositide 3-Kinase), *AKT* (V-Akt Murine Thymoma Viral Oncogene Homolog 1) and *MAPK* (MAP kinase phosphorylation) signaling and activation of *NFκB*, *IRF3* (Interferon Regulatory Factor 3), and AP-1 transcription factors were all defective. (Lee et al., 2012). Further investigations on the human monocytic THP1 cell line showed interactions of *TLR8* and *TLR9* with *BTK*, in which defective *BTK* lead to impaired *TLR8* and *TLR9* signaling and cause susceptibility of XLA patients to viral infections. (Doyle, Jefferies, Feighery, & O'Neill, 2007). Furthermore, it has been reported that *BTK* contributed in *TLR4* signaling to *NFκB*, and may also involve in signaling by ligands for *TLR2*, *TLR*6, *TLR8*, and *TLR9* and also with *MYD88*, *MAL* and *IRAK1* (Interleukin 1 Receptor Associated Kinase 1) (Horwood et al., 2006; Jefferies et al., 2003)

BTK was also reported to contribute to immune-complex mediated activation of *FcεR* (Fc Epsilon Receptor) and *FcγR* (Fc gamma Receptors) signaling pathways in mast cells and neutrophils (Hata et al., 1998; Jongstra-Bilen et al., 2008). Moreover, the decreased chemotaxis and defective *FcγR* (Fc gamma Receptors), *CR1* (Complement Receptor 1) and *CR3* (Complement Receptor 3)-mediated phagocytosis has been reported in monocytes from XLA patients compared healthy subjects (Amoras et al., 2003). Besides, it was demonstrated that in monocytes of XLA patients, defective *BTK* caused overexpression of *XBP1* (X-Box Binding Protein 1) , a key transcriptional factor for ER

stress and differentiation of plasma cell (Teocchi, Domingues Ramalho, Abramczuk, D'Souza-Li, & Santos Vilela, 2015).

2.4 RNA-Seq: a revolutionary tool for transcriptome study

In all organisms, the genes encoded by DNA is transcribed to RNAs. RNAs can be classified into coding and non-coding RNAs (Birney et al., 2007). Coding RNA comprises of the messenger RNAs (mRNAs) which can further translated into proteins, whereas the non-coding RNAs (ncRNAs) do not encode any proteins and consists of short non-coding and long non-coding RNAs (lncRNAs) (Gomes, Nolasco, & Soares, 2013). Short non-coding RNAs have below 200 nucleotides and comprise of small nucleolar RNAs (snoRNAs), small interfering RNAs (siRNA), piwi-interacting RNAs (piRNAs), and microRNAs (miRNAs) (Lindsay, 2008; O'Connell, Rao, Chaudhuri, & Baltimore, 2010). lncRNAs are the major class of non-coding genes in mammalian genomes, having over 200 nucleotides (Djebali et al., 2012). The transcriptome is a whole set of all RNA transcripts existing in a tissue or cell at a certain point of time under specific conditions (Wang, Gerstein, & Snyder, 2009). Transcriptome study is important for understanding the genome functional elements and the molecular components of cells and tissues, and also for understanding development of diseases. (Wang et al., 2009).

Microarray was commonly used method for transcriptome analysis which is based on the relative quantity of hybridization of complementary deoxyribonucleic acid (cDNA) to the investigations (Baldi & Hatfield, 2002). Microarray examination was used to understand gene expression, but it is biased and incomplete in most cases with failure to recognize novel transcripts (Clark, 2002). Recently, high-throughput RNA sequencing (RNA-Seq) has been introduced as a powerful alternative for large scale profiling of gene expressions (Nagalakshmi et al., 2008). The technology offers excellent base level resolution and genome coverage, which enable well-organized generation of

transcriptomic data in a laboratory setting. RNA-Seq provides a more inclusive insight into the complexity of eukaryotic transcriptomes than older techniques (such as microarray technology) as it can detect the quantitatively and qualitatively of various RNAs, including both mRNAs and ncRNAs, identification of 3` and 5` ends of genes (Tsuchihara et al., 2009), boundary mapping monitoring of allele expression and identification of new alternatives splicing events (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008; Nagalakshmi et al., 2008; Wang et al., 2009; Wilhelm et al., 2008). Several studies have applied RNA-Seq technology for transcriptome profiling of immune-related complement from various issues and cell types such as endometrium (Zieba et al., 2015), spleen (Dang et al., 2016), T cells (Mitchell et al., 2015), B cells (Toung, Morley, Li, & Cheung, 2011) and macrophages (Beyer et al., 2012). Also, by using the RNA-Seq approach, the global gene transcription changes that occur during the differentiation of monocyte to macrophage have been reported (Dong et al, 2013).

In the RNA-Seq approach, RNA sample [fractionated or total, such as poly(A)$^+$] is transformed into a collection of cDNA fragments with adapters binded to one or both ends. Then, every molecule is sequenced with or without amplification in a high-throughput way to obtain short sequences from one end (single-end sequencing) or both end (paired-end sequencing). The reads are normally 30 to 400 bp, according to the used sequencing platforms. The sequencing platforms are grouped as ensemble-based (several matching copies of a DNA molecule, such as SOLiD and Illumina) or single molecule-based (such as PacBio and Helicos) (Chu & Corey, 2012). An example of the steps of an RNA-Seq experiment is illustrated in Figure 2.4.

The raw reads produced from the sequencing platforms undergo quality assessment including trimming adapter sequences and eliminating low quality reads to generate high quality reads for additional computational examination (Wang et al., 2009). The high quality reads are either assembled *De novo* without the reference sequence or

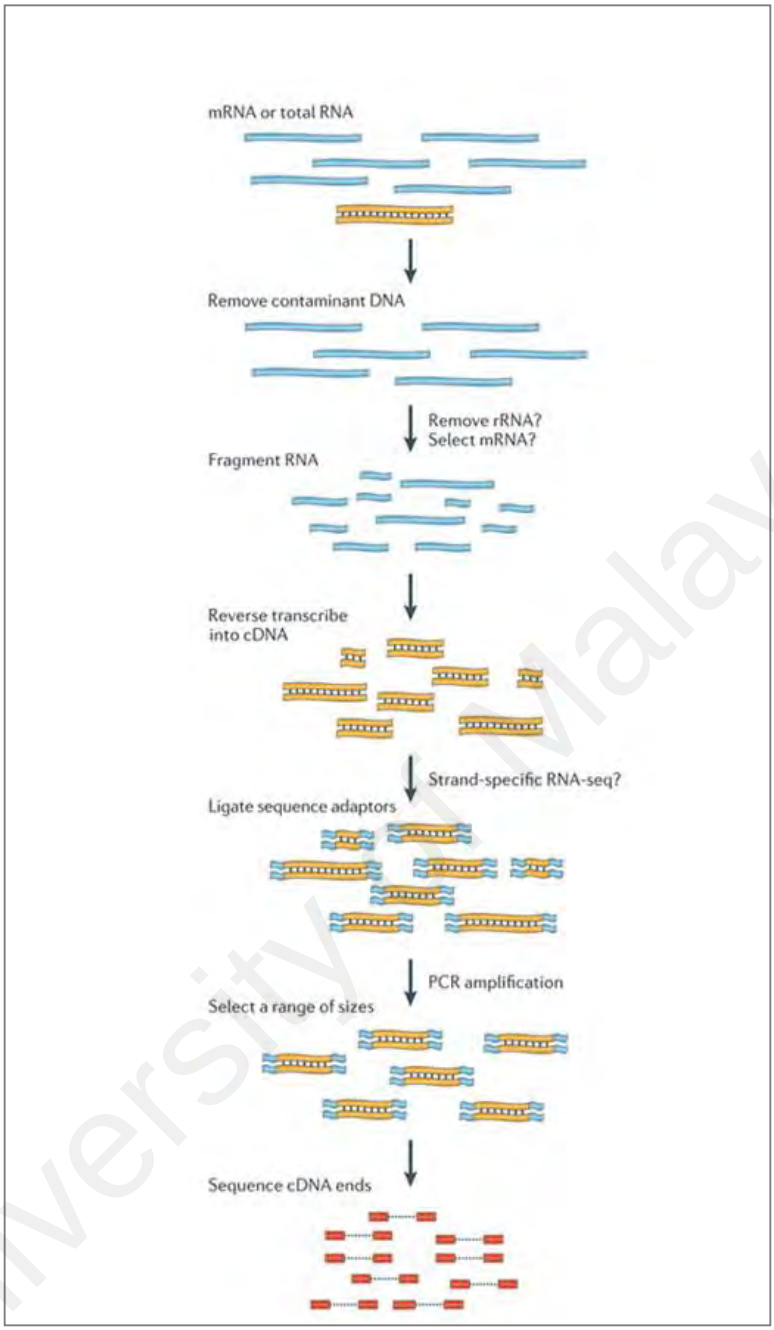

Figure 2.4: The RNA-Seq data generation. To generate an RNA sequencing (RNA-seq) data set, first the RNA (light blue) is extracted from target cell or tissue. DNA contamination is removed and the remaining RNA is cut into small fragments fragments. The RNA fragments are then reverse transcribed into cDNA (yellow), sequencing adaptors (blue) are ligated, and fragment size selection is undertaken. Finally, the ends of the cDNAs are sequenced using next-generation sequencing technologies to produce many short reads (adapted from Martin & Wang, 2011).

aligned to reference transcripts or a reference genome to generate a genome-scale transcriptome map that involves level of expression and/or the transcriptional structure for each gene (Wang et al., 2009). The expression level of the RNAs is defined by the quantity of sequenced fragments mapped to the transcript that is anticipated to associate directly with its level of abundance (Rapaport et al., 2013).

The sequencing depth (number of sequenced reads for a given sample) is an important factor relating to the quality quality of RNA-Seq analysis. Higher sequencing depth seemingly lead to a more precise approximation of the expression level and, alongside, inferential methods get higher power to recognize deferentially-expressed features. As a consequence, our ability to find low low abundance transcripts and detect deferential expression is very much determined by the sequencing depth (Sims, Sudbery, Ilott, Heger, & Ponting, 2014; Tarazona, Garcia-Alcalde, Dopazo, Ferrer, & Conesa, 2011)

2.4.1 Detecting differentially expressed (DE) genes

RNA-Seq can compare levels of gene expression among several samples such as healthy and disease states and two different tissues and detect the genes that have different expression levels between samples (Wang et al., 2009). Previously, microarray was the most commonly used and important approach in such study; however RNA-Seq has been a very powerful substitute method that has many advantages than microarrays, such as a lower background level, higher dynamic range and the ability to identify and measure the expression of formerly unidentified isoforms and transcripts (Agarwal et al., 2010; Bradford et al., 2010; Oshlack, Robinson, & Young, 2010). RNA-Seq data have contributed highly to cancer biology (Byron, Van Keuren-Jensen, Engelthaler, Carpten, & Craig, 2016). The whole transcriptome sequencing of tumor and normal tissues from patients with OSCC (oral squamous cell carcinoma) identified 515 DE genes between

tumors and normal tissues that comprised of genes associated with cell motility, adhesion and differentiation (Zhang et al., 2010). In the case of melanoma, 12 novel transcripts were detected, which might be beneficial for managing skin cancer (Berger et al., 2010). Furthermore, RNA-Seq can use to explore the molecular phenotype in new therapeutics development for inflammatory disease. For instance, RNA-seq has been successful in detecting higher expression of the *IL-36* (Interleukin 36) cytokine and *TREM1* (Triggering Receptor Expressed On Myeloid Cells 1) pathway in individuals having atopic dermatitis (Suárez-Fariñas et al., 2015). Also, in case of infected T cells with low passage HIV isolate, RNA-Seq analysis revealed robust changes in genes expression associated with immune response and patients infected with apoptosis HIV (Sherrill-Mix, Ocwieja, & Bushman, 2015).

2.4.2 Detecting long non-coding RNAs (lncRNAs)

RNA-Seq has significantly speed up the detection and classification of lncRNAs (Bowcock et al., 2001; Ponting, Oliver, & Reik, 2009). lncRNAs are the biggest class of ncRNAs in mammalians, having more than 200 nucleotides length and without coding potential (Djebali et al., 2012). The encoded lncRNAs in the human genome are localized both to the cytoplasm and nucleus. They are either polyadenylated (poly(A)$^{+}$) or non-polyadenylated (poly(A)$^{-}$) and usually have lower levels of expression than protein-coding genes (Guttman et al., 2009). They are classified based on the location of lncRNAs in relation to protein-coding genes as follows: intronic lncRNA, long intergenic non-coding RNA (lincRNA), antisense lncRNA (Cabili et al., 2011; Djebali et al., 2012; Guttman et al., 2009) enhancer RNA (eRNA) (Mousavi et al., 2013; Natoli & Andrau, 2012) sense-overlapping lncRNAs, lncRNAs-host and pseudogenes (Rapicavoli et al., 2013) (Table 2.3; Figure 2.5).

Table 2.3: Classification of human lncRNAs (adapted from Martens-Uzunova et al., 2014).

lncRNAs	Abbreviation	Description
Antisense transcripts	antisense RNA	Reside on the opposite strand of protein-coding genes and intersect their exons
Intronic transcripts	-	Reside within introns of a coding gene but do not intersect any exons
long intergenic non-coding RNA	lincRNA	Originate from protein noncoding genomic regions
Host genes	-	Primary hosts of small ncRNA genes nested within their introns
Enhancer RNAs	eRNAs	Originate at genomic enhancer regions. Boost gene transcription in tissue-specific and temporal manner
Pseudogenes	-	Highly similar to protein-coding genes that have lost their coding potential but can be activated in different tissues or in cancer

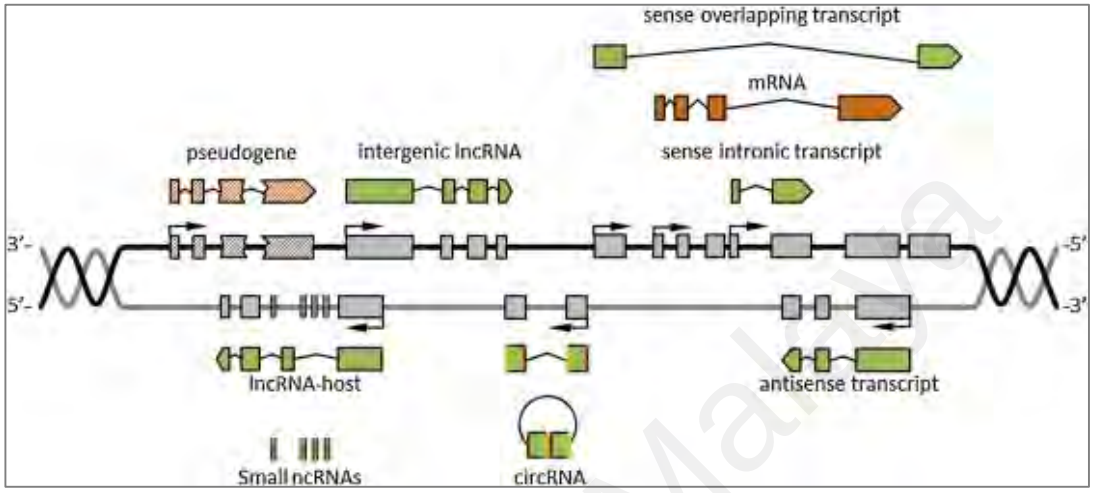

Figure 2.5: Genomic organization of different lncRNAs classes. A black and grey lines represent DNA strands. Grey boxes showed protein-coding genes or lncRNAs genomic exons. Thin black lines represent spliced introns. Arrows indicate direction of transcription. Protein-coding sequences are orange and lncRNAs are green. Pseudogenes have a diagonal stripe pattern. Intron boundaries of circular RNA precursors are shown in red (-5′) and yellow (-3′) (adapted from Martens-Uzunova et al., 2014).

RNA polymerase II often transcribed the lncRNAs. Posttranscriptional processing happens in several lncRNAs, such as polyadenylation, 5′-capping, and alternative splicing (Derrien et al., 2012). LncRNAs show more cell-specific and tissue-specific patterns of expression (Brunner et al., 2012; Gibb, Brown, & Lam, 2011). For example, lncRNA *MIAT* (Myocardial Infarction Associated Transcript) and *SOX2-OT* (SOX2 Overlapping Transcript) are specifically expressed in brain and while *H19* (H19, Imprinted Maternally Expressed Transcript*)* is highly expressed in the placenta (Derrien et al., 2012). lncRNAs control the expression of protein-coding genes via control of cellular or molecular processes; including protein localization, mRNA stability and epigenetic regulation of transcription (Karapetyan, Buiting, Kuiper, & Coolen, 2013).

2.4.2.1 lncRNAs in immune system

Immune cells development and activation depend on dynamic and integrated gene expression programs that are controlled by transcriptional and post-transcriptional processes. Our knowledge of the protein-coding genes functional roles in the transcriptional, posttranscriptional, regulation of gene expression is fairly well developed. However, the regulatory roles of lncRNAs have yet to be fully established (Atianand & Fitzgerald, 2014). Several types of lncRNAs exist in immune cells, including B cells, macrophages, monocytes, neutrophils, dendritic cells, and T cells (Geng & Tan, 2016). For example the development, differentiation as well as activation of immune cells said to be associated with lincRNAs expression levels (Atianand & Fitzgerald, 2014). Hence, the functional variety of these lncRNAs should be highlighted. Guttman et al. (2009) reported the first proof of lncRNAs' role in the innate immune response. They found *lincRNA-Cox2* that was induced in $CD11c^{+}$ dendritic cells once stimulated with LPS. It was also shown that *lincRNA-Cox2* highly controls innate immune gene expression in both negative and positive ways. The negative control happens via interactions of

lincRNA with *hnRNP* (Heterogeneous Nuclear Ribonucleoprotein)-*A*/*B* and *hnRNPA2*/*B1* to reduce expression of interferon stimulatory genes (like *IFN-β*, *IFNα* and *CCL5*) and chemokines in the murine macrophages. Moreover, *lincRNA-Cox2* has positive effect on the expression of *IL-6* and many more inducible immune genes through unknown mechanisms (Guttman et al., 2009). Another study examined the expression of monocytes lncRNAs in response to *TLR4* signaling through LPS stimulation and reported 221 differentially expressed (DE) lncRNAs (IIott et al., 2014). Recently, differential expression of lincRNAs was also demonstrated following triggering the innate immune reaction in THP1 macrophages. It was revealed that 159 lincRNAs (e.g. *THRIL* (TNF And HNRNPL Related Immunoregulatory Long Non-Coding RNA)) control *TNFα* expression via communication with hnRNPL (Li et al., 2014). Other lncRNAs related to innate immunity are *NEAT1* (Nuclear Paraspeckle Assembly Transcript 1), *PACER* (PTGS2 Antisense NFKB1 Complex-Mediated Expression Regulator RNA) and *Rps15a-ps4* (renamed *Lethe*) , which are involved in regulating immune cell functions (Imamura & Akimitsu, 2014; Zhonghan Li & Rana, 2014) and immune gene expression (Carpenter & Fitzgerald, 2015). The role of lncRNAs in development, differentiation and activation of lymphocytes in the adaptive immune responses has also been revealed by. Ranzani et al. (2015). Ranzani and colleagues were identified the total of 563 lincRNAs in subsets of B and T lymphocytes. They also reported the regulatory function of lincRNA *linc-MAF-4* (MAF Transcriptional Regulator RNA) in differentiation of T cells (Ranzani et al., 2015). In addition, Hu et al. (2013) detected 1,524 clusters of lincRNAs in samples of T cells, ranging from primary T cell progenitors to last differentiated helper subsets of T cell. Their analysis showed cell-specific and extremely dynamic lincRNAs expression patterns during differentiation of T cell.

2.4.2.2 lncRNAs and immune-related diseases

There is a link between immune-related diseases and lncRNAs (Geng & Tan, 2016). In rheumatoid arthritis, for instance, it was revealed that *HOTAIR* (HOX Transcript Antisense RNA) and *H19* are upregulated (Song et al., 2015). Genetic evidence suggests that the susceptibility for diabetes are associated with genomic regions for several lncRNAs, such as lncRNAs *MEG3* (Maternally Expressed 3) and *IGF2-AS* (IGF2 Antisense RNA) (Lee et al., 2005; Wallace et al., 2010). Recently, lncRNAs were shown to contribute in many immune-related disorders, such as rheumatoid arthritis, psoriasis, inflammatory bowel disease, primary sclerosing cholangitis, celiac disease, primary biliary cirrhosis and juvenile idiopathic arthritis (Hrdlickova et al., 2014). Shi et al. (2014) reported abnormal expression of few candidate lncRNAs in individuals with systemic lupus erythematosus through examination of transcriptome profiles in patients' peripheral blood mononuclear cells. Furthermore, the relation between irregular lncRNAs expression and inflammatory diseases, such as inflammatory bowel disease and obstructive pulmonary disease has been reported (Bi et al., 2015).

In this study, deep RNA-Seq analysis (200 million reads per sample) was carried out on a purified population of primary monocytes from 6 healthy subjects and 3 XLA patients. Using this datasets, comprehensive gene expression profile of human primary monocytes under healthy and disease states were generated. The differential gene expression analysis was also performed between healthy male and female subjects as well as healthy male subjects and XLA patients to look into possible differences in expression patterns of immune-related genes in healthy and disease conditions. Our deep RNA-Seq analysis allowed us to profile many transcripts that are expressed at low levels such as novel transcripts and lncRNAs which their detection requires deep sampling of the transcriptome (Sims et al., 2014). Also, our deep RNA-Seq also effectively increased the

power and accuracy of differentially genes expresstion identification in both conditions. (Tarazona et al., 2011).

CHAPTER 3: METHODOLOGY

3.1 Samples collection

Ethics approval to conduct this study was obtained from Medical Research and Ethics Committee (MREC), Malaysia with reference number NMRR-13-972-16921. The approval allows using of peripheral blood from healthy subjects and XLA patients.

3.1.1 Healthy subjects

Six healthy, non-blood related subjects with urban lifestyle (age range of 25 to 39 years) were selected in this study. All subjects fulfilled the criteria's set for the study: non-smokers, not having any medical illness, not prescribed any chronic medication, and not receiving any vaccination at least 6 months before the study. Each subject completed a consent form provided prior to the experiment.

3.1.2 X-linked agammaglobulinemia (XLA) patients

Three male patients (age range of 12 to 18 years) whom have been diagnosed as having XLA diseased by the PID unit of the Allergy and Immunology Research Centre (AIRC), Institute for Medical Research, Malaysia were selected for this study. The XLA patient's clinical history and cellular and molecular diagnostic tests results were traced back from the test record book of AIRC. The patients were diagnosed to have XLA disease based on the criteria set by the World Health Organization (WHO) scientific group for PIDs. The criteria are: low levels of circulating B cells (measured by levels of $CD19^+$ B cells in blood samples), reduced or absent of immunoglobulins in serum and a typical clinical history with recurrent bacterial infection or a positive family history (Anker, 2007). The monocytes *BTK* expression level from the XLA patients were evaluated using flow cytometry which showed a *BTK* downregulation for all the patients. The *BTK* gene mutations were detected by directly sequencing the amplified PCR

products which revealed the *BTK* gene mutations in the patients. Table 3.1 shows the history of the patient's serum immunoglobulin levels before receiving intravenous human immunoglobulin (IVIG) therapy, and the nucleotide change that occurred in each patient and its consequences in the protein synthesis. All the patients were in a stable clinical situation without fever and not hospitalized. They were under IVIG therapy monthly. Blood samples were collected before the administration of IVIG from the patients.

3.2 Isolation of the Peripheral Blood Mononuclear Cells (PBMCs)

A volume of 10 ml of peripheral blood from each healthy subjects and XLA patients were collected in Ethylenediaminetetraacetic acid (EDTA) glass tubes. The blood from each subjects was divided into 2 Falcon tubes (5 ml each tube), diluted with 5 ml of Phosphate-Buffered Saline buffer (PBS) and mixed well. A volume of 5 ml diluted cell suspension was layered over 3.5 ml of Ficoll-Paque in 15 ml conical tube and centrifuged at 400 g for 30 minutes, at 18 °C without acceleration and brake. After centrifugation, the upper layer was aspirated and the peripheral blood mononuclear cells (PBMCs) layer was collected into a fresh conical tube using pasteur pipette. The conical tube was filled with PBS buffer, mixed well and centrifuged at 300 g for 15 minutes at 18 °C with acceleration 5 and brake 5. The supernatant was removed and the cell pellet was resuspended with PBS buffer and centrifuged at 200 g for 15 minutes at 18 °C with acceleration 5 and brake 5. Then, the supernatant was removed and the cell pellet was resuspended with 400 µl of Bovine Serum Albumin (BSA) buffer and transferred into the 1.5 ml microfuge tube.

3.3 Isolation of monocytes

The classical monocytes ($CD14^{++}CD16^{-}$) were isolated from PBMCs using Monocyte Isolation Kit II (Miltenyi Biotec, Germany) with protocol that optimized in this

Table 3.1: Clinical and immunological data of patients with XLA.

					Ig levels at diagnosis (mg/dL)					Mutations		
Patient	**Age (years)**	**Age at onset (years)[i]**	**Age at diagnosis (years)[ii]**	**Family history[iii]**	**IgG**	**IgM**	**IgA**	**CD19+ (%)**	**BTK expression[iv]**	**Nucleotide**	**Protein**	**Protein Domain**
P1	12	1	4	−	N/A	N/A	N/A	1 (12−22)	7.7%	*c.1888A>T	*p.M630L	Kinase
P2	13	1	6	+	41(550−1200)	<12(40−95)	48 (60-170)	0 (12−22)	6%	IVS9+1G>C	Skipping of exon 9	SH3 & SH2
P3	18	2	7	−	91.1 (550−1200)	11.3 (40−95)	UD (60−170)	0 (12−22)	0.04%	*g.34430_34447 delCAAAGTCAT GATgtgagt	*p.A446_N451 ins(28 amino acids)	Kinase

N/A, not available
UD, undetectable
i, Age at the which an individual acquires, develops, or first experiences a condition or symptoms of a disease
ii, Age at the start of intravenous immunoglobulin replacement.
iii "+", indicates that family members [boy (s)] died at a young age because of infection.
iv, Normal expression is > 94%
*c, Coding DNA references sequences
*g, Genomic references sequences
*p, Protein references sequences

study. The kit contains: FcR Blocking Reagent (human Ig), Monocyte Biotin-Antibody Cocktail (Cocktail of biotin-conjugated monoclonal antibodies against CD7, CD16, CD56, CD19, CD3, CD123 and Glycophorin A) and Anti-Biotin MicroBeads (MicroBeads conjugated to a monoclonal antibiotin antibody). Using this kit, the monocytes were isolated from PBMCs by negative selection method through indirect magnetic labeling system. The non-monocyte cells including T cells, B cells, dendritic cells, NK cells, and basophils were indirectly magnetically labeled using a cocktail of biotin-conjugated antibodies and AntiBiotin MicroBeads. The monocytes were then isolated by depletion of the magnetically labeled cells.

After isolation of PBMC, the cell suspension was centrifuged at 300 g for 10 minutes at 8 °C. The supernatant was completely removed, and cell pellet was resuspended in 45 µl of BSA buffer. A volume of 15 µl of FcR Blocking Reagent and 15 µl of Biotin-Antibody Cocktail were added to the cells suspension, mixed well and incubated for 15 minutes at 4−8 °C. Then, another 45 µl of BSA buffer with 30 µl of Anti-Biotin MicroBeads per 10^7 total cells were added to the cell suspension, mixed well and incubated for additional 20 minutes at 4−8 °C. The cells were washed by adding 1 ml of BSA buffer and centrifuged at 300 g for 10 minutes at 8°C. The supernatant was removed completely and the cells resuspended in 500 µl of BSA buffer. The magnetic separation of monocytes from non-monocytes cells was performed using LS MACS column and MACS separator with 3 times wash in 500 µl of BSA buffer. The purified monocytes were not stimulated by any factors during the isolation process from the peripheral blood, therefore, the preparation can be considered as primary monocytes.

3.4 Evaluation of monocyte purity using flow cytometry

The purity of the enriched monocytes was evaluated by flow cytometry from 2 steps of monocyte isolation including before magnetically labeling the non-monocytes

cells (unstained) and after magnetically separation of monocytes from non-monocytes cells through negative selection (negative). A volume of 20 µl of cell suspension from each step was kept separately in 2 tubes for flow cytometry analysis. A volume of 80 µl of BSA buffer with 10 µl of FITC anti-human CD14 Antibody (BD Pharmingen^{TM}, USA) and 10 µl of PE-conjugated anti-human CD14 (BD Pharmingen^{TM}, USA) were added into each tubes and mixed well. The tubes were then incubated for 15 minutes at room temperature, in the dark. After incubation, 1 ml of BSA buffer was added into each tubes, centrifuged at 300 g for 15 minutes at 18 °C with acceleration 5 and brake 5. The supernatant was removed and the cells pellet was resuspended to 1 ml of BSA buffer. The purity of monocytes from each tube was evaluated using BD FacsCanto II Flow Cytometer (BD Biosciences, USA).

3.5 RNA extraction

Total RNA was extracted from purified monocytes using the RNeasy Mini-Kit (QIAGEN, Germany) according to manufacturer's protocol. The monocytes cell suspension was centrifuged at 300 g for 10 minutes at 8 °C. The supernatant was removed and the cell pellet diluted with 600 µl of Buffer RLT containing 10 µl of β-Mercaptoethanol (β-ME). The sample was then transferred onto a QIAShredder column sitting in 2 ml collection tube and centrifuged at 14000 rpm for 2 minutes to homogenize the cells. A volume of 600 µl of 70% ethanol was added into the 2 ml QIAshredder collection tube and mixed well. The total volume in the tube was 1200 µl. A volume of 600 µl of the sample was pipetted into RNeasy mini column sitting in a 2 ml collection tube and centrifuged at 1000 rpm for 15 seconds. The flow through was discarded. The remaining 600 µl of the sample in the QIAshredder collection tube was added into the RNeasy mini column. The sample was centrifuged at 1000 rpm for 15 seconds. The flow through was discarded and 700 µl of Buffer RW1 was added to the column and

centrifuged at 1000 rpm for 15 seconds. The flow through was discarded and 500 µl of Buffer RPE was added to the column and centrifuged at 1000 rpm for 15 seconds. The flow through was discarded and another 500 µl of Buffer RPE was added to the column and centrifuged at 1000 rpm for 2 minutes. The flow through was discarded and the RNeasy columns was transferred into the fresh 2 ml collection tube and centrifuged at 14000 rpm for 1 minutes. The RNeasy column was then transferred into fresh 1.5 ml nucleic acid free microfuge tube. A volume of 40 µl of RNase free water was directly added onto a column and centrifuged at 1000 rpm for 1 minute. To increase the RNA concentration, the eluate from 1.5 ml nucleic acid free microfuge tube was pipetted again into the column and centrifuged at 1000 rpm for another 1 minute.

3.6 RNA library preparation and sequencing

The quality and quantity of extracted RNA were measured using NanoDrop 2000 (Thermo Fisher Scientific Inc., USA) and Qubit 2.0 RNA Broad Range Assay (Invitrogen, USA). To determine the RNA Integrity Number (RIN) of extracted RNA, the RNA samples were run on Agilent Bioanalyzer RNA Nano Assay chip (Agilent Technologies, USA). A 1 µg volume of RNA (as measured by Qubit) with RIN number of 8 and above was used for sequencing. Messenger RNA (mRNA) isolation and complementary deoxyribonucleic acid (cDNA) synthesis were performed using TruSeq RNA Sample Preparation Kit (Illumina, USA) and SuperScript II Reverse Transcriptase (Invitrogen, USA) according to manufacturer's protocol. The synthesized cDNA was quantified using Qubit 2.0 DNA Broad Range Assay (Invitrogen, USA). A minimum of 40 ng cDNA was fragmented using Covaris S220 (Covaris Inc, USA) to a targeted size of 200−300 bp. The fragmented cDNAs were then end-repaired, ligated to Illumina TruSeq adapters, and PCR-enriched using TruSeq RNA Sample Preparation Kit (Illumina, USA) according to manufacturer's protocol. The final sequencing libraries

were quantified using KAPA kit (KAPA Biosystem, USA) on Agilent Stratagene Mx−3005p quantitative PCR (Agilent, USA) and sizes were confirmed using Agilent Bioanalyzer High Sensitivity DNA Chip (Agilent, USA). The resulting libraries were sequenced using an Illumina flow cell, with 209 cycles on the Illumina HiSeq 2000 platform (Illumina, USA). A Schematic representation of the experimental protocol used to isolate the RNA from all samples is shown in Figure 3.1.

3.7 RNA-Seq datasets

The RNA-Seq of 6 healthy subjects (3 male and 3 female) and XLA patients (3 male) generated in this study (Table 3.2) were used for data analysis and profiling the gene expression patterns of primary monocytes under healthy and XLA disease states.

3.8 Transcriptome profile of primary monocytes from healthy subjects

3.8.1 Alignment to genome and transcript assembly

The quality of all sequences reads from healthy subject's samples (Table 3.2) were assessed using FastQC (Andrews, 2010). The adaptors and low quality bases were trimmed from the sequences using Trimmomatic (Bolger, Lohse, & Usadel, 2014). For each sample, the standard score (mean and median) at base across reads was $Q > 20$. Quality trimmed raw reads from all samples were separately aligned to the human reference genome sequence (GRCH38.79) using HISAT (version 0.1.4) (Kim, Langmead, & Salzberg, 2015) with GENCODE junctions as a guided reference annotation (version 22). The aligned reads (BAM files) for each sample were assembled into transcripts by StringTie (version 1.3.3) (Pertea et al., 2015) using a GENCODE reference annotation GTF file (version 22) and separate GTF files were generated for each of the samples. The transcripts abundance was estimated as Fragments Per Kilobases of exon per Million fragments mapped (FPKM) (Trapnell et al., 2010).

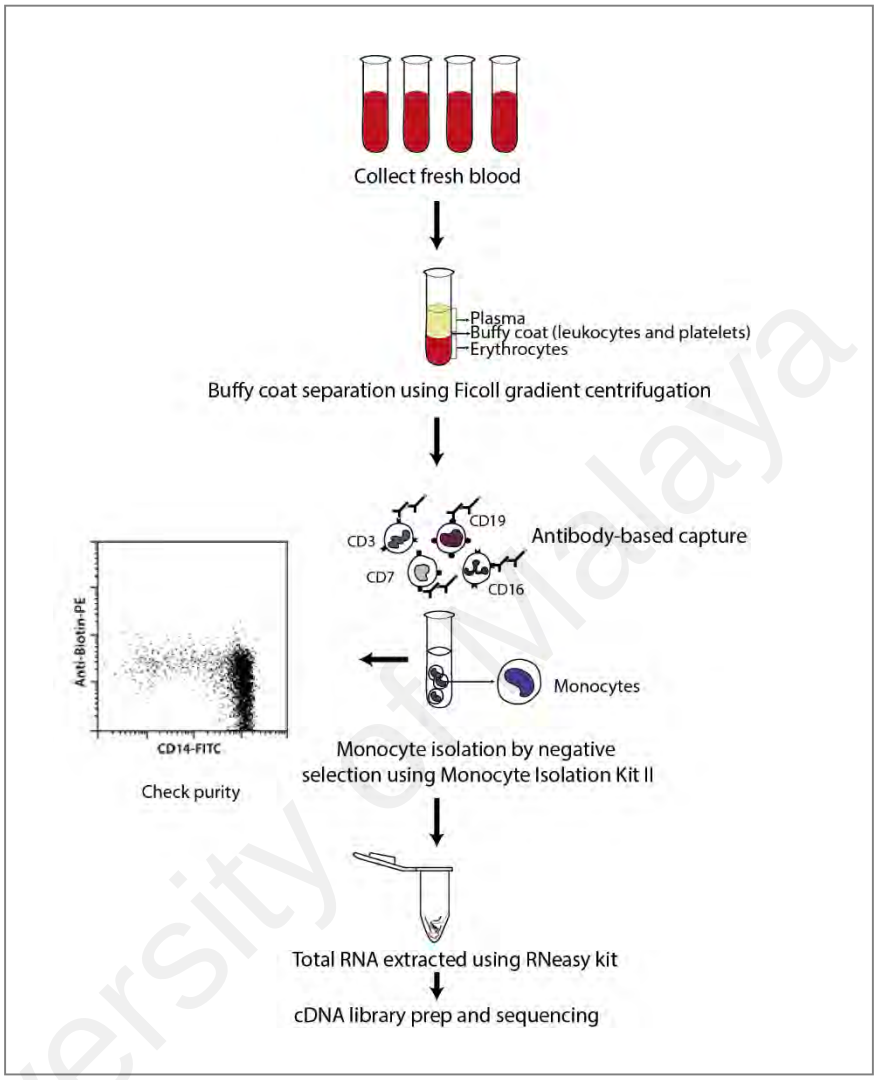

Figure 3.1: Schematic representation of experimental protocol used for isolation of RNAs prior to sequencing. The peripheral blood was collected from each subjects. The PBMC was isolated by density gradient centrifugation. The classical monocytes ($CD14^{++}CD16^{-}$) were isolated from PBMCs using a negative selection technique with Monocyte Isolation Kit II. RNA was extracted from purified monocytes using RNeasy kit and poly$(A)^{+}$ paired-end RNA-Seq was performed on the purified RNA.

Table 3.2: Summary of RNA-Seq datasets generated from primary monocytes.

Sample ID	Gender	Total number of reads
Healthy Subject_01	Male	283,190,913
Healthy Subject_02	Male	226,114,167
Healthy Subject_03	Male	204,978,859
Healthy Subject_04	Female	188,707,164
Healthy Subject_05	Female	199,019,074
Healthy Subject_06	Female	197,734,796
Patient_01	Male	149,591,888
Patient_02	Male	172,557,427
Patient_03	Male	155,750,350

3.8.2 Gene expression profiling

For detecting gene expression pattern in primary monocytes of healthy subjects, transcript assemblies (GTF files) of all samples were merged together to form a single set of non-redundant transcripts using Cuffmerge (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010). Cuffquant (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010) was used to quantify the expression levels of transcripts and to create individual binary files (CXB format). The Cuffnorm (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010) was used to normalize FPKM between the samples. The FPKM >0.1 threshold was used to determine expressed transcripts. The merged assembly was then compared with a GENCODE reference annotation GTF file (version 22), which contains protein-coding genes, non-coding genes, pseudogenes and, with their alternatively transcribed variants. From the comparative analysis results, the transcripts which were intergenic and not aligned to the reference annotation were considered as putatively novel transcripts. These transcripts were filtered against the non-redundant (NR) database from NCBI using Basic Local Alignment Search Tool for Nucleotide (BLASTN) (version 2.4.0) with an E-value <1e-10 threshold. TransDecoder (version 3.3.0) (http://transdecoder.github.io) were then used to identify potential novel transcripts coding for peptides [transcripts with open reading frames (ORFs)]. To further capture the ORFs that may have significant functions, the potential novel transcripts predicted with ORFs were searched against PFAM-A database (Finn et al., 2016) and the results obtained were filtered with an E-value <1e-10 threshold. To further investigate the biological importance of the identified protein-coding genes as well as detection of the immune-related genes which were expressed in primary monocytes, DAVID (Database for Annotation, Visualization and Integrated Discovery) functional annotation tool (Dennis et al., 2003) was used to perform Gene Ontology (GO) enrichment analysis (Ashburner et al., 2000). DAVID uses modified Fisher's exact test (known as EASE) (Hosack, Dennis, Sherman, Lane, & Lempicki,

2003) to measure enrichment against a background gene list and adjusting the resulting p-values (adjP) using a Benjamini-Hochberg method (Benjamini & Hochberg, 1995). The GO analysis was restricted to the category of biological processes as it is the most prominent for evaluation of genes' functions. The highest hierarchical level 1 of biological process ontology (DAVID category-GOTERM_BP_1) was chosen to provide a general description for genes' functions. The minimum number of genes for enrichment in each category was set at 2 and the significance cutoff was adjP < 0.01. Subsequently, the pathway analysis of identified protein-coding genes was conducted by applying the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000) in WEB-based GeneSet Analysis Toolkit (WebGestalt) (Wang, Duncan, Shi, & Zhang, 2013) using the Hypergeometric statistical test followed by the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) for adjusting the p-value (adjP).

3.8.3 Identification of differentially expressed (DE) genes between healthy male and female subjects

To identify the differentially expressed (DE) genes between healthy male and female subjects, Cuffdiff (a part of the Cufflinks, version 2.2.1) (Trapnell et al., 2010) was used to perform differentially gene expression analysis. In order to run Cuffdiff, the merged assembly file from male and female samples was used as the reference annotation and individuals CXB files (Cuffquant outputs) were used as input files. The Cuffdiff applies Benjamini-Hochberg method (Benjamini & Hochberg, 1995) to compute the false discovery rate (FDR) from the p-values (reported as q-value in cuffdiff). The DE genes were identified with q-value ≤ 0.01 and $\log_2$ fold-change ≥ 1 or ≤ -1. The GO analysis was conducted on identified DE genes using DAVID software (Dennis et al., 2003) by screen out the highest hierarchical level 1 of biological process ontology (DAVID category-GOTERM_BP_1). From the GO analysis results, the DE genes which related

to the immune system were selected for further analysis. The hierarchical clustering of the expression profiles of detected DE immune-related genes was generated with the heatmap function in the "NMF" R package (Gaujoux & Seoighe, 2010) using Pearson Correlation. In order to observe specific GO terms for DE immune-related genes, the DAVID category-GOTERM_BP_FAT was selected for displaying the results. Subsequently, pathway analysis of DE immune-related genes was conducted by applying KEGG database (Kanehisa et al., 2004) in WebGestalt (Wang, Duncan, Shi, & Zhang, 2013).

3.8.4 Gene interaction network construction

In order to predict the DE immune-related genes interaction, interaction networks were constructed on the altered genes using Cytoscape plug-in GeneMANIA (Montojo et al., 2010) by applying the information from the co-expression category from GeneMANIA.

3.8.5 Validation of RNA-Seq results by qRT-PCR

The expression levels of selected DE immune-related genes in male compared to female subjects were further validated through Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) analysis using the same set of total RNA from RNA-Seq experiment. First-strand cDNA was synthesized from 300 ng of RNA from each sample by using the High Capacity RNA to cDNA Kits (Applied Biosystems, USA). All the primers and probes for Taqman® Real-time PCR (Life Technologies, USA) were designed by Applied Biosystems as depicted in Table 3.3. The expression of target genes was assessed using the QuantStudio™ 12K Flex Real-Time PCR System. The PCR cycles conditions were: 50 °C for 2 minutes, and 95 °C for 20 second, followed by 40 cycles of 95 °C for 3 second, and 40 cycles of 60 °C for 30 second.

Table 3.3: List of selected DE immune-related genes for validation by qRT-PCR with their respective assay IDs.

Target ID	Assay ID
ACTB[*]	Hs01060665_g1 (Taqman gene expression Assay, Invitrogen)
PPIA[*]	Hs04194521_s1 (Taqman gene expression Assay, Invitrogen)
STAT1[*]	Hs01013996_m1 (Taqman gene expression Assay, Invitrogen)
JUN[*]	Hs01103582_s1 (Taqman gene expression Assay, Invitrogen)

*, Pre-designed qRT-PCR primers manufactured by Invitrogen, USA.

Each gene was analyzed in triplicate in each sample. *ACTB* (Actin, Beta) and *PPIA* (Peptidylprolyl Isomerase A) genes were used as endogenous controls. Fold-changes in gene expression between samples were calculated using $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001). For validation, qPCR derived $\log_2$ fold-changes values were compared with $\log_2$ fold-change values obtained through RNA-Seq analysis.

3.9 Transcriptome profile of primary monocytes from XLA patients

3.9.1 Alignment to genome and transcript assembly

The same pipelines as described in sections 3.8.1 were used to check the quality of RNA-Seq datasets of XLA patients (see Table 3.2, page 46), followed by alignment to a reference genome and constructing the assembled transcripts.

3.9.2 Gene expression profiling

To profile the gene expression patterns of primary monocytes from XLA patients, the same analysis as mentioned in sections 3.8.2 was conducted.

3.9.3 Identification of long non-coding RNAs (lncRNAs)

3.9.3.1 Annotated lncRNAs

To detect the expression of annotated lncRNAs in primary monocytes of XLA patients, the merged assembly file from 3 patient's samples was compared to annotated lncRNAs reference annotation from GENCODE (version 22). The transcripts which had an FPKM >0.1 in at least 1 sample were considered to be expressed.

3.9.3.2 Novel lincRNAs

A multi-step mapping and filtering criteria were employed to identify putative novel long intergenic non-coding RNAs (lincRNAs) in primary monocytes of XLA

patients. After aligning the samples and assembling them, the transcripts with 3 × coverage and above were filtered from individual GTF files. The filtered transcript files were merged to form a non-redundant set of transcripts using Cuffmerge (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010). Cuffquant (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010) was used to quantify the expression levels of each of the above transcripts and to create individual binary files (CXB format). The expression levels were calculated in FPKM after normalizing for total number of reads using Cuffnorm (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010). The transcripts were mapped to the known gene-annotation from GENCODE (version 22) and categorized as protein-coding, non-coding, pseudogenes, and novel loci. Further analysis was conducted on the novel transcripts which were intergenic to GENCODE transcript. All single exon and those that are less than 200 nucleotides were filtered out. The subset of transcripts (multi-exonic, longer than 200 nucleotides) were used for assessing coding potential. The coding potential for these transcripts was evaluated by Coding Potential Assessment Tool (CPAT) (Wang et al., 2013). The coding probability cut-off of 0.375 was used to detect any putative protein-coding sequences and exclude them from the analysis. The final set of filtered transcripts constituted the set of putative novel lncRNAs, which are addressed as lincRNAs. The nucleotide sequences of putative novel lincRNAs were searched for matched sequences against the non-redundant (NR) database from NCBI using BLASTN (version 2.4.0).

3.10 Identification of differentially expressed (DE) genes between XLA patients and healthy subjects

To identify the differences in gene expression patterns of primary monocytes between XLA patients and healthy subjects, our analysis focused on the protein-coding genes as well as lncRNAs. The transcript assembly files (GTF files) of 3 XLA patients

and 3 healthy male subjects (see Table 3.2, page 45) was merged together to form a single set of non-redundant transcripts using Cuffmerge (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010). Differential gene expression analysis was performed by employing Cuffdiff (a part of cufflinks, version 2.2.1) (Trapnell et al., 2010) using merged assembly file from XLA patients and healthy subjects as the reference annotation and individuals CXB files (Cuffquant outputs) as input. The DE genes were identified with q-value ≤ 0.01 and $\log_2$ fold- change ≥ 1 or ≤ -1. Hierarchical clustering of the expression profiles of identifed DE protein-coding genes and DE lncRNAs were done with the heatmap function in the R "NMF" package (Gaujoux & Seoighe, 2010) using Pearson Correlation. The functions of identified DE protein-coding genes were investigated through GO and KEGG pathway analysis using the same method as described in section 3.8.3

3.10.1 Co-location and co-expression analysis of DE lncRNAs and DE protein-coding genes

lncRNAs are presumed to regulate the expression of their neighboring genes (co-located genes) (Ørom et al., 2010). To predict the function of DE lncRNAs based on their co-located genes, the genomic coordinates of DE lncRNAs were imported to the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) to identify DE lncRNAs co-located genes. The functional enrichment analysis of DE lncRNAs co-located and co-expressed genes was performed by using DAVID software (Dennis et al., 2003). The identified DE lncRNAs co-located genes were then matched with DE protein-coding genes to obtain DE lncRNAs co-located and co-expressed genes.

3.10.2 Gene interaction network construction

In order to predict the DE protein-coding genes interaction networks, the genes interaction networks were constructed using Cytoscape plug-in GeneMANIA (Montojo

et al., 2010) for genes which were enriched in top significant upregulated and downregulated KEGG pathways (adjP <0.01). The network was generated by using the information from the co-expression category. In addition, genes interaction network between two subgroups of genes (DE lncRNA co-located genes and DE lncRNA co-expressed genes) were also constructed using Cytoscape (Shannon et al., 2003). The possible interaction between DE protein-coding genes in each subgroup was predicted using the Cytoscape plug-in GeneMANIA.

3.10.3 Validation of RNA-Seq results by qRT-PCR

The expression levels of selected DE protein-coding and DE lncRNAs in XLA patients compared to the healthy subjects were further measured by qRT-PCR. The same set of total RNA from RNA-Seq experiment was used for validation. A volume of 300 ng of the extracted RNA from each sample was converted to the first-strand cDNA using the High Capacity RNA to cDNA Kits (Applied Biosystems, USA). The primers and probes for Taqman® Real-time PCR (Life Technologies, USA) were designed by Applied Biosystems as presented in Table 3.4. The expression of target genes was assessed using the QuantStudio™ 12K Flex Real-Time PCR System as described in section 3.8.5. The *PPIA* (Peptidylprolyl Isomerase A) gene was used as endogenous control.

3.11 Gene catalogue and lncRNAs landscape in human primary monocytes

3.11.1 RNA-Seq datasets

In order to generate a comprehensive gene catalogue and also landscape of lncRNAs expression in human primary monocytes, the other publically available RNA-Seq datasets for monocytes from healthy subjects were added to the RNA-Seq datasets healthy subjects generated in this study (see Table 3.2, page 46). The raw RNA sequences of 10 human classical monocytes with 50 and 100 bases long, paired-end, and sequenced

Table 3.4: List of selected DE protein-coding genes and DE lncRNAs for validation by qRT-PCR with their respective IDs primer sequences/assay.

Target ID	Forward Primer (5`- 3`)	Reverse Primer (5`- 3`)
TCONS_00041961[*]	GTGCTTGCCCAGTGTCTCT	CCGGTCCCGCTGTAGAATTTATT
TCONS_00295657[*]	TCTCTTTCATACCGTTTCATTCCTAAGG	ATTTTATCATGCACCTCATTTCCAGTAGT
TCONS_00298577[*]	GATTTCCAGCGACTTTGTCAACAC	GGGAAAGACAAGGCAGTAAAGACAT
GAS5[*]	CAACTTGCCTGGACCAGCTTA	CCTTACCCAAGCAAGTCATCCAT
RMRP[*]	CGCTGTATGGGAACCTGCATTAT	TCTGGCTCTGGGTCTTGAGA
LINC-PINT[*]	CGAGGCAAGGAGCTAAAGCA	CCCAACTCTTCTAACTCGTAAAAGCA
HEIH[*]	TGGGATTTTCCAACCTTGAGATTCT	GCCAGAGACTTGAAAGGAAGCT
DANCR[#]	Hs03653830_g1 (Taqman gene expression Assay, Invitrogen)	
HOTAIRM1[#]	Hs03296533_g1 (Taqman gene expression Assay, Invitrogen)	
SOD1[#]	Hs00533490_m1(Taqman gene expression Assay, Invitrogen)	
TUG1[#]	Hs04404516_m1 (Taqman gene expression Assay, Invitrogen)	
PPIA[#]	Hs04194521_s1 (Taqman gene expression Assay, Invitrogen)	
FCGR2A[#]	Hs01013401_g1 (Taqman gene expression Assay, Invitrogen)	
ATP5D[#]	Hs00961521_m1 (Taqman gene expression Assay, Invitrogen)	
CXCR2[#]	Hs01891184_s1 (Taqman gene expression Assay, Invitrogen)	
BAX[#]	Hs00180269_m1(Taqman gene expression Assay, Invitrogen)	
TLR1[#]	Hs00413978_m1 (Taqman gene expression Assay, Invitrogen)	
TLR5[#]	Hs01920773_s1 (Taqman gene expression Assay, Invitrogen)	
UQCRB[#]	Hs01890823_s1 (Taqman gene expression Assay, Invitrogen)	
MTOR[#]	Hs00234522_m1 (Taqman gene expression Assay, Invitrogen)	
NDUFA1[#]	Hs00244980_m1 (Taqman gene expression Assay, Invitrogen)	

*, Pre-designed qRT-PCR primers manufactured by Invitrogen, USA.
#, Custom designed qRT-PCR primers based on the sequences target.

on Illumina platforms were downloaded from ENCODE and ArrayExpress databases with accession numbers accession numbers ENCSR000CUC and E-MTAB-2399 (Table 3.5).

3.11.2 Alignment to genome and transcript assembly

The same pipelines as described in sections 3.8.1 were used to check the quality of the RNA-Seq datasets, followed by alignment to a reference genome and constructing the assembled transcripts.

3.11.3 Gene catalogue of human primary monocytes

To generate a reference gene catalogue of human primary monocytes, the same pipelines as described in sections 3.8.2 were used to conduct the analysis. To detect the expression patterns of transcription factors (TFs) in monocytes, the merged transcripts assembly file from all datasets was compared with the list of human TFs which was compiled from literatures (Ravasi et al., 2010; Roach et al., 2007) and GO term 'transcription factor'. In order to detect the TFs-genes targets, the transcriptional regulatory relationships unraveled by sentence-based text-mining (TRRUST) database (Han et al., 2015). The interaction network between TFs with their genes target was constructed using Cytoscape plug-in GeneMANIA (Montojo et al, 2010) by applying the information from the co-expression category.

3.11.4 Identification of lncRNAs

To detect the expression patterns of annotated lncRNAs and novel lincRNAs in human primary monocytes, the same pipelines as described in sections 3.9.3.1 and 3.9.3.2 was used to conduct the analysis.

Table 3.5: Summary of RNA-Seq datasets obtained from public databases.

Study and sample ID	Gender	Age	Total number of Reads	Experiment
Iiott et al. 2014 (ERS422905)	Female	47	58,278,244	Gene catalogue/lncRNA landscape
Iiott et al. 2014 (ERS422908)	Male	32	60,554,892	Gene catalogue/lncRNA landscape
Iiott et al. 2014 (ERS422906)	Female	47	55,333,880	Gene catalogue/lncRNA landscape
Iiott et al. 2014 (ERS422910)	Male	42	54,265,696	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUY, ENCFF000HVE)	Female	N/A	59,181,719	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUX, ENCFF000HVD)	Female	N/A	58,610,690	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUW, ENCFF000HVC)	Female	N/A	86,613,622	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUU, ENCFF000HVA)	Female	N/A	82,389,934	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUZ, ENCFF000HVF)	Female	N/A	61,619,085	Gene catalogue/lncRNA landscape
Derrien et al. 2012 (ENCFF000HUV, ENCFF000HVB)	Female	N/A	61,619,085	Gene catalogue/lncRNA landscape

N/A, The information is not available in the public databases

3.11.4.1 Secondary structures of novel lincRNAs

The standalone version of RNAFold Vienna RNA package 2.0 (Lorenz et al., 2011) was downloaded and installed locally. All the sequences for novel lincRNAs were subjected to RNAfold secondary structure prediction and minimum stable energy calculations. The RNAfold was run with default parameters.

3.11.4.2 Validation of lncRNAs by RT-PCR

Five novel lincRNAs were randomly selected along with 2 previously annotated lncRNAs *DANCR* (differentiation antagonizing non-protein coding RNA) (Tong, Gu, Xu, & Lin, 2015) and *LINC01420* (long intergenic non-protein coding RNA 1420) (Zhu et al., 2015) for validation in monocytes and five other hematopoietic cells including T cell ($CD3^+$), T Helper cell ($CD4^+$), Regulatory T Cell ($CD4^+CD25^+$), Cytotoxic T cell ($CD8^+$) and B Cell ($CD19^+$), using Reverse Transcription Polymerase Chain Reaction (RT-PCR). RNAs of 5 examined hematopoietic cells with >90% purity were purchased from Miltenyi Biotec. Primers pairs were designed for each transcript using Primer3 (Untergasser et al., 2012) with the following parameters: primer length of 20−25 bp, melting temperature of 55−65 ºC and GC content of 40−60%. The primer sequences were presented in Table 3.6. The specificity of primers was checked by BLAST search against human genome and no significant homology was found with other genomic regions. cDNA synthesis and PCR amplification was performed using SuperScript III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (Invitrogen, USA) according to manufacturer's protocol. PCR product were resolved by electrophoresis on a 2% (w/v) agarose gel.

Table 3.6: List of selected annotated lncRNAs and novel lincRNAs for validation by RT-PCR and their respective primer sequences.

Target ID	Forward primer (5`- 3`)	Reverse primer (5`- 3`)
DANCR	TTGTCAGCTGGAGTTGCGCGG	TGTCACTGCTCTAGCTCCTGTGG
LINC01420	GTAATCCTTATGGGAGACCAACC	AAC TTC CAG GTT CAG GAC AC
TCONS_00056181	AAGCCGAGGCAGGGTGATCACG	CCATGCCAGAAGTTCTGCAGGCTG
TCONS_00008494	GCTGAGAAGAGAAGTCAGAGTTGAGC	CCAAAGGACTAATATTGGTACATCGC
TCONS_00226266	CTTATAGGAGACCTCCCAGTGC	ACAGCCACCTTGTCTTTGCT
TCONS_00282615	GCCAGACACATTCCTCTTCC	ACAAGAGCACAGCCTGGGAG
TCONS_00128310	GTGCAGTGACGTGATTCTGAC	CATCTTCCCAAATGGAAACTCG

CHAPTER 4: GENE EXPRESSION PROFILING OF HUMAN PRIMARY MONOCYTES FROM HEALTHY SUBJECTS

4.1 Introduction

Monocytes are crucial players in the innate immune system and essential for front-line defense against pathogens. While several studies have addressed the functional elements in monocytes subsets (Ancuta et al., 2009; Dong et al., 2013; Ziegler-Heitbrock et al., 2010). However, there is limited or no data exist on the genome-wide transcriptome expression profile of human primary monocytes under healthy state. Moreover, despite evidence of gender differences in immune responses (Lozano et al., 2012; Oertelt-Prigione, 2012), the knowledge of gender differences in expression patterns of innate immune-related genes of primary monocytes from healthy subjects is still limited. Thus, this chapter profiled a genome-wide transcriptome expression of primary monocytes from healthy subjects and provide comprehensive overview of immune-related genes, using deep RNA-Seq analysis. In addition, a comparative analysis of gene expression profiles between healthy male and female subjects has been conducted to identify possible differences of immune-related genes expression patterns based on genders. The schematic representation of workflow of the bioinformatics analysis procedure is described in Figure 4.1. The analysis results presented in this chapter is to achieve the first objective of the thesis as specified in Chapter 1.

4.2 Transcriptome profile of primary monocytes from healthy subjects

The peripheral blood samples from 6 healthy subjects (including 3 male and 3 female) were collected. The classical monocytes ($CD14^{++}CD16^{-}$) were isolated from peripheral PBMCs of subjects using a negative selection technique. The purity of isolated monocytes from all samples were checked separately by flow cytometry analysis and

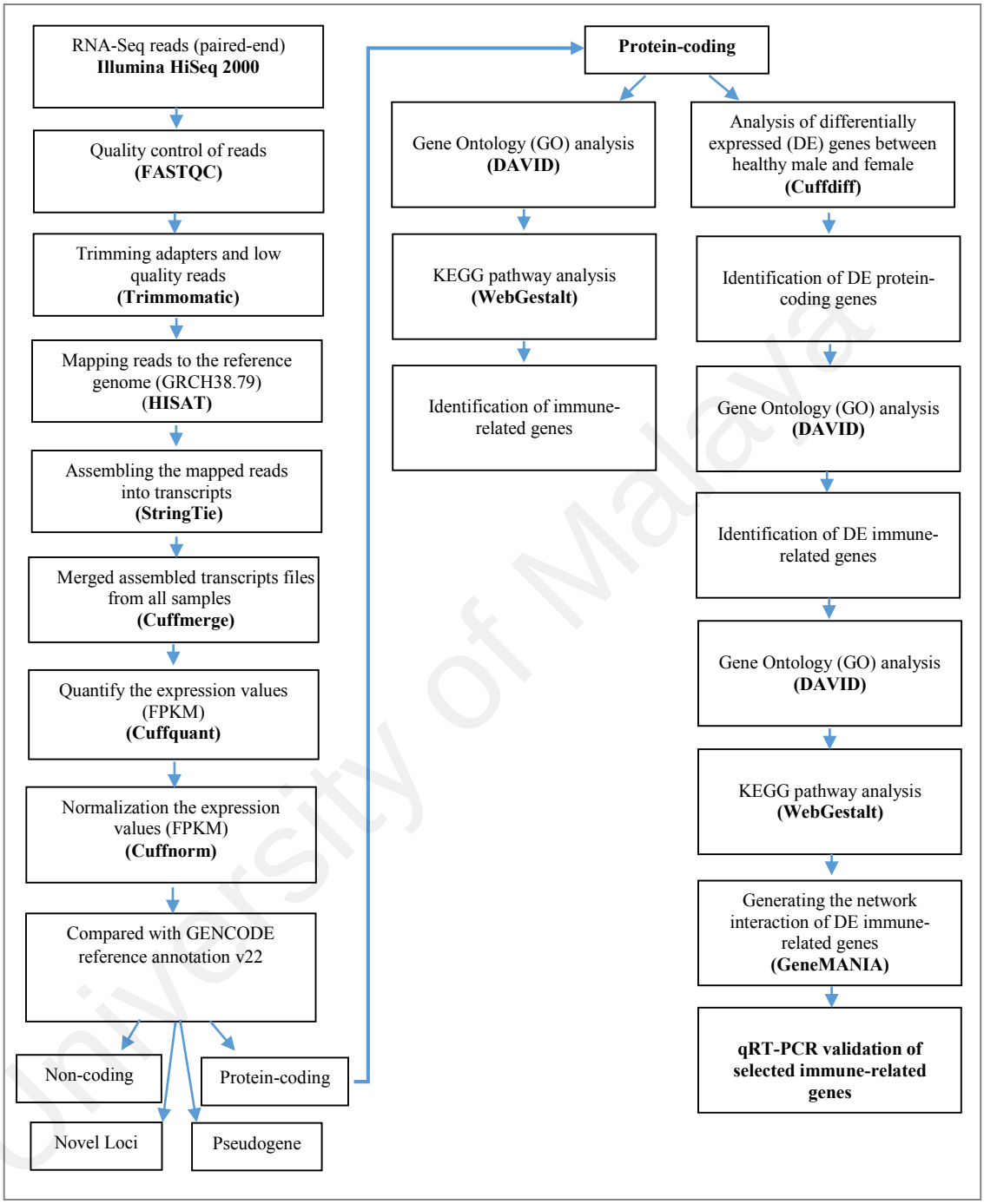

Figure 4.1: Schematic representation of workflow for gene expression profiling of primary monocytes from healthy subjects.

were found to be >90% for all samples. One presentation of flow cytometry analysis results is shown in Figure 4.2. The total RNA was extracted from monocytes and the quality, quantity, and integrity of the extracted RNA from all samples were check separately as presented in Table 4.1. The purified RNA was used to perform deep poly(A)$^+$ paired-end RNA sequencing.

All the reads were mapped to reference genome (Ensembl GRCH38.79) and assembled into transcriptome using the same pipeline to reduce any bias. An average 90% of the reads were aligned to the human reference genome (Ensembl GRCH38.79). The alignment summary from each sample is shown in Table 4.2. The transcripts were reconstructed for all aligned reads separately and then merged together to form a single non-redundant set of transcripts. The abundance of assembled transcripts was estimated using fragments per kilobase of exon per million fragments mapped (FPKM) value. By applying the FPKM >0.1 threshold, a total of 17,657 genes (including protein-coding genes, non-coding genes, pseudogenes) and 81,419 transcripts (including annotated and novel transcripts) were identified in our monocytes datasets. The distribution of identified genes and transcripts is shown [Figure 4.3 A]. The expression levels of the protein-coding genes were compared with non-coding and pseudogenes. The results showed that, the average expression of protein-coding genes was relatively higher than the pseudogenes and non-coding genes [Figure 4.3 B]. The potential functions of the identified protein-coding genes were determined using GO analysis based on the biological process categories using DAVID software. The GO analysis was performed by selecting the highest hierarchical level of biological process ontology (DAVID category-GOTERM_BP_1) which give a broad overview information for genes functions. The 11,644 protein-coding genes were assigned to 15 significant biological process terms (ajdP <0.01) mainly involved in “cellular process” “metabolic process”, “immune system

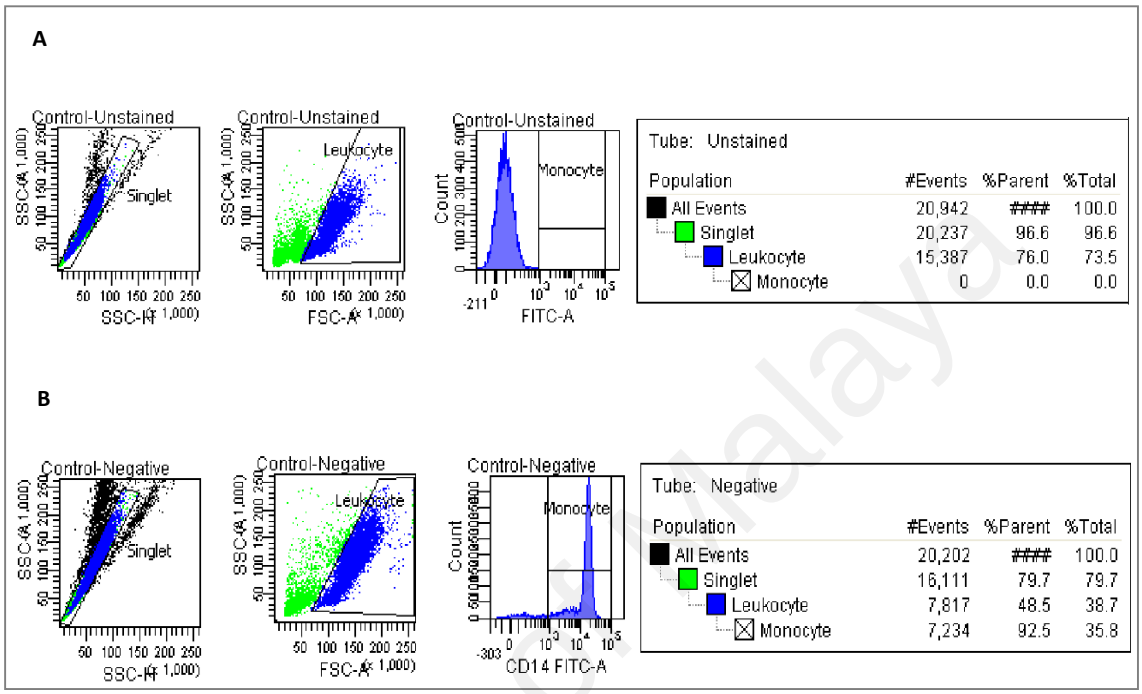

Figure 4.2: Flow cytometry analysis of isolated monocyte from healthy subject. The results show the purity and percentage of monocytes during a 2 steps monocyte isolation process involving, A: Before magnetically labeling the non-monocytes cells (labeled as Control-Unstained) and, B: After magnetically separation of monocytes from non-monocytes cells through negative selection (labeled as Control-Negative).

Table 4.1: The quality control results of RNA samples of primary monocytes from healthy subjects.

Sample_ID	Vol (µl)	Nano Drop				Qubit RNA BR		
		Con (ng/µl)	Total (µg)	A260/280	A260/230	Con (ng/µl)	Total (µg)	Bioanalyzer RIN No
Healthy Male_01	35	235.1	8.3	2.05	1.82	206.0	7.3	9.6
Healthy Male_02	30	89.6	2.7	2.04	1.79	95.2	2.8	9.8
Healthy Male_03	35	214.1	7.4	2.04	1.91	216.0	7.5	9.6
Healthy Female_01	35	149.7	5.2	2.07	1.90	159.0	5.5	10
Healthy Female_02	35	160.3	8.2	2.06	1.76	165.0	7.2	9.8
Healthy Female_03	35	138.0	4.5	2.03	1.96	145.0	4.7	9.3

Table 4.2: Summary of alignment results of RNA-Seq datasets generated from primary monocytes of healthy subjects.

Sample ID	Percentage of mapped reads
Healthy Male_01	92%
Healthy Male_02	91%
Healthy Male_03	90%
Healthy Female_01	90%
Healthy Female _02	92%
Healthy Female _03	92%

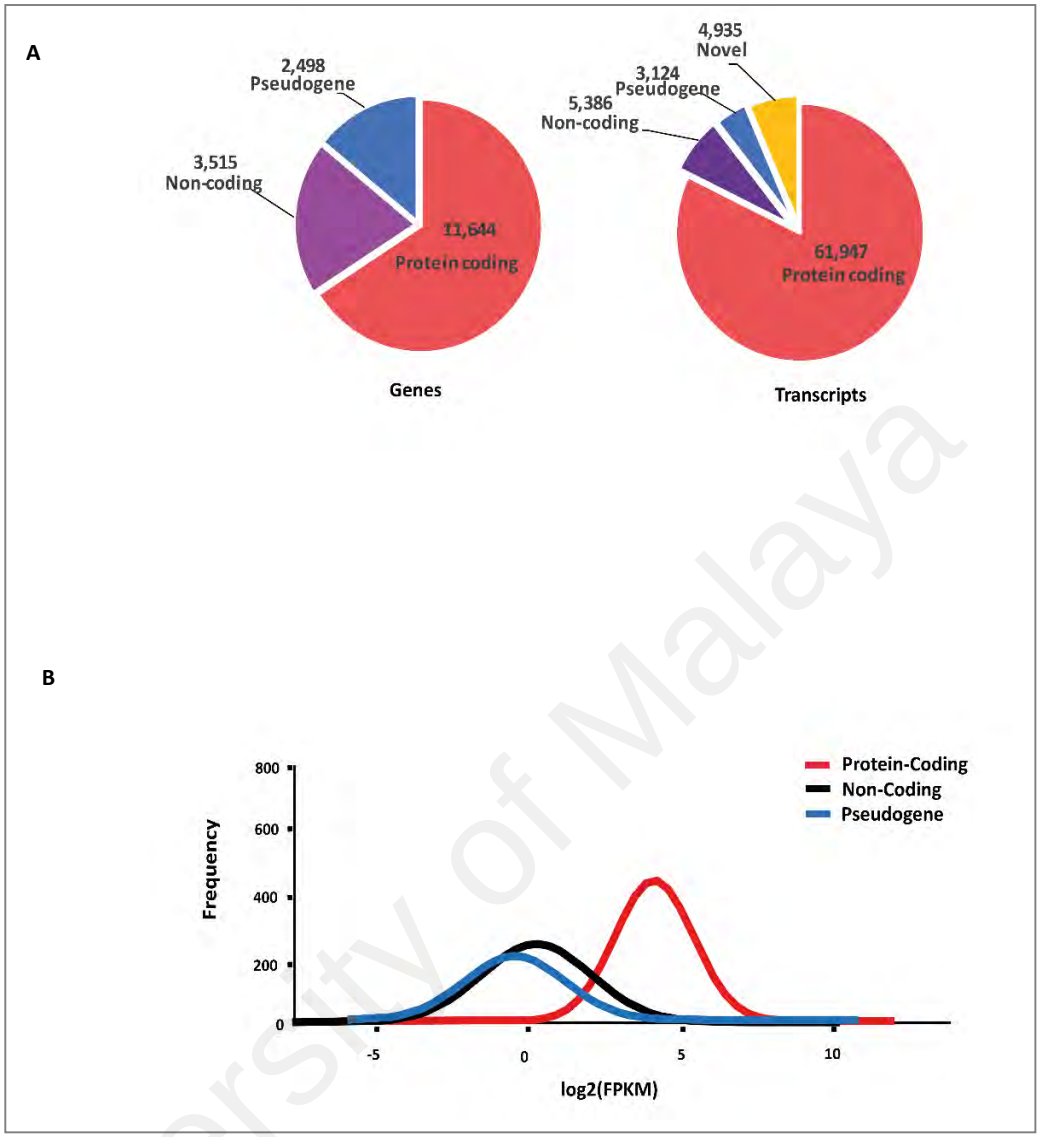

Figure 4.3: Transcriptome of primary monocytes from healthy subjects. A: Pie chart representing the number of diverse classes of identified genes and transcripts. B: Average expression levels of identified protein-coding, non-coding genes and pseudogenes.

process", "death" and "response to stimulus" [Figure 4.4 A]. Next, the protein-coding genes were mapped to KEGG database for signaling pathways analysis. The results showed that 11,644 protein-coding genes were mapped into 86 KEGG pathway. The top 10 significant pathways (ajdP <0.01) were mostly immune-related pathways such as "Fc gamma R-mediated phagocytosis", "Cytokine-cytokine receptor interaction", "Chemokine signaling pathway", "Toll-like receptor signaling pathway" and "RGI-like receptor signaling pathway" [Figure 4.4 B]. In total, 804 unique immune-related genes were identified in several KEGG immune system pathways. The detailed information regarding the chromosome position and expression level of these immune-related genes are summarized in APPENDIX A.

4.3 Differentially expressed (DE) genes between healthy male and female subjects

The Cuffdiff software was used to perform differentially gene expression analysis in monocytes between healthy male and female by applying q-value ≤ 0.01 and $\log_2$ fold-change ≥ 1 or ≤ -1. Based on this criteria, a total of 217 differentially expressed (DE) protein-coding genes were found from our datasets in which 169 and 48 genes were upregulated and downregulated in male compared to female, respectively. Fisher exact test (Fisher, 1922) was used to detect possible non-random chromosomal distribution of the identified DE protein-coding genes. The frequencies of DE protein-coding genes on each chromosome against the total number of expressed protein-coding genes were compared. The result showed that the DE protein-coding genes are distributed across all chromosomes (Figure 4.5). Chromosomes 7 and 21 were found to be enriched for male-biased genes, whereas chromosomes 4 and 22 contained more female-biased genes (p-value< 0.05; Fisher exact test). This results revealed that DE protein-coding genes on X chromosome of primary monocytes are not biased in female. Among the 8 DE protein-coding genes located on chromosome X, *TLR8* is the only gene that highly expressed in

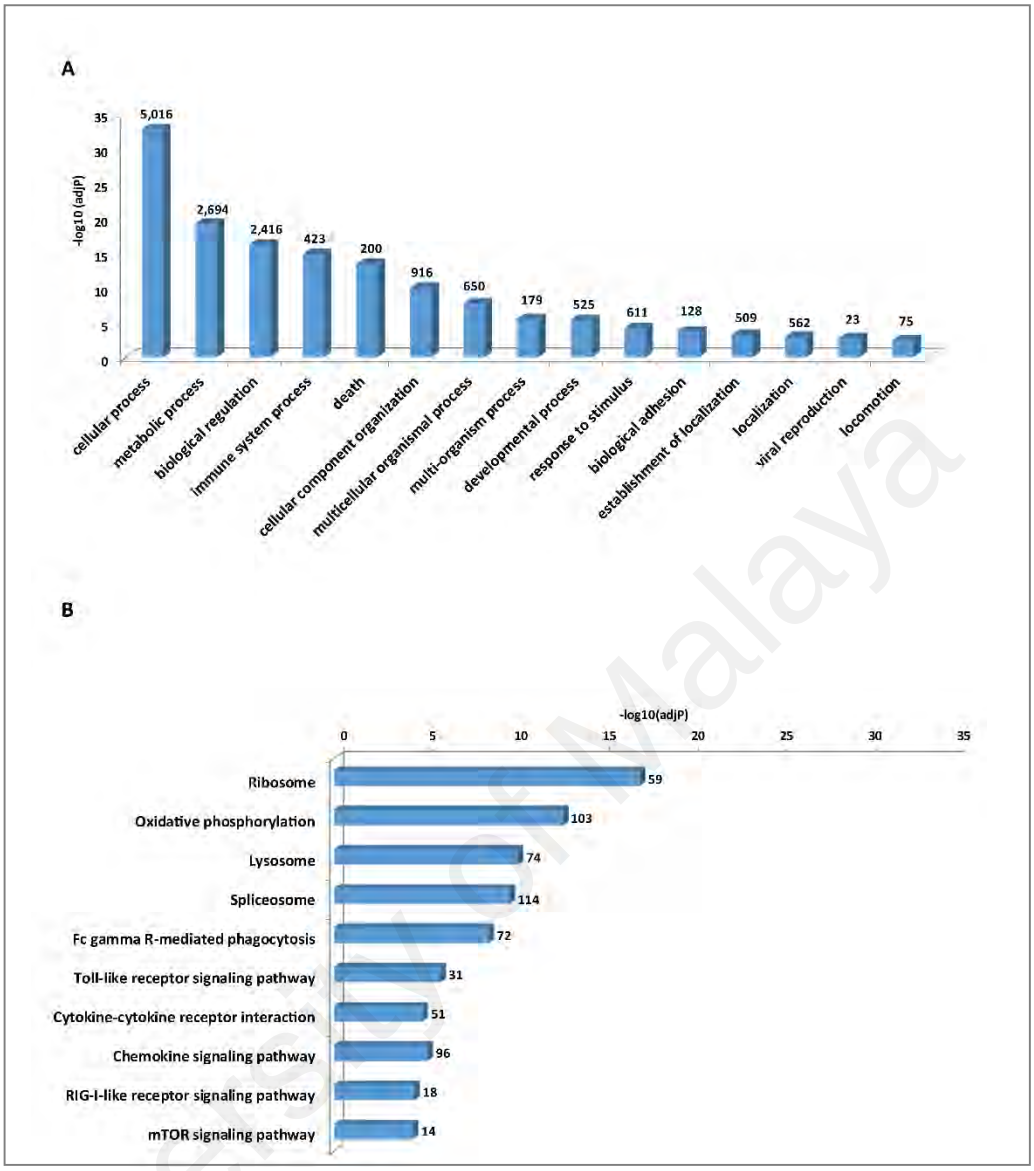

Figure 4.4: The GO and KEGG pathway analysis of protein-coding genes expressed in primary monocytes from healthy subjects. A: The significant GO biological process terms (DAVID category-GOTERM_BP_1) enriched for protein-coding genes. The number of identified protein-coding genes enriched in each GO terms is depicted above the bars in the figure. B: The significant KEGG pathway terms enriched for protein-coding genes. The numbers in the brackets indicated the total numbers of genes available in the KEGG database for each pathway terms. The number of identified protein-coding genes enriched in each KEGG pathways from our analysis is depicted above the bars in the figure.

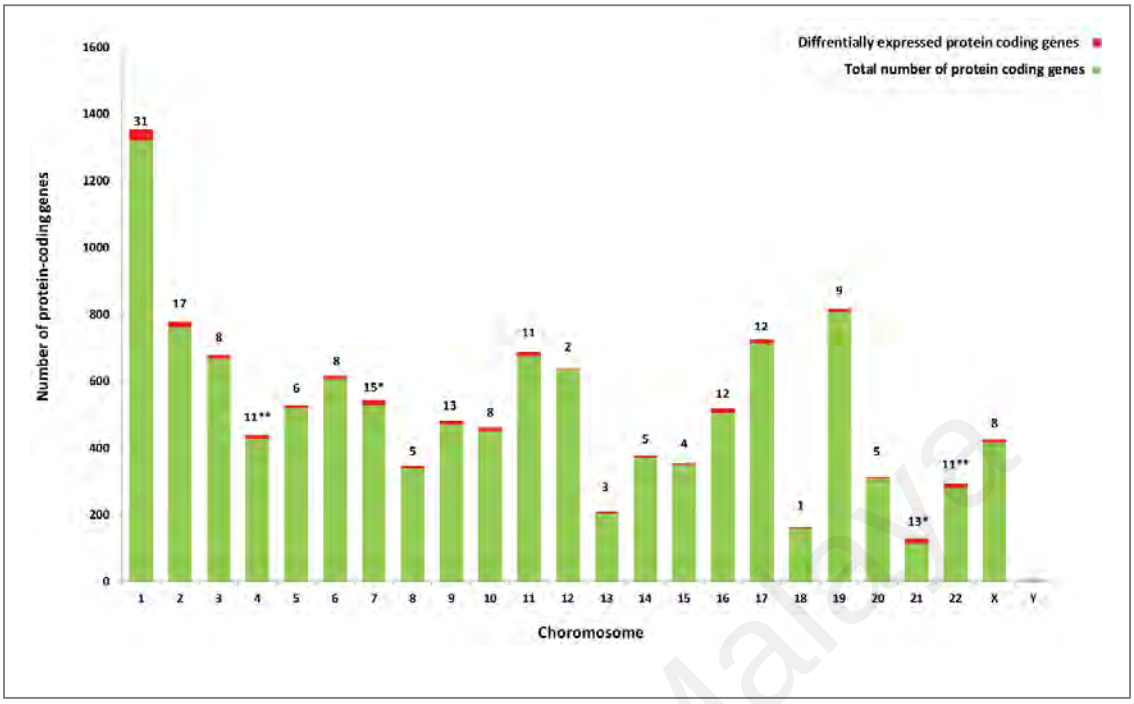

Figure 4.5: The chromosomal distribution of DE protein-coding genes in primary monocytes of male compared to female. The bars represented the total number of protein-coding genes on each chromosome. The red boxes at the top of each bar indicate the number of DE protein-coding genes on each chromosome. The chromosomes that passed the enrichment test ($P < 0.05$, Fisher's exact test) are symbolized as *, ** for male-biased genes and female-biased genes, respectively.

female compared to male. The GO analysis was used to characterize the functional consequences of identified 217 DE protein-coding genes. Our GO analysis was based on the biological processes category for both upregulated and downregulated genes in male compared to female. Initially the highest hierarchical level of biological process ontology (DAVID category-GOTERM_BP_1) was selected to identify the overview functional information for DE genes (Figure 4.6). The 169 upregulated genes were assigned to 5 general GO terms (ajdP <0.01): "immune system process", "biological adhesion", "multi-organism process", "response to stimulus" and "locomotion" [Figure 4.6 A]. While the 48 downregulated genes were assigned to 4 general significant GO terms (ajdP <0.01): "immune system process", "response to stimulus", "signaling" and "death" [Figure 4.6 B]. The results indicated that out of the 217 DE protein-coding genes, 40 unique DE genes were related to immune system. These DE immune-related genes were selected for further analysis (see below).

4.3.1 DE immune-related genes

Out of the 40 DE immune-related genes in primary monocytes between male and female, 23 and 17 genes were upregulated and downregulated in male compared to the female, respectively (Table 4.3). The hierarchical clustering analysis was applied on DE immune-related genes which revealed distinct transcription expression profiles for these genes between male and female groups (Figure 4.7). In order to identify the specific GO terms in the biological process for identified DE immune-related genes, the DAVID category-GOTERM_BP_FAT level were selected. The top 10 significant (adjP < 0.01) GO terms for upregulated and downregulated genes are shown in Figure 4.8. The upregulated genes in male were involved in "response to bacterium", "immune response", and "defense response" [Figure 4.8 A], whereas the downregulated genes were mainly

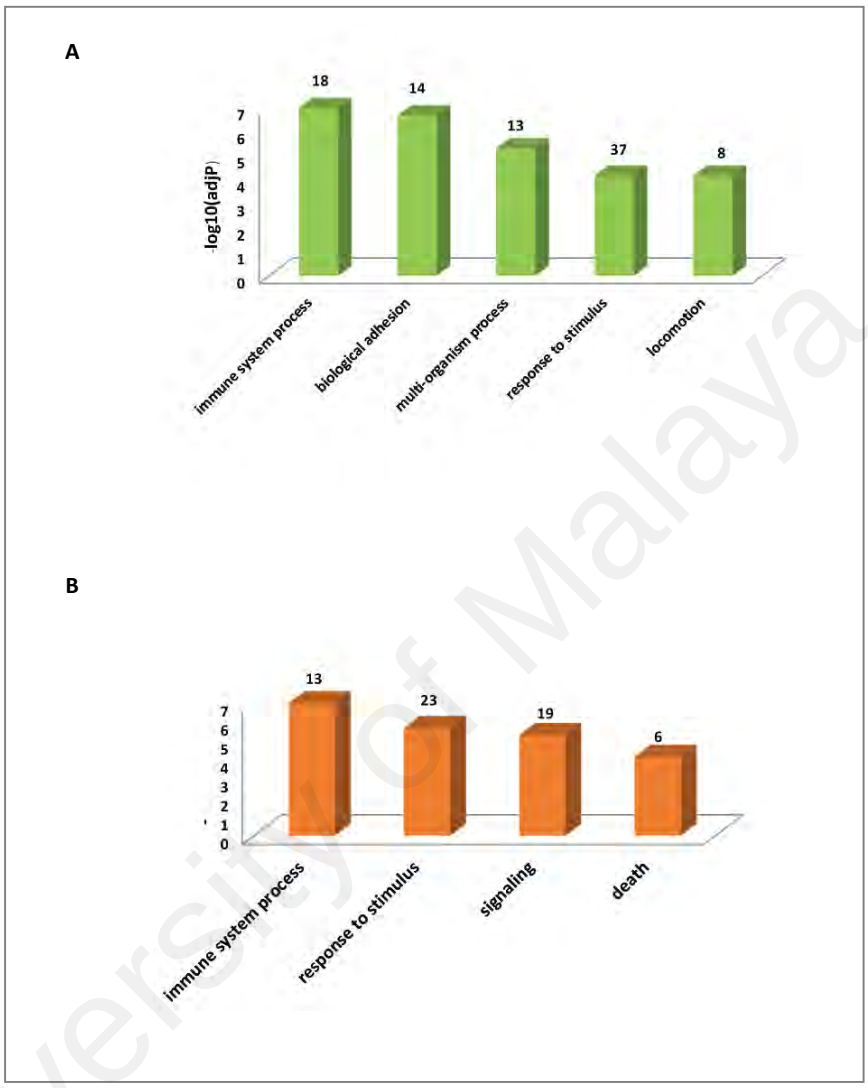

Figure 4.6: The GO analysis of DE protein-coding genes in primary monocytes of male compared to female. A: The significant GO biological process terms (DAVID category-GOTERM_BP_1) enriched for upregulated genes. B: The significant GO biological process terms (DAVID category-GOTERM_BP_1) enriched for downregulated genes. The number of DE protein-coding genes enriched in each GO term is depicted above the bars in the figure.

Table 4.3: List of DE immune-related genes in primary monocytes of male compared to female.

Gene_Name	Chromosome	FPKM_ Male	FPKM_ Female	log_2 (fold-change)	q-value
Upregulated					
DEFA1	8	1003.25	12.079	6.376036	0.001731
CTSG	14	9.52429	0.436636	4.447109	0.001731
ICOSLG	21	4.83809	0.279382	4.114127	0.001731
CHIT1	1	2.25221	0.149075	3.917231	0.001731
ELANE	19	7.98723	0.690534	3.531911	0.001731
AZU1	19	7.87397	0.740367	3.410779	0.001731
LTF	3	199.625	20.2574	3.300771	0.001731
CD24	6	27.3794	2.84339	3.267407	0.001731
RGS1	1	3.42094	0.481172	2.829768	0.001731
OLR1	12	5.27656	0.797704	2.725672	0.001731
CRISP3	6	12.4887	1.9078	2.710641	0.001731
HSPA1A	6	104.205	16.7022	2.641314	0.001731
PXDN	2	0.516054	0.083	2.636341	0.001731
DUSP1	5	1.19267	0.207375	2.523881	0.003217
GADD45G	9	1.01621	0.24455	2.054997	0.00831
JUN	1	27.3016	7.04482	1.954351	0.001731
MMP9	20	11.0684	3.23851	1.773045	0.001731
BPI	20	23.0461	6.82345	1.755949	0.001731
DUSP2	2	5.05191	1.81114	1.479931	0.005958
EGR1	5	10.426	4.05458	1.362561	0.001731
NFKBIA	14	326.44	135.716	1.266227	0.001731
FKBP1C	6	12.2016	5.20107	1.230190	0.001731
JUNB	19	333.361	166.21	1.004078	0.001731
Downregulated					
CXCL1	4	0.228006	1.42925	-2.648115	0.001731
TLR1	4	17.89413	87.32373	-2.286887	0.003439
HRH4	18	0.241441	1.05914	-2.133151	0.00831
SLC18A1	8	0.374821	1.42304	-1.924703	0.001731
CXCL10	4	9.95418	37.3327	-1.907065	0.001731

Table 4.3: Continued.

SERPING1	11	21.164	69.0734	-1.706518	0.001731
CD40	20	4.34771	14.1175	-1.699157	0.001731
TLR4	9	21.3456	67.4791	-1.660502	0.001731
TLR2	4	34.5845	96.645	-1.482569	0.00831
SLAMF7	1	26.6957	67.0692	-1.329043	0.001731
TNFSF10	3	155.197	339.094	-1.127585	0.001731
GCH1	14	2.75793	6.00395	-1.122326	0.003217
GIMAP7	7	51.2417	111.008	-1.115273	0.001731
DDX58	9	17.8983	38.0717	-1.088896	0.001731
STAT1	2	190.12	391.823	-1.043292	0.003217
IFIH1	2	23.9344	49.2822	-1.041981	0.001731
TLR8	X	39.8544	81.847	-1.038191	0.00831

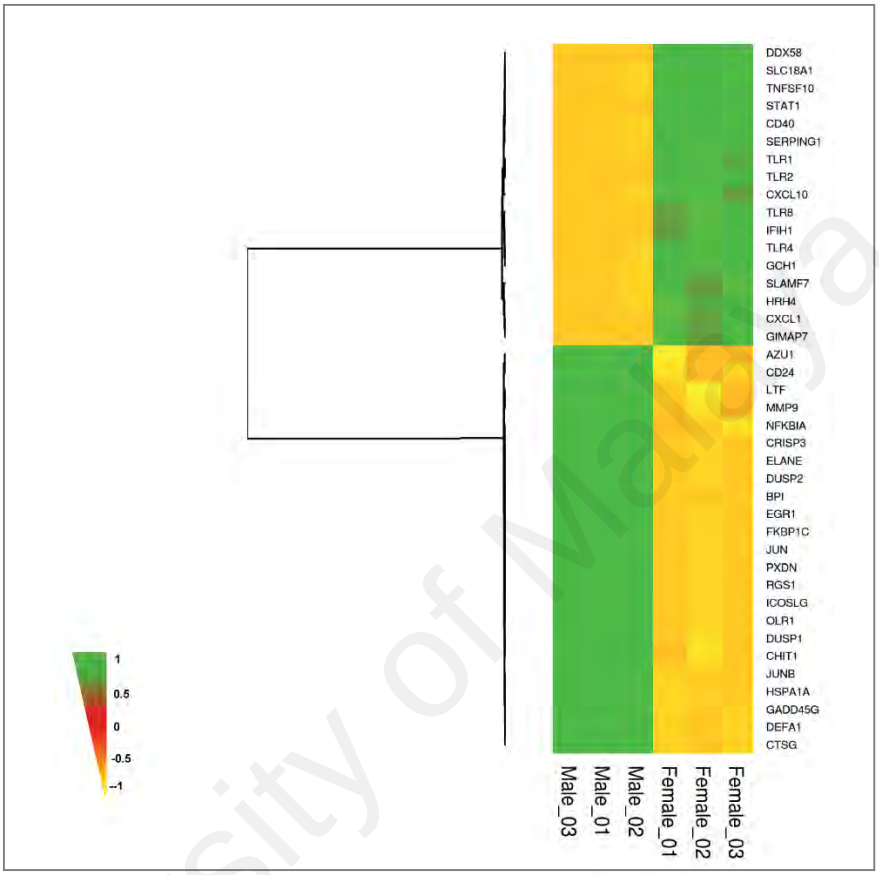

Figure 4.7: Hierarchical clustering of DE immune-related genes in primary monocytes of male compared to female. Each row in the heatmap represents a gene and each column represents a separate sample. The colours correspond to the $\log_2$ of FPKM values (green colour for upregulated genes and orange colour for downregulated genes).

related to “defense response”, “innate immune response” and “inflammatory response” [Figure 4.8 B]. The “defense response” and “immune response” categories were overlapped for both upregulated and downregulated genes but contained few upregulated genes in male compared to female. The results showed that the expression of innate immune-related genes; *DDX58* (DEXD/H-Box Helicase 58), *GCH1* (GTP Cyclohydrolase 1), *SLAMF7* (SLAM Family Member 7), *IFIH1* (Interferon Induced With Helicase C Domain 1), *SERPING1* (Serpin Family G Member 1), *TLR1*, *TLR2*, *TLR8* and inflammatory response-related genes; *CXCL1* [The chemokine (C-X-C motif) ligand 1], *HRH4* (Histamine Receptor H4), *CXCL10* (C-X-C Motif Chemokine Ligand 10), *IDO1* (Indoleamine 2,3-Dioxygenase 1), *APOL2* (Apolipoprotein L2) were significantly downregulated in male compared to female. To further understand the function of DE immune-related genes, a KEGG pathway enrichment analysis was performed for both upregulated and downregulated genes in male compared to female as shown in Figure 4.9. The most significant (adjP < 0.01) pathways enriched for upregulated genes were “MAPK signaling pathway” and “Osteoclast differentiation” [Figure 4.9 A], whereas the most significant pathways enriched for downregulated genes were “Toll-like receptor signaling pathway”, “RIG-I-like receptor signaling pathway”, “Cytokine-cytokine receptor interaction” and “chemokines signaling pathway” [Figure 4.9 B].

4.3.1.1 Gene interaction network of DE immune-related genes

The interaction network generated on all identified DE immune-related genes showed the relationships between upregulated and downregulated genes in terms of their co-expression. The apoptosis-related genes *JUN* (Jun Proto-Oncogene) and *STAT1* (Signal Transducer and Activator of Transcription 1, 91kDa) were found to be the core upregulated and downregulated genes in the network map where the other DE immune-related genes were directly or indirectly associated to these genes (Figure 4.10).

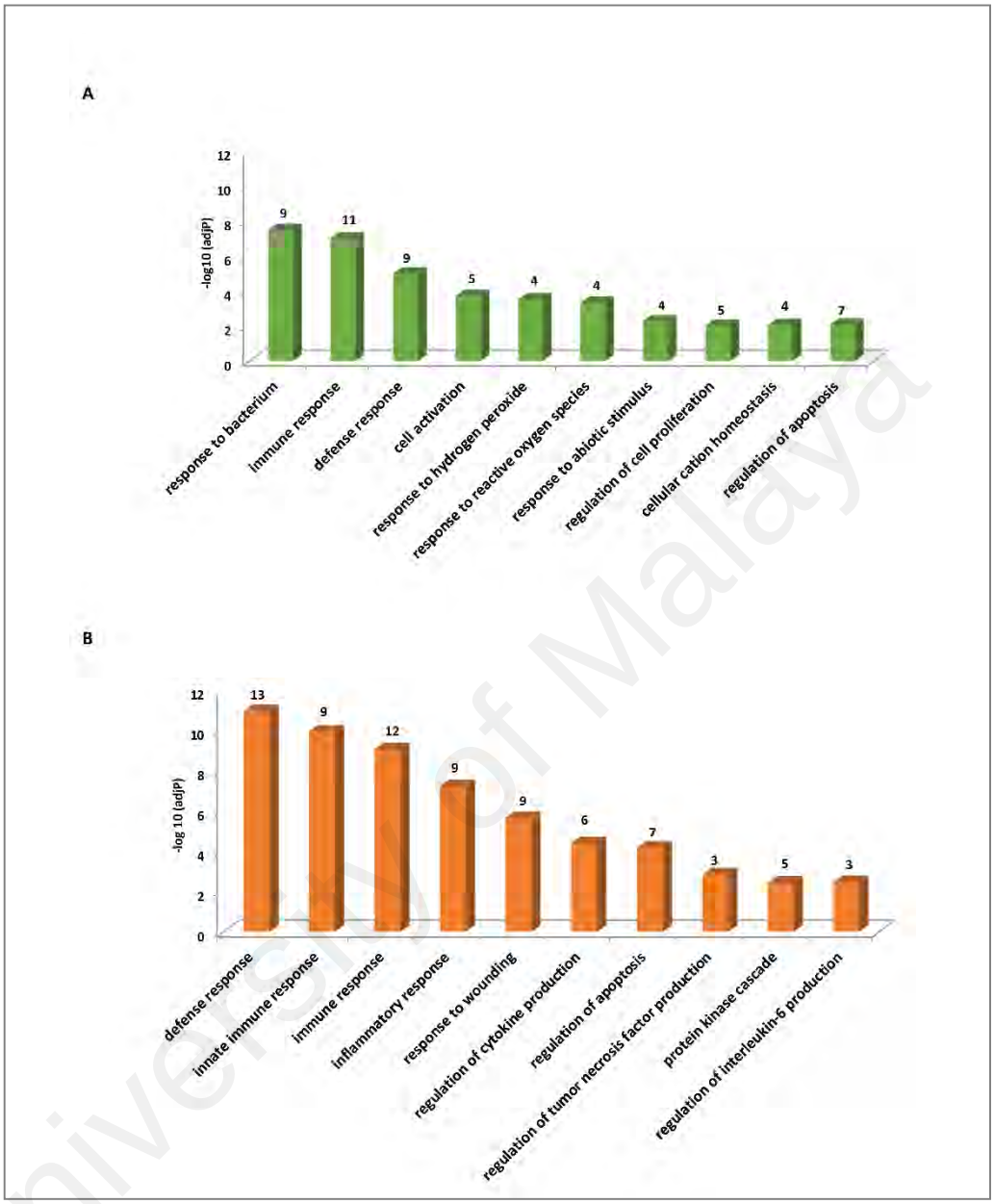

Figure 4.8: The GO analysis of DE immune-related genes in primary monocytes of male compared to female. A: The significant GO biological process terms (DAVID category-GOTERM_BP_FAT) enriched for upregulated genes. B: The significant GO biological process terms (DAVID category-GOTERM_BP_FAT) enriched for downregulated genes. The number of DE immune-related genes enriched in each GO term is depicted above the bars in the figure.

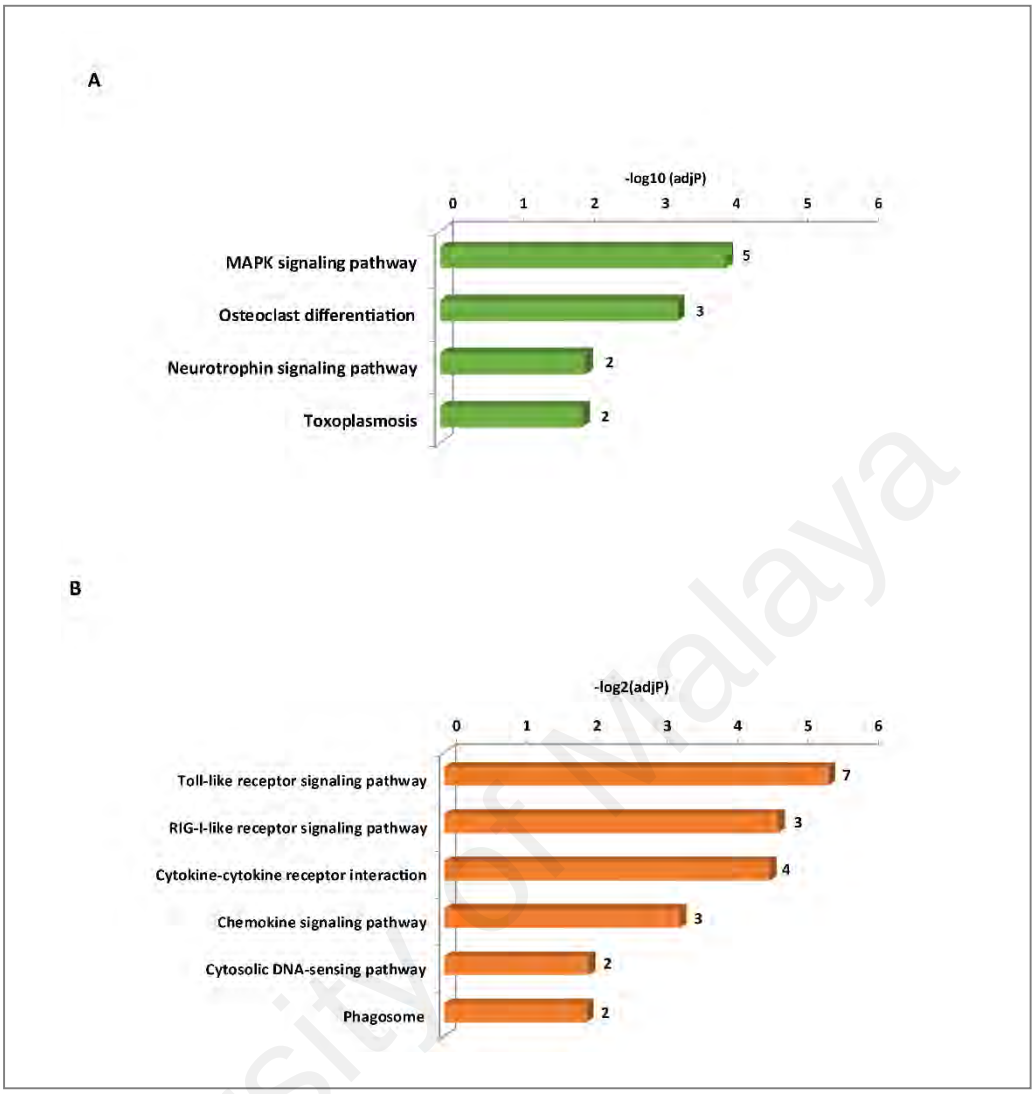

Figure 4.9: The KEGG pathway analysis of DE immune-related genes in primary monocytes of male compared to female. A: The significant KEGG pathway enriched for upregulated genes. B: The significant KEGG pathway enriched for downregulated genes. The numbers in the brackets indicated the total numbers of genes available in the KEGG database for each pathway terms. The number of identified DE immune-related genes enriched in each KEGG pathway is depicted above the bars in the figure.

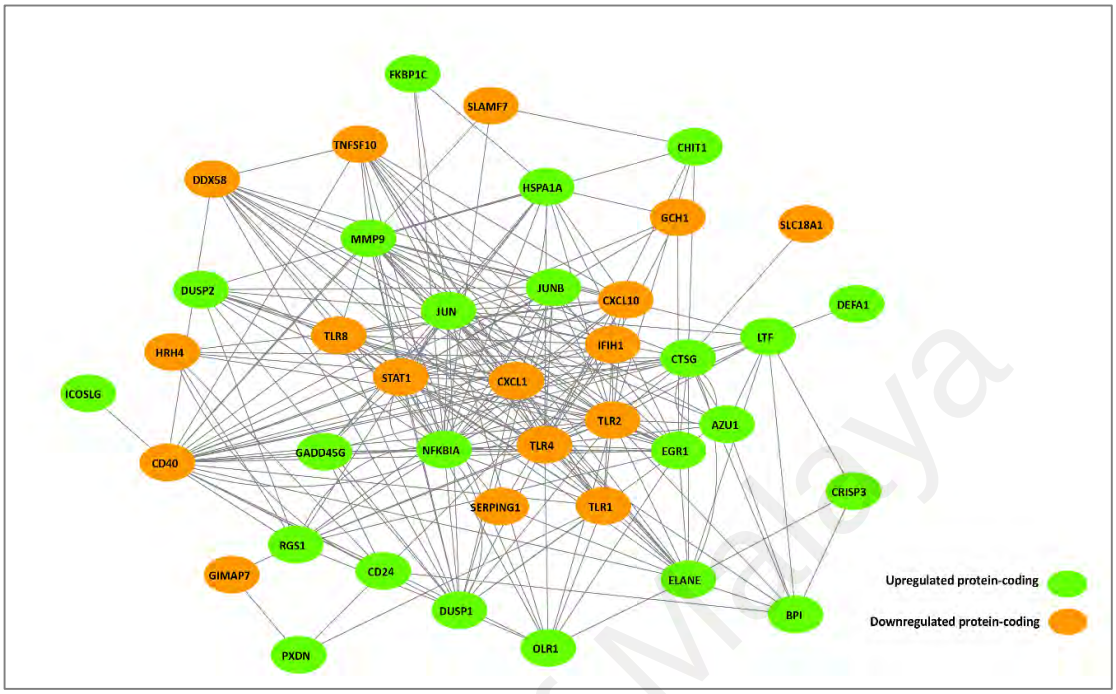

Figure 4.10: Interaction network analysis of DE immune-related genes in primary monocytes of male compared to female. The DE immune-related genes were connected in a network based on their co-expression.

4.3.1.2 qRT-PCR validation

The expression level of *JUN* and *STAT1* in male and female were further validated through qRT-PCR analysis using the *ACTB* (Actin, Beta) and *PPIA* (Peptidylprolyl Isomerase A) as endogenous controls. The comparison between qPCR derived $\log_2$ fold-changes values with $\log_2$ fold-change values obtained through RNA-Seq analysis demonstrated that our qRT-PCR results was in agreement with the RNA-Seq results where *JUN* and *STAT1* were upregulated and downregulated in primary monocytes of male compared to female, respectively (Figure 4.11).

4.4 Discussion

Deep RNA-Seq approach on primary monocytes from 6 healthy subjects was performed. The sequencing generated approximately 1.3 billion reads of 100 bp read length. Using this datasets, the expression of 17,657 genes (including 11,644 protein-coding, 3,515 non-coding and 2,498 pseudogenes) and 81,419 transcripts (including 70,457 annotated transcripts and 4,935 novel transcripts) were detected in primary monocytes. Protein-coding genes were expressed at different rates, and is influenced by different parameters. Identifying and measuring the protein-coding genes expression at transcriptome level is important to quantify which particular gene is expressed within a cell, tissue or organism under different conditions (Kryuchkova-Mostacci & Robinson-Rechavi, 2015). The functional analysis of identified protein-coding genes showed the expression of 804 immune-related genes in primary monocytes. Further comparative analysis of gene expression pattern between healthy male and female subjects revealed a total number of 217 DE protein-coding genes in male compared to female. The chromosomal distribution analysis of identified DE protein-coding genes indicated that these DE protein-coding genes were distributed across all chromosomes, not only on the sex chromosomes. Out of 8 DE protein-coding genes expressed on chromosome X, *TLR8*

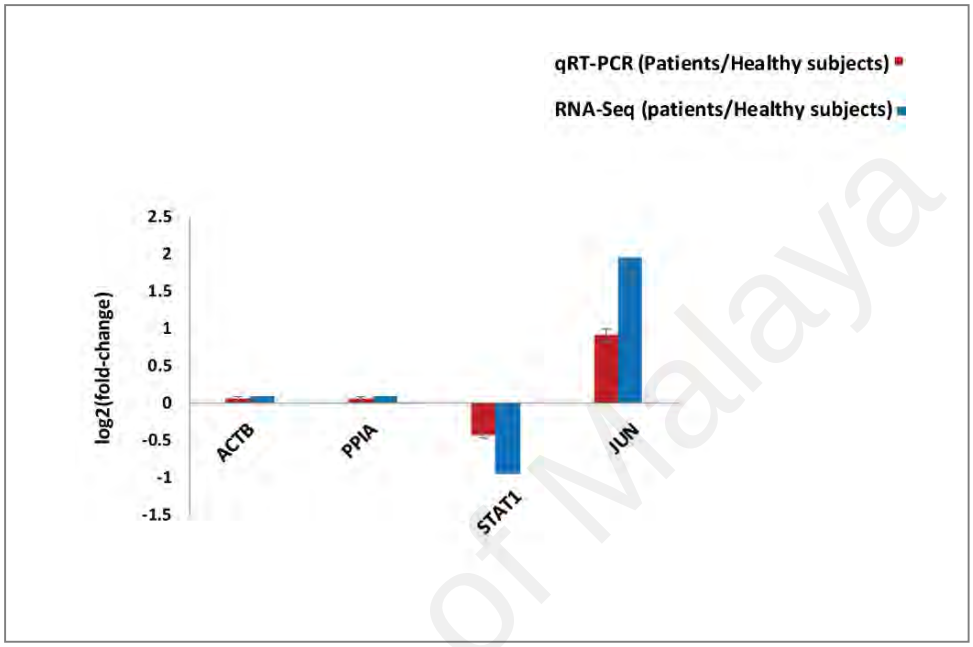

Figure 4.11: The qRT-PCR validation of *JUN* and *STAT1* expression patterns in primary monocytes of male compared to female. The comparison of $\log_2$ fold-change of *JUN* and *STAT1* were determined by RNA-Seq analysis (blue) and qRT-PCR validation (red). *ACTB* and *PPIA* were used as endogenous controls for normalizing the expression levels. x-axis shows genes; y-axis shows the $\log_2$ ratio of expression in male compared to female.

was the only gene that highly expressed in female compared to male. *TLR8* is known as X-linked mediator of innate immunity (Young et al., 2014) and mainly expressed in monocytes, macrophage and dendritic cells (Cervantes, Weinerman, Basole, & Salazar, 2012).

The functional analysis of DE protein-coding genes showed that 40 DE genes were related to immune system. Out of 40 DE immune-related genes, 23 and 17 genes were upregulated and downregulated in male compared to the female, respectively. The expression levels of innate immune-related genes were observed to be significantly higher in female compared to male. Female are reported to have stronger cellular and humoral immune response (Batchelor, 1968) and higher risk of autoimmune disease (Ngo, Steyn, & McCombe, 2014). Sex hormones are known to be attributed to the differences in innate and adaptive immune response in different gender by binding to the intracellular receptors of the immune cells, such as monocytes, B cells and T cells, leading to activation of immune-responsive genes (Ngo et al., 2014; Whitney et al., 2003). Estrogen regulate several immune molecules such as cytokines (Calippe et al., 2008), *DDX58* (Hewagama, Patel, Yarlagadda, Strickland, & Richardson, 2009), *SERPING1* (Sárvári et al., 2011) and *CXCL10* (Sentman, Meadows, Wira, & Eriksson, 2004). Estrogen also induces the expression of *TLR8* and all endosomal *TLRs* in female (Cervantes et al., 2012). Similar observations were discovered in our study regarding the higher expressions of these immune-related genes in primary monocytes of female compared to male, suggesting the role of sex hormones in regulating the expression of immune response-related genes in primary monocytes of female.

The results of this study also indicated that the innate immune-related pathways including "Toll-like receptor signaling pathway" and "Cytokine-cytokine receptor interaction" were the most significant pathways for highly expressed genes in female. It has been reported that the estrogen modulates several inflammatory pathways in female

(Chakrabarti, Lekontseva, & Davidge, 2008). Moreover, the global gene expression analysis of B cells in healthy male and female showed that "Cytokine-cytokine receptor interaction" and "Toll-like receptor signaling pathway" were involved in significant signaling pathways for highly expressed genes in healthy female compared to male and suggested that these pathways are associated with estrogen (Fan et al., 2014). Similarly, our analysis showed that in primary monocytes, the "Cytokine-cytokine receptor interaction" and "Toll-like receptor signaling pathways" were significant for highly expressed genes in female which may be related to estrogen level in female as reported by Fan et al. (2014).

In addition, the apoptosis-related genes *JUN* and *STAT1* were found to be highly upregulated and downregulated in male compared to female, respectively. The expression patterns of *JUN* and *STAT1* in male and female were further validated through qRT-PCR which confirmed the finding of our RNA-Seq results. *JUN* is the critical component of AP-1 transcription factors which regulates cell proliferation, transformation and apoptosis (Shaulian & Karin, 2002). *JUN* promotes the transition of cell cycle from G1 phase to S phase through upregulating cyclin D1 expression and suppress the *p53* (Tumor Protein P53) and *p21* (Cyclin-Dependent Kinase Inhibitor 1) functions. The *JUN* protein is involved in both induction and prevention of apoptosis. *JUN* has been shown to induce the pro-apoptotic proteins *FASL* and *BIM* (Tomicic et al., 2015). However, the *JUN* anti-apoptotic activity is related to the deficiency of *JUN* which causes massive hepatocyte apoptosis (Chen, 2003). *STAT1* has been known to regulate several growth factors, biological responses and cytokines in mammals (Lim et al., 2008). *STAT1* stimulate cell death through transcriptional activation of genes encoding proteins involved in regulating or promoting cell death, such as death receptors, *Bcl_xL* (Bcl-Associated Death Promoter) and *iNOS* (inducible Nitric Oxide Synthase*)*, as well as the genes which are involved in cell cycle arrest, such as *p21*. *STAT1* also interact with *TRADD* (Tumor Necrosis Factor

Receptor Type 1 Associated Death Domain Protein) and *p53* or *HATs* (Histones Acetyltransferase) and *HDACs* (Histones Deacetylase) proteins which are directly or indirectly involved in apoptotic cell death and increase the apoptotic response by inhibiting pro-survival *NF-κB* (Nuclear Factor Kappa B Subunit 1) signaling (Kim & Lee, 2007). The different expression patterns of *JUN* and *STAT1* in primary monocytes of male compared to female observed through our RNA-Seq analysis and qRT-PCR validation experiment may suggest the existence of sex differences in expression of those genes in primary monocytes.

In summary, utilization of deep RNA-Seq approach, a genome-wide transcriptome expression of human primary monocytes from healthy subjects was profiled and a comprehensive overview of immune-related genes in human primary monocytes under healthy state was provided. In addition, gender-based comparison of gene expression pattern in human primary monocytes revealed that the innate immune-related genes are not equally expressed in male and female. This finding provide new insights into gender disparity in innate immune response of human primary monocytes.

CHAPTER 5: GENE EXPRESSION PROFILING OF HUMAN PRIMARY MONOCYTES FROM XLA PATIENTS

5.1 Introduction

X linked agammaglobulinemia (XLA) is a rare X-linked genetic disorder that affects the male. XLA is cussed by mutations in gene coding for *BTK* (Bruton's tyrosine kinase) located in the Xq21.3-q22 region, which is essential for B cell development and function (Vetrie et al., 1993). XLA disorder is characterized by few or absent of peripheral B cells leading to decrease all serum immunoglobulin levels and recurrent infections with encapsulated bacteria and enteroviruses (Ochs & Smith, 1996). The expression of *BTK* is not limited to B cells. It is also expressed in other immune cells such as NK cells (Bao et al., 2012), neutrophils (Honda et al., 2012) and monocytes/macrophages (Koprulu & Ellmeier, 2009). It is well known that *BTK* mutations affected B cell development and functions in XLA patients (Lee et al., 2010; Mohamed et al., 2009; Ochs & Smith, 1996). However, the effect of *BTK* deficiency in primary monocytes of XLA patients is not fully understood and there is no or limit data available on transcriptome profile of primary monocytes from XLA patients. Moreover, despite the evidence for a role of lncRNAs in the immune system regulation (Derrien et al., 2012; Karapetyan et al.,2013; Guttman et al., 2009), the functions of lncRNAs in primary monocytes of XLA has not been studied yet. Through this study, the genome-wide transcriptome profile of primary monocytes from XLA patients was generated using deep RNA-Seq analysis. Furthermore, a comparative gene expression analysis of primary monocytes was conducted between RNA-Seq datasets of XLA patients and healthy male subjects to look into the possible differences in expression patterns of protein-coding genes and lncRNAs in XLA patients compared to healthy subjects which may affect the function of primary monocytes in these patients. The schematic representation of workflow of the bioinformatics analysis

procedure is described in Figure 5.1. The analysis results presented in this chapter is to achieve the second, third and fourth objectives of the thesis as specified in Chapter 1.

5.2 Transcriptome profile of primary monocytes from XLA patients

Three male patients with XLA disorder were involved in this study. The peripheral blood samples were collected and the classical monocytes ($CD14^{++}CD16^{-}$) were isolated from peripheral blood mononuclear cells (PBMCS) of patients using a negative selection technique. The purity of the isolated monocytes was checked by flow cytometry analysis. From each sample, more than 90% purity was obtained. One example of the flow cytometry analysis result is shown in Figure 5.2.

The total RNA was extracted from monocytes and the quality, quantity, and integrity of the extracted RNA were separately checked (Table 5.1). Deep poly$(A)^{+}$ paired-end RNA-Seq was conducted on the extracted RNAs. The quality of sequences reads from all samples were checked and the low quality reads and adaptors were removed from the reads. The trimmed reads from each samples were then mapped to the reference genome and assembled into a transcriptome. An average 90% of the reads were aligned to the human reference genome (Ensembl GRCH38.79) (Table 5.2). After aligning reads to the reference genome, the transcripts were reconstructed. The transcripts assembly files from all 3 patients were then merged together to form a single non-redundant set of transcripts. The expression levels of all the transcripts were quantified across samples in each dataset. By applying the FPKM >0.1 threshold, an average of 17,510 genes (including protein-coding, non-coding genes, and pseudogenes) and 62,367 transcripts (including annotated transcripts and novel transcripts) were obtained for the XLA patients. The distribution of identified genes and transcripts is shown in Figure 5.3.

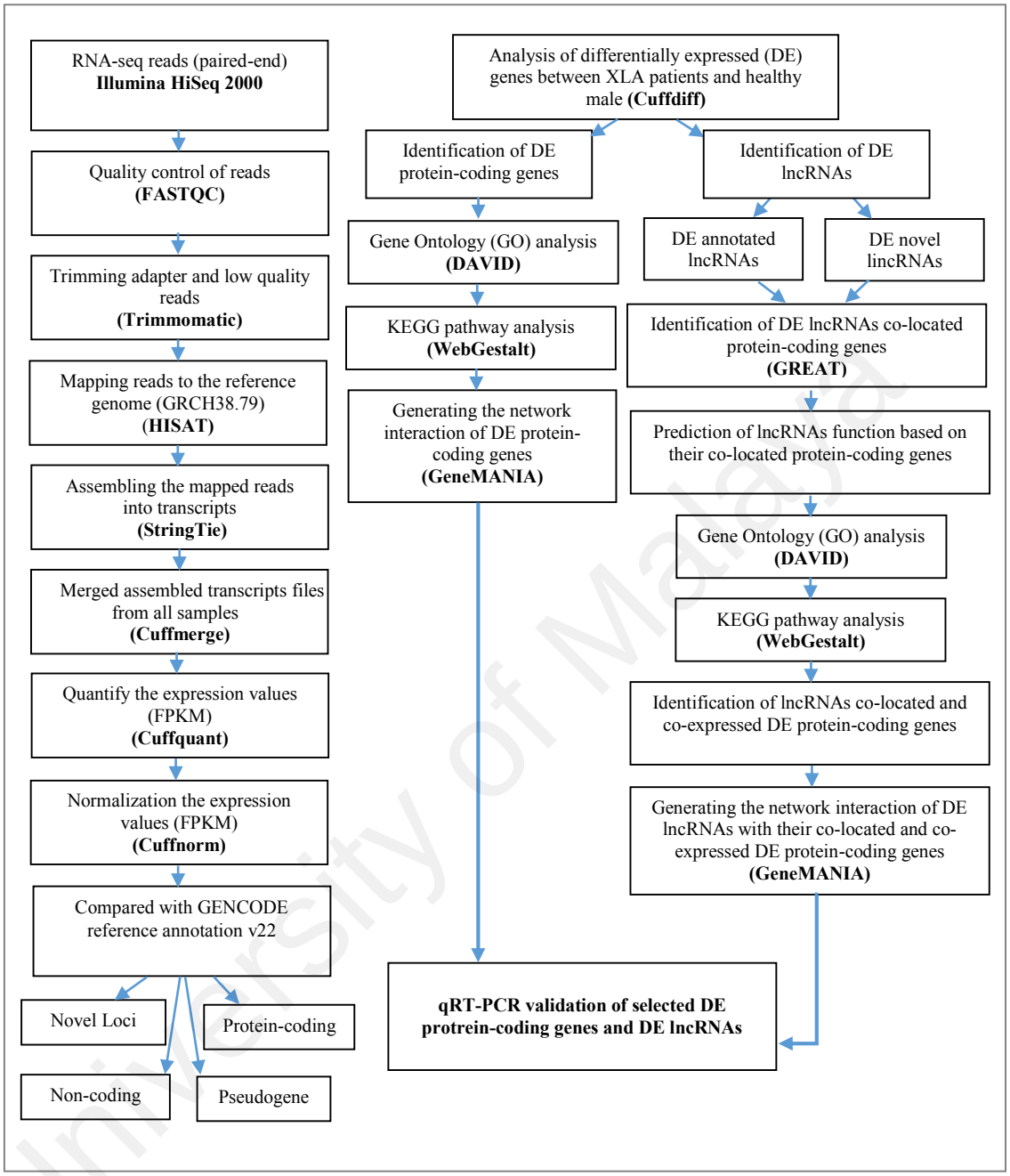

Figure 5.1: Schematic representation of workflow of bioinformatics analysis of RNA-Seq dataset of primary monocytes from XLA patients.

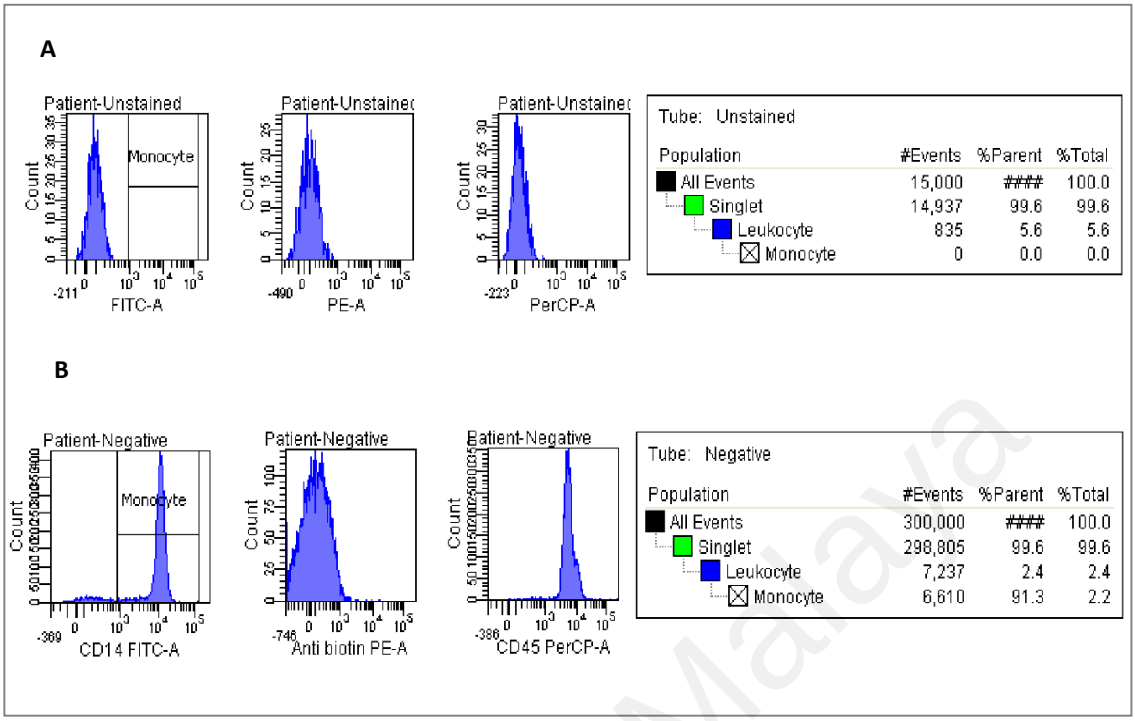

Figure 5.2: Flow cytometry analysis of isolated monocyte from XLA patient. The results show the purity and percentage of monocytes a during 2 steps monocyte isolation process involving, A: Before magnetically labeling the non-monocytes cells (labeled as Patient-Unstained) and, B: After magnetically separation of monocytes from non-monocytes cells through negative selection (labeled as Patient-Negative).

Table 5.1: The quality control results of RNA samples from patient's samples.

Sample_ID	Vol (µl)	Nano Drop				Qubit RNA BR		
		Con (ng/µl)	Total (µg)	A260/280	A260/230	Con (ng/µl)	Total (µg)	Bioanalyzer RIN no
Patient_01	30	98.2	2.946	2.06	1.73	102.0	3.060	9.5
Patient_0	35	94.4	3.304	2.06	1.92	86.2	3.017	9.9
Patient_03	35	214.3	7.5	2.07	1.86	294.0	10.29	9.7

Table 5.2: Summary of alignment results of RNA-Seq datasets from XLA patients.

Sample ID	Percentage of mapped reads
Patient_01	88%
Patient_02	94%
Patient_03	91%

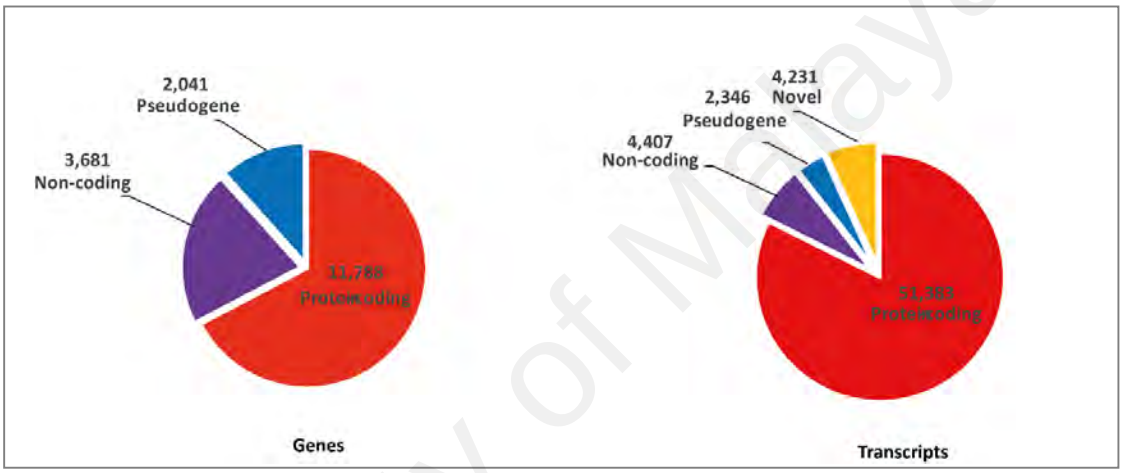

Figure 5.3: Transcriptome of primary monocytes from XLA patients. Pie chart representing the number of diverse classes of genes and transcripts identified in primary monocytes of XLA patients.

5.2.1 lncRNAs expressed in primary monocytes of XLA patients

A comparison of the merged assembled transcripts from XLA patients with the annotated lncRNAs from GENCODE (version 22) showed the expression of 3,363 annotated lncRNAs in patients. By using multi-step mapping and filtering criteria (as described in methodology section 3.9.3.2, pages 51-52) the expression of 430 potentials novel lincRNAs in the patients were also identified. The chromosome-wise distribution of identified annotated [Figure 5.4 A] and novel lincRNAs [Figure 5.4 B] showed that they are distributed across all chromosomes, mostly on chromosomes 1, 2, and 19.

5.3 Differentially expressed (DE) genes between XLA patients and healthy subjects

In order to investigate the differences of gene expression patterns in primary monocytes between XLA patients and healthy subjects, differentially gene expression analysis was performed between RNA-Seq datasets of 3 XLA patients and 3 healthy male subjects generated in our study. The analysis was conducted using Cuffdiff and the genes with q-value ≤ 0.01 and $\log_2$ fold-change ≥ 1 or ≤ -1 were defined as differentially expressed (DE) genes. Our analysis focused on protein-coding genes as well as lncRNAs, since they have been reported to contribute in immune-related disease (Hrdlickova et al., 2014; Wapinski & Chang, 2011). A total of 1,827 DE protein-coding genes, 95 DE annotated lncRNAs and 20 DE novel lincRNAs were identified in XLA patients compared to the healthy subjects.

5.3.1 DE protein-coding genes

Out of the 1,827 DE protein-coding genes, 859 genes were upregulated and 968 genes were downregulated in XLA patients compared to the healthy subjects. The chromosomal distribution of identified DE protein-coding genes is shown in Figure 5.5.

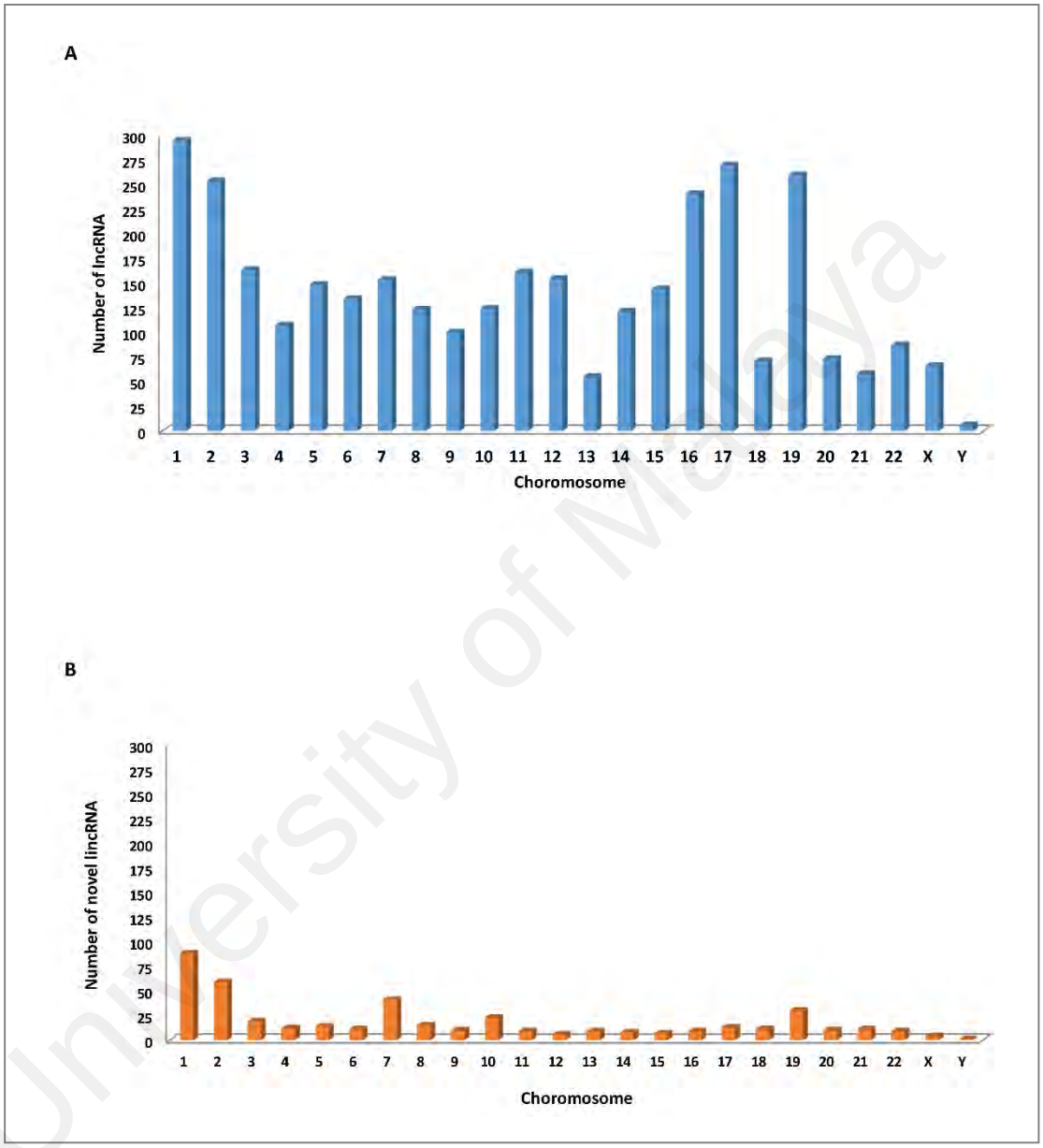

Figure 5.4: The chromosomal distribution of expreseed lncRNAs in primary monocytes of XLA patients. A: Annotated lncRNAs. B: Novel lincRNAs.

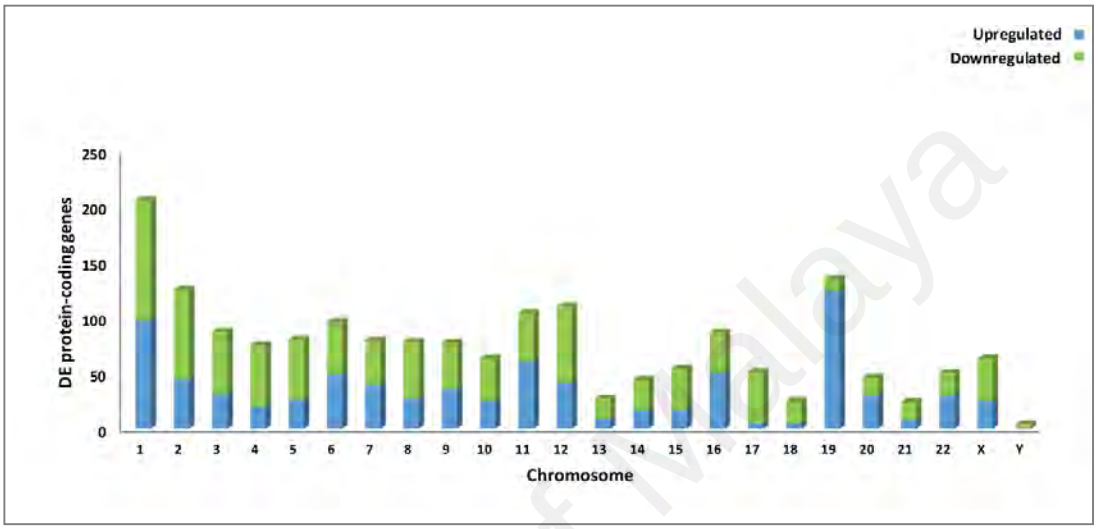

Figure 5.5: Chromosomal distribution of DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. DE protein-coding genes distributed across all chromosomes.

The top 10 upregulated and downregulated protein-coding genes ranked by $\log_2$ fold-change in XLA patients compared to the healthy subjects are presented in Table 5.3. The expression of *BTK* was detected to be significantly downregulated ($\log_2$ fold-change < -7) in XLA patients compared to healthy subjects (Table 5.3). The hierarchical clustering analysis was applied on DE protein-coding genes which revealed distinct transcription expression profiles between XLA patients and healthy subjects (Figure 5.6). The functional consequences of the DE protein-coding genes in XLA patients compared to healthy subjects was characterized through GO analysis based on the biological processes category, using DAVID software. First, the highest hierarchical level 1 of biological process ontology (DAVID category-GOTERM_BP_1) was selected which give a broad overview information for genes functions as presented in Figure 5.7. The 859 upregulated genes were assigned to 5 general GO terms including "metabolic process", "cellular process", "cellular component biogenesis", "death" and "localization" [Figure 5.7 A], while the 968 downregulated genes were assigned to 8 general GO terms mainly involved in "immune system process", "regulation of biological regulation" and "signaling" [Figure 5.7 B]. Next in order to characterize the specific GO terms in the biological process for both upregulated and downregulated genes, the DAVID category-GOTERM_BP_FAT level was analyzed. The top 10 significant (adjP < 0.01) GO terms for upregulated and downregulated genes are shown in Figure 5.8. The significant GO terms for upregulated genes were related to mitochondrial function and organization included; "oxidative phosphorylation", mitochondrial ATP synthesis coupled to electron transport", and "electron transport chain", as well as "apoptotic" process and "response to oxidative stress" [Figure 5.8 A]. The expression of several apoptosis-related genes such as *BAX* (BCL2 Associated X Protein) and *BAD* (BCL2 Associated Agonist Of Cell Death) and oxidative stress response-related genes such as *SOD1* (Superoxide Dismutase 1, Soluble), *GPX1* (Glutathione Peroxidase 1), *GPX4* (Glutathione Peroxidase 4), PRDX1

Table 5.3: List of top 10 upregulated and downregulated protein-coding genes in primary monocytes of XLA patients compared to healthy subjects.

Gene_Name	Chromosome	FPKM-Healthy	FPKM-Patients	log_2(fold-change)	q_value
Upregulated					
KCNMA1	10	0.0519476	1.49041	4.84251	0.000847
UTF1	10	0.122524	2.46573	4.33088	0.000847
LRRC26	9	0.0843423	1.49147	4.14434	0.000847
CLIC3	9	0.46531	8.22455	4.14367	0.000847
COL6A2	21	0.206435	3.4272	4.05327	0.000847
GZMM	19	0.560778	9.18781	4.03422	0.000847
CNIH3	1	1.50141	23.9122	3.99336	0.008879
SH2D2A	1	0.243494	2.91865	3.58334	0.000847
GZMB	14	5.9343	67.5878	3.50961	0.000847
RPS26	12	22.1783	250.317	3.49654	0.000847
Downregulated					
LIMS3	2	36.5789	0.0624726	-9.19357	0.000847
BTK	x	16.3307	0.0912267	-7.483914	0.004354
F8A2	X	38.4799	0.212644	-7.49952	0.000847
RASA4	7	27.2075	0.224622	-6.92036	0.00153
RGPD6	2	2.47584	0.0281969	-6.45624	0.006337
RPL41	7	527.127	6.35329	-6.3745	0.000847
GOLGA6L4	15	0.81077	0.0098742	-6.35947	0.000847
U2AF1	21	27.615	0.468034	-5.8827	0.000847
FAM156B	X	18.1565	0.398487	-5.50981	0.000847
AC138969.4	16	78.8136	1.73558	-5.50496	0.000847

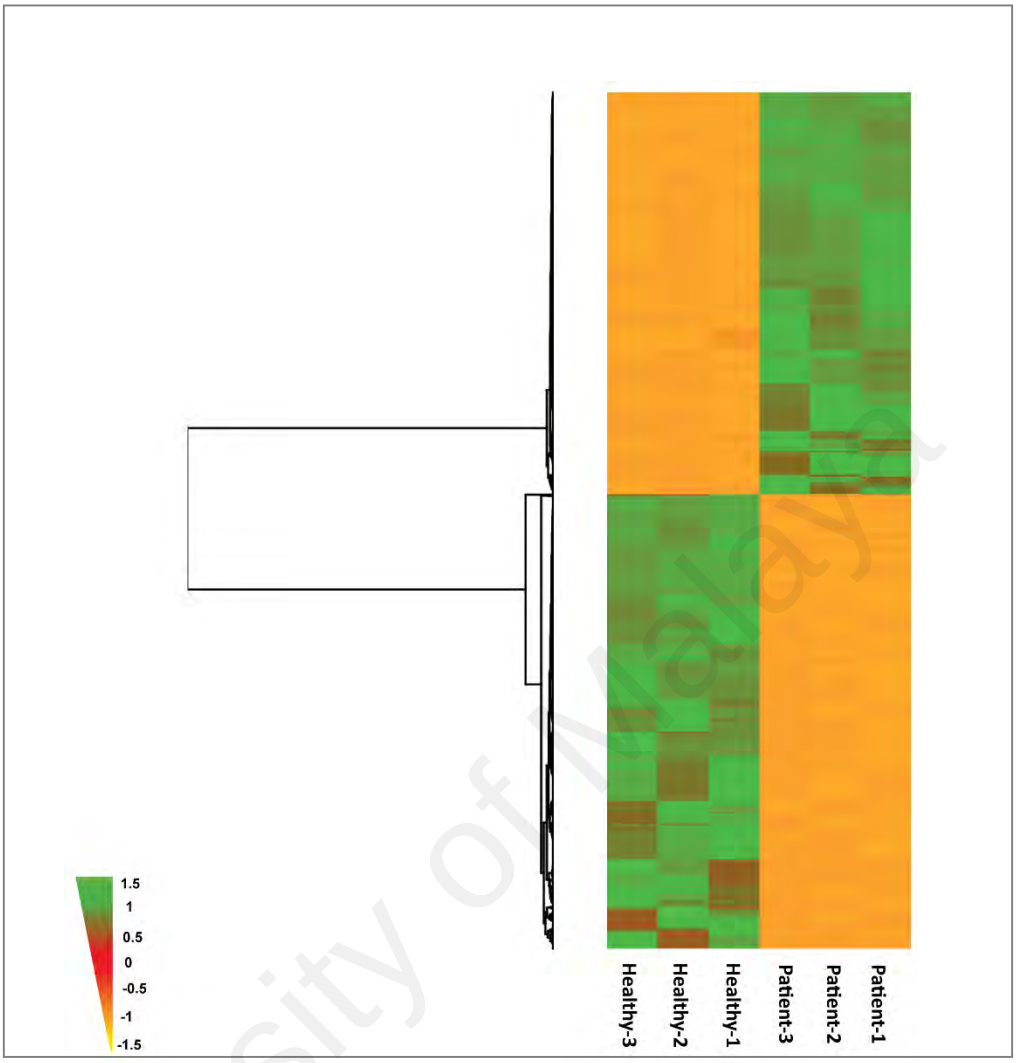

Figure 5.6: Hierarchical clustering of DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. Each row in the heatmap represents a gene and each column represents a separate sample. The colours correspond to the $\log_2$ of FPKM values (green colour for upregulated genes and orange colour for downregulated genes).

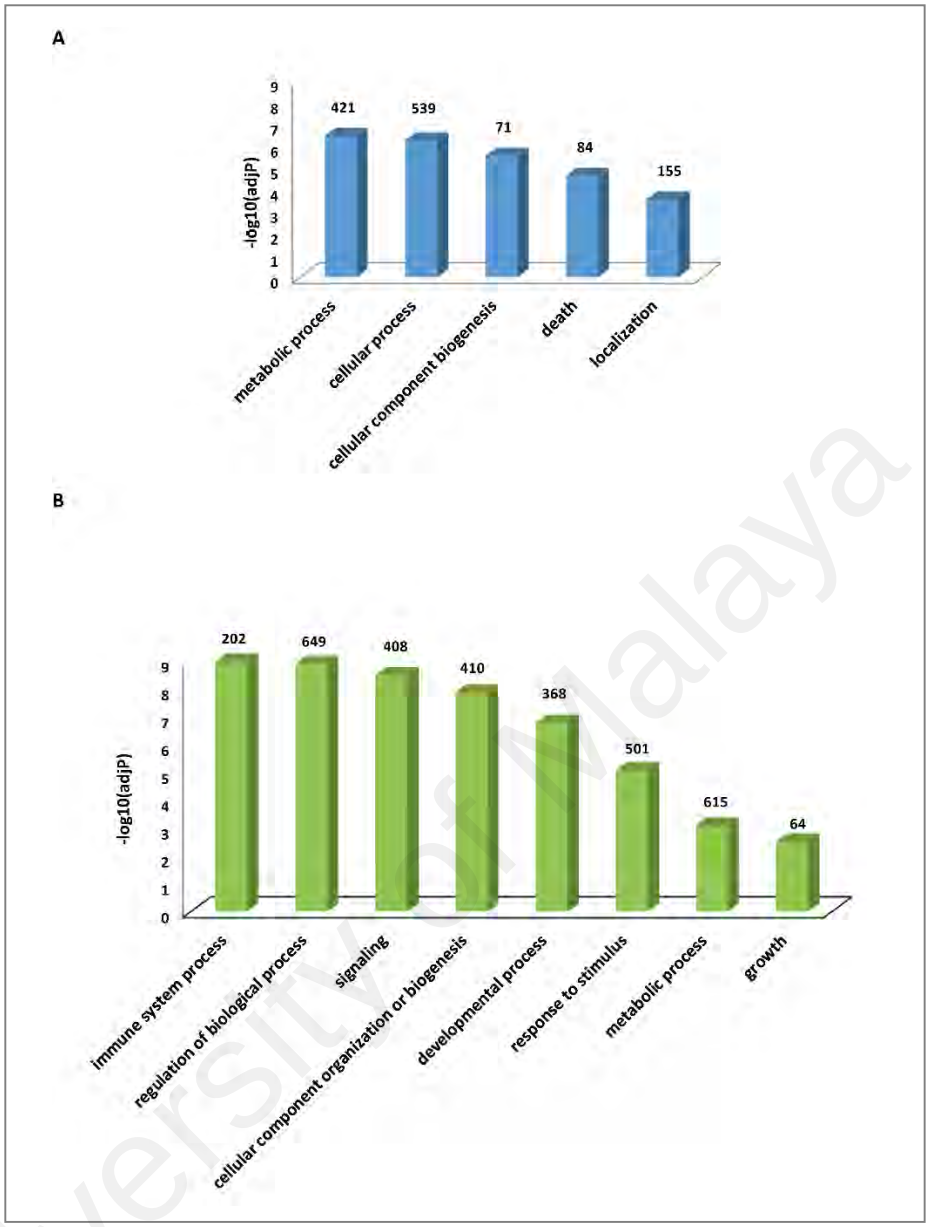

Figure 5.7: The GO analysis of DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. A: The significant GO biological process terms (DAVID category-GOTERM_BP_1) enriched for upregulated genes. B: The significant GO biological process terms (DAVID category-GOTERM_BP_1) enriched for downregulated genes. The number of identified DE protein-coding genes enriched in each GO terms is depicted above the bars in the figure.

(Peroxiredoxin 1), *PRDX5* (Peroxiredoxin 5) were observed to be significantly upregulated in XLA patients compared to healthy subjects. However, the GO terms for downregulated genes were significantly related to monocytes immune system functions including: "intracellular signaling cascade", "immune response" and "innate immune response" [Figure 5.8 B]. To further evaluate the biological roles of DE protein-coding genes, a KEGG pathway analysis was performed (Figure 5.9). The upregulated genes were enriched in 9 pathways, most significantly in "Oxidative phosphorylation" [Figure 5.9 A] which is related to the mitochondrial functions. The oxidative phosphorylation system consists of five protein complexes including I, II, III, IV, and V. Through our analysis, the upregulation of the several components of complexes I, III, IV, and V were observed in primary monocytes of XLA patients compared to the healthy subjects as presented in Table 5.4. The Go analysis also revealed that the downregulated genes were enriched in 29 pathways, most significantly in several immune-related pathways such as "Fc gamma R-mediated phagocytosis", "Chemokine signaling pathway", "Toll-like receptor signaling pathway" and "MTOR signaling pathway" [Figure 5.9 B]. The core downregulated genes contributing to the enrichment of the immune-related pathways in primary monocytes of XLA patients is presented in Table 5.5.

5.3.1.1 Gene interaction network of DE protein-coding genes

To explore the dysregulated gene interactions in XLA patients, the interaction network was generated on DE protein-coding genes which were enriched in significant upregulated and downregulated KEGG pathways in XLA patients compared to healthy subjects (Figure 5.10). The network contains 1,601 interactions between 78 upregulated and 103 downregulated genes and shows the relationships between DE protein-coding genes in terms of their co-expression.

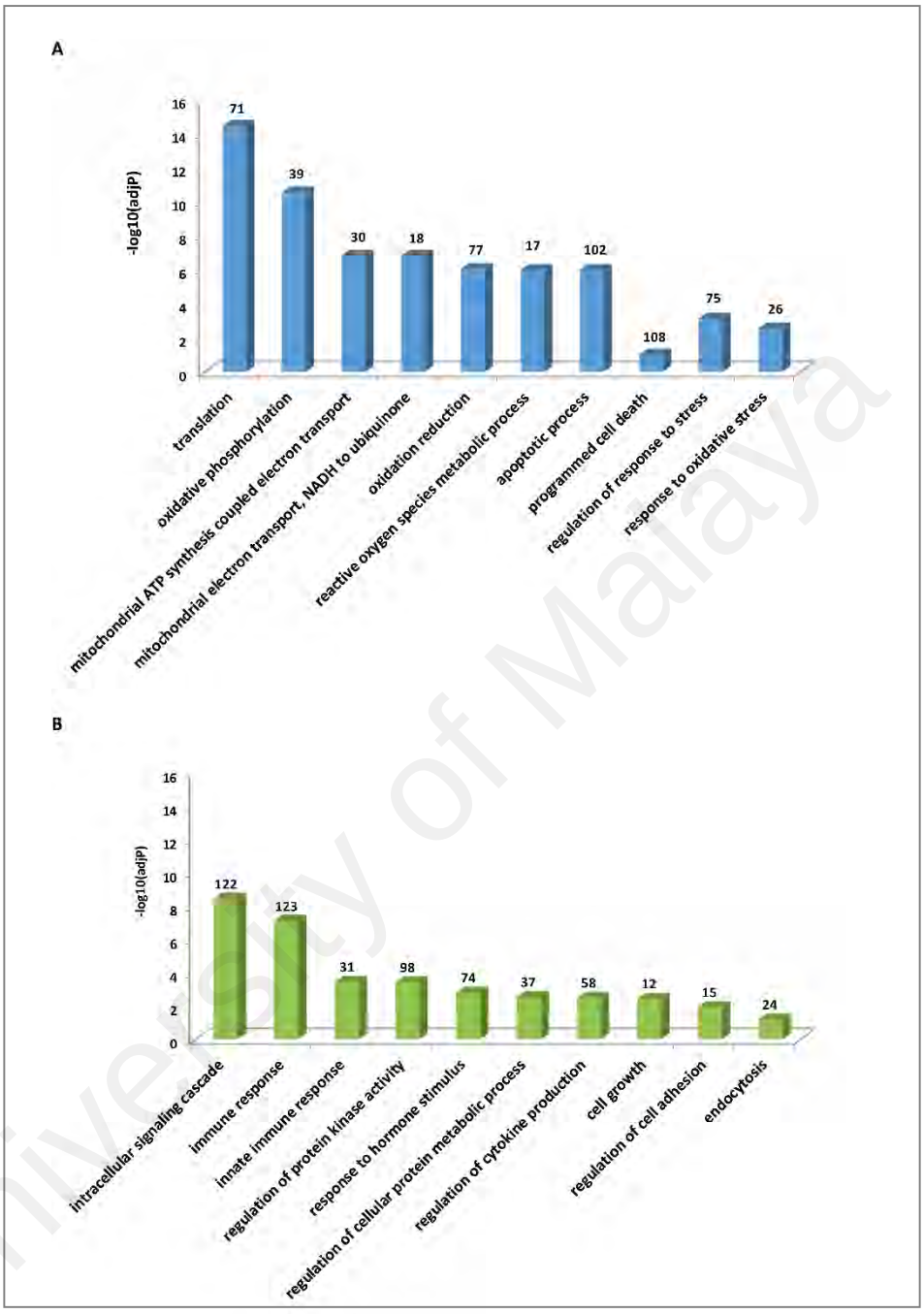

Figure 5.8: The specific GO biological process terms (DAVID category-GOTERM_BP_FAT) of DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. A: Upregulated genes. B: Downregulated genes. The number of identified DE protein-coding genes enriched in each GO term is depicted above the bars in the figure.

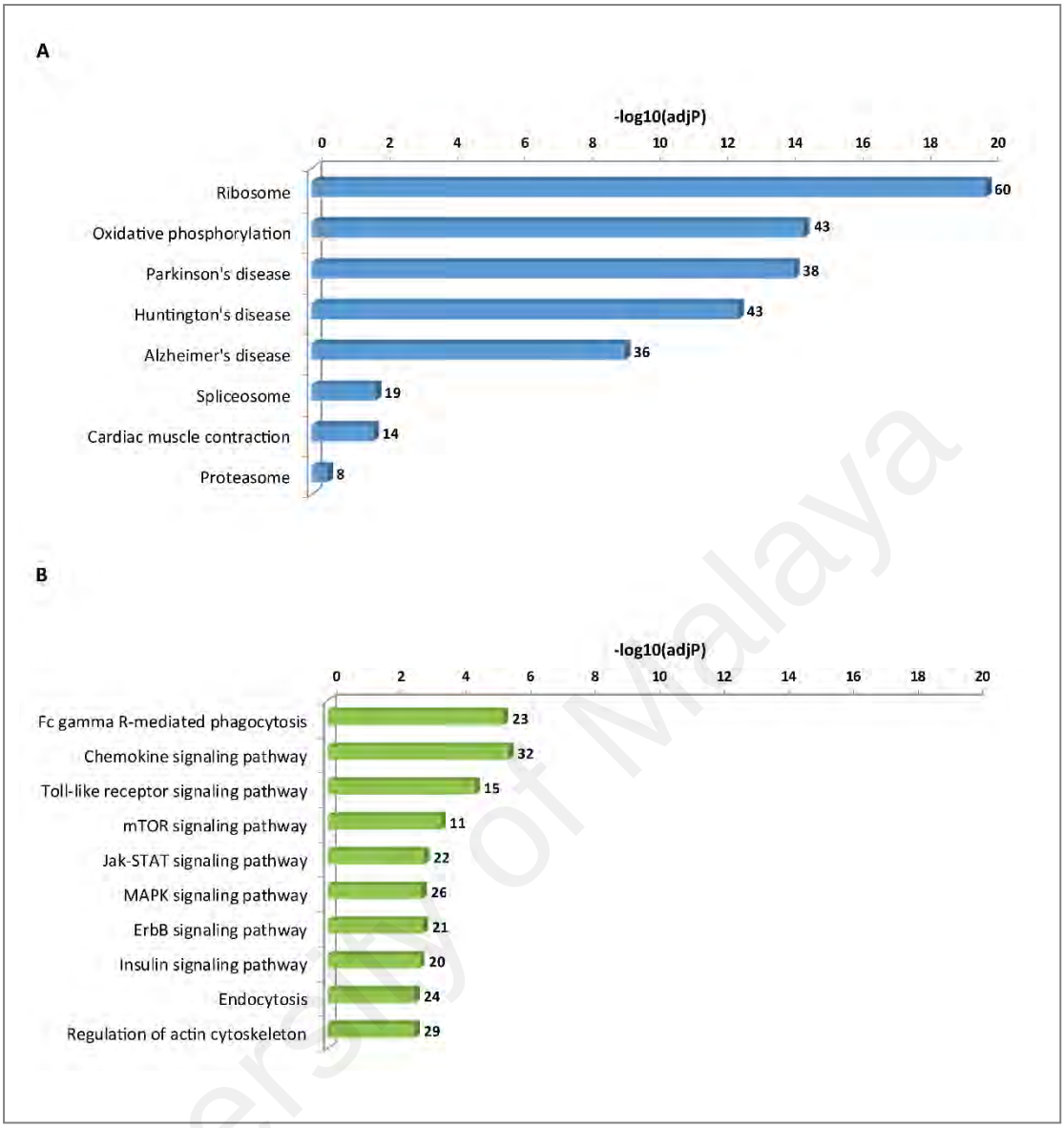

Figure 5.9: The KEGG pathway analysis of the DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. A: The significant KEGG pathway terms enriched for upregulated genes. B: The significant KEGG pathway terms enriched for downregulated genes. The numbers in the brackets indicated the total numbers of genes available in the KEGG database for each pathway terms. The number of identified DE protein-coding genes enriched in each KEGG pathway is depicted above the bars in the figure.

Table 5.4: The upregulated genes involved in Oxidative Phosphorylation pathway in primary monocytes of XLA patients compared to healthy subjects.

Oxidative Phosphorylation system subunits	**Genes**
Complex I: NADH Deshydrogenase	*NDUFA1, NDUFA12, NDUFA2, NDUFA3, NDUFA4, NDUFA6, NDUFA8, NDUFAB1, NDUFB11, NDUFB4, NDUFB5, NDUFB7, NDUFB8, NDUFS4, NDUFS5, NDUFS6, NDUFS7*
Complex III: Cytochorom C Reductase	*UQCR10, UQCRB, UQCRFS1, UQCRH, UQCRHL*
Complex IV: Cytochorom C Oxidase	*COX14, COX17, COX4I1, COX5A, COX5B, COX6B1, COX6C, COX7A2, COX7C*
Complex V: ATPase	*ATP1A3, ATP5D, ATP5E, ATP5G2, ATP5G3, ATP5H, ATP5I, ATP5J, ATP6V0B, ATP6V0E1, ATP6V1F, ATPIF1*

Table 5.5: The downregulated genes involved in immune-related pathways in primary monocytes of XLA patients obtained from comparison with healthy subjects.

Pathway	Genes
Fc gamma R-mediated phagocytosis	*ASAP1*, *FCGR2A*, *GAB2*, *DOCK2*, *DNM1L*, INPP5D, *MAPK1, MAP2K1, PAK1, PIKFYVE, PIK3CG, PIK3R5, PLCG1, PRKCB, PRKCE, PTPRC, SYK, LYN, VASP, VAV3, PIK3R1, RAF1, AKT2*
Chemokines signaling	*JAK2, ROCK2, ADCY6, ADCY7, ADCY9, ADRBK2, CXCL16, CXCR4, DOCK2, FOXO3, GRB2, GNB4, GNG12, GNG2, CXCR1, CXCR2, MAPK1, MAP2K1, NRAS, PAK1, PIK3CG, PIK3R1, PIK3R5, PRKCB, PRKX, STAT2, STAT5B, ROCK1, SOS1, SOS2, BRAF, VAV3*
Toll-like receptors signaling	*TBK1, IKBKE, JUN, MAPK1, MAP2K1, MAP2K4, MYD88, PIK3CG, PIK3R1, PIK3R5, TLR1, TLR2, TLR4, TLR5, TLR7, FOS*
MTOR signaling	*RICTOR, HIF1A, MTOR, MAPK1, PIK3CG, PIK3R1, PIK3R5, RPS6KA3, TSC1, ULK2, BRAF*
Jak-STAT signaling	*CREBBP, CBL, EP300, JAK1, JAK2, CSF2RB, GRB2, IL10RA, IL13RA1, IL6R, IL6ST, PIK3CG, PIK3R1, PIK3R5, PRLR, STAT2, STAT5A, STAT5B, SOS1, SOS2, SOCS4, SOCS7*
MAPK signaling	*RAPGEF2, ATF2, DUSP1, DUSP6, FLNB, GRB2, GNA12, GNG12, HSPA1A, JUN, MAPK1, MAP2K1, MAP2K4, MAP3K1, MAP4K4, NRAS, PAK1, PRKCB, PRKX, RPS6KA3, SOS1, SOS2, TGFBR1, FOS, BRAF*
ErbB signaling	*CBL, EREG, GRB2, JUN, MTOR, MAPK1, MAP2K1, MAP2K4, NRAS, PAK1, PIK3CG, PIK3R, PIK3CG, PIK3R, PIK3R5, PLCG1, PRKCB, STAT5A, STAT5B, SOS1, SOS2, ABL2, BRAF*
Insulin signaling	*CBL, GRB2, INPP5D, IRS2, MTOR, MAPK1, MAP2K1, NRAS, PDE3B, PIK3CG, PIK3R1, PIK3R5, PRKCI, PRKX, PTPRF, SOS1, SOS2, SOCS4, TSC1, BRAF*
Endocytosis	*ARAP2, ASAP1, ACAP2, CBL, DNAJC6, EHD3, IQSEC1, RAB11, FIP1, ADRBK2, CXCR4, CLTCL1, DNM1L, EPN2, HSPA1A, CXCR1, CXCR2, LDLR, NEDD4, PIKFYVE, PRKCI, RNF41, SMAP2, TGFBR1, VPS36*
Regulation of actin cytoskeleton	*IQGAP1, IQGAP2, ARHGEF12, ROCK2, ACTN1, CYFIP1, FGFR1, FN1, GNA12, GNA13, GNG12, ITGA4, ITGAV, ITGB1, MAPK1, MAP2K1, NRAS, PAK1, PIKFYVE, PIK3CG, PIK3R1, PIK3R5, PPP1R12A, ROCK1, SSH2, SOS1, SOS2, BRAF, VAV3*

5.3.2 DE lncRNAs

The total of 95 DE annotated lncRNAs were detected in primary monocytes between XLA patients and healthy subjects in which 56 and 39 lncRNAs were upregulated and downregulated in XLA patients, respectively (Table 5.6). The identified DE annotated lncRNAs dispersed across all chromosomes except chromosomes 11, 13 and Y [Figure 5.11 A]. Several lncRNAs with well-known functions were detected among DE lncRNAs in XLA patients. Such lncRNAs include; *HOTAIRM1* (HOXA Transcript Antisense RNA, Myeloid-Specific 1), *DANCR* (Differentiation Antagonizing Non-Protein Coding RNA), *GAS5* (Growth Arrest Specific 5), *LINC-PINT* (Long Intergenic Non-Protein Coding RNA, P53 Induced Transcript), *RMRP* (RNA Component Of Mitochondrial RNA Processing Endoribonuclease), and *HEIH* (Hepatocellular Carcinoma Associated Transcript) which were upregulated, while *TUG1* (Taurine Upregulated 1), was downregulated in XLA patients compared to the healthy subjects. In addition, the expression of 20 DE novel lincRNAs were identified between XLA patients and healthy subjects. Among the 20 DE novel lincRNAs, 5 lincRNAs were upregulated and 15 lincRNAs were downregulated in XLA patients compared to healthy subjects, respectively (Table 5.7). The chromosomal distribution of DE novel lincRNAs indicated that they are not dispersed across all chromosomes [Figure 5.11 B]. The hierarchical clustering analysis of DE annotated lncRNAs and DE novel lincRNAs revealed distinct transcription expression profiles between XLA patients and healthy subjects (Figure 5.12).

5.3.2.1 DE lncRNAs co-located and co-expressed with protein-coding genes

lncRNAs known to coordinate the regulation of neighboring protein-coding genes (co-located genes) (Wang et al., 2011). To identify the potential function of DE lncRNAs,

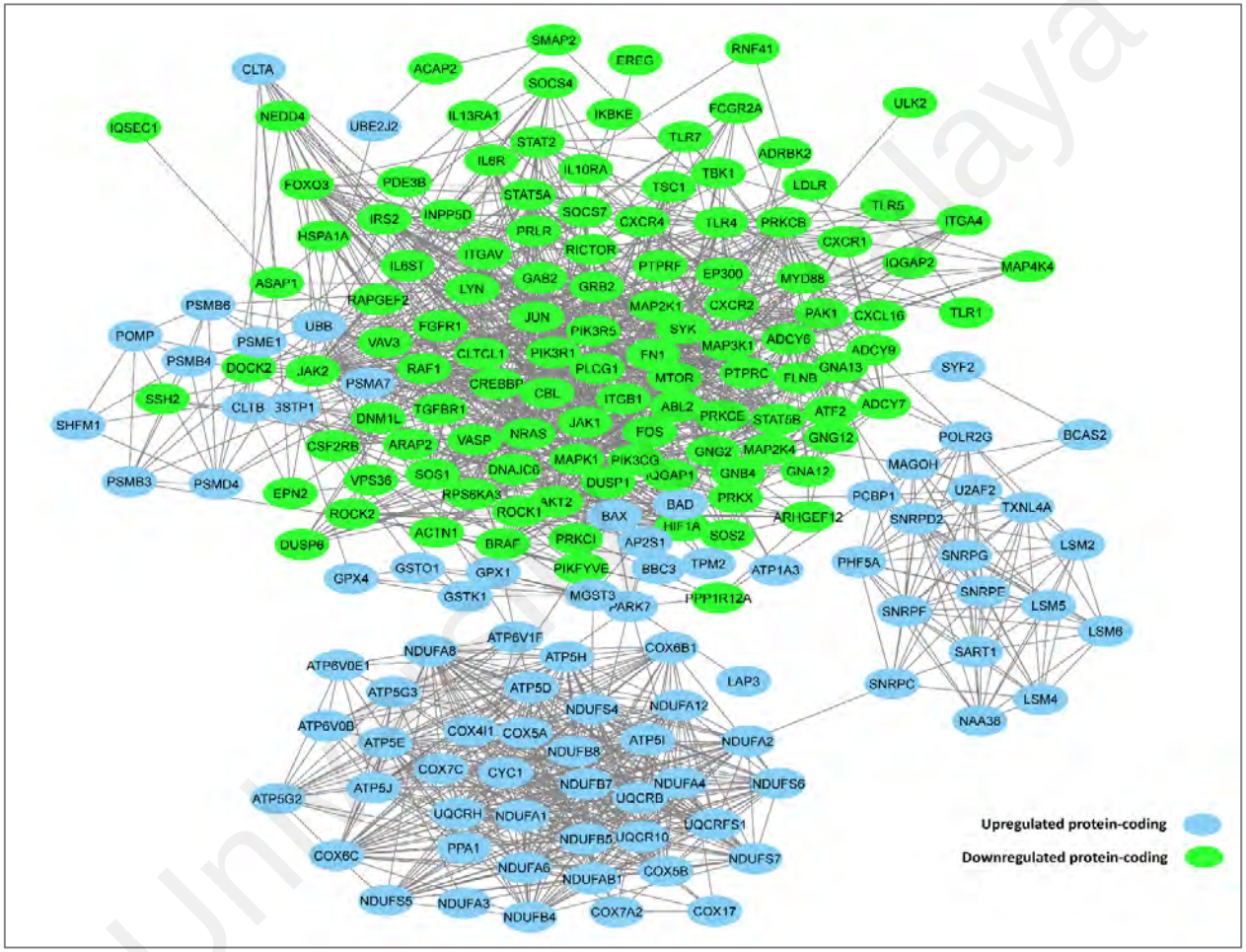

Figure 5.10: Interaction network analysis of DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects. The DE protein-coding genes were connected in a network based on their co-expression.

Table 5.6: List of identified DE annotated lncRNAs in primary monocytes of XLA patients compared to healthy subjects.

lncRNA_Name	Chromosome	FPKM-Healthy	FPKM-Patients	log_2(fold-change)	q-value
Upregulated					
RMRP	9	6.63546	73.1738	3.46306	0.000847
YTHDF3-AS1	8	0.105048	1.05009	3.32140	0.009267
AP001505.10	21	0.229011	1.96353	3.09996	0.005346
CTB-12A17.2	19	0.564883	4.15868	2.88010	0.000847
SNHG9	16	2.39478	15.6457	2.70780	0.002157
RP1-92O14.3	1	0.137191	0.887144	2.69298	0.006778
LINC01506	9	2.52923	16.0833	2.66880	0.000847
SNHG19	16	1.99338	12.6204	2.66246	0.000847
RP11-297C4.6	16	1.6566	10.1467	2.61472	0.000847
RP11-669M16.1	4	0.236648	1.41391	2.57887	0.003311
RP11-291B21.2	12	0.191131	1.05584	2.46576	0.006337
AC012442.5	2	6.04499	33.1964	2.45721	0.000847
RP11-410L14.2	8	0.588439	3.11348	2.40356	0.006778
PCED1B-AS1	12	2.85035	14.4957	2.34642	0.002157
RP5-1171I10.5	17	8.56208	43.068	2.33058	0.006337
LINC01023	5	0.67313	3.23143	2.26322	0.006778
SNHG8	4	19.6066	93.4341	2.25261	0.000847
RP11-672L10.6	18	0.504262	2.35742	2.22496	0.005846
AC106782.20	16	4.63293	19.6451	2.08417	0.000847
RP13-516M14.1	17	0.73916	3.12512	2.07996	0.000847
RP5-899E9.1	7	0.405633	1.7132	2.07845	0.008879
SNHG6	8	29.5726	116.211	1.97442	0.000847
SNHG5	6	51.7555	202.157	1.96569	0.000847
AC022154.7	19	331.948	1273.91	1.94023	0.000847
RP11-1398P2.1	4	2.65339	10.1417	1.93440	0.000847
HOXB-AS1	17	0.347057	1.27214	1.87402	0.00969
TP53TG1	7	2.39511	8.74862	1.86896	0.000847
CTA-29F11.1	22	1.08188	3.85636	1.83370	0.004354

Table 5.6: Continued.

TOLLIP-AS1	11	0.528472	1.85703	1.81310	0.007649
RP11-295G20.2	1	2.42254	8.31042	1.77840	0.002758
CTC-246B18.10	19	138.734	458.677	1.72516	0.000847
LINC00920	16	0.132636	0.437277	1.72107	0.00969
CTB-50L17.16	19	23.3181	76.2203	1.70872	0.000847
HOTAIRM1	7	9.13987	29.6788	1.69919	0.003835
RP11-879F14.1	18	7.10556	22.8448	1.68485	0.000847
CAHM	6	0.554187	1.77142	1.67646	0.008496
LINC01088	4	2.01749	6.34959	1.65410	0.006337
LINC-PINT	7	5.64058	17.1249	1.60218	0.008231
FLJ44511	7	3.04243	9.23649	1.60212	0.000847
EPB41L4A-AS1	5	8.09109	24.3278	1.58820	0.000847
LINC01003	7	4.20227	12.3802	1.55879	0.000847
LRRC75A-AS1	17	103.969	305.71	1.55601	0.002157
LINC01503	9	4.37042	12.5658	1.52366	0.00486
AC051649.12	11	213.581	606.795	1.50642	0.002157
RP11-162A12.2	18	2.19026	6.04719	1.46516	0.003835
LINC01420	X	4.73941	13.0777	1.46433	0.000847
AP001189.4	11	32.3012	84.9398	1.39485	0.000847
RP1-56K13.3	17	384.531	995.2	1.37189	0.003835
HEIH	5	17.6792	44.436	1.32968	0.006778
AD001527.4	19	52.2792	126.585	1.27580	0.000847
RP11-85B7.5	17	24.6231	58.8617	1.25732	0.000847
RP11-1094M14.11	17	2.86529	6.82755	1.25269	0.005846
AC144831.1	17	37.7989	88.1557	1.22171	0.000847
SCAMP1-AS1	5	2.53512	5.65985	1.15871	0.009267
DANCR	4	5.01543	10.3865	1.05026	0.001491
GAS5	1	55.2683	111.56	1.01330	0.002746
Downregulated					
CASC8	8	3.54842	0.063915	-5.79488	0.000847
FAM225B	9	2.76506	0.065556	-5.39845	0.000847
RP11-14N7.2	1	4.5988	0.158307	-4.86046	0.006778
RP11-1228E12.1	17	17.4914	0.673903	-4.69796	0.000847
CTD-2331C18.5	11	3.08677	0.143761	-4.42436	0.000847

Table 5.6: Continued.

RP11-244M2.1	18	23.8855	1.23125	-4.27794	0.000847
RP11-244M2.1	18	23.8855	1.23125	-4.27794	0.000847
RP11-498P14.3	9	2.4161	0.19552	-3.62729	0.000847
LINC00115	1	84.2572	6.94845	-3.60004	0.000847
RP11-1223D19.1	2	0.202411	0.018321	-3.46576	0.002157
AC145124.2	8	1.42568	0.136477	-3.38492	0.000847
LINC01347	1	1.10349	0.110021	-3.32622	0.000847
AC096579.13	2	67.9016	7.85069	-3.11255	0.000847
RP4-545C24.1	7	0.835281	0.119885	-2.80061	0.009267
FAM225A	9	2.88302	0.415071	-2.79615	0.000847
RP1-111C20.3	6	175.7551	38.1044	-2.20553	0.000847
CTD-2270P14.1	16	25.2213	5.66398	-2.15476	0.00969
MEG3	14	0.587536	0.132241	-2.15151	0.000847
AC073046.25	2	87.6153	20.4003	-2.10259	0.000847
LINC00926	15	3.07217	0.738531	-2.05653	0.002157
AC159540.1	2	0.854071	0.207273	-2.04283	0.002758
HYI-AS1	1	22.0722	6.90947	-1.67558	0.00153
CTD-2047H16.3	17	247.087	79.3875	-1.63804	0.004354
RP11-799D4.3	17	578.327	192.083	-1.59016	0.006337
LINC00657	20	296.924	100.944	-1.55654	0.004354
AC007879.2	2	27.7313	9.89716	-1.48643	0.000847
AF127936.7	21	29.7662	10.9513	-1.44257	0.000847
AP000240.9	21	134.737	51.1635	-1.39696	0.000847
RP5-1172N10.4	X	28.9615	11.496	-1.33300	0.003835
AF011889.2	X	150.882	59.9486	-1.33162	0.000847
AC096772.6	2	8.80365	3.52059	-1.32229	0.000847
AC108676.1	3	24.1436	9.65629	-1.32210	0.00486
TUG1	22	57.2615	23.6601	-1.27511	0.000847
AC092620.2	2	46.8885	20.3398	-1.20493	0.000847
RP11-875O11.1	8	84.9	37.6905	-1.17156	0.000847
RP11-1024P17.1	3	94.3899	42.471	-1.15215	0.000847
RP11-102F4.2	8	7051.42	3214.24	-1.13344	0.000847
AC004893.11	7	13.6059	6.33346	-1.10316	0.000847
AC005062.2	7	74.3774	35.5367	-1.06556	0.004354
ADAMTSL4-AS1	1	774.882	384.675	-1.01034	0.00153

Table 5.7: List of identified DE novel lincRNAs in primary monocytes of XLA patients compared to healthy subjects.

Transcript_ID	locus	FPKM_ healthy	FPKM_ patient	log_2(fold-change)	q-value
Upregulated					
TCONS_00347815	X:10014980-10015551	0.328617	3.14001	3.25629	0.000847
TCONS_00335887	9:127423334-127424338	0.210139	1.08451	2.36763	0.004354
TCONS_00008577	1:146387174-146388092	1.08843	3.58678	1.72044	0.007649
TCONS_00041961	10:129035316-129036770	4.46955	10.062	1.17072	0.007649
TCONS_00205060	2:157392899-157404794	0.608531	1.31521	1.11189	0.002758
Downregulated					
TCONS_00063576	11:57086190-57087643	1.87575	0.066164	-4.82527	0.005846
TCONS_00090109	13:76592360-76593698	2.5435	0.124231	-4.35572	0.000847
TCONS_00139067	17:42532730-42533453	6.04171	0.345591	-4.12782	0.000847
TCONS_00212764	20:45179562-45192881	1.2481	0.095338	-3.71055	0.000847
TCONS_00086992	13:109424117-109424380	1002.54	82.1261	-3.60968	0.000847
TCONS_00109575	15:20369804-20376362	0.223018	0.018473	-3.59368	0.008879
TCONS_00326114	8:114029500-114030153	8.38758	0.868887	-3.27101	0.000847
TCONS_00134089	17:488985-490396	2.00631	0.391009	-2.35927	0.000847
TCONS_00030433	1:108091096-108100990	0.24051	0.050152	-2.26171	0.000847
TCONS_00327731	9:65645189-65664577	1.02763	0.216316	-2.24811	0.006778
TCONS_00205058	2:157345321-157368674	0.175286	0.039985	-2.13218	0.002758
TCONS_00032359	10:31586861-31595354	0.24217	0.055692	-2.12048	0.002157
TCONS_00133967	16:55719783-55720456	3.11836	0.720532	-2.11366	0.003835
TCONS_00264979	4:143336124-143336701	3.24334	0.823761	-1.97718	0.00486
TCONS_00295657	6:359346-360215	4.74063	1.48483	-1.67478	0.009267

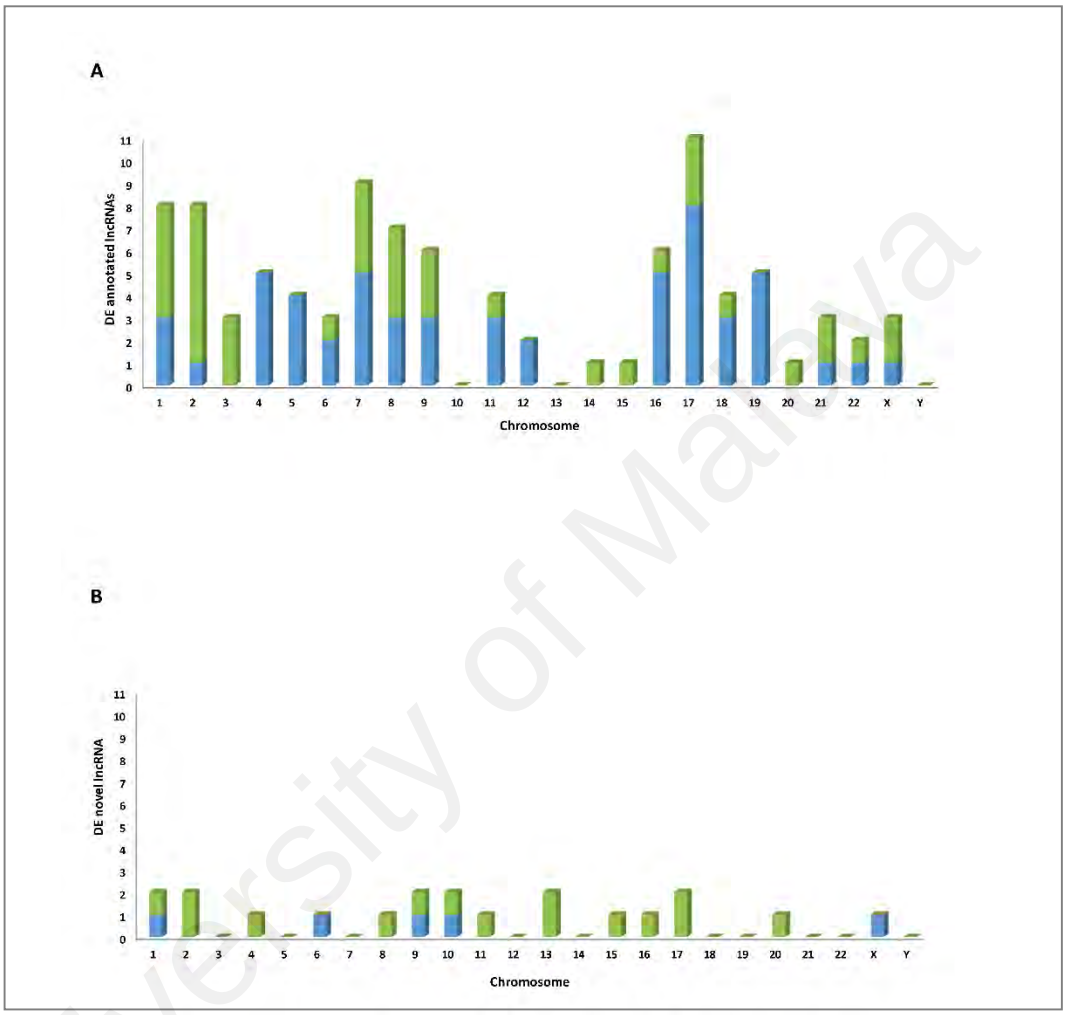

Figure 5.11: Chromosomal distributions of DE lncRNAs in primary monocytes of XLA patients compared to healthy subjects. A: DE annotated lncRNAs. B: DE novel lincRNAs.

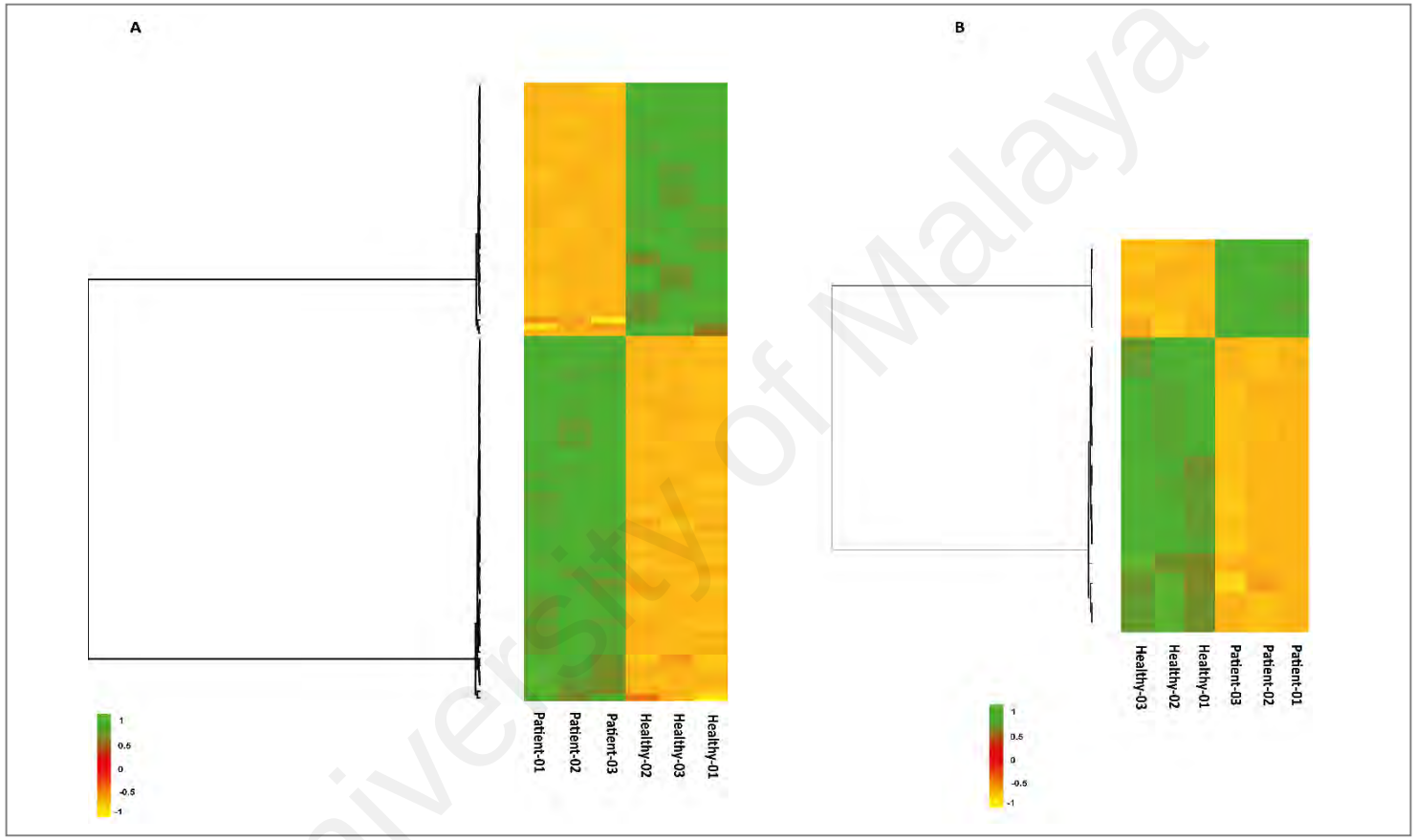

Figure 5.12: Hierarchical clustering of DE lncRNAs in primary monocytes XLA patients compared to healthy subjects. A: DE annotated lncRNAs. B: DE novel lincRNAs. Each row in the heatmap represents a gene and each column represents a separate sample. The colour scale indicates the log normalized FPKM values (green colour for upregulated genes and orange colour for downregulated genes).

the protein-coding genes whose genomic locations are within ~5 kb upstream and ~10 kb downstream of the DE annotated lncRNAs and DE novel lincRNAs were searched. The analysis revealed that out of 95 DE annotated lncRNAs, 85 lncRNAs corresponded to 144 protein-coding genes (Table 5.8). Moreover, among 20 DE novel lincRNAs, 18 novel lincRNAs were linked to 28 protein-coding genes (Table 5.9). This results, indicated that one lncRNA could target more than one protein-coding genes.

The function of DE lncRNAs was examined based on GO analysis of their co-located genes, using DAVID software. The GO analysis was restricted to the category of biological processes. In order to identify the specific GO terms, the DAVID category-GOTERM_BP_FAT level was analyzed (Figure 5.13). The results demonstrated that the DE annotated lncRNAs co-located genes were mainly involved in "reproductive process", "positive regulation of cell proliferation" and "induction of apoptosis by extracellular signals" [Figure 5.13 A] and the DE novel lincRNAs co-located genes were related to "regulation of cell differentiation" and "regulation of cell activation" [Figure 5.13 B]. Furthermore, the KEGG pathway analysis revealed that the DE annotated lncRNAs co-located genes were significantly involved in "Metabolic pathways", "Cytokine-cytokine receptor interaction", [Figure 5.14 A] and the DE novel lincRNAs co-located genes were related to "Focal adhesion" and "Regulation of actin cytoskeleton" [Figure 5.14 B].

Furthermore, our analysis examined whether any DE lncRNAs co-located genes are also differentially expressed (co-expressed genes) in XLA patients. The comparison of the DE annotated lncRNAs and DE novel lincRNAs co-located genes with DE protein-coding genes led to identification of 23 genomically co-located and co-expressed genes (Table 5.10). For instance, DE annotated lncRNAs; *HOTAIRM1*, *DANCR and GAS5* were co-located and co-expressed with DE protein-coding genes; *HOXA1*, *USP46*, *ZBTB48* respectively, which are involved in regulating the gene expression, morphogenesis and

Table 5.8: List of identified co-located genes with DE annotated lncRNAs in primary monocytes of XLA patients compared to healthy subjects.

lncRNA_Name	Chromosome	LncRNA associated genes
RMRP	9	*CCDC107*
YTHDF3-AS1	8	*GGH*, *NKAIN3*
AP001505.10	21	*SIK1*, *HSF2BP*
CTB-12A17.2	19	*IGFL2*, *IGFL3*
SNHG9	16	*HS3ST6*, *MEIOB*
RP1-92O14.3	1	N/A
LINC01506	9	*FOXD4L6*, *ANKRD20A1*
SNHG19	16	*NTHL1*, *PKD1*
RP11-297C4.6	16	*MYLPF*
RP11-669M16.1	4	*CPEB2*, *BOD1L1*
RP11-291B21.2	12	*KLRK1*
AC012442.5	2	*BCL2L11*, *ANAPC1*
RP11-410L14.2	8	*RPL30*
PCED1B-AS1	12	*SLC38A2*, *SLC38A4*
RP5-1171I10.5	17	*CA4*
LINC01023	5	*FER*, *PJA2*
SNHG8	4	*NDST3*, *TRAM1L1*
RP11-672L10.6	18	*YES1*
AC106782.20	16	*ZNF629*
RP13-516M14.1	17	N/A
RP5-899E9.1	7	*TMEM60*
SNHG6	8	*PDE7A*, *DNAJC5B*
SNHG5	6	*NT5E*, *TBX18*
AC022154.7	19	*PLA2G4C*, *CABP5*
RP11-1398P2.1	4	*FAM53A*, *NKX1-1*
HOXB-AS1	17	*RSAD1*, *MYCBPAP*
TP53TG1	7	*ABCB1*, *ABCB4*
CTA-29F11.1	22	*TRMU*, *CELSR1*
TOLLIP-AS1	11	*TOLLIP*, *MUC5B*
RP11-295G20.2	1	*EGLN1*, *SPRTN*
CTC-246B18.10	19	N/A
LINC00920	16	*CDH5*, *BEAN1*

Table 5.8: Continued.

CTB-50L17.16	19	*CHAF1A*, *UBXN6*
HOTAIRM1	7	*HOXA1*, *SKAP2*
RP11-879F14.1	18	*SERPINB2*
CAHM	6	*PARK2*, *QKI*
LINC01088	4	*ANXA3*, *FRAS1*
FLJ44511	7	*PDGFA*, *FAM20C*
EPB41L4A-AS1	5	*DCP2*, *REEP5*
LINC01003	7	*ACTR3B*
LINC01503	9	*LMX1B*, *MVB12B*
AC051649.12	11	*CTSD*, *SYT8*
RP11-162A12.2	18	*ATP9B*, *SALL3*
LINC01420	X	*UBQLN2*, *SPIN3*
AP001189.4	11	*ACER3*, *B3GNT6*
RP1-56K13.3	17	*KRT25*
AD001527.4	19	*RBM42*
RP11-85B7.5	17	*NTN1*, *STX8*
RP11-1094M14.11	17	*AATF*, *C17orf78*
AC144831.1	17	N/A
SCAMP1-AS1	5	*DMGDH*, *ARSB*
LRRC75A-AS1	17	*TRPV2*, *ZNF287*
HEIH	5	N/A
LINC00657	20	*SRC*, *NNAT*
RP5-1172N10.4	X	*DDX3X*, *USP9X*
HYI-AS1	1	N/A
RP11-1024P17.1	3	*TGFBR2*, *GADL1*
AF127936.7	21	N/A
CTD-2047H16.3	17	N/A
RP11-799D4.3	17	*LHX1*
RP1-111C20.3	6	*TMEM181*, *DYNLT1*
ADAMTSL4-AS1	1	*MCL1*
AC005062.2	7	*TMEM196*
AC004893.11	7	*PDAP1*
RP11-102F4.2	8	*PREX2*, *SULF1*
RP11-875O11.1	8	*TNFRSF10B*, *RHOBTB2*
AC092620.2	2	*HNMT*, *THSD7B*

Table 5.8: Continued.

TUG1	22	*GAL3ST1*, *PES1*, *MPST*, *H1F0*, *PRR14L*
AC108676.1	3	*XXYLT1*, *LSG1*
AC096772.6	2	*CPO*, *KLF7*
AF011889.2	X	*MAMLD1*, *MAGEA8*
AP000240.9	21	*N6AMT1*
AC007879.2	2	*ADAM23*, *DYTN*
AC159540.1	2	*CNNM4*
LINC00926	15	*TCF12*, *CGNL1*
AC073046.25	2	*DUSP11*, *STAMBP*
MEG3	14	*WARS*, *SLC25A47*
CTD-2270P14.1	16	*PDPK1*, *KCTD5*
FAM225A	9	N/A
RP4-545C24.1	7	*TPK1*, *NOBOX*
AC096579.13	2	*RPIA*
LINC01347	1	*CEP170*, *PLD5*
AC145124.2	8	*LONRF1*, *DEFB130*
RP11-1223D19.1	2	*SH3RF3*
LINC00115	1	N/A
RP11-498P14.3	9	*FBP2*, *HIATL1*
RP11-244M2.1	18	*PIK3C3*
CTD-2331C18.5	11	*SDHAF2, PPP1R32*
RP11-1228E12.1	17	*RPH3AL*, *DOC2B*
RP11-14N7.2	1	*PPIAL4B*, *NBPF8*
FAM225B	9	N/A
CASC8	8	*TRIB1*, *FAM84B*
GAS5	1	*ZBTB48*
LINC-PINT	7	*MKLN1*, *KLF14*
DANCR	4	*USP46*

#N/A, No protein-coding genes could find within ~5 kb upstream and ~10 kb downstream of this lncRNA.

Table 5.9: List of identified co-located genes with DE novel lincRNAs in primary monocytes of XLA patients compared to healthy subjects.

Transcript_ID	Genomic coordinate	LncRNA associated genes
TCONS_00347815	X:10014980-10015551	*CLCN4*
TCONS_00335887	9:127423334-127424338	*RLAGPS1*
TCONS_00008577	1:146387174-146388092	PRKAB2
TCONS_00295657	6:359346-360215	DUSP22, IRF4
TCONS_00041961	10:129035316-129036770	DOCK1, NPS
TCONS_00205060	2:157392899-157404794	N/A
TCONS_00264979	4:143336124-143336701	*INPP4B, IL15*
TCONS_00133967	16:55719783-55720456	*CES1, SLC6A2*
TCONS_00032359	10:31586861-31595354	*ZEB1,ZNF438*
TCONS_00205058	2:157345321-157368674	*ERMN*
TCONS_00327731	9:65645189-65664577	*SPATA31A7*
TCONS_00030433	1:108091096-108100990	*VAV3, NTNG1*
TCONS_00134089	17:488985-490396	*FAM101B, VPS53*
TCONS_00326114	8:114029500-114030153	*CSMD3*
TCONS_00109575	15:20369804-20376362	N/A
TCONS_00086992	13:109424117-109424380	*MYO16*
TCONS_00212764	20:45179562-45192881	*OCSTAMP, SLC13A3*
TCONS_00139067	17:42532730-42533453	*ITGA2B, GPATCH8*
TCONS_00090109	13:76592360-76593698	*LMO7, KCTD12*
TCONS_00063576	11:57086190-57087643	*APLNR, TNKS1BP1*

#N/A, No protein-coding genes could find within ~5 kb upstream and ~10 kb downstream of this lncRNA.

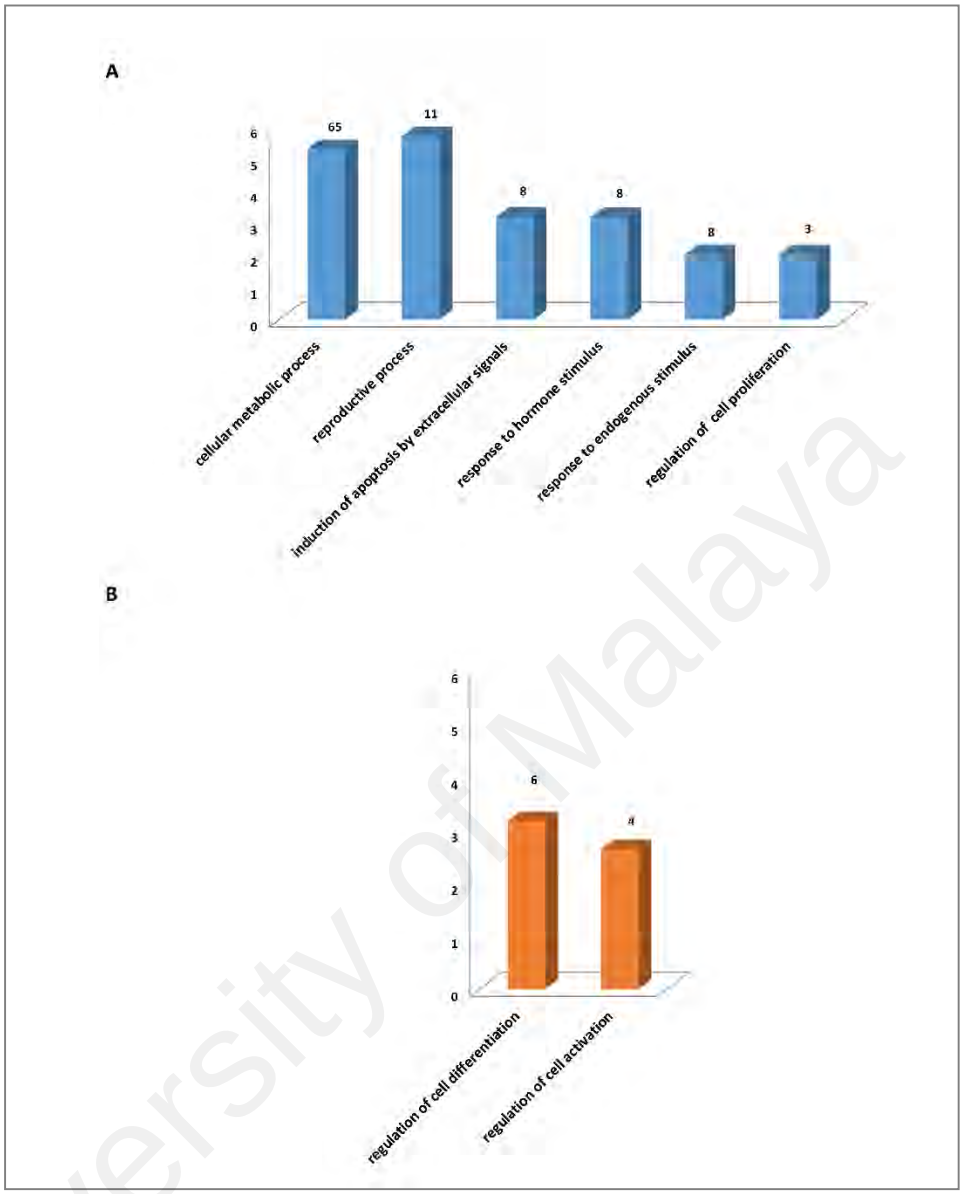

Figure 5.13: The GO analysis of identified DE lncRNAs co-located genes in primary monocytes of XLA patients compared to healthy subjects. A: The significant GO biological process terms (DAVID category-GOTERM_BP_FAT) enriched for DE annotated lncRNAs co-located genes. B: The significant GO biological process terms (DAVID category-GOTERM_BP_FAT) enriched for DE novel lincRNAs co-located genes. The number of DE lncRNAs co-located genes enriched in each GO terms is depicted above the bars in the figure.

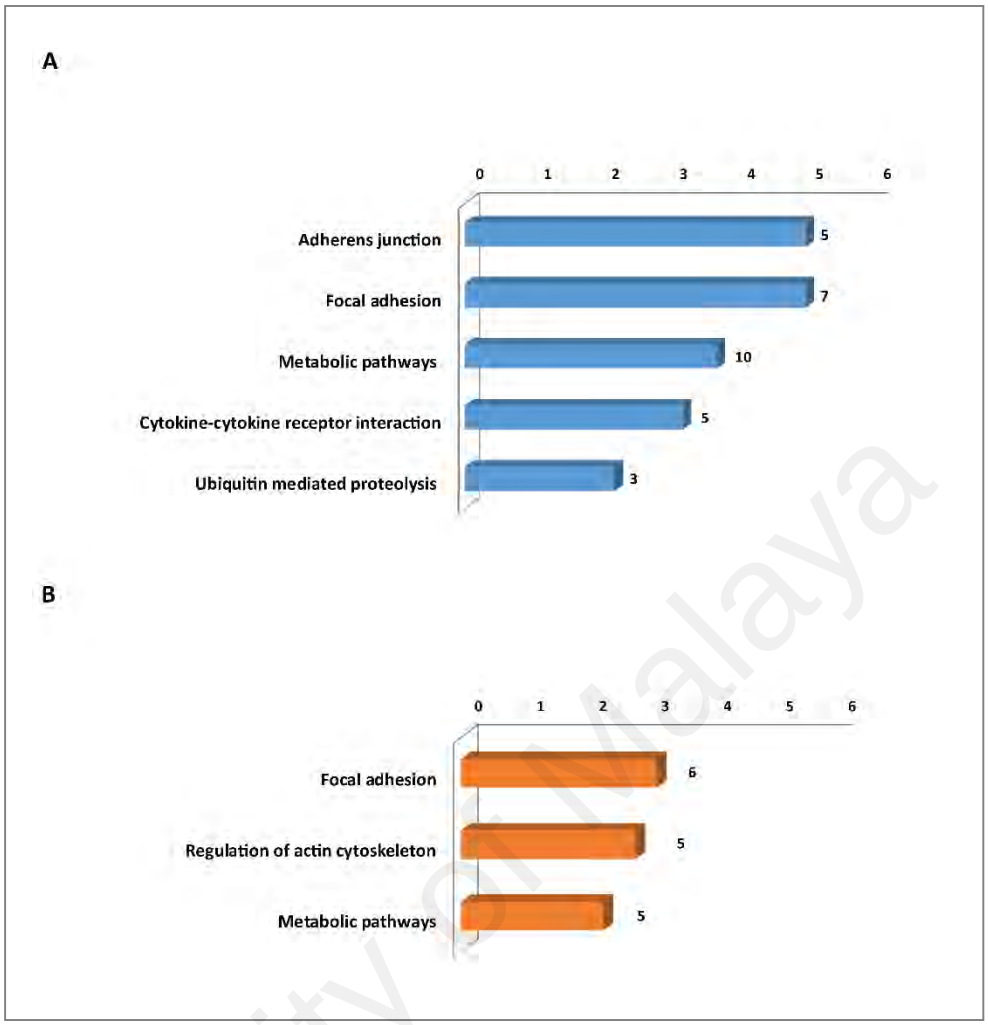

Figure 5.14: The KEGG pathway analysis for DE lncRNAs co-located genes in primary monocytes of XLA patients compared to healthy subjects. A: The significant KEGG pathway terms enriched for DE annotated lncRNAs co-located genes. B: The significant KEGG pathway terms enriched for DE novel lincRNAs co-located genes. The numbers in the brackets indicated the total numbers of genes available in the KEGG database for each pathway terms. The number of identified DE lncRNAs co-located genes enriched in each KEGG pathways is depicted above the bars in the figure.

Table 5.10: List of identified co-located and co-expressed genes with DE annotated lncRNAs and DE novel lincRNAs in primary monocytes of XLA patients compared to healthy subjects.

lncRNA_Name	Chromosome	LncRNA expression	LncRNA associated genes	Gene expression
DE annotated lncRNAs				
RP11-669M16.1	4	Up	*CPEB2*	Down
RP5-1172N10.4	X	Down	*DDX3X*	Down
CAHM	6	Up	*QKI*	Down
AP001505.10	21	Up	*SIK1*	Down
LINC00926	15	Down	*TCF12*	Down
CASC8	8	Down	*TRIB1*	Down
AC108676.1	3	Down	*XXYLT1*	Down
RMRP	9	Up	*CCDC107*	Up
RP11-410L14.2	8	Up	*RPL30*	Up
AD001527.4	19	Up	*RBM42*	Up
TUG1	22	Down	*MPST*	Up
TUG1	22	Down	*H1F0*	Up
TUG1	22	Down	*PRR14L*	Down
PcED1B-AS1	12	Up	*SLC38A2*	Down
HOTAIRM1	7	UP	*HOXA1*	Up
DANCR	4	Up	*USP46*	Down
GAS5	1	Up	*ZBTB48*	Up
LINC-PINT	7	Up	*MKLN1*	Down
DE novel lincRNAs				
TCONS_00139067	17	Down	*GPATCH8*	Down
TCONS_00030433	1	Down	*VAV3*	Down
TCONS_00008577	1	Up	*PRKAB2*	Down
TCONS_00295657	6	Down	*DUSP22*	Down
TCONS_00041961	10	UP	*DOCK1*	Down

differentiation, ubiquitin protease activity, and MHC II promoter compotes. The results also indicated that DE novel lincRNAs *TCONS_00030433*, *TCONS_00041961* and *TCONS_00295657* were co-located and co-expressed with DE protein-coding genes *VAV3* (Vav Guanine Nucleotide Exchange Factor 3); involve in phagocytosis, *DOCK1* (Dedicator of cytokinesis); involve in kinase activity, and *DUSP22* (Dual Specificity Phosphatase 22); involve in ubiquitin protease activity, respectively.

5.3.2.2 Gene interaction network of DE lncRNAs with co-located and co-expressed DE protein-coding genes

To unravel the interaction between DE annotated lncRNAs and DE novel lincRNAs with their co-located and co-expressed DE protein-coding genes, putative interactive networks were constructed using Cytoscape (Figure 5.15). The network contains 80 interactions between 21 DE annotated lncRNAs and DE novel lincRNAs with their 23 co-located and co-expressed DE protein-coding genes.

5.4 qRT-PCR validation

To further confirm the RNA-Seq analysis results presented in this chapter, the expression levels of selected DE protein-coding genes and DE annotated lncRNAs and DE novel lincRNAs were measured by qRT-PCR analysis (figure 5.16). The candidate genes included 10 DE protein-coding genes; *FCGR2A*, *CXCR2*, *TLR1*, *TLR5*, *ATP5D*, *NDUFA1*, *UQCRB*, *SOD, MTOR*, *BAX* which enriched in several significant upregulated and downregulated KEGG pathways, as well as 7 DE annotated lncRNAs; *HOTAIRM1*, *DANCR*, *GAS5*, *LINC-PINT*, *RMRP*, *HEIH*, *TUG1* and 3 DE novel lincRNAs; *TCONS_00041961*, *TCONS_00295657*, *TCONS_00298577*, which co-located and co-expressed with DE protein-coding genes in XLA patients compared to the healthy subjects. The $\log_2$ fold-change results obtained with qRT-PCR were fully consistent with

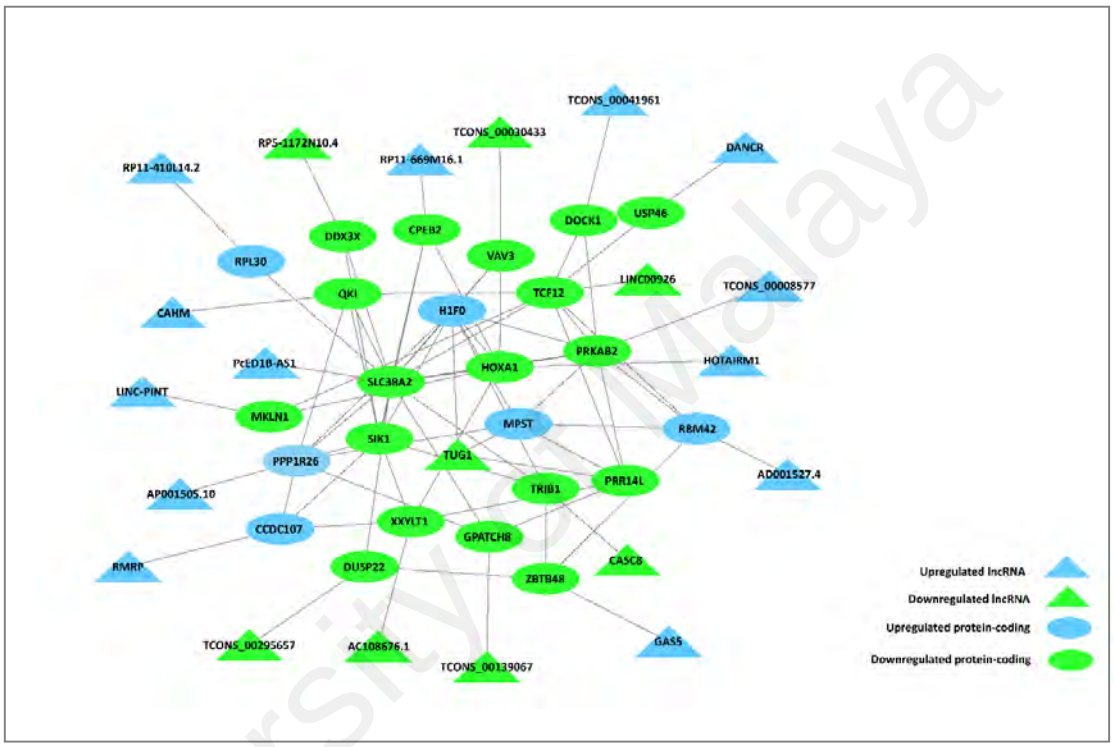

Figure 5.15: The interaction network of DE lncRNAs with their co-located and co-expressed DE protein-coding genes in primary monocytes of XLA patients compared to healthy subjects.

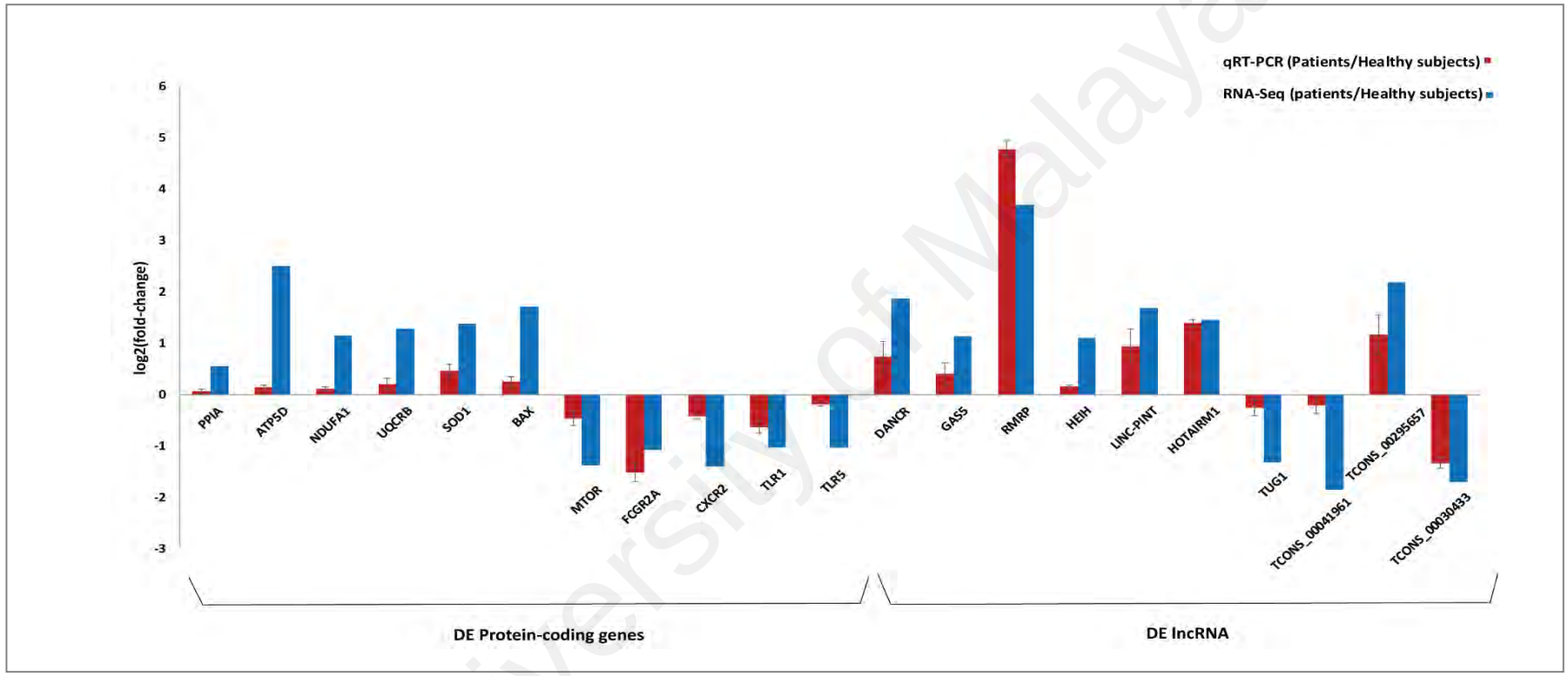

Figure 5.16: The qRT-PCR validation of DE protein-coding genes and DE lncRNAs in primary monocytes of XLA patients compared to the healthy subjects. The comparison of $\log_2$ fold-change of DE protein-coding genes and DE lncRNAs were determined by RNA-Seq analysis (blue) and qRT-PCR validation (red). *PPIA* was used as endogenous control for normalizing the expression levels. x-axis shows genes; y-axis shows the $\log_2$ ratio of expression in XLA patients compared to healthy subjects.

the $\log_2$ fold-change results obtained from RNA-Seq analysis, demonstrate the reliability of our RNA-Seq data analysis (Figure 5.16).

5.5 Discussion

Deep RNA-Seq approch on primary monocytes from 3 XLA patients was performed. The sequencing generated approximately 477 million reads of 100 bp read length which lead to the profile of 17,510 genes (including 11,788 protein-coding, genes 3,681 non-coding genes, 2,041 pseudogenes) and 62,367 transcripts (including 58,136 annotated and 4,231 novel transcripts) in primary monocytes of XLA patients. In addition, the lncRNAs expression patterns in primary monocytes of XLA patients was analyzed and a total of 3,363 annotated lncRNAs and 430 novel lincRNAs were identified. A comparative analysis on gene expressions profile of primary monocytes between XLA patients and healthy subjects was performed to examine possible differences in the gene expression patterns of primary monocytes in XLA patients. The comparative analysis showed the total of 1,827 DE protein-coding genes, in XLA patients compared to healthy subjects. Out of 1,827 DE protein-coding genes, 859 genes were upregulated and 968 genes were downregulated in XLA patients compared to the healthy subjects. Based on the GO and KEGG pathways analysis, the detailed information on the biological functions and potential mechanisms of detected DE protein-coding genes were obtained. The GO enrichment analysis showed that downregulated genes were mainly involved in regulation of immune response. Pathway analysis also revealed that downregulated genes mainly enriched in several pathways belonged to innate immune system such as; “Fc gamma R-mediated phagocytosis”, “Chemokine signaling pathways”, “Toll like receptors signaling pathway” and “MTOR signaling pathway”, which reflected the deficiencies of innate immune function in primary monocytes of XLA patients.

Phagocytosis plays a critical role in host-defense mechanisms through the uptake and destruction of infectious pathogens (Rosales, 2005). The expression of *FCGR2A* (also

known as *FcγRIIA*, CD32), the low affinity receptor for monomeric IgG, was significantly decreased in primary monocytes of the XLA patients, which is consistent with a previous study reporting about decreased expression of *FCGR2A* on monocytes from XLA patients upon *BTK* deficiency (Amoras et al., 2003). In addition, this study extended the previous finding regarding the impaired phagocytic functions of monocytes from XLA patients. The analysis results demonstrated the downregulation of 22 core enrichment genes involved in "Fc gamma R-mediated phagocytosis" pathway in XLA patients compared to healthy subjects (see Table 5.5, page 102). The identified downregulated genes encoding for kinases in early signaling events such as *SRC* (SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase) and *SYK* (Spleen Tyrosine Kinase) Kinases, as well as the genes encoding for proteins involved in cytoskeleton rearrangement. These results imply that, in addition to *FCGR2A*, *BTK* deficiency could affect an entire "Fc gamma R-mediated phagocytosis" pathway in monocytes of XLA patients.

The decreased expression of 32 core genes involved in "chemokines signaling pathway" including 4 chemokine receptors (*CXCL16*, *CXCR1*, *CXCR2*, *CXCR4*) was also observed in primary monocytes of the XLA patients (see Table 5.5, page 102). Chemokines signaling organizes cellular reactions to insult, injury or inflammation and has an essential role in the regulation of leukocyte extravasation and trafficking via the endothelial cells' luminal surface to the tissue inflammation sites (Phillips, Lutz, & Premack, 2005). *CXCL16* (C-X-C Motif Chemokine Ligand 16) is a membrane-bound chemokine, which facilitates the bacterial phagocytosis, adhesion and recognition (Tohyama et al., 2007). *CXCR1* (C-X-C Motif Chemokine Receptor 1) and *CXCR2* (C-X-C Motif Chemokine Receptor 2) are absorbed into inflammation sites in response to chemokines like *CXCL8* (C-X-C Motif Chemokine Ligand 8)/*IL-8* (Interleukin 8), *CCL2* (Chemokine (C-C motif) Ligand 20 and *CCL3* (Chemokine (C-C motif) Ligand 3) (Murdoch & Finn, 2000). *CXCR4* (C-X-C Motif Chemokine Receptor 4) has been shown

to help *SDF-1* (sodium:dicarboxylate-1) to recruit monocytes to the inflamed site (Caulfield, Fernandez, Snetkov, Lee, & Hawrylowicz, 2002). A direct role for *BTK* in signaling by *CXCR4* and in chemokine-controlled adhesion and migration in B cells of XLA patients has been shown previously (de Gorter et al., 2007). Similar observations regarding the regulatory role of *BTK* on *CXCR4* was found in monocytes of XLA patients in which the *BTK* deficiency lead to downregulation of *CXCR4*. In addition, the downregulation of *CXCL16*, *CXCR1,* and *CXCR2* observed in primary monocytes of the XLA patients indicated possible regulatory role of *BTK* on these chemokines receptors in primary monocytes.

The overall downregulation of the "Toll like receptors signaling pathway" was also observed in primary monocytes of the XLA patients compared to the healthy subjects, which was exemplified by the decreased expressions of *TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, and *TLR7* along with other 11 genes involved in "Toll like receptors signaling pathway" (see Table 5.5, page 102). *TLRs* are a main group of pathogen recognition receptors that are important as the first fighters against foreign microorganisms as well as modulators of the adaptive immune response (Kuby et al., 2007). The *BTK* has been shown to be involved in *TLR* signaling, interacts with *TLR2*, *TLR4*, *TLR6, TLR7*, *TLR8,* and *TLR9* and facilitates their transduction of downstream signals and phosphorylation (Doyle et al., 2007; Horwood et al., 2006b; Jefferies et al., 2003). In addition to those finding, our study demonstrates for the first time significantly decreased in expression of *TLR1* and *TLR5* in primary monocytes of XLA patients compared to healthy subjects. *TLR5* and *TLR1* are expressed both intracellularly (for phagolysosomes) and on the cell surface (O'Mahony, Pham, Iyer, Hawn, & Liles, 2008). *TLR1* identifies triacylated and diacylated lipopeptides in contract with *TLR2* (Takeuchi et al., 2002). *TLR5* has important role in innate immune system to recognize the bacteria and distinguish the flagellin (Hawn et al., 2003; Hayashi et al., 2001). The downregulation

of *TLR1* and *TLR5* in monocytes of XLA patients observed here suggest that in addition to other *TLRs* which has been reported previously, the *BTK* may also associates with *TLR1* and *TLR5* expression in primary monocytes of XLA patients, in which mutants of *BTK* may inhibit their signaling. Similar observation of *TLR1* signaling deficiency due to the *BTK* mutant was reported in mice (Alugupalli, Akira, Lien, & Leong, 2007).

Our analysis also revealed the downregulation of *MTOR* (Mechanistic Target Of Rapamycin) gene along with other 10 core genes involved in "MTOR signaling pathway" in primary monocytes of XLA patients compared to healthy subjects (see Table 5.5, page 102). This suggest the role of *BTK* in *MTOR* signaling in monocytes of XLA patients. Recently, Ezell et al., (2016) reported similar observation regarding the possible regulatory role of *BTK* on *MTOR* signaling in activated Diffuse large B cell lymphoma (DLBCL). *MTOR* is known as a central node in cell differentiation and growth, cellular metabolism and cancer metabolism. It can sense the growth factors, nutrients, insulin, environmental cues, and energy, then sends signals to downstream targets to effectuate the metabolic and cellular reaction (Soliman, 2013). Also, *MTOR* was recently found to be associated with the regulation of both the innate (Weichhart, Hengstschläger, & Linke, 2015) and adaptive immune response (Thomson, Turnquist, & Raimondi, 2009). Powell et al. (2012) reported that *MTOR* senses the immune microenvironment and guides the outcome of antigen recognition in an identical way to nutrient sensing mediated by *MTOR*, and in response to environmental cues and energy (Powell, Pollizzi, Heikamp, & Horton, 2012). The low expressions of *MTOR* is seen in cells that are more dependent on mitochondrial oxidative phosphorylation for energy supply (Keating & McGargill, 2016). The high demand on energy production as seen by overexpression of mitochondrial components has been reported in several disease states such as cancers (Griguer, Oliva, & Gillespie, 2005; Koppenol, Bounds, & Dang, 2011; Scott et al., 2011), human immunodeficiency virus (HIV) (Zhou et al., 2010), and Alzheimer's disease (AD)

(Manczak, Park, Jung, & Reddy, 2004). Our analysis also identified the low expression of *MTOR* and overexpression of multiple components of mitochondrial complexes I (*NDUF*), III (*UQCR*), IV (*COX*) and V (*ATP*) (see Table 5.4, page 101) in XLA patients compared to the healthy subjects. This would suggest the high energy demand production in primary monocytes of XLA patient compared to healthy subjects. Furthermore, the upregulation of several genes involved in production of reactive oxygen species (ROS), response to oxidative stress and apoptotic process were observed in XLA patients compared to healthy subjects. It is well known that during the oxidative phosphorylation, mitochondria consume most of the cellular oxygen and produce the majority of ROS. The high concentration of ROS in the cell would lead to state termed oxidative stress, in which the excess ROS induces oxidative damage on cellular components and activate apoptosis pathways and cell death (Finkel, 2012). This would explain the finding of upregulation of several genes involved in oxidative phosphorylation, ROS production, response to oxidative stress and apoptosis in XLA patients which caused monocytes to be vulnerable and more susceptible to apoptosis.

Evidence suggests that in addition to protein-coding genes, lncRNAs can act as key regulators of various biologic processes (Clark & Mattick, 2011; Kim & Sung, 2012). The lncRNAs functions via DNA–DNA, DNA–RNA or other kind of interactions (Satpathy & Chang, 2015). The lncRNAs dysregulated expression has been has reported in many human diseases (Hrdlickova et al., 2014; Wapinski & Chang, 2011). However, little is known about lncRNAs role in XLA patients. Through the comparative analysis of lncRNAs expressions in primary monocytes between XLA patients and healthy subjects, 95 DE annotated lncRNAs including 56 upregulated and 39 downregulated and 20 DE novel lincRNAs including 5 upregulated and 15 downregulated were obtained in XLA patients compared to healthy subjects. Several DE lncRNAs were identified in XLA patients, which were known to contribute to regulation of gene expressions and cell cycle.

Overexpression of these lncRNAs have been reported to suppress the cell growth, differentiation, proliferation, and promote apoptosis in variety of diseases. Such lncRNAs include: *HOTAIRM1* (Wan et al., 2016; X. Zhang et al., 2009), *DANCR* (Tong et al., 2015), *GAS5* (Tu, Li, Mei, & Li, 2014), *LINC-PINT* (Marín-Béjar et al., 2013), *HEIH* (Yang et al., 2011), *RMRP* (Elling, Chan, & Fitzgerald, 2016). These lncRNAs were also found to be overexpressed in primary monocytes of XLA patients compared to the healthy subjects. Our analysis also revealed significant decreased in expression level of lncRNA *TUG1* in XLA patients compared to the healthy subjects. The downregulation of *TUG1* has been reported to inhibits osteosarcoma cell proliferation and promotes apoptosis (Young, Matsuda, & Cepko, 2005). Similar observation regarding the dysregulated expression of these lncRNAs in primary monocytes of XLA patients suggest their possible role in regulating monocytes cell cycle and apoptosis in XLA patients. The analysis of DE novel lincRNAs also revealed that some of DE novel lincRNAs were co-located and co-expressed with DE protein-coding genes related to immune system. In particular, the novel DE lncRNA *TCONS_00030433* which was significantly downregulated in XLA patients, interacted with *VAV3*, which its expression was also detected to be downregulated in XLA patients. *VAV3* is known to be involved in "Fc gamma R-mediated phagocytosis" pathway (Hall et al., 2006). This results would suggest the possible role of *TCONS_00030433* in regulation of "Fc gamma R-mediated phagocytosis" pathway in primary monocytes of XLA patients. The expression of selected 10 DE protein-coding genes as well as 7 DE annotated and 2 DE novel lincRNAs in XLA patient compared to healthy subjects were further validated using qRT-PCR. The qRT-PCR results support the RNA-Seq data analysis and findings.

In summary, this study profiled the gene expressions of primary monocytes from XLA patients using deep RNA-Seq analysis. The comparative analysis was then conducted on gene expression profiles of primary monocytes between XLA patients and

healthy subjects. The results indicated the overall dysregulation of monocytes immune functions and the increased of susceptibility to apoptosis in monocytes of XLA patients. This suggest that *BTK* mutations is not only affecting the B cell development and differentiation, it would be also contributing in dysregulation of innate immune system in XLA patients. This study also revealed the differentially expression patterns of lncRNAs in primary monocytes of XLA patients that would suggest the potential role of lncRNAs in regulation of immune functions in primary monocytes of XLA patients.

CHAPTER 6: GENE CATALOGUE AND lncRNAs LANDSCAPE IN HUMAN PRIMARY MONOCYTES

6.1 Introduction

Transcriptome analyses based on high-throughput RNA-Seq provides powerful and quantitative characterization of cell types, and in-depth understanding of biological systems in health and disease (Sirri et al., 2008). Thus, establishment of a comprehensive gene reference catalogue of immune cells based on their transcriptome expression profiles would be useful for immune-related research, particularly for understanding disease states, pathogenesis and developing therapeutic biomarkers.

lncRNAs have been shown to possess a wide range of functions in both cellular and developmental processes in several immune-related diseases (Geng & Tan, 2016; Song et al., 2015; Lee et al., 2005; Wallace et al., 2010). Although some of the lncRNAs have been implicated in the regulation of the immune response (Hrdlickova et al., 2014), the exact function of the large majority of lncRNAs in immune system still remains unknown. In this study, the in-house deep RNA-Seq datasets of primary monocytes generated from healthy subjects were integrated with other publicly available RNA-Seq datasets for human monocytes to provide the comprehensive gene reference catalogue of human primary monocytes. Also, a landscape of lncRNAs expressions in human monocytes was generated which could facilitate future experimental studies to characterize the functions of these molecules in the innate immune system. The schematic representations of the bioinformatics analysis workflow are described in Figures 6.1 and 6.2.

6.2 Gene Catalogue of human primary monocytes

To generate the comprehensive landscape of transcriptome profile human primary monocytes under healthy states, 10 other publicly available RNA-Seq datasets from

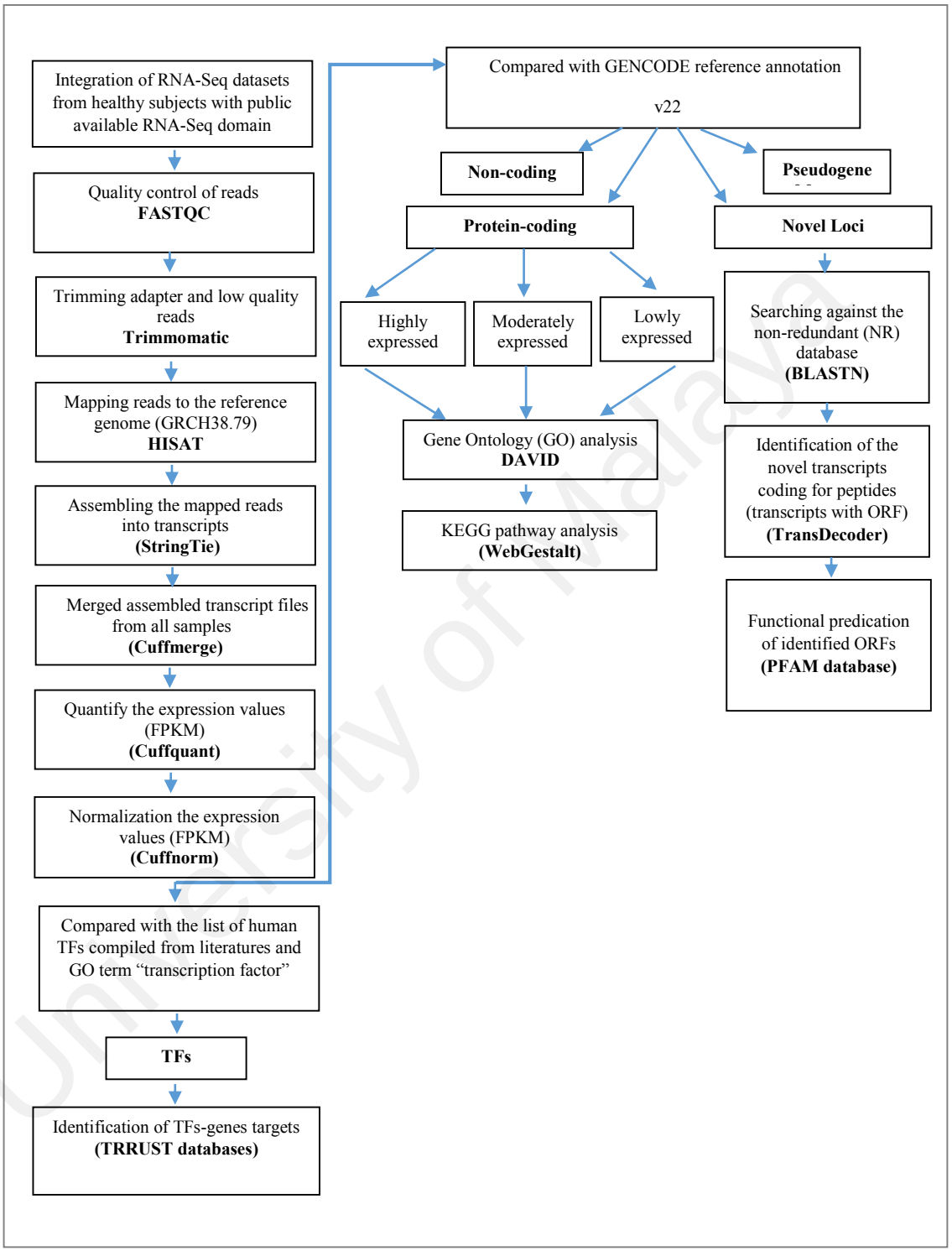

Figure 6.1: Schematic representation of workflow to generate the gene catalogue of primary monocytes from healthy subjects.

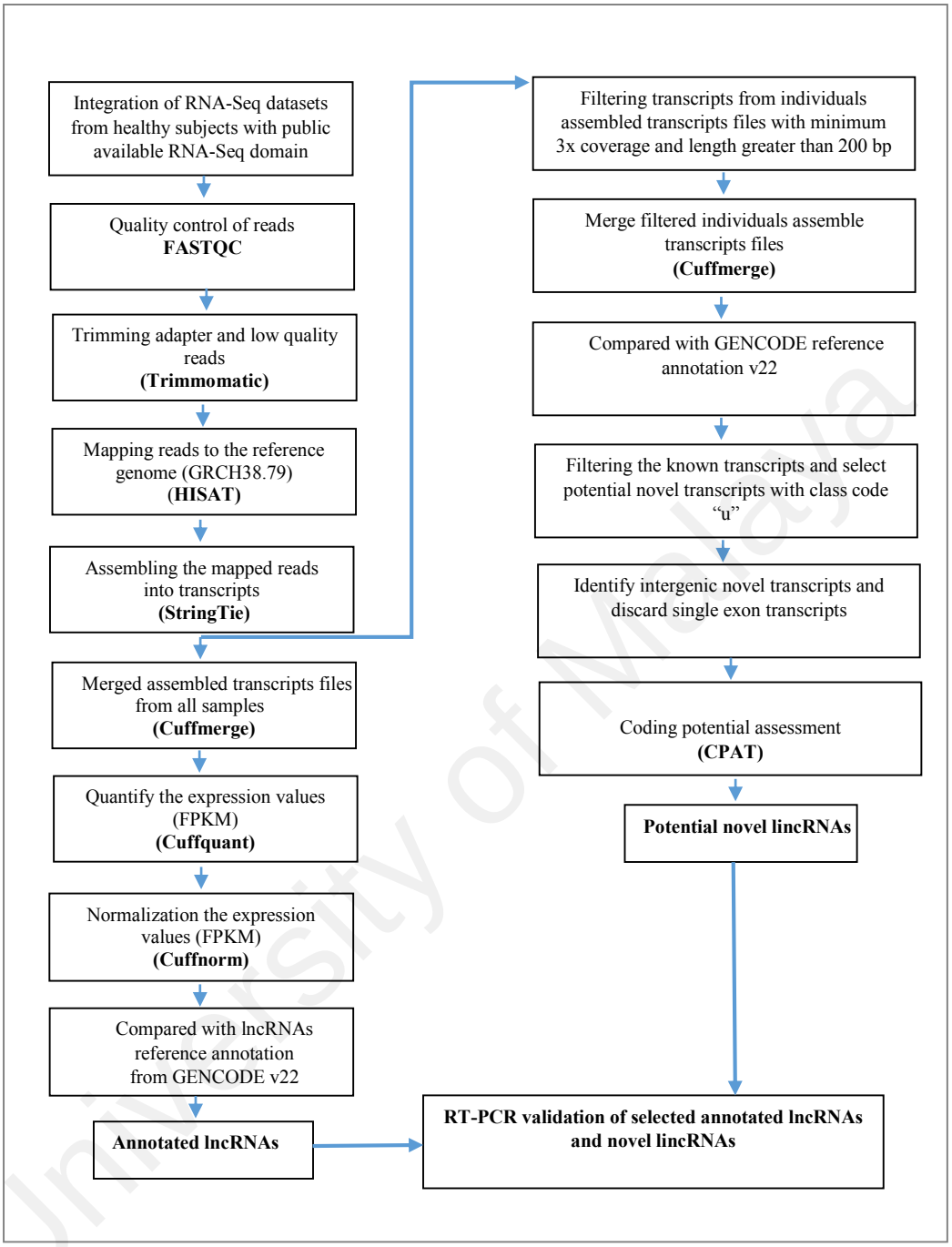

Figure 6.2: Schematic representation of workflow to profile the lncRNAs expression landscape of primary monocytes from healthy subjects.

human monocytes samples were added to in-house RNA-Seq datasets generated from healthy subjects (see Table 3.5, page 58) which provided approximately 1.9 billion reads. All the reads were mapped to human reference genome and assembled into transcriptome using the same pipeline (see Figure 6.1, page 130) to reduce any bias. An average 90% of the reads aligned to the human reference genome (Ensembl GRCH38.79) (Table 6.1). The abundance of assembled transcripts was estimated using FPKM (fragments per kilobase of exon per million fragments mapped) value. By applying the FPKM >0.1 threshold, a total of 20,371 genes and 82,996 transcripts were identified in human monocytes datasets. The summary of identified genes and transcripts with regard to their biotype is presented in Table 6.2.

6.2.1 Protein-coding genes

Out of 19,814 protein-coding genes reported in GENCODE database (version 22), the expression of total 11,994 protein-coding genes was detected in human monocytes. The identified protein-coding genes were divided into 3 groups based on their FPKM values: high expressed (top 25th percentile; FPKM >26.9) (3,009 genes), moderately expressed (middle 50th percentile; 1.6< FPKM ≤26.9) (6,008 genes) and lowly expressed (bottom 25th percentile; FPKM ≤1.6) (2,909 genes). The GO enrichment analysis based on the biological process categories (DAVID category-GOTERM_BP_1) revealed that the highly expressed genes were mainly enriched for “cellular process”, “immune system process” and “death” [Figure 6.3 A]. The moderately expressed genes were found to be mainly involved in “metabolic process”, “cellular components organization” and “cellular components biogenesis” [Figure 6.3 B]. Following that, the low expressed genes were found to be mainly enriched in “biological regulation”, “developmental process”, and “biological adhesion” [Figure 6.3 C]. The KEGG pathway enrichment analysis also showed that highly expressed genes mainly enriched in several significant pathways

Table 6.1: Summary of alignment results of primary monocyte's RNA-Seq datasets from public datasets.

Study and sample ID	Percentage of mapped reads
Iiott et al. 2014 (ERS422905)	90%
Iiott et al. 2014 (ERS422908)	89%
Iiott et al. 2014 (ERS422906)	88%
Iiott et al. 2014 (ERS422910)	93%
Derrien et al. 2012 (ENCFF000HUY, ENCFF000HVE)	93%
Derrien et al. 2012 (ENCFF000HUX, ENCFF000HVD)	88%
Derrien et al. 2012 (ENCFF000HUW, ENCFF000HVC)	89%
Derrien et al. 2012 (ENCFF000HUU, ENCFF000HVA)	90%
Derrien et al. 2012 (ENCFF000HUZ, ENCFF000HVF)	89%
Derrien et al. 2012 (ENCFF000HUV, ENCFF000HVB)	87%

Table 6.2: Summary of identified genes and transcripts in primary monocytes.

Gene Biotype	Protein-coding genes	Non-coding genes	Pseudogene	Novel
Number of genes	11,994	5,724	2,820	–
Number of transcripts	63,515	6,395	3,233	7,034
Distribution across chromosomes	All chromosomes	All chromosomes	All chromosomes	All chromosomes

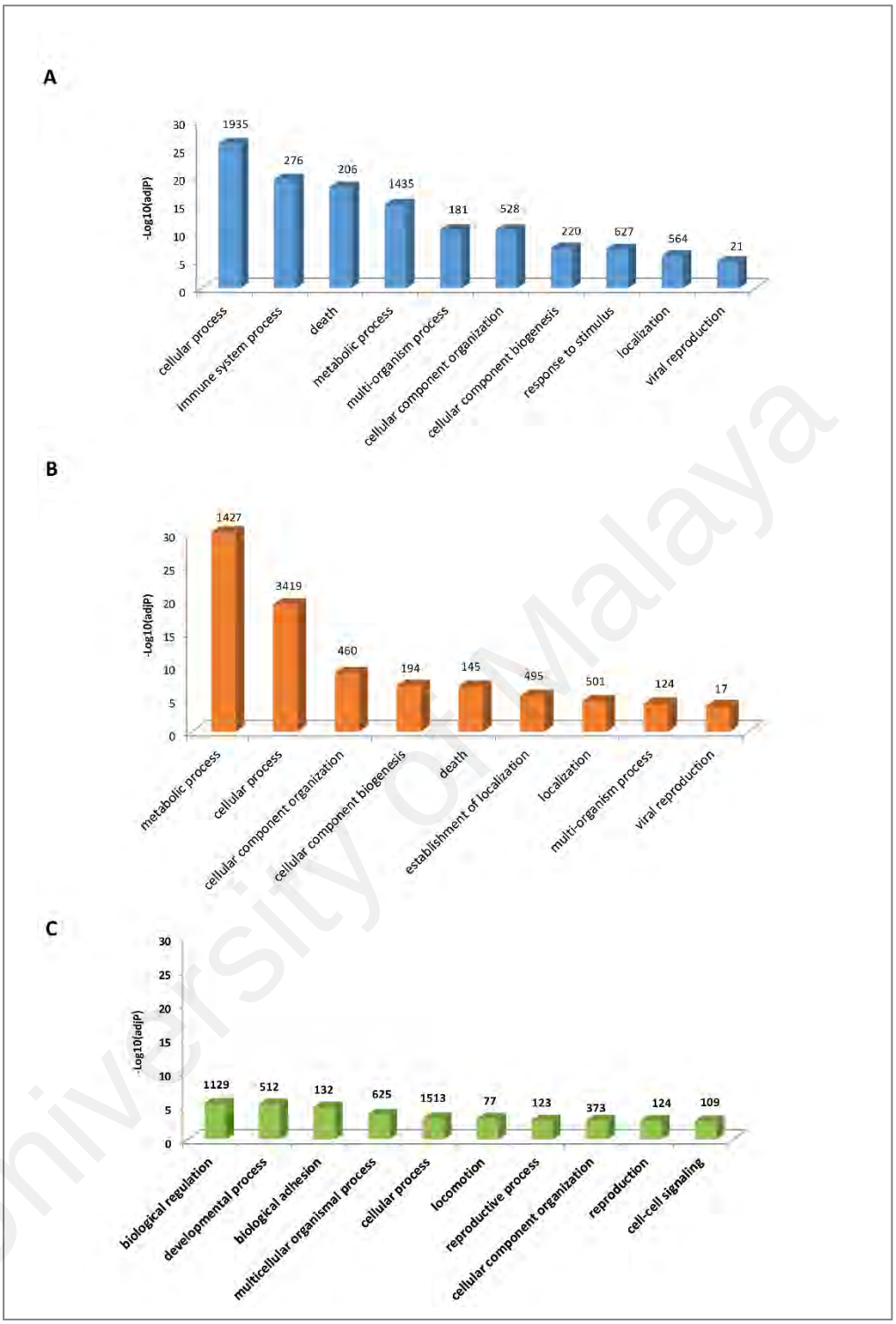

Figure 6.3: The GO analysis of protein-coding genes in primary monocytes. The significant GO biological process terms (DAVID category-GOTERM_BP_1) for A: highly expressed, B: moderately expressed, and C: lowly expressed protein-coding genes. The number of protein-coding genes enriched in each GO terms is depicted above the bars in the figure.

(ajdP <0.01) which belonged to immune system such as “Fc gamma R-mediated phagocytosis”, “chemokine signaling pathways” and “toll like receptors signaling pathway” and “apoptosis” [Figure 6.4 A]. While the lowly and moderately expressed genes were significantly enriched in “RNA degradation and glycine” and “serine and threonine metabolism” respectively [Figure 6.4 B, C].

6.2.2 Non-coding genes

A comparison of the assembled transcripts with the GENCODE reference annotation (version 22) showed evidence of expression for 5,558 non-coding genes across all the datasets studied.

6.2.3 Pseudogenes

Pseudogenes are copies of coding genes that come up from genomic duplication or retrotransposition of mRNA sequences into the genome followed by accumulation of harmful mutations because of loss of selection pressure, degenerating eventually into so-called hereditary fossils (Porter et al., 2014). The expression of 2,820 pseudogenes was detected across all the datasets studied. Although pseudogenes were known as “genomic fossils” for several years, some studies have reported that pseudogenes could play critical roles in regulation of their parent genes, and many pseudogenes were transcribed into RNA (Porter et al., 2014; Tian et al., 2007). The high expression of two functional pseudogenes which have been reported to act as transcription factors, including *MORF4* (Mortality Factor 4) (Yochum & Ayer, 2002) and *MEIS3P1* (Meis Homeobox 3 Pseudogene 1) (Tian et al., 2007) were identified in human primary monocytes.

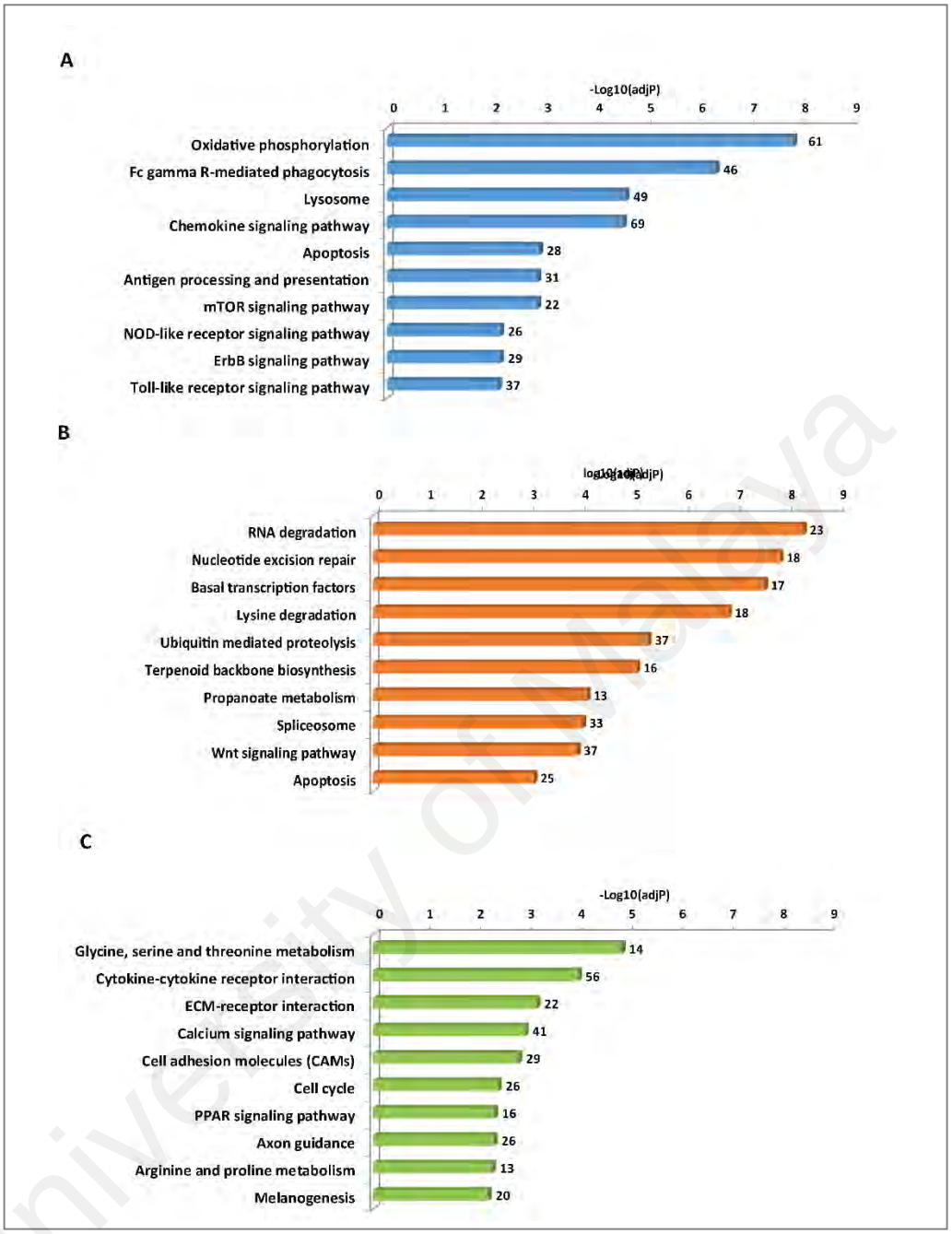

Figure 6.4: The KEGG pathways analysis of protein-coding genes in primary monocytes. The significant KEGG pathway terms enriched for; A: highly expressed protein-coding genes, B: moderately expressed protein-coding genes, and C: low expressed protein-coding genes. The numbers in the brackets indicated the total numbers of genes available in the KEGG database for each pathway terms. The number of identified protein-coding genes enriched in each KEGG pathways is depicted above the bars in the figure.

6.2.4 Novel transcripts

A salient feature of RNA-Seq is its ability to detect novel transcripts (Wang et al, 2009). Our RNA-Seq analysis detected the expression of potential 7,034 novel transcripts In monocytes which have not been previously annotated in database. Out of these, 1,362 transcripts could potentially code for peptides. A comparison of 1,362 novel transcripts against PFAM-A domain database resulted in 210 novel transcripts matching at least one protein domain model in which some of them associated with immune-related functions (APPENDIX B). However further functional studies are needed to identify the exact function and mechanism of these novel transcripts in human monocytes.

6.2.5 Transcription factors (TFs)

The expression of 1,155 TFs were identified in human primary monocytes (APPENDIX C). As the TFs are the major regulators of gene transcription, identification of the genes that are targeted by a specific TF is important for understanding cellular developmental processes, response to stimulates and disease etiology (Taverner, Smith, & Wardle, 2004). Using the TRURSUT database, 1,339 targeted genes were detected for 445 TFs in monocytes. Several TFs found to regulate a smaller number of genes such as *ZSCAN21* (Zinc Finger And SCAN Domain Containing 21) (1 target) and *CITED2* (Carboxy-Terminal Domain 2) (2 targets), while others regulate a larger number of genes such as *SP1* (Specificity Protein 1) (305 targets), *NF-κB1* (Nuclear Factor Kappa B Subunit 1) (226 targets) , *RELA* (RELA Proto-Oncogene, NF-KB Subunit) (223 targets), *TP53* (Tumor Protein P53) (132 targets), *E2F1* (E2F Transcription Factor 1) (108 targets), *JUN* (Jun Proto-Oncogene) (93 targets), and *STAT1* (Signal Transducer And Activator Of Transcription 1) (41 targets). The GO enrichment analysis based on the biological process categories (DAVID category-GOTERM_BP_FAT) on TFs-target genes showed that they are significantly (adjp <0.01) involved in 19 GO terms mainly in “response to

stimulus", "biological regulation", "immune system process" and "death" (Figure 6.5). The interaction network between top 20 highly expressed TFs with their targeted genes is presented in Figure 6.6. This network contains 226 interactions in between 20 TFs and 146 targeted genes. The TFs in the network which regulate the most immune system and death related genes were *STAT1*, *SATA6* (Signal Transducer and Activator of Transcription 6, Interleukin-4 Induced), *FOS* (FBJ Murine Osteosarcoma Viral Oncogene Homolog), *JUNB* (Jun B Proto-Oncogene), *FLI1*, *ZEP36* (Growth Factor-Inducible Nuclear Protein NUP475) and *DEK* (DEK Proto-Oncogene).

6.3 lncRNAs landscape in human primary monocytes

6.3.1 Annotated lncRNAs expressed in primary monocytes

Using the integrated RNA-Seq datasets (see Table 3.5, page 58), a landscape of lncRNAs expressions in human primary monocytes was generated. GENCODE version 22 contains a total of 15,900 genes coding for lncRNAs in humans. Since, the lncRNAs are known to be tissue/cell type specific and also developmentally regulated, our first goal was to look at the expression levels of the annotated lncRNAs in the monocytes. A comparison of the assembled transcripts with the annotated lncRNAs references from GENCODE (version 22) showed evidence of expression for 6,382 genes across all the datasets studied. Our results revealed that most of the detected lncRNAs expressed at low levels. however, the expression levels of some lncRNAs found to be high (with an average FPKM >20) across all 15 samples. Such lncRNAs included GAS5 (Growth arrest-specific 5), *CEBPB-AS1* (CCAAT/enhancer binding protein (C/EBP), beta anti-sense), *LINC-PINT* (Long Intergenic Non-Protein Coding RNA, P53 Induced Transcript), *ITGB2-AS1* (integrin, beta 2 complement component 3 receptor 3 and 4 subunit anti-sense) and *FGD5-AS1* (FYVE, RhoGEF and PH domain containing 5 antisense) among others.

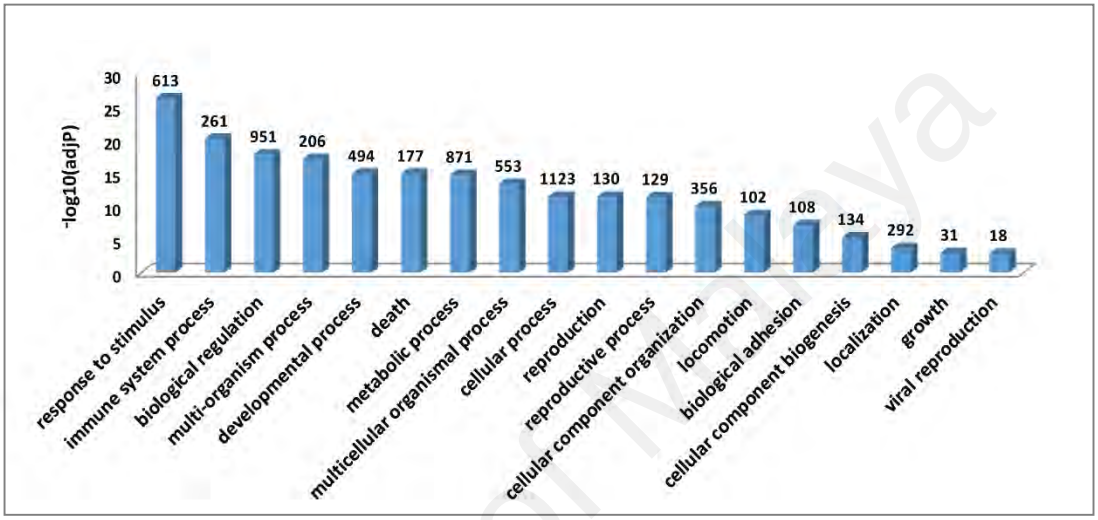

Figure 6.5: The GO analysis of TFs-target genes expressed in primary monocytes. The number of protein-coding genes enriched in each GO term is depicted above the bars in the figure.

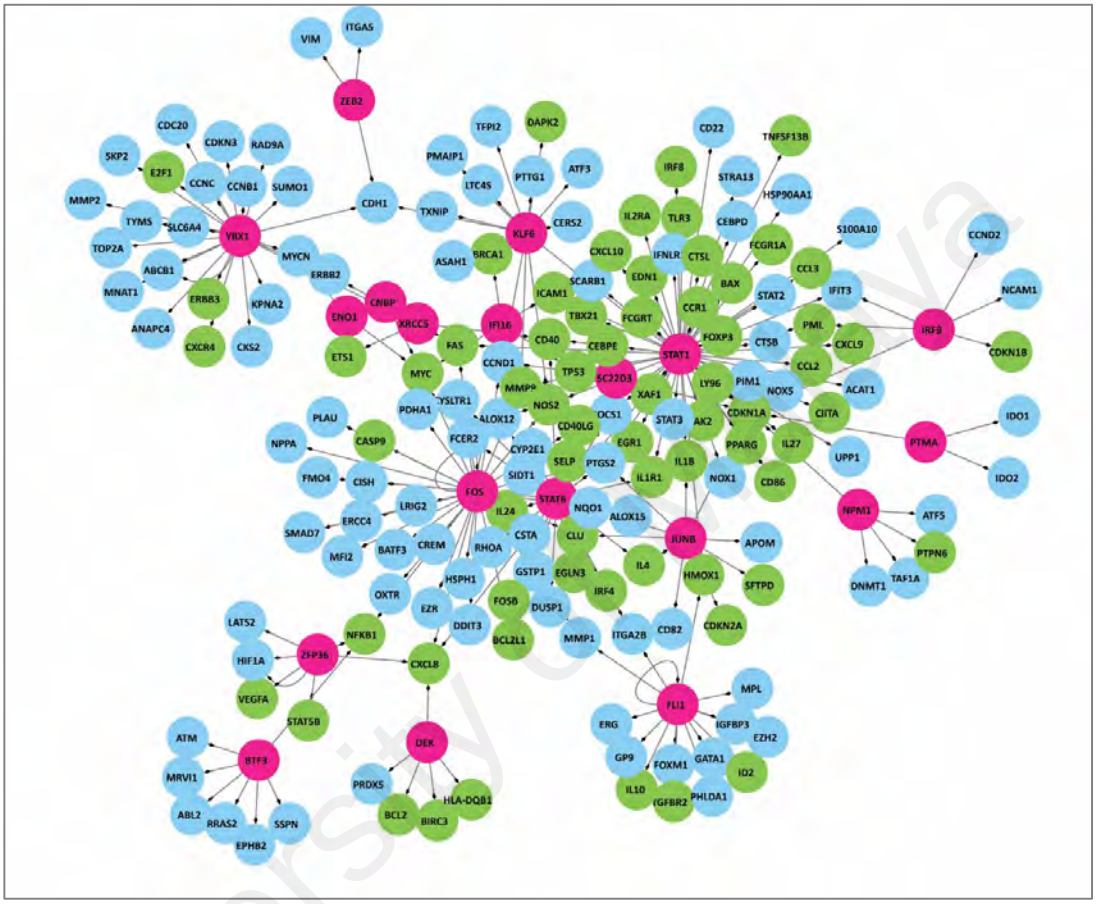

Figure 6.6: Interaction network of the top 20 highly expressed TFs with their target in primary monocytes. The network contains 226 interactions between 20 TFs and 146 targeted genes. The pink colour circle represents the TFs, while the blue and green colour circles represented the TF-target genes. The green colour circles represent the genes which are involved in immune system process and death.

6.3.2 Annotated lncRNAs expressed across hematopoietic cells

A recent study by Ranzani et al. (2015) characterized the lncRNAs in different subsets of lymphocytes. To investigate the expression patterns of the lncRNAs in various hematopoietic cells, a qualitative comparison was conducted on the expression patterns of annotated lncRNAs identified in our study with the lncRNAs expressed in lymphocyte subsets reported by Ranzani et al. (2015). The analysis showed several of the lncRNAs to be only specifically expressed in monocytes, these could either be truly monocyte-specific express in monocytes and other cell types that were not included in this comparison. Some lncRNAs were also identified to be only expressed in other hematopoietic cells and not monocytes. A total of 170 lncRNAs were identified to expressed in monocytes and lymphocyte subsets. A heat map showing the normalized expression values of these lncRNAs is presented in Figure 6.7. Several lncRNAs were identified, which had high expression levels across all monocytes and lymphocyte subsets. Such lncRNAs include *LINC-PINT, LINC01420* (long intergenic non-protein coding RNA 1420), *LINC00301* (Long Intergenic Non-Protein Coding RNA 301), *OIP5-AS1* (*OIP5* Antisense RNA 1), *RP13-297E16.4* (Clone-based (Vega)) and *MAPKAPK5-AS1* (*MAPKAPK5* antisense RNA 1). Some other lncRNAs including *AC021224.1* [Clone-based (Ensembl)] and *TPTEP1* (transmembrane phosphatase with tensin homology pseudogene 1) were found to be highly expressed in monocyte but lowly expressed in lymphocytes.

6.3.3 Novel lincRNAs expressed in primary monocytes

The multi-step mapping and filtering criteria were employed to identify putative novel lincRNAs in human monocytes. Briefly, the transcripts were mapped to known transcript annotations to filter out known transcripts and detect the potential novel transcripts. From novel transcripts, the transcripts with a minimum of 3x coverage and

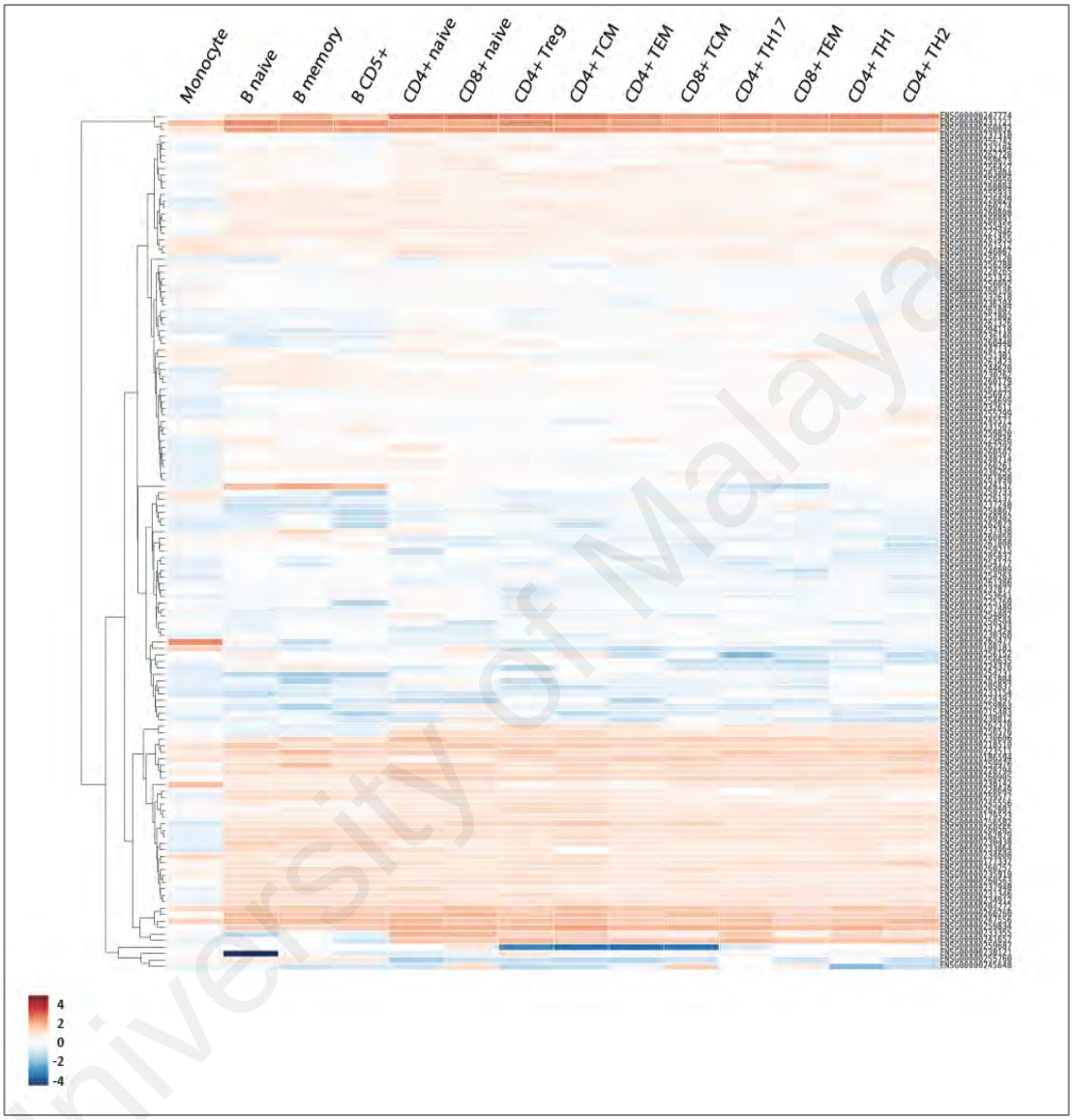

Figure 6.7: Heatmap showing normalized expression values (FPKM) of annotated lncRNAs across monocytes and other hematopoietic cell types. The colour scale indicates the $\log_2$ FPKM expression values (blue for low expression and red for high expression).

completely within the intergenic regions of annotated transcripts from GENCODE reference annotations were used for further analysis. Transcripts with single exons and less than 200 nucleotides were also excluded. The remaining subset of transcripts were assessed for their coding potential using coding potential assessment tool (CPAT). After applying this final coding potential filter, we could identify a total 1,032 potentials novel lincRNAs. A list of identified putative novel lincRNAs along with their expression levels as FPKMs is provided in APPENDIX D.

We examined if the exon distribution of identified novel lincRNAs follows a similar or different trend when compared to annotated lncRNAs from GENCODE. A pie chart showing the distribution of exons across the identified novel lincRNAs as compared to annotated lncRNAs is shown in [Figure 6.8 A]. Majority of the transcripts had two or three exons, which is also the case with annotated lncRNAs from GENCODE. The expression levels of the novel lincRNAs were then compared to the expression of protein-coding and annotated lncRNAs. The result showed that the average expression of all annotations lncRNAs and novel lincRNAs is relatively lower than the protein-coding genes across samples [Figure 6.8 B]. Also, a chromosome-wide distribution of all identified novel lincRNA transcripts showed majority of the novel lincRNAs from chromosomes 1, 2 and 7 [Figure 6.8 C] which may be due to the fact that these are among the longest chromosomes in humans.

6.3.4 Evolutionary conservation of novel lincRNAs

Using the RNAFold Vienna RNA package, the secondary structures of identified novel lincRNAs were predicted based on minimum free energies. The results indicated that these novel lincRNAs had stable negative energies enabling them to fold properly with an average minimum free energy of -508.42 kcal/ mol.

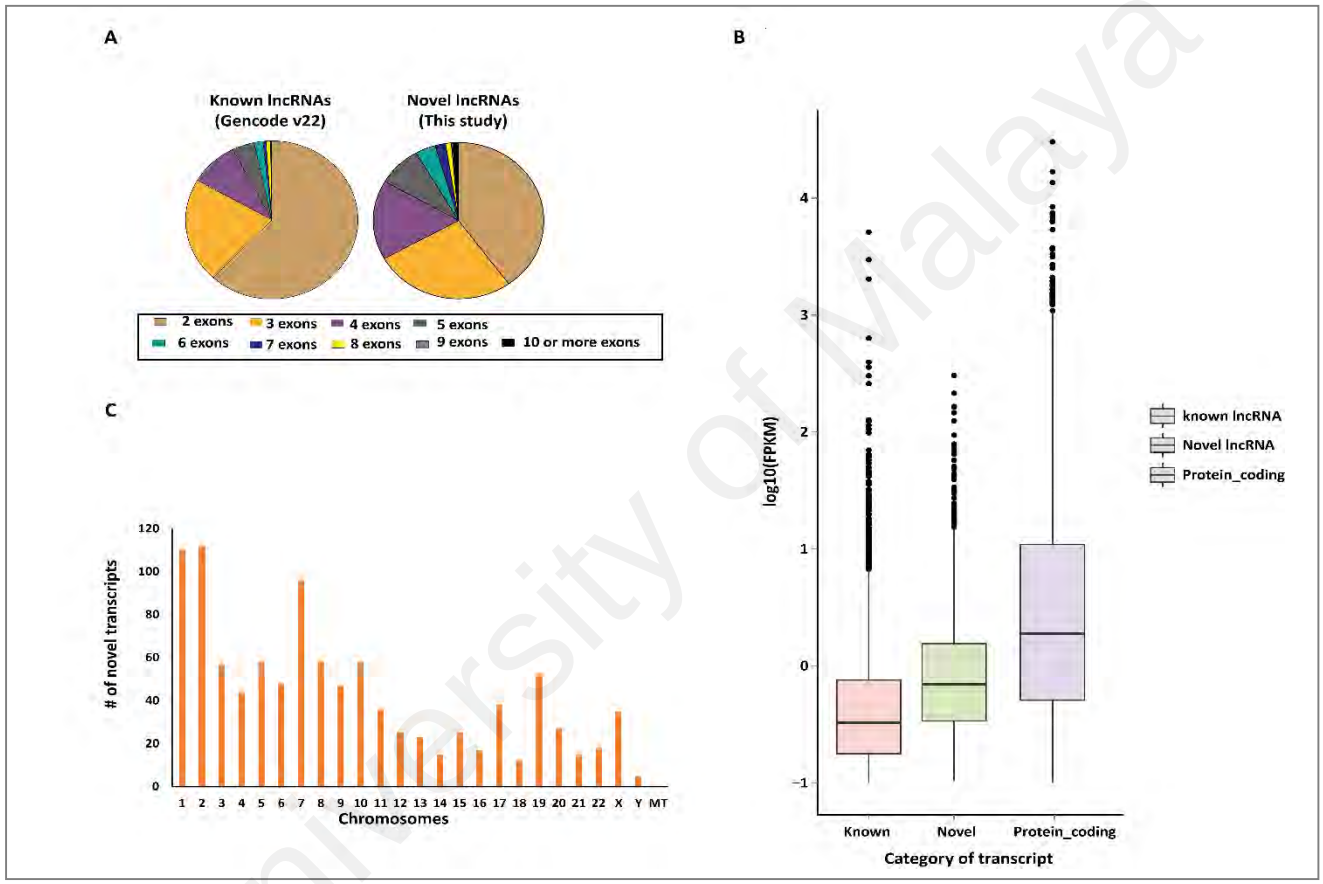

Figure 6.8: Characteristics of identified lncRNAs. A: Pie chart showing distribution of exons in annotated lncRNAs and novel lincRNAs. B: Box plot showing average expression levels of lncRNAs for protein-coding, previously annotated lncRNAs and newly identified lncRNAs in human monocytes. C: Number of novel lincRNAs identified for each chromosome.

6.3.5 Validation of lncRNAs by RT-PCR across hematopoietic cell types

A robust pipeline was developed for identification of potential novel lincRNAs from RNA-Seq data (see Figure 6.2, page 131). Using this strategy, several novel lincRNAs were identified which were not reported earlier in human monocytes. In order to see if the novel lincRNAs are only expressed in monocytes or are also expressed in other hematopoietic cells, a few subsets of annotated lncRNAs and novel lincRNAs were randomly selected for validation. The expression of five novel lincRNAs and two annotated lncRNAs were validated in monocytes and five other hematopoietic cell types using RT-PCR. The hematopoietic cells include T cell ($CD3^+$), T Helper cell ($CD4^+$), Regulatory T Cell ($CD4^+CD25^+$), Cytotoxic T cell ($CD8^+$) and B cell ($CD19^+$). Two annotated lncRNAs encoded by *DANCR* and *LINC01420* which were also detected in our datasets, were used as positive controls for RT-PCR. The RT-PCR analysis results (Figure 6.9) demonstrated that *DANCR* and *LINC01420* have moderate expression across all six hematopoietic cell types examined. The RT-PCR amplification and subsequent sequencing of the amplified cDNA fragments also supported the structures and expressions of 5 randomly selected novel lincRNAs in human primary monocytes. The schematic representation of validated novel lincRNAs structures and their positions are shown in Figure 6.10. Two of the five novel lincRNAs including TCONS_00282615 and TCONS_00008494 were found to be expressed only in monocytes but not in other hematopoietic cell types examined. The RT-PCR analysis also revealed high expression of novel lincRNAs TCONS_00128310 across monocytes, T cells (CD3+), T Helper cells (CD4+), and B Cell (CD19+), however moderate expression was observed in Regulatory T Cell (CD4+CD25+). Other novel lincRNAs including TCONS_00056181 and TCONS_00226266 exhibited moderate expression across all examined hematopoietic cells.

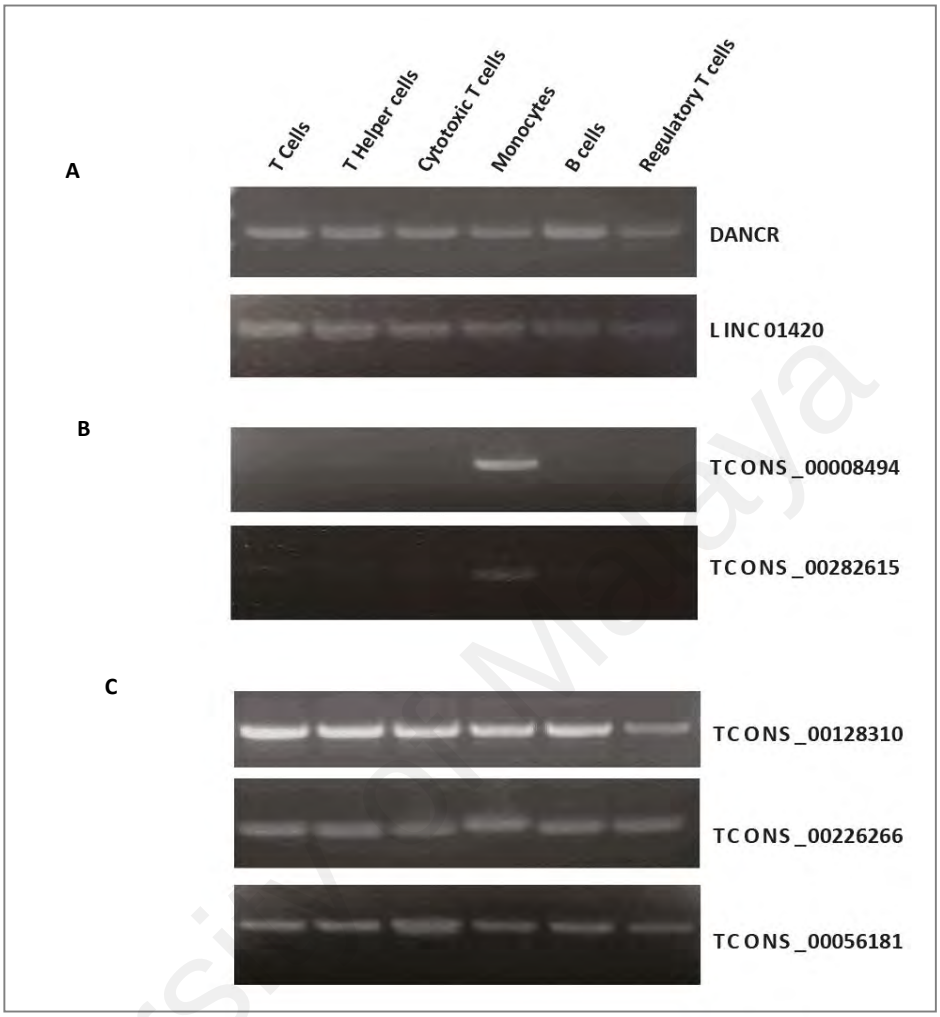

Figure 6.9: The RT-PCR validation of lncRNAs across hematopoietic cells. A: Two annotated lncRNAs (*DANCR*, *LINC01420*) expressed across all hematopoietic cells. B: Two novel lincRNAs showing expression only in monocytes. C: Three novel lincRNAs expressed ubiquitously across all hematopoietic cells. From left to right the lanes correspond to T cell ($CD3^+$), T Helper cell ($CD4^+$), Cytotoxic T cell ($CD8^+$), Monocyte ($CD14^+$), B cell ($CD19^+$) and Regulatory T cell ($CD4^+CD25^+$).

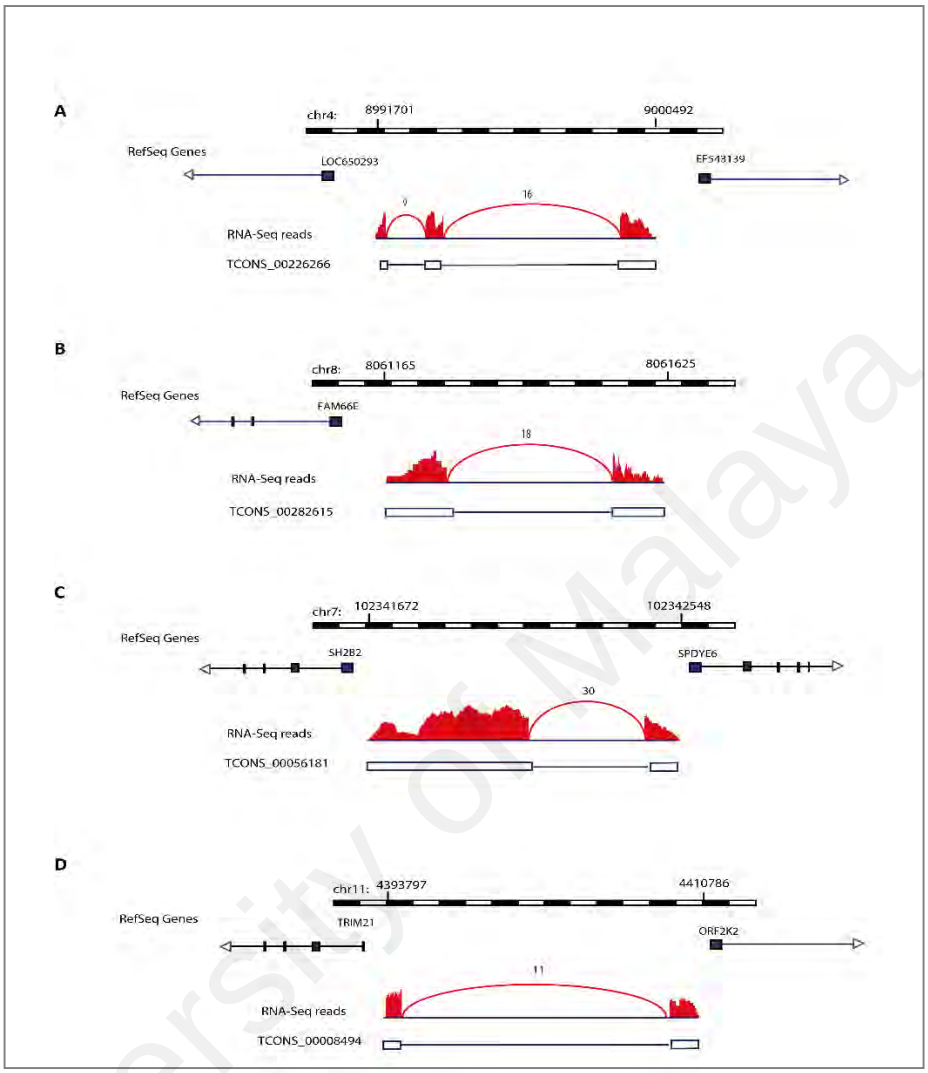

Figure 6.10: Schematic representation of novel lincRNAs exon architecture and genomic positions. RefSeq transcripts flanking the identified lincRNA are shown for each transcript (in blue colour). The exon model constructed by the Stringtie software is shown as white boxes. Read density for each exon is shown in red colour. Number of reads supporting each splice-junction are shown above each junction.

6.4 Discussion

Motivated by the ability of RNA-Seq technology to study gene expression, a deep RNA-Seq experiments on primary monocytes from 6 healthy subjects was performed. These RNA-Seq datasets were integrated with other publicly available RNA-Seq datasets for monocytes. Based on these datasets, we were able to capture most of genes transcribed in human monocytes including; 11,994 protein-coding genes, 5,558 non-coding genes, 2,820 pseudogenes, and 7,034 putative novel transcripts. The functional analysis of identified protein-coding genes expressed in monocytes revealed that highly expressed genes were mainly involved in several process belonged to immune system while the moderately expressed genes and lowly expressed genes were mainly involved in metabolic process and biological regulations. This results showed that genes within a particular process are expressed at similar levels which is in agreement with previous study (Toung et al., 2011).

The expression pattern of transcription factors (TFs) in human primary monocytes were profiled. TFs are key molecules that control gene transcriptions (Vaquerizas, Kummerfeld, Teichmann, & Luscombe, 2009). Over the past 30 years, several TFs involved in the immune system have been discovered and their mechanisms of action were studied (Smale, 2014). However, there is no report on the complete list of transcription factor expressed in monocytes. From this study, the expression of 1,155 TFs in human primary monocytes were detected. Among the highly expressed TFs, several TFs were found to be mainly involved in regulation of immune system such as *STAT1*, *STAT6*, *FOS*, *JUNB*, *FLI1*, *ZEP36* and *DEK*. *STAT1* is known as a regulator for several growth factors, cytokines and biological responses, based on phosphorylation and ligand-dependent tyrosine activation (Wen, Zhong, & Darnell, 1995). *FOS* is the component of *FOS* gene family which encode leucine zipper proteins that dimerize with JUN proteins and forming the transcription factor complex AP-1. These proteins have significant part

in regulation of cell corepressors, differentiation, and proliferation. *FOS* is also related to apoptotic cell decease in several diseases like cancer (Preston et al., 1996). The *FLI1* is a member of the *ETS* transcription factor (E26 transformation-specific) family, which expressed in haematopoietic cells. It plays a vital part in mononuclear phagocyte development and the *FLI1* C-terminal's transcriptional activation, which adversely controls the development of mononuclear phagocyte (Suzuki et al., 2013). *ZFP36* prevents pro-inflammatory expressions, such as transcriptional activation of *NF-κB* and by attaching to cytokine mRNAs, and also decreases transcript stability by binding to cytokine mRNAs (Zhang et al., 2013). *DEK* is a main component of metazoan chromatin. It is a nuclear phosphoprotein contributes to autoimmunity and oncogenesis (Lin et al., 2014). It influences cell-to-cell signaling, transcription regulation, mRNA splicing, cell viability and differentiation (Kappes et al., 2008). The high levels of DEK expression prevent apoptosis and senescence in the cells that are infected by the human papillomavirus E7 (Ageberg, Gullberg, & Lindmark, 2006).

Addition of publicly available RNA-Seq datasets for monocytes to our RNA-Seq dataset from healthy subjects also allowed us to provide a landscape of lncRNAs in human primary monocytes. Several recent studies have shown the role of lncRNAs in relation to immune regulation and their role in several autoimmune diseases like SLE (Shi et al., 2014) and rheumatoid arthritis (Müller et al., 2014). Recent studies have also indicated the possible role of lncRNAs expression and its correlation to immune cells differentiation and maturation (Stachurska, Zorro, van der Sijde, & Withoff, 2014). However most of the lncRNAs transcribed in the innate immune system remain unknown. From this study, the lncRNAs expression were characterized in human primary monocytes. The expression of 6,382 lncRNAs [40 % of the known lncRNAs reported in GENCODE database (version 22)] were identified across all the datasets studied. The majority of the lncRNAs are seen to be expressed at low levels which is in agreement

with previous reports. However, a fraction of these lncRNAs such as *GAS5*, *CEBPB-AS*, *LINC-PINT*, *ITGB2-AS1* and *FGD5-AS1* had high expression levels across all samples analyzed in the present study. The lncRNA encoded by *GAS5* is located on human chromosome 1q25.1. It is a multi-exonic (12 exons) transcript and is known to be a snoRNA host gene, which encodes multiple small nucleolar RNA (snoRNAs) within its introns. Three of these snoRNAs including *U44*, *U74* and *U78* may give rise to miRNAs. The exonic region of *GAS5* acts as a riborepressor of the glucocorticoid and related receptors (Mourtada-Maarabouni, Hedge, Kirkham, Farzaneh, & Williams, 2010). *GAS5* is reported to be an essential regulator of cell cycle and apoptosis in T cell lines and in non-transformed lymphocytes. *GAS5* is induced in cells during growth arrest, sensitized cells to apoptosis by inhibiting the transcription factor glucocorticoid receptor and preventing its localization to glucocorticoid-responsive elements (Kino, Hurt, Ichijo, Nader, & Chrousos, 2010). A high expression of *GAS5* across all monocytes samples warrants that it may be involved in regulation of monocytes cell cycle as well. The *LINC-PINT* is a gene encoding a nuclear lincRNA characterized by Marín-Béjar et al. (2013). It is located on human chromosome 7q32.3, between *MKLN1* (Muskelin 1) and *KLF14* (Kruppel-Like Factor 14) genes. The *LINC-PINT* is regulated by transcription factor *p53*, which is regulated by the expression of several lincRNAs and has a critical role in preservation of cellular homeostasis. *LINC-PINT* has been shown to be significantly downregulated in human colon cancer while its overexpression is reported to be a negative regulator of proliferation and survival of tumor cells (Marín-Béjar et al., 2013). Another highly expressed lncRNA identified in the monocytes is *ITGB2-AS1* which is located on human chromosome 21q22.3 and consists of 4 exons. The gene *ITGB2-AS1* is known to be involved in regulation of T cells and B cells activation. *ITGB2-AS1* exhibits high co-expression with multiple cancer genes such as *IKZF1* (IKAROS Family Zinc Finger 1) and *LCK* (LCK Proto-Oncogene, Src Family Tyrosine Kinase), which are

associated with acute lymphoid leukemia and *WAS* (Wiskott-Aldrich Syndrome) genes which is known to be associated with lymphoma. This anti-sense transcript has been reported to be an interesting candidate biomarkers for leukemia or lymphoma (Liangjiang Wang & Cogill, 2014). The high expression level of several uncharacterized lncRNAs were also observed in our datasets such as *LINC00678*, *LINC00657*, *LINC00989* and *LINC01420*, which may suggest their potential role in monocytes cell regulation. Further functional studies are needed to decipher the exact mechanism and function of these lncRNAs in monocytes.

lncRNAs have been reported to be specific at cell/tissue levels (Derrien et al., 2012). A comparison of the expression patterns of annotated lncRNAs identified in our study to the available lncRNAs catalogue of lymphocyte subsets reported by Ranzani et al, led to identification of several lncRNAs such as *AC021224.1* and *TPTEP1* which highly expressed in monocytes but exhibited lower expression in lymphocytes. Some other lncRNAs such as *LINC-PINT*, *LINC01420*, *LINC00301*, *OIP5-AS1*, *RP13-297E16.4* and *MAPKAPK5-AS1* were highly expressed in monocytes and lymphocytes suggest their possible roles in regulation of both innate and adaptive immune responses.

Using a computational pipeline developed in-house (see Figure 6.1 page 131), a total of 1,032 unannotated lincRNAs in monocytes were identified that has not been annotated in the public databases. lincRNAs have been reported as regulators of numerous physiological processes that involves gene regulation (Sánchez & Huarte, 2013). However, little is known about their function in the human immune system. A comparison of expression levels of the novel lincRNAs to the expression of protein-coding and annotated lncRNAs showed the both annotated and novel lincRNAs had a lower expression compared to the protein-coding genes which is consistent with previous reports (Derrien et al., 2012; Gonzalez-Porta, Calvo, Sammeth, & Guigo, 2012).

The expression of 2 annotated and 5 novel lincRNAs were validated using RT-PCR, across monocytes and five other hematopoietic cell types. The results showed that 2 novel lincRNAs *TCONS_00282615* and *TCONS_00008494* were only expressed in monocytes but not in other hematopoietic cell types examined. This suggests their specific functions in monocytes. Moreover 3 lincRNAs (*TCONS_00128310, TCONS_00056181* and *TCONS_00226266*) expressed across all examined hematopoietic cells, which indicated the possible role of these transcripts in regulation of hematopoietic cells.

In summary, this chapter provided gene reference catalogue of human primary monocytes under healthy states which could be used as resource for postgenomics and system biology research on human monocytes under healthy and diseased states. It would also offers significant promise for the fields of precision medicine, systems diagnostics, immunogenomics, and the development of innovative biomarkers and therapeutic monitoring strategies. In addition, a lncRNAs expressions landscape in human primary monocytes was provided which could be used as a starting point for designing functional studies and identify lncRNAs of immunopathlogical importance.

CHAPTER 7: GENERAL DISCUSSION AND CONCLUSION

Monocytes are essential components of the innate immune system. They are initiate from a common myeloid progenitor cells in the bone marrow and circulate in the blood vessels for short times, during inflammatory conditions they move into peripheral tissues and differentiate into macrophages and dendritic cells (Saha and Geissmann, 2011). Monocytes are heterogeneous and are divided into three groups based on their expression of CD14 and CD16 receptor markers; classical ($CD14^{++}CD16^{-}$), intermediate ($CD14^{++}CD16^{+}$) and non-classical ($CD14^{-}CD16^{++}$). Differences between monocytes subsets associate with differences in cytokine production, antigen presentation and antigen uptake (Ziegler-Heitbrock, 2010). The classical monocytes account for 90-95% of human blood monocytes. Several studies focused on the functional elements such as chemokine receptors, that distinguished monocytes subsets (Ancuta et al., 2009; Dong et al., 2013; Ziegler-Heitbrock et al., 2010), however, limited or no data exist on the genome-wide transcriptome expression profile of human primary monocytes.

X-linked agammaglobulinemia (XLA) is one of the inherited form of PIDs (Vetrie et al., 1993). XLA is caused by mutations in the *BTK* (Bruton Tyrosine Kinase), which is essential for B cell development and function (Noordzij et al., 2002; Ochs & Smith, 1996). The XLA disease is characterized by dramatic reduction of mature B cells in the peripheral blood as well as reduction of all serum immunoglobulin levels (Lopez-Herrera et al., 2014; Maas & Hendriks, 2001; Middendorp et al, 2003). The *BTK* expression is not restricted to B cells, it is also expressed in neutrophils (Honda et al., 2012), (Bao et al., 2012), and monocytes/macrophages (Koprulu & Ellmeier, 2009). However, up to now the role of *BTK* deficiency in primary monocytes of XLA patients is not well studied and there is no or limited data exist on the genome-wide transcriptome expression profile of primary monocytes under healthy and XLA disease states.

Transcription of eukaryotic genes is complex process. Although 90% of the eukaryotic genome is transcribed, only less than 2% has the capability to encode proteins (mRNAs) (Birney et al., 2007). The large fraction of eukaryotic transcripts represents non-coding RNAs which are classified into short non-coding RNAs (RNA species of length than 200 bp) and long non-coding RNAs (lncRNAs) (RNA species of length greater than 200bp) (Wilusz, Sunwoo, & Spector, 2009). lncRNAs are the major class of non-coding RNAs, localized in both nucleus and cytoplasmic, and be polyadenylated (poly(A)$^+$) or non-polyadenylated (poly(A)$^-$) (Yang, Duff, Graveley, Carmichael, & Chen, 2011).

Recently high-throughput RNA sequencing (RNA-Seq) based on next generation sequencing (NGS) technologies has been a popular approach for transcriptome profiling and identification of differentially expressed (DE) genes between different conditions (such as healthy and disease states) (Marioni, Mason, Mane, Stephens, & Gilad, 2008). RNA-Seq is superior to microarray analysis in terms of transcript quantity and quality (Balakrishnan, Lin, London, & Clayton, 2012). In addition, RNA-Seq provides better detection for gene fusion transcripts and splicing events in comparison with microarray. The poly(A)$^+$ library preparation for RNA-Seq enable the identification of expressed protein-coding genes as well as poly(A)$^+$ enriched lncRNAs (Yang et al., 2011).

A gene expression profiling of human primary monocytes was carried out from 6 healthy subjects and 3 XLA patients using deep RNA-Seq analysis. Total RNAs were extracted from classical monocytes (CD14^{++}CD16$^-$). Deep poly(A)$^+$ paired-end RNA-Seq was conducted on the extracted RNAs. The RNA-Seq datasets generated from healthy subjects were used to profile the expression of immune-relate genes in primary monocytes under healthy state. The analysis led to identify the expression of 804 genes related to immune system. Further comparative analysis of gene expression pattern between healthy male and female subjects revealed 217 DE protein-coding genes in male

compared to female in which 40 DE genes were related to immune system. Out of 40 DE immune-related genes, 23 and 17 genes were upregulated and downregulated in male compared to the female, respectively. The expression of several innate immune-related genes such as genes that involved in "Toll-like receptor signaling pathway" and "Cytokine-cytokine receptor interaction", were found to be significantly higher in female compared to male. These differences would provide important insights in understanding the disparity of innate immune response in male and female and could serve as target for drug discovery in the future. Moreover, the apoptosis-related genes *JUN* and *STAT1* were found to be significantly upregulated and downregulated in male compared to female respectively, which suggest that these genes could be involved in the sex predominance of immunity response. The expression patterns of *JUN* and *STAT1* in male compared to female were validated by qRT-PCR analysis which demonstrates reliability of our RNA-Seq data analysis. To the best of our knowledge, this is the first report on genome-wide transcriptome expression profile of human primary monocytes that exhibited differences between male and female for immune-related genes. Further large scale studies utilizing integrated omics approaches would advance our knowledge on the disparity of immune-related genes based on gender at the population levels. In addition, in order to provide a reference gene catalogue of human monocytes under healthy states, the RNA-Seq datasets generated from our healthy subjects were integrated with other publicly available RNA-Seq datasets of human monocytes. These datasets allowed us to capture most of genes transcribed in human monocytes, including 11,994 protein-coding genes, 5,558 non-coding genes, 2,819 pseudogenes, and 7,034 putative novel transcripts. In addition, the expression of 1,155 transcription factors (TFs) were identified in human monocytes, which are the main molecules in controlling the gene transcription. An interaction network was constructed among the top expressed TFs and their targeted genes which revealed the potential key regulatory genes in biological function of human monocytes

(Mirsafian, Ripen, Manaharan, Mohamad & Merican, 2016). Using the integrated RNA-Seq datasets, the expression patterns of lncRNAs were also characterized in human monocytes. Our analysis led to identification of around 6,382 annotated lncRNAs and 1,032 novel lincRNA which have not been previously reported in monocytes. The expression of 4 annotated and 3 novel lincRNAs were validated across various hematopoietic cells using RT-PCR (Mirsafian et al., 2016). The results showed that there are several lncRNAs which are expressed in either tissue/cell specific manner. Also, most of the lncRNAs have low expression, which could easily be missed by shallow sequencing. In this study, the poly(A)$^+$ enriched population of lncRNAs was examined. The another major fraction of lncRNAs that lack the poly(A)$^+$ tail (Yang et al., 2011) still remains to be explored as they are processed differently. Also, functional studies have to be designed which could help in identifying the role of these lncRNAs in immune cell regulation.

A comparative study has also been conducted on RNA-Seq datasets of 3 XLA patients and 3 healthy male subjects generated in our study. To obtain a representative picture of differentially expressed genes from the sample size used in this study, an adequately large fold-change cutoff threshold ($\log_2$ fold-change ≥ 1 or ≤ -1) was used to define the DE genes. The analysis revealed that the expression profiles of protein-coding genes and lncRNAs were significantly altered in primary monocytes of XLA patients compared to the healthy subjects. The total of 1,827 DE protein-coding genes including 859 upregulated and 968 downregulated genes were identified in XLA patients compared to the healthy subjects. Regardless type of mutations, loss of function mutation in *BTK* genes eventually induced dysfunction of *BTK* protein, block the development of B-cell and caused the disease. Functional annotation analysis indicated that several innate immune-related pathways such as “Fc gamma R-mediated phagocytosis”, “Chemokine signaling pathways”, “Toll like receptors signaling pathway” and “MTOR signaling

pathway" were enriched for downregulated genes. These results suggest the overall dysregulation of innate immune system in XLA patients. Additionally, the "oxidative phosphorylation" pathway, along with ROS production, response to oxidative stress and apoptotic process were significantly enriched for upregulated genes in XLA patients compared to the healthy subjects, suggesting a great demand of energy production in primary monocytes and their susceptibility to apoptosis in XLA patients. This showed that *BTK* is not only involved in the development and function of B cells, but it may play an important role in establishing the immunity function of monocytes. Moreover, the total of 95 DE annotated lncRNAs including 56 upregulated and 39 downregulated and 20 DE novel lincRNAs including 5 upregulated and 15 downregulated were identified in XLA patients compared to healthy subjects. lncRNAs have been shown to control several functions in both cellular and developmental processes including cancers (Tsai, Spitale, & Chang, 2011). Although some of the lncRNAs have been implicated in the regulation of the immune response (Yu, Wang, & Morris, 2015), the exact function of the large majority of lncRNAs still remains unknown. Our analysis detected the differential expression of lncRNAs *HOTAIRM1*, *DANCR*, *GAS5*, *LINC-PINT*, *HEIH*, *RMRP*, and *TUG1* in XLA patients, in which their dysregulated expressions have been reported to suppress cellular development, differentiation and proliferation, and apoptosis process in many diseases. The dysregulation of these lncRNAs observed in primary monocytes of XLA patients suggest their possible role in cell cycle regulation and apoptosis in monocytes of XLA patients. However, further functional characterization of these lncRNAs in primary monocytes of XLA patients would help us decipher their actual role. Interestingly, our results also identified the downregulation of novel lncRNA *TCONS_00030433* which interacted with phagocytosis-related gene *VAV3* (expression of VAV3 was observed to be downregulated in the XLA patients). This finding suggest the possible role of *TCONS_00030433* in the downregulation of phagocytosis pathway in

XLA patients. The expression pattern of selected of DE protein-coding genes and DE lncRNAs identified in XLA patients were validated by qRT-PCR, supporting the reliability of our RNA-Seq data analysis.

In summary, this study provides the first comprehensive gene expression signatures of human primary monocytes under healthy and XLA disease states based on deep RNA-Seq analysis which could be used as a starting point for postgenomics and system biology research on human monocytes. A set of DE protein-coding genes and DE lncRNAs identified in XLA patients compared to the healthy subjects opens several possible avenues of research that will help us to understand the complex pathophysiology in XLA provide compelling evidence for a potential genomic biomarker for XLA.

REFERENCES

Abbas, A. K., Lichtman, A. H., & Pillai, S. (2012). *Cellular and molecular immunology* (7th ed). Philadelphia: Elsevier/Saunders.

Abdullah, M., Chai, P.-S., Chong, M.-Y., Tohit, E. R. M., Ramasamy, R., Pei, C. P., & Vidyadaran, S. (2012). Gender effect on in vitro lymphocyte subset levels of healthy individuals. *Cellular Immunology*, *272*(2), 214–219.

Agarwal, A., Koppstein, D., Rozowsky, J., Sboner, A., Habegger, L., Hillier, L. W., … Gerstein, M. (2010). Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. *BMC Genomics*, *11*, 383.

Ageberg, M., Gullberg, U., & Lindmark, A. (2006). The involvement of cellular proliferation status in the expression of the human proto-oncogene DEK. *Haematologica*, *91*(2), 268–269.

Alugupalli, K. R., Akira, S., Lien, E., & Leong, J. M. (2007). MyD88- and Bruton's tyrosine kinase-mediated signals are essential for T cell-independent pathogen-specific IgM responses. *Journal of Immunology (Baltimore, Md.: 1950)*, *178*(6), 3740–3749.

Ameratunga, R., Woon, S.-T., Neas, K., & Love, D. R. (2010). The clinical utility of molecular diagnostic testing for primary immune deficiency disorders: a case based review. *Allergy, Asthma & Clinical Immunology*, *6*(1), 12.

Amoras, A. L., Kanegane, H., Miyawaki, T., & Vilela, M. M. (2003). Defective Fc-, CR1- and CR3-mediated monocyte phagocytosis and chemotaxis in common variable immunodeficiency and X-linked agammaglobulinemia patients. *Journal of Investigational Allergology & Clinical Immunology*, *13*(3), 181–188.

Ancuta, P., Liu, K.-Y., Misra, V., Wacleche, V., Gosselin, A., Zhou, X., & Gabuzda, D. (2009). Transcriptional profiling reveals developmental relationship and distinct biological functions of CD16+ and CD16- monocyte subsets. *BMC Genomics*, *10*(1), 403.

Anker, M. (2007). Addressing sex and gender in epidemic-prone infectious diseases. Genève (Suisse): World Health Organization.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.

Atianand, M. K., & Fitzgerald, K. A. (2014). Long non-coding RNAs and control of gene expression in the immune system. *Trends in Molecular Medicine*, *20*(11), 623–631.

Auffray, C., Sieweke, M. H., & Geissmann, F. (2009). Blood Monocytes: Development, Heterogeneity, and Relationship with Dendritic Cells. *Annual Review of Immunology*, *27*(1), 669–692.

Balakrishnan, C. N., Lin, Y.-C., London, S. E., & Clayton, D. F. (2012). RNA-seq transcriptome analysis of male and female zebra finch cell lines. *Genomics*, *100*(6), 363–369.

Baldi, P., & Hatfield, G. W. (2002). *DNA microarrays and gene expression: from experiments to data analysis and modeling* (1st paperback ed., Paperback reissue). Cambridge: Cambridge Univ. Press.

Bao, Y., Zheng, J., Han, C., Jin, J., Han, H., Liu, Y., … Cao, X. (2012). Tyrosine Kinase Btk Is Required for NK Cell Activation. *Journal of Biological Chemistry*, *287*(28), 23769–23778.

Barna, M., Komatsu, T., Bi, Z., & Reiss, C. S. (1996). Sex differences in susceptibility to viral infection of the central nervous system. *Journal of Neuroimmunology*, *67*(1), 31–39.

Batchelor, J. R. (1968). Hormonal control of antibody formation. Regulation of the antibody response. *C. B. Springfield and C. C. Thomas, Eds*, 276–293.

Benito-Martin, A., Di Giannatale, A., Ceder, S., & Peinado, H. (2015). The New Deal: A Potential Role for Secreted Vesicles in Innate Immunity and Tumor Progression. *Frontiers in Immunology*, *6*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Berger, M. F., Levin, J. Z., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., … Garraway, L. A. (2010). Integrative analysis of the melanoma transcriptome. *Genome Research*, *20*(4), 413–427.

Beyer, M., Mallmann, M. R., Xue, J., Staratschek-Jox, A., Vorholt, D., Krebs, W., … Schultze, J. L. (2012). High-Resolution Transcriptome of Human Macrophages. *PLoS ONE*, *7*(9), e45466.

Bhatia, A., Sekhon, H. K., & Kaur, G. (2014). Sex Hormones and Immune Dimorphism. *The Scientific World Journal*, *2014*, 1–8.

Bi, H., Zhou, J., Wu, D., Gao, W., Li, L., Yu, L., … Yao, X. (2015). Microarray analysis of long non-coding RNAs in COPD lung tissue. *Inflammation Research*, *64*(2), 119–126.

Bijl, M. (2006). Reduced uptake of apoptotic cells by macrophages in systemic lupus erythematosus: correlates with decreased serum levels of complement. *Annals of the Rheumatic Diseases*, *65*(1), 57–63.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., … de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816.

Bousfiha, A., Jeddane, L., Al-Herz, W., Ailal, F., Casanova, J., Chatila, T., … Tang, M. L. K. (2015). The 2015 IUIS Phenotypic Classification for Primary Immunodeficiencies. *Journal of Clinical Immunology*, *35*(8), 727–738.

Bowcock, A. M., Shannon, W., Du, F., Duncan, J., Cao, K., Aftergut, K., … Menter, A. (2001). Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies. *Human Molecular Genetics*, *10*(17), 1793–1805.

Bradford, J. R., Hey, Y., Yates, T., Li, Y., Pepper, S. D., & Miller, C. J. (2010). A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics*, *11*(1), 282.

Brunner, A. L., Beck, A. H., Edris, B., Sweeney, R. T., Zhu, S. X., Li, R., … West, R. B. (2012). Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biology*, *13*(8), R75.

Bruton, O. C. (1952). Agammaglobulinemia. *Pediatrics*, *9*(6), 722–728.

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics*, *17*(5), 257–271.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927.

Calippe, B., Douin-Echinard, V., Laffargue, M., Laurell, H., Rana-Poussine, V., Pipy, B., … Gourdy, P. (2008). Chronic estradiol administration in vivo promotes the proinflammatory response of macrophages to TLR4 activation: involvement of the phosphatidylinositol 3-kinase pathway. *Journal of Immunology (Baltimore, Md.: 1950)*, *180*(12), 7980–7988.

Candore, G., Balistreri, C. R., Colonna-Romano, G., Lio, D., Listì, F., Vasto, S., & Caruso, C. (2010). Gender-Related Immune-Inflammatory Factors, Age-Related Diseases, and Longevity. *Rejuvenation Research*, *13*(2–3), 292–297.

Carpenter, S., & Fitzgerald, K. A. (2015). Transcription of Inflammatory Genes: Long Noncoding RNA and Beyond. *Journal of Interferon & Cytokine Research*, *35*(2), 79–88.

Caulfield, J., Fernandez, M., Snetkov, V., Lee, T., & Hawrylowicz, C. (2002). CXCR4 expression on monocytes is up-regulated by dexamethasone and is modulated by autologous CD3+ T cells. *Immunology*, *105*(2), 155–162.

Cervantes, J. L., Weinerman, B., Basole, C., & Salazar, J. C. (2012). TLR8: the forgotten relative revindicated. *Cellular and Molecular Immunology*, *9*(6), 434–438.

Chakrabarti, S., Lekontseva, O., & Davidge, S. T. (2008). Estrogen is a modulator of vascular inflammation. *IUBMB Life*, *60*(6), 376–382.

Chear, C. T., Gill, H. K., Ramly, N. H., Dhaliwal, J. S., Bujang, N., Ripen, A. M., & Mohamad, S. B. (2013). A novel Bruton's tyrosine kinase gene (BTK) invariant splice site mutation in a Malaysian family with X-linked agammaglobulinemia. *Asian Pacific Journal of Allergy and Immunology*, *31*(4), 320.

Chen, F. (2003). JUN (v-Jun sarcoma virus 17 oncogene homolog (avian)), *7*(2), 98–99.

Chow, A., Brown, B. D., & Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. *Nature Reviews Immunology*, *11*(11), 788–798.

Chu, Y., & Corey, D. R. (2012). RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*, *22*(4), 271–274.

Chusid, M. J., Coleman, C. M., & Dunne, W. M. (1987). Chronic asymptomatic Campylobacter bacteremia in a boy with X-linked hypogammaglobulinemia. *The Pediatric Infectious Disease Journal*, *6*(10), 943–944.

Clark, M. B., & Mattick, J. S. (2011). Long noncoding RNAs in cell biology. *Seminars in Cell & Developmental Biology*, *22*(4), 366–376.

Clark, T. A. (2002). Genomewide Analysis of mRNA Processing in Yeast Using Splicing-Specific Microarrays. *Science*, *296*(5569), 907–910.

Conley, M. E., Broides, A., Hernandez-Trujillo, V., Howard, V., Kanegane, H., Miyawaki, T., & Shurtleff, S. A. (2005). Genetic analysis of patients with defects in early B-cell development. *Immunological Reviews*, *203*, 216–234.

Conley, M. E., & Howard, V. (2002). Clinical findings leading to the diagnosis of X-linked agammaglobulinemia. *The Journal of Pediatrics*, *141*(4), 566–571.

Conley, M. E., Rohrer, J., & Minegishi, Y. (2000). X-Linked Agammaglobulinemia. *Clinical Reviews in Allergy & Immunology*, *19*(2), 183–204.

Cook, M. B., McGlynn, K. A., Devesa, S. S., Freedman, N. D., & Anderson, W. F. (2011). Sex Disparities in Cancer Mortality and Survival. *Cancer Epidemiology Biomarkers & Prevention*, *20*(8), 1629–1637.

Cros, J., Cagnard, N., Woollard, K., Patey, N., Zhang, S.-Y., Senechal, B., … Geissmann, F. (2010). Human CD14dim Monocytes Patrol and Sense Nucleic Acids and Viruses via TLR7 and TLR8 Receptors. *Immunity*, *33*(3), 375–386.

Dang, Y., Xu, X., Shen, Y., Hu, M., Zhang, M., Li, L., … Li, J. (2016). Transcriptome Analysis of the Innate Immunity-Related Complement System in Spleen Tissue of Ctenopharyngodon idella Infected with Aeromonas hydrophila. *PLOS ONE*, *11*(7), e0157413.

de Gorter, D. J. J., Beuling, E. A., Kersseboom, R., Middendorp, S., van Gils, J. M., Hendriks, R. W., … Spaargaren, M. (2007). Bruton's Tyrosine Kinase and Phospholipase Cγ2 Mediate Chemokine-Controlled B Cell Migration and Homing. *Immunity*, *26*(1), 93–104.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, *4*(5), P3.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigo, R. (2012a). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigo, R. (2012b). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789.

Diodato, M. D., Knöferl, M. W., Schwacha, M. G., Bland, K. I., & Chaudry, I. H. (2001). Gender differences in the inflammatory response and survival following haemorrhage and subsequent sepsis. *Cytokine*, *14*(3), 162–169.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108.

Dong, C., Zhao, G., Zhong, M., Yue, Y., Wu, L., & Xiong, S. (2013). RNA sequencing and transcriptomal analysis of human monocyte to macrophage differentiation. *Gene*, *519*(2), 279–287.

Dong, Z., & Chen, Y. (2013). Transcriptomics: advances and approaches. *Science China. Life Sciences*, *56*(10), 960–967.

Doyle, S. L., Jefferies, C. A., Feighery, C., & O'Neill, L. A. J. (2007). Signaling by Toll-like receptors 8 and 9 requires Bruton's tyrosine kinase. *The Journal of Biological Chemistry*, *282*(51), 36953–36960.

Dunkelberger, J. R., & Song, W.-C. (2010). Complement and its role in innate and adaptive immune responses. *Cell Research*, *20*(1), 34–50.

Elling, R., Chan, J., & Fitzgerald, K. A. (2016). Emerging role of long noncoding RNAs as regulators of innate immune cell development and inflammatory gene expression. *European Journal of Immunology*, *46*(3), 504–512.

Elphinston, R., & Wickes, I. (1956). Familial Agammaglobulinaemia, *2*(4988), 336–338.

Ezell, S. A., Wang, S., Bihani, T., Lai, Z., Grosskurth, S. E., Tepsuporn, S., … Byth, K. F. (2016). Differential regulation of mTOR signaling determines sensitivity to AKT inhibition in diffuse large B cell lymphoma. *Oncotarget*, *7*(8), 9163–9174.

Fan, H., Dong, G., Zhao, G., Liu, F., Yao, G., Zhu, Y., & Hou, Y. (2014). Gender Differences of B Cell Signature in Healthy Subjects Underlie Disparities in Incidence and Course of SLE Related to Estrogen. *Journal of Immunology Research*, *2014*, 1–17.

Finkel, T. (2012). Signal transduction by mitochondrial oxidants. *The Journal of Biological Chemistry*, *287*(7), 4434–4440.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279-285.

Finotello, F., & Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Briefings in Functional Genomics*, *14*(2), 130–142.

Fish, E. N. (2008). The X-files in immunity: sex-based differences predispose immune responses. *Nature Reviews Immunology*, *8*(9), 737–744.

Fisher, R. A. (1922). On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, *85*(1), 87.

Fitzgerald, K. A., Palsson-McDermott, E. M., Bowie, A. G., Jefferies, C. A., Mansell, A. S., Brady, G., … O'Neill, L. A. J. (2001). Mal (MyD88-adapter-like) is required for Toll-like receptor-4 signal transduction. *Nature*, *413*(6851), 78–83.

Fried, A. J., & Bonilla, F. A. (2009). Pathogenesis, Diagnosis, and Management of Primary Antibody Deficiencies and Infections. *Clinical Microbiology Reviews*, *22*(3), 396–414.

Futatani, T., Miyawaki, T., Tsukada, S., Hashimoto, S., Kunikata, T., Arai, S., … Kishimoto, T. (1998). Deficient expression of Bruton's tyrosine kinase in monocytes from X-linked agammaglobulinemia as evaluated by a flow cytometric analysis and its clinical application to carrier detection. *Blood*, *91*(2), 595–602.

Gack, M. U. (2014). Mechanisms of RIG-I-Like Receptor Activation and Manipulation by Viral Pathogens. *Journal of Virology*, *88*(10), 5213–5216.

Gaspar, H. B., & Kinnon, C. (2001). X-linked agammaglobulinemia. *Immunology and Allergy Clinics of North America*, *21*(1), 23–43.

Gaujoux, R., & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, *11*(1), 367. https://doi.org/10.1186/1471-2105-11-367

Geissmann, F., Jung, S., & Littman, D. R. (2003). Blood monocytes consist of two principal subsets with distinct migratory properties. *Immunity*, *19*(1), 71–82.

Geissmann, F., Manz, M. G., Jung, S., Sieweke, M. H., Merad, M., & Ley, K. (2010). Development of Monocytes, Macrophages, and Dendritic Cells. *Science*, *327*(5966), 656–661.

Geng, H., & Tan, X.-D. (2016). Functional diversity of long non-coding RNAs in immune regulation. *Genes & Diseases*, *3*(1), 72–81.

Gibb, E. A., Brown, C. J., & Lam, W. L. (2011). The functional role of long non-coding RNA in human carcinomas. *Molecular Cancer*, *10*(1), 38.

Gomes, A., Nolasco, S., & Soares, H. (2013). Non-Coding RNAs: Multi-Tasking Molecules in the Cell. *International Journal of Molecular Sciences*, *14*(8), 16010–16039.

Gonzalez-Porta, M., Calvo, M., Sammeth, M., & Guigo, R. (2012). Estimation of alternative splicing variability in human populations. *Genome Research*, *22*(3), 528–538.

Griguer, C. E., Oliva, C. R., & Gillespie, G. Y. (2005). Glucose metabolism heterogeneity in human and mouse malignant glioma cell lines. *Journal of Neuro-Oncology*, *74*(2), 123–133.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., … Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227.

Hall, A. B., Gakidis, M. A. M., Glogauer, M., Wilsbacher, J. L., Gao, S., Swat, W., & Brugge, J. S. (2006). Requirements for Vav Guanine Nucleotide Exchange Factors and Rho GTPases in FcγR- and Complement-Mediated Phagocytosis. *Immunity*, *24*(3), 305–316.

Halliday, E., Winkelstein, J., & Webster, A. D. B. (2003). Enteroviral infections in primary immunodeficiency (PID): a survey of morbidity and mortality. *The Journal of Infection*, *46*(1), 1–8.

Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., … Lee, I. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. *Scientific Reports*, *5*, 11432.

Hanna, N., & Schneider, M. (1983). Enhancement of tumor metastasis and suppression of natural killer cell activity by beta-estradiol treatment. *Journal of Immunology (Baltimore, Md.: 1950)*, *130*(2), 974–980.

Hashimoto, S., Tsukada, S., Matsushita, M., Miyawaki, T., Niida, Y., Yachie, A., … Kishimoto, T. (1996). Identification of Bruton's tyrosine kinase (Btk) gene mutations and characterization of the derived proteins in 35 X-linked agammaglobulinemia families: a nationwide study of Btk deficiency in Japan. *Blood*, *88*(2), 561–573.

Hata, D., Kawakami, Y., Inagaki, N., Lantz, C. S., Kitamura, T., Khan, W. N., … Kawakami, T. (1998). Involvement of Bruton's tyrosine kinase in FcepsilonRI-dependent mast cell degranulation and cytokine production. *The Journal of Experimental Medicine*, *187*(8), 1235–1247.

Hawn, T. R., Verbon, A., Lettinga, K. D., Zhao, L. P., Li, S. S., Laws, R. J., … Aderem, A. (2003). A Common Dominant TLR5 Stop Codon Polymorphism Abolishes Flagellin Signaling and Is Associated with Susceptibility to Legionnaires' Disease. *The Journal of Experimental Medicine*, *198*(10), 1563–1572.

Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., … Aderem, A. (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, *410*(6832), 1099–1103.

Hendriks, R. W., Bredius, R. G., Pike-Overzet, K., & Staal, F. J. (2011). Biology and novel treatment options for XLA, the most common monogenetic immunodeficiency in man. *Expert Opinion on Therapeutic Targets*, *15*(8), 1003–1021.

Hernandez-Trujillo, V. P., Scalchunes, C., Cunningham-Rundles, C., Ochs, H. D., Bonilla, F. A., Paris, K., … Sullivan, K. E. (2014). Autoimmunity and Inflammation in X-linked Agammaglobulinemia. *Journal of Clinical Immunology*, *34*(6), 627–632.

Hewagama, A., Patel, D., Yarlagadda, S., Strickland, F. M., & Richardson, B. C. (2009). Stronger inflammatory/cytotoxic T-cell response in women identified by microarray analysis. *Genes and Immunity*, *10*(5), 509–516.

Honda, F., Kano, H., Kanegane, H., Nonoyama, S., Kim, E.-S., Lee, S.-K., … Morio, T. (2012). The kinase Btk negatively regulates the production of reactive oxygen species and stimulation-induced apoptosis in human neutrophils. *Nature Immunology*, *13*(4), 369–378.

Horwood, N. J., Page, T. H., McDaid, J. P., Palmer, C. D., Campbell, J., Mahon, T., … Foxwell, B. M. J. (2006a). Bruton's tyrosine kinase is required for TLR2 and TLR4-induced TNF, but not IL-6, production. *Journal of Immunology (Baltimore, Md.: 1950)*, *176*(6), 3635–3641.

Horwood, N. J., Page, T. H., McDaid, J. P., Palmer, C. D., Campbell, J., Mahon, T., … Foxwell, B. M. J. (2006b). Bruton's tyrosine kinase is required for TLR2 and TLR4-induced TNF, but not IL-6, production. *Journal of Immunology (Baltimore, Md.: 1950)*, *176*(6), 3635–3641.

Hosack, D. A., Dennis, G., Sherman, B. T., Lane, H. C., & Lempicki, R. A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biology*, *4*(10), R70.

Hrdlickova, B., Kumar, V., Kanduri, K., Zhernakova, D. V., Tripathi, S., Karjalainen, J., … Withoff, S. (2014). Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genome Medicine*, *6*(10).

Hu, G., Tang, Q., Sharma, S., Yu, F., Escobar, T. M., Muljo, S. A., … Zhao, K. (2013). Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nature Immunology*, *14*(11), 1190–1198.

Ilott, N. E., Heward, J. A., Roux, B., Tsitsiou, E., Fenwick, P. S., Lenzi, L., … Lindsay, M. A. (2014). Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nature Communications*, *5*.

Imamura, K., & Akimitsu, N. (2014). Long Non-Coding RNAs Involved in Immune Responses. *Frontiers in Immunology*, *5*.

Jacobs, Z. D., Guajardo, J. R., & Anderson, K. M. (2008). XLA-associated Neutropenia Treatment: A Case Report and Review of the Literature. *Journal of Pediatric Hematology/Oncology*, *30*(8), 631–634.

Janeway, C. A. (Ed.). (2001). *Immunobiology: the immune system in health and disease ; [animated CD-ROM inside]* (5. ed). New York, NY: Garland Publ. [u.a.].

Janeway, C. A., & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, *20*, 197–216.

Jefferies, C. A., Doyle, S., Brunner, C., Dunne, A., Brint, E., Wietek, C., … O'Neill, L. A. J. (2003). Bruton's Tyrosine Kinase Is a Toll/Interleukin-1 Receptor Domain-binding Protein That Participates in Nuclear Factor B Activation by Toll-like Receptor 4. *Journal of Biological Chemistry*, *278*(28), 26258–26264.

Jongstra-Bilen, J., Puig Cano, A., Hasija, M., Xiao, H., Smith, C. I. E., & Cybulsky, M. I. (2008). Dual functions of Bruton's tyrosine kinase and Tec kinase during Fcgamma receptor-induced signaling and phagocytosis. *Journal of Immunology (Baltimore, Md.: 1950)*, *181*(1), 288–298.

Kanegane, H., Futatani, T., Wang, Y., Nomura, K., Shinozaki, K., Matsukura, H., … Miyawaki, T. (2001). Clinical and mutational characteristics of X-linked agammaglobulinemia and its carrier identified by flow cytometric assessment combined with genetic analysis. *Journal of Allergy and Clinical Immunology*, *108*(6), 1012–1020.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.

Kappes, F., Fahrer, J., Khodadoust, M. S., Tabbert, A., Strasser, C., Mor-Vaknin, N., … Ferrando-May, E. (2008). DEK Is a Poly(ADP-Ribose) Acceptor in Apoptosis and Mediates Resistance to Genotoxic Stress. *Molecular and Cellular Biology*, *28*(10), 3245–3257.

Karapetyan, A., Buiting, C., Kuiper, R., & Coolen, M. (2013). Regulatory Roles for Long ncRNA and mRNA. *Cancers*, *5*(2), 462–490.

Kawai, T., & Akira, S. (2009). The roles of TLRs, RLRs and NLRs in pathogen recognition. *International Immunology*, *21*(4), 317–337.

Keating, R., & McGargill, M. A. (2016). mTOR Regulation of Lymphoid Cells in Immunity to Pathogens. *Frontiers in Immunology*, *7*.

Kerstens, P. J., Endtz, H. P., Meis, J. F., Oyen, W. J., Koopman, R. J., van den Broek, P. J., & van der Meer, J. W. (1992). Erysipelas-like skin lesions associated with Campylobacter jejuni septicemia in patients with hypogammaglobulinemia. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, *11*(9), 842–847.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360.

Kim, E.-D., & Sung, S. (2012). Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends in Plant Science*, *17*(1), 16–21.

Kim, H. S., & Lee, M.-S. (2007). STAT1 as a key modulator of cell death. *Cellular Signalling*, *19*(3), 454–465. https://doi.org/10.1016/j.cellsig.2006.09.003

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., & Chrousos, G. P. (2010). Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. *Science Signaling*, *3*(107), ra8-ra8.

Koppenol, W. H., Bounds, P. L., & Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Reviews. Cancer*, *11*(5), 325–337.

Koprulu, A. D., & Ellmeier, W. (2009). The role of Tec family kinases in mononuclear phagocytes. *Critical Reviews in Immunology*, *29*(4), 317–333.

Kraft-Terry, S. D., & Gendelman, H. E. (2011). Proteomic biosignatures for monocyte–macrophage differentiation. *Cellular Immunology*, *271*(2), 239–255.

Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2015). Tissue-Specific Evolution of Protein Coding Genes in Human and Mouse. *PLOS ONE*, *10*(6), e0131673.

Kuby, J., Kindt, T. J., Goldsby, R. A., & Osborne, B. A. (2007). *Immunology* (6. ed., 5. pr). New York: Freeman.

Lahn, B. T., Pearson, N. M., & Jegalian, K. (2001). The human Y chromosome, in the light of evolution. *Nature Reviews Genetics*, *2*(3), 207–216.

Lederman, H. M., & Winkelstein, J. A. (1985). X-linked agammaglobulinemia: an analysis of 96 patients. *Medicine*, *64*(3), 145–156.

Lee, H.-J., Kim, K.-J., Park, M.-H., Kimm, K., Park, C., Oh, B., & Lee, J.-Y. (2005). Single-Nucleotide Polymorphisms and Haplotype LD Analysis of the 29-kb IGF2 Region on Chromosome 11p15.5 in the Korean Population. *Human Heredity*, *60*(2), 73–80.

Lee, K.-G., Xu, S., Kang, Z.-H., Huo, J., Huang, M., Liu, D., … Lam, K.-P. (2012). Bruton's tyrosine kinase phosphorylates Toll-like receptor 3 to initiate antiviral response. *Proceedings of the National Academy of Sciences*, *109*(15), 5791–5796.

Lee, P. P. W., Chen, T.-X., Jiang, L.-P., Chan, K.-W., Yang, W., Lee, B.-W., … Lau, Y.-L. (2010). Clinical Characteristics and Genotype-phenotype Correlation in 62 Patients with X-linked Agammaglobulinemia. *Journal of Clinical Immunology*, *30*(1), 121–131.

Li, Z., Chao, T.-C., Chang, K.-Y., Lin, N., Patil, V. S., Shimizu, C., … Rana, T. M. (2014). The long noncoding RNA THRIL regulates TNF expression through its interaction with hnRNPL. *Proceedings of the National Academy of Sciences*, *111*(3), 1002–1007.

Li, Z., & Rana, T. M. (2014). Decoding the noncoding: Prospective of lncRNA-mediated innate immune regulation. *RNA Biology*, *11*(8), 979–985.

Libert, C., Dejager, L., & Pinheiro, I. (2010). The X chromosome in immune functions: when a chromosome makes the difference. *Nature Reviews Immunology*, *10*(8), 594–604.

Lim, W.-S., Timmins, J. M., Seimon, T. A., Sadler, A., Kolodgie, F. D., Virmani, R., & Tabas, I. (2008). Signal Transducer and Activator of Transcription-1 Is Critical for Apoptosis in Macrophages Subjected to Endoplasmic Reticulum Stress In Vitro and in Advanced Atherosclerotic Lesions In Vivo. *Circulation*, *117*(7), 940–951.

Lin, L., Piao, J., Ma, Y., Jin, T., Quan, C., Kong, J., … Lin, Z. (2014). Mechanisms Underlying Cancer Growth and Apoptosis by DEK Overexpression in Colorectal Cancer. *PLoS ONE*, *9*(10), e111260.

Lindsay, M. A. (2008). microRNAs and the immune response. *Trends in Immunology*, *29*(7), 343–351.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, *25*(4), 402–408.

López-Granados, E., Pérez de Diego, R., Ferreira Cerdán, A., Fontán Casariego, G., & García Rodríguez, M. C. (2005). A genotype-phenotype correlation study in a group of 54 patients with X-linked agammaglobulinemia. *Journal of Allergy and Clinical Immunology*, *116*(3), 690–697.

Lopez-Herrera, G., Vargas-Hernandez, A., Gonzalez-Serrano, M. E., Berron-Ruiz, L., Rodriguez-Alba, J. C., Espinosa-Rosales, F., & Santos-Argumedo, L. (2014). Bruton's tyrosine kinase--an integral protein of B cell development that also has

an essential role in the innate immune system. *Journal of Leukocyte Biology*, *95*(2), 243–250.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 26.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., … Murray, C. J. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2095–2128.

Maarschalk-Ellerbroek, L. J., Hoepelman, I. M., & Ellerbroek, P. M. (2011). Immunoglobulin treatment in primary antibody deficiency. *International Journal of Antimicrobial Agents*, *37*(5), 396–404.

Maas, A., & Hendriks, R. W. (2001). Role of Bruton's Tyrosine Kinase in B Cell Development. *Developmental Immunology*, *8*(3–4), 171–181.

Manczak, M., Park, B. S., Jung, Y., & Reddy, P. H. (2004). Differential expression of oxidative phosphorylation genes in patients with Alzheimer's disease: implications for early mitochondrial dysfunction and oxidative damage. *Neuromolecular Medicine*, *5*(2), 147–162.

Marín-Béjar, O., Marchese, F. P., Athie, A., Sánchez, Y., González, J., Segura, V., … Huarte, M. (2013). Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biology*, *14*(9), R104.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*(9), 1509–1517.

Marron, T. U., Martinez-Gallo, M., Yu, J. E., & Cunningham-Rundles, C. (2012). Toll-like receptor 4–, 7–, and 8–activated myeloid cells from patients with X-linked agammaglobulinemia produce enhanced inflammatory cytokines. *Journal of Allergy and Clinical Immunology*, *129*(1), 184–190.e4.

Marron, T. U., Rohr, K., Martinez-Gallo, M., Yu, J., & Cunningham-Rundles, C. (2010). TLR signaling and effector functions are intact in XLA neutrophils. *Clinical Immunology*, *137*(1), 74–80.

Martens-Uzunova, E. S., Böttcher, R., Croce, C. M., Jenster, G., Visakorpi, T., & Calin, G. A. (2014). Long Noncoding RNA in Prostate, Bladder, and Kidney Cancer. *European Urology*, *65*(6), 1140–1151.

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*(10), 671–682.

Martinez, F. O. (2009). The transcriptome of human monocyte subsets begins to emerge. *Journal of Biology*, *8*(11), 99.

McCusker, C., & Warrington, R. (2011). Primary immunodeficiency. *Allergy, Asthma & Clinical Immunology*, *7*(Suppl 1), S11.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*(5), 495–501.

Medzhitov, R., & Janeway, C. A. (1997). Innate immunity: the virtues of a nonclonal system of recognition. *Cell*, *91*(3), 295–298.

Middendorp, S., Dingjan, G. M., Maas, A., Dahlenborg, K., & Hendriks, R. W. (2003). Function of Bruton's Tyrosine Kinase during B Cell Development Is Partially Independent of Its Catalytic Activity. *The Journal of Immunology*, *171*(11), 5988–5996.

Mirsafian, H., Manda, S. S., Mitchell, C. J., Sreenivasamurthy, S., Ripen, A. M., Mohamad, S. B., … Pandey, A. (2016). Long non-coding RNA expression in primary human monocytes. *Genomics*, *108*(1), 37–45.

Mirsafian, H., Ripen, A. M., Manaharan, T., Mohamad, S. B., & Merican, A. F. (2016). Towards a Reference Gene Catalogue of Human Primary Monocytes. *OMICS*, *20*(11), 627–634.

Mitchell, C. J., Getnet, D., Kim, M.-S., Manda, S. S., Kumar, P., Huang, T.-C., … Pandey, A. (2015). A multi-omic analysis of human naïve CD4+ T cells. *BMC Systems Biology*, *9*(1).

Mohamed, A. J., Yu, L., Bäckesjö, C.-M., Vargas, L., Faryal, R., Aints, A., … Edvard Smith, C. I. (2009). Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunological Reviews*, *228*(1), 58–73.

Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S. L., … Bader, G. D. (2010). GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*, *26*(22), 2927–2928.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628.

Mourtada-Maarabouni, M., Hedge, V. L., Kirkham, L., Farzaneh, F., & Williams, G. T. (2010). Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *Journal of Cell Science*, *123*(7), 1181–1181.

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., … Sartorelli, V. (2013). eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Molecular Cell*, *51*(5), 606–617.

Muenchhoff, M., & Goulder, P. J. R. (2014). Sex Differences in Pediatric Infectious Diseases. *Journal of Infectious Diseases*, *209*(suppl 3), S120–S126.

Mukhopadhyay, S., Mohanty, M., Mangla, A., George, A., Bal, V., Rath, S., & Ravindran, B. (2002). Macrophage effector functions controlled by Bruton's tyrosine kinase are more crucial than the cytokine balance of T cell responses for microfilarial clearance. *Journal of Immunology (Baltimore, Md.: 1950)*, *168*(6), 2914–2921.

Müller, N., Döring, F., Klapper, M., Neumann, K., Schulte, D. M., Türk, K., … Laudes, M. (2014). Interleukin-6 and tumour necrosis factor-α differentially regulate lincRNA transcripts in cells of the innate immune system in vivo in human subjects with rheumatoid arthritis. *Cytokine*, *68*(1), 65–68.

Murdoch, C., & Finn, A. (2000). Chemokine receptors and their role in inflammation and infectious diseases. *Blood*, *95*(10), 3032–3043.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, *320*(5881), 1344–1349.

Natoli, G., & Andrau, J.-C. (2012). Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annual Review of Genetics*, *46*(1), 1–19. https://doi.org/10.1146/annurev-genet-110711-155459

Ngo, S. T., Steyn, F. J., & McCombe, P. A. (2014). Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology*, *35*(3), 347–369.

Noordzij, J. G., De Bruin-Versteeg, S., Comans-Bitter, W. M., Hartwig, N. G., Hendriks, R. W., De Groot, R., & van Dongen, J. J. M. (2002). Composition of Precursor B-Cell Compartment in Bone Marrow from Patients with X-Linked Agammaglobulinemia Compared with Healthy Children. *Pediatric Research*, *51*(2), 159–168.

O'Brien, J., & Sereda, M. (1956). Agammaglobulinæmia. *Canadian Medical Association Journal*, *74*(9), 723–724.

Ochs, H. D., & Smith, C. I. (1996). X-linked agammaglobulinemia. A clinical and molecular analysis. *Medicine*, *75*(6), 287–299.

O'Connell, R. M., Rao, D. S., Chaudhuri, A. A., & Baltimore, D. (2010). Physiological and pathological roles for microRNAs in the immune system. *Nature Reviews Immunology*, *10*(2), 111–122.

Oertelt-Prigione, S. (2012). The influence of sex and gender on the immune response. *Autoimmunity Reviews*, *11*(6–7), A479–A485.

O'Mahony, D. S., Pham, U., Iyer, R., Hawn, T. R., & Liles, W. C. (2008). Differential constitutive and cytokine-modulated expression of human Toll-like receptors in primary neutrophils, monocytes, and macrophages. *International Journal of Medical Sciences*, *5*(1), 1–8.

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., … Shiekhattar, R. (2010). Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell*, *143*(1), 46–58.

Ortona, E., Delunardo, F., Maselli, A., Pierdominici, M., & Malorni, W. (2015). Sex hormones and gender disparity in immunity and autoimmunity. *Italian Journal of Gender-Specific Medicine*, (2015October-December).

Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biology*, *11*(12), 220.

Passlick, B., Flieger, D., & Ziegler-Heitbrock, H. W. (1989). Identification and characterization of a novel monocyte subpopulation in human peripheral blood. *Blood*, *74*(7), 2527–2534.

Pérez de Diego, R., López-Granados, E., Rivera, J., Ferreira, A., Fontán, G., Bravo, J., … Bolland, S. (2008). Naturally occurring Bruton's tyrosine kinase mutations have no dominant negative effect in an X-linked agammaglobulinaemia cellular model: Gene therapy feasibility in XLA with Btk expression. *Clinical & Experimental Immunology*, *152*(1), 33–38.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295.

Phillips, R. J., Lutz, M., & Premack, B. (2005). Differential signaling mechanisms regulate expression of CC chemokine receptor-2 during monocyte maturation. *Journal of Inflammation (London, England)*, *2*, 14.

Picard, C., Al-Herz, W., Bousfiha, A., Casanova, J.-L., Chatila, T., Conley, M. E., … Gaspar, H. B. (2015). Primary Immunodeficiency Diseases: an Update on the

Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *Journal of Clinical Immunology*, *35*(8), 696–726.

Piras, V., & Selvarajoo, K. (2014). Beyond MyD88 and TRIF Pathways in Toll-Like Receptor Signaling. *Frontiers in Immunology*, *5*.

Plebani, A., Soresina, A., Rondelli, R., Amato, G. M., Azzari, C., Cardinale, F., … Italian Pediatric Group for XLA-AIEOP. (2002). Clinical, immunological, and molecular analysis in a large cohort of patients with X-linked agammaglobulinemia: an Italian multicenter study. *Clinical Immunology (Orlando, Fla.)*, *104*(3), 221–230.

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, *136*(4), 629–641.

Porter, K. A., Duffy, E. B., Nyland, P., Atianand, M. K., Sharifi, H., & Harton, J. A. (2014). The CLRX.1/NOD24 (NLRP2P) pseudogene codes a functional negative regulator of NF-κB, pyrin-only protein 4. *Genes and Immunity*, *15*(6), 392–403.

Powell, J. D., Pollizzi, K. N., Heikamp, E. B., & Horton, M. R. (2012). Regulation of Immune Responses by mTOR. *Annual Review of Immunology*, *30*(1), 39–68.

Preston, G. A., Lyon, T. T., Yin, Y., Lang, J. E., Solomon, G., Annab, L., … Barrett, J. C. (1996). Induction of apoptosis by c-Fos protein. *Molecular and Cellular Biology*, *16*(1), 211–218.

Ranzani, V., Rossetti, G., Panzeri, I., Arrigoni, A., Bonnal, R. J. P., Curti, S., … Pagani, M. (2015). The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nature Immunology*, *16*(3), 318–325.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., … Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*(9), R95.

Rapicavoli, N. A., Qu, K., Zhang, J., Mikhail, M., Laberge, R.-M., & Chang, H. Y. (2013). A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife*, *2*.

Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., … Hayashizaki, Y. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, *140*(5), 744–752.

Reddy, K. (2014). Host Autophagy Response: Friend or Foe in Reproductive Tract Infections. *SOJ Microbiology & Infectious Diseases*, *2*(2).

Roach, J. C., Smith, K. D., Strobe, K. L., Nissen, S. M., Haudenschild, C. D., Zhou, D., … Aderem, A. (2007). Transcription factor expression in lipopolysaccharide-activated peripheral-blood-derived mononuclear cells. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(41), 16245–16250.

Roifman, C. M., Rao, C. P., Lederman, H. M., Lavi, S., Quinn, P., & Gelfand, E. W. (1986). Increased susceptibility to Mycoplasma infection in patients with hypogammaglobulinemia. *The American Journal of Medicine*, *80*(4), 590–594.

Rosales, C. (Ed.). (2005). *Molecular mechanisms of phagocytosis*. Georgetown, Tex. : New York: Landes Bioscience/Eurekah.com ; Springer Science+Business Media.

Rudge, P., Webster, A. D. B., Revesz, T., Warner, T., Espanol, T., Cunningham-Rundles, C., & Hyman, N. (1996). Encephalomyelitis in primary hypogammaglobulinaemia. *Brain*, *119*(1), 1–15.

Sag, A. T., Saka, E., Ozgur, T. T., Sanal, O., Ayvaz, D. C., Elibol, B., & Kurne, A. T. (2014). Progressive Neurodegenerative Syndrome in a Patient with X-Linked Agammaglobulinemia Receiving Intravenous Immunoglobulin Therapy: *Cognitive And Behavioral Neurology*, *27*(3), 155–159.

Sánchez, Y., & Huarte, M. (2013). Long Non-Coding RNAs: Challenges for Diagnosis and Therapies. *Nucleic Acid Therapeutics*, *23*(1), 15–20.

Sarpong, S., Skolnick, H. S., Ochs, H. D., Futatani, T., & Winkelstein, J. A. (2002). Survival of wild polio by a patient with XLA. *Annals of Allergy, Asthma & Immunology*, *88*(1), 59–60.

Sárvári, M., Hrabovszky, E., Kalló, I., Solymosi, N., Tóth, K., Likó, I., … Liposits, Z. (2011). Estrogens regulate neuroinflammatory genes via estrogen receptors α and β in the frontal cortex of middle-aged female rats. *Journal of Neuroinflammation*, *8*(1), 82.

Satpathy, A. T., & Chang, H. Y. (2015). Long Noncoding RNA in Hematopoiesis and Immunity. *Immunity*, *42*(5), 792–804.

Schlee, M. (2013). Master sensors of pathogenic RNA–RIG-I like receptors. *Immunobiology*, *218*(11), 1322–1335.

Schmidt, N. W., Thieu, V. T., Mann, B. A., Ahyi, A.-N. N., & Kaplan, M. H. (2006). Bruton's tyrosine kinase is required for TLR-induced IL-10 production. *Journal of Immunology (Baltimore, Md.: 1950)*, *177*(10), 7203–7210.

Schroeder, H. W., Schroeder, H. W., & Sheikh, S. M. (2004). The complex genetics of common variable immunodeficiency. *Journal of Investigative Medicine: The*

Official Publication of the American Federation for Clinical Research, *52*(2), 90–103.

Scott, D. A., Richardson, A. D., Filipp, F. V., Knutzen, C. A., Chiang, G. G., Ronai, Z. A., … Smith, J. W. (2011). Comparative metabolic flux profiling of melanoma cell lines: beyond the Warburg effect. *Journal of Biological Chemistry*, *286*(49), 42626–42634.

Sentman, C. L., Meadows, S. K., Wira, C. R., & Eriksson, M. (2004). Recruitment of uterine NK cells: induction of CXC chemokine ligands 10 and 11 in human endometrium by estradiol and progesterone. *Journal of Immunology (Baltimore, Md.: 1950)*, *173*(11), 6760–6766.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.

Shaulian, E., & Karin, M. (2002). AP-1 as a regulator of cell life and death. *Nature Cell Biology*, *4*(5), E131–E136.

Shayakhmetov, D. M. (2010). Virus Infection Recognition and Early Innate Responses to Non-Enveloped Viral Vectors. *Viruses*, *2*(1), 244–261.

Sherrill-Mix, S., Ocwieja, K. E., & Bushman, F. D. (2015). Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. *Retrovirology*, *12*(1).

Shi, L., Zhang, Z., Yu, A. M., Wang, W., Wei, Z., Akhter, E., … Sullivan, K. E. (2014). The SLE Transcriptome Exhibits Evidence of Chronic Endotoxin Exposure and Has Widespread Dysregulation of Non-Coding and Coding RNAs. *PLoS ONE*, *9*(5), e93846.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121–132.

Sirri, V., Urcuqui-Inchima, S., Roussel, P., & Hernandez-Verdun, D. (2008). Nucleolus: the fascinating nuclear body. *Histochemistry and Cell Biology*, *129*(1), 13–31.

Smale, S. T. (2014). Transcriptional regulation in the immune system: a status report. *Trends in Immunology*, *35*(5), 190–194.

Sochorova, K., Horvath, R., Rozkova, D., Litzman, J., Bartunkova, J., Sediva, A., & Spisek, R. (2007). Impaired Toll-like receptor 8-mediated IL-6 and TNF-production in antigen-presenting cells from patients with X-linked agammaglobulinemia. *Blood*, *109*(6), 2553–2556.

Soliman, G. A. (2013). The role of mechanistic target of rapamycin (mTOR) complexes signaling in the immune responses. *Nutrients*, *5*(6), 2231–2257.

Song, J., Kim, D., Han, J., Kim, Y., Lee, M., & Jin, E.-J. (2015). PBMC and exosome-derived Hotair is a critical regulator and potent marker for rheumatoid arthritis. *Clinical and Experimental Medicine*, *15*(1), 121–126.

Souza, S. S., Castro, F. A., Mendonça, H. C., Palma, P. V., Morais, F. R., Ferriani, R. A., & Voltarelli, J. C. (2001). Influence of menstrual cycle on NK activity. *Journal of Reproductive Immunology*, *50*(2), 151–159.

Stachurska, A., Zorro, M. M., van der Sijde, M. R., & Withoff, S. (2014). Small and Long Regulatory RNAs in the Immune System and Immune Diseases. *Frontiers in Immunology*, *5*.

Suárez-Fariñas, M., Ungar, B., Correa da Rosa, J., Ewald, D. A., Rozenblit, M., Gonzalez, J., … Guttman-Yassky, E. (2015). RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications. *Journal of Allergy and Clinical Immunology*, *135*(5), 1218–1227.

Sun, D., & Ding, A. (2006). MyD88-mediated stabilization of interferon-γ-induced cytokine and chemokine mRNA. *Nature Immunology*, *7*(4), 375–381.

Suri, D., Rawat, A., & Singh, S. (2016). X-linked Agammaglobulinemia. *The Indian Journal of Pediatrics*, *83*(4), 331–337.

Suzuki, E., Williams, S., Sato, S., Gilkeson, G., Watson, D. K., & Zhang, X. K. (2013). The transcription factor Fli-1 regulates monocyte, macrophage and dendritic cell development in mice. *Immunology*, *139*(3), 318–327.

Takeuchi, O., Sato, S., Horiuchi, T., Hoshino, K., Takeda, K., Dong, Z., … Akira, S. (2002). Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *Journal of Immunology (Baltimore, Md.: 1950)*, *169*(1), 10–14.

Tao, L., Boyd, M., Gonye, G., Malone, B., & Schwaber, J. (2000). BTK mutations in patients with X-linked agammaglobulinemia: Lack of correlation between presence of peripheral B lymphocytes and specific mutations. *Human Mutation*, *16*(6), 528–529.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. *Genome Research*, *21*(12), 2213–2223.

Taverner, N. V., Smith, J. C., & Wardle, F. C. (2004). Identifying transcriptional targets. *Genome Biology*, *5*(3), 1.

Teocchi, M. A., Domingues Ramalho, V., Abramczuk, B. M., D'Souza-Li, L., & Santos Vilela, M. M. (2015). *BTK* mutations selectively regulate BTK expression and upregulate monocyte *XBP1* mRNA in XLA patients: Hyperexpression of monocyte *XBP1* in XLA. *Immunity, Inflammation and Disease*, *3*(3), 171–181.

Thomson, A. W., Turnquist, H. R., & Raimondi, G. (2009). Immunoregulatory functions of mTOR inhibition. *Nature Reviews Immunology*, *9*(5), 324–337.

Tian, L., Wang, P., Guo, J., Wang, X., Deng, W., Zhang, C., … Ma, D. (2007). Screening for novel human genes associated with CRE pathway activation with cell microarray. *Genomics*, *90*(1), 28–34.

Tizard, I. R. (2013). *Veterinary immunology* (9th ed). St. Louis, Mo: Elsevier/Saunders.

Tohyama, M., Sayama, K., Komatsuzawa, H., Hanakawa, Y., Shirakata, Y., Dai, X., … Hashimoto, K. (2007). CXCL16 is a novel mediator of the innate immunity of epidermal keratinocytes. *International Immunology*, *19*(9), 1095–1102.

Tomicic, M. T., Meise, R., Aasland, D., Berte, N., Kitzinger, R., Krämer, O. H., … Christmann, M. (2015). Apoptosis induced by temozolomide and nimustine in glioblastoma cells is supported by JNK/c-Jun-mediated induction of the BH3-only protein BIM. *Oncotarget*, *6*(32), 33755–33768.

Tong, X., Gu, P., Xu, S., & Lin, X. (2015). Long non-coding RNA-DANCR in human circulating monocytes: a potential biomarker associated with postmenopausal osteoporosis. *Bioscience, Biotechnology, and Biochemistry*, *79*(5), 732–737.

Toung, J. M., Morley, M., Li, M., & Cheung, V. G. (2011). RNA-sequence analysis of human B-cells. *Genome Research*, *21*(6), 991–998.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., … Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515.

Tsuchihara, K., Suzuki, Y., Wakaguri, H., Irie, T., Tanimoto, K., Hashimoto, S. -i., … Sugano, S. (2009). Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Research*, *37*(7), 2249–2263.

Tu, Z.-Q., Li, R.-J., Mei, J.-Z., & Li, X.-H. (2014). Down-regulation of long non-coding RNA GAS5 is associated with the prognosis of hepatocellular carcinoma. *International Journal of Clinical and Experimental Pathology*, *7*(7), 4303–4309.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804–810.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115–e115.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, *10*(4), 252–263.

Varol, C., Yona, S., & Jung, S. (2009). Origins and tissue-context-dependent fates of blood monocytes. *Immunology and Cell Biology*, *87*(1), 30–38.

Vetrie, D., Vořechovský, I., Sideras, P., Holland, J., Davies, A., Flinter, F., … Bentley, D. R. (1993). The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature*, *361*(6409), 226–233.

Vihinen, M., Kwan, S.-P., Lester, T., Ochs, H. D., Resnick, I., Väliaho, J., … Smith, C. I. E. (1999). Mutations of the humanBTK gene coding for bruton tyrosine kinase in X-linked agammaglobulinemia. *Human Mutation*, *13*(4), 280–285.

Vorechovský, I., Vihinen, M., de Saint Basile, G., Honsová, S., Hammarström, L., Müller, S., … Smith, C. I. (1995). DNA-based mutation analysis of Bruton's tyrosine kinase gene in patients with X-linked agammaglobulinaemia. *Human Molecular Genetics*, *4*(1), 51–58.

Voskuhl, R. (2011). Sex differences in autoimmune diseases. *Biology of Sex Differences*, *2*(1), 1.

Wallace, C., Smyth, D. J., Maisuria-Armer, M., Walker, N. M., Todd, J. A., & Clayton, D. G. (2010). The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nature Genetics*, *42*(1), 68–71.

Wan, L., Kong, J., Tang, J., Wu, Y., Xu, E., Lai, M., & Zhang, H. (2016). HOTAIRM1 as a potential biomarker for diagnosis of colorectal cancer functions the role in the tumour suppressor. *Journal of Cellular and Molecular Medicine*, *20*(11), 2036–2044.

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research*, *41*(W1), W77–W83.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., … Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, *472*(7341), 120–124.

Wang, L., & Cogill, S. (2014). Co-expression Network Analysis of Human lncRNAs and Cancer Genes. *Cancer Informatics*, 49.

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., & Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Research*, *41*(6), e74–e74.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63.

Wapinski, O., & Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends in Cell Biology*, *21*(6), 354–361. https://doi.org/10.1016/j.tcb.2011.04.001

Weber, M., Gawanbacht, A., Habjan, M., Rang, A., Borner, C., Schmidt, A. M., … Weber, F. (2013). Incoming RNA Virus Nucleocapsids Containing a 5′-Triphosphorylated Genome Activate RIG-I and Antiviral Signaling. *Cell Host & Microbe*, *13*(3), 336–346.

Weichhart, T., Hengstschläger, M., & Linke, M. (2015). Regulation of innate immune cell function by mTOR. *Nature Reviews Immunology*, *15*(10), 599–614.

Weinstein, Y., Ran, S., & Segal, S. (1984). Sex-associated differences in the regulation of immune responses controlled by the MHC of the mouse. *Journal of Immunology (Baltimore, Md.: 1950)*, *132*(2), 656–661.

Wen, Z., Zhong, Z., & Darnell, J. E. (1995). Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell*, *82*(2), 241–250.

Whitney, A. R., Diehn, M., Popper, S. J., Alizadeh, A. A., Boldrick, J. C., Relman, D. A., & Brown, P. O. (2003). Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences*, *100*(4), 1896–1901.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., … Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*(7199), 1239–1243.

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes & Development*, *23*(13), 1494–1504.

Wong, K. L., Tai, J. J.-Y., Wong, W.-C., Han, H., Sem, X., Yeap, W.-H., … Wong, S.-C. (2011). Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood*, *118*(5), e16–e31.

Yamamoto, M. (2003). Role of Adaptor TRIF in the MyD88-Independent Toll-Like Receptor Signaling Pathway. *Science*, *301*(5633), 640–643.

Yang, F., Zhang, L., Huo, X., Yuan, J., Xu, D., Yuan, S., … Sun, S. (2011). Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology*, *54*(5), 1679–1689.

Yang, J., Zhang, L., Yu, C., Yang, X.-F., & Wang, H. (2014). Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. *Biomarker Research*, *2*(1), 1.

Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G., & Chen, L.-L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biology*, *12*(2), R16.

Yona, S., & Jung, S. (2010). Monocytes: subsets, origins, fates and functions: *Current Opinion in Hematology*, *17*(1), 53–59.

Young, N. A., Wu, L.-C., Burd, C. J., Friedman, A. K., Kaffenberger, B. H., Rajaram, M. V. S., … Jarjour, W. N. (2014). Estrogen modulation of endosome-associated toll-like receptor 8: An IFNα-independent mechanism of sex-bias in systemic lupus erythematosus. *Clinical Immunology*, *151*(1), 66–77.

Young, T. L., Matsuda, T., & Cepko, C. L. (2005). The Noncoding RNA Taurine Upregulated Gene 1 Is Required for Differentiation of the Murine Retina. *Current Biology*, *15*(6), 501–512.

Zhang, H., Taylor, W. R., Joseph, G., Caracciolo, V., Gonzales, D. M., Sidell, N., … Kallen, C. B. (2013). mRNA-Binding Protein ZFP36 Is Expressed in Atherosclerotic Lesions and Reduces Inflammation in Aortic Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *33*(6), 1212–1220.

Zhang, X., Lian, Z., Padden, C., Gerstein, M. B., Rozowsky, J., Snyder, M., … Newburger, P. E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*, *113*(11), 2526–2534.

Zhang, Z.-Y., Zhao, X.-D., Jiang, L.-P., Liu, E.-M., Wang, M., Yu, J., … Yang, X.-Q. (2010). Clinical Characteristics and Molecular Analysis of 21 Chinese Children with Congenital Agammaglobulinemia: Clinical Characteristics and Molecular Analysis. *Scandinavian Journal of Immunology*, *72*(5), 454–459.

Zhou, L., Diefenbach, E., Crossett, B., Tran, S. L., Ng, T., Rizos, H., … Saksena, N. K. (2010). First evidence of overlaps between HIV-Associated Dementia (HAD) and non-viral neurodegenerative diseases: proteomic analysis of the frontal cortex

from HIV+ patients with and without dementia. *Molecular Neurodegeneration*, *5*(1), 27.

Zhu, Z., Liang, Z., Liany, H., Yang, C., Wen, L., Lin, Z., … Zhang, X. (2015). Discovery of a novel genetic susceptibility locus on X chromosome for systemic lupus erythematosus. *Arthritis Research & Therapy*, *17*(1).

Zieba, A., Sjöstedt, E., Olovsson, M., Fagerberg, L., Hallström, B. M., Oskarsson, L., … Ponten, F. (2015). The Human Endometrium-Specific Proteome Defined by Transcriptomics and Antibody-Based Profiling. *OMICS: A Journal of Integrative Biology*, *19*(11), 659–668.

Ziegler-Heitbrock, H. W. (2000). Definition of human blood monocytes. *Journal of Leukocyte Biology*, *67*(5), 603–606.

Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D. N., … Lutz, M. B. (2010). Nomenclature of monocytes and dendritic cells in blood. *Blood*, *116*(16), e74–e80.

Zimmerman, J. J. (Ed.). (2012). *Diseases of swine* (10th ed). Chichester, West Sussex: Wiley-Blackwell.

LIST OF PUBLICATIONS

Publications

1. Mirsafian, H., Manda, S. S., Mitchell, C. J., Sreenivasamurthy, S., Ripen, A. M., Mohamad, S. B., Merican, A. F., Pandey, A. (2016). Long non-coding RNA expression in primary human monocytes. *Genomics*, *108*(1), 37–45.

2. Mirsafian, H. Ripen, A. M., Manaharan, T. Mohamad, S. B., Merican, A. F. (2016). Towards a Reference Gene Catalogue of Human Primary Monocytes. *OMICS*, 20(11), 627–634.

Genomics

Volume 108, Issue 1, July 2016, Pages 37–45

Comprehensive functionality of genomic DNA

Long non-coding RNA expression in primary human monocytes

Hoda Mirsafian[a, 1], Srinivas Srikanth Manda[b, c, d, 1], Christopher J. Mitchell[c], Sreelakshmi Sreenivasamurthy[b, c], Adiratna Mat Ripen[e], Saharuddin Bin Mohamad[a, f], Amir Feisal Merican[a, f], Akhilesh Pandey[b, c, g, h]

Show more

http://dx.doi.org/10.1016/j.ygeno.2016.01.002

Highlights

- Deep RNA-sequencing of primary human monocytes
- Identification of ~ 8000 long non-coding RNAs (lncRNAs) across 15 different monocyte samples
- Identification of > 1000 potential novel lincRNAs
- A systematic pipeline to identify lincRNAs
- Validation of a subset of novel lncRNAs across various hematopoietic cells

Abstract

Long non-coding RNAs (lncRNAs) have been shown to possess a wide range of functions in both cellular and developmental processes including cancers. Although some of the lncRNAs have been implicated in the regulation of the immune response, the exact function of the large majority of lncRNAs still remains unknown. In this study, we characterized the lncRNAs in human primary monocytes, an essential component of the innate immune system. We performed RNA sequencing of monocytes from four individuals and combined our data with eleven other publicly available datasets. Our analysis led to identification of ~ 8000 lncRNAs of which > 1000 have not been previously reported in monocytes. PCR-based validation of a subset of the identified novel long intergenic noncoding RNAs (lincRNAs) revealed distinct expression patterns. Our study provides a landscape of lncRNAs in monocytes, which could facilitate future experimental studies to characterize the functions of these molecules in the innate immune system.

Keywords

Monocytes; RNA-Seq; LncRNAs; LincRNAs

OMICS A Journal of Integrative Biology
Volume 20, Number 11, 2016

DOI: 10.1089/omi.2016.0124

Toward a Reference Gene Catalog of Human Primary Monocytes

Hoda Mirsafian,[1] Adiratna Mat Ripen,[2] Thamilvaani Manaharan,[3] Saharuddin Bin Mohamad,[1,3] and Amir Feisal Merican[1,3]

Abstract

Transcriptome analyses based on high-throughput RNA sequencing (RNA-Seq) provide powerful and quantitative characterization of cell types and in-depth understanding of biological systems in health and disease. In this study, we present a comprehensive transcriptome profile of human primary monocytes, a crucial component of the innate immune system. We performed deep RNA-Seq of monocytes from six healthy subjects and integrated our data with 10 other publicly available RNA-Seq datasets of human monocytes. A total of 1.9 billion reads were generated, which allowed us to capture most of the genes transcribed in human monocytes, including 11,994 protein-coding genes, 5558 noncoding genes (including long noncoding RNAs, precursor miRNAs, and others), 2819 pseudogenes, and 7034 putative novel transcripts. In addition, we profiled the expression pattern of 1155 transcription factors (TFs) in human monocytes, which are the main molecules in controlling the gene transcription. An interaction network was constructed among the top expressed TFs and their targeted genes, which revealed the potential key regulatory genes in biological function of human monocytes. The gene catalog of human primary monocytes provided in this study offers significant promise and future potential clinical applications in the fields of precision medicine, systems diagnostics, immunogenomics, and the development of innovative biomarkers and therapeutic monitoring strategies.

Keywords: personalized medicine, multi-omics, transcriptome, RNA-Seq, gene catalogue, postgenomics

Introduction

MONOCYTES ARE ESSENTIAL CELLS of the innate immune system. They play important roles in the initiation and declaration of inflammation, generally through release of inflammatory cytokines, reactive oxygen species during phagocytosis, and activation of the adaptive immune system (Ziegler-Heitbrock, 2010). Monocytes initiate from a common myeloid progenitor cell in the bone marrow and circulate in the blood vessels for short times, and during inflammatory conditions, they move into peripheral tissues and differentiate into macrophages and dendritic cells (Saha and Geissmann, 2011). Monocytes are heterogeneous and are divided into three groups based on their expression of CD14 and CD16 receptor markers; classical ($CD14^{++}CD16^{-}$), intermediate ($CD14^{++}CD16^{+}$), and nonclassical ($CD14^{-}CD16^{++}$). Differences between monocyte subsets associate with differences in cytokine production, antigen presentation, and antigen uptake (Ziegler-Heitbrock, 2010). The classical monocytes account for 90–95% of human blood monocytes. Their major function is phagocytosis and they exhibit high peroxidase activity and produce high levels of *IL-10* and low levels of *TNF-α* in response to lipopolysaccharides (Yang et al., 2014).

Transcriptome study is important for understanding the genome functional elements, the molecular components of cells/tissues, and development of diseases. Previously, microarray was the commonly used method for transcriptome analysis; however, recently high-throughput RNA sequencing (RNA-Seq) has become a powerful alternative approach for transcriptome studies. RNA-Seq is able to qualitatively and quantitatively explore any RNA type, including messenger RNAs (mRNAs), long noncoding RNAs (lncRNAs), microRNAs (miRNAs), and small interfering RNAs (siRNAs), as well as novel transcripts (Dong and Chen, 2013). Recent studies have applied RNA-Seq technology for transcriptome profiling of several tissues and cell types such as endometrium

[1]Faculty of Science, Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia.
[2]Allergy and Immunology Research Centre, Institute for Medical Research, Jalan Pahang, Kuala Lumpur, Malaysia.
[3]Centre of Research for Computational Sciences and Informatics in Biology, Bioindustry, Environment, Agriculture and Healthcare (CRYSTAL), University of Malaya, Kuala Lumpur, Malaysia.

APPENDIX A

List of immune-related protein-coding genes expressed in human primary monocytes

Gene	Chr	FPKM
ACE	17	0.586237
ACIN1	14	52.66772
ADAM10	15	211.5107
ADAM17	2	28.82037
ADAM9	8	29.3759
ADD1	4	231.7057
ADSS	1	36.34807
ADSSL1	14	1.273058
AIMP1	4	32.07983
AKIRIN2	6	56.28375
AKT1	14	0.227527
ALAS2	X	0.203233
ANXA11	10	85.02763
AP3B1	5	33.45878
AP3D1	19	61.79758
APC	5	41.57535
AQP9	15	60.5002
ARHGDIB	12	1.006233
B2M	15	10823.64
BAX	19	46.1002
BCAP31	X	74.0168
BCL11B	14	0.30344
BCL6	3	0.151649
BLM	15	1.438325
BLNK	10	0.316407
BMI1	10	58.01472
BMPR1A	10	0.656758
BNIP3	10	1.650877
BNIP3L	8	137.7405
BRCA2	13	0.515149
BST1	4	136.3111
BST2	19	68.8477
BTLA	3	0.798848
C1QBP	17	36.565
C3AR1	12	65.19717
C5AR1	19	186.7257
C8G	9	0.200499
CACNB3	12	1.583641
CACNB4	2	1.19588
CADM1	11	0.801052
CALCOCO2	17	163.02
CALR	19	260.7467
CARD9	9	32.1991
CBFB	16	62.17693
CCL1	17	0.126719
CCL23	17	0.121361
CCL27	9	0.235336
CCL28	5	0.152019
CCL3L3	17	1.1828
CCL4	17	1.638007
CCL4L1	17	1.680079

Gene	Chr	FPKM
CCL5	17	97.68862
CCL8	17	0.306442
CCNB2	15	0.400625
CCND3	6	155.8147
CCNL2	1	65.29425
CCR1	3	75.77318
CCR2	3	149.0918
CCR4	3	0.953338
CCR7	17	0.920656
CCR9	3	0.191941
CD14	5	116.5953
CD164	6	209.3538
CD19	16	0.454884
CD1A	1	1.633547
CD1B	1	0.708462
CD1D	1	48.13053
CD27	12	34.7307
CD276	15	0.135606
CD28	2	0.823083
CD300A	17	107.5818
CD300C	17	99.92975
CD300E	17	389.4238
CD300LD	17	0.210319
CD300LF	17	0.500606
CD302	2	169.7947
CD34	1	1.297242
CD3D	11	0.845712
CD3E	11	1.592107
CD3G	11	0.860541
CD40LG	X	0.37021
CD46	1	166.811
CD48	1	285.2817
CD5	11	1.285111
CD55	1	105.0906
CD7	17	244.6638
CD74	5	68.44027
CD80	3	0.315509
CD86	3	113.8508
CD8A	2	0.885601
CD8B	2	0.421467
CD96	3	1.398303
CD97	19	382.156
CDC42	1	350.567
CEBPA	19	40.16788
CEBPB	20	69.23545
CFD	19	76.45935
CFH	1	0.560137
CFP	X	1.294826
CHIT1	1	1.185676
CHUK	10	34.15168
CIITA	16	99.6059

Gene	Chr	FPKM
CKLF	16	49.62243
CLEC4A	12	61.92043
CLEC4C	12	1.435893
CLEC4E	12	44.31265
CLEC7A	12	178.9765
CLNK	4	0.196761
CLU	8	283.0788
CMKLR1	12	0.358101
CNPY3	6	109.1518
CNR2	1	0.200022
COL4A3BP	5	53.84153
COLEC12	18	0.328715
CORO1A	16	377.4837
CR2	1	0.255068
CRIP2	14	0.288153
CRKL	22	35.14827
CSF1	1	1.192116
CTLA4	2	0.139414
CTNNB1	3	85.87722
CTNNBL1	20	30.54868
CTSC	11	0.118832
CTSS	1	621.283
CXCL1	4	0.79747
CXCL11	4	1.157181
CXCL16	17	72.5029
CXCL2	4	1.069155
CXCR4	2	59.73027
CYBA	16	294.8403
CYBB	X	789.7948
DDOST	1	112.085
DEFA1	8	510.5883
DEFB1	8	0.198113
DLL1	6	0.385071
DNASE2	19	32.22177
DOCK2	5	247.5572
DPP4	2	0.497605
DYRK3	1	0.11121
EBI3	19	0.270853
EDA	X	0.153972
EDN1	6	0.221328
EDNRB	13	0.296207
ELF4	X	64.81648
ENPP3	6	0.318281
EOMES	3	0.392511
ERAP1	5	89.32302
ERAP2	5	40.64854
ERCC1	19	30.55978
EXO1	1	0.314896
F12	5	0.682192
FANCC	9	1.231292
FAS	10	33.95655

Gene	Chr	FPKM
FCER1G	1	388.493
FCGR1A	1	43.49742
FCGRT	19	155.8513
FCN1	9	106.5991
FKBP1A	20	262.7465
FOXP3	X	0.16839
FTH1	11	2006.808
FYB	5	0.234143
FYN	6	234.6382
G6PD	X	75.52577
GALNT2	1	45.1099
GAPT	5	65.68303
GBP1	1	286.1252
GBP3	1	0.170714
GBP4	1	61.61317
GBP5	1	119.0342
GCH1	14	0.596471
GNL1	6	145.0643
GPI	19	76.11063
GPSM3	6	0.120248
GZMA	5	1.683981
HCLS1	3	338.882
HDAC9	7	43.22592
HELLS	10	0.277419
HLA-A	6	1056.711
HLA-B	6	3004.093
HLA-DOB	6	82.54188
HLA-DPA1	6	455.7492
HLA-DPB1	6	254.6522
HLA-DQA1	6	285.4907
HLA-DQB1	6	714.3683
HLA-DRA	6	1807.687
HLA-E	6	1518.303
HOXA3	7	0.332782
HOXB3	17	0.747608
HSH2D	19	152.5073
HSPD1	2	62.50355
ICAM1	19	45.06948
ID2	2	73.82327
IFI16	1	180.9417
IFI30	19	1055.32
IFI35	17	46.81407
IFI6	1	127.9699
IFIH1	2	41.64542
IFITM2	11	228.9053
IFNAR1	21	45.51988
IGBP1	X	39.9179
IGJ	4	115.6557
IGLL1	22	0.156518
IGSF6	16	103.2596
IK	5	86.16223

Gene	Chr	FPKM
IKBKB	8	46.10455
IKBKE	1	77.8543
IKZF1	7	136.5663
IL10	1	0.314554
IL12A	3	0.265218
IL17C	16	0.328636
IL18R1	2	1.230312
IL1RN	2	51.90648
IL20RB	3	37.77612
IL23A	12	0.861382
IL27	16	1.07094
IL27RA	19	42.26502
IL2RA	10	0.326374
IL2RG	X	53.39162
IL31RA	5	1.554525
IL4	5	0.146723
IL4R	16	53.72223
IL6R	1	190.9653
IL7	8	1.568548
ILF2	1	42.06493
IMPDH1	7	151.6738
IMPDH2	3	67.26107
INHBA	7	0.33526
INPP5D	2	123.4015
INPPL1	11	147.152
IRAK1	X	83.17877
IRAK2	3	1.615007
IRAK3	12	75.02385
IRF8	16	150.5813
ITGA4	2	258.2397
ITGAL	16	307.6928
ITGAM	16	368.5755
ITGB1	10	281.329
ITIH1	3	1.177581
JAK2	9	74.38647
KIRREL3	11	0.124425
KIT	4	0.633203
KLF1	19	0.297027
KLF6	10	261.9753
KYNU	2	42.85283
LAIR1	19	109.4923
LAT2	7	124.1527
LAX1	1	0.645321
LILRA6	19	524.0822
LMO2	11	74.20095
LST1	6	186.4355
LTB4R	14	46.9647
LTF	3	107.2816
LY86	6	69.32177
LY9	1	1.245665
LYN	8	393.3378

Gene	Chr	FPKM
LYST	1	81.64348
MALT1	18	30.64478
MAP4K2	11	44.24568
MAPK1	22	176.6083
MASP2	1	1.028553
MAVS	20	48.5458
MBP	18	104.027
MECOM	3	0.204043
MINK1	17	55.35597
MR1	1	38.0413
MSN	X	517.3898
MYLPF	16	0.180035
MYO1F	19	531.6478
NBN	8	37.73372
NCF1	7	159.8877
NCF2	1	493.6228
NCF4	22	58.07342
NCK2	2	84.46933
NCOA6	20	30.2147
NCR1	19	0.366397
NCR3	6	0.217971
NDRG1	8	77.08302
NFAM1	22	0.335407
NFKB2	10	33.40452
NFKBIA	14	1.061119
NLRC3	16	1.103987
NLRP3	1	48.95895
NOTCH1	9	0.275135
NOTCH2	1	1.356439
NTRK1	1	0.133577
OAS1	12	496.4908
OAS2	12	258.7711
OSTM1	6	69.56683
P2RX7	12	50.26382
PAX5	9	0.770402
PDCD1LG2	9	0.779974
PDGFB	22	0.245198
PF4	4	646.955
PGLYRP2	19	0.233394
PIK3R1	5	31.49577
PLEK	2	223.8522
PML	15	38.11943
PNP	14	37.09548
PODXL	7	0.556128
POU2AF1	11	1.190855
POU2F2	19	98.98807
PPARG	3	1.045849
PRDX1	1	74.76943
PRDX3	10	79.63792
PRELID1	5	125.0012
PREX1	20	0.190726

Gene	Chr	FPKM
PRG2	11	65.45282
PRKCD	3	112.9832
PROCR	20	0.483825
PSEN1	14	84.15998
PSMB8	6	223.1022
PSME1	14	157.3093
PTGER4	5	78.98257
PTPRC	1	918.8593
PXDN	2	0.290238
RASGRP4	19	71.24728
RB1	13	37.84023
RELA	11	1.010558
RNF125	18	68.75197
RNF19B	1	47.51273
ROCK1	18	64.91103
ROGDI	16	49.12883
RPA1	17	40.5687
RPL22	1	174.9698
RPS14	5	456.443
RPS19	19	284.689
RSAD2	2	41.26603
RUNX1	21	45.68727
S100A9	1	2689.97
SAMHD1	20	415.1952
SBNO2	19	0.149138
SEMA4D	9	98.89527
SERPING1	11	46.95954
SFTPD	10	1.280427
SH2B3	12	136.0938
SH2D1A	X	0.533553
SIT1	9	0.343806
SKAP1	17	0.691447
SKAP2	7	155.3588
SLAMF1	1	0.461652
SLAMF7	1	48.2271
SMAD3	15	40.94818
SMAD6	15	0.459514
SNCA	4	38.6794
SNRK	3	29.35745
SOD2	6	1.04394
SP1	12	62.133
SPG21	15	84.62165
SPN	16	0.350887
SPTA1	1	0.132325
SQSTM1	5	154.3677
ST6GAL1	3	55.08257
STAT5A	17	115.7063
STAT5B	17	84.9628
STXBP3	1	46.45952
SYK	9	158.2713
TACC3	4	45.00075

Gene	Chr	FPKM
TAPBP	6	0.320256
TAZ	X	33.39383
TCEA1	8	38.35138
TCF12	15	50.78518
TCF7	5	1.444199
TFE3	X	1.215815
TFEB	6	46.43997
TGFB1	19	149.683
TGFB2	1	0.414794
TGFBR2	3	76.26608
TGFBR3	1	0.723633
THEMIS	6	0.335669
TICAM2	5	76.00632
TIMP1	X	66.5022
TLR1	4	81.60893
TLR2	4	79.28142
TLR3	4	2.183876
TLR4	9	128.7649
TLR5	1	3.991353
TLR6	4	4.675892
TLR7	X	30.4726
TLR8	X	104.9199
TMEM173	5	86.77185
TMX1	14	50.0573
TNFRSF11A	18	0.202917
TNFRSF14	1	80.44225
TNFRSF17	16	0.923728
TNFRSF1A	12	194.1407
TNFSF10	3	253.4425
TNFSF12	17	216.6287
TNFSF13B	13	132.6227
TNFSF15	9	0.519099
TNFSF8	9	34.43168
TNFSF9	19	0.126224
TOLLIP	11	46.54945
TP53	17	31.39832
TRAT1	3	0.426688
TREM1	6	110.388
TREML1	6	50.82112
TRIM10	6	0.509626
TRIM22	11	0.222017
TRPC4AP	20	54.3165
TRPM4	19	1.110743
WAS	X	178.7477
TTC7A	2	125.636
TUBB	6	61.1152
UBE2N	12	62.8757
UNC13D	17	186.642
VAMP7	X	50.93507
VAV3	1	58.35602
VNN1	6	190.8205

Gene	Chr	FPKM
XBP1	22	63.75543
XRCC5	2	192.6528
XRCC6	22	108.4456
YTHDF2	1	34.37937
YWHAZ	8	619.1642
ZBTB32	19	0.106388
ZEB1	10	0.75667
CNP	17	34.12888
HMGCR	5	34.45682
APEX1	14	76.57758
ACTR3	2	365.1195
ATG12	5	33.53413
ATP5G3	2	47.45875
ABCC1	16	52.87673
ABCF1	6	29.85305
ATP6V0E1	5	137.9742
ATP6AP1	X	113.806
ATP1A1	1	128.0357
ATRIP	3	30.92405
BCL2L1	20	34.4299
BTG2	1	59.5325
BBS1	11	34.726
CD163	12	129.3708
CD36	7	492.8698
CD4	12	216.1165
CD44	11	528.9057
CD9	12	32.52337
CHEK1	11	45.07867
CMTM3	16	89.34072
CMTM6	3	133.2207
CMTM7	3	48.70833
COPS3	17	40.47955
CREBBP	16	54.01983
CITED2	6	76.35672
DHX9	1	57.66505
DEK	6	160.0825
DERL1	8	44.82355
DNAJA1	9	67.0653
DNAJB1	19	40.88982
DNAJB6	7	99.85037
DNAJC25	9	29.9992
DNAJC4	11	29.98965
EP300	22	35.80338
ELF3	1	81.81857
ELK3	12	31.13653
EDEM1	3	29.89883
ERO1L	14	48.20815
FBXO18	10	41.26337
FOSB	19	37.80448
FIG4	6	36.64403
FANCG	9	67.5906

Gene	Chr	FPKM
GPS1	17	39.05905
GPS2	17	55.28693
GM2A	5	40.10177
GNAS	20	727.5913
HERPUD2	7	39.19587
HPS1	10	82.19472
ISG15	1	43.25526
JAK1	1	168.56
KAT2B	3	31.62212
MKNK1	1	65.91037
MEFV	16	105.4973
MDFIC	7	40.90885
ASAH1	8	469.2023
NDST1	5	37.87162
NMI	2	45.07793
NDUFS2	1	57.81637
NLRC4	2	31.13222
OGT	X	93.25597
PDLIM1	10	33.47618
PRPF19	11	49.65622
PRPF8	17	78.47885
PTK2B	8	124.279
PXK	3	31.25363
PARK7	1	88.65283
RAB6A	11	106.5153
RAD23A	19	48.00525
RAD23B	9	71.09055
RBM3	X	253.23
RBM14	11	79.10048
ARHGEF6	X	60.69577
RECQL	12	38.16307
S100A12	1	203.3525
S100A8	1	1793.37
SH2D3C	9	50.05318
SMAD2	18	70.27083
SUMO1	2	60.76765
SLK	10	32.41295
TAF9	5	37.17985
TAOK3	12	72.10513
TIAL1	10	41.82792
TNIP1	5	118.8768
TYROBP	19	511.5823
UBXN4	2	62.23812
UPF1	19	67.38048
UVRAG	11	37.88857
ABHD2	15	81.44942
ACTB	7	150.4617
ACTN4	19	236.8332
ATF4	22	132.7715
ATF6B	6	138.9792
ACSL1	4	116.7291

Gene	Chr	FPKM
ACSL4	X	74.61392
AOAH	7	287.0133
APPL1	3	31.67843
AP3S1	5	50.23977
ADCY7	16	251.3398
ADIPOR1	1	134.297
ADIPOR2	12	32.38533
ADRBK1	11	289.2878
AGER	6	157.6393
ADH5	4	29.56507
AIF1	6	399.4167
AES	19	127.3908
APP	21	203.4877
APLP2	11	908.9762
ANXA1	9	462.4992
ANXA5	4	333.1552
ANXA7	10	99.73785
APOL2	22	138.0294
AATF	17	32.24427
APAF1	12	57.664
ALOX5	10	239.1443
ARRB2	17	353.8463
AHR	7	76.16878
AIP	11	32.52618
ARNT	1	43.41973
AHRR	5	39.37717
ATXN3	14	31.16077
AMFR	16	49.04143
BIRC2	11	36.81458
BSG	19	131.7467
BECN1	17	70.38152
BRE	2	53.89277
BCKDHA	19	56.11092
CREB1	2	66.09495
CIB1	15	31.43682
CAMK2D	4	29.80032
CALM3	19	34.65784
CAMTA2	17	49.85185
CAPN2	1	177.5862
CASC3	17	46.27953
CA2	8	40.76902
CES1	16	30.20958
CPT1A	11	61.38092
CAT	11	130.741
COMT	22	73.95175
CTSD	11	552.5767
CIDEB	14	39.30282
CCM2	7	35.19035
CX3CR1	3	169.3396
CLIC1	6	269.6032
CHRNA10	11	32.99862

Gene	Chr	FPKM
CHMP1A	16	52.21115
CHD1L	1	32.26658
LUC7L3	17	66.93565
CCDC47	17	205.2305
CSF3R	1	42.98435
CUL4A	13	43.1827
CDKN1A	6	51.62058
CDKN2D	19	75.34607
CYP1B1	2	68.64563
CYBRD1	2	39.25967
CYB5R4	6	72.56915
COX4I1	16	152.4237
DDB1	11	68.67812
DAXX	6	258.9572
DAD1	14	39.69578
DGKD	2	58.03067
DGKZ	11	62.18562
DICER1	14	50.33105
DUSP1	5	95.09155
DYNLRB1	20	78.39848
DYSF	2	58.53718
ERP44	9	56.32113
ENSA	1	66.04047
EIF2S1	14	59.92718
EIF2AK1	7	115.9058
EIF2AK2	2	46.23675
EIF2B1	12	38.9865
EIF4EBP2	10	42.31338
XPO1	2	60.11527
FAM129A	1	36.56267
FNTA	8	82.91418
FPR1	19	202.397
FPR2	19	51.52087
FXN	9	30.51133
GSDMD	8	43.76448
GSN	9	120.2558
GHRL	3	34.94387
GAA	17	188.3878
GLUL	1	458.917
GCLC	6	30.27322
GPX1	3	337.4627
GPX4	19	74.25607
GSK3B	3	32.7237
GP1BB	22	65.23445
GADD45B	19	34.01273
GNA12	7	34.7359
GNA13	17	74.3519
GNAI2	3	428.7853
GNB1	1	351.5892
GNB2	7	194.3038
GNB4	3	48.14503

Gene	Chr	FPKM
GNG2	14	80.20102
GNAQ	9	104.4432
HSPE1	2	42.77522
HSPA1A	6	60.85662
HSPA4	5	42.444
HSPA9	5	72.49697
HSP90AA1	14	214.8542
HSP90AB1	6	131.3772
HSP90B1	12	167.7015
HSF1	8	36.67937
HMOX1	22	92.10712
HCK	20	374.5395
HEXB	5	110.2758
HMGB1	13	238.3612
HMGB2	4	98.85848
HIATL1	9	85.22893
HIST1H2BJ	6	37.03945
HDAC2	6	39.81938
HDAC3	5	42.35218
HDAC6	X	33.53423
HIPK1	1	66.48825
HERPUD1	16	81.42763
HVCN1	12	62.3624
HADHA	2	102.9028
RALBP1	18	41.89658
HSPA5	9	88.13192
HIF1A	14	78.8005
HYOU1	11	35.26807
IVNS1ABP	1	51.7525
ING4	12	32.71257
INSIG1	7	44.62145
IRS2	13	42.21745
INTS3	1	58.15033
IFNGR1	6	233.1308
IFNGR2	21	138.4817
IFITM1	11	168.8549
IRF9	14	164.3558
IFI44	1	272.2635
IL6ST	5	32.12787
IDH1	2	39.00012
IDH3B	20	47.43872
JUNB	19	248.6805
LDHA	11	238.0508
LGALS3BP	17	34.04657
LEPR	1	65.37975
LTA4H	12	253.4737
LIPA	10	226.328
LYZ	12	1621.26
MIA3	1	43.6444
MGEA5	10	53.7494
MAT2A	2	83.89145

Gene	Chr	FPKM
MBD2	18	83.58965
MGST1	12	35.38062
MCM7	7	28.84433
MAPK14	6	89.1595
MAPK8IP3	16	44.30298
MAP2K1	15	93.89245
MAP2K3	17	103.6203
MAP3K1	5	61.71008
MAP3K11	11	51.48978
MAP3K2	2	37.65313
MAP4K4	2	93.94955
MAPKAPK3	3	201.024
MGLL	3	45.1893
MORF4L2	X	58.95868
MORF4L1	15	197.5692
MCL1	1	49.49312
MNDA	1	525.7072
MLF2	12	64.65938
MYOF	10	48.15763
LUZP6	7	245.9763
MX1	21	191.4697
MX2	21	140.4966
NINJ1	9	85.37188
NONO	X	184.2647
NME1	17	79.13287
NFKB1	4	48.90215
NFKBIZ	3	144.7393
NR3C1	5	44.43792
NPM1	5	212.6658
OR2AE1	7	30.9206
OR2W3	1	33.05632
OR56B1	11	382.2078
OPA1	3	54.12775
OS9	12	237.891
OXR1	8	44.52172
OXSR1	3	30.27313
PPT1	1	462.3202
PXN	12	110.7972
PRDX5	11	42.8049
PRDX6	1	79.177
PTEN	10	139.7055
PCYT1A	3	52.996
PFKL	21	73.16207
PIK3CB	3	38.3359
PLCB2	15	191.5337
PLSCR1	3	145.3071
GART	21	31.44293
PAFAH1B1	17	71.96062
PARP1	1	37.34928
PARP4	13	54.74207
PABPC4	1	90.85103

Gene	Chr	FPKM
POLG	15	33.83278
KCNQ1	11	71.2777
PRNP	20	70.51738
PTGS1	9	123.1525
PKN1	19	139.9287
PRKAR2B	7	104.7988
PPP1CB	2	108.1075
PPP1R15A	19	58.04253
PPP1R15B	1	35.47
PPP1R9B	17	83.78312
PPP2CB	8	33.72238
PPP2R1A	19	76.91322
PPP3CA	4	49.71435
PTPN11	12	33.20377
PTPN2	18	42.52177
PTPN6	12	341.0813
PTPRA	20	91.83222
PRKRIR	11	41.62857
PLP2	X	104.7461
RHOQ	2	112.6735
RAC1	7	130.0748
RAC2	22	248.3962
RETN	19	36.80824
RTN3	11	163.979
RP2	X	51.12178
RARA	17	105.8238
RXRA	9	120.9746
RNASE2	14	28.84743
RNASE6	14	56.73822
RPS3	11	486.65
SCFD1	14	39.3273
SEMA4A	1	105.4654
STK25	2	78.43078
SRRT	7	31.37085
SGK1	6	115.9845
SRP14	15	103.0344
SRP72	4	80.00087
STAT1	2	321.2763
STAT3	17	117.402
STAT6	12	238.7027
ATM	11	31.91877
EIF1	17	342.3603
SLC10A3	X	48.6038
SLC8A1	2	47.92808
SORT1	1	58.82863
SNX27	1	29.85025
SHFM1	7	46.1831
STAB1	3	94.69805
SNN	16	125.9799
SREBF2	22	40.44452
SERP1	3	97.0072

Gene	Chr	FPKM
STIP1	11	48.84743
SDF4	1	43.9627
SMC1A	X	53.26517
SMC5	9	32.77963
SSRP1	11	36.35008
SUPT16H	14	33.13037
TERF2IP	16	34.65883
TTC3	21	36.83672
TXN2	22	36.4841
TOR1A	9	35.08862
TOR1B	9	33.78812
TLK1	2	36.48815
TFEC	7	45.12493
TGFBI	5	271.8655
TRIB1	8	39.11398
TSC2	16	30.2318
MC1R	16	94.89745
UQCRC1	3	131.2945
UIMC1	5	31.1568
UBE2A	X	41.53348
UBE2B	5	49.77058
UNC119	17	42.61293
UCP2	11	353.0555
USF2	19	97.79038
AKT2	19	96.67512
ERBB3	12	63.1545
FOS	14	1042.202
MAFG	17	35.66073
SRC	20	38.33302
VPS4B	18	54.21067
VARS	6	34.75263
VEZF1	17	41.17277
VCAN	5	935.8262
VDAC1	5	56.56392
VDAC3	8	71.82875
VHL	3	30.8044
WAPAL	10	44.44218
XPC	3	52.56275
ZC3HAV1	7	30.78147

APPENDIX B

List of predicted PFAM domains for novel transcripts identified in human primary monocytes

Transcript_ID	Chr	Pfam Domain
TCONS_00000054	1	PF08039.9
TCONS_00000786	1	PF16625.3
TCONS_00000786	1	PF16503.3
TCONS_00008956	1	PF13567.4
TCONS_00015605	1	PF15788.3
TCONS_00016028	1	PF15788.3
TCONS_00016029	1	PF15788.3
TCONS_00016030	1	PF15788.3
TCONS_00029165	1	PF02994.12
TCONS_00029165	1	PF11250.6
TCONS_00029165	1	PF03372.21
TCONS_00029247	1	PF16637.3
TCONS_00032785	1	PF15788.3
TCONS_00032786	1	PF15788.3
TCONS_00033624	1	PF05356.9
TCONS_00034111	1	PF10003.7
TCONS_00034167	1	PF08333.9
TCONS_00034488	1	PF07780.10
TCONS_00034488	1	PF16100.3
TCONS_00034488	1	PF05812.10
TCONS_00034513	1	PF08333.9
TCONS_00034672	1	PF15742.3
TCONS_00034752	1	PF01525.14
TCONS_00034754	1	PF01525.14
TCONS_00034780	1	PF02093.14
TCONS_00035303	1	PF03372.21
TCONS_00035303	1	PF07780.10
TCONS_00041385	10	PF08513.9
TCONS_00041950	10	PF01469.16
TCONS_00041950	10	PF15320.4
TCONS_00041950	10	PF08666.10
TCONS_00041950	10	PF15140.4
TCONS_00041950	10	PF03811.11
TCONS_00041950	10	PF12118.6
TCONS_00041952	10	PF10584.7
TCONS_00047830	10	PF13765.4
TCONS_00048164	10	PF03105.17
TCONS_00048164	10	PF13908.4
TCONS_00048188	10	PF01469.16
TCONS_00048188	10	PF15320.4
TCONS_00048188	10	PF08666.10
TCONS_00048188	10	PF15140.4
TCONS_00048188	10	PF03811.11
TCONS_00048188	10	PF12118.6
TCONS_00048302	10	PF16360.3
TCONS_00048339	10	PF16975.3
TCONS_00048358	10	PF08865.9
TCONS_00048448	10	PF04513.10
TCONS_00048448	10	PF04928.15
TCONS_00048448	10	PF00175.19
TCONS_00048448	10	PF04049.11
TCONS_00048448	10	PF04156.12
TCONS_00048448	10	PF16360.3
TCONS_00048449	10	PF02994.12
TCONS_00048449	10	PF08614.9
TCONS_00048449	10	PF02403.20
TCONS_00048449	10	PF10205.7
TCONS_00048449	10	PF04136.13
TCONS_00048449	10	PF13851.4
TCONS_00048449	10	PF09177.9
TCONS_00048449	10	PF10337.7
TCONS_00048449	10	PF05478.9
TCONS_00048449	10	PF06160.10
TCONS_00048449	10	PF14425.4
TCONS_00048449	10	PF07439.9
TCONS_00048449	10	PF10186.7
TCONS_00048449	10	PF06320.11
TCONS_00048449	10	PF07889.10
TCONS_00048449	10	PF10046.7
TCONS_00048449	10	PF07200.11
TCONS_00048449	10	PF07106.11
TCONS_00048449	10	PF07055.10
TCONS_00048449	10	PF15278.4
TCONS_00048449	10	PF14712.4
TCONS_00048449	10	PF01166.16
TCONS_00048449	10	PF09969.7
TCONS_00048449	10	PF04156.12
TCONS_00048449	10	PF15136.4
TCONS_00048449	10	PF04012.10
TCONS_00048449	10	PF00992.18
TCONS_00048449	10	PF04111.10
TCONS_00048449	10	PF01519.14
TCONS_00048449	10	PF06156.11
TCONS_00048449	10	PF11932.6
TCONS_00048449	10	PF14077.4
TCONS_00048449	10	PF10779.7
TCONS_00048449	10	PF07303.11
TCONS_00048449	10	PF04977.13
TCONS_00048449	10	PF11214.6
TCONS_00048449	10	PF07851.11
TCONS_00048449	10	PF02320.14
TCONS_00048449	10	PF08333.9
TCONS_00048449	10	PF15386.4
TCONS_00048570	10	PF14386.4
TCONS_00048622	10	PF01388.19
TCONS_00048690	10	PF16650.3
TCONS_00048706	10	PF09900.7
TCONS_00048715	10	PF00574.21
TCONS_00049125	10	PF16118.3
TCONS_00049125	10	PF03105.17
TCONS_00049125	10	PF07423.9
TCONS_00049125	10	PF06024.10

Transcript_ID	Chr	Pfam Domain
TCONS_00049125	10	PF06024.10
TCONS_00049125	10	PF06024.10
TCONS_00049125	10	PF14575.4
TCONS_00049125	10	PF02790.13
TCONS_00049125	10	PF02487.15
TCONS_00049125	10	PF12326.6
TCONS_00049125	10	PF03348.13
TCONS_00049125	10	PF13314.4
TCONS_00049136	10	PF05297.9
TCONS_00064850	11	PF11057.6
TCONS_00071971	11	PF09843.7
TCONS_00072253	11	PF05901.9
TCONS_00072253	11	PF06906.9
TCONS_00072389	11	PF14529.4
TCONS_00072389	11	PF03372.21
TCONS_00073760	12	PF15040.4
TCONS_00094590	12	PF13439.4
TCONS_00095180	12	PF00077.18
TCONS_00095180	12	PF01585.21
TCONS_00095180	12	PF13975.4
TCONS_00095302	12	PF10761.7
TCONS_00095402	12	PF14529.4
TCONS_00095402	12	PF03372.21
TCONS_00097455	13	PF01343.16
TCONS_00098480	13	PF03732.15
TCONS_00098480	13	PF07995.9
TCONS_00101360	13	PF10939.6
TCONS_00101458	13	PF14529.4
TCONS_00101458	13	PF03372.21
TCONS_00101458	13	PF16455.3
TCONS_00101458	13	PF00010.24
TCONS_00101458	13	PF00175.19
TCONS_00101664	13	PF04433.15
TCONS_00101664	13	PF15386.4
TCONS_00101767	13	PF05019.11
TCONS_00101770	13	PF13465.4
TCONS_00101781	13	PF06449.9
TCONS_00101781	13	PF16801.3
TCONS_00101781	13	PF15315.4
TCONS_00102129	14	PF10001.7
TCONS_00115820	14	PF08309.9
TCONS_00116080	14	PF05410.11
TCONS_00116690	15	PF06836.10
TCONS_00116690	15	PF01007.18
TCONS_00121586	15	PF13094.4
TCONS_00121586	15	PF08317.9
TCONS_00121586	15	PF15070.4
TCONS_00121586	15	PF15905.3
TCONS_00121586	15	PF04751.12
TCONS_00121586	15	PF16674.3
TCONS_00128942	15	PF15070.4

Transcript_ID	Chr	Pfam Domain
TCONS_00128942	15	PF13094.4
TCONS_00128942	15	PF08317.9
TCONS_00128942	15	PF04751.12
TCONS_00128942	15	PF09726.7
TCONS_00128942	15	PF06160.10
TCONS_00128942	15	PF11744.6
TCONS_00128942	15	PF01519.14
TCONS_00128942	15	PF15905.3
TCONS_00128942	15	PF15456.4
TCONS_00128942	15	PF16278.3
TCONS_00130078	15	PF13705.4
TCONS_00130192	15	PF08411.8
TCONS_00134860	16	PF08468.9
TCONS_00145542	16	PF14529.4
TCONS_00145542	16	PF03372.21
TCONS_00145542	16	PF14335.4
TCONS_00145542	16	PF16455.3
TCONS_00145542	16	PF15625.4
TCONS_00145542	16	PF16360.3
TCONS_00149796	16	PF12860.5
TCONS_00149872	16	PF14386.4
TCONS_00149872	16	PF14003.4
TCONS_00149959	16	PF14529.4
TCONS_00149959	16	PF03372.21
TCONS_00149962	16	PF14136.4
TCONS_00149971	16	PF01726.14
TCONS_00149971	16	PF14136.4
TCONS_00150019	16	PF12990.5
TCONS_00150141	16	PF05568.9
TCONS_00150180	17	PF02070.13
TCONS_00151173	17	PF15235.4
TCONS_00157679	17	PF06398.9
TCONS_00161468	17	PF07330.10
TCONS_00161519	17	PF12838.5
TCONS_00161750	17	PF17188.2
TCONS_00161750	17	PF11665.6
TCONS_00161750	17	PF06809.9
TCONS_00161777	17	PF15276.4
TCONS_00161781	17	PF16594.3
TCONS_00165454	17	PF00078.25
TCONS_00165454	17	PF17063.3
TCONS_00167610	17	PF07740.10
TCONS_00172433	17	PF12922.5
TCONS_00172433	17	PF13728.4
TCONS_00172433	17	PF13728.4
TCONS_00173024	17	PF04265.12
TCONS_00174282	17	PF07330.10
TCONS_00174580	17	PF16594.3
TCONS_00174650	17	PF12360.6
TCONS_00174650	17	PF12360.6
TCONS_00174728	17	PF12911.5

Transcript_ID	Chr	Pfam Domain
TCONS_00245470	21	PF00711.17
TCONS_00245470	21	PF05680.10
TCONS_00247174	21	PF12861.5
TCONS_00247713	21	PF10895.6
TCONS_00247734	21	PF02434.14
TCONS_00247737	21	PF10895.6
TCONS_00247911	21	PF14929.4
TCONS_00247953	21	PF14054.4
TCONS_00247953	21	PF05814.9
TCONS_00252046	22	PF15788.3
TCONS_00256664	22	PF05001.11
TCONS_00256664	22	PF17041.3
TCONS_00256693	22	PF13559.4
TCONS_00256693	22	PF15365.4
TCONS_00256761	22	PF08501.9
TCONS_00256847	22	PF07289.9
TCONS_00256899	22	PF11762.6
TCONS_00256899	22	PF11143.6
TCONS_00259955	3	PF10270.7
TCONS_00260332	3	PF10361.7
TCONS_00263333	3	PF03878.13
TCONS_00263333	3	PF09578.8
TCONS_00263333	3	PF06014.9
TCONS_00279677	3	PF15788.3
TCONS_00280016	3	PF05023.12
TCONS_00280250	3	PF08333.9
TCONS_00280322	3	PF00078.25
TCONS_00280322	3	PF01585.21
TCONS_00280374	3	PF02994.12
TCONS_00280375	3	PF14529.4
TCONS_00280375	3	PF03372.21
TCONS_00280527	3	PF15423.4
TCONS_00280557	3	PF11065.6
TCONS_00280557	3	PF01335.19
TCONS_00280557	3	PF13365.4
TCONS_00280557	3	PF13786.4
TCONS_00280557	3	PF04347.11
TCONS_00280557	3	PF03721.12
TCONS_00282854	4	PF14529.4
TCONS_00282854	4	PF00078.25
TCONS_00282854	4	PF14003.4
TCONS_00294917	4	PF11162.6
TCONS_00294965	4	PF00574.21
TCONS_00294965	4	PF10851.6
TCONS_00294965	4	PF15386.4
TCONS_00295098	4	PF14386.4
TCONS_00295103	4	PF02994.12
TCONS_00295103	4	PF01031.18
TCONS_00295103	4	PF16043.3
TCONS_00295103	4	PF10205.7
TCONS_00295103	4	PF00992.18

Transcript_ID	Chr	Pfam Domain
TCONS_00295103	4	PF05667.9
TCONS_00295103	4	PF06657.11
TCONS_00295103	4	PF01519.14
TCONS_00295103	4	PF07851.11
TCONS_00295103	4	PF05478.9
TCONS_00295103	4	PF11559.6
TCONS_00295103	4	PF01105.22
TCONS_00295103	4	PF04156.12
TCONS_00295142	4	PF08333.9
TCONS_00295249	4	PF14335.4
TCONS_00295249	4	PF14362.4
TCONS_00295249	4	PF02515.15
TCONS_00295249	4	PF16360.3
TCONS_00295468	4	PF02994.12
TCONS_00295468	4	PF13851.4
TCONS_00295468	4	PF10498.7
TCONS_00295468	4	PF10186.7
TCONS_00295468	4	PF04156.12
TCONS_00295468	4	PF02403.20
TCONS_00295468	4	PF08614.9
TCONS_00295468	4	PF07926.10
TCONS_00295468	4	PF03962.13
TCONS_00295468	4	PF09177.9
TCONS_00295468	4	PF04136.13
TCONS_00295468	4	PF00435.19
TCONS_00295468	4	PF10205.7
TCONS_00295468	4	PF15390.4
TCONS_00295468	4	PF11214.6
TCONS_00295468	4	PF04111.10
TCONS_00295468	4	PF07106.11
TCONS_00295468	4	PF07889.10
TCONS_00295468	4	PF10046.7
TCONS_00295468	4	PF08317.9
TCONS_00295468	4	PF07200.11
TCONS_00295468	4	PF04582.10
TCONS_00295468	4	PF14523.4
TCONS_00295468	4	PF15908.3
TCONS_00295468	4	PF07303.11
TCONS_00295468	4	PF04728.11
TCONS_00295468	4	PF05478.9
TCONS_00295468	4	PF06160.10
TCONS_00295468	4	PF10337.7
TCONS_00295468	4	PF06005.10
TCONS_00295468	4	PF08397.9
TCONS_00295468	4	PF11690.6
TCONS_00295468	4	PF03954.12
TCONS_00295468	4	PF00170.19
TCONS_00295468	4	PF10779.7
TCONS_00295468	4	PF10805.6
TCONS_00295468	4	PF10473.7
TCONS_00295468	4	PF08700.9

Transcript_ID	Chr	Pfam Domain
TCONS_00295468	4	PF01763.14
TCONS_00295468	4	PF04048.12
TCONS_00295468	4	PF11932.6
TCONS_00295468	4	PF05667.9
TCONS_00295468	4	PF16471.3
TCONS_00295468	4	PF03961.11
TCONS_00295468	4	PF06148.9
TCONS_00295468	4	PF01486.15
TCONS_00295468	4	PF01920.18
TCONS_00295468	4	PF03938.12
TCONS_00295468	4	PF06156.11
TCONS_00295468	4	PF01724.14
TCONS_00295468	4	PF05266.12
TCONS_00295468	4	PF15136.4
TCONS_00295468	4	PF09744.7
TCONS_00295468	4	PF08609.8
TCONS_00295468	4	PF04678.11
TCONS_00295468	4	PF14712.4
TCONS_00295468	4	PF04977.13
TCONS_00295468	4	PF12329.6
TCONS_00295468	4	PF13874.4
TCONS_00295468	4	PF08941.8
TCONS_00295468	4	PF01895.17
TCONS_00295468	4	PF01895.17
TCONS_00295468	4	PF07439.9
TCONS_00295468	4	PF04799.11
TCONS_00295468	4	PF02601.13
TCONS_00295468	4	PF06818.13
TCONS_00295468	4	PF12732.5
TCONS_00295468	4	PF04859.10
TCONS_00295468	4	PF14077.4
TCONS_00295468	4	PF05377.9
TCONS_00295468	4	PF07851.11
TCONS_00295468	4	PF01166.16
TCONS_00295468	4	PF01544.16
TCONS_00295468	4	PF10428.7
TCONS_00295468	4	PF14584.4
TCONS_00295468	4	PF07989.9
TCONS_00295468	4	PF02349.13
TCONS_00295468	4	PF08286.9
TCONS_00295468	4	PF08286.9
TCONS_00295468	4	PF09969.7
TCONS_00295468	4	PF01025.17
TCONS_00295468	4	PF05529.10
TCONS_00295468	4	PF05529.10
TCONS_00295468	4	PF02183.16
TCONS_00295468	4	PF13815.4
TCONS_00295468	4	PF13600.4
TCONS_00295468	4	PF08647.9
TCONS_00295468	4	PF05837.10
TCONS_00295468	4	PF01496.17

Transcript_ID	Chr	Pfam Domain
TCONS_00330829	6	PF10205.7
TCONS_00330829	6	PF04111.10
TCONS_00330829	6	PF16740.3
TCONS_00330829	6	PF06401.9
TCONS_00330829	6	PF16471.3
TCONS_00330829	6	PF14523.4
TCONS_00330829	6	PF06160.10
TCONS_00330829	6	PF07200.11
TCONS_00330829	6	PF08317.9
TCONS_00330829	6	PF03961.11
TCONS_00330829	6	PF08663.8
TCONS_00330829	6	PF07106.11
TCONS_00330829	6	PF10337.7
TCONS_00330829	6	PF04048.12
TCONS_00330829	6	PF00170.19
TCONS_00330829	6	PF02320.14
TCONS_00330829	6	PF04582.10
TCONS_00330829	6	PF10046.7
TCONS_00330829	6	PF11690.6
TCONS_00330829	6	PF03954.12
TCONS_00330829	6	PF05929.9
TCONS_00330829	6	PF00435.19
TCONS_00330829	6	PF01920.18
TCONS_00330829	6	PF10473.7
TCONS_00330829	6	PF10805.6
TCONS_00330829	6	PF08657.8
TCONS_00330829	6	PF06148.9
TCONS_00330829	6	PF07055.10
TCONS_00330829	6	PF13874.4
TCONS_00330829	6	PF08066.10
TCONS_00330829	6	PF04977.13
TCONS_00330829	6	PF08609.8
TCONS_00330829	6	PF03938.12
TCONS_00330829	6	PF06005.10
TCONS_00330829	6	PF01486.15
TCONS_00330829	6	PF06156.11
TCONS_00330829	6	PF07989.9
TCONS_00330829	6	PF05377.9
TCONS_00330829	6	PF04799.11
TCONS_00330829	6	PF12329.6
TCONS_00330829	6	PF04859.10
TCONS_00330829	6	PF07851.11
TCONS_00330829	6	PF10779.7
TCONS_00330829	6	PF11932.6
TCONS_00330829	6	PF02183.16
TCONS_00330829	6	PF14584.4
TCONS_00330829	6	PF01025.17
TCONS_00330829	6	PF05266.12
TCONS_00330829	6	PF08286.9
TCONS_00330829	6	PF02601.13
TCONS_00330829	6	PF15456.4

Transcript_ID	Chr	Pfam Domain
TCONS_00330829	6	PF15456.4
TCONS_00330829	6	PF13600.4
TCONS_00330829	6	PF15898.3
TCONS_00330829	6	PF05529.10
TCONS_00330829	6	PF14077.4
TCONS_00330829	6	PF10199.7
TCONS_00330958	7	PF11630.6
TCONS_00334771	7	PF03221.14
TCONS_00334771	7	PF04218.11
TCONS_00334771	7	PF13384.4
TCONS_00334771	7	PF13936.4
TCONS_00334771	7	PF15030.4
TCONS_00334771	7	PF13518.4
TCONS_00334771	7	PF11242.6
TCONS_00334771	7	PF04545.14
TCONS_00338988	7	PF07686.15
TCONS_00338988	7	PF13927.4
TCONS_00338988	7	PF07679.14
TCONS_00338988	7	PF00047.23
TCONS_00348982	7	PF11270.6
TCONS_00348982	7	PF07829.9
TCONS_00349175	7	PF15742.3
TCONS_00349176	7	PF04433.15
TCONS_00349386	7	PF03372.21
TCONS_00349386	7	PF07780.10
TCONS_00349401	7	PF15788.3
TCONS_00349408	7	PF16823.3
TCONS_00349415	7	PF15788.3
TCONS_00349423	7	PF15788.3
TCONS_00349723	7	PF08562.8
TCONS_00349801	7	PF04193.12
TCONS_00349870	7	PF14529.4
TCONS_00349870	7	PF03372.21
TCONS_00349916	7	PF04695.11
TCONS_00355628	8	PF06897.10
TCONS_00360735	8	PF02093.14
TCONS_00360735	8	PF13696.4
TCONS_00360735	8	PF14787.4
TCONS_00360735	8	PF14392.4
TCONS_00363853	8	PF02337.15
TCONS_00363860	8	PF12356.6
TCONS_00363959	8	PF15449.4
TCONS_00363990	8	PF09907.7
TCONS_00364011	8	PF06637.9
TCONS_00364086	8	PF00078.25
TCONS_00364503	8	PF08333.9
TCONS_00369655	9	PF02687.19
TCONS_00378925	X	PF14691.4
TCONS_00378976	X	PF14670.4
TCONS_00378977	X	PF08333.9
TCONS_00379246	X	PF14386.4

Transcript_ID	Chr	Pfam Domain
TCONS_00379246	X	PF08333.9
TCONS_00384402	X	PF12727.5
TCONS_00387572	X	PF05363.10
TCONS_00389312	X	PF15747.3
TCONS_00389323	X	PF16042.3
TCONS_00389343	X	PF11095.6
TCONS_00389343	X	PF15827.3
TCONS_00389358	X	PF09972.7
TCONS_00389407	X	PF00078.25
TCONS_00389411	X	PF02518.24
TCONS_00389489	X	PF14815.4
TCONS_00389636	X	PF14003.4
TCONS_00389877	X	PF14386.4
TCONS_00379246	X	PF08333.9
TCONS_00384402	X	PF12727.5
TCONS_00387572	X	PF05363.10
TCONS_00389312	X	PF15747.3
TCONS_00389323	X	PF16042.3
TCONS_00389343	X	PF11095.6
TCONS_00389343	X	PF15827.3
TCONS_00389358	X	PF09972.7
TCONS_00389407	X	PF00078.25
TCONS_00389411	X	PF02518.24
TCONS_00389489	X	PF14815.4
TCONS_00389636	X	PF14003.4
TCONS_00389877	X	PF14386.4

APPENDIX C

List of transcription factors (TFs) expressed in human primary monocytes

TF_Name	Chr	FPKM
FOS	14	1042.202
DDX5	17	955.9827
DAZAP2	12	672.8758
MNDA	1	525.7072
PTMA	2	493.4772
ENO1	1	407.774
TSC22D3	X	407.734
CSDE1	1	376.0662
HCLS1	3	338.882
YBX1	1	334.219
STAT1	2	321.2763
KLF6	10	261.9753
DAXX	6	258.9572
CNBP	3	250.2028
JUNB	19	248.6805
LRRFIP1	2	244.1758
STAT6	12	238.7027
POLR2J2	7	238.6764
FLI1	11	214.936
NPM1	5	212.6658
BTF3	5	207.5807
ZFP36	19	202.7012
XRCC5	2	192.6528
ZEB2	2	190.6948
NONO	X	184.2647
IFI16	1	180.9417
SF1	11	178.9018
IRF9	14	164.3558
DEK	6	160.0825
CTBP1	4	152.1875
IRF8	16	150.5813
NFKBIZ	3	144.7393
MATR3	5	138.4562
ATF4	22	132.7715
ZBTB7B	1	130.5948
BZW1	2	129.8494
AES	19	127.3908
SP110	2	125.545
RXRA	9	120.9746
IRF5	7	118.8984
JUND	19	118.3665
STAT3	17	117.402
STAT5A	17	115.7063
HNRNPD	4	110.6388
PAPOLA	14	108.7945
XRCC6	22	108.4456
SND1	7	108.1236
PYGO2	1	107.5096
MEF2C	5	106.845
EWSR1	22	106.7126
RARA	17	105.8238

TF_Name	Chr	FPKM
CREG1	1	103.7564
ABTB1	3	103.3663
RBMX	X	103.3003
CHD4	12	100.5535
MAX	14	99.75922
CIITA	16	99.6059
KLF13	15	99.07043
POU2F2	19	98.98807
GTF2I	7	98.93038
TAF15	17	98.85947
HMGB2	4	98.85848
USF2	19	97.79038
CUX1	7	97.12292
CNOT8	5	95.36182
ILF3	19	94.77232
PNRC2	1	94.69955
TRIM28	19	94.4034
FUS	16	93.15243
KDM2A	11	92.30302
KHSRP	19	91.53088
IRF2	4	88.6939
SSB	2	88.68062
GON4L	1	88.07198
HAX1	1	87.49808
BTG1	12	86.15183
CTNNB1	3	85.87722
STAT5B	17	84.9628
MXD3	5	84.81863
HDAC1	1	84.63437
MBD2	18	83.58965
ELF3	1	81.81857
SATB1	3	79.36793
RBBP4	1	79.3381
CHD2	15	79.20178
RBM14	11	79.10048
HIF1A	14	78.8005
CALCOCO1	12	78.27125
APEX1	14	76.57758
RUNX3	1	76.43263
CITED2	6	76.35672
AHR	7	76.16878
ARID1A	1	76.0421
FOSL2	2	75.41178
KHDRBS1	1	74.70478
LMO2	11	74.20095
CBX3	7	74.19483
ID2	2	73.82327
SUB1	5	73.72447
TNFAIP3	6	73.6102
ARID4B	1	71.20523
MAFB	20	70.79457

TF_Name	Chr	FPKM
UBP1	3	70.39975
SMAD2	18	70.27083
MEF2A	15	69.5369
CEBPB	20	69.23545
RBL2	16	68.97265
DDB1	11	68.67812
KLF2	19	67.63495
UPF1	19	67.38048
E2F4	16	67.34277
CHURC1	14	66.16587
CREB1	2	66.09495
RGS14	5	65.43522
NCOR2	12	65.42433
NFE2	12	65.13302
POLR2E	19	64.85295
ELF4	X	64.81648
MED15	22	64.40745
XBP1	22	63.75543
SRRM1	1	62.6984
PHB2	12	62.50585
CBFB	16	62.17693
SP1	12	62.133
CEBPD	8	60.4411
NFYC	1	60.2628
OVOL2	20	59.88513
BTG2	1	59.5325
CNOT7	8	59.0944
TBL1X	X	58.7304
GRHL1	2	58.13448
MAF1	8	58.0814
NAP1L4	11	58.00868
MXD1	2	57.79998
RFX5	1	57.48113
HHEX	10	57.47582
GATAD2A	19	57.15055
BCLAF1	6	55.7295
SAFB	19	55.6497
DRAP1	11	55.31075
THRAP3	1	55.25855
HSBP1	16	54.45665
CREBBP	16	54.01983
DNMT1	19	53.57922
LYL1	19	53.54062
POLR2A	17	52.98218
XPC	3	52.56275
CNOT2	12	51.97958
RFWD2	1	51.84085
IVNS1ABP	1	51.7525
KLF3	4	51.33032
TCF12	15	50.78518
MKL1	22	50.45203

TF_Name	Chr	FPKM
PIAS1	15	50.34302
SMARCA2	9	50.30687
CBFA2T3	16	50.27498
POLR2B	4	50.19458
SMARCD3	7	49.9529
NCOR1	17	49.66162
FXR1	3	49.64757
YY1	14	49.1526
DR1	1	49.12953
LDB1	10	49.10723
NFKB1	4	48.90215
SMARCC2	12	48.86462
HMGN4	6	48.80078
FBXO7	22	48.76678
HMGN1	21	48.12883
ZNF217	20	48.04457
HBP1	7	47.46742
ZNF638	2	47.31525
ETS2	21	47.25657
BAZ1A	14	46.94467
NSD1	5	46.85027
SNW1	14	46.83863
GTF3A	13	46.49292
TFEB	6	46.43997
EIF2AK2	2	46.23675
PHF21A	11	46.16818
IKBKB	8	46.10455
SAP30BP	17	46.08268
ASCC2	22	45.94443
RUNX1	21	45.68727
DEDD	1	45.6181
COPS2	15	45.61618
RNF141	11	45.50125
RBPJ	4	45.4186
SMARCA5	4	45.23003
TFEC	7	45.12493
NMI	2	45.07793
WHSC1L1	8	44.90272
MBD1	18	44.85022
NCOA1	2	44.59147
NR3C1	5	44.43792
TSG101	11	43.47815
ARNT	1	43.41973
HDAC9	7	43.22592
TLE3	15	42.95517
SAP18	13	42.77353
POLR1D	13	42.76698
PHF12	17	42.62415
HDAC3	5	42.35218
ILF2	1	42.06493
CXXC1	18	42.00663

TF_Name	Chr	FPKM
PNN	14	42.00107
MLLT6	17	41.97328
TIAL1	10	41.82792
MBD3	19	41.72258
APC	5	41.57535
TCF7L2	10	41.48038
PRPF4B	6	41.40198
HTATIP2	11	41.25738
SMAD3	15	40.94818
WBP11	12	40.70338
SREBF2	22	40.44452
JDP2	14	40.18107
CEBPA	19	40.16788
HDAC2	6	39.81938
TFDP1	13	39.48107
YTHDC1	4	39.21903
CREBL2	12	39.20902
ATF1	12	39.1492
SMARCD2	17	39.05605
TAF7	5	39.01695
SUPT6H	17	38.93682
MCM5	22	38.50588
ZBTB7A	19	38.42653
TCEA1	8	38.35138
FUBP1	1	38.17082
PML	15	38.11943
SMARCD1	12	38.05122
RB1	13	37.84023
FOSB	19	37.80448
SKI	1	37.37657
PARP1	1	37.34928
TAF9	5	37.17985
BTAF1	10	37.11942
NELFCD	20	37.09092
HIST1H2BJ	6	37.03945
HSF1	8	36.67937
SAFB2	19	36.47027
SSRP1	11	36.35008
NCOA2	8	36.34383
RERE	1	36.31763
CNOT6L	4	36.31582
DDX54	12	36.27187
PHF20	20	36.09753
EP300	22	35.80338
MAFG	17	35.66073
ETV6	12	35.31202
SPEN	1	34.88793
ZBTB1	14	34.20897
TBC1D22A	22	34.03273
MEF2D	1	33.96097
DMTF1	7	33.88733

TF_Name	Chr	FPKM
SSBP4	19	33.7304
MLXIP	12	33.62937
JMJD1C	10	33.59448
PDLIM1	10	33.47618
NFKB2	10	33.40452
POLR2G	11	33.13857
SUPT16H	14	33.13037
HDAC7	12	32.99997
BRD8	5	32.86302
PHF3	6	32.8452
EHMT1	9	32.63145
AIP	11	32.52618
DHX38	16	32.39572
UBTF	17	32.37837
ZNF146	19	32.24932
AATF	17	32.24427
DPF2	11	32.20937
ATM	11	31.91877
PLRG1	4	31.91218
HTATSF1	X	31.88935
KAT2B	3	31.62212
ATF2	2	31.47092
HMG20B	19	31.40642
ELK3	12	31.13653
GTF3C2	2	31.12293
CHD1	5	31.12055
NRBF2	10	30.91497
VHL	3	30.8044
NR1H2	19	30.63942
CTNNBL1	20	30.54868
TCERG1	5	30.53632
PQBP1	X	30.5217
NCOA6	20	30.2147
NR4A1	12	30.1349
REL	2	29.70537
WWP1	8	29.63757
SMARCA4	19	29.61758
PSIP1	9	29.30023
FMR1	X	29.28762
GTF2A2	15	29.22125
SMARCC1	3	29.1504
TSC22D1	13	28.97472
CREB5	7	28.94242
UBN1	16	28.93523
ZNF467	7	28.91622
ASB8	12	28.86143
MCM7	7	28.84433
ZNFX1	20	28.7263
RBBP7	X	28.72052
ZNF641	12	28.52757
FOXK2	17	28.47753

TF_Name	Chr	FPKM
ATRX	X	28.13532
MED4	13	27.94945
TBC1D2B	15	27.9399
TRIM21	11	27.90017
TCF3	19	27.78943
CTNNBIP1	1	27.66428
ELF2	4	27.44943
BUD31	7	27.34595
ANKZF1	2	27.24748
GTF2F1	19	27.1242
ATF7IP	12	26.95662
ZDHHC2	8	26.83638
BCL3	19	26.69355
CHD8	14	26.64328
CIC	19	26.62027
ARID1B	6	26.46125
HELZ	17	26.29498
MTA1	14	26.2728
ARID3A	19	26.25402
ZNF592	15	26.25005
RREB1	6	26.01732
HMGA1	6	25.54487
SSBP3	1	25.49682
HMGN3	6	25.48303
COPS5	8	25.1739
KAT5	11	25.14922
ERCC3	2	25.03607
SP2	17	25.0211
PBX3	9	24.79918
HCFC1	X	24.73875
PREB	2	24.7249
MTF2	1	24.54718
TAF1C	16	24.53537
ZKSCAN1	7	24.45183
AFF1	4	24.2833
MXI1	10	24.22873
FOXO3	6	24.12502
ATF7	12	23.99618
CTDSPL	3	23.90551
THRB	3	23.8767
TRIM5	11	23.84735
RBM7	11	23.77287
MLLT10	10	23.69475
IRF3	19	23.6027
FOXJ3	1	23.57272
ASH2L	8	23.54827
FOXO4	X	23.37932
NR1D2	3	23.30835
EZH1	17	23.19673
JARID2	6	23.16047
RNF14	5	23.11878

TF_Name	Chr	FPKM
MED1	17	23.07177
TOB1	17	23.057
GTF2H1	11	22.93148
TCEB3	1	22.8463
ZNF276	16	22.76405
SMARCB1	22	22.71057
CRAMP1L	16	22.4508
CENPB	20	22.3835
POGK	1	22.37127
ADNP	20	22.3561
LMO4	1	22.31182
ZNF143	11	22.30585
KLF4	9	22.26334
TRIM33	1	22.25858
ATF5	19	22.16585
TRIP4	15	22.06207
PHF23	17	22.04917
GTF3C5	9	21.83233
ARID4A	14	21.65875
RELB	19	21.60398
BBX	3	21.44438
GTF2B	1	21.39963
SSBP2	5	21.24642
EP400	12	21.03405
FUBP3	9	20.96198
LIMD1	3	20.89983
H1F0	22	20.70123
SREBF1	17	20.68085
RNF2	1	20.6413
ZBTB14	18	20.29568
RB1CC1	8	20.24382
PHF8	X	20.1682
HIF1AN	10	20.1318
CTCF	16	19.82078
KDM4C	9	19.81285
BAZ1B	7	19.74157
ARNTL	11	19.73872
PARN	16	19.71165
ASXL1	20	19.68963
RXRB	6	19.6338
NFIC	19	19.6194
TRPS1	8	19.60423
MLLT1	19	19.57487
TAF1	X	19.54622
RFXANK	19	19.53908
SRF	6	19.39143
PWP1	12	19.26793
GABPA	21	19.1264
ZBTB38	3	19.10817
NAB1	2	18.94953
ATF6	1	18.94723

TF_Name	Chr	FPKM
DIDO1	20	18.88328
C1D	2	18.61645
KDM4B	19	18.61335
PHF2	9	18.41077
SUZ12	17	18.37042
AFF4	5	18.35312
ZNF148	3	18.32007
GTF3C1	16	18.06015
BRPF1	3	17.89145
SLA2	20	17.88842
EAPP	14	17.76853
SIRT1	10	17.7225
CNOT6	5	17.64857
ZNF267	16	17.6424
HIPK2	7	17.57143
MTF1	1	17.56497
CBX4	17	17.46585
GTF2E2	8	17.44725
PHF10	6	17.10105
JUN	1	17.08641
ING3	7	17.01735
GTF3C3	2	16.94342
HMG20A	15	16.85105
ZNF76	6	16.84488
ZNF639	3	16.83883
NCOA5	20	16.80533
MED8	1	16.77917
CNOT3	19	16.74178
POLR2K	8	16.65138
CBX6	22	16.59696
MED23	6	16.45168
RYBP	3	16.42358
YEATS2	3	16.34158
NFIL3	9	16.32176
CEBPZ	2	16.24078
ING1	13	16.21327
RBM15B	3	16.19317
NFATC1	18	16.15333
SERTAD2	2	15.99642
GFI1B	9	15.9416
BRD1	22	15.83993
MAML3	4	15.77632
RFC1	4	15.74038
SNAPC3	9	15.67778
POU4F3	5	15.63402
TAF6	7	15.57697
EHMT2	6	15.53577
ZFX	X	15.50212
PRDM2	1	15.43478
L3MBTL2	22	15.4013
TAF5L	1	15.31815

TF_Name	Chr	FPKM
MEF2B	19	15.30283
SERTAD3	19	15.23302
CTDP1	18	15.18957
EAF1	3	15.15352
TBPL1	6	15.03975
TGIF1	18	14.99333
ZMYND11	10	14.93203
NFYB	12	14.8973
NR2C2	3	14.86049
PIAS4	19	14.81877
NFRKB	11	14.78465
TAF12	1	14.77287
GTF2A1	14	14.74592
BCOR	X	14.6691
E2F3	6	14.62802
EIF5B	2	14.61633
GTF2H3	12	14.58532
ECD	10	14.54478
DMAP1	1	14.51617
TAF13	1	14.47393
RUVBL2	19	14.46592
TSHZ1	18	14.41878
ZBTB34	9	14.38477
ABL1	9	14.36222
ANKRD49	11	14.18637
DOT1L	19	13.9645
ZBTB11	3	13.8827
TCEB1	8	13.83955
ZBTB17	1	13.79278
POLR3E	16	13.70143
PIAS3	1	13.69903
WHSC1	4	13.6214
ZNF281	1	13.58147
DENND4A	15	13.54037
ZNF33A	10	13.4977
ZNF524	19	13.48913
ZBTB2	6	13.4732
SUGP1	19	13.47173
ETS1	11	13.39586
SAP30	4	13.36673
GTF2H4	6	13.29638
TAF11	6	13.28932
YAF2	12	13.28248
ZNF319	16	13.24247
HDAC10	22	13.1724
FOXJ2	12	13.13428
AEBP2	12	13.1285
TSHZ3	19	13.11451
MCM3	6	13.10758
VENTX	10	13.06113
CEBPG	19	13.0199

TF_Name	Chr	FPKM
SKIL	3	12.98227
ELK4	1	12.96085
ELL	19	12.8901
TCF20	22	12.83805
ZBTB33	X	12.80533
DDB2	11	12.72962
NFX1	9	12.6751
TSC22D2	3	12.6475
XAB2	19	12.63388
SUPT7L	2	12.62133
BRWD1	21	12.55973
ZNHIT3	17	12.53796
RBAK	7	12.48162
BRPF3	6	12.43003
MNT	17	12.38538
ZNF496	1	12.22827
REST	4	12.12622
DEAF1	11	12.02123
ZNF644	1	12.02011
DCP1A	3	11.9076
TRERF1	6	11.77942
RAI1	17	11.71026
ATF3	1	11.67196
ZZZ3	1	11.65973
MEN1	11	11.60667
TTF2	1	11.43412
ZFPL1	11	11.42458
GTF2H2	5	11.4135
OLIG1	21	11.29447
MITF	3	11.15535
SNRPD1	18	11.10654
ZNF3	7	11.08445
TCF4	18	10.96138
TAL1	1	10.92101
ERCC2	19	10.83025
SMAD5	5	10.77711
TRIM24	7	10.69025
ASCC3	6	10.63885
ANKRA2	5	10.63601
PHF14	7	10.55771
L3MBTL3	6	10.54261
GTF2F2	13	10.51892
TBP	6	10.51143
HINFP	11	10.37442
CDC5L	6	10.36172
INTS12	4	10.35686
BLZF1	1	10.35256
ZNF140	12	10.33838
POLR2L	11	10.28848
SNAPC2	19	10.25998
CDCA7L	7	10.21792

TF_Name	Chr	FPKM
ANKS1A	6	10.18169
ETV3	1	10.17895
SUFU	10	10.16437
BDP1	5	10.16345
CREB3	9	10.05761
TAF1B	2	10.00743
KLF16	19	10.0035
ZNF688	16	9.983877
HIST1H2BB	6	9.964935
ZDHHC16	10	9.939297
ANKRD42	11	9.930343
KLF9	9	9.85325
ZNF160	19	9.77207
HDAC8	X	9.706592
ZNF224	19	9.693385
HIST1H2BH	6	9.691845
DNMT3A	2	9.595917
ASH1L	1	9.529538
TARBP2	12	9.52945
ZBTB43	9	9.518798
NF1	17	9.368716
IRF4	6	9.251745
ZNF589	3	9.230753
FOXO1	13	9.195192
AHCTF1	1	9.187627
PDCD11	10	9.122602
ZNF177	19	9.088907
PPARGC1B	5	9.075703
TFCP2	12	9.070463
NDNL2	15	9.038128
ZNF646	16	8.97024
PPARA	22	8.941353
ANKRD22	10	8.917772
ZNF7	8	8.917472
ASCC1	10	8.915032
SMAD1	4	8.874687
ZBTB40	1	8.820803
RUVBL1	3	8.812065
POLR1B	2	8.792305
CNOT4	7	8.78585
ZNF211	19	8.746297
IGHMBP2	11	8.712798
BRF1	14	8.702055
SMARCAD1	4	8.698962
HIVEP3	1	8.694205
NRF1	7	8.665262
MCM4	8	8.658652
PBXIP1	1	8.639933
ZNF689	16	8.541023
FOXP4	6	8.524592
ARID3B	15	8.524358

TF_Name	Chr	FPKM
ZNF354A	5	8.523122
SCAND1	20	8.427132
PHF5A	22	8.412105
MYBBP1A	17	8.406397
PHF6	X	8.33036
THRA	17	8.324752
PAPOLG	2	8.293138
MYC	8	8.285762
GTF2E1	3	8.277103
POLR3D	8	8.253087
MZF1	19	8.235702
ZNF22	10	8.23077
EED	11	8.17005
CBFA2T2	20	8.165328
POLR2H	3	8.059885
RFX1	19	8.042833
HIVEP1	6	8.010728
ZNF75A	16	7.932922
PHTF1	1	7.915673
ZNF513	2	7.870305
POLR3H	22	7.870167
ZNF45	19	7.831978
SMAD7	18	7.658198
ZNF397	18	7.65529
BATF	14	7.596422
ZNF430	19	7.59012
PPARD	6	7.570563
ZNF100	19	7.567723
ZNF292	6	7.56267
GLI4	8	7.532827
SS18L1	20	7.494783
TCOF1	5	7.441417
ZNF12	7	7.435045
ZNF227	19	7.3889
ZNF282	7	7.37677
ASB1	2	7.360102
RUNX2	6	7.332825
NFAT5	16	7.311277
POLR1C	6	7.285232
ZFP90	16	7.263037
EGR1	5	7.2592
NFKBIB	19	7.250072
CREM	10	7.207432
SOX4	6	7.198172
ZNF274	19	7.160385
ABT1	6	7.141638
HLX	1	7.125048
ZNF350	19	7.093898
BTBD11	12	7.090518
NR2C1	12	7.052107
MED12	X	6.919143

TF_Name	Chr	FPKM
ZNF326	1	6.796357
GMEB1	1	6.767525
PIAS2	18	6.732197
ERCC6	10	6.714623
MYNN	3	6.710265
ZFP64	20	6.661277
CBX7	22	6.615738
RLF	1	6.560463
ING2	4	6.475408
CHAF1A	19	6.472162
NR4A2	2	6.46898
ZDHHC13	11	6.434525
HIST1H2BG	6	6.41599
ZNF302	19	6.392463
ZNF623	8	6.359547
ZNF213	16	6.345918
E2F6	2	6.335342
TCEAL1	X	6.289673
GMEB2	20	6.274485
ZNF577	19	6.17419
SHPRH	6	6.141523
SPOCD1	1	6.118833
BCL11A	2	6.102017
GATAD2B	1	5.95243
COIL	17	5.924193
THAP7	22	5.863555
RARG	12	5.855587
PMS1	2	5.830193
NFIX	19	5.814322
ZNF134	19	5.777648
ZNF136	19	5.772408
TARBP1	1	5.771468
GZF1	20	5.757373
CLOCK	4	5.732407
POLR1A	2	5.727168
ZNF32	10	5.684903
HKR1	19	5.653048
ANKS3	16	5.63257
ZNRD1	6	5.611885
CRY1	12	5.611005
ELL2	5	5.60035
FXR2	17	5.567132
HIVEP2	6	5.544215
BIN1	2	5.542155
ZNF615	19	5.521308
ERCC8	5	5.498868
TCEA2	20	5.494253
ZNF324	19	5.471603
ZNF226	19	5.465757
CRTC1	19	5.447235
TP53BP1	15	5.375022

TF_Name	Chr	FPKM
ZNF337	20	5.350092
PBX1	1	5.297303
DDX20	1	5.289873
TULP4	6	5.285422
ZNF398	7	5.23888
HIST1H1C	6	5.199417
PER2	2	5.173987
ZNF41	X	5.164853
ANKRD6	6	5.163648
ZNF331	19	5.125833
ZNF202	11	5.035065
POLR3A	10	5.022953
ZNF121	19	5.02253
ZNF101	19	5.009297
EZH2	7	4.981505
ZNF180	19	4.902832
PCBD1	10	4.895702
IRF2BP1	19	4.893188
DBP	19	4.852863
ZNF26	12	4.84231
ZNF189	9	4.794493
TOE1	1	4.788368
PGBD1	6	4.780962
GATA2	3	4.768877
GABPB2	1	4.75355
ZFHX3	16	4.746462
PKNOX1	21	4.714868
ZNF175	19	4.640315
ZNF44	19	4.624465
GTF3C4	9	4.610392
MAFK	7	4.60344
STAT2	12	4.51426
ZNF432	19	4.484232
TCFL5	20	4.48075
ZFP62	5	4.465067
ZNF671	19	4.42702
POLR3F	20	4.419153
DDIT3	12	4.332387
PCGF6	10	4.301487
ZNF124	1	4.214395
ZNF212	7	4.191885
HSF2	6	4.189172
ZNF141	4	4.162155
ZNF155	19	4.162117
TAF9B	X	4.143402
PRDM1	6	4.140938
ZNF408	11	4.105443
ETV7	6	4.102763
ZFP1	16	4.079258
L3MBTL1	20	4.05906
ZFY	Y	4.039953

TF_Name	Chr	FPKM
PRDM10	11	3.989862
XRCC3	14	3.986148
TIMELESS	12	3.969772
MYB	6	3.963463
SIRT5	6	3.960547
ZNF236	18	3.944182
ZNF169	9	3.91924
TEF	22	3.911363
SCMH1	1	3.836387
POLR1E	9	3.826485
ZNF444	19	3.774447
NAB2	12	3.754525
PHF7	3	3.72226
ASCL2	11	3.67031
NKRF	X	3.654192
HIST1H2BD	6	3.611473
FAIM3	1	3.609517
ZNF407	18	3.591705
POU6F1	12	3.585032
MED31	17	3.491865
SETBP1	18	3.477227
ZNF473	19	3.471923
ZNF92	7	3.463527
ZNF383	19	3.446647
TFAP4	16	3.426077
SNAPC4	9	3.39805
POLR2I	19	3.39547
MCM6	2	3.386335
JMY	5	3.349748
TAF1A	1	3.316455
TAF5	10	3.308342
SNAI3	16	3.301198
KLF12	13	3.29379
KLF5	13	3.287538
E2F2	1	3.243022
ZNF675	19	3.229225
HDAC11	3	3.211427
ZSCAN2	15	3.21104
ZSCAN21	7	3.207163
ZNF37A	10	3.161445
E2F1	20	3.159207
JRK	8	3.113103
PER3	1	3.091645
TRIB3	20	3.072555
TGFB1I1	16	3.022133
ZNF16	8	3.01744
ZNF550	19	3.016377
ZNF426	19	3.005772
HES4	1	2.965262
TFDP2	3	2.947622
ZBTB37	1	2.923305

TF_Name	Chr	FPKM
ZBTB10	8	2.915305
ZNF436	1	2.882082
SNIP1	1	2.860823
ZNF81	X	2.851915
ZFP36L1	14	2.836997
ZNF420	19	2.819607
ZHX3	20	2.801618
ZNF366	5	2.795743
ZNF677	19	2.782972
MAF	16	2.768385
ZNF627	19	2.763132
CREB3L4	1	2.746918
ZNF275	X	2.742315
ZNF555	19	2.734475
ZNF606	19	2.733797
RBL1	20	2.709932
ZNF133	20	2.654785
EGR2	10	2.650344
ZNF624	17	2.649982
HMGB3	X	2.601658
ZNF341	20	2.570897
ZNF219	14	2.570895
CITED4	1	2.561235
KLF10	8	2.56016
ZNF74	22	2.534872
ZNF24	18	2.520556
MKL2	16	2.511542
SCML1	X	2.508165
ZBTB5	9	2.456593
HIC2	22	2.452783
PHF13	1	2.438037
MLLT3	9	2.417715
CCRN4L	4	2.403613
L3MBTL4	18	2.325612
ZNF283	19	2.315342
BRCA1	17	2.305103
SFPQ	1	2.289523
MTA3	2	2.281195
POU2F1	1	2.281152
PAX8	2	2.277944
ARID5B	10	2.276418
SERTAD1	19	2.24038
CASZ1	1	2.239083
MCM2	3	2.234122
GATA1	X	2.205859
ZNF449	X	2.20426
CHD7	8	2.20101
EPAS1	2	2.199697
SP4	7	2.198245
ZNF441	19	2.182812
ZNF567	19	2.180138

TF_Name	Chr	FPKM
TRIM32	9	1.743565
MAFF	22	1.743225
MYBL1	8	1.733737
ZNF256	19	1.731406
ZNF507	19	1.727145
PAWR	12	1.716745
NCOA4	10	1.70485
NR3C2	4	1.69311
ZNF670	1	1.691313
ZNF416	19	1.687129
ELL3	15	1.662869
MYEF2	15	1.657553
E2F5	8	1.654377
ZNF611	19	1.605046
RORA	15	1.595837
HOXA1	7	1.58018
ZNF287	17	1.566381
ZNF132	19	1.564089
WDHD1	14	1.560531
ELF1	13	1.551223
ZNF257	19	1.542726
ISL2	15	1.526339
NDN	15	1.52305
ZNF25	10	1.516771
MYBL2	20	1.516526
BRCA2	13	1.515149
PRDM8	4	1.509526
NOTCH3	19	1.504016
FOSL1	11	1.498885
LHX4	1	1.494628
ZNF501	3	1.476536
AEBP1	7	1.469697
SMAD6	15	1.459514
ZNF445	3	1.455734
ZNF551	19	1.448111
TCEA3	1	1.438158
ZNF563	19	1.434911
TBX19	1	1.432405
SNAI1	20	1.432207
ZNF83	19	1.421951
ZNF382	19	1.420888
ZNF70	22	1.412167
ZNF221	19	1.405342
GATA3	10	1.401372
HES1	3	1.399835
ZNF384	12	1.399235
EOMES	3	1.392511
TOX	8	1.390068
PLAG1	8	1.388212
ZNF571	19	1.386229
HIST1H1T	6	1.385058

TF_Name	Chr	FPKM
ZBTB12	6	1.372284
ESPL1	12	1.364781
RAD51	15	1.363872
TRIP13	5	1.360384
ZNF599	19	0.354969
ZNF502	3	0.353657
PPARGC1A	4	0.347919
GLIS3	9	0.347459
ZNF154	19	0.347233
SCML2	X	0.344865
TBX6	16	0.341024
HOXA3	7	0.332782
ZNF540	19	0.329647
PRDM4	12	0.322759
ID3	1	0.320232
NR4A3	9	0.303908
BCL11B	14	0.30344
KLF1	19	0.297027
MYCN	2	0.29469
EHF	11	0.29447
ZNF558	19	0.294096
ZNF491	19	0.277293
PTRF	17	0.27402
BACH1	21	0.268478
TP63	3	0.268178
NR1I3	1	0.265708
MECP2	X	0.262916
PRDM7	16	0.26201
MSC	8	0.260378
HIST1H2BE	6	0.248291
ESR1	6	0.239568
VDR	12	0.23668
BHLHE41	12	0.233486
PAX6	11	0.228212
TFCP2L1	2	0.225602
ZNF43	19	0.217828
GLI1	12	0.217642
ZNF30	19	0.216908
APBB2	4	0.215144
SIRT4	12	0.209031
ZNF440	19	0.206988
IRF6	1	0.204887
OLIG2	21	0.191469
CRABP2	1	0.188078
ZNF451	6	0.187493
ZBTB20	3	0.187463
PLAGL2	20	0.178612
HOXA5	7	0.177446
HEY1	8	0.173259
BRIP1	17	0.169089
FOXP3	X	0.16839

TF_Name	Chr	FPKM
FOXD2	1	0.160157
TWIST2	2	0.15355
NHLH1	1	0.152655
BCL6	3	0.151649
ZNF215	11	0.150789
SOX13	1	0.149713
ANKRD10	13	0.145582
ST18	8	0.143108
PCGF2	17	0.140144
NAP1L3	X	0.138274
NFATC3	16	0.135482
PRDM5	4	0.135471
ZNF442	19	0.131887
NFIB	9	0.131578
ZNF114	19	0.125512
DTX1	12	0.125225
ETV2	19	0.122805
ZNF585B	19	0.12165
ZNF586	19	0.121439
HDGF	1	0.121302
EYA3	1	0.118461
ZNF266	19	0.111161
ZNF135	19	0.111067
NRL	14	0.110141
UTF1	10	0.109835
ZBTB32	19	0.106388
HIST1H1E	6	0.103497
RAI14	5	0.100619
EDF1	9	62.14123
TCEB2	16	48.90437
ING4	12	32.71257
TP53	17	31.39832
POLR2J	7	24.69587
FOXP1	3	22.18372
PHF19	9	21.56587
NFYA	6	20.87853
RING1	6	14.58423
ZNF142	2	11.09459
NFE2L3	7	10.89368
PHF1	6	10.73277
ZBTB16	11	10.62291
ASB6	9	10.49957
HEXIM1	17	10.16937
TOPORS	9	8.690143
REXO4	9	8.25187
NFXL1	4	6.976613
SIRT3	11	6.167273
ZNF480	19	5.077083
ZNF510	9	4.377527
ZBTB26	9	3.099088
TFPT	19	2.866213

TF_Name	Chr	FPKM
ZBTB6	9	2.7398
ZNF79	9	2.570045
ZNF669	1	1.92246
ZNF714	19	1.635685
ZNF273	7	1.533802
TLE1	9	1.390819
ZNF595	3	0.856381
MCM8	20	0.789255
PAX5	9	0.770402
MORF4	4	0.599239
MEIS3P1	17	0.447927
ID1	20	0.389662
PDLIM4	5	0.388846
MKX	10	0.234798
TRIM22	11	0.222017
IRX3	16	0.215956
ZFP37	9	0.203468
TP73	1	0.201627
HEYL	1	0.195776
ZNF471	19	0.187423
ZNF587	19	0.16987
DPF3	14	0.158747
TCF7L1	2	0.152407
RFX2	19	0.149448
HMGA2	12	0.149295
TLE2	19	0.141756
SOX30	5	0.14098
NFE2L1	17	0.129087
ZNF578	19	0.123457
FOXC1	6	0.118753
SALL2	14	0.116045
TRIM15	6	0.115446
SPIC	12	0.114988
MIXL1	1	0.106256
TAF1L	9	0.101194

APPENDIX D

List of identified novel lincRNAs expressed in human primary monocytes

Transcript_ID	Chr	FPKM
TCONS_00014270		1.73
TCONS_00000053	1	1.32
TCONS_00014404	1	3.45
TCONS_00014405	1	0.45
TCONS_00000069	1	0.48
TCONS_00000070	1	0.13
TCONS_00000076	1	0.17
TCONS_00000077	1	0.20
TCONS_00000082	1	0.42
TCONS_00000373	1	0.13
TCONS_00000374	1	0.24
TCONS_00000375	1	0.28
TCONS_00015036	1	0.16
TCONS_00015150	1	0.25
TCONS_00000698	1	0.11
TCONS_00000699	1	0.14
TCONS_00000700	1	0.47
TCONS_00015326	1	0.23
TCONS_00015327	1	0.14
TCONS_00015328	1	0.35
TCONS_00000919	1	0.10
TCONS_00016261	1	0.10
TCONS_00001979	1	0.36
TCONS_00018083	1	0.12
TCONS_00003523	1	0.25
TCONS_00018084	1	0.13
TCONS_00003524	1	0.44
TCONS_00003525	1	0.29
TCONS_00003526	1	0.12
TCONS_00018085	1	0.23
TCONS_00018807	1	0.10
TCONS_00018808	1	0.11
TCONS_00018809	1	0.22
TCONS_00018810	1	0.26
TCONS_00004987	1	0.27
TCONS_00004988	1	0.12
TCONS_00004989	1	0.56
TCONS_00019467	1	0.14
TCONS_00005092	1	0.15
TCONS_00005093	1	0.15
TCONS_00005260	1	0.18
TCONS_00005261	1	0.30
TCONS_00019656	1	0.38
TCONS_00005742	1	0.40
TCONS_00020684	1	0.30
TCONS_00021587	1	0.25
TCONS_00007497	1	0.19
TCONS_00007518	1	0.12
TCONS_00007556	1	0.58
TCONS_00021641	1	2.05
TCONS_00007557	1	1.64

Transcript_ID	Chr	FPKM
TCONS_00021643	1	1.87
TCONS_00021644	1	0.26
TCONS_00021645	1	0.22
TCONS_00021646	1	0.97
TCONS_00021647	1	0.27
TCONS_00007703	1	0.52
TCONS_00007708	1	0.15
TCONS_00007766	1	0.36
TCONS_00007779	1	0.23
TCONS_00022029	1	0.41
TCONS_00007958	1	1.73
TCONS_00007959	1	4.53
TCONS_00022379	1	0.10
TCONS_00008432	1	1.33
TCONS_00008433	1	0.16
TCONS_00008619	1	0.13
TCONS_00009720	1	0.03
TCONS_00024032	1	0.27
TCONS_00024107	1	0.26
TCONS_00024540	1	0.15
TCONS_00024580	1	0.14
TCONS_00024581	1	0.09
TCONS_00011121	1	0.10
TCONS_00025328	1	0.25
TCONS_00025495	1	0.15
TCONS_00011425	1	1.31
TCONS_00011444	1	0.41
TCONS_00025504	1	0.19
TCONS_00011462	1	0.14
TCONS_00011464	1	0.21
TCONS_00011495	1	0.11
TCONS_00012022	1	0.61
TCONS_00026310	1	0.30
TCONS_00026311	1	0.19
TCONS_00026312	1	0.37
TCONS_00026379	1	0.12
TCONS_00026392	1	0.12
TCONS_00012841	1	0.14
TCONS_00012842	1	0.10
TCONS_00026838	1	0.15
TCONS_00026854	1	1.35
TCONS_00026855	1	0.75
TCONS_00026856	1	0.59
TCONS_00027113	1	0.47
TCONS_00013181	1	0.67
TCONS_00013183	1	1.44
TCONS_00013184	1	0.23
TCONS_00013185	1	1.00
TCONS_00013186	1	0.54
TCONS_00027246	1	0.25
TCONS_00027247	1	0.15

Transcript_ID	Chr	FPKM
TCONS_00027248	1	0.37
TCONS_00027250	1	0.34
TCONS_00027251	1	0.19
TCONS_00027252	1	0.22
TCONS_00013450	1	0.11
TCONS_00027810	1	0.16
TCONS_00027811	1	0.58
TCONS_00013853	1	1.69
TCONS_00029690	1	0.12
TCONS_00035692	10	0.57
TCONS_00035789	10	0.14
TCONS_00030510	10	0.13
TCONS_00030512	10	0.16
TCONS_00030514	10	0.18
TCONS_00030546	10	30.61
TCONS_00035865	10	0.13
TCONS_00035866	10	0.11
TCONS_00035867	10	0.15
TCONS_00035868	10	0.11
TCONS_00035869	10	0.12
TCONS_00035870	10	0.25
TCONS_00035871	10	0.33
TCONS_00035872	10	0.31
TCONS_00035873	10	0.18
TCONS_00036041	10	0.25
TCONS_00036042	10	0.16
TCONS_00030839	10	0.28
TCONS_00031078	10	0.20
TCONS_00031080	10	2.71
TCONS_00031081	10	0.22
TCONS_00031082	10	0.23
TCONS_00031083	10	1.13
TCONS_00031090	10	0.22
TCONS_00031091	10	0.21
TCONS_00036328	10	0.25
TCONS_00031097	10	1.77
TCONS_00031098	10	10.38
TCONS_00031324	10	0.22
TCONS_00031331	10	0.51
TCONS_00031927	10	0.29
TCONS_00031929	10	0.18
TCONS_00032077	10	0.45
TCONS_00037475	10	3.88
TCONS_00037514	10	0.64
TCONS_00037690	10	0.77
TCONS_00032672	10	0.15
TCONS_00032905	10	0.42
TCONS_00032906	10	0.19
TCONS_00032912	10	0.13
TCONS_00038233	10	0.11
TCONS_00033833	10	0.43

Transcript_ID	Chr	FPKM
TCONS_00034057	10	0.15
TCONS_00034112	10	0.25
TCONS_00034113	10	0.25
TCONS_00034114	10	0.25
TCONS_00039283	10	0.12
TCONS_00039284	10	0.23
TCONS_00034117	10	0.35
TCONS_00034118	10	0.14
TCONS_00034119	10	0.17
TCONS_00039717	10	0.36
TCONS_00039939	10	0.10
TCONS_00039940	10	1.17
TCONS_00039941	10	0.15
TCONS_00039942	10	0.73
TCONS_00039969	10	0.86
TCONS_00040980	10	0.15
TCONS_00041631	11	0.46
TCONS_00041641	11	0.11
TCONS_00041725	11	0.28
TCONS_00051129	11	0.18
TCONS_00041982	11	0.11
TCONS_00042960	11	0.29
TCONS_00042961	11	0.38
TCONS_00043043	11	0.04
TCONS_00052548	11	0.26
TCONS_00052549	11	0.13
TCONS_00043376	11	0.17
TCONS_00043974	11	0.80
TCONS_00044014	11	0.13
TCONS_00044029	11	0.17
TCONS_00044149	11	2.81
TCONS_00044525	11	0.34
TCONS_00046640	11	0.14
TCONS_00046641	11	0.18
TCONS_00046642	11	0.14
TCONS_00046918	11	10.88
TCONS_00046921	11	0.12
TCONS_00046922	11	0.17
TCONS_00046924	11	7.18
TCONS_00056011	11	0.19
TCONS_00056012	11	0.68
TCONS_00047166	11	0.59
TCONS_00056340	11	0.16
TCONS_00056836	11	0.19
TCONS_00057598	11	1.18
TCONS_00048665	11	0.19
TCONS_00058090	11	0.31
TCONS_00058641	11	0.16
TCONS_00058744	11	6.57
TCONS_00058833	11	0.28
TCONS_00058945	11	0.18

Transcript_ID	Chr	FPKM
TCONS_00068147	11	0.17
TCONS_00059972	12	0.89
TCONS_00069048	12	1.39
TCONS_00069049	12	0.13
TCONS_00069237	12	0.15
TCONS_00069238	12	0.71
TCONS_00060725	12	0.10
TCONS_00060731	12	0.18
TCONS_00060780	12	0.20
TCONS_00061015	12	0.52
TCONS_00070459	12	0.13
TCONS_00071877	12	0.17
TCONS_00071959	12	0.24
TCONS_00073412	12	0.22
TCONS_00064867	12	0.19
TCONS_00074088	12	0.28
TCONS_00074307	12	0.19
TCONS_00065474	12	0.22
TCONS_00074488	12	0.13
TCONS_00074489	12	0.11
TCONS_00066122	12	0.20
TCONS_00067523	12	0.10
TCONS_00076547	12	0.17
TCONS_00076980	12	0.73
TCONS_00076996	12	0.35
TCONS_00077865	12	0.14
TCONS_00077896	13	0.19
TCONS_00077897	13	72.96
TCONS_00077898	13	0.22
TCONS_00077899	13	51.19
TCONS_00080476	13	3.59
TCONS_00078011	13	0.18
TCONS_00078366	13	0.53
TCONS_00078786	13	0.20
TCONS_00078787	13	0.13
TCONS_00078788	13	0.19
TCONS_00081514	13	0.10
TCONS_00081515	13	0.59
TCONS_00081939	13	0.12
TCONS_00079364	13	0.34
TCONS_00079365	13	0.14
TCONS_00079497	13	0.10
TCONS_00079532	13	0.84
TCONS_00082259	13	0.36
TCONS_00079898	13	0.20
TCONS_00080215	13	0.37
TCONS_00082657	13	0.14
TCONS_00082658	13	0.19
TCONS_00084135	13	0.24
TCONS_00084607	14	0.70
TCONS_00084613	14	0.11

Transcript_ID	Chr	FPKM
TCONS_00084614	14	0.10
TCONS_00090899	14	0.15
TCONS_00085159	14	0.11
TCONS_00085316	14	0.17
TCONS_00085694	14	0.27
TCONS_00091848	14	0.96
TCONS_00087210	14	0.92
TCONS_00092982	14	0.16
TCONS_00092987	14	2.69
TCONS_00092994	14	0.25
TCONS_00093552	14	0.71
TCONS_00093842	14	0.56
TCONS_00094223	14	0.65
TCONS_00094224	15	0.65
TCONS_00099918	15	0.11
TCONS_00100257	15	0.31
TCONS_00100258	15	0.41
TCONS_00100259	15	0.36
TCONS_00100260	15	0.15
TCONS_00100556	15	0.17
TCONS_00100557	15	0.19
TCONS_00100558	15	0.12
TCONS_00095296	15	0.15
TCONS_00100795	15	0.16
TCONS_00101563	15	0.13
TCONS_00096937	15	0.81
TCONS_00102502	15	0.35
TCONS_00102503	15	0.66
TCONS_00103021	15	0.15
TCONS_00098573	15	1.31
TCONS_00104483	15	0.12
TCONS_00104484	15	0.17
TCONS_00104485	15	0.06
TCONS_00098896	15	0.28
TCONS_00104690	15	0.49
TCONS_00105140	15	7.38
TCONS_00105150	15	0.57
TCONS_00114641	15	0.13
TCONS_00107428	16	0.33
TCONS_00108323	16	0.50
TCONS_00116426	16	0.28
TCONS_00109005	16	1.53
TCONS_00109047	16	0.28
TCONS_00116918	16	1.79
TCONS_00109105	16	5.12
TCONS_00109125	16	0.19
TCONS_00109346	16	0.16
TCONS_00109404	16	0.97
TCONS_00117779	16	0.18
TCONS_00111601	16	0.19
TCONS_00112412	16	0.16

Transcript_ID	Chr	FPKM
TCONS_00120289	16	0.13
TCONS_00113012	16	0.21
TCONS_00120497	16	0.15
TCONS_00121185	16	0.23
TCONS_00130426	17	0.13
TCONS_00130428	17	0.50
TCONS_00130446	17	0.19
TCONS_00130447	17	9.89
TCONS_00131111	17	0.24
TCONS_00121649	17	0.87
TCONS_00121787	17	0.37
TCONS_00122032	17	0.17
TCONS_00122859	17	0.68
TCONS_00123710	17	0.14
TCONS_00133268	17	0.13
TCONS_00133285	17	0.45
TCONS_00123754	17	9.64
TCONS_00123755	17	9.64
TCONS_00133325	17	0.29
TCONS_00124678	17	0.11
TCONS_00134904	17	0.14
TCONS_00135604	17	0.39
TCONS_00127132	17	0.85
TCONS_00136954	17	0.15
TCONS_00127311	17	0.14
TCONS_00137783	17	0.34
TCONS_00137962	17	0.11
TCONS_00128185	17	0.14
TCONS_00128303	17	0.13
TCONS_00128310	17	0.59
TCONS_00128454	17	0.19
TCONS_00138404	17	0.21
TCONS_00138415	17	0.29
TCONS_00138482	17	0.38
TCONS_00138743	17	1.17
TCONS_00128779	17	2.66
TCONS_00128780	17	0.72
TCONS_00139204	17	0.31
TCONS_00129526	17	0.64
TCONS_00130336	17	1.25
TCONS_00130359	17	0.29
TCONS_00140912	17	0.13
TCONS_00143854	18	0.19
TCONS_00141322	18	0.19
TCONS_00141323	18	0.41
TCONS_00141346	18	0.65
TCONS_00141367	18	0.99
TCONS_00143021	18	0.13
TCONS_00143088	18	0.14
TCONS_00145799	18	0.20
TCONS_00143265	18	0.15

Transcript_ID	Chr	FPKM
TCONS_00145926	18	0.54
TCONS_00143370	18	0.52
TCONS_00146155	18	1.17
TCONS_00155762	19	3.90
TCONS_00155763	19	2.01
TCONS_00155764	19	2.76
TCONS_00155766	19	0.41
TCONS_00155768	19	0.44
TCONS_00146163	19	3.99
TCONS_00146164	19	0.34
TCONS_00155804	19	0.82
TCONS_00146165	19	0.14
TCONS_00155805	19	0.34
TCONS_00155806	19	0.48
TCONS_00146166	19	0.48
TCONS_00155808	19	0.12
TCONS_00155809	19	0.21
TCONS_00155810	19	0.12
TCONS_00155811	19	1.73
TCONS_00155812	19	0.13
TCONS_00155813	19	0.30
TCONS_00155814	19	0.05
TCONS_00155815	19	0.08
TCONS_00155816	19	0.66
TCONS_00155817	19	0.19
TCONS_00155818	19	0.18
TCONS_00155819	19	0.19
TCONS_00155820	19	0.82
TCONS_00155821	19	0.87
TCONS_00155822	19	0.23
TCONS_00155823	19	0.25
TCONS_00155824	19	0.60
TCONS_00146169	19	0.19
TCONS_00146170	19	0.19
TCONS_00155827	19	0.38
TCONS_00155829	19	0.11
TCONS_00155830	19	0.10
TCONS_00155831	19	0.24
TCONS_00146693	19	0.19
TCONS_00147142	19	0.41
TCONS_00147144	19	0.22
TCONS_00157298	19	0.17
TCONS_00157370	19	0.16
TCONS_00147887	19	0.15
TCONS_00148964	19	0.36
TCONS_00160101	19	0.21
TCONS_00150482	19	0.11
TCONS_00150483	19	0.52
TCONS_00150484	19	0.37
TCONS_00150485	19	0.16
TCONS_00152443	19	0.11

Transcript_ID	Chr	FPKM
TCONS_00152616	19	4.05
TCONS_00154321	19	5.12
TCONS_00154325	19	4.11
TCONS_00154437	19	2.90
TCONS_00165621	19	0.13
TCONS_00165622	2	0.14
TCONS_00165623	2	4.98
TCONS_00165624	2	0.03
TCONS_00176764	2	0.65
TCONS_00176869	2	0.20
TCONS_00165829	2	0.74
TCONS_00177176	2	0.27
TCONS_00177207	2	0.11
TCONS_00177208	2	0.16
TCONS_00177209	2	0.15
TCONS_00177210	2	0.30
TCONS_00166052	2	0.13
TCONS_00166053	2	3.85
TCONS_00177333	2	0.39
TCONS_00177334	2	0.44
TCONS_00177632	2	0.15
TCONS_00166602	2	0.13
TCONS_00167282	2	0.19
TCONS_00178496	2	0.15
TCONS_00167328	2	0.16
TCONS_00167329	2	0.48
TCONS_00167330	2	0.57
TCONS_00167331	2	0.27
TCONS_00167332	2	0.27
TCONS_00167333	2	0.89
TCONS_00167335	2	0.22
TCONS_00178505	2	5.60
TCONS_00167597	2	0.20
TCONS_00167598	2	0.11
TCONS_00178733	2	2.04
TCONS_00167644	2	0.11
TCONS_00167645	2	0.23
TCONS_00167646	2	6.00
TCONS_00167647	2	0.93
TCONS_00167648	2	7.09
TCONS_00167649	2	0.89
TCONS_00167650	2	0.24
TCONS_00167651	2	0.26
TCONS_00167652	2	0.17
TCONS_00167653	2	0.14
TCONS_00167654	2	0.26
TCONS_00167655	2	0.19
TCONS_00167656	2	0.16
TCONS_00167982	2	0.45
TCONS_00167983	2	0.21
TCONS_00168098	2	0.12

Transcript_ID	Chr	FPKM
TCONS_00168099	2	0.77
TCONS_00168101	2	0.23
TCONS_00179334	2	6.56
TCONS_00179490	2	0.08
TCONS_00180254	2	1.39
TCONS_00168885	2	0.18
TCONS_00168963	2	0.18
TCONS_00180419	2	0.13
TCONS_00180420	2	0.11
TCONS_00180422	2	0.53
TCONS_00180423	2	0.44
TCONS_00169117	2	8.50
TCONS_00180600	2	0.21
TCONS_00180734	2	0.55
TCONS_00169316	2	0.68
TCONS_00169344	2	0.30
TCONS_00169345	2	0.17
TCONS_00180871	2	0.57
TCONS_00180872	2	4.54
TCONS_00180873	2	0.71
TCONS_00180874	2	3.39
TCONS_00180902	2	0.53
TCONS_00170121	2	0.39
TCONS_00170122	2	0.12
TCONS_00181568	2	0.33
TCONS_00170506	2	0.14
TCONS_00181966	2	0.13
TCONS_00170673	2	0.32
TCONS_00182030	2	0.39
TCONS_00170756	2	0.13
TCONS_00170796	2	0.39
TCONS_00182079	2	0.68
TCONS_00182099	2	0.56
TCONS_00171015	2	0.30
TCONS_00182530	2	0.11
TCONS_00182531	2	5.40
TCONS_00171548	2	0.26
TCONS_00171550	2	0.11
TCONS_00182686	2	0.20
TCONS_00182687	2	0.17
TCONS_00182905	2	0.21
TCONS_00183059	2	0.19
TCONS_00172349	2	0.37
TCONS_00184340	2	0.57
TCONS_00184371	2	0.34
TCONS_00184386	2	0.45
TCONS_00184442	2	0.11
TCONS_00184443	2	0.79
TCONS_00173439	2	0.17
TCONS_00185277	2	0.85
TCONS_00168099	2	0.77

Transcript_ID	Chr	FPKM
TCONS_00174367	2	5.48
TCONS_00175281	2	6.03
TCONS_00175282	2	0.08
TCONS_00175284	2	0.53
TCONS_00186583	2	2.09
TCONS_00186584	2	0.17
TCONS_00186586	2	0.36
TCONS_00175887	2	0.20
TCONS_00175930	2	0.28
TCONS_00176134	2	0.20
TCONS_00176136	2	0.18
TCONS_00176137	2	5.01
TCONS_00186798	2	0.29
TCONS_00186801	2	6.04
TCONS_00176522	2	0.19
TCONS_00188462	2	0.31
TCONS_00188556	20	0.32
TCONS_00188773	20	0.27
TCONS_00192615	20	0.84
TCONS_00192616	20	0.23
TCONS_00192960	20	0.77
TCONS_00192974	20	0.46
TCONS_00189405	20	0.63
TCONS_00189406	20	0.87
TCONS_00192980	20	0.98
TCONS_00192981	20	0.53
TCONS_00192998	20	3.62
TCONS_00193636	20	0.21
TCONS_00190066	20	0.42
TCONS_00190252	20	0.22
TCONS_00193766	20	0.21
TCONS_00193845	20	0.55
TCONS_00190980	20	0.27
TCONS_00190981	20	0.30
TCONS_00190982	20	0.44
TCONS_00190983	20	0.18
TCONS_00190984	20	0.12
TCONS_00190985	20	0.32
TCONS_00190987	20	0.34
TCONS_00194529	20	6.37
TCONS_00194566	20	0.36
TCONS_00194567	20	0.57
TCONS_00195409	20	0.15
TCONS_00195410	21	0.33
TCONS_00197354	21	0.14
TCONS_00197447	21	0.18
TCONS_00195601	21	0.25
TCONS_00195602	21	0.33
TCONS_00198180	21	0.76
TCONS_00196800	21	0.39
TCONS_00196853	21	0.22

Transcript_ID	Chr	FPKM
TCONS_00198850	21	0.58
TCONS_00198851	21	0.23
TCONS_00198852	21	0.10
TCONS_00198853	21	0.25
TCONS_00198857	21	0.55
TCONS_00198858	21	0.12
TCONS_00203038	21	0.18
TCONS_00199425	22	0.12
TCONS_00199426	22	0.30
TCONS_00199427	22	0.10
TCONS_00203419	22	0.11
TCONS_00200431	22	0.25
TCONS_00200432	22	0.11
TCONS_00200780	22	0.79
TCONS_00205439	22	0.77
TCONS_00201968	22	0.87
TCONS_00201969	22	0.41
TCONS_00205612	22	20.30
TCONS_00202345	22	0.13
TCONS_00205859	22	0.20
TCONS_00202667	22	1.27
TCONS_00202668	22	0.28
TCONS_00202669	22	0.37
TCONS_00206147	22	43.05
TCONS_00207581	22	1.23
TCONS_00216725	3	0.08
TCONS_00208141	3	4.05
TCONS_00208473	3	0.14
TCONS_00208474	3	0.12
TCONS_00208569	3	6.55
TCONS_00208586	3	5.04
TCONS_00217383	3	7.14
TCONS_00217711	3	3.45
TCONS_00210300	3	2.70
TCONS_00219588	3	0.63
TCONS_00219589	3	0.10
TCONS_00219781	3	0.35
TCONS_00219805	3	0.48
TCONS_00219806	3	2.54
TCONS_00219828	3	2.27
TCONS_00211057	3	0.94
TCONS_00211063	3	0.19
TCONS_00220116	3	0.17
TCONS_00211162	3	0.31
TCONS_00211163	3	0.14
TCONS_00211164	3	0.66
TCONS_00220228	3	0.83
TCONS_00211955	3	0.84
TCONS_00211957	3	0.95
TCONS_00221230	3	0.22
TCONS_00221259	3	0.23

Transcript_ID	Chr	FPKM
TCONS_00212144	3	0.47
TCONS_00212145	3	0.12
TCONS_00212146	3	0.15
TCONS_00221425	3	0.69
TCONS_00221426	3	0.44
TCONS_00221427	3	0.24
TCONS_00212267	3	0.33
TCONS_00212950	3	1.24
TCONS_00222409	3	0.22
TCONS_00222410	3	0.14
TCONS_00213248	3	0.10
TCONS_00222480	3	0.12
TCONS_00222547	3	0.40
TCONS_00223591	3	0.46
TCONS_00223592	3	0.41
TCONS_00223593	3	0.15
TCONS_00215284	3	0.35
TCONS_00215297	3	0.13
TCONS_00224586	3	0.21
TCONS_00224587	3	0.31
TCONS_00215436	3	0.15
TCONS_00215437	3	0.74
TCONS_00215439	3	0.52
TCONS_00215440	3	0.34
TCONS_00224598	3	0.22
TCONS_00224599	3	0.76
TCONS_00224600	3	0.19
TCONS_00224601	3	0.22
TCONS_00224602	3	4.00
TCONS_00215442	3	1.79
TCONS_00231635	3	0.43
TCONS_00225918	4	0.69
TCONS_00231672	4	0.65
TCONS_00231673	4	0.33
TCONS_00231674	4	0.62
TCONS_00231712	4	0.12
TCONS_00226266	4	16.23
TCONS_00226267	4	0.12
TCONS_00226321	4	0.74
TCONS_00226324	4	0.11
TCONS_00231956	4	0.15
TCONS_00232075	4	0.25
TCONS_00232174	4	0.26
TCONS_00226608	4	0.16
TCONS_00227186	4	0.10
TCONS_00227187	4	0.45
TCONS_00227188	4	0.33
TCONS_00233090	4	0.18
TCONS_00233331	4	0.15
TCONS_00227671	4	1.19
TCONS_00227901	4	0.29

Transcript_ID	Chr	FPKM
TCONS_00227902	4	0.14
TCONS_00227903	4	1.03
TCONS_00228170	4	0.41
TCONS_00228172	4	0.05
TCONS_00233778	4	0.17
TCONS_00234280	4	0.44
TCONS_00234281	4	0.33
TCONS_00234282	4	1.06
TCONS_00234283	4	0.35
TCONS_00234538	4	1.06
TCONS_00229272	4	0.13
TCONS_00229284	4	0.65
TCONS_00229348	4	0.66
TCONS_00235088	4	0.20
TCONS_00235089	4	0.26
TCONS_00235090	4	0.13
TCONS_00229804	4	1.50
TCONS_00229807	4	0.33
TCONS_00230372	4	0.11
TCONS_00236055	4	0.17
TCONS_00236336	4	0.17
TCONS_00236659	4	0.39
TCONS_00231182	4	0.54
TCONS_00244852	4	0.17
TCONS_00244914	5	0.17
TCONS_00237832	5	0.37
TCONS_00237833	5	0.65
TCONS_00245016	5	2.40
TCONS_00245341	5	0.24
TCONS_00245787	5	0.31
TCONS_00245886	5	0.11
TCONS_00245887	5	0.32
TCONS_00245888	5	2.02
TCONS_00245889	5	0.10
TCONS_00245890	5	0.98
TCONS_00245891	5	0.61
TCONS_00238665	5	5.39
TCONS_00238666	5	5.09
TCONS_00238667	5	0.14
TCONS_00238777	5	0.11
TCONS_00238885	5	3.47
TCONS_00238887	5	0.16
TCONS_00246377	5	0.59
TCONS_00246397	5	0.22
TCONS_00239253	5	0.17
TCONS_00239254	5	0.16
TCONS_00239497	5	0.42
TCONS_00246586	5	0.18
TCONS_00239541	5	1.04
TCONS_00246650	5	0.27
TCONS_00239611	5	0.61

Transcript_ID	Chr	FPKM
TCONS_00246677	5	6.61
TCONS_00239688	5	5.34
TCONS_00246683	5	0.30
TCONS_00246684	5	0.66
TCONS_00239694	5	0.46
TCONS_00239695	5	0.29
TCONS_00239698	5	0.48
TCONS_00239703	5	0.13
TCONS_00246721	5	0.46
TCONS_00246760	5	0.62
TCONS_00246762	5	0.21
TCONS_00246763	5	0.10
TCONS_00246772	5	0.13
TCONS_00240362	5	1.32
TCONS_00247446	5	0.23
TCONS_00241451	5	0.06
TCONS_00241452	5	0.20
TCONS_00248359	5	0.26
TCONS_00248794	5	0.65
TCONS_00249139	5	0.25
TCONS_00242666	5	0.16
TCONS_00249618	5	0.59
TCONS_00242888	5	0.14
TCONS_00250491	5	0.16
TCONS_00251057	5	9.05
TCONS_00251058	5	0.23
TCONS_00251060	5	0.66
TCONS_00251394	5	0.44
TCONS_00244698	5	0.15
TCONS_00244699	5	0.52
TCONS_00252347	5	0.28
TCONS_00252355	6	0.35
TCONS_00252356	6	0.33
TCONS_00252434	6	0.54
TCONS_00258947	6	0.20
TCONS_00252525	6	0.15
TCONS_00259069	6	0.40
TCONS_00252526	6	0.92
TCONS_00252785	6	0.23
TCONS_00252908	6	0.22
TCONS_00259741	6	0.01
TCONS_00253333	6	0.44
TCONS_00253334	6	0.45
TCONS_00259825	6	3.04
TCONS_00254804	6	2.21
TCONS_00261564	6	2.01
TCONS_00261712	6	0.27
TCONS_00255576	6	0.12
TCONS_00262238	6	0.67
TCONS_00262497	6	0.15
TCONS_00256117	6	0.64

Transcript_ID	Chr	FPKM
TCONS_00262780	6	0.77
TCONS_00262781	6	0.41
TCONS_00262843	6	2.42
TCONS_00263064	6	0.76
TCONS_00263074	6	0.58
TCONS_00263076	6	0.31
TCONS_00256655	6	0.12
TCONS_00256873	6	0.18
TCONS_00263712	6	0.85
TCONS_00257002	6	0.11
TCONS_00257311	6	0.64
TCONS_00257312	6	0.42
TCONS_00257662	6	0.26
TCONS_00257663	6	0.22
TCONS_00257747	6	0.45
TCONS_00257831	6	0.25
TCONS_00257865	6	0.59
TCONS_00264883	6	0.52
TCONS_00257991	6	0.28
TCONS_00257992	6	0.11
TCONS_00257993	6	0.47
TCONS_00265059	6	0.13
TCONS_00258162	6	0.27
TCONS_00258283	6	0.11
TCONS_00258740	6	0.25
TCONS_00258741	6	0.21
TCONS_00265719	6	22.47
TCONS_00274463	6	4.00
TCONS_00274465	7	0.22
TCONS_00274466	7	0.34
TCONS_00274745	7	0.47
TCONS_00274746	7	2.10
TCONS_00267120	7	0.74
TCONS_00275021	7	0.11
TCONS_00267634	7	0.13
TCONS_00267635	7	0.30
TCONS_00267636	7	0.18
TCONS_00267637	7	0.64
TCONS_00275217	7	0.19
TCONS_00275494	7	0.42
TCONS_00275495	7	0.25
TCONS_00275504	7	6.00
TCONS_00267945	7	0.23
TCONS_00267946	7	0.75
TCONS_00267947	7	0.83
TCONS_00275531	7	0.88
TCONS_00275532	7	0.60
TCONS_00275680	7	0.14
TCONS_00275681	7	0.13
TCONS_00275684	7	0.50
TCONS_00276283	7	0.85

Transcript_ID	Chr	FPKM
TCONS_00276546	7	0.27
TCONS_00276571	7	0.53
TCONS_00276611	7	0.33
TCONS_00276641	7	0.43
TCONS_00268966	7	0.26
TCONS_00268970	7	0.36
TCONS_00268973	7	0.21
TCONS_00268974	7	0.39
TCONS_00268975	7	0.14
TCONS_00268976	7	0.79
TCONS_00268977	7	0.15
TCONS_00268978	7	0.20
TCONS_00268979	7	0.56
TCONS_00268980	7	0.17
TCONS_00268981	7	0.14
TCONS_00268982	7	0.27
TCONS_00268983	7	0.20
TCONS_00268984	7	0.14
TCONS_00268985	7	0.35
TCONS_00268986	7	0.14
TCONS_00268987	7	0.23
TCONS_00268988	7	0.18
TCONS_00276758	7	0.21
TCONS_00276861	7	0.39
TCONS_00276862	7	0.53
TCONS_00276982	7	0.61
TCONS_00276983	7	0.13
TCONS_00276984	7	17.76
TCONS_00276985	7	5.24
TCONS_00276986	7	0.43
TCONS_00276987	7	18.76
TCONS_00276988	7	0.31
TCONS_00276989	7	0.42
TCONS_00276990	7	23.89
TCONS_00276991	7	0.51
TCONS_00276992	7	38.97
TCONS_00276993	7	1.15
TCONS_00276994	7	7.73
TCONS_00277157	7	0.59
TCONS_00277159	7	0.10
TCONS_00277160	7	0.17
TCONS_00277162	7	0.25
TCONS_00269465	7	0.15
TCONS_00277282	7	0.41
TCONS_00277602	7	0.46
TCONS_00270383	7	0.28
TCONS_00278599	7	1.02
TCONS_00278604	7	5.01
TCONS_00271274	7	4.15
TCONS_00271275	7	3.10
TCONS_00271720	7	7.86

Transcript_ID	Chr	FPKM
TCONS_00271721	7	0.28
TCONS_00279328	7	0.76
TCONS_00279329	7	3.33
TCONS_00279330	7	0.05
TCONS_00272473	7	0.22
TCONS_00280138	7	0.28
TCONS_00280222	7	1.67
TCONS_00280224	7	0.67
TCONS_00280227	7	0.24
TCONS_00280232	7	0.17
TCONS_00272967	7	0.06
TCONS_00272968	7	0.09
TCONS_00273158	7	0.28
TCONS_00273159	7	0.11
TCONS_00273378	7	0.33
TCONS_00273380	7	0.06
TCONS_00280941	7	0.13
TCONS_00273650	7	0.13
TCONS_00281430	7	0.15
TCONS_00281463	7	0.01
TCONS_00281489	7	0.15
TCONS_00287582	7	0.21
TCONS_00287594	8	4.73
TCONS_00282506	8	0.95
TCONS_00282547	8	10.42
TCONS_00282548	8	0.07
TCONS_00282553	8	0.56
TCONS_00282577	8	0.10
TCONS_00282578	8	12.96
TCONS_00282581	8	7.51
TCONS_00287790	8	2.67
TCONS_00282609	8	0.30
TCONS_00282610	8	0.30
TCONS_00282611	8	0.13
TCONS_00282613	8	0.86
TCONS_00287808	8	0.80
TCONS_00287809	8	7.02
TCONS_00282614	8	1.29
TCONS_00282615	8	1.04
TCONS_00282616	8	108.77
TCONS_00287810	8	1.95
TCONS_00282617	8	0.16
TCONS_00282645	8	0.11
TCONS_00282646	8	0.31
TCONS_00282651	8	0.29
TCONS_00282724	8	0.17
TCONS_00287980	8	5.60
TCONS_00287981	8	8.21
TCONS_00287982	8	7.06
TCONS_00282868	8	5.19
TCONS_00282870	8	0.55

Transcript_ID	Chr	FPKM
TCONS_00288042	8	0.70
TCONS_00288043	8	0.46
TCONS_00282881	8	0.37
TCONS_00288047	8	5.75
TCONS_00288048	8	0.90
TCONS_00288246	8	0.14
TCONS_00283033	8	0.16
TCONS_00288389	8	0.11
TCONS_00283500	8	0.12
TCONS_00283899	8	0.15
TCONS_00284071	8	0.27
TCONS_00284072	8	0.10
TCONS_00284785	8	0.15
TCONS_00285407	8	0.36
TCONS_00285437	8	0.74
TCONS_00285983	8	0.12
TCONS_00286164	8	0.05
TCONS_00291812	8	0.31
TCONS_00286660	8	0.12
TCONS_00292120	8	1.50
TCONS_00286903	8	0.36
TCONS_00292301	8	0.42
TCONS_00292302	8	0.12
TCONS_00292303	8	0.15
TCONS_00292305	8	0.23
TCONS_00292307	8	0.79
TCONS_00292559	8	0.40
TCONS_00293202	8	0.38
TCONS_00299325	8	0.18
TCONS_00299326	9	0.31
TCONS_00294236	9	0.14
TCONS_00294323	9	0.13
TCONS_00294962	9	0.15
TCONS_00300727	9	0.14
TCONS_00295088	9	0.32
TCONS_00300744	9	0.25
TCONS_00300752	9	0.34
TCONS_00295137	9	0.40
TCONS_00295194	9	0.55
TCONS_00300879	9	0.34
TCONS_00295289	9	0.14
TCONS_00300916	9	0.37
TCONS_00295348	9	0.37
TCONS_00295356	9	0.30
TCONS_00295358	9	0.75
TCONS_00295365	9	0.10
TCONS_00301001	9	0.53
TCONS_00301017	9	0.47
TCONS_00295428	9	0.19
TCONS_00295430	9	0.15
TCONS_00295479	9	0.13

Transcript_ID	Chr	FPKM
TCONS_00295695	9	0.17
TCONS_00301240	9	0.14
TCONS_00295996	9	0.47
TCONS_00296134	9	0.35
TCONS_00296135	9	0.36
TCONS_00296546	9	0.14
TCONS_00296547	9	0.26
TCONS_00302316	9	0.35
TCONS_00302319	9	0.27
TCONS_00297119	9	0.18
TCONS_00297149	9	0.22
TCONS_00297641	9	1.14
TCONS_00298585	9	0.11
TCONS_00298618	9	0.34
TCONS_00298623	9	0.66
TCONS_00303949	9	0.25
TCONS_00298809	9	0.10
TCONS_00304364	9	0.46
TCONS_00299207	9	0.36
TCONS_00299208	9	0.11
TCONS_00299212	9	0.37
TCONS_00299214	9	0.35
TCONS_00299215	9	0.12
TCONS_00304486	9	1.52
TCONS_00306049	9	0.50
TCONS_00306053	X	0.35
TCONS_00306054	X	0.44
TCONS_00306064	X	0.19
TCONS_00306123	X	3.04
TCONS_00306140	X	0.28
TCONS_00306141	X	0.14
TCONS_00306144	X	75.04
TCONS_00310591	X	0.66
TCONS_00310592	X	0.10
TCONS_00306901	X	0.49
TCONS_00306906	X	0.44
TCONS_00307674	X	0.29
TCONS_00312235	X	0.38
TCONS_00312257	X	0.26
TCONS_00307836	X	3.06
TCONS_00307837	X	2.57
TCONS_00307838	X	3.24
TCONS_00312349	X	0.46
TCONS_00312432	X	0.22
TCONS_00312527	X	0.83
TCONS_00312528	X	0.13
TCONS_00308334	X	0.35
TCONS_00308346	X	7.45
TCONS_00308361	X	0.11
TCONS_00308362	X	0.27
TCONS_00313389	X	0.57

Transcript_ID	Chr	FPKM
TCONS_00309200	X	0.12
TCONS_00309202	X	0.18
TCONS_00309203	X	0.13
TCONS_00313512	X	0.17
TCONS_00309302	X	1.22
TCONS_00313681	X	0.34
TCONS_00314044	X	0.14
TCONS_00314843	X	0.16
TCONS_00315687	X	0.26
TCONS_00316509	Y	0.14
TCONS_00316012	Y	0.11
TCONS_00316013	Y	0.17
TCONS_00316016	Y	0.15

APPENDIX E

List of Bioinformatics resourse that are used for this study

Software/Database	Description	Source
FastQC	A quality control tool for high throughput sequence data.	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trimmomatic	A flexible read trimming tool for Illumina NGS data	http://www.usadellab.org/cms/?page=trimmomatic
HISAT	Fast and sensitive spliced alignment program.	http://www.ccb.jhu.edu/software/hisat/index.shtml
StringTie	Fast and highly efficient assembler of RNA-Seq alignments into potential transcripts	https://ccb.jhu.edu/software/stringtie/
Cufflinks	Assembler of RNA-Seq alignments/perform differential expression analysis.	http://cole-trapnell-lab.github.io/cufflinks/
GENCODE	The reference human genome annotation database	https://www.gencodegenes.org/
DAVID	The Database for Annotation, Visualization and IntegratedDiscovery	https://david.ncifcrf.gov/
WebGestalt	Functional enrichment analysis web tool	http://www.webgestalt.org/option.php
"NMF" R package	A flexible R package for nonnegative matrix factorization	https://cran.r-project.org/web/packages/reshape2/index.html
Cytoscape	an open source software platform for visualizing molecular interaction networks	http://www.cytoscape.org/
RNAFold Vienna RNA package	Tool to predict minimum free energy structures and base pair probabilities from single RNA or DNA sequences.	http://rna.tbi.univie.ac.at/
CPAT	Coding-potential assessment tool using an alignment-free logistic regression model	http://rna-cpat.sourceforge.net/
GREAT	Tool to predict functions of cis-regulatory regions.	http://bejerano.stanford.edu/great/public/html/
BLASTN	Basic Local Alignment Search Tool to compare a nucleotide query sequence against a nucleotide sequence database.	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch