



UPPSALA
UNIVERSITET

UPTEC X15 035

Examensarbete 30 hp
Januari 2019

Genome-wide association study of drug-induced angioedema

Caroline Johansson



UPPSALA
UNIVERSITET

**Teknisk- naturvetenskaplig fakultet
UTH-enheten**

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
<http://www.teknat.uu.se/student>

Abstract

Genome-wide association study of drug-induced angioedema

Caroline Johansson

Angioedema is an adverse drug reaction of drugs commonly used for treatment of hypertension and heart failure. It involves a sudden swelling in the head and neck region and can be potentially life-threatening if affecting the upper airways. In this project, a genome-wide association study was done of ACEi- and ARB-induced angioedema, aiming to identify possible genetic risk factors predisposing patients to this rare but very critical adverse drug reaction.

Handledare: Mia Wadelius
Ämnesgranskare: Åsa Johansson
Examinator: Jan Andersson
ISSN: 1401-2138, UPTec X15 035

Populärvetenskaplig sammanfattning

Högt blodtryck och hjärtsvikt är vanliga åkommor som allt fler behandlas för. Möjligheterna till behandling har klart förbättrats, men en del patienter reagerar negativt på de läkemedel som ges. För ett fåtal läkemedel finns genetiska markörer som kan förutsäga risk för biverkningar, men för de flesta är detta ett outforskat område. Forskningsstudien SWEDEGENE samlar in kliniska uppgifter och DNA från drabbade patienter och undersöker om genetiska skillnader kan förklara varför vissa drabbas av allvarliga biverkningar. I denna studie har angioödem relaterat till ACE-hämmare och angiotensin II-receptorblockerare studerats. Angioödem innebär en svullnad i hud och slemhinnor, ofta i ansikte och hals, och reaktionen kan vara direkt livshotande om den drabbar de övre luftvägarna. Syftet med studien var att öka kunskaperna om angioödem utlöst av dessa läkemedel och målet var att hitta genetiska riskfaktorer som i framtiden skulle kunna användas för att utveckla kliniska test för att individualisera val av läkemedel.

En så kallad helgenomanalys (eng. genome-wide association study, GWAS) gjordes, vilket innebär en kartläggning av variationer i arvsmassan hos drabbade patienter och friska kontroller där man försöker hitta genetiska skillnader som kan förklara varför vissa drabbas. I studien kunde flera signifikanta varianter, så kallade SNPs ("snippar") i KCNMA1-genen identifieras som var associerade med angioödem relaterat till dessa blodtrycksmediciner.

Table of contents

List of abbreviations	1
1 Introduction	3
1.1 Adverse drug reactions	3
1.2 Angioedema	4
1.3 Swedegene	6
1.4 Aim	6
2 Background	7
2.1 Genetic association studies	7
2.1.1 Single nucleotide polymorphism (SNP)	7
2.1.2 Linkage disequilibrium	8
2.1.3 Haplotypes	8
2.1.4 The Hardy-Weinberg equilibrium	8
2.2 Genome-wide association studies (GWAS)	9
2.2.1 Study design	10
2.2.2 Genotyping	10
2.2.3 Analysis for association	10
2.2.4 Principal component analysis	11
2.2.5 Imputation	12
2.2.6 Functional annotation	12
2.3 GWAS in pharmacogenomics	14
2.3.1 GWAS of drugs acting on the angiotensin system	15
3 Methods	15
3.1 Selection of study participants	16
3.2 Genotyping	16
3.3 Quality control	16
3.3.1 Minor allele frequency	17
3.3.2 Missing rate per SNP	17
3.3.3 Missing rate per person	17
3.3.4 Hardy-Weinberg Equilibrium	17
3.3.5 Summary of quality control steps performed	17
3.4 Principal component analysis	18
3.5 Imputation	18
3.6 Analysis for association	18
3.6.1 Logistic regression	18
3.6.2 Sub Analysis	19

3.6.3	Visualisation.....	19
3.7	Functional annotation.....	20
3.7.1	ENCODE.....	20
3.7.2	Roadmap.....	20
3.7.3	USCS browser.....	20
3.7.4	HaploReg.....	20
4	Results	21
4.1	Part I: Genotyped data.....	21
4.1.1	Original genotyped data.....	21
4.1.2	Post-QC results.....	21
4.1.3	Population stratification.....	22
4.1.4	Analysis of genotyped data.....	23
4.2	Part II: Imputed data.....	27
4.2.1	Original imputed data.....	27
4.2.2	Post-QC results.....	27
4.2.3	Analysis of imputed data.....	27
4.3	Part III: Analysis of top SNP.....	30
4.4	Part IV: Regulatory potential of top SNPs.....	32
4.4.1	Chromatin state.....	32
4.4.2	Transcription factors	32
5	Discussion.....	35
6	Conclusion and outlook.....	36
7	Acknowledgements	37
	References	38
	Appendix A.....	41

List of abbreviations

ACEi	Angiotensin converting enzyme inhibitor
ADRs	Adverse drug reaction
ARBs	Angiotensin II type blocker
DNA	Deoxyribonucleic acid
GWAS	Genome-wide association study
HWE	Hardy–Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
OD	Odds ratio
PCA	Principal component analysis
QC	Quality control
SNP	Single nucleotide polymorphism

1 Introduction

Our population is getting older and older and more people need treatment for common medical conditions such as high blood pressure (hypertension) and heart failure. The possibilities of treatment have clearly improved, but some patients react adversely to the medicines given. For a few drugs, there are genetic markers that can predict the risk of adverse drug reactions (ADRs), but for most, the current knowledge of possible genetic causes of these reactions are inadequate (Daly 2013).

Angiotensin converting enzyme inhibitors (ACEi) and *Angiotensin II type blockers* (ARBs) are agents commonly used for treating hypertension and heart failure. In 2014, 7% of the Swedish population was treated with an ACE inhibitor and 6% was treated with an ARB (The National Board of Health and Welfare 2013). For most of these patients, these drugs are safe and effective, but some patients experience adverse drug reactions. A rare, but very critical ADR is angioedema which is a sudden swelling in the head and neck region that can be potentially life-threatening if affecting the upper airways. Angioedema occurs in 0.1-0.7% of patients treated with an ACEi and more rarely, 0.1%, in patients treated with ARB (Wadelius *et al.* 2014). This thesis aims to identify possible genetic risk factors predisposing patients to angioedema induced by these drugs.

1.1 Adverse drug reactions

Adverse drug reactions are an important clinical issue and a major health care problem and economic burden. ADRs are a common reason for hospitalisation and one of the leading cause of death in hospitalised patients in the US (Lazarou 1998). The World Health Organization's (WHO) definition of an adverse drug reaction is:

“A response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or for modification of physiologic function”. (Edwards & Aronson 2000)

This definition has been in use since 1972 and is commonly used. However, more exhaustive definitions of ADRs have been proposed, for instance by *Uppsala Monitoring Centre* (UMC), a WHO collaborating centre for international drug monitoring, which suggest the following definition of an ADR:

“An appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which predicts hazard from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product.”(Edwards & Aronson 2000)

So called Type A (augmented) reactions are common and constitute around 80% of all ADRs (Ritter *et al.* 2008). They are related to dose and the pharmacological action of the drug which make these types of reactions predictable. Bleeding from using the anticoagulant warfarin is an example of a type A reaction. Type B (bizarre) reactions on the other hand are uncommon,

non-dose related and not related to the pharmacological action, making them very unpredictable. Type B reactions can be immunological reactions like penicillin hypersensitivity or idiosyncratic reactions. Angioedema related to ACE inhibitors and ARBs is classified as a type B reaction. Sweden has a system for reporting adverse drug reactions, established in 1965, where reports are sent to the Swedish Medical Products Agency (MPA). According to Swedish law, (LVFS 2012:14, 19§), it is mandatory for health care professionals as physicians, dentists and nurses to report all suspected ADRs.

1.2 Angioedema

Angioedema (*angio* relating to blood vessels and *edema* to swelling) is a sudden swelling in the deep reticular dermis, subcutaneous or submucosal tissues (i.e. the deep and middle layers of the skin and the connective tissue supporting the mucosa) (Wadelius *et al.* 2014). It is caused by vasodilation i.e. widening of blood vessels and increased secretion of fluid into the interstitial compartment, i.e. the space surrounding tissue cells. The swelling can be potentially life-threatening if affecting the upper airways (Wadelius *et al.* 2014). Angioedema is an adverse drug reaction of ACEi and ARB, drugs commonly used for treatment of hypertension and heart. It is classed as a type B (bizarre) reaction since it is very unpredictable and not related to dose or the pharmacological action of the drugs. Other causes of angioedema include hereditary or acquired mediated by bradykinin or vasoactive molecules, where the hereditary angioedema is caused by mutations in the *SERPINE1* or *F12* gene (Wadelius *et al.* 2014). Other types are allergic or pseudoallergic angioedema dependent on mast cell degranulation, which usually can be treated with antihistamines and epinephrine (adrenaline), and idiopathic angioedema which has no known pathophysiology (Wadelius *et al.* 2014). The phenotype of drug-induced angioedema has been standardized for enabling global multicentre investigations of the underlying factors predisposing patients to this. The phenotype definition of angioedema induced by ACEi or ARB includes a swelling in the head and neck region that is first occurring during treatment of these drugs (Wadelius *et al.* 2014). It should not be mixed up with urticaria (hives) which is remarkably similar but affect more superficial layers of the skin.

Both *Angiotensin converting enzyme inhibitors* (ACEi) and *Angiotensin II type blockers* (ARBs) act on the angiotensin system which regulates blood pressure and fluid balance (homeostasis). ACEi and ARB cause vasodilation, a widening of blood vessels induced by a relaxation of smooth muscle cells in the cell walls, which lowers the blood pressure. The theoretical effect of these drugs and how they act on the angiotensin system and the angiotensin-bradykinin pathways, can be seen in Figure 1.

Blood pressure is naturally increased indirectly by the angiotensin-converting enzyme (ACE) which causes blood vessels to constrict by generating angiotensin II from angiotensin I. ACE also degrades the vasodilator bradykinin to inactive metabolites, see Figure 1a. ACE inhibitors work by inhibiting the enzyme ACE which results in a decrease of the formation of

angiotensin II. This in turn results in decreased degradation of bradykinin, leading to increased vasodilation and decreased blood pressure, see Figure 1b (Wadelius *et al.* 2014). Angioedema induced by ACEi can theoretically be caused by an accumulation of bradykinin that otherwise would have been degraded by ACE. There are alternative pathways, Aminopeptidase P (APP) and Membrane metallo-endopeptidase (MME), which can inactivate bradykinin (Wadelius *et al.* 2014). These usually step in, but if they are deficient, bradykinin could accumulate and genetic variants in these pathways could therefore explain why some patients develop angioedema (Mahmoudpour *et al.* 2013). The theoretical effect of ARBs is less clear and even though ARBs have no direct effect on ACE and the degradation of bradykinin, ARB seem to contribute to increased levels of bradykinin (Wadelius *et al.* 2014). By blocking the type 1 receptor for angiotensin II, more angiotensin II binds to the type 2 receptor instead, see Figure 1c. This in turn seem to inhibit ACE and MME causing increased bradykinin levels and vasodilation. So even though bradykinin is one of the therapeutic actions, accumulation of bradykinin may contribute to the development of angioedema.

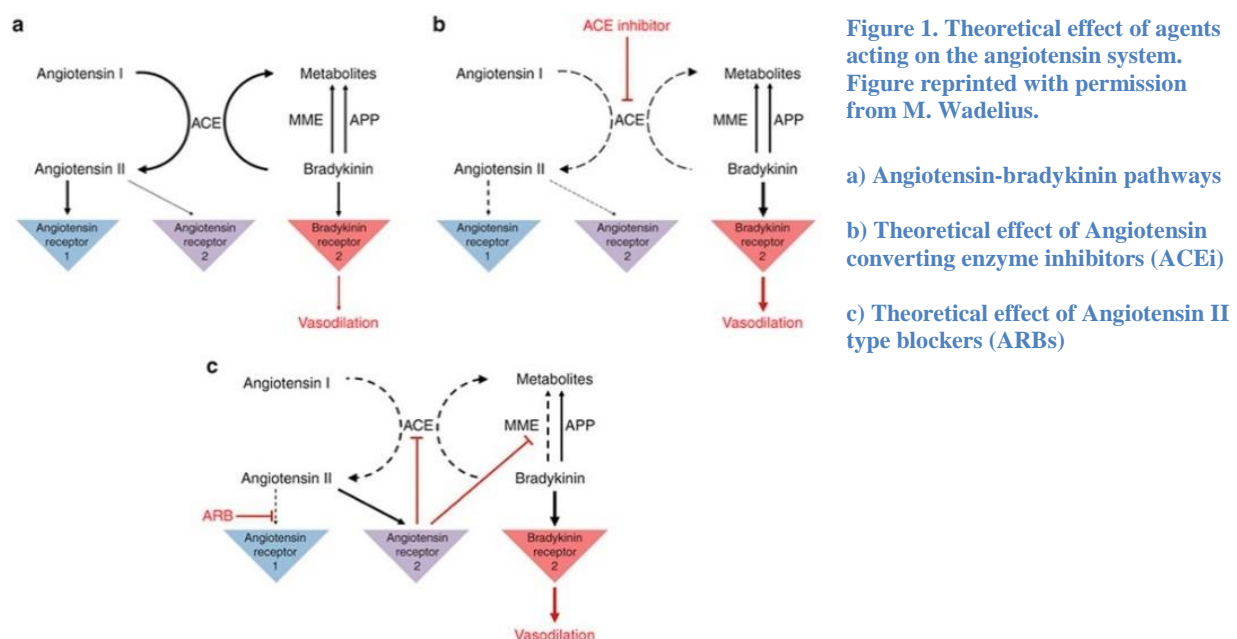


Figure 1. Theoretical effect of agents acting on the angiotensin system.
Figure reprinted with permission from M. Wadelius.

Examples of prescribed ACE inhibitors are *enalapril*, *perindopril*, *captopril*, *lisinopril*, and *ramipril*. Persistent dry cough is the most common ADR and occurs in about 9% of patients treated with an ACEi (Powers *et al.* 2012). Patients who experience this cough are often switched to the more expensive angiotensin II type blocker, which has less frequency of this ADR. Angioedema occurs in 0.1-0.7 % of patients treated with an ACEi (Wadelius *et al.* 2014). Prescribed ARBs include *candesartan*, *irbesartan*, *valsartan*, *fimasartan* and *losartan*. Angioedema is a much rarer ADR for ARB and occurs in about 0.1% of the patients (Wadelius *et al.* 2014).

1.3 Swedegene

SWEGEGENE is a project aiming to identify clinical and genetic risk factors predisposing patients to adverse drug reactions. It is a collaborative project between Uppsala University (Department of Clinical Pharmacology), the Swedish Medical Products Agency and Karolinska Institute (Department of Clinical Pharmacology) and aims to increase knowledge of adverse drug reactions. *SWEGEGENE* aims to establish a database of clinical data and a biobank of DNA from patients who have experienced ADRs to enable studies of both genetic and acquired risk factors.

Reported cases of ADRs are stored by the Swedish Medical Products Agency (MPA). Clinical data from cases in the *SWEGEGENE* project is obtained from medical records as well as from a standardised subject questionnaire which is sent to participating patients. A blood sample is then taken at a health-care facility and posted for storage at Uppsala University before DNA is extracted. Patients who have experienced ADRs are recruited to the project predominantly through MPA, but also through collaboration with health-care facilities and hospitals. Patients can also contact *SWEGEGENE* if they have experienced an ADR and want to participate in the study.

SWEGEGENE started as a part of the *EUDRAGENE* project, a European research project which focused on investigating genetic determinants of seven serious type B ADRs, for instance myopathy caused by statins or fibrates and agranulocytosis. The project started in 2005 and has since 2010 continued under the name *SWEGEGENE* in Sweden. *SWEGEGENE* is also a partner in the ongoing project *PREDICTION-ADR (Personalisation of tREatment In Cardiovascular disease through next generation sequencing in Adverse Drug Reactions)*, funded by the European Union's *7th Framework program for Research and Technological Development (FP7)* in 2013. It aims to find genetic factors that can explain ADRs from cardiovascular disease drugs, focusing on ACE-inhibitor induced angioedema and statin-induced myopathy. The *SWEGEGENE* project has ethics approval from the Uppsala Regional Ethical Review Board (EPN Uppsala Dnr 2008/213 and Dnr 2010/231).

1.4 Aim

In this thesis I aim to investigate possible genetic factors predisposing patients to angioedema triggered by treatment with drugs acting on the angiotensin system. A genome-wide association study (GWAS) is done based on data from the *SWEGEGENE* project and about 5000 population controls from the Swedish Twin Registry. Identification of possible risk factors could lead to an increase in the understanding of the pathogenesis behind angioedema, knowledge that consequently could be used to identify patients with increased risk of experiencing an ADR. The goal is to find genetic markers that can be used to develop clinical tests which in the future could be used to individualize the choice of drug, i.e. personalised medicine.

2 Background

The human genome consists of three billion base pairs distributed in 23 chromosome pairs, where one set is from the mother and the other from the father. Chromosomes are the packed and organised structure of DNA (deoxyribonucleic acid) which is made of nucleotides with four type of bases - adenine (A), thymine (T), cytosine (C) and guanine (G). The bases are distributed in two chains, where “A” only pairs with “T” and “C” only pairs with “G”, which is a key factor of DNA replication. Even though the nucleotides always are in fixed pairs the pairs can come in any order. This way, DNA can be likened to a recipe with chemical information where all instruction needed for making all the proteins and components a cell will ever need are encoded in the order of the bases. A single set of these instructions is called a gene. A gene is a functional region of DNA and includes sequences that regulate gene activity. Genes code for functional RNA (ribonucleic acid) or for proteins. The coding regions of a gene is called *exons* and the non-coding parts *introns*. The *Human Genome Project* first published the complete sequence of the human genome in 2001 and estimated that it consists of 20 000 to 25 000 protein coding genes, which is about 1-2 % of the human genome. The rest is so-called noncoding DNA even though it still can have biochemical activity and include functional and regulatory elements (The ENCODE Project Consortium 2012).

The genome contains the hereditary information encoded in the DNA and except for monozygotic twins, the genome for every individual human is unique. Even though two genomes are roughly 99.9% identical, genetic heritability among individuals are caused by the small fraction that differs. This variation can explain susceptibility to disease and why some people experience adverse drug reactions.

2.1 Genetic association studies

The principle of association studies in genetics is to link a phenotype, e.g. a disease or an adverse drug effect, to genetic variation. There is an association if the genetic variant is found more often than expected by chance in a person having the phenotype/trait of interest. A person who carries one or two copies of a high-risk variant could then be at increased risk of developing or having the associated trait, for instance increased risk of experiencing an adverse drug reaction of a certain drug. Most commonly used genetic variants in association studies are so called *single nucleotide polymorphisms* (SNPs; pronounced "*snips*"), which capture most of the common genetic variability (Eriksson 2012).

2.1.1 Single nucleotide polymorphism (SNP)

A SNP is a single base substitution that occurs commonly within a population, i.e. in 1 % of the population (Campbell & Reece 2008). For example, an “A” in the first chromosome can have been changed to a “G” in the second chromosome resulting in two different variants. The most common variant is called the major allele and the least common is the minor allele. *Minor allele frequency* (MAF) is a measurement of the frequency of a SNP (Eriksson 2012).

If “A” is the major allele and “G” the minor allele, a person carrying these alleles would be heterozygous G/A (minor/major allele). SNPs can be located anywhere in the genome but occur more frequently in non-coding regions than in coding regions. A SNP in a coding region of a gene that affects the amino acid sequence of protein is called non-synonymous and SNPs that do not change the protein are called synonymous. However, both synonymous SNPs and SNPs in non-coding regions can still be functional SNPs and affect the expression of a protein (Gibson & Muse 2009). Genetic variation between individuals can explain differences in susceptibility to disease or why some patients experience adverse drug reactions.

As of 21 June 2015, about 150 million SNPs were registered for the human genome in the National Centre for Biotechnology Information (NCBI) SNP database (dbSNP build 144). This database contains very rare SNPs as well, so called single nucleotide variants (SNVs) since they have no requirement or assumption about minimum allele frequency. Around 10 million SNPs are considered common in the human genome (i.e. present in at least one percent of the general population) and occur once every 100 to 300 bases (Campbell & Reece 2008).

2.1.2 Linkage disequilibrium

SNPs located in the same chromosomal region are often to, some degree, correlated and alleles can be inherited together in haplotype blocks. This is called linkage disequilibrium (LD) (Gibson & Muse 2009). The dependency is usually measured by D , D' or r^2 , used for different purposes where r^2 is useful in association analyses. r^2 is the squared correlation between SNPs and a value of 1 means that two SNPs are in total LD and thereby always inherited together. A value of 0 mean that the SNPs are completely independent. Since not all SNPs can be genotyped in a GWAS, LD is very helpful for enabling indirect association testing and find a functional SNP even though it has not been genotyped (Gibson & Muse 2009).

2.1.3 Haplotypes

Chromosomes occurs in pairs with one chromosome inherited from each ancestor. However, before they are passed on they undergo a process called recombination which slightly changes the copies (Campbell & Reece 2008). In this process two chromosomes in a pair exchange information (pieces of DNA) with each other resulting in a new chromosome pair containing pieces from both. A haplotype block is a region in which the frequency of recombination historically has been low and is therefore a region of high LD (Gibson & Muse 2009). The haplotype blocks are separated by regions where there has been recombination, called recombination hot-spots. The variation in a haplotype block can often be characterised by one SNP – a tag SNP or haplotype tagging SNP (Gibson & Muse 2009).

2.1.4 The Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium (HWE) is a principle that states the association between allele frequencies and genotypes in a population (Campbell & Reece 2008). When a SNP is in

Hardy-Weinberg equilibrium the genotype frequencies in a population will remain constant in successive generations, if random mating is assumed (Eriksson 2012). The expected genotype frequencies for SNPs in the Hardy-Weinberg equilibrium are calculated using the minor allele frequency. A SNP can be out of HWE for various reasons, for example a SNP could be selected for or against and genetic drift could change the allele frequency. The cause could also be genotyping errors and therefore, checking SNPs for HWE is a common quality control procedure in GWAS (Eriksson 2012).

2.2 Genome-wide association studies (GWAS)

Since 2007, a genome-wide approach where millions of SNPs are studied simultaneously has increasingly been applied to association studies. Genome-wide association studies have moved the focus from candidate gene studies, which are based on a priori knowledge of a gene's functional impact on the trait, to instead study the entire genome and detect common genetic variation in an unbiased way (Gibson & Muse 2009).

Factors enabling these studies include the availability of SNP databases and the expanding knowledge of haploblock structures and common patterns of variation generated by the *International HapMap Project*. *The 1000 Genomes Project* and other resources to identify SNPs as well as technical advances in SNPs genotyping arrays have enhanced the development even further and has been important prerequisites (Figure 2). Faster computers to carry out the millions of statistical tests and better ways to store the huge amount of data have made the analyses more amenable. Today's chips can genotype millions of SNPs which capture most of the common genetic variation.

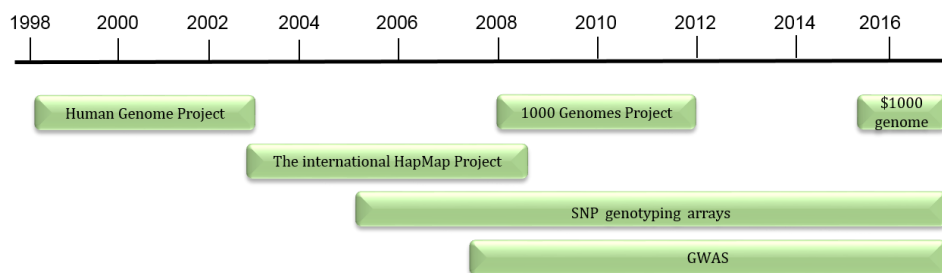


Figure 2. Timeline showing important events enabling genome-wide association studies.

A typical GWA study has distinct parts including selection of a large number of patients with the trait of interest and a suitable comparison group followed by DNA isolation and genotyping (Gibson & Muse 2009). The obtained data is reviewed to ensure good quality before the statistical tests for association is done. Principal component analysis is used to correct for population stratification and other cofounders and imputation can be done to include SNPs that have not been genotyped (Price *et al.* 2006). The last parts include replication of identified associations in independent samples to confirm the result and functional annotation of found top SNPs (Pare 2010).

2.2.1 Study design

The most frequently used study design in GWA studies is the case-control design, in which allele frequencies in patients who have experienced the trait in question, for instance an ADR, are compared to an unaffected control group (Pearson & Manolio 2008). It is important that cases are truly affected and therefore it is necessary to establish a strict criteria for the phenotype so that clinicians can judge the patients relevance for the study. Misclassification of patients is a problem and can markedly reduce power and also bias the results towards no association (Pearson & Manolio 2008). The control participants should be from the same population as the cases and should be at risk of developing the same disease or ADR (Pearson & Manolio 2008). Sample size is an important factor in the study design since a GWAS considers an enormous number of variables and enough individuals are needed to provide enough power to detect variants with a low or modest effect. The power needs to be adequate to compensate for the multiple testing that is carried out in an association study (Barrett *et al.* 2014). A big sample size is therefore usually needed but large effect phenotypes, often shown by pharmacogenomics traits, allow powerful studies to be carried out on smaller sample sizes (Daly 2010).

2.2.2 Genotyping

The genetic material collected from the study participants is genotyped using genotyping platforms such as *Illumina* or *Affymetrix*. Each DNA sample is run on a SNP array to genotype up to several millions of SNPs in a single assay. A SNP array is an array containing immobilised ASO (allele-specific oligonucleotide) probes that target a specific locus in the genome. Each probe will bind to a complementary sequence in the sample DNA and stop one base before the locus of interest. By extending with one single base that incorporates one of four differently colour-labelled nucleotides (A, T, C or G), the allele specificity can be confirmed by genotype calling. The nucleotide label can emit a signal that can be detected, and the allelic ratio of a given locus can be determined by a genotype calling software based on the intensity values for each colour. For instance if the colour representing “A” is approximately as intense as the colour representing “G”, the genotype of that SNP is A/G, i.e. heterozygote (Barrett *et al.* 2014). High intensity for only one colour indicates a homozygote for that allele. The arrays usually have a set of both common and rare variants and the SNPs typically have a MAF>2.5% to tag the most common haplotypes. Rarer SNPs with a MAF<1% are too difficult to measure reliably which can result in an underpowered GWAS.

2.2.3 Analysis for association

The basic principle of genome-wide association studies is to use statistical methods to test for association, i.e. to examine if there is a difference in genotype distribution between cases and controls. A common method used in the first genome-wide association studies was to count alleles, and Figure 3 shows the methodology for how the calculations could be done in a case-control study. A SNP where one allele is significantly more common in cases than in controls is associated with the trait in question. A statistical test is used to evaluate the allele count of each measured variant, usually a chi-squared test. The example in Figure 3, shows an

overrepresentation of individuals with the G-allele of SNP1 among patients, 52.6% compared to 44.6 % among controls and a p-value of 5.0×10^{-15} . The G-allele of that SNP is thereby thought to be associated with the trait in question. Today, a more flexible analysis is usually performed where a logistic regression model is fitted to the data (Gibson & Muse 2009). An additive model where each copy of the allele associated with ADR is assumed to increase the risk of ADR by the same amount, is the most common model (Pearson & Manolio 2008).

	SNP1	SNP2	SNP ...
	Cases	Cases	<i>Repeat for all SNPs</i>
	Count of G: 2104 of 4000	Count of G: 1648 of 4000	
	Frequency of G: 52.6%	Frequency of G: 41.2%	
	Controls	Controls	
	Count of G: 2676 of 6000	Count of G: 2532 of 6000	
	Frequency of G: 44.6%	Frequency of G: 42.2%	
	P-value: $5.0 \cdot 10^{-15}$	P-value: 0.33	

Figure 3. Example of a calculation showing the principle of association analysis in a case-control study. Individuals with the G-allele of SNP1 are overrepresented, and its p-value reach the genome-wide significance level. The G-allele of that SNP is thereby associated with the trait in question.

Attribution: "Method example for GWA study designs" by Lasse Folkersen (Own work) [Licenced under CC BY 3.0 (<http://creativecommons.org/licenses/by/3.0/>)], via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Method_example_for_GWA_study_designs.png#/media/File:Method_example_for_GWA_study_designs.png

In a case-control study, the strength of an association is measured by the odds ratio (OR). The odds ratio is a ratio of two odds; odds of angioedema for individuals having a specific allele and the odds of angioedema for individuals who do not have that same allele. When the allele frequency in the case group is much higher than in the control group, the odds ratio will be higher than 1, and vice versa for the lower allele frequency. Finding ORs that are significantly different from 1 is therefore the objective of a GWAS since that would indicate that a SNP is associated with the ADR in question.

2.2.4 Principal component analysis

The data can be adjusted for variables that potentially could confound the results, such as sex and age. Population stratification, i.e. differences in allele frequencies among populations of different ancestry, can also be corrected for since it can confound the association between the

trait in question and the SNP and thereby bias the observed association (Clarke *et al.* 2011). Principal component analysis (PCA) is a commonly used method for detecting and adjusting for population stratification. The basic principle is that people who are geographically close to each other are more likely to be closely related, i.e. be more correlated in terms of genotypes, than people living far apart. This correlation can be used to distinguish between subpopulations by clustering them, and thereby remove outliers. Applying PCA to the data will produce principal components, where the first gives the linear combination of genotypes that best explain the variation in the data. The second is the orthogonal combination that best captures the remaining variation and the third is the orthogonal to the second and so forth (Barrett *et al.* 2014).

Before doing PCA the data can be pruned to get a lower chance of regions with high LD having impact on the principal components. The principal components from PCA can be included as covariates in the logistic regression model when testing for association to adjust for population stratification (Clarke *et al.* 2011). Principal components are powerful for adjusting not only for population stratification but also for other confounders, such as whether all cases and controls were genotyped in the same genotype platform or with the same type of chip or if the DNA was collected by the same method (Clarke *et al.* 2011). If cases and controls are taken from the same population, population stratification is usually not really a problem, but there might still be some remaining differences between cases and controls that can be adjusted for. In this project, both cases and controls were taken from a Swedish population, and the data should form a distinct cluster when the first principal components from the PCA are plotted and compared with other populations.

2.2.5 Imputation

Genotyping arrays today include a huge number of SNPs, but these are still a subset of all known genetic variation. It is not feasible to sequence all participants in GWAS or even sequence a small region. An alternative strategy is therefore to impute the genotypes at those SNPs not on the genotype array to get a more exhaustive coverage. This might be interesting for various reasons, for instance if there is a region associated with the phenotype of interest where only ungenotyped SNPs have a strong enough signal to reach significance, the association might be missed using data from the array only. A whole-genome imputation approach is then of interest for increasing the number of SNPs to test for association and increase the power of the study. It might also be interesting to fine-map a region associated with the trait to better understand the source of the association and to find functional variants (Barrett *et al.* 2014). Imputation is based on statistical methods where genotypes from the GWAS data are phased into haplotypes and matched to reference panels of haplotypes generated from HapMap or the 1000 Genomes Project (Howie *et al.* 2009).

2.2.6 Functional annotation

Typical GWAS hits are SNPs in linkage disequilibrium (LD) distributed within intronic or intergenic regions where most variants are unknown, including the actual functional causal variant/variants of the ADR (Mortlock & Pregizer 2012). A huge challenge in GWAS is

therefore to go from a set of associated SNPs and prioritize the different genetic variants for further study to identify the causal SNP/SNPs. The function of noncoding genomic regions spanning associated variants can be characterized bioinformatically using public available data sets from different genome projects. Comparative genomics and biochemical data sets from the ENCODE (The ENCODE Project Consortium 2012) and Roadmap projects (Ward & Kellis 2012) are great bioinformatics resources for obtaining functional information, either experimentally verified or computationally predicted.

Even though most SNPs are found in non-coding regions of the genome, some can still alter cellular responses and increase susceptibility to disease or ADRs. SNPs can be regulatory when affecting non-coding regions such as enhancers, silencers and promoters, thereby modifying transcription factor binding sites or generating new binding sites which in turn can affect gene expression (Mortlock & Pregizer 2012). Coding SNPs alter amino acid sequences and can thereby directly modify structural and biological properties of the encoded proteins. The impact of non-coding regulatory SNPs on transcription factor binding and gene expression has been relatively unexplored, but bioinformatic analyses of data from different genome projects can be used for studying gene expression and give an indication of the regulatory potential of the identified SNPs (Mortlock & Pregizer 2012). Below are some bioinformatic resources that can be used for obtaining functional information about noncoding genomic regions.

Histone modifications

Gene expression is influenced by how accessible the chromatin is to transcription. Chromatin accessibility can be changed by chemical modifications, e.g. methylation, to the histone proteins in the chromatin. A modification of a specific histone protein is called a histone mark and patterns of modification are highly variable across different cell types (The ENCODE Project Consortium 2012). Histone modifications are associated with regulator binding, transcriptional initiation and elongation, enhancer activity and repression (Ernst *et al.* 2011). Examples of histone marks are H3K4Me1 (histone H3 monomethylated at lysine 4) and H3K27ac (histone H3 acetylated at lysine 27) which are associated with enhancers and H3K4Me3 (histone H3 trimethylated at lysine 4) and H3K9ac (histone H3 acetylated at lysine 9) associated with promoters (The ENCODE Project Consortium 2012). Mapping of chromatin marks in multiple cells types (i.e. chromatin profiling) is a powerful tool for the detection of regulatory activity in the genome (Ernst *et al.* 2011). The levels of enrichment of histone marks across the genome are determined by ChIP-seq assay (chromatin immunoprecipitation followed by high-throughput sequencing) (The ENCODE Project Consortium 2012).

DNase hypersensitivity

Chromatin accessibility can also be characterized by DNase I hypersensitivity, since regulatory regions in general and promoters in particular tend to be DNase sensitive (The ENCODE Project Consortium 2012). DNase-seq is used to define sites hypersensitive to DNase I which corresponds to open chromatin, by sequencing the cut points by the DNase

enzyme with high-throughput techniques in different cell types (The ENCODE Project Consortium 2012).

Transcription factor binding

Transcription factors (TFs) also regulate gene expression by binding to the DNA and interacting with RNA polymerase. As for chromatin marks, TF binding can be assayed experimentally using ChIP-seq. Alternatively, comparisons of multiple genomes (comparative genomics) can identify putative transcription factor binding sites as short stretches of conserved consensus binding sequences.

Conservation

Comparative genomics in general is a useful tool for identifying functional non-coding DNA variants. Sequence conservation in multiple distantly related species generally indicates purging of deleterious mutations from a functional DNA element.

The different features and signatures are useful for identifying regulatory elements but cannot work as proof of function on their own. For example, a regulatory element may not have a strong DNase hypersensitive site or histone marks in any of the cell types for which there is available data, but may be highly conserved, which indicates a regulatory potential. Some elements can be non-conserved but still be regulatory by being modified epigenetically. However, some features seem to be more reliable than other, for instance the above described histone modifications. Transcription factor occupancy is also useful indicators of regulatory potential and is more specific than histone modifications. The presence of a transcription factor in combination with histone modifications are often a good indication of regulatory function in that region (Mortlock & Pregizer 2012). Different data sets can therefore complement each other and together shed light on variants with regulatory potential (The ENCODE Project Consortium 2012).

2.3 GWAS in pharmacogenomics

Genome-wide association studies on adverse drug reactions and drug responses have been done frequently in pharmacogenomics last years, with varying degrees of success (Daly 2010, Daly 2013, Motsinger-Reif et al. 2013). The term pharmacogenomics is a result of merging pharmacology (the study of drug handling and action) and genomics (the study of genes and their function). Pharmacogenomics is the study of how genetics affect a patient's response to drugs by using a genome-wide association approach including genomics and epigenetics. The term is used interchangeable with pharmacogenetics, which is a candidate-gene approach focusing more on single drug-gene interactions. Today, a trial and error strategy is commonly used for finding the most effective treatment therapy for their patient. With knowledge of how genetic variation influences drug response, pharmacogenetic testing and personalised medicine could drastically improve treatment by giving the patient the best suited drug and dose (Mancinelli *et al.* 2000). An example of an earlier GWAS is *SEARCH* which studied

statin-induced myopathy (a muscular disease induced by cholesterol lowering drugs) and found GWAS variants in the *SLCO1B1* gene (Motsinger-Reif *et al.* 2013).

2.3.1 GWAS of drugs acting on the angiotensin system

There have been previous studies, mostly candidate gene studies, of angioedema induced by drugs acting on the angiotensin system (ACEi and ARB) but the results have been inconsistent. The GWAS performed so far has not resulted in any genome-wide significant associations with SNPs (Pare *et al.* 2013). A study called ONTARGET identified 16 SNPs in African Americans and 41 SNPs in European Americans that were moderately associated with angioedema (p-values between 10^{-4} and 10^{-6}). A major limit in this previous GWAS is the small sample size, which made it hard to achieve genome-wide significance. Also, the size of the replication cohort was very small (Pare *et al.* 2013). In candidate gene studies, rs989892 in MME (the gene encoding membrane metallo-endopeptidase that degrades bradykinin) has been significantly associated with angioedema (Pare *et al.* 2013). Another candidate gene study identified the gene region XPNPEP2 to be associated with ACEi-induced angioedema (Mahmoudpour *et al.* 2013).

3 Methods

An overview of the GWAS process used in this pharmacogenomic study can be seen in Figure 4.

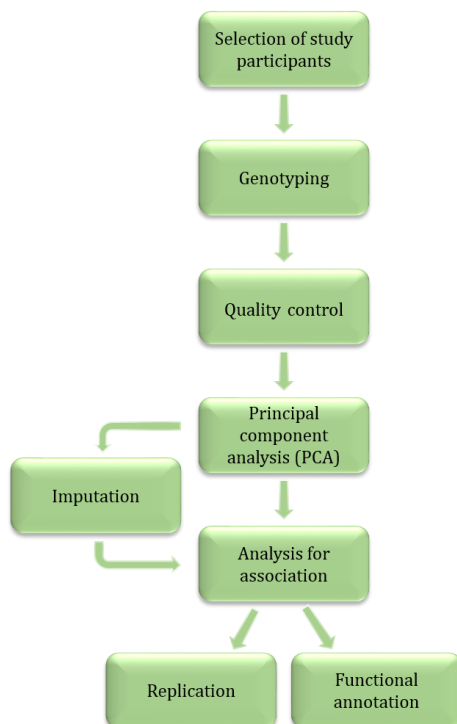


Figure 4. Overview of the GWAS process.

3.1 Selection of study participants

The GWAS was based on a case-control design with 173 patients (cases) from the SWEDEGENE project and genome-wide data from 4891 unrelated population controls from the Swedish Twin Registry. The cases had experienced drug-induced angioedema from either an ACEi or an ARB between 1999 and 2014. The patients had been recruited predominantly through cases reported to the Swedish Medical Products Agency. Patients were interviewed and, if necessary, their medical records were obtained. Three tubes of blood were drawn for each patient and stored at -70°C until analysis. All cases had been reviewed and adjudicated by a clinical expert (allergist) to exclude patients whose angioedema was thought of not being drug-induced, for instance angioedema associated with urticaria. For demographic data and clinical characteristics of the cases, see Table 1. Since we had access to the controls' medical records with drug prescriptions and diagnoses, we could select controls that had been treated with the same drugs as the cases but not experienced drug-induced angioedema. The association analysis was done with all controls as well as with just treated controls.

Table 1. Demographic data and clinical characteristics of the cases.

Cases (173)		
Age	65.6	[31, 91]
Sex		
Female	72	41.6 %
Male	101	58.4 %
Drug		
ACEi	144	83.2 %
ARB	31	17.9 %

3.2 Genotyping

SWEDEGENE patients were genotyped with the Illumina HumanOmni2.5 BeadChip which has 2,338,671 variants from the *1000 Genomes Project* with a MAF>2.5%. The controls had been genotyped with the Illumina HumanOmniExpress BeadChip 700K which has >713,014 variants. The genotyping was performed in different batches at the SNP&SEQ Technology Platform of Science for Life Laboratory (SciLifeLab) in Uppsala. SNP calling was performed by the same platform.

3.3 Quality control

To exclude false-positive signals of association resulting from unchecked systematic errors, the genotypic data obtained was cleaned using a quality control procedure which was performed using the software PLINK.

3.3.1 Minor allele frequency

SNPs with low MAF, $<1\%$, were removed since too rare SNPs are difficult to measure reliably (Pearson & Manolio 2008).

3.3.2 Missing rate per SNP

Individual SNPs were checked for missing genotypes. Only SNPs with a 98% genotyping rate or higher were included, i.e. SNPs with missing rate over 2 % were removed, which is a commonly used threshold (Pare 2010).

3.3.3 Missing rate per person

Individual samples were examined for completeness of genotyping. This was done by calculating the fraction of SNPs for which no allele could be called for each sample. Individuals with too much missing genotype data, over 2% missing genotypes, were excluded, since missing genotypes can reflect poor DNA quality due to, for instance, contamination (Pare 2010).

3.3.4 Hardy-Weinberg Equilibrium

Individual SNPs that deviated from the Hardy-Weinberg equilibrium (p-value cut-off at 10^{-6}) were excluded. SNPs that fail the Hardy-Weinberg test are excluded since a deviation could reflect selection, cryptic relatedness, a mixture of heterogeneous populations or genotyping errors. An exact test was performed on the controls only.

3.3.5 Summary of quality control steps performed

SWEDEGENE cases

- SNPs with a minor allele frequency (MAF) $< 1\%$ were removed
- SNPs with a missing rate $> 2\%$ were removed
- Individuals with a missing rate $> 2\%$ of markers were removed

Twingene controls

- SNPs with a minor allele frequency (MAF) $< 1\%$ were removed
- SNPs with a missing rate $> 2\%$ were removed
- Individuals with a missing rate $> 2\%$ of markers were removed
- SNPs with Hardy-Weinberg $p < 10^{-6}$ were removed

Merged data (cases+controls)

- SNPs with a missing rate $> 2\%$ were removed

Imputed data

- SNPs with a minor allele frequency (MAF) $< 1\%$ were removed
- SNPs with a missing rate $> 2\%$ were removed
- Individuals with a missing rate $> 2\%$ of markers were removed

3.4 Principal component analysis

The genotype data was pruned based on pairwise LD using a simple pairwise threshold. The window size was set to 100 SNPs, step to 5 and the r^2 threshold (pairwise SNP-SNP metric) to 0.2. Since LD is calculated between each pair of SNPs in the window, one of a pair of SNPs was therefore removed if the LD was greater than 0.2. The window was shifted 5 SNPs forward before the procedure was repeated. Principal components were calculated on the pruned data and later included as covariates in the logistic regression analysis. The pruning and principal component analysis were performed using the software PLINK.

The first two principal components were plotted to view the homogeneity of the samples and the SWEDEGENE data was compared with following populations from HapMap (The International HapMap 3 Consortium 2010):

- Yoruba in Ibadan, Nigeria (YRI)
- Japanese in Tokyo, Japan (JPT)
- Han Chinese in Beijing, China (CHB)
- CEPH, Utah residents with ancestry from northern and western Europe (CEU)

3.5 Imputation

Imputation had been performed previously using the software Impute2.

3.6 Analysis for association

Analysis for association was done with logistic regression aiming to identify SNPs where one allele was significantly more common in cases than in controls and might be associated with drug-induced angioedema.

3.6.1 Logistic regression

Logistic regression was used for performing the analysis for association and was implemented using the software package PLINK. Logistic regression is a flexible analysis for GWAS that is used for binary outcomes, such as angioedema (case) or no angioedema (control) in this study. The logistic regression was performed with a dependent variable of case-control status (1=case, 0=control) and a SNP genotype as an independent variable (measured as the number of copies of the minor allele, taking the value 0 for the common homozygote, 1 for the heterozygote and 2 for the rare homozygote) (Barrett *et al.* 2014). Multiple covariates, such as gender and the first principal components was also incorporated. For each SNP, the output from the logistic regression included odds ratio for the minor allele and the corresponding confidence interval as well as the p-value for association.

Studying 10 million SNPs requires that 10 million statistical tests are performed. The probability that some of the SNPs would reach the standard significance level of 0.05 by

chance could not be neglected. Therefore, this extreme number of statistical tests must be corrected for. In this study, *Bonferroni correction*, a simplified Šidák correction, was used to correct for multiple testing and reduce the false-positive rate. It is the conventional p-value (0.05) divided by the number of tests performed and was 8.7×10^{-8} for the genotyped data and 6.6×10^{-9} for the imputed data. 5×10^{-8} , which correspond to a correction for 1 million tests is usually considered a threshold for genome-wide significance in GWAS. The Bonferroni correction is very conservative, since it uses the actual number of SNPs that has been tested for and thereby assumes that each SNP is independently associated with the trait even though it is known that SNPs are correlated to some degree by linkage disequilibrium.

3.6.2 Sub Analysis

A sub analysis of the top association was performed with logistic regression using all controls as well as using 1106 treated controls, i.e. controls that have been treated with the same drugs as the cases but have not experienced drug-induced angioedema. The treated controls had been treated with 1 or more ACEi. In the sub analysis no model assumption was done, in contrast to the GWAS where the data was fitted to an additive model where each copy of the allele associated with ADR is assumed to increase the risk of ADR by the same amount. The results of an additive were also included in the sub analysis for comparison. The strength of an association was measured by the odds ratio (OR).

3.6.3 Visualisation

Results were presented with Q-Q plots and Manhattan plots generated with R, tables of top associations, region association plots generated from LocusZoom and LD-plots generated from HaploView.

A *quantile-quantile plot* (Q-Q plot) was used for assessing the inflation in low P-values from the statistical test. It was used to compare observed association test statistics with their expected values under the null hypothesis of no association. Most of the SNPs should follow the null distribution, which indicates that no population structure, or any other factors that dramatically influences the statistical test exists in the analysis. A deviation from the null distribution in the upper tails, on the other hand, would correspond to SNPs with strong association to the trait.

The result from the association study were visualised in so called *Manhattan plots* where P-values for all studied SNPs are plotted on a log scale against chromosomal position. The height corresponds to the strength of association to the trait in question and the peaks enable initial identification of significant regions. The Bonferroni correction was illustrated as a line in the Manhattan plots. A *region association plot* was used to view a smaller region of the genome and highlight the statistical strength of an associated SNP. Association results for surrounding SNPs as well as gene annotations, recombination rates that reflected the LD structure around the associated variant and pairwise correlations between the associated SNP and the nearby variants were visualised. A *LD-plot* was used for showing LD patterns,

chromosomal location and haplotype blocks of top SNPs, where the LD between SNPs was measured as r^2 (Clarke *et al.* 2011).

3.7 Functional annotation

Assessment of the biological impact of the top genetic variants associated with angioedema was done using public available bioinformatics resources described below.

3.7.1 ENCODE

Most biochemical functions of the genome are determined by the ENCODE-project (The Encyclopedia of DNA Elements) (The ENCODE Project Consortium 2012), which systematically maps regions of transcription, transcription factor occupancy, chromatin accessibility and histone modification (The ENCODE Project Consortium 2012). The different assays used in the ENCODE project (ChIP-seq, DNase-seq, RNA-seq, CAGE, assays of methylation etc.) complement each other and can together shed light on regions with regulatory potential. DNase hypersensitivity sites were studied using DNaseI Hypersensitivity Clusters in 125 cell types dataset with data from the University of Washington and Duke ENCODE groups. Levels of enrichment of histone marks were studied with data from the Bernstein Lab at the Broad Institute. Tracks showing peaks of TF binding, where 161 transcription factors had been assayed by ChIP-seq in different cell lines and with specific transcription-factor targeting antibodies, were used. The dataset was generated by five ENCODE TFBS ChIP-seq production groups (Broad, Stanford/Yale/UC-Davis/Harvard, HudsonAlpha Institute, University of Texas-Austin, University of Washington and University of Chicago).

3.7.2 Roadmap

Regulatory chromatin states from DNase and histone ChIP-Seq from the 2015 Roadmap Epigenomics Consortium, present in Haploreg version 4 (Update 2015.09.15) were used.

3.7.3 USCS browser

Almost all the public bioinformatics resources used in this project could be visualised in the UCSC Browser. It contains reference sequences and working draft assemblies for genomes and is integrated with a large set of aligned annotations (Mortlock & Pregizer 2012). The human genome assembly GRCh37/hg19 (Feb. 2009) and GRCh38/hg38 (Mar. 2013) were used.

3.7.4 HaploReg

HaploReg was also used in this project, which is a tool for annotation of non-coding variants. HaploReg uses information of linkage disequilibrium from the 1000 Genomes Project to view linked SNPs (Ward & Kellis 2012). Predicted chromatin state from the Roadmap Epigenomics project and the ENCODE project can be visualized as well as sequence conservation across mammals and the SNPs' effect on regulatory motifs (Ward & Kellis 2012, Kheradpour & Kellis 2014). Proteins bound in ChIP-Seq experiments from the

ENCODE Project Consortium (2012) can also be visualized as well as the effect of SNPs on expression from eQTL studies (Ward & Kellis 2012). Two versions of HaploReg were used: Version 3 (Update 2014.02.14) and version 4 (Update 2015.09.15).

4 Results

The results are divided into four parts, where the first part contains the results of the association analysis for genotyped data and part II the same analyses but for imputed data. A sub analysis of the top SNP is done in part III and finally a functional annotation of the top genotyped and imputed SNPs is done in part IV.

4.1 Part I: Genotyped data

4.1.1 Original genotyped data

The number of variants per chromosome that was genotyped for cases is visualised in [Figure 5](#).

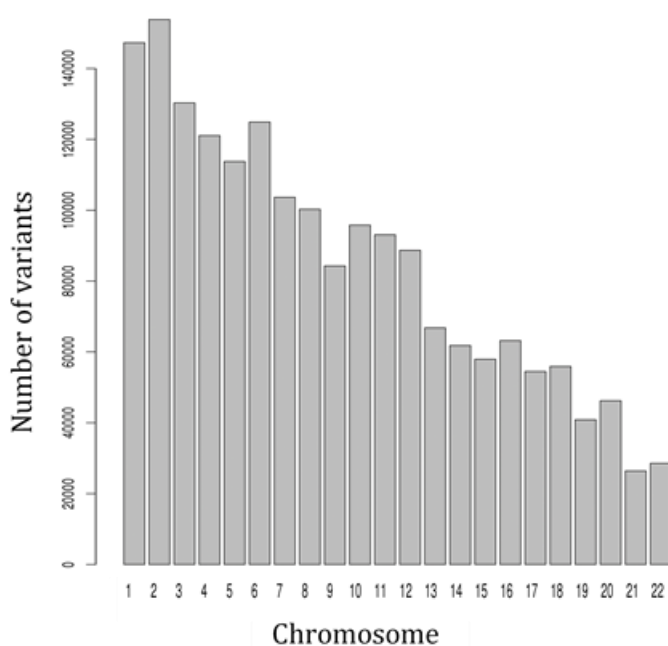


Figure 5. Number of variants for cases (Swedegene patients) per chromosome, in total 1,859,099 variants in 173 people.

4.1.2 Post-QC results

Cases

1,859,099 variants in 173 people before QC.

1,387,738 variants in 173 people passed filters and QC.

Controls

622,992 variants in 4891 people before QC.

620,252 variants in 4890 people passed filters and QC.

Merged data (cases+controls)

1,434,086 variants in 5063 people before QC (of these, 573,904 variants were present in both cases and controls).

573,904 variants in 5063 people passed filters and QC.

4.1.3 Population stratification

The SWEDEGENE cases and controls formed a distinct cluster (red) when compared with populations from the HapMap project (Figure 6). No obvious population stratification seemed to be present in the SWEDEGENE data set. However, there was one extreme outlier, close to the CHB and JPT clusters, but this single individual did probably not affect the results of the association study. SWEDEGENE cases and controls can be seen separately in Figure 7.

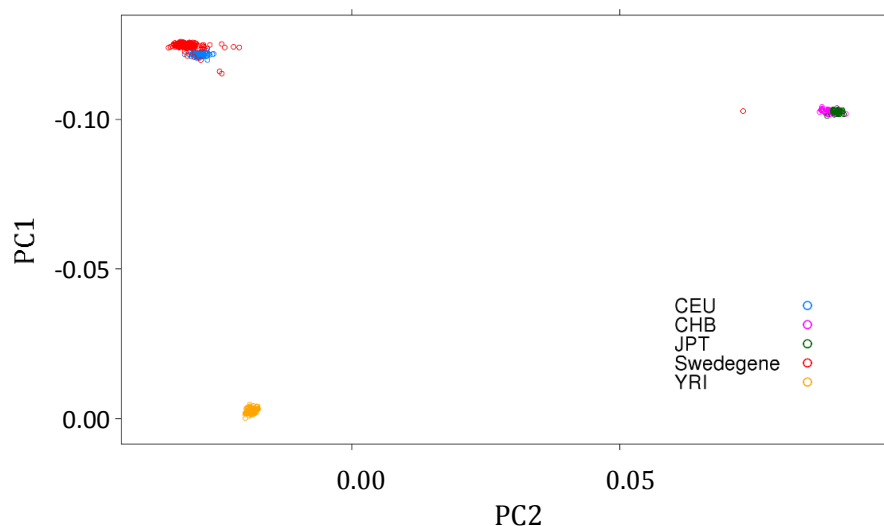


Figure 6. Plot of the first two genetic principal components for SWEDEGENE and HapMap data. Each circle corresponds to a single individual. The cases and controls from the SWEDEGENE data form a distinct cluster, except one outlier, and no obvious population stratification seem to be present in the data set.

CEU=Utah residents with ancestry from northern and western Europe, CHB=Han Chinese in Beijing, JPT=Japanese in Beijing and YRI=Yoruba in Ibadan, Nigeria.

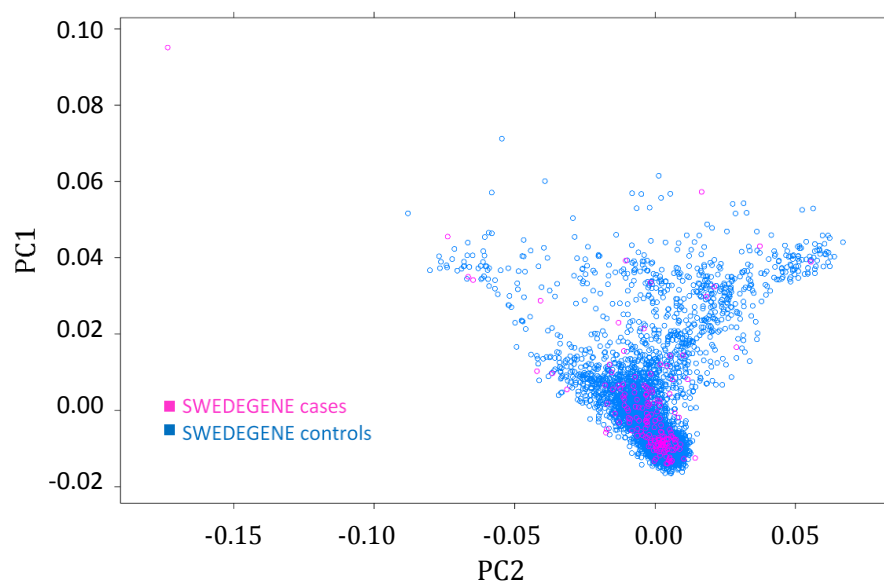


Figure 7. Plot of the first two genetic principal components for the SWEDEGENE data. Each circle correspond to a single individual. Pink=cases, blue=controls.

4.1.4 Analysis of genotyped data

The analysis of the genotyped data was first done by adjusting for gender and principal components 1-4. A second analysis was then performed where the data was adjusted also for the top hit from the first analysis.

4.1.4.1 Adjusted by gender and principal components.

The test statistics for most of the SNPs followed the null distribution and the lambda values were close to 1, which indicated that most of the population stratification was adjusted for in the analysis (Figure 8). On the other hand, a deviation from the null distribution could be seen in the upper tail, likely to correspond to SNPs with strong association to drug-induced angioedema (Figure 8). The result from the association analysis can be visualised in Figure 9, where P-values for all studied SNPs were plotted on a log scale against chromosomal position in a so-called *Manhattan plot*. A clear peak at chromosome 10 could be seen and three SNPs met the genome-wide significance threshold of 8.7×10^{-8} (Table 2) and were thus likely to be associated with drug-induced angioedema. All top SNPs on chromosome 10 were intronic variants in the *KCNMA1* gene (*Potassium Large Conductance Calcium-Activated Channel, Subfamily M, Alpha Member 1*) and they were in high LD, see Figure 10 and 11.

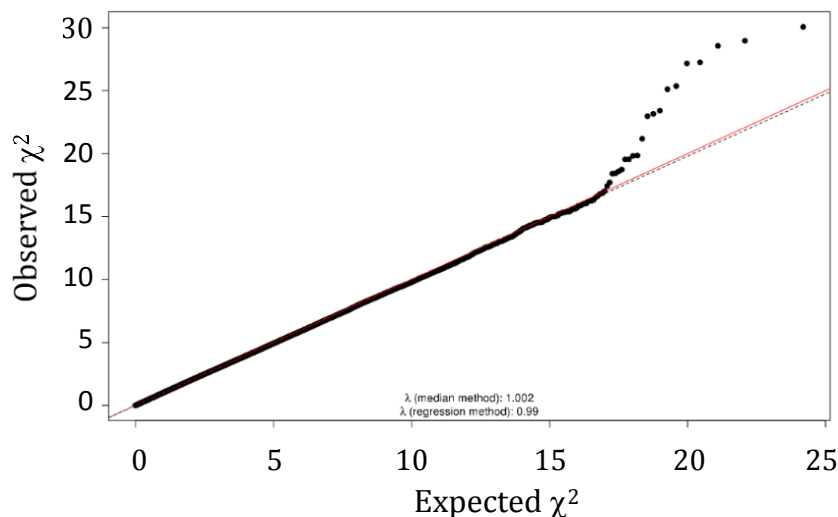


Figure 8. Q-Q plot, adjusted by gender and genetic principal components 1-4. The majority of the SNPs follow the null distribution and the lambda values are close to 1, which indicates that no population stratification is unadjusted for in the analysis. A deviation from the null distribution can be seen in the upper tail, likely to be SNPs with strong association to drug-induced angioedema.

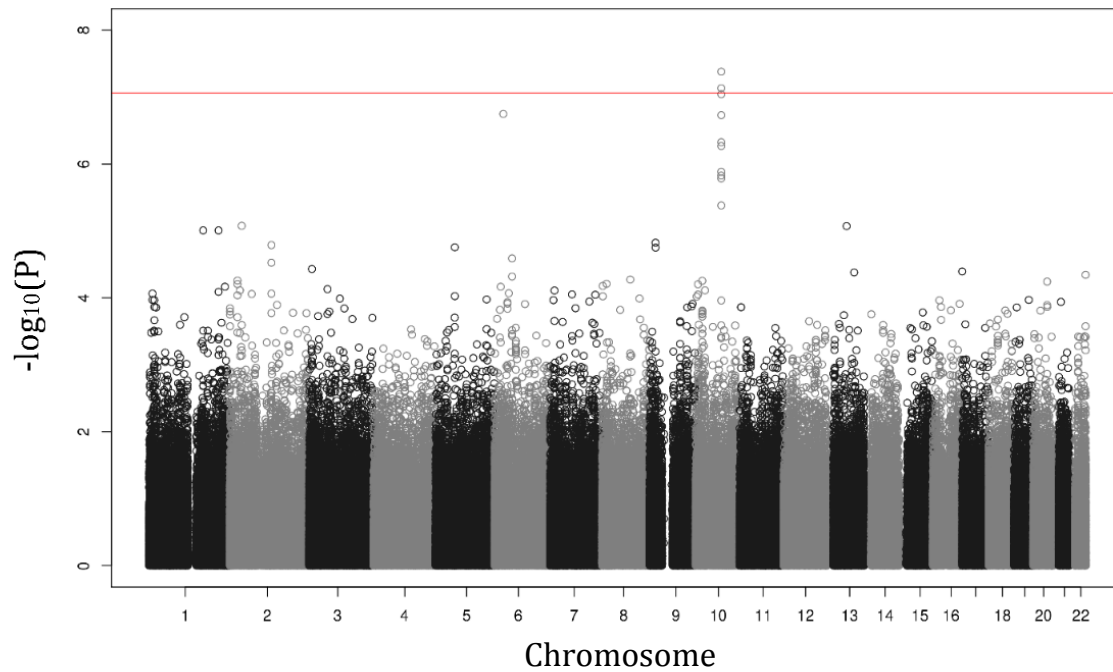


Figure 9. Manhattan plot adjusted by gender and principal components 1-4. The red line corresponds to the genome-wide significance threshold 8.7×10^{-8} (Bonferroni correction). A clear peak at chromosome 10 can be seen, with three SNPs above the red line.

Table 2. Table of genotyped top SNPs. Three SNPs meet the genome-wide significance threshold with a p-value below 8.7×10^{-8} . All top SNPs on chromosome 10 have a low MAF (minor allele frequency) and also a clear difference in MAF between cases and controls.

CHR	SNP	BP	N	OR	L95	U95	P	MAF	GTPS	MAF case	MAF control
10	rs2253201	79356397	5062	2.474	1.79	3.419	4.174e-08	0.063	G/A	0.136	0.06
10	rs1949352	79249522	5062	2.399	1.744	3.299	7.372e-08	0.068	C/T	0.142	0.065
10	rs2253202	79356393	5061	2.433	1.756	3.37	9.077e-08	0.063	G/A	0.134	0.06
6	rs6913724	27254843	5062	1.804	1.446	2.252	1.788e-07	0.432	A/T	0.572	0.428
10	rs2255649	79343812	5063	2.387	1.721	3.312	1.865e-07	0.064	C/T	0.133	0.061
10	rs1464111	79342926	5063	2.332	1.678	3.243	4.732e-07	0.063	T/C	0.13	0.061
10	rs2670164	79327231	5057	2.342	1.679	3.266	5.392e-07	0.063	A/G	0.129	0.06

Abbreviations: CHR= chromosome, BP=base pair (chromosome position), N=number of study participants, OR= odds ratio, L95 and U95 = lower and upper confidence interval, P= p-value, MAF= Minor allele frequency, GTPS=minor/major alleles.

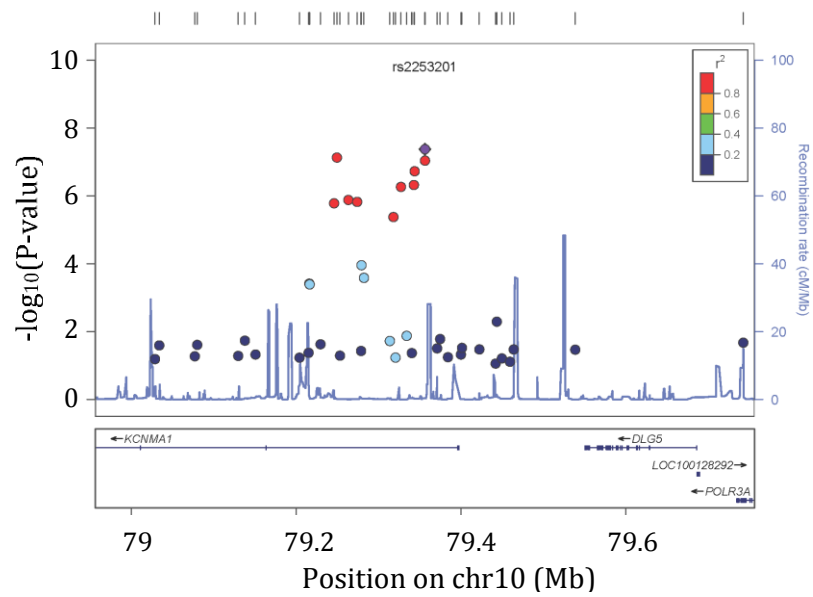


Figure 10. Region plot of top associations generated from LocusZoom with rs2253201, shown in purple, as reference. All top SNPs on chromosome 10 are intronic variants in the KCNMA1 gene.

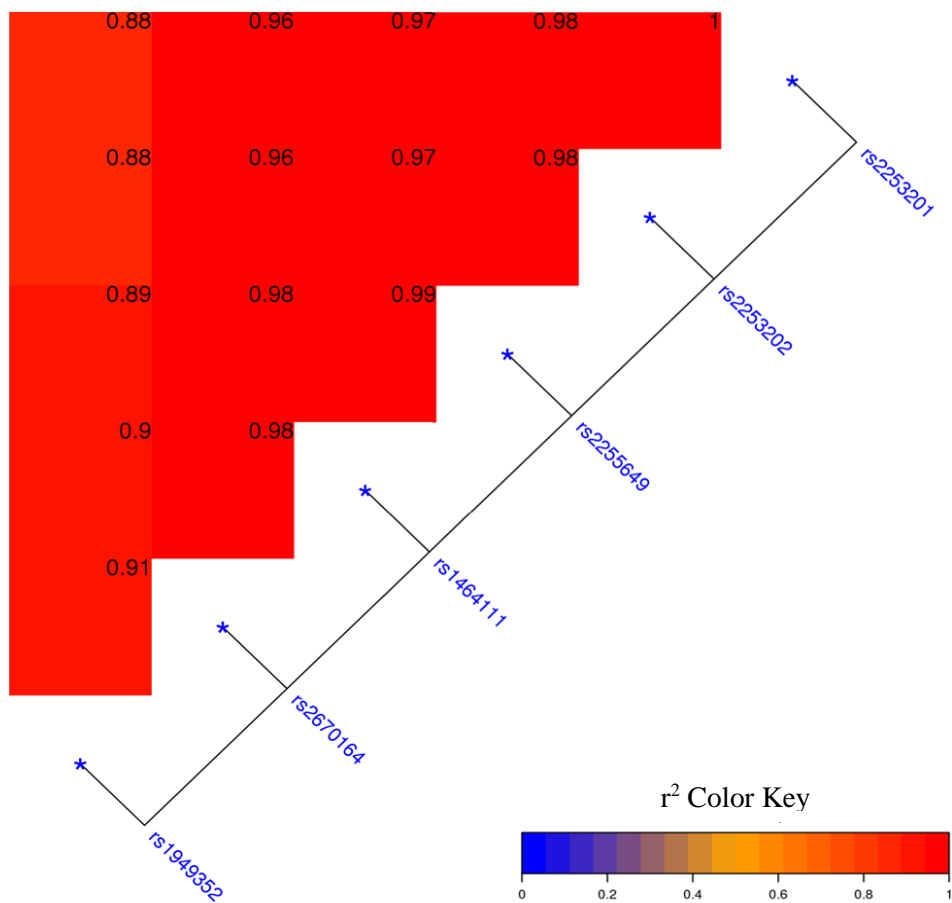


Figure 11. LD heat map showing LD patterns among the top SNPs on chromosome 10. The pairwise LD is measured as r^2 and is shown at the intersection of the diagonals from each SNP. $r^2 = 0$ is shown in blue, $0 < r^2 < 1$ in orange and $r^2 = 1$ is shown in red.

4.1.4.2 Adjusted by top association

When the top hit rs2253201 was adjusted for in the logistic regression, no deviation from the null distribution could be seen in the upper tail in the Q-Q plot (Figure 12) and no residual SNPs were associated with drug-induced angioedema (Figure 13). This indicated that all the top SNPs were in high linkage disequilibrium.

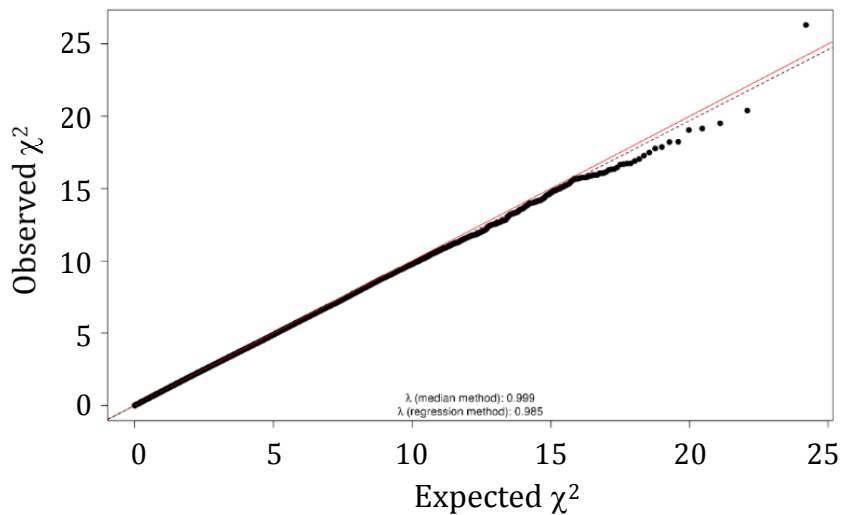


Figure 12. Q-Q plot, adjusted by gender, genetic principal components 1-4 and top SNP rs2253201.

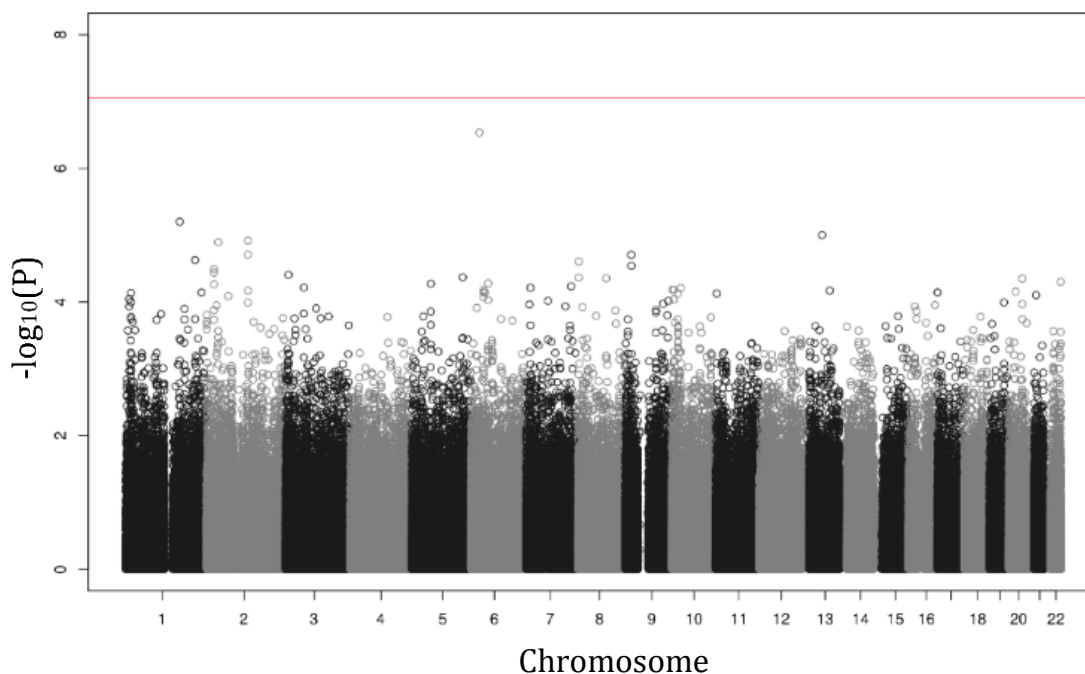


Figure 13. Manhattan plot, adjusted by gender, principal components 1-4 and top SNP rs2253201.

4.2 Part II: Imputed data

4.2.1 Original imputed data

The number of variants per chromosome after imputation for the merged data (cases and controls) can be visualised in Figure 14.

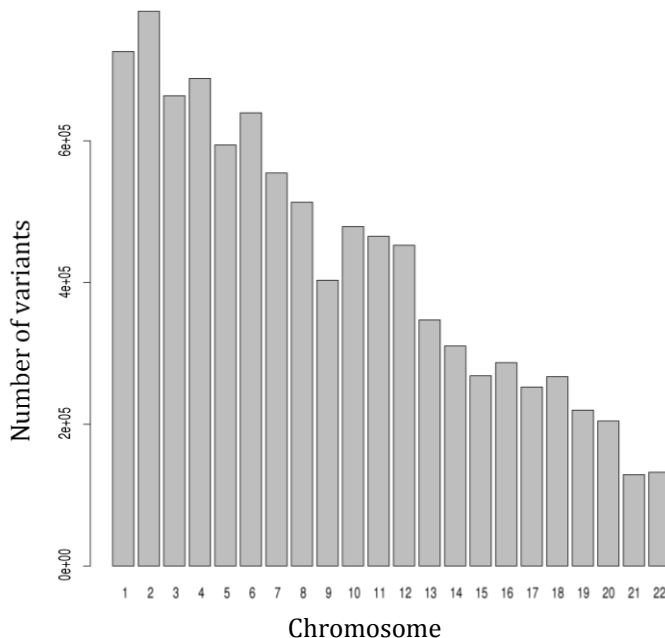


Figure 14. Number of variants per chromosome for imputed data, in total 9,380,034 variants in 5064 people.

4.2.2 Post-QC results

Imputed data

9,380,034 variants in 5064 people before QC.

7,585,599 variants in 5064 people passed filters and QC.

4.2.3 Analysis of imputed data

The same analyses made for the genotyped data was performed for the imputed data with a first analysis where data was adjusted by gender and principal components 1-4, and a second analysis where the data was adjusted for the top hit.

4.2.3.1 Adjusted by gender and principal components.

Most of the SNPs followed the null distribution and the lambda values are close to 1, which indicated that no population stratification was unadjusted for in the analysis (Figure 15). On the other hand, a deviation from the null distribution could be seen in the upper tail, likely to correspond to SNPs with strong association to drug-induced angioedema (Figure 15). The result from the association analysis can be visualised in Figure 16. A clear peak at chromosome 10 could still be seen, indicating that there was an association to drug-induced angioedema. However, no SNP met the genome-wide significance threshold of 6.6×10^{-9} (Bonferroni correction), which was much lower than for genotyped data because of the increased

number of statistical tests performed for the imputed data. Still many SNPs had very low p-values (Table 3). All top SNPs were intronic variants in the KCNMA1 gene.

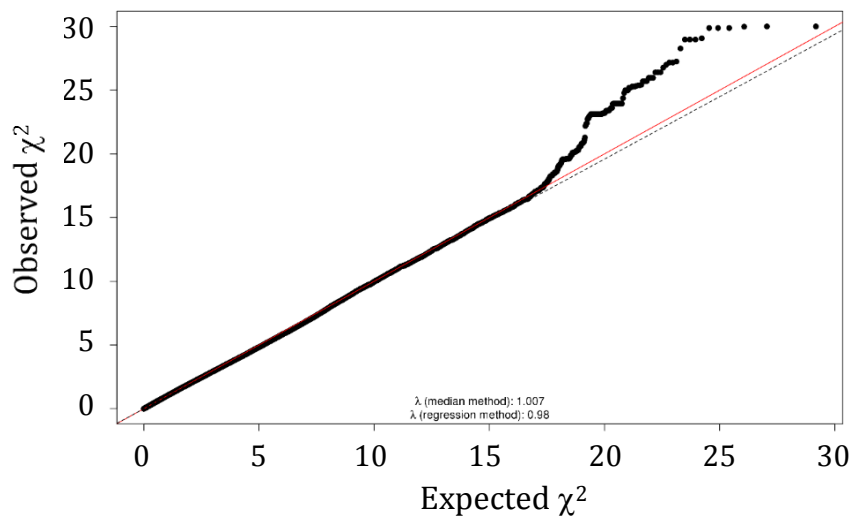


Figure 15. Q-Q plot adjusted by sex and genetic principal components 1-4. Most of the SNPs follow the null distribution and the lambda values are close to 1, which indicates that no population stratification is unadjusted for in the analysis. A deviation from the null distribution can be seen in the upper tail, likely to be SNPs with strong association to drug-induced angioedema.

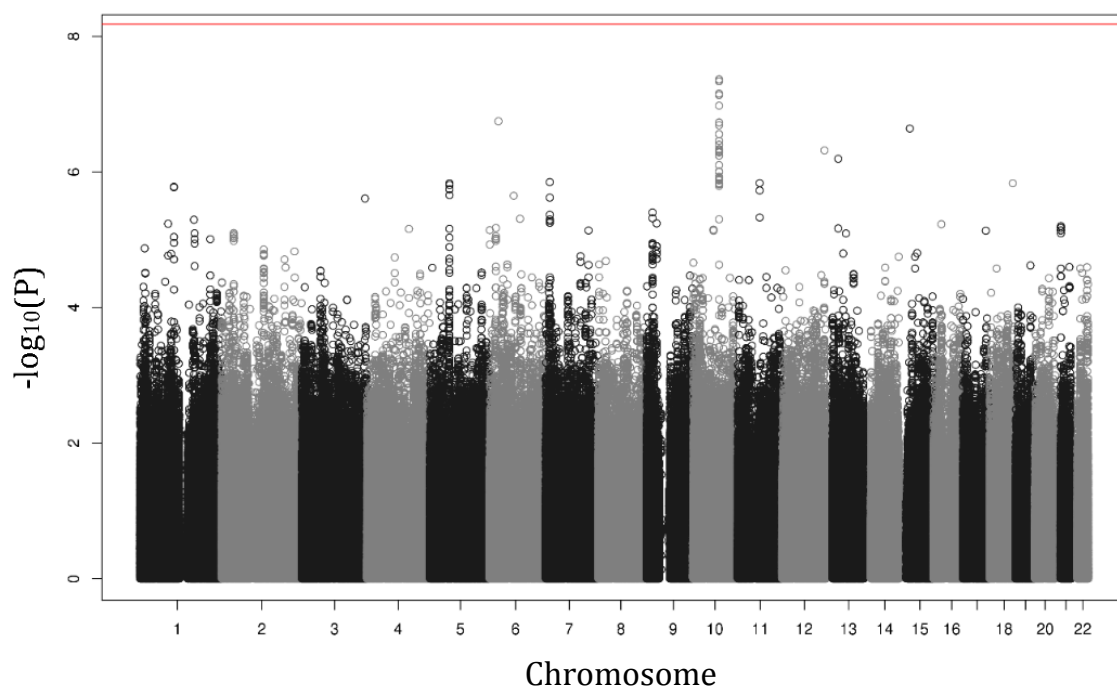


Figure 16. Manhattan plot adjusted by gender and principal components 1-4. The red line corresponds to the genome-wide significance threshold 6.6×10^{-9} (Bonferroni correction). A clear peak at chromosome 10 can be seen, but no SNP is above the red line.

Table 3. Table of top SNPs after imputation. No SNP meet the genome-wide significance threshold (p-value below 6.6×10^{-9}), even though all top SNPs have a very low p-value. All top SNPs are on chromosome 10, have a low MAF (minor allele frequency) and also a clear difference in MAF between cases and controls.

CHR	SNP	BP	N	OR	L95	U95	P	MAF	GTPS	MAF case	MAF control
10	rs2253201	79356397	5063	2.471	1.788	3.416	4.314E-08	0.063	G/A	0.136	0.06
10	rs2253202	79356393	5063	2.471	1.788	3.416	4.314E-08	0.063	G/A	0.136	0.06
10	rs2673471	79357323	5062	2.471	1.787	3.415	4.35E-08	0.063	A/G	0.136	0.06
10	rs2619635	79358602	5063	2.467	1.785	3.41	4.592E-08	0.063	G/A	0.136	0.06
10	rs2670121	79358889	5063	2.467	1.785	3.41	4.592E-08	0.063	A/G	0.136	0.06
10	rs2673455	79359111	5063	2.467	1.785	3.41	4.592E-08	0.063	C/G	0.136	0.06
10	rs865293	79258692	5062	2.402	1.747	3.304	6.979E-08	0.068	G/A	0.142	0.065
10	rs1949352	79249522	5063	2.399	1.744	3.299	7.331E-08	0.068	C/T	0.142	0.065
10	rs816847	79250577	5063	2.399	1.744	3.299	7.331E-08	0.068	G/C	0.142	0.065
10	rs866539	79249804	5063	2.399	1.744	3.299	7.331E-08	0.068	C/T	0.142	0.065

Abbreviations: CHR= chromosome, BP=base pair (chromosome position), N=number of study participants, OR= odds ratio, L95 and U95 = lower and upper confidence interval, P= p-value, MAF= Minor allele frequency, GTPS=minor/major alleles.

4.2.3.2 Adjusted by top association

When the top hit rs2253201 was adjusted for in the logistic regression, no deviation from the null distribution could be seen in the upper tail in the Q-Q plot (Figure 17), and no residual SNPs were associated with drug-induced angioedema (Figure 18). This indicated that all the top SNPs after imputation were in high linkage disequilibrium.

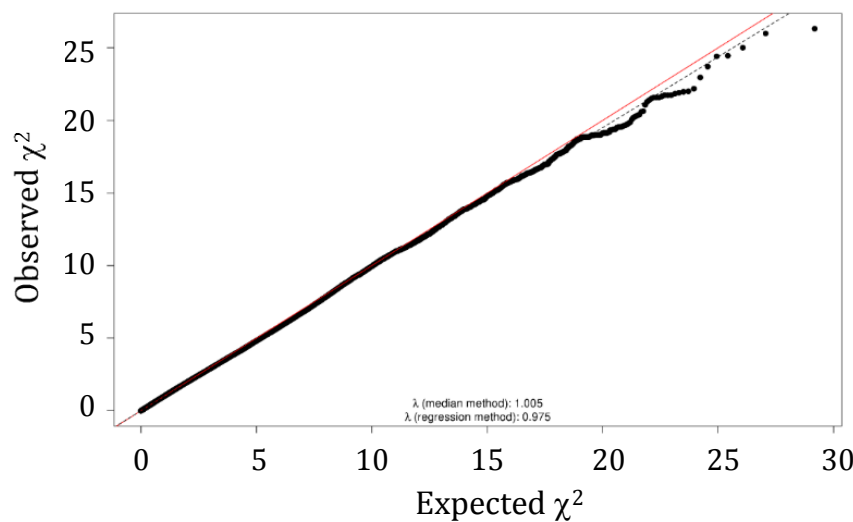


Figure 17. Q-Q plot adjusted by sex and genetic principal components 1-4 and top hit rs2253201.

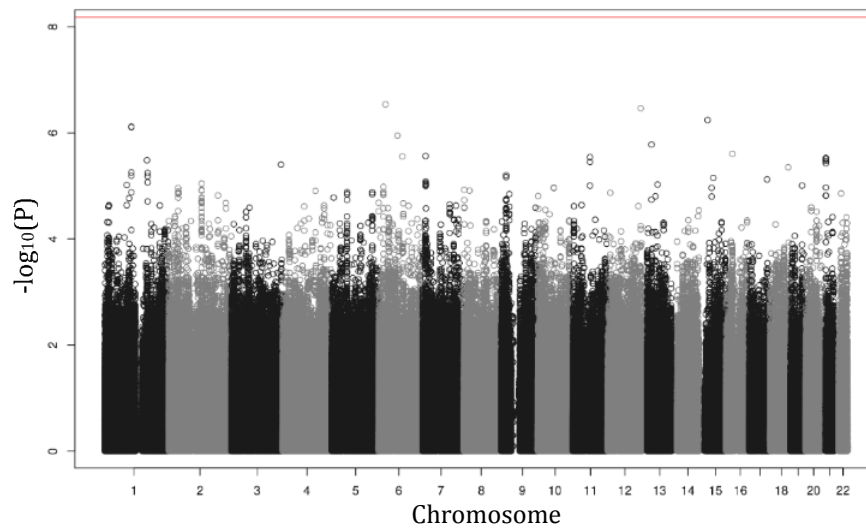


Figure 18. Manhattan plot adjusted by sex and genetic principal components 1-4 and top hit rs2253201.

4.3 Part III: Analysis of top SNP

The analysis of the top association rs2253201 revealed an association between homozygous G/G carriers of SNP rs2253201 and drug-induced angioedema, with an odds ratio (OR) of 16.2, when using all controls, and 12.7 when only treated controls were included (Figure 19). No model assumption was done for these. An additive model (as used in the GWAS) was also included resulting in an OR of 2.45 for one copy of the G-allele for all controls and 2.20 when using treated controls. No model assumption and the additive model resulted in similar results. Since all top SNPs were in high LD, only one SNP (rs2253201) was studied further with treated controls.

The distribution of SNP rs2253201 genotype, seen in Table 4 and 5, revealed a clear difference in minor allele frequency (i.e. frequency of allele G) between cases and controls. 23.1 % of the cases carried one or two copies of the G allele whereas 11.8 % of the controls and 12.6 % of treated controls carried the G allele.

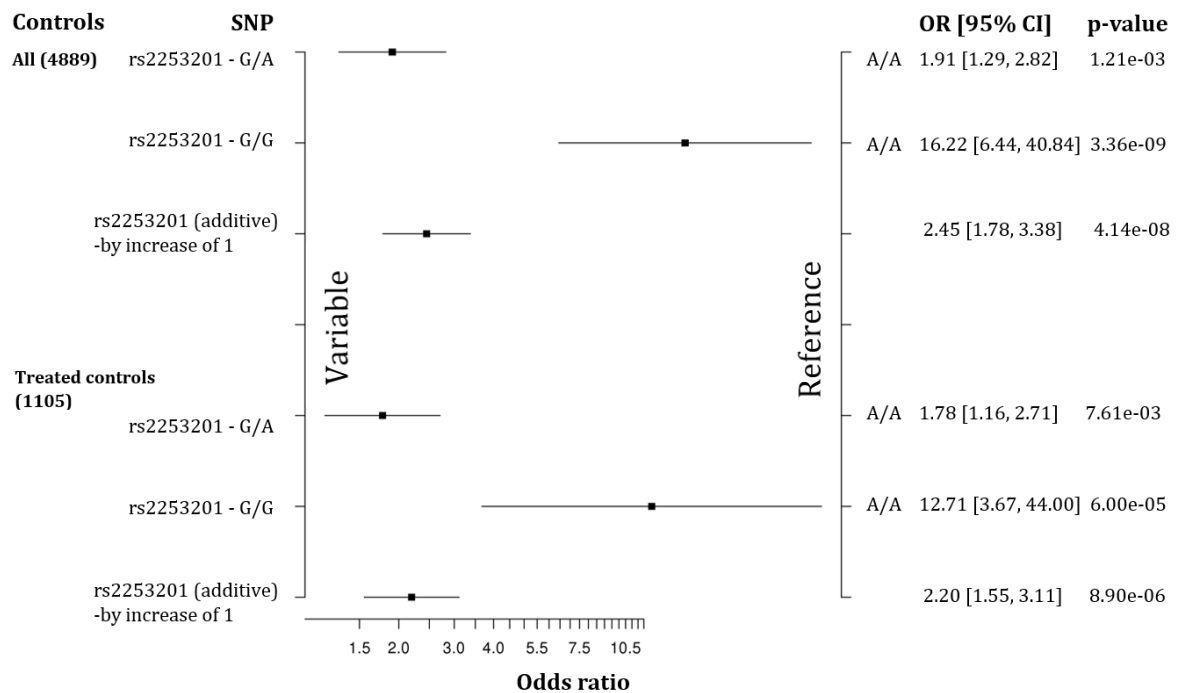


Figure 19. Forest plot. There is an association between homozygous G/G carriers of SNP rs2253201 and drug-induced angioedema, with an odds ratio (OR) of 16.2, when using all controls, and 12.7 when only treated controls were included. No model assumption was made for these analyses. The additive model revealed an association with an odds ratio of 2.45 for each copy of the G-allele for all controls and 2.20 for treated controls.

Table 4. Distribution of SNP rs2253201 genotype among cases and 4889 controls.

	A/A	G/A	G/G	MAF
Controls (4889)	4314 (88.2%)	561 (11.5%)	14 (0.3%)	0.060
Cases (173)	133 (76.9%)	33 (19.1%)	7 (4.0%)	0.136

Table 5. Distribution of SNP rs2253201 genotype among cases and 1105 treated controls.

	A/A	G/A	G/G	MAF
Treated controls (1105)	966 (87.4%)	135 (12.2%)	4 (0.4%)	0.065
Cases (173)	133 (76.9%)	33 (19.1%)	7 (4.0%)	0.136

4.4 Part IV: Regulatory potential of top SNPs

This part contains an assessment of the biological impact of the top genetic variants associated with angioedema. All the identified top SNPs were intronic and located in the beginning of the *KCNMA1* gene (*Potassium Large Conductance Calcium-Activated Channel, Subfamily M, Alpha Member 1*), see Figure 10.

The functional annotation was based on data from the ENCODE and Roadmap projects available in the UCSC and Haploreg browsers and included chromatin states from DNase and histone ChIP-Seq, transcription factors bound in ChIP-seq experiments and conservation for all top SNPs. Table 6 contains the genotyped top SNPs while Table 7 contains top SNPs after imputation. SNPs already described in Table 6 were excluded in Table 7. The tables only include data for cell lines thought to be relevant for drug-induced angioedema. These cell lines included immunological and blood cell lines, smooth muscle cell lines, heart and digestive cell lines. No data for endothelial cell lines (large-vessel or microvascular), which would have been relevant, was available. The tables only show chromatin states in groups of cell lines. For specific cell lines, see Appendices.

4.4.1 Chromatin state

All top SNPs lied within promoter and enhancer histone mark sites in relevant cell lines. Four SNPs (rs1949352, rs2670164, rs2670121 and rs816847) were located in DNase hypersensitive sites in either blood or digestive cell lines.

4.4.2 Transcription factors

Regulatory motifs altered could be found for almost all SNPs. For SNP rs1464111 six transcription factors (POLR2A, GATA2, TEAD4, MAZ, RCOR1 and TAL1) bound in ChIP-seq experiments in the K562 and NB4 cell lines, and for SNP rs865293, GR bound in the A549 cell line. Worth mentioning is that GR was found also for SNP rs1464111 and rs2673471. GR is a glucocorticoid receptor that binds cortisol and other glucocorticoids and controls immune responses.

Table 6. Functional annotation for genotyped top SNPs. Chromatin states comes from Roadmap unless otherwise stated.

		rs2253201	rs1949352	rs2253202	rs2255649	rs1464111	rs2670164
CHROMATIN STATE	Potentially active promoter region <i>H3K4me3</i>	-	-	-	-	-	Smooth muscle, digestive
	Potentially active promoter region <i>H3K9ac</i>	Smooth muscle	Blood (ENCODE)	Smooth muscle	Blood, smooth muscle	Blood + blood (ENCODE)	Smooth muscle
	Potentially active enhancer region <i>H3K4me1</i>	Smooth muscle, heart	Smooth muscle, digestive	Smooth muscle	Blood, Smooth muscle, digestive	Blood, heart, smooth muscle, digestive	Heart, smooth muscle, digestive
	Potentially active enhancer region <i>H3K27ac</i>	Smooth muscle	Smooth muscle, digestive	Smooth muscle	Smooth muscle	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive
	DNase hyper-sensitivity	-	Digestive (small intestine, fetal intestine large)	-	-	-	Blood (ENCODE)
TRANSCRIPTION FACTORS	TF ChIP-seq, Factorbook Motifs (ENCODE)	-	-	-	-	POLR2A (NB4), GATA2 (K562), TEAD4 (K562), MAZ (K562), RCOR1 (K562), TAL1 (K562)	-
	Regulatory motifs altered (HaploReg v.4)	-	HP1-site-factor	-	Pax-2, Pax-8	Ets, GR, Myf, TATA	Myb
CONS.	Conservation GERP blocks (ENCODE)	NO	NO	NO	NO	NO	NO

Abbreviations: POLR2A=DNA-directed RNA polymerase II subunit RPB1, GATA2=GATA binding protein 2, TEAD4=Transcriptional enhancer factor TEF-3, MAZ=Myc-associated zinc finger protein, RCOR1=REST corepressor 1, TAL1= T-cell acute lymphocytic leukemia protein 1, HP1-site-factor=Heterochromatin Protein 1, Pax=Paired box, Ets=E-twenty six, GR=Glucocorticoid receptor, Myf=Myogenic factor, Myb=Myb proto-oncogene protein.

Table 7. Functional annotation for top SNPs after imputation. Top SNPs already presented in Table 8 are excluded in this table. Chromatin states comes from Roadmap unless otherwise stated.

		rs2673471	rs2619635	rs2670121	rs2673455	rs865293	rs816847	rs866539
CHROMATIN STATE	Potentially active promoter region <i>H3K4me3</i>	-	-	-	-	Digestive	Heart	-
	Potentially active promoter region <i>H3K9ac</i>	Smooth muscle	Smooth muscle + blood (ENCODE)	Smooth muscle + blood (ENCODE)	Smooth muscle + blood (ENCODE)	-	Blood (ENCODE)	Blood (ENCODE)
	Potentially active enhancer region <i>H3K4me1</i>	Smooth muscle	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Smooth muscle, digestive
	Potentially active enhancer region <i>H3K27ac</i>	Smooth muscle	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Heart, smooth muscle, digestive	Smooth muscle, digestive	Smooth muscle, digestive	Smooth muscle, digestive
	DNase hyper-sensitivity	-	-	Blood (K562)	-	-	Digestive (gastric)	-
TRANSCRIPTION FACTORS	TF ChIP-seq, Factorbook Motifs (ENCODE)	-	-	-	-	GR (A549)	-	-
	Regulatory motifs altered (HaploReg v.4)	Dobox4, GR	Pax-4, PREB-1, p300	-	Dbx1&2, Hoxa5, Hoxd8, Ncx, Pou1f1, Pou6f1, TATA, Zfp187	Sin3Ak-20_disc1, Spdef	STAT	TFIIA
CONS.	Conservation GERP blocks (ENCODE)	NO	NO	NO	YES	NO	NO	NO

Abbreviations: GR= Glucocorticoid receptor, Pax=Paired box, PREB-1=Prolactin regulatory element binding protein 1, p300=E1A binding protein p300, Dbx= Developing brain homeobox protein 1, Hox=Homeobox protein, Ncx=neural crest homeobox, Pou1f1=POU domain class 1 transcription factor 1, Pou6f1=POU domain class 6 transcription factor 1, Zfp187=zinc finger protein 187, Sin3Ak=SIN3 homolog A (k-20), Spdef= SAM pointed domain containing ETS transcription factor, STAT=Signal Transducer and Activator of Transcription, TFIIA=Transcription factor TFIIA.

5 Discussion

In contrast to the previous GWAS of angioedema induced by ACE inhibitors (Pare *et al.* 2013), several associations of genome-wide significance were found in this study. Three of the genotyped SNPs, all located in the KCNMA1 gene, met the genome-wide threshold with a p-value below 8.7×10^{-8} . KCNMA1 encodes BK (Big Potassium) channels also called Maxi-K channels which have large conductance for potassium (K⁺) ions through cell membranes. These channels are fundamental for regulation of various key physiological processes including contraction of smooth muscle and neuronal excitability (National Center for Biotechnology Information 2010). As known previously, the vasodilation caused by ACEi and ARB, is induced by relaxation of smooth muscle cells in the cell walls which widens blood vessels and decreases blood pressure. Consequently, there might be a connection between the associated gene and the therapeutic effect of the drugs regarding smooth muscle control. In the functional annotation, data in smooth muscle cell lines was therefore of interest along with immunological, blood and digestive cell lines. The result of the functional annotation, summarised in Table 6 and 7, revealed that all top SNPs lie within promoter and enhancer histone mark sites in relevant cell lines. Four SNPs are also located in DNase hypersensitive sites. Almost all alter regulatory motifs and for two SNPs, rs1464111 and rs865293, transcriptions factors were found to bind in ChIP-seq experiments. From this functional information of the SNPs it is hard to draw conclusions of the regulatory potential and identify the causal variant or variants of the ADR. The different features and biochemical signatures are very useful for identifying regulatory elements but cannot work as proof of function on their own. Therefore, additional experimental assays are needed as a complement for studying the function of the associated variants. This could include experimental studies in additional cell lines like endothelial cell lines and more extensive analyses of the transcription factors.

A limitation of this study is the small sample size, which is not unusual for these types of pharmacogenomic studies of rare ADRs. Nevertheless, the number of study participants in this study is comparatively large and genome-wide significance was achieved. All 173 cases were also from the same population and genotyped at the same platform, providing good quality of data which strengthens the study. The reports of adverse drug reactions from MPA as well as the many controls from the Swedish twin registry were important resources for this study and enabled a satisfying study design. The strict criteria for the phenotype, which made it possible to judge the relevance of the angioedema cases for the study and just include truly affected patients in the study, was also an important factor which probably enhanced the power of the study.

The distribution of the top SNP rs2253201 genotype revealed a clear difference in minor allele frequency between cases and controls, where 23.1 % of the cases carried one or two copies of the G allele whereas only 11.8 % of the controls and 12.6 % of treated controls carried the same allele. This means 23% of the cases could be identified by genotyping this

SNP. But what would the clinical utility of genetic testing be? The *Number Needed to Genotype* (NNG), which is based on the *Number Needed to Treat* (NNT), could be used to evaluate this. It is the average number of patients who need to be genotyped to prevent one additional ADR (de Graaff *et al.* 2013). If we assume that 0.7% is affected (i.e. 1 of 142 patients is affected) and that the sensitivity of genotyping of the top SNP allele is 23% it would mean that the number needed to genotype is 617 people ($142/0.23 = 617$), to prevent one case of angioedema (assuming avoidance of the causing drug). Since healthy controls also were carriers of the G-allele, we have to assume that some unnecessarily will avoid the causing drug. Worth discussing is the costs of genotyping around 600 patients to be able to prevent one patient from getting the ADR in contrast to the costs for treating a patient with drug-induced angioedema. Rough assumption of the costs of genotyping one patient is around 500 SEK, resulting in a total cost of around 300 000 SEK (617×500 SEK) for genotyping. The exact costs of treatment of angioedema patients is not known and depends on the severity of the ADR and if the patient need intensive care medicine or not. The assumption that an angioedema patient would need three days of intensive care medicine would result in a cost of around 150 000 SEK. The genotyping costs, in this example, would therefore be higher than the costs of treatment. What is not measurable but worth mentioning is the harm and suffering of patients and their families that could be prevented by individualising medicine. Also, 73 people (617×0.118) would unnecessarily change drug, since 12.6 % of the treated controls carried one or two copies of the G-allele in the top SNP without having the ADR. Worth discussing is also that it is not known whether the controls carrying the risk alleles might get the ADR in the future.

An essential part of a GWAS is to replicate the results in independent samples to confirm positive association signals and exclude possible false-positive results to provide more definitive conclusions of associated SNPs. It was, however, beyond the scope of this thesis. A replication could be done in additional patients collected by SWEDEGENE as well as in patients from collaborating groups, to get a larger cohort. Also, an additional analysis could be done with separated GWAS of ACEi and ARB and see if the association signal changes.

6 Conclusion and outlook

In conclusion, several SNPs in the *KCNMA1* gene were associated with angioedema induced by ACE inhibitors and ARB. However, a replication of the results is needed to confirm these positive association signals before definitive conclusions of the impact of these SNPs can be drawn. Additional analyses of the associated SNPs are also needed to improve the functional annotation and identify the causal SNP. A pathway-based analysis could also be used to go from a set of associated SNPs to a causal biological process.

By identifying possible triggers, this project, along with other pharmacogenomics studies, have the potential of increasing the understanding of adverse drug reactions. Identified genetic

risk markers could be used to develop clinical tests, which in the future could be used to individualise the choice of drug and dose for each patient, i.e. personalised medicine. With these clinical tests, a physician would be able to determine a patient's SNP profile, compare it with known data, and predict how the patient will respond to a drug and then decide which drug to give accordingly. This strategy would improve the treatment for the patients and replace the trial-and-error method of drugs used today. By studying the genetic basis of ADRs associated with certain drugs it might also be possible to identify new molecular targets for treatment of the syndrome itself, when not drug-induced. For instance, by studying cough associated with ACEi-inhibitors it is possible to find new therapeutic targets for treatment of cough. Increasing knowledge of genetic causes of ADRs makes it easier to in an early stage of a drug development processes screen for molecules likely to cause these reactions and thereby design safer drugs.

7 Acknowledgements

Mia Wadelius, Niclas Eriksson
and Pär Hallberg

for their support and guidance as supervisors

Åsa Johansson

*for her useful input as a scientific reviewer and
valuable feedback on the report*

SWEDEGENE

for collecting all the data

Patients and twins

without them – no study!

Simon Lind

for accepting to act as opponent

References

- Barrett J.H., Taylor J.C., and Iles M.M. (2014). Statistical perspectives for genome-wide association studies (GWAS). *Methods Mol Biol*: 1168: 47-61.
- Campbell N.A. and Reece J.B. (2008). *Biology*. 8 ed. Pearson Education.
- Clarke G.M., Anderson C.A. and Pettersson F.H., Cardon L.R., Morris A.P. and Zondervan K.T. (2011). Basic statistical analysis in genetic case-control studies. *Nature Protocols*: 6(2): 121-133.
- Daly A.K. (2010). Genome-wide association studies in pharmacogenomics. *Nat Rev Genet*: 11(4): 241-246.
- Daly A.K. (2013). Pharmacogenomics of adverse drug reactions. *Genome Med*. 5(1): 5.
- de Graaff L.C., van Schaik R.H. and van Gelder T. (2013). A clinical approach to pharmacogenetics. *The Netherlands Journal of Medicine*: 71(3): 145-152
- Edwards I.R. and Aronson J.K. (2000). Adverse drug reactions: definitions, diagnosis, and management. *The Lancet* 356(9237): 1255-1259.
- Eriksson N. (2012). On the Prediction of Warfarin Dose. Diss., Uppsala University.
- Ernst J., Kheradpour P., Mikkelsen T.S., Shores N., Ward L.D., Epstein C.B., Zhang X., Wang L., Issner R., Coyne M., Ku M., Durham T., Kellis M. and Bernstein B.E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*: 473(7345): 43-49.
- Gibson G. and Muse S.V. (2009). *A Primer of Genome Science*. 3 ed. Sinauer Associates.
- Howie B.N., Donnelly P., and Marchini J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*: 5(6): e1000529.
- Kheradpour P. and Kellis M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res*: 42(5): 2976-2987.
- Lazarou J. (1998). Incidence of Adverse Drug Reactions in Hospitalized Patients. *American Medical Association*, 1998: 279.
- Mahmoudpour S.H., Leusink M., van der Putten L., Terreehorst I., Asselbergs F.W., de Boer A. and Maitland-van der Zee A.H. (2013). Pharmacogenetics of ACE inhibitor-induced angioedema and cough: a systematic review and meta-analysis. *Pharmacogenomics*: 14(3): 249-260.

Mancinelli L., Cronin M. and Sadée W. (2000). Pharmacogenomics - the promise of personalized medicine. AAPS PharmSci.

Mortlock D.P. and Pregizer S. (2012). Identifying functional annotation for noncoding genomic sequences. Curr Protoc Hum Genet: Chapter 1: Unit1 10.

Motsinger-Reif A.A., Jorgenson E., Relling M.V., Kroetz D.L., Weinshilboum R., Cox N.J. and Roden D.M. (2013). Genome-wide association studies in pharmacogenomics: successes and lessons. Pharmacogenet Genomics: 23(8): 383-394.

National Center for Biotechnology Information. (2015). Gene database search: KCNMA1 potassium large conductance calcium-activated channel, subfamily M, alpha member 1. <http://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=3778>. Derived 2015-10-01.

Pare G. (2010). Genome-wide association studies--data generation, storage, interpretation, and bioinformatics. J Cardiovasc Transl Res: 3(3): 183-188.

Pare G., Kubo M., Byrd J.B., McCarty C.A., Woodard-Grice A., Teo K.K., Anand S.S., Zuvich R.L., Bradford Y., Ross S., Nakamura Y., Ritchie M. and Brown N.J. (2013). Genetic variants associated with angiotensin-converting enzyme inhibitor-associated angioedema. Pharmacogenet Genomics: 23(9): 470-478.

Pearson T.M. and Manolio T. (2008). How to Interpret a Genome-wide Association Study. JAMA: 299: 1335-1344.

Powers B.J., Coeytaux R.R., Dolor R.J., Hasselblad V., Patel U.D., Yancy W.S. Jr, Gray R.N., Irvine R.J., Kendrick A.S. and Sanders G.D. (2012). Updated report on comparative effectiveness of ACE inhibitors, ARBs, and direct renin inhibitors for patients with essential hypertension: much more data, little new information. J Gen Intern Med: 27(6): 716-29.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. and Reich D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet: 38(8): 904-909.

Ritter J.M., Lewis L., Mant T. and Ferro A. (2008). A Textbook of Clinical Pharmacology and Therapeutics. Hodder Arnold.

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature: 489(7414): 57-74.

The International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. Nature: 467(7311): 52-58.

The National Board of Health and Welfare. (2013). Swedish Prescribed Drug Register.
<http://www.socialstyrelsen.se/statistik/statistikdatabas/lakemedel>. Derived 2015-10-01.

Wadelius M., Marshall S.E., Islander G., Nordang L., Karawajczyk M., Yue Q.Y.,
Terreehorst I., Baranova E.V., Hugosson S., Sköldefors K., Pirmohamed M., Maitland-van
der Zee A.H., Alfrevic A., Hallberg P. and Palmer C.N. (2014). Phenotype standardization of
angioedema in the head and neck region caused by agents acting on the angiotensin system.
Clin Pharmacol Ther, 96(4): 477-481.

Ward L.D. and Kellis M. (2012). HaploReg: a resource for exploring chromatin states,
conservation, and regulatory motif alterations within sets of genetically linked variants.
Nucleic Acids Res: 40(Database issue): D930-4.

Appendix A

Table 8. Cell lines.

SNP	<i>H3K4me3</i>	<i>H3K9ac</i>	<i>H3K4me1</i>	<i>H3K27ac</i>
rs2253201		Smooth muscle: Colon smooth muscle	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle
rs1949352		Blood Monocytes-CD14+ RO01746 Primary Cells (ENCODE)	Smooth muscle: Stomach smooth muscle Digestive: Fetal intestine small & large, gastric	Smooth muscle: Stomach smooth muscle Digestive: Fetal intestine small & large, gastric
rsS2253202		Smooth muscle: Colon Smooth Muscle	Smooth muscle: Colon, rectal and stomach smooth muscle	Smooth muscle: Colon, rectal and stomach smooth muscle
rsS2255649		Blood: Primary mononuclear cells from peripheral blood Smooth muscle: Stomach smooth muscle	Blood: Primary monocytes, B-cells and NK cells from peripheral blood, primary hematopoietic stem cells Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Fetal intestine small & large, stomach & duodenum mucosa, esophagus	Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Esophagus, gastric
rs1464111		Blood: Primary mononuclear cells from peripheral blood, monocytes-CD14+ RO01746 primary cells (ENCODE)	Blood: Primary monocytes and NK cells from peripheral blood, primary hematopoietic stem cells Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Fetal intestine small & large, stomach & duodenum mucosa, esophagus, gastric	Heart: Aorta Smooth muscle: Colon, and stomach smooth muscle Digestive: Fetal stomach, intestine large, esophagus, gastric

SNP	<i>H3K4me3</i>	<i>H3K9ac</i>	<i>H3K4me1</i>	<i>H3K27ac</i>
rs2670164	Smooth muscle: Colon smooth muscle Digestive: Stomach mucosa	Smooth muscle: Colon and rectal smooth muscle Blood: Monocytes-CD14+ RO01746 primary cells (ENCODE)	Blood: Primary monocytes from peripheral blood, primary hematopoietic stem cells, primary NK cells from peripheral blood Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Fetal stomach, colonic & rectal mucosa	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Fetal stomach
rs2673471		Smooth muscle: Colon smooth muscle	Smooth muscle: Colon, rectal and stomach smooth muscle	Smooth muscle: Colon, rectal and stomach smooth muscle
rs2619635		Smooth muscle: Colon and stomach smooth muscle Blood: Monocytes-CD14+ RO01746 primary cells (ENCODE)	Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Stomach mucosa	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Rectal mucosa
rs2670121		Smooth muscle: Colon and stomach smooth muscle Blood: Monocytes-CD14+ RO01746 primary cells (ENCODE)	Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Stomach mucosa	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Rectal mucosa
rs2673455		Smooth muscle: Stomach smooth muscle Blood: Monocytes-CD14+ RO01746 primary cells (ENCODE)	Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Stomach mucosa	Heart: Aorta Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Rectal mucosa

SNP	<i>H3K4me3</i>	<i>H3K9ac</i>	<i>H3K4me1</i>	<i>H3K27ac</i>
rs865293	Digestive: Gastric		Heart: Aorta Smooth muscle: Duodenum, colon, rectal and stomach smooth muscle Digestive: Stomach, duodenum mucosa, esophagus, gastric	Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Esophagus, gastric
rs816847	Heart: Fetal heart	Blood Monocytes-CD14+ RO01746 Primary Cells (ENCODE)	Heart: Aorta, fetal heart, left and right ventricle Smooth muscle: Colon, rectal and stomach smooth muscle Digestive: Fetal intestine small & large, esophagus, gastric	Smooth muscle: Colon, and stomach smooth muscle Digestive: Stomach, gastric
rs866539		Blood Monocytes-CD14+ RO01746 Primary Cells (ENCODE)	Smooth muscle: Stomach smooth muscle Digestive: Gastric	Smooth muscle: Colon and stomach smooth muscle Digestive: Gastric