TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

Bachelor's Thesis

Gene mapping for type 1 diabetes using a high density genotyping array platform

Inga Pukonen

Biotechnology and Food Technology

2009

TURUN AMMATTIKORKEAKOULU

TIIVISTELMÄ

Bio- ja elintarviketekniikan koulutusohjelma

Tekijä: Inga Pukonen

Työn nimi: 1 tyypin diabeteksen geenikartoitus

Biotekniikka	Ohjaajat: FT Ilari Suominen, FT Robert Hermann
Opinnäytetyön valmistumisajankohta: Kesäkuu 2009	Sivumäärä: 48

Työn tarkoituksena oli tehdä koko ihmisgenomin genotyyppaus käyttäen alustana Affymetrixin SNP 5.0 mikrosirua. Tavoitteena oli toteuttaa koko genomin assosiaatiotutkimus isolle väestökohortille, käyttäen aineistona DIPP-projektissa 15 vuoden aikana kerättyjä diabetes potilasnäytteitä. Sirun avulla on tarkoitus löytää tilastollisesti merkittäviä SNP:ä eli yhden emäksen muutoksesta aiheutuvia geenien rakenne-eroja, jotka voisivat auttaa paremmin ymmärtämään 1 tyypin diabetesta.

Työssä määritettiin koko genomi yhteensä 245:sta potilasnäytteestä, joista 153 oli diabetekseen sairastunutta potilasnäytettä ja 92 kontrollinäytettä. Mikrosiruista saatua dataa analysoitiin käyttäen Affymetrix Genotyping Console (GTC) ohjelmaa, ja lopuksi saadulle SNP-aineistolle tehtiin yhden pisteen assosiaatiotesti käyttäen tilastollisella R ohjelmointikielellä toimivaa Bioconductoria.

Työn päämäärä saavutettiin ja tuloksena tutkimuksessa löydettiin useita kymmeniä tilastollisesti merkittäviä SNP:ä. Saatujen tuloksien avulla tiedämme missä geeneissä löydetyt SNP:t esiintyvät ja näin ollen voimme jatkossa tutkia kyseisten geenien merkitystä ja roolia 1 tyypin diabeteksessa.

Hakusanat: mikrosiru, genotyyppaus, data analysointi

Säilytyspaikka: Turun ammattikorkeakoulun kirjasto

TURKU UNIVERSITY OF APPLIED SCIENCES ABSTRACT

Degree Programme in Biotechnology and Food Technology

Author: Inga Pukonen

Title: Gene mapping for type 1 diabetes using a high density genotyping array platform

Biotechnology	Instructors: Ilari Suominen PhD Robert Hermann PhD
Date: June 2009	Total number of pages: 48

The aim of the study was to carry out the genotyping of the human genome by using Affymetrix Genome-Wide Human Array SNP 5.0 as the array platform. The aim was further to carry out the whole genome association study to a large Finnish population cohort by using DIPP project material which had been collected over a period of 15 years from diabetes patient samples. With the help of the array it will be possible to find statistically significant SNP that can help us for gain deeper understanding of type 1 diabetes genes.

The whole genome was determined from a total of 245 patient samples, of which 153 were diabetes positive samples and the remaining 92 samples were controls. The genotyping calls were determined from the gathered array picture data by using Affymetrix Genotyping Console (GTC) software. From the results of the genotyping calls, single point association tests were performed by using Bioconductor. With the help of the software it was possible to determine statistically the most significant SNPs of all 500,568.

The goal of this study was achieved and as a result, dozens of statistically significant SNPs were found. With the help of these results we can determine the genes where these SNPs are located and explore the role of these genes in type 1 diabetes.

Keywords: high density array, microarray, genotyping, data analysis

Deposit at: Library of Turku University of Applied Sciences

CONTENTS

FIGURES

GRAPHS

TABLES

ABBREVIATIONS

Ab	antibody
cDNA	complementary DNA
CEL	file type that contain the raw probe level data
CNV	copy number variation
DIPP	The Diabetes Prediction and Prevention project
DM	Dynamic Model algorithm
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
GADA	glutamic acid decarboxylase
GTC	Genotyping Console
HLA	human leukocyte antigen
HSDNA	Herring Sperm DNA
HW	Hardy-Weinberg equilibrium
IAA	insulin auto antibodies
IA-2A	tyrosine phosphatase-related IA-2 molecule
ICA	islet cells auto antibodies
IDDM	insulin-dependent diabetes mellitus, type 1 diabetes
MHC	major histocompatibility complex
MM	mismatch
MQW	Milli-Q water
PCR	polymerase chain reaction
PM	perfect mach
SAPE	Streptavidin Phycoerythin
SNP	single nucleotide polymorphisms
T1D	type 1 diabetes
WGA	Whole genome association
WHO	World Health Organization

1 AIM OF THE STUDY

The disease association studies that have been done during the last 20 years were done using only few SNPs because genotyping has been too expensive and slow. This has been a major limitation in understanding disease coding genes. Now that the technology of microarrays has been developed rapidly producing fast and cheap platforms, genotyping of hundreds of thousands of SNPs is feasible. This study aimed to carry out the first whole genome scan in the large population cohort -the DIPP cohort- to help further understanding of type 1 diabetes genes. Such a study has never been condacted before.

2 INTRODUCTION

The Diabetes Prediction and Prevention Project (DIPP) is a study where general population newborns are screened of increased risk for type 1 diabetes in the University Hospitals of Turku, Tampere and Oulu. DIPP was launched in Finland in 1994 and since then over 100 000 newborns have been screened. Infants with risk HLA-DQB1 genotypes (DR3-DQ2/DR4-DQ8, DR4-DQ8/X, DR3-DQ2/Y, where X≠DR3-DQ2, DQB1*0301, DQB1*0602 and Y≠DR7-DQ2, DQB1*0301, *0302, *0602 or *0603), are followed and sampled at intervals of 3–12 months. Islet cell antibodies (ICA) are analyzed from all serum samples and if they are found positive, IAA, GADA and IA-2A are tested in all samples available from that ICA-positive subject (Hermann 2006a). To date over 8 500 children carrying increased genetic risk of type 1 diabetes have participated in the study and over 110 of them have progressed to clinical diabetes (http://research.utu.fi/dipp).

3 THEORETICAL BACKROUND

3.1 Diabetes

Diabetes is a chronic metabolic disorder characterized by elevated levels of blood glucose, or sugar. It occurs when the body produces little or no insulin or when cells not respond appropriately to the insulin that is produced. Hyperglycaemia, or raised blood sugar, is a common phenomenon in uncontrolled diabetes and over time it leads to serious damage to many of the body's systems, especially the nerves and blood vessels (World Health Organization 2008).

According to the International Diabetes Institute there are three main diabetes types: type 1, also known as juvenile onset diabetes (T1D), type 2 also known as adult-onset diabetes and gestational diabetes which is first diagnosed during pregnancy. Type 1 diabetes is characterized by lack of insulin production. Without daily administration of insulin, Type 1 diabetes is rapidly fatal. Type 2 diabetes results from the body’s ineffective use of insulin. Type 2 diabetes comprises 90% of people with diabetes around the world, and is largely the result of excess body weight and physical inactivity. The symptoms may be similar to those of Type 1 diabetes, but are often less marked. As a result, the disease may be diagnosed several years after onset, once complications have already arisen. Until recently, type 2 diabetes was seen only in adults but it is now also occurring in obese children (World Health Organization 2009).

The World Health Organization (WHO) estimates that more than 180 million people worldwide have diabetes. This number is likely to more than double by the year 2030.

3.1.1 Type 1 diabetes

Type 1 diabetes occurs in about 10-15% of all cases of diabetes. It usually occurs in people under the age of 30, but can breakout at any age.

Type 1 diabetes occurs when the body’s immune system destroys the cells of the pancreas that produce insulin (autoimmune response). Specific antibodies may be present in the blood during this time. This process may take several years. It is thought that a virus or chemical may trigger this reaction in people who have a genetic predisposition. Only a small number of people have this genetic risk.

In the disease the pancreas no longer produces insulin and so the glucose cannot enter the muscle and other body cells, resulting in a rapid buildup of glucose and ketones in the blood stream. The kidneys attempt to wash this excess glucose out of the body so there is an increase in the amount of urine produced, and the person becomes very thirsty. If glucose cannot be used by the cells, the body breaks down fat as an alternative energy source. The by-products of fat breakdown are ketones. If too many ketones accumulate in the bloodstream they may cause serious illness, and result in a medical emergency. The onset of type 1 diabetes may be quite sudden and often the person shows rapid and unplanned weight loss over several weeks. In adults it may appear more slowly (Internation Diabetes Institute 2009).

3.1.2 The genetics of type 1 diabetes

In general T1D is considered as a complex genetic trait, i.e., not only do multiple genetic loci contribute to susceptibility, but environmental factors also play a major role in determining risk. A large body of evidence indicates that inherited genetic factors influence both susceptibility and resistance to the disease. Genetic susceptibility in family members is clearly dependent on the degree of genetic identity with the proband (the first affected family member), and in fact the risk of T1D in families has a non-linear correlation with the number of alleles shared with the proband; the highest risk is observed in monozygotic twins (100% sharing), followed by first and second degree relatives (50% and 25% sharing respectively).

In studies evaluating genes for association with disease status in either case-control or family-based studies two chromosomal regions have emerged with consistent and significant evidence of association with T1D across multiple studies. These are the human leukocyte antigen (HLA) region at chromosome 6p21.3 and the insulin gene region at chromosome 11p15 (Table 3.).

Tabel 1. The HLA Classes (Al-Mutairi H. & Mohnsen A. 2007)

Class I genes:	HLA-A, HLA-B, HLA-C	encode class I HLA antigens, located on the surface of all nucleated cells
Class II genes:	HLA-DR, HLA-DQ, HLA-DP	produce class II HLA antigens that are found exclusively on B-lymphocytes, macrophages, epithelial cells of Langerhans, and activated T-lymphocytes
Class III genes:	C2, properdin factor B, C4D and C4B	code for complement components, 21-hydroxylase and products involved in T-cell-mediated inflammation, such as TNF-A and TNF-B, and acute phase protein

In a recently conducted genome-wide linkage analysis, which included five complete genome scans (Davies et al. 1994, Hashimoto et al. 1994, Concannon et al. 1998, Mein et al. 1998, Nerup 2001 and Pociot 2002) and a combined analysis of the UK and US genome scan data (Cox et al. 2001), identification of the genetic determinants for T1D was attempted. These studies suggest that close to 20 loci, with variable degrees of linkage evidence, might be influencing T1D risk. The largest contribution from a single locus (IDDM1) comes from several genes located in the MHC (major histocompatibility complex) complex on chromosome 6p21.3 accounting for at least 40% of familial aggregation of this disease. The genes of the MHC region are classified into four families, classes I, II, III and IV. The strongest genetic susceptibility to T1D is conferred by HLA class II gene alleles (Table 1). HLA class II molecules, particularly DR and DQ, account for approximately 40% out of the genetic risk for T1D development (Al-Mutairi H. & Mohnsen A. 2007).

Tabel 2. HLA haplotypes and type 1 diabetes (Al-Mutairi H. & Mohnsen A. 2007)

DR 3	DQA1*0501	DQB1*0201	DRB1*0301
DR 4	DQA1*0301	DQB1*0302	DRB1*0401
DR 4	DQA1*0301	DQB1*0302	DRB1*0405
Predisposing haplotypes			
DR 2	DQA1*0102	DQB1*0502	DRB1*1601
DR 4	DQA1*0301	DQB1*0302	DRB1*0402
DR 4	DQA1*0301	DQB1*0302	DRB1*0404
Protective haplotypes			
DR 2	DQA1* 0102	DQB1*0602	DRB1*1501
DR 6	DQA1*0101	DQB1*0503	DRB1*1401
DR 7	DQA1*0201	DQB1*0303	DRB1*0701

Approximately 30% of T1D patients are heterozygous for HLA-DQA1*0501-DQB1*0201/DQA1*0301-DQB1*0302 alleles which have formerly been referred to as HLA-DR3/4 and are for simplification usually shortened to HLA-DQ2/DQ8 (Table 2). A particular HLA-DQ6 molecule (HLA-DQA1*0102-DQB1*0602) is associated with dominant protection from the disease (Pociot et al. 2002).

In the recent study (Hermann et al. 2006b) evidence was found for the assumption that the PTPN22 C1858T variant regulates type 1 diabetes-specific autoimmunity and strongly affects the progression from preclinical to clinical diabetes in ICA+ individuals. It was shown that PTPN22, INS and HLA-DRB1 had an additive effect on the emergence of IAA. The strong effect of PTPN22 on disease susceptibility (p = 2.1 x 10(-8)) was more pronounced in males (p = 0.021) and in subjects with non-DR4-DQ8/low-risk HLA genotypes (p = 0.0004).

Tabel 3. Susceptibility loci of type 1 diabetes (Al-Mutairi H. & Mohnsen A. 2007)

Loci	Chromosome	Candidate Genes
IDDM1	6p21.3	HLADR/DQ
IDDM2	11p15.5	INSULIN (INS) VNTR
PTPN22	1p13	PTPN22 (LYP)
SUMO4	6q25 (IDDM5)	SUM04
IDDM3	15q26	-
IDDM4	11q13	LRP5, FADD
IDDM5	6q25	MnSOD, SUMO4
IDDM6	18q12-q21	JK(Kidd), ZNF236, BLC2
IDDM7	2q31-33	NEUROD
IDDM8	6q25-27	-
IDDM9	3q21-25	-
IDDM10	10p11-q11	GAD2
IDDM11	14q24-q31	ENSA, SEL-1L
IDDM12	2q33	CTLA-4, CD28
IDDM13	2q34	
IDDM15	6q21	
IDDM16	14q32	
IDDM17	10q25	
IDDM18	5q31.1-33.1	ILI2B

3.2 Single Nucleotide Polymorphisms

The variations in our DNA sequence that cause or contribute to disease are called either mutations or polymorphisms, based on their frequency in the population. DNA sequence variants that occur in > 1% of the population are called polymorphisms, and those that occur in less than one percent of individuals are called mutations. Mutations are responsible for the relatively rare single-gene Mendelian disorders, while polymorphisms are associated with the more common complex genetic disorders (Tebbutt S. 2007). Single nucleotide polymorphisms, SNPs, are common, small variations that occur in human DNA throughout the genome. These polymorphic markers can be used to map and identify important genes associated with diseases. There are around 10 million common SNPs that constitute 90% of the variation in the current human population (The International HapMap Consortium, 2003). There are SNPs approximately every 200 or 300 base-pairs in the human genome. It has been estimated that the human genome contains more than 5 million common SNPs with minor allele frequencies (MAF) ≥ 10% [1–3], and 7.5 million common SNPs with MAF ≥ 5% (Hao K. et al 2008). Recently high-density SNP arrays have allowed researchers to conduct whole-genome association studies (WGAS). Using these arrays we measure the probability of association between a linked causative SNP marker and a trait.

3.3 DNA Microarray

DNA microarrays can be classified according to the type of probes on the array (cDNA, oligonucleotides, and genomic fragments), their generation and immobilization. In many cases presynthesized molecules, such as PCR products, oligonucleotides or isolated DNA are deposited on the array either by contact printing (using metal pins that carry small volumes of probe solution due to capillary action) or by non-contact printing, when probe solution is dispensed by ink-jet printing. Several companies generate high-density microarrays by synthesizing oligonucleotides *in situ*. The synthesis is either based on specific base deprotection by light which is coordinated by photomasks or digital micromirror devices, or on chemical deprotection and the use of ink-jet technology. SNP genotyping microarrays can be manufactured in different ways, including *in situ* synthesis or immobilization of the locus-/allele-specific oligonucleotides on the array and by using tag arrays, which act as hybridization partners for the allele-/locus-specific oligonucleotides, tailed with sequence complementary to the tag (DNA Microarrays, Nuber 2005).

Affymetrix chips use short oligonucleotide probe quarters to interrogate each dimorphic site and include up to 500 000 SNPs. Each quarter consist of a perfect match (PM) and a mismatch (MM) 25-mer, corresponding to both alleles (arbitrarily named allele A and allele B) of a known SNP, yielding four different probes – PMA, PMB, MMA and MMB. This forms the basic unit of quantifying allele-specific hybridization. Each SNP has multiple quartets querying different strands and shifts surrounding the polymorphic site (Xiao Y. et al. 2007).

3.3.1 High-density SNP genotyping array

High-density single nucleotide polymorphism (SNP) genotyping arrays have been used for copy number variation (CNV) detection and analysis, because the arrays can serve a dual role for SNP- and CNV-based association studies. They also can provide considerably higher precision and resolution than traditional techniques. CNV refers to genomic segments of at least one kb in size, for which copy number differences have been observed in comparison to reference genome assemblies. Multiple large-scale studies have reported prevalent CNVs in humans, suggesting that they may account for a significant portion of phenotypic variation. The precise and comprehensive identification of CNVs would greatly benefit the functional analysis of human genome variation, and complement current genome-wide association studies that use SNPs. (Wang K. 2008)

3.3.2 Selection of the array type

After making decision of starting to use the DNA microarray technology, one of the first considerations to make is to select an appropriate array type. The next step is to consider where the array will be acquired from. The most usual choice is to buy a ready-made array from some commercial chip producer. The most known commercial ready-made array producers are Affymetrix, Illumina, Agilent Technologies, Applied Biosystems and Roche (Chu W. 2005). It is also possible to buy a custom-made array where the buyer can decide the genes that will be spotted on to chip. (Kulta A. 2002).

Figure 1 Affymetrix Genome-Wide Human SNP Array 5.0 (Nature Methods, June 2008)

The selected Affymetrix Genome-Wide Human SNP Array 5.0 (Figure 1) is a high density array that evenly covers coding and non coding regions. This way array provides high information content in most of the genome. The SNPs were selected using known linkage disequilibrium patterns. The pattern of linkage disequilibrium (LD) is critical for association studies, in which disease-causing variants are identified by allelic association with adjacent markers (Mueller JC. 2005). In the present study this array type was selected, since it has the required SNP density to capture most of the genetic variation across genome.

3.3.3 Principle of the Affymetrix Genome-Wide Human SNP Array 5.0

The Affymetrix Genome-Wide Human SNP Array 5.0 features 500,568 single nucleotide polymorphisms (SNPs) from the original two-chip Mapping 500K Array Set, as well as additional 420,000 non-polymorphic probes that can measure other genetic differences, such as copy number variations (CNV). SNPs on the array are present on 200 to 1,100 base pair (bp) Nsp I or Sty I digested fragments in the human genome, and are amplified using the fifth generation of the Whole-genome Sampling Assay (WGSA). Using the current 3.0.1 version of the Affymetrix Genotyping Console Software, a set of 440,794 SNPs on the array exhibit the performance capabilities.

As it can be seen from the Figure 2, total genomic DNA (500 ng) is digested with Nsp I and Sty I restriction enzymes and ligated to adaptors that recognize the cohesive 4 bp overhangs. All fragments resulting from restriction enzyme digestion, regardless of size, are substrates for adaptor ligation. A genetic primer that recognizes the adaptor sequence is used to amplify adaptor-ligated DNA fragments. PCR conditions have been optimized to preferentially amplify fragments in the 200 to 1,100 bp size range. PCR amplification products for each restriction enzyme digest are combined and purified using polystyrene beads. The amplified DNA is then fragmented, labeled and hybridized to a Genome-Wide Human SNP Array 5.0. (Genome-Wide Human Array SNP 5.0 Data Sheet)

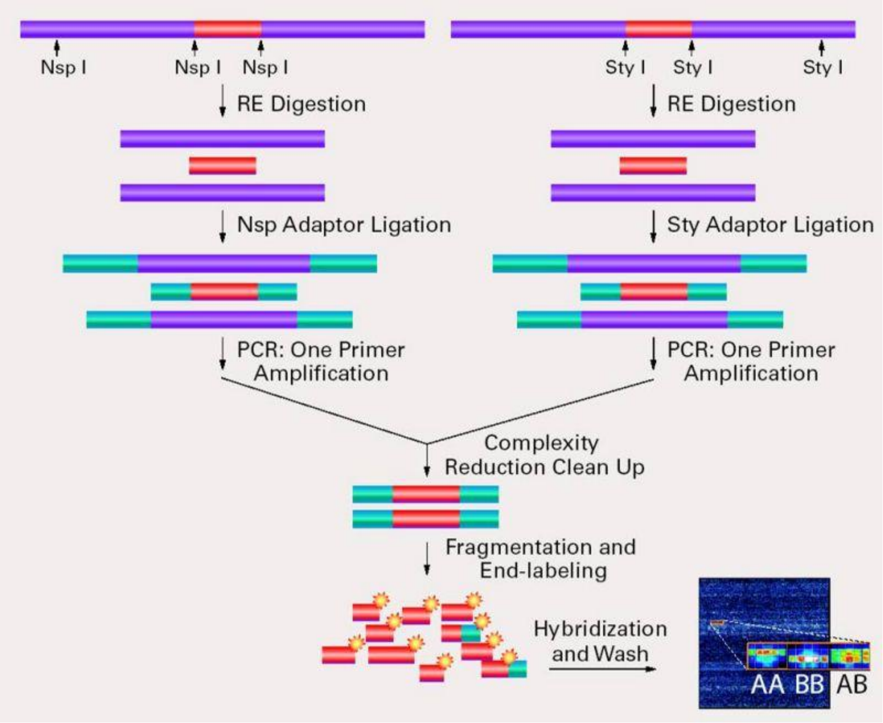


Figure 2. The fifth generation Whole-genome Sampling assay (Data Sheet Affymetrix Genome-Wide Human SNP Array 5.0)

3.4 Hybridization

Hybridization is a reaction where complementary DNA strands bind together. The sample DNA, or more usually PCR product derived from it, is labeled with a fluorophore so that the specific sites of hybridization on the microarray can be visualized and the intensity of the hybridization signal quantified. Success of hybridization requires suitable conditions and enough time. Conditions can be adjusted with hybridization buffers (e.g. pH) and hybridization temperature.

3.5 Principles of fluorescent microarray scanning

Currently the technologies used to image microarrays are detection systems for absorbance, fluorescence or luminescense quantification. Fluorescently labeled microarrays can be easily “read” with commercially available scanners. Laser-based systems use motion control elements to scan the laser beam across the sample and photomultiplier tudes (PMTs) to collect emitted light one pixel at a time. White-light-based systems use mercury or xenon arc lamps to excite an entire field of view and a chargedcoulped device (CCD) to capture emitted light from the entire field of view.

In fluorescence imaging system excitation light is provided by either a halogen arc lamp or a laser. The light is delivered to the sample through a series of lenses and filters. The fluorophore on the sample emits light, which then travels through additional lenses and filters to the detector (a CCD or PMT). The analog signal from the detector is converted into a digital signal, which is then used to display an image of the sample on the computer screen.

3.5.1 Background substraction

The fluorescence intensity that is measured in a feature usually includes a certain amount of a stray light from various sources which can be auto-fluorescence of the slide or a non-specific binding of a labeled sample to the microarray substrate. This stray fluorescence, known as background, needs to be accounted for in order to calculate a true measure of the fluorescence in a feature. There are many methods for removing background from microarray images but each of them has advantages and disadvantages.

Local background substraction methods, which are most commonly used, calculate a unique background value for each feature from a region near the feature. The advantage of local methods is that they substract only the nearby background from each feature. However, if there are artifacts or binding variations near or within a feature, local methods may produce extremely high or low background values.

Global background substraction methods calculate a single value for each wavelength. The advantage of these methods is that they provide single background estimation for the whole slide but as a disadvantage, if the background varies significantly across a slide, one single estimate may not accurately represent the background contribution to all features.

Negative control background subtraction methods calculate a background value from the intensity of specified negative-control features. These methods have several advantages over local and global methods. For example, non-specific binding may differ where features have been printed on a slide, compared to the space between features. In such cases one can estimate non-specific fluorescent background from negative-control features rather than from local regions between features. However, unlike all the other background subtraction methods, the disadvantage to using these methods is that they must be included in the microarray slide design from the beginning. They should also be distributed widely on the array and can therefore take up a lot of space.

There are also other methods, such as morphological methods, in which a copy of each single-wavelength image is created and then each image is filtered to construct a background image for each wavelength. The two standard morphological methods are Opening (where a local minimum filter is applied to the whole image) and Closing followed by Opening (where small dark regions are filled in on the background image, and then a local minimum filter is applied to the whole image).

The problem with background subtraction is that it can add noise and dye bias to a microarray. Another problem with some background subtraction methods is that the result can depend upon the segmentation method used to separate foreground signal from background. If the segmentation method is wrong for the image, foreground intensity can be counted as background. For all these reasons there are these days a model-based method that estimates the true intensity of a spot by modeling the contribution of the background, or background adjustments are not used at all (DNA Microarrays, Nuber 2005).

3.6 Data analysis

Data analyzing is the stage of the microarray technology where results are to be analyzed and the meaning of the results is worked out. A DNA microarray can contain hundreds of thousands of spots with a huge amount of data from just one sample. All of this data needs to be somehow handled and for this purpose special computer programs have been developed. There are quite many different kind of programs available and even programs that do all that is needed to analyze a DNA chip. Most of the smaller programs are available for academic use only and usually larger programs are commercial (Kulta A. 2002).

3.6.1 Normalization

Normalization is the process of removing errors (bias) from a measurement. The data from each spot on the array is usually normalized, which means that intensity ratio value is set to some specific range and so different microarray results become comparable to each other. Data on a microarray could be biased for several reasons such as differences in dye properties, probe labeling or for example in hybridization efficiencies. Normalization can be made using either data from user-selected probe sets or all probe sets.

3.6.2 Clustering

Clustering is a commonly used term in microarray technology. Clustering means grouping of the similar genes into groups with a mathematical algorithm. Clustering is the next thing to do after the spots have been normalized.

There are different types of clustering. Commonly used algorithms are Hierarchical clustering, K-means clustering and SOM (self-organizing map based clustering). Hierarchical clustering algorithm clusters data of N items into a single cluster of size N by pairing two most similar items (or clusters) into a single cluster. This is repeated so many times that there is only one cluster left. K-means clustering algorithm clusters data into a predetermined number (K) of clusters. Self-organized map is an unsupervised neural network algorithm. Self-organized map clusters data into cells and the distance between cells refers to the similarity of the data in cells: the higher the distance is between two cells, the more different is the data in those cells. (Pavan K. 2008)

3.6.3 Quality control

Quality control are performed using various measures and visualization techniques for estimating the quality and sample relationships in the data. If some samples clearly deviate from others, these samples may be considered to be called outliers and excluded from further analysis.

Quality control performed before genotyping is called Single Chip Quality Control. The QC call rate for individual array is used to determine whether a sample should be repeated or used for downstream analysis. In this trial run 3022 SNPs tiled as PM (perfect mach) and MM (mismatch) probes (1511 for Nsp and 1511 for Sty) are tested (Figure 3). These 3022 SNPs are specifically chosen for evaluating data quality and QC call rate is generated by using the Dynamic Model algorithm (DM). Cutoff for passing samples is set to 86%. The arrays that pass the QC call rate will be analyzed using the BRLMM-P algorithm and the majority of the Genome_Wide SNP 5.0 samples that meet this 86% specification will have a BRLMM-P genotyping call rate of at least 96% and accuracy. The QC analysis provides an estimate of the overall quality for a sample. This analysis provides a quick preview of data quality before performing a full clustering analysis. (Affymetrix Genome-Wide Human SNP Array 5.0 Data Sheet)

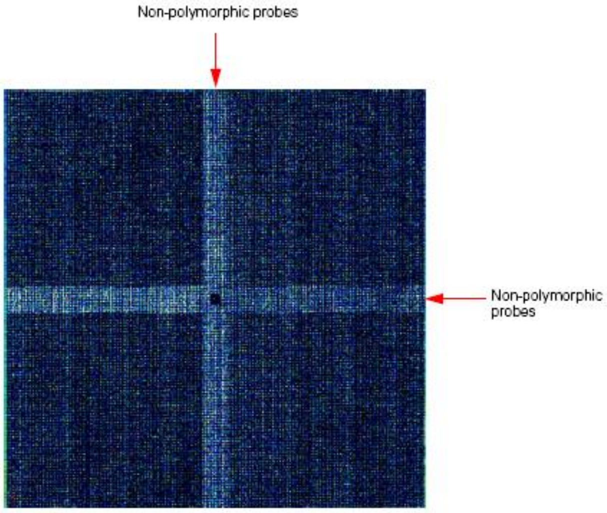


Figure 3. Scanned image of Genome-Wide Human SNP Array 5.0. Array is divided into four quadrants and the genotyping probes are tiled within each quadrant. (Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual)

3.6.4 Genotype calling

Highly accurate and reliable genotype calling is an essential component of high throughput SNP genotyping technology. In this genotyping calling determination was used BRLLM-P algorithm. BRLLM-P performs a multiple chip analysis, enabling the simultaneous estimation of probe effects and allele signals for each SNP. This step is retained even in arrays employing multiple copies of the same probe, even though the probe specific effects are only minimally different between copies. The distribution of summary values across arrays is then used to evaluate the likely genotypes.

BRLMM-P algorithm call genotypes by a template-matching procedure comparing the transformed allele signal values observed in an experiment to the typical values, prototypes that are expected for each genotype. The genotype that is estimated to have the highest probability of having produced the data point is reported as the call. The approximate confidence that is reported for that call is the estimated probability that the data point belongs to one of the other clusters. These allow ranking the genotype assignments by quality, and therefore make the decision not to call in cases of ambiguity.

Every SNP is expected to have three genotypes, “AA”, “AB”, and “BB”. For each genotype for a given SNP is expected to have a prototype with some scatter of values around the prototype. Prototype is a typical observed value for that specific genotype or cluster center. The scatter is approximated by a normal distribution. The relative probability of belonging to a cluster is calculated as a function of the distance from the cluster center and the variation within the cluster. In the end, the standard settings of BRLMM-P calculate a common variance for all clusters (BRLMM-P: Genotype Calling Method for the SNP 5.0 Array, 2007).

3.6.5 Hardy-Weinberg equilibrium

Hardy-Weinberg equilibrium (HW) is a basic principle of population genetics. According to HW the genotype frequencies at particular gene locus will become fixed at certain equilibrium value resulting from random mating under conditions were sexually reproducing organism is diploid and the species has distinct generations. It is however important to understand that outside the lab, one or more of disturbing influences like mutations, non-random mating, selection or limited population size are always in effect. Genetic equilibrium is an ideal state that provides a baseline to measure genetic change against. The equilibrium values are depicted by a mathematical function ($p^2+2pq+q^2=1$) of the allele frequencies (p and q) at a particular locus. (Hardy 2003).

4 METHODS OF LABORATORY WORK

4.1 Case and control selection

From the table below (Table 4) details of the sample group specification can be seen.

Tabel 4. Group description

Group of cases	Inclusion criteria
A: DIPP Ab positive cohort	HLA Genotype DRB1*0401-DQB1*0302 T1D IAA as the first positive biochemical Ab: IAA positive 2x consecutively
B: DIPP Ab positive cohort-GAD antibody group	HLA Genotype DR3-DQA1*05-DQB1*0201 positive (DR3/DR401; DR3/DR404; DR3 boys) T1D and at least three Abs positive two times consecutively (GADA + two others)
C: DIPP Ab positive cohort	Like group A, but no T1D and at least three Abs positive two times consecutively (IAA + two others)
D: DIPP Ab positive cohort-GAD antibody group	Like group B, but no T1D but at least three Abs positive two times consecutively (GADA + two others)
Genetic isolates	T1D Kuusamo, Koillismaa Kainuu 2 or more grandparents are from these regions

Controls were matched by HLA, gender, geographic region and by following up time which should be longer than 10 years. The project covered 154 case samples and 92 control samples.

4.2 Sample collection

DNA was collected from the DIPP-project library. The DNA had originally been isolated from newborns' navel blood samples and preserved in MediCity in -70 ºC. General requirements for the sample DNA were that the DNA had to be double-stranded and free of PCR inhibitors. The DNA must not have been contaminated by other human genomic DNA sources, or by genomic DNA from other organisms. The DNA must not have been degraded. The success of this assay required the amplification of PCR fragments between 200 and 1100 bp in size throughout the genome. To reach this requirement, the genomic DNA needed to be of high quality and free of contaminants because it would otherwise affect on the enzymatic reactions.

4.3 Sample preparation

DNA was isolated by following the "DNA isolation protocol" by CellScreen Laboratory of Immunogenomics. 8ml of ice-cold redblood-lysis buffer (2x blood volume) was added in to 4 ml of blood (in a 15 ml falcon-tube) and centrifuged at 4ºC, 2 500 rpm 10 min. Supernatant was poured out into a Heamasol bucket and 1 ml of MQW was added into the falcon-tube. The tube was centrifuged at 4ºC, 13 000 rpm, 1 min and the wash step was repeated until the color of the pellet had turned homogeneously salmon. 370 µl of proteinase K-mix was added to the sample, shaken and incubated for 15 min at 55 ºC. After that 100 µl of 5M NaCl was added and centrifuged at 4ºC 13 000 rpm for 5 min. Then supernatant was transferred into a new 2 ml eppendorf tube, 1 ml of ice-cold abs. Ethanol was added and the sample was mixed. In this step DNA precipitation was visible and after that DNA was transferred into a new 1.5 ml eppendorf tube containing 1 ml of ice-cold 70% Ethanol. In the end DNA is moved from the ethanol and dried at 37 ºC. 200 µl of TE was added to each sample DNA and incubated at 37 ºC over night.

After DNA isolation samples were prepared by diluting each sample to 50 ng/µl using reduced EDTA TE buffer.

4.4 Sample Quality Control

Sample quality assurance was performed in two ways. Samples were tested by allele specific PCR using HLA-B*62 gene as a control band and genomic DNA was run on a gel to assure that the DNA was intact.

4.4.1 Allele specific PCR

Samples were tested by allele specific PCR (Table 5) using the HLA-B*62 gene as a control band. The purpose of this method was to assure that the DNA was in a good condition and would be duplicated in the actual PCR.

Tabel 5. Allele specific PCR program

	For one sample / (µl)	For 200 samples / (µl)
H_2O	12.52	2504
10 x ABgene buffer	2.10	420
10 mM dNTP-mix	0.42	84
10 µM B62 243 F	0.63	126
10 µM B62 250 R	0.63	126
10 µM C63 F	0.32	64
10 µM C64 R	0.32	64
25 mM $MgCl_2$	3.36	672
5 U/µl ABgene Taq polymerase	0.10	20

PCR MasterMix was pipeted on a 96-well plate 20.4 µl/well. The A1 well contained 1.5 µl B62+ control (20ng/µl), the A2 well 1.5 µl B62- control (20ng/µl) and the A3 well 0.6 µl H_2O. The samples were pipeted in the same order in which they were on the sample-plate. The sample wells contained 0.6 µl (50ng/µl) sample solution + 0.6 µl H_2O. PCR was performed by using the B62NEW66 PCR protocol (Table 6).

PCR protocol: Program name B62NEW66

Duration 1h 30min

Control method: Block

Tabel 6. PCR program of B62NEW66

1	96 ºC	1 min
2	96 ºC	25 s
3	66 ºC	50 s
4	72 ºC	45 s
5	Go to 2	30 times
6	7 ºC	for ever
7	END	

After PCR the samples were run and analyzed on 2% agarose gel.

4.4.2 Agarose gel electrophoresis for Genomic DNA

The prepared 50 ng/µl sample-DNA was run and analyzed on 2% agarose gel. The DNA in the gel was detected by using Invitrogen's SYBR Safe DNA gel stain, which glow brightly when bound to double stranded DNA and exposed to blue light. Samples were prepared by adding 5 µl of sample-DNA (250 ng) and 1 µl of a loading dye (6xBromofenolisine-sucrose). The gel was loaded totally of 249 6 µl samples and GeneRuler 1kb DNA ladder in the both sites of the gel. The gel was run at 160 V for an hour. Before viewing the gel it was divided into four pieces for fitting into a BioRad Imager.

4.5 Restriction enzyme digestion

The genomic DNA was digested by the Sty I restriction enzyme. After preparation of the Sty Digestion Master Mix, 14.75 µl of Master Mix was added to 5 µl of each DNA sample. The 48 samples included 46 genomic DNA samples and one positive and one negative control. Each plate was placed onto a thermal cycler and run by the GW5.0

Digest Program (Table 7). Digestion of genomic DNA by Nsp I restriction enzyme was performed simultaneously with Sty I digestion. It was performed by a different person to avoid contamination. After preparation of the Nsp Digestion Master Mix, 14.75µl of it was added to one set of 48 samples (5µl of each sample). The plate was placed onto a thermal cycler and run by the GW5.0 Digest Program (Table 7).

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise amounts and workflow of this step can be found in the manual.

Tabel 7. Digestion PCR program

GW5.0 Digest Program	
Temperature	Time
37 ºC	120 minutes
65 ºC	20 minutes
4 ºC	Hold

4.6 Ligation

The digested samples were ligated simultaneously using Sty and Nsp Adaptor. After preparing Sty Ligation Master Mix and Nsp Ligation Master Mix, they were both added to their own plates with samples. The samples were placed onto a thermal cycler and run by the GW5.0 Ligate Program (Table 8). After the program the ligated samples were diluted with 75 µl of AccuGENE water.

Tabel 8. Ligate PCR program

GW5.0 Ligate Program	
Temperature	Time
16 ºC	180 minutes
70 ºC	20 minutes
4 ºC	Hold

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise workflow of this step can be found in the manual.

4.7 Sty and Nsp PCR

Equal amounts of each Sty ligated sample were transferred into three fresh 96-well plates (Figure 4). Sty PCR Master Mix was added to each sample and each plate was placed on a thermal cycler and GW 5.0 PCR Program was run (Table 9).

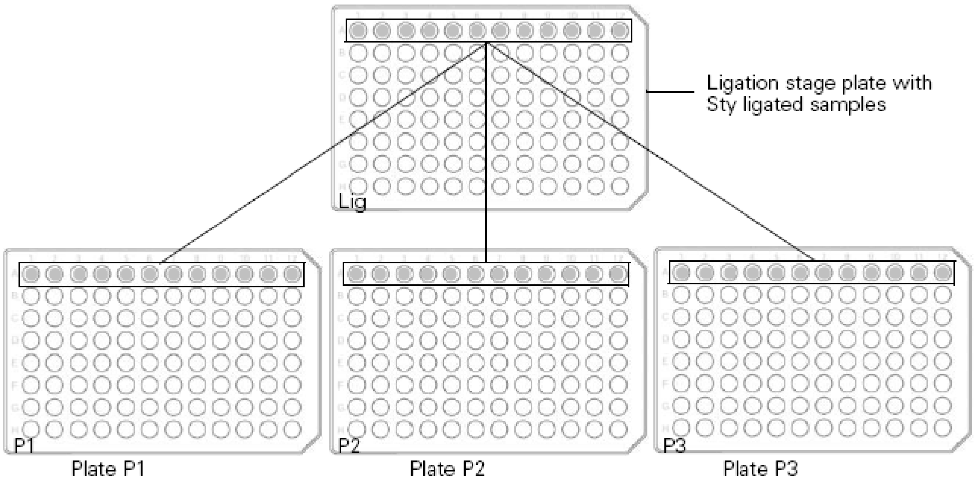


Figure 4. Transferring equal aliquots of diluted, ligated Sty samples to three reaction plates (Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual)

Likewise, equal amounts of each Nsp ligated sample was transferred, but in this case into four fresh 96-well plates. Nsp PCR Master Mix was added to each sample and each plate was placed on a thermal cycler and GW 5.0 PCR Program (Table 9) was run.

Tabel 9. Sty and Nsp PCR program

GW 5.0 PCR Program		
Temperature	Time	Cycles
94 ºC	3 minutes	1 X
94 ºC	30 sec	
60 ºC	30 sec	30 X

68 ºC	15 sec	
68 ºC	7 minutes	1 X
4 ºC	Hold	

After finishing the GW 5.0 PCR Program, PCR products were run on 2% TBE agarose gel at 160 V for 1 hour for verifying that the PCR product distribution is between ~ 250 bp to 1100 bp.

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise workflow of this step can be found in the manual.

4.8 PCR-product pooling and purification

Sty and Nsp PCR reactions were pooled to a single deep well pooling plate (Figure 5), for a total of 700 µl/well. The magnetic beads were added to each pool and incubated. During incubation, the DNA binds to the magnetic beads. Each pool was transferred to a filter plate and dried down on a vacuum manifold. PCR products were washed with EtOH and dried down again. PCR products were eluted using Buffer EB and vacuumed end spin transferred to a new 96-well plate.

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise workflow of this step can be found in the manual.

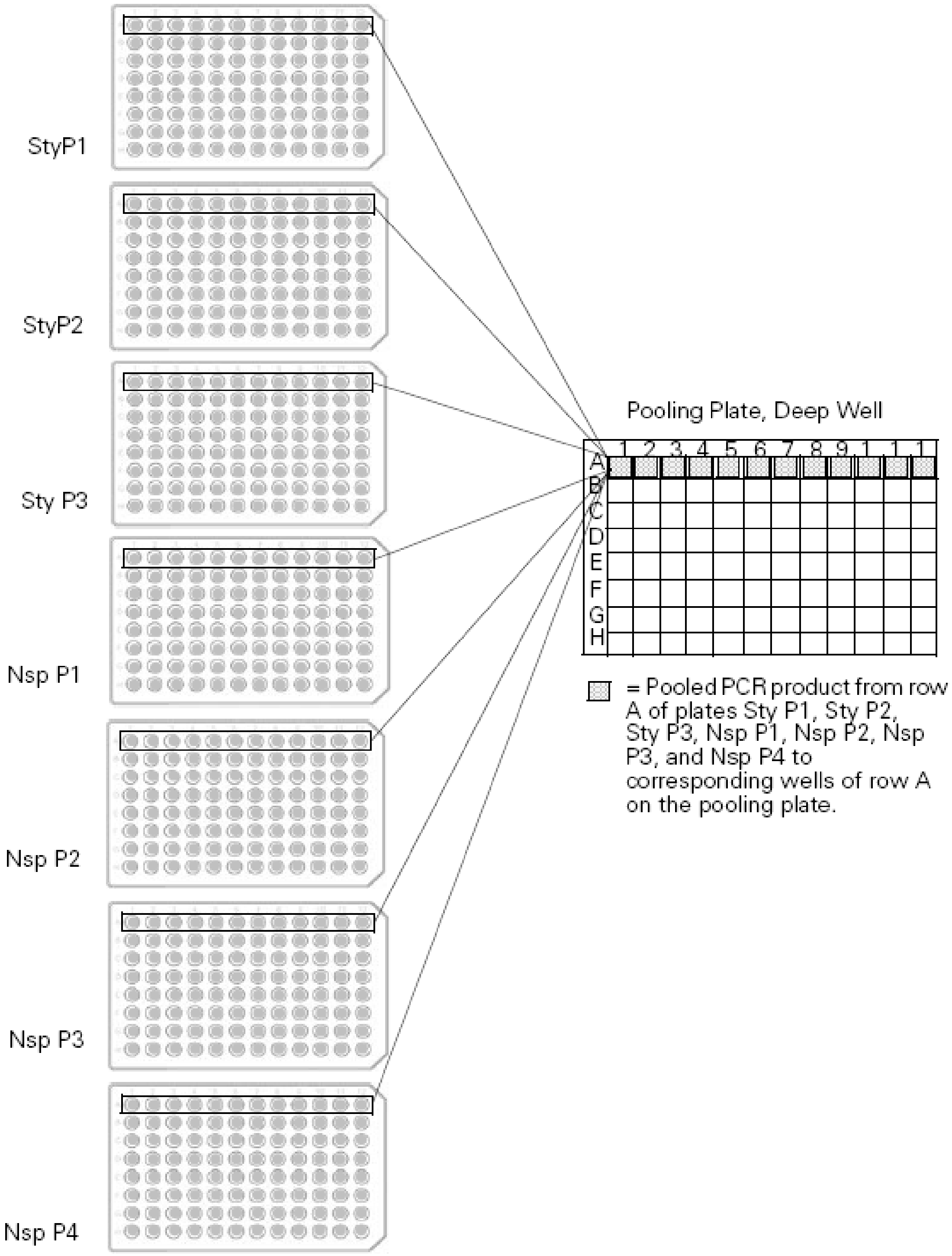

StyP1
StyP2
Sty P3
Nsp P1
Nsp P2
Nsp P3
Nsp P4
Pooling Plate, Deep Well
1 2 3 4 5 6 7 8 9 1 1 1
A
B
C
D
E
F
G
H
▨ = Pooled PCR product from row A of plates Sty P1, Sty P2, Sty P3, Nsp P1, Nsp P2, Nsp P3, and Nsp P4 to corresponding wells of row A on the pooling plate.

Figure 5. Pooling Sty and Nsp PCR product on a deep well pooling plate (Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual)

4.9 Qantitation

The concentration of PCR products after purification with magnetic beads was measured using a spectrophotometer (NanoDrop®). One dilution of each PCR product was prepared in optical plates as a 100-fold dilution. The OD of each PCR product was measured at 260 nm.

4.10 Fragmentation

PCR products were fragmented using Fragmentation Reagent. First Fragmentation Reagent was diluted by adding the Fragmentation Buffer and AccuGENE water. After that the diluted reagent was quickly added to each reaction, placed the plate onto a thermal cycler and the GW5.0 Fragment Program (Table 10) was run. When the program was completed the results of this stage was checked by running 1.5 µl of each reaction on a 4% agarose gel.

Tabel 10. Fragmentation PCR program

GW5.0 Fragment Program	
Temperature	Time
37 ºC	35 minutes
95 ºC	15 minutes
4 ºC	Hold

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise workflow of this step can be found in the manual.

4.11 DNA labeling

The fragmented samples were labeled using the DNA Labeling Reagent. After preparing the Labeling Master Mix the 19.5 µl of Master Mix was added to each sample. The samples were placed onto a thermal cycler and the GW5.0 Label Program (Table 11.) was run.

Tabel 11. Labeling PCR program

GW5.0 Label Program	
Temperature	Time
37 ºC	4 hours
95 ºC	15 minutes
4 ºC	Hold

Laboratory working was performed by following the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual. The precise workflow of this step can be found in the manual.

4.12 Target hybridization

Hybridization is the most critical step in microarray experiment. In this stage, each reaction was loaded onto a Genome-Wide Human SNP Array 5.0.

Before loading arrays they were prepared by unwrapping and marking. Arrays were allowed to warm to room temperature by leaving them on the table for 15 minutes. A 200 µl pipette tip was inserted into the upper right septum of the each array.

The Hybridization Master Mix was prepared according to Table 12.

Tabel 12. Hybridization Master Mix

Reagent	1 Array	48 Arrays (15 % extra)
MES (12X;1.25 M)	12 µl	660 µl
Denhardt’s Solution (50X)	13 µl	715 µl
EDTA (0.5 M)	3 µl	165 µl
HSDNA (10 mg/ml)	3 µl	165 µl
OCR, 0100	2 µl	110 µl

Human Cot-1 DNA® (1 mg/ml)	3 µl	165 µl
Tween-20 (3%)	1 µl	55 µl
DMSO (100%)	13 µl	715 µl
TMACL (5 M)	140 µl	7.7 ml
Total	190 µl	10.45 ml

After preparation of Hybridization Master Mix, 190 µl of Master Mix was added to each sample on the Label Plate. Now the total volume in each well was 263 µl. After vortexing and spinning the plate to assure that the sample was mixed properly, the plate was placed onto the thermal cycler and the GW5.0 Hyb Program (Table 13).

Tabel 13. Hybridization PCR program

GW5.0 Hyb Program	
Temperature	Time
95 ºC	10 minutes
49 ºC	Hold

When the plate reached 49 ºC, 200 µl of denatured sample was removed from each well and immediately injected – one sample per array. The loading of the arrays was processed in 4 array sets so that 4 arrays were loaded at a time and then immediately placed into the hybridization oven. The hybridization oven was preheated to 50 ºC. The goal was not to allow loaded arrays to stand at room temperature for more than 1.5 minutes. Samples were left to hybridize for 16 to 18 hours. The target hybridization was performed in two days so that 32 arrays were hybridized on one day and on the other 14 arrays. Positive and negative controls were not loaded on the arrays.

The precise workflow of this step can be found in the Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual.

4.13 Washing and staining arrays

Washing and staining steps were carried out by using GeneChip® Fluidics Station 450 (Figure 6) and it was operated using GeneChip® Operating Software 1.4. The Affymetrix staining protocol for mapping arrays was a three stage process which consisted of 1) Streptavidin Phycoerythin (SAPE) stain, 2) an antibody amplification step, and 3) a final stain with Streptavidin Phycoerythin (SAPE). All three solutions were placed on the Fluidics Stations into the sample holders 1, 2 and 3 in vials. Before starting the program, the fluidics station was primed to ensure that the lines were filled with the appropriate buffers and the fluidics station was ready to run fluidics station protocols.

Figure 6. Fluidics Station 450

(http://www.affymetrix.com/products_services/instruments/specific/fs450.affx)

When the hybridization was ready, the hybridization cocktail was removed from the array and transferred to the corresponding well of a 96-well plate. The array was then filled completely with 270 µl of Array Holding Buffer and placed into the designated module of the Fluidics Station. The Fluidics Station method protocol is shown in Table 14.

Tabel 14. Fluidics Station 450 Protocol – Antibody Amplification for Mapping targets (Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual)

49 Format (Standard) GenomeWideSNP5v1_450	
Post Hyb Wash #1	6 cycles of 5 mixes/cycle with Wash Buffer A at 25 ºC
Post Hyb Wash #2	24 cycles of 5 mixes/cycle with Wash Buffer B at 45 ºC
Stain	Array is stained for 10 minutes in SAPE solution at 25 ºC
Post Stain Wash	6 cycles of 5 mixes/cycle with Wash Buffer A at 25 ºC
2nd Stain	Array is stained for 10 minutes in Antibody Stain Solution at 25 ºC
3rd Stain	Array is stained for 10 minutes in SAPE solution at 25 ºC
Final Wash	10 cycles of 6 mixes/cycle with Wash Buffer A at 30 ºC. The final holding temperature is 25 ºC
Filling Array	Array is filled with Array Holding Buffer.

Wash Buffer A = non-stringent wash buffer
Wash Buffer B = stringent wash buffer

After followed staining, the array was filled with Array Holding Buffer prior to scanning. Before loading array in scanner, arrays window was checked for large bubbles or air pockets. If bubbles or air pockets were present, the array was placed back in Fluidics Station and refilled with Array Holding Buffer.

4.14 Scanning arrays

Scanning was performed by using the GeneChip® Scanner 3000 7G and it was also controlled by GCOS Software 1.4. The glass surface of the array was cleaned with the non-abrasive towel before scanning and both septa on the array were carefully covered with Tough Spots (Figure 7).

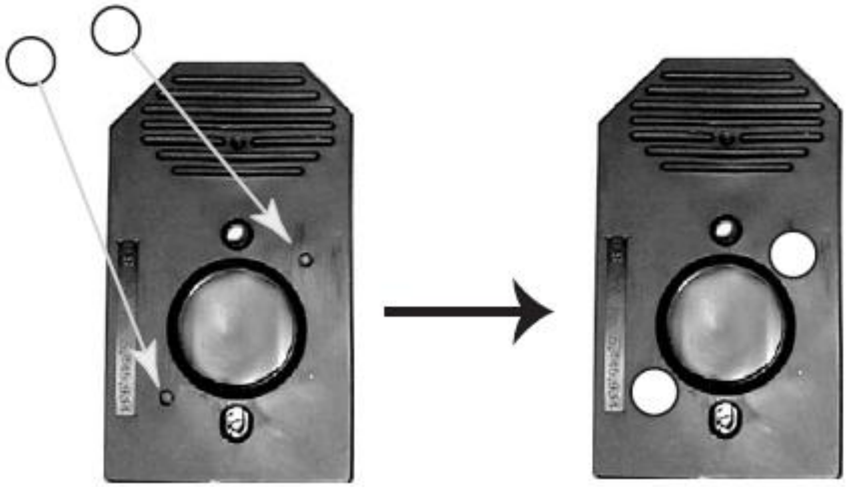


Figure 7. Applying Tough Spots to arrays (Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual)

After placing the arrays in the scanner and selecting the right experiment, the Start button was clicked and scanning was started. The scanning of one array lasted approximately 11 minutes.

5 DATA ANALYSIS

5.1 Scanning the arrays

Microarrays were scanned by using GeneChip® Scanner 3000 7G which is controlled by GCOS Software 1.4. After scanning, the program produced among other file types an image file of each scanned array (.dat file) and after that the array data were ready for analysis.

5.2 Genotyping

The Affymetrix Genotyping Console (GTC) was used to determine genotyping calls for the analyzed samples. Samples were processed in four groups (Table 15). The GTC 3.0.1 version included a BRLMM-P algorithm, which was used to determine genotyping calls for SNP Array 5.0. With the help of the GTC it was possible to display metrics and annotation information in standard tabular form, to evaluate the data quality for each array.

Tabel 15. Sample groups

Samples were processed in four groups:	
Group 1	21 DR3 cases + 23 DR3 controls
Group 2	59 DR4 cases + 62 DR4 controls
Group 3	63 Genetic Isolate cases + all 85 controls (DR3 and DR4)
Group 4	All samples: 153 cases + 92 controls

For each group its own Workspace was created as a working platform. The data files CEL, GQC and ARR were added in to each Workspace from selected arrays. Before running the genotyping, Quality Control Call Rate was determined and the threshold for passing samples was set to be >86%. All arrays that passed this value continued to the genotyping part. Genotyping analysis configurations were set to use the BRLMM-P algorithm, the score threshold was set as 0.05 and Block size as 0. After genotyping, results were reviewed by running per-SNP QC filtering were SNPs that had less per-SNP call rate than 95% threshold were removed.

5.3 Single Point Association Test

Single point association tests were performed by using Bioconductor which is based on the statistical R programming language. With the help of this software it was possible to determine statistically the most significant SNPs of all 500,568. At the end the results were additionally examined with the Hardy-Weinberg p-value.

6 RESULTS

6.1 Results of the laboratory experiments

6.1.1 Sample Quality Control

Results of sample quality control were determined by running 2% agarose gel electrophoresis. There were totally 6 quality control plates.

6.1.1.1 Allele specific PCR

In the gel picture (Figure 8) it can be seen that if there is an upper bright band (~1000 bp), the sample has been amplified. The second band (~150 bp) tells that the sample is positive for B*62 (which is irrelevant in this determination). If there is no band, the sample has been amplified and so it has to be disqualified. All samples that were amplified were forwarded to agarose gel electrophoresis for Genomic DNA.

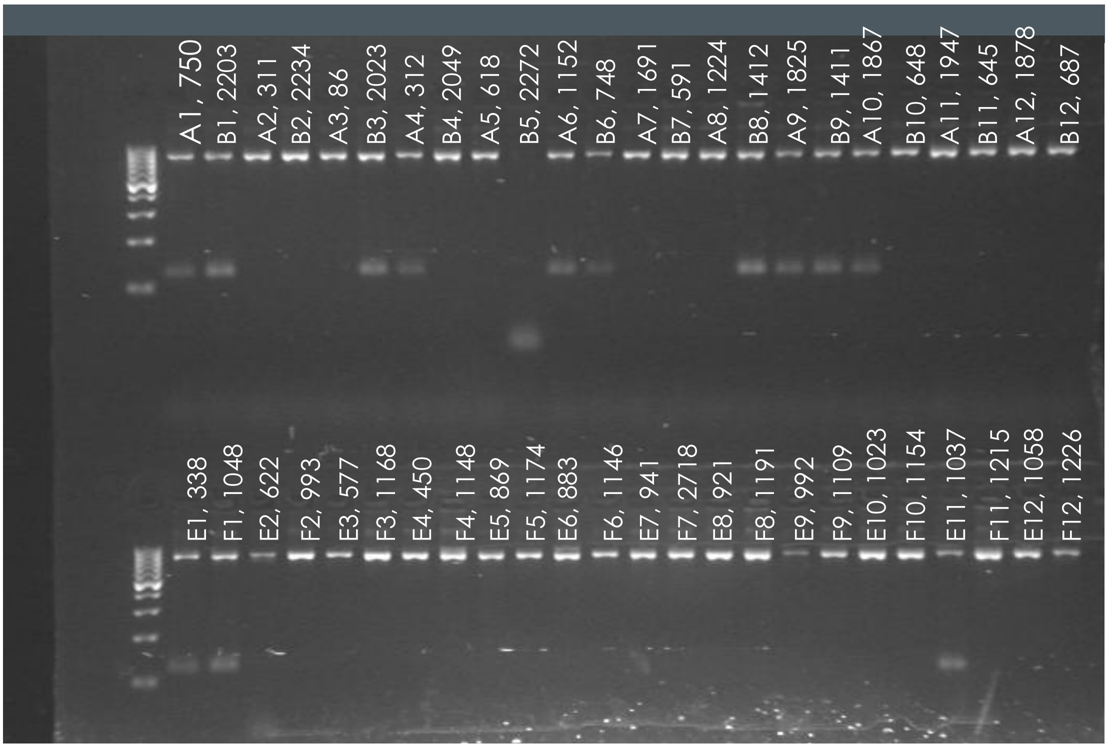


Figure 8. Agarose gel results of allele specific PCR, plates 5 and 6

6.1.1.2 Agarose gel electrophoresis for Genomic DNA

As an assumption, high quality genomic DNA should be run as a major band at approximately 10-20 kb on the gel. In the gel picture it can be seen that if there is the upper bright band (~10 kb), the sample has multiplied and its quality is high.

In the picture (Figure 9), in the second row of samples, sample E5 883 has not amplified and is seen as a smear. This means that the sample DNA is degraded which is why the sample is disqualified. All samples that had major band at 10kb were continued to restriction enzyme digestion.

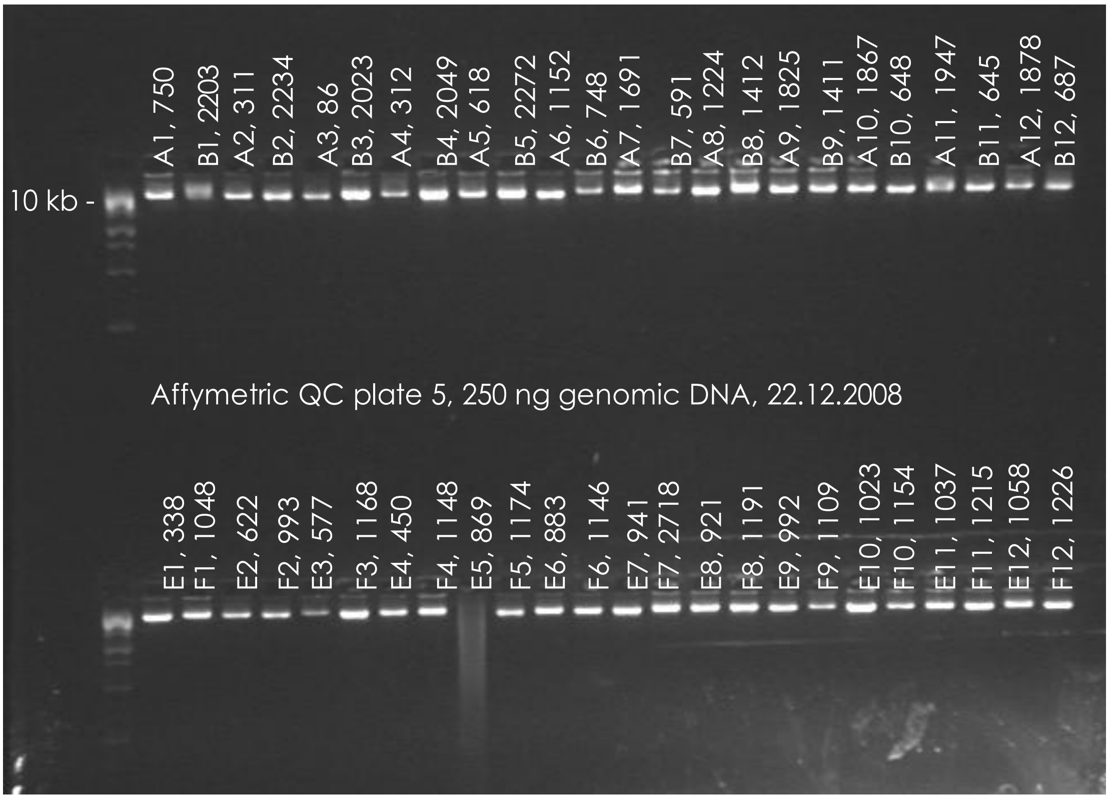


Figure 9. Agarose gel results of genomic DNA, plates 5 and 6

6.1.2 Sty and Nsp PCR

To verify the Sty and Nsp PCR reactions, samples were run on 2% TBE agarose gel (Figure 10). By reading the gel from right to left, the negative control is seen at the fourth well in every row, at the second well the positive control and the last well of every row is GeneRuler 100 bp DNA Ladder. The highest band of the ladder is 1000 bp and the lowest 100. As it can be seen from the picture, the average product distribution is approximately between ~200 to 1100bp as it was required. The samples that had wide smear were multiplied and continued to pooling and purification.

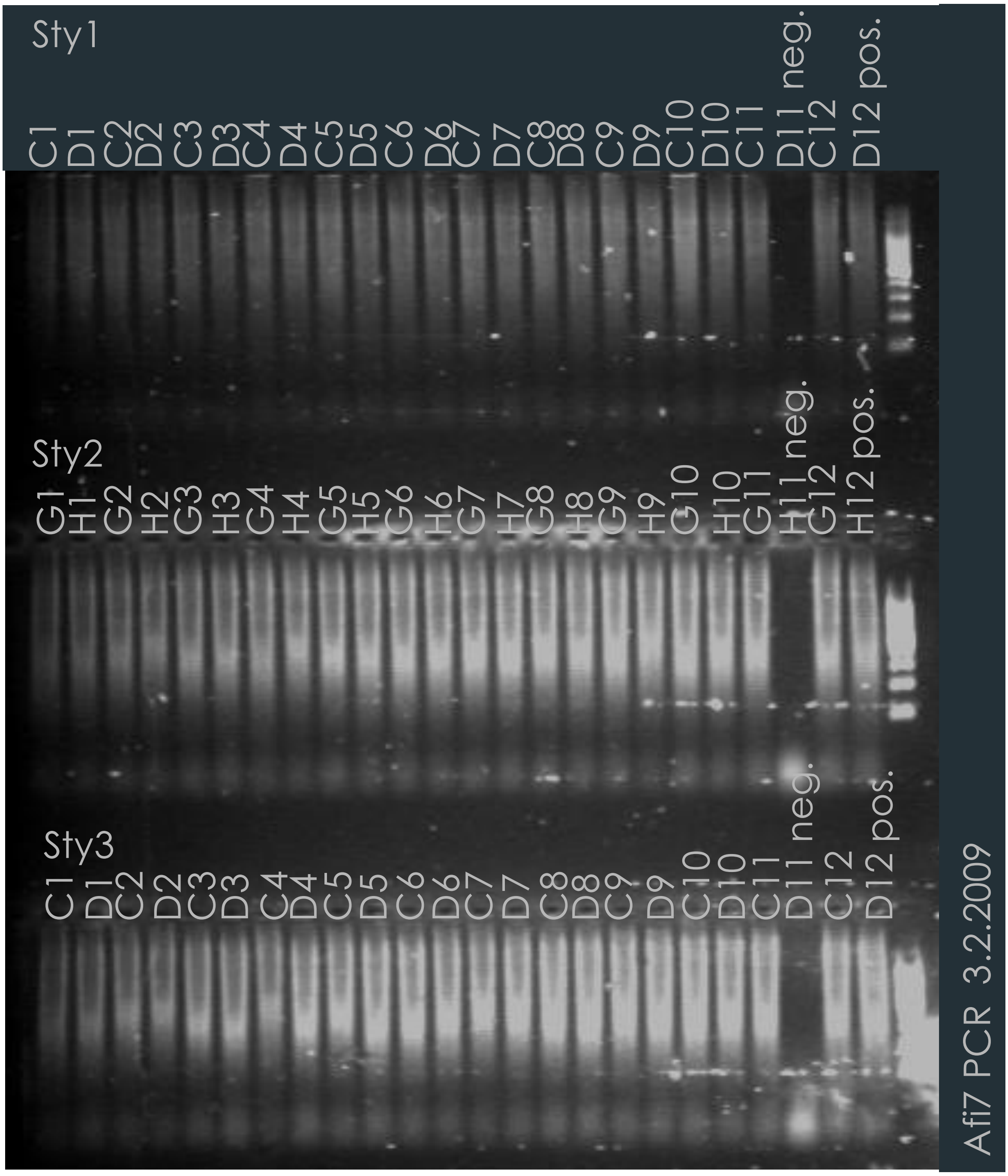


Figure 10. plates 1,2 and 3 PCR products on 2% TBE agarose gel at 160V for 1 hour.

6.1.3 Fragmentation

To ensure that fragmentation was successful, results were checked by running each reaction on ready-made 4 % TBE gel (Figure 11). First and last samples at the gel were GeneRuler 100 bp DNA ladder. By reading the gel from the right to left, the second well was positive control and third well negative. Those samples that multiplied continued to labeling stage.

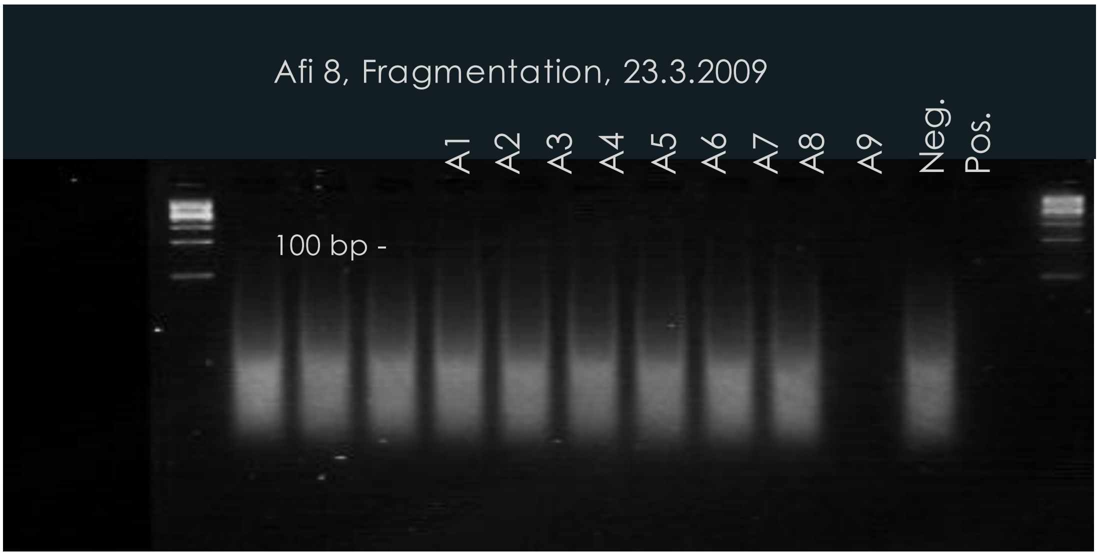


Figure 11. Fragmentation PCR product run on 4 % TBE agarose gel at 120V for 1 hour. Average fragment size is < 200 bp.

6.2 Microarray scanning

GeneChip Scanner 3000 7G automatically scanned microarrays and produced scanned image file (.dat file) for analysis. Totally of all 300 scanned microarrays 246 passed the quality controls and automatic generation of CEL. Files were made. Most of the failed arrays had low SNP call rates which might had been result of not optimal genomic DNA.

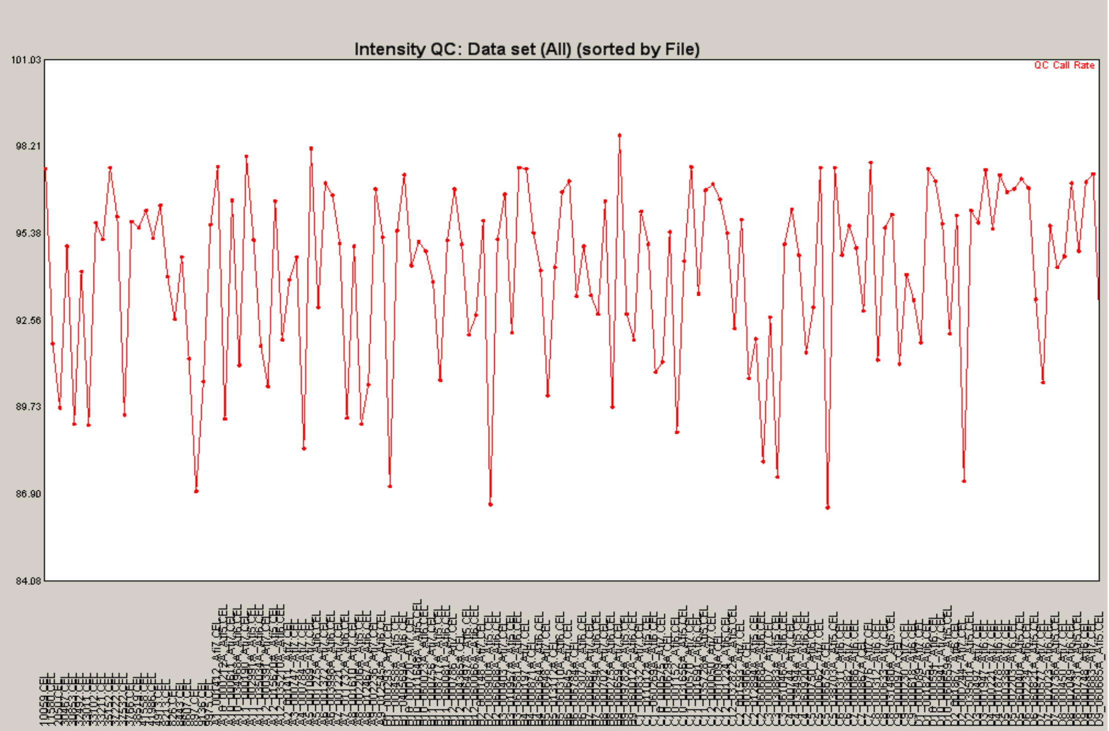
There were also problem with one series of 48 arrays, in which after scanning occurred that no labeling had happened. The problem appeared in the wrong order of Fluidic Station solutions. Because of that arrays were rehybridized, washed, stained and scanned again.

6.3 Data analysis

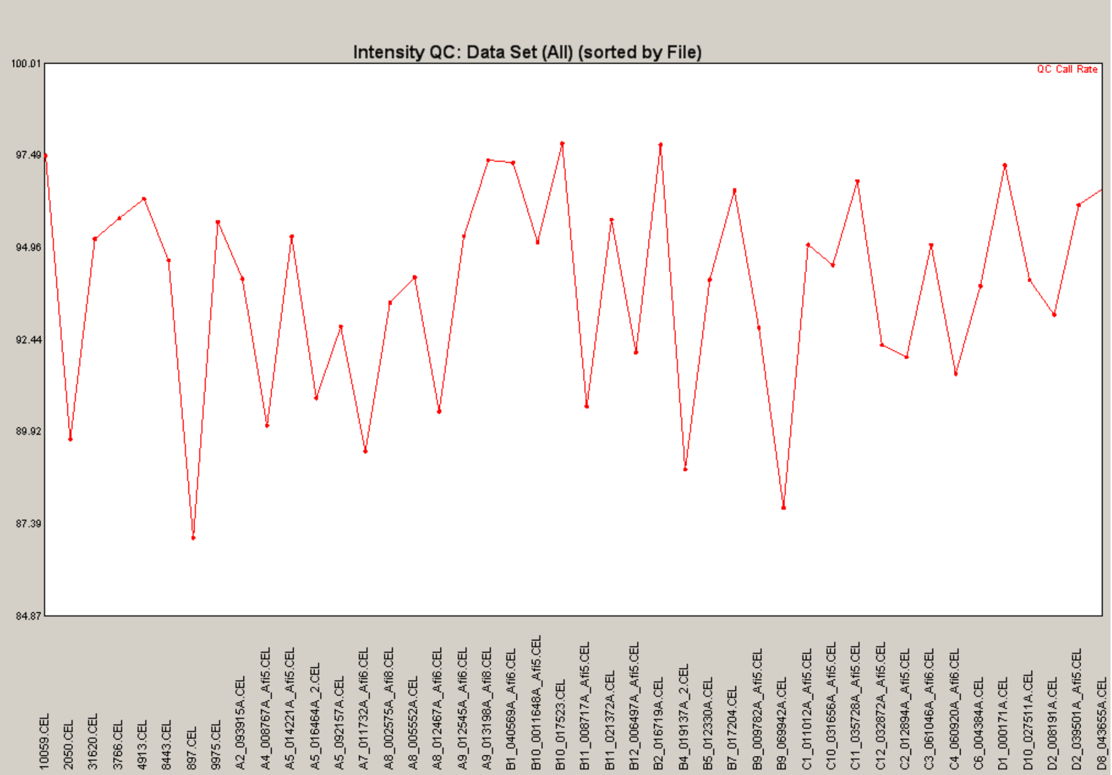
Data produced by GeneChip Scanner 3000 7G and GCOS Software 1.4 was transferred to Affymetrix Genotyping Console 3.0.1 for further processing.

6.3.1 Reviewing CEL data Quality

All of 246 arrays passed the set quality control threshold of > 86%. The arrays were processed in four groups (see table 15) with their own Workspaces. Each Intensity Quality Control Data was demonstrated as a matrix and graph. Intensity Quality Control graphs of groups 1 and 3 can be seen below in Graphs 1 and 2.



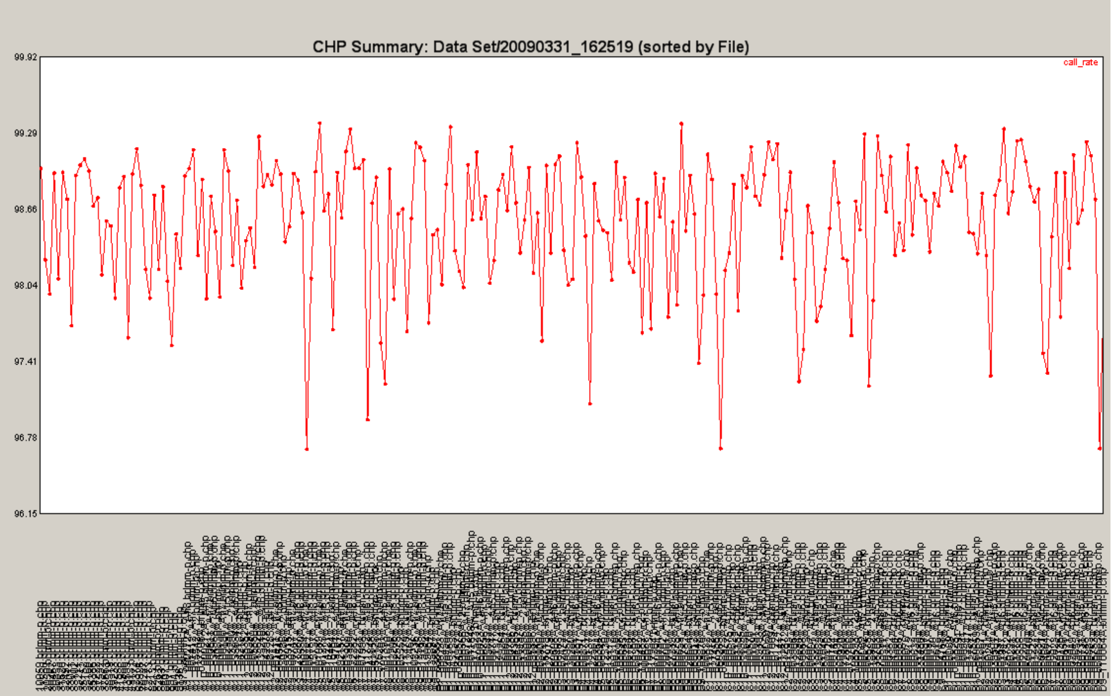
Graph 1. Intensity QC data of Group 3 (Genetic Isolates + all controls)



Graph 2. Intensity QC data of Group 1 (DR3 cases + DR3 controls)

6.3.2 Genotyping

When the genotyping analysis was completed, the CHP summary statistics (genotype calls) was displayed as a table and the line graph (Graph 3) could be drawn from this data. Standard table operations included selecting which columns to display, sort or search and export to clipboard or file.



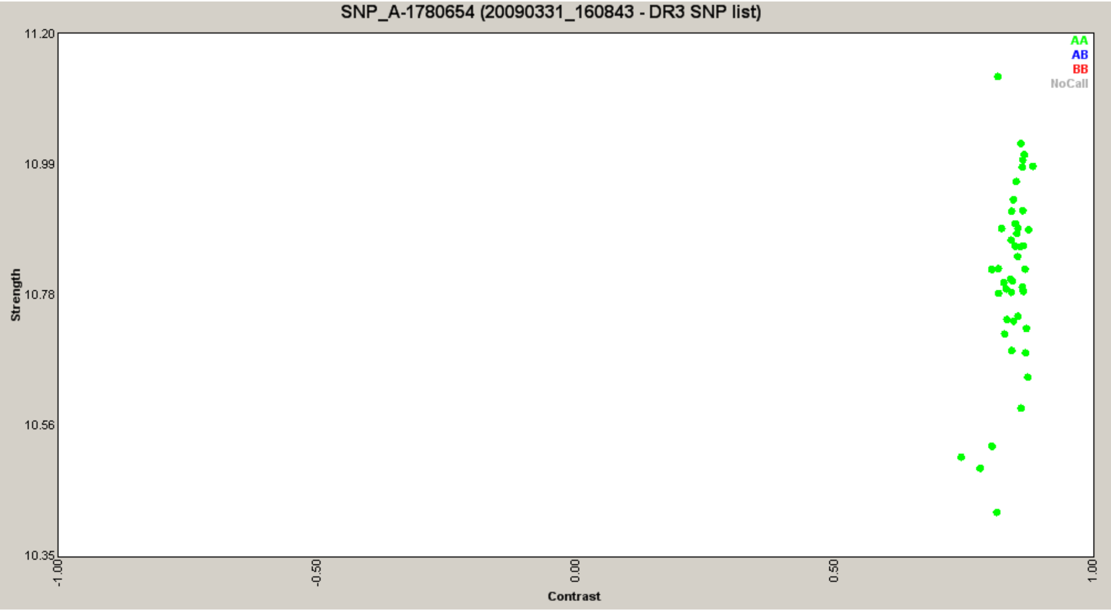
Graph 3. CHP summary data of genotyping results of Group 4 (All samples)

6.3.3 Reviewing CEL data quality

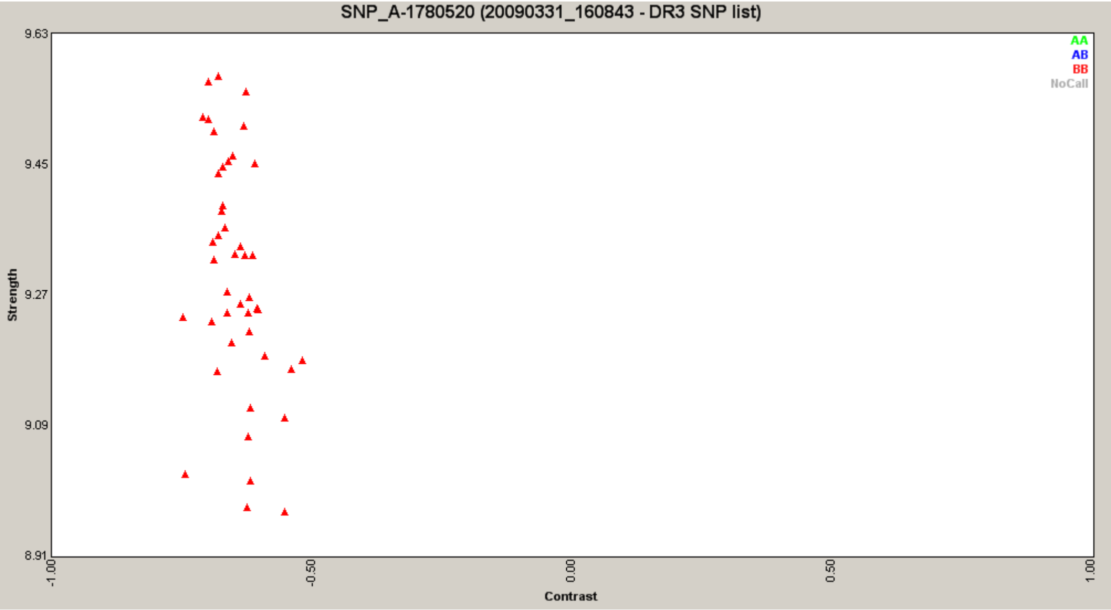
After genotyping of all four groups, results were reviewed by running per-SNP QC filtering were SNPs that had less per-SNP call rate than 95% threshold were removed (Table 16). Per-SNP clusters could visualize for every SNP (Graphs 4-6).

Tabel 16. SNPs passed 95% call rate from total 500,568 SNPs

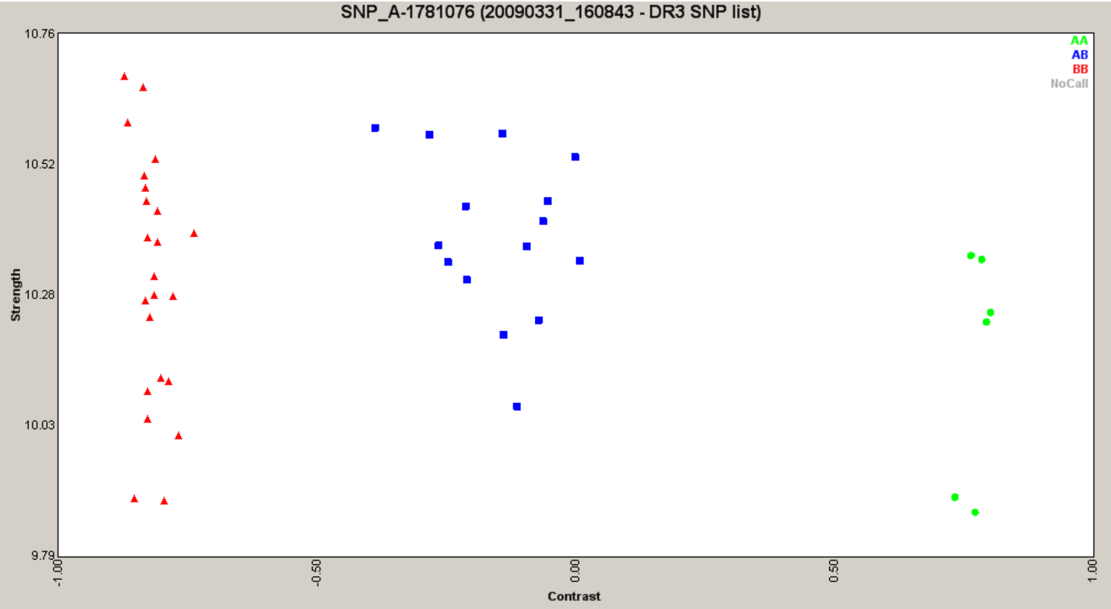
Group 1	417 203
Group 2	414 946
Group 3	417 472
Group 4	415 176



Graph 4. SNP_A-1780654 cluster graph from Group 1 (DR3 cases + DR3 controls). SNP_A1780654 is homogeneous to signal A



Graph 5. SNP_A-1780520 cluster graph from Group 1 (DR3 cases + DR3 controls). SNP_A-1780520 is homogeneous to signal B



Graph 6. SNP_A-1781076 cluster graph from Group 1 (DR3 cases + DR3 controls)

CHP file data and SNP summary data were exported to a text file for association analyses.

6.3.4 Association analyses

6.3.4.1 Single point association tests

Results of Single Point Association tests could be seen in Table 17.

Tabel 17. Results of single point association tests

Group	Significant SNPs (p-value threshold 5.0E-3)	Significant SNPs (p-value threshold 5.0E-4)	Significant SNPs (p-value threshold 5.0E-5)
Group 1	2171	189	19
Group 2	3886	340	25
Group 3	4391	445	39
Group 4	4369	494	66

Tabel 18. Final results of statistically significant SNPs from all arrays

p-value limit	Number of significant SNPs
0.05	40 257
0.01	8 757
0.005	4 369
0.001	944
5.0E-4	494
1.0E-4	107
5.0E-5	66

7 DISCUSSION

7.1 Laboratory work

Working with blood samples that are several years old is always a challenge and sets sometimes limitations in the starting material. In this case finding good samples in the database and isolating the DNA was a somewhat challenging because of this problem. However, the required amount of samples was successfully collected and the DNA fulfilled the requirements. The quality of the DNA was tested by two different methods to ensure that the DNA was free of inhibitors and not highly degraded.

The stages involving enzymatic reactions were the most critical part of the assay. Successful performing of the various molecular biology steps in this protocol required accuracy and attention to details because many of the stages involved specific enzymatic reactions. For example in Sty and Nsp digestion and ligation two different persons performed these steps to avoid any kind of contamination.

In the stage of purifying pooled PCR products by magnetic beds, problem appeared with some samples. During the filtering of the samples, some samples dried more quickly than others though the concentration of samples should have been the same. The problem might have been in the magnetic beads because they expired before all the series were accomplished. Another defect may have been in the vacuum equipment, the filtering plate did not always fit perfectly to the vacuum manifold causing deficiency in vacuum pressure. However, a more likely reason might still have been the deviation between the sample DNA concentrations.

Other problems that were encountered during the protocol were the complication of staining during the washing step. Two SAPE staining solutions and an antibody stain solution got mixed up and were placed in the wrong order in the Fluidics Station. For that reason the staining of the arrays failed and during the scanning of the arrays it looked like no hybridization had taken place. When the problem occurred, the arrays were once again rehybridized, washed and scanned.

In the few unsuccessful array image scannings an error occurred on the image (Figure 12). It seemed like there was a bubble or an air pocket in the array window although it didn’t look like it. The solution was found in the amount of the loaded Array Holding Buffer and the pipeting amount was increased with 10 µl.

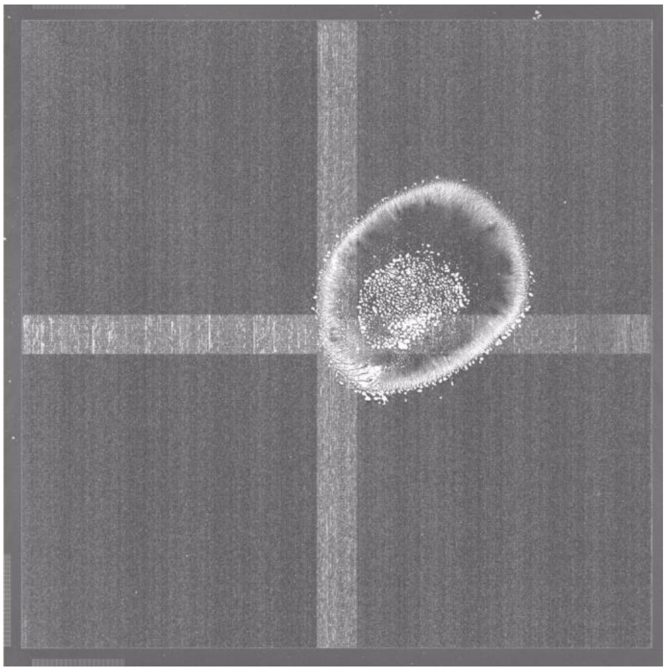


Figure 12. Array image problem

7.2 Data analysis

Affymetrix Genotyping Console (GTC) and Bioconductor were easy and comfortable to use and array analysis did not present any bigger difficulties. GTC program demanded however efficient computer to process and produce such big amounts of data but after acquiring more memory the work was much readable.

In this study, deviating from many other similar WGA studies, sample data was very strictly defined, which may bring us to more specific results concerning T1D. Results can be assumed reliable because they have been statistically examinated with the help of Hardy-Weinberg p-value and critically analyzed.

By results of this study we now know statistically significant SNPs and by that genes where they appear. In the future, with the help of these results, we can study role of those genes in T1D and possible find new essential genes which impact on this disease.

8 REFERENCES

Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0 Manual (2006-2007), P/N 702419 Rev 2

Al-Mutairi H., Mohsen A., Al-Mazidi Z. (2007), Genetics of Type 1 Diabetes Mellitus, Kuwait Medical Journal 39, (2):107-115

BRLMM-P: Genotype Calling Method for the SNP 5.0 Array (2007), Revision Version: 1.0

Chu W. (2005), Affymetrix leads DNA microarray sector, DrugReseacher.com [online 13.03.2009]. Could be found also in www-form: http://www.drugresearcher.com/Tools-and-techniques/Affymetrix-leads-DNA-microarray-sector

Concannon, P., Gogolin-Ewens, K.J., Hinds, D.A., Wapelhorst, B., Morrison, V.A., Stirling, B., Mitra, M., Farmer, J., Williams, S.R., Cox, N.J., Bell, G.I., Risch, N. & Spielman, R.S. (1998): A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat Genet 19(3):292-6.

Davies J.L., Kawaguchi Y., Bennett S.T., Copeman J.B., Cordell H.J., Pritchard L.E., Reed P.W., Gough S.C., Jenkins S.C., Palmer S.M. et al. (1994), A genome-wide search for human type 1 diabetes susceptibility genes, Nature 371 (6493):130-136

Internation Diabetes Institute (2009), Type 1 Diabetes, [online 12.02.2009]. Could be found also in www-form: http://www.diabetes.com.au

HapMap 2003: The International HapMap Project. Nature 426(6968):789-96

Hardy, G.H. (2003), Mendelian proportions in a mixed population. 1908. Yale J Biol Med 76(2):79-80

Hashimoto, L., Habita, C., Beressi, J.P., Delepine, M., Besse, C., Cambon-Thomsen, A., Deschamps, I., Rotter, J.I., Djoulah, S., James, M.R. & et al. (1994): Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 371(6493):161-4.

Hermann R., Lipponen K., Kiviniemi M., Kakko T., Veijola R., Simell O., Knip M., Ilonen J.(2006a) Lymphoid tyrosine phosphatase (LYP/PTPN22) Arg620Trp variant regulates insulin autoimmunity and progression to type 1 diabetes, Diabetologia volume 49, 1198-1208

Hermann R., Lipponen K., Kiviniemi M., Kakko T., Veijola R., Simell O., Knip M., Ilonen J. (2006b) Lymphoid tyrosine phosphatase (LYP/PTPN22) Arg620Trp variant regulates insulin autoimmunity and progression to type 1 diabetes, Diabetologia volume 49, 1198-1208

Kulta A. (2002), Data mining of the DNA microarray experiment in drug development, Opinnäytetyö, Turun Ammattikorkeakoulu, Bio- ja elintarviketekniikka

Mein, C.A., Esposito, L., Dunn, M.G., Johnson, G.C., Timms, A.E., Goy, J.V., Smith, A.N., Sebag-Montefiore, L., Merriman, M.E., Wilson, A.J., Pritchard, L.E., Cucca, F., Barnett, A.H., Bain, S.C. & Todd, J.A. (1998): A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat Genet 19(3):297-300.

Mueller JC, Lohmussaar E, Magi R *et al* (2005) Linkage disequilibrium patterns and tagSNP transferability among European populations. American Journal of Human Genetics 76: 387–398

Nerup, J. & Pociot, F. (2001): A genomewide scan for type 1-diabetes susceptibility in Scandinavian families: identification of new loci with evidence of interactions. Am J Hum Genet 69(6):1301-13.

Numer U. (2005) DNA Microarrays, Taylor & Francis Group, UK

Pavan K.K. , Rao A.A., Rao D.A., Sridhar G.R. (2008), Automatic Generation of Merge Factor for Clustering icroarray Data, International Journal of Computer Science and Network Security vol 8(9)

Pociot, F. & McDermott, M.F. (2001): Genetics of type 1 diabetes mellitus. Genes Immun 3(5):235-49.

Scott J. Tebbutt, James A., Paré P. (2007), Single-Nucleotide Polymorphisms and Lung Disease, Cheast 131(4):1216-1223

Wang K., Bucan M. (2008), Copy Number Variation Detection via High-Density SNP Genotyping, Cold spring harbour protocols 7

World Health Organisation (2008), Definition, diagnosis and classification of diabetes mellitus and its complications, Fact sheet Nº312

Xiao Y., Segal M., Yang Y., Yeh R. (2007), A multi-array multi-SNP genotyping algorithm for Affymetrix SNP microarrays, Bioinformatics 23(12):1459-1467

APPENDICES

APPENDIX 1.

SNP ID	p-value	Gene description
1	4.24E-5	downstream//U2 spliceosomal RNA// Eukaryotic type signal recognition particle RNA
2	2.40E-5	intron//CUB and Sushi multiple domains 2
3	1.66E-5	upstream//Mal, T-cell differentiation protein 2// Collectin sub-family member 10 (C-type lectin)
4	4.24E-5	intron//Pecanex-like 2 (Drosophila)
5	4.05E-5	intron// Methionine sulfoxide reductase A
6	4.24E-5	upstream// Ras association (RalGDS/AF-6) domain family (N-terminal) member 8
7	1.66E-5	upstream// Adenylate cyclase 2 (brain)// LOC442132 // 442132 // Similar to hypothetical protein FLJ36144
8	1.97E-5	upstream// 7SK RNA
9	2.49E-5	intron// ADAM metallopeptidase domain 23
10	4.24E-5	downstream// Septin 8// Ankyrin repeat domain 43// Cyclin I family, member 2///upstream// Ankyrin repeat domain 43
11	3.13E-5	upstream// Noggin//downstream// Ankyrin-repeat and fibronectin type III domain containing 1
12	5.97E-5	downstream// Eukaryotic type signal recognition particle RNA// U2 spliceosomal RNA
13	2.40E-5	upstream// CDC42 binding protein kinase alpha (DMPK-like)// Serine/threonine-protein kinase MRCK alpha
14	1.49E-5	downstream// Eukaryotic type signal recognition particle RNA// Branched chain aminotransferase 1, cytosolic
15	1.97E-5	intron// Solute carrier family 39 (metal ion transporter), member 11
16	2.40E-5	intron// Kazrin// kazrin isoform B
17	2.40E-5	intron// Vacuolar protein sorting 53 homolog (S. cerevisiae)

18	4.24E-5	upstream// Y RNA///downstream// POU domain, class 4, transcription factor 1 (Brain-specific homeobox/POU domain protein 3A)
19	3.83E-5	downstream// B-cell translocation gene 4

APPENDIX 2.

SNP ID	p-value	Gene description
111	4,38E-05	downstream// U7 small nuclear RNA//upstream// Phosphoinositide-3-kinase, class 3
112	4,42E-05	downstream// Neurotrophic tyrosine kinase, receptor, type 3// Transmembrane protein 83
113	4,01E-05	downstream// U7 small nuclear RNA// CDNA FLJ45625 fis, clone BRTHA3028505
114	4,44E-05	upstream// 7SK RNA// U4 spliceosomal RNA
115	8,68E-06	downstream// Zinc finger protein 273
116	4,82E-06	intron// Cadherin 13, H-cadherin (heart)
117	3,61E-05	intron// Fibroblast growth factor 14 (FGF-14) (Fibroblast growth factor homologous factor 4) (FHF-4)
118	1,58E-05	Receptor accessory protein 1
119	2,76E-05	intron// FERM domain-containing protein 5
1110	2,48E-05	downstream// Periostin, osteoblast specific factor//
1111	2,66E-05	intron// Tensin 1
1112	8,26E-06	downstream// IMP2 inner mitochondrial membrane protease-like//upstream// Leucine rich repeat neuronal 3// IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)// Hypothetical LOC154907
1113	3,66E-05	upstream//- - - //
1114	1,27E-05	intron// Cytoplasmic tyrosine-protein kinase BMX (EC 2.7.10.2) (Bone marrow tyrosine kinase gene in chromosome X protein)
1115	3,87E-05	downstream// U1 spliceosomal RNA// 5S ribosomal RNA
1116	3,66E-05	Cannabinoid receptor 1//downstream// U17/E1 small nucleolar RNA// mRNA capping enzyme (HCE) (HCAP1)
1117	3,82E-05	upstream//- - -//
1118	4,44E-05	intron// Sorbin and SH3 domain-containing protein 1

		(Ponsin) (c-Cbl-associated protein) (CAP) (SH3 domain protein 5) (SH3P12)
1119	4,62E-05	downstream// Prostaglandin E receptor 3 (subtype EP3)
1120	1,41E-06	intron// DiGeorge syndrome critical region gene 2// Integral membrane protein DGCR2/IDD precursor
1121	8,98E-06	downstream// Chromosome 3 open reading frame 55
1122	3,69E-05	upstream//- - -//
1123	7,26E-06	upstream// Metallophosphoesterase domain containing 2// Doublecortin domain containing 5
1124	1,16E-05	downstream// Hydroxysteroid (17-beta) dehydrogenase 6 homolog (mouse)
1125	3,66E-05	upstream// Adenosine deaminase, RNA-specific, B2 (RED2 homolog rat)// Double-stranded RNA-specific editase B2

APPENDIX 3.

SNP ID	p-value	Gene description
1201	3,98E-05	downstream//CCR4-NOT transcription complex, subunit 6-like//Chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)
1202	1,15E-06	intron//Sarcoglycan zeta
1203	1,79E-06	intron//Chondroitin beta-1,4-N-acetylgalactosaminyltransferase 1
1204	1,88E-05	upstream//Steroid sulfatase (microsomal), isozyme S//Haloacid dehalogenase-like hydrolase domain containing 1A
1205	1,02E-05	downstream//- - - //
1206	1,05E-06	upstream//D4, zinc and double PHD fingers, family 3//WD repeat domain 21A
1207	2,56E-06	downstream//2-oxoisovalerate dehydrogenase subunit beta, mitochondrial precursor (EC 1.2.4.4) (Branched-chain alpha-keto acid dehydrogenase)
1208	2,56E-06	intron//mitogen-activated protein kinase kinase kinase 15
1209	1,88E-05	intron// RAB33A, member RAS oncogene family
1210	2,76E-05	downstream//Diacylglycerol O-acyltransferase 2-like protein 6
1211	4,34E-06	intron//Nuclear transcription factor, X-box binding-like 1
1212	3,99E-05	intron//Ankyrin repeat and IBR domain containing 1
1213	5,23E-06	downstream//Mdm1 nuclear protein homolog (mouse)
1214	1,07E-05	intron//Slit homolog 3 (Drosophila)// Slit homolog 3 protein precursor (Slit-3) (Multiple epidermal growth factor-like domains 5)
1215	2,70E-05	upstream//RAB1A, member RAS oncogene family//U19 small nucleolar RNA//downstream//Centrosomal protein of 68 kDa (Cep68 protein)
1216	8,08E-06	intron//DDHD domain containing 1

1217	4,15E-05	downstream//Hyaluronan synthase 2//Syntrophin, beta 1 (dystrophin-associated protein A1, 59kDa, basic component 1)
1218	4,20E-05	downstream//Interleukin 17F//Interleukin-17F precursor (IL-17F) (Interleukin-24) (IL-24) (Cytokine ML-1)
1219	2,15E-05	downstream//- - -//
1220	3,24E-05	upstream//- - - //
1221	1,99E-05	intron//C1orf112 protein//SCY1-like 3 (S. cerevisiae)// Chromosome 1 open reading frame 112
1222	3,52E-05	SASH1//AM and SH3 domain containing 1//upstream//Uronyl-2-sulfotransferase
1223	1,38E-05	intron//Collagen, type XXI, alpha 1
1224	3,74E-05	intron//Interleukin 18 (interferon-gamma-inducing factor)
1225	1,79E-05	intron//Hypothetical protein FLJ21511
1226	2,33E-05	downstream//U23 small nucleolar RNA//Transcribed locus//upstream//LIM domain-binding protein 2 (Carboxyl-terminal LIM domain-binding protein 1)
1227	3,33E-06	intron//Platelet-activating factor acetylhydrolase 2 cytoplasmic, 40kDa//PAFAH2//
1228	1,05E-06	intron//Cadherin 13, H-cadherin (heart)
1229	1,21E-05	intron//SEC16 homolog A (S. cerevisiae)
1230	4,94E-06	upstream//Cell adhesion molecule 2//immunoglobulin superfamily, member 4D
1231	3,18E-05	downstream//Ubiquitin specific peptidase 31//Heparan sulfate (glucosamine) 3-O-sulfotransferase 2
1232	1,05E-06	downstream//Family with sequence similarity 148, member B
1233	3,33E-06	intron//F-box and leucine-rich repeat protein 7
1234	2,56E-06	downstream//Chromosome 21 open reading frame 91//B lymphocyte activation-related protein BC-2048
1235	2,15E-05	intron//Dachshund homolog 2 (Drosophila)
1236	4,68E-06	upstream//- - -//

1237	8,08E-06	upstream//Oligonucleotide/oligosaccharide-binding fold-containing protein 1//STE20-like serine/threonine-protein kinase
1238	3,31E-06	downstream//5S ribosomal RNA//V-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)
1239	8,16E-06	intron//Schwannomin interacting protein 1

APPENDIX 4.

SNP ID	p-value	Gene description
1301	1,24E-05	downstream//Ubiquitin ligase protein FANCL (EC 6.3.2.-) (Fanconi anemia group L protein)// Vaccinia related kinase 2//
1302	8,76E-06	intron//Target of myb1-like 2 (chicken)
1303	3,79E-05	upstream//Solute carrier family 10 (sodium/bile acid cotransporter family), member 2//D-amino acid oxidase activator
1304	1,99E-05	downstream//Activating transcription factor 1
1305	4,28E-05	downstream//TWIST neighbor//Fer3-like (Drosophila)
1306	4,17E-07	intron//Protein tyrosine phosphatase, receptor type, K
1307	3,56E-05	upstream//D-amino acid oxidase activator//Solute carrier family 10 (sodium/bile acid cotransporter family), member 2
1308	4,25E-05	downstream//PDCD6 protein (Fragment)// Lysophosphatidylcholine acyltransferase 1
1309	4,32E-05	downstream//U6 spliceosomal RNA//zona pellucida-like domain containing 1
1310	3,17E-05	intron//CREB regulated transcription coactivator 3
1311	4,64E-06	downstream//Interleukin 6 (interferon, beta 2)
1312	4,28E-05	upstream//Y RNA// Meis homeobox 2
1313	2,59E-05	downstream//Chromosome 10 open reading frame 136//Chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)// Stromal cell-derived factor 1 precursor (SDF-1) (CXCL12) (Pre-B cell growth-stimulating factor) (PBSF) (hIRH)
1314	4,25E-05	60S ribosomal protein L37
1315	1,29E-05	upstream//- - -//
1316	6,56E-06	upstream//Lymphocyte antigen 86//Coagulation factor XIII, A1 polypeptide

1317	2,08E-05	upstream// Nuclear receptor interacting protein 1
1318	3,44E-06	upstream//Guanylate cyclase 1, soluble, alpha 2//CWF19-like 2, cell cycle control (S. pombe)
1319	1,09E-05	upstream//Y RNA
1320	3,94E-05	intron//Neurexin 1
1321	4,91E-05	Serpin peptidase inhibitor, clade B (ovalbumin), member 8
1322	4,33E-06	intron//RNA binding motif, single stranded interacting protein
1323	2,79E-07	intron//Latrophilin 2
1324	4,76E-05	upstream//Family with sequence similarity 86, member A
1325	2,17E-05	intron//Nibrin// Homo sapiens nibrin (NBN), transcript variant 2, mRNA
1326	2,20E-05	intron//Forkhead box P1
1327	4,33E-06	downstream//Poly(A) binding protein interacting protein 2B//M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)
1328	2,38E-05	intron//Heparan sulfate 6-O-sulfotransferase 3
1329	3,31E-05	downstream// Zic family member 1 (odd-paired homolog, Drosophila)
1330	4,30E-05	upstream//ADAM metallopeptidase with thrombospondin type 1 motif, 3
1331	4,17E-07	intron// Chromosome 12 open reading frame 50
1332	7,38E-06	upstream//TRNA splicing endonuclease 2 homolog (S. cerevisiae)// Peroxisome proliferator-activated receptor gamma
1333	9,23E-06	upstream//- - -//U6 spliceosomal RNA
1334	4,01E-05	intron//Transcription factor 7-like 1 (T-cell specific, HMG-box)
1335	4,17E-07	downstream//CDNA FLJ46681 fis, clone TRACH3010382// Polypeptide N-

		acetylgalactosaminyltransferase 17//Heat shock protein 90Af
1336	1,94E-05	Family with sequence similarity 19 (chemokine (C-C motif)-like), member A4
1337	1,49E-06	downstream//Receptor accessory protein 3// Leucine rich repeat transmembrane neuronal 3//Catenin (cadherin-associated protein), alpha 3
1338	3,59E-05	upstream//C-type lectin domain family 2, member A
1339	4,17E-07	intron// Monoacylglycerol O-acyltransferase 1
1340	4,17E-07	upstream//Par-3 partitioning defective 3 homolog B (C. elegans)// Amyotrophic lateral sclerosis 2 chromosome region candidate gene 19 protein (Partitioning-defective 3-like protein)// Inducible T-cell co-stimulator
1341	3,57E-05	upstream//VPS10 domain-containing receptor SorCS1 precursor (hSorCS)// 5S ribosomal RNA
1342	1,08E-06	upstream//Serpin B5 precursor (Maspin) (Protease inhibitor 5)// Vacuolar protein sorting 4 homolog B (S. cerevisiae)
1343	1,29E-05	upstream//Syntaxin 3// Olfactory receptor, family 10, subfamily V, member 1//Fatty acid-binding protein, epidermal (E-FABP) (Psoriasis-associated fatty acid-binding protein homolog) (PA-FABP)
1344	2,65E-05	upstream//Solute carrier family 19 (folate transporter), member 1//Chromosome 21 open reading frame 123//Collagen, type XVIII, alpha 1
1345	4,17E-07	upstream//Chromosome 15 open reading frame 57//RNA pseudouridylate synthase domain containing 2
1346	2,79E-05	downstream//U6 spliceosomal RNA//SP100 nuclear antigen
1347	2,25E-05	downstream//U6 spliceosomal RNA//Chromosome X open reading frame 27//Huntingtin interacting protein HYPM (Fragment)

1348	4,01E-05	intron//X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining)
1349	3,31E-05	upstream//Spermatogenesis associated 13
1350	1,16E-05	intron// Plexin C1
1351	3,31E-06	intron//V-set domain containing T cell activation inhibitor 1
1352	4,01E-05	Hermansky-Pudlak syndrome 1 protein
1353	1,83E-06	upstream//OPALIN//Oligodendrocytic myelin paranodal and inner loop protein//Tolloid-like 2//Transmembrane protein 10
1354	7,91E-07	intron//Neurexin 3
1355	4,87E-05	upstream//Pellino homolog 1 (Drosophila)
1356	5,41E-06	intron//Mediator complex subunit 7
1357	3,26E-05	intron//Na+/K+ transporting ATPase interacting 2
1358	2,28E-05	intron//CDNA FLJ31737 fis, clone NT2RI2007084//Chromosome 10 open reading frame 72
1359	1,26E-05	intron//Roundabout, axon guidance receptor, homolog 2 (Drosophila)
1360	4,25E-05	intron//Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3
1361	2,71E-05	intron//Catenin (cadherin-associated protein), alpha 3
1362	1,47E-05	intron//ATPase, Ca++ transporting, plasma membrane 2
1363	6,98E-06	upstream//- - -//downstream//5S ribosomal RNA
1364	4,17E-07	downstream//Family with sequence similarity 84, member A//Neuroblastoma-amplified protein
1365	9,36E-07	downstream//Iroquois homeobox 5