CANDIDATE GENE STUDY ON NICOTINE DEPENDENCE IN FINNISH SIBPAIRS

Katri Heikkilä

PREFACE

This study has been performed at the National Public Health Institute, Department of Molecular Medicine, Biomedicum, Helsinki, during April 2004 and June 2005. The epidemiological data collection has been performed by University of Helsinki, Department of Public Health. Writing of the thesis has been finalized in 2008.

I want to thank Professor Leena Peltonen-Palotie for the possibility to participate in this interesting and captivating research. The facilities in Biomedicum created amazing surroundings to work and made everything very comfortable. This project taught me a lot about the Finnish research world and made me understand it much better.

Professor Jaakko Kaprio is thanked for enjoyable collaboration and for providing the excellent study sample for this study. The world of epidemiology was totally new for me and prof. Kaprio was able to make it quite easily understandable.

My supervisor, Ph.D. Anu Loukola…well, how can I thank you enough! It has been a long way until this point and it has been a very memorable time. Your enthusiasm towards this project and science is something I had never seen before. Your drive, usually without coffee, kept also me going and made me try harder, even at the times when everything did not look quite good.

All the people at the Department of Public Health, and especially Ulla, thank you for the nice meetings and all the work every one has done for the nicotine project. And Ulla, the secrets of genetics were difficult in the beginning for me too.

And then the best workmates ever! Thank you Heidi for helping me to get started and sharing the same aspects on nicotine study project. Marika, William, Juho, Joni, Emma N, Pia, Helena, Annu and all the others, thank you for the coffee breaks and the loud conversations that maybe not everyone could understand.

Emma P and Olli, although time has passed and we all are doing different things at the moment, it is amazing how the friendships seem to stay the same. Without you two, this project would have felt much different and the life after it would be little emptier.

Last, but not the least, my parents. Thank you for all the support during this project and making it possible to be able to go back to school.

And Tommi, thank you for being there in good and bad times.

Author:	Katri Heikkilä
Title of thesis:	Candidate gene study on nicotine dependence in Finnish sibpairs
Finnish title:	Nikotiiniriippuvuuden ehdokasgeenitutkimus suomalaisilla sisaruspareilla
Date:	27.5.2008 **Pages:** 43 + 14

Department:	Department of Biological and Environmental Science
Chair:	Biotechnology
Supervisor(s):	FT Anu Loukola

Abstract:

Smoking is the most common reason for preventable untimely death. Tobacco products are related to one fourth of the self-inflicted diseases. Smoking is the most important known risk factor of cancers. About 30 % of the diagnosed cancers in Finland are related to smoking. Smoking increases heart attack risk by 2-fold. According to WHO approximately five million people die due to tobacco use every year. It has been estimated that this number will increase to ten million people by 2030.

Very often the closest relatives of a smoker also become smokers. Therefore, it is believed that also genetic variants are related to smoking. Tobacco consists of over 4000 different chemical compounds, but only one of these compounds, nicotine, is believed to be addictive.

Family, twin and adoption studies have revealed a strong genetic component in nicotine addiction. However, surprisingly little is known about the genetics of smoking and nicotine dependence. In this candidate gene study the aim was to identify genes involved in nicotine dependence in Finnish twin families.

The study sample consisted of Finnish twins and their siblings born in 1945 – 1955. They were selected based on history of heavy smoking in at least two sibs. Nicotine dependence has been evaluated from a diagnostic telephone interview using the DSM-IV criteria and the Fagerström score. 720 individuals were included in this study. Studied phenotypes were nicotine dependence (DSM-IV score) and regular smoker (smoker or non-smoker).

The allelic spectrum of candidate genes was analyzed by using single nucleotide polymorphisms (SNPs). Six candidate genes with functional relevance were selected from the literature: four nicotinic acetylcholine receptor genes and two cytochrome P450 genes. Sequenom's Mass ARRAY technology was utilised to genotype 26 SNPs that were selected based on their genetic location, polymorphic nature and linkage disequilibrium in Finnish families. Family based association analyses were then used to evaluate the role of specific alleles.

The results show statistically significant association in three genes with nicotine dependence: *CHRNA5* (rs667282, rs3743078), *CHRNA4* (rs2273502, rs2273505) and *CHRNB1* (rs2302763, rs871990) and two genes with regular smoker phenotype *CHRNA5* (rs3743078) and *CHRNB1* (rs2302767, rs2302762). These results with nicotine dependence phenotype are similar to the ones reported earlier.

In short, these results show that genetic components play a role in nicotine dependence and give us more perspective in revealing them in more detail.

Tekijä:	Katri Heikkilä
Tutkielman nimi:	Nikotiiniriippuvuuden ehdokasgeenitutkimus suomalaisilla sisaruspareilla
English title:	Candidate gene study on nicotine dependence in Finnish sibpairs
Päivämäärä:	27.5.2008　　　　　　　　　**Sivumäärä:** 43 + 14

Laitos:	Bio- ja ympäristötieteiden laitos
Oppiaine:	Biotekniikka
Tutkielman ohjaaja(t):	FT Anu Loukola

Tiivistelmä:

Tupakointi on yleisin estettävissä oleva ennenaikaisen kuoleman syy. Noin neljäsosaan itseaiheutetuista sairauksista syyllisiä ovat tupakkavalmisteet. Suomessa noin 30 % syöpäsairauksista aiheutuu tupakasta. Tupakointi alentaa myös merkittävästi elinkaaren pituutta. WHO:n mukaan tupakoinnista johtuvista syistä joka vuosi kuolee noin viisi miljoonaa ihmistä. Arvion mukaan luku nousee vuoteen 2030 mennessä 10 miljoonaan. Teollisuusmaissa tupakka aiheuttaa joka viidennen kuolemantapauksen. Tupakointi lähes kaksinkertaistaa sydäninfarktin riskin.

Tupakoijien lähisukulaisista tulee hyvin usein myös tupakoitsijoita. Tämän vuoksi uskotaan, että henkilön perimä, vanhemmilta peritty geenistö, vaikuttaa osaltaan taipumukseen tupakoida. Vaikka tupakoidessa elimistöön tulee yli 4000 erilaista yhdistettä, on nikotiini se yhdiste joka aiheuttaa riippuvuuden. Nikotiini on hermojen välittäjäaineen, asetykoliinin, kaltainen keskushermostoa stimuloiva aine.

Kaksos- ja perhetutkimukset kehittyneitä molekyylibiologisia menetelmiä hyväksikäyttäen, mahdollistavat altistavien geenien selvittämisen. Assosiaatiotutkimusten perusteella voidaan pyrkiä selvittämään nikotiiniriippuvuuden monitekijäistä periytymistä. Assosiaatiotutkimuksissa pyritään osoittamaan että tietty geneettinen osoitin, eli geenissä tapahtunut luonnollinen yhden emäksen muunnos, ja sairastumisalttius liittyvät toisiinsa siten, että osoitin havaitaan useammin sairastuneissa yksilöissä kuin terveissä yksilöissä.

Tutkimuksessa oli mukana 720 henkilöä aineistosta, joka koostuu vuosina 1945–1955 syntyneistä suomalaisista kaksosista ja heidän sisaruksistaan. Valintakriteerinä on käytetty erittäin runsaan tupakoinnin esiintymistä vähintään kahdessa sisaruksessa samassa perheessä. Ilmiasuina analyyseissä käytettiin sekä nikotiiniriippuvuutta että säännöllistä tupakointia. Nikotiiniriippuvuuden mittarina käytettiin DSM-IV kriteerejä sekä Fagerströmin asteikkoa.

Tutkittavia ehdokasgeenejä oli kuusi, *CHRNA4, CHRNA5, CHRNA7, CHRNB, CYP1A2* ja *CYP2A6*. Yhden emäksen monimuotoisuuden osoittimia valittiin geneettisen sijainnin, monimuotoisuuden sekä kytkentäepätasapainon perusteella 26 kappaletta. Osoittimien tutkimiseen tutkimusaineistosta käytettiin Sequenom Mass Array – tekniikkaa sekä tietokoneavusteisia analyysejä.

Kvalitatiiviset assosiaatioanalyysit osoittavat tilastollisesti merkittävää, $p < 0.05$, assosiaatiota useiden osoittimien ja ilmiasujen väillä, kolmella geenillä nikotiiniriippuvuuden suhteen: *CHRNA5* (rs667282, rs3743078), *CHRNA4* (rs2273502, rs2273505) ja *CHRNB1* (rs2302763, rs871990) sekä kahdella geenillä säännöllisen tupakoinnin suhteen: *CHRNA5* (rs3743078) ja *CHRNB1* (rs2302767, rs2302762). Nikotiiniriippuvuusilmiasulla saadut tulokset ovat hyvin samankaltaisia kuin mitä aiemmin on raportoitu.

Tutkimus antaa vahvan näytön siitä, että perinnöllisillä tekijöillä on vaikutusta nikotiiniriippuvuuteen. Perinnöllisten tekijöiden tämän kaltaista sekä laajempaa tutkimusta tulisi jatkaa ja tutkittavien ilmiasujen määrää tulisi lisätä.

TABLE OF CONTENTS

ABBREVIATIONS

AA	African-American
BLAST	Basic Local Alignment Search Tool
cM	centi Morgan
CNS	central nervous system
COGA	Collaborative Study on the Genetics of Alcoholism
dom	dominant model
DSM	Diagnostic and Statistical Manual of Mental Disorders
EA	European-American
EDTA	ethylene diamine tetra acetic acid
FBAT	Family-Based Association Test
FTND	Fagerström Test for Nicotine Dependence
FTQ	Fagerström Tolerance Questionnaire
hME	homogenous MassEXTEND assay
HVA	homovanillic acid
HWE	Hardy-Weinberg equilibrium
ICD	International Classification of Diseases
LD	linkage disequilibrium
LOD score	logarithm of odds (a measure of the likelihood of genetic linkage between loci)
MAF	minor allele frequency
MALDI-TOF MS	matrix-assisted laser desorption / ionisation time-of-flight mass spectrometry
nAChRs	nicotinic acetylcholine receptors
NDSS	Nicotine Dependence Syndrome Scale
NPHI	National Public Health Institute, Finland
rec	recessive model
RTQ	Revised Tolerance Questionnaire
SAP	shrimp alkaline phosphatase
SNP	single nucleotide polymorphism
WHO	World Health Organisation

Abbreviations of the gene names are provided in the context of each of the genes described in the text

1   INTRODUCTION

1.1   Smoking and public health

World Health Organisation has estimated that one third of the adult population worlwide smokes. In Finland, 23.5% of the adult population smoke. The annual cigarette consumption in Finland is 1 351 cigarettes per person. (Mackay and Eriksen, 2002). Smoking is the most common reason for preventable untimely death. Tobacco products are related to one fourth of the self-inflicted diseases. Smoking is the most important known risk factor for cancers. About 30 % of the diagnosed cancers in Finland are related to smoking. Smoking increases heart attack risk by 2-fold. Every other smoker will die of smoking, if they do not quit early enough. Thus, smoking is a significant causing factor in morbidity and work disability.

During the last 50 years smoking habit among men has decreased in Finland, but at the same time smoking among women has increased. In the year 2000, 20% and 28% of working aged females and males, respectively, were daily smokers. In the future the increase on women's smoking habit can be seen in the death statistics, where the number of deaths caused by smoking related diseases will increase among women and will even the numbers between sexes. Compared to the rest of Europe the Finns smoke little, but we have as our speciality the problem with very low initiation age. (Patja K and Haukkala A, 2004)

According to a study made by WHO, approximately 5 million people die due to tobacco use every year. (Mackay and Eriksen, 2002) It has been estimated that this number will increase to 10 million people by 2030. In industrialised countries tobacco causes one fifth of the deaths per year. Smoking related diseases cost to the Finnish society about 2 billion euros per year. A smoker uses the public health care 22% more often than a non-smoker. The Finnish General Medicine association defines nicotine addiction as a chronic disease, which requires treatment.

1.2   Nicotine dependence and addiction

Tobacco is the most commonly used heavily addictive substance around the world (Salaspuro M et al., 2003). It generates physiological, social, psychological and emotional

addiction. People smoke to maintain nicotine levels in the brain, to avoid the negative effects of nicotine withdrawal and to control their mood. Some smoke to calm them down, to focus better, to solve problematic situations or to cheer themselves up.

Tobacco consists of over 4000 different chemical compounds, but only one of these compounds, nicotine, is believed to be addictive. Nicotine addiction is very problematic because of its individual nature. The possible genetic determinants behind the addiction are still largely unknown. According to the international psychiatric classification system nicotine addiction fulfils the criteria for chemical addiction. The affects on behaviour are not as severe as with classical drugs, but the addiction itself can be as severe. Because of the strong addictiveness of nicotine, quitting smoking is very difficult. About 60% of the individuals that smoke in Finland would like to quit, but only 3-8 % of the ones who attempt to quit succeed. (Salaspuro M et al., 2003)

Nicotine addiction is a complex multifactorial behaviour with both genetic and environmental determinants (Feng et al., 2004; Niu et al., 2000).

1.3 Dependence scales

There are several dependence scales available for the evaluation of nicotine addiction. American Psychiatric Association has established the Diagnostic and Statistical Manual of Mental Disorders (DSM), also containing a diagnostic scale for the recognition of nicotine addiction. Originally this criterion was designed for a wide range of substances and nicotine was added afterwards. In the fourth edition of DSM (DSM-IV) nicotine dependence and nicotine withdrawal are included as disorders. Nicotine abuse is not included since psychosocial problems and intoxication are not clinically significant. The DSM-IV questionnaire consists of seven criteria and three of them must be fulfilled within a twelve-month period. Some example questions are attached in Appendix 1a. (American Psychiatric Association, 1994)

WHO created an International Classification of Diseases (ICD) to consistently diagnose diseases clinically worldwide. In the tenth revision of the ICD (ICD-10) also nicotine addiction was included. In ICD-10 there are six criteria, and three or more of them need to

be fulfilled simultaneously at some time during the previous year. See appendix 1b for the criteria. (WHO, 1992)

In 1978 an eight-question Fagerström Tolerance Questionnaire (FTQ) was introduced, attempting to assess physical dependence on nicotine. FTQ is the most widely used dependence questionnaire for tobacco and nicotine (Fagerström, 1978). FTQ has been revised in 1994 when the eight questions were reduced to six; the revised version is called the Fagerstöm Test for Nicotine Dependence (FTND). The scale was improved by renewal of scoring in questions of time of the first cigarette of the day and number of cigarettes smoked per day. (Heatherton et al., 1991) The maximum score is ten and scores seven or more identify physical dependence on nicotine. The questionnaire is in Appendix 1c.

FTQ has been criticized to have somewhat poor psychometric properties. In the Revised Tolerance Questionnaire (RTQ) wording, scale format and modification of some items were enhanced. This scale uses Likert scales, in which the subjects indicate the level of agreement or disagreement using a five-point scale. Maximum score is therefore 30 and this way it gives more accurate response in behaviour. (Tate and Schmitz, 1993)

Nicotine Dependence Syndrome Scale (NDSS) evaluates, in a multifactor way, tobacco dependence. It consists of 19 questions and scores are divided in 5 subscales. NDSS measures drive, priority, tolerance, continuity and stereotypy. (Shiffman et al., 2004)

1.4   Nicotine

Nicotine as a clean substance is toxic. 60 mg is a lethal dose for an adult. Nicotine increases the pulse rate and blood pressure. As large doses it causes symptoms of toxication and the acute toxicity is so high that even a person who has smoked for a long time typically can smoke no more than 60 cigarettes per day.

When the cigarette smoke is inhaled into the lungs, 25% of the nicotine dose is absorbed into the blood stream. It takes only seven seconds for nicotine to reach the brain. The affects in the central nervous system (CNS) are immediate. The nicotine level in the plasma and the brain is very high and this causes a rewarding effect. A remarkable part of

the effects of nicotine are due to the increased release of neurotransmitters in the CNS. Nicotine also increases the secretion of glucocorticoids and vasopressin, which has an anti-diuretic effect. The positive and negative effects of nicotine are listed in Table I.

Table I Positive and negative effects of nicotine in the CNS

Positive effects	Negative effects
Improvement of cognitive functions	Hypothermia
Improvement of attentiveness	Ataxia
Decrease of anxiety	Cramps
Neuroprotection	Nausea
Analgesia	Vomiting
	Development of addiction

The rewarding effect vanishes quite quickly because the nicotinic receptors desensitise and nicotine metabolises. The main metabolite is cotinine, which is metabolised in 24 hours. Compared to the half-life of nicotine, which is 30 minutes, the difference is huge. The amount of smoked cigarettes per day can be measured from cotinine levels in blood or urine.

Nicotine binds to nicotinic acetylcholine receptors. These receptors are located in skeletal muscle tissue, medulla, pre-synaptically throughout the central nervous system and post-synaptically in autonomic nervous system. These nervous systems modulate the release of neurotransmitters and ganglionic potentials. (Role and Berg, 1996)

1.5 Twin and adoption studies

Two commonly used approaches to examine genetic and environmental components of phenotypic variance are twin and adoption studies. Twin studies examine the concordance rates for traits of interest. They are based on the fact that monozygotic twins share identical genetic material whereas dizygotic twins are genetically comparable to non-twin siblings. Both monozygotic and dizygotic twins share the environment during prenatal period;

furthermore, twins typically share environment to a great deal also during childhood and adolescence.

Adoption studies limit the effect of family environment and give more information about how far the component is genetic. The challenges are the lack of information about the biological family and difficulties to get this information. In the ideal case of adoption studies the behaviour and genetic information of children who grew up separated from their biological parents is compared with those growing with the biological parents.

A meta-analysis, which involved 14 twin studies including 17 500 twins with smoking dependence, indicated that 56% of the variance is genetic, 24% familial-environmental and 20% individual specific. (Kendler et al., 1999) There have been several twin studies which have examined the genetic influence to nicotine dependence. The most recent ones have shown that the heritability varies from 50-70% giving a clear indication of a genetic factor for smoking persistence in regular smokers. (Lessov et al., 2004; Maes et al., 2004; True et al., 1997)

1.6 Linkage studies

The development of molecular biology and biocomputing has made it possible to look for genome-wide localisation of genes that contribute to certain phenotypes. Genome-wide linkage studies aim to localize chromosomal regions in which affected family members share alleles. Those regions showing statistically significant linkage may harbour gene variants affecting the studied phenotype. Traditionally genome-wide linkage studies have been performed with a commercially available set of microsatellite markers. Approximately 400 markers, evenly distributed throughout the genome, have been used. There are reports of several genome-wide linkage scans related to smoking and nicotine dependence from the recent years.

With never/ever-smoked phenotype, some evidence of linkage has been reported in chromosomes 6, 9 and 14 (Bergen et al., 1999) (Bierut et al., 2004) and in chromosomes 3, 4, 15 and 17. (Duggirala et al., 1999) All of these studies have utilised the same study material, the Collaborative Study on the Genetics of Alcoholism (COGA) families collected in the U.S.A. In another study, two different populations, one from New Zealand

and one from Richmond, U.S.A. were used to study linkage for nicotine dependence. Linkage was found in chromosomes 2, 4, 10, 16, 17 and 18. (Straub et al., 1999) Significant linkage for smoking rate has been reported in the Framingham Heart Study population (U.S.A.) in chromosome 11 and also some suggestive linkage was found for the same phenotype in chromosomes 4, 7 and 17, and for nicotine dependence in chromosomes 7, 11 and 17. (Li et al., 2003) Linkage for the number of cigarettes smoked has been reported in chromosomes 3 and 10 and for smoking initiation in chromosomes 6, 10 and 14 in Dutch sibling pairs. (Vink et al., 2004)

The latest genome-wide scan was reported in year 2007 and it was a joint study of Australian and Finnish families. The phenotype was maximum number of cigarettes smoked in a 24-hour period. Both of the data sets were analysed separately as well as combined. The combined scan showed linkage to chromosome 22 with a LOD score of 5.21. In the Finnish study sample also chromosome 20 showed a LOD score of 4.22 (Saccone et al., 2007) Part of the Finnish families in this genome-wide scan are also included in this study.

Another study published in 2007 replicated some previously reported linkage results. This study utilised Finnish twins. Studied phenotypes were smoker, nicotine dependence and nicotine dependence combined with alcohol use. Smoker phenotype gave a LOD score of 3.12 in chromosome 10, nicotine dependence a LOD score of 2.50 on chromosome 7 and 2.25 on chromosome 11. The combined phenotype showed evidence of linkage at the same loci with nicotine dependence. (Loukola et al., 2007)

1.7 Association studies

It is known that several genes with a small overall contribution and relative risk most likely cause complex diseases, like addictions and mental disorders. Association studies look for specific alleles that are shared among affected individuals, either across families or by using a case-control cohort. Association studies can either be genome-wide association (GWA) or candidate gene studies.

1.7.1   Candidate gene approach

Candidate gene studies focus on genes that have been selected either because of their location within an identified linkage region (i.e., positional candidate genes) or because of their hypothetical aetiological role in disease (i.e., functional candidate genes). Functional candidate gene selection process is started with a careful investigation of earlier published studies and by evaluating known of hypothesised biological pathways involved in the studied phenotype. There might be biological, aetiological and pathological models of the disease and e.g. expression studies can give an idea of the genes involved in the pathology of certain diseases. It is important to examine if the chosen genes have functional consequences and therefore animal models also give more information of the involved pathways.

After the candidate genes have been chosen, the type of polymorphisms that will be used need to be decided. In the genome there are several types of polymorphisms: SNPs (single nucleotide polymorphisms), repeats and insertions or deletions, ranging widely in size. Most common sequence variations in DNA are SNPs, with 10 million SNPs estimated in the human genome. (Strachan T and Read A, 2004; dbSNP, 2007) The number of SNPs in a given gene can be remarkably high, which makes the selection of SNPs very important. Traditionally it has been thought that SNPs that affect the function of a protein or its expression should be chosen. On the other hand, functional data is not always available. Therefore, Tabor and colleagues suggest that SNPs should be prioritised based on their likelihood to affect gene function. DNA variations that affect amino acid sequence of a protein (missense variant) or result in a premature stop codon (nonsense variant) should be given the highest priority. These variations are most likely to affect the function of the protein, and to be involved in disease aetiology. Also the selected SNPs should have a reasonable minor allele frequency (e.g. over 5%), in order to have enough statistical power to detect the plausible association. (Tabor et al., 2002)

Currently, the SNP selection is typically based on the HapMap information. The HapMap project has gathered genotype data from four geographically diverse populations using over 3.4 million SNPs. (HapMap Consortium, 2005 and 2007) Using the HapMap database, the tagging SNPs at the region of interest are chosen. For example minor allele

frequencies, linkage disequilibrium and population can be used as selective criteria. (HapMap.org, 2007) The tagging SNPs can be considered as the most informative SNPs, i.e., the minimum set of SNPs needed to extract the allelic diversity of the studied region.

1.7.2   Smoking and nicotine dependence candidate gene studies

One of the first association studies related to smoking and nicotine dependence was published in 2000. In the study most of the introns and exons of *CHRNB2* were screened and five novel SNPs were found. Four out of those five SNPs were tested in a case-control study of 317 non-smokers, 238 regular smokers with low levels of nicotine dependence and 317 regular smokers with high levels of nicotine dependence. No association was found with the four SNPs or their estimated haplotypes with smoking initiation or progression to nicotine dependence. (Silverman et al., 2000)

In year 2004 a study with six SNPs in *CHRNA4* and four SNPs in *CHRNB2* was published. The study sample consisted of 901 individuals from 222 nuclear families from Anhui province in China. Only males were included in the analysis because of the low frequency of female smokers (5.8%). Statistically significant association between two SNPs in *CHRNA4* was found in univariate FBAT (family based association test) analysis and in haplotype FBAT. These two SNPs, rs1044396 and rs1044397, showed protective effects against nicotine addiction in an additive model. *CHRNB2* showed no association with the four chosen SNPs. (Feng et al., 2004)

Another study on nicotinic acetylcholine receptors was published in 2005. In this study six SNPs in *CHRNA4* and four in *CHRNB2* were studied for association with nicotine dependence. Three SNPs in both of the genes were the same ones as Feng et al. had studied. The study sample consisted of 2037 individuals from 602 nuclear families. Individuals were recruited based on their ancestry, African-American (AA) or European-American (EA), and 75.7% of the recruited families had at least two siblings whose FTND score was ≥ 5. Number of individuals with AA and EA ancestry was 1366 and 671, respectively. The study sample was investigated by dividing it by ethnicity and gender. Two SNPs in *CHRNA4* were associated in the EA sample (rs2273504 and rs1044396) and in the AA sample (rs3877137 and rs2236196). Also haplotype analysis showed significant association in AA females with one *CHRNA4* haplotype. (Li et al., 2005)

CHRNA7 has been linked to smoking and schizophrenia and other mood disorders. A 2-bp deletion in exon 6 of *CHRNA7* was studied with schizophrenia and major depressive order related to heavy smoking. No association was seen in schizophrenic patients; however, a modest allele frequency difference (p = 0.037) between major depressive patients and controls was seen. (Lai et al., 2001)

There have been two studies in the Japanese population investigating association between smoking and a *CYP2A6* deletion polymorphism. Ando and colleagues genotyped 240 Japanese adults and looked for signs of protective effect against smoking behaviour, but no clear effect was seen between homozygous and heterozygous deletion allele carriers (Ando et al., 2003). Zhang and colleagues studied 237 individuals for the same deletion, but did not find any kind of association on smoking behaviour (Zhang et al., 2001). Schoedel and colleagues have also studied *CYP2A6* for association of genetically slow nicotine metabolism and smoking. They investigated four alleles of *CYP2A6* associated with decreased nicotine metabolism and found out that slow nicotine inactivators are less likely to become adult smokers and they also smoke fewer cigarettes per day. (Schoedel et al., 2004)

In another study utilizing the same sample set of 1050 cases and 879 controls, SNPs were selected from gene families known to relate with nicotine dependence. A total of 296 genes from nicotinic receptors, dopaminergic receptors and gamma-aminobutyric acid (GABA) receptors were included in the study and the total number of genotyped SNPs was 3713. The strongest association signals were shown in *CHRNB3*. Also the *CHRNA5-CHRNA3-CHRNB4* gene cluster showed significant association. Furthermore, *KCNJ6*, which is related to synaptic regulation of transmission, and *GABRA4*, part of the inhibitory transmitter system of CNS, also showed association. (Saccone et al., 2007)

A study focusing on smokers vs. non-smokers and successful vs. unsuccessful cessation was published in 2007. A set of 520 000 SNPs was genotyped utilizing the Affymetrix platform in 134 nicotine dependent cases having dependence on other substances and in 320 non-smoker controls. The most interesting associations arose in genes involved in cell adhesion processes (e.g., *CDH13)*, enzymatic activity and G-protein coupled protein (e.g.,

cyclic g dependent protein kinase gene, *PRKG1)*. These genes were significant with both studied phenotypes. (Uhl et al., 2007)

1.7.3   Genome-wide association studies (GWA)

Recently, genome-wide association studies (GWA) have become available as the HapMap project has proceeded. Commercially available genotyping platforms include hundreds of thousands of pre-defined SNPs evenly spaced along the genome. The GWA approach proceeds in two stages. In the first stage a group of evenly distributed cases and controls are genotyped using the selection of SNPs. Based on the p-values of association analysis, the significant SNPs or genes are selected for the second phase. This second SNP selection is then genotyped with larger sample size to keep the power of the analyses high. (Affymetrix, 2007 and Bierut et al., 2007)

Nicotine dependence has been studied also by using the GWA approach. A large study sample of 1050 cases and 879 controls from U.S.A. and Australia was genotyped with 2.4 million SNPs utilizing the Perlegen platform. A total of 41 042 SNPs were selected for Stage II and genotyped in 1050 cases and 879 controls. The most interesting results were seen for *Neurexin 1* (*NRXN1*), a gene coding for a vascuolar sorting protein (VSP13A), and *beta3 nicotinic receptor subunit* (*CHRNB3*). (Bierut et al., 2007)

1.8   Nicotinic acetylcholine receptors

Neuronal nicotinic acetylcholine receptors (nAChRs) form a superfamily of pentameric ligand-gated ion channels that mediate fast signal transmission at synapses. The endogenic agonist of nAChRs is acetylcholine. Epibatidine is also known to activate nAChRs. When nicotine binds to a post-synaptic receptor, the cation-ion channel opens and allows Na^+-ions to flow into the cell. Depolarisation of the cell then occurs and creates an action potential. Depolarisation keeps the receptor partly active for an extended time and because of this, the receptor becomes desensitised. Desensitisation causes very rapid occurrence of tolerance which is a strategic implication in nicotine dependence.(Leonard and Bertrand, 2001; Wonnacott, 1997) Long-term exposure to nicotine causes an increase in the number of nAChRs in the brain. This increase is specific to high affinity nAChRs, such as alpha4beeta2-heteropentamers. (Buisson and Bertrand, 2001; Wonnacott et al., 2000)

Nicotine receptors are also involved in development and neuronal plasticity. Nicotine can abnormally alter cell proliferation and differentiation.(Dani and De Biasi, 2001; Slotkin, 1998)

Currently 12 neuronal nicotinic acetylcholine receptors, alpha2-alpha10 and beeta2-beeta4, are known in human. These subunits typically form heteropentameric receptors. Subunits alpha7-alpha10 form homomeric channels but only alpha7 is widely distributed in the mammalian CNS. Different subunit combinations have some what different properties. Alpha7 homo-oligomeric receptors have faster kinetics and higher calcium permeability than other nAChRs whereas alpha4beeta2 receptors have higher affinity for nicotine.

1.9 Metabolising genes

The cytochrome P450 enzymes are expressed in the liver. They play a major role in metabolising e.g. steroids and drugs. Cytochrome P450 enzymes are responsible for most of the phase I metabolising events of drugs and environmental pollutants.

The majority of nicotine is metabolized into cotinine by cytochrome P450 2A6 enzyme (*CYP2A6*). It has been shown that individuals lacking fully functional *CYP2A6* are significantly protected against becoming tobacco dependent smokers. Individuals carrying *CYP2A6* null-alleles smoke less and they should have decreased risk of tobacco smoking related diseases. Two studies have investigated the possible association between smoking behaviour and *CYP2A6*; however, no significant association has been found. (Piannezza et al. 1998; London et al. 1999; Sabol and Hamer, 1999)

CYP2D6 is involved in the oxidation of nicotine into cotinine (Batra et al., 2003). It has been proposed that individuals with different allelic variants of this gene differ also by their nicotine metabolising rates. However, the results so far have been controversial.

The role of *CYP1A2* in nicotine metabolism is unclear, but it has a role in drug metabolism. Nicotine affects by increasing *CYP1A2* function in metabolising some pharmaceuticals like

benzodiazepine, haloperidol and clozapine. As a result e.g. the calming effect of benzodiazepine decreases. (Liang et al., 1996)

1.10 Other candidate genes

Dopamine genes: receptors and transporter

Similarly to other addictive substances, nicotine reinforces the release of dopamine in the mesolimbic dopaminergic system. It binds to the neuronal nicotinic acetylcholine receptors in dopamine neurons of the ventral tegmental area. When exposure to nicotine is continuous, this action is followed by desensitisation. When short exposure to nicotine occurs, the dopamine route creates a rewarding effect, but when the exposure is chronic, the desensitisation causes tolerance. According to recent studies dopamine might also have a role on nicotine's aversive properties. (Laviolette and van der Kooy, 2004)

There are five dopamine receptor genes: *DRD1, DRD2, DRD3, DRD4* and *DRD5*. *DRD2* and *DRD4*, as well as the dopamine transporter gene, *SLC6A3,* have been reported to show association with smoking (Shields et al., 1998; Spitz et al., 1998); (Lerman et al., 1999; Sabol et al., 1999). The remaining dopamine receptor genes have either not been extensively studied so carefully or the results have been controversial. Therefore, all the dopamine genes are worth of further studying.

Serotonin receptor and transporter

Nicotine increases serotonin secretion and withdrawal reduces it. Based on these facts the serotonergic system may be related to nicotine dependence. There are at least three known serotonin receptor genes and one transporter gene, but none of them has been reported to show any significant association with smoking; however, some association has been found with alcohol dependence among Japanese alcoholics with inactive aldehyde dehydrogenase-2 and heroin dependence. (Nakamura et al., 1999; Sáiz et al., 2008)

Other interesting candidate genes

Genes that encode enzymes involved in the metabolism of neurotransmitters could also be excellent candidate genes in nicotine dependence. *Monoamine oxidase* genes A and B (*MAOA, MAOB*) have been shown to be related with cigarette smoking. Smoking decreases the levels of *MAOA* and *MAOB* in the brain. (Fowler et al., 1996a; Fowler et al.,

1996b; Fowler et al., 1998) Catechol-O-methyltransferase (*COMT*) metabolises dopamine to homovanillic acid (HVA) (Lundstrom et al., 1995). The level of HVA is lower in cerebral spinal fluid of smokers than non-smokers. This could be because of the chronic inhalation of nicotine. (Geracioti et al., 1999) Possible association of *COMT* with substance abuse has been suggested. (Uhl et al., 1998; Vandenbergh et al., 1997)

2   AIM OF THE STUDY

The goal is to identify associations in functional candidate genes contributing to nicotine dependence and smoking.

3   MATERIALS AND METHODS

3.1   Study sample and phenotype information

The participants to this study have been selected from the Finnish Twin cohort questionnaires in 1975, 1981, 1990 and 1996 to 1997. The index cases are twins from pairs that are concordant for heavy smoking. The siblings and parents of the twins have also been invited to participate in this study. After the families have been recruited, family members are interviewed using structured telephone interview and invited to give a blood sample. The telephone interview is a diagnostic interview with over 250 questions from nine different sections (e.g. alcohol use, tobacco use, mental disorders). Blood samples are drawn from smoking and non-smoking individuals. This project is a joint project between National Public Health Institute, Department of Molecular Medicine, which is responsible for the genetic data and University of Helsinki, Department of Public Health, which is responsible of the phenotypic data.

Table II describes the situation in the data collection in May 2005. Altogether 1816 persons have been interviewed and the blood samples have been also drawn. From this whole study sample, 720 individuals were chosen for this study.

Table II Data collection in May 2005

Data collection (May 2005)
1160 families invited
671 families recruited with 2131 family members (1175 men, 956 women)
Data collection complete for 1816 persons

Table III describes the familial structure of the study sample. Approximately half of the individuals are twins and half are siblings. The number of parents is low (only 24 parents are included) due to the high age of the recruited twins.

Table III The structure of the study sample

Candidate gene analysis sample	Number of persons
Twins	419
Non-twin sisters	136
Non-twin brothers	133
Others (half-sibs, parents, adult children)	32
Total number of individuals	720
Total number of families	240

Smoking history of the individuals has been carefully studied. Table IV shows the smoking history of the individuals chosen to this candidate gene study. Over 90% of the individuals have tried smoking and 53 % have become nicotine dependent. Thus, our study sample is enriched for smokers and nicotine dependent individuals.

Table IV Smoking history of the current study sample

Smoking history	Candidate gene analysis sample (n=720)
Never smoked, even not one puff	2.1 %
Smoked one whole cigarette ever	93 %
Regular smoked (at least weekly for 2 months)	90 %
Average level of smoking: > 15 cigarettes/ day	52 %
First cigarette within 30 min of waking up	56 %
Average age of onset	18 years
Fagerström (FTND) score (0 – 10)	3.7
Diagnosed as nicotine dependent based on DSM-IV criteria	53 %

3.2 Laboratory methods

3.2.1 Blood samples and DNA extraction

Trained health care professionals around Finland have drawn the blood samples into standard EDTA tubes. The tubes have been posted to National Public Health Institute (NPHI), Helsinki, and the DNA has been extracted at NPHI by the DNA extraction unit. For the extraction procedure Gentra System Incorporation's Autopure LS® automated DNA purification system was used. The purified total DNA was diluted to 500 or 1000 µl with 1x TE buffer (Gentra system Inc.). 10 µl of the total volume is used for concentration measurement of the sample. The measurement is performed with a flow through cuvette spectrophotometer. The extraction unit stores the DNA in -20 °C.

3.2.2 DNA dilutions

From the DNA extraction unit the DNA samples are brought to the DNA dilution unit. As presented in Figure 1 the first dilution is done from the original DNA sample tubes.
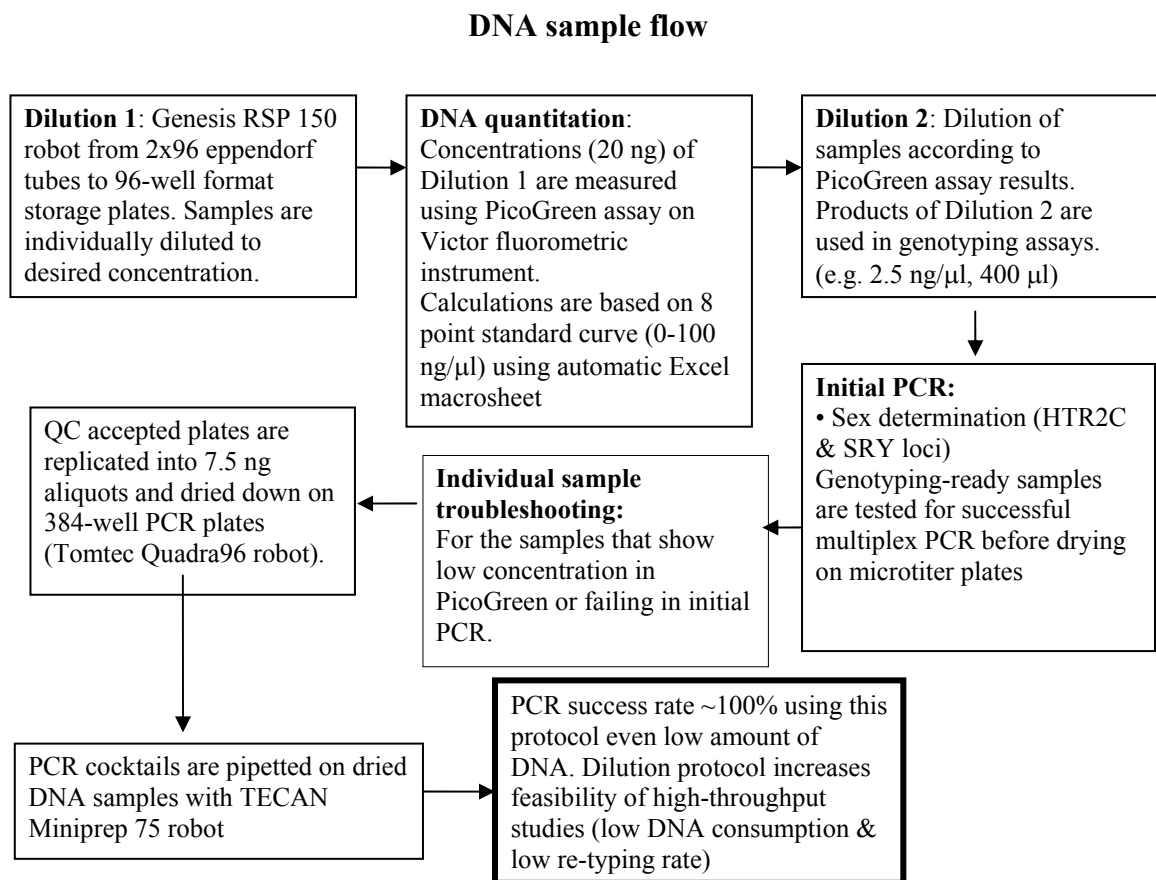
DNA sample flow

Figure 1 DNA sample flow (modified from P. Ellonen's flowchart, NPHI, 27.1.2003)

The concentration of the DNA is targeted to be 20ng/µl (50-500 µl) after the first dilution and this is confirmed with PicoGreen assay (Molecular Probes). If the concentration does not match with the desired one, it is corrected by adding more of the stock DNA. The second dilution is made from Dilution 1 and the concentration is 2.5 ng/µl (volume 50-500 µl). A sex-PCR is carried out at this point to verify the functionality of the DNA and to control the gender validity. See appendix 3 for PCR reactions, PCR program and primer sequences. All the dilutions are made in 96-well format.

After the quality controls, plates can be dried in a 96- or 384-well format with an automated robotic system. One 384-well plate consists of four different 96-well plates. 2-3 µl of the Dilution 2 is dried down to secure adequate DNA concentration in each well. In order to control for possible problems on the plates every 96-well Dilution 1 plate consists of 2-4 plate specific blind duplicates and 2 water controls. Thus, in a 384-well plate there are 8 water controls and 8-12 duplicates.

3.2.3   Candidate gene selection

The candidate genes to be studied were selected based on available literature in year 2002, before this thesis project was started. Four nicotinic acetylcholine receptor genes were selected: *CHRNA4, CHRNA5, CHRNA7* and *CHRNB1*. The alpha receptors are all expressed in the central nervous system whereas *CHRNB1* is expressed in the muscle tissue. Also, human cytochrome P450 genes *CYP1A2* and *CYP2A6*, which are expressed in the liver, were chosen. The full names and abbreviations with sizes and chromosomal positions are shown in Table V.

Table V Location information of the chosen candidate genes

Gene		Chromosomal position	Size
CHRNA4	Nicotinic acetylcholine receptor α4	20q13.2-q13.3	17 kb
CHRNA5	Nicotinic acetylcholine receptor α5	15q24	29 kb
CHRNA7	Nicotinic acetylcholine receptor α7	15q14	139 kb
CHRNB1	Nicotinic acetylcholine receptor β1	17p13.1	13 kb
CYP1A2	Human cytochrome P450	15q22-qter	8 kb
CYP2A6	Human cytochrome P450	19q13.2	7 kb

Intragenic single nucleotide polymorphisms (SNPs) were selected for genotyping to characterize the allelic diversity of the candidate genes. SNPs were selected based on availability in year 2002, the knowledge of sufficient information content and aiming to even distribution across the candidate genes.

3.2.4 Assay design

Assay design for the Sequenom MassARRAY™ system uses AssayDESIGN 2.0 software. A screen shot of the program is shown in Figure 2.
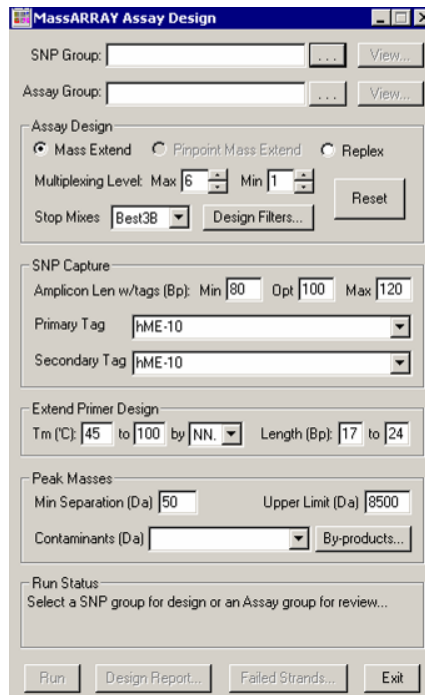

Figure 2 Assay DESIGN 2.0 software screen shot

The software is set to design six SNPs in a group, i.e. a multiplex of six SNPs. The actual groups are often smaller depending upon sequences. The software uses masses for each SNP assay product to look for possible multiplex combinations within the same termination mix group of ddNTPs. The termination point and the number of incorporated nucleotides are sequence specific. This mass difference makes it possible to produce allele specific products to be identified. The mass difference between all possible products in the same assay is set to 50 Daltons. Important SNPs are run singly, if necessary. The software designs forward and reverse primers and extension primers for each SNP in the multiplex group. (Sequenom Inc., 2004)

All the primer sequences proposed by the software are checked against the human genome sequence using BLAST (Basic Local Alignment Search Tool) in order to verify their uniqueness. The primers, forward, reverse and extension, were ordered from commercial manufacturers, Metabion or Proligo. 2 OD of PCR and extension primers are ordered.

3.2.5   PCR primer testing

The PCR primers are tested with a few DNA samples (e.g. three samples and one water control) in a specific primer test DNA 384-well plate. Working dilution for PCR primers is 10 μmol in primer testing and 100 μmol for validation and optimisation and extension primers. The dilutions are made in commercial water (Fluka). Each primer pair is verified to give a band of expected size and that the water controls (i.e. no DNA template) are clean. The PCR protocol is shown in appendix 4. All the samples are run on a 3-4 % SeaKem GTG agarose gel (Lonza) for 2 hours and the DNA band sizes are controlled.

3.2.6   SNP validation

On a single 384-well plate 24 SNPs can be validated. There is a specific validation plate for this purpose; in each row there is one blank well and 15 wells with individual DNA samples. The DNA dilution unit produces these plates. The PCR protocol is shown in appendix 4. Every second well in each column has a double volume (10 μl) of PCR mix and of these, 4 μl of the PCR product is run on a 3-4 % SeaKem GTG agarose gel for 2 hours and the DNA band sizes are controlled. After this control the plates are processed according to Sequenom MassARRAY™ protocol (chapter 3.3). The data is carefully

reviewed to ensure the assays are working properly. Quality control steps are described more precisely in section 3.4 Quality control and statistical analysis.

3.2.7 Multiplex redesigning and tests

Polymorphic and well working SNPs are combined to new multiplex assay groups according to the analysis in the validation step. This step is made with the AssayDesign 2.0 software's replex function. Now the SNPs are tested in multiplex assay groups by doing PCR on few samples (e.g. 3 samples and a water control) on the same DNA plate as in the primer tests. The PCR protocol is shown in appendix 4. All the samples are run on a 3-4 % SeaKem GTG agarose gel for 2 hours. The PCR products are monitored carefully for expected bands and even intensity of the bands. If the multiplex assay PCR seems weak, some modifications like enzyme amount or primer concentration can be done to improve the balance of the assay.

3.2.8 Optimisation

In optimisation the success rate of the multiplex assays are tested. Twelve multiplex assays can be tested on one 384-well optimisation plate, which consists of 31 unrelated DNAs, 15 of which are identical to the samples used in the validation. A blank well (i.e. water control) is used for contamination monitoring. The PCR protocol is shown in appendix 4. Every second well in every second column contains double volume (10 µl) of the PCR mix and 4 µl of the volume is run on an agarose gel as in multiplex testing step. The PCR products are verified and the 384-well plate is further processed. The profiles of each multiplex assay group are monitored for allele peak heights and successful extension reactions. Quality control checks for the optimisation are explained in the section 3.4 Quality control and statistical analysis.

3.2.9 Throughput testing and genotyping

Before proceeding with the actual sample material, one 384-well sample plate is genotyped to see that all assays work well. Only one multiplex is run on this stage. PCR primer concentration in this stage is 100 µM. The PCR protocol is shown in appendix 4. The PCR pipeting robot adds double volume (10 µl) of the PCR mix in 8 wells in 6 columns and the corresponding samples are run on a 3-4 % SeaKem GTG agarose gel for 2 hours and the

PCR products are carefully monitored. Again the quality control steps are done, see chapter 3.4.

3.3   Sequenom MassARRAY™ genotyping system

Sequenom MassARRAY™ system is used for genotyping. The flowchart of the protocol is shown in Table VI. The system consists of an automated 96-channel pipettor, a 24-pin nanoliter dispenser and a mass spectrometer. Mass spectrometry in this system is based on matrix-assisted laser desorption / ionisation time-of-flight mass spectrometry (MALDI-TOF MS) analysis of nucleic acids cleaved at specific bases. A homogenous MassEXTEND assay (hME) from Sequenom is used to analyse single nucleotide polymorphisms. The hME assay is based on the annealing of the extension primer to the polymorphic site of interest. The extension of the primer goes through the polymorphic site. The result is allele-specific extension products, which have a unique molecular mass. The masses are then analyzed by the MALDI-TOF MS and genotypes are assigned. (Sequenom Inc., 2003)

Table VI MassARRAY™ homogenous MassEXTEND™ Assay flowchart

1. Template Amplification
2. Dephosphorylation
3. hME reaction
4. Sample conditioning and transfer
5. Genotype calling and bioinformatics

3.3.1   Template amplification

Genomic DNA samples must be amplified with PCR prior to using the MassARRAY system. The amplification is processed in 384-well format and it will be assisted with robotics. The PCR primers have a 10-mer tag on the 5´ ends of each primer which significantly improves the performance of the PCR reaction. The tags also balance the amplification and increase the masses of unused PCR primers and this they will fall outside the detection window. (Sequenom Inc., 2003)

3.3.2 Dephosphorylation

Unincorporated dNTPs are removed by using shrimp alkaline phosphatase (SAP). This step prevents incorporation and interference of residual nucleotides with the primer extension assay. (Sequenom Inc., 2003)

3.3.3 hME reaction

In mass extension reaction, extension primer is added to the template. This extension primer is complementary to the template at the region directly or close to the SNP site and binds at this location. This reaction cocktail contains DNA polymerase, a mixture of three dideoxynucleotides and one deoxynucleotide, and an extension primer. DNA polymerase uses the available nucleotides to extend the primer through the polymorphic site. Extension continues until a single dideoxynucleotide is incorporated and the reaction is terminated. As the termination point and number of nucleotides are sequence specific, the size of the extension product generated can be used for identifying the allele variants. (Sequenom Inc., 2003)

3.3.4 Sample conditioning and transfer

The purpose of this step is to optimise mass spectrometry analysis by removing extraneous salts that interferes it. After the extension reaction is finished, 3 mg Clean Resin (Sequenom Inc.) is added to the reaction. The resin is added with a 384-format dimple plate and then 80 µl of nanopure water is added into each well. The plate is rotated with 360 degrees around its long axis for five minutes in room temperature. Then it is centrifuged for three minutes at 1600 RPM.

After desalting 15 nl of the samples are transferred from the 384-well plate and spotted onto the pad of the 394 SpectroCHIP® bioarray (Sequenom Inc.). It is a small silicon chip that incorporates a high-density array of mass spectrometry analysis sites. Each site is preloaded with light absorbing crystals that form a matrix with the dispensed sample. (Sequenom Inc., 2003)

3.3.5 MALDI-TOF mass spectrometry analysis

The MALDI-TOF mass spectrometry analysis contains an adsorption process, in which the matrix is hit with a pulse from a laser beam. The sample and the matrix are vaporized and the primer extension products are expelled into the flight tube. As primer extension

products are negatively charged, they are launched down the flight tube towards the detector. The time between application of the electrical field pulse and the collision of the primer extension products with the detector is referred as the "time-of-flight". This is a very precise measurement method for the products' molecular weight. Sequenom system's SpectroTYPER-RT software collects in real time the information on the time-of-flight measurements and applies algorithms to provide automated genotype calling. This information is then displayed in TyperAnalyzer software (Figure 3). (Sequenom Inc., 2003 and 2004)


Figure 3 Screenshot of TyperAnalyzer software

3.4 Quality control

The genotype results of each marker are accepted only if the success rate is at least 80% (preferably at least 90%), all duplicates were identical, water controls were clean, marker was in Hardy-Weinberg equilibrium and no Mendelian errors were observed.

3.4.1 KariOTyper

KariOTyper is a database established for semi-automated checks for various factors of peak and assay quality. (Finnish Genome Center, 2004 and 2005)

In the validation step SNPs are genotyped as singleplex and reviewed carefully. Following aspects are verified:

- Minor allele frequency > 5%
- Heterozygotes and homozygotes are identified among the genotyped individuals
- Genotype profiles are reliable and there are no extra alleles or noise peaks
- Heterozygote alleles are not variably skewed

In the multiplex optimisation step are genotyped in multiplexes and reviewed carefully. Following aspects are verified:

- Assays are successful and balanced for all SNPs
- Genotypes match among the individuals genotyped on both validation and optimisation plates
- Genotype profiles are reliable and there are no extra alleles or noise peaks
- Heterozygote alleles are not variably skewed

After genotyping the first sample plate (i.e., throughput plate testing) the following aspects are verified:

- If family samples are included, no inheritance inconsistencies are found
- Genotype profiles look good and there are no extra alleles or noise peaks
- Heterozygote alleles are not variably skewed
- Success rate is > 90%
- All duplicate genotypes are identical
- Water controls are clean
- Genotype frequencies are consistent with Hardy-Weinberg equilibrium

Each throughput plate is carefully reviewed and monitored for the same aspects as the throughput testing plate.

3.5 Hardy-Weinberg equilibrium

Hardy-Weinberg's equilibrium (HWE) gives a possibility to study if an evolutionary event has influenced the allele frequencies. Theoretically calculated, expected and seen genotype frequencies are compared to each other and a chi squared test is used to test whether they significantly differ from each other. The estimated genotype frequencies are calculated as follows: $p^2 + 2pq + q^2 = 1$, in which p represents the frequency of one allele, and q represents the frequency of the other allele. (Strachan and Read, 2004) Significant deviations from HWE may indicate genotyping errors.

3.6 Linkage disequilibrium

Linkage disequilibrium (LD) is a population based statistical association between alleles or an allele and the disease locus. It measures the co-segregation of alleles or an allele and the disease locus. It is based on the assumption that all individuals are related.

The alleles of two loci can form four haplotypes (e.g., SNP#1 A/C and SNP#2 A/G can form the following haplotypes: AA, AG, CA, and CG). Haplotype is a combination of alleles at linked loci (on one chromosome). If there is no LD the frequencies of the haplotypes are directly derived from allele frequencies. A D-value describes the difference between the probabilities of observing the two marker alleles on the same haplotype and observing them independently in the population. There is a more simple way describing LD, the D´-value. It varies between 0 (independent) and ±1 (full LD, alleles occur together). Another used parameter is r^2 which describes the correlation co-efficient between pairs of loci. The scale is the same as in D´, but in order to have $r^2 = 1$ the allele frequencies need to be exactly the same and this makes it very strict. (Strachan and Read, 2004; Zondervan and Cardon, 2004)

3.7 Statistical methods

3.7.1 LINKAGE format parameter files

Parameter files from the genotyping results were created using MAKEDATA 1.4 and MAKEPED 2.21 programs. MAKEDATA is used to create LINKAGE format locus files. Information about number of markers in pedigree file, autosomal or sex-linked data, frequency of disease allele, liability classes, maximum number of alleles in any locus and

possible map file are given to this program. MAKEPED program converts pedigree files from pre-makeped format to LINKAGE format. It also breaks marriage loops in the file and recodes the pedigree and individual identifiers. (Lathrop et al, 1984)

3.7.2   DOWNFREQ

DOWNFREQ 2.1 program was used to estimate allele frequencies from the data and to remove missing alleles from the pedigree file. This program calculates heterozygosity, allele frequencies for founders and all individuals and polymorphism information content. (Lathrop et al, 1984)

3.7.3   PEDCHECK

Pedcheck is a program for identifying marker typing incompatibilities in family data. It examines mendelian inheritance patterns and can also take X chromosome into account. Program was used for checking possible Mendelian errors. (O'Connell & Weeks, 1998)

3.7.4   PSEUDOMARKER

Pseudomarker program was selected to perform association analyses to look for LD between genetic markers and nicotine dependence in the study sample. Association analysis look for specific genetic variants that are shared among affected individuals. This type of analysis can have greater statistical power to detect several genes of small effect. Association is a statistical observation between alleles or haplotypes and the disease. This analysing method is based on the likelihood between linkage and / or linkage disequilibrium. Linkage and LD analysis are performed for each marker locus vs. pseudomarker locus. Pseudomarker uses a special version of ILINK from FASTLINK 4.1P software package. The program provides for each marker separate statistical evaluations of evidence for linkage, LD given linkage, LD given no linkage, linkage given LD and a joint test of linkage and LD. (Göring and Terwilliger, 2000a and b)

3.7.5   LDA

This program computes LD values of D prime and r-squared between two markers. First haplotypes are estimated from the data and based on that information maximum likelihood is used to estimate LD values. (Ding et al. 2003)

4   RESULTS

4.1   The study sample

Altogether 720 individuals were analysed in this project to identify potential candidate genes related to nicotine dependence and smoking. The study sample consists of 240 families mostly of twins and their siblings and only few parents. The individuals were chosen based on their smoking status and scores in the dependence scaling. During the genotyping procedures it was noticed that some of the twins, 24 pairs, were monozygotic. From the 24 monozygotic pairs the other twin was removed before the analyses were performed. Two twin pairs, expected to be monozygotic twins, were found to be discordant in their genotypes. The two discordant pairs were removed and this decreased the number of families by two and the number of individuals by four resulting 716 genotyped individuals and 238 families. After removal of the discordant pairs and other twin of the monozygotic twins, 705 individuals from 227 families were analysed. As seen in Table IV 53 % of the study sample is nicotine dependent measured with DSM-IV criteria and 90 % are regular smokers. These distributions make the study biased.

4.2   The candidate genes and the SNPs

The NCBI and dbSNP databases were used for the marker selection. The goal was to select SNPs that would have significant functional relevance, would cover the genes as widely as possible and would have been genotyped already by someone else. The SNP selection was also done, like the gene selection, in 2002 and the knowledge about certain SNPs was not adequate at that time. See appendix 2 for the figures of the SNP distribution in each of the chosen genes.

In the beginning of the project a total of 60 SNPs were selected. Altogether 28 SNPs passed all the quality controls to the final throughput genotyping. As the SNPs had not been previously genotyped at NPHI, they needed to be tested and optimized before large-scale genotyping. After the SNPs were selected, their sequence was taken from the SNPper database (Riva and Kohane, 2002) and modified for the AssayDESIGN 2.0 software. The modification of the sequence is shown in the Appendix 5. The proposed primer sequences from the AssayDESIGN 2.0 software were checked for the possible homologues in the

human genome by BLAST. The possible homologues could interfere with the interpretation of the results.

As mentioned in section 3 the primers were ordered according to the sequences from the AssayDESIGN 2.0 software. Each of the primer pairs (reverse and forward) were tested with PCR and controlled with an agarose gel electrophoresis as described in chapter 4.2.5. See appendix 4 for the PCR protocol. If any extra bands were seen in the sample wells or the water well, the primer pair was abandoned.

The multiplexes designed by the AssayDESIGN 2.0 software are shown in the Appendix 5. The coding of the multiplexes has been chosen during the designing of the multiplexes. Nine functional multiplexes were used. Some of the available, functional multiplexes were not used for the throughput genotyping, because the physical location of SNPs was so close to each other.

4.3   Allele frequencies and success rates

The minor allele frequencies and success rates for each of the genotyped SNPs (see Table VII) were calculated manually and compared to results given by the DOWNFREQ program. The average success rate of the markers was 90.5% (range 79.5% - 97.8%).

Table VII Minor allele frequencies and success rates for the genotyped SNPs

Gene	SNP	MAF (%)	Success rate (%)
CHRNA4	rs2236196	25,1	93,8
CHRNA4	rs3827020	20,0	95,4
CHRNA4	rs1044396	45,7	92,9
CHRNA4	rs1044397	43,2	89,5
CHRNA4	rs2273502	5,5	88,2
CHRNA4	rs2273505	5,0	91,1
CHRNA4	rs755203	36,0	79,5
CHRNA4	rs3810471	36,6	81,7
CHRNA5	rs667282	23,0	93,1
CHRNA5	rs680244	32,9	91,6
CHRNA5	rs621849	31,9	88,6
CHRNA5	rs1051730	32,1	82,7
CHRNA5	rs3743078	22,4	90,8
CHRNA7	rs3087454	14,0	91,5
CHRNA7	rs868437	13,1	80,7
CHRNA7	rs883473	15,1	89,8
CHRNA7	rs2337506	10,5	93,8
CHRNA7	rs904952	39,1	83,4
CHRNB1	rs3853818	44,4	94,8
CHRNB1	rs2302767	27,4	93,7
CHRNB1	rs2302762	26,5	90,9
CHRNB1	rs2302763	21,3	92,7
CHRNB1	rs2302764	16,1	89,2
CHRNB1	rs871990	21,6	93,7
CYP1A2	rs2069522	27,1	97,8
CYP1A2	rs4646427	4,3	93,8
CYP1A2	rs4886406	5,8	94,8
CYP2A6	rs2644914	22,2	93,1

4.4   Quality control

All the genotyped SNPs were checked for Hardy-Weinberg equilibrium and Mendelian errors. Using PEDCHECK program it was found out that there were 24 monozygotic twin pairs and two monozygotic twin pairs that were discordant, were removed. Only one individual from each of the monozygotic twin pairs was included in the analyses.

4.5 Linkage disequilibrium

The linkage disequilibrium of the genotyped SNPs in each of the candidate genes were determined by using LDA software. D' and r^2 values were calculated. One individual per family was included to the analysis in order to have a representative and non-skewed data set. Each of the genes is separately presented in Figures 4-13.

D'

	rs2236196	rs3827020	rs1044396	rs1044397	rs2273502	rs2273505	rs755203	rs3810471
rs3827020	1							
rs1044396	0,95	0,59						
rs1044397	0,95	0,65	1					
rs2273502	0,61	0,12	1	1				
rs2273505	0,6	0,01	1	1	1			
rs755203	0,88	0,53	0,85	0,85	1	1		
rs3810471	0,91	0,49	0,85	0,85	1	1	0,99	

Figure 4 D prime values of SNPs genotyped from *CHRNA4*

r^2

	rs2236196	rs3827020	rs1044396	rs1044397	rs2273502	rs2273505	rs755203	rs3810471
rs3827020	0,1							
rs1044396	0,37	0,1						
rs1044397	0,34	0,12	0,96					
rs2273502	0,07	0	0,06	0,06				
rs2273505	0,06	0	0,06	0,06	0,94			
rs755203	0,29	0,08	0,72	0,69	0,07	0,06		
rs3810471	0,31	0,07	0,72	0,69	0,07	0,06	0,99	

Figure 5 r-squared values of SNPs genotyped from *CHRNA4*

D'

	rs667282	rs680244	rs621849	rs1051730	rs3743078
rs680244	1				
rs621849	1	1			
rs1051730	1	0,98	0,99		
rs3743078	1	1	1	1	

Figure 6 D prime values of SNPs genotyped from *CHRNA5*

r^2

	rs667282	rs680244	rs621849	rs1051730	rs3743078
rs680244	0,21				
rs621849	0,21	1			
rs1051730	0,22	0,3	0,31		
rs3743078	0,99	0,23	0,22	0,21	

Figure 7 r-squared values of SNPs genotyped from *CHRNA5*

D'

	rs3087454			
rs868437	1	rs868437		
rs883473	1	1	rs883473	
rs2337506	1	0,98	0,99	rs2337506
rs904952	0,61	0,56	0,56	0,71	rs904952

Figure 8 D prime values of SNPs genotyped from *CHRNA7*

r²

	rs3087454			
rs868437	1	rs868437		
rs883473	0,96	0,95	rs883473	
rs2337506	0,73	0,73	0,66	rs2337506
rs904952	0,09	0,08	0,09	0,1	rs904952

Figure 9 r-squared values of SNPs genotyped from *CHRNA7*

D'

	rs3853818				
rs2302767	0,99	rs2302767			
rs2302762	0,99	1	rs2302762		
rs2302763	1	0,96	1	rs2302763	
rs2302764	1	1	0,99	0,89	rs2302764
rs871990	1	1	1	1	1	rs871990
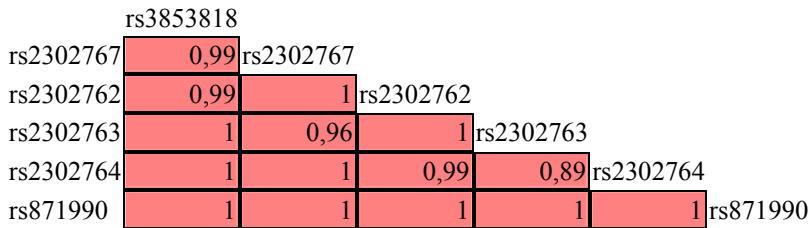
Figure 10 D prime values of SNPs genotyped from *CHRNB1*

r²

	rs3853818				
rs2302767	0,35	rs2302767			
rs2302762	0,33	0,96	rs2302762		
rs2302763	0,27	0,12	0,13	rs2302763	
rs2302764	0,2	0,55	0,56	0,06	rs2302764
rs871990	0,27	0,13	0,13	0,98	0,07	rs871990
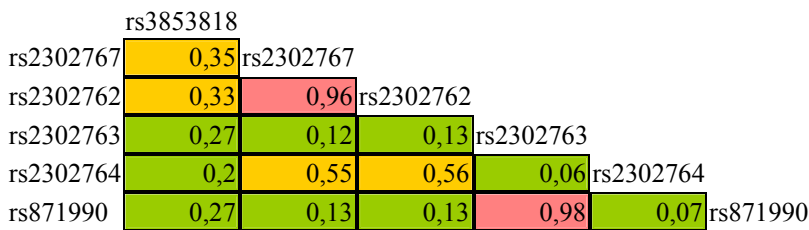
Figure 11 r-squared values of SNPs genotyped from *CHRNB1*

D'

	rs4886406	
rs4646427	1	rs4646427
rs2069522	1	1	rs2069522
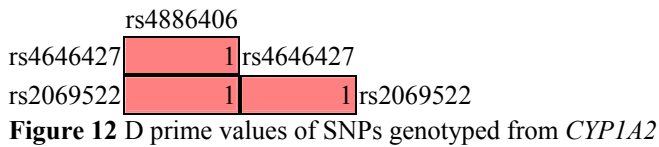
Figure 12 D prime values of SNPs genotyped from *CYP1A2*

r²

	rs4886406	
rs4646427	0,57	rs4646427
rs2069522	0,02	0,01	rs2069522
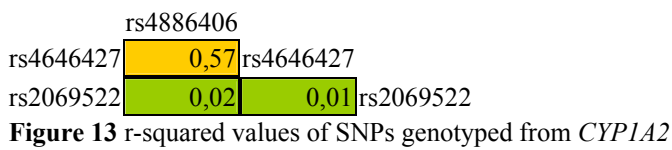
Figure 13 r-squared values of SNPs genotyped from *CYP1A2*

4.6   Association studies

Association analyses were performed by Pseudomarker software. The phenotypes regular smoker and nicotine dependence were used. Both of the phenotypes were dichotomised (discontinuous), i.e., individual either has the phenotype or not. In regular smoker all individuals that had smoked weekly for 2 at least months were considered as affected and all others were assigned as "unknown". Individuals with a scored more or equal to 4 points in the assessment with DSM-IV criteria were considered as nicotine dependent whereas individuals with scored 0-3 were considered as "unknown".

Association results for nicotine dependence with recessive (rec) and dominant (dom) models are shown in Table VIII. The results are presented as p-values and only the SNPs that gave a p-value less than 0.05 with some of the analysed models are presented in Table VIII. Similarly, association results for the regular smoker are shown in Table IX. All the original results are presented in appendix 6. Results for *CYP2A6* are not presented as there was only one analysed SNP and the genotypes of this SNP were considered as non-reliable.

Table VIII Association analysis results from the study sample with nicotine dependence phenotype using Pseudomarker software. Results are presented as p-values. p-values less than 0.05 are bolded.

SNP	Model	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
CHRNA4						
rs3827020	dom	**0.031**	0.57	0.75	0.054	0.10
	rec	**0.021**	0.50	0.51	**0.042**	**0.067**
rs2273502	dom	0.38	**0.027**	**0.028**	0.70	0.053
	rec	0.31	0.066	0.063	0.67	0.11
rs2273505	dom	0.37	**0.011**	**0.010**	0.89	**0.023**
	rec	0.33	**0.039**	**0.035**	1.0	0.071
CHRNA5						
rs667282	dom	0.50	**0.0028**	**0.0068**	0.20	**0.0071**
	rec	0.50	**0.039**	**0.046**	0.59	0.078
rs1051730	dom	0.13	0.13	0.15	0.21	0.11
	rec	0.17	**0.049**	**0.046**	0.38	0.061
rs3743078	dom	0.37	**0.0012**	**0.0048**	0.10	**0.0030**
	rec	0.42	**0.023**	**0.030**	0.49	**0.048**
CHRNB1						
rs2302763	dom	0.50	0.16	0.17	0.75	0.26
	rec	0.50	**0.026**	**0.030**	0.64	0.055
rs871990	dom	0.50	0.11	0.12	0.75	0.20
	rec	0.50	**0.012**	**0.014**	0.62	**0.028**
CYP1A2						
rs2069522	dom	**0.035**	0.43	0.89	0.048	0.094
	rec	**0.041**	0.78	1.0	0.078	0.15

Table IX Association analysis results from the study sample with regular smoker phenotype using Pseudomarker software. Results are presented as p-values. p-values less than 0.05 are bolded.

SNP	Model	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
CHRNA4						
rs2273502	dom	0.50	**0.0096**	**0.0096**	1.0	**0.022**
	rec	0.50	0.19	0.19	1.0	0.31
rs2273505	dom	0.50	**0.039**	**0.039**	1.0	0.079
	rec	0.50	0.29	0.34	0.65	0.44
CHRNA5						
rs3743078	dom	0.20	**0.0039**	0.066	**0.018**	**0.0068**
	rec	0.38	0.19	0.65	0.21	0.29
CHRNA7						
rs2337506	dom	0.39	0.33	0.33	0.82	0.46
	rec	0.50	**0.032**	**0.036**	0.65	**0.066**
rs2302767	dom	0.50	0.14	0.15	0.76	0.24
	rec	0.50	**0.0097**	**0.010**	0.75	**0.022**
rs2302762	dom	0.50	0.15	0.16	0.73	0.25
	rec	0.50	**0.010**	**0.011**	0.75	**0.024**
rs2302764	dom	0.50	0.071	0.10	0.43	0.13
	rec	0.50	**0.032**	0.10	0.17	**0.066**
CYP1A2						
rs2069522	dom	0.24	0.81	0.92	0.45	0.60
	rec	0.38	**0.034**	**0.043**	0.48	
rs4886406	dom	0.056	0.080	1.0	**0.018**	**0.039**
	rec	0.057	0.44	0.32	0.15	0.15

5 DISCUSSION

Nicotine is the main addictive substance responsible for maintaining and regulating use of tobacco. Both dopamine and non-dopamine neurochemical pathways through functionally diverse nicotinic acetylcholine receptors are influenced by nicotine and its psychoactive and addictive effects. (Picciotto et al., 2000) It has been suggested that nicotinic receptor genes such as *CHRNA4* and *CHRNB1* play a role in the genetics of nicotine dependence. (Feng et al., 2004 and Lou et al., 2007)

In this study 720 individuals were genotyped using 26 SNPs from six selected candidate genes. Phenotypes for the analyses were regular smoker and nicotine dependence. They were chosen as the starting point for this study from several available ones obtained from a wide collection of data from telephone interview based study.

Significant association for nicotine dependence phenotype was seen in *CHRNA5*: rs667282 (p< 0.003) and rs3743078 (p< 0.002) with a dominant model (LD given linkage). It has been reported that *CHRNA5* is a part of a gene cluster of *CHRNA5 – CHRNA3 - CHRNB4* in chromosome 15. (Berrettini et al., 2008) *CHRNA5* rs680244 has shown association with nicotine dependence on study sample of Israeli women (Greenbaum et al., 2006) and rs1051730 on an American study sample (Saccone et al., 2007). In this study sample rs680244 is included and is in strong LD with rs667282 and rs3743078.

Association with nicotine dependence was seen in *CHRNA4* rs2273502 (p=0.027) and rs2273505 (p=0.010), located only 9.7 kb from the previously reported association to nicotine dependence. (Feng et al., 2004 and Li et al., 2005) Also *CHRNA4* rs3827020 gave indication of association to nicotine dependence. SNPs rs2273502 and rs2273505 are in strong LD with each other and also with the other previously reported associating SNPs, rs1044397 and rs1044396. (Feng et al. 2004) This indicates that this study should be continued and expanded to haplotype/haploblock studies, which would give more information of the contribution of these SNPs to nicotine dependence. (Clark, 2004)

Association with nicotine dependence was also seen in *CHRNB1* rs2302763 (p=0.026) with the recessive model, confirming the previously reported association with dominant

and recessive models to nicotine dependence measured with the FTND scale. (Lou et al., 2007) Also *CHRNB1* rs871990 showed a p-value of 0.012.

With the regular smoker phenotype all of the genes showed some association. Strongest associations were seen in *CHRNA5* (rs3743078, p=0.0039) and *CHRNB1* (rs2302767, p=0.0097; rs2302762, p=0.010). The same SNPs that gave indication of association with nicotine dependence, also showed association with the regular smoker phenotype. The regular smoker phenotype has been less published than the nicotine dependence phenotype. The stratified nature of the phenotype and also the wide interpretation of the phenotype – smoked regularly at least weekly for two months – do not give very wide perspective to the phenotype. Therefore a continuous scale and quantitative trait analysis would give more detailed information.

Both of *CYP1A2* and *CYP2A6* genes had very problematic SNPs. Mostly the chosen SNPs turned out be monomorphic during validation or the allele frequencies on the test plates showed them to be very rare. As these genes, and also other genes from the CYP-gene family, have very important role in the metabolism of nicotine, more optimisation should be done with the chosen SNPs. Also new database search should be done to identify SNPs that have previously been successfully genotyped.

Some of the genotyped SNPs had very low minor allele frequencies, in some cases less than 5%. As Finnish population has been quite isolated, the chance of detecting rare disease alleles may be increased. However, there is a risk that these rare variants are hard to replicate in other populations. (Risch, 2000)

This candidate gene study gave strong evidence that nicotine dependence can be studied using candidate gene approach. It also gives real data about the genetics of the addiction and the associated genetic variants in this study sample. Also multi-variant analyses with more phenotypes should be included.

Interestingly while this project has been under its way, a genome wide scan results including partly overlapping study sample from Finland was published. In that study

chromosome 22q12 gave linkage with heavy smoking quantitative trait (maximum number of cigarettes smoked during 24 hours). A LOD score of 4.22 in chromosome 20 and a LOD 3.03 in chromosome 22 was reported. (Saccone et al., 2007) *CHRNA4* gene is located in chromosome 20; however, in the current study *CHRNA4* showed only weak association. Perhaps this gene should be more carefully analysed with more markers. The signal from the chromosome 22 came from an intron in *ADRBK2*. This gene codes for beta-adrenergic receptor kinase 2, which is involved in desensitisation of several G-protein receptor systems, such as dopamine factors. As nicotine influences the neurochemical pathway of dopamine, *ADRBK2* is an interesting gene to be studied. (Carman et al. 1998)

Isolated populations, such as the Finnish study sample, offer an advantage for genetic studies. The number of different variations in the genes behind phenotypes are expected to be smaller than in more heterogeneous populations. Thus, the probability to find genetic associations increases. (Peltonen, 2000) Therefore even a small study sample from an isolated population, like in this study, can give very realistic results.

Also the used study sample is strongly enriched with regular smokers and nicotine dependent individuals giving more statistical power. The careful and extensive telephone interview based phenotypic data collection has been performed by highly trained interviewers, yielding exceptionally reliable phenotype data.

The HapMap data was not available at the time when the SNPs were selected. The distribution of SNPs throughout the selected genes thus is not optimal. The tagSNP data was checked on May 2008 and only five of the used SNPs were tagSNPs: *CHRNA4* rs1044396, *CHRNA5* rs621849, *CHRNB1* rs2302767 and rs2302764 and *CYP1A2* rs4646427. (HapMap.org, 2008)

In the current study, several quality control measures were used, such as, HWE, check for skewed allele callings, manual checking of the alleles and PCR products run on agarose gels. However, possibly remaining genotyping errors may have influenced the results, either by yielding false positive of false negative results.

As a conclusion, this study showed very interesting and promising results of the candidate genes involved in genetics of nicotine dependence. In the future the project should be continued with new candidate genes and a denser set of SNPs, possibly with genome-wide association and including more specific case-control study with a wider set of phenotypes.

REFERENCES

Affymetrix, Incorporation. 2007. Application note: A new paradigm for two-stage association studies. Affymetrix, Incorporation, Santa Clara, CA, U.S.A.

American Pshychiatric Association. 1994. Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). American Psychiatric Press Inc, Washington, D.C., USA.

Ando, M., N. Hamajima, N. Ariyoshi,T. Kamataki, K. Matsuo, Y. Ohno. 2003. Association of CYP2A6 gene deletion with cigarette smoking status in Japanese adults. *J Epidemiol*. 13:176-81.

Batra, V., A.A. Patkar, W.H. Berrettini, S.P. Weinstein, and F.T. Leone. 2003. The genetic determinants of smoking. *Chest*. 123:1730-9.

Bergen, A.W., J.F. Korczak, K.A. Weissbecker, and A.M. Goldstein. 1999. A genome-wide search for loci contributing to smoking and alcoholism. *Genet Epidemiol*. 17 Suppl 1:S55-60.

Berrettini W., X. Yuan, F. Tozzi, K. Song, C. Francks, H. Chilcoat, D. Waterworth, P.Muglia, V. Mooser. 2008. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase a risk on heavy smoking. *Mol Psychiatry*. 13: 368-373.

Bierut, L.J., J.P. Rice, A. Goate, A.L. Hinrichs, N.L. Saccone, T. Foroud, H.J. Edenberg, C.R. Cloninger, H. Begleiter, P.M. Conneally, R.R. Crowe, V. Hesselbrock, T.K. Li, J.I. Nurnberger, Jr., B. Porjesz, M.A. Schuckit, and T. Reich. 2004. A genomic scan for habitual smoking in families of alcoholics: common and specific genetic factors in substance dependence. *Am J Med Genet A*. 124:19-27.

Bierut, L.J., P.A.F. Madden, N. Breslau, E.O. Johnson, D. Hatsukami, O.F. Pomerleau, G.E. Swan, J. Rutter, S. Bertelsen, L.Fox, D. Fugman, A.M. Goate, A.L. Hinrichs, K.Konvicka, N.G. Martin, G.W. Montgomery, N.L. Saccone, S.F. Saccone, J.C. Wang, G.A. Chase, J.P. Rice, D.G. Ballinger. 2007. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 16: 24-35.

Buisson, B., and D. Bertrand. 2001. Chronic exposure to nicotine upregulates the human (alpha)4((beta)2 nicotinic acetylcholine receptor function. *J Neurosci*. 21:1819-29.

Carman, C.V., J.L Benovic. 1998. G-protein-coupled receptors: turn-ons and turn-offs. *Curr Opin Neurobiol*. 8:335-44.

Clark, A.G. 2004. The role of haplotypes in candidate gene studies. *Genet Epidemiol*. 27:321-33.

Dani, J.A., and M. De Biasi. 2001. Cellular mechanisms of nicotine addiction. *Pharmacol Biochem Behav*. 70:439-46.

Ding, K., K. Zhou, F. He, and Y. Shen. 2003. LDA - A java-based linkage disequilibrium analyzer. *Bioinformatics*. 19(16): 2147-2148.

Duggirala, R., L. Almasy, and J. Blangero. 1999. Smoking behavior is under the influence of a major quantitative trait locus on human chromosome 5q. *Genet Epidemiol*. 17 Suppl 1:S139-44.

Fagerstrom, K.O. 1978. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict Behav*. 3:235-41.

Feng, Y., T. Niu, H. Xing, X. Xu, C. Chen, S. Peng, L. Wang, and N. Laird. 2004. A common haplotype of the nicotine acetylcholine receptor alpha 4 subunit gene is associated with vulnerability to nicotine addiction in men. *Am J Hum Genet*. 75:112-21.

Fowler, J.S., N.D. Volkow, G.J. Wang, N. Pappas, J. Logan, R. MacGregor, D. Alexoff, C. Shea, D. Schlyer, A.P. Wolf, D. Warner, I. Zezulkova, and R. Cilento. 1996a. Inhibition of monoamine oxidase B in the brains of smokers. *Nature*. 379:733-6.

Fowler, J.S., N.D. Volkow, G.J. Wang, N. Pappas, J. Logan, R. MacGregor, D. Alexoff, A.P. Wolf, D. Warner, R. Cilento, and I. Zezulkova. 1998. Neuropharmacological actions of cigarette smoke: brain monoamine oxidase B (MAO B) inhibition. *J Addict Dis*. 17:23-34.

Fowler, J.S., N.D. Volkow, G.J. Wang, N. Pappas, J. Logan, C. Shea, D. Alexoff, R.R. MacGregor, D.J. Schlyer, I. Zezulkova, and A.P. Wolf. 1996b. Brain monoamine oxidase A inhibition in cigarette smokers. *Proc Natl Acad Sci U S A*. 93:14065-9.

Geracioti, T.D., Jr., S.A. West, D.G. Baker, K.K. Hill, N.N. Ekhator, M.D. Wortman, P.E. Keck, Jr., and A.B. Norman. 1999. Low CSF concentration of a dopamine metabolite in tobacco smokers. *Am J Psychiatry*. 156:130-2.

Gibbs J.R. and A. Singleton. 2006. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet*. 2:e150.

Greenbaum L., K. Kanyas, O. Karni, Y. Merbl, T. Olender, A. Horowitz, A. Yakir, D. Lancet, E. Ben-Asher, B. Lerer. 2006. Why do young women smoke? I. Direct and interactive effects of environment, psychological characteristics and nicotinic cholinergic receptor genes. *Mol Psychiatry*. 11: 312-322, 223.

Göring H.H. and Terwilliger, J.D. 2000a. Linkage Analysis in the Presence of Errors III: Marker Loci and Their Map as Nuisance Parameters *Am J Hum. Genet*. 66(4):1298-1309.

Göring H.H. and Terwilliger, J.D. 2000b. Linkage Analysis in the Presence of Errors IV: Joint Pseudomarker Analysis of Linkage and/or Linkage Disequilibrium on a Mixture of Pedigrees and Singletons When the Mode of Inheritance Cannot Be Accurately Specified. *Am J Hum. Genet*. 66(4):1310-1327.

HapMap Consortium Inc. 2005. A haplotype map of the human genome. *Nature*. 437: 1299-1320.

HapMap Consortium Inc. 2007. A second generation human haplotype map over 3.1 million SNPs. *Nature*. 449: 851-862.

Heatherton, T.F., L.T. Kozlowski, R.C. Frecker, and K.O. Fagerstrom. 1991. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict*. 86:1119-27.

Kendler, K.S., M.C. Neale, P. Sullivan, L.A. Corey, C.O. Gardner, and C.A. Prescott. 1999. A population-based twin study in women of smoking initiation and nicotine dependence. *Psychol Med*. 29:299-308.

Lai, I.C., C.J. Hong, S.J. Tsai. 2001. Association study of nicotinic-receptor variants and major depressive disorder. *J Affect Disord*. 66:79-82.

Lathrop G.M., J.M. Lalouel, C. Julier and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. U.S.A. 81(11):3443-3446.

Laviolette, S.R., and D. van der Kooy. 2004. The neurobiology of nicotine addiction: bridging the gap from molecules to behaviour. *Nat Rev Neurosci*. 5:55-65.

Leonard, S., and D. Bertrand. 2001. Neuronal nicotinic receptors: from structure to function. *Nicotine Tob Res*. 3:203-23.

Lerman, C., N.E. Caporaso, J. Audrain, D. Main, E.D. Bowman, B. Lockshin, N.R. Boyd, and P.G. Shields. 1999. Evidence suggesting the role of specific genetic factors in cigarette smoking. *Health Psychol*. 18:14-20.

Lessov, C.N., N.G. Martin, D.J. Statham, A.A. Todorov, W.S. Slutske, K.K. Bucholz, A.C. Heath, and P.A. Madden. 2004. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med*. 34:865-79.

Li, M.D., J.Z. Ma, R. Cheng, R.T. Dupont, N.J. Williams, K.M. Crews, T.J. Payne, and R.C. Elston. 2003. A genome-wide scan to identify loci for smoking rate in the Framingham Heart Study population. *BMC Genet*. 4 Suppl 1:S103.

Li, M.D., J. Beuten, J.Z. Ma, T.J. Payne, X.Y. Lou, V. Garcia, A.S. Duenes, K.M. Crews, R.C. Elston. 2005. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. *Hum Mol Genet*. 14:1211-9.

Liang, H.C., H. Li, R.A. McKinnon, J.J Duffy, S.S. Potter, A. Puga, D.W. Nebert. 1996. Cyp1a2(-/-) null mutant mice develop normally but show deficient drug metabolism. *Proc Natl Acad Sci U.S.A*. 93:1671-6.

London, S.J., J.R. Idle, A. K. Daly, G.A. Coetzee. 1999. Genetic variation of CYP2A6, smoking, and risk of cancer. *Lancet*. 353:898-9.

Lou, X.Y., J.Z. Ma, T.J. Payne, J. Beuten, K.M. Crew, M.D. Li. 2006. Gene-based analysis suggests association of the nicotinic acetylcholine receptor beta1 subunit (CHRNB1) and M1 muscarinic acetylcholine receptor (CHRM1) with vulnerability for nicotine dependence. *Hum Genet*. 120:381-9.

Loukola, A., U. Broms, H. Maunu, E. Widén, K Heikkilä, M. Siivola, A. Salo, M.L. Pergadia, E. Nyman, S. Sammalisto, M. Perola, A. Agrawal, A.C. Heath, N.G. Martin, P. A. Madden, L. Peltonen, J. Kaprio. 2007. *Pharmacogenomics J*. published online 5 June 2007.

Lundstrom, K., J. Tenhunen, C. Tilgmann, T. Karhunen, P. Panula, and I. Ulmanen. 1995. Cloning, expression and structure of catechol-O-methyltransferase. *Biochim Biophys Acta*. 1251:1-10.

Mackay, and Eriksen. 2002. The Tobacco Atlas. World Health Organisation.

Maes, H.H., P.F. Sullivan, C.M. Bulik, M.C. Neale, C.A. Prescott, L.J. Eaves, and K.S. Kendler. 2004. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med*. 34:1251-61.

Nakamura T., S. Matsushita, N. Nishiguchi, M. Kimura, A. Yoshino, S. Higuchi. 1999. Association of a polymorphism of the 5HT2A receptor gene promoter region with alcohol dependence. *Mol Psychiatry*. 4: 85-88.

Niu, T., C. Chen, J. Ni, B. Wang, Z. Fang, H. Shao, and X. Xu. 2000. Nicotine dependence and its familial aggregation in Chinese. *Int J Epidemiol*. 29:248-52.

O'Connell J.R., Weeks D.E. 1998. PedCheck: A program for identifying genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* 63(1): 259-266

Patja K., and Haukkala A. 2004. Tupakkakertomus 2003. National Public Health Institute, Helsinki.

Peltonen L., 2000. Positional Cloning of Disease Genes: Advantages of Genetic Isolates. *Hum Hered*. 50:66-75.

Pianezza, M.L., E.M. Sellers, R.F. Tyndale. 1998. Nicotine metabolism defect reduces smoking. *Nature*. 393:750.

Picciotto, M.R., B.J. Caldarone, S.L King, V. Zachariou. 2000. Nicotinic receptors in the brain. Links between molecular biology and behavior. *Neuropsychopharmacology*. 22:451-65.

Risch, N.J., 2000. Searching for genetic determinants in the new millennium. *Nature*. 405:847-56.

Riva, A., Kohane, I. S. 2002. SNPper: retrieval and analysis of human SNPs. *Bioinformatics*. 18 (12): 1681-1685.

Role, L.W., and D.K. Berg. 1996. Nicotinic receptors in the development and modulation of CNS synapses. *Neuron*. 16:1077-85.

Rustin T.A. 2001. Accessing nicotine dependence. *Am. Fam. Phycisian.*62:579-84.

Sabol, S.Z., M.L. Nelson, C. Fisher, L. Gunzerath, C.L. Brody, S. Hu, L.A. Sirota, S.E. Marcus, B.D. Greenberg, F.R.T. Lucas, J. Benjamin, D.L. Murphy, and D.H. Hamer. 1999. A genetic association for cigarette smoking behavior. *Health Psychol*. 18:7-13.

Sabol, S.Z. and D.H. Hamer. 1999. An improved assays shows no association between CYP2A6 gene and cigarette smoking behavior. *Behav. Genet.* 4: 257-261.

Saccone, S. F., M. L. Pergadia, A. Loukola, U. Broms, G. W. Montgomery, J. C. Wang, A. Agrawal, D. M. Dick, A. C. Heath, A. A. Todorov, H. Maunu, K. Heikkila, K. I. Morley, J. P. Rice, R. D. Todd, J. Kaprio, L. Peltonen, N. G. Martin, A. M. Goate, P. A. F. Madden. 2007. Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *Am. J. Hum. Genet.* 80: 856-866.

Saccone S.F., A.L. Hinrichs, N.L. Saccone, G.A. Chase, K. Konvicka, P.A.F Madden, N. Breslau, E.O. Johnson, D. Hatsukami, O. Pomerleau, G.E. Swan, A.M. Goate, J. Rutter, S. Bertelsen, L.Fox, D. Fugman, N.G. Martin, G.W. Montgomery, J.C. Wang, D.G. Ballinger, J.P. Rice, L.J. Bierut. 2007. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 16:36-49.

Sáiz, P.A., M.P. Garciá-Portilla, C. Arango, B. Morales, S. Martínez-Barrondo, C. Alvarez, G. San Narciso, E. Carreno, V. Alvarez, E. Coto, J. Bobes. 2008. Association between heroin dependence and 5-HT2A receptor gene. *Eur Addict Res*. 14: 47-52.

Salaspuro M, K. Kiianmaa, and K. Seppä. 2003. Päihdelääketiede. Kustannus Oy Duodecim, Helsinki.

Sequenom Inc., 2004. MassARRAY Assay Design Software User's Guide. Version 2.0.7 Revision 3

Sequenom Inc., 2003. MassARRAY Typer User's Guide Version 3.1 Revision 3

Shields, P.G., C. Lerman, J. Audrain, E.D. Bowman, D. Main, N.R. Boyd, N.E. Caporaso. 1998. Dopamine D4 receptors and the risk of cigarette smoking in African-Americans and Caucasians. *Cancer Epidemiol Biomarkers Prev*. 7:453-8.

Shiffman, S., A. Waters, and M. Hickcox. 2004. The nicotine dependence syndrome scale: a multidimensional measure of nicotine dependence. *Nicotine Tob Res*. 6:327-48.

Schoedel, K.A., E.B. Hoffmann, Y. Rao, E.M. Sellers, R.F. Tyndale. 2004. Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians. *Pharmacogenetics*. 14:615-26.

Silverman, M.A., M.C. Neale, P.F. Sullivan, C. Harris-Kerr, B. Wormley, H. Sadek, Y. Ma, K.S. Kendler, R.E. Straub. 2000. Haplotypes of four novel single nucleotide polymorphisms in the nicotinic acetylcholine receptor beta2-subunit (CHRNB2) gene show no association with smoking initiation or nicotine dependence. *Am J Med Genet*. 96:646-53.

Slotkin, T.A. 1998. Fetal nicotine or cocaine exposure: which one is worse? *J Pharmacol Exp Ther*. 285:931-45.

Spitz, M.R., H. Shi, F. Yang, K.S. Hudmon, H. Jiang, R.M. Chamberlain, C.I. Amos, Y. Wan, P. Cinciripini, W.K. Hong, and X. Wu. 1998. Case-control study of the D2 dopamine receptor gene and smoking status in lung cancer patients. *J Natl Cancer Inst*. 90:358-63.

Strachan, T. and Read A.P. 2004. Human molecular genetics. 3rd edition. Garland Science, London U.K. and New York U.S.A.

Straub, R.E., P.F. Sullivan, Y. Ma, M.V. Myakishev, C. Harris-Kerr, B. Wormley, B. Kadambi, H. Sadek, M.A. Silverman, B.T. Webb, M.C. Neale, C.M. Bulik, P.R. Joyce, and K.S. Kendler. 1999. Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample

suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Mol Psychiatry*. 4:129-44.

Tabor, H.K., N.J. Risch, R.M. Myers. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet.* 3:391-7.

Tate, J.C., and J.M. Schmitz. 1993. A proposed revision of the Fagerstrom Tolerance Questionnaire. *Addict Behav*. 18:135-43.

True, W.R., A.C. Heath, J.F. Scherrer, B. Waterman, J. Goldberg, N. Lin, S.A. Eisen, M.J. Lyons, and M.T. Tsuang. 1997. Genetic and environmental contributions to smoking. *Addiction*. 92:1277-87.

Uhl, G.R., D.J. Vandenbergh, L.A. Rodriguez, L. Miner, N. Takahashi. 1998. Dopaminergic genes and substance abuse. *Adv Pharmacol*. 42:1024-32.

Uhl G.R., Q.R. Liu, T Drgon, C. Johnson, D. Walther, J.E. Rose. 2007. Molecular genetics of nicotine dependence and abstinence: whole genome assiciation using 520,000 SNPs. *BMC Genetics*. 8:10.

World Health Organization. 1992. International Classification of Diseases, Clinical Descriptions and Diagnostic Guidelines (ICD-10). World Health Organization, Geneva, Switzerland.

Vandenbergh, D.J., L.A. Rodriguez, I.T. Miller, G.R. Uhl, H.M. Lachman. 1997. High-activity catechol-O-methyltransferase allele is more prevalent in polysubstance abusers. *Am J Med Genet*. 74:439-42.

Vink, J.M., A.L. Beem, D. Posthuma, M.C. Neale, G. Willemsen, K.S. Kendler, P.E. Slagboom, and D.I. Boomsma. 2004. Linkage analysis of smoking initiation and quantity in Dutch sibling pairs. *Pharmacogenomics J*. 4:274-82.

Wonnacott, S. 1997. Presynaptic nicotinic ACh receptors. *Trends Neurosci*. 20:92-8.

Wonnacott, S., S. Kaiser, A. Mogg, L. Soliakov, and I.W. Jones. 2000. Presynaptic nicotinic receptors modulating dopamine release in the rat striatum. *Eur J Pharmacol*. 393:51-8.

Zhang, X., K. Amemo, S. Ameno, K. Iwahashi, H. Kinoshita, T. Kubota, J. Mostofa, I. Ijiri. 2001. Lack of association between smoking and CYP2A6 gene polymorphisms in A Japanese population. *Nihon Arukoru Yakubutsu Igakkai Zasshi*. 36:486-90. Abstract only available.

Zondervan K.T. and L.R. Cardon. 2004. The complex interplay among factors that influence allelic association.*Nat Rev Genet*. 5:89-100.

Unpublished sources

National centre for biotechnology information (NCBI)
www.ncbi.nlm.nih.gov

Basic Local Alignment Search Tool (BLAST)
www.ncbi.nlm.nih.gov/blast

Single nucleotide polymorphism database (dbSNP)
www.ncbi.nlm.nih.gov/projects/SNP

International HapMap Project
www.hapmap.org
HapMap Genome Browser B53, 17.5.2008

APPENDICES

APPENDIX 1a DSM-IV CRITERIA

Examples of used questions when diagnosing nicotine dependence according to DSM-IV criteria directly quoted from Rustin T.A. (2001).

Compulsion--the intensity with which the desire to use a chemical overwhelms the patient's thoughts, feelings and judgment

Do you ever smoke more than you intend?

Have you ever neglected a responsibility because you were smoking, or so you could smoke?

Control--the degree to which patients can (or cannot) control their chemical use once they have started using

Have you felt the need to control how much you smoke but were unable to do so easily?

Have you ever promised that you would quit smoking and bought a pack of cigarettes that same day?

Cutting down -- withdrawal symptoms

Have you ever tried to stop smoking? How many times? For how long?

Have you ever had any of the following symptoms when you went for a while without a cigarette: agitation, difficulty concentrating, irritability, mood swings? If so, did the symptom go away after you smoked a cigarette?

Consequences--denial or acceptance of the damage caused by the chemical

How long have you known that smoking was hurting your body? If you continue to smoke, how long do you expect to live? If you were able to quit smoking today and never start again, how long do you think you might live?

APPENDIX 1b ICD-10 CLASSIFICATION

Nicotine dependence diagnostic criteria according to ICD-10 classification (WHO, 1992).

F17.2 Tobacco Dependence Syndrome

A cluster of physiological, behavioral, and cognitive phenomena in which the use of tobacco takes on a much higher priority for a given individual than other behaviors that once had greater value. A central descriptive characteristic of the dependence syndrome is the desire (often strong, sometimes overpowering) to take tobacco. There may be evidence that return to substance use after a period of abstinence leads to a more rapid reappearance of other features of the syndrome than occurs with nondependent individuals.

Diagnostic Guidelines

A definite diagnosis of dependence should usually be made only if three or more of the following have been experienced or exhibited at some time during the previous year:

(a) A strong desire or sense of compulsion to take tobacco;

(b) Difficulties in controlling tobacco-taking behavior in terms of its onset, termination, or levels of use;

(c) A physiological withdrawal state when tobacco use has ceased or been reduced, as evidenced by: the characteristic withdrawal syndrome for tobacco; or use of the same (or a closely related) substance with the intention of relieving or avoiding withdrawal symptoms;

(d) Evidence of tolerance, such that increased doses of tobacco are required in order to achieve effects originally produced by lower doses;

(e) Progressive neglect of alternative pleasures or interests because of tobacco use, increased amount of time necessary to obtain or take the substance or to recover from its effects;

(f) Persisting with tobacco use despite clear evidence of overtly harmful consequences, such as depressive mood states consequent to periods of heavy substance use, or drug-related impairment of cognitive functioning; efforts should be made to determine that the user was actually, or could be expected to be, aware of the nature and extent of the harm. Narrowing of the personal repertoire of patterns of tobacco use has also been described as a characteristic feature.

It is an essential characteristic of the dependence syndrome that either tobacco taking or a

desire to take tobacco should be present; the subjective awareness of compulsion to use drugs is most commonly seen during attempts to stop or control substance use.

APPENDIX 1c FAGERSTRÖM TEST

Modified Fagerström Test for Nicotine Dependence (Heatherton et al., 1991)

1. How soon after you wake up do you smoke your first cigarette?

Within 5 minutes (3 points)

5 to 30 minutes (2 points)

31 to 60 minutes (1 point)

After 60 minutes (0 points)

2. Do you find it difficult not to smoke in places where you shouldn't, such as in church or school, in a movie, at the library, on a bus, in court or in a hospital?

Yes (1 point)

No (0 points)

3. Which cigarette would you most hate to give up; which cigarette do you treasure the most?

The first one in the morning (1 point)

Any other one (0 points)

4. How many cigarettes do you smoke each day?

10 or fewer (0 points)

11 to 20 (1 point)

21 to 30 (2 points)

31 or more (3 points)

5. Do you smoke more during the first few hours after waking up than during the rest of the day?

Yes (1 point)

No (0 points)

6. Do you still smoke if you are so sick that you are in bed most of the day or if you have a cold or the flu and have trouble breathing?

Yes (1 point)

No (0 points)

Scoring: 7 to 10 points = highly dependent; 4 to 6 points = moderately dependent; less than 4 points = minimally dependent

APPENDIX 2 FIGURES OF THE SNP DISTRIBUTIONS IN THE CANDIDATE GENES

CHRNA4 20q13.2-q13.3

SNP 8 rs3810471 MAF 0,37
SNP 7 rs755203 MAF 0,36
SNP 6 rs2273505 MAF 0,05
SNP 5 rs2273502 MAF 0,06
SNP 4 rs1044397 MAF 0,43
SNP 3 rs1044396 MAF 0,46
SNP 2 rs3827020 MAF 0,20
SNP 1 rs2236196 MAF 0,26
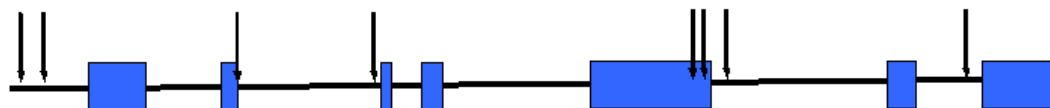
CHRNA5 15q24

SNP 1 rs667282 MAF 0,23
SNP 2 rs680244 MAF 0,33
SNP 3 rs621849 MAF 0,32
SNP 4 rs1051730 MAF 0,32
SNP 5 rs3743078 MAF 0,22

CHRNA7 15q14

SNP 1 rs3087454 MAF 0,14
SNP 2 rs868437 MAF 0,13
SNP 3 rs883473 MAF 0,15
SNP 4 rs2337506 MAF 0,11
SNP 5 rs904952 MAF 0,39

CHRNB1 17p13.1
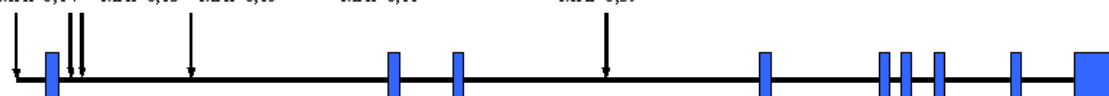
SNP1 rs3853818 MAF 0,44
SNP2 rs2302767 MAF 0,27
SNP3 rs2302762 MAF 0,27
SNP4 rs2302763 MAF 0,21
SNP5 rs2302764 MAF 0,16
SNP6 rs871990 MAF 0,22

CYP1A2 15q22 – qter

SNP 1 rs4886406 MAF 0,27
SNP 2 rs4646427 MAF 0,04
SNP 3 rs2069522 MAF 0,06

CYP2A6 19q13.2

SNP 1 rs2644914 MAF 0,22
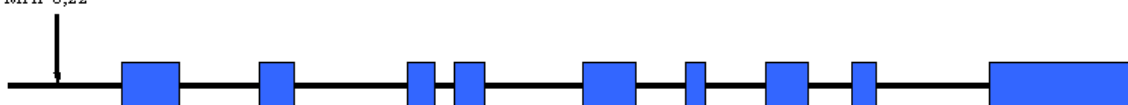
APPENDIX 3 SEX-PCR PROTOCOL

Sex-PCR master mixes, primers and PCR protocol

Component	stock concentration	1x mix	Primers	stock concentration	1x mix
10x buffer	10,00 x	1,5	SRY-F	20,0 μM	0,15
dNTP	2,00 mM	1,5	SRY-R	20,0 μM	0,15
MgCl2	25,00 mM	0,9	HTR2C-F	20,0 μM	0,15
dH2O	μl	8,34	HTR2C-R	20,0 μM	0,15
AmpliTaqGold	5,00 U/μl	0,16	total volume		15,00 μl
DNA	2,50 ng/μl	2			
total volume		17,4 μl			

Primer name	Gene	Size (bp)	Size with tails	forward primer name	reverse primer name
HTR2C	HTR2C	158	206	HTR2C-F	HTR2C-R
SRY	SRY	91	137	STY-F	SRY-R

SRY-F
5'-TAATACGACTCACTATAGGGAGAATAAGTATCGACCTCGTCGGAA-3'

SRY-R
5'-AATTAACCCTCACTAAAGGGAGACACTTCGCTGCAGAGTACCGA-3'

HTR2C-SNP1F
5'-TAATACGACTCACTATAGGGAGAGTGGTTTCAGATCGCAGTAA-3'

HTR2C-SNP1R
5'-AATTAACCCTCACTAAAGGGAGAATATCCATCACGTAGATGAGAA-3'
Program: SexPCR

Step	Temp °C	Time
1	95	11 min
2	95	30 s
3	65	1 min - 1.0 C per cycle
4	Go to step 2, 4 times	
5	95	30 s
6	60	30 s - 0.5 C per cycle
7	72	30 s
8	Go to step 5, 14 times	
9	95	30 s
10	53	30 s
11	70	30 s
12	Go to step 9, 13 times	
13	72	6 min
14	10	forever

APPENDIX 4 PCR PROTOCOLS FOR PRIMER TESTS AND GENOTYPING

Primer tests and validation:

Singleplex

SNP 5 µl PCR reaction:	final	1 reaction (µl)	primer tests 25x master mix	validation 1/24
dH2O		3,58	89,50	158
10x Qiagen Buffer	1x	0,50	12,50	22
25 mM MgCl2	1 mM	0,20	5,00	9
2mM dNTPs	200 µM	0,50	12,50	22
10 µM primer-Reverse	200 nM	0,10		4,4
10 µM primer-Forward	200 nM	0,10		4,4
HotStarTaq polymerase (5U/µl)	0.1 units	0,02	0,50	0,9
Total volume mix:		5	120	220

Aliquot 30 µl/ tube master mix and add primers 0,6 µl each

PCR program: PCR-45

95°C for 15 minutes

95°C for 20 seconds
56°C for 30 seconds
72°C for 1 minute

45 cycles

72°C for 3 minutes

10°C forever

PCR-45 takes about 2.5 hours to run.

Multiplex testing and optimisation

4-plex

SNP 5 µl PCR reaction:	final	1 reaction (µl)	1/12 + extra
dH2O		2,98	178,80
10x Qiagen Buffer	1x	0,50	30,00
25 mM MgCl2	1 mM	0,20	12,00
2mM dNTPs	200 µM	0,50	30,00
10 µM each reverse primer	200 nM	0,10	6,00
10 µM each forward primer	200 nM	0,10	6,00
HotStarTaq polymerase (5U/µl)	0.1 units	0,02	1,20
Total volume mix:		5	300

96-well PCR-mix-plate preparation, 12 different mixes:
Aliquot 252 µl/tube for each of the master mixes and add primers
Divide 35 µl of master mix to eight wells (A1-H1) of the 96-well plate

APPENDIX 4 PCR PROTOCOLS FOR PRIMER TESTS AND GENOTYPING

Throughput PCR

4-plex   1x384 plate

SNP 5 µl PCR reaction:	final	1 reaction (µl)	1,28x
dH2O		3,70	1819
10x Qiagen Buffer	1x	0,50	246
25 mM MgCl2	1 mM	0,20	98
2mM dNTPs	200 µM	0,50	246
100 µM each reverse primer	200 nM	0,01	5
100 µM each forward primer	200 nM	0,01	5
HotStarTaq polymerase (5U/µl)	0.1 units	0,02	10
Total volume mix:		5	2459

APPENIDX 5 PRIMER SEQUENCES AND MULTIPLEX COMPOSITIONS AFTER ASSAY DESIGN

GENE	MPLEX	TERM	SNP_ID	2nd-PCRP	1st-PCRP	UEP_SEQ
CYP1A2	S1	ACT	rs2069522	ACGTTGGATGAACCTGTGAAGATGCCAAGG	ACGTTGGATGTTCTCCCATTCATGGCCTTC	TGGATGGGGAATCCAATAGAG
CHRNA5	S1	ACT	rs680244	ACGTTGGATGCTTATGCTACTACAGAGCTC	ACGTTGGATGTTCCTTAAGTGATCCCAGGC	TACTACAGAGCTCAAATAAGC
CHRNA7	S1	ACT	rs3087454	ACGTTGGATGCCTTCTAGAAGAGCATGAGC	ACGTTGGATGGGAATGTGATCCTGATAGCC	CCATGGGGCCAAGGGAATTTATT
CYP1A2	S2	ACT	rs4886406	ACGTTGGATGATTCTTCAGGCACTACCCTC	ACGTTGGATGTCTGTGCATCATTCAGGGTC	CGTTCCCTGCCTAACCCC
CHRNA4	S2	ACT	rs2236196	ACGTTGGATGAACCAAACACAATCCCTGCC	ACGTTGGATGTGTGTGGGCTTCAACTTCTC	CTAGCGAAGCAGATTGGAGC
CHRNA4	S2	ACT	rs3827020	ACGTTGGATGCACAGACTTCTCGGTAAGTC	ACGTTGGATGTACACACCAGGAAGAAAGGG	CTCGGTAAGTCCCGCCC
CHRNA5	S3	ACT	rs1051730	ACGTTGGATGCCTCGCAGCAGTTGTACTTG	ACGTTGGATGTCAAGGACTATTGGGAGAGC	TGTACTTGATGTCGTGTTT
CHRNA5	S3	ACT	rs667282	ACGTTGGATGTGTGATACACTGACCAACAG	ACGTTGGATGATCCAGTAGGTCAGAAGCAG	TACACTGACCAACAGTATTCAC
CHRNA7	S4	ACG	rs2337506	ACGTTGGATGATTTCCTACTCCTCGTCCAC	ACGTTGGATGATGTGTGGCTGACTACTGGC	ACAGGTCAGCTCCTACACC
CHRNA4	S4	ACG	rs2273505	ACGTTGGATGAGACCCCTGTGCTCCTTGCA	ACGTTGGATGCAGCTCATTGACGTGGTAGG	GCGACCTCAGTCACAGTGCA
CHRNA4	S4	ACG	rs2273502	ACGTTGGATGGGTGACCCCTTGGTGTCTTTC	ACGTTGGATGTCCACCATATCTTGCCCTGG	CCCCTTGGTGTCTTTCTCTGGC
CHRNA5	S4	ACG	rs621849	ACGTTGGATGTAGAATAAGGGACACTCTGG	ACGTTGGATGTGTTTCTTAAGGAGCTCAGC	AAGGGACACTCTGGTTCAGATA
CYP1A2	S5	ACT	rs4646427	ACGTTGGATGATAGCAATTGAGGTCCCCTC	ACGTTGGATGAATGTGTGCAGGTTCAGCAG	GAGGTCCCCTCCTTTCATTAT
CYP2A6	S5	ACT	rs2644914	ACGTTGGATGAAGCCAACACGATGCCTTTC	ACGTTGGATGAGCAGTGGACACACTGATGA	CTGATGCTGATATCAGAACCC
CHRNA7	S5	ACT	rs904952	ACGTTGGATGAGTAGATGGACTGAAGGAAG	ACGTTGGATGCCACGACACAGTCAAATTGG	AATGAAATTTAGCAGTTTAGGGA
CHRNB1	S5	ACT	rs3853818	ACGTTGGATGAGTGCTCCAAAAGGGCTCCC	ACGTTGGATGAGGAAGGCTTTGCCCTTTGG	GCTCCAAAAGGGCTCCCTTCCCGT
CHRNB1	S7	ACT	rs871990	ACGTTGGATGAAGTCAGAGTCCTTCCAGCC	ACGTTGGATGACAGCAGTCAGCCCAGCCAC	CTCCTCCGTGGTGGTGGC
CHRNB1	S7	ACT	rs2302763	ACGTTGGATGAGTGCTGGTGCGGAATTTGG	ACGTTGGATGAGGACCACGATGCGGTATGTC	CTAGACCTGGCCTTGTTCC
CHRNB1	S7	ACT	rs2302767	ACGTTGGATGCCCTTTCCTTAGACATCCTG	ACGTTGGATGTAGACGAGGTGAGGGTCATG	CCTTAGACATCCTGACTCCCC
CHRNA5	K1	ACT	rs3743078	ACGTTGGATGAAGGCACTGGAAGATGAGAG	ACGTTGGATGGCCTCAGTATCCCTGTTGGT	GTCACCCATTTCCTGGC
CHRNA4	K1	ACT	rs1044397	ACGTTGGATGACGGTCAAGACCCGCAGCAC	ACGTTGGATGTCTGCAATGTACTGGACGCC	AGACCCGCAGCACCAAAGC
CHRNA7	K1	ACT	rs883473	ACGTTGGATGGGGCTATGAGGAAAACCAAC	ACGTTGGATGTCTGCCGAAGTTACCAAGTC	TCGGACAAGTCCTTTGTCTTTC

APPENIDX 5 PRIMER SEQUENCES AND MULTIPLEX COMPOSITIONS AFTER ASSAY DESIGN

x

GENE	MPLEX	TERM	SNP_ID	2nd-PCRP	1st-PCRP	UEP_SEQ
CHRNA4	K2	ACG	rs3810471	ACGTTGGATGCCTCTGACCCACTCACAGAA	ACGTTGGATGTCTCTGTCCCCTTCCTAAGC	CACTCACAGAACCCCAC
CHRNA4	K2	ACG	rs755203	ACGTTGGATGGGCTCCAGGAAAGAGTTCAG	ACGTTGGATGTGTAGTTTTGGGCTGACGTG	AGAGTTCAGGAAAGTAGAC
CHRNA7	K2	ACG	rs868437	ACGTTGGATGCCTGGCAAAGTGGAAAAAGG	ACGTTGGATGGTGATATACCATTCATCAAC	TCCTACAAATAAACAGTTACC
CHRNB1	K3	ACG	rs2302764	ACGTTGGATGTTGAAGACTGGAGGGTTGAG	ACGTTGGATGGAAGGAGTTAAGAGGCAGAG	GTGATACTGTCAAGCCC
CHRNA7	K3	ACG	rs2337980	ACGTTGGATGGCCATAGTGGCCAAAATGAG	ACGTTGGATGCCACCTCAAAAAAACACAGG	GGGTGCTGGTACTTTGA
CHRNB1	K3	ACG	rs2302762	ACGTTGGATGAGAAGAGTGTGCGGAAGAAG	ACGTTGGATGAGCAGCAGGAGTACTACAAC	GAAGAAGCGGCCCATGCTC
CHRNA4	K3	ACG	rs1044396	ACGTTGGATGCGTGCAAATGCACATGCAAG	ACGTTGGATGCTTTGGTGCTGCGGGTCTTG	CCCTCTTCGGTGTCCCCGAG

APPENDIX 6 RESULTS FROM THE PSEUDOMARKER ANALYSES

All the results of the Pseudomarker association analysis first with nicotine dependence phenotype and both dominant and recessive models and then with regular smoker

Locus	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
dominant pseudomarker results, nicotine dependence					
A4_1	0.5	0.554135	0.617098	0.751836	0.696796
A4_2	0.0310632	0.57163	0.751836	0.0544269	0.100418
A4_3	0.335696	0.920323	1	0.662937	0.786155
A4_4	0.5	0.698551	0.698551	1	0.813147
A4_5	0.382091	0.0265692	0.0275079	0.698551	0.0534494
A4_6	0.370076	0.0106262	0.0101028	0.887517	0.0230715
A4_7	0.335696	0.29425	0.261632	0.887517	0.392586
A4_8	0.364523	0.192269	0.17849	0.920323	0.28991
A5_1	0.5	0.00280013	0.00683534	0.203072	0.00712373
A5_2	0.5	0.423716	0.631546	0.521988	0.574932
A5_3	0.5	0.654742	1	0.654742	0.779789
A5_4	0.127097	0.126081	0.153195	0.205883	0.109223
A5_5	0.370076	0.00118043	0.00484916	0.100347	0.00300255
A7_1	0.420734	0.29425	0.298682	0.806491	0.42558
A7_2	0.5	0.646789	0.671392	0.862474	0.773557
A7_3	0.5	0.402785	0.406166	0.920323	0.553736
A7_4	0.395667	0.617098	0.590244	0.862474	0.711887
A7_5	0.5	0.920323	0.920323	1	0.957668
B1_1	0.5	0.252375	0.254194	0.920323	0.385908
B1_2	0.263554	0.920323	0.920323	0.527109	0.668318
B1_3	0.308549	0.751836	0.806491	0.590244	0.696796
B1_4	0.5	0.155224	0.16584	0.751836	0.259722
B1_5	0.170052	0.823054	0.806491	0.342772	0.472979
B1_6	0.5	0.113121	0.120555	0.751836	0.199098
1A2_1	0.5	0.365175	0.377138	0.841468	0.514412
1A2_2	0.127097	0.718445	0.806491	0.241792	0.360468
1A2_3	0.0346445	0.427361	0.887517	0.0480162	0.0938023
2A6_1	0.5	0.254194	0.277338	0.729045	0.38812

APPENDIX 6 RESULTS FROM THE PSEUDOMARKER ANALYSES

Locus	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
recessive pseudomarker results, nicotine dependence					
A4_1	0.5	0.537624	0.624229	0.708295	0.682291
A4_2	0.0210739	0.502352	0.512007	0.041653	0.0668172
A4_3	0.271513	0.887517	0.841468	0.554135	0.677577
A4_4	0.5	0.841468	0.841468	1	0.910833
A4_5	0.308549	0.0664039	0.0625029	0.698551	0.110378
A4_6	0.331468	0.0390392	0.0349215	1	0.0714945
A4_7	0.26616	0.512007	0.409589	0.708295	0.514412
A4_8	0.295122	0.462444	0.374096	0.841468	0.511303
A5_1	0.5	0.0390392	0.0463349	0.590244	0.0789382
A5_2	0.5	0.462444	0.631546	0.577703	0.612912
A5_3	0.5	0.887517	0.920323	0.920323	0.938784
A5_4	0.172734	0.0491727	0.0460607	0.377138	0.0608502
A5_5	0.420734	0.0230021	0.0296622	0.488437	0.0481912
A7_1	0.420734	0.225325	0.230119	0.791335	0.344565
A7_2	0.5	0.521988	0.543025	0.841468	0.668318
A7_3	0.5	0.337465	0.342772	0.887517	0.484374
A7_4	0.403245	0.380214	0.380214	0.806491	0.511303
A7_5	0.5	0.662937	0.671392	0.920323	0.786155
B1_1	0.5	0.168061	0.170317	0.887517	0.277401
B1_2	0.256003	0.729045	0.617098	0.583905	0.60895
B1_3	0.308549	0.920323	0.751836	0.689174	0.744119
B1_4	0.5	0.025962	0.0294904	0.639062	0.0548526
B1_5	0.174097	0.662937	0.823054	0.312505	0.4433
B1_6	0.5	0.0122308	0.0140855	0.617098	0.0277568
1A2_1	0.5	0.188392	0.203072	0.740153	0.304722
1A2_2	0.10437	0.751836	0.920323	0.196239	0.313308
1A2_3	0.0408714	0.777299	1	0.0778191	0.144504
2A6_1	0.5	0.617098	0.662937	0.806491	0.749797

APPENDIX 6 RESULTS FROM THE PSEUDOMARKER ANALYSES

Locus	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
dominant pseudomarker results, regular smoker					
A4_1	0.5	0.806491	0.806491	1	0.888468
A4_2	0.323394	0.662937	0.565681	0.791335	0.67292
A4_3	0.5	0.646789	0.777299	0.718445	0.773557
A4_4	0.420734	0.198939	0.345468	0.371089	0.311569
A4_5	0.5	0.00955226	0.00955226	1	0.0221438
A4_6	0.5	0.0392695	0.0392695	1	0.0793512
A4_7	0.5	1	1	1	1
A4_8	0.5	0.777299	0.791335	0.920323	0.869044
A5_1	0.460162	0.0869504	0.399444	0.135343	0.15817
A5_2	0.5	1	1	1	1
A5_3	0.5	0.543025	0.920323	0.548527	0.687065
A5_4	0.070357	1	0.920323	0.141632	0.239308
A5_5	0.203083	0.00391189	0.0656034	0.0176772	0.00683941
A7_1	0.5	0.170317	0.18712	0.708295	0.280472
A7_2	0.364523	0.493003	0.483942	0.751836	0.593479
A7_3	0.319531	0.164742	0.153195	0.740153	0.241928
A7_4	0.388649	0.334852	0.327175	0.823054	0.459197
A7_5	0.5	0.791335	0.791335	1	0.87847
B1_1	0.5	0.764181	0.791335	0.887517	0.860089
B1_2	0.5	0.139802	0.148252	0.764181	0.238009
B1_3	0.5	0.149226	0.161497	0.729045	0.25134
B1_4	0.5	0.610143	0.662937	0.791335	0.744119
B1_5	0.5	0.0714318	0.104201	0.434794	0.134172
B1_6	0.5	0.610143	0.662937	0.791335	0.744119
1A2_1	0.235427	0.806491	0.920323	0.450268	0.597292
1A2_2	0.0807487	0.30093	1	0.0817427	0.150775
1A2_3	0.055495	0.0802469	1	0.0179821	0.0393961
2A6_1	0.5	2.29727e-06	0.000227965	0.0030899	8.12385e-06

APPENDIX 6 RESULTS FROM THE PSEUDOMARKER ANALYSES

Locus	Linkage	LD|Linkage	LD|NoLinkage	Linkage|LD	LD+Linkage
recessive pseudomarker results, regular smoker					
A4_1	0.5	0.179694	0.179694	1	0.293132
A4_2	0.327371	0.113841	0.121328	0.583905	0.179794
A4_3	0.5	0.583905	0.583905	1	0.722307
A4_4	0.5	1	1	1	1
A4_5	0.5	0.192269	0.192269	1	0.309842
A4_6	0.5	0.29425	0.342772	0.654742	0.4356
A4_7	0.5	0.342772	0.371089	0.751836	0.4902
A4_8	0.5	0.254194	0.254194	1	0.38812
A5_1	0.5	0.29425	0.317297	0.751836	0.4356
A5_2	0.5	0.0885819	0.0942648	0.751836	0.161576
A5_3	0.5	0.254194	0.254194	1	0.38812
A5_4	0.0501734	0.751836	1	0.0942648	0.170431
A5_5	0.375918	0.192269	0.654742	0.205883	0.293132
A7_1	0.5	0.0512681	0.0577928	0.654742	0.100418
A7_2	0.375918	0.157284	0.179694	0.583905	0.248612
A7_3	0.5	0.0613808	0.069289	0.654742	0.117577
A7_4	0.5	0.0319928	0.035959	0.654742	0.0661258
A7_5	0.5	0.583905	0.583905	1	0.722307
B1_1	0.5	0.317297	0.342772	0.751836	0.461914
B1_2	0.5	0.00965987	0.0102168	0.751836	0.0223721
B1_3	0.5	0.0102168	0.0108067	0.751836	0.02355
B1_4	0.5	0.273303	0.317297	0.654742	0.411057
B1_5	0.5	0.0319928	0.100347	0.168061	0.0661258
B1_6	0.5	0.254194	0.29425	0.654742	0.38812
1A2_1	0.375918	0.0339153	0.0429014	0.479514	0.0661258
1A2_2	0.0646822	0.179694	0.438585	0.0613808	0.0858181
1A2_3	0.0569204	0.438585	0.317297	0.147286	0.145273
2A6_1	0.5	0.527109	0.751836	0.583905	0.67292