Aging and Longevity in Humans: Functional Annotation of Single Nucleotide Variants in 1000 Genomes

By

Syed Aleem Haider

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2015

Aging and Longevity in Humans: Functional Annotation of Single Nucleotide Variants in 1000 Genomes

By

Syed Aleem Haider

A thesis

In the partial fulfillment of the

requirements for the degree of

MASTER OF PHILOSOPHY

In

BIOINFORMATICS

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2015

CERTIFICATE

This thesis submitted by **Mr. Syed Aleem Haider** is accepted in its present form by the National Center for Bioinformatics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad as satisfying the thesis requirements for the degree of Master of Philosophy (M.Phil) in Bioinformatics.

Internal Examiner: _____

Dr. Amir Ali Abbasi

Assistant Professor& Supervisor

National Center for Bioinformatics

Quaid-i-Azam University, Islamabad

External Examiner: _____

Chairman: _____

Prof. Dr. Wasim Ahmed

Chairperson & Dean,

Faculty of Biological Sciences

Dated: _____

Submission to Allah's Will is the best companion; wisdom is the noblest heritage; theoretical and practical knowledge are the best signs of distinction; deep thinking will present the clearest picture of every problem.

Hazrat Ali (A.S)

Nahjul Balagha

Dedication

To my late father Syed Zameer Haider

His words of inspiration and encouragement still remain fresh in my heart.

LIST OF ABBREVIATIONS

AD	**Alzheimer's Disease**
Chip-seq	**Chromatin Immunoprecipitation Sequencing**
CDKs	**Cyclin Dependent Kinases**
DNA	**Deoxyribonucleic Acid**
ENCODE	**Encyclopaedia of DNA Elements**
eQTL	**Expression Quantitative Trait Loci**
GERP	**Genomic Evolutionary Rate Profiling**
GWAS	**Genome Wide Association Studies**
LD	**Linkage Disequilibrium**
MAF	**Minor Allele Frequency**
MODY	**Maturity Onset Diabetes of the Young**
NHGRI	**National Human Genome Research Institue**
PWMs	**Position Weight Matrices**
RBCs	**Red Blood Cells**
SiPhy	**Site-specific Phylogenetic analysis**
SNP	**Single Nucleotide Polymorphisms**

ACKNOWLEDGEMENTS

Foremost, I would like to thank Almighty **ALLAH** for giving me much more than I deserve. My sincere thanks to my parents for their endless support and love day in and day out throughout my career. Thank you both for providing me a conducive environment and giving me strength to achieve my goals.

I would like to express deep gratitude to my advisor **Dr. Muhammad Faisal** for his persistent encouragement and supervision during my M.Phil research work. His motivation, confidence in me, and guidance has undoubtedly become an integral component of this manuscript.

Besides, I also thank **Prof. Dr. Waseem Ahmed**, Dean Biological Sciences and chairperson of National Center of Bioinformatics (NCB), and all the faculty members of NCB particularly **Dr. Amir Ali Abbasi** who initially introduced me to this wonderful place - his guidance throughout the research work is surely commendable.

Moreover, my sincere thanks also goes to all the staff members of NCB for their kind facilitation in each and every relevant issue.

To all my friends especially my lab mates **Rabia Tahir**, **Saima Nawaz** and **Raiha Mumtaz**, thank you for sharing a wonderful time at Statistical Bioinformatics Lab, NCB.

Syed Aleem Haider

ABSTRACT

Human aging is a gradual decrease in cellular integrity that contributes to multiple complex disorders such as Neurodegenerative disorders, Cancer, Diabetes and Cardiovascular diseases. Since the completion of Human Genome Project nearly a decade ago, the focus has shifted towards the identification of Single Nucleotide Polymorphisms (SNPs) that are correlated with complex disorders. In this regard, Genome-wide association studies (GWAS) play a key role in discovering genetic variations that may contribute towards disease vulnerability. However, mostly disease-associated SNPs lie within non-coding part of the genome; majority of the variants are also present in Linkage Disequilibrium (LD) with the genome-wide significant SNPs (GWAS lead SNPs). We asked whether non-coding variants play a crucial role in the cellular homeostasis; we performed functional annotation of SNPs that lie in the non-coding genomic region ─ using ENCODE datasets via RegulomeDB, HaploregV2 and rSNPBase. Moreover, various bioinformatics tools including STRING, DISEASES, and Gene Network informs us about known and predicted protein-protein interactions, disease-gene associations and gene networks with shared pathways respectively. Overall 600 SNPs were analyzed, out of which 291 returned RegulomeDB scores of 1-6. It was observed that just 4 out of those 291 SNPs show strong evidence for potential regulatory effects with RegulomeDB score < 3, while none of them includes any GWAS lead SNP. Nevertheless, this study demonstrates that by utilizing epigenetic data sets, it is possible to discover potential regulatory variants – moving from GWAS towards understanding disease pathways.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1. INTRODUCTION

1.1. Human Aging: An Overview

Aging is a prolonged mechanism that affects the life span of almost every living organism on earth. In a given population, the rate of aging varies with some individuals that age slowly than others [1]. Under these circumstances, scientists use separate terms related with aging – such as chronological and biological aging – in order to better describe the process of aging. While chronological aging refers to the overall age of an individual in years, the biological aging simply means how fit an individual is, moving from birth towards death in old age [2]. To better understand this phenomenon, we can look at different individuals who may be in his/her 60s but look quite younger than their chronological age. In this study, I have regarded aging as an overall physiological process. There are several parameters that affect aging ranging from organism's body mass, size, composition and its environment. When we talk about the overall life expectancy, it has been discussed rather noticeably in terms of medium life expectancy versus maximum life span of the specie. With the advent of modern healthcare equipment and better hygienic conditions, the contemporary medicine has actually contributed towards better life extent of humans. We ought to completely identify the fundamental mechanisms that ultimately contribute towards aging in order to improve the longevity of humans. Additionally, we can say that the numerous complex life-threatening diseases of the present time are playing a vital role in the overall life expectancy of the individual. Recent advances in human genetics ever since the completion of Human Genome Project [3-5] have paved the way for better understanding of the cellular mechanisms altered in quite a lot of complex diseases. Understanding the genes implicated in disease phenotypes will finally contribute towards increased life span. Lately, some focus has also shifted towards the consideration of epigenetic perturbations of the human genome that could also be responsible for aging. Finally, it has been observed that aging is not a general phenotype of an organism; rather it has consequences at cellular, tissue, organ, and system level. As a result, keeping this in view, one should recognize aging as cellular aging and to comprehend aging at smallest possible cellular levels.

Human aging has been associated with an increased risk of multiple complex disorders, but the fact is that we do not know much about the molecular mechanisms behind these correlations. With the better understanding of aging and its mechanisms it would be unproblematic to avoid these ailments. However as mentioned previously, this may possibly be done once we acknowledge the fact that aging is basically a multi-factorial process that involves organ systems as a whole. Consequently, a systems genetics approach to reveal the hidden genetic markers associated with aging would definitely improve the prevention against multifaceted disorders.

Meanwhile, when we talk about aging, we must also think of it at the basic cellular level, and in this study we have considered this process as cellular aging. Moreover, we need to study different biochemical changes associated with aging at cellular level in order to pinpoint crucial pathways leading to disturbed cellular homeostasis. Besides this, Alzheimer's disease (AD) is still incurable ailment and the statistical data confirms the fact that there is strong association between AD and aging. Therefore, we need to look for different biological mechanisms that

gets altered with increasing age; only then we might be able to find the causal link behind AD. Several types of Cancer are also frequently reported as the organism ages; cancer treatment somehow suppresses the intensity of ailment, but again, the causal link between cancer and aging is not well understood. With the advent of modern sequencing technologies and the availability of genotype data of thousands of individuals, we are edging closer towards better understanding of disease pathogenesis. In addition, Cancer is often termed as disease of the genome and with the successful completion of International HapMap Project [6-8], 1000Genomes Project [9-12] - and the ongoing ENCODE Project [13-16] - more knowledge has been gained regarding the structure of human genome. Genome-wide association studies (GWAS) [17-20] have also successfully identified thousands of loci linked with complex disorders. In this context, more research is required particularly towards the identification of causal genetic variations linked with diseases as shown in figure 1.1; this could perhaps assist us in getting an even closer look at the factors related with these disorders.

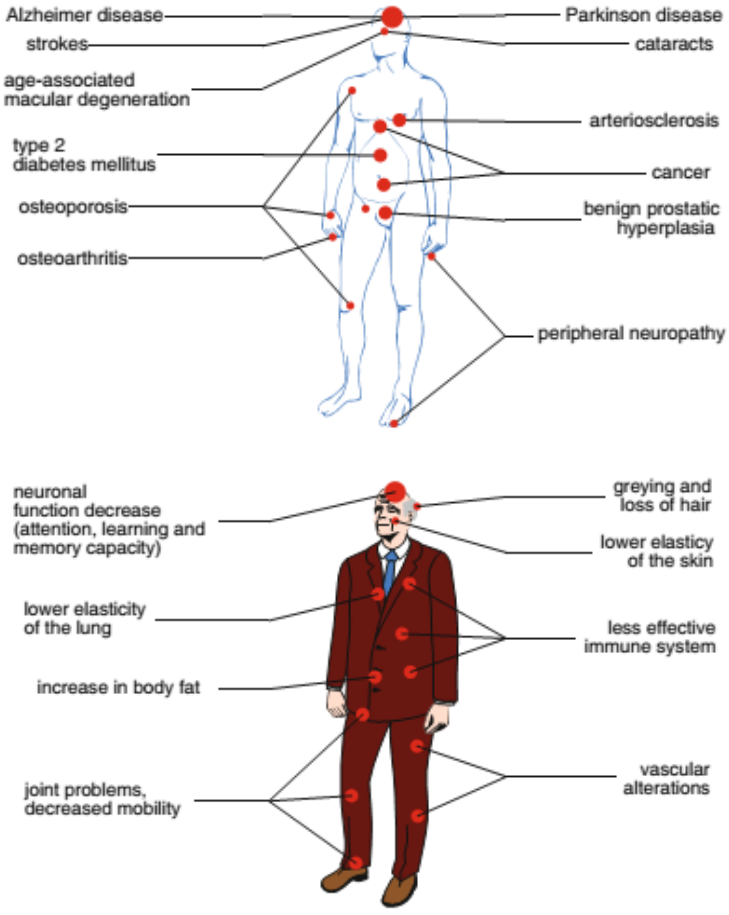

Figure 1.1: Major disorders and functional changes associated with Aging
(Reprinted from C. Behl and C. Ziegler, Cell Aging: Molecular Mechanisms and Implications for Disease)

There are a few fundamental changes reported in the biochemistry of cells, tissues and even organs that experience effects of aging. Some of the key modifications comprise oxidation of bio-molecules such as proteins, lipids and carbohydrates etc. as it has been shown that there is remarkable decline in the levels of estrogen produced in human females during and even past

the menopause. This reduced concentration of estrogen is primarily responsible for one of the most important age-related pathology i.e. Osteoporosis. Nevertheless, we should focus towards the aging mechanism of a single cell specifically to identify the essential biological changes taking place inside it so as to have a broader view of aging in general. Considering the median life span of different cells in a human body, a number of the cells have longer life span than others. For instance, liver cells stay alive for as long as 5 months, while erythrocytes (red blood cells-RBCs) gets renewed after about 4 months. Yet, some cells remain alive for the entire life of an organism such as the neuronal cells. Consequently, the chief motivating force behind the life and death of a single cell lies in the cell cycle that performs a central role in determining the destiny and existence of a single cell. A better knowledge of the cell cycle would beyond doubt throw some light on the processes coupled with cellular aging.

Cell division is a gradual process that finally directs the development of two daughter cells from a distinct parent cell. This procedure is exceedingly structured into separate phases so that the reliability of cell division remains intact. Let's talk about few of the key steps of division. There are primarily two central steps in this namely; interphase and mitosis. Figure 1.2 explains various stages of cell cycle.

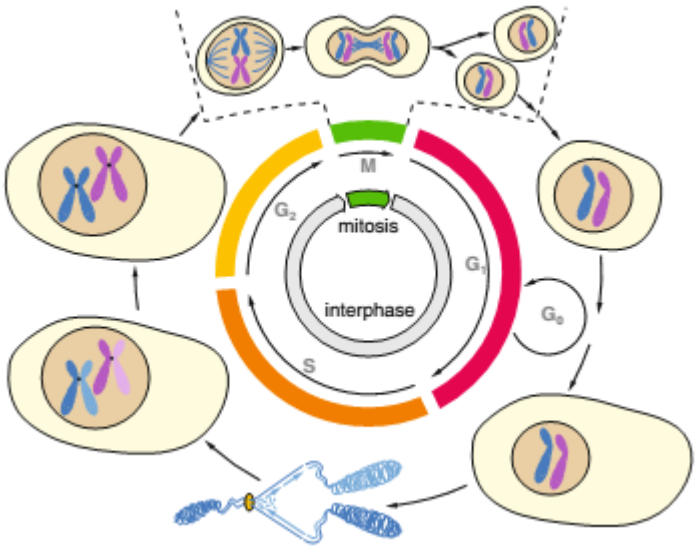

Figure 1.2: Stages of Eukaryotic cell division
(Reprinted from C. Behl and C. Ziegler, Cell Aging: Molecular Mechanisms and Implications for Disease)

Interphase involves the period during which a cell organizes itself for division shortly by copying its genome and producing essential biomolecules before entering the cell cycle. Conversely, mitosis includes the splitting of a single cell into two identical ones. Aberrations in this cycle are associated with aging and numerous other complex disorders of the current age. The main focus of cell cycle in the perspective of aging is towards the ability of cells to repair damaged parts in an attempt to maintain a healthy number of cells in the body. It has been noted that the reliability of cell cycle rests upon the successful error free duplication of DNA before the cell go through mitotic phase. Additionally, DNA repair machinery guarantees typical duplication of DNA. For this reason, it has been observed that studying the numerous

checkpoints in the cell cycle in addition to the DNA repair mechanism might give us some evidence about the mistakes in cell cycle that sooner or later give rise to cellular aging. Talking about these checkpoints, we first need to expand appropriate awareness of the fundamental steps that direct the correct cell division. There are essentially two main phases one before and one after the mitotic M phase. The first phase named as G1 phase is also recognized as post-mitotic phase wherein cell initiates its development together with the conclusion of differentiation process. It also assembles necessary cellular equipment for the next stage and it generally takes roughly 3hrs. G2 phase is another segment that is also related with the M phase. It is identified as pre-mitotic stage when cell gets ready to divide into two in the upcoming M phase. In the meantime, there is one more significant phase termed S phase, through which DNA gets duplicated and histones are also formed that largely ensure the correct packaging and folding of genetic matter once the cell gets divided. This segment continues for about 7hrs. Another important phase is acknowledged as G0 phase, also known as the dormant state whereby cell stops dividing and enters a paused state. Some cells gets differentiated and then by no means come back to regular division process while some cells stop dividing courtesy unfavorable environmental conditions. However, once this condition subsides, cell re-enter the cycle. Lastly, a cell enters the M phase followed by karyokinesis during which nucleus split into two and finally cytokinesis ensures equal division of the cellular cytoplasm. This more often than not puts an end to the cycle. In the context of aging, it must be clear that senescence and dormant states are two discrete ones. While senescence (observed in aging) is a permanent state that in due course ends up on apoptosis (programmed cell death), the dormant condition of a cell as stated in case of G0 phase is a transitory one. When the circumstances becomes favorable, cell re-enters the cycle.

The integrity of cell cycle involves appropriate command and control system, so as to carry on this process in a regular way. This control system is chiefly composed of particular proteins known as Cyclins and cyclins-dependent kinases (CDKs). These molecules fit in to a unique group of proteins known as kinases. Kinases are special enzymes which are involved in the shifting of phosphate groups primarily from the energy rich ATP molecules to particular substrates. This method is termed as phosphorylation and is one of the most important steps regarded in the context of post-translational modifications. In excess of 500 kinases have been registered in the research that generally contributes in the phosphorylation of typically three amino acids (threonine, serine and tyrosine). This phosphorylation practice is very much sensitive taking into consideration at what time it occurs. Moreover, there are a few other proteins operating alongside known as phosphatases. These are also distinctive enzymes that essentially functions opposite to kinases by removing phosphate groups from the substrates. Together kinases and phosphatases regulate standard functioning of the cell cycle. The concentration, as well as the proper communication between these two different types of enzymes determines correct cell division as shown in figure 1.3. Talking about cell cycle control, there also exist two crucial players including retinoblastoma protein (Rb) and p53 protein. Both these proteins are one way or another linked with cancer, as studies established that they are related with the cell cycle control mechanism. Additionally, cancer has been reported to mainly involve the interruption of S-phase initiation and G1 progression. When these proteins were studied deeply, they were found to be linked with the same stages of cell

cycle. Therefore, appropriate knowledge of their function inside the cell was required. This encouraged a comprehensive investigation of both of these proteins at biochemical level.

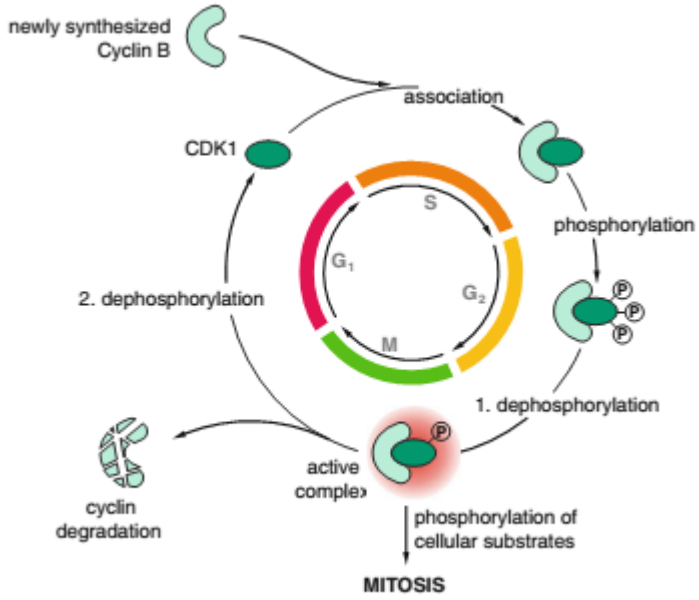

Figure 1.3: Cyclins and CDKs control cell cycle

(Reprinted from C. Behl and C. Ziegler, Cell Aging: Molecular Mechanisms and Implications for Disease)

P53 is a tumor suppressor protein that operates by preventing cell division once DNA gets damaged. These damages mainly consist of UV radiations, chemicals and toxins etc. that eventually influence the three dimensional structure of DNA. These involve breaks in DNA in addition to cross-linking of double helical structure. These changes will in due course pass on to the daughter cells if left un-noticed. However, a powerful DNA repair mechanism ensures correct repair of these alterations to guarantee typical cell performance. DNA repair mechanism is aided by unique p53 check points in the cell cycle that are largely responsible for promoting apoptosis, cellular senescence and growth arrest in particular every time DNA damage is encountered. It also repairs double stranded cuts in DNA by boosting the performance of DNA recombination process. This would finally make sure the proper cell division and inhibits the development of tumors once the cell gets damaged. Yet, alterations in the p53 at gene level may stop the progress of this whole process of repair. This is where the story gets complicated and cell loses control of itself. Unfortunately, many of the tumors involve mutations in p53 gene that in some way make p53 dysfunctional. While these changes upset normal functioning of the cell, p53 has one more role to play as mentioned before. Under such circumstances, when its ordinary functioning is disrupted, it forces the cell to progress towards ultimate programmed death as depicted in figure 1.4. In this fashion, it attempts to decrease some of the adverse effects of uncontrolled cell division. It has been recently observed that "it is being recognized as a critical feature of mammalian cells to suppress tumorigenesis, acting alongside cell death programs" [21]. So what precisely is going on inside a cell that has stopped dividing rather entered in to a long-drawn-out senescence state? First, we need to spot

these senescent cells and why they have entered this state. A number of studies already conferred that the buildup of certain oncogenes forces the cell to move away from regular proliferative state. However, some fresh literature argues that standard immune system simply eliminate these cells. So there is need to examine the indispensable systems coupled with cellular senescence as well as apoptosis.

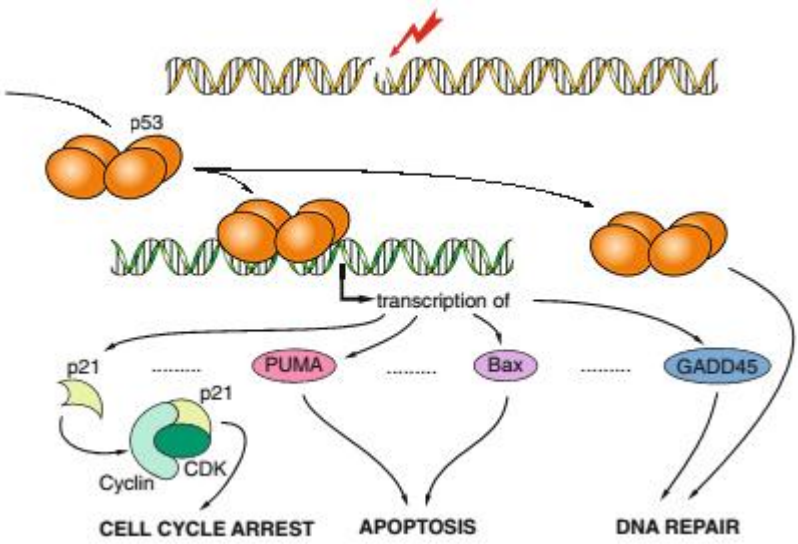

Figure 1.4: p53 and its role in inducing apoptosis upon stress

(Reprinted from C. Behl and C. Ziegler, Cell Aging: Molecular Mechanisms and Implications for Disease)

The progressive loss of physiological integrity associated with aging accompanied with a gradual decline in cellular function is somehow referred as the primary risk factor for major ailments. Several age-related pathologies have been identified including diabetes, neurodegenerative disorders, cardiovascular disorders and cancer [22]. Meanwhile, it has been also suggested that these major diseases can be somehow controlled by looking for common pathways associated with these disorders. In addition to that, it was also suggested that genetic perturbations located in these pathways could play a crucial role. For this purpose, several studies and reviews were focused on the identification of these common pathways and hallmarks linked with these complex disorders. Recently, it was reported that there exists nine potential hallmarks of aging that contributes most towards the on-set of age-related pathologies [22]. The hallmarks include telomere attrition, genomic instability, loss of proteostasis, epigenetic alterations, mitochondrial dysfunction, deregulated nutrient sensing, stem cell exhaustion, cellular senescence and altered intercellular communication [22]. A summary of these mentioned hallmarks is shown in figure 1.5. Furthermore, these nine hallmarks are divided into 3 main categories: primary hallmarks, antagonistic hallmarks and integrative hallmarks. The primary hallmarks consist of telomere attrition, genomic instability, epigenetic alterations and loss of proteostasis. Deregulated nutrient sensing, mitochondrial dysfunction and cellular senescence constitute antagonistic hallmarks whereas the integrative ones are supposed to be stem cell exhaustion and the altered intercellular communication. Moreover, the primary hallmarks can be regarded as the ones causing the initial cellular damage leading

to a response by antagonistic hallmarks. Finally, the actual disease phenotype observed is supposed to be contributed by the integrative hallmarks. The functional link between these hallmarks is shown in figure 1.6. Talking about the primary hallmarks, they possess negative effects, whereas the antagonistic ones have divergent effects based on their respective concentration; they are beneficial at low levels whereas become deleterious when the concentration goes high. Lastly, the integrative hallmarks are directly involved in disturbing cellular homeostasis and normal functioning. Understanding the functional link among these hallmarks is currently progressing and looks promising for future studies.

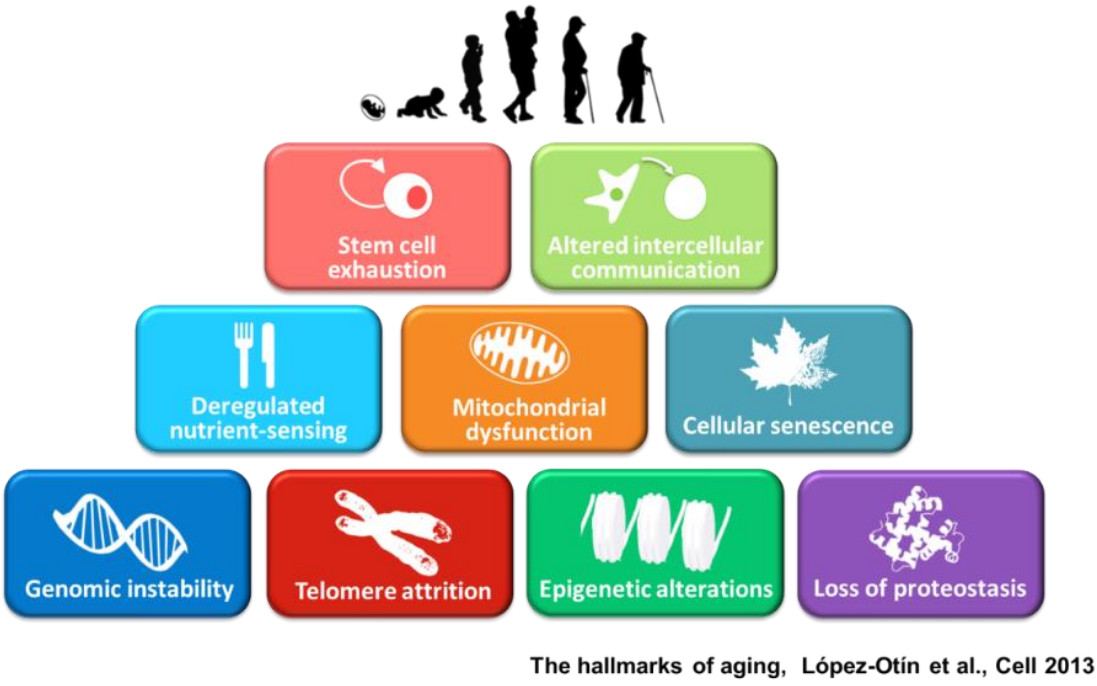

Figure 1.5: The hallmarks of Aging
(Reprinted from Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. Cell 153, 1194–1217 (2013)

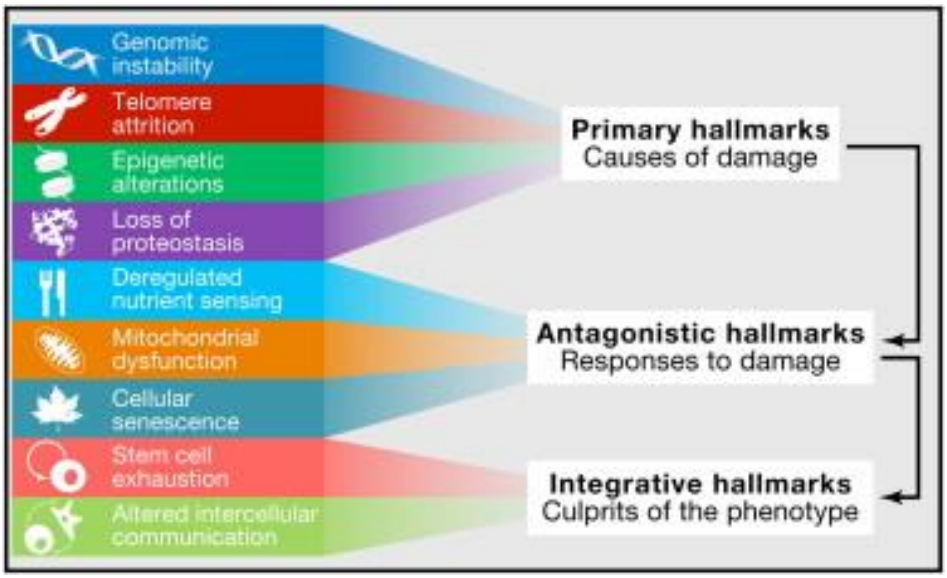

Figure 1.6: Functional link between hallmarks
(Reprinted from Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. Cell 153, 1194–1217 (2013).

1.2. Post-Genomic Era: Applications to Human Genetics

In 1990, a group of scientists started a project that aims to sequence and study the nucleotide base pairs present in the human genome – thereby attaching some meaning to the overall genetic makeup of humans. This venture termed as Human Genome Project [3-5] was supposed to become a new approach in biomedical sciences by unlocking large piece of unknown biological information. The major work was divided into two phases: an early phase – known as draft sequence – with ~90% of the genome sequenced at low coverage, an a final phase – known as complete sequence – with ~99% of the genome sequenced at high coverage. This brings forward a complete euchromatic portion of the human genome. The draft sequence was initially released by joint effort of International Human Genome Sequencing Consortium [3] and Celera Genomics [4] in February 2001 involving 20 centers in 6 countries. Later in 2004, they released the complete version [5]. The completion of Human Genome Project was undoubtedly a tremendous achievement in the context of biomedical research that opened new venues for studies such as: search of disease genes, locate crucial heritable markers for complex diseases, identify somatic mutations associated with major ailments: including Alzheimer's disease, Diabetes, and Breast cancer etc. The Human Genome Project also facilitates in understanding the genome structure during the course of evolution. For this purpose, scientists planned to sequence many mammalian genomes – using Human Genome draft sequence as model – in order to perform comparative genomic studies; thereby locating important functional elements of the human genome. The next big thing after the Human Genome Project was to identify nearly all Single Nucleotide Polymorphisms (SNPs) in human populations to study correlations between the SNPs and diseases. This was a wonderful idea in the context that human genome was already sequenced; if we could somehow identify the polymorphisms present in humans, we will be able to somehow fill the gap between genotype and phenotype. Of course, this must be realized that this work requires lots and lots of sequencing effort –

involving hundreds and even thousands of genomes to be sequenced. The idea of sequencing many more human genomes prompted few projects such as the International HapMap Project [6-8] and the renowned 1000Genomes Project [9-12] – aimed to identify SNPs among different human populations around the globe. These projects are further discussed later in this manuscript. The International Human Genome Sequencing Consortium published a paper marking almost the completion of Human Genome Project that can be viewed for further information [5].

A single nucleotide polymorphism is a sort of DNA sequence variation that is believed to arise within no less than 1% of the population. In this kind of variation, a particular DNA base-pair (A, T, G or C) in the genome differs among individuals of that population. The frequency of occurrence of these variants differs too among different population groups. To understand the frequency of these variants in a given population, the concept of "Minor Allele Frequency (MAF)" is often used as a standard. This frequency essentially refers to the minimum incidence of a specific allele that is being reported in a particular population. Due to the difference present in the MAF of SNPs among dissimilar populations, it is frequently observed that a SNP reported as uncommon in one population might not occur as rare in another. There are quite a few types of these variants that arise in coding regions of the genome as well as in the non-coding part of the genome.

Single nucleotide polymorphisms that lie in the coding region are mainly of two types consisting of synonymous and non-synonymous variants. Before talking about each one of these variants, it must be clear that more often than not SNPs occurring in the coding regions have little or no effect courtesy codon degeneracy, whereby even a single nucleotide change would have no effect on the amino acid formed as a result of codon translation. Such SNPs that have no consequence on the amino acid productivity are in fact termed as synonymous variants. On the other hand, some SNPs modify the codon in such a way that alters amino acid series and therefore influence the overall function of the protein. These sorts of variants are usually known as non-synonymous variants and are of two types too: nonsense and missense. Nonsense variants are those SNPs due to which a normal codon gets converted into a stop codon, whereas missense variants are those SNPs that merely change the amino acid produced after translation of codon.

Moreover, SNPs that lie in the non-coding region of the genome still have a major role in the normal cellular functioning. It has been observed of late that such SNPs have a much bigger role to play in the cellular well-being. These SNPs are usually referred to as the regulatory SNPs as they are mainly found in that region of the genome that is linked with the regulation of gene expression. Hence, its role has been really important in this context. Several studies have already reported important SNPs that alter these regulatory portions of the genome and ultimately affecting normal cellular function.

Since the completion of Human Genome Project, the focus was mainly shifted to identify SNPs present in different human populations. Because Human Genome Project didn't show much about the root cause of complex disorders, it was needed to locate genetic variations that may guide in understanding the diseases [7]. In addition, there also exist heritable risk factors that

play a crucial role in the onset of many complex disorders: such as Alzheimer's disease, Diabetes and Cardiovascular disorders. Therefore, it was planned to discover causal genes and variants that are responsible for disease pathogenesis; this might be an important step towards better understanding of disease mechanism and a much better prevention later [7]. More than 1000 genes have been identified – in rare Mendelian disorders – in which variation in a single gene can lead to disease phenotype; however there also exists common disorders, in which there lies a combined effect of multiple genes, variants and even environmental factors that contributes towards disease pathogenesis [7]. Linkage studies in the past have successfully managed to identify few disease associated loci, however the overall power of the study remains quite low unless a single locus explains much heritability; this requires a strategy that examine genetic variations in a large population consisting of controls and cases (affected individuals) [7]. Availability of complete genome re-sequencing platforms have made this possible. Further insights reveal that common variants play a crucial role in multiple disorders: including Type II Diabetes, Alzheimer's disease, Rheumatoid Arthritis, and Cardiovascular disorders [7]. Linkage disequilibrium (LD) is a phenomenon closely associated with the variants in which specific alleles close to the nearby SNP always travel as a group; this correlation is often used quite a lot these days in scientific research to functionally annotate the phenotypic changes associated with these LD SNPs [7].Moreover, a specific pattern of alleles on a chromosome is termed as haplotype. This notion of developing haplotype maps initiated the International HapMap Project in October 2002 – to create a publicly available catalog of common human genetic variations that eventually guide us towards better understanding of disease mechanisms [6].Results from the project confirms: generality of recombination hotspots, limited diversity of haplotypes along the genome, strong LD patterns, greater redundancy in closely positioned SNPs. However, the mapping of rare, less common variants was still not achieved at that time, making it difficult to solve the missing heritability problem. Nevertheless, this was a big step towards the making of a high resolution haplotype map of human genetic variations; with the cost of sequencing technologies decreasing, it was expected that low frequency rare variants with MAF < 5% will soon be available that may solve the missing heritability phenomenon [8]. In this perspective, the 1000Genomes Project was the next big thing – whose major aim was to detect rare variants.

In order to characterize rare variants, an International collaboration termed as 1000Genomes Project [9-12] among research groups from China, Germany, UK and USA, was started – extending further the ongoing HapMap Project [12]. Initially the project was launched to locate variants in 1,092 human genomes from 14 different populations in order to create a validated haplotype map of about 38 million SNPs [10]. Moreover, it identify variants that occur at least once in 50 individuals; the variants identified, includes some that increases the risk for a disease or some that decreases it whereas most of the variants remain neutral [11]. The pattern of variants located as a result of 1000Genomes Project becomes that much useful in revealing the genetic underpinnings of disease. Meanwhile, more than 2500 genomes from 26 populations will be sequenced at the end of this project – explaining disease susceptibility, drug response and response to environmental cues [11]. Table 1.1 gives a detailed list of 1000Genomes sample populations. Consequently, almost all variants will be available making it easier for

Genome-wide Association Studies (GWAS) to find regions of genome associated with diseases. [12].

1000Genomes Samples					
Population	**Pilot Samples**	**Phase 1 Samples**	**Final Phase Discovery Sample**	**Final Release Sample**	**Total**
Chinese Dai in Xishuangbanna, China (CDX)	0	0	99	93	99
Han Chinese in Beijing, China (CHB)	91	97	103	103	106
Japanese in Tokyo, Japan (JPT)	94	89	104	104	105
Kinh in Ho Chi Minh City, Vietnam (KHV)	0	0	101	99	101
Southern Han Chinese, China (CHS)	0	100	108	105	112
Total East Asian Ancestry (EAS)	**185**	**286**	**515**	**504**	**523**
Bengali in Bangladesh (BEB)	0	0	86	86	86
Gujarati Indian in Houston (GIH)	0	0	106	103	106
Indian Telugu in the UK (ITU)	0	0	103	102	103
Punjabi in Lahore, Pakistan (PJL)	0	0	96	96	96
Sri Lankan Tamil in the UK (STU)	0	0	103	102	103
Total South Asian Ancestry (SAS)	**0**	**0**	**494**	**489**	**494**
African Ancestry in Southwest US (ASW)	0	61	66	61	66
African Caribbean in Barbados (ACB)	0	0	96	96	96
Esan in Nigeria (ESN)	0	0	99	99	99
Gambian in Western Division, The Gambia (GWD)	0	0	113	113	113
Luhya in Webuye, Kenya (LWK)	102	97	101	99	116
Mende in Sierra Leone (MSL)	0		85	85	85
Yoruba in Ibadan, Nigeria (YRI)	106	88	109	108	116
Total African Ancestry (AFR)	**208**	**246**	**669**	**661**	**691**
British in England and Scotland (GBR)	0	89	92	91	94
Finnish in Finland (FIN)	0	93	99	99	100
Iberian populations in Spain (IBS)	0	14	107	107	107
Toscani in Italy (TSI)	66	98	108	107	110
Utah residents with Northern and Western European ancestry (CEU)	94	85	99	99	103
Total European Ancestry (EUR)	**160**	**379**	**505**	**503**	**514**
Colombian in Medellin, Colombia (CLM)	0	60	94	94	95
Mexican Ancestry in Los Angeles, California (MXL)	0	66	67	64	69
Peruvian in Lima, Peru (PEL)	0	0	86	85	86
Puerto Rican in Puerto Rico (PUR)	0	55	105	104	105
Total Americas Ancestry (AMR)		**181**	**352**	**347**	**355**
Total	**553**	**1092**	**2535**	**2504**	**2577**

Table 1.1: Detailed list of samples genotypes for 1000Genomes Project
(Reprinted from http://www.1000genomes.org/about)

The human genome project was a remarkable effort that manages to sequence the entire genome; however the next step that attributes function to each and every region of the genome was not straightforward. Due to the fact that majority of the genome consists of non-coding portion, it was needed to somehow devise a strategy in order to decode the information. In September 2003, National Human Genome Research Institute (NHGRI) initiated a project termed as **Enc**yclopedia **o**f **D**NA **E**lements (ENCODE) [13-16] – aimed at identifying almost all functional elements in the human genome. As part of the pilot phase, the ENCODE project utilizes diverse experimental, computational and evolutionary data in order to define functional elements in 1% of the genome [14]; moreover the findings of ENCODE project has several advantages: determine changes associated with the chromatin region such as methylation and acetylation patterns, identify critical histone marks (H3K4me1, H3k4me2, H3K4me3 and H3K27ac [23-25]) in a wide variety of human cell types, predicts particular DNAse I hypersensitive sites, and locate particular segments of the non-coding genome involved in controlling normal gene transcription. Besides, the physical location of enhancer elements is cell-type specific [26] suggesting that different cells selectively activate particular regulatory regions of the genome in order to maintain their integrity [27]. Thus, it becomes easier to map genetic variants to putative regulatory portions of the genome using experimental datasets from multiple cell types. Figure 1.7 shows an overview of the regulatory elements identified by ENCODE. Nevertheless, the ENCODE project has already made substantial progress in locating the genes, transcripts and regulatory elements of the human genome; it also maps potential chromatin states as well as different DNA methylation patterns associated with the genome [15]. Finally, the ENCODE findings facilitate the annotation of non-coding variants characterized by the GWA studies [28], and the examination of gene regulatory mechanisms. This will further promote better understanding of the disease pathogenesis in the post-GWAS era.

1.3. Genome-wide Association Studies (GWAS)

Once the HapMap [6-8] and 1000Genomes Project [9-12] successfully identified millions of SNPs, the stage was set for another exciting work known as GWAS [17-19]; this sort of approach is based on the idea of surveying complete genome of many individuals in search of finding correlations between SNPs and diseases [20]. In order to perform GWAS, two group of individuals are selected: a group of cases (people with the disease) and a second group of normal controls; then SNPs were analyzed in both the groups. Some genetic variations are observed more often in cases than controls suggesting their association with the disease [20]. These variations basically mark that part of the genome, which plays a crucial role in the onset of the disease. Once these associations are identified, follow up strategies contribute towards better understanding, treatment and prevention of disorders; in this way, GWAS have already laid the foundation that ultimately drives the notion of personalized medicine [20]. Consequently, individualized patient treatment will soon become a reality where we have at our disposal, knowledge of individual's risk of developing a disease, or response against a particular drug [20]. Meanwhile, more than 1500 GWA studies have been published since 2005; however more than 80% of the SNPs identified to be associated with the diseases lie in the non-coding part of the genome; associated variants are also linked with the nearby SNPs

making it difficult to identify the causal variant among apool of linked SNPs, this phenomenon is termed as linkage disequilibrium [20]. Therefore, scientists look for the causal variants in the post-GWAS era utilizing additional rare variants identified as a result of 1000Genomes Project.

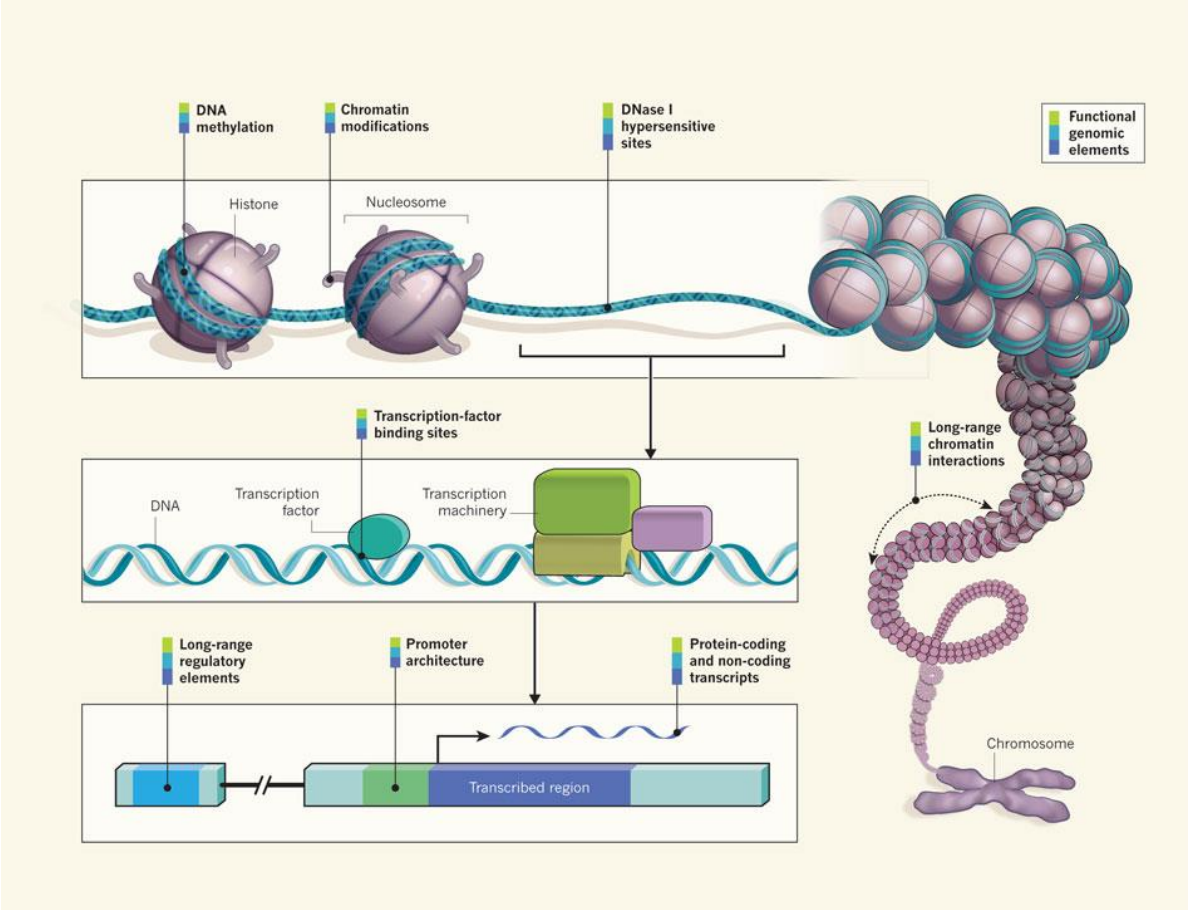

Figure 1.7: An overview of the regulatory elements located by ENCODE Project
(Reprinted from ENCODE Consortium, Bernstein et al., 2012)

1.4. Regulatory Variants and their impact on cellular physiology

In the past, non-coding portion of the genome was ignored in the sense that almost no information was available regarding that so-called "junk region" of the genome. However, recent advances in sequencing technologies have already benefitted scientific research such as the initiation and progress of International HapMap Project [6-8] and 1000Genomes Project [9-12]. Subsequently the ENCODE Project [13-16] adds to it, an enormous amount of data particularly attributed to the non-coding genomic portion. Due to this, growing evidence has suggested that non-coding variations – specifically those residing close to the coding region – are involved in controlling different gene regulatory mechanisms and chromatin structures of the genome; these variations are often termed as regulatory variants and are also associated with differential gene expression patterns [29]. Recent evidences generated as a result of Genome-wide association studies has also suggested that regulatory SNPs may have a role in controlling quantitative gene expression levels. Variations may change the expression levels among individuals by altering the binding affinity of transcriptional machinery; it may also

affect chromatin structure besides influencing the methylation pattern [29] of relevant genomic loci. In this scenario, one could argue that the GWAS results – that marks more than 80% of the variants in the non-coding region of the genome – have a substantial influence on disease phenotype. Meanwhile, Expression Quantitative Trait Loci (eQTLs) are specific regions of the genome whose expression levels are altered by genetic variations. Different experiments have already found large number of eQTLs for many human genes in different cell types [30]. With the availability of eQTL information, the functional annotation of regulatory variants in post-GWAS analysis has certainly improved. However, due to the phenomenon of linkage disequilibrium, one may find it difficult to pinpoint the causal variant that contributes to eQTL among a pool of possible variants [30]. However, with the successful progress of 1000Genomes Project [9-12] many rare variants have been identified that will solve this problem. Nevertheless, eQTLs will definitely prove to be quite effective in determining causal variants in post-GWA studies; it also helps in providing an insight of different gene regulatory mechanisms linked with disease pathogenesis.

Chapter 2

Materials and Methods

2. MATERIALS AND METHODS

2.1. NHGRI GWAS Catalog

Over the years, there is significant increase in the number of GWA studies published in peer-reviewed journals; however very little genotypic data were available publicly to the scientific community. In this context, National Human Genome Research Institute (NHGRI) initiated a project to provide public access to each and every GWA study that gets published. This initiative then develops a manually curated catalog of GWA studies [31]; more than 1500 published GWA studies have been maintained in the catalog [31]. Still, a major challenge lies ahead: over the past few years there is an increase in the complexity of GWA studies, mostly the studies include gene-gene interactions and gene-environment associations; therefore GWAS Catalog needs to accommodate all these updates and the associated web interface needs further improvements as well [31].

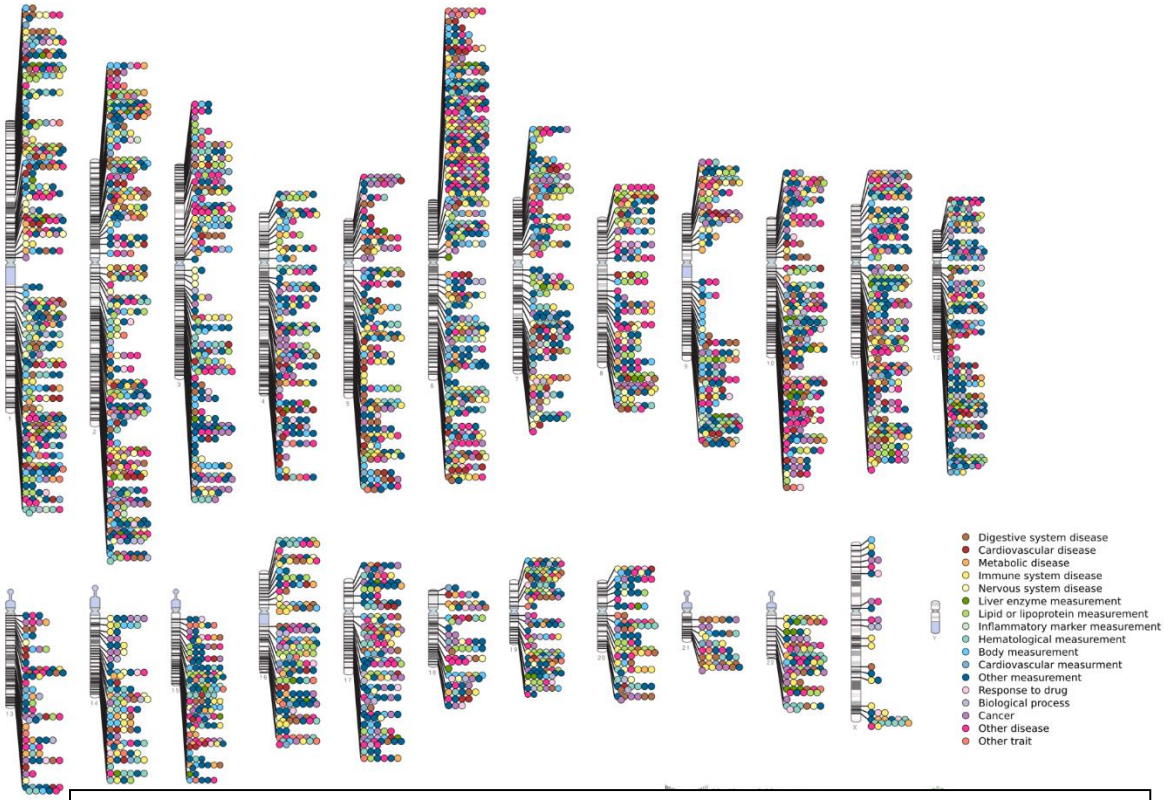

Figure 2.1: NHGRI GWAS Catalog, Published Genome-wide Association Studies (2013)
(Reprinted from http://www.genome.gov/multimedia/illustrations/GWAS_2013-12.pdf)

We considered 3 GWA studies of Aging [32-34] from the NHGRI GWAS Catalog; a total of 20 genome-wide significant SNPs were selected initially belonging to different risk loci (Table 2.1). However, SNPs present in strong LD ($r^2 \geq 0.80$) with the GWAS lead SNPs were also included using Haploreg [35] tool: the same ethnic groups – studied in the reported GWA studies of aging – were selected. The overall number of variants then becomes 600, referred to as LD80 SNPs (Table S1).

Chr	Pos (hg19)	SNP	Ref	Alt	RefSeq genes	dbSNP functional annotation
6	161333937	rs1247318	G	C	79kb 5' of *MAP3K4*	----------
5	162998515	rs294588	T	C	52kb 3' of *MAT2B*	----------
6	858970	rs1572438	C	T	102kb 3' of *LOC285768*	----------
6	80466124	rs6925255	A	G	14kb 3' of *RNY4*	----------
1	229019148	rs11585386	C	T	137kb 3' of *RHOU*	----------
6	161933935	rs16892673	G	A	*PARK2*	intronic
7	93691744	rs9918668	G	A	58kb 5' of *BET1*	----------
12	129475939	rs643473	A	G	6.4kb 3' of *GLT1D1*	----------
2	141721142	rs12474609	A	T	*LRP1B*	intronic
12	51703834	rs766903	A	G	*BIN2*	intronic
3	162681995	rs1425609	G	A	213kb 3' of *LOC647107*	----------
11	124017493	rs4936894	G	A	*VWA5A*	3'-UTR
2	238270894	rs10202497	C	A	*COL6A3*	intronic
1	44215828	rs2367725	C	A,T	*ST3GAL3*	intronic
19	3927771	rs10412199	G	A	*ATCAY*	3'-UTR
5	152639677	rs3112530	A	G	230kb 5' of *GRIA1*	----------
12	14115482	rs4764043	C	T	*GRIN2B*	intronic
13	48387722	rs8001976	C	T	129kb 3' of *SUCLA2*	----------
3	168686676	rs16852912	C	T	115kb 3' of *MECOM*	----------
1	80734581	rs11162963	C	T	1.3Mb 5' of *ELTD1*	----------

Table 2.1: Details of 20 Genome-wide significant SNPs selected for study

2.2. Regulatory Variant Annotation Tools:

2.2.1. Haploreg

Haploreg [35] contains information about SNPs present in LD with the GWAS lead SNPs: utilizes evolutionary conserved genome sequences across mammals (GERP and SiPhy scores), considers epigenomic alterations associated with the variants, uses information coming from 1000Genomes Project to demonstrate SNPs present in LD, utilize ENCODE datasets to visualize SNPs and their predictive chromatin states in several cell types, and uses experimental datasets (such as position weight matrices (PWMs) generated from JASPAR, TRANSFAC and other protein binding microarrays) to predict changes affecting regulatory motifs and their binding affinities. Open chromatin regions were detected by DNAse I hypersensitive sites; possible promoter regions marked with H3K4me3 peaks were identified using Chip-seq data, whereas potential enhancer regions were also discovered using H3K4me1 and H3K27ac peaks.

2.2.2. RegulomeDB

RegulomeDB [36] investigates regulatory variants present in the human genome. Currently, it is equipped with high-throughput experimental data – coming from the ENCODE Project [13-16]. Several computational and manually performed annotations are also included in the knowledgebase of RegulomeDB. Experimentally performed annotations include: Chip-seq data for numerous transcription factors across many cell types, chromatin state information across various cell types, and expression quantitative trait loci (eQTL) information helping in the functional annotation of variants. Computational predictions involve: DNAse footprinting, which allows accurate examination of protein binding regions and nucleotide variations

responsible for binding motif alterations. Subsequently, using the predictions and relevant annotations – a scoring scheme was developed, allotting variants with particular scores. A summary of RegulomeDB scoring scheme is shown in Table 2.2. In this way, it becomes feasible to filter out potential regulatory variants from a large group of variants reported in the GWAS.

Category	Description
	Likely to affect binding and linked to expression of a gene target
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
	Likely to affect binding
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
	Less likely to affect binding
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
	Minimal binding evidence
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit
**** Lesser the scores, more likely it would be that variant lies within a potential functional region***	
Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22: 1790–1797.	

Table 2.2: Summary of RegulomeDB scoring scheme

2.2.3. rSNPBase

Functional interpretation of regulatory variants in the human genome was further facilitated by rSNPBase [37]. Using ENCODE and other experimental resources, it labels SNPs using 4 different criteria: proximal transcriptional regulation, distal transcriptional regulation, RNA binding protein-mediated post-transcriptional regulation and miRNA mediated post-transcriptional regulation. It also informs about eQTLs for regulatory SNPs.

In this study, we initially employed RegulomeDB [36] to score variants followed by the filtration of particular variants having strong regulatory potential (RegulomeDB score < 3). Furthermore, the filtered variants were annotated to discover their potential causal link with the disease pathogenesis, using Haploreg [35] and rSNPBase [37].

2.2.4. Supporting Databases

Finally, few other tools also played a role in deciphering disease-SNP association. The corresponding tools include a protein-protein interaction prediction database STRING [38] that locates physical and indirect functional associations, develops broad protein networks covering more than 1100 living organisms, and makes good use of 4 different sources such as previous knowledgebase, conserved co-expression data, high-throughput testing and genomic context of genes. Besides, we also considered Gene Network (www.genenetwork.nl) – to detect any shared pathways between risk loci for complex diseases. In addition to this, DISEASES [39] database helps in the identification of any gene-disease associations extracted from literature. These extractions were promoted by automatic text mining, manually curated literature, cancer mutation data, and GWAS; it also allot confidence scores to different types of evidences present in the database that further aids in sorting the best one.

Chapter 3

Results

3. RESULTS

3.1. Potential SNPs identification using RegulomeDB

We examined 600 variants using RegulomeDB: out of which 291 returned RegulomeDB scores of 1-6; whereas remaining variants had scores of "No Data". Table S2shows a detailed list of all variants and their scores – with 4 variants possessing strong regulatory potential having the RegulomeDB score < 3 (rs3791051, rs9822393, rs159972 and rs932492). Unexpectedly, we failed to find any genome-wide significant variant among the potential variants (Table 3.1); each variant has the RegulomeDB score of 2b (likely to affect binding).

3.2. Functional annotation of potential SNPs

Using bioinformatics tools, we found that rs3791051 is an intronic variant located in the *ST3GAL3* gene, within a DNAse I hypersensitive site reported in about 31 different cell types. Sequence constraint information suggests that it lies in a conserved genomic location across mammals. Chip-seq data indicate, variant is situated within the binding site of MAX, HNF4A, USF1, MYC, BHLHE40, MXI1 and USF2 protein. Histone modification data confirm its presence within a transcriptionally active locus in multiple cell lines including hepatocellular carcinoma (HepG2), neuroblastoma (SK-N-SH-RA), astrocytes in brain and lymphoblastoid (GM12878). Moreover, the variant significantly disrupt E2A/TCF3, MYF5, NRSF/REST and STAT3 transcription factor binding site. According to rSNPBase, the SNP locus is associated with the distal transcriptional regulation of *ARTN*; this SNP is an eQTL for *ST3GAL3* and *MOB3C*.

Chr	Pos	Functional SNP	Refseq gene	*LD (r^2)	RegulomeDB score	GWAS lead SNP	Refseq gene
1	44252908	rs3791051	*ST3GAL3*	0.87	2b	rs2367725	*ST3GAL3*
3	162702771	rs9822393	192kb 3' of *LOC647107*	0.83	2b	rs1425609	213kb 3' of *LOC647107*
5	152732406	rs159972	138kb 5' of *GRIA1*	0.88	2b	rs3112530	230kb 5' of *GRIA1*
6	80474572	rs932492	23kb 3' of *RNY4*	0.84	2b	rs6925255	14kb 3' of *RNY4*

Linkage disequilibrium between functional SNPs and GWAS lead SNPs were reported by Haploreg.

Ward, L.D. and Kellis, M. (2012) Haploreg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res., 40, D930–D934.

Table 3.1: Details of potential regulatory SNPs reported by RegulomeDB with score < 3

Similarly, rs9822393 lies immediately 3′ to the *LOC647107* gene – within a DNAse I hypersensitive region – altering GR/NR3C1 and KLF4 transcription factor binding sites. Histone marks spotted this variant in an active locus in Hela-S3 and choroid plexus epithelial cell line (Hcpe). This variant also overlaps binding sites of MAX and POLR2A in cervical carcinoma cell-line (Hela-S3).

Moreover, in case of rs159972,it is present 138kb 5′ to the *GRIA1* transcription start site – in an evolutionary conserved region of the genome – within binding sites of 5 proteins namely BATF, EBF1, EGR1, PBX3 and SPI1 in B-lymphocytes/lymphoblastoid cell lines; it changes

the IRF1, IRF2 and NFKB transcription factor binding site, and situated in an active locus in multiple cell lines such as astrocytes in brain, B lymphocytes/lymphoblastoid (GM12878), osteoblasts (Osteobl) and embryonic stem cells (H1 hESC) etc.

According to RegulomeDB, rs932492 is likely to affect binding with RegulomeDB score of 2b: present immediately to the RNY4 transcription start site, changes the binding affinity of ATF3 and HSF1, lies in a transcriptionally active locus in the embryonic stem cell line (H1 hESC), and maps to a locus where GATA1 transcription factor binds. Table 3.2 highlights the annotated information regarding variants.

Chr	Pos	SNP	Ref	Alt	Gene	eQTL (genes)	Proteins bound (transcription factor binding site)	Motifs altered	Ref	Alt	Histone marked active loci (Cell types)
1	44252908	rs3791051	C	T	*ST3GAL3*	*ST3GAL3* *MOB3C*	HNF4A, MAX, MXI1, c-MYC, BHLHE40, USF1, USF2	Ascl2 E2A/TCF3 MYF5 NRSF/REST PU1 PAX6 STAT3	12.3 13.7 13.4 -17.6 11.3 8.7 14	10.8 2.6 2.5 -5.6 11.3 7 2.7	astrocytes (brain), HUVEC, HSMM, GM12878, NHLF, HepG2, Osteobl, NHDF-Ad, A549, K562, Monocytes-CD14+, Dnd41
3	162702771	rs9822393	C	T	192kb 3' of *LOC647107*	----------	POLR2A, MAX	CTCF GR/NR3C1 GR/NR3C1 KLF4	11.5 12.2 12.4 11.5	11.4 0.9 11.4 5.9	HeLa-S3, HCPEpiC
5	152732406	rs159972	G	A	138kb 5' of *GRIA1*	----------	SPI1, EBF1, BATF, EGR1, PBX3	IRF1 IRF2 IRF2 IRF1 IRF NFKB PAX4 PAX4	3.6 -10.2 -13.1 3.5 14 2.3 11.5 11.7	15.6 1.1 -1.2 15.3 14.4 14.3 11.8 12.7	HMEC, GM12878, HSMMtube, H1-hESC, NH-A, NHEK, Osteobl, HEK293, HVMF,AG10803, MCF-7, NHLF
6	80474572	rs932492	C	T	23kb 3' of *RNY4*	----------	GATA1	ATF3 HSF1	-9.3 11.9	-21.2 -0.1	H1-hESC

Joint investigation of variants was performed by RegulomeDB, Haploreg and rSNPBase.

Table 3.2: Annotation of potential regulatory variants using bioinformatics tools

Chapter 4

Discussion

4. DISCUSSION

We identified four SNPs that have a strong regulatory potential (score < 3), but none of them were the genome-wide significant SNP – nevertheless, the potential variants were in strong LD with the reported GWAS lead SNPs (Table 3.1). Moreover, we found rs3791051 as a potential regulatory variant confirmed by all 3 databases (RegulomeDB, Haploreg and rSNPBase). This SNP resides in *ST3GAL3*; this gene associates with few age-related disorders such as cancer [40-43] and amyotrophic lateral sclerosis [44], and express most often in the hypothalamus, thalamus, putamen and heart ventricles.

According to Haploreg, rs3791051/ST3GAL3 changes the binding affinity of STAT3 (transcription factor) in mammary gland, non-tumorigenic epithelial inducible cell line (MCF10A-Er-Src). We also found, STAT3 interacts with BHLHE40 [45] sharing JAK-STAT signaling pathway. Possibly, this variant may affect *STAT3*-mediated *BHLHE40* binding. However, it is not yet clear whether this activity will increase or decrease the normal functioning of *ST3GAL3*. Multiple types of cancer (one of the major hallmarks of aging) and Alzheimer's disease also relates with *STAT3* [46-48]; so in this perspective, *STAT3* may influence *ST3GAL3* expression. However, Gene Network failed to report any association between *ST3GAL3* and *STAT3*.

STRING also shows that STAT3 interacts with c-MYC (one of the target genes of STAT3) in promoting the transition of cell cycle from G1 to the S-phase [49]. In this regard, it could be possible that the reduction in STAT3 binding affinity due to rs3791051 may affect ST3GAL3 transcription; however this needs to be investigated experimentally. We also found that rs3791051/*ST3GAL3* reduces the binding affinity of TCF3 (transcription factor). According to GeneCards [50], TCF3 initiates neuronal differentiation and associates with multiple types of cancers [51-52]. As a result, TCF3 may interact with ST3GAL3; however we failed to detect any shared pathway between the two genes.

We also identified that rs3791051 decreases the binding affinity of MYF5, which is a transcriptional activator that plays a part in muscle cells differentiation; it also associates with numerous disorders including neuropathy, which is a nervous system disorder located in nerves (another hallmark of aging). Consequently, it may indirectly relate with *ST3GAL3*. Again, Gene Network failed to provide any shared pathway between the two genes. Apart from that, TCF3 and MYF5 also interact with each other [53]; our SNP of interest rs3791051 may affect this interaction – which could indirectly affect *ST3GAL3* expression. Furthermore, rs3791051 affects NRSF/REST; Haploreg reports a three-fold decline in the binding affinity of NRSF/REST (Table 3.2). A transcriptional repressor – REST – reduces the expression of neuronal genes in nervous system and involves negative regulation during neurogenesis: alterations in *REST* associates with Huntington's disease [54], a neurodegenerative disorder (often seen in aged people) affecting normal muscle functioning that leads to dementia. Thus, *REST* may cause aberrant functioning of *ST3GAL3* in promoting age-associated phenotype; *REST* also interacts with Huntingtin *HTT* [55], a gene that is reported to be involved in the on-set of Huntington's disease. Again, one can argue in this context; *ST3GAL3* may interact with the *HTT*, however Gene Network didn't spotted any shared pathway between the two genes.

Another transcription factor, HNF4A, is affected by rs3791051 in cancer cell lines (HepG2 and Caco-2). This transcription factor plays a role in the development of liver, kidney and intestine. Moreover, mutations in this gene associate with MODY [56-57], cancer [58-60], and type-II diabetes mellitus [56, 61-62] etc. Diabetes mellitus with insulin resistance in the body relates with Alzheimer's disease [63], so *HNF4A* may affect the development of age-related pathologies – which need to be investigated in future. Alongside this, rs3791051affects the binding of three more proteins: MAX, MYC and MXI1. These proteins frequently associate with cancer [64-68] and affected in HepG2 cells. So it could be that the corresponding genes somehow interact with the *ST3GAL3* gene.

There exists few histone marks that suggest active gene loci; in case of rs3791051/*ST3GAL3*, it resides in an actively labeled genomic locus in multiple cell types (Table 3.2) including cancer cell types and astrocytes in brain. The variant, rs3791051, also regulates distal transcriptional regulation of *ARTN*: this gene correlates with multiple types of cancer [69-70] as well as neurological disorders [71-72]. In this scenario, *ARTN* may associate with *ST3GAL3* in promoting age-related disorders. This hypothesis needs further examination in future.

Finally, we also observed that rs3791051 is an eQTL for *ST3GAL3* and *MOB3C*, which confirms the change in expression level of *ST3GAL3* – the SNP, rs3791051 somehow correlates with aging phenotype. Whereas, according to GeneCards, MOB3C is a MOB Kinase activator 3C, which is similar to the yeast MOB1 protein. In yeast, this protein kinase controls the cell cycle. Consequently, it may contribute towards tumor formation and act as a potential player in aging, which needs further investigation.

According to Haploreg, rs9822393 significantly reduces the binding affinity of NR3C1/GR in lung carcinoma tissues (A549): NR3C1/GR can act as a transcription factor that binds to glucocorticoid response elements and regulate other transcription factors, and also associates with several age-related disorders such as: Cancer [73-75], Diabetes mellitus [76-77], Osteoporosis [77-78], Arthritis [79-80], and Schizophrenia [81-82] etc. Keeping this in view, it may interact with the SNP locus and may play a role in the on-set of age-related pathologies. Interestingly, *GR* (glucocorticoid receptor) interacts with *SMARCA4* [83] – another cancer risk gene [84-85] – while both share chronic myeloid leukemia pathway. Subsequently, this indicates a potential functional link between the SNP locus and *SMARCA4* that needs to be studied. Another transcription factor whose binding affinity is reduced at rs9822393 locus isKLF4; this gene correlates with numerous age-related disorders including cancer [86-87], atherosclerosis [88] and neurodegenerative disease [89]. In this context, it may interact with ST3GAL3 in promoting aging phenotypes. Additionally, rs9822393 also affect MAX and POLR2A binding in cervical carcinoma cells; both proteins are associated with cancer [64-65, 90-91]. Again, the SNP locus may relate with MAX and POLR2A, which needs experimental verification. Keeping this information and the observation regarding the presence of this SNP in an active locus in choroid plexus epithelial cells (HCPE) and cervical carcinoma cells (HelaS3) as shown in table 3.2, we suggest that the variant may directly or indirectly affects cancer and nervous system disorders.

We also found rs159972 as a potential variant (RegulomeDB score of 2b). Histone modification data suggests that it lies in a transcriptionally active locus in multiple cell lines including astrocytes in brain, embryonic stem cells and skeletal muscle myotubes. Haploreg informs that the variant is present near *GRIA1* transcription start site; *GRIA1* associates with numerous age-related neurodegenerative disorders including Alzheimer's disease [92-93], Schizophrenia [94-95], Parkinson's disease [96] and Vascular disease [97-98]. The variant, rs159972, alters the binding affinity of NFKB in B lymphocytes/lymphoblastoid cell line (GM10847). NFKB1 has been demonstrated in literature to correlate with some age-related pathologies including Schizophrenia [99] and vascular disease [100]. In this regard, it may interact with GRIA1 in promoting age-related diseases, meanwhile we failed to report any shared pathway between the two genes. However, NFKB1 interacts with RELA [101] (STRING) and shares NF-KB signaling pathway according to BioCarta (http://genenetwork.nl). RELA also correlate with vascular disease [102] and neurodegenerative disease [103] suggesting a potential functional link between RELA and GRIA1; still Gene Network didn't report any shared pathway. Alongside this, the binding of EGR1 is also affected by the SNP in B-lymphocytes/lymphoblastoid cells. DISEASES reports that *EGR1* correlates with several age-related diseases that includes: Vascular disease [104], Alzheimer's disease [105], Heart disease [106-107], and Schizophrenia [108]. This suggests potential interaction between GRIA1 and EGR1; however no shared pathway was detected.

Finally, rs932492 is also shown as a potential regulatory variant by RegulomeDB. It is located in a transcriptionally active locus in embryonic stem cells (H1 hESC). Furthermore, it is situated near *RNY4* transcription start site. The variant, rs932492, changes the ATF3 motif binding affinity; whereas ATF3 also associates with multiple age-related pathologies including: Cancer [109-110], Vascular disease [111], Amyotrophic lateral sclerosis [112] and Diabetes mellitus [113]. In this scenario, ATF3 may interact with RNY4 to enhance the aging process; however, it is yet to be investigated. Moreover, ATF3 interacts with DDIT3 [114] (STRING) while sharing p53 signaling pathway. *DDIT3* associates with several age-related phenotypes namely: Cancer [115], Diabetes mellitus [116], Alzheimer's disease [117] and Parkinson's disease [118] etc. Considering this, it could be that DDIT3 interacts with RNY4 in promoting aging; still this is not yet experimentally verified. Similarly, HSF1 binding affinity is also affected at rs932492 locus. HSF1 is reportedly linked previously with numerous age-related pathologies which include: Cancer [119], Huntington's disease [120], Amyotrophic lateral sclerosis [121], Parkinson's disease [121] and Alzheimer's disease [121] etc. This suggests potential functional link between *HSF1* and *RNY4* that needs to be analyzed in future.

In conclusion, we found few potential regulatory SNPs, which are associated mostly with the Neurological and Cancer related pathways. We also observed that related proteins interact with each other in shared biochemical pathways; aging may be delayed once we will be able to identify crucial pathways associated with it. Moreover, the current study utilized RegulomeDB to score variants according to the experimental evidence present. However, RegulomeDB did include some variants with score of "No Data" making it difficult to establish their involvement in aging. Nevertheless, Haploreg and rSNPBase improves the annotation of studied SNPs.

Utilizing LD SNPs alongside genome-wide significant SNPs will assist in moving from GWAS to disease pathways – making it feasible to fill the gap between genotype and phenotype.

4.1. Genomic Medicine gets Personal

In the post-genomic era, several exciting projects such as 1000Genomes Project and the renowned ENCODE Project have already revolutionized the field of medicine and personal healthcare. This has created enough room for the notion of personalized medicine to become a reality now. With an increased knowledge of patient's genetic as well as the epigenomic data, disease management would surely get better in near the future. Knowledge of genetic variations among the individuals from almost all parts of the globe has the potential to tailor conventional medicine for a patient's own needs. Genetic variations are also well known to contribute towards varying drug response among different individuals. Meanwhile, the current genomics research has failed to identify the causes of complex diseases. This has hampered the quest for better understanding of disease pathogenesis; the completion of Human Genome Project was a remarkable feat but the fact is we are still dealing with the problem of characterizing accurate structure of the human genome. Figure 4.1 shows where we are standing right now in terms of having an insight of disease mechanisms.

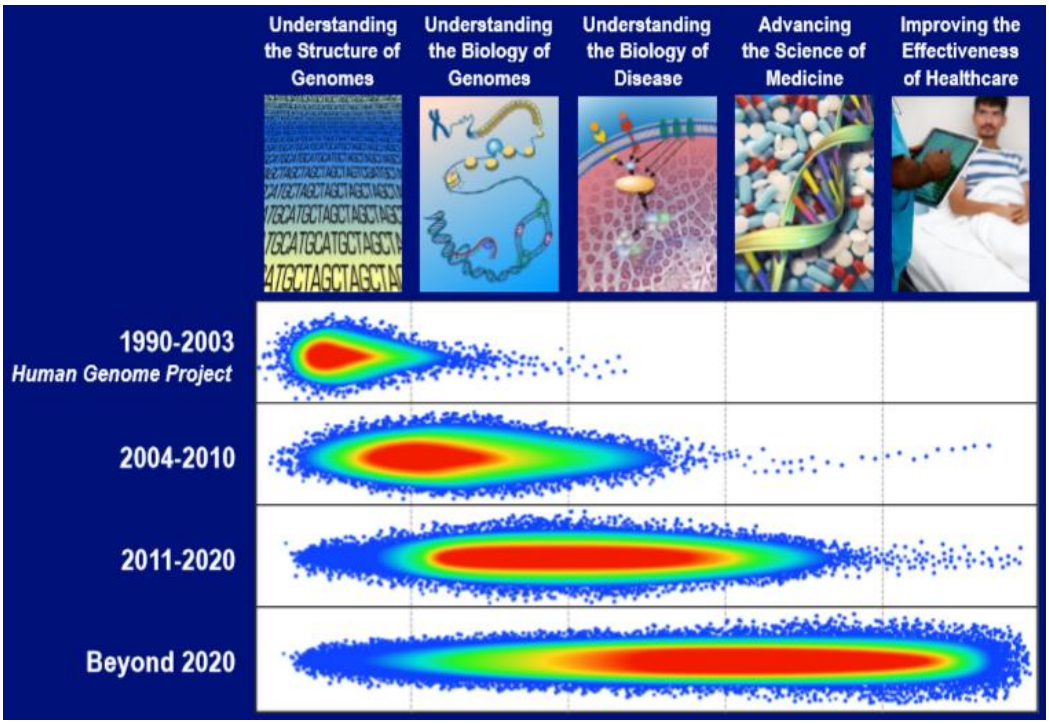

Figure 4.1: Genomics to Personalized Medicine: The Future looks promising
(Reprinted from http://www.genome.gov/12514288)

Since the completion of Human Genome Project, many parallel strategies were applied to gain further knowledge of how genome functions at different levels of organization; in doing so there are few well known efforts already discussed in previous chapters such as the International HapMap Project, Human Epigenome Project and the renowned one – the ENCODE Project. During the last few years, focus of biomedical research has shifted towards

the analysis of epigenomic changes associated with the human genome. This was the basic idea behind ENCODE Project, and due to this we already have gain momentum towards better analysis and knowledge of crucial cellular pathways. With the initiation of 100,000Genomes Project – genotyping cancer patient's genome to identify crucial variations – this is surely an exciting time to be part of the global scientific research.

CONCLUSION

In conclusion, we found few potential regulatory SNPs, which are associated mostly with the Neurological and Cancer related pathways. We also observed that related proteins interact with each other in shared biochemical pathways; aging may be delayed once we will be able to identify crucial pathways associated with it. Moreover, the current study utilized RegulomeDB to score variants according to the experimental evidence present. However, RegulomeDB did include some variants with score of "No Data" making it difficult to establish their involvement in aging. Nevertheless, Haploreg and rSNPBase improves the annotation of studied SNPs. Utilizing LD SNPs alongside genome-wide significant SNPs will assist in moving from GWAS to disease pathways – making it feasible to fill the gap between genotype and phenotype.

REFERENCES

1. Kirkwood TB (1999) Time of our lives: the science of human aging, 1st edn. Oxford University Press, New York.

2. Montesanto A, Dato S, Bellizzi D, Rose G, Passarino G (2012) Epidemiological, genetic and epigenetic aspects of the research on healthy ageing and longevity. Immun Ageing 9(1):6 Scully T (2012) To the limit. Nature 492:S2.

3. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).

4. Venter, J. C. et al. The sequence of the human genome. Science 291, 1304–1351 (2001).

5. The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945 (2004).

6. The International HapMap Consortium. The International HapMap Project. Nature 426, 789-796. 2003.

7. The International HapMap Consortium. A Haplotype Map of the Human Genome. Nature 437, 1299-1320. 2005.

8. The International HapMap Consortium. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52-58. 2010.

9. The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

10. 1000 Genomes Project Consortium, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo et al., 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 490: 56–65.

11. The 1000 Genomes Project more than doubles catalog of human genetic variation. (2013). In NIH News. Retrieved from http://www.genome.gov/27551417.

12. 1000 Genomes Project (2012). Retrieved from http://www.genome.gov/27528684.

13. The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306: 636–640. 17. Birney, E. et al.

14. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816 (2007)

15. The ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 9: e1001046.

16. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al: An integrated encyclopedia of DNA elements in the human genome. Nature 2012, 489(7414):57–74.

17. Hardy, J. & Singleton, A. Genome-wide association studies and human disease. N.Engl. J. Med.360, 1759–1768 (2009).

18. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012; 90: 7-24.

19. Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12): e1002822.

20. Genome-wide Association Studies (2014) Retrieved from http://www.genome.gov/ 20019523

21. Kuilman T, Michaloglou C, Mooi WJ, Peeper DS (2010). The essence of senescence.Genes Dev24 (22):2463.

22. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. Cell 153, 1194–1217 (2013).

23. Talbert PB, Henikoff S (2010) Histone variants–ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11: 264–275.

24. Pekowska A, Benoukraf T, Ferrier P, Spicuglia S (2010). A unique H3K4me2 profile marks tissue-specific gene regulation. Genome Res 20: 1493–1502.

25. Zentner GE, Tesar PJ, Scacheri PC (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 21: 1273–1283.

26. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108–112.

27. Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, et al. (2010) Genome-wide analysis of histone modifications in human pancreatic islets. Genome Res 20: 428–433.

28. Kellis, M. et al. Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. U.S.A. 111, 6131–6138 (2014).

29. Gongcheng Li, Tiejun Pan, Dan Guo, and Long-Cheng Li, "Regulatory Variants and Disease: The E-Cadherin −160C/A SNP as an Example," Molecular Biology International, vol. 2014, Article ID 967565, 9 pages, 2014.

30. Battle A, Montgomery SB: Determining causality and consequence of expression quantitative trait loci. Hum Genet 2014, 133:727-735.

31. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Research, 2014, Vol. 42, Database issue D1001–D1006.

32. Edwards DRV, Gilbert JR, Hicks JE, Myers JL, Guo S, Gallins PJ, et al. (2013) Linkage and association of successful aging to the 6q25 region in large Amish kindreds. AGE 35:1467–1477.

33. Poduslo SE, Huang R, Spiro A III. 2010. A Genome Screen of Successful Aging Without Cognitive Decline Identifies LRP1B by Haplotype Analysis. Am J Med Genet Part B 153B:114–119.

34. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, et al. (2011) A Genome-Wide Association Study of Aging. Neurobiol Aging 32(11): 2109.e15–2109.e28.

35. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res., 40, D930–D934.

36. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22: 1790–1797.

37. Guo L, Du Y, Chang S, Zhang K, Wang J (2014) rSNPBase: A database for curated regulatory SNPs. Nucleic Acids Res 42: D1033–9.

38. Franceschini,A., Szklarczyk,D., Frankild,S., Kuhn,M., Simonovic,M., Roth,A., Lin,J., Minguez,P., Bork,P., von Mering,C. et al. (2013) STRING v9.1: protein-protein

interaction networks, with increased coverage and integration. Nucleic Acids Res., 41, D808–D815.

39. DISEASES: Text mining and data integration of disease-gene associations. bioRxiv doi: 10.1101/008425.

40. Pousset, D., Piller, V., Bureaud, N., Monsigny, M., and Piller, F (1997) Increased alpha 2,6 sialylation of N-glycans in a transgenic mouse model of hepatocellular carcinoma. Cancer Res. 57, 4249–4256.

41. Kudo, T., Ikehara, Y., Togayachi, A., Morozumi, K., Watanabe, M., Nakamura, M., Nishihara, S., and Narimatsu, H. Up-regulation of a set of glycosyltransferase genes in human colorectal cancer. Lab. Investig., 78: 797–811, 1998.

42. Hebbar M, Krzewinski-Recchi MA, Hornez L, Verdiere A, Harduin-Lepers A, Bonneterre J, Delannoy P, Peyrat JP: Prognostic value of tumoral sialyltransferase expression and circulating E-selectin concentrations in node-negative breast cancer patients. Int J Biol Markers 2003, 18:116-122.

43. Gretschel S, Haensch W, Schlag PM, Kemmner W: Clinical relevance of sialyltransferases ST6GAL-I and ST3GAL-III in gastric cancer. Oncology 2003, 65(2):139-145.

44. ALSGEN Consortium, Ahmeti KB, Ajroud-Driss S, Al-Chalabi A, Andersen PM, et al. (2013) Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. Neurobiol Aging 34: 357.e7–19.

45. Ivanova AV, Ivanov SV, Zhang X, Ivanov VN, Timofeeva OA, Lerman MI. STRA13 interacts with STAT3 and modulates transcription of STAT3-dependent targets. J Mol Biol. 2004 Jul 16; 340(4):641-53.

46. Rivat C., S. Rodrigues, E. Bruyneel, G. Pietu, A. Robert, G. Redeuilh, M. Bracke, C. Gespach & S. Attoub: Implication of STAT3 signaling in human colonic cancer cells during intestinal trefoil factor 3 (TFF3) -- and vascular endothelial growth factor-mediated cellular invasion and tumor growth. Cancer Res 65, 195-202 (2005)

47. Klampfer L: The role of signal transducers and activators of transcription in colon cancer. Front Biosci 2008, 13:2888-99.

48. Wan J, Fu AK, Ip FC, Ng HK, Hugon J, Page G, Wang JH, Lai KO, Wu Z, Ip NY. Tyk2/STAT3 signaling mediates beta-amyloid-induced neuronal cell death: implications in Alzheimer's disease.J Neurosci. 2010 May 19; 30(20):6873-81.

49. Wang H, Lafdil F, Kong X, Gao B. Signal transducer and activator of transcription 3 in liver diseases: a novel therapeutic target. Int J Biol Sci 2011; 7(5):536-550.

50. Safran M, Solomon I, Shmueli O, Lapidot M et al.: GeneCards 2002: towards a complete, object-oriented, human gene compendium. Bioinformatics 2002, 18(11):1542-1543.

51. Tijchon E, Havinga J, van Leeuwen FN, Scheijen B. B-lineage transcription factors and cooperating gene lesions required for leukemia development. Leukemia. 2013 Mar; 27(3):541-52.

52. McWhirter JR, Neuteboom ST, Wancewicz EV, Monia BP, Downing JR, Murre C. Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia. Proc Natl Acad Sci U S A. 1999 Sep 28; 96(20):11464-9.

53. Braun T, Winter B, Bober E, Arnold HH. Transcriptional activation domain of the muscle-specific gene-regulatory protein myf5. Nature. 1990 Aug 16; 346(6285):663-5.

54. Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D, Cattaneo E. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. Nat Genet. 2003 Sep; 35(1):76-83. Epub 2003 Jul 27.

55. Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D, Cattaneo E. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. Nat Genet. 2003 Sep; 35(1):76-83.

56. Thanabalasingham G, Huffman JE, Kattla JJ, Novokmet M, Rudan I, et al. (2013) Mutations in HNF1A result in marked alterations of plasma glycan profile. Diabetes 62: 1329–1337.

57. Frayling TM, Evans JC, Bulman MP, Pearson E, Allen L, Owen K, Bingham C, Hannem ann M, ShepherdM, Ellard S, Hattersley AT.(2001) Beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. Diabetes 50, Suppl1:S94–S100.

58. Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, et al. (2009) Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet 41: 1330–1334.

59. Pugh, T.J. *et al*. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 488, 106–110 (2012).

60. Walesky C, Edwards G, Borude P, Gunewardena S, O'Neil M, Yoo B, Apte U. Hepatocyte nuclear factor 4 alpha deletion promotes diethylnitrosamine-induced hepatocellular carcinoma in rodents.Hepatology. 2013 Jun; 57(6):2480-90.

61. Cho YS, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in East Asians. Nat. Genet. 2012; 44:67–72.

62. Kooner, S. J. *et al*. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* 43, 984–989 (2011).

63. de la Monte SM, Tong M. (2014) Brain metabolic dysfunction at the core of Alzheimer's disease. Biochem Pharmaco 88: 548–559. , Review.

64. Crona J, Maharjan R, Delgado Verdugo A, Stålberg P, Granberg D, Hellman P, Björklund P. MAX mutations status in Swedish patients with pheochromocytoma and paraganglioma tumours. Fam Cancer. 2014 Mar; 13(1):121-5.

65. Wenzel A, Schwab M. The mycN/max protein complex in neuroblastoma. Short review. Eur J Cancer. 1995; 31A (4):516-9.

66. Li Y, Guessous F, Johnson EB, Eberhart CG, Li XN, Shu Q, Fan S, Lal B, Laterra J, Schiff D, Abounader R. Functional and molecular interactions between the HGF/c-Met pathway and c-Myc in large-cell medulloblastoma.Lab Invest. 2008 Feb; 88(2):98-111.

67. Solé X, Hernández P, de Heredia ML, Armengol L, Rodríguez-Santiago B, Gómez L, Maxwell CA, Aguiló F, Condom E, Abril J, Pérez-Jurado L, Estivill X, Nunes V, Capellá G, Gruber SB, Moreno V, Pujana MA. Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility. BMC Genomics. 2008 Jan 11; 9:12.

68. Engstrom LD, Youkilis AS, Gorelick JL, Zheng D, Ackley V, Petroff CA, Benson LQ, Coon MR, Zhu X, Hanash SM, Wechsler DS. Mxi1-0, an alternatively transcribed Mxi1 isoform, is overexpressed in glioblastomas. Neoplasia. 2004 Sep-Oct; 6(5):660-73.

69. Liebl F, Demir IE, Rosenberg R, Boldis A, Yildiz E, Kujundzic K, Kehl T, et al. The severity of neural invasion is associated with shortened survival in colon cancer. Clin Cancer Res. 2013 Jan 1; 19(1):50-61.

70. Banerjee A, Wu ZS, Qian P, Kang J, Pandey V, Liu DX, Zhu T, Lobie PE. ARTEMIN synergizes with TWIST1 to promote metastasis and poor survival outcome in patients with ER negative mammary carcinoma. Breast Cancer Res. 2011; 13(6):R112.

71. Bespalov MM, Saarma M. GDNF family receptor complexes are emerging drug targets. Trends Pharmacol Sci. 2007 Feb; 28(2):68-74.

72. Ceyhan GO, Schäfer KH, Kerscher AG, Rauch U, Demir IE et al. Nerve growth factor and artemin are paracrine mediators of pancreatic neuropathy in pancreatic adenocarcinoma. Ann Surg. 2010 May; 251(5):923-31.

73. Roca R, Kypta RM, Vivanco Md. Loss of p16INK4a results in increased glucocorticoid receptor activity during fibrosarcoma development. Proc Natl Acad Sci U S A. 2003 Mar 18; 100(6):3113-8.

74. Sinclair AJ, Jacquemin MG, Brooks L, Shanahan F, Brimmell M, Rowe M, Farrell PJ. Reduced signal transduction through glucocorticoid receptor in Burkitt's lymphoma cell lines. Virology. 1994 Mar; 199(2):339-53.

75. Teulings FA, van Gilse HA. Demonstration of glucocorticoid receptors in human mammary carcinomas. Horm Res. 1977; 8(2):107-16.

76. Jacobson PB, von Geldern TW, Ohman L, Osterland M, Wang J.; et al. Hepatic glucocorticoid receptor antagonism is sufficient to reduce elevated hepatic glucose output and improve glucose control in animal models of type 2 diabetes. J Pharmacol Exp Ther. 2005 Jul; 314(1):191-200.

77. Chrousos GP, Kino T. Glucocorticoid action networks and complex psychiatric and/or somatic disorders. Stress. 2007 Jun; 10(2):213-9.

78. Liu YZ, Dvornyk V, Lu Y, Shen H, Lappe JM, Recker RR, Deng HW. A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes. J Biol Chem. 2005 Aug 12; 280(32):29011-6.

79. Gossye V, Elewaut D, Bougarne N, Bracke D, Van Calenbergh S, Haegeman G, De Bosscher K. Differential mechanism of NF-kappaB inhibition by two glucocorticoid receptor modulators in rheumatoid arthritis synovial fibroblasts. Arthritis Rheum. 2009 Nov; 60(11):3241-50.

80. Tohyama CT, Yamakawa M, Murasawa A, Nakazono K, Ishikawa H. Localization of human glucocorticoid receptor in rheumatoid synovial tissue of the knee joint. Scand J Rheumatol. 2005 Nov-Dec; 34(6):426-32.

81. Guidotti A, Auta J, Davis JM, Dong E, Gavin DP, Grayson DR, Sharma RP, Smith RC, Tueting P, Zhubi A. Toward the identification of peripheral epigenetic biomarkers of schizophrenia. J Neurogenet. 2014 Mar-Jun; 28(1-2):41-52.

82. Sinclair D, Fillman SG, Webster MJ, Weickert CS. Dysregulation of glucocorticoid receptor co-factors FKBP5, BAG1 and PTGES3 in prefrontal cortex in psychotic illness. Sci Rep. 2013 Dec 18; 3:3539.

83. Wallberg AE, Neely KE, Hassan AH, Gustafsson JA, Workman JL, Wright AP. Recruitment of the SWI-SNF chromatin remodeling complex as a mechanism of gene activation by the glucocorticoid receptor tau1 activation domain. Mol Cell Biol. 2000 Mar; 20(6):2004-13.

84. Jones DT, Jäger N, Kool M, Zichner T, Hutter B, Sultan M, Cho YJ, Pugh TJ, et al. Dissecting the genomic complexity underlying medulloblastoma. Nature. 2012 Aug 2; 488(7409):100-5.

85. Bartlett C, Orvis TJ, Rosson GS, Weissman BE. BRG1 mutations found in human cancer cell lines inactivate Rb-mediated cell-cycle arrest. J Cell Physiol. 2011 Aug; 226(8):1989-97.

86. Evans PM, Zhang W, Chen X, Yang J, Bhakat KK, Liu C. Kruppel-like factor 4 is acetylated by p300 and regulates gene transcription via modulation of histone acetylation. J Biol Chem. 2007 Nov 23; 282(47):33994-4002.

87. Clark VE, Erson-Omay EZ, Serin A, Yin J, Cotney J, Ozduman K, et al. Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. Science. 2013 Mar 1; 339(6123):1077-80.

88. Jiang YZ, Jiménez JM, Ou K, McCormick ME, Zhang LD, Davies PF. Hemodynamic disturbed flow induces differential DNA methylation of endothelial Kruppel-Like Factor 4 promoter in vitro and in vivo. Circ Res. 2014 Jun 20; 115(1):32-43.

89. Su C, Sun F, Cunningham RL, Rybalchenko N, Singh M. ERK5/KLF4 signaling as a common mediator of the neuroprotective effects of both nerve growth factor and hydrogen peroxide preconditioning. Age (Dordr). 2014 Aug; 36(4):9685.

90. Zhou Y, Du WD, Chen G, Ruan J, Xu S, Zhou FS, Zuo XB, Lv ZJ, Zhang XJ. Association analysis of genetic variants in microRNA networks and gastric cancer risk in a Chinese Han population. J Cancer Res Clin Oncol. 2012 Jun; 138(6):939-45.

91. Mook OR, Baas F, de Wissel MB, Fluiter K. Allele-specific cancer cell killing in vitro and in vivo targeting a single-nucleotide polymorphism in POLR2A. Cancer Gene Ther. 2009 Jun; 16(6):532-8.

92. Wakabayashi K, Narisawa-Saito M, Iwakura Y, Arai T, Ikeda K, Takahashi H, Nawa H. Phenotypic down-regulation of glutamate receptor subunit GluR1 in Alzheimer's disease. Neurobiol Aging. 1999 May-Jun; 20(3):287-95.

93. Pellegrini-Giampietro DE, Bennett MV, Zukin RS. AMPA/kainate receptor gene expression in normal and Alzheimer's disease hippocampus. Neuroscience. 1994 Jul; 61(1):41-9.

94. Lang UE, Puls I, Muller DJ, Strutz-Seebohm N, Gallinat J. Molecular mechanisms of schizophrenia. Cell Physiol Biochem. 2007; 20(6):687-702.

95. Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, et al. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol Psychiatry. 2012 Sep; 17(9):887-905.

96. Ba M, Kong M, Yang H, Ma G, Lu G, Chen S, Liu Z. Changes in subcellular distribution and phosphorylation of GluR1 in lesioned striatum of 6-hydroxydopamine-lesioned and l-dopa-treated rats. Neurochem Res. 2006 Nov; 31(11):1337-47.

97. Li DP, Byan HS, Pan HLSwitch to glutamate receptor 2-lacking AMPA receptors increases neuronal excitability in hypothalamus and sympathetic drive in hypertension. J Neurosci. 2012 Jan 4; 32(1):372-80.

98. Morrell CN, Sun H, Ikeda M, Beique JC, Swaim AM, Mason E, Martin TV, et. al. Glutamate mediates platelet activation through the AMPA receptor. J Exp Med. 2008 Mar 17; 205(3):575-84.

99. Liou YJ, Wang HH, Lee MT, Wang SC, Chiang HL, et al. (2012) Genome-wide association study of treatment refractory schizophrenia in Han Chinese. PLoS One 7: e33598.

100. Nair J, Ghatge M, Kakkar VV, Shanker J. Network analysis of inflammatory genes and their transcriptional regulators in coronary artery disease. PLoS One. 2014 Apr 15; 9(4):e94328.

101. Malek S, Huxford T, Ghosh G. Ikappa Balpha functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-kappaB. J Biol Chem. 1998 Sep 25; 273(39):25427-35.

102. Tragante V, Barnes MR, Ganesh SK, Lanktree MB, Guo W, Franceschini N, et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. Am J Hum Genet. 2014 Mar 6; 94(3):349-60.

103. Ghose J, Sinha M, Das E, Jana NR, Bhattacharyya NP. Regulation of miR-146a by RelA/NFkB and p53 in STHdh(Q111)/Hdh(Q111) cells, a cell model of Huntington's disease. PLoS One. 2011; 6(8):e23837.

104. Kim J, Min JK, Park JA, Doh HJ, Choi YS, Rho J, Kim YM, Kwon YG. Receptor activator of nuclear factor kappaB ligand is a novel inducer of tissue factor in macrophages. Circ Res. 2010 Oct 1; 107(7):871-6.

105. Koldamova R, Schug J, Lefterova M, Cronican AA, Fitz NF, et al. Genome-wide approaches reveal EGR1-controlled regulatory networks associated with neurodegeneration. Neurobiol Dis. 2014 Mar; 63:107-14.

106. Ghazvini-Boroujerdi M, Clark J, Narula N, Palmatory E, Connolly JM, et al. Transcription factor Egr-1 in calcific aortic valve disease. J Heart Valve Dis. 2004 Nov; 13(6):894-903.

107. Zhang T, Zhao LL, Cao X, Qi LC, Wei GQ, Liu JY, Yan SJ, Liu JG, Li XQ. Bioinformatics analysis of time series gene expression in left ventricle (LV) with acute myocardial infarction (AMI). Gene. 2014 Jun 15; 543(2):259-67.

108. Kurian SM, Le-Niculescu H, Patel SD, Bertram D, Davis J, et al.Identification of blood biomarkers for psychosis using convergent functional genomics. Mol Psychiatry. 2011 Jan; 16(1):37-58.

109. St Germain C, O'Brien A, Dimitroulakos J. Activating Transcription Factor 3 regulates in part the enhanced tumour cell cytotoxicity of the histone deacetylase inhibitor M344 and cisplatin in combination. Cancer Cell Int. 2010 Sep 9; 10:32.

110. Buganim Y, Madar S, Rais Y, Pomeraniec L, Harel E, Solomon H, et al. Transcriptional activity of ATF3 in the stromal compartment of tumors promotes cancer progression. Carcinogenesis. 2011 Dec; 32(12):1749-57.

111. Nawa T1, Nawa MT, Adachi MT, Uchimura I, Shimokawa R, Fujisawa K, et al. Expression of transcriptional repressor ATF3/LRF1 in human atherosclerosis:

colocalization and possible involvement in cell death of vascular endothelial cells.Atherosclerosis. 2002 Apr; 161(2):281-91.

112. de Waard MC, van der Pluijm I, Zuiderveen Borgesius N, Comley LH, et al. Age-related motor neuron degeneration in DNA repair-deficient Ercc1 mice. Acta Neuropathol. 2010 Oct; 120(4):461-75.

113. Hartman MG, Lu D, Kim ML, Kociba GJ, Shukri T, Buteau J, et al. Role for activating transcription factor 3 in stress-induced beta-cell apoptosis. Mol Cell Biol. 2004 Jul; 24(13):5721-32.

114. Chen BP, Wolfgang CD, Hai T. Analysis of ATF3, a transcription factor induced by physiological stresses and modulated by gadd153/Chop10. Mol Cell Biol. 1996 Mar; 16(3):1157-68.

115. Zang C, Liu H, Bertz J, Possinger K, Koeffler HP, Elstner E, Eucker J. Induction of endoplasmic reticulum stress response by TZD18, a novel dual ligand for peroxisome proliferator-activated receptor alpha/gamma, in human breast cancer cells. Mol Cancer Ther. 2009 Aug;8(8):2296-307.

116. Oyadomari S, Araki E, Mori M. Endoplasmic reticulum stress-mediated apoptosis in pancreatic beta-cells. Apoptosis. 2002 Aug; 7(4):335-45.

117. Ono Y, Tanaka H, Tsuruma K, Shimazawa M, Hara H. A sigma-1 receptor antagonist (NE-100) prevents tunicamycin-induced cell death via GRP78 induction in hippocampal cells. Biochem Biophys Res Commun. 2013 May 17; 434(4):904-9.

118. Gómez-Santos C, Barrachina M, Giménez-Xavier P, Dalfó E, Ferrer I, Ambrosio S. Induction of C/EBP beta and GADD153 expression by dopamine in human neuroblastoma cells. Relationship with alpha-synuclein increase and cell damage. Brain Res Bull. 2005 Feb 15; 65(1):87-95.

119. Khaleque MA, Bharti A, Gong J, Gray PJ, Sachdev V, Ciocca DR, et al. Heat shock factor 1 represses estrogen-dependent transcription through association with MTA1.Oncogene. 2008 Mar 20; 27(13):1886-93.

120. Riva L, Koeva M, Yildirim F, Pirhaji L, Dinesh D, Mazor T, Duennwald ML, Fraenkel E.Poly-glutamine expanded huntingtin dramatically alters the genome wide binding of HSF1. J Huntingtons Dis. 2012 Jan; 1(1):33-45.

121. Neef DW, Jaeger AM, Thiele DJ. Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. Nat Rev Drug Discov. 2011 Dec 1; 10(12):930-44.

Table S1: Details of SNPs present in LD with the reported GWAS SNPs as provided by Haploreg.

GWAS SNP	Proxy SNPs in LD	LD (r2)
rs1247318 /79kb 5' of MAP3K4	rs2489955	0.83
	rs2465876	0.87
	rs1937474	0.88
	rs2465877	0.88
	rs2465878	0.88
	rs9355851	0.88
	rs9364580	0.9
	rs1247329	0.91
	rs1247328	0.92
	rs1247327	0.92
	rs1247326	0.89
	rs1247325	0.92
	rs1247323	0.98
	rs1247322	0.98
	rs1247321	1
	rs1247320	1
	rs1247319	1
	rs1247318	**1**
	rs1247317	1
	rs61717948	1
	rs935183	1
	rs935182	1
	rs35668205	1
	rs1247313	1
	rs1247312	1
	rs1781546	0.98
	rs1663025	0.98
	rs1247310	1
	rs1247309	1
	rs1247307	1
	rs1247305	1
	rs146425745	0.82
	rs200682997	0.81
	rs1247360	1
	rs1247361	1
	rs35877341	1
	rs1247362	1
	rs1247363	1
	rs1247364	1
	rs1247365	1
	rs1247302	1
	rs1247300	0.99
	rs1247298	1
	rs1247294	0.99
	rs591676	0.83
	rs687701	0.82
	rs638141	0.82
	rs585649	0.83
	rs581947	0.83
rs294588 /52kb 3' of MAT2B	rs985253	0.81
	rs294588	**1**

GWAS SNP	Proxy SNPs in LD	LD (r2)
rs1572438 / 102kb 3' of LOC285768	rs11243209	0.82
	rs10901016	0.87
	rs9392933	0.87
	rs9406070	0.94
	rs9405362	0.98
	rs4959463	0.99
	rs34751630	0.86
	rs9406073	1
	rs9406075	0.89
	rs17139759	0.89
	rs12202771	0.89
	rs4959466	0.9
	rs4959467	0.89
	rs4959468	0.89
	rs9378515	0.89
	rs9405366	0.85
	rs1572438	**1**
	rs3799325	0.89
	rs3799324	1
	rs9406088	0.82
	rs9406089	0.99
	rs9392198	0.88
	rs67507315	0.96
	rs9378520	0.96
	rs9328455	0.89
	rs9379167	0.88
	rs873559	0.83
	rs873560	0.83
rs6925255 /14kb 3' of RNY4	rs12212008	0.92
	rs7740895	0.93
	rs4706795	0.93
	rs4706796	0.92
	rs9343929	0.93
	rs9343930	0.93
	rs4706101	0.85
	rs6900699	0.85
	rs6925170	0.81
	rs9359390	0.94
	rs199658477	0.8
	rs9350823	0.89
	rs4706797	0.89
	rs9350824	0.82
	rs9352759	0.86
	rs9341788	0.86
	rs2207369	0.94
	rs9448772	0.93
	rs2057155	0.98
	rs10738002	0.99
	rs9294155	0.8
	rs4642424	0.99
	rs9341789	0.8
	rs7774837	0.8
	rs10943662	0.8

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs9361557	0.99
	rs4706798	0.99
	rs761605	0.98
	rs6925255	**1**
	rs7768071	0.99
	rs6936836	0.87
	rs932492	0.84
	rs6924740	0.89
	rs13197772	0.89
	rs9350826	0.88
	rs6921089	0.88
	rs6941826	0.88
	rs9352764	0.83
	rs17514496	0.88
	rs12660740	0.88
	rs12663653	0.88
	rs9443716	0.88
	rs714701	0.84
	rs714702	0.87
	rs714704	0.88
	rs714703	0.88
	rs4467741	0.88
	rs9448820	0.8
	rs9352765	0.81
rs11585386 /137kb 3' of RHOU	**rs11585386**	**1**
rs16892673 /PARK2	rs6922154	0.98
	rs55694434	0.8
	rs57350839	0.8
	rs73782913	0.8
	rs16892658	0.98
	rs73782918	0.98
	rs16892668	0.98
	rs73782921	1
	rs16892673	**1**
	rs76203566	1
	rs142446701	0.88
	rs11965294	0.88
	rs11965303	0.88
	rs73782924	0.88
	rs142941238	0.88
	rs56085387	0.88
	rs59341317	0.88
	rs16892698	0.88
	rs16892700	0.88
rs9918668 /58kb 5' of BET1	**rs9918668**	**1**
	rs9918681	0.9
rs643473 /6.4kb 3' of GLT1D1	rs470593	0.88
	rs470427	0.87
	rs470424	0.86
	rs470597	0.87

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs569352	0.89
	rs3914957	0.82
	rs3858583	0.89
	rs3858584	0.87
	rs641711	0.99
	rs1795681	0.88
	rs7970002	0.99
	rs676500	0.98
	rs643473	**1**
	rs9668939	0.88
	rs674729	1
	rs487407	1
	rs511863	1
	rs489270	1
	rs489304	1
	rs470758	1
	rs510134	1
	rs509934	0.96
	rs470834	1
	rs470837	0.99
	rs12811232	0.99
	rs12812667	0.99
	rs12811264	0.99
	rs12818591	0.89
	rs12811474	0.89
	rs470785	0.99
	rs504501	0.87
	rs1621538	0.98
	rs1627443	0.85
	rs1718513	0.85
	rs1718514	0.92
	rs1718515	0.96
	rs2702212	0.99
	rs470429	0.99
	rs470386	0.99
	rs558831	0.91
	rs551816	0.85
	rs2266466	0.99
	rs2675927	0.96
	rs604517	0.97
	rs470458	0.98
	rs470405	0.85
	rs4307761	0.87
	rs486420	0.85
	rs470519	0.86
	rs470400	0.84
	rs470521	0.86
	rs470529	0.86
	rs497428	0.86
	rs470592	0.86
	rs470955	0.85
	rs470602	0.86
	rs470609	0.86
	rs470619	0.86

GWAS SNP	Proxy SNPs in LD	LD (r2)
rs12474609 /LRP1B	rs10928081	0.95
	rs12474609	**1**
	rs10928082	1
	rs12620377	0.97
	rs12614087	0.96
	rs12614115	0.98
	rs12617569	0.97
rs766903 /BIN2	rs7134625	0.9
	rs4762016	0.96
	rs4761832	0.96
	rs2011124	0.97
	rs9788096	0.94
	rs9788179	0.97
	rs766903	**1**
	rs7971409	0.92
	rs4761846	0.8
rs1425609 /213kb 3' of LOC647107	rs13094783	0.83
	rs9866192	0.83
	rs9816086	0.83
	rs11914874	0.84
	rs9854933	0.84
	rs13074225	0.8
	rs72391779	0.97
	rs11917929	0.99
	rs9290153	0.99
	rs9836765	0.99
	rs3844503	0.84
	rs7355808	0.84
	rs7355972	0.98
	rs12494878	1
	rs6790172	0.85
	rs9851215	0.99
	rs7640988	0.86
	rs4099484	1
	rs9867215	0.85
	rs9867485	0.86
	rs9845104	1
	rs9865699	0.98
	rs9875594	1
	rs1120906	1
	rs1120907	1
	rs34861578	0.99
	rs994408	1
	rs994410	1
	rs9847253	1
	rs11925890	1
	rs11919599	1
	rs34012392	1
	rs9830442	1
	rs9867733	0.85
	rs12152484	1
	rs12152310	1

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs10936318	1
	rs10936319	0.99
	rs10936320	0.85
	rs144050070	0.99
	rs76229349	0.87
	rs4476530	0.89
	rs4510403	0.89
	rs77529894	0.98
	rs140145501	0.98
	rs199836029	0.86
	rs77420327	0.94
	rs148119770	0.99
	rs62396422	0.98
	rs141210220	1
	rs1425609	**1**
	rs9884105	1
	rs9847141	0.85
	rs10212418	1
	rs10212444	0.85
	rs10212178	1
	rs28898349	0.99
	rs75411070	0.85
	rs114910966	0.85
	rs9877131	0.86
	rs9839942	0.85
	rs28873116	0.84
	rs76083821	0.81
	rs138502324	0.82
	rs10513596	0.83
	rs1488168	0.82
	rs1386854	0.8
	rs1386856	0.83
	rs981916	0.83
	rs981917	0.83
	rs981918	0.83
	rs9866542	0.83
	rs12496047	0.83
	rs1548001	0.83
	rs1548002	0.83
	rs2362561	0.83
	rs13073152	0.82
	rs2114021	0.83
	rs1386857	0.83
	rs1386858	0.83
	rs1946342	0.83
	rs1844074	0.83
	rs1844075	0.82
	rs9879670	0.83
	rs9860284	0.83
	rs1386859	0.83
	rs5854008	0.83
	rs1386860	0.83
	rs9864922	0.83
	rs988191	0.83

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs12630510	0.82
	rs12632934	0.82
	rs1955090	0.83
	rs1955091	0.83
	rs1955093	0.83
	rs13096974	0.83
	rs74797128	0.83
	rs6548324	0.83
	rs4611845	0.83
	rs4276209	0.83
	rs12486590	0.82
	rs12486540	0.83
	rs12486577	0.82
	rs62396435	0.83
	rs13095136	0.83
	rs62396436	0.83
	rs9822393	0.83
	rs9822721	0.83
	rs1488169	0.83
	rs1488170	0.83
	rs9852069	0.83
	rs13318277	0.81
	rs9869731	0.83
	rs9870585	0.83
	rs9877942	0.81
	rs9858129	0.8
	rs9820405	0.96
	rs9820439	0.82
	rs9858617	0.8
	rs9858618	0.82
	rs9858779	0.81
	rs61207972	0.81
	rs13077542	0.82
	rs13100199	0.8
	rs9863458	0.81
	rs9864520	0.81
	rs12489054	0.81
	rs6774016	0.82
	rs1011306	0.82
	rs1030361	0.82
	rs901819	0.82
	rs12488794	0.82
	rs1157780	0.82
	rs1580514	0.82
	rs1580515	0.82
	rs60130055	0.81
	rs1488181	0.8
	rs1488180	0.82
	rs12632097	0.82
	rs13084207	0.82
	rs6798456	0.82
	rs1038635	0.82
	rs1038634	0.82
	rs2019760	0.82

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs1030362	0.82
	rs5854010	0.82
rs4936894 /VWA5A	**rs4936894**	**1**
rs10202497 /COL6A3	rs2645764	0.83
	rs2645763	0.83
	rs2645777	0.84
	rs10202497	**1**
	rs10167850	0.99
rs2367725 /ST3GAL3	rs6683825	0.89
	rs140287770	0.89
	rs3791037	0.89
	rs3791038	0.89
	rs3791039	0.9
	rs3815268	0.89
	rs2240517	0.93
	rs34550543	0.91
	rs7528454	0.94
	rs6429635	0.94
	rs7536221	0.94
	rs11210903	0.94
	rs12139239	0.93
	rs12140156	0.95
	rs6693799	0.95
	rs11210906	0.95
	rs7528907	0.93
	rs11210908	0.94
	rs11210909	0.95
	rs11210910	0.96
	rs12118274	0.96
	rs3838465	0.95
	rs11210911	0.97
	rs10890276	0.97
	rs144951225	0.96
	rs10890277	0.96
	rs35052091	0.96
	rs12403488	0.96
	rs1990195	0.95
	rs2108419	1
	rs2158955	1
	rs2367725	**1**
	rs12125928	0.99
	rs12129369	1
	rs12564694	1
	rs71579308	0.95
	rs11579637	1
	rs10890279	1
	rs3791045	0.99
	rs11590088	1
	rs3791048	1
	rs12142533	1
	rs11210918	0.99

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs12140180	0.98
	rs12750525	0.81
	rs7549701	0.98
	rs6692652	0.84
	rs3791049	0.86
	rs3791051	0.87
	rs1875653	0.87
	rs10890283	0.87
	rs6701645	0.87
	rs3791053	0.87
	rs12121688	0.87
	rs12122504	0.87
	rs71579311	0.83
	rs6688144	0.86
	rs35314257	0.86
	rs7553818	0.86
	rs11210924	0.86
	rs11210925	0.85
	rs11210926	0.86
	rs12116642	0.86
	rs34330931	0.86
	rs975761	0.86
	rs7515003	0.86
	rs3828140	0.86
	rs976313	0.86
	rs976312	0.86
	rs11210927	0.86
	rs12128485	0.86
	rs12128547	0.85
	rs3791057	0.85
	rs6672795	0.85
	rs6665149	0.85
	rs3791058	0.85
	rs74070702	0.85
	rs812490	0.81
	rs11580676	0.85
	rs12126352	0.85
	rs35747575	0.83
	rs12122867	0.83
	rs200285982	0.82
	rs12123718	0.81
	rs3791059	0.85
	rs1012125	0.84
	rs3838468	0.83
	rs12125838	0.85
	rs12133508	0.84
	rs3791062	0.85
	rs3791063	0.85
	rs3791064	0.85
	rs972444	0.85
	rs972443	0.85
rs10412199 /ATCAY	rs10424392	0.97
	rs10412199	**1**

GWAS SNP	Proxy SNPs in LD	LD (r2)
rs3112530 /230kb 5' of GRIA1	rs2546328	0.91
	rs2615153	0.91
	rs3101131	0.91
	rs2974110	0.91
	rs2349576	1
	rs2061638	1
	rs1462126	1
	rs3101485	1
	rs3101130	0.98
	rs3112530	**1**
	rs3111205	1
	rs3101482	0.98
	rs35122664	1
	rs2926288	1
	rs2973148	0.98
	rs982494	1
	rs2927172	1
	rs2973149	1
	rs2964817	1
	rs2926287	1
	rs1381540	1
	rs2964813	1
	rs1824311	0.98
	rs2926286	0.98
	rs2964812	1
	rs2927174	1
	rs2199123	1
	rs1462108	1
	rs1462109	1
	rs2125515	1
	rs2964821	1
	rs2973152	1
	rs1462122	0.98
	rs1462123	0.98
	rs2964816	0.98
	rs10463328	0.98
	rs2262709	0.93
	rs4958634	0.95
	rs6580010	0.91
	rs151011228	0.82
	rs2609665	0.95
	rs2560229	0.95
	rs2560231	0.95
	rs112668900	0.81
	rs308267	0.95
	rs308270	0.95
	rs1462113	0.95
	rs1462114	0.93
	rs1026615	0.95
	rs2615178	0.95
	rs2615176	0.91
	rs2560243	0.95
	rs300328	0.95
	rs300327	0.95

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs300326	0.95
	rs12655396	0.95
	rs2615173	0.95
	rs2609661	0.95
	rs2260878	0.95
	rs2609667	0.95
	rs2262710	0.95
	rs2262711	0.95
	rs2262712	0.85
	rs138445161	0.85
	rs2617267	0.95
	rs2927584	0.95
	rs2913310	0.95
	rs2964820	0.95
	rs10073797	0.95
	rs2560245	0.95
	rs2615172	0.95
	rs2615170	0.95
	rs2615169	0.95
	rs2617268	0.95
	rs2446429	0.95
	rs2560247	0.91
	rs2609673	0.95
	rs2560248	0.95
	rs2560249	0.95
	rs170027	0.84
	rs167357	0.91
	rs304856	0.95
	rs304857	0.84
	rs304858	0.88
	rs304859	0.91
	rs170028	0.91
	rs2560252	0.91
	rs304860	0.91
	rs172569	0.91
	rs172570	0.91
	rs10531959	0.83
	rs304861	0.91
	rs170029	0.91
	rs304862	0.91
	rs304863	0.91
	rs34031062	0.83
	rs304864	0.91
	rs150196369	0.88
	rs300331	0.88
	rs300332	0.88
	rs186183	0.88
	rs300335	0.88
	rs159759	0.88
	rs160163	0.88
	rs160066	0.88
	rs150618	0.88
	rs149095	0.88
	rs302738	0.88

GWAS SNP	Proxy SNPs in LD	LD (r2)
	rs159972	0.88
	rs159973	0.88
	rs160172	0.88
	rs159978	0.88
rs4764043 /GRIN2B	**rs4764043**	**1**
rs8001976 /129kb 3' of SUCLA2	rs17424137	0.86
	rs8001976	**1**
	rs10507562	0.88
rs16852912 /115kb 3' of MECOM	rs113340333	0.9
	rs9831461	0.9
	rs13316589	1
	rs139566210	1
	rs9879474	1
	rs16852912	**1**
	rs75809958	1
	rs79574287	0.93
	rs115076174	0.93
	rs9861283	0.93
rs11162963 /1.3Mb 5' of ELTD1	**rs11162963**	**1**

Table S2: Summary of all GWAS and proxy SNPs in LD (r2 >= 0.8) with RegulomeDB score. ***Bolded SNPs are GWAS reported SNPs.***

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr1:44252907	rs3791051	2b
chr3:162702770	rs9822393	2b
chr5:152732405	rs159972	2b
chr6:80474571	rs932492	2b
chr12:129481865	rs470619	3a
chr1:44245776	rs7549701	3a
chr3:162696257	rs9864922	3a
chr3:168706811	rs9861283	3a
chr6:161919525	rs6922154	3a
chr6:875335	rs873559	3a
chr1:44172457	rs34550543	4
chr1:44174080	rs7528454	4
chr1:44236589	rs3791048	4
chr1:44242466	rs12140180	4
chr1:44249769	rs6692652	4
chr3:162691290	rs9866542	4
chr3:162696085	rs1386860	4
chr3:162702813	rs9822721	4
chr3:168686675	***rs16852912**	4
chr5:152719987	rs304863	4
chr6:161342435	rs35877341	4
chr12:129475349	rs3858584	5
chr12:129477130	rs12811264	5
chr12:129477202	rs12818591	5
chr12:129477203	rs12811474	5
chr12:129477281	rs470785	5
chr12:129480300	rs486420	5
chr12:129480323	rs470519	5
chr12:129480419	rs470400	5
chr12:129480423	rs470521	5
chr12:129481598	rs470609	5
chr12:51709161	rs7971409	5
chr19:3925412	rs10424392	5
chr19:3927770	***rs10412199**	5
chr1:44171265	rs2240517	5
chr1:44185167	rs6693799	5
chr1:44200170	rs12118274	5
chr1:44215219	rs2108419	5
chr1:44215827	***rs2367725**	5
chr1:44223844	rs71579308	5
chr1:44254144	rs10890283	5
chr1:44258930	rs6688144	5
chr1:44263713	rs34330931	5
chr1:44265466	rs3828140	5
chr1:44266194	rs976313	5
chr1:44266309	rs976312	5
chr1:44270560	rs74070702	5
chr1:44280424	rs3791064	5
chr2:238262253	rs2645764	5
chr2:238270893	***rs10202497**	5
chr3:162637804	rs9816086	5
chr3:162650451	rs9836765	5

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr3:162688216	rs10513596	5
chr3:162689804	rs1386854	5
chr3:162695980	rs5854008	5
chr3:162699265	rs74797128	5
chr3:162699290	rs6548324	5
chr3:162705392	rs9858617	5
chr3:162705397	rs9858618	5
chr3:162705481	rs9858779	5
chr3:162705731	rs61207972	5
chr3:162705863	rs13077542	5
chr3:162705878	rs13100199	5
chr3:162707977	rs1011306	5
chr3:162708444	rs1030361	5
chr3:162710619	rs1157780	5
chr3:162714486	rs12632097	5
chr3:162715926	rs1038635	5
chr3:162716003	rs1038634	5
chr3:162716101	rs2019760	5
chr3:168702770	rs115076174	5
chr5:152619463	rs2546328	5
chr5:152679175	rs2560231	5
chr5:152727357	rs160066	5
chr5:152727393	rs150618	5
chr5:152727444	rs149095	5
chr5:152737574	rs159978	5
chr6:161346443	rs1247302	5
chr6:161928545	rs73782918	5
chr6:80424011	rs12212008	5
chr6:80478782	rs6924740	5
chr6:80485219	rs17514496	5
chr6:80486195	rs12660740	5
chr6:80486205	rs12663653	5
chr6:854450	rs9406073	5
chr6:858295	rs9405366	5
chr6:858969	***rs1572438**	5
chr6:866371	rs9379167	5
chr7:93691743	***rs9918668**	5
chr12:129474866	rs470593	6
chr12:129475023	rs470597	6
chr12:129475114	rs569352	6
chr12:129475276	rs3858583	6
chr12:129475535	rs641711	6
chr12:129475611	rs7970002	6
chr12:129475703	rs676500	6
chr12:129476579	rs489270	6
chr12:129476587	rs489304	6
chr12:129476868	rs470834	6
chr12:129477066	rs12811232	6
chr6:844121	rs11243209	6
chr6:80486610	rs714702	6
chr6:80486490	rs714701	6
chr6:80484528	rs6921089	6
chr6:80460522	rs761605	6
chr6:80459235	rs9361557	6

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr6:80457516	rs10943662	6
chr6:80450778	rs4642424	6
chr6:80449231	rs9294155	6
chr6:80439297	rs9341788	6
chr6:80438870	rs9350824	6
chr6:80437227	rs4706797	6
chr6:80435515	rs4706101	6
chr6:80434123	rs9343929	6
chr6:80433731	rs4706796	6
chr6:161940898	rs142941238	6
chr6:161939594	rs11965294	6
chr6:161938679	rs142446701	6
chr6:161934259	rs76203566	6
chr6:161933934	***rs16892673**	6
chr6:161931651	rs73782921	6
chr6:161365628	rs581947	6
chr6:161364783	rs585649	6
chr6:161350478	rs1247294	6
chr6:161342689	rs1247363	6
chr6:161341268	rs1247361	6
chr6:161339377	rs200682997	6
chr6:161339291	rs146425745	6
chr6:161338000	rs1247305	6
chr6:161336635	rs1247309	6
chr6:161336408	rs1663025	6
chr6:161335409	rs35668205	6
chr6:161335252	rs935183	6
chr6:161333936	***rs1247318**	6
chr6:161333859	rs1247319	6
chr6:161332209	rs1247321	6
chr6:161330012	rs1247323	6
chr6:161316322	rs2465877	6
chr6:161315399	rs2465876	6
chr5:152725254	rs160163	6
chr5:152723199	rs300331	6
chr5:152721516	rs150196369	6
chr5:152720861	rs304864	6
chr5:152720720	rs34031062	6
chr5:152719261	rs304861	6
chr5:152718506	rs304860	6
chr5:152718343	rs304859	6
chr5:152717731	rs304856	6
chr5:152716309	rs2560247	6
chr5:152716101	rs2615169	6
chr5:152715537	rs2560245	6
chr5:152715444	rs10073797	6
chr5:152715434	rs2964820	6
chr5:152715389	rs2913310	6
chr5:152713832	rs2262712	6
chr5:152713690	rs2262710	6
chr5:152713246	rs2609667	6
chr5:152701345	rs2560243	6
chr5:152685429	rs2615150	6
chr5:152665884	rs2262709	6

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr5:152651537	rs2927174	6
chr5:152648707	rs2964813	6
chr5:152643973	rs2973148	6
chr5:152643968	rs2926288	6
chr5:152642146	rs35122664	6
chr5:152641606	rs3111205	6
chr5:152638131	rs2061638	6
chr5:152638037	rs2349576	6
chr3:168699619	rs79574287	6
chr3:168677983	rs9879474	6
chr3:162716491	rs1030362	6
chr3:162713842	rs1488180	6
chr3:162713817	rs1488181	6
chr3:162713155	rs60130055	6
chr3:162711326	rs1580514	6
chr3:162708973	rs901819	6
chr3:162707562	rs6774016	6
chr3:162705184	rs9820405	6
chr3:162704822	rs9870585	6
chr3:162704324	rs9869731	6
chr3:162703179	rs1488169	6
chr3:162701852	rs13095136	6
chr3:162701538	rs62396435	6
chr3:162701230	rs12486540	6
chr3:162698132	rs1955090	6
chr3:162697947	rs12632934	6
chr3:162697911	rs12630510	6
chr3:162695482	rs9860284	6
chr3:162693960	rs1946342	6
chr3:162693699	rs1386858	6
chr3:162693419	rs1386857	6
chr3:162693189	rs2114021	6
chr3:162692255	rs12496047	6
chr3:162690508	rs981918	6
chr3:162689868	rs1386856	6
chr3:162688459	rs1488168	6
chr3:162687094	rs76083821	6
chr3:162686367	rs28873116	6
chr3:162684920	rs114910966	6
chr3:162684339	rs28898349	6
chr3:162684128	rs10212178	6
chr3:162682127	rs9884105	6
chr3:162681994	***rs1425609**	6
chr3:162681224	rs148119770	6
chr3:162680503	rs79352825	6
chr3:162680156	rs4510403	6
chr3:162680135	rs4476530	6
chr3:162679122	rs10936318	6
chr3:162678992	rs12152310	6
chr3:162678966	rs12152484	6
chr3:162676567	rs34012392	6
chr3:162673705	rs11925890	6
chr3:162668911	rs1120906	6
chr3:162664344	rs9845104	6

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr3:162661570	rs9867215	6
chr3:162659704	rs7640988	6
chr3:162652029	rs7355972	6
chr3:162648687	rs11917929	6
chr3:162647626	rs61468644	6
chr3:162642754	rs13074225	6
chr2:238266145	rs2645777	6
chr2:238263298	rs2645763	6
chr2:141721598	rs12617569	6
chr2:141721502	rs12614115	6
chr1:80734580	***rs11162963**	6
chr1:44282354	rs972443	6
chr1:44277745	rs12133508	6
chr1:44277169	rs3838468	6
chr1:44277098	rs1012125	6
chr1:44275688	rs12123718	6
chr1:44275688	rs12123718	6
chr1:44275440	rs12122867	6
chr1:44275376	rs35747575	6
chr1:44270309	rs3791058	6
chr1:44268784	rs6665149	6
chr1:44268698	rs6672795	6
chr1:44268011	rs3791057	6
chr1:44266419	rs11210927	6
chr1:44261521	rs11210926	6
chr1:44261194	rs11210924	6
chr1:44260311	rs7553818	6
chr1:44260112	rs35314257	6
chr1:44258024	rs71579311	6
chr1:44256818	rs12122504	6
chr1:44256457	rs12121688	6
chr1:44255274	rs3791053	6
chr1:44251118	rs3791049	6
chr1:44237119	rs12142533	6
chr1:44230257	rs10890279	6
chr1:44223885	rs11579637	6
chr1:44222909	rs12564694	6
chr1:44221528	rs12129369	6
chr1:44217707	rs12125928	6
chr1:44213088	rs12403488	6
chr1:44210665	rs144951225	6
chr1:44204300	rs11210911	6
chr1:44193455	rs7528907	6
chr1:44184624	rs12140156	6
chr1:44160057	rs3791037	6
chr1:44153400	rs140287770	6
chr13:48388547	rs10507562	6
chr13:48387721	***rs8001976**	6
chr12:51713768	rs4761846	6
chr12:51701154	rs9788179	6
chr12:51700924	rs9788096	6
chr12:14115481	***rs4764043**	6
chr12:129481354	rs470955	6
chr12:129481125	rs470592	6

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr12:129478864	rs551816	6
chr12:129478838	rs558831	6
chr12:129478532	rs470386	6
chr12:129478448	rs470429	6
chr12:129478266	rs2702212	6
chr12:129478249	rs1718515	6
chr12:129478141	rs1718513	6
chr12:129478140	rs1627443	6
chr12:129477066	rs12811232	6
chr12:129476868	rs470834	6
chr12:129476587	rs489304	6
chr12:129476579	rs489270	6
chr12:129475703	rs676500	6
chr12:129475611	rs7970002	6
chr12:129475535	rs641711	6
chr12:129475276	rs3858583	6
chr12:129475114	rs569352	6
chr12:129475023	rs470597	6
chr12:129474866	rs470593	6
chr7:93692061	rs9918681	No Data
chr6:875558	rs873560	No Data
chr6:861664	rs67507315	No Data
chr6:859800	rs9406089	No Data
chr6:859789	rs9406088	No Data
chr6:858055	rs4959468	No Data
chr6:857976	rs4959467	No Data
chr6:856753	rs17139759	No Data
chr6:856360	rs9406075	No Data
chr6:853029	rs34751630	No Data
chr6:851067	rs9405362	No Data
chr6:849230	rs10901016	No Data
chr6:80488986	rs9352765	No Data
chr6:80488589	rs9448820	No Data
chr6:80486718	rs4467741	No Data
chr6:80486681	rs714703	No Data
chr6:80486643	rs714704	No Data
chr6:80486348	rs9443716	No Data
chr6:80484971	rs9352764	No Data
chr6:80484569	rs6941826	No Data
chr6:80484291	rs9350826	No Data
chr6:80479372	rs13197772	No Data
chr6:80473534	rs6936836	No Data
chr6:80467653	rs7768071	No Data
chr6:80466123	***rs6925255**	No Data
chr6:80459367	rs4706798	No Data
chr6:80456491	rs7774837	No Data
chr6:80454550	rs9341789	No Data
chr6:80448811	rs10738002	No Data
chr6:80448392	rs2057155	No Data
chr6:80443733	rs9448772	No Data
chr6:80442332	rs2207369	No Data
chr6:80438924	rs9352759	No Data
chr6:80436670	rs9350823	No Data
chr6:80436651	rs199658477	No Data

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr6:80435850	rs6925170	No Data
chr6:80435784	rs6900699	No Data
chr6:80434181	rs9343930	No Data
chr6:80433324	rs4706795	No Data
chr6:80432687	rs7740895	No Data
chr6:161944578	rs16892700	No Data
chr6:161943364	rs16892698	No Data
chr6:161943071	rs59341317	No Data
chr6:161941809	rs56085387	No Data
chr6:161940858	rs73782924	No Data
chr6:161939646	rs11965303	No Data
chr6:161931123	rs16892668	No Data
chr6:161928033	rs16892658	No Data
chr6:161924380	rs73782913	No Data
chr6:161923404	rs57350839	No Data
chr6:161921805	rs55694434	No Data
chr6:161361887	rs638141	No Data
chr6:161359557	rs687701	No Data
chr6:161358807	rs591676	No Data
chr6:161347688	rs1247298	No Data
chr6:161346889	rs1247300	No Data
chr6:161342927	rs1247365	No Data
chr6:161342737	rs1247364	No Data
chr6:161342659	rs1247362	No Data
chr6:161341207	rs1247360	No Data
chr6:161337378	rs1247307	No Data
chr6:161336527	rs1247310	No Data
chr6:161336406	rs1781546	No Data
chr6:161336190	rs1247312	No Data
chr6:161336019	rs1247313	No Data
chr6:161335375	rs935182	No Data
chr6:161335144	rs1247316	No Data
chr6:161334413	rs1247317	No Data
chr6:161333780	rs1247320	No Data
chr6:161330651	rs1247322	No Data
chr6:161328669	rs1247325	No Data
chr6:161328555	rs1247326	No Data
chr6:161328455	rs1247327	No Data
chr6:161328070	rs1247328	No Data
chr6:161327756	rs1247329	No Data
chr6:161322926	rs9364580	No Data
chr6:161319793	rs9355851	No Data
chr6:161318424	rs2465878	No Data
chr6:161315894	rs1937474	No Data
chr6:161315215	rs2489955	No Data
chr5:162998514	***rs294588**	No Data
chr5:162997441	rs985253	No Data
chr5:152735784	rs160172	No Data
chr5:152732840	rs159973	No Data
chr5:152728481	rs302738	No Data
chr5:152724543	rs159759	No Data
chr5:152723632	rs300335	No Data
chr5:152723559	rs186183	No Data
chr5:152723340	rs300332	No Data

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr5:152719588	rs170029	No Data
chr5:152719062	rs10531959	No Data
chr5:152718835	rs172570	No Data
chr5:152718704	rs172569	No Data
chr5:152718481	rs2560252	No Data
chr5:152718436	rs170028	No Data
chr5:152718104	rs304858	No Data
chr5:152718050	rs304857	No Data
chr5:152717401	rs167357	No Data
chr5:152717343	rs170027	No Data
chr5:152716650	rs2560249	No Data
chr5:152716434	rs2560248	No Data
chr5:152716370	rs2609673	No Data
chr5:152716206	rs2446429	No Data
chr5:152716159	rs2617268	No Data
chr5:152716080	rs2615170	No Data
chr5:152715562	rs2615172	No Data
chr5:152715354	rs2927584	No Data
chr5:152715084	rs2617267	No Data
chr5:152714137	rs138445161	No Data
chr5:152713774	rs2262711	No Data
chr5:152712582	rs2260878	No Data
chr5:152712554	rs2609661	No Data
chr5:152711946	rs2615173	No Data
chr5:152710987	rs12655396	No Data
chr5:152704460	rs300326	No Data
chr5:152703973	rs300327	No Data
chr5:152703034	rs300328	No Data
chr5:152701152	rs2615176	No Data
chr5:152700032	rs2615178	No Data
chr5:152699648	rs1026615	No Data
chr5:152693366	rs1462114	No Data
chr5:152693290	rs1462113	No Data
chr5:152686943	rs308270	No Data
chr5:152685454	rs308267	No Data
chr5:152677992	rs2560229	No Data
chr5:152677966	rs2609665	No Data
chr5:152676683	rs151011228	No Data
chr5:152671348	rs6580010	No Data
chr5:152667576	rs4958634	No Data
chr5:152664041	rs10463328	No Data
chr5:152660234	rs2964816	No Data
chr5:152658667	rs1462123	No Data
chr5:152658547	rs1462122	No Data
chr5:152657104	rs2973152	No Data
chr5:152653524	rs2964821	No Data
chr5:152652975	rs2125515	No Data
chr5:152652413	rs1462108	No Data
chr5:152651764	rs2199123	No Data
chr5:152650210	rs2964812	No Data
chr5:152649497	rs2926286	No Data
chr5:152649412	rs1824311	No Data
chr5:152647989	rs1381540	No Data
chr5:152647817	rs2926287	No Data

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr5:152645407	rs2973149	No Data
chr5:152645389	rs2927172	No Data
chr5:152645212	rs982494	No Data
chr5:152641907	rs3101482	No Data
chr5:152639676	***rs3112530**	No Data
chr5:152639494	rs3101130	No Data
chr5:152639234	rs3101485	No Data
chr5:152638352	rs1462126	No Data
chr5:152632472	rs2974110	No Data
chr5:152625451	rs3101131	No Data
chr5:152619754	rs2615153	No Data
chr3:168687799	rs75809958	No Data
chr3:168677567	rs139566210	No Data
chr3:168673120	rs13316589	No Data
chr3:168662496	rs9831461	No Data
chr3:168661890	rs113340333	No Data
chr3:162716703	rs5854010	No Data
chr3:162714735	rs6798456	No Data
chr3:162714633	rs13084207	No Data
chr3:162711438	rs1580515	No Data
chr3:162710156	rs12488794	No Data
chr3:162707044	rs12489054	No Data
chr3:162706791	rs9864520	No Data
chr3:162706072	rs9863458	No Data
chr3:162705232	rs9820439	No Data
chr3:162705062	rs9858129	No Data
chr3:162704997	rs9877942	No Data
chr3:162704253	rs13318277	No Data
chr3:162704222	rs9852069	No Data
chr3:162703242	rs1488170	No Data
chr3:162702309	rs62396436	No Data
chr3:162701302	rs12486577	No Data
chr3:162701211	rs12486590	No Data
chr3:162699910	rs4276209	No Data
chr3:162699812	rs4611845	No Data
chr3:162698848	rs13096974	No Data
chr3:162698351	rs1955093	No Data
chr3:162698252	rs1955091	No Data
chr3:162697650	rs988191	No Data
chr3:162695902	rs1386859	No Data
chr3:162695319	rs9879670	No Data
chr3:162694179	rs1844075	No Data
chr3:162694175	rs1844074	No Data
chr3:162693177	rs13073152	No Data
chr3:162693152	rs2362561	No Data
chr3:162693062	rs1548002	No Data
chr3:162692995	rs1548001	No Data
chr3:162690366	rs981917	No Data
chr3:162690327	rs981916	No Data
chr3:162687312	rs138502324	No Data
chr3:162686145	rs9839942	No Data
chr3:162686051	rs9877131	No Data
chr3:162684901	rs75411070	No Data
chr3:162683855	rs10212444	No Data

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr3:162682264	rs9847141	No Data
chr3:162681267	rs141210220	No Data
chr3:162681231	rs62396422	No Data
chr3:162680571	rs77420327	No Data
chr3:162680501	rs140145501	No Data
chr3:162680483	rs77529894	No Data
chr3:162679861	rs76229349	No Data
chr3:162679462	rs144050070	No Data
chr3:162679424	rs10936320	No Data
chr3:162679241	rs10936319	No Data
chr3:162678882	rs9867733	No Data
chr3:162678857	rs9830442	No Data
chr3:162676081	rs11919599	No Data
chr3:162673400	rs9847253	No Data
chr3:162672866	rs994410	No Data
chr3:162672689	rs994408	No Data
chr3:162670731	rs34861578	No Data
chr3:162669091	rs1120907	No Data
chr3:162668181	rs9875594	No Data
chr3:162666486	rs9865699	No Data
chr3:162660850	rs4099484	No Data
chr3:162658440	rs9851215	No Data
chr3:162657856	rs6790172	No Data
chr3:162656951	rs12494878	No Data
chr3:162651944	rs7355808	No Data
chr3:162651107	rs3844503	No Data
chr3:162649357	rs9290153	No Data
chr3:162641776	rs9854933	No Data
chr3:162640496	rs11914874	No Data
chr3:162636655	rs9866192	No Data
chr3:162633606	rs13094783	No Data
chr2:238271283	rs10167850	No Data
chr2:141721469	rs12614087	No Data
chr2:141721364	rs12620377	No Data
chr2:141721158	rs10928082	No Data
chr2:141721141	***rs12474609**	No Data
chr2:141720639	rs10928081	No Data
chr1:44282203	rs972444	No Data
chr1:44279710	rs3791063	No Data
chr1:44279222	rs3791062	No Data
chr1:44277599	rs12125838	No Data
chr1:44276061	rs3791059	No Data
chr1:44273196	rs12126352	No Data
chr1:44272528	rs11580676	No Data
chr1:44271655	rs812490	No Data
chr1:44267810	rs12128547	No Data
chr1:44267618	rs12128485	No Data
chr1:44264497	rs7515003	No Data
chr1:44264434	rs975761	No Data
chr1:44263615	rs12116642	No Data
chr1:44261207	rs11210925	No Data
chr1:44254513	rs6701645	No Data
chr1:44253297	rs1875653	No Data
chr1:44244339	rs12750525	No Data

Coordinate (0-based)	dbSNP ID	RegulomeDB score
chr1:44233531	rs11590088	No Data
chr1:44230884	rs3791045	No Data
chr1:44215393	rs2158955	No Data
chr1:44214195	rs1990195	No Data
chr1:44212282	rs35052091	No Data
chr1:44211250	rs10890277	No Data
chr1:44208121	rs10890276	No Data
chr1:44202713	rs3838465	No Data
chr1:44200101	rs11210910	No Data
chr1:44195301	rs11210909	No Data
chr1:44194511	rs11210908	No Data
chr1:44189694	rs11210906	No Data
chr1:44184060	rs12139239	No Data
chr1:44179232	rs11210903	No Data
chr1:44178845	rs7536221	No Data
chr1:44178831	rs6429635	No Data
chr1:44163885	rs3815268	No Data
chr1:44162906	rs3791039	No Data
chr1:44161979	rs3791038	No Data
chr1:44152287	rs6683825	No Data
chr1:229019147	***rs11585386**	No Data
chr13:48385772	rs17424137	No Data
chr12:51703833	***rs766903**	No Data
chr12:51698769	rs2011124	No Data
chr12:51694277	rs4761832	No Data
chr12:51694155	rs4762016	No Data
chr12:51691412	rs7134625	No Data
chr12:129481439	rs470602	No Data
chr12:129481021	rs497428	No Data
chr12:129480704	rs470529	No Data
chr12:129480152	rs4307761	No Data
chr12:129479806	rs470458	No Data
chr12:129479298	rs604517	No Data
chr12:129479033	rs2675927	No Data
chr12:129479005	rs2266466	No Data
chr12:129478177	rs1718514	No Data
chr12:129478087	rs1621538	No Data
chr12:129477351	rs504501	No Data
chr12:129477080	rs12812667	No Data
chr12:129476936	rs470837	No Data
chr12:129476786	rs509934	No Data
chr12:129476717	rs510134	No Data
chr12:129476668	rs470758	No Data
chr12:129476522	rs511863	No Data
chr12:129476387	rs487407	No Data
chr12:129476096	rs674729	No Data
chr12:129476078	rs9668939	No Data
chr12:129475938	***rs643473**	No Data
chr12:129475548	rs1795681	No Data
chr12:129475251	rs3914957	No Data
chr12:129475001	rs470424	No Data
chr12:129474955	rs470427	No Data
chr11:124017492	***rs4936894**	No Data