From the Institute of Experimental Dermatology

Director: Prof. Saleh Ibrahim

Systems genetics in polygenic autoimmune diseases

Dissertation

for Fulfillment of Requirements

for the Doctoral Degree of the University of Lübeck

from the Department of Medicine

Submitted by

Yask Gupta

from New-Delhi, India

Lübeck 2015

First referee:- Prof. Dr. med Saleh Ibrahim

Second referee:- Prof. Dr. rer. nat. Jeanette Erdman

Chairman:- Prof. Dr. Horst-Werner Stüzbecher

Data of oral examination:- 7 June 2016

Contents

1. Introduction

The complexity of the human body has fascinated scientists for centuries. To study it, the biology-based interdisciplinary field called 'systems biology' was introduced (Kitano, 2002a, b). According to its main principle, in order to understand a biological system, one can modulate it, monitor the response, and formulate a mathematical model to describe those semantics. In system genetics, natural variation in the whole population perturbs biological processes (Civelek and Lusis, 2014; Farber, 2012; Feltus, 2014). Thus, the ultimate goal of system genetics is to understand the complexities of biological traits, by the identification of genes, pathways and networks associated with human diseases (van der Sijde et al., 2014). Genome-wide association studies (GWAS) have enabled the identification of genetic polymorphisms (SNPs) that are correlated to disease phenotypes (Gregersen et al., 2012). However, while data from GWAS studies may statistically determine single gene associations, it is not sufficient to explain molecular mechanisms (Farber, 2012).

Thus, the development of adequate mouse models mimicking human diseases is required (Kung and Huang, 2001). Previously, the generation of recombinant inbred mouse strains was used as a powerful tool for mapping the genomic loci underlying

the complex traits (QTLs) (Gonzales and Palmer, 2014). However, this data requires further fine mapping to eventually indicate the exact causal genes of deviant phenotypes (Mott and Flint, 2002). To shorten this process and dissect the molecular effects, gene expression data from inbred strains can be combined with genomic variations to identify the genes that control the traits (eQTL) (Cookson et al., 2009).

The discovery of micro RNA (miRNA) introduced an additional level of complexity to the molecular mechanisms leading to disease (Lee et al., 1993). This class of small non-coding RNA molecules has been reported to control gene expression by diverse mechanisms (Capuano et al., 2011; Dai et al., 2013; Paraboschi et al., 2011). Similar to eQTL, the genomic loci controlling the miRNA expression can also be derived using chip technology (Liu et al., 2012). An observed statistical association of miRNA levels with mRNA levels is indicative of a mutual or joint genetic control. We seek those genetic loci for which variations affect these statistical interactions. Preferably, that locus is close ('cis') to the regions coding either of the two interacting genes. This, in turn, can lead to the identification of possible molecular pathways and help to unravel new therapeutic approaches.

Recently, additional classes of non-coding RNAs such as snoRNA, snRNA, telomerase RNA etc. have been discovered to contribute to disease phenotypes (Esteller, 2011). Therefore, to gain an overview of the complete gene network, the integration of additional classes of non-coding RNA beyond miRNA is required.

However, the current computational algorithms cannot yet predict all the existing classes of non-coding RNA. Hence, in this work, I sought to address the following aims: i) To find the possible interaction network of genes and miRNAs that play an important role in the pathogenesis of autoimmune skin disease, using experimental Epidermolysis Bullosa Acquisita (EBA) as a model for autoimmune diseases. ii) To develop software which can predict different classes of post-transcriptional RNAs (a class of non-coding RNA) based on their sequences and structural properties using machine learning algorithms.

1.1 Autoimmune diseases

One of the main functions of the immune system is to discriminate between self and non-self antigens (Parkin and Cohen, 2001). Failure to differentiate between self and foreign antigens can lead to a sustained immune response and development of autoimmune diseases (ADs). More than 80 ADs have been identified, affecting nearly 100 million people worldwide (Cooper et al., 2009; Cooper and Stroehla, 2003; Jacobson et al., 1997). ADs are differentiated from other types of diseases on the basis of of Witebsky´s postulates. According to these postulates ADs should fulfil at least two or more of the following criteria: 1) a specific adaptive immune response directed to an affected tissue or organ; 2) presence of auto-reactive T cells or auto-antibodies (AAbs) in the affected organ or tissue; 3) induction of the disease in healthy humans or animals by transfer of auto-reactive T cells or AAbs; 4) induction

of the disease in animals by immunization with auto-antigen; 5) immune-modulation and suppression of autoimmune response. Numerous risk factors have been associated with ADs, however the complete pathoaetiology of these diseases remains unknown (Rioux and Abbas, 2005). Nonetheless, it is clear that both genetic and environmental factors play a critical role in the pathology of autoimmunity.

1.2 Autoimmune bullous diseases

Production of auto-antibodies (AAbs) directed against different structural proteins of the skin and the mucous membranes lead to the development of organ-specific autoimmune disorders in the skin called autoimmune bullous disease (AIBDs) (Sticherling and Erfurt-Berge, 2012). AIBDs can be classified into four distinct groups based on their clinical, histological and immunopathological criteria, i.e. pemphigus and the group of pemphigoid diseases, and dermatitis herpetiformis Duhring (Zillikens, 2008). In the group of pemphigus diseases, AAbs are directed against desmosomal proteins, leading to a loss of adhesion between adjacent epidermal keratinotcytes and intra-epithelial blister formation (Amagai et al., 1995). In pemphigoid diseases, AAbs are formed against hemidesmosomal proteins, resulting in the dysadhesion of basal keratinocytes to the underlying basement membrane (BM) and formation of subepidermal blisters (Borradori and Sonnenberg, 1999).

1.3 Epidermolysis bullosa acquisita (EBA): an autoantibody-mediated autoimmune skin disease

Epidermolysis bullosa acquisita (EBA) is a severe chronic inflammatory subepidermal ABD which is characterized by the production of AAbs against type VII collagen (Ludwig, 2013). Type VII collagen is a major component of anchoring fibrils that is situated at the DEJ (dermal epidermal junction). Though some EBA patient sera show antibody response to NC2 domain, but most patient's IgGs recognize epitopes mapped to the NC1 domain of type VII collagen (Csorba et al., 2014). EBA can be classified into two main categories: the mechanobullous (classical) and inflammatory forms. In the prior form, the blister appears on the sites of friction followed by erosions, scarring, and milia formation (Niedermeier et al., 2007). Mucosal lesions are also seen in this form. The inflammatory type of EBA, accounting for 2/3 of EBA cases, is associated with widespread vesiculobullous eruptions and characterized by neutrophilic infiltration (Jonkman et al., 2000).

1.4 Role of miRNA in autoimmune diseases

The discovery of microRNA, a class of small endogenous non-coding molecules ranging from 18-24 nt, brought a new level of complexity into understanding the mechanisms underlying various biological processes (Lee et al., 1993). The involvement of these small non-coding elements in the control of gene expression has been thoroughly defined in cell cycle, metabolism, immunity, and cancer studies

(Abu-Hijleh et al., 1997; Calin and Croce, 2006; Rottiers and Naar, 2012; Wang and Blelloch, 2011). They can regulate multiple pathways by binding to the 3' UTR region of the DNA transcript, thereby leading to changes in target gene expression levels (Gommans and Berezikov, 2012).

Approximately half of miRNAs are coded by intergenic regions in the genome (Issabekova et al., 2012). Many miRNAs are located within introns present in the genes (Lin et al., 2006). They are thought to have their own enhancers and promoters and are transcribed by RNA polymerase II (Kurihara and Watanabe, 2010). However, it remains unclear whether they are produced as by-products of protein-coding gene transcription or whether they have their own machinery (Kim et al., 2009). Recent studies in murine and human fibroblasts in liver tissues have revealed that regulation of miRNA can either be controlled from their genomic location or from other regions in the genome (Kim et al., 2009). While, advancements have been achieved in the general understanding of regulatory mechanisms for the expression of miRNA in various tissues, very little is known about their own regulation in the skin.

In vivo studies suggest that a lack of enzymes such as Dgcr8 and Dicer contributes to miRNA processing in the skin, causing severe phenotypes (Krill et al., 2013; Wang et al., 2007). They are also found to have important regulatory functions in the morphogenesis and homeostasis of the skin (Schneider, 2012). Differentially

expressed miRNAs are associated with different physiological and pathological skin processes such as melanoma, Sézary syndrome, psoriasis and atopic dermatitis (Ballabio et al., 2010; Bonazzi et al., 2012; Ichihara et al., 2011; Sonkoly et al., 2010).

1.5 Genes in autoimmune diseases

The development of high-throughput technologies led to the identification of multiple genes that are associated with autoimmunity (Richard-Miceli and Criswell, 2012). As expected, the HLA locus has been confirmed as one of the dominant regions in the genome, which is thought to contribute to autoimmune diseases (Eastmond, 1994; Kollaee et al., 2012; Trachtenberg et al., 2000). Apart from the HLA locus, scientists have also identified non-HLA related genes, including interleukins, that play an important role in the induction of autoimmune diseases (Beebe et al., 2002; Geng et al., 2012; Tsokos and Fleming, 2004; Wang et al., 2010; Yao et al., 2014). Other genes such as RGS1, LPP, and MYO9B, etc. have been found to be associated with celiac disease (Dubois et al., 2010). NOD and TNF-alpha genes have been associated with Crohn's disease and diabetes (Koulmanda et al., 2012; Lala et al., 2003). Transcriptional factors such as STAT4 have been connected to the pathophysiology of Sjogren syndrome and lupus erythematosus (Gestermann et al., 2010; Namjou et al., 2009). These genes have been shown to be involved in multiple pathways where the cytokine-cytokine pathways and chemokine pathways have

been the major focus (Blanco et al., 2008; Carlson et al., 2008). To study their molecular mechanisms and putative pathways mouse models have been implemented. Data generated from these studies have been deposited in common databases, from where they can be derived to further study the underlying molecular mechanisms using experimental and computational methods (Beck et al., 2014; Welter et al., 2014).

1.6 Gene networks

The availability of the aforementioned databases allows researchers to describe putative hypothetical networks between genes and their interactions on different cellular levels across tissues. These networks are called gene regulatory networks (van Someren et al., 2002). They provide systematic molecular mechanisms that underlie various biological processes (Karlebach and Shamir, 2008). Moreover, based on the number of interacting partners, one can further determine the causal genes and the drivers that influence genes downstream (Studham et al., 2014). Similar studies have been implicated in cancer studies (Delfino and Rodriguez-Zas, 2013; Emmert-Streib et al., 2014). The identified networks are based on the Bayesian network approach, information theory, correlation and partial correlation (Johansson et al., 2011; Reverter and Chan, 2008; Schmitt et al., 2004; Zou and Conzen, 2005). These methods use expression profiles in combination with available

knowledge to drive putative hypothetical networks that can be subsequently studied in experimental set-ups.

1.6.1 Gene and protein interaction database

A protein-protein interaction (PPI) database includes a wide range of information about cell-cell interaction, metabolism, development and biological processes across various phenotypes (Ooi et al., 2010). The major objective of these databases is to provide abstract knowledge about mechanical pathways. Therefore they are considered to be an important part of systems genetics (Xenarios and Eisenberg, 2001). Based on these databases one can derive functional knowledge of unknown proteins in a cell (Pattin and Moore, 2009). They can also be used to determine their molecular targets and identify new potential drugs (Fuentes et al., 2009). The major source of information for these databases has been derived from data generated in two-hybrid system, mass spectrometry, phage display, and protein chip technology (Beckmann et al., 2005; Bruckner et al., 2009; Gstaiger and Aebersold, 2009; Nariai et al., 2004). Well known databases include BIOGRID, MINT, and STRING. (Chatr-Aryamontri et al., 2013; von Mering et al., 2003; Zanzoni et al., 2002). Databases such as STRING predict gene and protein interactions based on data mining algorithms that derives information from publicly available scientific journals (Franceschini et al., 2013). The score for each protein or gene interaction depends upon the validity of the interactions in the experimental set ups and the number of

times it has reappeared in various texts (Zhang et al., 2010). These databases also use algorithms that can verify the same information across various species.

Furthermore, they have added statistical methods for the functional enrichment of given interaction network. Apart from text based databases for protein-protein interactions (PPI), computational scientists have also designed databases which contain PPI information from protein structures, as found in the DOMINE database (Raghavachari et al., 2008). This database contains information about domain interactions derived from a protein structure database, the Protein data bank (PDB) (Berman et al., 2000). Furthermore, it also predicts domain interactions based on eight different algorithms. The protein and gene interaction information serves as an important base for removing the falsely identified interactions that are predicted by statistical methods. Some of the commercial databases and software combine all this information into one meta-database such as IPA (Kramer et al., 2014). The IPA software can be used for various purposes and it is composed of many different databases including BIOGRID, MINT and its own independent dataset of protein interaction networks (Chatr-Aryamontri et al., 2013; Zanzoni et al., 2002). It also contains tools for the functional enrichment of different networks. The database can assist in identifying a novel network for a specific phenotype.

1.6.2 *Ab intio* prediction of gene networks

In addition to the aforementioned databases, various statistical methods can be used to generate gene networks. Gene expression profiles can be deduced using high-throughput chip-based technologies. Subsequently, statistical methods are used to cluster these profiles and generate putative networks (Eisen et al., 1998). Methods for generating such hypotheses include 'clustering' and 'information theory' based approaches. These approaches have been implemented in R packages such as WGCNA and c3net (Altay and Emmert-Streib, 2011; Langfelder and Horvath, 2008). WGCNA (weighted gene co-expression network analysis) is one of the most commonly used cluster based methods. In brief, the genes are grouped into small clusters using their co-expression profiles. This is done through hierarchal clustering and dynamic cut-tree algorithms (Langfelder et al., 2008). The algorithm uses a scale-free topology to identify the average interaction of each gene. Then, the first principle component from each cluster is correlated to the phenotypic score. The cluster showing a significant correlation to phenotype is considered to contain genes and pathways required for the variation in the phenotype. Despite the high success and importance of such algorithms, the interaction among the genes within a cluster was based upon hypothetical scores generated by algorithm-based topological overlap methods. Therefore, the involvement of additional algorithms for further identifying relationships between genes can remove some false positives. The use of information theory and regression-based methods has been helpful in identifying the

networks in a small subset of genes. Some of the well-known statistical methods such as c3net and PLS regression have shown their utility in constructing gene networks and discovering causal genes. C3net infers the direct physical network of genes from expression data using information theory. It helps in identifying the causal network for a set of genes. PLS regression is based on partial least square regression for generating putative hypotheses for interaction (Pihur and Datta, 2008). Applying all these methods together can generate a putative gene network, which could be referred to as a mechanical pathway for a phenotype. The interactome produced by these methods can be confirmed by the database, and novel interaction can be identified.

1.7 QTL and expression QTL

Mouse models serve as an important source for understating how genes control various molecular phenotypes. Using mouse models and system genetics, diseases have been investigated for genetic loci contributing to quantitative traits (QTL) (Broman, 2001). The QTLs which represent a region of the chromosome are further fine mapped to identify the possible causal genes contributing to variation in traits (Gonzales and Palmer, 2014). The techniques have been applied across various species such as yeast, mouse, and plants (Liti and Louis, 2012; Young, 1996). Recently, due to the availability of deep sequencing technology and chip-based technology, the traits not only include molecular phenotypes and disease score but

the expression of individual genes. Therefore, quantitative trait loci which can regulate the expression of the genes are termed expression QTL or eQTL (Nica and Dermitzakis, 2013). Expression QTL is primarily used to identify the transcriptional activity of genes. The identification of the eQTL for a specific molecular phenotype not only reveals the causal gene marker, but also sheds light on various mechanisms involved in the generation of the phenotype (Li et al., 2014). eQTL data can be combined with the co-expression data generated from the same cohort to provide vital knowledge about the putative mechanistic pathways contributing to the phenotype (Grieve et al., 2008). For example, various mechanism in multiple tissues leading to abnormal phenotype have been studied using eQTL analysis (Flutre et al., 2013).

1.8. Prediction of ncRNA

There are many different types of RNA with multiple functions in the cell. Some RNA molecules contribute to the translation of genetic information into proteins and the regulation of genes. Others function as enzymes by catalysing biological reactions. While the non-coding regions in the genome were first believed to be senseless sequences, they have been shown to code for RNA families that play important roles in the eukaryotic cell. These non-coding RNAs (ncRNAs) include types of RNA that do not code for protein, but are involved in many regulatory processes and can be divided into a tremendous variety of highly plethoric and

versatile families that are essential for cellular function (Carninci et al., 2005). Other than their general function, they are also found to be associated with different phenotypes of diseases (Esteller, 2011). Hence, they form a vast and to a large extent unexplored reservoir of potentially valuable medical biomarkers (Beck et al., 2011; Kim and Reitmair, 2013). In order to identify these RNA-based biomarkers, modern techniques such as next-generation-sequencing and microarray-technologies are employed (Bompfunewerer et al., 2005; Jung et al., 2010). These techniques provide an immense amount of data and offer ample opportunities to identify novel classes of non-coding RNA. However, the experimental analysis of new sequences is time-consuming and complex, indicating the need to find alternative approaches for their analysis. Promising and auspicious approaches are given by *in-silico* methods. Due to phylogenetic relationships, sequences of non-coding RNA show similarities regarding their properties. They can be divided into subclasses based on their conserved properties, meaning sequence conservation and structural conservation (Lu et al., 2011). Computational methods, such as classification tools, offer a fast and reliable way to analyse and classify sequences by exploiting conserved properties among the sequences (Artzi et al., 2008).

Various classification systems have been developed to predict different subsets of RNA, using machine-learning and phylogenetic approaches (Hertel and Stadler, 2006; Lagesen et al., 2007; Laslett and Canback, 2004; Lowe and Eddy, 1999). So far, tRNAs can be detected reliably using tRNAScan-SE (Lowe and Eddy, 1997).

Furthermore, various approaches have been established to detect miRNA (Yoon and De Micheli, 2005) and other small RNA subsets. Recently, snoReport was introduced, which is designed to recognize small nucleolar RNA (snoRNA) from the genome without using any target information (Hertel et al., 2008). Most of these systems achieve a satisfying accuracy. However, not every RNA family can be predicted. For example, to this point, there is no tool for the prediction of small nuclear RNA (snRNA), Ribonuclease P (RNase P), Ribonuclease MRP (RNase MRP), Y RNA and telomerase RNA. Facing the continuing increase in the number of human RNAs in databases such as Rfam (Griffiths-Jones et al., 2003) , it is necessary to extend the current possibilities of RNA prediction. SnRNA), RNase P, RNase MRP, Y RNA and telomerase RNA have the common characteristic, besides snoRNA, they are involved in post-transcriptional modification or DNA replication in eukaryotes (Kiss, 2001; Lustig, 1999; Pannucci et al., 1999; Perreault et al., 2007; Thore et al., 2003).

1.8.1 Support vector machine and random forest

Various methods have been developed for solving classification problems. Computational algorithms such as random forest and support vector machines have been the primary choice of researchers (Byvatov and Schneider, 2003; Sun, 2010; Ziegler Andreas, 2014). These algorithms are based on the concept that labeled training data is first used to develop a classification model. Afterwards, the

classifier is used for classifying novel data. A classification system requires two kinds of input. The first is the description of the object, i.e. features or values by which the object can be described. These features are used to train the classifier and create a model. The second is the group label of each object. Classification analysis, then finds the mapping function that can feature the class labels. Different methods used for formulating the classifier aims in essence at minimizing the error probability of the training set. One such method is support vector machines (SVM).

As compared to other algorithms, SVMs have an outstanding track record for classifying biological data (Statnikov and Aliferis, 2007). Although many important mathematical models present in SVM were used in machine learning from the 1960s, they were officially proposed by Vladimir Vapnik and his co-workers in 1992 (Boser, 1992). Since then, it has been widely used in classification of biological data (Naul, 2009; Yang, 2004). One of the best-known biological applications of SVM is the molecular classification of microarray gene expression data (Mukherjee, 2003), where it has shown statistical and clinical relevance for a variety of tumour types: Leukaemia (Golub et al., 1999), Lymphoma (Shipp et al., 2002), Brain cancer (Pomeroy et al., 2002), Lung cancer (Bhattacharjee et al., 2001) and the classification of multiple primary tumour (Ramaswamy et al., 2001). Additionally, it is used for recognition of translation initiation sites in DNA, protein fold recognition, protein-protein interaction, protein secondary structure prediction, protein localization, etc. (Bock and Gough, 2001; Ding and Dubchak, 2001; Hua and Sun,

2001; Zien et al., 2000). Detailed implementation of SVM for this work is provided in the *method section* 3.9.3. An important aspect of SVMs is the 'features' used to classify the label of the object. Multiple numbers of irrelevant features can lead to 'over-fitting' describing noise or random error, resulting in a false positive model. Therefore, feature selection is a crucial process when designing a classifier. Methods such as PCA, PLS and random forest etc. are employed for feature selection in many studies (Le Cao et al., 2011; Menze et al., 2009; Mishra et al., 2011). In addition to their importance in modeling classifiers, these methods have been used for the identification of causal genes from gene expression profiling (Chen and Ishwaran, 2012). One of the most used algorithms is random forest. Random forest is an ensemble learning method for classification and regression that was developed by Leo Breiman in 2001 (Breiman, 2001). In the last decade, the algorithm has shown excellent performance where the number the variables are much larger than the number of observations. Furthermore, it can efficiently deal with complex interaction structures and highly correlated variables and return measures of variable importance (Boulesteix, 2012). The algorithm has been used in multiple biological application such as data mining (Ziegler Andreas, 2014), classification of microarray data (Cutler and Stevens, 2006), GWAS data (Botta et al., 2014; Schwarz et al., 2010), etc. In brief, the algorithm draws several bootstrap samples from the training sets and aggregates many binary decision trees. The samples are drawn uniformly with repetitions. In the algorithm, the dataset is divided based on

randomly selected features and the best split is calculated. The process is repeated several times until the best split for the dataset is observed. The trees are aggregated and the best features for the splits are calculated through the majority of votes. The out-of-bag (OOB) error rate is calculated for variable selection. In this, data in the training sample are fixed and all bootstrap samples not having this class are considered (data out of the bag). Then the majority of the votes only among these samples are determined and compared to the real class to get the prediction error (cross-validation). To evaluate the variable importance, the procedure is permutated. The larger the error becomes upon permutation of values assigned to a single attribute, the more important the variable is considered to be. Therefore, this approach helps in the selection of variables which can be used for classification problems in biological systems. Additional details of random forest and implementation of algorithm in this work are provided in the Method section 3.9.3.

1.9. Structure of the thesis

The thesis is divided into three major parts. The first part describes the eQTL mapping and networks derived from the miRNA expression profiling. Methods and protocols for the analysis are described in sections 3.4, 3.5, 3.6.1, 3.7 and 3.8. The results of the analysis are presented in section 4.1. The second part of the thesis describes the eQTL mapping and networks derived from gene expression profiling. The methods and protocols implemented for the second part are identical to the first

part apart from the thresholds for generation of *de novo* gene networks, which are described in section 3.6.2. The results of the second part are presented in section 4.2. The application of databases in prediction of gene networks has been published in the journal *Exp. Dermatol* (Gupta et al., 2013). This part of the thesis also consists of meta-networks of genes and miRNAs. These results are described in section 5.3. The hub genes (protein-coding genes and miRNAs) and the network are under experimental validation. The third part of the thesis describes a bioinformatics tool for predicting post-transcriptional non-coding RNA, *ptRNApred*. The methods and protocols for this work are provided in section 3.9 and results are presented in section 4.4. The work described in the fourth section has been published in *NAR (Nucleic Acid Research)* (Gupta et al., 2014).

2. Aims of the thesis

The primary aim of the work is to develop an analytical workflow to understand gene networks involved in the pathogenesis of skin autoimmune blistering diseases such as EBA. Therefore, this work proposes the integration of novel genetic data generated in the lab (SNPs) with multiple molecular and clinical phenotypes. This is achieved by statistically associating miRNAs and gene expression profiles with SNPs and by further integrating co-expression analysis to identify gene-gene and gene-miRNA networks. Additionally, gene networks need to be verified using multiple publicly available genes and protein interaction databases. The identification of novel putative gene networks would provide novel gene candidates (hub genes in gene networks) and suggest molecular mechanisms for experimental intervention.

Recently, it has been identified that, in addition to miRNAs, other classes of non-coding RNAs known as post-transcriptional RNAs may also play an important role in the mechanisms of different clinical and molecular phenotype. Therefore, the second aim of the work is to develop a bioinformatics tool to differentiate between non-post-transcriptional RNA and post-transcriptional RNA and to identify subclasses of post-transcriptional RNA, which includes snoRNA, snRNA, telomerase RNA, Y RNA, RNase P and RNase MRP. The identification of novel non-coding RNA

can further facilitate the understanding of the molecular mechanisms behind clinical phenotypes.

3. Materials and methods

To present the complete context of this thesis and for scientific reproducibility, sections 3.1 to 3.3 describe the work performed by the institute's technicians that generated the data on which this thesis is based. The thesis work itself is described in sections 3.4 and after.

3.1 Generation of a four way advanced inter-cross line

An advanced inter-cross line (AIL) was generated by inter-crossing parental strains (MRL/MpJ, NZM2410/J, BXD2/TyJ, CAST/EiJ) that were purchased from the Jackson laboratory (Maine, USA). Briefly, these strains were intercrossed at an equal strain and sex distribution to generate a genetically diverse mouse line. 50 breeding pairs from each generation (from F1-G4) were maintained to keep an equal distribution of parental alleles. Males and female offspring were separated into distinct cages after weaning. The genetic diversity between mice was observed by different morphological-phenotypic traits such as weight, coat color and tail length. The animals were held under specific pathogen-free conditions in a 12-hour light/dark cycle with food and water *ad libitum*. All animals in this study were taken from the fourth generation of AIL. All animal experiments were approved by the state of Schleswig-Holstein, Germany.

3.2 Experimental EBA and observation protocol

Active EBA was induced by the immunization of 300 mice with murine collagen 7 (mCol7c-GST). In brief, 60µg of mCol7c-GST were emulsified in 60µl of adjuvant (TiterMax, Alexix, Lörrach, Germany) and injected subcutaneously into the footpads of mice. Following immunization mice were assessed for the EBA phenotype. The extent of disease was determined by the percentage of affected body surface area every 4th week for a period of 12 weeks. The observed scores were later classified into 'Low' (0>x<5), 'Moderate' (5>x315) and 'Severe' EBA (x>15) categories, where x is the percentage of surface, body area affected by erosions, lesions and alopecia. At the end of week 12, the mice were sacrificed and, ear and skin samples were obtained for further analysis. Overall, in this study, 300 mice were immunized with mcol7C-GST protein. Ear skin samples were snap-frozen at -80^{o}C for subsequent processing.

3.2.1 Generation of recombinant peptides

The immunodominant mcol7c epitope of the murine NC1 domain (aa 757-967) was fused to the GST tag and expressed in prokaryotic systems. The recombinant protein (mcol7c-GST) was further purified using glutathione-affinity chromatography as described before (Ludwig et al., 2012).

3.3 Extraction of genomic DNA for genotypic analysis

Genomic DNA was isolated from tail clippings and incubated in 500µl 50mM NaOH at 95°C for 1 hour. The reaction was neutralized by the posterior addition of 50 µl 1M Tris- HCl (pH 8.0). The DNA was further processed with the DNeasy Blood & Tissue Kit according to the manufacturer's instructions. The extracted DNA was quantified using Nanodrop and normalized to 50ng/µl in TE buffer (10 mM Tris; 1mM EDTA; pH= 8). Agarose gel electrophoresis was performed for quality control.

3.4 Generation of microarray data (miRNAs/genes) and bioinformatics analysis

To monitor gene expression in the skin samples derived from the AIL population, total RNA was extracted and hybridized to the Affymetrix miRNA 2.0 Array according to the manufacturer's protocol (Chalaris et al., 2010). The raw data was pre-processed using the R 'affy' package and normalized using RMA (Robust Multi-array analysis) to generate normalized expression levels across the samples (n =100) (Gautier et al., 2004).

For gene expression profiling, total RNA was hybridized to the Affymetrix Mouse Gene 1.0 ST Array. The raw data was processed using the 'oligo' R package. The RMA method incorporated in the R package was used for normalization of the probe intensities for 200 samples (Carvalho and Irizarry, 2010). For gene expression

profiling, the low intensity probes were omitted to avoid putative false positive signals and lower the barrier for the correction of multiple testing, i.e. these were filtered out using the median-based method as implemented in the function 'expressionBasedfilter' from the R DCGL package (Liu et al., 2010). Additionally, genes which were not showing significant variations ($p\text{-}value < 0.05$) across the samples were also filtered using the function 'varianceBasedfilter' of the same package. This function reduces the data to the most variable genes which are presumably critical for the phenotype.

3.5 Software for expression QTL analysis

In an expression QTL analysis, one describes the association of gene expression levels with a sequence variation (SNP) across individuals. The term differs from QTL analysis because the association is performed for genome-wide gene expression levels rather than a clinical phenotype. The analysis finds its edge, as it can identify causal genes, regulating the expression of other critical genes for a specific phenotype. Such analyses can be used to construct statistical networks, which in turn may yield vital information about various biological mechanisms underlying the molecular phenotypes.

To perform eQTL analysis, two software packages were used, i.e. HAPPY and EMMA (Kang et al., 2008; Mott et al., 2000). HAPPY first computes haplotype probabilities for all samples across all SNPs, without taking into account the

relationship between different individuals. Thereafter, it determines the association between the haplotype probabilities and the respective quantitative trait. Unlike HAPPY, EMMA uses mixed modeling to find associations between genotype and quantitative traits and integrates the kinship matrix in the analysis, which represents the degree of relatedness between individuals to correct for biases in the population structure between cases and controls. Thus, the combination of the two approaches can be used to perform eQTL mapping while filtering out false positive eQTL reported due to bias in the family structure of the cohort (Iancu et al., 2012).

3.5.1 HAPPY

HAPPY is an R package wrapping the initial C implementation (Mott et al., 2000) to calculate the haplotype probability for each sample and each marker on the basis of the haplotype distributions among the founder strains. The probability is calculated for all the genotypes by taking two consecutive SNPs into account using a hidden Markov model. Once the probabilities have been calculated, linear regression is performed between the calculated genotype probabilities and quantitative traits. To determine the threshold that is ensuring the statistical significance, the performance on permuted phenotype assignments is evaluated, i.e. the analysis is repeated 1000 times after a respective shuffling. The software needs to be provided with two files: (i) a marker file, containing information about the fraction of alleles for each SNP across the founder genotypes and (ii) a pedigree file, containing the

phenotype and genotype for all the samples. The input from the pedigree and marker files allows the calculation of likelihood for each SNP and each sample derived from a particular founder strain. Subsequently, for every SNP, statistical significance is computed by an ANOVA between the null hypothesis (covariates contribute to the phenotype alone) and alternate hypothesis (the genotype is added as an additional factor to the null model) to find an association between the phenotype and genotype probabilities of the founder strain using an additive Gaussian model. The null model is formulated as:

$$\boldsymbol{y} = \boldsymbol{X} + \boldsymbol{\varepsilon}$$

and the alternative hypothesis

$$\boldsymbol{y} = \boldsymbol{X}_{+}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$\boldsymbol{y}$ is a $n \times 1$ vector of observed phenotypes (miRNA/mRNA expression levels), X is a vector of fixed effects such as sex (0 = male, 1 = female), β is a n x q matrix for founder haplotype probabilities across strains, ε is residual error, n number of samples and q is number of founder strains.

3.5.2 EMMA

EMMA, like HAPPY, is implemented in R. Its distinguishing feature is revisiting disease associations under the consideration of random effects (Kang et al., 2008). The association is performed for each marker, not by haplotype group. Unlike HAPPY, EMMA does not require founder genotype information but derives the

relationship among individuals by a kinship matrix. The genotypes are coded 0 (homozygous dominant), 0.5 (heterozygous) and 1 (homozygous recessive). The matrix is decomposed using spectral decomposition. Spectral decomposition is a factorization of the matrix in canonical form whereby the matrix is represented in the form of eigengene values and eigengene vectors. The associations between each marker and quantitative trait are performed using linear mixed modelling. The decomposed kinship matrix was used as a random effect. The use of the kinship matrix while performing association captures variability due to the same family. Therefore, the only variation in the genotype which affects the quantitative traits is accessed. The association is calculated using a linear mixed model which is described by equation:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zh} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of phenotypes, β is a vector of fixed marker effects (i.e., single SNP), $\mathbf{h}$ is a vector of polygenic effects caused by relatedness, $\mathbf{e}$ is vector of residual effects, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating $\mathbf{y}$ to β and $\mathbf{y}$ to $\mathbf{h}$, respectively. It is assumed $\mathbf{h} \sim N(0, 2\mathbf{K}\sigma^2_G)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$, where $\mathbf{K}$ is an allele-sharing matrix i.e. kinship matrix calculated from the SNP data, σ^2_G is the genetic variance, $\mathbf{I}$ is an identity matrix, and σ^2_e is the residual variance (Lorenz et al., 2010).

3.5.3 Implementation of software for expression QTL mapping

Both software tools, HAPPY and EMMA were applied. Primary eQTL analysis was performed using HAPPY for both miRNA and mRNA expression levels. The miRNA and mRNA expression levels of mice (609 probes and 34000 probes) were considered as quantitative traits and linkage analysis was performed using an additive Gaussian model with 1200 SNPs across the whole genome. 1000 permutations were performed for each quantitative trait to detect if the associations are random or false positives. A significance threshold of $\alpha = 0.05$ (95th percentile) was defined as the cut-off for detecting true positive association for eQTL. Additionally, HAPPY provides a routine to determine combined SNP effects, i.e. epistasis, which was applied for miRNAs. The function "epistasis" from the HAPPY R package was used to evaluate SNP-SNP interactions (Civelek and Lusis, 2014). The interaction model for epistasis is calculated similarly to the additive model in HAPPY; the alternate hypothesis consists of an additional interaction term between the haplotype probabilities for two interacting SNPs. To access the statistical significance, *p-values* between the null and alternate hypothesis are calculated using ANOVA. All the *p-values* were corrected for multiple testing using the Bonferroni correction method.

EMMA, the R package for mixed modelling, was used to filter the eQTL which could be possibly obtained due to a family effect. We defined the confidence interval of the

given eQTL based on a -log *p-value* drop of 1.5. An eQTL was said to be cis-regulating when the probe was mapped within 15 cM on the same chromosome as its genomic location (Su et al., 2011). Otherwise it is referred to as trans-regulating.

To visualize the eQTL results we used the software Circos (Krzywinski et al., 2009). Circos uses a circular ideogram layout that facilitates the display of relationships between pairs of positions by the use of ribbons, which encode the position, size, and orientation of related genomic elements. Circos is open source software written in the programming language *Perl* and takes GFF (General feature format)-styled data as input. It is capable of displaying data as scatter- or line-plots, histograms, heat maps, tiles, connectors, or plain text.

3.6 *Ab initio* gene and microRNA network prediction

As described in *section 1.6.2*, in addition to eQTL mapping, we implemented various statistical methods to predict gene interaction networks from gene and miRNA expression profiles. These methods include; *Ab initio* methods such as WGCNA, c3net, PLS, etc. and database-driven methods such as STRING, IPA, etc. The various parameters and implementation of the algorithm for interaction networks in the current work is described below.

3.6.1 MicroRNAs

For *Ab initio* prediction, the genes were first clustered together based on their co-expression profiles across different samples. The standard WGCNA procedure was used for cluster or module detection (Langfelder and Horvath, 2008). For miRNA we considered 97 samples, where 3 samples were excluded as these were not phenotype for EBA. A weighted adjacency matrix of pair-wise connection strengths (correlation coefficients of gene expression levels) was constructed using the soft threshold approach with a scale-free topological power $\beta = 7$ (miRNA). A scale free topology is a network whose degree of distribution follows a power law. For each probe, the connectivity was defined as the sum of all connection strengths with all others. Probes were aggregated into modules by hierarchical clustering and refined using the dynamic cut-tree algorithm (Langfelder et al., 2008). The Pearson correlation coefficient was determined for each phenotype-module pair. The representative module expression profiles, or module eigengene values, are the first principal component of the gene expression profiles within a module. The correlation between the module eigengene and the sample trait of interest yields the eigengene significance, as assessed by a correlation test. The modules were assigned by different colors where grey was assigned to traits that could not be clustered in any other module. To determine the important pathways involved in the important cluster we used the online miSystems tool (Lu et al., 2012). Additionally, causal networks of the miRNA expression profiles were generated using the c3net R

package that uses the maximum mutual information value to calculate gene networks based on expression profiles (Altay and Emmert-Streib, 2011).

3.6.2 Genes

To detect the modules of gene expression profiles, we utilized a similar approach to that described in *section 3.6.1* with a modified threshold for scale-free topology i.e. the power of β was 6. Additionally, PLS regression was used to filter out false positive interaction among the genes in each module (Pihur and Datta, 2008). In brief, the PLS regression method measures associations between each pair of genes under the influence of all other genes present in the dataset. Thus, it assigns a weight or numerical measurement for each edge/interaction for every pair of genes. The statistical significance of these edges is calculated using an empirical Bayes technique which uses an fdr to assess significance (Efron, 2004).

For pathway and gene ontology the DAVID online software was used (Huang et al., 2007). This software is based on the modified Fisher exact test. The software performs gene ontology for the group of genes based on metabolic pathways, cellular component and biological function. In brief, the software has a pre-defined database in which various set of genes are assigned to ontology terms such as metabolic pathways: chemokine pathway, cancer pathway etc., and cellular compartment: mitochondria, nucleus, etc., and biological processes, etc. A list of genes is provided in the software by the user. Thereafter, the software calculates how many genes are

found for a specific ontology term. The significance of the ontology term for these genes is calculated using the modified Fisher exact test. The *p-values* obtained for each ontology term are corrected for multiple testing using Bonferroni corrections.

3.6.3 MicroRNA gene target prediction

Multiple approaches were applied to predict the gene targets for miRNAs. Since it is expected that miRNA down-regulates the gene expression, we calculated the Pearson correlation coefficient between all the miRNAs (609) and all the differentially expressed genes (1065) in the EBA disease phenotype. The *p-values* were corrected for multiple testing using Bonferroni's correction. The genes and miRNAs showing significant negative correlation (adjusted *p-value* < 0.05) were considered for further analysis. The miRNAs which passed the above described threshold were queried in database for their known and predicted gene targets. To retrieve gene targets we used the miRWalk database (Dweep et al., 2011). miRWalk is a publicly available database, which features predicted as well as experimentally validated microRNA (miRNA)-target interaction pairs. It can predict all miRNA binding sites for all the genes of three mammalian genomes (human, mouse, and rat). Furthermore, it allows miRNA-gene target prediction comparison between ten different databases (Dweep et al., 2014). It also hosts experimentally validated data derived from cell lines, diseases, miRNA processing proteins, etc.

3.7 Visualization of interaction networks

There are various tools for visualization of gene-networks. In this work, the two most popular tools, Cytoscape and Visant were used. These tools are designed on the Java platform, rendering those OS independent. Cytoscape is one of the most widely used piece of software for visualization of gene networks in the scientific community (Shannon et al., 2003). It is open source software which also provides additional plug-ins for different kinds of analysis such as hierarchical clustering, data mining for protein-protein interaction network, etc. Moreover, the software supports wide range of data structures coming from different platforms such as expression profiling data, Chip-seq data, etc. Finally, it can be used as a meta-base for different PPI databases. Despite its advantages, the major drawback of the software is that it is memory inefficient. Hence, it is unable to handle large amounts of data. To overcome this difficulty, other software like Visant can be used that can handle large amounts of data (Hu et al., 2013). The input format of Visant is XML and text format. The tools can color code the different interactions derived from different resources and algorithms, e.g. co-expression analysis, text mining, etc. It also offers various tools such as spring embedded relaxing, circular diagrams etc. for clear visualization of the interaction networks.

3.8 Network databases (STRING and IPA)

To detect physical and biochemical interactions among the interacting genes in the networks we refer to curated gene and protein interaction databases. As discussed before, there are many protein-protein interaction databases available. One of the most popular is STRING (von Mering et al., 2003). Its data is obtained from both in-silico and experimental set-ups. Accepted techniques in STRING databases include neighbourhood genes, gene fusion, co-expression, co-occurrence, and text mining. To the degree that it is known, the database also annotates the biological function for each interaction such as activation, inhibition, or catalysis, etc. Each interaction offers a confidence score to reflect the scientific evidence supporting it. Subsequently, interactions are colour coded for visualization. Additionally, the database can be used to provide gene ontology terms for a given network. The output can be exported as a text file to support further local investigations.

Another commonly used tool to study protein and gene interaction is the Ingenuity Pathway Assistant (IPA). With IPA the export of data to a text file is not supported, but the software provides additional information such as causality, classes of molecules, promoters, motifs etc. in a given network(Kramer et al., 2014) : It not only searches for direct interactions, but also provides evidence for indirect interactions. Furthermore, the software also offers to select the organisms from which the user can derive information (e.g. human and mouse) and overlap them in

their own networks. In analogy to the STRING companion STICH (Kuhn et al., 2014) and IPA can also suggest drugs for gene targets in a network. Users can design their own networks and can generate publication-quality images. Altogether, both STRING and IPA provide ample resources for verifying and discovering new mechanisms.

3.9 Prediction of ncRNA tools

Accurate prediction of function of non-coding RNA (ncRNA) remains a challenge for researchers, as their importance in various mechanisms leading to abnormal phenotypes has been revealed in the last decade. With advancements in sequencing technology, new classes and subclasses of non-coding RNA based on their functionality have been added to existing databases such Rfam (Burge et al., 2013). The prediction of novel sequences discovered by scientists largely depends upon querying against these databases by sequence similarity, e.g. with tools such as BLAST (Altschul et al., 1990). Non-coding RNA sequences that lack such similarity remain unclassified, which would require expensive and uncertain, and hence risky wet-lab experiments to determine their function. One can computationally speed up the process in a cost and time-effective manner by additionally integrating different properties such as secondary structure to the sequence similarity algorithms. To develop such software, a machine-learning method is used to predict a specific group of post-transcriptional non-coding RNA in eukaryotes which outperforms existing

methods (Gupta et al., 2014). A two-layer classification system was developed comprising a first layer as a binary classifier to discriminate post-transcriptional non-coding RNA from other classes of non-coding RNA and a multi-class classifier to further distinguish within the subclasses.

3.9.1 Dataset

The NONCODE database is the collection of all non-coding RNAs (except tRNAs and rRNAs) where all the sequences of non-coding RNAs are manually curated (more than 80% experimentally verified) (Liu et al., 2005). We obtained 2,040 sequences of ncRNA from the NONCODE database, including 268 sequences of RNase P, 14 sequences of RNase MRP, 1,443 sequences of snoRNA (1,430 + 13 scaRNA), 46 sequences of telomerase RNA, 14 sequences of Y RNA and 255 sequences of snRNA (Table 1). These sequences were used as both a dataset for the multi-class classifier and as a positive set for our binary classifier. The negative set was made up of sequences of tRNA, 5s ribosomal RNA and miRNA that were derived from the Rfam database (Burge et al., 2013). Since our classifier focuses on the eukaryotic system, our selections of miRNA-sequences were restricted to sequences from the species *Homo* sapiens, *Mus* musculus and *Drosophila* melanogaster. The redundancy of the sequences within a set of sequences was removed using CD-Hit (Li and Godzik, 2006) at a 0.9 threshold for the positive set and at 0.8 for the negative set of sequences. CD-Hit is a clustering program that was

originally designed to identify non-redundant sequences in the amino acid sequence database. It clusters the sequence database into subgroups, providing a representative longest nucleotide sequence for each group. While training machine-learning algorithm, redundant sequences in a dataset can lead to over-fitting of the classifier, hence resulting in biasness towards specific classes. Therefore, it is essential to train the classifier with non-redundant sequences for reduced false positives and optimized results. We used approximately $^{2}/_{3}$ of the sequences for training and $^{1}/_{3}$ for testing the classifiers. A detailed list of the sequences is provided online at http://www.ptrnapred.org/ and in Table 1.

Table 1 : Total number of test and training sequences.

Non coding RNA	Training sequences	Testing sequences
RNase P	178	90
RNase MRP	9	5
snoRNA + scaRNA	978 + 9	452 + 4
telomerase RNA	29	17
Y RNA	9	5
snRNA	170	85

3.9.2 Features for classification

Feature selection and training of classifier were performed using two sets of input parameters: The first set was based on the primary sequence and the second set considered the secondary structure which was predicted with RNAfold (Ding, 2006).

Primary sequence properties were mainly derived from dinucleotide properties employing DiProGB (Friedel et al., 2009), using a sliding window approach (window size: 2 nucleotides). Some of the properties in DiProGB are highly correlated to each other. The use of highly correlating features for classification would not only be redundant, but would overfit the classifier. In order to determine which of the features that we derived from DiProGB were correlating, we determined the Pearson correlation coefficients among all possible di-nucleotide properties. Two features were considered as strongly correlated when the absolute coefficient of the Pearson correlation was > 0.9. As an example, the di-nucleotide property 'stacking energy' was highly correlated with the property 'melting temperature'. Whenever two features were strongly correlated, one of them was randomly discarded. In our example, we used 'stacking energy' as an input feature and discarded 'melting temperature' from consideration as a feature for classification. A table of the selected di-nucleotide properties as well as their di-nucleotide values is provided in the supplement (Supplement Table 1).

The secondary structures of every sequence were calculated via RNAfold (Ding, 2006), RNAfold is an algorithm implemented in the Vienna RNA package which predicts secondary structures of nucleotide sequences using thermodynamic measures of nucleotides and base pairing probabilities which is often well conserved in evolution (Lorenz et al., 2011). The software produces various additional descriptions of nucleotide sequences such as minimum free energy, enthalpy etc.,

which are hidden at the primary sequence level. The parsing of features derived from RNAfold output was automated using self-developed Perl scripts.

52 different properties were derived from the secondary structure, e.g. the number of loops, the number of bulges, the number of hairpins or the frequency of nucleotides involved in substructures (Supplement Section 1).

Additionally, we included 32 triplet element properties employed by miPred, a triplet SVM for the classification of miRNA (Jiang et al., 2007). MiPred considers the middle nucleotide among the triplet elements, resulting in 32 (4 × 8) possible combinations, which are denoted as 'U(((', 'A((.', etc.

Altogether, ptRNApred uses 91 features for classification. A detailed description of the feature selection is provided in the supplement (Supplement Table 2 and Supplementary Section S1).

3.9.3 Classification system

For classifying different classes of post transcriptional non coding RNA, we compared the outcome of two different algorithms i.e. random forest and support vector machines. First, we employed a random forest (Breiman, 2001) as a sophisticated classification method. Random forests operate by constructing a multitude of classification and regression trees at training time and suggest the class supported most frequently by individual trees. Secondly, we employed support Vector Machines (SVM) (Berwick, 2002), which are supervised learning models with

associated learning algorithms that analyze data and recognize patterns, and used for classification and regression.

3.9.4 Implementation of support vector machines and random forest on the post transcriptional non coding RNA dataset

Support vector machines are supervised learning algorithms which are used for classification and regression analysis (Boser, 1992). The algorithms are used for pattern and data analysis (Byvatov and Schneider, 2003). To build a SVM classifier, the reference data was divided into a training set and testing set. In support vector machines, support vectors are the data points that lie nearest to margins or the decision surface. The decision surface that differentiates between the different classes with maximum margins is a hyper plane (Berwick, 2002). A SVM model postulates data points in the feature space so that differentiation between different classes can be attained with maximum margins. For nonlinear classification, SVMs can use kernel functions that implicitly map feature vectors into a higher-dimensional feature space. Kernels are functions that return inner products between all pairs of data points without computing its coordinates in high dimensional feature space. In this work, we used the standard radial basis function kernel, (RBF) from the LibSVM package to train the SVM model (Kaminski and Strumillo, 1997).

To optimize RBF kernel for training SVM model, users must specify some parameters. LibSVM provides a python script to optimize the grid parameters C and gamma (γ) (Chang CC, 2001,2007). C stands for cost function, i.e. penalty for the misclassification in the training set and γ defines influence of each training vector on the model.

To decide optimal choice of parameters for C and γ, LibSVM obtains cross-validation (CV) accuracy using a grid search (Chang CC, 2001,2007). Regarding our binary classifier, the highest CV accuracy was achieved when C was set to 32768 and γ was set to 0.008 (Figure 20a). These parameters were used to train the whole training set and generate the final model.

For multi-class classification, under a given (C, γ), LibSVM can use one-against-one, one against all and sparse methods to obtain the CV accuracy. Hence, the parameter selection tool suggests the same (C, γ) for all k (k-1)/2 decision functions. Chen et al. discuss issues of using the same or different parameters for the k(k-1)/2 two-class problems (Chen et al., 2008). We obtained best CV accuracy for our dataset using sparse method for the multi-class classification where C=4 and γ=0.5 (Figure 20b).

The random forests method as an ensemble learning method is a collection of classification and regression trees algorithms to acquire better predictive performance than a single tree (Breiman, 2001; Opitz D., 1999). In random forest, a decision tree is a decision making information system, which recursively splits cases

from controls with an improving accuracy, thus forming a tree structure. In random forest, multiple decision trees are constructed using bootstrap samples from the training data set. While 2/3 of the samples are randomly selected to construct a tree, 1/3 samples are left out which are called out of bag (OOB) data (Breiman L., 1996). The left out samples are put down the constructed tree to get a classification. The process is repeated for each tree and majority vote is returned as classification of the whole forest. The proportion of time that a class for a sample is not equal to the true assigned class is averaged over all samples and called an OOB error estimate (Breiman L., 1996b). The feature importance can be calculated by permutation where randomly shuffled values are assigned to each sample and thereby trees are constructed. This estimates the mean decrease in accuracy of that particular predictor variable which has been permuted. Thus, the importance of the predictor variables can be ranked by their mean decrease in accuracy. Another way of assessing variable importance is by the Gini-index criterion (Breiman L., 1984). When a split of a node is made on a variable, the Gini impurity criterion for the two descendent nodes is below the same of the parent node is calculated. The Gini importance for a variable is the sum for all splits based on that variable across all the trees. One of the most useful tools in random forest is proximities calculation. It is determined by examining the terminal node membership of the data (Chen and Ishwaran, 2012). If two cases occupy the same terminal node, then proximity is increased by one. At the end, the proximities are normalized by dividing by the

number of trees. In this work the R package 'Boruta' was used to train random forests for classification and to perform feature selection (Fortino et al., 2014; Miron B, 2010). The R package Boruta, is a feature selection algorithm wrapped around the random forest classification algorithm to find all relevant features. It iteratively removes the irrelevant features which are determined by a statistical test. The function 'Boruta' in the R package was used to construct 10,000 trees. Variable importance was determined using function 'getImpRf' where node impurity was measured by the Gini - index criterion. The other parameters were kept as default.

3.9.5 Work flow and output of ptRNApred

The web server implementation takes sequences in a FASTA-format, which can be either uploaded as a file or pasted into the text box. By checking "Post-Transcriptional RNA", an inbuilt Perl script calculates input vectors for the pre-trained model to predict whether or not the input sequence belongs to the group of post-transcriptional RNA. Additionally checking "RNA family", the server also predicts the RNA subclass.

Altogether, the output includes the prediction for ptRNA as well as the classification of the RNA class within the ptRNA (Gupta et al., 2014). Additionally, it displays the minimum free energy using RNAfold (Ding, 2006) as well as the secondary structure, using VARNA (Darty et al., 2009). The output can be downloaded directly.

4. Results

Combining mRNA expression data, genotype data and phenotype data from the experimentally crossed population of mouse have achieved success in the identification of the causal genes for various complex phenotypes. Despite its success, the understanding of systems biology for occurrence of such phenotypes still resides in its early stages, as many other classes of small RNA, such as miRNA, are involved in complex traits. Here, our study concentrates upon understanding complex traits by evaluating miRNA and mRNA expression levels using an integrative genomics approach. To apprehend such an inference, we crossed four mouse strains for three generations (G4 stage) and induced EBA. We collected skin tissue for the study. For statistical accuracy and generality in the population, the mice were selected randomly from the gene pool. We chose 100 samples for miRNA profiling and 200 samples for mRNA profiling.

4.1 MicroRNA expression profiling

4.1.1 Expression QTL mapping

A total of 100 mice were used from the fourth generation of the AIL to study the variability occurring in miRNA expression due to genetic heterogeneity. A full genome-wide scan was conducted to find genetic loci associated with miRNAs. Hence, here the miRNA expression levels were treated as a quantitative trait, i.e.

eQTL. At significant thresholds derived from the genome-wide permutation test ($\alpha < 0.05$), 42 eQTL for 38 miRNAs were mapped. This corresponds to 6.83% of a total amount of murine miRNA present in the Affymetrix chip (miRNA 2.0 Array) (Figure 1 and Table 2). The highest -log *p-value* of 6.57 corresponds to miR-298 on chromosome 9 at a confidence interval of 68-100 Mb, explaining 20.76% of the phenotypic variance. The peak SNP (rs3700596) associated within this genomic region was found near the Ube2cbp gene, an ubiquitin-conjugating enzyme E2C binding protein (~1 kb from peak SNP). In this study we could observe only trans-eQTL, i.e. regulate miRNAs that are regulated from other loci than their own transcriptional site. Only one miRNA miR-486 (-log *p-value* = 4.10, rs13479880) was mapped on the same chromosome of its transcriptional site (Chr 8, ~89 Mb), suggesting a possible cis-eQTL..

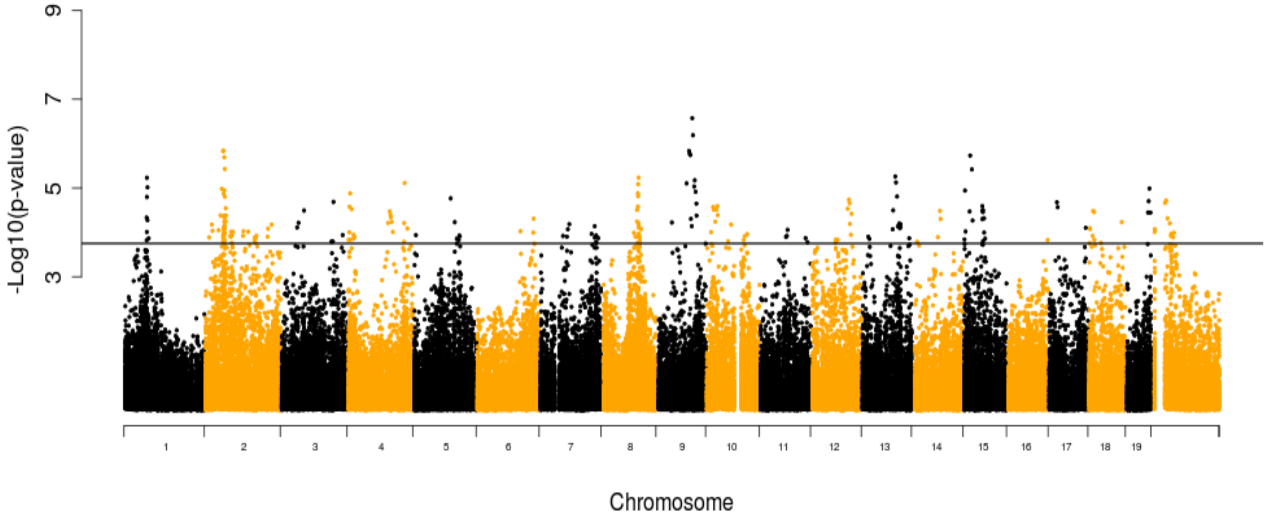


Figure 1: Manhattan plot showing the eQTL for miRNA. The black line represents the genome wide significant threshold ($\alpha < 0.05$).

Table 2 : Significant miRNA eQTL at genome wide significance (α < 0.05)

AffyID	PEAK SNP	Chr	Pos(Mb)	CI	-log *p-value*	α
miR-322_st	rs6386920	1	54.1525	51.4554-58.4393	4.797697778	0.02
miR-431_st	rs6386920	1	54.1525	51.4554-59.7799	5.231417376	0.006
miR-26a_st	rs6265423	2	47.02	28.1446-51.3211	4.376166299	0.021
miR-291a-3p_st	CEL.2_50605053	2	50.5005	28.1446-59.8562	4.546953451	0.018
miR-423-3p_st	rs6265423	2	47.02	33.85-51.3211	4.948513523	0.015
miR-671-5p_st	rs13476472	2	45.5486	33.85-50.5005	5.838570402	0.004
miR-23b_st	rs6250599	2	48.5289	44.2572-51.3211	4.335454904	0.022
miR-409-5p_st	rs13476874	2	159.495	135.954-162.978	4.175101842	0.035
miR-409-5p_st	rs13477083	3	43.3298	27.5001-45.4009	4.215972522	0.033
miR-546_st	rs13477126	3	56.4461	52.8217-58.2937	4.493016911	0.029
miR-200a-star_st	rs3671119	3	126.116	117.357-131.302	4.688008826	0.011
miR-339-5p_st	rs3660863	4	7.12733	3.64972-19.5083	4.878132165	0.003
miR-465c-5p_st	rs13477873	4	101.103	82.8343-118.065	4.472478928	0.01
miR-295_st	rs3663950	4	135.285	129.391-141.126	4.209508065	0.04
miR-878-3p_st	rs13478002	4	136.23	135.285-141.126	5.110152979	0.01
miR-742_st	rs3673049	5	90.1166	87.521-96.6196	4.765858909	0.035
miR-379_st	rs6208251	6	104.839	98.3634-116.707	4.024954772	0.034
miR-154_st	rs13479063	6	136.34	133.918-142.369	4.309450382	0.032
miR-425-star_st	rs3663988	7	146.505	140.19-146.505	4.342220512	0.035
miR-486_st	rs13479880	8	89.2709	72.486-95.0537	4.108040959	0.033
miR-487b_st	rs6257357	8	88.0714	77.6509-90.1831	4.81779528	0.019
miR-501-3p_st	rs13479880	8	89.2709	81.6926-95.0537	4.205948891	0.03
miR-130b_st	rs6413270	9	37.749	36.7545-44.4448	4.222559895	0.032
miR-298_st	rs3700596	9	86.1986	85.0493-90.5779	6.573678024	0.001
miR-466c-3p_st	rs3712394	10	17.6633	14.1231-24.2073	4.569565941	0.03
miR-466c-3p_st	rs13480563	10	27.8549	24.2073-38.6853	4.591526295	0.029
miR-126-5p_st	rs6374078	10	60.568	30.7258-65.9068	4.177297455	0.031
miR-681_st	rs13481076	11	66.5323	40.1175-71.2932	4.055742697	0.043
miR-20a-star_st	rs3712881	11	120.929	112.095-120.929	4.160612966	0.03
miR-203_st	mCV22351241	12	60.0416	55.0208-72.5617	3.838027866	0.039
miR-542-3p_st	CEL.12_84750094	12	91.3972	79.7044-103.767	4.735149963	0.023
miR-341_st	gnf13.079.671	13	80.5444	69.4679-88.2773	5.260091403	0.008
miR-449b_st	rs13482231	14	67.5971	50.848-72.2967	4.485232377	0.019
miR-7a_st	CEL.15_4222769	15	4.34915	3.22903-9.71766	4.941341377	0.044
miR-7a_st	rs13482455	15	16.6609	14.813-24.5234	5.73214998	0.022
miR-337-3p_st	rs13482549	15	45.4745	38.4141-53.8672	4.298224443	0.029
miR-673-5p_st	rs13482549	15	45.4745	38.9557-56.6379	4.592339382	0.014
miR-136_st	rs13482914	17	20.9827	16.5395-27.5883	4.679083445	0.026
miR-466b-5p_st	rs13483212	18	12.2691	0-20.9827	4.48515896	0.016
miR-493_st	rs6211533	19	57.0668	50.2034-60.1574	4.98913787	0.004
miR-26a_st	gnfX.023.543	X	36.2528	0-49.3187	4.718906745	0.01
miR-466e-5p_st	rs13483712	X	9.12938	0-33.548	4.013512497	0.05

The presence of only trans-eQTL indicates that other, possibly tissue-specific, genes contribute to the regulation of the miRNA expression. Previous studies suggest that genes encoding for various classes of argonaute and helicases play a major role in the regulation of miRNAs (Beitzinger and Meister, 2011; Chu and Rana, 2006). In line with this notion, the coordinates of all helicases and other genes involved in the biogenesis of miRNA pathways were obtained from the databases and mapped to the identified eQTL (Jankowsky et al., 2011). Helicases such as Ddx39, Ddx49, CD97 and Upf1 were mapped within the confidence interval of miR-486, miR-487b and miR-501 on chromosome 8. Furthermore, four helicases i.e., Ddx50, Ascc3, Ddx21 and DNA2, were mapped within a confidence interval of eQTL observed for miR-126. Several genes including Polr3f, Polr2a, Polr3g and Polr2a that were previously shown to play an important role in transcriptional machinery were also mapped to the eQTL for miR-409, miR-681, miR-34 and miR-449. Other genes such as lin28a and its homolog lin28b, which have been shown to modulate let-7a, were mapped within eQTL for miR-290 on chromosome 4 and miR-126 on chromosome 10, further inferring a possible expansion of lin28 to modulate other miRNAs.

Some miRNA eQTL were confined to specific locations in the genome, hence indicating eQTL hotspots. On chromosome 2, five miRNAs (miR-26a, mir-291a, miR-423, miR-671 and miR-23b) were mapped between 28-51 Mb (Figure 2). Three nearby SNPs (rs6250599 ~48.5 Mb, rs6265423 ~47 Mb and rs13476472 ~45Mb) showed a significant association (-log *p-value* > 4) with all five miRNAs mapped to

this region. A lincRNA 1700019E08Rik was located near SNP rs13476472 (~3Kb). As for another two SNPs, the nearest gene to rs6265423 was mapped 6.8kb apart, coding for the snRNA U7.39-201, while the nearest coding gene for SNP rs6250599 was a pseudogene Gm13489-001(~13kb upstream). We further investigated the five co-regulated miRNAs in chromosome 2 to identify the possible common molecular functions or pathways. The analysis suggested that these miRNAs are associated with mucin type o-glycan biosynthesis (*p-value* = 6.27e-11) and glycosphingolipid biosynthesis - lacto and neolacto series (*p-values* < 0.01).

Similarly, on chromosome 8 we identified eQTL for miRNAs (miR-486, miR487b and miR-501) within a confidence interval of 72-95 Mb. Two genes, Gm1068 and DNAJA2 were located near peak SNPs (rs13479880 and rs6257357). Pathway ontology analysis suggests that these three miRNAs are involved in the B-cell receptor pathway (*p-values* < 0.05) and endocytosis pathway (*p-values* < 0.05).

In our analysis, we found several miRNAs were regulated by more than one locus in the genome, i.e. multi-locus genetic loci. For example miR-7a showed significant association with two loci present on chromosome 15 (3-9 Mb, CEL.15_4222769 and 14-24 Mb, rs13482455). Similarly, miR-466-3c was regulated by two nearby loci on chromosome 10 (14-24 Mb, rs3712394 and 24-38, rs13480563). In contrast, miR-26a was mapped to two different loci on chromosomes 2 and X (28-51 Mb, rs6265423 and 0-49Mb, gnfX.023.543). Multi-locus control of miRNAs and regulation by trans-

eQTL infers complex machinery for the regulation of miRNAs. Therefore we investigated the epistatic control of miRNAs to derive other interacting loci that can possibly play a vital role in the regulation of miRNAs.

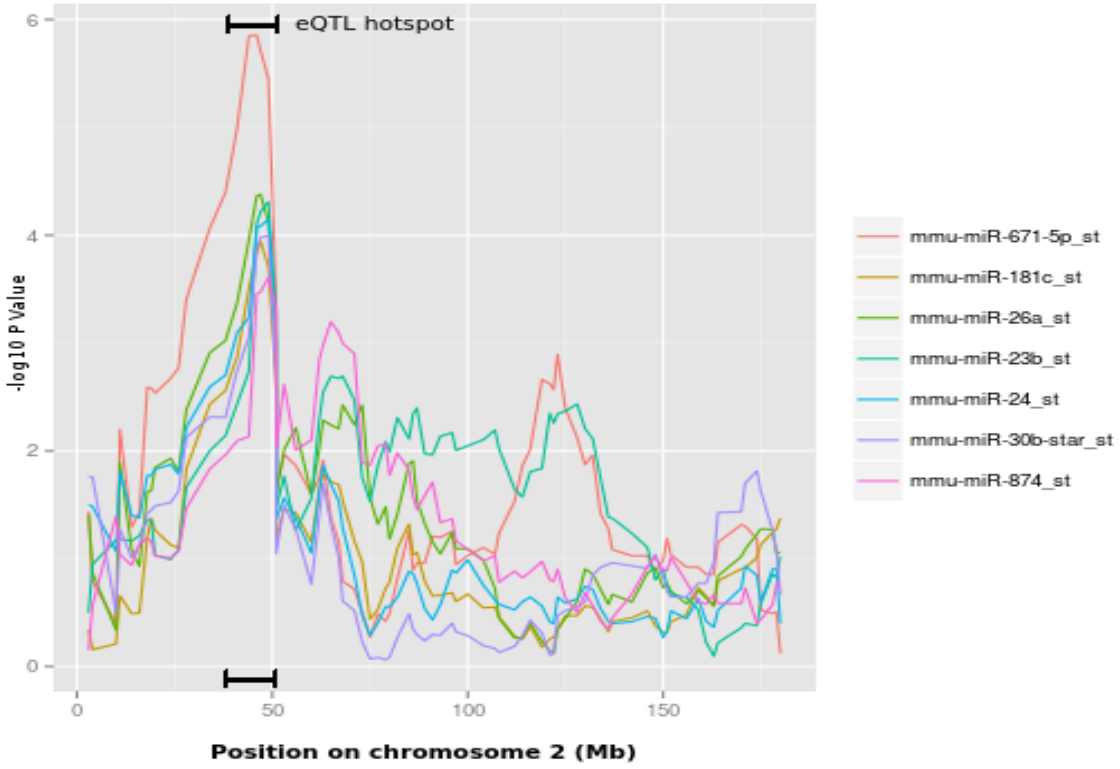


Figure 2: eQTL hot spots on chromosome 2. Position on x axis represents the coordinates in million base pairs on chromosome 2. The y axis is –log *p-value*. The overlapping peaks on y axis represents eQTL hot spot between 28-51 Mb. miR-181, miR-30b* and miR-874 are suggestive eQTL with significant genome-wide ($\alpha = 0.1$) threshold after permutation.

4.1.2 Epistasis in miRNA

Multi-locus control of miRNAs and regulation by trans-eQTL shows that miRNA expression is regulated by complex machinery. We investigated an epistatic control of miRNA expression, aiming to identify interacting loci that might play a role in the regulation of miRNAs. Therefore, we analysed the epistasis between each SNP pair for all the miRNAs for which we found a significant eQTL in the single locus association studies. As a result, we identified 200 SNP pairs for 8 miRNAs below the significance level (Bonferroni adjusted *p-value* < 0.05) (Figure 3).

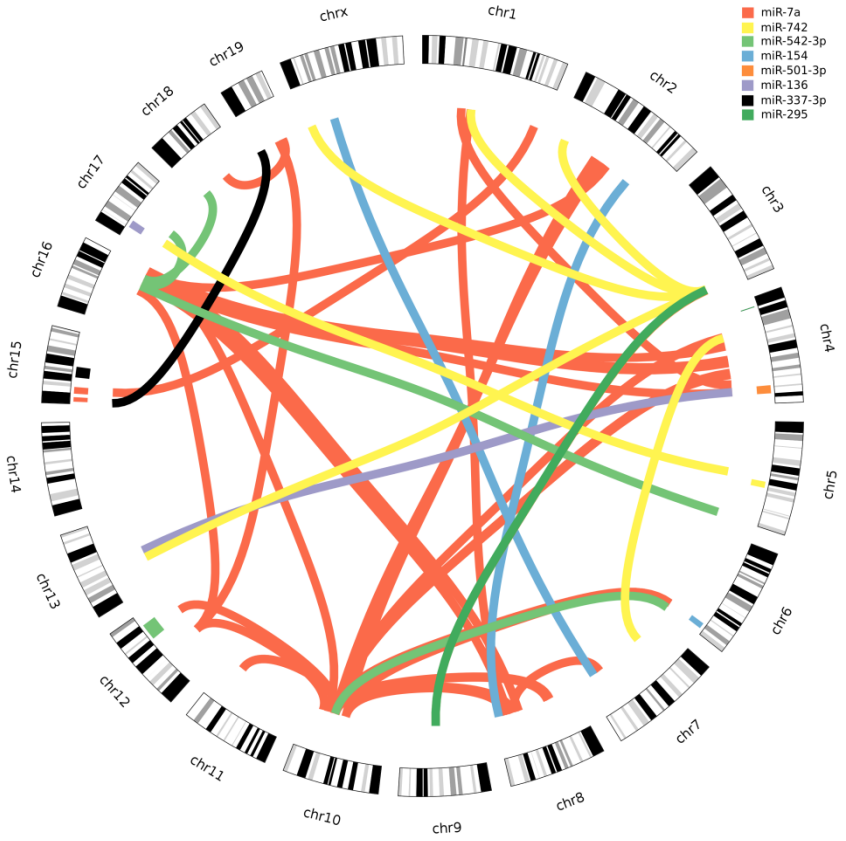


Figure 3: Epistasis in miRNA eQTL. The circular plot shows the chromosomes in its circumference. Each line between the chromosomes represents SNP pair interaction above the significance level (adjusted *p-value* < 0.05). Each interaction is color coded for different miRNAs. The boxes adjacent to the chromosomal band show eQTL for miRNAs mapped for single locus scans.

The highest -log *p-values* of 10.38 were found between the SNP pairs rs13480360 (Chr 10, ~67 Mb, nearest gene: AK139516) and rs3689658 (Chr 2, ~ 85 Mb, nearest gene: Olfr1006) for miR-7a (Table 3). In total, we found 119 SNP pairs for miRNA miR-7a. The hub locus (i.e. SNP with the maximal number of interactions) for miR-7a was observed on chromosome 16 (rs3680665 ~84 Mb, nearest gene: AK04263). The same SNP (rs3680665) also showed a high number of interactions (n=39) for

miR-542 and with additional SNP (rs4200124, nearest gene: Gbe1) present nearby. For miR-742, we observed 47 SNP pairs with SNP rs3657112 (~ 148 Mb, nearest gene: Snora17) on chromosome 3 showing the highest number of interactions (40 SNP pairs). The same SNP, i.e. rs3657112, also showed interactions with chromosome 9 loci 67-72 Mb for miR-295. We also found multiple SNPs on chromosome 2 (13-17 Mb) interacting with SNPs on chromosome 1(3-11 Mb) for miR-501. Two miRNAs (miR-136 and miR-337) had only one significant SNP pair: for miR-136 the interaction was found between SNPs rs37113033 (Chr 19, ~5Mb, nearest gene: Slc29a2) and rs13459176 (Chr 15, ~3 Mb, nearest gene: Sepp1), while miR-337 had SNP pair rs3693942 (Chr 13, ~55 Mb, nearest gene: Unc5A) and rs3663950 (chr4, ~135Mb, nearest gene: Il22ra) (Table 2).

Table 3 Epistasis in miRNA. The table includes only the interacting SNPs for highest interacting –log 10 *p-value* for each miRNAs.

miRNA	Top interacting SNP 1					Top interacting SNP 2					I. *p-value* (log)
	SNP ID	Chr	Pos(Mb)	Ref Genes	*p-value* (log)	SNP ID	Chr	pos(Mb)	Refseq Genes	*p-value* (log)	
miR-7a	rs13480630	10	67.29		0.89	rs3689658	2	85.52	Olfr1006	1.12	10.4
miR-136	rs3713033	19	5.03	Slc29a2	0.073	rs13459176	15	3.23	Ccdc152, Sepp1	1.56	7.2
miR-154	rs3708073	8	124.28	Gm20388,Jph3	0.3	rs8246404	2	136.7	Mkks	0.57	7.67
miR-295	rs13480271	9	72.68	RP23-461P14	0.018	rs3657112	3	148.03		0.32	8.064
miR-337-3p	rs3693942	13	55.05	Unc5a	0.026	rs3663950	4	135.29	Il22ra1	0.2	7.17
miR-501-3p	CZECH-2_15618849	2	15.5	Gm13364	1.33	rs3716083	1	9.01	Sntg1	0.72	8.46
miR-542-3p	rs4200124	16	70.7		1.12	rs3718776	5	150.4	Wdr95	0.14	8.5
miR-742	rs3657112	3	148.028		0.35	CEL.1_49993068	1	49.68		0.33	9.16

In order to verify interacting genes from our analysis, we investigated the epistatic control of miR-501-3p. Using Ingenuity Pathway Analysis (IPA), we searched for all the interacting genes for interacting loci present on chromosome 1 and 2 [89]. We found three interacting gene pairs: Commd3 with Cops5, Cacnb2 with Vopip1 and Commd3-Bmi1 with Rb1cc. Interestingly, Cops5 in the Nfkb1 pathway, possibly regulates miR-501 via Tp53 (Figure 4).

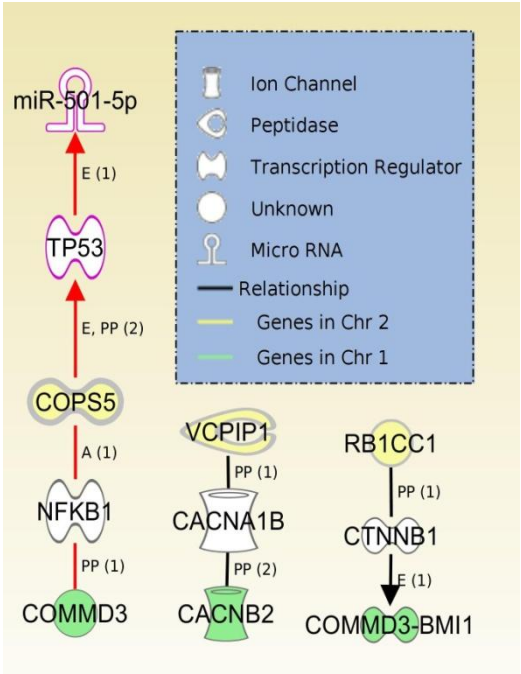


Figure 4 : Interaction network accessed via IPA software for epistasis of miR-501. The graph depicts the interacting genes identified from epistasis scan of miRNA miR-501 in chromosome 1 and chromosome 2. The graph shows all known gene interactions between the two loci .Genes colored in yellow are located on chromosome 2, while genes colored in green are encoded from chromosome 1. The red line shows the possible pathway for the regulation of miR-501.

4.1.3 Genetic overlap of ASBD QTL (EBA) and expression QTL (miRNAs)

A genetic variation leading to a clinical phenotype with an effect penetrating to a clinical phenotype by means of inflicting miRNA levels should map to the same chromosomal region. Such consensus QTL indicate clinical relevance over intrinsic differences of the parental strains. Previously, genetic loci for EBA an autoimmune skin blistering disease were studied using the same cohort of mice for AIL (Ludwig et al., 2012). We found 4 eQTL for miRNAs mapped to QTLs for EBA (Figure 6). The eQTL for miR130b (Chr 9: 36-44 Mb, -log *p-value* = 4.42), miR-542-3p (Chr 12: 79-103Mb, -log *p-value* = 4.73) and miR-449b (Chr 14: 50-72 Mb, -log *p-value* = 4.49) were mapped on the QTLs for disease onset on chromosome 9, 12 and 14. Additionally, we mapped eQTL for miR-493 (50-60 Mb, -log *p-value* = 4.98) to the QTLs for both disease severity and onset on chromosome 19. Therefore, all onset QTLs in the previous study were mapped on miRNA eQTL, while one miRNA eQTL mapped on the QTL for severity of disease, inferring a role of miRNAs in the induction of disease (Figure 5).

MiRNAs eQTL overlapping with QTL of clinical phenotypes have been described in various other autoimmune diseases. For example, miR-130b has been reported in plasma of rheumatoid arthritis patients in previous studies (Murata et al., 2013). Moreover, studies have shown over-expression of miR-542-3p in other autoimmune diseases such as multiple sclerosis. The over-expression of miRNAs, for example

miR-449b over-expression, leads to a reduction of NOTCH1 signalling in celiac patients (Capuano et al., 2011). In addition, in some cases of multiple sclerosis another miRNA, miR-493, was found to be significantly up-regulated (Paraboschi et al., 2011). The data indicate that miRNA eQTL which overlapped with ASBD QTL has been explained in different autoimmune disorders. Thereby, to gain more insight for the role of miRNA in autoimmune disease, we looked into expression levels of miRNAs.

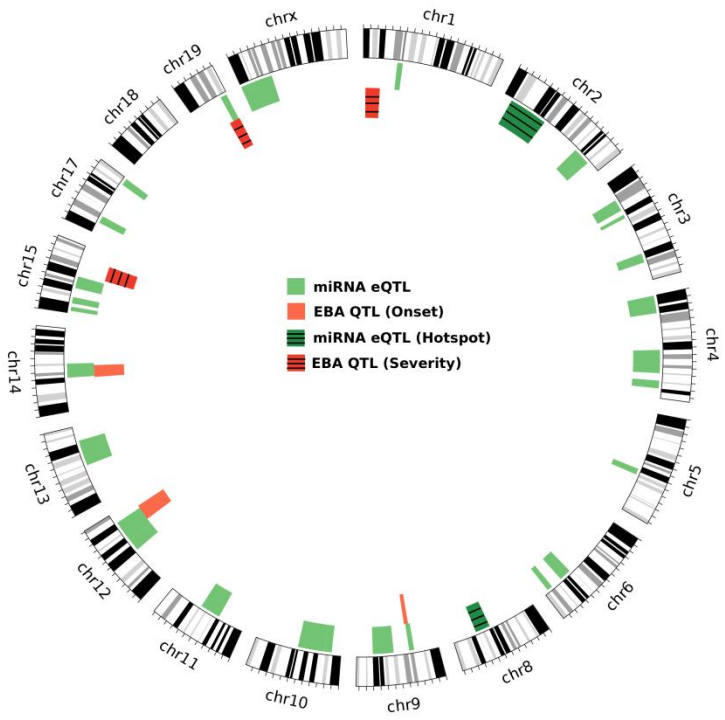


Figure 5 : Overlapping QTL for EBA disease and miRNA eQTL. The circular plot shows all the eQTL for miRNAs (green) and QTL for EBA (red). It also presents the eQTL hot spots (dark green) and EBA QTL for onset (red) and severity (dark red). Each circular band represents a chromosome on which QTL and eQTL are mapped. The region within the chromosome which has either red or dark red and green or dark green bands is overlapping eQTL with EBA QTL.

4.1.4 Expression and co-expression of miRNAs in ASBD (EBA)

To deduce differentially expressed miRNAs, we divided the G4 mouse cohort into two groups: EBA diseased and non diseased mice, irrespective of their severity score. We found 2 miRNAs, miR-379 (adjusted *p-value* = 0.044) and miR-223 (adjusted *p-*

value = 0.044) to be significantly differentially expressed between the two groups. Both miRNAs (miR-223 and miR-379) were up-regulated in the diseased mouse cohort. The up-regulation of both miRNAs has been previously described in the context of other autoimmune diseases. While miR-223 was shown to be over-expressed in peripheral $CD4^+$ T-lymphocytes from RA patients (Fulci et al., 2010), miR-379 has been highly up-regulated in splenocytes in lupus (Dai et al., 2013). Moreover, eQTL for miR-379 (-log *p-value* = 4.7, 98-116 Mb) was also mapped to chromosome 6 with a peak at ~104 Mb (rs6208251, nearest gene: Cntn6).

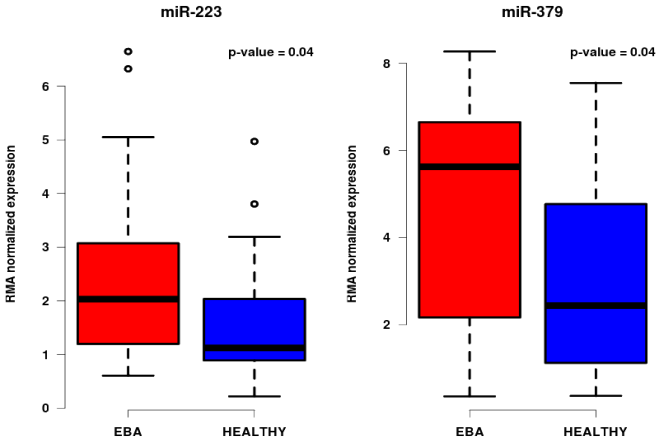


Figure 6 : Box-plots showing differentially expressed miRNAs. The box-plots show the most differentially expressed miRNAs miR-223 and miR-379 for the disease phenotype EBA. The plots in blue color show the expression of miRNAs in mice that did not have disease while mice with inflammation are shown in red color.

Dividing the cohort of mice into diseased and non-diseased has its drawbacks, as the severity of the disease observed in individual mice differs depending upon the genetic architecture present in them. Therefore, we performed a co-expression analysis, with expression levels of miRNAs (assuming the co-expressed miRNA are part of the same pathway) in correlation with quantitative scores for severity and onset week of the disease. For this purpose, we used the WGCNA approach which clusters miRNA into modules and associates them with different phenotypic scores, including EBA severity and onset. We identified 11 clusters in our analysis (Supplement Figure 1). Out of the 11 co-expressed modules, only the 'black' module, consisting of 23 miRNAs, was significantly associated with the severity of EBA (ρ= 0.28, *p-value* = 0.005). Additionally, it was also significant for the onset of EBA (ρ = 0.29, *p-value* = 0.005). Due to the fact that the 'black' module is marginally more strongly correlated to the onset of the disease than the maximum score, one can speculate that miRNAs from this module are involved in the onset rather than the severity of the disease. This corresponds to previous observations of overlapping miRNA eQTL with EBA onset QTL. The pathways associated with the 'black' module were, for example, the MAPK signalling pathway, cytokine-cytokine receptor pathway and focal adhesion pathway (*p-value* < 0.01). All these pathways have been previously described in different autoimmune disorders.

Individual correlation of miRNA expression levels with disease severity showed that 24 miRNAs were significant (*p-value* < 0.05) (Table 2), with miR-223 showing the

strongest association (ρ = 0.4, *p-value* = 4.93e-05). Another miRNA which was highly correlated with EBA was miR-21 (ρ = 0.36, *p-value* = 0.00036). An association of miR-21 with the disease phenotype has been previously demonstrated in multiple autoimmune disorders, such as type1 diabetes, multiple sclerosis, systemic lupus erythematosus, systemic sclerosis, and psoriasis (Kumarswamy et al., 2011). Furthermore, investigations of the co-expressed module showed that the miRNAs were also co-regulated, i.e. there were overlapping genetic loci among co-expressed miRNAs. We observed that miRNAs in a given locus were either clustered within the same module or showed a stronger inter-module membership for a specific module, even if they were assigned to different modules. As an example, miR-322 and miR-431, which were mapped to chromosome 1 (51-59 Mb), are clustered in the 'red' module. In eQTL hot spots such as chromosome 2 (28-51 Mb), miR-423-3p and miR-23b are clustered in the 'yellow' module. Even though other miRNAs in this locus were assigned to a different module, they also show significance for the 'yellow' module. Examples are given by miR-671-5p (*p-value* = 1.22e-12), miR-26a (*p-value* = 2.65e-19) and miR-291a-3p (*p-value* = 3.53e-02). In line with this observation, for the eQTL mapped to chromosome 8, two miRNAs (miR-501-3p and miR-486) were clustered with the 'brown' module while miR-487 was assigned to the 'red' module, but had a significant module membership with the brown module as well (*p-value* = 0.004). The results suggest that certain miRNAs might not only be co-expressed, but

also co-regulated by the same locus, further implying that genomic loci might be controlling the pathways of miRNAs.

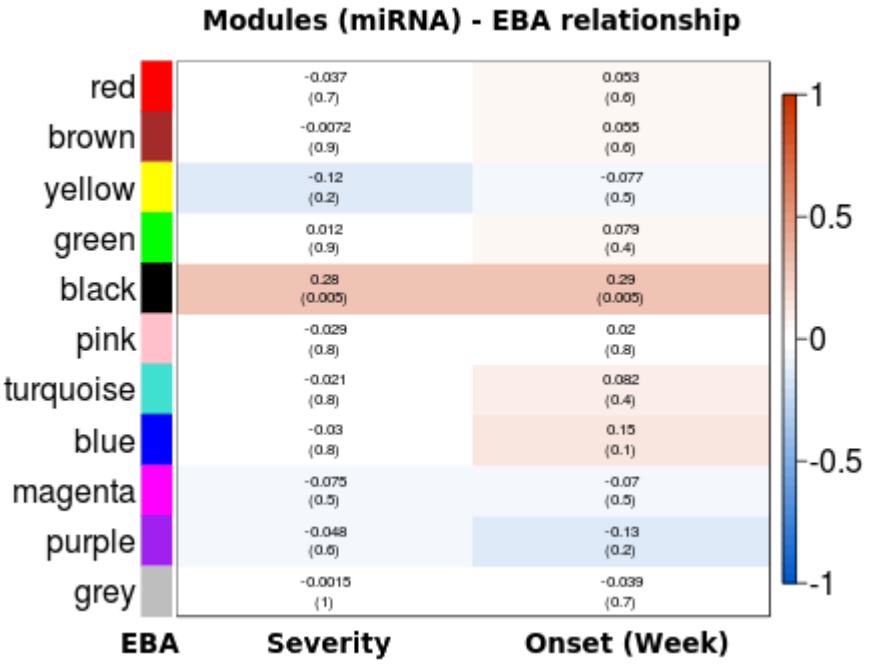


Figure 7 : Module-trait relationships between clusters of miRNAs with EBA severity and onset. The graph is a representation of co-expression analyses of miRNAs. The clusters of miRNAs are called modules; the colors are coded on the y-axis. The disease phenotypes (traits), EBA onset and EBA severity are represented on the x-axis. Blocks represent the correlation of modules with phenotypes using a Pearson correlation coefficient. The correlation range is color coded, with red indicating a positive correlation and blue showing negative correlation. '*p-values*' of the correlation coefficient are given in brackets below the correlation coefficients.

4.1.5 Causal miRNA network in skin

Recent advancements in the statistical models for defining gene networks from varying gene expression profiling have allowed understanding of putative pathways involved in various phenotypes. Therefore, to understand possible miRNA networks

in the skin, we employed the c3net algorithm, using their expression levels in the skin. The largest interconnected network consisted of 168 miRNAs (Figure 8). The hub node miRNA let-7c was connected to 21 other miRNA. In this network, 8 miRNA were significantly correlated with the onset of EBA, while 5 miRNA were associated with its severity. The pathway ontology terms for networks suggest an association with the Neurotrophin (*p-value* < 0.0001) and Notch signalling pathway (*p-value* = 0.0059).

The second largest subnetwork of miRNA consisted of 101 miRNAs (Figure 8). The subnetwork consists of 8 miRNAs for onset, 3 miRNAs for severity and 4 miRNAs significantly correlating with both phenotypes. The pathway associated with this network was the Fc epsilon RI signalling pathway (*p-value* = 0.0099). This subnetwork is important because it consists of miRNAs that are differentially expressed in diseased and non-diseased mice, such as miR-379 and miR-223. Fc receptors have been described to play an important role in skin diseases like EBA (Sesarman et al., 2008). Another pathway which was found to be significantly associated with the subnetwork was a regulation of the actin cytoskeleton pathway (*p-value* = 0.0468), further providing additional evidence for network authenticity, as expression levels are derived from skin tissue.

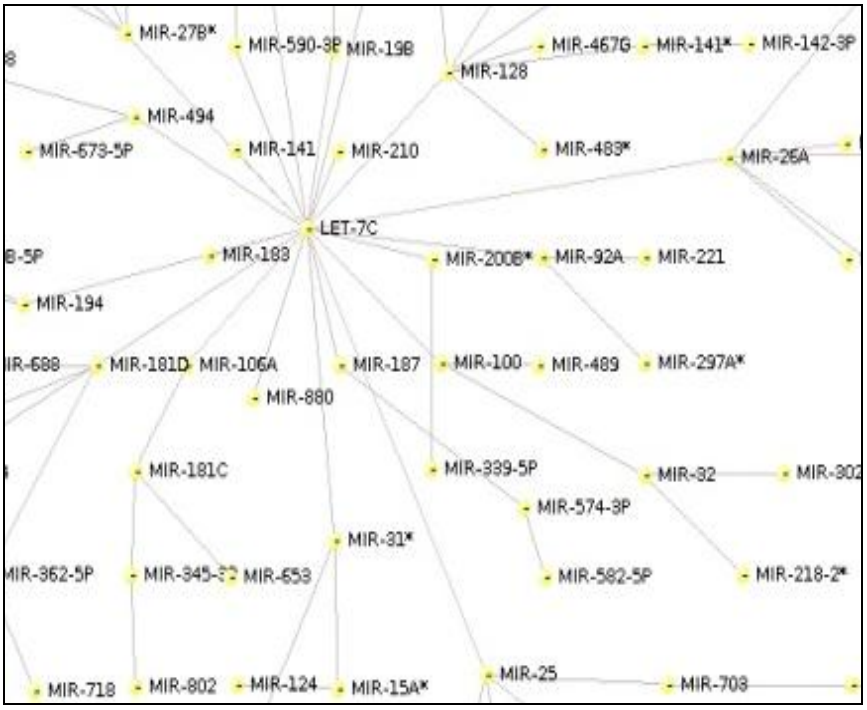


Figure 8 : Causal network for miRNA generated using c3net algorithm. The figure shows interactions of different miRNAs indicating the role of let7c as a hub.

4.2 Gene expression profiling

The Affymetrix chip consists of ~34000 probes coding for most of the genes across the genome. The probes were first checked for low expression in the skin tissue, leaving only 17,778 probes above median intensities. In addition, the probes were also checked for their variance (p-*value* < 0.05) across samples, thereby leaving 15,179 probes. The intersection of both expression and variance-based filtration, i.e. 7,043 probes, were used in further analysis. Amongst them, almost 2,000 probes were used as controls in the chip, i.e. for technical reason in the Affymetrix chip,

leaving only 5,043 probes consisting of 4,073 genes. The tools and software for the filtration of probes are described in the Material and method (section 3.4).

4.2.1 Differentially expressed and co-expressed genes in EBA

In order to find differentially expressed genes, we divided the samples into two sets, non-EBA samples (max score = 0) and EBA samples (max score > 0). Out of the 200 samples we used 190 samples (122 controls, 68 EBA mice) for performing differential expression analysis, as phenotype information was not available for the other 10 samples. We found 1,039 mRNA probes out of 5,054 probes to be significantly (adjusted *p-value* < 0.05, Bonferroni corrected) differentially expressed for the EBA and non-EBA mice samples. 425 probes were down-regulated and 613 probes were up-regulated for disease (EBA) phenotype (Table 4 and Supplement Table 3).

Table 4: Top 20 differentially expressed gene between EBA and non EBA mice. ID stands for Affymetrix ID, Gene name stands for official gene symbol, positive log FC show up-regulation and negative log FC shows down-regulation. The adjusted *p-values* are *p-value* after correction for multiple testing using Bonferroni correction method.

ID	Gene Name	log FC	*p-value*	*Adjusted p-value*
10380419	Col1a1	1.092457892	1.71E-11	2.97E-07
10529457	Cpz	0.476105771	1.83E-11	2.97E-07
10595211	Col12a1	0.937846131	3.26E-11	2.97E-07
10346015	Col3a1	0.986338829	3.42E-11	2.97E-07
10531724	Plac8	1.014783794	1.21E-10	8.43E-07
10583056	Mmp12	0.831790171	2.46E-10	1.20E-06
10536220	Col1a2	0.933247962	2.64E-10	1.20E-06
10460782	Gpha2	-0.588016109	2.77E-10	1.20E-06
10560919	Atp1a3	0.485810749	3.51E-10	1.36E-06
10560685	Bcl3	0.577733423	7.41E-10	2.58E-06
10354309	Col5a2	0.496367606	1.15E-09	3.39E-06
10546450	Adamts9	0.501134973	1.17E-09	3.39E-06
10556082	Ppfibp2	-0.28734775	1.29E-09	3.44E-06
10379636	Slfn4	1.289214032	1.60E-09	3.96E-06
10572949	Nr3c2	-0.319316147	1.90E-09	4.41E-06
10367400	Mmp19	0.798947908	2.12E-09	4.62E-06
10352143	Kif26b	0.383798765	2.39E-09	4.75E-06
10557895	Itgax	0.490973272	2.46E-09	4.75E-06
10403743	Inhba	0.618346825	3.67E-09	6.63E-06

Gene ontology for the up-regulated genes suggested that these genes are involved in pathways such as the chemokine signalling pathway (*p-value* = 1.1e-6), leukocyte trans-endothelial migration (*p-value* = 3.3e-6), Fc gamma R-mediated phagocytosis (*p-value* = 3.3e-6) and cytokine-cytokine receptor interaction (*p-value* = 7.2e-6). A more detailed list of pathways is given in Table 5.

Table 5: Pathways ontology for the up-regulated genes. P-values are calculated using fisher exact test and adjusted using Bonferroni's correction.

KEGG PATHWAY	*p-value*	adjusted *p-value*
Chemokine signaling pathway	8.6e-9	1.1e-6
Leukocyte trans-endothelial migration	5.1e-8	3.3e-6
Fc gamma R-mediated phagocytosis	7.6e-8	3.3e-6
Cytokine-cytokine receptor interaction	2.2e-7	7.2e-6
ECM-receptor interaction	1.6e-6	4.2e-5
Jak-STAT signaling pathway	2.5e-6	5.6e-5
Focal adhesion	3.6e-5	6.7e-4
Natural killer cell mediated cytotoxicity	3.7e-5	6.0e-4
Regulation of actin cytoskeleton	6.2e-3	8.7e-2
Pathways in cancer	8.7e-2	7.0e-1

Next, we constructed gene co-expression network of differentially expressed genes. Next, we constructed a gene co-expression network of differentially expressed genes. We applied hierarchical clustering to cluster gene profiles and a dynamic cut-tree method (WGCNA) to cut and combine the branches to further define a co-expressed group (modules) from 1,039 gene expression levels. The method is unbiased and does not use any prior biological information about the genes. We found 12 modules and assigned them into different colours with grey for the genes that could not be clustered in any of the other modules (Figure 9).

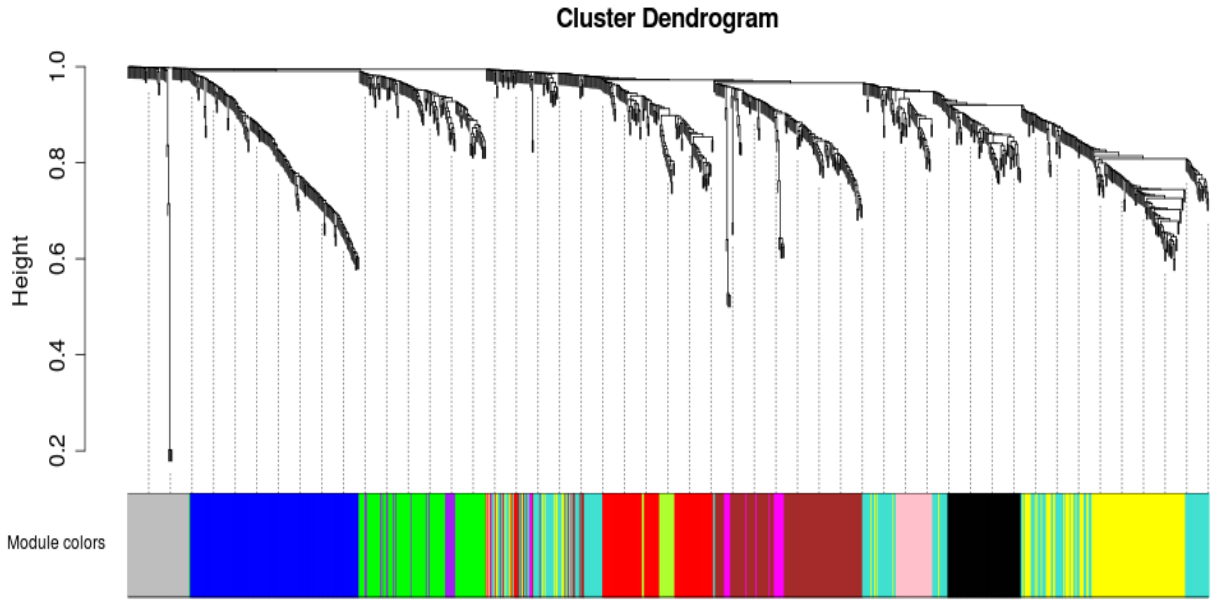


Figure 9: The figure shows dendrogram of different mRNA clustered together. Modules names are derived from their colors.

The module expression levels were abstracted by standardizing gene expression profiles to the first principal component, defined as the 'module eigengene'. The module eigengene can define a weighted average expression of the genes within a given module. To find disease-related modules, we correlated module eigenvalues to the disease score for EBA severity and onset (week) (Figure 10).

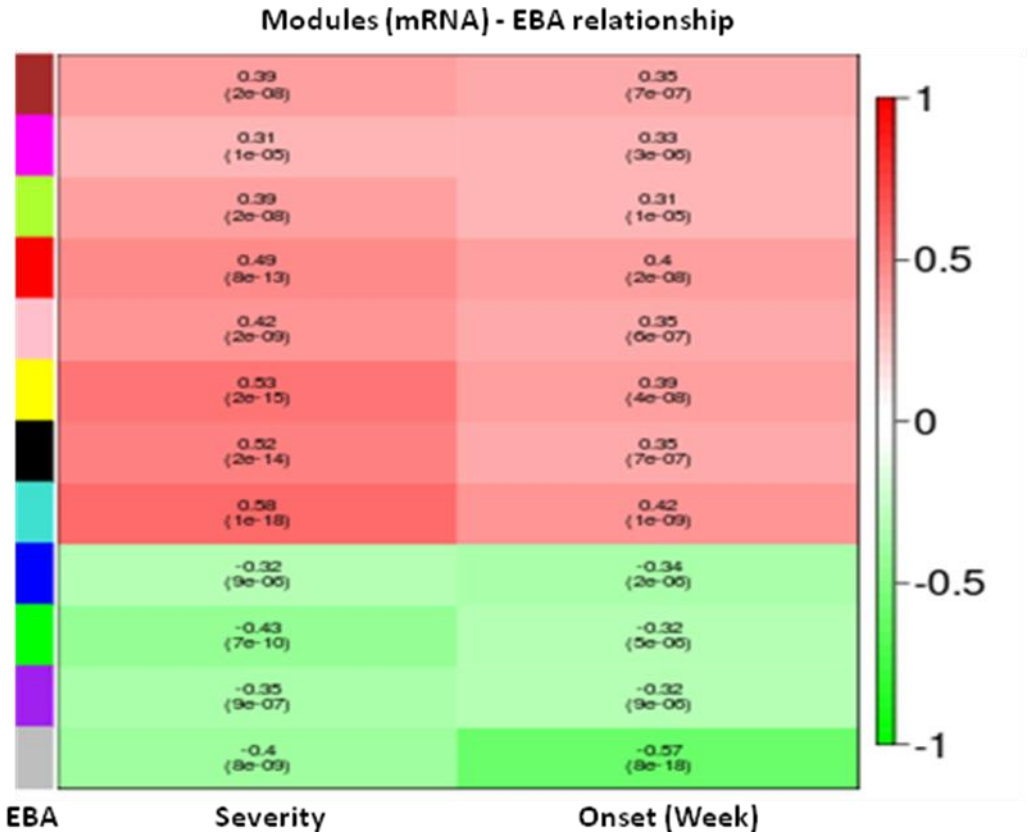


Figure 10 : The figure shows different modules by colors on y axis (vertically left) and its correlation with EBA max score (left) and week of EBA onset (right) on x axis (bottom). The right vertical bar displays the color range for Pearson correlation coefficient ranging from -1 to 1.

We found that 8 modules were significantly positively correlated, while 4 were significantly negatively correlated with EBA disease severity scores (*p-value* < 0.05). The 'turquoise' module was found to be the most significant amongst them all (ρ = 0.58, *p-value* = 1e-18), containing 178 genes. Gene ontology terms associated with this module were immune response (*p-value* = 1.8e-10), cell activation (*p-value* = 1.2e-8), response to wounding (*p-value* = 2.1e-8), leukocyte activation (*p-value* =

2.2e-7), and inflammatory response (*p-value* = 2.2e-6). Moreover the module also contains some important pathways such as inflammatory response (*p-value* = 1.1e-5) and cytokine-cytokine receptor interaction (*p-value* = 2.5e-5).

The second module which showed a significant positive relationship with EBA disease was the 'black' module. The correlation coefficient of this module with the disease was 0.52 with *p-value* = 2e-14. The module corresponds to biological functions similar to the turquoise module with cell activation (*p-value* = 1.59e-08) and immune response (*p-value* = 1.78e-08). Similar to the 'turquoise' module, the 'black' module also shows that genes in this module are involved in the cytokine-cytokine receptor interaction pathway (*p-value* = 7.44e-07).

The 'green yellow' module produced by the clustering method consisted of 16 Affymetrix probes coding for Hmnc1 genes verifying our clustering threshold.

We found 3 down-regulated modules in our dataset, with grey being the group of un-clustered genes. The cluster with the smallest *p-value* for negative correlation was the 'green' module (ρ = -0.35, *p-value* = 7e-10). The GO terms predict that the genes clustered in this module are associated with the cellular process (*p-value* = 6.62e-05) and metabolic process (*p-value* = 9.89e-05). This was expected, as the up-regulation of disease would result in the degradation of cell growth and other metabolic functions in cells. Other modules, such as the 'purple' module also showed a significant negative correlation with EBA (ρ = -0.32, *p-value* = 9e-07). This module is

associated with biological processes, such as the steroid hormone-mediated signalling pathway (*p-value* = 1.68e-05) and hormone-mediated signalling pathway (*p-value* = 6.56e-05).

The 'blue' module (ρ = -0.32, *p-value* = 9e-05) included 165 probes and was found to be associated with muscle structure development (*p-value* = 2.55e-14). Since EBA is a blistering disease, the down-regulation of the myosin binds ATP pathway (*p-value* = 4.27e-12) might lead to degradation of the skin.

4.2.2 Expression QTL mapping

To find a gene's eQTL we picked thresholds (α = 0.05, *p-value* (genome-wide) < 0.05) across the genome. To further remove false positives, an additional threshold cut-off was set (α= 0.01, *p-value* (point-wise) < 0.01). Therefore, -log *p-value* cut-off > 4.36 (Figure 11). We found 424 eQTL for 260 mRNA expression levels (Affymetrix probes) associated with 251 genes from 1,039 differentially expressed genes (Table 6 and Supplement Table 4). The highest –log *p-value* of 9.8 was observed for Scgb2b2 on chromosome 7, which was cis-regulated with phenotypic variance of 21.92. We found that 83/260 genes were cis-regulated while 177/260 were trans-regulated.

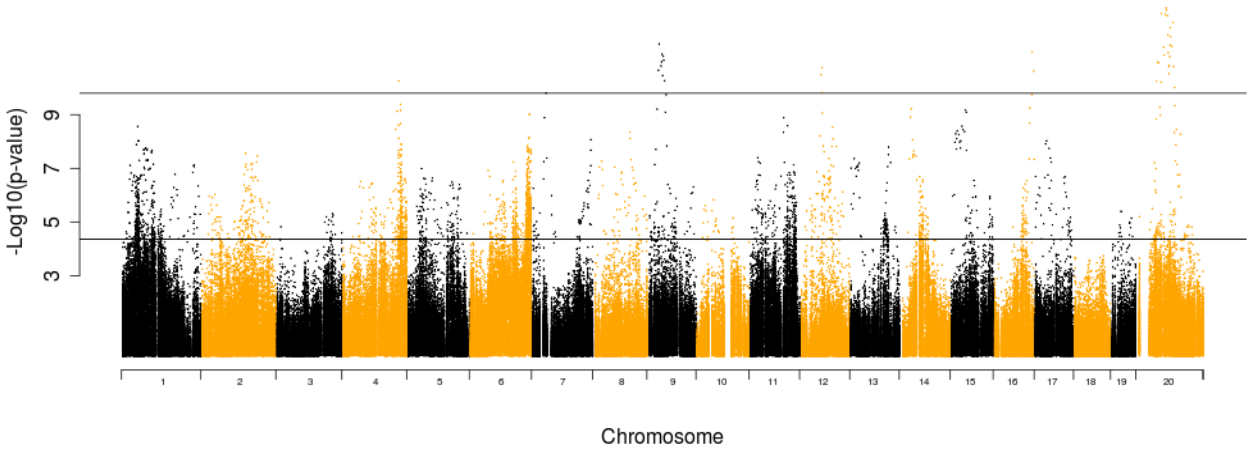


Figure 11 : Manhattan plot for 424 mRNA. The black line across the plot shows significant threshold of –log *p-value* (4.36) and upper threshold –log *p-value* (9.8).

Table 6: eQTL with -log *p-value* > 8.5. The table shows most significant eQTL. In table Probe ID is Affymetrix probe ID, Peak SNP is SNP with highest –log ***p-value***, and CI is confidence interval for eQTL.

Probe ID	Peak SNP	Chr	Peak (SNP)	-log10(*p-value*)	C.I.	Length QTL (Mb)	Gene Name
10552090	rs3719311	7	35	9.8	0-60	21.92	Scgb2b2
10592336	rs13480173	9	46	9.75	18-61	21.81	Spa17
10510482	rs3688566	4	141	9.39	135-156	21.14	Clstn1
10423471	CEL.15_36490596	15	36	9.18	3-61	20.74	Ctnnd2
10517250	UT_4_132.137715	4	133	9.13	128-135	20.64	Extl1
10528159	rs3658783	6	144	9.01	142-149	20.43	Gm10482
10379646	rs13481127	11	83	8.89	71-109	20.19	Slfn3
10386495	rs3688566	4	141	8.67	135-156	19.78	Tom1l2
10587880	rs13478002	4	136	8.61	135-142	19.66	Pcolce2
10392207	rs13481161	11	92	8.59	88-110	19.62	Tex2
10528159	rs4222295	1	39	8.56	13-79	19.56	Gm10482
10400510	gnf12.077.067	12	80	8.54	47-108	19.51	Clec14a
10605328	gnfX.084.751	X	98	8.45	10-135	19.35	Fam3a /// Fam3a
10572693	rs13479880	8	89	8.34	48-105	19.14	Jak3 /// Insl3

Similar to the miRNA protocol, we found eQTL hot spots for mRNA expression levels. We found eQTL hot spots on chromosome 6, 1, 4 and 11.

On chromosome 6, the eQTL hot spot was found to be between 134-149 Mb. The locus was found to control 85 eQTL for 72 genes, in which 84/85 were trans-regulated and 1 eQTL for gene Akr1b8 was cis-regulated. Previous reports suggest that experimental autoimmune myocarditis is regulated by chromosome 6 from 130-149 Mb (Guler et al., 2005). The overlap of the eQTL hot spot with a similar autoimmune disease suggests that the hot spot could be regulating multiple autoimmune diseases. The highest -log *p-value* 7.64 was observed for gene Krt7 with a phenotypic variance of 17.76, which was trans-regulated. The gene ontology terms indicate that genes regulated by loci are involved in the actin filament-based process (*p-value* = 1.9e-1) and regulation of actin cytoskeleton pathways (*p-value* = 6.0e-1). It was also observed that 61 genes were clustered in the 'brown' module, while 10 were clustered in the 'magenta' module in co-expression analysis.

On chromosome 1, the eQTL hot spot was found between 23-69 Mb. The locus was observed to regulate 42 eQTL for 39 genes in which 41/42 was trans-regulated and 1 eQTL for gene Aox4 was cis-regulated. The highest -log *p-value* of 8.56 was observed for Gm14082 with phenotypic variance of 19.56, which was trans-regulated. The

gene ontology terms suggest that the associated genes are involved in leukocyte activation (*p-value* = 4.5e-1) and cell activation (*p-value* = 3.7e-1). 32 genes were clustered in the 'brown' module and 5 in the 'turquoise' module.

On chromosome 4, the eQTL hot spot was mapped between 128-156 Mb. The loci was found to be regulating 57 eQTL for 52 genes; 55 eQTL were trans-regulated and 2 eQTL for genes Extl1 and Clstn1 were cis-regulated. The highest –log *p-value* (9.39) in this hot spot was observed for cis-regulated Clstn1 gene with a phenotypic variance of 21.14. The gene ontology terms show that these genes are associated with processes such as glucose metabolic process (*p-value* = 8e-1) and glycolysis/ gluconeogenesis pathway (*p-value* = 4.4e-1). 21 of these genes were clustered in the 'green' module, 15 in the 'blue' module and 11 in the 'purple' module. The data suggest that the hot spot could possibly be the master regulator for the down-regulation of genes.

On chromosome 11, the eQTL hot spot was detected at 87-103 Mb. The loci was found to regulate 21 eQTL for 18 genes, in which 20/21 were trans-regulated and 1 eQTL for gene Tex2 was cis-eQTL. The highest –log *p-value* 6.75 was observed for the trans-regulated gene Thrap3 with a phenotypic variance of 19.62. The gene ontology terms suggest that the genes regulated by this locus are involved in molecular functions such as ion binding (*p-value* = 1.1e-1) and calcium ion binding (*p-value* = 1.7e-1). We observed 18 genes that were clustered in the 'brown' module.

Although eQTL studies provide a broad understanding of the genetic regulation of different genes, since array-based technology is prone to detecting false positive signals, the outcome cannot be trusted. Therefore, identification of accurate eQTL analysis needs additional justification. In our study we found that gene Stat3, a well-known gene for immunity, is regulated by chromosome 1 (33–44 Mb), chromosome 6 (104-116 Mb) and chromosome 6 (136-147 Mb). We found all three loci controlling Stat3 had 'Stat3 transcription binding sites', thus appropriately confirming the reliability of this study.

We observed that genes regulated by the same eQTL hot spot regions were also co-expressed. Therefore, we took all the SNPs that regulate at least 5 genes and at –log P score >= 3.75 ($\alpha = 0.1$) and redefined the hot spots (Figure 12).

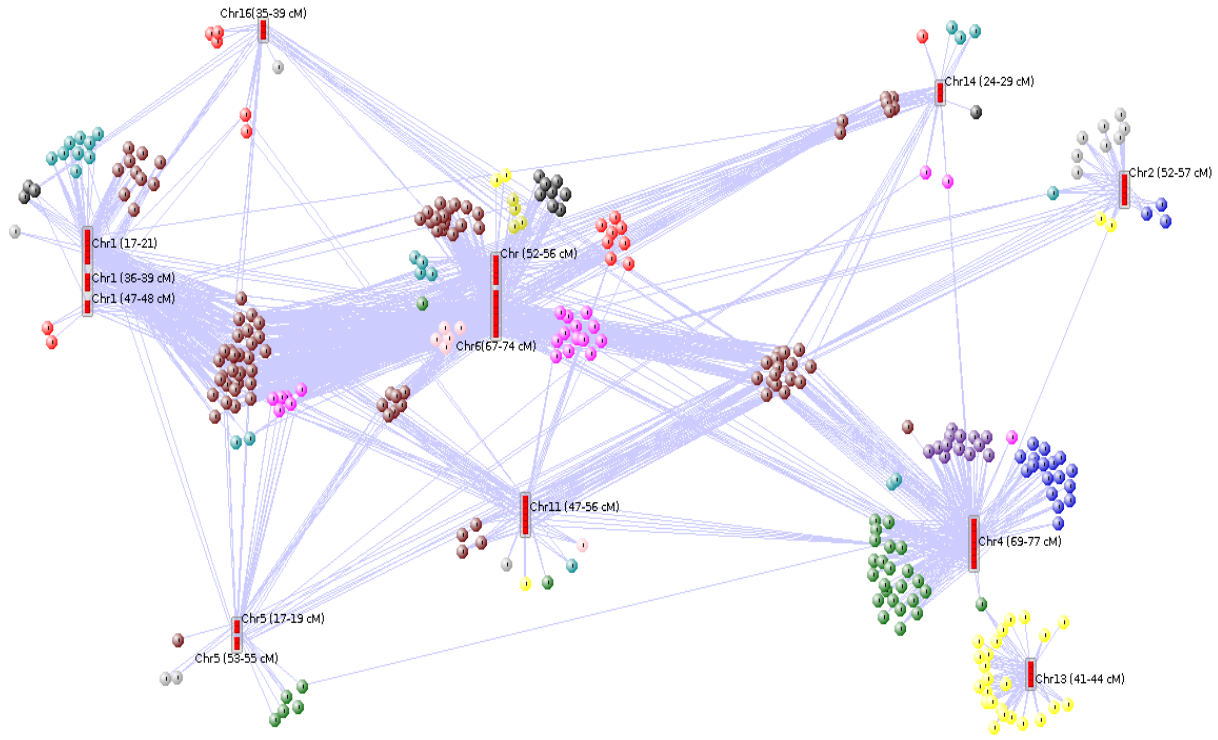


Figure 12 : eQTL hot spots at –log P score >= 3.75. The red blocks represent the hot spots in chromosome and circle represents genes. The colors display the associated module.

The analysis suggested that the co-expressed modules were controlled by specific loci. For example, in chromosome 13, an eQTL hot spot was detected at a confidence interval of 41-44 cM. The eQTL hot spot controls 28 probes (14 genes) for which gene ontology term suggests its involvement in immune system process (*p-value* = 5.5e-1). Most of the genes controlled by this locus are found in a co-expressed group ('yellow' module). When these genes were queried to the STRING database for known interaction data, we found that most of these genes are connected to each other

(Figure 13). This infers that genes that are co-expressed could be co-regulated by the same genetic locus. For this reason, we included the genetic data to find protein interactions among the genes in the co-expressed group.

.

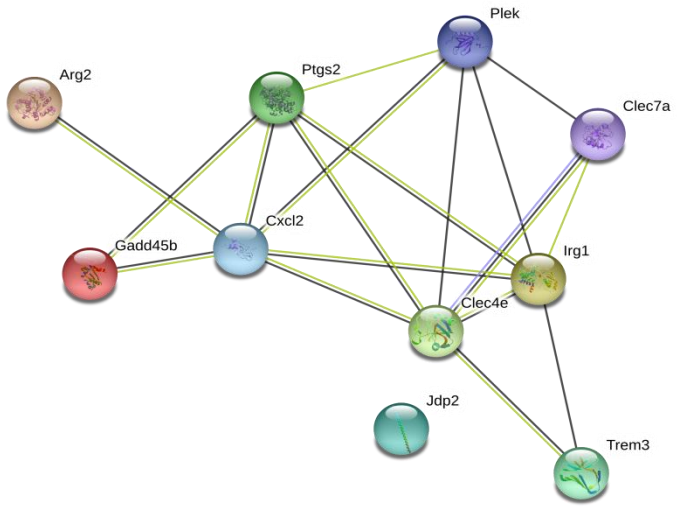


Figure 13 : STRING layout for eQTL hot spot. Layout depicts interaction between the genes that were controlled by eQTL hot spot in chromosome 13 in the yellow module.

4.2.3 Combining protein interaction networks with eQTL

To generate gene interaction networks, we took all the genes which were present within the co-expressed groups (modules). We combined the genetic data into the module by including eQTL information, considering two genes to be interacting only if they are controlled by the same SNP or other SNPs which are 10cM apart from each other. In this case, we considered all the SNPs that correspond to –log *p-value*

>3.75 (log point $P < 0.001$ and $\alpha = 0.1$ genome-wide). Adding SNP information to the co-expressed module decreases the false positive edges in a given module.

Furthermore, to determine the known interactions, we used known interactions from the STRING database. To identify new interactions which are not yet reported in the database, we used the DOMINE database for interacting protein-domains. Additionally, we used the PLS algorithm to determine statistically significant interactions. The details are provided in Material and methods (sections 3.6.2 and 3.8)

Black module

In the black module, 29 co-expressed-genetic interactions were confirmed by different methods. From 29 interactions, 11 interactions were found from the database and 19 interactions were predicted using domain interaction information (Figure 14). The interaction between Rab31- Arpc1b was confirmed by both methods. Based on the degree of connectivity (edges >= 5) we found two hub gene candidates, i.e. Mlkl and controlling the network. Additionally, using the domain interaction approach, we found Mlkl and Arhgap25 as hub genes controlling the major part of the network. Mlkl's role in multiple autoimmune diseases has been previously described (Zhao et al., 2012). As for Arhgap25, the gene showed 5 interactions. The interacting genes are associated with skeletal muscles and have

been shown to be important in various diseases (Katoh, 2004). The genes in this module are regulated by locus in chromosomes 1 and 6.

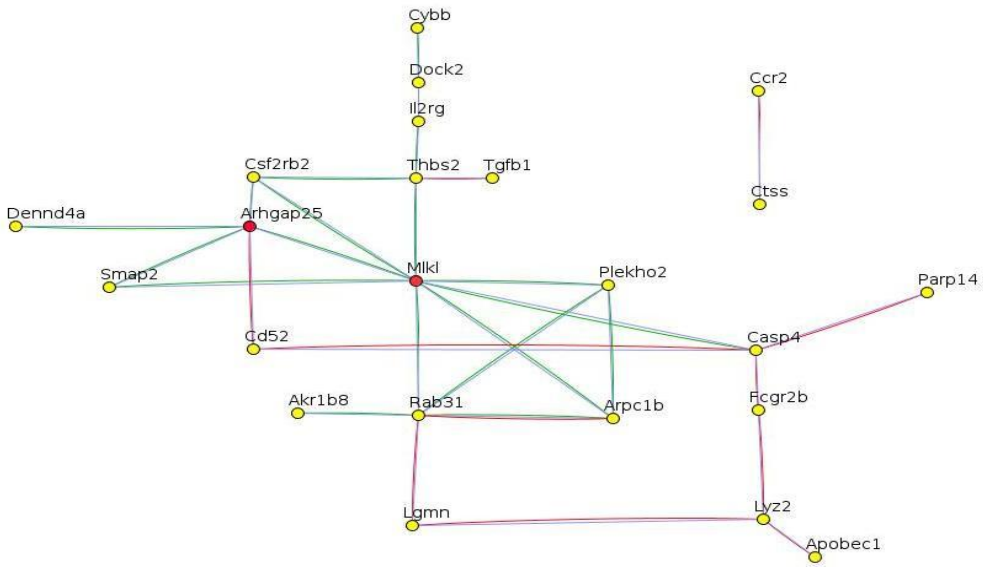


Figure 14 : Network of genes in black module. The red lines display the interaction information from STRING database. The blue lines display predicted interaction using domain interaction information. Red circle defines hub genes.

Turquoise module

In this module, 31 interactions were verified by database and domain interaction. Out of 31 interactions, 24 interactions were validated by domain interaction and 11 were found in the database (Figure 15). 4 interactions, Ralb-Nfkb2, Nfkb2-Nfkbiz, Nfkb2-Cebpb and Cebpb-Ptk2b, were confirmed by both methods. In this module, Ptk2b was suggestive as a hub gene with 10 interactions with other genes. The gene is involved in the MAP kinase pathway (Lev et al., 1995). The gene is up-regulated

in disease and connected to Nfkb2, which is another suggestive hub gene (edges = 7). NfKb2 plays an important role in immune deficiency diseases (O'Sullivan et al., 2007). Similarly, we also found the Nfkbiz gene as a possible hub candidate (edges = 6), which is also suggested to be associated with multiple autoimmune diseases (Okuma et al., 2013). Another suggestive hub gene candidate was Pard3b (edges = 5). The gene was down regulated in the module with respect to EBA (ρ = -0.26, *p-value* = 0.0002). Down-regulation of this gene leads to down-regulation of the actin cytoskeleton pathway (Lucas et al., 2013). This gene has been described as crucial in the context of autoimmune diseases (Below et al., 2011). The co-expressed genes were regulated by two loci on chromosomes 1 and 6.

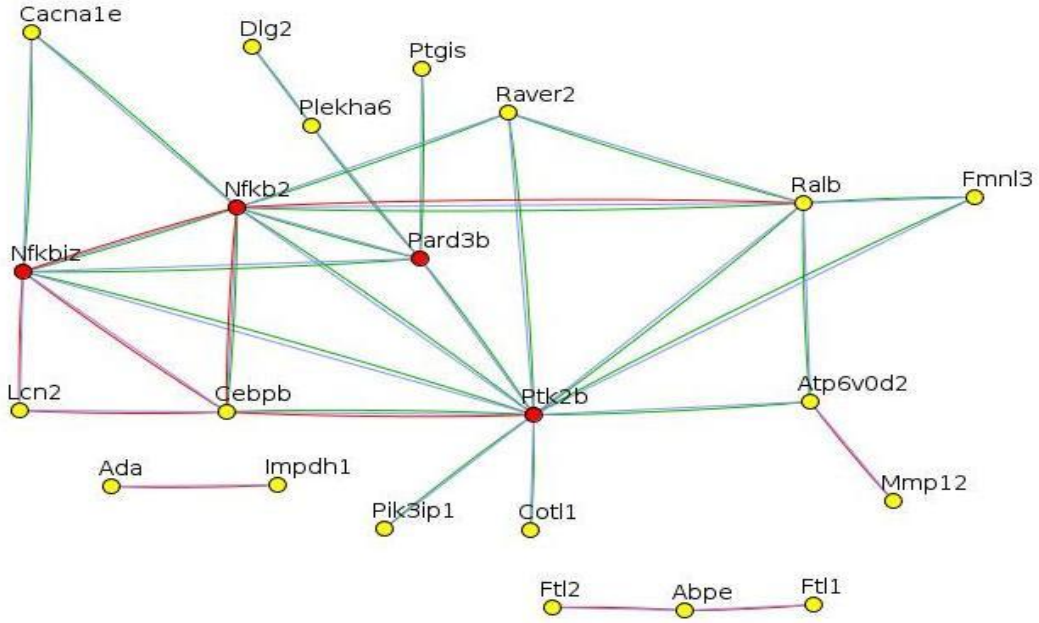


Figure 15 : Network of genes in turquoise module. The red lines display the interaction information from STRING database. The blue lines display predicted interaction using domain interaction information. Red circle defines hub genes.

Yellow module

86 co-expressed genetic interactions were identified by different methods in this module. 37 interactions were validated by domain interaction, 41 by the database and 17 by the PLSR algorithm. 11 interactions were validated by both domain interaction and the database (Figure 16). As stated before, the module is regulated by a locus on chromosome 13. Since it is a highly interconnected module, we defined the threshold for hub genes as edges >=10. Under the defined criteria, we found 4 hub genes Sykb (16 interactions), Ccrl2 and Sell (12 interactions) and Trem3 (11 interactions). Sykb and Ccrl2 are well-established therapeutic targets for various

autoimmune diseases (Wong et al., 2004; Zabel et al., 2008). Sell (L- Selectin) gene is important for immune response and a possible regulator for primary immune cells (Marschner et al., 1999). Trems are well known targets for the immunological disease (Colonna, 2003).

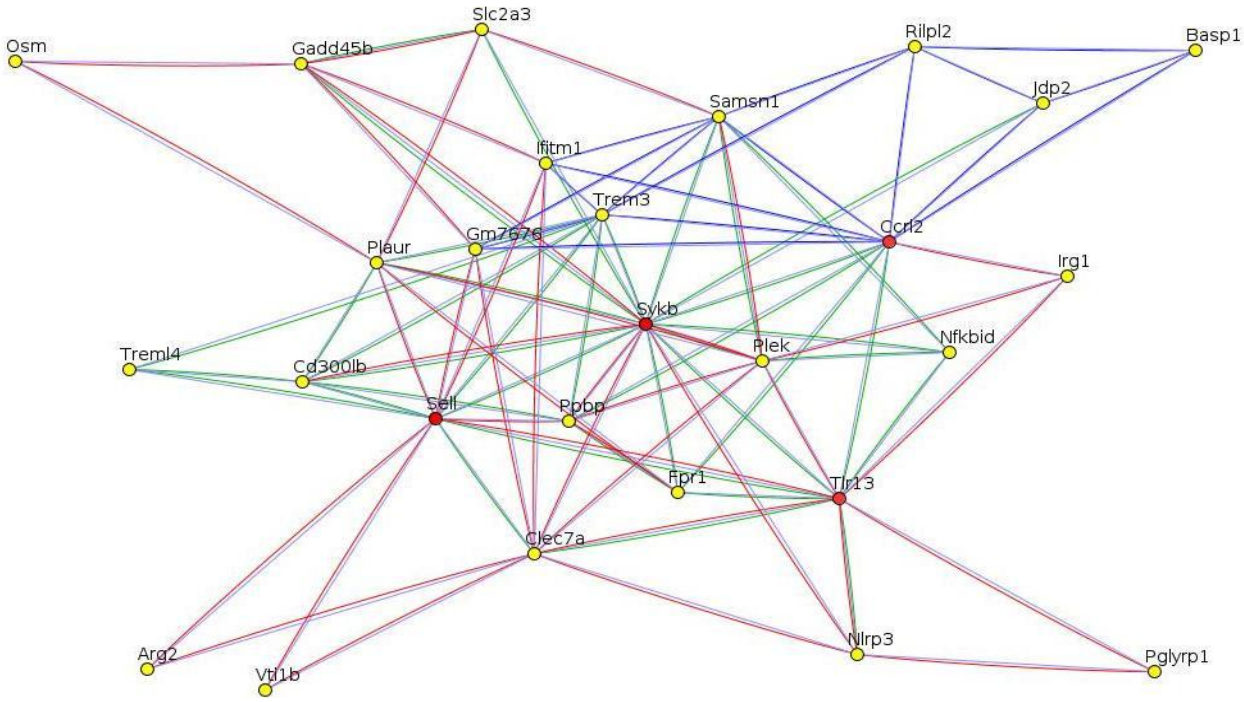


Figure 16 : Network for genes of yellow module. The red lines display the interaction information from STRING database. The green lines display predicted interaction using domain interaction information. Dark blue line represents PLSR based interaction. Red circle defines hub genes.

Red module

In this module, there are 78 interactions, out of which 65 interactions were validated by domain interaction and 13 were found in the database (Supplement

Figure 2). The hub genes characterizing this module are Pxdn, Notch (edges = 13) and CD93 (edges= 11). Notch is a known target for autoimmune diseases (Ma et al., 2010). Various drugs have been suggested for targeting Notch protein, which have proved to be effective in controlling autoimmune disease (Ma et al., 2010). Cd93 has been recently discovered to be associated with various autoimmune diseases (Greenlee-Wacker et al., 2012). We found a novel potential drug target such as Pxdn in this module. Chromosomes 11 and 15 are associated with this module.

Green module

We verified 17 co-expressed gene interactions, 16 were verified by domain interaction (Figure 17). Arhgef and Rgnef interaction was verified by both domain interactions and the database. Additionally, we found that Zfp652 was interacting with genes such as Rrp7a, Ptpn21, Crim, Serhl, Rgnef, Etv3 and Lmo7. The gene Zfp652 is significantly down-regulated in EBA (ρ = -0.30, *p-value* = 2.4e-05). The module is controlled by two loci on chromosomes 4 (primarily) and 5.

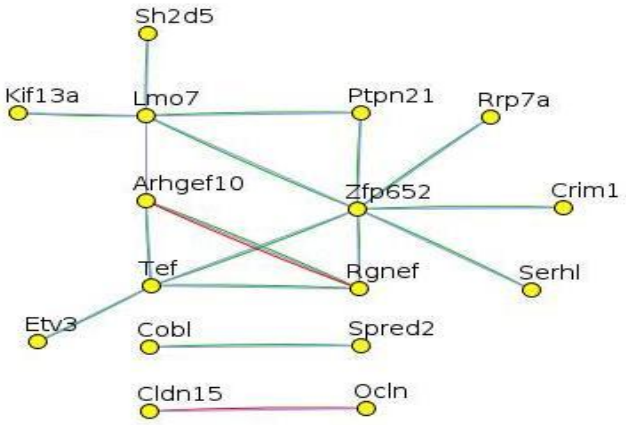


Figure 17 : Network for genes of green module. The red lines display the interaction information from STRING database. The blue lines display predicted interaction using domain interaction information.

Purple module

12 genetic interactions in this module were validated by various methods. In brief, 9 were validated by domain interaction and 6 were found in the database (Figure 18). 3 were predicted by both methods. The genes Thra, Nr1d1 and Rorc were the most important genes and have 5 edges each. It was found that Nr1d1, a nuclear receptor, is involved in the regulation of the cytokine pathway (Gibbs et al., 2012) (Figure 16).

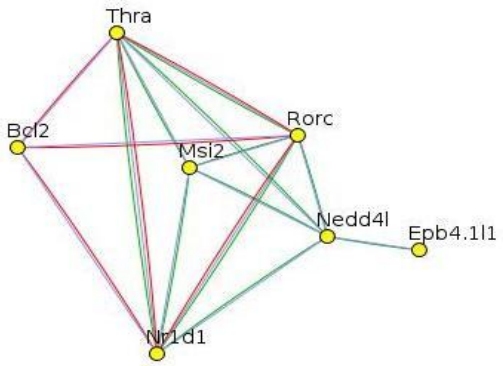


Figure 18 : Network for genes of purple module. The red lines display the interaction information from STRING database. The blue lines display predicted interaction using domain interaction information.

4.3 MicroRNA – Gene target prediction

MiRNAs are known to regulate gene expression levels. Briefly, a gene promoter binds to the mRNA via 3-8 nucleotides, making a perfect Watson-Crick pairing and regulating its transcription. This binding region is called a seed region, and is highly conserved across species. Using sequence complementarity between miRNAs and gene promoters and in addition to experiments such as reporter assay, many databases contain predicted and experimentally verified gene targets for different miRNAs such as miRanda, PICTAR, TARGETSCAN, etc. (John et al., 2004; Krek et al., 2005; Lewis et al., 2005). Additionally, many software programs follow the same principle and also include thermodynamic properties of the miRNA-gene promoter structure such as RNAhybrid, mirWALK, etc. (Issabekova et al., 2012) (Dweep et al.,

2011). Therefore, in this work, we used a combination of such databases and software to search for known targets and predict new miRNA targets respectively. The details are provided in Material and methods (section 3.6.3).

It has been recently demonstrated that miRNA eQTL could be accurately identified if their target gene eQTL information is added. This approach also decreases the false positive eQTL for miRNA (Su et al., 2011) , thus suggesting that gene targets for different miRNAs could possibly be regulated by the same locus. Conversely, the genetic information can be further used to predict possible accurate targets for diseases like EBA in which the role of miRNAs is still largely unknown.

First, we used the Pearson correlation coefficient measure to calculate the correlation between all the miRNAs which are significantly correlated with EBA score and the differentially expressed genes (i.e. 1,065 genes x 30 miRNAs). We found that a total of 16,128 miRNA-gene pairs were negatively correlated ($\rho < 0$). After correcting for multiple testing using Bonferroni corrections, 3,941 pairs of miRNA-genes were selected ($\rho < 0$, *p-value* < 0.05). To further narrow down our list, we applied an additional filter criterion. According to this criterion, miRNA-gene pairs were selected only if they show association with SNPs that are less than 10 cM apart. As a result, we could further discard many pairs, reducing our list to 249 miRNA-gene pairs.

In parallel, we downloaded all the gene targets for 30 miRNAs from databases. In order to find new targets, we used RNA hybrid software with 'mfe' (minimum free energy) <= -0.25. This software uses the secondary structure properties of the miRNA-gene duplex to predict gene targets. Additionally, we used tools such as miRwalk, RNA 22 and Targetscan to find complementary sequences between the seed region of miRNA and 3'UTR region of the gene. We allowed only 2 mismatches between the target 3'UTR region and seed region of miRNA. Overall, we identified 471 pairs between miRNA and genes from our dataset of 3,941 pairs of miRNA-genes.

The intersection between negatively correlated and genetically controlled miRNA-gene pairs and the output of the software and databases narrowed down our list to 76 miRNA-gene pairs.

We found miR-1195 (1 target), miR-1224 (8 targets), miR-1272 (2 targets), miR-134 (12 targets), miR-183 (6 targets), miR-21 (8 targets), miR-214 (10 targets), miR-223 (1 target), miR-23a (3 targets), miR-27a (1 target), miR-379 (18 targets), and miR-467g (6 targets) to be important miRNA (Figure 19).

As we stated before, we observed multiple mRNA expression levels regulated by eQTL hot spot in chromosome 6. The eQTL observed for miR-379 maps to the hot spot in chromosome 6. One of the targeted genes, Pcbd1, shares the same eQTL region 97-112 Mb with –log *p-value* of 5.56 and a phenotype variance of 13.53.

MiR-134 was found to be targeting 12 genes, including Bcl2 (also a gene target of miR-127), Drp2 (also a gene target of miR-218) and Oxtr (also a gene target of miR-379). Moreover, it has been previously reported that miR-134 down-regulates translation in different autoimmune diseases such as multiple sclerosis (Ma et al., 2014). Gene targets like Bcl2 gene were found to be hub gene in the previously described 'purple' module (Results, section 4.2.3).

MiR-214 targets 10 genes, out of which it shares 3 genes (Rab11fip4, Gm12824 and Atp5s) also targeted by miR-379 and 1 by miR-1195. miRNA is well-known for its over-expression in cancer (Penna et al., 2014). We found miRNA to be positively correlated (up-regulated) in EBA ($\rho = 0.25$, *p-value* = 0.0145). We found miRNA to be positively correlated (up-regulated) in EBA ($\rho = 0.25$, *p-value* = 0.0145). MiR-21 and miR-1224 both had 8 targets each. miR-21 is a well-established miRNA known for its regulation in different autoimmune diseases (Xu et al., 2013). In inflammation it is up-regulated and controls the interleukin pathway (Guinea-Viniegra et al., 2014).

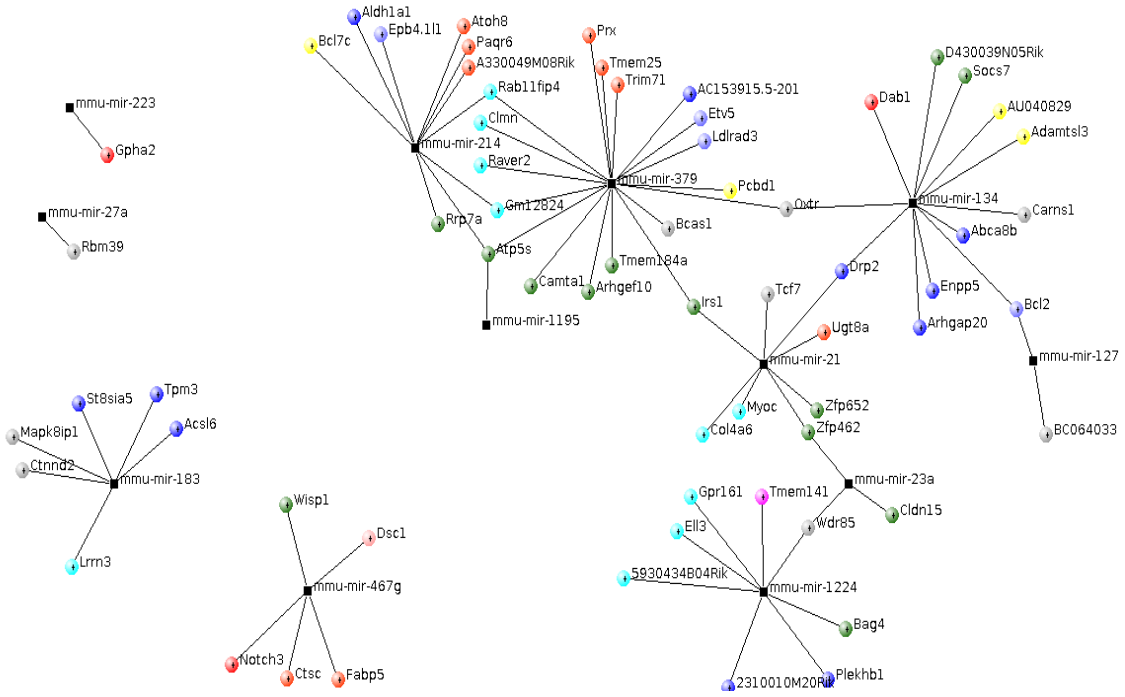


Figure 19 : miRNA – gene module interactions. The figure displays miRNA as black and other colors represent genes coming from different clusters (modules).

4.4 Prediction of non-coding RNA

Our results suggest that there are additional potent classes of non-coding RNA that can regulate miRNAs such as snoRNA, snRNA and linc RNA, etc. The understanding of these mechanisms, such as the regulation of coding transcripts by novel classes of non-coding transcript, remains elusive. Given the rapid advances in sequencing technology, it is now possible to detect new and unknown RNA molecules. Such molecules can influence transcription, translation, and epigenetic modification in genes that regulate the immune system. Hence, we set out to design

a predictor that can classify one of the subclasses of non-coding RNA known as post-transcriptional RNAs, i.e. ptRNApred (Gupta et al., 2014).

First, a binary classifier was created. This classifier distinguishes sequences of ptRNA and non-ptRNA in a binary classification. In the second step, a multi-class classifier separates 6 classes of post-transcriptional RNA (snRNA, snoRNA, RNase P, RNase MRP, Y RNA or telomerase RNA). The accuracy of 5-fold cross-validation for binary was 93% within the training set (Figure 20a). The number of sequences present in the training and testing set is provided in Table 1. The multi-class classifier yields a 5-fold cross-validation accuracy of 87% (Figure 20b).

While testing the classifiers with the test set of sequences, the binary classifier had an accuracy of 93%, with a sensitivity of 91%, a specificity of 94%, and an overall precision of 90%. The multi-class classifier had an accuracy of 91% (Table 7).

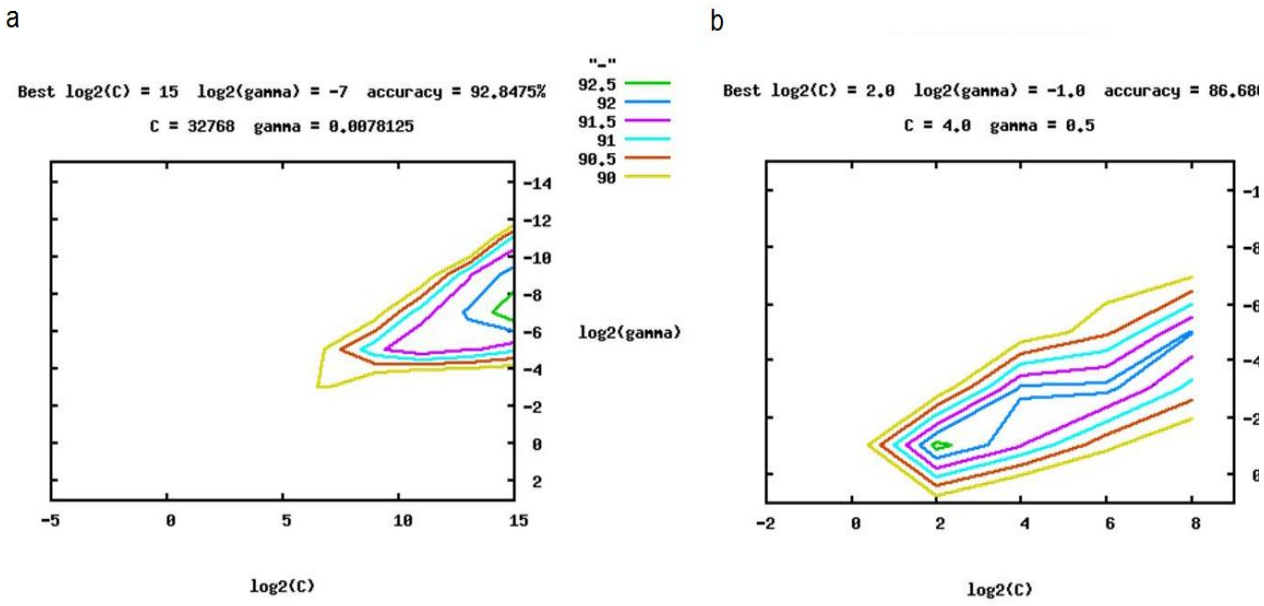


Figure 20 : C and γ determination and 5 fold cross validation using LibSVM. The figure shows graphs for different values of parameters C (a trade-off for misclassification) and γ (inverse width of RBF kernel) on a logarithmic X and Y axis. The ranges of the axes describe the different values that were tested, searching the optimal C and γ values in the grid space. The different colors in the diagram display the different accuracies obtained while optimizing C and γ values. We chose the C and γ values according to the green graphs, respectively, representing the C and γ value with the highest accuracy. A C and γ determination and 5 fold cross validation of the two-class SVM. The green graph represents the optimal values for C and gamma. In this case, the highest 5 fold cross validation accuracy (92.89%) is achieved when C=32768 and γ=0.008. b C and γ determination and 5 fold cross validation of the multi-class SVM. The green graph represents the optimal values for C and gamma. In this case, the highest 5 fold cross validation accuracy (86.69%) is achieved when C=4 and γ=0.5.

The fact that the accuracy of the test set is higher than the 5-fold cross-validation accuracy in the training set suggests that an increase in the number of sequences leads to a more accurate prediction. Consequently, we observed an increase of the accuracy of the multi-class classification when adding more sequences to the training set.

Table 7: The results of multiclass classifier are presented in a confusion matrix. As a result, implementation of Random Forest yields an overall accuracy of 82%. In comparison, our multi-class classifier developed using LibSVM yields an accuracy of 91%.

		Predicted class						
		RNase MRP	RNAse P	snoRNA	snRNA	telomerase RNA	YRNA	Accuracy
Actual class	RNase MRP	3	2	1	0	0	0	50
	RNAse P	0	100	6	0	3	0	91,7
	snoRNA	0	11	1123	75	1	0	92,8
	snRNA	0	4	194	713	2	0	78,1
	telomerase RNA	0	4	4	2	8	0	44,4
	YRNA	0	0	10	1	0	0	0
								82

4.4.1 Validation of the method

In order to validate our tool, we compared it to existing tools. We found snoReport (Hertel et al., 2008) is an advanced tool for prediction of snoRNA. It predicts orphan snoRNA without using target information, thus making it similar to our approach. To compare both tools, we derived snoRNA sequences from a mouse genome from

Ensembl (Flicek et al., 2013) and used them as an independent set. In total, we used an input of 1,603 sequences of snoRNA. As a result, snoReport identified 733 sequences correctly while our classifier could identify 1,589 sequences correctly.

Furthermore, we abstracted a human dataset with 1,641 sequences of snoRNA from Ensembl. While snoReport identified 852 of the snoRNA sequences correctly, our tool identified 1,611 sequences (Table 8).

In order to analyse the low sensitivity of snoReport, we inspected the sequences that it failed to classify. We found that snoReport was not able to detect a major snoRNA-subclass, snoU13. SnoU13 was identified in 1989 (Tyc and Steitz, 1989) . It has been well-characterized in 35 species by both functional assay and prediction. It is involved in the nucleolytic cleavage at the 3' end of 18S rRNA, where it works as a trans-acting factor (Cavaille et al., 1996).

SnoReport was unable to assign any of 245 snoU13-sequences in a human cohort to snoRNA. Our tool ptRNApred, however, identified all of them correctly. Current approaches to identifying different RNA families rely heavily on their secondary structure conservation. Consequently, these approaches are accurate as long as the RNA families show high secondary structure conservation. However, as soon as a RNA family lacks a conserved secondary structure, it will be misclassified. This can explain why snoReport failed to identify snoU13: snoU13 does not form any secondary structure conservation, as it forms a loop.

Table 8: Comparison between snoReport and ptRNApred. A murine and a human dataset of snoRNA was abstracted from Ensembl (49) and performance of ptRNApred was compared to snoReport, as a well-established tool for snoRNA prediction. ptRNApred achieved higher sensitivity than snoReport (99% vs. 46% on the murine and 98% vs. 52% on the human set of sequences). Regarding snoU13, a member of the snoRNAs, there is even larger difference in the sensitivity (100% vs. 0%).

Organism	RNA class	Total number of sequences1	Number of sequences identified by snoReport (% of total number of sequences)	Number of sequences identified by ptRNApred (% of total number of sequences)
Mus musculus	snoRNA	1,603	737(46%)	1589(99%)
H. sapiens	snoRNA	1,641	852(52%)	1611(98%)
	snoU132	2452	0(0%)2	245(100%)2

4.4.2 Validation of the algorithm

As mentioned in Materials and methods (section 3.9.4), we compared the algorithm implemented in our tool to a random forest classification. The implementation of random forest yielded an overall accuracy of 82%. In comparison, our multi-class classifier developed using LibSVM yielded an accuracy of 91%.

4.4.3 Validation of the feature number

As mentioned in Materials and methods (section 3.9.2), a general concern for all machine-learning approaches is that one has too many features, i.e. that one trains on features that are not relevant, referred to as over-fitting. This limitation was excluded by the above-mentioned cross-validation test. On the other hand, too few features would lead to loss of (overall) accuracy. In order to confirm that using fewer

features would lead to loss of accuracy, we selected the 78 most discriminating features based on random forest prediction, using the R package 'Boruta'. When using these 78 features instead of 91 features, the 5-fold cross-validation accuracy decreased from 92.89 to 74.46% (Figure 21).

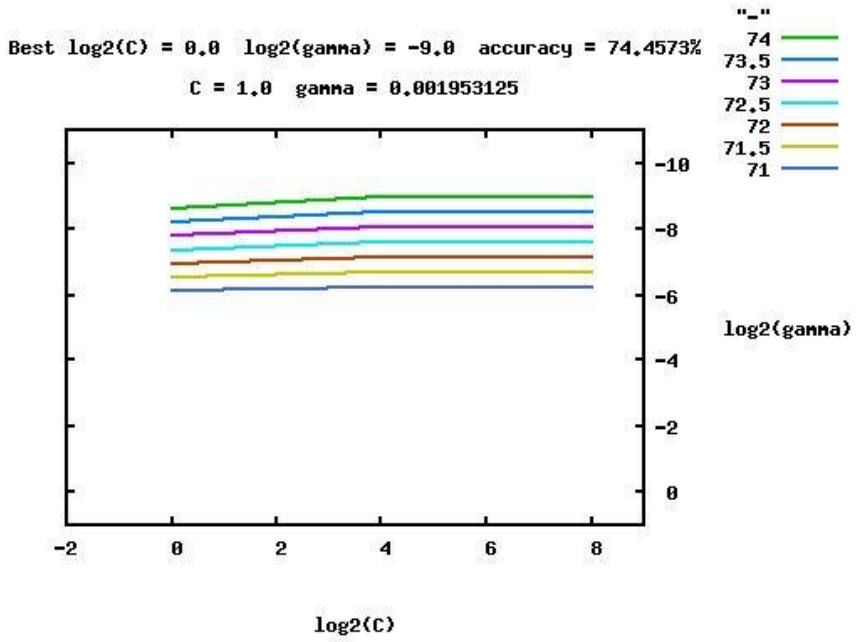


Figure 21: C and γ determination and 5 fold cross validation when using 78 instead of 91 features. The green graph represents the optimal values for C and gamma. In this case, the highest 5 fold cross validation accuracy (74.46%) is achieved when C=1 and γ=0.002.

4.4.4 Variable importance

Classification features have an individual impact on the differentiation of RNA classes. To determine the importance of each of the 91 features for classifying ptRNA, an F-score was calculated for each feature, using LibSVM. F-scores can be interpreted as a weighted average of precision and recall, where an F-score reaches its best value at 1 and worst score at 0. Supplement Table 2 depicts the F-score corresponding to every feature.

Additionally, even though the random forest method was not implemented in ptRNApred, it provided useful information on variable importance. One of the measures of variable importance in random forest is the mean decrease in accuracy, calculated using the out-of-bag sample. The difference between the prediction accuracy on the untouched out-of-bag sample and that on the out-of-bag sample permuted on one predictor variable is averaged over all trees in the forest and normalized by the standard error. This gives the mean decrease in the accuracy of that particular predictor variable which has been permuted. Thus, the importance of the predictor variables can be ranked by their mean decrease in accuracy. Table 8 depicts the Gini index corresponding to every feature.

Interestingly, comparing the 25 most discriminated feature variables according to the F-score and Gini index (Table 9), dinucleotide properties achieve high ranks: 9 of the 10 most discriminative features according to the F-score are composed of

dinucleotide properties. Furthermore, all of the 15 dinucleotide properties can be found among the 25 most discriminative properties. According to the Gini index, 12 properties can be found among the 25 most discriminative properties, whereas only 3 of them can be found among the top 10, further indicating the importance of the secondary structure.

Table 9: Table of top 25 properties ranked by their importance for discrimination among ptRNA according to F-score and Gini-Index.

Rank of importance for discrimination	Property ranked by F-score	Property ranked by Gini-Index
1	Keto_content	value_in_3rd_rnafold
2	Guanine_content	Shift
3	GC_content	Adenine_content
4	Hydrophilicity	U...
5	Roll	number_of_U_in_first_complementary_strand
6	Slide	value_MFE_RNAfold
7	U...	value_line_number_4(second value)
8	Entropy_2	G...
9	Rise	Stacking_energy
10	Twist	(((
11	Entropy_1	G(((
12	Stacking_energy	Twist
13	Cytosine_content	Roll
14	number_of_U_hairpin	Hydrophilicity
15	Adenine_content	U..(
16	G_in_buldges	value_line_no_3_RNAfold
17	number_of_buldges_in_sec_struc	Entropy_2
18	Tilt	number_of_AU
19	A_in_buldges	Tilt
20	Shift	Rise
21	number_of_A_pyramidine	Entropy_1
22	C(((	Slide
23	..(	value_line_number_4
24	(..	Guanine_content
25	value_in_3rd_rnafold	value_line_no_3_RNAfold(second value)

4.4.5 Performance on a non-eukaryotic system

Even though ptRNApred is designed to primarily predict eukaryotic sequences, ptRNApred was tested for performance on RNase P sequences, using 329 RNase P sequences from the Ribonuclease P Database (Brown, 1999). RNAse P has not only been described in eukaryotic systems (Jarrous and Reiner, 2007), but rather distributes among different organisms (Pannucci et al., 1999). Interestingly, our tool predicted the RNase P sequences with an accuracy of 97.3%.

4.4.6 Performance on mRNA

Over the last few years, several tools have been developed to distinguish coding from non-coding RNA (Badger and Olsen, 1999; Gaspar et al., 2013; Liu et al., 2006). Our aim was to develop a novel tool that can differentiate between subclasses of non-coding RNAs instead of distinguishing coding from non-coding RNAs. Nevertheless, ptRNApred was tested for performance on mRNA. Therefore, ptRNApred was challenged by 10,000 mRNA randomly downloaded sequences from Ensembl. Surprisingly, only 15 of the sequences were misclassified as ptRNA. Therefore, the accuracy of separating out mRNA is 99.85%.

5. Discussion

5.1 Candidate gene for regulation of miRNA

Small non-coding RNAs such as miRNAs are reported to contribute to the onset and severity of a vast array of diseases as well as to the defense against them (Baumjohann and Ansel, 2013). Thus, miRNAs function as tissue-specific key regulators, affecting some of the major pathways towards an aggravation of disease severity when aberrantly expressed (Liu and Kohane, 2009). Accordingly, it is not surprising to now see miRNAs discussed as potential therapeutic targets (De Guire et al., 2013; Schmidt, 2014).

However, the underlying mechanisms of such dysregulated miRNA expression patterns are not well-characterized. Multiple studies have shown that gene expression alterations across tissues are genetically derived (Liu and Kohane, 2009). Thus, it is plausible that not only the regulation of gene expression, but also the expression of miRNAs, is genetically controlled. In this study, we explore the diversity of miRNAs in inflamed skin tissue and genetic loci that control variations in miRNA expression levels across a mouse cohort (Gupta et al., submitted BMC genomics, May 2015). We provide evidence that miRNA levels in skin tissue are not genetically controlled on the transcriptional level but rather on the post-transcriptional level mediated by different regulatory factors encoded by the

genome. Furthermore, we found that some of the miRNA eQTL are restricted to one particular locus (e.g. chromosome 2, 8 and 15, Results section 4.1.1) in the genome (eQTL hot spots). Deeper investigation revealed that these miRNAs are under multi-locus and/or epistatic control.

Interestingly, eQTL hot spots were predominantly found in non-protein-coding genomic regions. Hence, it is tempting to speculate that miRNA expression might not necessarily be solely controlled by protein-coding RNA, but rather by non-coding RNA, which would impose an additional level of post-transcriptional regulation. This in turn leads to the tempting hypothesis that non-coding RNAs at least in part regulate miRNA expression. Such a scenario is supported by the fact that some non-coding RNAs have been shown to bind to miRNAs at the functional level, as demonstrated by an interaction of linc-MD1 with miR-133 and miR-135 (Cesana et al., 2011). Here, linc-MD1 works as a sponge and traps these miRNAs preventing them from binding to the canonical targets. Moreover, a recent study even shows an interaction network between lncRNAs and miRNAs (Cesana et al., 2011).

Based on our observation of an overlap between QTL controlling miRNA expression and those affecting the autoimmune blistering skin disease EBA, we conclude that there are interconnected pathways that simultaneously regulate both disease development and miRNA expression. This might explain the findings of earlier studies that show a clear correlation between aberrant miRNA expression and

autoimmune diseases (Ma et al., 2014; Marcet et al., 2011; Murata et al., 2013; Otaegui et al., 2009). Accordingly, initiation and/or progression of the disease do not primarily appear to be caused by aberrant miRNA expression. Quite to the contrary, aberrant miRNA expression could be a consequence of the disease which in turn would lead to a downward spiral. Hence, miRNAs could provide a large and unexplored reservoir of potential biomarkers for EBA and related cutaneous autoimmune skin blistering diseases, and an interesting target for therapeutic intervention.

Recently, microbiome QTL from skin samples of the same murine cohort have been identified (Srinivas et al., 2013). The QTLs observed for the microbiome also overlap with miRNA eQTL. For instance, a QTL for bacteroidales on chromosome 2 (132-152 Mb) was found to overlap with eQTL for miR-409-5p (135-152) with –log *p-value* of 4.17. An eQTL for mir-449b (50-70 Mb) on chromosome 14, as discussed before overlaps with QTL for EBA onset and also overlaps with QTL for Nisseria (56-69 Mb). On chromosome 18, we found an overlap for Prevotella (0-12 Mb) and miR-466-5p (0-20 Mb) with –log *p-value* of 4.49. On chromosome X, a QTL for streptococcus (9-34 Mb) and clostridiales (9-36 Mb) overlaps with an eQTL for miR-26a (0-36 Mb). However, limited information is available about the cause or effect of the observed statistical associations between miRNA and microbiome. But our results provide first directions for further investigation.

Overall, this study sets a complex framework of gene-gene and miRNA-gene-interaction, which eventually leads to disease development and progression. Furthermore, it gives evidence that miRNAs are important drivers of cutaneous autoimmune diseases. Moreover, the study strongly implies there is yet another, so far largely unexplored, level of regulatory networks, possibly instrumental non-coding RNAs to affect miRNA expression. In this sense, the aberrant miRNA expression would indeed be one the responsible elements for the disease progression, however, the driving force behind it might be a different one.

5.2 Regulating QTL for gene-expression

Expression QTL (eQTL) have proven a major source for information to dissect the effects from multiple genes influencing the behavior of the phenotype (Cookson et al., 2009). In autoimmune diseases like EBA they could contribute to the interpretation of molecular pathways known to contribute to the pathogenesis of the disease. Using a network approach from eQTL studies and in combination with prior knowledge from databases, we can predict possible mechanisms shaping both the onset and severity of the disease. Based on the network analysis, we were also able to determine possible hub gene candidates, which are now due to be experimentally confirmed. These new candidate genes provide new potential targets for therapeutic intervention in autoimmune diseases. Our results show that we were able to identify existing pathways that contribute to the pathophysiology in autoimmune diseases

including the cytokine-cytokine pathway, Notch signaling pathway, MAP kinase pathway, etc. Moreover, the study also provides putative interactions between genes on the basis of their control from the same chromosomal loci (Breitling et al., 2008). Such hot spots are likely to be explained by transcriptional factors to be further modelled by the transcriptional factor binding sites and regulatory RNAs, etc. However, any such immediate effects may also affect the downstream pathways. For example, in the hot spot region present on chromosome 6, a cis-regulated gene Akr1b8 was identified. Akr1b8 (aldo-keto reductase family 1, member B8) has been shown to be important in the pathogenesis of the autoimmune form of diabetes (Thessen Hedreul et al., 2013), and in an EAE (autoimmune encephalomyelitis) mouse model. Drugs such as furosemide increase the level of aldo-keto reductase (Lee et al., 2007). The drug has been previously shown to induce autoimmune skin blistering diseases such as bullous pemphigoid (Lee and Downham, 2006).

On chromosome 4, the gene Clstn1 was found to be cis-regulated. Mutations in Clstn1 have been associated with Alzheimer disease (Vagnoni et al., 2012). Chemicals like 4-hydroxytamoxifen have been shown to decrease the level of Clstn1 (Scafoglio et al., 2006). Another cis-regulating gene in the same loci was Ext1 (exostosin glycosyltransferase 1). The gene regulates the production of exostosin 1 and is involved in forming heparan sulfate complex, which is required for the formation of blood vessels and blood clotting (Simmons et al., 1999). A genetic study of psoriasis suggests Ext1 to be one of the candidate markers for this disease

(Alshobaili et al., 2010). On chromosome 11, Tex2 was a cis-regulated gene. Tex2 is expressed in the testis and is responsible for phospholipid binding. Its expression has been also found in the skin and other tissues (Lee and Hong, 2006). Although the role of the gene has not yet been associated with autoimmune diseases, it could be an important marker for further investigation.

Since we performed co-expression analysis in combination with eQTL analysis, we were able to determine genes that may underlie the disease phenotype (EBA). The module or group of genes following the same pathway led us to identify the hub candidate genes, which could be investigated in the future. For example, Mlkl was a candidate hub gene the 'black' module. The gene has mixed lineage kinase like the domain and prognostic bio-marker in pancreatic adenocarcinoma (Colbert et al., 2013). The gene causes necrotic membrane disruption upon phosphorylation. The gene is also required for Tnf-induced necroptosis (Cai et al., 2014). Tnf-alpha has been the key marker for many autoimmune diseases. In addition, the gene has been found to be very important in other chronic inflammatory diseases such as inflammatory bowel disease (Bradley, 2008).

Moreover, we found 4 hub genes in the 'yellow' module, i.e. Sykb, Trem3, Sell and Ccrl2. All four genes are known to play critical roles in autoimmune diseases in both humans and mice. A closer analysis shows that these genes are connected to each other via IL1, IL5 and Tyrobp (Figure 22). The 6 genes present in this pathway are

known for their function in leukocytes, cell movement of phagocytes and inflammatory response. A Syk inhibitor such as R406 is a known therapy for treating patients with autoimmune and allergic inflammatory diseases (Lhermusier et al., 2011).

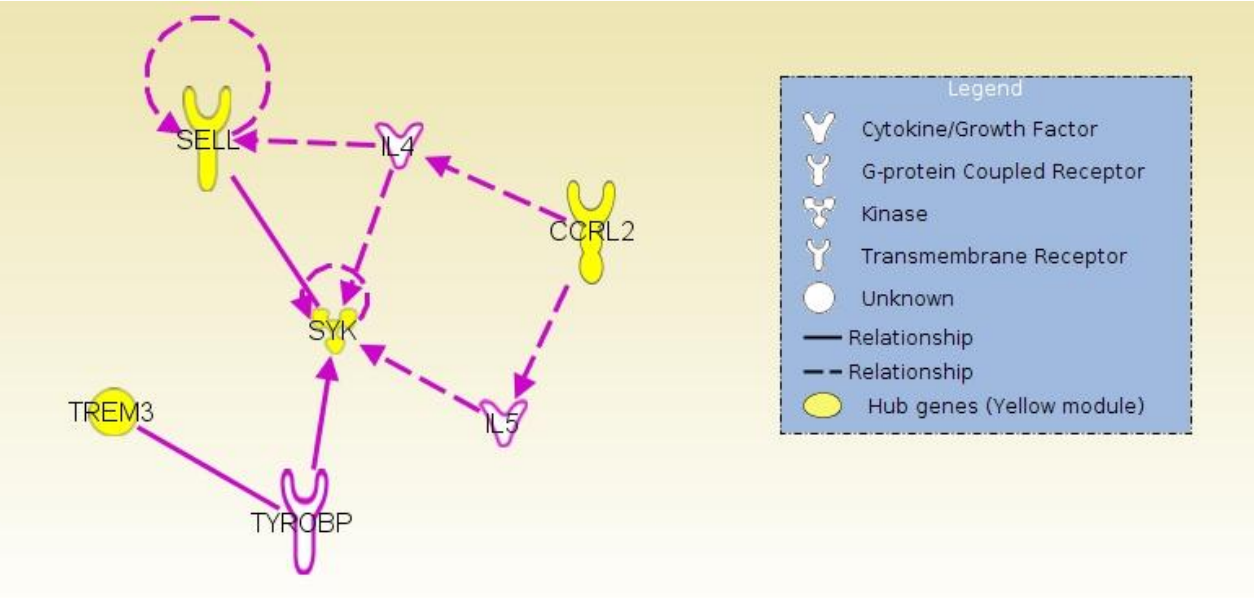


Figure 22: IPA generated pathway for yellow module hub genes. The hub genes are colored in yellow. Gene such as Tyrobp, Il4 and Il5 are signifies intermediate genes connecting hub genes. The dotted arrow show indirect interaction and plain arrows show direct interaction. The above gene symbol is for mouse identifiers.

Known genes such as Notch3 were also captured in our analysis. The gene was one of the hub genes in the 'red' module together with Pxdn and Cd93. The Notch signalling pathway is known for T-cell development and activation (Dongre et al., 2014). Many researchers have reported this pathway to be critical for the establishment of autoimmune disorders (Ma et al., 2010). However, very little is

known about the interaction of the hub genes in this module. From the database, we determined that Notch3 interacts with Pxdn with Hex1 and Nuerog1 and CD93 via the Igf1r and Esr1 gene. Pharmacological inhibitors such as gamma-secretase inhibitor and Dapt are known to suppress the Notch signalling pathway in autoimmune diseases like SLE (Jiao et al., 2014).

Thereby, understanding different gene networks provides a unique opportunity to study various new mechanisms *in silico* and identify candidate genes for the disease. Furthermore, our system genetics approach can easily be employed for any disease model. In addition to protein-coding genes, we used a similar approach to address the role of miRNAs in the pathogenesis of autoimmune disease.

5.3 miRNA-targets in EBA

MiRNAs and their targets have been one of the major focuses of the scientific community in search of new therapeutic approaches. In this work we were able to not only find the direct target of the miRNAs, but also to determine the group of co-expressed genes regulated by them. Using statistical genetics, we removed the false positive genes identified by various target prediction tools. One of the most important miRNAs identified by this approach was miR-379. Although it has not been thoroughly studied in autoimmune disease, its regulation of genes that are important in autoimmune disease makes it an interesting candidate for further investigation. The genes targets of miR-379 were found to be involved in diseases

such as cancer (*p-value* = 2.73e-02) and epithelial neoplasmia (*p-value* = 5.9e-03) (Argef10, Atp5s, Clmn, Etv5, Ldlrad3, Prx, Rab11fip4 and Tmem184a). The ontology terms for the set of genes targeted by miR-379 suggest a potential role in functions such as organism injury and abnormalities (*p-value* = 4.73e-02). Moreover, the interactions between Tmem184a and Pcbd1 in mouse and Argef10 and Prx in humans have been validated by curated databases. Much of the information about the gene targets regulated by miR-379 is not known, and hence it needs further interrogation. Another miRNA that was identified from our analysis is miR-134. The gene ontology terms suggest that genes targets of this miRNA are responsible for organism injury and abnormalities (*p-value* = 1.48e-08) and cellular assembly and organization (*p-value* = 2.9e-08). The key molecule targeted by miR-134 is Bcl2, which is also targeted by miR-127. Interestingly, the genes also interact with other gene targets of miR-134 (Figure 23). For example, Bcl2 interacts with Oxtr via Prkz, Carns1 via the Myc gene, Dab1 via Cav1 and Enpp5 via Creb. Other gene targets of the miRNA such as Dab1 were also found to be interacting with targets such as Abca8 and Enpp5. The interactions amongst the targeted genes suggest the possible regulation of molecular pathways by miRNAs in autoimmune disease.

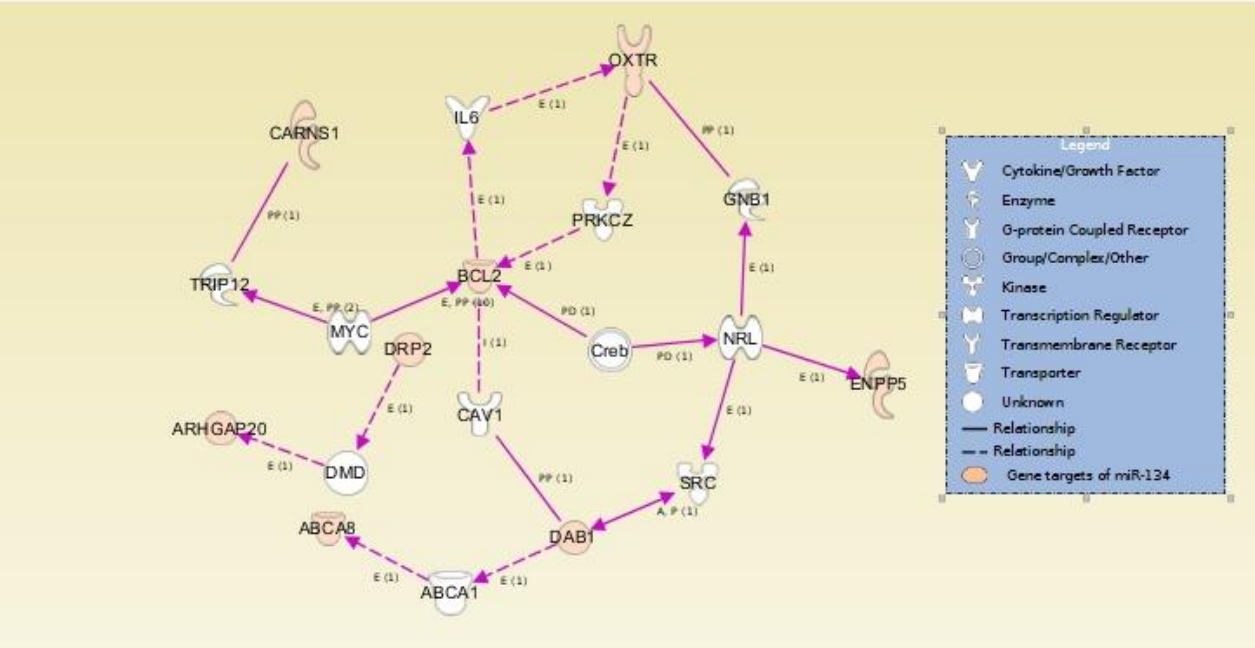


Figure 23: Gene target of miR-134 pathway generated by IPA software. The gene identifiers represented in the interaction network are for species mouse.

5.4 ptRNA-prediction tool

For the aims of this thesis we crafted a novel user-friendly tool that employs discriminative properties to (i) distinguish what we here call 'post-transcriptional RNA' from other classes of ncRNA, and (ii) discriminate between the different types of post-transcriptional RNA (Gupta et al., 2014).

An advantage of the tool is its high accuracy. This is based on its working principle: More than 90 features that are derived from the primary sequence and secondary structure are used to define properties for characterization and differentiation between the subclasses. Analysing the most discriminating feature variables

according to F-score and Gini index, the most important features are not only based on the secondary structure, but even more importantly on dinucleotide properties. This might be due to the fact that many nucleic acid properties such as nucleic acid stability, for example, seem to depend primarily on the identity of nearest-neighbour nucleotides (SantaLucia, 1998). Furthermore, the corresponding nearest-neighbour model is also the basis for RNA secondary structure prediction by free-energy minimization (Mathews and Turner, 2006). It has long been known that thermodynamic, but also conformational, nucleotide properties may be of functional relevance. It has been shown, for example, that promoter locations can be predicted adopting dinucleotide stiffness parameters derived from molecular dynamic simulations (Goni et al., 2007). Our tool shows the value of these properties.

Recently, a major effort has been focused on the characterization of snoRNA. However, there has been no classifier that could predict snRNA, RNase P or RNase MRP, even though these subclasses have conserved secondary structures. The identification of those RNA classes has as yet been dependent on sequence alignment. This technique frequently leads to misidentification, especially if the particular homologous sequence is not present in any database.

Furthermore, ptRNApred can be used to elucidate unknown relations and derivations of RNA classes. Based on the assumption that evolutionary close RNA

families have similar sequence properties, one can speculate that tools like ptRNApred will falsely arrange evolutionary close RNA families in the same group.

A deficiency of this tool is that its accuracy is dependent on the amount of published ptRNA sequences. Some classes of post-transcriptional RNA, for example Y RNA, at to this point rarely available in the NONCODE database, making it hard to define discriminative sequence properties. In the current era of high-throughput next-generation sequencing, where a large amount of genomic data is generated each day, post-transcriptional RNA sequences that will be added to the database in the future can be used to increase training and test sets, setting a basis to improve the classifier. On the other hand, the discovery of new candidates for post-transcriptional RNA requires a method that can classify them rapidly and reliably. Our tool offers a solution to this problem. In addition, facing the huge amount of new sequences that are found in NGS or RNA-Seq data (Morin et al., 2008), it is important to include such algorithms into NGS pipelines. For such purposes, we provide a standalone version.

We implemented our method as a web-based server for free public use. The transparent and user-friendly design makes it possible for everyone to understand and employ the tool. Data and scripts for the development of the tool can be downloaded, allowing anyone to acquire the working principles and improve ptRNApred.

Collectively, our tool offers a fast and reliable way to analyse DNA sequences and outperforms the existing classifiers. Furthermore, the tool provides comprehensive annotations. Therefore, ptRNApred introduces different opportunities for identifying and classifying new and un-annotated RNA sequences.

6. Conclusions

This work presents the first description of the genetic control of microRNAs (miRNA) expression in the murine skin. With a systems genetics approach, many key regulators and potential biomarkers for autoimmune skin blistering diseases were revealed. These were found both among protein-coding transcripts and the regulatory non-coding RNA.

The introduced predictor 'ptRNApred' in broadest the best performing classifier for prediction of post transcriptional non coding RNA. Focusing on features in sequence and structures, it depends less on current knowledge and thus facilitates the analyses to migrate from biased hybridization-based analyses to the whole-transcriptome RNA-Seq technology.

7. Summary

This work describes the role of miRNA and genes with their interactions in autoimmune skin blistering disease, and newly developed software to identify non-coding RNAs. In this work, first evidence for the genetic regulation of miRNA and protein-coding gene expression in skin is provided. In brief, 42 eQTL for 38 cutaneous miRNAs have been identified. Four of these miRNA i.e. eQTL for miR130b (Chr 9: 36-44 Mb, -log *p-value* = 4.42), miR-542-3p (Chr 12: 79-103Mb, -log *p-value* = 4.73) and miR-449b (Chr 14: 50-72 Mb, -log *p-value* = 4.49) were mapped on the QTLs for EBA (an autoimmune skin blistering disease) disease onset on chromosome 9, 12 and 14. Additionally, eQTL for miR-493 (50-60 Mb, -log *p-value* = 4.98) was mapped on EBA QTLs for both disease severity and onset on chromosome 19 suggesting common genetic regulation between EBA disease and miRNAs. Thereafter, the data is integrated using systems genetics approach (weighted gene co-expression networks) to define plausible miRNA pathways involved in disease phenotype. Specifically, miRNAs such as miR-379, miR-223 and miR-21 were speculated to play important role in the pathogenesis of EBA. Moreover, based on the overlap of eQTL and stronger correlation with onset week of EBA, it is speculated that miRNAs play crucial role in onset of EBA rather than severity of the EBA.

To further investigate, the role of protein coding genes in the EBA disease phenotype, eQTL for gene expression were also studied. First, 1,039 genes were identified differentially expressed between EBA and healthy skin. Further, to investigate molecular pathways in EBA, we found 8 differentially co-expressed gene modules were significantly positively correlated, while 4 were significantly negatively correlated with EBA disease severity scores ($p\text{-}value < 0.05$). Additionally, 424 eQTL for 260 mRNA expression levels (Affymetrix probes) associated with 251 genes from differentially expressed genes. It was observed 83/260 genes were cis-regulated i.e. regulated by its own transcription site, while 177/260 were trans-regulated. We found four eQTL hot spots (genetic loci regulating expression of more than 20 genes) for gene expression on chromosome 6, 1, 4 and 11. Combining eQTL, co-expression and manually curated interaction database such as STRING and DOMINE database we derived hypothetical gene network for EBA disease phenotype. Genes such as Syk, Notch1, Trem3 and etc. were identified as potential regulators (hub genes) in disease network. Among the gene such as Syk and Notch1 which have been already described many autoimmune diseases; new biomarkers such as Mlkl, Ptk2b, and Sell were suggested by the statistical analysis contributing to skin blistering diseases phenotype.

Since the data for miRNA and gene expression was derived from the same cohort of advance inter-cross line, it was possible to define miRNA gene targets in EBA. Therefore we defined miRNA genes target using negative correlation between

miRNA and gene, prediction using database and overlap of eQTL between miRNA and gene. Using the described criteria, miRNA targets genes such as Pcbd1 and Bcl2 for miR-379 and miR-134 are suggested in the study for the pathogenesis of skin blistering disease.

Many non-coding classes of RNA other than miRNA which are known as post transcriptional RNA have been described in last decade. These post transcriptional RNA includes snoRNA, snRNA, telomerase RNA and etc. Though regulations of these RNA are poorly understood, but they affect various disease phenotypes. Therefore, it is important to identify them in order to understand the pathogenesis of disease. A self-developed software "ptRNApred" categorizes post-transcriptional non-coding RNA employing a machine learning algorithm. It distinguishes multiple sub-classes such as miRNA, snoRNA, snRNA as possible regulators of disease and prepares the analyses on non-coding RNA interference with genes to be repeated on whole transcriptome data derived with RNAseq data and various sequencing technologies. Moreover, it surpasses the sensitivity and specificity of the existing tool such as snoReport software designed to predict snoReport, a sub class of post transcriptional RNA.

The work finds its importance as it is first step to understand the complete pathways of genes, miRNA and other non-coding RNA in pathogenesis of an

autoimmune disease of the skin. As such, it will guide respective investigations in autoimmune diseases also for tissues that are less easily accessible.

8. Zusammenfassung

Diese Dissertation beschreibt die Rolle von miRNA und Genen, sowie deren Interaktionen in blasenbildenden Autoimmunerkrankungen in einer Kohorte von AIL Mäusen[1]. Des Weiteren wird eine neu entwickelte Software zur Identifikation von nicht-kodierender RNA beschrieben. Diese Arbeit liefert erste Hinweise auf die genetische Regulation von miRNA und Protein-kodierenden Genen in der Haut. Es wurden 42 eQTL[2] zu 38 kutan-exprimierten miRNAs identifiziert. Drei dieser miRNAs, genauer, die eQTL für miR-130b (Chr 9: 36-44 Mb, -log p-Wert = 4.42), miR-542-3p (Chr 12: 79-103Mb, -log p-Wert = 4.73) und miR-449b (Chr 14: 50-72 Mb, -log p-Wert = 4.49) überlagerten sich mit drei QTL, welche mit dem Krankheitsbeginn der EBA (Epidermolysis bullosa acquisita, eine blasenbildende Autoimmunerkrankung) assoziiert sind und entsprechend ebenfalls auf den Chromosomen 9, 12 und 14 liegen. Zusätzlich wurde eine Überschneidung des eQTL für miR-493 (Chr:19 50-60 Mb, -log p-Wert = 4.98) mit einem QTL entdeckt, welcher mit dem Schweregrad und Beginn der EBA assoziiert ist und ebenfalls auf Chromosome 19 liegt. Zusammengenommen deuten die Ergebnisse auf eine gemeinsame genetische Regulation der vier identifizierten miRNAs und den Pathomechanismen der EBA hin. Die gewonnenen Daten wurden darauffolgend nach einem systemgenetischen Ansatz (gewichtete Genkoexpressionsnetzwerke) integriert, um plausible miRNA Signalwege zu definieren, welche an der

Ausprägung des Krankheitsbildes der EBA beteiligt sind. Insbesondere bei den miRNAs miR-379, miR-223 und miR-21 kann über eine wichtige Rolle in der Pathogenese der EBA spekuliert werden. Die gefundenen Überschneidungen zwischen miRNA eQTL und den QTL für EBA, sowie eine starke Korrelation der eQTL mit dem Krankheitsbeginn der EBA lassen auf eine bedeutende Rolle der miRNAs bei der Induktion der EBA schließen und weniger auf eine Korrelation mit dem Schweregrad der Erkrankung.

Um zusätzlich auch die Rolle von Protein-kodierenden Genen im Krankheitsbild der EBA genauer zu überprüfen wurden eQTL für Genexpression untersuchen. Zunächst wurde die Genexpression in gesunder Haut mit der Genexpression in Haut mit EBA Läsionen verglichen, wobei 1039 differentiell exprimierte Gene identifiziert wurden. Um die zugrundeliegenden molekularen Signalwege für die EBA zu identifizieren wurden differenziell koexprimierte Genmodule untersucht, von denen 8 signifikant positiv und 4 signifikant negativ mit dem Schweregrad der EBA korrelierten (p-Wert < 0.05). Zusätzlich konnten 424 eQTL, welche mit den Expressionsniveaus (Affymetrix Gensonden) von 260 mRNAs korrelierten, mit 251 der in EBA differenziell exprimierten Gene assoziiert werden. 83 der 260 Gene waren cis-reguliert[3], wohingegen 177 der 260 Gene als trans-reguliert[4] identifiziert wurden. Es wurden vier eQTL Hotspots (genetische Loci, welche die Expression von mehr als 20 Gene regulieren) entdeckt, welche an der Genexpression auf den Chromosomen 1, 4, 6, und 11 beteiligt sind. Durch die Kombination von eQTL

Positionen, Koexpressionsdaten von Genen und der Verwendung von manuell kuratierter Interaktionsdatenbanken, wie die STRING und DOMINE Datenbanken, konnte ein hypothetisches Gennetzwerk für das Krankheitsbild der EBA erstellt werden. Gene wie *Syk*, *Notch1*, *Trem3* und weitere Gene wurden als potentielle Regulatoren (Hub-Gene) im Gennetzwerk für die Erkrankung identifiziert. Neben Genen wie *Syk* und *Notch1*, deren Bedeutung schon für diverse Autoimmunerkrankungen beschrieben wurden, konnten durch Anwendung von statistischen Analysen neue Biomarker, wie *Mlkl*, *Ptk2b*, und *Sell*, vorgeschlagen werden, die Einfluss auf die Ausprägung von blasenbildenden Autoimmunerkrankungen haben könnten.

Die Datensätze zu miRNA und Genexpression wurden in der selben Kohorte von AIL Mäusen generiert, wodurch es möglich war Gene zu identifizieren die im Kontext des Krankheitsbildes der EBA durch miRNA beeinflusst wurden. Um die Interaktionspartner von miRNA und dem entsprechenden Zielgen zu definieren wurde nach negativen Korrelationen zwischen miRNA und Genexpression gesucht. Weiterhin wurden Vorhersagen mit Online-Datenbanken erstellt und die Überlappung von eQTL zwischen miRNA und Gensequenzen bestimmt. Mit den genannten Kriterien wurden für die miRNAs miR-379 und miR-134 die entsprechenden Zielgene *Pcbd1* und *Bcl2* vorhergesagt, welche Einfluss auf die Pathogenese von blasenbilden Autoimmunerkrankungen haben könnten.

Im Laufe der letzten 10 Jahre wurden, neben miRNAs, viele neue Klassen von nicht-kodierender RNA definiert und werden allgemein als posttranskriptionelle RNA beschrieben. Diese posttranskriptionellen RNAs beinhalten unter anderem snoRNA, snRNA und die RNA-Komponente der Telomerase (Telomerase RNA component; TERC). Obwohl über die Regulation dieser RNA Varianten nur wenig bekannt ist, gibt es viele Hinweise darauf, dass sie die Pathogenese verschiedendster Erkrankungen beeinflussen. Eine, während dieser Dissertation eigens entwickelte, Software namens "ptRNApred" kategorisiert posttranskriptionelle nicht-kodierende RNA mittels eines Algorithmus aus dem Bereich des maschinellen Lernens. Die Software wurde programmiert um verschiedene Unterklasse von RNAs, wie etwa miRNA, snoRNA, snRNA und mehr, zu unterscheiden. Die Vorhersage von nicht-kodierenden RNAs ermöglicht eine darauffolgende Untersuchung von Genexpressionsinterferenz durch nicht-kodierenden RNAs und kann durch Transkriptionsdaten, generiert aus RNA Sequenzierungen und anderen Sequenziertechnologien, erweitert und validiert werden. Weiterhin sind Sensitivität und Spezifität dieser Software höher, als in anderen vorhandenen Programmen, wie etwa der Software "snoReport", welche lediglich zur Vorhersage von snoRNA dient.

Diese Dissertation trägt zum Verständnis der Pathogenese von Autoimmunerkrankungen der Haut bei und ist ein erster Schritt zur Identifikation der vollständigen Signalwege und Wechselwirkungen, über die Gene mit miRNA und andere nicht-kodierenden RNAs interagieren und die Ausprägung von Erkrankungen beeinflussen. Die Untersuchungen in dieser Dissertation können richtungsweisend auch für anderen Autoimmunerkrankungen angewendet werden.

9. References

Abu-Hijleh, G., Reid, O., and Scothorne, R.J. (1997). Cell death in the developing chick knee joint: I. Spatial and temporal patterns. Clin Anat *10*, 183-200.

Alshobaili, H.A., Shahzad, M., Al-Marshood, A., Khalil, A., Settin, A., and Barrimah, I. (2010). Genetic background of psoriasis. Int J Health Sci (Qassim) *4*, 23-29.

Altay, G., and Emmert-Streib, F. (2011). Structural influence of gene networks on their inference: analysis of C3NET. Biol Direct *6*, 31.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Amagai, M., Hashimoto, T., Green, K.J., Shimizu, N., and Nishikawa, T. (1995). Antigen-specific immunoadsorption of pathogenic autoantibodies in pemphigus foliaceus. J Invest Dermatol *104*, 895-901.

Artzi, S., Kiezun, A., and Shomron, N. (2008). miRNAminer: a tool for homologous microRNA gene search. BMC Bioinformatics *9*, 39.

Badger, J.H., and Olsen, G.J. (1999). CRITICA: coding region identification tool invoking comparative analysis. Mol Biol Evol *16*, 512-524.

Ballabio, E., Mitchell, T., van Kester, M.S., Taylor, S., Dunlop, H.M., Chi, J., Tosi, I., Vermeer, M.H., Tramonti, D., Saunders, N.J.*, et al.* (2010). MicroRNA expression in Sezary syndrome: identification, function, and diagnostic potential. Blood *116*, 1105-1113.

Baumjohann, D., and Ansel, K.M. (2013). MicroRNA-mediated regulation of T helper cell differentiation and plasticity. Nat Rev Immunol *13*, 666-678.

Beck, D., Ayers, S., Wen, J., Brandl, M.B., Pham, T.D., Webb, P., Chang, C.C., and Zhou, X. (2011). Integrative analysis of next generation sequencing for small non-coding RNAs and transcriptional regulation in Myelodysplastic Syndromes. BMC Med Genomics *4*, 19.

Beck, T., Hastings, R.K., Gollapudi, S., Free, R.C., and Brookes, A.J. (2014). GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet *22*, 949-952.

Beckmann, C., Brittnacher, M., Ernst, R., Mayer-Hamblett, N., Miller, S.I., and Burns, J.L. (2005). Use of phage display to identify potential Pseudomonas aeruginosa gene products relevant to early cystic fibrosis airway infections. Infect Immun *73*, 444-452.

Beebe, A.M., Cua, D.J., and de Waal Malefyt, R. (2002). The role of interleukin-10 in autoimmune disease: systemic lupus erythematosus (SLE) and multiple sclerosis (MS). Cytokine Growth Factor Rev *13*, 403-412.

Beitzinger, M., and Meister, G. (2011). Experimental identification of microRNA targets by immunoprecipitation of Argonaute protein complexes. Methods Mol Biol *732*, 153-167.

Below, J.E., Gamazon, E.R., Morrison, J.V., Konkashbaev, A., Pluzhnikov, A., McKeigue, P.M., Parra, E.J., Elbein, S.C., Hallman, D.M., Nicolae, D.L.*, et al.* (2011). Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. Diabetologia *54*, 2047-2055.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.

Berwick, R. (2002). An Idiot's guide to Support vector machines (SVMs).
Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M.*, et al.* (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A *98*, 13790-13795.
Blanco, P., Palucka, A.K., Pascual, V., and Banchereau, J. (2008). Dendritic cells and cytokines in human inflammatory and autoimmune diseases. Cytokine Growth Factor Rev *19*, 41-52.
Bock, J.R., and Gough, D.A. (2001). Predicting protein--protein interactions from primary structure. Bioinformatics *17*, 455-460.
Bompfunewerer, A.F., Flamm, C., Fried, C., Fritzsch, G., Hofacker, I.L., Lehmann, J., Missal, K., Mosig, A., Muller, B., Prohaska, S.J.*, et al.* (2005). Evolutionary patterns of non-coding RNAs. Theory Biosci *123*, 301-369.
Bonazzi, V.F., Stark, M.S., and Hayward, N.K. (2012). MicroRNA regulation of melanoma progression. Melanoma Res *22*, 101-113.
Borradori, L., and Sonnenberg, A. (1999). Structure and function of hemidesmosomes: more than simple adhesion complexes. J Invest Dermatol *112*, 411-418.
Boser, B., Guyon, I., Vapnik,V (1992). A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Work-shop on Computational Learning Theory,.
Botta, V., Louppe, G., Geurts, P., and Wehenkel, L. (2014). Exploiting SNP correlations within random forest for genome-wide association studies. PLoS One *9*, e93379.
Boulesteix, A., Janitza S, König, I. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Int Rev Data Min and Knowl Disc *2*, 493-507.
Bradley, J.R. (2008). TNF-mediated inflammatory disease. J Pathol *214*, 149-160.
Breiman, L. (2001). Random Forests. Machine Learning *45*, 5-32.
Breiman L. (1996). Bagging Predictors. Machine Learning *26*, 123-140.
Breiman L. (1996b). Out-of-bag estimation. Technical report, Department of Statistics: University of California, Berkeley.
Breiman L., F.J.H., Olshen R.A. and Stone C.J. (1984). Classification and Regression Trees. Belmont, California.
Breitling, R., Li, Y., Tesson, B.M., Fu, J., Wu, C., Wiltshire, T., Gerrits, A., Bystrykh, L.V., de Haan, G., Su, A.I.*, et al.* (2008). Genetical genomics: spotlight on QTL hotspots. PLoS Genet *4*, e1000232.
Broman, K.W. (2001). Review of statistical methods for QTL mapping in experimental crosses. Lab Anim (NY) *30*, 44-52.
Brown, J.W. (1999). The Ribonuclease P Database. Nucleic Acids Res *27*, 314.
Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci *10*, 2763-2788.
Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. Nucleic Acids Res *41*, D226-232.
Byvatov, E., and Schneider, G. (2003). Support vector machine applications in bioinformatics. Appl Bioinformatics *2*, 67-77.
Cai, Z., Jitkaew, S., Zhao, J., Chiang, H.C., Choksi, S., Liu, J., Ward, Y., Wu, L.G., and Liu, Z.G. (2014). Plasma membrane translocation of trimerized MLKL protein is required for TNF-induced necroptosis. Nat Cell Biol *16*, 55-65.
Calin, G.A., and Croce, C.M. (2006). MicroRNA signatures in human cancers. Nat Rev Cancer *6*, 857-866.

Capuano, M., Iaffaldano, L., Tinto, N., Montanaro, D., Capobianco, V., Izzo, V., Tucci, F., Troncone, G., Greco, L., and Sacchetti, L. (2011). MicroRNA-449a overexpression, reduced NOTCH1 signals and scarce goblet cells characterize the small intestine of celiac patients. PLoS One *6*, e29094.
Carlson, T., Kroenke, M., Rao, P., Lane, T.E., and Segal, B. (2008). The Th17-ELR+ CXC chemokine pathway is essential for the development of central nervous system autoimmune disease. J Exp Med *205*, 811-823.
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C*., et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.
Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363-2367.
Cavaille, J., Hadjiolov, A.A., and Bachellerie, J.P. (1996). Processing of mammalian rRNA precursors at the 3' end of 18S rRNA. Identification of cis-acting signals suggests the involvement of U13 small nucleolar RNA. Eur J Biochem *242*, 206-213.
Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell *147*, 358-369.
Chalaris, A., Adam, N., Sina, C., Rosenstiel, P., Lehmann-Koch, J., Schirmacher, P., Hartmann, D., Cichy, J., Gavrilova, O., Schreiber, S*., et al.* (2010). Critical role of the disintegrin metalloprotease ADAM17 for intestinal inflammation and regeneration in mice. J Exp Med *207*, 1617-1624.
Chang CC, L.C. (2001,2007). LIBSVM: a library for supportvector machines.
Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L*., et al.* (2013). The BioGRID interaction database: 2013 update. Nucleic Acids Res *41*, D816-823.
Chen, S.H., Sun, J., Dimitrov, L., Turner, A.R., Adams, T.S., Meyers, D.A., Chang, B.L., Zheng, S.L., Gronberg, H., Xu, J*., et al.* (2008). A support vector machine approach for detecting gene-gene interaction. Genet Epidemiol *32*, 152-167.
Chen, X., and Ishwaran, H. (2012). Random forests for genomic data analysis. Genomics *99*, 323-329.
Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol *4*, e210.
Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat Rev Genet *15*, 34-48.
Colbert, L.E., Fisher, S.B., Hardy, C.W., Hall, W.A., Saka, B., Shelton, J.W., Petrova, A.V., Warren, M.D., Pantazides, B.G., Gandhi, K*., et al.* (2013). Pronecrotic mixed lineage kinase domain-like protein expression is a prognostic biomarker in patients with early-stage resected pancreatic adenocarcinoma. Cancer *119*, 3148-3155.
Colonna, M. (2003). TREMs in the immune system and beyond. Nat Rev Immunol *3*, 445-453.
Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. Nat Rev Genet *10*, 184-194.
Cooper, G.S., Bynum, M.L., and Somers, E.C. (2009). Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun *33*, 197-207.
Cooper, G.S., and Stroehla, B.C. (2003). The epidemiology of autoimmune diseases. Autoimmun Rev *2*, 119-125.

Csorba, K., Chiriac, M.T., Florea, F., Ghinia, M.G., Licarete, E., Rados, A., Sas, A., Vuta, V., and Sitaru, C. (2014). Blister-inducing antibodies target multiple epitopes on collagen VII in mice. J Cell Mol Med *18*, 1727-1739.
Cutler, A., and Stevens, J.R. (2006). Random forests for microarrays. Methods Enzymol *411*, 422-432.
Dai, R., McReynolds, S., Leroith, T., Heid, B., Liang, Z., and Ahmed, S.A. (2013). Sex differences in the expression of lupus-associated miRNAs in splenocytes from lupus-prone NZB/WF1 mice. Biol Sex Differ *4*, 19.
Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics *25*, 1974-1975.
De Guire, V., Robitaille, R., Tetreault, N., Guerin, R., Menard, C., Bambace, N., and Sapieha, P. (2013). Circulating miRNAs as sensitive and specific biomarkers for the diagnosis and monitoring of human diseases: promises and challenges. Clin Biochem *46*, 846-860.
Delfino, K.R., and Rodriguez-Zas, S.L. (2013). Transcription factor-microRNA-target gene networks associated with ovarian cancer survival and recurrence. PLoS One *8*, e58608.
Ding, C.H., and Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. Bioinformatics *17*, 349-358.
Ding, Y. (2006). Statistical and Bayesian approaches to RNA secondary structure prediction. RNA *12*, 323-331.
Dongre, A., Surampudi, L., Lawlor, R.G., Fauq, A.H., Miele, L., Golde, T.E., Minter, L.M., and Osborne, B.A. (2014). Non-Canonical Notch Signaling Drives Activation and Differentiation of Peripheral CD4(+) T Cells. Front Immunol *5*, 54.
Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A.*, et al.* (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat Genet *42*, 295-302.
Dweep, H., Gretz, N., and Sticht, C. (2014). miRWalk database for miRNA-target interactions. Methods Mol Biol *1182*, 289-305.
Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform *44*, 839-847.
Eastmond, C.J. (1994). Psoriatic arthritis. Genetics and HLA antigens. Baillieres Clin Rheumatol *8*, 263-276.
Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION *99*, 96-104.
Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.
Emmert-Streib, F., de Matos Simoes, R., Mullan, P., Haibe-Kains, B., and Dehmer, M. (2014). The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. Front Genet *5*, 15.
Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet *12*, 861-874.
Farber, C.R. (2012). Systems genetics: a novel approach to dissect the genetic basis of osteoporosis. Curr Osteoporos Rep *10*, 228-235.
Feltus, F.A. (2014). Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. Plant Sci *223*, 45-48.
Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S.*, et al.* (2013). Ensembl 2013. Nucleic Acids Res *41*, D48-55.

Flutre, T., Wen, X., Pritchard, J., and Stephens, M. (2013). A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet *9*, e1003486.
Fortino, V., Kinaret, P., Fyhrquist, N., Alenius, H., and Greco, D. (2014). A robust and accurate method for feature selection and prioritization from multi-class OMICs data. PLoS One *9*, e107801.
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*, D808-815.
Friedel, M., Nikolajewa, S., Suhnel, J., and Wilhelm, T. (2009). DiProGB: the dinucleotide properties genome browser. Bioinformatics *25*, 2603-2604.
Fuentes, G., Oyarzabal, J., and Rojas, A.M. (2009). Databases of protein-protein interactions and their use in drug discovery. Curr Opin Drug Discov Devel *12*, 358-366.
Fulci, V., Scappucci, G., Sebastiani, G.D., Giannitti, C., Franceschini, D., Meloni, F., Colombo, T., Citarella, F., Barnaba, V., Minisola, G.*, et al.* (2010). miR-223 is overexpressed in T-lymphocytes of patients affected by rheumatoid arthritis. Hum Immunol *71*, 206-211.
Gaspar, P., Moura, G., Santos, M.A., and Oliveira, J.L. (2013). mRNA secondary structure optimization using a correlated stem-loop prediction. Nucleic Acids Res *41*, e73.
Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.
Geng, X., Zhang, R., Yang, G., Jiang, W., and Xu, C. (2012). Interleukin-2 and autoimmune disease occurrence and therapy. Eur Rev Med Pharmacol Sci *16*, 1462-1467.
Gestermann, N., Mekinian, A., Comets, E., Loiseau, P., Puechal, X., Hachulla, E., Gottenberg, J.E., Mariette, X., and Miceli-Richard, C. (2010). STAT4 is a confirmed genetic risk factor for Sjogren's syndrome and could be involved in type 1 interferon pathway signaling. Genes Immun *11*, 432-438.
Gibbs, J.E., Blaikley, J., Beesley, S., Matthews, L., Simpson, K.D., Boyce, S.H., Farrow, S.N., Else, K.J., Singh, D., Ray, D.W.*, et al.* (2012). The nuclear receptor REV-ERBalpha mediates circadian regulation of innate immunity through selective regulation of inflammatory cytokines. Proc Natl Acad Sci U S A *109*, 582-587.
Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A.*, et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.
Gommans, W.M., and Berezikov, E. (2012). Controlling miRNA regulation in disease. Methods Mol Biol *822*, 1-18.
Goni, J.R., Perez, A., Torrents, D., and Orozco, M. (2007). Determining promoter location based on DNA structure first-principles calculations. Genome Biol *8*, R263.
Gonzales, N.M., and Palmer, A.A. (2014). Fine-mapping QTLs in advanced intercross lines and other outbred populations. Mamm Genome *25*, 271-292.
Greenlee-Wacker, M.C., Galvan, M.D., and Bohlson, S.S. (2012). CD93: recent advances and implications in disease. Curr Drug Targets *13*, 411-420.
Gregersen, P.K., Diamond, B., and Plenge, R.M. (2012). GWAS implicates a role for quantitative immune traits and threshold effects in risk for human autoimmune disorders. Curr Opin Immunol *24*, 538-543.
Grieve, I.C., Dickens, N.J., Pravenec, M., Kren, V., Hubner, N., Cook, S.A., Aitman, T.J., Petretto, E., and Mangion, J. (2008). Genome-wide co-expression analysis in multiple tissues. PLoS One *3*, e4033.
Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Gstaiger, M., and Aebersold, R. (2009). Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nat Rev Genet *10*, 617-627.
Guinea-Viniegra, J., Jimenez, M., Schonthaler, H.B., Navarro, R., Delgado, Y., Concha-Garzon, M.J., Tschachler, E., Obad, S., Dauden, E., and Wagner, E.F. (2014). Targeting miR-21 to treat psoriasis. Sci Transl Med *6*, 225re221.
Guler, M.L., Ligons, D.L., Wang, Y., Bianco, M., Broman, K.W., and Rose, N.R. (2005). Two autoimmune diabetes loci influencing T cell apoptosis control susceptibility to experimental autoimmune myocarditis. J Immunol *174*, 2167-2173.
Gupta, Y., Moller, S., Zillikens, D., Boehncke, W.H., Ibrahim, S.M., and Ludwig, R.J. (2013). Genetic control of psoriasis is relatively distinct from that of metabolic syndrome and coronary artery disease. Exp Dermatol *22*, 552-553.
Gupta, Y., Witte, M., Moller, S., Ludwig, R.J., Restle, T., Zillikens, D., and Ibrahim, S.M. (2014). ptRNApred: computational identification and classification of post-transcriptional RNA. Nucleic Acids Res *42*, e167.
Hertel, J., Hofacker, I.L., and Stadler, P.F. (2008). SnoReport: computational identification of snoRNAs with unknown targets. Bioinformatics *24*, 158-164.
Hertel, J., and Stadler, P.F. (2006). Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. Bioinformatics *22*, e197-202.
Hu, Z., Chang, Y.C., Wang, Y., Huang, C.L., Liu, Y., Tian, F., Granger, B., and Delisi, C. (2013). VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. Nucleic Acids Res *41*, W225-231.
Hua, S., and Sun, Z. (2001). A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. J Mol Biol *308*, 397-407.
Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol *8*, R183.
Iancu, O.D., Darakjian, P., Kawane, S., Bottomly, D., Hitzemann, R., and McWeeney, S. (2012). Detection of expression quantitative trait Loci in complex mouse crosses: impact and alleviation of data quality and complex population substructure. Front Genet *3*, 157.
Ichihara, A., Jinnin, M., Yamane, K., Fujisawa, A., Sakai, K., Masuguchi, S., Fukushima, S., Maruo, K., and Ihn, H. (2011). microRNA-mediated keratinocyte hyperproliferation in psoriasis vulgaris. Br J Dermatol *165*, 1003-1010.
Issabekova, A., Berillo, O., Regnier, M., and Anatoly, I. (2012). Interactions of intergenic microRNAs with mRNAs of genes involved in carcinogenesis. Bioinformation *8*, 513-518.
Jacobson, D.L., Gange, S.J., Rose, N.R., and Graham, N.M. (1997). Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin Immunol Immunopathol *84*, 223-243.
Jankowsky, A., Guenther, U.P., and Jankowsky, E. (2011). The RNA helicase database. Nucleic Acids Res *39*, D338-341.
Jarrous, N., and Reiner, R. (2007). Human RNase P: a tRNA-processing enzyme and transcription factor. Nucleic Acids Res *35*, 3519-3524.
Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X., and Lu, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res *35*, W339-344.

Jiao, Z., Wang, W., Hua, S., Liu, M., Wang, H., Wang, X., Chen, Y., Xu, H., and Lu, L. (2014). Blockade of Notch signaling ameliorates murine collagen-induced arthritis via suppressing Th1 and Th17 cell responses. Am J Pathol *184*, 1085-1093.
Johansson, A., Loset, M., Mundal, S.B., Johnson, M.P., Freed, K.A., Fenstad, M.H., Moses, E.K., Austgulen, R., and Blangero, J. (2011). Partial correlation network analyses to detect altered gene interactions in human disease: using preeclampsia as a model. Hum Genet *129*, 25-34.
John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. PLoS Biol *2*, e363.
Jonkman, M.F., Schuur, J., Dijk, F., Heeres, K., de Jong, M.C., van der Meer, J.B., Yancey, K.B., and Pas, H.H. (2000). Inflammatory variant of epidermolysis bullosa acquisita with IgG autoantibodies against type VII collagen and laminin alpha3. Arch Dermatol *136*, 227-231.
Jung, C.H., Hansen, M.A., Makunin, I.V., Korbie, D.J., and Mattick, J.S. (2010). Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. BMC Genomics *11*, 77.
Kaminski, W., and Strumillo, P. (1997). Kernel orthonormalization in radial basis function neural networks. IEEE Trans Neural Netw *8*, 1177-1183.
Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. Genetics *178*, 1709-1723.
Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol *9*, 770-780.
Katoh, M. (2004). Identification and characterization of ARHGAP24 and ARHGAP25 genes in silico. Int J Mol Med *14*, 333-338.
Kim, T., and Reitmair, A. (2013). Non-Coding RNAs: Functional Aspects and Diagnostic Utility in Oncology. Int J Mol Sci *14*, 4934-4968.
Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.
Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J *20*, 3617-3622.
Kitano, H. (2002a). Computational systems biology. Nature *420*, 206-210.
Kitano, H. (2002b). Systems biology: a brief overview. Science *295*, 1662-1664.
Kollaee, A., Ghaffarpor, M., Ghlichnia, H.A., Ghaffari, S.H., and Zamani, M. (2012). The influence of the HLA-DRB1 and HLA-DQB1 allele heterogeneity on disease risk and severity in Iranian patients with multiple sclerosis. Int J Immunogenet *39*, 414-422.
Koulmanda, M., Bhasin, M., Awdeh, Z., Qipo, A., Fan, Z., Hanidziar, D., Putheti, P., Shi, H., Csizuadia, E., Libermann, T.A*., et al.* (2012). The role of TNF-alpha in mice with type 1- and 2- diabetes. PLoS One *7*, e33254.
Kramer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523-530.
Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M*., et al.* (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.
Krill, K.T., Gurdziel, K., Heaton, J.H., Simon, D.P., and Hammer, G.D. (2013). Dicer deficiency reveals microRNAs predicted to control gene expression in the developing adrenal cortex. Mol Endocrinol *27*, 754-768.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res *19*, 1639-1645.
Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T.H., von Mering, C., Jensen, L.J., and Bork, P. (2014). STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res *42*, D401-407.
Kumarswamy, R., Volkmann, I., and Thum, T. (2011). Regulation and function of miRNA-21 in health and disease. RNA Biol *8*, 706-713.
Kung, H.F., and Huang, J.D. (2001). [The mouse model and human disease]. Zhongguo Yi Xue Ke Xue Yuan Xue Bao *23*, 2-7.
Kurihara, Y., and Watanabe, Y. (2010). Processing of miRNA precursors. Methods Mol Biol *592*, 231-241.
Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res *35*, 3100-3108.
Lala, S., Ogura, Y., Osborne, C., Hor, S.Y., Bromfield, A., Davies, S., Ogunbiyi, O., Nunez, G., and Keshav, S. (2003). Crohn's disease and the NOD2 gene: a role for paneth cells. Gastroenterology *125*, 47-57.
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.
Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719-720.
Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res *32*, 11-16.
Le Cao, K.A., Boitard, S., and Besse, P. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics *12*, 253.
Lee, H.W., Kim, W.Y., Song, H.K., Yang, C.W., Han, K.H., Kwon, H.M., and Kim, J. (2007). Sequential expression of NKCC2, TonEBP, aldose reductase, and urea transporter-A in developing mouse kidney. Am J Physiol Renal Physiol *292*, F269-277.
Lee, I., and Hong, W. (2006). Diverse membrane-associated proteins contain a novel SMP domain. FASEB J *20*, 202-206.
Lee, J.J., and Downham, T.F., 2nd (2006). Furosemide-induced bullous pemphigoid: case report and review of literature. J Drugs Dermatol *5*, 562-564.
Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.
Lev, S., Moreno, H., Martinez, R., Canoll, P., Peles, E., Musacchio, J.M., Plowman, G.D., Rudy, B., and Schlessinger, J. (1995). Protein tyrosine kinase PYK2 involved in Ca(2+)-induced regulation of ion channel and MAP kinase functions. Nature *376*, 737-745.
Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.
Lhermusier, T., van Rottem, J., Garcia, C., Xuereb, J.M., Ragab, A., Martin, V., Gratacap, M.P., Sie, P., and Payrastre, B. (2011). The Syk-kinase inhibitor R406 impairs platelet activation and monocyte tissue factor expression triggered by heparin-PF4 complex directed antibodies. J Thromb Haemost *9*, 2067-2076.
Li, Q., Stram, A., Chen, C., Kar, S., Gayther, S., Pharoah, P., Haiman, C., Stranger, B., Kraft, P., and Freedman, M.L. (2014). Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. Hum Mol Genet *23*, 5294-5302.
Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.

Lin, S.L., Miller, J.D., and Ying, S.Y. (2006). Intronic microRNA (miRNA). J Biomed Biotechnol *2006*, 26818.
Liti, G., and Louis, E.J. (2012). Advances in quantitative trait analysis in yeast. PLoS Genet *8*, e1002912.
Liu, B.H., Yu, H., Tu, K., Li, C., Li, Y.X., and Li, Y.Y. (2010). DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. Bioinformatics *26*, 2637-2638.
Liu, C., Bai, B., Skogerbo, G., Cai, L., Deng, W., Zhang, Y., Bu, D., Zhao, Y., and Chen, R. (2005). NONCODE: an integrated knowledge database of non-coding RNAs. Nucleic Acids Res *33*, D112-115.
Liu, C., Zhang, F., Li, T., Lu, M., Wang, L., Yue, W., and Zhang, D. (2012). MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. BMC Genomics *13*, 661.
Liu, H., and Kohane, I.S. (2009). Tissue and process specific microRNA-mRNA co-expression in mammalian development and malignancy. PLoS One *4*, e5436.
Liu, J., Gough, J., and Rost, B. (2006). Distinguishing protein-coding from non-coding RNAs through support vector machines. PLoS Genet *2*, e29.
Lorenz, A.J., Hamblin, M.T., and Jannink, J.L. (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS One *5*, e14079.
Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26.
Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res *25*, 955-964.
Lowe, T.M., and Eddy, S.R. (1999). A computational screen for methylation guide snoRNAs in yeast. Science *283*, 1168-1171.
Lu, T.P., Lee, C.Y., Tsai, M.H., Chiu, Y.C., Hsiao, C.K., Lai, L.C., and Chuang, E.Y. (2012). miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. PLoS One *7*, e42390.
Lu, Z.J., Yip, K.Y., Wang, G., Shou, C., Hillier, L.W., Khurana, E., Agarwal, A., Auerbach, R., Rozowsky, J., Cheng, C.*, et al.* (2011). Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome Res *21*, 276-285.
Lucas, E.P., Khanal, I., Gaspar, P., Fletcher, G.C., Polesello, C., Tapon, N., and Thompson, B.J. (2013). The Hippo pathway polarizes the actin cytoskeleton during collective migration of Drosophila border cells. J Cell Biol *201*, 875-885.
Ludwig, R.J. (2013). Clinical presentation, pathogenesis, diagnosis, and treatment of epidermolysis bullosa acquisita. ISRN Dermatol *2013*, 812029.
Ludwig, R.J., Muller, S., Marques, A., Recke, A., Schmidt, E., Zillikens, D., Moller, S., and Ibrahim, S.M. (2012). Identification of quantitative trait loci in experimental epidermolysis bullosa acquisita. J Invest Dermatol *132*, 1409-1415.
Lustig, A.J. (1999). Crisis intervention: the role of telomerase. Proc Natl Acad Sci U S A *96*, 3339-3341.
Ma, D., Zhu, Y., Ji, C., and Hou, M. (2010). Targeting the Notch signaling pathway in autoimmune diseases. Expert Opin Ther Targets *14*, 553-565.
Ma, X., Zhou, J., Zhong, Y., Jiang, L., Mu, P., Li, Y., Singh, N., Nagarkatti, M., and Nagarkatti, P. (2014). Expression, Regulation and Function of MicroRNAs in Multiple Sclerosis. Int J Med Sci *11*, 810-818.
Marcet, B., Chevalier, B., Luxardi, G., Coraux, C., Zaragosi, L.E., Cibois, M., Robbe-Sermesant, K., Jolly, T., Cardinaud, B., Moreilhon, C.*, et al.* (2011). Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. Nat Cell Biol *13*, 693-699.

Marschner, S., Freiberg, B.A., Kupfer, A., Hunig, T., and Finkel, T.H. (1999). Ligation of the CD4 receptor induces activation-independent down-regulation of L-selectin. Proc Natl Acad Sci U S A *96*, 9763-9768.
Mathews, D.H., and Turner, D.H. (2006). Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol *16*, 270-278.
Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F.A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics *10*, 213.
Miron B, K., Witold R, Rudnicki (2010). Feature Selection with the Boruta Package. Journal of Statistical Software *36*.
Mishra, D., Dash, R., Rath, A.K., and Acharya, M. (2011). Feature selection in gene expression data using principal component analysis and rough set theory. Adv Exp Med Biol *696*, 91-100.
Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., and Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques *45*, 81-94.
Mott, R., and Flint, J. (2002). Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics *160*, 1609-1618.
Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. Proc Natl Acad Sci U S A *97*, 12649-12654.
Mukherjee, S. (2003). Understanding And Using Microarray Analysis Techniques : A Practical Guide (Boston, MA Kluwer Academic Publishers).
Murata, K., Furu, M., Yoshitomi, H., Ishikawa, M., Shibuya, H., Hashimoto, M., Imura, Y., Fujii, T., Ito, H., Mimori, T.*, et al.* (2013). Comprehensive microRNA analysis identifies miR-24 and miR-125a-5p as plasma biomarkers for rheumatoid arthritis. PLoS One *8*, e69118.
Namjou, B., Sestak, A.L., Armstrong, D.L., Zidovetzki, R., Kelly, J.A., Jacob, N., Ciobanu, V., Kaufman, K.M., Ojwang, J.O., Ziegler, J.*, et al.* (2009). High-density genotyping of STAT4 reveals multiple haplotypic associations with systemic lupus erythematosus in different racial groups. Arthritis Rheum *60*, 1085-1095.
Nariai, N., Kim, S., Imoto, S., and Miyano, S. (2004). Using protein-protein interactions for refining gene networks estimated from microarray data by Bayesian networks. Pac Symp Biocomput, 336-347.
Naul, B. (2009). A review of Support Vector Machines in Computational Biology. Protein Similarities and Homologies.
Nica, A.C., and Dermitzakis, E.T. (2013). Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci *368*, 20120362.
Niedermeier, A., Eming, R., Pfutze, M., Neumann, C.R., Happel, C., Reich, K., and Hertl, M. (2007). Clinical response of severe mechanobullous epidermolysis bullosa acquisita to combined treatment with immunoadsorption and rituximab (anti-CD20 monoclonal antibodies). Arch Dermatol *143*, 192-198.
O'Sullivan, B., Thompson, A., and Thomas, R. (2007). NF-kappa B as a therapeutic target in autoimmune disease. Expert Opin Ther Targets *11*, 111-122.
Okuma, A., Hoshino, K., Ohba, T., Fukushi, S., Aiba, S., Akira, S., Ono, M., Kaisho, T., and Muta, T. (2013). Enhanced apoptosis by disruption of the STAT3-IkappaB-zeta signaling pathway in epithelial cells induces Sjogren's syndrome-like autoimmune disease. Immunity *38*, 450-460.
Ooi, H.S., Schneider, G., Chan, Y.L., Lim, T.T., Eisenhaber, B., and Eisenhaber, F. (2010). Databases of protein-protein interactions and complexes. Methods Mol Biol *609*, 145-159.

Opitz D., M.R. (1999). Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 169-198.
Otaegui, D., Baranzini, S.E., Armananzas, R., Calvo, B., Munoz-Culla, M., Khankhanian, P., Inza, I., Lozano, J.A., Castillo-Trivino, T., Asensio, A.*, et al.* (2009). Differential micro RNA expression in PBMC from multiple sclerosis patients. PLoS One *4*, e6309.
Pannucci, J.A., Haas, E.S., Hall, T.A., Harris, J.K., and Brown, J.W. (1999). RNase P RNAs from some Archaea are catalytically active. Proc Natl Acad Sci U S A *96*, 7803-7808.
Paraboschi, E.M., Solda, G., Gemmati, D., Orioli, E., Zeri, G., Benedetti, M.D., Salviati, A., Barizzone, N., Leone, M., Duga, S.*, et al.* (2011). Genetic association and altered gene expression of mir-155 in multiple sclerosis patients. Int J Mol Sci *12*, 8695-8712.
Parkin, J., and Cohen, B. (2001). An overview of the immune system. Lancet *357*, 1777-1789.
Pattin, K.A., and Moore, J.H. (2009). Role for protein-protein interaction databases in human genetics. Expert Rev Proteomics *6*, 647-659.
Penna, E., Orso, F., and Taverna, D. (2014). miR-214 as a Key Hub that Controls Cancer Networks: Small Player, Multiple Functions. J Invest Dermatol.
Perreault, J., Perreault, J.P., and Boire, G. (2007). Ro-associated Y RNAs in metazoans: evolution and diversification. Mol Biol Evol *24*, 1678-1689.
Pihur, V., and Datta, S. (2008). Reconstruction of genetic association networks from microarray data: a partial least squares approach. Bioinformatics *24*, 561-568.
Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y., Goumnerova, L.C., Black, P.M., Lau, C.*, et al.* (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature *415*, 436-442.
Raghavachari, B., Tasneem, A., Przytycka, T.M., and Jothi, R. (2008). DOMINE: a database of protein domain interactions. Nucleic Acids Res *36*, D656-661.
Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P.*, et al.* (2001). Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A *98*, 15149-15154.
Reverter, A., and Chan, E.K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics *24*, 2491-2497.
Richard-Miceli, C., and Criswell, L.A. (2012). Emerging patterns of genetic overlap across autoimmune disorders. Genome Med *4*, 6.
Rioux, J.D., and Abbas, A.K. (2005). Paths to understanding the genetic basis of autoimmune disease. Nature *435*, 584-589.
Rottiers, V., and Naar, A.M. (2012). MicroRNAs in metabolism and metabolic disorders. Nat Rev Mol Cell Biol *13*, 239-250.
SantaLucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A *95*, 1460-1465.
Scafoglio, C., Ambrosino, C., Cicatiello, L., Altucci, L., Ardovino, M., Bontempo, P., Medici, N., Molinari, A.M., Nebbioso, A., Facchiano, A.*, et al.* (2006). Comparative gene expression profiling reveals partially overlapping but distinct genomic actions of different antiestrogens in human breast cancer cells. J Cell Biochem *98*, 1163-1184.
Schmidt, M.F. (2014). Drug target miRNAs: chances and challenges. Trends Biotechnol.
Schmitt, W.A., Jr., Raab, R.M., and Stephanopoulos, G. (2004). Elucidation of gene interaction networks through time-lagged correlation analysis of transcriptional data. Genome Res *14*, 1654-1663.

Schneider, M.R. (2012). MicroRNAs as novel players in skin development, homeostasis and disease. Br J Dermatol *166*, 22-28.
Schwarz, D.F., Konig, I.R., and Ziegler, A. (2010). On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. Bioinformatics *26*, 1752-1758.
Sesarman, A., Sitaru, A.G., Olaru, F., Zillikens, D., and Sitaru, C. (2008). Neonatal Fc receptor deficiency protects from tissue injury in experimental epidermolysis bullosa acquisita. J Mol Med (Berl) *86*, 951-959.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S*., et al.* (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med *8*, 68-74.
Simmons, A.D., Musy, M.M., Lopes, C.S., Hwang, L.Y., Yang, Y.P., and Lovett, M. (1999). A direct interaction between EXT proteins and glycosyltransferases is defective in hereditary multiple exostoses. Hum Mol Genet *8*, 2155-2164.
Sonkoly, E., Janson, P., Majuri, M.L., Savinko, T., Fyhrquist, N., Eidsmo, L., Xu, N., Meisgen, F., Wei, T., Bradley, M*., et al.* (2010). MiR-155 is overexpressed in patients with atopic dermatitis and modulates T-cell proliferative responses by targeting cytotoxic T lymphocyte-associated antigen 4. J Allergy Clin Immunol *126*, 581-589 e581-520.
Srinivas, G., Moller, S., Wang, J., Kunzel, S., Zillikens, D., Baines, J.F., and Ibrahim, S.M. (2013). Genome-wide mapping of gene-microbiota interactions in susceptibility to autoimmune skin blistering. Nat Commun *4*, 2462.
Statnikov, A., and Aliferis, C.F. (2007). Are random forests better than support vector machines for microarray-based cancer classification? AMIA Annu Symp Proc, 686-690.
Sticherling, M., and Erfurt-Berge, C. (2012). Autoimmune blistering diseases of the skin. Autoimmun Rev *11*, 226-230.
Studham, M.E., Tjarnberg, A., Nordling, T.E., Nelander, S., and Sonnhammer, E.L. (2014). Functional association networks as priors for gene regulatory network inference. Bioinformatics *30*, i130-138.
Su, W.L., Kleinhanz, R.R., and Schadt, E.E. (2011). Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. Mol Syst Biol *7*, 490.
Sun, Y.V. (2010). Multigenic modeling of complex disease by random forests. Adv Genet *72*, 73-99.
Thessen Hedreul, M., Moller, S., Stridh, P., Gupta, Y., Gillett, A., Daniel Beyeen, A., Ockinger, J., Flytzani, S., Diez, M., Olsson, T*., et al.* (2013). Combining genetic mapping with genome-wide expression in experimental autoimmune encephalomyelitis highlights a gene network enriched for T cell functions and candidate genes regulating autoimmunity. Hum Mol Genet *22*, 4952-4966.
Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003). Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J Biol Chem *278*, 1239-1247.
Trachtenberg, E.A., Yang, H., Hayes, E., Vinson, M., Lin, C., Targan, S.R., Tyan, D., Erlich, H., and Rotter, J.I. (2000). HLA class II haplotype associations with inflammatory bowel disease in Jewish (Ashkenazi) and non-Jewish caucasian populations. Hum Immunol *61*, 326-333.
Tsokos, G.C., and Fleming, S.D. (2004). Autoimmunity, complement activation, tissue injury and reciprocal effects. Curr Dir Autoimmun *7*, 149-164.

Tyc, K., and Steitz, J.A. (1989). U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. EMBO J *8*, 3113-3119.
Vagnoni, A., Perkinton, M.S., Gray, E.H., Francis, P.T., Noble, W., and Miller, C.C. (2012). Calsyntenin-1 mediates axonal transport of the amyloid precursor protein and regulates Abeta production. Hum Mol Genet *21*, 2845-2854.
van der Sijde, M.R., Ng, A., and Fu, J. (2014). Systems genetics: From GWAS to disease pathways. Biochim Biophys Acta *1842*, 1903-1909.
van Someren, E.P., Wessels, L.F., Backer, E., and Reinders, M.J. (2002). Genetic network modeling. Pharmacogenomics *3*, 507-525.
von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. Nucleic Acids Res *31*, 258-261.
Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nat Rev Genet *11*, 843-854.
Wang, Y., and Blelloch, R. (2011). Cell cycle regulation by microRNAs in stem cells. Results Probl Cell Differ *53*, 459-472.
Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.
Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L.*, et al.* (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res *42*, D1001-1006.
Wong, B.R., Grossbard, E.B., Payan, D.G., and Masuda, E.S. (2004). Targeting Syk as a treatment for allergic and autoimmune disorders. Expert Opin Investig Drugs *13*, 743-762.
Xenarios, I., and Eisenberg, D. (2001). Protein interaction databases. Curr Opin Biotechnol *12*, 334-339.
Xu, W.D., Pan, H.F., Li, J.H., and Ye, D.Q. (2013). MicroRNA-21 with therapeutic potential in autoimmune diseases. Expert Opin Ther Targets *17*, 659-665.
Yang, Z.R. (2004). Biological applications of support vector machines. Brief Bioinform *5*, 328-338.
Yao, X., Huang, J., Zhong, H., Shen, N., Faggioni, R., Fung, M., and Yao, Y. (2014). Targeting interleukin-6 in inflammatory autoimmune diseases and cancers. Pharmacol Ther *141*, 125-139.
Yoon, S., and De Micheli, G. (2005). Prediction and Analysis of Human microRNA Regulatory Modules. Conf Proc IEEE Eng Med Biol Soc *5*, 4799-4802.
Young, N.D. (1996). QTL mapping and quantitative disease resistance in plants. Annu Rev Phytopathol *34*, 479-501.
Zabel, B.A., Nakae, S., Zuniga, L., Kim, J.Y., Ohyama, T., Alt, C., Pan, J., Suto, H., Soler, D., Allen, S.J.*, et al.* (2008). Mast cell-expressed orphan receptor CCRL2 binds chemerin and is required for optimal induction of IgE-mediated passive cutaneous anaphylaxis. J Exp Med *205*, 2207-2220.
Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., and Cesareni, G. (2002). MINT: a Molecular INTeraction database. FEBS Lett *513*, 135-140.
Zhang, S.W., Li, Y.J., Xia, L., and Pan, Q. (2010). PPLook: an automated data mining tool for protein-protein interaction. BMC Bioinformatics *11*, 326.
Zhao, J., Jitkaew, S., Cai, Z., Choksi, S., Li, Q., Luo, J., and Liu, Z.G. (2012). Mixed lineage kinase domain-like is a key receptor interacting protein 3 downstream component of TNF-induced necrosis. Proc Natl Acad Sci U S A *109*, 5322-5327.
Ziegler Andreas, K.I.R. (2014). Mining data with random forests: current options for real-world applications. WIREs Data Mining Knowl Discov, 8.

Zien, A., Ratsch, G., Mika, S., Scholkopf, B., Lengauer, T., and Muller, K.R. (2000). Engineering support vector machine kernels that recognize translation initiation sites. Bioinformatics *16*, 799-807.
Zillikens, D. (2008). Diagnosis of autoimmune bullous skin diseases. Clin Lab *54*, 491-503.
Zou, M., and Conzen, S.D. (2005). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics *21*, 71-79.

10. Appendix

10.1 List of abbreviations

AAbs	Autoimmune antibodies
ABDs	Autoimmune bullous disease
AIL	Advanced intercross line
Biogrid	Biological general repository for interaction datasets
BM	Basement membrane
cDNA	Complementary DNA
Chr	Chromosome
CV	Cross-validation
DEJ	Dermal-epidermal junction
DNA	Deoxyribonucleic acid
DOMINE	Database of protein-domain interactions
EBA	Epidermolysis bullosa acquisita
EMMA	Efficient mixed-model association
eQTL	Expression quantitative trait loci
G4	Generation 4
GST	Glutathione S-transferase
GWAS	Genome-wide association studies
HLA	Human leukocyte antigen
IPA	Ingenuity pathway analysis
LibSVM	Library of support vector machines
MINT	Molecular interaction database
miRNA	Micro RNA
MS	Multiple sclerosis
ncRNA	Non-coding ribonucleic acid
OOB	Out of bag
PCA	Principle component analysis
PDB	Protein data bank
PLS	Partial least squares
PPIs	Protein-protein interactions
ptRNApred	Post transcriptional RNA prediction
QTL	Quantitative trait loci
RA	Rheumatoid arthritis
RBF	Radial basis function kernel

RMA	Robust multi-array analysis
RNA	Ribonucleic acid
MRP	Ribonuclease MRP
RNase P	Ribonuclease P
SLE	Systemic lupus erythematosus
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
STRING	Search tool for the retrieval of interacting genes
SVM	Support vector machines
tRNAs	Transfer RNA
WGCNA	Weighted gene co-expression network analysis

10.2 List of figures

10.3 List of tables

11. Acknowledgements

I would like to acknowledge all the people who have supported me during my PhD work. First, I would like to share my gratitude towards Prof. SaIeh Ibrahim, director of the Institute of Experimental Dermatology, for hosting my thesis and giving me opportunity to work with tremendous supervision throughout my working period. His guidance, appreciation, constructive criticism, motivation and knowledge helped me to pursue my work in the appropriate direction.

I would like to show my sincerest appreciation to Prof. Detlef Zillikens, director of the Department of Dermatology, and Prof. Ralf Ludwig for scientific discussion and support.

I would also like to thank Dr. Steffen Möller for guiding me in every stage as my mentor and for fruitful scientific discussion and input.

Moreover, I would also thank Mareike Witte, Tanya Sezin, Katrina Vafia, Laura Mellado, Unni Samadevan, Siegfried Bezdek, Julia Bischof, Axel Künstner, Dr. Misa Hirose, Dr. Girish Srinivas and Paul Schiff for providing me with biological insights, innovative scientific ideas, moral support and reviewing my manuscript and thesis.

I would also like to share my gratitude to my parents, family and my friends Tanya Sezin, Ahmet Bicer Emmerich, Wladmir Helbrecht, Katrina Vafia, Dr. Misa Hirose, Ashraf Hadnah, and Siegfried Bezdek for providing me with moral support in my good and bad days.

12. Curriculum vitae

Personal data

Family name:	Gupta
Given name:	Yask
Date of birth:	19/01/1987
Gender:	Male
Nationality:	Indian
Telephone:	+49 (0) 500 2433
E-mail:	Yask.Gupta@uksh.de yask.gupta87@gmail.com

Academic background

X C.B.S.E. (2000-2002)	Delhi Police Public School, Safdarjung Enclave New Delhi, India
XII C.B.S.E (2003-2004)	Delhi Police Public School, Safdarjung Enclave New Delhi, India
Graduation (2004 – 2007)	B.Sc. (H) Botany A.N.D.College, Kalkaji, Delhi University, New Delhi, India
Post Graduation (2007-2009)	M.Sc. (Bioinformatics) Jamia Millia Islamia, New Delhi, India

Research Background

8/2009 – 10/2010	International Center for Genetic Engineering and Biotechnology, New Delhi, India
Since 11/2010	Department of Dermatology, University of Lübeck, Germany

Publications

First Author

1. Gupta Y, Witte M, Möller S, Ludwig RJ, Restle T, Zillikens D and Ibrahim SM. ptRNApred: Computational identification and classification of post-transcriptional RNA. **Nucleic Acid Research,** PMID: 25303994.

2. Gupta Y, Möller S, Zillikens D, Boehncke WH, Ibrahim SM, Ludwig RJ. Genetic control of psoriasis is relatively distinct from that of metabolic syndrome and coronary artery disease. Exp Dermatol. 2013 Aug; 22(8):552-3.PMID: 23879815.

3. Gupta Y et al. Systems genetics of miRNA in skin autoimmune skin blistering disease (Submitted BMC Genomics).

Co-Author

4. Kasperkiewicz M, Nimmerjahn F, Wende S, Hirose M, Iwata H, Jonkman MF,Samavedam U, **Gupta Y**, Möller S, Rentz E, Hellberg L, Kalies K, Yu X, Schmidt E,Häsler R, Laskay T, Westermann J, Köhl J, Zillikens D, Ludwig RJ. Genetic identification and functional validation of FcγRIV as key molecule in autoantibody-induced tissue injury. J Pathol. 2012 Sep. PMID: 22430937.

5. Samavedam UK, Kalies K, Scheller J, Sadeghi H, **Gupta Y**, Jonkman MF, Schmidt E,Westermann J, Zillikens D, Rose-John S, Ludwig RJ. Recombinant IL-6 treatment protects mice from organ specific autoimmune disease by IL-6 classical signalling-

dependent IL-1ra induction. J Autoimmun. 2013 Feb; PMID: 22980031.

6. Thessen Hedreul M, Möller S, Stridh P, **Gupta Y**, Gillett A, Daniel Beyeen A,Öckinger J, Flytzani S, Diez M, Olsson T, Jagodic M. Combining genetic mapping with genome-wide expression in experimental autoimmune encephalomyelitis highlights a gene network enriched for T cell functions and candidate genes regulating autoimmunity. Hum Mol Genet. 2013 Dec 15; PMID: 23900079

7. Ranea LM, de Castro Marques A, Möller S, **Gupta Y**, Ibrahim SM. Genetic control of spontaneous arthritis in a four-way advanced intercross line. PLoS One, 2013 Oct 11. PMID: 24146764

8. Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, Farber M,Gahlot S, Gamble J, Gupta D, **Gupta Y**, Jackson L, Malandrin L, Malas TB, Moussa E,Nair M, Reid AJ, Sanders M, Sharma J, Tracey A, Quail MA, Weir W, Wastling JM,Hall N, Willadsen P, Lingelbach K, Shiels B, Tait A, Berriman M, Allred DR, Pain A. The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. Nucleic Acids Res. PMID: 24799432.

13. Supplements

Section S1: Features for classification.

Feature selection and SVM training was performed using two sets of input parameters: The first set is based on the primary sequence and the second set considers the secondary structure which is predicted with RNAfold. All training steps were automated by a Perl script.

Set 1: Dinucleotide properties

Each sequence was divided into its dinucleotides, using the sliding window approach (window size: 2 nucleotides). In total, 16 different dinucleotides are possible: AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC and GG. Each of the 16 dinucleotides can be assigned distinct properties, including thermodynamic features (e.g. stacking energy, free energy), structural features (e.g. twist, roll) and sequences based features. These features have been described in previous experimental or computational work. The dinucleotide property database (DiProDB) contains information on dinucleotides and a collection of more than 100 published dinucleotide property sets. In order to determine whether different ptRNA-subclasses can be distinguished via specific dinucleotide properties, 125 dinucleotide properties were abstracted from DiProDB and individually correlated with every

ptRNA-subclass. Properties were clustered and a representative property was selected from each of the 16 resulting clusters (Supplement Table 1).

Set 2: Secondary structure properties

Secondary structures of every sequence were calculated by RNAfold, accessing the Vienna RNA Package. RNAfold provides structures according to different parameters. Various properties were derived from the RNAfold output:

a. The Minimum free energy (MFE) structure

The MFE structure of a RNA sequence is a secondary structure that contributes a minimum of free energy. For MFE structure prediction, RNAfold uses a loop-based energy model and the dynamic programming algorithm introduced. As RNA secondary structure can be uniquely decomposed into loops and external bases, the loop-based energy model treats the free energy of an RNA secondary structure as the sum of the contributing free energies of the loops contained in the secondary structure. According to the chosen energy parameter set and a given temperature (defaults to 37 °C) a secondary structure with minimal free energy is computed. The minimum free energy was selected as property in our SVM. Additional features were deducted from the MFE structure, which is denoted by brackets '(or)' and dots '.' Brackets indicate paired nucleotides, whereas dots represent unpaired nucleotides. The left bracket '(' means the paired nucleotide is located near the 5′-end and can be

paired with another nucleotide at the 3′-end, which is indicated as a right bracket ‘)’. In our script, we do not distinguish these two situations and use ‘(’ for both situations. Brackets and dots were counted within each sequence, yielding two additional properties.

Different features were selected combining secondary structure and primary sequence. In this context, the number of bulges and hairpins were counted, as well as the four nucleotides A, G, C and U in every bulge and every hairpin, yielding ten additional properties. Furthermore, purine and pyrimidine contents were examined and the number of mismatches was determined. Additionally, paired bases were considered alongside, counting AU, CG and GU pairs. All in all, this section of sequence examination yields 18 properties.

Further information was gained for every three adjacent nucleotides, which we call triplet elements for the convenience of the discussion. Eight additional properties were given by counting of the eight possible triplet element compositions ‘(((’, ‘((.’, ‘(..’, ‘…’, ‘.((’, ‘..(’, ‘.(.’ and ‘(.(’ within every sequence.

The nucleotide composition of the triplet elements was not regarded and the compositions were counted using the sliding window approach.

32 further triplet element properties were derived from miPred, a triplet SVM for the classification of miRNA. MiPred considers the middle nucleotide among the triplet elements, resulting in 32 (4 × 8) possible combinations, which are denoted as

‘U(((’, ‘A((.’, etc. The number of appearance of each triplet element is counted for each hairpin to produce the 32-dimensional feature vector and used as input features for SVM.

b. The ensemble free energy

RNAfold provides an ensemble structure, considering probabilities of the presence of certain base pairs. Bases with a strong preference (more than 2/3) to pair upstream (with a partner further 3'), pair downstream or not pair are represented by the usual symbols '(',')' or '.'. Additional symbols '{','}’ or ',' reflect bases with a weaker preference and thus are a weaker version of the above respectively. '|' represents a base that is mostly paired but has pairing partners both upstream and downstream. In this case open and closed brackets do not need to match up. This pseudo bracket notation is followed by the ensemble free energy. The numbers of '{', '}’ and ',' as well as the ensemble free energy were taken as features for our SVM.

c. The centroid structure

RNAfold further provides a centroid structure that is given by its secondary structure with minimal base pair distance to all other secondary structures in the Boltzmann ensemble (5). The values of the centroid structure’s free energy as well as its distance to the ensemble were taken as features for our SVM.

d. The maximum expected accuracy (MEA) structure

RNAfold further outputs a MEA structure, in which each base pair (i,j) gets a score 2*gamma*p_ij and the score of an unpaired base is given by the probability of not forming a pair. Subsequently, the expected accuracy is computed from the pair probabilities. The MEA as well as the MEA structure's free energy serve as additional features for our SVM.

e. The frequency of the MFE representative in the complete ensemble of secondary structures and the ensemble diversity

Two additional features are given by the frequency of the MFE representative in the complete ensemble of secondary structures and the ensemble diversity.

Supplement Table 1: Table of selected dinucleotide properties. The table shows the dinucleotide properties selected as vectors for the SVM. 15 distinct properties (left column), ranging from the shift score to the entropy, are assigned to the16 possible dinucleotides: AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC and GG. All information was derived from DiProGB (1). Further information is provided in Section S1.

Property Name	AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
Shift[1]	-0.08	0.23	-0.04	-0.06	0.11	-0.01	0.3	-0.04	0.07	0.07	-0.01	0.23	-0.02	0.07	0.11	-0.08
Hydrophilicity[2]	0.023	0.083	0.035	0.09	0.118	0.349	0.193	0.378	0.048	0.146	0.065	0.16	0.112	0.359	0.224	0.389
GC_content[3]	0	1	1	0	1	2	2	1	1	2	2	1	0	1	1	0
Keto_content[4]	0	0	0	1	0	0	1	1	1	1	2	2	1	1	1	2
Adenine_content[5]	2	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0
Guanine_content[6]	0	0	1	0	0	0	1	0	1	1	2	1	0	0	1	0
Cytosine_content[7]	0	1	0	0	1	2	1	1	0	1	0	0	0	1	0	0
Slide[1]	-1.27	-1.43	-1.5	-1.36	-1.46	-1.78	-1.89	-1.5	-1.7	-1.39	-1.78	-1.43	-1.45	-1.7	-1.46	-1.27
Rise[1]	3.18	3.24	3.3	3.24	3.09	3.32	3.3	3.3	3.38	3.22	3.32	3.24	3.26	3.38	3.09	3.18
Tilt[1]	-0.8	0.8	0.5	1.1	1	0.3	-0.1	0.5	1.3	0	0.3	0.8	-0.2	1.3	1	-0.8

Roll[1]	7	4.8	8.5	7.1	9.9	8.7	12.1	8.5	9.4	6.1	12.1	4.8	10.7	9.4	9.9	7
Twist[1]	31	32	30	33	31	32	27	30	32	35	32	32	32	32	31	31
Stacking_energy[8]	-13.7	-13.8	-14	-15.4	-14.4	-11.1	-15.6	-14	-14.2	-16.9	-11.1	-13.8	-16	-14.2	-14.4	-13.7
Entropy_1[9]	-18.4	-26.2	-19.2	-15.5	-27.8	-29.7	-19.4	-19.2	-35.5	-34.9	-29.7	-26.2	-22.6	-26.2	-19.2	-18.4
Entropy_2[10]	-19	-29.5	-27.1	-26.7	-26.9	-32.7	-26.7	-27.1	-32.5	-36.9	-32.7	-29.5	-20.5	-32.5	-26.9	-19

1. Perez, A. The relative flexibility of B-DNA and A-RNA duplexes: database analysis. Nucleic Acids Res. 32, 6144–6151 (2004).

2. Weber, A. L. & Lacey, J. C., Jr. Genetic code correlations: amino acids and their anticodon nucleotides. J. Mol. Evol. 11, 199–210 (1978).

3. Friedel, M. Each C or G counts +1.

4. Friedel, M. G and T (U) counts +1.

5. Friedel, M. Each A counts +1.

6. Friedel, M. Each G counts +1.

7. Friedel, M. Each C counts +1.

8. Encyclopedia of Life Sciences. (John Wiley & Sons, Ltd, 2001).

9. Freier, S. M. et al. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. U. S. A. 83, 9373–9377 (1986).

10. Xia, T. et al. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry (Mosc.) 37, 14719–14735 (1998).

Supplement Table 2: Gini index for feature selection. The table provides Gini-index for each features which relates to the importance of each feature to differentiate between classes of non-coding RNAs.

Property number	Property description	F-Score	Gini-Index
1	Adenine_content[1]	0,064268	2,41E-44
2	Cytosine_content[1]	0,083006	6,03E-15
3	Entropy_1[1]	0,088785	4,34E-19
4	Entropy_2[1]	0,09256	5,93E-21
5	GC_content[1]	0,106083	5,43E-11
6	Guanine_content[1]	0,12987	2,75E-18
7	Hydrophilicity[1]	0,098519	4,37E-23
8	Keto_content[1]	0,131688	4,12E-16
9	Rise[1]	0,091877	8,37E-20
10	Roll[1]	0,098188	6,50E-26
11	Shift[1]	0,057281	5,62E-49
12	Slide[1]	0,09489	4,93E-19
13	Stacking_energy[1]	0,088663	1,80E-30
14	Tilt[1]	0,060113	6,03E-20
15	Twist[1]	0,090472	3,98E-27
16	value_in_3rd_rnafold[2]	0,039964	2,48E-51
17	count_star_bracket_{[2]	0,032699	0,002720791
18	count_comma_in_rnafold[2]	0,032914	0,000482739
19	value_line_no_3_RNAfold[2]	0,037566	2,54E-21
20	value_line_no_3_RNAfold(second value)[2]	0,034666	7,25E-18
21	value_line_number_4[2]	0,037892	1,68E-18
22	value_line_number_4(second value) [2]	0,034268	6,11E-32
23	frequency_of_mfe_structure_in_ensemble[2]	0,035204	1,12E-16
24	ensemble diversity[2]	0,039227	1,21E-11
25	value_MFE_RNAfold[2]	0,038989	1,18E-34
26	Number_of_loops[2]	0,03857	1,12E-12
27	(((2	0,032912	3,18E-29
28	((.2	0,037534	7,90E-11
29	(..2	0,039972	3,42E-07
30	...2	0,029854	7,49E-10
31	.((2	0,038714	4,07E-11
32	..(2	0,040519	0,000102735
33	.(.2	0,023965	0,996225164

34	(.(^{2}	0,031401	0,003872273
35	A(((2	0,003259	5,34E-15
36	A((.2	0,003767	1,49E-06
37	A(.(2	0,00501	0,037528552
38	A(..2	0,014225	2,74E-05
39	A.((2	0,001732	4,20E-06
40	A.(.2	0,001482	0,119541752
41	A..(2	0,015074	2,38E-12
42	A...2	0,029134	5,04E-12
43	U(((2	0,003228	1,41E-09
44	U((.2	0,00721	2,26E-05
45	U(.(2	0,00737	0,000196773
46	U(..2	0,010258	4,59E-09
47	U.((2	0,002741	0,00024243
48	U.(.2	0,000326	0,996205638
49	U..(2	0,017021	1,90E-21
50	U...2	0,093364	3,83E-43
51	G(((2	0,021702	3,72E-27
52	G((.2	0,024863	0,000494941
53	G(.(2	0,005776	3,58E-08
54	G(..2	0,010109	6,84E-14
55	G.((2	0,003169	0,000144641
56	G.(.2	0,001119	0,403681085
57	G..(2	0,010458	2,26E-06
58	G...2	0,006942	1,12E-30
59	C(((2	0,044132	4,38E-07
60	C((.2	0,009389	2,98E-09
61	C(.(2	0,001738	0,98836726
62	C(..2	0,002454	1,76E-08
63	C.((2	0,016575	0,017213739
64	C.(.2	0,000654	0,000405832
65	C..(2	0,006828	0,003771578
66	C...2	0,019666	2,19E-12
67	number_of_AU^2	0,018431	1,55E-20
68	number_of_CG^2	0,008113	1,63E-10
69	number_of_GU^2	0,005051	4,01E-09
70	number_of_mistatches_in_sec_struc^2	0,036243	1,48E-12
71	number_of_buldges_in_sec_struc^2	0,06064	4,57E-09
72	A_in_buldges^2	0,059369	2,94E-13

73	G_in_buldges[2]	0,064056	0,168681863
74	C_in_buldges[2]	0,037917	2,02E-05
75	U_in_buldges[2]	0,035724	1,48E-09
76	length_of_hairpin[2]	0,029	4,29E-12
77	number_of_sub_sec_structure[2]	0,025913	1,59E-11
78	number_of_A_hairpin[2]	0,026994	8,85E-09
79	number_of_G_hairpin[2]	0,010075	0,021895501
80	number_of_C_hairpin[2]	0,027103	0,006452735
81	number_of_U_hairpin[2]	0,066878	1,63E-12
82	number_of_A_purine[2]	0,021688	6,41E-07
83	number_of_A_pyramidine[2]	0,055579	1,70E-14
84	number_of_A_in_first_complementary_strand[2]	0,002319	1,32E-16
85	number_of_G_in_first_complementary_strand[2]	0,005858	4,21E-14
86	number_of_C_in_first_complementary_strand[2]	0,001151	1,07E-07
87	number_of_U_in_first_complementary_strand[2]	0,004656	1,13E-34
88	number_of_A_in_second_complementary_strand[2]	0,003052	2,39E-14
89	number_of_G_in_second_complementary_strand[2]	0,003099	0,00106611
90	number_of_C_in_second_complementary_strand[2]	0,012551	2,20E-08
91	number_of_U_in_second_complementary_strand[2]	0,035877	9,40E-16

Supplement Table 3: List of differentially expressed genes in EBA and non EBA mouse.

ID	Gene Name	logFC	*p-value*	Adjusted *p-value*
10380419	Col1a1	1.092457892	1.71E-11	2.97E-07
10529457	Cpz	0.476105771	1.83E-11	2.97E-07
10595211	Col12a1	0.937846131	3.26E-11	2.97E-07
10346015	Col3a1	0.986338829	3.42E-11	2.97E-07
10531724	Plac8	1.014783794	1.21E-10	8.43E-07
10583056	Mmp12	0.831790171	2.46E-10	1.20E-06
10536220	Col1a2	0.933247962	2.64E-10	1.20E-06
10460782	Gpha2	-0.588016109	2.77E-10	1.20E-06
10560919	Atp1a3	0.485810749	3.51E-10	1.36E-06
10560685	Bcl3	0.577733423	7.41E-10	2.58E-06
10354309	Col5a2	0.496367606	1.15E-09	3.39E-06
10546450	Adamts9	0.501134973	1.17E-09	3.39E-06
10556082	Ppfibp2	-0.28734775	1.29E-09	3.44E-06

10379636	Slfn4	1.289214032	1.60E-09	3.96E-06
10572949	Nr3c2	-0.319316147	1.90E-09	4.41E-06
10367400	Mmp19	0.798947908	2.12E-09	4.62E-06
10352143	Kif26b	0.383798765	2.39E-09	4.75E-06
10557895	Itgax	0.490973272	2.46E-09	4.75E-06
10403743	Inhba	0.618346825	3.67E-09	6.63E-06
10578045	Nrg1	0.58092562	3.82E-09	6.63E-06
10523128	Ppbp	0.790749348	4.69E-09	7.77E-06
10400126	Lrrn3	-0.240696539	5.18E-09	8.18E-06
10513739	Tnc	1.169230251	5.81E-09	8.57E-06
10478633	Mmp9	1.123929228	5.92E-09	8.57E-06
10393449	Socs3	0.639354861	6.96E-09	9.68E-06
10470462	Col5a1	0.323101365	7.78E-09	1.04E-05
10360040	Fcgr3	0.693844879	8.33E-09	1.07E-05
10351182	Sele	0.640160808	1.03E-08	1.28E-05
10494262	Ctsk	0.919182273	1.07E-08	1.28E-05
10407286	BC067074 /// LOC101055997	-0.276250362	1.22E-08	1.41E-05
10534202	Ncf1	0.450832781	1.30E-08	1.45E-05
10466735	Fam189a2	-0.361487314	1.33E-08	1.45E-05
10570434	Ifitm1 /// Gm7676	1.051091959	1.51E-08	1.51E-05
10461558	Slc15a3	0.912327937	1.51E-08	1.51E-05
10368277	Rps12 /// Snora33	0.502620387	1.52E-08	1.51E-05
10421309	Slc39a14	0.401603199	1.56E-08	1.51E-05
10490923	Car2	0.897349277	1.73E-08	1.59E-05
10518147	Pdpn	0.706520792	1.74E-08	1.59E-05
10541614	Clec4d	1.334967414	1.86E-08	1.62E-05
10413726	Tnnc1	-0.706765141	1.86E-08	1.62E-05
10351551	Adamts4	0.497613129	1.98E-08	1.65E-05
10426425	Pdzrn4	-0.326182757	2.00E-08	1.65E-05
10598976	Timp1	0.794947804	2.15E-08	1.71E-05
10364262	Itgb2	0.753250668	2.18E-08	1.71E-05
10546454	Adamts9	0.411697522	2.21E-08	1.71E-05
10518300	Tnfrsf1b	0.546329087	2.26E-08	1.71E-05
10504775	Col15a1	0.408820674	2.32E-08	1.72E-05
10601729	Drp2	-0.481376088	2.39E-08	1.73E-05
10436100	Retnlg	1.483732955	2.56E-08	1.81E-05
10392815	AF251705	0.44425203	2.85E-08	1.98E-05
10546434	Adamts9	0.319294623	2.93E-08	2.00E-05
10425053	Ncf4	0.564342209	3.01E-08	2.00E-05

10505489	Pappa	0.41352215	3.06E-08	2.00E-05
10387909	Chrne	-0.466539878	3.18E-08	2.05E-05
10551009	---	0.441853258	3.58E-08	2.26E-05
10375307	C1qtnf2	0.274185976	3.93E-08	2.35E-05
10498024	Slc7a11	1.288478178	4.01E-08	2.35E-05
10472050	Tnfaip6	0.778311584	4.04E-08	2.35E-05
10460585	Fosl1	0.492664814	4.08E-08	2.35E-05
10432767	Gm5478	0.508658887	4.08E-08	2.35E-05
10432852	Krt1	1.0092609	4.13E-08	2.35E-05
10396831	Arg2	0.965296433	5.10E-08	2.77E-05
10490159	Pmepa1	0.399646146	5.30E-08	2.77E-05
10469786	Il1f9	1.070296747	5.33E-08	2.77E-05
10420362	Gjb2	0.783128927	5.45E-08	2.77E-05
10553042	Rasip1 /// Izumo1	0.368560716	5.46E-08	2.77E-05
10391066	Krt17	0.695176974	5.53E-08	2.77E-05
10398665	Tnfaip2	0.650135568	5.56E-08	2.77E-05
10347291	Cxcr2	1.344378458	5.56E-08	2.77E-05
10406334	Mctp1	0.552234795	5.59E-08	2.77E-05
10525256	Tmem116	-0.263637327	5.95E-08	2.91E-05
10547664	Clec4e	1.481126891	6.09E-08	2.92E-05
10567580	Igsf6	0.715295729	6.13E-08	2.92E-05
10558769	Ifitm1	0.912770287	6.34E-08	2.96E-05
10358408	Rgs1	0.552416512	6.39E-08	2.96E-05
10350516	Ptgs2	1.264735585	6.91E-08	3.12E-05
10503098	Lyn /// Lyn /// Gm11787	0.741080309	6.92E-08	3.12E-05
10499861	S100a9	1.623253128	7.35E-08	3.26E-05
10435501	Stfa1	1.148132173	7.41E-08	3.26E-05
10481627	Lcn2	1.041958733	7.52E-08	3.27E-05
10350742	Rnasel	0.51575754	7.81E-08	3.35E-05
10392484	Abca8b	-0.397645889	8.20E-08	3.45E-05
10347335	Slc11a1	0.586758934	8.25E-08	3.45E-05
10477717	Procr	0.421396828	8.39E-08	3.47E-05
10439936	Nfkbiz	0.402627223	8.53E-08	3.47E-05
10568668	Adam12	0.37473823	8.59E-08	3.47E-05
10427895	Basp1	0.377187144	8.70E-08	3.48E-05
10461721	Mpeg1	0.895559695	9.49E-08	3.75E-05
10499902	Sprr4	0.512118605	9.62E-08	3.76E-05
10389222	Ccl6	0.727703084	1.01E-07	3.82E-05
10554599	Adamtsl3	-0.370355486	1.01E-07	3.82E-05

10384458	Plek	1.093469683	1.01E-07	3.82E-05
10496727	Ddah1	0.515135084	1.03E-07	3.82E-05
10601834	Gprasp2	-0.269879367	1.03E-07	3.82E-05
10351206	Selp	0.840694083	1.12E-07	4.11E-05
10572693	Jak3 /// Insl3	0.244932151	1.15E-07	4.17E-05
10391061	Krt16	1.424727578	1.17E-07	4.19E-05
10579872	---	0.382223174	1.20E-07	4.26E-05
10354432	Myo1b	0.463945082	1.34E-07	4.72E-05
10472538	Dhrs9	0.686685774	1.40E-07	4.84E-05
10519717	Sema3a	0.414908406	1.42E-07	4.84E-05
10364361	Icosl	0.242111932	1.43E-07	4.84E-05
10537146	Akr1b8	0.572195382	1.44E-07	4.84E-05
10374236	Upp1	0.728124576	1.46E-07	4.84E-05
10425066	Csf2rb	0.727245204	1.46E-07	4.84E-05
10435565	Hcls1	0.526324484	1.49E-07	4.89E-05
10581813	Mlkl	0.34156443	1.54E-07	4.97E-05
10499899	Sprr1a	0.945940891	1.54E-07	4.97E-05
10351197	Sell	0.871802891	1.56E-07	4.98E-05
10412123	---	0.742054083	1.61E-07	5.10E-05
10464529	Tcirg1	0.310796341	1.65E-07	5.16E-05
10591884	Glb1l2	-0.246828369	1.68E-07	5.18E-05
10582997	Casp4	0.666068612	1.68E-07	5.18E-05
10579636	Cyp4f18	0.833030258	1.74E-07	5.26E-05
10376778	Mfap4	0.680017715	1.74E-07	5.26E-05
10395520	Immp2l	-0.179472549	1.76E-07	5.27E-05
10539244	Tacr1	0.352472014	1.78E-07	5.27E-05
10445753	Trem3	1.008970487	1.85E-07	5.46E-05
10585555	Pstpip1	0.280956985	1.98E-07	5.67E-05
10352928	Rp1	-0.302908271	1.99E-07	5.67E-05
10463070	Entpd1	0.83436244	2.01E-07	5.67E-05
10398326	Meg3	0.54241179	2.02E-07	5.67E-05
10585286	Arhgap20	-0.297809586	2.03E-07	5.67E-05
10534935	Pilrb1	0.478488714	2.03E-07	5.67E-05
10519140	Mmp23	0.350933281	2.04E-07	5.67E-05
10545921	Mxd1	0.555461741	2.08E-07	5.73E-05
10405216	Syk	0.52223161	2.10E-07	5.73E-05
10487597	Il1b	1.660901165	2.11E-07	5.73E-05
10489107	Samhd1	0.569661024	2.15E-07	5.78E-05
10475517	AA467197 /// Mir147	0.56866572	2.26E-07	5.95E-05

10603551	Cybb	0.77594813	2.27E-07	5.95E-05
10601385	Tlr13	0.955615482	2.28E-07	5.95E-05
10389759	Ankfn1	-0.26774664	2.28E-07	5.95E-05
10416837	Irg1	1.381808438	2.30E-07	5.95E-05
10499904	Ivl	0.488833309	2.33E-07	6.00E-05
10493831	S100a8	1.680062139	2.40E-07	6.12E-05
10419034	2610528A11Rik	0.857401658	2.47E-07	6.26E-05
10370721	Sbno2	0.422520702	2.57E-07	6.45E-05
10371846	Apaf1	0.258320034	2.58E-07	6.45E-05
10527158	Fscn1	0.411573415	2.75E-07	6.83E-05
10575799	Plcg2	0.268749357	2.86E-07	7.06E-05
10557326	Il4ra	0.455492588	2.94E-07	7.16E-05
10351525	Mpz	-0.675290888	3.02E-07	7.30E-05
10410931	Vcan	0.776732878	3.25E-07	7.60E-05
10564938	Fes	0.266985694	3.27E-07	7.60E-05
10511779	Atp6v0d2	0.628719695	3.28E-07	7.60E-05
10451953	Lrg1	0.776695891	3.29E-07	7.60E-05
10560862	Pinlyp	0.521015713	3.30E-07	7.60E-05
10578264	Msr1	0.591858532	3.31E-07	7.60E-05
10508392	Rnf19b	0.456960455	3.32E-07	7.60E-05
10517488	Ephb2	0.227161244	3.33E-07	7.60E-05
10515803	Wdr65	-0.337186502	3.42E-07	7.63E-05
10427369	Pde1b	0.244701591	3.42E-07	7.63E-05
10551883	Tyrobp	0.852729487	3.42E-07	7.63E-05
10476321	Prnd	0.378948304	3.44E-07	7.63E-05
10538100	Repin1	-0.189593542	3.45E-07	7.63E-05
10530145	Tlr1	0.529042667	3.47E-07	7.63E-05
10553261	Kcnc1	-0.318324135	3.53E-07	7.71E-05
10550906	Plaur	0.86584396	3.59E-07	7.79E-05
10476301	Smox	0.555777919	3.63E-07	7.84E-05
10425092	Cyth4	0.497007822	3.71E-07	7.95E-05
10397346	Fos	0.655477724	3.77E-07	7.99E-05
10597648	Myd88	0.397992988	3.77E-07	7.99E-05
10545045	Fam13a	-0.307047666	3.94E-07	8.27E-05
10521667	Bst1	0.543602269	3.96E-07	8.27E-05
10432780	Krt6a	1.019134925	3.97E-07	8.27E-05
10459866	Slc14a1	0.44709575	4.10E-07	8.43E-05
10535174	Tmem184a	-0.382185982	4.13E-07	8.45E-05
10433114	Itga5	0.44596887	4.21E-07	8.56E-05

10346611	Gm973	-0.285836062	4.24E-07	8.57E-05
10606714	Gla	0.546680978	4.27E-07	8.58E-05
10494271	Ctss	0.722132714	4.30E-07	8.59E-05
10367919	Stx11	0.62448849	4.39E-07	8.71E-05
10474836	Ivd	-0.264061185	4.46E-07	8.80E-05
10563597	Saa3	1.191171188	4.56E-07	8.96E-05
10490150	Zbp1	0.341564603	4.60E-07	8.97E-05
10471844	Nek6	0.324868451	4.69E-07	9.10E-05
10364375	Cstb	0.477464027	4.83E-07	9.28E-05
10416800	Lmo7	-0.31727954	4.83E-07	9.28E-05
10534667	Serpine1	0.495286179	5.05E-07	9.65E-05
10597323	Arpp21	-0.267902143	5.08E-07	9.65E-05
10592084	St3gal4	0.331406025	5.11E-07	9.65E-05
10583044	Mmp13	1.026985255	5.25E-07	9.79E-05
10569877	1810033B17Rik	0.576723994	5.26E-07	9.79E-05
10518069	Efhd2	0.386988715	5.27E-07	9.79E-05
10545479	Tmsb10	0.498617093	5.37E-07	9.91E-05
10406676	Lhfpl2	0.350053392	5.46E-07	9.99E-05
10397112	Papln	-0.28666379	5.72E-07	0.000104157
10439292	BC100530 /// BC117090 /// Stfa1	1.480561659	5.80E-07	0.000104985
10379630	Slfn2	0.57522639	5.90E-07	0.000106213
10548375	Clec7a	0.957893555	6.22E-07	0.000110765
10404904	Rbm24	-0.431987316	6.26E-07	0.000110765
10452485	Rab31	0.42381586	6.27E-07	0.000110765
10447951	Thbs2	0.626875226	6.28E-07	0.000110765
10598004	Ccr1	1.010350929	6.36E-07	0.000111614
10570068	Col4a2	0.313926631	6.48E-07	0.000112773
10375751	Adamts2	0.550193292	6.53E-07	0.000112773
10592251	Pknox2	-0.304538831	6.54E-07	0.000112773
10466210	Ms4a6d	0.836294497	6.55E-07	0.000112773
10416437	Lcp1	0.90749695	6.72E-07	0.000115145
10510700	Gpr153	0.264052593	6.87E-07	0.000117104
10529979	Ppargc1a	-0.538258494	7.04E-07	0.000119309
10595979	Mras	-0.215022996	7.25E-07	0.000121859
10473809	Sfpi1	0.669549808	7.26E-07	0.000121859
10499891	Sprr1b	1.30595022	7.35E-07	0.000122603
10427796	Npr3	-0.389921778	7.37E-07	0.000122603
10366293	Csrp2	0.299241174	7.52E-07	0.00012284
10551891	Nfkbid	0.578488679	7.54E-07	0.00012284

10451974	Sema6b	0.27567543	7.55E-07	0.00012284
10561453	Zfp36	0.592168812	7.57E-07	0.00012284
10550509	Pglyrp1	0.746003328	7.60E-07	0.00012284
10443527	Pim1	0.433785336	7.63E-07	0.00012284
10498992	Tlr2	0.535686132	7.66E-07	0.00012284
10439299	Stfa3	1.186464483	7.67E-07	0.00012284
10358978	Ier5	0.374736347	7.91E-07	0.000125777
10586242	Dennd4a	0.390417842	7.93E-07	0.000125777
10456400	Tubb6	0.378083995	7.96E-07	0.000125777
10508772	Fgr	0.440983486	8.06E-07	0.000126831
10490061	Bcas1	-0.246233761	8.45E-07	0.000132154
10360028	Fcgr2b	0.688726241	8.48E-07	0.000132154
10536483	Tes	0.345269751	9.06E-07	0.00014
10363082	Lilrb4	1.016299643	9.22E-07	0.000141843
10353844	Neurl3	0.437939191	9.29E-07	0.00014185
10381809	Itgb3	0.311387399	9.30E-07	0.00014185
10486833	Ell3 /// Serinc4	-0.274251245	9.44E-07	0.000143314
10588037	Rbp1	0.27269474	9.59E-07	0.000144911
10406270	Glrx	0.63873625	9.72E-07	0.000146236
10432439	Fmnl3	0.263223368	1.01E-06	0.000150271
10477250	Hck	0.526098613	1.01E-06	0.000150271
10392845	Cd300lf	0.551957954	1.01E-06	0.000150271
10520452	Il6	0.687497684	1.02E-06	0.000150271
10588043	Rbp2	0.585566606	1.02E-06	0.000150929
10557862	Itgam	0.751058423	1.03E-06	0.000151157
10357472	Cxcr4	0.480166181	1.05E-06	0.000152433
10360070	Fcer1g	0.845083798	1.06E-06	0.000152433
10440393	Samsn1	0.775266199	1.06E-06	0.000152433
10349427	---	0.600079588	1.06E-06	0.000152433
10601980	Mum1l1	-0.196462789	1.06E-06	0.000152433
10430302	Csf2rb2	0.591250201	1.08E-06	0.000154065
10341410	---	0.402946827	1.11E-06	0.000157634
10423836	Cthrc1	0.394464835	1.12E-06	0.000158217
10476395	Bmp2	-0.354070876	1.12E-06	0.000158217
10488378	Thbd	0.370060904	1.12E-06	0.000158217
10358601	Hmcn1	0.346905061	1.14E-06	0.000159578
10370210	Col6a1	0.358702043	1.15E-06	0.000160684
10435497	Stfa2l1	1.278896745	1.16E-06	0.000161968
10428943	Gsdmc	0.982036011	1.19E-06	0.000165098

10462281	Vldlr	-0.351968359	1.20E-06	0.000166197
10415991	Zfp395	-0.199781401	1.24E-06	0.000170058
10352152	Kif26b	0.236103175	1.26E-06	0.000172098
10572897	Hmox1	0.5935614	1.27E-06	0.000172545
10574276	Gpr97	0.334181052	1.27E-06	0.000172545
10582303	Cyba	0.584486374	1.31E-06	0.00017784
10426894	Mettl7a3 /// Mettl7a2	-0.210126586	1.38E-06	0.000186337
10517274	Sepn1 /// Sepn1	0.403418265	1.39E-06	0.000186464
10566132	Rhog	0.413281473	1.41E-06	0.000188036
10541246	Il17ra	0.316525269	1.41E-06	0.000188036
10467139	Lipa	0.481396021	1.44E-06	0.000190869
10432774	Krt6b	1.372516593	1.46E-06	0.00019298
10504402	Tmem8b	-0.19910478	1.48E-06	0.000195259
10358666	Hmcn1	0.472639164	1.51E-06	0.000198214
10509901	Mfap2	0.439632891	1.53E-06	0.000200566
10354247	Fhl2	0.314896127	1.55E-06	0.000201178
10576973	Col4a1	0.320312964	1.56E-06	0.000202405
10425031	Apol6	-0.297230743	1.57E-06	0.000202814
10358545	Hmcn1	0.518657431	1.58E-06	0.000202814
10370180	Col6a2	0.360547168	1.59E-06	0.000203538
10566583	Gm8995	0.601523166	1.60E-06	0.000204005
10534303	Lat2	0.21558713	1.60E-06	0.000204096
10534585	Sh2b2	0.237122588	1.62E-06	0.000205628
10358543	Hmcn1	0.341857399	1.64E-06	0.000207236
10378857	Coro6	-0.44282884	1.65E-06	0.000207236
10357952	Ppp1r12b	-0.209878708	1.66E-06	0.000207954
10435345	Mylk	0.249587661	1.67E-06	0.000209131
10457022	Mbp	-0.263651595	1.68E-06	0.000209586
10591988	Adamts15	0.366745746	1.69E-06	0.000209978
10398364	DQ267102	0.482444924	1.72E-06	0.000212621
10445758	Treml4	0.829440123	1.73E-06	0.000212939
10370037	Mmp11	0.309401328	1.74E-06	0.000212939
10566358	Trim30a	0.620674259	1.75E-06	0.000213629
10523156	Cxcl2	1.55383049	1.80E-06	0.000219033
10358599	Hmcn1	0.343090259	1.82E-06	0.000219465
10404686	Bmp6	-0.206104134	1.82E-06	0.000219465
10453717	---	0.326316532	1.83E-06	0.000219839
10421361	Bmp1	0.291148686	1.84E-06	0.000220347
10425078	Mpst	-0.182442095	1.86E-06	0.000221583

10416215	Loxl2	0.482660336	1.87E-06	0.000222051
10394770	Odc1	0.370351644	1.87E-06	0.000222051
10396079	Klhdc1	-0.25708649	1.90E-06	0.000224743
10377804	Arrb2	0.293136155	1.96E-06	0.000230667
10552516	Klk6	0.725410549	1.96E-06	0.000230667
10490212	Ctsz	0.384028874	2.09E-06	0.000244474
10416181	Stc1	0.521451081	2.10E-06	0.000245071
10432661	Galnt6	0.351225826	2.13E-06	0.000246566
10586252	Dennd4a	0.48383767	2.14E-06	0.000246597
10565152	Homer2	-0.341610229	2.14E-06	0.000246597
10586250	Dennd4a	0.516656903	2.15E-06	0.000246739
10586227	Dennd4a	0.505726331	2.16E-06	0.000246908
10600836	Msn	0.429556858	2.18E-06	0.000247878
10546432	Adamts9	0.359930339	2.18E-06	0.000247878
10412260	Fst	0.293649257	2.24E-06	0.000253359
10382106	Milr1	0.441019956	2.35E-06	0.000264762
10596718	Slc38a3	-0.384090589	2.38E-06	0.000267233
10430372	Rac2	0.531974823	2.40E-06	0.000267884
10586254	Dennd4a	0.412231235	2.43E-06	0.000271065
10351603	Arhgap30	0.274207641	2.47E-06	0.000273225
10515399	Plk3	0.703783383	2.48E-06	0.000273225
10424370	Trib1	0.418556797	2.48E-06	0.000273225
10448124	Fpr1	0.860338087	2.51E-06	0.000276412
10358605	Hmcn1	0.298323858	2.53E-06	0.000276916
10363070	Gp49a	1.106310152	2.61E-06	0.0002852
10543058	Dlx5	-0.264342935	2.64E-06	0.000287694
10365729	Cdk17	0.267055516	2.70E-06	0.000293333
10384423	Cobl	-0.265438443	2.73E-06	0.000295429
10529515	Sorcs2	0.24986189	2.74E-06	0.000295429
10497372	Gm5150	0.767893041	2.75E-06	0.000295429
10478415	Wisp2	0.348272312	2.80E-06	0.000299776
10505073	Zfp462	-0.249638599	2.82E-06	0.000300378
10481101	Snora43 /// Snhg7	0.242028593	2.87E-06	0.000305473
10344506	---	-0.44297024	2.90E-06	0.000307723
10593050	Il10ra	0.436389691	2.92E-06	0.00030876
10557111	Scnn1g	-0.213739458	3.00E-06	0.00031587
10400510	Clec14a	0.436059106	3.04E-06	0.000316108
10420114	Tgm1	0.502074872	3.04E-06	0.000316108
10511923	Pm20d2	-0.186049458	3.04E-06	0.000316108

10370946	Mob3a	0.195071747	3.05E-06	0.000316108
10420198	Ripk3	0.238158472	3.10E-06	0.000320808
10413492	Lrtm1	-0.307351155	3.17E-06	0.000326926
10430358	C1qtnf6	0.256004133	3.20E-06	0.000328461
10370259	Col18a1	0.255230855	3.20E-06	0.000328461
10507671	Guca2a	-0.452441088	3.22E-06	0.000329647
10418053	Kcnma1	-0.386438602	3.33E-06	0.000338351
10487645	Cpxm1	0.571597388	3.34E-06	0.000338351
10586240	Dennd4a	0.494109458	3.39E-06	0.000342677
10574166	Cpne2	0.194427339	3.41E-06	0.000342677
10357944	Ppp1r12b	-0.252748584	3.42E-06	0.000342677
10349118	Serpinb12	0.668048169	3.43E-06	0.000342677
10358589	Hmcn1	0.287180049	3.47E-06	0.000345253
10373515	Suox	-0.32065768	3.49E-06	0.00034651
10392808	Cd300ld	0.646402513	3.51E-06	0.000347163
10511416	Tox	-0.195443949	3.52E-06	0.000347338
10397351	Jdp2	0.332632758	3.53E-06	0.000347805
10567108	Sox6	-0.280460288	3.54E-06	0.000347942
10437687	Litaf	0.402022173	3.63E-06	0.00035552
10348244	Inpp5d	0.404033612	3.64E-06	0.00035583
10473058	Osbpl6	-0.306323168	3.66E-06	0.000356427
10485405	Cd44	0.530300016	3.67E-06	0.000356563
10435704	Cd80	0.580587843	3.70E-06	0.000356997
10383194	Rnf213	0.170254299	3.70E-06	0.000356997
10574220	Cx3cl1	-0.1625985	3.76E-06	0.000362071
10469695	Apbb1ip	0.361329358	3.80E-06	0.000364405
10449284	Dusp1	0.637180143	3.81E-06	0.000364405
10506254	Raver2	-0.16941291	3.82E-06	0.000364549
10372410	Glipr1	0.617154076	3.84E-06	0.000364549
10508465	Marcksl1	0.566593044	3.84E-06	0.000364549
10582162	Cotl1	0.328355583	3.90E-06	0.000369122
10379511	Ccl2	0.437954636	3.94E-06	0.000371778
10435641	Fstl1	0.680332933	3.97E-06	0.000371977
10473367	Slc43a1	-0.242499319	3.98E-06	0.000371977
10351509	Fcgr4	0.555657604	3.98E-06	0.000371977
10376513	Nlrp3	0.87626806	3.98E-06	0.000371977
10505623	Lurap1l	0.234357606	4.00E-06	0.000372785
10537742	Clcn1	-0.326404281	4.05E-06	0.000376365
10523182	Areg	0.341812605	4.07E-06	0.000377586

10583100	Mmp8	0.836923476	4.21E-06	0.000388788
10521678	Cd38	0.341964963	4.22E-06	0.000388788
10569485	Tnfrsf26	0.357808588	4.27E-06	0.000392325
10481164	Slc2a6	0.357104026	4.28E-06	0.000392325
10373912	Osm	0.815474635	4.33E-06	0.000396509
10584674	Mcam	0.394477803	4.39E-06	0.000400629
10358573	Hmcn1	0.302881795	4.43E-06	0.000403537
10363455	Pcbd1	-0.31484185	4.59E-06	0.000416368
10547621	Apobec1	0.379825716	4.66E-06	0.00042068
10379633	Slfn1	0.619137657	4.86E-06	0.000436164
10395103	Pxdn	0.441316311	4.89E-06	0.000438242
10386965	Arhgap44	-0.261517788	4.93E-06	0.000440522
10379535	Ccl8	0.879642306	4.95E-06	0.000440678
10493925	Hrnr /// Flg	0.705181886	4.96E-06	0.000440678
10401296	Slc8a3	-0.293765554	5.00E-06	0.000443645
10358549	Hmcn1	0.434635933	5.16E-06	0.000454815
10446253	Vav1	0.436125578	5.17E-06	0.000455007
10364593	Cnn2	0.478393476	5.21E-06	0.000457184
10358928	Cacna1e	-0.249944922	5.28E-06	0.000460901
10580282	Junb	0.472504078	5.29E-06	0.000460901
10444890	Ier3	0.58332504	5.29E-06	0.000460901
10354588	Stk17b	0.602321129	5.33E-06	0.00046288
10357950	Ppp1r12b	-0.281116507	5.37E-06	0.000465623
10603492	Porcn	0.187713164	5.40E-06	0.000466894
10409876	Ctla2a	0.514458519	5.44E-06	0.000469005
10594540	Plekho2	0.377123539	5.48E-06	0.000471509
10350149	Tnni1	-0.42698613	5.52E-06	0.00047351
10351679	Cd84	0.500239348	5.55E-06	0.000475323
10583669	AB124611	0.432030187	5.58E-06	0.000476926
10447006	Vit	-0.184990512	5.70E-06	0.0004844
10550574	Dmpk	-0.320090968	5.70E-06	0.0004844
10517165	Cd52	0.711656196	5.79E-06	0.000489281
10441718	Park2 /// Park2	-0.191825241	5.80E-06	0.000489722
10571653	Actg1 /// Gm12715 /// Actg-ps1 /// Gm8399	0.398473725	5.94E-06	0.000499651
10525397	Arpc3 /// Arpc3	0.413489976	5.96E-06	0.000500333
10420957	Ptk2b	0.268588884	5.99E-06	0.000501335
10383799	Tcn2 /// Tcn2	0.455536526	6.02E-06	0.000503255
10557591	Itgal	0.387010815	6.06E-06	0.000504875
10554204	Agbl1	-0.592771813	6.15E-06	0.000510549

10439442	Pla1a	0.290741175	6.15E-06	0.000510549
10409118	Wnk2	-0.274361804	6.17E-06	0.000510875
10519747	Sema3e	-0.297592089	6.23E-06	0.000514469
10590631	Ccr2	0.461912688	6.26E-06	0.000515952
10501063	Cd53	1.042565773	6.34E-06	0.000520711
10358224	Ptprc	0.637881818	6.38E-06	0.000522775
10463599	Nfkb2	0.229332455	6.40E-06	0.00052376
10510509	Gpr157	-0.169242496	6.46E-06	0.00052711
10541605	Clec4n	0.728064034	6.68E-06	0.000544043
10544913	Crhr2	-0.323849905	6.72E-06	0.000545657
10492091	Smad9	-0.22708658	6.75E-06	0.000546687
10526277	Mlxipl	-0.313925481	6.88E-06	0.000554984
10569504	Tnfrsf23	0.30891062	6.94E-06	0.000557403
10547641	Slc2a3	0.596649759	6.94E-06	0.000557403
10521205	Sh3bp2	0.19696392	6.98E-06	0.000559213
10523451	Anxa3	0.347677573	7.02E-06	0.000559431
10403584	Nid1	0.563005425	7.17E-06	0.00057018
10523120	Cxcl5	0.646686641	7.21E-06	0.000572292
10351099	Tnfsf18	-0.331427391	7.27E-06	0.00057547
10485013	1110051M20Rik /// 1110051M20Rik	-0.20820284	7.29E-06	0.000576038
10451580	Bysl	0.171755948	7.53E-06	0.000589882
10344981	Pi15	0.459669145	7.60E-06	0.000593157
10475866	Bcl2l11	0.365236948	7.61E-06	0.000593157
10546010	Arhgap25	0.277263824	7.71E-06	0.000599073
10489891	B4galt5	0.356811819	7.72E-06	0.000599073
10344725	Adhfe1	-0.305199638	8.04E-06	0.00061783
10521984	G6pd2	0.208075723	8.04E-06	0.00061783
10411668	Ocln	-0.240747616	8.05E-06	0.00061783
10543572	Impdh1	0.176035708	8.12E-06	0.000619891
10363901	Etv5	-0.232348177	8.13E-06	0.000619891
10358660	Hmcn1	0.540531141	8.16E-06	0.000620407
10541587	Clec4a2	0.389778952	8.32E-06	0.000631628
10545760	Paip2b	-0.140527322	8.35E-06	0.000632566
10445412	Nfkbie	0.374317811	8.45E-06	0.000638418
10531887	Slc10a6	0.224850308	8.51E-06	0.000641906
10379321	Rab11fip4	-0.242456696	8.56E-06	0.00064351
10440926	Dnajc28	-0.232747696	8.60E-06	0.000644491
10369290	Ddit4	0.265368588	8.74E-06	0.00065302
10490972	Trim55	-0.317025402	8.78E-06	0.000654566

10527441	Arpc1b /// Gm5637	0.325592211	8.79E-06	0.000654566
10515771	Tie1	0.322140898	8.95E-06	0.000663311
10488382	Cd93	0.62005639	8.96E-06	0.000663311
10494174	Sema6c	-0.223399711	8.97E-06	0.000663311
10376096	Acsl6	-0.248854995	9.02E-06	0.000665313
10531560	Antxr2	0.490579384	9.15E-06	0.000673448
10400357	Baz1a /// LOC100048557	0.258672531	9.28E-06	0.000680325
10540085	Fbln2	0.491457728	9.38E-06	0.000686644
10508069	Ftl1 /// Mir692-1	0.44907503	9.49E-06	0.000689939
10358656	Hmcn1	0.488107118	9.53E-06	0.000690394
10371332	Aldh1l2	0.212971218	9.54E-06	0.000690394
10607012	Col4a6	-0.264025747	9.55E-06	0.000690394
10358658	Hmcn1	0.47385586	9.58E-06	0.000691203
10497364	Sirpb1a /// Sirpb1b /// LOC100038947	0.648983635	9.64E-06	0.000693559
10591494	S1pr5	-0.204022521	9.90E-06	0.000711331
10493633	Tpm3	-0.314389203	9.93E-06	0.000711358
10569656	Tpcn2	0.164687168	9.96E-06	0.000712598
10463670	Sfxn2	-0.160920917	1.00E-05	0.000716892
10593668	Dmxl2	0.329138942	1.01E-05	0.000717025
10507477	---	0.585994099	1.01E-05	0.000717352
10600144	F8a	-0.150790958	1.01E-05	0.000718615
10380956	Gsdma3	0.250626638	1.02E-05	0.000718615
10519578	Abcb4	-0.299114587	1.02E-05	0.000719521
10357965	Lgr6	-0.374891246	1.02E-05	0.000721106
10415396	Nfatc4	0.253696042	1.02E-05	0.000721106
10554789	Ctsc /// Ctsc	0.662108988	1.04E-05	0.000727234
10554945	Prcp	0.407947223	1.04E-05	0.000728022
10419934	Myh7	-0.552055885	1.05E-05	0.000734693
10592266	Slc37a2	0.308998144	1.06E-05	0.000736195
10501164	Csf1	0.323343081	1.06E-05	0.000738504
10420846	Fzd3	-0.287805343	1.07E-05	0.000739252
10358637	Hmcn1	0.356924603	1.08E-05	0.000747894
10424584	Dennd3	0.315093878	1.09E-05	0.000748788
10422760	Fyb	0.46405481	1.10E-05	0.000755483
10445293	Pla2g7	0.52447263	1.11E-05	0.000760366
10460385	Clcf1	0.204773463	1.11E-05	0.000761521
10360248	Atp1a4	-0.27102708	1.14E-05	0.000780468
10380403	Lrrc59	0.236478209	1.15E-05	0.00078677
10511333	Plag1	-0.247747704	1.17E-05	0.000800273

10461636	Gm10212	0.394067803	1.18E-05	0.000801377
10428809	Klhl38	-0.446000016	1.19E-05	0.000806093
10379736	1100001G20Rik	0.320396779	1.20E-05	0.000809262
10374560	Zrsr1	-0.191964509	1.20E-05	0.000811936
10525542	Bcl7a	-0.151133306	1.21E-05	0.000811936
10597279	Ccrl2	0.72136536	1.21E-05	0.000816524
10519951	Gsap	0.353924456	1.25E-05	0.000840582
10417212	Itgbl1	0.520838417	1.27E-05	0.000849897
10538082	Atp6v0e2	-0.213262936	1.28E-05	0.000852651
10589602	Myl3	-0.379728818	1.32E-05	0.000881421
10373740	Pik3ip1	-0.204243595	1.34E-05	0.000893503
10411680	Marveld2	-0.227393031	1.35E-05	0.000893503
10418921	Sncg	-0.300834693	1.35E-05	0.000893503
10595768	Pls1	-0.227441854	1.36E-05	0.000899765
10448743	Fahd1	-0.186106766	1.37E-05	0.000906238
10390691	Nr1d1	-0.384094862	1.38E-05	0.000909144
10382316	Kcnj16	-0.253683226	1.38E-05	0.000910628
10360270	Atp1a2	-0.547963768	1.40E-05	0.000917701
10598013	Ccr5	0.466419008	1.41E-05	0.000922637
10468893	Csf2ra	0.243451665	1.42E-05	0.000930215
10544732	Skap2	0.264980436	1.43E-05	0.000932384
10385036	Fgf18	-0.275314185	1.44E-05	0.000932384
10523190	Parm1	0.299820054	1.44E-05	0.000932409
10497358	Sirpb1b /// LOC100038947 /// Sirpb1a	0.73771508	1.44E-05	0.000933334
10494781	Igsf3	0.18031277	1.46E-05	0.00094102
10528480	---	0.240297763	1.46E-05	0.00094102
10380976	Gsdma	0.415495928	1.46E-05	0.00094102
10464642	Carns1	-0.508583956	1.48E-05	0.000947891
10464560	Aldh3b1	0.261188735	1.49E-05	0.000948008
10476021	Sirpa	0.355711438	1.49E-05	0.000948008
10556509	Spon1	0.320141508	1.49E-05	0.000948008
10430201	Myh9	0.302808673	1.50E-05	0.000954409
10540935	Cand2	-0.238834729	1.53E-05	0.000975268
10391870	Map3k14 /// 1700028N14Rik	0.157616082	1.55E-05	0.000984248
10495967	Tifa	0.24677101	1.56E-05	0.000984573
10478890	Cebpb	0.361418058	1.58E-05	0.000998678
10519983	Fgl2	0.588334524	1.59E-05	0.001000427
10544932	Inmt	-0.464597165	1.59E-05	0.001003563
10500276	BC028528	0.364261747	1.60E-05	0.001006753

10385656	Zfp2 /// Zfp2	-0.177827641	1.63E-05	0.001019601
10563295	Ftl1 /// Mir692-1 /// LOC101056386 /// LOC100862446 /// Gm20746	0.431220788	1.63E-05	0.001021776
10559547	Tnnt1	-0.585892373	1.64E-05	0.001023895
10368144	Tnfaip3	0.358746645	1.65E-05	0.00102512
10570606	Defb14	0.596638613	1.65E-05	0.00102512
10404702	Gcnt2	0.303653006	1.66E-05	0.00102512
10562192	Fxyd5	0.527859468	1.68E-05	0.00103621
10501555	Amy1	-0.318165045	1.68E-05	0.00103621
10485151	Mapk8ip1	-0.152727086	1.68E-05	0.00103621
10585048	Cadm1	-0.223755568	1.73E-05	0.00106266
10520862	Fosl2	0.313199336	1.74E-05	0.001069613
10379646	Slfn3	0.454816046	1.75E-05	0.001073317
10423471	Ctnnd2	-0.191023732	1.78E-05	0.001083605
10346191	Stat1	0.278378281	1.83E-05	0.001110897
10467470	Aldh18a1	0.18123087	1.84E-05	0.001114966
10532628	Myo18b	-0.511962212	1.84E-05	0.001115521
10580139	Zswim4	0.256617435	1.85E-05	0.001116407
10509514	Sh2d5	0.225289568	1.86E-05	0.001116407
10415319	Irf9	0.201216006	1.87E-05	0.00111747
10392796	Cd300lb	0.357868097	1.87E-05	0.001117748
10358559	Hmcn1	0.343859455	1.91E-05	0.001139112
10602977	Scml2	-0.215639696	1.92E-05	0.001140394
10518686	Pik3cd	0.189473837	1.92E-05	0.001142702
10489878	Ptgis	-0.31870243	1.93E-05	0.001145206
10568282	Bcl7c	-0.156132458	1.94E-05	0.001146189
10413381	Asb14	-0.414923589	1.95E-05	0.001151805
10593713	Cib2	-0.196745208	1.95E-05	0.001152907
10522217	Limch1	-0.360612432	1.97E-05	0.001160498
10391025	Krt15	-0.492350314	1.98E-05	0.001166055
10605571	Gyk	0.494195778	1.99E-05	0.001166055
10572146	Atp6v1b2 /// Atp6v1b2	0.366500403	1.99E-05	0.001166055
10496605	Ccbl2	-0.222438163	2.00E-05	0.001168346
10353460	Kcnq5	-0.328272445	2.01E-05	0.00117307
10401473	Aldh6a1	-0.328147045	2.01E-05	0.001174149
10390090	Sgca	-0.430710538	2.04E-05	0.001184873
10554900	Dlg2	-0.162248269	2.04E-05	0.001184873
10575363	Zfp612	-0.17010693	2.05E-05	0.001187574
10362701	Ddo	-0.207595588	2.07E-05	0.001193873
10367224	Stat2	0.175278246	2.07E-05	0.001195378

10351224	F5	0.378173074	2.09E-05	0.001199902
10422059	Kctd12 /// Mir5130	0.23217724	2.11E-05	0.00120603
10462035	Ldhb	-0.377757424	2.20E-05	0.001249492
10402473	Clmn	-0.279413056	2.20E-05	0.001249492
10473312	Fam171b	-0.390113595	2.20E-05	0.001249492
10546430	Adamts9	0.270024774	2.21E-05	0.001249754
10415052	Mmp14	0.380232514	2.24E-05	0.001259499
10439296	Stfa2	0.626158797	2.24E-05	0.001259499
10500469	Pde4dip	-0.456811256	2.24E-05	0.001259499
10458731	Mcc	-0.205191563	2.26E-05	0.001267967
10484472	Smtnl1	-0.317880722	2.27E-05	0.001269218
10525343	Myl2	-0.63590777	2.27E-05	0.001269218
10518570	Pgd	0.316084384	2.27E-05	0.001269218
10458906	Ppic	0.357682262	2.29E-05	0.001275847
10526559	Ache	-0.325346118	2.33E-05	0.001294423
10533844	Rilpl2	0.379852529	2.33E-05	0.00129568
10555174	Lrrc32	0.3798279	2.34E-05	0.001297054
10607361	Gm4750 /// Gm8464	0.235184362	2.38E-05	0.001318989
10469856	Wdr85	-0.126923499	2.39E-05	0.001320578
10415980	Fbxo16	-0.180510349	2.40E-05	0.001322164
10351491	Olfml2b	0.419520215	2.41E-05	0.001322164
10495976	Pitx2	-0.206066314	2.41E-05	0.001322164
10473240	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735	0.350874685	2.42E-05	0.001325941
10431962	Endou	0.392433959	2.42E-05	0.001326257
10424731	Gsdmd	0.197437058	2.44E-05	0.001328996
10470388	Cacfd1	-0.164925031	2.44E-05	0.001328996
10357946	Ppp1r12b	-0.324053845	2.45E-05	0.001330775
10474526	Lpcat4	0.16320546	2.45E-05	0.001330775
10571514	Gm6180	0.295010549	2.50E-05	0.001353301
10592847	Myl6	0.232978461	2.53E-05	0.001367905
10493235	Paqr6	-0.201568285	2.56E-05	0.001379252
10445774	B430306N03Rik	0.35705353	2.56E-05	0.001379252
10552090	Scgb2b2	-0.476005048	2.57E-05	0.001380049
10573583	Man2b1	0.282476158	2.59E-05	0.001388353
10514779	Prkaa2	-0.375418423	2.64E-05	0.00141066
10461587	Ms4a4a	0.812420426	2.65E-05	0.00141558
10380761	Socs7	-0.152177013	2.66E-05	0.00141558
10434778	Rtp4	0.393680172	2.68E-05	0.001427859
10506433	Dab1	-0.251797815	2.69E-05	0.001427859

10570855	Plat	0.328504878	2.71E-05	0.001437652
10539739	Asprv1	0.61546138	2.71E-05	0.001437652
10579347	Ifi30	0.370301366	2.72E-05	0.001439167
10502830	Nexn	-0.564622088	2.75E-05	0.001455141
10494023	Rorc	-0.303595828	2.76E-05	0.001455306
10463282	Entpd7	0.162810313	2.77E-05	0.001455306
10534889	Agfg2	-0.149340081	2.77E-05	0.001455306
10517287	Man1c1	0.242332308	2.77E-05	0.001455972
10569385	Ascl2	-0.21845359	2.81E-05	0.001468923
10584883	Fxyd6	-0.349663299	2.83E-05	0.001474986
10425781	Serhl	-0.141496645	2.84E-05	0.001478572
10359948	Uap1	0.217112622	2.85E-05	0.001482477
10495562	Lrrc39	-0.353758759	2.88E-05	0.001493974
10548552	Klra2	0.429060919	2.88E-05	0.001496117
10513666	Akna	0.221071475	2.92E-05	0.001510381
10435504	Gm5416	0.70395993	2.95E-05	0.001524125
10464999	Cst6	0.451224693	2.97E-05	0.00153323
10357158	Ralb	0.188507944	2.98E-05	0.001533942
10503134	Sdcbp	0.368011036	3.02E-05	0.001549526
10577315	Angpt2	0.268991921	3.02E-05	0.001549526
10364950	Gadd45b	0.554983993	3.03E-05	0.001555568
10523175	Ereg	0.345581091	3.04E-05	0.001557755
10494332	---	0.225897411	3.09E-05	0.001578402
10505187	Ugcg	0.422661648	3.11E-05	0.001587138
10463476	Kazald1	-0.281483053	3.16E-05	0.001608701
10568873	Adam8	0.525514957	3.23E-05	0.001630683
10398267	Evl	0.136094981	3.23E-05	0.001630683
10582378	Piezo1	0.181270685	3.26E-05	0.001642768
10426557	Pfkm	-0.545371481	3.33E-05	0.001675838
10598178	Disp1	-0.284941517	3.35E-05	0.00167994
10407797	Prl2c4 /// Prl2c2 /// Prl2c3	0.481542956	3.35E-05	0.00167994
10338479	---	-0.423537668	3.39E-05	0.001699498
10486697	Tgm7	-0.408318121	3.46E-05	0.001728859
10509204	Tcea3	-0.298597966	3.47E-05	0.001728863
10520574	Agbl5	-0.122139163	3.47E-05	0.001729747
10579313	Ssbp4	0.14740258	3.50E-05	0.001737049
10487605	F830045P16Rik	-0.380347334	3.50E-05	0.001737049
10366528	Best3	-0.241813562	3.50E-05	0.001737049
10403871	Aoah	0.31999844	3.52E-05	0.001743089

10552760	Pnkp	0.157031292	3.55E-05	0.001757431
10355278	Erbb4	-0.259093113	3.58E-05	0.001765525
10581645	Marveld3	-0.159052568	3.60E-05	0.00177566
10506274	Dnajc6	-0.198634519	3.62E-05	0.001781389
10443470	Rab44	0.222571344	3.64E-05	0.001788066
10390175	Ngfr	-0.15207976	3.65E-05	0.00179096
10462005	Tmem2	0.186415823	3.72E-05	0.001822065
10526693	Zcwpw1	-0.211096245	3.79E-05	0.001849959
10452815	Xdh	0.367955077	3.82E-05	0.001860461
10507273	Pik3r3	0.250914756	3.83E-05	0.001860461
10432756	Gm5414 /// Gm5476	0.219047893	3.84E-05	0.001861081
10606102	Phka1	-0.31912242	3.84E-05	0.001861081
10601178	Itgb1bp2	-0.280114103	3.85E-05	0.001862097
10362372	9330159F19Rik	-0.400082235	3.96E-05	0.001910433
10426467	Tmem117	-0.193719137	3.97E-05	0.001913261
10389786	Hlf	-0.33523888	3.99E-05	0.001913662
10438904	Lrrc15	0.511157019	3.99E-05	0.001913662
10384064	Camk2b	-0.330405151	3.99E-05	0.001913662
10428336	---	0.246073525	4.00E-05	0.001913662
10460616	Cfl1	0.310063178	4.05E-05	0.001937606
10577517	Slc25a15	-0.116538546	4.22E-05	0.002011634
10393754	Actg1	0.391725975	4.26E-05	0.002022454
10519196	Vwa1	0.166608289	4.26E-05	0.002022454
10524621	Oasl2	0.373881017	4.28E-05	0.002025947
10588836	Gmppb	0.18915958	4.30E-05	0.002032215
10541260	Cecr2	-0.303994459	4.35E-05	0.002050959
10558150	Htra1	0.425937954	4.39E-05	0.002065688
10387180	Ndel1 /// Ndel1	0.309613368	4.41E-05	0.002071992
10545672	Mthfd2	0.211508177	4.41E-05	0.002071992
10583163	Trpc6	0.273564506	4.43E-05	0.002076896
10459075	Myoz3	-0.280978523	4.44E-05	0.002082256
10465844	Asrgl1	-0.161276931	4.47E-05	0.002093257
10488195	Rrbp1	0.214500237	4.48E-05	0.002094881
10513957	Ptprd	-0.187381701	4.50E-05	0.002101442
10557285	Lcmt1	-0.174574337	4.54E-05	0.002115855
10368356	Akap7	-0.238120387	4.57E-05	0.002127724
10419015	Cdhr1	0.234195006	4.72E-05	0.002188566
10439249	Parp14	0.270104283	4.74E-05	0.002189261
10553299	---	0.39666405	4.74E-05	0.002189261

10339140	---	-0.188882656	4.87E-05	0.002240246
10553477	Ano5	-0.405786753	4.98E-05	0.002278985
10365983	Lum	0.779895956	5.02E-05	0.002292652
10358635	Hmcn1	0.283281308	5.06E-05	0.00230079
10356084	Irs1	-0.320011553	5.07E-05	0.002300917
10383556	Fn3krp	-0.15073472	5.07E-05	0.002300917
10470564	Ralgds	0.232693017	5.09E-05	0.002303752
10377018	Myh3	-0.626783054	5.16E-05	0.002325433
10392364	Cacng1	-0.316510372	5.16E-05	0.002325433
10449775	Notch3	0.190068291	5.18E-05	0.00232648
10476740	Slc24a3	-0.3238715	5.23E-05	0.002341829
10409486	Pdlim7	0.228845898	5.23E-05	0.002341829
10565910	Plekhb1	-0.270100215	5.23E-05	0.002341829
10484195	Ttn	-0.481634681	5.28E-05	0.002353786
10517364	Ncmap	-0.242245887	5.28E-05	0.002353786
10468909	Disp1	-0.150068912	5.29E-05	0.002353786
10490838	Fabp5	0.727402911	5.30E-05	0.002353786
10491300	Skil	0.244074966	5.30E-05	0.002353786
10419578	Ndrg2	-0.514722861	5.31E-05	0.002355501
10402268	Lgmn	0.507982549	5.39E-05	0.002384159
10419854	Slc7a8	0.366750317	5.44E-05	0.002399743
10586591	Car12	0.441322749	5.51E-05	0.00241579
10350199	Cacna1s	-0.524211376	5.51E-05	0.00241579
10606016	Il2rg	0.578662909	5.51E-05	0.00241579
10357043	Bcl2	-0.219001755	5.58E-05	0.002442259
10433403	Rbfox1	-0.413279362	5.60E-05	0.002442268
10407792	Gpr137b-ps /// Gpr137b	0.474291278	5.60E-05	0.002442268
10345869	Tmem182	-0.607094509	5.61E-05	0.002442268
10483163	Grb14	-0.311438799	5.72E-05	0.002477753
10392936	Nt5c	0.192142686	5.72E-05	0.002477753
10394778	Hpcal1	0.202937216	5.82E-05	0.002516875
10585699	---	0.64868113	5.83E-05	0.002516875
10608715	---	-0.319042392	5.87E-05	0.002528613
10349782	Nuak2	0.210713805	5.88E-05	0.002529097
10532626	Myo18b	-0.410208554	5.91E-05	0.002538047
10343861	---	-0.380800417	5.97E-05	0.002554517
10443120	Ggnbp1	-0.227665055	6.02E-05	0.002573857
10372648	Lyz2	0.752293225	6.06E-05	0.002587634
10543939	Fam180a	-0.170595231	6.08E-05	0.002589853

10346838	Pard3b	-0.185418021	6.12E-05	0.002601794
10503448	Mmp16	0.217376219	6.14E-05	0.002607624
10361906	Il22ra2	-0.300411851	6.19E-05	0.00262262
10587818	Plscr4	-0.223285779	6.35E-05	0.002687868
10383756	---	0.388597838	6.36E-05	0.002687868
10532584	Myo18b	-0.273078059	6.39E-05	0.00269602
10461614	Ms4a6c	0.68648215	6.47E-05	0.002721766
10448202	---	0.358965595	6.51E-05	0.00273313
10589886	4930520O04Rik	-0.214896625	6.56E-05	0.002750734
10489391	Ada	0.187573001	6.63E-05	0.002778061
10585860	Adpgk	0.189197674	6.65E-05	0.002778061
10582337	Piezo1	0.241581204	6.65E-05	0.002778061
10427471	Osmr	0.337068982	6.77E-05	0.002823672
10572747	Tpm4	0.320622792	6.89E-05	0.002870133
10436809	Eva1c	-0.130964207	6.99E-05	0.002903559
10370522	Casp14	0.539779613	7.00E-05	0.002904353
10439289	BC117090	0.518459516	7.07E-05	0.002925872
10368495	Rspo3	-0.270808799	7.09E-05	0.002927482
10517213	Cnksr1	-0.25721141	7.09E-05	0.002927482
10409866	Ctla2b	0.316863642	7.11E-05	0.002929508
10602692	Rragb	-0.16936181	7.23E-05	0.002974692
10440300	---	0.267014201	7.42E-05	0.003038439
10551185	Tgfb1	0.267871523	7.44E-05	0.003040692
10351131	Myoc	-0.694197686	7.46E-05	0.003048036
10357553	Il24	0.357210037	7.49E-05	0.003056727
10340624	---	-0.396739756	7.50E-05	0.003058264
10362422	Trdn	-0.862179088	7.54E-05	0.003064686
10532339	Pxmp2	-0.214063265	7.65E-05	0.003095746
10456836	St8sia5	-0.236287679	7.65E-05	0.003095746
10583312	Taf1d	0.335553136	7.70E-05	0.003108038
10387372	Kdm6b	0.256750951	7.72E-05	0.003113921
10457707	Dsc1	0.622406854	7.82E-05	0.00314273
10606333	Fndc3c1	-0.237251638	7.83E-05	0.003142799
10506424	Actg1	0.3687786	7.84E-05	0.003142799
10349876	Plekha6	-0.179954065	7.86E-05	0.003147907
10596925	Ndufaf3	-0.117727161	7.87E-05	0.003147907
10587616	Prss35	0.395864916	7.93E-05	0.003163174
10362959	Popdc3	-0.252818325	7.94E-05	0.003164002
10414527	Pnp2 /// Pnp	0.183720079	7.98E-05	0.003176745

10605815	Asb12	-0.590591349	8.02E-05	0.003189823
10550482	Igfl3	0.298002122	8.06E-05	0.003194112
10536697	Asb15	-0.280125846	8.07E-05	0.003194112
10425601	Tef	-0.246182668	8.08E-05	0.003194572
10607467	Sat1	0.328079682	8.11E-05	0.003203991
10506714	Lrp8	0.155813978	8.15E-05	0.003213915
10378649	Slc43a2	0.223077737	8.15E-05	0.003213915
10581378	Psmb10	0.150214992	8.24E-05	0.003234608
10364559	Arid3a	0.148015381	8.37E-05	0.003284679
10534085	Phkg1	-0.380412198	8.45E-05	0.003301976
10504902	Murc	-0.356789215	8.45E-05	0.003301976
10472933	Scrn3	-0.257726473	8.54E-05	0.003330796
10371627	Mybpc1	-0.643084525	8.57E-05	0.003340064
10432619	Pou6f1	-0.1531049	8.74E-05	0.003399149
10532578	Myo18b	-0.323538891	8.89E-05	0.003443336
10517250	Extl1	-0.283274989	8.92E-05	0.00344376
10592289	Ccdc15	-0.158771755	8.95E-05	0.00345212
10579307	Kxd1 /// Uba52	-0.127622631	8.99E-05	0.003462059
10341686	---	0.266562899	9.02E-05	0.003464967
10411395	Arhgef28	-0.244074614	9.03E-05	0.003464967
10358664	Hmcn1	0.508587313	9.06E-05	0.003472874
10352661	Ptpn14	-0.272809883	9.11E-05	0.003487504
10564211	Snurf /// Snrpn	-0.280637997	9.27E-05	0.003534766
10607189	Amot	-0.303255021	9.30E-05	0.003541867
10520706	Trim54	-0.290722223	9.47E-05	0.003594564
10605338	G6pdx	0.311753201	9.50E-05	0.003599503
10596680	Sema3b	-0.158950101	9.54E-05	0.00360683
10522009	Pgm1	0.213025426	9.55E-05	0.003608647
10563050	Prr12	-0.180466129	9.69E-05	0.003653187
10359644	Mettl11b	-0.197927273	9.73E-05	0.003663045
10392152	Scn4a	-0.406041313	9.79E-05	0.003683981
10483809	Nfe2l2	0.373660514	9.84E-05	0.003698987
10546829	Oxtr	-0.230296624	9.92E-05	0.003714655
10445338	Enpp5	-0.269068231	9.94E-05	0.003715776
10605256	Flna	0.333304095	9.95E-05	0.003715776
10383233	Rnf213	0.196717666	9.99E-05	0.003725954
10467425	Sorbs1	-0.219647535	0.000101116	0.00376556
10518812	Camta1	-0.218521631	0.000101181	0.00376556
10401997	Ptpn21	-0.209723424	0.000103667	0.00385217

10586244	Dennd4a	0.543013262	0.000104062	0.00385217
10424392	9930014A18Rik	-0.182346673	0.000105545	0.003902923
10434133	Dgcr6	-0.12411165	0.000106147	0.003920909
10362462	Trdn	-0.601586085	0.000106975	0.003942768
10403352	Klf6	0.328786686	0.00010716	0.003942768
10381211	Naglu	0.156903857	0.000107247	0.003942768
10442584	Rpl3l	-0.480902936	0.000107303	0.003942768
10459496	Ccbe1	0.238081286	0.000108796	0.003984983
10391301	Stat3 /// Stat3	0.255982825	0.000109009	0.003988594
10392601	Abca6	-0.233061706	0.000109621	0.004006753
10351298	Gpr161	-0.150027484	0.000110212	0.004024135
10434860	Ostn	-0.308442216	0.000115035	0.004187047
10396730	---	-0.143391359	0.000115999	0.004217704
10386262	Obscn	-0.381301377	0.000116467	0.004230306
10379731	Wfdc18	0.22123651	0.000117236	0.004240843
10461028	Trpt1	-0.135012262	0.000117573	0.004240843
10418950	Ldb3	-0.483919789	0.000117611	0.004240843
10577824	Letm2	-0.128772234	0.000118114	0.004252843
10385118	Dock2	0.396543744	0.000118189	0.004252843
10346260	Osgepl1	-0.199062895	0.000120067	0.004311478
10492341	Arhgef26	-0.181127657	0.00012075	0.004331555
10374035	Xbp1	0.298010801	0.000121339	0.0043482
10375167	Fam196b	-0.228015293	0.000122172	0.004373537
10455212	0610009O20Rik	-0.144911203	0.000123055	0.004387073
10404402	Foxq1	-0.228128339	0.000123294	0.004391093
10588203	Ky	-0.324097762	0.000123916	0.00440871
10554819	Me3	-0.192307009	0.000124594	0.00441942
10532586	Myo18b	-0.384652156	0.000124598	0.00441942
10347218	---	0.48973462	0.00012568	0.004451254
10494085	Selenbp2	-0.19445157	0.000126697	0.00448014
10365627	Sycp3	-0.190469805	0.000126999	0.004481898
10450116	Slc39a7 /// Gm20427	0.127339136	0.000127004	0.004481898
10461856	Gna14	-0.224132043	0.000127629	0.004490256
10583318	Taf1d	0.336990995	0.000128231	0.004506876
10507101	Trabd2b	-0.260058672	0.000128868	0.004520119
10388185	Smtnl2	-0.271007143	0.000130028	0.004556238
10583316	Taf1d	0.504188572	0.000130266	0.004559963
10358585	Hmcn1	0.236944884	0.00013076	0.004572665
10404783	Edn1	0.242010383	0.000130926	0.004573848

10396652	Hspa2	-0.174475363	0.000131506	0.004580371
10573427	G430095P16Rik	-0.157550872	0.000132492	0.004596234
10475144	Ganc /// Capn3	-0.22402144	0.000132798	0.00459766
10357948	Ppp1r12b	-0.243155534	0.000133245	0.004603981
10604076	Snora69	0.312068779	0.000134217	0.004620924
10600372	B230340J04Rik	-0.209915894	0.000136244	0.004670465
10378555	Smyd4	-0.114135984	0.000136638	0.00467936
10362432	Trdn	-0.563758079	0.000137719	0.004711719
10435043	Tm4sf19	0.326138675	0.000142387	0.004847559
10400984	Tmem30b	-0.216697584	0.000143588	0.004867313
10525553	---	-0.20370552	0.0001439	0.004867313
10366938	Stac3	-0.331087502	0.000143947	0.004867313
10363856	2310015B20Rik /// 2310015B20Rik	-0.256059053	0.000144166	0.004869993
10458999	Fbn2	0.291630035	0.000144663	0.004877276
10439357	Fbxo40	-0.255149616	0.000146934	0.00494425
10492021	Postn	0.481625703	0.000147175	0.004944436
10398388	Mir380	0.274714449	0.000149808	0.00500037
10572419	Ell	0.19090879	0.000149896	0.00500037
10500100	Tnfaip8l2	0.19179743	0.000150823	0.005016843
10428983	Fam49b	0.306990132	0.000151439	0.00503224
10496748	Syde2	-0.162706251	0.000151641	0.00503224
10375065	Sh3pxd2b	0.198111137	0.00015403	0.005091014
10532574	Myo18b	-0.347836911	0.000154445	0.005098313
10386238	Obscn	-0.42885953	0.000155339	0.005122952
10583090	Mmp10	0.277569321	0.000156865	0.005153723
10359793	Dusp27	-0.254251463	0.000158043	0.00518263
10554839	Picalm	0.36474787	0.000158802	0.005197712
10418842	3425401B19Rik	-0.534474199	0.000159678	0.005209615
10549097	Ldhb	-0.365184624	0.000159921	0.005209615
10384780	6820445E23Rik	-0.172054823	0.000160802	0.005228707
10493903	Lce3d	0.368732065	0.000161341	0.005237616
10497381	Cyp7b1	0.447265242	0.000161378	0.005237616
10435019	Smco1	-0.302661597	0.000163747	0.005289814
10559919	2810047C21Rik1 /// Zfp772 /// Gm20482 /// Gm3912	-0.285436581	0.000163983	0.005292525
10552469	Klk13	0.488111425	0.000166599	0.005371953
10353192	Eya1	-0.297521217	0.000168211	0.005407579
10499095	Fam160a1	-0.178102256	0.000168483	0.005407642
10458875	Dtwd2	-0.126524974	0.000174862	0.005581439
10515994	Smap2	0.187751071	0.000177272	0.00565317

10359867	Lrrc52	-0.234786854	0.000179604	0.005706613
10380887	Tcap	-0.51125758	0.000181466	0.005735865
10369301	Chst3	-0.17764962	0.00018168	0.005735865
10447591	Ftl1	0.35556077	0.00018221	0.005742177
10594988	Mapk6	0.235309939	0.000182905	0.005748689
10357954	Ppp1r12b	-0.241748452	0.000183458	0.005760609
10438575	Ehhadh	-0.198429278	0.000184087	0.00577514
10349968	Chi3l1	0.588713819	0.00018571	0.00581556
10369715	Mypn	-0.524846195	0.000186464	0.005833911
10469571	Otud1	-0.246890298	0.000188604	0.00588809
10495539	Extl2	-0.131534104	0.000188704	0.00588809
10546624	Lmod3	-0.442361175	0.00019086	0.005934061
10362420	Trdn	-0.685369073	0.000192068	0.005960959
10338923	---	0.449712849	0.000194524	0.006015714
10376868	Trpv2	0.235352302	0.000199272	0.006146127
10392221	Pecam1	0.456832147	0.000201381	0.006189209
10477813	---	-0.179024667	0.000204782	0.006278494
10408543	Mylk4	-0.757933472	0.000204828	0.006278494
10450699	Gm7030 /// Gm11127	0.290010139	0.000205943	0.006290086
10488060	Jag1	0.277988232	0.000206048	0.006290086
10552919	Hrc	-0.357410432	0.000206051	0.006290086
10530371	Yipf7	-0.308150408	0.000208313	0.006335047
10374529	Wdpcp	-0.12640172	0.000210558	0.006375966
10454254	Dtna	-0.262528095	0.000210576	0.006375966
10461979	Aldh1a1	-0.331646183	0.000212356	0.00641312
10599487	Sash3	0.186383214	0.000214333	0.006461597
10360001	---	0.199065184	0.000218183	0.006560597
10525932	Tmem132c	0.146717672	0.000219805	0.006597934
10580649	Ces1e	-0.251721455	0.000221879	0.006646109
10476728	Dtd1	-0.107191296	0.000222338	0.006651008
10418016	Dusp13	-0.235340045	0.000223753	0.006681828
10517401	Grhl3	0.269993296	0.000224534	0.006693659
10478594	Ctsa	0.273026329	0.000229204	0.006798936
10572485	Rab3a	-0.138457148	0.000229626	0.006798936
10501302	Sypl2	-0.312300593	0.00022963	0.006798936
10346448	Aox4	0.694386123	0.000232717	0.006861094
10545372	Atoh8	-0.178241629	0.000236979	0.006933821
10485388	Ldlrad3	-0.273403283	0.000237732	0.006950003
10464391	Emx2	-0.238933966	0.000241803	0.007038817

10449545	---	-0.1668789	0.000243222	0.007062992
10551365	Prx	-0.136308231	0.000244203	0.007085544
10491106	Pld1	0.175502157	0.000247573	0.007171375
10438854	Atp13a4	0.308874111	0.000250705	0.007238288
10431210	Wnt7b	-0.152120926	0.000255369	0.007355473
10398390	Mir323	0.250881631	0.000261302	0.007486135
10345492	Cnnm3	-0.114911212	0.000265224	0.007575469
10498371	P2ry12	0.217231267	0.000267917	0.007602292
10553140	Tmem143	-0.142545734	0.000274665	0.007749678
10587880	Pcolce2	-0.378394483	0.000279928	0.007866049
10546349	Xpc	-0.189391043	0.000284295	0.007937435
10593413	2310030G06Rik	-0.138222473	0.000289631	0.008066955
10569017	Ifitm3	0.529663669	0.000293198	0.008140224
10477206	Mylk2	-0.489405713	0.000296924	0.008210882
10490304	Sycp2	0.287157741	0.000298751	0.008241729
10474520	Tdh /// Gm13929	0.370017055	0.000300631	0.008280445
10594183	Senp8	-0.159882008	0.000303325	0.008341423
10402575	Degs2	0.264763411	0.000304929	0.008369313
10405729	---	0.179543335	0.000308004	0.008416844
10577858	Bag4	-0.172460594	0.000309036	0.008437169
10555681	Stim1	-0.238597073	0.000309524	0.008438484
10345482	Cnnm4	-0.115000997	0.000314927	0.008536711
10508953	Trim63	-0.298293418	0.000316391	0.008558564
10450484	Aif1	0.220357832	0.000322098	0.008674974
10588509	Pcbp4	-0.156805156	0.000323188	0.008674974
10559837	Vmn2r29 /// Vmn2r46 /// Gm3912	-0.339220545	0.000323371	0.008674974
10582376	Piezo1	0.258186228	0.000326738	0.008749941
10497001	Cryz	-0.115343471	0.000331062	0.008831698
10529402	Tnip2	0.129273117	0.000339779	0.009002075
10487945	Gpcpd1	0.340103078	0.000345344	0.009073447
10338491	---	-0.253990562	0.000347742	0.009122643
10379650	---	0.365820492	0.000348614	0.009138637
10504692	Tmod1	-0.466422777	0.000353867	0.009241492
10605811	Amer1	-0.138673658	0.000355628	0.009273527
10532580	Myo18b	-0.225666609	0.000357483	0.009307957
10555460	Stard10	-0.183123973	0.000362667	0.009400669
10463517	Pprc1	0.185131946	0.000366917	0.009482223
10565193	Hdgfrp3	-0.192244326	0.000367845	0.009483285
10426891	Mettl7a1	-0.231174648	0.00037374	0.009601777

10563766	---	-0.312659862	0.000376736	0.009656447
10496405	---	-0.27993514	0.000377168	0.009656447
10359307	Tnn	0.263910388	0.000377931	0.00966017
10376564	Med9	-0.115347469	0.000382567	0.009742148
10396112	Atp5s	-0.1476935	0.000383931	0.009755507
10392687	BC006965	-0.158147342	0.000385531	0.009760409
10377148	Myh8	-0.227759757	0.000388852	0.009823044
10589909	C130032M10RIK	-0.167787502	0.000391983	0.009880598
10435697	Popdc2	-0.256124191	0.000398032	0.009989593
10466606	Anxa1	0.402785827	0.000400947	0.010026551
10468974	Cdnf	-0.163890456	0.000405777	0.010131923
10501963	Ugt8a	-0.235392247	0.000408282	0.010173397
10462697	---	-0.169416867	0.000411978	0.010250799
10459455	Alpk2	-0.306377994	0.000412843	0.010257636
10390507	Fbxo47	-0.188760011	0.000414794	0.010284045
10417700	---	0.199831667	0.000440204	0.010722835
10575095	Has3	0.225939	0.00044496	0.010793599
10385776	Tcf7	-0.140639583	0.000445647	0.010801464
10449807	Ephx3	0.423823474	0.000445995	0.010801464
10441339	A630089N07Rik	0.26260589	0.000446229	0.010801464
10596117	Cep63	-0.184242537	0.000452323	0.010911
10538087	Lrrc61	-0.170199182	0.000456026	0.010992687
10456254	Nedd4l	-0.243780073	0.000459201	0.011038617
10592336	Spa17	-0.173513292	0.00046378	0.011107063
10375537	Mgat1 /// Mgat1	0.129538553	0.00046784	0.011192094
10503023	Cth	-0.195854152	0.000474947	0.011307633
10494043	Tdrkh	-0.160962661	0.000492682	0.011665959
10490502	Tcfl5	0.157133683	0.000494447	0.011691823
10512156	Aqp3	0.37211728	0.000497162	0.011742313
10431659	Kif21a	-0.306798022	0.000500545	0.011803897
10357072	Serpinb3a	0.635663232	0.000509696	0.011962889
10419073	Tspan14	0.177554747	0.000516798	0.012105063
10395672	Ap4s1	-0.124105841	0.000556006	0.012799195
10400967	Six1	-0.302654049	0.000578048	0.013245194
10493108	Crabp2	0.300340904	0.000580087	0.013272215
10508436	Sync	-0.251769063	0.000584524	0.013349567
10496796	Ssx2ip	-0.153392547	0.000602156	0.013671423
10447224	Dync2li1	-0.138993913	0.000613325	0.013883711
10515861	4930538K18Rik /// AU022252	-0.14979231	0.000627681	0.014098016

10554367	Mesp2	-0.158631725	0.000629895	0.014125901
10424543	Wisp1	0.198064937	0.000639314	0.014300238
10568436	Fgfr2	-0.309787942	0.000642768	0.014359006
10590821	9230110C19Rik	-0.176959322	0.000645817	0.01439012
10433373	Sec14l5	-0.139601206	0.000646863	0.014394968
10588786	Uba7 /// Gm20661 /// Cdhr4	0.128155703	0.000655047	0.014534112
10432682	Krt80	-0.252409628	0.000655763	0.014534112
10342466	---	0.177381233	0.00065707	0.014538346
10422962	Nadk2	-0.142786488	0.000672108	0.014786364
10513362	Susd1	0.190832196	0.000690437	0.015094082
10409414	Rab24	0.162186139	0.000692141	0.015110591
10601569	Pcdh11x	-0.221814757	0.000708901	0.015354759
10389680	Msi2	-0.158245749	0.000744432	0.015865388
10526514	Cldn15	0.166232999	0.000762028	0.016190766
10381006	Thra	-0.217682313	0.000764838	0.016232565
10607841	Tceanc	-0.147283632	0.000784447	0.016495694
10339611	---	-0.269631169	0.000788971	0.016547434
10341641	---	0.235142131	0.000801117	0.01671478
10570483	Arhgef10	-0.19684312	0.00080748	0.016827331
10548871	Smco3	-0.183548484	0.000814196	0.0169267
10398360	Rian	0.300862713	0.000848004	0.017474091
10477854	Epb4.1l1 /// Epb4.1l1	-0.105022622	0.000910136	0.018318659
10407209	Slc38a9	-0.15785544	0.00091239	0.018353401
10433433	Mettl22 /// Mettl22	-0.179786082	0.0009192	0.018453016
10604508	Frmd7	-0.163580614	0.000938235	0.01874313
10591947	Acad8	-0.133125121	0.000950114	0.018897004
10605328	Fam3a /// Fam3a	-0.134880267	0.000968703	0.019219251
10387983	Gm12318	0.172514019	0.00097513	0.019302689
10411532	Mccc2	-0.153243573	0.000985715	0.019435448
10585428	Dnaja4	-0.179520124	0.00104103	0.020212902
10453887	Cables1	-0.189012919	0.001142528	0.021630871
10607317	Tsr2	-0.150716358	0.001183886	0.022208239
10607300	Gm10437	-0.275093048	0.001253654	0.023056628
10476969	Pygb	-0.195687804	0.001454207	0.025590942
10445006	Gm6623	0.142237152	0.001476155	0.025810431
10480842	Tmem141	-0.130294392	0.001537491	0.026575436
10561025	Cnfn	0.36661943	0.001547351	0.026692766
10497349	LOC100038947 /// Sirpb1b /// Sirpb1a	0.342661229	0.001572556	0.027060416
10406417	Actg1	0.400587083	0.001606362	0.027398015

10503399	---	0.166729266	0.001715133	0.028626392
10537712	Gstk1	-0.238704334	0.001723502	0.028747086
10374455	Spred2	-0.212127785	0.001756039	0.029149906
10408975	Kif13a	-0.185845405	0.001883616	0.030566987
10553967	Pcsk6	-0.1691985	0.001970468	0.031491251
10392207	Tex2	-0.195162434	0.002102469	0.032981645
10380560	Zfp652	-0.188964657	0.002306811	0.03499495
10592926	Tmem25	-0.148622247	0.002363763	0.035615263
10510482	Clstn1	-0.182139796	0.002836685	0.040778813
10452030	Plin3	-0.110975266	0.002856798	0.040916515
10386495	Tom1l2	-0.135420812	0.003589772	0.048103492

Supplement Table 4: The table shows the most significant eQTL. In table column Probe ID is the Affymetrix probe ID, Peak SNP is the SNP with the highest –log *p-value*, and CI is the confidence interval defined by a drop by 1.5 of the –log P score.

Probe ID	Peak SNP	Chr	Peak (SNP)	P(log)	C.I.	Length QTL (Mb)	Gene Name
10552090	rs3719311	7	35	9.8	0-60	21.92	Scgb2b2
10592336	rs13480173	9	46	9.75	18-61	21.81	Spa17
10510482	rs3688566	4	141	9.39	135-156	21.14	Clstn1
10423471	CEL.15_36490596	15	36	9.18	3-61	20.74	Ctnnd2
10517250	UT_4_132.137715	4	133	9.13	128-135	20.64	Extl1
10528159	rs3658783	6	144	9.01	142-149	20.43	Gm10482
10379646	rs13481127	11	83	8.89	71-109	20.19	Slfn3
10386495	rs3688566	4	141	8.67	135-156	19.78	Tom1l2
10587880	rs13478002	4	136	8.61	135-142	19.66	Pcolce2
10392207	rs13481161	11	92	8.59	88-110	19.62	Tex2
10528159	rs4222295	1	39	8.56	13-79	19.56	Gm10482
10400510	gnf12.077.067	12	80	8.54	47-108	19.51	Clec14a
10605328	gnfX.084.751	X	98	8.45	10-135	19.35	Fam3a /// Fam3a
10572693	rs13479880	8	89	8.34	48-105	19.14	Jak3 /// Insl3
10427471	rs3658783	6	144	8.15	136-149	18.77	Osmr
10488060	rs6329892	6	142	8.12	136-147	18.7	Jag1
10553967	rs13478002	4	136	8.12	128-156	18.72	Pcsk6

10559919	CEL.7_122752866	7	142	8.07	135-147	18.61	2810047C21Rik1 /// Zfp772 /// Gm20482 /// Gm3912
10445338	rs3693494	17	30	8.02	21-55	18.52	Enpp5
10387983	rs3658783	6	144	7.87	140-149	18.22	Gm12318
10411395	gnf13.093.328	13	93	7.79	74-117	18.07	Arhgef28
10391301	rs6329892	6	142	7.7	136-147	17.88	Stat3 /// Stat3
10452030	rs3023025	4	143	7.68	135-155	17.84	Plin3
10446986	rs13478002	4	136	7.66	128-156	17.81	Crim1
10391066	rs6265387	6	147	7.64	136-149	17.76	Krt17
10572146	rs3672808	6	140	7.63	136-147	17.74	Atp6v1b2 /// Atp6v1b2
10605256	rs6329892	6	142	7.6	136-147	17.69	Flna
10527732	rs13478002	4	136	7.55	135-147	17.59	Fry
10368277	rs13476689	2	107	7.55	91-130	17.59	Rps12 /// Snora33
10381006	rs13478002	4	136	7.49	135-147	17.48	Thra
10445338	CEL.2_135876979	2	136	7.46	87-175	17.41	Enpp5
10386495	rs3023251	11	21	7.4	12-65	17.29	Tom1l2
10456254	rs13478002	4	136	7.39	135-143	17.27	Nedd4l
10407792	rs13481689	13	11	7.36	0-26	17.21	Gpr137b-ps /// Gpr137b
10577824	mCV24845756	8	22	7.26	8-46	17.02	Letm2
10582997	rs13478952	6	106	7.23	104-112	16.96	Casp4
10604508	rs3023025	4	143	7.21	135-156	16.91	Frmd7
10384780	rs3723990	11	27	7.19	12-38	16.88	6820445E23Rik
10474836	rs6228179	2	123	7.17	116-130	16.84	Ivd
10572747	rs13479071	6	138	7.15	135-144	16.79	Tpm4
10360248	rs3707910	1	174	7.13	164-178	16.75	Atp1a4
10449807	rs6386362	11	107	7.13	90-114	16.75	Ephx3
10346838	rs6404446	1	21	7.11	13-44	16.71	Pard3b
10346448	rs6386920	1	54	7.05	33-98	16.59	Aox4
10526514	rs3673363	5	36	6.99	28-83	16.47	Cldn15
10538087	rs13478732	6	45	6.92	45-51	16.32	Lrrc61
10396112	rs13481514	12	71	6.9	54-77	16.3	Atp5s
10488060	rs13478971	6	111	6.89	110-116	16.28	Jag1
10412260	rs6386362	11	107	6.87	105-112	16.22	Fst
10395672	rs13481408	12	41	6.87	10-80	16.22	Ap4s1
10349876	rs6228473	1	129	6.78	114-137	16.04	Plekha6
10412260	rs13481161	11	92	6.75	90-103	15.99	Fst
10430201	rs13479071	6	138	6.7	135-149	15.87	Myh9
10452815	rs13483103	17	75	6.69	63-87	15.86	Xdh
10486858	rs3023025	4	143	6.64	135-156	15.75	Mfap1b
10582997	rs3672808	6	140	6.63	134-149	15.73	Casp4

10439936	rs13475827	1	41	6.63	33-43	15.75	Nfkbiz
10419073	rs13482193	14	56	6.58	40-72	15.64	Tspan14
10587880	rs6181382	6	81	6.55	59-94	15.58	Pcolce2
10521984	rs6400804	15	57	6.55	43-68	15.58	G6pd2
10408975	rs3688566	4	141	6.54	135-156	15.55	Kif13a
10441339	rs4211364	16	81	6.54	77-85	15.56	A630089N07Rik
10450699	rs3702604	17	41	6.54	33-42	15.56	Gm7030 /// Gm11127
10544732	rs3683997	1	36	6.54	23-69	15.56	Skap2
10466606	rs3672808	6	140	6.52	135-148	15.52	Anxa1
10346838	gnf01.075.385	1	78	6.51	63-81	15.5	Pard3b
10505073	rs13477676	4	45	6.51	40-55	15.5	Zfp462
10415319	rs13482193	14	56	6.49	42-70	15.46	Irf9
10375065	rs6329892	6	142	6.48	135-149	15.44	Sh3pxd2b
10525256	rs13478451	5	109	6.48	97-135	15.44	Tmem116
10525397	rs3658783	6	144	6.47	135-149	15.42	Arpc3 /// Arpc3
10352320	rs13478002	4	136	6.47	135-142	15.41	Tmem63a
10487945	rs6329892	6	142	6.46	135-147	15.39	Gpcpd1
10505073	rs13477774	4	73	6.45	62-83	15.37	Zfp462
10485388	rs13478002	4	136	6.44	135-141	15.36	Ldlrad3
10346448	rs3718160	16	77	6.42	69-81	15.31	Aox4
10585860	rs4222295	1	39	6.42	34-43	15.31	Adpgk
10393754	rs3672808	6	140	6.39	135-147	15.26	Actg1
10386965	rs13481071	11	65	6.37	61-67	15.22	Arhgap44
10555681	rs3023025	4	143	6.35	141-154	15.16	Stim1
10360248	rs13476273	1	184	6.34	183-191	15.15	Atp1a4
10568873	rs3683997	1	36	6.33	31-43	15.12	Adam8
10596925	rs13480421	9	112	6.31	102-113	15.08	Ndufaf3
10583316	rs6386362	11	107	6.31	103-114	15.09	Taf1d
10384780	rs6329892	6	142	6.29	135-147	15.05	6820445E23Rik
10466606	rs3683997	1	36	6.28	33-43	15.03	Anxa1
10445338	rs3662820	17	13	6.28	0-18	15.02	Enpp5
10352143	rs13478971	6	111	6.27	104-117	15.01	Kif26b
10449807	rs4201998	16	73	6.27	69-81	15.01	Ephx3
10485405	rs6329892	6	142	6.26	135-149	14.98	Cd44
10583316	rs3672808	6	140	6.26	135-148	14.99	Taf1d
10406417	rs6329892	6	142	6.25	136-148	14.96	Actg1
10545372	rs3698446	11	105	6.24	88-114	14.95	Atoh8
10364375	rs3658783	6	144	6.23	138-149	14.93	Cstb
10392936	rs3658783	6	144	6.23	138-149	14.92	Nt5c

10515861	rs6355837	4	129	6.22	128-135	14.91	4930538K18Rik /// AU022252
10370721	rs4222295	1	39	6.21	33-43	14.89	Sbno2
10575799	rs13475827	1	41	6.19	34-44	14.84	Plcg2
10398390	rs6215373	5	45	6.18	31-51	14.83	Mir323
10476969	rs13478002	4	136	6.17	128-156	14.79	Pygb
10583318	rs13476689	2	107	6.16	100-114	14.77	Taf1d
10346448	rs13481161	11	92	6.15	88-99	14.75	Aox4
10507671	rs13477959	4	124	6.14	118-128	14.74	Guca2a
10389680	rs13478002	4	136	6.12	135-147	14.69	Msi2
10515994	rs6208251	6	105	6.12	104-111	14.71	Smap2
10400984	rs13481514	12	71	6.11	51-82	14.68	Tmem30b
10370721	rs3658783	6	144	6.1	134-149	14.65	Sbno2
10346191	rs3672808	6	140	6.09	134-148	14.63	Stat1
10427471	rs13478971	6	111	6.07	104-117	14.59	Osmr
10558150	rs13479071	6	138	6.06	135-147	14.58	Htra1
10400126	rs13481408	12	41	6.05	41-45	14.55	Lrrn3
10546829	rs3658927	2	121	6.04	116-130	14.53	Oxtr
10373740	rs6181382	6	81	6.04	61-94	14.52	Pik3ip1
10469856	gnf02.035.469	2	34	6.03	16-51	14.51	Wdr85
10607300	rs13481161	11	92	5.98	90-100	14.42	Gm10437
10521205	rs3700706	5	33	5.98	21-46	14.4	Sh3bp2
10491300	rs3658783	6	144	5.97	136-149	14.38	Skil
10467425	rs3023025	4	143	5.97	141-151	14.39	Sorbs1
10575799	rs6377872	8	125	5.96	92-130	14.36	Plcg2
10373740	rs3719217	15	95	5.95	90-101	14.35	Pik3ip1
10483809	rs3658783	6	144	5.93	136-147	14.31	Nfe2l2
10559837	rs3023025	4	143	5.93	141-154	14.3	Vmn2r29 /// Vmn2r46 /// Gm3912
10352143	rs13481173	11	96	5.93	90-100	14.3	Kif26b
10541246	rs4222295	1	39	5.92	33-43	14.28	Il17ra
10543939	rs4211364	16	81	5.91	71-86	14.26	Fam180a
10548871	rs13479071	6	138	5.9	131-149	14.24	Smco3
10583316	rs6341620	5	37	5.9	29-45	14.25	Taf1d
10569017	rs6329892	6	142	5.89	134-147	14.22	Ifitm3
10364593	rs3672808	6	140	5.89	134-148	14.21	Cnn2
10526514	rs6355445	9	70	5.88	63-79	14.21	Cldn15
10452815	rs3658783	6	144	5.87	138-149	14.18	Xdh
10352143	rs13479071	6	138	5.87	136-144	14.18	Kif26b
10432619	rs6285067	15	95	5.87	92-103	14.18	Pou6f1
10546829	rs13476810	2	142	5.86	136-152	14.16	Oxtr

10394770	rs6329892	6	142	5.85	135-147	14.14	Odc1
10515861	rs6381371	4	116	5.84	106-121	14.13	4930538K18Rik /// AU022252
10362959	rs3716113	10	42	5.83	31-64	14.1	Popdc3
10392207	rs3723990	11	27	5.83	18-31	14.09	Tex2
10412260	rs3658783	6	144	5.82	136-149	14.07	Fst
10452030	rs13477749	4	66	5.82	62-106	14.08	Plin3
10492021	rs6329892	6	142	5.8	135-147	14.03	Postn
10392207	rs6199956	11	51	5.79	41-67	14	Tex2
10537712	rs3023025	4	143	5.78	141-151	14	Gstk1
10432852	rs3672808	6	140	5.78	136-149	14	Krt1
10374035	rs13481161	11	92	5.78	90-96	13.98	Xbp1
10488060	rs13481161	11	92	5.77	90-99	13.97	Jag1
10464999	rs13475919	1	73	5.75	65-79	13.93	Cst6
10392936	rs4222295	1	39	5.75	25-60	13.92	Nt5c
10468974	rs3088801	9	25	5.75	12-29	13.93	Cdnf
10383233	rs3699056	11	114	5.74	109-116	13.9	Rnf213
10394770	rs13478952	6	106	5.74	104-114	13.91	Odc1
10462140	rs13478068	4	155	5.73	151-156	13.88	Dock8
10360248	CEL.X_77780392	X	83	5.73	82-99	13.89	Atp1a4
10441339	rs3671849	2	163	5.72	151-170	13.86	A630089N07Rik
10583316	rs13478971	6	111	5.72	104-118	13.87	Taf1d
10526514	CEL.5_93945748	5	97	5.71	97-100	13.84	Cldn15
10604076	rs3711088	6	148	5.7	140-149	13.82	Snora69
10460616	rs6329892	6	142	5.69	135-148	13.8	Cfl1
10463517	rs3658783	6	144	5.68	136-149	13.78	Pprc1
10473240	rs13479071	6	138	5.68	135-148	13.78	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735
10455595	rs3672808	6	140	5.66	135-147	13.74	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735
10554789	rs6329892	6	142	5.64	135-147	13.69	Ctsc /// Ctsc
10370522	rs6386362	11	107	5.63	100-114	13.68	Casp14
10591988	rs4200124	16	71	5.62	66-77	13.65	Adamts15
10368356	rs3679120	10	23	5.62	14-66	13.65	Akap7
10447951	rs6329892	6	142	5.61	134-147	13.64	Thbs2
10607300	rs3688566	4	141	5.61	136-154	13.63	Gm10437
10447951	rs13478971	6	111	5.61	105-114	13.65	Thbs2
10400126	rs3089800	12	30	5.61	17-39	13.64	Lrrn3
10384423	rs3023251	11	21	5.61	15-30	13.65	Cobl
10596117	rs3658783	6	144	5.6	135-149	13.62	Cep63
10404702	rs3683997	1	36	5.6	31-46	13.62	Gcnt2

10380403	rs6329892	6	142	5.59	135-148	13.6	Lrrc59
10504402	rs6181382	6	81	5.57	73-94	13.55	Tmem8b
10586591	rs3700706	5	33	5.57	29-41	13.56	Car12
10363455	rs4138572	6	98	5.56	97-112	13.53	Pcbd1
10464999	rs13481161	11	92	5.55	87-96	13.52	Cst6
10592926	rs13479071	6	138	5.53	120-149	13.48	Tmem25
10423471	rs3667334	5	83	5.53	71-109	13.47	Ctnnd2
10357946	rs3023025	4	143	5.52	141-151	13.45	Ppp1r12b
10564211	rs3023025	4	143	5.52	135-147	13.46	Snurf /// Snrpn
10374035	rs6329892	6	142	5.5	136-147	13.4	Xbp1
10487945	rs13478971	6	111	5.5	104-116	13.42	Gpcpd1
10521205	rs3711088	6	148	5.49	142-149	13.39	Sh3bp2
10374035	rs6370458	11	109	5.48	105-114	13.37	Xbp1
10463517	rs6316774	2	97	5.47	87-119	13.34	Pprc1
10582378	rs6268443	1	95	5.47	65-105	13.34	Piezo1
10392936	rs13481161	11	92	5.47	90-99	13.34	Nt5c
10419578	rs3023025	4	143	5.46	141-154	13.33	Ndrg2
10488060	rs6386362	11	107	5.46	105-114	13.32	Jag1
10391301	rs4222295	1	39	5.45	33-44	13.31	Stat3 /// Stat3
10607300	rs13480854	11	8	5.45	4-12	13.3	Gm10437
10554839	rs6329892	6	142	5.44	136-147	13.28	Picalm
10490838	rs4201998	16	73	5.44	69-81	13.28	Fabp5
10583669	rs6174757	9	68	5.44	58-80	13.27	AB124611
10425781	rs13478089	4	154	5.43	136-156	13.25	Serhl
10409414	rs13475827	1	41	5.43	34-44	13.26	Rab24
10565193	rs3688566	4	141	5.42	135-151	13.23	Hdgfrp3
10469581	rs13478002	4	136	5.42	135-147	13.24	Etl4
10559837	rs13482170	14	49	5.42	36-68	13.25	Vmn2r29 /// Vmn2r46 /// Gm3912
10432852	rs3683997	1	36	5.42	31-43	13.24	Krt1
10528159	rs13481161	11	92	5.4	90-96	13.2	Gm10482
10559919	rs3720897	19	25	5.4	20-30	13.19	2810047C21Rik1 /// Zfp772 /// Gm20482 /// Gm3912
10538087	rs6330932	6	37	5.39	29-41	13.17	Lrrc61
10541260	rs13478971	6	111	5.38	105-138	13.15	Cecr2
10480842	rs3658783	6	144	5.36	142-149	13.11	Tmem141
10554789	rs3711357	11	61	5.36	51-65	13.11	Ctsc /// Ctsc
10352143	rs6386362	11	107	5.35	103-110	13.09	Kif26b
10580649	rs13482193	14	56	5.35	32-69	13.1	Ces1e
10456254	rs13478089	4	154	5.34	151-156	13.08	Nedd4l
10490773	rs3672808	6	140	5.34	134-148	13.08	Hnrnph2 /// Hnrnph2 /// Gm8775

10424392	rs13475827	1	41	5.33	31-46	13.05	9930014A18Rik
10487945	rs3723990	11	27	5.33	21-41	13.05	Gpcpd1
10559837	rs13478451	5	109	5.32	98-125	13.02	Vmn2r29 /// Vmn2r46 /// Gm3912
10373396	rs6265387	6	147	5.31	135-149	13	Myl6
10473240	rs3023025	4	143	5.31	135-155	13.02	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735
10586244	rs3672808	6	140	5.31	135-147	13.01	Dennd4a
10419073	rs3676039	3	136	5.31	125-146	13	Tspan14
10464391	rs3023025	4	143	5.3	135-151	12.98	Emx2
10387180	rs4222295	1	39	5.29	33-43	12.96	Ndel1 /// Ndel1
10554789	rs6268443	1	95	5.28	90-100	12.95	Ctsc /// Ctsc
10428983	rs3683997	1	36	5.28	33-43	12.95	Fam49b
10605571	rs3683997	1	36	5.28	31-43	12.95	Gyk
10489107	rs3683997	1	36	5.28	31-43	12.95	Samhd1
10435641	rs6329892	6	142	5.27	134-147	12.93	Fstl1
10419073	rs6329892	6	142	5.27	136-147	12.92	Tspan14
10577824	rs3719217	15	95	5.27	90-101	12.92	Letm2
10425601	rs6271003	4	93	5.27	76-103	12.93	Tef
10455595	rs13483783	X	61	5.27	36-71	12.93	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735
10455595	rs6268364	4	151	5.25	135-155	12.87	Eno1 /// Gm5506 /// LOC101056352 /// Gm4735
10445753	mCV24625340	13	85	5.25	81-93	12.88	Trem3
10535174	rs3705275	8	106	5.24	105-113	12.86	Tmem184a
10430201	rs3683997	1	36	5.24	33-43	12.87	Myh9
10590631	mCV23053483	9	7	5.24	0-9	12.86	Ccr2
10370721	rs13478952	6	106	5.23	104-110	12.85	Sbno2
10583316	rs13481161	11	92	5.23	90-99	12.84	Taf1d
10403584	rs4200124	16	71	5.23	66-79	12.84	Nid1
10532339	rs13478801	6	65	5.23	59-71	12.83	Pxmp2
10596117	rs13482170	14	49	5.23	36-53	12.84	Cep63
10492021	rs3700706	5	33	5.23	29-36	12.83	Postn
10457022	rs13478002	4	136	5.22	135-155	12.82	Mbp
10380560	rs4139273	4	81	5.22	63-90	12.82	Zfp652
10572693	rs3714664	9	41	5.22	22-48	12.81	Jak3 /// Insl3
10463517	rs13478971	6	111	5.21	105-127	12.79	Pprc1
10528159	rs13475980	1	94	5.21	81-99	12.79	Gm10482
10476969	rs6315002	4	103	5.2	90-106	12.78	Pygb
10506254	rs6316774	2	97	5.2	91-119	12.78	Raver2
10594988	rs13481161	11	92	5.2	90-100	12.78	Mapk6
10487945	rs4222295	1	39	5.2	33-50	12.77	Gpcpd1

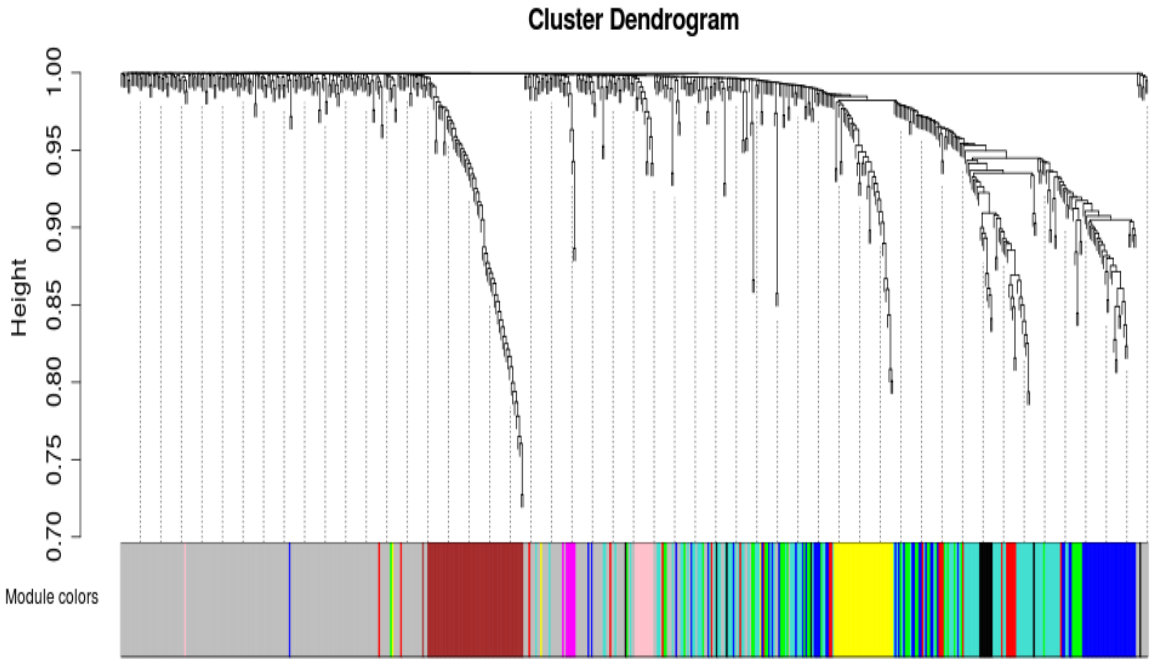
10407797	rs13483712	X	9	5.2	0-10	12.77	Prl2c4 /// Prl2c2 /// Prl2c3
10425601	rs3659836	3	121	5.19	109-126	12.76	Tef
10508392	rs6208251	6	105	5.19	104-111	12.75	Rnf19b
10360248	rs13459097	6	85	5.19	78-94	12.75	Atp1a4
10585860	rs3676039	3	136	5.18	126-143	12.74	Adpgk
10469581	rs6265423	2	47	5.18	20-51	12.72	Etl4
10432682	rs13482744	15	101	5.17	92-103	12.72	Krt80
10546829	rs4231720	17	87	5.17	73-95	12.71	Oxtr
10528159	rs3719581	11	87	5.17	82-88	12.71	Gm10482
10396652	rs13481521	12	73	5.17	51-80	12.71	Hspa2
10509514	rs13476439	2	38	5.17	26-51	12.71	Sh2d5
10401473	rs13478002	4	136	5.16	135-147	12.68	Aldh6a1
10519717	mCV23053483	9	7	5.16	0-11	12.69	Sema3a
10366528	rs3089366	10	89	5.15	85-98	12.67	Best3
10439296	rs3695416	15	38	5.15	28-52	12.68	Stfa2
10508392	rs3683997	1	36	5.15	33-43	12.66	Rnf19b
10585428	rs6230717	4	147	5.14	142-154	12.65	Dnaja4
10579347	rs6329892	6	142	5.14	134-148	12.65	Ifi30
10398360	rs3688566	4	141	5.14	135-156	12.65	Rian
10600144	rs6194426	19	50	5.14	43-55	12.66	F8a
10419073	rs4222295	1	39	5.14	33-46	12.65	Tspan14
10384064	rs3023025	4	143	5.13	141-154	12.63	Camk2b
10403352	rs3672808	6	140	5.13	135-148	12.63	Klf6
10358660	rs4201998	16	73	5.13	66-81	12.64	Hmcn1
10554789	rs13478801	6	65	5.13	59-97	12.63	Ctsc /// Ctsc
10480842	rs8239888	5	122	5.12	100-130	12.61	Tmem141
10577858	rs13481161	11	92	5.12	88-109	12.61	Bag4
10396831	mCV24625340	13	85	5.12	76-93	12.61	Arg2
10607361	rs13478626	6	10	5.12	4-18	12.6	Gm4750 /// Gm8464
10411395	rs3688566	4	141	5.11	135-143	12.59	Arhgef28
10581813	rs4201998	16	73	5.11	69-81	12.59	Mlkl
10559919	rs13482194	14	56	5.11	42-69	12.58	2810047C21Rik1 /// Zfp772 /// Gm20482 /// Gm3912
10492997	rs3659933	5	46	5.11	31-51	12.59	Etv3
10518570	rs13475931	1	76	5.1	65-99	12.57	Pgd
10582337	rs6329892	6	142	5.09	135-149	12.53	Piezo1
10523128	rs3686443	13	87	5.09	76-93	12.53	Ppbp
10364950	mCV24625340	13	85	5.09	76-93	12.55	Gadd45b
10487945	rs3700706	5	33	5.09	29-41	12.54	Gpcpd1
10553140	rs13478068	4	155	5.08	142-156	12.52	Tmem143

10586250	rs3658783	6	144	5.08	135-149	12.52	Dennd4a
10594988	rs6370458	11	109	5.08	105-112	12.51	Mapk6
10354432	rs4201998	16	73	5.08	67-81	12.51	Myo1b
10449775	rs4201998	16	73	5.08	66-83	12.53	Notch3
10525397	rs3683997	1	36	5.08	31-43	12.51	Arpc3 /// Arpc3
10580649	rs3670168	3	128	5.07	125-138	12.5	Ces1e
10364593	rs3683997	1	36	5.07	31-44	12.5	Cnn2
10533844	rs13481908	13	82	5.06	69-90	12.48	Rilpl2
10490838	rs13475931	1	76	5.06	69-99	12.47	Fabp5
10511416	rs3681732	1	34	5.06	6-43	12.47	Tox
10489107	rs3672808	6	140	5.05	135-147	12.45	Samhd1
10554789	rs13478971	6	111	5.05	104-116	12.46	Ctsc /// Ctsc
10415319	rs13483877	X	84	5.05	82-96	12.46	Irf9
10582162	rs3683997	1	36	5.05	11-43	12.45	Cotl1
10402268	rs6329892	6	142	5.04	134-147	12.43	Lgmn
10427471	rs3672597	11	112	5.04	105-114	12.44	Osmr
10411668	rs3705275	8	106	5.04	105-112	12.43	Ocln
10485013	rs3667007	2	82	5.04	75-108	12.43	1110051M20Rik /// 1110051M20Rik
10464999	rs3718160	16	77	5.04	69-81	12.43	Cst6
10605256	rs13482193	14	56	5.04	42-72	12.43	Flna
10523128	rs3705458	5	100	5.02	97-111	12.38	Ppbp
10397351	rs13481908	13	82	5.02	69-91	12.4	Jdp2
10483809	rs3683997	1	36	5.02	31-44	12.39	Nfe2l2
10415980	rs13482193	14	56	5.01	42-69	12.38	Fbxo16
10572146	rs3683997	1	36	5.01	33-43	12.37	Atp6v1b2 /// Atp6v1b2
10390691	rs13478002	4	136	5	135-141	12.35	Nr1d1
10572146	rs13478952	6	106	5	104-116	12.36	Atp6v1b2 /// Atp6v1b2
10587818	rs6385971	9	100	5	86-112	12.35	Plscr4
10600836	rs6329892	6	142	4.99	134-149	12.34	Msn
10415980	rs3672808	6	140	4.99	135-144	12.33	Fbxo16
10582378	rs6370458	11	109	4.99	105-112	12.33	Piezo1
10559919	rs13478433	5	104	4.99	98-112	12.34	2810047C21Rik1 /// Zfp772 /// Gm20482 /// Gm3912
10487597	rs3686443	13	87	4.99	81-93	12.33	Il1b
10363455	rs6174757	9	68	4.99	58-79	12.33	Pcbd1
10554789	rs13478753	6	51	4.99	45-57	12.33	Ctsc /// Ctsc
10463517	gnf01.036.770	1	40	4.99	34-44	12.32	Pprc1
10452030	rs3658783	6	144	4.97	136-149	12.28	Plin3
10551365	rs13478971	6	111	4.97	110-116	12.28	Prx
10463517	rs6386362	11	107	4.97	103-114	12.29	Pprc1

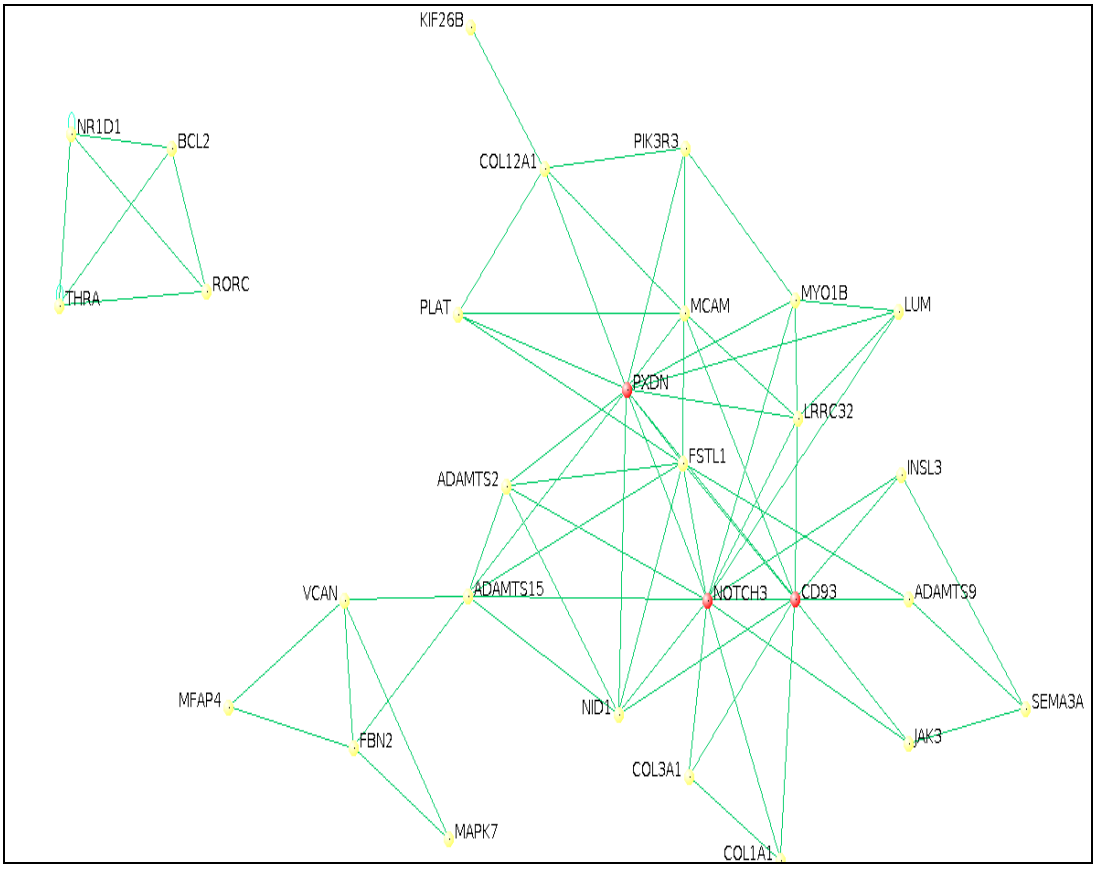
10525397	rs13475931	1	76	4.97	56-89	12.29	Arpc3 /// Arpc3
10591494	gnf09.012.310	9	18	4.97	11-34	12.29	S1pr5
10580649	rs3023025	4	143	4.96	135-155	12.26	Ces1e
10346191	rs6355719	6	56	4.95	49-97	12.24	Stat1
10463517	mCV23390953	19	46	4.95	43-48	12.24	Pprc1
10380560	rs3659933	5	46	4.95	33-51	12.24	Zfp652
10393754	rs3683997	1	36	4.95	31-43	12.24	Actg1
10453887	rs3688566	4	141	4.94	135-156	12.23	Cables1
10523128	rs3714012	9	61	4.94	50-79	12.23	Ppbp
10586591	rs3672808	6	140	4.93	136-149	12.2	Car12
10548375	mCV24625340	13	85	4.93	76-93	12.2	Clec7a
10491300	rs4222295	1	39	4.93	33-44	12.2	Skil
10416837	mCV24625340	13	85	4.92	76-93	12.19	Irg1
10358660	rs6329892	6	142	4.91	134-149	12.16	Hmcn1
10604076	CEL.2_111981058	2	112	4.91	100-119	12.16	Snora69
10400357	rs13481161	11	92	4.91	88-99	12.15	Baz1a /// LOC100048557
10464642	rs3658927	2	121	4.9	116-128	12.14	Carns1
10547664	mCV24625340	13	85	4.9	74-93	12.13	Clec4e
10537146	rs3658783	6	144	4.88	134-149	12.09	Akr1b8
10374455	rs3688566	4	141	4.88	135-155	12.1	Spred2
10392207	rs13482744	15	101	4.88	95-103	12.09	Tex2
10573583	rs3683997	1	36	4.88	33-43	12.1	Man2b1
10461979	rs6372656	19	22	4.88	4-26	12.09	Aldh1a1
10409414	rs13478451	5	109	4.87	98-124	12.08	Rab24
10452815	rs13475919	1	73	4.87	51-79	12.07	Xdh
10463599	rs3658783	6	144	4.86	138-149	12.04	Nfkb2
10369715	rs3023025	4	143	4.86	141-154	12.04	Mypn
10358664	rs13478971	6	111	4.86	110-116	12.06	Hmcn1
10452815	gnf01.036.770	1	40	4.86	34-50	12.05	Xdh
10507273	rs13478971	6	111	4.85	110-116	12.04	Pik3r3
10557326	rs3683997	1	36	4.85	33-43	12.03	Il4ra
10503134	rs3672808	6	140	4.84	135-147	12.01	Sdcbp
10349876	rs13476012	1	102	4.84	99-107	12	Plekha6
10570483	rs13478002	4	136	4.83	135-142	11.99	Arhgef10
10468974	rs3658927	2	121	4.83	85-123	11.99	Cdnf
10505187	rs13481161	11	92	4.83	90-100	11.99	Ugcg
10346015	rs3723062	1	88	4.83	76-105	11.99	Col3a1
10346191	rs13483135	17	84	4.83	79-95	11.98	Stat1
10607467	rs13478002	4	136	4.82	135-142	11.97	Sat1

10375065	rs13478971	6	111	4.82	110-117	11.96	Sh3pxd2b
10443120	rs3089366	10	89	4.82	85-98	11.95	Ggnbp1
10487605	rs6248752	3	12	4.82	8-16	11.97	F830045P16Rik
10604076	rs13478068	4	155	4.81	151-156	11.95	Snora69
10439299	rs6265387	6	147	4.81	138-149	11.94	Stfa3
10557326	rs3672808	6	140	4.81	134-149	11.95	Il4ra
10492997	rs13478002	4	136	4.81	135-142	11.95	Etv3
10591988	rs13484024	X	135	4.81	106-139	11.95	Adamts15
10572747	rs13478952	6	106	4.81	104-116	11.94	Tpm4
10470388	rs6174757	9	68	4.81	56-79	11.95	Cacfd1
10496605	rs4183448	16	47	4.81	42-57	11.94	Ccbl2
10511416	rs13478952	6	106	4.8	98-114	11.93	Tox
10426557	rs3023025	4	143	4.79	141-154	11.9	Pfkm
10553967	rs13477873	4	101	4.79	93-111	11.89	Pcsk6
10546430	rs6341620	5	37	4.79	29-51	11.9	Adamts9
10428983	rs3672808	6	140	4.78	134-147	11.89	Fam49b
10463517	rs13475919	1	73	4.78	62-79	11.87	Pprc1
10601569	rs13481173	11	96	4.77	88-103	11.85	Pcdh11x
10403352	rs3683997	1	36	4.77	31-43	11.86	Klf6
10400357	rs13478971	6	111	4.76	104-116	11.83	Baz1a /// LOC100048557
10391301	rs13478971	6	111	4.76	104-116	11.84	Stat3 /// Stat3
10350516	rs3686443	13	87	4.76	81-93	11.83	Ptgs2
10598178	rs13466711	1	174	4.74	164-178	11.8	Disp1
10458731	rs3688566	4	141	4.74	135-147	11.8	Mcc
10409414	rs3672808	6	140	4.74	135-149	11.79	Rab24
10601729	rs13478002	4	136	4.74	135-147	11.79	Drp2
10523156	rs3686443	13	87	4.74	76-93	11.8	Cxcl2
10526514	rs13479922	8	95	4.73	89-116	11.78	Cldn15
10384458	rs3686443	13	87	4.73	76-93	11.78	Plek
10600836	rs13483756	X	50	4.72	34-71	11.75	Msn
10582376	rs6329892	6	142	4.7	136-148	11.71	Piezo1
10521678	UT_8_46.498556	8	48	4.7	32-64	11.71	Cd38
10407792	rs13482194	14	56	4.69	53-70	11.68	Gpr137b-ps /// Gpr137b
10605328	rs3725341	2	51	4.69	41-73	11.69	Fam3a /// Fam3a
10437687	rs3683997	1	36	4.69	31-43	11.69	Litaf
10408975	rs13476399	2	28	4.69	20-51	11.69	Kif13a
10537742	rs3688955	11	90	4.68	87-100	11.67	Clcn1
10403352	rs13482194	14	56	4.68	42-70	11.67	Klf6
10493235	rs3700706	5	33	4.67	29-41	11.64	Paqr6

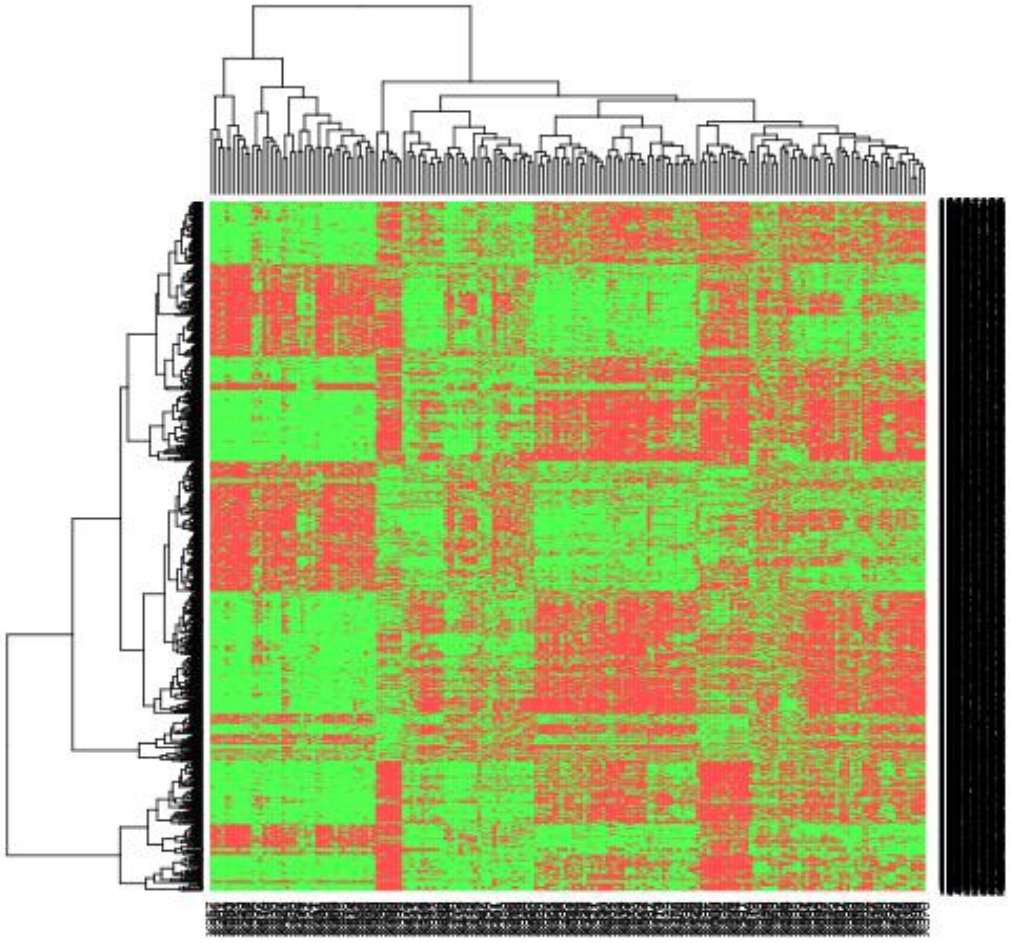
10411395	rs13478068	4	155	4.66	151-156	11.62	Arhgef28
10439249	rs13479063	6	136	4.65	134-149	11.6	Parp14
10346191	CEL.X_77780392	X	83	4.65	62-97	11.6	Stat1
10358605	rs3659551	7	7	4.65	4-17	11.6	Hmcn1
10594988	CEL.10_86322095	10	86	4.62	85-108	11.53	Mapk6
10452815	rs13478753	6	51	4.6	45-59	11.5	Xdh
10582376	rs13478002	4	136	4.59	133-141	11.47	Piezo1
10596117	rs13481127	11	83	4.59	73-88	11.48	Cep63
10411395	petM.05810.1	X	67	4.59	61-93	11.48	Arhgef28
10360270	rs3023025	4	143	4.58	141-154	11.45	Atp1a2
10359948	rs6329892	6	142	4.58	134-148	11.45	Uap1
10427471	rs13482571	15	50	4.58	43-54	11.46	Osmr
10543572	rs13482170	14	49	4.54	42-55	11.36	Impdh1
10543572	rs3714664	9	41	4.53	25-48	11.35	Impdh1
10375537	rs6385855	9	29	4.52	25-37	11.32	Mgat1 /// Mgat1
10358658	rs13479071	6	138	4.51	134-149	11.29	Hmcn1
10554789	rs4200124	16	71	4.51	66-81	11.29	Ctsc /// Ctsc
10549097	rs3023025	4	143	4.41	135-156	11.08	Ldhb
10473058	rs6368370	1	104	4.36	90-109	10.97	Osbpl6



Supplement Figure 1: Cluster dendrogram of miRNAs. The tree-structure shows group of genes hierarchically clustered together. Colors indicate to the module to which genes is assigned to.



Supplement Figure 2: Network of Red module. Green lines display predicted interaction using domain interaction information and predicted interactions. Red circles define candidate hub genes for the red module.



Supplement Figure 3: Heatmap of differentially expressed genes in EBA. The figure shows the genes cluster on the x-axis and samples cluster on the y axis. The genes up-regulated are presented in red while down regulated are presented in green.