An Investigation into the MicroRNA-gene interactions involved in the pathogenesis of Systemic Lupus Erythematosus

STEPHANIE JULIA PITTS (3050456)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR A DEGREE OF MAGISTER SCIENTIAE

at the South African National Bioinformatics Institute,

University of the Western Cape

Supervisor: Dr Nicki Tiffin

Co-Supervisor: Dr Junaid Gamieldien

November 2015

"An Investigation into the MicroRNA-gene interactions involved in the pathogenesis of Systemic Lupus Erythematosus"

Stephanie Julia Pitts

Keywords:

Systemic Lupus Erythematosus, *microRNA*, *genes*, *microRNA-Target Interactions*

1. ABSTRACT

"An Investigation into the MicroRNA-gene interactions involved in the pathogenesis of Systemic Lupus Erythematosus"

S.J. Pitts

MSc Thesis,

South African National Bioinformatics Institute,

University of the Western Cape

Systemic lupus erythematosus is a chronic, inflammatory disease characterised by the production of autoantibodies which target particularly the nuclear components of multiple cell types throughout the body. MicroRNA's have been well-established to regulate gene function by partial-, or complete binding to the 3'-UTR of the target genes, causing repression or complete degradation of the target gene. As a result, proteins normally produced by the targeted mRNA would exhibit a decrease in production.

The aim of this study was to investigate the interactions between genes and microRNAs implicated in the pathogenesis of SLE. Objectives included curating lists of miRNAs and genes associated with lupus pathogenesis, to identify regulatory targets of miRNAs and genes targeted by miRNAs, and to find the intersections of these outputs. By examining the intersections of the resultant targets, we aimed to identify novel interactions using Pathway Analysis, which have not been previously reported in scientific literature, to be associated with the pathogenesis of SLE.

Understanding the miRNA-gene target interactions in the progression of SLE may provide us with essential biomarkers and targets for disease diagnosis and therapy.

Date: November 2015

2. DECLARATION

I declare that *"An Investigation into the MicroRNA-gene interactions involved in the pathogenesis of Systemic Lupus Erythematosus"* is my own work, that it has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name: Stephanie Julia Pitts Date: November 2015

Signed: ……

3. ACKNOWLEDGEMENTS

The Almighty – for without His grace and blessings, this thesis study would not have been possible. To my parents: Thank you for your love, patience, and concern for me throughout this degree. To aunty Charmaine and uncle Malcolm (my second-parents) – Thank you for your encouragement and support throughout my endeavours to obtain higher degrees.

To Francuois Müller- you have been my rock and support. You have been my inspiration, and kept the fire within me alive to strive for bigger and better achievements in life, and in my research.

To Dr Nicki Tiffin – thank you for your faith in me as a student; for allowing me to be a free-thinker throughout this project, yet pulling in the reigns when I strayed too far. Your support throughout this project has made me a stronger person, and helped develop me into someone I hope will become a true scientist.

To Dr Junaid Gamieldien – thank you for your pearls of wisdom at random moments throughout my MSc degree. Your door was always open, and your enthusiasm for bigger and better projects will always inspire me to push the boundaries of my research.

Tiffin Lab group (2014-2015): Jean-Baka, Galen, Darlington, Hocine, and Larry. Thank you for your willingness to assist me at any hour, with any problem. Without your patience and guidance, this project may not have seen completion. Thank you to the DAAD-NRF for funding this Masters project.

To Tracey Calvert Joshua - for your invaluable advice, patience, and health tips ☺. To Clint – For always reminding me about 'second breakfast' and lunch ;), and Ereshia – thank you for a friendship that fate must've wished for, and which promises to be everlasting. As we often ask ourselves, how did we manage to get through our degrees if we went out to lunch so often ;) ? It was with your support that I've made it through this year, and I thank you all for that.

4. CONTENTS

5. TABLE OF FIGURES

6. LIST OF TABLES

7. LIST OF ACRONYMS

5-mC	5-methylcytosine
ANA	Anti-Nuclear Antibody
anti-ds DNA	anti-double stranded Deoxyribonucleic acid
ATP	Adenosine Triphosphate
CpG	Cytosine-phosphate Guanine
DIL	Drug-Induced Lupus
DNA	Deoxyribonucleic acid
DNMT1	DNA Cytosine-S-Methyltransferase 1
EBV	Epstein-Barr Virus
ERV3	Endogenous Retrovirus Group 3
GUI	Graphical User Interface
GWAS	Genome Wide Association Study
HRES1	Human T cell lymphotrophic Virus
HRT	Hormone Replacement Therapy
IFN-I	Type 1 Interferon
IFNγ	Interferon gamma
LMP1	Latent Membrane Protein 1
miRNA	micro-Ribonucleic acid
mRNA	messenger Ribonucleic acid
MS	Multiple Sclerosis
NGS	Next-Generation Sequencing
NO	Nitric Oxide
OCP	Oral Contraceptive Pill
PAMPs	Pathogen-Associated Molecular Patterns
PBMC	Peripheral Blood Mononuclear Cells
PKC δ	Protein kinase C δ
pre-miRNA	precursor-miRNA
pri-miRNA	primary-miRNA
PS	Phosphatidylserine
pSILAC	pulsed Stable Isotope Labelling with Amino acids in Culture
qPCR	quantitative Polymerase Chain Reaction
RA	Rheumatoid Arthritis
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic acid
RNAi	Ribonucleic Acid interference
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
SAM	S-adenosylmethionine
SELENA	Safety of Estrogen in Lupus National Assessment
SLE	Systemic Lupus Erythematosus
SLEDAI	Systemic Lupus Erythematosus Disease Activity Index
SLICC	Systemic Lupus International Collaborating Clinics
SNP	Single Nucleotide Polymorphism

T_c	Cytotoxic T cell
T_h	Helper T cell
TET	ten-eleven translocation
TIFF	Tagged Image File Format
TLR	Toll-Like Receptor
TNF	Tumour Necrosis Factor
TNFR	Tumour Necrosis Factor Receptor
Treg	Regulatory T cell
UK	United Kingdom
USA	United States of America
UTR	Untranslated Region
UV	Ultraviolet

8. INTRODUCTION

8.1. Maintenance of Immune System Homeostasis

8.1.1. Normal functioning of the immune system

8.1.1.1. Innate immunity

The innate immune system serves as the first line of defence against pathogens which have breached the protective barriers of the human body. The innate immune system is made up of mast cells, natural killer cells, complement proteins, and granulocytes which comprise the largest percentage of peripheral blood mononuclear cells (PBMC's) in the circulatory blood. Granulocytes, arising from immune stem cells, further subdivide into neutrophils, eosinophils, basophils, mast cells, and monocytes which subdivide into dendritic cell and macrophage populations, as can be seen in Figure 1 (Dranoff, 2004).

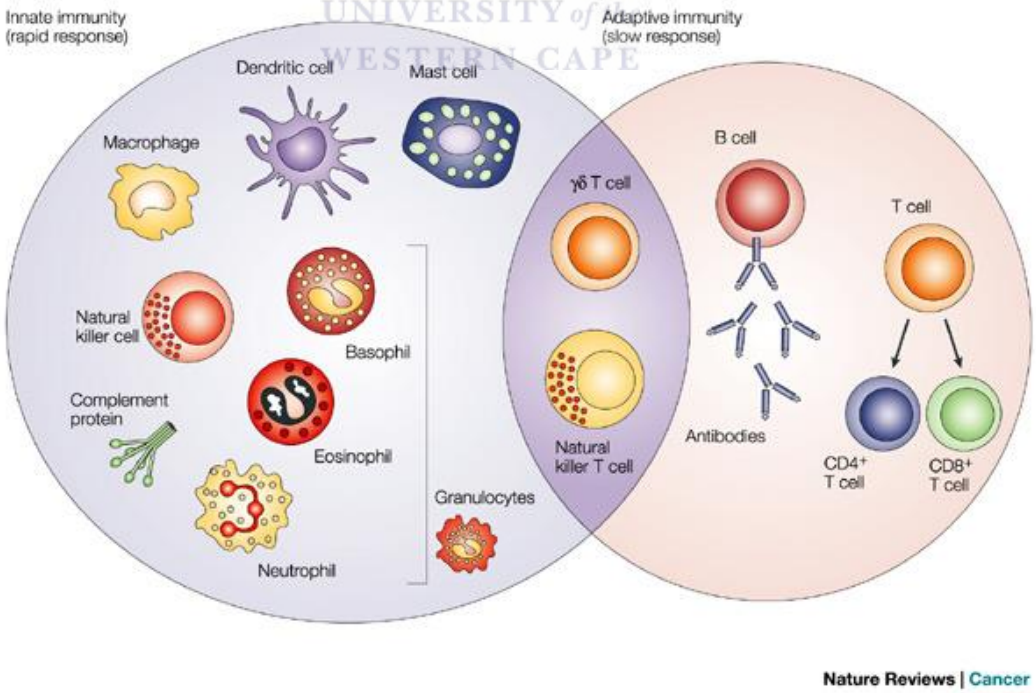

Figure 1: The immune system consists of two core components: the Innate immune system, and the Adaptive immune system. The Innate immune system provides a rapid, first-line response to invading pathogens, and the adaptive immune system responds progressively, with increasing memory for antigens (Dranoff, 2004)

The innate immune system plays the important role of producing responses against specialized classes of pathogens, via the activation of pattern recognition receptors, such as Toll-like receptors (TLRs) (Fairweather, 2007). Found on the membrane of monocytes, TLRs recognize and bind to specific microbial products called pathogen-associated molecular patterns (PAMPS), and subsequently trigger downstream signalling pathways to initiate inflammatory responses. To prevent excessive inflammation, various classes of negative regulators control TLR signalling, and recent studies have revealed that miRNAs both regulate and fine-tune innate immune responses to pathogenic insult (Dai and Ahmed, 2011). TLRs 7, 8, and 9 have been shown to be associated with Systemic Lupus Erythematosus (SLE) (De Azevedo Silva *et al.*, 2014)

8.1.1.2. Adaptive immunity

Cells of the adaptive immune system are activated for proliferation in response to a secondary response of stimulus, hence the synonymous name of acquired immunity. Arising from immune stem cells, lymphocytes are subdivided into B cell progenitors, and T cell progenitors. B cell progenitors produce antibodies upon stimulation by helper T cells (T_h) and thus fight extracellular pathogens. T cell progenitors mature into T_h and cytotoxic T cells (T_c), and B cell progenitors mature into plasma cells and memory cells. T_c cells release toxic compounds directed at the infected cell and thus fight intracellular pathogens (Murphy *et al.*, 2011). Regulatory T cells (Tregs) suppress the activation of effector T cells to maintain immune system homeostasis and tolerance to self-antigens. Various miRNAs have been shown to play critical roles in the regulation of T cell development and/or function (Dai and Ahmed, 2011).

8.1.1.3. Apoptosis

An intricate balance between the proliferation of new cells and the death of aged cells is required for tissue homeostasis. Dead or dying cells – due to age, exposure to toxicants, or due to other factors – require clearance from the functional component of the human body, by means of controlled degradation

and elimination by cells of the immune system. During normal functioning of the immune system, a dying cell may undergo one of several pathways of cell death: apoptosis, necrosis or autophagy.

The physiological process of apoptosis is considered to be a fundamental component of mammalian development of the embryo, the immune system and cell death induced by harmful agents (Elmore, 2007). There are three main processes of apoptosis: the extrinsic or death receptor pathway, the intrinsic or mitochondrial pathway, and the perforin/granzyme pathway. Two key pathways of apoptosis have been identified to date: a p53-dependent pathway and a p53-independent pathway. In the p53-dependent pathway, irradiation or drug-induced DNA damage (e.g. corticosteroids) may result in activation of this apoptotic pathway. Apoptosis can also be induced by ligand-binding of Fas or TNF receptors expressed on the surfaces of various cells. In some cells, a default death pathway governs cell survival and a hormone or growth factor is required to block the death pathway (Elmore, 2007).

During the early stages of apoptosis the cell shrinks considerably, resulting in dense packaging of the cytoplasm and organelles. Chromatin condensation, termed 'Pyknosis', is the most common identifiable feature of apoptosis. The apoptotic cell appears as a round or oval mass with dense nuclear chromatin fragments. In the subsequent stages of apoptosis, phosphatidylserine (PS) molecules usually embedded in the inner membrane of the cell, translocate to the outer membrane. The negative charges which these PS molecules carry, results in repulsion between adjacent PS molecules. These repulsive forces result in the observation of plasma membrane blebbing. During the process of plasma membrane blebbing (or budding), apoptotic bodies consisting of cytoplasm with tightly packed organelles with or without a nuclear fragment are formed, enclosed in an intact plasma membrane. Now present on the outer membrane, PS acts as "eat me" signals for circulating macrophages (Hochreiter-Hufford and Ravichandran, 2013).

Macrophages, parenchymal cells, or neoplastic cells engulf and digest these apoptotic bodies by means of phagolysosomal organelles. These macrophages are termed “tingible body macrophages” because these bodies are the fragments of nuclear debris within the apoptotic cell. These macrophages are commonly found within the lymphoid follicles or the thymic cortex (Elmore, 2007).

The process of apoptosis does not result in any inflammatory response being evoked for a number of reasons. Firstly, apoptotic cells contain their debris within the cell membrane, and thus do not release any debris into the surrounding tissues. Apoptotic cells are also rapidly phagocytosed by circulating macrophages, thereby preventing the onset of secondary necrosis. Lastly, macrophages engulf these apoptotic bodies without producing any cytokines which may promote inflammation (Elmore, 2007).

8.1.1.4. Secondary necrosis: Chronic inflammation

During the observation of cell death, there is often an overlap between the process of apoptosis and secondary necrosis. Injury-causing stimuli such as radiation, heat, hypoxia and cytotoxic anticancer drugs could induce apoptosis at low dosages, whereas higher doses could result in activation of a necrotic form of cell death (Elmore, 2007). Necrosis is an alternative cell removal process to apoptosis and is considered to be a toxic process as it involves swelling, and subsequent bursting of the cell (Hochreiter-Hufford and Ravichandran, 2013).

The process of necrosis is poorly controlled and progresses passively, usually affecting a large group of cells. In contrast, apoptosis is a tightly-controlled and energy-dependent process clearing specific cells (Elmore, 2007). The interplay between these two pathways has led to the description of an “apoptosis-necrosis continuum” for the shared biochemical network. An example of the relationship between these two pathways exists such that an ongoing apoptotic process will be converted into a necrotic process when there is a decrease in the availability intracellular adenosine triphosphate (ATP) and caspases (Elmore, 2007). Thus,

necrosis can be seen as an inflammatory activity that adversely affects the host, whereas apoptosis is a cell-clearing process aimed at protecting the host by removing damaged or infected cells.

8.1.2. Dysregulation of the immune system in SLE

Using both mouse and human models for lupus, numerous immunological abnormalities have been identified, including dysregulations observed in B - and T-lymphocyte functions, TLR functioning, NF-κβ signalling pathways, abnormal interactions among immune cells, and aberrations in the Type I IFN network (Dai and Ahmed, 2011). In addition to genetic loci which have been identified to be associated with the onset of SLE, dysregulation in epigenetic mechanisms such as histone modifications and DNA methylation have also been linked to the pathogenesis of lupus (Dai and Ahmed, 2011).

8.1.2.1. Impaired removal of apoptotic bodies

One hypothesis of the cellular processes driving the onset of SLE, is that the inadequate removal of apoptotic bodies results in the formation of self-antigens which are recognized by the immune system as "danger signalling" molecules (Tiffin *et al.*, 2013). Apoptotic cells which are ready for engulfment by phagocytes present numerous molecules on their surface, such as altered sugar molecules, ICAM3, and C1q. In response, phagocytic cells present lectins on their cell surface, which bind to sugar molecules on the apoptotic cell surface. CD14, and a LRP1/CD91 complex, are also present on the surface of phagocytic cells, binding to ICAM3, and C1q, respectively (Hochreiter-Hufford and Ravichandran, 2013).

In contrast, healthy, living cells express CD47 – an integrin-associated protein. CD47 inhibits the engulfment of the living cell by phagocytes. However, during apoptosis, the expression of CD47 is suppressed thereby allowing for the effective removal of the dying cells (Hochreiter-Hufford and Ravichandran, 2013). A study by Junker *et al.* described the critical findings of three miRNAs which targeted the CD47 protein in the lesions of patients with active multiple sclerosis (MS) (Junker *et al.*, 2009). Thus, it may be hypothesised that the miRNA-

mediated regulation of CD47 protein expression may have increased the engulfment of undamaged cells by phagocytes in these MS patients.

8.1.2.2. Autoantibody production

Autoantibodies can circulate within a patient for many years prior to the onset of an active autoimmune disease (Arbuckle *et al.*, 2003). These autoantibodies are produced in response to misrecognizing self-antigens as foreign bodies, thereby eliciting an immune response, and SLE is biochemically characterized by the presence of vast amounts of circulating autoantibodies (De Azevedo Silva *et al.*, 2014).

The production of anti-nuclear antibodies (ANAs) often results in a systemic end-organ damage phenotype caused by the immune complex (Apostolidis *et al.*, 2011). Numerous miRNAs have been described to influence the production of various autoantibodies, including anti-ds DNA antibody, anti-histone antibody, and anti-ribonucleoprotein antibody (Zan *et al.*, 2014). Lupus patients commonly test positive for ANAs, whereas anti-Smith and anti-ds DNA autoantibodies have demonstrated a high degree of specificity amongst lupus patients (Apostolidis *et al.*, 2011).

A phenomenon known as 'inflammaging' first described in 2000, was coined to describe the observations of a loss of protective immunity and an increase in low-grade chronic inflammation, accompanying the process of ageing. An increase in ANAs has also been observed with a progression in age (Jeffries and Sawalha, 2011). Late-onset SLE, in which patients are 50 years old and above at first diagnosis, has been noted to be particularly severe in terms of the morbidities associated with the disease (Pons-Estel *et al.*, 2010). Thus, the increased production of ANAs with progressing age may result in a more severe manifestation of lupus.

8.1.2.3. Interferon signalling

The pathogenesis of SLE is largely characterised by the production of ANAs against various cell types, as well as a sustained immune activation of the Type 1

interferon (IFN-I) pathway. The sustained response of the IFN-I pathway results in the upregulation of numerous genes which respond to IFN-I and Interferon-α-inducible cytokines, and central to this response are plasmacytoid dendritic cells (PDCs) which are activated during lupus flare to rapidly secrete IFNα (De Azevedo Silva *et al.*, 2014). A key event in the onset of SLE is nucleic acids binding to TLRs. This binding results in the activation of plasmacytoid dendritic cells – an event which links the adaptive and innate arms of the immune system. Specific miRNA's have been identified to contribute to the IFNα regulation of expression by plasmacytoid dendritic cells, IFNα sensitivity, and receptor expression for cytokines in target cells (Carlsen *et al.*, 2013).

8.2. Systemic Lupus Erythematosus

Systemic Lupus Erythematosus (SLE or lupus) is a chronic autoimmune disease, characterised by the production of autoantibodies against degraded nucleic acids (DNA and RNA), and associated proteins. During periods of lupus-flare (or disease state) inflammation of the skin, central nervous system, kidneys, lungs and joints may be experienced by the patient. The onset of active lupus is largely multifactorial and varies between patients. However, it is generally accepted that environmental- and genetic factors, and sex hormones may be responsible for the onset of the disease (Tiffin *et al.*, 2013).

8.2.1. Prevalence

Lupus predominantly affects women of child-bearing age, with the ratio of females to males affected with SLE estimated to be as high as 12:1 (Aggarwal *et al.*, 2010). However, the prevalence of SLE has been approximated at eight females affected for every male affected, per 100 000 individuals in the United Kingdom (UK) (Tiffin *et al.*, 2014).

Women of African-American origin have shown to be three to four times more affected with SLE than Caucasian women (Kyttaris, 2010). However, the highest recognized incidence of SLE is amongst women of Afro-Caribbean origin (Danchenko *et al.*, 2006). Women of African-American ethnicity are more

commonly affected by lupus, followed by Hispanic, Asian, and lastly Caucasian women, as observed in the United States (Fessel, 1974; Weckerle and Niewold, 2011). Racial admixture has also been suggested to increase the risk of developing SLE (Molokhia *et al.*, 2001; Tikly and Navarra, 2008).

In the UK, the incidence of SLE has been estimated at 25 people affected for every 100 000 individuals in the country. An investigation into the epidemiology of lupus revealed that the highest incidence of lupus was found in the United States of America (USA) and France, whilst the lowest incidence was found in Iceland and Japan (Jeffries and Sawalha, 2011). These observations suggest possible genetic and environmental factors influencing the geographical incidence of lupus; although socio-economic factors and access to health care may also affect incidence rates in different geographical groups (Danchenko *et al.*, 2006).

8.2.2. Diagnosis

Lupus is a multi-systemic disorder affecting populations of varying ethnicities and socio-economic statuses. The disease is characterised by, and often diagnosed by (but not restricted to) inflammation of various tissues throughout the body. The aberrant activation of autoreactive T- and B cells leads to the production of pathogenic autoantibodies against the nuclei and phospholipid antigens of various cells, and their binding proteins, leading subsequently to tissue injury (Choi *et al.*, 2012).

The procedure for diagnosing a patient with SLE is a complex process involving the close observation of both physical changes, and biochemical changes within patients. Numerous scoring systems have been developed, and used over the years. The SLEDAI (SLE Disease Activity Index) score is one of the most commonly used scoring systems, rating the state of disease activity on a scale of 0-10. A weighted index across nine organ systems determines the outcome of the SLEDAI score (Bombardier *et al.*, 1992). In 1971, the American Rheumatism Association developed the first set of 11 criteria used in identifying, and

diagnosing active SLE, in which a minimum of four out of the 11 criteria had to be met (Pons-Estel *et al.*, 2010). With the advancements made in medical technology, and the unveiling of new mechanisms involved in diseases processes, revision of the criteria was conducted in 1982, and again in 1997. Lastly, the Systemic Lupus International Collaborating Clinics (SLICC) criteria was derived and validated in a recent study by (Petri *et al.*, 2013). The SLICC criterion for SLE assessment makes use of 11 clinical criteria, and six immunologic criteria, of which at least four criteria must be met for the diagnosis of active SLE.

There is also a likelihood for the under diagnosis, and over diagnosis of the disease using these scoring methods. Due to the gold standard (SLEDAI) being set at four out of the 11 criteria observed in order to make a positive diagnosis, under diagnosis may occur when any mild cases present, in which less than four of the criteria are clinically and/or biochemically observed. Over diagnosis of the disease is also possible if clinicians only use the presence of ANAs in the blood sample as an indicator for active disease (Liu *et al.,* 2013). The degree to which SLE affects African populations is largely unknown due to a variety of reasons. These include varying qualities of available health care, under-diagnosis, and poor recognition of the disease, making lupus a supposed “rarity” in southern Africa and other tropical regions. The disparity of disease prevalence amongst different races is possibly due to poor diagnostic tools available in different geographical regions, and poor disease recognition due to the contrast in phenotype of any two patients diagnosed with SLE (Tiffin *et al.*, 2014).

8.2.3. Onset of SLE

A self-antigen is an antigen, as a constituent of the normal bodily tissues, which provokes the response of the innate immune system to produce antibodies against it, and are therefore termed autoantibodies. The general loss of recognition and tolerance for auto-antigens, as the key cause for the inflammation observed during lupus-flare, may occur as result of various factors including environmental, hormonal, and epigenetic factors.

Many checkpoints exist to maintain immunological tolerance at both central and peripheral lymphoid organ systems and mechanisms of these checkpoints are aimed at ensuring that randomly autoreactive B and T lymphocytes are either silenced or deleted. However, in exceptional cases, where autoimmunity is imminent, these auto-reactive lymphocytes manage to escape the checkpoints and survive in peripheral lymphoid tissues. Once activated, these autoreactive immune cells launch a detrimental attack on normal tissues, resulting in the onset of an autoimmune disease such as lupus (Dai and Ahmed, 2011).

Various hormones including estrogens, androgens and dihydrotestosterone have been shown to influence the risk for developing SLE (Jeffries and Sawalha, 2011). A prominent observation in many autoimmune diseases in both animal models and humans is that females are more susceptible to autoimmune diseases than their male counterparts, and more than 80% of patients suffering from scleroderma, thyroiditis, SLE, and Sjögrens syndrome are women. Thus, in addition to the genetic factors associated with autoimmune disease susceptibility, sex hormones such as androgens and estrogens are believed to play an important role in the development of this subset of diseases (Dai and Ahmed, 2011).

B cell tolerance has been shown to be broken with estrogen treatment, even in normal, non-autoimmune prone mice, resulting in the production of several lupus-related autoantibodies including those against cardiolipin, phospholipids and double-stranded DNA. Two investigations into the influence of estrogens on lupus onset revealed contradictory results. The Nurses' cohort studies reported that the use of hormone replacement therapy (HRT) and oral contraceptive pills (OCP) increased the risk for developing SLE (Bernier *et al.*, 2009). In contrast, studies by the Safety of Estrogen in Lupus: National Assessment (SELENA) trials on using HRT's and OCP's reported that the use of OCP has no significant effect on global lupus disease activity and incidence of flares, and the use of HRT is only

associated with a small risk for mild and moderate lupus flare, but not severe flare (Bernier *et al.*, 2009; Petri *et al.*, 2005; Sanchez-Guerrero *et al.*, 1995).

Endogenous retroviruses and retroviral elements are found in abundance throughout the genetic codes of vertebrates, and may constitute up to 40% of the human genome. Numerous studies have implicated the retroviral group in the pathogenesis of SLE, as viral components have been found in the sera and organs of lupus patients. Retroviruses are particularly implicated in SLE by means of the detection of the p30 gag protein in renal glomeruli of SLE patients. Along with this, SLE patients have been reported to demonstrate reactivity to the p30 gag antigen (Francis and Perl, 2010). It has been illustrated that retroviruses may contribute to the onset of autoimmune diseases by either directly encoding reactive auto-antigens, or by altering the expression of immune-regulatory genes (Jeffries and Sawalha, 2011)

The Human T cell lymphotrophic virus 1 (HRES-1) and Endogenous Retrovirus Group 3 (ERV-3) genes are endogenous retroviral auto-antigens which have been identified in lupus patients. HRES-1 has been found in 21-52% of lupus patients, while ERV-3 has been found in up to 32% of lupus patients. The expression of endogenous retroviral sequences has in part been shown to be regulated by DNA methylation processes previously discussed (Jeffries and Sawalha, 2011).

Infection with the Epstein-Barr virus (EBV) affects more that 90% of the world's adult human population and as with other members of its gamma herpes virus family, EBV typically results in a lifelong, mostly asymptomatic infection in its host (Cameron *et al.*, 2008; Dolcetti and Masucci, 2003). The latent membrane protein 1 (LMP1) gene is the latency viral gene most frequently implicated in tumour formation due to EBV-associated malignancies. LMP1 is a six-transmembrane, active signalling molecule that functionally mimics members of the cellular tumour necrosis factor receptor (TNFR) family (Cameron *et al.*, 2008).

Among others, LMP1 regulates the interferon type I (IFN alpha/beta) pathway, which is partially induced by IFN regulatory factor 7 (IRF7). The BIC transcript,

which encodes miR-155, was shown to be upregulated in B cell lymphomas, including lymphoproliferative diseases associated with EBV infection (Cameron *et al.*, 2008). EBV infection was also shown to induce the expression of miR-155 and miR-146a in a NF-kB dependent manner (Dai and Ahmed, 2011). Dysregulation of miR-155, and miR-146a expression as result of EBV infection was shown to be associated with SLE (Dai and Ahmed, 2011; Tang *et al.*, 2009). It has also been reported that the anti-Smith antigen, a biomarker highly specific for SLE disease flare, is mimicked by the EBV-nuclear antigen-1 (Jeffries and Sawalha, 2011).

8.2.4. Clinical presentation and treatments

The clinical manifestations often observed for SLE are a butterfly rash (across the cheekbones and forehead), as well as the appearance of oral ulcers, and an unusual sensitivity to light (De Azevedo Silva *et al.*, 2014). Diseases which may occur secondarily to the onset of SLE include pericarditis, arthritis, pleuritis, neural abnormalities including seizures and psychosis, and kidney disease, including nephritis (Maidhof and Hilas, 2012; Weckerle and Niewold, 2011).

Infection with malaria-carrying parasites has also been suggested to contribute the onset of SLE. The low reported incidence of SLE occurrence in tropical regions of Africa has resulted in numerous scientific literature reviews suggesting that malaria infection confers some form of protection against the onset of SLE (Borchers *et al.*, 2010; Gilkeson *et al.*, 2011; Tikly and Navarra, 2008) . It is hypothesised that malaria, and other infections caused by parasites may alter the natural immune response, thereby conferring a form of tolerance to the onset of SLE (Borchers *et al.*, 2010). An infection predominantly occurring in tropical regions, malaria has been shown to protect individuals in these regions against developing SLE and there are two main mechanisms proposed to explain these observations (Clatworthy et al 2007).

First, there is a hypothesis that infection with malaria confers some form of resistance due to the activation of the production for the tumour necrosis factor (TNF) cytokine. The activation of macrophages and neutrophils by TNF results in

a rapid response to the control of the replicating *Plasmodium* parasite during the early stages of the disease (Francis and Perl, 2010). It is proposed that when infected with the malaria parasite, patients who are asymptomatic for the disease produce high concentrations of reactive nitrogen intermediates by means of activated macrophages. These reactive nitrogen species are thereby proposed to minimise autoreactive T cell proliferation , and subsequently reduce the chance of onset of autoimmune diseases (Clark *et al.*, 1996).

A second hypothesis is that resistance to SLE results from the activation of the FcγRIIb receptor. It is hypothesised that the activation of this receptor by circulating auto-antigens may result in the hyper-activation of T- and B immune cells, a typical observation made in lupus patients (Choi *et al.*, 2012).

Anti-malarial drugs have been used for decades to treat rheumatic diseases, including SLE and rheumatoid arthritis (RA) (Rynes, 1997). The mechanism of effectivity has not yet been well-described. Treatments include the administration of hydroxychloroquine, steroids, and cytotoxic drugs such as cyclophosphamide and methotrexate (Mishra, 2008; Wakeland *et al.*, 2001).

Table 1: An extensive review of genes associated with numerous SLE phenotypes (De Azevedo Silva *et al.*, 2014)

Clinical Manifestations and Susceptibility Genes			
SLE Features	**Definition**	**Gene**	**Referencies**
Skin Lesions	A rash on the face (cheeks and nose), often in a butterfly shape	*IL-10, RASGRP3, VDR, CTLA-4, XRCC1, STK17A, TYK2.*	[24, 78, 100, 105, 111, 114, 119, 146]
Photosensitivity	Reaction to sunlight leading to a new rash or worsen a present one	*XRCC1, TYK2.*	[111, 114, 146].
Oral ulcers	Ulcers inside the mouth or naso-pharyngeal ulcers.	*STAT4.*	[159].
Arthritis	Joint pain and swelling of two or more joints	*TNFSF4, VDR, STK17A.*	[69, 74, 100, 119].
Serositis	Pleuritis or pericarditis	*RASGRP3.*	[78].
Renal Disorders	Persistent protein or cellular casts in the urine	*PDCD1, PTPN22, TNFSF4, VDR, CTLA-4, IRF5,TLRs, STAT4.*	[33, 34, 45, 49, 73, 74, 96, 99, 105, 142, 153, 154, 159].
Neurologic disorder	Seizures or psychosis	*IL-10, XRCC1, XRCC3, XRCC4.*	[21, 25, 107].
Hematological Disorders	Anemia, leukopenia, lymphopenia or thrombocytopenia	*FYB.*	[61].
Immunologic disorder	Positivity for anti–dsDNA, anti-Sm, or antiphospholipid antibodies	*IL-10, PRL, FYB, TNFSF4, RASGRP3, BANK1, VDR, XRCC1, XRCC3 and XRCC4, STK17A, IRF5, STAT4.*	[21, 26, 27, 39, 61, 65, 78, 84, 88, 100, 107, 119, 139, 159].
Disease´s Severity and activity	Active phase modulation with symp-toms worsening	*IFIH1, VDR, IRF5 and STAT4.*	[67, 100, 138, 159].

The prevalence of SLE has been demonstrated with up to 10% of SLE patients having a first degree relative with the disorder, as well as up to 20% of monozygotic twins showing disease concordance (Aggarwal *et al.*, 2010). SLE has been well-described as an autoimmune disease of polygenic aetiology. Biomedical research has focused on numerous aspects of SLE in an attempt to find therapeutic targets to alleviate the morbidity, and mortality caused by this autoimmune disease. Through multiple research studies, many genes in the mammalian genome have been identified to play a role in numerous biological processes dysregulated during the active disease state. A recent review by De Azevedo Silva *et al.* tabulated the genes (see Table 1) found in various biomedical research studies, to be associated with defining phenotypes experienced by lupus patients in the active disease state.

8.2.5. Epigenetic mechanisms contributing to the onset of SLE

Epigenetics can be described as the biological study of heritable alterations in gene-expression patterns that do not alter the underlying DNA sequence. Three mechanisms most frequently involved in epigenetic control of gene expression are DNA methylation, histone modifications and the involvement of RNA interference (RNAi). These mechanisms, along with other epigenetic factors such as pharmaceutical drugs, infection with viruses, and endogenous hormones are also believed to influence the onset of SLE.

8.2.5.1. DNA methylation

A well-known example of epigenetic modification of the mammalian genome is the DNA modifications at the 5th position carbon in the pyramidine ring, which leads to the formation of 5-methylcytosine (5mC) (Balada *et al.*, 2007). DNA methylation processes are catalysed by DNA methyltransferases (DNMTs), facilitating the regulation of chromatin modelling and transcriptional silencing (Tu *et al.*, 2015). DNA methylation is a specific mechanism of epigenetic modifications in which a methyl group is added to the 5' group of cytokines within the Cytosine-phosphate-Guanine (CpG) dinucleotide sequence. In

mammals, DNA methylation is performed by *de novo* methyltransferases, namely DNMT3a and DNMT3b and is maintained by DNA Cytosine-5-Methyltransferase 1 (DNMT1) which has an affinity for hemi-methylated daughter strands produced during DNA replication (Jeffries and Sawalha, 2011). DNMT2 functions to methylate RNA (Jeffries and Sawalha, 2011). CpG dinucleotides are clustered into islands, often at the promoter regions of genes where transcription factors bind to initiate replication. These islands tend to be protected from methylation but when methylation does occur at the CpG island, there is a silencing of gene expression. DNA methylation is a mitotically heritable modification and is thought to be irreversible, except by failure of DNMT1 to maintain methylation. Active DNA methylation appears to play essential roles in a variety of cellular responses to environmental stimuli, including hypoxia, hormonal signalling, viral latency and reactivation, and X-chromosome inactivation (Jeffries and Sawalha, 2011). It has also been found that T cells isolated from SLE patients exhibited lower levels of methylated DNA (Balada *et al.*, 2007)

Early studies into the epigenetic mechanisms associated with lupus showed that $CD4^+$ T cells treated with 5-azacytidine (5-azaC), a DNA methylation inhibitor, lost antigen selectivity properties and responded to the presentation of self-antigens. Thus, it was suggested that drugs causing DNA demethylation resulted in the onset of autoimmune-like diseases (Jeffries and Sawalha, 2011). Between 15 000 and 30 000 cases of Drug-induced Lupus (DIL) are reported in the United States each year. DIL has shown decreased bias in women, and in African-Americans, when compared to idiopathic SLE (Borchers *et al.*, 2007).

Procainamide and hydralazine both resulted in the production of positive antinuclear antibodies in patients taking them, thereby resulting in the onset of a lupus-like syndrome in a genetically-susceptible subset of patients. Both drugs were found to inhibit deoxyribonucleic acid (DNA) methylation and induce murine lupus, however, the mechanism used was distinct for both drugs. Procainamide demonstrated to act as a competitive antagonist in the

maintenance of DNA methylation by DNMT1, whereas Hydralazine was found to inhibit the ERK pathway by blocking signalling through protein kinase C δ (PKC δ) (Jeffries and Sawalha, 2011).

DNA demethylation (hypomethylation) is a process in which a methyl group is removed from the DNA sequence, and is shown to occur in early developmental stages such as during primordial germ cell development. DNA demethylation processes can also be catalysed by a family of enzymes known as the ten-eleven translocation (TET) enzymes. TET enzymes have shown involvement in DNA demethylation processes via the oxidation of 5-mC (Kohli and Zhang, 2013). T cells isolated from SLE patients have demonstrated elevated concentrations of TET1 and TET2 protein expression, which subsequently is likely to affect DNA demethylation, thereby altering gene expression (Tsokos, 2016).

For a parent cell to faithfully copy DNA methylation patterns to the daughter cell, DNMT1 requires a source of methyl group donors, specifically *S*-adenosylmethionine (SAM), to perform the reaction needed to methylate a CpG island. A study by Liu *et al.* demonstrated that a restriction of methionine via supplementation with homocysteine resulted in functional hypomethylation *TNFSF* (Liu *et al.*, 2013). By examining T cells of healthy adults over the age of 50, it was noted that there was a reduced level of methylation of the *KIR* gene family and *TNFSF* (Liu *et al.*, 2013). The concentration of DNMT1 in T cells declines notably with age and studies have shown that T cells become hypersensitive to their environmental micronutrient levels, and when methyl donors become restricted, may result in an up-regulation of methylation-sensitive genes. It is also observed that DNA methylation is globally reduced in T cells of lupus patients, as in that of ageing T cells (Jeffries and Sawalha, 2011).

A secondary mechanism by which DNA methylation could lead to gene silencing is when methylated CpG molecules prohibit transcription factor binding and thus directly alter gene expression (Liu *et al.*, 2013; Ramos *et al.*, 2010). Over-expression of specific hypomethylated genes are often observed in lupus

patients. In T cells, these genes include ITGAL (*CD11a*), CD40LG (*CD40L*), TNFSF7 (*CD70*), KIR2D44, PRF1 (*perforin*), and in B cells, CD5 (Jeffries and Sawalha, 2011).

Ultraviolet (UV) radiation – more specifically UVB, can induce the expression and secretion of various interleukins. These include IL-1 and TNF-α in keratinocytes, mast cells, and Langerhans cells, and these cells may function in recruiting dendritic and T cells in the mounting of an immune response. When compared to baseline controls, lupus patients exhibited significantly reduced global DNA methylation levels when exposed to both moderate and high doses of UVB radiation. This observation is independent of DNMT1 levels, which remained stable after exposure to UV radiation. Healthy controls only demonstrated significant reduction in DNA methylation when exposed to high doses of UVB radiation (Jeffries and Sawalha, 2011). It was also found that following exposure to UV light, an increase in the expression of Gadd45A resulted in a reduction in genomic DNA methylation – a process counteracted in part by the knockdown of Gadd45A. Early knock-out studies of *Gadd45a* revealed a protective role as exerted by the gene product Gadd45A. Knockout of *Gadd45a* in mice resulted in the development of an SLE-like autoimmune disorder (Hughes and Sawalha, 2011).

Early epigenetic studies of SLE onset revealed that the treatment of $CD4^+$ T cells with the DNA methylation inhibitor 5-azaC resulted in a loss of antigen selectivity and an increased response to self-antigens (Patel and Richardson, 2010). It has been well-established that a core mechanism in lupus pathogenesis, in comparison to healthy individuals, is the impaired clearance of apoptotic DNA, resulting in the presence of self-antigens. Isolation of DNA-anti-DNA antigen-antibody complexes from lupus patients were exactly the correct size to have resulted from the extracellular breakdown of apoptotic chromatin. Hypomethylation of either normal or necrotic DNA due to treatment with 5-azaC greatly increased its potential to stimulate an autoimmune response. This selective response of 5-azaC to hypomethylated DNA is likely initiated by TLR 9

signalling (Jeffries and Sawalha, 2011). DNA demethylation is a factor known to influence lupus progression, and thus a greater level of demethylation may be correlated with increased levels of disease activity.

8.2.5.2. Histone modifications

Histone modifications by means of acetylation and methylation have been described in the aetiology of some lupus cases, and are another example of an epigenetic factor which may influence the onset of SLE (Patel and Richardson, 2013). Histone modifications such as an increased global H4 acetylation in monocytes of SLE patients are often observed (Jeffries and Sawalha, 2011; Patel and Richardson, 2010; Wu *et al.*, 2015). It has been hypothesised that over-expression of the miRNA miR-101 may contribute to the decrease in repression observed in the CD4+ T cells of lupus patients (Hughes and Sawalha, 2011). Furthermore, study by Sun *et al.* 2008 also revealed a close association between the production of anti-histone antibodies and the disease activity index of SLE patients tested (Sun *et al.*, 2008).

8.2.6. MicroRNAs in SLE

MicroRNA's (miRNAs) are small, non-protein-coding RNA molecules ranging between 20 and 25 nucleotides in length (Alevizos and Illei, 2010a). Previously termed to be part of the "junk DNA" of the human genome, miRNAs have proven to be considerable contributors to the transcriptional and post-transcriptional regulation of gene expression. MiRNAs perform these functions by inducing translational repression on the target messenger RNA (mRNA) or by directly facilitating the degradation of the mRNA strand. The first miRNA identified was *lin-4* in the nematode *Caenorhabditis elegans*, and was implicated in mechanisms controlling the progression and timing of this organism's life cycle (Lee *et al.*, 1993.

More than 1000 human miRNAs have since been identified to be able to influence the translation of mRNA's, thereby depicting a strong role of miRNAs in gene regulation (Dai and Ahmed, 2011). The duality of miRNA-based translation

regulation is demonstrated by a single miRNA being able to bind to and regulate multiple mRNA's in one instance; while in another, many miRNAs may collectively perform regulatory activities on a single mRNA (Alevizos and Illei, 2010a; Lee *et al.*, 2004). An analysis of 72 lupus susceptibility genes using computational prediction analyses revealed that 71 out of those 72 genes examined contained at least one miRNA target site for more than 140 miRNAs (Vinuesa *et al.*, 2009).

8.2.6.1. Biogenesis

The biogenesis of miRNAs involves intricate processing in both the nucleus and the cytoplasm. Genomic miRNA is transcribed into primary-miRNA (pri-miRNA) by RNA polymerase II. Pri-miRNA transcripts vary from 100's to 1000's of nucleotides in length and may contain a single or several hairpin-loop structures (Lee *et al.*, 2004). Pri-miRNA may then follow either a canonical or non-canonical pathway of nuclear-cytoplasmic processing. In the canonical pathway, pri-miRNAs undergo nuclear cleavage by the RNAse III enzyme, Drosha, and the double-stranded RNA-binding protein, DGCRB. This cleavage produces nuclear, hairpin-shaped precursor-miRNAs (pre-miRNAs) (Dai and Ahmed, 2011; Lee *et al.*, 2004). In some instances, the primary transcript may contain several hairpin structures, and will subsequently result in different miRNAs being produced via transcription of the different hairpins. These pri-miRNA transcripts are termed "polycistronic" and each resultant miRNA is given a unique name (Ambros *et al.*, 2003).

In the non-canonical pathway of processing, genomic miRNA is directly translated by RNA polymerase II into mirtrons which lack a lower stem-loop structure, as well as the single-stranded flanking segments which are essential for processing by the Drosha RNAse III enzyme. Mirtrons therefore do not undergo processing by Drosha and Dicer, but undergo splicing by a spliceosome to produce branched precursor-mirtrons (pre-mirtrons). Pre-mirtrons then

undergo lariate-mediated debranching to produce pre-miRNAs as a result of the non-canonical pathway of processing (Dai and Ahmed, 2011).

Following both pathways of processing, pre-miRNAs are exported from the nucleus via the nucleoplasmic shuttle protein, Exportin-5. Pre-miRNAs are processed by the RNAse III enzyme, Dicer, to produce miRNA/miRNA* duplexes of approximately 22 nucleotides long, with imperfectly matched sequences. The miRNA/miRNA duplex is loaded into the Argonaute protein, forming an RNA-induced silencing complex (RISC). Within the RISC complex, the passenger miRNA (miRNA*) is degraded by the RISC complex, whilst the guided miRNA strand remains localized within the RISC complex in the form of mature miRNA (Dai and Ahmed, 2011).

The strand which is to become a mature miRNA (guided strand) is determined by the relative internal stability of the two ends of the duplex, and the strand with a less stable 5' end (e.g. the strand with a G:U pair versus a G:C pair at the 5'end) will survive preferably over the passenger strand which undergoes degradation by an unknown nuclease (Lee *et al.*, 2004). This mature miRNA is capable of interacting with the 3' untranslated region (3' UTR) of its target mRNA, thereby inducing either its translational repression or degradation activities, possibly by transporting the complex to a perinuclear compartment referred to as GW or P bodies, and thereby sequestering the mRNA away from the Golgi apparatus (Cameron *et al.*, 2008; Dai and Ahmed, 2011).

Many human miRNA molecules have been shown to exert their regulatory effects by having perfect or imperfect complementarity to the 3'UTR of the mRNA strand. For imperfect complementarity to be successful, target matching between the miRNA and the target mRNA needs to occur perfect. Starting from the 5' end of the miRNA, this matching occurs from the 2nd through to the 7th nucleotides (Alevizos and Illei, 2010b). In contrast, most plant miRNAs exert their regulatory functions through binding sites located in the exons of target genes (Ehrenreich and Purugganan, 2008).

8.2.6.2. miRNA action as a mechanism of RNA interference (RNAi)

Transcriptional gene silencing of newly transcribed mRNAs can be performed by a number of biological processes, including RNA interference (RNAi). Central to this mechanism of RNAi, is the binding of a dsRNA to a target mRNA, resulting in translational inhibition of the target protein (de França *et al.*, 2010). RNA interference is a mechanism of gene regulation performed by non-coding small RNA's, including miRNAs. In SLE, miR-21, miR-148a and miR-126 have been shown to perform these regulatory activities on the expression of certain proteins by targeted mRNA translational repression or degradation (Jeffries and Sawalha, 2011; Wu *et al.*, 2015).

8.2.6.3. miRNA nomenclature

MiRNA naming should follow a standard procedure as aptly described in Ambros *et al.* 2003 and Griffiths-Jones *et al.* 2006. Standard naming procedures of experimentally-validated miRNAs are meant to be employed prior to the publication of the author's discovery. The naming procedure should follow several steps which will be described here.

The use of 'miR-' indicates a mature miRNA sequence, whereas 'mir-' indicates a precursor hairpin miRNA which has not yet given rise to a mature miRNA (Wright and Bruford, 2011). The three letters preceding 'miR' or 'mir' and joined by a dash '-' to indicate the species from which the miRNA is derived, e.g. 'hsa-' is used for human miRNAs, 'mmu-' for mouse, and 'oar-' for sheep. Other prefixes include the addition of a 'v' for viral miRNAs and a 'd' for miRNAs of the *Drosophila* fruit fly species (Griffiths-Jones *et al.*, 2006).

The prefix 'miR-'or 'mir' is followed by a numerically assigned value, indicating the order in which the miRNA was discovered. For example, 'miR-101' was discovered and named prior to 'miR-456'. Pre-miRNAs and pri-miRNAs which have different loci in the genome but lead to the production of an identical mature miRNA are given an additional dash-numbered suffix. An example of such

a case is with miRNAs 'hsa-mir-194-1' and 'hsa-mir-194-2' (Griffiths-Jones *et al.*, 2006).

Numerous studies have shown that many miRNA sequences are evolutionarily conserved (Ambros *et al.*, 2003; Park *et al.*, 2002). Orthologous miRNA sequences are given the same numeric suffix, but are differentiated by the species prefix, e.g. 'hsa-miR-101' and 'mmu-miR-101'. In contrast, paralogue miRNA sequences which are very near to being completely identical (i.e. the sequences only differ by one or two nucleotides), are annotated with an additional lettered suffix, e.g. 'mmu-miR-101a' and 'mmu-miR-101b' which are paralogues (Griffiths-Jones *et al.*, 2006).

Lastly, when two mature miRNAs originate from opposite arms of the same precursor hairpin miRNA, and both mature miRNAs are produced in a relatively equal abundance, these mature miRNAs are denoted by a '-3p' or a '-5p' (Griffiths-Jones *et al.*, 2006). However, in nature one of the mature miRNAs produced is often in a greater abundance. In this case, the less abundant miRNA is indicated by an asterisk (*) (Ambros *et al.*, 2003).

8.2.6.4. Other causes of miRNA dysfunction in SLE

Genetic and environmental factors may contribute to lupus pathogenesis through altering the expression and/or function of miRNAs. The genetic polymorphisms in a miRNA gene and the 3'UTR sequences of its target genes, gene translocation, gene amplification or deletion in a host genome can all contribute to dysregulated miRNA expression and/or function in diseases. For example, a single nucleotide polymorphism (SNP) in pre-miR-146a decreased the expression of mature miR-146a, which has been shown to be related to the genetic predisposition to papillary thyroid carcinoma (Alevizos and Illei, 2010a). Pre-miR-146a is known to specifically target the *IRAK1* gene (which encodes the interleukin-1 receptor-associated kinase 1 protein) and *TRAF6* gene (which encodes the TNF-receptor-associated factor 6 protein) (Alevizos and Illei, 2010a). Both *IRAK1* and *TRAF6* are known lupus-associated genes (Tiffin *et al.*, 2013).

Estrogen has been shown to exert both anti-inflammatory and pro-inflammatory immune responses in different studies and it was found that estrogen not only modulated key transcription factors involved in inflammation, but also regulated miRNA expression to promote inflammatory responses in activated spleen cells from estrogen-treated mice. Using a combination of high throughput microarrays and Realtime RT-PCR as a confirmatory test, a number of miRNAs were identified to be dysregulated in splenocytes from estrogen-treated orchi-ectomised male C57BL/6 mice when compared to placebo controls (Dai and Ahmed, 2011).

In a study by Dai and others, it was found that miR-451, miR-486, miR-223, miR-148a, miR-18a, and miR-708 were up-regulated in response to estrogen treatment. They also found that miR-146a, miR-125a, miR-125b, miR-143, miR-145, *let-7e*, miR-126, miR-181a were down-regulated in response to treatment with this sex-hormone which plays a critical role during the onset of puberty in females (Dai *et al.*, 2008). Down-regulation of miR-146a and up-regulation of miR-223 was found to contribute to enhanced interferon gamma (IFNγ) production in LPS activated splenocytes from estrogen-treated mice (Dai *et al.*, 2008). The decreased levels of miR-146a and miR-125a, and increased level of miR-148a have been identified in human lupus patients and reported to contribute to lupus pathogenesis by regulating the type I IFN pathway, the inflammatory chemokine RANTES and DNA methylation, respectively (Pan *et al.*, 2010; Tang *et al.*, 2009; Zhao *et al.*, 2010).

8.2.6.5. miRNAs dysregulated in SLE

Numerous miRNAs have been identified to be associated with the onset and progression of SLE. The critical regulatory functions of miRNAs - by means of direct binding to the target gene - results in decreased production of the intended protein. This potent regulatory mechanism has implicated numerous miRNAs in biological pathways which are dysregulated during lupus flare. A recent review by Zan *et al.* summarised the functions of miRNAs implicated through various studies with lupus pathogenesis.

The *DNMT1* gene was reported to be a target for miR-21, miR-126, and miR-148a. MiR-155 was reported to target *CD62L* and *SHIP1*, regulating Treg phenotype and B cell activation, respectively. MiR-23b was reported to target IL-17 and *Blimp1*, suppressing inflammation in autoimmunity, and regulating the production of antibodies, respectively. MiR-146a was reported to target both *TRAF6* and *IRAK1*, and as a consequence, negatively regulated the activities of *NFKβ* (Zan *et al.*, 2014). Numerous other miRNAs have been described in their functioning as regulators of target genes which have been shown to be critical in the onset and progression of SLE (Carlsen *et al.*, 2013; Lu *et al.*, 2012; Pan *et al.*, 2010; Stagakis *et al.*, 2011; Tiffin *et al.*, 2013).

8.2.6.6. Methods used to identify dysregulated miRNAs

Using microarray assays, numerous miRNAs have been found to be dysregulated in lupus. It has however been proposed that these results require further verification using Real-Time RT-PCR and Northern blotting laboratory techniques. It has been demonstrated that results from one study showing dysregulated miRNAs in lupus, is not often reproduced in follow-up studies. One example of this is that Dai *et al.* reported a contradiction in the expression levels of hsa-miR-146a in PBMCs of lupus patients, between their study and the study performed by Tang *et al.* (Dai *et al.*, 2007; Tang *et al.*, 2009). The discrepancies in the results retrieved from these human lupus studies may be due to a number of variable factors. It may be due to the type of samples collected, and the subsequent storage methods, or it may be due to the sensitivity of the methods used to detect the miRNAs. The differences in the results may also be due to the variability of patient health histories, the length of the disease state, the severity of the disease at the time of collecting the sample, as well as the ethnic race of the patient (Dai and Ahmed, 2011; Yan *et al.*, 2014).

8.3. Introduction to Methods

8.3.1. Databases used to verify the IDs of miRNAs of interest

8.3.1.1. miRBase

The miRBase database is a searchable, freely-accessible online database containing curated miRNA sequences, their annotation, standardised nomenclature, and target prediction. Numerous versions of the miRBase database have been released since its conception in 2002. Formally known as the microRNA Registry, the creators of the database aimed to create a repository of miRNA data which could undergo updating as new information became available for these important regulators of gene expression (Kozomara and Griffiths-Jones, 2011). The home page can be accessed at http://www.mirbase.org/index.shtml, and Release 21 (June 2014) of the miRBase database is reported to contain 28 645 entries.

8.3.1.2. Ensembl BioMart

The Ensembl BioMart database is a collaborative project between EMBL-EBI and the Wellcome Trust Sanger Institute. The main goal of the Ensembl project, started in 1999, was to create a system of automated annotation of the human genome. The current version of the database, Release 82, became publically available in September 2015, and can be accessed at (http://www.ensembl.org). The database provides recent updated annotations, which include comparative genomics, gene models, variations in the genome, and regulatory regions (Cunningham *et al.*, 2014). The BioMart tool allows users to search for a user-defined dataset, across numerous species, and to filter across multiple parameters. The BioMart tool hosted on the Ensembl website also allows users to convert between dataset formats.

8.3.2. MiRNA target prediction

8.3.2.1. Algorithm-based miRNA target prediction

8.3.2.1.1. Principles and examples of Algorithms

Predicting miRNA targets in animals using computational tools is often observed to be much more challenging than predicting miRNA targets in plants. This is mainly because in animals, miRNAs often perform imperfect base-pairing with the target sites of the target mRNA, whereas in plants, regulatory miRNAs mostly perform perfect base pairing to their targets (Wang *et al.*, 2007). Over the past few years, the scientific community has seen a growth spurt in the field of miRNA target prediction algorithm development. Some of the earlier algorithms include DIANA-microT, MicroInspector, PicTar, TargetScan, and TargetScanS (Xiao *et al.*, 2009).

These algorithms are based on the use of one or more core principles derived from observations of known miRNA-target interactions. Not listed in any particular order of importance, these principles include:

- Near-perfect complementarity between the 6-8 nucleotide region ('seed' region) of the miRNA - near the miRNA's 5' end - with the 3' UTR of the sequence targeted.
- Evolutionary conservation of the binding sites between species
- Strong thermodynamic stability for the miRNA-mRNA duplex structure
- Cooperation between multiple target sites in close proximity to each other
- Existence of a centralised, non-matched area within the sequence, forming a loop structure, or bulge.
- Dinucleotide composition of the flanking sequence (Rehmsmeier *et al.*, 2004)
- Secondary structure accessibility (Leclercq *et al.*, 2013)
- Target-site accessibility (Long *et al.*, 2007).

Two examples of miRNA target prediction algorithms are RNAHybrid and RNA22. The RNAHybrid algorithm searches for the most favourable hybridizations

according to the free energy available between the miRNA and its target mRNA/s (Rehmsmeier *et al.*, 2004). The Rna22 algorithm attempts to identify significant sequence motifs amongst all of the known miRNA sequences. Rna22 then proceeds to define specific regions within the target mRNA, where reverse complementarity of the miRNA sequences accumulate. The algorithm then centres its search for miRNA target sites on these selected regions (Miranda *et al.*, 2006).

8.3.2.2. Experimental validation of miRNA targets

Experimental validation of gene regulation by miRNAs can be a time-consuming and expensive process. Thus, an essential part of the research process is to first perform miRNA target prediction utilising algorithm-based methods. Numerous algorithms, following various principles have been developed over the past few years to provide insight into the putative binding sites of these regulatory elements. Once the miRNA binding sites have been identified within the mRNA target by one or more computational algorithms, the predicted interactions can be validated using a variety of molecular experiments. These molecular experiments include Reporter assays, Northern blots, and quantitative Polymerase Chain Reaction (qPCR).

8.3.2.2.1. Principles and types of Experimental validation methods

The rationale of the reporter assay technique used for validating a miRNA-Target Interaction (MTI), is that the interaction of a specific miRNA to its targeted binding site on the mRNA will repress the production of the protein produced, thereby resulting in a reduced expression of the protein which can then be quantitatively compared to the expression of the control (Hsu *et al.*, 2014). Northern blot analysis makes use of the isolation of total RNA from a chosen cell type and examines the co-expression of both the miRNA, and its targeted mRNA.

Molecular experiments for identifying interactions between miRNAs and their associated targets can become an expensive and time-consuming process, which

until recently was unable to produce meaningful results for large-scale experiments. Recent studies have shown that miRNA-target interactions can be discovered using next-generation sequencing (NGS) as a high-throughput screening method of choice (Hsu *et al.*, 2014).

8.3.2.2.2. Examples of databases integrating data generated from these methods

miRTarBase

miRTarBase is a database housing a vast number of experimentally validated MTI's. The database is accessible at the following website: http://miRTarBase.mbc.nctu.edu.tw/. Version 1 of the miRTarBase database was launched in 2010, comprising of data curated from more than 100 published studies. miRTarBase is built using several algorithms and database updates are performed every two months (Hsu *et al.*, 2014). Algorithms include miRBase version 20, HMDD version 2.0, miR2Disease, Gene Ontology and KEGG pathway analysis. MiRNA-target interactions can be downloaded in Excel format, catalogued by species, experimental evidence, or as a complete dataset for all species, at the following web link: http://miRTarBase.mbc.nctu.edu.tw/php/download.php?opt=list.

The current curation of the database, version 4.5, was released on November 01, 2013, and cites 2 636 articles for the curation of miRNA-target interactions (MTIs). The database contains interaction data for 17 520 target genes and 1 232 miRNAs, across 18 species with a total number of 51 460 miRNA-target interactions presented, curated from 2 636 research articles.

Each publication of interest was perused by at least two of the database developers to confirm an experimentally-validated miRNA-target interaction. Molecular validation experiments include reporter assay, microarray, Western blots, pSILAC (pulsed stable isotope labelling with amino acids in culture), or qPCR. In the publication of the updated version of miRTarBase, Hsu *et al.* describes the database as having a 14-fold increase in the miRNA-target

interactions present in the updated version versus the 2010 version, as well as the inclusion of NGS data generated using Degradome-seq or CLIP-seq methods as support types. In the latest version of the database, 38 113 human miRNA-target interactions (MTIs) with experimental support were collected from 2 143 research articles. These MTIs were reported as the cumulative interactions between 587 miRNAs and 12 194 targeted genes (Hsu *et al.*, 2014).

Experimental validations are split into two categories. The first is "Strong evidence", which is comprised of experiments validated with Western blot analyses and/or Reporter assays. Less strong evidence consists of experimental validations performed by pSILAC, microarray, or NGS assays. This updated version of miRTarBase contains 1 778 and 3 026 MTIs, experimentally validated by Western blot, and reporter assay techniques, respectively (Hsu *et al.*, 2014).

The current version of miRTarBase database (Version 4.5) integrates human MTI expression profiled using 21 GEO datasets, all of which used microarray platforms for molecular experimentation. MiRNA-target networks can also be visualised through integration of the interactive web interface of CytoscapeWeb (Lopes *et al.*, 2010). Visualization of neighbouring interactions help researchers to understand complex MTIs, and assists in unravelling the physiological mechanisms associated with these networks (Hsu *et al.*, 2014).

Each MTI is allocated a unique miRTarBase identification number. Within the graphical user interface (GUI) of miRTarBase, searching for "hsa-miR-21" yielded numerous matches, not strictly for "hsa-miR-21" as an exact string match. The first match made was to the mature miRNA "hsa-miR-21-5p", interacting with the gene target, "RASGRP1". This MTI has the miRTarBase entry ID, MIRT000019. Opening the link of this miRTarBase ID brought up a host of information, including tabs for the miRNA, the target gene, the evidences collected, expression values, and gene set enrichments.

miRecords

miRecords is a database housing both a curation of experimentally validated MTIs, and a miRNA-target prediction tool, which utilises 11 miRNA target prediction algorithms. The miRecords database can be accessed at http://c1.accurascience.com/miRecords/.

The current curation of miRecords' *Validated Targets*, released on April 27 2013, contains 2 705 records of miRNA-target interactions for 644 miRNAs and 1901 gene targets across 9 animal species. This release of the miRecords database of validated-targets can be downloaded in Excel file format at http://c1.accurascience.com/miRecords/download_data.php?v=4.

The *Predicted Targets* section of miRecords utilises the miRNA-target prediction algorithms of 11 software tools: DIANA-microT, MicroInspector, miRanda, MirTarget2, miTarget, NBmiRTar, PicTar, PITA, RNA22, RNAhybrid, and TargetScan/TargetScanS. The miRecords database is constructed in the form of a relational database using MySQL software (Xiao *et al.*, 2009).

9. AIMS AND OBJECTIVES

The main aim of this study was to investigate the miRNA-gene interactions potentially involved in the pathogenesis of Systemic Lupus Erythematosus. Using curated datasets of miRNAs and genes which have been associated with lupus in largely unrelated studies, the associations of these molecules to their targets was determined in the miRTarBase and miRecords databases of experimentally-validated targets. The objectives of this study included generating miRNA-target interaction networks for the data generated from the analyses, to identify new, candidate genes and miRNAs which were placed in interaction networks built using the curated data, and to identify genes and miRNAs which interacted with known lupus-associated genes and miRNAs, but have not been reported in recent biomedical literature to have associations with SLE-pathogenesis. Through further assessment of the proposed novel interactions, the new, candidate genes and miRNAs found through the analyses described herein hold the potential to be therapeutic targets and biomarkers for lupus.

10. METHODOLOGY

10.1. Overview

For this study, recent biomedical literature was carefully perused for the curation of miRNAs and genes which have been associated with lupus pathogenesis as shown in Figure 2. A list of differentially-expressed lupus-associated genes was also included in this study. Two databases containing experimentally-validated target prediction for miRNAs – miRTarBase and miRecords – were interrogated using a bioinformatics scripting language, Python (iPython Version: 2.2.0, Python Version: 2.7.8, and all self-written Python scripts can be viewed at the following GitHub web link: https://github.com/StephP15/MSc_Thesis_2015. SLE-associated gene names selected directly from literature, and Ensembl gene IDs for the differentially-expressed genes, were both verified for their existence as records in the Ensembl BioMart database. With a written Python script, miRNA names were verified for their existence as IDs in a downloaded file containing the miRBase database.

For both the curated genes and miRNAs, corresponding targets were identified in miRTarBase, and miRecords. Subsequently, a cross-analysis of the input data, with the targets found, revealed which miRNAs, and genes were present in both the starting lists and the outputs generated. The cross-analysis therefore revealed how many genes were reported to be associated with SLE in recent literature, and were found to be targets for lupus-associated miRNAs. Similarly, the cross-analysis revealed how many miRNAs reported to be associated with SLE, were also found to target known, lupus-associated genes.

The genes regulated by ten or more miRNAs, and the miRNAs targeting ten or more genes were then prioritized as the top regulators in this study. Using Ingenuity Pathway Analysis (IPA) software, analyses were performed on the top regulators to generate biological networks influenced by these genes and miRNAs. By overlaying the core networks with various prepared datasets, I was

able to visually identify nodes within the networks containing integral genes or miRNAs.

I was also able to identify key genes or miRNAs which were supplied by the IPA database to complete the networks, but were not present in the curated lists of lupus-associated genes, and miRNAs, or differentially-expressed lupus genes. These key genes or miRNAs can be considered as candidate molecules integral to pathways largely associated with known lupus genes and miRNAs. These molecules may provide insight into the missing links in recent biomedical research performed for lupus pathogenesis, and can be considered as potential targets for therapeutics, and further research.

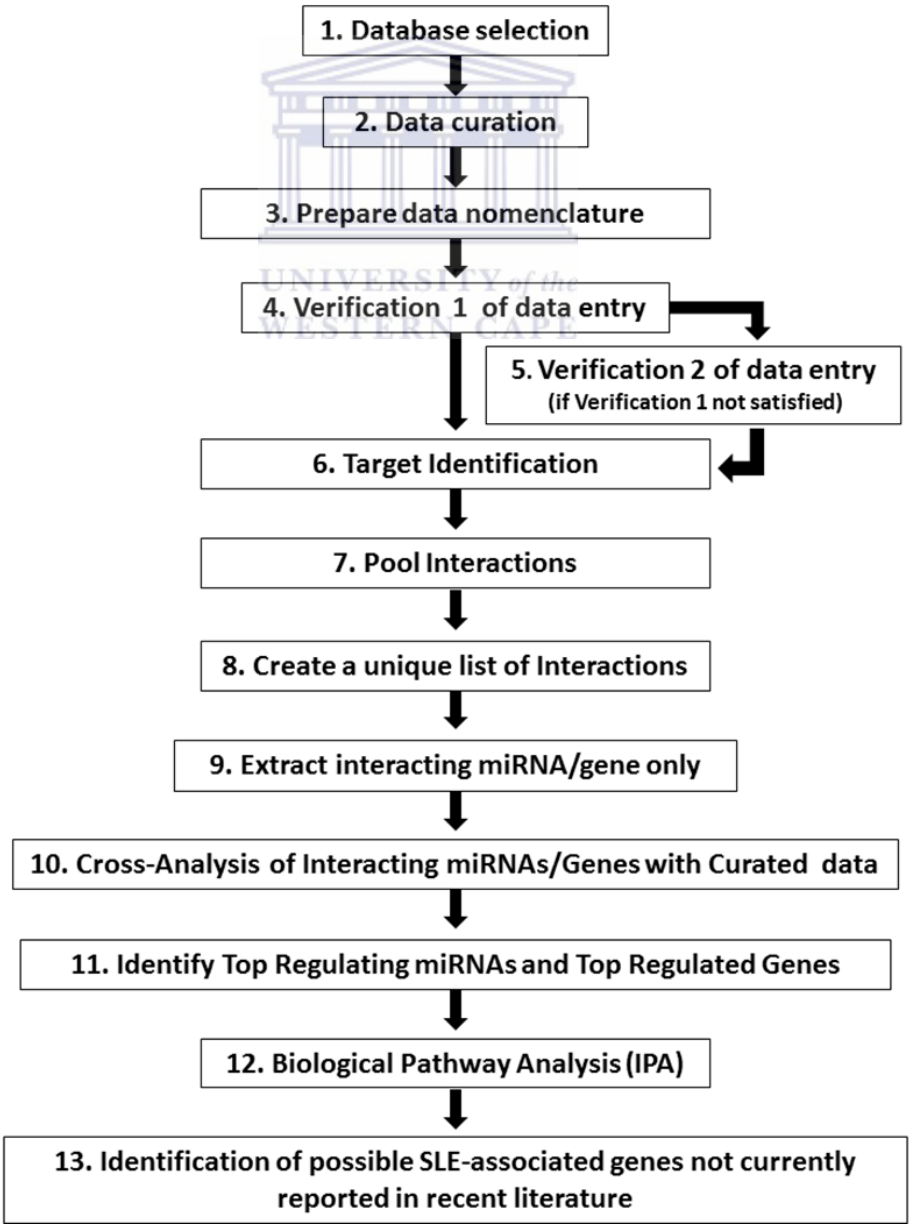

Figure 2: An overview of the methods designed, and used in this study

10.2. Database selection

10.2.1. Database for miRNA verification

10.2.1.1. miRBase database

Due to the complexity of miRNA annotation as described in 8.2.6.3, I chose to verify the SLE-associated miRNA IDs identified from literature. In order to perform this verification, I identified miRBase as a highly suitable database for this important preparatory step, which precedes target identification.

The latest curation of miRBase, version 21, was released in June 2014. miRBase is a database which houses annotated miRNA sequences and target data. On the 20th of May 2015, annotated miRNA sequences were downloaded at the following web link: http://www.mirbase.org/ftp.shtml. Mature miRNA sequences for all species in this database, listed in fasta format, were downloaded as the original file, 'mature.fa', at ftp://mirbase.org/pub/mirbase/CURRENT/mature.fa.gz. An entry in 'mature.fa' contained the following six fields: miRNA, miRBase ID, genus, species, miRNA name, and sequence.

A GFF3 file containing the genome coordinates for all human miRNA sequences was downloaded on the 27th of June at the following web link: ftp://mirbase.org/pub/mirbase/-CURRENT/genomes/hsa.gff3. This file contained the following nine fields: seq_id which contained the chromosome number, source, type=> miRNA or miRNA_primary_transcript, start position, end position, score, strand, phase, and attributes => ID; Alias; Name; Derives_from. The curation date for this data is the 22nd of June 2014, and the data is stored in GFF3 format. Genome build ID: GRCh38, with the Genome-build-accession: NCBI_Assembly:GCA_000001405.15 was used. This file contains information for both predicted pre-miRNA sequences, as well as mature miRNA sequences.

10.2.2. Database for gene name/ gene ID verification

10.2.2.1. Ensembl BioMart

Release 80 (May 2015) of the Ensembl BioMart database was the most updated version of this genome repository at the time of analyses, and was used to verify the existence of the lupus-associated gene names retrieved from literature, as HGNC symbols. Ensembl BioMart was also used to convert the Ensembl Gene IDs of the differentially-expressed lupus genes provided by Dr Wendy Kröger's meta-analysis. As a consequence of verification in the database, the differentially-expressed genes were converted from ENSG IDs to HGNC symbols.

10.2.3. Databases for miRNA/gene target identification

As reported in Table 2, several databases containing MTI data were carefully assessed based on relevance of data for this study, accessibility of data, the year of last curation of the database, the type of target prediction reported, and the number of miRNA-target interactions reported for human. The miRTarBase, miRecords, microRNA.org, and TargetScan.human databases were considered as possible databases to be used as sources for the miRNA-target interactions needed for this study.

Table 2: Summary of database assessment at the time of database selection (March 2015). A total number of MTIs for human was not available for miRecords, or TargetScan.human. A total count of MTIs was also not available for TargetScan.human.

Database	Accessibility of data	Version	Curation	Type of Target Prediction	Total number of MTIs reported	Number of MTIs reported for human	Ref.
miRTarBase	http://miRTarBase.mbc.nctu.edu.tw/php/download.php?opt=list in Excel format	4.5	November 01, 2013	Experimentally validated	51 460	38 113	(Hsu *et al.*, 2014)
miRecords	http://c1.accurascience.com/miRecords/download_data.php?v=4 in Excel format	4	April 27, 2013	• Experimentally validated • Algorithm-Predictions	2705		(Xiao *et al.*, 2009)
microRNA.org	http://www.microrna.org/microrna/getDownloads.do in text file format	August 2010	November 01, 2010	Experimentally validated	2623	1100	(Betel, Wilson, Gabow, Marks, and Sander, 2008)
TargetScan.human	http://www.targetscan.org/cgi-bin/targetscan/data_download.cgi?db=vert_40 data can be downloaded in text file format	4.1	January, 2008	TargetScan algorithm			http://www.targetscan.org/vert_40/docs/help.html

10.3. Data curation

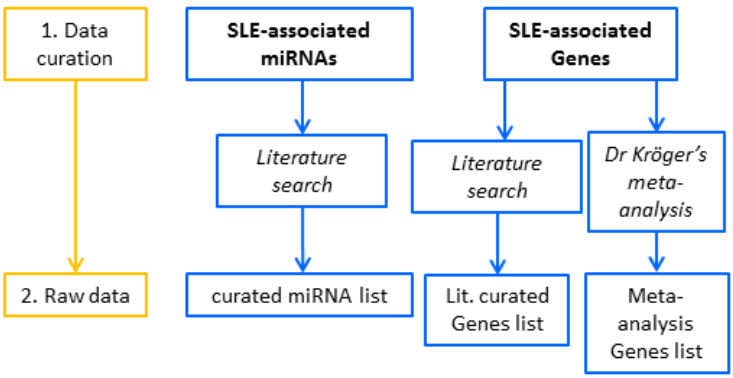

Figure 3: An overview of the methods used for data curation. The numbered orange boxes describe the data curation process, while the blue boxes on the right describe the origin of the miRNA, and gene lists used as source data in this study.

10.3.1. SLE-associated miRNAs

As seen in Figure 3, recent biomedical literature was perused to identify miRNAs which have been associated with lupus pathogenesis. When a miRNA was reported to be associated with lupus pathogenesis, attempts were made to find a secondary article also reporting similar findings for the given miRNA. No editing was made to the miRNA name before adding to the list of lupus-associated miRNAs.

10.3.2. SLE-associated Genes: from literature

A list of lupus-associated genes was compiled by manual curation from recent biomedical literature. Attempts were made to find a secondary article which reported a similar finding for the gene reported to be associated with SLE pathogenesis.

10.3.3. Differentially-expressed SLE-associated genes

A list of genes which demonstrated differential expression with lupus pathogenesis, was obtained from Dr Wendy Kröger (manuscript submitted), and utilised in this study with her permission. A full list of these Ensembl IDs can be found in ADDENDUM A.

Using a collection of existing microarray data, Dr Kröger's meta-analysis entailed identifying differentially-expressed genes in the PBMCs obtained from lupus patients, and which were compared to control samples. The meta-analysis revealed that the differentially-expressed genes identified from the study were implicated in interferon signalling, which was in concordance with the pathways reported in the data source.

However, augmenting the findings in the individual sourced studies, the pathway analysis performed on the meta-analysis output revealed that TLR signalling, diapedesis, oxidative phosphorylation, and adhesion regulatory networks, were identified as the key pathways associated with the differentially-expressed genes found to be associated with the lupus patients in this study.

10.4. Preparation of Data Nomenclature

10.4.1.1. Preparation of SLE-associated miRNA IDs

MiRNA ID assignment should follow the standard annotation rules described in the literature study. Thus, as shown in Figure 4, no changes were made to the miRNA IDs in preparation for their verification in miRBase, and their target identification.

10.4.1.2. Preparation of SLE-associated gene names

The list of lupus genes was prepared for analysis by firstly splitting the aliases in parentheses from the preceding gene which it was reported to be an alias for. Secondly, genes which had been paired were also separated for the purpose of this study, as shown in Figure 4.

10.4.1.3. Preparation of differentially-expressed SLE-associated gene IDs

The 'EnsemblGeneID.txt' file contained the Ensembl gene IDs of differentially-expressed lupus-associated genes as collected by Dr Kröger. Because of the standard annotation used in Ensembl Gene ID nomenclature, no manual preparation or editing was performed on this file, as shown in Figure 4.

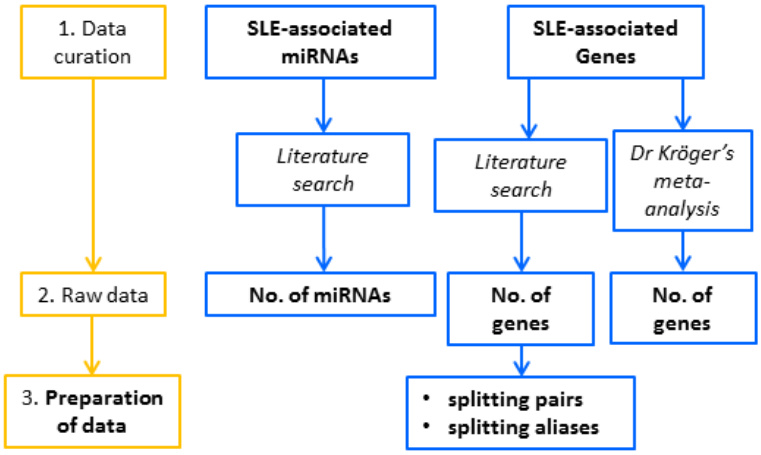

Figure 4: An overview of the methods for data preparation. The numbered orange boxes show the transition of collected raw data being prepared for analyses. The blue boxes show how only the genes curated from literature underwent nomenclature preparation by means of splitting of paired genes and genes with aliases' as retrieved from the source.

10.5. Verification of data

10.5.1. Data set: SLE-associated miRNAs

10.5.1.1. Python Script 1: 'PS01_June_28_miRBase_1.py'

The first three steps of this python script prepared the miRBase database for this verification step by merging the sequence annotation file, and the genome coordinates data file. Thus, the above python script performed the following steps sequentially:

1) Filtered the contents of the 'mature.fa' file to extract only entries containing information for genus = 'Homo', species = 'sapiens', and for mature miRNA IDs, i.e. miRNA names starting with 'hsa-miR-'.
2) Filtered the contents of the miRNA genome coordinates file ('hsa.gff3'), to extract entries only for mature human miRNA sequences, in which the miRNA name started with 'hsa-miR- '.

The unique miRBase accession numbers are the one truly stable characteristic of miRBase entries in which miRNA names may change over time due to advances in miRNA research, http://www.mirbase.org/help/nomenclature.shtml.

Thus,

3) The ‘mature.fa’ file (containing the accession numbers for mature miRNA sequences was cross-referenced with the ‘hsa.gff3’ file (genome coordinates), on the basis of the miRBase accession number, provided as an ID in Attributes.

The ‘mature.fa’ file contained miRNA sequences where the ‘hsa.gff3’ file did not, and the ‘hsa.gff3’ file contained genome coordinates, where the ‘mature.fa’ file did not. Therefore, on cross-referencing with the unique miRBase accession number, the two files were merged to create a more comprehensive file containing the following fields: microRNA_name, miRNA_accession number, Genus, Species, microRNA, miRNA sequence, chromosome, start genome coordinate, and end genome coordinate. This data was recorded as a file named ‘mirbase_complete.txt’.

4) The first check for whether the miRNA names curated from literature were present as a record in the miRBase database was then performed. The list of lupus-associated miRNA names, curated from literature was fed to the python script ‘PS01_June_28_miRBase_1.py’ to test if the miRNA names were found as an exact string match to the entries in the miRNA name column of the miRBase database.

As shown in Figure 5, if the input miRNA from the curated list ‘microRNAs_assoc_SLE.txt’ matched perfectly as a string to the miRNA record in the ‘mirbase_complete.txt’ file, the results were written out to a file named ‘perfect_match_in_miRBase.txt’. If the input miRNA name from the curated list ‘microRNAs_assoc_SLE.txt’ did not make an exact string match to a miRNA record in the ‘mirbase_complete.txt’ file, the results were written out to a file named ‘no_match_in_miRBase.txt’.

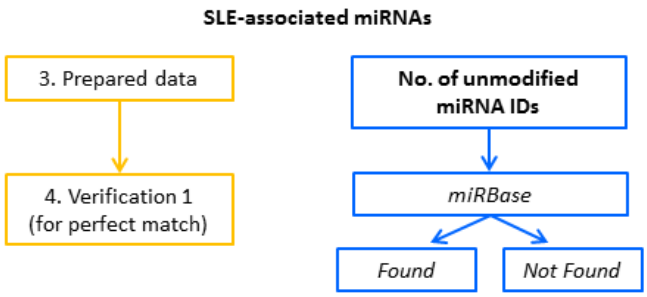

Figure 5: The numbered orange boxes shows how data proceeded to the first verification step for a perfect string match in miRBase. The blue boxes on the right shows how the miRNA list which did not undergo any nomenclature preparation proceeded to undergo the first verification step in the miRBase database (shown in blue). The attempt to identify a perfect string match for the unmodified miRNA IDs resulted in two files generated which are described as "Found", and "Not Found".

10.5.1.2. Python Script 2: 'PS02_July_01_new_regexes.py'

Due to the complex annotation rules associated with miRNAs as described, for those miRNAs which did not match as a perfect string match in miRBase, I decided to employ regular expression code to perform a secondary search for miRNAs which may be related to those which did not match with the first verification step. Regular expression testing maximised the chance of retrieving closely related miRNA names for the curated miRNA IDs. Regular expressions allow the user to query for a defined partial match to a string sequence. Thus, to query miRNA IDs with regular expressions, I wrote three Python regular expressions for the miRBase database column containing the miRNA names, namely 'hsa-miR-\d+', 'hsa-miR-\d\d\D', and 'hsa-miR-\d\d\d\D'.

As shown in Figure 6, if the input miRNA name from the list of miRNAs which did not make an exact string match to a miRBase record ('no_match_in_miRBase.txt') matched with at least one of the three regular expressions described above, the miRNA was written to the output file named 'found_regex_matches_in_mirbase.txt'. If the input miRNA name from the list of miRNAs which did not make an exact string match to a miRBase record ('no_match_in_miRBase.txt') did not match with any of the three regular expressions described above, the miRNA was written to the output file named 'not_found_regex_matches_in_mirbase.txt'. MiRNAs which did not match a

miRNA entry in the miRBase database using one of the three regular expressions were then manually checked for, using the Find tool in Microsoft Excel.

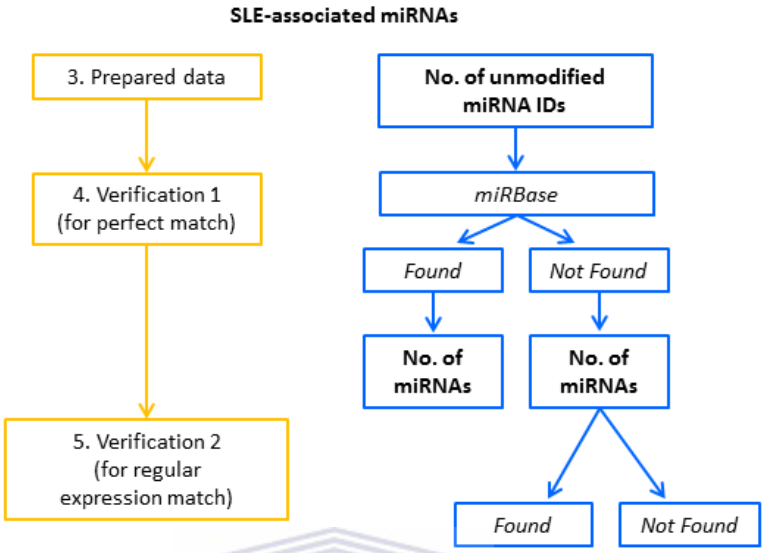

Figure 6: The orange boxes on the left of the figure shows the transition of miRNA IDs from the first verification step (as a perfect string match) to the second verification step (as a regular expression string match). The blue boxes on the right of the figure describes how miRNA IDs which did not produce a perfect string match during the first verification step in miRBase proceeded to undergo a second verification attempt using regular expression string matching in Python.

10.5.2. Data set: SLE-associated gene names curated from literature

10.5.2.1. Verification of the SLE-associated gene names

On the 14th of July 2015, the curated genes were searched for in the Ensembl BioMart database, Release 80 (May 2015), to validate the existence of these genes in this repository of known, annotated genes as shown in Figure 7. The following parameters were selected:

- Database tab: Ensembl Genes 80
- Dataset tab: Homo sapiens genes (GRCh38.p2)
- Filters tab:
    - Input external references ID list:
        - Selected to search as HGNC symbol
        - 'new_list.txt' was the input text file containing the names of lupus-associated genes curated from literature
- Attributes tab: desired information for output

- associated gene name
- Ensembl gene ID
- gene start bp
- gene end bp

The result of this search, 'mart_export.txt', was saved to my local machine.

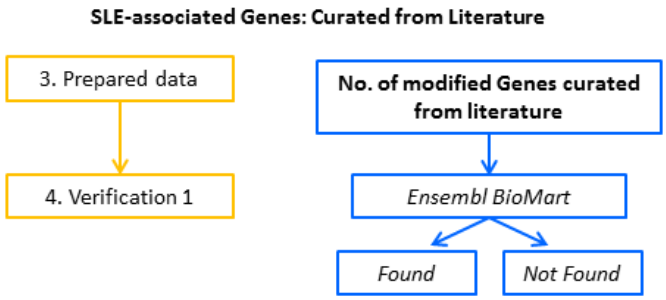

Figure 7: The numbered orange boxes shows how data proceeded to the verification step. The blue boxes show how genes curated from literature underwent data preparation, followed by the verification of the individual gene names as HGNC symbols in Ensembl BioMart, resulting in two files containing the names of the genes which were "Found" in the BioMart database, and those which were "Not Found".

10.5.2.2. Python Script 3:

'PS03_15_July_assoc_genes_prep_and_Biomart.py'

Using the above Python script, the 'mart_export.txt' file was analysed to determine which of the curated SLE-associated genes were found in BioMart, and which genes were not. The Python script employed exact-match filtering using "==". If the associated gene name in 'new_list.txt' was present in the output of the BioMart search named 'mart_export.txt', then the gene name, verified in BioMart was written to the text file, 'Biomart_assoc_gene_ MATCH_2.txt'. If a gene name was not found in the 'mart_export.txt' file, then the gene name was written to the 'Not_Found_assoc_genes_in_BioMart_2.txt' file, and was also searched for manually in the Ensembl BioMart search pane, in order to verify that the gene was indeed not present in the repository.

10.5.3. Data set: Differentially-expressed SLE-associated gene IDs

10.5.3.1. Verification of the existence of the differentially-expressed lupus-associated gene names

On the 16th of June 2015, the differentially-expressed genes were searched for in Ensembl BioMart, to validate these Ensembl IDs in this repository of known, annotated genes. The parameters used were:

- Database tab: Ensembl Genes 80
- Dataset tab: Homo sapiens genes (GRCh38.p2)
- Filters tab:
    - Input external references ID list:
        - Selected to search as Ensembl ID
        - 'EnsemblGeneID.txt' was the input text file containing the names of differentially-expressed lupus-associated genes
- Attributes tab: desired information for output
    - ensembl gene ID
    - associated gene name
    - gene start bp
    - gene end bp
    - phenotype
    - status

The result of this search, 'mart_export.txt', was saved to my local machine, and renamed as 'BioMart_DE_genes.txt'.

10.5.3.2. Python Script 4: 'PS04_16_July_DE_genes_thru_BioMart.py'

This script used the following files in the analysis:

- 'EnsemblGeneID.txt' => the original file containing Ensembl Gene IDs of differentially-expressed SLE-associated genes

- ‘BioMart_DE_genes.txt’ => The output file of searching for those genes in BioMart

Because of the standard annotation used in Ensembl Gene IDs, no manual preparation or editing was performed on this file. Using the above Python script, the file containing gene IDs which were matched in BioMart, ‘BioMart_DE_genes.txt’, was then run through the above script and filtered against the original starting list of differentially-expressed gene IDs in ‘EnsemblGeneID.txt’ to determine which genes were not found in BioMart.

A sorted list of the output was generated, and furthermore, the first occurrence of each Ensembl gene ID match was retained for further analyses and written to a list, ‘ensembl_list_new.txt’. As shown in Figure 8, for each Ensembl Gene ID queried and found in BioMart, the Ensembl Gene ID searched for, along with all the attribute data accompanying it, was written to the output file ‘Found_DE_genes_ALLinfo_in_BioMart.txt’. A subset of this output data, containing only the associated gene name for the matched Ensembl Gene IDs was written out to a separate file named ‘Found_DE_genes_in_BioMart.txt’.

If the Ensembl Gene ID was not found in BioMart, the ID was written to a file named ‘Not_Found_DE_genes_in_BioMart.txt’. To confirm the results of the IDs written to ‘Not_Found_DE_genes_in_BioMart.txt’, a manual search was conducted for the IDs not found in BioMart.

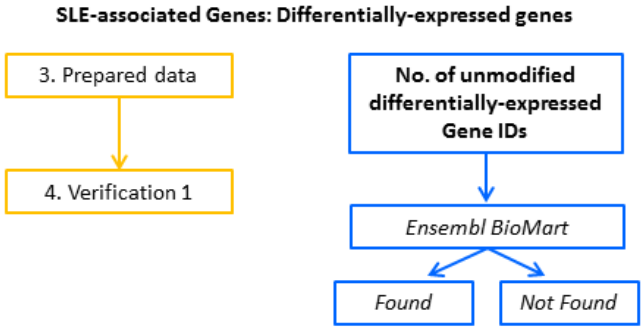

Figure 8: The numbered orange boxes describe how the differentially-expressed genes list in Ensembl ID format proceeded to the verification step for these IDs. The blue boxes show how this unmodified genes list - provided by Dr Kröger – underwent Ensembl gene ID verification as ENSG IDs in Ensembl BioMart, producing two lists - IDs which were found in the database, and IDs which were not.

10.6. Target identification

10.6.1. Identification of gene targets for SLE-associated miRNAs

10.6.1.1. Python Script 5: 'PS05_05_July_miRTarBase_miRecords_for_PERFECT_MATCH.py'

10.6.1.1.1. miRTarBase

The text file '*perfect_match_in_miRBase.txt*', produced in 10.5.1.1, containing the names of the input miRNAs which matched as a perfect string match to miRNA names in the miRBase database, was then run through the above Python script. This Python script took each of the miRNA names and attempted to make an exact string match to the miRNA name present in the same 'hsa-miR-' format in both the miRTarBase and the miRecords databases as shown in Figure 9. If the miRNA name was matched in the miRTarBase database, and the length of the corresponding gene name in the fourth column was greater than 1 string character, then the miRNA name, and the corresponding gene name was written out as a result to the output file 'found_miRNA1_ALL_targets_in_ miRTarBase.txt'. If however, the miRNA name was not matched in the miRTarBase database, then the miRNA name was written to the output file 'NOT_found_miRNA1_ALL_ targets_in_miRTarBase.txt'.

10.6.1.1.2. miRecords

The text file '*perfect_match_in_miRBase.txt*' was then also processed through the miRecords database, to identify miRNA-gene interactions in this validated-targets database. If found in the miRecords database file, the miRNA and its associated gene targets were written to the text file 'found_miRNA1_ALL_targets _in_miRecords.txt'. If the miRNA was not found in the miRecords database, the miRNA was written to 'NOT_found_miRNA1_ALL_targets_in_miRecords.txt'.

10.6.1.2. Python Script 6: 'PS06_05_July_miRTarBase_miRecords_for_NOT_FOUND_IN_MIRBASE.py'

10.6.1.2.1. miRTarBase

The text file '*no_match_in_miRBase.txt*', produced in 10.5.1.1., containing the names of the input miRNAs which did not match as an exact string match to miRNA names in the miRBase database, was then run through the above Python script. This Python script took each of the miRNA names and attempted to make an exact string match to the miRNA name, present in the same 'hsa-miR-' format in both the miRTarBase and the miRecords databases. This script attempted to identify gene targets for the miRNAs which were not validated as exact string matches in the miRBase database.

If the miRNA name was matched in the miRTarBase database, and the length of the corresponding gene name in the fourth column was greater than 1 string character, then the miRNA name, and the corresponding gene name was written out as a result to the output file 'found_miRNA0_ALL_targets_in_ miRTarBase.txt'. If however, the miRNA name was not matched in the miRTarBase database, then the miRNA name was written to the output file 'NOT_found_miRNA0_ALL_ targets_in_miRTarBase.txt'.

10.6.1.2.2. miRecords

The text file '*no_match_in_miRBase.txt*' was then also processed through the miRecords database, to identify miRNA-gene interactions in this validated-targets database. If found in the miRecords database, the miRNA and its associated targets were written to the text file 'found_miRNA0_ALL_ targets_in_miRecords.txt'. If the miRNA was not found in the miRecords database, the miRNA was written to 'NOT_found_miRNA0_ALL_targets_in_ miRecords.txt'.

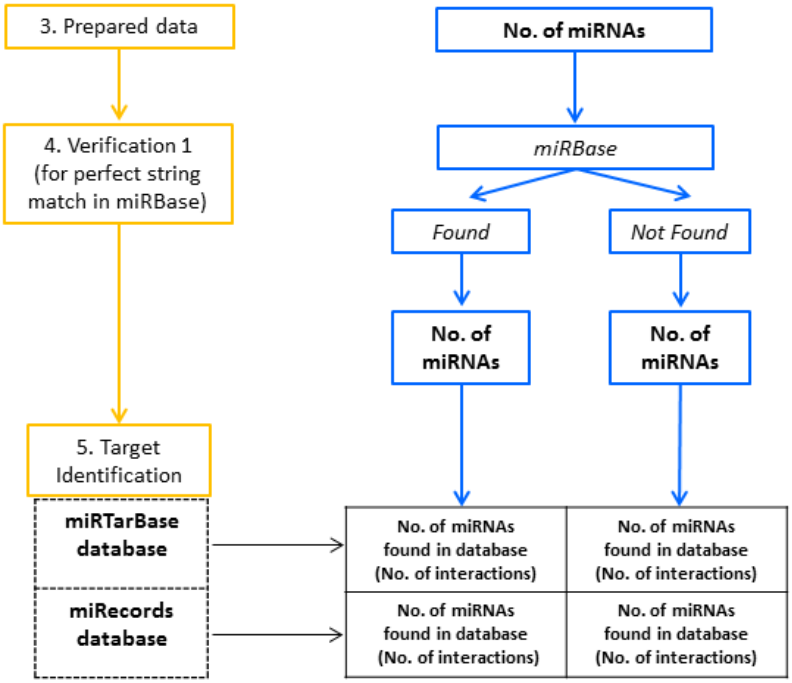

Figure 9: The numbered orange boxes shows how the data which was verified as a perfect string match proceeded to target identification in two curated miRNA-Target Interactions (MTI) databases. The blue boxes describe how miRNAs which were found as a perfect string match in miRBase, underwent target identification in the miRTarBase and miRecords databases.

10.6.1.3. Python Script 7: 'PS07_05_July_miRTarBase_miRecords_for_FOUND_REGEX_MATCH.py'

10.6.1.3.1. miRTarBase

The text file *'found_regex_matches_in_miRBase.txt'*, produced in 10.5.1.2., containing the names of the input miRNAs which matched as regular expressions for the miRNAs in the miRBase database, was then run through the above script. As shown in Figure 10, gene targets were identified for the regular expression matched miRNAs in the miRTarBase and miRecords databases.

If the miRNA name was matched in the miRTarBase database, and the length of the corresponding gene name in the fourth column was greater than 1 string character, then the miRNA name, and the corresponding gene name was written out as a result to the output file 'found_miRNA2_ALL_targets_in_ miRTarBase.txt'. If however, the miRNA name was not matched in the

miRTarBase database, then the miRNA name was written to the output file 'NOT_found_miRNA2_ALL_ targets_in_miRTarBase.txt'.

10.6.1.3.2. miRecords

The text file '*found_regex_matches_in_miRBase.txt*' was then also processed through the miRecords database, to identify miRNA-gene interactions in this validated-targets database. If found in the miRecords database, the miRNA and its associated gene targets were written to the text file 'found_miRNA2_ALL_ targets_in_miRecords.txt'. If however, the miRNA was not found in the miRecords database, the miRNA name was written to 'NOT_found_miRNA2_ ALL_targets_in_miRecords.txt'.

10.6.1.4. Python Script 8: 'PS08_05_July_miRTarBase_miRecords_for_NOT_FOUND _REGEX_ MATCH.py'

10.6.1.4.1. miRTarBase

The text file '*not_found_regex_matches_in_mirbase.txt*', produced in 10.5.1.2., containing the names of the miRNAs which did not match with the second chance attempt of regular expression matching in miRBase, was then run though the above Python script, to determine if any of these miRNAs, not validated in miRBase, were recorded in the miRTarBase and miRecords database as having gene targets.

If the miRNA name was matched in the miRTarBase database, and the length of the corresponding gene name in the fourth column was greater than 1 string character, then the miRNA name, and the corresponding gene name was written out as a result to the output file 'found_miRNA3_ALL_targets_in_ miRTarBase.txt'.

If however, the miRNA name was not matched in the miRTarBase database, then the miRNA name was written to the output file 'NOT_found_miRNA3_ALL_ targets_in_miRTarBase.txt'. A manual check was also performed to confirm any

miRNAs which were not found in the miRTarBase database using the Python script.

10.6.1.4.2. miRecords

The text file '*not_found_regex_matches_in_mirbase.txt*', was then also processed through the miRecords database, to identify miRNA-gene interactions in this validated-targets database. Importantly, the miRNAs not found as regular expression match in miRBase, is in fact a subset of the miRNAs not found as an exact string match in miRBase. Thus, the result of this match is a duplication of the match found in 'no_match_in_mirbase.txt'.

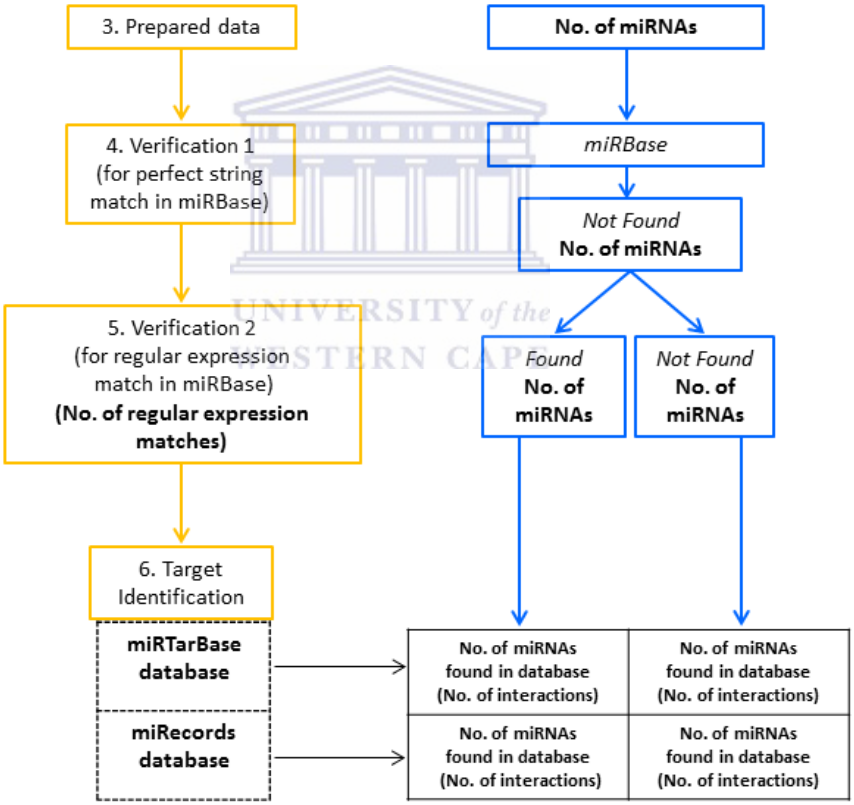

Figure 10: The numbered orange boxes describe the transition of data which has passed through the second verification step for regular expression string matches, to the target identification step. The blue boxes describe how unmodified miRNA IDs which did not pass the perfect string matching step during the first verification step, proceeded to a second verification step, producing two lists for miRNAs which were found in miRBase using the regular expressions, and the miRNAs which were not found in miRBase using this second method for ID verification.

10.6.2. Identification of miRNAs interacting with SLE-associated genes

10.6.2.1. Python Script 9: 'PS09_15_July_Genes1_thru_DBs.py'

The text file 'BioMart_assoc_gene_MATCH_2.txt', containing the names of the genes which passed both the matching in BioMart, and the case-matching in the Python script 'PS03_15_July_assoc_genes_prep_and_Biomart.py', was then run through the Python script 'PS09_15_July_Genes1_thru_DBs.py'. This Python script took each of the gene names and attempted to make an exact string match to the gene name present in the same associated gene name format, in both the miRTarBase and the miRecords databases, in the fourth column and second column, respectively.

10.6.2.1.1. miRTarBase

As shown by the orange block numbered 5 in Figure 11, if the gene name was matched in the miRTarBase database, and the length of the corresponding miRNA name in the second column was greater than 1 string character, then the gene name, and the corresponding miRNA name was written out as a result to the output file 'found_Genes1_ALL_targets_in_miRTarBase.txt'. The remaining genes which were found in Ensembl BioMart, but were not found in miRTarBase to interact with one or many miRNAs were written to the text file, 'NOT_found_Genes1_targets_in_miRTarBase.txt'.

10.6.2.1.2. miRecords

The text file 'BioMart_assoc_gene_MATCH_2.txt' was then also processed through the miRecords database to identify gene-miRNA interactions in this database of experimentally-validated targets, as shown in Figure 11. If the gene name was matched in the miRTarBase database, and the length of the corresponding miRNA name in the second column was greater than 1 string character, then the gene name, and the corresponding miRNA name was written out as a result to the output file 'found_Genes1_ALL_targets_in_miRecords.txt'. The remaining genes which were found in Ensembl BioMart, but were not found in miRecords to interact with one or many miRNAs were written to the text file,

'NOT_found_Genes1_ targets_in_miRecords.txt'. Selections of the genes not found in miRecords were manually checked for in the miRecords database file to confirm the findings.

10.6.2.2. Python Script 10: 'PS10_15_July_Genes2_thru_DBs.py'

The text file 'Not_found_assoc_genes_in_BioMart_2.txt', containing the names of the genes which failed both the matching in BioMart, and the case-matching in the Python script 'PS03_15_July_assoc_genes_prep_and_BioMart.py', was then run through the Python script 'PS10_15_July_Genes2_thru_DBs.py'. This Python script took each of the gene names and attempted to make an exact string match to the gene name present in the same associated gene name format, in both the miRTarBase and the miRecords databases, in the fourth column and second column, respectively as shown in the bottom far right of Figure 11.

10.6.2.2.1. miRTarBase

If the gene name was matched in the miRTarBase database, and the length of the corresponding miRNA name in the second column was greater than 1 string character, then the gene name, and the corresponding miRNA name was written out as a result to the output file 'found_Genes2_ALL_targets_in_miRTarBase.txt'. If the gene name was not matched in the miRTarBase database, the gene was written out to a text file 'NOT_found_Genes2_targets_in_miRTarBase.txt'.

10.6.2.2.2. miRecords

The text file 'Not_found_assoc_genes_in_BioMart_2.txt' was then also processed through the miRecords database, to identify gene-miRNA interactions in this validated-targets database. If gene interactions were found for these genes, the results were written out to the file 'found_Genes2_ALL_targets_in_miRecords.txt'. If gene interactions were not found for these genes in the miRecords database, these genes were written out to the file 'NOT_found_Genes2_targets_in_miRecords.txt'. Selections of these were manually checked to confirm the findings.

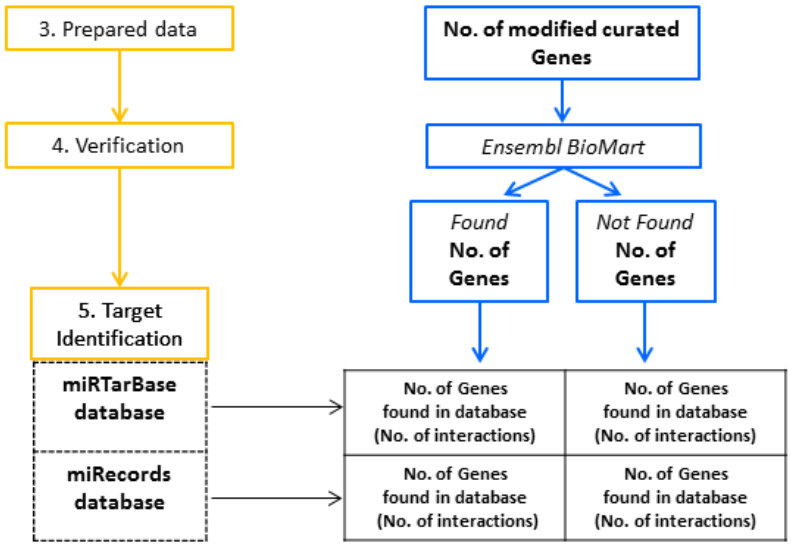

Figure 11: The orange boxes describe how data proceeded from verification to target identification in two curated MTIs databases. The blue boxes show how genes which underwent nomenclature preparation proceeded from verification in BioMart to the target identification in miRTarBase and miRecords. The number of genes curated from literature found, and not found in each of the MTI databases was recorded, along with the number of interactions with miRNAs found for each input gene, as indicated in parentheses.

10.6.3. Identification of miRNAs interacting with the Differentially-expressed lupus-associated genes

10.6.3.1. Python Script 11: 'PS11_16_July_DE_Genes_thru_DBs.py'

10.6.3.1.1. miRTarBase

As shown in Figure 12, if the gene name, in HGNC gene name format from the file 'Found_DE_genes_in_BioMart.txt', was matched in the miRTarBase database, and the length of the corresponding miRNA name in the second column was greater than 1 string character, then the gene name, and the corresponding miRNA name was written out as a result to the output file 'found_DE_genes_ALL_targets_in_miRTarBase.txt'. The remaining genes which were found in Ensembl BioMart, but were not found as an entry in miRTarBase to interact with one or many miRNAs were written to the text file, 'NOT_found_DE_genes_targets_in_miRTarBase.txt'.

10.6.3.1.2. miRecords

The text file 'Found_DE_genes_in_BioMart.txt' was then also processed through the miRecords database, as shown in Figure 12, to identify gene-miRNA

interactions in this validated-targets database. If a differentially-expressed gene was found in the miRecords database, the gene along with its interactions was written to the text file 'found_DE_genes_ALL_targets_in_miRecords.txt'. The remaining genes which were found in Ensembl BioMart to exist, but were not found as an entry in miRecords to interact with one or many miRNAs were written to the text file, 'NOT_found_DE_genes_targets_in_miRecords.txt'.

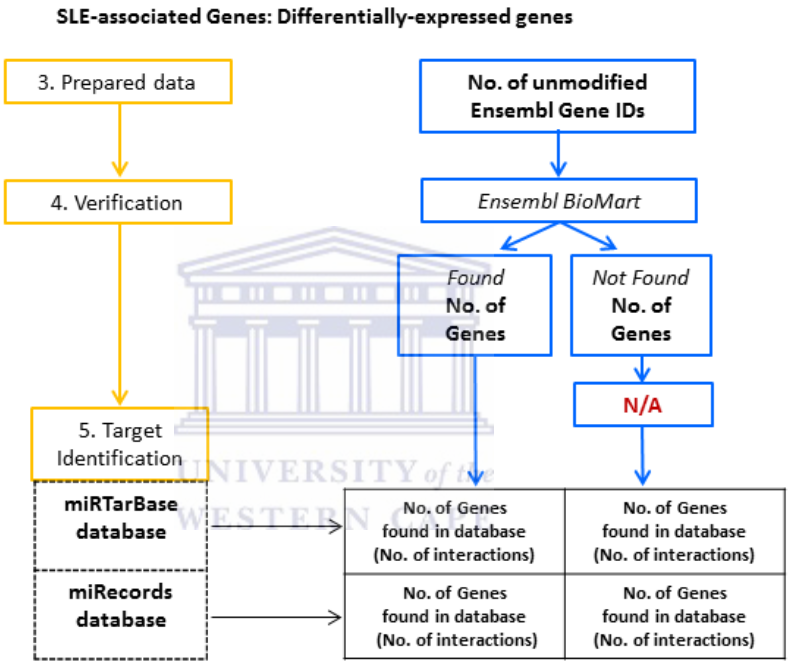

Figure 12: The orange boxes describe the transition of verified data to target identification. The blue boxes describe how unmodified Ensembl Gene IDs were verified in Ensembl BioMart. This verification step also provided a conversion of the Ensembl IDs to HGNC symbols. Thus, the Ensembl IDs which were "Not Found" in BioMart were excluded from Target Identification as indicated by the "N/A" on the right hand side of the figure.

10.7. Cross-Analysis of Interactions

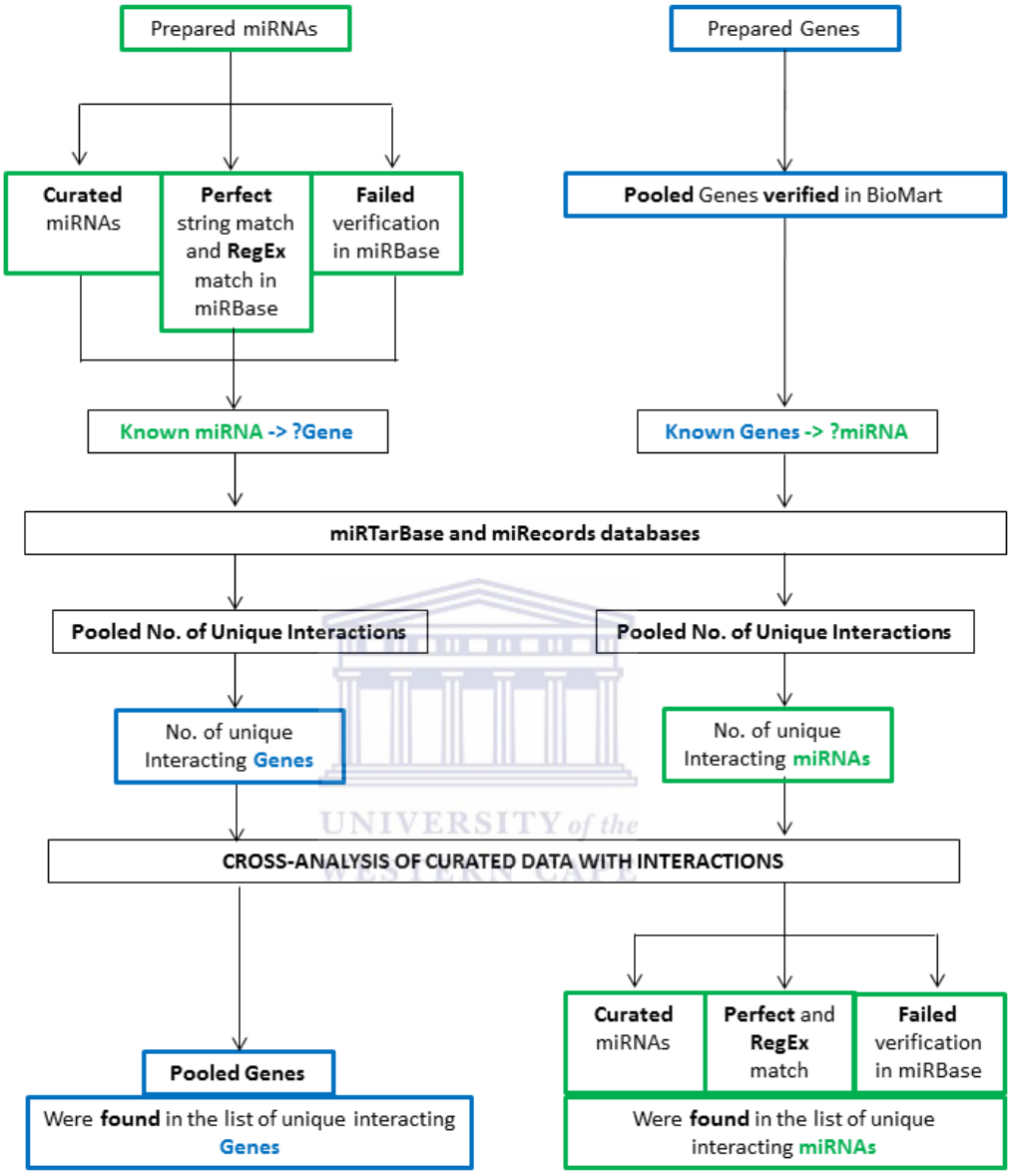

Figure 13: This figure describes the cross-analysis processing of the curated SLE-associated miRNAs and genes. Starting with data curation, proceeding to target identification, and lastly identifying which genes and miRNAs were found to interact with the curated data, and were also found in the initial curated data sets.

As shown in Figure 13, the 3 top-left green boxes joined together were cross-analysed with the pooled unique interacting miRNAs in the bottom of the right-hand side process. Similarly, the pooled genes which were verified in BioMart as shown by the blue box in the top right of the image were cross-analysed with the unique genes found to interact with the curated SLE-associated miRNAs.

10.7.1. Cross-analysis of Lupus-associated miRNAs

10.7.1.1. Pooling of miRNAs interacting with input gene lists

All interacting miRNAs obtained in the analysis of lupus-associated genes and differentially-expressed genes were pooled into a single list for the cross-analysis. This single list consisted of the curated list of miRNAs combined with the interacting miRNAs found in miRTarBase and miRecords. Each of the test scripts which are described below performed the pooling of interacting miRNAs and was identical in the three cross-analysis tests described. Thus, the only variable files were the input miRNAs to cross-analyse.

The set() function in Python was used to create a list of unique gene-miRNA interactions – as the associated genes lists and differentially-expressed genes lists may contain overlapping interactions. To determine how many unique miRNAs were interacting with the input list of genes, the miRNA names were extracted from the interactions, and appended to a list named 'mirnas_only.txt'. The items of the list of unique miRNAs were also written to a file named 'unique_interacting_miRNAs.txt' for further pathway analysis.

10.7.1.2. Test 1: starting list of VS interacting miRNAs

Python Script 12: 'PS12_Test_miRNA_1.py'

The starting list of miRNA names curated from literature were analysed here to identify how many, and which of these miRNAs was present in the list of interacting miRNAs. These miRNAs had not yet been processed for verification as miRNA IDs in miRBase. If the miRNA in the curated/starting list was found in the list of miRNAs identified to interact with the curated list of SLE-associated genes, the miRNA-gene interaction was written to the output file, 'starting_mirnas_ found_interactions.txt'. The miRNAs from the starting list which were found in the list of interacting miRNAs was written to the output file, 'starting_mirnas_IN_interacting.txt'. The miRNAs from the starting list which were not found in the list of interacting miRNAs was written to the output file, 'starting_mirnas_NOT_IN_interacting.txt'.

10.7.1.3. Test 2: Verified (==) and (=regex) VS interacting miRNAs

Python Script 13: 'PS13_Test_miRNA_2.py'

In this script, the miRNAs verified as an exact string match in miRBase, and the miRNAs verified as a regular expression in miRBase were pooled. Using the set() function in Python, a unique list of miRNAs was generated thus removing any duplicate miRNA names. The miRNAs from the list of verified genes associated, and differentially-expressed in lupus, which were not found in the list of interacting miRNAs were written to the output file 'verified_mirnas_NOT_IN_ interacting.txt'.

10.7.1.4. Test 3: Not verified VS interacting miRNAs

Python Script 14: 'PS14_Test_miRNA_3.py'

This script analysed the miRNAs which did not pass the verification step as an exact string match in the collection of miRBase entries ('no_match_in_ miRBase.txt') For these miRNAs, if the miRNAs were found to interact with the SLE-associated-, and differentially-expressed genes, the miRNA-gene interactions were written to the output file, 'not_verified_mirnas_found_interactions.txt'. Of these miRNAs not verified in miRBase, the miRNA names which were found in the list of interacting miRNAs were written to the output file 'not_verified_mirnas_IN_interacting.txt'. The miRNA names from the curated list of miRNAs, which were not found in the list of interacting miRNAs was written to the output file, 'not_verified_mirnas_NOT_IN_interacting.txt'.

10.7.2. Cross-analysis of Lupus-associated Genes

10.7.2.1. Pooling of genes interacting with input miRNA list

All gene interactions obtained in miRTarBase and miRecords from the analysis of the lupus-associated miRNAs were pooled into a single list for the cross-analysis of the associated lupus genes curated from literature, and the differentially-expressed lupus genes obtained from Dr Wendy Kröger's meta-analysis. Each of the test scripts which are described below performed the pooling of interacting

genes, and this process was identical in the two cross-analysis tests described. Thus, the only variable files were the input genes lists to cross-analyse.

The set() function in Python was then used to create a list of unique miRNA-gene interactions (as shown in Figure 13), as the exact string match miRNAs and regular expression match miRNAs may have generated overlapping miRNA-gene interactions. To determine how many unique genes were interacting with the input list of miRNAs, the gene symbols were extracted from the interactions, and appended to a list named 'genes_only.txt'.

10.7.2.2. Test 1: Verified in BioMart VS interacting genes

Python Script 15: 'PS15_Test_Gene_1.py'

This script analysed the list of unique genes created, by pooling the curated SLE-associated, and differentially-expressed SLE-associated genes. These interactions were written to the output file, 'verified_Genes_found_interactions.txt'. The genes from the pooled list of curated genes, which were verified in BioMart, and were found in the list of interacting genes, were written to the output file, 'verified_Genes_IN_interacting.txt'. The genes from the pooled list of curated genes, which were verified in BioMart, and were not found in the list of interacting genes, were written to the output file, 'verified_Genes _NOT_IN_interacting.txt'.

10.7.2.3. Test 2: NOT Verified in BioMart VS interacting genes

Python Script 16: 'PS16_Test_Gene_2.py'

The differentially-expressed SLE-associated genes were parsed to BioMart in Ensembl ID (ENSG) format. Thus, any gene IDs which were not found in BioMart could not be processed with the above script, because a conversion to HGNC symbol could not be performed.

A manual search was performed to confirm the findings of Ensembl gene IDs not found in BioMart. If a gene not verified in BioMart was found in the pooled list of interacting genes, the interactions were written to the output file, 'verified_Genes_found_interactions.txt'. If any of the genes not verified in

BioMart were found in the pooled list of interacting genes, these genes were written to a file named 'NOT_verified_Genes_IN_interacting.txt'. The genes from the list of genes not verified in BioMart, and which were not found in the list of interacting genes were written to the output file 'NOT_verified_Genes_ NOT_IN_interacting.txt'.

10.8. Prioritization of data for Pathway Analysis

As described in the literature review (8.2.6), it has been well-established that a single miRNA has the ability to regulate multiple genes, whilst a single gene can also be targeted for regulation by multiple miRNAs. As this project involved creating a novel bioinformatics pipeline, a number of criteria were decided on at my own discretion, as described in detail below. These cut-off values can be adjusted in order to manipulate the amount of data included in the generation of the core networks.

10.8.1. Identification of Top Regulating miRNAs

From the cross-analysis of the curated SLE-associated miRNAs, a subset of these miRNAs was found to interact with the curated SLE-associated genes. For these miRNAs, those which were found to interact with ten or more SLE-associated genes were identified as the top regulating miRNAs for this dataset. This cut-off was a pragmatic choice, allowing for the generation of well-populated networks that were neither too complex, nor too sparse. The top regulating SLE-associated miRNAs were pooled with their corresponding interactions into a single text file named 'all_top_regulating_miRNAs.txt'.

10.8.2. Identification of Top Regulating Genes

From the cross-analysis of the curated SLE-associated genes, a subset of these genes was found to interact with the curated SLE-associated miRNAs. For these genes, those which were found to interact with four or more SLE-associated miRNAs were identified as the top regulated genes for this dataset. This cut-off was also a pragmatic choice, allowing for the generation of well-populated networks that were neither too complex, nor too sparse. The top regulated SLE-

associated, and/or differentially-expressed genes were pooled with their corresponding interactions into a single text file named 'all_top_regulating_genes.txt'.

10.9. Pooling of all top regulators

The top regulated miRNAs and top regulating genes, along with their associated targets were pooled into a single data file named 'all_top_regulated.txt'.

10.10. Pathway Analysis

10.10.1. Preparation of dataset files

Two python scripts, Python Script 17: 'PS17_prep_genes_for_IPA.py' and Python Script 18: 'PS18_prep_miRNAs_for_IPA.py', were written to prepare the individual dataset files into lists of unique identifiers.

10.10.1.1. Starting Genes and miRNAs

All curated genes and miRNAs were pooled to create a non-unique list. Using the set() function in Python in each of the two scripts above, I created a unique list of curated identifiers. The unique lists of starting genes, and miRNAs were then copied and pasted into a text file named, 'known_list.txt'.

10.10.1.2. Interacting Genes and miRNAs

The genes found to interact with the curated miRNAs, and the miRNAs found to interact with the curated genes, were pooled to create a list of IDs for the interacting molecules only. Using the set() function, a unique list of identifiers was created. The genes and miRNAs obtained from the analysis of interactions were then copied and pasted into a text document named, 'new_list.txt'.

10.10.1.3. All Genes and miRNAs (starting, and interacting)

A list of unique starting genes which were verified in BioMart, along with the miRNAs verified in miRBase, along with their associated targets identified in miRTarBase and miRecords, were pooled into a single file named 'big_list.txt' to create a comprehensive dataset of all the molecules in this study.

10.10.1.4. All Top Regulating miRNAs and associated genes

The text file prepared in 10.8.1, named 'all_top_regulating_miRNAs.txt' was used as the dataset file for all top regulating miRNAs identified, and the corresponding genes which they interacted with.

10.10.1.5. All Top Regulated Genes and associated miRNAs

The text file prepared in 10.8.2, named 'all_top_regulating_genes.txt' was used as the dataset file for all the top regulated genes identified from the cross-analysis. This file also contained the miRNAs which the genes interacted with.

10.10.1.6. All Top Regulated Genes and miRNAs and associations

The text file prepared in 10.9, named 'all_top_regulated.txt' consisted of all the top regulated genes, and top regulating miRNAs, as well as their interactions. This dataset was used to form the skeleton of the regulatory networks.

10.10.2. Uploading of datasets into Ingenuity Pathway Analysis

Data was analysed through the use of QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). The networks, functional analyses, and other results were generated through the use of QIAGEN's Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).

All prepared data sets were uploaded into IPA. The default Flexible Format was used as the format for data upload. For this study, the Identifier types selected were 'Gene symbol – human (Hugo / HGNC, Entrez Gene)' for the gene names, and miRBase (mature) for the miRNA IDs. Array platform used was kept at the default as 'Not specified/applicable'. A dataset summary was also provided, showing how many of the raw data entries were mapped, and unmapped in the IPA database. Unmapped identifiers, not recognized by IPA, were not included in further analyses by the software program. For all mapped IDs, the recognized Symbol, Entrez Gene name, biological location, molecule type, and associated drugs, were also reported.

10.10.3. Core Analysis

The Core Analysis in IPA was performed on all uploaded dataset files, and the following parameters were used when running the Core Analysis: the Reference Set used was the default Ingenuity Knowledge Base (Genes only), whereas I selected the relationships to consider as 'Direct' relationships only (Default: Direct and Indirect). The Interaction networks to generate were set at 70 molecules per network (Default: 35), and 10 networks per analysis (Default: 10). I kept the default of using all data sources provided, and adjusted the Confidence to include Experimentally Observed and High (predicted) data (Default: Experimentally Observed only). I selected the species to only include Human (Default: Human, Mouse, Rat, Uncharacterized), and kept the default setting of no selected tissues and cell lines. Mutations were also kept at default, in which all were selected.

10.10.3.1. Summary of Core Analysis

The summary provided on the first tab contained an overview of the outputs described below.

10.10.3.2. Canonical Pathways

The canonical pathways tab provided a default bar chart view of the biological pathways associated with the dataset on which the Core Analysis was run. In IPA, Canonical Pathways can be viewed as a bar chart, line chart, or stacked bar chart. The orientation of the chart can also be presented as horizontal or vertical. On the Customize Chart tab, score cut-off values were selected at p-value greater than 1.3, using the Benjamini-Hochberg multiple testing correction. The graph style was selected to colour by default, with a threshold value of 0.05. The sort order selected was by Observation or Time-point, and the Score was selected as the Primary Y-axis parameter. A graph showing the overlap of pathways can also be viewed and exported.

10.10.3.3. Upstream Analysis

The Upstream Analysis tab provided upstream regulators identified by IPA. The molecule type, p-value of overlap, and target molecules in the dataset was provided. This data was sorted by descending p-value, and exported in Excel format for reporting.

10.10.3.4. Diseases and Functions

The Diseases and Functions tab provided a graphical representation of the number of diseases annotated in categories, using data from the given dataset. All the data was exported in Excel format. In Excel, the data was sorted in descending order of number of molecules from my dataset found to be associated with these disease categories. Thereafter, disease categories were selected where ten or more of the molecules from my dataset were placed in these disease categories. Then, categories that had a p-value association of greater than 0.05 were removed from this list. A heat map was also generated and reported. Disease categories which appeared to be highly associated with SLE were selected for reporting.

10.10.3.5. Networks

The networks tab provided access to the networks generated by the Core Analysis. Networks were individually exported in the original image. Information regarding which molecules were placed in a generated network, along with the score, number of focus molecules, and the top diseases and functions for that network, were recorded. Focus molecules are the molecules from my input dataset and are displayed by IPA in bold font. The molecules in normal font are those added by IPA to complete the relationships between the focus molecules.

10.10.4. Selection of networks

Only networks which had more than two molecules in a network were selected as core networks for further processing. All information regarding the network formation was also recorded.

10.10.5. Viewing of original core networks

Each core network was viewed in its original output as presented by IPA. This image was then exported in PNG format, without the inclusion of the analysis details.

10.10.6. Expanding of network nodes

For each network, nodes were identified and at my own discretion, spread out to create a network image with fewer overlapping relationship lines than in the original core network, thus making it clearer for viewing.

10.10.7. Expanding of selected networks using the Grow Tool

Using the Grow tool in the Build tab, molecules which appeared to be nodes (i.e. were linked to multiple molecules, and appeared to be a core molecule) were identified. Attempts were then made to expand these nodes, as well as molecules represented in white (those molecules added in by IPA), with the dataset of All Genes and miRNAs. This allowed for the maximisation of including as many molecules from my dataset of starting and interacting genes and miRNAs.

Parameter settings for using Grow were set at adding direct interactions only, and adding a maximum of 25 molecules at a time. The dataset to be used for growing was limited to my specified **All Genes and miRNAs** dataset. Confidence level was kept at Experimentally Observed, and High (predicted), while species was selected as Human only, no Tissues and cell lines were selected, and all Mutations, and Relationship Types were kept at Default. .

Where molecules were added, IPA represented these in dark pink. The purpose of using the Grow tool was to maximise the illustration of the interactions between the molecules within the dataset being analysed. Importantly, I only grew on nodes and molecules presented by the core networks, and I did not expand further on molecules which had been added as a result of Growing. Results of addition of molecules to the core networks were manually recorded

for reporting in the methods section. Expanded networks were then exported as images in PNG format.

10.10.8. Preparation for overlays in Path Designer Mode

Once each core network from the **Top Regulated** dataset was grown to include more elements from the **All Genes and miRNAs** dataset, the expanded network was taken into the Path Designer tool in IPA. After clearing the default background, all molecules were selected. For all the molecules, the fill colour was adjusted to white, the font colour remained the default black, the line colour remained the default dim grey, and the line thickness was adjusted to the second line thickness. This image was then exported in tagged image file format (TIFF) as the image prepared for overlays.

10.10.9. Overlays on expanded core networks

The skeleton dataset networks described in 10.10.8 were then overlaid with each of the sub-datasets shown in the left hand side column of Table 3. When moving from one overlay to the next, the previous overlay was cleared. When an overlay dataset was selected from the Overlay panel on the left of the screen, the overlay was shown in grey. Each molecule in grey was then manually selected and the fill colour was changed as shown in Table 3. As an exception, for the **All Genes and miRNAs** overlay, the molecules which were not overlaid, i.e. were in white, these molecules were selected and the fill colour changed to pink. This was to highlight the molecules which had been added in by IPA to connect the focus molecules in the network.

Table 3: The image fill colours used to highlight the overlays of sub-datasets on the Core Networks generated by the Core Analysis of the skeleton dataset.

Sub-dataset which the Core Network was overlaid with:	**Fill Colour:**
All Top Regulated **Genes and miRNAs** and associations	light green
All Top Regulating **miRNAs** and associated genes	blue
All Top Regulated **Genes** and associated miRNAs	orange
All Genes and miRNAs **(starting, and interacting)**	pink
Starting Genes and miRNAs	yellow
Interacting Genes and miRNAs	light wood

10.11. Identification of possible SLE-associated genes (from Pathway Analysis) not reported in recent biomedical literature

By careful examination of the differences between the network-overlays generated, candidate genes which were not present in the complete list of starting and interacting molecules, but were interacting with numerous molecules present in my dataset, were considered to be novel candidate SLE-associated genes or miRNAs. Furthermore, molecules which were added by the IPA database, to complete the network, were also considered as potential novel candidate SLE-associated molecules.

11. RESULTS

11.1. Database selection

11.1.1. Database for miRNA verification

11.1.1.1. miRBase database

Version 21 of the miRBase database was downloaded for analyses in May 2015. The mature miRNA sequence annotation for all the species reported in this database was downloaded as the original file, 'mature.fa'. This file contained 71 656 lines, in fasta format. As fasta format entails two lines per data entry, this file contained 35 828 mature miRNA annotation entries.

A GFF3 format file containing the genome coordinates of the miRNAs in the miRBase database was downloaded in June 2015. This file contained 4 707 lines, of which the first 13 lines were descriptive for file format and data sources. Thus, this file named 'hsa.gff3' contained 4 694 lines of genome coordinate data for both mature- and precursor human miRNA sequences.

11.1.2. Database for gene name/ Ensembl Gene ID verification

11.1.2.1. Ensembl BioMart

At the time of analyses, Ensembl BioMart Release 80 was available for gene name and Ensembl Gene ID verification as HGNC symbols.

11.1.3. Databases for miRNA/gene target identification

I chose to work with miRTarBase and miRecords as the databases containing experimentally-validated miRNA-target interactions. These two databases, containing downloadable miRNA-target interactions, were the most frequently updated and curated databases amongst those described in Table 2.

11.2. Data curation

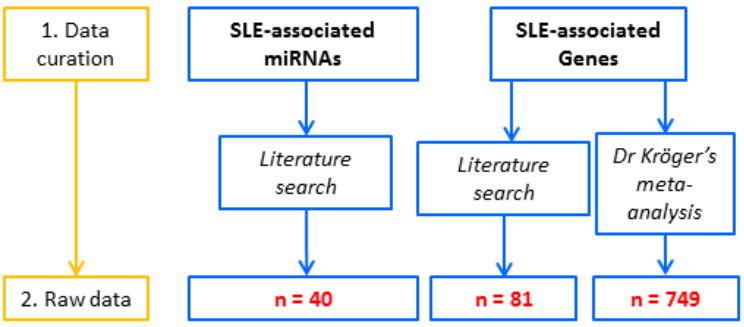

Figure 14: This figure illustrates the data curation steps used in this study, along with the results of the curation step for each data set. Raw data was collected from literature for miRNAs and genes, and a list of differentially-expressed SLE-associated genes was provided by Dr Wendy Kröger, and included in this study with her permission.

11.2.1. SLE-associated miRNAs

Through the careful perusal of recent scientific biomedical literature, 40 human miRNAs were reported to have been associated with lupus pathogenesis as shown in Figure 14, most reporting dysregulation in the various experiments performed. The forty miRNAs curated from literature and included as raw data in this study were: hsa-miR-21, hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-126, hsa-miR-141, hsa-miR-142, hsa-miR-145, hsa-miR-150, hsa-miR-155, hsa-miR-181, hsa-miR-184, hsa-miR-189, hsa-miR-198, hsa-miR-224, hsa-miR-298, hsa-miR-342, hsa-miR-383, hsa-miR-629, hsa-miR-638, hsa-miR-663, hsa-miR-3148, hsa-miR-1224-3p, hsa-miR-1224-3P, hsa-miR-125a, hsa-miR-142-3p, hsa-miR-146a, hsa-miR-148a, hsa-miR-17-5p, hsa-miR-17-92, hsa-miR-182-96-183, hsa-miR-196a, hsa-miR-23b, hsa-miR-299-3p, hsa-miR-371-5p, hsa-miR-409-3p, hsa-miR-423-5p, hsa-miR-483-5p, hsa-miR-513-5p, hsa-miR-516a-5p.

These 40 unique miRNA IDs were recorded in a text file named 'microRNAs_assoc_SLE.txt', for miRNAs which have been associated with SLE pathogenesis. The different sources for the miRNAs associated with lupus are reported in Table 4.

Table 4: The distribution of sources for the curated list of 40 miRNAs.

Source article		Number of miRNAs contributed
Lu et al. (2012)	*- Primary source*	*6*
Tiffin et al. (2013)	*- Review*	*28*
Wang et al. (2014)	*- Primary source*	*1*
Zan et al. (2014)	*- Review*	*5*

11.2.2. SLE-associated Genes: from literature

Three peer-reviewed journal articles (see Table 5) contributed to a starting list of 81 unique names of genes, as shown in Figure 14, which have been associated with lupus.

Table 5: The distribution of sources for curated list of 81 lupus-associated genes.

Source article		Number of genes contributed
Lu et al. (2012)	*- Primary source*	*1*
De Azevedo Silva et al. (2014)	*- Review*	*8*
Tiffin et al. (2013)	*- Review*	*72*

11.2.3. Differentially-expressed SLE-associated genes

A total of 749 Ensembl Gene IDs of differentially-expressed SLE-associated genes as collected by Dr Wendy Kröger was included in this study with her permission.

11.3. Preparation of Data Nomenclature

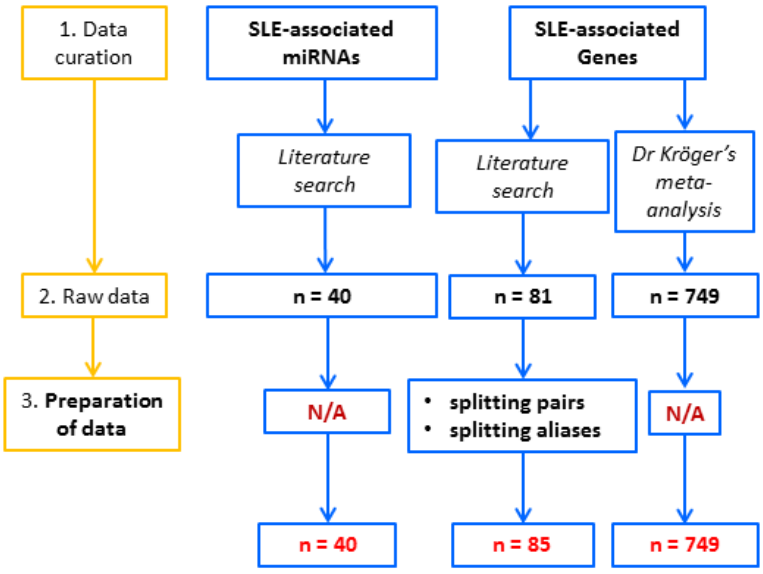

Figure 15: The orange boxes on the left describe the methods used in the preparation of data nomenclature. The blue boxes illustrate how the raw data of curated miRNAs, and the raw data of differentially-expressed SLE-associated genes did not undergo data preparation as indicated by an "N/A". SLE-associated genes curated from literature underwent nomenclature processing, thereby increasing the count of raw data from 81 to 85 genes.

11.3.1. Preparation of SLE-associated gene names

The 81 lupus-associated genes as collected from literature were: *ACTN4*, *API5*, *ATF6B (CREBL1)**, *ATG5*, *BANK1*, *BLK*, *C1QA*, *C1QB*, *C2*, *C4A*, *C4B*, *CD44*, *CRP*, *DNAJA1*, *ETS1*, *FCGR1A*, *FCGR1B*, *FCGR2A*, *FCGR2B*, *FCGR3A*, *FRK-PTPRD pair***, *HLA-DRB1*, *HLA-DRB2*, *HLA-DRB3*, *ICA1*, *IKZF1*, *IL10*, *IL21*, *IRAK1 (MECP2)**, *IRF5*, *IRF53*, *IRF8*, *ITGAM*, *JAZF1*, *KPNA1*, *LRRC18 (WDFY4)**, *Lyn*, *MECP2*, *MICB*, *NMNAT2*, *NOTCH4*, *OR2H2*, *PDCD1*, *PHRF1*, *PPP2CA*, *PRDM1 (ATG5)**, *PTN*, *PTPN22*, *PTTG1*, *PXK*, *RABGAPIL*, *RASGRP3*, *RPS6KA1*, *SCUBE1*, *SERPIND1-PARVB pair***, *SKIV2L*, *SLC15A4*, *SPP1*, *STAT1*, *STAT4*, *STK17A*, *SYK* ,*TAP2*, *TLR7* ,*TLR8*, *TLR9*, *TNFAIP3*, *TNFSF4*, *TNIP1*, *TRAF6*, *TREX1*, *TRIM27*, *TRIM31*, *TYK2*, *UBE2L3*, *UHRF1BP1*, *VDR*, *XKR6*, *XRCC1*, *XRCC3*, *XRCC4.*

The gene names indicated by an '*' are directly followed by an alias gene name enclosed in parentheses, as taken from the source article. Thus, as shown in Figure 15, the list of lupus genes was then prepared for analysis by firstly splitting the aliases in parentheses.

The gene names indicated by a '**' are joined by a hyphen, and have been noted in their literature sources actively being associated with lupus, as a pair. The second step of preparing the SLE-associated genes curated from literature was to split these paired items, as shown in Figure 15. Thus, *ATF6B* (*CREBL1*) became two separate genes to be analysed: *ATF6B*, *and CREBL1. IRAK1* (*MECP2*) became two separate genes to be analysed: *IRAK1*, *and MECP2* (which was already in this list of associated lupus genes. *LRRC18* (*WDFY4*) became *LRRC18*, and *WDFY4*. *PRDM1* (*ATG5*) became *PRDM1*, and *ATG5* (which was also already present in this list of genes. *FRK-PTPRD* pair became *FRK*, and *PTPRD*. *SERPIND1-PARVB* pair became *SERPIND1*, and *PARVB*.

On the 14th of July, I accessed the GeneCards GUI at http://www.genecards.org/ to verify that the gene in parentheses was in fact a verified alias for the accompanying gene. Searching for the main gene, not in parentheses, I found that *ATF6B* gene (GCID: GC06M032065) had 11 alias' of which *CREBL1* was one.

IRAK gene (GCID: GC0XM154010) had 7 aliases, of which *MECP2* was not listed as a known alias. *LRRC18* (GCID: 10M048909) has 5 alias', of which *WDFY4* was not listed as a known alias. *PRDM1* (GCID: GC06P106086) had 13 alias', of which *ATG5* was not listed as a known alias for this gene.

Even though these genes were not confirmed as aliases, they were included in the list of associated lupus genes, as they were collected from literature stating that they were involved in the pathogenesis of lupus. Thus, *MECP2*, *WDFY4*, and *ATG5* were flagged for attention when miRNA interactions were determined in downstream analyses. Thus, the new list which was to be used in analyses consisted of 85 gene names: *ACTN4*, *API5*, *ATF6B*, *ATG5*, *BANK1*, *BLK*, *C1QA*, *C1QB*, *C2*, *C4A*, *C4B*, *CD44*, *CREBL1*, *CRP*, *DNAJA1*, *ETS1*, *FCGR1A*, *FCGR1B*, *FCGR2A*, *FCGR2B*, *FCGR3A*, *FRK*, *HLA-DRB1*, *HLA-DRB2*, *HLA-DRB3*, *ICA1*, *IKZF1*, *IL10*, *IL21*, *IRAK1*, *IRF5*, *IRF53*, *IRF8*, *ITGAM*, *JAZF1*, *KPNA1*, *LRRC18*, *Lyn*, *MECP2*, *MICB*, *NMNAT2*, *NOTCH4*, *OR2H2*, *PARVB*, *PDCD1*, *PHRF1*, *PPP2CA*, *PRDM1*, *PTN*, *PTPN22*, *PTPRD*, *PTTG1*, *PXK*, *RABGAPIL*, *RASGRP3*, *RPS6KA1*, *SCUBE1*, *SERPIND1*, *SKIV2L*, *SLC15A4*, *SPP1*, *STAT1*, *STAT4*, *STK17A*, *SYK* ,*TAP2*, *TLR7* ,*TLR8*, *TLR9*, *TNFAIP3*, *TNFSF4*, *TNIP1*, *TRAF6*, *TREX1*, *TRIM27*, *TRIM31*, *TYK2*, *UBE2L3*, *UHRF1BP1*, *VDR*, *WDFY4*, *XKR6*, *XRCC1*, *XRCC3*, *XRCC4*.

11.4. Verification of data

11.4.1. Dataset: SLE-associated miRNAs

11.4.1.1. Python Script 1: 'PS01_June_28_miRBase_1.py'

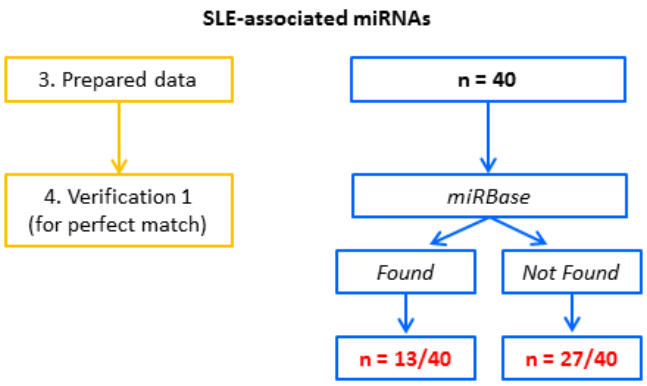

Figure 16: The numbered orange boxes describe the methods used for the first attempt at verifying the data set. The blue boxes describe how the raw miRNA data undergoes verification, producing the results of the verification attempt for the miRNAs as perfect string matches in miRBase.

From the 40 miRNAs curated from literature, 14 miRNAs were found as an exact string match in miRBase. As 14 + 27 does not equate to 40 (starting number of miRNAs), I checked for duplicate miRNA entries in both output files. In this file, I found that 'hsa-miR-516a-5p' had been found twice in the 'mirbase_ complete.txt' file. All entry information is identical (most importantly the miRNA accession number), except for the start and end bp genome coordinates. Thus, the latter miRNA entry was removed from the dataset, and the first entry was retained for further analyses. The dataset was written to a file named 'new_ perfect_match_in_miRBase.txt', and contained 13 miRNAs. These 13 miRNAs were: hsa-miR-184, hsa-miR-198, hsa-miR-298, hsa-miR-638, hsa-miR-3148, hsa-miR-1224-3p, hsa-miR-142-3p, hsa-miR-17-5p, hsa-miR-299-3p, hsa-miR-409-3p, hsa-miR-423-5p, hsa-miR-483-5p, and hsa-miR-516a-5p.

As shown in Figure 16, from the starting list of 40 miRNAs, 27 miRNAs were not found as an exact string match in the miRBase database. These 27 miRNAs were: hsa-miR-21, hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-126, hsa-miR-141, hsa-miR-142, hsa-miR-145, hsa-miR-150, hsa-miR-155, hsa-miR-181, hsa-miR-189, hsa-miR-224, hsa-miR-342, hsa-miR-383, hsa-miR-629, hsa-miR-663, hsa-miR-1224-3P, hsa-miR-125a, hsa-miR-146a, hsa-miR-148a, hsa-miR-17-92, hsa-miR-182-96-183, hsa-miR-196a, hsa-miR-23b, hsa-miR-371-5p, hsa-miR-513-5p. The 27 miRNAs listed here did not match as an exact string match to any of the miRNA IDs in the miRBase database. Thus, these miRNAs were re-evaluated using regular expression analysis in Python.

11.4.1.2. Python Script 2: 'PS02_July_01_new_regexes.py'

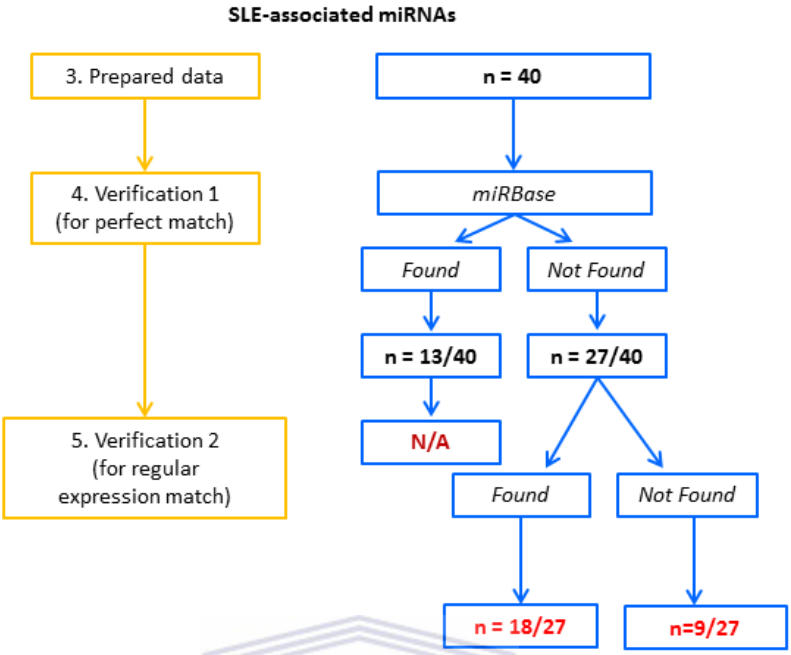

Figure 17: The orange boxes on the left describe the methods used to process the raw data through the sequential verification steps in miRBase. The blue boxes on the right describe the results of both verification steps for the raw miRNA data. miRNAs which passed the perfect string matching did not proceed to the second verification step, as indicated by "N/A". The results of the regular expression matching in miRBase are reported in red font for the miRNAs which failed perfect string matching.

Of the 27 miRNAs which did not make an exact string match to a miRNA name in miRBase, 18 miRNAs matched to at least one of the regular expressions for the miRNA names in the miRBase entries, as shown in Figure 17. The combination of these three regular expressions allowed for the capturing of all possible miRNA nomenclature derivatives in miRBase records. The 18 miRNAs queried in the regular expression miRBase script resulted in 44 regular expression matches being made, as summarized in Table 7. The results are reported in Table 6.

Table 6: Results of miRNA name matching in miRBase using Python Regular Expressions

Original miRNA	Regular expression miRNA match in miRBase	Original miRNA	Regular expression miRNA match in miRBase
hsa-miR-21	hsa-miR-21-5p hsa-miR-21-3p	hsa-miR-342	hsa-miR-342-5p hsa-miR-342-3p
hsa-miR-126	hsa-miR-126-5p hsa-miR-126-3p	hsa-miR-383	hsa-miR-383-5p hsa-miR-383-3p
hsa-miR-141	hsa-miR-141-5p hsa-miR-141-3p	hsa-miR-629	hsa-miR-629-5p hsa-miR-629-3p
hsa-miR-142	hsa-miR-142-5p hsa-miR-142-3p	hsa-miR-663	hsa-miR-663a hsa-miR-663b
hsa-miR-145	hsa-miR-145-5p hsa-miR-145-3p	hsa-miR-125a	hsa-miR-125a-5p hsa-miR-125a-3p
hsa-miR-150	hsa-miR-150-5p hsa-miR-150-3p	hsa-miR-146a	hsa-miR-146a-5p hsa-miR-146a-3p
hsa-miR-155	hsa-miR-155-5p hsa-miR-155-3p	hsa-miR-148a	hsa-miR-148a-5p hsa-miR-148a-3p
hsa-miR-181	hsa-miR-181a-5p hsa-miR-181a-3p hsa-miR-181a-2-3p hsa-miR-181b-5p hsa-miR-181b-3p hsa-miR-181b-2-3p hsa-miR-181c-5p hsa-miR-181c-3p hsa-miR-181d-5p hsa-miR-181d-3p	hsa-miR-196a	hsa-miR-196a-5p hsa-miR-196a-3p
hsa-miR-224	hsa-miR-224-5p hsa-miR-224-3p	hsa-miR-23b	hsa-miR-23b-5p hsa-miR-23b-3p

As shown in Table 6, regular expression matching detected the original miRNA name as recorded with a '-3p' or a '-5p' denoting the arm of origin on the precursor hairpin miRNA structure. Furthermore, regular expression matching generated several paralogues (miRNAs differing by only one or two nucleotides in their sequence) for the original miRNA 'hsa-miR-181'.

Of the 27 miRNAs that did not make an exact string match in miRBase with the 'PS01_June_28_miRBase_1.py' script, nine miRNAs did not match a miRBase entry using one of the three regular expressions in the 'PS02_July_01_new_ regexes.py' script. These nine miRNAs were: hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-189, hsa-miR-1224-3P, hsa-miR-17-92, hsa-miR-182-96-183, hsa-miR-371-5p, and hsa-miR-513-5p.

Each of the nine miRNAs was then searched for manually in the miRBase database using the Find tool in Excel. Eight of the nine miRNAs were confirmed to not be in the miRBase database, in any format (i.e. with a -3p, or 5p suffix). Only one miRNA, 'hsa-miR-1224-3P', was found with the manual search in miRBase as 'hsa-miR-1224-3p'. Even though the miRNA name was found in this format, the miRNA 'hsa-miR-1224-3P' was not changed and resubmitted to the script, as this was an indication of a nomenclature error from source.

Table 7 : Summary of results for checking if ID is recorded in miRBase

File	Number of miRNAs
'perfect_match_in_miRBase.txt'	n = 13 / 40
'no_match_in_miRBase.txt'	n = 27 / 40
'found_regex_match_in_mirbase.txt'	n = 18 / 27 {44 regular expression matches}
'not_found_regex_match_in_mirbase.txt'	n = 9 / 27

11.4.2. Dataset: SLE-associated gene names curated from literature

11.4.2.1. Verification of the SLE-associated gene names

On the 14th of July 2015, the 85 lupus-associated genes curated from literature, were searched for in Ensembl BioMart, to validate the existence of these genes in this repository of known, annotated genes.

11.4.2.2. Python Script 3:

'PS03_15_July_assoc_genes_prep_and_Biomart.py'

The 'ORIGINAL_assoc_SLE_genes.txt' file contained the names of 82 lupus-associated genes. Manual preparation of this genes list, by splitting paired genes, and aliases, resulted in the list being extended to contain 85 unique gene names, and this file was named 'new_list.txt' as shown as the result of Step 3 in Figure 18.

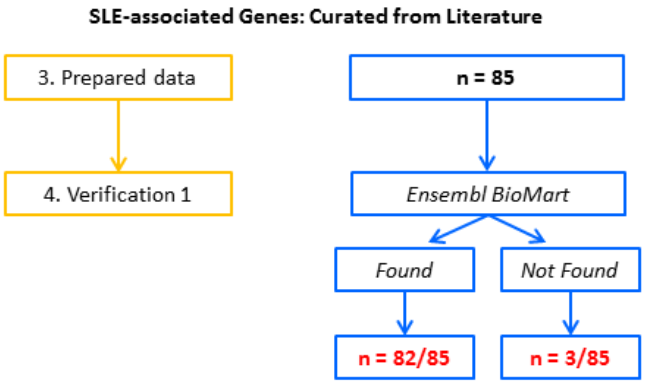

Figure 18: The numbered orange boxes describe the methods used in the verification of curated SLE-associated genes. The blue boxes report the results of the verification step for the prepared gene names in Ensembl BioMart.

Running the list of prepared, SLE-associated genes though Ensembl BioMart resulted in 163 non-unique gene matches being made in the Ensembl database. The sorting, and set() function in Python determined that 82 unique genes were matched in Ensembl BioMart as reported in Table 8.

Table 8: Distribution of the 163 gene matches made in Ensembl BioMart, for 82 SLE-associated genes curated from literature. Each of the 82 genes was matched at least once as HGNC entries in Ensembl BioMart. "LYN" was highlighted in red in the table to indicate that it was not an exact string match to the curated gene "Lyn", and required further assessment.

Associated Gene Name	Number of matches made in Ensembl	Associated Gene Name	Number of matches made in Ensembl	Associated Gene Name	Number of matches made in Ensembl
ACTN4	1	*IRAK1*	1	*SLC15A4*	1
API5	1	*IRF5*	1	*SPP1*	1
ATF6B	4	*IRF8*	2	*STAT1*	2
ATG5	1	*ITGAM*	1	*STAT4*	1
BANK1	1	*JAZF1*	1	*STK17A*	1
BLK	1	*KPNA1*	1	*SYK*	1
C1QA	2	*LRRC18*	1	*TAP2*	8
C1QB	2	*LYN*	1	*TLR7*	1
C2	8	*MECP2*	1	*TLR8*	1
C4A	5	*MICB*	8	*TLR9*	1
C4B	6	*NMNAT2*	1	*TNFAIP3*	1
CD44	1	*NOTCH4*	8	*TNFSF4*	1
CRP	1	*OR2H2*	7	*TNIP1*	1
DNAJA1	1	*PARVB*	1	*TRAF6*	1
ETS1	1	*PDCD1*	2	*TREX1*	2
FCGR1A	1	*PHRF1*	2	*TRIM27*	7
FCGR1B	2	*PPP2CA*	1	*TRIM31*	8
FCGR2A	1	*PRDM1*	1	*TYK2*	2
FCGR2B	1	*PTN*	1	*UBE2L3*	1
FCGR3A	2	*PTPN22*	1	*UHRF1BP1*	1

FRK	1	*PTPRD*	1	*VDR*	1
HLA-DRB1	6	*PTTG1*	1	*WDFY4*	1
HLA-DRB2	1	*PXK*	1	*XKR6*	1
HLA-DRB3	2	*RASGRP3*	1	*XRCC1*	1
ICA1	1	*RPS6KA1*	2	*XRCC3*	1
IKZF1	1	*SCUBE1*	1	*XRCC4*	1
IL10	1	*SERPIND1*	1		
IL21	1	*SKIV2L*	6		

The python script determined that 81 out of the 85 genes were found in BioMart, and four out of the 85 starting genes were not found. These four genes were *Lyn*, *CREBL1*, *RABGAPIL*, and *IRF53*. Because the number of unique matches found in BioMart (n = 82) did not match the filtering of the Python script on exact string matching (n = 81), I checked which of the four genes were not found using the exact string match, but were in fact found in BioMart.

I then found that because Ensembl BioMart performs string matching as not being Case-sensitive, Lyn was detected as '*LYN*', and reported as being a match in BioMart. Thus, '*LYN*', as highlighted in red in the above table, was flagged for attention, in case a match was made in the subsequent target-matching scripts.

On the 14th of July, I also manually checked in the Ensembl BioMart database at http://www.ensembl.org/BioMart/martview/e92cd50c9a75820c693577f48e739a5a to see if *CREBL1*, *IRF53*, or *RABGAPIL* were indeed not found as entries in the database. I found that *CREBL1* was one of two aliases for the *ATF6B* gene (ENSG00000213676), which is 'activating transcription factor 6 beta'. This is in concordance with the aliasing of *ATF6B* in the original list of 82 SLE genes. *CREBL1* was however, not listed as its own gene entry in Ensembl, nor was *IRF53* found with the manual search in the Ensembl BioMart database.

When searching manually for '*RABGAPIL*', this gene was not found in Ensembl. However, the top suggestion made by the database was '*rabgap1l*', a lowercase form of what was likely to have been a typographical error of this gene. Upon searching the source data, (Tiffin *et al.*, 2013), this gene was listed as '*RABGAPIL*' in Supplementary data table S-1. Referring to the source article for this table, (Ramos *et al.*, 2010) , supplementary table S1-S2, downloaded on the 15th of July

at http://www.ncbi.nlm.nih.gov/pmc/articles/-PMC3128183/bin/NIHMS279925-supplement-Supp_Table_S1-S2.doc , revealed that '*RABGAP1L'* had indeed been misspelt when carried over to Tiffin *et al.* 2013. Thus, *RABGAPIL* was removed from the list of 85 lupus-associated genes, and *RABGAP1L* was added. This genes list was recorded as 'new_list2.txt'.

The text file, 'new_list2.txt', was then rerun through Ensembl BioMart, and subsequently the Python script '15_July_assoc_genes_prep_and_BioMart.py', to identify which of the genes were found as entries in BioMart, with being aware of the capitalization of '*Lyn'* as '*LYN'* in the BioMart match, and the updated spelling of '*RABGAP1L'*. Of the 85 starting genes, 82 genes were found to be exact string matches in the output of BioMart, 'mart_export2.txt'. The 82 genes were written to the output file 'BioMart_assoc_gene_MATCH_2.txt'. A sorted list was made in this Python script, and the first occurrence of each gene match was recorded and written out with all the selected accompanying BioMart information to the output file 'BioMart_assoc_gene_MATCH_All_Info_2.txt'. This data was used to compile Table 9.

Table 9: The 82 SLE-associated genes curated from literature which were found in BioMart, and were retained for further analyses

Associated Gene Name	First Ensembl ID output	Gene start bp	Gene end bp
ACTN4	ENSG00000130402	38647649	38731583
API5	ENSG00000166181	43311963	43344529
ATF6B	ENSG00000228628	32191691	32204680
ATG5	ENSG000000057663	106184476	106325820
BANK1	ENSG00000153064	101411286	102074812
BLK	ENSG00000136573	11494001	11564604
C1QA	ENSG00000173372	22636506	22639608
C1QB	ENSG00000173369	22652762	22661538
C2	ENSG00000206372	31888164	31936056
C4A	ENSG00000206340	31972402	31993060
C4B	ENSG00000228454	32007416	32028073
CD44	ENSG00000026508	35138870	35232402
CRP	ENSG00000132693	159712289	159714589
DNAJA1	ENSG00000086061	33025211	33039907
ETS1	ENSG00000134954	128458761	128587558

FCGR1A	ENSG00000150337	149782671	149792518
FCGR1B	ENSG00000198019	121087345	121096310
FCGR2A	ENSG00000143226	161505430	161524013
FCGR2B	ENSG00000072694	161581311	161678654
FCGR3A	ENSG00000203747	161541759	161550737
FRK	ENSG00000111816	115931149	116060758
HLA-DRB1	ENSG00000206306	32449770	32595320
HLA-DRB2	ENSG00000227442	32488465	32502952
HLA-DRB3	ENSG00000231679	32444128	32521486
ICA1	ENSG00000003147	8113184	8262687
IKZF1	ENSG00000185811	50304124	50405101
IL10	ENSG00000136634	206767602	206772494
IL21	ENSG00000138684	122612628	122621069
IRAK1	ENSG00000184216	154010500	154019980
IRF5	ENSG00000128604	128937612	128950035
IRF8	ENSG00000140968	85898803	85922609
ITGAM	ENSG00000169896	31259990	31332892
JAZF1	ENSG00000153814	27830573	28180743
KPNA1	ENSG00000114030	122421949	122514945
LRRC18	ENSG00000165383	48909483	48935190
MECP2	ENSG00000169057	154021573	154137103
MICB	ENSG00000231179	31561856	31578081
NMNAT2	ENSG00000157064	183248237	183418602
NOTCH4	ENSG00000204301	32194843	32224067
OR2H2	ENSG00000229680	29587411	29588994
PARVB	ENSG00000188677	43999211	44172949
PDCD1	ENSG00000276977	241849881	241858908
PHRF1	ENSG00000274780	576461	612472
PPP2CA	ENSG00000113575	134194334	134226142
PRDM1	ENSG00000057657	106086320	106109939
PTN	ENSG00000105894	137227341	137343865
PTPN22	ENSG00000134242	113813811	113871759
PTPRD	ENSG00000153707	8314246	10612723
PTTG1	ENSG00000164611	160421822	160428744
PXK	ENSG00000168297	58332880	58426126
RABGAP1L	ENSG00000152061	174159410	174995308
RASGRP3	ENSG00000152689	33436324	33564750
RPS6KA1	ENSG00000117676	26529761	26575030
SCUBE1	ENSG00000159307	43197283	43343388
SERPIND1	ENSG00000099937	20773879	20787720
SKIV2L	ENSG00000204351	31959080	31969755
SLC15A4	ENSG00000139370	128793191	128823983
SPP1	ENSG00000118785	87975650	87983426

STAT1	ENSG00000115415	190964358	191020960
STAT4	ENSG00000138378	191029576	191151596
STK17A	ENSG00000164543	43582758	43626786
SYK	ENSG00000165025	90801787	90898549
TAP2	ENSG00000225967	32861254	32878050
TLR7	ENSG00000196664	12867083	12890380
TLR8	ENSG00000101916	12906620	12923169
TLR9	ENSG00000239732	52221080	52226163
TNFAIP3	NSG00000118503	137867188	137883312
TNFSF4	ENSG00000117586	173183734	173207313
TNIP1	ENSG00000145901	151029945	151093577
TRAF6	ENSG00000175104	36487027	36510272
TREX1	ENSG00000280804	48446710	48467645
TRIM27	ENSG00000204713	28903002	28923989
TRIM31	ENSG00000233573	30093221	30103423
TYK2	ENSG00000105397	10350529	10380676
UBE2L3	ENSG00000185651	21549447	21624034
UHRF1BP1	ENSG00000065060	34792015	34883138
VDR	ENSG00000111424	47841537	47943048
WDFY4	ENSG00000128815	48684876	48982956
XKR6	ENSG00000171044	10896045	11201366
XRCC1	ENSG00000073050	43543040	43580473
XRCC3	ENSG00000126215	103697609	103715504
XRCC4	ENSG00000152422	83077498	83353787

This run of the script also revealed that three genes were not matched in BioMart. These genes were CREBL1, Lyn, and IRF53 and were written to the output file 'Not_Found_assoc_genes_in_BioMart_2.txt'. As previously reported, manual checking in Ensembl confirmed that these three genes were not present as entries in this database of curated genes.

11.4.3. Dataset: Differentially-expressed SLE-associated gene IDs

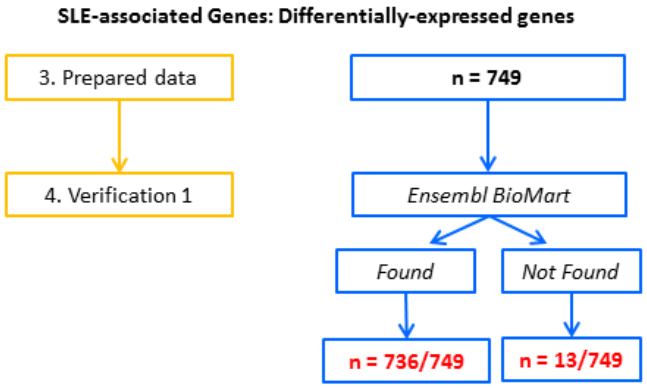

Figure 19: The orange boxes on the left describe the methods used to verify the curated data of differentially-expressed genes. The blue boxes on the right describe the results of verifying the Ensembl gene IDs in the Ensembl BioMart database. Most of the curated gene IDs in this list was found as HGNC symbols in BioMart, and only 13 were reported as deprecated IDs.

11.4.3.1. Verification of the existence of the differentially-expressed lupus associated gene IDs

On the 16th of June 2015, 749 Ensembl Gene IDs were searched for as HGNC symbols in Ensembl BioMart to validate the existence of these Ensembl IDs in this repository of known, annotated genes. Of the 749 differentially-expressed genes, 736 genes were found as HGNC symbols in BioMart, and 13 Ensembl Gene IDs were flagged as deprecated IDs. The results of the verification step are reported in Figure 19.

11.4.3.2. Python Script 4: 'PS04_16_July_DE_genes_thru_BioMart.py'

The 'EnsemblGeneID.txt' file contained the Ensembl gene IDs of 749 differentially-expressed lupus-associated genes. Of the 749 Ensembl Gene IDs submitted for verification as records, 1151 non-unique matches were made in BioMart. Using the sort() function in Python, a sorted list was generated, and furthermore, code was included to retain only the entry of the first occurrence of each Ensembl gene ID matched. This sorted list was written to 'ensembl_list_new.txt'. With sorting, and filtering, 1151 non-unique ID matches was reduced to 736 unique ID matches in BioMart.

In concordance with the number of unique gene IDs which were found in BioMart (n = 736), the remaining 13 Ensembl genes which were not found in BioMart, were appropriately written out to the text file named 'Not_Found_DE_genes_in_BioMart.txt'. These 13 differentially-expressed gene IDs are listed in Table 10.

Table 10: The 13 differentially-expressed gene IDs which were reported as being removed from the BioMart database

ENSG00000108294	ENSG00000203813
ENSG00000155130	ENSG00000211630
ENSG00000168242	ENSG00000235604
ENSG00000171282	ENSG00000236249
ENSG00000185044	ENSG00000256018
ENSG00000198366	ENSG00000262556
ENSG00000198518	

Upon manually searching in the BioMart GUI for each of these 13 Ensembl gene IDs, it was found that these were now listed as deprecated IDs and had been removed from the latest version of the database.

11.5. Target Identification

11.5.1. Identification of gene targets for SLE-associated miRNAs

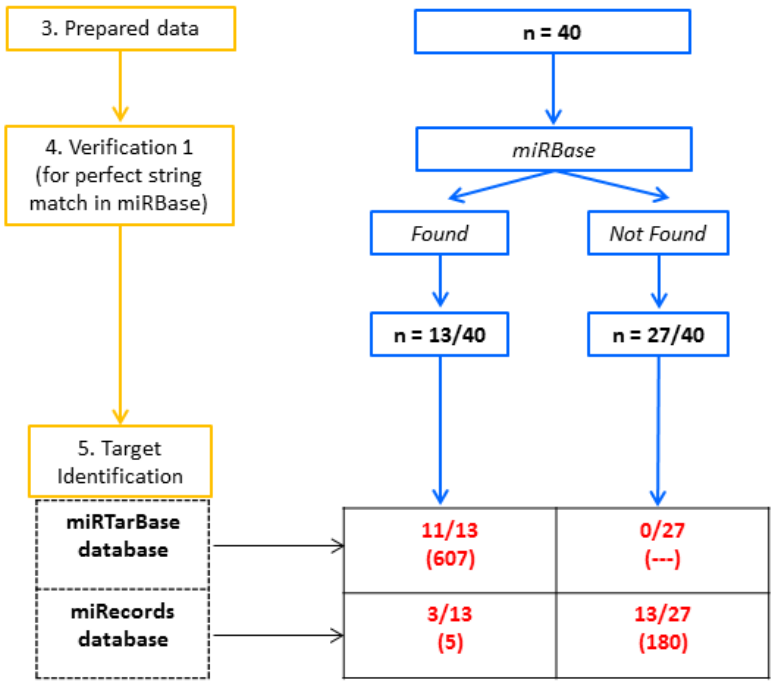

Figure 20: This figure describes the processing of the curated SLE-associated miRNAs, starting with data curation, proceeding to target identification, and lastly identifying which genes which were found to interact with the curated data. A (---) indicates that no tally of interactions was possible, due to no targets being found.

11.5.1.1. Python Script 5: 'PS05_05_July_miRTarBase_miRecords_for_PERFECT_MATCH.py'

11.5.1.1.1. miRTarBase

Of the 13 miRNAs found as an exact string match in miRBase, 11 were found as exact string matches, and having corresponding gene targets, in the experimentally validated targets database miRTarBase. These 11 miRNAs collectively matched to 607 gene targets, as shown in Figure 21. Thus, 607 miRNA-gene interactions for these 11 miRNAs were detected in the miRTarBase database, as shown in Table 11. These results were written to the output file 'found_miRNA1_ALL_targets_in_miRTarBase.txt'.

Table 11: The 11 miRNAs found in miRTarBase, and their associated gene targets. A total of 607 MTIs are reported for 11 miRNAs. Shown below is an abridged list of these 11 miRNAs, for which ten or less interacting genes were found. The full list can be found in ADDENDUM B.

miRNA	Number of gene targets	Gene targets
hsa-miR-184	3	*AKT2*, *INPPL1*, *NFATC2*
hsa-miR-198	3	*CCNT1*, *NTRK3*, *MYB*
hsa-miR-298	2	*BACE1*, *CDKN1A*
hsa-miR-638	1	*OSCP1*
hsa-miR-1224-3p	5	*KMT2D*, *IGSF3*, *TCF12*, *HNRNPA3*, *CDC42EP1*
hsa-miR-142-3p	107	
hsa-miR-17-5p	310	
hsa-miR-299-3p	1	*FXN*
hsa-miR-409-3p	6	*PHF10*, *FGB*, *FGA*, *FGG*, *ANG*, *IFNG*
hsa-miR-423-5p	168	
hsa-miR-483-5p	1	*MAPK3*

Two of the 13 miRNAs queried were not found in the miRTarBase database. These two miRNAs are 'hsa-miR-3148' and 'hsa-miR-516a-5p', and were written to 'NOT_found_miRNA1_ALL_targets_in_miRTarBase.txt'. A manual check in the downloaded miRTarBase database dataset, 'hsa_MTI.xls', revealed that those two miRNAs were indeed not listed in the miRTarBase database for human miRNA-gene interactions.

11.5.1.1.2. miRecords

Of the 13 miRNAs, for whom their existence was validated in the miRBase registry, only three miRNAs were identified in the miRecords database as having interacting gene targets. These results were written to the text file 'found_miRNA1_ALL_targets_in_miRecords.txt'. Collectively, these three miRNAs regulated five genes as shown in Table 12.

Table 12: The 3 miRNAs found in miRecords, and their associated gene targets.

miRNA	Number of gene targets	Gene targets
hsa-miR-516a-5p	2	*KLK10*, *SULF1*
hsa-miR-184	2	*AKT2*, *NFAT1*
hsa-miR-17-5p	1	*IL-8*

The remaining ten miRNAs from this list which were not listed in miRecords as having corresponding gene targets were: hsa-miR-198, hsa-miR-298, hsa-miR-638, hsa-miR-3148, hsa-miR-1224-3p, hsa-miR-142-3p, hsa-miR-299-3p, hsa-miR-409-3p, hsa-miR-423-5p, and, hsa-miR-483-5p. These miRNA names were written to 'NOT_found_miRNA1_ALL_targets _in_miRecords.txt'.

11.5.1.2. *Python Script 6: 'PS06_05_July_miRTarBase_miRecords_for_NOT_FOUND_IN_MIRBASE.py'*

11.5.1.2.1. miRTarBase

Of the 27 miRNAs that were not found as an exact string match in miRBase, none of these miRNAs were found as miRTarBase entries, and thus all were written out to the text file 'NOT_found_miRNA0_ ALL_targets_in_miRTarBase.txt'. These miRNAs were: hsa-miR-21, hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-126, hsa-miR-141, hsa-miR-142, hsa-miR-145, hsa-miR-150, hsa-miR-155, hsa-miR-181, hsa-miR-189, hsa-miR-224, hsa-miR-342, hsa-miR-383, hsa-miR-629, hsa-miR-663, hsa-miR-1224-3P, hsa-miR-125a, hsa-miR-146a, hsa-miR-148a, hsa-miR-17-92, hsa-miR-182-96-183, hsa-miR-196a, hsa-miR-23b, hsa-miR-371-5p, hsa-miR-513-5p.

11.5.1.2.2. miRecords

Of the 27 miRNAs, for whom their existence was not validated in the miRBase registry, 13 miRNAs were identified in the miRecords database as having interacting gene targets, as shown in Table 13. Collectively, these 13 miRNAs regulated 180 target genes.

Table 13: The 13 miRNAs found in miRecords, and their associated gene targets. Shown below is an abridged list of these 13 miRNAs, for which 10 or less interacting genes were found. The full list which was written to the text file 'found_miRNA0_ALL_ targets_in_ miRecords.txt' can be found in ADDENDUM C.

miRNA	Number of gene targets	Gene targets
hsa-miR-21	49	
hsa-miR-126	6	*VCAM1*, *PIK3R2*, *CRKL*, *IRS1*, *VEGFA*, *TOM1*
hsa-miR-141	13	
hsa-miR-145	19	
hsa-miR-150	7	*gag-pol*, *cmyb*, *MYB*, *CCNE1*, *P2RX7*, *EGR2*, *MUC4*
hsa-miR-155	24	
hsa-miR-224	4	*API5*, *KLK10*, *KLK1*, *AP2M1*
hsa-miR-342	1	*DNMT1*
hsa-miR-146a	38	
hsa-miR-148a	6	*DNMT3B*, *NR1I2*, *Rps6ka5*, *DNMT1*, *MSK1*, *CDNK1B*,
hsa-miR-17-92	2	*PTPRO*, *HIF1A*
hsa-miR-196a	9	*HOXB8*, *HOXC8*, *HOXD8*, *HOXA7*, *IKBKB*, *ANXA1*, *KRT5*, *SPRR2C*, *S100A9*
hsa-miR-23b	2	*PLAU*, *MET*

The remaining 14 miRNAs from this list which were not listed in miRecords as having corresponding gene targets were written out to a text file named 'NOT_found_miRNA0_ALL_targets_in_miRecords.txt'. They were: hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-142, hsa-miR-181, hsa-miR-189, hsa-miR-383, hsa-miR-629, hsa-miR-663, hsa-miR-1224-3P, hsa-miR-125a, hsa-miR-182-96-183, hsa-miR-371-5p, and hsa-miR-513-5p. Selections of these were manually checked to confirm these findings.

11.5.1.3. *Python Script 7: 'PS07_ 05_July_miRTarBase_miRecords_for_FOUND_ REGEX_MATCH.py'*

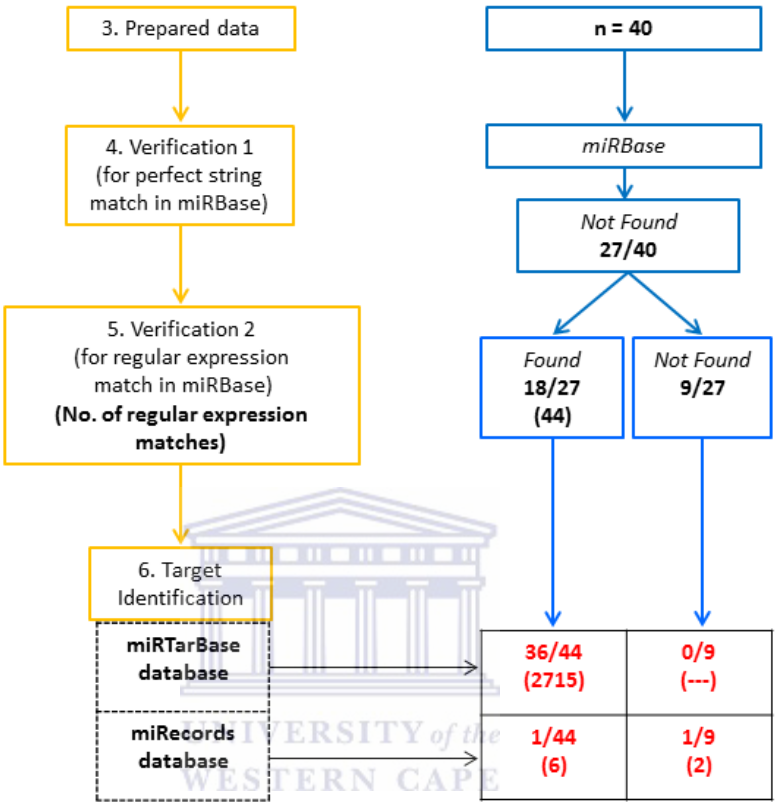

Figure 21: The orange boxes describe the sequential processing of data from verification as a regular expression match in miRBase to the Target Identification in two curated MTI databases. The blue boxes on the right of the figure report the results of the target identification for the two datasets created as result of the second verification step. These miRNAs which were either matched as a regular expression string match, or were not matched as a regular expression, were then assessed for target identification in miRTarBase and miRecords. The number of miRNAs found in both miRTarBase, and miRecords were recorded, along with the total number of interactions found (in parentheses) for that dataset. Where targets were not found in a database, a "---" was used to indicate that no count of interactions could be derived.

11.5.1.3.1. miRTarBase

Of the 44 miRNAs that were found as regular expression string matches in miRBase, 36 miRNAs were found as exact string matches and having corresponding gene targets in the experimentally validated targets database miRTarBase. These 36 miRNAs collectively matched to 2715 gene targets. Thus, 2715 miRNA-gene interactions for these 36 miRNAs were detected in the miRTarBase database and are reported in Table 14.

Table 14: The 36 miRNAs found in miRTarBase, and their associated gene targets. Shown below is an abridged list of these 36 miRNAs, for which 10 or less interacting genes were found. The full list can be found in ADDENDUM D.

miRNA	Number of gene targets	Gene targets
hsa-miR-21-5p	489	
hsa-miR-21-3p	2	*SUV420H1*, *MED25*
hsa-miR-126-5p	6	*SLC45A3*, *PTPN7*, *ADAM9*, *MMP7*, *CXCL12*, *CSGALNACT1*
hsa-miR-126-3p	30	
hsa-miR-141-3p	37	
hsa-miR-142-5p	1	*NFE2L2*
hsa-miR-142-3p	107	
hsa-miR-145-5p	89	
hsa-miR-150-5p	15	
hsa-miR-155-5p	717	
hsa-miR-155-3p	1	*IRAK3*
hsa-miR-181a-5p	237	
hsa-miR-181a-2-3p	1	*ARID1A*
hsa-miR-181b-5p	97	
hsa-miR-181c-5p	17	
hsa-miR-181c-3p	1	*MMP25*
hsa-miR-181a-3p	2	*NANOG*, *HOOK3*
hsa-miR-181d-5p	25	
hsa-miR-224-5p	17	
hsa-miR-342-5p	29	
hsa-miR-342-3p	83	
hsa-miR-383-5p	4	*IRF1*, *VEGFA*, *IGF1R*, *PRDX3*
hsa-miR-629-5p	4	*HNF4A*,*HIST1H2AC*,*ATP5G2*, *ZCCHC6*
hsa-miR-629-3p	2	*DDX3X*, *ATP5B*
hsa-miR-663a	5	*JUNB*, *JUNB*, *TCEAL1*, *HSPG2*, *CDK1*
hsa-miR-663b	1	*HRASLS5*
hsa-miR-125a-5p	137	
hsa-miR-125a-3p	10	*IP6K2*, *PFAS*, *TTC8*, *ARFGEF1*, *E2F8*, *GLUL*, *LEPRE1*, *IGF2BP2*, *NUP62*, *ATP6V1B2*
hsa-miR-146a-5p	66	
hsa-miR-146a-3p	1	*PLXNA1*
hsa-miR-148a-5p	2	*GALK2*, *ACTR1A*
hsa-miR-148a-3p	81	
hsa-miR-196a-5p	222	
hsa-miR-196a-3p	1	*RPS26*
hsa-miR-23b-5p	3	*PRODH*, *SPEN*, *EIF4B*
hsa-miR-23b-3p	173	

11.5.1.3.2. miRecords

Of the 44 miRNAs for whom their existence was validated in the miRBase registry as matching to a regular expression, only one miRNA was identified in the miRecords database as having interacting gene targets. This miRNA had 6 gene targets in the miRecords database and they are listed in Table 15.

Table 15: The single miRNA found in miRecords, and its associated gene targets

miRNA	Number of gene targets	Gene targets
hsa-miR-125a-5p	6	*ERBB2*, *ERBB3*, *LIN28*, *ARID3B*, *TP53*, *HuR*

The other 43 miRNAs matched with a regular expression in miRBase, not listed in miRecords as having corresponding gene targets were: hsa-miR-21-5p, hsa-miR-21-3p, hsa-miR-126-5p, hsa-miR-126-3p, hsa-miR-141-5p, hsa-miR-141-3p, hsa-miR-142-5p, hsa-miR-142-3p, hsa-miR-145-5p, hsa-miR-145-3p, hsa-miR-150-5p, hsa-miR-150-3p, hsa-miR-155-5p, hsa-miR-155-3p, hsa-miR-181a-5p, hsa-miR-181a-2-3p, hsa-miR-181b-5p, hsa-miR-181b-3p, hsa-miR-181c-5p, hsa-miR-181c-3p, hsa-miR-181a-3p, hsa-miR-181b-2-3p, hsa-miR-181d-5p, hsa-miR-181d-3p, hsa-miR-224-5p, hsa-miR-224-3p, hsa-miR-342-5p, hsa-miR-342-3p, hsa-miR-383-5p, hsa-miR-383-3p, hsa-miR-629-5p, hsa-miR-629-3p, hsa-miR-663a, hsa-miR-663b, hsa-miR-125a-3p, hsa-miR-146a-5p, hsa-miR-146a-3p, hsa-miR-148a-5p, hsa-miR-148a-3p, hsa-miR-196a-5p, hsa-miR-196a-3p, hsa-miR-23b-5p, hsa-miR-23b-3p. Selections of these were manually checked to confirm these findings.

11.5.1.4. Python Script 8: 'PS08_05_July_miRTarBase_miRecords_for_NOT_FOUND_REGEX_MATCH.py'

11.5.1.4.1. miRTarBase

Of the nine miRNAs that were not found as an exact string match, or regular expression string match in miRBase, none of these miRNAs were found in the miRTarBase database and thus no corresponding gene targets were established. These nine miRNAs were: hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-189,

hsa-miR-1224-3P, hsa-miR-17-92, hsa-miR-182-96-183, hsa-miR-371-5p, hsa-miR-513-5p.

A manual check of these miRNAs revealed that all nine miRNAs were not present in the miRTarBase database. However, 'hsa-miR-1224-3P' was present in the miRTarBase database as 'hsa-miR-1224-3p' with a lowercase –p suffix. Because 5 gene targets (*KMT2D*, *IGSF3*, *TCF12*, *HNRNPA3*, *CDC42EP1*) were found for 'hsa-miR-1224-3p', this miRNA was flagged for checking in source literature, and not resubmitted to the miRTarBase database as 'hsa-miR-1224-3p'.

11.5.1.4.2. miRecords

Of the nine miRNAs, for whom their existence was not validated as a regular expression string match in the miRBase registry, one miRNA was identified in the miRecords database as having two interacting gene targets as shown in Table 16.

Table 16: Of the 40 curated miRNAs, nine miRNAs were not verified in the miRBase database. Of these 9 miRNAs, only one miRNA was found in the miRecords database, having two gene targets.

miRNA	Number of gene targets	Gene targets
hsa-miR-17-92	2	PTPRO, HIF1A

The remaining eight miRNAs from this list which were not listed in miRecords as having corresponding gene targets were: hsa-miR-61, hsa-miR-78, hsa-miR-112, hsa-miR-189, hsa-miR-1224-3P, hsa-miR-182-96-183, hsa-miR-371-5p, hsa-miR-513-5p.

11.5.2. Identification of miRNAs interacting with SLE-associated genes

11.5.2.1. Python Script 9: 'PS09_15_July_Genes1_thru_DBs.py'

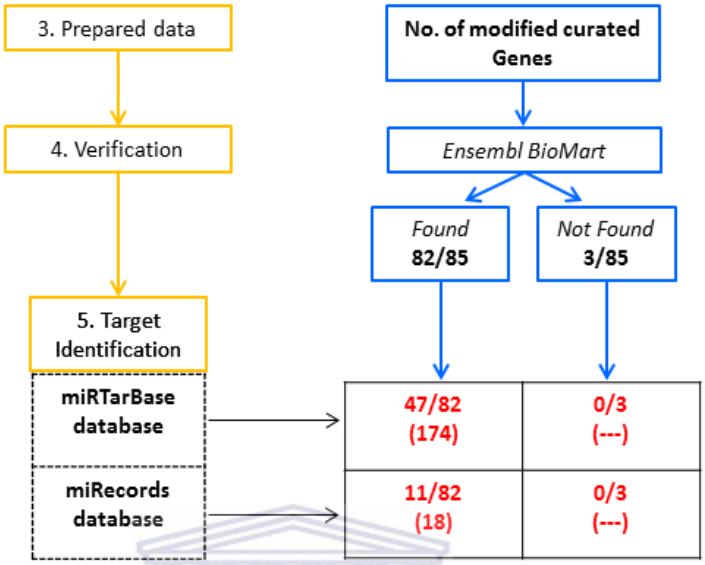

Figure 22: The orange boxes on the left describe the sequential processing of verified data to the target identification. The blue boxes describe how the associated gene names which were "Found" and "Not Found" in Ensembl BioMart underwent target identification in miRTarBase and miRecords. For both datasets, the number of genes found in each database was reported, along with the count of the interactions reported in parentheses. Where no genes were found in miRTarBase, or miRecords, a "---" indicated that no count of miRNA-Target Interactions could be reported.

11.5.2.1.1. miRTarBase

From the 83 genes which were found as non-case sensitive matches in the BioMart search, 82 genes were found as case-sensitive, exact string matches in the filtering Python script 'PS09_15_July_Genes1_thru_DBs.py'.

Of these 82 genes, 47 genes were found in the miRTarBase database. These 47 genes collectively matched to 174 interacting miRNAs. Thus, 174 miRNA-gene interactions for these 47 genes were detected in the miRTarBase database and are reported in Table 17.

Of the 82 lupus-associated genes, 35 genes were not identified in miRTarBase to have target interactions with miRNAs. These genes were: *ATG5*, *BANK1*, *BLK*, *C4A*, *C4B*, *CRP*, *FCGR1A*, *FCGR2A*, *FCGR2B*, *FCGR3A*, *HLA-DRB1*, *HLA-DRB2*, *HLA-DRB3*, *ICA1*, *IL21*, *IRF8*, *ITGAM*, *LRRC18*, *MICB*, *OR2H2*, *PARVB*, *PDCD1*, *PTN*, *SCUBE1*, *STAT4*, *STK17A*, *TAP2*, *TLR7*, *TLR8*, *TLR9*, *TNFSF4*, *TREX1*, *TRIM31*, *XRCC3*, *XRCC4*, and were written to a text file named 'NOT_found_Genes1_targets _in_miRTarBase.txt'.

Table 17: A complete list 47 lupus-associated genes curated from literature, and verified in Ensembl BioMart along with the corresponding interacting miRNAs identified in miRTarBase.

Gene	Number of interacting miRNAs	Interacting miRNAs
API5	2	hsa-miR-224-5p, hsa-miR-183-5p
ATF6B	1	hsa-miR-197-3p
C1QA	1	hsa-miR-335-5p
C1QB	1	hsa-miR-26b-5p
C2	2	hsa-miR-335-5p, hsa-miR-142-3p
DNAJA1	3	hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-193b-3p
ETS1	10	hsa-miR-125b-5p, hsa-miR-9-5p, hsa-miR-222-3p, hsa-miR-200b-3p, hsa-miR-31-5p, hsa-miR-155-5p, hsa-miR-193b-3p, hsa-miR-208a-3p, hsa-miR-10a-5p, hsa-miR-30c-5p
FCGR1B	1	hsa-miR-26b-5p
FRK	1	hsa-miR-335-5p
IKZF1	1	hsa-miR-26b-5p
IL10	1	hsa-miR-106a-5p
IRF5	1	hsa-miR-22-3p
JAZF1	2	hsa-miR-31-5p, hsa-miR-96-5p
KPNA1	5	hsa-miR-34a-5p, hsa-miR-103a-3p, hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-186-5p
NMNAT2	3	hsa-miR-9-5p, hsa-miR-122-5p, hsa-miR-615-3p
NOTCH4	3	hsa-miR-34c-5p, hsa-miR-181c-5p, hsa-miR-18a-3p
PHRF1	4	hsa-miR-1229-3p, hsa-miR-92b-3p, hsa-miR-193b-3p, hsa-miR-222-3p
PPP2CA	2	hsa-miR-340-3p, hsa-miR-197-3p
PRDM1	5	hsa-miR-9-5p, hsa-miR-127-3p, hsa-miR-125b-5p, hsa-miR-30a-5p, hsa-miR-877-3p
PTPN22	2	hsa-miR-181a-5p, hsa-miR-1
PTPRD	6	hsa-miR-429, hsa-miR-200a-3p, hsa-miR-200c-3p, hsa-miR-141-3p, hsa-miR-1, hsa-miR-200b-3p
PTTG1	5	hsa-miR-423-5p, hsa-miR-320a, hsa-miR-186-5p, hsa-miR-26a-5p, hsa-miR-17-5p
PXK	1	hsa-miR-744-5p

RABGAP1L	1	hsa-miR-1
RASGRP3	1	hsa-miR-21-5p
RPS6KA1	2	hsa-miR-125b-5p, hsa-miR-193b-3p
SERPIND1	1	hsa-miR-335-5p
SKIV2L	1	hsa-miR-222-3p
SLC15A4	3	hsa-miR-124-3p, hsa-miR-151a-5p, hsa-miR-186-5p
SPP1	3	hsa-miR-299-5p, hsa-miR-335-5p, hsa-miR-146a-5p
STAT1	4	hsa-miR-145-5p, hsa-miR-146a-5p, hsa-miR-34a-5p, hsa-miR-615-3p
SYK	3	hsa-miR-99b-3p, hsa-miR-615-3p, hsa-miR-331-3p
TNFAIP3	6	hsa-miR-125a-5p, hsa-miR-125b-5p, hsa-miR-29a-3p, hsa-miR-21-5p, hsa-miR-26b-5p, hsa-miR-24-3p
TNIP1	3	hsa-miR-221-3p, hsa-miR-324-3p, hsa-miR-181a-5p
TRAF6	2	hsa-miR-146a-5p, hsa-miR-146b-5p
TRIM27	2	hsa-miR-421, hsa-miR-1260b
TYK2	1	hsa-miR-124-3p
UBE2L3	3	hsa-miR-122-5p, hsa-miR-423-5p, hsa-miR-93-5p
UHRF1BP1	3	hsa-miR-34a-5p, hsa-miR-192-5p, hsa-miR-652-3p
VDR	2	hsa-miR-125b-5p, hsa-miR-27b-3p
WDFY4	1	hsa-miR-335-5p
XKR6	1	hsa-miR-335-5p
XRCC1	4	hsa-miR-34a-5p, hsa-miR-92b-3p, hsa-miR-193b-3p, hsa-miR-186-5p

11.5.2.1.2. miRecords

The text file 'BioMart_assoc_gene_MATCH_2.txt' was then also processed through the miRecords database, to identify gene-miRNA interactions in this validated-targets database. Of the 82 genes which were validated in both Ensembl BioMart search, and the Python script filtering step, 11 genes were identified in the miRecords database as being regulated by 18 unique miRNAs, as shown in Table 18.

The remaining 71 genes from this list which were not found in the miRecords database as having interactions with miRNAs were: *ACTN4*, *ATF6B*, *ATG5*, *BANK1*, *BLK*, *C1QA*, *C1QB*, *C2*, *C4A*, *C4B*, *CRP*, *DNAJA1*, *FCGR1A*, *FCGR1B*, *FCGR2A*, *FCGR2B*, *FCGR3A*, *FRK*, *HLA-DRB1*, *HLA-DRB2*, *HLA-DRB3*, *ICA1*, *IKZF1*, *IL10*, *IL21*, *IRF5*, *IRF8*, *ITGAM*, *JAZF1*, *KPNA1*, *LRRC18*, *MECP2*, *MICB*, *NMNAT2*, *OR2H2*, *PARVB*, *PDCD1*, *PHRF1*, *PPP2CA*, *PTN*, *PTPN22*, *PTTG1*, *PXK*, *RASGRP3*, *RPS6KA1*, *SCUBE1*, *SERPIND1*, *SKIV2L*, *SPP1*, *STAT4*, *STK17A*, *SYK*, *TAP2* ,*TLR7*, *TLR8*, *TLR9*, *TNFAIP3*, *TNFSF4*, *TNIP1*, *TREX1*, *TRIM27*, *TRIM31*, *TYK2*, *UBE2L3*, *UHRF1BP1*,

VDR, *WDFY4*, *XKR6*, *XRCC1*, *XRCC3*, *XRCC4*. These results were written to the text file named 'found_Genes1_ALL_targets_in_miRecords.txt'.

Table 18: The 11 genes from the differentially-expressed SLE-associated genes dataset, for which interacting miRNAs were found in miRecords.

Gene	Number of interacting miRNAs	Interacting miRNAs
API5	1	hsa-miR-224
CD44	2	hsa-miR-373, hsa-miR-34a
ETS1	1	hsa-miR-155
IRAK1	1	hsa-miR-146a
NOTCH4	3	hsa-miR-34b, hsa-miR-34c, hsa-miR-181c
PRDM1	1	hsa-miR-9
PTPRD	5	hsa-miR-141, hsa-miR-200a, hsa-miR-200b, hsa-miR-200c, hsa-miR-429
RABGAP1L	1	hsa-miR-1
SLC15A4	1	hsa-miR-124
STAT1	1	hsa-miR-146a
TRAF6	1	hsa-miR-146a

11.5.2.2. Python Script 10: 'PS10_15_July_Genes2_thru_DBs.py'

11.5.2.2.1. miRTarBase

No gene interactions were found for the 3 genes '*Lyn*', '*CREBL1*', '*IRF53*'. These three genes were curated from literature sources, but were not verified in Ensembl BioMart to exist as Ensembl entries. Thus, these gene names were written to the file 'NOT_found_Genes2_targets_in _miRTarBase.txt'.

11.5.2.2.2. miRecords

No gene interactions were found for '*Lyn*', '*CREBL1*', or '*IRF53*' in miRecords, and thus this output file contained no entries. These three genes '*Lyn*', '*CREBL1*', '*IRF53*' were written to the file 'NOT_found_Genes2_targets_in_miRecords.txt'. A manual check using the Find tool in Excel confirmed that these three genes were not present in the miRecords database.

11.5.3. Identification of miRNAs interacting with the Differentially-expressed lupus-associated genes

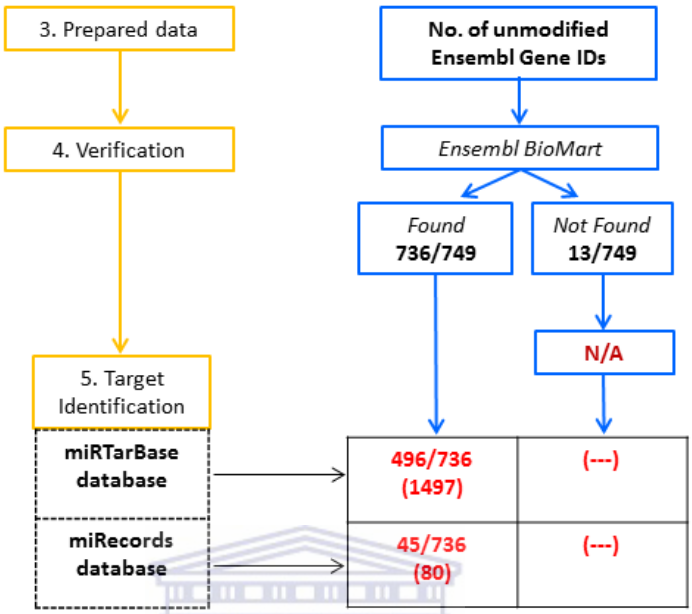

Figure 23: The orange boxes on the left describe the methods used to proceed from verification of data to target identification. The blue boxes on the right describe the results of the verification of Ensembl Gene IDs in Ensembl BioMart, followed by the target identification in miRTarBase and miRecords. The number of genes found in each database is reported, along with the count of the interactions reported in parentheses. Ensembl Gene IDs which did not pass the verification in BioMart were not processed further as no conversion could be made to HGNC symbols for target identification. "N/A" indicates that the IDs not found in BioMart were not processed further, therefore not reporting any target interactions, as illustrated by "---".

11.5.3.1. Python Script 11: 'PS11_16_July_DE_Genes_thru_DBs.py'

11.5.3.1.1. miRTarBase

From the input list of 749 Ensembl Gene IDs, 736 IDs were matched, as shown in Figure 23, and written with the corresponding associated gene name for further analysis. Of the 736 genes found, 496 genes were found in the miRTarBase database and were written to a text file named 'found_DE_genes_ALL_targets_in_miRTarBase.txt'. These 496 genes collectively matched to 1497 non-unique interacting miRNAs. Thus, 1497 miRNA-gene interactions for these 496 genes were detected in the miRTarBase database and can be viewed in ADDENDUM E. The remaining 240 differentially-expressed SLE-

associated genes were not found as records in the miRTarBase database and were written to the text file, 'NOT_found_DE_genes_targets_in_miRTarBase.txt'.

11.5.3.1.2. miRecords

Of the 736 genes validated in Ensembl BioMart search, 45 genes were identified in the miRecords database as having interacting miRNAs. Collectively, these 45 genes had 80 target interactions listed in this database. These 80 MTIs are shown in Table 19. A total of 691 of the 736 differentially-expressed SLE-associated genes were not found as gene records in the miRecords database and were written to the error file 'NOT_found_DE_genes_targets_in_miRecords.txt'.

Table 19: The 45 out of 736 differentially-expressed SLE genes which were found to have interacting miRNAs in the miRecords database

Gene	Number of interacting miRNAs	Interacting miRNAs
ABCA1	3	hsa-miR-33a,hsa-miR-33b, hsa-miR-758
ADAR	1	hsa-miR-1
AP1M2	1	hsa-miR-124
ATP11A	1	hsa-miR-122
CCND1	13	hsa-miR-302a, hsa-miR-195, hsa-miR-424, hsa-miR-16, hsa-miR-20a, hsa-miR-17, hsa-miR-34a, hsa-miR-296-5p, hsa-miR-19a, hsa-miR-15a, hsa-miR-503, hsa-miR-155, hsa-miR-302c
CCND3	2	hsa-miR-424, hsa-miR-16
CDH1	1	hsa-miR-9
CDKN1A	5	hsa-miR-21, hsa-miR-17, hsa-miR-106a, hsa-miR-93, hsa-miR-106b
COL4A1	1	hsa-miR-29c
CTGF	4	hsa-miR-124, hsa-miR-18a hsa-miR-19a, hsa-miR-19b
ERBB4	3	hsa-miR-372, hsa-miR-302d, hsa-miR-19a
FA2H	1	hsa-miR-124
FAM3C	1	hsa-miR-21
FGF20	1	hsa-miR-433
FSCN1	2	hsa-miR-145, hsa-miR-133a
H3F3B	4	hsa-miR-1, hsa-miR-16, hsa-miR-15a, hsa-miR-125b
HERC6	2	hsa-miR-15a, hsa-miR-16
HNRNPK	1	hsa-miR-21
HOMER2	1	hsa-miR-125b
HOXA7	1	hsa-miR-196a

IGFBP5	1	hsa-miR-140-5p
IRAK1	1	hsa-miR-146a
KRT7	3	hsa-miR-30a, hsa-miR-133a, hsa-miR-145
LRRC1	1	hsa-miR-124
MAFB	1	hsa-miR-130a
MATR3	1	hsa-miR-155
MBNL2	2	hsa-miR-302d, hsa-miR-372
MCPH1	1	hsa-miR-146a
MMP1	1	hsa-miR-222
MMP16	1	hsa-miR-146b-5p
NETO2	1	hsa-miR-1
OSMR	1	hsa-miR-122
PARP16	1	hsa-miR-124
PPP2R2A	1	hsa-miR-222
PRKD1	2	hsa-miR-124, hsa-miR-17
RBL1	1	hsa-miR-17
RELN	1	hsa-miR-128
SLC16A1	2	hsa-miR-376a*, hsa-miR-124
SNCA	1	hsa-miR-7
STOM	1	hsa-miR-124
STX11	1	hsa-miR-373
SULF1	1	hsa-miR-516a-5p
UBE2Q1	1	hsa-miR-338-3p
UGP2	2	hsa-miR-15a, hsa-miR-16
YWHAZ	1	hsa-miR-375

11.6. Pooling of Interactions

11.6.1. Pooling of miRNAs interacting with input gene lists

With pooling, the list of all interactions contained 1769 gene-miRNA interactions for both databases. From this, 1738 unique gene-miRNA interactions were identified from the pooled list of all interacting miRNAs. From the 1738 unique gene-miRNA interactions, 291 unique miRNAs were found to interact with input genes being evaluated.

11.6.2. Pooling of genes interacting with input miRNA list

With pooling, the list of all interactions contained 3515 miRNA-gene interactions for both databases. To determine how many unique genes were interacting with the input list of miRNAs, the gene symbols were extracted from the interactions,

and appended to a list named genes_only. From the 3401 unique gene-miRNA interactions, 2702 unique genes were found to interact with miRNAs being evaluated.

11.7. Cross-Analysis of Interactions

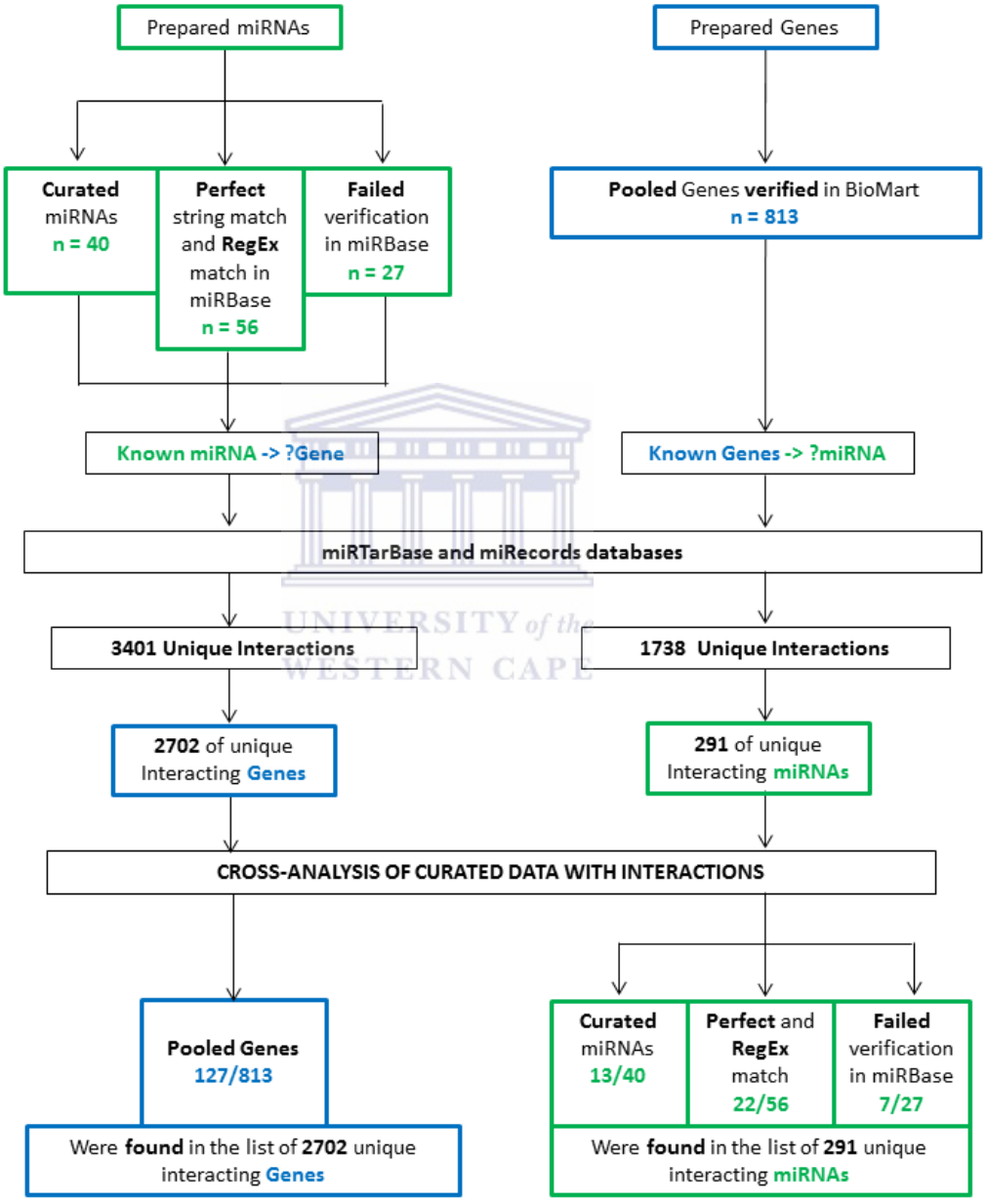

Figure 24: This figure describes the results of the processing the curated SLE-associated miRNAs and genes, starting with data curation, proceeding to target identification, and lastly identifying which gene and miRNAs which were found to interact with the curated miRNAs, and genes, respectively.

11.7.1. Cross-Analysis of Lupus-associated miRNAs

11.7.1.1. Test 1: starting list of n= 40 VS interacting miRNAs n = 291

Python Script 12: 'PS12_Test_miRNA_1.py'

From the starting list of 40 miRNAs shown in the green box in the top far left of Figure 24, 13 miRNAs were found in the list of interacting miRNAs, and were written to the output file 'starting_mirnas_IN_interacting.txt'. Shown in the bottom far left of Figure 24, these 13 miRNAs were: hsa-miR-141, hsa-miR-142-3p, hsa-miR-145, hsa-miR-146a, hsa-miR-155, hsa-miR-17-5p, hsa-miR-196a, hsa-miR-21, hsa-miR-224, hsa-miR-298, hsa-miR-409-3p, hsa-miR-423-5p, hsa-miR-516a-5p. The miRNA-gene interactions were written to the text file 'starting_ mirnas_found_interactions.txt'.

Table 20: The 13/40 starting miRNAs - not yet verified in miRBase - which were found in the list of interacting miRNAs as output from the analysis of the genes' list of interactions

13 miRNAs from starting list of n = 40	No. of Interactions	Gene interactions
hsa-miR-141	1	*PTPRD*
hsa-miR-142-3p	2	*C2*, *C4BPB*
hsa-miR-145	2	*FSCN1*, *KRT7*
hsa-miR-146a	4	*IRAK1*, *MCPH1*, *STAT1*, *TRAF6*
hsa-miR-155	3	*CCND1*, *ETS1*, *MATR3*
hsa-miR-17-5p	19	*ARIH1*, *ATXN7*, *CCL1*, *CCND1*, *CDKN1A*, *CTSA*, *DCBLD2*, *GNAS*, *IGFBP5*, *IRAK1*, *KAT2A*, *PTTG1*, *RBL1*, *RPL37*, *RPS27A*, *RPSA*, *SMAD4*, *SYNDIG1*, *TRA2B*
hsa-miR-196a	1	*HOXA7*
hsa-miR-21	3	*CDKN1A*, *FAM3C*, *HNRNPK*
hsa-miR-224	1	*API5*
hsa-miR-298	1	*CDKN1A*
hsa-miR-409-3p	1	*FGG*
hsa-miR-423-5p	11	*CCNI*, *DRAP1*, *HIST2H2BE*, *HOXA7*, *MYBL2*, *NFKB2*, *PTTG1*, *RPL18A*, *SLC7A2*, *UBE2L3*, *YWHAZ*
hsa-miR-516a-5p	1	*SULF1*

For these 13 starting miRNAs, 50 interacting genes were found for the pooled list of associated- and differentially-expressed genes as shown in Table 20. From these 13 starting miRNAs, unverified in miRBase, hsa-miR-17-5p and hsa-miR-423-5p had the most gene interactions in the two databases, interacting with 19 and 11 genes, respectively. The miRNAs from the starting list which were not found in the list of interacting miRNAs was written to the output file 'starting_mirnas_NOT_ IN_interacting.txt'. These 27 miRNAs were: hsa-miR-112, hsa-miR-1224-3P, hsa-miR-1224-3p, hsa-miR-125a, hsa-miR-126, hsa-miR-142, hsa-miR-148a, hsa-miR-150, hsa-miR-17-92, hsa-miR-181, hsa-miR-182-96-183, hsa-miR-184, hsa-miR-189, hsa-miR-198, hsa-miR-23b, hsa-miR-299-3p, hsa-miR-3148, hsa-miR-342, hsa-miR-371-5p, hsa-miR-383, hsa-miR-483-5p, hsa-miR-513-5p, hsa-miR-61, hsa-miR-629, hsa-miR-638, hsa-miR-663, hsa-miR-78.

11.7.1.2. Test 2: Verified (==) and (=regex) VS interacting miRNAs

Python Script 13: 'PS13_Test_miRNA_2.py'

In this script, the miRNAs verified as a perfect string match in miRBase, and the miRNAs verified as a regular expression in miRBase were pooled. A total of 57 miRNAs - 13 found as an exact string match, and 44 regular expression matches, see step 1 and step 2 of miRNA verification in BioMart – were cross-analysed for being present in the list of 291 unique interacting miRNAs. By using the Python set() function, a unique set of miRNAs was created to remove any duplicate miRNA names. The list of 57 miRNAs, verified in BioMart, was reduced to 56 miRNAs. From this list of 56 unique miRNAs, 22 miRNAs were found in the list of interacting miRNAs, and were written to the output file 'verified_mirnas_IN_ interacting.txt', as shown in the bottom right of Figure 25. These 22 miRNAs were: hsa-miR-125a-5p, hsa-miR-126-3p, hsa-miR-141-3p, hsa-miR-142-3p, hsa-miR-145-5p, hsa-miR-146a-5p, hsa-miR-148a-3p, hsa-miR-155-5p, hsa-miR-17-5p, hsa-miR-181a-5p, hsa-miR-181b-5p, hsa-miR-181c-5p, hsa-miR-196a-5p, hsa-miR-21-5p, hsa-miR-224-5p, hsa-miR-23b-3p, hsa-miR-298, hsa-miR-342-3p, hsa-miR-409-3p, hsa-miR-423-5p, hsa-miR-516a-5p, hsa-miR-629-5p. These

interactions were written to the output file 'verified_mirnas_found_ interactions.txt', and the interactions are shown in Table 20.

Table 21: The 22/56 starting miRNAs, which were verified in miRBase, and were found in the list of interacting miRNAs as output from the analysis of the genes lists.

22 miRNAs from verified list (n = 56)	No. of Interactions	Gene interactions
hsa-miR-125a-5p	8	*ATP5G2, CDKN1A, COL4A1, FERMT1, LAMB1, MACF1, OTUB2, TNFAIP3*
hsa-miR-516a-5p	1	*SULF1*
hsa-miR-141-3p	2	*PTPRD, ZFPM2*
hsa-miR-148a-3p	4	*AMELX, CCNI, CYCS, MPP5*
hsa-miR-196a-5p	9	*CCND1, EPHA7, HOXA7, NAP1L4, RANBP9, TAB2, TRA2B, TSPAN12, ZBTB24*
hsa-miR-142-3p	4	*C2, C4BPB, C2, C4BPB*
hsa-miR-342-3p	2	*ATXN7, FAM3C*
hsa-miR-409-3p	1	*FGG*
hsa-miR-21-5p	20	*ARID4A, CCR1, DDX46, EIF4A2, ELOVL4, FAM3C, FAM46A, HNRNPK, MPP5, MYEF2, NETO2, PHACTR2, PLAT, RASGRP3, RUFY3, SYNE2, TCF21, TNFAIP3, ZNF587, ZNF667*
hsa-miR-298	1	*CDKN1A*
hsa-miR-145-5p	8	*CDKN1A, CTGF, FAM3C, FSCN1, FZD7, KRT7, MMP1, STAT1*
hsa-miR-629-5p	2	*ATP5G2, HIST1H2AC*
hsa-miR-17-5p	19	*ARIH1, ATXN7, CCL1, CCND1, CDKN1A, CTSA, DCBLD2, GNAS, IGFBP5, IRAK1, KAT2A, PTTG1, RBL1, RPL37, RPS27A, RPSA, SMAD4, SYNDIG1, TRA2B*
hsa-miR-146a-5p	15	*CDKN1A, ERBB4, IFI27, IFI44, IFI44L, IFIT1, IFIT3, IRAK1, ISG15, OASL, OLFML2A, SMAD4, SPP1, STAT1, TRAF6*
hsa-miR-181a-5p	10	*GCNT1, H3F3B, HEY2, KAT2B, PRLR, PTPN22, SIK2, TAB2, TCF21, TNIP1*
hsa-miR-181b-5p	2	*KAT2B, RPL18A*
hsa-miR-23b-3p	7	*ACTN4, CCT7, GHITM, HNRNPK, RPS27A, SLC16A1, TAB2*
hsa-miR-181c-5p	1	*NOTCH4*
hsa-miR-155-5p	35	*CCND1, CDH6, CHAF1B, CLDN1, CTSA, DRAP1, ETS1, FAM135A, FAM3C, GHITM, GNAS, HAX1, HSPA4L, IMPAD1, MAFB, MATR3, MECP2, MPP5, NCAPD2, PDE3A, PHF14, PPFIBP1, PPP2R2A, RBM22, RIF1, SLC35F2, SMAD4, SMARCE1, STIM1, SYNE2, TAB2, UGT8, VAV2, YWHAZ, ZNF273*
hsa-miR-423-5p	11	*CCNI, DRAP1, HIST2H2BE, HOXA7, MYBL2, NFKB2, PTTG1, RPL18A, SLC7A2, UBE2L3, YWHAZ*
hsa-miR-224-5p	2	*API5, SMAD4*
hsa-miR-126-3p	1	*PGR*

For these 22 miRNAs verified for their existence as a record in miRBase, 165 interactions were found for the pooled list of associated- and differentially-expressed genes, as shown in Table 21. From these 22 miRNAs, hsa-miR-155-5p had the most interactions, regulating 35 genes. This miRNA was followed by hsa-miR-21-5p and hsa-miR-7-5p, which regulated 20 and 19 genes, respectively.

The 34 miRNAs which were verified in miRBase, but were not found to interact with the SLE-associated genes were written to the text file, verified_mirnas_NOT_IN_interacting.txt. The miRNA IDs written to this file were: hsa-miR-1224-3p, hsa-miR-125a-3p, hsa-miR-126-5p, hsa-miR-141-5p, hsa-miR-142-5p, hsa-miR-145-3p, hsa-miR-146a-3p, hsa-miR-148a-5p, hsa-miR-150-3p, hsa-miR-150-5p, hsa-miR-155-3p, hsa-miR-181a-2-3p, hsa-miR-181a-3p, hsa-miR-181b-2-3p, hsa-miR-181b-3p, hsa-miR-181c-3p, hsa-miR-181d-3p, hsa-miR-181d-5p, hsa-miR-184, hsa-miR-196a-3p, hsa-miR-198, hsa-miR-21-3p, hsa-miR-224-3p, hsa-miR-23b-5p, hsa-miR-299-3p, hsa-miR-3148, hsa-miR-342-5p, hsa-miR-383-3p, hsa-miR-383-5p, hsa-miR-483-5p, hsa-miR-629-3p, hsa-miR-638, hsa-miR-663a, hsa-miR-663b.

11.7.1.3. Test 3: Not verified (n=27) VS interacting miRNAs

Python Script 14: 'PS14_Test_miRNA_3.py'

The above script analysed 27 miRNAs ('no_match_in_miRBase.txt') which failed the verification step as an exact string match in the miRBase database. From this list of 27 unique miRNAs, seven miRNAs were found in the list of interacting miRNAs, as shown in the far bottom right of Figure 24, and were written to the output file 'not_verified_mirnas_IN_interacting.txt'. These seven miRNAs were: hsa-miR-141, hsa-miR-145, hsa-miR-146a, hsa-miR-155, hsa-miR-196a, hsa-miR-21, hsa-miR-224.

Table 22: The 7/27 miRNAs - NOT verified in miRBase - which were found in the list of interacting miRNAs as output from the analysis of the genes lists.

7 miRNAs from unverified list of n = 27	No. of Interactions	Gene interactions
hsa-miR-141	1	*PTPRD*
hsa-miR-145	2	*FSCN1*, *KRT7*
hsa-miR-146a	4	*IRAK1*, *MCPH1*, *STAT1*, *TRAF6*
hsa-miR-155	3	*CCND1*, *ETS1*, *MATR3*
hsa-miR-196a	1	*HOXA7*
hsa-miR-21	3	*CDKN1A*, *FAM3C*, *HNRNPK*
hsa-miR-224	1	*API5*

The gene interactions for the seven miRNAs which failed verification in BioMart were written to the output file 'not_verified_mirnas_found_interactions.txt' and are reported in Table 22. For these seven miRNAs not found as a record in miRBase, 15 interactions were found for the pooled list of SLE-associated- and differentially-expressed genes. From these seven miRNAs, hsa-miR-146a had the most interactions, regulating four genes.

The remaining 20 miRNAs which were not found in the list of interacting miRNAs was written to the output file, 'not_verified_mirnas_NOT_IN_interacting.txt'. These 20 miRNAs were: hsa-miR-112, hsa-miR-1224-3P, hsa-miR-125a, hsa-miR-126, hsa-miR-142, hsa-miR-148a, hsa-miR-150, hsa-miR-17-92, hsa-miR-181, hsa-miR-182-96-183, hsa-miR-189, hsa-miR-23b, hsa-miR-342, hsa-miR-371-5p, hsa-miR-383, hsa-miR-513-5p, hsa-miR-61, hsa-miR-629, hsa-miR-663, hsa-miR-78.

11.7.2. Cross-Analysis of Pooled Lupus-associated and Differentially-Expressed Genes:

11.7.2.1. Test 1: Verified in BioMart, n= 813 VS interacting genes

Python Script 15: 'PS15_Test_Gene_1.py'

The list of SLE-associated genes (n = 82), and the differentially-expressed SLE genes (n = 736), which passed the verification step in BioMart were pooled into a single list of test genes (n = 818). Using the set() function in Python, a unique list of test genes was created. The list of 818 starting genes, verified in BioMart, was reduced to a list of 813 unique genes.

From the list of 813 unique genes verified in BioMart, 127 genes were found in the list of 2702 unique interacting genes, and were written to the output file 'verified_Genes_IN_interacting.txt'. For these 127 genes, 183 miRNA interactions were found cumulatively. Their interactions identified with the SLE-associated- and differentially-expressed genes being assessed, were written to the output file 'verified_Genes_found_interactions.txt' and were reported in Table 23.

The 686 genes from the unique starting list of 813 genes which were not found in the list of interacting genes were written to the output file 'verified_Genes_NOT_IN_interacting.txt'.

Table 23: The 127/813 lupus-associated genes - verified in BioMart - which were found in the result of interacting genes as output from the analysis of the lupus-associated miRNA list.

Gene name	No. of Interactions	miRNA interactions	Gene name	No. of Interactions	miRNA interactions
ACTN4	1	hsa-miR-23b-3p	*MCPH1*	1	hsa-miR-146a,
AMELX	1	hsa-miR-148a-3p	*MECP2*	1	hsa-miR-155-5p
API5	2	hsa-miR-224, hsa-miR-224-5p	*MMP1*	1	hsa-miR-145-5p
ARID4A	1	hsa-miR-21-5p,	*MPP5*	3	hsa-miR-148a-3p, hsa-miR-155-5p, hsa-miR-21-5p
ARIH1	1	hsa-miR-17-5p	*MYBL2*	1	hsa-miR-423-5p
ATP5G2	2	hsa-miR-125a-5p, hsa-miR-629-5p	*MYEF2*	1	hsa-miR-21-5p

ATXN7	2	hsa-miR-17-5p, hsa-miR-342-3p	*NAP1L4*	1	hsa-miR-196a-5p
C2	1	hsa-miR-142-3p	*NCAPD2*	1	hsa-miR-155-5p
C4BPB	1	hsa-miR-142-3p	*NETO2*	1	hsa-miR-21-5p
CCL1	1	hsa-miR-17-5p	*NFKB2*	1	hsa-miR-423-5p
CCND1	4	hsa-miR-155, hsa-miR-155-5p, hsa-miR-17-5p, hsa-miR-196a-5p	*NOTCH4*	1	hsa-miR-181c-5p
CCNI	2	hsa-miR-148a-3p, hsa-miR-423-5p	*OASL*	1	hsa-miR-146a-5p
CCR1	1	hsa-miR-21-5p	*OLFML2A*	1	hsa-miR-146a-5p
CCT7	1	hsa-miR-23b-3p	*OTUB2*	1	hsa-miR-125a-5p
CDH6	1	hsa-miR-23b-3p	*PDE3A*	1	hsa-miR-155-5p
CDKN1A	6	hsa-miR-125a-5p, hsa-miR-145-5p, hsa-miR-146a-5p, hsa-miR-17-5p, hsa-miR-21, hsa-miR-298	*PGR*	1	hsa-miR-126-3p
CHAF1B	1	hsa-miR-155-5p	*PHACTR2*	1	hsa-miR-21-5p
CLDN1	1	hsa-miR-155-5p	*PHF14*	1	hsa-miR-155-5p
COL4A1	1	hsa-miR-125a-5p	*PLAT*	1	hsa-miR-21-5p
CTGF	1	hsa-miR-145-5p	*PPFIBP1*	1	hsa-miR-155-5p
CTSA	2	hsa-miR-155-5p, hsa-miR-17-5p	*PPP2R2A*	1	hsa-miR-155-5p
CYCS	1	hsa-miR-148a-3p	*PRLR*	1	hsa-miR-181a-5p
DCBLD2	1	hsa-miR-17-5p	*PTPN22*	1	hsa-miR-181a-5p
DDX46	1	hsa-miR-21-5p	*PTPRD*	2	hsa-miR-141, hsa-miR-141-3p
DRAP1	2	hsa-miR-155-5p, hsa-miR-423-5p	*PTTG1*	2	hsa-miR-17-5p, hsa-miR-423-5p
EIF4A2	1	hsa-miR-21-5p	*RANBP9*	1	hsa-miR-196a-5p
ELOVL4	1	hsa-miR-21-5p	*RASGRP3*	1	hsa-miR-21-5p
EPHA7	1	hsa-miR-196a-5p	*RBL1*	1	hsa-miR-17-5p

ERBB4	1	hsa-miR-146a-5p	*RBM22*	1	hsa-miR-155-5p
ETS1	2	hsa-miR-155, hsa-miR-155-5p	*RIF1*	1	hsa-miR-155-5p
FAM135A	1	hsa-miR-155-5p	*RPL18A*	2	hsa-miR-181b-5p, hsa-miR-423-5p
FAM3C	5	hsa-miR-145-5p, hsa-miR-155-5p, hsa-miR-21, hsa-miR-21-5p, hsa-miR-342-3p	*RPL37*	1	hsa-miR-17-5p
FAM46A	1	hsa-miR-21-5p	*RPS27A*	2	hsa-miR-17-5p, hsa-miR-23b-3p
FERMT1	1	hsa-miR-125a-5p	*RPSA*	1	hsa-miR-17-5p
FGG	1	hsa-miR-409-3p	*RUFY3*	1	hsa-miR-21-5p
FSCN1	2	hsa-miR-145, hsa-miR-145-5p	*SIK2*	1	hsa-miR-181a-5p
FZD7	1	hsa-miR-145-5p	*SLC16A1*	1	hsa-miR-23b-3p
GCNT1	1	hsa-miR-181a-5p	*SLC35F2*	1	hsa-miR-155-5p
GHITM	2	hsa-miR-155-5p, hsa-miR-23b-3p	*SLC7A2*	1	hsa-miR-423-5p
GNAS	2	hsa-miR-155-5p, hsa-miR-17-5p	*SMAD4*	4	hsa-miR-146a-5p, hsa-miR-155-5p, hsa-miR-17-5p, hsa-miR-224-5p
H3F3B	1	hsa-miR-181a-5p	*SMARCE1*	1	hsa-miR-155-5p
HAX1	1	hsa-miR-155-5p	*SPP1*	1	hsa-miR-146a-5p
HEY2	1	hsa-miR-181a-5p	*STAT1*	3	hsa-miR-145-5p, hsa-miR-146a, hsa-miR-146a-5p,
HIST1H2AC	1	hsa-miR-629-5p	*STIM1*	1	hsa-miR-155-5p
HIST2H2BE	1	hsa-miR-423-5p	*SULF1*	1	hsa-miR-516a-5p
HNRNPK	3	hsa-miR-21, hsa-miR-21-5p, hsa-miR-23b-3p	*SYNDIG1*	1	hsa-miR-17-5p,
HOXA7	3	hsa-miR-196a, hsa-miR-196a-5p, hsa-miR-423-5p	*SYNE2*	2	hsa-miR-155-5p, hsa-miR-21-5p
HSPA4L	1	hsa-miR-155-5p	*TAB2*	4	hsa-miR-155-5p, hsa-miR-181a-5p, hsa-miR-196a-5p, hsa-miR-23b-3p

IFI27	1	hsa-miR-146a-5p	*TCF21*	2	hsa-miR-181a-5p, hsa-miR-21-5p
IFI44	1	hsa-miR-146a-5p	*TNFAIP3*	4	hsa-miR-125a-5p, hsa-miR-21-5p, hsa-miR-125a-5p, hsa-miR-21-5p
IFI44L	1	hsa-miR-146a-5p	*TNIP1*	1	hsa-miR-181a-5p
IFIT1	1	hsa-miR-146a-5p	*TRA2B*	2	hsa-miR-17-5p, hsa-miR-196a-5p
IFIT3	1	hsa-miR-146a-5p	*TRAF6*	2	hsa-miR-146a, hsa-miR-146a-5p
IGFBP5	1	hsa-miR-17-5p	*TSPAN12*	1	hsa-miR-196a-5p
IMPAD1	1	hsa-miR-155-5p	*UBE2L3*	1	hsa-miR-423-5p
IRAK1	6	hsa-miR-146a, hsa-miR-146a-5p, hsa-miR-17-5p, hsa-miR-146a, hsa-miR-146a-5p, hsa-miR-17-5p	*UGT8*	1	hsa-miR-155-5p
ISG15	1	hsa-miR-146a-5p	*VAV2*	1	hsa-miR-155-5p
KAT2A	1	hsa-miR-17-5p	*YWHAZ*	2	hsa-miR-155-5p, hsa-miR-423-5p,
KAT2B	2	hsa-miR-181a-5p, hsa-miR-181b-5p	*ZBTB24*	1	hsa-miR-196a-5p
KRT7	2	hsa-miR-145, hsa-miR-145-5p	*ZFPM2*	1	hsa-miR-141-3p
LAMB1	1	hsa-miR-125a-5p	*ZNF273*	1	hsa-miR-155-5p
MACF1	1	hsa-miR-125a-5p	*ZNF587*	1	hsa-miR-21-5p
MAFB	1	hsa-miR-155-5p	*ZNF667*	1	hsa-miR-21-5p
MATR3	2	hsa-miR-155, hsa-miR-155-5p			

11.7.2.2. Test 2: NOT Verified in BioMart, n = 3 VS interacting genes

Python Script 16: 'PS16_Test_Gene_2.py'

Of the 749 differentially-expressed lupus genes, in Ensembl ID format, 13 gene IDs were reported as deprecated IDs within the database and thus were not processed further. Of the 85 lupus-associated genes curated from literature, three genes, namely *CREBL1*, *IRF53*, and *Lyn*, were not found in the BioMart database. None of these three genes were found in the list of interacting genes,

and an empty file was created and named 'NOT_verified_Genes_IN_interacting.txt', demonstrating no interactions being obtained. All three genes curated from literature and not verified in BioMart, were written to the file 'NOT_verified_Genes_NOT_IN_interacting.txt'.

11.8. Identification of Top Regulating miRNAs

From the cross-analysis of the curated SLE-associated miRNAs, 42/123 non-unique miRNAs was found to interact with the curated SLE-associated genes. Of the 42 non-unique starting miRNAs, six miRNAs were identified to interact with ten or more SLE-associated genes, as shown in Table 24.

Table 24: The six miRNAs which were identified to interact with 10 or more SLE-associated genes

	miRNA name:	No. of miRNA -> gene interactions:
	hsa-miR-17-5p	19
	hsa-miR-423-5p	11
	hsa-miR-21-5p	20
	hsa-miR-146a-5p	15
	hsa-miR-181a-5p	10
	hsa-miR-155-5p	35
Total :	**6**	**110**

Thus, the six top regulating SLE-associated miRNAs were pooled with their corresponding interactions and written to a file named 'all_top_regulating_ miRNAs.txt'. This file contained 116 raw data entries to be used in pathway analysis.

11.9. Identification of Top Regulated Genes

From the cross-analysis of the pooled list of curated SLE-associated- and differentially-expressed genes, 127/813 genes were found to interact with the curated SLE-associated miRNAs. Of the 127 genes, seven genes were identified to interact with four or more SLE-associated miRNAs as shown in Table 25.

Table 25: The seven genes which were identified to interact with 4 or more SLE-associated miRNAs

	Gene name:	No. of gene -> miRNA interactions:
	CCND1	4
	CDKN1A	6
	FAM3C	5
	SMAD4	4
	TAB2	4
	TNFAIP3	4
	IRAK1	6
Total:	**7**	**33**

Thus, the seven top regulating SLE-associated genes were pooled with their corresponding interactions and written to a file named 'all_top_regulating_ genes.txt', to be used in biological pathway analysis.

11.10. Pooling of all Top Regulators

The top regulated miRNAs and top regulating genes were pooled into a single data file named 'all_top_regulated.txt'. The six top regulating miRNAs along with the 110 associated gene targets, and the seven top regulated genes along with the 33 interacting miRNAs were pooled to create a file of all the top regulated elements from this study. Thus, the file contained a list of 156 raw data entries which was used in pathway analysis.

11.11. Pathway Analysis

11.11.1. Preparation of dataset files

Two python scripts, Python Script 17: 'PS17_prep_genes_for_IPA.py' and Python Script 18: 'PS18_prep_miRNAs_for_IPA.py', were written to prepare the individual dataset files into lists of unique identifiers.

11.11.1.1. Starting Genes and miRNAs

Using Python Script 17: 'PS17_prep_genes_for_IPA.py', the starting list of genes and miRNAs was created by pooling the known SLE-associated genes from literature (n = 82), the known list of differentially-expressed genes, verified in BioMart (n = 736), and the known miRNAs curated from literature (n = 57). Using the set() function in this Python script, the list of 875 items was reduced to a list

of 869 unique identifiers. These 869 identifiers were then uploaded into IPA as a dataset. Of the total, 866 identifiers were mapped, and three identifiers were unmapped in the IPA database. The three unmapped IDs were: '*HLA-DRB2*', '*RP1-152L7.5*', and '*RP4-802A10.1*'.

11.11.1.2. Interacting Genes and miRNAs

The list of interacting genes and miRNAs was produced as a result of pooling the interacting targets, as obtained by the analysis of the miRTarBase, and miRecords databases. For the curated miRNAs which were verified in miRBase, 2702 unique gene interactions were identified. For the starting genes verified in Ensembl BioMart, 291 unique interacting miRNAs were identified collectively in miRTarBase and miRecords. Thus, the raw data file of interacting miRNAs and genes contained 2993 identifiers. Of the 2993 identifiers, 2969 IDs were mapped, and 24 IDs were unmapped in the IPA database.

11.11.1.3. All Genes and miRNAs (starting, and interacting)

A complete list of all verified, starting miRNAs and genes, as well as the interacting miRNAs and genes were pooled into a single list. As shown in 11.6.1 and 11.6.2 respectively, 291 unique interacting miRNAs and 2702 unique interacting genes were pooled along with the starting, and verified differentially-expressed genes, and genes curated from literature. This list contained a total of 3868 entries.

In a Python script, the set() function was employed to create a unique list of identifiers, and remove any duplicate entries. A unique list of 3713 entries was created and these were then analysed by IPA in preparation for analysis. Of the 3713 entries, 3687 were mapped to identifiers in the IPA database, and 26 identifiers remained unmapped.

11.11.1.4. All Top Regulating miRNAs and associated genes

As described in Table 24, six top regulating miRNAs and the 110 interacting genes were included for this dataset. Thus, this file contained 116 non-unique identifiers. All 116 identifiers were mapped in IPA, and duplicate entries were

flagged. Post preparation, this file contained 101 analysis-ready molecules, as determined by IPA.

11.11.1.5. All Top Regulated Genes and associated miRNAs

As shown in Table 25, seven top regulated genes were identified from the cross-analysis, along with 33 interacting miRNAs. Therefore, this file contained 40 identifiers. Duplicate entries were flagged and as a result, this dataset contained 22 analysis-ready molecules.

11.11.1.6. All Top Regulated Genes and miRNAs and associations

This dataset comprised of the pooled top regulators and their associations, as described in 11.10. The raw data list contained 156 identifiers, and after processing by IPA, 108 unique analysis-ready molecules were reported.

11.11.2. Uploading of datasets into IPA

All prepared data sets were uploaded into IPA. A dataset summary was also provided, showing how many of the raw data entries were mapped, and unmapped in the IPA database, as shown in Table 26.

Table 26: The dataset summary provided for the skeleton dataset submitted to IPA for Core Analyses

Dataset	Raw data	No. of entries **mapped** in IPA	No. of entries **unmapped** in IPA
All Top Regulated Genes and miRNAs and associations	156	156	0
All Top Regulating miRNAs and associated genes	116	116	0
All Top Regulated Genes and associated miRNAs	40	40	0
All Genes and miRNAs (starting, and interacting)	3713	3687	26
Starting Genes and miRNAs	869	866	3
Interacting Genes and miRNAs	2993	2969	24

For the ‘All Genes and miRNAs (starting, and interacting)’ dataset of 3713 entries, 3687 data entries were mapped to identifiers in the IPA database. However, I noticed that not all the data entries were exact string matches to the curated data identifiers. Thus, I wrote two Python scripts named ‘PS19_ALL_mapped_IDs_prepped.py’, and ‘PS20_Test_IPA_mapping’.

The ‘PS19_ALL_mapped_IDs_prepped.py’ script analysed the dataset summary which I exported from IPA. I copied the contents of the ‘mapped’ IDs tab to a standard text file. The first column on the file was tested to identify which entries started with ‘hsa-‘. These entries were then stripped of the ‘hsa-‘ part, to standardise it against the second column, which was the mapped IDs identified by IPA, and already specified to be human. The modified first column was then manually copied, and pasted in an empty Excel spreadsheet.

The second column, containing the IPA mappings, was then searched, to identify any IDs which were separated from additional information by an empty space, i.e. (‘ ‘). These mapped IDs were stripped of the empty space, and thus the first column retained. This modified second column was then manually copied, and pasted into the Excel document, in the column directly adjacent to the column which had been stripped of the ‘hsa-‘ string. This Excel sheet was then saved as a text file named ‘mapped_IDs.csv’.

Thereafter, the ‘PS20_Test_IPA_mapping.py’ script was used to open the ‘mapped_IDs.csv’ file, and then attempt to perform a perfect string match between the first column containing the input curated ID, and the corresponding second column containing the ID mapped in the IPA database.

For the 3687 IDs mapped in the IPA database, 3466 input entries were matched as an exact string match, and were written to an output text file named ‘mapped_IDs_exact_match.txt’, while 221 IDs were mapped to IDs in IPA which were not identical to the input ID, and were written to the output text file ‘mapped_IDs_NOT_matched.txt’. These 221 curated entries, which were mapped to IDs not having exactly the same gene name in the IPA database, can

be found in ADDENDUM K. For the same dataset, 26 data entries were not mapped to the IPA database, and thus were excluded from the Core Analysis. They were: *SHIP*, *IRF-5*, *pten*, *RP1-152L7.5*, *cmyb*, *RP4-802A10.1*, *MeCP2*, *HLA-DRB2*, *MKK4*, *MCP2*, *tp63*, *Peli1*, *KGF*, *CTLA-4*, *MyD88*, *IL-8*, *MSK1*, *CDNK1B*, *IGF-IR*, *NFAT1*, *HuR*, *Rps6ka5*, *gag-pol*, *FOG2*, *NT-3*, *hsa-miR-1274a*.

11.11.3. Core Analysis

A core analysis, as shown in Figure 25, was run on the **'All Top Regulated Genes and miRNAs and associations'** dataset file, as this dataset formed the skeleton of the regulatory networks. Using the parameters previously described, 156 entries of raw data were assessed in preparation for Core Analysis. IPA flagged and removed any duplicate entries creating a list of 108 Analysis-ready molecules.

These were: *ARID4A*, *ARIH1*, *ATXN7*, *CCL1*,*CCND1*, *CCNI*, *CCR1*, *CDH6*, *CDKN1A*, *CHAF1B*, *CLDN1*, *CTSA*, *DCBLD2*, *DDX46*, *DRAP1*, *EIF4A2*, *ELOVL4*, *ERBB4*, *ETS1*, *FAM135A*, *FAM3C*, *FAM46A*, *GCNT1*, *GHITM*, *GNAS*, *H3F3B*, *HAX1*, *HEY2*, *HIST2H2BE*, *HNRNPK*, *HOXA7*, *hsa-miR-125a-5p*, *hsa-miR-145-5p*, *hsa-miR-146a-5p*, *hsa-miR-155*, *hsa-miR-17-5p*, *hsa-miR-181a-5p* ,*hsa-miR-196a-5p*, *hsa-miR-21-5p*, *hsa-miR-224-5p*, *hsa-miR-23b-3p*, *hsa-miR-298*, *hsa-miR-342-3p*, *hsa-miR-423-5p*, *HSPA4L*, *IFI27*, *IFI44*, *IFI44L*, *IFIT1*, *IFIT3*, *IGFBP5*, *IMPAD1*, *IRAK1*,*ISG15*, *KAT2A*, *KAT2B*, *MAFB*, *MATR3*, *MECP2*, *MPP5*, *MYBL2*, *MYEF2*, *NCAPD2*, *NETO2*, *NFKB2*, *OASL*, *OLFML2A*, *PDE3A*, *PHACTR2*, *PHF14*, *PLAT*, *PPFIBP1*, *PPP2R2A*, *PRLR*, *PTPN22*, *PTTG1*, *RASGRP3*, *RBL1*, *RBM22*,*RIF1*, *RPL18A*, *RPL37*, *RPS27A*, *RPSA*, *RUFY3*, *SIK2*, *SLC35F2*, *SLC7A2*, *SMAD4*, *SMARCE1*, *SPP1*, *STAT1*, *STIM1*, *SYNDIG1*, *SYNE2*, *TAB2*, *TCF21*, *TNFAIP3*, *TNIP1*, *TRA2B*, *TRAF6*, *UBE2L3*, *UGT8*, *VAV2*, *YWHAZ*, *ZNF273*, *ZNF587*, *ZNF667* . The Core Analysis was then run on this dataset of 108 molecules.

The canonical pathways tab provided a default bar chart view of the biological pathways associated with the dataset on which the Core Analysis was run. Using the parameters described in the Methods section, a bar graph was generated for the Canonical Pathways identified for this dataset. The first quarter of the bar

graph is shown in Figure 26. Overlapping pathways were viewed and are shown in Figure 27. The heat map generated for this dataset is reported in Figure 28.

11.11.3.1. Networks

For the '**All Top Regulated Genes and miRNAs and associations' dataset,** consisting of 108 analysis- ready molecules, the Core Analysis generated five networks. The network details for the skeleton dataset are shown in Table 29.

11.11.3.2. Summary of Core Analysis

Top Canonical Pathways		
Name	p-value	Overlap
Interferon Signaling	5.48E-05	11.1 % 4/36
Toll-like Receptor Signaling	6.59E-05	6.8 % 5/73
Role of PKR in Interferon Induction and Antiviral Response	8.36E-05	10.0 % 4/40
iNOS Signaling	1.11E-04	9.3 % 4/43
IL-1 Signaling	1.88E-04	5.5 % 5/91

Top Upstream Regulators		
Upstream Regulator	p-value of overlap	Predicted Activation
CNOT7	5.58E-07	
RBL2	7.09E-06	
POLR2A	9.89E-06	
HDAC1	2.18E-05	
SNW1	2.74E-05	

Top Diseases and Bio Functions		
Diseases and Disorders		
Name	p-value	#Molecules
Cancer	1.16E-02 - 1.74E-10	98
Organismal Injury and Abnormalities	1.16E-02 - 1.74E-10	100
Endocrine System Disorders	1.16E-02 - 6.86E-10	26
Gastrointestinal Disease	1.07E-02 - 6.86E-10	61
Connective Tissue Disorders	1.16E-02 - 1.14E-09	32

Figure 25: Summary of the Core Analysis performed by IPA. The core analysis reports on multiple parameters. The Top Canonical pathways, top upstream regulators, and top diseases for this dataset were selected for reporting.

11.11.3.3. Canonical Pathways

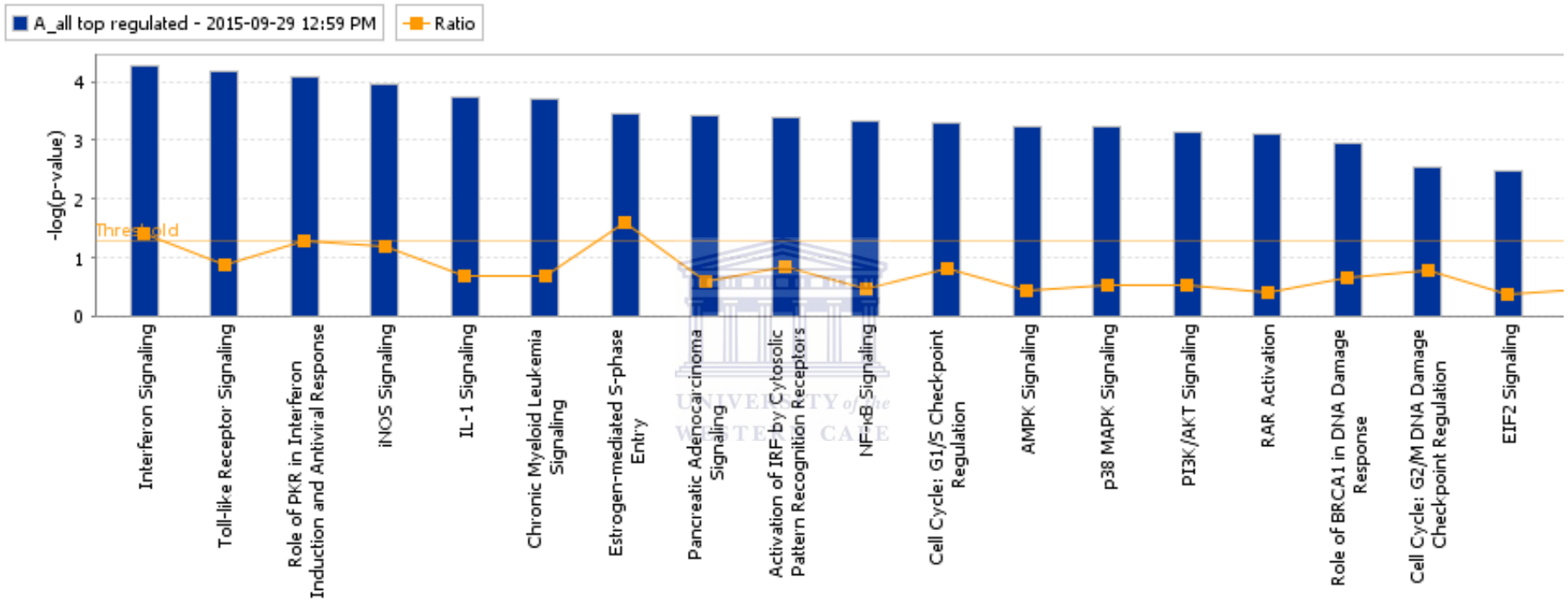

Figure 26: The top 18 canonical pathways impacted by the 'All Top Regulated' dataset. This image shows that many of the genes and miRNAs in the 'All Top Regulated' dataset were mapped to important immune-regulatory pathways. These particularly included Interferon signalling, and Toll-like receptor signalling.

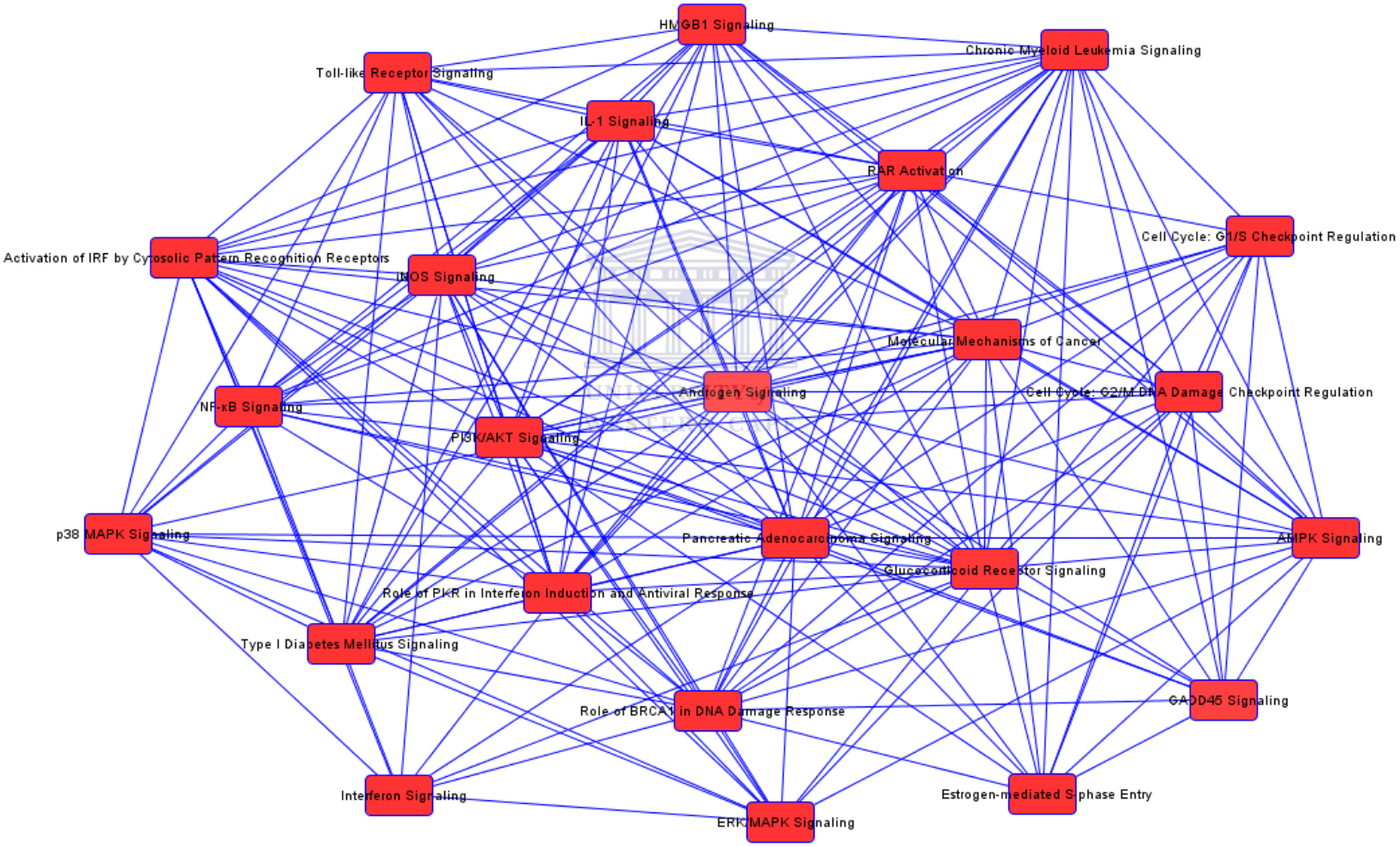

Figure 27: Generated in IPA, this image shows the overlap of 24 biological pathways.

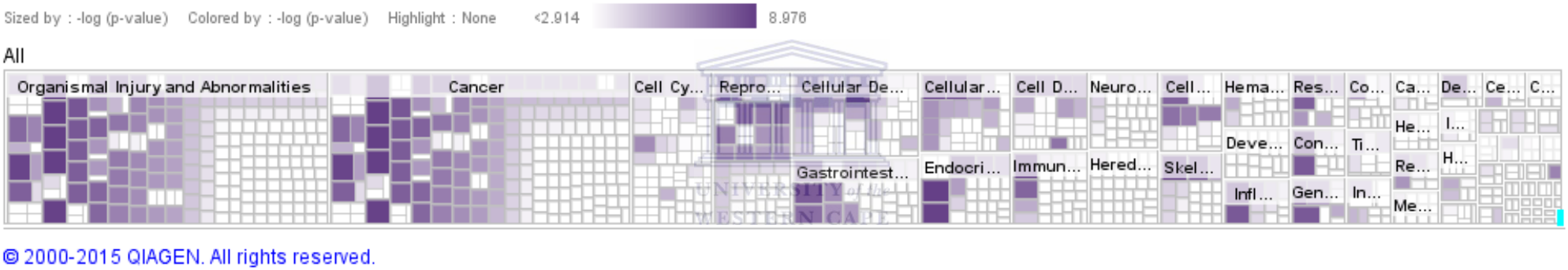

Figure 28: A Heat map of all 25 diseases associated with this dataset. The heat map illustrated that many of the molecules in the skeleton dataset are associated with organismal injury processes, cancer, and cell cycle processes.

11.11.3.4. Upstream Analysis

Table 27: The Top 25 Upstream regulators as determined by the IPA Core Analysis

Upstream Regulator	Molecule Type	p-value of overlap	Target molecules in dataset
Histone h4	group	1.61E-04	*CCND1,CDKN1A,PRLR,RBL1*
N-cor	group	1.20E-03	*CCND1,CDKN1A*
estrogen receptor	group	1.75E-03	*CCND1,ERBB4*
Creb	group	1.94E-03	*CCND1,CDKN1A,GNAS*
Histone h3	group	1.95E-03	*CCND1,CDKN1A,PRLR,RBL1,SYNE2*
FKHR	group	5.26E-03	*CCND1*
TFIIH	complex	5.26E-03	*CDKN1A*
ISGF3	complex	5.26E-03	*ISG15*
CALCOCO1	transcription regulator	5.26E-03	*CCND1*
ZBTB7C	other	5.26E-03	*CDKN1A*
CXXC1	transcription regulator	5.26E-03	*CDKN1A*
VENTX	transcription regulator	5.26E-03	*CCND1*
PADI4	enzyme	1.05E-02	*CDKN1A*
THRAP3	transcription regulator	1.05E-02	*CCND1*
TRIM29	transcription regulator	1.05E-02	*CDKN1A*
RPRD1B	other	1.05E-02	*CCND1*
FOXG1	transcription regulator	1.05E-02	*CDKN1A*
PRDM16	transcription regulator	1.05E-02	*CDKN1A*
HSF1	transcription regulator	1.38E-02	*EIF4A2,HSPA4L,TRA2B*
ZMIZ2	transcription regulator	1.57E-02	*CCND1*
TIP60	complex	1.57E-02	*RBL1*
Cebp	complex	2.60E-02	*PRLR*
E2f	group	3.74E-02	*MYBL2,RBL1*
E2F8	transcription regulator	4.13E-02	*CCND1*
NR0B1	ligand-dependent nuclear receptor	4.13E-02	*HSPA4L*

The analysis of upstream regulators showed the upstream regulators of the data in the skeleton dataset. The results are shown in Table 27.

11.11.3.5. Diseases and Functions

A selection of the diseases and functions with a p-value below 0.05, which appeared to be associated with SLE pathogenesis can be seen in Table 28. A complete list of the result of the diseases and functions associated with this dataset can be found in ADDENDUM G.

Table 28: Disease associations, generated by the Core Analysis of the All Top Regulated dataset, which at my discretion appeared to be associated with SLE.

Categories	Diseases or Functions Annotation	p-Value	Molecules	# Molecules
Immunological Disease	systemic autoimmune syndrome	1.59E-07	*ARIH1*,*CCL1*,*CCR1*,*CTSA*, *H3F3A/H3F3B*,*IFIT1*, *IGFBP5*,*ISG15*,*MAFB*, miR-155-5p,miR-17-5p, miR-196a-5p,miR-21-5p, miR-298, *OASL*,*PTPN22*, *RPL18A*,*RPSA*,*SLC7A2*, *SPP1*,*STAT1*,*TNFAIP3*, *UBE2L3*	23
Inflammatory Disease	chronic inflammatory disorder	5.77E-04	*ARIH1*,*CCL1*,*CCR1*, *H3F3A/H3F3B*,*IGFBP5*, *MAFB*,miR-155-5p,miR-17-5p ,*PDE3A*,*PTPN22*,*RASGRP3*, *RPL18A*,*RPSA*,*SLC7A2*,*SPP1*, *STAT1*,*TNFAIP3*	17
Inflammatory Response	inflammation of organ	4.63E-05	*ARIH1*,*H3F3A/H3F3B*, *IFI27*,miR-125b-5p, miR-145-5p, miR-155-5p,miR-17-5p , miR-21-5p,miR-224-5p, miR-23a-3p,miR-423-5p, *PDE3A*,*SPP1*,*STAT1*, *SYNE2*,*TNFAIP3*	16
Inflammatory Response	inflammation of body region	1.31E-03	miR-125b-5p,miR-145-5p, miR-17-5p,miR-21-5p, miR-224-5p, miR-23a-3p, miR-423-5p,*PDE3A*,*SPP1*, *STAT1*	10

Table 29: The five networks generated by IPA Core Analyses, using the skeleton dataset of Top regulating miRNAs and Top Regulated genes

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	*ABCC5*,*AGO2*,*ALOX12B*,*ARPC2*,*ATP6V0A1*,*BACH2*,*BRCA1*, *CARS*,*CDKN2A*,*CLDN1*,*DDX46*,*ERG*,*ESR1*,*ETS1*,*FAM3C*,*FOXD1*, *GAK*,*glutathione peroxidase*,*GTSE1*,*H3F3A/H3F3B*,*HEY2*,*HIF1A*, *HOXA7*,*IFI44*,*IFI44L*,*ISG15*,*MAGI1*,*MAPK12*,*MBD1*,*MECP2*,mir-324, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-194-5p (miRNAs w/seed GUAACAG), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *NDUFA13*,*NFKB2*,*NOTCH1*,*NUCKS1*,*NUPR1*,*P4HA2*, *PFKFB3*,*PHACTR2*,*PSEN1*,*PTPN22*,*RBMS1*,*RPS27A*, *SESN2*,*SIK2*,*SLC2A12*,*SLC39A8*,*SLC7A2*,*SMARCE1*, *STAT3*,*STIM1*,*STX3*,*TMEM19*,*TMEM97*,*TNFAIP8L1*,*TNIP1*,*TP53*, *UNC5B*,*VCP*,*WDR91*	43	29	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
2	*AHSP*,*AKAP8*,*ASF1B*,*ATP13A2*,*BTG3*,*CCND1*,*CD274*,*CDCA7L*, *CDK14*,*CDKN1A*,*Creb*,*Cyclin A*,*DCBLD2*,*DCTPP1*,*DRAP1*, *ERBB4*,*ERMAP*,*estrogen receptor*,*EXOSC8*,*FANCI*,*G6PD*,*GNAS*, *HDAC9*,*HDAC10*,*HIST2H2BE*,*Histone H1*,*Histone h3*,*Histone h4*, *HNRNPK*,*HSD11B2*,*HSPH1*,*IFI27*,*IGFBP5*,*INS*,*KAT2B*, *LOC102724788/PRODH*,	39	27	Cell Cycle, Cancer, Organismal Injury and Abnormalities

	miR-145-5p (and other miRNAs w/seed UCCAGUU), *MTBP,MYBL2,MYC,MYEF2,NCAPD2,NPAT,Patched, PDE3A,PPP2R2A,PRLR,PTTG1,RBL1,RBM39,RIF1, RNA polymerase II,RPL10A,RPRD1B,RUFY3,SERINC3, SFRP1,SMAD4,SPTA1,STAT1,STOM,SYNE2,TCF21, TFAP2A,TPM3,WNT7B,YWHAQ,YWHAZ,ZMIZ2,ZNF451*			
3	*ARID4A,ARIH1,ATXN7,CCL1,CCR1,CD44,CHAF1B,CSF2RA, DDX58,EDARADD,EFTUD2,EIF4A2,GRB2,HAX1,HCLS1, HNRNPA3,HNRNPL,HSF1,HSPA4L,IFIT1,IFNB1,IL1R1,IL36B, IL7R,IRAK1,IRAK2,IRAK3,IRF4,KARS,KAT2A,KHDRBS1,KHDRBS3, Laminin,MAFB,MATR3,* miR-146a-5p (and other miRNAs w/seed GAGAACU), *OASL,PAX5,PELI1,PELI3,PHLDB1,PLAT,RBBP4,RPL18A,RPLP2, RPSA,SAP30,SERPINH1,SMARCA4,SNRNP200,SPP1,SWISNF, SYNCRIP,TAB2,TADA2A,TADA2B,TFF1,TLR7,TLR8,TNFAIP3, TNFRSF9,TNFRSF11A,TRA2B, TRAF6,TSPAN8,UBE2L3,VAV2, WDR34,WIPF1,ZC3HAV1*	37	26	Cellular Function and Maintenance, Cellular Development, Hematological System Development and Function
4	*CCNI,LIN28A*	2	1	Cancer, Organismal Injury and Abnormalities, Cell Death and Survival
5	*CRB3,MPP5,MPP7*	2	1	Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization, Cellular Function and Maintenance

11.11.4. Selection of networks

Networks 1, 2, and 3 were selected for further analyses as they consisted of 29, 27, and 26 of my focus molecules, respectively.

11.11.5. Preparation of selected networks

For each network, nodes were identified and at my own discretion, spread out to create a network image with fewer overlapping relationship lines than in the original core network.

11.11.6. Expanding of selected networks using the Grow Tool

Using the Grow tool in the Build tab, the complete dataset of starting miRNAs, and genes, and interacting miRNAs, and genes found in miRecords and miRTarBase, was used to expand each of the core networks. The results of expanding the 3 selected networks with the **All Genes and miRNAs** dataset are shown in Table 30, Table 31, and Table 32.

11.11.6.1. Core Network 1

Table 30: This table reports the number of molecules which were linked from the All Genes and miRNAs dataset to selected nodes in Core Network 1. The additions were performed by IPA using existing knowledge on gene interactions in the database.

Node selected	Number of molecules added	Node selected	Number of molecules added
SMARCE1	11	*NUPR1*	25
AGO2	25	*RBMS1*	1
BACH2	0	*FOXD1*	1
NOTCH1	17	miR-155-5p	3
HEY2	2	miR-23a-5p	0
miR-194-5p	1	miR-224-5p	0
NDUFA13	0	miR-17-5p	8
miR-298	0	*FAM3C*	0
STIM1	2	*MBD1*	6
PTPN	1	*ATP6V0A1*	0
WDR91	0		

11.11.6.2. Core Network 2

Table 31: This table reports the number of molecules which were linked from the All Genes and miRNAs dataset to selected nodes in Core Network 2

Node selected	Number of molecules added	Node selected	Number of molecules added
HDAC9	3	*BTG3*	0
CDKN1A	1	*INS*	0
estrogen receptor	12	*ASF1B*	1
IRPRD1B	1	*AHSP*	1
MTBP	2	*RPL10A*	1
ZMIZ2	0	*G6PD*	1
HSPH1	0	*SPTA1*	0
SERPINC3	0	*CD274*	2
CDCA7L	7	*ATP13A2*	0
YWHAQ	11	*EXOSC8*	0
TFAP2A	16	*ZNF451*	0
Creb	19	*SMAD4*	25
AKAP8	0	*KAT2B*	3
RBM39	1	*PRCR*	1

11.11.6.3. Core Network 3

Table 32: This table reports the number of molecules which were linked from the All Genes and miRNAs dataset to selected nodes in Core Network 3

Node selected	Number of molecules added	Node selected	Number of molecules added
TADA2A	2	*SNRNP200*	2
KAT2A	4	*EFTUD2*	1
ATXN	7	*WIPF1*	3
PAX	5	*PHLDB1*	1
TADA2B	1	*TSPAN8*	1
CHAF1B	2	*IL1R1*	7
RBBP4	9	*HCLS1*	1
SAP30	0	*KHDRB53*	3
IRF4	19	*CSF2RA*	0
TNFRSF9	2	*KHDRBS1*	25
TFF1	18	*KARS*	3
EDARADD	0	*PEL13*	1
TNFRSF11A	0	*RPLP2*	2
HSF1	25	*SYNCRIP1*	1
ZC3HAV1	0	miR-146a-5p	13

A full-page print out of these three expanded networks, prepared for the overlays with the subsets of data, can be viewed in ADDENDUM H, ADDENDUM I, and ADDENDUM J.

11.11.7. Overlaying of Core Network 1

As shown in Table 30, Core Network 1 was generated having 29 molecules from the dataset of All Top Regulated Genes and miRNAs being placed into a network of 70 molecules. The top diseases and functions found to be associated with molecules in Core Network 1 were cancer, organismal injuries and abnormalities, and reproductive system diseases.

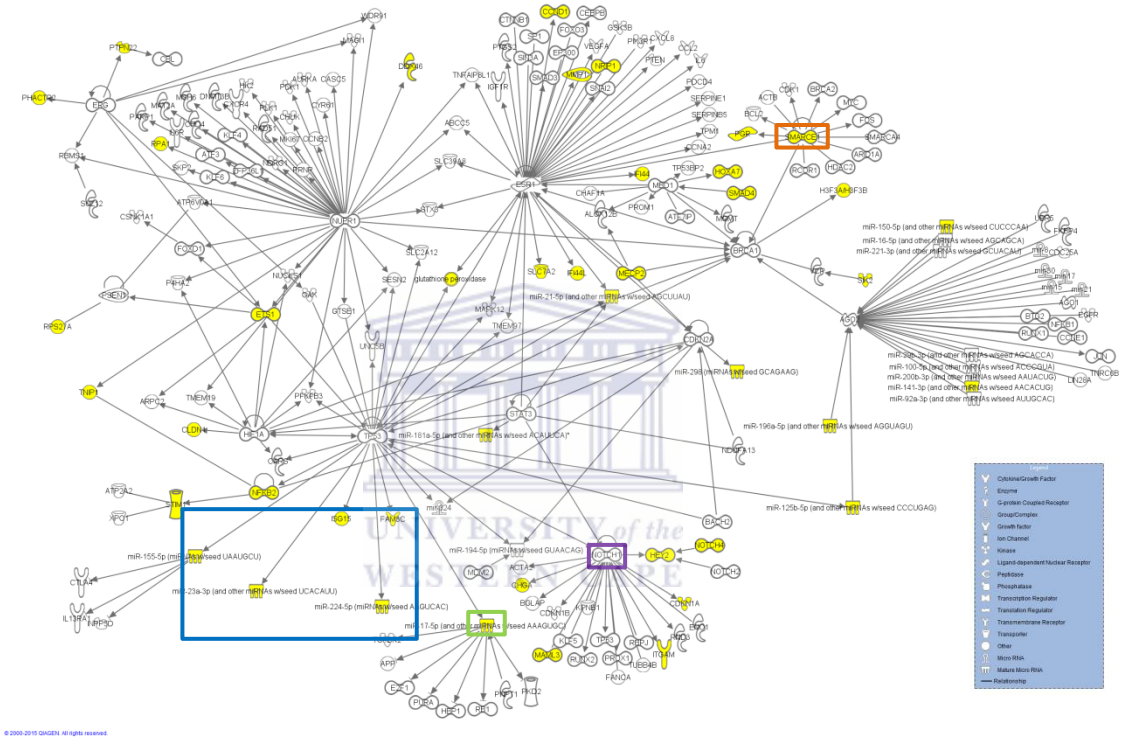

Figure 29: Core Network 1 overlaid with Starting Genes and miRNAs. Of the 866 Starting Genes and miRNAs mapped in IPA, 43 IDs were found in Core Network 1, composed of 198 molecules.

As can be seen in Figure 29, 43 of the curated genes and miRNAs were mapped onto Core Network 1, as highlighted in yellow. Most of these curated genes and miRNAs lie on the outskirts of the network, and are regulated by genes or miRNAs not in the starting list. Two genes and three miRNAs (outlined by the blue box) are present in the list of curated genes and miRNAs, but are all regulated by *TP53*, a gene not present in the starting list.

In contrast, miR-17-5p, a curated miRNA ID (light green box), was shown to directly interact with eight genes not present in the starting list of IDs. Similarly, *SMARCE1* (orange box) appeared as a node in this network, as it directly interacted with PGR, a curated gene, but was also shown to interact with 12 genes not associated with SLE in recent literature. The NOTCH1 gene (outlined by the purple box) exhibited direct interactions with five known lupus-associated genes. One of the interactions, with the gene *HEY2*, was expanded and then further shown to also have direct interactions with the known lupus gene *NOTCH4*, as well as with *NOTCH2*, a gene not yet associated with lupus pathogenesis.

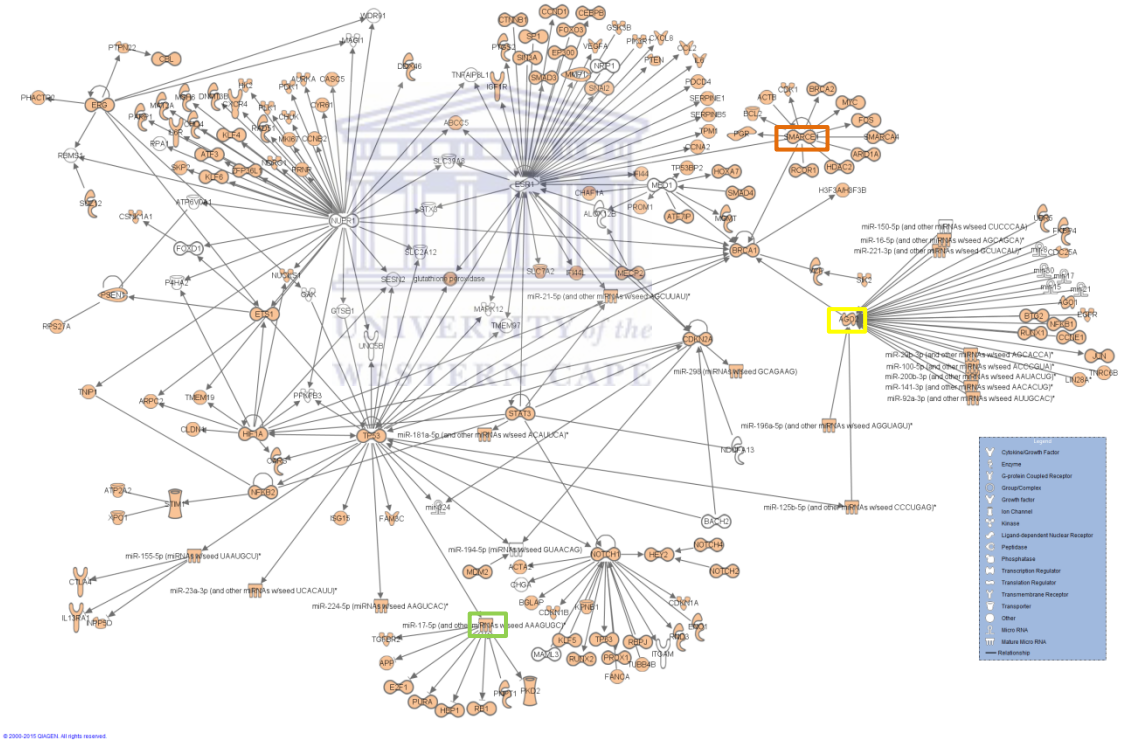

Figure 30: Core Network 1 overlaid with Interacting Genes and miRNAs. Of the 2969 Interacting Genes and miRNAs mapped in IPA, 162 IDs were found in Core Network 1, composed of 198 molecules.

It is evident from Figure 30 that most of the molecules in the list of interacting genes and molecules can be seen in Core Network 1, as highlighted in brown. As shown in Figure 29 and now in Figure 30, miR-17-5p, and *SMARCE1* appear as a nodal miRNA, and nodal gene, respectively. However, in contrast to Figure 29, miR-17-5p and *SMARCE1* are now also shown to each interact with numerous

genes from the list of interacting genes, gathered from the analysis through miRTarBase, and miRecords.

The gene *AGO2* was not identified as a starting gene in Figure 29. However, in Figure 30, *AGO2* (yellow box) is identified as one of the genes shown to interact with a known lupus-associated miRNA. Here, *AGO2* is also shown to interact directly with numerous miRNAs, which were found as new interactions for the known lupus-associated genes, and were not present in the starting list of lupus-associated miRNAs.

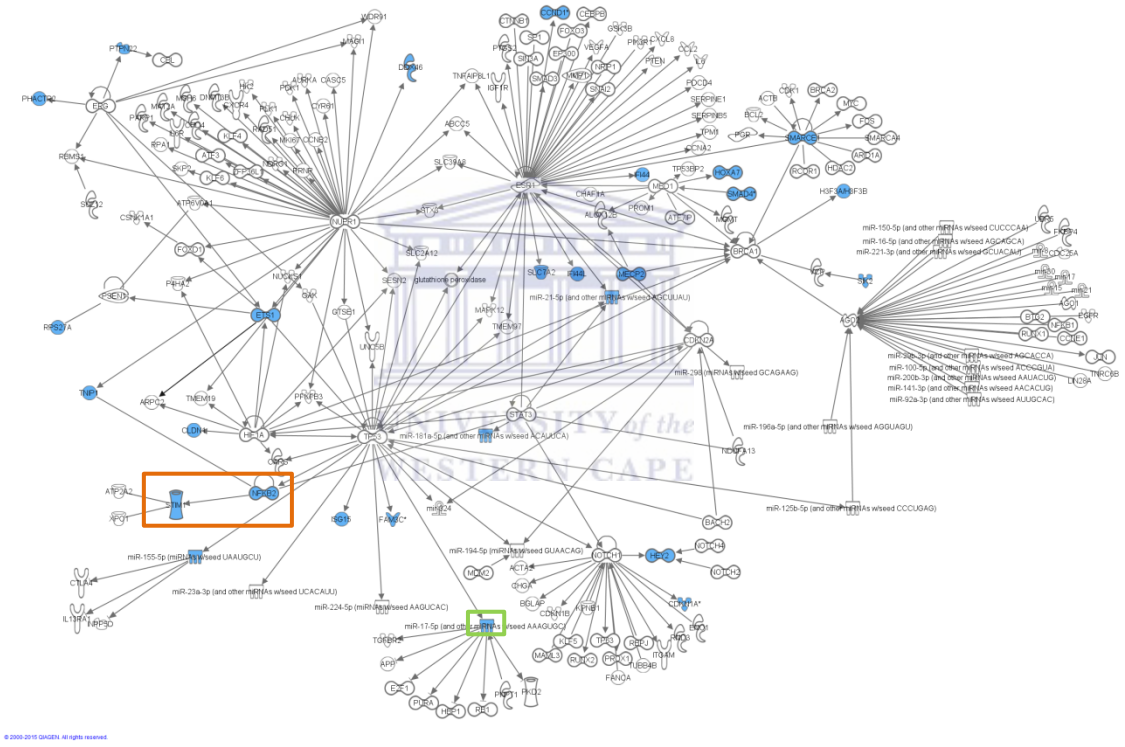

Figure 31: Core Network 1 overlaid with All Top Regulating miRNAs and associated genes. Of the 116 All Top Regulating miRNAs, and associated gene targets mapped in IPA, 27 IDs were found in Core Network 1, composed of 198 molecules.

Figure 31 is overlaid with the 'All Top Regulating miRNAs and associated genes' list. These miRNAs were selected as they interacted with ten or more of the known lupus-associated genes. Of the six top regulating miRNAs selected to build the core networks (see 11.8), four miRNAs can be seen above. MiRNA miR-17-5p (light green box), which is also a miRNA from the starting list of miRNAs (see

Figure 29) is shown here to be one of the top regulating miRNAs for the starting list of lupus-associated genes.

Interestingly, *STIM1* and *NFKB2* (orange box) are both shown here to be two genes which interact with the top regulating miRNAs. Unlike all of the other genes overlaid in Figure 31, these two genes are also shown to directly interact with each other, as well as with *TP53*, identified in Figure 29 to be a gene found to interact with a known lupus-associated miRNA.

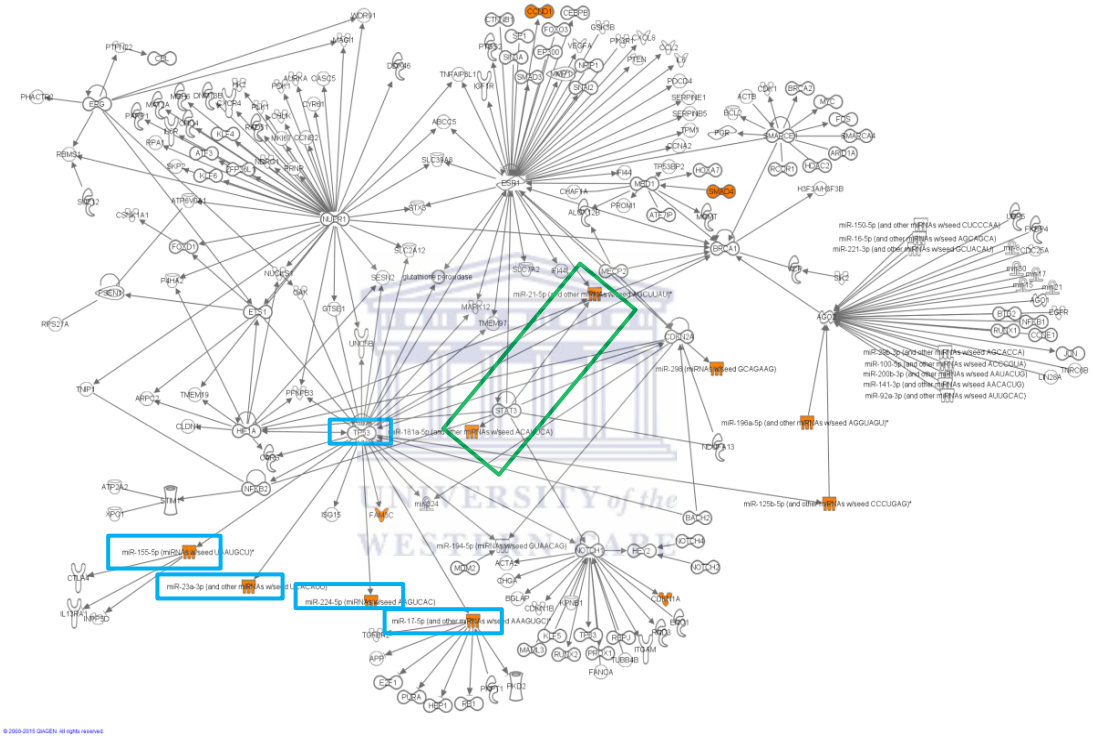

Figure 32: Core Network 1 overlaid with Top Regulated Genes and associated miRNAs. Of the 40 All Top Regulating genes, and associated miRNA targets mapped in IPA, 13 IDs were found in Core Network 1, composed of 198 molecules.

Figure 32 is overlaid with the 'All Top Regulating genes and associated miRNAs' list. These genes were selected as they interacted with ten or more of the known lupus-associated miRNAs. Of the seven top regulating genes selected to build the core networks (see 11.9), four genes can be seen above. Of the 33 miRNAs shown in 11.9 to be regulated by the seven top regulating genes, nine miRNAs can be identified here in Figure 32. *STAT3*, identified in Figure 32 as a gene interacting with a known lupus-associated miRNA, is also shown here to directly interact with miR-181a-5p, and miR-21-5p (outlined by a dark green box). *TP53*

(outlined by a blue box), was also identified in Figure 30 as a gene interacting with a known lupus-associated miRNA, is shown in Figure 32 to directly interact with four miRNAs (blue box) which interact with top regulated genes.

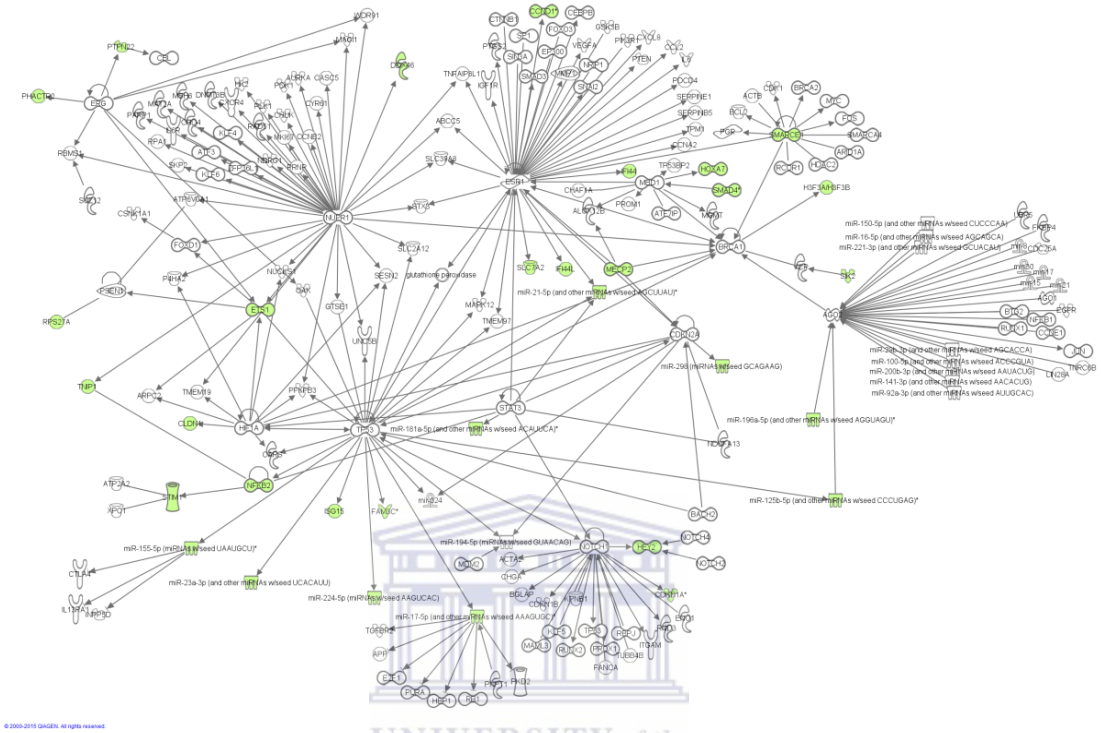

Figure 33: Core Network 1 overlaid with All Top Regulated Genes and miRNAs and associations. Of the 156 All Top Regulating Genes and miRNAs, and associated targets mapped in IPA, 32 IDs were found in Core Network 1, composed of 198 molecules.

Figure 33 is overlaid with all the top regulating genes and miRNAs, as well as their associations. In Core Network 1, this overlay shows that most of the genes and miRNAs from the 'All Top Regulated' dataset are widespread across the network.

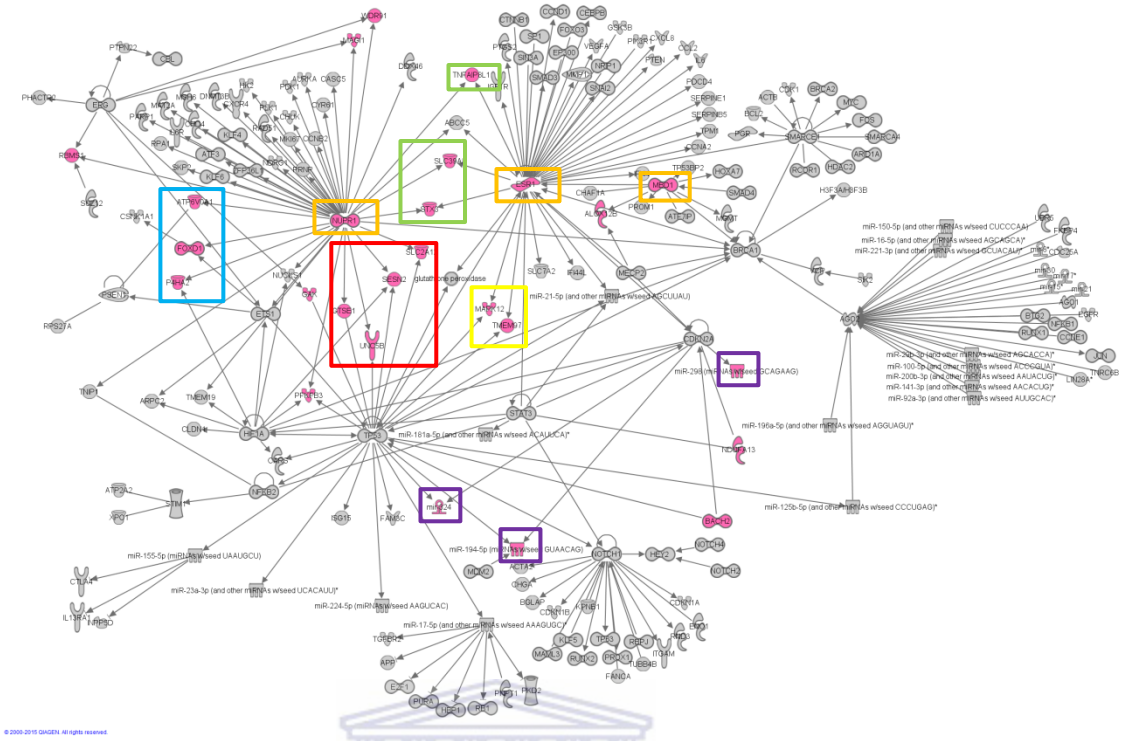

Figure 34: Core Network 1 overlaid with All Genes and miRNAs, IPA-added molecules in Pink. Of the 3687 All Genes and miRNAs, and associated targets mapped in IPA, 26 IDs were not overlaid in Core Network 1, and thus allowed for the identification of the molecules added

Figure 34 was overlaid with the complete list of all curated genes and miRNAs, and their corresponding interactions obtained from the analysis of miRTarBase and miRecords. The genes and miRNAs which were not overlaid by the dataset were then selected and highlighted in pink, as shown above.

A number of new genes, added in by IPA to complete Core Network 1, were placed centrally within the network. Of note, *NUPR1*, *ESR1*, and *MED1* (all outlined by an orange box), are all genes which were not present in the starting list of genes, or in the list of genes interacting with known lupus-associated miRNAs. However, as demonstrated by the many genes connected to, and overlaid by the complete dataset of all the curated, and interactions data, *NUPR1*, *ESR1*, and *MED1*, are shown in this network to interact with many of the known lupus-associated genes and miRNAs, as well as those found to interact with the curated dataset.

Three miRNAs (outlined in purple), miR-298, miR-194-5p, and mir-324, were also added by IPA to complete this network. These three miRNAs are all linked to *CDKN2A*, a gene which in Figure 30 was illustrated to be a new gene found by assessment of miRTarBase and miRecords, to interact with a known lupus-associated miRNA.

Three genes (outlined by a light green box), *STX3*, *SLC39A8*, and *TNFAIP8L1*, are all shown in Figure 34 to be important linking genes between the *NUPR1*, and *ESR1* nodal genes, which have also been added in by IPA to complete this network.

Another three genes (outlined by a blue box), *P4HA2*, *FOXD1*, and *ATP6V0A1*, are also genes which have been added in by IPA. However, it is important to also note that these three genes have direct interactions with genes which have been gathered through the interactions assessments of the curated miRNAs list with miRTarBase and miRecords.

Similarly, a five of genes (outlined by a red box) were shown in Figure 34 to be key genes linking *NUPR1* (an IPA-added gene) and *TP53* (a gene found to interact with a known lupus-associated miRNA). Two genes, *MAPK12* and *TMEM97* (yellow box), were shown to link the nodal genes *TP53*, and *ESR1*.

11.11.8. Overlaying of Core Network 2

As can be seen in Table 29, Core Network 2 was generated as having 27 molecules from the dataset of All Top Regulated Genes and miRNAs being placed into a network of 70 molecules. The top diseases and functions found to be associated with molecules in Core Network 2 were cell cycle functioning, cancer, and organismal injury and abnormalities.

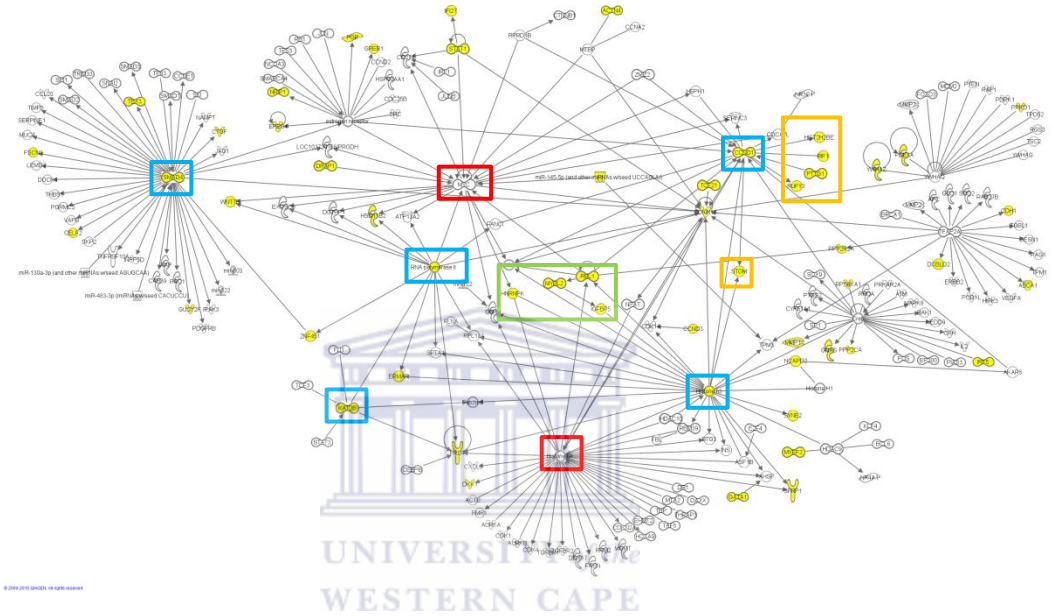

Figure 35: Core Network 2 overlaid with Starting Genes and miRNAs. Of the 866 Starting Genes and miRNAs mapped in IPA, 54 IDs were found in Core Network 2, composed of 207 molecules.

As can be seen in Figure 35, 54 of the curated genes and miRNAs were mapped onto Core Network 2. The distribution of these genes and miRNAs is widespread, with most of the molecules lying as inter-connectors between defined nodes.

Of note, *SMAD4*, *KAT2B*, *Histone h3*, *RNA polymerase II*, and *CCND1* (all outlined by a blue box), are all present in the curated list of genes and miRNAs, and they appear ad nodal genes in Core Network 2. Each of these nodal genes clearly interacted with a number of genes, which are not present in the starting list of genes.

Interestingly, *CCND1* also illustrates direct interactions with genes from the curated genes and miRNAs list, namely *HIST2H2BE*, *RIF1*, *PTTG1*, *RUFY3*, and *STOM* (orange boxes).

Four genes, *HNRNPK*, *MYBL2*, *RBL1*, and *IGFBBP5* (all outlined by a light green box) are also shown here to be present in the starting list of curated genes and miRNAs, but form integral connections between *Histone h4* and *MYC* (both outlined by a red box), which are two genes not present in the starting list of genes.

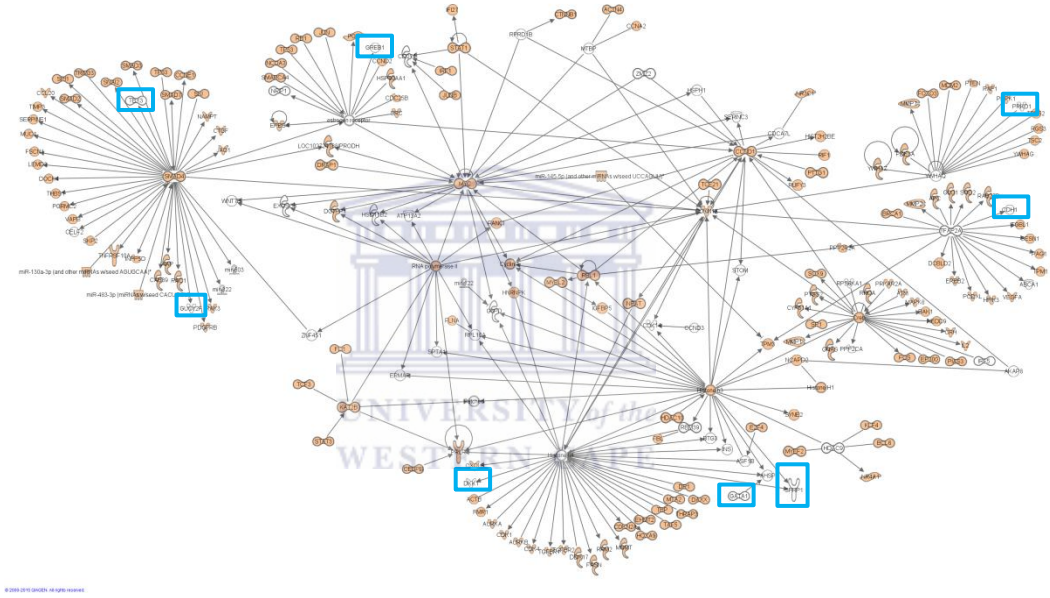

Figure 36: Core Network 2 overlaid with Interacting Genes and miRNAs. Of the 2969 Interacting Genes and miRNAs mapped in IPA, 157 IDs were found in Core Network 2, composed of 207 molecules.

Figure 36 was generated by overlaying Core Network 2 with the list of all interacting miRNAs and genes obtained from the analysis of the curated genes and miRNAs, in miRTarBase, and miRecords. Most of the molecules in Core Network 2 were shown to consist of genes and miRNAs from the list of interactions. However, eight genes (outlined by blue boxes), were present in the curated list of genes (see Figure 35), but were not found to interact with the list of miRNAs curated from literature.

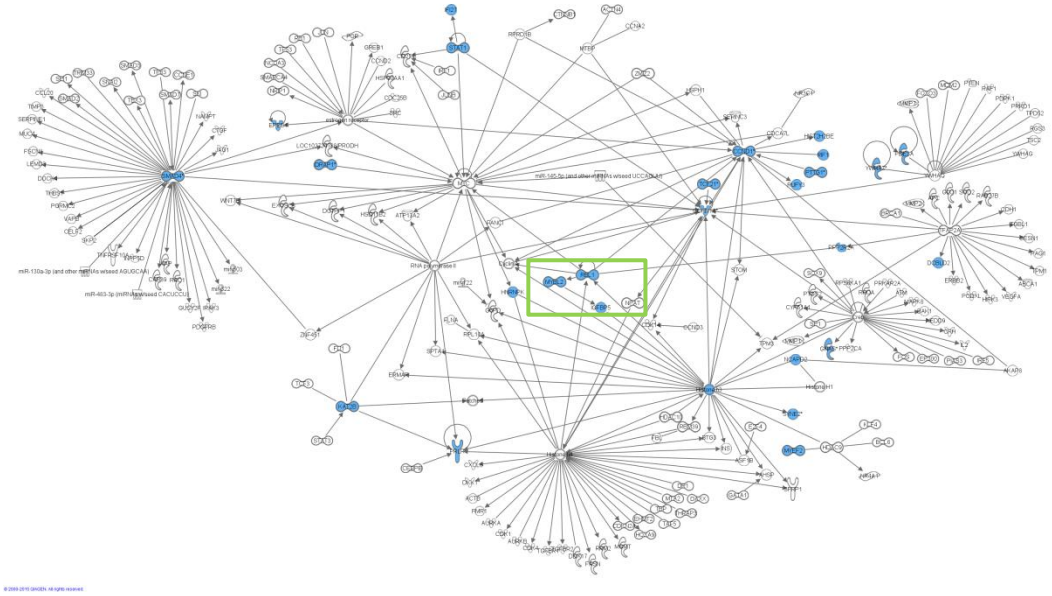

Figure 37: Core Network 2 overlaid with All Top Regulating miRNAs and associated genes. Of the 116 All Top Regulating miRNAs, and associated gene targets mapped in IPA, 27 IDs were found in Core Network 2, composed of 207 molecules.

Figure 37 is overlaid with the 'All Top Regulating miRNAs and associated genes' list. Of the six top regulating miRNAs selected to build the core networks (see 11.8), none of the miRNAs were overlaid above. A total of 27 genes, which were shown to interact with the top regulating miRNAs, were overlaid in Figure 35.

Four genes, *HNRNPK*, *MYBL2*, *RBL1*, and *IGFBP5* (all outlined by a light green box), were shown here to be regulated by the Top Regulating miRNAs. Interestingly, these four genes, were also overlaid with the starting list of lupus-associated genes and miRNAs (see Figure 35), as well as by the list of interacting molecules (see Figure 36). Thus, these four genes can be said to be lupus-associated in terms of literature source, as well as associated with lupus, because they have been identified as interacting with known lupus-associated miRNAs.

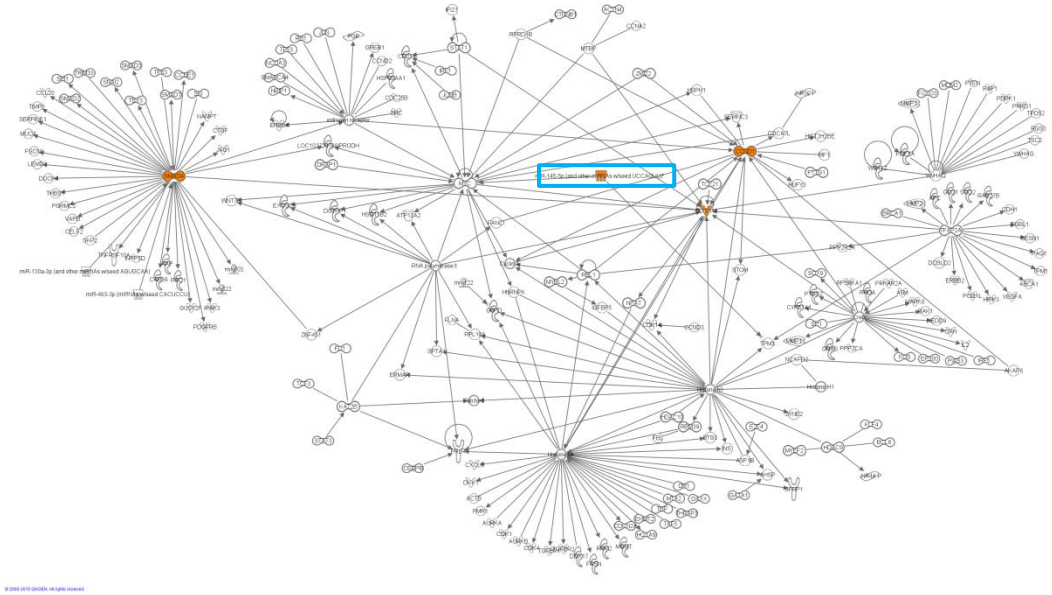

Figure 38: Core Network 2 overlaid with Top Regulated Genes and associated miRNAs. Of the 40 All Top Regulating genes, and associated miRNA targets mapped in IPA, 4 IDs were found in Core Network 2, composed of 207 molecules.

Figure 38 is overlaid with the 'All Top Regulating genes and associated miRNAs' list. Of the seven top regulating genes selected to build the core networks (see 11.9), three genes can be seen above. Of the 33 miRNAs shown in 11.9 to be regulated by the seven top regulating genes, one miRNA miR-145-5p can be identified Figure 38, outlined by a blue box. Compared to Core Network 1 (Figure 32), Core Network 2's overlay with the All Top Regulated Genes, shows much fewer of this subset of data being present in Figure 38.

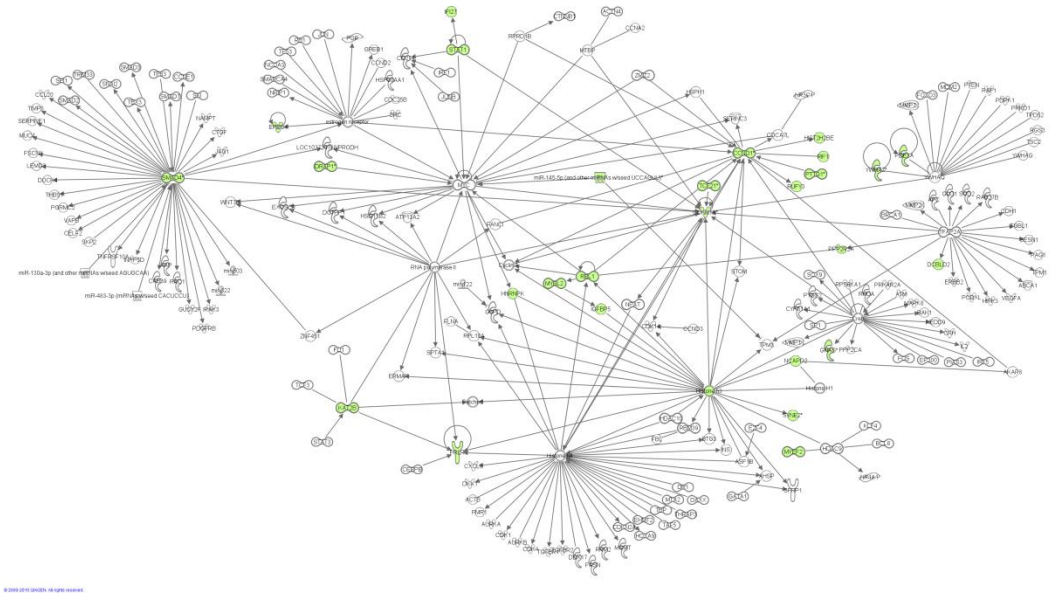

Figure 39: Core Network 2 overlaid with All Top Regulated Genes and miRNAs and associations. Of the 156 All Top Regulating Genes and miRNAs, and associated targets mapped in IPA, 28 IDs were found in Core Network 2, composed of 207 molecules.

Figure 39 shows Core Network 2 overlaid with all the Top Regulated genes, miRNAs, and their corresponding interactions, as determined by analysis of miRTarBase and miRecords. All of the genes highlighted by the overlay above appear in either Figure 37 or Figure 38.

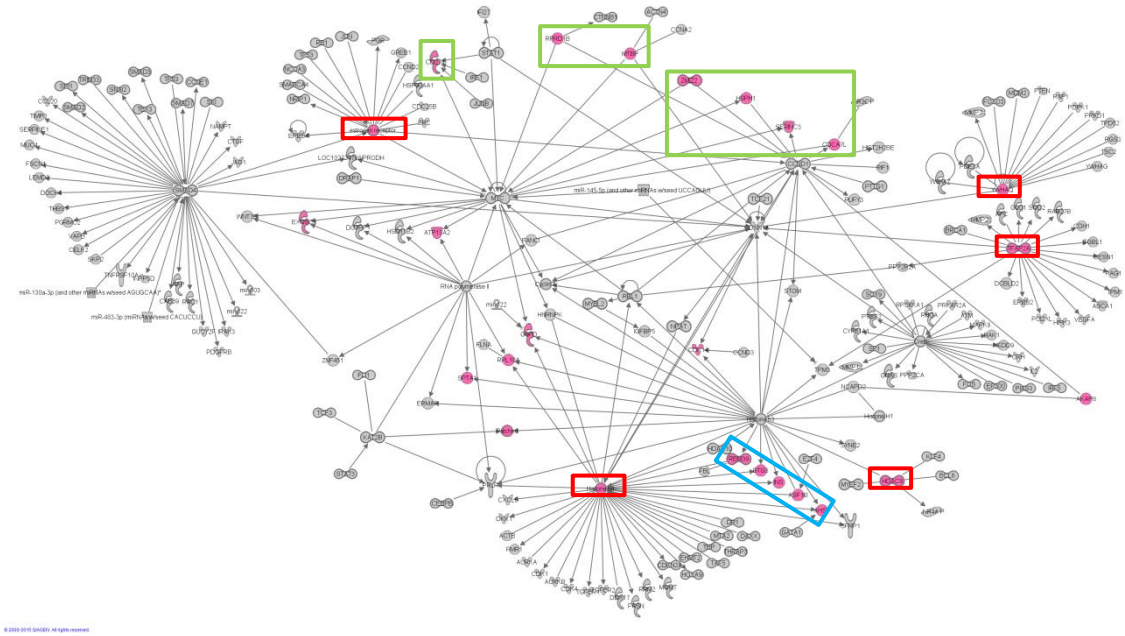

Figure 40: Core Network 2 overlaid with All Genes and miRNAs, IPA-added molecules in Pink. Of the 3687 All Genes and miRNAs, and associated targets mapped in IPA, 25 IDs were not overlaid in Core Network 2, and thus allowed for the identification of the molecules added by IPA.

Figure 40 was overlaid with the complete list of all curated genes and miRNAs, and their corresponding interactions. Those genes not overlaid by the dataset, but added in by the IPA database, were selected and highlighted in pink.

A cluster consisting of five genes (outlined by a blue box) namely RBM39, BTG3, INS, ASF1B, AHSP, were all shown to interact with a known lupus-associated gene, Histone h3 (as seen in Figure 35), and also with *Histone h4*, a new gene added in by IPA to complete this section of Core Network 2.

A number of genes (outlined by a light green box), were also added by IPA, and can be seen here to interact either directly or indirectly, with a starting- and interacting gene named *CCND1*, and *MYC*- which was only found to interact with a known lupus-associated miRNA.

Of importance, Figure 40 shows how, along with *Histone h4* already described, *HDAC9*, *TFAP2A*, *YWHAQ*, and *estrogen receptor* (all outlined by a red box), are new genes added by IPA, and appear to be nodal genes, interacting with numerous genes from the complete dataset of starting, and interacting genes and miRNAs.

11.11.9. Overlaying of Core Network 3

As can be seen in Table 29, Core Network 3 was generated having 26 molecules from the dataset of All Top Regulated Genes and miRNAs being placed into a network of 70 molecules. The top diseases and functions found to be associated with molecules in Core Network 3 were cellular function and maintenance, cellular development, and haematological system development and function.

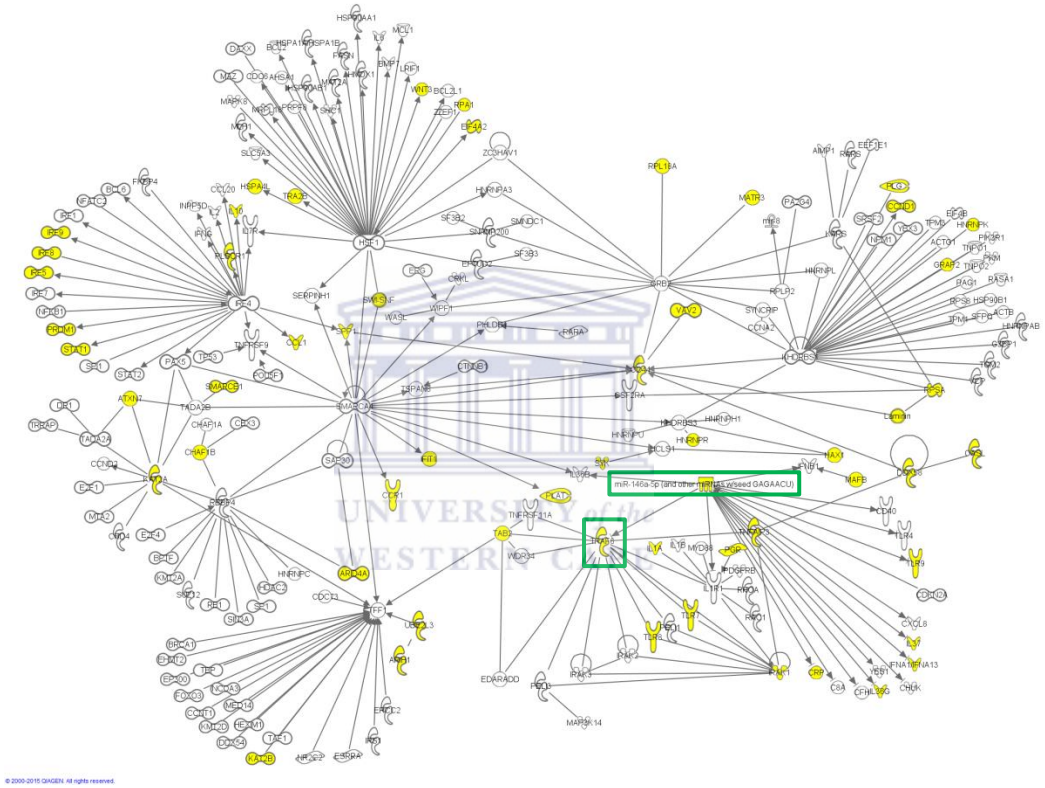

Figure 41: Core Network 3 overlaid with Starting Genes and miRNAs. Of the 866 Starting Genes and miRNAs mapped in IPA, 56 IDs were found in Core Network 3, composed of 216 molecules

Figure 41 is overlaid with the dataset containing all the curated genes and miRNAs. A total of 35 genes and one miRNAs were mapped onto Core Network 3 above. Most of the genes mapped from this dataset lay on the periphery of the network. Of the 36 identifiers overlaid on this core network, only *TRAF6*, and miR-146a-5p appeared from this dataset to be a nodal gene and miRNA, respectively.

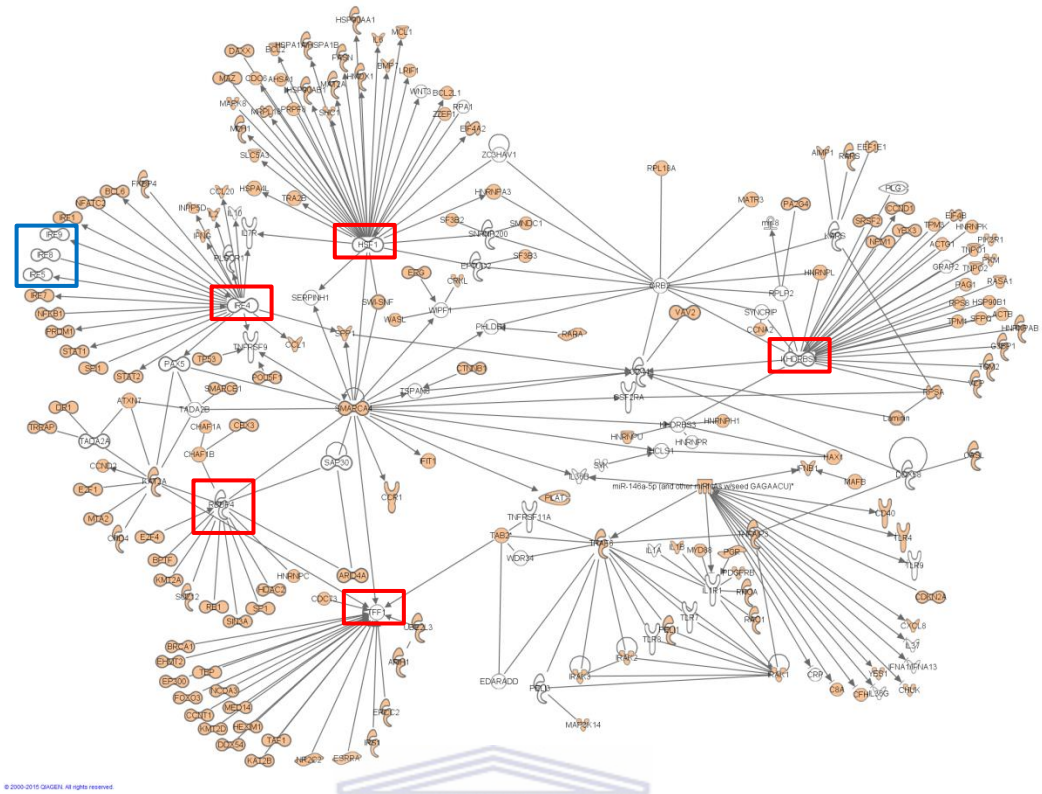

Figure 42: Core Network 3 overlaid with Interacting Genes and miRNAs. Of the 2969 Interacting Genes and miRNAs mapped in IPA, 164 IDs were found in Core Network 3, composed of 216 molecules

Figure 42 shows Core Network 3 overlaid with all the genes and miRNAs found to interact with the set of curated genes and miRNAs, showing an association with lupus pathogenesis. The distribution of the overlay is widespread across the network. However, it is interesting to note that three genes, *IRF5*, *IRF8*, and *IRF9*, are not overlaid by the interactions dataset. These three genes were however, overlaid by the starting genes and miRNAs dataset in Figure 41. Thus, these three genes are considered in recent literature, or reported in the meta-analysis, to be associated with SLE, but these three genes were according to the analysis of miRTarBase, and miRecords, not found to be regulated by any of the curated lupus-associated miRNAs. Most of the genes from the overlay dataset of interacting genes and miRNAs are shown to be interacting with nodal genes (outlined by a red box), which were not found in miRTarBase or miRecords, to interact with any of the known lupus-associated miRNAs.

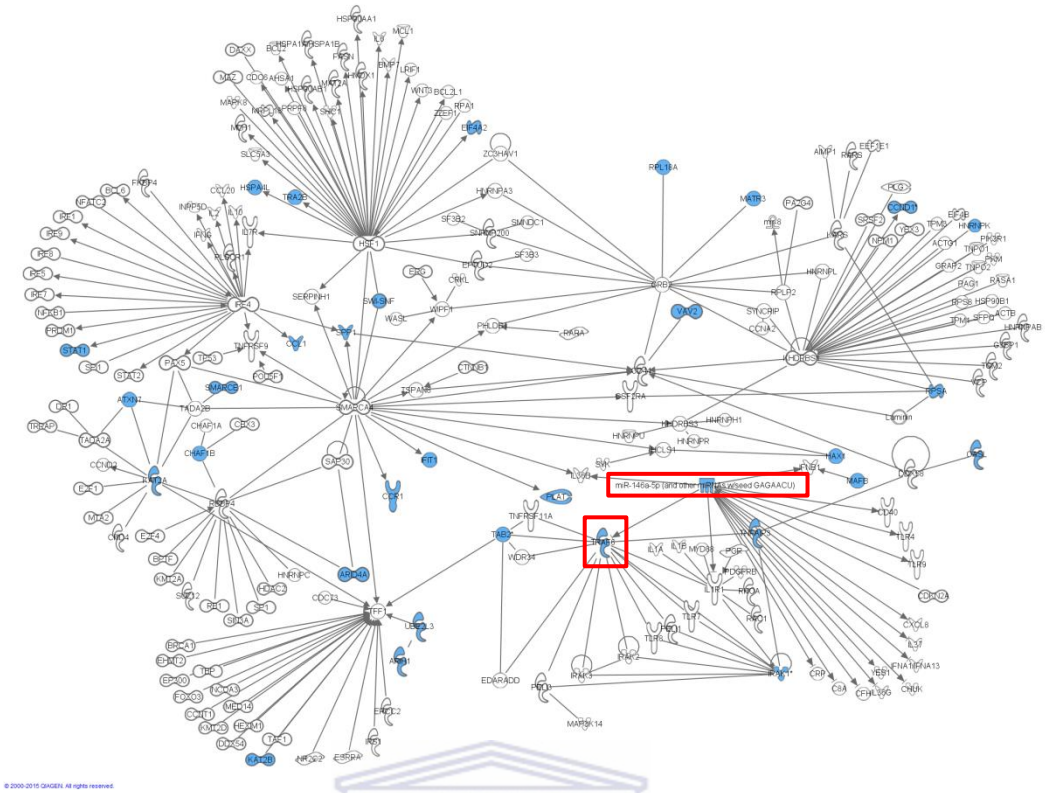

Figure 43: Core Network 3 overlaid with All Top Regulating miRNAs and associated genes. Of the 116 All Top Regulating miRNAs, and associated gene targets mapped in IPA, 32 IDs were found in Core Network 3, composed of 216 molecules

Figure 43 is overlaid with the dataset of all the miRNAs regulating 10 or more genes. Of the 32 identifiers overlaid, 31 were genes, and only one top regulating miRNA appeared in this network. The miRNA, miR-146a-5p, appeared in both the overlays for the starting, and the interacting datasets, see Figure 41 and Figure 42, respectively. This miRNA also appears as a nodal miRNA, and the only nodal gene overlaid in Figure 43 is *TRAF6* (both outlined by a red box).

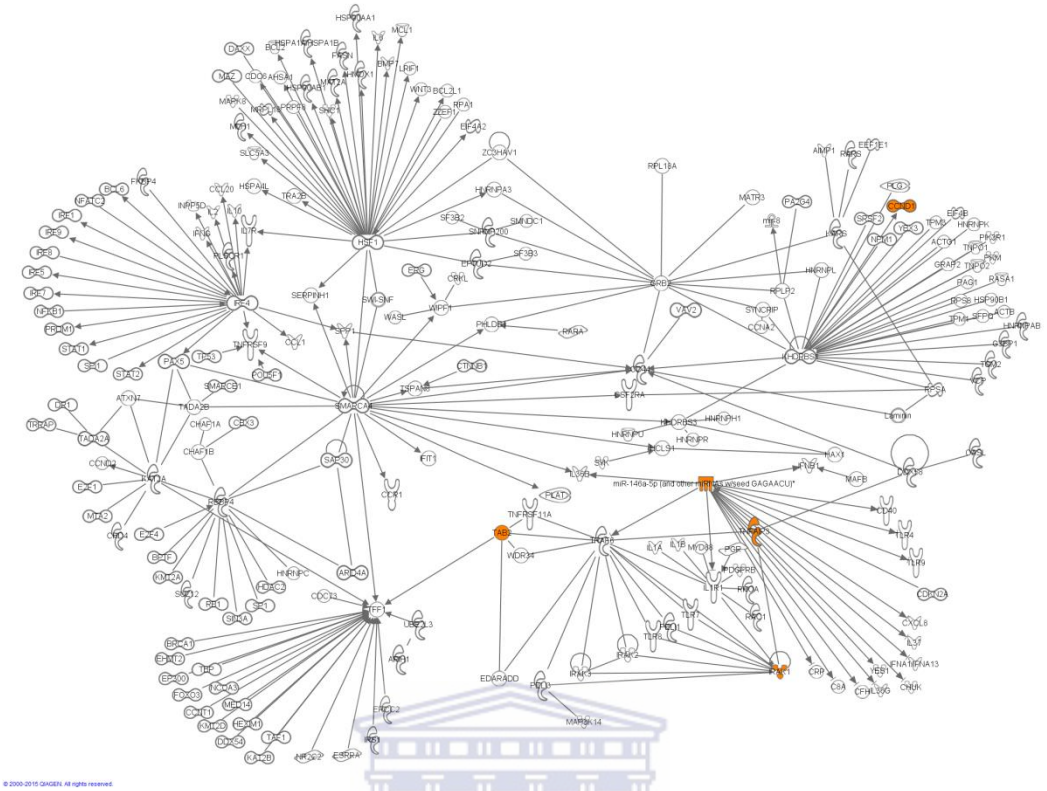

Figure 44: Core Network 3 overlaid with Top Regulated Genes and associated miRNAs. Of the 40 All Top Regulating genes, and associated miRNA targets mapped in IPA, 5 IDs were found in Core Network 3, composed of 216 molecules

Figure 44 shows the result of overlaying the dataset containing all genes regulated by ten or more miRNAs. Four genes and one miRNA from this dataset appeared in Core Network 3. It is interesting to note that these five molecules all appeared in Figure 43, in which all the top regulated miRNAs and the genes they regulated were overlaid, unlike that seen in Core Networks 1, and 2, in which there was a noticeable difference between the two overlays, see Figure 31 and Figure 32, Figure 37 and Figure 38.

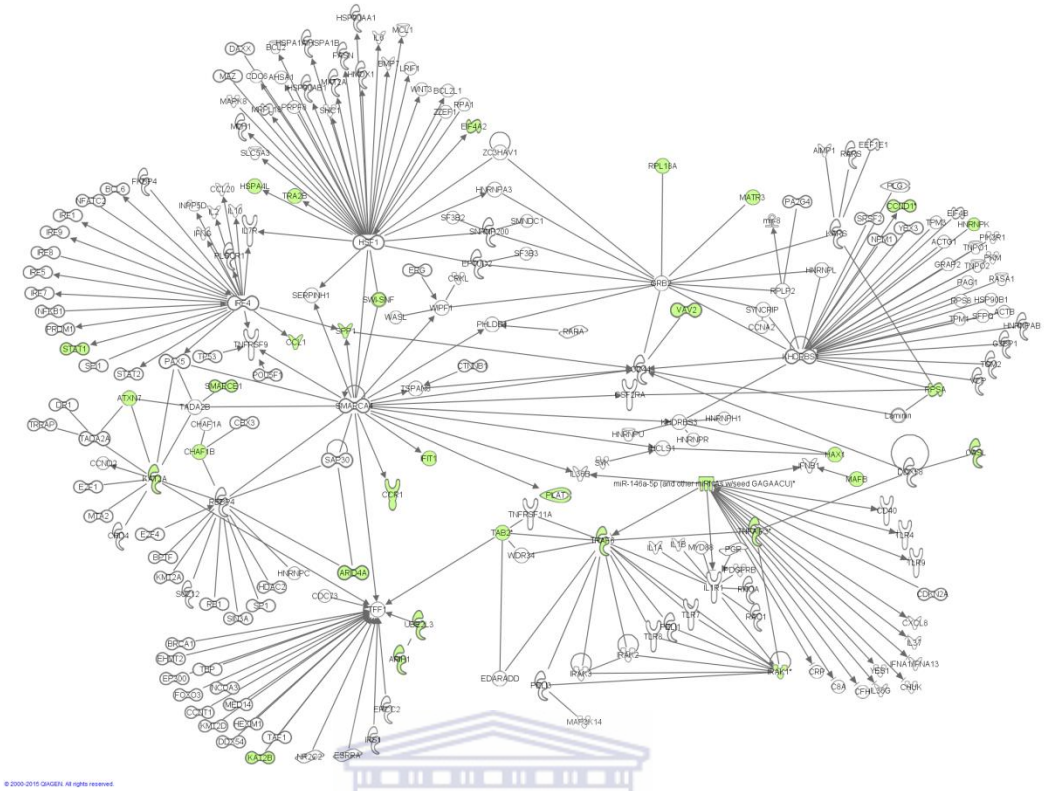

Figure 45: Core Network 3 overlaid with All Top Regulated Genes and miRNAs and associations. Of the 156 All Top Regulating Genes and miRNAs, and associated targets mapped in IPA, 32 IDs were found in Core Network 3, composed of 216 molecules.

Figure 45 is a result of Core Network 3 being overlaid with the complete dataset of all the top regulated genes and their associated miRNAs, and all the top regulating miRNAs and their association. Thus, Figure 45 is a combined overlay of both datasets used in Figure 43 and Figure 44. Figure 45 thus shows that of the top regulators and their interactions, most of the molecules are widespread across the network, and few form nodes. Most molecules can be found at the periphery of nodal interactions.

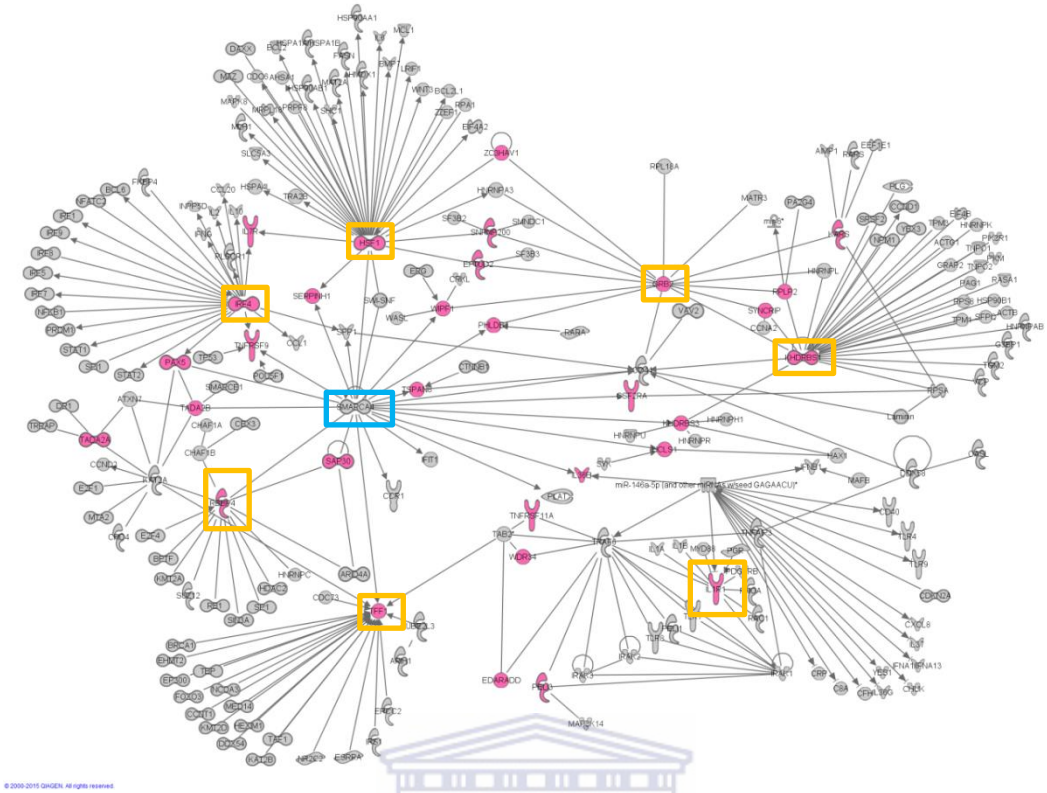

Figure 46: Core Network 3 overlaid with All Genes and miRNAs, IPA added molecules in Pink. Of the 3687 All Genes and miRNAs, and associated targets mapped in IPA, 31 IDs were not overlaid in Core Network 3, and thus allowed for the identification of the molecules added in by IPA to complete the network structure.

Figure 46 was generated by first overlaying Core Network 3 with the complete list of starting, and interacting genes and miRNAs. Those molecules which were not overlaid were then manually selected and highlighted in pink.

Of the 31 molecules highlighted in pink, seven genes were identified as nodal genes (all outlined by an orange box), which were not present in the curated genes and miRNAs list, or the interactions list derived from miRTarBase, and miRecords. These six genes were: *TFF1*, *RBBP4*, *IRF4*, *HSF1*, *GRB2*, *KHDRBS1*, and *IL1R1*.

It is important to note that *SMARCA4* (outlined by a blue box), is a gene overlaid only in Figure 42, yet it appears as an important node, interacting with numerous genes from the complete dataset of all curated and interaction data, as well as with new genes added by IPA.

11.12. Identification of candidate SLE-associated genes from pathway analysis

11.12.1. Core Network 1

Core Network 1 consisted of 70 molecules, of which 29 were focus molecules from **'All Top Regulated Genes and miRNAs and associations'** dataset file, and 26 molecules were added by the IPA database to complete the network. A number of nodal genes, added in by IPA, could be identified in Figure 34. *NUPR1*, *ESR1*, and *MED1*, were identified as the set of nodal genes for this network. Linking *NUPR1* and *ESR1* were 3 more IPA-added genes, *STX3*, *SLC39A8*, and *TNFAIP8L1*.

Another set of what may be described as "connector" genes, *ATP6V0A1*, *FOXD1*, and *P4HA2*, were also added in by IPA. They did not appear as nodes, however they were shown to interact with *CSNK1A1*, *PSEN1*, *HIF1A*, *ETS1*, and *NUCKS1*, all of which appeared in the overlay of interacting genes and miRNAs (see Figure 30), but only *ETS1* also appeared as a curated gene (see Figure 29)

Four genes were also found to be highlighted in the overlay of interacting genes (see Figure 30), and were not overlaid by the starting dataset of curated genes and miRNAs, as seen in Figure 29. These four genes were *AGO2*, *TP53*, *STAT3*, and *CDKN2A*. These four genes were not found in the recent biomedical literature perused for this study, but were found in miRTarBase and/or miRecords, to interact with known, curated lupus-associated miRNAs. *CDKN2A*, a gene added in by IPA, was also shown to interact with three miRNAs, added in by IPA. These three miRNAs had the following identifiers: miR-298, miR-194-5p, and mir-324. Thus, the 28 molecules from Core Network 1 were considered as potential molecules which may be associated with important biological processes in SLE pathogenesis, but may not have been explored as such in recent lupus studies.

11.12.2. Core Network 2

Core Network 2 consisted of 70 molecules, of which 27 were focus molecules from **'All Top Regulated Genes and miRNAs and associations'** dataset file, and 25 molecules were added by the IPA database to complete the network. A number of nodal genes, added in by IPA, could be identified in Figure 40.

For Core Network 2, five nodal genes added in by the IPA database could be identified, namely YWHAQ, TFAP2A, HDAC9, estrogen receptor, and Histone h4. Histone h4 was also directly linked to Histone h3, a gene present in the starting list of curated genes and miRNAs (see Figure 35).

CCND1 was overlaid as both a starting and interacting gene, while MYC was only overlaid as an interacting gene. Four IPA-added genes, namely *RPRD1B*, *ZMIZ2*, *HSPH1*, and *SERINC3* were shown to link the interactions of *CCND1* and *MYC*. *ATP13A2* and *EXOSC8* are two genes which were also added in by IPA, and shown to link *MYC* and *RNA polymerase II*. *HDAC9*, added in by IPA, was also shown to directly interact with *KLF4*, *BCL6*, and *NR4A1*, three genes which were not present in the starting list of genes, but were found to interact with known lupus-associated miRNAs curated from literature.

A set of five genes, all added in by IPA, were illustrated in Figure 40 to interact directly with a known lupus-associated gene, *Histone h3*, as well as *Histone h4*, a gene added in by IPA. These five genes, *RBM39*, *BTG3*, *INS*, *ASF1B*, and *AHSP* thus formed an important link between *Histone h3* and *Histone h4*.

Thus, 20 genes from Core Network 2 were considered as potential molecules which may be associated with important biological processes in SLE pathogenesis, but may not have been explored as such in recent lupus studies.

11.12.3. Core Network 3

Core Network 3 consisted of 70 molecules, of which 26 were focus molecules from **'All Top Regulated Genes and miRNAs and associations'** dataset file, and 31 molecules were added by the IPA database to complete the network. A number of nodal genes, added in by IPA, could be identified in Figure 46.

Seven nodal genes were added by the IPA database and identified in Core Network 3. These seven genes were: *HSF1*, *IRF4*, *RBBP4*, *TFF1*, *IL1R1*, *KHDRBS1*, and *GRB2*. Each of these genes was shown to have direct associations with numerous genes found in miRTarBase and miRecords, interacting with data from the list of curated SLE-associated miRNAs.

TADA2A and *TADA2B*, are two genes added in by IPA, which were also shown to interact with two known SLE-associated genes, *KAT2A* and *ATXN7* (see Figure 41). With the addition of TADA2A and TADA2B to the network, RBBP4, and TFF1 were also added to the network by IPA, and allowed for the addition of numerous genes from the list of interacting genes and miRNAs.

EDARADD and *PELI3* were also added by IPA and formed connections with *TAB2*, and *IRAK1* – two genes which were present both as starting genes and interacting genes. Interacting with *TRAF6* and *IRAK1*, *IL1R1* was added by IPA, and extended to interact with six genes found to be associated with known lupus miRNAs, and one gene (*IL1A*) only found as a curated gene. *PGR* was present as both a curated gene and a gene interacting with a known lupus –associated miRNA, and was found to be directly associated with *IL1R1* – and IPA-added gene.

SMARCA4, *CD40*, *TLR4*, *AIMP1*, and *KARS* were all found as genes interacting with known SLE-associated miRNAs (see Figure 42), but were not identified in the list of curated gene associated with SLE.

KHDRBS1 and *KHDRBS3* are two genes also added by IPA. *KHDRBS1* appeared as a nodal gene, directly interacting with numerous genes found in miRTarBase and

miRecords, to be associated with known lupus miRNAs. *KHDRBS3* was also found to directly interact with *HNRNPR* which was only overlaid as a starting gene, as well as with *HNRNPH1* and *HNRNPU*, both which were not overlaid as starting genes, but were found to interact with known lupus miRNAs.

Thus, 19 genes from Core Network 3 were considered as potential molecules which may be associated with important biological processes in SLE pathogenesis, but may not have been explored as such in recent lupus studies.

12. DISCUSSION

12.1. Overview of this study

MiRNAs are small molecules, ranging from 20 to 25 nucleotides in length. These biomolecules perform critical regulatory activities by binding to the target mRNA. By means of partially-complementary binding to the mRNA, miRNAs in animals are able to repress the production of the protein which would have been produced by the targeted mRNA. As a result, an abundance of studies have implicated aberrant gene expression in the onset and progression of numerous diseases, including Systemic Lupus Erythematosus, and at the centre of this gene regulation, lies the potent regulatory activities performed by miRNAs.

This study aimed to develop, and subsequently employ a novel bioinformatics workflow, which used existing knowledge to identify potentially novel candidate molecules, which may be associated with SLE. The workflow described within this thesis followed seven sequential steps, namely: data curation, data verification, target identification, cross-analysis of interactions, prioritization of top regulators, biological pathway analysis and lastly, the identification of novel candidate SLE-associated genes. Using curated lists of miRNAs and genes, I aimed to identify miRNA-target interactions in two well-established databases of experimentally-validated MTIs, namely miRTarBase and miRecords. Using both the miRTarBase and miRecords databases proved useful, as the former integrated the miRBase database, whilst the latter database reported no inclusion of the miRBase database. This difference between the two databases became evident during the target identification process.

A total of 40 miRNA IDs, and 81 gene names (in associated gene name format) were curated from recent biomedical literature as molecules which have been associated with SLE pathogenesis. A total of 749 gene IDs (in Ensembl Gene ID format) which were found to be differentially-expressed in a meta-analysis of lupus studies, were also included in this study with the permission of Dr Wendy Kröger.

12.2. Lupus-associated miRNAs

A total of 40 miRNA IDS were curated from four recently published peer-reviewed articles, of which two were primary sources for data, and two articles were reviews of existing data. In adherence to the standardised nomenclature of miRNAs described in the literature study, no adjustments were made to any of the miRNA IDs. Before identifying in miRTarBase and miRecords which genes were known to interact with these curated miRNAs, the miRNA names underwent what proved to be an important verification step in a downloaded version of the miRBase database.

The verification step aimed to first identify the curated miRNA IDs as exact (perfect) string matches in the miRBase database. If the curated miRNA ID was not found as is in the miRBase database, the ID then went through the second attempt verification step to identify the ID as a regular expression string. Thus, the aim of the first verification step (perfect string match) of the miRNAs in the miRBase database was not to discard any miRNAs that were not found in the database. It was rather to flag those that did not match the records in miRBase for attention, so as to invoke a check in the literature source for any typographical errors. It was also important not to deliberately alter the name of the miRNA, so as to avoid future errors. Thus, any alterations were only performed in a forward direction in terms of miRNA nomenclature, and by using the regular expressions described.

To support the identification of paralogues of the original miRNA curated from literature, regular expression string testing was employed using the Python script 'PS02_July_01_new_regexes.py'. Regular expression testing allowed for the identification of closely-related sequences (paralogues) which differed by only one or two nucleotides, and are indicated by the addition of a lettered suffix following its numeric value. Thus, regular expression testing allowed for the identification of records in the miRBase database which included a '-3p' or a '-5p' suffix, proving itself an asset in this workflow as it enabled the identification of

the original curated miRNA with its origin hairpin arm, where it would not have been identified as is in the miRBase database.

Of the 40 curated miRNAs, 13 miRNAs were found as exact string matches in the miRBase database. Of these 13 miRNAs, 11 were found as records in the miRTarBase database, and two miRNAs were not. In contrast, only three of these 13 miRNAs were found as records in the miRecords database. Of significance, two of the three records found in miRecords, namely for 'hsa-miR-184' and 'hsa-miR-17-5p', were also identified as part of the 11 miRNAs found in the miRTarBase database. However, when assessing the target identification, I found that 310 gene targets were identified for the miRNA 'hsa-miR-17-5p' in miRTarBase, but only one gene target, '*IL-8*' was found for this miRNA in the miRecords database. Of importance, this single gene target, '*IL-8*', identified for 'hsa-miR-17-5p' in miRecords, was not one of the 310 gene targets identified in miRTarBase. Thus, despite the large difference in the number of reported miRNA-target interactions in miRTarBase and miRecords (51 460, and 2705 respectively), it proved valuable to use more than one database for the detection of MTIs.

Subsequently, the remaining 27 curated miRNA IDs, not found as exact string matches in miRBase, were subjected to regular expression testing, thereby increasing the chances of retrieving paralogue derivatives of the original curated miRNA. Of these 27 miRNAs, 18 miRNAs were matched in miRBase with a cumulative 44 regular expressions. Each of the 18 miRNAs matched using this secondary verification method yielded at least two resultant matches ('original miRNA' + '-3p' or '-5p'). In addition, the original curated miRNA 'hsa-miR-181', which did not match as an exact string match in miRBase, revealed itself as a highly paralogous miRNA. This miRNA had seven known paralogues in miRBase, of which 4 paralogues had a record for both a '-3p' and a '-5p' miRNA sequence. The remaining 3 paralogues had a record for either a '-3p' or a '-5p' miRNA.

The regular expression string testing also identified two paralogues for 'hsa-miR-181' which were identical in sequence, but which arose from a primary- or pre-miRNAs which had a different locus in the genome. Subsequent target identification also revealed that the paralogues of 'hsa-miR-181' cumulatively matched to 380 gene targets in miRTarBase. However, none of the regular expressions for 'hsa-miR-181' were found as targets in the miRecords database. If the regular expression testing step had not been performed, these 380 gene targets for the regular expressions of this single miRNA would not have been included in the downstream analyses of this study, thereby limiting the degree to which biological networks could be generated for SLE.

The regular expression miRNA ID testing also served to report possible nomenclature errors from the source article. This was demonstrated in (Tiffin *et al.*, 2013), where both 'miR-1224-3P' and 'miR-1224-3p' were listed in Supplementary data file S-4, as miRNAs that have been implicated in the progression of lupus. Upon referring back to the source data for those two miRNAs, it was found in (Te *et al.*, 2010) that a miRNA, referred to as 'hsa-miR-1224-3P' was found in the article on nine occasions. There was no occurrence of 'hsa-miR-1224-3p'. It must however be noted that the nomenclature of miRNAs used in this paper does not follow convention of using lowercase p's in miRNAs having a -3p or -5p suffix, nor does this paper follow convention of using lowercase letters, directly following the numerical naming of the miRNA, as can be seen in multiple instances in the paper, e.g. 'hsa-miR-125A-3P'. However, 'hsa-miR-1224-3p' was found as an exact string match miRNA in the 'PS01_June_28_miRBase_1.py' Python script. The miRNA 'hsa-miR-1224-3P' was found in the miRTarBase database as 'hsa-miR-1224-3p' with a lowercase –p suffix. Because 5 gene targets (*KMT2D*, *IGSF3*, *TCF12*, *HNRNPA3*, *CDC42EP1*) were found for 'hsa-miR-1224-3p', this miRNA was simply flagged for checking in source literature, and not resubmitted to the miRTarBase database as 'hsa-miR-1224-3p'.

As shown in Table 13, 'hsa-miR-21' was found to have 49 gene targets in the miRecords database. The regular expression matched and verified miRNA 'hsa-miR-21-5p' illustrated the dominance described for the mature miRNA produced from one arm of a precursor hairpin miRNA. Regular expression matching revealed that 'hsa-miR-21-5p' had 439 gene targets in miRTarBase, and Table 14 showed 'hsa-miR-21-3p' having only two targets In miRTarBase. Thus, regular expression testing lead to a vast increase in the number of targets obtained for the dominant miRNA 'hsa-miR'21-5p'. If only the original miRNA, 'hsa-miR-21' had been considered for target identification, almost 400 gene targets would have been missed and not included in the downstream analyses.

Even though the miRNA 'hsa-miR-516a-5p' was not found in miRTarBase (see Table 11) this miRNA was found in miRecords to regulate two genes, *KLK10* and *SULF1* (see Table 12). This further demonstrated that even though miRecords had much fewer MTIs than miRTarBase, the miRecords database still had the potential to have MTIs that miRTarBase did not have. Therefore, considering multiple databases for obtaining miRNA-Target interactions has proven to be a valuable step prior to performing biological analyses.

With the knowledge that miRNAs are able to regulate numerous genes, it may be hypothesised that the more interactions reported by a database increases the chances of obtaining a more realistic view of the regulatory capabilities of a given miRNA. This hypothesis was substantiated by the target identification for the 13 miRNAs which matched perfectly to entries in the miRBase database. Of these 13 miRNAs, 11 were found in miRTarBase, yielding a total of 607 miRNA-target interactions. In contrast, only three out of the 13 verified miRNAs were found to have targets reported in miRecords, and these three miRNAs cumulatively had five target interactions. Target identification in miRTarBase and miRecords revealed that miRNAs targeted one or multiple genes, e.g. 'hsa-miR-342' and 'hsa-miR-17-5p', respectively.

The cross-analysis step in this bioinformatics workflow aimed to identify which genes and miRNAs from the curated datasets were in the output lists of interacting molecules retrieved from the miRTarBase and miRecords databases. The different sets of miRNAs which resulted from the verification step in miRBase, were cross-analysed with the outputs of the miRNA target identification step in miRTarBase and miRecords for the SLE-associated-, and differentially-expressed genes.

For the cross-analysis of miRNAs, three datasets were selected: the raw curated list of miRNAs (n = 40), and the miRNAs which matched perfectly as a string record in miRBase (n = 13) pooled with the 44 miRNAs which were a result of regular expression verification in miRBase. The third dataset was the miRNAs which failed the verification step as a perfect string match in miRBase (n = 27). The data which matched as a perfect string match, and those which failed the perfect string match verification step in miRBase, are subsets of the curated dataset of miRNAs (n = 40). However, each of these datasets was subjected to cross-analysis in order to draw meaningful conclusions from the subsets of data, instead of only drawing a conclusion from the results of the starting dataset.

Before performing the cross-analysis, the miRNAs which matched as a perfect string match (n = 13) in miRBase were pooled along with the miRNAs which were verified as a regular expression (n = 44). The pooled list (n = 57) was then subjected to the set() function in Python to create a unique list of miRNA names. As a result, the list of 57 miRNAs was reduced to 56 miRNAs. This observation lead to a further investigation into these two datasets because a perfect string match (first verification step) should never appear as a duplicate entry in the result of the regular expression match (second chance at verification), as these are sequential steps, i.e. these steps do not take place in parallel. Thus, upon investigation I found that both 'hsa-miR-142', and 'hsa-miR142-3p' were included as individual miRNAs curated from two separate literature sources, Zan *et al.* 2014 and Tiffin *et al.* 2013, respectively. Thus, when each of these two miRNA

IDs was tested for being a perfect string match in the miRBase database, 'hsa-miR-142-3p' was found to exist as such, whereas 'hsa-miR-142' did not. However, when the second verification step (regular expression) was employed on those miRNAs which did not match as an exact string match, 'hsa-miR-142' was then found in miRBase as two regular expression records, i.e. hsa-miR-142-5p' and 'hsa-miR-142-3p'. Thus, when pooling these two subsets of data verified in miRBase, a duplication of 'hsa-miR-142-3p' occurred, resulting in the need to create a unique list of miRNAs, which ordinarily should not be needed for these two distinct datasets.

In the first cross-analysis test, the starting miRNAs (n = 40) were assessed against the interacting miRNAs (n = 291). Of these 40 miRNAs directly curated from literature, 13 miRNAs were identified in the list of 291 miRNAs found to interact with the SLE-associated genes curated for this study (from literature, and from Dr Kröger). Thus, 27 of these starting miRNAs, which have been described as associated with lupus pathogenesis, were not identified in the list of 291 miRNAs which were found to interact with the 813 unique genes which are known to be associated with SLE.

Test 2 of the cross-analysis of miRNAs involved testing the pooled dataset consisting of a unique list (n = 56) of the miRNAs which had passed either the first or second verification step in miRBase, against the interacting miRNAs (n = 291). Of these 56 miRNAs, 22 were found in the list of interacting miRNAs. Thus, 34 miRNAs, which were verified in miRBase as either an exact string match or a regular expression match for the original curated lupus-associated miRNA, were not identified in the list of 291 miRNAs found to interact with the known lupus-associated genes dataset.

Lastly, Test 3 assessed the 27 miRNAs which failed verification in miRBase against the 291 miRNAs found to interact with the known lupus-associated genes. Seven out of the 27 miRNAs were identified in the list of 291 interacting miRNAs. However, 20 of these miRNAs, which were reported in literature to be associated

with SLE pathogenesis but were not verified as either a perfect string match or regular expression string match in miRBase, were not found in the list of 291 miRNAs which were demonstrated to interact with the known lupus-associated genes list (n = 813).

As previously described, the perfect string match sub-dataset (n = 13) and the miRNAs which failed verification in miRBase (n = 27) are both subsets of the list of starting miRNAs (n = 40) assessed in Test 1 of the cross-analysis. Thus, excluding the miRNAs for these datasets which were not found in the list of interacting miRNAs, it can be concluded that 34 miRNAs (Test 2), although curated from recent literature as being associated with SLE pathogenesis, were not identified in the list of 291 miRNAs found to interact with the curated SLE-associated genes (n = 818). Thus, further investigations need to be made into assessing the miRNA-target interactions, and the biological relevance of the regulatory activities of these 34 miRNAs.

12.3. Lupus-associated genes

A total of 830 non-unique gene identifiers were originally collected for this study. A total of 81 genes (in associated gene name format) were curated from three recently published peer-reviewed biomedical research articles. This list comprised of one gene curated from a primary source article, and 80 unique gene names curated from two review articles. The remaining 746 genes (in Ensembl Gene ID format) were included in this study with the permission of Dr Wendy Kröger. These 749 genes were reported as being differentially-expressed during the observations of lupus pathogenesis.

With the preparation of the associated gene names, as described in the Methods chapter, the starting list of 81 genes was adjusted to a new list of 85 unique genes, which had been split from their aliases and associated pairs. Once again, in adherence to the standardisation of Ensembl Gene ID notation, no manual alterations were made to these IDs collected during Dr Kröger's meta-analysis of differentially-expressed lupus genes.

Both gene lists were individually submitted to the Ensembl BioMart tool in order to identify which of the genes submitted were matched as HGNC symbols in the database. The verification of the SLE-associated gene names as HGNC symbols revealed that 82 out of 85 genes and three out of 85 genes were found, and not found respectively in the BioMart database. The three genes which were not found were *CREBL1*, *Lyn*, and *IRF53*. This may be in concordance with the fact that GeneCards reported *CREBL1* to be an alias for *ATF6B*, and *ATF6B* was verified in BioMart to exist. The curated gene "*Lyn*" was not found in BioMart as is. A gene named "*LYN*" was found in BioMart, but because it did not match the font case of "*Lyn*", when searched for against the original input list of starting miRNAs, '*Lyn'* was reported as an error as one of the three genes not found in BioMart. Because '*Lyn'* was curated as such from the source, no adjustments were made to uppercase to suit the findings in BioMart.

As described in 11.4.2.2, 'RABGAP1L' was initially incorrectly reported in the source article as '*RABGAPIL'*. Upon manual referral back to the original source article, the correct spelling of the gene was confirmed. '*RABGAP1L'* was then confirmed as a gene in the Ensembl BioMart database, and subsequent target identification revealed a single miRNA interaction with 'hsa-miR-1'. The manual checking of unverified gene names ensured that this gene was included in the analyses, with the correct nomenclature from the source article, and that targets were correctly reported. Thus, the BioMart verification of the Ensembl IDs for the differentially-expressed genes as a step prior to target identification, allowed for the accountability of all the curated Ensembl IDs with the conversion to HGNC symbols.

The verification of the 749 Ensembl Gene IDs as HGNC symbols in BioMart also proved helpful as it identified 13 genes, which upon manual checking were found to be deprecated IDs. If this verification step had not been included as part of the workflow, and if target identification had been performed directly by simply converting all the ENSG IDs to their respective gene names using a tool such as

BioBDNet (http://biodbnet.abcc.ncifcrf.gov/db/db2db.php), these 13 deprecated genes would have either been incorrectly reported as "Not Found" in the miRTarBase or miRecords databases, or it may have been reported as "Found", but may not have been a true gene name. Thus, using an updated version of Ensembl BioMart to verify the existence of these ENSG symbols as records in the database ensured that each of the gene IDs were correctly converted to HGNC symbols, and that the genes which had been removed from the database were reported as such. Thus, ensuring that target identification was only performed on the genes which had been verified to exist, and excluding those which were no longer recognised.

Of the 85 genes curated from literature, 82 genes were verified in BioMart to exist, and three genes (*CREBL1*, *IRF53*, and *Lyn*). Of the 82 genes verified in BioMart, 47 genes were found to have interactions with miRNAs in the miRTarBase database, whereas only 11 of these 82 genes were found in the miRecords database. In contrast, the three genes which were not verified in BioMart, were also not found as records in either miRTarBase or miRecords. As in the verification step of the miRNAs, this verification of the genes in BioMart proved useful in identifying why these genes were not found in the target interactions databases; whether it was because they were not records in the databases (without verification, and submitting the complete list of 85 genes to the target databases), or more specifically, whether they were not confirmed as actual genes, and as a subset were submitted to the target databases. As seen in this study, the inclusion of the gene name verification in BioMart proved its worth as described by the latter scenario.

For the cross analysis of the input genes with the output genes found to interact with the known SLE-associated miRNAs curated from literature, two tests were designed and performed. Test 1 consisted of testing the pooled SLE-associated genes and differentially-expressed genes which were verified to exist in BioMart ($Total_n$ = 813), against the 2 702 genes found to interact with the known SLE-

associated miRNAs. It was found that 127 out of 813 known lupus-associated genes were identified in the list of 2702 genes found to interact with known lupus-associated miRNAs. Thus, of the 2702 unique genes found to interact with the known lupus miRNAs, 686 genes were not found in the starting list of curated genes. Therefore, it is suggested that the 686 genes which were found to interact with the known SLE-miRNAs, but were not identified in the starting list of genes, be further researched to identify the biological networks influenced by the activities of these 686 possibly dysregulated genes.

Test 1 also revealed that a number of genes identified as targets for a given miRNA in miRTarBase or miRecords, were in fact not present in the starting list of curated genes. An example of this is where 'hsa-miR-17-5p' was found to have 310 unique gene targets in miRTarBase, but in the cross-analysis, only 19 genes out of those 310 identified were in the list of curated genes. Because a multi-systemic autoimmune disease is likely to be affected by the dysregulation of an abundance of genes, it is interesting that only 19 known lupus-associated genes out of the 310 targets identified for this miRNA was observed. This was again seen in the analysis of 'hsa-miR-423-5p' where this miRNA was shown in miRTarBase to target 168 genes, but the cross-analysis showed that only 11 of these 168 where known-lupus associated genes. These findings suggest that there are possibly an abundance of genes regulated by the same miRNAs which may be involved in biological processes associated with lupus pathogenesis.

Test 2 of the cross-analysis aimed to identify whether the three genes not verified in BioMart (*CREBL1*, *IRF53*, and *Lyn*) were present in the list of 2 702 genes found to interact with the known SLE-miRNAs. It was subsequently found that none of these three genes were in the list of 2 702 genes.

12.4. Pathway Analysis

Of the 813 unique genes evaluated in this study for MTIs, 736 genes were reported by the study performed by Dr Kröger, to be involved in biological pathways including that of interferon signaling, TLR signaling, diapedesis, oxidative phosphorylation, and adhesion regulatory networks. From this study, six datasets were created, all derived from (and including) the dataset containing all the genes and miRNAs which were curated, and those obtained from the determination of targets in miRTarBase and miRecords. The Core Analysis was run on the dataset file containing the top regulated genes and miRNAs, and their associated targets. These 108 molecules, determined by IPA as suitable for analysis, contained a mixture of genes and miRNAs which would form the skeleton of the overall network.

For these top regulators and their associated targets, the Core Analysis revealed that majority of these molecules were implicated in one of five top canonical pathways. These top canonical pathways were Interferon signaling, TLR signaling (both in concordance with the findings of Dr Kröger), as well as three new pathways, namely iNOS signaling, *IL11* signaling, and the role of *PKR* in Interferon induction and antiviral response.

Numerous reviews have evaluated the role of nitric oxide (NO) in the onset of autoimmunity (Nagy *et al.*, 2010; Oates and Gilkeson, 2006; Oates *et al.*, 2010). It has been well-established that NO production in response to pathogen invasion assists with controlling T cell proliferation. However, when NO production shifts to an uncontrolled manner, there is a break down in immune self-tolerance as result of T cell dysfunction (Nagy *et al.*, 2010). NO may, depending on its circulating concentration, either inhibit or promote the apoptotic pathway (Nagy *et al.*, 2010). Therefore, with the knowledge that impaired clearance of dead cells by means of apoptosis is a hallmark of lupus onset, it is likely that NO production through iNOS signaling plays an important role in the progression of lupus.

Another top canonical pathway listed was that of Interleukin-1 (IL-1) signaling. IL-1 is a potent inducer of acute inflammation – a phenotype defining lupus-flare (Feghali and Wright, 1997). When involved in the NF-κB pathway, IRAK1 (IL-1 receptor associated kinase 1) has also been implicated in the activities of interferon production. It has been proposed that an overproduction of IFN may also contribute to the pathogenesis of SLE (Ceribelli *et al.*, 2011). TLR signalling has also been described in numerous studies to play an important role in the maintenance of normal immune system functioning, and conversely, in the dysfunctioning of immune receptors during lupus flare. Genetic variations in TLRs have been described for various populations where lupus studies have been conducted, such as the study conducted on a Taiwanese population (Wang *et al.*, 2014). TLRs have been described extensively with regards to their role in the pathogenesis of SLE, as well as their regulation by a number of miRNAs (Deng *et al.*, 2013; Quinn and O'Neill, 2011).

In contrast to the top canonical pathways described for the skeleton dataset, the top diseases and biological functions shown to be associated with this dataset were all largely different from each other. However, due to multi-systemic nature of SLE pathogenesis, the disparity in the pathways described can be justified. Organismal injuries and connective tissue disorders are likely to describe pathways related to the overall damage and destruction inflicted by auto-antigens produced during the abnormal immune functioning. Endocrine system disorders describes the abnormalities of immune cell production and circulation, an observation characteristic of lupus disease state.

Lastly, the implication of cancer as a disease describes the overall malfunction of cell cycle regulation, and aptly describes the level to which the immune system can become dysregulated during lupus flare. The heat map in Figure 28 once again illustrated the degree to which the pathways implicated differed. Organismal injury and cancer pathways were shown to be predominantly

affected by the molecules present in the Core Analysis dataset, followed by cell cycle, and lastly, varieties of biological pathways.

In coherence with the findings described above, Table 28 described that of the dataset of top regulated genes and miRNAs, and their associations, a number of molecules were identified to be associated with numerous immunological disease categories, including 'systemic autoimmune syndrome' in which 23 molecules out of the dataset were implicated.

12.5. Identification of novel, candidate SLE-associated genes from pathway analysis

12.5.1. Genes not in the starting list of genes, but were found to interact with known lupus-associated miRNAs

Across all three Core Networks, eighteen genes were found to be potentially novel SLE-associated candidate genes. These genes were not in the list of curated genes from recent biomedical literature, or Dr Kröger's meta-analysis of lupus-associated genes. A manual search in the data file containing all the curated genes and miRNAs revealed that these 18 genes were in fact not found in the datasets. Each of the genes were searched for in BioMart (September 2015 version) to identify which phenotypes and biological processes, if any, have already been described and associated with the gene. Using the search terms "SLE", "lupus", and "systemic lupus erythematosus" on the Excel file exported from BioMart, it was found that none of these genes had been associated with the lupus phenotype. Thus, dysregulation of these genes by miRNAs – as determined by the initial target identification in miRTarBase and miRecords, may suggest a dysregulation in biological processes described by BioMart. For reporting purposes, I have grouped the results of the phenotype descriptions into three categories, namely Cell cycle processes, Cancers, and Neurological diseases.

Of the 18 genes which were not in the list of 818 curated SLE-associated genes, but were found to interact with known lupus-associated miRNAs, 10 genes were reported by BioMart to be associated with cell cycle processes. Some of these genes were also reported to specifically be involved in the positive regulation of transcription from the RNA polymerase II promoter. These genes were: *AGO2*, *CSNK1A1*, *HIF1A*, *NUCKS1*, *KLF3*, *NRF4A1*, *AIMP1*, *KARS*, *TLR4*, and *STAT3* (which was also implicated in an infantile onset form of multi-systemic autoimmune disease (OMIM: #615952)).

Of the 18 genes, five genes were described to be associated with some form of cancer. *CDKN2A* was reported to be associated with melanomas (OMIM: #606719), *BCL6* with lymphomas (OMIM: *109565), *MYC* with Burkett's lymphoma (OMIM: #113970), and *HNRNPH1* with leukaemia (OMIM: none available). Lastly, *TP53* was found to be associated with a variety of cancers (OMIM: #614740), including hereditary breast and ovarian cancer syndrome, oesophageal cancer, lung cancer, and basal cell carcinoma.

Three genes were reported to be associated with neurological diseases. *PSEN1* was reported to be associated numerous forms of neurological disorders such as semantic dementia (OMIM: 600274), progressive non-fluent aphasia (OMIM: 600274), and Alzheimer disease 3 (OMIM: 607822). *HNRNPU* was associated with epileptic encephalopathy (OMIM: none available), and *SMARCA4* was associated with mental retardation autosomal dominant 16 (OMIM: 614609).

12.5.2. Genes which were added by IPA to complete the Core Networks

Across the three Core Networks, a total of 44 genes, and three miRNAs were found to be added in by IPA to complete these networks. The genes were once again searched for as HGNC symbols in Ensembl BioMart (September 2015 version), and a search in the exported Excel file revealed that the search phrases "systemic", "autoimmune", "lupus", and "SLE" were not present in this data file for the IPA-added molecules. Of the three miRNAs added by IPA, only hsa-miR-298 was found in the BioMart search. The miRNA was only reported to be found

on Chromosome 20, and no phenotypic information or PubMed source was available for this miRNA. The other two miRNAs were not found in this BioMart search.

Of the 44 genes added in by IPA, 26 genes were identified as nodal genes, interacting with numerous other SLE-associated genes, and 18 genes were identified as 'connector' genes. Of the 26 nodal genes, two genes were not found in BioMart, namely *estrogen receptor*, and *Histone h4*. Thirteen genes were not found to have any phenotypic description in BioMart. These 13 genes were *NUPR1*, *YWHAQ*, *ZMIZ2*, *HSPH1*, *SERINC3*, *HSF1*, *RBBP4*, *TFF1*, *IL1R1*, *KHDRBS1*, *KHDRBS2*, *TADA2A*, and *TADA2B*.

The remaining 11 genes were found in BioMart and associated with at last one known phenotype or biological process. *ESR1* was described to be involved in estrogen resistance syndrome (Green *et al.*, 1986). In the network overlay analyses, *ESR1* was shown to interact with a number of IPA-added genes, as well as with three known lupus-associated genes (*CCND1*, *MMP1*, and *NRIP1*), and a number of genes found to interact with known lupus-associated miRNAs. Because of the strong prevalence of lupus in females compared to males, it is to be expected that various forms of the estrogen hormone may play a critical role in the pathogenesis of lupus. However, the *ESR1* gene is yet to be reported in BioMart as being associated with SLE. *MED1* was described in this BioMart search as being involved in numerous biological processes including the process of androgen biosynthesis, chromatin binding, cell morphogenesis, and the differentiation of keratinocytes. *MED1* was also described in Cui *et al.* 2012 to interact closely with *HER2* in the regulation of Tamoxifen-resistant breast cancer cells in humans. *STX3* was described as being involved in the positive regulation of numerous biological processes including that of cell adhesion, cell proliferation, chemotaxis, and protein localization (Darios and Davletov, 2006).

ATP13A2 was reported to be associated with the Parkinson disease 9, and Kufor-Rakeb Syndrome phenotypes (OMIM: #606693). The manifestation of Parkinson

disease 9 was described here to be as a result of *ATP13A2* deficiency. *EXOSC8* was reported to be involved in another neurological disorder, pontocerebellar hypoplasia (OMIM: #616081). *TFAP2A* was described to participate in numerous biological processes (OMIM: #113620), including kidney development, hard palate development, and mRNA export from the nucleus. Dysregulation in *TFAP2A* may thus result in the pathologies of the kidneys observed as a secondary disease in lupus patients. *EDARADD* was reported by this BioMart search to be associated with the phenotype of autosomal dominant- and recessive forms of a skin disorder named hypohidrotic ectodermal dysplasia (OMIM: 614900). *IRF4* the interferon regulatory factor 4 gene was described to be associated with the pathology of myelomas, including the multiple systemic amyloidosis (OMIM: 254500).

HDAC9 was described to be associated with a phenomenon known as Peter's Anomaly (OMIM: none available), as well as with the positive regulation of the cell migration in sprouting angiogenesis, and *RPRD1B* was described to be involved in the dephosphorylation of RNA polymerase II at the C-terminal domain. *GRB2* was described to be associated with cellular response to ionizing radiation, positive regulation of reactive oxygen species metabolic processes, the negative regulation of receptor internalization, and signal transduction in response to DNA damage *PELI3* the retinoblastoma binding protein 4 gene, was described to be associated with the negative regulation of the extrinsic apoptotic signalling pathway, as well as the negative regulation of TNF-mediated signalling pathway.

Across the three networks, a total of 18 genes were selected as 'connector' genes. Of these 18, 15 genes were found in BioMart, having no known phenotype description. These 15 genes were *SLC39A8*, *TNFAIP8L1*, *SLCA12*, *SESN2*, *GTSE1*, *GAK*, *MAPK12*, *TMEM97*, *RBM39*, *BTG3*, *ASF1B*, *AHSP*, *ATP6V01*, *FOXD1*, and *P4HA2*. A search in Google Scholar for *GTSE1*, and *GAK* with the phrase "gene in systemic lupus erythematosus" also revealed no published

literature on the associations of these two genes to lupus. Only three of the connector genes had phenotypes described for them in BioMart. These three genes were *STX3*, *UNC5B*, and *INS*. *INS* was reported by BioMart to have a known phenotypic description. *INS*, the gene encoding for insulin was described as being involved in numerous metabolic regulatory processes. *STX3* was described to be involved in cell cycle processes and *UNC5B* was described to be a negative regulator of apoptosis.

12.6. Bioinformatics pipeline review

Linkage and association studies have been frequently used to identify rare variants exhibiting high penetrance in common disorders. However, the identification of common variants with decreased penetrance has proven problematic using these aforementioned methods (Armstrong *et al.*, 2008). In a disease with complex aetiology such as SLE, many researchers have resorted to genome–wide approaches for elucidating the SNPs which may be associated with a disease phenotype.

A disadvantage of GWAS methods is the high cost associated with the requirement to scan the entire genome in order to perform a comparison against a reference, in order to identify the SNPs (Armstrong *et al.*, 2008). In order to reduce this level of expense, Armstrong *et al.* developed a software programme, Function2Gene, which integrates the searching of biomedical literature for SNPs associated with the query disease, followed by prioritization tactics defined by the user, and performed by the programme (Armstrong *et al.*, 2008). Another disadvantage of GWAS studies is the high cost associated with genotyping a large number of samples in order to have adequate power required for statistical analyses, and for obtaining results with statistical significance (Hunter and Kraft, 2007).

A recent study by Song *et al.* described a bioinformatics pipeline for identifying miRNA-regulated RA-associated genes (Song *et al.*, 2015). This study involved performing conditional literature search in PubMed for RA-associated genes

target by miRNAs, and downloading microarray and RNA-seq datasets from NCBI's GEO database. miRNA target genes were predicted using TargetScan and miRDB. Once miRNA target genes were predicted and RA-associated miRNAs were identified, differentially-expressed RA genes were derived. Subsequently, functional analysis was performed on the differentially-expressed RA genes, and GO molecular functions were reported (Song *et al.*, 2015).

In contrast to the study by Armstrong *et al.*, the pipeline designed and employed within this dissertation aimed to identify novel miRNAs and genes which were potentially associated with SLE, and did not include the analysis of SNP data. Similar to the study by Song *et al.*, I performed an extensive literature search, and included a second dataset of differentially-expressed SLE genes, as provided by Dr Wendy Kröger. In contrast to the study by Song *et al.*, I used the latest versions of miRTarBase and miRecords databases to identify miRNA-target interactions for my source data. There is also no mention by Song *et al.* that verification of miRNA IDs was performed, as I had completed as the initial verification step prior to target identification.

The study by Song *et al.* also reported on significantly enriched biological pathways using GO terms, whereas I reported the diseases and functions of the network datasets using the results generated from pathway analysis. Lastly, unlike either of these two studies, I included a comprehensive biological pathway analysis of my interaction-derived data using Ingenuity Pathway Analysis software. This pathway analysis generated graphical interaction networks for the input data, a heat map of the canonical pathways impacted, and subsequently allowed for the identification of potentially novel SLE-associated genes and miRNAs.

12.7. Results in comparison to other SLE gene target identification studies

This bioinformatics-based approach to identifying novel SLE-associated genes and miRNAs reported five top canonical pathways affected by the prioritized

dataset of top regulated genes and top regulating miRNAs. These five pathways were: Interferon signalling, TLR signalling, iNOS signalling, IL11 signalling, and the role of PKR in interferon induction and antiviral response. In agreement with these observations, it was also reported that TLR-signalling and Type-1 Interferon signalling are two of the canonical pathways containing SLE susceptibility loci (Harley *et al.*, 2009). In contrast, Harley *et al*. also reported "immune complex processing" and "immune signal transduction" to be critical pathways involved in SLE progression, whereas single, specific components of the immune system were reported in this thesis to participate in SLE processes, and not at the more general functional pathway level.

A total of 18 genes which were not present in the curated list were found in miRTarBase and/or miRecords to interact with known SLE-associated miRNAs. These were: *AGO2*, *CSNK1A1*, *HIF1A*, *NUCKS1*, *KLF3*, *NRF4A1*, *AIMP1*, *KARS*, *TLR4*, *STAT3*, *CDKN2*, *BCL6*, *MYC*, *HNRNPH1*, *TP53*, *PSEN*, *HNRNPU*, and *SMARCA4*. From the pathway analysis in IPA, a total of 44 genes, and three miRNAs were added by the IPA database to complete the core networks generated. These genes and miRNAs were not present in the curated lists, or in the lists of interactions. These 62 genes and three miRNAs are considered potentially novel SLE-associated genes which might be investigated in future experimental studies as potential biomarkers for the disease.

As previously described, GWAS studies have dominated the bioinformatics-based approaches to identifying novel disease-associated genes. One such study identified *AFF1* as a candidate SLE-disease gene in a Japanese population (Okada *et al.*, 2012). This gene was however, not identified in my analysis as a gene to be interacting with a known SLE-associated miRNA, nor reported as an SLE-associated, or differentially-expressed SLE gene. This clearly demonstrates that even with the extensive database knowledge used in Ingenuity Pathway Analysis software, as well as the integration of multiple data sources, the knowledge of

SLE-associated genes continues to be elucidated as more experiments are performed and reported on.

A number of genes have been identified from GWAS studies as being candidate SLE-associated genes. These include *TREX1*, *C1Q*, *TNFAIP3*, *IRF5*, and many others (Harley *et al.*, 2009). A number of these genes were also reported in the Results chapter of this dissertation as interacting with a known SLE-associated miRNA, thus adding an additional line of validation for the current study.

13. CONCLUSION

To conclude, each of the IPA-added molecules described by the BioMart search could be placed in the three categories described for the genes which were found to interact with known lupus-associated genes. The three categories were cell cycle regulation, cancers, and neurological diseases. These IPA-added genes, in addition to the interacting genes which have been described to play significant roles in these three biological processes and pathologies, were through network analysis also shown to interact extensively with members of the curated genes and miRNA data set. Experimental validation is known to be labour-intensive, time-consuming, and costly. Thus, the *in silico* methods described in this dissertation allude to the potential benefits to performing these analyses for candidate molecules prior to experimentation, by improving the chances of investigating the true aetiological gene in the laboratory for validation

13.1. Limitations of this study

A few limitations can be described which may be addressed in order to improve the workflows developed and utilised in this bioinformatics-based study. First, the number of experimentally-validated miRNA targets reported in databases such as miRTarBase and miRecords used in this study, are limited by the current state of knowledge about miRNAs and their targets. These targets are also largely limited by the molecular experiments employed to validate miRNA functionality, but will improve over time as new techniques are developed and included in these experimental validations of miRNA targets. Thus, the bioinformatics methodologies proposed in this study should be reused in future, as more updated versions of these target-prediction databases are released.

Second, the validation of interactions between miRNAs and the target genes are still not extensive, but this will improve with knowledge improving over time. Lastly, because of the complex aetiology associated with lupus, it is believed that the interplay between genetic susceptibility and environmental factors plays a critical role in the onset and progression of this disease. Thus, it is proposed that

at best, one can aim to identify general susceptibility genes and miRNAs associated with lupus, but that how these complex interactions play out is more difficult to determine.

The approaches used in this study contribute at a level of mechanisms driving lupus, and gene networks with key nodes, rather than looking for specific causative genes. Thus, it may be more beneficial to use the data to identify general cellular mechanisms which may be relevant to lupus, as shown in this study, rather than attempting to isolate a few specific genes that drive the pathogenesis of the SLE.

13.2. Future directions

Further studies span across both biological experiments and bioinformatics analyses. For biological experiments, the nodal genes may be investigated in expression studies to determine whether they are in fact linked to the onset of SLE, and the progression of the disease. In terms of bioinformatics, it is proposed that the methodologies developed in this study be integrated into an automated database query system which may be freely accessible for researchers interested in investigating biological networks in this context.

With the permission of the respective database or software creators, the automated system could integrate the freely-accessible and downloadable data from the miRBase, Ensembl BioMart, miRTarBase, and miRecords databases, as described in the Methods chapter of this study. The query system would also implement the automatic update of database versions, thereby ensuring that miRNA ID verification, and gene name or ID verification and conversion are performed using the latest versions of the databases. The automated update would also allow users to perform their analyses preferably with the updated miRTarBase, and miRecords databases, or with archived versions of these two databases. The Python scripts written for the Methods chapter of this study would be integrated into this query system to ease the identification of miRNA, or gene targets for the input list of genes, or miRNAs, respectively.

Because IPA is a licensed resource, automated integration of this tool would not be possible. However, an open-source tool such as Gene Ontology can be integrated to obtain the annotation of identified molecules according to known biological functions. Furthermore, using Cytoscape open-source software with the CyTargetLinker add-on application, KEGG or Reactome pathway analysis may be performed to visualise gene-miRNA interactions. Another open-source software tool, STRINGdb, may be integrated into the query system to provide protein-protein interaction visualisation for the dataset uploaded by the user.

An important feature of this bioinformatics pipeline is that it integrates the verification of miRNA IDs and gene IDs using well-established and utilised open-source resources. Thus, all subsequent analyses including the target identification and pathway analyses can be trusted to be accurate based on the fact that each ID was verified to exist as an entry in its respective data repository. Lastly, the query system would not be restricted to SLE, as it is purely based on the miRNA-gene interactions derived from the source databases, and thus can prove to be useful in its application to multiple fields of biomedical research.

14. REFERENCES

Aggarwal, R., Namjou, B., Li, S., D'Souza, A., Tsao, B. P., Bruner, B. F., … Scofield, R. H. (2010). Male-only Systemic Lupus. *The Journal of Rheumatology*, *37*(7), 1480–1487. http://doi.org/10.3899/jrheum.090726

Alevizos, I., and Illei, G. G. (2010a). MicroRNAs as biomarkers in rheumatic diseases. *Nature Reviews Rheumatology*, *6*(7), 391–398. http://doi.org/10.1038/nrrheum.2010.81

Alevizos, I., and Illei, G. G. (2010b). MicroRNAs in Sjögren's syndrome as a prototypic autoimmune disease. *Autoimmunity Reviews*, *9*(9), 618–621. http://doi.org/10.1016/j.autrev.2010.05.009

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., … Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA*, *9*(3), 277–279. http://doi.org/10.1261/rna.2183803

Apostolidis, S. A., Lieberman, L. A., Kis-Toth, K., Crispín, J. C., and Tsokos, G. C. (2011). The Dysregulation of Cytokine Networks in Systemic Lupus Erythematosus. *Journal of Interferon and Cytokine Research*, *31*(10), 769–779. http://doi.org/10.1089/jir.2011.0029

Arbuckle, M. R., McClain, M. T., Rubertone, M. V., Scofield, R. H., Dennis, G. J., James, J. A., and Harley, J. B. (2003). Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *New England Journal of Medicine*, *349*(16), 1526–1533. Retrieved from http://www.nejm.org/doi/full/10.1056/nejmoa021933

Armstrong, D. L., Jacob, C. O., and Zidovetzki, R. (2008). Function2Gene: a gene selection tool to increase the power of genetic association studies by utilizing public databases and expert knowledge. *BMC Bioinformatics*, *9*, 311. http://doi.org/10.1186/1471-2105-9-311

Balada, E., Ordi-Ros, J., and Vilardell-Tarrés, M. (2007). DNA methylation and

systemic lupus erythematosus. *Annals of the New York Academy of Sciences*, *1108*, 127–136. http://doi.org/10.1196/annals.1422.015

Bernier, M., Mikaeloff, Y., Hudson, M., and Suissa, S. (2009). Combined oral contraceptive use and the risk of systemic lupus erythematosus. *Arthritis and Rheumatism*, *61*(4), 476–481. http://doi.org/10.1002/art.24398

Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Research*, *36*(Database issue), D149–53. http://doi.org/10.1093/nar/gkm995

Bombardier, C., Gladman, D. D., Urowitz, M. B., Caron, D., Chang, C. H., Austin, A., … Schur, P. H. (1992). Derivation of the SLEDAI. A disease activity index for lupus patients. *Arthritis and Rheumatism*, *35*(6), 630–640. http://doi.org/10.1002/art.1780350606

Borchers, A. T., Keen, C. L., and Gershwin, M. E. (2007). Drug-Induced Lupus. *Annals of the New York Academy of Sciences*, (1108), 166–182.

Borchers, A. T., Naguwa, S. M., Shoenfeld, Y., and Gershwin, M. E. (2010). The geoepidemiology of systemic lupus erythematosus. *Autoimmunity Reviews*, *9*(5), A277–A287. http://doi.org/10.1016/j.autrev.2009.12.008

Cameron, J. E., Yin, Q., Fewell, C., Lacey, M., McBride, J., Wang, X., … Flemington, E. K. (2008). Epstein-Barr Virus Latent Membrane Protein 1 Induces Cellular MicroRNA miR-146a, a Modulator of Lymphocyte Signaling Pathways. *Journal of Virology*, *82*(4), 1946–1958. http://doi.org/10.1128/JVI.02136-07

Carlsen, A. L., Schetter, A. J., Nielsen, C. T., Lood, C., Knudsen, S., Voss, A., … Heegaard, N. H. H. (2013). Circulating MicroRNA Expression Profiles Associated With Systemic Lupus Erythematosus. *Arthritis and Rheumatism*, *65*(5), 1324–1334. http://doi.org/10.1002/art.37890

Ceribelli, A., Yao, B., Dominguez-Gutierrez, P. R., Nahid, M. A., Satoh, M., and Chan, E. K. (2011). MicroRNAs in systemic rheumatic diseases. *Arthritis Res*

Ther, *13*(4), 229. Retrieved from http://www.biomedcentral.com/content/pdf/ar3377.pdf

Choi, J., Kim, S. T., and Craft, J. (2012). The pathogenesis of systemic lupus erythematosus—an update. *Current Opinion in Immunology*, *24*(6), 651–657. http://doi.org/10.1016/j.coi.2012.10.004

Clark, I., Al-Yaman, F., Cowden, W., and Rockett, K. (1996). Does malarial tolerance, through nitric oxide, explain the low incidence of autoimmune disease in tropical Africa? *The Lancet*, *348*(9040), 1492–1494. http://doi.org/10.1016/S0140-6736(96)07342-4

Cui, J., Germer, K., Wu, T., Wang, J., Luo, J., Wang, S., … Zhang, X. (2012). Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Research*, *72*(21), 5625–34. http://doi.org/10.1158/0008-5472.CAN-12-1305

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., … Flicek, P. (2014). Ensembl 2015. *Nucleic Acids Research*, *43*(Database issue), D662–9. http://doi.org/10.1093/nar/gku1010

Dai, R., and Ahmed, S. A. (2011). MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases. *Translational Research*, *157*(4), 163–179. http://doi.org/10.1016/j.trsl.2011.01.007

Dai, R., Phillips, R. A., Zhang, Y., Khan, D., Crasta, O., and Ahmed, S. A. (2008). Suppression of LPS-induced Interferon- and nitric oxide in splenic lymphocytes by select estrogen-regulated microRNAs: a novel mechanism of immune modulation. *Blood*, *112*(12), 4591–4597. http://doi.org/10.1182/blood-2008-04-152488

Dai, Y., Huang, Y.-S., Tang, M., Lv, T.-Y., Hu, C.-X., Tan, Y.-H., … Yin, Y.-B. (2007). Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients. *Lupus*, *16*(12), 939–946.

http://doi.org/10.1177/0961203307084158

Danchenko, N., Satia, J., and Anthony, M. (2006). Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus*, *15*(5), 308–318. http://doi.org/10.1191/0961203306lu2305xx

Darios, F., and Davletov, B. (2006). Omega-3 and omega-6 fatty acids stimulate cell membrane expansion by acting on syntaxin 3. *Nature*, *440*(7085), 813–7. http://doi.org/10.1038/nature04598

De Azevedo Silva, J., Addobbati, C., Sandrin-Garcia, P., and Crovella, S. (2014). Systemic Lupus Erythematosus: Old and New Susceptibility Genes versus Clinical Manifestations. Current Genomics.

de França, N. R., Mesquita, D., Lima, A. B., Pucci, F. V. C., Andrade, L. E. C., and Silva, N. P. (2010). RNA interference: A new alternative for rheumatic diseases therapy. *Bras J Rheumatol*, *50*(6), 702–709.

Deng, Y., Zhao, J., Sakurai, D., Kaufman, K. M., Edberg, J. C., Kimberly, R. P., … Tsao, B. P. (2013). MicroRNA-3148 Modulates Allelic Expression of Toll-Like Receptor 7 Variant Associated with Systemic Lupus Erythematosus. *PLoS Genetics*, *9*(2), e1003336. http://doi.org/10.1371/journal.pgen.1003336

Dolcetti, R., and Masucci, M. G. (2003). Epstein-Barr virus: Induction and control of cell transformation. *Journal of Cellular Physiology*, *196*(2), 207–218. http://doi.org/10.1002/jcp.10263

Dranoff, G. (2004). Cytokines in cancer pathogenesis and cancer therapy. *Nature Reviews Cancer*, *4*(1), 11–22. http://doi.org/10.1038/nrc1252

Ehrenreich, I. M., and Purugganan, M. D. (2008). Sequence variation of MicroRNAs and their binding sites in Arabidopsis. *Plant Physiology*, *146*(4), 1974–1982. http://doi.org/10.1104/pp.108.116582

Elmore, S. (2007). Apoptosis: a review of programmed cell death. *Toxicologic*

Pathology, *35*(4), 495–516. Retrieved from http://tpx.sagepub.com/content/35/4/495.short

Fairweather, D. (2007). *Encyclopedia of Life Sciences*. (L. John Wiley and Sons, Ed.). Chichester, UK: John Wiley and Sons, Ltd. http://doi.org/10.1002/9780470015902.a0020193

Feghali, C. A., and Wright, T. M. (1997). Cytokines in acute and chronic inflammation. *Front Biosci*, *2*(1), d12–26. Retrieved from http://immuneweb.xxmu.edu.cn/ckine-infec.pdf

Fessel, W. J. (1974). Systemic lupus erythematosus in the community: incidence, prevalence, outcome, and first symptoms; the high prevalence in black women. *Arch Intern Med*, *134*, 1027–1035.

Francis, L., and Perl, A. (2010). Infection in systemic lupus erythematosus: friend or foe? *International Journal of Clinical Rheumatology*, *5*(1), 59–74. http://doi.org/10.2217/ijr.09.72

Gilkeson, G., James, J., Kamen, D., Knackstedt, T., Maggi, D., Meyer, A., and Ruth, N. (2011). The United States to Africa lupus prevalence gradient revisited. *Lupus*, *20*(10), 1095–1103. http://doi.org/10.1177/0961203311404915

Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J. M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature*, *320*(6058), 134–9. http://doi.org/10.1038/320134a0

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, *34*(90001), D140–D144. http://doi.org/10.1093/nar/gkj112

Harley, I., Kaufman, K. M., Langefeld, C. D., Harley, J. B., and Kelly, J. A. (2009). Genetic susceptibility to SLE: new insights from fine mapping and genome-

wide association studies. *Nature Reviews Genetics*, *10*(5), 285–290. http://doi.org/10.1038/nrg2571

Hochreiter-Hufford, A., and Ravichandran, K. S. (2013). Clearing the Dead: Apoptotic Cell Sensing, Recognition, Engulfment, and Digestion. *Cold Spring Harbor Perspectives in Biology*, *5*(1), a008748–a008748. http://doi.org/10.1101/cshperspect.a008748

Hsu, S.-D., Tseng, Y.-T., Shrestha, S., Lin, Y.-L., Khaleel, A., Chou, C.-H., … Huang, H.-D. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research*, *42*(Database issue), D78–85. http://doi.org/10.1093/nar/gkt1266

Hughes, T., and Sawalha, A. H. (2011). The role of epigenetic variation in the pathogenesis of systemic lupus erythematosus. *Arthritis Res Ther*, *13*(245), 1–11. Retrieved from http://www.biomedcentral.com/content/pdf/ar3484.pdf

Hunter, D., and Kraft, P. (2007). Drinking from the fire hose--statistical issues in genomewide association studies. *The New England Journal of Medicine*, *357*, 436–439. http://doi.org/10.1056/NEJMp078120

Jeffries, M. A., and Sawalha, A. H. (2011). Epigenetics in systemic lupus erythematosus: leading the way for specific therapeutic agents. *International Journal of Clinical Rheumatology*, *6*(4), 423–438. http://doi.org/10.2217/ijr.11.32

Junker, A., Krumbholz, M., Eisele, S., Mohan, H., Augstein, F., Bittner, R., … Meinl, E. (2009). MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain*, *132*(12), 3342–3352. http://doi.org/10.1093/brain/awp300

Kohli, R. M., and Zhang, Y. (2013). TET enzymes , TDG and the dynamics of DNA demethylation. *Nature*, *502*, 472–479. http://doi.org/10.1038/nature12750

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, *39*(Database), D152–D157. http://doi.org/10.1093/nar/gkq1027

Kyttaris, V. C. (2010). Systemic Lupus Erythematosus: From Genes to Organ Damage. *Methods Mol Biol*, *662*(1), 167–178. http://doi.org/10.1007/978-1-60761-800-3

Leclercq, M., Diallo, A. B., and Blanchette, M. (2013). Computational prediction of the localization of microRNAs within their pre-miRNA. *Nucleic Acids Research*, *41*(15), 7200–7211. http://doi.org/10.1093/nar/gkt466

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75: 843–85*, 843–854. http://doi.org/10.1016/0092-8674(93)90529-Y

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, *23*(20), 4051–4060. Retrieved from http://onlinelibrary.wiley.com/doi/10.1038/sj.emboj.7600385/full

Liu, C.-C., Kao, A. H., Manzi, S., and Ahearn, J. M. (2013). Biomarkers in systemic lupus erythematosus: challenges and prospects for the future. *Therapeutic Advances in Musculoskeletal Disease*, *5*(4), 210–233. Retrieved from http://tab.sagepub.com/content/5/4/210.short

Long, D., Lee, R., Williams, P., Chan, C. Y., Ambros, V., and Ding, Y. (2007). Potent effect of target structure on microRNA function. *Nature Structural and Molecular Biology*, *14*(4), 287–294. http://doi.org/10.1038/nsmb1226

Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., and Bader, G. D. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics (Oxford, England)*, *26*(18), 2347–8. http://doi.org/10.1093/bioinformatics/btq430

Lu, M.-C., Lai, N.-S., Chen, H.-C., Yu, H.-C., Huang, K.-Y., Tung, C.-H., … Yu, C.-L. (2012). Decreased microRNA(miR)-145 and increased miR-224 expression in T cells from patients with systemic lupus erythematosus involved in lupus immunopathogenesis: Altered expression of miRNAs in SLE T cells. *Clinical and Experimental Immunology*, *171*(1), 91–99. http://doi.org/10.1111/j.1365-2249.2012.04676.x

Maidhof, W., and Hilas, O. (2012). Lupus: An overview of the disease and management options. *Pharmacy and Therapeutics*, *37*(4), 240. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351863/

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., … Rigoutsos, I. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*, *126*(6), 1203–1217. http://doi.org/10.1016/j.cell.2006.07.031

Mishra, N. (2008). *MicroRNA biomarkers in lupus*. Google Patents. Retrieved from http://www.google.com/patents/US20090123933

Molokhia, M., McKeigue, P. ., Cuadrado, M. J., and Hughes, G. (2001). Systemic lupus erythematosus from migrants in west Africa compared with afro-caribbean people in the UK. *Lancet*, *357*, 1414–1415.

Murphy, K., Travers, P., and Walport, M. (2011). Innate Immunity : The First Lines of Defense. *Janeway's Immunobiology*, 37–73. http://doi.org/9780815342434

Nagy, G., Koncz, A., Telarico, T., Fernandez, D., Érsek, B., Buzás, E., and Perl, A. (2010). Central role of nitric oxide in the pathogenesis of rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Research and Therapy*, *12*. http://doi.org/10.1186/ar3045

Oates, J. (2010). The biology of reactive intermediates in systemic lupus erythematosus. *Autoimmunity*, *43*(1), 56–63.

http://doi.org/10.3109/08916930903374683.The

Oates, J. C., and Gilkeson, G. S. (2006). The biology of nitric oxide and other reactive intermediates in systemic lupus erythematosus. *Clinical Immunology*, *121*(3), 243–50. http://doi.org/10.1016/j.clim.2006.06.001

Okada, Y., Shimane, K., Kochi, Y., Tahira, T., Suzuki, A., Higasa, K., … Yamamoto, K. (2012). A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. *PLoS Genetics*, *8*(1). http://doi.org/10.1371/journal.pgen.1002455

Pan, W., Zhu, S., Yuan, M., Cui, H., Wang, L., Luo, X., … Shen, N. (2010). MicroRNA-21 and MicroRNA-148a Contribute to DNA Hypomethylation in Lupus CD4+ T Cells by Directly and Indirectly Targeting DNA Methyltransferase 1. *The Journal of Immunology*, *184*(12), 6773–6781. http://doi.org/10.4049/jimmunol.0904060

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY , a Dicer Homolog , and HEN1 , a Novel Protein , Act in microRNA Metabolism in Arabidopsis thaliana. *Current Biology*, *12*(02), 1484–1495.

Patel, D. R., and Richardson, B. C. (2010). Epigenetic mechanisms in lupus. *Current Opinion in Rheumatology*, *22*(5), 478–82. http://doi.org/10.1097/BOR.0b013e32833ae915

Patel, D. R., and Richardson, B. C. (2013). Dissecting complex epigenetic alterations in human lupus. *Arthritis Research and Therapy*, *15*(1), 201. Retrieved from http://arthritis-research.com/content/15/1/201/abstract

Petri, M., Kim, M. Y., Kalunian, K. C., Grossman, J., Hahn, B. H., Sammaritano, L. R., … Askanase, A. D. (2005). Combined oral contraceptives in women with systemic lupus erythematosus. *New England Journal of Medicine*, *353*(24), 2550–2558. Retrieved from http://www.nejm.org/doi/full/10.1056/NEJMoa051135

Petri, M., Orbai, A.-M., Alarcon, G. S., Gordon, C., Merrill, J. T., Fortin, P. R., … Nived, O. (2013). Derivation and Validation of Systemic Lupus International Collaborating Clinics Classification Criteria for Systemic Lupus Erythematosus. *Arthritis and Rheumatology*, *64*(8), 2677–2686. http://doi.org/10.1002/art.34473.Derivation

Pons-Estel, G. J., Alarcón, G. S., Scofield, L., Reinlib, L., and Cooper, G. S. (2010). Understanding the Epidemiology and Progression of Systemic Lupus Erythematosus. *Seminars in Arthritis and Rheumatism*, *39*(4), 257–268. http://doi.org/10.1016/j.semarthrit.2008.10.007

Quinn, S. R., and O'Neill, L. A. (2011). A trio of microRNAs that control Toll-like receptor signalling. *International Immunology*, *23*(7), 421–425. http://doi.org/10.1093/intimm/dxr034

Ramos, P. S., Brown, E. E., Kimberly, R. P., and Langefeld, C. D. (2010). Genetic Factors Predisposing to Systemic Lupus Erythematosus and Lupus Nephritis. *Seminars in Nephrology*, *30*(2), 164–176. http://doi.org/10.1016/j.semnephrol.2010.01.007

Rehmsmeier, M., Steffen, P., Hochmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, *10*(10), 1507–1517. http://doi.org/10.1261/rna.5248604

Rynes, R. I. (1997). Antimalarial drugs in the treatment of rheumatological diseases. *British Journal of Rheumatology*, *36*(November 1996), 799–805.

Sanchez-Guerrero, J., Liang, M. H., Karlson, E. W., Hunter, D. J., and Colditz, G. A. (1995). Postmenopausal estrogen therapy and the risk for developing systemic lupus erythematosus. *Annals of Internal Medicine*, *122*(6), 430–433. Retrieved from http://annals.org/article.aspx?articleid=708498

Song, Y. J., Li, G., He, J. H., Guo, Y., and Yang, L. (2015). Bioinformatics-based identification of microRNA-regulated and rheumatoid arthritis-associated

genes. *PLoS ONE*, *10*(9), 1–19. http://doi.org/10.1371/journal.pone.0137551

Stagakis, E., Bertsias, G., Verginis, P., Nakou, M., Hatziapostolou, M., Kritikos, H., … Boumpas, D. T. (2011). Identification of novel microRNA signatures linked to human lupus disease activity and pathogenesis: miR-21 regulates aberrant T cell responses through regulation of PDCD4 expression. *Annals of the Rheumatic Diseases*, *70*(8), 1496–1506. http://doi.org/10.1136/ard.2010.139857

Sun, X.-Y., Shi, J., Han, L., Su, Y., and Li, Z.-G. (2008). Anti-histones antibodies in systemic lupus erythematosus: prevalence and frequency in neuropsychiatric lupus. *Journal of Clinical Laboratory Analysis*, *22*(4), 271–7. http://doi.org/10.1002/jcla.20248

Tang, Y., Luo, X., Cui, H., Ni, X., Yuan, M., Guo, Y., … Shen, N. (2009). MicroRNA-146a contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins. *Arthritis and Rheumatism*, *60*(4), 1065–1075. http://doi.org/10.1002/art.24436

Te, J. L., Dozmorov, I. M., Guthridge, J. M., Nguyen, K. L., Cavett, J. W., Kelly, J. A., … Ojwang, J. O. (2010). Identification of Unique MicroRNA Signature Associated with Lupus Nephritis. *PLoS ONE*, *5*(5), e10344. http://doi.org/10.1371/journal.pone.0010344

Tiffin, N., Adeyemo, A., and Okpechi, I. (2013). A diverse array of genetic factors contribute to the pathogenesis of systemic lupus erythematosus. *Orphanet J Rare Dis*, *8*(1), 2. Retrieved from http://www.biomedcentral.com/content/pdf/1750-1172-8-2.pdf

Tiffin, N., Hodkinson, B., and Okpechi, I. (2014). Lupus in Africa: can we dispel the myths and face the challenges? *Lupus*, *23*(1), 102–111. http://doi.org/10.1177/0961203313509296

Tikly, M., and V. Navarra, S. (2008). Lupus in the developing world – is it any

different? *Best Practice and Research Clinical Rheumatology*, *22*(4), 643–655. http://doi.org/10.1016/j.berh.2008.05.003

Tsokos, G. C. (2016). Systemic Lupus Erythematosis Basic, Applied and Clinical Aspects. In *Systemic Lupus Erythematosis Basic, Applied and Clinical Aspects* (pp. 260–262).

Tu, J., Liao, J., Chun, A., Luk, S., Leung, N., Tang, S., … Lee, T. (2015). MicroRNAs mediated targeting on the Yin-yang dynamics of DNA methylation in disease and development. *International Journal of Biochemistry and Cell Biology*, *67*, 115–120. http://doi.org/10.1016/j.biocel.2015.05.002

Vinuesa, C. G., Rigby, R. J., and Yu, D. (2009). Logic and Extent of miRNA-Mediated Control of Autoimmune Gene Expression. *International Reviews of Immunology*, *28*(3-4), 112–138. http://doi.org/10.1080/08830180902934909

Wakeland, E. K., Liu, K., Graham, R. R., and Behrens, T. W. (2001). Delineating the Genetic Basis of Systemic Lupus Erythematosus. *Immunity*, *15*(3), 397–408. http://doi.org/10.1016/S1074-7613(01)00201-1

Wang, C.-M., Chang, S.-W., Wu, Y.-J. J., Lin, J.-C., Ho, H.-H., Chou, T.-C., … Chen, J.-Y. (2014). Genetic variations in Toll-like receptors (TLRs 3/7/8) are associated with systemic lupus erythematosus in a Taiwanese population. *Scientific Reports*, *4*. http://doi.org/10.1038/srep03792

Wang, Y., Stricker, H. M., Gou, D., and Liu, L. (2007). MicroRNA: past and present. *Frontiers in Bioscience*, *12*, 2316–2329. http://doi.org/10.2741/2234

Weckerle, C. E., and Niewold, T. B. (2011). The Unexplained Female Predominance of Systemic Lupus Erythematosus: Clues from Genetic and Cytokine Studies. *Clinical Reviews in Allergy and Immunology*, *40*(1), 42–49. http://doi.org/10.1007/s12016-009-8192-4

Wright, M. W., and Bruford, E. a. (2011). Naming “junk”: human non-protein

coding RNA (ncRNA) gene nomenclature. *Human Genomics*, *5*(2), 90–98. http://doi.org/10.1186/1479-7364-5-2-90

Wu, H., Zhao, M., Chang, C., and Lu, Q. (2015). The Real Culprit in Systemic Lupus Erythematosus: Abnormal Epigenetic Regulation. *International Journal of Molecular Sciences*, *16*(5), 11013–11033. http://doi.org/10.3390/ijms160511013

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*, *37*(Database issue), D105–10. http://doi.org/10.1093/nar/gkn851

Yan, S., Yim, L. Y., Lu, L., Lau, C. S., and Chan, V. S.-F. (2014). MicroRNA Regulation in Systemic Lupus Erythematosus Pathogenesis. *Immune Network*, *14*(3), 138. http://doi.org/10.4110/in.2014.14.3.138

Zan, H., Tat, C., and Casali, P. (2014). microRNAs in lupus. *Autoimmunity*, *47*(4), 272–285. http://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Zhao, X., Tang, Y., Qu, B., Cui, H., Wang, S., Wang, L., … Shen, N. (2010). MicroRNA-125a contributes to elevated inflammatory chemokine RANTES levels via targeting KLF13 in systemic lupus erythematosus. *Arthritis and Rheumatism*, *62*(11), 3425–3435. http://doi.org/10.1002/art.27632

15. ADDENDUM A: The complete list of 749 Ensembl Gene Identifiers of differentially-expressed lupus genes obtained from Dr Wendy Kröger

ENSG00000002549	ENSG00000149636	ENSG00000102145	ENSG00000088325	ENSG00000105402	ENSG00000135604
ENSG00000003137	ENSG00000150244	ENSG00000102384	ENSG00000099139	ENSG00000105819	ENSG00000136527
ENSG00000005075	ENSG00000150281	ENSG00000103740	ENSG00000100003	ENSG00000106100	ENSG00000136928
ENSG00000005108	ENSG00000150403	ENSG00000104177	ENSG00000100206	ENSG00000107807	ENSG00000138050
ENSG00000007944	ENSG00000151834	ENSG00000104331	ENSG00000102970	ENSG00000108294	ENSG00000138617
ENSG00000011143	ENSG00000153292	ENSG00000105366	ENSG00000103942	ENSG00000108702	ENSG00000138642
ENSG00000013725	ENSG00000153993	ENSG00000105374	ENSG00000105700	ENSG00000109819	ENSG00000138646
ENSG00000018189	ENSG00000154114	ENSG00000105679	ENSG00000107957	ENSG00000111666	ENSG00000138698
ENSG00000023228	ENSG00000156103	ENSG00000106443	ENSG00000108176	ENSG00000111670	ENSG00000140464
ENSG00000040731	ENSG00000156976	ENSG00000107014	ENSG00000109182	ENSG00000111701	ENSG00000141646
ENSG00000065325	ENSG00000157014	ENSG00000108379	ENSG00000110448	ENSG00000111783	ENSG00000143198
ENSG00000066583	ENSG00000157216	ENSG00000108622	ENSG00000110841	ENSG00000112541	ENSG00000146425
ENSG00000067064	ENSG00000157399	ENSG00000108773	ENSG00000113389	ENSG00000112936	ENSG00000151789
ENSG00000068650	ENSG00000157601	ENSG00000109193	ENSG00000113494	ENSG00000115461	ENSG00000152377
ENSG00000074660	ENSG00000158901	ENSG00000112200	ENSG00000113889	ENSG00000115616	ENSG00000154736
ENSG00000075618	ENSG00000159224	ENSG00000112293	ENSG00000115761	ENSG00000118816	ENSG00000158863
ENSG00000076003	ENSG00000160200	ENSG00000112773	ENSG00000116209	ENSG00000119771	ENSG00000159259
ENSG00000080839	ENSG00000160870	ENSG00000112941	ENSG00000117069	ENSG00000121578	ENSG00000160613
ENSG00000081154	ENSG00000160932	ENSG00000113924	ENSG00000118503	ENSG00000121769	ENSG00000160868
ENSG00000082269	ENSG00000162976	ENSG00000115226	ENSG00000118523	ENSG00000124508	ENSG00000162747
ENSG00000085733	ENSG00000163132	ENSG00000118194	ENSG00000119042	ENSG00000127540	ENSG00000162775
ENSG00000087460	ENSG00000163251	ENSG00000118640	ENSG00000120149	ENSG00000128714	ENSG00000162992
ENSG00000088280	ENSG00000163254	ENSG00000118702	ENSG00000120805	ENSG00000128872	ENSG00000164418

ENSG00000089012	ENSG00000163347	ENSG00000119509	ENSG00000122140	ENSG00000130396	ENSG00000164604
ENSG00000089127	ENSG00000163635	ENSG00000119917	ENSG00000122971	ENSG00000131264	ENSG00000164751
ENSG00000089723	ENSG00000164399	ENSG00000122592	ENSG00000124688	ENSG00000132475	ENSG00000165678
ENSG00000091181	ENSG00000164924	ENSG00000123096	ENSG00000124762	ENSG00000134115	ENSG00000166741
ENSG00000095066	ENSG00000165029	ENSG00000123500	ENSG00000128652	ENSG00000136153	ENSG00000167755
ENSG00000100249	ENSG00000165629	ENSG00000125944	ENSG00000128805	ENSG00000136950	ENSG00000168484
ENSG00000100351	ENSG00000165949	ENSG00000127603	ENSG00000129028	ENSG00000137269	ENSG00000169397
ENSG00000100490	ENSG00000166780	ENSG00000128408	ENSG00000129596	ENSG00000137804	ENSG00000173805
ENSG00000100604	ENSG00000168028	ENSG00000128928	ENSG00000130813	ENSG00000139180	ENSG00000176387
ENSG00000100902	ENSG00000168453	ENSG00000129221	ENSG00000132330	ENSG00000139209	ENSG00000178498
ENSG00000101311	ENSG00000168505	ENSG00000130489	ENSG00000132383	ENSG00000142676	ENSG00000180530
ENSG00000101850	ENSG00000169035	ENSG00000131097	ENSG00000132530	ENSG00000144191	ENSG00000181433
ENSG00000101890	ENSG00000169760	ENSG00000133048	ENSG00000132561	ENSG00000144566	ENSG00000183648
ENSG00000102057	ENSG00000169764	ENSG00000133872	ENSG00000133026	ENSG00000145592	ENSG00000184304
ENSG00000102290	ENSG00000170231	ENSG00000134013	ENSG00000133636	ENSG00000152049	ENSG00000185404
ENSG00000102871	ENSG00000170476	ENSG00000134245	ENSG00000134240	ENSG00000157152	ENSG00000185670
ENSG00000104368	ENSG00000171557	ENSG00000134760	ENSG00000137500	ENSG00000160145	ENSG00000187607
ENSG00000104497	ENSG00000171953	ENSG00000135114	ENSG00000138069	ENSG00000160710	ENSG00000197165
ENSG00000104723	ENSG00000172016	ENSG00000135480	ENSG00000138092	ENSG00000160796	ENSG00000197641
ENSG00000105619	ENSG00000172183	ENSG00000135547	ENSG00000140398	ENSG00000163823	ENSG00000198039
ENSG00000107018	ENSG00000172466	ENSG00000136141	ENSG00000140577	ENSG00000164080	ENSG00000198176
ENSG00000107317	ENSG00000172724	ENSG00000136688	ENSG00000143318	ENSG00000164708	ENSG00000198445
ENSG00000107438	ENSG00000173064	ENSG00000136888	ENSG00000144285	ENSG00000164818	ENSG00000204936
ENSG00000107890	ENSG00000173163	ENSG00000137225	ENSG00000144366	ENSG00000165119	ENSG00000205420
ENSG00000108684	ENSG00000174963	ENSG00000137720	ENSG00000145335	ENSG00000165973	ENSG00000242600
ENSG00000108924	ENSG00000177910	ENSG00000137860	ENSG00000145833	ENSG00000166908	ENSG00000243130
ENSG00000109163	ENSG00000180573	ENSG00000137877	ENSG00000146378	ENSG00000168066	ENSG00000008952
ENSG00000109832	ENSG00000182168	ENSG00000137959	ENSG00000148344	ENSG00000169704	ENSG00000016602

ENSG00000110063	ENSG00000182378	ENSG00000138101	ENSG00000149922	ENSG00000171282	ENSG00000035720
ENSG00000110660	ENSG00000182415	ENSG00000142515	ENSG00000150636	ENSG00000171467	ENSG00000038295
ENSG00000110777	ENSG00000184216	ENSG00000143303	ENSG00000152952	ENSG00000172572	ENSG00000066468
ENSG00000111046	ENSG00000184979	ENSG00000143556	ENSG00000153560	ENSG00000172987	ENSG00000069696
ENSG00000111331	ENSG00000185044	ENSG00000143575	ENSG00000156802	ENSG00000173812	ENSG00000081248
ENSG00000111335	ENSG00000185201	ENSG00000143631	ENSG00000158373	ENSG00000174607	ENSG00000087494
ENSG00000111775	ENSG00000185652	ENSG00000144810	ENSG00000160293	ENSG00000174720	ENSG00000088386
ENSG00000112144	ENSG00000185745	ENSG00000145623	ENSG00000163106	ENSG00000176834	ENSG00000091513
ENSG00000112365	ENSG00000185875	ENSG00000145826	ENSG00000164070	ENSG00000177853	ENSG00000099622
ENSG00000112406	ENSG00000186111	ENSG00000145990	ENSG00000164136	ENSG00000178403	ENSG00000102854
ENSG00000112414	ENSG00000187140	ENSG00000147316	ENSG00000164342	ENSG00000179455	ENSG00000105472
ENSG00000112419	ENSG00000187210	ENSG00000147381	ENSG00000164626	ENSG00000183207	ENSG00000106304
ENSG00000112576	ENSG00000187446	ENSG00000149573	ENSG00000164821	ENSG00000185088	ENSG00000107984
ENSG00000112983	ENSG00000187513	ENSG00000150394	ENSG00000165055	ENSG00000185585	ENSG00000108039
ENSG00000113361	ENSG00000187608	ENSG00000153822	ENSG00000166233	ENSG00000185973	ENSG00000112706
ENSG00000114120	ENSG00000187621	ENSG00000155130	ENSG00000167618	ENSG00000187801	ENSG00000113141
ENSG00000114166	ENSG00000188064	ENSG00000155380	ENSG00000169908	ENSG00000189056	ENSG00000115008
ENSG00000114487	ENSG00000188937	ENSG00000155760	ENSG00000169933	ENSG00000196159	ENSG00000115520
ENSG00000114491	ENSG00000196116	ENSG00000158636	ENSG00000171189	ENSG00000196639	ENSG00000116830
ENSG00000116748	ENSG00000196428	ENSG00000159450	ENSG00000172115	ENSG00000197465	ENSG00000118402
ENSG00000117477	ENSG00000196611	ENSG00000160714	ENSG00000173826	ENSG00000198393	ENSG00000126214
ENSG00000117501	ENSG00000196655	ENSG00000163624	ENSG00000175183	ENSG00000198466	ENSG00000127928
ENSG00000118307	ENSG00000197446	ENSG00000163737	ENSG00000176024	ENSG00000204104	ENSG00000130377
ENSG00000118492	ENSG00000197540	ENSG00000163815	ENSG00000176293	ENSG00000205531	ENSG00000132432
ENSG00000118526	ENSG00000197912	ENSG00000164402	ENSG00000182631	ENSG00000206190	ENSG00000134516
ENSG00000119508	ENSG00000198001	ENSG00000166923	ENSG00000183873	ENSG00000211630	ENSG00000135624
ENSG00000120963	ENSG00000198498	ENSG00000167106	ENSG00000184908	ENSG00000249859	ENSG00000139292
ENSG00000122194	ENSG00000198610	ENSG00000167323	ENSG00000185070	ENSG00000250588	ENSG00000141012

ENSG00000122566	ENSG00000198670	ENSG00000167863	ENSG00000185567	ENSG00000256018	ENSG00000145779
ENSG00000123843	ENSG00000198692	ENSG00000168135	ENSG00000187391	ENSG00000007952	ENSG00000147246
ENSG00000124535	ENSG00000204941	ENSG00000168242	ENSG00000189143	ENSG00000008300	ENSG00000154240
ENSG00000124678	ENSG00000204983	ENSG00000168653	ENSG00000196208	ENSG00000015479	ENSG00000156076
ENSG00000125363	ENSG00000211677	ENSG00000168938	ENSG00000196433	ENSG00000022267	ENSG00000159346
ENSG00000125571	ENSG00000213079	ENSG00000169919	ENSG00000197417	ENSG00000039068	ENSG00000161217
ENSG00000125843	ENSG00000213689	ENSG00000170145	ENSG00000197496	ENSG00000043462	ENSG00000162368
ENSG00000126267	ENSG00000213931	ENSG00000171208	ENSG00000197822	ENSG00000064012	ENSG00000164010
ENSG00000126549	ENSG00000213949	ENSG00000171431	ENSG00000197919	ENSG00000065923	ENSG00000166526
ENSG00000126709	ENSG00000216775	ENSG00000171786	ENSG00000197977	ENSG00000073584	ENSG00000166704
ENSG00000129951	ENSG00000221818	ENSG00000172232	ENSG00000198366	ENSG00000075391	ENSG00000167395
ENSG00000130193	ENSG00000221826	ENSG00000172476	ENSG00000198952	ENSG00000077150	ENSG00000168065
ENSG00000131370	ENSG00000235604	ENSG00000173702	ENSG00000203747	ENSG00000078114	ENSG00000169946
ENSG00000133020	ENSG00000236249	ENSG00000174669	ENSG00000203813	ENSG00000080345	ENSG00000170689
ENSG00000134602	ENSG00000241322	ENSG00000175279	ENSG00000204103	ENSG00000082068	ENSG00000170950
ENSG00000135063	ENSG00000257017	ENSG00000175550	ENSG00000204519	ENSG00000083454	ENSG00000171475
ENSG00000136213	ENSG00000260230	ENSG00000178568	ENSG00000204852	ENSG00000091704	ENSG00000173786
ENSG00000136271	ENSG00000010017	ENSG00000182512	ENSG00000213160	ENSG00000100207	ENSG00000174255
ENSG00000136732	ENSG00000010292	ENSG00000183036	ENSG00000213398	ENSG00000100949	ENSG00000174469
ENSG00000136866	ENSG00000032219	ENSG00000184330	ENSG00000213722	ENSG00000101981	ENSG00000175213
ENSG00000137198	ENSG00000042813	ENSG00000184678	ENSG00000221914	ENSG00000103051	ENSG00000180209
ENSG00000137573	ENSG00000054654	ENSG00000185313	ENSG00000270742	ENSG00000105640	ENSG00000181619
ENSG00000137628	ENSG00000055208	ENSG00000187498	ENSG00000003989	ENSG00000106025	ENSG00000181634
ENSG00000137634	ENSG00000055732	ENSG00000188313	ENSG00000006788	ENSG00000106153	ENSG00000183778
ENSG00000137672	ENSG00000059378	ENSG00000196150	ENSG00000022840	ENSG00000106819	ENSG00000185825
ENSG00000137965	ENSG00000065371	ENSG00000196371	ENSG00000042304	ENSG00000107201	ENSG00000186001
ENSG00000138792	ENSG00000067646	ENSG00000196937	ENSG00000048740	ENSG00000109061	ENSG00000187688
ENSG00000138944	ENSG00000068489	ENSG00000198046	ENSG00000057019	ENSG00000110092	ENSG00000187758

ENSG00000139793	ENSG00000070019	ENSG00000198258	ENSG00000067842	ENSG00000110195	ENSG00000187954
ENSG00000140365	ENSG00000070495	ENSG00000203857	ENSG00000070778	ENSG00000113525	ENSG00000189266
ENSG00000141542	ENSG00000072786	ENSG00000204616	ENSG00000073350	ENSG00000120937	ENSG00000196782
ENSG00000142973	ENSG00000073614	ENSG00000213231	ENSG00000080823	ENSG00000129354	ENSG00000197079
ENSG00000143772	ENSG00000073910	ENSG00000213699	ENSG00000082175	ENSG00000130299	ENSG00000198518
ENSG00000143786	ENSG00000077522	ENSG00000221946	ENSG00000083782	ENSG00000130303	ENSG00000198753
ENSG00000143947	ENSG00000078018	ENSG00000228315	ENSG00000089053	ENSG00000130529	ENSG00000204568
ENSG00000144426	ENSG00000078579	ENSG00000262556	ENSG00000089820	ENSG00000130768	ENSG00000204592
ENSG00000145287	ENSG00000087842	ENSG00000267272	ENSG00000091136	ENSG00000131471	ENSG00000204610
ENSG00000145864	ENSG00000089775	ENSG00000267855	ENSG00000092377	ENSG00000131831	ENSG00000205084
ENSG00000145945	ENSG00000091164	ENSG00000003249	ENSG00000099251	ENSG00000132429	ENSG00000205413
ENSG00000146938	ENSG00000091527	ENSG00000048545	ENSG00000099812	ENSG00000132703	ENSG00000211815
ENSG00000147255	ENSG00000095752	ENSG00000064601	ENSG00000101463	ENSG00000133110	ENSG00000213928
ENSG00000147804	ENSG00000100156	ENSG00000064886	ENSG00000102243	ENSG00000133773	ENSG00000221955
ENSG00000148175	ENSG00000100385	ENSG00000069956	ENSG00000103089	ENSG00000135111	ENSG00000233791
ENSG00000148484	ENSG00000101057	ENSG00000072415	ENSG00000104332	ENSG00000135333	ENSG00000241484
ENSG00000148700	ENSG00000101448	ENSG00000086589	ENSG00000104938	ENSG00000135390	ENSG00000242689
ENSG00000149596	ENSG00000101608	ENSG00000086991	ENSG00000104967	ENSG00000135482	

16. ADDENDUM B: A complete list of the genes found in miRTarBase to be targeted by miRNAs which were verified as a perfect string match in miRBase

miRNA	Number of gene targets	Gene targets
hsa-miR-184	3	*AKT2, INPPL1, NFATC2*
hsa-miR-198	3	*CCNT1, NTRK3, MYB*
hsa-miR-298	2	*BACE1, CDKN1A*
hsa-miR-638	1	*OSCP1*
hsa-miR-1224-3p	5	*KMT2D, IGSF3, TCF12, HNRNPA3, CDC42EP1*
hsa-miR-142-3p	107	*RAC1, ARNTL, TNIP3, VSIG1, PSMB5, HOXA13, NAP1L2, CLOCK, ARHGAP12, IRF2BP1, KCND3, MESP2, CRH, PITPNC1, SLC37A3, PPP2R2C, ARL14EP, CDCA4, SH2D1A, MROH7, ETNK2, PCDHB14, C2, FAM209A, PCDHGA2, KRTAP4-5, C4BPB, ZNF608, GPC4, GPD1L, EVI2B, ZNF236, NEFH, HSFY2, ZNF701, UGT2B17, INSIG2, TAS2R7, RGS5, TPSG1, ARL6IP6, PCDHGA10, SIGLEC10, MRGPRX1, AGBL2, PSMD12, BMP8A, C11orf16, INCENP, ALAD, TRIM36, AGTR2, TUBAL3, RNF170, NCKAP1, FAM65A, C7orf31, CLUAP1, ENAH, TMEM136, CLEC4D, KCTD12, ABCC9, CDC6, MCFD2, DOCK10, DCTN4, SECISBP2, KIAA1191, BRMS1L, SYT4, ARF4, FAM98A, RRM2B, UBE2W, ARL10, PCMTD2, CALCOCO2, MORF4L1, TSPAN6, GOLGA5, YES1, ARL15, FAM177A1, CA5B, NSFL1C, MAN1A2, SKP2, FBXO3, EDEM3, DNAJB4, CNIH4, CCNJ, RPE, CLIC4, TGFBR1, ZCCHC14, C5orf24, MORF4L2, SLC35F5, TAOK1, GNAQ, ITGAV, ATP2A2, BOD1, C9orf72, PROM1*
hsa-miR-17-5p	310	*ZNFX1, CCL1, GPR137B, NABP1, NPAT, YES1, JAK1, PTEN, CDKN1A, PTPRO, PKD2, BCL2L11, E2F1, MAP3K12, BCL2, MEF2D, RUNX1, APP, VEGFA, MAPK9, DNAJC27, FBXO31, TGFBR2, TNFSF12, MUC17, BMPR2, CCND1, MYC, NCOA3, THBS1, SMAD4, ICAM1, SELE, CCND2, E2F3, RB1, RBL1, RBL2, WEE1, RND3, TCF3, HSPB2, MMP2, HBP1, SIRPA, MOB1B, FBXO28, TJP1, SLC35E2B, ZC3H18, UBE4A, DHX33, RPS27A, HIST1H4C, GDAP1, TMEM165, SOX4, TRIM11, MRPL40, PIGS, ND4, PPP1R15A, USP38, RAB23, LARP1, ZSWIM3, ADARB1, MEN1, FAM8A1, NPAS2, ATP6, STK11, ACOT2, ATF3, ASNS, RTTN, ZNF689, HTT, IMMT, TBC1D15, PSD3, HIST1H2AM, HUWE1, HIST3H2A, APEX1, AZIN1, RPS15A, COA1, CPE, MBNL1, IGFBP5, HNRNPU, ND2, PIK3CA, PAIP1, EEF1A1, MTRF1L, KAT2A, POGK, OPTN, MYCBP2, SYNDIG1, UBE2S, RYR2, HIST2H2AA3, COPS3, MGEA5, PELI1, CETN2, AGO1, RPL37, REEP5, MRPS6, CHD4, SLC25A3, MUC21, KDM4A, HIST2H4B, COX2, ATP5B, CBL, CNEP1R1, NOC2L, CANX, ENPP5, EARS2, RPL7, CAP1, UBE2C, RMDN1, SLC25A37, NFAT5, DCTPP1, MFHAS1, ASH1L, HDAC10, RBM5, UBE3C, DEPDC1, TRUB1, TMED10, MDK, ZBED3, LCOR, NUCKS1, KIF5C, WDR82, COX7B, SH3GLB2, TAF9B, NAPEPLD, ATRX,*

		PTBP1, FTH1, PLXNA1, OFD1, PER1, ARL9, ARHGAP5, SIPA1L3, QARS, ELP2, SURF4, RPL21, CTSA, EPB41L5, ABI2, UQCRFS1, TMEM9, MED12, NAGK, TSC2, GPM6A, SEPT2, TEFM, FASN, AMD1, TRA2B, DCUN1D4, PRICKLE1,SEPT11, CHST14, TRIM44, SON, SMIM15, LAPTM4A, CHTF8, SMAD3, NONO, EIF4G3, NOTCH2, SDHA, XIAP, DNMBP, TRAP1, EPB41L2, MED13, TMTC4, BACE1, POGZ, GDF11, GPI, MNT, VPRBP, PGAM1, KANSL1,GAPDH, PRPF8, FOPNL, TUBB4B, DPYSL2, NBR1, KIAA1919, ZNF507, BLVRA, PPP2R1A, DCBLD2, CLPTM1, ARHGEF7, HIST2H3A, FANCA, MORF4L2, RLF, RAB5B, TOB1, EFHC1, MTMR3, TNFRSF10B, GANAB, SLC16A2, RTN3, ERCC2, ATXN7, HSP90AA1, GLO1, LPIN1, LSM14A, ILF3, ANKRD27, TNPO3, NAT8L, PTTG1, ARPC2, HN1, SLC25A28, TRRAP, ELMO2, PCNX, TMEM19, PBXIP1, CCT6A, HDHD1, PLS1, RNF145, AK1, TCEA2, PLAG1, PDPK1, TNFAIP1, RTCA, TUT1, LAS1L, RPSA, TMSB10, ERLIN1, EHMT2, BAZ2A, ARHGAP10, PDLIM5, CCDC47, CPSF1, MTF1, C15orf41, NUFIP2, GNB1, IRAK1, CPS1, FXR2, SUOX, STIL, WAC, FER, EIF4G2, BTN3A1, TRIM8, RUNDC1, ZFYVE26, RNF146, PPP1CA, NAP1L1, ZIK1, MKI67, MAP3K8, ZNF598, TRIM71, KIF1A,MAP3K9, KMT2A, RALGAPA1,ZBTB5, PCGF5, GNAS, WASL, RBM12B, TXLNA, ARIH1, KCTD7, HEXIM1, DAPK3, PLEKHM1
hsa-miR-299-3p	1	*FXN*
hsa-miR-409-3p	6	*PHF10, FGB, FGA, FGG, ANG, IFNG*
hsa-miR-423-5p	168	*RABAC1, NMD3, YWHAZ, RGPD2, TSPYL6, HIST2H2BE, USP6, PGAM4, DIP2A, NOS1AP, IGDCC4, RC3H1, KIF1A, SRSF11, BCLAF1, RPS15A, CCNI, GDF11, FAM127A, NME2, COX1, ARPC1A, MYBL2, ND5, NDE1, CDKN2A, TDRKH, AKAP13, RPS10, PRPF38B, PPIAL4G, MRPL34, NDRG1, ZBTB34, ATP6, ZFP36L1, COQ6, EPM2AIP1, REPS1, AAR2, PPIA, MKNK2, DDX21, SLC6A6, RPS2, CSTB, ABCC5, PAK1, ERRFI1, CEACAM6, FHOD1, LMNB2, NECAP1, NPLOC4, PMPCA, C12orf10, DDX54, VAC14, RUNX1, FAM98A, EPN1, TMUB1, UBE2L3, FAM110B, MICALL1, PNRC1, DLX5, ZNF655, FAAH, ATN1, ATP6V1E1, SLC9A3R1, NOLC1, PLCB1, SQSTM1, CLSTN1, ULK1, ATP2B4, EPB41, HIST1H1C, AGPAT6, TPGS2, FKBP4, GNG12, RPS9, RPL18A, CCND2, PCBP1, AP3M1, PDPK1, WBP2, CLSPN, PTTG1, GGA3, GAPDH, ARHGDIA, KCTD2, RNF139, ARHGAP32, HIST1H2BK, SAR1B, PA2G4, UBE2D3, GTPBP1, HIST2H2BF, NUTM1, COPS7A, POLR3H, HIST1H1E, NFKB2, SLC7A2, TIMP3, PTMA, CSDE1, CD81, CCNF, HNRNPUL1, RCE1, FAM83H, HES4, FAM189B, SLC48A1, SNX8, ZYX, DRAP1, TEX2, C15orf39, LRCH4, CHD8, MYC, TRIM28, SEPN1, MIER2, HOXA7, PRRC2A, SRP54, PABPC1, SRM, GTF2F1, HIPK2, NEDD4L, COPZ1, GLUL, PPARGC1B, PKM, MTCH2, ACTG1, MTMR3, ARL6IP1, DDX24, TUBB, NSD1, PAX2, ATP5B, CEBPB, YARS, STRIP2, KCTD15, RPL23A, MFGE8, SF3B3, RPL10, SALL1, BCCIP, PSMD6, MED28, MESDC1, DAG1*
hsa-miR-483-5p	1	*MAPK3*

17. ADDENDUM C: A complete list of the 13 unverified miRNAs, for which gene targets were found in miRecords

miRNA	Number of gene targets	Gene targets
hsa-miR-21	49	*TPM1, NFIB, PDCD4, SERPINB5, CDKN1A, FAS, FAM3C, HIPK3, PRRG4, ACTA2, BTG2, BMPR2, SESN1, IL6R, SOCS5, GLCCI1, APAF1, SLC16A10, SGK3, RP2, CDK6, CFL2, RECK, TIMP3, MTAP, SOX5, JMY, TGFBR2, TGFBR3, HNRNPK, tp63, TOPORS, DAXX, TP53BP2, PPIF, PTEN, pten, TGFB1, RASA1, RASGRP1, E2F1, CDC25A, LRRFIP1, JAG1, MARCKS, Peli1, PPARA, STAT3, NT-3*
hsa-miR-126	6	*VCAM1, PIK3R2, CRKL, IRS1, VEGFA, TOM1*
hsa-miR-141	13	*CLOCK, TGFB2, ZEB2, FOG2, ERBB2IP, BAP1, KLHL20, PTPRD, ELMO2, WDR37, MKK4, SIP1, JAG1*
hsa-miR-145	19	*PARP8, IRS1, MYC, FSCN1, CCNA2, MUC1, CLINT1, CBFB, PPP3CA, FBXO28, ACBD3, USP46, RASA1, CCDC25, C11orf58, IGF-IR, KRT7, RTKN, SOX9*
hsa-miR-150	7	*gag-pol, cmyb, MYB, CCNE1, P2RX7, EGR2, MUC4*
hsa-miR-155	24	*AGTR1, BACH1, LDOC1, MATR3, TM6SF1, RHOA, ETS1, MEIS1, MAP3K7IP2, CYR61, MyD88, FOXO3, SOCS1, SHIP, CEBPB, JARID2, KGF, IKBKE, SMAD2, FADD, MeCP2, CCND1, SPI1, CTLA-4*
hsa-miR-224	4	*API5, KLK10, KLK1, AP2M1*
hsa-miR-342	1	*DNMT1*
hsa-miR-146a	38	*IRAK1, TRAF6, BRCA2, BRCA1, CXCR4, TBP, MCP2, CCNA2, PA2G4, FADD, METTL7A, CDKN3, PRR15, MCPH1, MCM10, POLE2, PDIK1L, VWCE, LTB, ELL3, RAD54L, KIF22, CFH, TMSL8, MR1, ATOH8, UHRF1, TRIM14, COL13A1, TIMELESS, PBLD, BLMH, NFIX, PEX11G, SDCBP2, PLEKHA4, IRF-5, STAT1*
hsa-miR-148a	6	*DNMT3B, NR1I2, Rps6ka5, DNMT1, MSK1, CDNK1B,*
hsa-miR-17-92	2	*PTPRO, HIF1A*
hsa-miR-196a	9	*HOXB8, HOXC8, HOXD8, HOXA7, IKBKB, ANXA1, KRT5, SPRR2C, S100A9*
hsa-miR-23b	2	*PLAU, MET*

18. ADDENDUM D: The 36 regular expression string matched miRNAs for which gene targets were found in the miRTarBase database

miRNA	Number of gene targets	Gene targets
hsa-miR-21-5p	489	*RASGRP1, CDC25A, BCL2, WIBG, WFS1, TM9SF3, RTN4, RPS7, PLOD3, PDHA2, NCAPG, DERL1, BASP1, JAG1, REST, SMARCA4, SPRY2, DUSP10, TIMP3, SOX5, MTAP, DOCK7, DOCK5, RECK, PIAS3, E2F2, FMOD, TGFBR2, PTEN, E2F1, TGFBI, LRRFIP1, MARCKS, SP1, CCL20, TPM1, NFIB, CDK6, RP2, SGK3, SLC16A10, APAF1, GLCCI1, SOCS5, SESN1, BTG2, FAM3C, FAS, PDCD4, RHOB, ANP32A, SERPINB5, NFKB1, BMPR2, TGIF1, NCOA3, MYC, ERBB2, SPATS2L, EIF2S1, PCBP1, JMY, TOPORS, HNRNPK, DAXX, TP53BP2, TP63, TGFBR3, PPIF, MSH2, MSH6, TIAM1, ISCU, MEF2C, EIF4A2, ANKRD46, EGFR, IL1B, ICAM1, PLAT, PTX3, TNFAIP3, CCR1, CDK2AP1, DOCK4, PPARA, FASLG, SOD3, TCF21, SMAD7, PHIP, PALLD, BOC, DAAM1, C20orf194, COL5A2, TPRG1L, NIN, PPAP2A, TRIM38, DDR2, VPS54, KIAA1715, BAZ1B, LIN7C, USP34, GNE, ZBTB8A, CLOCK, LPGAT1, NIPBL, DDHD2, C2orf43, SGCB, FANCI, EDIL3, DTX3L, IREB2, MGAT4A, RAPH1, CYBRD1, SLAIN2, KIAA1551, SOX2, GTF2A1, RRAGC, PIGN, HPS5, SESTD1, MEF2A, ZBTB47, ELOVL4, EPHA4, ENAH, TBL1XR1, NBEA, TAF5, AP3M1, CD47, LATS1, CPEB3, RMND5A, KBTBD7, MEIS1, GNB4, MMP2, VEGFA, REV3L, TXLNG2P, SCAF11, HNRNPH1, PARP1, CERS6, SNRNP48, FUBP1, TGFB1, ANKRD28, SYNE2, OSBPL1A, RSPRY1, SRSF11, ASRGL1, RB1, EXOC8, IVNS1ABP, UTRN, ACAT1, PFKFB2, PDGFD, COBLL1, SAMD5, MOAP1, KLHL24, ZNF326, WHSC1, BTBD7, GAPVD1, TNS3, ESYT2, MYEF2, PKD2, PTAR1, PAN3, ZRANB1, MON2, RASEF, GNAQ, HERPUD2, MAP3K2, THOC2, RASGRP3, PHF17, USP7, GTF2I, MKNK2, CLCN5, TRIM2, LONRF2, EIF4EBP2, PHF16, ELOVL7, SASH1, RAB22A, SUZ12, SPIN1, POLR3B, ATP2B4, CCDC34, PGRMC2, TRPM7, GID4, PLEKHA1, MALT1, ZNF367, TMEM245, MBNL1, MORC3, ZNF35, FOXN3, NR2C2, FBXO3, GPAM, PM20D2, CLIP4, DMTF1, DDX46, ATRX, LMBR1, MUC1, TUBGCP5, RHOQ, TET1, CASC5, ARHGAP21, MDM4, LARS, FAXDC2, CDK19, LAMP2, MOXD1, AHSA2, PIK3C2A, TMEM2, IPP, RAI14, MYO9A, MRPS10, FBXL17, AGGF1, SLC26A2, PRKAB2, SREK1, FAM20B, ARMCX3, SERPINI1, DDAH1, PROSER1, PRICKLE2, JPH1, NAA30, FKBP5, TGFB2, ZMYM2, TRIM33, RAPGEF6, MTPN, GOLGA4, SACM1L, ATAD2B, FIGN, EPM2A, RAB6C, TSNAX, TRAPPC2, LIFR, CALD1, LYRM7, PIK3R1, RPS6KA3, ADNP, LRRC57, DUSP8, KLHL15, GXYLT2, PBRM1, SMNDC1, NCSTN,*

		PPM1L, MMP9, PARP9, CKAP5, CAPRIN1, CNTRL, PHTF1, MRAP2, MYCBP2, ITSN2, BCAT1, PPFIA4, YME1L1, ZNF667, ETNK1, CYP4V2, BRCA1, DOCK10, DMD, TMX4, PTBP3, APOLD1, CCT6P1, UBR3, NEK1, DYNC1LI2, CSNK1A1, ACTR2, TRIM59, SPTLC3, TMEM56, ZADH2, DDX3X, ZNF217, FAM217B, ZBTB20, FNBP1, ZFYVE16, VPS13A, B3GALNT1, SLC5A3, GPD1L, STRBP, PRKCE, TOP2A, SEC63, TAF1, MPP5, PTK2, GPD2, SSFA2, HECTD1, PAG1, SFXN1, HAPLN1, EIF5, KLHL42, NFAT5, SLC9A6, ZNF207, RALGPS2, FAM46A, MTMR12, KBTBD6, TNRC6B, ABCD3, PKNOX1, RAB11FIP2, PURB, LCORL, KLF5, ST6GAL1, GPR64, SKP2, ECI2, NUFIP2, GRPEL2, HS3ST3B1, ELAVL4, CORO2A, TTC33, AUTS2, AKAP9, NKTR, PTPN14, STXBP5, MGA, KIFAP3, SPG11, PER3, ZNF292, ZNF587, AIM1, DSE, SRPK2, TSHZ3, CEP97, DCAF10, PHF20, SEMA5A, PHACTR2, RABGAP1, VPS36, WNK1, NUBPL, LIMCH1, MTMR9, SERAC1, RUFY3, APPL1, PBX1, OSBPL3, WNT5A, REV1, FBXL2, HOXA9, MIB1, FMR1, SNX30, SLC17A5, PREPL, CCDC14, ARHGEF12, USP47, BTBD3, ARID4A, ATP11B, SLC31A1, DLG1, SOWAHC, RSF1, SNRK, PLD1, TLR4, VPS26A, ZYG11B, AGO4, TESK2, MAP3K1, RAB6A, PURA, OLR1, KLF9, ATMIN, PRPF39, AGO2, STAG2, ACBD5, AKT2, UGGT1, PRRC1, NT5C2, ZNF532, DCAF8, UBR5, PTPN3, ATF2, EXOC5, AFTPH, WDR7, FAM126B, PTGFR, PHF20L1, B3GNT5, GLG1, CEP152, HMGB3, BDH2, NETO2, ALMS1, SMC1A, PIGX, SAR1A, APC, SLMAP, SOCS4, ZBTB38, OSR1, FBXO11, SCRN1, PTPDC1, CCNG1, MEGF9, TNPO1, FERMT2, KAT6A, TNFRSF11B, WWC2, WNK3, WHSC1L1, VASH2, SLK, FILIP1L, DCP1A, RNF11, RDH11, PELI1, YOD1, STAT3, BCL6, OTUD1, TMEM147
hsa-miR-21-3p	2	*SUV420H1, MED25*
hsa-miR-126-5p	6	*SLC45A3, PTPN7, ADAM9, MMP7, CXCL12, CSGALNACT1,*
hsa-miR-126-3p	30	*PITPNC1, IGFBP2, KRAS, SPRED1, PLK2, EGFL7, SLC45A3, CCNE2, RGS3, TOM1, HOXA9, MERTK, CRK, VEGFA, PIK3R2, VCAM1, IRS1, E2F1, SOX2, TWF1, TWF2, PTPN7, DNMT1, SLC7A5, TEK, ADAM9, MMP7, CXCL12, PGR, RBMX*
hsa-miR-141-3p	37	*MAPK14, ZEB2, ZEB1, DLX5, BAP1, ACVR2B, KLF5, STK3, TGFB2, SFPQ, CLOCK, BRD3, UBAP1, PTEN, ZFPM2,TRAPPC2P1, EIF4E, PPARA, NR0B2, YWHAG, ELAVL4, MAPK9, TFDP2, E2F3, SHC1, VAC14, TCF7L1, ELMO2, RASSF2, KLHL20, RIN2, SEPT7, HOXB5, ERBB2IP, KLF11, PTPRD, WDR37*
hsa-miR-142-5p	1	*NFE2L2*
hsa-miR-142-3p	107	*RAC1, ARNTL, TNIP3, VSIG1, PSMB5, HOXA13, NAP1L2, CLOCK, ARHGAP12, IRF2BP1, KCND3, MESP2, CRH, PITPNC1, SLC37A3, PPP2R2C, ARL14EP, CDCA4, SH2D1A, MROH7, ETNK2, PCDHB14, C2, FAM209A, PCDHGA2, KRTAP4-5, C4BPB, ZNF608, GPC4, GPD1L, EVI2B, ZNF236, NEFH, HSFY2, ZNF701, UGT2B17, INSIG2, TAS2R7, RGS5, TPSG1, ARL6IP6, PCDHGA10,*

		SIGLEC10, MRGPRX1, AGBL2, PSMD12, BMP8A, C11orf16, INCENP, ALAD, TRIM36, AGTR2, TUBAL3, RNF170, NCKAP1, FAM65A, C7orf31, CLUAP1, ENAH, TMEM136, CLEC4D, KCTD12, ABCC9, CDC6, MCFD2, DOCK10, DCTN4, SECISBP2, KIAA1191, BRMS1L, SYT4, ARF4, FAM98A, RRM2B, UBE2W, ARL10, PCMTD2, CALCOCO2, MORF4L1, TSPAN6, GOLGA5, YES1, ARL15, FAM177A1, CA5B, NSFL1C, MAN1A2, SKP2, FBXO3, EDEM3, DNAJB4, CNIH4, CCNJ, RPE, CLIC4, TGFBR1, ZCCHC14, C5orf24, MORF4L2, SLC35F5, TAOK1, GNAQ, ITGAV, ATP2A2, BOD1, C9orf72, PROM1
hsa-miR-145-5p	89	*BNIP3, SOX2, KLF4, MUC1, MYO6, CDKN1A, ITGB8, STAT1, YES1, CBFB, PPP3CA, CLINT1, IRS1, IRS2, VEGFA, TMOD3, HOXA9, FSCN1, MYC, FLI1, IFNB1, TIRAP, POU5F1, IGF1R, KRT7, PPM1D, MYRF, CPEB4, FZD7, ROBO2, SRGAP1, EIF4E, CDK4, SERPINE1, SWAP70, NEDD9, PAK4, DDX17, ERG, NRAS, ILK, CTGF, SOCS7, MDM2, ADAM17, CDH2, HDAC2, RTKN, F11R, ARL6IP5, AKR1B10, C11orf65, HLTF, GMFB, SERINC5, MEST, ALPPL2, NDRG2, DTD1, TPM3, MAP2K6, CEP19, TPRG1, GOLM1, CCDC43, MMP1, PTP4A2, TMEM9B, MMP12, MTMR14, ALDH3A1, NDUFA4, FAM3C, LYPLA2, FAM45A, PIGF, AP1G1, PHF17, NIPSNAP1, KREMEN1, MMP14, ABRACL, MIXL1, TSPAN6, PODXL, APH1A, ABHD17C, NANOG, MYO5A*
hsa-miR-150-5p	15	*MYB, EGR2, VEGFA, P2RX7, IGF2, CXCR4, ZEB1, NOTCH3, FLT3, EP300, PTPRR, MS4A3, AGA,ATP13A3, TP53*
hsa-miR-155-5p	717	*UQCRFS1, MEIS1, TAB2, MECP2, SOCS1, MSH6, MSH2, MLH1, INPP5D, DET1, SMAD5, HIVEP2, ZNF652, ZIC3, BACH1, JARID2, CSNK1A1, APC, WDFY1, VAMP3, UBE2J1, TXNRD1, TXNDC12, TRIP13, TRIM32, TRAM1, TNFRSF10A, TBCA, TACSTD2, SYPL1, SYNE2, SNAP29, SMAD1, SLC30A1, SH3BP4, SDCBP, SCAMP1, RHEB, RCOR1, RCN2, RAI14, RAB6A, RAB5C, RAB34, RAB27B, RAB23, PTPRJ, PRKCI, PRAF2, PPP5C, PPL, POLE4, POLE3, PODXL, PLXND1, PKN2, PICALM, PHC2, PDLIM5, PDE3A, NT5E, NARS, MYO1E, MYO10, MSI2, MPZL1, MOSPD2, MARC1, METTL7A, LY6K, LPL, UFL1, HSDL1, HSD17B12, GNA13, FMNL2, FADS1, DSG2, DPP7, DNAJC19, DNAJB1, DHX40, CYP51A1, CUL4B, CTNNB1, CLDN1, CHAF1A, CEBPB, CDK5RAP3, CBFB, BRPF3, BET1, ATP6V1C1, ATG3, ARL5B, ARL10, ARID2, ARFIP2, ARFIP1, ANKFY1, AMIGO2, TM6SF1, MATR3, LDOC1, PHF17, RHOA, AGTR1, PKIA, TP53INP1, IKBKE, FGF7, KDM3A, NFATC2IP, CUX1, BCAT1, SPI1, CTLA4, EDN1, FOXO3, MAFB, TSHZ3, RUNX2, JUN, IFNGR1, ZNF236, LAT2, PAPOLA, EHD1, SERTAD2, PELI1, KBTBD2, HNRNPA3P1, SLC39A10, KRAS, CAMTA1, NAMPT, CREBRF, ETS1, TLE4, FAR1, EDEM3, TWF1, C3orf58, SLC25A40, PSMG1, IKBIP, LCLAT1, VEZF1, SACM1L, DCAF7, ERMP1, KRT80, FLI1, DOCK4, CYR61, ICAM1, SELE, SMAD2, YB, SKI, GCSAM, CKAP5, IL13RA1, BCL6, MITF, MAP3K10, NOS3, UBR4, 9orf142,*

MECR, LUC7L2, TCEA1, ANAPC16, C17orf80, SLC27A2, XPC, EIF3G, EIF2B5, UBA3, EIF3E, PCYOX1, EXOC7, QPCTL, DYNC2H1, GLIPR2, HEATR2, UGT8, HSPB11, CCDC137, GLB1, LRRC40, VBP1, MGST2, GNB4, JUP, DNAJC2, GOLPH3, DRAP1, NOB1, SLC1A5, AB14, YBX3, SLC39A14, UBE2J2, EIF4A1, CAB39L, EIF2B2, CNDP2, RTN3, TMBIM6, EIF4G2, CDC42BPB, MYD88, ADH5, EIF3C, QRICH1, CYFIP1, PSIP1, SIN3A, STK24, AGL, METAP2, SPIN1, FAM98B, ERI1, COLGALT1, UGDH, KIAA1715, XPO1, PALLD, MUT, RAP1B, BAG5, LPGAT1, GCFC2, EXOSC2, LNX2, ZNF248, CHD9, MEF2A, CAB39, CLUAP1, CARD11, PCDH9, ZNF561, CARHSP1, C16orf62, LIN7C, CBR4, ECI1, OSBPL10, EIF4E2, TGM2, SLC9A3R2, CHAF1B, PPFIBP1, UBXN1, ESRRA, GAPVD1, CTAGE5, RBM42, MFF, NEU1, SCD, UBE2D2, OVCA2, EIF3CL, AURKA, LEPREL2, FUBP1, RETSAT, MPP2, KIAA0368, NES, KIF22, PACSIN2, SLC25A19, IPO8, GPT2, OGFOD1, AKR7A2, TRMT1, MCAM, INA, SNX6, PDAP1, PCNT, FIP1L1, MTAP, CEP55, AIFM1, PPM1G, PAM16, PLK1, AURKB, NASP, NUPL2, NCKAP1, EXOC2, SEC24B, RRM2, GRPEL1, ALDH9A1, AP1G1, LUZP1, RPRD1A, AA25, RP2, NUP155, FAM105B, TMX3, ERGIC1, NFYC, BE2D3, SUZ12, GLIPR1, GPM6B, LRIF1, TAF5L, HERC4, MORC3, MBNL3, UPF2, KRT6B, FLNA, DAG1, TIMM13, BUD31, CUTA, STIM1, EOGT, GEMIN5, GNPNAT1, STXBP2, PCCB, EIF3J, CHRAC1, PLAUR, CISD2, BCL7C, TICAM2, KDM1A, PSEN1, STRN, SMARCD2, RARS, RBPJ, LARS, ALDH1A2, CNOT10, UBE2H, TNKS1BP1, PUS7, RPL39, CTSA, MBLAC2, GAR1, RAD1, NOLC1, FGF2, KDELC2, HSPA4L, PEBP1, PNPLA4, GHITM, FAM3C, PGRMC2, RPAP1, NCAPG, PDE12, TFCP2, L2HGDH, PRKAR2A, ARL5A, CD109, TOMM34, CFL2, AIMP1, CDC73, TRIO, PAXBP1, TSPAN14, INTS6, YWHAZ, PRKAR1A, SSX2IP, FAM199X, RAC1, PLS1, SAP30L, MRPS27, TNFAIP2, JUNB, SPCS1, CTNNA1, ASNS, H2AFY, NR3C1, AXL, RING1, CLTA, COL4A2, CNNM3, CTNNBL1, RQCD1, ATL2, CDK5, PPP2R2A, THOC7, DNMT1, MRPL16, RAB6C, HAX1, ABHD16A, ASB6, GPAM, CECR5, WDR11, NMD3, IGF2R, ACOT7, RRAGA, INTS7, YARS, FLNB, CORO1B, SUPT5H, KANK2, DIAPH3, FASTKD1, NSA2, FMNL3, CHTOP, STRBP, MARCKS, POLR1B, FNDC3B, CD3EAP, PSAT1, COPS3, DEGS1, OXCT1, PDK1, EIF3A, GMPS, OLR1, SMAD4, CD68, FLT1, CEP41, CIAPIN1, CCDC82, ACTR2, TRAK1, CYP2U1, SLC35F2, ZNF493, HAL, IL17RB, TBC1D14, ZNF254, GABARAPL1, IGJ, RAPGEF2, WBP1L, PBRM1, ABCC4, SPECC1, MAT2B, TFPI, EXOC3, PFDN4, HK2, TYSND1, C12orf10, CRAT, PLEKHA5, PACSIN3, F5, SMARCE1, CD81, VAV2, SLC7A1, OBSCN, MAVS, DMD, CARS2, SLC12A4, MRPS34, B4GALT1, KDELC1, PDCD10, NCAPD2, UBA2, ALDH5A1, FUBP3, MYO6, NAA50, MARC2, RIOK2, OSBPL9, DDX10, ATXN10, RAB30, DEK, PHF6, ARL8B, LEMD3, ZNF207, CSE1L, CPT1A, TTC37, MAN1A2, RICTOR, IMPAD1,

		VPS4B, CLINT1, UBQLN2, RIF1, PNPT1, MRPL18, MAP3K14, ARMC2, LCORL, APAF1, MPP5, RAB11FIP2, NOVA1, RBAK, ARL15, MYO1D, LRRC59, CMSS1, LONP2, MUS81, ITGB4, DDB2, CAT, ATPAF1, TPBG, INTS4, TIMM8A, RBM22, MTFMT, WRB, DDRGK1, GMPPA, EEF1A2, RDH13, CLTC, NOTCH2, CARS, PTMS, CSNK1A1L, LUC7L3, TJP1, FDFT1, CDKN2A, AKR1C3, PNPLA8, S100A11, IER3IP1, FSTL1, SLC38A5, ATP6V1H, ITGB5, SRSF2, SLC7A11, ANXA2, DDX17, RAB2A, FOXK1, TMOD3, PSME4, ZNF384, GNAS, SNTB2, TMTC3, SLC30A7, ASPH, INTS8, CPD, EGFR, SRPK2, STAG2, TPRKB, VPS36, VCAM1, SMAD3, TTF1, FAM91A1, CCDC41, KIAA0430, CDC40, DCUN1D2, KLHL5, AGO4, HBP1, WWC1, WEE1, GOLT1B, PALD1, THBS1, DBN1, DHCR24, HSD17B7, NSUN5, TPP2, UAP1, OXNAD1, SSSCA1, EEF1E1, PHGDH, CCND1, ALDH3A2, CDK2, SGPL1, TSPAN3, ANPEP, PSME3, AGRN, GNL3L, VANGL1, RTFDC1, WNT5A, TMEM167A, CLIC4, MEST, TRIM24, CDH6, MMS22L, SKIV2L2, CCR9, DR1, RSF1, ANTXR1, SEPT11, HNRNPA3, IL8, ZNF83, PHF14, TBC1D8B, INPP5F, ARPC3, KRCC1, FAM177A1, UBTD2, SECISBP2, PAK2, SLC33A1, ZNF28, MCM8, SMARCA4, TCF12, TOMM20, UBQLN1, PDPR, MOV10, FAM120A, FAM96B, FKBP3, STAT3, NEUROG1, EPRS, CDH2, CDH13, ATP13A1, EIF3F, NUP62, GCLC, LSM3, HTRA1, CDK4, EPB41L2, COG2, CCT2, NUCKS1, PRSS21, INTS10, STX5, CIAO1, GLG1, TROVE2, ARGLU1, CALU, BRI3BP, RAD23B, PYGL, MRS2, KIF14, POLR2C, TPD52, MTHFD2, RNF2, LTN1, TMEM33, TNPO1, GSK3B, ADAM10, ERBB2IP, CCL2, NFKB1, IL6, CD36, RREB1, ZKSCAN5, ZNF611, ZNF273, VPS18, WHSC1L1, MASTL, MYBL1, GATM, E2F2, FAM135A, C3orf18, ARL6IP5
hsa-miR-155-3p	1	*IRAK3*
hsa-miR-181a-5p	237	*DUSP6, NLK, GATA6, CDX2, PRAP1, PLAG1, RALA, PTPN22, PTPN11, DUSP5, BCL2, PROX1, KAT2B, CDKN1B, ZNF763, DDIT4, ATM, HIPK2, BCL2L11, HRAS, SIRT1, FOS, MTMR3, KLF6, MCL1, XIAP, GPR78, LFNG, LRRC17, CHRFAM7A, CD46, RASSF6, FXYD6, KCTD3, TSHR, ZNF558, C8A, ARL6IP6, ZNF426, ATF7IP2, PRR4, TCF21, PHOX2A, PROSC, PTPLAD1, GSTM2, FSIP1, KBTBD3, PTPRZ1, WNT3A, TUSC1, LRRN3, TMEM45A, ARF6, C1orf109, TAF15, PLXDC2, NMRK2, WNT2, ATG10, PRDX3, ZNF652, RTEL1-TNFRSF6B, GCNT1, PCDHB8, ENAH, ZNF25, S100A1, PLA2G4C, NOL4, SIX6, FKBP10, SMCHD1, OR11A1, INCENP, LPGAT1, CLUAP1, LYSMD3, CCDC6, BAG2, GPR83, PTGS2, ANKRD13C, RLF, FBXO28, ZNF350, TIAL1, RNF34, LCLAT1, KIAA1462, ZNF35, PITPNB, SCD, H3F3B, GATAD2B, LGALSL, TGIF2, MOB1A, SLC35B4, FAM160A2, NUPL1, GPRIN3, H1F0, ARHGAP12, SPRY2, TGFBR3, TMED4, MAP2K1, PUM1, TRIM2, FBXO33, NRP1, FAM47B, CCNG1, BRMS1L, OTUD1, ATP6V0E1, WNT16, CST5, SH3BGRL, GPR137B, OFCC1, IQCG, NKX3-2, OTX2, ROPN1L, TMEM14A, TAF2, IDS, FRA10AC1, COL27A1,*

		EPHA5, DCST1, ZNF562, EYA4, CHL1, TAAR6, SLCO2A1, TMEM257, HMGB2, HERC3, BTBD3, SRPK2, DNAJC7, ANKRD1, CFI, MRPS14, HEY2, MTMR12, ACOT12, KIAA0101, USP28, AMMECR1, BPGM, DSCR8, UGT3A1, HSD17B3, GADD45G, FBXO34, C1QTNF9, KLRC4, MOB3B, FKBP7, TBX4, TMPRSS11A, SNAI2, SLC7A11, NUDT12, COPS2, ZNF12, PRLR, PLCL2, ZNF594, METAP1, HSPA13, NR6A1, YOD1, SLC37A3, FBXO11, ZNF445, TM9SF3, ATP8A1, TMEM64, MOB1B, GNAI3, TAB2, SRSF7, DDX3X, KRAS, LBR, GIGYF1, KLHL42, TMEM132B, AFTPH, ZNF148, NOTCH2, NFYB, NOTCH1, SIK2, HOOK3, FAM222B, RPS8, STAG2, SMG1, PFKFB2, ZEB2, MAZ, RPL14, KCTD2, UBA2, DDX27, FAT1, HDAC6, TMEM192, LAMA3, HUWE1, ND2, HNRNPAB, OCA2, AP1M1, UCHL1, PGD, ZFP36L2, AKAP12, PABPC1, GANAB, PHPT1, H2AFY, TEAD4, KLHL15, PRRC2B, BRCA1, SOGA2, KIAA0100, PPP1R9A, MGAT5, TNIP1
hsa-miR-181a-2-3p	1	*ARID1A*
hsa-miR-181b-5p	97	*NLK, GATA6, CDX2, CBX7, TIMP3, PLAG1, SIRT1, GRIA2, VSNL1, BCL2, KAT2B, TCL1A, MAP3K10, CYLD, TMED7, XIAP, MCL1, IGF1R, E2F1, NRD1, C20orf173, FAM5B, NDE1, KRI1, RBMX, DNAJB11, PTEN, EIF5A, KLHL7, NDC1, TMEM189, KPNB1, CAT, DTD2, USP10, LMCD1, MTOR, RPRD1A, THOC3, ZFHX3, MED14, PDHX, KPNA2, RPL18A, RPS7, UWE1, RNF187, ZNF526, HIST1H1E, EP300, INO80B, C5orf34, IFNGR1, ZXDC, MEPCE, NRBP1, ZNF503, UBB, SRRM2, NR2F1, RAP1GAP2, ARHGAP35, XPO6, ZNF747, MRPS23, SCAP, XRCC6, TRRAP, GANAB, C17orf100, RPS12, FBL, FUS, NEK3, TAF5L, RTFDC1, ZBTB20, DIDO1, ZBTB4, HIST2H4B, SEC24C, BPTF, DDHD1, ILF2, USP24, EIF4E2, OSBPL8, MYO1C, RPS8, RQCD1, FGFR1, PWP2, EIF3C, ATP6, ASB13, KIAA1524, C2orf69*
hsa-miR-181c-5p	17	*NLK, GATA6, CDX2, NOTCH4, KRAS, NOTCH2, BCL2, TRIM2, SIRT1, BTBD3, BMPR2, IL2, TUB, UBR7, TIAL1, KIT, ISCA1*
hsa-miR-181c-3p	1	*MMP25*
hsa-miR-181a-3p	2	*NANOG, HOOK3*
hsa-miR-181d-5p	25	*BCL2, HRAS, MGMT, NRD1, C20orf173, HMGB1, HSP90AB1, BORA, FCHO2, ENO1, HIST1H3D, PBX2, RPS14, PGAM1, DHRS4, CTC1, NID2, ATP6, PDHX, GAB1, RQCD1, MGRN1, C10orf35, EP300, ARID3B*
hsa-miR-224-5p	17	*KLK10, CXCR4, CDC42, RAB9B, PDGFRB, SMAD4, NCOA6, FOSB, NIT1, AP2M1, CD40, API5, EYA4, EDNRA, PEBP1, ARSB, ADNP2*
hsa-miR-342-5p	29	*UCP3, ATRX, PHF17, MCMBP, ROBO2, USP36, NUFIP2, VPS13D, EIF3D, CCNE1, AKR1B1, NFE2L1, KIF7, PTPRG, ELP3, SMARCA4, TBC1D22A, BCLAF1, ND2, TMEM109, PRRC2C, RNPS1, DDX1, MMACHC, GAPDH, PPIA, FAM65A, PSMC1, ANKRD12*
hsa-miR-342-3p	83	*GEMIN4, BMP7, DNMT1, TMEM98, IDH3B, KPNA2, PRKCE, TRMT61A, ZNF766, FIGN, FIBP, ARIH2, KCTD15, RPS3A, NCAN, SLC6A12, RPS9, ATXN7, ISOC2,*

		CRB2, PPIE, PSAT1, GJA1, CANX, NOP2, VAPB, ATF4, BPTF, RPL26, NGRN, TNS3, RPL15, FN3KRP, METTL2A, VPS35, IDS, HNRNPC, ABL1, SLC30A9, RPL27A, GOT2, PRKCSH, RRM2, RFX3, CCND2, GART, ZNF430, NAPG, GTF2H2C, FAM3C, PWP2, PTRF, ZC3H12C, ACTG1, TRERF1, CBX1, TOMM70A, ZSCAN29, NABP2, HARS, INO80D, RALGAPB, EEF1A2, PA2G4, STRN, CCDC6, EXOSC1, EPB41, ZFC3H1, XPOT, UN, SOD2, EEF2, TCP1, TCF3, TUT1, RAP1, FOSL2, APTX, SF3B3, SF3B2, TACC1, CXADR
hsa-miR-383-5p	4	*IRF1, VEGFA, IGF1R, PRDX3*
hsa-miR-629-5p	4	*HNF4A,HIST1H2AC,ATP5G2, ZCCHC6*
hsa-miR-629-3p	2	*DDX3X, ATP5B*
hsa-miR-663a	5	*JUNB, JUNB, TCEAL1, HSPG2, CDK1*
hsa-miR-663b	1	*HRASLS5*
hsa-miR-125a-5p	137	*TNFAIP3, CDKN1A, LIN28A, NTRK3, CD34, TP53, LIF, MTUS1, RARA, VEGFA, ERBB3, ERBB2, KLF13, BAK1, ARID3B, ELAVL1, CCL5, MMP11, HK2, ARPC5, E2F7, ZBTB10, CREB5, OMA1, OSTM1, MAPK8, RBM17, PUM2, NR1D2, HOXC4, SNX4, NDUFB6, NOP16, CRK, SLC1A5, PARD6B, TCEB2, PANX1, ZFAND1, CORO1C, CLDN12, ZNF776, TNPO2, EIF4EBP1, PDPK1, YES1, DUSP3, YOD1, ATXN1, TMEM101, ERMP1, GUCD1, RBAK, ATL2, OGT, PREPL, PPP2R5E, TFRC, G2E3, PLS3, TTC30B, RRP1B, COIL, NDEL1, ATXN7L3, PRC1, LIN28B, TOR1AIP2, KIAA0232, TOR4A, TPI1, PCDHGB2, UBE2G2, SPRTN, XPO1, THRAP3, OSBPL9, LBX2, SEPW1, PGM3, MACF1, PC, EEF1A1, TBC1D1, PLA2G4F, PEG10, ATP5G2, FBXO10, TDG, SP1, DLST, WNK3, RALBP1, RPS2, MRPL50, NIN, MORC2, ZZEF1, PDCL3, RPL35A, ND6, TPM4, RPL9, ENO1, CLCN7, COL4A1, NME2, GGCT, CGNL1, ZMYND19, MAT2A, NPM1, NNT, SRGAP2, EMID1, ATP5J2, STAT2, FERMT1, DIP2A, MAP3K1, SLFN11, AGTRAP, KIF6, TRMT2A, PAM, UNC45A, MAP2K7, LYPLA1, LAMB1, NCKIPSD, OTUB2, CASC4, ZDHHC9, ARHGAP12, YIPF4, IP6K1, C17orf80*
hsa-miR-125a-3p	10	*IP6K2, PFAS, TTC8, ARFGEF1, E2F8, GLUL, LEPRE1, IGF2BP2, NUP62, ATP6V1B2*
hsa-miR-146a-5p	66	*CXCR4, CFH, IRAK2, TLR2, FADD, TRAF6, IRAK1, ROCK1, BRCA2, BRCA1, TLR4, FAF1, CCNA2, PA2G4, IL8, NFKB1, CDKN1A, EGFR, MTA2, CD40LG, FAS, CDKN3, KIF22, ERBB4, SMAD4, WASF2, IRF7, PPP1R1C, ITGB2, SNTG2, RSAD2, BGLAP, SYT12, STAT1, IFIT1, IFITM1, OLFML2A, ISG15, IFITM3, RGS13, MTUS2, NMI, HSPA1A, TRIM22, ITGBL1, SPATS2L, IFI44, OASL, IFI44L, IFI27, LINC00304, STON2, BCL2A1, MX2,EPSTI1, OSBPL1A, IFIT5, C1orf21, GIMAP4, CCR9, SAMD9L, IFIT3, SPP1, SLPI, L1CAM, MSC*
hsa-miR-146a-3p	1	*PLXNA1*
hsa-miR-148a-5p	2	*GALK2, ACTR1A*
hsa-miR-148a-3p	81	*DNMT1, HLA-G, TGIF2, DNMT3B, NR1I2, RPS6KA5, CCKBR, IRS1, ACVR1, BCL2, TMED7, GPATCH8,*

		TMEM14A, ANP32A, RAB1B, HSP90B1, POFUT1, CYCS, ADARB1, CBX3, UQCRQ, SPRY2, PAN3, KANSL1, GAS1, PTPN4, ZNF92, RAB10, PAPD4, HCCS, WAPAL, MPP5, ZNF490, RAB12, GNB5, SNAPIN, PSMD9, TRIM59, DYNLL2, SECISBP2L, LYSMD1, PBXIP1, MTMR9, DNAJB4, DSTYK, LBR, KIAA1549, DYRK1A, CDK19, RAB34, ARRDC3, PRNP, HOXC8, TMEM9B, RASSF8, BTBD3, TNRC6A, SESTD1, CDC25B, MRPL45, DENR, APPBP2, SLC2A3, PTPN23, VPS41, MSL3, AMELX, OR2C3, SLC25A3, APC, GOLIM4, MYCBP2, RPS17, HSPA4, WDTC1, HMGB1, MAP3K4, USP38, NONO, CCNI, AURKB
hsa-miR-196a-5p	222	*SPRR2C, S100A9, KRT5, ANXA1, HOXB8, HOXA7, HOXD8, HOXC8, HOXB7, BACH1, HMOX1, CDKN1B, HOXA5, PDCD4, SRRT, DIEXF, ZBTB24, NR4A1, ABT1, SPATA2, NHLRC3, RPUSD2, CEP120, PHC3, ZBTB6, C9orf41, PNP, COPS3, NUP50, ZNF763, FOXJ3, BIN3, C12orf4, RAD9A, PCGF3, KATNAL1, SLC10A7, LBR, BCL11A, GLMN, STK40, TMEM135, GGA3, POLR2D, LGR4, IGF2BP3, ESPL1, RFX5, SYT9, PEX13, IGF1R, SMCR7L, SLC30A6, RAB31, KPNA5, SLC20A1, TIMM23, IGDCC4, SMARCAD1, CCNT2, CCND2, USP24, TRPC3, SMC3, TMEM2, RDH10, C11orf57, FAM127A, PRUNE2, PRYD4, TMEM161B, CPEB3, FAM104A, RAB7L1, ITGAV, CPD, ZNF354B, TMEM194A, EPHA7, KIAA1804, DFFA, TSPAN12, LIN28B, ARHGAP28, IGF2BP1, CCDC47, HAND1, LLGL1, PGAM1, PSMC3, TMX2, KLHL7, RFC2, GLUL, NRXN1, GSTK1, HIST2H4B, SPEN, TAF15, SAP18, SNRPD1, MSL3, TSKU, KMT2C, LYRM2, UBE2Z, C19orf55, LRP2, ND4L, NKX6-1, BRMS1L, RPS2, RBMX, ZFP64, REEP2, MED13, MYCBP2, TRAPPC9, DNTTIP2, VDAC3, ZNF529, EWSR1, ATP6V1B2, RDH11, APP, PROSER1, HIST1H2BB, TAB2, TUBA1B, USP19, COX3, PSMD8, TRA2B, MRPL35, KCTD1, UBE2C, ZNF581, CNOT11, POTEG, MTRF1L, SRP9, BCORL1, NRBP1, LRRC41, ATP6, FKTN, RANBP9, SYVN1, NR2F6, ATG16L1, EIF2S3, ECHDC1, HAUS6, CKAP2L, IFNGR1, ANXA7, ENAH, FOXO1, NDFIP1, LSM14A, ATP1A1, SLC25A17, ZNF398, RASSF7, SBF1, TP53RK, MED12, FLNA, CCNE2, CANX, PRPF8, FRS2, KIF18B, U2AF2, SH3GL3, VCP, ETV3, BCS1L, ND5, ZCCHC11, RALGPS2, HUWE1, ACTB, UQCRC2, EIF2B4, GMFB, PALLD, CASP3, VCL, IPO5, PATL1, NAP1L4, RUFY2, TUBB, CCND1, GID8, VDAC2, SLC6A8, FXR2, SCN11A, AHSA1, ATG9A, NOTCH2, DYRK2, ND4, OAT, TRAP1, SKI, GLTP, EEF2, LSM3, GRIK4, PDE6D, MPP2, SAR1B, MAP4, RAB21, BUB1, OGFRL1, GOT2, NRDE2, SBNO1*
hsa-miR-196a-3p	1	*RPS26*
hsa-miR-23b-5p	3	*PRODH, SPEN, EIF4B*
hsa-miR-23b-3p	173	*PRAP1, NOTCH1, E2F1, RB1, PLAU, MET, PTK2B, VHL, TAB3, TAB2, CHUK, SRC, PTEN, DHX57, CCDC88C, ACAT1, ZIC5, CEPT1, KIAA1279, G3BP1, BLOC1S2, CPEB3, ANKRD17, SPTY2D1, MED4, STAT5B, PNN, ARFIP1, AKAP11, SNRPC, PCMTD2, MCFD2, SOS1,*

	CATSPERB, RPS27A, MSH6, RBPMS2, HK1, BCL7B, MED13L, ATXN10, CDC73, FLNA, CUL4A, ZNFX1, WDR54, UQCRC2, HERC3, TNRC6B, CDK2, RPL19, ACTN4, ATP1A1, DDX47, CEP104, PPM1A, PPP1CB, CPSF7, GNS, SORCS2, LASP1, CAP1, RRAGD, TCF25, FAM168B, BCS1L, RMND5A, ILF2, GAPDH, ABCF1, CCT7, SLC9A6, DNAJC5, KDM1A, SSRP1, INPP4A, SBF1, PRPF6, GNB2L1, TCP1, ILVBL, PI4KB, AMOTL1, ATF7IP, BCL2L1, ZMYM2, GANAB, IMPDH2, GJA1, COX1, HIST2H4B, CDCA8, DIAPH1, PLOD1, HNRNPUL1, CRKL, AP4B1, RAB5B, STK11IP, PSMD14, CLUAP1, CD302, XPO1, GNA12, JAK1, PPARGC1B, CYB561D1, PRDX1, ABCC1, NHP2L1, AARS2, UBR3, RPN2, C11orf68, CPSF3L, MLLT6, DDX27, ITPR3, SYDE2, PRR14L, SUPT7L, BBS12, ATP5A1, EEF1A1, GHITM, EWSR1, EP300, PDCD6IP, SLC16A1, DMTN, CSDE1, HIF1AN, DONSON, DNMBP, KLHL2, RAF1, KANSL1, COPA, PHF10, CCNB2, LBR, DHX33, ECH1, RPL22, HNRNPK, ITFG1, UNC45A, C8orf33, SLC6A15, MAP4, RAD51, SOCS6, PSAP, WWP2, CLTA, ALDOA, HNRNPL, TPCN1, MPV17, NSD1, DGCR2, CDC20, PANK2, USP31, NMT1, TSR1, EIF4A1, E2F4, GTF2I, NPM1, EMC1, ABCC4, CMTM4

19. ADDENDUM E: The complete list of differentially-expressed genes for which miRNA targets were found in the miRTarBase database

Gene	Interacting miRNAs
ABCA1	hsa-miR-33b-5p, hsa-miR-33a-5p, hsa-miR-26a-5p, hsa-miR-26b-5p, hsa-miR-19a-3p, hsa-miR-27a-3p
ACADS	hsa-miR-26b-5p, hsa-miR-484, hsa-miR-92a-3p
ACTN2	hsa-miR-7-5p
ADAMTS5	hsa-miR-98-5p
ADAR	hsa-miR-1, hsa-miR-130b-3p, hsa-miR-93-5p, hsa-miR-1226-3p, hsa-miR-625-5p, hsa-miR-484, hsa-miR-197-3p
ADD3	hsa-miR-374b-5p, hsa-miR-124-3p, hsa-miR-24-3p
ADGB	hsa-miR-335-5p, hsa-miR-26b-5p
ADH1A	hsa-miR-132-3p
ADIPOR1	hsa-miR-197-3p
AHNAK2	hsa-miR-124-3p, hsa-miR-1, hsa-miR-16-5p
AIPL1	hsa-miR-3615
AMELX	hsa-miR-320a, hsa-miR-148a-3p
AMPD1	hsa-miR-10a-5p, hsa-miR-16-5p
ANAPC5	hsa-miR-124-3p, hsa-miR-1227-3p, hsa-miR-26b-3p, hsa-miR-34a-5p, hsa-miR-93-5p, hsa-miR-30a-5p
ANKRD26	hsa-miR-30a-5p
AP1M2	hsa-miR-124-3p
AP5S1	hsa-miR-15b-5p, hsa-miR-10a-5p
APITD1	hsa-miR-182-5p
ARHGAP22	hsa-miR-124-3p, hsa-miR-615-3p
ARID4A	hsa-miR-335-5p, hsa-miR-21-5p
ARIH1	hsa-miR-425-5p, hsa-miR-17-5p
ARL1	hsa-miR-130b-3p, hsa-miR-93-5p, hsa-miR-24-3p, hsa-miR-769-5p
ARPC5L	hsa-miR-335-5p, hsa-miR-124-3p, hsa-miR-16-5p, hsa-miR-615-3p, hsa-miR-25-3p
ASAP3	hsa-miR-335-5p
ASPN	hsa-miR-335-5p, hsa-miR-124-3p, hsa-miR-26b-5p
ATAD2	hsa-miR-106b-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-941, hsa-miR-193b-3p, hsa-miR-17-3p
ATP10A	hsa-miR-335-5p
ATP11A	hsa-miR-744-5p, hsa-miR-769-5p, hsa-miR-484, hsa-miR-197-3p, hsa-miR-92a-3p
ATP5C1	hsa-miR-30b-5p, hsa-miR-16-5p
ATP5G2	hsa-miR-629-5p, hsa-miR-185-5p, hsa-miR-149-5p, hsa-miR-125a-5p
ATP5H	hsa-miR-324-5p
ATP6V1G1	hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-98-5p
ATPAF2	hsa-miR-320a
ATXN7	hsa-miR-192-5p, hsa-miR-423-3p, hsa-miR-342-3p, hsa-miR-92a-3p, hsa-miR-17-5p
BCAP31	hsa-miR-346, hsa-miR-29a-3p
BRD8	hsa-miR-26b-5p, hsa-miR-24-3p, hsa-miR-92b-3p

C11orf30	hsa-miR-26b-5p
C4BPB	hsa-miR-142-3p, hsa-miR-124-3p
CACNA1S	hsa-miR-26b-5p
CAPN11	hsa-miR-335-5p, hsa-miR-124-3p
CASP8	hsa-miR-19b-1-5p, hsa-miR-26b-5p
CASQ1	hsa-miR-335-5p, hsa-miR-26b-5p
CBS	hsa-miR-106b-3p, hsa-miR-652-3p, hsa-miR-615-3p, hsa-miR-193b-3p, hsa-miR-324-5p, hsa-miR-149-5p, hsa-miR-92a-3p
CCDC102B	hsa-miR-124-3p
CCDC181	hsa-miR-26b-5p, hsa-miR-92a-3p
CCDC59	hsa-miR-375, hsa-miR-24-3p, hsa-miR-16-5p
CCL1	hsa-miR-17-5p
CCL17	hsa-miR-148b-3p, hsa-miR-128-3p
CCL19	hsa-miR-335-5p, hsa-miR-148b-3p, hsa-miR-9-5p
CCND1	hsa-miR-335-5p, hsa-miR-148b-3p, hsa-miR-9-5p, hsa-miR-20a-5p, hsa-miR-195-5p, hsa-miR-193b-3p, hsa-miR-424-5p, hsa-miR-16-1-3p, hsa-miR-34a-5p, hsa-miR-520b, hsa-miR-503-5p, hsa-miR-19b-1-5p, hsa-miR-16-5p, hsa-miR-15a-5p, hsa-miR-365a-3p, hsa-miR-19a-3p, hsa-miR-302a-3p, hsa-miR-15b-5p, hsa-miR-449a, hsa-miR-17-5p, hsa-miR-302c-3p, hsa-miR-106b-5p, hsa-miR-34b-3p, hsa-miR-338-3p, hsa-miR-9-5p, hsa-miR-374b-5p, hsa-miR-425-5p, hsa-miR-155-5p, hsa-miR-1, hsa-miR-183-5p, hsa-miR-93-5p, hsa-miR-26b-5p, hsa-miR-196a-5p, hsa-miR-92a-3p
CCND3	hsa-miR-424-5p, hsa-miR-16-5p, hsa-miR-34a-5p, hsa-miR-195-5p, hsa-miR-138-5p, hsa-miR-615-3p, hsa-miR-324-5p, hsa-miR-27b-3p
CCNI	hsa-miR-374b-5p, hsa-miR-26b-5p, hsa-miR-423-5p, hsa-miR-320a, hsa-miR-148a-3p, hsa-miR-92a-3p
CCR1	hsa-miR-21-5p
CCT7	hsa-miR-26b-5p, hsa-miR-877-3p, hsa-miR-615-3p, hsa-miR-598-3p, hsa-miR-505-3p, hsa-miR-23b-3p, hsa-miR-92a-3p, hsa-miR-1260b
CD177	hsa-miR-335-5p
CDH1	hsa-miR-9-5p, hsa-miR-25-3p, hsa-miR-193b-3p, hsa-miR-138-5p, hsa-miR-26b-5p, hsa-miR-30c-5p
CDH6	hsa-miR-155-5p
CDKN1A	hsa-miR-10b-5p, hsa-miR-942-5p, hsa-miR-145-5p, hsa-miR-182-5p, hsa-miR-93-5p, hsa-miR-657, hsa-miR-654-3p, hsa-miR-639, hsa-miR-572, hsa-miR-520h, hsa-miR-520b, hsa-miR-520a-3p, hsa-miR-519e-3p, hsa-miR-519d-3p, hsa-miR-519b-3p, hsa-miR-515-3p, hsa-miR-423-3p, hsa-miR-372-3p, hsa-miR-363-3p, hsa-miR-345-5p, hsa-miR-299-5p, hsa-miR-298, hsa-miR-28-5p, hsa-miR-20b-5p, hsa-miR-20a-5p, hsa-miR-208b-3p, hsa-miR-208a-3p, hsa-miR-17-5p, hsa-miR-132-3p, hsa-miR-125a-5p, hsa-miR-106b-5p, hsa-miR-503-5p, hsa-miR-106a-5p, hsa-miR-519a-3p, hsa-miR-96-5p, hsa-miR-302a-3p, hsa-miR-146a-5p, hsa-miR-146b-5p, hsa-miR-519c-3p, hsa-miR-335-5p, hsa-miR-98-5p, hsa-miR-1229-3p, hsa-miR-505-5p, hsa-miR-1260b
CDO1	hsa-miR-124-3p
CDS1	hsa-miR-335-5p, hsa-miR-324-5p
CDV3	hsa-miR-29c-3p, hsa-miR-124-3p, hsa-miR-103a-3p, hsa-miR-16-5p, hsa-miR-92a-3p
CDX4	hsa-miR-335-5p
CELF2	hsa-miR-425-5p, hsa-miR-375, hsa-miR-324-5p
CELSR3	hsa-miR-335-5p, hsa-miR-30c-5p

CENPI	hsa-miR-215-5p, hsa-miR-192-5p
CENPO	hsa-miR-218-5p
CEP112	hsa-miR-331-3p
CHAF1B	hsa-miR-193b-3p, hsa-miR-155-5p, hsa-miR-1, hsa-miR-484
CHCHD2	hsa-miR-320b, hsa-miR-1229-3p, hsa-miR-484, hsa-miR-221-3p, hsa-miR-92a-3p
CHI3L2	hsa-miR-335-5p
CHP1	hsa-miR-16-5p
CHPT1	hsa-miR-16-5p, hsa-miR-375, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-16-5p
CHST12	hsa-miR-122-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-26b-5p
CIRBP	hsa-miR-30a-5p, hsa-miR-24-3p
CLCA4	hsa-miR-335-5p, hsa-miR-128-3p
CLDN1	hsa-miR-155-5p, hsa-miR-335-5p, hsa-miR-375, hsa-miR-29b-3p
CLDN4	hsa-miR-335-5p
CLEC11A	hsa-miR-335-5p, hsa-miR-122-5p
CLEC3B	hsa-miR-335-5p
CMPK1	hsa-miR-374b-5p, hsa-miR-590-3p, hsa-miR-215-5p, hsa-miR-34a-5p, hsa-miR-10a-5p, hsa-miR-192-5p, hsa-miR-26b-5p, hsa-miR-16-5p, hsa-miR-484
CNIH3	hsa-miR-335-5p
CNP	hsa-miR-124-3p, hsa-miR-874-3p, hsa-miR-615-3p, hsa-miR-345-5p, hsa-miR-30a-5p, hsa-miR-16-5p
CNTNAP2	hsa-miR-218-5p, hsa-miR-99b-5p
COG4	hsa-miR-615-3p, hsa-miR-18a-3p
COL10A1	hsa-miR-335-5p
COL4A1	hsa-miR-29c-3p, hsa-miR-29a-3p, hsa-miR-29b-3p, hsa-miR-124-3p, hsa-miR-218-5p, hsa-miR-769-3p, hsa-miR-18a-3p, hsa-miR-193b-3p, hsa-miR-125a-5p, hsa-miR-100-5p, hsa-miR-99a-5p, hsa-miR-92a-3p, hsa-miR-16-5p
COMMD4	hsa-miR-502-3p, hsa-miR-500a-3p, hsa-miR-320a
COQ10B	hsa-miR-98-5p
COX6B1	hsa-miR-320b, hsa-miR-130b-5p, hsa-miR-320a
CPA1	hsa-miR-26b-5p
CRTC3	hsa-miR-124-3p, hsa-miR-7-5p, hsa-miR-671-5p, hsa-miR-92a-3p
CRYGC	hsa-miR-335-5p, hsa-miR-26b-5p
CTGF	hsa-miR-124-3p, hsa-miR-18a-5p, hsa-miR-26a-5p, hsa-miR-205-5p, hsa-miR-145-5p, hsa-miR-375
CTSA	hsa-miR-155-5p, hsa-miR-1296-5p, hsa-miR-503-5p, hsa-miR-130a-3p, hsa-miR-20a-5p, hsa-miR-17-5p
CTTN	hsa-miR-340-5p, hsa-miR-1, hsa-miR-34a-5p, hsa-miR-185-5p, hsa-miR-92a-3p
CYCS	hsa-miR-148a-3p, hsa-miR-16-5p, hsa-miR-320c, hsa-miR-93-3p, hsa-miR-361-5p
CYP26B1	hsa-miR-103a-3p, hsa-miR-16-5p
CYP2F1	hsa-miR-335-5p, hsa-miR-26b-5p
CYP3A4	hsa-miR-27b-3p
DCBLD2	hsa-miR-197-3p, hsa-miR-192-5p, hsa-miR-33a-5p, hsa-miR-26b-5p, hsa-miR-106a-5p, hsa-miR-17-5p
DCPS	hsa-miR-193b-3p
DDAH2	hsa-miR-34a-5p
DDX25	hsa-miR-26b-5p
DDX46	hsa-miR-93-5p, hsa-miR-21-5p, hsa-miR-423-3p, hsa-miR-99a-5p

DDX56	hsa-miR-34a-5p, hsa-miR-423-3p, hsa-miR-331-3p, hsa-miR-328-3p, hsa-miR-92a-3p
DDX60	hsa-miR-1, hsa-miR-26b-5p
DEFA4	hsa-miR-335-5p
DKK1	hsa-miR-29a-3p, hsa-miR-31-5p, hsa-miR-1
DNAJC12	hsa-miR-26b-5p
DRAP1	hsa-miR-155-5p, hsa-miR-1, hsa-miR-423-5p
DRD4	hsa-miR-124-3p
DSG1	hsa-miR-148b-3p
DSN1	hsa-miR-193b-3p, hsa-miR-215-5p, hsa-miR-192-5p
DTNB	hsa-miR-26b-5p
DTX3	hsa-miR-320a
DYNLT1	hsa-miR-326
EIF1	hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-193b-3p, hsa-miR-186-5p, hsa-miR-10b-5p, hsa-miR-10a-5p
EIF1AY	hsa-miR-26b-5p
EIF4A2	hsa-miR-21-5p, hsa-miR-744-5p, hsa-miR-331-3p
ELOVL2	hsa-miR-10a-5p, hsa-miR-26b-5p, hsa-miR-17-3p
ELOVL4	hsa-miR-21-5p, hsa-miR-19b-3p, hsa-miR-30b-5p, hsa-miR-30c-5p
ENPEP	hsa-miR-320a
EPHA7	hsa-miR-196a-5p, hsa-miR-93-5p, hsa-miR-455-3p, hsa-miR-615-3p, hsa-miR-331-3p, hsa-miR-125b-5p, hsa-miR-92a-3p
EPPIN	hsa-miR-26b-5p
ERBB4	hsa-miR-372-3p, hsa-miR-19a-3p, hsa-miR-302d-3p, hsa-miR-146a-5p
FABP3	hsa-miR-1
FABP6	hsa-miR-335-5p
FAM102A	hsa-miR-335-5p, hsa-miR-122-5p, hsa-miR-1, hsa-miR-296-3p, hsa-miR-320a
FAM135A	hsa-miR-335-5p, hsa-miR-155-5p, hsa-miR-26b-5p, hsa-miR-92a-3p
FAM160B2	hsa-miR-484, hsa-miR-92a-3p
FAM189A2	hsa-miR-215-5p, hsa-miR-192-5p
FAM3C	hsa-miR-21-5p, hsa-miR-375, hsa-miR-155-5p, hsa-miR-145-5p, hsa-miR-34a-5p, hsa-miR-1227-3p, hsa-miR-1226-3p, hsa-miR-769-3p, hsa-miR-92b-3p, hsa-miR-342-3p, hsa-miR-93-5p
FAM46A	hsa-miR-9-5p, hsa-miR-21-5p, hsa-miR-19b-3p, hsa-miR-484 hsa-miR-346, hsa-miR-34a-5p
FAT4	hsa-miR-132-3p
FERMT1	hsa-miR-9-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-125a-5p
FGF20	hsa-miR-433-3p
FGFR2	hsa-miR-19b-1-5p, hsa-miR-125b-5p, hsa-miR-1, hsa-miR-186-5p
FGG	hsa-miR-409-3p, hsa-miR-29a-3p, hsa-miR-29b-3p, hsa-miR-144-3p, hsa-miR-29c-3p
FHL1	hsa-miR-335-5p
FLG	hsa-miR-124-3p
FNDC4	hsa-miR-124-3p, hsa-miR-7-5p
FOLR1	hsa-miR-375, hsa-miR-1
FOXD3	hsa-miR-335-5p
FRMPD4	hsa-miR-185-5p
FRY	hsa-miR-132-3p
FSCN1	hsa-miR-133b, hsa-miR-133a-3p, hsa-miR-145-5p, hsa-miR-143-3p, hsa-miR-30a-5p, hsa-miR-221-3p, hsa-miR-17-3p
FUT4	hsa-miR-199a-3p

FZD5	hsa-miR-26b-5p, hsa-miR-24-3p, hsa-miR-149-5p
FZD7	hsa-miR-145-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-940
GABBR2	hsa-miR-124-3p
GAPDHS	hsa-miR-335-5p
GATA1	hsa-miR-335-5p
GBX2	hsa-miR-335-5p
GCNT1	hsa-miR-132-3p, hsa-miR-181a-5p
GFOD1	hsa-miR-130b-3p, hsa-miR-9-5p, hsa-miR-215-5p, hsa-miR-26b-5p, hsa-miR-24-3p, hsa-miR-19b-3p, hsa-miR-1276, hsa-miR-484, hsa-miR-100-5p
GHITM	hsa-miR-155-5p, hsa-miR-374b-5p, hsa-miR-484, hsa-miR-23b-3p, hsa-miR-92a-3p, hsa-miR-31-5p
GIP	hsa-miR-335-5p, hsa-miR-9-5p
GLRX5	hsa-miR-222-3p, hsa-miR-18a-5p
GNAS	hsa-miR-155-5p, hsa-miR-320b, hsa-miR-615-3p, hsa-miR-18a-3p, hsa-miR-324-3p, hsa-miR-331-3p, hsa-miR-320a, hsa-miR-17-5p
GNGT1	hsa-miR-335-5p
GNPTAB	hsa-miR-98-5p, hsa-miR-615-3p
GNRHR	hsa-miR-124-3p, hsa-miR-1, hsa-miR-26b-5p
GP9	hsa-miR-335-5p
GPLD1	hsa-miR-505-3p
GPR135	hsa-miR-26b-5p
GPR143	hsa-miR-148b-3p
GREB1	hsa-miR-335-5p, hsa-miR-26b-5p
GREM1	hsa-miR-124-3p
GRIK2	hsa-miR-26b-5p
GULP1	hsa-miR-335-5p, hsa-miR-215-5p, hsa-miR-192-5p
GUSB	hsa-miR-26b-5p, hsa-miR-760
H3F3B	hsa-miR-16-5p, hsa-miR-15a-5p, hsa-miR-1, hsa-miR-181a-5p, hsa-miR-34a-5p, hsa-miR-32-5p, hsa-miR-22-3p, hsa-miR-1236-3p, hsa-miR-93-3p, hsa-miR-652-3p, hsa-miR-324-5p
HAX1	hsa-miR-155-5p, hsa-miR-331-3p
HBE1	hsa-miR-335-5p
HECA	hsa-miR-26b-5p
HECTD4	hsa-miR-335-5p, hsa-miR-296-3p, hsa-miR-149-5p
HERC5	hsa-miR-1
HERC6	hsa-miR-16-5p, hsa-miR-15a-5p
HEY2	hsa-miR-148b-3p, hsa-miR-128-3p, hsa-miR-181a-5p
HGD	hsa-miR-335-5p, hsa-miR-26b-5p
HIST1H2AC	hsa-miR-34a-5p, hsa-miR-98-5p, hsa-miR-629-5p, hsa-miR-504-5p, hsa-miR-330-3p, hsa-miR-221-3p, hsa-miR-92a-3p
HIST1H2BD	hsa-miR-98-5p, hsa-miR-744-5p, hsa-miR-455-3p, hsa-miR-296-3p, hsa-miR-652-3p, hsa-miR-615-3p, hsa-miR-378a-3p, hsa-miR-3140-3p
HIST2H2BE	hsa-miR-34a-5p, hsa-miR-103a-3p, hsa-miR-16-5p, hsa-miR-877-3p, hsa-miR-423-5p, hsa-miR-296-3p, hsa-miR-615-3p
HLA-E	hsa-miR-296-3p, hsa-miR-92a-3p
HNRNPA2B1	hsa-miR-103a-3p, hsa-miR-16-5p, hsa-miR-1296-5p, hsa-miR-877-3p, hsa-miR-615-3p, hsa-miR-191-5p, hsa-miR-27b-3p, hsa-miR-30c-5p, hsa-miR-197-3p, hsa-miR-92a-3p, hsa-miR-1260b
HNRNPK	hsa-miR-450a-5p, hsa-miR-21-5p, hsa-miR-877-3p, hsa-miR-106b-3p, hsa-miR-615-3p, hsa-miR-503-5p, hsa-miR-484, hsa-miR-328-3p, hsa-miR-125b-5p, hsa-miR-23b-3p, hsa-miR-15b-5p, hsa-miR-30c-5p, hsa-

	miR-1260b
HNRNPR	hsa-miR-18a-5p, hsa-miR-940, hsa-miR-130b-5p, hsa-miR-615-3p
HOMER2	hsa-miR-375, hsa-miR-103a-2-5p
HOXA7	hsa-miR-196a-5p, hsa-miR-423-5p, hsa-miR-31-5p
HOXB9	hsa-miR-215-5p, hsa-miR-192-5p
HOXD13	hsa-miR-103a-3p, hsa-miR-423-3p, hsa-miR-149-5p, hsa-miR-10a-5p, hsa-miR-16-5p, hsa-miR-1260b
HOXD3	hsa-miR-17-3p
HRH1	hsa-miR-192-5p, hsa-miR-124-3p
HSD17B7P2	hsa-miR-335-5p
HSD3B1	hsa-miR-26b-5p
HSPA4L	hsa-miR-155-5p, hsa-miR-16-5p
HTR2C	hsa-miR-22-3p
ICK	hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-101-3p
IDI1	hsa-miR-193b-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-26b-5p
IFI27	hsa-miR-146a-5p
IFI44	hsa-miR-146a-5p, hsa-miR-1, hsa-miR-26b-5p
IFI44L	hsa-miR-146a-5p, hsa-miR-124-3p
IFI6	hsa-miR-375
IFIT1	hsa-miR-375, hsa-miR-146a-5p, hsa-miR-1
IFIT3	hsa-miR-146a-5p, hsa-miR-124-3p, hsa-miR-1, hsa-miR-92b-3p
IFNA1	hsa-miR-122-5p
IGFBP5	hsa-miR-140-5p, hsa-miR-331-3p, hsa-miR-17-5p
IGLC2	hsa-miR-335-5p
IGSF1	hsa-miR-425-5p, hsa-miR-577, hsa-miR-148b-3p, hsa-miR-10b-5p, hsa-miR-10a-5p, hsa-miR-92a-3p, hsa-miR-16-5p
IK	hsa-miR-34a-5p, hsa-miR-92b-3p, hsa-miR-92a-3p
IL11	hsa-miR-204-5p, hsa-miR-211-5p, hsa-miR-379-5p, hsa-miR-124-3p, hsa-miR-1
IL15	hsa-miR-215-5p, hsa-miR-192-5p
IL1A	hsa-miR-191-5p, hsa-miR-335-5p
IL3	hsa-miR-26b-5p
IL37	hsa-miR-335-5p
IL5	hsa-miR-335-5p, hsa-miR-454-5p
IMPAD1	hsa-miR-155-5p, hsa-miR-16-5p
IMPG1	hsa-miR-26b-5p
IRAK1	hsa-miR-146a-5p, hsa-miR-146b-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-1226-3p, hsa-miR-744-5p, hsa-miR-330-5p, hsa-miR-93-3p, hsa-miR-193b-3p, hsa-miR-423-3p, hsa-miR-328-3p, hsa-miR-132-3p, hsa-miR-222-3p, hsa-miR-92a-3p, hsa-miR-17-5p
IRAK4	hsa-miR-215-5p, hsa-miR-192-5p
IRF9	hsa-miR-106b-5p
ISG15	hsa-miR-146a-5p, hsa-miR-1
ISG20	hsa-miR-335-5p, hsa-miR-1
ISOC1	hsa-miR-652-3p
ITGA1	hsa-miR-335-5p, hsa-miR-1
ITPKB	hsa-miR-186-5p, hsa-miR-183-5p
IVD	hsa-miR-124-3p, hsa-miR-503-5p, hsa-miR-484
JMJD6	hsa-miR-484, hsa-miR-92a-3p, hsa-miR-1260b
KALRN	hsa-miR-335-5p
KAT2A	hsa-miR-124-3p, hsa-miR-1, hsa-miR-615-3p, hsa-miR-331-3p, hsa-miR-183-5p, hsa-miR-92a-3p, hsa-miR-17-5p

KAT2B	hsa-miR-181a-5p, hsa-miR-181b-5p, hsa-miR-25-3p, hsa-miR-32-5p, hsa-miR-92a-3p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-19a-3p, hsa-miR-19b-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-26b-5p
KCND1	hsa-miR-1, hsa-miR-26b-5p
KCNJ16	hsa-miR-148b-3p
KCNK5	hsa-miR-335-5p
KDM5A	hsa-miR-335-5p
KIAA1644	hsa-miR-335-5p, hsa-miR-26b-5p, hsa-miR-1226-3p
KLC1	hsa-miR-197-3p, hsa-miR-34a-5p
KLHL23	hsa-miR-128-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-101-3p, hsa-miR-24-3p, hsa-miR-183-5p, hsa-miR-10a-5p
KLK3	hsa-miR-335-5p
KLK7	hsa-miR-335-5p
KNG1	hsa-miR-128-3p
KRT7	hsa-miR-133a-3p, hsa-miR-199a-3p, hsa-miR-125b-5p, hsa-miR-195-5p, hsa-miR-145-5p, hsa-miR-30a-3p, hsa-miR-7-5p
KXD1	hsa-miR-505-5p, hsa-miR-193b-3p, hsa-miR-378a-5p, hsa-miR-10a-5p
LAIR2	hsa-miR-335-5p
LAMB1	hsa-miR-335-5p, hsa-miR-16-5p, hsa-miR-760, hsa-miR-500a-3p, hsa-miR-193b-3p, hsa-miR-125a-5p
LCAT	hsa-miR-26b-5p
LGR5	hsa-miR-124-3p
LLGL2	hsa-miR-335-5p
LMO7	hsa-miR-98-5p, hsa-miR-16-5p
LOXL2	hsa-miR-192-5p
LRCH3	hsa-miR-320a
LRRC1	hsa-miR-124-3p, hsa-miR-22-3p
LY6E	hsa-miR-877-3p, hsa-miR-744-5p
MACF1	hsa-miR-335-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-16-5p, hsa-miR-296-3p, hsa-miR-615-3p, hsa-miR-125a-5p, hsa-miR-222-3p, hsa-miR-30d-5p, hsa-miR-92a-3p
MAD2L1BP	hsa-miR-151a-5p, hsa-miR-615-3p
MAFB	hsa-miR-130a-3p, hsa-miR-155-5p, hsa-miR-135b-5p, hsa-miR-335-5p
MAP2	hsa-miR-335-5p
MAPK6	hsa-miR-128-3p, hsa-miR-100-5p, hsa-miR-99a-5p, hsa-miR-15a-5p
MATR3	hsa-miR-155-5p, hsa-miR-200b-3p, hsa-miR-1, hsa-miR-7-5p, hsa-miR-26b-5p, hsa-miR-24-3p, hsa-miR-744-5p, hsa-miR-615-3p, hsa-miR-186-5p, hsa-miR-185-5p, hsa-miR-149-5p, hsa-miR-125b-5p, hsa-miR-30b-5p, hsa-miR-30c-5p, hsa-miR-18a-5p
MBNL2	hsa-miR-372-3p, hsa-miR-302d-3p, hsa-miR-218-5p, hsa-miR-335-5p, hsa-miR-101-3p, hsa-miR-98-5p
MCM6	hsa-miR-193b-3p, hsa-miR-1, hsa-miR-215-5p, hsa-miR-34a-5p, hsa-miR-192-5p, hsa-miR-1180-3p
MCPH1	hsa-miR-335-5p, hsa-miR-15b-5p, hsa-miR-10a-5p
METTL2B	hsa-miR-375, hsa-miR-484, hsa-miR-92a-3p
MGST3	hsa-miR-26b-5p
MISP	hsa-miR-335-5p, hsa-miR-124-3p
MLLT4	hsa-miR-335-5p, hsa-miR-124-3p, hsa-miR-346, hsa-miR-15b-5p, hsa-miR-92a-3p
MMP1	hsa-miR-222-3p, hsa-miR-145-5p
MMP16	hsa-miR-146b-5p, hsa-miR-31-5p
MPP5	hsa-miR-375, hsa-miR-130b-3p, hsa-miR-155-5p, hsa-miR-7-5p, hsa-miR-

	148a-3p, hsa-miR-192-5p, hsa-miR-21-5p, hsa-miR-93-3p
MROH9	hsa-miR-335-5p
MRPS18B	hsa-miR-26b-5p, hsa-miR-615-3p, hsa-miR-484
MRPS2	hsa-miR-98-5p, hsa-miR-26b-5p, hsa-miR-16-5p, hsa-miR-744-5p
MSLN	hsa-miR-335-5p
MUC13	hsa-miR-148b-3p, hsa-miR-132-3p, hsa-miR-128-3p
MYBL2	hsa-miR-149-3p, hsa-miR-423-5p, hsa-miR-505-3p, hsa-miR-92a-3p
MYEF2	hsa-miR-26b-5p, hsa-miR-21-5p, hsa-miR-125b-5p, hsa-miR-10a-5p, hsa-miR-92a-3p
MYF6	hsa-miR-9-5p
MYH1	hsa-miR-23a-3p
MYH10	hsa-miR-1274a, hsa-miR-106b-3p, hsa-miR-92b-3p, hsa-miR-505-3p, hsa-miR-197-3p, hsa-miR-26a-5p, hsa-miR-23a-3p
MYLIP	hsa-miR-20b-5p, hsa-miR-106a-5p, hsa-miR-19b-3p, hsa-miR-92a-3p, hsa-miR-340-5p, hsa-miR-93-5p, hsa-miR-186-5p
MYLPF	hsa-miR-335-5p
NAP1L4	hsa-miR-196a-5p
NBEAL2	hsa-miR-93-5p
NCAPD2	hsa-miR-193b-3p, hsa-miR-155-5p, hsa-miR-16-5p, hsa-miR-1303, hsa-miR-93-3p, hsa-miR-615-3p, hsa-miR-92b-3p, hsa-miR-484, hsa-miR-331-3p, hsa-miR-186-5p, hsa-miR-222-3p, hsa-miR-3620-3p
NDUFA7	hsa-miR-186-5p, hsa-miR-92a-3p
NDUFA9	hsa-miR-16-5p, hsa-miR-100-5p
NDUFB1	hsa-miR-26b-5p
NDUFS1	hsa-miR-615-3p, hsa-miR-324-5p, hsa-miR-221-3p
NDUFS5	hsa-miR-1296-5p
NEBL	hsa-miR-125b-5p
NEIL1	hsa-miR-335-5p
NETO2	hsa-miR-1, hsa-miR-375, hsa-miR-24-3p, hsa-miR-21-5p, hsa-miR-20a-5p
NEUROD1	hsa-miR-30a-5p
NFKB2	hsa-miR-193b-3p, hsa-miR-98-5p, hsa-miR-423-5p
NKG7	hsa-miR-335-5p
NLGN4X	hsa-miR-148b-3p, hsa-miR-26b-5p
NNMT	hsa-miR-124-3p, hsa-miR-98-5p
NOL10	hsa-miR-148b-3p, hsa-miR-124-3p, hsa-miR-935
NOX4	hsa-miR-148b-3p
NPPB	hsa-miR-375
NPR3	hsa-miR-16-5p, hsa-miR-30a-5p, hsa-miR-335-5p, hsa-miR-98-5p
NR4A3	hsa-miR-124-3p, hsa-miR-10b-5p, hsa-miR-26b-5p
NRIP1	hsa-miR-590-3p, hsa-miR-24-3p, hsa-miR-615-3p, hsa-miR-15b-5p
NTF3	hsa-miR-200c-3p, hsa-miR-335-5p
NYX	hsa-miR-26b-5p
OAS1	hsa-miR-335-5p
OAS2	hsa-miR-335-5p, hsa-miR-132-3p, hsa-miR-7-5p
OASL	hsa-miR-146a-5p, hsa-miR-124-3p, hsa-miR-1
OCLN	hsa-miR-335-5p
OLFML2A	hsa-miR-335-5p, hsa-miR-146a-5p
OSMR	hsa-miR-335-5p, hsa-miR-98-5p
OTUB2	hsa-miR-335-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-125a-5p
P2RX5	hsa-miR-98-5p
PAPD7	hsa-miR-423-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-32-5p, hsa-miR-106b-3p, hsa-miR-93-5p
PARP12	hsa-miR-98-5p

PARP16	hsa-miR-124-3p, hsa-miR-590-3p, hsa-miR-193b-3p, hsa-miR-215-5p, hsa-miR-192-5p
PCNP	hsa-miR-18a-5p, hsa-miR-186-5p
PCSK5	hsa-miR-107
PCSK7	hsa-miR-652-3p
PCYT1A	hsa-miR-26a-5p
PDE3A	hsa-miR-155-5p, hsa-miR-128-3p
PDLIM1	hsa-miR-26b-5p, hsa-miR-504-5p, hsa-miR-320a, hsa-miR-92a-3p
PGR	hsa-miR-126-3p
PHACTR2	hsa-miR-124-3p, hsa-miR-21-5p, hsa-miR-378a-3p, hsa-miR-185-5p
PHF14	hsa-miR-155-5p, hsa-miR-128-3p, hsa-miR-122-5p, hsa-miR-183-5p
PIP4K2C	hsa-miR-193b-3p, hsa-miR-877-3p, hsa-miR-125b-5p, hsa-miR-101-3p, hsa-miR-32-5p, hsa-miR-26b-5p, hsa-miR-193b-3p
PIR	hsa-miR-1
PLAT	hsa-miR-204-5p, hsa-miR-21-5p, hsa-miR-335-5p
PLCXD1	hsa-miR-98-5p, hsa-miR-296-3p
PLOD2	hsa-miR-221-3p, hsa-miR-98-5p, hsa-miR-26b-5p, hsa-miR-425-5p, hsa-miR-615-3p, hsa-miR-183-5p, hsa-miR-1260b
PLSCR1	hsa-miR-132-3p, hsa-miR-18a-3p, hsa-miR-10a-5p, hsa-miR-99a-5p
PML	hsa-miR-335-5p, hsa-miR-423-3p, hsa-miR-378a-3p
POLR2J	hsa-miR-124-3p
PPARGC1A	hsa-miR-23a-3p, hsa-miR-335-5p, hsa-miR-98-5p, hsa-miR-421, hsa-miR-425-3p, hsa-miR-378a-5p
PPFIBP1	hsa-miR-335-5p, hsa-miR-155-5p
PPIC	hsa-miR-26b-5p
PPP2R2A	hsa-miR-31-5p, hsa-miR-222-3p, hsa-miR-155-5p, hsa-miR-1, hsa-miR-30a-5p, hsa-miR-26b-5p, hsa-miR-92a-3p
PQLC3	hsa-miR-9-5p, hsa-miR-124-3p
PRLR	hsa-miR-181a-5p
PRR11	hsa-miR-128-3p, hsa-miR-122-5p
PSG3	hsa-miR-124-3p, hsa-miR-1
PSMA6	hsa-miR-30e-5p
PTGDS	hsa-miR-335-5p
PTGES	hsa-miR-335-5p, hsa-miR-124-3p, hsa-miR-125b-5p
PTHLH	hsa-miR-33a-5p
PTPN21	hsa-miR-615-3p
RAB1A	hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-19b-3p, hsa-miR-16-5p
RAB40A	hsa-miR-335-5p
RAB40B	hsa-miR-204-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-26b-5p
RAB5A	hsa-miR-101-3p, hsa-miR-130a-3p, hsa-miR-130b-3p, hsa-miR-33a-5p, hsa-miR-18a-5p
RABGGTA	hsa-miR-26b-5p
RAD54L2	hsa-miR-26b-5p
RAI2	hsa-miR-186-3p
RANBP9	hsa-miR-101-3p, hsa-miR-196a-5p
RASAL2	hsa-miR-98-5p, hsa-miR-26b-5p, hsa-miR-15a-3p, hsa-miR-125b-5p, hsa-miR-99a-5p, hsa-miR-92a-3p
RBL1	hsa-miR-17-5p, hsa-miR-20a-5p, hsa-miR-106b-5p
RBM22	hsa-miR-155-5p
RELN	hsa-miR-128-3p, hsa-miR-138-5p, hsa-miR-10a-5p
RIBC2	hsa-miR-124-3p
RIF1	hsa-miR-155-5p, hsa-miR-103a-3p, hsa-miR-16-5p, hsa-miR-421, hsa-miR-186-5p, hsa-miR-183-5p, hsa-miR-93-5p, hsa-miR-24-3p

RNF10	hsa-miR-877-3p, hsa-miR-130b-5p, hsa-miR-484, hsa-miR-320a, hsa-miR-222-3p
RPA1	hsa-miR-1301-3p, hsa-miR-615-3p, hsa-miR-339-5p, hsa-miR-31-5p
RPL11	hsa-miR-505-5p, hsa-miR-92a-3p
RPL18A	hsa-miR-744-5p, hsa-miR-423-5p, hsa-miR-421, hsa-miR-652-3p, hsa-miR-503-5p, hsa-miR-301a-3p, hsa-miR-106b-5p, hsa-miR-186-5p, hsa-miR-149-5p, hsa-miR-181b-5p, hsa-miR-92a-3p, hsa-miR-20a-5p
RPL37	hsa-miR-744-5p, hsa-miR-324-5p, hsa-miR-191-5p, hsa-miR-34a-5p, hsa-miR-17-5p
RPS27A	hsa-miR-30b-5p, hsa-miR-23b-3p, hsa-miR-30c-5p, hsa-miR-26a-5p, hsa-miR-17-5p
RPS27L	hsa-miR-484
RPSA	hsa-miR-16-5p, hsa-miR-1227-3p, hsa-miR-769-3p, hsa-miR-615-3p, hsa-miR-132-3p, hsa-miR-107, hsa-miR-25-3p, hsa-miR-22-3p, hsa-miR-17-5p, hsa-miR-3200-3p
RSU1	hsa-miR-34a-5p, hsa-miR-7-5p
RUFY3	hsa-miR-124-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-21-5p, hsa-miR-505-3p
RUVBL2	hsa-miR-1229-3p, hsa-miR-324-3p, hsa-miR-92a-3p
RXFP3	hsa-miR-335-5p, hsa-miR-26b-5p
S100A7	hsa-miR-26b-5p
S100A7A	hsa-miR-26b-5p
SAMD9	hsa-miR-30a-5p
SATB2	hsa-miR-31-5p, hsa-miR-34a-5p, hsa-miR-192-5p, hsa-miR-320b, hsa-miR-505-5p, hsa-miR-320a, hsa-miR-149-5p
SCAF8	hsa-miR-877-3p, hsa-miR-29b-3p, hsa-miR-92a-3p
SCARF1	hsa-miR-548q
SCLY	hsa-miR-375, hsa-miR-193b-3p, hsa-miR-92a-3p
SCN1A	hsa-miR-93-3p
SCO2	hsa-miR-26b-5p
SEC61G	hsa-miR-26b-5p, hsa-miR-1260b
SEC62	hsa-miR-30a-5p, hsa-miR-1
SEMA3D	hsa-miR-193a-3p
SEPT8	hsa-miR-324-5p
SF1	hsa-miR-186-5p, hsa-miR-935, hsa-miR-320a, hsa-miR-221-3p
SFRP1	hsa-miR-335-5p
SH3BP5	hsa-miR-130b-3p, hsa-miR-103a-3p, hsa-miR-193b-3p
SH3PXD2A	hsa-miR-335-5p, hsa-miR-130b-3p, hsa-miR-124-3p, hsa-miR-93-5p, hsa-miR-33a-5p, hsa-miR-24-3p, hsa-miR-331-3p
SHPK	hsa-miR-124-3p, hsa-miR-615-3p
SIGLEC8	hsa-miR-335-5p
SIK2	hsa-miR-1229-3p, hsa-miR-125b-5p, hsa-miR-181a-5p
SLC15A1	hsa-miR-92b-3p
SLC16A1	hsa-miR-376a-5p, hsa-miR-124-3p, hsa-miR-484, hsa-miR-376a-3p, hsa-miR-128-3p, hsa-miR-615-3p, hsa-miR-320a, hsa-miR-23b-3p
SLC25A36	hsa-miR-124-3p, hsa-miR-222-3p
SLC29A2	hsa-miR-93-5p
SLC2A10	hsa-miR-335-5p
SLC35F2	hsa-miR-155-5p, hsa-miR-98-5p
SLC35F6	hsa-miR-373-3p, hsa-miR-1226-3p
SLC38A4	hsa-miR-335-5p
SLC7A2	hsa-miR-9-5p, hsa-miR-192-5p, hsa-miR-101-3p, hsa-miR-24-3p, hsa-miR-423-5p, hsa-miR-30c-2-3p, hsa-miR-615-3p

SLC9A2	hsa-miR-16-5p
SLC9A7	hsa-miR-335-5p
SMAD4	hsa-miR-224-5p, hsa-miR-26a-5p, hsa-miR-483-3p, hsa-miR-18a-5p, hsa-miR-17-5p, hsa-miR-19a-3p, hsa-miR-20a-5p, hsa-miR-92a-3p, hsa-miR-146a-5p, hsa-miR-199a-5p, hsa-miR-130a-3p, hsa-miR-301a-3p, hsa-miR-454-3p, hsa-miR-182-5p, hsa-miR-130b-3p, hsa-miR-155-5p, hsa-miR-186-5p, hsa-miR-93-5p, hsa-miR-19b-3p
SMARCE1	hsa-miR-155-5p, hsa-miR-103a-3p, hsa-miR-26b-5p, hsa-miR-328-3p, hsa-miR-30c-5p, hsa-miR-1260b
SMG5	hsa-miR-708-5p, hsa-miR-130b-5p, hsa-miR-106b-3p, hsa-miR-193b-3p, hsa-miR-331-3p
SNCA	hsa-miR-7-5p, hsa-miR-153-3p
SNX16	hsa-miR-124-3p, hsa-miR-103a-3p, hsa-miR-96-5p, hsa-miR-16-5p
SPTBN5	hsa-miR-335-5p
SSPN	hsa-miR-335-5p
ST6GALNAC5	hsa-miR-148b-3p
STIM1	hsa-miR-155-5p, hsa-miR-769-3p, hsa-miR-17-3p
STK10	hsa-miR-484
STOM	hsa-miR-124-3p
STX11	hsa-miR-373-3p
SULF1	hsa-miR-516a-3p, hsa-miR-124-3p
SULT1E1	hsa-miR-199a-5p, hsa-miR-124-3p
SYNDIG1	hsa-miR-17-5p
SYNE2	hsa-miR-155-5p, hsa-miR-193b-3p, hsa-miR-148b-3p, hsa-miR-9-5p, hsa-miR-1, hsa-miR-34a-5p, hsa-miR-26b-5p, hsa-miR-21-5p, hsa-miR-484
TAAR2	hsa-miR-335-5p
TAB2	hsa-miR-155-5p, hsa-miR-23b-3p, hsa-miR-181a-5p, hsa-miR-98-5p, hsa-miR-15b-5p, hsa-miR-196a-5p
TATDN2	hsa-miR-423-3p, hsa-miR-34a-5p, hsa-miR-615-3p, hsa-miR-222-3p
TBCEL	hsa-miR-92a-3p
TBX3	hsa-miR-128-3p, hsa-miR-22-3p, hsa-miR-1226-3p
TCF20	hsa-miR-149-5p
TCF21	hsa-miR-21-5p, hsa-miR-181a-5p
TCL1B	hsa-miR-9-5p, hsa-miR-7-5p
TCP11	hsa-miR-122-5p
TCTN1	hsa-miR-324-3p
TDRD7	hsa-miR-26b-5p
TF	hsa-miR-335-5p
TFDP1	hsa-miR-935
TFPT	hsa-miR-26b-5p
THAP10	hsa-miR-26b-5p, hsa-miR-92b-3p
THEM6	hsa-miR-124-3p, hsa-miR-215-5p, hsa-miR-192-5p
TLL1	hsa-miR-335-5p
TLR3	hsa-miR-124-3p
TLX1	hsa-miR-340-5p, hsa-miR-26b-5p
TM4SF1	hsa-miR-128-3p
TMA16	hsa-miR-92a-3p
TMCO3	hsa-miR-26b-5p, hsa-miR-30c-5p
TMEM59	hsa-miR-30a-5p, hsa-miR-149-5p, hsa-miR-1260b
TMOD2	hsa-miR-335-5p, hsa-miR-34a-5p, hsa-miR-484
TNFAIP3	hsa-miR-125a-5p, hsa-miR-125b-5p, hsa-miR-29a-3p, hsa-miR-21-5p, hsa-miR-26b-5p, hsa-miR-24-3p
TNFSF15	hsa-miR-26b-5p

TNNT2	hsa-miR-335-5p
TPX2	hsa-miR-193b-3p
TRAF2B	hsa-miR-10b-5p, hsa-miR-10a-5p, hsa-miR-1, hsa-miR-378a-3p, hsa-miR-30c-5p, hsa-miR-196a-5p, hsa-miR-17-5p
TRAF3IP1	hsa-miR-877-3p, hsa-miR-92a-3p, hsa-miR-25-3p
TRPC6	hsa-miR-124-3p
TRPM4	hsa-miR-335-5p, hsa-miR-1
TSC22D2	hsa-miR-375, hsa-miR-98-5p, hsa-miR-26b-5p, hsa-miR-1227-3p, hsa-miR-320a, hsa-miR-221-3p
TSPAN12	hsa-miR-196a-5p, hsa-miR-101-3p, hsa-miR-26b-5p
TTF2	hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-16-5p, hsa-miR-183-5p
TUSC3	hsa-miR-128-3p, hsa-miR-26b-5p, hsa-miR-1229-3p, hsa-miR-25-3p
TXNL1	hsa-miR-26a-5p, hsa-miR-16-5p
UBE2Q1	hsa-miR-338-3p, hsa-miR-103a-3p, hsa-miR-16-5p, hsa-miR-1226-3p, hsa-miR-93-3p, hsa-miR-186-5p
UBL5	hsa-miR-18a-5p, hsa-miR-877-5p, hsa-miR-324-5p
UBP1	hsa-miR-183-5p, hsa-miR-34a-5p, hsa-miR-182-5p
UGP2	hsa-miR-16-5p, hsa-miR-15a-5p, hsa-miR-193b-3p, hsa-miR-32-5p, hsa-miR-106b-5p
UGT8	hsa-miR-148b-3p, hsa-miR-155-5p, hsa-miR-1, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-16-5p, hsa-miR-1260b
UMPS	hsa-miR-197-3p, hsa-miR-124-3p
UNC5C	hsa-miR-335-5p
UQCR11	hsa-miR-106b-3p, hsa-miR-615-3p
USP18	hsa-miR-221-3p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-26b-5p
VAMP8	hsa-miR-193b-3p, hsa-miR-124-3p, hsa-miR-7-5p, hsa-miR-16-5p
VAV2	hsa-miR-155-5p, hsa-miR-877-3p, hsa-miR-193b-3p, hsa-miR-320a
WIF1	hsa-miR-335-5p
WIPF2	hsa-miR-92a-3p
WNT3	hsa-miR-335-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-1301-3p
WNT7B	hsa-miR-335-5p
XAF1	hsa-miR-335-5p
YWHAZ	hsa-miR-193b-3p, hsa-miR-375, hsa-miR-155-5p, hsa-miR-1, hsa-miR-30a-5p, hsa-miR-1229-3p, hsa-miR-877-3p, hsa-miR-455-3p, hsa-miR-423-5p, hsa-miR-615-3p, hsa-miR-484, hsa-miR-340-3p, hsa-miR-378a-3p
ZBTB24	hsa-miR-151a-3p, hsa-miR-196a-5p, hsa-miR-98-5p
ZBTB25	hsa-miR-26b-5p
ZBTB3	hsa-miR-93-5p
ZFP37	hsa-miR-935, hsa-miR-320a
ZFPM2	hsa-miR-141-3p, hsa-miR-200a-3p, hsa-miR-200b-3p, hsa-miR-200c-3p, hsa-miR-429, hsa-miR-26b-5p
ZFY	hsa-miR-26b-5p
ZNF135	hsa-miR-335-5p
ZNF273	hsa-miR-155-5p, hsa-miR-215-5p, hsa-miR-192-5p
ZNF286A	hsa-miR-331-3p
ZNF3	hsa-miR-193b-3p
ZNF318	hsa-miR-744-5p, hsa-miR-320a, hsa-miR-149-5p, hsa-miR-10b-5p, hsa-miR-10a-5p, hsa-miR-197-3p
ZNF408	hsa-miR-339-5p
ZNF451	hsa-miR-335-5p, hsa-miR-124-3p
ZNF518A	hsa-miR-335-5p

ZNF551	hsa-miR-26b-5p, hsa-miR-92a-3p
ZNF587	hsa-miR-21-5p, hsa-miR-106b-5p, hsa-miR-548aa
ZNF606	hsa-miR-103a-3p
ZNF646	hsa-miR-378a-5p, hsa-miR-320a, hsa-miR-30b-5p
ZNF667	hsa-miR-21-5p, hsa-miR-345-5p
ZNF706	hsa-miR-106b-5p, hsa-miR-93-5p, hsa-miR-20a-5p
ZNF80	hsa-miR-186-5p
ZPBP	hsa-miR-335-5p

20. ADDENDUM F: A complete list of the miRNAs interacting with 47 lupus-associated genes curated from literature, and verified in Ensembl BioMart

Gene	Number of interacting miRNAs	Interacting miRNAs
ACTN4	21	hsa-miR-1, hsa-miR-935, hsa-miR-744-5p, hsa-miR-296-3p, hsa-miR-769-5p, hsa-miR-652-3p, hsa-miR-615-3p, hsa-miR-92b-3p, hsa-miR-193b-3p, hsa-miR-339-5p, hsa-miR-324-3p, hsa-miR-324-5p, hsa-miR-378a-5p, hsa-miR-320a, hsa-miR-185-5p, hsa-miR-23b-3p, hsa-miR-182-5p, hsa-miR-96-5p, hsa-miR-26a-5p, hsa-miR-23a-3p, hsa-miR-16-5p
API5	2	hsa-miR-224-5p, hsa-miR-183-5p
ATF6B	1	hsa-miR-197-3p
C1QA	1	hsa-miR-335-5p
C1QB	1	hsa-miR-26b-5p
C2	2	hsa-miR-335-5p, hsa-miR-142-3p
CD44	15	hsa-miR-708-5p, hsa-miR-328-3p, hsa-miR-373-3p, hsa-miR-199a-3p, hsa-miR-520c-3p, hsa-miR-34a-5p, hsa-miR-608, hsa-miR-330-3p, hsa-miR-216a-5p, hsa-miR-1, hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-744-5p, hsa-miR-320a, hsa-miR-15b-5p
DNAJA1	3	hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-193b-3p
ETS1	10	hsa-miR-125b-5p, hsa-miR-9-5p, hsa-miR-222-3p, hsa-miR-200b-3p, hsa-miR-31-5p, hsa-miR-155-5p, hsa-miR-193b-3p, hsa-miR-208a-3p, hsa-miR-10a-5p, hsa-miR-30c-5p
FCGR1B	1	hsa-miR-26b-5p
FRK	1	hsa-miR-335-5p
IKZF1	1	hsa-miR-26b-5p
IL10	1	hsa-miR-106a-5p
IRAK1	15	hsa-miR-146a-5p, hsa-miR-146b-5p, hsa-miR-215-5p, hsa-miR-192-5p, hsa-miR-1226-3p, hsa-miR-744-5p, hsa-miR-330-5p, hsa-miR-93-3p, hsa-miR-193b-3p, hsa-miR-423-3p, hsa-miR-328-3p, hsa-miR-132-3p, hsa-miR-222-3p, hsa-miR-92a-3p, hsa-miR-17-5p
IRF5	1	hsa-miR-22-3p
JAZF1	2	hsa-miR-31-5p, hsa-miR-96-5p
KPNA1	5	hsa-miR-34a-5p, hsa-miR-103a-3p, hsa-miR-30a-5p, hsa-miR-16-5p, hsa-miR-186-5p
MECP2	13	hsa-miR-802, hsa-miR-155-5p, hsa-miR-199a-3p, hsa-miR-195-5p, hsa-miR-19a-3p, hsa-miR-122-5p, hsa-miR-212-3p, hsa-miR-199a-5p, hsa-miR-124-3p, hsa-miR-148b-3p,hsa-miR-425-3p, hsa-miR-331-3p, hsa-miR-92a-3p
NMNAT2	3	hsa-miR-9-5p, hsa-miR-122-5p, hsa-miR-615-3p
NOTCH4	3	hsa-miR-34c-5p, hsa-miR-181c-5p, hsa-miR-18a-3p
PHRF1	4	hsa-miR-1229-3p, hsa-miR-92b-3p, hsa-miR-193b-3p, hsa-

		miR-222-3p
PPP2CA	2	hsa-miR-340-3p, hsa-miR-197-3p
PRDM1	5	hsa-miR-9-5p, hsa-miR-127-3p, hsa-miR-125b-5p, hsa-miR-30a-5p, hsa-miR-877-3p
PTPN22	2	hsa-miR-181a-5p, hsa-miR-1
PTPRD	6	hsa-miR-429, hsa-miR-200a-3p, hsa-miR-200c-3p, hsa-miR-141-3p, hsa-miR-1, hsa-miR-200b-3p
PTTG1	5	hsa-miR-423-5p, hsa-miR-320a, hsa-miR-186-5p, hsa-miR-26a-5p, hsa-miR-17-5p
PXK	1	hsa-miR-744-5p
RABGAP1L	1	hsa-miR-1
RASGRP3	1	hsa-miR-21-5p
RPS6KA1	2	hsa-miR-125b-5p, hsa-miR-193b-3p
SERPIND1	1	hsa-miR-335-5p
SKIV2L	1	hsa-miR-222-3p
SLC15A4	3	hsa-miR-124-3p, hsa-miR-151a-5p, hsa-miR-186-5p
SPP1	3	hsa-miR-299-5p, hsa-miR-335-5p, hsa-miR-146a-5p
STAT1	4	hsa-miR-145-5p, hsa-miR-146a-5p, hsa-miR-34a-5p, hsa-miR-615-3p
SYK	3	hsa-miR-99b-3p, hsa-miR-615-3p, hsa-miR-331-3p
TNFAIP3	6	hsa-miR-125a-5p, hsa-miR-125b-5p, hsa-miR-29a-3p, hsa-miR-21-5p, hsa-miR-26b-5p, hsa-miR-24-3p
TNIP1	3	hsa-miR-221-3p, hsa-miR-324-3p, hsa-miR-181a-5p
TRAF6	2	hsa-miR-146a-5p, hsa-miR-146b-5p
TRIM27	2	hsa-miR-421, hsa-miR-1260b
TYK2	1	hsa-miR-124-3p
UBE2L3	3	hsa-miR-122-5p, hsa-miR-423-5p, hsa-miR-93-5p
UHRF1BP1	3	hsa-miR-34a-5p, hsa-miR-192-5p, hsa-miR-652-3p
VDR	2	hsa-miR-125b-5p, hsa-miR-27b-3p
WDFY4	1	hsa-miR-335-5p
XKR6	1	hsa-miR-335-5p
XRCC1	4	hsa-miR-34a-5p, hsa-miR-92b-3p, hsa-miR-193b-3p, hsa-miR-186-5p

21. ADDENDUM G: Diseases and Functions which were associated with the Core Networks generated by the All Top Regulators (Genes and miRNAs, and associated targets) dataset

Categories	Diseases or Functions Annotation	p-Value	Molecules	# Molecules
Cancer, Endocrine System Disorders, Gastrointestinal Disease, Organismal Injury and Abnormalities	pancreatic cancer	1.85E-03	*CCND1*,*CCR1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC),PLAT, *SMAD4*,*SPP1*,*SYNE2*,*YWHAZ*	19
Cancer, Endocrine System Disorders, Gastrointestinal Disease, Organismal Injury and Abnormalities	ductal pancreatic carcinoma	6.86E-10	*CDKN1A*, ,*SMAD4*,*SPP1* miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC)	10
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	gastrointestinal tract cancer	1.52E-03	*ARID4A*,*ATXN7*,*CCND1*,*CDH6*,*CDKN1A*,*CLDN1*, *DCBLD2*,*DRAP1*,*EIF4A2*,*ERBB4*,*ETS1*,*FAM135A*, *FAM46A*,*GNAS*,*H3F3A/H3F3B*,*HAX1*,*HOXA7*,*IFI27*, *IGFBP5*,*KAT2B*,*MAFB*,*MATR3*,miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU),	58

			miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-423-5p (and other miRNAs w/seed GAGGGGC), *MYBL2*,*MYEF2*,*NCAPD2*,*NFKB2*,*PDE3A*,*PHACTR2*, *PLAT*,*PPFIBP1*,*PTTG1*,*RBL1*,*RBM22*,*RIF1*,*RPL18A*, *SIK2*,*SMAD4*,*SPP1*,*STAT1*,*SYNDIG1*,*SYNE2*,*TCF21*, *TNFAIP3*,*TRA2B*,*UGT8*,*VAV2*,*YWHAZ*,*ZNF667*	
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	colorectal cancer	5.08E-03	*ARID4A*,*ATXN7*,*CCND1*,*CDH6*,*CDKN1A*,*CLDN1*, *DRAP1*,*EIF4A2*,*ERBB4*,*ETS1*,*FAM135A*,*FAM46A*, *GNAS*,*IFI27*,*IGFBP5*,*KAT2B*,*MAFB*,*MATR3*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-423-5p (and other miRNAs w/seed GAGGGGC), *MYBL2*,*MYEF2*,*NCAPD2*,*NFKB2*,*PDE3A*,*PHACTR2*, *PLAT*,*PTTG1*,*RBL1*,*RBM22*,*RIF1*,*SIK2*,*SMAD4*,*SPP1*, *STAT1*,*SYNE2*,*TCF21*,*TNFAIP3*,*UGT8*,*VAV2*,*YWHAZ*	49
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	upper gastrointestinal tract cancer	2.73E-06	*ARID4A*,*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*FAM135A*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HOXA7*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU),	32

			miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*PDE3A*,*PLAT*,*PPFIBP1*,*PTTG1*,*RIF1*,*RPL18A*, *SIK2*,*SMAD4*,*SPP1*,*SYNDIG1*,*SYNE2*,*TRA2B*,*VAV2*,*ZNF667*	
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	upper gastrointestinal tract tumor	5.42E-06	*ARID4A*,*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*FAM135A*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HOXA7*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*PDE3A*,*PLAT*,*PPFIBP1*,*PTTG1*,*RIF1*,*RPL18A*, *SIK2*,*SMAD4*,*SPP1*,*SYNDIG1*,*SYNE2*,*TRA2B*,*VAV2*, *ZNF667*	32
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	gastroesophageal cancer	2.33E-05	*ARID4A*,*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*FAM135A*, *GNAS*,*HOXA7*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),*MYBL2*,*PDE3A*,*PLAT*,*PPFIBP1*,*PTTG1*, *RIF1*,*RPL18A*,*SIK2*,*SMAD4*,*SPP1*,*SYNDIG1*,*SYNE2*, *TRA2B*,*VAV2*,*ZNF667*	28
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	gastro-esophageal carcinoma	2.77E-04	*ARID4A*,*DCBLD2*,*ERBB4*,*FAM135A*,*GNAS*,*HOXA7*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),	24

			MYBL2,*PDE3A*,*PLAT*,*PPFIBP1*,*PTTG1*,*RIF1*,*RPL18A*, *SIK2*,*SMAD4*,*SPP1*,*SYNDIG1*,*SYNE2*,*TRA2B*,*ZNF667*	
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	gastroesophageal adenocarcinoma	3.90E-03	*ARID4A*,*DCBLD2*,*ERBB4*,*FAM135A*,*GNAS*,*HOXA7*, miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*PDE3A*,*PLAT*,*PPFIBP1*,*RIF1*,*RPL18A*,*SIK2*, *SMAD4*,*SYNDIG1*,*SYNE2*,*TRA2B*,*ZNF667*	20
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	gastric cancer	5.56E-03	*ARID4A*,*DCBLD2*,*ERBB4*,*GNAS*,*HOXA7*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),*MYBL2*,*PDE3A*,*PLAT*,*PPFIBP1*,*RIF1*, *SIK2*,*SMAD4*,*SPP1*,*SYNDIG1*,*SYNE2*,*ZNF667*	19
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	esophageal cancer	4.08E-06	*CCND1*,*CDKN1A*,*FAM135A*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PDE3A*,*PTTG1*,*RPL18A*,*SMAD4*,*SPP1*,*SYNE2*, *TRA2B*,*VAV2*	14
Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities	esophageal carcinoma	1.15E-04	*FAM135A*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PDE3A*,*PTTG1*,*RPL18A*,*SMAD4*,*SPP1*,*SYNE2*,*TRA2B*	11
Cancer, Hematological Disease, Immunological Disease, Organismal Injury and Abnormalities	leukemia	4.65E-03	*ARIH1*,*CCND1*,*CDKN1A*,*GNAS*,*HEY2*,*IFI27*,*IFI44L*, *IFIT1*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), *RIF1*,*RPL37*,*SMAD4*,*SYNE2*,*TNIP1*	19

Cancer, Hematological Disease, Immunological Disease, Organismal Injury and Abnormalities	lymphocytic cancer	4.72E-04	*CCND1*,*CDKN1A*,*ETS1*,*IRAK1*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *TNFAIP3*,*TNIP1*,*YWHAZ*	12
Cancer, Hematological Disease, Immunological Disease, Organismal Injury and Abnormalities	lymphoproliferative malignancy	1.29E-03	*CCND1*,*CDKN1A*,*IFI27*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *RIF1*,*SYNE2*,*TNFAIP3*	12
Cancer, Hematological Disease, Immunological Disease, Organismal Injury and Abnormalities	non-Hodgkin's disease	7.71E-04	*CCND1*,*CDKN1A*,*ETS1*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *TNFAIP3*,*TNIP1*,*YWHAZ*	11
Cancer, Hematological Disease, Organismal Injury and Abnormalities	hematological neoplasia	4.63E-05	*ARIH1*,*CCND1*,*CDKN1A*,*ETS1*,*GNAS*,*HEY2*, *IFI27*,*IFI44L*,*IFIT1*,*IFIT3*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *PDE3A*,*RIF1*,*RPL37*,*RPS27A*,*SMAD4*,*SPP1*,*STIM1*,	30

			SYNE2,*TNFAIP3*,*TNIP1*,*YWHAZ*	
Cancer, Hematological Disease, Organismal Injury and Abnormalities	hematologic cancer	4.10E-04	*ARIH1*,*CCND1*,*CDKN1A*,*ETS1*,*GNAS*,*HEY2*,*IFI27*, *IFI44L*,*IFIT1*,*IFIT3*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *PDE3A*,*RIF1*,*RPL37*,*SMAD4*,*STIM1*,*SYNE2*,*TNFAIP3*, *TNIP1*,*YWHAZ*	27
Cancer, Hematological Disease, Organismal Injury and Abnormalities	myeloproliferative disorder	1.24E-03	*ARIH1*,*CCND1*,*GNAS*,*HEY2*,*IFI44L*,*IFIT1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), *PDE3A*,*RPL37*,*RPS27A*,*SMAD4*,*SYNE2*,*TNIP1*	17
Cancer, Immunological Disease, Organismal Injury and Abnormalities	lymphatic node tumor	2.44E-04	*CCND1*,*CDKN1A*,*ETS1*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *SPP1*,*TNFAIP3*,*TNIP1*,*YWHAZ*	12
Cancer, Neurological Disease, Organismal Injury and Abnormalities	central nervous system tumor	8.58E-04	*CDKN1A*,*ERBB4*,*FAM3C*,*GNAS*,*H3F3A/H3F3B*, *IGFBP5*,*IRAK1*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*SMAD4*,*SMARCE1*,*SPP1*,*UGT8*,*YWHAZ*	16

Cancer, Neurological Disease, Organismal Injury and Abnormalities	gliomatosis	4.39E-03	*CDKN1A*,*ERBB4*,*FAM3C*,*GNAS*,*H3F3A/H3F3B*, *IGFBP5*, miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*SMAD4*,*SPP1*,*UGT8*,*YWHAZ*	13
Cancer, Neurological Disease, Organismal Injury and Abnormalities	central nervous system cancer	7.72E-03	*CDKN1A*,*ERBB4*,*FAM3C*,*GNAS*,*H3F3A/H3F3B*, *IGFBP5*,*IRAK1*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*SMAD4*,*UGT8*	13
Cancer, Neurological Disease, Organismal Injury and Abnormalities	astrocytoma	3.44E-03	*CDKN1A*,*ERBB4*,*FAM3C*,*GNAS*,*H3F3A/H3F3B*,*IGFBP5*, miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*SMAD4*,*UGT8*,*YWHAZ*	12
Cancer, Organismal Injury and Abnormalities	tumorigenesis of tissue	7.26E-05	*ARID4A*,*ARIH1*,*ATXN7*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*, *CHAF1B*,*CLDN1*,*CTSA*,*DCBLD2*,*DRAP1*,*EIF4A2*, *ELOVL4*,*ERBB4*,*ETS1*,*FAM135A*,*FAM46A*,*GHITM*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HEY2*,*HIST2H2BE*, *HNRNPK*,*HOXA7*,*HSPA4L*,*IFI27*,*IFI44*,*IFI44L*,*IFIT1*, *IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2A*,*KAT2B*,*MAFB*,*MATR3*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG),	93

			MPP5,*MYBL2*,*MYEF2*,*NCAPD2*,*NETO2*,*NFKB2*,*OASL*, *OLFML2A*,*PDE3A*,*PHACTR2*,*PLAT*,*PPFIBP1*,*PPP2R2A*, *PRLR*,*PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RBM22*,*RIF1*, *RPL18A*,*RPL37*,*RPS27A*,*RUFY3*,*SIK2*,*SLC7A2*,*SMAD4*, *SPP1*,*STAT1*,*STIM1*,*SYNDIG1*,*SYNE2*,*TAB2*,*TCF21*, *TNFAIP3*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*YWHAZ*, *ZNF273*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities	epithelial cancer	4.43E-05	*ARID4A*,*ARIH1*,*ATXN7*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*, *CHAF1B*,*CLDN1*,*CTSA*,*DCBLD2*,*DRAP1*,*EIF4A2*,*ELOVL4*, *ERBB4*,*ETS1*,*FAM135A*,*FAM46A*,*GHITM*,*GNAS*, *H3F3A/H3F3B*,*HAX1*,*HIST2H2BE*,*HNRNPK*,*HOXA7*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2A*, *KAT2B*,*MAFB*,*MATR3*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*MYEF2*,*NCAPD2*,*NETO2*,*NFKB2*,*OASL*, *OLFML2A*,*PDE3A*,*PHACTR2*,*PLAT*,*PPFIBP1*,*PPP2R2A*, *PRLR*,*PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RBM22*,*RIF1*, *RPL18A*,*RPL37*,*RPS27A*,*RUFY3*,*SIK2*,*SLC7A2*,*SMAD4*, *SPP1*,*STAT1*,*STIM1*,*SYNDIG1*,*SYNE2*,*TAB2*,*TCF21*, *TNFAIP3*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*YWHAZ*, *ZNF273*,*ZNF667*	92

Cancer, Organismal Injury and Abnormalities	abdominal cancer	3.94E-03	*ARID4A*,*ARIH1*,*ATXN7*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*, *CHAF1B*,*CLDN1*,*CTSA*,*DCBLD2*,*DRAP1*,*EIF4A2*,*ERBB4*, *ETS1*,*FAM135A*,*FAM46A*,*GHITM*,*GNAS*,*H3F3A/H3F3B*, *HNRNPK*,*HOXA7*,*HSPA4L*,*IFI27*,*IFI44*,*IFI44L*,*IFIT1*,*IFIT3*, *IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*,*MAFB*,*MATR3*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), miR-423-5p (and other miRNAs w/seed GAGGGGC), *MPP5*,*MYBL2*,*MYEF2*,*NCAPD2*,*NETO2*,*NFKB2*,*OASL*, *OLFML2A*,*PDE3A*,*PHACTR2*,*PLAT*,*PPFIBP1*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RBM22*,*RIF1*,*RPL37*, *RUFY3*,*SIK2*,*SLC7A2*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*, *SYNDIG1*,*SYNE2*,*TAB2*,*TCF21*,*TNFAIP3*,*TNIP1*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*YWHAZ*,*ZNF273*,*ZNF667*	86
Cancer, Organismal Injury and Abnormalities	adenocarcinoma	2.13E-03	*ARID4A*,*ARIH1*,*ATXN7*,*CCND1*,*CCR1*,*CDKN1A*, *CHAF1B*,*CLDN1*,*CTSA*,*DCBLD2*,*DRAP1*,*EIF4A2*, *ELOVL4*,*ERBB4*,*ETS1*,*FAM135A*,*FAM46A*,*GHITM*, *GNAS*,*HOXA7*,*HSPA4L*,*IFI27*,*IFI44*,*IFI44L*,*IFIT3*, *IGFBP5*,*KAT2B*,*MATR3*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU),	70

			miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*MYEF2*,*NCAPD2*,*NETO2*,*NFKB2*, *OLFML2A*,*PDE3A*,*PHACTR2*,*PLAT*,*PPFIBP1*,*PRLR*, *PTPN22*,*RASGRP3*,*RBL1*,*RBM22*,*RIF1*,*RPL18A*,*SIK2*, *SLC7A2*,*SMAD4*,*SPP1*,*STAT1*,*SYNDIG1*,*SYNE2*, *TNFAIP3*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*YWHAZ*, *ZNF273*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities	abdominal carcinoma	6.39E-03	*ARID4A*,*ARIH1*,*ATXN7*,*CCND1*,*CCR1*,*CDKN1A*, *CHAF1B*,*CLDN1*,*CTSA*,*DCBLD2*,*DRAP1*,*EIF4A2*, *ERBB4*,*ETS1*,*FAM135A*,*FAM46A*,*GHITM*,*GNAS*, *HOXA7*,*HSPA4L*,*IFI27*,*IFI44*,*IFIT3*,*KAT2B*,*MATR3*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NCAPD2*,*NETO2*,*NFKB2*,*OLFML2A*, *PDE3A*,*PHACTR2*,*PLAT*,*PPFIBP1*,*PRLR*,*PTPN22*, *RASGRP3*,*RBL1*,*RBM22*,*RIF1*,*SIK2*,*SLC7A2*,*SMAD4*, *SPP1*,*STAT1*,*SYNDIG1*,*SYNE2*,*TNFAIP3*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*YWHAZ*,*ZNF273*,*ZNF667*	64

Cancer, Organismal Injury and Abnormalities	urogenital cancer	1.97E-06	*ARID4A*,*ARIH1*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*, *CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*,*ERBB4*,*FAM135A*, *GHITM*,*GNAS*,*HNRNPK*,*HSPA4L*,*IFI27*,*IFI44*,*IFI44L*, *IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PDE3A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*, *SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*, *TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF273*,*ZNF667*	61
Cancer, Organismal Injury and Abnormalities	pelvic cancer	2.51E-05	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG),	57

			MPP5,*MYBL2*,*NETO2*,*NFKB2*,*OLFML2A*,*PDE3A*, *PLAT*,*PRLR*,*PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*, *SIK2*,*SLC7A2*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*, *TCF21*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities	breast or colorectal cancer	2.62E-03	*ARID4A*,*ATXN7*,*CCND1*,*CDH6*,*CDKN1A*,*CLDN1*, *DRAP1*,*EIF4A2*,*ERBB4*,*ETS1*,*FAM135A*,*FAM46A*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HEY2*,*HIST2H2BE*, *IFI27*,*IFIT1*,*IGFBP5*,*KAT2B*,*MAFB*,*MATR3*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-423-5p (and other miRNAs w/seed GAGGGGC), *MYBL2*,*MYEF2*,*NCAPD2*,*NFKB2*,*PDE3A*,*PHACTR2*, *PLAT*,*PRLR*,*PTTG1*,*RBL1*,*RBM22*,*RIF1*,*SIK2*,*SMAD4*, *SPP1*,*STAT1*,*SYNE2*,*TCF21*,*TNFAIP3*,*TRA2B*,*UGT8*, *VAV2*,*YWHAZ*	56
Cancer, Organismal Injury and Abnormalities	endometrioid carcinoma	9.96E-04	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*, *ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*,*IFIT3*, *KAT2B*,miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*,*SPP1*,*STAT1*,*SYNE2*, *TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	36

Cancer, Organismal Injury and Abnormalities	breast or ovarian cancer	1.01E-04	*CCND1*,*CDKN1A*,*CLDN1*,*EIF4A2*,*ERBB4*,*ETS1*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HEY2*,*HIST2H2BE*,*IFI27*, *IFIT1*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*PLAT*,*PRLR*,*PTTG1*,*SIK2*,*SMAD4*,*SPP1*, *STAT1*,*TRA2B*,*YWHAZ*	30
Cancer, Organismal Injury and Abnormalities	lymphohematopoietic cancer	2.02E-04	*ARIH1*,*CCND1*,*CDKN1A*,*ETS1*,*GNAS*,*HEY2*,*IFI27*, *IFI44L*,*IFIT1*,*IFIT3*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *PDE3A*,*RIF1*,*RPL37*,*SMAD4*,*SPP1*,*STIM1*,*SYNE2*, *TNFAIP3*,*TNIP1*,*YWHAZ*	28
Cancer, Organismal Injury and Abnormalities	cancer of secretory structure	1.35E-03	*CCND1*,*CCR1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*,*IFI44L*, *IRAK1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU),	28

			miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *MPP5*,*MYBL2*,*PLAT*,*RBL1*,*SMAD4*,*SPP1*,*STAT1*, *STIM1*,*SYNE2*,*YWHAZ*	
Cancer, Organismal Injury and Abnormalities	Lymphoid Cancer and Tumors	1.46E-04	*ARIH1*,*CCND1*,*CDKN1A*,*ETS1*,*GNAS*,*HEY2*,*IFI27*, *IFI44L*,*IFIT1*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *PDE3A*,*RIF1*,*RPL37*,*RPS27A*,*SMAD4*,*SPP1*,*SYNE2*, *TNFAIP3*,*TNIP1*,*YWHAZ*	27
Cancer, Organismal Injury and Abnormalities	head and neck neoplasia	1.01E-02	*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*ETS1*,*FAM135A*, *FAM3C*,*FAM46A*,*GNAS*,*H3F3A/H3F3B*,*HAX1*, *HNRNPK*,*HOXA7*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU),*PTTG1*,*SMAD4*,*SPP1*,*STAT1*,*TNFAIP3*	27
Cancer, Organismal Injury and Abnormalities	head and neck cancer	1.44E-06	*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*ETS1*,*FAM3C*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*HNRNPK*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU),	24

			miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU),*PTTG1*,*SMAD4*,*SPP1*,*STAT1*,*TNFAIP3*	
Cancer, Organismal Injury and Abnormalities	lymphoid cancer	1.18E-03	*ARIH1*,*CCND1*,*CDKN1A*,*ETS1*,*GNAS*,*HEY2*,*IFI27*, *IFI44L*,*IFIT1*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC*)*,*RIF1*,*RPL37*,*SMAD4*,*SPP1*,*SYNE2*, *TNFAIP3*,*TNIP1*,*YWHAZ*	24
Cancer, Organismal Injury and Abnormalities	malignant neoplasm of aerodigestive tract	6.50E-10	*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*FAM135A*, *GNAS*,*H3F3A/H3F3B*,*HAX1*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PDE3A*,*PTTG1*,*RPL18A*,*SMAD4*,*SPP1*,*SYNE2*, *TRA2B*,*VAV2*	23
Cancer, Organismal Injury and Abnormalities	squamous-cell carcinoma	3.60E-08	*CCND1*,*CDKN1A*,*DCBLD2*,*EIF4A2*,*ERBB4*,*HAX1*, *HIST2H2BE*,*IFI27*,*IFI44*,*IFIT1*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU),	23

			miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-298 (miRNAs w/seed GCAGAAG),*PTTG1*, *SMAD4*,*SPP1*	
Cancer, Organismal Injury and Abnormalities	advanced malignant tumor	1.74E-10	*CCND1*,*CDH6*,*ERBB4*,*IGFBP5*,*KAT2B*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*PLAT*,*PTTG1*,*RBL1*,*RPS27A*,*SMAD4*,*SPP1*, *STAT1*,*TCF21*,*YWHAZ*	22
Cancer, Organismal Injury and Abnormalities	advanced stage solid tumor	2.43E-09	*CCND1*,*CDH6*,*ERBB4*,*IGFBP5*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),*MYBL2*,*PLAT*,*PTTG1*,*RPS27A*,*SPP1*, *STAT1*,*TCF21*,*YWHAZ*	19
Cancer, Organismal Injury and Abnormalities	upper aerodigestive tract carcinoma	7.83E-08	*DCBLD2*,*ERBB4*,*FAM135A*,*GNAS*,*HAX1*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),*PDE3A*,*PTTG1*,*RPL18A*,*SMAD4*,*SPP1*, *SYNE2*,*TRA2B*	19

Cancer, Organismal Injury and Abnormalities	metastatic solid tumor	6.48E-10	*CCND1*,*CDH6*,*ERBB4*,*IGFBP5*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),*PLAT*,*PTTG1*,*RPS27A*,*SPP1*,*STAT1*, *TCF21*,*YWHAZ*	18
Cancer, Organismal Injury and Abnormalities	benign neoplasia	1.01E-05	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*,*IFI44L*, *IFIT1*,*IFIT3*,*IGFBP5*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*NETO2*,*PRLR*,*PTTG1*,*SPP1*,*STAT1*	18
Cancer, Organismal Injury and Abnormalities	head and neck carcinoma	5.37E-08	*CCND1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*ETS1*,*GNAS*, *HAX1*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC),*PTTG1*, *SMAD4*,*SPP1*	17
Cancer, Organismal Injury and Abnormalities	connective or soft tissue tumor	2.99E-05	*CCND1*,*CDKN1A*,*ERBB4*,*GNAS*,*IFI44L*,*IFIT1*,*IFIT3*, *IGFBP5*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PRLR*,*PTTG1*,*SPP1*	15

Cancer, Organismal Injury and Abnormalities	serous neoplasm	1.98E-04	*CDKN1A*,*EIF4A2*,*GNAS*,*IGFBP5*,*IRAK1*,*ISG15*, *MYBL2*,*PRLR*,*PTTG1*,*SIK2*,*SPP1*,*STAT1*,*TCF21*, *TNIP1*	14
Cancer, Organismal Injury and Abnormalities	clear-cell adenocarcinoma	4.15E-03	*ARIH1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), *NETO2*,*PRLR*,*SMAD4*,*STAT1*,*ZNF273*	12
Cancer, Organismal Injury and Abnormalities	ductal carcinoma	9.87E-07	*CCND1*,*CDKN1A*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC),SMAD4,SPP1	11
Cancer, Organismal Injury and Abnormalities	benign connective or soft tissue neoplasm	4.44E-05	*CCND1*,*ERBB4*,*GNAS*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*, miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PRLR*,*PTTG1*	11
Cancer, Organismal Injury and Abnormalities	upper aero-digestive squamous cell carcinoma	1.07E-05	*DCBLD2*,*ERBB4*,*HAX1*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*SPP1*	10
Cancer, Organismal Injury and Abnormalities	primary neoplasm	2.82E-05	*IFI27* ,miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU),	10

			PRLR,*SMAD4*,*SPP1*	
Cancer, Organismal Injury and Abnormalities	adenoma	7.81E-04	*CCND1*,*CDKN1A*,*EIF4A2*,*GNAS*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*PRLR*,*SPP1*,*STAT1*	10
Cancer, Organismal Injury and Abnormalities, Renal and Urological Disease	urinary tract tumor	6.28E-06	*ARIH1*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*,*EIF4A2*,*ERBB4*, *GNAS*,*HNRNPK*,*IFIT3*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *NETO2*,*PDE3A*,*PLAT*,*PTPN22*,*SMAD4*,*SPP1*,*SYNE2*, *TCF21*,*ZNF273*	26
Cancer, Organismal Injury and Abnormalities, Renal and Urological Disease	urinary tract cancer	3.01E-05	*ARIH1*,*CCND1*,*CCR1*,*CDH6*,*CDKN1A*,*EIF4A2*,*ERBB4*, *GNAS*,*HNRNPK*,*IFIT3*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), *NETO2*,*PDE3A*,*PLAT*,*PTPN22*,*SMAD4*, *SPP1*,*SYNE2*,*TCF21*,*ZNF273*	24
Cancer, Organismal Injury and Abnormalities, Renal and Urological Disease	Bladder Cancer and Tumors	1.31E-05	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PDE3A*,*PLAT*,*PTPN22*,*SMAD4*,*SPP1*,*SYNE2*,*TCF21*	16
Cancer, Organismal Injury and Abnormalities, Renal and Urological Disease	renal cancer	1.09E-03	*ARIH1*,*CCR1*,*CDH6*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*, *HNRNPK*,*IFIT3*,*IGFBP5*,*IRAK1*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU),*NETO2*, *SMAD4*,*ZNF273*	16

Cancer, Organismal Injury and Abnormalities, Renal and Urological Disease	bladder cancer		4.15E-04	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), *PDE3A*,*PLAT*,*PTPN22*,*SMAD4*,*SPP1*,*SYNE2*,*TCF21*	13
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	genital tumor		4.21E-06	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*,*IRAK1*, *ISG15*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*, *SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*, *TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	56
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	female genital neoplasm		1.29E-07	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*, *KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU),	55

			miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*,*SMAD4*, *SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	genital tract cancer	3.08E-06	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*,*SMAD4*, *SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	55
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	tumorigenesis of genital organ	3.57E-06	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFI44L*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC),	55

			miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG),MPP5, *MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*, *SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*, *TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	female genital tract cancer	2.20E-07	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*, *SLC7A2*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*, *TCF21*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	53
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	tumorigenesis of reproductive tract	2.40E-07	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI27*,*IFI44*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC),	53

			miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SIK2*, *SLC7A2*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*,*SYNE2*, *TCF21*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	uterine tumor	1.98E-07	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*, *ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*, *IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*, *SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	51
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	uterine cancer	2.96E-07	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2* ,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*, *IFIT3*,*IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC),	49

			miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*, *PTPN22*,*PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*, *SPP1*,*STAT1*,*STIM1*,*SYNE2*,*TCF21*,*TNIP1*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	endometrial cancer	8.38E-06	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*, *ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*,*IFIT3*, *IGFBP5*,*IRAK1*,*ISG15*,*KAT2B*, miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *PTTG1*,*RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*,*STAT1*,*SYNE2*, *TCF21*,*TNIP1*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	44
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	uterine carcinoma	2.97E-04	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*,*ERBB4*, *FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*,*IFIT3*,*KAT2B*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-298 (miRNAs w/seed GCAGAAG), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*,*STAT1*,*SYNE2*,*TRA2B*,	

			TRAF6,*UGT8*,*VAV2*,*ZNF667*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	female genital tract adenocarcinoma	1.13E-03	*ARID4A*,*CCND1*,*CDKN1A*,*CHAF1B*,*CTSA*,*DRAP1*, *EIF4A2*,*ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*, *IFI44*,*IFIT3*,*KAT2B*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *RASGRP3*,*RBL1*,*RIF1*,*SIK2*,*SLC7A2*,*SPP1*,*STAT1*, *SYNE2*,*TRA2B*,*TRAF6*,*UGT8*,*VAV2*,*ZNF667*	38
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	endometrial carcinoma	2.34E-03	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*, *ERBB4*,*FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*, *IFIT3*,*KAT2B*,miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PRLR*,*PTPN22*, *RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*,*STAT1*,*SYNE2*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	35
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	adenocarcinoma in endometrium	3.66E-03	*ARID4A*,*CCND1*,*CHAF1B*,*CTSA*,*DRAP1*,*EIF4A2*,*ERBB4*, *FAM135A*,*GHITM*,*GNAS*,*HSPA4L*,*IFI44*,*IFIT3*,*KAT2B*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MPP5*,*MYBL2*,*NETO2*,*OLFML2A*,*PLAT*,*PTPN22*, *RASGRP3*,*RBL1*,*RIF1*,*SLC7A2*,*STAT1*,*SYNE2*,*TRA2B*, *TRAF6*,*UGT8*,*VAV2*,*ZNF667*	34
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	breast cancer	7.50E-06	*CCND1*,*CDKN1A*,*CLDN1*,*EIF4A2*,*ERBB4*,*ETS1*,*GNAS*, *H3F3A/H3F3B*,*HAX1*,*HEY2*,*HIST2H2BE*,*IFI27*,*IFIT1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC),	28

			miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*PLAT*,*PRLR*,*PTTG1*,*SMAD4*,*SPP1*,*STAT1*, *TRA2B*,*YWHAZ*	
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	Prostate Cancer and Tumors	2.30E-03	*CCND1*,*CDKN1A*,*GNAS*,*IFI44L*,*IRAK1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU),*MPP5*,*MYBL2*,*RBL1*,*SPP1*,*STAT1*,*STIM1*	16
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	prostate cancer	4.95E-03	*CCND1*,*CDKN1A*,*IFI44L*,*IRAK1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU),*MPP5*,*MYBL2*,*RBL1*,*SPP1*,*STAT1*,*STIM1*	15
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	carcinoma in breast	1.11E-02	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*ETS1*,*IFI27*,*IFIT1*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*PRLR*,*SMAD4*,*SPP1*,*TRA2B*	14
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	benign neoplasm of female genital organ	1.30E-05	*CDKN1A*,*EIF4A2*,*GNAS*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*, miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PRLR*,*PTTG1*,*SPP1*,*STAT1*	13
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	female genital tract serous cancer	2.15E-05	*CDKN1A*,*EIF4A2*,*IGFBP5*,*IRAK1*,*ISG15*,*MYBL2*, *PRLR*,*PTTG1*,*SPP1*,*STAT1*,*TCF21*,*TNIP1*	12

Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	HER2 negative hormone receptor negative breast cancer		9.56E-08	*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), *MYBL2*,*PLAT*,*SPP1*,*STAT1*	10
Cancer, Organismal Injury and Abnormalities, Respiratory Disease	respiratory system tumor		8.89E-08	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*,*HIST2H2BE*, *IRAK1*,*KAT2A*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYEF2*,*PPP2R2A*,*RPS27A*,*SMAD4*,*SPP1*,*STAT1*, *SYNE2*,*YWHAZ*	26
Cancer, Organismal Injury and Abnormalities, Respiratory Disease	lung cancer		7.10E-08	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*,*HIST2H2BE*, *IRAK1*,*KAT2A*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYEF2*,*PPP2R2A*,*RPS27A*,*SMAD4*,*SPP1*,*STAT1*, *SYNE2*,*YWHAZ*	25

Cancer, Organismal Injury and Abnormalities, Respiratory Disease	non-small cell lung cancer	1.61E-05	*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*HIST2H2BE*, *KAT2A*,*MAFB*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *MYEF2*,*PPP2R2A*,*RPS27A*,*SMAD4*,*SPP1*,*STAT1*, *SYNE2*,*YWHAZ*	19
Cancer, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders	smooth muscle tumor	4.47E-05	*CDKN1A*,*ERBB4*,*IFI44L*,*IFIT1*,*IFIT3*,*IGFBP5*, miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PTTG1*,*SPP1*	10
Cell Cycle	cell cycle progression	4.77E-04	*CCND1*,*CDKN1A*,*ETS1*,*IGFBP5*,*KAT2B*,*MECP2*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *PTTG1*,*RBL1*,*SMAD4*,*SMARCE1*,*STAT1*	14
Cell Cycle	interphase of tumor cell lines	6.83E-06	*CCND1*,*CDKN1A*,*ETS1*,*IGFBP5*,*KAT2A*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*PPP2R2A*,*RBL1*,*SPP1*,*STAT1*	12
Cell Death and Survival	cell death	2.70E-06	*ATXN7*,*CCND1*,*CDKN1A*,*ELOVL4*,*ERBB4*,*ETS1*, *GNAS*,*HAX1*,*HNRNPK*,*HOXA7*,*IFI27*,*IFIT3*,*IGFBP5*, *IRAK1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU),	36

			MYBL2,*NFKB2*,*PLAT*,*PPP2R2A*,*PTTG1*,*SMAD4*, *SMARCE1*,*SPP1*,*STAT1*,*STIM1*,*TNFAIP3*,*TRAF6*, *UBE2L3*,*YWHAZ*	
Cell Death and Survival	apoptosis	1.64E-07	*CCND1*,*CDKN1A*,*ELOVL4*,*ERBB4*,*ETS1*,*GNAS*, *HAX1*,*HNRNPK*,*HOXA7*,*IFI27*,*IFIT3*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*NFKB2*,*PLAT*,*PPP2R2A*,*PTTG1*,*SMAD4*, *SMARCE1*,*SPP1*,*STAT1*,*STIM1*,*TNFAIP3*,*TRAF6*, *YWHAZ*	33
Cell Death and Survival	necrosis	7.73E-07	*ATXN7*,*CCND1*,*CDKN1A*,*ELOVL4*,*ERBB4*,*ETS1*, *GNAS*,*HAX1*,*HNRNPK*,*HOXA7*,*IGFBP5*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*NFKB2*,*PLAT*,*PPP2R2A*,*PTTG1*,*SMAD4*, *SMARCE1*,*SPP1*,*STAT1*,*STIM1*,*TNFAIP3*,*TRAF6*, *UBE2L3*,*YWHAZ*	33
Cell Death and Survival	cell death of tumor cell lines	3.04E-06	*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*,*HAX1*, *HNRNPK*,*HOXA7*,*IGFBP5*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU),	28

			miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*NFKB2*,*PPP2R2A*,*PTTG1*,*SMAD4*,*SMARCE1*, *SPP1*,*STAT1*,*STIM1*,*TNFAIP3*,*TRAF6*,*UBE2L3*,*YWHAZ*	
Cell Death and Survival	apoptosis of tumor cell lines	1.08E-06	*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*,*HAX1*, *HNRNPK*,*HOXA7*,*IGFBP5*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *NFKB2*,*PPP2R2A*,*PTTG1*,*SMAD4*,*SMARCE1*,*SPP1*, *STAT1*,*STIM1*,*TNFAIP3*,*TRAF6*,*YWHAZ*	25
Cell Death and Survival	cell viability	1.43E-03	*CCND1*,*CCR1*,*CDKN1A*,*ERBB4*,*IRAK1*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*NFKB2*,*PPP2R2A*,*PTPN22*,*RPSA*,*SPP1*, *TNFAIP3*,*UBE2L3*	16
Cell Death and Survival	cell viability of tumor cell lines	3.53E-03	*CCND1*,*CDKN1A*,*ERBB4*,*IRAK1*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *NFKB2*,*PPP2R2A*,*PTPN22*,*RPSA*,*SPP1*,*UBE2L3*	13
Cellular Development	differentiation of cells	3.95E-09	*CCND1*,*CCR1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*ETS1*, *HAX1*,*HOXA7*,*IGFBP5*,*KAT2B*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU),	23

			miR-21-5p (and other miRNAs w/seed AGCUUAU), *PLAT*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*,*TRAF6*	
Cellular Development, Cellular Growth and Proliferation	proliferation of tumor cell lines	2.64E-08	*ARID4A*,*ARIH1*,*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*, *GNAS*,*HNRNPK*,*HOXA7*,*IGFBP5*,*KAT2A*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *PLAT*,*PTPN22*,*PTTG1*,*RBL1*,*SMAD4*,*SMARCE1*, *SPP1*,*STAT1*,*STIM1*,*TAB2*,*TRAF6*,*YWHAZ*	32
Cellular Development, Cellular Growth and Proliferation	proliferation of breast cancer cell lines	8.83E-06	*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*,*HNRNPK*, *IGFBP5*,miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *SMARCE1*,*SPP1*	13
Cellular Development, Cellular Growth and Proliferation	proliferation of carcinoma cell lines	6.40E-05	*CCND1*,*CDKN1A*,*KAT2A*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *SPP1*,*TAB2*,*TRAF6*,*YWHAZ*	11
Cellular Development, Cellular Growth and Proliferation	proliferation of lung cancer cell lines	1.04E-05	*CCND1*,*CDKN1A*,*KAT2A*, miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-21-5p (and other miRNAs w/seed AGCUUAU), *SPP1*,*TAB2*,*TRAF6*,*YWHAZ*	10

Cellular Growth and Proliferation	proliferation of cells	1.00E-08	*ARID4A*,*ARIH1*,*CCND1*,*CCNI*,*CCR1*,*CDKN1A*, *DCBLD2*,*ERBB4*,*ETS1*,*GNAS*,*H3F3A/H3F3B*, *HNRNPK*,*HOXA7*,*IFIT3*,*IGFBP5*,*ISG15*,*KAT2A*,*KAT2B*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), *MYBL2*,*NFKB2*,*PLAT*,*PRLR*,*PTPN22*,*PTTG1*,*RBL1*, *RPSA*,*SMAD4*,*SMARCE1*,*SPP1*,*STAT1*,*STIM1*, *TAB2*,*TNFAIP3*,*TRAF6*,*UBE2L3*,*YWHAZ*	45
Cellular Movement	cell movement	1.64E-06	*CCL1*,*CCND1*,*CCR1*,*CDKN1A*,*DCBLD2*,*ERBB4*,*ETS1*, *GNAS*,*HAX1*,*HNRNPK*,*HOXA7*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *NFKB2*,*PLAT*,*PTTG1*,*RUFY3*,*SMAD4*,*SPP1*,*STAT1*, *STIM1*,*TAB2*,*YWHAZ*	26
Cellular Movement	migration of cells	6.23E-07	*CCL1*,*CCND1*,*CCR1*,*CDKN1A*,*DCBLD2*,*ETS1*, *GNAS*,*HAX1*,*HNRNPK*,*HOXA7*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *NFKB2*,*PLAT*,*PTTG1*,*RUFY3*,*SMAD4*,*SPP1*,*STAT1*, *STIM1*,*TAB2*,*YWHAZ*	25
Cellular Movement	invasion of cells	1.26E-07	*CCL1*,*CCND1*,*CCR1*,*CDKN1A*,*ERBB4*,*ETS1*,	20

			GNAS,*HAX1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *PLAT*,*PTTG1*,*RPSA*,*SMAD4*,*SPP1*,*STAT1*	
Cellular Movement	invasion of tumor cell lines	1.37E-07	*CCND1*,*CCR1*,*CDKN1A*,*ERBB4*,*ETS1*,*GNAS*,*HAX1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *PLAT*,*PTTG1*,*RPSA*,*SMAD4*,*SPP1*,*STAT1*	19
Cellular Movement	migration of tumor cell lines	2.56E-06	*CCL1*,*CCND1*,*CDKN1A*,*DCBLD2*,*ETS1*,*GNAS*,*HAX1*, *HNRNPK*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), *PLAT*,*PTTG1*,*SMAD4*,*SPP1*,*STAT1*,*STIM1*	18
Connective Tissue Disorders, Immunological Disease, Inflammatory Disease, Skeletal and Muscular Disorders	rheumatoid arthritis	3.87E-04	*ARIH1*,*CCL1*,*CCR1*,*H3F3A/H3F3B*,*IGFBP5*,*MAFB*, miR-155-5p (miRNAs w/seed UAAUGCU), *PTPN22*,*RPL18A*,*RPSA*,*SLC7A2*,*SPP1*,*STAT1*, *TNFAIP3*	14

Connective Tissue Disorders, Immunological Disease, Inflammatory Disease, Skeletal and Muscular Disorders	systemic lupus erythematosus	1.14E-09	*CTSA*,*IFIT1*,*ISG15*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-298 (miRNAs w/seed GCAGAAG), *OASL*,*PTPN22*,*SPP1*,*TNFAIP3*,*UBE2L3*	12
Connective Tissue Disorders, Inflammatory Disease, Skeletal and Muscular Disorders	Rheumatic Disease	3.68E-08	*ARIH1*,*CCL1*,*CCR1*,*CTSA*,*H3F3A/H3F3B*,*IFIT1*, *IGFBP5*,*ISG15*,*MAFB*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-298 (miRNAs w/seed GCAGAAG), *OASL*,*PTPN22*,*RPL18A*,*RPSA*,*SLC7A2*,*SPP1*,*STAT1*, *TNFAIP3*,*UBE2L3*	25
Dermatological Diseases and Conditions	psoriasis	9.70E-09	*CCND1*,*HAX1*,*IFI27*,*IFI44*,*IFI44L*,*IFIT1*,*IFIT3*,*IRAK1*, *ISG15*,*MECP2*, miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *OASL*,*PLAT*,*STAT1*,*TNFAIP3*,*TNIP1*	20
Gene Expression	expression of RNA	3.90E-08	*ARID4A*,*CCND1*,*CDKN1A*,*EIF4A2*,*ERBB4*,*ETS1*, *HAX1*,*HEY2*,*HNRNPK*,*HOXA7*,*IFI27*,*IGFBP5*,*IRAK1*, *KAT2A*,*KAT2B*,*MECP2*,*MYBL2*,*NFKB2*,*PRLR*,*PTPN22*, *PTTG1*,*RBL1*,*RPL37*,*SMAD4*,*SMARCE1*,*STAT1*,*TAB2*, *TCF21*,*TNFAIP3*,*TNIP1*,*TRAF6*,*UBE2L3*	32

Gene Expression	transcription	4.31E-06	*ARID4A*,*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*HAX1*,*HEY2*, *HNRNPK*,*HOXA7*,*IFI27*,*KAT2A*,*KAT2B*,*MECP2*,*MYBL2*, *NFKB2*,*PRLR*,*PTPN22*,*PTTG1*,*RBL1*,*SMAD4*,*SMARCE1*, *STAT1*,*TAB2*,*TCF21*,*TRAF6*,*UBE2L3*,*ZNF417/ZNF587*	27
Gene Expression	transcription of RNA	1.27E-05	*ARID4A*,*CCND1*,*CDKN1A*,*ERBB4*,*ETS1*,*HAX1*,*HEY2*, *HNRNPK*,*HOXA7*,*IFI27*,*KAT2A*,*KAT2B*,*MECP2*,*MYBL2*, *NFKB2*,*PRLR*,*PTTG1*,*RBL1*,*SMAD4*,*SMARCE1*, *STAT1*,*TAB2*,*TCF21*,*TRAF6*,*UBE2L3*	25
Gene Expression	transcription of DNA	3.82E-05	*ARID4A*,*CCND1*,*ERBB4*,*ETS1*,*HAX1*,*HEY2*, *HNRNPK*,*HOXA7*,*IFI27*,*KAT2A*,*KAT2B*,*MECP2*,*MYBL2*, *NFKB2*,*PTTG1*,*SMAD4*,*SMARCE1*,*STAT1*,*TCF21*,*TRAF6*,*UBE2L3*	21
Gene Expression	activation of DNA endogenous promoter	5.50E-05	*ARID4A*,*CCND1*,*ETS1*,*HAX1*,*HEY2*,*HNRNPK*, *HOXA7*,*IFI27*,*KAT2A*,*KAT2B*,*MYBL2*,*NFKB2*, *PTTG1*,*SMAD4*,*STAT1*,*TCF21*,*TRAF6*	17
Hematological Disease, Immunological Disease	lymphoproliferative disorder	3.33E-04	*CCND1*,*CDKN1A*,*ETS1*,*IFI27*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-342-3p (miRNAs w/seed CUCACAC), *RIF1*,*SYNE2*,*TNFAIP3*,*TNIP1*,*YWHAZ*	16
Hereditary Disorder	autosomal dominant disease	3.08E-04	*ATXN7*,*CCND1*,*ELOVL4*,*ERBB4*,*HEY2*,*MAFB*, *NFKB2*,*PTTG1*,*RPSA*,*SMAD4*,*STAT1*,*STIM1*,*SYNE2*	13
Immunological Disease	systemic autoimmune syndrome	1.59E-07	*ARIH1*,*CCL1*,*CCR1*,*CTSA*,*H3F3A/H3F3B*,*IFIT1*, *IGFBP5*,*ISG15*,*MAFB*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-196a-5p (and other miRNAs w/seed AGGUAGU), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-298 (miRNAs w/seed GCAGAAG),	23

			OASL,*PTPN22*,*RPL18A*,*RPSA*,*SLC7A2*,*SPP1*, *STAT1*,*TNFAIP3*,*UBE2L3*	
Infectious Diseases	Viral Infection	6.96E-04	*FAM135A*,*H3F3A/H3F3B*,*HIST2H2BE*,*HNRNPK*, *HOXA7*,*IFIT1*,*IFIT3*,*ISG15*,*KAT2A*, miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYEF2*,*OASL*,*PDE3A*,*PLAT*,*PPP2R2A*,*PTTG1*, *RPS27A*,*RPSA*,*SPP1*,*STAT1*,*UBE2L3*,*ZNF417/ZNF587*	23
Inflammatory Disease	chronic inflammatory disorder	5.77E-04	*ARIH1*,*CCL1*,*CCR1*,*H3F3A/H3F3B*,*IGFBP5*,*MAFB*, miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC) ,*PDE3A*,*PTPN22*,*RASGRP3*,*RPL18A*,*RPSA* ,*SLC7A2*,*SPP1*,*STAT1*,*TNFAIP3*	17
Inflammatory Response	inflammation of organ	4.63E-05	*ARIH1*,*H3F3A/H3F3B*,*IFI27*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-423-5p (and other miRNAs w/seed GAGGGGC), *PDE3A*,*SPP1*,*STAT1*,*SYNE2*,*TNFAIP3*	16
Inflammatory Response	inflammation of body region	1.31E-03	miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-21-5p (and other miRNAs w/seed AGCUUAU), miR-224-5p (miRNAs w/seed AAGUCAC), miR-23a-3p (and other miRNAs w/seed UCACAUU), miR-423-5p (and other miRNAs w/seed GAGGGGC), *PDE3A*,*SPP1*,*STAT1*	10

Metabolic Disease	amyloidosis	2.89E-04	*CCND1*,*DDX46*,*FAM3C*,*IGFBP5*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), *PDE3A*,*PLAT*,*SLC7A2*,*YWHAZ*	13
Metabolic Disease, Neurological Disease, Psychological Disorders	Alzheimer's disease	2.33E-03	*DDX46*,*FAM3C*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-17-5p (and other miRNAs w/seed AAAGUGC), miR-181a-5p (and other miRNAs w/seed ACAUUCA), *PDE3A*,*PLAT*,*SLC7A2*,*YWHAZ*	11
Neurological Disease	progressive motor neuropathy	8.72E-05	*CDKN1A*,*EIF4A2*,*ERBB4*,*GNAS*,*H3F3A/H3F3B*, *IFIT1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC),SPP1	13
Neurological Disease, Skeletal and Muscular Disorders	neuromuscular disease	2.74E-04	*CHAF1B*,*EIF4A2*,*FAM3C*,*GHITM*,*GNAS*, *H3F3A/H3F3B*,*IFIT1*,*ISG15*, miR-125b-5p (and other miRNAs w/seed CCCUGAG), miR-145-5p (and other miRNAs w/seed UCCAGUU), miR-146a-5p (and other miRNAs w/seed GAGAACU), miR-155-5p (miRNAs w/seed UAAUGCU), miR-17-5p (and other miRNAs w/seed AAAGUGC), *SPP1*,*TRAF6*,*YWHAZ*	16
Organismal Survival	organismal death	3.33E-06	*CCND1*,*HNRNPK*,*HOXA7*,*IRAK1*, miR-125b-5p (and other miRNAs w/seed CCCUGAG),	12

				miR-155-5p (miRNAs w/seed UAAUGCU), miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-21-5p (and other miRNAs w/seed AGCUUAU), *MYBL2*,*SPP1*,*STAT1*,*TNFAIP3*	
Protein Synthesis	metabolism of protein		5.39E-04	*ARIH1*,*EIF4A2*,*HNRNPK*,*IGFBP5*,*KAT2A*,*PLAT*, *RPL37*,*SMAD4*,*SPP1*,*TNFAIP3*,*TNIP1*,*UBE2L3*	12

22. ADDENDUM H: Core Network 1 expanded, and prepared for overlays

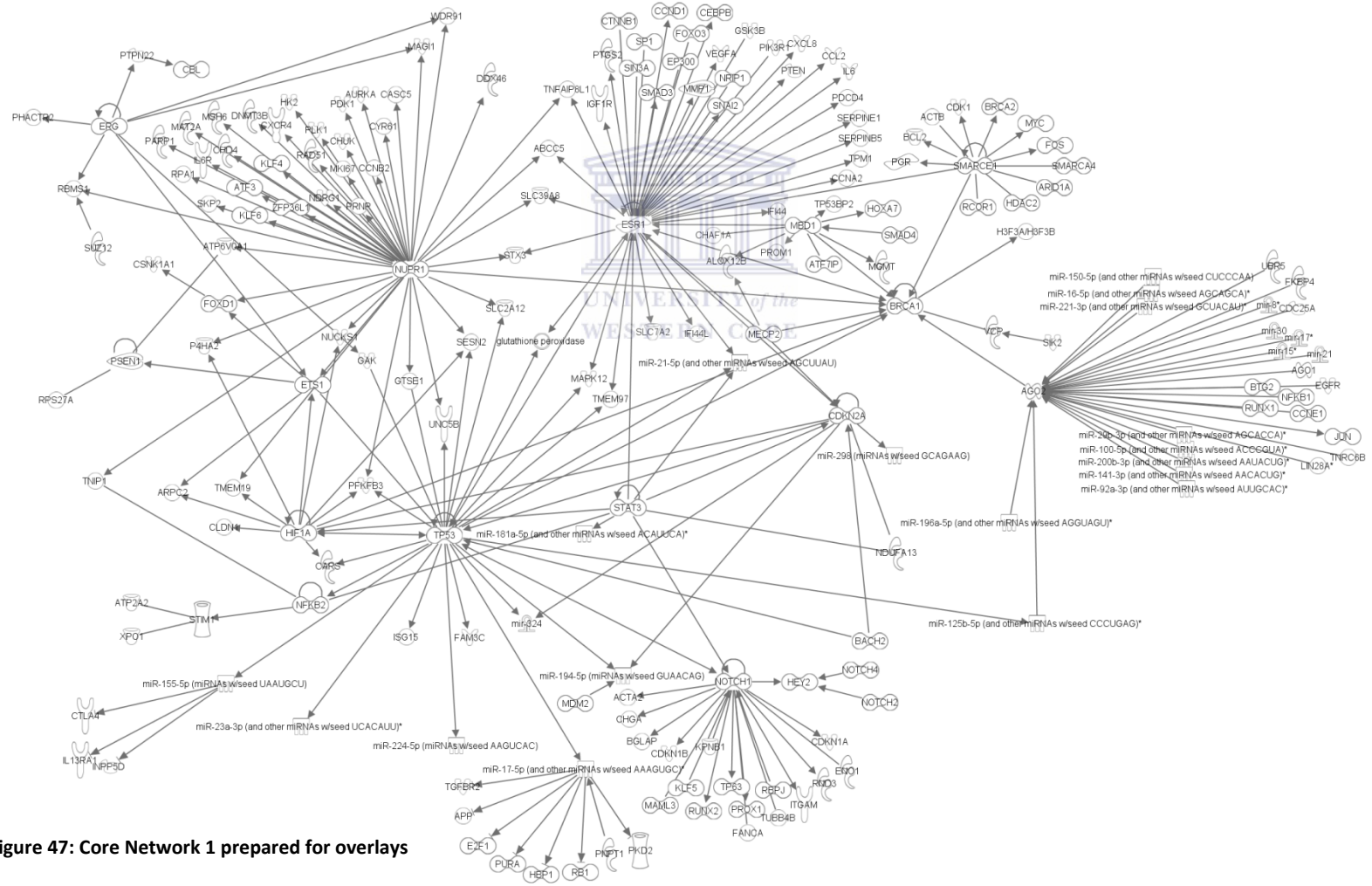

Figure 47: Core Network 1 prepared for overlays

23. ADDENDUM I: Core Network 2 expanded, and prepared for overlays

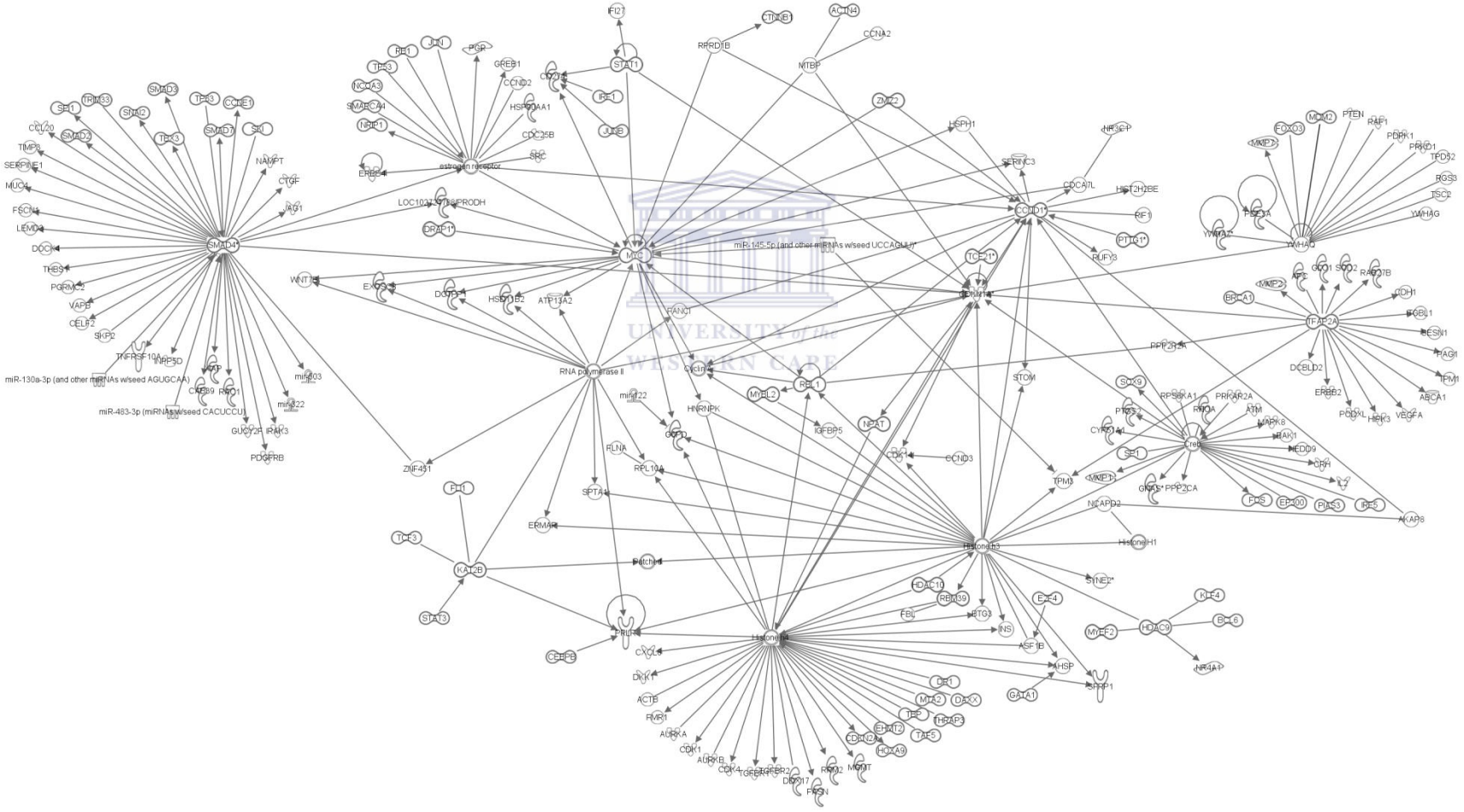

Figure 48: Core Network 2 prepared for overlays

24. ADDENDUM J: Core Network 3 expanded, and prepared for overlays

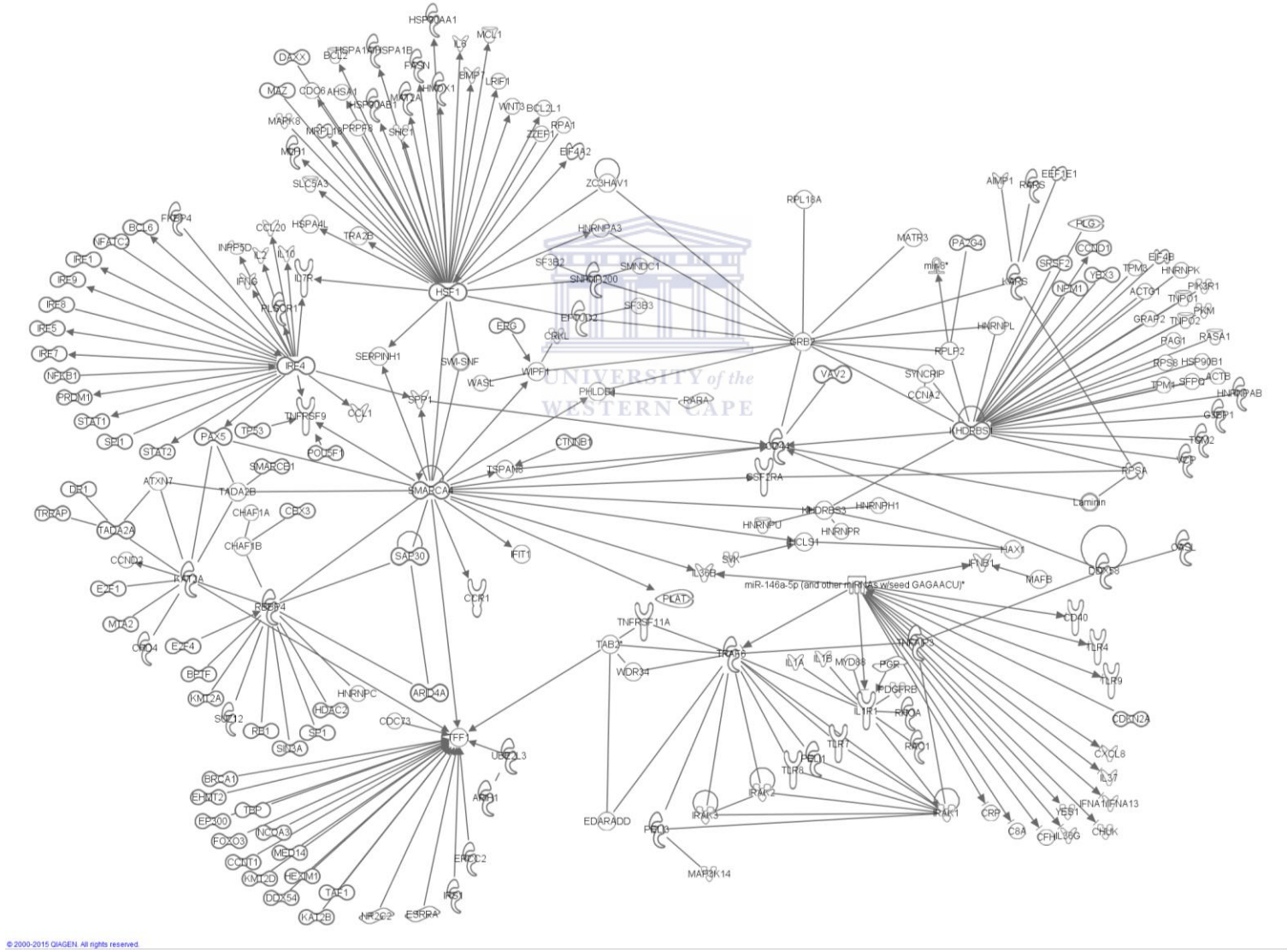

Figure 49: Core Network 3 prepared for overlays

25. ADDENDUM K: The 221 curated, and prepared identifiers, which were mapped to identifiers in IPA not having exactly the same name

Raw Data ID	ID Mapped in IPA	Raw Data ID	ID Mapped in IPA
APITD1	*APITD1/APITD1-CORT*	miR-200	mir-8*
ARHGAP8	*ARHGAP8/PRR5-ARHGAP8*	miR-200a-3p	miR-141-3p
ATP6	*MT-ATP6*	miR-200b	mir-8*
C19orf55	*PROSER3*	miR-200c	mir-8*
C2orf43	*LDAH*	miR-200c-3p	miR-200b-3p
C4A	*C4A/C4B**	miR-208b-3p	miR-208a-3p
C4B	*C4A/C4B**	miR-20	mir-17*
CBS	*CBS/LOC102724560*	miR-20a-5p	miR-17-5p
CCDC41	*CEP83*	miR-20b-5p	miR-17-5p
CDY2A	*CDY1*	miR-21	mir-21
COX1	*MT-CO1*	miR-211-5p	miR-204-5p
COX2	*MT-CO2*	miR-212-3p	miR-132-3p
COX3	*MT-CO3*	miR-215-5p	miR-192-5p
DNAAF5	*DNAAF5**	miR-222	mir-221
FAM105B	*OTULIN*	miR-222-3p	miR-221-3p
FAM5B	*BRINP2*	miR-224	mir-224
FCGR3A	*FCGR3A/FCGR3B**	miR-23b-3p	miR-23a-3p
FCGR3B	*FCGR3A/FCGR3B**	miR-23b-5p	miR-23a-5p
GPR64	*ADGRG2*	miR-25-3p	miR-92a-3p
H3F3B	*H3F3A/H3F3B*	miR-26b-5p	miR-26a-5p
HEATR2	*DNAAF5**	miR-27b-3p	miR-27a-3p
HIST2H2AA3	*HIST2H2AA3/HIST2H2AA4*	miR-28-5p	miR-708-5p
HSFY2	*HSFY1/HSFY2*	miR-298	mir-298
HSPA1A	*HSPA1A/HSPA1B*	miR-299-3p	miR-299a-3p
IFNA1	*IFNA1/IFNA13*	miR-299-5p	miR-299a-5p
IGJ	*JCHAIN*	miR-29a-3p	miR-29b-3p
IL8	*CXCL8*	miR-29c	mir-29
KIAA1279	*KIF1BP*	miR-29c-3p	miR-29b-3p
KLHL23	*KLHL23/PHOSPHO2-KLHL23*	miR-301a-3p	miR-130a-3p
LEPRE1	*P3H1*	miR-302	mir-302*
LEPREL2	*P3H3*	miR-302a-3p	miR-291a-3p
LIN28	*LIN28A**	miR-302c	mir-302*
LIN28A	*LIN28A**	miR-302c-3p	miR-291a-3p

MAP3K7IP2	*TAB2**	miR-302d	mir-302*
ND2	*MT-ND2*	miR-302d-3p	miR-291a-3p
ND4	*MT-ND4*	miR-30	mir-30
ND4L	*MT-ND4L*	miR-30a-5p	miR-30c-5p
ND5	*MT-ND5*	miR-30b-5p	miR-30c-5p
ND6	*MT-ND6*	miR-30c-2-3p	miR-30c-1-3p
PHF16	*JADE3*	miR-30d-5p	miR-30c-5p
PHF17	*JADE1*	miR-30e-5p	miR-30c-5p
PRODH	*LOC102724788/PRODH*	miR-3148	MIR3148
PTPLAD1	*HACD3*	miR-32-5p	miR-92a-3p
PWP2	*LOC102724159/PWP2*	miR-320	mir-320
RAB6C	*RAB6C/WTH3DI*	miR-320c	miR-320b
RAB7L1	*RAB29*	miR-324-3p	miR-1913
RGPD2	*RGPD3*	miR-326	mir-326
SGK3	*C8orf44-SGK3/SGK3*	miR-33	mir-33*
SIP1	*GEMIN2*	miR-33a-5p	miR-33-5p
SMCR7L	*MIEF1*	miR-33b	mir-33*
SOGA2	*MTCL1*	miR-33b-5p	miR-33-5p
TAB2	*TAB2**	miR-346	mir-346
TMSB10	*TMSB10/TMSB4X*	miR-34	mir-34*
TMSL8	*TMSB15A*	miR-34b	mir-34*
TRAPPC2P1	*TRAPPC2B*	miR-34c	mir-34*
TXLNG2P	*TXLNGY*	miR-34c-5p	miR-34a-5p
ZNF587	*ZNF417/ZNF587*	miR-3615	mir-3615
miR-1	miR-1-3p	miR-363-3p	miR-92a-3p
miR-103a-2-5p	miR-103-1-5p	miR-365a-3p	miR-365-3p
miR-103a-3p	miR-103-3p	miR-372	mir-290
miR-106	mir-17*	miR-372-3p	miR-291a-3p
miR-106a-5p	miR-17-5p	miR-373	mir-373
miR-106b	mir-17*	miR-373-3p	miR-291a-3p
miR-106b-5p	miR-17-5p	miR-375	mir-375
miR-107	mir-103	miR-376a*	miR-376a-5p
miR-10b-5p	miR-10a-5p	miR-421	mir-95
miR-122	mir-122	miR-424	mir-322
miR-124	miR-124-3p	miR-424-5p	miR-16-5p
miR-125a-5p	miR-125b-5p	miR-429	mir-8*
miR-125b	miR-125b-5p	miR-433	mir-433
miR-126-3p	miR-126a-3p	miR-449	mir-449
miR-126-5p	miR-126a-5p	miR-454-3p	miR-130a-3p
miR-1260b	mir-1260b	miR-484	mir-484
miR-1276	mir-1276	miR-502-3p	miR-501-3p
miR-128	miR-128-3p	miR-503	mir-503
miR-1303	mir-1303	miR-519b-3p	miR-519a-3p

miR-130	mir-130	miR-519c-3p	miR-519a-3p
miR-130b-3p	miR-130a-3p	miR-519d-3p	miR-17-5p
miR-133	miR-133a-3p	miR-519e-3p	miR-515-3p
miR-133b	mir-133	miR-520a-3p	miR-291a-3p
miR-135b-5p	miR-135a-5p	miR-520b	mir-515*
miR-141	mir-8*	miR-520c-3p	miR-291a-3p
miR-145	mir-145	miR-520	mir-515*
miR-146	mir-146	miR-548	miR-548aa
miR-146b-5p	miR-146a-5p	miR-548q	mir-548
miR-148b-3p	miR-148a-3p	miR-572	mir-572
miR-151a-3p	miR-151-3p	miR-577	mir-577
miR-151a-5p	miR-151-5p	miR-608	mir-608
miR-155	mir-155	miR-638	mir-638
miR-15	mir-15*	miR-639	mir-639
miR-15a-5p	miR-16-5p	miR-657	mir-657
miR-15b-5p	miR-16-5p	miR-663	mir-663*
miR-16	miR-16-5p	miR-663b	mir-663*
miR-17	mir-17*	miR-7	miR-7a-5p
miR-181a-3p	miR-181a-1-3p	miR-7-5p	miR-7a-5p
miR-181b-2-3p	miR-181b-1-3p	miR-758	mir-379
miR-181b-3p	miR-181b-1-3p	miR-760	mir-760
miR-181b-5p	miR-181a-5p	miR-769-3p	miR-450b-3p
miR-181c	mir-181	miR-802	mir-802
miR-181c-5p	miR-181a-5p	miR-9	miR-9-5p
miR-181d-5p	miR-181a-5p	miR-92b-3p	miR-92a-3p
miR-184	mir-184	miR-93	mir-17*
miR-18	mir-17*	miR-93-5p	miR-17-5p
miR-193b-3p	miR-193a-3p	miR-935	mir-935
miR-195	mir-15*	miR-940	mir-940
miR-195-5p	miR-16-5p	miR-942-5p	miR-7028-3p
miR-196	miR-196a-5p	miR-98-5p	let-7a-5p
miR-198	mir-198	miR-99a-5p	miR-100-5p
miR-19	mir-19	miR-99b-3p	miR-99a-3p
miR-19a-3p	miR-19b-3p	miR-99b-5p	miR-100-5p
miR-19b	miR-19b-3p		