Summary and discussion

The research described in this thesis was characterized by two major themes: the first was related to the question to which extent variation between individuals in inflammation biomarkers and metabolic syndrome traits are caused by the impact of genetic and environmental differences between people. The second important focus of this thesis went beyond the influence of the DNA sequence and examined epigenetic variation. I addressed the question how important genetic and non-genetic sources of variation are for individual differences in DNA methylation. DNA methylation is an epigenetic mechanism that receives increasing attention as it may provide novel insights into human disease and represents an extra layer causing differences between people. In this chapter, I summarize the most important results and discuss these findings in the broader context of the current state and future directions of research on complex trait genetics.

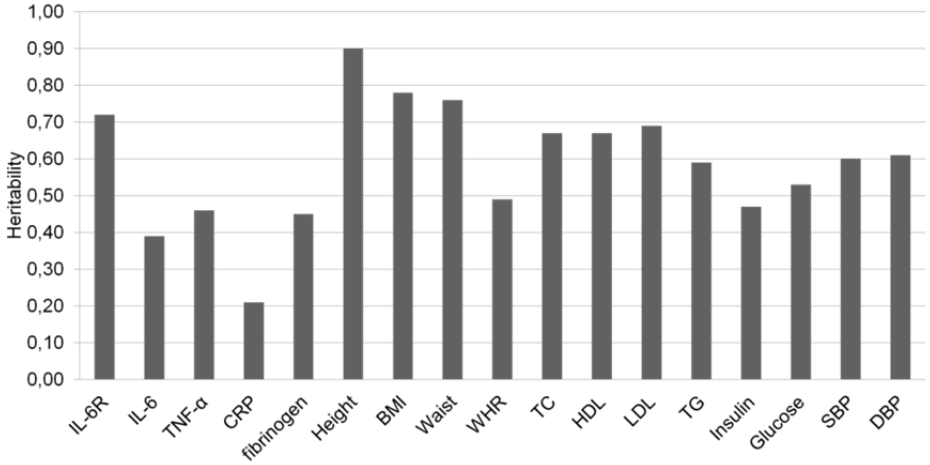
Part 1: Characterizing the genetic architecture of inflammation biomarkers

In the first part of this thesis (chapters 2 and 3), I examined the importance of genetic and environmental influences for individual differences in inflammation biomarkers, including pro-inflammatory cytokines (IL-6 and TNF-α), a soluble cytokine receptor (sIL-6R) and acute-phase proteins fibrinogen and CRP. Extended twin-family models showed that variation in all of these biomarkers of inflammation is to an important extent explained by genetic variation (figure 1). Moderate heritabilities were found for the concentrations of IL-6 (H^2=21%), TNF- α (H^2=39%), CRP (H^2=45%), and fibrinogen (H^2=46%). The levels of soluble IL-6 receptor levels in blood were highly heritable (H^2=72%). The heritability reflects the overall proportion of variation of a trait in the population that can be attributed to genetic variation, but the value of this statistic does not give insight in the number of genes involved, or in the molecular pathways that give rise to the inheritance of traits.

The importance of currently identified genetic variants

The contribution of particular genetic variants to the heritability of a trait can be examined if classical estimation of heritability is combined with the analysis of measured DNA-sequence variants. In chapter 3, I applied this combination of methods to the concentration of sIL-6R in blood. The variance of sIL-6R was largely explained by a single SNP in the IL6R gene that influences the production of sIL-6R (rs2228145, total variance explained=51%; 71% of the total heritability). Through linkage analysis, we found evidence that the remaining heritability is mostly attributable to other genetic variants within the

Figure 1: Heritability of inflammation biomarkers and metabolic syndrome traits.



Heritability=Broad-sense heritability, WHR=Waist-to-hip-ratio, TC=Total cholesterol, TG=Triglycerides, SBP=Systolic blood pressure, DBP=diastolic blood pressure.

IL6R gene region on chromosome 1 and detected novel SNPs at the 3'end that were associated with IL6R gene expression. Of the inflammation biomarkers examined in chapter 2, GWA studies have thus far been published for CRP, fibrinogen and IL-6. A genetic risk score based on 18 genome-wide significant SNPs for CRP explained 5% of the variance in CRP levels[1]. For fibrinogen, four genome-wide significant loci have been reported to date that together explain < 2% of the variance in fibrinogen levels [2]. For IL-6, two loci (*ABO* and *IL6R*) have thus far been reported. The top SNPs in these loci together explain 2.2% of the variance of IL-6 levels [3]. The findings illustrate the difference in the genetic architecture of sIL-6R (for which a single common SNP explains > 50% of the total variance) versus the other inflammation biomarkers. In fact, the variance of sIL-6R levels that can be explained by rs2228145 is very large compared to single SNP effect sizes observed for quantitative traits in general.

The difference between the population explained variance by common SNPs for sIL-6R and IL-6 could be related to the biological consequences of a certain degree of change in concentration of these molecules. Importantly, it is thought that the level of IL-6, and not sIL-6R determines the degree of IL-6 signalling under 'normal conditions'. There is typically a buffer of sIL-6R and soluble gp130 in blood that block the activity of IL-6: the concentrations of sIL-6R and sgp130 (which acts as an antagonist of the active membrane-bound transducer protein gp130) are generally 1000 times higher than the concentration of IL-6 [4]. Importantly, the dynamic range of sIL-6R concentration is more restricted compared to IL-6: during inflammation, sIL-6R levels may rise in the range of 2-3 fold[5]. By contrast, the levels of IL-6 can rise as much as 1 million-fold (under severe conditions)[6]. It may be hypothesized that a genetic

variant with an effect so large that it would explain 50% of the variance of IL-6 levels would have much more serious health consequences compared to the effect of SNP rs2228145, because it would cause serious chronic inflammation in part of the population. Of note, the top SNP for IL-6 may be just as important for disease outcomes as the top SNP of sIL-6R. These inflammation biomarkers exemplify the differences in genetic architecture that may exist between different complex traits (or between different biological pathways that together lead to a disease) and how these differences affect the effort to identify the underlying genes. Even though SNPs identified through GWAS of complex diseases typically explain only a small proportion of the variance, they may point at biological pathways that provide novel insight into disease mechanisms. For example, based on the finding that the *IL6R* SNP is associated with coronary heart disease, it was suggested that tocilizumab, a monoclonal antibody against IL-6R that was already used for treatment of rheumatoid arthritis, should be tested in randomised trials for prevention of coronary heart disease [7]. SNPs in *IL6R* are also associated with asthma risk, and trials investigating tocilizumab for treatment of asthma are currently underway in Australia[8, 9]. It is expected that GWAS findings will lead to novel drug targets for other complex diseases as well.

Part 2 Genetic and environmental influences on BMI and other metabolic syndrome traits

Heritability

In the second part of this thesis, the extended twin-family design was applied to examine the heritability of height and individual metabolic syndrome traits including BMI, waist circumference, waist-to-hip-ratio, metabolic biomarkers and blood pressure (chapter 4, figure 1). This study showed that height, BMI and waist circumference were the most heritable traits (H^2=0.90, H^2=0.78, and H^2=0.76, respectively). Lipids and blood pressure were also highly heritable (H^2=0.59- H^2=0.69), and WHR, insulin and glucose levels were moderately heritable (H^2=0.49, H^2= 0.47, H^2= 0.53, respectively).

An important finding of chapter 4 was that all metabolic syndrome traits were influenced by non-additive genetic effects. For example, the total (broad) heritability of BMI was 78%, with 41% of the variance explained by additive genetic effects and 37% of the variance explained by non-additive genetic effects. Possible explanations for the observed non-additive genetic effects may include the effects of dominant alleles, epistasis, or any other type of genetic effect not acting in an additive manner, causing lower phenotypic similarity between parents and offspring compared to the similarity of DZ twins and siblings. We examined qualitative age differences as a possible alternative explanation for the observed non-additive genetic variance, however, we found no significant qualitative genetic differences between age groups (i.e. different genes contributing to a trait at different ages) for any trait. We did find that the total heritability was slightly lower in older subjects for several traits and this

was mainly related to an increase in unique environmental variance at higher ages. The non-additive genetic effects in our study could also potentially involve genetic effects that depend on environmental influences that are shared to a greater extent between siblings and twins than between children and their parents (gene X (sibling-shared) environment interaction). For example, lifestyle conditions have changed between the time when the parents in our study grew up and the time when their offspring grew up[10]. If the impact of genetic influences on metabolic syndrome traits depends on environmental exposures (e.g. caloric intake) that have changed between generations, such effects may be to a greater extent shared between twins and sibs than between parents and offspring. A final possible explanation is that parents and offspring have been equally exposed to relevant environmental influences during their lives, but that the impact of these exposures depends on genotype plus developmental stage, age, or length of the exposure (the latter three are similar among siblings but differ between parents and children).

In chapter 5, the influence of genes and environment on BMI was studied from a different angle by examining longitudinal BMI discordance in MZ pairs who participated in NTR studies between 1991 and 2011. The most important finding was that large BMI discordance (within-pair BMI difference > 3 kg/m^2), especially long-term discordance, in MZ twin pairs is rare, although discordance became more frequent at later NTR surveys - as the mean BMI and age of twin pairs increased (mean age at survey 1=17.2, SD=2.4, mean age at survey 8=34.6, SD=15.0). We observed that of the MZ twins who became discordant at some point in their lives, most converged to the same weight quickly. These findings illustrated that the heritability of BMI, at least in part, reflects that people tend to have a certain set-point of BMI that is strongly genetically determined, and that this set-point may diffuse after substantial weight gain. Another important finding of this study was that BMI discordant twins show clinically relevant differences in metabolic and inflammation biomarkers. These differences were not present in MZ twins who developed BMI discordance later in their life, illustrating that the level of these biomarkers cannot predict whether a person will develop a high BMI but rather the levels of these biomarkers are changed in response to changes in BMI. The discordant MZ twin design will continue to provide insight into the causes of obesity-associated disease in the future. The EUroDiscoTwin consortium plans to characterize obesity- concordant twins with differences in metabolic biomarkers, including Dutch twins from the NTR, to gain insight into the causes of 'metabolically healthy' versus 'unhealthy obesity'.

Current state and future directions of genetic research on complex (metabolic) traits

Large genome-wide association meta-analyses for the traits studied in chapters 6 and 7 have identified a large number of loci to date [11-14]. For example, 32 loci have been published for BMI to date, which explain 1.45% of the variance of

BMI [13], and 164 loci have been identified in the largest to date meta-analysis of BMI based on 320,485 individuals (manuscript submitted[15]). Importantly, as I explained earlier in this chapter, a small proportion of variance explained by a variant in the population certainly does not mean that the pathway affected by this variant is not that important for the phenotype or disease, although extremely large samples may be required to identify all common genetic variation through GWAS [13]. Of note, for adult height, as much as 697 common variants in 423 loci have been identified with genome-wide significance to date[16], explaining one-fifth of the heritability of height (16% of the total variance). The height paper illustrates that larger samples continue to provide novel biological insights for the trait by pointing at novel relevant genes, and that phenotypic variation is related to multiple variants in the same gene or pathway, each with small individual effects[16].

While genome-wide significant SNPs from GWAS point at locations in the genome where something is happening, the molecular action of these SNPs is often not understood. The majority of currently identified SNPS associated with complex diseases are located outside gene-coding regions[17], and the pathways through which the risk allele influences disease risk is often unknown. This issue is exemplified by the *FTO* locus, which harbours the strongest genetic associations with BMI and obesity [13, 18]. The SNPs are located in intron 1 and 2 of the *FTO* gene. The biological relationship between this SNP and BMI was thought for years to be related to the expression of *FTO*, but a recent study demonstrated that the obesity-associated SNPs are associated with the expression of the homeobox gene *IRX3*, not *FTO* in human brain tissue[19]. Thus, the obesity-associated region within FTO contains enhancers that form long-range interactions with the promoter of *IRX3*. It has become clear that studies that provide insight into how DNA sequence and gene regulation are connected may be crucial to our understanding of complex diseases. The ENCODE project has greatly enhanced our knowledge of the regulatory regions encoded in the human genome by large-scale mapping of functional elements such as promoters, enhancers and open chromatin regions[20]. Further insight may come from studies that map genetic variants that impact on epigenetic regulation. For example, methylation QTL analysis aims to identify DNA sequence variants associated with variation in DNA methylation, and is expected to provide insight into the mechanisms affected by complex disease-associated SNPS that are identified through GWAS [21, 22]. The results of our study from chapter 6 in part 3 of my thesis give insight in the extent to which DNA methylation targeted by the now widely used Illumina 450k array shows variation due to common genotyped SNPs.

Part 3 Beyond DNA sequence: Epigenetic variation

A growing body of evidence highlights the importance of epigenetic regulation for the phenotypes examined in part 1 and 2 of my thesis (see for example [23-29]). DNA methylation is one of such epigenetic mechanisms, which may

mediate DNA sequence effects on complex traits as well as the effects of environmental exposures. In part 3 of this thesis, I examined the causes of variation between people in DNA methylation across the genome. In chapter 6, I describe that methylation at many CpG sites in blood is heritable and sites where a large proportion of variance is due to non-genetic influences (including environment, stochastic variation and measurement error) are also abundant. Of note, the average genome-wide variance of DNA methylation across individuals is low, because at a large number of CpGs, most individuals have very similar methylation levels. I also extended current knowledge on the heritability of DNA methylation with estimates of the variance explained by common genome-wide SNPs. These estimates of 'SNP heritability' suggested that DNA methylation level at a large number of CpGs measured by the 450k array displays genetic variation that is tagged by common genotyped SNPs, which provides guidance for mQTL studies and suggests that the 450k array will provide insight into disease-associated SNPs that act through methylation. In addition, I identified a large number of CpGs where the heritability of DNA methylation decreased with age, a small subset where the heritability increased with age, and a number of CpGs where the heritability of DNA methylation differed between males and females. The overall (twin-based) heritability and 'SNP' heritability of DNA methylation were on average 22 % and 7%, respectively, across genome-wide CpGs.

In chapter 7, I examined MZ twin correlations for DNA methylation measured using the 450k array in buccal samples. In many studies, for example those that focus on development and childhood traits and disorders, buccal is an attractive biological sample to study. I found that the MZ twin correlation is high at a number of CpGs sites in the genome (average correlation=0.31 across all CpGs, and average correlation=0.54 at sites where methylation showed the largest variance across subjects). In comparison, the average MZ twin correlation for DNA methylation level in blood was 0.20 across all genome-wide sites, and 0.57 for sites with the largest between-individual variance in blood. These findings illustrate that DNA methylation variation in buccal cells, similar to our finding in blood, is likely to be influenced by genetic variation. Of note, in chapter 7 I also found that the MZ twin correlation for DNA methylation level shows variation between genomic regions: Regions with low CpG density showed lower MZ twin correlations, suggesting that the variation in DNA methylation in these regions to a larger extent reflects non-genetic sources including environmental and stochastic influences. Although most studies of DNA methylation are currently performed using blood samples, our results from chapter 7 suggest that buccal samples are also suitable for obtaining genome-wide DNA methylation data.

Epigenetic studies and tissue

Large-scale epigenetic studies in living humans are only possible if DNA is obtained from easily accessible tissues. In chapter 6, I examined DNA

methylation in blood samples and in chapter 7 I studied DNA methylation levels based on DNA extracted from buccal swabs. Buccal samples are easier to collect than blood samples, particularly in (young) children. It has also been suggested that buccal-derived DNA may be better suited compared to blood-derived DNA for studying epigenetics in connection to behavioural/psychiatric traits [30]. A recent study that compared methylation measured with the 450k array based on DNA from human post-mortem brain tissues from four different regions, blood and saliva samples found that saliva samples, in particular those containing the largest proportion of buccal epithelial cells showed an overall genome-wide methylation pattern that was closest to the brain tissues [31]. This greater similarity may relate to the fact that buccal cells and brain cells are both derived from the ectodermal layer, whereas blood cells are derived from the mesodermal layer. Nevertheless, there are also indications that DNA methylation levels at relevant candidate genes in blood may respond in a similar way to relevant exposures compared to methylation in the relevant disease tissue: Allele-specific methylation of the *FKBP5* gene (which has been implicated as a mediator of the effects of childhood traumatic life events on stress-related psychiatric disorders in adulthood) responds similarly in blood and brain cells to stimulation by glucocorticoids [32]. Although well-suited for DNA methylation studies, a potential drawback of buccal samples is that, because they contain a relatively small number of cells and thus a small amount of proteins, they are currently not suited for studies of other epigenetic marks such as histone modifications. In conclusion, it is expected that both blood and buccal will contribute to insights in the role of DNA methylation in human complex traits. In the nearby future, DNA methylation will be measured in buccal samples from twins who participate in the NTR as part of an FP7 EU-funded ACTION project on aggression in children.

Part 4: Twin studies and complex traits: Future and further considerations

While GWASs for complex diseases and traits such as type II diabetes, insulin response and triglyceride levels were an instant success (see for example [33-35]), early GWASs for common mental disorders were initially less successful, however, these studies were conducted in much smaller cohorts. For example, early GWASs of schizophrenia did not find genome-wide significant SNPs [36-38]. The finding that SNPs identified through GWAS explained only a small proportion of the total variance fuelled the discussion about the “missing heritability”[39], which was at the time an important inspiration for chapter 8, in which I reviewed evolutionary perspectives on schizophrenia. It had been postulated that the difficulty to find genetic variants that explain a substantial part of the risk for psychiatric disorders may be related to the evolutionary history of these disorders [40]. In chapter 8, I reviewed evolutionary perspectives that have been proposed to explain why schizophrenia, one of the most detrimental common mental disorders, persists in the population despite

conferring a significant reproductive disadvantage to patients. Proposed evolutionary explanations include balancing selection, fitness trade-offs, fluctuating environments, sexual selection, mutation-selection balance and genomic conflicts. The evolutionary perspectives of schizophrenia may to some extent also apply to other common (psychiatric) disorders. In July 2014, a paper came out that reported 108 loci for schizophrenia, which were identified based on 36,989 cases and 113,075 controls and more loci will undoubtedly be found in even larger samples [41]. Yet, the principles discussed in chapter 8 are still relevant as the 108 identified loci explain just 3.4% of the variation of disease liability, and all common SNPs together have been estimated to account for 23% of the variation in liability to schizophrenia, while the total heritability of this disease is ~70%.

For decades, the classical twin study has provided insight into the importance of genes and environment to individual differences in human complex traits. In chapter 9, I described the types of questions that can be examined with twin studies and review studies that applied either the classical twin study or the discordant MZ twin design to complex molecular traits such as metabolomics data, gene expression, the microbiome and epigenetics. The chapter illustrates that twin studies will continue to be highly valuable in the future. In chapters 6 and 7, I studied the importance of genes and environment to the variance of methylation at individual CpG sites in the genome. In chapter 9, I mentioned other interesting questions that may be examined in future studies with multivariate twin designs, for example: to what extent is epigenetic regulation and gene expression across genomic regions influenced by shared genetic or environmental factors? And to what degree do common genetic and environmental mechanisms underlie biological variation across different cells and tissues? MZ twins, who are often discordant for complex diseases, are expected to provide novel insights into epigenetic mechanisms involved in complex disease in the future.

On the interpretation of the heritability of DNA methylation

In chapter 9, I highlighted several assumptions of the twin design that are important to (re) evaluate if this design is applied to molecular data, including the assumption that MZ twins share 100% of their genetic material and the assumption that MZ twins share environmental influences to the same extent as DZ twins. Two other important issues for the interpretation of heritability in the light of epigenetic studies are parental effects and the potential existence of trans-generational epigenetic inheritance. The question whether trans-generational epigenetic inheritance exists in humans is perhaps one of the hottest topics these days in human epigenetic research. Although the answer to this question is currently unknown, a non-trivial question is how parental and trans-generational effects would influence the interpretation of the research described in this thesis, including the heritability of DNA methylation (part 3) and the heritability of phenotypes such as BMI (part 1 and 2).

Parental effects on epigenetic regulation imply that an exposure of the parent causes an epigenetic change in the germ line that is transmitted to the offspring or that maternal exposure during pregnancy causes an epigenetic change in the embryo. These epigenetic changes in turn may cause phenotypic changes in the offspring. There is increasing evidence for the existence of such effects in mammals. For example, several studies have demonstrated that in-utero exposure of rodents to either maternal low protein diet or maternal high-fat diet causes an increased risk of obesity and metabolic syndrome in the offspring through changes in epigenetic regulation [42-46], and effects of maternal under-nutrition on epigenetic regulation in offspring have also been demonstrated in humans [47]. In rodents, similar effects have been demonstrated for paternal exposures [48, 49]. Although there is no doubt that these effects exist, an important question that remains to be unravelled is how much they contribute to variation in human phenotypes at the population level. Trans-generational epigenetic inheritance implies that epigenetic marks are transmitted through the germ line from parents to children, from these children to the grand children etc. (also in the absence of in-utero or germ line exposures in each generation, and independently of the DNA sequence). Rodent studies have shown that the phenotypic effects of in-utero exposure to maternal diet may be transmitted at least up to the third generation, depending on the sex of the offspring and the parent-of-origin [50]. Although the mechanisms behind the transmission of these effects are not well understood, a recent study of paternal prediabetes in rats that was transmitted to offspring revealed alterations in the fathers sperm methylome [51] and transmission of traits to offspring after paternal exposure have also been linked to microRNAs in sperm [52]. Parental and grandparental effects have also been observed in human epidemiological data. For example, effects of exposure to famine of paternal grandfathers is still evident in the grand-children, who have an increased risk of diabetes, cardiovascular disease and mortality [53, 54]. Yet, whether trans-generational inheritance of acquired epigenetic marks occurs in humans remains currently unknown.

If the classical twin model is applied to estimate the sources of variation underlying DNA methylation level, the effects of parental exposures on DNA methylation level that act independently of genotype are expected to end up in the component referred to as “the common environment”, because such effects are equally shared between members of both types of twins. However, if the effect of parental exposures (or other shared environmental influences) on DNA methylation depends on genotype of the offspring, then these effects will manifest as part of the genetic variance. Also, if the if the parental exposure affects DNA methylation in an allele-specific manner and this allele-specific methylation pattern is maintained in the offspring, methylation variation will be statistically correlated with genetic variation (even though genetic variation in this case is not the primary cause of the variation in methylation); such effects will be included in the heritability of DNA methylation. Examples of allele-

specific exposure effects, where a certain exposure alters only the methylation level of a particular genetic allele, have been published (see for example [32]), but it is unknown how often these effects occur. Allele-specific methylation implies that individuals have intermediate methylation levels and the fact that not that many CpGs show intermediate methylation levels suggests that allele-specific exposure effects do not have a great impact on DNA methylation at the genome-wide level, although it may affect a proportion of CpGs in the genome. A special but abundant type of allele-specific methylation occurs when the CpG site itself is a genetic variant (e.g. the C or G may 'contain' a SNP; only the "CG allele" has the potential to become methylated) [55]. These CpG-SNPs also have implications for GWA studies of complex traits: if disease risk depends on the methylation status of the "CG" allele, associations of such SNPs with complex phenotypes are statistically weak if methylation status is not taken into account. Such mechanisms may be identified by integrated analysis of DNA sequence variants and epigenetic variation [56, 57].

If DNA methylation is indeed influenced by trans-generational inheritance of epigenetic mechanisms, variation in DNA methylation attributable to this mechanism of inheritance is only included in the heritability of DNA methylation if transmitted in an allele-specific fashion. The existence of allele-specific trans-generational inheritance of DNA methylation is well-established in plants. For example, fruit-ripening in tomato plants is affected by an "epi-allele" involving methylation of a locus that shows a Mendelian inheritance pattern despite not being related to a nucleotide sequence difference [58]. Similar to human MZ twins, isogenic *Arabidopsis* lines can show substantial variation in DNA methylation [59]. Some differentially methylated regions that reside in isogenic lines (called epigenetic quantitative trait loci) account for substantial parts (60% to 90%) of the heritability of the complex traits flowering time and primary root length [59]. Although these findings by no means imply that similar mechanisms occur in humans (because in humans, DNA methylation is largely erased in the germ line and in the early embryo, whereas DNA methylation is not erased between generations in plants), the findings from plants illustrate the implications for the heritability of complex traits if trans-generational epigenetic inheritance (independent of the DNA sequence) would exist in humans.

Partitioning complex trait variation into genetic and epigenetic effects

It is clear that individual differences in complex traits are related to genetic variation and epigenetic variation, and that these components are connected to each other in complex ways. While it was already possible to estimate the proportion of variation in complex traits due to measured genetic variation (as illustrated in chapter 3), it is currently unknown how much of the variation in complex traits is related to epigenetic variation. To explore this question, my colleague Michel Nivard and I applied a novel method based on genome-wide SNP data and genome-wide blood methylation data, to simultaneously estimate the proportion of variation in BMI explained by SNPs and the proportion of

variation explained by DNA methylation in blood, while accounting for the correlation between the two. Preliminary results of this analysis indicated that 42% of the variation in BMI is explained by common genotyped SNPs, 39% is explained by DNA methylation variation in blood and the correlation between the effect of genome-wide SNPs on BMI and methylation variation associated with BMI was 0.36. Importantly, while the association between SNPs and the trait is inherently causal (the level of a trait cannot affect the genotype), the association between DNA methylation and the trait is more complex: this variance component may include effects of DNA methylation on the trait, effects of the trait on DNA methylation, and effects of a common underlying mechanisms on the trait and DNA methylation (with the exception of shared effects of common genotyped SNPs as these are included in the correlation of 0.36). Longitudinal phenotypic and epigenetic data will be highly valuable to distinguish between epigenetic variation that is induced by the phenotype/disease state and epigenetic variation that is causally related to the trait. In the future, the novel method may be applied to other complex traits to gain insight into the degree to which variation is related to DNA methylation variation in a target tissue, and into the extent to which genetic and epigenetic effects related to the trait of interest are correlated.

Conclusions

In conclusion, I found that inflammation biomarkers and metabolic syndrome traits are characterized by significant heritability. These findings illustrate that individual differences in the vulnerability to develop metabolic disease are to a large extent explained by differences in genetic susceptibility. Many loci have been identified where genetic or epigenetic variation is connected to these traits. I found that DNA methylation is itself to an important extent heritable, but environmental influences also account for a large part of the variation. Interestingly, I found that at a substantial number of CpGs in the genome the variation in DNA methylation due to environmental or stochastic influences increases with age. I also found that a substantial part of the heritability of DNA methylation can be explained by common SNPs, and that that blood and buccal samples are both informative to inter-individual differences in methylation due to genetic and environmental influences, although they may not hold the same information with respect to epigenetic variation involved in complex traits. The extent to which the phenotypes examined in this thesis (i.e. inflammation biomarkers and metabolic syndrome traits) are influenced by epigenetic variation, possibly in interaction with the DNA sequence and environment, awaits further examination. Future studies including Epigenome-Wide Association Studies (EWAS), and studies that integrate epigenetic and genetic information will provide further insight into the epigenetic mechanisms involved in complex traits and into the interplay of genetic and epigenetic variation within and across generations.

Reference List

1. Dehghan,A. *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731-738 (2011).
2. Dehghan,A. *et al.* Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ. Cardiovasc. Genet.* **2**, 125-133 (2009).
3. Naitza,S. *et al.* A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS. Genet.* **8**, e1002480 (2012).
4. Rose-John,S. IL-6 trans-signaling via the soluble IL-6 receptor: importance for the pro-inflammatory activities of IL-6. *Int. J. Biol. Sci.* **8**, 1237-1247 (2012).
5. Scheller,J., Garbers,C., & Rose-John,S. Interleukin-6: From basic biology to selective blockade of pro-inflammatory activities. *Semin. Immunol.*(2013).
6. Waage,A., Brandtzaeg,P., Halstensen,A., Kierulf,P., & Espevik,T. The complex pattern of cytokines in serum from patients with meningococcal septic shock. Association between interleukin 6, interleukin 1, and fatal outcome. *J. Exp. Med.* **169**, 333-338 (1989).
7. IL6R Genetics Consortium and Emerging Risk Factors Collaboration. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet* **379**, 1205-1213 (2012).
8. Ferreira,M.A. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006-1014 (2011).
9. Revez,J.A. *et al.* A new regulatory variant in the interleukin-6 receptor gene associates with asthma risk. *Genes Immun.* **14**, 441-446 (2013).
10. Food and Agriculture Organization of the United Nations. World agriculture: towards 2015/2030 (Summary report). 2012. Rome.
11. Dupuis,J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105-116 (2010).
12. Global Lipids Genetics Consortium Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274-1283 (2013).
13. Speliotes,E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937-948 (2010).
14. Levy,D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677-687 (2009).
15. GIANT consortium. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study (manuscript submitted). 2014.
16. Wood,A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173-1186 (2014).
17. Hindorff,L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A* **106**, 9362-9367 (2009).
18. Frayling,T.M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
19. Smemo,S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-375 (2014).

20. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
21. Bell,J.T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
22. Gibbs,J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS. Genet.* **6**, e1000952 (2010).
23. Dayeh,T. *et al.* Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS. Genet.* **10**, e1004160 (2014).
24. Dick,K.J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990-1998 (2014).
25. Nilsson,E. *et al.* Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* **63**, 2962-2976 (2014).
26. Sinclair,K.D. *et al.* DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proc. Natl. Acad. Sci. U. S. A* **104**, 19351-19356 (2007).
27. Stenvinkel,P. *et al.* Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease? *J. Intern. Med.* **261**, 488-499 (2007).
28. Agarwal,S. & Rao,A. Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. *Immunity.* **9**, 765-775 (1998).
29. Plagemann,A. *et al.* Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J. Physiol* **587**, 4963-4976 (2009).
30. Lowe,R. *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics.* **8**, 445-454 (2013).
31. Smith,A.K. *et al.* DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*(2014).
32. Klengel,T. *et al.* Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat. Neurosci.* **16**, 33-41 (2013).
33. Saxena,R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-1336 (2007).
34. Scott,L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-1345 (2007).
35. Steinthorsdottir,V. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770-775 (2007).
36. Lencz,T. *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol. Psychiatry* **12**, 572-580 (2007).
37. O'Donovan,M.C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053-1055 (2008).
38. Sullivan,P.F. *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol. Psychiatry* **13**, 570-584 (2008).
39. Maher,B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21 (2008).
40. Uher,R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol. Psychiatry* **14**, 1072-1082 (2009).

41. Schizophrenia Working Group of the Psychiatric Genomics Consortium Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
42. Guo,F. & Jen,K.L. High-fat feeding during pregnancy and lactation affects offspring metabolism in rats. *Physiol Behav.* **57**, 681-686 (1995).
43. Khan,I.Y. *et al.* A high-fat diet during rat pregnancy or suckling induces cardiovascular dysfunction in adult offspring. *Am. J. Physiol Regul. Integr. Comp Physiol* **288**, R127-R133 (2005).
44. Lillycrop,K.A., Phillips,E.S., Jackson,A.A., Hanson,M.A., & Burdge,G.C. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J. Nutr.* **135**, 1382-1386 (2005).
45. Vucetic,Z., Kimmel,J., Totoki,K., Hollenbeck,E., & Reyes,T.M. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology* **151**, 4756-4764 (2010).
46. Zambrano,E. *et al.* Sex differences in transgenerational alterations of growth and metabolism in progeny (F2) of female offspring (F1) of rats fed a low protein diet during pregnancy and lactation. *J. Physiol* **566**, 225-236 (2005).
47. Heijmans,B.T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U. S. A* **105**, 17046-17049 (2008).
48. Carone,B.R. *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-1096 (2010).
49. Ng,S.F. *et al.* Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* **467**, 963-966 (2010).
50. Dunn,G.A. & Bale,T.L. Maternal high-fat diet effects on third-generation female body size via the paternal lineage. *Endocrinology* **152**, 2228-2236 (2011).
51. Wei,Y. *et al.* Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc. Natl. Acad. Sci. U. S. A* **111**, 1873-1878 (2014).
52. Gapp,K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.* **17**, 667-669 (2014).
53. Kaati,G., Bygren,L.O., & Edvinsson,S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur. J. Hum. Genet.* **10**, 682-688 (2002).
54. Pembrey,M.E. *et al.* Sex-specific, male-line transgenerational responses in humans. *Eur. J. Hum. Genet.* **14**, 159-166 (2006).
55. Shoemaker,R., Deng,J., Wang,W., & Zhang,K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20**, 883-889 (2010).
56. Guintivano,J. *et al.* Identification and Replication of a Combined Epigenetic and Genetic Biomarker Predicting Suicide and Suicidal Behaviors. *Am. J. Psychiatry*(2014).
57. Reynard,L.N., Bui,C., Syddall,C.M., & Loughlin,J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum. Genet.* **133**, 1059-1073 (2014).
58. Manning,K. *et al.* A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* **38**, 948-952 (2006).
59. Cortijo,S. *et al.* Mapping the epigenetic basis of complex traits. *Science* **343**, 1145-1148 (2014).